

Mari Buseth Helland

Validering av modellerte forsenkninger i Trondheim ved bruk av folkeinnsamlet data om overvann: en GIS-studie

Masteroppgave i Geografi

Veileder: Jan Ketil Rød

Mai 2022

Mari Buseth Helland

Validering av modellerte forsenkninger i Trondheim ved bruk av folkeinnsamlet data om overvann: en GIS-studie

Masteroppgave i Geografi
Veileder: Jan Ketil Rød
Mai 2022

Norges teknisk-naturvitenskapelige universitet
Fakultet for samfunns- og utdanningsvitenskap
Institutt for geografi



Kunnskap for en bedre verden

Sammendrag

Klimaendringene vil føre til flere og mer intense nedbørshendelser, noe som gjør at overvannsproblematikk blir mer aktuelt, spesielt i urbane områder hvor skadepotensialet er stort. Urban flom-modellering baserer seg på en forenkling av urbane dreneringsforhold, noe som skaper usikkerheter rundt resultatenes troverdighet. Data samlet inn av befolkning uten forskningsbakgrunn kan generere data for store studieområder på kort tid, og har i tidligere studier vist å kunne validere GIS-modellerte forsøkninger.

Denne masteroppgaven diskuterer derfor om folkeinnsamlet data om overvann fra prosjektet *Citizen Sensing* kan brukes til å bekrefte at modellerte forsøkninger i GIS er reelle vannoppsamlinger. Oppgaven beskriver kjennetegn ved det folkeinnsamlede datasettet, og diskuterer ut ifra disse om datasettet er egnet til sitt formål. I tillegg reflekterer oppgaven rundt mulige feilkilder knyttet til metodene, både modelleringen av forsøkningene og sorteringen av folkeinnsamlet data.

Resultatene viser at det folkeinnsamlede datasettet om overvann er lite egnet til å validere om modellerte forsøkninger er reelle vannoppsamling. Dette fordi datasettet inneholder en stor andel feilkilder som gjør at datasettet ikke vil gi troverdige resultat uten en omfattende sortering. De største feilkildene knyttes til brukerfeil som fører til feilregistreringer og lav posisjonsnøyaktighet. Overregistrering av observasjoner blant enkelte deltakere, samt overrepresentasjon i enkelte deler av studieområdet gjør også at datasettet vurderes som lite egnet. Sluttresultatet viser at så mye som 92 prosent av det opprinnelige datasettet ikke kan brukes som valideringsgrunnlag på grunn av både tydelige og potensielle feilkilder.

Abstract

Climate change will lead to even heavier and more intense precipitation events, which makes issues concerning surface water more relevant, especially in urban areas where the potential damage is considered to be high. Urban flood-modeling is based on a simplification of urban drainage conditions, which pose uncertainties about the credibility of the results. Observations collected by citizens can generate data for larger study areas in short periods of time, and previous studies have shown that such data are able to validate modeled depressions.

This master thesis therefore discusses whether data collected by citizens regarding surface water during the project *Citizen Sensing* can be used to confirm that modeled depressions in GIS are accumulated surface water. Based on the characteristics of the observations, the thesis discusses whether the data is suitable for its purpose. In addition, the thesis reflects on possible errors related to the methods used, both the modeling of the depressions and the sorting of observations.

The results conclude that the dataset collected by citizens is unsuitable to validate whether modeled depressions are accumulated surface water. This is because the dataset contains different sources of error which makes it hard to trust the final results without comprehensive sorting to eliminate these. The largest sources of error are related to user errors that lead to observations being wrongly submitted, as well as the low position accuracy of the observations. Over-registration of observations among some participants, as well as over-representation in some parts of the study area leads to the conclusion that the data set is considered unsuitable. The end result shows that as much as 92 percent of the original dataset cannot be used as a validation basis due to both clear and potential sources of error.

Forord

Masteroppgaven markerer slutten på fem fine og interessante år som student ved NTNU. Tiden har gått fort, og det er både godt og vemodig å sitte her og skrive forordet.

Jeg vil rette en stor takk til min veileder Jan Ketil Rød ved institutt for geografi ved NTNU for god veiledning gjennom hele arbeidet med masteroppgaven. Du har vært en god støtte helt fra de første diskusjonene om aktuell problemstilling, avgrensning av oppgaven og fram til innlevering. Vil også rette en takk til Tomasz Opach for data fra prosjektet *Citizen Sensing* og for at du stilte opp når jeg hadde spørsmål angående prosjektet.

Takk til mamma for mange timer med korrekturlesing.

Takk til Maia og Thale for støtte i masterskrivingens mange frustrasjoner, men også i alle oppturer. Takk til gjengen på kontoret for hyggelige (og litt for mange) pauser i løpet av året. Dere har gjort det morsomt å være student!

Til slutt vil jeg takke den harde kjernen på GIS-labben for mange sene kvelder i masterinnspurten, det hadde ikke vært det samme uten dere ved min side.

Mari Buseth Helland

Innhold

Figurer.....	xv
Tabeller	xvii
Forkortelser.....	xvii
1. 1	Innledning
.....	1
1.1 Problemstilling	2
2. 2	Bakgrunn
.....	3
2.1 Overvann som konsekvens av klimaendringer og fortetting.....	3
2.2 Studieområde.....	4
2.3 Citizen Sensing.....	5
2.4 Tidligere studier	7
2.4.1 Utlandet.....	7
2.4.2 Norge.....	8
2.4.3 I studieområdet.....	8
2.5 Hva skiller denne masteroppgaven fra tidligere studier?	10
3. 3	Teori
.....	11
3.1 GIS for å representere den virkelige verden.....	11
3.2 Vektor- og rasterdata	11
3.3 Digital terrengmodell (DTM).....	12
3.4 Hvordan representeres overvann i GIS?.....	13
3.4.1 Forskningsanalyse	14
3.5 Folkeinnsamlet data.....	15
3.5.1 Volunteered geographic information (VGI)	15
3.5.2 Citizen science	15
3.5.3 Fordeler og ulemper med citizen science.....	16
3.6 Posisjonering	18
3.7 Flerveisinterferens	19
3.8 Nøyaktighet og presisjon.....	19
4. 4	Metode
.....	21
4.1 ArcGIS Pro	21

4.2	Dataforberedelse – modellering av forsenkninger	21
4.2.1	Vektordata.....	21
4.2.2	Rasterisering	21
4.3	Digital terrengmodell (DTM).....	22
4.3.1	Forskningsanalyse i urbane strøk.....	22
4.3.2	Hvilken oppløsning egner seg?.....	22
4.3.3	Valg av oppløsning	23
4.4	Bluespot-modell	23
4.5	Arbeidsflyt for modellering av forsenkninger.....	24
4.6	Forskningsanalyse – første del	24
4.6.1	Fylle groper.....	25
4.6.2	Vertikal nøyaktighet.....	26
4.6.3	Hvis/eller-evaluering.....	27
4.7	Forskningsanalyse – andre del.....	28
4.8	Forskningsanalyse – tredje del	29
4.9	Folkeinnsamlede observasjoner	31
4.10	Sortering av observasjoner	31
4.11	Grov sortering.....	33
4.11.1	Fjerne observasjoner utenfor det opprinnelige studieområdet	33
4.11.2	Fjerne observasjoner som ikke har hendelse <i>kraftig regn</i>	34
4.11.3	Fjerne observasjoner i bygg	35
4.11.4	Fjerne observasjoner uten bilde.....	35
4.11.5	Fjerne observasjoner som ligger på broer	35
4.12	Detaljert sortering.....	36
4.12.1	Rutenett	36
4.12.2	Fjerne observasjoner der bildet ikke viser vann.....	38
4.12.3	Fjerne observasjoner som ikke viser omgivelser	40
4.12.4	Fjerne observasjoner der omgivelsene er vanskelig å identifisere	41
4.12.5	Fjerne observasjoner med feil posisjon.....	42
4.12.6	Finne vannoppsamlingens reelle posisjon.....	44
4.13	Avstandsmål	45
4.13.1	Manuell avstandsmåling.....	45
4.13.2	Verktøyet <i>near</i>	46

4.13.3	Romlig utvalg basert på avstand og lokasjon.....	47
4.14	Bekreftede vannoppsamlinger.....	47
4.14.1	En meter fra observasjon.....	48
5.	5.....	Resultat
	49
5.1	Bekreftede forsøkninger.....	49
5.1.1	Eksempel på bekreftede forsøkninger.....	51
5.2	Antall observasjoner av bekreftede forsøkninger.....	52
5.2.1	Forsøkninger bekreftet av flest observasjoner.....	54
5.2.2	Forsøkningsområdet med sju observasjoner.....	55
5.3	Sorteringen.....	56
5.4	Observasjoner i studieområdet før sortering, etter grov sortering og etter detaljert sortering.....	57
5.5	Antall observasjoner per deltaker.....	59
5.6	Fordeling av observasjoner i studieområdet.....	60
5.7	Observasjoner med samme koordinat.....	61
5.8	Avstand fra observasjoner til nærmeste forsøkningspunkt.....	63
5.8.1	Grov sortering.....	63
5.8.2	Detaljert sortering.....	64
5.9	Feilplasserte observasjoner.....	65
6.	6.....	Diskusjon
	66
6.1	Bekreftede vannoppsamlinger.....	66
6.2	Folkeinnsamlet data.....	68
6.3	Fordeling av observasjoner.....	69
6.3.1	Ulikt antall observasjoner per deltaker.....	69
6.3.2	Ujevn fordeling av observasjoner i studieområdet.....	71
6.4	Ulik størrelse på vannoppsamlingene.....	72
6.5	Bildene har varierende kvalitet.....	74
6.6	Brukerfeil.....	74
6.6.1	Observasjoner som ikke viser overvann.....	75
6.6.2	Registrering i bygg.....	75
6.6.3	Observasjoner med samme koordinat.....	76
6.7	Posisjonsnøyaktighet.....	76

6.7.1	Nøyaktighet.....	77
6.7.2	Presisjon.....	79
6.8	Små lokasjonsfeil	80
6.8.1	Flerveisinterferens som mulig problem	81
6.9	Store feilplasseringer av observasjoner.....	82
6.10	Refleksjon rundt metoder	83
6.11	Sortering av observasjoner	83
6.11.1	En omfattende sorteringsjobb.....	84
6.11.2	For strenge krav i sorteringen?.....	85
6.11.3	Subjektivitet som mulig feilkilde	86
6.12	Forskningsanalyse.....	87
7.	7..... Avslutning	89
7.1	Konklusjon	89
7.2	Videre studier	90
8.	Referanser	91

Figurer

Figur 1:	Masteropgavens studieområde	4
Figur 2:	Citizen sensing-applikasjonen.....	6
Figur 3:	Modellerte forskninger og registrerte flomobservasjoner (Yu et al., 2016).....	7
Figur 4:	a) GIS-modellerte flomveier og forskninger nær bygning i Trondheim, b) bilde fra befolkningen som viser en flomhendelse i samme bygning (Meiforth, 2013).....	9
Figur 5:	Bygninger representere med vektordata og terrenghøyden representeres med rasterdata	12
Figur 6:	Illustrasjon av hvordan flerveisinterferens kan påvirke GPS-signal i urbane områder (Argust tracking, 2020).	19
Figur 7:	Illustrasjon av begrepene nøyaktighet og presisjon: a) lav nøyaktigheten, lav presisjon, b) lav nøyaktigheten, høy presisjon, c) høy nøyaktighet, lav presisjon og d) høy nøyaktighet, høy presisjon (Esri, 2017)	20
Figur 8:	Første del av forsknings-analysen	24
Figur 9:	Fylling av groper (Esri, 2021a)	25
Figur 10:	Anbefalt vertikal nøyaktighet i følge ASPRS standarden fra 2014 (American Society for Photogrammetry, 2015).....	26
Figur 11:	Verktøyet <i>con</i> utfører en hvis/eller-evaluering av pikslene i studieområdet.....	27

Figur 12: Andre del av forsenkningsanalysen.....	28
Figur 13: Verktøyet minus endrer verdien til piksler som samlokaliserer med bygg til 0.....	28
Figur 14: Verktøyet minus endrer verdien til piksler lokalisert i Nidelva til 0.....	29
Figur 15: Tredje del av forsenkningsanalysen	29
Figur 16: Pikslene blir tildelt en felles gridcode basert på deres romlige tilknytning	30
Figur 17: Piksler med lik gridcode aggregeres sammen til ett sammenhengende polygon	30
Figur 18: Det opprinnelige studieområdet i Citizen Sensing-prosjektet.....	34
Figur 19: Støy fra biler på a) Elgeseter bro og b) Bakke bro	35
Figur 20: To rutenett med ulik størrelse ble brukt til sortering av observasjoner.....	37
Figur 21: Bilder som ikke viste overvann: a) tett sluk, b) stor ansamling løv og c) ødelagt asfalt	38
Figur 22: Vannoppsamlinger av ulik størrelse	39
Figur 23: Den samme vannoppsamlingen observert av flere deltakere på ulike tidspunkt viser forskjell i størrelse.....	39
Figur 24: Bilder fra deltakere som ikke viste omgivelser	40
Figur 25: Bilder fra deltakere der a) bilder var for mørkt til at omgivelsene kunne identifiseres og b) bildet viste en byggeplass som ikke lenger var gjenkjennbar.....	41
Figur 26: Tre observasjoner av den samme vannoppsamlingen har noe ulikt koordinat.....	42
Figur 27: Kartet viser polygoner som ble brukt til å markere observasjoner av samme vannoppsamling og knytte observasjoner med feil posisjon til vannoppsamlingens reelle posisjon.....	43
Figur 28: Vannoppsamlingens reelle posisjon ble bestemt ved bruk av a) observasjoner i ArcGIS Pro, b) bilde fra deltakerne og c) Google Street View.....	44
Figur 29: Manuell avstandsmåling ble gjort mellom observasjoner og vannoppsamlingens reelle posisjon.....	46
Figur 30: I verktøyet <i>generate near table</i> måles avstanden fra inngangsdataen til det nærmeste spesifiserte objektet	47
Figur 31: Forsenkninger en meter eller nærmere blir bekreftet å være reelle vannoppsamlinger	48
Figur 32: Bekreftede forsenkninger i studieområdet.....	50
Figur 33: Eksempel på bekreftede forsenkninger	51
Figur 34: Kartet viser antall observasjoner av bekreftede forsenkninger	52
Figur 35: Histogrammet viser antall observasjoner av hver bekreftede forsenkning	53
Figur 36: Forsenkninger bekreftet av fem, seks og sju observasjoner	54
Figur 37: Forsenkning med sju observasjoner	55
Figur 38: 2554 observasjoner før sortering	57
Figur 39: 1020 observasjoner etter grov sortering	58
Figur 40: 206 observasjoner etter detaljert sortering	59
Figur 41: Histogrammet viser hvor mange observasjoner som ble registrert av hver deltaker	60
Figur 42: Kartet viser hvilke områder det var registrert flest og færrest observasjoner	61
Figur 43: Kartet viser hvor i studieområdet det ble funnet overlappende observasjoner	62
Figur 44: Histogrammet viser avstanden fra observasjoner til nærmeste forsenkning etter den grove sorteringen. Grønn søyle viser observasjoner som overlapper med forsenkninger	63

Figur 45: Histogrammet viser avstanden fra observasjoner til nærmeste forsenkning etter den detaljerte sorteringen. Grønn søyle viser observasjoner som overlapper med forsenkninger..	64
Figur 46: Histogrammet viser avstanden fra observasjoner til vannoppsamlingens reelle posisjon.....	65
Figur 47: Eksempel på en observasjon som regnes som nøyaktig.....	77
Figur 48: Observasjoner med lav nøyaktighet og lav presisjon hvor a) viser observasjoner som er lite nøyaktige, mens b) viser observasjoner som er svært lite nøyaktige.....	79
Figur 49: a) deltakeren står omtrent en meter unna vannoppsamlingen og b) deltakeren står omtrent to meter unna vannoppsamlingen	81

Tabeller

Tabell 1: Sortering av folkeinnsamlede observasjoner, delt inn i grov og detaljert sortering .	32
Tabell 2: Tabellen viser hvor mange observasjoner som ble fjernet i hvert steg av sorteringen	56

Forkortelser

VGI	Volunteered Geographic Information
ASPRS	American Society for Photogrammetry and Remote Sensing
GIS	Geografiske informasjonssystemer
DTM	Digital terrengmodell
GPS	Global Positioning System
NTNU	Norges teknisk-naturvitenskapelige universitet
NVE	Norges vassdrags- og energidirektorat
ASPRS	American Society for Photogrammetry and Remote Sensing
LiDAR	Light Detection and Ranging
FKB	Felles kartdatabase
ETRS 1989	European Terrestrial Reference System 1989
UTM	Universal Transverse Mercator

1 Innledning

Urban flom regnes som en av de største klimatrusslene i vår tid (Skaaraas et al., 2015). I en temaplan om klimatilpasning publisert av Trondheim kommune våren 2022 blir det regnet som meget sannsynlig at oversvømmelse som en konsekvens av ekstremnedbør kommer til å inntreffe i Trondheim i nærmeste framtid. Det antas at store mengder overvann har høy risiko for å gjøre skade på materielle verdier, og kommunen mener det er middels risiko for liv og helse (Trondheim kommune, 2022). Det er i byene en finner de største materielle verdiene, og det er også disse områdene som er mest utsatt for store mengder overvann. Klimaendringene gir fremtidsbilder som gjør at det er viktig å kartlegge områder hvor overvannet vil kunne samle seg ved kraftig nedbør.

Geografiske informasjonssystem (GIS) har i flere år blitt hyppig brukt til hydrologisk modellering og for å identifisere forsøkninger der nedbør kan akkumulere. Ved å representere den virkelige verden i GIS kan vi tilegne oss mer kunnskap enn det vi klarer på egen hånd (Longley et al., 2005, s. 65). Informasjon om blant annet urban flom vil ifølge Skaaraas et al. (2015) ha en viktig rolle i klimatilpasningen og arealplanleggingen i årene som kommer. Til tross for dette vil modellene vi bruker kunne inneholde feil da det ofte ikke tas hensyn til infiltrasjonsrate og dreneringsforhold.

I følge Wang et al. (2018) er en av de fremste årsakene til usikkerhet i modellering av urban flom mangelen på valideringsdata. Valideringsmetoder er en viktig del for å sørge for at de modelleringene vi gjør gir troverdige resultat. Ved institutt for geografi ved NTNU har det blitt publisert en rekke masteroppgaver som tar for seg hydrologisk modellering. Typisk for de fleste studiene er at de har brukt en eller annen form for valideringsmetode for å bekrefte at modelleringen gir troverdige resultat i studieområdet. Som eksempel brukte Viréhn (2014) skadepunkt fra en tidligere flomhendelse for å validere GIS-modellerte flomveier i Gudbrandsdalen. Skaderegistreringer samlet inn av eksperter har altså vist seg å kunne bekrefte hydrologisk modellerte flomveier.

De siste årene ser en at antall studier hvor folkeinnsamlet data blir integrert i modellering av flom er økende (Assumpção et al., 2018). Fotografier og informasjon samlet inn av befolkning uten forskningsbakgrunn har vist seg å være effektive til å validere modellering av flom (Kutija et al., 2014). Folkeinnsamlet data kan gi observasjoner som enten bekrefter eller avkrefter usikkerhetene som oppstår på grunn av den varierende topografien og forskjellene i arealdekke i urbane områder (Helmrich et al., 2021). Folkeinnsamlet data er i tillegg ofte nødvendig for å kunne validere flommodeller som strekker seg over store områder (Alabri & Hunter, 2010). Ved å inkludere befolkningen er det mulig å samle inn store mengder informasjon om overvann i by. Dette åpner dørene for en mer effektiv form for datainnsamling som muligens kan hjelpe oss i klimatilpasningen i årene som kommer.

1.1 Problemstilling

I denne masteroppgaven er programvaren ArcGIS Pro brukt til å modellere forsøkninger i terrenget i den urbane delen av Trondheim. Dette gjør at en kan identifisere områder hvor overvann kan akkumulere ved nedbørsmengder. Modellen er en grov forenkling av urbane avrenningsforhold, noe som fører med seg usikkerhet rundt resultatets troverdighet. På grunn av dette vil folkeinnsamlet data om overvann fra prosjektet *Citizen Sensing* blir brukt til å validere om modellerte forsøkninger faktisk er reelle vannoppsamlinger.

Ved hjelp av GIS-modellering av forsøkninger og bruk av folkeinnsamlet data om overvann, vil jeg i denne masteroppgaven besvare følgende hovedproblemstilling:

Hvor egnet er folkeinnsamlet data om overvann til å validere at GIS-modellerte forsøkninger er reelle vannoppsamlinger?

For å kunne besvare dette har jeg definert en underproblemstilling:

Hva kjennetegner det folkeinnsamlede datasettet om overvann?

2 Bakgrunn

2.1 Overvann som konsekvens av klimaendringer og fortetting

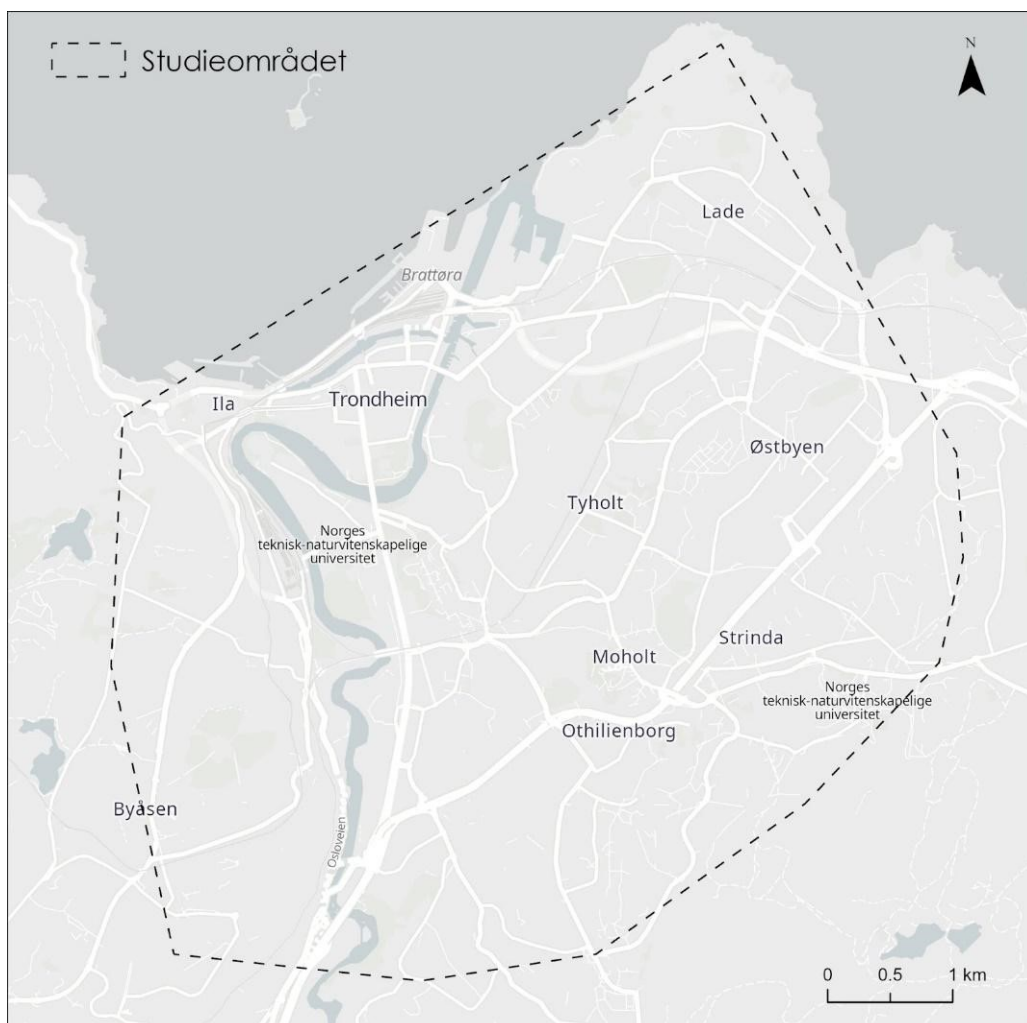
Menneskeskapte klimaendringer vil trolig føre til høyere temperaturer og mer totalnedbør. For Norge anslås det at årsnedbøren vil øke med opptil 18 prosent fram til 2100. I tillegg vil vi oppleve hyppigere perioder med styrtregn, og disse vil være kraftigere og mer uforutsigbare enn tidligere (Hanssen-Bauer et al., 2009).

Vi blir stadig flere innbyggere i Norge og mange av oss vil bo sentralt. Byenes fortettes og dette fører til færre grøntareal og flere asfalterte områder, såkalte impermeable flater. Denne fortettingen er med på å svekke jordas naturlige evne til å ta til seg vann, noe som gjør at mer av nedbøren samles på overflaten. Samtidig ser en at dagens ledningsnett ikke alltid er tilstrekkelig til å lede vekk store mengder nedbør (Skaaraas et al., 2015). Kombinasjonen av økte mengder nedbør over kort tid, fortetting i byene og et ledningsnett som ikke er tilpasset denne utviklingen, gjør at vi kan forvente at store deler av nedbøren samler seg som overvann. Overvann defineres av Skaaraas et al. (2015) som overflateavrenning forårsaket av nedbør og smeltevann. Et endret klima med større mengder nedbør vil trolig føre til flere utfordringer knyttet til overvann og håndteringen av dette, noe vi allerede ser. Trondheim kommune har for eksempel en plan for håndtering av overvann som del av sin temaplan for klimatilpasning (Trondheim kommune, 2022).

I Norge er overvann som oftest forårsaket av kraftige nedbørsperioder eller smeltevann. I tillegg til dette spiller terrengets topografi, bakkens infiltrasjonsevne og dreneringsmulighetene til ledningsnettene en stor rolle for mengden overvann og hvordan vannet oppfører seg (Skaaraas et al., 2015). Overvann vil typisk samle seg i forsenkninger, altså lavpunkt i terrenget (Balstrøm, 2021). Det er intens nedbør i løpet av noen få timer (korttidsnedbør) som forårsaker de fleste skadene i tettbygde strøk (Hanssen-Bauer et al., 2009). Det er i de urbane områdene vi finner de største verdiene, og i Norge fører urban flom til store materielle skader hvert år (Skaaraas et al., 2015). I urbane områder hvor infiltrasjonen av nedbør er minimal, vil dette kunne føre til store mengder overvann på avveie, samt at vann akkumuleres i nye områder.

2.2 Studieområde

Trondheim er Norges tredje største by og ligger i Trøndelag fylke. Nidelva renner gjennom byen og ut i Trondheimsfjorden og de fleste innfartsårene til sentrum går over broer. Studieområdet er urbant og preges av bygninger, infrastruktur og asfalterte områder. Særlig de sentrumsnære delene av byen har tett bebyggelse som gir mange små urbane høydeforskjeller. Selve sentrumskjernen inneholder få grøntarealer, men i ytterkanten av sentrum er det flere rekreasjonsområder, blant annet på Marinen, oppe ved Kristiansten festning og på toppen ved Tyholt. Masteroppgavens studieområde inkluderer sentrumskjernen Midtbyen, men strekker seg også fra Ila, opp mot Byåsen, Stubban, Dragvoll, hele Østbyen og helt ut til Ladehammeren (se figur 1). Ytre del av byen er i all hovedsak villastrøk med eneboliger, rekkehus og flere grøntområder. Disse områdene må likevel beskrives som urbane. De preges av asfalterte flater og en godt utbygd infrastruktur.



Figur 1: Masteroppgavens studieområde

I Trondheim bor det litt over 210 000 innbyggere (Esri, 2017). Av disse er 33 000 studenter, noe som gjør Trondheim til en av Norges største studentbyer. Trondheim er derfor en by hvor vilkårene for folkeinnsamling av data er gode. Her er det potensielt mange bidragsytere både blant den generelle befolkningen og blant studentene.

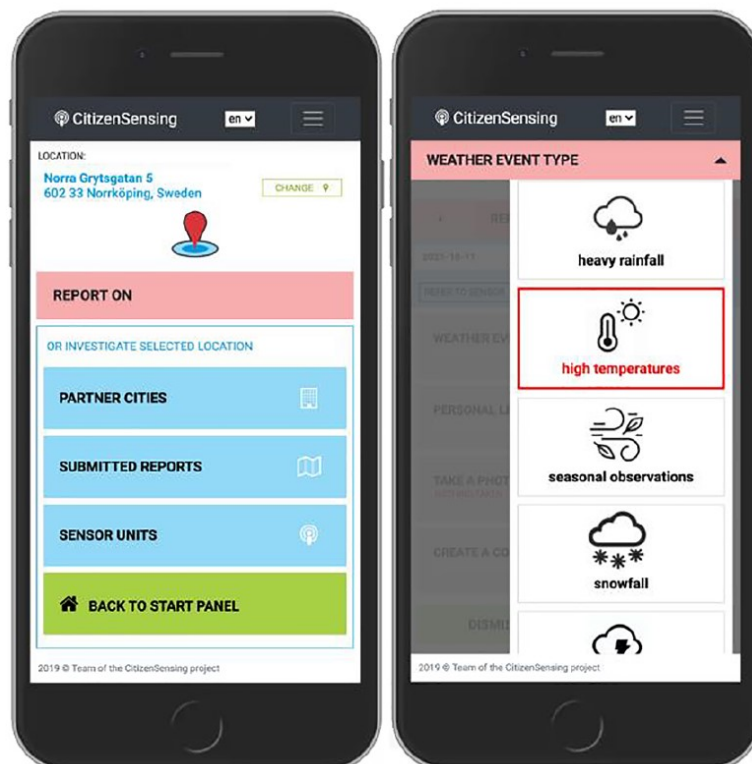
Problemstillingen i denne masteroppgaven er besvart med utgangspunkt i et mindre studieområde i Trondheim, vist i figur 1. Studieområdets grense er definert av sorterte observasjoner gjort i forbindelse med to kampanjer i et prosjekt med folkeinnsamlet data, Citizen Sensing-prosjektet, som ble gjennomført høsten 2019 og høsten 2020. Som oppgavens resultatdel vil vise, er flere av observasjonene lokalisert i områder nært universitetscampus, spesielt på Gløshaugen, Dragvoll og Kalvskinnet. Det er også gjort observasjoner i andre deler av byen, og dette har påvirket avgrensingen av det endelige studieområdet.

Klimaet i Trøndelag er mildt og fuktig, og det kjennetegnes av store klimaforskjeller gjennom året. I følge Klimaservicesenter (2016) vil det bli 20 prosent mer nedbør i kommunen fram til 2100. Klimaprofilen for Sør-Trøndelag viser at nedbørsperiodene vil øke i både intensitet og hyppighet gjennom hele året. I Trøndelag er det den kraftige og kortvarige nedbøren som fører til overvann og urban flom, noe som igjen kan gi skade på infrastruktur og bebyggelse (Klimaservicesenter, 2016). Det gjør at tettbebygde strøk som Trondheim vil ha økt behov for tilpasning på grunn av kraftig nedbør og overvann (Skaaraas et al., 2015).

2.3 Citizen Sensing

Citizen Sensing var et samarbeid mellom fire europeiske byer: Trondheim, Norrköping, Porto og Rotterdam. Formålet med prosjektet var å utvikle og evaluere et *participatory risk management system*, et webbasert system som gjør det mulig for befolkningen å rapportere inn informasjon og observasjoner om klima og vær. Systemet skulle også bidra til refleksjon om hvordan de ulike klimahendelsene påvirker omgivelsene og hvordan en kan arbeide for en bedre klimatilpasning. Et annet mål i prosjektet var å undersøke i hvilken grad et slikt system kan bidra til å øke befolkningens og myndighetenes beredskap mot klimaendringer, såkalt klimaresiliens. Citizen Sensing-applikasjonen ble designet for å tilrettelegge for kommunikasjon mellom deltakerne og på den måten bidra til deling av både kunnskap og erfaringer (Neset et al., 2021).

Datasettet som er benyttet i denne masteroppgaven er samlet inn gjennom den andre og tredje kampanjen i Citizen Sensing-prosjektet. Høsten 2019 og høsten 2020 brukte mer enn hundre studenter på bachelornivå Citizen Sensing-applikasjonen til å rapportere inn observasjoner om vær- og klimarelaterte hendelser i Trondheim. Deltakerne var studenter i faget *Introduksjon til geografiske informasjonssystem* (GEOG1012) ved institutt for geografi ved NTNU, og datainnsamlingen var del av en øving i faget. Selve deltakelsen var frivillig, men den studenten som registrerte flest observasjoner kunne vinne en Netatmo værstasjon.



Figur 2: Citizen sensing-applikasjonen

Figur 2 viser hvordan Citizen Sensing-applikasjonen så ut for brukeren. Her ser en at posisjonen til observasjonene enten kunne fastsettes ved bruk av GPS i smarttelefonen når deltakeren observerte fenomenet, men at det også var mulig å registrere observasjonene i ettertid. Da ble lokasjon for registreringene lagt inn manuelt av brukeren ved å plassere en pil på kartet. I tillegg til posisjon ble det samlet inn informasjon om hvilken type hendelse deltakerne hadde observert. Her kunne deltakerne kategorisere hendelser i kategoriene kraftig regn, snøfall, storm, stormflo, sterk vind, sesongvarierte hendelser, kalde temperaturer, høye temperaturer og tørke. Deltakerne hadde også mulighet til å bestemme hvilken påvirkning de mente hendelsen hadde for omgivelsene. Dette valget var basert på de forhåndsbestemte kategoriene i applikasjonen -

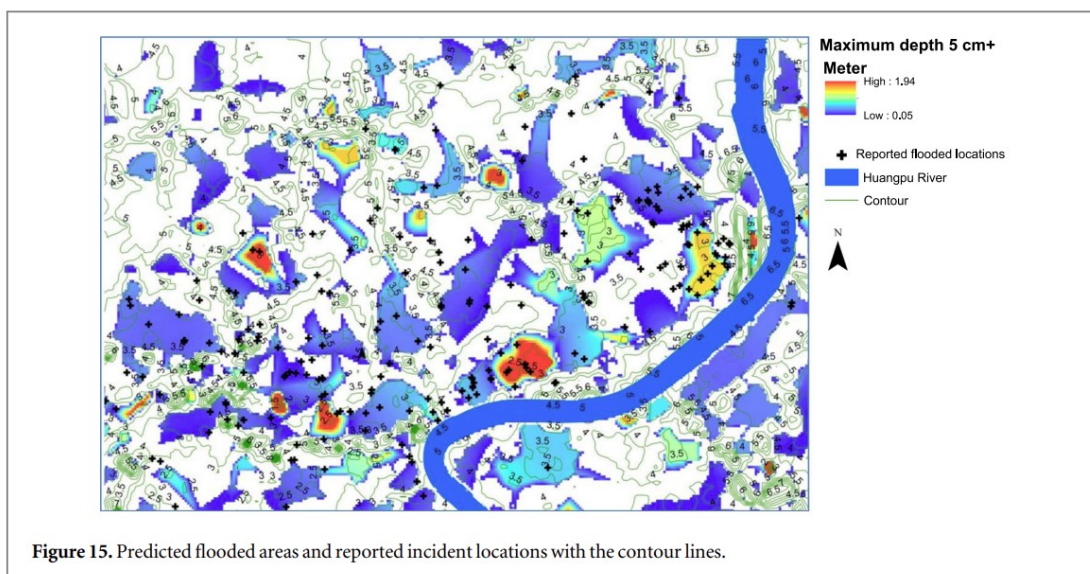
negativ, nøytral eller positiv påvirkning på omgivelsene. Deltakerne måtte deretter vurdere påvirkningsgraden til lav, moderat eller høy. I tillegg var det mulig å legge igjen en kommentar for å beskrive hendelsen mer detaljert. Til slutt kunne deltakeren legge til et bilde som viste hendelsen.

2.4 Tidligere studier

Det er flere tidligere studier som har undersøkt om data om overvann og flom samlet inn av befolkningen kan validere forsenkninger i GIS. Det finnes studier både i utlandet, i Norge og i studieområdet i Trondheim. I disse studiene er modellering av forsenkninger validert ved bruk av både observasjoner av flomhendelser og skaderegistreringer etter overvann.

