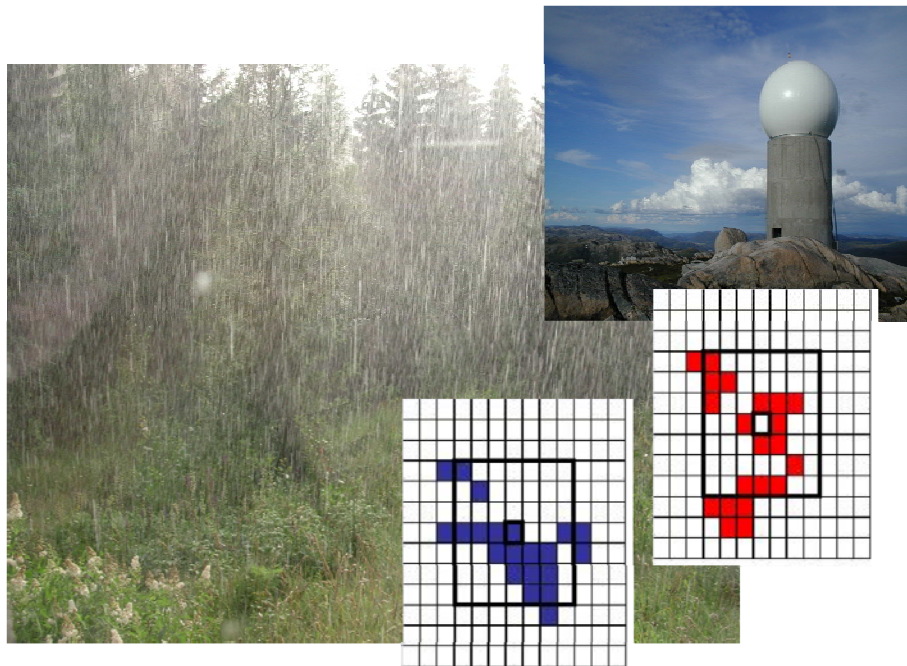


En skalaavhengig verifikasjonsmetode anvendt på finskalamodeller.

Bente Marie Wahl



UNIVERSITETET I OSLO

DET MATEMATISK-NATURVITENSKAPELIGE FAKULTET

En skalaavhengig verifikasjonsmetode anvendt på finskalamodeller.

Bente Marie Wahl



Masteroppgave i geofag: meteorologi og oseanografi
Institutt for geofag, Matematisk-naturvitenskaplig fakultet

UNIVERSITETET I OSLO

Januar 2010

© **Bente Marie Wahl, 2010**

Veiledere: Thor Erik Nordeng, UiO, Jørn Kristiansen, Meteorologisk Institutt, Oslo

Dette eksamensarbeidet er publisert elektronisk i DUO – Digitale Utgivelser ved UiO

<http://www.duo.uio.no>

Det er også katalogisert i BIBSYS (<http://www.bibsys.no/>)

All rights reserved. No part of this publication may be reproduced or transmitted, in any form or by any means, without permission.

Forord

Det er med en uvirkelig følelse jeg sitter her og skriver de siste ordene til denne oppgaven. Hadde jeg vist hva jeg gikk til for litt over to år siden hadde jeg ikke sittet her nå. Jeg har lært utrolig mye denne tiden men også hatt mange tunge og frustrerende stunder.

Så det er mange som bør takkes for at jeg til slutt kom i mål. Jeg bør begynne med min sjef, Tor Helge Skaslien (Meteorologisk Institutt, Værvarslingsavdelingen). Underveis har jeg ikke alltid vært like glad for at han fikk meg til å begynne på dette studiet. Men i dag kan jeg ikke annet enn å si takk. Hadde ikke han ”pushet” litt på og lagt ting til rette så jeg kunne gjøre dette ved siden av jobben, hadde jeg aldri sittet her med en ferdig oppgave i hånden. Så må jeg gi en stor takk til mine veiledere Thor Erik Nordeng (Universitet i Oslo, Institutt for Geofag, Avdeling MetOs) og Jørn Kristiansen (Meteorologisk Institutt, Forskningsavdelingen) som har holdt ut med meg disse årene og tok på seg oppgaven. Stor takk til Thor Erik som virkelig stilte opp da jeg trengte det som mest. Og stor takk til Jørn som har vært veldig tålmodig og til stor hjelp med programmeringen. Takk til dere begge for all hjelp i innspurten hvor jeg virkelig har trengt litt veiledning.

En stor takk til Trygve Aas (Meteorologisk Institutt, Observasjonsdivisjonen) som alltid har vært tilgjengelig, spesielt i radarspørsmål. Jeg må også gi en stor takk til mange gode kollegaer som har vært en stor støtte for meg og hatt troen på meg hele tiden. Alltid stilt velvillig opp når jeg har trengt hjelp og ikke minst holdt ut med mitt varierende humør på vakt. Og sist men ikke minst må jeg gi en kjempestor takk til min mann og mine to barn som virkelig har vært tålmodige med meg disse årene og stilt opp for meg på alle de måter de har hatt mulighet til.

Oslo, 11.1.2010 Bente Marie Wahl

En blir ikke skikkelig klok før en har vært tre ganger gal!

Abstract

High resolution forecast models have the potential to produce more detailed forecasts of, for example precipitation, than comparable models with a coarser resolution. This is because high resolution forecast models, amongst other advantages, give a more accurate description of weather phenomena such as intense convective precipitation systems. However, high resolution forecast models do not always score better with standard verification methods than models with a coarser resolution. This can be explained by the fact that individual phenomena, for example showers, cannot be forecasted at the exact place at the right time, even though the structure and intensity of the precipitation system is properly forecasted and in the right area. The forecast itself is good, even though the details are not entirely correct. We can easily assess this subjectively in each separate case. However, we need objective methods in order to assess the models qualitatively. Therefore, in the last few years, several new verification methods have been developed to provide more diagnostic and informative evaluations of high resolution forecasts of spatial fields, such as precipitation.

In this thesis such a verification method is tested using Fraction Skill Score, FSS, which takes into consideration the spatial scale of the parameter to be forecasted, a so-called neighbourhood-area method. The method is tested on the UK Met. Office model UM 4 km and verified against Norwegian radars over an area which covers most of Southern Norway. Approximately 500 forecasts from July to August 2009 have been used as the data basis.

The method shows an improvement of verification scores in relation to grid scale verification (forecasted point values against observations). For all percentiles and all threshold values we see a strong improvement if we increase the neighbourhood area. The model reaches an acceptable score for scales of the order of 50-100 km for the 90-95 % percentile. In order to reach the limit for the 99 % percentile we must increase the neighbourhood area to 230 km.

The method can also be used to calculate the probability for an event to occur within a geographical area, and some examples of this are given.

Innhold

FORORD	2
ABSTRACT.....	3
INNHold	4
1. INNLEDNING	6
2. VERIFIKASJONSMETODER	12
2.1 TRADISJONELLE METODER	14
2.2 NYE METODER – ROMLIG VERIFIKASJON.....	20
2.2.1 <i>Objektbaserte metoder</i>	21
2.2.2 <i>Feltbaserte metoder</i>	24
2.2.3 <i>Nabobaserte (fuzzy) metoder</i>	25
2.2.4 <i>Skalanedbrytende metoder</i>	28
2.3 VALG AV METODE FOR VERIFIKASJON	29
3. METODE	31
3.1 KORT BESKRIVELSE AV ROBERTS OG LEAN (2008)	31
3.1.1 <i>Steg 1: Generere fraksjonene</i>	31
3.1.2 <i>Steg 2: “Fraction Skill Score” FSS</i>	33
3.2 ENKLE TESTER.....	36
4. DATA.....	41
4.1 MODELL	41
4.2 RADARDATA	41
4.3 DATAGRUNNLAG – JULI OG AUGUST 2009.....	43
4.4 KORRIGERING AV DATAENE	44

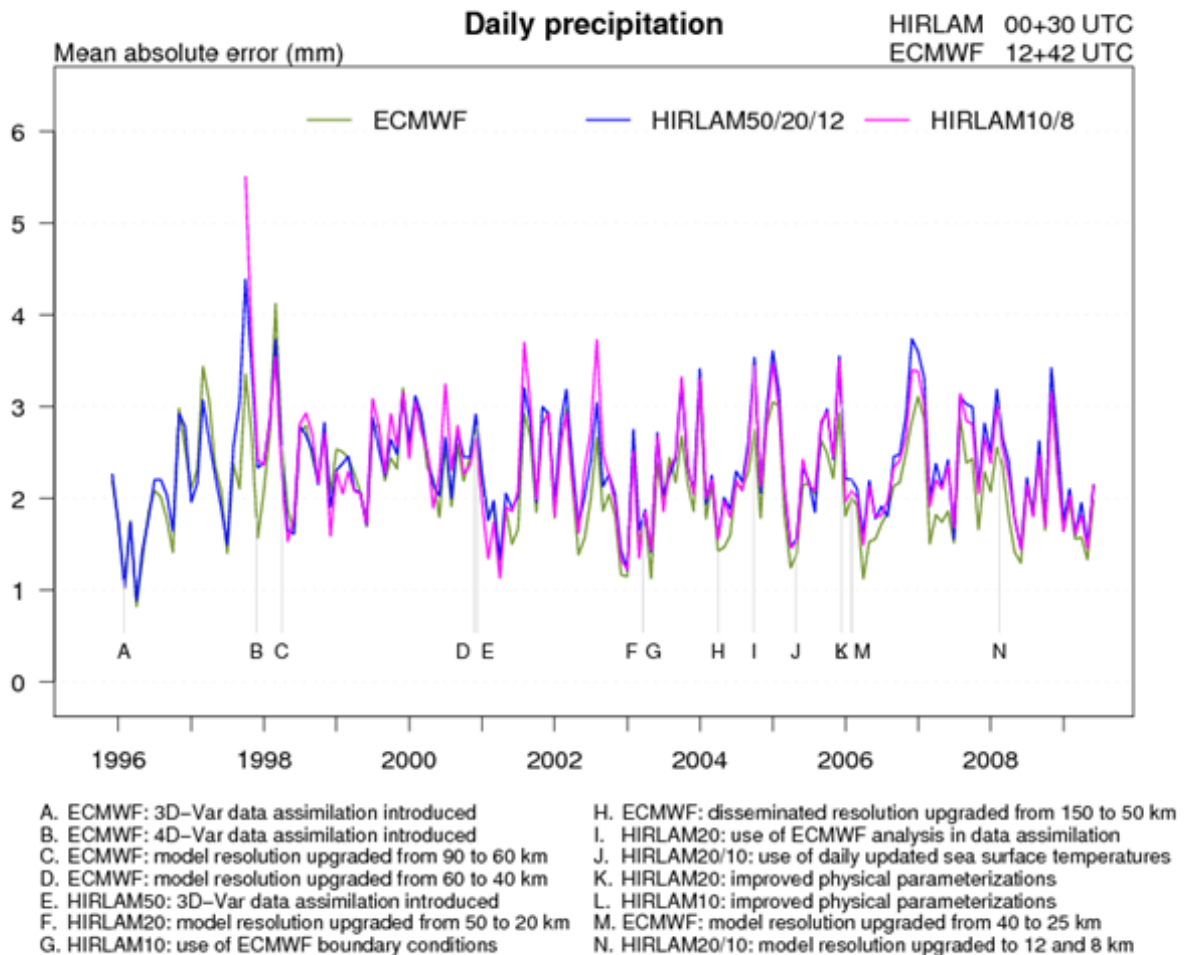
5.	RESULTATER OG DISKUSJON	49
5.1	ENKELTVARSLER	49
5.2	GJENNOMSNIITT JULI OG AUGUST.....	55
5.3	LAV VÅTANDEL.....	59
5.4	PREDIKTABILITET.....	61
5.5	STORT MOT LITE OMRÅDE	64
5.6	SAMMENLIGNING HIRLAM4 MOT UM4.....	70
6.	METODEN SOM SANNSYNLIGHETSVARSEL	73
7.	OPPSUMMERING OG KONKLUSJON	80
	APPENDIKS:	84
	FEILKILDER RADARDATA.....	84
	KILDELISTE	87

1. Innledning

Nedbør er et av de elementene i værvarsling som er av størst interesse for brukerne. Å kunne si noe om hvor og hvor mye nedbør som kommer har stor nytteverdi for beslutninger som skal tas, enten det er av samfunnsnyttig karakter eller til privat bruk. Avanserte numeriske modeller brukes i dag rutinemessig for å varsle været frem i tid. Regionale værmodellene har de siste 10 årene fått stadig finere oppløsning. For eksempel har HIRLAM gått fra omkring 50 km til 10 km (4 km oppløsning for mindre områder) i horisontal avstand mellom gridrutene (beregningsspunktene) og fra 31 til 60 vertikale lag. Den globale modellen ECMWF har gått fra 90 km til 25 km i horisontal avstand, 31 til 91 i vertikale lag, i samme tidsrom. Informasjon om modellene kan finnes på met.no sin hjemmeside (http://fou.oslo.dnmi.no/seksjoner/meteorologi/opr_log.html) og hjemmesiden til ECMWF (http://www.ecmwf.int/products/data/operational_system/evolution). Disse finskalamodellene har potensial til å produsere mer ”riktige” varsler enn tilsvarende modeller med grovere oppløsning; finskalamodeller kan blant annet gi en mer nøyaktig beskrivelse av værphenomener slik som intense konvektive nedbørssystemer. Mye av nedbøren som nødvendigvis er parametrisert, dvs. implisitt representert, i de grovere oppløste modellene er dermed eksplisitt representert i finskalamodellene. I tillegg kan man pga den kortere avstanden mellom gridpunktene beskrive topografien bedre. På tross av dette, viser flere verifikasjonsstudier der man sammenligner for eksempel modellert og observert nedbør at varslingsfeilen fortsatt er betydelig på de minste skalaene (Zepeda-Arce et al., 2000, Casati et al., 2004). Dette skyldes at feil/usikkerheter på de minste skalaene vokser raskest og det vil derfor alltid være en grense mellom hvilke skalaer det er mulig å varsle og de som er tilfeldige. Det medfører at finskalamodeller har kortere varselbarhet (”prediktabilitet”) enn modeller med grovere skala. Man må derfor være forsiktig med hvordan man tolker, verifiserer, samt presenterer varslene (Done et al., 2004).

For nedbør bør forbedringen i de numeriske prognosene som skyldes høyere oppløsning, hovedsakelig komme fra modellens beskrivelse av nedbørsprossene, modellens initialverdier gjennom bedre observasjoner og dataassimilasjonsmetoder samt mer representative topografiske data. Mye er gjort på dette området de siste 10 årene, men hvorfor ser vi da ikke dette igjen i verifikasjonsresultatene? Figur 1.1 viser ”Mean Absolute Error” (MAE) fra 1996 til nå, med avmerking for når nye modellimplementeringer er satt i

drift. Det er ikke noen tydelig tendens til forbedringer over tid selv om modellene gradvis har fått bedre oppløsning. Vi ser også at modellen med grovest oppløsning, ECMWF, ofte skårer best. Har ikke forbedringene som er gjort i modellene ført til bedre varsler?

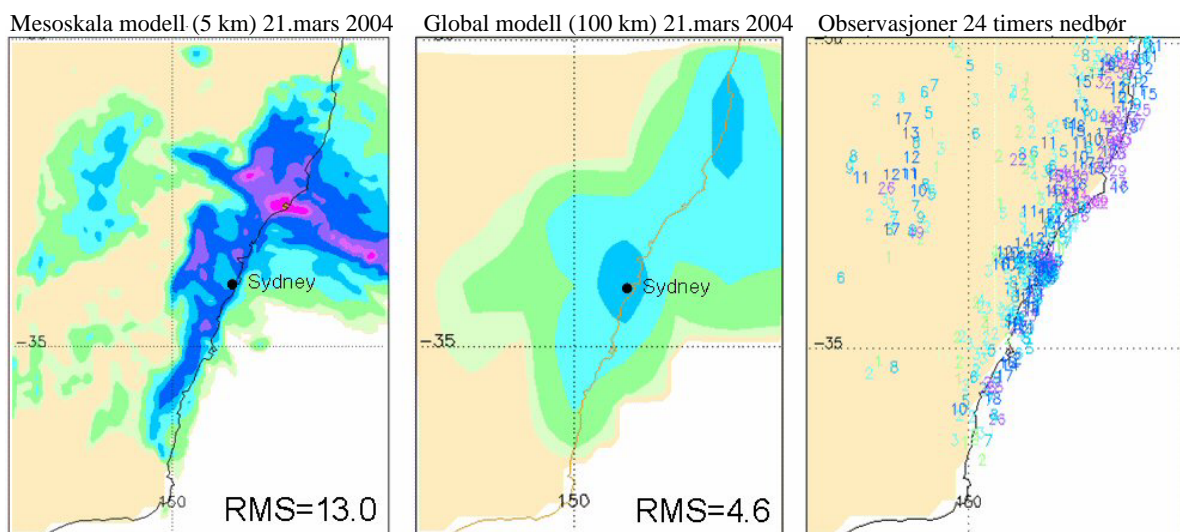


Figur 1.1: Verifikasjon av 24-timers nedbør for Norge ved bruk av MAE fra 1996 til 2009 for tre forskjellige modeller ECMWF og to HIRLAM modeller med forskjellig skala. Implementering av nye versjoner og finere oppløsning er forklart under figuren (Bremnes og Homleid, 2009).

Bruk av nyere verifikasjonsmetoder viser at det er stor sannsynlighet for at finskalerte modeller varsler bedre enn de med grovere oppløsning (Lean et al., 2008, Mass et al, 2002). Men de viser også hvor viktig det er å se på et område rundt punktet av interesse for å få med variansen i nedbøren. Nedbørfeltet kan ha store forskjeller i både utbredelse og mengde

over korte avstander. Et gridpunkt kan ha verdien null mens nabogridpunktet kan ha mange mm nedbør.

Individuelle småskala fenomener, for eksempel byger, har som nevnt kort prediktabilitet og er derfor ikke alltid varslet på eksakt riktig sted til rett tid. For eksempel viser simuleringer av konvektive stormer at de kan være noen timer feil eller noen kilometer vekk fra den observerte hendelsen (Bernardet et al., 2000). Imidlertid kan ofte strukturen og intensiteten til nedbørssystemet være godt varslet og plassert i omtrent rett område, men til feil tid. Selve varslet kan være godt, mens detaljene ikke er helt korrekte. Dette kan vi lett vurdere subjektivt for hvert enkelt tilfelle. Men de tradisjonelle verifikasjonsmetodene klarer ikke å få med variansen på de mindre skalaene, så vi trenger nye og mer egnede verifikasjonsmetoder for å kunne gi en objektiv vurdering av kvaliteten på prognosene der en tar hensyn til at modellene er blitt mer finskalerte. For å illustrere dette er det i figur 1.2 vist nedbørvarslet fra en mesoskala og en global modell og observert 24 timers nedbør for samme tidspunkt. Hvilket av de to varslene ville man helst brukt?

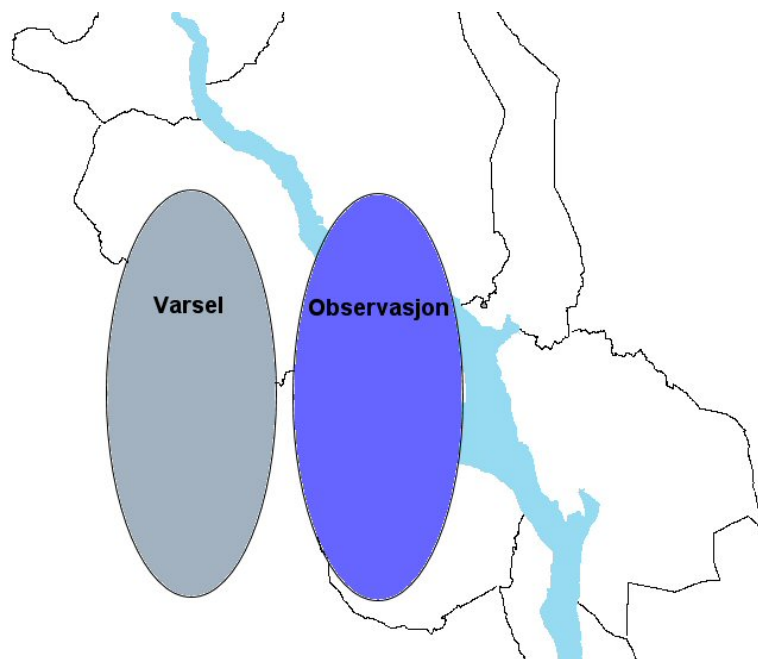


Figur 1.2: Eksempel på nedbørfordeling fra en mesoskala modell, en global modell og tilsvarende observert nedbør for samme tidspunkt. (Ebert, 2009).

”Root Mean Square” (RMS) feilen for begge varslene står på figuren. Et lavt RMS er bedre enn et høyt skår. Intuitivt ville man nok velge å bruke varslet fra den mesoskala modellen

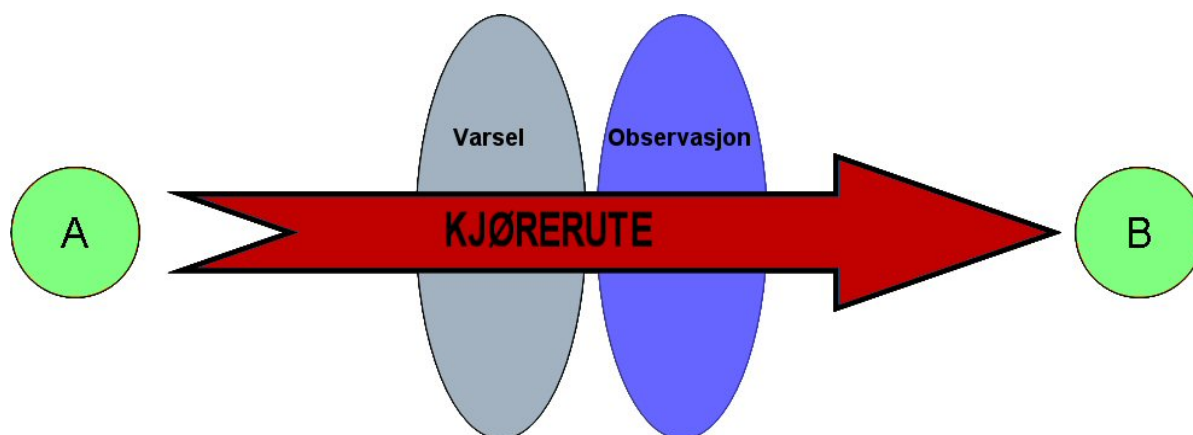
som ved første øyekast gir det mest nøyaktige varslet. Men dette varslet har også størst RMS dvs. størst feil, mye større enn den globale modellen. Sannsynligvis er den finskalerte modellen bedre totalt sett, men dersom man verifiserer gridpunkt for gridpunkt kan man få slike utslag. Dette viser hvor viktig det er å finne verifikasjonsmetoder som på riktig måte kan gjengi hvor godt varslet egentlig er. Dette kan gjøres ved å ta hensyn til at finskalerte modeller kan plassere nedbøren noe feil, og at den ikke ”smører” nedbøren utover på samme måte som en global modell.

Med den høyere detaljeringsgraden i varslene fra finskalamodeller øker også mulighetene for å kunne varsle været lokalt. Et eksempel på dette er yr.no. Brukerne får været akkurat der de er til et gitt tidspunkt. Dette representerer et paradigmeskifte i forhold til de tidligere grove tekstvarslene som beskrev været over et stort område og over en tidsperiode på flere timer. Men med dette paradigmeskiftet klarer vi nå å formidle til brukerne den underliggende usikkerheten som ligger i punktvarslene pga avtagende prediktabilitet ved høyere oppløsning? Spesielt på parametere som nedbør er dette vanskelig. Ikke bare fordi modellene varierer over små områder, men også fordi forskjellige brukere tolker et varsel forskjellig, avhengig av hva varslet skal brukes til. Dersom vi har en situasjon som nedenfor (figur 1.3) vil en som har ansvaret for vannmagasinene i elven si at dette var et dårlig varsel.



Figur 1.3: Illustrasjon av varslet og observert nedbør og viktigheten for brukeren.

Mens en som skal fly eller kjøre bil fra A til B (figur 1.4) ikke vil oppfatte varslet på samme måten, men tolke det som et relativt godt varsel. Den vil passere nedbørområdet, kanskje ikke til akkurat riktig tid eller sted, men dette var kanskje ikke så viktig.



Figur 1.4: Illustrasjon av varslet og observert nedbør og viktigheten for brukeren.

Dette viser hvor forskjellig vi kan oppfatte et varsel og at forskjellige brukere har forskjellig oppfatning av hva et godt varsel er, ut i fra hva varslet skal brukes til. Derfor er det viktig å identifisere og evaluere egenskaper ved varslene som er meningsfulle for forskjellige brukere. Altså vil ulike brukere kreve ulike verifikasjonsmetoder eller mål som vektlegger ulike sider ved et varsel. Hvor stor feilplassering er vanlig? Hvor gode er varslene på intensiteten av nedbøren? Innenfor hvor stort område kan man forvente byger? osv.

Gode verifikasjonsmetoder kan gi oss denne kunnskapen. Men tradisjonelle verifikasjonsmetoder "straffer" finskala forskjeller som man ikke finner i modeller med grovere oppløsning. Tradisjonelle metoder gir derfor ikke et godt mål på den merkbare fordelene finskalamodeller har. Derfor er nye, mer egnede metoder utviklet for spesielt finskalamodeller. De forskjellige metodene fokuserer på forskjellige egenskaper ved modellene slik som type nedbør (kontinuerlig nedbør kontra konvektiv nedbør), hvilke type feil man er interessert i (f.eks. forflytningsfeil eller intensitetsfeil), gridskala på modellen og hvilket område man verifiserer over.

Denne oppgaven vil vi gi en kort gjennomgang av de viktigste nye verifikasjonsmetodene. Videre vil en av disse nyere metodene, først beskrevet i Roberts og Lean (2008), videre i oppgaven referert til som RL08, bli anvendt på varsler fra met.no sin operasjonelle modell UM4 der 4 står for den horisontale gridavstanden i km. RL08 er beregnet brukt på nedbørsvarsler fra finskalamodeller og tar hensyn til den romlige skalaen til parameteren som skal varsles. Den inneholder også en mulighet for biaskorreksjon. Man vil ved å bruke en slik såkalt "fuzzy" eller naboområde verifikasjonsmetode, som gir verdi for varsler som er nærme i rom, kunne bedømme mangel på prediktabilitet på gridskalaen og dermed finne usikkerheten som er i varsler og observasjoner. Observasjonsdataene er hentet fra norske radarer. Det vil bli lagt vekt på å teste metoden mot radardata da 1) verifikasjon mot radarene ikke har vært gjort systematisk for norske forhold, spesielt ikke for finskalamodeller med gridavstand tilnærmet den man finner i radarobservasjonene, 2) det er sparsomt med tradisjonelle nedbørstasjoner over våre områder og for å få god romlig observasjonsdekning er radar et bedre alternativ, og 3) denne metoden kan være et godt alternativ til mer tradisjonelle verifikasjonsmetoder. Grunnen til å velge denne metoden er at den egner seg godt for konvektiv nedbør og for å se hvilken skala man må opp i for å få akseptable skår for nedbør. På denne måten kan vi definere en optimal skala for den aktuelle modellen vi verifiserer når det gjelder nedbør. For andre parametere vil man kunne få andre resultater. Indirekte gir den også et mål på lokaliseringsfeil.

For å kunne varsle usikkerhet benytter man typisk sannsynlighetsvarsler i tillegg til de deterministiske varslene. Siden RL08 også kan gi en sannsynlighetsvurdering av nedbør i finskalamodeller, vil vi kort undersøke denne muligheten.

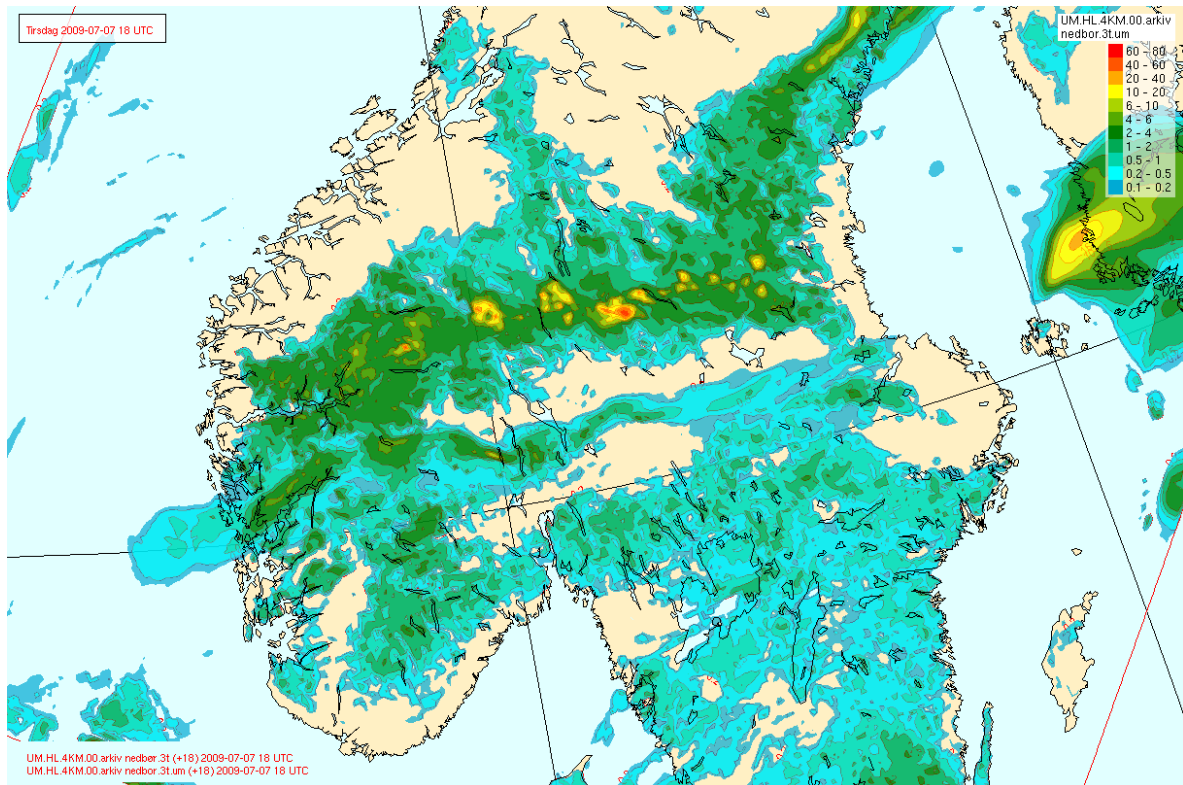
Kapittel 2 gir en kort litteraturgjennomgang av de viktigste nye ikke-tradisjonelle verifikasjonsmetodene, deres hovedpoenger, likheter og ulikheter. For fullstendighet er tradisjonelle metoder også beskrevet. I kapittel 3 forklares selve metoden, RL08, som blir brukt i denne oppgaven. Kapittel 4 tar for seg modellen og dataene som benyttes. I kapittel 5 testes metoden ut på norske forhold og kapittel 6 ser på muligheten til å bruke metoden som sannsynlighetsvarsel. Oppsummering og konklusjon følger i kapittel 7.

2. Verifikasjonsmetoder

Verifikasjon er prosessen med å sammenligne varsler med relevante observasjoner, og er et mål på varslens riktighet. Hensikten med å verifisere kan være administrative, vitenskaplige eller økonomiske. Den administrative delen går på å måle varslets prestasjon, om varslet har blitt bedre. Dette kan være viktig for de som skal betale og støtte utvikling av modellene og bruk av disse. Den vitenskaplige delen er viktig for å kunne korrigere modellene og forbedre dem. Det økonomiske formålet er å kunne ta bedre avgjørelser som sparer samfunnet for penger. Andre grunner til å verifisere kan være å hjelpe brukerne, for eksempel meteorologene, til å forstå og diagnostisere modellenes styrker og svakheter bedre og hjelpe i prosessen med å velge den "beste" modellen til bestemte formål. Verifikasjon av for eksempel nedbør kan også brukes for å finne systematiske feil som man lett kan korrigere for i ettertid, men usystematiske feil er det vanskeligere å korrigere for.

