



Norges miljø- og  
biovitenskapelige  
universitet

**Masteroppgave 2020 30 stp**  
Fakultet for realfag og teknologi

## **Støyfiltrering av sanntidsmålinger fra norske automatiske nedbørstasjoner -**

En beskrivelse og vurdering av Meteorologisk  
Institutt sin nye korrigeringsalgoritme

Filtering noise from real-time automatic precipitation  
measurements in Norway - A description and evaluation of the  
Meteorological Institute's new correction algorithm

**Maylinn Cecilie Dramstad**  
Miljøfysikk og fornybar energi



## Forord

Med denne oppgaven fullfører jeg 5 år på siving i miljøfysikk og fornybar energi ved NMBU. De fem årene har gått skremmende fort, selv om det siste halvåret har følt litt langt. Arbeidet med denne masteroppgava har lært meg mer om meg selv enn jeg trodde var mulig, og enn jeg egentlig ønsket å vite. Det har vært spennende, lærerikt og slitsomt, hvilket også oppsummerer de andre årene på NMBU.

Aller først vil jeg få takke min fantastiske hovedveileder Mareile A. Wolff for uvurderlig hjelp og støtte. Dette hadde aldri gått uten henne. Jeg vil også benytte anledningen til å takke biveileder Arne Auen Grimenes som gjennom alle fem år har svart på dumme spørsmål og bidratt med faglig glede. Lasse Stenberg og Søren Raunsbæk Jørgensen på Meteorologisk Institutt fortjener også applaus for tålmodig hjelp med datatilgang og tekniske spørsmål.

Det er ikke fritt for at dette har vært en mentalt utfordrende reise. Jeg vil derfor takke mamma for korrekturlesing, motivasjonssitater og trøstende ord når masterskrivingen har blitt litt mye. Jeg vil også takke Sondre for å være husfar og stille med middager de siste ukene, i tillegg til store klemmer og ett dytt i siden når prokrastineringa tar overhånd. Til sist vil jeg takke mine kjære vestlendinger; Vilde for gåturer og Inger for å sitte sammen med meg i det til siste slutt. Og Malin selvsagt, for å vise meg at det finnes et liv også etter master.



## Sammendrag

Det er flere ulike måter å måle nedbør på, både manuelle og automatiske metoder, og de ulike metodene benyttes på ulike målestasjoner. I Norge benyttes norske- og svenske målere på manuelle målestasjoner, mens Geonor T-200B er den mest brukte måleren på automatiske målestasjoner. Ulike målemetoder påvirkes av ulike former for målefeil. Sensorene som benyttes på automatiske målestasjoner kan i tillegg være utsatt for både lavfrekvent og høyfrekvent støy. Dette kan resultere i unøyaktige målinger og kan gjøre sammenligningen av målinger utført med forskjellige metoder vanskelig. Det finnes mange studier som ser på effekten av støy og støyreducerende tiltak, men for automatiske målestasjoner er det fortsatt behov for en algoritme som filtrerer vekk støy fra målingene i etterkant og dermed gir mer nøyaktige resultater.

Meteorologisk Institutt har utviklet en ny støyfiltrerende algoritme med mål om å ha pålitelige målinger med en tidsoppløsning helt ned til ti minutter. Ved årsskiftet 2019/2020 ble denne tatt i bruk på et par utvalgte automatiske værstasjoner og planen er at den etter hvert skal erstatte den gamle filtreringsalgoritmen (*AlgGML*). For å evaluere den nye algoritmen er den blitt testet for avvik mellom implementering i ulike programmeringsspråk, i tillegg til at de to algoritmene (ny og gammel) ble sammenlignet med hverandre. Filtreringsevnen til den nye algoritmen ble testet på et syntetisk datasett tilsatt ti ulike støykarakteristikker. Det syntetiske datasettet er basert på målinger gjort av en Geonor T-200B værstasjon på Haukelisetter i 2019. Støyen som ble lagt til hadde ulikt standardavvik på mellom 0,001 mm og 1,00 mm. Algoritmen ble testet i tre ulike forsøkskonfigurasjoner.

Resultatene fra sammenligningen mellom *AlgPYTN* og *AlgGML* viser ingen tydelig forbedring ved bruk av den nye algoritmen, men det anbefales å sammenligne ny og gammel algoritme på et lengre datasett med mer støy. I denne studien ble det ikke observert en signifikant forskjell mellom implementering av den nye algoritmen i ulike programmeringsspråk. Støyanalysen indikerer at det ikke er en forskjell mellom de testede konfigurasjonene av algoritmen. Ikke uventet ble det høyest avvik og *RMSE*-verdier for støytypen med størst standardavvik. Sammenlignet med analysen av de fire ulike algoritmene som ble testet av Environment Canada, er den nye algoritmen fra Meteorologisk Institutt bedre på støyfiltrering av datasett med tilsatt ikke-normalisert støy.



## Abstract

There are several different ways to measure precipitation, both manual and automatic methods, and they are both in use at different precipitation stations. In Norway, Norwegian and Swedish gauges, are used at manual precipitation stations, while the Geonor T-200B is most widely used at automatic measuring stations. Different measurement methods are affected by different types of measurement errors. The sensors used at automatic precipitation stations can also be exposed to both low-frequency and high-frequency noise. This can result in inaccurate measurements and can make the comparison of measurements difficult. Many studies examine the effect of noise and noise reduction measures. However, for automatic precipitation stations, there is still a need for an algorithm that filters out noise from the recorded measurements and thus provides more accurate results.

The Norwegian Meteorological Institute has developed a new noise filtering algorithm with the aim of having reliable measurements with a time resolution as small as ten minutes. The use of this was initiated at the end of 2019/beginning of 2020, at a couple of selected automatic weather stations. The intention is that this new algorithm eventually shall replace the old filtering algorithm (*AlgGML*). To evaluate the new algorithm, it has been tested for discrepancies between implementation in different programming languages. In addition, the two algorithms (new and old) has been compared with each other. The filtering ability of the new algorithm was tested on a set of synthetic data to which ten different noise characteristics was added. The synthetic data set is based on measurements made by a Geonor T-200B weather station at Haukelisetter in 2019. The noise that was added had different standard deviations ranging from 0.001 mm to 1.00 mm. The algorithm was tested in three different configurations.

The results from the comparison between the new (*AlgPYTN*) and old (*AlgGML*) algorithms show no particular improvement through use of the new algorithm, but it is recommended to compare the new and old algorithm on a data set covering a longer time period and with more noise. Further, no significant difference was observed between implementation of the new algorithm in different programming languages in this study. The noise analysis indicates that there is no difference between the tested configurations of the algorithm. As could be expected, the highest deviation and RMSE values were found for the noise type with the largest standard deviation. Compared with the analysis of the four different algorithms tested by Environment Canada, the new algorithm from the Norwegian Meteorological Institute is better at filtering noise from data sets with added non-normalized noise.

# Innholdsfortegnelse

Forord .....	i
Sammendrag .....	iii
Abstract .....	v
1 Introduksjon .....	1
1.1 Bakgrunn .....	1
1.2 Motivasjon.....	3
1.3 Problemstilling.....	4
2 Teori.....	5
2.1 Skydannelse.....	5
2.2 Nedbørsdannelse .....	5
2.3 Nedbørmålere .....	6
2.3.1 Manuelle Nedbørmålere .....	6
2.3.2 Automatiske nedbørmålere .....	6
2.4 Feilkilder .....	10
2.5 Støy.....	11
2.5.1 Lavfrekvent støy .....	12
2.5.2 Høyfrekvent støy .....	13
2.6 Feilreduserende tiltak .....	14
3 Metode .....	15
3.1 Algoritmen.....	16
3.1.1 Variabler .....	16
3.1.2 Beskrivelse av algoritmen.....	18
3.1.3 Beskrivelse av den gamle algoritmen.....	24
3.1.4 Algoritmeversjoner.....	24
3.2 Værstasjoner og datasett .....	25
3.2.1 Hammerfest.....	26
3.2.2 Bjorli.....	27
3.2.3 Haukeliseter.....	27
3.3 Syntetisk datasett.....	28
3.4 Fordampning .....	29
3.5 Støyanalyse.....	30
4 Resultat.....	34
4.1 Hammerfest.....	34
4.2 Bjorli .....	38
4.3 Fordampning .....	40



4.4	Støyreduksjon.....	42
4.4.1	Støytype 1-6 .....	42
4.4.2	Støytype 7-10 .....	47
5	Diskusjon .....	52
5.1	Hammerfest.....	52
5.2	Bjorli .....	53
5.3	Fordampning .....	55
5.4	Støyreduksjon.....	56
5.4.1	Referanseverdien .....	56
5.4.2	Støytype 1-6 .....	57
5.4.3	Støytype 7-10 .....	59
5.5	Videre arbeid .....	61
6	Konklusjon .....	62
	Referanser .....	63
	Vedlegg.....	65
	A – Flytskjema .....	65
	B – Python script for AlgPYTN .....	65

# 1 Introduksjon

## 1.1 Bakgrunn

Det er vann i alt rundt oss, og det er et viktig element i mange av jordens prosesser. 71 % av jordens overflate er vann og opp til 4 % av atmosfæren er vann i form av vanndamp (NCSU, n.d.; Williams, 2014). Vannets kretsløp beskriver mekanismene rundt transport av vann på jorden/atmosfæren og de ulike tilstandene. Mengden nedbør er også viktig for varmetransporten i atmosfæren. Vann som fordampes fra jordoverflaten tar med seg latent varmeenergi opp i atmosfæren. Dette bidrar til å senke temperaturen ved overflatene, og gir opphav til meteorologiske systemer slik som lavtrykk eller stormsystemer (se f.eks. NASA, n.d.).

Nedbør og nedbørsmønstre er avhengig av geografisk plassering og topografi. I tropiske områder har man lange perioder med tørke mellom perioder med mye nedbør, mens andre steder har hyppigere og jevnere fordelte nedbørsituasjoner gjennom et år. Jorden sett under ett har en årlig gjennomsnittlig nedbørmengde på rundt 953 mm/m<sup>2</sup> (se f.eks. Barry & Chorley, 1968). Dette er skjevt fordelt, med noen områder, for eksempel Sahara og Arktis, som har årlig nedbør på under 50 mm, mens det i tropiske regnskogsområder faller det opp mot 2000 mm/m<sup>2</sup> i året (Hjermann & Larsen, 2019). Både nedbørmengde og nedbørsmønster varierer fra sted til sted.

De siste årene ser man et endret nedbørsmønster der man generelt får flere tilfeller av ekstremvær; noen områder får kraftigere nedbørsituasjoner, mens andre områder opplever mer ekstreme temperaturer (Benestad et al., 2019; IPCC, 2014). Slike endrede værmønstre påvirker oss på mange ulike måter. I 2018 var det en ekstremt tørr og varm sommer i Norge og store deler av Europa. Svært lite nedbør i perioden mai til juli resulterte i dårlige avlinger. Dette fikk konsekvenser for mange innen landbruket den påfølgende vinteren, da det ble mangel på fôr til mange husdyr. Flere måtte importere fôr fra utlandet, hvilket ga en økt bekymring for spredning av sykdommer og skadeorgansimer, i tillegg til økte kostnader (Norsk Landbruk, 2018). Samtidig ser man at klimaet i Norge de siste 100 årene har blitt våtere og våtere (Meteorologisk Institutt, 2017).

Nedbør og endringer i nedbør påvirker ikke bare landbrukets avlinger og dyrehold, men har også betydning for mange andre viktige samfunnsøkonomiske forhold. For eksempel har det en direkte påvirkning på strømproduksjonen, transport og infrastruktur. Uforutsigbare mengder nedbør kan få dramatiske konsekvenser i form av oversvømmelser, flom og jordskred. Dette er også forhold med stor økonomisk betydning, både for de som blir direkte berørt og for samfunnet som helhet. Dermed er riktig dimensjonering av vann- og avløpssystemer viktig da dette kan bidra til å forebygge flomsituasjoner. En annen viktig forebyggende faktor er riktig varsling av kommende nedbørhendelser.

Store nedbørmengder over et kort tidsrom kan være utslagsgivende for flom og skred.

Flomhendelser.no er en databaseportal levert av Norges vassdrag- og Energidirektorat (NVE) der de har samlet informasjon om flomhendelser i Norge. De siste ti årene er det registrert 21 flommer fordelt over store deler av landet og i 19 av disse tilfellene er regn og/eller bygenedbør listet opp som årsak til flommen (NVE, 2010-2020). I følge tall fra Norsk Naturskadepool og Finans Norge stod flom for et anslått erstatningskrav på rundt 111 millioner norske kroner i 2019, mens skred utgjorde ca. 77 millioner norske kroner samme år (Neverdal, 2020).

Store vannmengder kan med andre ord være en stor samfunnsøkonomisk kostnad, men det kan også være en stor ressurs i form av energiproduksjon. I Norge er det 335 vannverk og mer enn 1000 vannmagasiner fordelt rundt om i landet. Dette gjør at 99% av all kraftproduksjon i Norge kommer fra vannkraft (Statkraft, n.d.). Statkraft melder at det de senere årene har blitt registrert flere ekstreme nedbørsituasjoner hvilket har skapt utfordringer for vannmagasinene. For at vannmagasinene skal kunne ta unna for flom, er det viktig at man kan forutsi når vannføringen vil øke slik at man kan justere magasin størrelsen i forkant. Samtidig er det viktig at det er nok vann i magasinene til enhver tid for å dekke behovet for kraftproduksjon i Norge. God prediksjon av kommende nedbørhendelser og potensielle flomsituasjoner gjør det enklere å ta denne avveiningen. Dette er en annen årsak til at presise og pålitelige målinger av klimatiske faktorer er samfunnsøkonomisk viktig (Wolff, Isaksen, Ødemark, et al., 2013).

Observasjoner av vær og klima har blitt gjort i hundrevis av år, men metodene har utviklet seg. Forbedringer i observasjoner og mer nøyaktige resultater gjør oss bedre rustet til å forstå fremtiden og de kommende utfordringene. Å vite hva som er typiske nedbørmengder og nedbørmønstre i ulike områder gir forskere et nyttig verktøy for å analysere trender, lage gode modeller og gi gode varsler for kommende nedbørsituasjoner. Samtidig gir det bedre grunnlag for dimensjonering av for eksempel rørsystemer og åpne vannveier. Mer digitalisering av datainnsamling og prosessering gjør dette mulig, men man er fortsatt avhengig av at dataene er nøyaktige og pålitelige. Tidligere ble det i hovedsak brukt manuelle målinger av nedbør, men i senere tid er automatiske målestasjoner blitt mer og mer vanlig. Norge begynte å ta i bruk automatiske målestasjoner i 1993 (Dyngeseth, 2004). Det finnes ulike former for automatiske målestasjoner, men hovedprinsippene som følges er enten måling av nedbør ved å måle et oppsamlet volum, en oppsamlet vekt eller intensiteten på nedbøren som faller. De ulike måleinstrumentene har sine styrker og svakheter, men det er noen utfordringer som gjelder mer generelt.

En utfordring med automatiske nedbørmålere er støy. Det finnes ulike former for støy som påvirker målingene litt ulikt. Ofte deles støy inn i høyfrekvent og lavfrekvent støy (Nitu et al., 2018). En kan også oppleve usikkerhet i målingene som en ikke vet den klare årsaken til. Den samme typen påvirkning kan gi ulike resultater for ulike nedbørstyper eller ulike nedbørmålere (se f.eks. Sevruck et al., 2009).

Det er mange utfordringer ved måling av nedbør. Noen av utfordringene følger målemetode, mens andre følger nedbørtype. Vind er et eksempel som påvirker flere målemetoder og nedbørstyper, på ulikt vis. Forsøk fra WMOs SPICE-rapport peker på at turbulente vindstrømmer er noe som er avhengig av topografien rundt måleren og nedbørmålere plassert i samme område kan få ulike utslag (Nitu et al., 2018; Sevruck et al., 2009).

Vind kan bidra med høyfrekvent støy i form av vibrasjoner i armatur på automatiske nedbørmålere, eller ved å direkte påvirke nedbøren ved å gjøre at den ikke treffer måleren. Sistnevnte er spesielt en utfordring på stasjoner som måler fast nedbør og det er gjort mye forsøk som tar for seg akkurat dette (se f.eks. Rasmussen et al., 2012). World Meteorological Organization (WMO) ga i 2018 ut en rapport der de hadde testet ulike måleinstrumenter og instrumentkonfigurasjoner i forhold til snømåling på 20 testfelt fordelt over 15 land, med mål om å utvikle en anbefaling for måleinstrument og vindskjermingskonfigurasjon for måling av snø (Nitu et al., 2018). Det finnes med andre ord ulike måter å begrense feilen for eksempel vind skaper på nedbørmålingene.

Med gode korrigeringsalgoritmer kan mye av effekten støy har på målinger reduseres. I Canada økte årlig nedbør med rund 16,5 % i perioden 1960-2009, og mye av denne økningen skyldes korrigerings av dataene i etterkant (Mekis & Vincent, 2011). Leeper and Kochendorfer (2015) beskriver hvordan ulike filtreringsalgoritmer har en stor betydning for korrigeringen av fordampning. Yang et al. (2005) viser til store ulikheter i nøyaktighet i nedbørmålinger som følge av ulikt utstyr og ulike korrigeringsalgoritmer på tvers av landegrensene. Derfor er det behov for grundig dokumentasjon og informasjon om de ulike praksisene, slik at sammenligninger og modeller for store områder blir enklere.

Våren 2020 ble det publisert en artikkel i Atmospheric Measurement Techniques der Environment and Climate Change Canada presenterte sin postprosesseringsalgoritme (Ross et al., 2020). Her gikk de gjennom hvordan en gruppe kanadiske forskere har testet fire ulike algoritmer på syntetiske datasett med ulike former for støy, deriblant fordampning. I rapporten ser de på målinger gjort av Geonor T-200B nedbørmålere. Craig D. Smith (2009) så på forholdet mellom denne sensoren, underestimeringer av snø som følge av vind og ulike vindskjerminger i sin artikkel. Duchon (2008) antyder at feil som følge av døgnvariasjoner kan være på opp mot 10% av nedbørtotalen for denne type måleinstrument. Konklusjonen fra disse undersøkelsene er generelt at det som følge av støy og måleusikkerhet er et behov for å forbedre målingene ved hjelp av robuste prosesseringsmetoder (se f.eks. Sevruk et al., 2009).

## 1.2 Motivasjon

I Norge er det Meteorologisk institutt som har hovedansvaret for målingen og lagringen av klimatiske parameter slik som nedbør, temperatur og vind, i tillegg til at det er de som står for værvarsling gjennom blant annet yr.no. De innsamlede dataene brukes til forskning og modellering av nedbørstrenger, hvilket igjen gjør det mulig å sammenligne og analysere årsaker og konsekvenser av eventuelle endringer som skjer. For at disse analysene og modelleringene skal bli så gode som mulig, trengs det nøyaktige og pålitelige målinger av tidligere hendelser.

I 2004 ble det gjennomført en sammenligning mellom manuelle målinger og Geonormålinger fra tre ulike stasjoner i Norge (Dyngeseth, 2004). Denne sammenligningen viste at Geonormålinger og manuelle målinger stort sett stemte godt overens, men med noen unntak. Disse unntakene ble gjennomgått i en ny rapport samme år. Her viste det seg at mange tilfellene der det var større differanser mellom manuelle måleverdier og Geonormålinger skyldtes ulike støyfaktorer slik som ekstreme temperaturer eller veldig nedbørsrike perioder (Mathisen, 2004). På bakgrunn av resultatene fra disse rapportene ser man at det er nødvendig med en algoritme som reduserer denne type støy i målingene for å gjøre målingene mer nøyaktige.

Den algoritmen som ble valgt tok utgangspunkt i en algoritme brukt på svenske Geonor-målestasjoner. Den svenske algoritmen ble utviklet ved Sveriges meteorologiska och hyderologiska institutt (SMHI), men ble tilpasset norske forhold før den ble implementert på norske stasjoner. Algoritmen var utviklet for en tidsoppløsning på timer, men da det var ønske om å øke innsamlingsfrekvensen til ti minutter rundt 2016 viste algoritmen svakheter med tanke på støyreduisering.

Den optimale algoritmen må kunne gi presise resultater ved flere ulike nedbørsituasjoner på ulike lokasjoner og under ulike forhold. Meteorologisk Institutt begynte derfor utviklingen av en ny

algoritme med mulighet for en tidsoppløsning på ti minutter. I tillegg så man et behov for en algoritme som var pålitelige både i normale nedbørsituasjoner og i situasjoner med mer ekstreme værforhold. Den nye algoritmen ble tatt i bruk i januar 2020 og tar utgangspunkt i værstasjoner som benytter den automatiske nedbørsensoren Geonor T-200B.

### 1.3 Problemstilling

Hovedmålsetningen for denne masteroppgaven er å evaluere Meteorologisk Institutt sin nye filtreringsalgoritme benyttet på nedbørdata fra værstasjoner som benytter Geonor T-200B som nedbørmåler. Fremgangsmåten er inspirert av den fremgangsmåten Ross et al. (2020) beskriver i sin artikkel.

Oppgavens hovedproblemstilling er derfor

«Hvor godt filtrerer den nye algoritmen for ulike støytyper på nedbørdata hentet fra værstasjoner med Geonor T-200B sensorer?»

Det er også sett på følgende underproblemstillinger

- Er det store forskjeller i resultater som følge av ulik software på loggerne plassert på værstasjonene?
- Filtrerer algoritmen like godt for ulike typer støy?
- Er algoritmens evne til støyfiltrering avhengig av tidsoppløsning?
- Hvordan fungerer algoritmen på datasett med fordampning?
- Hvor stort standardavvik kan det være i et datasett før algoritmens avvik blir for store?

## 2 Teori

Denne masteroppgaven ser på korrigeringen av nedbørsmålinger hentet fra tre værstasjoner plassert på ulike steder i Norge, med ulik topografi og værforhold. Denne delen tar for seg den grunnleggende teorien og bakgrunnskunnskapen som er nødvendig for å forstå den videre analysen. Teorien vil blant annet gå inn på ulike nedbørstyper og hvordan disse dannes, hvilke målemetoder som finnes for nedbør og hvilke utfordringer som følger med dem.

Dersom ikke andre referanser er oppgitt er teorien hentet fra boken *Atmosphere, weather and climate* skrevet av Roger G. Barry og Richard J. Chorley (1968).

### 2.1 Skydannelse

Atmosfæren består av ulike gasser, deriblant opp til 4 % vanndamp. Også iskrystaller og vanndråper er representert i atmosfæren. Konsentrasjonen av vanndamp endrer seg med vertikal og horisontal plassering, og vil for eksempel være høyere i luftlagene over områder med høy fordampning. Vanndampkonsentrasjonen er svært viktig for dannelsen av skyer i atmosfæren, da den avgjør den relative fuktigheten som igjen er avgjørende for kondensasjonen. Ved en gitt temperatur kan en luftpakke holde på en viss mengde vanndamp før den blir overmettet og vanndampen kondenserer. Denne grensen for mengden vanndamp en luftpakke kan holde på, kalles luftas metningstrykk. Metningstrykket øker med økt temperatur. Forholdet mellom den faktiske mengden vanndamp i luftpakka og den maksimale mengden vanndamp luftpakka kan holde på, kalles den relative fuktigheten. Økt mengde vanndamp som følge av for eksempel økt fordampning, eller avkjøling av luftpakka, øker den relative fuktigheten.

Skyer består av små vanndråper og skydannelsen avhenger dermed av at vanndampen kondenserer. For at dette skal skje, må det bli en overmetning av vanndamp, altså at må den relative fuktigheten øke til over 100 % (Andersen, 2011). Dersom det er kondensasjonskjerner til stede, dvs. små luftforurensningspartikler, kan kondensasjonen skje til tross for at luftpakka ikke er overmettet. Eksempler på luftforurensningspartikler er støv, aske eller salt, hvilket også kalles aerosoler. Vannløselige aerosoler har et lavere metningstrykk på overflaten enn hva metningstrykket i lufta rundt er. Dette gjør at den relative fuktigheten blir høyere akkurat på overflaten av disse, og vanndampen kondenserer lettere. Luftforurensninger kan dermed bidra til skydannelse.

### 2.2 Nedbørsdannelse

I forbindelse med nedbørsmålinger deles nedbør inn i underkategoriene fast og flytende nedbør. Flytende nedbør refererer til nedbør i form av regn, mens snø og hagl defineres som fast nedbør.

Nedbørsdannelse kan skje gjennom ulike prosesser. Hvilken prosess som finner sted avhenger av temperaturen i skyen (National Geographic Society, 2019). Sluttproduktet avhenger av temperaturen i skyen der nedbørdannelsen skjer, samt temperaturen i atmosfæren mellom skyen og bakken. Dersom frysepunktet er nærme nok jordoverflaten, dvs. lavere enn 300 m og med en bakketemperatur på under 4°C, faller nedbøren som snø heller enn regn. Dette er fordi ispartiklene som ble laget i skyen ikke rekker å smelte fullstendig mellom skyen og bakken.

Bergeron-Findeisen-prosessen finner sted i kalde skyer. Denne prosessen forklarer mesteparten av nedbøren ved høyere breddegrader. Her er skytemperaturen så lav at noen av skydråpene har frosset til ispartikler. Disse ispartiklene samler mer vanndamp enn de omkringliggende vanndråpene fordi metningstrykket er lavere over is enn over vann. Vanndampen fryser på ispartiklene, og når ispartiklene blir tunge nok faller de mot bakken. Det er lufttemperaturen mellom skyen og bakken som avgjør om nedbøren kommer som regn eller snø. Er temperaturen høy nok, smelter iskrystallene på vei ned og nedbøren faller som regn.

Den andre prosessen foregår i varmere skyer, hvilket betyr at det ikke er iskrystaller tilstede i skyen. Her har vanndamp kondensert til tunge vanndråper, og disse faller mot bakken. I fallet kolliderer de med mindre vanndråper som absorberes. Dette gjør at den opprinnelige vanndråpen vokser i størrelse og tyngde. På grunn av turbulens i skyene faller ikke dråpene rett, og det skjer dermed flere kollisjoner.

Dannelsen av hagl finner sted i blandingskyer med sterke vertikale vinder. Disse skyene har kuldegrader i de øvre skylagene, mens de nedre har varmegrader. Vanndråper fra de nedre skylagene fraktes opp i kulden av den sterke vertikale bevegelsen i skyen. Der fryser dråpene til is, før de igjen faller nedover i skyen. I fallet kolliderer ispartiklene med vanndråper som legger seg i et lag rundt ispartiklene. Vindene i skyen frakter dem på nytt opp til de kaldere skylagene der de fryser til. Denne prosessen med frysing i de øvre skylagene og fusjon med vanndråper i de nedre skylagene gjentas flere ganger frem til ispartiklenes tyngde overviner de vertikale vindene og de faller mot bakken.

## 2.3 Nedbørmålere

Det finnes ulike måter å måle nedbør på med utgangspunkt i ulike måleprinsipper. De forskjellige metodene har ulike styrker og svakheter. Mer om generelle feilkilder kan leses i kapittel 2.4. Det skilles mellom manuelle og automatiske nedbørmålere, i tillegg til at det skilles på oppsamlende og ikke-opsamlende målere.

### 2.3.1 Manuelle Nedbørmålere

Manuelle målestasjoner er nedbørmålere der den oppsamlede nedbøren registreres av en observatør. På slike målestasjoner samles nedbøren opp i en beholder. Innholdet tømmes i et målebeger én til fire ganger daglig til bestemte tidspunkter, og volumet registreres. Dersom nedbøren har kommet som fast nedbør, smeltes denne før avlesning. Dette gjøres for å korrigere for ulik tetthet hos fast og flytende nedbør. Manuelle nedbørmålere benyttes i store deler av verden. I USA har de et eget nettverk kalt COOP (The National Weather Service's Cooperative Observer Program) som baserer seg på manuelle observasjoner av temperatur og nedbør gjort rundt om i USA (Holder et al., 2006). I Norge er det vanlig å benytte «norske og svenske målere» (Førland et al., 1996).

### 2.3.2 Automatiske nedbørmålere

Automatiske nedbørmålere kommer i flere varianter som kategoriseres etter om de er oppsamlende eller ikke-opsamlende. Oppsamlende nedbørmålere registrerer enten vekt eller volum, mens de

fleste ikke-oppsamlende nedbørmålere er optiske målere. Disse registrerer for eksempel hastighet eller størrelse på nedbørspartiklene, hvilket gjør nedbørmengde til en avledet størrelse.

### *Vippepluviometer*

Vippepluviometre, avbildet i figur 1, er et av instrumentene som benyttes til volummåling av nedbør. Denne nedbørmåleren består av to måleskjeer balanserende på et vippepunkt. Måleskjeene har et standardisert volum. Nedbøren fanges opp av en trakt og ledes ned i en av måleskjeene. Når denne blir full, vipper den over slik at den andre måleskjeen vender opp og innholdet i den første måleskjeen tømmes ut. Det er denne vippemekanismen som har gitt måleinstrumentet navnet vippepluviometer. Det er antallet ganger instrumentet vipper som registreres, og derfor er det viktig at måleskjeens volum er fastsatt (Meteorologisk Institutt, n.d.). Vippepluviometer er godt egnet til å måle intensiteten på nedbøren, og det er høy oppløsning på målingene. Dersom det er en periode med høy intensitet, blir det imidlertid også en større usikkerhet i forhold til hvor mye nedbør som ikke fanges opp av måleskjeene i løpet av de tidels sekundene vippepluviometeret bruker på å tømmes (Duchon & Essenberg, 2001). Vippepluviometre er godt egnet i sommermånedene der nedbøren faller som regn. Dersom nedbøren faller som snø, må den smeltes før den kan måles av instrumentet. Dette gjøres ved at trakta varmes opp. Det å finne riktig smeltetemperatur er en vanskelig prosess, da for høy temperatur fører til at noe av væska fordamper, mens for lav temperatur resulterer i snøen ikke smelter.