2.4.1 Utlandet

Studien til Yu et al. (2016) undersøkte om folkeinnsamlede observasjoner for en spesifikk flomhendelse kunne validere modellerte forsenkninger i GIS. Datasettet var samlet inn av befolkningen gjennom en webbasert nødportal utviklet av myndighetene. Resultatet fra studien viste at så mye som 91,6 prosent av de registrerte observasjonene overlappet med modellerte forsenkninger som var dypere enn to cm, mens 54,7 prosent av punktene overlappet med forsenkninger dypere enn fem cm. Det var spesielt godt samsvar mellom de rapporterte hendelsene og de modellerte flomområdene i den indre delen av byen. Dette kan muligens skyldes at det i dette området var flest observasjoner. Figur 3 viser de modellerte flomområdene sammen med rapporterte flomhendelser fra befolkningen.



Figur 3: Modellerte forsenkninger og registrerte flomobservasjoner (Yu et al., 2016)

Noe av det samme ble gjort i studien til Di Salvo et al. (2017) i Roma. Her ble det brukt observasjonsdata om flomhendelser i perioden mellom 2014 og 2015 til å validere et farekart for flom. Datasettet bestod av totalt 50 registrerte flomhendelser. Studien viste at 57 prosent av observasjonene befant seg i modellerte forsenkninger. Studien til Di Salvo et al. (2017) konkluderte derfor med at både GIS-basert modellering av forsenkninger og observasjonsdata om flom er nødvendig for å identifisere områder utsatt for urban flom.

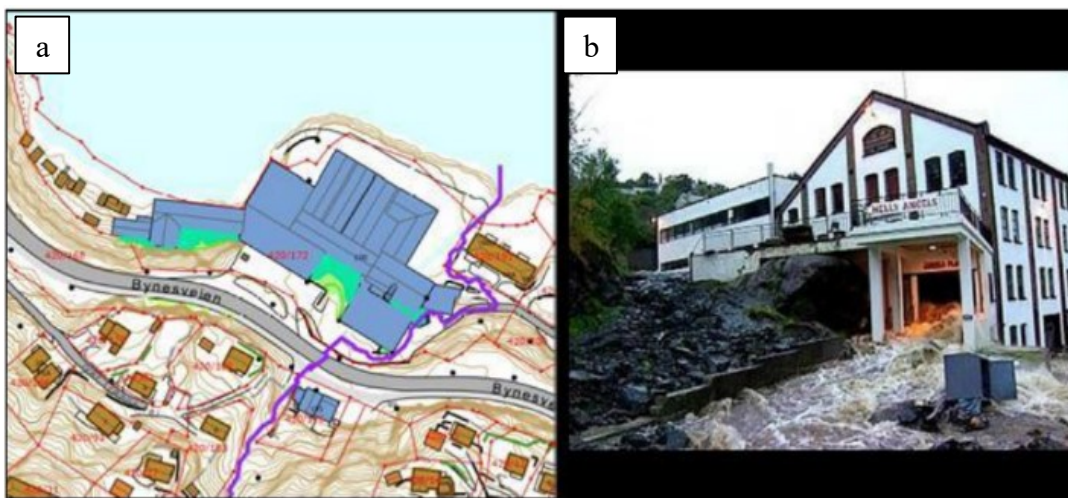
2.4.2 Norge

I studien til Aano et al. (2019) ble registrerte overvannsskader for Oslo kommune brukt for å validere forsenkninger modellert i GIS. Datasettet for skadene var levert av Finans Norge, men samlet inn av takstmenn fra seks ulike forsikringsselskap over en tidsperiode på seks år (2008 til 2014). Resultatene fra studien til Aano et al. (2019) viste ikke noe tydelig forhold mellom registrerte overvannsskader og modellerte forsenkninger, og det ble derfor vanskelig å bekrefte forsenkninger med datasettet. Studien fant likevel at antallet overvannsskader minker med økende antall meter fra modellerte forsenkninger. Den minste avstanden mellom overvannsskader og bluespots var én meter, noe som vil si at ingen skaderegistreringer overlappet direkte med modellerte forsenkninger. Dette kan skyldes at registreringer ble gjort på adresse, noe som er problematisk fordi en adresse kan dekke et stort område (Aano et al., 2019). På grunn av dette foreslår Aano et al. (2019) at GPS kan være nyttig å bruke for å registrere koordinaten til skaden. Det blir også poengtert at menneskelige feil og variasjoner i registreringsvaner er en potensiell feilkilde. De forskjellige forsikringsselskapene hadde ulike retningslinjer for hvordan takstmenn skulle registrere skader, og i tillegg kan boligens forsikring ha påvirket hvordan skaden ble registrert. Bilder av skadene viste seg å kunne gi nyttig informasjon. Studien til Aano et al. (2019) konkluderte med at til tross for at valideringen ikke var fullt så enkel på grunn av adresseregistrering, så ga likevel skadedata kombinert med modellerte forsenkninger et unikt bilde på hvor en kan finne flomutsatte områder.

2.4.3 I studieområdet

Meiforth (2013) utførte en hydrologisk modellering i GIS for de mest sentrale delene av Trondheim. For å validere om modelleringen viste troverdige resultat, ble det på en dag med kraftig nedbør foretatt feltarbeid for å validere flomveiene. I tillegg ble bilder fra tidligere flomhendelser brukt for å validere flomutsatte områder, som vist i figur 4. Denne delen av

studien til Meiforth (2013) har likhetstrekk med arbeidet utført i denne masteroppgaven hvor det også er brukt folkeinnsamlede bilder. Studien til Meiforth (2013) viser at bilder fra befolkningen kan være nyttige for å validere GIS-basert modellering. På bakgrunn av validering i felt konkluderer rapporten med at modelleringen i de fleste tilfeller virker å stemme med virkeligheten (Meiforth, 2013). Kartet i figur 4 viser modellerte forsengkninger og flomveier nær en bygning i Bynesveien i Trondheim. Bildet til høyre viser at det tidligere har vært en flomhendelse i akkurat samme område.



Figur 4: a) GIS-modellerte flomveier og forsengkninger nær bygning i Trondheim, b) bilde fra befolkningen som viser en flomhendelse i samme bygning (Meiforth, 2013)

Som vist i tidligere studier er folkeinnsamlet data brukt flere ganger for å validere modellerte forsengkninger i GIS. Studiene nevnt ovenfor viser at det er gjennomførbart, selv om det ikke alltid er like lett (Aano et al., 2019). Både Aano et al. (2019) og Di Salvo et al. (2017) konkluderer med at data samlet inn av befolkning uten forskningsbakgrunn, både i form av skadedata og observasjonsdata, gir et godt bilde av og ny innsikt i områder hvor det kan samle seg overvann. Det å ha kjennskap til skader og hendelser fra virkeligheten har altså vist seg å være nyttig i validering av hydrologisk modellering (Meiforth, 2013). Det er også nyttig for å kunne prioritere hvor forebyggende tiltak burde settes inn (Aano et al., 2019). Både studien til Aano et al. (2019) og Yu et al. (2016) viser imidlertid at sortering av datasettet før bruk er en viktig del av arbeidet med å sikre data av god kvalitet. Sortering av folkeinnsamlet data har også blitt gjort i denne masteroppgaven, og vil bli gjennomgått i metodekapittelet (kapittel 3).

2.5 Hva skiller denne masteroppgaven fra tidligere studier?

I denne masteroppgaven vil folkeinnsamlet data benyttes til å validere om modellerte forsenkninger i GIS er reelle vannoppsamlinger, noe som ligger nært det overnevnte studier har gjort tidligere. Det er derimot ingen tidligere studier i dette studieområdet som har brukt folkeinnsamlet data om overvann som hovedmetode til å bekrefte forsenkninger. Som nevnt brukte Meiforth (2013) bilder fra tidligere flomhendelser til å bekrefte modelleringen, men dette var ikke studiens hovedfokus og ble derfor lite vektlagt.

Denne masteroppgaven skiller seg derfor fra tidligere studier i studieområdet fordi det er brukt et datasett samlet inn som en del av et citizen science-prosjekt. Masteroppgaven skiller seg også fra tidligere arbeid fordi det blir brukt data samlet inn av studenter. Som teorien i kapittel 5 vil vise, er det flere faktorer som kan påvirke motivasjonen til deltakere i frivillige prosjekt, og ulik grad av motivasjon kan ha stor påvirkning på datasettets kvalitet. Denne masteroppgaven fokuserer på vannoppsamlinger av *alle* størrelser, og skiller seg derfor fra Yu et al. (2016) sin studie der de modellerte flomområdene var større i omfang. Siden observasjonene viser vannoppsamlinger av alle størrelser, har det vært viktig å ta hensyn til dette i modelleringen av forsenkninger. Masteroppgaven vil derfor se på om folkeinnsamlet data kan brukes til å bekrefte om forsenkninger, i alle størrelser, er reelle vannoppsamlinger. Dette betyr at vannoppsamlingene inkluderes selv om de ikke nødvendigvis trenger å utgjøre en fare for samfunnet. Hvis folkeinnsamlet data i denne masteroppgaven kan validere om forsenkninger av ulike størrelser er reelle vannoppsamlinger, kan en anta at dette også vil gjelde for forsenkninger av stor karakter, som i Yu et al. (2016). I motsetning til Meiforth (2013) sin studie har det ikke blitt gjennomført feltarbeid for å bekrefte resultatene. Denne masteroppgaven er en ren GIS-studie.

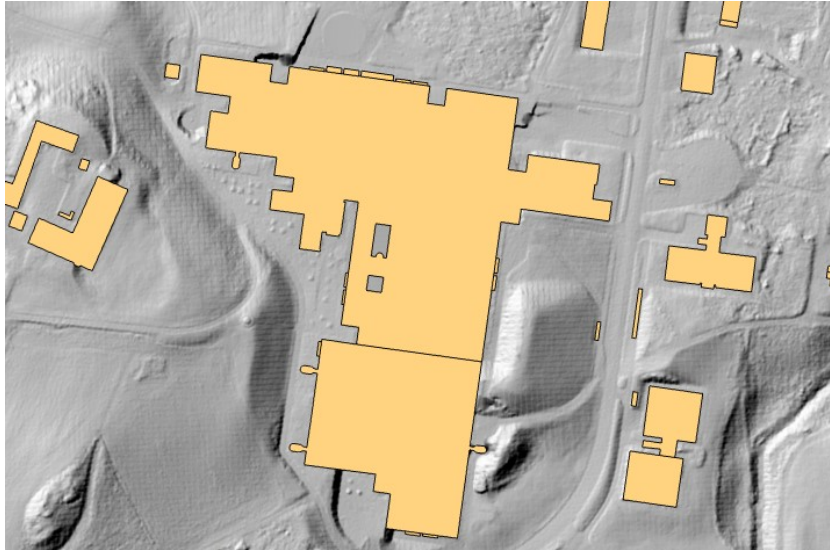
3 Teori

3.1 GIS for å representere den virkelige verden

Et geografisk informasjonssystem (GIS) defineres i Rød (2015, s. 17) som “en samling verktøy for å samle inn, lagre, endre, analysere og presentere romlige data fra den virkelige verden for et eller flere bestemte formål”. Verden er kompleks og detaljert, og vi må derfor bruke representasjoner og mindre detaljerte modeller for at datasystemene vi bruker skal fungere (Longley et al., 2005, s. 70). I GIS jobber en derfor med forenklinger av fenomener fra den virkelige verden (Rød, 2015).

3.2 Vektor- og rasterdata

I GIS representeres fenomener fra den virkelige verden ved bruk av enten vektor- eller rasterdata (Rød, 2015, s. 23). Vektordata brukes som oftest til å representere *diskrete* fenomen, altså fenomen med en tydelig avgrensning, som for eksempel en bygning. Her representeres verden ved bruk av punkt, linjer og polygon. Ved å representere virkeligheten med vektordata har en i GIS mulighet til å studere relasjonen mellom ulike datalag basert på romlig beliggenhet og egenskapsdata. Dette gjør at en lett kan gjøre utvalg av data med både romlige og ikke-romlige egenskaper. Rasterdata benyttes når en ønsker å representere *kontinuerlige* fenomen, som for eksempel høydeverdier. Rasterdata representerer verden som et rutenett hvor hver piksel i rasteret har sin egen verdi (Rød, 2015, s. 36). I denne masteroppgaven er verden representert med både vektor- og rasterdata. Folkeinnsamlede observasjoner er representert som punkt, hvor hvert punkt har et koordinat. Bygninger og elver er brukt i modelleringen av forsengkninger, og disse er representert som polygon. En digital høydemodell er et kontinuerlig raster og brukes i masteroppgaven til å modellere forsengkninger. I figur 5 representeres NTNU-campus Dragvoll som et diskret fenomen (vektor), mens terrenget med høydeverdier representeres som et kontinuerlig fenomen (raster).



Figur 5: Bygninger representere med vektordata og terrenghøyden representeres med rasterdata

For å sikre den beste representasjonen av virkeligheten som mulig, og for at resultatet av analysene skal bli så korrekt som mulig, er det viktig å finne den mest passende datatypen til å representere fenomenet (Longley et al., 2005, s. 68). Uansett hvilken datatype en bruker må en huske på at fenomen har mange ulike former i virkeligheten, mens de samme fenomenene vil være betydelig forenklet i GIS.

3.3 Digital terrengmodell (DTM)

En digital terrengmodell representerer terrenget ved hjelp av et rutenett hvor hver piksel har sin unike høydeverdi (Heywood, 2011, s. 92). I denne masteroppgaven brukes en digital høydmodell generert fra LiDAR-data til å modellere forsenkninger i terrenget. LiDAR er forkortelse for *Light Detection And Ranging*, og flybåren LiDAR regnes som en effektiv måte å produsere terrengmodeller av høy kvalitet (Yu et al., 2016). Slike skanninger gjøres ved at en sensor i flyet sender ut laserpulser som treffer bakken, hvor refleksjonene senere kan brukes til å generere en punktsky. Punktskyen over reflekterte signaler fra bakkenivå brukes til å generere en terrengmodell (Rød, 2015, s. 101).

3.4 Hvordan representeres overvann i GIS?

Overvann og flomproblematikk er et tema hvor GIS har vist seg å være et nyttig verktøy. I dag finnes det allerede store mengder geografisk informasjon om flom tilgjengelig for offentligheten. Norges vassdrags- og energidirektorat (NVE) har tilgjengelig data om flomsone og tidligere flomhendelser. Denne informasjonen tar hovedsakelig for seg elveflom, som er flom som skyldes at større vassdrag går over sine bredder på grunn økt vannføring. Når det gjelder overvann og urban flom, er det derimot lite offentlig data å finne. Dette kan muligens skyldes at det er vanskeligere å generere riktige data i et urbant miljø hvor landskapet er komplekst. Med klimaendringene er det grunn til å tro at urban flom vil bli vanligere (Skaaraas et al., 2015), og på grunn av det vil trolig mengden offentlig data som omhandler urban flom øke.

Overvann kan representeres i GIS ved såkalte *bluespots*, som er en forsenkning i terrenget hvor vann vil kunne samle seg ved nedbør (Balstrøm, 2021). Begrepet ble først brukt av det danske vegdirektoratet for å beskrive forsenkninger som kan fylles eller oversvømmes og gjøre skade på motorvei i nærheten (Hansson et al., 2010). Her var det kun forsenkninger med volum over ti kubikkmeter og som lå nær motorvei som ble inkludert. Begrepet bluespot handlet i denne konteksten spesielt om risikoen for skade fra flom. Begrepet ble da definert som et område som sannsynligvis vil oversvømmes under kraftige nedbørshendelser og som vil være til fare for infrastruktur som ligger i nærheten (Balstrøm, 2021). I denne masteroppgaven er ikke vannoppsamlings skadepotensial vektlagt, og begrepet bluespot blir derfor ikke brukt. Betegnelsen *forsenkning* blir i stedet brukt for alle områder som identifiseres som lavpunkt i terrenget. En forsenkning defineres av Di Salvo et al. (2017) som et område som ligger lavere enn sine nabopiksler. Masteroppgaven vil inkludere forsenkninger av ulike volum, og ikke kun forsenkninger med volum over ti kubikkmeter som det danske vegdirektoratet.

3.4.1 Forsenkingsanalyse

Ved hjelp av en digital terrengmodell kan en i GIS identifisere forsenkninger i terrenget hvor det kan tenkes at overvann vil kunne samle seg. Selv om modellering av forsenkninger i GIS baserer seg på en grov forenkling av urbane avrenningsforhold, har erfaring vist at resultatene samsvarer godt med observasjoner av overvann og flom i virkeligheten (Skaaraas et al., 2015, s. 62). Modellering av forsenkninger har vist seg å være nyttig i flere sammenhenger, og som den følgende teksten vil vise er dette gjort av flere. Modellering av forsenkninger kan blant annet brukes som et nyttig første steg i arealplanlegging for å få oversikt over områder som kan være utsatt for overvann. Dette er viktig informasjon fordi overvann som tidligere nevnt kan gjøre skade på bygg og infrastruktur (Skaaraas et al., 2015). Overvann kan også gi indirekte skader ved at befolkningens tilgjengelighet til ulike områder begrenses. I studien til Opach et al. (2021) ble det modellert forsenkninger som ble brukt til å undersøke hvilke ruter fotgjengere ville ta ved vannoppsamlinger. Terrengmodeller med høy oppløsning kan i urbane områder gi såpass gode detaljer at en kan identifisere oppsamlinger av overvann langs sykkel- og bilveier. Dette kan være av interesse for de med ansvar for å opprettholde god infrastruktur i kommunen (Balstrøm, 2021).

Som vist har modellering av forsenkninger i GIS flere bruksområder og kan være nyttig for å gi oversikt over sårbare områder. Det er samtidig viktig å huske at en i GIS kun bruker representasjoner av virkeligheten (Rød, 2015). Analyser og modelleringer vil alltid inneholde usikkerheter (Longley et al., 2005, s. 128). For å trekke riktige konklusjoner, må derfor de representasjonene vi bruker være så lik virkeligheten som mulig (Rød, 2015, s. 18). Som nevnt tidligere vil ansamling av overvann påvirkes av topologien på terrenget, infiltrasjonen til bakken og dreneringsmulighetene i området (Skaaraas et al., 2015). Ved modellering av forsenkninger i denne masteroppgaven blir det verken tatt hensyn til infiltrasjon eller drenering, noe som kan medføre at resultatene avviker fra virkeligheten. Selv om resultatet fra en modellering viser at det finnes en forsenkning, vil det ikke nødvendigvis bety at det blir akkumulert overvann der.

3.5 Folkeinnsamlet data

Prosjekt som baserer seg på folkeinnsamlet data er ikke noe nytt (J. Silvertown, 2009), men med fremveksten av internett, sosiale plattformer og smarttelefoner har innbyggere blitt en viktigere kilde til informasjon (Degrossi et al., 2014; See et al., 2016). Et bredt spekter av begrep har i litteraturen blitt brukt til å beskrive forskningsprosjekt hvor deltakere uten forskningsbakgrunn samler inn informasjon. *Crowdsourcing* brukes for å beskrive forskningsprosjekt på nett der et stort antall deltakere samler informasjon og meninger (Wiggins & Crowston, 2011). *Peer production* er et annet eksempel på folkeinnsamling av data og kan best beskrives som et samarbeid mellom et høyt antall deltakere på nett. Et eksempel på et slikt prosjekt er Wikipedia, der hvem som helst kan publisere faktabasert informasjon til nettsiden (Goodchild, 2007). Disse formene for prosjekt skiller seg fra mer geografiske prosjekt der det romlige aspektet står i fokus.

3.5.1 Volunteered geographic information (VGI)

Volunteered geographic information (VGI) ble først definert av Goodchild (2007) som en samlebetegnelse for å beskrive prosjekt der frivillige innbyggere samler inn og deler geografisk informasjon. Wikimapia trekkes fram som et eksempel på VGI. Her kan hvem som helst med internettilgang dele bilder og komme med beskrivelser av spesifikke steder i verden. Avsnittene nedenfor beskriver en form for VGI kalt *citizen science*. Dette begrepet passer godt som beskrivelse på prosjektet som datasettet i denne masteroppgaven er hentet fra.

3.5.2 Citizen science

Citizen science er forskning hvor personer uten noen spesiell forskningsbakgrunn bidrar til å samle inn eller prosessere data i et forskningsprosjekt (Alabri & Hunter, 2010). Det er den aktive involveringen av deltakeren som skiller citizen science fra mer passive forskningsformer som peer production og crowdsourcing (Wiggins & Crowston, 2011).

Citizen science-prosjekt kan ha et sterkt romlig aspekt med geografisk informasjon, slik som prosjektet som er brukt som datakilde i denne masteroppgaven. I Citizen Sensing-prosjektet skulle deltakerne observere klimarelaterte hendelser som ble lagt inn som punkt på et kart. Dette prosjektet passer godt inn i det Haklay (2013) betegner som et aktivt og geografisk citizen science-prosjekt. Typiske kjennetegn er at det samles inn romlig data og i tillegg informasjon om personlige oppfatninger og beskrivelser av fenomenet som observeres. Citizen

Sensing-prosjektet passer godt inn i beskrivelsen av et aktivt og undersøkende prosjekt, hvor data samles inn fra de fysiske omgivelsene. Slike prosjekt har ofte et underliggende mål om læring og utdanning (Wiggins & Crowston, 2011). Begrepet *participatory sensing* beskriver også prosjektet godt. Dette regnes ifølge Haklay (2013) som en av de nyeste formene for citizen science. Her bruker deltakerne sensorer i mobiltelefonen, som trådløst nett, kamera og GPS for å samle informasjon om den fysiske verden.

3.5.3 Fordeler og ulemper med citizen science

Citizen science kan være en effektiv og ressursbesparende måte å samle inn data på (Alabri & Hunter, 2010). Ved å inkludere et høyt antall deltakere kan det gjennomføres en mer omfattende datainnsamling, og det kan samles inn data over store studieområder (Wiggins & Crowston, 2011). Fremveksten av ny teknologi som smarttelefoner med innebygd GPS gjør det mulig med effektiv datainnsamling over et stort geografisk område. Citizen science er også en billigere måte å samle inn geografisk data på enn tradisjonelle forskningsprosjekt (Alabri & Hunter, 2010).

Med et gjennomtenkt studiedesign og under de rette omstendighetene kan citizen science generere data av høy kvalitet som kan gi pålitelige og gyldige resultat, samt uventet innsikt og nye innfall til forskningsfeltet (Wiggins & Crowston, 2011). Med riktig oppsett og god opplæring av deltakerne, kan data som samles inn gjennom citizen science-prosjekt være like god som den som blir samlet inn av eksperter (Bonney et al., 2014). Ifølge See et al. (2016) finnes det også studier som har vist at folkeinnsamlet data kan være bedre enn data samlet inn fra forskere.

En annen fordel med citizen science er at slike prosjekt kan skape økt engasjement og bevissthet om miljøutfordringer og andre utfordringer knyttet til dagsaktuell forskning (Alabri & Hunter, 2010). Den beste måten å få befolkningen til å forstå og sette pris på forskning er at de selv deltar i den (Silvertown, 2009). Citizen science gir befolkningen mulighet til å gjøre en innsats innenfor et felt de interesserer seg for, eller et geografisk område de oppholder seg mye i (Bonney et al., 2014). Til tross for fordelene som er trukket fram ved slike prosjekt, poengteres det at citizen science-data i de fleste tilfeller ikke burde erstatte annen data, men heller supplere data som allerede finnes (See et al., 2016).

Til tross for fordelene nevnt ovenfor, er ikke citizen science akseptert hos alle (Bonney et al., 2014). Flere forskningssamfunn har kritisert citizen science og mener det ikke burde brukes i ekte forskning (Delaney et al., 2008). Dette fordi flere bekymrer seg for at citizen science kan generere data av dårlig kvalitet (Bonney et al., 2014). Bakgrunnen for dette er hovedsakelig at den manglende kunnskapen og ekspertisen til innsamlere kan føre til data som er mangelfull og misledende (Alabri & Hunter, 2010). Ny teknologi gjør at teknologiske feil minimeres, men brukerfeil vil fortsatt være til stede. Som eksempel kan det nevnes at det i Alabri og Hunter (2010) ble identifisert flere former for brukerfeil i citizen science-prosjektet *CoralWatch*. Blant annet måtte 64 prosent av datasettet fjernes på grunn av feil bruk av GPS blant deltakerne. Alabri og Hunter (2010) understreker derfor at tydelige retningslinjer i forkant av datainnsamlingen er svært viktig for å sikre god data og minimere brukerfeil. I tillegg er det nødvendig at forskere gjennomfører kvalitetssikring og validering av datasettet før det benyttes i forskningen (Delaney et al., 2008).

3.5.3.1 Representativitet

Den romlige fordelingen av deltakere sees vanligvis på som en positiv del av citizen science-prosjekt, men dette kan også føre til romlig skjevhet (Alabri & Hunter, 2010). I citizen science-prosjekt vil datainnsamlingen typisk være konsentrert i områder hvor populasjonen er høy (Haklay, 2013). Her vil ofte byer og tettsteder ha et høyere antall observasjoner enn det fjellområder og avsidesliggende steder har. I følge Wiggins og Crowston (2011) kan denne ujevne fordelingen av mennesker føre til problemer med representativitet i datasettet. Dette problemet kan minimeres ved å utarbeide et godt forskningsdesign i forkant av studien, eller ved å inkludere flere deltakere (Wiggins & Crowston, 2011). Haklay (2013) legger fram at det må nøye planlegging til i forkant av prosjektet hvis en skal dekke områder der det vanligvis oppholder seg få personer. Mangel på representativitet kan også oppstå for enkelte tidsperioder. Det vil være enkelte måneder, dager og tidspunkt der flere deltakere samler inn data (Haklay, 2013).

3.5.3.2 Motivasjon og forpliktelse

Siden deltakerne ikke får betaling en indre motivasjon og interesse for prosjektets tema er ofte bakgrunnen for at folk ønsker å delta i citizen science-prosjekt (Tönurist & Surva, 2017). Typisk for slike prosjekt er at enkelte deltakere bidrar mye, mens mange bidrar lite (Haklay, 2013). Som oftest er deltakerne i citizen science anonymisert, noe som kan føre til at flere deltakere mangler forpliktelse til prosjektet (Alabri & Hunter, 2010). Den anonyme datainnsamlingen gir

ikke mulighet for selvpromotering, noe som kan være en viktig motivator for flere (Goodchild, 2007). Anonymisering kan også føre til at enkelte deltakere ikke tar oppgaven like seriøst som de ville gjort hvis identiteten hadde vært kjent. I slike tilfeller kan det bli samlet inn data som er misledende og som kan skade datasettet (Alabri & Hunter, 2010).

3.5.3.3 Heterogen data

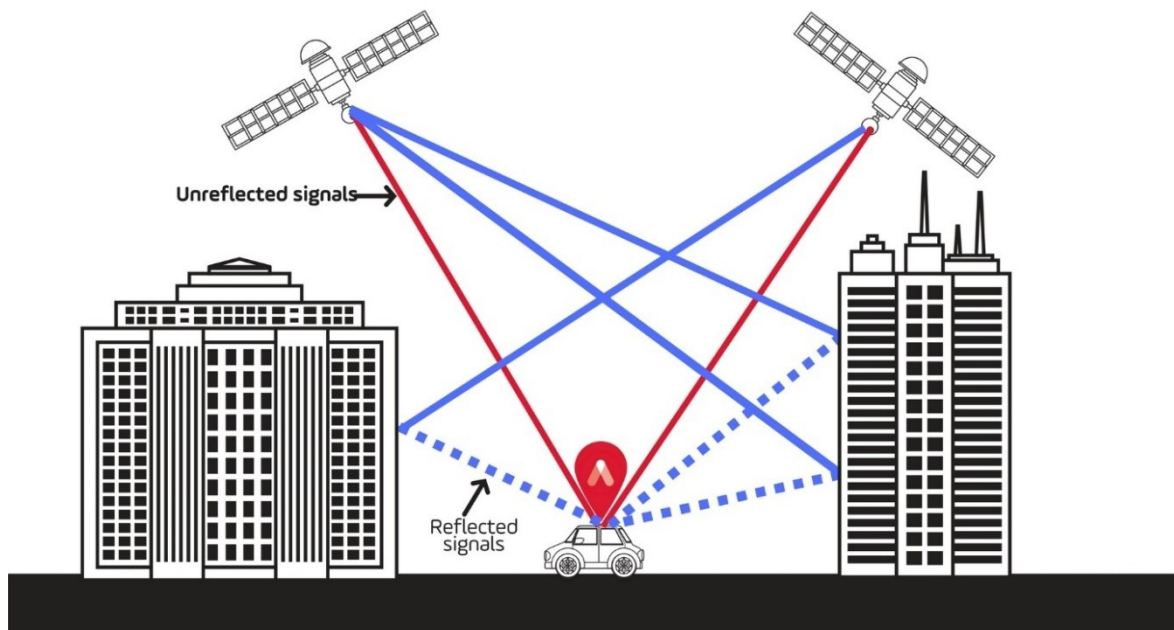
Som en har sett er data generert fra citizen science-prosjekt ofte preget av heterogenitet (Haklay, 2013). Datakvaliteten vil variere avhengig av hvor mange deltakere det er i prosjektet, og i tillegg vil hver enkelt deltakers erfaring og kunnskap spille en viktig rolle. På grunn av disse forskjellene understreker Haklay (2013) at det er viktig at datakvaliteten ikke bedømmes som en helhet, men at hvert enkelt bidrag vurderes for seg.

3.6 Posisjonering

Globale posisjoneringssystem (GPS) brukes for å registrere lokasjonen til fenomen i den fysiske verden. GPS består av totalt 24 satellitter fordelt på seks baner, noe som gjør det mulig å motta signal fra minst fire satellitter uansett hvor i verden du befinner deg (Rød, 2015). Satellittene sender radiobølger til mottakeren i smarttelefonen, og slik kan posisjonen bestemmes. Som tidligere nevnt har internett, smarttelefoner og GPS gjort at citizen science-prosjekt har blitt mer utbredt (Silvertown, 2009). Med fremveksten av smarttelefoner er GPS blitt en stor del av hverdagen til folk flest, noe som gjør det enkelt å ta i bruk GPS til hverdagslige gjøremål, men også til folkeforskning. Det folkeinnsamlede datasettet brukt i denne masteroppgaven ble levert som koordinatfestede punktdata. Punktene er koordinatbestemt både gjennom GPS i felt og ved manuell innsetting på kart.

3.7 Flerveisinterferens

I urbane områder hvor hustakene står tett og det er mange blokkerende objekter i terrenget, kan det som kalles *flerveisinterferens* være et problem. GPS-målinger gjort nært bygg kan føre til feil posisjonsbestemmelse fordi reflekterte signaler kan påvirke lokasjonen (Rød, 2015, s. 75). Dette er vist i figur 6. For stasjonære brukere kan flerveisinterferens skape koordinatfeil på noen få meter (Kos et al., 2010). De fleste observasjonene i Citizen Sensing-prosjektet ble registrert i urbane strøk der husene står tett.

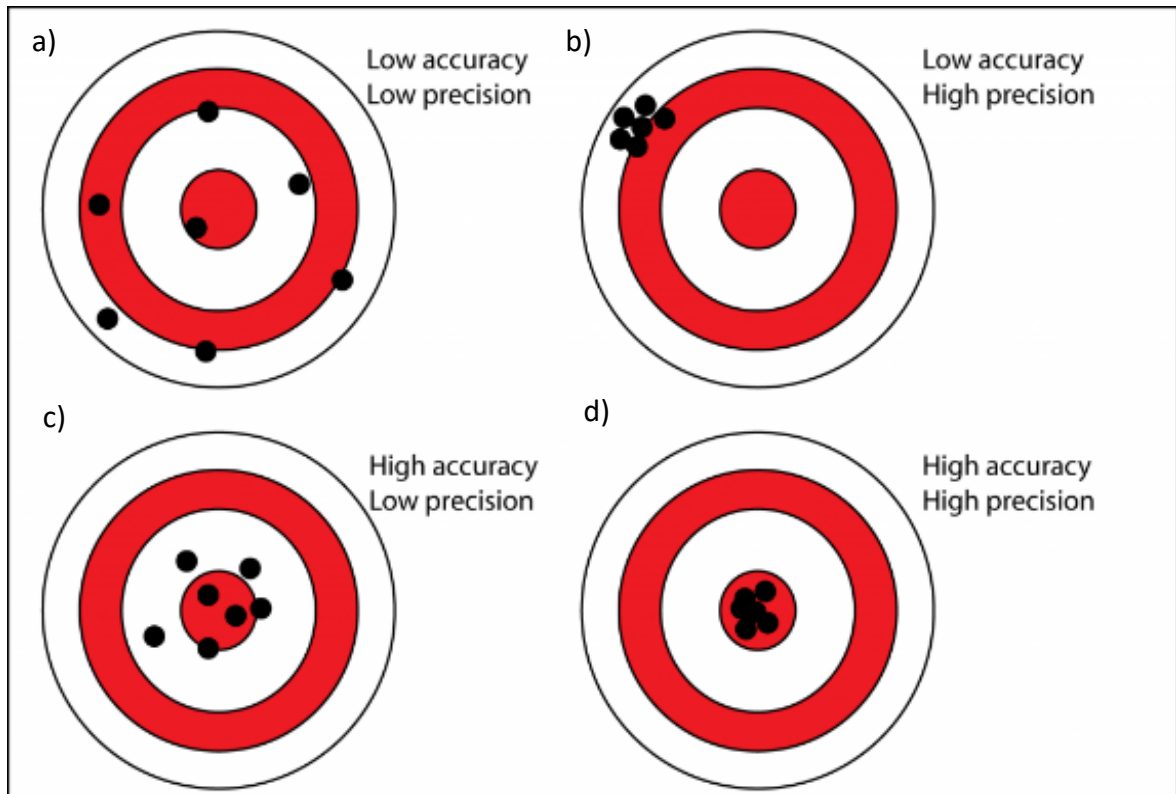


Figur 6: Illustrasjon av hvordan flerveisinterferens kan påvirke GPS-signal i urbane områder (Argust tracking, 2020).

3.8 Nøyaktighet og presisjon

Nøyaktighet og presisjon er to begrep som kan brukes til å beskrive kvaliteten på data i GIS (Longley et al., 2005). Nøyaktighet handler om hvor lik dataverdien er den faktiske verdien i virkeligheten (Heywood, 2013). Dette er illustrert i figur 7, hvor en ser at jo nærmere punktene er midten, jo mer nøyaktig er dataen. Presisjon handler om hvor stort avvik det er mellom data av samme verdi, og punktene som er registrert nært hverandre har høy presisjon. Nøyaktighet påvirkes i stor grad av metoden som er brukt for å samle inn data, og her vil måleinstrumentet spille inn (Longley et al., 2005, s. 140).

Ideelt sett burde datasettet en bruker ha både høy nøyaktighet og høy presisjon for at resultatet av analysen skal bli så korrekt og virkelighetsnært som mulig. Dette ser en eksempel på nederst til høyre i figur 7, der alle data er samlet i midten. Motsatt situasjon ser en øverst til venstre, her er punktene verken nøyaktige eller presise, men langt fra målet og med stor avstand til hverandre.



Figur 7: Illustrasjon av begrepene nøyaktighet og presisjon: a) lav nøyaktigheten, lav presisjon, b) lav nøyaktigheten, høy presisjon, c) høy nøyaktighet, lav presisjon og d) høy nøyaktighet, høy presisjon (Esri, 2017)

4 Metode

I dette kapittelet beskrives først arbeidsflyten for modellering av forsengkninger i terrenget. En digital terrengmodell kombineres med en rekke verktøy, og resultatet er en oversikt over forsengkninger der en kan anta at det vil samle seg overvann ved kraftig regn. Videre beskrives stegene som er gjort for å sortere ut deler av de folkeinnsamlede observasjonene for å minimere potensielle feilkilder. Til slutt beskrives fremgangsmåten for validering av overvann i de modellerte forsengkningene.