Av alle værelementer som blir varslet på daglig basis er det kanskje nedbør som er av størst interesse for publikum. Både tid, sted og mengde er viktig informasjon. Nedbør varierer mye i tid og rom og det er derfor en utfordring å varsle nedbør korrekt, som beskrevet i innledningen. Et eksempel er vist i figur 2.1 Vi ser av figuren at nedbøren kan variere mye fra sted til sted. Et gridpunkt kan ha verdien null mens et annet bare noen mil unna kan ha høye verdier. For eksempel ser vi at Lillehammer ikke har nedbør, mens litt lenger nord i Rondane har enkelte steder verdier på 20-40 mm. Nedbør er derfor også vanskelig å verifisere pga diskontinuiteten i tid og rom og mange områder med nullverdier. Man trenger derfor verifikasjonsmetoder som tar høyde for disse spesielle egenskapene ved nedbør. For 10-15 år siden ble det for det meste brukt *tradisjonelle verifikasjonsmetoder*, mens det de senere årene har vokst frem mange *nye metoder* som i større grad tar høyde for de spesielle egenskapene ved nedbør og nye typer nedbørmålinger. Tidligere hadde man stort sett bare manuelle, lokale observasjoner, å sammenligne med, mens man de senere årene også har fått tilgang til nye typer nedbørsobservasjoner fra bl.a. radar og satellitt, som gir et større omfang av observasjoner å verifisere mot.

De nye metodene for verifikasjon er motivert ut i fra ønske om å lage verifikasjonsresultater som er tilpasset spesielle brukere og brukergrupper, og for å ta høyde for utviklingen til modeller med høy oppløsning.



Figur 2.1: Varslet nedbør fra UM4 7.juli 2009. Bilde viser 3 timers akkumulert nedbør 18 UTC. Modellkjøringen er startet 00 UTC.

Disse nye verifikasjonsmetodene og tradisjonell verifikasjon har forskjellig filosofi. Nyere metoder bygger på muligheten til å forutsi et system med lokalisering og intensitet nær det observerte, så lenge varslet er ”nær nok” teller det som et godt varsel. Tradisjonelle verifikasjonsmetoder sammenligner varslet og observert nedbør gridpunkt for gridpunkt. Dersom varslet nedbør avviker fra observert nedbør, resulterer denne verifikasjonsmetoden i dårligere skår. Verifikasjon over store områder kan samlet gi gode resultat hvis nedbøren er overestimert i noen områder og underestimert i andre, derfor er det viktig å utføre mer lokal verifikasjon for å få frem variasjonene. Med det blotte øye ser man ofte at modellen varsler nedbøren ganske godt, men på feil sted. Allikevel kan varslet inneholde viktig informasjon slik som intensitet.

Tidligere studier av kvantitativ nedbørvarsling har vist at det er viktig å skille mellom de forskjellige feilkildene (Browning et al., 1982). Mange av de senere årenes verifikasjonsteknikker skiller mellom type feil, posisjonsfeil og mengdefeil, og behandler disse separat. Tabell 2.1 viser en oversikt over hovedgruppene av verifikasjonsmetoder og

hvilke type feil de kan si noe om. Tradisjonell verifikasjon og noen av de nyere verifikasjonsmetodene vil bli nærmere beskrevet i resten av kapitlet og til slutt begrunnes valget av metoden som er brukt videre i denne oppgaven.

<i>Kjennetegn</i>	<i>Tradisjonell</i>	<i>Objektbasert</i>	<i>Feltbasert</i>	<i>Skala- nedbrytende</i>	<i>Nabo- område</i>
<i>Hvor god modellen er på forskjellig skala</i>	Indirekte	Indirekte	Nei	Ja	Ja
<i>Lokaliseringsfeil</i>	Nei	Ja	Ja	Indirekte	Indirekte
<i>Intensitetsfeil</i>	Ja	Ja	Ja	Ja	Ja
<i>Strukturfeil</i>	Nei	Ja	Ja	Nei	Nei
<i>Treff, miss, falsk alarm.</i>	Ja	Ja	Ja	Indirekte	Ja

Tabell 2.1: Oversikt over hovedgruppene av verifikasjonsmetoder og hvilke type feil de kan si noe om.

2.1 Tradisjonelle metoder

Tradisjonell nedbørverifikasjon sammenligner varslet og observert nedbør gridpunkt for gridpunkt for et gitt tidspunkt eller periode. Dersom varslet nedbør avviker fra observert nedbør resulterer det i dårlig skår. I tradisjonell verifikasjon sammenligner man statistikk og utvalgte størrelser og fordelinger. Disse fremstilles ofte i spredningsplott, box-plott, qq-plott m.m. De statistiske variablene kan være "Root Mean Square Error" (RMSE), "Correlation" (COR) og "Mean Absolute Error" (MAE). Svakheterne ved disse kontinuerlige variablene er at de er sensitive til fordelinger som ikke er normalfordelt, store verdier, naturlig varians, diskontinuitet og bratte gradienter. De tar heller ikke hensyn til klustering/romlige fenomener. Vanlige kontinuerlige verifikasjonsskår er vist i tabell 2.2.

Fordi nedbør ikke er normalfordelt og kan ha svært høye verdier, er ikke de kontinuerlige verifikasjonsskårene best egnet for verifisering av nedbør.

Navn	Forkortelse	Formel	Intervall	Optimalt skår
Mean Error	ME	$\frac{1}{n} \sum_{i=1}^n (R_i - M_i)$	$-\infty$ til ∞	0
Mean Absolute Error	MAE	$\frac{1}{n} \sum_{i=1}^n R_i - M_i $	0 til ∞	0
Standard Deviation of Error	SDE	$\left(\frac{1}{n} \sum_{i=1}^n (R_i - M_i - ME)^2 \right)^{1/2}$	0 til ∞	0
Root mean Square Error	RMSE	$\left(\frac{1}{n} \sum_{i=1}^n (R_i - M_i)^2 \right)^{1/2}$	0 til ∞	0
Correlation	COR	$\frac{\frac{1}{n} \sum_{i=1}^n (R_i - \bar{R})(M_i - \bar{M})}{SD(R)SD(M)}$	-1 til 1	1

Tabell 2.2: Kontinuerlige verifikasjonsskår. R er i denne oppgaven radarfeltets nedbørverdi i et gridpunkt i, M er modellens nedbørverdi i samme gridpunkt. (Bremnes og Homleid, 2009)

Kategoriske variable og metoder tar bedre hensyn til slike variasjoner i nedbørfordelingen og brukes derfor ofte. Man lager en kontingenstabell ved å kategorisere nedbøren i en tabell ut i fra om den overstiger en viss terskelverdi eller ikke. Det enkleste er en 2x2 kontingenstabell som vist i tabell 2.3 (f.eks. 24 timers nedbør mindre eller lik 1 mm mot større enn 1 mm).

	Observert \leq 1mm	Observert $>$ 1mm
Varsel \leq 1 mm	TREFF / a	MISS / b
Varsel $>$ 1 mm	FALSK ALARM / c	TREFF / d

Tabell 2.3: 2x2 kontingenstabell som plasserer et varsel i bokser ettersom de oppfyller en terskelverdi.

Her sammenlignes gridpunkt for gridpunkt og man setter skår etter prinsippet treff, miss, falsk alarm. Fordelen med denne metoden er at man kan verifisere for forskjellige

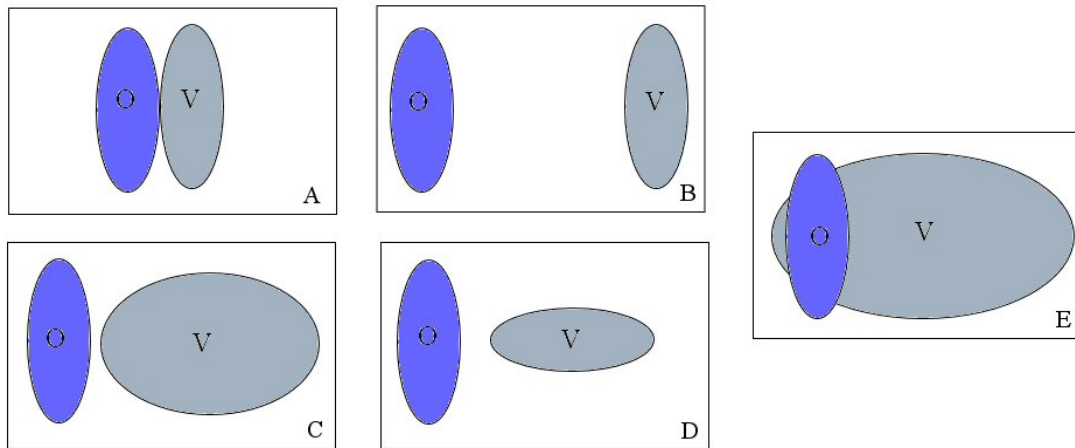
intensiteter. Men den har også svakheter bl.a. ved at den er sensitiv til bias (fremmer overvarsling), straffer ekstremverdier og glatter feltet. Kategoriske verifikasjonsskår er vist i tabell 2.4.

Navn	Forkortelse	Formel	Intervall	Optimalt skår
Hit rate	HR	$\frac{a}{a+c}$	0 til 1	1
False alarm rate	F	$\frac{b}{b+d}$	0 til 1	0
False alarm ratio	FAR	$\frac{b}{a+b}$	0 til 1	0
Equitable threat score	ETS	$\frac{a-ar}{a+b+c-ar}$	-1/3 til 1	1 (0= null skår)
Hanssen-Kuipers skill score	KSS	$HR - F$	-1 til 1	1 (0= null skår)

Tabell 2.4: Kategoriske verifikasjonsskår for en kontingenstabell. Se tabell 2.3 for forklaring av a, b, c og d. I formelen ETS er $ar=(a+b)(a+c)/(a+b+c+d)$. (Bremnes og Homleid, 2009)

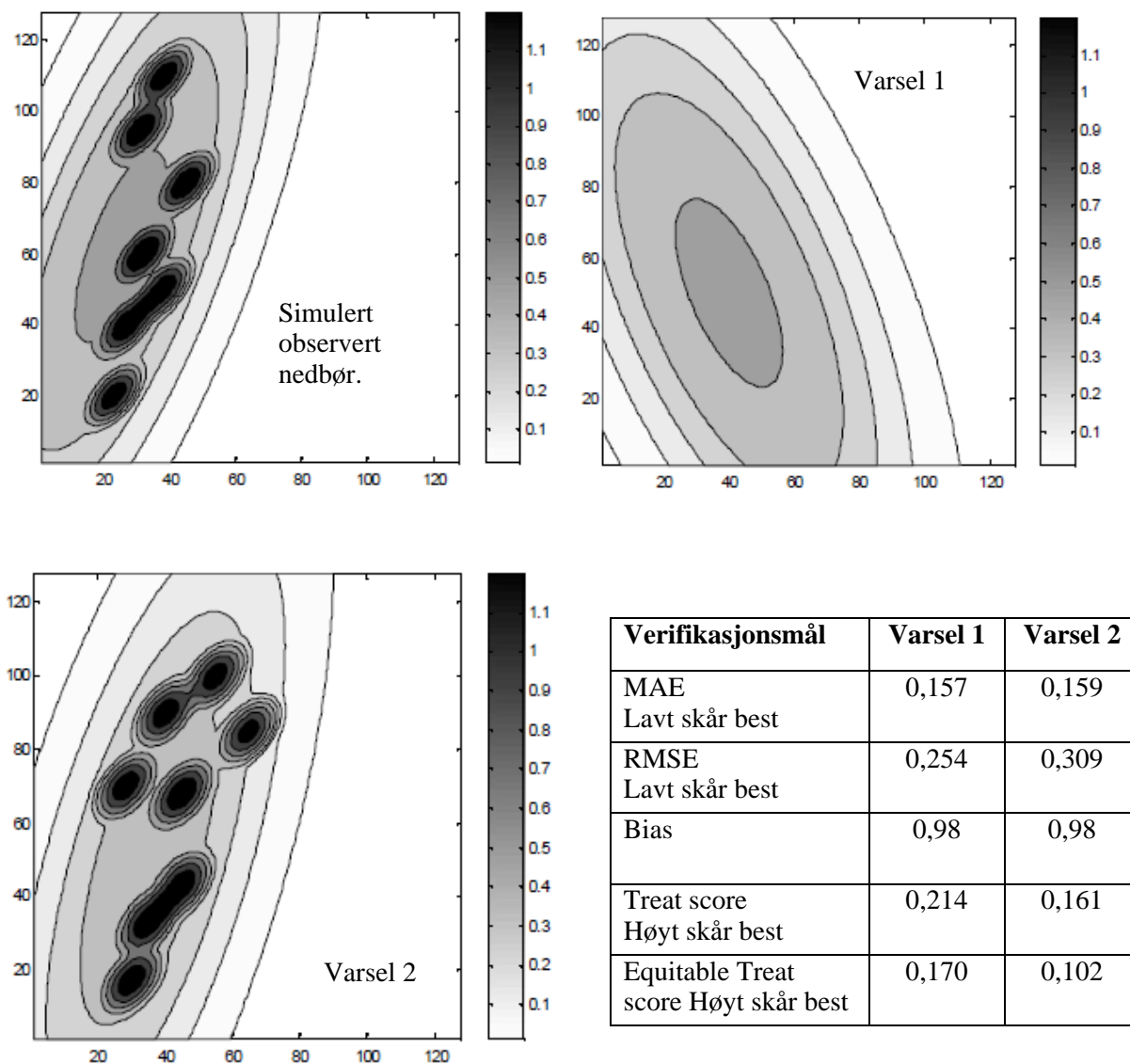
Feilplassering blir imidlertid hardt "straffet" av de tradisjonelle kategoriske skårene og RMSE. Et varsel som plasserer nedbøren feil vil få like dårlig skår som et varsel som ikke varsler hendelsen i det hele tatt, uavhengig av hvor stor feilplasseringen er.

Et klassisk eksempel på tradisjonell verifikasjon og hvor mye den straffer varsler som er feil plassert, selv om de er riktig både i form, intensitet og tid, er vist i figur 2.2. Her vil eksemplene i A, B, C og D alle gi null i skår mellom varsel og observasjon mens E gir en positiv skår. Tradisjonelle verifikasjonsmetoder forteller oss ikke hva som er feil med varslene eller hva som er riktig. Som figuren illustrerer er det ikke intuitivt at E er det "beste" varsel selv om dette er det eneste med skår større enn null.



Figur 2.2: Illustrasjon på nedbørfelt fra modellvarsel (V) og observert felt (O). Figur A, B, C og D vil få null treff varsel/observasjon, mens figur E vil få treff > 0 , modifisert fra Davis et al. (2006I).

Et annet eksempel er vist i Baldwin et al.(2001), figur 2.3. Tradisjonelle skår ble testet mot et simulert nedbørområde på to modeller, en finskalamodell og en med grovere oppløsning. Subjektiv verifikasjon sier oss at varsel 2 er bedre enn varsel 1 for det observerte nedbørområdet. Men som vi ser i tabellen nederst til høyre i figuren er alle de tradisjonelle verifikasjonsskårene bedre for varsel 1 enn for varsel 2. Begge varslene har samme bias. De tradisjonelle skårene "straffer" her den finskalerte modellen selv om vi ved subjektiv verifikasjon synes dette er det beste varslet.



Figur 2.3: Eksempel på simulert nedbørfelt og varsler fra to modeller, en finskala modell og en med grovere oppløsning. Nederst til høyre tabell over forskjellige tradisjonelle verifikasjonsskår fra de to varslene. Hentet fra Baldwin et al. (2001).

I figur 2.4 ser vi de kategoriske og kontinuerlige skårene i gjennomsnitt for hele Norge for månedene juli og august 2009. Dette er den samme perioden som RL08 er testet mot i resten av oppgaven, og vi ser hvordan modeller med forskjellig oppløsning verifiserer i forhold til hverandre. Verifikasjonen er utført mot norske synopstasjoner.

Number of stations with best score

85 Nor. stations		00+30 UTC							
	ME		SDE		MAE		COR		
EC H8	40	45	61	24	57	28	66	19	
EC H4	65	20	62	23	69	16	69	16	
EC UM4	47	38	61	24	55	30	59	26	
H8 EC	45	40	24	61	28	57	19	66	
H8 H4	66	19	43	42	65	20	39	46	
H8 UM4	48	37	50	35	45	40	35	50	
H4 EC	20	65	23	62	16	69	16	69	
H4 H8	19	66	42	43	20	65	46	39	
H4 UM4	28	57	43	42	18	67	32	53	
UM4 EC	38	47	24	61	30	55	26	59	
UM4 H8	37	48	35	50	40	45	50	35	
UM4 H4	57	28	42	43	67	18	53	32	

Daily precipitation ≤ 0.1 mm

5227 cases		00+30 UTC				
	HR	F	FAR	ETS	KSS	
EC	0.37	0.04	0.12	0.22	0.33	
H12	0.45	0.05	0.14	0.26	0.39	
H8	0.43	0.05	0.14	0.25	0.38	
H4	0.25	0.02	0.09	0.15	0.23	
UM4	0.5	0.07	0.15	0.29	0.43	

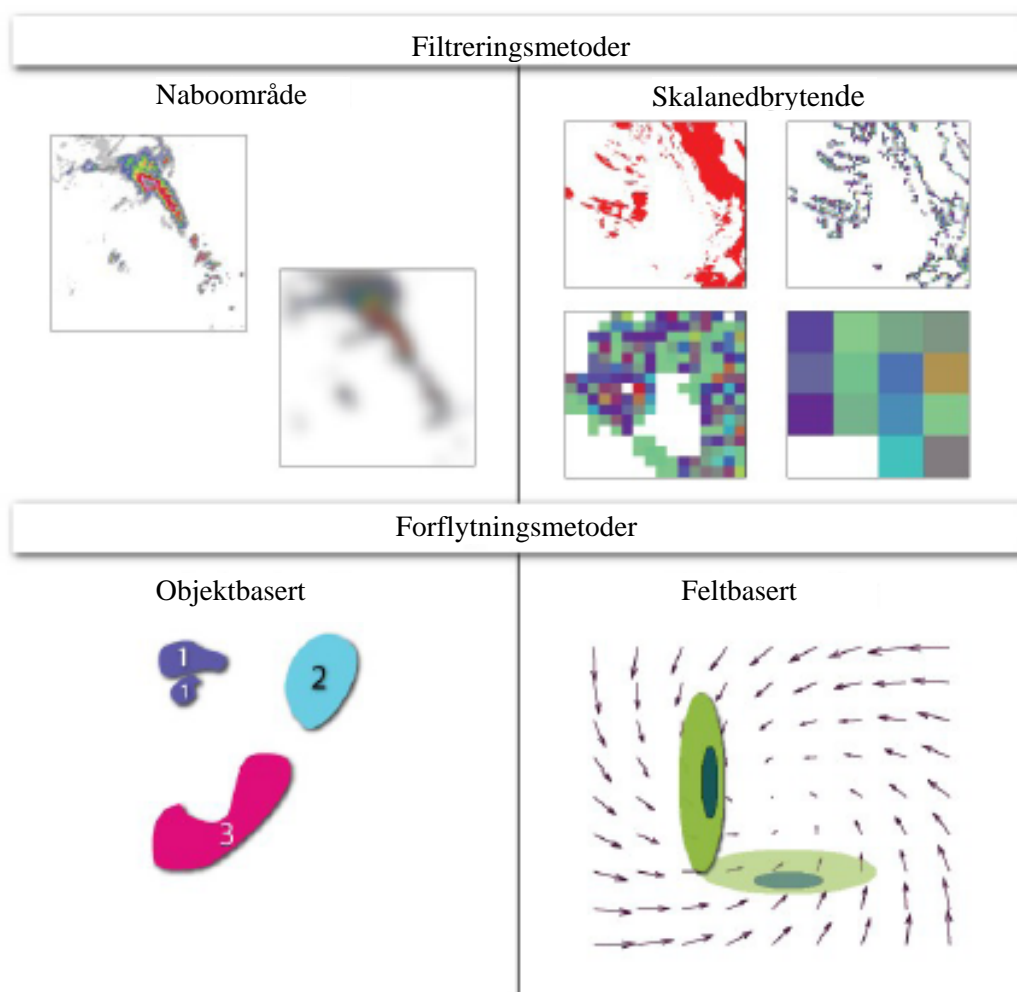
Figur 2.4: Kategoriske og kontinuerlige verifikasjonsskår gjennomsnittlig for Norge for juli og august 2009. Verifikasjonen er gjort mot norske synopstasjoner. De gule merkene i tabellene viser hvilken modell som kommer best ut ved de forskjellige skårene (John Bjørnar Bremnes).

I disse beregningene verifiserer UM4 dårligere i forhold til modellene med grovere oppløsning, H8 og ECMWF, når en ser på de kontinuerlige skårene. ECMWF har bedre skår enn UM4 på alle de 5 kontinuerlige skårene (nederste tabell i figur 2.4). UM4 kommer bedre ut i forhold til ECMWF (som har en oppløsning på 25 km) ved bruk av de kategoriske skårene, figur 2.4 topp, men fortsatt skårer ECMWF bedre på 2 av de 5 skårene. Dette viser igjen at finskalamodeller ikke alltid gir bedre skår enn modeller med grovere oppløsning ved bruk av tradisjonelle verifikasjonsmetoder.

For de kontinuerlige skårene verifiserer UM4 bedre på de fleste skårene sammenlignet med HIRLAM4, (H4), som har samme gridavstand. Også for de kategoriske skårene skårer UM4 bedre enn HIRLAM4 på 3 av 5 skår. I kapittel 5.6 er UM4 testet mot HIRLAM4 ved bruk av RL08.

2.2 Nye metoder – romlig verifikasjon

De nye verifikasjonsmetodene kan grovt sett deles inn i to overordnede kategorier: *Forflytningsmetoder* og *filtreringsmetoder*. Forflytningsmetodene kan igjen deles inn i *feltbaserte metoder* og *objektbaserte metoder*. Filtreringsmetodene kan deles inn i *skalanedbrytende metoder* og *naboområde(fuzzy) metoder* (Gielland et al., 2009a; Casati et al., 2008). Skjematisk oversikt over de fire hovedkategoriene er illustrert i figur 2.5.



Figur 2.5: Skjematisk oversikt over de fire hovedkategoriene av nyere verifikasjonsmetoder. Naboområde metodene og skalanedbrytende metodene kan begge ses på som filtreringsmetoder (øverste rad). De objektbaserte og feltbaserte metodene kommer under forflytningskategorien (nederste raden). Fra Gielland et al. (2009a).

Det må også bemerkes at enkelte metoder ikke passer godt inn i noen av kategoriene eller ligger i grenseland mellom flere. Felles for metodene er at de tar hensyn til usikkerhet i tid, lokalisering, strukturen i feltet og de gir oss informasjon om feil i fysiske vilkår. De gir oss informasjon som kjennetegner fenomenet vi ser på og er meningsfulle for modellens brukere.

Forflytningsmetodene prøver å tilpasse varslet til observasjonene så godt som mulig og gir så informasjon om hvor mye varslet må "tunes" for å tilpasses observasjonene (horisontal forflytning, rotasjon, skalering osv.) og kvantifiserer de forskjellige feilene for å gi et mest mulig meningsfylt skår på varslets kvalitet. Hovedforskjellen mellom *feltbaserte metoder* og *objektbaserte metoder* er at den førstnevnte analyserer hele feltet, mens den sistnevnte først identifiserer hendelser av interesse og så analyserer disse separat.

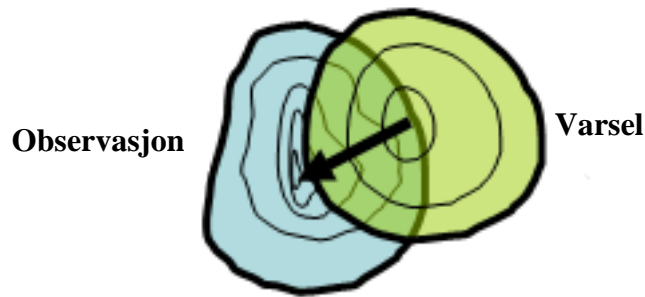
Filtreringsmetodene er som navnet tilsier, metoder som benytter romlige filtre på enten både modell og observasjoner eller kun et av feltene (og noen ganger til differansen mellom feltene). Filteret blir som regel brukt på stadig økende/grovere skalaer for å gi informasjon om på hvilken skala varslet har brukbart skår. Naboområdet metodene bruker et glattende filter mens de skalanedbrytende metodene bruker flere romlige filtre (f.eks Fourier og Wavelets) slik at de forskjellige skalaene kan ses på uavhengig av hverandre.

I det følgende vil vi kort komme inn på noen av de mest kjente metodene i hver kategori.

2.2.1 Objektbaserte metoder

Med objektbaserte metoder evaluerer man egenskaper til identifiserbare fenomener ved at man definerer ett eller flere bestemte objekter i feltet. Det er flere tilnærminger til denne metoden, og teknikken som blir brukt til å identifisere objektene i varsel og observasjon, varierer. Det som skiller de er: i) hva som kjennetegner et objekt, ii) hvorvidt objekter som ikke er kontinuerlige i rommet skal behandles som et objekt eller flere objekter, iii) hvordan de tilpasser objektene fra et felt til et annet og iv) hva slags type mål man får ut. De fleste metodene avgrensner et objekt ved å sette en terskelverdi.

Ebert og McBride (2000) gjorde et viktig studie i objektbasert verifikasjon. Deres metode, kalt "Contiguous Rain Areas" (CRA), isolerer et spesielt nedbørtilfelle ved å bruke terskelverdier for nedbøren.



Figur 2.6: Illustrasjon av objektbasert metode. Varsel flyttes til det passer best mulig med observasjonen etter bestemte kriterier. (Ebert og McBride, 2000)

Det kriteriet som blir brukt for å finne ”best fit” mellom varsel og observasjon kan imidlertid være forskjellig, for eksempel minste MSE mellom varsel og observasjon, maksimal korrelasjon, maksimal overlapp etc. Etter å ha kalkulert forskyvningen (for eksempel ved MSE) flyttes nedbørområdet til posisjonen ”null forskyvning” eller ”beste tilpasning”, så regner man ut en ny MSE, figur 2.6. Ebert og McBride (2000) målte forflytning, volum og strukturfeil mellom varsel og observasjon (hva som ligger i struktur er et definisjonsspørsmål og er ulikt i de forskjellige metodene). De så på kontinuerlige nedbørområder, og ikke konvektiv nedbør som ofte er det som er dårligst plassert og vanskeligst å varsle. I en studie med flere tilfeller fra Australia viste verifiseringen at lokalisering var den dominerende feilen (opp mot 60 %), mens volumfeilen var 7 % eller mindre i alle soner (McBride og Ebert, 2000). Å verifisere nedbør ut i fra CRA-metoden har vist seg å være effektiv når metoden utvides til også i inneholde en kategorisk verifikasjon. Kriterier kan settes for maksimum intensitetsfeil og lokaliseringsfeil. Når varslet tilfredsstillende begge kravene, er varslet ansett for å være et ”treff”. Konseptet av ”treff”, ”feil” og ”falske alarmer” kan dermed brukes på slike hendelsesvarsler. En nyere studie med denne metoden, der den er prøvd ut på de konstruerte tilfellene fra ”Spatial Forecast Verification Methods Intercomparison Project” finnes i Ebert og Gallus (2009).

En annen objektbasert metode er ”Method for Object-based Diagnostic Evaluation” (MODE), Davis et al. (2006I,II). De brukte også terskelverdi, men først etter å ha anvendt et konvoluttprinsipp hvor feltet blir glattet i rommet. Denne metoden identifiserer og verifiserer intensiteten og senteret til lokaliseringen, knyttet til par av varsel og observerte objekter. Metoden inneholder også kontingenstabeller (jfr. kapittel 2.1) basert på

lokaliseringsfeil av hendelsene, og regner ut tilhørende kategoriske skår.

Verifikasjonsskårene presenteres som en funksjon av hendelsens størrelse slik at denne objektbaserte metoden også kan forklares i et skalaorientert perspektiv. MODE tar også hensyn til tidsperspektivet.

Wernli et al. (2008) har valgt en litt annen tilnærming enn MODE og CRA. De definerer tilfeller av nedbør innenfor et relativt lite areal av interesse, men verken samkjører eller flytter objektene for å tilpasse hverandre slik som i CRA. Metoden til Wernli et al. (2008) kalles "Structure, Amplitude and Location" (SAL), og betrakter disse tre komponentene, struktur, amplitude og lokalitet, separat slik at et perfekt varsel vil ha verdien null for alle tre komponenter. Først identifiserer man objektet man ønsker å verifisere ved hjelp av en terskelverdi som er en viss prosent av maksimalverdien for nedbør i domenet man ser på. Objektet blir gridpunktene som overstiger terskelverdien i området rundt maksimalverdien. Dernest beregner man de tre komponentene hver for seg. Positive (negative) verdier av A (amplitude) indikerer overestimering (underestimering) av nedbøren.

Lokaliseringskomponenten L kombinerer informasjon om feilplassering av nedbørfeltets massemidtpunkt og gjennomsnittsfelen til de forskjellige objektene fra massemidtpunktet. Strukturkomponenten S er laget slik at man for hvert objekt konstruerer et "skalert volum" og et gjennomsnittsvolum for både varsel og observasjoner. Positive verdier oppstår dersom nedbørobjektet er for stort og/eller for glatt, og negativ dersom nedbørobjektet er for lite og/eller for ujevnt. Denne muligheten til å si noe om type feil er en av hovedkarakteristikkene for SAL-metoden.

A og S i SAL-metoden er forskjellig fra volum og struktur i CRA-metoden til Ebert og McBride (2000). En annen forskjell mellom Wernli et al. (2008) sin SAL-metode og andre objektbaserte metoder (f.eks Ebert og McBride, 2000; Davis et al., 2006) er at SAL ikke bruker en fast terskelverdi til å identifisere objektene, men en fleksibel terskelverdi som ofte kan variere mellom varslet og observert felt. Fordelen er at man da ikke mister så mange varsler. Spesifisering av en fast terskelverdi for alle varslene innenfor en gitt tidsperiode av interesse ekskluderer dårlige varsler i situasjoner der terskelverdien ikke oppnås i enten varsel ("miss") eller observert felt ("falsk alarm"). Dette kan gi en positiv bias siden bare rimelig gode varsler blir med i den statistiske beregningen. Fordelen med en fast terskelverdi kan være at verifiseringen kan fokusere på spesielle hendelser, for eksempel intense nedbørtillfeller, og ikke bli svekket av tilfeller som er av mindre interesse.

Nachamkin (2005) utviklet en metode som identifiserer alle observerte nedbørtilfeller, sentrerer dem på et felles grid og setter de så sammen. Forskjellen mellom den observerte sammensetningen og den varslede sammensetningen gir kunnskap om modellens tendens til å overvarsle, undervarsle og feilplassere nedbørtilfeller.

Styrken til de objektbaserte metodene er at de beregner varslets treff både før og etter forflytningen og deler opp feilen i komponenter (lokalisering, volum og struktur) som dermed kan identifisere den største feilkilden. Metoden kan også identifisere ”falske alarmer” og ”bommer”.

Svakheten er at varslede og observerte nedbørområder må være kontinuerlige eller i det minste relativt nære i geografisk plassering for å bli sammenlignet. Dette kan være et problem for relaterte nedbørområder som ikke er nær nok hverandre. To eller flere observerte nedbørfelter kan også bli del av ett verifiseringsområde (CRA), da det ikke er noen måte å splitte overvarslede områder inn i deler som passer separate observerte hendelser.

Objektbaserte metoder er blitt mye brukt. Men nettopp fordi man ser på enkeltobjekter, eller en samling av objekter, egner denne metoden seg best til å se på spesielle situasjoner og ikke så godt til verifikasjon over et helt område for lengre tidsrom.

2.2.2 Feltbaserte metoder

De feltbaserte metodene evaluerer skjevheter og forflytningsfeil ved hele feltet. Hoffman et al. (1995) var de første som introduserte denne verifikasjonsteknikken. De spalter varslingsfeilen i forflytningsfeil, mengdefeil og strukturfeil (det som blir igjen).

Forflytningsfeilen og mengdefeilen måles ved å bruke et såkalt ”best fit-kriterie” (f.eks. minimere MSE); man ”flytter” varslet til det passer best mulig med observasjonene og måler så feilen. Mengdefeilen får man så ved å sammenligne dette ”oversatte” varslet med observasjonen. Denne måten å verifisere på satte en milepæl i romlig verifisering ved å måle feilen direkte i fysiske enheter, som feilplassering i kilometer, og inspirerte mange videre studier slik som Du et al. 2000 og Nehr Korn et al. 2003. Alle metodene som baserer seg på Hoffman et al. (1995) sin metode utføres på hele feltet man verifiserer over. De fleste nyere feltbaserte metoder er kombinert med en skalanedbrytende tilnærming (se kapittel 2.2.4 nedenfor), og da hovedsakelig til dataassimilering og nowcasting (Casati et al., 2008).

