



Evaluering av flomvarslings usikkerhetsmodell

Rapport fra delprosjekt 4, FoU-prosjekt 81057 Forbedring av
flomvarslings verktøy

Elin Langsholt og Gusong Ruan

47
2015



R
A
P
P
O
R
T

Evaluering av flomvarslingas usikkerhetsmodell

Utgitt av: Norges vassdrags- og energidirektorat

Redaktør:

Forfattere: Elin Langsholt og Gusong Ruan

Trykk: NVEs hustrykkeri

Opplag: 10

Forsidefoto: Elinn Langsholt

ISBN 978-82-410-1094-1

Sammendrag: I denne rapporten evalueres flomvarslingstjenestens modellsystem for kvantifisering av usikkerheten i vannføringsprognosene. Når prognosen korrigeres for den estimerte feilen, får den mindre avvik fra observert vannføring enn den ukorrigerede de første 3-4 dagene. Den estimerte sannsynlighetsfordelingen gir en god beskrivelse av vær og vannføring. Modellfeilen bidrar med den største feilandelen i starten av perioden, mens værprognosen bidrar med det største feilbidraget i slutten av prognoseperioden.

Emneord: Flomvarsling, hydrologisk modellering, hydrologiske prognoser, kvantifisering av usikkerhet, evaluering

Norges vassdrags- og energidirektorat
Middelthunsgate 29
Postboks 5091 Majorstua
0301 OSLO

Telefon: 22 95 95 95
Telefaks: 22 95 90 00
Internett: www.nve.no

Mai 2015


Innhold

Forord	4
Sammendrag	5
1 Innledning	6
1.1 Historikk	6
1.2 Flomvarslings modellsystem og dagens usikkerhetsestimering.....	6
2 Flomvarslings usikkerhetsmodell	8
2.1 Usikkerheten i vannføringsprognosene.....	8
2.2 Modellering av feilen i den hydrologiske modellen.....	9
2.3 Modellering av feilen i værprognosen	9
2.3.1 Temperatur.....	9
2.3.2 Nedbør	9
2.4 Implementering	9
2.5 Kalibrering av usikkerhetsmodellen	11
3 Evaluering av usikkerhetsmodellen	12
3.1 Oversikt over evalueringen	12
3.2 Evaluering av feilmodellen for nedbørprognosen.....	13
3.2.1 Hvor godt treffer modellen?.....	13
3.2.2 Varsling av nedbørmengder	17
3.2.3 Nedbør - ikke nedbør	19
3.2.4 Sannsynlighetsfordelingen for nedbør.....	20
3.2.5 Treff, bom og falske alarmer	21
3.3 Evaluering av feilmodellen for temperaturprognosen.....	25
3.3.1 Hvor godt treffer modellen?.....	25
3.3.2 Sannsynlighetsfordelingen for temperatur.....	27
3.4 Vannføringsprognosen	29
3.4.1 Hvor godt treffer modellen?.....	29
3.4.2 Kvantilplott.....	32
3.4.2.1 Narsjø	32
3.4.2.2 Farstad	34
3.4.2.3 Øvrevatn	35
3.4.3 Treffrate og falske alarmer	37
3.4.4 Delfeilmodellenes bidrag.....	43
3.4.5 Sannsynlighet for overskridelse av flomnivåer	48
3.4.6 Sannsynlighetsfordeling for vannføring	52
4 Konklusjoner og anbefalinger	54
4.1 Oppsummering nedbør	54
4.2 Oppsummering temperatur	54
4.3 Oppsummering vannføring.....	54
4.4 Anbefalinger	56
5 Referanser	58

Forord

Varsling av naturfare, inkludert videreutvikling av modeller og rutiner for flomvarsling, er ett av NVEs strategiske fokusområder. Det er en overordnet målsetting at det hele tiden er et best mulig samsvar mellom flomvarslingas modellprognoser og observert vannføring. Men i erkjennelsen av at det alltid vil være et element av usikkerhet forbundet med varslete hendelser, vil det å kvantifisere denne usikkerheten være med og gi et bedre beslutningsgrunnlag for myndighetene når beredskapssituasjoner kan være på vei. NVEs flomvarslingstjeneste har anvendt probabilitiske metoder og estimert usikkerheten i vannføringsprognosene siden 2000. Utviklingen og anvendelsen av probabilitiske metoder i hydrologien har utviklet seg mye siden den gangen, i takt med den teknologiske utviklingen, og det er behov for en omfattende gjennomgang og videreutvikling av dagens metodikk og rutiner. Som et ledd i denne gjennomgangen er det, innenfor FoU-prosjektet 81057, Forbedring av flomvarslingas verktøy, gjort en evaluering av hvor godt dagens usikkerhetsmodell beskriver usikkerheten i vannføringsprognosene. Resultatene fra evalueringen er presentert i denne rapporten.

Oslo, mai 2015


for Morten Johnsrud
avdelingsdirektør


Hege Hisdal
seksjonssjef

Sammendrag

I denne rapporten presenteres og evalueres flomvarslingstjenestens modellsystem for kvantifisering av usikkerheten i vannføringsprognosene. Systemet estimerer forventet feil i vær- og vannføringsprognosene, og korrigerer for denne. Denne usikkerhetsmodellen ble implementert på flomvarslinga i 2000, og har gått operasjonelt siden den gangen, tilpasset HBV-modellen og ulike versjoner av værprognoser. Modellsystemet er i denne runden evaluert på data for perioden 2011 – 2012. Evalueringen tar sikte på å kvantifisere hvor godt de korrigerte prognosene sammenfaller med observasjonene, hvor godt usikkerhetsmodellens estimerte sannsynlighetsfordeling beskriver vær og vannføring i prognoseperioden, og størrelsesforholdet mellom feilbidraget til vannføringsprognosene fra værprognosene og fra HBV-modellen.

Resultatene av evalueringen viser at delfeilmodellen for nedbør feiler, ved at den korrigerte nedbørprognosen faktisk gir dårligere resultater enn den ukorrigerte. Feilmodellen for temperatur, derimot, gir en forbedring i temperaturprognosen, som i situasjoner der temperaturen spiller en rolle for tilsiget må antas å ha en positiv effekt på vannføringsprognosen. Modellen som korrigerer HBV-modellens vannføringsprognose med hensyn til siste observerte vannføringsverdi gir også en betydelig forbedret vannføringsprognose de første prognosedagene. Den positive effekten av modellfeilkorrigeringen dominerer over det negative utslaget av nedbørprosesseringen, og gir en prognose med totalt sett mindre avvik de første 3 - 4 dagene.

For alle delfeilmodeller gjelder det at den estimerte sannsynlighetsfordelingene gir en relativt god beskrivelse av vær og vannføring. Analysen av delfeilmodellenes bidrag viser at modellfeilen har et konstant bidrag til totalfeilen gjennom hele prognoseperioden, og bidrar med den største andelen i starten av perioden. Værprognosefeilens andel av totalfeilen øker utover i prognoseperioden, i takt med at usikkerheten i værprognosen øker, og bidrar med det største feilbidraget i slutten av prognoseperioden. Dette er et gjennomgående trekk for alle modellfelt.

Resultatene fra evalueringen er nyttige i den daglige tolkingen av resultatene fra usikkerhetsmodellen, og i arbeidet med å kartlegge forbedringspotensialer og legge en strategi for videreutvikling av probabilistisk varsling i flomvarslingstjenesten.

1 Innledning

1.1 Historikk

Siden 1989 har NVEs flomvarslingstjeneste publisert daglige, landsdekkende vannføringsprognoser. Slike prognoser har som regel relativt begrenset interesse og anvendelse, men i flomsituasjoner når de ut i offentligheten, og kan bli gjenstand for omfattende kritisk vurdering. Vesleofsen i 1995, som rammet store deler av Østlandet, var en slik hendelse. Prognosene som ble publisert mens flommen sto på varierte, naturlig nok, fra dag til dag, og i tillegg ble prognoser fra NVE og fra vassdragsregulanten publisert ukoordinert. Dette skapte forvirring og frustrasjon blant publikum, og offentlig debatt i media. Det ble åpenbart at mye av denne forvirringen kunne vært unngått hvis man hadde klart å kommunisere noe av usikkerheten som alltid vil være til stede i slike prognoser. Det forutsetter imidlertid at man er i stand til å gi et kvantitativt mål på denne usikkerheten.

Flommen satte fokus på tiltak som kan redusere flomskader og gjøre oss bedre forberedt foran en flom, og forskningsprogrammet HYDRA startet opp januar 1996. I HYDRA var det et uttalt mål å bli bedre til å forebygge skadeflommer og redusere omfanget av flomskader, utfra et forbedret kunnskapsgrunnlag. Flomvarsling var ett av temaene i HYDRA, og flere delprosjekter ble initiert for å forbedre grunnlaget for hydrologiske prognoser. Det ble gjort flere analyser for å identifisere forhold som kan redusere eller si noe om usikkerheten i prognosene. En av leveransene herfra var en modell for kvantifisering av usikkerhet i vannføringsprognosene, som fremdeles er grunnlaget for sannsynlighetsestimering og beregning av utfallsrom i NVEs vannføringsprognoser. Utviklingen av usikkerhetsmodellen ble initiert og administrert av NVE. Norsk Regnesentral utviklet de stokastiske komponentene som brukes i beregningen av usikkerheten.

Metoden er den samme i dag som den gangen den ble utviklet, men er blitt tilpasset ulike versjoner av HBV-modellen og ulike generasjoner av meteorologiske inndata. Den ble implementert i flomvarslingas modellsystem første gang rundt 2000, da i en halvautomatisert versjon for et utvalg av modellfeltene. Metoden ble første gang presentert internasjonalt i 2000 (Langsholt, 2000) Usikkerhetsmodellen har blitt evaluert en gang tidligere (Langsholt og Væringstad, 2004). Resultatene fra den gang stemmer godt overens med resultatene fra denne evalueringen, men denne rundens evaluering går mer i detalj og studerer flere aspekter ved usikkerhetsmodellen.

1.2 Flomvarslingas modellsystem og dagens usikkerhetsestimering

De hydrologiske nedbør-avløpsmodellene utgjør et grunnleggende beslutningsstøtteverktøy i NVEs flomvarslingstjeneste. Da usikkerhetsmodellen ble utviklet besto modellparken av HBV-modeller som var kalibrert for ca. 80 nedbørfelt. Meteorologiske inndata ble beregnet ved hjelp av observasjoner ca. 150 stasjoner og fra korttids- (HIRLAM) og langtidsprognoser (ECMWF) for totalt 6 prognosedager. Inndata til modellene ble beregnet hver morgen og tidsskrittet for modellene var ett døgn.

De tradisjonelle HBV-modellene er fortsatt i daglig bruk, nå for 145 nedbørfelt. I tillegg er det utviklet en modulær variant av HBV-modellen for en ny plattform, Open Distributed Model (ODM), som kjører for de samme 145 feltene. Denne kjører på døgnbasis og er klargjort for 3-timersbasis. Nok et supplement er en alternativ nedbør-avløpsmodell, Distance Distribution Dynamics (DDD), som er kalibrert og kjører på både døgn- og 3-timersbasis for ca. 80 nedbørfelt. Alle modellene kjøres på nytt hver gang det kommer inn nye observasjoner eller resultater fra en ny værprognosekjøring, dvs. fire ganger om dagen. Observerte inndata beregnes ikke lenger direkte fra stasjonsobservasjoner, men hentes fra interpolerte grid, som er basert på det meteorologiske stasjonsnettet, som i dag omfatter ca. 400 nedbørstasjoner og 230 temperaturstasjoner (se Norge-griddet). Interpolasjonsmetodene er beskrevet i Mohr (2008). Værprognosene hentes fra samme type grid, interpolert fra prognoseresultater med en grovere romlig skala.

Det er mange kilder til usikkerhet i de modellerte vannføringsprognosene. Usikkerhetsmodellen er basert på at det er 2 hovedkilder til usikkerhet: det at værprognosen avviker mer eller mindre fra været som kommer, og den tilnærmingen til den avløpsgenererende prosessen som modellens matematiske formelverk representerer. Bidragene fra disse to feilkildene vil avhenge av hvilken værprognosemodell som brukes og hvor god den hydrologiske modellen/parametertilpasningen er. Usikkerhetsmodellen må altså tilpasses de til enhver tid gjeldende værprognoser, observasjonsgridversjoner og modellkalibreringer. For at usikkerhetsmodellen skal fungere optimalt, kreves det derfor at den recalibreres hver gang det skjer en endring i disse rammebetingelsene, og at det eksisterer tidsserier å kalibrere på. Å skaffe til veie tidsserier av tilstrekkelig lengde til kalibrering er et problem når det gjelder meteorologiske prognoser. Dette, i kombinasjon med at det er tidkrevende å kalibrere usikkerhetsmodellen, har medvirket til at den i lange perioder ikke har vært optimalt tilpasset det øvrige modellsystemet.

Usikkerhetsmodellen er nå bare implementert på de tradisjonelle HBV-modellene. Kvantifisering av usikkerheten i våre hydrologiske prognoser er imidlertid en funksjonalitet som må være operasjonell på alle modellplattformer. Men metoden er tungdrevet og det finnes alternativ metodikk som bør gjennomgås med tanke på å erstatte eller supplere eksisterende system. Kvantifisering av usikkerhet i meteorologiske og hydrologiske prognoser har vært et sentralt tema for forskning og utvikling i mange år. Mye velprøvd metodikk er tilgjengelig og klar for utprøving. Men før vi introduserer nye metoder, må vi, som referanse, vite hvor godt dagens system fungerer. Det er lenge siden forrige evaluering av metoden, og mye har skjedd siden den gang. Derfor ble evaluering av usikkerhetsmodellen satt opp som et delprosjekt i det interne FoU-prosjektet 81057 Forbedring av flomvarslingas verktøy.

Evalueringen tar for seg usikkerhetsestimeringen mht. værprognosene og den hydrologiske modellen hver for seg og samlet. Den omfatter treff/sammenfall av prognoser, simuleringer og observasjoner generelt og flommer spesielt, konfidensintervallenes kvantifisering av spredningen i utfallene og treffsikkerheten til risikoestimatene. Kriteriene som evalueringen bygger på er i stor grad de samme som er foreslått (delprosjekt 1, 81057) for å evaluere flomvarslingsmodellene (Haddeland, 2014).

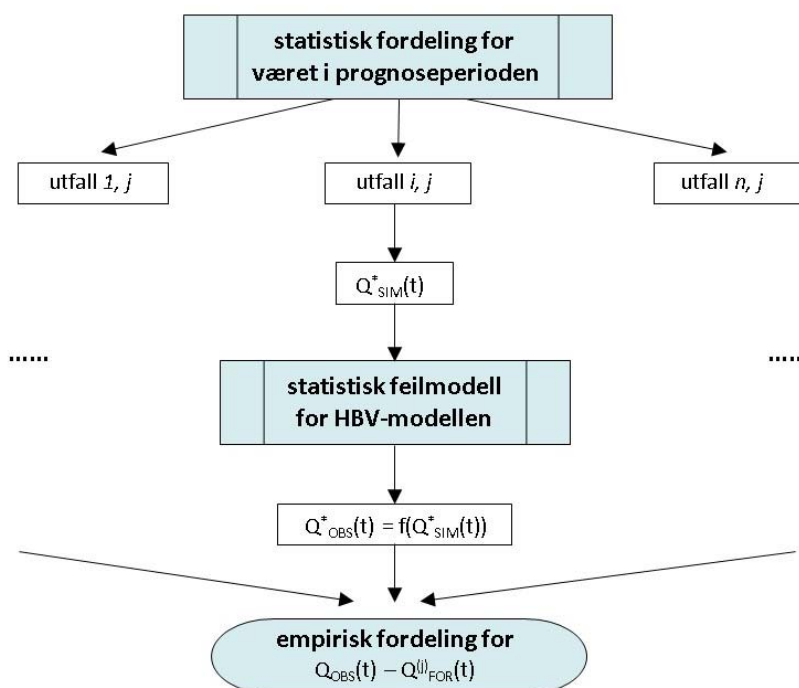
2 Flomvarslings usikkerhetsmodell

2.1 Usikkerheten i vannføringsprognosene

Dagens metode for beregning av usikkerhet i vannføringsprognosene bygger på en hypotese om at det er to hovedkilder til feil i vannføringsprognosene, den hydrologiske modellen, som er en grov tilnærming til virkeligheten, og usikkerheten i værvarslene. Den totale feilen i vannføringsprognosene kan etter denne modellen uttrykkes slik:

$$Q_{OBS}(t) - Q_{SIM}(t) + Q_{SIM}(t) - Q_{FOR}^{(j)}(t) = Q_{OBS}(t) - Q_{FOR}^{(j)}(t).$$

$Q_{OBS}(t) - Q_{SIM}(t)$, differansen mellom observert og modellsimulert, dvs. med observert nedbør og temperatur som inndata, vannføring ved tiden t , beskriver modellfeilen. $Q_{SIM}(t) - Q_{FOR}^{(j)}(t)$, differansen mellom simulert vannføring med observerte inndata og simulert vannføringsprognose laget j dager før, med nedbør- og temperaturprognose som inndata, beskriver feilen som skyldes værprognosen. Den totale feilen i vannføringsprognosen kan uttrykkes som $Q_{OBS}(t) - Q_{FOR}^{(j)}(t)$, altså differansen mellom vannføringen som til syvende og sist blir observert ved tiden t og prognosen for samme tidspunkt, lager j dager før.



Figur 2.1. En illustrasjon av dataflyten i usikkerhetsmodellen

Usikkerhetsmodellen kombinerer en feilmodell for HBV-modellen med modeller for feilen i værvarslene sammenliknet med været slik det blir. Det beregnes en sannsynlighetsfordeling for nedbør og temperatur hver dag t i prognoseperioden, gitt dagens værvarsel. Disse fordelingene er avhengige av vær og snøforhold. Det trekkes et antall utfall fra disse to fordelingene, uavhengig av hverandre, som gis som inndata til HBV-modellen og videreprosesserer gjennom feilmodellen for HBV-modellen. Et typisk

antall utfall brukt i flomvarslinga er 100. Dette gir opphav til et tilsvarende antall mulige vannføringer for hver prognosedag t . Ut fra disse utfallene kan forventningsverdi, konfidensintervaller og sannsynlighet for overskridelse av flomnivåer estimeres. Figur 2.1 gir et bilde av dataflyten i systemet. Feilmodellene er utviklet basert på data fra nedbørfeltene Knappom på Finnskogen (vannmerke 2.142 Knappom) og Røykenes sørøst for Bergen (vannmerke 55.4 Røykenes).

2.2 Modellering av feilen i den hydrologiske modellen

Den isolerte usikkerheten med hensyn til HBV-modellene ble studert ved å sammenlikne observasjoner med simuleringer beregnet ved å bruke værobservasjoner, som kan ses på som perfekte prognoser, som inndata. Variabiliteten i avviket mellom simulert og observert vannføring er typisk større ved høye vannføringer, så det var hensiktsmessig å utvikle feilmodellen ved bruk av logtransformerte data. Det ble funnet en første ordens autoregressiv struktur på avviket, det vil si at avviket den ene dagen er en funksjon av avviket dagen før, og følgende modell for feilen ved dag t ble foreslått

$$d_t = \alpha_t d_{t-1} + \sigma_t \varepsilon_t$$

der d_t er avviket mellom den logtransformerte simulerte og observerte vannføringen ved dag t , den autoregressive parameteren α_t og standardavviket σ_t er betinget av den logtransformerte vannføringen og temperaturen, nedbøren og snødekket på dag t , og ε_t er et standard normalfordelt feilledd. Modellen skiller mellom 5 ulike meteorologiske regimer, og totalt 12 parametere må tilpasses modellen i hvert nedbørfelt.

2.3 Modellering av feilen i værprognosen

2.3.1 Temperatur

Temperaturmodellen følger en normalfordelingen og har også en autoregressiv struktur. Den har ulike parametersett ved prognoser under og over null, med større variabilitet ved negative temperaturprognoser. 8 parametere for hvert av de 9 tidsskrittene må estimeres ved kalibrering mot historiske prognoser og observasjoner.

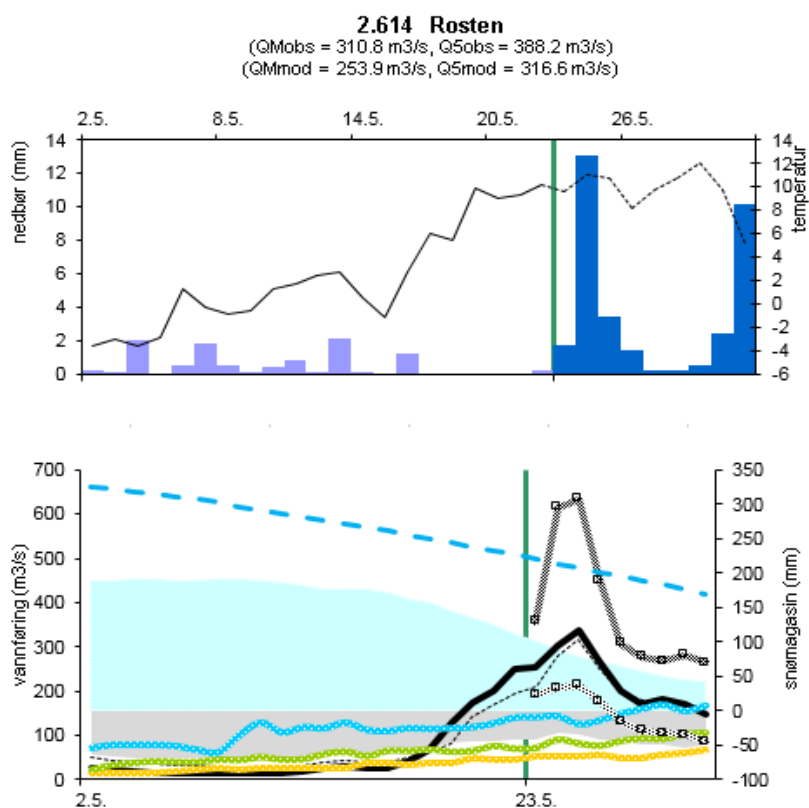
2.3.2 Nedbør

Nedbørmodellen er todelt. Først modelleres sannsynligheten for nedbør, og deretter mengden nedbør, gitt at sannsynligheten for nedbør er større enn null. Nedbørmengden antas å følge en gammafordeling med en forventningsverdi som bare er avhengig av prognosen for samme tidspunkt. Totalt 7 parametere for hvert tidsskritt må estimeres ved hjelp av historiske data.

2.4 Implementering

Usikkerhetsmodellen er i dag implementert på flomvarslingas tradisjonelle HBV-modellsystem, som går på linux-plattformen og omfatter 145 nedbørfelt. Historiske inndata til disse modellene hentes fra tidsserien av seNorge-gridd for nedbør og temperatur. Modellene kjøres operativt på nedbørprognoser fra AROME MetCoOp første to døgn og EC de neste sju. Temperaturprognosene beregnes av Hirlam8 første to døgn

og EC de neste sju. AROME står for Application of Operations at MESoscale, og er en numerisk værvarslingsmodell med 2,5 km oppløsning. Versjonen er utviklet innen MetCoOp-samarbeidet mellom MET og SMHI (Sveriges meteorologiska och hydrologiska institut). EC står for European Centre for Medium-Range Weather Forecasts' deterministiske værmodell, som har en horisontal oppløsning på 16 km. Hirlam8 betegner High Resolution Limited Area Model med 8 km oppløsning. Usikkerhetsmodellen ble kalibrert siste gang i 2011. Da ble nedbørprognoser fra UM4 (Unified Model, 4 km oppløsning) og temperaturprognoser fra Hirlam8 benyttet de første to dagene og EC-resultater de neste sju. De nedbørprognosene som brukes nå er altså fra en annen værvarslingsmodell, øvrige prognoser er fra nyere versjoner av samme værvarslingsmodell.



Figur 2.2. Modellresultatet fra HBV-modellen med usikkerhetsestimering for stasjonen 2.614 Rosten, 22. mai 2014. Den øverste figuren viser observert nedbør og temperatur i historisk periode og prognose 9 dager framover. Den nederste figuren viser observert (heltrukken svart linje) og simulert (stipla linje) vannføring i historisk periode. Modellsimuleringen fortsetter i prognoseperioden med prognoserte inndata, mens fortsettelsen av observasjonene viser medianen i utfallene fra usikkerhetsmodellen. De grå linjene som omslutter vannføringsprognosen i prognoseperioden viser 25- og 75%-persentilen i utfallsensebelet, og angir dermed et 50% konfidensintervall for vannføringen i prognoseperioden.

HBV-modellene med usikkerhetsmodellens postprosessering kjøres 4 ganger i døgnet, dvs. når nye versjoner av værprognosen og/eller siste observasjonsgridd er på plass. Modellresultatene med usikkerhetsestimater presenteres i regneark, se figur 2.2. Det

beregnes dessuten en tabell med sannsynlighet for overskridelse av middel- 5- og 50-årsflom for hver dag i prognoseperioden.

2.5 Kalibrering av usikkerhetsmodellen

Som det framgår av beskrivelsen av feilmodellene, er det en rekke parametere som må tilpasses for at de skal fungere optimalt på et spesifikt modellsystem. Modellsystemet defineres av hvilken hydrologisk modell som kjøres, parameterverdiene i modellen, hva slags inndata som brukes og hvordan de prosesseres. Hver gang det skjer endringer i noen av disse systemkomponentene bør usikkerhetsmodellen recalibreres, med data som representerer det nye regimet. Endringer i flomvarslingssystemet skjer typisk når HBV-modellene recalibreres, når det skjer endringer i historiske data, som når det kommer en ny versjon av seNorge-griddene, når noen av vannføringsstasjonene har fått ny vannføringskurve, eller når nye eller nye versjoner av værvarslingsmodellene blir tatt i bruk for å generere værprognoser. Det er tilrettelagt for kalibrering med halvautomatiske rutiner, men dette er likevel en tidkrevende prosess. Et større problem er at homogene tidsseriene med værprognoser, som er nødvendige for å få kalibrert modellen for feilen i værprognosen, er svært korte. Dette problemet er størst når en ny værprognoseversjon, som gjerne har gått operativt en kort periode, tas i bruk. Det at kalibreringen er ressurskrevende og at det ikke eksisterer tidsserier av tilfredsstillende lengde å kalibrere på, medfører at usikkerhetsmodellen ikke nødvendigvis er optimalt tilpasset modellsystemet den skal beregne usikkerheten til. Dette kan gi seg utslag i usikkerhetsestimater med lav presisjon eller som i verste fall bommer.

3 Evaluering av usikkerhetsmodellen

3.1 Oversikt over evalueringen

Datagrunnlaget for denne evalueringen er vannføringsprognoser med usikkerhetsberegninger for perioden 2011 – 2012, som sammenliknes med den tilsvarende observerte vannføringen. I den perioden ble det beregnet vannføringsprognoser for 127 flomvarslingsfelt, og 109 av disse er med i evalueringen. Vannføringsprognosene ble beregnet med nedbørprognoser fra UM4 og EC og temperaturprognoser fra Hirlam8 og EC som inndata. Evalueringen omfatter alle delmodeller, dvs. feilmodellen for nedbør- og temperaturprognoser og feilmodellen for HBV-modellen. Datagrunnlaget for evalueringen omfatter dermed i tillegg nedbør- og temperaturprognoser og utfall fra de statistiske fordelingene for nedbør og temperatur i prognoseperioden for 2011-2012. Analysen er gjort med 100 utfall fra fordelingene. Prognosene sammenliknes med interpolerte observasjoner av nedbør og temperatur fra seNorge-griddet. Dette er et datasett som, på tross av at det er basert på observasjoner, i stor grad er modellert, og forventes å ha et avvik i forhold til nedbøren og temperaturen som reelt representerer de samme arealene, men som er en ukjent størrelse. Verifikasjonsdatasettet gir dermed ikke den fulle sannheten om været slik det ble. Men sett i sammenheng med vannføringsmodellene, som er tilpasset seNorge-griddet ved kalibreringen, er de et tilstrekkelig relevant sammenlikningsgrunnlag.