4.1 ArcGIS Pro

ArcGIS Pro er den nyeste GIS-programvaren til selskapet Esri og ble lansert som en erstatning for den eldre versjonen kalt ArcMap. I denne masteroppgaven er versjon 2.9.1 av programvaren brukt til å gjennomføre alle undersøkelser og analyser samt til å visualisere resultatene i form av 2D-kart. Tilleggspakken *spatial analyst* har gitt tilgang til en rekke geoprosesseringsverktøy som har vært nødvendige for blant annet modelleringen av forsengkninger. Flere av disse verktøyene vil beskrives nærmere senere i metodekapittelet.

4.2 Dataforberedelse – modellering av forsengkninger

4.2.1 Vektordata

I denne masteroppgaven ble vektordata brukt til å fjerne forsengkninger som er lokalisert i bygg, innsjøer og elver. Det samme ble gjort med folkeinnsamlede observasjoner som overlappet med bygg. Vektordata i form av polygon for bygg og vann fra Felles kartdatabase (FKB) ble lagt inn i ArcGIS Pro. FKB er en samling av noen av de mest detaljerte kartdata som finnes i Norge og var derfor godt egnet til formålet. FKB-data er produsert og forvaltes av Geovekst-partene, som er Statens kartverk, Statens vegvesen, fylkeskommunene, Energi Norge, Landbruksdepartementet, BaneNOR, Telenor og NVE (Kartverket, 2022).

4.2.2 Rasterisering

Vektordata for bygg og vann ble omgjort til et rasterlag hvor pikselverdiene i rasteret ble bestemt av objektklassen som dekket det største arealet av pikselen. Her ble terrengmodellen brukt som snappe-raster slik at pikslene i utgangsrastere lå kant i kant med pikslene i terrengmodellen. Pikselstørrelse ble satt lik terrengmodellen. De to nye rasterlagene for bygg

og vann ble reklassifisert, slik at piksler som representerte bygg og vann fikk verdien 1 og alle andre piksler i rasteret fikk verdien 0.

4.3 Digital terrengmodell (DTM)

Masteroppgavens digitale terrengmodell ble lastet ned fra hoydedata.no og levert som separate TIF-filer. Filene ble satt sammen til et raster i ArcGIS Pro, hvor koordinatsystemet ble satt til *ETRS 1989 UTM Zone 32N* for å være egnet for Midt-Norge og Trondheim. Pikseltypen ble definert til *32 bit float* for å bevare terrengmodellens opprinnelige verdibredde. For å spare prosesseringstid i den videre analysen ble terrengmodellen klippet til å dekke det opprinnelige studieområdet for Citizen Sensing-prosjektet.

4.3.1 Forsenkningsanalyse i urbane strøk

GIS-basert modellering i urbane strøk kan være mer utfordrende da landskapet her er komplekst og ofte består av mange små høydeforskjeller. Bygninger, veier og broer påvirker hvor modellen identifiserer forsenkninger. Det å ikke få representert slike urbane objekter godt nok er ifølge Wang et al. (2018) en av de største kildene til usikkerhet i urban flom-modellering. Det er derfor viktig å velge en terrengmodell som har såpass god oppløsning at den tar hensyn til urbane objekter.

4.3.2 Hvilken oppløsning egner seg?

Terrengmodellens oppløsning og kvalitet kan ha stor påvirkning på analysens resultat (Bratlie, 2015). Ifølge Murphy et al. (2008) har pikselstørrelse større betydning for resultatet av en hydrologisk modellering enn måten terrengmodellen ble produsert på. Hvilken pikselstørrelse som er egnet er avhengig av størrelsen på studieområdet, hvordan terrenget i studieområdet er og ønsket resultat av analysen. Typisk er grove pikselstørrelser brukt for store studieområder, mens fine pikselstørrelser for små studieområder (Hengl & Evans, 2009). I et slakt terreng uten store høydeforskjeller vil en grov pikselstørrelse som regel være godt nok (Vaze et al., 2010), og dette vil også gi gevinst i form av kortere prosesseringstid (Hengl & Evans, 2009).

I denne masteroppgaven baseres modelleringen på Balstrøm (2021) sin modell. Her modelleres forsenkninger ved bruk av en terrengmodell med oppløsning på 40 cm, noe som kan regnes som en fin oppløsning. Likevel poengterer Balstrøm (2021) at terrengmodeller med grovere oppløsning også kan gi gode resultat i modellering av forsenkninger. Albă et al. (2018) brukte også Balstrøm (2021) sin modell, men her ble det brukt en terrengmodell med grov oppløsning

for deler av studieområdet. Alba et al. (2018) poengterer at bruk av en grov oppløsning påvirket resultatene negativt. Studien til Di Salvio (2017) fant ut at nøyaktigheten til terrengmodellen er svært viktig for at modelleringen av forsengkninger skal bli så korrekt som mulig.

4.3.3 Valg av oppløsning

Det virker å være nokså stor enighet om at finere pikselstørrelser har bedre evne til å representere terrenget detaljert. Dette støttes blant annet av Yang et al. (2014) som sier at en finere oppløsning på terrengmodellen vanligvis vil føre til en mer korrekt representasjon av terrenget. For å få en mest mulig korrekt representasjon av terrenget har terrengmodellen brukt i denne masteroppgaven en oppløsning på 25 cm og en punkttetthet på 5 punkt per m², som er den beste oppløsningen tilgjengelig for masteroppgavens studieområde. Oppløsningen er valgt fordi studieområdet er urbant med store høydeforskjeller og har infrastruktur som er viktig å inkludere i modelleringen. I følge Heywood (2011) beskriver oppløsningen det minste fenomenet som kan representeres i datasettet. Dette betyr at dersom en modellerer forsengkninger med en terrengmodell med oppløsning på én meter, vil ingen forsengkninger mindre enn dette inkluderes. Oppløsningen på 25 cm er derfor valgt for å kunne modellere så små forsengkninger som mulig, og dermed gi mulighet til å inkludere også små vannoppsamlinger i analysen. Dette fordi det folkeinnsamlede datasettet brukt i masteroppgaven viser bilder av et bredt spekter av vannoppsamlinger, alt fra små sølepytter til dammer som strekker seg over flere meter.

4.4 Bluespot-modell

Etter en kraftig nedbørshendelse i København som førte til oversvømmelse av flere kjellere, utviklet Balstrøm (2021) en modell for å identifisere områder utsatt for akkumulering av overvann, såkalte bluespots. Bluespot-modellen identifiserer bluespots, deres lokale nedbørsfelt og hvordan de fylles eller oversvømmes ved nedbørshendelser, men uten å ta hensyn til infiltrasjonsrate og tid (Balstrøm, 2021). Modellen har blitt brukt i flere studier, blant annet i studien til Albă et al. (2018) hvor den ble brukt til å identifisere forsengkninger i sentrum av Craiova i Romania.

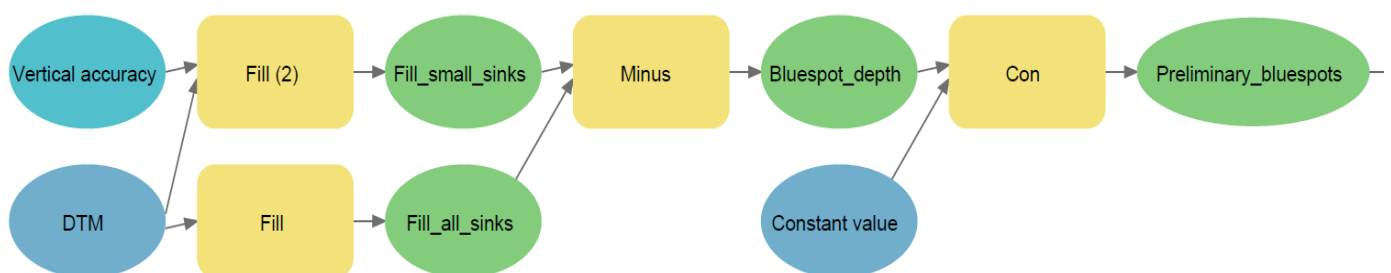
Denne masteroppgaven tar utgangspunkt i en forenklet versjon av Balstrøm (2021) sin opprinnelige modell. Dette er en modell som ble brukt i en øvingsoppgave i faget «GIS Tools for Climate Change Studies» (GEOG3527) ved NTNU i 2020. Arbeidsflyten i modellen er igjen forandret noe for å passe bedre til denne oppgavens formål. Blant annet tar modellen verken for

seg beregning av volumet til forsenkningene eller hvor mange millimeter med nedbør som trengs for å fylle forsenkningene. I denne oppgaven identifiseres *alle* forsenkninger, uavhengig av deres størrelse eller volum. Som tidligere nevnt er dette fordi vannoppsamlingene observert i Citizen Sensing-prosjektet varierer i størrelse og volum, og det var derfor viktig å inkludere alle forsenkninger. I senere tid ser jeg likevel at mitt valg om en justert modell ble tidkrevende og at Balstrøm (2021) sin opprinnelige modell trolig kunne gitt de samme resultatene. Men en justering av Balstrøms modell har vært nyttig læring og gitt god forståelse av hvordan de ulike verktøyene i modellen fungerer. Justeringen av modellen krevde at jeg måtte sette meg godt inn i de ulike stegene som kreves for å modellere forsenkninger.

4.5 Arbeidsflyt for modellering av forsenkninger

Arbeidsflyten er bygget i *Model Builder*, et visuelt programmeringsspråk i ArcGIS Pro. En slik modell gir god oversikt og gjør det enkelt å forandre parameter, finne feilkilder og med det oppdage hvilke steg som må forandres. Når arbeidsflyten først er bygget i modellen, kan den enkelt gjenbrukes til samme analyse i andre studieområder. I Model Builder symboliseres inngangsdata med blå symboler, verktøy med gule symboler og utgangsdata med grønne symboler. Figuren som symboliserer parameter har en lysere blåfarge. De kommende sidene vil beskrive arbeidsflyten i modellen stegvis.

4.6 Forsenkingsanalyse – første del

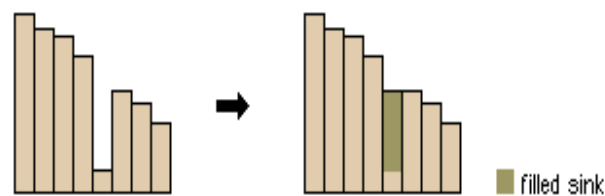


Figur 8: Første del av forsenknings-analysen

Som figur 8 viser består den første delen av modellen av de tre verktøyene *fill*, *minus* og *con*. Avsnittene som kommer vil gå nærmere inn på verktøyene og hvilke inngangsdata og utgangsdata disse har.

4.6.1 Fylle groper

I en digital terrengmodell kan en finne feilkilder i form av små groper eller forhøyninger som kan ha stor betydning for analysens resultat (Vaze et al., 2010). En grop er en piksel uten definert dreneringsretning (Rød, 2015, s. 241). Dette kan oppstå under datainnsamlingen dersom høydeverdier rundes opp til heltall (Esri, 2021a). Ofte ser en at antallet groper øker i takt med rasterets pikselstørrelse og en antar derfor at antallet groper forårsaket av feil er mindre med dagens høydemodeller som har finere oppløsning (Rød, 2015, s. 242).



Figur 9: Fylling av groper (Esri, 2021a)

Figur 9 illustrerer hvordan verktøyet *fill* fungerer. Kort oppsummert så fyller verktøyet groper slik at en kontinuerlig flyt i dreneringsnettverket opprettholdes. I masteroppgavens forsenkningsanalyse ble groper fylt to ganger med samme terrengmodell som inngangsdata. Den første gangen ble det kun fylt groper som en kunne anta var resultat av feil i dannelsen av høydemodellen, og her ble det tatt utgangspunkt i den vertikale nøyaktigheten til terrengmodellen. Vertikal nøyaktighet handler om hvor store feil en kan regne med å finne når høydeverdiene i terrengmodellen sammenlignes med virkeligheten. I følge Balstrøm (2021) er det god praksis å ikke lete etter bluespots som er grunnere enn den vertikale nøyaktigheten til terrengmodellen. Dette fordi disse kan representere feil i datasettet.

4.6.2 Vertikal nøyaktighet

American Society for Photogrammetry and Remote Sensing (ASPRS) kom i 2014 med en standard for vertikal nøyaktighet for terrengmodeller generert fra LiDAR-data. Som en ser i figur 10 (røde markeringer), vil en digital terrengmodell med en punkttetthet på mellom 2-8 punkt per m² typisk ha en vertikal nøyaktighet på mellom 5-10 cm.

TABLE B.9 EXAMPLES ON VERTICAL ACCURACY AND RECOMMENDED LIDAR POINT DENSITY FOR DIGITAL ELEVATION DATA ACCORDING TO THE NEW ASPRS 2014 STANDARD

Vertical Accuracy Class	Absolute Accuracy		Recommended Minimum NPD ⁸ (pls/m ²)	Recommended Maximum NPS ⁸ (m)
	RMSE _z Non-Vegetated (cm)	NVA at 95% Confidence Level (cm)		
1-cm	1.0	2.0	≥20	≤0.22
2.5-cm	2.5	4.9	16	0.25
5-cm	5.0	9.8	8	0.35
10-cm	10.0	19.6	2	0.71
15-cm	15.0	29.4	1	1.0
20-cm	20.0	39.2	0.5	1.4
33.3-cm	33.3	65.3	0.25	2.0
66.7-cm	66.7	130.7	0.1	3.2
100-cm	100.0	196.0	0.05	4.5
333.3-cm	333.3	653.3	0.01	10.0

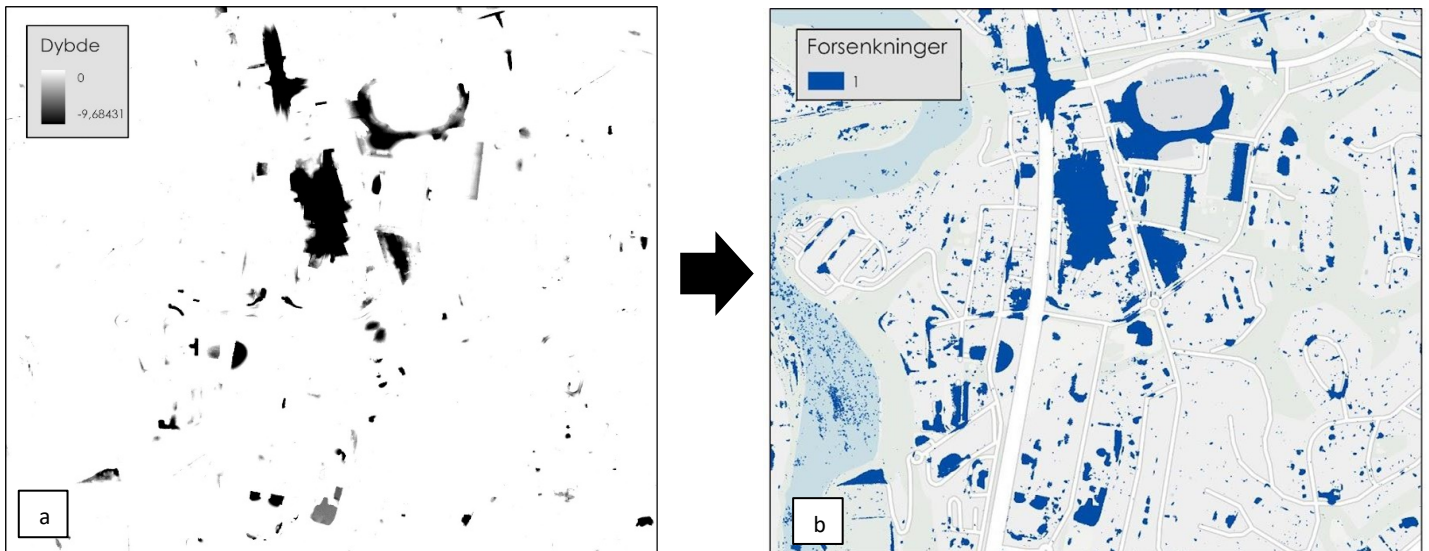
Figur 10: Anbefalt vertikal nøyaktighet ifølge ASPRS standarden fra 2014 (American Society for Photogrammetry, 2015)

Balstrøm (2021) definerte vertikal nøyaktighet til 5 cm i sin *bluespot-modell*. Det må nevnes at Balstrøms modell bruker en terrengmodell med oppløsning på 40 cm, sammenlignet med 25 cm oppløsning som er brukt i analysen i denne masteroppgaven. Valget av vertikal nøyaktighet er likevel basert på ASPRS-standarden samt Balstrøm (2021) sin studie, med bruk av en vertikal nøyaktighet på 5 cm. Dette betyr at alle groper under 5 cm regnes som potensielle feilkilder og derfor blir fylt.

Den andre gangen fyllverktøyet ble kjørt fyltes groper uten noen satt vertikal nøyaktighet – da blir altså alle groper uavhengig av høyde fylt. Pikseldverdiene for første terrengmodell ble så trukket fra pikseldverdiene til andre terrengmodell. Resultatet ble et raster som viser dybden til hver piksel i terrengmodellen (se figur 11a).

4.6.3 Hvis/eller-evaluering

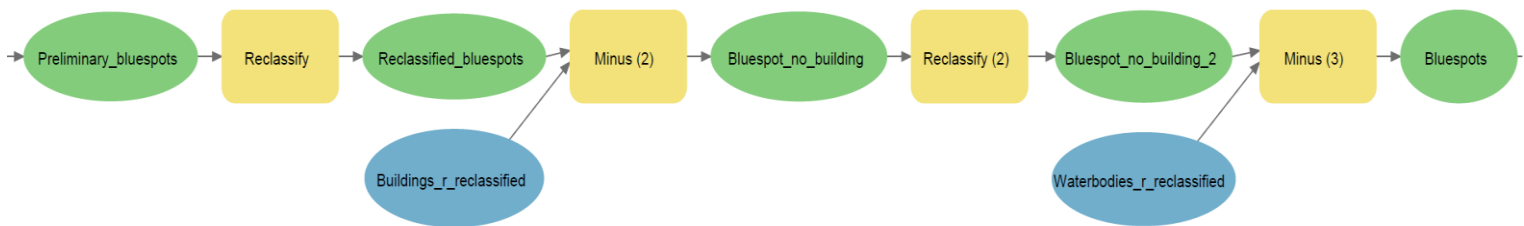
Verktøyet *con* er en kartalgebrafunksjon som utfører en hvis/eller-evaluering (Esri, 2021a), og i denne masteroppgaven var det dybden til forsenkningene som skulle evalueres. Her skulle piksler med negativ verdi velges ut. Uttrykket «VALUE < 0» ble brukt til å evaluere om pikselen skulle gis konstant verdi 1 dersom uttrykket var sant, eller kodes som *NoData* dersom uttrykket var usant.



Figur 11: Verktøyet *con* utfører en hvis/eller-evaluering av pikslene i studieområdet

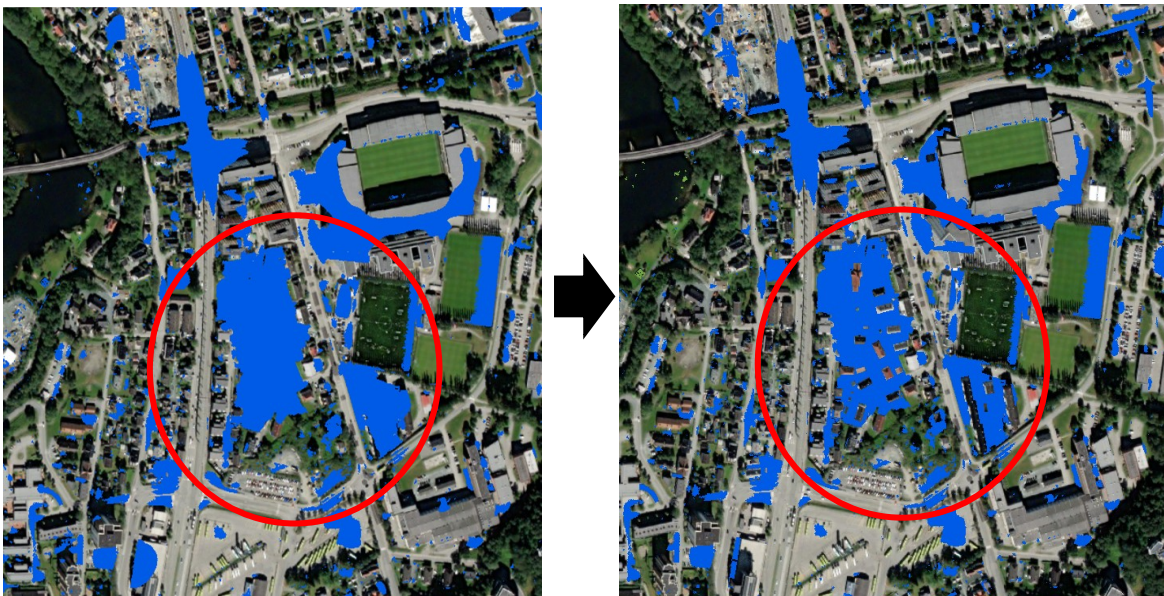
Figur 11 viser en illustrasjon av hvordan *con*-verktøyet fungerer i masteroppgavens studieområde. Her evalueres dybden til forsenkningene ved bruk av uttrykket nevnt tidligere, hvor resultatet er et nytt raster med verdien 1 for piksler som møter betingelsen og *NoData* for piksler som ikke møter betingelsen. I dette tilfellet regnes alle piksler med verdien 1 som forsenkninger i terrenget.

4.7 Forskningsanalyse – andre del



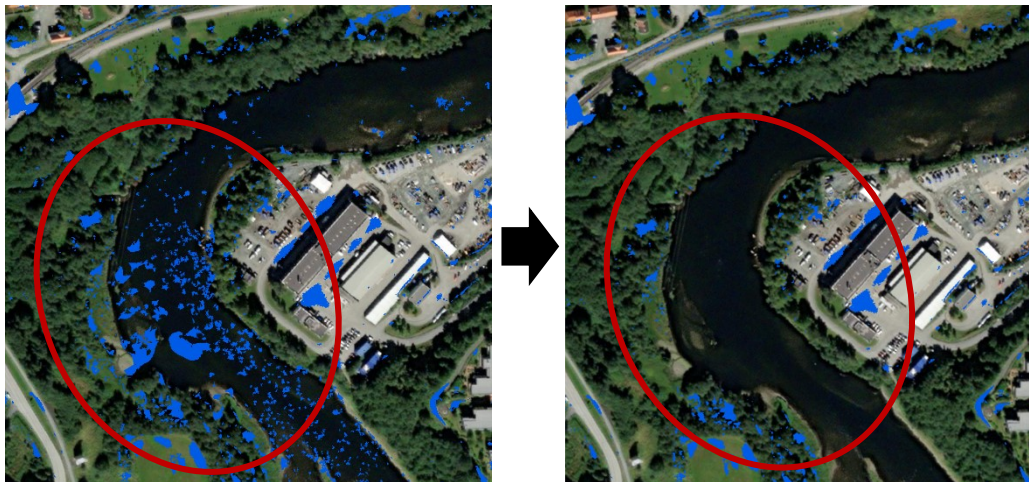
Figur 12: Andre del av forskningsanalysen

Figur 12 viser arbeidsflyten i den andre delen av bluespot-modellen. Modellen starter med at rasterlaget fra forrige steg ble reklassifisert slik at forskninger fikk verdien 1 og alle andre piksler verdien 0. Kartalgebrafunksjonen *minus* ble utført med rasterlaget med forskninger som første inngangstema og rasterlaget for bygg som andre inngangstema. Her ble pikselverdiene i inngangstemaet trukket fra pikselverdiene i andre inngangstema (1 minus 1 er lik 0), og resultatet ble et raster der kun forskninger som ikke lokaliserer med bygg har verdien 1. Figur 13 viser forskninger før og etter verktøyet *minus*. Her ser en at piksler som hadde verdien 1 (bygg) får verdien 0 etter subtraksjonen.



Figur 13: Verktøyet minus endrer verdien til piksler som samlokaliserer med bygg til 0

Andre verdier enn 1 ble omkodet til NoData. Dette rasterlaget ble første inngangstema i en subtraksjon med rasterlaget for innsjøer og elver. Dette resulterte i et nytt raster hvor forsenkninger som er lokalisert i innsjøer og elver ble kodet som NoData. Det var viktig å fjerne forsenkninger lokalisert i bygg og vann ettersom disse feilaktige forsenkningene kunne ført til feil i den romlige analysen og avstandsmålingen.



Figur 14: Verktøyet minus endrer verdien til piksler lokalisert i Nidelva til 0

Figur 14 viser hvordan verktøyet minus endret verdiene til pikslene som var lokalisert i Nidelva i studieområdet. Pikslene har på første bilde verdien 1, men når de ble trukket fra rasterlaget for elv fikk pikslene verdien 0.

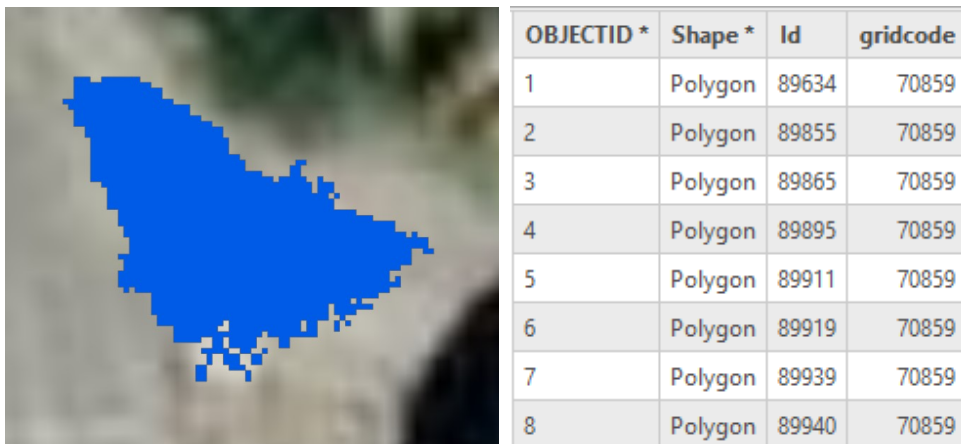
4.8 Forsenkningsanalyse – tredje del



Figur 15: Tredje del av forsenkningsanalysen

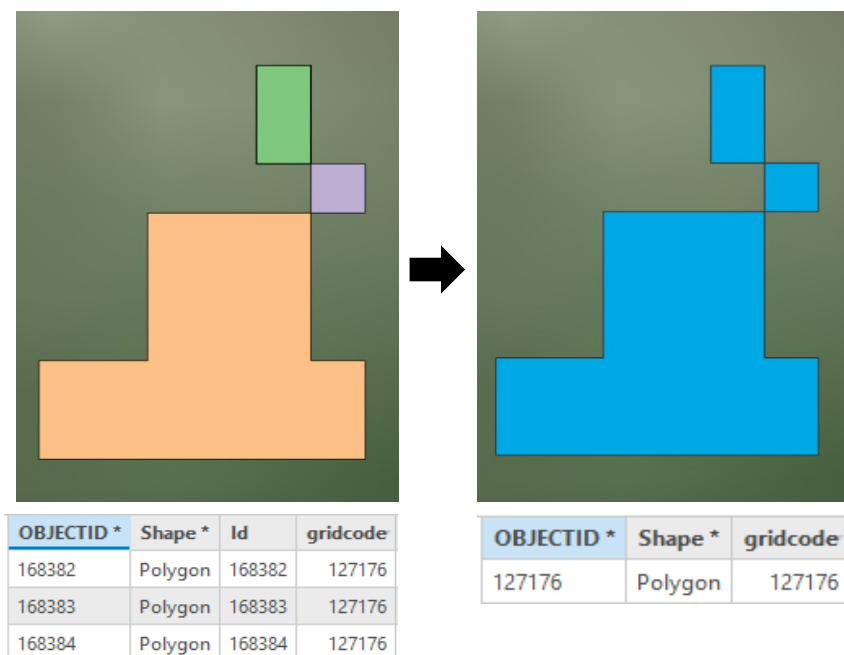
Figur 15 viser arbeidsflyten i den tredje og siste delen av modellen. Forsenkningene ble reklassifisert og gitt verdien 1, mens alle andre piksler i rasteret fikk verdien 0. Videre ble det definert regioner for pikslene, hvor hver region ble gitt en unik verdi. Her ble alle pikslene i rasteret skannet og verktøyet tilegnet en region til hver piksel basert på hvilke piksler som hadde

en romlig tilknytning til hverandre. Figur 16 viser hvordan piksler som er lokalisert ved siden av hverandre får tildelt samme *gridcode*.



Figur 16: Pikslene blir tildelt en felles gridcode basert på deres romlige tilknytning

I den siste delen av modellen ble rasterlaget for forsenkninger omgjort til et polygonlag. Forsenkninger som hadde lik gridcode ble deretter aggregert sammen, slik at en fikk færre store polygoner i stedet for flere små. Figur 17 illustrerer hvordan aggregeringen fungerer. Her aggregeres flere små polygon som har lik gridcode til en sammenhengende polygon.



Figur 17: Piksler med lik gridcode aggregeres sammen til ett sammenhengende polygon

4.9 Folkeinnsamlede observasjoner

Datasettet som brukes i denne oppgaven inneholder opprinnelig 2554 registreringer samlet inn under den andre og tredje kampanjen til Citizen Sensing-prosjektet i 2019 og 2020. De stedfestede observasjonsdataene ble lagt inn som shapefil i ArcGIS Pro. For denne masteroppgaven var det kun informasjon om objekt-ID, brukernavn, dato og tidspunkt for registrering, bilde lagt inn av deltakeren og koordinatene til punktene som ble brukt. Det var kun denne informasjonen som ble vurdert som relevant for masteroppgavens problemstilling.

4.10 Sortering av observasjoner

Silvertown (2009) poengterer at før folkeinnsamlet data kan benyttes er det viktig at datasettet blir validert og gjennomgått av noen som har god kjennskap til temaet. I følge Longley et al. (2005, s. 152) skal en aldri stole på data som ikke har gjennomgått kvalitetssjekk, spesielt ikke data fra kilder som er kjent for å ha dårlig kvalitet. Som nevnt i teorikapittelet mener flere at data generert i citizen science preges av lav kvalitet (Alabri & Hunter, 2010). I studien til Bowser et al. (2020) ble det funnet at 56 prosent av totalt 36 undersøkte citizen science-prosjekt benyttet seg av en ekspert til gjennomgang av dataen før bruk. I samme studien nevnes det også at i 39 prosent av prosjektene ble data som ikke kunne regnes som pålitelige fjernet før bruk i forskning. Flertallet av prosjektene fant det nødvendig med en profesjonell gjennomgang og sortering for å fjerne data som kan regnes som feilkilder før det tas i bruk i forskning.

I studien til Aano et al. (2019) måtte registreringer av overvannsskader geokodes og sorteres slik at kun relevant data var igjen. I studien gikk sorteringen ut på å fjerne skaderegistreringer som ikke lå innenfor studieområdet Oslo kommune, fjerne registreringer som ikke lot seg geokode og til slutt fjerne skader som ikke var relatert til overvann. Ofte kan sortering innebære at en stor del av det opprinnelige datagrunnlaget forsvinner, slik som i studien til Aano et al. (2019) hvor kun 9,3 prosent av det opprinnelige datasettet var igjen. I studien til Yu et al. (2016) ble også folkeinnsamlede observasjoner av flomhendelser sortert før de ble tatt i bruk. Her ble registreringer som ikke var relatert til overvannsfloam eller doble registreringer fjernet. I den samme studien ble det opprinnelig registrert 490 flomhendelser, men etter å ha fjernet dubletter og registreringer som ikke viste overvann var det 298 observasjoner igjen.

På bakgrunn av hva tidligere studier har sagt om usikkerhet rundt folkeinnsamlet data, ble sortering også vektlagt i denne masteroppgaven. Dette for å minimere feilkilder og sikre best mulig kvalitet på datasettet. Gjennomgangen var også viktig for å få mer kunnskap og oversikt

over datasettets kjennetegn. Dette er informasjon som senere ble brukt til å vurdere hvor egnet observasjonene var til å validere om modellerte forsenkninger er reelle vannoppsamlinger. Hvordan sorteringen kan ha påvirket det endelige resultatet - både i en positiv, men også negativ retning, vil diskuteres i kapittel 6.

Sorteringen av observasjoner ble delt inn i to deler. Betegnelsen *grov sortering* brukes om de stegene av sorteringen som krever liten vurdering. Her ble det sortert ut observasjoner som ikke ble regnet som relevante for masteroppgavens studieområde eller problemstilling, og som ikke kunne gi troverdige resultat i den senere analysen. Betegnelsen *detaljert sortering* brukes om sorteringen som krever en grundigere vurdering.

Tabell 1: Sortering av folkeinnsamlede observasjoner, delt inn i grov og detaljert sortering

	Steg	Fremgangsmåte
Grov sortering	1. Fjerne observasjoner utenfor det opprinnelige studieområdet	Romlig utvalg
	2. Fjerne observasjoner som ikke har hendelse <i>kraftig regn</i>	Utvalg fra tabell
	3. Fjerne observasjoner som ligger i bygningspolygon	Romlig utvalg
	4. Fjerne observasjoner uten bilde	Utvalg fra tabell
	5. Fjerne observasjoner som ligger på broer	Manuell fjerning
Detaljert sortering	6. Fjerne observasjoner der bildet ikke viser overvann	Manuell fjerning
	7. Fjerne observasjoner hvor bildet ikke viser omgivelser	Manuell fjerning
	8. Fjerne observasjoner der omgivelsene på bildet er vanskelig å identifisere	Manuell fjerning
	9. Fjerne observasjoner med feil posisjon	Manuell fjerning
	10. Erstatte observasjoner med feil posisjon med nye punkt	Manuell punktsetting

Tabell 1 viser hvilke steg som inngår i den grove og detaljerte sorteringen av folkeinnsamlede observasjoner. Den grove sorteringen bestod av å fjerne observasjoner utenfor prosjektets opprinnelige studieområde, fjerne observasjoner som ikke var kategorisert med hendelse kraftig regn, fjerne observasjoner som lå i bygg, fjerne observasjoner som manglet bilde og til slutt fjerne observasjoner som lå på broer. Den detaljerte sorteringen gikk ut på å fjerne observasjoner hvor bildet ikke var vannrelatert, fjerne observasjoner der bildet ikke viste

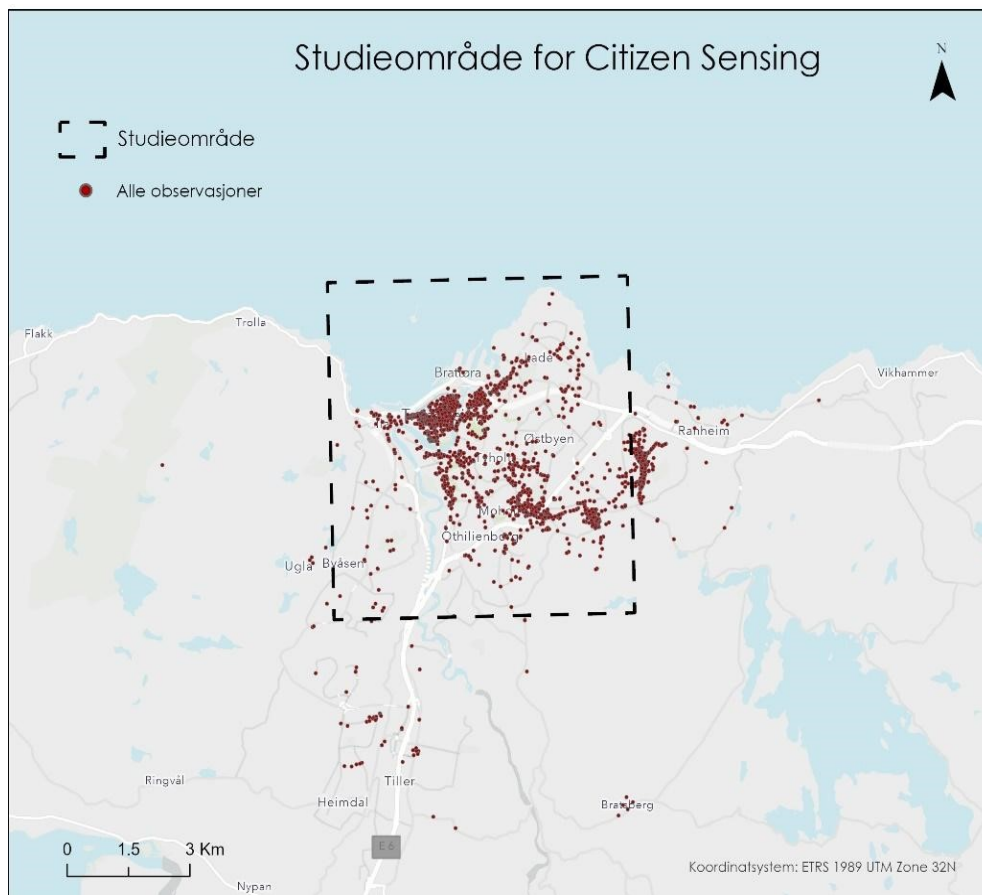
omgivelser, fjerne observasjoner der omgivelsene var vanskelig å identifisere, og til slutt fjerne observasjoner der bildet viser omgivelser, men der observasjonen hadde feil posisjon.

