

## Samkjøring av vær- og vannføringsdøgnet i hydrologiske modeller

.....  
EN TEST MED BRUK AV ULIKE DØGNINDELINGER

*Elin Langsholt*



## Rapport nr 71-2018

# Samkjøring av vær- og vannføringsdøgnet i hydrologiske modeller

**Utgitt av:** Norges vassdrags- og energidirektorat

**Forfatter:** Elin Langsholt

**Trykk:** NVEs hustrykkeri

**Forsidefoto:** Drammenselva. Foto: Elin Langsholt

**ISBN:** 978-82-410-1728-5

**Sammenheng:** Rapporten presenterer en test der HBV-modellen kjøres med fire ulike kombinasjoner av døgninndelinger av datasettene for vær- og vannføringsdata, og et utvalg kvalitetskriterier sammenliknes. Testen konkluderer med en anbefaling om å bruke samkjørt døgninndeling for datasettene i den videre utviklingen av flomvarslingas modellsystem. Studien er gjort innenfor det interne FoU-prosjektet 80103, Treffsikkerhet og presisjon i hydrologiske prognoser.

**Emneord:** hydrologiske modeller nedbørdata temperaturdata vannføringsdata godhetskriterier

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

Telefon: 22 95 95 95  
Epost: [nve@nve.no](mailto:nve@nve.no)  
Internett: [www.nve.no](http://www.nve.no)

August 2018

# Innhold

<b>Forord</b> .....	<b>4</b>
<b>Sammendrag</b> .....	<b>5</b>
<b>1 Innledning</b> .....	<b>6</b>
<b>2 Metode</b> .....	<b>7</b>
2.1 Oppsett av testen.....	7
2.2 Datasett med sammenfallende døgn .....	8
2.2.1 Døgninndelingen av nedbør- og temperaturdøgnet.....	8
2.2.2 Døgninndelingen av vannføringsdøgnet.....	10
2.3 Kalibrering .....	11
2.4 Evalueringskriterier .....	11
2.4.1 Nash-Sutcliffekoeffisienten.....	11
2.4.2 Kling-Guptakriteriet (KGE).....	12
2.4.3 Taylordiagrammet .....	13
2.4.4 Treffrate/falsk alarm/suksessindeks-kriteriet .....	14
<b>3 Resultater</b> .....	<b>15</b>
3.1 Første runde: modellparametere tilpasset tradisjonell døgninndeling.....	15
3.1.1 Nash-Sutcliffekoeffisienten.....	15
3.1.2 Kling-Guptakriteriet .....	17
3.1.3 Taylordiagrammet.....	19
3.1.4 Treffrate/falsk alarm/suksessindeks-kriteriet .....	20
3.2 Andre runde: modellparametere tilpasset hvert enkelt datasett.....	23
3.2.1 Nash-Sutcliffekoeffisienten.....	23
3.2.2 Kling-Guptakriteriet .....	24
3.2.3 Taylordiagrammet.....	26
3.2.4 Treff/falsk alarm/suksessindeks-kriteriet.....	27
<b>4 Drøfting og konklusjon</b> .....	<b>29</b>
<b>5 Referanser</b> .....	<b>30</b>

# Forord

Innenfor rammene av det interne FoU-prosjektet 80103, Treffsikkerhet og presisjon i hydrologiske prognoser, arbeides det med å forbedre de hydrologiske prognosene på ulike måter. Denne rapporten presenterer en studie som tester ut ulike måter å dele inn og samkjøre døgnnet på for vær- og vannføringsdata, og hvilken effekt dette har på HBV-modellkjøringene for flomvarslingsfeltene. Bakgrunnen er at datagrunnlaget for modellene stadig forbedres i retning av større tilfang av data med fin tidsoppløsning og lenger dataserier, og vi er nå i en situasjon der dataene vi bruker i de operasjonelle modellkjøringene ikke lenger nødvendigvis må være bundet av observasjonstidspunkter og tradisjonell organisering av tidsseriene. Formålet med studien er å komme med en anbefaling for hvordan modelldatasettene skal organiseres for de døgnbaserte modellene i den videre utviklingen av systemet.