*Figur 1: Figuren viser et vippepluviometer. Når instrumentet er installert, er det plassert i en beskyttende beholder som skjermer mot ytre påvirkninger.*

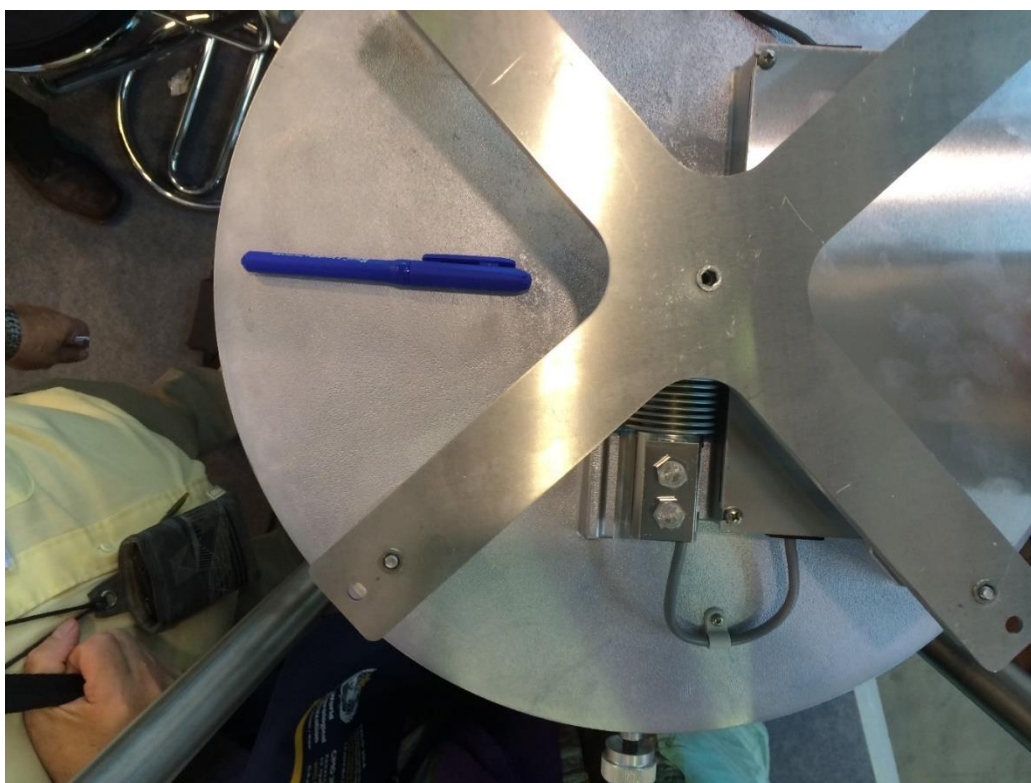


### Vektmålere

Den andre formen for oppsamlede nedbørmålere er vektmålere. Disse samler opp nedbør i en bøtte og måler vekten av bøtta. Denne metoden tar ikke hensyn til volum, og dermed trenger ikke nedbøren smeltes før den registreres (Nitu et al., 2018). Dette er årsaken til at vektmålere er bedre egnet som helårsmålere enn for eksempel vippepluviometre. En annen fordel med disse nedbørmålerne er at de samler nedbør selv om de ikke er operasjonelle, for eksempel som følge av strøm- eller kommunikasjonsbrudd. Dette er fordi estimeringen skjer på bakgrunn av vektendring over et tidsrom, hvilket betyr at det er mulig å estimere nedbørsmengde også i disse periodene. Disse målerne er derfor mindre utsatt for feilestimeringer som følge av operasjonell svikt. Avhengig av nedbørsmengde og oppsamlerens kapasitet må måleren tømmes to til seks ganger per år (Nitu et al., 2018).

Vekten av den oppsamlede nedbøren kan enten måles ved bruk av en vektcelle under bøtta eller ved hjelp av vibrerende strenger. I Norge er det det sistnevnte prinsippet som er vanligst på automatiske målestasjoner. Instrumentet som benyttes på norske automatiske værstasjoner er av typen Geonor T-200B.

Nedbørmåleren Ott Pluvio2 er et eksempel på en nedbørmåler som benytter en vektcelle til å måle oppsamlet nedbør. Vektcellen er avbildet i figur 2. Vektcellen fungerer på samme måte som en kjøkkenvekt, men er kalibrert for vekten av oppsamleren når den er tom eller inneholder frostvæske eller olje. Korrigeringen av målingene skjer i selve instrumentet, slik at verdiene som hentes ut er ferdig korrigert.



Figur 2: Figuren illustrerer vektcellen som benyttes til nedbørmålinger som benytter vektcelleprinsippet

Værstasjonene det er hentet data fra til denne oppgaven benytter Geonor T-200B som nedbørmåler, avbildet i figur 3. Denne nedbørmåleren er utviklet av det norske geotekniske instituttet (NGI) på 1960-tallet, og tar utgangspunkt i prinsippet om bruken av vibrerende strenger som målemetode. Oppsamleren henger stabilt i tre oppheng av kjetting. Deretter bytter man ut opptil tre av disse med en stålstring koblet til en sensor. Resonansfrekvensen av stålstringene kan bestemmes på grunn av kjent tykkelse og lengde av strengen, og stålstringene blir satt til å vibrere med denne (Bakkehøi et al., 1985). Maksimal kapasitet for oppsamleren kan variere fra værstasjon til værstasjon, men de har et fastsatt oppsamlerareal på 200 cm<sup>2</sup>. Når innholdet i bøtta endrer seg, endres spenningen i disse stålstringene og dermed endres også frekvensen vekk fra resonansfrekvensen. Sensoren registrerer frekvensen stålstringen den er koblet på vibrerer med. Denne frekvensen konverteres til millimeter ved hjelp av et annengradspolynom som vist i formel 1

$$P = A \times (f - f_0) + B \times (f - f_0)^2, \quad (1)$$

der  $P$  er lasten gitt i cm,  $f$  er registrert frekvens og  $f_0$  er resonansfrekvens. Det er de tre parameterne  $A$ ,  $B$  og  $f_0$  som bestemmes ved kalibrering av sensoren og som er unike for hver streng. En endring i bøttas tyngde som følge av endret innhold vil gi en endring i registrert frekvens,  $f$ . Ved hjelp av formel 1 kan denne endringen beregnes i centimeter, og deretter millimeter. Geonor-sensorene har en sensitivitet på 0,1 mm.



Figur 3: Bildet viser en Geonor nedbørmåler med Alter-skjerming plassert på NMBU sitt testområde på Søråsjordet, Ås.

### Optiske metoder

Distrometer er en form for optiske nedbørmålere som benytter laser. Disse er ikke-opsamlende nedbørmålere og måler hovedsakelig nedbørtype ved å se på fallhastighet og partikkelstørrelse av det som bryter laserstrålen. Ved å gjøre antagelser for nedbørtype, form og tetthet kan man avlede nedbørmengden fra antallet målte partikler. Dette blir da en verdi med større usikkerheter enn om nedbørmengde måles direkte.

En annen type optisk måler er avbildet i figur 4. Disse er også ikke-opsamlende og sender ut et strømsignal som endres når laserstrålen brytes. Disse registrerer dermed intensiteten på det som bryter laserstrålen. Ved å benytte en terskelverdi for dette strømsignalet kan man bruke disse som Ja/Nei-sensorer for nedbør. Et strømsignal høyere enn den fastsatte terskelverdien betyr ja (det har falt nedbør) og under terskelverdien betyr nei (det har ikke falt nedbør). En kan avlede nedbørmengde fra intensiteten, men dette blir sjelden gjort da det blir mye usikkerhet i verdiene – opp mot 15% for regn og opp mot 30% for snø (Thies GmbH). Til tross for at disse sensorene egentlig måler intensitet benyttes de kun som Ja/Nei-sensorer på norske værstasjoner.



Figur 4: Bildet viser en Thies Clima benyttet som Ja/Nei- sensor på stasjonen Haukeliseter Testfelt (stasjonsnummer 33950).

## 2.4 Feilkilder

Nedbørmålinger er utsatt for feilkilder, hvor noen følger målemetode mens andre følger nedbørtype. Vind er en kilde til målefeil og usikkerhet i nedbørmålinger av flere årsaker. Vektmålere kan få målefeil som følge av vind-pumping, det vil si at oppsamlerbøtta løftes av vinden (Goodison et al., 1998). Dette vil gi negative vektsignaler og er dermed en feilkilde spesielt på vindutsatte stasjoner. Enkelte forsøk (se f.eks. Strangeways, 2004) viser at vind også kan skape turbulens i selve

nedbørmåleren. Dette påvirker både bøttas vekt, men kan også gjøre at oppsamlet nedbør forsvinner ut av måleren.

En annen feilkilde er underestimering av nedbør som følge av vind. En studie fra Duchon & Essenberg (2001) viser at uskjermede vippepluviometer i gjennomsnitt hadde en underestimering på 4% i regnsituasjoner, mens uskjermede vektmaalere i gjennomsnitt hadde en underestimering på 5%. Rasmussen et al. (2012) viser til underestimeringer av fast nedbør på rundt 30% for vindskjermede nedbørmålere.

'Wetting loss' er en feilkilde som gjelder de fleste nedbørmålere, og den skyldes vanndråper eller snøpartikler som henger igjen i måleren etter tømning (se f.eks. Mekis & Vincent, 2011). Dette kan være tømning av oppsamlere i automatiske vektmaalere eller i manuelle nedbørmålere. Det gjelder også vanndråper som henger igjen i måleskjeen på vippepluviometerne etter at de tømmes (WMO, 2018). Disse dråpene blir dermed ikke medregnet i nedbørmålingene. Dette kan også være en feilkilde ved manuell måling av fast nedbør, slik at volumetriske snømålinger må korrigeres for en 'wetting loss error' som kan være så høy som 0,3 mm per observasjon (Goodison et al., 1998).

I situasjoner med mye nedbør i form av sterkt regn, eller i byge-situasjoner, kan oppsamlet nedbør sprute ut av oppsamleren. Dette gir en underestimering av falt nedbør. Målerens utforming har en betydning for denne typen målefeil (Strangeways, 2004). Forsøkene beskrevet av Strangeways (2004) viser at nedbørmålere med lave kanter har større risiko for denne typen målefeil. Ved sterke nedbørhendelser, 200 mm/h, kan tiden det tar å vippe måleskjeene i vippepluviometere føre til en akkumulert feil på opptil 10% (Duchon & Essenberg, 2001).

Unøyaktighet ved avlesning kan gi feil for målinger gjort på manuelle målestasjoner. Forskjeller mellom observeringstidspunkt og registreringstidspunkt er en annen feilkilde på manuelle målestasjoner. Holder et al. (2006) sammenlignet manuelle og automatiske nedbørmålinger i Nord-Carolina. I tilfellet beskrevet i artikkelen ble målingene korrigert for forskjell i måletidspunkt, hvilket gav en forbedret korrelasjonskoeffisient på 0,50 for de aktuelle dataene (Holder et al., 2006).

I noen tilfeller har man sett at manuelle målestasjoner registrerer mer nedbør enn automatiske. Holder et al. (2006) antyder at dette kan skyldes tetting trakten på vippepluviometerne, heller enn mikrometeorologiske forskjeller mellom plasseringen av de automatiske og manuelle målestasjonene. Tetting kan oppstå som følge av at død vegetasjon faller ned i oppsamleren og hindrer den påfølgende nedbøren i å bli registrert. Dersom et dødt løv legger seg slik at det tetter for åpningen i trakta, blir nedbøren forhindret i å nå målebegeret, og det blir dermed feilregistreringer. Tetting kan også oppstå som følge av fast nedbør som fryser fast i oppsamleren (Rasmussen et al., 2012). Disse situasjonene er vanskelig å oppdage uten manuelle undersøkelser av måleren (Nitu et al., 2018).

Andre feilkilder er kalibreringsfeil av instrumenter, eller situasjoner der kalibreringen endrer seg over tid (Strangeways, 1996).

## 2.5 Støy

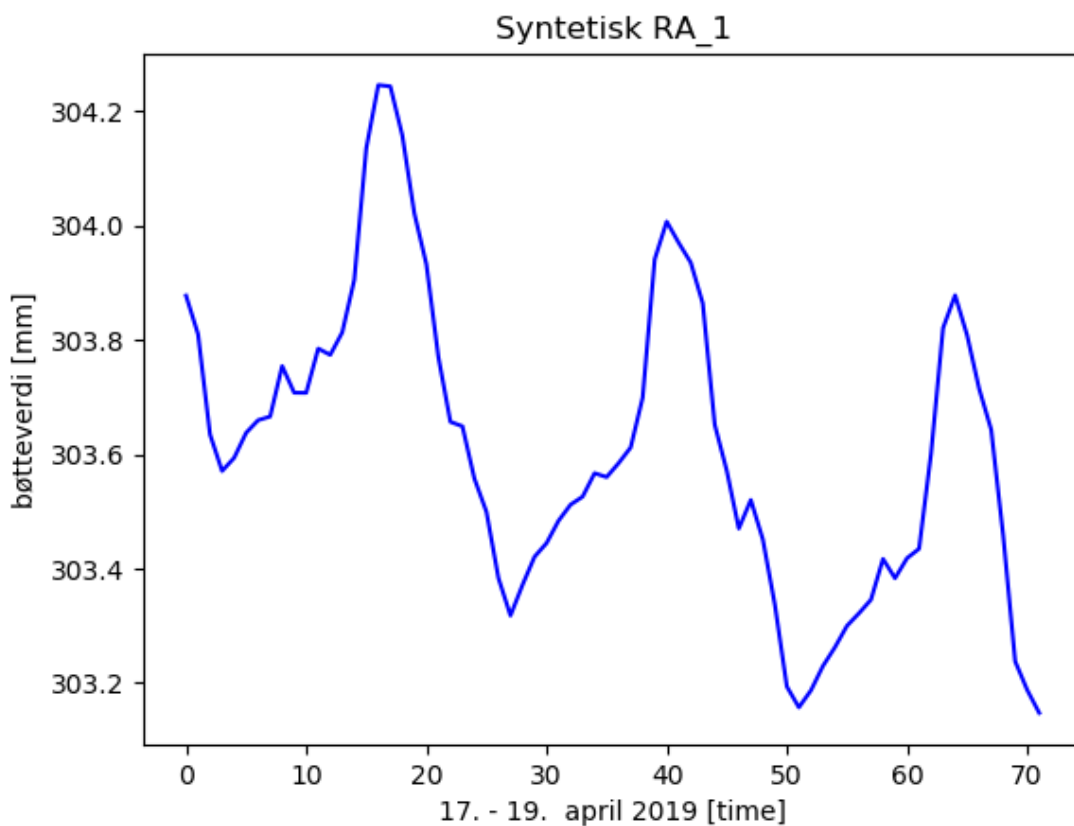
Geonor T-200B nedbørmålerne bruker frekvens til å estimere hvor mye nedbør som er samlet opp. En endring av denne frekvensen kan enten komme av endret vekt i bøtta eller det kan være støy som følge av ytre faktorer. Hvilke ytre faktorer som skaper støy varierer fra stasjon til stasjon, men

eksempelvis kan det være vind, temperaturendringer, elektromagnetisk støy eller fordampning. Støy kan deles inn i lavfrekvent og høyfrekvent støy (Nitu et al., 2018).

### 2.5.1 Lavfrekvent støy

Lavfrekvent støy er støy som påvirker målingene på en lavere tidsoppløsning, for eksempel timer eller dager. De to vanligste formene for lavfrekvent støy på nedbørmålinger er døgnvariasjon og fordampning (Nitu et al., 2018).

Døgnvariasjon i nedbørmålinger oppstår som følge av at temperaturen varierer gjennom et døgn, og påvirker målemetoder der sensoren har en form for temperaturavhengighet. For Geonor T-200B endres strengens egenskaper med temperaturen. Det gjør at strengens spenning og dermed resonansfrekvens endres, hvilket igjen endrer registrert frekvens. Figur 5 viser et signal fra en Geonor-stasjon over tre døgn. Her ser en tydelig at signalet endres med timene i løpet av et døgn, og den samme endringen gjentas de to påfølgende dagene. Disse svingningene er ikke store i utslag (opptil 1 mm i løpet av 24 t, avlest i figur 5). Ved direkte solinnstråling øker døgnvariasjonene, mens de blir jevnere på overskyede dager. På bakgrunn av dette er det ønskelig med en lav temperatursensitivitet i instrumentet, i tillegg er det nødvendig med en korrigeringsalgoritme som tar hensyn til døgnvariasjoner i nedbørestimeringene. En grundigere gjennomgang av denne typen problematikk på Geonor-sensorer er gitt i (Brækkan, 2007, 2013).



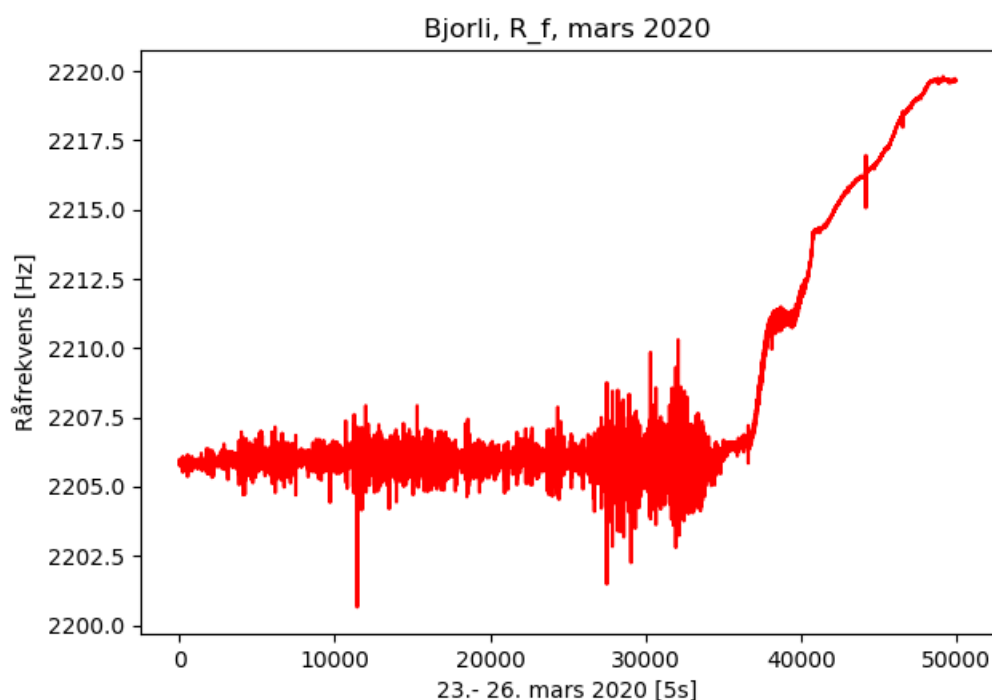
Figur 5: Figuren viser nedbørdata (bøtteverdi) fra tre dager i april 2019. Dataene er hentet fra værstasjonen på Haukelisetestfelt, en stasjon med Geonor T-200B med tre strenger og DFAR vindskjerming. Dette var en nedbørfri periode, slik at nedbørsignalet påvirkes av døgnvariasjon, og fordampning fører til en avtakende bøtteverdi over tid.

En annen form for lavfrekvent støy er fordampning. Fordampning er av større betydning for vektmaalere, da disse har et oppsamlet volum over lengre tid, og dermed har et lengre tidsrom der det kan skje en fordampning. Til sammenligning er ikke dette en like stor utfordring for manuelle maalere som tømmeres opp til fire ganger daglig, men noe forskning viser at fordampning likevel er av betydning for manuelle værstasjoner (Goodison et al., 1998). Utfordringen med fordampning er at nedbør samlet opp under en hendelse fordamper vekk før neste hendelse, og dermed blir estimert nedbørmengde feil. I figur 5 ser man tydelig fordampningen ved at bunnpunktet for hvert døgn synker slik at det blir en lineær nedgang fra døgn til døgn. Et forsøk gjort i Marshall, Colorado, sammenlignet fordampning fra to Geonor T-200B sensorer der den ene ble tilsatt olje for å begrense fordampningen. Resultatene herfra viser at det i løpet av sommermånedene juni til august ble en total fordampning på 228,5 mm i periodene uten nedbørhendelser for måleren uten olje (Leeper & Kochendorfer, 2015). Denne artikkelen peker imidlertid også på at ulike algoritmer kan korrigere godt for fordampning i situasjoner der det ikke er ønskelig å bruke olje, for eksempel av miljøhensyn. Strangeways (2004) og Leeper & Kochendorfer (2015) peker dessuten på at fordampning er avhengig av nedbørmålerens utforming.

### 2.5.2 Høyfrekvent støy

Høyfrekvent støy er støy som har innvirkning på og endres med en høy tidsoppløsning, dvs. målinger gjort med sekundintervaller eller minuttsintervaller. Årsaker til høyfrekvent støy kan være vind eller elektromagnetiske signaler.

Vind kan påvirke nedbørmålinger gjort på sekundnivå, men utslagene avhenger av vindstyrken (Wolff, Isaksen, Ødemark, et al., 2013). Sterk vind eller vindkast kan gi ekstra vibrasjoner i armaturen eller i selve strengene slik at registreringene blir feil. Dette er hovedsakelig et problem på de målerne som benytter vibrerende strenger, slik som nedbørmålerne fra Geonor. Vindpumping, det vil si effekten der vindkast løfter opp bøtta, kan også være en utfordring for nedbørmålere og føre til støy på signalene. Figur 6 viser et signal med høyfrekvent støy.



Figur 6: Denne figuren illustrerer høyfrekvent støy. Dette er målinger gjort hvert femte sekund, og en ser at bølgeverdien svinger ganske mye selv ved så hyppige målinger.

## 2.6 Feilreduserende tiltak

En måte å unngå tetting som følge av fastfrossen snø er å benytte oppvarmede målere (se f.eks. Rasmussen et al., 2001).

Fordampning er hovedsakelig et problem i perioder med høye lufttemperaturer, for eksempel sommermånedene. Det finnes ulike måter å hindre fordampning på, der det å tilsette olje i nedbørmålerne er et tiltak som er mye brukt på værstasjoner utsatt for fordampning. Oljen har en lavere tetthet enn vann og legger seg som en beskyttende hinne på vannflaten og hindrer dermed den oppsamlede nedbøren fra å fordampe. Om vinteren kan det være et problem at nedbøren fryser i nedbørmåleren. For å unngå dette tilsettes det frostvæske. Frostvæska blander seg med nedbøren og senker frysepunktet slik at den blandede væska ikke fryser til is ved 0°C.

Det er gjort flere forsøk der man tester ulike former for vindskjerming for å redusere de usikkerhetene som skyldes vind. Eksempler på slike forsøk er beskrevet i Rasmussen et al. (2012), Wolff et al. (2015) samt Nitu et al. (2018). Det finnes også ulike former for vindskjerming, men de vanligste formene er Alter-skjerming, Double Fence Intercomparison Reference (DFIR) eller stasjoner der man graver ned måleren slik at åpningen er på nivå med bakken. Forsøk viser at DFIR gir en vesentlig nedgang i vindhastighet på innsiden av vindskjermingen (Wolff, Isaksen, Ødemark, et al., 2013). I WMOs SPICE-rapport anbefaler de dobbelt skjerming (DFIR) over enkeltskjerming (Alter-skjerming), og enkeltskjerming over ingen skjerming (Nitu et al., 2018). WMO benytter DFIR som sin klare anbefaling for vindskjerming, og det er ofte nedbørmålere med denne typen vindskjerming som benyttes som referanse i forsøk (se f.eks. Smith, 2009).

### 3 Metode

Dette kapitlet beskriver de ulike aspektene rundt forsøkene. Mye av fremgangsmåten og flere av forsøkene er basert på fremgangsmåten Ross et al. (2020) har benyttet for å evaluere flere algoritmer som er i bruk i Canada. Med dette var det mulig å vurdere algoritmens evne til å filtrere støy, i tillegg til å sammenligne resultatene direkte med resultatene fra Ross et al. (2020) sine forsøk.

Algoritmen med forklaring av ulike variabler og algoritmeversjoner er gitt i kapittel 3.1. En gjennomgang av værstasjonene og hensikten med analysene gjort på Hammerfest Lufthavn og Bjorli er gitt i henholdsvis kapittel 3.2.1 og 3.2.2. Genereringen av det syntetiske datasettet brukt i flere av analysene er forklart i kapittel 3.3. Kapittel 3.4 forklarer metoden brukt for å teste algoritmens evne til å filtrere datasett med fordampning, mens kapittel 3.5 gjennomgår støyanalysene som er gjort.

Resultatene fra Hammerfest Lufthavn og Bjorli evalueres ved hjelp av absolutt og relativt avvik. Avvikene beregnes opp mot referanseverdien for hver av periodene, hvilket er oppgitt sammen med resultatene i kapittel 4.1 og 4.2. Referanseverdiene har en usikkerhet beregnet ved formelen

$$usikkerhet = \sqrt{std\_start^2 + std\_slutt^2}, \quad (2)$$

der  $std\_start$  er standardavviket for de første 60 målingene i en periode og  $std\_slutt$  er standardavviket for de siste 60 målingene i den samme perioden.

For støyanalysene benyttes i tillegg root mean square error,  $RMSE$ , og Pearsons korrelasjonskoeffisient,  $r$ , som evalueringemetode. Disse beregnes opp mot det syntetiske datasettet, beskrevet i kapittel 3.3. De ulike evalueringemetodene belyser ulike styrker og svakheter med modellen.  $RMSE$  tallfester standardavviket i residualene ved hjelp av formelen

$$RMSE = \sqrt{\frac{\sum(\text{faktisk verdi} - \text{beregnet verdi})^2}{N}}, \quad (3)$$

der  $N$  er antall observasjoner.  $RMSE$ -verdien forteller hvor godt de filtrerte verdiene stemmer overens med de faktiske resultatene, og en ønsker derfor en lav  $RMSE$ -verdi (Moody, 2019).

Pearsons korrelasjonskoeffisient,  $r$ , beskriver hvor sterkt det lineære forholdet mellom det filtrerte støysettet og det syntetiske settet er. Det er ønskelig med en høy  $r$ -verdi. Korrelasjonskoeffisienten beregnes med formelen

$$r = \frac{\sum(x - m_x)(y - m_y)}{\sqrt{\sum(x - m_x)^2 \sum(y - m_y)^2}}, \quad (4)$$

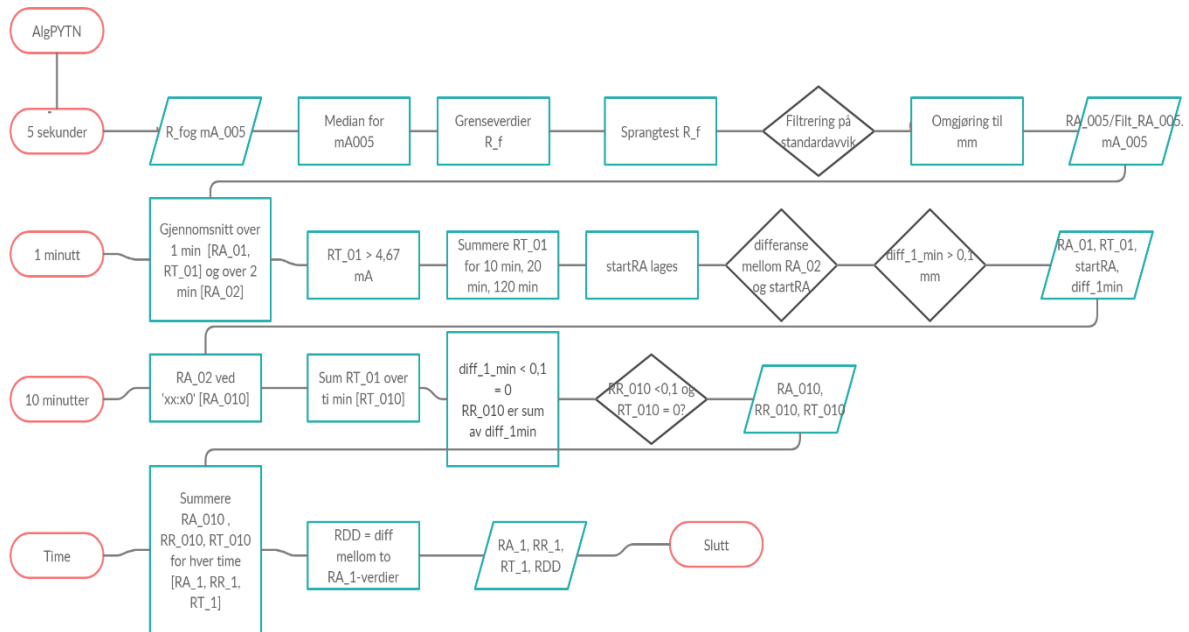
der  $x$  er en vektor med referanseverdier som har et gjennomsnitt  $m_x$  og  $y$  er en vektor med beregnede verdier med gjennomsnitt  $m_y$ .

Alle resultater avrundes til to desimaler, med mindre de første desimalene er null. Hvis det er tilfellet benyttes det to siffer fra første siffer forskjellig fra null.



### 3.1 Algoritmen

Det fullstendige flytskjemaet for algoritmen ligger vedlagt (vedlegg A). Figur 7 er et forenklet bilde av algoritmen. Variablene forklares i kapittel 3.1.1, mens algoritmen gjennomgås trinnvis i kapittel 3.1.2. De ulike værstasjonene har ulike loggere, og det kan derfor være små variasjoner i algoritmene på bakgrunn av ulik software. De ulike algoritmeversjonene som benyttes her er forklart i kapittel 3.1.4.



Figur 7: Figuren viser et forenklet flytskjema for algoritmen beskrevet i kapittel 3.1.2. Det fullstendige flytskjemaet ligger vedlagt (vedlegg A).

Algoritmen er designet av Lasse Stenberg ved Meteorologisk Institutt. Koden som programmert i Python ligger vedlagt (vedlegg B).

#### 3.1.1 Variabler

Algoritmen tar utgangspunkt i fire eksternt bestemte parametere og to målte variabler. De resterende variablene beregnes underveis. De eksternt bestemte parameterne er:

- Grenseverdier; De høyfrekvente målte verdiene sjekkes opp mot satte grenseverdier. Grenseverdiene settes slik at de stemmer overens med oppsamlerens kapasitet, dvs. med nedre grense tilsvarende 0 mm og øvre grense tilsvarende 600 eller 1000 mm. Alle målinger som går utenfor disse grensene settes til 0 og tas ikke med i de videre beregningene.
- Terskelverdi for sprangtest; Algoritmen gjennomfører en sprangtest der den sjekker differansen mellom to etterfølgende høyfrekvente målinger. Dersom som dette spranget er større enn  $|3|$  Hz, hvilket tilsvarende ca. 1,3 mm på værstasjonene brukt her, regnes målingen som usannsynlig og målingen settes til 0 og blir ikke med i de videre beregningene. Dette er for å fjerne store støtutslag, både positive og negative.
- Terskelverdi for nedbør; Denne er satt til 0,1 mm. Terskelverdien brukes til å bestemme hvilke utslag som ansees som nedbør og hvilke som trolig skyldes støt. Dersom utslagene er

lavere enn terskelverdien, antar algoritmen at dette er støy, og de tas ikke med i de videre beregningene. Denne korrigeringen skjer for målinger med minutt oppløsning.