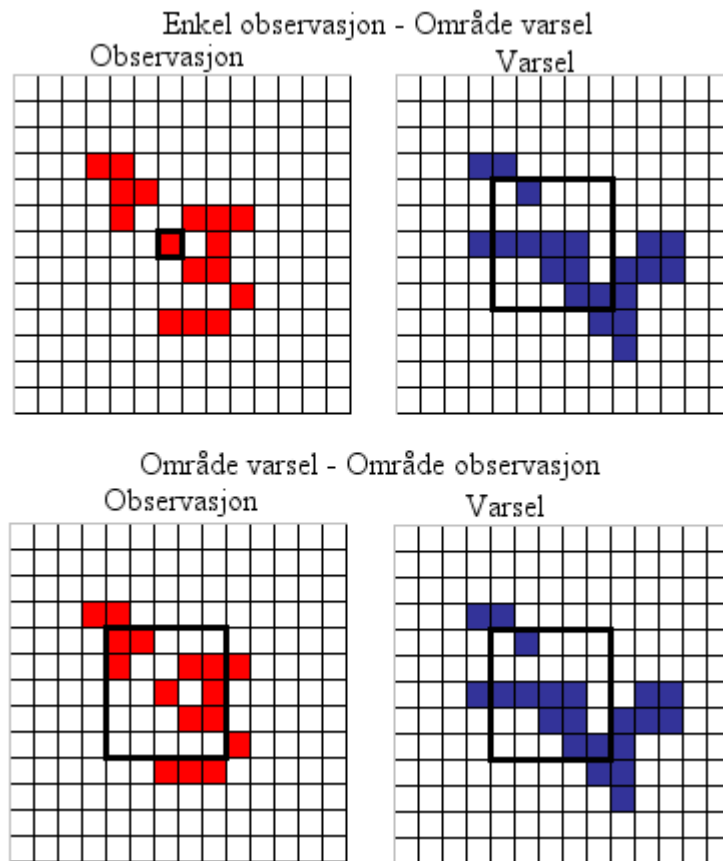
De fleste forflytningsmetodene er best egnet for å studere enkeltsituasjoner i likhet med de objektbaserte metodene. Feltbaserte metoder og objektbaserte metoder er kategorisert i samme hovedkategori da de er veldig like. Den største forskjellen mellom feltbaserte metoder og objektbaserte metoder er at feltbaserte metoder ser på hele feltet på en gang i stedet for å identifisere individuelle objekter og sammenligne disse. Både objektbaserte metoder og feltbaserte metoder egner seg godt til alle typer feil, men feltbaserte metoder egner seg lite til å si noe om på hvilken skala det verifiseres best, mens objektbaserte metoder lettere kan brukes på forskjellig skala og indirekte gi et mål på skalafeil, jamfør tabell 2.1. Styrkene og svakhetene til disse metodene er mye de samme som for de objektbaserte metodene (se kapittel 2.2.1).

2.2.3 Nabobaserte (fuzzy) metoder

Naboområde eller ”fuzzy” verifikasjonsmetoder evaluerer ”nærhet” mellom varselet og observert nedbør innenfor et tid-rom-vindu i stedet for på gridskalaen, dvs. verifikasjonen tas over gridpunkter i nærområdet, det såkalte naboområdet, i motsetning til de tradisjonelle metodene som sammenligner gridpunkt for gridpunkt. Nabobaserte metoder krever ikke eksakte treff mellom varsel og observasjon.

Man kan tilnærme seg denne metoden på to måter: en enkelt observasjon verifisert mot et område i varslet, eller et område rundt punktet man er interessert i både i varsel og observasjon (figur 2.7). I det første tilfellet får man treff dersom observasjonen har treff og ett punkt i området man har valgt i varslet også har det. I det andre tilfellet er det full skår dersom det er like mange treff i begge områdene.

Fordelen med denne type verifisering er at man skalerer opp dataene og ser på en gjennomsnittsverdi. Man konsentrerer seg om en hendelse inntreffer et eller annet sted innenfor området man ser på. Ved å variere størrelsen på naboområdet kan denne metoden i likhet med skalanedbrytende metoder (2.2.4) si noe om på hvilken skala varslet oppnår akseptabelt skår (se tabell 2.1).



Figur 2.7: Verifikasjon av enkelt observasjon mot nabo område varsel (øverst), og nabo område observasjon mot nabo område varsel (nederst).

Den eldste og kanskje mest kjente av de nabobaserte metodene er såkalt ”oppskalering eller glatting” (Zepeda-Arce et al. 2000), av varsel og observasjoner til en grovere oppløsning i både tid og rom for så å verifisere ved å bruke vanlige kategoriske eller kontinuerlige variable, se kapittel 2.1. Dette er en god verifikasjonstilnærming når man ønsker et mål på hvor godt et område av varslet passer i forhold til et område av observasjonene, for eksempel til å evaluere nedbørvarsler for hydrologiske formål. Men med denne metoden mister man mye informasjon om detaljene i de observerte og varslede nedbørfeltene og er derfor ikke alltid ideell.

Roberts og Lean (2008) og Roberts (2008) ser på hvor stor andel av gridpunktene i et gitt nabo område som overstiger en viss verdi for nedbørsintensitet, og definerer en ”Fraction Skill Score” (FSS), som varslet oppnår som funksjon av intensitet og størrelse på nabo området (også kalt skala). Jo større området blir, jo mer glatter man feltet og mer informasjon går tapt. Ved å fokusere på nedbørfordelingen gir denne metoden bedre informasjon om den ser realistisk ut sammenlignet med observasjonene. Bruken av

fraksjoner er også mer robust og tar høyde for tilfeldige feil i observasjonsgrunnlaget. FSS er sensitiv til falske alarmer og misser, spesielt for store områder, siden den verifiserer i et område rundt et gridpunkt både for observasjon og varsel. Metoden vil bli beskrevet i mer detalj i kapittel 3. Både Theis et al. (2005) og Roberts og Lean (2008) tok deterministiske varsler og gjorde dem om til sannsynlighetsvarsler. Måten Theis og Roberts og Lean utleder sannsynlighetsvarsler er veldig lik, men Theis ser også på tidsperspektivet og bruker "Brier skill score" (BSS) for å beregne sannsynlighetene.

Atger (2001) utviklet "Multi-Event Contingency Table" (MECT). MECT måler om varslet lykkes i å fortutsi minst en forekomst av den parameteren man vil verifisere i nærheten av en observasjon der observasjonen er plassert midt i naboområdet og representerer en lokasjon av interesse (øverste del av figur 2.7). Nærheten kan spesifiseres som avstand, tid, intensitet eller andre viktige aspekter i varslet. Denne metoden er mer krevende enn de foregående siden den kun ser på et punkt i observasjonene, men er allikevel ikke så streng som tradisjonelle metoder siden den kan gi noe utslag dersom varslet er i nærheten av observasjonspunktet.

Ebert (2009) sammenlignet 4 forskjellige naboombaserte metoder. Oppskalering (Zepeda-Arce et al., 2000), metoden til Roberts og Lean (2008), MECT (Atger, 2001), og "Practically perfect hindcast" (PP) metoden av Brooks et al. (1998) (ikke beskrevet i denne oppgaven), ved å bruke konstruerte nedbørs- og observasjonsfelt. Ingen av metodene ga veldig gode resultater når det gjaldt å karakterisere hovedtrekkene til nedbørområdet. Men hensikten med naboombaserte metoder er å gi bedre skår til varsler som er nærme observasjonene, ikke diagnostisere kilden til feilen. Ebert (2009) viser at naboombaserte metodene gir bedre skår på skalaer større enn selve gridskalaen, og at den optimale skalaen for hver metode varierer som en funksjon av nedbørintensiteten. Et litt overraskende resultat ble funnet når det gjaldt metoden som sammenligner et naboombasert i varslet med en enkelt observasjon (MECT). Tester på konstruerte varsler viste at varsler som var plassert feil med høy bias fikk bedre skår enn varsler med samme forflytningsfeil men uten bias for de fleste skaler og intensiteter. Dette går mot vår intuitive formening om at et varsel med null bias bør gi bedre skår enn et varsel med høy bias. Men spørsmålet er hva som menes med "bedre"? Modellutviklere ønsker å forutsi nedbør med så lite feil som mulig i lokasjon og intensitet. Fra deres synspunkt er nok et varsel uten bias klart bedre enn et varsel med bias. Men for en bruker som ønsker det beste varsel for en bestemt lokasjon behøver ikke dette være tilfelle (jamfør figur 1.3 og 1.4). MECT ga heller ikke "perfekte"

skår når metoden ble brukt på et perfekt varsel, dette pga innflytelse av gridboksene i området rundt punktet man ser på.

Den virkelige styrken til de nabobaserte metodene er å vise på hvilken skala varslet har brukbart skår for forskjellige terskelverdier. Disse metodene egner seg spesielt godt til bruk på finskalerte modeller. Og i motsetning til forflytningsmetodene egner filtreringsmetodene seg godt til å verifisere over hele områder og er derfor best egnet til operasjonell verifikasjon. I motsetning til objektbaserte metoder egner naboområde metoder seg til å verifisere nedbørvarsler som er ”rotete” og ikke inneholder veldefinerte nedbørområder.

Svakheten ved disse metodene i motsetning til forflytningsmetodene er at de ikke klart skiller på type feil, slik som forflytningsfeil og strukturfeil (tabell 2.1), og man får derfor ikke noe klart bilde av hvilken type feil som er dominerende.

2.2.4 Skalanedbrytende metoder

Skalanedbrytende metoder gir informasjon om hvordan varslet verifiserer på forskjellig skala. Casati et al. (2004) utviklet en metode kalt ”Intensity-Scale” (IS). Først rekalibreres både varsel og observasjonene for å ta bort støy (i den forbedrede utgaven av modellen beskrevet i Casati (2009) blir ikke denne rekalibreringen lenger gjort). Deretter brukes binærverdier (likt som i Roberts og Lean (2008) beskrevet nedenfor) og man lager et nytt binært bilde ved å bruke differansen av binærverdiene i modellfelt og det observerte feltet. Dette bilde blir så delt opp i nye bilder på forskjellige skaler ved bruk av ”Haar wavelets”. Til slutt beregnes Skill Score ved bruk av MSE. Skåret beregnes for hver terskelverdi og hvert bilde. Resultatene viser at MSE minker for stor skala og store terskelverdier. Det viktige er da å finne hvilken skala man må ha for å få en brukbar treff. Teknikken tillater treff å bli diagnostisert som en funksjon av både skala og intensitet. Resultater viser at reduksjon i prediktabilitet hovedsaklig kommer av feil plassering av mer intense (sjeldne) nedbørtilfeller. Denne metoden gir også informasjon om prediktabilitet for forskjellige nedbørintensiteter og romlig skala, mens Zepeda-Arce et al. (2000) og Briggs og Levine (1997), to tilsvarende metoder, bare gir prediktabilitet for romlig skala.

Skalanedbrytende metoder isolerer individuelle bølgelengder, assosiert med f. eks værphenomener på forskjellig skala (frontsoner, konvektive celler m.m.) og kan gi informasjon om varslets treff separat for hver skala. Naboområde verifikasjon, på den annen

side, separerer ikke treff direkte på hver skala, men vurderer ved hvilken oppløsning den ønskede treffprosent oppnås. Skalaen i den skalaorienterte tilnærmingen relaterer seg til skalaen på hendelsen man ser på og feilen.

Styrkene og svakhetene til de skalanedbrytende metodene er de mye samme som de nabobaserte metodene, se 2.2.3., men de nabobaserte metodene er ofte noe enklere og dermed lettere å forstå og implementere.

2.3 Valg av metode for verifikasjon

De forskjellige metodene som er nevnt i dette kapitlet vektlegger forskjellige typer feil i nedbørsvarene og man må derfor velge metode etter hva man ønsker å se på, se tabell 2.1.

Filtreringsmetodene gir *skalaen* hvor varslet oppnår en bestemt skår, men objektbaserte metoder kan også potensielt gi informasjon på forskjellig skala ved at man anvender metoden på feltet med forskjellig skala.

Lokaliseringsfeil er ganske vanlig i finskalamodeller og de objektbaserte og feltbaserte metodene gir direkte informasjon om lokaliseringsfeil. Nachamkin (2005) sin metode er for eksempel en god måte å summere den systematiske lokaliseringsfeilen for en stor andel tilfeller. Nabobaserte metoder og IS-metoden til Casati et al. (2004) gir imidlertid ikke direkte informasjon om lokaliseringsfeil.

Alle metodene, inklusiv de tradisjonelle, gir informasjon om *intensiteten* og fordelingen av områder med stor nedbørintensitet.

De fleste metodene kan si lite om *strukturen* på nedbøren. Men dette avhenger av hvordan man definerer struktur. Metodene som definerer objekter av nedbør i feltet kan bestemme geometriske effekter til disse objektene, som areal, lengde, orientering osv. Dette er egenskaper som summerer noe struktur på nedbøren. To metoder som gir slik informasjon er MODE og CRA. Dette er begge objektbaserte metoder.

Både tradisjonelle og flere av naboområde metodene gir gode svar på *treff, miss og falske alarmer*, mens andre metoder er mer sensitive til disse verdiene uten å måle dem direkte. De objektbaserte metodene er generelt gode på å gi slik informasjon, bortsett fra SAL metoden.

Alle metodene kan brukes på forskjellige *tidspunkt*, men de metodene som har potensial til å gi direkte informasjon om tidsfeil er IS (Casati et al., 2004), sammensatt metode (Nachamkin, 2005) og tradisjonelle metoder. For mange av metodene kan denne informasjonen regnes ut indirekte. For FSS og andre naboområdetmetoder kan tidsfeil implementeres hvis forskjellige tidsvinduer blir brukt. Men praktiske betraktninger, inkludert datakraft, begrenser muligheten til å utvide en del av metodene for også å gi informasjon om tidsdimensjonen.

Hvis hensikten med verifikasjonen er å finne kilden til feilen, er filtreringsmetoder ikke en passende tilnærming siden de kun identifiserer skalaen hvor feilen er tilstrekkelig liten for en viss terskelverdi. Imidlertid kan skalanedbrytende metoder isolere de skalaavhengige feilene. For varsler med godt definerte objekter vil objektbaserte metoder være effektive i å diagnostisere feil i lokalisering, størrelse, intensitet og andre parametere.

I denne oppgaven er naboområdetmetoden til Roberts og Lean (2008) valgt for videre studie. Den viktigste grunnen til å velge denne metoden er at den gir oss på hvilken skala varslet har brukbart skår for forskjellige terskelverdier og egner seg spesielt godt for å verifisere finskalamodeller. På denne måten kan vi definere en optimal skala for den aktuelle modellen vi verifiserer. Siden vi ønsker å verifisere et helt felt over en lengre periode og ikke er like interessert i å studere enkeltsituasjoner, egner denne metoden seg bedre enn for eksempel de objektbaserte metodene. RL08 egner seg godt for vær-situasjoner som er ”rotete” og ikke inneholder veldefinerte objekter. En annen grunn til å velge RL08 er at den kan brukes til å lage sannsynlighetsvarsler fra en deterministisk modell og dette kan være til stor nytte i daglig varsling. For å kunne bruke metoden som en sannsynlighetstilnærming, må vi vite på hvilken skala den verifiserer tilfredsstillende og det er derfor ikke like viktig å skille på type feil til dette formålet (jamfør tabell 2.1).

3. Metode

I dette kapitlet vil det først bli gitt en kort beskrivelse av verifikasjonsmetoden til Rogers og Lean (2008). Deretter illustrer vi dens anvendelse ved enkle eksempler.

3.1 Kort beskrivelse av Roberts og Lean (2008)

Både modelldataene og radardataene (se kapittel 5) er projisert til samme grid for direkte å kunne sammenlignes. Verifikasjonen gjøres i to steg: Først genereres fraksjonsverdier for både radarfelt og modellfelt så sammenlignes disse ved bruk av "Fraction Skill Score" (FSS).

3.1.1 Steg 1: Generere fraksjonene

En terskelverdi q for absolutt nedbørmengde over en gitt tidsperiode velges, for eksempel 1 mm/3t. Terskelverdien brukes både på modell- (M) og radarfeltet (R) til å regne ut binærverdien i hver gridrute, hhv. I_M og I_R . De gridrutene som overstiger eller er lik den valgte terskelverdien får verdien 1 alle andre 0:

$$I_R = \begin{cases} 1 & R \geq q \\ 0 & R < q \end{cases} \quad I_M = \begin{cases} 1 & M \geq q \\ 0 & M < q \end{cases} . \quad (1)$$

Dette gjentas for et sett med terskelverdier. Istedenfor absolutt nedbørmengde kan man velge relativ mengde, såkalt persentil. Vi kan for eksempel sette q til 95 % persentilen, som velger ut de 5 % høyeste verdiene over alle gridrutene i modell og radarfelt. Ved å bruke persentiler vil antall ruter med verdi 1 være det samme for både radar og modell og man får fjernet en eventuell bias mellom de to feltene, dvs. vi fokuserer på den romlige nøyaktigheten til varslet. På grunn av manglende presisjon på feltfilene som brukes til å lagre radar- og modelldata, har vi imidlertid erfart vanskeligheter med å definere et eksakt skille ved bruk av persentiler. For en gitt persentil velger algoritmen i (1) flere enn det angitte andelen ruter. Det ligger derfor inne en liten usikkerhet i denne beregningsmetoden. Alternativet er å holde seg til den angitt andelen ved å ta et tilfeldig utvalg av de rutene som har samme verdi. Etter

å ha testet på en del tilfeller finner vi at forskjellen mellom disse to alternativene blir veldig liten og ikke større enn andre kilder til usikkerhet.

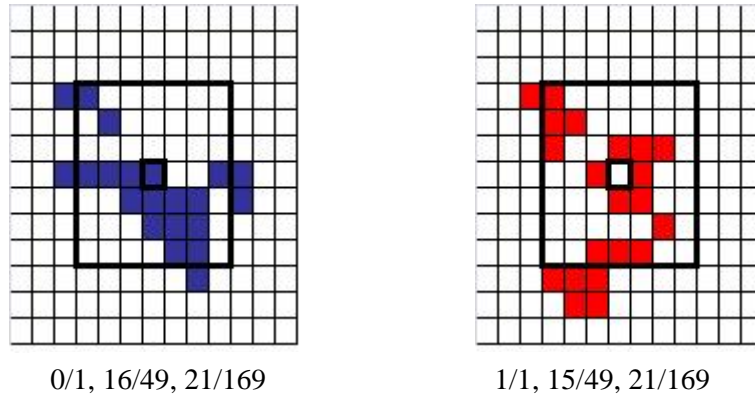
For hver terskelverdi har vi nå et par (radar og modell) assosierte binære felter. For hvert gridpunkt i domenet beregnes en fraksjonsverdi over forskjellige valgte kvadratiske naboområdet med n antall gridpunkter i hver retning. Naboområdene er sentrert omkring gridpunktet som gis fraksjonsverdiene R_n (radar) og M_n (modell):

$$R_n(i, j) = \frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n I_R \left[i+k-1-\frac{(n-1)}{2}, j+l-1-\frac{(n-1)}{2} \right] \quad (2)$$

$$M_n(i, j) = \frac{1}{n^2} \sum_{k=1}^n \sum_{l=1}^n I_M \left[i+k-1-\frac{(n-1)}{2}, j+l-1-\frac{(n-1)}{2} \right] \quad (3)$$

der i og j er koordinaten til gridruten. Fraksjonene kan så regnes ut for alle naboområdet fra $n=1$ til størrelsen på domenet. Merk at for å kunne beregne R og M i randsonen til domenet gjøres en kunstig utvidelse av domenet der nedbørintensiteten settes lik null i det utvidede området. Hvis utvidelsen ikke er symmetrisk, slik som i kapittel 5, kan man når naboområdet blir stort, i enkelte situasjoner få urealistiske resultater fra verifikasjonen.

Metoden er illustrert i figur 3.1. De gridrutene der hhv. varslet (blå) og observert (rød) nedbørsintensitet overstiger terskelverdien q , er vist i farger og får verdien 1, de andre rutene settes lik 0. På gridskalaen ($n=1$), sort firkant i sentrum av domenet, ser vi at radarfeltet har fraksjonen $0/1=0$ mens modellfeltet har fraksjon $1/1=1$, dvs. varslet er feil. Setter vi isteden $n=7$, dvs. et naboområdet på 7×7 gridruter, markert med stor sort firkant i figur 3.1, får radarfeltet fraksjonsverdi $15/49=0,31$ mens modellfeltet har fraksjonsverdi $16/49=0,33$. Altså et ganske godt varsel. Over hele domenet som i figuren er 13×13 gridruter har radarfeltet fraksjonsverdien $21/169=0,124$ og modellfeltet har også verdien $21/169=0,124$. Varslet er korrekt på denne skalaen og vi sier at biasen er null. Beregningene gjentas for hvert gridpunkt i domenet.



Figur 3.1: Grafisk fremstilling av fraksjonene for gridskalaen 1x1 gridruter, og domenene, 7x7 og 13x13 gridruter. Modellfelt til venstre i blått Radarfelt til høyre i rødt.

3.1.2 Steg 2: “Fraction Skill Score” FSS

Vi har nå for alle n funnet fraksjonsverdiene R_n og M_n i alle gridpunktene $i=1,2,\dots,N_x$ og $j=1,2,\dots,N_y$ der N_x og N_y er hele domenet det verifiseres over. For hver n sammenlignes radar- og modellfeltet ved å bruke MSE (Mean Square Error):

$$MSE_{(n)} = \frac{1}{N_x N_y} \left[\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} R_{(n)i,j} - M_{(n)i,j} \right]^2 \quad (4)$$

En MSE ”skill skår”, FSS, har så blitt utledet ut i fra et ”lav skår” referanse varsel (Murphy og Epstein, 1989). Denne er definert som:

$$FSS_{(n)} = \frac{MSE_{(n)} - MSE_{(n)ref}}{MSE_{(n)perfekt} - MSE_{(n)ref}} = 1 - \frac{MSE_{(n)}}{MSE_{(n)ref}}. \quad (5)$$

Referansevarslet brukt for et bestemt område er:

$$MSE_{(n)ref} = \frac{1}{N_x N_y} \left[\sum_{i=1}^{N_x} \sum_{j=1}^{N_y} R_{(n)i,j}^2 + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} M_{(n)i,j}^2 \right] \quad (6)$$

$MSE_{(n)ref}$ kan tolkes som den høyeste verdien MSE kan oppnå over et gitt naboområde.

$MSE_{(n)perfekt} = 0$ for et perfekt varsel over et bestemt naboområde, som vi så ovenfor. Høyest mulig skår har vi når FSS er lik 1 og det dårligste skåret når FSS er lik 0. Dårligst skår oppnås når $n=1$ (naboområdet er lik gridlengden) og binærverdiene er forskjellige mellom radar og modell (se 1x1 gridrute i figur 3.1). Skåret vil øke etter som området rundt gridpunktet (naboområdet) øker, og nærmer seg til slutt en asymptotisk verdi. Denne verdien kan beregnes ved å sammenligne radarfeltets fraksjonsverdi over hele domenet, f_R (ligning (8)), med modellens fraksjonsverdi, f_M (ligning (9)). Fra ligningene (4), (5) og (6) får vi den asymptotiske FSS:

$$AFSS = 1 - \frac{(f_R - f_M)^2}{f_R^2 + f_M^2} = \frac{2f_R f_M}{f_R^2 + f_M^2} \quad (7)$$

der

$$f_R = \frac{\text{ant.ruter.nedbør.radarfelt}}{\text{ant.ruter.nedbør.hele.verifikasjonsområde}} \quad (8)$$

$$f_M = \frac{\text{ant.ruter.nedbør.modellfelt}}{\text{ant.ruter.nedbør.hele.verifikasjonsområde}} \quad (9)$$

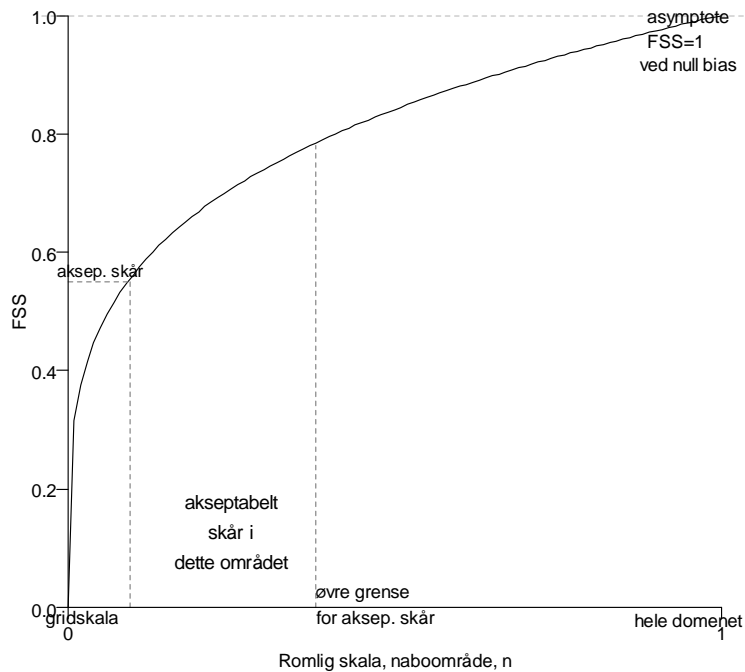
f_R kan sees på som domenets *våtandel*, dvs. hvor mye av radarfeltet som er dekt av nedbør. AFSS er det beste skåret vi kan oppnå for et gitt varsel og prognoselengde. Dersom varslet ikke har noen bias, dvs. like mange ruter med binærverdier lik 1 i modell- og radarfelt innenfor domenet, vil den asymptotiske verdien gå mot 1.

Figur 3.2 viser FSS som funksjon av n . Vi har et intervall langs n -aksen på kurven som går fra der FSS passerer akseptabel skår til naboområdet er utvidet til en øvre grense. Den øvre grensen kan være der feltet er glattet så mye at vi kan si at effekten av en finskalamodell blir borte. Akseptabelt skår er gitt ved:

$$f_{aksep} = 0,5 + \frac{f_R}{2} \quad (10)$$

Vi ser at grensen for akseptabelt skår vil øke når våtandelen er stor, for da er sjansen for å oppnå gode skår større enn dersom våtandelen er liten og det er få ruter med nedbør i radarfeltet. Våtandelen vil som regel minke med økende terskelverdi og f_{aksep} vil nærme seg 0,5. Dersom vi ser på lave terskelverdier og varsler med mye nedbør vil akseptabelt skår

kunne ligge en del over 0,5. Som vi skal se senere i kapittel 5.2 er f_{aksep} i gjennomsnitt for juli og august 2009 omkring 0,6 for lave terskelverdier q . Dette viser at juli og august 2009 har hatt nedbør over store områder og dermed en høy våtandel for lave terskelverdier.



Figur 3.2: Grafisk fremstilling av FSS mot romlig skala. Tegnet etter figur fra Roberts og Lean (2008).

FSS vil nærme seg AFSS når vi utvider naboområdet, og dersom $AFSS > f_{aksep}$ vil vi alltid få et mål på når akseptabel grense oppnås. Hvis $f_R = f_M$ er det ingen bias og f_{aksep} vil alltid være mindre eller lik AFSS. Men dette behøver ikke være tilfelle dersom det er en bias i varslene. For eksempel hvis 20 % av gridrutene i varslet overstiger terskleverdien, men bare 2 % av gridrutene i radarfeltet gjør det, vil $f_M = 10 * f_R$ og $AFSS \approx 0,198$. Da vil $f_{aksep} = 0,5 + (0,02/2) = 0,501$ og vi ser at akseptabelt skår blir mye høyere enn AFSS. Varslet har så stor bias at akseptabelt skår ikke kan oppnås for noe naboområde. FSS er sensitiv til biasen i varslene og er mer sensitiv til varsler med lav våtandel fordi det da er mer sannsynlig at f_R / f_M kan bli mye større enn 1.

I kapittel 6 skal vi se på hvordan man kan bruke sannsynlighetsverdier utledet fra denne metoden i daglig varsling. Vi kan bruke RL08 til å finne en sannsynlighetsverdier i hvert gridpunkt ved å bruke fraksjonene som ble beregnet ved ligning (2) og (3) for hvert naboombåde med størrelse n :

$$R_{sams} = R_n(i, j) * 100 \quad (11)$$

$$M_{sams} = M_n(i, j) * 100 \quad (12)$$

Ligning (11) og (12) gir da en sannsynlighetsverdi for hvert gridpunkt for en bestemt naboombådestørrelse og fra disse verdiene kan man lage et romlig kart som viser sannsynlighetsfordelingen av nedbøren utledet fra det deterministiske varslet (se figur 6.1).

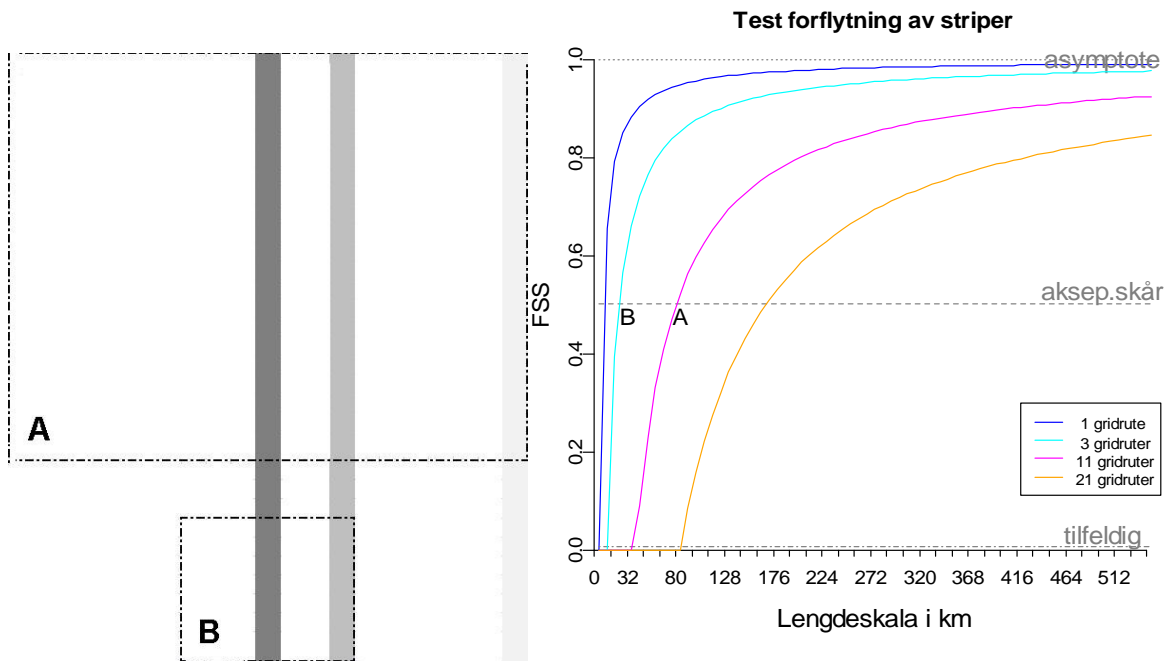
For å oppsummere: FSS har følgende karakteristikk:

- Tar verdier i intervallet 0-1; 0 for bomvarsel og 1 for et perfekt varsel
- Hvis det ikke er noen gridverdier i modellen som overstiger terskelverdien mens noen i radarfeltet gjør det, eller omvendt, er FSS alltid lik 0.
- Etter som størrelsen på naboombådene som brukes for å beregne fraksjonsverdiene blir større vil skåret gå mot en øvre verdi som er et forhold mellom den observerte frekvensen og modellens frekvens. Dvs. at jo nærmere denne asymptotiske verdien er 1 jo mindre bias er det. Ved å bruke persentiler sikrer en at FSS går mot 1 når naboombåde går mot hele verifikasjonsområdet.
- FSS er mest sensitiv til ”lav våtandel”.

3.2 Enkle tester

I de følgende to eksemplene er det regnet ut fraksjonsverdier for naboombåder fra 1 gridrute opp til 131 (tilsvarende 524 km^2 , ved et 4 km grid) for kunstige datasett.

Først ser vi hvordan FSS varierer når observasjonen og varselet viser et bånd av regn, en gridrute bredt, med uniform intensitet, men er plassert i ulike avstander fra hverandre, figur 3.3. Det er ingen bias og AFSS er derfor lik 1. Avstanden l mellom radar og modell er 1, 3, 11 og 21 gridruter.



Figur 3.3: Illustrasjon av forflytningsfeil. Henholdsvis 3 og 11 gridruter fra radarfeltet. I bildet til venstre er den mørkeste linjen radarfeltet, den mellom grå varslet nedbør 3 gridruter feil, den lys grå varslet nedbør 11 gridruter feil. (Roberts og Lean, 2008). Til høyre vises FSS for stripene med forskjellig antall gridruter forflytning. A og B viser hvor kurvene krysser akseptabelt skår, og firkantene til venstre tilsvarende naboområde. De horisontale stiplede linjene er AFSS (øverst), f_{aksep} (midten), $FSS_{tilfeldig} = f_R$ (nederst).