Oslo, august 2018

  
avdelingsdirektør

  
seksjonssjef

# Sammendrag

Hydrologiske modeller med 24 timers tidsskritt spiller en viktig rolle som beslutningsstøtteverktøy i NVEs flomvarslingstjeneste. Inndataene til 24-timersmodellene har tradisjonelt vært definert av et viktig observasjonstidspunkt for de manuelle værobservasjonene, kl. 07. 24-timersnedbøren som brukes i modellene er den akkumulerte nedbøren over de 24 timene forut for kl. 07, og 24-timerstemperaturen er middeltemperaturen over de samme 24 timene. Denne nedbøren og temperaturen tilordnes det døgnet som omslutter observasjonstidspunktet og beregner avløpet for dette døgnet. Nedbør- og temperaturdøgnet ligger altså 17 timer foran døgnet de tilordnes, og dette bidrar til at den simulerte vannføringen som regel ligger litt etter den observerte. Det oppstår dessuten lett forvirring rundt tidsintervallet for nedbør, temperatur og vannføring og sammenhengen mellom dem når modellresultater skal formidles.

Det skjer en stadig forbedring av datagrunnlaget, og vi er nå i en situasjon der dataene vi bruker i de operasjonelle modellkjøringene ikke lenger nødvendigvis må være bundet av observasjonstidspunkter og tradisjonell inndeling av tidsseriene. Vi kan redefinere vannføringsdøgnet ved å bruke findata, eller alternativt redefinere nedbør- og temperaturdøgnet ved å bruke data fra automatiske værstasjoner med fin tidsoppløsning eller ved å bruke disaggregerte døgndata. De forskjellige metodene er imidlertid beheftet med ulemper i form av dårligere datakvalitet og kortere tidsserier. Det bør derfor undersøkes om ulempene ved å samkjøre vær- og vannføringsdøgnet medfører dårligere modellsimuleringer i en slik grad at ulempene ved omleggingen mer enn oppveier fordelene.

Innenfor det interne FoU-prosjektet 80103 – Treffsikkerhet og presisjon i hydrologiske prognoser er det derfor satt opp en test der en av de operative hydrologiske modellene (HBV) kjøres med fire ulike kombinasjoner av døgninndelinger av datasettene, og et utvalg kvalitetskriterier sammenliknes. Formålet med studien er å komme med en anbefaling for hvordan modelldatasettene skal organiseres for de døgnbaserte modellene i den videre utviklingen av systemet.

Det er ikke veldig stor forskjell på hvordan modellene presterer med de ulike datasettene. Rangeringen av datasettene varierer etter kvalitetskriterium og testoppsett. Datasettene med samkjørte døgn scorer stort sett hårfint bedre enn det tradisjonelle datasettet. Effekten av å samkjøre døgnet gir størst forbedring i små nedbørfelt. Datasettene med samkjørte data gir vannføringssimuleringer med best timing med hensyn til flomtopper. Testen konkluderer med en klar anbefaling om å bruke samkjørte data i den videre driften av flomvarslingsmodellene. Det er ikke mulig å gi en entydig anbefaling når det gjelder hvilket av de samkjørte datasettene som bør brukes.

# 1 Innledning

Hydrologiske modeller spiller en viktig rolle som beslutningsstøtteverktøy i NVEs flomvarslingstjeneste. Fortsatt er det 24-timersmodellene som er mest brukt i den daglige vurderingen av flomfare, selv om modeller med 3 timers tidsskritt er tilgjengelige og et verdifullt verktøy spesielt for intensive og kortvarige hendelser. Modellene kjøres med nedbør og temperatur som inndata, og modellenes simulerte vannføring sammenliknes med observert vannføring. I flomvarslingstjenesten kjøres modellene med nedbør- og temperaturprognoser, og de simulerte vannføringsprognosene brukes til å vurdere flomfare.

Inndataene til 24-timersmodellene har tradisjonelt vært definert av observasjonstidspunktet for de manuelle værobservasjonene. Disse er koordinert internasjonalt, og de viktigste observasjonstidspunktene, norsk normaltid, er kl. 07, 13 og 19. Det er tidspunktet 07 som definerer inndataene til 24-timersmodellene. 24-timersnedbøren som brukes i modellene er den nedbøren som ble observert kl. 07, det vil si den akkumulerte nedbøren over de 24 timene forut for kl. 07. 24-timerstemperaturen er middeltemperaturen over de samme 24 timene. Modellene tilordner denne nedbøren og temperaturen til det døgnet som omslutter observasjonstidspunktet og beregner avløpet for dette døgnet. Nedbør- og temperaturdøgnet ligger altså i realiteten 17 timer foran døgnet de tilordnes, og dette bidrar til at den simulerte vannføringen som regel ligger litt etter den observerte, spesielt i de mindre nedbørfeltene med raskere responstid. Det oppstår dessuten lett forvirring rundt tidsintervallet for nedbør, temperatur og vannføring og sammenhengen mellom dem når modellresultater skal formidles.