- Terskelverdi for optisk sensor; Der det er et signal fra den optiske sensoren, benyttes en terskelverdi for disse målingene. Denne er satt til 4,67 mA, hvilket betyr at dersom målingene har en verdi høyere enn 4,67 mA antar algoritmen at det har falt nedbør og alle målinger under dette settes til 0. Den optiske sensoren er en Ja/Nei-sensor. Dette gjøres for målinger med minutt oppløsning.

De to høyfrekvente målte variablene er strømsignalet  $mA_{005}$  og frekvensen  $R_f$ . Strømsignalet kommer fra den optiske sensoren og har enheten mA. Frekvensen registreres av Geonor-sensoren og har enheten Hz. For Haukeliseter skjer omgjøringen til millimeter før algoritmens start.

De andre variablene beregnes underveis i algoritmen. Tabell 1 inneholder en beskrivelse av variablene som benyttes i algoritmen, med variabelnavn, enhet og eventuell beregningsmåte. En gjennomgang av algoritmen er beskrevet i kapittel 3.1.2. Variabelnavnene er organisert slik at  $RA$  ('rain accumulation') refererer til bømteverdi,  $RR$  ('rain amount') refererer til beregnet nedbør og  $RT$  ('rain time') er antall minutter med regn. De påfølgende tallene i parameternavnet forklarer om variabelen gjelder time, timinuttersperiode, minutt eller sekunder.

Tabell 1: Tabellen inneholder de ulike variabelnavnene med forklaring på hva de representerer og hvordan de er beregnet. En grundigere forklaring av beregningsmetodene er gitt i kapittel 3.1.2.

Variabelnavn	Variabeltype, enhet	Forklaring
$R_f$	Målt, [Hz]	Registrert frekvens fra Geonor-sensoren. Måles hvert femte/sjette sekund.
Filtrert_frekvens	Beregnet, [Hz]	Tilsvare $R_f$ etter at denne er sjekket mot grenseverdier, sprangtest og filtrering på standardavvik.
$mA_{005}$	Målt, [mA]	Strømsignalet fra den optiske sensoren logget hvert femte sekund. Tas løpende gjennomsnitt over to minutter.
$RA_{005}$	Beregnet, [mm]	Bøttas innhold hvert femte/sjette sekund i millimeter. Omgjøres fra $Filtrert\_frekvens$ for stasjonene Hammerfest Lufthavn og Bjorli. Er inputvariabel for Haukeliseter.
Filt_ $RA_{005}$	Beregnet, [mm]	Gjelder bare for Haukeliseter der $RA_{005}$ er inputvariabel. Er $RA_{005}$ etter at denne er sjekket for grenseverdier, sprangtest og filtrering på standardavvik.
$RA_{01}$	Beregnet, [mm]	Gjeldende bømteverdi for hvert minutt. Tilsvare gjennomsnittet av alle $RA_{005}$ med samme minutt i tidsstemplingen.
$RA_{02}$	Beregnet, [mm]	Løpende to minutters gjennomsnitt over $RA_{01}$ .
start $RA$	Beregnet, [mm]	Siste godkjente nivå i bøtta. Holder oversikt over eventuelle bømteømminger eller feilverdier. start $RA$ settes i utgangspunktet lik $RA_{02}$ , men justeres opp

		eller ned ved registrert nedbør, fordampning eller ved mye støy. Er en verdi per minutt.
<b>diff_1min</b>	Beregnet, [mm]	Er differansen mellom RA_02 og startRA, og gir dermed en endring i bømteverdi på minutt-nivå. Kan være både positiv og negativ. Er utgangspunktet for videre beregninger av nedbør.
<b>RT_01</b>	Beregnet, [boolean]	Beregnes ved å ta gjennomsnittet av alle mA_005 med samme minutt i tidsstemplingen. Deretter omgjøres denne verdien til 1 dersom gjennomsnittet er større en terskelverdien for strømsignalet, ellers settes den til 0. Forteller om det er et minutt med nedbør eller ikke.
<b>RA_010</b>	Beregnet, [mm]	Aktuelt bømtenivå ved slutten av timinuttersperioden, tilsvarer RA_02 registrert tt:m0.
<b>RR_010</b>	Beregnet, [mm]	Beregnet mengde nedbør som har falt de foregående ti minuttene, tilsvarer summen av diff_1min-verdiene i perioden som er større enn 0,1 mm og der RT_01 er forskjellig fra null minst en gang de siste 20 minuttene.
<b>RT_010</b>	Beregnet, [min]	Antall minutter med nedbør i den foregående timinuttersperioden. Er en summering av RT_01 for de siste ti minuttene.
<b>RA_1</b>	Beregnet, [mm]	Aktuelt bømtenivå ved slutten av en time, tilsvarer RA_02 registrert tt:00.
<b>RR_1</b>	Beregnet, [mm]	Beregnet mengde nedbør som har falt den foregående timen. Beregnes ved å summere RR_010 for de seks timinuttersperiodene i den aktuelle timen.
<b>RT_1</b>	Beregnet, [min]	Antallet minutter med nedbør i den foregående timen. Er en summering av RT_010 for de seks timinuttersperiodene i den aktuelle timen.
<b>RDD</b>	Beregnet, [mm]	Differansen mellom to etterfølgende bømteverdier RA_1. Kan brukes som estimat på nedbørmengde dersom RR_1 faller bort som følge av feil.

### 3.1.2 Beskrivelse av algoritmen

Hensikten med algoritmen er å filtrere vekk støy fra nedbørmålinger gjort på stasjoner med nedbørmålere av typen Geonor T-200B for å få mer korrekte målinger. Inngangsverdiene er målte frekvenser,  $R_f$ , og strømsignal, mA\_005, med en tidsoppløsning på fem eller seks sekunder, avhengig av stasjonens oppsett. En oversikt over hvilke stasjoner som har hvilken tidsoppløsning kan leses i kapittelet 3.2.

Trinnene i algoritmen avhenger av om de målte inputvariablene er frekvensen  $R_f$  (trinnene markert med a) eller allerede konvertert til nedbørenhet millimeter, RA\_005 (trinnene markert med b). Trinn 1-5 har en tidsoppløsning på fem eller seks sekunder, trinn 6- 1 har en tidsoppløsning på minutter, trinn 12-14 har en tidsoppløsning på ti minutter og trinn 15-16 har en tidsoppløsning på timer.

### Trinn 1

- a) Frekvens (målt i Hz) fra Geonorsensoren og et strømsignal (målt i mA) fra den optiske sensoren sendes inn i algoritmen. Medianen i et løpende to-minutters-intervall for strømsignalet velges ut. Disse variablene har henholdsvis navnene  $R\_f$  og  $mA\_005$ .
- b) Høyfrekvent målt bølgeverdi,  $RA\_005$ , sammen med strømsignalet sendes inn i algoritmen. Medianen i et løpende to-minutters-intervall for strømsignalet velges ut og blir variabelen  $mA\_005$ .

### Trinn 2

- a) Frekvensen  $R\_f$  sendes gjennom en grenseverditest og en sprangtest. Verdier utenfor de satte grensene eller med et sprang større enn terskelverdien for sprangtesten settes til 0. Grenseverdier og terskelverdi for sprangtesten beskrives i kapittel 3.1.1.
- b)  $RA\_005$  sendes gjennom en grenseverditest og en sprangtest. Verdier utenfor de satte grensene eller med et sprang større enn terskelverdien for sprangtesten settes til 0. Grenseverdier og terskelverdi for sprangtesten beskrives i kapittel 3.1.1.

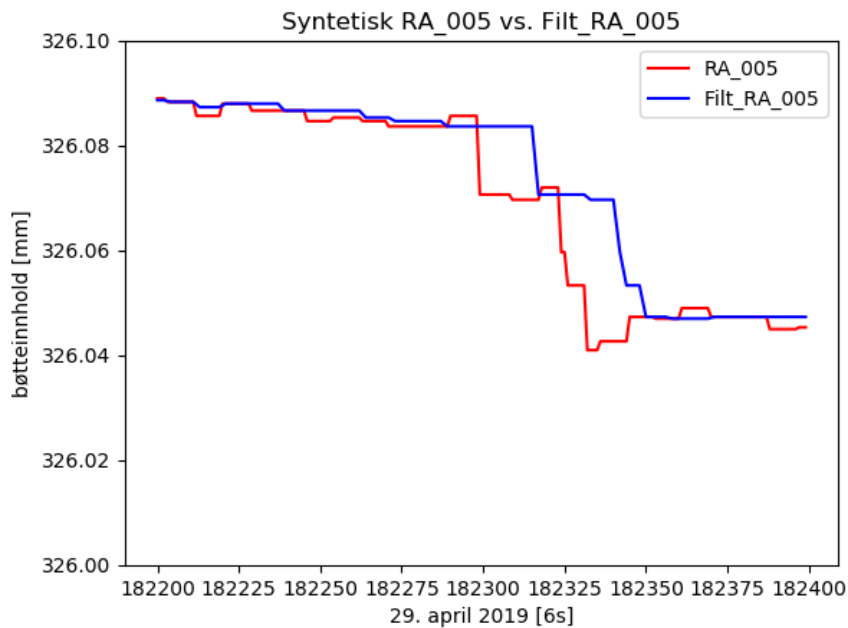
### Trinn 3

- a) Standardavviket for frekvensmålingene ( $R\_f$ ) beregnes over et løpende treminuttersintervall. Verdiene i intervallet sorteres i stigende rekkefølge.
- b) Standardavviket for bølgeverdimålingene ( $RA\_005$ ) beregnes over et løpende treminuttersintervall. Verdiene i intervallet sorteres i stigende rekkefølge.

### Trinn 4

- a) Avhengig av størrelsen på standardavviket beregnet i trinn 3a velges det ut en kvantilverdi. Et økt standardavvik gjør at en lavere kvantilverdi velges ut. Den valgte kvantilverdien blir den filtrerte frekvensverdien  $Filtrert\_frekvens$ . Tidsoppløsningen på fem sekunder beholdes i dette trinnet.
- b) Avhengig av størrelsen på standardavviket beregnet i trinn 3b velges det ut en kvantilverdi. Et økt standardavvik gjør at en lavere kvantilverdi velges ut. Den valgte kvantilverdien blir den filtrerte bølgeverdien  $Filt\_RA\_005$ . Tidsoppløsningen på seks sekunder beholdes i dette trinnet.

Figur 8 illustrerer hvordan signalet jevnes ut for de høyfrekvente målingene. I figuren tilsvarer den røde grafen det ufiltrerte signalet, mens den blå grafen har vært gjennom grenseverditesten, sprangtesten og filtreringen basert på standardavvik.



Figur 8: I figuren er bøtteverdien RA\_005 (rød) for det syntetiske settet plottet sammen med den filtrerte bøtteverdien Filt\_RA\_005 (blå). Den blå grafen har vært gjennom trinn 1-5 i algoritmebeskrivelsen.

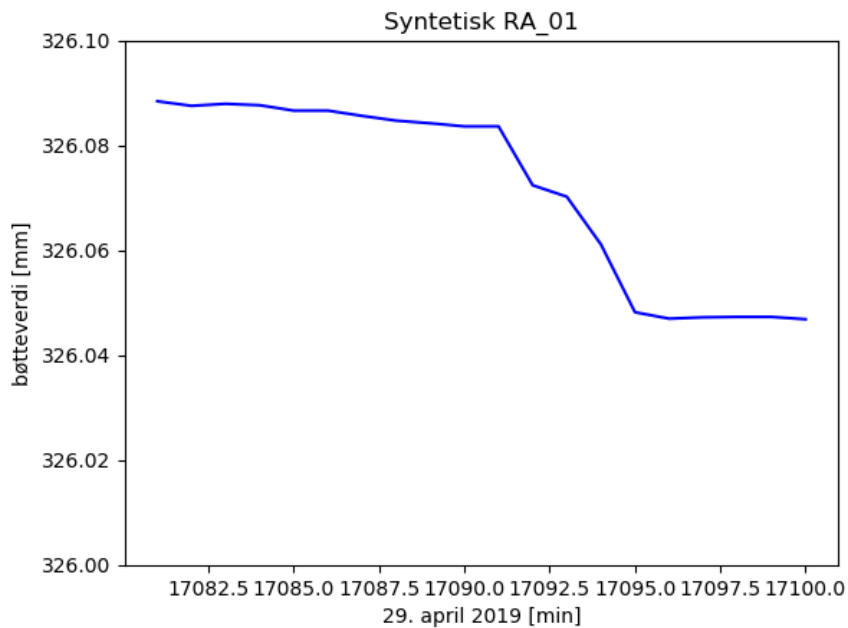
### **Trinn 5**

- a) Filtrert\_frekvens konverteres til bøtteverdi RA\_005 i millimeter ved hjelp av kalibreringskoeffisienter for målestrengen.

### **Trinn 6**

De RA\_005-verdiene (dersom det er b-trinnene som er fulgt, Filt\_RA\_005-verdiene) med samme minuttstempling, samles. Det tas et gjennomsnitt over disse verdiene og dette gjennomsnittet blir variabelen RA\_01. På samme måte tas det et gjennomsnitt over de tilsvarende mA\_005-verdiene, og denne variabelen får navnet RT\_01.

Figur 9 viser hvordan det samme tidsrommet som er plottet i figur 8 ser ut når signalet er behandlet til å være representative for minutter. Som en ser fra figuren er verdien på signalet beholdt, men signalet er jevnere og ikke like hakkete som i forrige figur.



Figur 9: Figuren viser grafen til bøtteverdien RA\_01 for en periode 29. april 2019. Dette er det samme tidsintervallet som vist i figur 8.

### **Trinn 7**

Det tas et løpende tominutters gjennomsnitt over RA\_01, denne får navnet RA\_02. RT\_01 beregnet i trinn 6 sjekkes opp mot terskelverdien for strømsignalet. RT\_01 bytter verdi til 1 dersom gjennomsnittet fra trinn 6 er over terskelverdien (dvs. over 4,76 mA), de resterende settes til 0. RT\_01 er dermed en booleansk variabel. Minutter der RT\_01 er lik 1 telles som nedbørminutter.

### **Trinn 8**

Det lages en oversikt for hvert minutt med hvor mange minutter det har regnet de foregående 10, 20 og 120 minuttene. Disse variablene får henholdsvis navnene `yes_RT10`, `yes_RT20` og `yes_RT120`.

### **Trinn 9**

Variabelen `startRA` opprettes. Denne er lik det siste godkjente nivået i oppsamleren, og er i utgangspunktet lik RA\_02 dersom det ikke er feilsignaler eller registrert bøttetømming.

### **Trinn 10**

For hvert minutt beregnes det en differanse mellom `startRA` og RA\_02. Denne får navnet `diff_1min`. Dersom denne differansen er større enn terskelverdien for nedbør (som er lik 0,1 mm),

og variabelen `yes_RT20` for dette minuttet er forskjellig fra 0, oppjusteres `startRA` for neste minutt med det hele antall terskelsteg `diff_1min` tilsvarer.

Dersom `diff_1min > terskelsteg`, men `yes_RT20 = 0`, oppdateres `startRA` for neste måling med et terskelsteg. Et terskelsteg er lik terskelverdien for nedbør, altså 0,1 mm.

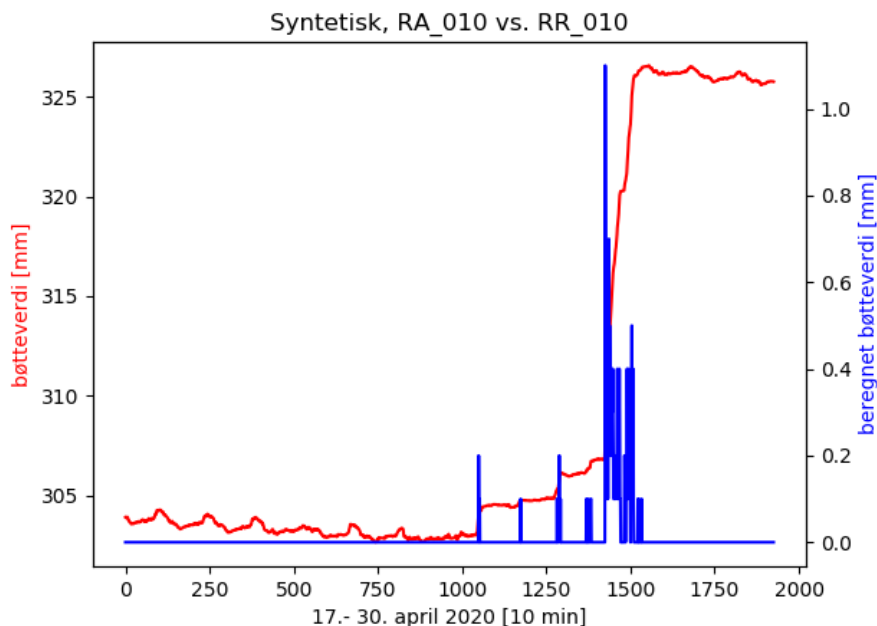
### Trinn 11

Samtidig som beregningene av `diff_1min` og oppdateringen av `startRA` skjer i trinn 10, sjekker algoritmen om `RA_02` er mer enn et terskelsteg lavere enn `startRA` (`RA_02 < startRA - terskelsteg`). Dersom dette er tilfelle, nedjusteres `startRA` for neste måling med ett terskelsteg. Dette er for å ta hensyn til en eventuell fordampning.

### Trinn 12

Algoritmen går over til å se på timinuttersperioder. Bøtteverdien `RA_010` defineres hvert tiende minutt (`tt:m0`), og settes lik `RA_02` registrert i det aktuelle minuttet. `RT_010` er summen av alle `RT_01` for de foregående ti minuttene.

Figur 10 illustrerer hvordan bøtteverdisignalet, `RA_010`, blir sendt ut med en tidsoppløsning på ti minutter for det syntetiske datasettet over perioden 17.- 30. april. I samme figur er den beregnede nedbøren, `RR_010`, plottet som blå graf mot høyre y-akse.



Figur 10: Figuren viser bøtteverdien `RA_010` (rød graf mot venstre y-akse) mot beregnet nedbør `RR_010` (blå graf mot høyre y-akse) for det syntetiske datasettet etter prosessering av algoritmen. Verdiene plottet er timinuttersverdier for perioden 17.- 30. april 2019.

### **Trinn 13**

Alle  $diff\_1min$  med en verdi lavere enn terskelverdien for nedbør (dvs. mindre enn 0,1 mm) settes til 0. De  $diff\_1min$ -verdiene som er ulik 0 trunkeres til en desimal og summeres for hver timinuttersperiode. Dette utgjør variabelen  $RR\_010$ . Den verdien som er registrert på tidspunktet  $tt:50$  er beregnet nedbør for perioden  $tt:41$ -  $tt:50$ .

### **Trinn 14**

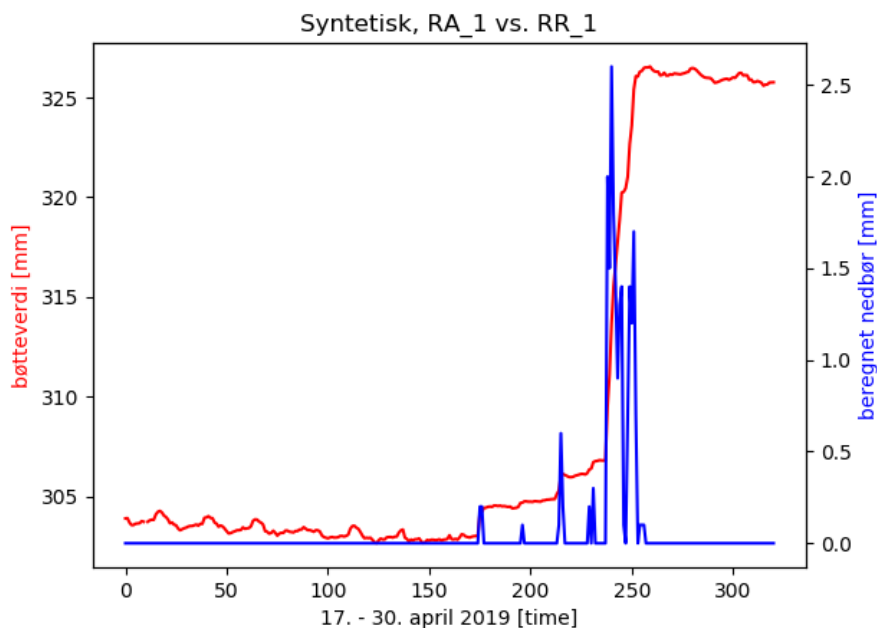
For hver av de registrerte verdiene  $RR\_010$ , sjekker algoritmen om  $RT\_010$  er lik 0. Dersom  $RT\_010$  er lik 0, settes  $RR\_010$  for samme periode til 0.

### **Trinn 15**

Bøtteverdien for hver time,  $RA\_1$ , settes lik  $RA\_010$ -verdien som er registret på hel time, dvs. ved tidspunktet  $tt:00$ .  $RR\_1$  og  $RT\_1$  beregnes som en summing av henholdsvis  $RR\_010$  og  $RT\_010$  for den foregående timen.

Bøtteverdiene  $RA\_1$  og beregnet nedbørmengde  $RR\_1$  med timesoppløsning for 17.- 30. april 2019 vises i figur 11. Den blå kurven for beregnet nedbørmengde per time viser de samme variasjonene som kurven for beregnet nedbørmengde per timinuttersperioder vist i figur 10. Kurven med timesoppløsning er som forventet glattere enn kurven for signalet med timinuttersoppløsning vist i figur 10. Se for eksempel perioden 1500-2000 i figur 10 sammenlignet med perioden 250-300 i figur 11. Skalaen til høyre i figur 11 går opp til 2,5 mm for å ta hensyn til høyere nedbørtilvekst per time enn nedbørtilvekst per ti minutter, der sistnevnte har en maksimumsverdi på høyre y-akse på bare 1,0 mm.





Figur 11: Figuren viser bøtteveerdi  $RA_1$  (rød graf mot venstre y-akse) mot beregnet nedbørmengde  $RR_1$  (blå graf mot høyre y-akse) for det syntetiske datasettet etter behandling av algoritmen. Verdiene plottet er timesverdier for perioden 17.-30. april 2019.

## Trinn 16

RDD beregnes som differansen mellom to etterfølgende bøtteveerdi,  $RA_1$ . Dersom  $RR$ -verdiene faller vekk som følge av kommunikasjonsfeil eller loggerfeil, benyttes RDD som estimat på falt nedbør.

### 3.1.3 Beskrivelse av den gamle algoritmen

*AlgGML*, den gamle algoritmen benyttet for korrigerende av nedbørmålinger, var basert på en algoritme utviklet hos Sveriges meteorologiska och hydrologiska institut (SMHI). Denne algoritmen behandler en timinuttersperiode av gangen. Utgangspunktet er bøtteveerdi i begynnelsen av perioden. Deretter bestemmes maksimal bøtteveerdi i perioden og påfølgende minimumsverdi innenfor den samme timinuttersperioden. Denne minimumsverdien benyttes som korrigeret verdi, og er startverdi i neste timinuttersperiode. Algoritmen godtar bare en akkumulering i nedbør dersom det er registrert nedbør av Ja/Nei-sensoren. Akkumuleringen beregnes som differansen mellom startverdi i intervallet og den korrigerte verdien. Det er denne akkumuleringen som benyttes som beregnet nedbør.

### 3.1.4 Algoritmeversjoner

*AlgGML* henviser til den gamle korrigeringsalgoritmen Meteorologisk Institutt benyttet på sine værstasjoner. Denne algoritmeversjonen er fortsatt i bruk på noen stasjoner, men blir byttet ut med den nye versjonen beskrevet i kapittel 3.1.2. Prinsippene bak *AlgGML* er beskrevet i kapittel 3.1.3.

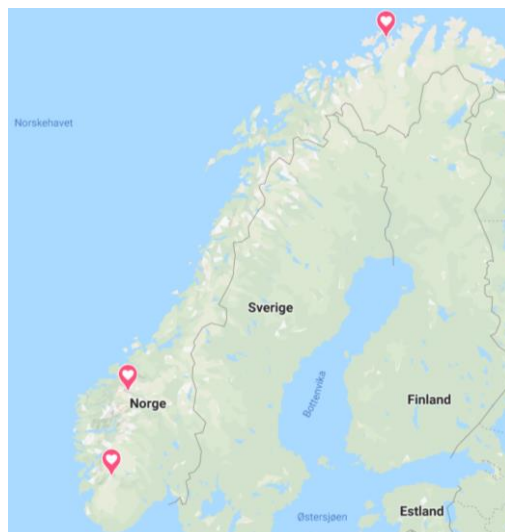
På værstasjonen Hammerfest Lufthavn blir det som forklart nærmere i kapittel 3.2.1 benyttet en pc-løsning som logger, og algoritmeversjonen som benyttes her refereres til som *AlgFLOS* i den videre teksten. Denne loggerløsningen benytter programmeringsspråket Pascal, Kylix, for Linux.

Algoritmen som beskrevet i kapittel 3.1.2 ble overført til Python. Denne algoritmeversjonen har fått navnet *AlgPYTN*. *AlgPYTN* blir brukt til å gjennomføre støyanalysene som beskrevet i kapittel 3.5. I tillegg blir datasettene fra Hammerfest og Bjorli testet for forskjeller knyttet til software ved hjelp av denne versjonen. Pythonversjonen som blir brukt er for programmering og kjøring av *AlgPYTN* er Python 3.7.1 (Van Rossum & Drake, 2009).

På noen værstasjoner, eksempelvis Bjorli, benyttes det Campbell-loggere. Algoritmeversjonen benyttet på disse loggerne er versjonen *AlgCAMP*. Disse benytter programmeringsspråket CRBasic, hvilket er et dataprogrammeringsspråk utviklet spesielt for Campbell-loggere.

### 3.2 Værstasjoner og datasett

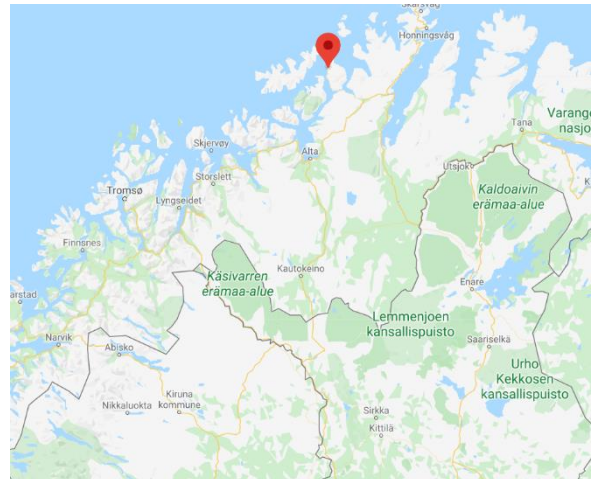
Det er benyttet datasett fra tre ulike værstasjoner i denne oppgaven, plassert som vist i figur 12. Datasettene er valgt slik at det ikke er gjort en bøttetømming i tidsrommet som benyttes. Dette er gjort for å forenkle forarbeidet før selve analysen.



Figur 12: Kartet viser plasseringen av de tre værstasjonene Haukelisetter (helt sør), Bjorli og Hammerfest (helt nord). Det er nedbørdata fra disse tre stasjonene som ble analysert i denne studien.

### 3.2.1 Hammerfest

Værstasjonen Hammerfest Lufthavn, stasjonsnummer 94280, er montert ved sørvestenden av rullebanen på Hammerfest Flyplass på Kvaløya, Troms og Finnmark. Denne værstasjonen har koordinatene  $70^{\circ} 40,84' N$  og  $23^{\circ} 40,63' \text{Ø}$ . Plasseringen er vist i figur 13. Stasjonen er utstyrt med en nedbørmåler av typen Geonor T-200B med en streng og oppsamlingskapasitet på 600 mm. I tillegg er det en optisk nedbørdetektor fra Thies Clima. Nedbørmåleren står i lett hellende terreng.



Figur 13: Figuren viser posisjonen til værstasjonen Hammerfest Lufthavn.

Den operasjonelle datainnsamlingen skjer via en PC-løsning med algoritmeversjonen *AlgFLOS*.

Frekvensen  $R_{\text{f}}$  og strømsignalet  $mA_{005}$  logges

med en tidsoppløsning på fem sekunder, og det er disse som er inputvariabler i algoritmen.

Rådataene ble i denne oppgaven filtrert med algoritmeversjonen *AlgPYTN*. Resultatene ble sammenlignet med de operasjonell filtrerte verdiene fra stasjonen. Målet var å sikre at algoritmene behandler nedbørsignalene tilnærmet likt uavhengig av programmeringsspråk. I tillegg gir analysen et estimat på størrelsesorden i eventuelle avvik mellom algoritmen programmert i ulike programmeringsspråk. Dataene fra perioden 5. januar til 20. mars 2020 ble analysert fordelt over de fire periodene 5.- 17. januar 2020, 17.- 29. januar 2020, 1.- 29. februar 2020 og 1.- 20. mars 2020.

Referanseverdien bestemmes ved å beregne gjennomsnittene og standardavvikene av henholdsvis de første 60 og siste 60 frekvensmålingene i perioden. Deretter gjøres disse gjennomsnittene om til millimeter og differansen mellom dem benyttes som referanseverdi. Standardavvikene brukes til å beregne usikkerheten i referanseverdien. Avvikene mellom bømteverdiendring, beregnet nedbørmengde og referanseverdien i perioden brukes for å evaluere algoritmeversjonene. Det er ønskelig at den beregnede nedbørmengden, dvs. summen av  $RR$ -verdiene i perioden, blir lik referanseverdien.

Resultatene fra analysen gjort av dataene fra Hammerfest Lufthavn kan leses i kapittel 4.1.