4.11 Grov sortering

Den grove sorteringen ble gjort ved romlig utvalg og attributtseleksjon. Ved romlig utvalg selekteres data basert på deres relative lokasjon til andre datasett, mens ved attributtseleksjon gjøres utvalg basert på egenskapsinformasjon. I GIS kan begge disse operasjonene enkelt utføres for å velge ut det datasettet en ønsker. Stegene i den grove sorteringen var nødvendig for at datasettet kunne brukes til oppgavens formål, da disse dataene ble regnet som potensielt store feilkilder i den videre analysen.

4.11.1 Fjerne observasjoner utenfor det opprinnelige studieområdet

Før Citizen Sensing-kampanjens oppstart ble det definert et studieområde for å avgrense registreringene av observasjoner. Som en ser i figur 18 er de fleste observasjonene registrert innenfor studieområdets grenser. Likevel finner en også observasjoner på Ranheim, i Tillerområdet og ved Bratsberg. Det ble også registrert observasjoner utenfor Trondheim kommunes grenser. For å avgrense masteroppgavens studieområde ble alle observasjoner utenfor det opprinnelige studieområdet til Citizen Sensing-prosjektet fjernet.



Figur 18: Det opprinnelige studieområdet i Citizen Sensing-prosjektet

4.11.2 Fjerne observasjoner som ikke har hendelse *kraftig regn*

Som nevnt i masteroppgavens bakgrunnsdel handlet Citizen Sensing-prosjektet om mer enn observasjoner av overvann, noe som gjorde det nødvendig å sortere ut observasjoner som ikke omhandlet dette. For oppgavens formål var det kun hendelser kategorisert som *kraftig regn* som ble beholdt da det kun er disse observasjonene som er relevante for masteroppgavens problemstilling. Dette samsvarer med det som ble gjort både i studien til Aano et al. (2019) og Yu et al. (2016). Her ble også hendelser som ikke kunne knyttes til overvann fjernet. Dette steget ble regnet som grunnleggende for å minimere irrelevant data som kan skape feil i analysen.

4.11.3 Fjerne observasjoner i bygg

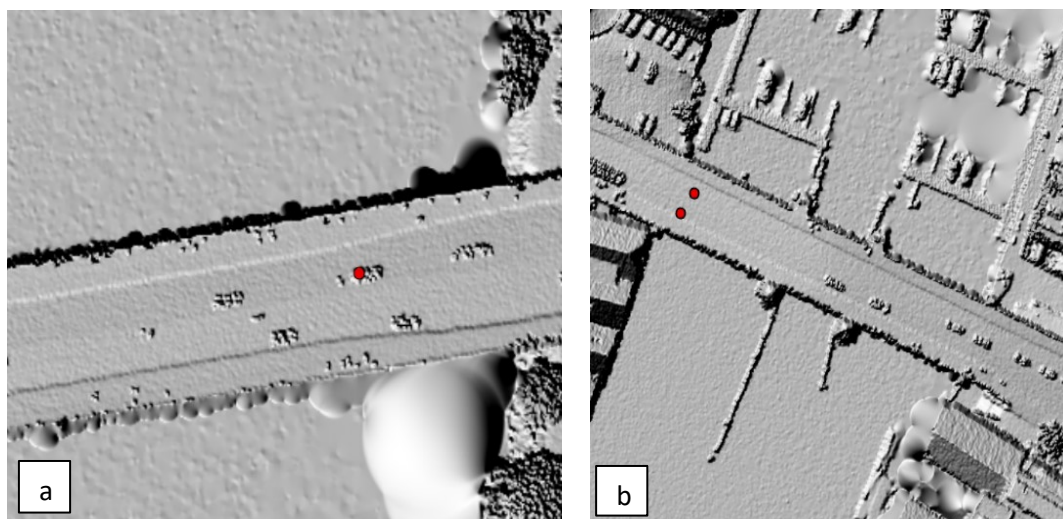
I det opprinnelige datasettet var flere observasjoner registrert inne i bygg. Dette skyldes trolig at deltakeren ikke har gjennomført registreringen på stedet hvor vannoppsamlingen ble observert, men først etter å ha kommet hjem eller i et annet bygg. Uavhengig av årsak vil ikke bluespot-modellen ta hensyn til forsenkninger lokalisert i bygg. Derfor ble alle observasjoner lokalisert i bygg fjernet ved bruk av bygningspolygon fra Felles kartdatabase (FKB).

4.11.4 Fjerne observasjoner uten bilde

Totalt 23 observasjoner manglet bilde, noe som skapte usikkerhet om observasjonene faktisk var av overflatevann eller om det var en annen hendelse. Bildene koblet til observasjonene var en viktig del av informasjonen, og disse ble senere brukt til å stedfeste observasjonene som hadde feil koordinat. Observasjoner uten bilde ble fjernet for å forhindre unødvendig usikkerhet rundt datasettet.

4.11.5 Fjerne observasjoner som ligger på broer

Totalt tre observasjoner ble registrert på broer, ett på Elgeseter bro og to på Bakke bro. Den digitale terrenngmodellen inkluderer kun det laveste punktet i terrenget, dermed vil ikke broer representeres. I mange tilfeller er det mulig å erstatte terrenngmodellen med data fra en overflatemodell (DOM). Ved å bruke en overflatemodell til bluespot-modellering kan en oppleve at falske bluespots kan modelleres, for eksempel store bluespots mellom trær eller bygg (Balstrøm, 2021). I masteroppgavens studieområde var det store mengder støy på broene. Figur 19 viser en overflatemodell med skyggerelieff, der biler skaper høydeforskjeller midt på broene og dermed gir falske høydeforskjeller i en GIS-modellering.



Figur 19: Støy fra biler på a) Elgeseter bro og b) Bakke bro

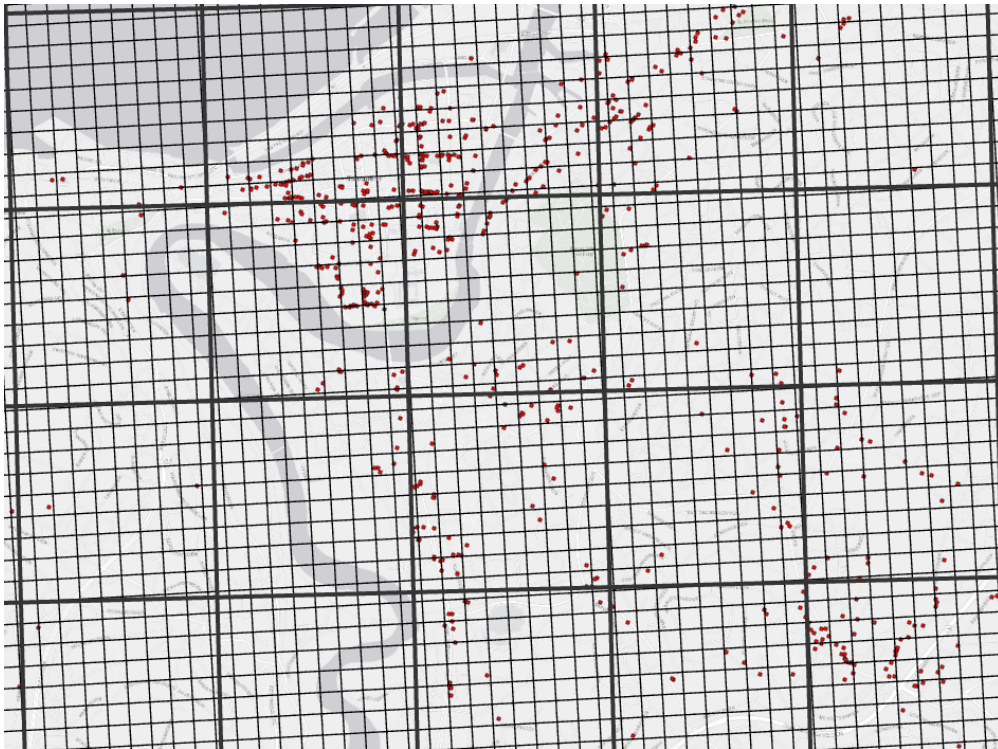
I tillegg kan en anta at disse broene er bygd for å unngå oppsamling av store mengder overvann, og at det derfor ikke vil være det mest relevante stedet for en bluespot-modellering. De tre punktene som befant seg på broene ble derfor fjernet. Ettersom antallet observasjoner var lavt, vil dette trolig ikke ha hatt stor innvirkning på resultatet. Problemet viser likevel en av utfordringene en kan møte på ved modellering av bluespots i urbane områder med broer og andre objekt som ikke inkluderes i terrengmodellen.

4.12 Detaljert sortering

Den detaljerte sorteringen krevde en grundigere gjennomgang av observasjonene enn den grove sorteringen. Det var nødvendig med en manuell gjennomgang av hver observasjon for å sortere bort uriktige observasjoner. I tillegg ble denne sorteringen viktig for å kunne se sammenheng mellom observasjonene, blant annet for å finne ut hvilke vannoppsamlinger som var observert av flere deltakere. I sorteringen ble Google Street View brukt som verktøy for å stedfeste den reelle lokasjonen til vannoppsamlingene. Observasjonene med feil posisjon ble erstattet med et nytt punktlag.

4.12.1 Rutenett

I enkelte områder var tettheten av observasjoner høyere, som vist i varmekartet i resultatkapittelet (figur 42). Dette krevde at det ble etablert et system for gjennomgangen. Nye felt i attributt Tabellen ble brukt til å holde kontroll på egenskaper ved observasjonene samt oversikt over hvilke punkter som var gjennomgått og hvilke som ikke var det. Det ble laget to rutenett av ulik størrelse for studieområdet.



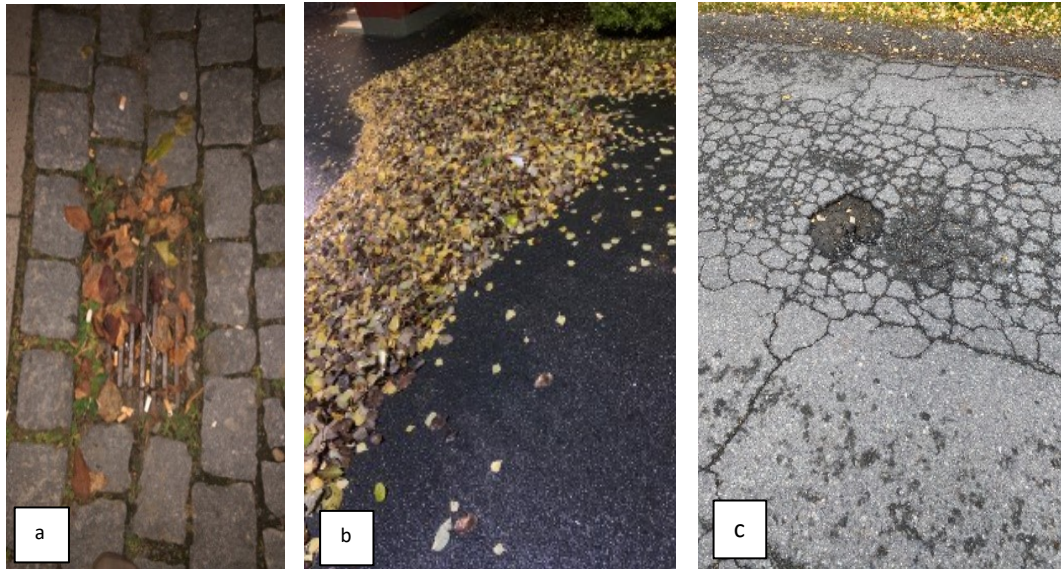
Figur 20: To rutenett med ulik størrelse ble brukt til sortering av observasjoner

Figur 20 viser de to rutenettene som ble brukt til å holde kontroll på observasjonene i studieområdet. Det minste rutenettet besto av ruter på 40 m^2 og ble hovedsakelig brukt til å holde kontroll på de observasjonene som befant seg i områder med høy tetthet av observasjoner. Det andre rutenettet besto av ruter på 400 m^2 og ble brukt til å holde kontroll på hvilke områder av studieområdet som var gjennomgått og ikke. Rutenettene skulle sikre at gjennomgangen ble gjort på en strukturert måte for å sikre at ingen observasjoner ble utelatt. Etter hvert som rutene ble gjennomgått ble symbologien endret slik at det var lett å holde kontroll på hvilke områder som gjensto.

Ved bruk av rutenett til sortering var det lettere å finne likheter mellom observasjonene, for eksempel ulike deltakers observasjoner av samme vannoppsamling. I tillegg gjorde rutenettene det enklere å finne observasjoner som var posisjonert feil. Rutenettet gjorde det mulig å gå gjennom observasjonene på en strukturert måte samtidig som de ble studert romlig ved hjelp av Ortofoto som bakgrunnskart. Attributt Tabellen er bygd opp etter Objekt-ID, som vil si at observasjonene ligger i den rekkefølgen de ble registrert og ikke etter posisjon i studieområdet. Gjennomgang i attributt Tabellen ga derfor ikke samme mulighet til å oppdage observasjoner som var plassert feil fordi tabellen som nevnt manglet det romlige perspektivet.

4.12.2 Fjerne observasjoner der bildet ikke viser vann

Totalt 176 observasjoner hadde bilder som ikke viste vann, og disse ble derfor fjernet. Denne sorteringen ble gjort ved en manuell gjennomgang av hvert enkelt bilde.



Figur 21: Bilder som ikke viste overvann: a) tett sluk, b) stor ansamling løv og c) ødelagt asfalt

Som vist i figur 21 ser en at bildene ikke viser vann, men sluk som er tett av løv, stor ansamling av løv på veien og en ødelagt gangsti hvor det er sprekker i asfalten. Disse observasjonene hadde ingen betydning for oppgavens problemstilling og ble sett på som potensielle feilkilder i den videre analysen. For å sikre så objektiv sortering som mulig ble alle bilder som viste overvann beholdt, uavhengig av vannmengden. Dette fordi bildene er tatt over en tidsperiode på seks måneder, og i tillegg på ulike dager og tider på døgnet. Selv om bildet viser små mengder overvann, kan den potensielle vannsopsamlingen være større på dager med mer nedbør. I tillegg må en ta i betraktning at tidsperioden mellom nedbørshendelsen og billedtakingen varierer. Enkelte bilder kan være tatt rett etter nedbørshendelsen, mens andre bilder kan være tatt flere timer etter at det har vært nedbør. Figur 22 viser eksempel på vannopsamlinger av ulik størrelse.



Figur 22: Vannoppsamlinger av ulik størrelse

Figur 23 viser tre observasjoner av den samme vannoppsamlingen sendt inn av ulike deltakere på ulike tidspunkt. Figuren viser at den samme vannoppsamlingen kan ha ulik størrelse avhengig av når bildet er tatt. Det må også nevnes at bildene er tatt fra ulike vinkler og har ulik bildekvalitet, noe som igjen gjør at tolkningen av vannoppsamlingens størrelse vil variere. Her vil også den subjektive meningen til den som foretar vurderingen spille en rolle, dette vil diskuteres i kapittel 6.



Figur 23: Den samme vannoppsamlingen observert av flere deltakere på ulike tidspunkt viser forskjell i størrelse

4.12.3 Fjerne observasjoner som ikke viser omgivelser

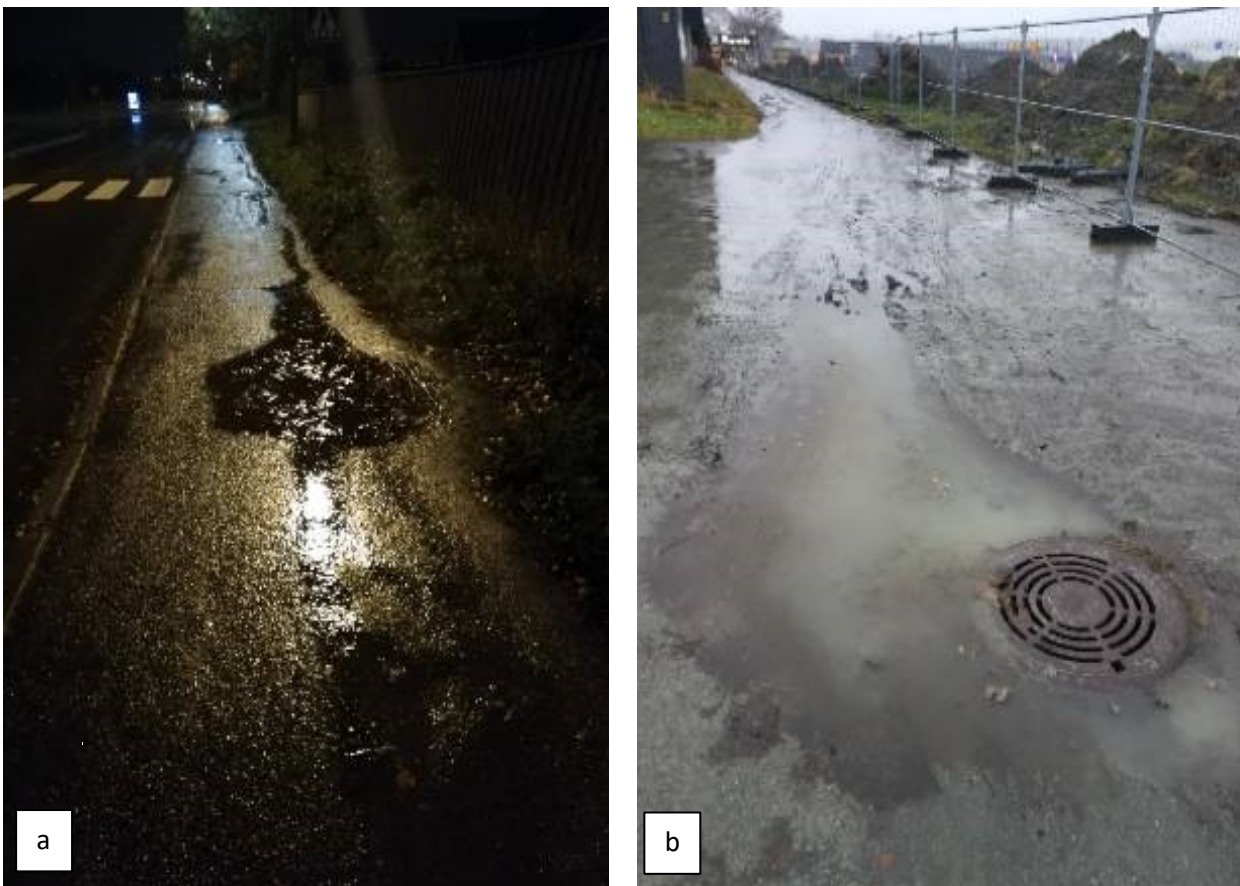
Flere av bildene fra deltakere viste ikke omgivelser, men kun selve vannoppsamlingen (se figur 24). Her ser en at deltakerne har tatt bildet med en vinkel som gjør at omgivelsene rundt vannoppsamlingen ikke vises. Dette gjør det vanskelig å gjenkjenne hvor bildet er tatt, og det blir derfor ikke mulig å vurdere om posisjonen er riktig eller ikke. Siden deltakerne ikke trengte å markere riktig posisjon da bildet ble tatt, men hadde mulighet til å markere posisjonen i ettertid på et kart, var det stor usikkerhet knyttet til disse observasjonene. Disse observasjonene ble derfor fjernet. Hvilken mulig innvirkning dette kan ha hatt for den videre analysen og resultatet vil diskuteres i kapittel 6.



Figur 24: Bilder fra deltakere som ikke viste omgivelser

4.12.4 Fjerne observasjoner der omgivelsene er vanskelig å identifisere

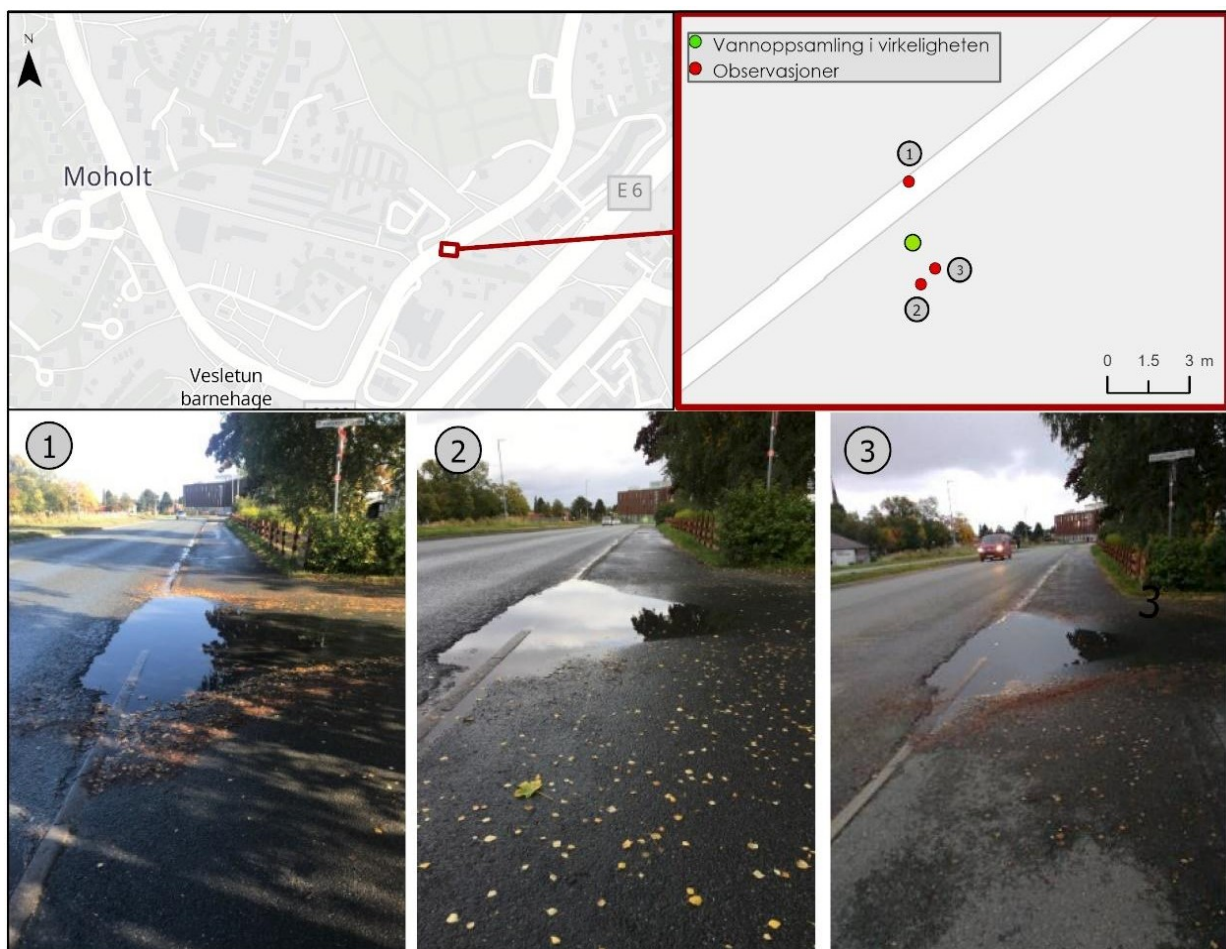
Totalt 133 observasjoner hadde bilder der lokasjonen var vanskelig å fastslå basert på bildet. Dette fordi bildet ikke var av god nok kvalitet, som vist i figur 25a er bildet mørkt og derfor vanskelig å tolke. Det var også områder som hadde blitt endret siden bildet ble tatt, dette kunne typisk være byggeplasser (Figur 25b). Disse bildene kunne ikke identifiseres ved bruk av ortofoto og Google Street View, dette går nærmere inn på i avsnitt 4.12.6.



Figur 25: Bilder fra deltakere der a) bilder var for mørkt til at omgivelsene kunne identifiseres og b) bildet viste en byggeplass som ikke lenger var gjenkjennbar

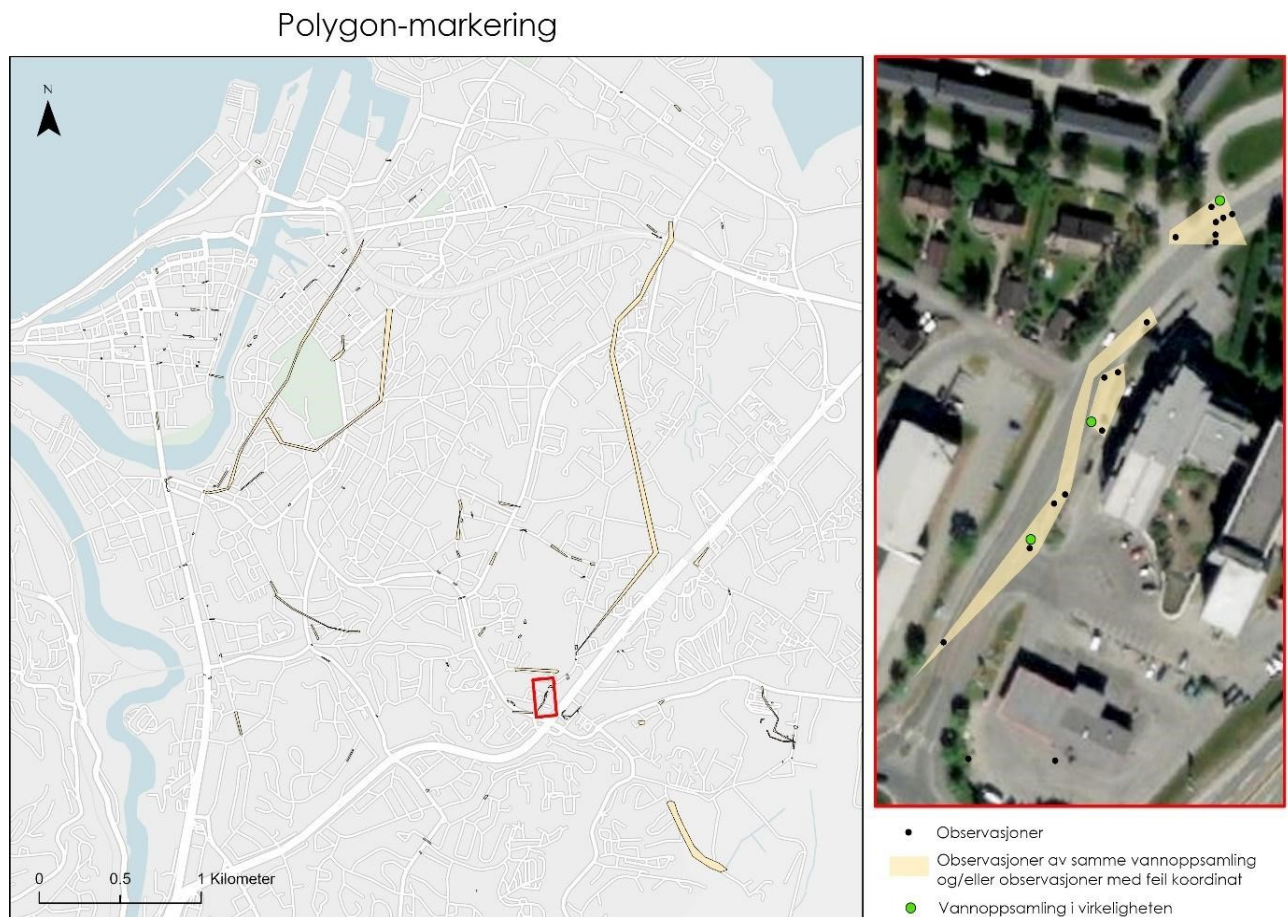
4.12.5 Fjerne observasjoner med feil posisjon

Totalt 263 observasjoner hadde feil posisjon sammenlignet med vannoppsamlingens reelle posisjon. Flere steder i studieområdet hadde ulike deltakere gjort observasjoner av samme vannoppsamling. Disse observasjonene hadde ulik koordinat - i enkelte tilfeller var det snakk om noen centimeters avvik, mens i andre tilfeller var det flere hundre meter. Polygon-markering ble brukt til å markere observasjoner av samme vannoppsamling og til å knytte observasjoner med feil posisjon til den reelle posisjonen. Figur 26 viser et tilfelle der den samme vannoppsamlingen er observert av tre ulike deltakere. Bildene er tatt fra samme sted med samme vinkel, men har likevel noe ulik posisjon - her vist med røde punkt i kartutsnittet.



Figur 26: Tre observasjoner av den samme vannoppsamlingen har noe ulikt koordinat

Som figur 27 viser ble det laget et polygonlag for å holde kontroll på hvilke observasjoner som viste samme vannoppsamling. Etter hvert som observasjonene ble gjennomgått ble det tegnet polygon for å markere observasjoner av samme vannoppsamling. Det samme polygonlaget ble også brukt til å sammenkoble feilplasserte observasjoner med vannoppsamlingens reelle posisjon. Som tidligere nevnt ble den reelle posisjonen markert med et nytt punktlag.

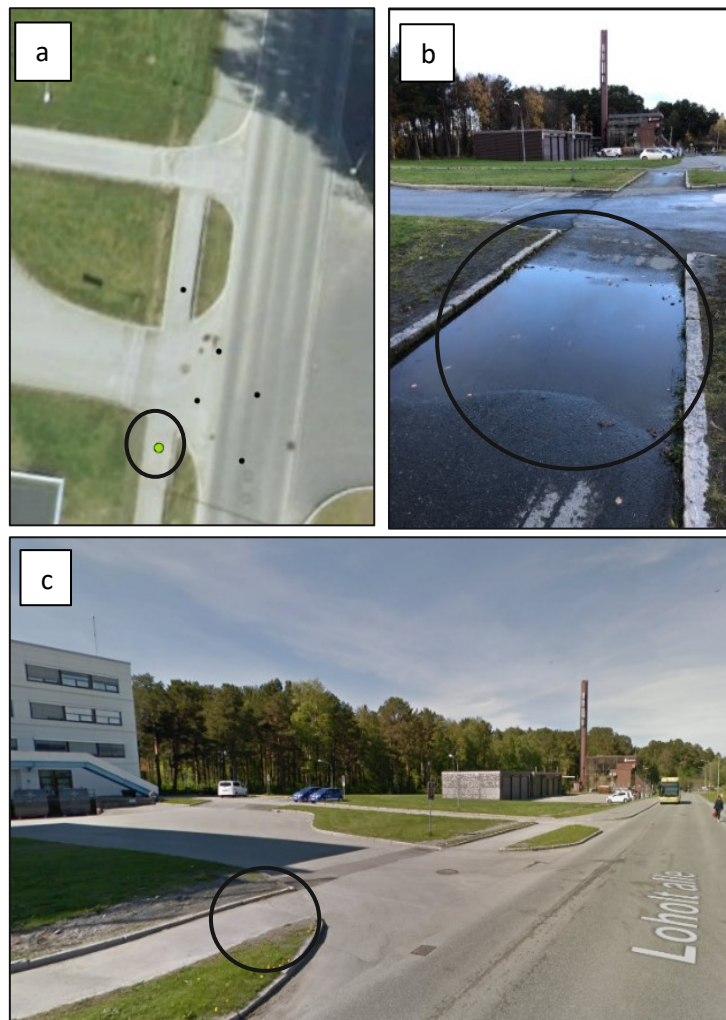


Figur 27: Kartet viser polygoner som ble brukt til å markere observasjoner av samme vannoppsamling og knytte observasjoner med feil posisjon til vannoppsamlingens reelle posisjon

Figur 27 viser hvordan polygonmarkering ble brukt flere steder i studieområdet. Utsnittskartet viser et mindre område på Moholt, hvor en ser at observasjoner fra ulike deltakere av samme vannoppsamling har ulik koordinat. For enkelte vannoppsamlinger er observasjonene registrert nært vannoppsamlingens reelle posisjon, mens andre observasjoner har stor avstand, noe en kan se i kartet. Polygonmarkeringene ble derfor en viktig måte å få oversikt over hvilke observasjoner som viste samme vannoppsamling og hvilke observasjoner som var lokalisert feil.

4.12.6 Finne vannoppsamlingens reelle posisjon

For å finne den reelle posisjonen til observasjonene ble bildene og lokasjonen til de opprinnelige observasjonene brukt. Dette er likt det som ble gjort i studien til Yu et al. (2016) hvor Google maps ble brukt til å tolke posisjonen til flomobservasjonen. Studien viser at metoden kan ha ført til feil som påvirket resultatet. For enkelte observasjoner var likevel ikke den opprinnelige lokasjonen til stor hjelp, da denne kunne være fastsatt flere tusen meter unna vannoppsamlingens virkelige lokasjon. For de observasjonene hvor bildet viser omgivelser ble det vurdert om observasjonen var lokalisert riktig i forhold til der vannoppsamlingen på bildet var. Her ble det laget et nytt punktlag som skulle erstatte observasjonen dersom lokasjonen viste seg å være feil. Dette punktlaget viste vannoppsamlingenes virkelige posisjon.



Figur 28: Vannoppsamlingens reelle posisjon ble bestemt ved bruk av a) observasjoner i ArcGIS Pro, b) bilde fra deltakerne og c) Google Street View

For å vurdere om observasjonen var riktig lokalisert ble det brukt topografisk kart og satellittbilde, samt Google Street View. Street View gjør det mulig å se omgivelsene i panorama, noe som gir et bedre bilde av omgivelsene. Det er lettere å se hvor objekt er lokalisert og avstandene mellom dem, og det gir mulighet til å oppleve et sted på en mer realistisk måte enn et todimensjonalt bilde gir. Figur 28c viser hvordan Street View kan se ut for brukeren. De to øverste bildene viser et todimensjonalt ortofoto og et todimensjonalt fotografi.