Når naboområdet er mindre enn avstanden mellom radar og modellområde ($n < 1$), er $FSS = 0$. Etter som vi øker naboområdet oppnår vi høyere skår og når til slutt verdien 1, asymptotisk grense (den øverste stiplede linjen i figuren, se formel (7)). Jo mindre forflytningsfeilen på modellen er jo raskere øker varslets skår med skalaen. Formel (10) gir oss et mål på akseptabelt skår (stiplet linje midt på figuren). I vårt eksempel er $f_R = 500 / 150000 = 0,003$ altså dekker nedbøren 0,33 % av hele domenet, våtandelen er lav og $f_{aksep} = 0,503$. Den siste stiplede linjen på figuren (merket "tilfeldig") er den FSS som oppnås av et tilfeldig varsel med samme våtandel som radarfeltet, denne er gitt ved $FSS_{tilfeldig} = f_R$. FSS er over akseptabelt skår ved naboområder på 10 km når $l=1$, ved ca 25 km $l=3$, 85 km for $l=11$ og

170 km for $l=21$ gridruter forflytting. Vi ser at FSS gir et tydelig mål på forflytningsfeil når nedbørsområdene ellers er like.

I neste eksempel er nedbørområdene forskjellige mellom radar og modell. Utgangspunkt er tatt i de geometriske områdene som er brukt i deler av "Spatial Forecast Verification Methods Inter-comparison Project" (ICP) (Ahijevych et al., 2009, Gilleland et al., 2009a), for å undersøke forflytningsfeil, orienteringsfeil (rotasjon) og bias. Figur 2.2 og tabell 3.1 viser hvordan områdene av varsel og observert nedbør er plassert i forhold til hverandre samt biasen. Alle punkter i radarfelt og modellfelt som har nedbør har fått verdien 4 mm og det er testet på terskelverdien 1 mm. A og B representerer forflytningsfeil, henholdsvis liten og stor (tilsvarende testen over for striper), C og E gir feil pga bias, der E har mye større bias enn C (biasen er kun i form av større område og ikke større verdier av nedbøren). D ser på rotasjonsfeil. Hensikten med å teste FSS på disse geometriske tilfellene er å vurdere om FSS gir korrekte egenskaper ved feilene og en klar representasjon av varslets riktighet.

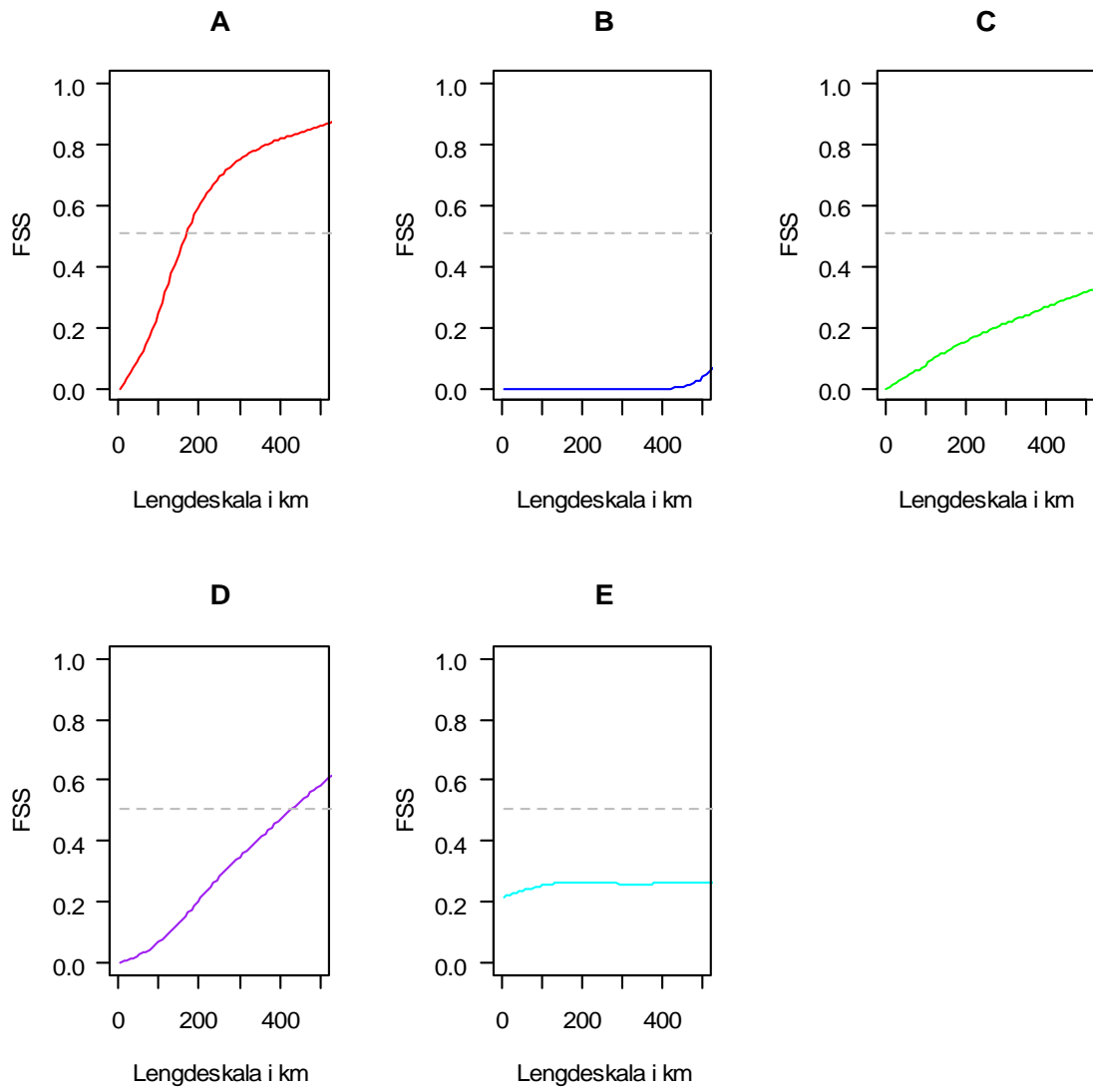
Våtandelen til radarfeltet er den samme i alle 5 tilfellene slik at $f_{aksept} = 0,508$.

FSS for hvert av tilfellene er vist i figur 3.4 og viser hvordan FSS varierer for de forskjellige typer feil. At figurene her for tilfelle A, B og D ikke går mot 1 er kun fordi vi pga manglende datakapasitet ikke har beregnet FSS for naboområder større enn 500 km. Kun tilfelle A og D får FSS-skår som når over akseptabel grense. Dette viser at det kun er tilfellene som har varsel og radarfelt nærme hverandre og med liten eller ingen bias som gir tilfredsstillende skår. Varsel E er det eneste varsel som har noe skår helt fra minste naboområde, altså gridstørrelsen. Dette fordi det her er noe overlapp av modell- og observasjonsfelt, men skåren øker ikke noe særlig og kommer aldri over grensen for akseptabelt skår. Varsel B som er langt unna har nesten ikke økning av skåren i det hele tatt.

Tilfelle	Beskrivelse
RADAROMRÅDE	2500 gridruter
A	25 punkt til høyre (~100 km) – begge felt 2500 gridruter
B	125 punkt til høyre (~500 km) – begge felt 2500 gridruter
C	25 punkt til høyre (~100 km) og for stort – radar 2500 gridruter, modell 10000 gridruter LITEN BIAS
D	25 punkt til høyre (~100 km) og feil orientering – begge felt 2500 gridruter
E	75 punkt til høyre (~300 km) og for stort m/ overlapp – radar 2500 gridruter, modell 20000 gridruter STOR BIAS

Tabell 3.1: Beskrivelse av de 5 geometriske områdene i figur 2.2.

Disse eksemplene viser at FSS kan gi et sannferdig bilde av forflytningsfeil. I tillegg kan man bruke FSS til å bestemme for hvilken skala man oppnår akseptable skår for varslets prediktabilitet og dette er kanskje vel så informativt som den absolutte verdien. Det viser også at FSS reagerer på varslets bias. Et varsel med høy bias vil alltid ha lavere verdier for FSS på store skalaer og som regel også på små skalaer. Men det er mulig å få høyere verdier for et varsel med høy bias på de lavere skalaene, hvis modellen overlapper den observerte nedbøren som i tilfelle E. Men med et tilstrekkelig antall varsler vil FSS ikke gi høyere skår for et varsel med større bias.



Figur 3.4: Grafer som vise FSS for de 5 forskjellige geometriske tilfellene A, B, C, D og E.

4. Data

4.1 Modell

Meteorologisk Institutt har kjørt UK Met Office sin modell "Unified Modell" (UM) rutinemessig siden 2003/4. UM er en ikke-hydrostatisk numerisk værvarslingsmodell som er godt egnet for å varsle værphenomener på mesoskala. Den bruker et semi-implisitt, semi-Lagrangian numerisk skjema (Davis et al., 2005). UM bruker et Arkawa C grid horisontalt og vertikalt. Prognostiske modellparametere inkluderer vind, potensiell temperatur, spesifikk fuktighet, tetthet, skypartikler og nedbør. Fysisk parametrisering tar hensyn til skyer, gasser og iskrystaller, langbølget og kortbølget stråling, mikrofysikk i skyer og kondensasjon. I denne oppgaven er varslene hentet fra den såkalte UM4-modellen som er satt opp for et større geografisk område som dekker Norge, Sverige og deler av Finland, Russland og Danmark og tilhørende havområder. UM4 er satt opp med 4 km horisontal gridavstand og 38 nivåer i vertikalen opp til ca. 40 km. Når modellen kjøres med 4 km oppløsning, parametriseres konveksjon ved et CAPE-basert oppdriftsskjema som tar hensyn til at på denne oppløsningen er nedbøren både eksplisitt og implisitt representert i modellen.

Randverdier (hver time) og initialbetingelser (overflate og atmosfære) tas fra den norske modellen HIRLAM 8 km, dvs UM4 er en dynamisk nedskalering av HIRLAM. UM4 kjøres to ganger i døgnet (00 og 12 UTC) begge ganger med en prognoselengde på 66 timer. I oppgaven er varsler for hver 3 time frem til 24 timer benyttet, bortsett fra i kapittel 5.4 der vi har brukt varsler frem til 27 timer. Vi har kun benyttet kjøringene fra 00 UTC. I denne perioden var det referanseversjon 6.1 som ble benyttet ved met.no. Utover referanseversjonen har met.no gjort endringer i modellen med hensyn på behandlingen av prosesser relatert til snø på bakken og det stabile grenselaget.

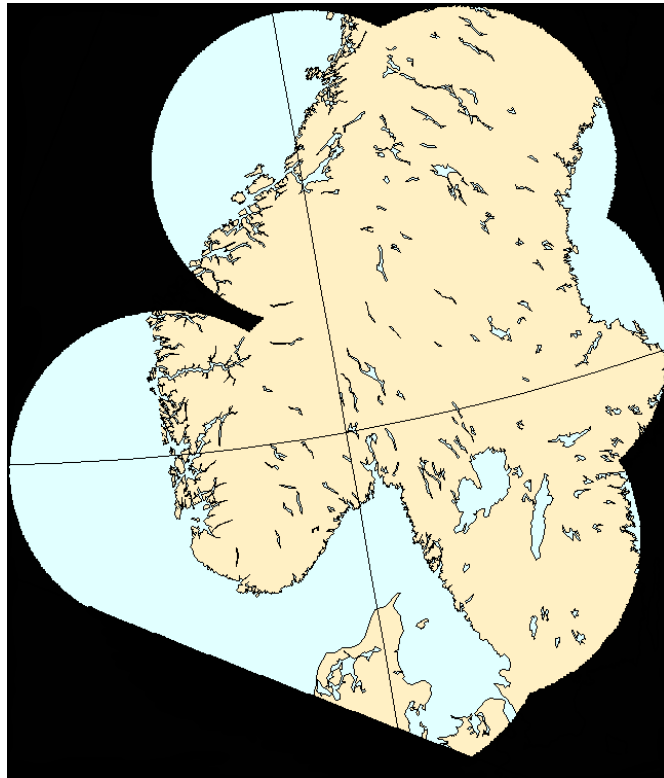
4.2 Radardata

Det Norske Meteorologisk Institutt har i dag åtte operative radarer som dekker store deler av Sør-Norge, Trøndelag samt mesteparten av Nordland, Troms og kysten av Finnmark frem til

Nordkapp (http://met.no/Meteorologi/A_male_varet/Observasjoner_fra_land/Radar/).

Radarene som er brukt i denne oppgaven er Hagahogget i Asker, Hægebostad i Vest-Agder, Bømlo i Hordaland, Rissa i Sør-Trøndelag og i tillegg tre svenske radarer. Radaren ved Stad kom ikke i operativ drift før medio september 2009 og er derfor ikke med.

Verifikasjonsdomenet er vist i figur 4.1 og er på ca $800\,000\text{ km}^2$, og utgjør det geografiske området som radarene dekker, samt UM4. Den horisontale oppløsningen på radardataene er 1 km.



Figur 4.1: Utsnittet det verifiseres over, begrenset av radardekning og UM4.

Radardataene er uten noen form for korleksjoner for feilkilder hvilket reduserer kvaliteten på målingene av reflektert intensitet (se Appendix for feilkilder). Den viktigste og vanligste feilkilden er den komplekse topografien vi har i Norge. Dette medfører at radarene har større eller mindre blokkeringer over land. Det ble utviklet en metode for å korrigere for dette for radar Oslo og Hægebostad (Gjertsen og Dahl, 2002), men denne metoden benyttes imidlertid ikke lenger. Radar Oslo, den eldste radaren i drift, mangler dopplerfiltrering av reflektiviteten. Resultatet er at det er rikelig med markekko, eller «ground clutter», rundt Oslofjorden. For å redusere dette brukes et statisk filter, eller en maske (cluttermap), som

fjerner pikselverdier permanent der man vet at det er falske ekko. Her blir det da en avveining, fjerner man for mye kan man også fjerne ekko fra nedbør, fjerner man for lite må man leve med en del faste ekko. Dopplerfilter derimot fjerner bare piksler når ekkoene ikke flytter seg og er derfor en bedre måte å fjerne falske ekko på (personlig kommunikasjon Trygve Aas).

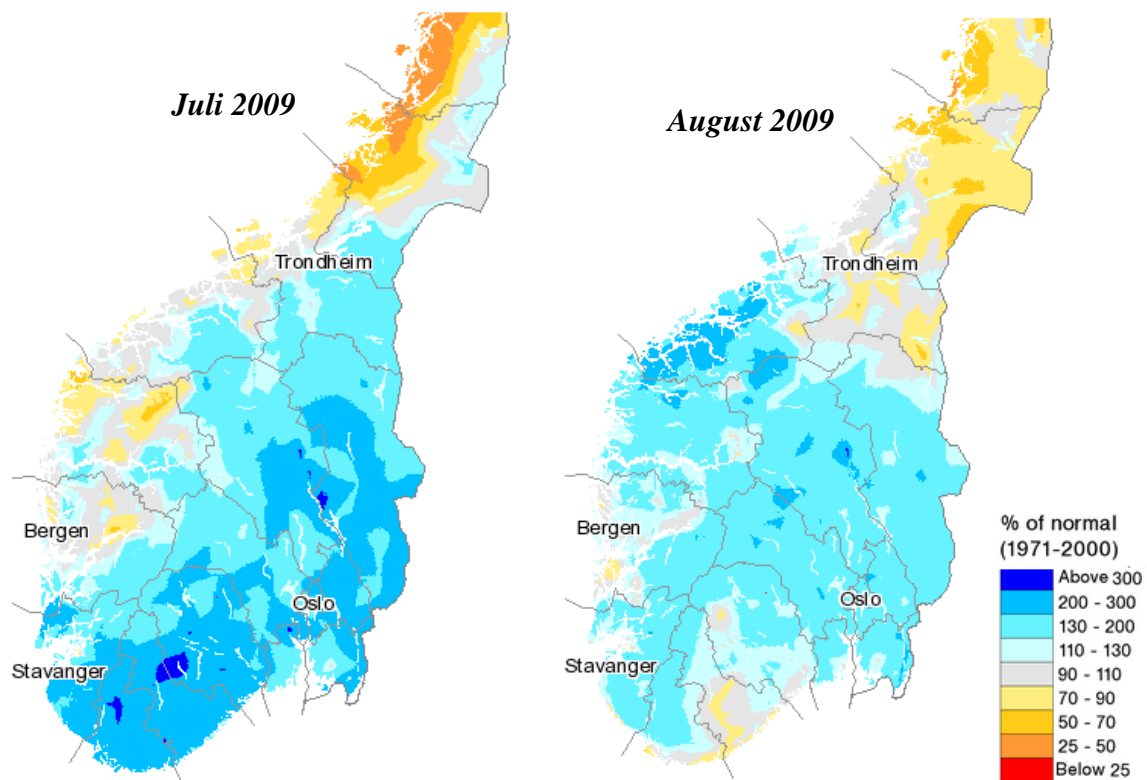
Radarene måler reflektert stråling (se Appendiks). Disse verdiene konverteres så til akkumulert nedbør typisk over en periode på 15 eller 7,5 minutter, som igjen summeres over lengre tidsperioder, for eksempel 1 eller 3 timer. Algoritmen for konvertering av refleksivitet til mm per time er beheftet med usikkerhet pga. antagelser og kalibreringer (Reinhart, 2004). Det er lite dokumentasjon på kvaliteten av akkumulert nedbør, spesielt for Norge. Men i samarbeidsprosjektet "Operational Programme for the Exchange of weather RADar information" (OPERA) er det gjort en studie på å verifisere radarene i Europa mot tre forskjellige datasett for nedbør: satellittbaserte observasjoner, synoper og ECMWFs operasjonelle modell (Lopez, 2008). Resultatene viser at kvaliteten på radardataene varierer geografisk. Spesielt over Skandinavia er ikke radarestimatene for nedbør spesielt tilfredsstillende (noe kan skyldes snødekte områder da studien dekker perioden 10.april til 8.juli 2008). Studien viser også at radaren hovedsakelig underestimerer nedbørsmengdene, bortsett fra over deler av Frankrike og Nordsjøen. Lopez (2008) viste imidlertid at "feilen" i radardataene var mindre når man sammenlignet med både synoper og satellittobservasjoner, enn når man sammenlignet med modellen, ECMWF, noe som tyder på at radardataene er nærmere virkeligheten enn modellen.

I denne oppgaven er det brukt 3 timers akkumulert nedbør, og hver modellkjøring er startet 00 UTC og kjørt 24 timer, dvs. 8 felt hver dag, ca 500 felt totalt. Radardataene er interpolert til UM4 sitt gitter. Området i Fig. 4.1 utgjør dermed ca 50 000 gridpunkter.

4.3 Datagrunnlag – juli og august 2009

Juli og august 2009 var begge nedbørrike måneder i Sør-Norge og gir derfor et godt grunnlag for testing av metoden. Det var ingen dager som var helt uten nedbør i området vi verifiserer over. Mye lavtrykksaktivitet med tilhørende frontsystemer passerte over Sør-Norge disse månedene. Det var mye konvektiv aktivitet i disse frontene selv om den "typiske" sommernedbøren med oppvarming over land og byger utover ettermiddagen ikke

var like fremtredende. Jevnt over kan vi si at nedbøren kom ofte og mye. Observasjoner fra synoper og nedbørsstasjoner viser at månedsnedbøren for Norge som helhet var 125 % av normalen i juli og 120 % av normalen for august. Mest nedbør hadde Sør-Norge. I juli fikk store områder av Sørlandet og Østlandet over det dobbelte av normalen. Det samme gjaldt deler av Oppland, Sogn og Fjordane og Møre og Romsdal i august, mens deler av Rogaland og Østlandet fikk 150-200 % av normalen, figur 4.2.

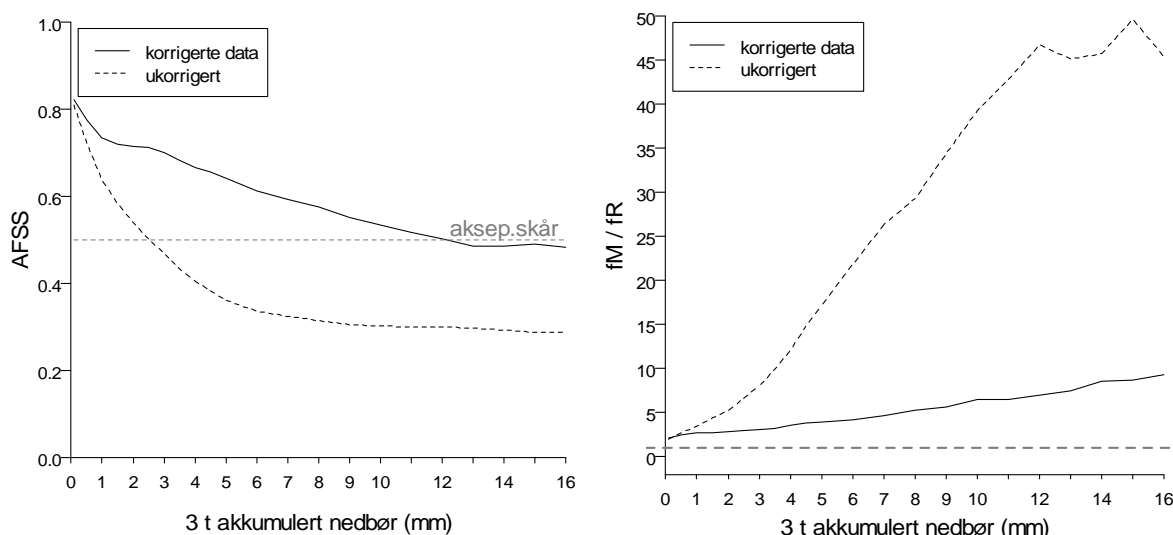


Figur 4.2: Prosentvis avvik fra nedbørnormalen for juli (venstre) og august (høyre) 2009 for Sør-Norge (<http://senorge.no/>).

4.4 Korrigerering av dataene

AFSS gir oss et mål på biasen i et enkelt modellvarsel sammenlignet med radarmålingene for gitt prognoselengde og terskelverdi (se ligning (7) i kapittel 3). Den stiplede linjen i figur 4.3 viser AFSS for ukorrigererte data, midlet over alle varslene for juli og august. For små terskelverdier er AFSS i overkant av 0,8 men synker raskt for høyere terskelverdier, altså

øker biasen raskt når terskelverdien øker. Men AFSS sier oss ikke noe om det er radarfeltet eller modellfeltet som har størst utbredelse av nedbøren for den enkelte terskelverdigrense. Til høyre i figur 4.3 vises forholdet mellom frekvensen av nedbøren i modellen mot frekvensen i radarfeltet, f_M / f_R (se ligningene (8) og (9)) som funksjon av terskelverdi. Ved like stor utbredelse av nedbøren i både modell og radarfelt for hver terskelverdi skulle forholdet mellom frekvensene vært 1 (den stiplede horisontale linjen), mens når $f_M / f_R > 1$ (< 1) har modellen mer (mindre) nedbør enn radarfeltet. For ukorrigerede data (stiplet kurve) er f_M / f_R omkring 1,8 for de laveste terskelverdiene, dvs. at vi har nesten dobbelt så mange gridruter med nedbør i modellen sammenlignet med radaren, og forholdet øker raskt med økende terskelverdi. Dette viser at modellen enten overvarsler og/eller radaren har for få gridpunkter med nedbør og/eller for lave verdier. Dette er konsistent med Lopez (2008).



Figur 4.3: Venstre: AFSS i forhold til økende terskelverdier for ukorrigerede data (stiplet linje) og korrigerede data (heltrukken linje). Høyre: Forholdet mellom frekvensen av modellnedbør (f_M) mot frekvensen av radarnedbør (f_R) som funksjon av terskelverdi. Terskelverdiene er lik for begge feltene. Den stiplede horisontale linjen i figuren til høyre viser om modell eller radarfelt har mest nedbør (over har modellen mest nedbør, under radaren). Begge grafer viser akkumulert nedbør over 3 timer og gjennomsnitt over ca 500 varsler for juli og august.

At antall gridpunkter med nedbør i radarfeltet i gjennomsnitt er mye lavere enn i modellen vanskeliggjør bruk av absolutte terskelverdier (ikke vist). Derfor ble radardataene korrigeret lineært etter formelen:

$$\text{Verdi radar} = (\text{verdi radar} / \text{middel radar}) * \text{middel modell} \quad (13)$$

Middelverdien i modellfelt og radarfelt er regnet ut for hvert tidssteg (akkumulert 3 timers nedbør) og radarfeltet er så korrigert for hvert tidssteg underveis.

Denne korreksjonen kan anses som en forenklet korreksjon av de usikkerheter som skyldes konverteringen fra radarreflektivitet til akkumulert nedbør samt feilkilder som påvirker målt reflektivitet. Typisk korrigerer man modelldataene siden radarobservasjonene ligger nærmere sannheten (Lopez, 2008). Imidlertid har varslerne, forskerne og brukere (gjennom for eksempel yr.no) mer følelse på absolutte terskelverdier enn relative terskelverdier (persentiler). Vi valgte derfor å korrigere radarobservasjonene for å se om det var mulig å bruke absolutte terskelverdier i verifikasjonen. Dessverre foreligger ikke dataene fra seNorge på et slikt format at en korreksjon av både radar og UM4 mot disse observasjonene var mulig innenfor rammene av denne oppgaven.

Logaritmisk korreksjon er ofte brukt for nedbør og dette ble derfor også testet men viste seg å ha liten effekt. Ved logaritmisk korreksjon ble det først tatt naturlig logaritme av alle verdiene i radarfeltet deretter korrigert etter formel (13) ovenfor, så ble verdiene i hver gridrute konvertert tilbake eksponentielt.

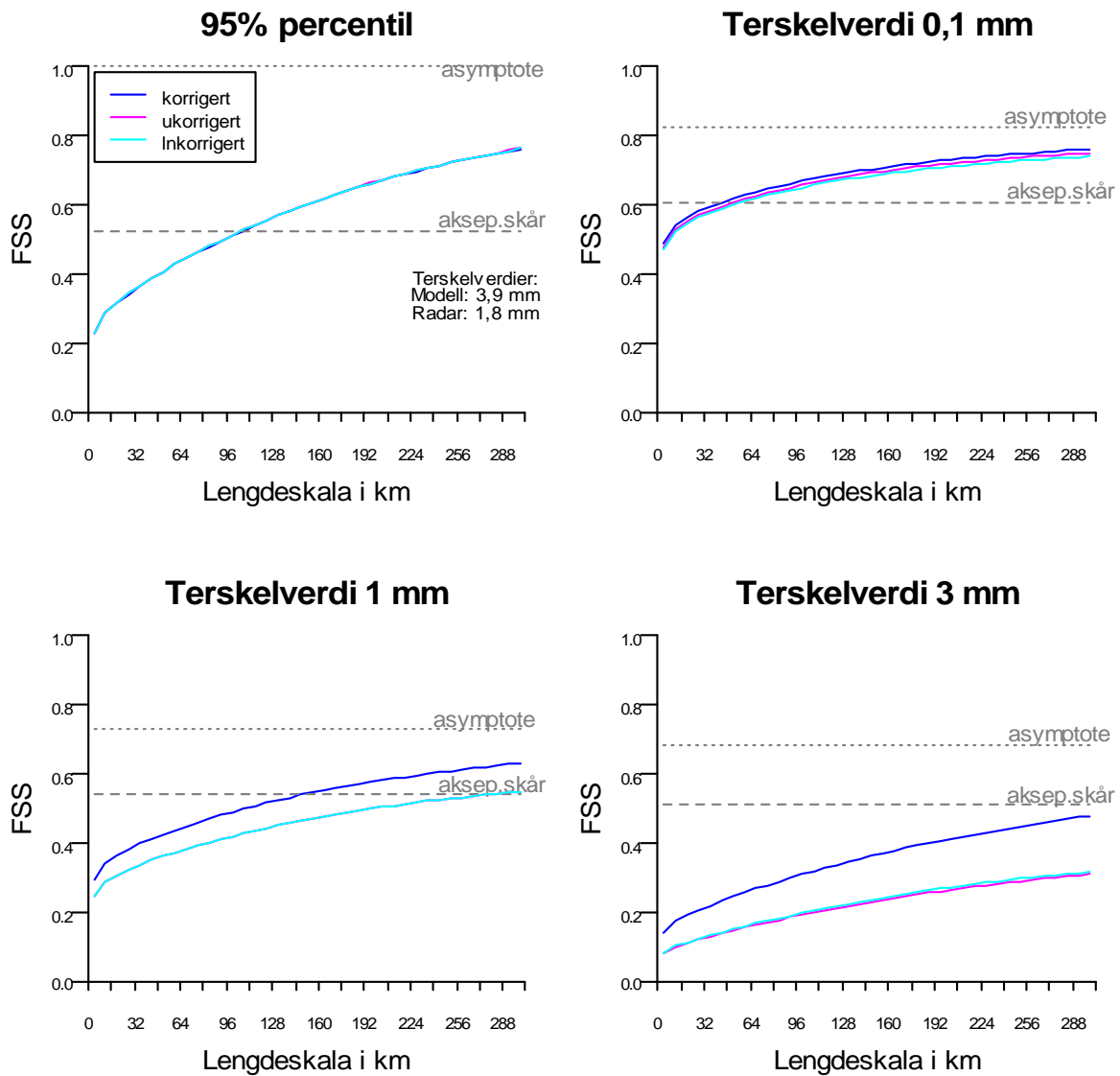
På figur 4.3 ser vi at biasen i varslene ble betydelig mindre ved å korrigere radardataene. For terskelverdi 2 mm økte AFSS fra 0,3 for ukorrigerede data til 0,5 for korrigerede data. Også forholdet mellom frekvensen av nedbøren i modellfelt og radarfelt (til høyre i figur 4.3) ble betraktelig mindre ved den lineære korrigerings. For terskelverdien 2 mm gikk f_M / f_R fra 5,3 til 2,8, dvs. at etter radarfeltet ble korrigert hadde modellen 2,3 ganger så mye nedbør mot før korrigeringsen 5,3.

Figur 4.4 viser hvilke utslag de to korrigeringsene ga for FSS-skåret sammenlignet med ukorrigerede data. For lave terskelverdier er det liten forskjell mellom opprinnelige verdier og de to typer korrigerings mens denne forskjellen øker når terskelverdien øker. For terskelverdi 0,1 mm får vi en forbedring på ca 10 km mellom ukorrigerede data og lineært korrigerede data, mens vi får hhv. 90 km og 200 km forbedring for 1 mm og 3 mm terskelverdier. Vi ser liten effekt av den logaritmiske korrigeringsen.

For 95 % persentilen (relativ terskelverdi) gir korrigeringsen som ventet ikke noe utslag verken ved lineær korrigerings eller logaritmisk korrigerings. Bruk av persentilene er i seg selv en type biaskorrigerings, der vi kun tar hensyn til de prosentvis høyeste verdiene. Selv

etter å ha korrigert radardataene, er det fortsatt bias i mange av varslene, så persentilene gir for våre data det beste bilde av hvilken skala modellen gir akseptable verdier. Absolutte terskelverdier er lett å forstå og viktige for mange brukere siden nedbøren er målt i mengde og denne er den samme i både radarfelt og modellfelt. For varsler med liten eller ingen bias er dette en god tilnærming å bruke for verifikasjon.

Det bør også bemerkes at ved bruk av persentiler er gjennomsnittet for juli og august blitt tatt over noe færre varsler enn totalen på ca 500 varsler, dette gjelder for de laveste persentilene. For eksempel er 90 % persentilen tatt over et gjennomsnitt på ca 450 varsler. Dette er fordi noen varsler ikke har nok gridruter med nedbør til at det kan bli fastsatt en terskelverdi. For eksempel skal 90 % persentilen ha et utvalg av de 5000 høyeste verdiene. Dersom et av feltene ikke har så mange ruter med nedbør kan ikke en terskelverdi fastsettes.