### 3.2.2 Bjorli

Værstasjonen på Bjorli, stasjonsnummer 62630, ligger på 579 m.o.h. i Lesja kommune. Denne værstasjonen har koordinatene  $62^{\circ} 15,5' N$  og  $8^{\circ} 12,0' \text{Ø}$ , som vist i figur 14. Stasjonen er utstyrt med en nedbørmåler av typen Geonor T-200B med sensor på en streng og en oppsamlingskapasitet på 600 mm. I tillegg benyttes det en optisk nedbørdetektor fra Thies Clima.

Den operasjonelle datainnsamlingen skjer på en logger fra Campbell av typen CR1000.

Frekvensen  $R_{\text{f}}$  og strømsignalet  $mA_{005}$  logges med en tidsoppløsning på fem

sekunder, og det er disse to som er inputvariabler for algoritmen. De filtrerte verdiene ble sammenlignet med operasjonelle filtrerte verdier fra stasjonen. På denne stasjonen er de operasjonelle verdiene filtrert med den gamle algoritmen, *AlgGML*, hvilket gjør det mulig å evaluere om den nye algoritmen gir en forbedring i beregnet nedbørmengde. Referanseverdien ble beregnet som for hver av periodene i dataene fra Hammerfest Lufthavn, med utgangspunkt i gjennomsnittene av de første 60 og siste 60 målingene i perioden. Datasettet benyttet på denne stasjonen er hentet ut i perioden 23. mars til 3. april 2020. Referanseverdien med tilhørende usikkerhet og resultatene fra analysen av dataene fra Bjorli er gitt i kapittel 4.2.



Figur 14: Figuren viser plasseringen av værstasjonen Bjorli.

### 3.2.3 Haukeliseter

Værstasjonen på Haukeliseter testfelt, stasjonsnummer 33950, er knyttet til Meteorologisk Instituttets testfelt for nedbørmålinger som ble opprettet i samarbeid med forskjellige aktører i kraftbransjen (Wolff, Isaksen, Brækkan, et al., 2013). Feltet ble også benyttet for å skaffe data for WMO SPICE.

Haukeliseter testfelt er utstyrt med referanse for fast nedbør, dvs. en 'double fence automatic reference' (DFAR) (Nitu et al., 2018).

Værstasjonen ligger på 990 m.o.h. med koordinatene  $59^{\circ}48'36.0''N$  og  $7^{\circ}12'36.0'' \text{Ø}$ , som vist i figur 15. Denne stasjonen ble også benyttet under WMO sitt SPICE-prosjekt. Testfeltet på

Haukeliseter er omkranset av lave fjelltopper og to innsjøer. Selve værstasjonen ligger i lett kupert terreng. Stasjonen benytter nedbørmåleren Geonor T-200B i DFAR med tre strenger og en oppsamlingskapasitet på 1000 ml. I tillegg benyttes det en optisk sensor fra Thies Clima.



Figur 15: Figuren viser plasseringen av Haukeliseter Testfelt.

Datainnsamlingen skjer på en logger av typen SM4049 fra Scanmatic. Bøtteverdien logges med en tidsoppløsning på seks sekunder, mens signalet fra den optiske sensoren har en oppløsning på ti

minutter, dvs.  $RT_{010}$ . I motsetning til på Bjorli og Hammerfest Lufthavn, gjøres frekvensmålingene om til millimeter før algoritmens start. Datasettet som er benyttet er hentet fra perioden 17.- 30. april 2019. Dette datasettet ble brukt til å generere det syntetiske datasettet som er grunnlaget for de videre støyanalysene. En nærmere forklaring av det syntetiske datasettet er gitt i kapittel 3.3.

### 3.3 Syntetisk datasett

Det syntetiske datasettet er basert på gjennomsnittet av målingene fra de tre strengene fra referansemåleren (Geonor T-200B i DFAR) på Haukelisetter. Dataene er hentet fra perioden 17. – 30. april 2019. Som beskrevet i kapittel 3.2.3 logges bølgeverdien med en oppløsning på seks sekunder til forskjell fra Hammerfest og Bjorli som har en oppløsning på fem sekunder. Dette tas det hensyn til i algoritmen, men de høyoppløste dataene betegnes fortsatt med  $\Delta RA_{005}$  og  $\Delta Filt_{RA_{005}}$ .

Det syntetiske datasettet ble behandlet av algoritmeversjonen *AlgPYTN*, hvilket gav verdier som i tabell 2. De behandlede datasettene benyttes som referanse i beregning av  $RMSE$  og korrelasjonskoeffisient  $r$  i støyanalysene. Referanseverdien i avviksberegningene for bølgeverdiendringene er gitt i tabell 3, sammen med standardavvikene og usikkerheten i denne referanseverdien. Denne verdien er beregnet ved at det er tatt differansen mellom gjennomsnittet av de første 60 og siste 60 målingene i perioden. Deretter er det beregnet standardavviket for hvert av disse gjennomsnittene, som benyttes til å beregne usikkerheten ved hjelp av formel 2.

Referanseverdien for bølgeverdiendring i støyanalysen er dermed på 21,90 mm med en usikkerhet på  $\pm 0,012$  mm. For beregnet nedbørmengde i støyanalysene er referanseverdien satt til 20,90 mm med en usikkerhet på  $\pm 0,012$  mm, på bakgrunn av  $\Sigma RR_{010}$  som gitt i tabell 2. Dette er diskutert grundigere i kapittel 5.4.1.

Tabell 2: Tabellen inneholder de beregnede verdiene for det syntetiske datasettet.

Variabel	Verdi
$\Delta RA_{010}$	21,88 mm
$\Sigma RR_{010}$	20,90 mm
$\Delta RA_1$	21,86 mm
$\Sigma RR_1$	20,90 mm

Tabell 3: Tabellen inneholder referanseverdien for bølgeverdiendringen, med tilhørende standardavvik og usikkerhet. Usikkerheten er beregnet med formel 2.

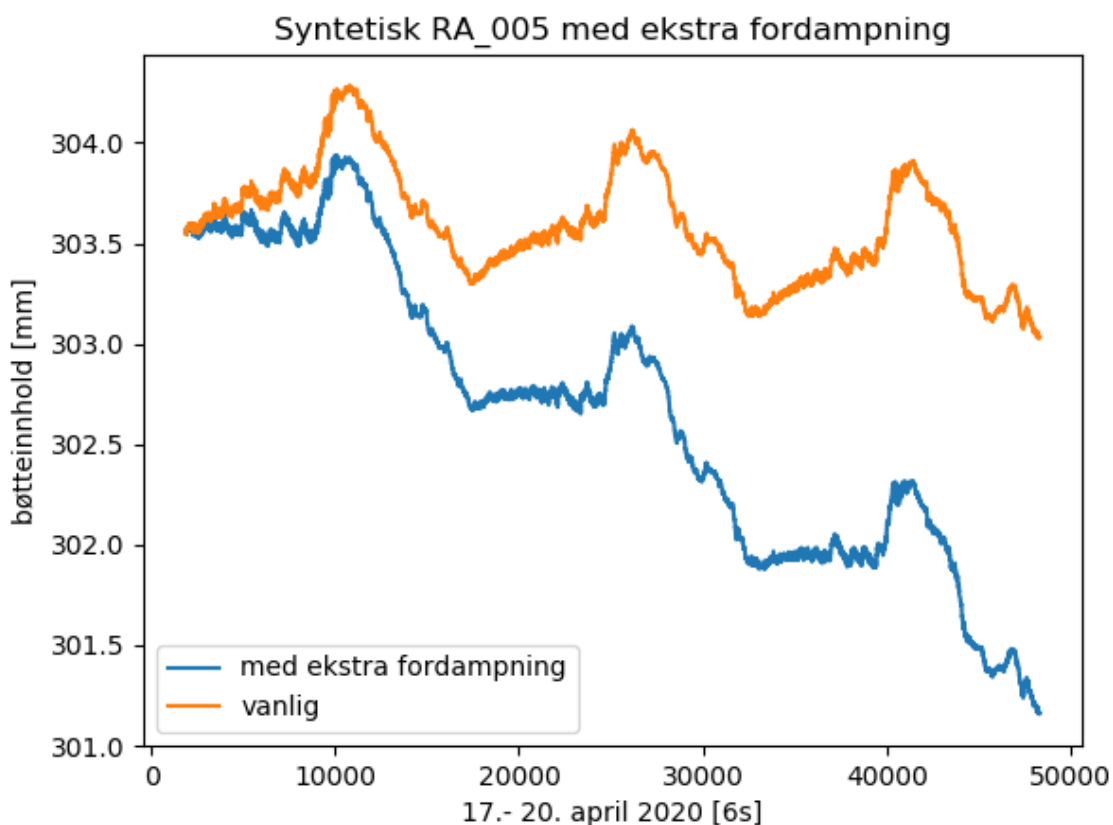
	Verdi [mm]
Referanseverdi bølgeverdiendring	21,90
Std_start	0,0096
Std_slutt	0,0079
Usikkerhet	$\pm 0,012$

### 3.4 Fordampning

Ross et al. (2020) testet de kanadiske algoritmene på et syntetisk datasett med tillagt ekstra fordampning. Fordampningen utgjorde totalt 10% av den totale nedbøren, og ble fordelt utfra VPD ('Vapor Pressure Deficit') i ett minutts intervaller.

For dette datasettet ble det tatt ut de tre første døgnene i det syntetiske datasettet, fra 17.- 19. april 2019. Dette var en periode uten nedbørhendelser. Denne perioden hadde allerede en fordampning på 0,53 mm, men med utgangspunkt i det kanadiske forsøket ble det lagt til ekstra fordampning slik at den totale fordampningen ble på 2,4 mm over disse tre døgnene. Med ønske om at fordampningen skulle skje lineært for å ikke forstyrre variasjonen som allerede var i settet, ble den totale fordampningen fordelt over antallet målinger i settet. Deretter ble det trukket fra en økende andel av totalen. Dette gav et fordampningssett lik det blå signalet i figur 16.

Målet er at kontrollvariabelen `startRA` skal følge bømteverdiendringen gjennom filtreringen, slik at dersom det kommer en nedbørhendelse, er startpunktet for beregningen riktig. Resultatene er gitt i kapittel 4.3.



Figur 16: Figuren viser det opprinnelige syntetiske signalet som oransje graf og signalet tilsatt ekstra fordampning som blå graf. Det er tatt utgangspunkt i en periode uten nedbørhendelser.

### 3.5 Støyanalyse

Det produseres et syntetisk datasett som forklart i kapittel 3.3. Dette brukes som referanse og utgangspunkt for den videre støyanalysen. Det produseres i alt ti støysett som behandles av algoritmen. Kjennetegn for de ulike støysettene er listet opp i tabell 4.

For støytype 1-6 gjennomføres det tre ulike forsøkskonfigurasjoner for filtrering. Dette gjøres for å evaluere hvordan de ulike delene av filtreringen påvirker resultatet. De tre konfigurasjonene som gjennomføres er:

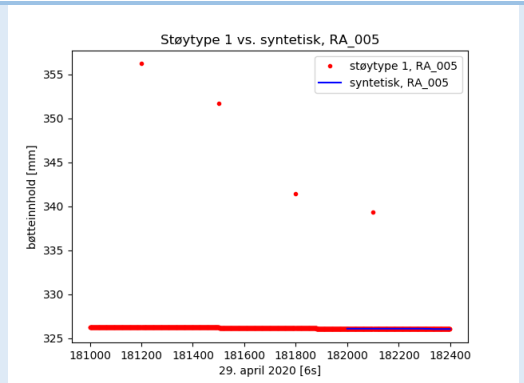
- Forsøk 1; I de løpende treminuttersintervallene i trinn 4 velges alltid medianen som gjeldende filtrert verdi. Ellers kjøres algoritmen som vanlig.
- Forsøk 2; Algoritmen kjøres akkurat som beskrevet i kapittel 3.1.2.
- Forsøk 3; Filtreringen skjer som normalt, men det gjennomføres ikke en sprangtest i trinn 2.

For støytype 7-10 benyttes bare forsøkskonfigurasjon 1 og 2, da resultatene fra de tidligere forsøkene tydet på at det ikke var nødvendig å gjennomføre forsøk 3.

Støytype 7-10 benytter en ekstra variabel som beregnes utenfor selve algoritmen. Denne er summen av differansene mellom utplukkede døgnverdier. RA\_005-verdien registrert kl 06:00:00 ble valgt, og differansene mellom disse over hele perioden ble beregnet og summert til variabelen  $\Delta 24hdata$ . I løpet av perioden var det to døgn der det ikke var registrert en verdi kl 06:00:00. I disse to tilfellene ble verdien registrert kl 05:00:00 benyttet i stedet.

Tabell 4: Tabellen beskriver hvilke støytyper som er benyttet og hvordan disse er produsert. Støytype 1 og 2 er etterligninger av støytyper observert på stasjoner. Støytype 3-6 er laget med inspirasjon fra det kanadiske forsøket beskrevet i Ross et al. (2020). Disse finnes det derfor sammenligningsgrunnlag for. Støytype 7-10 er laget for å finne grensen for algoritmens funksjonsområde. Alle støytypene bruker det syntetiske datasettet som utgangspunkt og senere som referanse når evalueringsverdiene skal beregnes.

Støynavn	Støytype
<b>Støytype 1</b>	Det er lagt til ren positiv støy på opptil 10% av målingens verdi på hver 300. måling, hvilket tilsvarer hver halvtime.



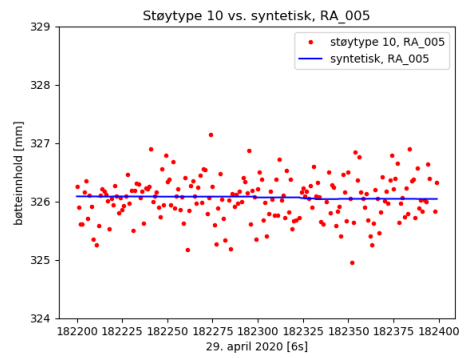
<p><b>Støytype 2</b></p>	<p>Det er lagt til ren negativ støy med en absoluttverdi på opptil 10% av målingens verdi på hver 300. måling, hvilket tilsvarer hver halvtime.</p>	
<p><b>Støytype 3</b></p>	<p>Lagt til gaussisk fordelte verdier ved hjelp av en innebygd generator i Python. Forventningsverdien, my, satt til 0 og standardavvik satt til 0,1 mm. Gir både positive og negative verdier som er lagt til på det syntetiske settet.</p>	
<p><b>Støytype 4</b></p>	<p>Lagt til gaussisk fordelte verdier ved hjelp av en innebygd generator i Python. Forventningsverdien, my, satt til 0 og standardavvik satt til 0,01 mm. Gir både positive og negative verdier som er lagt til på det syntetiske settet.</p>	
<p><b>Støytype 5</b></p>	<p>Lagt til gaussisk fordelte verdier ved hjelp av en innebygd generator i Python. Forventningsverdien, my, satt til 0 og standardavvik satt til 0,001 mm. Gir både positive og negative verdier som er lagt til på det syntetiske settet.</p>	



<p><b>Støytype 6</b></p>	<p>Lagt til gaussisk fordelte verdier ved hjelp av en innebygd generator i Python. Forventningsverdien, <math>\mu</math>, satt til 0 og standardavvik satt til 0,1 mm. Gjennomsnittet av disse verdiene trekkes fra hver enkelt verdi før de legges til på det syntetiske settet. Dette gjøres for å tvinge snittet nærmere null og unngå innføringen av en bias.</p>	
<p><b>Støytype 7</b></p>	<p>Lagt til gaussisk fordelte verdier ved hjelp av en innebygd generator i Python. Forventningsverdien, <math>\mu</math>, satt til 0 og standardavvik satt til 1,00 mm. Gir både positive og negative verdier som er lagt til på det syntetiske settet.</p>	
<p><b>Støytype 8</b></p>	<p>Lagt til gaussisk fordelte verdier ved hjelp av en innebygd generator i Python. Forventningsverdien, <math>\mu</math>, satt til 0 og standardavvik satt til 0,75 mm. Gir både positive og negative verdier som er lagt til på det syntetiske settet.</p>	
<p><b>Støytype 9</b></p>	<p>Lagt til gaussisk fordelte verdier ved hjelp av en innebygd generator i Python. Forventningsverdien, <math>\mu</math>, satt til 0 og standardavvik satt til 0,5 mm. Gir både positive og negative verdier som er lagt til på det syntetiske settet.</p>	

## Støytype 10

Lagt til gaussisk fordelte verdier ved hjelp av en innebygd generator i Python. Forventningsverdien,  $\mu$ , satt til 0 og standardavvik satt til 0,4 mm. Gir både positive og negative verdier som er lagt til på det syntetiske settet.



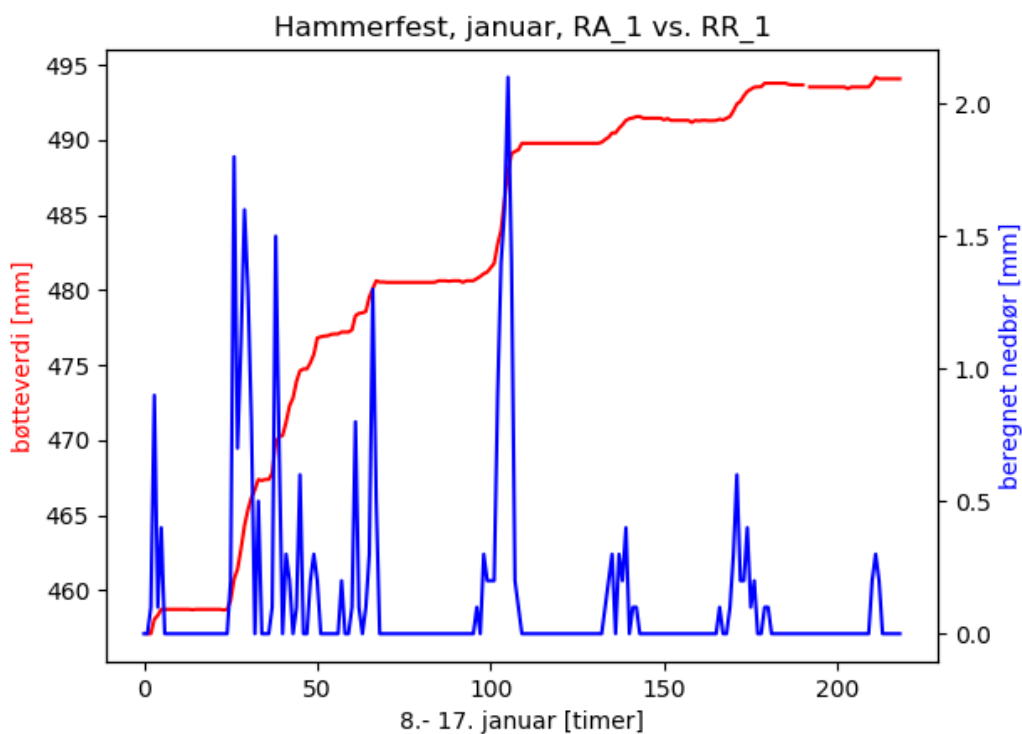
## 4 Resultat

I dette kapittelet er nøkkelresultatene fra de ulike forsøkene fremstilt. Første delkapittel inneholder analysen fra Hammerfest Lufthavn med de fire delperiodene. Deretter følger et delkapittel med analysen gjort av dataene fra Bjorli, et delkapittel med fordampning, et delkapittel med analysen av støytype 1-6 og til sist følger analysen av støytype 7-10. Figurer utover de som er presentert her kan oversendes på etterspørsel.

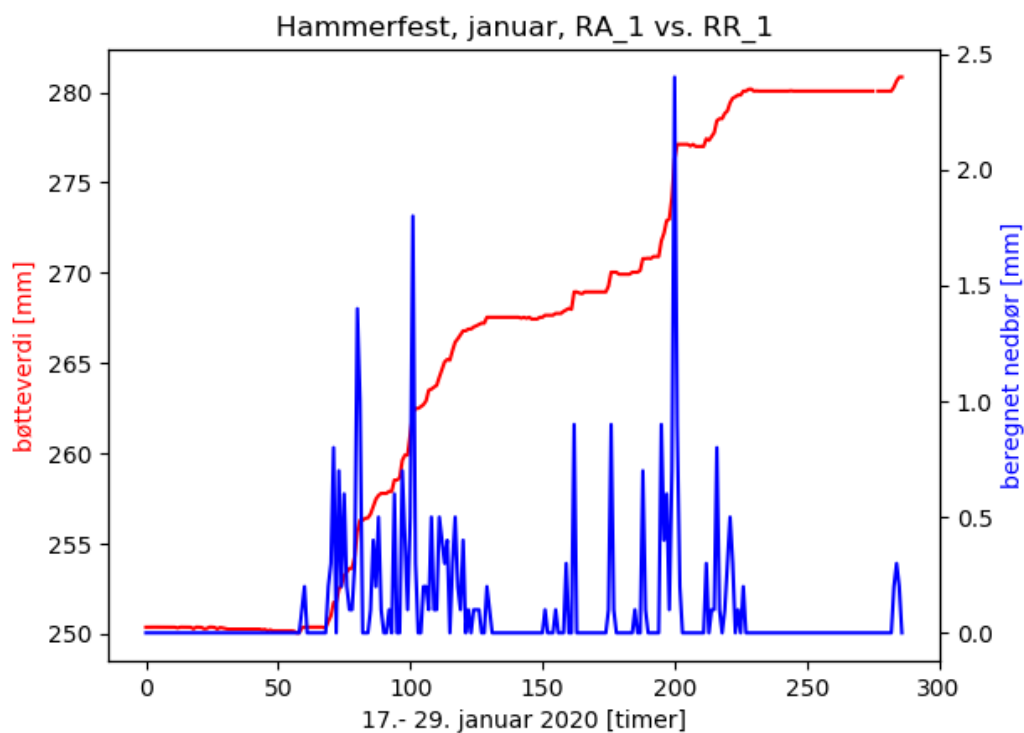
### 4.1 Hammerfest

Målet med analysene gjort på datasettene fra Hammerfest Lufthavn er å validere algoritmeversjonen *AlgPYTN* opp mot *AlgFLOS*, samt å se om det er et avvik som følge av ulike programmeringsspråk.

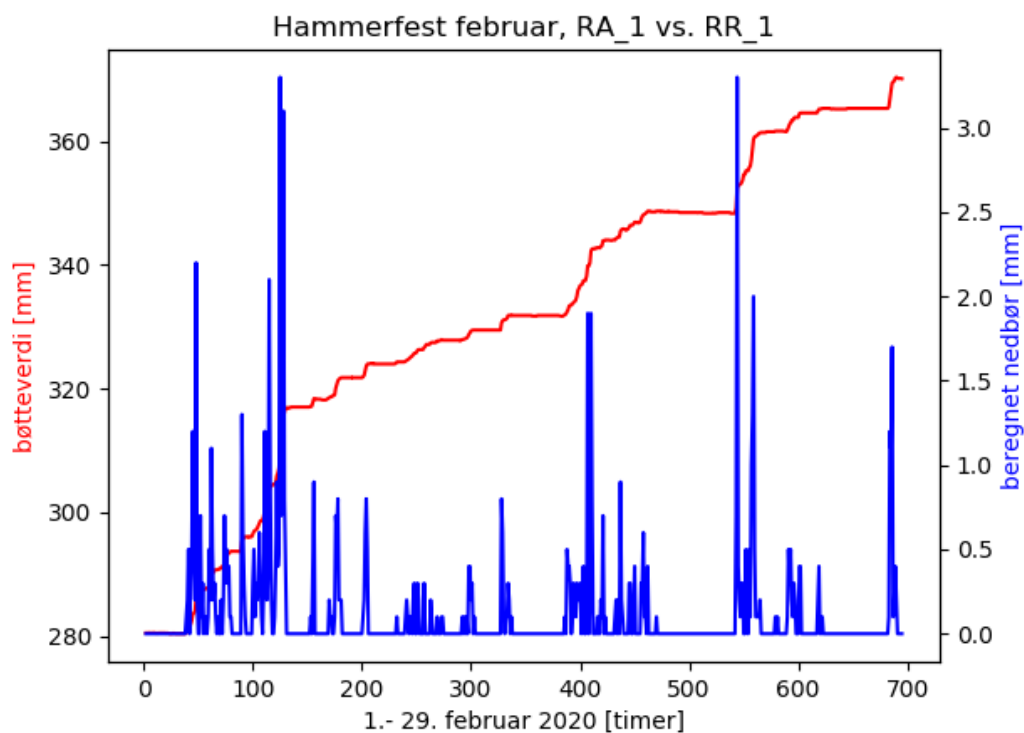
Dataene fra Hammerfest Lufthavn (stasjonsnummer 94280) er delt inn i fire ulike perioder, navngitt periode 1-4. Frekvensen  $R_f$  og strømsignalet  $mA_{005}$  har en tidsoppløsning på fem sekunder. Bøtteverdisignalet for de fire periodene er illustrert i figur 17 til 20. Her ser vi at periode 1 har åtte nedbørhendelser, periode 2 har seks nedbørhendelser der den første varer i nesten to døgn. Periode 3 har fem større nedbørhendelser med ytterligere noen mindre hendelser. Periode 4 har ni nedbørhendelser. En periode med høyt stigningstall i rød graf regnes som en nedbørhendelse.



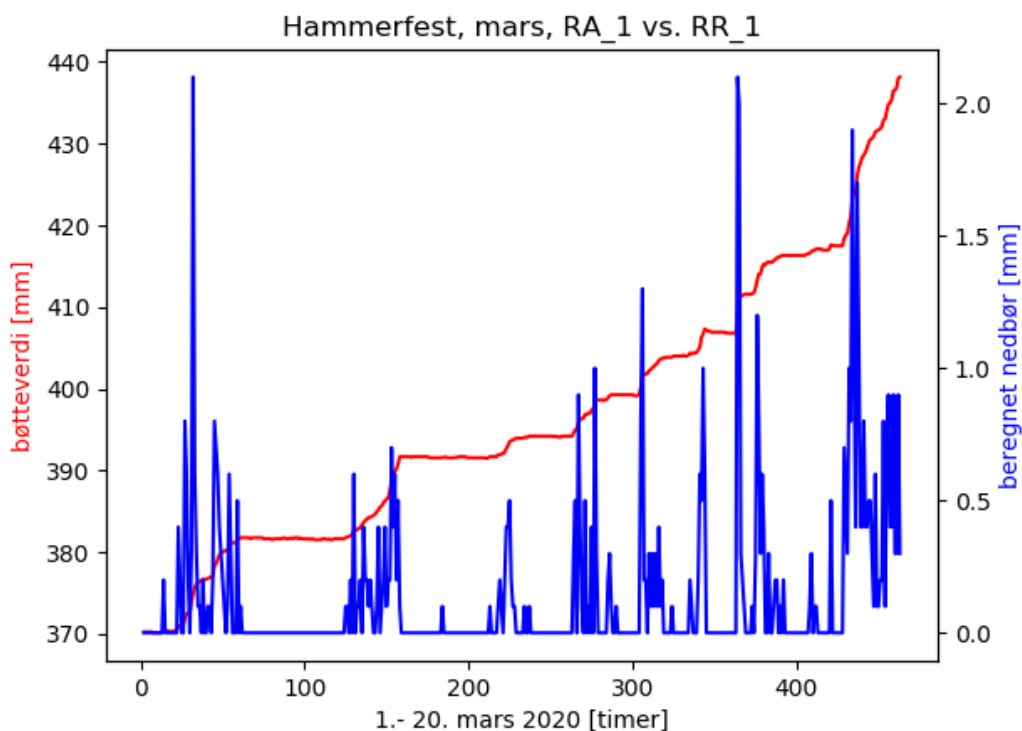
Figur 17: Figuren viser bøtteverdisignalet  $RA_1$  som rød graf mot venstre y-akse og beregnet nedbørmengde  $RR_1$  som blå graf mot høyre y-akse. En ser at det falt mye nedbør i starten av perioden, og at det var noen mindre nedbørhendelser mot slutten av perioden.



Figur 18: Figuren viser bøtteverdisignalet RA\_1 som rød graf mot venstre y-akse og beregnet nedbørmengde RR\_1 som blå graf mot høyre y-akse. En ser at så godt som all nedbøren falt midt i perioden, med en lang nedbørhendelse i starten avløst av mindre hendelser med korte oppholdsperioder.



Figur 19: Figuren viser bøtteverdisignalet RA\_1 som rød graf mot venstre y-akse og beregnet nedbørmengde RR\_1 som blå graf mot høyre y-akse. I februar falt nedbøren jevnt over hele måneden på værstasjonen Hammerfest Lufthavn.



Figur 20: Figuren viser bøtteverdisignalet RA\_1 som rød graf mot venstre y-akse og beregnet nedbørmengde RR\_1 som blå graf mot høyre y-akse. En ser at det falt mye nedbør helt i starten av perioden, men at mesteparten av nedbøren kom i siste halvdel av mars.

For å bestemme referanseverdien for hver periode er det beregnet gjennomsnitt av de 60 første frekvensmålingene og de 60 siste frekvensmålingene for hver periode. Disse gjennomsnittsverdiene er omgjort til millimeter, og differansen mellom dem benyttes som referanseverdi for perioden. Standardavviket for de 60 verdiene beregnes, og disse standardavvikene benyttes til å beregne usikkerheten i referanseverdien. Disse verdiene er gitt i tabell 5.

Tabell 5: Tabellen inneholder en oversikt over de fire periodene med referanseverdier, standardavvik og referanseverdiens usikkerhet for hver av periodene fra Hammerfest Lufthavn.

	Dato	Referanseverdi [mm]	Standardavvik start [mm]	Standardavvik slutt [mm]	Usikkerhet [mm]
<b>Periode 1</b>	5.-17. januar 2020	36,99	0,15	0,15	± 0,21
<b>Periode 2</b>	17.-29. januar 2020	30,44	0,080	0,41	± 0,42
<b>Periode 3</b>	1.-29. februar 2020	89,59	0,34	0,13	± 0,36
<b>Periode 4</b>	1.-20. mars 2020	68,09	0,18	0,15	± 0,23

For å validere AlgPYTN er resultatene fra denne oppgaven sammenlignet med de offisielle dataene fra værstasjonen Hammerfest Lufthavn. Verdiene beregnet med AlgFLOS på Hammerfest Lufthavn

har en tidsoppløsning på ti minutter.  $\Delta RA_{010}$  og  $\Sigma RR_{010}$  beregnet med *AlgFLOS* for hver periode er gitt i tabell 6. De absolutte og relative avvikene for disse verdiene er beregnet opp mot referanseverdiene i tabell 5.