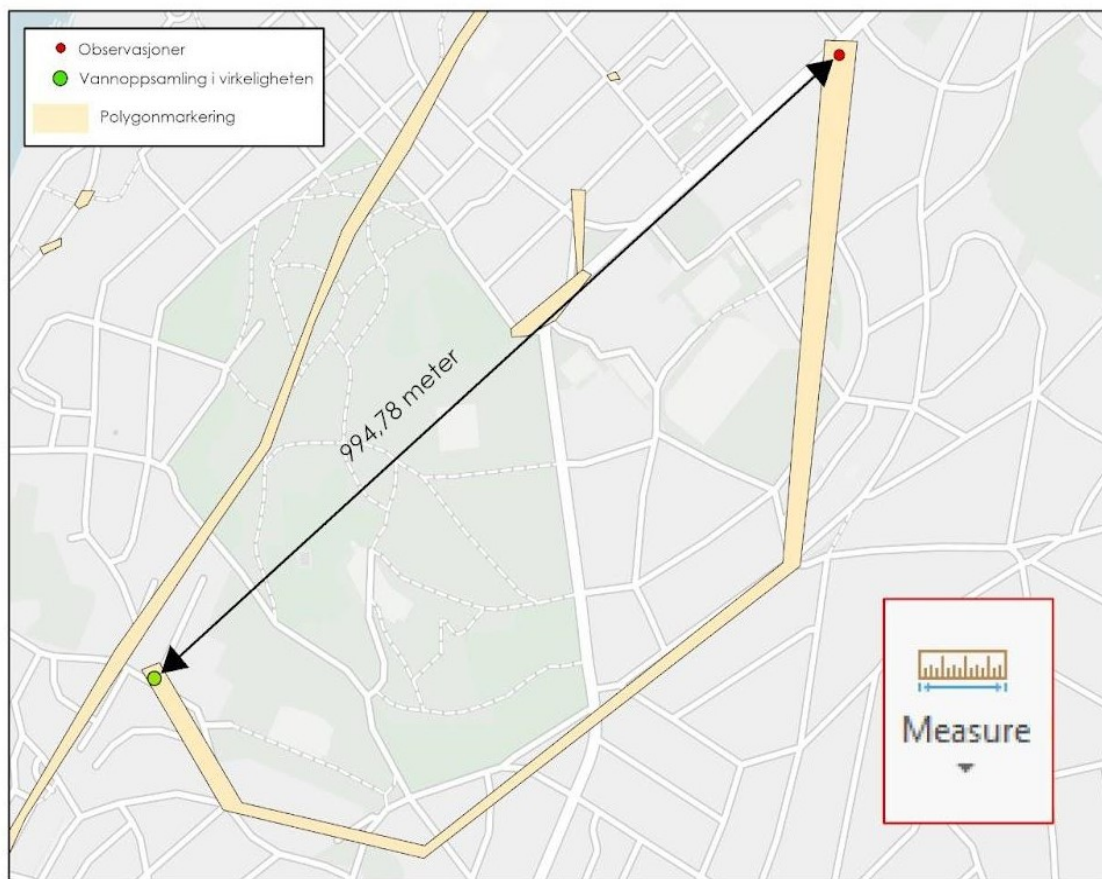
For punkter med åpenbart feilregistrert lokasjon ble koordinatene erstattet med nye som stemte med vannoppsamlingens virkelige lokasjon. Denne lokasjonen var som nevnt basert på bildene fra deltakerne og Google Street View. Observasjonene med feil lokasjon ble fjernet, men informasjon om antall observasjoner ble lagret som informasjon i det nye punktlaget. Denne informasjonen var viktig for den senere valideringen, siden flere observasjoner av samme vannoppsamling styrket observasjonen, spesielt hvis vannoppsamlingen var observert på ulike tidspunkt.

4.13 Avstandsmål

I GIS er distansen mellom to objekt målt som den korteste avstanden mellom dem (Esri, 2021b). I denne masteroppgaven er avstandsmåling brukt som metode for å beskrive lokasjonen til folkeinnsamlet data sammenlignet med modellerte forsenkninger, både etter grov og detaljert sortering. Dette var interessant for å se hvilken påvirkning sorteringen hadde på avstanden mellom observasjonene og forsenkningene. I tillegg har avstandsmåling blitt brukt etter detaljert sortering til å måle avstanden fra observasjoner til vannoppsamlingens reelle posisjon. Dette var viktig for å kunne si noe om nøyaktigheten og presisjonen til datasettet.

4.13.1 Manuell avstandsmåling

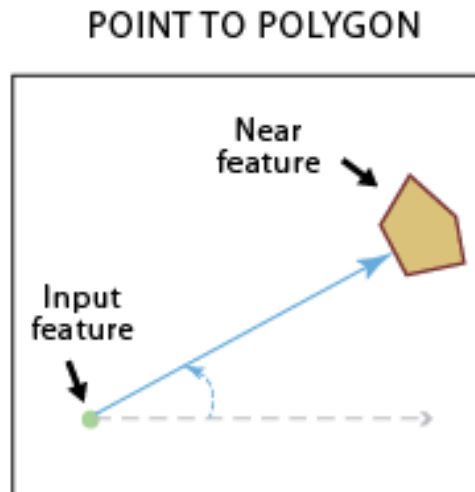
Det ble utført manuell avstandsmåling for å finne avstanden mellom observasjoner med lokasjonsfeil og punktet som representerte vannoppsamlingens reelle posisjon. Dette gjøres ved bruk av linjal i ArcGIS Pro, her må en manuelt trykke på de punktene en ønsker å måle avstanden mellom. Manuell avstandsmåling ble brukt fordi verken romlig utvalg eller verktøyet *near* hadde ført til riktig resultat her. Det nærmeste punktet er ikke alltid det en ønsker å måle avstand fra. Figur 29 viser hvordan manuell avstandsmåling foregår. Her trekkes en linje fra observasjonsdataen til vannoppsamlingens reelle posisjon.



Figur 29: Manuell avstandsmåling ble gjort mellom observasjoner og vannoppsamlingens reelle posisjon

4.13.2 Verktøyet *near*

Den tredje formen for avstandsmåling er bruk av verktøyet *generate near table* som er et romlig verktøy i GIS som automatisk kalkulerer distansen mellom to inngangslag. Ved å bruke dette verktøyet kan en effektivt finne distansen mellom et stort antall objekt. I denne masteroppgaven er verktøyet brukt som metode for å finne avstanden observasjoner har til nærmeste forsenkning, både etter grov sortering og etter detaljert sortering. Formålet var å se hvordan avstanden mellom observasjoner og forsenkninger forandret seg ved ulike grader av sortering. Dette for å kunne trekke konklusjoner rundt betydningen av sortering av folkeinnsamlet data. Figur 30 viser hvordan *near* fungerer. Her spesifiseres det at en ønsker å finne nærmeste polygon fra punktlaget. I denne masteroppgaven var polygonene forsenkninger, mens punktlaget var observasjoner fra deltakerne.



Figur 30: I verktøyet *generate near table* måles avstanden fra inngangsdataen til det nærmeste spesifiserte objektet (Esri 2021b)

4.13.3 Romlig utvalg basert på avstand og lokasjon

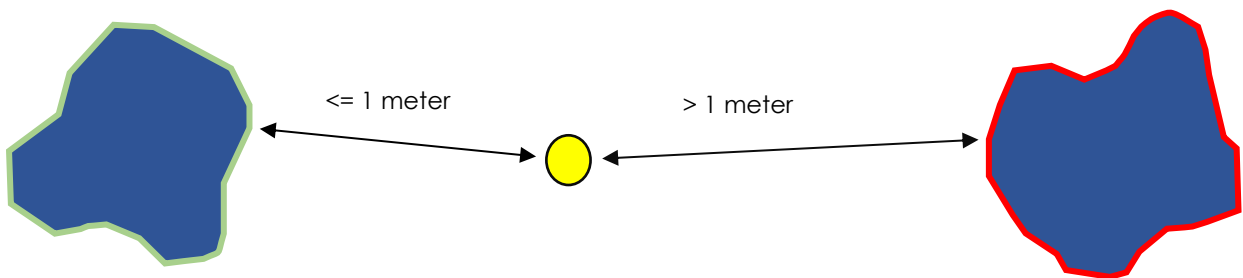
Det ble gjort romlige utvalg basert på objekters relative posisjon til hverandre. I denne masteroppgaven har romlig utvalg basert på avstand blitt brukt både til sortering og til å bekrefte forsøkninger. I sorteringen var det nyttig å kunne gjøre romlige utvalg for å sortere bort observasjoner utenfor det opprinnelige studieområdet og observasjoner lokalisert i bygg. I valideringen av forsøkninger ble romlige utvalg brukt til å velge ut forsøkninger to meter eller nærmere fra en observasjon, noe som vil bli omtalt nærmere i avsnittet som kommer.

4.14 Bekrefte at modellerte forsøkninger er reelle vannoppsamlinger

Valideringen i denne masteroppgaven handler om å bruke folkeinnsamlet data om overvann til å bekrefte eller avkrefte om modellerte forsøkninger er reelle vannoppsamlinger. I studien til Yu et al. (2016) ble folkeinnsamlede observasjoner om flomhendelser brukt til å validere modellerte forsøkninger. Her ble forsøkningene bekreftet dersom de hadde en overlapp med observasjonsdataen. I studien til Yu et al. (2016) må det nevnes at de modellerte forsøkningene var større enn de modellerte forsøkningene i denne masteroppgaven, noe som trolig gjorde at sjansene for overlapp var større.

4.14.1 Én meter fra observasjon

I denne masteroppgaven måtte det beregnes en feilmargin for å ta hensyn til mulige posisjonsfeil. Disse posisjonsfeilene er både knyttet til usikkerhet rundt observasjoner som ble bekreftet å ha riktig lokasjon, men også mulige feilkilder knyttet til fastsetting av nye posisjoner ved bruk av bilder fra deltakeren. Posisjonsbestemmelsen kan preges av subjektiviteten til den som setter punktet, noe som kan gjøre at posisjonen må antas å kunne være noe feilplassert sammenlignet med hvor vannoppsamlingen befinner seg i virkeligheten. På bakgrunn av dette ble en feilmargin på én meter vurdert som tilstrekkelig for å bekrefte om forsøkningene er reelle vannoppsamlinger. Det er mulig denne feilmarginen er for lav eller for høy, noe som vil diskuteres i kapittel 6.



Figur 31: Forsøkninger en meter eller nærmere blir bekreftet å være reelle vannoppsamlinger

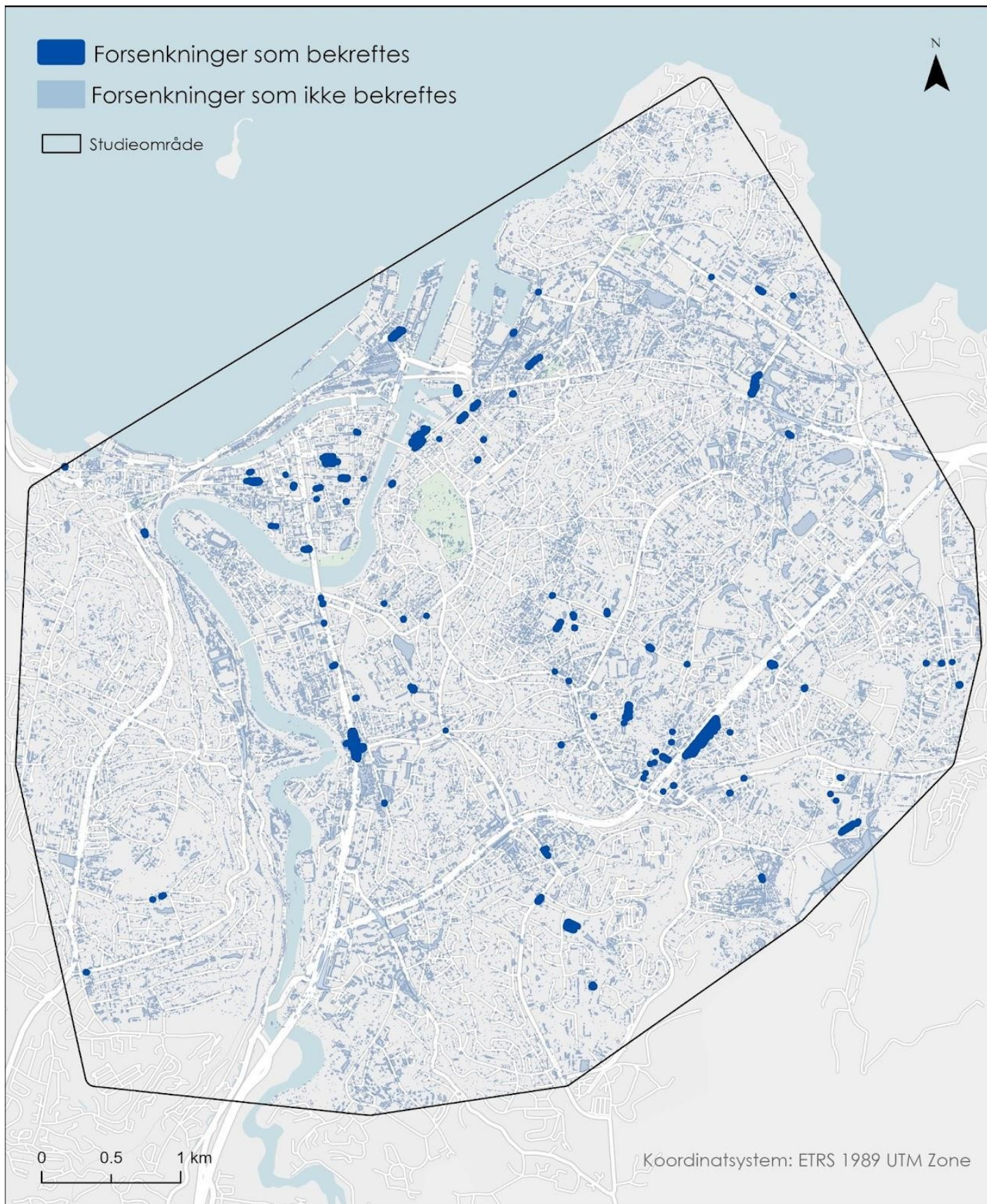
Figur 31 viser en illustrasjon av hvordan bekrefting av forsøkninger fungerte. Det ble gjort et romlig utvalg hvor kun forsøkninger som befinner seg en meter eller nærmere i avstand til en observasjon ble beholdt. Figuren viser to forsøkninger med ulik avstand til det gule observasjonspunktet. Her ser en at forsøkningen med grønt omriss bekrefte, mens den med rødt omriss ikke møter det første kravet i valideringen og derfor fjernes fra datasettet. Resultatet er et nytt datasett som inneholder bekreftede forsøkninger. I hver forsøkning lagres informasjon om hvor mange observasjoner det finnes av vannoppsamlingen, noe som senere vil bli brukt til å kunne si noe om hvilke forsøkninger som kan bekrefte med større sikkerhet enn andre.

5 Resultat

I dette kapitlet vil masteroppgavens funn presenteres. Dette er resultat fra modelleringen av forsengkninger i GIS og funn av kjennetegn for det folkeinnsamlede datasettet. Første del av kapitlet vil presentere resultat som beskriver den folkeinnsamlede datasettet. Deretter vil resultatene fra valideringen av forsengkninger presenteres. Her blir resultatene fra både grov og detaljert sortering av datasettet lagt fram. Sluttresultatet er 96 forsengkninger med overvann som er bekreftet ved bruk av folkeinnsamlet data.

5.1 Bekreftede forsengkninger

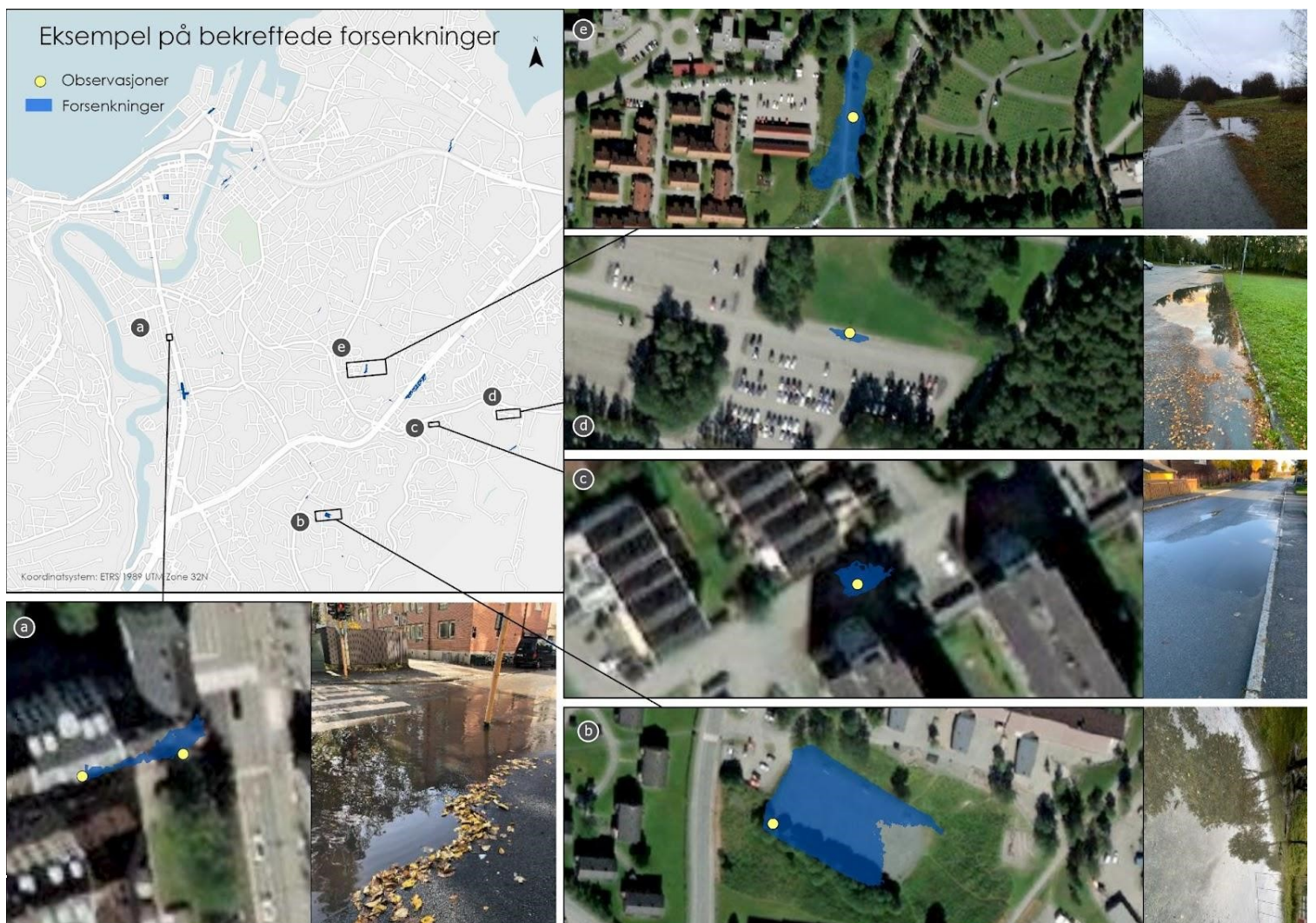
Det var totalt 96 forsengkninger som hadde en maksimumavstand på én meter til observasjonen etter den detaljerte sorteringen og derfor ble bekreftet å inneholde overvann. Som en ser i karte i figur 32 var dette en liten andel av alle forsengkningene som ble modellert i studieområdet. Det var totalt 96 149 forsengkninger som ble modellert i studieområdet. Det høye antallet skyldes at mange forsengkninger var veldig små.



Figur 32: Bekreftede forskningsområder i studieområdet

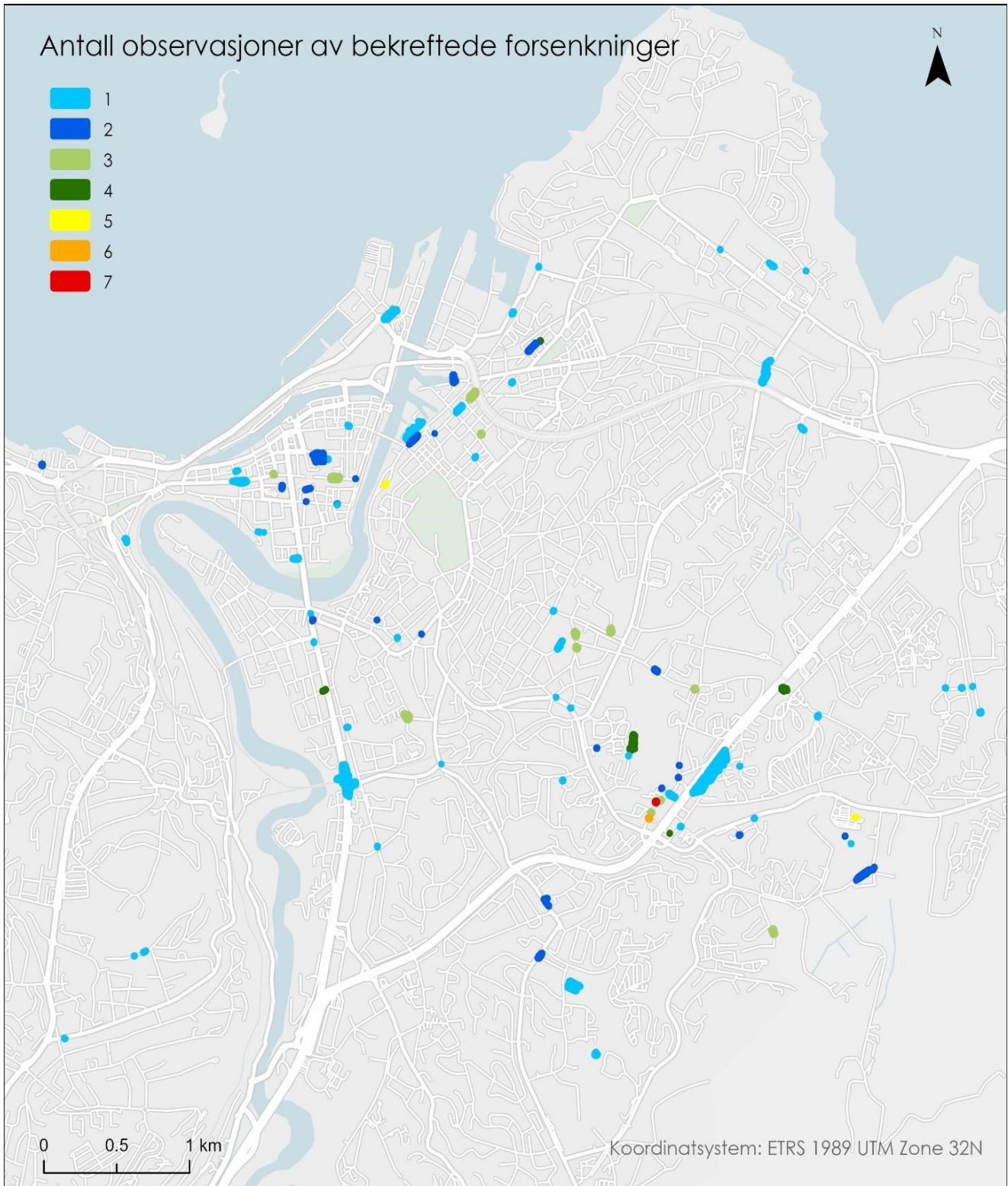
5.1.1 Eksempel på bekreftede forsengkninger

Figur 33 viser eksempel på fem forsengkninger som lå maksimum én meter fra det detaljert sorterte datasettet og derfor ble bekreftet å inneholde overvann. Kartet viser lokasjonen til de fem eksemplene. Kartutsnittene viser hvor observasjonene ble lokalisert (gult punkt) i forhold til forsengkningene (blå polygon). Ved kartutsnittet finner en bildet av vannoppsamlingen som ble brukt til å bekrefte forsengkningen.



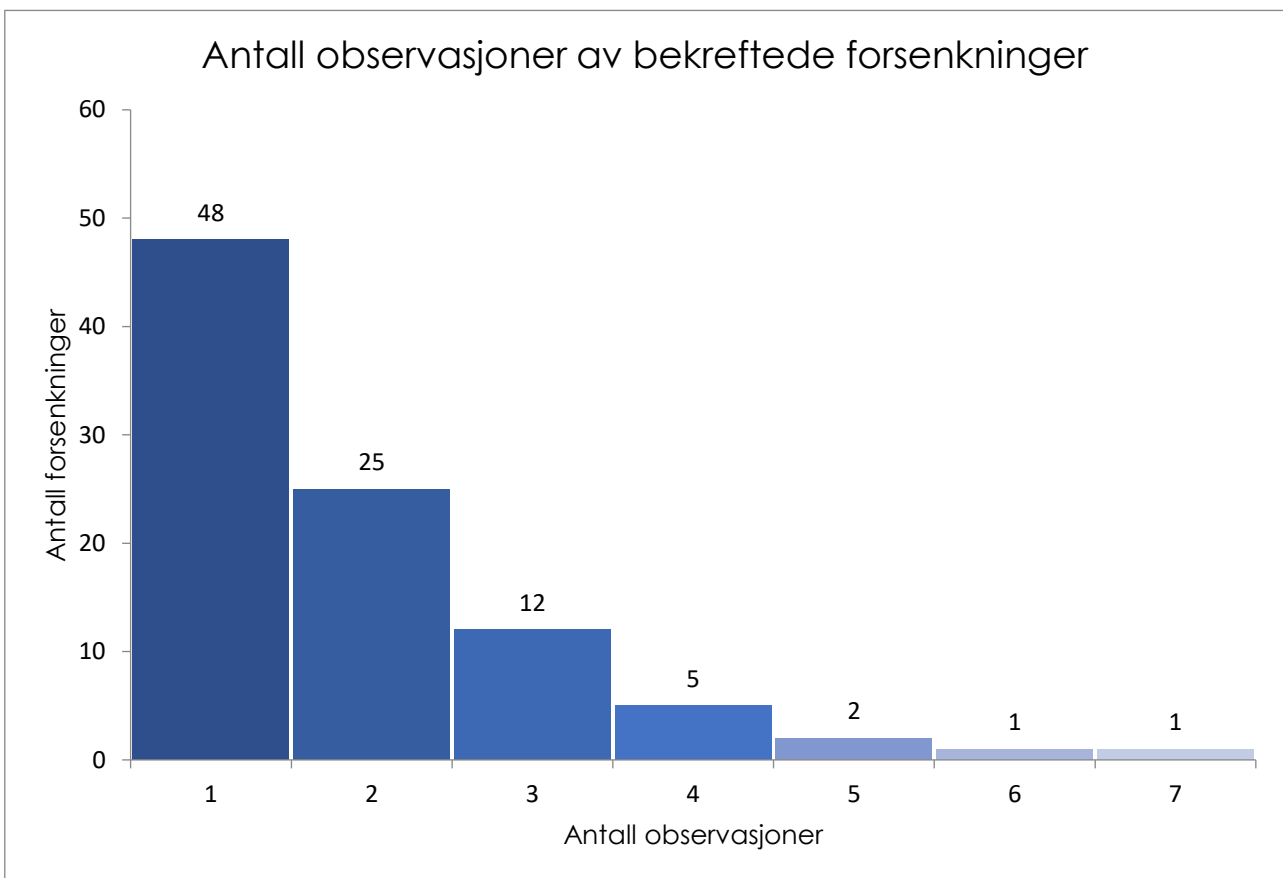
Figur 33: Eksempel på bekreftede forsengkninger

5.2 Antall observasjoner av bekreftede forsennkninger



Figur 34: Kartet viser antall observasjoner av bekreftede forsennkninger

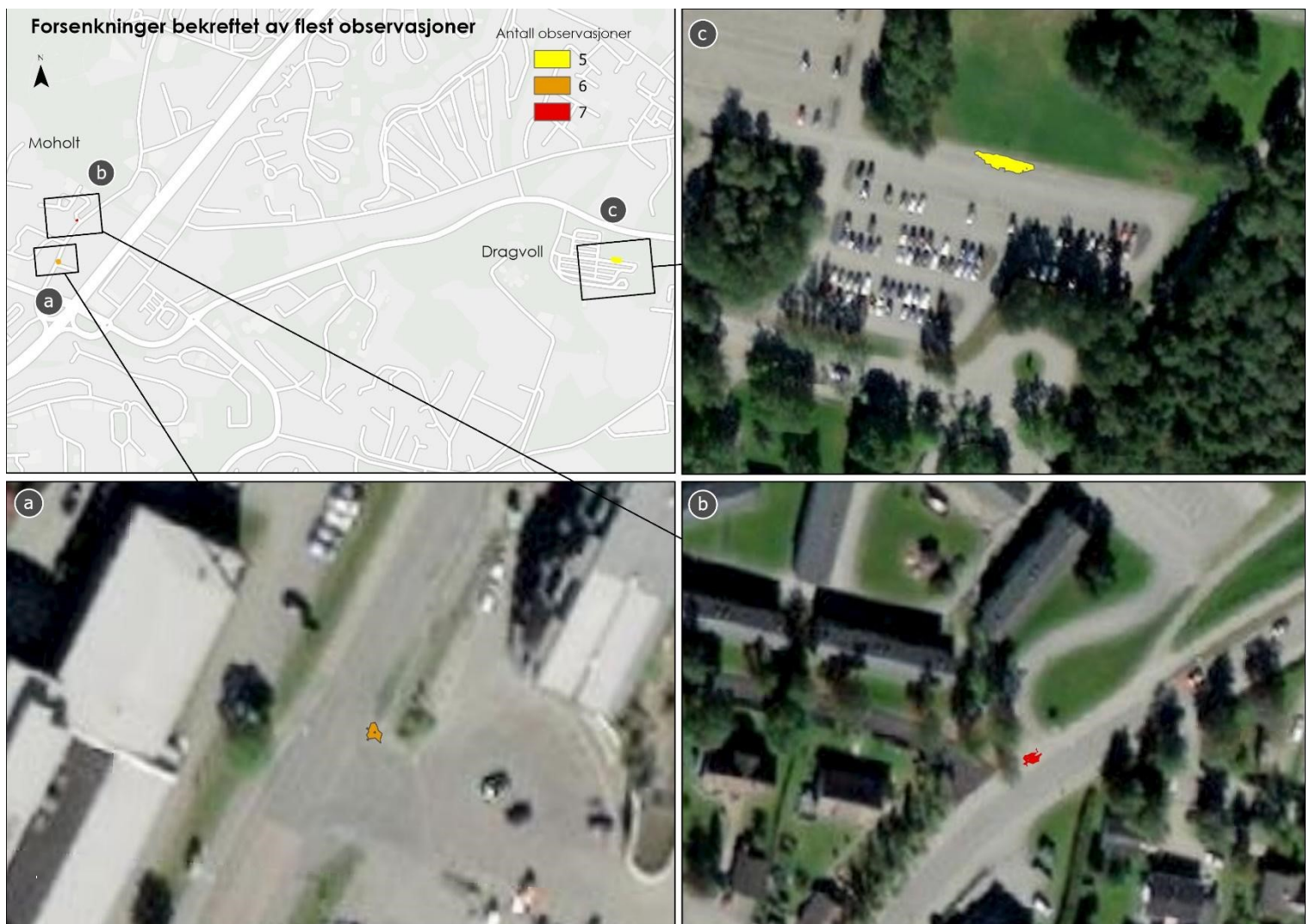
Kartet i figur 34 viser hvor mange observasjoner det er av hver bekreftede forsøknings og hvor disse ligger. Her ser en at det finnes forsøknings i studieområdet som ble observert av opptil sju deltakere (markert i kartet med rød farge). Kartet og histogrammet i figur 35 viser likevel at flesteparten av de bekreftede forsøkningsene kun ble observert én gang av hver deltaker (markert i kartet med lys blå farge). Det var 48 forsøknings som kun hadde én observasjon, to forsøknings var observert av fem deltakere, en forsøknings av seks deltakere og en forsøknings var observert av så mange som sju deltakere.



Figur 35: Histogrammet viser antall observasjoner av hver bekreftede forsøknings

5.2.1 Forsenkninger bekreftet av flest observasjoner

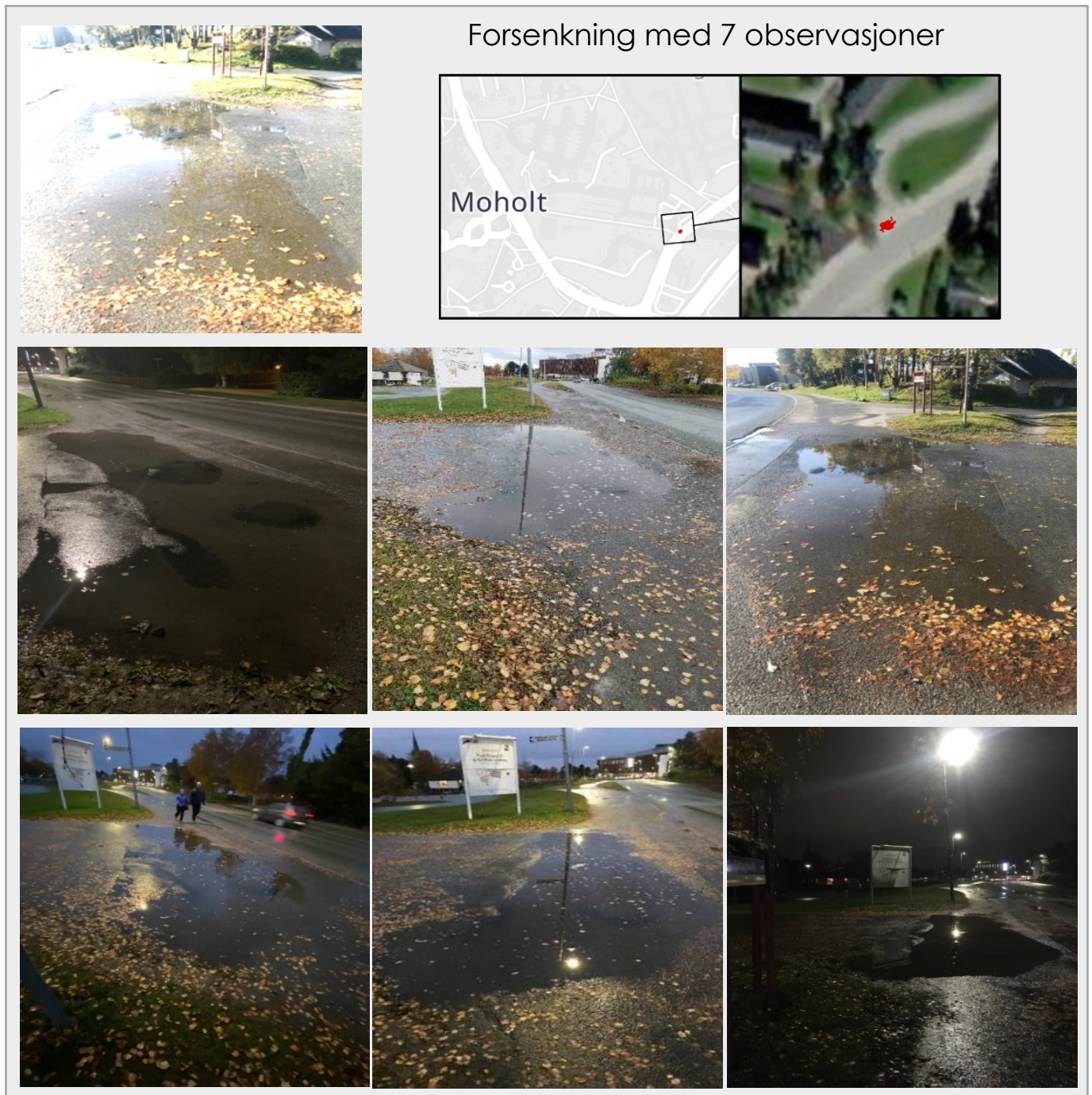
Kartet viser tre forsenkninger som ble observert av flest deltakere – forholdsvis 5 (gul polygon), 6 (oransje polygon) og 7 deltakere (rød polygon). De to forsenkningene med 6 og 7 observasjoner var lokalisert på Moholt, mens en av forsenkningene med 5 observasjoner befant seg utenfor NTNU Dragvoll.



Figur 36: Forsenkninger bekreftet av fem, seks og sju observasjoner

5.2.2 Forsenkningen med sju observasjoner

Figur 37 viser forsenkningen som hadde sju og dermed flest observasjoner av de bekreftede forsenkningene. Kartet viser lokasjonen til forsenkningen på Moholt, mens kartutsnittet til høyre viser den modellerte forsenkningen mer detaljert. De sju bildene viser den samme vannoppsamlingen sendt inn av sju ulike deltakere.



Figur 37: Forsenkning med sju observasjoner

5.3 Sorteringen

Som tidligere beskrevet var sorteringen av folkeinnsamlet data et viktig, men omfattende arbeid for å sikre best mulig kvalitet på datasettet. Grov og detaljert sortering utgjorde til sammen ni steg frem mot det endelige datasettet som ble brukt i validering av forsøkningene. Tabell x viser hvor mange observasjoner som ble fjernet i hvert sorteringssteg.