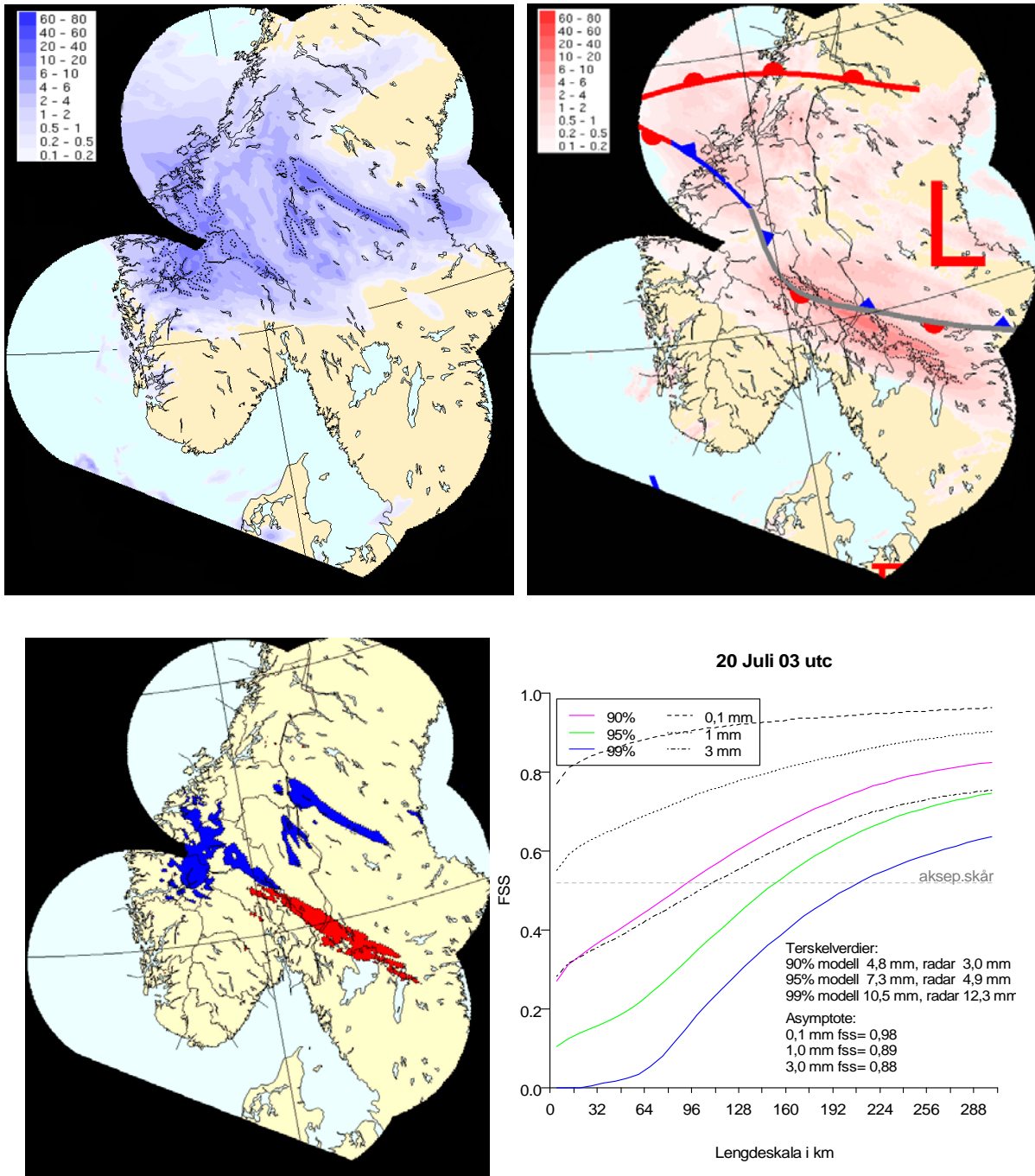
Figur 4.4: Forskjell mellom ukorrigerte radarverdier (rosa), lineært korrigerede (blå) og logaritmisk korrigerede (lys blå). De horisontale stiplede linjene er AFSS (asymptote) og f_{aksep} (aksep. skår) for de korrigerede dataene, som vil avvike noe fra de ukorrigerede. Gjennomsnitt er tatt over varslene fra UM4 og radarfelt for juli, 247 varsler.

5. Resultater og diskusjon

5.1 Enkeltvarsler

For å illustrere hvordan FSS oppfører seg på virkelige varsler, har vi sett på noen enkeltsituasjoner. Et varsel som er subjektivt ”dårlig” for høye persentiler og terskelverdier, men relativt godt for lavere terskelverdier (20.juli 03 UTC). Et varsel som er subjektivt godt for alle persentiler og terskelverdier, bortsett fra den høyeste terskelverdien hvor varslene har stor bias (11.juli 21 UTC). Og et siste varsel som får dårlige skår for nesten alle persentilene og terskelverdiene, men som er et godt eksempel på feilplassering (13.august 03 UTC). Alle varslene viser 3 timers akkumulert nedbør.

I eksemplet fra 20. juli er det ekstremene i nedbøren som er feilplassert mens selve utbredelsen av nedbørsfeltet stemmer rimelig godt mellom UM4 og radarfeltet. Figur 5.1 viser nedbørsfeltet både i modell (øverst til venstre) og radar (øverst til høyre). På denne dagen lå det en nesten stasjonær front over Norge og midtre deler av Sverige (skissert i radarfeltet) med svakt vindfelt på bakken og i 850 hPa-flaten (ikke vist). I 500 hPa-flaten var det østlig vind, men heller ikke her var vindfeltet veldig sterkt (ikke vist). Radarfeltet har hovedtyngden av nedbøren plassert over østlige deler av Østlandet og innover i Sverige, mens UM4 har hovedtyngden i Møre og Romsdal og vestlige deler av fjellet der. I denne situasjonen kan det virke som modellen gir for mye nedbør på lesiden av fjellet (på vestsiden), og at modellen varsler mye på frontsonen og dermed tar mer hensyn til de meteorologiske parametrene enn topografien. Visuelt ser det ut til at den geografiske fordelingen av nedbøren stemmer relativt godt mellom radar og modell, men om vi sammenligner de høyeste nedbørsmengdene ser vi at det er mindre overensstemmelse. De stiplede linjene på de øverste figurene viser 95 % persentilen, de 5 % høyeste verdiene i henholdsvis modell og radarfelt, disse er også vist nederst til venstre i figuren. Det kan virke som modellen plasser frontsonen over Østlandet litt for langt nord, mens nedbøren på Vestlandet er litt for sterk og plassert for langt sør. At det i modellen er to områder med maksima har nok også innvirkning på FSS-skåret.



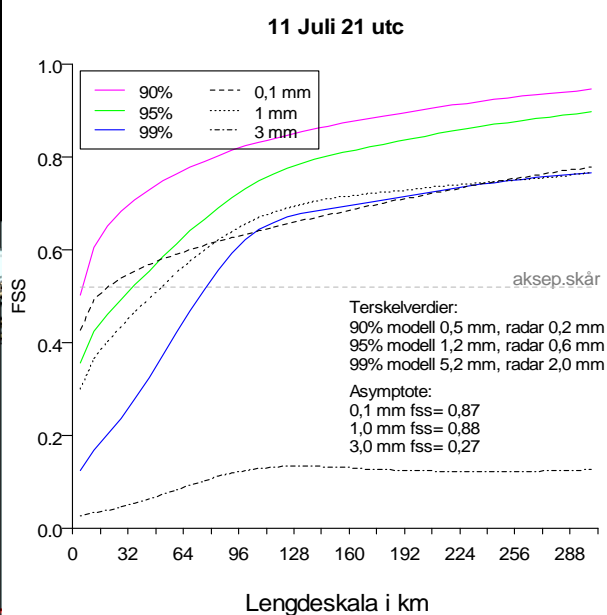
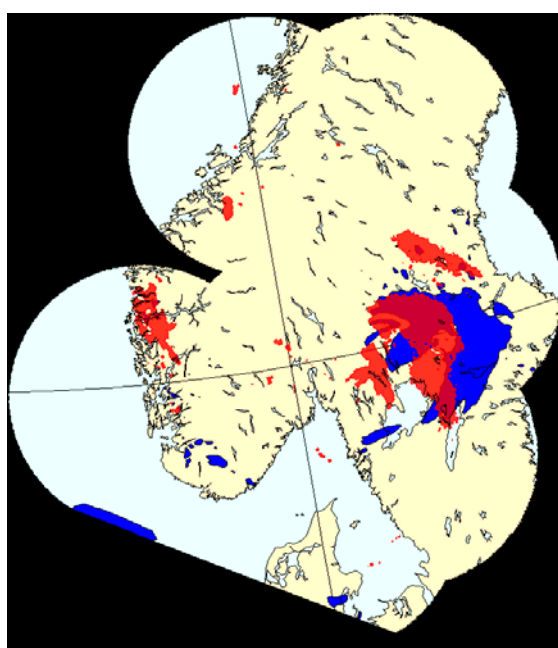
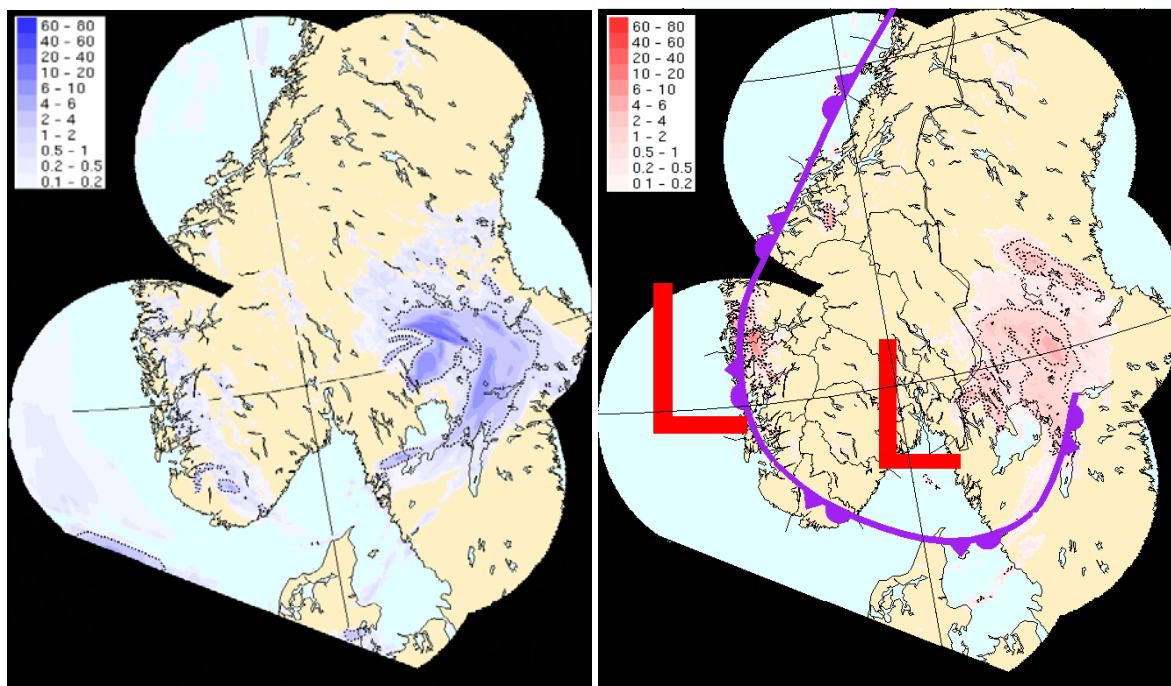
Figur 5.1: 3 timers akkumulert nedbør fra 20.juli 03 UTC. UM4 øverst til venstre i blått og radarfeltet øverst til høyre i rødt. 95 % persentilen er merket med stiptet linje. På radarfeltet er subjektiv bakkeanalyse fra met.no lagt på. Nederst til venstre vises 95 % persentilen for modell og radarfelt. For UM4 tilsvarende dette en terskelverdi på 7,3 mm og for radaren 4,9 mm. Rødt er radarfeltet og blått modellfeltet. Nederst til høyre FSS for 3 terskelverdier og 3 persentiler. Den horisontale stiplede linjen er f_{aksept} (aksept.skår).

Siden det var store nedbørsmengder over store geografiske områder, ble terskelverdiene for 95 % persentilen forholdsvis høye, henholdsvis 7,3 mm for modellen og 4,9 mm for radaren. Det er ikke mye overlapp mellom 95 % persentilene, kun litt i området nord for Mjøsa. Med lite overlapp av nedbørsfeltene blir FSS for små naboområder alltid lav (se geometriske tester i kapittel 3.2). Kurven for FSS (nederst til høyre i figur 5.1) viser at vi må opp i et naboområdet på nærmere 140 km før vi når akseptabelt skår for 95 % persentilen.

For lave terskelverdier har vi FSS-skår over akseptabel grense helt fra gridskalaen, 4 km. Grunnen til dette er lav bias, og mye overlapp for lave terskelverdier. AFSS for terskelverdi 0,1 mm er 0,98 og viser at radar og modellfelt har nesten like stor utbredelse av nedbøren, dvs. konsistent med det visuelle inntrykket. Vi får et godt FSS-resultat for alle de absolutte terskelverdiene (AFSS er skrevet på figuren, ikke plottet, fordi den er forskjellig for alle terskelverdiene). Den høyeste absolutte terskelverdien (3 mm) er mindre enn terskelverdien for 95 % persentilen (7,3 mm for modell og 4,9 mm for radar) og vi kan se av figuren at det sannsynligvis er mye overlapp av nedbøren for terskelverdier under denne grensen, som gir gode FSS-skår. 90 % persentilen har terskelverdier nærme 3 mm og vi ser at kurven for terskelverdi 3 mm og 90 % persentilen er ganske like, her er det biasen som skiller resultatene.

Eksemplet fra 11.juli 21 UTC, figur 5.2, viser et subjektivt ”godt” varsel. I denne situasjonen var det et eldre lavtrykk over Skagerrak med en tilhørende okklusjon som strakk seg fra Sverige, over Skagerrak og opp langs Vestlandskysten. I løpet av dagen ga okklusjonen en del tordenbyger i hele Sør-Norge, spesielt i Agder, og inn i Sverige. Bygeaktiviteten ble redusert utover kvelden.

FSS når grensen for akseptabelt skår for naboområder mindre enn ca 70 km for alle persentilene, også 99 % persentilen som representerer lokal nedbør, figur 5.2 nede til høyre. For alle terskelverdiene og persentilene øker skåret sakte, etter de første 100 km, når naboområdet utvides. For den høyeste terskelverdien, 3 mm, når FSS-skåren ikke høyere enn 0,13 for naboområder opp til 300 km (AFSS=0,27 for terskelverdi 3 mm). Ved bruk av absolutte terskelverdier vil vi aldri kunne nå FSS=1 dersom det er en bias, men vil maksimalt nå de respektive asymptotiske verdiene. Metoden kan på denne måten si oss noe om biasen i varslene.



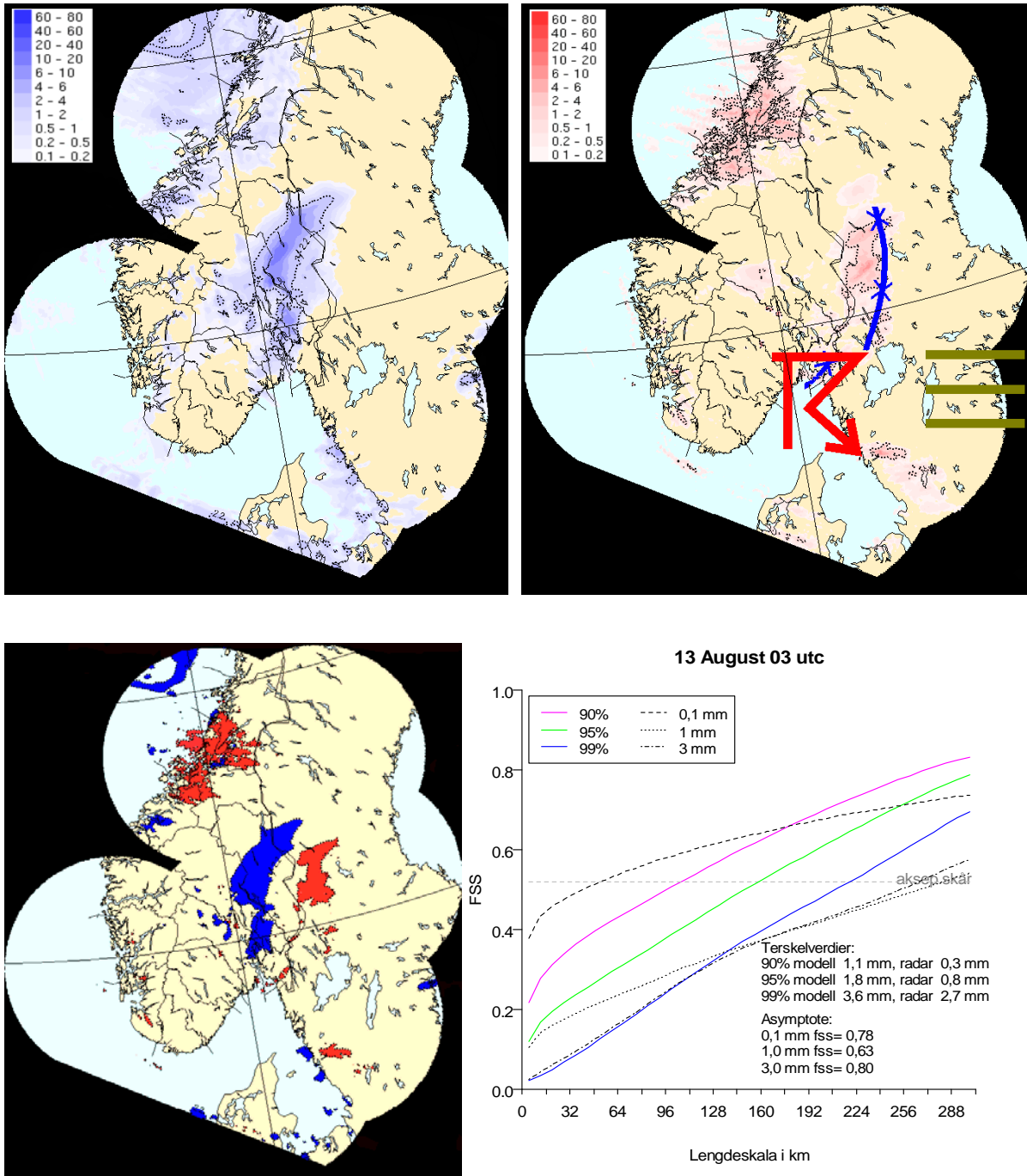
Figur 5.2: 3 timers akkumulert nedbør fra 11.juli 21 UTC. UM4 feltet øverst til venstre i blått og radarfeltet øverst til høyre i rødt. På radarfeltet er subjektiv bakkeanalyse fra met.no lagt på. 95 % percentilen er merket med stiplet linje. Nederst til venstre vises 95 % percentilen for modell og radarfelt. Rødt er radarfeltet og blått modellfeltet. For modell tilsvarende dette en terskelverdi på 1,2 mm og for radaren 0,6 mm. Nederst til høyre FSS for 3 terskelverdier og 3 percentiler. Den horisontale stiplede linjen er f_{aksept} (aksept.skår).

At AFSS er så lav som 0,27 for terskelverdien 3 mm viser at det er stor bias for denne terskelverdien og varslet vil aldri kunne nå grensen for akseptabelt skår (se kapittel 3.1.2). For de mindre terskelverdiene er ikke biasen spesielt stor. Totalt var våtandelen, f_{rad} , i dette varslet omkring 1/4 av våtandelen i varslet fra 20.juli. FSS er sensitiv til varsler med lav våtandel, likevel er dette et relativt godt varsel.

I varslet fra 11.juli når 95 % persentilen grensen for akseptabelt skår ved ca 35 km. Dette er en stor forbedring sammenlignet med situasjonen den 20.juli (155 km). Varslet fra 11.juli er ikke perfekt, men viser at varsler med noe overlapp (nederst til venstre i figur 5.2) gir positive utslag på FSS i forhold til varsler med lite overlapp (samme bilde i figur 5.1).

Varsler med lite overlapp av nedbørområder har lave verdier for FSS for små naboombråder, men kan øke raskt dersom varslet er nærme observasjonen. Varsler med mere overlappende nedbørområder begynner på høyere verdier for FSS, men kan flate ut når man utvider naboombrådet dersom det er bias i systemet. Dette stemmer overens med de geometriske tilfellen vi så på i kapittel 3.2.

Figur 5.3 viser et varsel hvor modellen har plassert nedbøren feil i forhold til radarobservasjonene. På nordvestlandet har vi pålandsvind med orografisk nedbør, men radardataene viser sterkere nedbør i dette området enn det modellen gir. På Østlandet har vi en bygelinje som modellen plasserer for langt vest i forhold til radarfeltet. Nederst til venstre i figuren, som viser 95 % persentilene, ser vi at det nesten ikke er overlapp for denne bygelinjen. Nedbørområdet på Østlandet er plassert omrent 120 km for langt vest i forhold til radarfeltet. Forflytningsfeilen kan vi indirekte trekke ut fra kurven for FSS, hvor vi må opp i et naboombråde på omkring 160 km før vi når grensen for akseptabelt skår. Nedbøren over nordvestlandet vil også påvirke dette resultatet og vi får derfor ikke eksakt mål på forflytningsfeilen for denne situasjonen. Vi ser at kurven for FSS er bratt når vi utvider naboombrådet. Dette tyder på at varslet er ”godt” men plassert noe feil (jamfør tilfelle A i figur 3.4) og at skåret derfor øker raskt etter som naboombrådet blir utvidet. På denne måten gir metoden oss et bilde på forflytningsfeil. Verifikasjonen gir dårlige skår for alle persentilene og de høye terskelverdiene, men gode skår for den laveste terskelverdien. Dette varslet har forholdsvis høy bias, de asymptotiske grenseverdiene er fra 0,6 til ca 0,8. Det forklarer hvorfor vi får gode skår for lave terskelverdier og små naboombråder, når det er noe overlapp av modellfelt og radarfelt (jamfør geometrisk tilfelle E i figur 3.4).



Figur 5.3: 3 timers akkumulert nedbør fra 13.august 03 UTC. UM4 feltet øverst til venstre i blått og radarfeltet i rødt øverst til høyre. På radarfeltet er subjektiv bakkeanalyse fra met.no lagt på. 95 % persentilen er merket med stiptet linje. Nederst til venstre vises 95 % persentilen for modell og radarfelt. Rødt er radarfeltet og blått modellfeltet. For modell tilsvarer dette en terskelverdi på 1,8 mm og for radaren 0,8 mm. Nederst til høyre FSS for 3 terskelverdier og 3 persentiler. Den horisontale stiplede linjen er f_{aksep} (aksep.skår).

Vi har sett at det er store forskjeller på når et varsel når grensen for akseptabelt skår i enkeltsituasjoner. Spesielt varsler som ikke har overlapp gir dårlige skår for små terskelverdier, men skåren kan øke raskt dersom varslene er ”nærme”. Det er derfor viktig å se på en lengre periode med flere tilfeller for å kunne anslå på hvilken skala metoden kan brukes. Spesielt for høye terskelverdier ser vi at det er varslene med liten bias som gir best resultat, siden man da får minst innvirkning fra andre feilkilder (f. eks falske ekko). I alle figurene 5.1, 5.2, og 5.3 ser vi tydelig de falske ekkoene ved Gaustatoppen og Vassfaret, spesielt i figur 5.2 kommer disse tydelig frem. Falske ekko i disse områdene har ofte verdier opp mot 40 mm på 3 timer og kan derfor gi et betydelig bidrag til skårene for de høyeste terskelverdiene og persentilene.

5.2 Gjennomsnitt juli og august

Figur 5.4 og 5.5 viser FSS for persentilene 80, 90, 95 og 99 % og terskelverdiene 0,1, 0,5, 1 og 3 mm. Vi ser fortsatt på 3-timers akkumulert nedbør og gjennomsnittene er tatt over alle varslene i juli og august 2009, gjennomsnittet for juli og august hver for seg er også vist. Det er litt forskjell i resultatene for juli og august. August har noe bedre skår enn juli. Dette kan skyldes at modellen traff bedre på værtilstandene i august enn juli. Men hvis vi sammenligner AFSS for de to månedene (som sier noe om biasen) ser vi at juli har større bias i varslene og vi vet at dette også har innvirkning på resultatene (for terskelverdi 0,1 mm er $AFSS = 0,79$ for juli og $AFSS = 0,85$ for august), så dette kan også være en årsak. Tabell 5.1 viser når FSS når grensen for akseptabelt skår for de forskjellige persentilene. Som vi så i kapittel 5.1 kan det være stor forskjell fra et enkelt varsel til et annet og vi ser at gjennomsnittet for månedene også skiller en del. August når grensen for akseptabelt skår for mindre naboområder (10-35 km mindre) enn for juli.

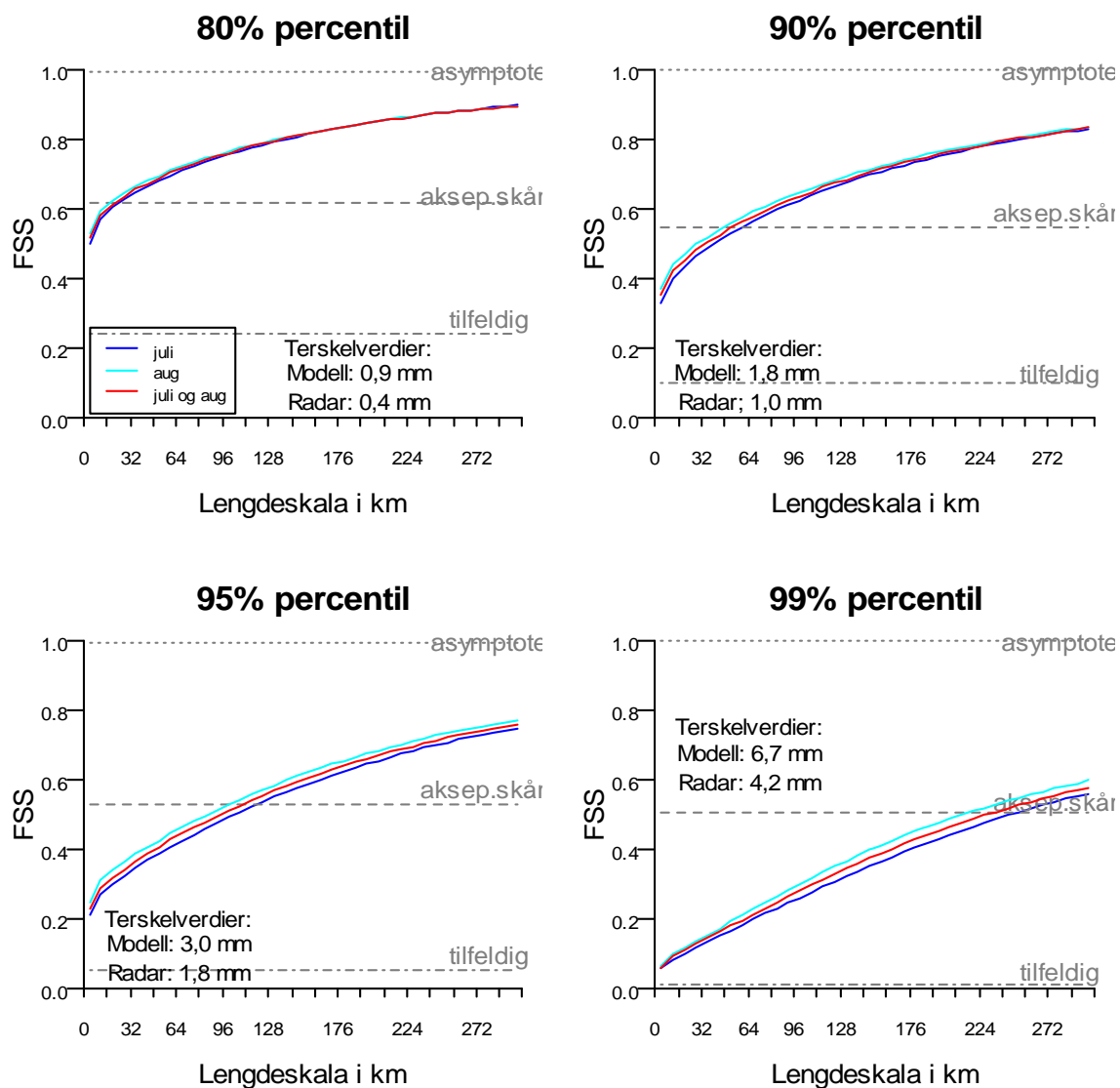
For de høyeste persentilene må vi opp i ganske store naboområder før vi når grensen for akseptabelt skår. Noe av dette kan forklares ved å se på hvilke terskelverdier disse høyeste persentilene representerer.

Persentil	Grense for når akseptabelt skår oppnås (km)		
	Juli	August	Alle varsler
80 %	28	20	25
90 %	60	45	50
95 %	116	98	105
99 %	252	220	235

Tabell 5.1: Viser for hvilket naboombåde FSS når akseptabelt skår for juli, august og totalt for begge månedene.

Terskelverdiene for 95 % persentilen er henholdsvis 3,0 mm for modellen og 1,8 mm for radarfeltet. For 99 % persentilen er terskelverdien 6,7 mm for modellen og 4,2 mm for radaren. Dette er høye terskelverdier over våre områder og representerer lokal nedbør som er vanskelig å varsle korrekt. Fra figur 5.4 og 5.5 ser vi at FSS når akseptabelt skår ved mindre naboombåde når vi bruker persentiler enn ved bruk av terskelverdier. For eksempel når ikke kurven for terskelverdien 3 mm akseptabelt skår i det hele tatt og da har vi valgt naboombåder opp til 300 km størrelse. Dette viser at biasen har stor innvirkning. Både persentilene og terskelverdiene viser at modellen verifiserer dårligere jo mer lokal nedbøren er.

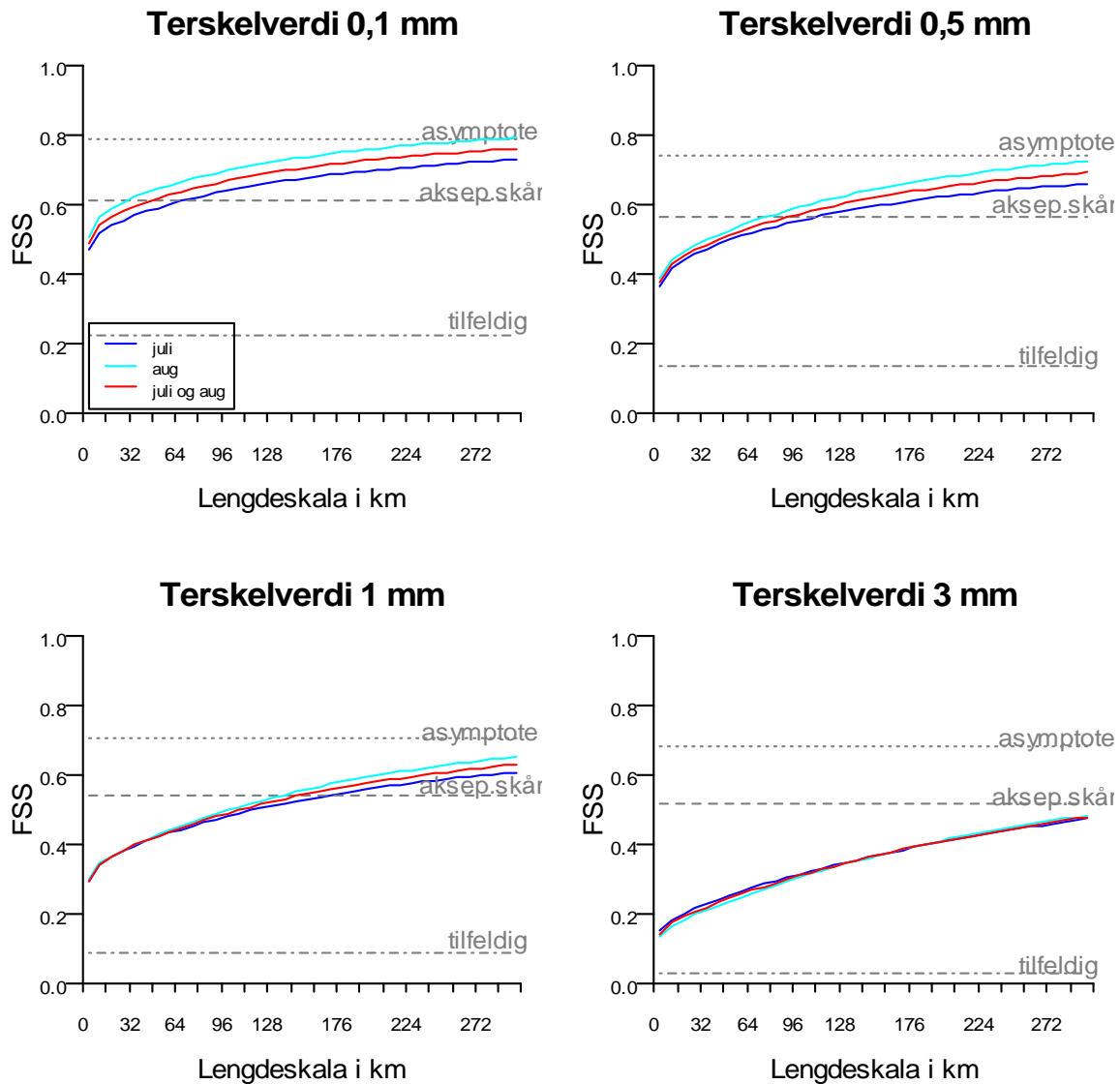
Figur 5.5 viser skårene for 4 forskjellige terskelverdier. Her kommer biasen inn og siden den ikke er neglisjerbar for våre varsler påvirker den skårene betydelig. AFSS, asymptoten, sier oss noe om biasen i varslene, jo lavere denne er, jo større er biasen. Vi ser at denne er rundt 0,8 for terskelverdi 0,1 mm og blir mindre for høyere terskelverdier. Biasen øker når terskelverdiene blir høyere. Kurvene når akseptabelt skår, totalt for alle varslene, omkring 44 km for terskelverdi 0,1 mm, 92 km for terskelverdi 0,5 mm og 156 km for terskelverdi 1 mm. For terskelverdi 3 mm når ikke kurven akseptabelt skår selv ikke for 300 kms naboombåde. Å se på terskelverdiene egner seg derfor ikke så godt for vårt område, i hvertfall ikke for høye terskelverdier. Selv om vi får noen høye verdier for det aller mest lokale regnet (99 % persentilen), gir metoden gode skår for 80, 90 og 95 % persentilene på mellom 30-100 km. For 90 % persentilen gir et naboombåde på 52 akseptabelt skår for FSS= 0,54 mot FSS=0,35 på gridskalaen (naboombåde 4 km), en forbedring mot verifikasjon på gridskalaen på over 50 %.