Tabell 6: Tabellen viser en oversikt over verdiene beregnet med *AlgFLOS*, i tillegg til de absolutte og relative avvikene for hver av de fire periodene.

	$\Delta RA_{010}$ [mm]	Absolutt avvik $\Delta RA_{010}$ [mm]	Relativt avvik $\Delta RA_{010}$ [%]	$\Sigma RR_{010}$ [mm]	Absolutt avvik $\Sigma RR_{010}$ [mm]	Relativt avvik $\Sigma RR_{010}$ [%]
<b>Periode 1</b>	37,30	+0,31	+ 0,84	32,82	-4,17	- 11,27
<b>Periode 2</b>	30,75	+0,31	+ 1,02	30,20	-0,24	- 0,79
<b>Periode 3</b>	90,43	+0,84	+ 0,94	84,85	-4,74	- 5,29
<b>Periode 4</b>	68,76	+0,67	+ 0,97	64,90	-3,19	- 4,77

Av tabellen fremkommer det at *AlgFLOS* overestimerer bøtteverdien  $\Delta RA_{010}$  for de fire periodene, men underestimerer den beregnede nedbørmengden  $\Sigma RR_{010}$ . De relative avvikene for  $\Delta RA_{010}$  er svært like, alle rundt 1%. Likevel er avvikene i  $\Delta RA_{010}$  for periode 1, 3 og 4 signifikante. I de relative avvikene for  $\Sigma RR_{010}$  er det større variasjon; - 11,27 % for periode 1 mot - 0,79 % for periode 2. Det er bare periode 2 som har et avvik i  $\Sigma RR_{010}$  som ikke er signifikant.

De fire datasettene ble behandlet med *AlgPYTN* som beskrevet i kapittel 3.1.2. Nøkkelforholdene fra disse beregningene er fremstilt i tabell 7. Tabell 8 inneholder de absolutte og prosentvise avvikene for hver av periodene, beregnet opp mot referanseverdiene gitt i tabell 5.

Tabell 7: Tabellen inneholder en oversikt over de beregnede verdiene  $\Delta RA_{010}$ ,  $\Sigma RR_{010}$ ,  $\Delta RA_1$  og  $\Sigma RR_1$  for de fire periodene. Verdiene er beregnet med *AlgPYTN*.

Variabel	Periode 1	Periode 2	Periode 3	Periode 4
$\Delta RA_{010}$ [mm]	36,99	30,47	89,55	67,93
$\Sigma RR_{010}$ [mm]	30,60	30,20	85,00	64,80
$\Delta RA_1$ [mm]	36,99	30,47	89,55	67,93
$\Sigma RR_1$ [mm]	30,60	30,20	85,00	64,80

Tabell 8: Tabellen inneholder de absolutte og relative avvikene for hver av de fire periodene for beregningene gjort med *AlgPYTN*.

Variabel	Avvik $\Delta RA_{010}$ [mm (%)]	Avvik $\Sigma RR_{010}$ [mm (%)]	Avvik $\Delta RA_1$ [mm (%)]	Avvik $\Sigma RR_1$ [mm (%)]
<b>Periode 1</b>	0 (0)	-6,39 (-17,27)	0 (0)	-6,39 (-17,27)
<b>Periode 2</b>	+0,03 (+0,099)	-0,24 (-0,79)	+0,03 (+0,099)	-0,24 (-0,79)
<b>Periode 3</b>	-0,04 (-0,045)	-4,59 (-5,12)	-0,04 (-0,045)	-4,59 (-5,12)
<b>Periode 4</b>	-0,16 (-0,23)	-3,29 (-4,83)	-0,16 (-0,23)	-3,29 (-4,83)

Felles for alle de fire periodene er at bøtteverdiene beregnes til det samme uavhengig av tidsoppløsning, og det samme gjelder for beregnet nedbørmengde RR. Dermed er også avvikene de

samme, uavhengig av tidsoppløsning. Avvikene for bølgeverdiene er lavere enn avvikene for beregnet nedbør. For enkelhetsskyld henvises det bare til verdiene med en tidsoppløsning på ti minutter.

For periode 1 ser vi at *AlgPYTN* estimerer  $\Delta_{RA\_010}$  til referanseverdien, men underestimerer  $\Sigma_{RR\_010}$  med 6,39 mm tilsvarende 17,27 %. *AlgFLOS* overestimerte bølgeverdiendringen med 0,31 mm (0,84 %), og underestimerer den beregnede nedbørmengden med 4,17 mm (11,27 %). *AlgPYTN* treffer bedre for bølgeverdiendringen, men får et større avvik i beregnet nedbørmengde. Usikkerheten i referanseverdien for periode 1 er  $\pm 0,21$  mm, gitt i tabell 5. Dette betyr at avviket i  $\Delta_{RA\_010}$  ikke er signifikant for beregningen med *AlgPYTN*, men er signifikant for beregningen med *AlgFLOS*. For  $\Sigma_{RR\_010}$  er begge avvikene signifikante.

For periode 2 overestimerer *AlgPYTN*  $\Delta_{RA\_010}$  minimalt, med et absolutt avvik på 0,03 mm (0,099 %). Avviket i  $\Sigma_{RR\_010}$  er det laveste av avvikene for de fire periodene, med en verdi på -0,24 mm eller -0,79 %. Avviket i  $\Sigma_{RR\_010}$  er likt for beregningen med *AlgFLOS*, mens det absolutte avviket i  $\Delta_{RA\_010}$  er noe større, +0,31 mm tilsvarende +1,02 %. Periode 2 har en usikkerhet i referanseverdien lik  $\pm 0,42$  mm. Det betyr at ingen av avvikene for periode 2 er signifikante.

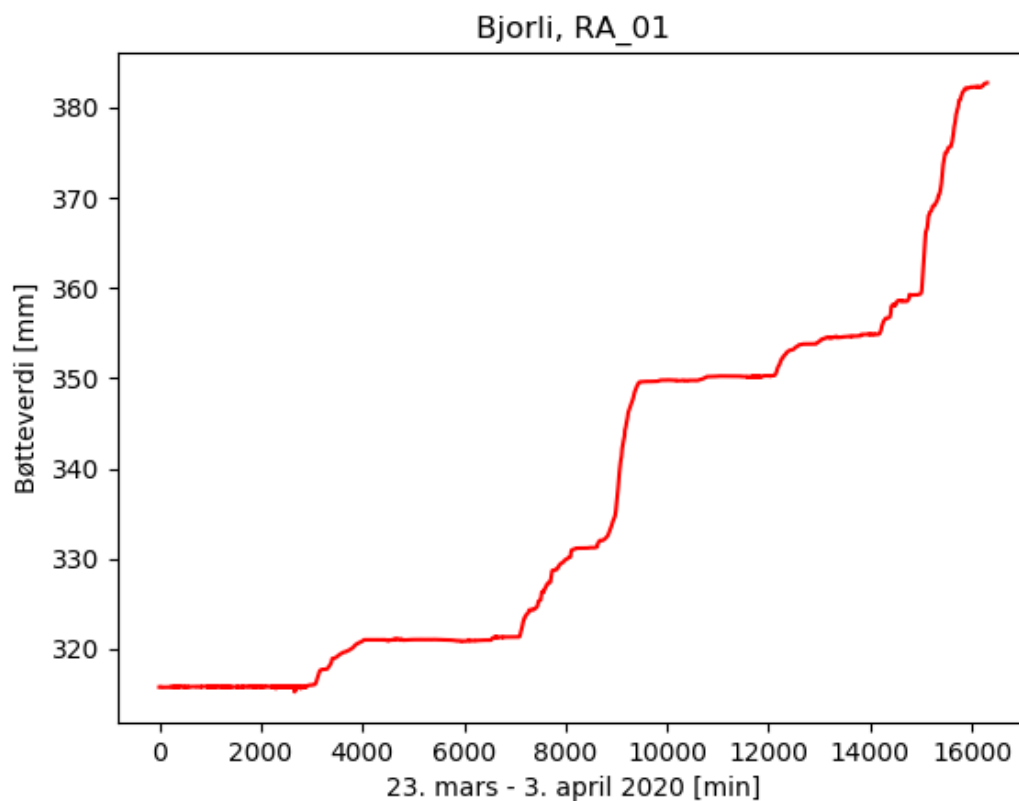
For periode 3 ser vi at *AlgPYTN* underestimerer  $\Delta_{RA\_010}$  med 0,04 mm (0,045 %) og  $\Sigma_{RR\_010}$  med 4,59 mm (5,12 %). *AlgFLOS* overestimerer bølgeverdiendringen med 0,84 mm (0,94 %) og underestimerer beregnet nedbørmengde med 4,74 mm (tilsvarende 5,29 %). Referanseverdien har en usikkerhet på  $\pm 0,36$  mm, hvilket gjør at alle avvikene, med unntak av avviket for  $\Delta_{RA\_010}$  for *AlgPYTN*, er signifikante.

For periode 4 ser vi at *AlgPYTN* underestimerer  $\Delta_{RA\_010}$  med 0,16 mm (tilsvarende 0,23 %) og i  $\Sigma_{RR\_010}$  med 3,29 mm, hvilket tilsvarer 4,83 %. *AlgFLOS* overestimerer  $\Delta_{RA\_010}$  med 0,67 mm (0,97 %), men underestimerer  $\Sigma_{RR\_010}$  med 3,19 mm (4,77 %). Usikkerheten for periode 4 er  $\pm 0,23$  mm, hvilket betyr avviket i  $\Sigma_{RR\_010}$  beregnet med *AlgPYTN* og begge avvikene beregnet med *AlgFLOS* er signifikante.

## 4.2 Bjorli

Målsetningen for analysen av datasettet fra Bjorli værstasjon er å sammenligne de to algoritmeversjonene *AlgGML* og *AlgPYTN* sine evner til å filtrere støy.

Datasettet fra Bjorli er hentet fra perioden 23. mars 2020 til 3. april 2020. Råfrekvensen  $R_f$  og strømsignalet  $mA_{005}$  har en tidsoppløsning på fem sekunder. Figur 21 viser bølgeverdien per minutt,  $RA_{01}$ , for hele tidsperioden fra Bjorli. Av figuren ser man at det er totalt seks mer langvarige nedbørhendelser hvilket gir en total økning i bølgeverdi på ca. 70 mm.



Figur 21: Figuren viser bøtteverdien RA\_01 for hele perioden på værstasjonen på Bjorli.

Referanseverdien er differansen mellom gjennomsnittet av de henholdsvis første 60 og siste 60 målingene. Standardavviket beregnes for hver av de 60 målingene, og usikkerheten i referanseverdien beregnes på bakgrunn av dette. Referanseverdien med tilhørende standardavvik og usikkerhet kan leses i tabell 9.

Tabell 9: Tabellen inneholder referanseverdien for analysen gjort på Bjorli, samt standardavvikene og usikkerheten for referanseverdien.

	Dato	Referanseverdi [mm]	Standardavvik start [mm]	Standardavvik slutt [mm]	Usikkerhet [mm]
<b>Bjorli</b>	23. mars – 3. april 2020	67,00	0,053	0,054	± 0,076

Datasettet ble behandlet av *AlgPYTN* som beskrevet i kapittel 3.1.2. Dette gav verdier som gitt i tabell 10. Tabellen inneholder også de absolutte og relative avvikene for hver av variablene.



Tabell 10: Tabellen inneholder  $\Delta RA_{010}$ ,  $\Sigma RR_{010}$ ,  $\Delta RA_1$  og  $\Sigma RR_1$  med tilhørende absolutt og relativt avvik. Avvikene er beregnet opp mot referanseverdien gitt i tabell 9.

Variabel	Verdi [mm]	Absolutt avvik [mm]	Relativt avvik [%]
$\Delta RA_{010}$	67,00	0	0
$\Sigma RR_{010}$	66,50	-0,50	-0,75
$\Delta RA_1$	66,94	-0,06	-0,090
$\Sigma RR_1$	66,50	-0,50	-0,75

Av verdiene i tabell ser man at *AlgPYTN* estimerer  $\Delta RA_{010}$  til referanseverdien, mens  $\Delta RA_1$  underestimeres med 0,06 mm (0,090 %). Sammenligner man med usikkerheten listet opp i tabell 9, ser man at dette avviket ikke er signifikant. Avvikene i beregningen av nedbørmengde er lik uavhengig av tidsoppløsning på ti minutter eller en time. Disse avvikene er signifikante, med en verdi på - 0,50 mm tilsvarende 0,75 % av referanseverdien.

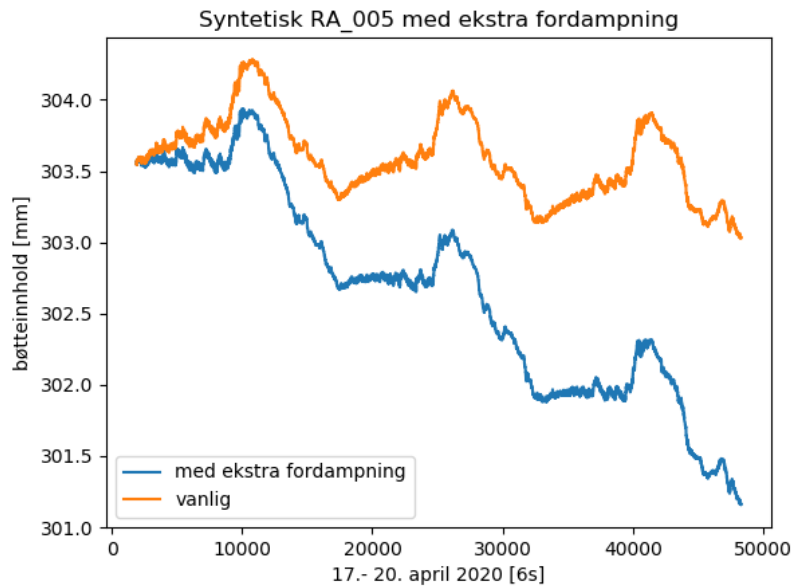
De offisielle verdiene er listet opp i tabell 11. Disse er behandlet med *AlgGML*. De absolutte og relative avvikene er listet opp i samme tabell. Sammenligner man disse med verdiene fra *AlgPYTN* ser man at bølgeverdiendringen  $\Delta RA_{010}$  har et større avvik med *AlgGML* enn med *AlgPYTN*. Usikkerheten i referanseverdien er lav ( $\pm 0,076$  mm) hvilket betyr at dette avviket er signifikant. *AlgGML* estimerer beregnet nedbørmengde til referanseverdien, mens *AlgPYTN* underestimerer med 0,50 mm. Ser man bare på akkurat denne perioden fra Bjorli værstasjon, er estimatene på timinutters tidsoppløsning for *AlgGML* totalt sett bedre enn de beregnet med *AlgPYTN*.

Tabell 11: Tabellen inneholder verdiene beregnet av *AlgGML* for perioden på Bjorli.

Variabel	Verdi [mm]	Absolutt avvik [mm]	Relativt avvik [%]
$\Delta RA_{010}$	67,17	+0,17	+0,25
$\Sigma RR_{010}$	67,00	0	0
$\Delta RA_1$	67,68	+0,68	+1,01
$\Sigma RR_1$	67,50	+0,50	+0,74

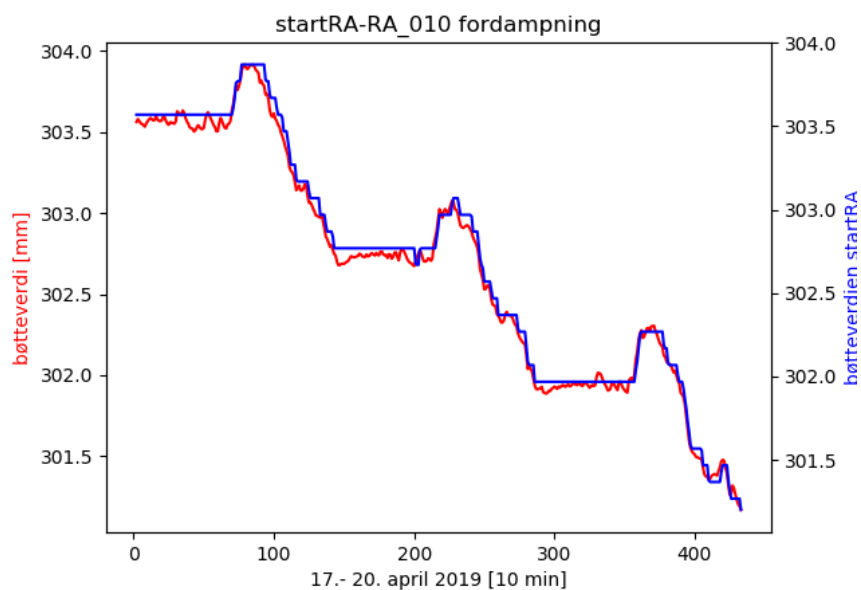
### 4.3 Fordampning

Som forklart i kapittel 3.4 ble det lagt til ekstra fordampning på de første 72 timene av det syntetiske datasettet. Det var ingen nedbør i løpet av disse 72 timene, slik at de resterende svingningene skyldes temperatursvingninger gjennom døgnet slik forklart i kapittel 2.5.1. Målet med dette var å se om algoritmen klarte å ta hensyn til fordampning i dataene. Figur 22 viser det opprinnelige signalet som oransje og signalet med ekstra fordampning som blå. Den ekstra fordampningen som er lagt til utgjør ca. 2 mm i løpet av perioden.



Figur 22: Figuren viser det originale signalet for de første 72 timene av det syntetiske datasettet som oransje signal, og signalet tilsatt ekstra fordampning som blått signal.

Variabelen `startRA` inneholder til enhver tid den siste korrekte bøtteverdien, og denne er utgangspunktet for beregningen av nedbør. Dette forsøket gikk dermed ut på å se om denne variabelen ble redusert når det var fordampning i datasettet, for å dermed ha riktig utgangspunkt i de kommende nedbørsituasjonene. Figur 23 viser `RA_010` for signalet med ekstra fordampning som rød graf, og variabelen `startRA` med oppløsning på ti minutter som blå graf. Av figuren ser man at den røde og den blå grafen følger hverandre godt, og ender i samme verdi. Dette antyder at algoritmen tar godt hensyn til fordampning.



Figur 23: Figuren viser `RA_010` for signalet med ekstra fordampning som rød graf mot venstre y-akse, og `startRA` med tidsoppløsning på ti minutter som blå graf mot høyre y-akse. Tidsoppløsningen er ti minutter og signalene går over tre døgn.

## 4.4 Støyreduksjon

For å teste algoritmens evne til å filtrere ulike støytyper ble det laget totalt ti ulike støysett. Disse er basert på det syntetiske datasettet beskrevet i kapittel 3.3. Hvordan de ulike støytypene er produsert, i tillegg til en illustrasjon av støysignalet før filtrering, er vist i tabell 4 i kapittel 3.5. Dette kapittelet er delt opp i to underkapitler, der underkapittel 4.4.1 inneholder resultatene for støytype 1-6 og underkapittel 4.4.2 inneholder resultatene for støytype 7-10.

Referanseverdien for bølgeverdiendringen for de ti støysettene er beregnet som beskrevet i kapittel 3.3. Denne har en verdi på 21,90 mm med en usikkerhet på  $\pm 0,12$  mm. I tabell 2 ser man at *AlgPYTN* beregner nedbørmengden,  $RR$ , til 20,90 mm, uavhengig av tidsoppløsning. Det betyr at det for støyanalysene vil bli benyttet to ulike referanseverdier – en for bølgeverdiendringene og en for beregnet nedbørmengde. Det antas at usikkerheten er lik for begge verdiene.

### 4.4.1 Støytype 1-6

Støyanalysen ble gjennomført som beskrevet i kapittel 3.5. For støytype 1-6 ble det gjennomført tre ulike forsøk for støyreduksjon. De tre ulike forsøkene er beskrevet i kapittel 3.5, og de ulike støytypene er beskrevet i tabell 4.

Tabell 12 inneholder maksimums- og minimums- verdiene for de ulike støysettene i tillegg til maksimum og minimum for det syntetiske datasettet.

Tabell 12: Tabellen inneholder de ulike maksimums- og minimumsverdiene for de ulike støytypene. Hvordan hver støytype er laget kan leses i tabell 4.

Støytype	Syntetisk	1	2	3	4	5	6
Maksimum [mm]	326,60	358,46	326,58	326,84	326,60	326,58	326,86
Minimum [mm]	302,64	302,64	272,80	302,36	302,61	302,64	302,24

Resultatene for denne delen av støyanalysen er delt inn etter de tre forsøkene for å enklere kunne sammenligne filtreringsevnen for de ulike støytypene som er testet her. Resultatene for de tre forsøkene er ganske like, men det er små endringer i beregnet nedbørmengde for et par av støytypene i tillegg til at  $RMSE$  for støytype 6 endres.

#### Forsøk 1

I forsøk 1 av støyfiltreringen ble algoritmen kjørt som beskrevet i kapittel 3.1.2, men i trinn 4 i ble medianen i de løpende treminuttersintervallene valgt i stedet for at kvantilverdien ble valgt på bakgrunn av standardavvik. Endring i bølgeverdi og sum av beregnet nedbørmengde for de seks støytypene er listet opp i tabell 13. Her ser man at den beregnede nedbørmengden er uavhengig av tidsoppløsning for alle støysettene. Bølgeverdiendringen varierer med 0,01- 0,04 mm fra  $\Delta RA_{010}$  til  $\Delta RA_1$  for de seks støytypene. Tabell 14 inneholder de beregnede  $RMSE$ -verdiene, avvikene og korrelasjonskoeffisientene for bølgeverdiendringene  $\Delta RA_{010}$  og  $\Delta RA_1$ . Støytype 6 skiller seg ut da denne har de høyeste avvikene og den laveste korrelasjonskoeffisienten, uavhengig av

tidsoppløsning. Støytype 6 har også en svært høy  $RMSE$ -verdi for  $RA_{-1}$ , mens ingen av de andre støytypene har en  $RMSE$ -verdi høyere enn 0,1 mm. Støytype 1, 2 og 4 har de samme estimerte verdiene og dermed de samme avvikene. Korrelasjonskoeffisienten  $r$  tyder på at bølgeverdiendringen både for timinuttersperiodene og for timene er perfekt positivt korrelert for støytype 1-5. Referanseverdien har en usikkerhet på  $\pm 0,12$  mm, hvilket betyr at ingen av avvikene i bølgeverdiendring for disse seks støytypene er signifikante.

Tabell 15 inneholder de absolutte og relative avvikene for de beregnede nedbørmengdene  $\Sigma RR_{-010}$  og  $\Sigma RR_{-1}$  for de seks støytypene fra forsøk 1. Alle avvikene er positive og de er målt opp mot referanseverdien for beregnet nedbør, lik 20,90mm. Antas det samme usikkerhet i denne referanseverdien som usikkerheten i referanseverdien for bølgeverdiendring (dvs.  $\pm 0,12$  mm) er avvikene for støytype 3, 4 og 6 signifikante.

Tabell 13: Tabellen inneholder de beregnede variablene  $\Delta RA_{-010}$ ,  $\Sigma RR_{-010}$ ,  $\Delta RA_{-1}$  og  $\Sigma RR_{-1}$  for de seks ulike støytypene etter forsøkskonfigurasjon 1.

Støytype	1	2	3	4	5	6
$\Delta RA_{-010}$ [mm]	21,86	21,86	21,86	21,86	21,86	21,82
$\Sigma RR_{-010}$ [mm]	21,00	21,00	21,30	21,10	21,00	21,30
$\Delta RA_{-1}$ [mm]	21,87	21,87	21,89	21,87	21,88	21,84
$\Sigma RR_{-1}$ [mm]	21,00	21,00	21,30	21,10	21,00	21,30

Tabell 14: Tabellen inneholder de beregnede  $RMSE$ -verdiene, absolutte og relative avvik og korrelasjonskoeffisienten  $r$  for bølgeverdiendringene  $\Delta RA_{-010}$  og  $\Delta RA_{-1}$  for de seks støytypene. Verdiene er målt opp mot referanseverdien 21,90 mm og de syntetiske settene med oppløsning på ti minutter og en time.

Støytype	1	2	3	4	5	6
$RMSE_{RA_{-010}}$ [mm]	0,011	0,011	0,022	0,011	0,011	0,077
$RMSE_{RA_{-1}}$ [mm]	0,0086	0,0086	0,021	0,0087	0,0085	23,98
Abs.avvik $_{\Delta RA_{-010}}$ [mm]	-0,040	-0,040	-0,040	-0,040	-0,040	-0,080
Rel.avvik $_{\Delta RA_{-010}}$ [%]	-0,18	-0,18	-0,18	-0,18	-0,18	-0,37
Abs.avvik $_{\Delta RA_{-1}}$ [mm]	-0,030	-0,030	-0,010	-0,030	-0,020	-0,060
Rel.avvik $_{\Delta RA_{-1}}$ [%]	-0,14	-0,14	-0,046	-0,14	-0,091	-0,27
$r_{RA_{-010}}$ [mm]	1,00	1,00	1,00	1,00	1,00	0,57
$r_{RA_{-1}}$ [mm]	1,00	1,00	1,00	1,00	1,00	0,25

Tabell 15: Tabellen inneholder de absolutte og relative avvikene for de beregnede nedbørmengdene  $\Sigma_{RR\_010}$  og  $\Sigma_{RR\_1}$  for de seks støytypene støyytype 1- 6. De absolutte og relative avvikene er beregnet opp mot referanseverdien for beregnet nedbørmengde lik 20,90 mm.

Støyytype	1	2	3	4	5	6
<b>Abs.avvik</b> $_{\Sigma_{RR\_010}}$ [mm]	+0,10	+0,10	+0,40	+0,20	+0,10	+0,40
<b>Rel.avvik</b> $_{\Sigma_{RR\_010}}$ [%]	+0,48	+0,48	+1,91	+0,96	+0,48	+1,91
<b>Abs.avvik</b> $_{\Sigma_{RR\_1}}$ [mm]	+0,10	+0,10	+0,40	+0,20	+0,10	+0,40
<b>Rel.avvik</b> $_{\Sigma_{RR\_1}}$ [%]	+0,48	+0,48	+1,91	+0,96	+0,48	+1,91

### Forsøk 2

I forsøk 2 av støyyfiltreringen ble algoritmen kjørt som beskrevet i kapittel 3.1. Endring i bøtteverdi og sum av beregnet nedbørmengde for de seks støytypene er listet opp i tabell 16. Her ser man at bøtteverdiendringen varierer med tidsoppløsning, mens den beregnede nedbørmengden er uavhengig av tidsoppløsningen for alle de seks støytypene. Tabell 17 inneholder de beregnede  $RMSE$ -verdiene, avvikene og korrelasjonskoeffisientene  $r$  for bøtteverdiendringene  $\Delta_{RA\_010}$  og  $\Delta_{RA\_1}$ . Støyytype 6 skiller seg ut da denne har de høyeste avvikene og den laveste korrelasjonskoeffisienten, uavhengig av tidsoppløsning. Støyytype 1- 5 har de samme absolutte og relative avvikene, men støyytype 3 har forholdsvis mye større  $RMSE$ -verdi sammenlignet med 1, 2, 4 og 5. Støyytype 6 har den klart høyeste  $RMSE$ -verdien, med  $RMSE_{RA\_1} = 23,98$  mm. Korrelasjonskoeffisienten  $r$  tyder på at bøtteverdiendringen både for timinuttersperiodene og for timene er perfekt positivt korrelert for støyytype 1- 5. Referanseverdien har en usikkerhet på  $\pm 0,12$  mm, hvilket betyr at ingen av avvikene i bøtteverdiendring for disse seks støytypene er signifikante.

Tabell 18 inneholder de absolutte og relative avvikene for de beregnede nedbørmengdene  $\Sigma_{RR\_010}$  og  $\Sigma_{RR\_1}$  for de seks støytypene fra forsøkskonfigurasjon 2. Alle avvikene er positive og de er målt opp mot referanseverdien for beregnet nedbør, lik 20,90 mm. Dersom det antas samme usikkerhet i denne referanseverdien som usikkerheten i referanseverdien for bøtteverdiendring (dvs.  $\pm 0,12$  mm) er avvikene i beregnet nedbørmengde for støyytype 2, 3, 4 og 6 signifikante. Det er støyytype 3 og 6 som har de største avvikene i beregnet nedbørmengde.

Tabell 16: Tabellen inneholder de beregnede variablene  $\Delta_{RA\_010}$ ,  $\Sigma_{RR\_010}$ ,  $\Delta_{RA\_1}$  og  $\Sigma_{RR\_1}$  for de seks ulike støytypene etter forsøkskonfigurasjon 2.

Støyytype	1	2	3	4	5	6
<b><math>\Delta_{RA\_010}</math></b> [mm]	21,86	21,86	21,86	21,86	21,86	21,82
<b><math>\Sigma_{RR\_010}</math></b> [mm]	21,00	21,20	21,30	21,10	21,00	21,30
<b><math>\Delta_{RA\_1}</math></b> [mm]	21,87	21,87	21,89	21,87	21,88	21,84
<b><math>\Sigma_{RR\_1}</math></b> [mm]	21,00	21,20	21,30	21,10	21,00	21,30

Tabell 17: Tabellen inneholder de beregnede *RMSE*-verdiene, absolutte og relative avvik og korrelasjonskoeffisienten  $r$  for bøtteverdiendringene  $\Delta_{RA\_010}$  og  $\Delta_{RA\_1}$  for de seks støttypene. Verdiene er målt opp mot referanseverdien 21,90 mm og de syntetiske settene med oppløsning på ti minutter og en time.

Støytype	1	2	3	4	5	6
<b>RMSE<sub>RA_010</sub></b> [mm]	0,011	0,011	0,022	0,011	0,011	0,077
<b>RMSE<sub>RA_1</sub></b> [mm]	0,0086	0,0086	0,021	0,0087	0,0085	23,98
<b>Abs.avvik<sub>ΔRA_010</sub></b> [mm]	-0,040	-0,040	-0,040	-0,040	-0,040	-0,080
<b>Rel.avvik<sub>ΔRA_010</sub></b> [%]	-0,18	-0,18	-0,18	-0,18	-0,18	-0,37
<b>Abs.avvik<sub>ΔRA_1</sub></b> [mm]	-0,030	-0,030	-0,010	-0,030	-0,020	-0,060
<b>Rel.avvik<sub>ΔRA_1</sub></b> [%]	-0,14	-0,14	-0,048	-0,14	-0,091	-0,27
<b>r<sub>RA_010</sub></b> [mm]	1,00	1,00	1,00	1,00	1,00	0,57
<b>r<sub>RA_1</sub></b> [mm]	1,00	1,00	1,00	1,00	1,00	0,25

Tabell 18: Tabellen inneholder de absolutte og relative avvikene for de beregnede nedbørmengdene  $\Sigma_{RR\_010}$  og  $\Sigma_{RR\_1}$  for de seks støttypene støytype 1- 6. De absolutte og relative avvikene er beregnet opp mot referanseverdien for beregnet nedbørmengde lik 20,90 mm.