Tabell 2: Tabellen viser hvor mange observasjoner som ble fjernet i hvert steg av sorteringen

	Steg	N fjernet	N igjen
Grov sortering	Original		2554
	1. Fjerne observasjoner utenfor det opprinnelige studieområdet	827	1727
	2. Fjerne observasjoner som ikke har hendelse <i>kraftig regn</i>	159	1568
	3. Fjerne observasjoner som ligger i bygningspolygon	522	1046
	4. Fjerne observasjoner uten bilde	23	1023
	5. Fjerne observasjoner som ligger på broer	3	1020
Detaljert sortering	6. Fjerne observasjoner der bildet ikke viser overvann	176	844
	7. Fjerne observasjoner hvor bildet ikke viser omgivelser	392	452
	8. Fjerne observasjoner der omgivelsene på bildet er vanskelig å identifisere	133	319
	9. Fjerne observasjoner med feil posisjon	263	56
	10. Erstatte observasjoner med feil posisjon med nye punkt	+ 150	206
	Ferdig sortert		206

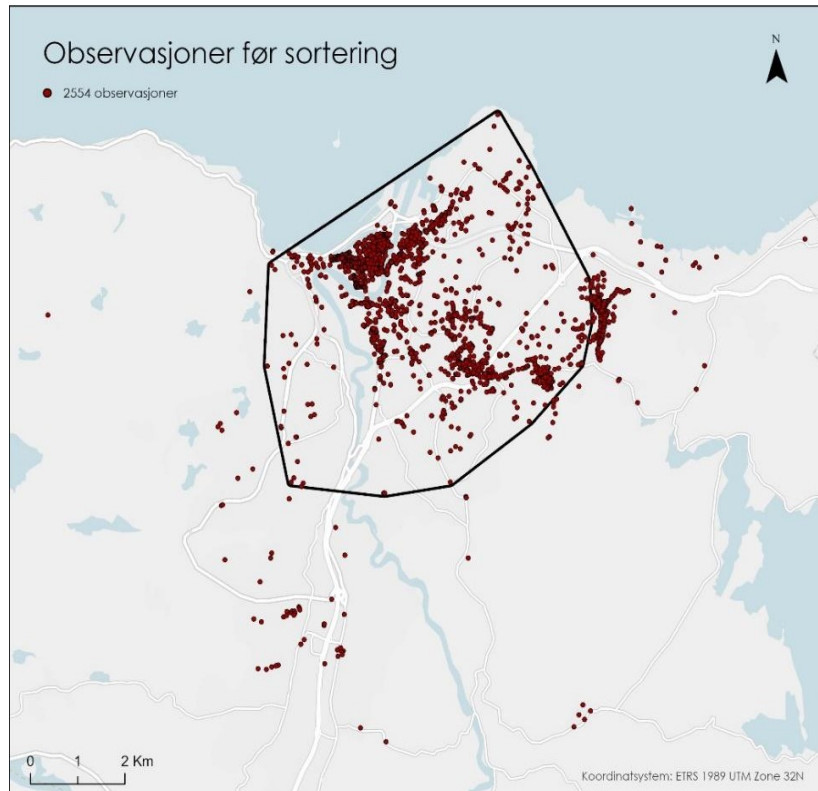
I den grove sorteringen ble det fjernet totalt 1534 observasjoner. 827 observasjoner ble fjernet fordi de lå utenfor det opprinnelige studieområdet. 159 observasjoner ble fjernet fordi de ikke hadde hendelse *kraftig regn*. Så mange som 522 observasjoner ble fjernet fordi de lå inne i bygninger. 23 observasjoner ble fjernet fordi de manglet bilde og 3 observasjoner ble fjernet fordi de var lokalisert på broer hvor det ikke ble modellert noen forsøkninger.

I den detaljerte sorteringen ble det sortert ut totalt 964 observasjoner. 176 observasjoner ble fjernet fordi bildet ikke viste overvann. 392 observasjoner ble fjernet fordi bildet ikke viste omgivelser. 133 observasjoner ble fjernet fordi omgivelsene på bildet var vanskelig å

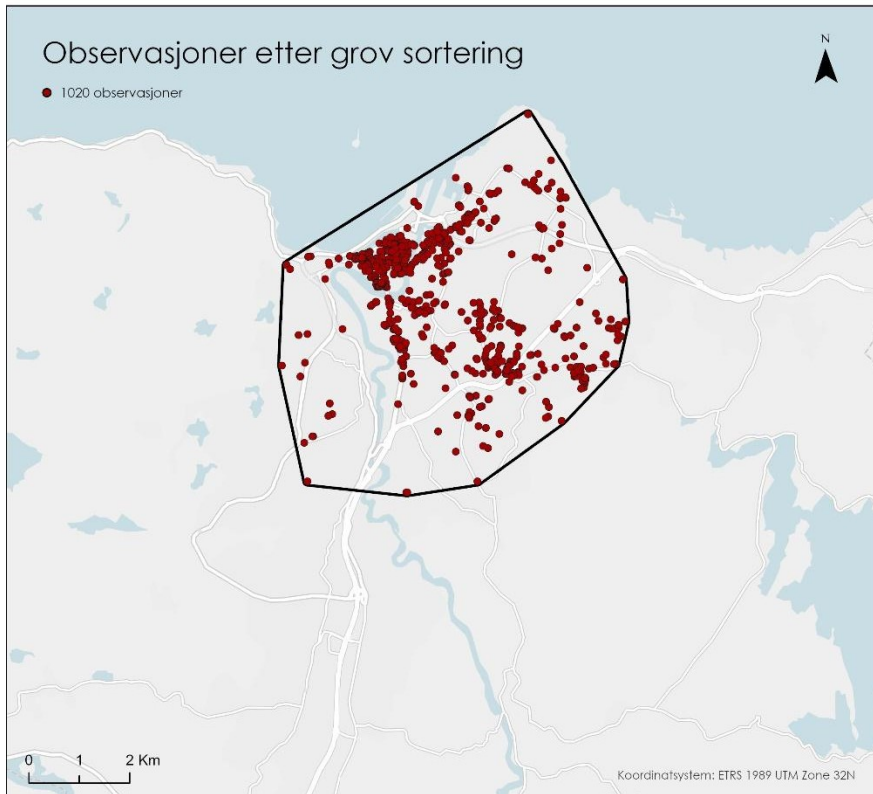
identifisere, og 263 observasjoner ble fjernet og erstattet med nye punkt fordi observasjonene hadde feil posisjon sammenlignet med vannoppsamlingens reelle posisjon. Disse observasjonene ble erstattet med 150 nye punkt. Årsaken til det er flere observasjoner som fjernes enn det som erstattes er at observasjoner av samme vannoppsamling går inn i samme punkt. Det endelige antallet observasjoner er 206.

5.4 Observasjoner i studieområdet før sortering, etter grov sortering og etter detaljert sortering

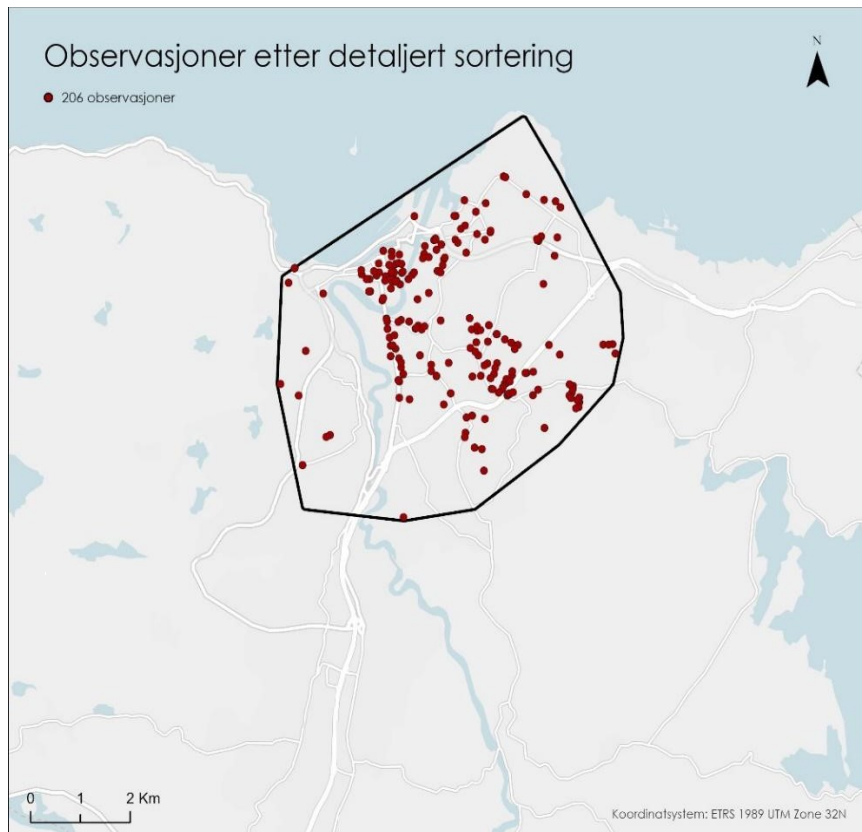
Resultatet av sorteringen er fremstilt i figur 38, 39 og 40 som viser den romlige fordelingen av observasjoner i Trondheim kommune før sortering, etter grov sortering og etter detaljert sortering. Før sortering var det totalt 2552 observasjoner i datasettet, etter grov sortering reduseres antallet til 1020, og etter den detaljerte sorteringen er det kun 206 observasjoner igjen som kan brukes i validering av forsenkningene.



Figur 38: 2554 observasjoner før sortering



Figur 39: 1020 observasjoner etter grov sortering

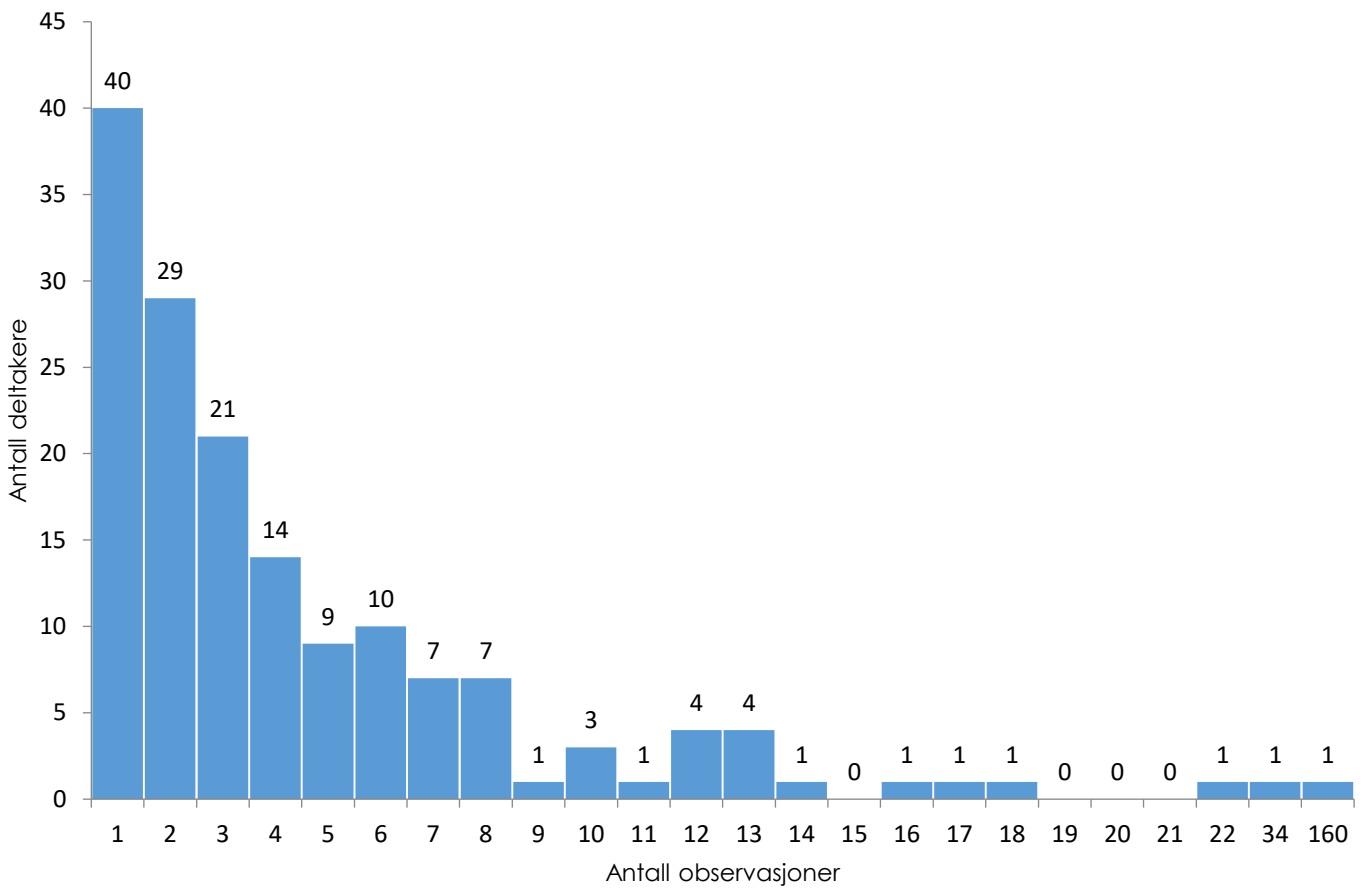


Figur 40: 206 observasjoner etter detaljert sortering

5.5 Antall observasjoner per deltaker

Som kartene over viste var det stor spredning i observasjonene. Det var også stor variasjon i hvor mange observasjoner som ble registrert per deltaker. Som figur 41 viser så har flest deltakere (40) registrert inn kun én observasjon. Det vanligste var at deltakerne registrerte mellom 1-8 observasjoner hver. Tre deltakere skiller seg ut med mange observasjoner - en har registrert 22 observasjoner, en 34 observasjoner og en deltaker har registrert hele 160 observasjoner.

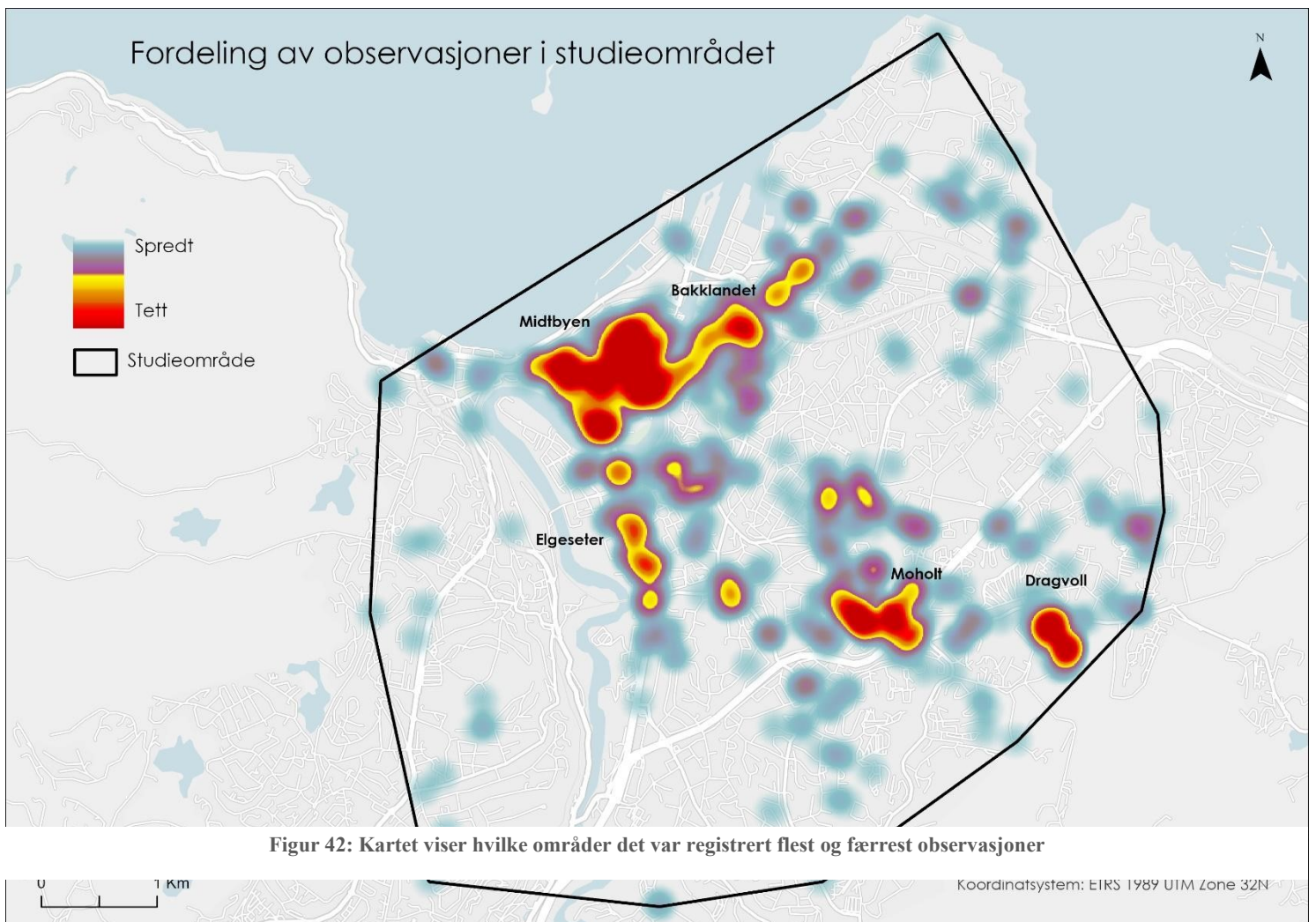
Antall observasjoner per deltaker



Figur 41: Histogrammet viser hvor mange observasjoner som ble registrert av hver deltaker

5.6 Fordeling av observasjoner i studieområdet

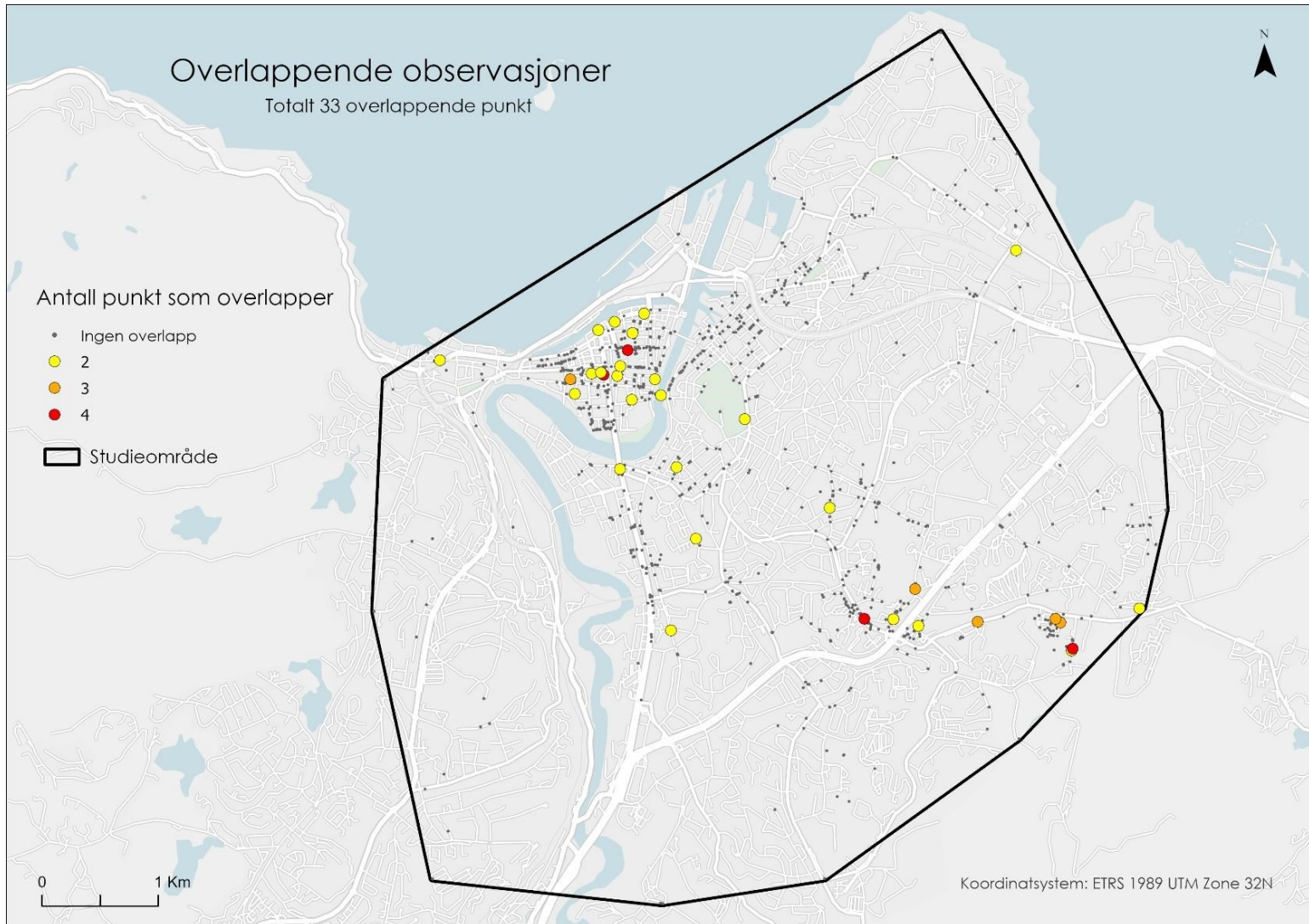
Varmekartet viser fordelingen av observasjoner i studieområdet etter den grove sorteringen. Rød farge viser områder med flest observasjoner, mens blå farge indikerer få observasjoner. De fleste observasjonene ble registrert i Midtbyen, men i områdene Bakklandet/Møllenberg, Elgeseter, Moholt og Dragvoll var det også mange observasjoner.



5.7 Observasjoner med samme koordinat

For 33 observasjoner i studieområdet var det overlapp, det vil si at det var registrert flere observasjoner i samme koordinat. Flestparten av de overlappende observasjonene var

registrert av ulike personer, men noen var også registrert av samme deltaker. På det meste var det fire observasjoner i samme koordinat.



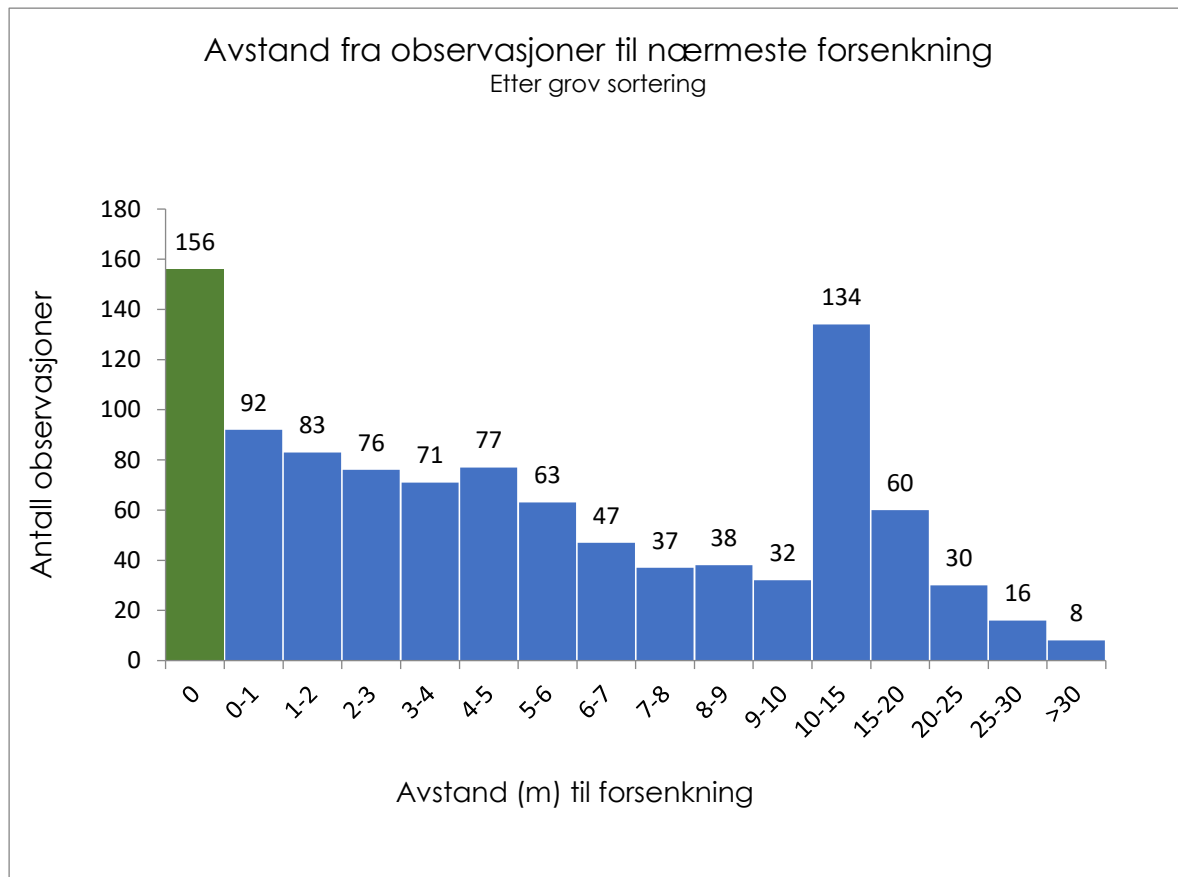
Figur 43: Kartet viser hvor i studieområdet det ble funnet overlappende observasjoner

5.8 Avstand fra observasjoner til nærmeste forsenkning

Ved bruk av verktøyet *near* var det mulig å finne avstanden fra observasjonene til nærmeste forsenkning. Dette ble gjort for å kunne gi en beskrivelse av forholdet mellom observasjoner og forsenkninger.

5.8.1 Grov sortering

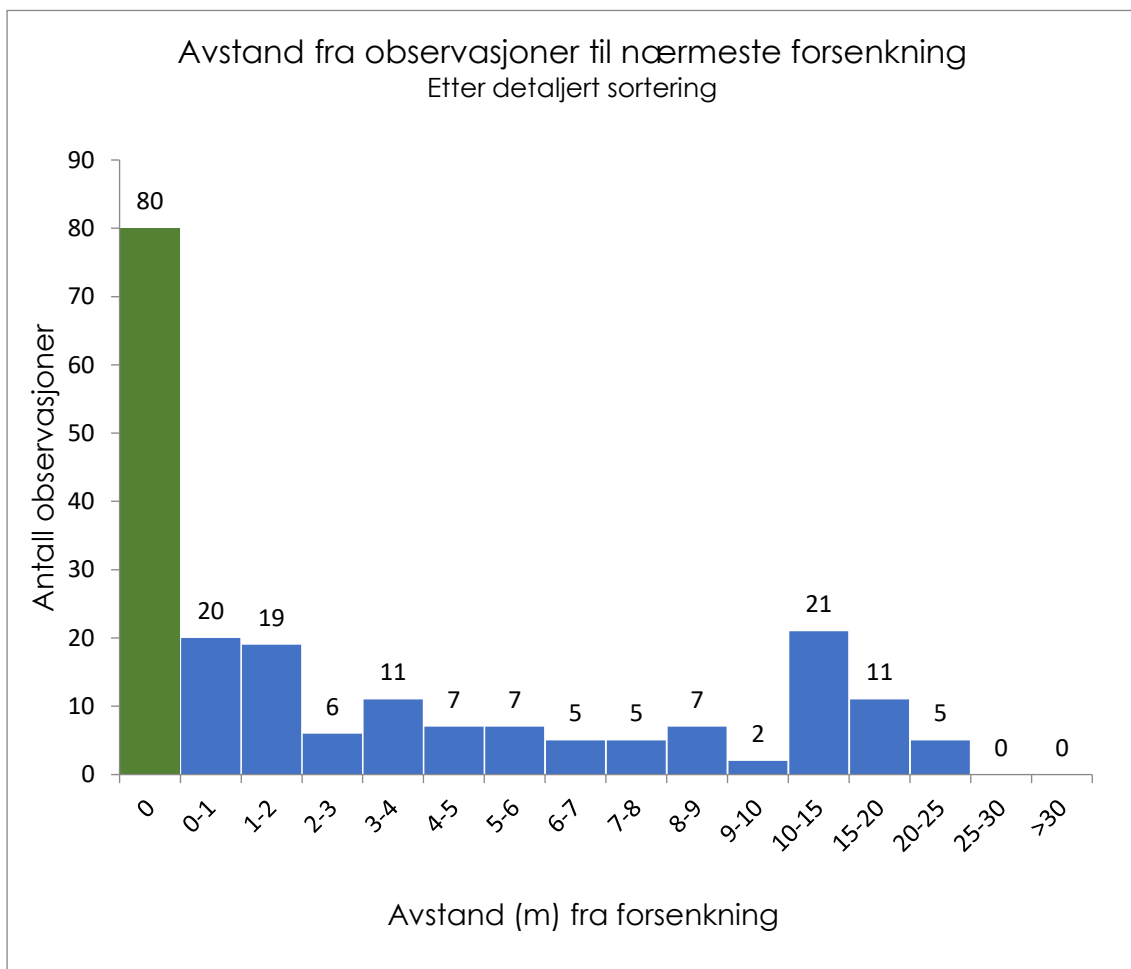
Figur 44 viser avstanden fra observasjoner til nærmeste forsenkning etter at den grove sorteringen var gjennomført og det gjenstod totalt 1020 observasjoner. Gjennomsnittsavstanden var 6,9 meter fra observasjonene til nærmeste forsenkning. Det var 156 observasjoner som overlappet med forsenkninger - avstanden var null meter (se grønn søyle). Dette er 15 prosent av observasjonene (totalt 1020). Den største avstanden som ble registrert var 413 meter til nærmeste forsenkning.



Figur 44: Histogrammet viser avstanden fra observasjoner til nærmeste forsenkning etter den grove sorteringen. Grønn søyle viser observasjoner som overlapper med forsenkninger

5.8.2 Detaljert sortering

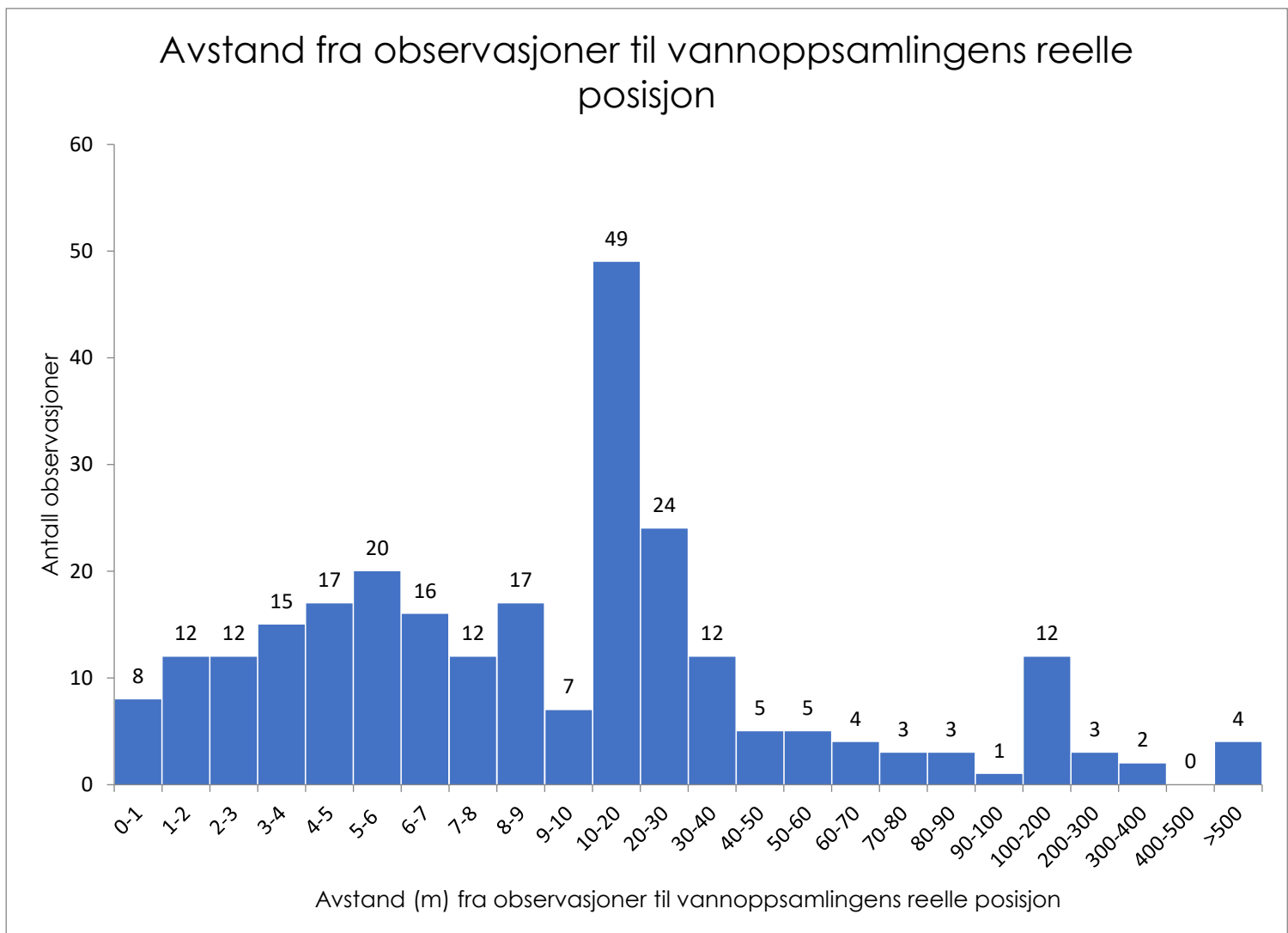
Histogrammet viser avstanden fra observasjoner til nærmeste forsenkning etter at den detaljerte sorteringen var gjennomført. Etter at den detaljerte sortering var gjennomført bestod det gjenværende datasettet av 206 observasjoner. Gjennomsnittsavstanden var 4,2 meter fra observasjonene til nærmeste forsenkning. Det var 80 observasjoner som overlappet med forsenkninger og dermed hadde avstand null meter (grønn søyle). Dette er 39 prosent av observasjonene (totalt 206), en økning på 24 prosent fra den grove sorteringen. Etter den detaljerte sorteringen var det ikke lenger noen observasjoner som hadde en avstand på over 25 meter til nærmeste forsenkning.



Figur 45: Histogrammet viser avstanden fra observasjoner til nærmeste forsenkning etter den detaljerte sorteringen. Grønn søyle viser observasjoner som overlapper med forsenkninger

5.9 Feilplasserte observasjoner

I den detaljerte sorteringen ble observasjonenes posisjon validert. Her ble observasjoner som var feilplassert i forhold til posisjonen på bilde erstattet med nye punkt. Figur 46 viser avstanden fra feilplasserte observasjoner til vannoppsamlingens reelle posisjon. Den gjennomsnittlige avstanden fra feilplasserte observasjoner til vannoppsamlingens reelle posisjon var 49 meter. Den minste avstanden var 0,7 meter. Fire observasjoner hadde en feilplassering på over 500 meter, der den største avstanden var på 2660 meter fra vannoppsamlingens reelle posisjon.



Figur 46: Histogrammet viser avstanden fra observasjoner til vannoppsamlingens reelle posisjon

6 Diskusjon

I dette kapitlet diskuteres resultatene fra kapittel 5. Her blir det sett nærmere på kjennetegn som beskriver det folkeinnsamlede datasettet brukt i masteroppgaven. Relevant teori blir brukt for å se om masteroppgavens resultat samsvarer med det andre har funnet ut, eller om masteroppgaven har avdekket andre kjennetegn ved folkeinnsamlet data. Kjennetegnene ved det folkeinnsamlede datasettet vil så brukes til å diskutere om folkeinnsamlet data er egnet til masteroppgavens formål, som er til å validere modellerte forsøkninger for å finne ut om de er reelle vannoppsamlinger. Diskusjonen vil rette et kritisk blikk på metodene brukt i masteroppgaven, altså GIS-basert modellering av forsøkninger og sortering av folkeinnsamlede observasjoner. Det vil bli diskutert om masteroppgavens metoder har vært godt egnet til formålet, hva som kunne vært gjort annerledes og hva som fungerte bra.