Figur 5.4: Sammenligning av FSS for juli og august og totalt for begge månedene. De respektive terskelverdier for hver persentil er kommentert på hver figur. Dette er terskelverdiene totalt for alle varslene samlet og kan avvike fra terskelverdiene for juli og august for seg. AFSS er 1 siden vi ser på persentiler og vi da har tilnærmet null bias mellom modell og radarfelt. De horisontale stiplede linjene er AFSS (asymptote), f_{aksep} (aksep.skår), $FSS_{tilfeldig} = f_R$ (tilfeldig).

Resultatene viser at skårene til metoden er avhengig av den romlige fordelingen av selve nedbøren. Det er lettere å forutsi med rimelig nøyaktighet sannsynligheten for utbredt nedbør enn mer lokal nedbør. Men det er som oftest det mer lokale regnet som gir mest nedbør og det er derfor ønskelig å finne ut hvor lokalt som er "for lokalt" til å kunne si noe generelt.

Hvis vi vurderer dette ut i fra persentilene, siden disse gir det beste bilde, vil grensen for terskelverdier mellom 1 og 3 mm (sett i forhold til modellen) ligge mellom 25-100 km. For høyere terskelverdier må vi opp mot 200 km.



Figur 5.5: Sammenligning av FSS for juli og august og totalt for begge månedene for terskelverdier. Her kommer biasen inn, jo lavere asymptote jo høyere bias. Asymptoten som er tegnet på figuren er for gjennomsnittet juli og august og kan avvike fra asymptoten for juli eller august hver for seg. De horisontale stiplede linjene er AFSS (asymptote), f_{aksep} (aksep.skår),

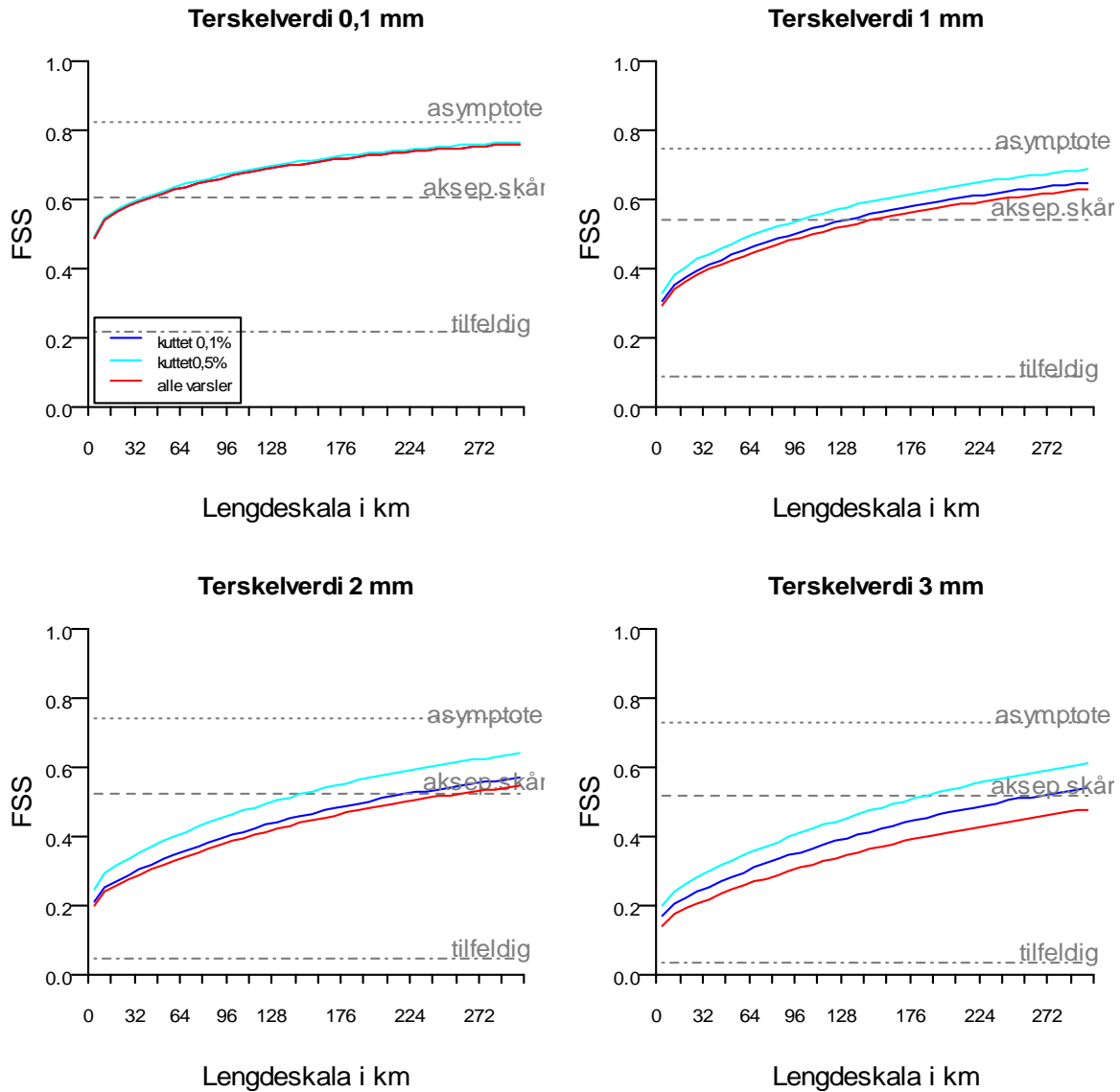
$$FSS_{tilfeldig} = f_R(\text{tilfeldig}).$$

5.3 Lav våtandel

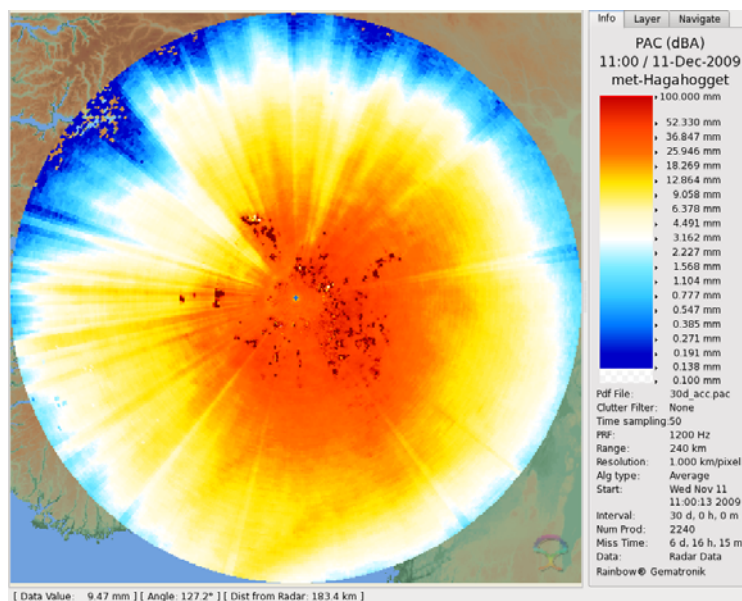
Våtandelen, f_R og f_M (ligning (8) og (9)), viser hvor mye nedbør som er i verifikasjonsområdet. Men med de usikkerhetene som ligger i radardataene (se appendiks) kan det se ut som resultatene er sensitive til varsler med liten utbredelse av nedbøren, og at dette kan innvirke på resultatene. I figur 5.6 er de varsler hvor radarfeltet har under 0,1 % og 0,5 % nedbør i forhold til hele feltet blitt fjernet. Det er ikke tatt hensyn til at enkelte av disse varslene kan være gode (noe de også er) der det er lite nedbør og varslet treffer godt. Allikevel ser vi at dette gir store utslag spesielt for terskelverdiene 1 mm og 3 mm. Når terskelverdien øker får vi flere varsler som har liten våtandel og flere varsler blir fjernet. For terskelverdi 1 mm forbedrer denne korrigeringen grensen for akseptabelt skår med hele 50 km og for 3 mm mellom 100 og 150 km. Dette viser at metoden er sensitiv til lav våtandel og at dette gir størst utslag for store terskelverdier, der våtandelen naturlig er mindre. På den annen side fjerner man på denne måten kunstig de ”dårlige” varslene og forbedrer skårene, så et bedre alternativ er å bruke persentiler så lenge biasen er stor.

En grunn til at vi forbedrer skårene radikalt ved å fjerne varsler med lav våtandel kan være innvirkningen av falske ekko. Vi vet at det ved en del av våre radarer er betydelige falske ekko som gir svært høye verdier. Når vi øker terskelverdigrensen og det blir færre gridruter med nedbør vil disse få stor innvirkning. I tillegg har de falske ekkoene størst innvirkning i situasjoner med lite nedbør fra før (personlig kommunikasjon Trygve Aas). Denne avbøyingen av radarstrålen som fører til at den treffer andre ting enn nedbøren (i disse tilfellene fjell) er størst i høytrykksituasjoner med pent og varmt vær (den relative fuktigheten avtar med høyden og temperaturen synker mindre med høyden enn i standardatmosfæren).

Som eksempel kan nevnes at over en 30 dagers periode (se figur 5.7) er akkumulert nedbør i området rundt Vassfaret ca 36 mm og de faste falske ekkoene i dette området gir over 100 mm (skalaen går ikke lenger enn til 100). Det samme forholdet gjelder også for de falske ekkoene rundt Gaustatoppen. Dette viser at de faste falske ekkoene kan gi et betydelig bidrag.



Figur 5.6: Sammenligning av FSS for et gjennomsnitt over alle varsler mot et utvalg av varsler der varsler med mindre enn 0,1 % og 0,5 % nedbør på radardataene er fjernet. Her kommer biasen inn, jo større bias jo lavere asymptote. De horisontale stiplede linjene er AFSS (asymptote), f_{aksep} (aksep. skår), $FSS_{tilfeldig} = f_R$ (tilfeldig).

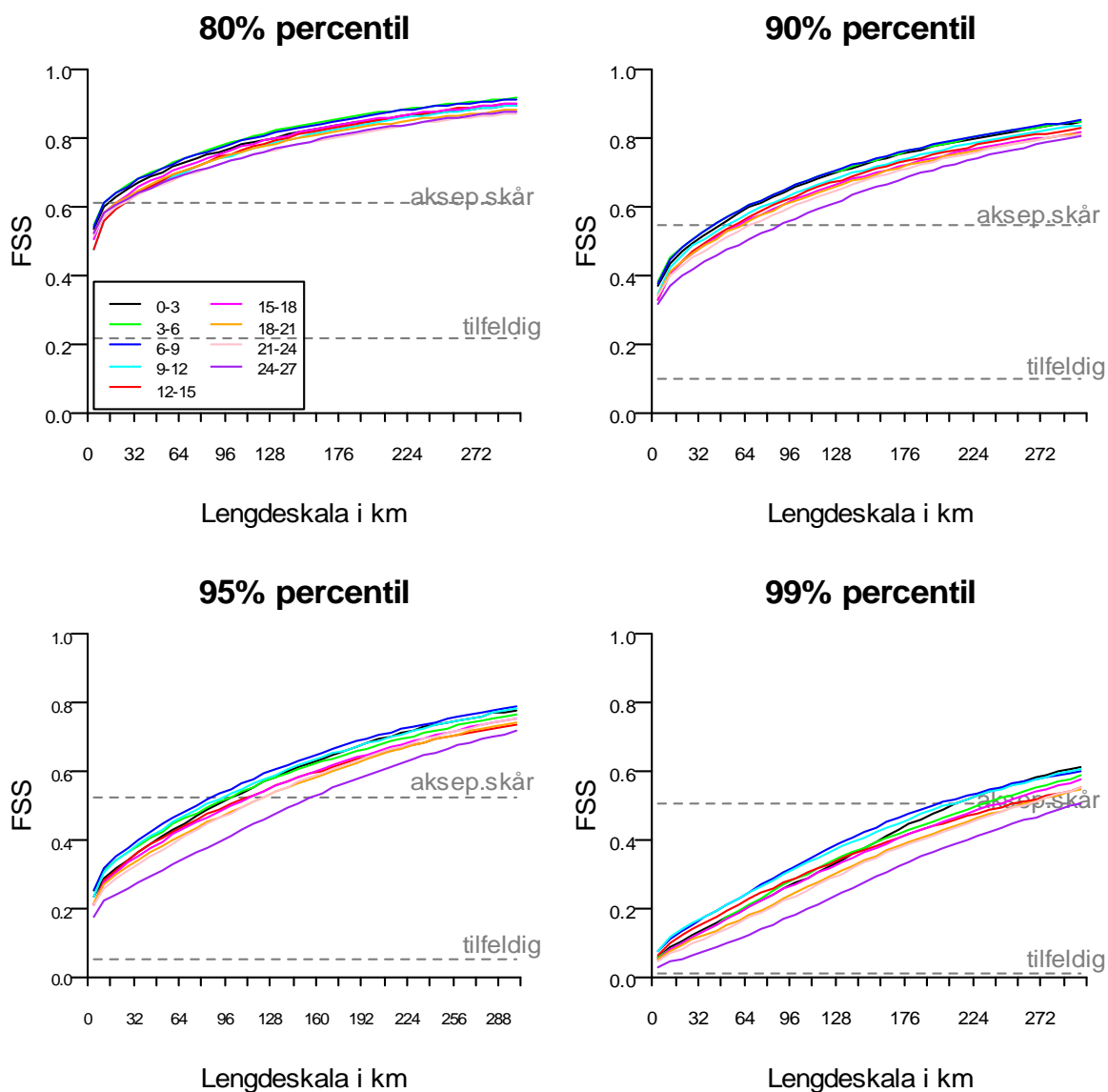


Figur 5.7: Akkumulert nedbør over en 30 dagers periode fra radaren Hagahogget. (6 dager og 16 timer mangler i akkumuleringen). De falske ekkoene kommer tydelig frem ved at de gir mer enn 100 mm nedbør. Skalaen stopper på 100 mm.

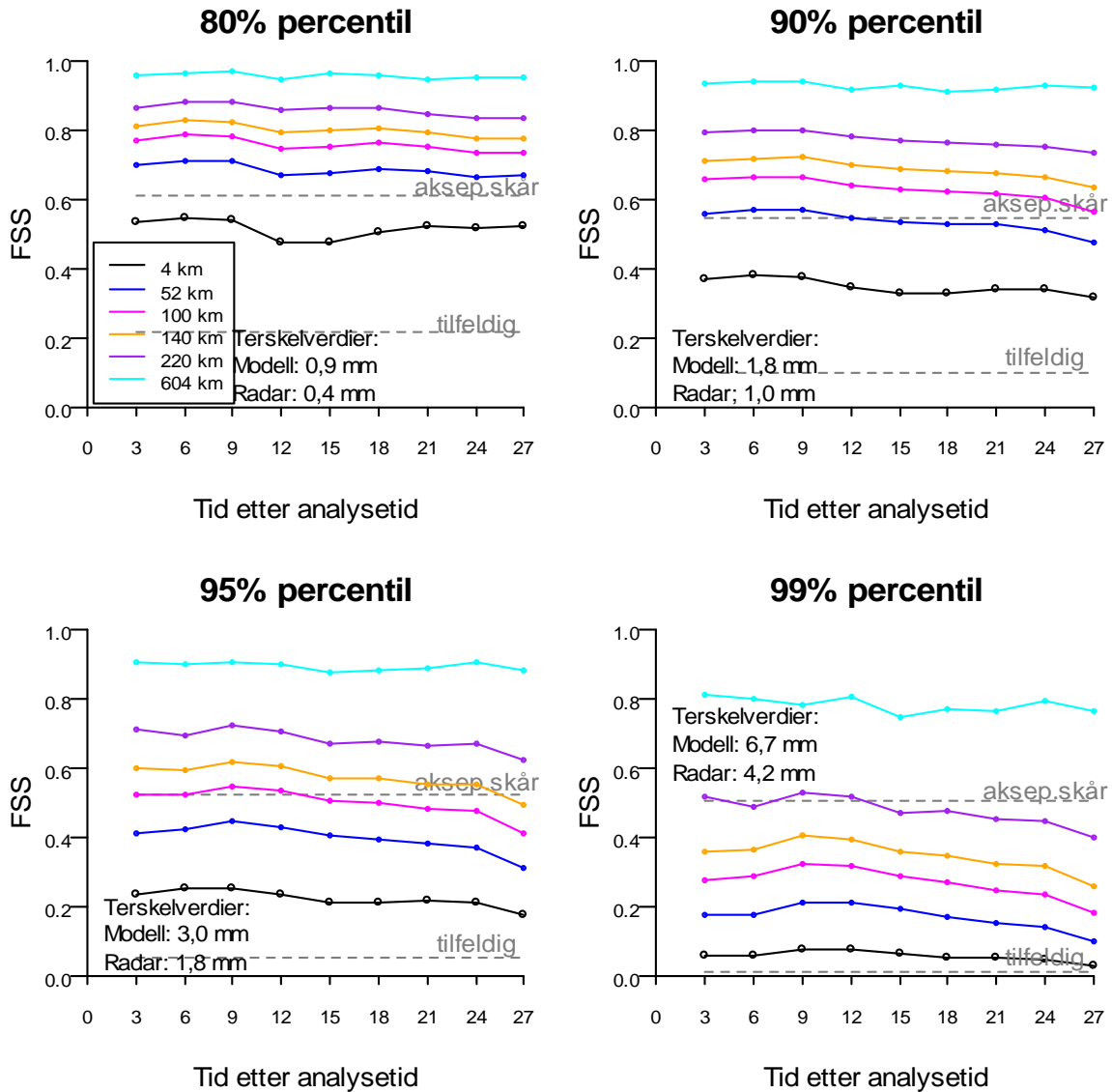
5.4 Prediktabilitet

For å se om metoden kan si noe om prediktabilitet har jeg sett på gjennomsnittet for juli og august for hvert tidssteg frem til 27 timer (3 timers akkumulert nedbør, alle kjøringene startet fra 00 UTC). Variasjon av FSS med lengden av varslet er vist i figur 5.8. De fire panelene i figur 5.8 viser FSS for 80 %, 90 %, 95 %, 99 % persentilen. 99 % persentilen representerer den mer lokale nedbøren og generelt mer lokale nedbørtilfeller som byer eller de lokale maksimumene innenfor større nedbørområder. 80 % persentilen representerer mere utbredt nedbør over større områder.

Figur 5.8 viser at varslets riktighet øker med økende naboerområde for alle persentiler. Vi ser at skåret minsker med varslets lengde bortsett fra de første timene frem til ca 6 timer. Dette ses tydeligere i figur 5.9.



Figur 5.8: Plott av FSS for 4 persentiler som funksjon av naboområdet for hvert varslingstidspunkt, med 3 timers intervaller. Gjennomsnitt av varsler for hele juli og august 2009. Kjøringene startet 00 UTC og kjørt 27 timer frem. De horisontale stiplede linjene er f_{aksep} (aksep.skår), $FSS_{tilfeldig} = f_R$ (tilfeldig).



Figur 5.9: Plott som viser forandring i FSS for 3 timers akkumulert nedbør. Hver graf viser naboområdene 4, 52, 100, 140, 220 og 604 km. De horisontale stiplede linjene er

$$f_{aksep} \text{ (aksep.skår)}, FSS_{tilfeldig} = f_R \text{ (tilfeldig)}.$$

Varslene mister en del prediktabilitet i løpet av 27 timer, og vi ser at prediktabiliteten er størst for varslene fra 6-9 timer etter starttidspunktet for kjøringen. Videre skjer en gradvis minking i skårene og et mere markert fall fra 24-27 timer. For lavere terskelverdier (80 % og 90 % persentilen) ser vi også at vi har en forbigående minking i prediktabiliteten på ettermiddagen, mest markert for lave persentiler, spesielt 12 UTC og 15 UTC terminen, men også til en viss grad 18 UTC. Dette kan skyldes konvektiv nedbør pga soloppvarming. Også

her ser vi at modellen gir best skår for utbredt nedbør (80 % persentilen) i forhold til mer lokal nedbør (99 % persentilen). Et naboområde på 52 km gir en FSS på mellom 0,1 og 0,2 for 99 % persentilen og mellom 0,65 og 0,70 for 80 % persentilen. Vi ser også at på gridskalaen (naboområde 4 km) er kurvene noe flatere med tiden, spesielt for 99 % persentilen. For det største naboområdet (604 km) ser vi også at kurven er nærmest flat og ikke mister mye skår med tiden. Det største fallet i FSS med tiden har vi for de midterste naboområdene (50 km til 220 km). Roberts (2008) antyder at dette kan være de skalaene hvor dataassimilasjon tilfører mest skår.

I tabell 5.1 ser vi når de forskjellige persentilene når grensen for akseptabelt skår for det første varslet (0-3 timer), tidsintervallet 6-9 timer og det siste varslet (24-27 timer). Vi ser at forholdet mellom tidsperioden med høyest prediktabilitet og det siste varslet ligger på mellom 1,5 til 3.

Persentil grense	Skala hvor akseptabelt skår oppnås (km) +/- 10 km			Tilnærmet forhold mellom beste og siste varslet
	Første varslet 0-3 t	Beste varslet 6-9 t	Siste varslet 24-27 t	
99 % (lokal nedbør)	212	196	300	~1,4
95%	100	90	156	~1,6
90%	60	44	110	~1,8
80% (utbredt nedbør)	12	12	36	~3,0

Tabell 5.1: Viser på hvilken skala første, beste og siste varslet når akseptabelt skår for 4 persentiler, 80 %, 90 %, 95 % og 99 %. Hvilke terskelverdier disse persentilene representerer er skrevet på figur 5.14.

5.5 Stort mot lite område

To mindre områder er også undersøkt, et innenfor Oslo radaren og et innenfor Rissa radaren. Dette illustrerer hvordan metoden fungerer for mindre områder og vil vise eventuelle forskjeller i radardataene. Området innenfor Osloradaren og Rissaradaren er valgt fordi

Osloradaren er den eldste radaren vi har med mye falske ekko og blokkeringer, mens Rissa er den radaren som ligger gunstigst til med minst blokkeringer pga topografi og er av nyere dato. Bildene i figur 5.10 er hentet fra disse radarene og viser akkumulert nedbør for 30 dager. Figuren gir et bilde av de blokkeringer som finnes på disse to radarene.

Som man ser av bildet fra Rissaradaren (til venstre) er verdiene tilnærmet sirkulære, dvs at det er lite eller ingen blokkeringer. Bilde fra Osloradaren (til høyre) er mere "frynsete". Mot NW fra senteret kan vi tydelig se at det er en blokkering. Ikke en fullstendig blokkering, men en partiell. Også mot NE er det en partiell blokkering. Den smale blokkeringen mot SW er trolig en Telenor mast som står noen 100 m fra Osloradaren. Rissaradaren (2002) har i praksis ikke blokkeringer. Den ligger høyt og fritt på Olsøyheia 611 moh. Osloradaren (1989) derimot har dårligere oppløsning og blokkeringer. For å luke ut faste falske ekko benytter Rissa radaren et dopplerfilter. Med det kan man fjerne ekko som ligger fast. Oslo radaren derimot har cluttermap (se kapittel 4.2).

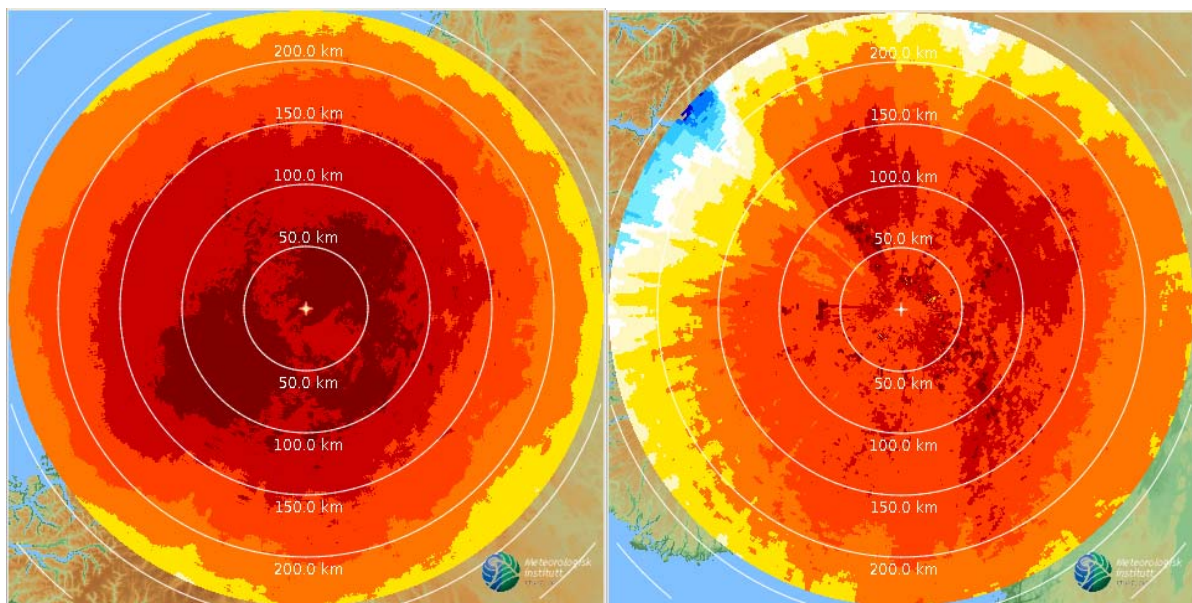
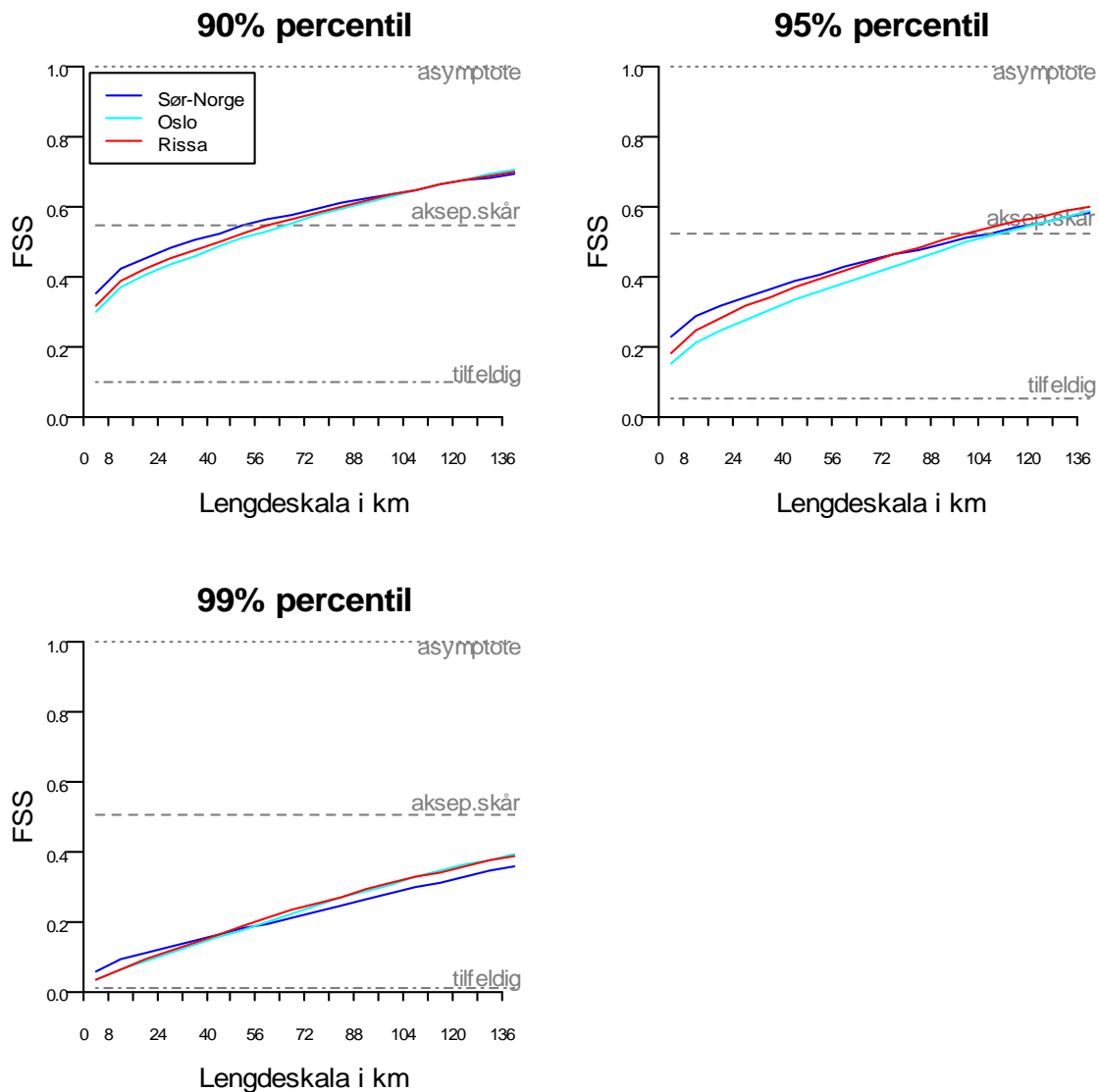


Fig 5.10: 30 døgn akkumulert nedbør for Rissa radaren (til venstre) og Oslo radaren (til høyre).

Som tidligere nevnt er radardataene brukt i denne oppgaven helt uten noen form for korrigering, og med de topografiske forhold vi har i Norge var det ikke uventet at radaren ville gi færre verdier enn modellen siden vi har en del blokkeringer på nesten alle radarene,

og i tillegg dekker radarene dårlig i fjellet. Vi har også en del falske ekko, som øker nedbørverdiene vi får fra radarene. Eksempel på slike falskt ekko som gir noen få gridruter med høye verdier finner vi for eksempel innenfor Osloradaren i området rundt Vassfaret og rundt Gaustatoppen. Vi har tatt ut et mindre område innenfor hver av radarene på 80x80 gridruter (320x320 km), for å se om disse områdene verifiserer forskjellig fra hverandre, men også om for å se om små områder verifiserer forskjellig fra hele Sør-Norge utsnittet (figur 4.1).

I figur 5.11 ser vi FSS for de to mindre områdene og hele Sør-Norge utsnittet. Vi får ikke de store forskjellene, men vi ser at Rissaområdet har litt bedre skår enn Osloområdet for de fleste naboområder. For alle persentilene ser vi at de små verifikasjonsområdene får dårligere skår enn hele Sør-Norge utsnittet, men når naboområdet øker verifiserer de små områdene bedre enn det store. Dette er naturlig siden naboområdet da blir stort i forhold til det totale området det verifiseres over og man nærmer seg raskere den asymptotiske verdien ($AFSS=1$ siden vi ser på persentiler). Osloområdet når akseptabelt skår ved et naboområde på omkring 84 km for 90 % persentilen, Rissaområdet omkring 80 km og hele Sør-Norge utsnittet ved omkring 68 km. De små områdene når grensen for akseptabelt skår på en skala 10-15 km høyere enn for hele Sør-Norge utsnittet. For 95 % persentilen er tallene 116 km for Osloområdet, 104 km for Rissaområdet og 100 km for hele Sør-Norge utsnittet.