Støytype	1	2	3	4	5	6
<b>Abs.avvik<sub>ΣRR_010</sub></b> [mm]	+0,10	+0,30	+0,40	+0,20	+0,10	+0,40
<b>Rel.avvik<sub>ΣRR_010</sub></b> [%]	+0,48	+1,44	+1,91	+0,96	+0,48	+1,91
<b>Abs.avvik<sub>ΣRR_1</sub></b> [mm]	+0,10	+0,30	+0,40	+0,20	+0,10	+0,40
<b>Rel.avvik<sub>ΣRR_1</sub></b> [%]	+0,48	+1,44	+1,91	+0,96	+0,48	+1,91

### Forsøk 3

I forsøk 3 av støyfiltreringen ble algoritmen kjørt som beskrevet i kapittel 3.1, men det ble ikke gjennomført en sprangtest som i trinn 2. Endring i bøtteverdi og sum av beregnet nedbørmengde for de seks støttypene er listet opp i tabell 19. Her ser man at bøtteverdiendringen er avhengig av tidsoppløsning for alle de seks støttypene, mens beregnet nedbørmengde er uavhengig av tidsoppløsning. Tabell 20 inneholder de beregnede *RMSE*-verdiene, avvikene og korrelasjonskoeffisientene for bøtteverdiendringene  $\Delta_{RA\_010}$  og  $\Delta_{RA\_1}$ . Støytype 6 skiller seg ut da denne har de største avvikene og laveste korrelasjonskoeffisientene, uavhengig av tidsoppløsning, i tillegg til den klart høyeste *RMSE*-verdien for  $RA\_1$ . Støytype 1, 2 og 4 har de samme estimerte verdiene og dermed de samme avvikene på både ti minutters oppløsning og timesoppløsning.  $\Delta_{RA\_1}$  for støytype 3 har det laveste avviket, med et avvik på - 0,01 mm. Med en

usikkerhet i referanseverdien på  $\pm 0,12$  mm, betyr det at ingen av avvikene i bølgeverdiendring for disse seks støtytypene er signifikante. Korrelasjonskoeffisienten  $r$  tyder på at bølgeverdiendringen både for timinuttersperiodene og for timene er perfekt positivt korrelert for støttype 1- 5.

Tabell 21 inneholder de absolutte og relative avvikene for de beregnede nedbørmengdene  $\Sigma RR_{010}$  og  $\Sigma RR_1$  for de seks støtytypene fra forsøkskonfigurasjon 3. Alle avvikene er positive og de er målt opp mot referanseverdien for beregnet nedbør, lik 20,90 mm. Støttype 1, 3 og 6 har de høyeste avvikene, med + 0,40 mm. Antas det samme usikkerhet i denne referanseverdien som usikkerheten i referanseverdien for bølgeverdiendring (dvs.  $\pm 0,12$  mm) er avvikene i beregnet nedbørmengde for alle støtytypene med unntak av støttype 5 signifikante.

Tabell 19: Tabellen inneholder de beregnede variablene  $\Delta RA_{010}$ ,  $\Sigma RR_{010}$ ,  $\Delta RA_1$  og  $\Sigma RR_1$  for de seks ulike støtytypene etter forsøkskonfigurasjon 3.

Støttype	1	2	3	4	5	6
$\Delta RA_{010}$ [mm]	21,86	21,86	21,86	21,86	21,86	21,82
$\Sigma RR_{010}$ [mm]	21,30	21,20	21,30	21,10	21,00	21,30
$\Delta RA_1$ [mm]	21,87	21,87	21,89	21,87	21,88	21,84
$\Sigma RR_1$ [mm]	21,30	21,20	21,30	21,10	21,00	21,30

Tabell 20: Tabellen inneholder de beregnede *RMSE*-verdiene, absolutte og relative avvik og korrelasjonskoeffisienten  $r$  for bølgeverdiendringene  $\Delta RA_{010}$  og  $\Delta RA_1$  for de seks støtytypene. Verdiene er målt opp mot referanseverdien 21,90 mm og de syntetiske settene med oppløsning på ti minutter og en time.

Støttype	1	2	3	4	5	6
$RMSE_{RA_{010}}$ [mm]	0,012	0,012	0,022	0,011	0,011	11,99
$RMSE_{RA_1}$ [mm]	0,0094	0,0096	0,021	0,0087	0,0086	23,98
Abs.avvik $_{\Delta RA_{010}}$ [mm]	-0,040	-0,040	-0,040	-0,040	-0,040	-0,080
Rel.avvik $_{\Delta RA_{010}}$ [%]	-0,18	-0,18	-0,18	-0,18	-0,18	-0,37
Abs.avvik $_{\Delta RA_1}$ [mm]	-0,030	-0,030	-0,010	-0,030	-0,030	-0,060
Rel.avvik $_{\Delta RA_1}$ [%]	-0,14	-0,14	-0,046	-0,14	-0,14	-0,27
$r_{RA_{010}}$ [mm]	1,00	1,00	1,00	1,00	1,00	0,57
$r_{RA_1}$ [mm]	1,00	1,00	1,00	1,00	1,00	0,25

Tabell 21: Tabellen inneholder de absolutte og relative avvikene for de beregnede nedbørmengdene  $\Sigma_{RR\_010}$  og  $\Sigma_{RR\_1}$  for de seks støytypene 1- 6. De absolutte og relative avvikene er beregnet opp mot referanseverdien for beregnet nedbørmengde lik 20,90 mm.

Støytype	1	2	3	4	5	6
<b>Abs.avvik</b> $_{\Sigma_{RR\_010}}$ [mm]	+0,40	+0,30	+0,40	+0,20	+0,10	+0,40
<b>Rel.avvik</b> $_{\Sigma_{RR\_010}}$ [%]	+1,91	+1,44	+1,91	+0,96	+0,48	+1,91
<b>Abs.avvik</b> $_{\Sigma_{RR\_1}}$ [mm]	+0,40	+0,30	+0,40	+0,20	+0,10	+0,40
<b>Rel.avvik</b> $_{\Sigma_{RR\_1}}$ [%]	+1,91	+1,44	+1,91	+0,96	+0,48	+1,91

En sammenligning på tvers av de tre forsøkene viser at de tre forsøkskonfigurasjonene gir de samme avvikene i bøtteverdiendring for alle støytypene. Ingen av avvikene er signifikante. Beregnet nedbørmengde for støytype 1 og 2 endrer seg med forsøkskonfigurasjonene. For støytype 1, som er ren positiv støy som gjentas hver halvtime, er beregnet nedbørmengde lik for forsøk 1 og 2, men øker med 0,30 mm i forsøk 3. Støytype 2 er ren negativ støy som gjentas hver halvtime. Her blir det en økning i beregnet nedbørmengde fra forsøk 1 til forsøk 2, mens forsøk 2 og 3 får like verdier. Økningen er på 0,20 mm. Disse økningene i beregnet nedbørmengde for støytype 1 og 2 gir dem større absolutte avvik da de beveger seg vekk fra referanseverdien for beregnet nedbørmengde, men de kommer nærmere endringen i bøtteverdi. For de fire andre støytypene er beregnet nedbørmengde lik for alle tre forsøkskonfigurasjoner.  $RMSE$ -verdier og korrelasjonskoeffisienter er like gjennom alle tre forsøkene, med unntak av  $RMSE_{RA\_010}$  for støytype 6 som endres fra forsøk 2 til forsøk 3.

#### 4.4.2 Støytype 7-10

Da resultatene fra de seks første støytypene ga veldig like verdier for de tre forsøkene, ble det bare gjennomført to ulike forsøkskonfigurasjoner for støytypene 7- 10, altså forsøk 1 og 2 som beskrevet i kapittel 3.5.

De fire støytypene 7- 10 inneholder en større mengde støy enn de seks første settene. Alle inneholder gaussisk fordelt støy, men med ulike standardavvik. Støytypene er beskrevet i tabell 4 med illustrasjon av det ufiltrerte høyfrekvente signalet. Målet med analysen er å finne en øvre terskel for standardavvik i støyen på signalet som algoritmen filtrerer godt.

Denne analysen benytter de samme referanseverdiene som de tidligere støysettene. Referanseverdien for bøtteverdiendringene  $\Delta_{RA\_010}$  og  $\Delta_{RA\_1}$  er lik 21,90 mm med en usikkerhet på  $\pm 0,12$  mm. Referanseverdien for beregnet nedbørmengde  $\Sigma_{RR\_010}$  og  $\Sigma_{RR\_1}$  er lik 20,90 mm. Det antas samme usikkerhet i denne verdien som i referanseverdien for bøtteverdiendring.

Maksimums- og minimumsverdier for de fire støysettene, i tillegg til maksimums- og minimumsverdier for det syntetiske settet er gitt i tabell 22. Støytype 7 har det høyeste standardavviket, mens støytype 10 har det laveste. Det ser man også av verdiene i tabellen, da støytype 7 har det største spennet i verdier, mens støytype 10 har det minste spennet.



Tabell 22: Tabellen inneholder maksimums- og minimumsverdier for det syntetiske settet og for støytpe 7- 10.

Støytpe	Syntetisk	7	8	9	10
Minimum [mm]	302,64	298,60	299,59	300,81	301,27
Maksimum [mm]	326,60	330,14	329,17	328,28	328,08

### Forsøk 1

I forsøk 1 av støyfiltreringen ble algoritmen kjørt som beskrevet i kapittel 3.1, men i trinn 4 ble medianen i de løpende treminuttersintervallene valgt i stedet for at kvantilverdi basert på standardavvik velges. Endring i bølgeverdi og sum av beregnet nedbørmengde for de fire støytpeene er listet opp i tabell 23. Her ser man at bølgeverdiendringen for de fire støytpeene er avhengig av tidsoppløsningen, mens beregnet nedbørmengde er uavhengig av tidsoppløsning. Beregnet nedbørmengde synker med synkende standardavvik i støyen. Tabell 24 inneholder de beregnede *RMSE*-verdiene, avvikene og korrelasjonskoeffisientene for bølgeverdiendringene  $\Delta RA_{010}$ ,  $\Delta RA_1$  og  $\Delta 24hdata$ . Støytpe 7 har de desidert høyeste *RMSE*-verdiene og avvikene, uavhengig av tidsoppløsning. Av tabellen ser man at både *RMSE*-verdiene og avvikene synker med synkende standardavvik i støyen, med lavest verdier for støytpe 10. Korrelasjonskoeffisienten er lavest for støytpe 8, og er tilnærmet 1 for de tre andre støytpeene. Referanseverdien har en usikkerhet på  $\pm 0,12$  mm. Det er bare avviket i  $\Delta RA_{010}$  for støytpe 9 som ikke er signifikant.

Tabell 25 inneholder de absolutte og relative avvikene for de beregnede nedbørmengdene  $\Sigma RR_{010}$  og  $\Sigma RR_1$  for de fire støytpeene fra forsøkskonfigurasjon 1. Alle avvikene er positive og de er målt opp mot referanseverdien for beregnet nedbør, lik 20,90mm. Avvikene i beregnet nedbørmengde er uavhengig av tidsoppløsning for disse fire støytpeene. Dersom det antas samme usikkerhet i denne referanseverdien som usikkerheten i referanseverdien for bølgeverdiendring (dvs.  $\pm 0,12$  mm) er alle avvikene signifikante. Avvikene er størst for støytpe 7, og selv om de avtar med avtagende standardavvik i støyen, er avviket for støytpe 10 fortsatt på + 6,90 mm hvilket tilsvarer + 33,01 %.

Tabell 23: Tabellen inneholder de beregnede variablene  $\Delta RA_{010}$ ,  $\Sigma RR_{010}$ ,  $\Delta RA_1$ ,  $\Sigma RR_1$  og  $\Delta 24hdata$  for de fire ulike støytpeene etter forsøkskonfigurasjon 1 med *AlgPYTN*.

Støytpe	7	8	9	10
$\Delta RA_{010}$ [mm]	38,82	22,26	21,93	21,77
$\Sigma RR_{010}$ [mm]	78,30	57,80	35,50	27,80
$\Delta RA_1$ [mm]	38,35	21,49	22,17	21,74
$\Sigma RR_1$ [mm]	78,30	57,80	35,50	27,80
$\Delta 24h data$ [mm]	75,32	58,58	40,08	33,60

Tabell 24: Tabellen inneholder de beregnede  $RMSE$ -verdiene, absolutte og relative avvik og korrelasjonskoeffisienten  $r$  for bøtteverdiendringene for de fire støytypene. Verdiene er målt opp mot referanseverdien 21,90 mm og de syntetiske settene med oppløsning på ti minutter og en time.

Støtype	7	8	9	10
$RMSE_{RA\_010}$ [mm]	4,14	0,32	0,11	0,078
$RMSE_{RA\_1}$ [mm]	1,97	0,30	0,11	0,078
Abs.avvik $_{\Delta RA\_010}$ [mm]	+16,92	+0,36	+0,03	-0,13
Rel.avvik $_{\Delta RA\_010}$ [%]	+77,26	+1,64	+0,14	-0,59
Abs.avvik $_{\Delta RA\_1}$ [mm]	+16,45	-0,41	+0,27	-0,16
Rel.avvik $_{\Delta RA\_1}$ [%]	+75,11	-1,87	+1,23	-0,73
Abs.avvik $_{\Delta 24hdata}$ [mm]	+53,42	+36,68	+18,18	+11,70
Rel.avvik $_{\Delta 24hdata}$ [%]	+243,93	+167,49	+83,01	+53,42
$r_{RA\_010}$	1,00	0,57	1,00	1,00
$r_{RA\_1}$	0,95	0,25	1,00	1,00

Tabell 25: Tabellen inneholder de absolutte og relative avvikene for de beregnede nedbørmengdene  $\Sigma_{RR\_010}$  og  $\Sigma_{RR\_1}$  for de fire støytypene støtype 7-10. De absolutte og relative avvikene er beregnet opp mot referanseverdien for beregnet nedbørmengde lik 20,90 mm.

Støtype	7	8	9	10
Abs.avvik $_{\Sigma_{RR\_010}}$ [mm]	+57,40	+36,90	+14,60	+6,90
Rel.avvik $_{\Sigma_{RR\_010}}$ [%]	+274,64	+176,56	+69,86	+33,01
Abs.avvik $_{\Sigma_{RR\_1}}$ [mm]	+57,40	+36,90	+14,60	+6,90
Rel.avvik $_{\Sigma_{RR\_1}}$ [%]	+274,64	+176,56	+69,86	+33,01

### Forsøk 2

I forsøk 2 på støyfiltreringen ble algoritmen kjørt som beskrevet i kapittel 3.1. Endring i bøtteverdi og sum av beregnet nedbørmengde for de fire støytypene er listet opp i tabell 26. Her ser man at bøtteverdiendringen for alle fire støytypene er avhengig av tidsoppløsningen, mens beregnet nedbørmengde er uavhengig av tidsoppløsningen. Bøtteverdiendringen  $\Delta_{RA\_1}$  for støtype 7 blir negativ. Årsaken til dette er diskutert i kapittel 5.4.3. Beregnet nedbørmengde synker med synkende standardavvik i støyen. Tabell 27 inneholder de beregnede  $RMSE$ -verdiene, avvikene og korrelasjonskoeffisientene for bøtteverdiendringene  $\Delta_{RA\_010}$ ,  $\Delta_{RA\_1}$  og  $\Delta_{24hdata}$ . Støtype 7 har de desidert høyeste  $RMSE$ -verdiene og avvikene, uavhengig av tidsoppløsning. Av tabellen ser

man at både  $RMSE$ -verdiene og avvikene synker med synkende standardavvik i støyen, med de laveste verdiene for støytype 10. Korrelasjonskoeffisienten er lavest for støytype 7 og 8, mens støytype 9 og 10 har  $r = 1,00$ . Referanseverdien har en usikkerhet på  $\pm 0,12$  mm. Det er bare avviket i  $\Delta RA_{010}$  for støytype 9 som ikke er signifikant.

Tabell 28 inneholder de absolutte og relative avvikene for de beregnede nedbørmengdene  $\Sigma RR_{010}$  og  $\Sigma RR_1$  for de fire støytypene. Alle avvikene er positive og de er målt opp mot referanseverdien for beregnet nedbør, lik 20,90 mm. Avvikene i beregnet nedbørmengde er uavhengig av tidsoppløsning for disse fire støytypene. Dersom det antas samme usikkerhet i denne referanseverdien som usikkerheten i referanseverdien for bølgeverdiendring (dvs.  $\pm 0,12$  mm) er alle avvikene signifikante. Avvikene er størst for støytype 7, og selv om de avtar med avtagende standardavvik i støyen, er avviket for støytype 10 fortsatt på + 6,90 mm hvilket tilsvarer + 33,01%.

Tabell 26: Tabellen inneholder de beregnede variablene  $\Delta RA_{010}$ ,  $\Sigma RR_{010}$ ,  $\Delta RA_1$  og  $\Sigma RR_1$  for de fire ulike støytypene etter forsøkskonfigurasjon 2 med *AlgPYTN*.

Støytype	7	8	9	10
$\Delta RA_{010}$ [mm]	53,58	22,26	21,93	21,77
$\Sigma RR_{010}$ [mm]	87,90	57,90	35,50	27,80
$\Delta RA_1$ [mm]	-198,22	21,49	22,17	21,74
$\Sigma RR_1$ [mm]	87,90	57,90	35,50	27,80
$\Delta 24h$ data [mm]	58,58	58,58	40,08	33,60

Tabell 27: Tabellen inneholder de beregnede  $RMSE$ -verdiene, absolutte og relative avvik og korrelasjonskoeffisienten  $r$  for bølgeverdiendringene  $\Delta RA_{010}$  og  $\Delta RA_1$  for de fire støytypene. Verdiene er målt opp mot referanseverdien 21,90 mm og de syntetiske settene med oppløsning på ti minutter og en time.

Støytype	7	8	9	10
$RMSE_{RA_{010}}$ [mm]	65,34	2,32	0,11	0,078
$RMSE_{RA_1}$ [mm]	71,40	0,31	0,11	0,078
Abs.avvik $_{\Delta RA_{010}}$ [mm]	+31,68	0,36	+0,03	-0,13
Rel.avvik $_{\Delta RA_{010}}$ [%]	+144,66	+1,64	+0,14	-0,59
Abs.avvik $_{\Delta RA_1}$ [mm]	-220,12	-0,41	+0,27	-0,16
Rel.avvik $_{\Delta RA_1}$ [%]	-1005,11	-1,87	+1,23	-0,73
Abs.avvik $_{\Delta 24hdata}$ [mm]	+36,68	+36,68	+18,18	+11,70
Rel.avvik $_{\Delta 24hdata}$ [%]	+167,49	+167,49	+83,01	+53,42
$r_{RA_{010}}$	0,18	0,56	1,00	1,00
$r_{RA_1}$	0,25	0,25	1,00	1,00

Tabell 28: Tabellen inneholder de absolutte og relative avvikene for de beregnede nedbørmengdene  $\Sigma_{RR\_010}$  og  $\Sigma_{RR\_1}$  for de fire støytypene støytype 7- 10. De absolutte og relative avvikene er beregnet opp mot referanseverdien for beregnet nedbørmengde lik 20,90 mm.

Støytype	7	8	9	10
<b>Abs.avvik</b> $_{\Sigma_{RR\_010}}$ [mm]	+67,00	+37,00	+14,60	+6,90
<b>Rel.avvik</b> $_{\Sigma_{RR\_010}}$ [%]	+320,57	+177,03	+69,86	+33,01
<b>Abs.avvik</b> $_{\Sigma_{RR\_1}}$ [mm]	+67,00	+37,00	+14,60	+6,90
<b>Rel.avvik</b> $_{\Sigma_{RR\_1}}$ [%]	+320,57	+177,03	+69,86	+33,01

Sammenligner man resultatene fra de to forsøkene på støyreduksjon for de fire støytypene beskrevet her, ser man at det ikke er forskjell i resultatene for støytype 9 og 10. Det er derimot stor variasjon i de beregnede verdiene for støytype 7. De beregnede nedbørmengdene for støytype 7 øker med nesten 10 mm fra forsøk 1 til forsøk 2. Det skjer også en økning for støytype 8, men den er på bare 0,10 mm. Dette indikerer at for støy med et høyt standardavvik (støytype 7 har et standardavvik i støyen på 1,00 mm), er forsøkskonfigurasjon 1 bedre enn forsøkskonfigurasjon 2, men algoritmen er uegnet til å filtrere støy med så høyt standardavvik. Dersom man kun ser på avvikene i bølgeverdiendring, er disse forholdsvis små, men signifikante, for støytype 8, 9 og 10. En kan derfor tenke at algoritmen er brukbar til å beregne bølgeverdiendring for støy med et standardavvik på opp til 0,75 mm. Avvikene i beregnet nedbørmengde er for høy for alle støytypene, men er lavest for støytype 10. Dette er logisk da denne støytypen har det laveste standardavviket i den tillagte støyen, med et standardavvik på 0,40 mm.

## 5 Diskusjon

### 5.1 Hammerfest

Analysen av datasettene fra værstasjonen Hammerfest Lufthavn ble gjort for å se på avvik i algoritmen som følge av ulik software. Dette ble gjort ved å sammenligne resultatene av *AlgPYTN* og *AlgFLOS* brukt på datasett fra totalt fire perioder. De fire periodene inneholdt ulikt antall nedbørhendelser og ulike totale nedbørmengder. Avvikene i bømteverdieregninger er generelt små, også mellom beregningene gjort av de to algoritmene. Tabell 29 inneholder den totale nedbøren i perioden, dvs. referanseverdien, med tilhørende usikkerhet og antall nedbørhendelser i perioden. Referanseverdien er beregnet ved å ta differansen mellom gjennomsnittet av de første 60 målingene og de siste 60 målingene. Usikkerheten er beregnet utfra standardavvikene for disse to gjennomsnittene og formel 2. Referanseverdien er beregnet slik for å redusere usikkerheten ved at hvert av endepunktene blir mindre påvirket av støyhendelser som kan finnes i de enkelte målingene.

Tabell 29 inneholder også avvikene mellom bømteverdiendringen beregnet med de to algoritmeversjonene og avvikene i beregnet nedbørmengde mellom de to algoritmeversjonene, relative til referanseverdiene. Begge variablene er beregnet for ti minutters tidsoppløsning.

Tabell 29: Tabellen inneholder de fire periodene evaluert fra Hammerfest Lufthavn med tilhørende referanseverdi, usikkerhet og antallet nedbørhendelser. Avvikene er forskjellen mellom verdiene beregnet med hver algoritme for hver periode, relativ til referanseverdien.

	Total nedbør [mm]	Usikkerhet [mm]	Antall nedbørhendelser	Avvik mellom $\Delta RA_{010}$ [%]	Avvik mellom $\Sigma RR_{010}$ [%]
<b>Periode 1</b>	36,99	± 0,21	8	0,84	6,00
<b>Periode 2</b>	30,44	± 0,42	6	0,92	0,00
<b>Periode 3</b>	89,59	± 0,36	5	0,98	-0,17
<b>Periode 4</b>	68,09	± 0,23	9	1,22	0,15

Avvikene er relative til referanseverdien, beregnet ved *AlgFLOS*-verdi minus *AlgPYTN*-verdi. For bømteverdiendringen er dette avviket relativt sett størst for periode 4. Dette kan komme av lengden på perioden, da periode 4 er den nest lengste av periodene. Det kan også skyldes antallet nedbørhendelser i denne perioden eller fordelingen av dem utover i perioden. Periode 4 har nedbør fordelt utover hele perioden, med bare et par opphold på noen timer. Dette skiller seg fra karakteristikken til periode 1, som er perioden med lavest avvik mellom de to periodene.

Avvikene i beregnet nedbørmengde,  $\Sigma RR_{010}$ , mellom de to algoritmene er generelt mindre enn for bømteverdiendringen. Unntaket er periode 1 som har et avvik mellom de to algoritmene på 6,00 % relativt til referanseverdien. Dette avviket tilsvarer ca. 2 mm, og er unaturlig stort sammenlignet med de andre avvikene i  $\Sigma RR_{010}$ . Av verdiene i tabell 6 og 8 ser man at hver av algoritmene har et stort avvik sammenlignet med de andre periodene. Dette kan tyde på at det er noe i datasettet fra denne perioden, men dette er ikke analysert nærmere. *AlgPYTN* beregner verdier med størst avvik mot referanseverdien for periode 1 og 4. I periode 2 beregner begge algoritmeversjonene samme verdi og i periode 3 har *AlgFLOS* det største avviket.

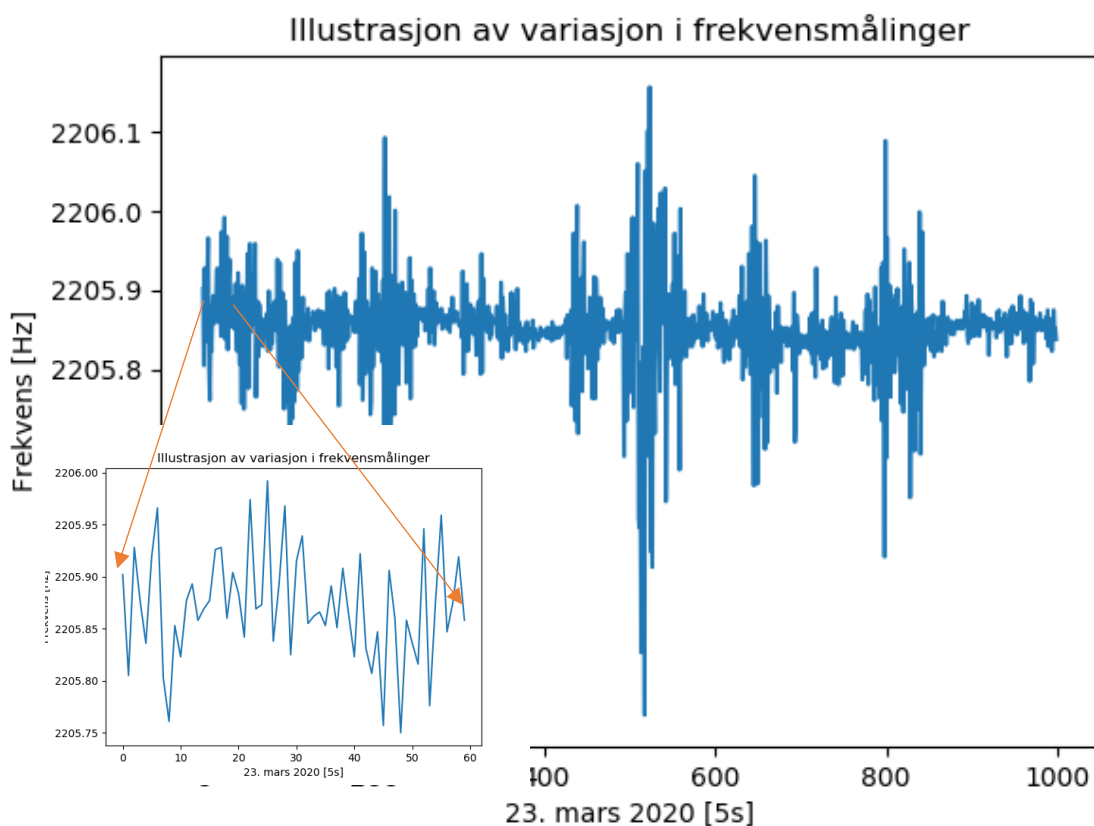
Hva avvikene mellom de to algoritmene skyldes er ikke analysert spesifikt, men det kan skyldes ulik software og programmeringsspråk for de to algoritmene. Programmeringsspråk har ofte innebygde kommandoer og funksjoner, og det kan være små forskjeller i hvordan disse funksjonene beregnes. Eksempelvis kan språkene se på et ulikt antall desimaler eller runde av på forskjellige steder underveis i beregningene. Dette er små differanser som over lengre tidsperspektiv kan akkumuleres opp til større avvik. Datasettene som blir benyttet er de samme, slik at det er lite trolig det er noe i dataene som skaper avvikene.

Som vist i tabell 6 i kapittel 4.1 er et par av avvikene i bøttestverdiendring for *AlgFLOS*, og så godt som alle beregnede nedbørmengdene for begge algoritmene, signifikante i forhold til usikkerhetene for hver av periodene. Av avvikene for  $\Delta_{RA\_010}$  i tabell 29 er avviket for periode 1, 3 og 4 signifikante. For beregnet nedbørmengde er det bare avviket i beregnet nedbørmengde for periode 1 som er signifikant. Avvikene ansees likevel for så små at *AlgPYTN* og *AlgFLOS* ansees som likeverdige versjoner.

## 5.2 Bjorli

Den nye algoritmen ble utarbeidet da man hadde et ønske om å gå fra en tidsoppløsning på en time til en tidsoppløsning på ti minutter. På værstasjonen på Bjorli er algoritmeversjonen *AlgGML* fortsatt operativ, derfor ble denne værstasjonen valgt for en sammenligning av gammel og ny algoritme. Analysen ble gjennomført for å se på differansen i beregningene gjort av *AlgPYTN* og *AlgGML* over den samme tidsperioden. De offisielle dataene fra værstasjonen i perioden er behandlet med *AlgGML*. Alle resultatene kan leses i kapittel 4.2.

Dataene fra Bjorli værstasjon er hentet fra ti dager i månedsskiftet mars/april 2020. Perioden inneholder seks nedbørhendelser. Disse hendelsene er avløst av lengre perioder uten registrert nedbør. Referanseverdien ble beregnet til 67,00 mm med en usikkerhet på  $\pm 0,076$  mm. Denne ble beregnet med utgangspunktet i gjennomsnittene av de første og siste 60 målingene i perioden. Den lave usikkerheten tyder på at det er lite støy i målingene som brukes til å sette referanseverdien. Dette ser man også av figur 24, som viser signalet for de første 60 målingene.