Evalueringen har omfattet følgende analyser:

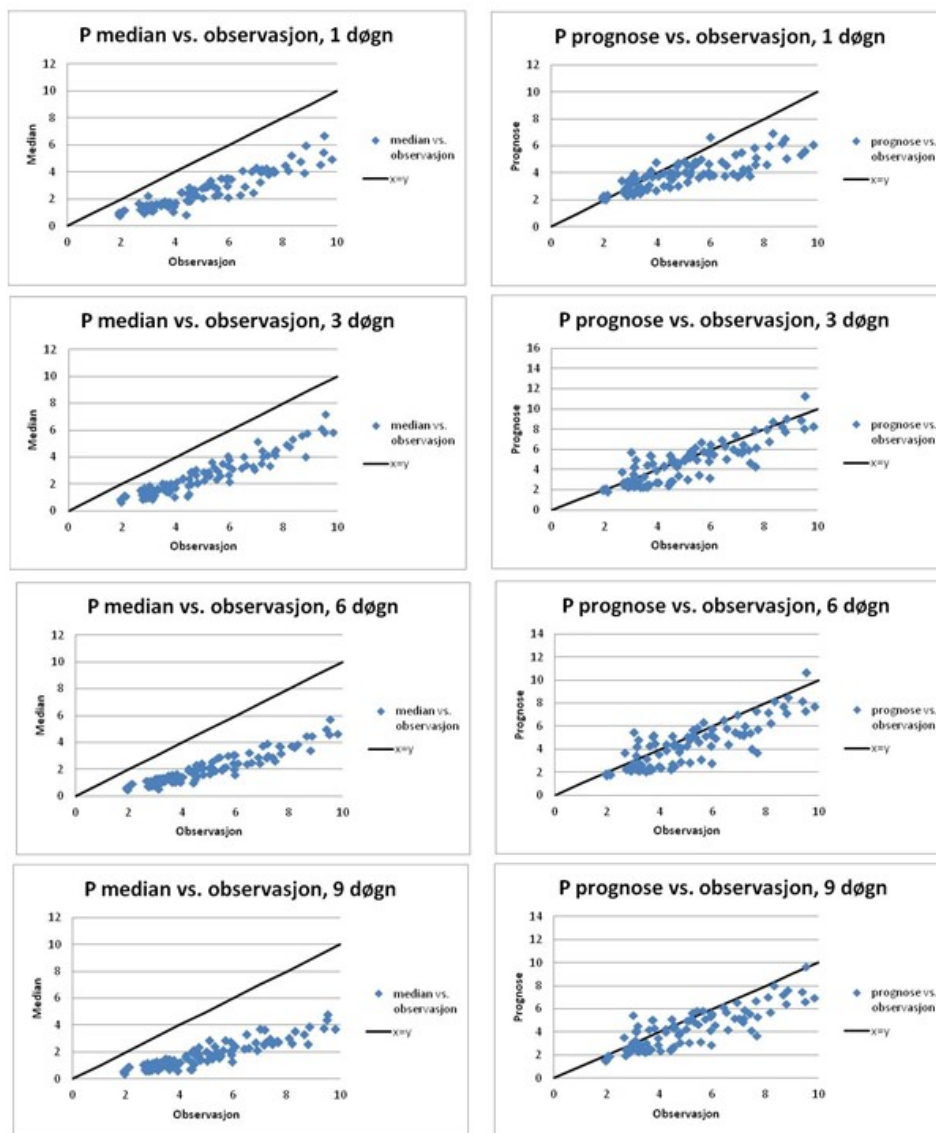
- 1) sammenlikning av hvor godt medianen i sannsynlighetsfordelingene for nedbør, temperatur og vannføring fra usikkerhetsmodellen verifiserer mot observasjonene, i forhold til de uprosesserte prognosene,
- 2) evaluering av usikkerhetsmodellens sannsynlighetsfordelinger for vær og vannføring og
- 3) kvantifisering av størrelsesforholdet mellom feilbidraget til vannføringsprognosene fra værprognosene og fra HBV-modellen.

Medianen og prognosen er sammenliknet ved å se på sammenfallet med observasjonene ved hjelp av ulike feilfunksjoner. Evalueringen omfatter også dikotome hendelser, dvs. hendelser med to utfall, i form av treff/bom-analyser og hvordan varselet treffer mht. fordeling (resolution), dvs. i forhold til hendelsene inndelt i kategorier mht. intensitet/størrelse. I tillegg er det undersøkt hvordan sannsynlighetsfordelingene for nedbør, temperatur og vannføring treffer mht. pålitelighet (reliability), dvs. sammenfall mellom varslet sannsynlighetsfordeling og observert frekvens, treffsikkerhet (sharpness), dvs. hvor godt ekstremer varsles, og fordeling. Mye av metodikken er beskrevet i WWRP/WGNE Joint Working Group on Forecast Verification Research (2013).

Resultatene fra evalueringen er nyttige i den daglige tolkingen av resultatene fra usikkerhetsmodellen, og ikke minst med tanke på å kartlegge forbedringspotensialer og legge en strategi for videreutvikling av probabilistisk varsling i flomvarslingstjenesten.

3.2 Evaluering av feilmodellen for nedbørprognosen

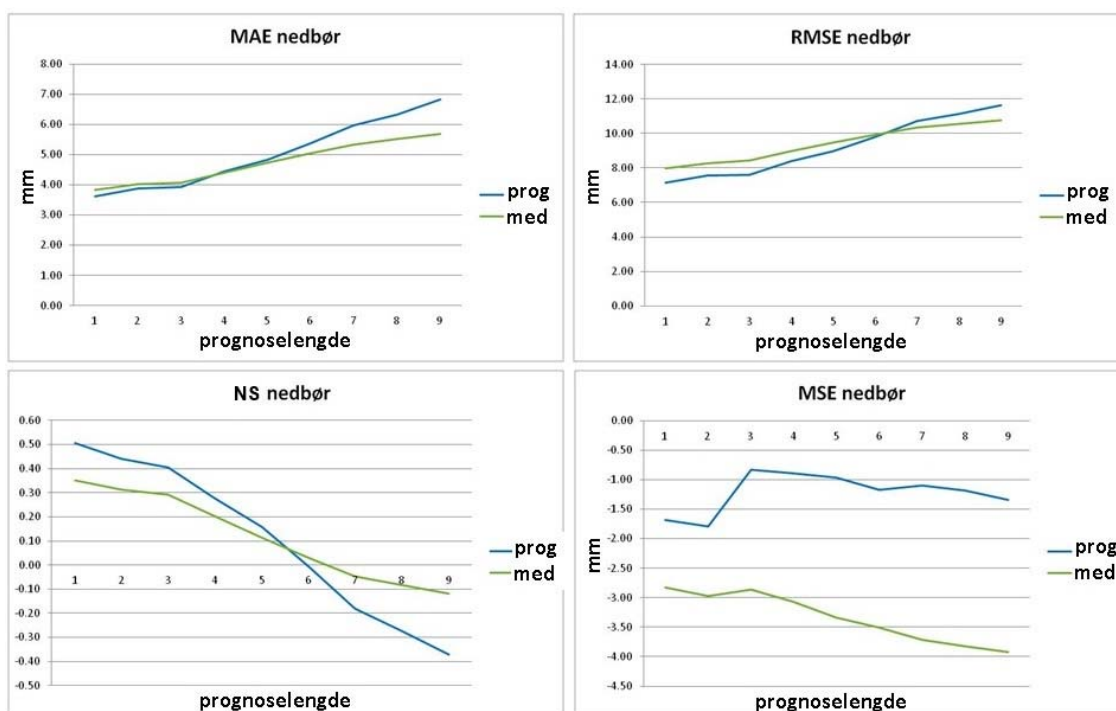
3.2.1 Hvor godt treffer modellen?



Figur 3.1. Punktdiagrammer som viser medianen i utfallene fra usikkerhetsmodellen for nedbør plottet mot tilsvarende observasjon (venstre kolonne) og den uprosesserte prognosen plottet mot observasjon (høyre kolonne). Punktene representerer middelverdien over evalueringsperioden for hvert nedbørfelt, og viser resultatene for prognoselengde 1, 3, 6 og 9 døgn.

Punktdiagrammer gir et visuelt inntrykk av hvor godt prognosene sammenfaller med observasjonene. Figur 3.1 viser slike punktdiagrammer for henholdsvis medianen i utfallene fra usikkerhetsmodellen og den uprosesserte prognosen. Figuren synliggjør et

trekk ved nedbørmedianen som har stor betydning for kvaliteten på usikkerhetsmodellens korrigerte prognoser. Medianen ligger betydelig lavere enn observasjonen, og forskjellen øker med gjennomsnittsnedbøren i feltet. Forskjellen øker også med prognoselengden etter 3 dager. Den uprosesserte prognosen har bedre overensstemmelse med observasjonene, men også denne ligger lavt for mer nedbørrike felt. Både medianen og prognosen ligger lavt i forhold til observasjonene for nedbørrike felt ved en dags varslingslengde. Det er UM4-modellen som brukes for 1-dags varsler, mens EC brukes ved de andre varslingslengdene som er vist i figuren. Figur 3.1 viser at EC gir nedbørprognoser som i mindre grad enn UM4 underestimerer nedbøren i nedbørrike felt. Usikkerhetsmodellens "korrigering" av nedbørprognosen har, i henhold til dette visuelle inntrykket, en tendens til å feile, og gjøre den dårligere enn den ukorrigerte.



Figur 3.2. Fire feilestimater som viser hvor godt den uprosesserte prognosen (blå strek) og medianen (grønn strek) i utfallene fra usikkerhetsmodellen treffer observasjonene. Prognoselengden er vist langs x-aksen.

Kvantitative estimater på sammenfallet mellom nedbørprognoser og observasjoner er vist på figur 3.2. Her er feilfunksjonene anvendt på tidsseriene over evalueringsperioden for hver stasjon, og gjennomsnittet over alle stasjoner er vist på figuren. De fire feilfunksjonene som er presentert er:

$$\text{Gjennomsnittlig absoluttavvik (MAE)} = \frac{1}{N} \sum_{i=1}^N |P_i - O_i|,$$

der P_i er prognosen og O_i er observasjonen ved tidsskritt i og N er antall tidsskritt. MAE beregner den gjennomsnittlige størrelsen på feilen uten å ta hensyn til fortegnet.

$$\text{Additivt avvik (MSE (mean signed error), balanse eller bias)} = \frac{1}{N} \sum_{i=1}^N (P_i - O_i),$$

som viser i hvilken retning nettofeilen går.

$$\text{Middelavvik (RMSE)} = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2},$$

som estimerer et gjennomsnittlig avvik, der store avvik har større vekt enn små avvik.

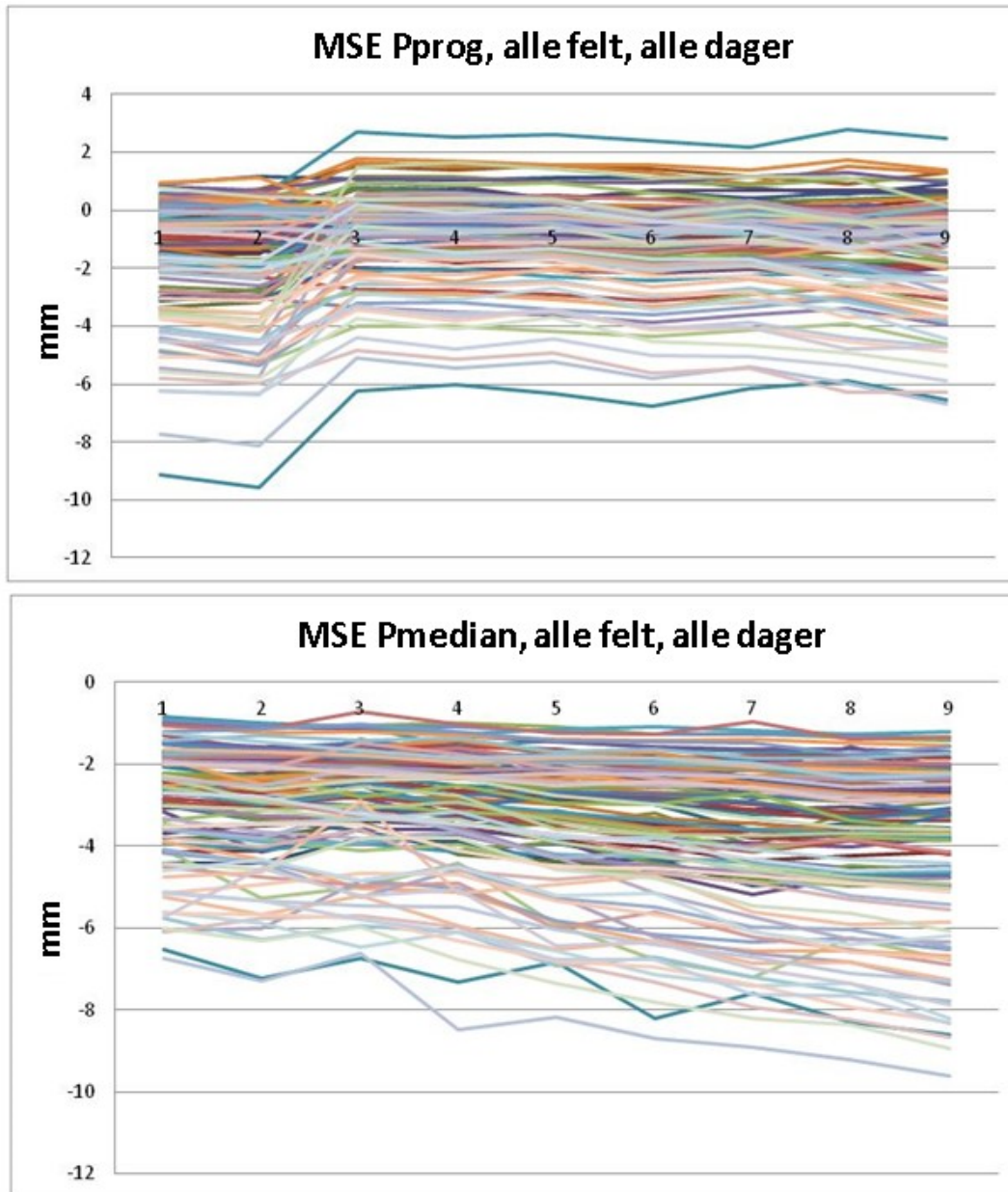
$$\text{Nash-Sutcliffe-koeffisienten (NS eller determinanskoeffisienten } R^2) = 1 - \frac{\sum_{i=1}^N (P_i - O_i)^2}{\sum_{i=1}^N (\bar{O} - O_i)^2},$$

som i likhet med middelavviket vektlegger store avvik med enn små. NS kan variere fra $-\infty$ til 1, der 1 tilsvarer perfekt tilpasning mellom prognose og observasjon. Hvis NS er lik 0, har prognosen en treffsikkerhet som tilsvarer gjennomsnittet av observasjonene.

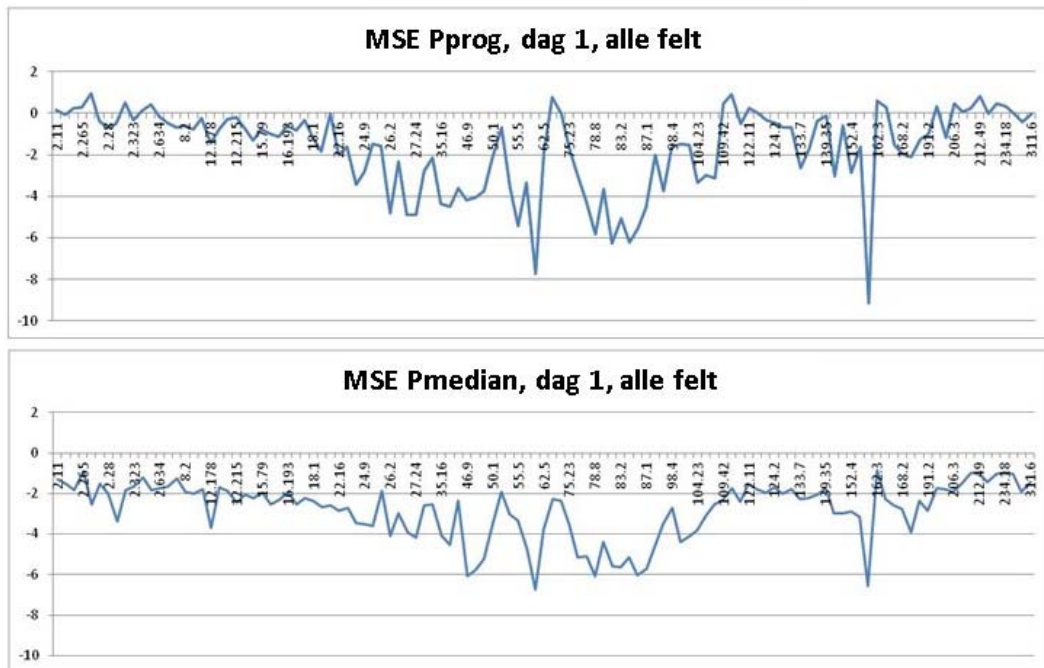
Det er et gjennomgående trekk at medianen i utfallene fra usikkerhetsmodellen skårer dårligere enn den uprosesserte prognosen de første prognosedagene. Middelavviket skårer relativt dårligere enn absoluttavviket, noe som viser at medianen har store avvik som bidrar til et høyt middelavvik. NS-koeffisienten viser et tilsvarende lavt skår for medianen, der store avvik trekker ned. Det additive avviket viser at både medianen og prognosen tenderer mot å ligge lavere enn den realiserte nedbøren, og medianen enda levere enn prognosen, en tendens som også er tydelig i figur 3.1. Tendensen øker utover i prognoseperioden. Fra prognosedag 5-6 gjør medianen det bedre enn prognosen målt med tre av feilfunksjonene. Prognosen har altså et større avvik enn medianen i absolutt forstand, selv om medianen underestimerer mer. Så langt ut i prognoseperioden er imidlertid prognosene relativt dårlige. NS-koeffisienten viser at gjennomsnittsnedbøren gir en like treffsikker prognose fra dag 6-7.

De to første prognosedagene brukes prognoser fra UM4, mens resten av perioden bruker EC-prognoser. Det additive avviket for den uprosesserte prognosen viser at UM4 har en sterkere tendens til å underestimere nedbøren enn EC. Figur 3.1, øverste høyre diagram, viser at det gjelder i størst grad for dager med mye nedbør. Usikkerhetsmodellens korrigering utligner noe av forskjellen mellom de to modellene.

Figur 3.3 viser spredningen i det additive avviket over alle prognosefeltene. De enkelte prognosefeltene følger i hovedsak samme variasjonsmønster som gjennomsnittet, men noen felt har store avvik. Figur 3.4 viser det additive avviket for første prognosedag som funksjon av vassdragsnummer. Både den uprosesserte prognosen og medianen ha en klar tendens til å ha størst avvik i nedbørrike deler av landet, på Vestlandet og i Nordland, men utslagene varierer mye fra felt til felt. Vestlandsfelt med små avvik har en tendens til å ligge langt øst eller ha stort areal (f.eks. Bulken). Tilsvarende er det negativt avvikende Nordlandsfeltet et lite, vestlig eksponert felt (Vassvatn). Usikkerhetsmodellen har en viss evne til å redusere variasjonen fra felt til felt, men nedbørprognosen blir i det store og hele ikke bedre.



Figur 3.3. Additivt avvik for alle prognosefelt som funksjon av prognoselengden. Øverste figur viser resultatene for den uproesserte prognosen og nederste for medianen i utfallene fra usikkerhetsmodellen.



Figur 3.4. Additivt avvik for prognosedag 1 for alle prognosefelt sortert etter vassdragsnummer. Øverste figur viser resultatene for den uprosesserte prognosen og nederste for medianen i utfallene fra usikkerhetsmodellen.

3.2.2 Varsling av nedbørmengder

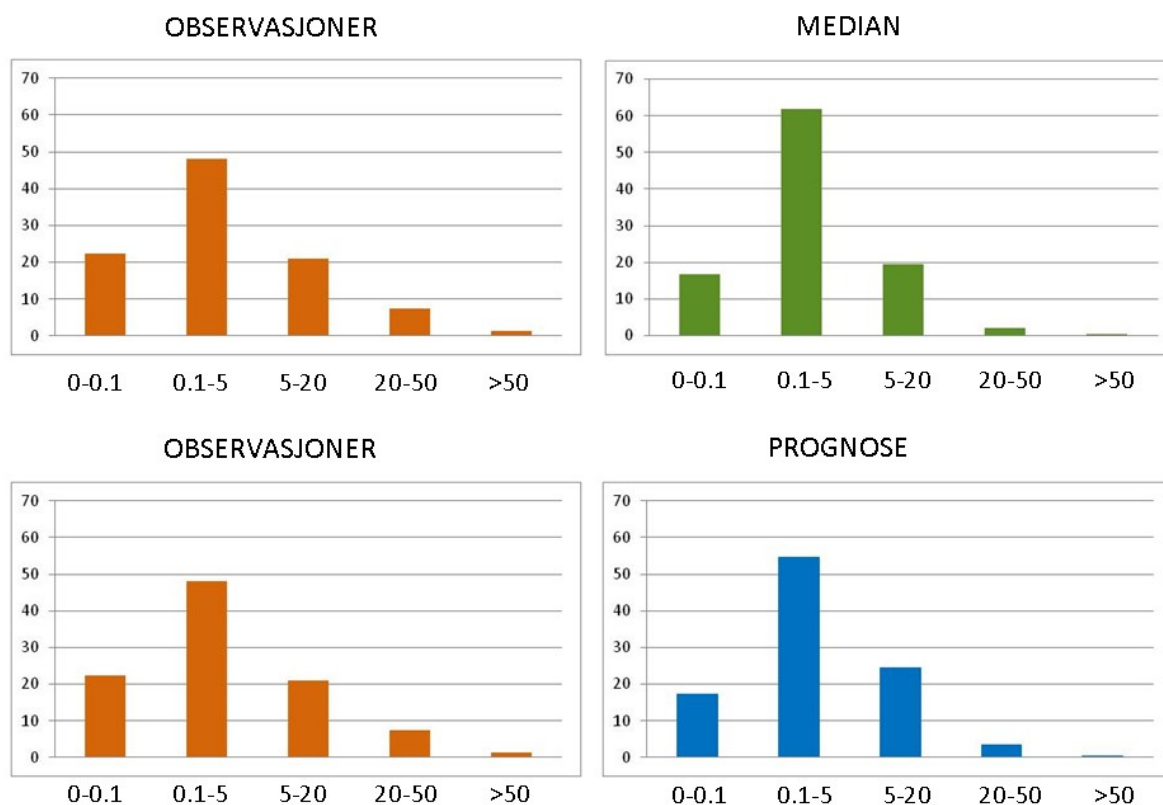
Selv om prognosene underestimerer nedbøren, gjelder ikke det alle nedbørhendelser. Prognosene har en tendens til å overestimere nedbøren på dager med lite eller ikke noe nedbør. Tabellen i figur 3.5 viser hvor mange av de varsle nedbørhendelsene (døgnnedbør) som faller i samme mengdekategori som observert nedbør. Første kolonne viser hvor mange dager det ble observert 0.1 mm eller mindre nedbør, altså i praksis dager med oppholdsvær (134 dager). Medianen og prognosen ligger på tilsvarende nivå i snaut halvparten av disse tilfellene. I de fleste tilfellene varsles det mer nedbør enn det faller i denne kategorien. Det varsles like mange (prognosen) eller flere (medianen) dager med 0.1 – 5 mm nedbør, og til og med noen dager med 5 – 20 mm nedbør. Totalt varsles det færre dager med oppholdsvær enn det observeres, nærmere bestemt 25% færre. Både for prognosen og medianen faller over halvparten av dagene i småregnkategorien, 0.1 – 5 mm. I de fleste tilfellene sammenfaller mengden med det som blir observert, men rundt en tredjedel fordeler seg på nabokategoriene. Regn i kategorien 5 - 20 mm varsles under halvparten av gangene av medianen, som oftere varsler disse hendelsene som småregn. Prognosen har en bedre treffprosent, og varsler småregn for en drøy tredjedel av dagene. Dagene med mye regn varsles i liten grad av både medianen og prognosen, selv om prognosen har flere sammenfallende varsler enn medianen. Av de 8 dagene med mer enn 50 mm nedbør varsles ingen av medianen, og bare en av prognosen.

Vi ser altså at både medianen og prognosen har en tendens til å varsle småregn på bekostning av både antall dager med oppholdsvær og dager med mere regn. Denne tendensen vises tydelig i figur 3.6, der andelen dager i de ulike kategoriene er plottet. Tendensen er tydeligst for medianen. Figur 3.7 viser hvordan tendensen forsterkes med prognoselengden.

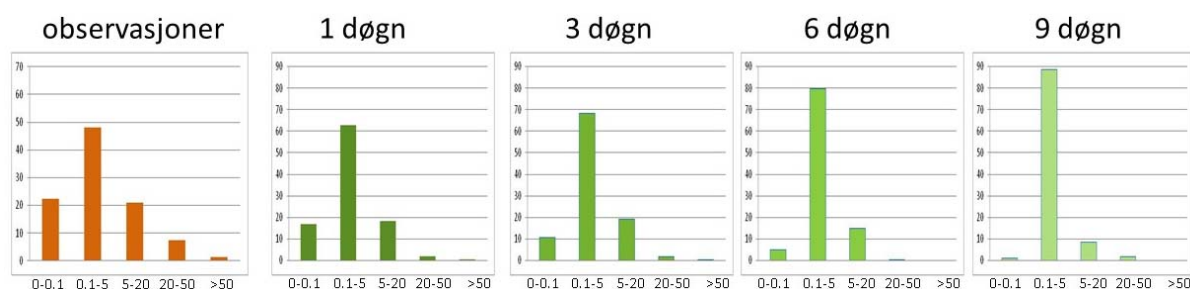
OBSERVASJONER						
	0-0.1	0.1-5	5-20	20-50	>50	
MEDIAN	63	38	1	0	0	101
	70	232	68	6	0	376
	1	17	55	33	4	110
	0	0	1	5	4	11
	0	0	0	0	0	0
	134	287	125	44	8	599

OBSERVASJONER						
	0-0.1	0.1-5	5-20	20-50	>50	
PROGNOSE	66	37	0	0	0	103
	66	212	45	4	0	327
	2	37	75	30	3	147
	0	1	5	11	5	21
	0	0	0	0	1	1
	134	287	125	44	8	599

Figur 3.5. Tabell som viser antall nedbørhendelser (døgnedbør) med mengder innenfor de angitte intervallene (mm), gjennomsnitt for alle felt. Figuren til venstre viser medianen sammenstilt med observasjonene, og figuren til høyre prognosen sammenstilt med observasjonene. Antallet i diagonalene viser hvor mange prognosehendelser som sammenfaller med observasjoner mht. nedbørmengde.



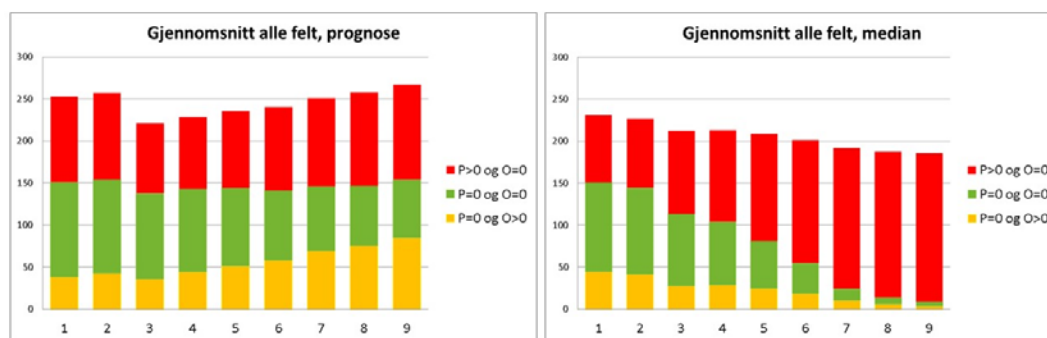
Figur 3.6. Andelen nedbørhendelser (døgnedbør) fordelt på mengdekategorier (mm). Observasjoner er sammenstilt med medianen i utfallene fra usikkerhetsmodellen (øverst) og med den uprosesserte prognosen (nederst). Prognoselengden er en dag. Figuren er basert på samme datagrunnlaget som figur 3.5.