6.1 Bekreftede vannoppsamlinger

Det er totalt 96 forsøkninger som ved bruk av folkeinnsamlet data er bekreftet å være reelle vannoppsamlinger. Dette taler for at folkeinnsamlet data har fungert til formålet, til å bekrefte at modellerte forsøkninger er reelle vannoppsamlinger. Som den videre diskusjonen vil vise, betyr ikke dette nødvendigvis at folkeinnsamlet data er godt egnet til formålet.

Av de 96 bekreftede forsøkningene er enkelte forsøkninger bekreftet av flere deltakere. I denne masteroppgaven blir det faktum at flere deltakere har observert den samme vannoppsamlingen ansett å styrke observasjonen og med det øke sannsynligheten for at den bekreftede forsøkingen faktisk er en reell vannoppsamling. Figur 35 viser at det er flest bekreftede forsøkninger som kun har én observasjon, mens det på den andre ytterkanten finnes forsøkninger med både 5, 6 og 7 observasjoner. En antar derfor at det er større sikkerhet for at disse forsøkningene er reelle vannoppsamlinger. Det er altså større usikkerhet knyttet til de forsøkningene hvor kun én deltaker har observert vannoppsamlingen.

De bekreftede forsøkningsene med flest observasjoner befinner seg hovedsakelig i områdene Dragvoll, Moholt, Midtbyen og Møllenberg. Dette er også områder hvor en kan anta at flest studenter ferdes, både i forbindelse med skole, men også fordi mange studenter bor i disse områdene. Studentene som deltok i Citizen Sensing-prosjektet var antagelig knyttet til enten campus Gløshaugen eller Dragvoll. De forsøkningsene som er observert flest ganger ligger i disse områdene. Dette betyr at forsøkningsene som ikke er lokalisert der deltakerne ferdes mye, har mindre sjanse for å bli observert mer enn én gang. Det er altså ikke tilfeldig at noen observasjoner er observert flere ganger, men et resultat av at det folkeinnsamlende datasettet mangler representativitet - det er en overrepresentasjon av observasjoner i enkelte områder. En ser at bekreftede forsøkningsene i ytterkanten av studieområdet gjerne har færre observasjoner.

Hvis en skal stille krav om at valideringen skal være helt pålitelig, vil det være nødvendig med feltarbeid. Dette er den eneste måten en kan være helt sikker på at de bekreftede forsøkningsene faktisk er reelle vannoppsamlinger. En kan likevel gå ut ifra at forsøkningsene med flest observasjoner har mindre usikkerheter, og om det skal foretas undersøkelser i felt vil det derfor kunne være naturlig å prioritere de forsøkningsene som færre har observert og som det derfor knyttes større usikkerhet til. Det er likevel viktig å huske på at observasjonene i datasettet i utgangspunktet var ujevnt fordelt, noe som påvirker hvor mange observasjoner det er av hver vannoppsamling i de ulike områdene. Dette handler om mangel på representativitet, som tidligere er nevnt som et typisk kjennetegn ved Citizen Science (Wiggins & Crowston, 2011). Hvis mer av det opprinnelige datasettet hadde vært brukbart å inkludere som en del av valideringsgrunnlaget, er det mulig at det hadde ville vært flere og andre forsøkningsene som var observert flere ganger.

6.1.2 Terskelverdi på én meter

Kun de forsøkningsene som lå én meter eller nærmere fra observasjonen etter den detaljerte sorteringen ble bekreftet som reelle vannoppsamlinger. Som refleksjonen rundt metodene i avsnitt 6.10 vil vise, er det usikkerheter knyttet til både sorteringen, fastsettingen av nye punkt og forsøkningsanalysen. Dette gjorde det nødvendig å fastsette en terskelverdi for hvilke forsøkningsene som kunne bekreftes og ikke.

Det ble ikke funnet teori eller tidligere studier som kunne gi konkret støtte til valg av riktig terskelverdi. I studien til Aano et al. (2019) ble det funnet at skaderegistreringer av overvann lå minimum én meter fra modellerte forsøkningsene. Det var altså ingen registreringer som

overlappet direkte med forsenkninger. Selv om studien til Aano et al. (2019) ikke ser på nøyaktig det samme som denne masteroppgaven, kan resultatene tyde på at én meter fra forsenkningene er en god terskelverdi. Det ble her altså akseptert en viss avstand til forsenkningen. I studien til Yu et al. (2016) derimot, ble direkte overlapp stilt som krav for å bekrefte forsenkningene.

Valg av én meter som terskelverdi kan enten ha vært for strengt eller vært for mye, men det er vanskelig å vite helt sikkert. Det er mulig at nye punkt satt på bakgrunn av deltakernes bilder avviker mer enn én meter fra den egentlige posisjonen. Valget ble gjort ut fra en skjønnsvurdering av hva som var logisk plassering av vannoppsamlingen, og med formål om å minimere feil i datasettet. Det var 110 forsenkninger som ble utelukket ved en terskelverdi på én meter. Hvis terskelverdien hadde vært justert til to meter, ville det vært 136 forsenkninger som ble bekreftet. To meters avstand ville imidlertid økt sjansene for at forsenkningen ikke var den samme vannoppsamlingen som vist på bilde. Det er likevel mulig at to meter hadde vært en bedre terskelverdi fordi det da ville ha vært ytterligere 40 forsenkninger som ble bekreftet. Dette er såpass mange at det kunne hatt betydning for masteroppgavens resultat. Det ville gitt en indikasjon på at folkeinnsamlet data er mer egnet til formålet enn slik resultatet nå tyder på. Kun 96 av de 206 modellerte forsenkningene ble bekreftet, noe som er under halvparten av forsenkningene. Med en terskelverdi på to meter ville denne andelen økt til 66 prosent.

6.2 Folkeinnsamlet data

Som nevnt i teorien er ikke Citizen Science godtatt i alle forskningssamfunn (Bonney et al., 2014). Dette skyldes skepsisen rundt den manglende ekspertisen, kunnskapen og erfaringen hos deltakerne (Alabri & Hunter, 2010). Denne formen for datainnsamling kan generere data som gir nye innblikk i studiefeltet (Wiggins & Crowston, 2011), og som i enkelte tilfeller kan være like god (Bonney et al., 2014), eller bedre enn data som produseres av eksperter (See et al., 2016). Citizen science blir likevel kritisert for å produsere data av kvalitet som ikke er god nok til bruk i forskning (Alabri & Hunter, 2010; Bonney et al., 2014). Haklay (2013) argumenterer for at datakvaliteten vil variere etter hvor mange deltakere som inkluderes i prosjektet, samt hvilken erfaring og kunnskap hver enkelt deltaker har. Dette betyr at en bør vurdere datakvaliteten til hver observasjon for seg selv, og ikke bedømme datasettets kvalitet som en helhet.

De følgende avsnittene i diskusjonskapittelet vil se på hvordan det folkeinnsamlede datasettet om overvann kan beskrives. Generelt sett kan en si at datasettet preges av heterogenitet - det finnes store forskjeller mellom observasjonene. Dette samsvarer med det Haklay (2013) sier om folkeinnsamlet data generert i citizen science-prosjekt. Disse forskjellene handler om hvor mange observasjoner som er registrert per deltaker, den romlige representativiteten til observasjonene, størrelsen på vannoppsamlingen, kvaliteten på bildene, mulige brukerfeil og posisjonsnøyaktigheten til observasjonene. Oppsummert viser den kommende diskusjonen at det folkeinnsamlede datasettet inneholder flere typer feilkilder, noe som understreker viktigheten av sortering før bruk av denne typen data. Disse usikkerhetene gjør også at en må stille spørsmålsteget ved om datasettet er egnet som valideringsgrunnlag.

6.3 Fordeling av observasjoner

Det folkeinnsamlede datasettet i denne oppgaven preges av en ujevn fordeling. Med dette menes at det er stor variasjon i hvor mange observasjoner hver deltaker har registrert, og i tillegg har observasjonene en ujevn romlig fordeling i studieområdet. Denne ujevnheten kan ha påvirket datasettets kvalitet i positiv, men også negativ retning. Dette igjen påvirker hvor egnet datasettet er som valideringsgrunnlag, noe de følgende avsnittene vil komme inn på.

6.3.1 Ulikt antall observasjoner per deltaker

Det var stor variasjon i antall registrerte observasjoner per deltaker. De fleste deltakerne registrerte kun én observasjon hver, mens noen deltakere skilte seg ut med veldig mange observasjoner - en deltaker hadde så mange som 160 observasjoner (se figur 41). Som nevnt i teorikapittelet er et typisk kjennetegn ved folkeinnsamlet data at enkelte deltakere bidrar mer enn andre (Haklay, 2013), noe som også kjennetegner datasettet brukt i denne oppgaven.

Det kan være flere mulige årsaker til at deltakerne registrerer ulikt antall observasjoner. Tonurist (2017) skiller mellom indre og ytre motivasjon i prosjekt som involverer frivillige innbyggere. Deltakelsen i Citizen Sensing-prosjektet som denne masteroppgaven baserer seg på var frivillig arbeid, men de som deltok fikk muligheten til å være med i trekningen av en Netatmo værstasjon. Den deltakeren som registrerte flest observasjoner ble vinneren, og det er grunn til å anta at dette var en ytre motiverende faktor, særlig for de med mange registreringer.

I Neset (2021) sin artikkel om Citizen Sensing-prosjektet poengteres det at *spillifisering* (gameification) var en årsak til forlenget interesse og engasjement blant flere deltakere i prosjektet. Dette betyr at motivasjonen hos enkelte deltakere skyldes at datainnsamlingen har preg av konkurranse. I dette tilfellet handler det om å registrere flest observasjoner for å vinne premien. For andre deltakere igjen virket det som konkurransen om å registrere flest observasjoner hadde motsatt effekt (Neset, 2021). Muligens hadde disse deltakerne en sterkere indre motivasjon, noe som kunne bety at konkurranseaspektet ble en demotiverende faktor.

Når motivasjonen går fra å komme innenfra til kun å være basert på ytre faktorer, poengterer Neset (2021) at dette kan ha ført til overregistrering av observasjoner. Hvis dette er tilfellet, kan det ha ført til et datasett av lavere kvalitet. Dette gjelder særlig hvis deltakeren som har registrert 160 observasjoner har slurvet med registreringen og det finnes store feil i observasjonene som skyldes unøyaktig fastsetting av koordinat, dårlig bildekvalitet eller feil tolkning av oppgaven. På den andre siden kan mange observasjoner fra samme deltaker påvirke datasettet positivt hvis registreringen er gjort på en god måte og observasjonene bærer preg av høy nøyaktighet. En enkelt deltaker som registrerte 160 observasjoner vil påvirke det helhetlige resultatet i større grad enn deltakere som kun har registrert én observasjon. Etter den detaljerte sorteringen var det ingen av observasjonene fra deltakeren som registrerte totalt 160 som var igjen. Dette betyr at disse observasjonene var preget av lav kvalitet, mer spesifikt handlet det om at observasjonene ikke viste omgivelser og derfor var usikre. I dette tilfellet kan en derfor anta at Neset et al. (2021) sin teori om overregistrering stemmer.

For de deltakerne som kun registrerte én enkelt observasjon, kan en anta at premien ikke var motivasjon nok. For de deltakerne som registrerte det mest vanlige antallet observasjoner (1-8), kan det tenkes å ha vært andre faktorer som påvirket motivasjonen. Dette kan for eksempel være en interesse for faget og et ønske om å lære mer om temaet (Bonney et al., 2014). For studenter som er interessert i hvilke klimahendelser som finnes i nærområdet, var det muligens lettere å finne motivasjon til å rapportere inn hendelser, og også gjøre dette på en ordentlig måte.

Anonymisering er typisk for Citizen Science-prosjekt, og i enkelte tilfeller kan dette føre til en mangel på forpliktelse blant deltakerne (Alabri & Hunter, 2010). Uten muligheten til selvpromotering eller følelsen av å få noe igjen, er det ikke sikkert motivasjonen er høy nok for enkelte (Goodchild, 2007). Deltakeren får ikke betaling slik eksperter trolig ville fått for å gjøre den samme oppgaven. Dersom deltakerne ikke ser hvilken nytte prosjektet kan ha for dem selv,

kan det føre til at deltakere føler mindre ansvar for at data som registreres er av god kvalitet og oppfyller de kravene som er satt.

6.3.2 Ujevn fordeling av observasjoner i studieområdet

Fordelingen av observasjoner i studieområdet kan beskrives som ujevn. Figur 42 viser fordelingen av observasjoner i studieområdet, hvor en ser at områdene Midtbyen, Bakklandet, Elgeseter, Moholt og Dragvoll har større tetthet av observasjoner enn andre steder i studieområdet. Deltakerne i Citizen Sensing-prosjektet var studenter, og de nevnte områdene er bydeler i Trondheim som studenter gjerne oppholder seg i. Her finnes både universitetscampus, studentboliger og populære kulturtilbud. Det er grunn til å anta at dette kan ha påvirket tettheten av observasjoner i disse områdene. Det er mulig mangel på motivasjon og forpliktelse også har påvirket dette. Det er ikke sikkert deltakerne finner motivasjon til å observere i andre områder enn de selv ferdes i løpet av dagen. Mangelen på romlig representativitet viser et av problemene som kan oppstå når en bruker folkeinnsamlet data i forskning. Mangel på representativitet av data i studieområdet er i litteraturen trukket fram som et hovedproblem med datainnsamling i Citizen Science (Wiggins & Crowston, 2011). Det samme problemet kan en se på artsdatabanken.no, hvor befolkningen typisk registrerer mange observasjoner av ulike arter i bebygde områder eller typiske rekreasjonsområder. I avsidesliggende områder, som for eksempel fjellpartier, er antall observasjoner færre fordi det ikke ferdes like mange her.

Overrepresentasjon av observasjoner i enkelte områder og underrepresentasjon i andre kan ha påvirket hvilke forsøkninger som ble bekreftet i valideringen. Bekreftelsen betinger at det er registrert en observasjon som ligger maksimum én meter fra forsøkningspunktet. Det vil trolig være færre bekreftede forsøkninger i de områdene hvor det er få observasjoner. Dette er en tydelig svakhet ved datasettet og kan gjøre at det er mindre egnet som valideringsgrunnlag for et helt studieområde. Det er likevel usikkert om mangelen på representativitet har påvirket besvarelsen av selve problemstillingen. Det har til tross for manglende representativitet vært mulig å bekrefte forsøkninger ved bruk av det folkeinnsamlede datasettet. Hvis en for eksempel ønsker å bruke datasettet til å prioritere områder der overvann må forebygges eller håndteres, kan mangelen på representativitet gjøre det vanskelig å få full oversikt over hele studieområdet. Dette er en utfordring som trolig kunne blitt minimert med en tydelig plan for datainnsamling (Wiggins & Crowston, 2011), der det på forhånd ble definert hvilke områder deltakerne skulle undersøke.

6.4 Ulik størrelse på vannoppsamlingene

Det var stor variasjon i størrelsen på vannoppsamlingene som deltakerne registrerte, et eksempel på dette ser en i figur 22. Dette kan tyde på at deltakerne ikke hadde fått noen klar beskjed over hvilke fenomen som skulle inkluderes og hvilke som skulle utelates. Stor variasjon i størrelsen på de registrerte vannoppsamlingene kan ha påvirket oppgavens resultat. Det er mulig at observasjoner som viser veldig små vannoppsamlinger ikke vil modelleres i GIS som en forsenkning, rett og slett fordi de er for små og dermed ikke anses som en forsenkning i GIS. Det er likevel ikke sikkert at dette er tilfellet for denne masteroppgaven, ettersom oppløsningen på terrengmodellen var 25 cm. Som tidligere nevnt definerer Heywood (2011) oppløsning som det minste fenomenet som kan representeres i datasettet. For denne masteroppgaven betyr det at ingen forsenkninger mindre enn 25 cm vil modelleres, kun forsenkninger større enn dette identifiseres i analysen. Det er viktig å være klar over denne begrensningen når en inkluderer observasjoner av vannoppsamlinger i alle størrelser.

Deltakerne skulle registrere sine observasjoner under kategorien *kraftig regn*, men det var ikke spesifisert nærmere hvilke fenomen som inngår i denne betegnelsen. Det bidro til at det ble registrert vannoppsamlinger i alle størrelser, men som nevnt også fenomen som ikke var vannrelatert. Forskere eller eksperter som jobber med et fagfelt kan bruke sine erfaringer til å avgjøre hvilke observasjoner som er nyttige eller ei. Innbyggere som deltar i datainnsamlingen har ikke samme erfaringsgrunnlag for å kunne skille nyttige og unyttige observasjoner. Enkelte deltakere registrerer muligens alle former for overvann de ser, mens andre kun de vannoppsamlingene som er store nok til å ha en tydelig påvirkning på omgivelsene. For å gjøre en vurdering av hvilke vannoppsamlinger som kan gjøre skade og ikke, trengs noe bakgrunnskunnskap om temaet, og det er det ikke sikkert deltakerne hadde. Hvilke vannoppsamlinger som ble registrert og hvilke som ble utelatt, varierte altså fra person til person. Dette gjør at kvaliteten på datasettet vil variere for hver enkelt observasjon, fordi individets kunnskap og erfaringer vil spille inn (Haklay, 2013). Det må også nevnes at Citizen Sensing-prosjektet i utgangspunktet ikke hadde samme formål som denne masteroppgaven, og deltakerne hadde derfor trolig ikke fått beskjed om å for eksempel utelate de minste vannoppsamlingene. Formålet med Citizen Sensing-prosjektet var å samle inn data om ulike klimarelaterte hendelser, og her bestemte deltakerne selv hvilke hendelser som skulle registreres. Dette viser at folkeinnsamlet data muligens er best egnet hvis datainnsamlingen er tilpasset oppgavens egentlige formål. På denne måten sørger en for at vannoppsamlingene som

observeres har større likhet, noe som skaper mer homogenitet i kvaliteten på datasettet og gjør det lettere å trekke generelle konklusjoner.

Som nevnt i metodekapittelet ble det ved sortering valgt å inkludere vannoppsamlinger av alle størrelser. Dette skyldtes at det var vanskelig å sette en grense for hvilke vannoppsamlinger som var for små og ikke. Ifølge Bayr (2021) kan det være vanskelig å tolke størrelsen til fenomen ut ifra bildet fordi vinkel, mangelen på geografisk referanse, ulik bildekvalitet og objekter som blokkerer spiller en rolle. Dette kan også tenkes å gjelde vurdering av størrelsen på vannoppsamlingene. Dette er bakgrunnen for valget om å inkludere vannoppsamlinger av alle størrelser i valideringen av forsøkninger. I studien til Bayr (2021) ble *monoplotting*, en metode for å omgjøre todimensjonale bilder slik at de passer inn i et geografisk koordinatsystem, dette er brukt for å bestemme størrelsen på arealklasser. Det er mulig en slik metode kunne vært brukt til å bestemme størrelsen på vannoppsamlingene. Da kunne en hatt mer informasjon om vannoppsamlingenes størrelse, noe som kunne fungert som en ekstra sikkerhet i valideringen. Selv om observasjonen overlappet med den modellerte forsøkningskningen, så kunne det oppstå usikkerhet i valideringen. Når vannoppsamlingen på bildet var mye mindre enn den modellerte vannoppsamlingen, skapte dette usikkerhet om bildet faktisk viste den modellerte forsøkningskningen, eller om det viste en annen vannoppsamling. En slik tidskrevende metode er likevel langt utenfor denne masteroppgavens omfang, men kan være en interessant problemstilling for videre forskning.

En annen mulighet for utfyllende informasjon over vannoppsamlingenes størrelse er at deltakerne selv kunne lagt inn informasjon på hvor bred vannoppsamlingen var. Her vil derimot subjektiviteten til hver enkelt deltaker være en utfordring. Ettersom deltakerne ikke er forskere som har mye kunnskap på feltet, er det mulig motivasjonen kan svekkes dersom forskningsopplegget er for komplisert. Det er mulig deltakere ikke lenger ser på oppgaven som et frivillighetsprosjekt, men heller et arbeid de burde fått betaling for å utføre. En strukturert plan for datainnsamling har allerede blitt trukket fram som en viktig del av å sikre god data i citizen science (Wiggins & Crowston, 2011). Det å følge en strukturert plan kan føre til at alle vannoppsamlinger som registreres holder en viss størrelse, noe som kan sikre at observerte vannoppsamlinger kan modelleres som en forsøkningskning i GIS. For å sikre data som er egnet i validering av forsøkninger vil det være viktig å informere deltakere om hvilke observasjoner som har høyest verdi. Opplæring av deltakere i forkant av prosjektet (Alabri & Hunter, 2010), kan sikre et mer homogent datasett uten store variasjoner i størrelsen på vannoppsamlingene.

For å sikre en objektiv datainnsamling som resulterer i mer homogent resultat, er det ifølge Hunter (2010) viktig med tydelige retningslinjer i forkant av datainnsamlingen. Dette bidrar til at deltakere blir klar over hvordan datainnsamlingen bør gjennomføres. Det er likevel viktig å huske på at folkeinnsamlet data som krever refleksjon og vurdering fra den enkelte deltaker alltid vil være preget av en viss grad av subjektivitet.

6.5 Bildene har varierende kvalitet

Bildene koblet til observasjonene var en viktig kilde til informasjon, både for å vurdere om observasjonene viste overvann og om lokasjonen til observasjonene var korrekt. Det var totalt 23 observasjoner som ikke hadde bilde, og disse ble fjernet for å minimere usikkerhet rundt observasjonene. Den manuelle gjennomgangen viste at det var varierende kvalitet på bildene fra deltakerne, noe som gjorde at enkelte bilder var bedre å bruke i fastsettingen av vannoppsamlingens reelle posisjon. Bildets vinkel hadde betydning - enkelte bilder viste omgivelsene rundt vannoppsamlingen, mens andre bilder viste kun selve vannoppsamlingen. Hvis en skal inkludere omgivelsene rundt vannoppsamlingen kan dette kreve at bildet tas et stykke unna selve vannoppsamlingen. Eksempler på bilder som ikke viser omgivelser ble presentert i metodekapittelet i figur 24. Bilder som viste omgivelser som gjorde det enkelt å identifisere lokasjonen var svært nyttige, og ble derfor vurdert som godt egnet til bruk i validering av forsøkninger. Det må nevnes at kvaliteten på bildene også var varierende, enkelte bilder var for mørke til at omgivelsene kunne identifiseres. Dette kan skyldes varierende kvalitet på kameraet, men også ulik kunnskap om fotografering og justering av lysstyrke hos deltakerne.

6.6 Brukerfeil

Mangelen på ekspertise blant deltakere har allerede blitt trukket fram som en av de viktigste årsakene til at citizen science data kan generere data som er mangelfull og av lav kvalitet (Alabri & Hunter, 2010). En lite strukturert plan for datainnsamling kan også føre til at resultatene preges av brukerfeil som muligens ikke ville vært til stede dersom innsamlingen var gjort av forskere (Wiggins & Crowston, 2011). Når data samles inn av et stort antall mennesker uten forskningsbakgrunn på feltet, er det fare for feil i datainnsamlingen grunnet manglende kunnskap og erfaring (Alabri & Hunter, 2010), og kanskje forståelse for hvordan data burde samles inn. I denne masteroppgaven har det blitt brukt folkeinnsamlet data samlet inn i felt

gjennom GPS fra mobil, men enkelte observasjoner har også blitt lokalisert manuelt på kart av deltakerne i ettertid. Spesielt sistnevnte åpner for mulige brukerfeil hvis deltakerne ikke har fått god nok opplæring i forkant. I de følgende avsnittene vil ulike brukerfeil i det folkeinnsamlede datasettet diskuteres, og det blir drøftet om disse brukerfeilene har ført til så store feilkilder at datasettet blir lite egnet til sitt formål.

6.6.1 Observasjoner som ikke viser overvann

Som tabell 2 viser var det 176 observasjoner som måtte fjernes fordi bildene ikke var vannrelaterte. Som vist i metodekapittelet kunne dette være bilder av løv og hull i asfalten. Disse observasjonene hadde vært mulige store feilkilder i analysen hvis de ikke hadde blitt sortert ut. Det er mulig deltakerne har vært usikre på hvilken kategori observasjonen deres passet inn i og derfor registrert de under kategorien *kraftig regn*. Kanskje var det ingen andre kategorier som passet bedre. Dette er en indikasjon på at kategoriene i Citizen Sensing-applikasjonen ga rom for stor frihet blant deltakerne, noe som førte til mindre struktur i datasettet. Uten sortering hadde en ikke blitt klar over disse brukerfeilene, og det kunne ført til at valideringen ble gjort på grunnlag av observasjoner av andre fenomen enn overvann. Dette viser igjen at sortering har vært en svært viktig del for å minimere feilkilder i analysen, og kan tyde på at det opprinnelige datasettet ikke var særlig egnet til validering av forskninger. Observasjoner registrert i feil kategori kunne ført til at valideringen ble gjort på grunnlag av observasjoner som ikke handlet om overvann.

6.6.2 Registrering i bygg

Totalt 522 observasjoner var registrert inne i bygg. Dette skyldes antageligvis at deltakerne ikke har fullført registreringen i applikasjonen før de enten har kommet hjem eller fram til destinasjonen sin. Dette er et eksempel på en typisk brukerfeil i dette prosjektet, og det er grunn til å anta at samme feil kan oppstå i lignende prosjekt. Det er mulig disse observasjonene inneholdt brukbar informasjon, men siden lokasjonen var feil ble de fjernet. Resultatet av analysen kunne blitt noe annerledes hvis disse punktene var mer nøyaktige og inkludert i analysen, men dette er det vanskelig å si noe sikkert om. Studien til Wiggins og Crowston (2011) sier at flere observasjoner kan være med på å styrke representativiteten til datasettet, noe som også kan stemme for denne masteroppgaven. Andelen observasjoner som går tapt ved brukerfeil som dette viser hvor store feilkilder en kan finne i folkeinnsamlet data, og understreker igjen viktigheten av sortering før datasettet tas i bruk.

6.6.3 Observasjoner med samme koordinat

Et annet eksempel på mulige brukerfeil i datasettet kan være at totalt 33 observasjoner var registrert i samme koordinat. De fleste av de overlappende observasjonene var registrert av ulike deltakere, mens noen få var av samme deltaker. Ved den manuelle gjennomgangen ble det som nevnt i metodekapittelet brukt rutenett for å sørge for en strukturert gjennomgang av observasjonene. Ettersom de overlappende observasjonene ikke ville blitt lagt merke til ved gjennomgang i attributtavhengig tabell, understreker dette viktigheten av å bruke rutenett som metode for å sørge for struktur. Mulige årsaker til overlappende observasjoner kan være brukerfeil ved registrering eller en mulig svikt i Citizen Sensing-applikasjonen. Uansett grunn så understreker funnet av overlappende observasjoner at sortering av denne typen data er viktig både for å minimere feilkilder, men også for å bli klar over hvilke feilkilder som finnes og som det må tas hensyn til.

6.7 Posisjonsnøyaktighet

Usikkerhet i datasettet en bruker som inngangsdata i GIS vil føre til usikkerhet i resultatdatasettet (Longley et al., 2005, s. 145). For å kunne validere om forsøkninger inneholder overvann, har det vært viktig at observasjonen er lokalisert der bildet viser at vannoppsamlingen befinner seg. Som tabell 2 viser var det totalt 263 observasjoner som var lokalisert feil sammenlignet med hvor vannoppsamlingen befant seg i virkeligheten. Gjennomsnittsavstanden var på 49 meter, men her trakk enkelte observasjoner opp gjennomsnittet en god del – en observasjon var lokalisert så mye som 2660 meter fra vannoppsamlingens reelle posisjon. Posisjonsfeilene understreker viktigheten av å sortere folkeinnsamlet data. Observasjonsdata med posisjonsfeil medfører såpass stor usikkerhet at de blir lite egnet som valideringsgrunnlag. Som de kommende avsnittene vil vise, ble det både identifisert observasjoner med små og store posisjonsfeil. Når en skal bekrefte at en forsøknings er en reell vannansamling er det viktig at posisjonen stemmer. Toblers *første regel for geografi* (first law of geography) sier at alt har en relasjon til alt, men fenomen som ligger nært romlig er mer relatert enn fenomen lengre unna (Longley et al., 2005, s. 65). Dette gjelder også for valideringen i denne masteroppgaven. Det vil ikke være mulig å bekrefte forsøkningsene hvis observasjonene og forsøkningsene ikke har en romlig nærhet.

Som presentert i teorikapittelet (kapittel 3) brukes begrepene *nøyaktighet* og *presisjon* for å beskrive kvaliteten på stedfestet data i GIS. I Longley et al. (2005, s. 140) er begrepene brukt

om nøyaktigheten og presisjonen som oppnås ved instrumentmåling. I denne masteroppgaven vil begrepene brukes for å diskutere den romlige kvaliteten til det folkeinnsamlede datasettet om overvann, og dette påvirkes stort sett av de ulike deltakerne som registrert observasjonene.

6.7.1 Nøyaktighet

I denne masteroppgaven brukes begrepet nøyaktighet om de observasjonene som har samme posisjon som vannoppsamlingen i virkeligheten (Longley et al., 2005, s. 140). Etter den detaljerte sorteringen var det totalt 56 observasjoner som ble beholdt slik deltakerne opprinnelig registrerte dem. Dette fordi de ble bekreftet å ha samme posisjon som vannoppsamlingens reelle posisjon. Figur 47 viser eksempel på en slik observasjon. Her ligger observasjonen (grønt punkt) nøyaktig der vannoppsamlingen befinner seg i virkeligheten sett ut ifra bildet fra deltakeren. Denne observasjonen ble derfor beholdt som den var og ingen lokasjonsendring var nødvendig.



Figur 47: Eksempel på en observasjon som regnes som nøyaktig

Det var totalt 263 observasjoner som ble fjernet og erstattet med et nytt punkt lokalisert der vannoppsamlingen befant seg virkeligheten. Disse observasjonene vil derfor kunne beskrives som lite nøyaktige. Ut ifra dette funnet kan en anta at det sannsynligvis er flere observasjoner i det opprinnelige datasettet som er unøyaktige. Dette støtter argumentet om at en detaljert sortering av datasettet er nødvendig for å minimere feilkilder. Uten denne sorteringen ville nok datasettet vært uegnet til bruk i valideringen fordi feilkildene er mange, og en kunne ikke stolt på resultatets troverdighet.

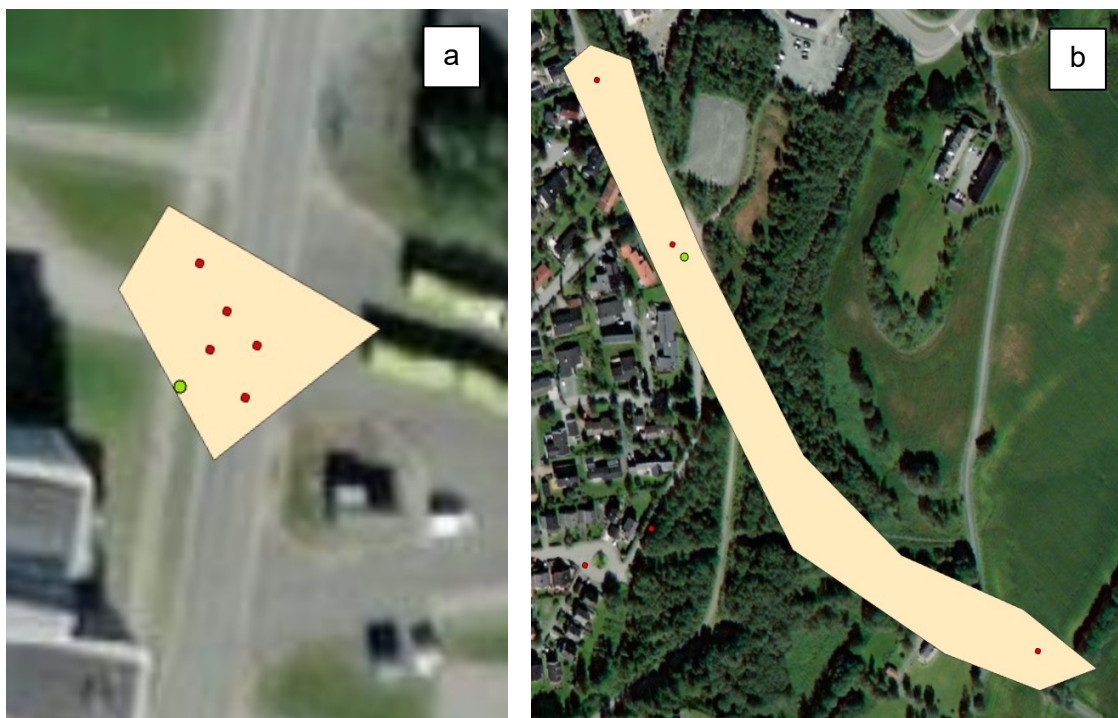
Blant de feilplasserte observasjonene må det nevnes at det var stor forskjell på nøyaktigheten - noen var lite nøyaktige, mens andre var svært lite nøyaktige. Histogrammet i figur 46 viser avstanden fra observasjoner til vannoppsamlingens reelle posisjon, og her ser en at åtte observasjoner har en nøyaktighet på under én meter, men at det også finnes flere observasjoner med en nøyaktighet på over 500 meter – på det meste var en observasjon feilplassert med 2660 meter. Her kommer en tilbake til det Haklay (2013) sier om heterogenitet i folkeinnsamlet data – det vil være store forskjeller på kvaliteten innad i datasettet.

6.7.1.1 Økt nøyaktighet etter detaljert sortering

Histogrammene i figur 44 og 45 viser avstanden fra observasjoner til nærmeste forsenkning, både etter grov sortering og detaljert sortering. Her ser en at den detaljerte sorteringen har ført til en økning i antall observasjoner som overlapper med modellerte forsenkninger. Dette tyder på at sorteringen har ført til økt nøyaktighet, noe som indikerer at den detaljerte sorteringen har vært viktig for datasettets kvalitet. Videre viser histogrammet i figur 45 at det etter den detaljerte sorteringen ikke lenger finnes observasjoner som er over 25 meter unna forsenkningen. Dette viser det samme som nevnt over, den detaljerte sorteringen forbedrer kvaliteten på datasettet når det kommer til nøyaktighet. Disse funnene indikerer at den detaljerte sorteringen har vært svært viktig for å sikre troverdig data, og som følge av dette kan en anta at datasettet i sin opprinnelige stand ikke var særlig egnet til formålet. Dette fordi en romlig analyse i stor grad baserer seg på nærhet mellom objekt, og avstanden mellom observasjoner og forsenkninger minker tydelig fra den grove sorteringen til den detaljerte sorteringen.

6.7.2 Presisjon

Begrepet *presisjon* kan brukes til å beskrive hvor samlet ulike observasjoner av den samme vannoppsamlingen er (Longley et al., 2005, s. 140) I denne masteroppgaven kan begrepet være noe vanskelig å anvende. Data er registrert av mange ulike deltakere med hver sine subjektive vurderinger av observasjonen. Dette påvirker blant annet hvor nær vannoppsamlingen de stod da bildet ble tatt. Begrepet presisjon blir likevel aktuelt når flere deltakere har observert samme vannoppsamling. I alle tilfeller hvor flere deltakere har observert samme vannoppsamlingen ser en at observasjonene ligger med noe avstand fra hverandre, det er altså ingen observasjoner av samme vannoppsamling som ligger i nøyaktig samme koordinat. Dette kan skyldes variasjon i hvordan observasjonene registreres blant deltakerne, hvor nærme de står til vannoppsamlingen og om de har registrert observasjonen på riktig måte (og ikke etter at de har flyttet seg vekk fra vannoppsamlingen). Figurene under (48a og 48b) viser eksempler på observasjoner av samme vannoppsamling fra ulike deltakere.