Det skjer en stadig forbedring av datagrunnlaget, og vi er nå i en situasjon der dataene vi bruker i de operasjonelle modellkjøringene ikke lenger nødvendigvis må være begrenset av observasjonstidspunkter og tradisjonell datastruktur. Vi kan redefinere vannføringsdøgnet ved å bruke findata, som kan aggregeres til døgndata med vilkårlig døgnskilletidspunkt, for eksempel det som sammenfaller med værdøgnet (kl. 07-07). Alternativt kan vi redefinere nedbør- og temperaturdøgnet ved å bruke data fra automatiske værstasjoner med fin tidsoppløsning eller ved å bruke disaggregerte, tradisjonelle døgndata som aggregeres med ny tidsinndeling, for eksempel den som sammenfaller med det tradisjonelle vannføringsdøgnet (kl. 00-00).

På ulike måter er vi altså nå i stand til å samkjøre vær- og vannføringsdøgnet. De forskjellige metodene er imidlertid beheftet med ulemper i form av dårligere datakvalitet og kortere tidsserier. Før man eventuelt går over til å bruke sammenfallende vær- og vannføringsdøgn, bør det derfor undersøkes om disse ulempene medfører dårligere modellsimuleringer i en slik grad at ulempene ved omleggingen mer enn oppveier fordelene ved å samkjøre døgnet. Innenfor det interne FoU-prosjektet 80103 – Treffsikkerhet og presisjon i hydrologiske prognoser er det derfor satt opp en test der en av de operative hydrologiske modellene (HBV) kjøres med fire ulike kombinasjoner av døgninndelinger av datasettene, og et utvalg kvalitetskriterier sammenliknes. Formålet med studien er å komme med en anbefaling for hvordan modelldatasettene skal organiseres for de døgnbaserte modellene i den videre utviklingen av systemet.

# 2 Metode

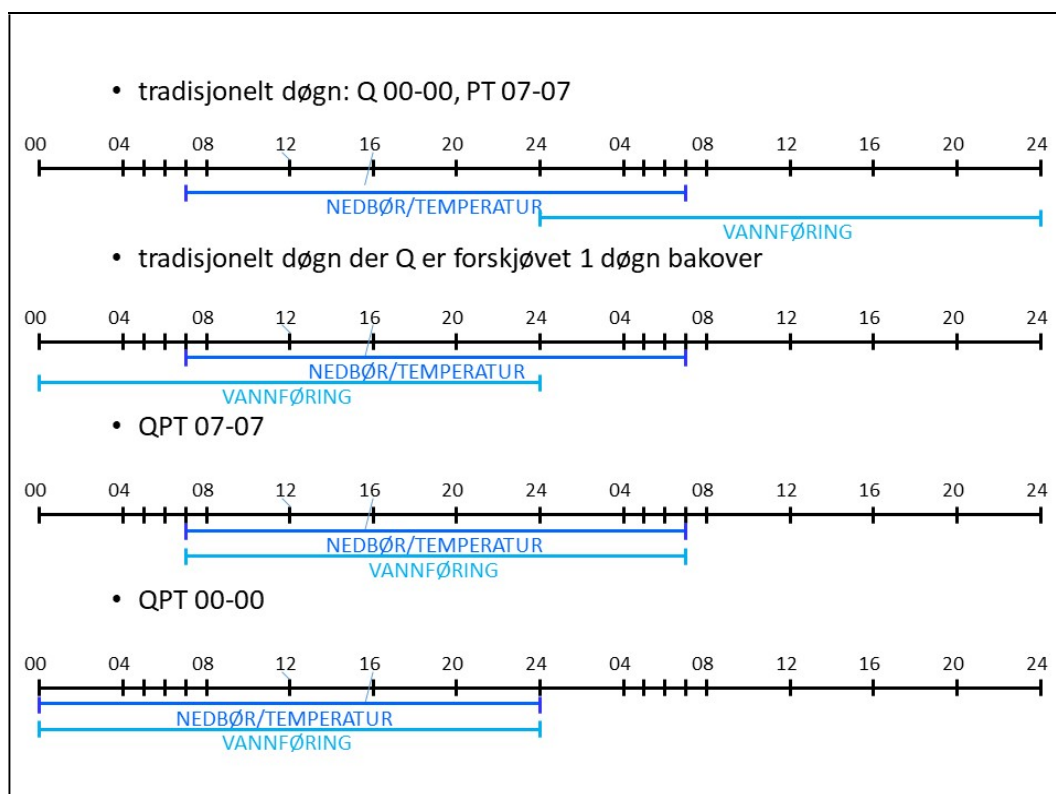
## 2.1 Oppsett av testen

Fire ulike datasett sammenliknes:

beskrivelse	døgninndeling for vannføringsdata	døgninndeling for værdata
tradisjonelt døgn	$00_t - 00_{t+1}$	$07_{t-1} - 07_t$
forskjøvet vannføringsdøgn	$00_{t-1} - 00_t$	$07_{t-1} - 07_t$
samkjørt døgn 07-07	$07_{t-1} - 07_t$	$07_{t-1} - 07_t$
samkjørt døgn 00-00	$00_{t-1} - 00_t$	$00_{t-1} - 00_t$

Tabell 1. Døgninndelingen for de fire ulike datasettene som er sammenliknet i testen. Klokkeslettet merket med datoen i forhold til referansedatoen t er vist.

Figur 1 viser en grafisk sammenlikning av døgninndelingen som er brukt i de fire datasettene.



Figur 1. Grafisk framstilling av døgninndelingen i de fire datasettene.

Det tradisjonelle døgnet er den døgninndelingen som kjøres operasjonelt nå. Det tradisjonelle døgnet der vannføringsdøgnet er forskjøvet ett døgn bakover ble tatt med i testen for å se effekten av denne veldig enkle forskyvningen, der datasettene har bedre overlapp, samtidig som vannføringen, som jo er en konsekvens av været, kommer noen

timer før korresponderende vær, noe som er litt ulogisk. Samkjørte døgn krever en litt mer omfattende restrukturering av dataene. Dette er forklart nedenfor.

Testen ble utført i to runder:

1. Først ble modellene kjørt med modellparameterne som er beregnet i den foreliggende kalibreringen, som er utført på datasettene med tradisjonell døgninndeling
2. Deretter ble modellene kalibrert for hvert datasett før modellene ble kjørt og kvalitetskriteriene beregnet

Den første delen av testen forskjellsbehandler datasettene og gir en fordel til det tradisjonelle datasettet ved at parameterne ved kalibreringen er tilpasset dette datasettet. Runden ble likevel kjørt for å gi en første antydning om hvor følsomme modellene er med hensyn til døgninndelingen. En relativt kort periode med en stor andel stasjoner med findata ble brukt.

I den andre delen av testen ble en lenger periode med lavere findatadekning brukt, for å få rom til både en kalibreringsperiode og en uavhengig valideringsperiode.

## 2.2 Datasett med sammenfallende døgn

Det er to naturlige valg når det gjelder felles tidsinndeling av døgnet for 24-timers vær- og vannføringsdata: døgninndelingen som brukes i databasen (hydra2) for å definere vannføringsdøgnet (kl. 00-00) og den tradisjonelle døgninndelingen av værdøgnet (kl. 07-07). Ved det første alternativet må disaggregerte værdøgndata eller data fra automatiske værstasjoner med finere tidsoppløsning benyttes istedenfor døgnobservasjoner. Ved alternativ to må vannføringsdata med fin tidsoppløsning benyttes og aggregeres over perioden kl. 07 – 07.