Figur 3.7. Andelen nedbørhendelser (døgnnedbør) fordelt på mengdekategorier (mm). Observasjoner er sammenstilt med medianen i utfallene fra usikkerhetsmodellen, for 1, 3, 6 og 9 døgns prognoselengde.

3.2.3 Nedbør - ikke nedbør

Feilmodellen for nedbørprognosen er todelt, slik at det først avgjøres hvorvidt det blir nedbør eller ei, ut fra en beregnet sannsynlighet, og deretter hvor mye nedbør som kommer, gitt at det blir nedbør. Figur 3.8 viser resultatene av en analyse av hvorvidt prognosen og feilmodellen treffer når det gjelder å varsle nedbør kontra ikke nedbør, avhengig av prognosedag. Alle hendelser der enten prognosen/medianen eller observasjonen er 0.0 er analysert.



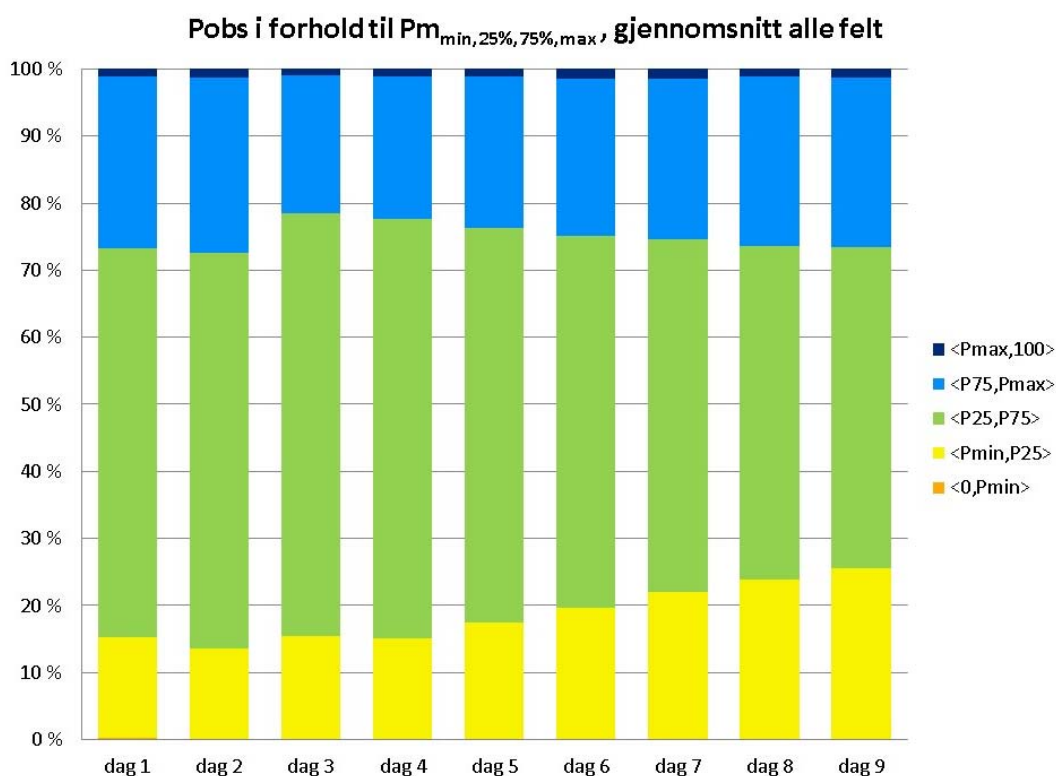
Figur 3.8. Sammenfall av hendelser (døgnnedbør) sortert etter nedbør eller ikke nedbør. Grønne søyler viser antall dager med oppholdsvær som var riktig varslet, røde søyler viser dager med oppholdsvær der det var varslet nedbør, og gule søyler viser regnværsdager der det var varslet oppholdsvær. Antall regnværsdager med varslet nedbør er ikke vist i figuren.

Den uprosesserte prognosen har en tendens til å gi dårligere treff når det gjelder oppholdsdager når prognoselengden øker. Antallet oppholdsdager er omtrent konstant, men færre av dem treffer realiserte oppholdsdager. Dette kan tolkes som et utslag av at prognosen generelt treffer dårligere ved større prognoselengde.

Medianen i usikkerhetsmodellen gir et dramatisk fall i varsle oppholdsdager med prognoselengden, både de som treffer og de som bommer på realiserte oppholdsdager. Dette er også et utslag av generelt økt usikkerhet med økende prognoselengde. Flere dager fram i tid, vil generelt sannsynligheten for null nedbør avta, dvs. det er mer åpent for innslag av litt regn, i forhold til ved et sannsynlig oppholdsvær i nær framtid. Etter hvert vil det være få tilfeller der 50% av utfallene gir oppholdsvær, og medianen vil ligge

over null, altså i det røde søyleområdet eller i den gruppen av hendelser som ikke er presentert i diagrammet (både medianen og observasjonen gir nedbør), og som vi ser øker i antall utover i prognoseperioden, siden de representerte gruppene avtar. Medianen i utfallene fra denne sannsynlighetsbaserte feilmodellen gir altså et svært skjevt bilde av andelen dager med oppholdsvær etter hvert som prognoselengden og prognoseusikkerheten øker. Den generelle usikkerheten langt fram i tid, vil med denne feilmodellen gi flere midt-på-treet-hendelser (dager med småregn) på bekostning av hendelser i ytterkanten av utfallsrommet, altså dager med oppholdsvær og dager med mye nedbør. Trenden som viser økende andel nedbørhendelser i 0.1 – 5 mm-kategorien i figur 3.7 er en del av det samme bildet.

3.2.4 Sannsynlighetsfordelingen for nedbør



Figur 3.9. Fordelingen av observert nedbør i forhold til utfallene fra usikkerhetsmodellen som funksjon av prognosedag, gjennomsnitt for alle felt. Oransje søyle viser andelen nedbørobservasjoner som ligger under laveste utfall, gul søyle viser andelen observasjoner mellom laveste utfall og 25-persentilen, grønn søyle viser andelen observasjoner innenfor medianen +/- 25% av utfallene, lys blå søyle viser andelen observasjoner mellom 75-persentilen og høyeste utfall og mørk blå søyle viser andelen observasjoner over høyeste utfall.

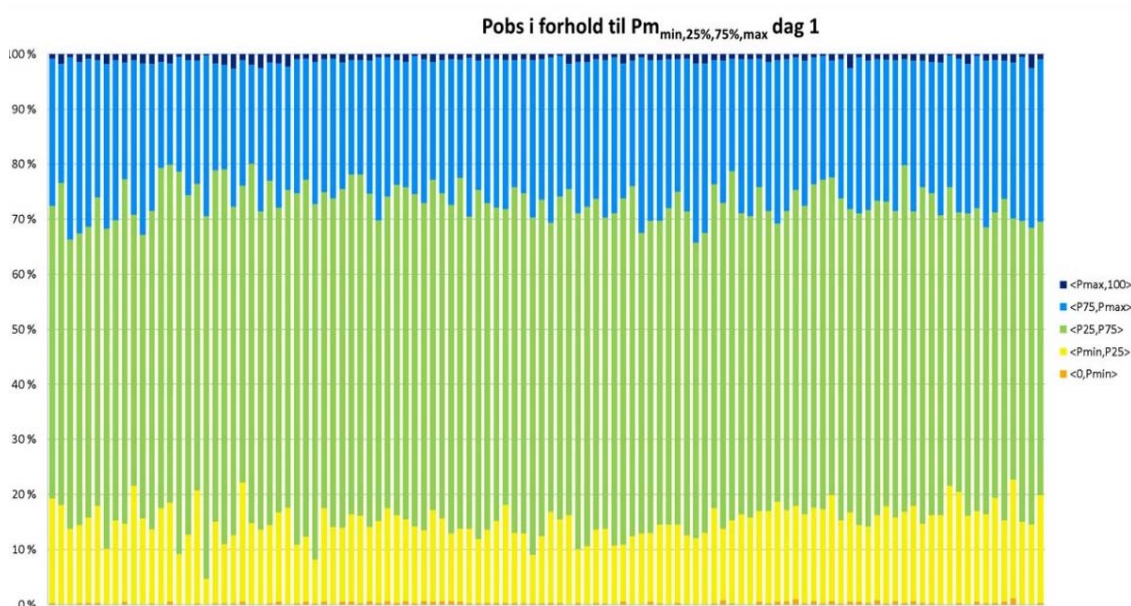
Det er undersøkt hvordan den observerte nedbøren fordeler seg i forhold til usikkerhetsmodellens daglige estimerte sannsynlighetsfordeling.

Sannsynlighetsfordelingen beskriver relativt bra det observerte været, selv om tendensen til å varsle mye småregn på bekostning av oppholdsvær og store nedbørmengder gir fordelingen en tydelig skjevhet. Det er betydelig færre enn 25% av de observerte nedbørmengder som ligger under 25-persentilen. Parallelt med at medianen har en

tendens til å ligge i småregnsjiktet, definerer også de 25% laveste utfallene en for lav nedre grense på konfidensintervallet. Omtrent ingen ligger under laveste utfall, fordi laveste utfall ofte er null.

Usikkerhetsmodellen treffer bedre i overkant av fordelingen, og omtrent 25% av observasjonene ligger over 75-persentilen. De mørkeblå søylene viser at usikkerhetsmodellen ikke fanger opp de største nedbørmengdene, dvs. ingen av utfallene når de høyeste verdiene. Konfidensintervallet smalner utover i prognoseperioden, og definerer faktisk det beste konfidensintervallet på dag ni. Selv om medianen har en tendens til å legge seg for ofte i småregnkategorien (jfr. fig. 3.7), definerer spredningen av utfallene likevel et brukbart 50% konfidensintervall.

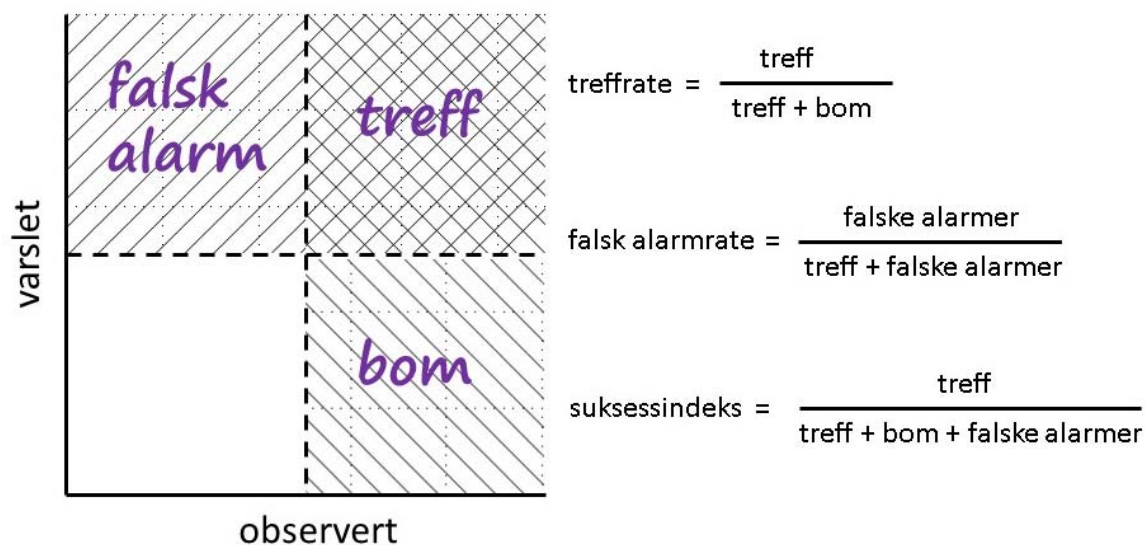
Figur 3.10. Viser resultatene for alle felt for første prognosedag. Spredningen rundt gjennomsnittet, som er vist i figur 3.9, kan sies å være moderat, det er få iøynefallende avvik fra middelverdien.



Figur 3.10. Fordelingen av observert nedbør i forhold til utfallene fra usikkerhetsmodellen for prognosedag 1. Resultatene for alle felt er vist. Se figur 3.9 for forklaring.

3.2.5 Treff, bom og falske alarmer

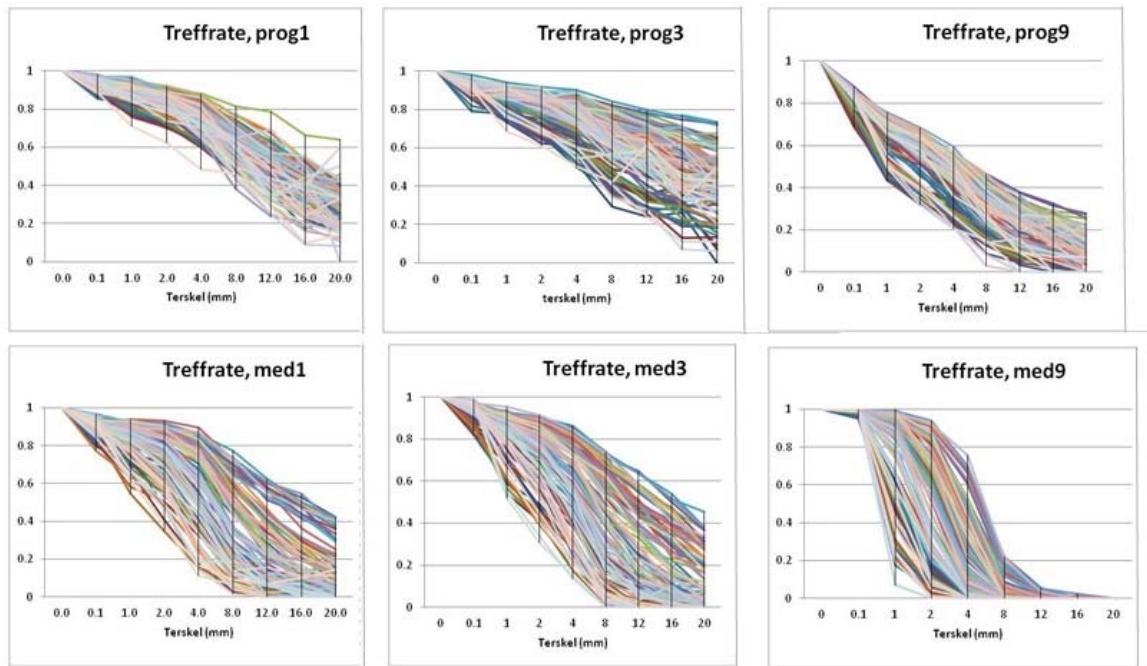
En annen måte å måle prognosenes evne til å varsle ulike nedbørmengder på er å se på treffraten til prognosen som funksjon av nedbørmengden. Treffraten bør alltid ses i sammenheng med falsk alarmraten, for å ta i betraktning varsler som ikke korresponderer med noen observasjon. Figur 3.11 illustrerer begrepene treff, bom og falsk alarm, og definerer treffrate og falsk alarmrate. I tillegg defineres en syntese av treff- og falsk alarmraten, suksessindeksen, der både bommer og falske alarmer trekker ned.



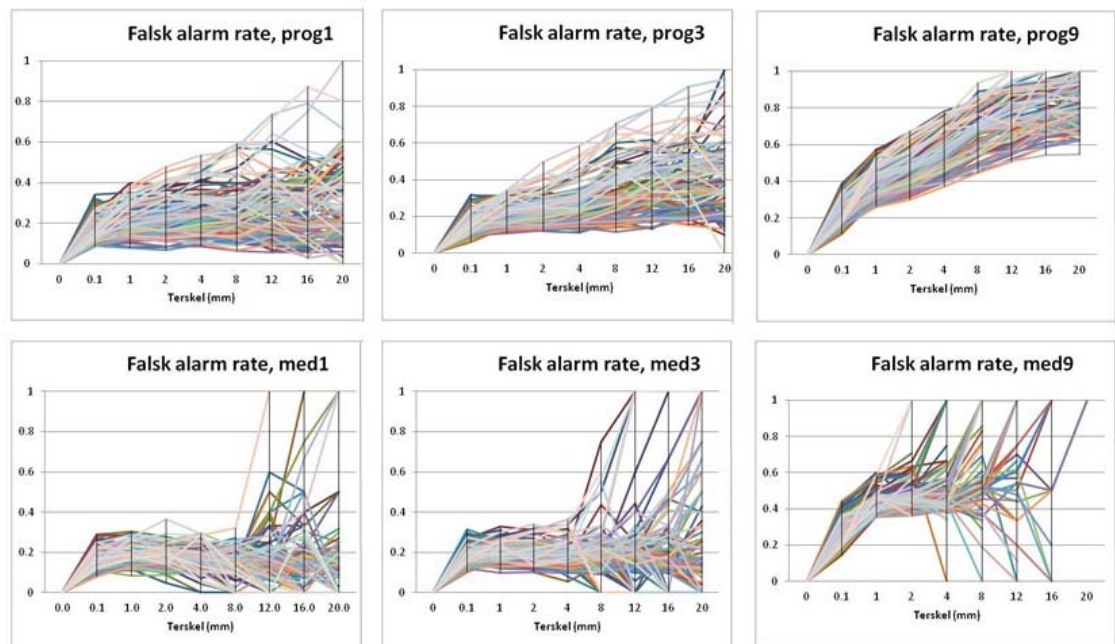
Figur 3.11. Illustrasjon av begrepene treff, bom og falsk alarm. Prognosen treffer når det observeres tilsvarende, den bommer hvis observasjonen ligger over terskelen mens varselet ligger under, og den gir en falsk alarm hvis varselet ligger over terskelen mens observasjonen ligger inder. Figuren kan gjelde for forskjellige variabler, og terskelene kan defineres på ulike måter.

Figur 3.12 viser treffraten for alle flomvarslingsfeltene som funksjon av økende terskelverdi (logaritmisk) og for prognoselengde 1, 3 og 9 dager. Den uprosesserte prognosen ligger på et høyere nivå enn medianen, og har en treffrate på over 50% for døgnnedbør under ca. 12 mm. 3-dagersprognosen fra EC-modellen har samme nivå, men litt større spredning, mens treffraten for 9-dagersprognosen raskt synker under 50%. Treffraten for medianen har stor spredning over feltene, og ligger i gjennomsnitt under 50% for nedbørverdier over ca. 4 mm ved prognoselengde 1 dag. Nivået er omtrent tilsvarende for 3-dagersprognosen, men raten avtar svært raskt for mange av feltene for prognoselengde 9 dager.

Falsk alarmraten, som vises på figur 3.13, er lav for prognoselengde 1 og 3 dager for både den uprosesserte prognosen og medianen, og betydelig høyere for 9-dagersprognosen, der rundt halvparten av prognosene over 2 mm er falske alarmer, for begge prognoser. Grafene som viser resultatene for medianen har mange tilfeller av 0- og 1-verdier, som oppstår når datamaterialet ikke inneholder hendelser over terskelverdien. Som vi har sett tidligere, er det få tilfeller av større nedbørhendelser i prognosene, spesielt medianen. Både treff- og falsk alarmraten har en tydelig tendens til å vise svært dårlige resultater for de høyere terskelverdiene, men resultatene må tolkes med forbehold om at datamaterialet inneholder få større nedbørhendelser.

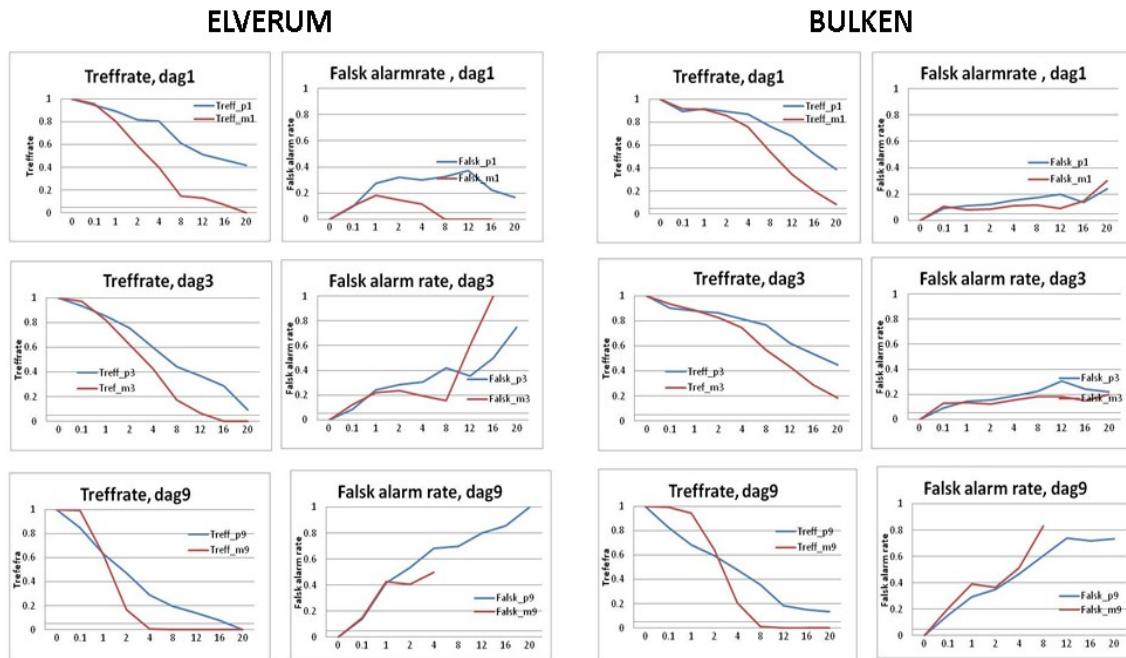


Figur 3.12. Treffrate for nedbørprognosene som funksjon av ulike terskelverdier, logaritmisk skala. Den uprosesserte prognosen vises i øverste rad og medianen i utfallene fra usikkerhetsmodellen i nederste. Figurene viser resultatene for alle modellfelt og for prognoselengde 1, 3 og 9 dager.



Figur 3.13. Falsk alarmrate for nedbørprognosene som funksjon av ulike terskelverdier, logaritmisk skala. Den uprosesserte prognosen vises i øverste rad og medianen i utfallene fra usikkerhetsmodellen i nederste. Figurene viser resultatene for alle modellfelt og for prognoselengde 1, 3 og 9 dager

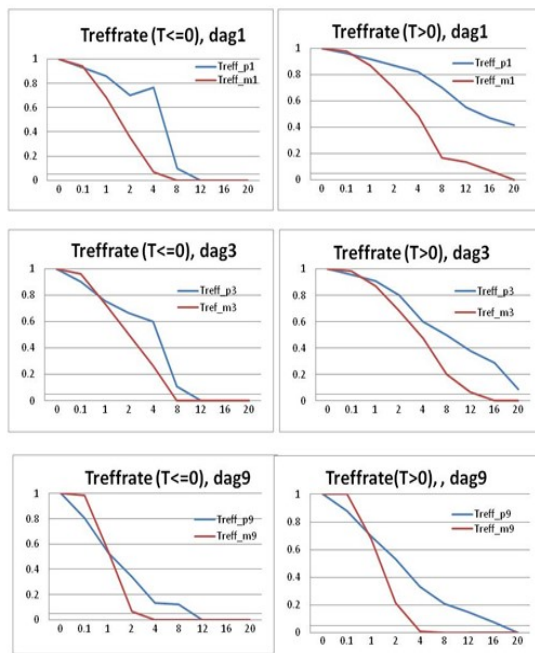
I figur 3.14 er resultatene for et Østlandsfelt, Elverum, og et Vestlandsfelt, Bulken, presentert. Treffraten ligger høyere i det nedbørrike Vestlandsfeltet enn i Elverum, men selv der treffer medianen under 50% av hendelsene over 10 mm første prognosedøgn. Den uprosesserte prognosen klarer 50% av hendelsene opp til 18 mm. Til gjengjeld er falsk alarmraten lav de første prognosedøgnene. For begge felt gjelder det at både treff- og falsk alarmraten skårer svært dårlig for prognosedag 9.



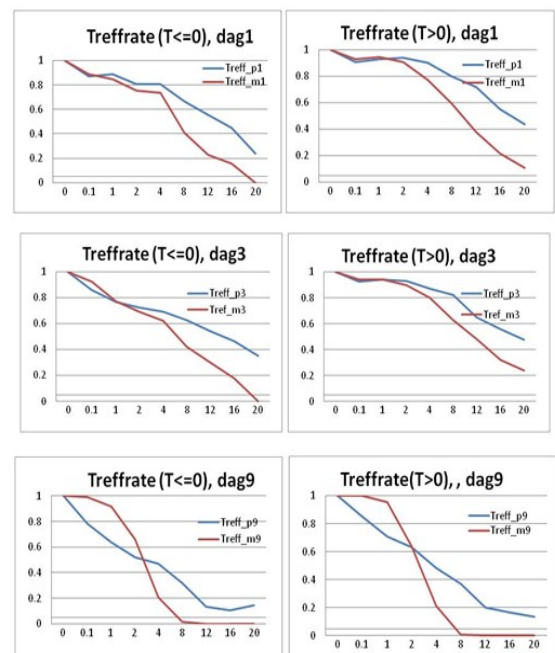
Figur 3.14. Treffrate og falsk alarmrate for nedbørfeltene Elverum og Bulken. Blå linjer viser resultatene for den uprosesserte prognosen og røde linjer for medianen i utfallene fra usikkerhetsmodellen.

Figur 3.15 viser at det gjennomgående er høyere treffrate for større nedbørmengder på dager med varmegrader enn på dager med minusgrader. Større nedbørmengder opptrer trolig oftere i varmt vær, noe det tidlige null-nivået på kuldegradskurvene for Elverum også antyder. Dette er et gunstig sammenfall med når prognosen treffer best. For øvrig er det ingen iøynefallende forskjeller på hvordan treffraten varierer for temperaturer under og over null.