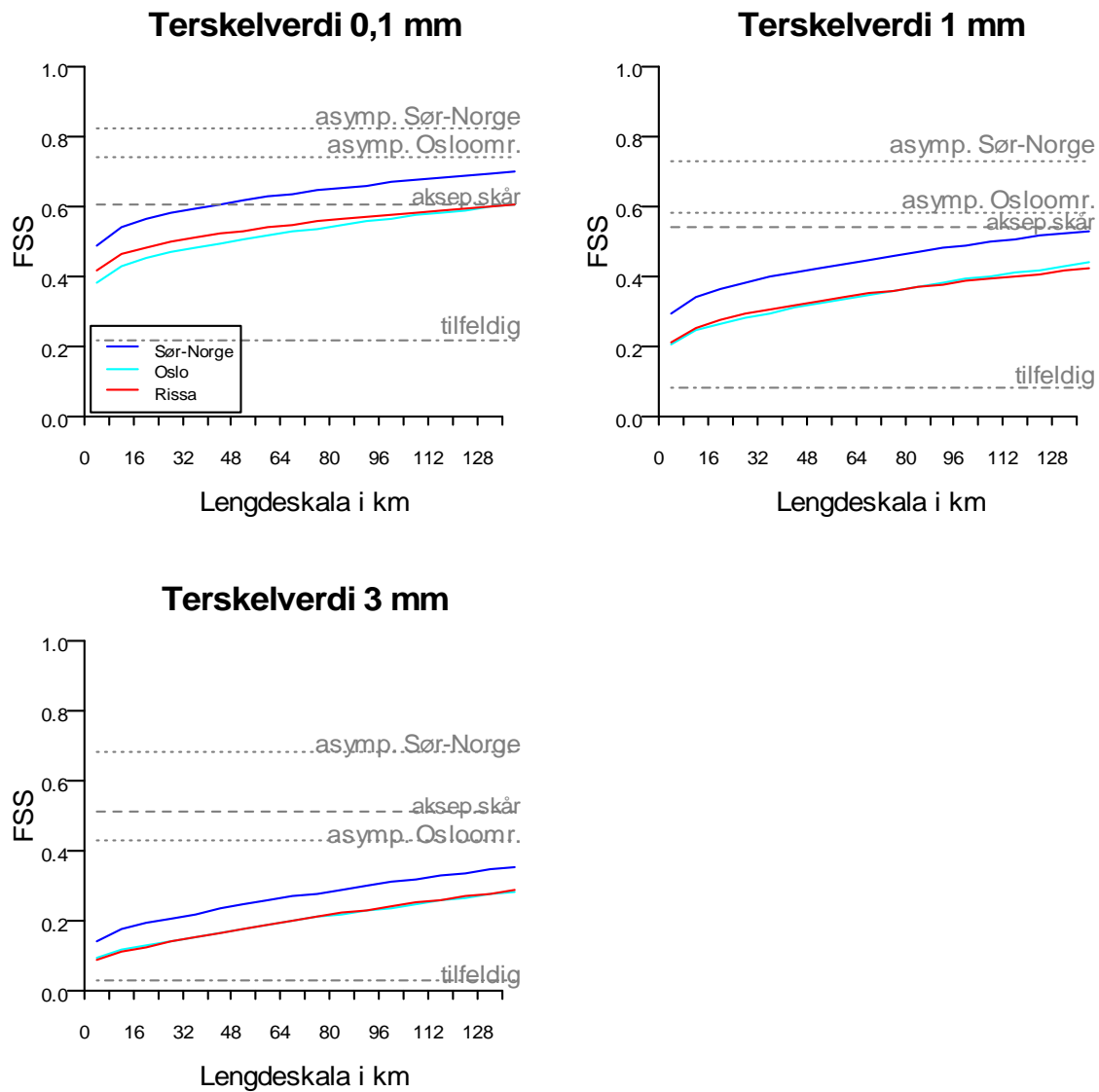
Figur 5.11: Plott av FSS for persentilene 90 %, 95 % og 99 % for Rissaområdet, Osloområdet og hele Sør-Norge utsnittet. De horisontale stiplede linjene er AFSS (asymptote),

$$f_{aksep} \text{ (aksep. skår)}, FSS_{tilfeldig} = f_R \text{ (tilfeldig)}.$$

Når vi bruker persentiler ser vi bort fra biasen, så det er ikke uventet at vi da ikke finner de store forskjellene mellom Osloområdet og Rissaområdet (figur 5.11). Ved bruk av terskelverdier har biasen større innvirkning og man kan forvente å finne større forskjeller mellom de to områdene. Men som vi ser i figur 5.12 er det heller ikke her store forskjeller. Årsaker til dette kan være at Osloradaren har betydelig falske ekko som gir et tilskudd til

nedbøren, Rissaradaren derimot har mindre falske ekko. I tillegg har Osloradaren en del blokkeringer som minsker nedbørverdiene, noe Rissaradaren ikke har (se figur 5.10). Muligens oppveier feilkildene på Osloradaren hverandre noe (falske ekko øker nedbørverdiene, mens blokkeringer minsker dem). Men til tross for at forskjellen ikke er betydelig kan vi se at både ved bruk av terskelverdier og persentiler har Osloradaren litt dårligere skår enn Rissaradaren.

For terskelverdiene, figur 5.12, ser vi derimot større forskjell mellom de små områdene og Sør-Norge utsnittet enn vi gjorde for persentilene. For terskelverdi 0,1 mm skiller grensen for når områdene når akseptabelt skår med over 80 km (mot 16 km for 90 % persentilen). En årsak kan være at for små verifikasjonsområder får vi mange varsler med liten våtandel. På figuren er asymptotisk skår for Sør-Norge utsnittet og Osloområdet tegnet inn og gir oss et bilde på biasen. Grafen viser at biasen er større for de små områdene og vi får derfor totalt lavere skår for FSS. Forskjellen i asymptotene mellom Sør-Norge utsnittet og Osloområdet blir større med større terskelverdi. Spesielt for Osloområdet har radaren nesten alltid noen gridruter med nedbør selv om modellen ikke har noen, sannsynligvis skyldes en del av disse falske ekko, og dette gjør seg utslag i stor bias. Resultatene viser at skårene til metoden er avhengig av domenet varslet blir evaluert over. Det er lettere å forutsi med rimelig nøyaktighet sannsynligheten for nedbør over et stort område enn et lite.



Figur 5.12: Plott av FSS for terskelverdiene 0,1 mm, 1 mm og 3 mm for Rissa området, Osloområdet og hele Sør-Norge utsnittet. De horisontale stiplede linjene er AFSS (asymptote), f_{aksep} (aksep. skår), $FSS_{tilfeldig} = f_R$ (tilfeldig). Asymptotisk grense for Sør-Norge utsnittet og Osloområdet er begge tegnet inn. Rissa området har ganske lik asymptotisk grense som Osloområdet bortsett fra for terskelverdien 3 mm, der Rissa har litt høyere, 0,48.

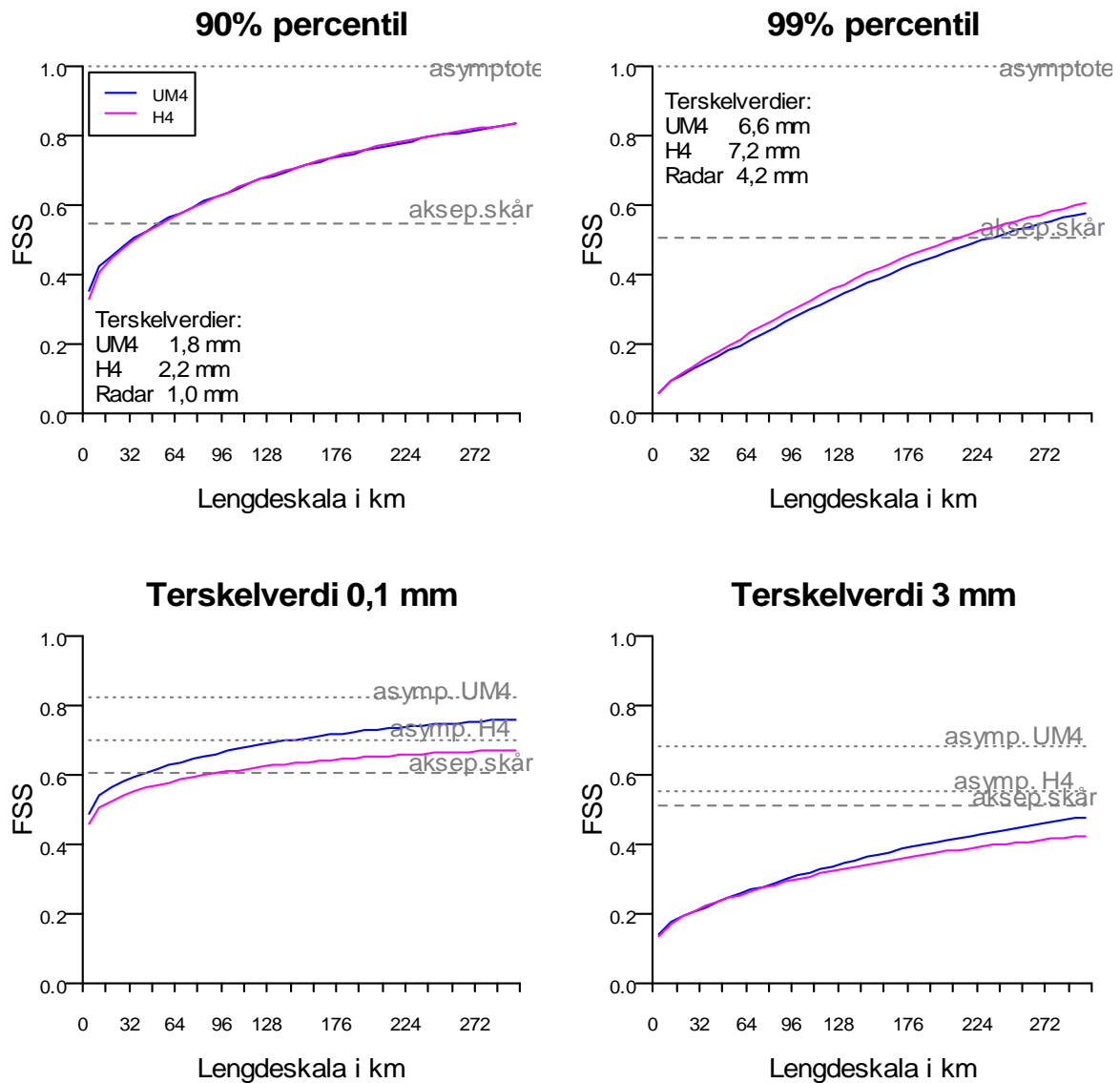
5.6 Sammenligning HIRLAM4 mot UM4

I Lean et al. (2008) ble metoden testet på UM modellen med 1, 4 og 12 km oppløsning. Det er derfor ikke gjort her, men vi har valgt å teste den mot ”High Resolution Limited Area Modell” (HIRLAM) som også har 4 km oppløsning (Undén et al., 2002). Denne brukes i dag operativt på Meteorologisk Institutt. HIRLAM er en hydrostatisk modell som ikke er tenkt brukt for 4 km oppløsning (HIRLAM4) når det gjelder nedbør. HIRLAM4 bør være dårligere på nedbør siden denne for 4 km oppløsning er parametrisert som HIRLAM 12 km.

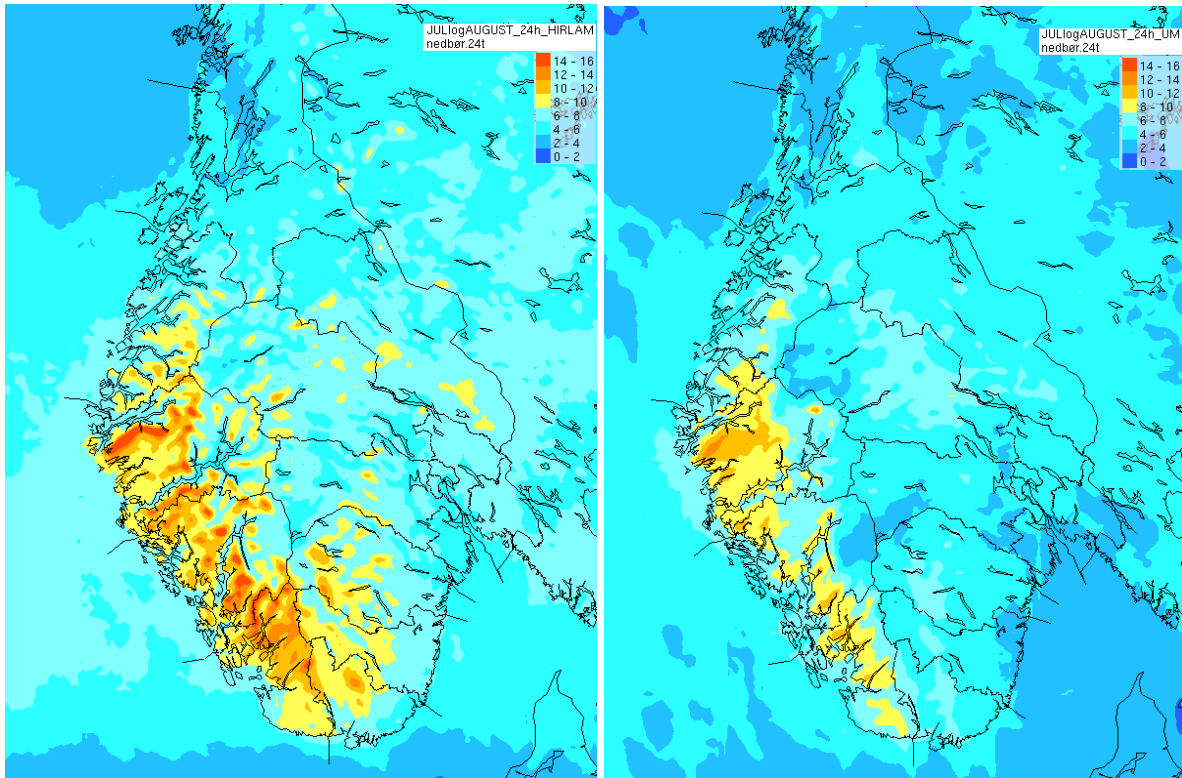
I figur 5.13 ser vi at UM4 er en tanke bedre enn HIRLAM4 opp til naboområdet på ca 80-100 km hvis vi ser på 90 % persentilen, deretter er HIRLAM4 bedre, men forskjellene er små. For 99 % persentilen er modellene like for de aller minste naboområdene, ellers er HIRLAM4 bedre. HIRLAM4 er som nevnt en hydrostatisk modell og ”glatter” feltet mer enn UM4. I figur 5.14 er det vist gjennomsnittlig 24 timers nedbør for UM4 og HIRLAM4 modellen. Gjennomsnittet er tatt over hele juli og august 2009. Dette viser at HIRLAM4 har mer utbredt nedbør og høyere verdier enn UM4 for den perioden vi verifiserer over. Dette kan vi også se ut fra terskelverdiene vi får ved bruk av persentiler (disse er skrevet på figur 5.13). Det er derfor ikke overraskende at HIRLAM4 skårer bedre for høye persentiler, både fordi HIRLAM4 har større utbredelse av nedbøren, men også fordi HIRLAM4 generelt har høyere nedbørverdier enn UM4.

Ved bruk av terskelverdier (nederst i figur 5.13) er bildet litt annerledes. Her verifiserer UM4 betydelig bedre enn HIRLAM4 for lave terskelverdier mens forskjellen er mindre for høyere terskelverdier (UM4 når akseptabelt skår for et naboområde på 45 km mens HIRLAM4 må opp i ca 92 km.) Dette er fordi UM4 har mindre bias enn HIRLAM4 og derfor skårer bedre når det gjelder terskelverdier. Et varsel med stor bias vil ikke skåre bedre enn et med mindre når vi utvider naboområdene.

Det må også nevnes at 4 km ikke er en optimal oppløsning på modellen for varsling av konvektiv nedbør (Lean et al. 2008).



Figur 5.13: Plott av FSS mot naboområde for UM4 og HIRLAM4 (H4), for to persentiler og to terskelverdier. De horisontale stiplede linjene er AFSS (asymptote), for både UM4 og HIRLAM4 og f_{aksep} (aksep. skår).



Figur 5.14: Gjennomsnittlig akkumulert 24 timers nedbør for juli og august 2009 for HIRLAM4 modellen til venstre og UM4 til høyre (Morten Køltzow)

6. Metoden som sannsynlighetsvarsel

Nedbørvarsler fra mesoskala numeriske værmodeller inneholder ofte trekk som ikke er deterministisk varslingsbare og krever en sannsynlighetstilnærming. Derfor har man de senere årene begynt å bruke mer ensembler og sannsynlighetsberegninger for varsling av nedbør. Ensembler for langtidsvarsling har vært brukt en god stund, men bruk av ensembler for finskalamodeller er fortsatt et kostnads- og ressurs spørsmål og det kan ennå gå noe tid før dette er vanlig i daglig værvarsling. Fordelen med å lage et sannsynlighetsvarsel er at en avgjørelse i forhold til en hendelse kan baseres på brukeren og hans behov og gi varsleren bedre grunnlag for å vurdere hvor sikkert akkurat dette varslet er.

Sannsynlighetsvarsler/betraktninger kan gjøres på forskjellige måter:

- *Ensemblekjøringer*: krever mange kjøring og for finskalamodeller krever dette stor datakapasitet og indirekte store kostnader.
- *Subjektive vurderinger*: krever økte menneskelige ressurser, da samfunnet krever mer og mer lokalt tilpassede varsler.
- *Skjemaer basert på statistikk*: krever ikke mye kostnader, men krever til gjengjeld en lang datarekke av modellkjøringer som ikke alltid er tilgjengelige.

Ensemblekjøringer krever mye regnekapasitet. Subjektiv tilnærming krever store menneskelige ressurser. Skjemaer basert på historiske data er kostnadsbesparende, men krever lange rekker med data som ikke alltid er tilgjengelig. Metoden beskrevet i denne oppgaven kan brukes på et deterministisk varsel i etterkant og vi får ut en sannsynlighet for hvert enkelt gridpunkt som baserer seg på verdiene i området rundt punktet (ligning (12), kapittel 3). Dette baseres på hypotesen om at de finskalerte modellene er gode, men at hendelser plasseres litt feil i tid eller rom. Verdien man får fra den finskalerte modellen blir ikke sett på som en eksakt verdi, men mer som en indikasjon på hva som vil skje i området rundt. Hensikten med denne postprosesseringen er å kunne gjøre mer nytte av finskalamodellene uten for store kostnader. For å ta en optimal beslutning vil et sannsynlighetsvarsel gi mer fleksibilitet enn et kategorisk varsel, og en beslutning trenger ikke bare avhenge av varslers synspunkt, men også på brukeren og hans behov. Det

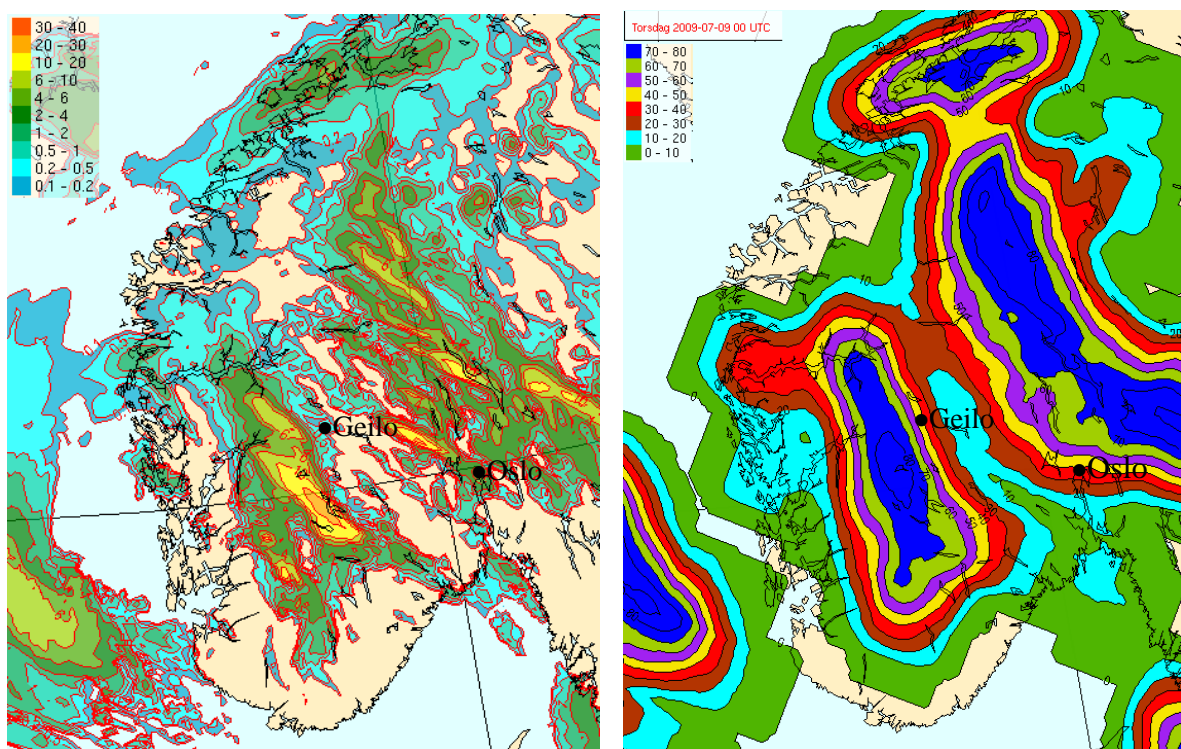
viktige er å finne for hvilken skala (naboområde) metoden verifiserer godt nok. Da kan vi bruke denne skalaen til å lage slike sannsynlighetsvarsler.

I resten av kapitlet vil vi vise noen eksempler på hvordan sannsynlighetstilnærmingen ved bruk av RL08 kan brukes.

I kapittel 5 fant vi at et naboområde på mellom 50-100 km kan være en brukbar skala for et gjennomsnitt av nedbørtilfeller. I figur 6.1 er RL08 brukt til å utlede et sannsynlighetsvarsel for 8.juli 2009 for et naboområde på 76 km. Den deterministiske kjøringen fra UM4 vises til venstre i figuren, til høyre sannsynligheten for å overstige 1 mm for samme tidspunkt.

Figuren viser 3 timers akkumulert nedbør for tidspunktet 21-00 UTC. Modellkjøringen er startet 00 UTC 8.juli. Sannsynligheten refererer til hver enkelt rute i modellen.

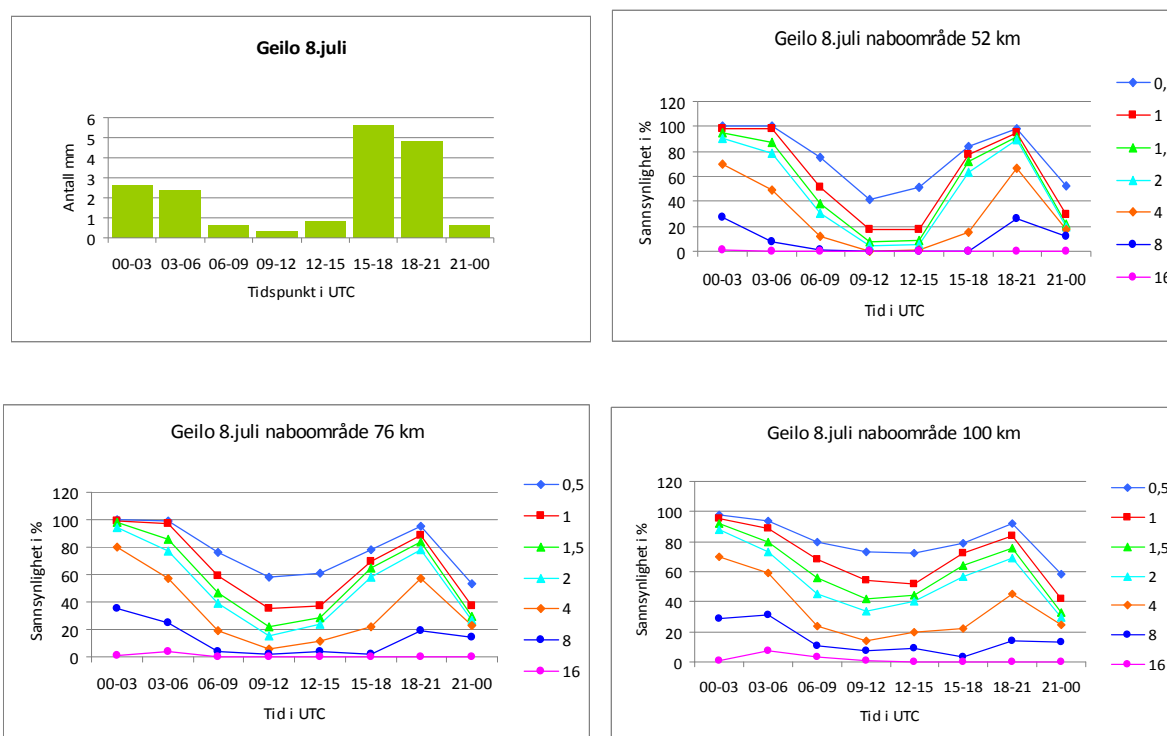
Sannsynlighetsvarslet er ”smurt” mer utover enn det deterministiske varslet. Allikevel blir ikke de romlige detaljene fra det deterministiske varslet helt borte ved denne postprosesseringen, men de er fremstilt på en ny måte som usikkerheter.



Figur 6.1: Til venstre: Eksempel på en modellkjøring fra UM4, 3 timers akkumulert nedbør fra 21 UTC 8.juli 2009 tom 00 UTC 9.juli 2009. Til høyre: Sannsynlighetsvarsel utledet fra UM4 modellen ved bruk av RL08. Sannsynligheten i % for å overstige 1 mm i løpet av en 3 timers periode. Naboområde er 76 km.

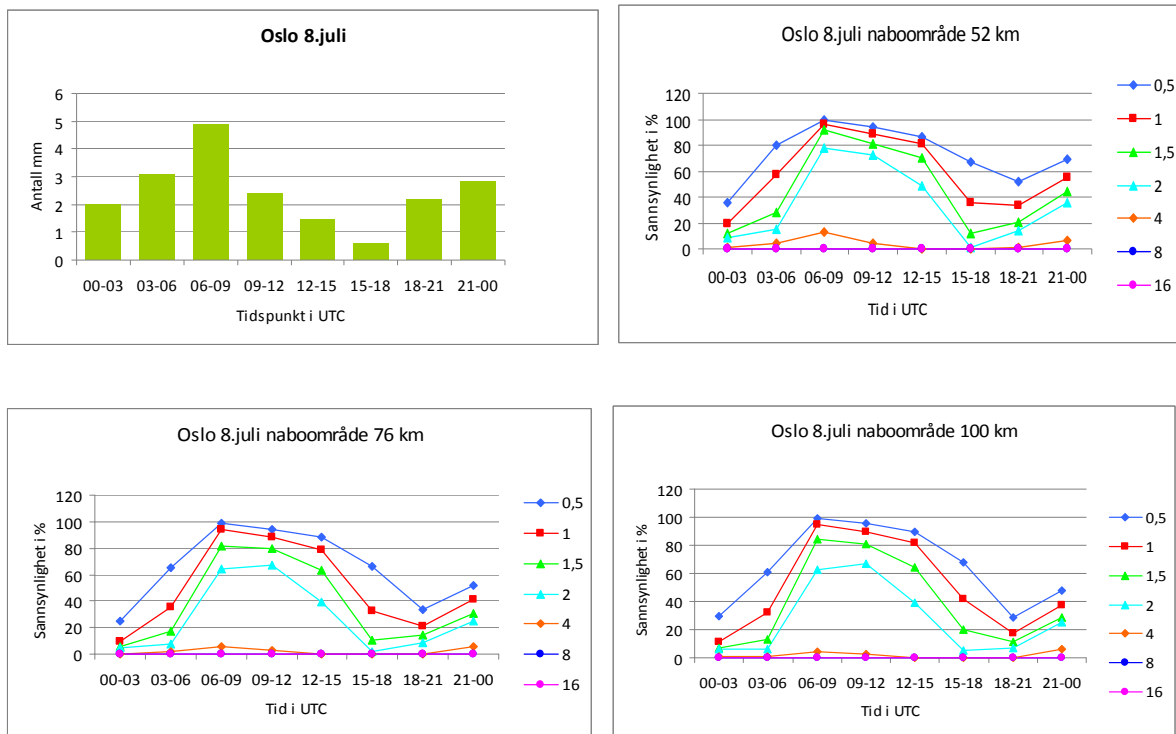
En vanlig bruker uten meteorologisk bakgrunn vil ofte bruke et varsel ved å se på punktet nærmest det han er interessert i. I figur 6.2 og 6.3 er det vist et varsel for to gridpunkter, Oslo og Geilo (meteogram) for den aktuelle dagen. Det er også laget sannsynlighetsdiagrammer for flere terskelverdier og naboombråder ved bruk av RL08. En slik type sannsynlighetsdiagram kan være til hjelp for en bruker når beslutninger skal tas. En dristig bruker vil kun ta forholdsregler når varslet indikerer en høy sannsynlighet, mens en mer forsiktig bruker vil reagere ved mye mindre sannsynligheter. Alt avhengig av hva man risikerer og kostnader forbundet med beslutningstakingen.

For Geilo, figur 6.2, ser vi at sannsynlighetene øker for de fleste terskelverdiene når vi utvider naboombrådet, men det er også enkelte terminer og enkelte terskelverdier hvor sannsynligheten ikke øker selv om vi utvider naboombrådet. For eksempel øker sannsynligheten for å overstige 1 mm lite i perioden 15-18 UTC selv om vi utvider naboombrådet fra 52 km til 100 km. I tidsrommet 9-15 UTC derimot er det store forskjeller i sannsynlighet avhengig av hvor stort naboombråde vi velger.



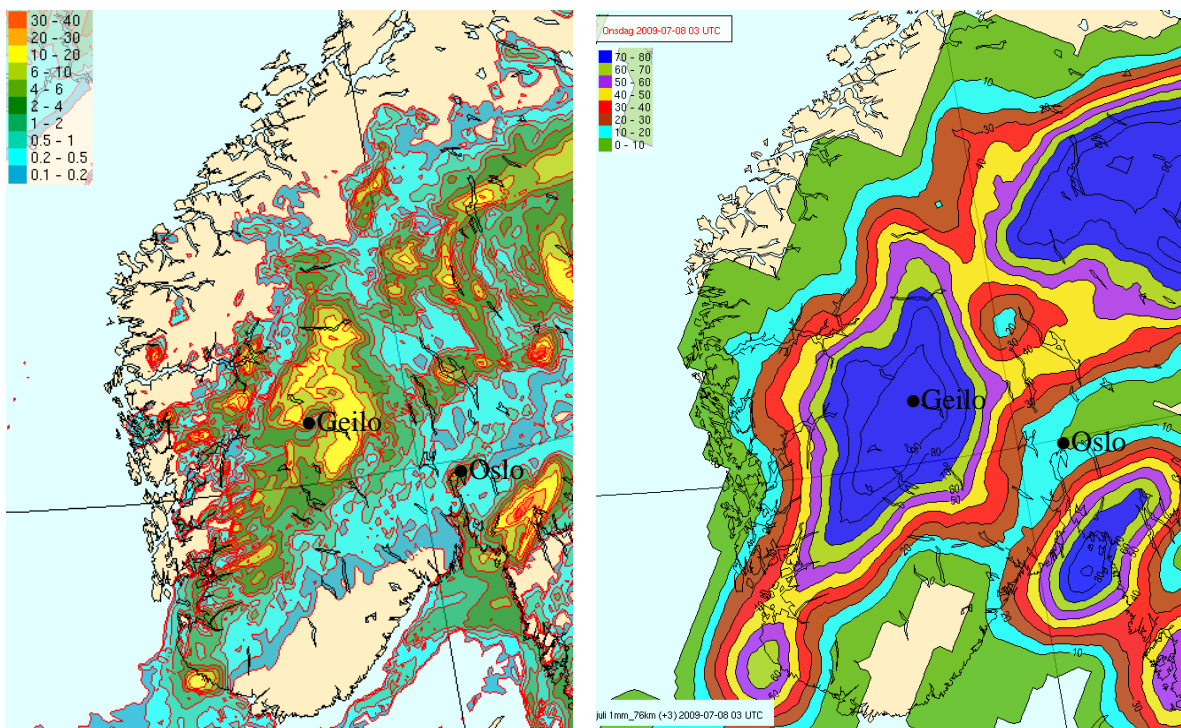
Figur 6.2: Meteogram fra UM4 modellen (øverst til venstre) og sannsynlighetsdiagrammer (resten av figurene) for punktet Geilo. Varslene viser 3 timers akkumulert nedbør 8.juli 2009, kjøringen starter 00 UTC. Sannsynlighetsdiagrammene viser 7 forskjellige terskelverdier i mm og 3 forskjellige naboombråder.

For Oslo, figur 6.3, er sannsynlighetsverdiene mer like selv om vi utvider naboområdet. For tidspunktet, 21-00 UTC (også illustrert ved kart i figur 6.1), har Oslo 2,8 mm nedbør og Geilo 0,6 mm, allikevel viser sannsynlighetene i figur 6.2 og 6.3 ganske like verdier. Omtrent 40 % sannsynlighet for å overstige 1 mm både for Geilo og Oslo (naboområde 76 km og 100 km, for 52 km har Geilo 30 % og Oslo 60 %). For begge stedene ser vi at sannsynlighetene også kan bli lavere når vi utvider naboområdet. Et eksempel på dette er Oslo for tidsrommet 18-21 UTC. Her går sannsynligheten ned for de fleste terskelverdiene når vi utvider naboområdet fra 52 km til 100 km. Dette viser at man har flere gridruter med verdien null, eller lave verdier, i området rundt punktet og derfor får lavere sannsynlighet selv om man utvider området rundt punktet.