### 2.2.1 Døgninndelingen av nedbør- og temperaturdøgnet

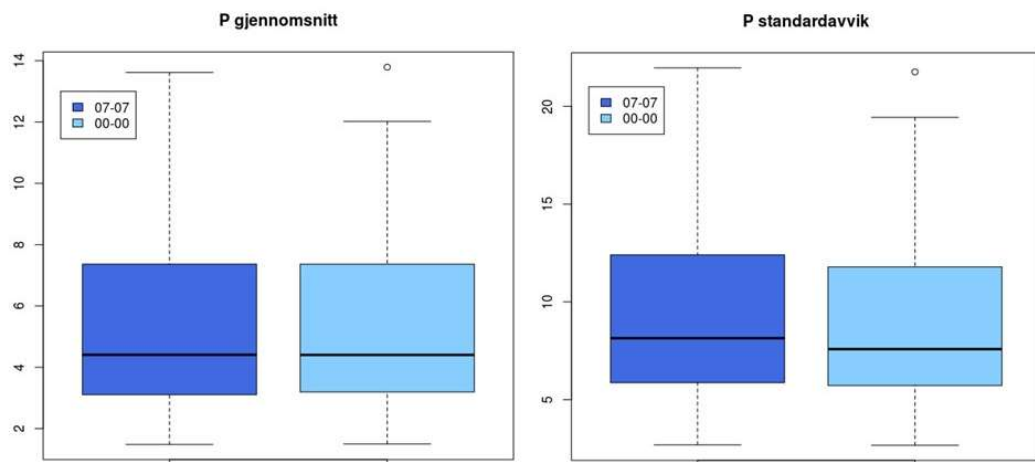
Når det gjelder nedbør- og temperaturdata, er det slik at den beste romlige dekningen oppnås ved å bruke de målingene som er gjort kl. 07/19, som omfatter både automatiske og manuelle målinger. Standard døgndata, som vises i xgeo.no og som tradisjonelt har vært brukt i de operasjonelle kjøringene av de døgnbaserte hydrologiske modellene, er basert på 07-observasjonene. Det er utviklet en metodikk for disaggregering av dette datasettet ved å bruke data fra værprognosemodellene med finere tidsoppløsning (Vormoor and Skaugen, 2013), og den brukes til å generere findataserier med 3 timers oppløsning for nedbør og temperatur. Den fungerer slik at den observerte 24-timersverdien fordeles over døgnet etter samme mønster som findataene fra prognosemodellen for det samme døgnet. Denne metodikken er brukt til å generere et historisk findatasett for nedbør og temperatur (1957-2010), og den brukes operasjonelt for å generere inndata til den ene hydrologiske modellen (DDD) som per våren 2018 går med et tidsskritt på 3 timer. Den variasjonen over døgnet som oppnås gjennom disaggregering er basert på modellering, og disse findataene tilsvarer ikke fullgode observasjoner. Datakvaliteten for det redefinerte døgnet vil dermed være noe redusert.

Stadig flere av de meteorologiske stasjonene blir automatisert. Data fra slike stasjoner har fin tidsoppløsning, og kan brukes til å lage døgndata for 00-00-døgnet. Da metodikken for disaggregering av døgndataene ble utviklet, fantes det færre automatstasjoner enn det gjør

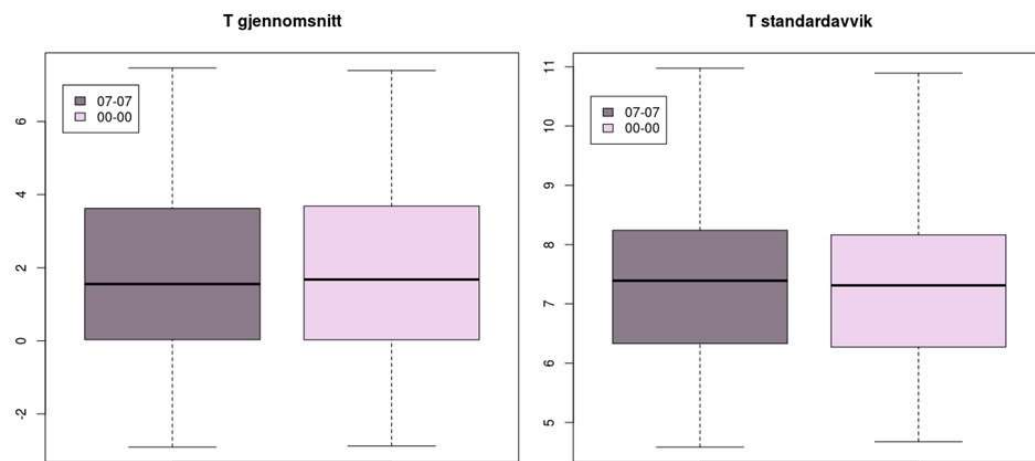


nå. Den begrensede romlige dekningen av automatstasjoner gjorde at kvaliteten på døgndataene for det redefinerte døgnet, basert på dette findatasettet, ble dårligere enn døgndata basert på det disaggregerte datasettet. Derfor ble det disaggregerte datasettet valgt som utgangspunkt for det redefinerte værdøgnet i denne studien. Etter hvert som antallet automatstasjoner øker, vil trolig disse danne et bedre utgangspunkt for 00-00 døgndata enn det disaggregerte datasettet, og kvaliteten på værddata for 00-00-døgnet vil øke.