ELVERUM



BULKEN

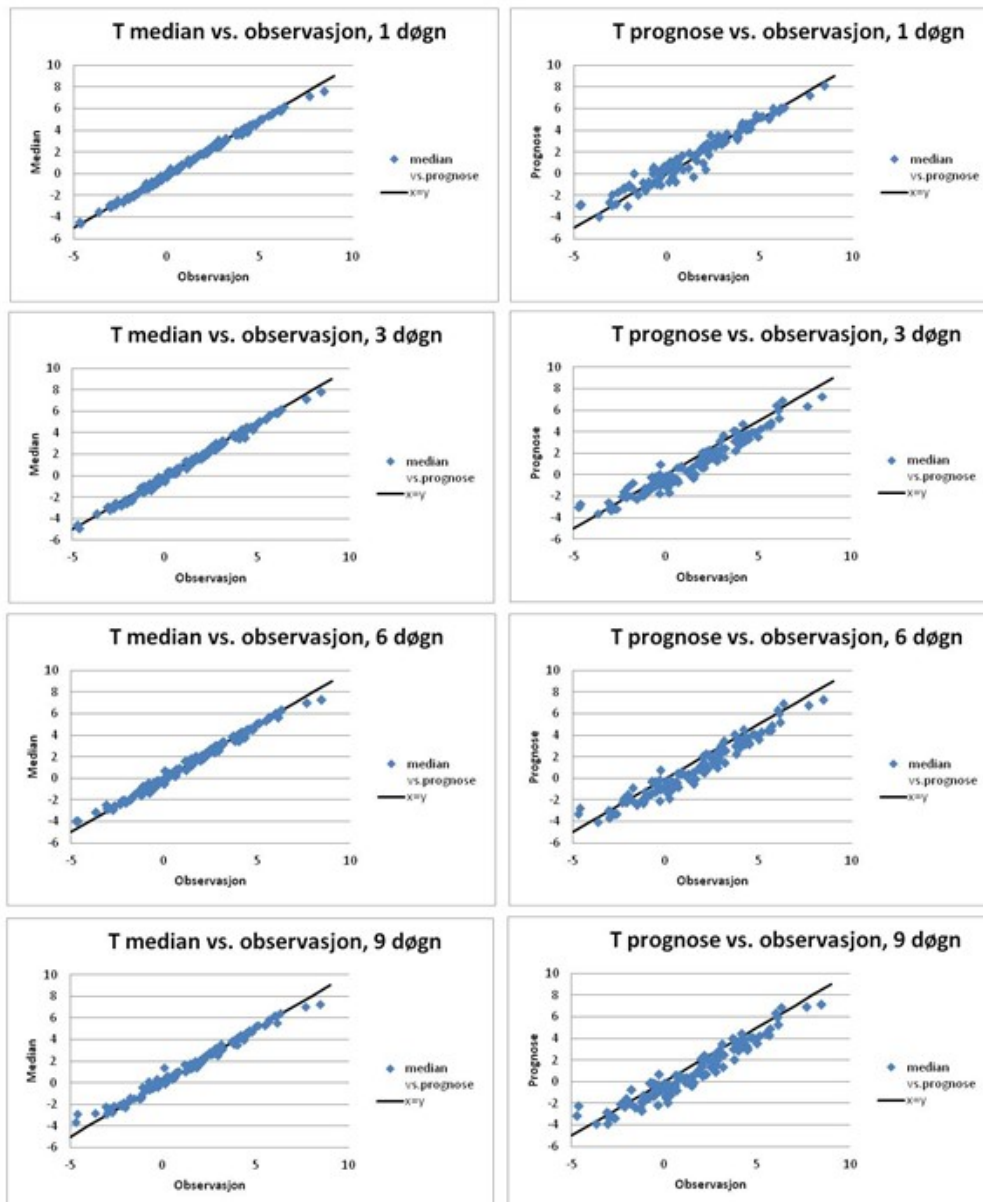


Figur 3.15. Treffrate for nedbørfeltene Elverum og Bulken. Treffraten for dager med temperatur lavere enn null er vist til venstre og for dager med temperatur høyere enn null til høyre for hvert felt. Blå linjer viser resultatene for den uprosesserte prognosen og røde linjer for medianen i utfallene fra usikkerhetsmodellen.

3.3 Evaluering av feilmodellen for temperaturprognosen

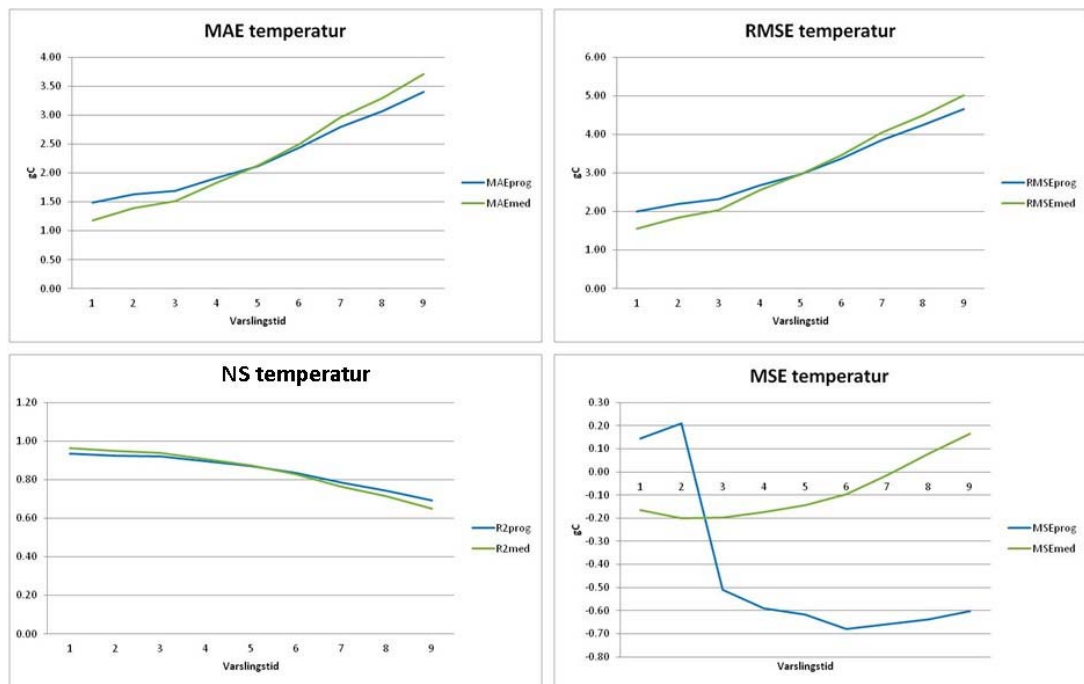
3.3.1 Hvor godt treffer modellen?

Feilmodellen for temperaturprognosen fungerer, i motsetning til feilmodellen for nedbørprognosen, etter hensikten. Korrigeres det for den estimerte feilen, får vi en forbedret prognose. Figur 3.16 viser punktdiagrammer av medianen i temperaturutfallene fra usikkerhetsmodellen, som er korrigert for estimert feil, plottet mot observasjoner (venstre kolonne). Punktssvermen er konsentrert rundt diagonalen i diagrammet, og sammenfallet er betydelig bedre enn for den uprosesserte prognosen i høyre kolonne, som har større spredning og viser en underestimering av temperaturen for varslingslengde fra tre dager. Det er middeltemperaturen over evalueringsperioden for alle modellfelt som vises.



Figur 3.16. Punktdiagrammer som viser medianen i utfallene fra sannsynlighetsfordelingen for temperatur plottet mot tilsvarende observasjoner (venstre kolonne) og den uprosesserte prognosen plottet mot observasjoner (høyre kolonne). Punktene viser middelveidien over evalueringsperioden for hvert nedbørfelt, og viser resultatene for prognoselengde 1, 3, 6 og 9 døgn.

Kvantitative mål på sammenfallet mellom temperaturprognoser og observasjoner er vist på figur 3.17. Se avsnitt 3.2.1 for definisjon av feilfunksjonene. Medianen viser best sammenfall med observasjonene de første 5 dagene, målt med alle feilfunksjoner. De siste prognosedagene gir medianen en dårligere skår enn den uprosesserte prognosen i absolutt forstand. Det additive avviket viser imidlertid at prognosen underestimerer temperaturen, mens dette er korrigerte for i medianen, slik at positive og negative avvik fra observasjonene utligner hverandre i stor grad.



Figur 3.17. Fire feilestimater som viser hvor godt den uprosesserte prognosen og medianen i utfallene fra usikkerhetsmodellen treffer observasjonene. Prognoselengden er vist på x-aksen.

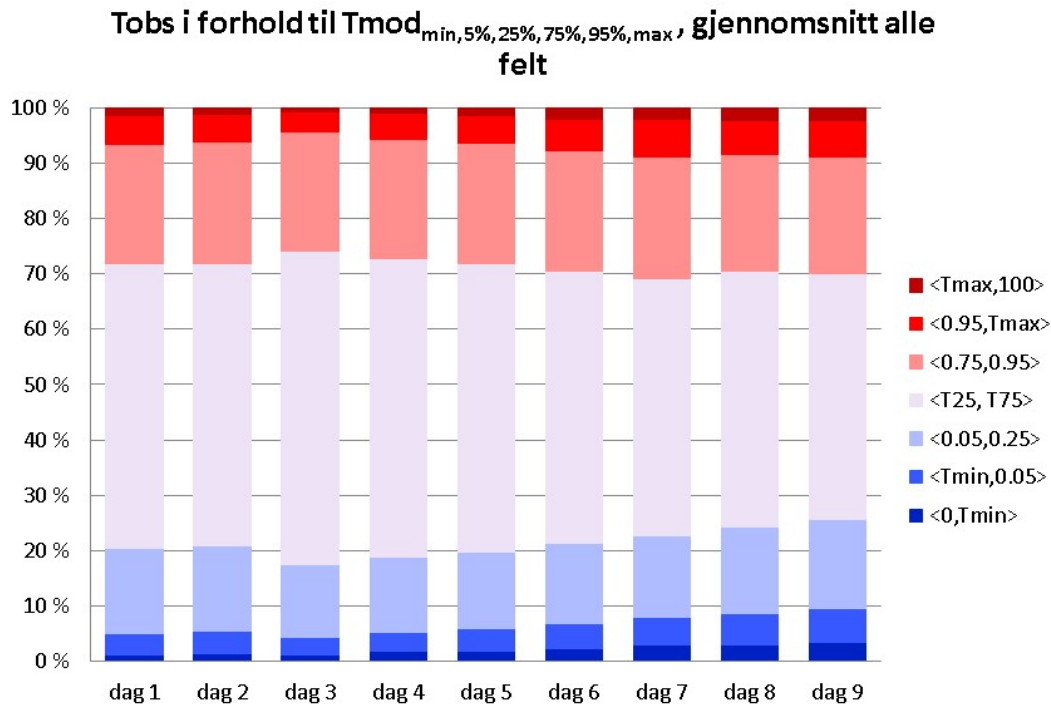
3.3.2 Sannsynlighetsfordelingen for temperatur

Usikkerhetsmodellen genererer hver dag en empirisk sannsynlighetsfordeling for temperaturen i prognoseperioden. Figur 3.18 viser hvordan den observerte døgntemperaturen fordeler seg i forhold til usikkerhetsmodellens persentiler og minimums- og maksimumsutslag hver dag i prognoseperioden.

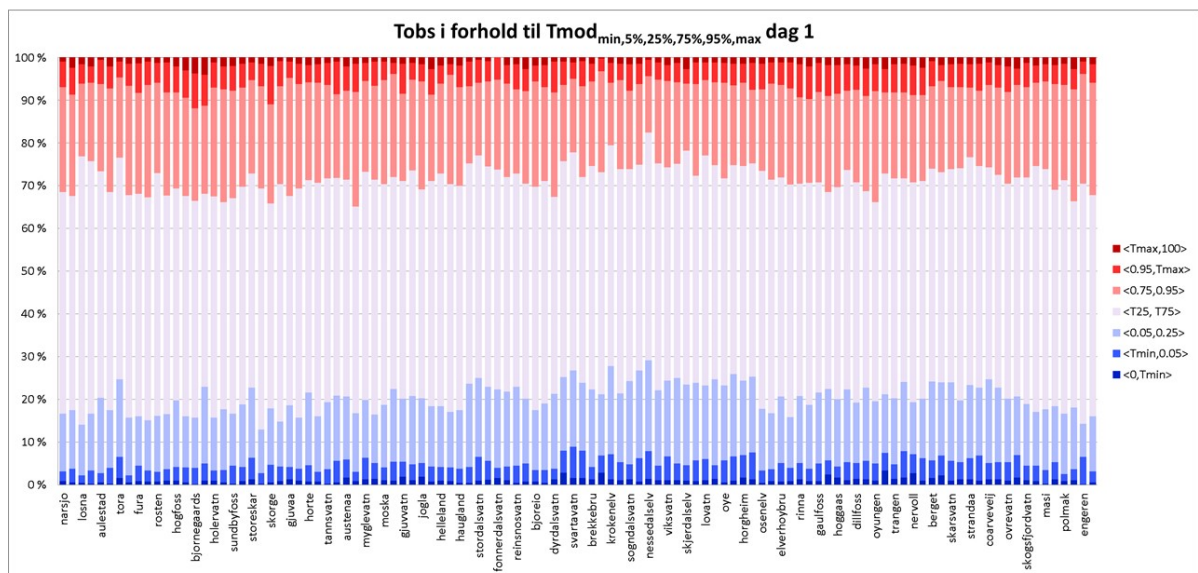
Vi ser at rundt halvparten av observasjonene havner i 50% konfidensintervallet, mellom 25- og 75-persentilen. De øvrige 50 prosentene fordeler seg skjevt over og under konfidensintervallet. Rundt 30% av observasjonene ligger over 75-persentilen, og et par prosent ligger over høyeste utfall. Tilsvarende ligger ca. 20% under 25-persentilen, og også her i nedre temperatursjikt ligger et par prosent utenfor utfallsrommet.

Utfallsrommet er relativt smalt, og gjennomsnittsmmedianen viser ikke noen tydelig underestimering av temperaturen i figur 3.16. Også de kvantitative feilestimatene viser at avvikene er små. Alt i alt beskriver den korrigerede temperaturen og usikkerhetsmodellen den observerte temperaturen relativt bra, og det er ikke gjort ytterligere analyser av temperaturmodellens egenskaper.

Figur 3.19 viser den tilsvarende fordelingen av døgntemperaturen i forhold til utfallene i usikkerhetsmodellen for alle felt den første prognosedagen. Tendensen som er beskrevet for gjennomsnittet av feltene er tydelig også for de aller fleste enkeltfeltene, og er relativt konsistent over alle feltene.



Figur 3.18. Fordelingen av observert temperatur i forhold til utfallene fra usikkerhetsmodellen som funksjon av prognosedag, gjennomsnitt for alle felt. De lyse søylene viser andelen observerte verdier innenfor 50%-konfidensintervallet. De røde søylene viser andelen observerte temperaturer mellom 75- og 95-persentilen og høyeste utfall og over høyeste utfall. Blå søyer viser andelen observerte temperaturer mellom 25- og 5-persentilen og laveste utfall og under laveste utfall.



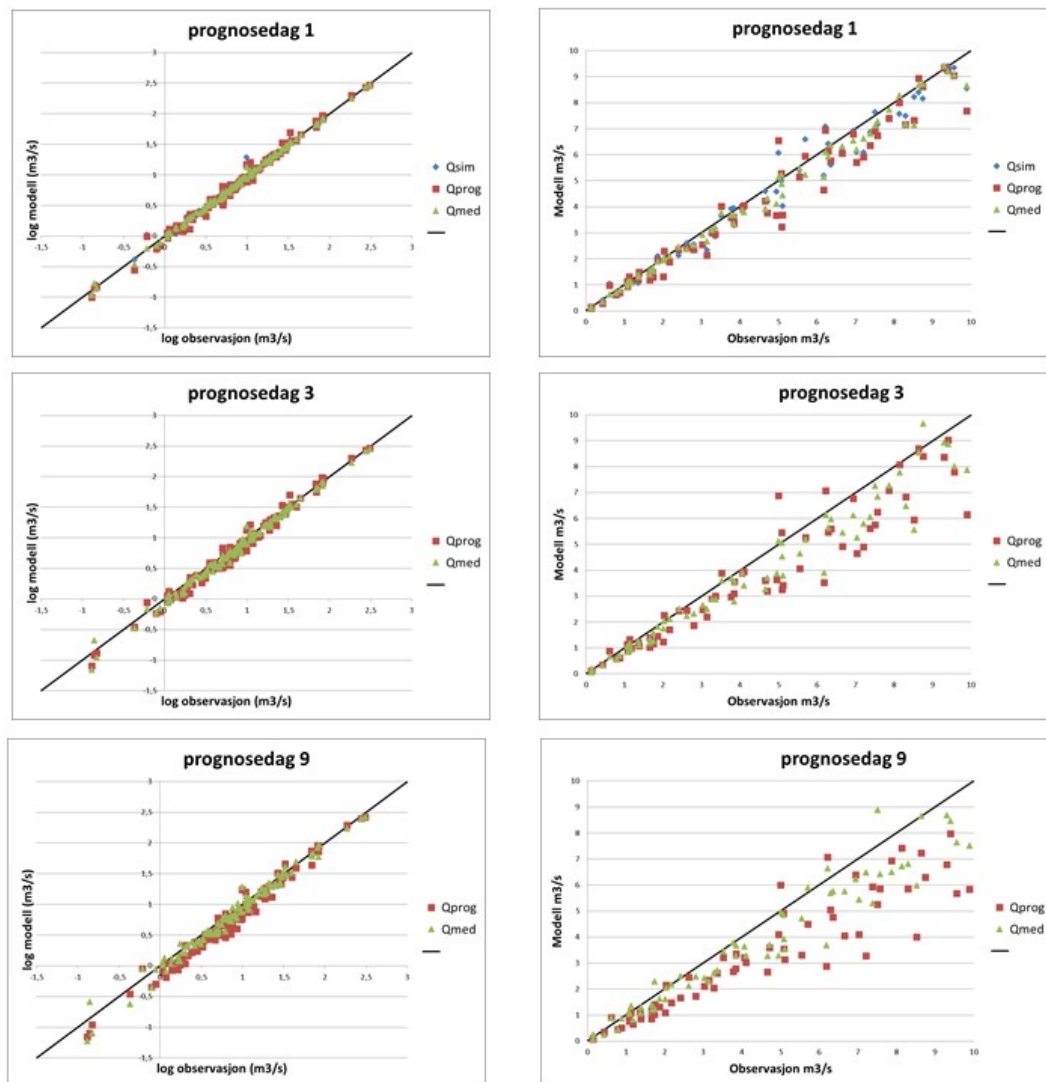
Figur 3.19. Fordelingen av observert temperatur i forhold til utfallene fra usikkerhetsmodellen for prognosedag 1. Resultatene for alle felt er vist. Se figur 3.18 for forklaring.

3.4 Vannføringsprognosen

3.4.1 Hvor godt treffer modellen?

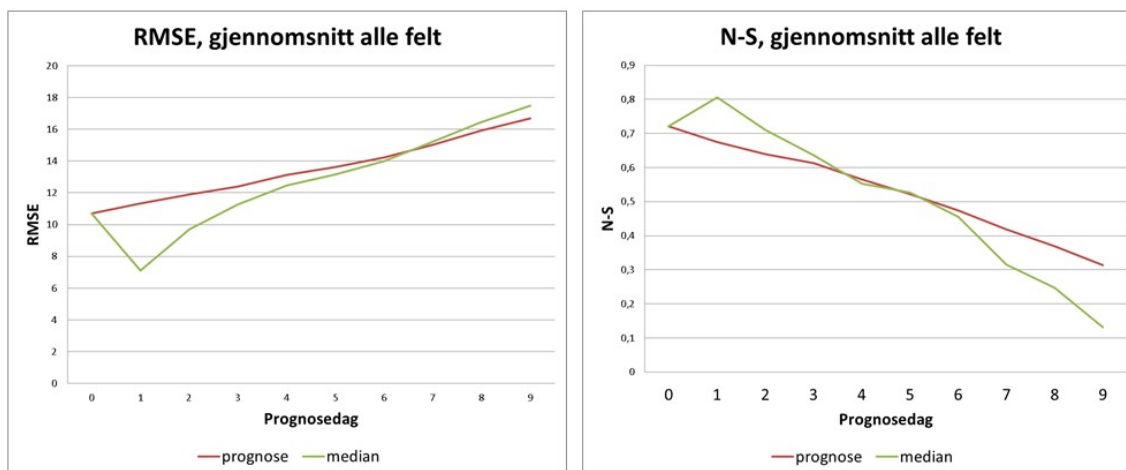
Figur 3.20 viser punktdiagrammer som gir et visuelt inntrykk av hvor godt middelvannføringen over evalueringsperioden korresponderer med observasjonene for den uprosesserte prognosen og medianen i utfallene fra usikkerhetsmodellen for prognosedag 1, 3 og 9. I diagrammet for prognosedag 1 er i tillegg resultatet for modellsimuleringen, altså modellen kjørt med observert vær som inndata.

Alle modellvarianter underestimerer vannføringen, og tendensen øker med prognosedagen. Den uprosesserte prognosen har størst spredning, og underestimerer vannføringen mer enn medianen for prognosedag 9.

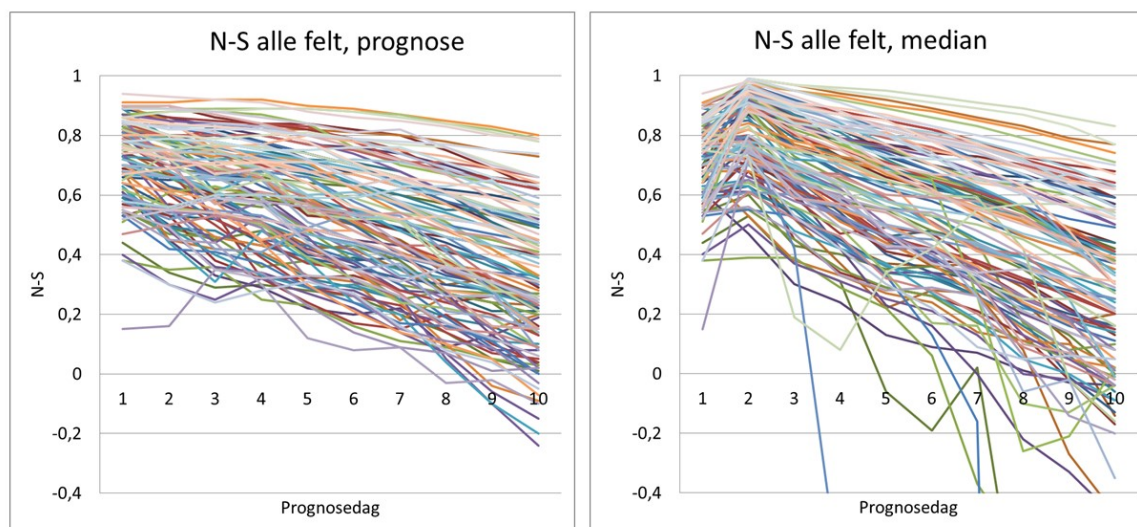


Figur 3.20. Punktdiagrammer for prognosedag 1, 3 og 9. Diagrammet for dag 1 viser modellsimuleringen (med værobservasjoner som inndata, blå ruter), den uprosesserte prognosen (røde firkanter) og medianen i utfallene fra usikkerhetsmodellen (grønne trekanten) plottet mot tilsvarende observasjoner. Punktene viser middelverdien over evalueringsperioden for hvert nedbørfelt. Venstre kolonne viser alle verdier, logtransformert, og høyre kolonne viser verdier under $10\text{m}^3/\text{s}$.

Figur 3.21 viser to kvantitative estimater på sammenfallet mellom prognoser og observasjoner. Her er feilfunksjonene anvendt på tidsseriene over evalueringsperioden for hver stasjon, og gjennomsnittet over alle stasjoner er vist på figuren. Verdien for prognosedag 0 representerer simuleringen med observert vær som inndata. Vi ser at den uprosesserte prognosen har større middelavvik og lavere N-S-skår enn simuleringen for alle tidsskritt. Medianen fra usikkerhetsmodellen er nærmere observasjonen første prognosedag, og er bedre enn prognosen de første dagene. De siste dagene fungerer ikke usikkerhetsmodellen etter hensikten, og den uprosesserte prognosen er best. N-S-kriteriet viser at både modellsimuleringene og prognosene fungerer bra de første 2-3 dagene.



Figur 3.21. To feilestimater som viser hvor godt den uprosesserte prognosen og medianen i utfallene fra usikkerhetsmodellen treffer observasjonene. Prognoselengden er vist på x-aksen.

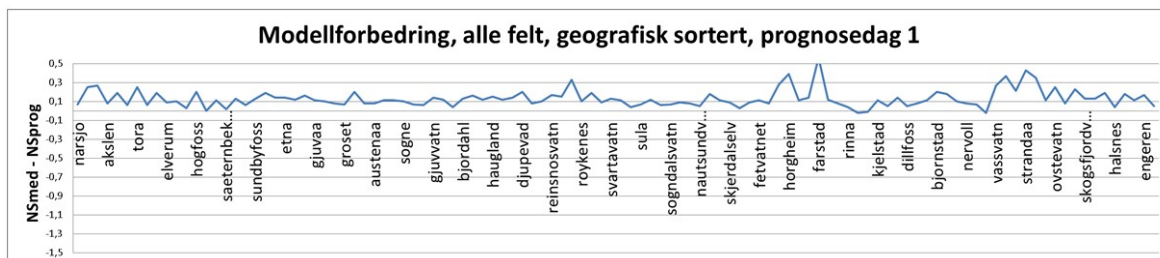


Figur 3.22. NS-kriteriet for alle modellfelt. Resultatene for den uprosesserte prognosen er vist til venstre, medianen fra usikkerhetsmodellen til høyre.

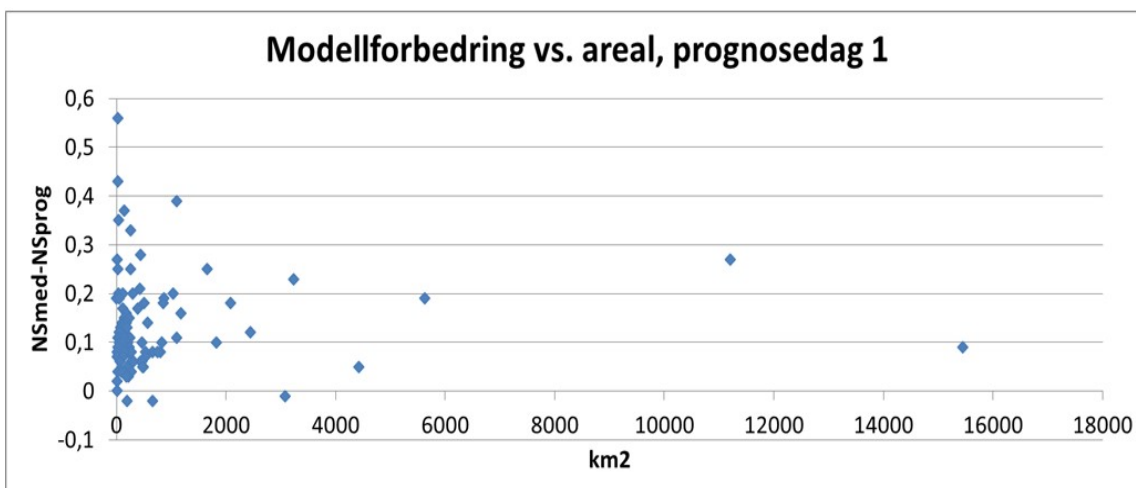
N-S-kriteriet for alle felt er vist på figur 3.22. På tross av stor spredning mellom feltene, ligger N-S-verdien på et akseptabelt nivå for de aller fleste feltene de første dagene. Noen felt har en svært god verdi, og noen få avviker mye i negativ retning, særlig utover i

prognoseperioden. I mediandiagrammet til høyre ser vi noen eksempler på at usikkerhetsmodellen kommer skjevt ut og genererer en dårligere prognose enn den uprosesserte.

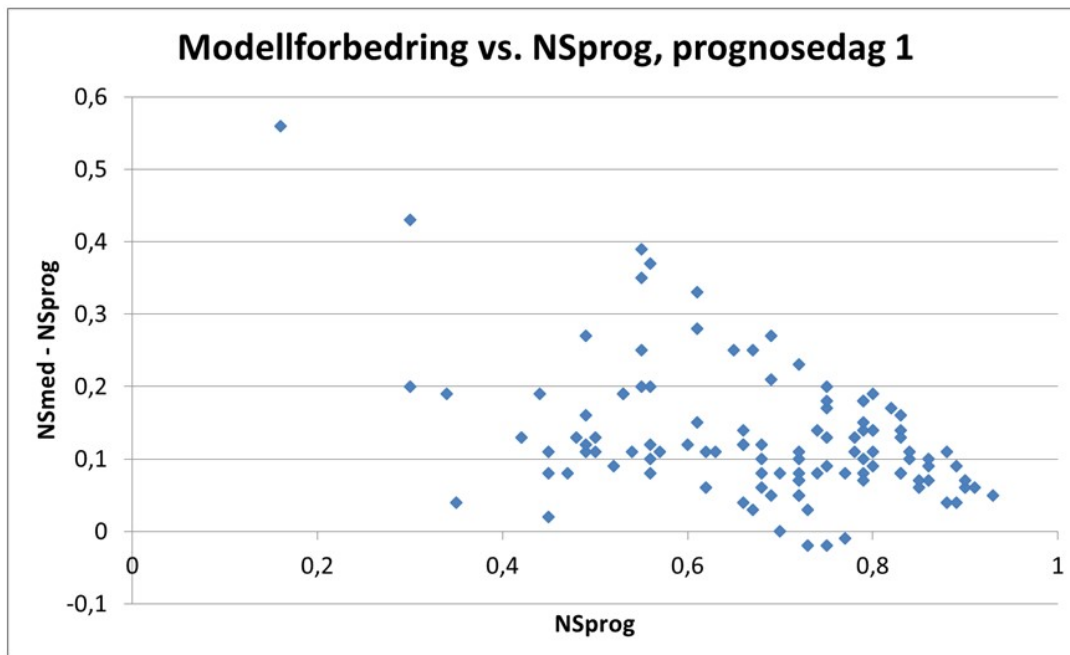
I et forsøk på å se et systematisk mønster i hvilke forutsetninger som skal til for å forbedre prognosen, er modellforbedringen, uttrykt ved differansen i N-S-verdi mellom median og uprosessert prognose, plottet mot geografisk plassering av nedbørfeltene, feltareal og N-S-verdi for prognosen. Figur 3.23 til 3.25 viser resultatene. Det framgår ikke noe åpenbart geografisk mønster av grafen i figur 3.23. Det er vanskelig å trekke noen slutninger om modellforbedringens avhengighet av feltarealet fra figur 3.24, da antall større felt er lite. Figuren antyder at variasjonen er størst for små felt, men bildet påvirkes av at tettheten av små felt er mye større enn for store felt. Forbedringen som funksjon av NS-verdien til prognosen begrenses oppad av linja Forbedring = $1 - NS_{\text{prog}}$, da NS-verdien ikke kan overstige 1.0. Dette begrenser forbedringspotensialet i de gode modellene, men utover det er det ingen åpenbar sammenheng mellom hvor god/dårlig prognosen er i utgangspunktet og hvor mye feilkorrigeringen forbedrer den.