Figur 24: Figuren viser frekvenssignalet fra Bjorli den 23. mars 2020. De første 60 målingene som benyttes til å beregne referanseverdi og usikkerhet er forstørret.

De absolutte avvikene i beregningene gjort med de to algoritmene er gitt i tabell 30. Resultatene av analysen viser at avvikene i  $\Sigma_{RR\_010}$  og  $\Sigma_{RR\_1}$  beregnet med *AlgPYTN* ble beregnet til 0,50 mm, hvilket betyr at de er signifikante. *AlgGML* beregnet  $\Sigma_{RR\_010}$  til referanseverdien, men beregnet samme verdi som *AlgPYTN* for  $\Sigma_{RR\_1}$ . Dermed er avvikene for  $\Sigma_{RR\_1}$  like for de to algoritmene. For bøttestørrelsendringene  $\Delta_{RA\_010}$  og  $\Delta_{RA\_1}$  hadde ingen av beregningene gjort av *AlgPYTN* signifikante avvik, mens begge avvikene i beregningene gjort av *AlgGML* var signifikante.

Det er vanskelig å konkludere med at *AlgPYTN* er bedre enn *AlgGML* bare basert på denne analysen, da begge algoritmene beregner et par av verdiene med signifikante avvik. Dette er en kort periode, med relativt lite støy. Derfor kan det være interessant å behandle data fra en lengre periode eller en periode med mer støy, med begge algoritmeversjonene før man konkluderer endelig på hvilken av algoritmene som er best på støyreduksjon. Dersom man ser på avvikene for de to algoritmene, har *AlgPYTN* totalt sett de laveste avvikene, men *AlgGML* beregner mest nøyaktige resultat for  $\Sigma_{RR\_010}$ . Dette er litt overraskende, da det var forventet at *AlgPYTN* skulle være best på timinutters tidsoppløsning. Dette kan endres dersom man ser på en lengre periode med mer støy, og bør dermed analyseres nærmere.

Tabell 30: Tabellen inneholder de absolutte avvikene i beregningen i  $\Delta RA_{010}$ ,  $\Sigma RR_{010}$ ,  $\Delta RA_1$  og  $\Sigma RR_1$  beregnet med de to algoritmene *AlgPYTN* og *AlgGML*.

Variabel	Absolutt avvik <i>AlgPYTN</i> [mm]	Absolutt avvik <i>AlgGML</i> [mm]
$\Delta RA_{010}$	0	+0,17
$\Sigma RR_{010}$	-0,50	0
$\Delta RA_1$	-0,06	+0,68
$\Sigma RR_1$	-0,50	+0,50

### 5.3 Fordampning

Fordampning resulterer i at bølgeverdien synker i perioder uten nedbør. Dermed vil avlest bølgeinnhold avta, og det er en risiko for at nedbør ikke registreres før bølgeinnholdet er tilbake på det nivået det var før fordampningen fant sted. Dette vil gi feil i beregningene og føre til at mye nedbør ikke registreres. Dette forsøket ble gjennomført for å sjekke hvor vidt algoritmens kontrollnivåvariabel, *startRA*, fulgte bølgeinnholdet i fordampningsprosessen og dermed ville ha riktig utgangspunkt i en påfølgende nedbørhendelse.

De første 72 timene av det syntetiske datasettet ble benyttet som utgangspunkt for analysen. Inspirert av Ross et al. (2020) ble det tilsatt ekstra fordampning som beskrevet i kapittel 3.4. Dette ble gjort lineært. Fordampning ble lagt til slik at den totale fordampningen i perioden tilsvarte 10 % av nedbørmengden i det syntetiske settet, det vil si en total fordampning på 2,4 mm. Dette gav et signal som vist i figur 22 i kapittel 4.3.

Datasettet ble sendt gjennom *AlgPYTN* med forsøkskonfigurasjon 2 som forklart i kapittel 3.1.2. For denne analysen var det utviklingen til *startRA* i forhold til  $RA_{010}$  som var interessant. Figur 23 viser de to variablene *startRA* og  $RA_{010}$  for en tidsoppløsning på ti minutter. Som en ser av figuren følger de to variablene hverandre ganske tett, fra samme startpunkt til samme sluttspunkt, men det er mindre variasjoner underveis. Det at disse følger hverandre såpass godt betyr at algoritmen gjenkjenner fordampning og dermed tar dette med i beregningene.

I denne analysen ble det ikke testet med en påfølgende nedbørhendelse av kontrollert størrelse. Det antas likevel at algoritmen ville beregne denne korrekt, da nedbørberegningene tar utgangspunkt i differansen mellom kontrollvariabelen *startRA* og bølgeverdisignalet  $RA_{02}$ . Dermed vil beregningene bli riktig så lenge *startRA* følger bølgeverdien også i fordampningen.

Ross et al. (2020) ser på tidsintervaller på 1 minutt og beregner den kumulative fordampningen i perioden på bakgrunn av disse. Ved å beregne den kumulative fordampningen i en periode, kan man etter hvert få statistikk og trender også på dette. Dette kunne derfor vært interessant å innføre og lagre også i den norske algoritmen.



## 5.4 Støyreduksjon

### 5.4.1 Referanseverdien

Det syntetiske settet som er utgangspunktet for støyanalysen er beskrevet i kapittel 3.3. I tabell 3 er referanseverdien med tilhørende usikkerhet listet opp. Referanseverdien er differansen mellom gjennomsnittene av henholdsvis de første 60 og siste 60 målingene. Usikkerheten er beregnet ved hjelp av formel 2. Tabell 2 viser de beregnede variablene  $\Delta_{RA\_010}$ ,  $\Sigma_{RR\_010}$ ,  $\Delta_{RA\_1}$  og  $\Sigma_{RR\_1}$  for det syntetiske settet etter at det er behandlet med forsøkskonfigurasjon 2 slik den er beskrevet i kapittel 3.1.2. Sammenligner man disse beregnede verdiene med referanseverdien ser man at det er en differanse på 1,00 mm mellom referanseverdien og  $\Sigma_{RR\_010}$ .

Optimalt sett ville algoritmen beregnet  $\Sigma_{RR\_010}$  og  $\Sigma_{RR\_1}$  til referanseverdien dersom det syntetiske signalet hadde vært perfekt og helt uten støy. Denne differansen tyder dermed på at det er en liten mengde støy i det syntetiske signalet, hvilket ikke er utenkelig da det syntetiske settet er basert på virkelige målinger gjort av en værstasjon på Haukelisetet. På grunn av denne differansen som allerede er i det syntetiske settet, er det urealistisk å anta at algoritmen skal klare å beregne nedbørmengden lik bømteverdiendringen, uavhengig av karakteristikken i den tilsatte støyen. Et best mulig resultat for *AlgPYTN* i denne analysen vil derfor være en underestimering av beregnet nedbørmengde mot bømteverdiendring på 1,00 mm. Dette er årsaken til at det er benyttet to ulike referanseverdier; 21,90 mm for bømteverdiendringene  $\Delta_{RA\_010}$  og  $\Delta_{RA\_1}$ , og 20,90 mm for de beregnede nedbørmengdene  $\Sigma_{RR\_010}$  og  $\Sigma_{RR\_1}$ .

Hva dette avviket på 1,00 mm kommer av, er ikke analysert her. Som nevnt kan det skyldes støyen som allerede ligger i det syntetiske datasettet, men det kan også være et såkalt 'trace'-avvik. 'Trace'-avvik er avvik som kommer av at det er registrert nedbør på Ja/Nei-sensoren, men nedbørmengden er ikke stor nok til å overskride sensorens sensitivitet eller terskelverdien som benyttes i algoritmen (se f.eks. Mekis, 2005). I dette tilfellet er både sensitiviteten til sensoren og terskelverdien som benyttes i algoritmen lik 0,1 mm, hvilket betyr at mengder under 0,1 mm ikke blir fanget opp. En korrigering av trace-avvik gir alltid økte målinger. Mekis & Vincent (2011) opplevde en økning i årlig nedbørmengde på 18 % for en målestasjon som følge av korrigering av trace-avvik. Trace-avvik har naturlig nok større innvirkning på stasjoner i områder med lite årlig nedbør, da andelen trace utgjør av totalen blir større. Det kunne også tenkes at fordampning kunne gi ulike verdier for beregnet nedbørmengde og bømteverdiendring, men da ville beregnet nedbørmengde være større enn bømteverdiendringen. Dermed er det lite trolig at det er fordampning som er skyld i dette avviket.

Dersom man ser på resultatene generelt, ser en at beregnet nedbørmengde stort sett underestimeres (unntaket er støytype 7- 10, der de overestimeres). Dette kan komme av at differansene som er utgangspunktet for nedbørsberegningene trunkeres til en desimal. Siden verdiene trunkeres og ikke avrundes, vil utslag på 0,19 mm telle som 0,1 mm i beregningene. Dette kan potensielt sett føre til at deler av nedbørmengden ikke blir med i beregningene. Dette kan være årsaken til differansen på 1,00 mm som en ser i det syntetiske datasettet. Samtidig er sensitiviteten i Geonor-sensorene på 0,1 mm, slik at det vil være større usikkerhet i andre desimal. På den annen side registrerer sensoren frekvensen med to desimaler og dermed ned til 0,01 Hz, slik at to desimaler burde være pålitelige dersom en tar hensyn til litt større usikkerhet. Det kan derfor være en mulighet å trunkere verdier til to desimaler, heller enn én, og dermed potensielt fange opp mer av nedbørsignalet, men potensielt sett også mer av støyen. Dette er noe som burde testes grundigere før det eventuelt innføres i algoritmen.

Det er antatt at referanseverdien for den beregnede nedbørmengden  $\Sigma_{RR}$  har samme usikkerhet som referanseverdien for bømteverdiendringen  $\Delta_{RA}$ . Dette er en antagelse som trolig ikke stemmer helt, da den beregnede nedbøren benytter signalet fra Ja/Nei-sensoren i beregningen i tillegg til bømteverdisignalet. Det ligger en usikkerhet i Ja/Nei-sensorens sensitivitet som ikke tas hensyn til i beregningen av usikkerheten for bømteverdiendringen. Forsøk gjort av Leeper et al. (2015) viser en økt usikkerhet i distrometeret ved lavere temperaturer. Dette både grunnet usikkerheten de lavere temperaturene gir i instrumentet, i tillegg til at det er en usikkerhet i oppfangingen av fast nedbør (Leeper et al., 2015). En mulighet hadde vært å beregne usikkerheten i strømsignalet på samme måte som man beregnet usikkerheten i bømteverdisignalet, og at summen av disse to vil bli usikkerheten i referanseverdien for den beregnede nedbørmengden  $\Sigma_{RR}$ . På denne måten vil usikkerheten tilpasses perioder med ulike temperaturer. Dette er et av punktene som foreslås å undersøke nærmere i kapittel 5.5.

#### 5.4.2 Støytype 1-6

Det ble produsert seks ulike støysett for å teste algoritmens toleranse for ulike støykarakteristikker. Støysettene er lagd med mål om å etterligne reelle støysituasjoner. Støytype 1 og 2 inneholder henholdsvis ren positiv og ren negativ støy, der et støysignal på opptil 10 % av målingens verdi gjentas hver halvtime. Dette er støytyper som er blitt observert på værstasjoner tidligere. Algoritmen reduserer støyen i målingene godt og bømteverdisignalet går tilbake til det syntetiske signalet. Dette vises av lave avvik, en lav *RMSE*-verdi og en høy korrelasjonsfaktor. For positiv støy øker beregnet nedbørmengde i forsøk 3, mens det for negativ støy øker fra forsøk 1 til forsøk 2, mens 2 og 3 er like. Denne økningen fører den beregnede nedbørmengden lenger vekk fra referanseverdien for beregnet nedbør, men nærmere referanseverdien for bømteverdiendring. Dette er en svakhet med det syntetiske settet, da man i en perfekt situasjon vil at beregnet nedbørmengde skal være lik bømteverdiendring. Mulige årsaker til denne svakheten er diskutert i kapittel 5.4.1.

De resterende fire støytypene er produsert på bakgrunn av forsøkene gjort i Ross et al. (2020). I denne rapporten lagde de tre syntetiske signaler med gaussisk fordelt støy. Alle tre hadde en forventningsverdi lik 0, men standardavvikene var henholdsvis 0,1 mm, 0,01 mm og 0,001 mm. I de kanadiske forsøkene ble støyen normalisert før den ble lagt på det syntetiske signalet. Dette ble gjort for å unngå og innføre en bias. Støytype 3-5 tilsvarer disse settene, men støyen ble ikke normalisert før den ble lagt til på det syntetiske signalet. Støytype 6 inneholder gaussisk fordelt støy med forventningsverdi 0 og standardavvik 0,1 mm, der støyen ble normalisert. Avvikene mellom støytype 3 og 6 kan derfor antas å være en følge av normaliseringen som ble gjort. Resultatene av støytype 6 er derfor helt sammenlignbare med de kanadiske resultatene for støytypen med standardavvik 0,1 mm listet opp i tabell 31. For øvrig ble det ikke observert en bias under filtreringen av datasettene uten normalisert støy.

Det ble liten til ingen endring i de beregnede verdiene,  $\Delta_{RA}$  og  $\Sigma_{RR}$ , for de fire støytypene som følge av ulike forsøkskonfigurasjoner. Det er også små forskjeller i resultatene mellom støytype 3- 5, men ikke overaskende er det støytype 3 som får høyest *RMSE*-verdier av disse tre. *RMSE*-verdiene er fortsatt så små at støyredueringen kan ansees som vellykket. Mer overraskende er det at de er støytype 3 som har det laveste avviket for  $\Delta_{RA\_1}$ , da dette er settet har det høyeste standardavviket i støyen. Da mye av filtreringen baseres på gjennomsnitt, median og standardavvik ville det vært naturlig at et høyt standardavvik i støyen er mer krevende å filtrere tilbake til det opprinnelige

syntetiske signalet. Derfor var det forventet at støytype 3 ville ha de høyeste avvikene og  $RMSE$ -verdiene av disse seks støysettene.

Differansen i resultatene mellom støytype 3 og støytype 6 er relativt sett stor. De har de samme beregnede verdiene når det kommer til beregnet nedbørmengde, men støytype 6 har større avvik i bølgeverdiendring (0,04 mm større for  $\Delta RA_{010}$  og 0,05 mm større for  $\Delta RA_1$ ). Når det gjelder  $RMSE$ -verdier har støytype 6 de klart største verdiene på rundt 24 mm for  $RA_1$ .  $RMSE_{RA_{010}}$  endrer seg med forsøkskonfigurasjonene, fra 0,077 mm til 11,99 mm. Til sammenligning har støytype 3  $RMSE$ -verdier på 0,022 mm for  $RA_{010}$  og 0,021 mm for  $RA_1$ . Også korrelasjonen er dårligere for støytype 6, med  $r_{RA_{010}} = 0,57$  og  $r_{RA_1} = 0,25$  for de tre forsøkskonfigurasjonene. For støytype 3 er både  $r_{RA_{010}}$  og  $r_{RA_1}$  lik 1 ved alle tre forsøkskonfigurasjoner. Det kan være interessant å se om det blir tilsvarende endringer dersom man normaliserer støyene i støytype 4 og 5.

I Ross et al. (2020) benytter de akkumulerte verdier for 30 minutters intervaller i stedet for ti minutter eller timer som er benyttet her. I det kanadiske forsøket ser de på 44 tidsserier i perioden oktober til april 2014- 2017. Det antas at resultatene oppgitt i tabellene (gjengitt i tabell 31- 33) er gjennomsnittlige avvik,  $RMSE$  og korrelasjonskoeffisient  $r$  over disse 44 tidsseriene. Ross et al. (2020) analyserer fire ulike algoritmer, der  $NAF-SEG$  er den algoritmen som ligner mest på  $AlgPYTN$ . For  $AlgPYTN$  er både støytype 3 og støytype 6 listet opp, adskilt av skråstrek. For avvikene er det kun de relative avvikene som er sammenlignet, da det er heller usikkert hvorvidt den totale nedbørmengden i de kanadiske settene og settene brukt her er like. For  $AlgPYTN$  er det resultatene fra forsøkskonfigurasjon 2 som er benyttet.

Dersom man sammenligner avvikene i tabell 31 ser en at  $AlgPYTN$  har de laveste relative avvikene uavhengig av standardavviket i støyen. Ikke uventet er  $NAF-SEG$  den kanadiske algoritmen som gir avvik nærmest  $AlgPYTN$  i relativ størrelse. Når det gjelder korrelasjonskoeffisienter er disse likere hverandre, da de fleste er tilnærmet lik 1. Av de kanadiske algoritmene er det  $r_{O15}$  for støy med standardavvik 0,1 mm som skiller seg ut med en verdi på 0,87. Det er likevel støytype 6 filtrert med  $AlgPYTN$  som har de klart laveste korrelasjonskoeffisientene med  $r_{RA_{010}} = 0,57$  og  $r_{RA_1} = 0,25$ .  $RMSE$ -verdiene er generelt sett veldig lave for alle beregningene, men igjen er unntaket for støytype 6 som har  $RMSE_{RA_1} = 23,98$ . Det som skiller støytype 6 fra de andre støytypene behandlet med  $AlgPYTN$  er at støyen i dette er normalisert. Det kan derfor virke som at den kanadiske algoritmen  $NAF-SEG$  er bedre enn  $AlgPYTN$  på normalisert støy, men at  $AlgPYTN$  generelt sett er bedre med bakgrunn i jevnt over lavere avvik, høyere korrelasjon og lavere  $RMSE$ -verdier.

Tabell 31: Tabellen gir de relative avvikene for de fire kanadiske algoritmene i tillegg til  $AlgPYTN$  med tidsoppløsningene på 10 minutter og time. Verdiene for de fire kanadiske algoritmene er hentet fra tabell 2 i Ross et al. (2020). For  $AlgPYTN$  er både støytype 3 og støytype 6 listet opp med standardavvik 0,1 mm. Det er støytype 6 som er det tilsvarende settet til hva Ross et al. (2020) har testet.

Standardavvik	NAF [%]	NAF-S [%]	O15 [%]	NAF-SEG [%]	AlgPYTN $\Delta RA_{010}$ [%]	AlgPYTN $\Delta RA_1$ [%]
<b>0,001</b>	-10,1	-2,5	0,6	-1,1	-0,18	-0,14
<b>0,01</b>	-10,1	-1,3	12,9	0,4	-0,18	-0,14
<b>0,1</b>	-10,2	-1,0	33,2	1,9	-0,18/ -0,37	-0,046/ -0,27

Tabell 32: Tabellen gir Pearsons korrelasjonskoeffisient,  $r$ , for de fire kanadiske algoritmene i tillegg til  $r$  for *AlgPYTN* med tidsoppløsningene på 10 minutter og time. Verdiene for de fire kanadiske algoritmene er hentet fra tabell 3 i Ross et al. (2020). For *AlgPYTN* er både støytype 3 og støytype 6 listet opp med standardavvik 0,1 mm. Det er støytype 6 som er det tilsvarende settet til hva Ross et al. (2020) har testet.

	$r_{NAF}$	$r_{NAF-S}$	$r_{O15}$	$r_{NAF-SEG}$	<b>AlgPYTN</b> $r_{RA\_010}$	<b>AlgPYTN</b> $r_{RA\_1}$
<b>0,001</b>	0,97	0,99	0,94	0,99	1,00	1,00
<b>0,01</b>	0,97	0,98	0,92	0,98	1,00	1,00
<b>0,1</b>	0,95	0,96	0,87	0,96	1,00/0,57	1,00/0,25

Tabell 33: Tabellen gir *RMSE*-verdiene for de fire kanadiske algoritmene i tillegg til *AlgPYTN* med tidsoppløsningene på 10 minutter og time. Verdiene for de fire kanadiske algoritmene er hentet fra tabell 3 i Ross et al. (2020). For *AlgPYTN* er både støytype 3 og støytype 6 listet opp med standardavvik 0,1 mm. Det er støytype 6 som er det tilsvarende settet til hva Ross et al. (2020) har testet.

	$RMSE_{NAF}$ [mm]	$RMSE_{NAF-S}$ [mm]	$RMSE_{O15}$ [mm]	$RMSE_{NAF-SEG}$ [mm]	<b>AlgPYTN</b> $RMSE_{RA\_010}$ [mm]	<b>AlgPYTN</b> $RMSE_{RA\_1}$ [mm]
<b>0,001</b>	0,029	0,020	0,044	0,019	0,011	0,0085
<b>0,01</b>	0,032	0,025	0,053	0,024	0,011	0,0087
<b>0,1</b>	0,041	0,038	0,069	0,037	0,022/0,077	0,021/23,98

#### 5.4.3 Støytype 7-10

Analysen av støytype 7- 10 ble gjennomført for å finne en øvre grense for hva standardavviket i støyen kan være samtidig som algoritmen gir tilfredsstillende resultater. Det ble her testet fire ulike standardavvik; 1 mm, 0,75 mm, 0,5 mm og 0,4 mm. Resultatene ble som vist i tabell 23-28 i kapittel 4.4.2. Det ble bare testet to forsøkskonfigurasjoner på disse settene, forsøk 1 og 2 som forklart i kapittel 3.5. Disse to ble valgt på bakgrunn av resultatene fra analysen av støytype 1-6. Det kunne vært interessant å teste forsøkskonfigurasjon 3 også på disse settene. Dette fordi et større standardavvik i støyen betyr større sprang mellom to etterfølgende verdier, og dermed vil trolig sprangtesten i trinn 2 i algoritmen ha en større betydning.

Støytype 7 hadde det største standardavviket, og ikke overraskende var det denne støytypen som fikk de dårligste resultatene i begge forsøkene. I tabell 26 i kapittel 4.4.2 ser man at bølgeverdiendringen  $\Delta RA\_1$  for denne støytypen er på -198,22 mm, altså negativ. Dette stemmer naturligvis ikke med realiteten, men dette viser effekten av store støyutslag i enkeltmålinger som påvirker resultatene. Ved å undersøke det ferdige datasettet manuelt, ser man at ved å benytte differansen mellom nest siste og andre  $RA\_1$ -verdi, blir bølgeverdiendringen mye bedre; 21,89 mm. I forsøkskonfigurasjon 1 får støytype 7 korrelasjonskoeffisienter på rundt 1, hvilket er rart med tanke på at de relative avvikene er på over 75 %. I forsøkskonfigurasjon 2 er korrelasjonen mer som forventet, men  $r_{RA\_010} = 0,18$  og  $r_{RA\_1} = 0,25$ . Også støytype 8 har lave korrelasjonskoeffisienter (henholdsvis 0,56 og 0,25 for  $RA\_010$  og  $RA\_1$ ) og høye *RMSE*-verdier. *RMSE*-verdiene for støytype 7 og 8 øker fra forsøkskonfigurasjon 1 til forsøkskonfigurasjon 2, spesielt for  $RMSE_{RA\_1}$  for støytype 7 som går fra 1,97 til 71,40.

Resultatene forbedres med redusert standardavvik i støyen. For støytpe 9 og 10 er avvikene i bølgeverdiendringer akseptable, henholdsvis + 0,03 mm og - 0,13 mm for  $\Delta_{RA\_010}$  og + 0,27 mm og -0,16 mm for  $\Delta_{RA\_1}$ . Med en usikkerhet på  $\pm 0,12$  mm betyr dette at alle avvikene i bølgeverdiendringer er signifikante med unntak av avviket i  $\Delta_{RA\_010}$  for støytpe 9. Disse to støyettene er godt korrelerte med  $r = 1,00$ , og med en lav *RMSE*-verdi uavhengig av forsøkskonfigurasjon og tidsoppløsning. Dersom man i stedet ser på den beregnede nedbørmengden,  $\Sigma_{RR\_010}$  og  $\Sigma_{RR\_1}$ , er avvikene i disse fortsatt altfor store for disse to støyettene. Støytpe 10 har det laveste avviket, med et absolutt avvik på + 6,90 mm. Dette tyder på at et standardavvik på 0,4 mm fortsatt er for høyt til at algoritmen klarer å beregne nedbørmengden på en tilfredsstillende måte.

I analysen av støyet 7- 10 ble det beregnet en ekstra variabel,  $\Delta_{24hdata}$ , på bakgrunn av hva som har vært gjort i reelle situasjoner tidligere. Denne variabelen er en summering av alle differansene i bølgeverdiendring over 24 timer. Alle høyfrekvente bølgeverdier med registreringstidspunkt '06:00:00' ble plukket ut, og differansene mellom disse ble summert. Det var to dager der registreringen ikke hadde skjedd på helt klokkeslett, og der ble bølgeverdien med registreringstidspunkt '05:00:00' benyttet i stedet.

På forhånd ble det antatt at  $\Delta_{24hdata}$  ville gi et bedre estimat på beregnet nedbørmengde i datasett med mye støy. Dette stemmer til en viss grad. For støytpe 7, som inneholdt mest støy, ble denne nærmere referanseverdien enn beregnet nedbørmengde, men avviket er fortsatt på ca. 53 mm for forsøkskonfigurasjon 1 og ca. 36 mm for forsøkskonfigurasjon 2. For de resterende støyettene er avvikene i  $\Delta_{24hdata}$  større enn beregnet nedbørmengde med ti minutters eller times-oppløsning.  $\Delta_{24hdata}$  kan, på bakgrunn av disse resultatene, ikke anbefales som et estimat på beregnet nedbørmengde med disse tidsoppløsningene. Det kan være resultatet ville vært bedre dersom man sammenligner summen av  $\Sigma_{RR\_010}$  eller  $\Sigma_{RR\_1}$  med  $\Delta_{24hdata}$  over de samme 24 timene, men det er ikke gjort i denne oppgaven.

Som nevnt ble det bare benyttet to forsøkskonfigurasjoner i denne delen av analysen. Det var bare støytpe 7 som opplevde en vesentlig endring i resultatene mellom de to forsøkskonfigurasjonene, mens det for de tre andre støyettene var liten til ingen endring i resultatene. Leeper et al. (2015) testet en algoritmeversjon der de brukte vektet gjennomsnitt av tre strengsignaler, og opplevde denne som god på både filtrering av støy og fordampning, med små relative avvik selv ved store nedbørmengder og sterke støysituasjoner. En mulighet hadde vært å lage en forsøkskonfigurasjon som ligner på den de benytter i Leeper et al. (2015) for å se om filtreringen av store støymengder slik som de i støytpe 7 blir bedre. Dette vil imidlertid kreve at man har nedbørmålere med flere sensorer installert.

## 5.5 Videre arbeid

Oppgaven har et begrenset omfang og det er flere punkter det kunne vært interessant å undersøke nærmere.

For å evaluere prestasjonen av den nye algoritmen kontra den gamle algoritmen, *AlgGML*, bør begge algoritmene kjøres på et utvidet datasett. Dette kan gjerne være et syntetisk datasett, der man har en form for et fasitsignal og kan kontrollere mengden støy. Da vil det også være mulig å se på ulike støykarakteristikker. En tanke kan være å teste *AlgGML* på de ti støysettene som er benyttet i analysen forklart i kapittel 3.5.

For å optimalisere algoritmen er det nødvendig å vite effekten av de ulike trinnene og parameterne i algoritmen. Det burde derfor gjøres forsøk der man endrer de konfigurerbare parameterne i algoritmen for å se hvordan dette endrer filtreringen. Det kan da også være interessant å se på tidsintervallene det midles over og filtreres på bakgrunn av.

I denne oppgaven ble de første seks støytypene testet på tre ulike forsøkskonfigurasjoner. Disse tre ga stort sett svært like resultater, hvilket var bakgrunnen for at de fire siste støysettene bare ble testet på to ulike forsøkskonfigurasjoner. Da de siste fire støysettene inneholder støy med et større standardavvik, kan det være interessant å også kjøre forsøkskonfigurasjon 3, dvs. uten sprangtest, på disse settene.

Videre kan det være interessant å se på viktigheten av de ulike instrumentene. For eksempel bør effekten av en Ja/Nei-sensor utforskes nærmere. En bør både se på hvordan støy og usikkerhet i dette signalet påvirker algoritmen og den totale usikkerheten i resultatene. I tillegg kan det være interessant å se hvordan algoritmen opererer uten et signal fra Ja/Nei-sensoren. For støyanalysen i denne oppgaven hadde Ja/Nei-signalet en tidsoppløsning på ti minutter, slik at algoritmen ble kjørt uten et Ja/Nei-signal frem til dette. Det kan også være interessant å teste andre typer Ja/Nei-sensorer for å se om dette gir en bedre total. I den forbindelse kan man også se på rollen variabelen  $RDD$  vil ha. I disse forsøkene er denne beregnet, men ikke benyttet, da algoritmen er kjørt med et Ja/Nei-signal hele veien.

Forsøkene gjort på støytype 7-10 hadde som mål å finne en øvre grense for standardavvik i støyen som algoritmen fortsatt håndterer på en tilfredsstillende måte. Støytype 10 hadde et standardavvik på 0,4 mm. Algoritmen klarte å filtrere tilbake bølgeverdisignalet slik at avvikene i bølgeverdiendring ble relativt små, men avviket i beregnet nedbørmengde  $RR$ , ble fortsatt for høyt. Derfor bør det testes støysett med standardavvik mellom 0,1 mm og 0,4 mm for å finne den mengden der både bølgeverdiendringen og beregnet nedbørmengde har avvik som ikke er signifikante. Dette bør gjøres etter at forsøkskonfigurasjon 3 er testet på disse settene.

Meteorologisk institutt måler og lagrer også andre parametere på de ulike værstasjonene. Det kunne derfor vært interessant å legge inn målinger på disse, for eksempel vindhastighet eller temperatur, for å se om dette gir en forbedring av algoritmen og resultatene. Dersom man klarer å koble støyfiltreringen opp mot vindhastighet vil dette trolig være til hjelp i filtreringen av støy på stasjoner som er svært utsatt for vind.

## 6 Konklusjon

Det er små avvik i resultatene beregnet med algoritmen implementert i ulike programmeringsspråk, men disse avvikene er så små at algoritmeversjonene likestilles.

I analysen gjort på Bjorli blir det ikke signifikante forbedringer i beregningene gjort av den nye algoritmen kontra den gamle algoritmen. Det anbefales å teste begge algoritmene på en lengre tidsserie med kontrollert støy for å kunne gi et sikrere svar på avvikene mellom de to algoritmene.