Figur 6.3: Meteogram fra UM4 modellen (øverst til venstre) og sannsynlighetsdiagrammer (resten av figurene) for punktet Oslo. Varslene viser 3 timers akkumulert nedbør 8.juli 2009, kjøringen starter 00 UTC. Sannsynlighetsdiagrammene viser 7 forskjellige terskelverdier i mm og 3 forskjellige naboområder.

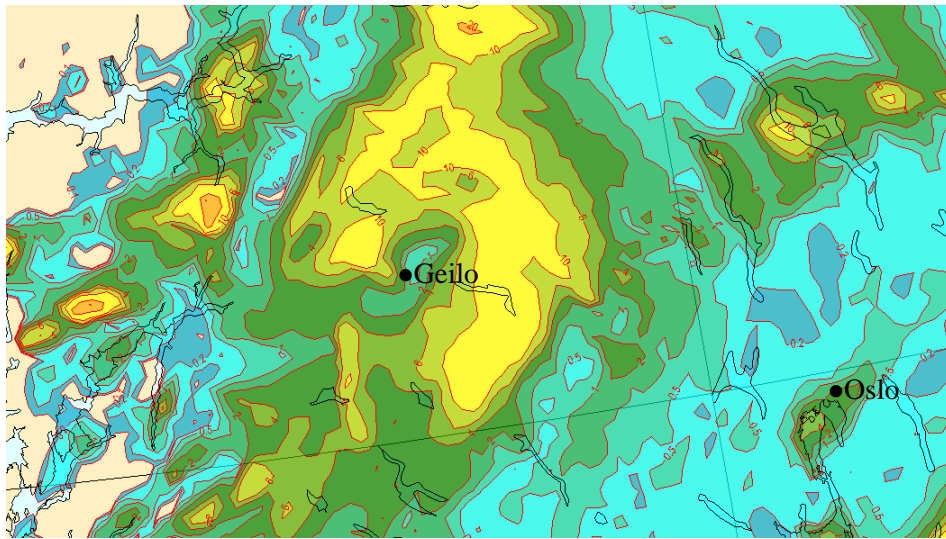
Sannsynligheten kan være størst for små naboområder og små terskelverdier, men også størst for høye terskelverdier innenfor samme tidsintervall. Dette viser hvor stor variasjoner det er i finskalamodeller over små områder. Det er derfor viktig å finne den riktige størrelsen på naboområdet som gjenspeiler usikkerheten i de finskalerte modellene, men ikke gletter feltet for mye. På figur 6.4 ser vi 3 timers akkumulert nedbør for samme dag som de andre figurene men for tidspunktet 03 UTC. Fra meteogrammet på figur 6.2 og 6.3 har Geilo 2,7 mm nedbør og Oslo 2,0 mm for dette tidspunktet. Sannsynlighetsdiagrammet i figur 6.2 og 6.3 gir 98 % sannsynlighet for å overstige 1 mm nedbør for Geilo og 10 % for Oslo for naboområde 76 km. Sannsynlighetene er veldig forskjellige selv om verdien fra den deterministiske modellen er ganske like.



Figur 6.4: 3 timers akkumulert nedbør fra 8.juli 03 UTC. Til venstre UM4 modellen. Til høyre sannsynlighetsvarslet for samme tidsrom med naboområde 76 km. Viser sannsynligheten for å overstige 1 mm.

Dette viser at nedbøren har større utbredelse rundt Geilo med flere høye verdier, mens rundt Oslo må det være mindre nedbør og/eller områder med lavere verdier i nærheten. Dette ser vi igjen i figur 6.4. Vi ser at modellen hadde mer nedbør i et lite område rundt Oslo som ga en såpass høy verdi som 2,0 mm på meteogrammet, selv om det egentlig er lite nedbør i området rundt og dette fører til en lav sannsynlighet. Geilo ligger inne i nedbørområdet, så

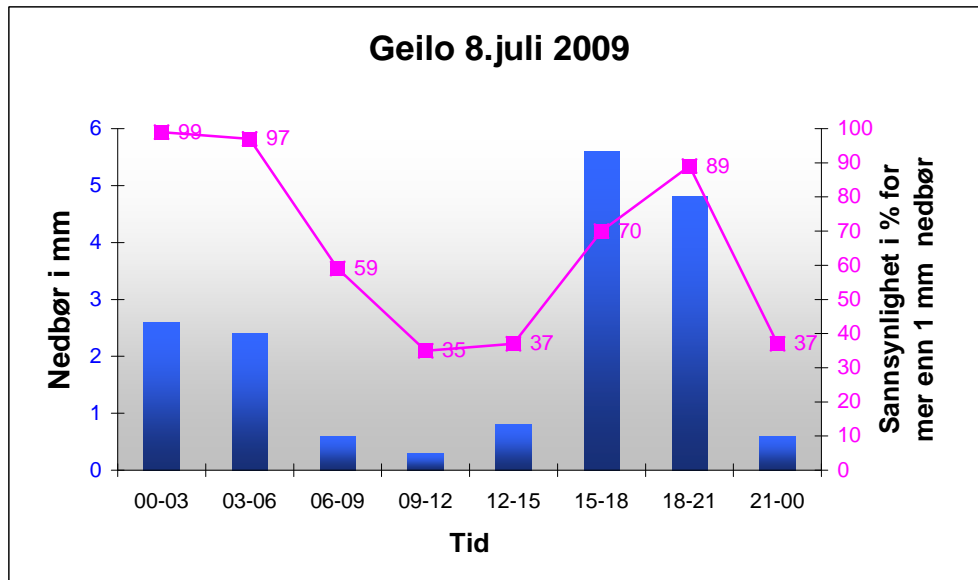
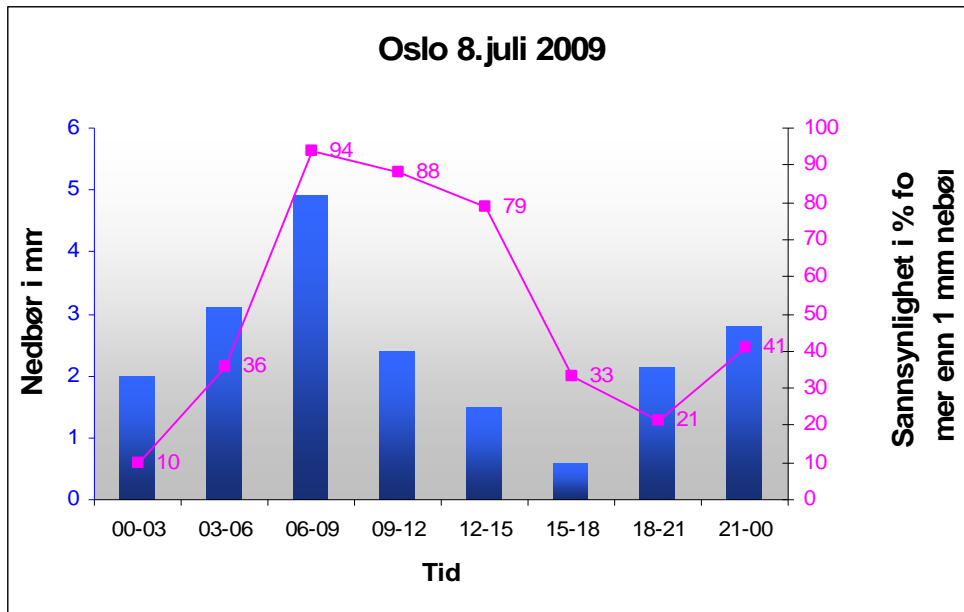
her blir sannsynligheten større, men vi ser at det er en liten ”lomme” med lavere verdier rundt Geilo (zoomet bilde av situasjonen 03 UTC vises i figur 6.5).



Figur 6.5: Zoomet bilde av modellkjøringen fra UM4 for situasjonen 03 UTC 8.juli 2009 som dekker meteogrampunktet Geilo og Oslo. UTC.

Figur 6.6 viser en annen type meteogram hvor opplysningene fra det deterministiske varslet er flettet sammen med sannsynlighetsvarslet. Også her er det valgt et fast naboområde på 76 km, og vi ser på sannsynligheten for å overstige 1 mm nedbør i et 3 timers intervall. Legg merke til tidspunktet 00-03 UTC som vi har kart for i figur 6.4. Her er det lett å se at de deterministiske verdiene er ganske like men sannsynligheten svært forskjellige. Hadde vi valgt et punkt litt lenger øst enn Geilo hadde forskjellene blitt enda mer tydelige. Andre gridpunkter i nærheten av Geilo har verdier helt ned i 0,7 mm, men sannsynligheten hadde fortsatt vært 99 %. Dette illustrerer hvor store utslag det kan gi å bare se på et punkt.

Ved å legge sannsynlighetsvarsling til det deterministiske på denne måten, enten ved bruk av diagrammer (figur 6.2 og 6.6) eller ved bruk av kart (figur 6.4), kan vi informere brukeren om varslets usikkerhet og gi et verdifullt bidrag til en beslutningstaking. Så må det være opp til brukeren hvor de setter grensen for usikkerheten ut i fra hva formålet med varslet skal være.



Figur 6.6: Varsel fra UM4 (blått) for punktene Geilo og Oslo 8.juli 2009 og sannsynlighetsvarsel for naboombåde 76 km (rosa). Viser sannsynligheten for å overstige 1 mm nedbør i 3 timers intervaller.

7. Oppsummering og konklusjon

Utvikling av modeller til bruk i operativ værvarsling går raskt mot mer finskalerte modeller. Disse har potensialet til å produsere mer ”riktige” varsler enn tilsvarende modeller med grovere oppløsning. Dette fordi finskalamodeller blant annet gir en mer nøyaktig representasjon av værphenomener slik som intense konvektive nedbørssystemer. Men tradisjonelle verifikasjonsmetoder viser ikke alltid noen forbedring i forhold til modeller med grovere oppløsning (figur 1.1). Det har derfor vokst frem et behov for nye verifikasjonsmetoder som kan ta hensyn til egenskapene ved disse finskalerte modellene. Vi har i denne oppgaven gått kort igjennom noen av disse nyere verifikasjonsmetodene og gitt begrunnelse for valg av naboområdemetoden som er brukt i denne oppgaven. Denne metoden er så testet mot norske radardata. I motsetning til en tradisjonell verifikasjonsmetode hvor man sammenligner et punkt i modellen med et punkt i observasjonsfeltet sammenligner man ved denne metoden et område rundt et punkt i modellen med et område rundt det samme punktet i observasjonsfeltet (figur 3.1). Det er benyttet ca 500 varsler fra juli og august 2009. Det er også gitt en kort beskrivelse av hvordan metoden kan benyttes som sannsynlighetsvarsel i daglig værvarsling.

Ved bruk av metoden viser verifikasjonsskårene stor forbedring i forhold til om verifikasjonen hadde vært gjort på gridskalaen (naboområde lik 1). I tabell 7.1 ser vi hvor mye FSS forbedrer seg fra gridskalaen ved å utvide naboområdet til henholdsvis 52, 76 og 100 km. Ved å utvide naboområdet til 76 km (tilsvarende et område på 19x19 gridruter) får vi en forbedring i skåret på 100 %. Det viktige er å finne hvor mye man må utvide naboområdet før man kan si at akseptabelt skår er nådd, uten å utvide så mye at forbedringene ved bruk av finskalerte modeller blir borte. Ut fra formelen for akseptabelt skår som er brukt i denne oppgaven (basert på våtandelen i radarfeltet) når modellen denne grensen for naboområder mellom 30-100 km for 80 %, 90 % og 95 % persentilen. For 99 % persentilen må vi opp i 230 km for å nå akseptabelt skår. Et naboområde på mellom 30-100 km er akseptabelt uten at vi glatter feltet for mye og vi kan si at metoden verifiserer tilfredsstillende på disse skalaene.

FSS ved forskjellige naboombråder					
Percentil	4 km Gridskalaen	52 km	76 km	100 km	Prosentvis forbedring 4 til 76 km
80%	0,52	0,69	0,73	0,76	40 %
90%	0,35	0,54	0,60	0,64	71 %
95%	0,23	0,41	0,46	0,51	100 %
99%	0,06	0,18	0,23	0,28	283 %

Tabell 7.1: Viser verdier for FSS for gridskalaen 4 km, og 3 forskjellige naboombråder. Forbedringen i % mellom 4 km og 76 km naboombåde er vist i siste kolonne.

Resultatene viser at skårene til metoden er avhengig av både domenet varslet blir evaluert over (stort kontra lite område), og den romlige fordelingen av selve nedbøren (våtandelen). Det er lettere å forutsi med rimelig nøyaktighet sannsynligheten for nedbør over et stort område enn et lite, og utbredt nedbør er lettere å forutsi enn lokal nedbør. Lav våtandelen gjør som nevnt skåret sensitivt og dette slår sterkest inn for høye terskelverdier og persentiler. En del falske ekko i radardataene, med svært høye verdier, slår også sterkt ut for de høye terskelverdiene og persentilene. Ved å se på enkeltvarsler ser man at de varslene som gir gode skår er de som er nærme hverandre og har liten eller ingen bias. Stor bias slår sterkt ut når man ser på absolutte terskelverdier.

Det viser seg at metoden verifiserer noe dårligere for norske forhold enn for eksempel over UK (Roberts og Lean, 2008). Dette kan skyldes at de skandinaviske radarene har større utfordringer, kanskje spesielt pga topografi og tetthet (Lopez, 2008). Modellen har også noe større utfordringer med skandinavisk og spesielt norsk topografi og klima.

En styrke ved metoden er at ved å vite på hvilken skala akseptabelt skår oppnås, kan vi si noe om på hvilken skala varslet kan presenteres på. Metoden er indirekte egnet for å finne plasseringsfeil i varsler og den viser kvantitative feil og bias (tabell 2.1). Den minsker problemet med "dobbel straff" ved at den samler rundt nedbørområdet. Den gir oss en mulighet til å måle riktigheten av varslet ut i fra riktige forutsetninger, oppløsningen i modellen, og er derfor spesielt brukbar for finskalerte modeller der vi forventer at detaljene

er mer uforutsigbare. Metoden kan brukes for både enkle og mer sammensatte tilfeller. Den er godt egnet for diskontinuerlige variable som nedbør og kan brukes for å sammenligne forskjellige modeller (kapittel 5.6) eller modeller forskjellig oppløsning, vurdere kvaliteten på radardataene, forskjell i modell og radar med årstidene, prediktabilitet (kapittel 5.4) m.m. Det er viktig at man verifiserer med de samme initialbetingelsene og over samme område hele tiden.

Svakheter ved metoden er at den kan være mindre intuitiv enn objektbaserte metoder. Stor bias kan gi merkelige verdier og den er som nevnt sensitiv mot få data for høyere terskelverdier. Den er med andre ord terskelavhengig, som mange andre metoder som bruker terskelverdier. Resultatene kan være avhengig av størrelsen på domenet (et større domene gir muligheten for større romlige feil). Når våtandelen i området er liten kan man få store avvik i FSS primært fordi biasen forandres. Valg av størrelse på verifikasjonsområde kan derfor ha betydning (Mittermaier og Roberts, 2009). For finskalamodeller må verifikasjonsområde være stort nok til å omfatte skalaen på et typisk mesoskala system. Hvis område er for stort vil våtandelen alltid bli liten, og hvis det blir for lite vil man kunne gå glipp av større romlige feil. Mittermaier og Roberts (2009) har antydnet en ideell domenestørrelse med lengde 800-1200 km. Og ikke minst som alle andre skår, forteller den ikke hele historien alene. Totalt sett er FSS et godt mål på romlig nøyaktighet av nedbørvarsler. Andre metoder trengs for å bestemme andre typer feil.

I denne oppgaven er grensen for akseptabelt skår satt ut fra romlig utbredelse av nedbøren i radarfeltet, formel (10) i kapittel 3. Men det kan det stilles spørsmål ved om denne formelen for akseptabelt skår er den eneste rette. Dette ble diskutert sent i prosessen (personlig kommunikasjon Nigel Roberts) og er derfor ikke testet ut: Alternative formler for f_{aksep} kan ta hensyn til utbredelsen av nedbør i modellfeltet (fraksjonen):

$$f_{aksep} = 0,5 + f_M / 2 \quad (13)$$

Eller man kunne bruke en blanding av fraksjonen for radarfelt og modellfelt:

$$f_{aksep} = 0,5 + ((f_R + f_M) / 2) / 2 \quad (14)$$

Hvis vi har null bias, $f_R = f_M$, vil formel (10), som er brukt i denne oppgaven, alltid bli riktig. For varsler med stor bias vil det ha liten betydning hvem formel man bruker, men med en mindre bias vil valg av formel for akseptabelt skår kunne ha betydning.

Forflytningsfeil av et nedbørområde kan også være tegn på tidsfeil, det vil si at modellen varsler litt før eller litt for sent. Tidsfeil kan også være en feilkilde i resultatene, men er ikke vurdert i denne oppgaven. Mange av de nyere metodene har begynt å implementere også tidsfeil.

Det kan se ut som radardataene ikke er helt optimale og at mer kan gjøres for å forbedre kvaliteten på disse. Allikevel vil verifisering med denne metoden mot radardata gi verdifullt bidrag til verifisering av nedbør ved at man har mye større omfang av "observasjoner" å verifisere mot. Radardataene er i denne oppgaven ikke testet mot andre typer observasjoner, noe som er et viktig arbeid for å optimalisere radardataene. Så gode radardata som mulig er viktig for verifikasjon og forbedring av modellene.

En annen interessant ting med metoden er at den kan brukes som en sannsynlighetstilnærming. Når man har funnet hvilken skala metoden oppnår akseptabelt skår, kan man utlede sannsynlighetsvarsler fra det deterministisk varslet på en enkel og billig måte. Slik kan metoden være et supplement til andre typer sannsynlighetsvarsler og være en enkelt måte å utlede sannsynlighetsvarsler til bruk for eksempel på yr.no. Her kunne det vært interessant å se hvordan varslet sannsynlighet verifiserte mot observert frekvens over et lengre tidsrom, men dette har ikke vært prioritert i denne oppgaven.

Det er vanskelig å sammenligne skårene på denne metoden med tradisjonelle skår for å gi svar på om modellene verifiserer bedre nå enn for 10 år siden, eller om de finskalerte modellene gir bedre skår enn de med grovere oppløsning. Det er en kjensgjerning at vi ved tradisjonell verifikasjon ikke kan påvise noen historisk forbedring i modellenes evne til å varsle nedbør over det siste tiåret, selv om verifikasjon av andre værparametre viser en markant forbedring. Vi vet imidlertid ikke om dette skyldes problemer med å beskrive nedbørens fysikk (parametrisering) eller om det er den romlige strukturen på nedbøren som skaper problemer siden høyere oppløsning er "straffende" for tradisjonelle verifikasjonsmetoder. Men dette kan testes ved å anvende metoden på lagrede historiske data og se om vi får bedre resultater når man tar hensyn til skalaen og den romlige fordelingen av værelementene. Lean (2008) brukte metoden for UM 1,4 og 12 km over UK og viste at høy oppløsning ga bedre resultat enn modeller med grovere oppløsning.

Appendiks:

Feilkilder radardata

Å måle mengde nedbør med radar har en del feilkilder (Salek et al., 2004). Hvor stor innvirkning disse feilkildene har varierer med værforholdene, avstanden til radaren, ”scan strategy”, oppløsning i tid og rom, topografi, data prosessering og mengde og kvalitet på verdiene radaren gir. Dette gjør det vanskelig å si hvor gode radarene er. Det vi måler i en radar er reflektiviteten (Z), det er den delen av den utsendte strålingen som kommer tilbake til mottakeren. Reflektivitet beregnes av en empirisk formel som inneholder den strålingsmengde som mottas (Watt), avstand, målets egenskaper og en del andre faktorer. Formelen for reflektivitet er gitt ved Probert-Jones radarligning:

$$P_r = \frac{P_t G^2 \theta^2 H \pi^3 K^2 L}{1024 (\ln 2) \lambda^2} * \frac{Z}{R^2}$$

P_r	Power return
G	Antenna gain
H	Puls length
K	Physical const.
Z	Target reflectivity
R	Target range
P_t	Peak trans. power
θ	Angular beamwidth
π	pi
L	Signal loss
λ	Energy wavelength

Man kan skille mellom systematiske og tilfeldige feil og korrigere disse på forskjellige måter. Tilfeldige feil kan minskes ved å redusere område og tiden man samler over. Systematiske feil kan minskes ved å sammenligne med andre type målinger og korrigere for slike feil. Kvaliteten på radaren måles ofte ved å sammenligne målinger gjort med radar med målinger gjort på bakken. Man sammenligner målinger fra et punkt i radaren med en måling på bakken. En radarverdi er gjerne over et område av størrelsesorden 1 km^2 mens en måling på bakken er veldig mye mindre. I tillegg er bakkemålinger som regel veldig spredt. Dette gjør at slike sammenligninger har store svakheter. Største avviket fra radarmålinger mot bakkebaserte observasjoner er fordi radaren måler nedbøren over et relativt stort volum i en høyde over bakken. Mengden av flytende vann kan variere med høyden og vil derfor ikke uten videre være representativ for hva som når bakken. Spesielt i situasjoner med konvektiv nedbør kan dette ha stor innvirkning.

Man kan også skille mellom instrument feil (ikke meteorologiske feil) og feil pga forandringer i den meteorologiske situasjonen (meteorologiske feil). Ikke meteorologiske feil kan være forårsaket av radarstrålens forplantning (underestimerer nedbøren på lange avstander, over 100 km), clutter, radar hardware og generell data prosessering. Også unormal bøyning av strålen pga vertikalfordeling av trykk, temp og fuktighet kan gi måling i høyere eller lavere nivå (ducting) enn det standardatmosfæren antyder. Men lagdelingen i atmosfæren under nedbørsituasjoner er som regel ikke langt fra standardatmosfæren og derfor blir ikke disse effektene så store. Dersom strålen blir hindret av bakken eller andre gjenstander og ingen ting eller lite av strålen passerer, fører dette til hel eller delvis radarskygge og dette gir underestimering av nedbør. Korreksjon kan gjøres dersom mesteparten av strålen passerer. Falske ekko oppstår når strålen treffer bakken, bygninger, trær eller andre gjenstander som reflekterer strålen. Falske ekko kan være en stor feilkilde i radarmålingene. Ekkoene kan være sterkere enn det sterkeste meteorologiske signal. Det er ikke lett å fjerne falske ekko, men man har flere metoder for å minske dette som falske ekko kart, statistisk filtrering, doppler filtrering, doble polariserings målinger osv. I områder der falske ekko blir observert regelmessig vil nedbøren bli systematisk overestimert. Andre feilkilder kan være antennens rotasjonshastighet, romlig oppløsning (den fysiske egenskapen som skal måles er nedbørintensiteten mens det som virkelig måles er reflektiviteten, som ikke er lineært relatert til intensiteten, dette fører til en bias), valg av bølgelengde (lengre bølgelengde utvider Raylightområde og omfatter da større partikler, på den annen side øker problemet med falske ekko med økende bølgelengde), insekter, fugler, fly, solstråling m.m. Den meteorologiske situasjonen (meteorologiske feil) påvirker også målingene. Om

nedbørsystemet utvikler seg fort eller beveger seg fort innvirker dette på radarstrålen. Det store omfanget av meteorologiske forhold som påvirker radarobservasjonen kan noen ganger forbedre nedbørestimatet og noen ganger forverre det. Z-R relasjonen (omregning av reflektivitet til nedbørintensitet) er kanskje den best kjente feilkilden. Man trenger et estimat på dråpestørrelsesfordelingen for å kunne regne om reflektiviteten. I dag er det vanlig å bruke en enkel, tids konservativ, sesonguavhengig Z-R relasjon gitt ved:

$$Z = aR^b \quad (12)$$

der a og b er empiriske konstanter, Z er reflektiviteten i dBZ = $10 \cdot \log_{10}(Z)$ og R er nedbør i mm/t. $Z = 200R^{1.6}$ blir brukt i Norge. Den passer bra for regn, men dårlig for snø og er derfor ikke gunstig over våre områder i vintermånedene. Det finnes metoder for å forbedre kvaliteten på Z-R relasjonen ved å tune den på forskjellige måter, men det blir foreløpig ikke gjort for de norske radarene. Den fysiske forandringen av nedbørpartikler, størrelse og fase, under dens livsutvikling påvirker den målte reflektiviteten og derved den utledede nedbørintensiteten. Det vi ønsker å måle er reflektivitetsverdiene i bakkenivå, men i virkeligheten måler vi en gjennomsnittlig reflektivitetsverdi i en viss høyde. Den observerte reflektiviteten viser seg å minke med økende høyde. Et kjent fenomen av denne vertikprofilen er det maksimum som man finner i smeltelaget, der fallende ispartikler går over til regn, en typisk tykkelse på noen hundre meter kalt "bright band". Den økte reflektiviteten i smeltelaget kan føre til overestimering av nedbøren (Joss og Waldvogel, 1990). Innflytelsen av dette smeltelaget er størst nær radaren. Dette gjør at snø og lavt frysenivå er en utfordring på våre breddegrader. Noen av de meteorologiske feilkildene som sterkest påvirker målingen av nedbøren er variasjonen av den fysiske tilstanden til nedbørpartikkelen, variasjon i antall partikler pga variasjon i hastighet til partiklene, partikler som smelter sammen og brytter opp og fordampning i tørr luft under skyen.

Kildeliste

- Ahijevych, D., Gilleland, E., Brown, B.G. og Ebert, E.E., 2009: Application of spatial verification methods to idealized and NWP gridded precipitation forecasts. *Wea. Forecasting*, (Available at Early Online Release).
- Atger, F., 2001: Verification of intense precipitation forecasts from single models and ensemble prediction systems. *Nonlin. Proc. Geophys.* 8, pp 401-417.
- Baldwin, M. E., Lakshminarayanan, S. og Kain, J. S., 2001: Verification of mesoscale features in NWP models. Preprints, 9th Conference on Mesoscale Processes, 30 July-2 August, Ft. Lauderdale, FL, American Meteorological society, pp 255-258.
- Bernardet, L. R., Grasso, L. D., Nachamkin, J. E., Finley, C.A. og Cotton, W. R., 2000: Simulating convective events using a high-resolution mesoscale model. *J. Geophys. Res.*105(D11), pp 14,963–14,982.
- Bremnes, J.B. og Homleid, M., 2009: Verification of Operational Numerical Weather Prediction Models June to August 2009. 23/2009
<http://met.no/?module=Articles;action=Article.publicShow;ID=1257>
- Briggs, W. M. og Levine, R. A., 1997: Wavelets and field forecast verification. *Mon. Wea. Rev.*, 125, pp 1329-1341.
- Brooks, H.E., Kay, M. og Hart, J.A., 1998: Objective limits on forecasting skill of rare events. 19th Conf. Severe Local Storms, Amer. Met. Soc., Minneapolis, MN, 14-18 September 1998, pp 552-555.
- Browning, K. A., Collier, C. G., Larke, P. R., Menmuir, P., Monk, G. A. og Owens, R. G., 1982: On the forecasting of frontal rain using a weather radar network. *Mon. Wea. Rev.*, 110, pp 534–552.
- Casati, B., Ross, G. og Stephenson, D. B., 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts, *Meteorol. Appl.*, 11, pp 141-154.
- Casati, B. et al. 2008: Review Forecast verification: current status and future directions. *Meteorol. Appl.*, 15, pp 3-18.
- Casati, B., 2009: New Developments of the Intensity-Scale Technique within the Spatial Verification Methods Inter-Comparison Project. *Wea. Forecasting*, (Available at Early Online Release).
- Davies, T., Cullen, M. J. P., Malcolm, A. J., Mawson, M. H., Staniforth, A., White, A. A. og Wood, N., 2005: A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Quart. J. Roy. Meteor. Soc.*, 131, pp 1759–1782.
- Davis, C., Brown, B. og Bullock, R., 2006I: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, 134, pp 1772-1784.

- Davis, C., Brown, B. og Bullock, R., 2006II: Object-Based Verification of Precipitation Forecasts. Part II: Application to Convective Rain Systems. *Mon. Wea. Rev.*, 134, pp 1785–1795.
- Done, J., Davis, C. og Weisman, M., 2004: The next generation of nwp: Explicit forecasts of convection using the weather research and forecasting (WRF) model. *Atmos. Sci. Lett.*, 5, pp 110-117.
- Du, J., Mullen, S.L. og Sanders, F., 2000: Removal of Distortion Error from an Ensemble Forecast. *Mon. Wea. Rev.*, 128, pp 3347-3351.
- Ebert, E.E., 2009: Neighborhood verification - a strategy for rewarding close forecasts. *Wea. Forecasting*, (Available at Early Online Release).
- Ebert, E.E. og Gallus, Jr. W.A., 2009: Toward better understanding of the contiguous rain area (CRA) method for spatial forecast verification. *Wea. Forecasting*, 24 (5), pp 1401-1415.
- Ebert, E. E. og McBride, J. L., 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J.Hydrol.*, 239, pp 179-202.
- Gilleland, E., Ahijevych, D., Brown, B. G., Casati, B. og Ebert E. E., 2009a: Intercomparison of spatial forecast verification methods. *Wea. Forecasting*, 24 (5), pp 1416-1430.
- Gjertsen, U. og Dahl, J.I., 2002: Challenges for precipitation estimation in mountainous regions. *Proceedings of ERAD 2002*, ERAD publication series, 1, pp 250-254.
- Hoffman, R. N., Liu, Z., Louis, J.-F. og Grassotti, C. 1995: Distortion representation of forecast errors. *Mon. Wea. Rev.*, 123, pp 2758–2770.
- Joss, J. og Waldvogel, A., 1990: Precipitation measurement and hydrology. *Radar in Meteorology*, D. Atlas (Ed.), AMS, pp 577-606.
- Lean, H.W., Clark, P.A., Dixon, M., Roberts, N.M., Fitch, A., Forbes, R. og Halliwell, C., 2008: Characteristics of High-Resolution Versions of the Met Office Unified Model for Forecasting Convection over the United Kingdom. *Mon. Wea. Rev.*, 136, pp 3408–3424.
- Lopez, P., 2008: Comparison of OPERA precipitation radar composites to CMORPH, SYNOP and ECMWF model data (technical memorandum 569).
- Mass, C. F., Ovens, D., Westrick, K. og Colle, B. A., 2002: Does increasing horizontal resolution produce more skillful forecasts? *Bull. Amer. Meteor. Soc.*, 83, pp 407-430.
- McBride, J.L. og Ebert, E.E., 2000: Verification of quantitative precipitation forecasts from operational numerical weather prediction models over Australia. *Weather Forecasting*, 15, pp 103-121.
- Mittermaier, M. P. og Roberts, N., 2009: Intercomparison of spatial forecast verification methods: Identifying skillful spatial scales using the Fractions Skill Score. *Wea. Forecasting*, (Available at Early Online Release)

- Murphy, A. H. og Epstein, E. S., 1989: Skill scores and correlation coefficients in model verification. *Mon. Wea. Rev.*, 117, pp 572–581.
- Nachamkin, J.E., Chen, S. and Schmidt, J., 2005: Evaluation of Heavy Precipitation Forecasts Using Composite-Based Methods: A Distributions-Oriented Approach. *Mon. Wea. Rev.*, 133, pp 2163–2177.
- Nehrkorn, T., Hoffman, R., Grassotti, C. og Louis, J.-F., 2003: Feature calibration and alignment to represent model forecast errors: Empirical regularization. *Q.J.R Meteorol. Soc.* 129, 195–218.
- Reinehart, R.E., 2004: *Meteorological Targets in Radar for meteorologists*, 4rd edition, Rinehart Publications.
- Roberts, N., 2008: Assessing the spatial and temporal variation in the skill of precipitation forecasts from an nwp model. *Meteorol. Appl.*, 15, pp 163-169.
- Roberts N.M. og Lean, H.W., 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, 136, pp 78-96.
- Salek, M., Cheze, J-L., Handwerker, J., Delobbe, L. og Uijlenhoet, R., 2004: Radar techniques for identifying precipitation type and estimating quantity of precipitation. Document of COST-717, WG1., http://www.smhi.se/cost717/doc/WDF_01_200407_1.pdf
- Theis, S. E., Hense, A., og Damrath, U., 2005: Probabilistic precipitation forecasts from a deterministic model: A pragmatic approach. *Meteor. Appl.*, 12, pp 257-268.
- Undén, P., Rountu, L., Järvinen, H., Lynch, P., J. Calvo, J. og co-authors, 2002: HIRLAM-5 Scientific Documentation, HIRLAM-5 project, c/o Per Undén SMHI, S-60176 Norrköping, Sweden.
- Wernli, H., Paulat, M., Hagen, M. og Frei, C., 2008: SAL—A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts. *Mon. Wea. Rev.*, 136, pp 4470–4487.
- Zepeda-Arce, J., Fofoula-Georgiou, E. og Droegemeier, K. K., 2000: Space-time rainfall organization and its role in validating quantitative precipitation forecasts. *J. Geophys. Res.*, 105(D8), 10,12910,146.