Figur 3.23. Modellforbedring, uttrykt som differansen i NS-verdi mellom medianen i utfallene fra usikkerhetsmodellen og den uprosesserte prognosen, plottet for alle modellfelt sortert etter vassdragsnummer.



Figur 3.24. Modellforbedring plottet mot nedbørfeltets areal.



Figur 3.25. Modellforbedring plottet mot N-S-verdien til den uprosesserte prognosen.

3.4.2 Kvantilplott

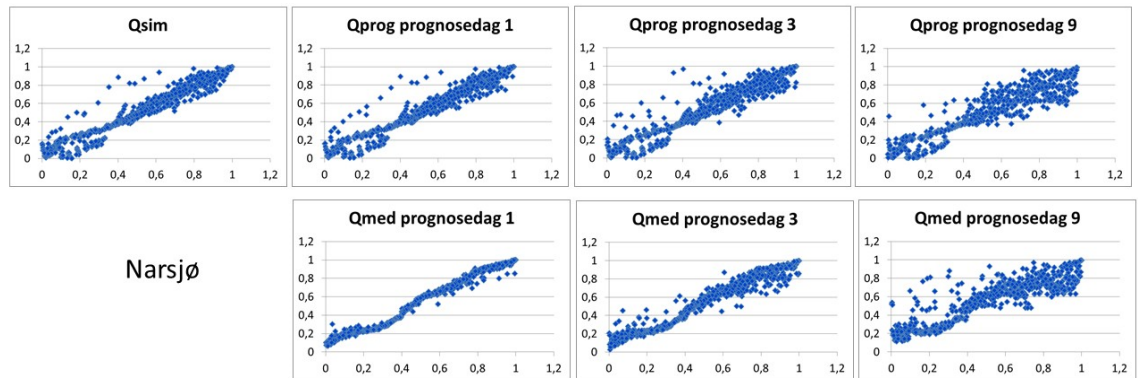
I kvantilplottene sammenliknes fordelingen av prognosene med fordelingen av observasjonene. Døgnvannføringene plottes mot hverandre som kvantiler istedenfor absolute verdier. Dataene sorteres i stigende rekkefølge og fordeles uniformt over utfallsrommet, og korresponderende verdier er plottet mot hverandre. Når det gjelder medianen i utfallene fra usikkerhetsmodellen og den uprosesserte prognosen, rangeres disse i henhold til sorteringen av modellsimuleringene, dvs. modellen kjørt med observert vær som inndata, for at det skal være lettere å sammenlikne resultatene for de forskjellige prognosevariantene og prognosedagene. Modellsimuleringene definerer dermed modellklimaet, på samme måten som de definerer modellens flomnivåer (se kapittel 3.4.3), uavhengig av observerte flomnivåer. Kvantilplottene synliggjør karakteristiske avvik mellom observasjoner og modellsimuleringer. Kvantilplottene for tre modellfelt er presentert, som eksempler. Kvantilplottene har trolig størst potensiale som verktøy for enkeltfelt, og er ikke så godt egnet til å se på de samlede resultatene.

3.4.2.1 Narsjø

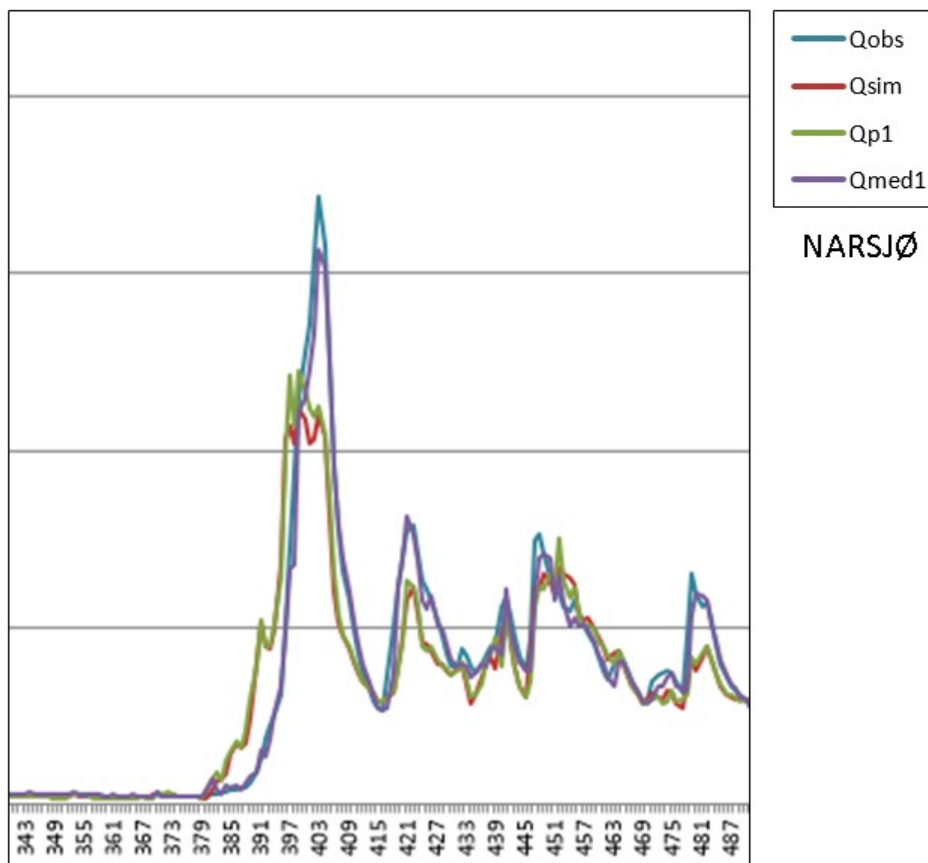
Figur 3.26 viser kvantilplott for Narsjø. Den uprosesserte prognosen viser samme karakteristika som simuleringen den første prognosedagen, som en følge av at værprognosen første dag er godt korrelert med det observerte været. Medianen, som er korrigert for modellfeilen med en autoregressiv modell som tar hensyn til siste observerte vannføring, og gir et kvantilplott med mindre spredning første prognosedag.

Figur 3.27 viser en del av hydrogrammet for Narsjø, der vi ser hvordan medianen følger observasjonen bedre enn simuleringen og prognosen, noe som forklarer det bedre samsvaret i kvantilplottet. Kvantilplottene for simuleringen og den uprosesserte prognosen for dag 1 viser en tydelig rad med for høye kvantiler over den tette prikkvermen. Denne raden er et resultat av en vårflo som ble simulert for tidlig, og som også vises på figur 3.27. Spredningen i kvantilplottene øker med prognosedagen,

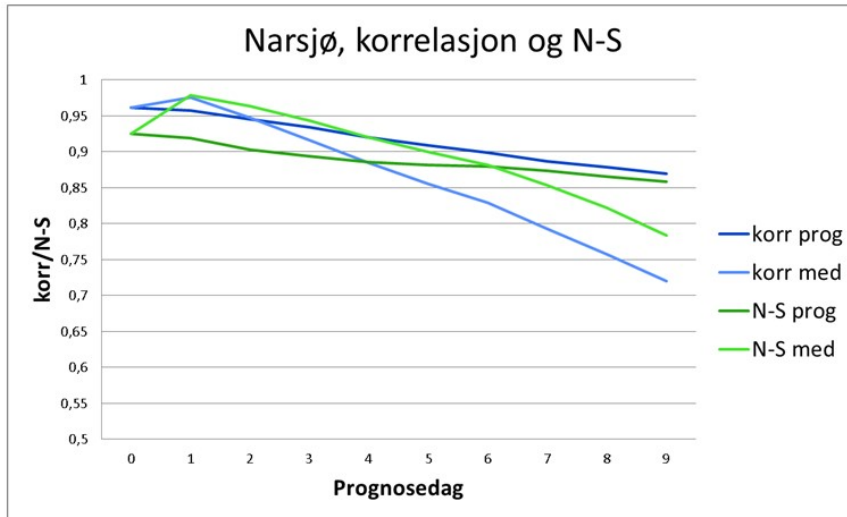
spesielt for medianen, og figur 3.28 viser at feilmodellen beskriver feilen dårligere utover i prognoseperioden. Både korrelasjonskoeffisienten og Nash-Sutcliffe-kriteriet basert på kvantilene avtar bratt med prognosedagen for medianen, men holder seg på et brukbart nivå for den uprosesserte prognosen.



Figur 3.26. Kvantilplott for Narsjø. Observasjonene vises på x-aksen og modellverdiene langs y-aksen. Qsim-plottet viser resultatene for modellsimuleringer med observert vær som inndata, Qprog viser den uprosesserte prognosen og Qmed viser medianen i utfallene fra usikkerhetsmodellen. Resultatene for prognosedag 1, 3 og 9 er vist.



Figur 3.27. Detalj fra hydrogrammet for Narsjø. Observert og simulert vannføring er vist, sammen med den uprosesserte prognosen og medianen i utfallene fra usikkerhetsmodellen for første prognosedag.

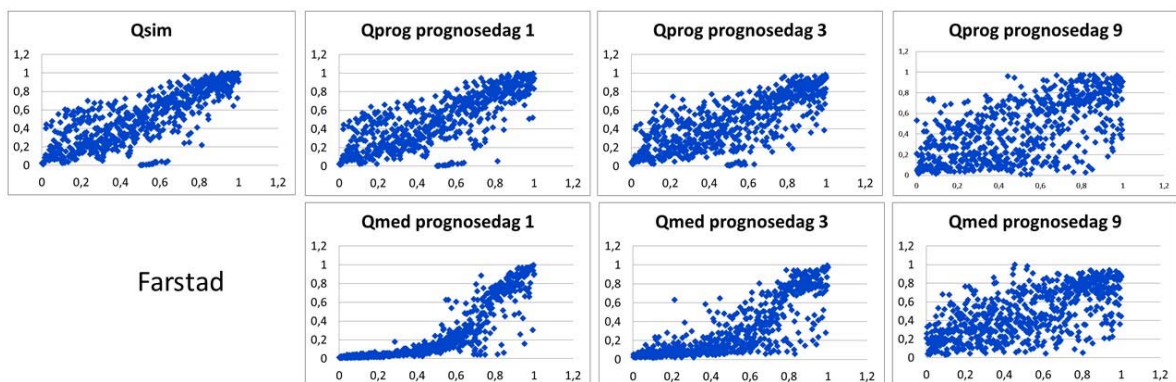


Figur 3.28. Korrelasjonskoeffisienten og Nash-Sutcliffe-kriteriet for den den uprosesserte prognosen og medianen fra usikkerhetsmodellen for Narsjø, basert på kvantilene.

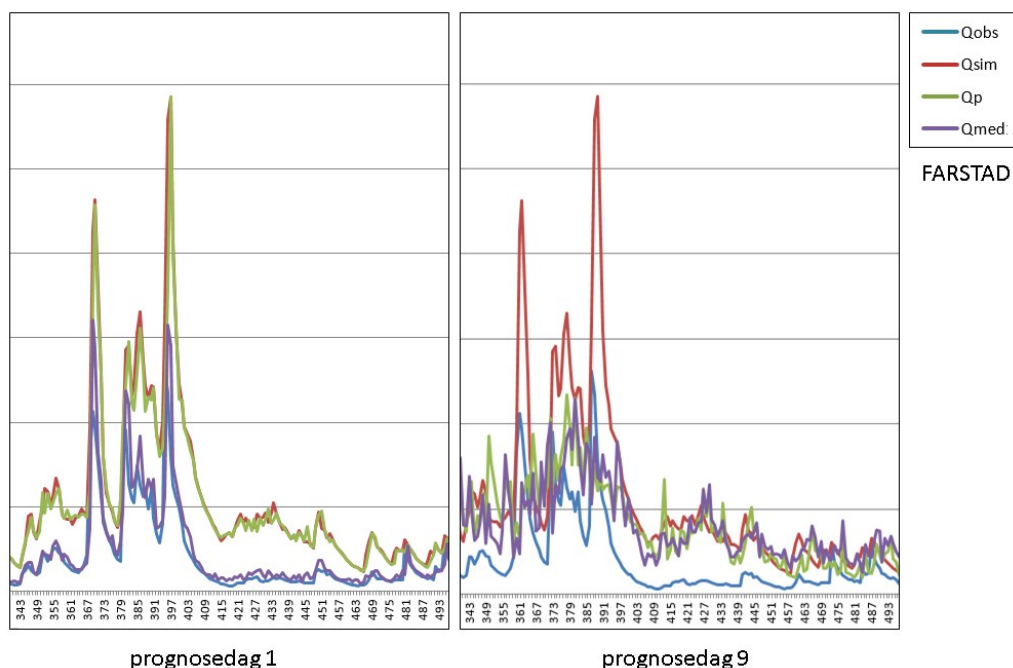
3.4.2.2 Farstad

Kvantilplottet for Farstad viser et endret mønster i punktsvermen for den korrigerte prognosen første prognosedag, særlig for de lavere verdiene, se figur 3.29. Punktene er mer samlet, og det er en større konsentrasjon av punkter rundt lave kvantiler. Hydrogrammet i figur 3.30 viser årsaken til dette mønsteret. Den autoregressive feilmodellen for modellfeilen, oppdaterer vannføringsprognosen med siste observerte verdi. Dette resulterer i at medianen ligger nærmere observasjonene enn den uprosesserte prognosen, som ligger på nivå med simuleringene, ofte på et høyere nivå enn observasjonene. Medianen blir dermed liggende relativt lavt i forhold til simuleringene, som kvantilene er definert ut fra, i store deler av lavvannssesongen, og vi får mange lave kvantilverdier.

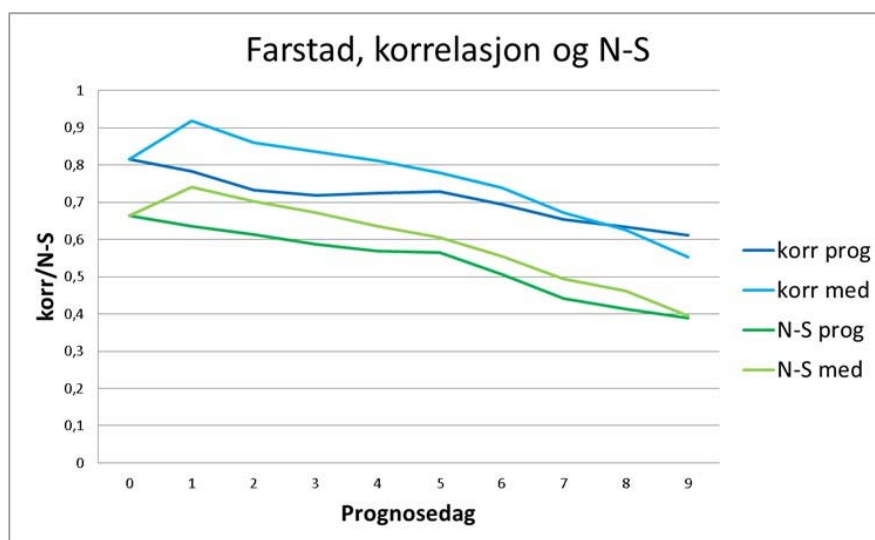
Spredningen i plottene er svært stor den niende prognosedagen. Dette henger sammen med den lave treffsikkerheten til prognosene, som vi ser i høyre del av figur 3.30, både når det gjelder medianen og den uprosesserte prognosen. Den niende prognosedagen ligger langt utenfor rekkevidden til den første ordens autoregressive feilmodellen, som effektivt korrigerer prognosen på dag 1. Medianen har en bedre Nash-Sutcliffe-verdi enn den uprosesserte prognosen gjennom hele prognoseperioden, men ligger på bare 0.4 for dag ni.



Figur 3.29. Kvantilplott for Farstad. Observasjonene vises på x-aksen og modellverdiene langs y-aksen. Qsim-plottet viser resultatene for modellsimuleringer med observert vær som inndata, Qprog viser den uprosesserte prognosen og Qmed viser medianen i utfallene fra usikkerhetsmodellen. Resultatene for prognosedag 1, 3 og 9 er vist.



Figur 3.30. Detalj fra hydrogrammet for Farstad. Observert og simulert vannføring er vist, sammen med den uprosesserte prognosen og medianen i utfallene fra usikkerhetsmodellen for første og niende prognosedag.

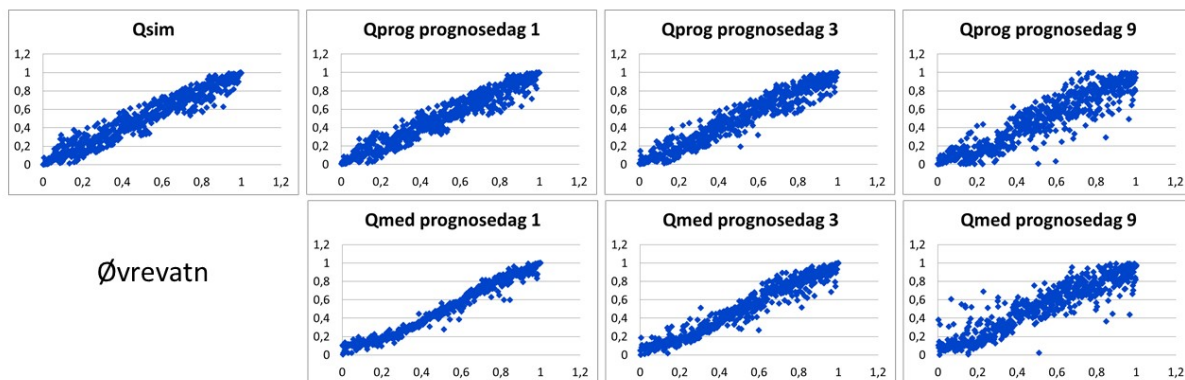


Figur 3.31. Korrelasjonskoeffisienten og Nash-Sutcliffe-kriteriet for den den uprosesserte prognosen og medianen fra usikkerhetsmodellen for Farstad, basert på kvantilene.

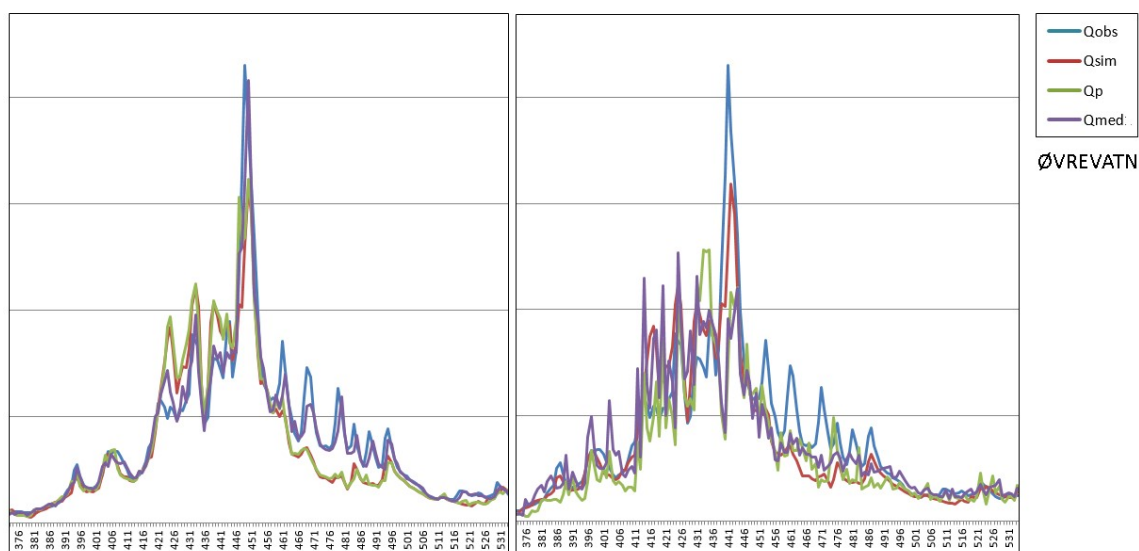
3.4.2.3 Øvrevatn

Øvrevatn er et eksempel på en modell som gir gode prognoser for hele prognoseperioden, og usikkerhetsmodellen kvantifiserer feilen og korrigerer effektivt for den i flere dager.

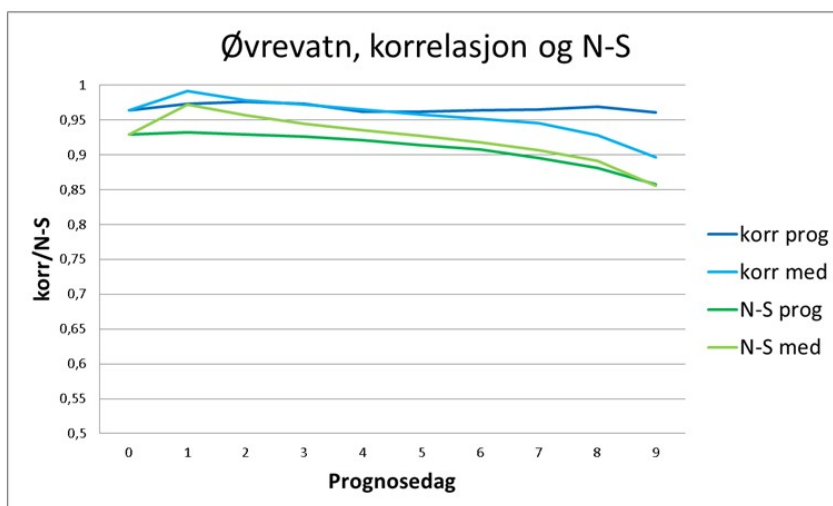
Figur 3.32 viser hvordan medianen samler punktene i diagrammet langs diagonalen, og figur 3.34 viser at Nash-Sutcliffe-koeffisienten ligger høyt hele prognoseperioden, og høyest for medianen. Figur 3.33 viser hvordan den feilkorrigerte medianen følger observasjonene godt den første prognosedagen, og hvordan både prognosen og medianen opptrer mer tilfeldig og med dårligere treffsikkerhet på alle vannføringsnivåer på prognosedag 9, på grunn av både dårligere værvarsel og at prognosen er utenfor rekkevidden til den autoregressive modellfeilkorrigeringen.



Figur 3.32. Kvantilplott for Øvrevatn. Observasjonene vises på x-aksen og modellverdiene langs y-aksen. Qsim-plottet viser resultatene for modellsimuleringer med observert vær som inndata, Qprog viser den uprosesserte prognosen og Qmed viser medianen i utfallene fra usikkerhetsmodellen. Resultatene for prognosedag 1, 3 og 9 er vist.



Figur 3.33. Detalj fra hydrogrammet for Øvrevatn. Observert og simulert vannføring er vist, sammen med den uprosesserte prognosen og medianen i utfallene fra usikkerhetsmodellen for første og niende prognosedag.



Figur 3.34. Korrelasjonskoeffisienten og Nash-Sutcliffe-kriteriet for den den uprosesserte prognosen og medianen fra usikkerhetsmodellen for Øvrevatn, basert på kvantilene.

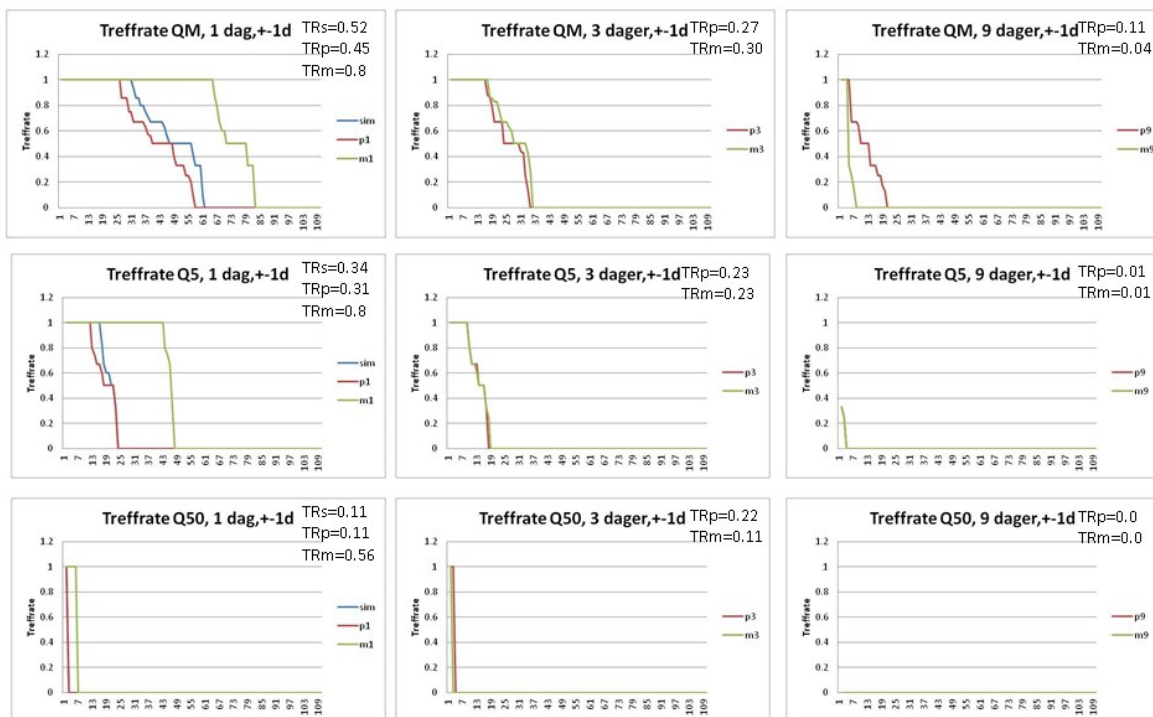
3.4.3 Treffrate og falske alarmer

Som for nedbør, er en analyse av treffrate og falsk alarmrate gjennomført for vannføringsprognosene. Mens nedbøren ble analysert med hensyn til en diskretisert skala over alle nedbørnivåer, er treff og falske alarmer for vannføringen vurdert i forhold til varslingsnivåene for flom, dvs. middelflom (Q_m), 5-årsflom (Q_5) og 50-årsflom (Q_{50}). Med en evalueringsperiode på to år, er imidlertid antall flomhendelser begrenset, og det kan være vanskelig å dra konklusjoner ut fra et såpass begrenset datamateriale. Alle flomhendelser i alle felt i løpet av evalueringsperioden er analysert. Modellsimuleringene definerer flomverdiene som prognosene klassifiseres etter. Disse ligger litt lavere enn flomverdiene som er definert fra observerte data.