Det er kun blitt testet en kort periode med kunstig fordampning, 72 timer med høyfrekvente målinger, men algoritmen lokaliserte fordampningen på en tilfredsstillende måte. Det kan være lurt å enten teste et større datasett med mer ujevn fordampning eller et datasett med fordampning med en påfølgende kjent nedbørmengde.

Støyanalysen viser at algoritmen takler ulike støysituasjoner godt. Det ble testet tre ulike forsøkskonfigurasjoner, hvilket alle ga svært like resultater. Dette tyder på at det ikke er nødvendig med filtrering basert på standardavvik eller sprangtest for at støyfiltreringa skal bli god. Det anbefales likevel å teste effekten av hvert enkelt element i algoritmen.

I de fleste støyforsøkene er det kun små forskjeller mellom resultatene med de ulike tidsoppløsningene. For beregnet nedbørmengde er resultatet uavhengig av tidsoppløsning, mens det for bølgeverdiendringer kan være forskjell på under 1 mm. Sammenlignet med de kanadiske algoritmene, viste *AlgPYTN* seg bedre på filtrering av ikke-normalisert støy, mens den kanadiske *NAF-SEG* var bedre på normalisert støy.

Analysen av støysett 7- 10 viser at bølgeverdiendringen på timinuttersoppløsning for støy med standardavvik på 0,4 mm gir et godt estimat på nedbørmengde. For beregnet nedbørmengde,  $\Sigma_{RR\_010}$ , er ikke algoritmen god nok selv for støysett med standardavvik 0,4 mm. Algoritmen bør derfor kjøres på støysett der støyen har et standardavvik mellom 0,1 mm og 0,4 mm for å kunne fastsette grensen sikkert.

Algoritmen filtrerer godt for de ulike støysettene som er analysert her, i tillegg til dataene hentet inn fra Hammerfest Lufthavn og Bjorli. Det bør likevel gjøres ytterligere undersøkelser der en ser på hvert av elementene i algoritmen, for å se effekten av dem og dermed kunne optimalisere algoritmen ytterligere.

## Referanser

- Andersen, M. S. (2011). Hvordan danner vann skyer? Retrieved from yr.no/artikkel/hvordan-danner-vann-skyer\_-1.7838779
- Bakkehøi, S., Øien, K., & Førland, E. J. (1985). An Automatic Precipitation Gauge Based on Vibrating-Wire Strain Gauges. *Hydrology Research*, 16(4), 193-202. doi:10.2166/nh.1985.0015
- Barry, R. G., & Chorley, R. J. (1968). *Atmosphere, weather and climate* (9. ed.). USA: Routledge.
- Benestad, R. E., Parding, K. M., Erlandsen, H. B., & Mezghani, A. (2019). A simple equation to study changes in rainfall statistics. *Environmental Research Letters*, 14 084017.
- Brækkan, R. (2007). Test av temperaturstabilitet for Geonorstrenger. Meteorologie og Klimadivisjon, Avdeling for Observasjonstjenester, intern rapport Nr. 57. Meteorologisk institutt. (ObsR\_057), 13.
- Brækkan, R. (2013). VKN prosjektet: Målekvalitet og støyanalyser. Meteorologie og Klimadivisjon, Avdeling for Observasjonstjenester, intern rapport Nr. 87. Meteorologisk institutt.
- Duchon, C. E. (2008). Using vibrating-wire technology for precipitation measurements. In M. S. (Ed.), *Precipitation: Advances in Measurement, Estimation and Prediction*. (pp. 33-58). Berlin, Heidelberg: Springer.
- Duchon, C. E., & Essenberg, G. R. (2001). Comparative rainfall observations from pit and aboveground rain gauges with and without wind shields *Water Resources Research*, 37(12), 3253-3263.
- Dyngeseth, H. (2004). *Nedbørmålingar frå Geonor T-200 samanlikna med manuelle målingar*.
- Førland, E. J., Allerup, P., Dahlström, B., Elomaa, E., Jónsson, T., Madsen, H., . . . Vejen, F. (1996). *Manual for operational correction of Nordic precipitation data* Oslo:
- Goodison, B. E., Louie, P. Y. T., & Yang, D. (1998). *WMO Solid Precipitation Measurement Intercomparison*.
- Hjermann, D. Ø., & Larsen, H. (2019). Regnskog. Retrieved from <https://snl.no/regnskog>
- Holder, C., Boyles, R., Syed, A., Niyogi, D., & Raman, S. (2006). Comparison of Collocated Automated (NCECNet) and Manual (COOP) Climate Observations in North Carolina. *Journal of Atmospheric and Oceanic Technology*, 23(5), 671-682. doi:10.1175/jtech1873.1
- IPCC. (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland*.
- Leeper, R. D., & Kochendorfer, J. (2015). Evaporation from weighing precipitation gauges: impacts on automated gauge measurements and quality assurance methods. *Atmospheric Measurement Techniques*, 8(6), 2291-2300. doi:10.5194/amt-8-2291-2015
- Leeper, R. D., Pakecki, M. A., & Davis, E. (2015). Methods to Calculate Precipitation from Weighing-Bucket Gauges with Redundant Depth Measurements. *Journal of Atmospheric and Oceanic Technology*, 32, 1179-1190. doi:10.1175/JTECH-D-14-00185.1
- Mathisen, T. I. (2004). *Studie av enkelttilfeller for nedbørmålinger fra Geonor T-200 sammenlignet med manuelle målinger* (ObsR\_014).
- Mekis, É. (2005). *Adjustments for Trace Measurements in Canada*. Paper presented at the 15th Conference on Applied Climatology, Savannah, GA: Am. Meteorol. Soc.
- Mekis, É., & Vincent, L. A. (2011). An Overview of the Second Generation Adjusted Daily Precipitation Dataset for Trend Analysis in Canada. *Atmosphere-Ocean*, 49(2), 163-177. doi:10.1080/07055900.2011.583910
- Meteorologisk Institutt. (2017, 21.10.2019). Klima fra 1900 til i dag. Retrieved from <https://www.met.no/vaer-og-klima/klima-siste-150-ar>
- Meteorologisk Institutt. (n.d.). Værstasjoner i norske kommuner.
- Moody, J. (2019, 05.09.2019). What does RMSE really mean?
- NASA. (n.d.). Total Rainfall. Retrieved from [https://earthobservatory.nasa.gov/global-maps/TRMM\\_3B43M](https://earthobservatory.nasa.gov/global-maps/TRMM_3B43M)
- National Geographic Society, N. (2019, 07.08.2019). Precipitation. Retrieved from <https://www.nationalgeographic.org/encyclopedia/precipitation/>



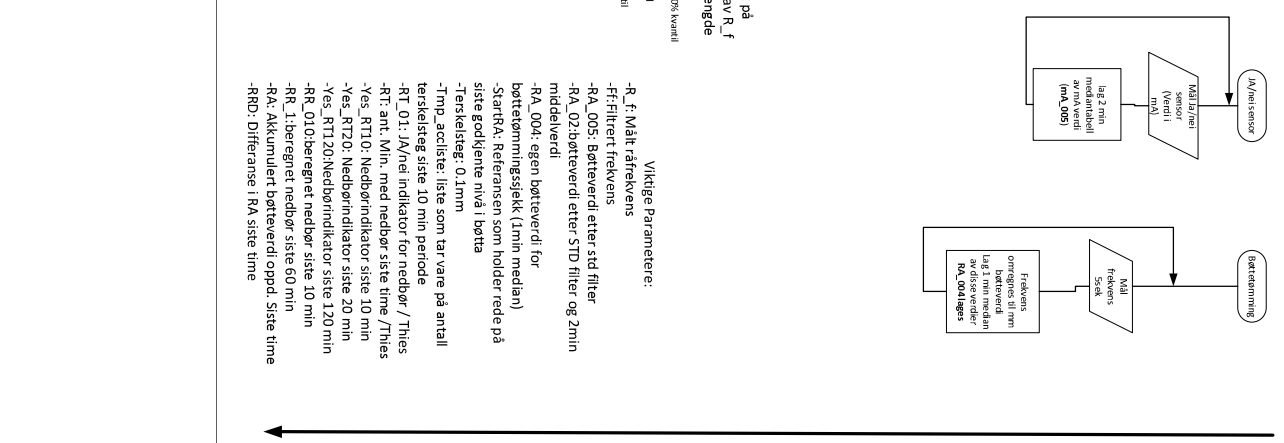
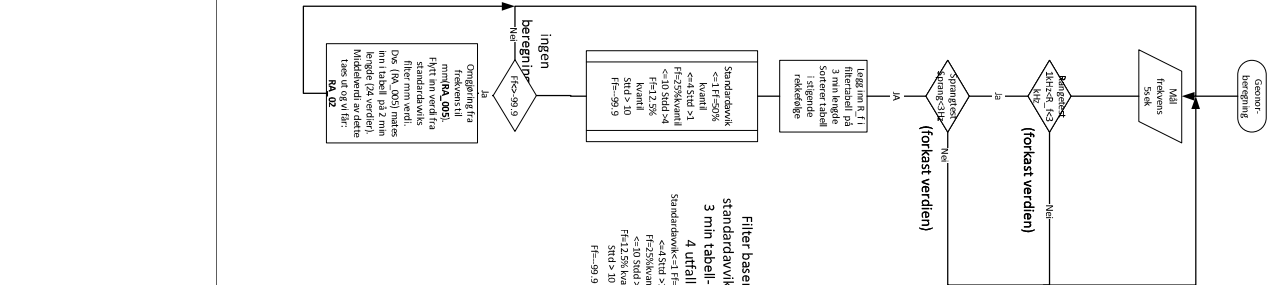
- NCSU. (n.d.). Composition of the atmosphere. Retrieved from <https://climate.ncsu.edu/edu/Composition>
- Neverdal, S. (2020, 08.01.2020). Naturskader for over en halv milliard i 2019. Retrieved from <https://www.finansnorge.no/aktuelt/nyheter/2020/01/naturskader-for-over-en-halv-milliard/>
- Nitu, R., Roulet, Y.-A., Wolff, M., Earle, M., Reverdin, A., Smith, C., . . . Yamashita, K. (2018). *WMO Solid Precipitation Intercomparison Experiment (SPICE)*. Geneva, Switzerland:
- Norsk Landbruk. (2018). 4 av 10 slakter flere dyr på grunn av førmangel. Retrieved from <https://www.norsklandbruk.no/husdyr/fire-av-ti-vil-slakte-flere-dyr-i-host/>
- NVE. (2010-2020). Flomhendelser. from Norges vassdrag- og energidirektorat
- Rasmussen, R., Baker, B., Kochendorfer, J., Meyers, T., Landolt, S., Fischer, A. P., . . . Gutmann, E. (2012). How Well Are We Measuring Snow: The NOAA/FAA/NCAR Winter Precipitation Test Bed. *Bulletin of the American Meteorological Society*, 93(6), 811-829. doi:10.1175/bams-d-11-00052.1
- Rasmussen, R., Dixon, M., Hage, F., Cole, J., Wade, C., Tuttle, J., . . . Rehak, N. (2001). Weather Support to Deicing Decision Making (WSDDM): A Winter Weather Nowcasting System. *Bulletin of the American Meteorological Society*, 82(4), 579-596. doi:10.1175/1520-0477(2001)082<0579:Wstddm>2.3.Co;2
- Ross, A., Smith, C. D., & Barr, A. (2020). An improved post-processing technique for automatic precipitation gauge time series. *Atmospheric Measurement Techniques*, 13(6), 2979-2994. doi:10.5194/amt-13-2979-2020
- Sevruk, B., Ondrás, M., & Chvíla, B. (2009). The WMO precipitation measurement intercomparisons. *Atmospheric Research*, 92(3), 376 - 380. doi:<https://doi.org/10.1016/j.atmosres.2009.01.016>
- Smith, C. D. (2009, 20.-23. april 2009). *The Relationships Between Snowfall Catch Efficiency and Wind Speed for the Geonor T-200B Precipitation Gauge Utilizing Various Wind Shield Configurations*. Paper presented at the 77th Western Snow Conference, Canmore, BC, Canada.
- Statkraft. (n.d.). Vår virksomhet - Vannkraft. Retrieved from <https://www.statkraft.no/var-vmirksomhet/vannkraft/>
- Strangeways, I. (1996). Back to basics. *Weather*, 51(8), 274-279. doi:10.1002/j.1477-8696.1996.tb05519.x
- Strangeways, I. (2004). Improving precipitation measurement. *International Journal of Climatology*, 24, 1443-1460. doi:10.1002/joc.1075
- Thies GmbH. (n.d.). Thies Clima Precipitation Retrieved from <https://www.thiesclima.com/en/Products/Precipitation-Electrical-devices/>
- Van Rossum, G., & Drake, F. L. (2009). *Python 3 Reference Manual*. GreateSpace, Scotts Valley, CA. ISBN 1441412697.
- Williams, M. (2014). What percent of Earth is water? Retrieved from <https://phys.org/news/2014-12-percent-earth.html>
- WMO. (2018). *Guide to Instruments and Methods of Observation*. Geneva, Switzerland:
- Wolff, M., Isaksen, K., Brækkan, R., Alfnes, E., Petersen-Øverleir, A., & Ruud, E. (2013). Measurements of wind-induced loss of solid precipitation: description of a Norwegian field study. *Hydrology Research*, 44(1), 35-43. doi:10.2166/nh.2012.166
- Wolff, M., Isaksen, K., Petersen-Øverleir, A., Ødemark, K., Reitan, T., & Brækkan, R. (2015). Derivation of a new continuous adjustment function for correcting wind-induced loss of solid precipitation: results of a Norwegian field study. *Hydrol. Earth Syst. Sci.*, 19(2), 951-967. doi:10.5194/hess-19-951-2015
- Wolff, M., Isaksen, K., Ødemark, K., Petersen-Øverleir, A., Reitan, T., & Brækkan, R. (2013). *Vindkorreksjon av nedbør* (22/2013). Oslo:
- Yang, D., Kane, D., Zhang, Z., Legates, D., & Goodison, B. (2005). Bias corrections of long-term (1973–2004) daily precipitation data over the northern regions. *Geophysical Research Letters*, 32(19). doi:10.1029/2005gl024057

## Vedlegg

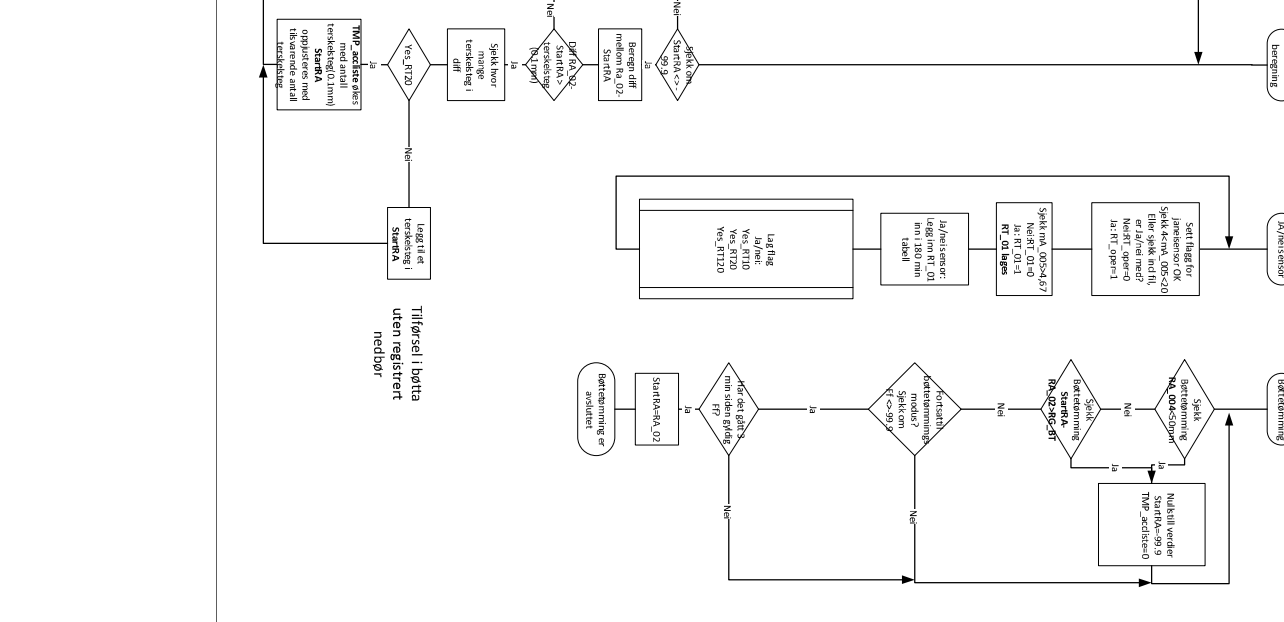
A – Flytskjema

B – Python script for AlgPYTN

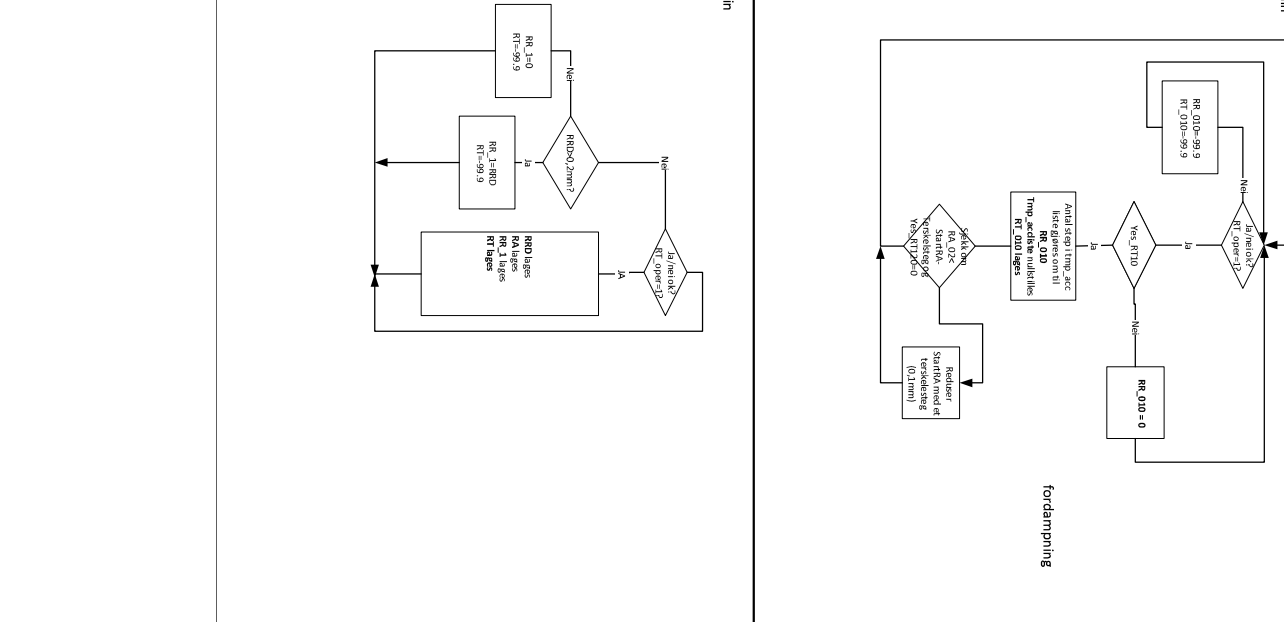
5 sek



10 min



60 min



fordampning

```

# -*- coding: utf-8 -*-
"""
Created on Thu Aug  6 20:16:35 2020

@author: maced
"""

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

#%%

data_f = pd.read_csv('localhost_ParkenGeonorHammerfest_2087_02.csv', sep = ',',
                    names=['timestamp', 'mA_005', 'R_f', 'Ff', 'midlet_ff'])

data_f = data_f.drop(['Ff', 'midlet_ff'], axis = 1)

#%%
#sjekker frekvensen opp mot grenseverdier
data_f.loc[data_f.R_f > 3000, 'R_f'] = 0
data_f.loc[data_f.R_f < 1000, 'R_f'] = 0

#%%
#tar medianen av strømsignalet i et tominutters-intervall
data_f['mA_005'] = data_f['mA_005'].rolling(window = 24, min_periods = 1).median()

#%%
#Sprangtest
data_f.loc[data_f.R_f.diff() > 3, 'R_f'] = 0
data_f.loc[data_f.R_f.diff() < -3, 'R_f'] = 0

#%%
"""
Standardavvikfilter:
    Velger ut kvantilverdi basert på standardavviket i et løpende treminuttersintervall
"""

Filter_3min = []
E_frekvenser = [0]*len(data_f)
kvantile_f = [0]*len(data_f)

for i in range(35, len(data_f)):
    Filter_3min = data_f.loc[i-35:i]['R_f']
    Filter_3min.sort_values()
    std_3min = Filter_3min[Filter_3min != 0].std()

    if std_3min <= 1:
        E_frekvenser[i]= np.quantile(Filter_3min, 0.50, axis = 0)
        kvantile_f[i] = 0.5
    elif std_3min > 1 and std_3min <= 4:
        E_frekvenser[i]= np.quantile(Filter_3min, 0.25, axis = 0)
        kvantile_f[i] = 0.25
    elif std_3min > 4 and std_3min <= 10:
        E_frekvenser[i]= np.quantile(Filter_3min, 0.125, axis = 0)
        kvantile_f[i] = 0.125

Filter = []
for i in range(35):

```

```

Filter = data_f.loc[:,i]['R_f']
Filter.sort_values()
std = Filter[Filter!=0].std()

if std <= 1:
    E_frekvenser[i]= np.quantile(Filter, 0.50, axis = 0)
    kvantile_f[i] = 0.5
elif std > 1 and std <= 4:
    E_frekvenser[i]= np.quantile(Filter, 0.25, axis = 0)
    kvantile_f[i] = 0.25
elif std > 4 and std <= 10:
    E_frekvenser[i]= np.quantile(Filter, 0.125, axis = 0)
    kvantile_f[i] = 0.125

E_frekvenser[0] = data_f.R_f[0]

data_f.insert(3, 'Filtrert_frekvens', E_frekvenser)

###
"""
Omregning fra Hz til mm vha. kalibreringssertifikat for stasjonen.
Input:
    ff - filtrert frekvenssignal
Output:
    mm - bøttestverdi i millimeter
"""
RA_005 = []

def hammerfest(ff):
    A = 1.65064*pow(10, -2)
    B = 9.15152*pow(10, -6)
    f0 = 1060.7
    cm = A*(ff - f0) + B*pow((ff - f0), 2)
    mm = cm * 10
    return mm

RA_005 = hammerfest(ff = data_f.Filtrert_frekvens)
data_f.insert(4, 'RA_005', RA_005)

###
"""
Beregner bøttestverdi og strømsignal for minutter ved å ta gjennomsnittet av
registrerte verdier innenfor samme minutt.
"""
RG_BT = 10
terskelverdi = 0.1

RA_01 = []
RT_01 = []
RA_004 = []

tid = data_f['timestamp'].astype(str)
tid = tid.str[:2]
tid = pd.unique(tid).astype(str)

for i in tid:
    RT = []
    ra = []

```

```

RT.append(data_f.loc[data_f.timestamp.astype(str).str[: -2] == i, 'mA_005'])
ra.append(data_f.loc[data_f.timestamp.astype(str).str[: -2] == i, 'RA_005'])

RA_004.append(ra[0].median())
RA_01.append(ra[0].sum()/len(ra[0]))
RT_01.append(RT[0].sum()/len(RT[0]))

###
"""
Gjør om midlet strømsignal til booleansk variabel vha. terskelverdi på 4,67 mA
"""
for i in range(len(RT_01)):
    if RT_01[i] >= 4.67:
        RT_01[i] = 1
    else:
        RT_01[i] = 0

kombi = list(zip(tid, RA_01, RT_01, RA_004))
minutt_f = pd.DataFrame(kombi, columns = ['Tid', 'RA_01', 'RT', 'RA_004'])

###
"""
Sjekker hvor mange minutter det har regnet de siste 10, 20 og 120 minuttene
"""

yes_RT20 = [0]*len(minutt_f)
yes_RT10 = [0]*len(minutt_f)
yes_RT120 = [0]*len(minutt_f)

yes_RT10 = minutt_f['RT'].rolling(window = 10, min_periods = 1).sum()
yes_RT20 = minutt_f['RT'].rolling(window = 20, min_periods = 1).sum()
yes_RT120 = minutt_f['RT'].rolling(window = 120, min_periods = 1).sum()

minutt_f.insert(4, 'yes_RT10', yes_RT10)
minutt_f.insert(5, 'yes_RT20', yes_RT20)
minutt_f.insert(6, 'yes_RT120', yes_RT120)

###
"""
Midler bølgeverdi over to minutter
"""

RA_02 = []
RA_02 = minutt_f['RA_01'].rolling(window = 2, min_periods = 1).mean()
minutt_f.insert(2, 'RA_02', RA_02)

###
"""
Lager kontrollvariabelen startRA
"""
startRA_f = [0]*len(minutt_f)
TMP_acclist_f = 0
i = 3
startRA_f[0] = minutt_f.RA_02[0]
startRA_f[1] = minutt_f.RA_02[1]
startRA_f[2] = minutt_f.RA_02[2]
startRA_f[3] = minutt_f.RA_02[3]

```

```

while i < len(minutt_f):
    if minutt_f.RA_004[i] < 50:
        startRA_f[i+1] = -99.9
        TMP_acclist_f = 0
        print(i, 'a')
    elif startRA_f[i] - minutt_f.RA_02[i] > RG_BT:
        startRA_f[i+1] = -99.9
        TMP_acclist_f = 0
        print(i, 'b')
    else:
        if 0 not in minutt_f.RA_01[i-3:i]:
            startRA_f[i+1] = minutt_f.RA_02[i]
        else:
            startRA_f[i+1] = -99.9
            print(i, 'c')
    i = i+1

###
"""
Beregner differansen mellom startRA og RA_02 for hvert minutt.
"""
diff_1min = []
i=0
while i < len(minutt_f)-1:
    if startRA_f[i] != -99.9:
        diff_1min.append(minutt_f.RA_02[i]- startRA_f[i])
        if diff_1min[i] > terskelverdi and minutt_f.yes_RT20[i] >= 1:
            a = diff_1min[i]//terskelverdi
            TMP_acclist_f = TMP_acclist_f + a
            startRA_f[i+1] = startRA_f[i] + a*terskelverdi
        elif diff_1min[i] > terskelverdi and minutt_f.yes_RT20[i] < 1:
            startRA_f[i+1] = startRA_f[i] + terskelverdi
        else:
            startRA_f[i+1] = startRA_f[i]

        if minutt_f.RA_02[i] < (startRA_f[i] - terskelverdi) and minutt_f.yes_RT120[i] == 0:
            startRA_f[i+1] = startRA_f[i] - terskelverdi

    else:
        diff_1min.append(-999)

    i = i + 1

diff_1min.append(0)
minutt_f.insert(8, 'diff_1min', diff_1min)

###
"""
Kode for 10 min

Plukker ut RA_02 og startRA registrert ved 'tt:m0'. Summerer RT_01 innenfor hver
10 min-periode. Differansen trunkeres før RR_010 beregnes.
"""

tid10 = minutt_f['Tid'].astype(str)
tid10 = tid10.str[: -1]
tid10 = pd.unique(tid10).astype(str)

RR_010 = [0]*(len(tid10)+1)

```

```

RT_010 = [0]*(len(tid10)+1)
RA_010 = [0]*(len(tid10)+1)
k=0

for i in tid10:
    k = k+1
    rain_mean_10min = []
    RT10 = []

    rain_mean_10min.append(minutt_f.loc[minutt_f.Tid.astype(str).str[:] == i + '0', 'RA_02'])
    RT10.append(minutt_f.loc[minutt_f.Tid.astype(str).str[:-1] == i, 'RT'])
    rr10 = pd.DataFrame(minutt_f.loc[minutt_f.Tid.astype(str).str[:-1] == i, 'diff_1min'], columns=['diff_1min'])

    rr10.loc[rr10.diff_1min<0.1, 'diff_1min'] = 0
    rr10 = rr10['diff_1min']//terskelverdi
    rr10 = rr10*terskelverdi

    RR_010[k]= sum(rr10)
    RA_010[k-1] = rain_mean_10min[0].sum()
    RT_010[k] = RT10[0].sum()

tid10 = np.append(tid10, '20200229236')

kombi = list(zip(tid10, RA_010 , RT_010, RR_010))
timinutter_f = pd.DataFrame(kombi, columns = ['Tid10', 'RA_010', 'RT_010', 'RR_010'])

timinutter_f = timinutter_f[:][:-1]
###
"""
Setter alle RR_010 der RT_010 == 0 til 0
"""
timinutter_f.loc[timinutter_f.RT_010 == 0, 'RR_010'] = 0

###
"""
Kode for time

Plukker ut RA_010 og registrert ved 'tt:00'. Summerer RT_010 og RR_010
innenfor hver time
"""

tid60 = []
tid60 = timinutter_f.Tid10.str[:-1]
tid60 = pd.unique(tid60).astype(str)

RR_1 = [0]*(len(tid60)+1)
RT_1 = [0]*(len(tid60)+1)
RA = [0]*(len(tid60)+1)
k=0

for i in tid60:
    k = k+1
    rain_mean_60min = []
    RT60 = []
    rr60 = []

    rain_mean_60min.append(timinutter_f.loc[timinutter_f.Tid10.astype(str).str[:] == i + '0', 'RA_02'])
    RT60.append(timinutter_f.loc[timinutter_f.Tid10.astype(str).str[:-1] == i, 'RT_010'])
    rr60.append(timinutter_f.loc[timinutter_f.Tid10.astype(str).str[:-1] == i, 'RR_010'])

```



```

if 'NaN' in rr60[0] or 'NaN' in rain_mean_60min[0] or 'NaN' in RT60[0]:
    break
else:
    RR_1[k]= rr60[0].sum()
    RA[k-1] = rain_mean_60min[0].sum()
    RT_1[k] = RT60[0].sum()

tid60 = np.append(tid60, '2020023000')

kombi2 = list(zip(tid60, RA, RT_1, RR_1))
time_f = pd.DataFrame(kombi2, columns = ['Time', 'RA_1', 'RT_1', 'RR_1'])

time_f = time_f[:][:-1]

#%%
"""
Beregner differansen mellom to RA_1-verdier
"""
RDD = time_f.RA_1.diff()
time_f.insert(4, 'RDD', RDD)

```



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

Postboks 5003  
NO-1432 Ås  
Norway