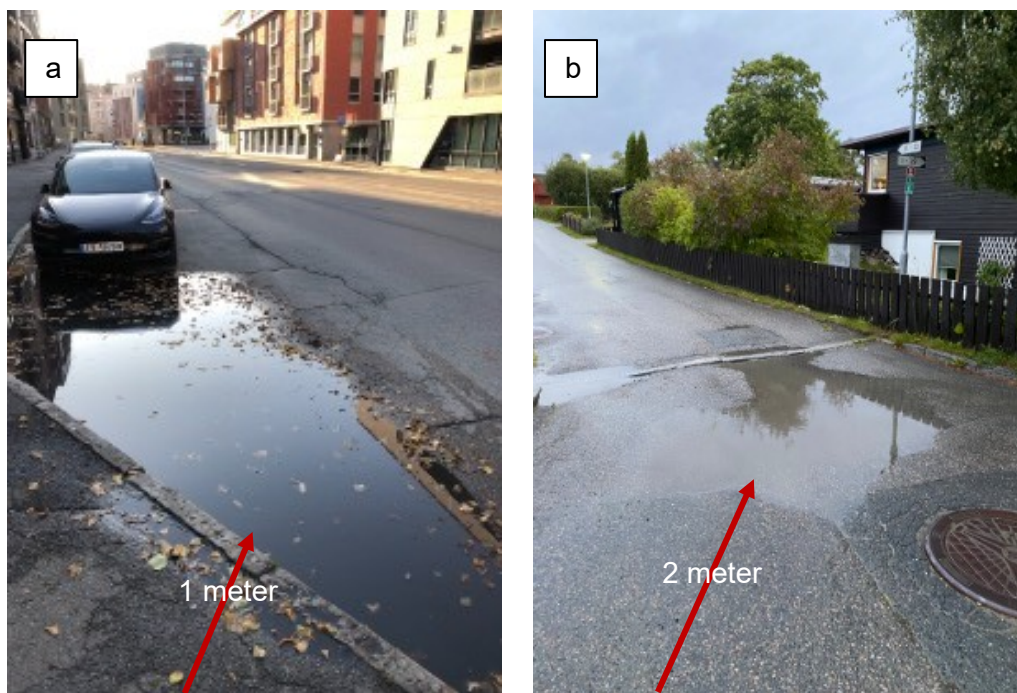


Figur 48: Observasjoner med lav nøyaktighet og lav presisjon hvor a) viser observasjoner som er lite nøyaktige, mens b) viser observasjoner som er svært lite nøyaktige

De røde punktene er deltakernes observasjoner og det grønne punktet viser hvor vannoppsamlingen er lokalisert i virkeligheten. I figur 48a har observasjonene en gjennomsnittlig avstand til hverandre på omtrent 5 meter. Observasjonene beskrives derfor som lite presise. I tillegg har de samme observasjonene en gjennomsnittsavstand på 8,5 meter fra vannoppsamlingen i virkeligheten, og kan i dette tilfellet også betegnes som lite nøyaktige. I figur 48b er avstanden mellom observasjonene og vannoppsamlingen større. Gjennomsnittlig er disse observasjonene 169 meter unna vannoppsamlingens reelle posisjon, og ligger som en ser forholdsvis langt fra hverandre. Det er likevel én observasjon som kan betegnes som mer nøyaktig enn de to andre, men denne ligger likevel 12 meter fra vannoppsamlingens reelle posisjon. I dette tilfellet kan en derfor også beskrive observasjonene som verken nøyaktige eller presise.

6.8 Små lokasjonsfeil

Som vist i histogrammet i figur 46 er det 8 observasjoner som befinner seg én meter eller mindre fra vannoppsamlingen i virkeligheten, og det er 12 observasjoner som ligger mellom 1-2 meter fra vannoppsamlingens reelle posisjon. Den korte avstanden fra observasjon til vannoppsamlingens reelle posisjon kan skyldes at deltakerne av naturlige årsaker ikke ønsker å stå midt i vannoppsamlingen. I tillegg står de gjerne et stykke unna selve vannoppsamlingen for å få med omgivelsene på bildet. En kan anta at deltakerne registrerer posisjonen kort tid etter at bildet er tatt. Dette kan bety at registreringen skjer der bildet ble tatt, og at deltakerne ikke går nærmere vannoppsamlingen for å registrere en mer nøyaktig posisjon. Dette kan som sagt skape mindre og større avstander mellom observasjonspunkt og vannoppsamling. Dette ser en eksempel på i figur 49, hvor a) viser at deltakeren står omtrentlig en meter unna vannoppsamlingen når bilde tas, mens b) viser at deltakeren står omtrent to meter unna. Disse korte avstandene er i seg selv nok til at det ikke registreres overlapp mellom observasjoner og modellerte forsenkninger i romlige utvalg, eller ved bruk av avstandsverktøy som vist i histogrammene i figur 44 og 45 i resultatdelen.



Figur 49: a) deltakeren står omtrent en meter unna vannoppsamlingen og b) deltakeren står omtrent to meter unna vannoppsamlingen

Avstanden fra deltakeren som tar bilde til vannoppsamlingen viser at validering av forsøkninger ved bruk av folkeinnsamlet data kan være problematisk da punktet representerer deltakerens ståsted ved observasjonen og ikke lokasjonen til selve vannoppsamlingen. Uten sortering og erstatting av unøyaktige observasjoner, vil denne avstanden kunne føre til at færre forsøkninger bekreftes eller at de som bekreftes egentlig ikke er reelle vannoppsamlinger.

6.8.1 Flerveisinterferens som mulig problem

Som presentert i teorikapittelet kan det en kaller *flerveisinterferens* være et problem i urbane områder hvor bygninger, biler og andre urbane objekt kan reflektere satellittsignaler (Rød, 2015, s. 75). For de deltakerne som registrerte observasjonene ved bruk av GPS i telefonen, kan det være fare for at koordinatene i urbane deler av studieområdet ble basert på reflekterte signaler og derfor avvek fra den faktiske posisjonen. Ettersom studieområdet for denne masteroppgaven er lagt til den urbane delen av Trondheim, må en derfor ta med i betraktningen at flerveisinterferens kan ha hatt innvirkning på posisjonsnøyaktigheten til observasjonene. Spesielt kan dette tenkes å gjelde Midtbyen, hvor bygningene ligger tett og skaper mange refleksjonsflater. Dette er også det området de fleste observasjonene er registrert.

I følge Kos et al. (2010) kan flerveisinterferens skape koordinatfeil på et par meter. Som vist i histogrammet i figur 46 er det totalt 20 observasjoner som har en posisjonsfeil på mellom 0-2 meter. Om posisjonsfeilen til disse observasjonene skyldes at brukeren står et stykke unna

vannoppsamlingen under registreringen eller skyldes flerveisinterferens er vanskelig å si helt sikkert. Det er likevel viktig å være klar over denne potensielle feilkilden når en foretar datainnsamling med GPS i urbane områder. Flerveisinterferens kan også hatt påvirkning på observasjoner med større posisjonsfeil, men for disse er det nok mer sannsynlig at brukerfeil ved registrering har hatt en innvirkning.

6.9 Store feilplasseringer av observasjoner

Det var også flere observasjoner som hadde større feilplasseringer enn en og to meter fra vannoppsamlingens reelle posisjon. I histogrammet i figur 46 ser en blant annet at så mange som 49 observasjoner hadde en feilplassering på mellom 10-20 meter. Den maksimale avstanden var som sagt 2660 meter fra den virkelige posisjonen til vannoppsamlingen. Det betyr at for denne observasjonen er det usikkert om vannoppsamlingen er i nærheten - den kan like gjerne være lokalisert på den andre siden av byen. En kan anta at store avstander trolig ikke kun kan skyldes feil innført ved flerveisinterferens eller at deltakeren tar bilde et stykke unna vannoppsamlingen.

Disse feilplasseringene skyldes trolig brukerfeil ved registrering. Deltakeren har observert vannoppsamlingen og tatt bilde på stedet, men trolig ikke fullført registreringen før senere når lokasjonen til deltakeren er endret. Det var også mulig for deltakerne å registrere observasjonens lokasjon i ettertid på et kart, hvor en også kunne legge til bilde i etterkant. Denne formen for registrering kan tenkes å ha generert de største feilkildene ettersom posisjonen til disse ikke baserer seg på GPS-signal, men manuell posisjonsbestemmelse av deltakeren selv. Ved GPS-registrering er det en viss sikkerhet i at instrumentet sørger for nøyaktige registreringer uten at deltakeren trenger å foreta seg noe. Når deltakeren bestemmer posisjonen manuelt på et kart i ettertid er det imidlertid flere faktorer som påvirker hvor nøyaktig posisjonen blir. Hver enkelt deltaker vil ha ulik kompetanse både når det gjelder teknologisk forståelse, kunnskap om kart og kjennskap til studieområdet. Noen deltakere er lokalkjente, mens andre er innflyttere uten samme lokalkunnskap. Dersom deltakeren ikke selv husker helt hvor observasjonen er, kan lokasjonen blir satt noe tilfeldig. Det kan være vanskelig for deltakeren å bestemme posisjonen til observasjonen helt nøyaktig kun ved å se på et kart. Deltakerens forpliktelse til prosjektet og motivasjonen for å levere data av god kvalitet vil også påvirke nøyaktigheten. Her kan det tenkes at deltakere som kun motiveres av premien ikke vil

legge like stor innsats i at posisjonen bestemmes nøyaktig. De er kanskje mer opptatt av å få registrert selve observasjonen. Deltakere som har en genuin interesse for prosjektet kan ha større interesse for å levere data av god kvalitet. Store posisjonsfeil forsterker poenget med at sortering av folkeinnsamlet data er svært viktig før datasettet tas i bruk i forskning, spesielt for prosjekt som inkluderer deltakere uten forskningsbakgrunn.

6.10 Refleksjon rundt metoder

I denne masteroppgaven er det brukt to metoder for å komme fram til et resultat som kan besvare problemstillingen – den ene metoden er modellering av forsøkninger i GIS og den andre er sortering av folkeinnsamlet data. Hver av metodene inneholder potensielle feilkilder som kan ha påvirket det endelige resultatet. For sorteringen av observasjoner vil det diskuteres om sorteringsjobben er for omfattende til at det lønner seg å bruke folkeinnsamlet data som valideringsgrunnlag. En potensiell feilkilde med denne metoden er subjektivitet, som er spesielt framtrekkende i den detaljerte sorteringen. En annen feilkilde som kan ha påvirket resultatet er sorteringskravene. Det blir diskutert om det har blitt fjernet observasjoner av god kvalitet fordi kravene for sortering har vært for strenge. I modelleringen av forsøkninger har en terrengmodell med god oppløsning blitt brukt for å representere det ujevne terrenget en finner i det urbane studieområdet. Til tross for dette er modelleringen en grov forenkling, og en kan derfor anta at det vil finnes noen feil.

6.11 Sortering av observasjoner

På bakgrunn av det tidligere studier, som for eksempel Alabri & Hunter (2010), har sagt om feilkilder i folkeinnsamlet data, har sortering som metode blitt vurdert som nødvendig for å minimere mengden feil i datasettet. I diskusjonen om kjennetegn ved datasettet avdekkes det ulike sider ved datasettet som gjør det mindre egnet til formålet - som valideringsgrunnlag for modellerte forsøkninger. Funnene i denne masteroppgaven viser hvor viktig det er med sortering av folkeinnsamlet data, noe som også er understreket i tidligere studier (Delaney et al., 2008). Som nevnt tidligere ble det både i studien til Aano et al. (2019) og Yu et al. (2016) viktig å sortere det folkeinnsamlede datasettet før bruk. Dette viser at det ikke bare er i denne masteroppgaven sortering har blitt sett på som en nødvendighet, men viser at dette er et typisk kjennetegn for prosjekt som benytter data samlet inn av folk uten forskningsbakgrunn.

6.11.1 En omfattende sorteringsjobb

Citizen science har blitt trukket fram som en effektiv måte å samle inn store mengder data over store geografisk områder, data som ikke er like lett å skaffe for forskere (Alabri & Hunter, 2010). Datasettet brukt i denne masteroppgaven inneholdt opprinnelig 2554 observasjoner, men etter den detaljerte sorteringen var gjennomført var det kun 206 observasjoner igjen. Som nevnt var det kun 56 av disse som var fra det opprinnelige datasettet, mens 150 av dem var punkt som i ettertid ble fastsatt ved bruk av bildene fra deltakerne.

Mange observasjoner er positivt hvis disse holder høy kvalitet, men nytten av mange observasjoner vil bli mindre hvis datasettet preges av lav kvalitet og krever en stor sorteringsjobb. I denne masteroppgaven er det skilt mellom grov sortering og detaljert sortering. Den grove sorteringen har som nevnt ikke krevd like nøye vurdering som den detaljerte sorteringen, og denne jobben har gått forholdsvis raskt. Den grove sorteringen har fjernet de mest åpenbare feilkildene som kunne hatt stor innvirkning på resultatet. Det var også andre store feilkilder i datasettet som ikke ville blitt oppdaget uten den detaljerte sorteringen. Eksempler på dette er observasjoner som ikke var vannrelaterte og observasjoner som var feil posisjonert.

Den detaljerte sorteringen har imidlertid vist seg å være svært tidkrevende fordi det har vært nødvendig med en manuell gjennomgang av hver enkelt observasjon. Hver observasjon har blitt vurdert som enten gyldig eller ugyldig for videre bruk. Dette viser en svakhet ved folkeinnsamlet data som nødvendigvis ikke vil være til stede dersom forskere med god kunnskap og erfaring på feltet foretar datainnsamlingen. Hvis datasettet var samlet inn av forskere, er det grunn til å tro at det hadde blitt betydelig færre brukerfeil og posisjonsfeil. Det hadde nok blitt lagt stor vekt på å minimere disse feilkildene allerede i datainnsamlingen. Forskere har gjerne god kjennskap til hva analysen vil kreve av datasettet i form av kvalitet, nøyaktighet og presisjon. Deltakere uten forskningsbakgrunn har nok ikke like tydelig oppfatning av hvor viktig det er at observasjonene er nøyaktige og blir riktig kategorisert.

Bruk av rutenett og polygon-markeringer har vært nødvendig for å holde struktur på sorteringen av observasjonene. Uten slike metoder hadde sorteringsjobben blitt enda mer krevende. Denne erfaringen viser at sortering av folkeinnsamlet data, eller annen data som er posisjonert på et kart, krever struktur slik at ingen data utelates. Hvis dårlig struktur medfører at observasjoner blir utelatt, kan dette i seg selv være en feilkilde som kan påvirke resultatet av analysen. Dette

viser at forskerens metodevalg kan ha stor innvirkning på resultatet. Det er ikke kun datasettets opprinnelige kvalitet som har betydning.

6.11.2 For strenge krav i sorteringen?

Det er kun åtte prosent av det opprinnelige antallet observasjoner som er igjen etter at den detaljerte sorteringen er gjennomført, noe som betyr at størsteparten av observasjonene har blitt sortert bort. Dette resultatet samsvarer med studien til Aano et al. (2019), hvor kun ni prosent av de opprinnelige skaderegistreringene var igjen etter sorteringen. Her kan det diskuteres om sorteringen i denne masteroppgaven har vært for streng. Ble det fjernet flere observasjoner enn nødvendig, eller var det nødvendig å få sortert ut disse observasjonene for å minimere feilkilder så mye som mulig? Longley et al. (2005, s. 152) poengterer at data aldri automatisk skal bli tolket som sannhet, men at det er svært viktig å samle informasjon om kvaliteten til datasettet for å kunne vurdere om det faktisk er egnet til formålet. De observasjonene som er fjernet hadde usikkerheter som gjorde at de ble vurdert som uegnet til formålet. Sorteringen har blitt forsøkt holdt så objektiv som mulig. Et eksempel på det er at alle observasjoner som viste vannansamling, uavhengig av størrelse, ble beholdt. Det samme gjelder alle observasjoner lokalisert i bygg. Her har det vært en "en eller alle-logikk" for å holde sorteringen objektiv. Observasjoner der bildene ikke viste omgivelser har blitt fjernet, men det kan diskuteres om denne sorteringen har vært for streng. Disse har ikke blitt fjernet fordi de var tydelige feilkilder, men fordi det ikke var mulig å bekrefte om posisjonen var nøyaktig eller unøyaktig.

Sorteringen har fjernet observasjoner som regnes som tydelige feilkilder, men også observasjoner en kunne anta inneholdt mulige feilkilder. Et eksempel på en tydelig feilkilde er observasjoner lokalisert i bygg. Disse observasjonene kan ikke brukes til å bekrefte forsøkninger da det ikke vil modelleres forsøkninger i bygninger. Feilkilder en kun kan anta er til stede kan for eksempel være observasjoner som ikke viser omgivelser. Her antas det at disse observasjonene kan være feilplasserte, og fra det en ser i histogrammet i figur 46, kan det være kun små feilplasseringer eller feilplasseringer så store som nærmere tre kilometer.

Sett i ettertid har sorteringen vært en nødvendighet for å sørge for at observasjonsdataen er egnet som valideringsgrunnlag, fordi uten sortering hadde det vært såpass store feilkilder at en ikke kunne trukket riktige konklusjoner. Dette viser at det folke innsamlede datasettet brukt i masteroppgaven ikke er egnet uten en detaljert sortering, hvor hele 92 prosent av det opprinnelige datasettet ble fjernet.

6.11.3 Subjektivitet som mulig feilkilde

Subjektivitet preger sorteringen av dette folkeinnsamlede datasettet i stor grad, spesielt den detaljerte sorteringen hvor det har blitt gjennomført en mer nøye vurdering av hver enkelt observasjon. I sorteringssteg 9 ble observasjoner med bilder hvor omgivelsene var vanskelig å tolke fjernet. Hvilke observasjoner som fjernes og ikke vil her være opp til individet som tolker. For enkelte er det kanskje mulig å gjenkjenne stedet selv med et mørkt bilde eller i et område i forandring, mens for andre vil dette være nærmest umulig. Muligheten til å tolke omgivelsene kan påvirkes av hvem som tolker bildet, og av kjennskapen denne personen har til studieområdet.

En stor del av den detaljerte sorteringen gikk ut på å identifisere observasjoner som var posisjonert feil i forhold til det bildet fra deltakerne viste. Erfaring med sortering i denne masteroppgaven tilsier at denne typen sortering krever god kjennskap til studieområdet og ulike kjennetegn ved omgivelsene. Metoden for identifisering av observasjoner som er posisjonert feil har svakheter ved at den mest sannsynlig ikke ville ha vært like enkel å gjennomføre for en som er ukjent i studieområdet. For min egen del har jeg bodd i Trondheim hele livet, og det gjør at jeg lett gjenkjenner ulike steder i studieområdet. Dette gjorde det lettere for meg å finne ut hvor i byen bildene var tatt. Metoden kan bli mer tidkrevende for personer som er ukjent i studieområdet, og det kan bli vanskeligere å oppdage feilplasserte observasjoner fordi manglende lokalkunnskap gjør at en ikke vet hvor den faktiske posisjonen burde være.

Google Earth og Google Street View ble brukt til å stedfeste observasjonene som hadde feil posisjon. Bildet av observasjonen ble sammenlignet med Google Earth og Google Street View for å finne vannoppsamlingens reelle posisjon. Dette ble også brukt som metode i studien til Yu et al. (2016) hvor det poengteres at metoden kan ha ført til potensielle feil i posisjonen til flomobservasjonene. Som nevnt tidligere har Google Street View gjort det mulig å studere omgivelsene i panorama, noe som gjør at en får en opplevelse av å være fysisk til stede i studieområdet. Det gjorde det mulig å anta deltakerens ståsted og vinkel da bildet ble tatt, noe som igjen gjorde det enklere å forstå hvor vannoppsamlingen var lokalisert. Det er likevel viktig å huske på at det fortsatt bare er bilder en studerer, noe som gjør at avstanden til ulike objekt i omgivelsene kan mistolkes til å være nærmere eller lengre borte enn de faktisk er. En annen ting som gjorde Google Street View til en noe problematisk metode å bruke til fastsetting av vannoppsamlingens reelle posisjon, var at bildene på Google Street View i flere tilfeller enten var eldre eller nyere enn bildene deltakerne hadde tatt. Dette kunne medføre at omgivelsene

rundt vannoppsamlingen ikke lenger var gjenkjennbar. Det kunne for eksempel være en byggeplass da bildet fra deltakeren ble tatt i 2019 eller 2020, men nå et ferdigbygd leilighetskompleks på bildene i Google Street View. Dette gjorde det noe vanskelig å validere om observasjonen lå på riktig plass, og spesielt vanskelig å bestemme vannoppsamlingens reelle posisjon. Som nevnt i metodekapittelet ble bilder av byggeplasser fjernet fordi omgivelsene ikke var gjenkjennbare.

Subjektiviteten i valideringen av om observasjonene hadde riktig posisjon, samt fastsetting av nye punkt som representerte vannoppsamlingens reelle posisjon, har derfor skapt flere potensielle feilkilder i datasettet. På grunn av dette ble det som tidligere nevnt satt en feilmargin på én meter i valideringen av forsøkninger. Dette betyr at forsøkninger ikke trengte å overlape med observasjoner - alle inntil én meter fra observasjoner ble godkjent. Diskusjonen rundt subjektivitet viser hvor problematisk det kan være å sortere folkeinnsamlet data.

6.12 Forsøkningsanalyse

Som Rød (2015) nevner bør målet være at de representasjonene vi bruker er så virkelighetsnære som mulig for at vi skal trekke riktige konklusjoner. Men tidligere studier fra Norge og utlandet har vist at også forenklede modeller av virkeligheten samsvarer godt med reelle tilfeller av overvann og flom (Skaaraas et al., 2015). Modelleringen av forsøkninger i denne masteroppgaven baserer seg som sagt på en forenkling av urbane dreneringsforhold. Dette gjør at forsøkningene som modelleres ikke nødvendigvis akkumulerer overvann, men det er mulig vannet dreneres bort kjapt. Som nevnt i masteroppgavens teoridel vil dreneringsmuligheter og bakkens infiltrasjonsrate spille en rolle for om nedbør akkumulerer i forsøknungen.

Som nevnt i metodekapittelet er en av de største utfordringene i urban flom-modellering at landskapet ikke representeres godt nok. Eksempelet med tre observasjoner registrert på broer viser at validering av forsøkninger i by ikke alltid er like enkelt. Urbane miljøer er komplekse, og det har derfor vært viktig å velge en god oppløsning på terrenngmodellen som gjør at bygninger, veier og annen infrastruktur representeres best mulig. Her har det vært viktig å velge den best mulige oppløsningen, 25 cm, for å sikre så god representasjon som mulig. Balstrøm (2021) sier at terrenngmodeller med oppløsning på én meter kan gi gode resultat, og det er mulig at det også kunne vært tilfelle i denne masteroppgaven. Men siden studieområdet ikke er så stort, vil jeg likevel si at mitt valg om 25 cm har vært en riktig vurdering. En god oppløsning er

også tilpasset det faktum at flere av de registrerte vannoppsamlingene er svært små, mer som sølepytter, noe som er vist i figur 22a.

Til tross for dette kan det likevel ha blitt modellert falske forsenkninger i studieområdet. Med dette menes at det ble identifisert forsenkninger som ikke var en forsenkning i virkeligheten, men ble identifisert på grunn av for eksempel feil ved terrengmodellen. Det er likevel grunn til å anta at de største feilene med terrengmodellen brukt i denne masteroppgaven er fjernet av Terratec før høydedataene ble offentlig publisert. Et eksempel på falske forsenkninger kan likevel være forsenkningene som modelleres i bygninger og i Nidelva. For å besvare denne masteroppgavens problemstilling har det vært nødvendig å gjennomføre romlige analyser. Det har derfor vært svært viktig å få fjernet disse forsenkningene som ikke regnes som gyldige forsenkninger.

7 Avslutning

7.1 Konklusjon

Innledningsvis ble det presentert en hovedproblemstilling som skulle besvares i masteroppgaven. *Hvor egnet er folkeinnsamlet data om overvann til å validere at GIS-modellerte forskninger er reelle vannoppsamlinger?* For å kunne besvare dette ble det definert en underproblemstilling: *Hva kjennetegner det folkeinnsamlede datasettet om overvann?*

Det folkeinnsamlede datasettet brukt i denne masteroppgaven kjennetegnes av heterogenitet. Bildene fra deltakerne holder varierende kvalitet, enkelte bilder er nyttige i valideringen mens andre er sortert ut da de regnes som potensielle feilkilder. Datasettet preges av brukerfeil som har ført til feilregistrering av observasjoner, noe som gjør at store deler av datagrunnlaget fjernes. I tillegg er det avdekket både små og store posisjonsfeil, det er kun 56 observasjoner som regnes som nøyaktige. I enkelte områder av studieområdet er det overrepresentasjon av observasjoner, noe som gjør at valideringen ikke kan gjennomføres likt for hele studieområdet. I tillegg er det overregistrering av enkelte deltakere, noe som trolig har påvirket observasjonens kvalitet i en negativ retning.

Masteroppgavens resultat viser at folkeinnsamlet data om overvann er lite egnet til å validere om modellerte forskninger er reelle vannoppsamlinger. Mange feilkilder har gjort det nødvendig med et omfattende sorteringsarbeid som har bestått av både en grov og detaljert sortering. Så mye som 92 prosent av det opprinnelige datasettet er sortert bort for å minimere både tydelige og potensielle feilkilder. Det er også potensielle feilkilder knyttet til metodene brukt i oppgaven. Sorteringen av observasjonene preges i stor grad av subjektivitet, noe som kan ha ført til feil posisjonering av punkt. Terskelverdien på én meter for å bekrefte observasjoner regnes også som en potensiell feilkilde for resultatet. Ettersom kun åtte prosent av det opprinnelige datasettet har vært mulig å bruke i valideringen, gir dette grunnlag for å si at det folkeinnsamlede datasettet er lite egnet. Til tross for dette har 96 forskninger blitt bekreftet å være reelle vannoppsamlinger. Dette viser at det har vært mulig å bruke datasettet som valideringsgrunnlag, så det er ikke grunnlag for å si at det er helt uegnet. Selv om denne masteroppgaven viser at folkeinnsamlet data om overvann har vært uegnet til formålet, er det

trukket fram andre studier der resultatene har vært bedre. Dette viser at folkeinnsamlet data har et potensial innenfor overvannsmodellering.

7.2 Videre studier

Som konklusjonen viser er det flere usikkerheter knyttet til det folkeinnsamlede datasettet brukt i denne masteroppgaven, noe som gjør at datasettet vurderes å være lite egnet til validering av modellerte forsøknings i GIS. Som masteroppgaven har vist, har folkeinnsamlet data potensiale til å brukes som valideringsgrunnlag hvis feilkildene minimeres. En aktuell problemstilling for videre studier kan derfor være å undersøke hvordan datainnsamlingen kan forbedres slik at folkeinnsamlet data blir bedre egnet til formålet. Her kan en se på hvordan en skal minimere feilkilder i alle ledd av datainnsamlingen slik at datasettet krever mindre etterarbeid. Utvikling av en strukturert plan for datainnsamling som fokuserer på punktene som er trukket fram i konklusjonen kan være et viktig grep. Med en god struktur i forkant slipper en forhåpentligvis å legge ned tidkrevende arbeid i sortering. I tillegg kan dette føre til at en større andel av det opprinnelige datagrunnlaget kan beholdes, noe som gjør at en vil få en bedre representativitet i studieområdet. Dette kan gi mulighet til å trekke mer generelle konklusjoner.

Referanser

- Alabri, A. & Hunter, J. (2010). Enhancing the quality and trust of citizen science data. 2010 IEEE sixth international conference on e-science,
- Albă, C. D., Zamfir, A. G. & Boengiu, S. (2018). Urban hydrography and Bluespots map of Craiova (Romania). *Journal of the Geographical Institute "Jovan Cvijic", SASA*, 68(3), 321-332.
- American Society for Photogrammetry. (2015). Positional Accuracy Standards for Digital Geospatial Data. *Edition 1, Version 1, 81*, A1-A26.
- Argust tracking. (2020). *GPS Accuracy & Bouncing (Multipath)*. Argust tracking. <https://argustracking.zendesk.com/hc/en-us/articles/333757037696-GPS-Accuracy-Bouncing-Multipath->
- Assumpção, T. H., Popescu, I., Jonoski, A. & Solomatine, D. P. (2018). Citizen observations contributing to flood modelling: opportunities and challenges. *Hydrology and Earth System Sciences*, 22(2), 1473-1489. <https://doi.org/10.5194/hess-22-1473-2018>
- Balstrøm, T. (2021, 26. august). *Model bluespots to map flood risk*. LearnArcGIS. <https://learn.arcgis.com/en/projects/model-bluespots-to-map-flood-risk/>
- Bonney, R., Shirk, J. L., Phillips, T. B., Wiggins, A., Ballard, H. L., Miller-Rushing, A. J. & Parrish, J. K. (2014). Next steps for citizen science. *Science*, 343(6178), 1436-1437.
- Bratlie, R. (2015). Beregning av flomveier med eksempler på bruk i kommunal forvaltning. *Kart og Plan*, 75(1), 24-34.
- Degrossi, L. C., de Albuquerque, J. P., Fava, M. C. & Mendiondo, E. M. (2014). Flood Citizen Observatory: a crowdsourcing-based approach for flood risk management in Brazil. SEKE.
- Delaney, D. G., Sperling, C. D., Adams, C. S. & Leung, B. (2008). Marine invasive species: validation of citizen science and implications for national monitoring networks. *Biological Invasions*, 10(1), 117-128.
- Di Salvo, C., Ciotoli, G., Pennica, F. & Cavinato, G. P. (2017). Pluvial flood hazard in the city of Rome (Italy). *Journal of Maps*, 13(2), 545-553. <https://doi.org/10.1080/17445647.2017.1333968>
- Esri. (2017, 10. oktober). Analytics Geocoding: Delivering High Location Accuracy,. *ArcGIS Blog*. <https://www.esri.com/arcgis-blog/products/analytics/analytics/geocoding-delivering-high-location-accuracy/>

- Esri. (2021a). *How Fill works*. ArcGIS Pro. <https://pro.arcgis.com/en/pro-app/2.8/tool-reference/spatial-analyst/how-fill-works.htm>
- Esri. (2021b). *Near (Analysis)*. ArcGIS Pro. <https://pro.arcgis.com/en/pro-app/2.8/tool-reference/analysis/near.htm>
- Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), 211-221.
- Haklay, M. (2013). Citizen Science and Volunteered Geographic Information: Overview and Typology of Participation. I *Crowdsourcing Geographic Knowledge* (s. 105-122). https://doi.org/10.1007/978-94-007-4587-2_7
- Hanssen-Bauer, I., Drange, H., Førland, E., Roald, L., Børsheim, K., Hisdal, H., Lawrence, D., Nesje, A., Sandven, S. & Sorteberg, A. (2009). Klima i Norge 2100. *Bakgrunnsmateriale til NOU Klimatilpassing., Norsk klimasenter, Oslo, Norway.*
- Helmrich, A. M., Ruddell, B. L., Bessem, K., Chester, M. V., Chohan, N., Doerry, E., Eppinger, J., Garcia, M., Goodall, J. L. & Lowry, C. (2021). Opportunities for crowdsourcing in urban flood monitoring. *Environmental modelling & software*, 143, 105124.
- Hengl, T. & Evans, I. S. (2009). Mathematical and digital models of the land surface. *Developments in soil science*, 33, 31-63.
- Heywood, I. (2011). *An introduction to geographical information systems* (4. utg.). Pearson Education
- Kartverket. (2022). *Geovekst*. Kartverket. <https://www.kartverket.no/geodataarbeid/geovekst>
- Klimaservicesenter, N. (2016). Klimaprofil Sør-Trøndelag: Et kunnskapsgrunnlag for klimatilpassing. Norsk Klimaservicesenter.
- Kos, T., Markežic, I. & Pokrajcic, J. (2010). Effects of multipath reception on GPS positioning performance. Proceedings ELMAR-2010.
- Kutija, V., Bertsch, R., Glenis, V., Alderson, D., Parkin, G., Walsh, C., Robinson, J. & Kilsby, C. (2014). Model validation using crowd-sourced data from a large pluvial flood. International Conference on Hydroinformatics.
- Longley, P. A., Goodchild, M. F., Maguire, D. J. & Rhind, D. W. (2005). *Geographic information systems and science* (2. utg.). John Wiley & Sons.
- Meiforth, J. (2013). Modelling flood streams for urban planning in Trondheim, Norway. *UniGIS professional*.
- Murphy, P. N. C., Ogilvie, J., Meng, F.-R. & Arp, P. (2008). Stream network modelling using lidar and photogrammetric digital elevation models: a comparison and field verification. *Hydrological Processes*, 22(12), 1747-1754. <https://doi.org/10.1002/hyp.6770>

- Neset, T.-S., Wilk, J., Cruz, S., Graca, M., Rød, J., Maarse, M., Wallin, P. & Andersson, L. (2021). Co-designing a citizen science climate service. *Climate Services*, 24, 100273.
- Opach, T., Navarra, C., Rød, J. K. & Neset, T.-S. (2021). Pedestrian Routing and Perspectives: WayFinder's Route down the Lane—Come on with the Rain. *ISPRS International Journal of Geo-Information*, 10(6). <https://doi.org/10.3390/ijgi10060365>
- Rød, J. K. (2015). *GIS: verktøy for å forstå verden*. Fagbokforlaget.
- See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B., Laakso, M., Liu, H.-Y., Milčinski, G., Nikšič, M., Painho, M., Pödör, A., Olteanu-Raimond, A.-M. & Rutzinger, M. (2016). Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information. *ISPRS International Journal of Geo-Information*, 5(5), 55. <https://www.mdpi.com/2220-9964/5/5/55>
- Silvertown. (2009). A new dawn for citizen science. *Trends Ecol Evol*, 24(9), 467-471. <https://doi.org/10.1016/j.tree.2009.03.017>
- Skaaraas, H., Hansen, A., Riise, E., Stenersen, J., Refling, D., Johansen, R., Hjelle, H., Ebeltoft, M., Skofteland, H. & Kippeberg, G. (2015). Overvann i byer og tettsteder: som problem og ressurs. *NOU (Norges offentlige utredninger)*.
- Tönurist, P. & Surva, L. (2017). Is volunteering always voluntary? Between compulsion and coercion in co-production. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, 28(1), 223-247.
- Trondheim kommune. (2022). Temaplan for klimatilpasning. *Miljøenheten*.
- Vaze, J., Teng, J. & Spencer, G. (2010). Impact of DEM accuracy and resolution on topographic indices. *Environmental modelling & software*, 25(10), 1086-1098. <https://doi.org/10.1016/j.envsoft.2010.03.014>
- Viréhn, P. L. E. (2014). *Water on devious ways: a GIS analysis* [NTNU].
- Wang, Y., Chen, A. S., Fu, G., Djordjević, S., Zhang, C. & Savić, D. A. (2018). An integrated framework for high-resolution urban flood modelling considering multiple information sources and urban features. *Environmental modelling & software*, 107, 85-95.
- Wiggins, A. & Crowston, K. (2011). From conservation to crowdsourcing: A typology of citizen science. 2011 44th Hawaii international conference on system sciences.
- Yang, P., Ames, D. P., Fonseca, A., Anderson, D., Shrestha, R., Glenn, N. F. & Cao, Y. (2014). What is the effect of LiDAR-derived DEM resolution on large-scale watershed model results? *Environmental modelling & software*, 58, 48-57. <https://doi.org/10.1016/j.envsoft.2014.04.005>

- Yu, D., Yin, J. & Liu, M. (2016). Validating city-scale surface water flood modelling using crowd-sourced data. *Environmental Research Letters*, 11(12). <https://doi.org/10.1088/1748-9326/11/12/124011>
- Aano, A. H., Mora, R. Z. O., Lawrence, D. & Skaugen, T. (2019). *Bruk av registrerte overvannskader for validering av beregnede vannveier og overvannansamlinger (bluespots)* (nr. 44). Norges vassdrags- og energidirektorat.