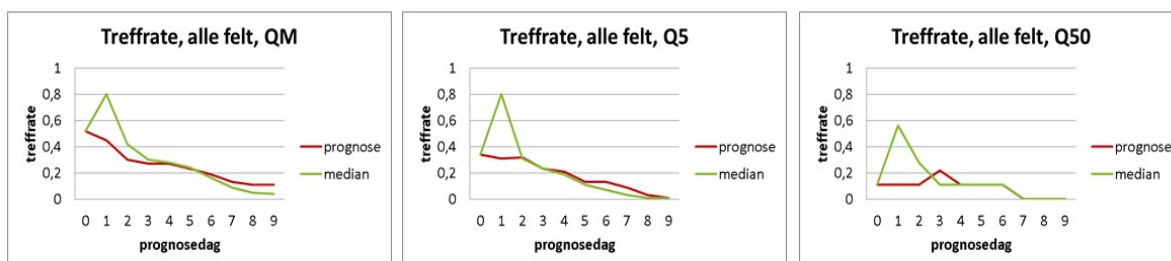
Figur 3.35 framstiller resultatet for treffraten. Treffraten for feltene er sortert, og kurvene viser fordelingen av treffraten over feltene. Vi ser at medianen har 100% treff i 66 av feltene, mens 27 felt har null treff, når det gjelder middelflom første prognosedag. Dette er betydelig bedre enn for den uprosesserte prognosen, som har 100% treff i 25 av feltene og null treff i 52 av feltene. Simuleringen som er kjørt med observerte inndata er litt bedre enn prognosen, med 100% treff i 31 av feltene og null treff i 48 av feltene. Feltene med null treff inkluderer de feltene der det ikke er observert middelflom i evalueringsperioden (18 felt). Tallene oppe til høyre i hvert diagram viser gjennomsnittlig treffrate for medianen og prognosen (og simuleringen i diagrammet for første prognosedag).

Treffraten avtar med prognosedagen, og fra prognosedag 3 har prognosen en bedre treffrate enn medianen, men begge er svært lave for prognoselengde over et par dager. Treffraten for 5-årsflom viser samme mønster som for middelflom, men her er antall hendelser mindre, og mange felt (53) har ingen flomhendelser. Få flomhendelser gir et uhyre tynt analysegrunnlag når det gjelder 50-årsflom. Det er bare observert 50-årsflom i 9 av feltene i løpet av evalueringsperioden. Figur 3.36 viser den gjennomsnittlige treffratens variasjon med prognosedagen, og vi ser tydelig den avtakende trenden, som figur 3.35 viste for dag 1, 3 og 9.

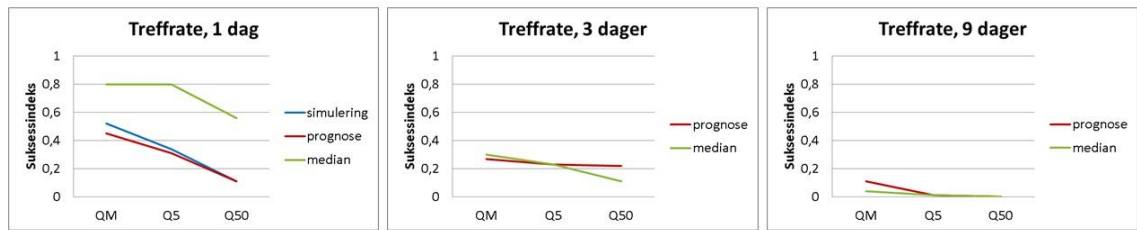
Figur 3.37 viser treffratens variasjon med flomstørrelse for 1, 3 og 9 dagers prognoselengde. Medianen gir brukbare resultater for alle de analyserte flomstørrelser første prognosedag, med forbehold om tynt datagrunnlag. Prognosen og simuleringen gir svært dårlige resultater for flomstørrelser over middelflom. For flere dagers prognoselengde er resultatene svært dårlige for både den prosesserte og den uprosesserte prognosen.



Figur 3.35. Treffrater for medianen i utfallene fra usikkerhetsmodellen (grønn linje) og den uprosesserte prognosen (rød linje) for første, tredje og niende prognosedag. For første prognosedag er også treffraten for simuleringen med observerte inndata vist (blå linje). Treffraten (y-aksen) vises sortert og fordelt over alle prognosefelt (x-aksen). Oppe til høyre i hvert diagram er gjennomsnittlig treffrate for median (TRm), prognose (TRp) og simulering (TRs) vist.

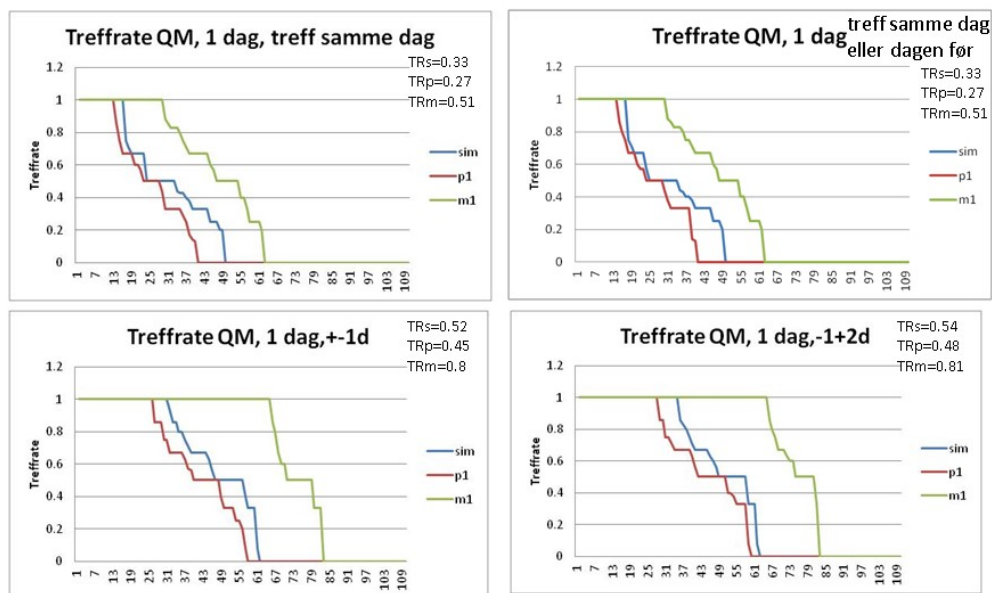


Figur 3.36. Variasjonen i gjennomsnittlig treffrate med prognosedag mht. middelflom, 5-årsflom og 50-årsflom. Grønn linje viser medianen og rød linje viser den uprosesserte prognosen. Verdien for dag 0 er treffraten for simuleringen med observerte inndata.



Figur 3.37. Variasjonen i gjennomsnittlig treffrate med flomstørrelsen for første, tredje og niende prognosedag. Grønn linje viser medianen og rød linje viser den uprosesserte prognosen, og blå linje simuleringen med observerte inndata.

Effekten av å legge inn et slingringsmonn i tiden for sammenfall mellom observasjon og prognose ble undersøkt. HBV-modellen har en klar tendens til å ligge en dag på etterskudd med vannføringsresponsen, dels på grunn av at værdøgnet er definert slik at det ligger nesten ett døgn foran vannføringsdøgnet (7 timers overlapp), og dels fordi HBV-modellens struktur og numeriske løsningsalgoritmer bidrar til forsinket respons. En simulert flom har likevel stor informasjonsverdi, selv om den plasseres en dag feil i tid. For å gi et riktig bilde av nytteverdien til modellen, vil det derfor være en pragmatisk løsning å gi treffberegningen en dags slingringsmonn.



Figur 3.38. Effekten av slingringsmonn i tid ved treffrateberegninger. Treffraten mht. middelflom for prognoselengde 1 dag og for simuleringen med observerte inndata er vist. Diagrammene viser treffraten når treff defineres som sammenfall samme dag, samme dag eller dagen før, samme dag eller dagen før eller dagen etter og samme dag eller dagen før, dagen etter eller to dager etter. Gjennomsnittlig treffrate er vist oppe til høyre i hvert diagram.

Figur 3.38 viser resultatene for beregninger av treffraten gjort med ulike definisjoner av hva som gjelder som treff, eksemplifisert med resultatene for middelflom og en dags prognoselengde. Vi ser at å inkludere sammenfall dagen før flommen ikke har noen effekt på treffraten. Det er fordi modellen så å si aldri simulerer respons på forskudd. Det har stor effekt å ta med dagen etter en observert flom, av grunnene gitt over, men det gir liten

tilleggsgevinst å legge til ytterligere en dag i etterkant til slingringsmonnet. I analysene som er presentert over og i de videre analysene er det funnet hensiktsmessig å definere sammenfall med en dags slingringsmonn.

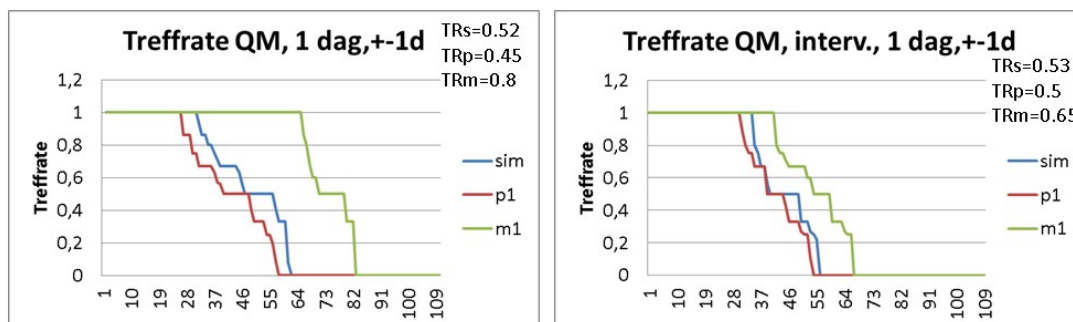
Treffberegninger gjøres oftest ved å definere alt over terskelverdien som treff. Man kan alternativt tenke seg at man definerer treff ut ifra et intervall rundt en verdi. Det er sett på forskjellen når vi definerer treff over en terskelverdi og i et intervall rundt verdien. Figur 3.39 viser resultatene mht. middelflom. Diagrammet til venstre er det samme som vises i figur 3.35 og figur 3.38, og definerer treff som

$$Q_m \leq Q_{obs} \cap Q_m \leq Q_{prog/med(sim)},$$

og i diagrammet til høyre er treff definert som

$$Q_m \leq Q_{obs} \leq Q_5 \cap 0.9 * Q_m \leq Q_{prog/med(sim)} \leq Q_5.$$

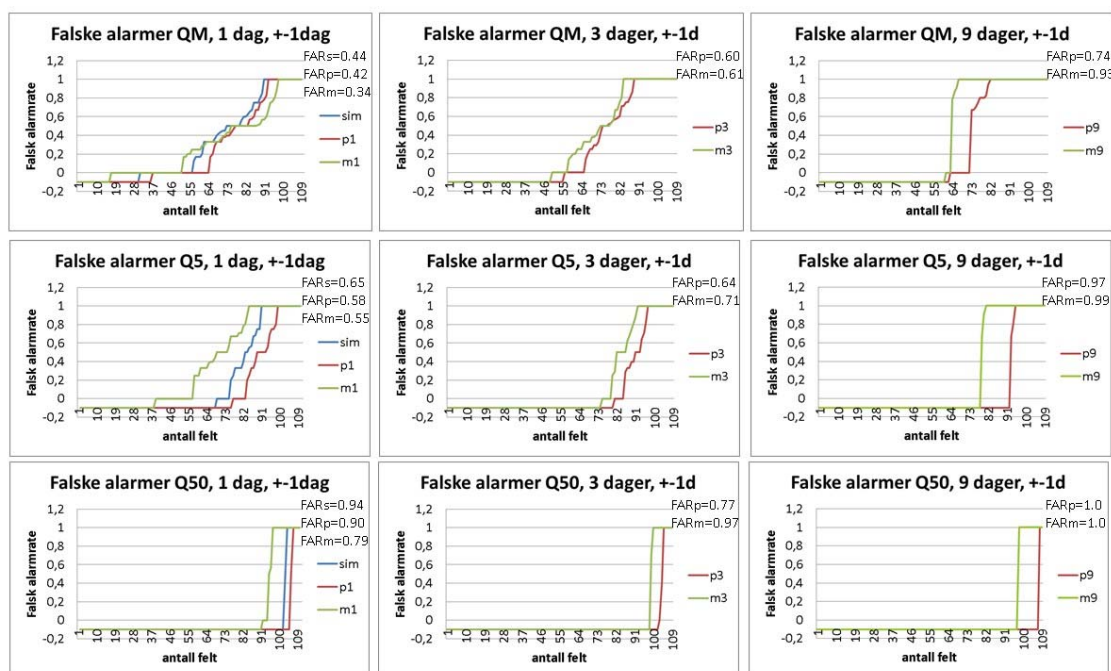
Vi ser en liten økning i treffraten for simuleringen og prognosen ved intervallberegning, i diagrammet til høyre. Dette har å gjøre med at intervallet tar med verdier ned til $0.9 * Q_m$, noe som bidrar med flere treff enn intervallgrensen i overkant utelukker. Medianen gjør det betydelig dårligere med intervallgrenser. Det er fordi medianprognosene ligger høyere enn modellsimuleringene, som definerer modellflomverdiene, og intervallgrensen avskjærer mange av treffene. For øvrig er antall flomhendelser mindre ved intervalldefinisjon, og analyse materialet blir mer begrenset og sårbart for tilfeldig variasjon. For øvrige analyser av treffrater og falske alarmer er det funnet hensiktsmessig å benytte den vanligere definisjonen basert på terskelverdier.



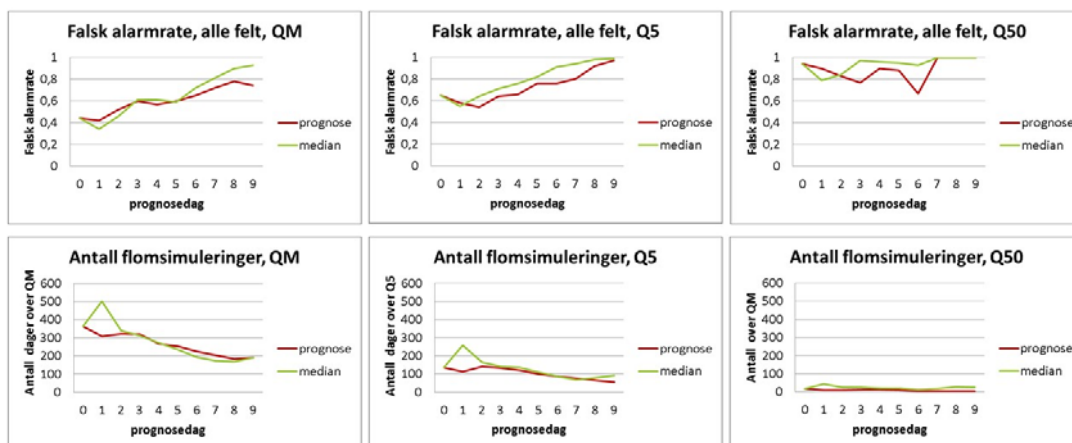
Figur 3.39. Treffrater når treff er definert i forhold til en terskelverdi (diagrammet til venstre) og i forhold til et intervall rundt verdien som skal sammenliknes (til høyre). Resultater for middelflommen første prognosedag er vist. Gjennomsnittlig treffrate er vist oppe til høyre i hvert diagram.

Treffrate må alltid ses i sammenheng med falsk alarmrate. Høy treffrate er bra, men hvis modellen samtidig genererer alt for mange flommer og har en høy falsk alarmrate, er presisjonen og nytten begrenset. Figur 3.40 viser falsk alarmraten mht. middel-, 5-års- og 50-årsflommen for første, tredje og niende prognosedag. Modelfelt som det ikke er simulert det aktuelle flomnivået for er representert med verdien -0.1. Medianen gir det beste resultatet for korttidsprognoser, med en gjennomsnittlig falsk alarmrate på 0.34 for de feltene der det er simulerte flomhendelser, mot 0.42 for den uprosesserte prognosen og 0.44 for simuleringen. Medianen simulerer også flest flomhendelser, f.eks simuleres det middelflom i 93 av feltene, mens prognosen og simuleringen gir middelflomhendelser i hhv. 73 og 77 av feltene. Antallet simulerte flommer reduseres med prognoselengden.

Medianen gir dårligere skår for lenger prognosetid, i likhet med hva som er tilfelle med treffraten. Igjen er datagrunnlaget svært tynt for store flommer, med få simulerte 50-årsflommer, se figur 3.41. 9-dagersprognosen for 50-årsflom gir hendelser i hhv. 12 (medianen) og 2 (prognosen) felt, og alle bommer.



Figur 3.40. Falsk alarmrate for medianen i utfallene fra usikkerhetsmodellen (grønn linje) og den uprosesserte prognosen (rød linje) for første, tredje og niende prognosedag. For første prognosedag er også falsk alarmraten for simuleringen med observerte inndata vist (blå linje). Falsk alarmraten (y-aksen) vises sortert og fordelt over alle prognosefelt (x-aksen). For felt der flomhendelser ikke er simulert, er falsk alarmraten satt til -0.1. Oppe til høyre i hvert diagram er gjennomsnittlig falsk alarmrate for median (FARm), prognose (FARp) og simulering (FARs) vist.

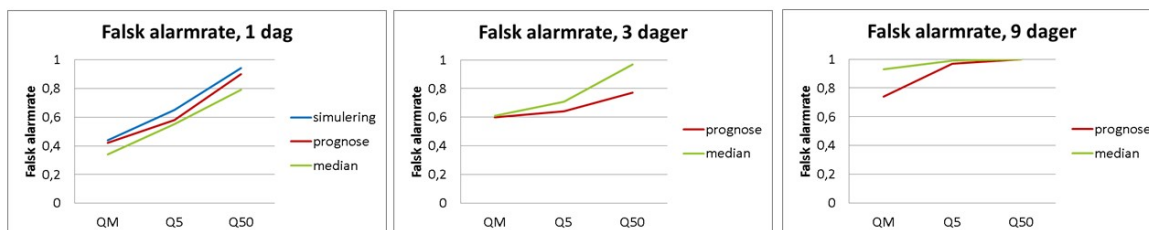


Figur 3.41. Variasjonen i gjennomsnittlig falsk alarmrate med prognosedag mht. middelflom, 5-årsflom og 50-årsflom. Grønn linje viser medianen og rød linje viser den uprosesserte prognosen. Verdien for dag 0 er falsk alarmraten for simuleringen med observerte inndata.

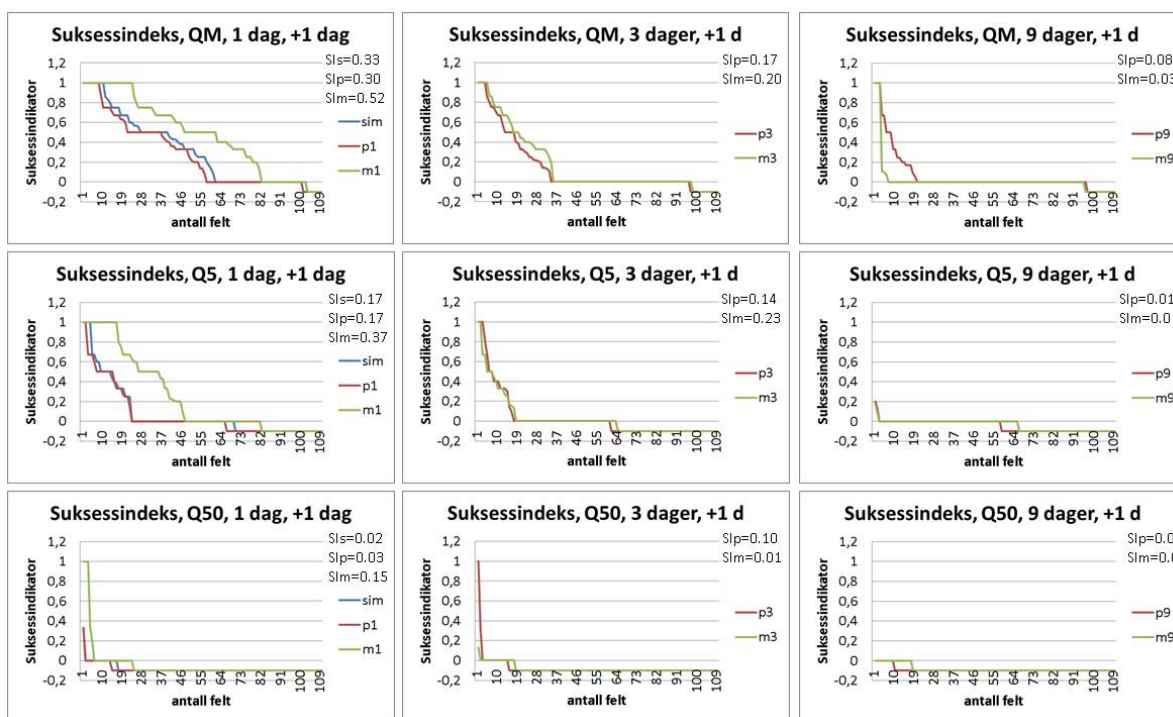
observerte inndata. Antall simulerte flomhendelser er presentert i nederste rad, for å vise beregningsgrunnlaget.

Falsk alarmraten øker med prognosedagen, og nærmer seg 1 utover i prognoseperioden. Sett i sammenheng med treffraten, bærer langtidvarsling av flomhendelser preg av å være skudd i blinde, og har liten nytteverdi. Varsling av mer moderate flomhendelser på en til to dagers basis, fungerer relativt godt.

Resultatene som funksjon av prognosedag er sammenfattet i figur 3.42. Falsk alarmraten degraderer raskt med flomstørrelse og varslingstid. En kuriositet er at simuleringen gir en høyere falsk alarmrate enn både den uprosesserte og prosesserte prognosen.



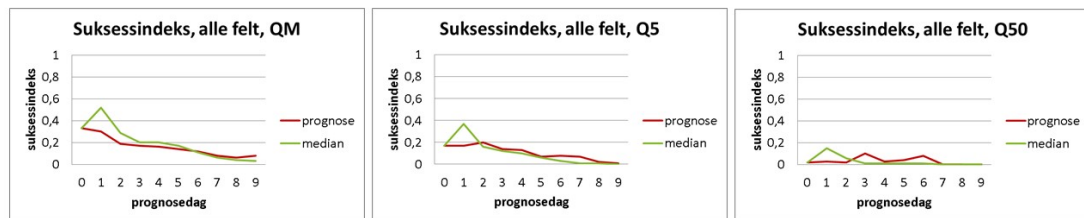
Figur 3.42. Variasjonen i gjennomsnittlig Falsk alarmrate med flomstørrelsen for første, tredje og niende prognosedag. Grønn linje viser medianen og rød linje viser den uprosesserte prognosen, og blå linje simuleringen med observerte inndata.



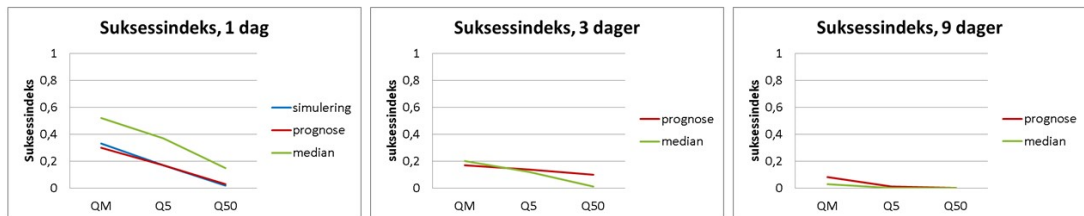
Figur 3.43. Suksessindeks for medianen i utfallene fra usikkerhetsmodellen (grønn linje) og den uprosesserte prognosen (rød linje) for første, tredje og niende prognosedag. For første prognosedag er også suksessindeksen for simuleringen med observerte inndata vist (blå linje). Suksessindeksen (y-aksen) vises sortert og fordelt over alle prognosefelt (x-aksen). For felt der flomhendelser ikke er simulert, er suksessindeksen satt til -0.1. Oppe

til høyre i hvert diagram er gjennomsnittlig suksessindeks for median (SI_m), prognose (SI_p) og sumulering (SI_s) vist.

Suksessindeksen er formulert slik at den tar hensyn til både modellens evne til å treffe flomhendelser og tendensen til å gi falske alarmer, og sammenfatter dermed resultatene ovenfor. Figur 3.43 viser suksessindeksen for samme flomverdier og prognosedager som over. Beste verdi er 1.0, som for treffraten, men verdiene er ellers lavere fordi antall treff angis i forhold til både antall treff, bom og falske alarmer. En suksessindeks over 0.5 indikerer at modellen oppnår flere treff enn bom og falske alarmer, og medianen runder så vidt denne milepælen for middelflom første prognosedag, betydelig bedre enn simuleringen og den uprosesserte prognosen, som har dobbelt så mange bom og falske alarmer som treff. Suksessindeksen viser ellers, nødvendigvis, de samme tendenser som treffraten og falsk alarmraten: den avtar raskt med prognosedagen og flomstørrelsen, og medianen gir best resultater ved kort prognoselengde. Suksessindeksen viser, for øvrig et bredere spekter av variasjon mellom 0.0 og 1.0 enn de bratte treffrate- og falsk alarmkurvene. Figurene 3.44 og 3.45 sammenfatter henholdsvis suksessindeksens variasjon med prognosedagen og flomstørrelsen.



Figur 3.44. Variasjonen i gjennomsnittlig suksessindeks med prognosedag mht. middelflom, 5-årsflom og 50-årsflom. Grønn linje viser medianen og rød linje viser den uprosesserte prognosen. Verdien for dag 0 er suksessindeksen for simuleringen med observerte inndata.



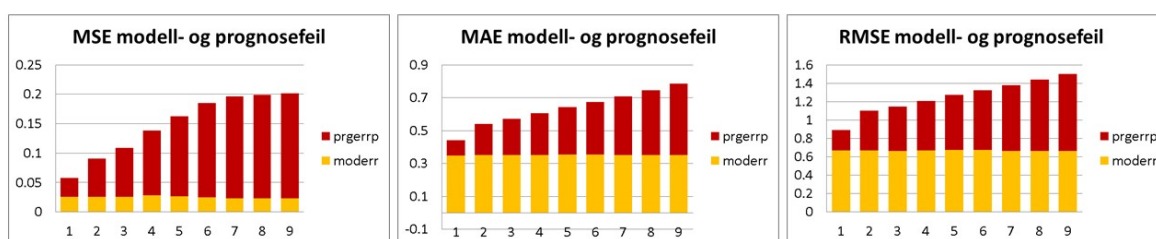
Figur 3.45. Variasjonen i gjennomsnittlig suksessindeks med flomstørrelsen for første, tredje og niende prognosedag. Grønn linje viser medianen, rød linje viser den uprosesserte prognosen, og blå linje viser simuleringen med observerte inndata.

3.4.4 Delfeilmodellenes bidrag

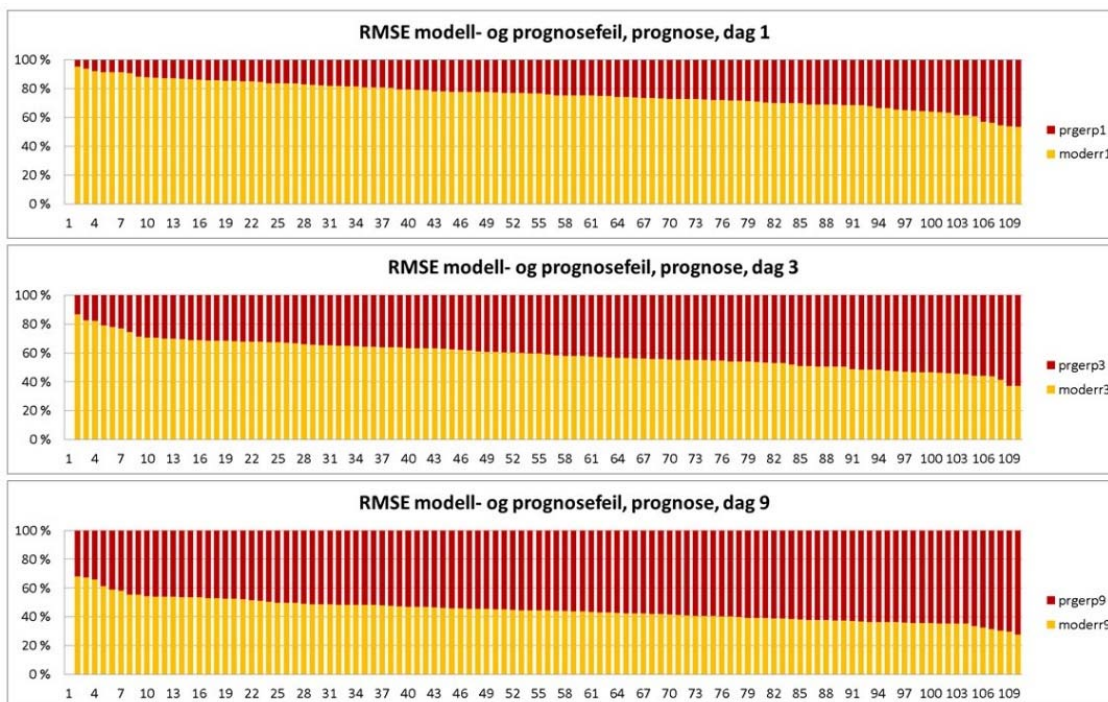
Som beskrevet i kapittel 2, er feilmodellen bygget opp slik at den korrigerer vannføringsprognosen for to typer feil, feilen i værprognosen og feilen som skyldes den hydrologiske modellen. Likningen i avsnitt 2.1 gir den totale feilen i vannføringsprognosen som summen av de to komponentene. Figur 3.46 viser søylediagrammer for modellfeilen, prognosefeilen og totalfeilen uttrykt som additivt avvik, gjennomsnittlig absoluttavvik og middelavvik, se avsnitt 3.2 for forklaring av feilestimatene. Diagrammene representerer den uprosesserte prognosen, som er utgangspunktet for korrigeringen. Feilkomponentene for de forskjellige modellfeltene er

beregnet relativt til feltets middelavløp i valideringsperioden, for at ikke resultatet skal domineres av felt med stor vannføring.

Modellfeilen ($Q_{OBS}(t) - Q_{SIM}(t)$) er uavhengig av prognosedag, og gir et konstant bidrag gjennom hele prognoseperioden (gule søyler). Feilbidraget fra værprognosen ($Q_{SIM}(t) - Q_{FOR}^{(j)}(t)$) er lite første prognosedag og øker med prognoselengden, en logisk konsekvens av at usikkerheten i værprognosen øker med prognosedagen, se kapittel 3.2 og 3.3. Modellfeilen bidrar lite til det additive avviket, samtidig som det utgjør en betydelig andel av absoluttavviket. Det viser at modellfeilen har varierende fortegn, og i stor grad nulles ut over valideringsperioden. Prognosefeilen har en tendens til å være positiv, dvs. at modellsimulert vannføring med værprognose som inndata ligger lavere enn når observert vær brukes som inndata. Dette henger sammen med at nedbørprognosen har en tendens til å underestimere nedbøren, og særlig de store nedbørtilfellene, se avsnitt 3.2.1 og 3.2.2.



Figur 3.46. Modellfeil (gule søyler), prognosefeil (røde søyler) og total feil (summen av gul og rød søyle) for den ukorrigerte prognosen, uttrykt ved additivt avvik (MSE), gjennomsnittlig absoluttavvik (MAE) og middelavvik (RMSE).

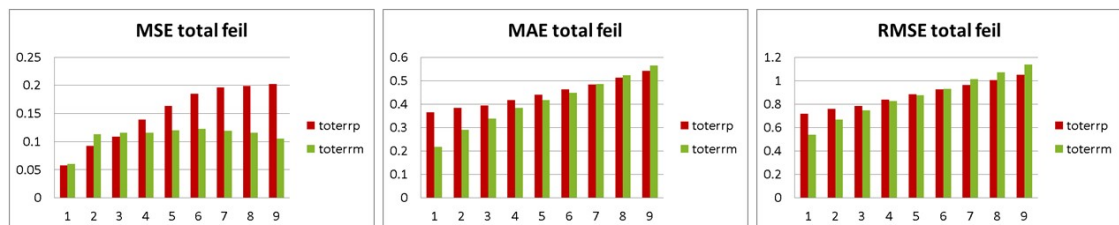


Figur 3.47. Andelen modell- (gul søyle) og prognosefeil (rød søyle) for alle felt og for prognosedag 1, 3 og 9, uttrykt ved middelavviket.

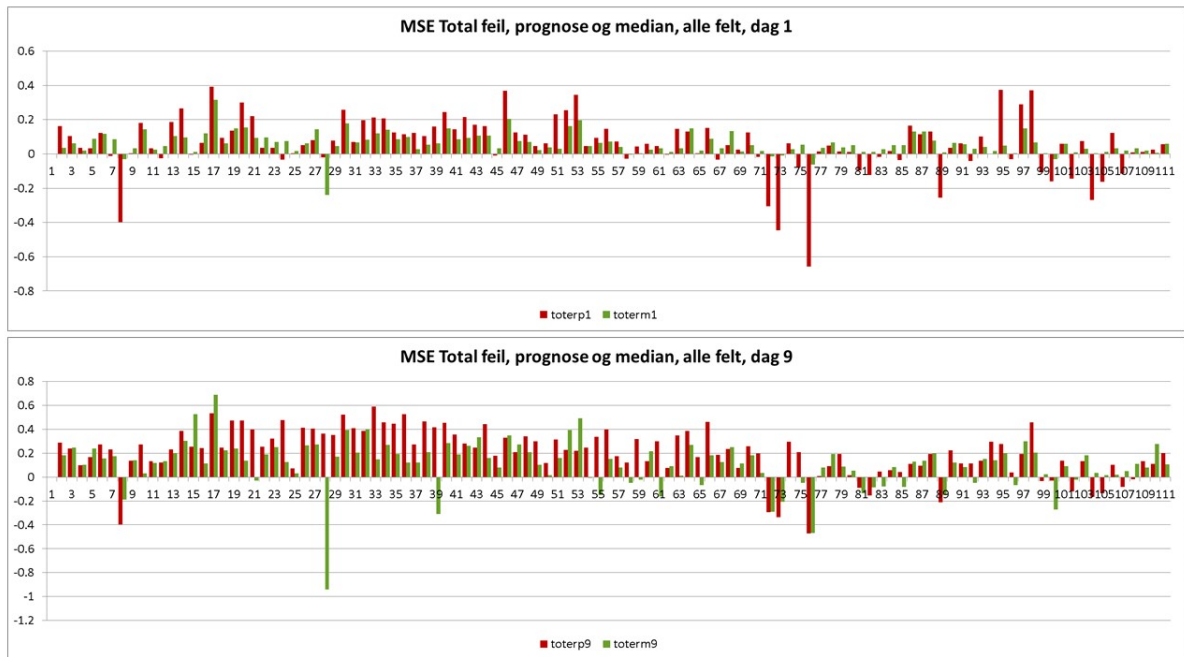
I figur 3.47 er andelen modell- og prognosefeil, uttrykt ved middelavviket, vist for alle felt for prognosedag 1, 3 og 9. For alle modellfelt gjelder det at modellfeilen er større enn prognosefeilen når det gjelder 1-dagsprognosen, og for de fleste feltene er prognosefeilen større enn modellfeilen for 9-dagersprognosen. Trenden mot økende prognosefeilandel med prognosedagen gjelder for alle felt.

I figur 3.48 er totalfeilen ($Q_{OBS}(t) - Q_{FOR}^{(j)}(t)$) før og etter korrigering vist for alle prognosedager. Når det gjelder det additive avviket er totalfeilen før korrigering (røde søyler) det samme som summen av modell- og prognosefeilen i figur 3.46. Det gjelder ikke det gjennomsnittlige absoluttavviket og middelavviket, som involverer transformasjoner som gjør at gjennomsnittet av addendene i likningen i kapittel 2 ikke er lik gjennomsnittet av summen. Det additive avviket viser at positive avvik dominerer, dvs. at observert vannføring i gjennomsnitt ligger over prognosert, både for den uprosesserte og prosesserte prognosen. Det additive avvik for medianen flater ut med prognoselengden. Det skyldes at et større innslag av modellfelt med negative feil bidrar til å nulle ut feiløkningen utover i prognoseperioden, se figur 3.49. Den uprosesserte prognosen har et større innslag av negative avvik de første prognosedagene (fig. 3.49), som nuller ut avvikene i noen grad og gir et additivt avvik sammenliknbart med medianens avvik, mens absolutt- og middelavviket er større disse første prognosedagene. Absolutt- og middelavviket til medianen øker relativt mer enn prognosen utover i prognoseperioden, og er på størrelse med den uprosesserte prognosen fra dag 5-6. Middelavviket øker relativt mer enn absoluttavviket, på grunn av noen store avvik som vektet mer av middelavviket. Figur 3.49 viser illustrerer for prognosedag 1 og 9 forekomsten av avvik som styrer mønsteret i de ulike målene av totalfeilen.

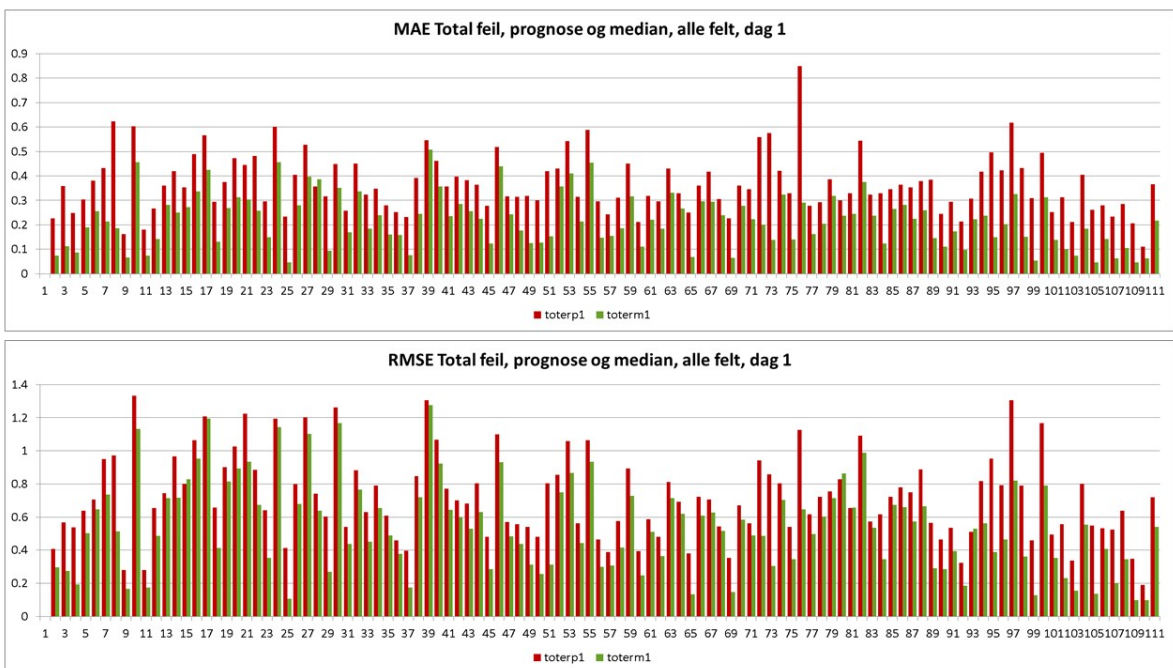
Figur 3.50 viser at tendensen til lavere absoluttavvik for medianen enn prognosen første prognosedag er gjennomgående for alle felt. Noen store avvik gjennom evalueringsperioden gjør at forspranget ikke er så stort når det gjelder middelavvik.



Figur 3.48. Totalfeilen for den uprosesserte prognosen (røde søyler) og medianen i utfallene fra usikkerhetsmodellen (grønne søyler) for alle prognosedager. Additivt avvik, absoluttavvik og middelavvik.



Figur 3.49. Totalfeilen for den uprosesserte prognosen (røde søyler) og medianen i utfallene fra usikkerhetsmodellen (grønne søyler) uttrykt ved det additive avviket vist for alle felt for prognosedag 1 og 9.



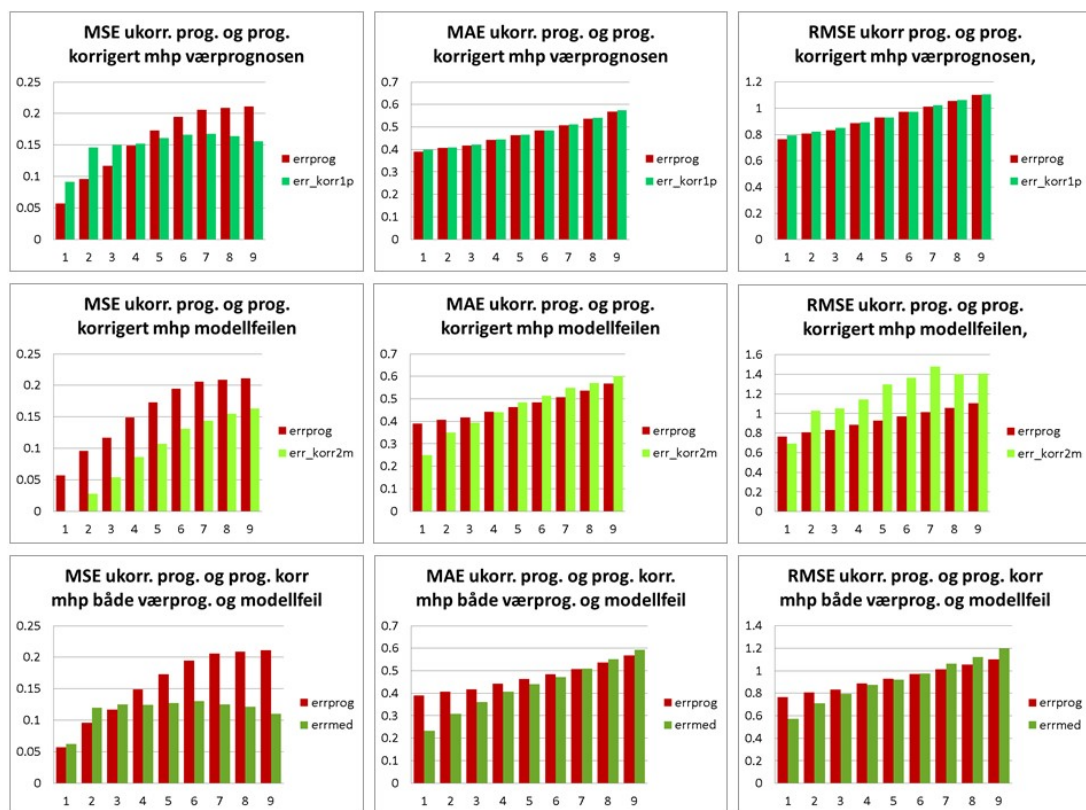
Figur 3.50. Totalfeilen for den uprosesserte prognosen (røde søyler) og medianen i utfallene fra usikkerhetsmodellen (grønne søyler) uttrykt ved absolutt- og middelavviket for alle felt for prognosedag 1.

Feilmodellen korrigerer prognosen i to omganger. Det genereres et sett statistiske utfall som skal representere været i prognoseperioden på en bedre måte enn den operasjonelle værprognosen, og HBV-modellen kjøres med disse utfallene som inndata. Medianen i

disse utfallene er korrigert med hensyn til feil i værprognosen. I tillegg prosesseres utfallene fra HBV-modellen gjennom den autoregressive modellen, som korrigerer for modellfeilen. Utfallene etter denne operasjonen er korrigert både med hensyn til feil i værprognosen og modellen. Dersom den uprosesserte prognosen bare prosesseres gjennom den autoregressive modellen, er den bare korrigert med hensyn til modellfeilen. Figur 3.51 viser totalfeilen for prognoser korrigert bare for feilen i værprognosen, bare for modellfeilen og for begge deler.

Øverste rad i figur 3.51 viser at korrigerings med hensyn til værprognosen ikke har vesentlig positiv effekt på prognosen. Absolutt- og middelavviket er i stor grad uendret, mens det additive avviket øker de første prognosedagene og avtar de siste. Grunnen til at korrigerings for feil i værprognosen ikke fungerer etter hensikten, henger sammen med at feilmodellens nedbørprognose skårer dårligere enn den uprosesserte nedbørprognosen. Feilmodellens temperaturprognose gir en forbedring i forhold til den uprosesserte, men det veier ikke opp for at nedbørrkorrigerings feiler.

Korrigerings med hensyn på modellfeilen, derimot, gir en bedre prognose de første prognosedagene. Den autoregressive modellen, som justerer for avvik mellom den siste observerte vannføringen og tilsvarende simulering, har en åpenbar positiv effekt de første prognosedagene. Effekten av den autoregressive modellen når ikke langt ut i prognoseperioden, og noen store avvik forårsaker høye middelavvik, men nulles i noen grad ut i det additive avviket, da de til dels er negative.



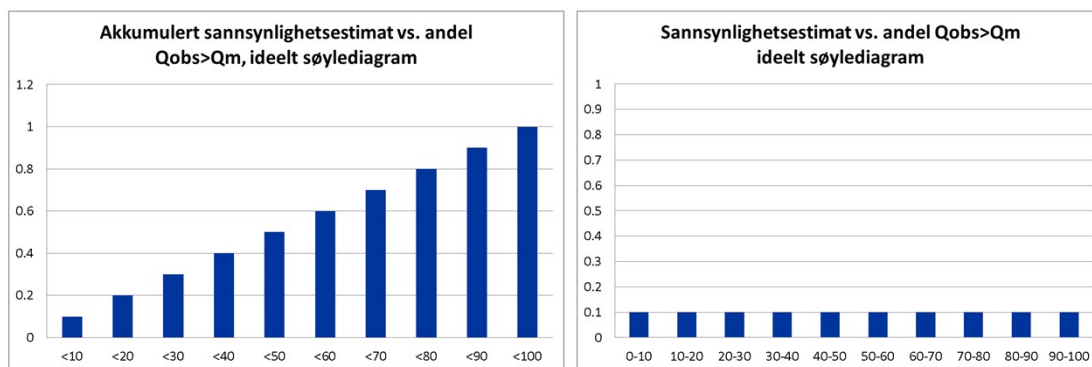
Figur 3.51. Totalfeilen (røde søyler) og totalfeilen korrigert med hensyn til værprognosen (grønne søyler øverst), modellfeilen (grønne søyler i midten) og begge deler (grønne søyler nederst).

Nederste rad i figur 3.51 viser den samlede effekten av korrigerende for feil i værprognosen og vannføringsmodellen. De to delkorrigeringene er ikke additive, så summen av dem er ikke en ren superponering av deffektene. Den positive effekten av modellfeilkorrigeringen dominerer over det negative utslaget av nedbørprosesseringen, og gir en prognose med mindre avvik de første 3 - 4 dagene.

3.4.5 Sannsynlighet for overskridelse av flomnivåer

Hver gang flomvarslingsmodellene og usikkerhetsmodellen kjøres, beregnes også sannsynligheten for overskridelse av flomnivåene middel-, 5-års og 50-årsflom. Disse sannsynlighetsestimatene, som presenteres samlet for alle felt, alle prognosedager og for de tre flomnivåene, gir flomvarsleren et første signal om at en flomsituasjon kan være under oppseiling. I dette avsnittet undersøkes det om disse sannsynlighetsestimatene gir et representativt bilde av antall påfølgende forekomster av flommer.

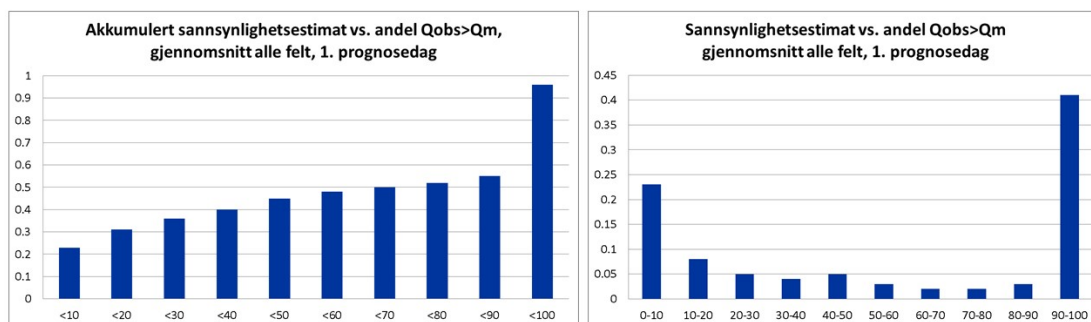
Sannsynlighetsestimatene fra 0 til 100 er sortert i 10 klasser, som analysen er basert på. Antall hendelser der vannføringen overstiger middel-, 5-års og 50-årsflom er telt opp for hvert felt, og hver hendelse er tilordnet sannsynlighetsklassen som 1-dagsprognosen sorterer i. På grunn av vannføringsmodellens tendens til å ligge en dag etter observasjonen, slik det er forklart i avsnitt 3.4.3, er det her den observerte vannføringen dagen før prognosedagen som er brukt. Figur 3.52 viser det ideelle resultatet når det gjelder forekomst av flommer, eksemplifisert ved middelflommen, ved gitte sannsynlighetsestimater. Diagrammet til venstre, som viser akkumulerte sannsynlighetsestimater, tar utgangspunkt i at alle tilfeller av varslet sannsynlighet opp til f.eks. 40%, ideelt sett skal fange opp 40% av flomtilfellene, varslene opp til 90%, skal fange opp 90% av flomtilfellene osv. Andelen flommer som skal fanges opp innenfor hver av de 10 intervallklassene blir tilsvarende 10%, som vist i diagrammet til høyre.



Figur 3.52. Ideell forekomst av flommer innenfor sannsynlighetsklassene. Akkumulert til venstre og fordelt over hver klasse til høyre.

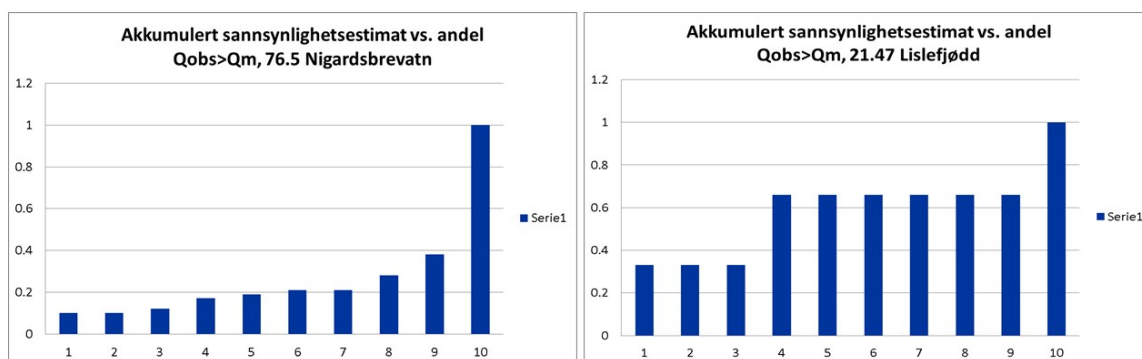
Virkeligheten fordeler seg litt annerledes, og figur 3.53 viser resultatene for middelflommen i form av gjennomsnittet over alle felt. Sannsynligheten summerer her ikke opp til 1.0 på grunn av innslaget av felt som ikke har hendelser. Vi ser en overrepresentasjon i den laveste og i den høyeste klassen. Som vi har sett tidligere, er det relativt mange tilfeller av undervarsla flommer. Mange av disse er representert ved en lav sannsynlighet for flom, og bidrar i den laveste klassen. Den høyeste klassen teller med alle flomhendelser av flere dagers varighet. Hvis det er flom, og modellen, oppdatert med siste flomobservasjon, opprettholder flomvannføringen neste dag, vil den dagens

sannsynlighet ligge på 100%. Dette bidrar til overrepresentativitet i denne klassen. For øvrig er flommene relativt jevnt fordelt over sannsynlighetsklassene.



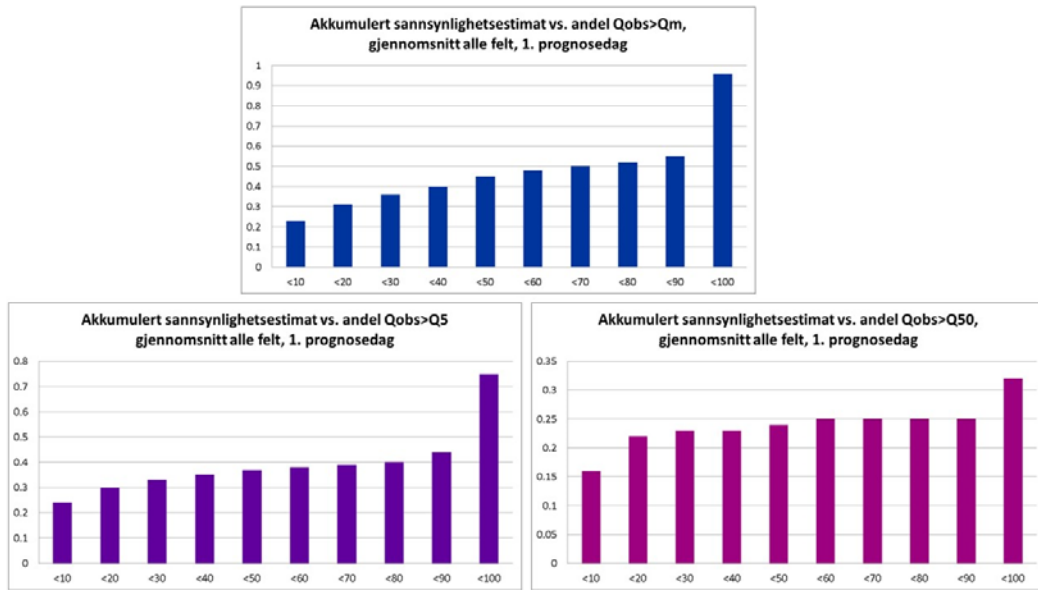
Figur 3.53. Forekomst av middelflommer innenfor sannsynlighetsklassene. Gjennomsnitt over alle felt. Akkumulert til venstre og fordelt over hver klasse til høyre.

Figur 3.54 gir to eksempler på resultater for enkeltfelt. Tendensen til overrepresentativitet i den øverste klassen er tydelig i eksempelet Nigardsbreen. Lislefjodd er et eksempel på et felt med få hendelser. Her er det bare observert 3 tilfeller av vannføringer over middelflom, og de fordeler seg relativt spredt over sannsynlighetsklassene.



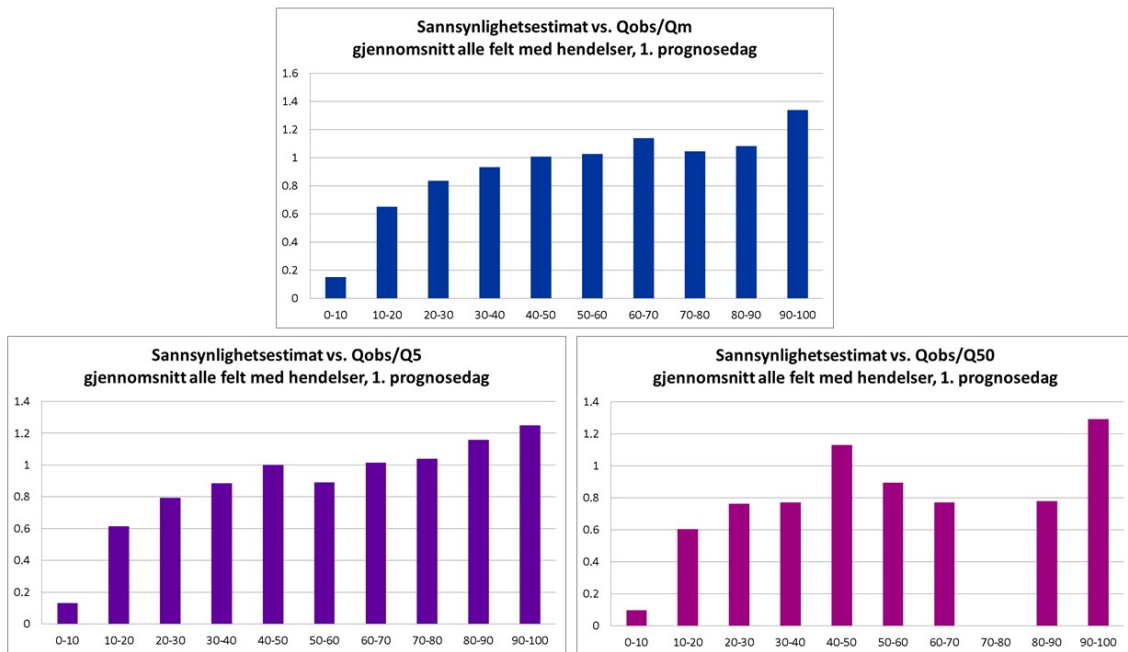
Figur 3.54. Akkumulert forekomst av middelflommer innenfor sannsynlighetsklassene for to enkeltfelt, Nigardsbrevatn og Lislefjodd.

Figur 3.55 viser den skkumulerte andelen flommer over sannsynlighetsklassene for 5- og 50-årsflom i tillegg til middelflom. Antall ganger det er varslet om sannsynlighet for flom over 10% er i gjennomsnitt 10.8 per felt for middelflom, 5.9 for 5-årsflom og 1.3 for 50-årsflom. Det statistiske materialet å drive analyse på er dermed svært begrenset for større flommer. Tendensen til å ha for mange hendelser i den nederste og øverste risikoklassen er likevel tydelig også for de høyeste flomklassene. Søylediagrammet flater ut noe, slik at det blir mindre bidrag fra de midterste sannsynlighetsklassene med flomstørrelsen.



Figur 3.55. Akkumulert forekomst av middel- 5- og 50-årsflommer innenfor sannsynlighetsklassene. Gjennomsnitt over alle felt.

Figur 3.56 viser tilsvarende figurer med flomstørrelse, uttrykt ved forholdet mellom vannføringen og den aktuelle flomstørrelsen, på y-aksen. Flomstørrelsen har en logisk og god fordeling, og stiger med sannsynlighetsestimatet. De midlere sannsynlighetsklassene har et forholdstall rundt 1, som betyr at det opptrer like mange vannføringer over som under middelflom når disse sannsynlighetene er meldt. Diagrammet for 50-årsflom bærer preg av at det er få hendelser å gjøre statistikk på.

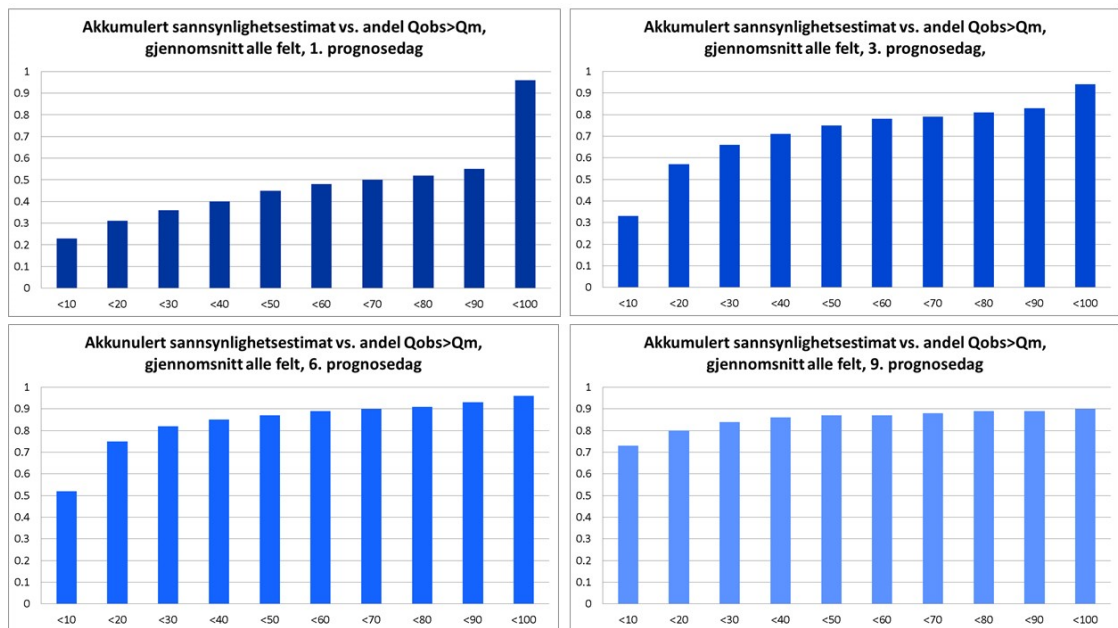


Figur 3.56. Flomstørrelse, gitt som forholdet til flomnivået, for flomhendelsene tilordnet de ulike sannsynlighetsklassene. Resultater for middel-, 5- og 50-årsflommen.

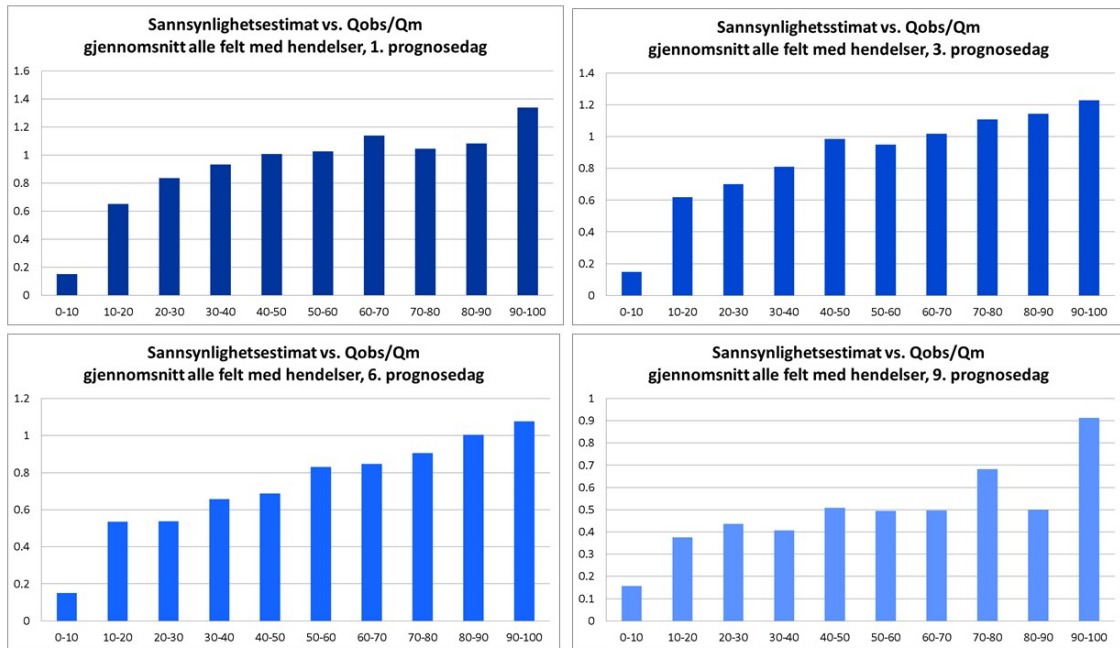
Middelflom varsles oftere i langtidsprognosene, fordi usikkerheten er større, utfallene får større spredning og flere av utfallene overstiger middelflom. Antall ganger (gjennomsnitt per felt) det er varslet om sannsynlighet for middelflom over 10% er 10.8 i 1-dagsprognosen, 21.4 i 3-dagersprognosen, 24.1 i 6-dagersprognosen og 18.8 i 9-dagersprognosen. Spredningen i utfallene veier i noen grad opp for at fleredagersprognosene har mindre nedbørmengder enn korttidsprognosen, men for 9. prognosedag ser vi at effekten av mindre nedbørmengder gjør seg mer gjeldende.

Figur 3.57 viser hvordan forekomsten av flommer i forhold til sannsynlighetsestimatene endrer seg med prognosedagen. Den varslete sannsynligheten for flom i fleredagersprognosene er som regel lav, og det gir seg utslag i at de to laveste sannsynlighetsgruppene bidrar med brorparten av flomhendelsene. Søylediagrammene flater ut ved at bidragene fra de høyere sannsynlighetsklassene er minimale. Tendensen øker med prognosedagen. Effekten av modelloppdatering med siste observerte vannføringsverdi som gjorde at høyeste sannsynlighetsklasse var overrepresentert med flommer i 1-dagsprognosen, er ikke så utpreget ved lengere prognosetider, og effekten forsvinner for de høyeste prognoseklassene.

I figur 3.58 er gjennomsnittlig flomstørrelsen i forhold til middelflom som er observert ved varsel i de ulike sannsynlighetsklassene vist for ulike prognoselengder. Flomstørrelsen øker med sannsynlighetsklassen også for prognoselengde på 3, 6 og 9 dager, men den ligger lavere enn for 1-dagsprognosen. Dette er et utslag av at flomprognoser har lavere presisjon i langtidsprognosen.



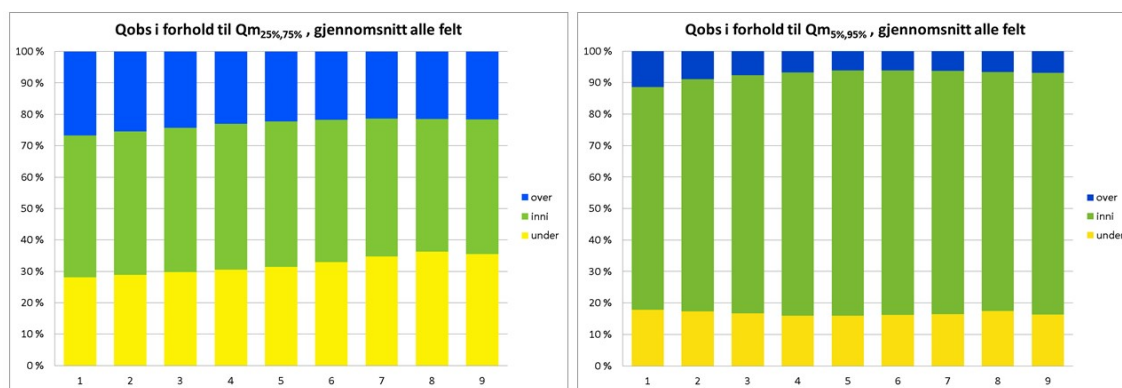
Figur 3.57. Akkumulert forekomst av middelflommer innenfor sannsynlighetsklassene forprognosedag 1, 3, 6 og 9. Gjennomsnitt over alle felt.



Figur 3.58. Flomstørrelse, gitt som forholdet til middelflom tilordnet de ulike sannsynlighetsklassene. Resultater for 1-, 3-, 6- og 9-dagersprognosen.

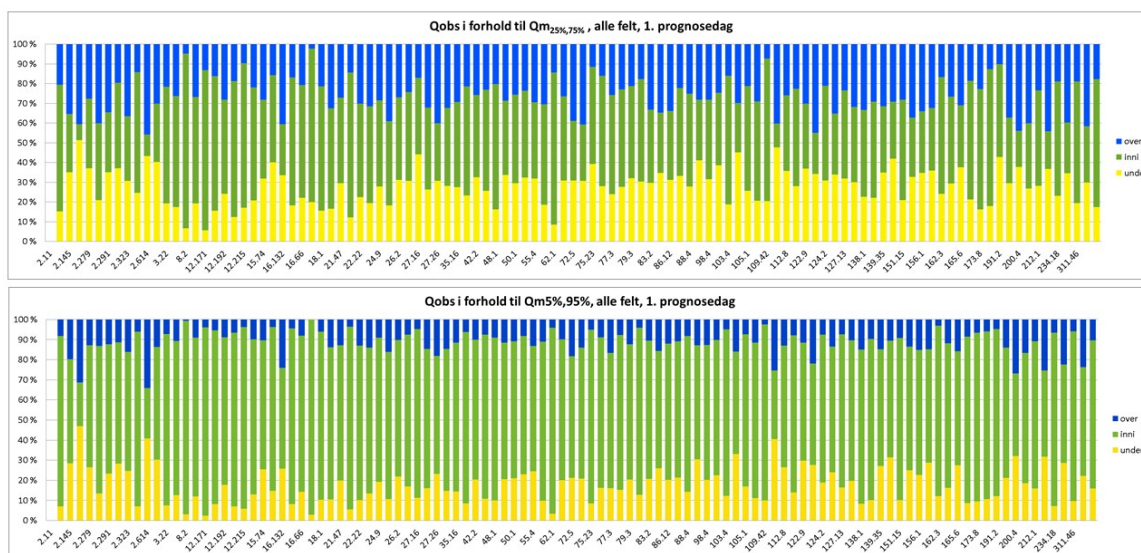
3.4.6 Sannsynlighetsfordeling for vannføring

Som for nedbør og temperatur, er det undersøkt hvordan observert vannføring fordeler seg i forhold til usikkerhetsmodellens sannsynlighetsfordeling. Figur 3.59 viser fordelingen i forhold til de 50% og 90% midterste utfallene. Både 50%- og 90%-intervallene fanger opp for få av de observerte vannføringene. Ca. 45% av de observerte vannføringene faller innenfor 50%-intervallet. 75-persentilen er godt samstemt med observasjonene første prognosedag, men ligger litt høyt utover i prognoseperioden. Flere enn 25% av de observerte vannføringene faller under 25-persentilen, og antallet øker utover i prognoseperioden. 90%-intervallet fanger opp relativt færre av de observerte vannføringene enn 50%-intervallet, med rundt 70% av observasjonene innenfor. 12% faller over 95-persentilen første prognosedag og 17% faller under 5-persentilen. Siden samtidig medianen i utfallene fra usikkerhetsmodellen har en tendens til å ligge for lavt i forhold til observasjonen, se avsnitt 3.4.4, tyder det på at avvikene i overkant av usikkerhetens fordeling, som jo er færre enn avvikene i underkant, er desto større enn avvikene i underkant. Fordelingen av vannføringsobservasjonene i forhold til utfallene fra usikkerhetsmodellen i endrer seg ikke vesentlig utover i prognoseperioden.



Figur 3.59. Fordelingen av observert vannføring i forhold til utfallene fra usikkerhetsmodellen som funksjon av prognosedag, gjennomsnitt over alle felt. Diagrammet til venstre viser observasjonene i forhold til 25- og 75-persentilen og diagrammet til høyre i forhold til 5- og 95-persentilen. Gule søyler viser andelen observasjoner under hhv. 25- og 5-persentilen, blå søyler viser andelen observasjoner over hhv. 75- og 95-persentilen og grønne søyler viser andelen innenfor respektive persentiler.

Spredningen over modellfeltene er betydelig, og noen få felt fanger opp svært få observasjoner innenfor de markerte persentilene, se figur 3.60. De fleste feltene har imidlertid en rimelig fordeling av observasjoner i forhold til persentilene fra usikkerhetsmodellen. Spredningen over feltene er større når det gjelder 50%-intervallet enn for 90%-intervallet, trolig et utslag av at utfallene fra usikkerhetsmodellen ligger tett i midtsjiktet, og at plasseringen av observasjonene i forhold til 25/75-persentilene er mer sårbar for tilfeldigheter enn plasseringen i forhold til de mer perifere 5/95-persentilene.



Figur 3.60. Fordelingen av observert vannføring i forhold til utfallene fra usikkerhetsmodellen første prognosedag, vist for alle felt. Øverste diagram viser fordelingen i forhold til 25- og 75-persentilen og nederste diagram i forhold til 5- og 95-persentilen. Se for øvrig forklaringen under figur 3.59.

4 Konklusjoner og anbefalinger

4.1 Oppsummering nedbør

Usikkerhetsmodellens "korrigerende" av nedbørprognosen har en tendens til å feile, og gjøre den dårligere enn den ukorrigerede. Nedbørprognosene fra medianen i utfallene fra usikkerhetsmodellen ligger gjennomgående lavere enn de uprosesserte prognosene, og forskjellen øker med gjennomsnittenedbøren i feltet, og med prognoselengden over tre dager. Begge prognosene ligger lavere enn observasjonene. Både målt i gjennomsnittlig absolutt avvik, additivt avvik, middelavvik og Nash-Sutcliffe-verdi skårer medianen dårligere enn den uprosesserte prognosen de første prognosedagene. I absolutt forstand har medianen et mindre avvik fra observasjonene enn prognosen de siste prognosedagene. Men da er begge prognosene relativt dårlige, og nærmer seg gjennomsnittsobservasjonen i treffsikkerhet. Usikkerhetsmodellen reduserer forskjellene i prediksjonsevne fra felt til felt og mellom værvarslingsmodellene, men prognosene blir i det store og hele ikke bedre.

Både medianen og prognosen har en tendens til å varsle småregn på bekostning av både antall dager med oppholdsvær og dager med mye regn, og tendensen er tydeligst for medianen. Tendensen forsterkes med prognoselengden. Andelen varsla oppholdsdager er for prognosen omtrent konstant utover i prognoseperioden, men færre av dem treffer realiserte oppholdsdager. Medianen ligger sjelden på oppholdsvær utover i prognoseperioden, siden ofte over 50% av utfallene vil gi litt regn som en følge av økt usikkerhet lengt fram i tid.

Usikkerhetsmodellens 50% konfidensintervall treffer jamt over godt 50% av de observerte nedbørhendelsene. Men konfidensintervallet ligger litt skjevt. Færre enn 25% av nedbørdagene faller under 25-persentilen. 75-persentilen avgrenser bedre de rundt 25% av nedbørobservasjonene som faller over konfidensintervallet.

Både treffraten og falsk alarmraten viser svært dårlige resultater for de høyere terskelverdiene, men resultatene må tolkes med forbehold om at datamaterialet inneholder få større nedbørhendelser. Det gjelder i størst grad for medianen.

4.2 Oppsummering temperatur

Feilmodellen for temperaturprognosen fungerer etter hensikten, slik at når det korrigeres for den estimerte feilen, så får vi en forbedret prognose. Medianen i utfallene fra usikkerhetsmodellen har et betydelig bedre sammenfall med observasjonene enn den uprosesserte prognosen. Usikkerhetsmodellens sannsynlighetsfordeling for temperaturen treffer litt lavt i forhold til den observerte temperaturen, men gir ellers en relativt god beskrivelse av temperaturen gjennom hele prognoseperioden.

4.3 Oppsummering vannføring

Vannføringsmodellen har en tendens til å underestimerer vannføringen, og tendensen er større for den uprosesserte enn for den korrigerede prognosen. Tendensen øker med prognosedagen. Medianen i utfallene fra usikkerhetsmodellen skårer også bedre på godhetskriteriene RMSE og N-S-kriteriet enn den uprosesserte prognosen. Det er stor variasjon i hvor godt prognosen treffer fra felt til felt, men de aller fleste feltene har en

akseptabel N-S-verdi de første prognosedagene. Hvor mye den korrigerte prognosen forbedrer N-S-verdien varierer mye over modellfeltene, men det er vanskelig å finne et mønster i hvor mye den forbedres relatert til geografi, arealstørrelse eller hvor god modellen er i utgangspunktet.

Kvantilplott er et evalueringskriterium som viser karakteristiske avvik mellom observasjoner og modellsimuleringer for enkeltfelt. De er ikke så godt egnet til å vurdere de samlede resultatene. Treffraten i forhold til flomnivåer er analysert, men evalueringsperioden på to år gir få hendelser på de høyeste nivåene, så analyse materialet er begrenset. Tendensen er at medianen har betydelig bedre treffrate enn den uprosesserte prognosen de to første prognosedagene. Treffraten avtar med prognosedagen og med flomstørrelsen, men treffraten for middelflom første prognosedag ligger på et brukbart nivå. Treffraten må ses i sammenheng med falsk alarmraten, som viser samme tendens som treffraten: medianen gir best skår de to første prognosedagene, og fungerer brukbart for middelflom og kort varslings tid. Langtidsvarsling av flomhendelser, derimot, bærer preg av å være skudd i blinde og har liten nytteverdi.

Analysen av delfeilmodellenes bidrag viser at modellfeilen har et konstant bidrag til totalfeilen gjennom hele prognoseperioden, og bidrar med den største andelen, målt i absoluttfeil, i starten av perioden. Prognosefeilens andel av totalfeilen øker utover i prognoseperioden, i takt med at usikkerheten i værprognosen øker. Dette gjelder alle modellfelt. Den korrigerte prognosen har en mindre absolutt- og middelfeil enn den uprosesserte de første 3-4 prognosedagene. På grunn av store negative avvik utover i prognoseperioden, har medianen en relativt lav additiv feil mot slutten av prognoseperioden. På grunn av at feilmodellens nedbørprognose treffer dårligere enn den uprosesserte, har den isolerte korrigeringen av vannføringsprognosen med hensyn til værprognosen ingen positiv effekt. Korrigering med hensyn til modellfeilen, derimot, gir en bedre prognose de første dagene. Den positive effekten av modellfeilkorrigeringen dominerer over det negative utslaget av nedbørprosesseringen, og gir en prognose med totalt sett mindre avvik de første 3 - 4 dagene.

Feilmodellens sannsynlighetsestimater for de ulike flomnivåene gir, for første prognosedag, et brukbart representativt bilde av antall påfølgende forekomster av flommer, men med en overrepresentativitet i gruppene for lav og høy sannsynlighet for flom. Denne tendensen forsterker seg med flomstørrelsen. Antall hendelser å analysere er imidlertid lite for større flommer. Flomstørrelsen har en logisk og god fordeling i forhold til sannsynlighetsestimatet. Fleredagersprognosene viser en fordeling av flommer i forhold til sannsynlighetsestimatene med stort innslag av lave sannsynlighetsestimater, og lite bidrag fra høyere sannsynlighetsklasser.

Usikkerhetsmodellens sannsynlighetsfordeling for vannføringen fanger opp litt for få hendelser innenfor 50- og 90%-intervallene. Avvikene i underkant av intervallene er flere enn avvikene i overkant, og tendensen er klar gjennom hele prognoseperioden. Ser man på sannsynlighetsfordelingen for alle modellfeltene, er spredningen stor, men de aller fleste feltene har en rimelig fordeling av observasjoner i forhold til persentilene fra usikkerhetsmodellen.

4.4 Anbefalinger

Evalueringen av usikkerhetsmodellen har gitt verdifull erfaring i bruk av ulike evalueringskriterier, både for evaluering av visuell tilpasning, sammenfall, treffsikkerhet og presisjon. Sammen med rapporten fra FoU-prosjekt 81057s delprosjekt 1 (Haddeland et al. 2014), gir rapporten et praktisk bidrag til videreutvikling av flomvarslingas prosedyrer for evaluering og forbedring av modeller og prognoser. Bruken av ulike kriterier i evalueringen av modeller og prognoser i disse to delprosjektene har eksemplifisert hvordan de forskjellige kvalitetsmålene gir utfyllende informasjon om modellenes styrker og svakheter, og hvordan ulike kvaliteter kan visualiseres på en kompakt og informativ måte.

Når det gjelder usikkerhetsmodellens evne til å kvantifisere forventet feil i vannføringsprognosene, og deretter korrigere for denne feilen, viser resultatene av evalueringen at delfeilmodellen for nedbør feiler, ved at den korrigerte nedbørprognosen faktisk gir dårligere resultater enn den ukorrigerte. Feilmodellen for temperatur, derimot, gir en forbedring i temperaturprognosen, som i situasjoner der temperaturen spiller en rolle, dvs. når temperaturen ligger rundt null og det er nedbør og/eller det ligger snø, må antas å ha en positiv effekt på vannføringsprognosen. Den autoregressive modellen som korrigerer HBV-modellens vannføringsprognose med hensyn til siste observerte vannføringsverdi gir også en betydelig forbedret vannføringsprognose de første prognosedagene.

Som det ble beskrevet i avsnitt 2.5, er kalibrering av usikkerhetsmodellen omstendelig og ressurskrevende, og dessuten eksisterer det ofte ikke tidsserier av værprognoser av tilfredsstillende lengde å kalibrere på. Dette kan være noe av årsaken til nedbørmodellens dårlige resultater i denne evalueringen, men det er også et generelt driftsproblem ved bruk av denne usikkerhetsmodellen at den ikke nødvendigvis er optimalt tilpasset modellsystemet den skal beregne usikkerheten til. Dette forholdet, sammen med resultatene fra evalueringen, bidrar til en anbefaling om å ta i bruk nye produkter og metoder for kvantifisering av usikkerheten i flomvarslingas vannføringsprognoser heller enn å videreutvikle dagen usikkerhetsmodell. Tilgjengelige produkter som beskriver usikkerheten i værprognosene er modellensembler og romlige sannsynlighetsestimater. Ensemble fra EC-modellen har vært tilgjengelige lenge, mens ensemble som representerer korttidsmodellen (nå AROME) vil bli tilgjengelige i løpet av 2015. Romlige estimater av kvantilfelt for nedbør, ved bruk av en nabolagsmetode, har også vært tilgjengelige en tid. Også når det gjelder dataassimilering/oppdatering er det god anledning til å ta i bruk alternativ metodikk, både med hensyn til vannføring og andre tilstandsvariable. Denne anbefalingen er allerede implementert i et oppfølgende FoU-prosjekt, 80103, Treffsikkerhet og presisjon i hydrologiske prognoser, der arbeidspakke 2 har fokus på kvantifisering av usikkerhet i vannføringsprognoser som har sitt opphav i usikre værprognoser, og arbeidspakke 3 skal teste ut alternativ metodikk for dataassimilering.

Det vil imidlertid ta noe tid å få i operasjonell drift et nytt system for kvantifisering av usikkerhet i vannføringsprognosene. Inntil det er på plass, tas resultatene fra denne evalueringen til følge ved å stoppe prosedyren som estimerer feil og korrigerer for den i nedbørprognosen. Isteden vil nedbørinndata til HBV-modellen inntil videre være operasjonell nedbørprognose i tillegg til nedbørprognose +/- en viss prosent. Når romlige

kvantilfelt fra nabolagsmetoden er implementert i vårt system, vil disse brukes i tillegg til operasjonell prognose for å si noe om usikkerheten i nedbørprognosen. Temperaturprognosen vil korrigeres som før, og det samme vil modellfeilen.

5 Referanser

Bremnes, J.B. 1998. Statistical forecasting of precipitation conditional on numerical weather prediction models. HYDRA-notat 8/98, 15 s.

Follestad, T. & Høst, G. 1998. A statistical model for the uncertainty in meteorological forecasts. HYDRA-notat 12/98, 46 s.

Haddeland, I., T. Væringstad, E. Langsholt og T. Skaugen. 2014. Evaluering av flomvarslingas verktøy. NVE-rapport 84–2014.

Langsholt, E. 2000. Quantifying uncertainty in runoff forecasts: application to an HBV-model based forecasting and flood warning routine. Proceedings from the European Conference on Advances in Flood Research. Potsdam Institute for Climate Impact Research, November 2000. PIK Report No. 65.

Langsholt, E. og T. Væringstad. 2004. Quantifying uncertainty in stream flow forecasts – validation of an operative routine. 2nd ACTIF Workshop Delft, 22 – 23th November 2004.

Langsrud, Ø., Frigessi, A. & Høst, G. 1998. Pure model error of the HBV-model. HYDRA-notat 4/98, 28 s.

Langsrud, Ø., Høst, G., Follestad, T., Frigessi, A. & Hirst, D. 1999. Quantifying uncertainty in HBV runoff forecasts by stochastic simulations. HYDRA-notat 2/99, 38 s.

Mohr, M. 2008. New Routines for Gridding of Temperature and Precipitation observations for “seNorge.no”. Met.no note No. 08/2008.

WWRP/WGNE Joint Working Group on Forecast Verification Research. 2013. Forecast Verification: Issues, Methods and FAQ. <http://www.cawcr.gov.au/projects/verification/>



Norges
vassdrags- og
energidirektorat

Norges vassdrags- og energidirektorat

Middelthunsgate 29
Postboks 5091 Majorstuen
0301 Oslo

Telefon: 09575
Internett: www.nve.no