Gjennomsnitt og standardavvik for de to nedbør- og temperaturdatasettene (07-07-døgnet og 00-00-døgnet) ble sammenliknet, alle modellfelt sett under ett. Figur 2 viser bokplott av gjennomsnitt og standardavvik for nedbørseriene til modellfeltene, og Figur 3 viser tilsvarende plott for temperaturseriene. Gjennomsnittene er svært sammenfallende, mens det er litt avvik i standardavvikene. Forskjellene er så små at de regnes som ubetydelige for denne analysen.



Figur 2. Bokplott som viser fordelingen av gjennomsnittet og standardavviket til nedbørdataene for 07-07 (mørk blått)- og 00-00 (lys blått)-døgnet over alle modellfeltene.



Figur 3. Bokplott som viser fordelingen av gjennomsnittet og standardavviket til temperaturdataene for 07-07 (mørk blått)- og 00-00 (lys blått)-døgnet over alle modellfeltene.

## 2.2.2 Døgninndelingen av vannføringsdøgnet

Findata for vannføring er beheftet med flere observasjonsfeil og er i mindre grad kvalitetskontrollert enn døgndata. Tidsseriene er dessuten stor sett betydelig kortere enn seriene med døgndata.

Når det gjelder observasjonsfeil, er det største problemet observasjoner i vinterhalvåret der dataene er påvirket av isoppstuvning. Dette gjelder mange stasjoner. Når vannføringsdataene kvalitetskontrolleres, er det døgndataene som gjennomgås og isreduseres. Findataene isreduseres i utgangspunktet ikke. Det er imidlertid utviklet et program som er tilgjengelig fra start-systemet, og som utfører en automatisk isreduksjon av findataene med utgangspunkt i isreduksjonen som er gjort på døgndataene. Hvilke data som er ispåvirka er definert av døgndatagjennomgangen, og findataene justeres med samme faktor som er funnet for døgndataene. De iskorrigerte dataene legges på et virtuelt arkiv, «virtuelt isreduserte findata» (35 HYKVALP\_ICECORR). Det finnes også et arkiv der de ispåvirka dataene er fjernet (34 HYKVALP\_RM\_ICECORR).

For å hente ut døgndata som er aggregert over en annen 24-timersperiode enn kl. 00 til 00, anvender man en verdi for «offset» i kommandoen som henter dataene ut fra findataarkivet i databasen. En offset på 420 minutter (7 timer) sammen med en aggregeringstid (base\_time) på 1440 minutter (24 timer) tar ut findata for et døgn som varer fra kl. 07 til kl. 06:59 neste døgn. For at døgnet skal sammenfalle med det tradisjonelle meteorologiske døgnet, må verdien tilordnes det døgnet som omslutter sluttidspunktet for det aggregerte døgnet.

Det er først i de siste 30-40 årene at utstyr for å registrere vannføringsdata med fin tidsoppløsning har blitt installert i stort omfang. Tidsseriene med findata er derfor i de fleste tilfeller betydelig kortere enn tidsseriene med døgndata. For å få et bredest mulig vurderingsgrunnlag ble testperioden for den første del av testen valgt ut på bakgrunn av hvor langt tilbake modellstasjonene hadde findata. Tabell 2 viser hvor mange av modellstasjonene som har findata i utvalgte tidsperioder. I 20-årsperioden 1994 – 2014 har 127 av de 145 modellstasjonene findata i hele perioden, og denne perioden ble valgt. Modellene ble kjørt fra 01.09.1994 og evalueringskriteriene ble beregnet fra 01.01.1995.

periode	antall modellstasjoner med findata
1961 – 2014	35
1984 – 2014	105
1994 – 2014	127

Tabell 2. Antall modellstasjoner med findata i forskjellige perioder.

Når det gjelder den andre delen av testen, måtte det være rom for både en kalibreringsperiode og en uavhengig valideringsperiode, så en lenger periode enn 20 år var ønskelig. En periode som har fungert som en standardperiode ved modellsammenlikning tidligere ble valgt:

- kalibrering: 01.09.2000 – 31.12.2014
- validering: 01.09.1985 – 31.08.2000

## 2.3 Kalibrering

Modellene ble kalibrert med optimaliseringsalgoritmen PEST, Model-Independent Parameter Estimation & Uncertainty Analysis ([www.pesthomepage.com](http://www.pesthomepage.com)), som i flere år har vært brukt i NVE til kalibrering av de operasjonelle HBV-modellene. PEST optimaliserer modellparameterne ved å minimere objektfunksjonen

$$\Phi = \sum_{i=1}^n [w^i (obs^i - sim^i)]^2$$

der  $w^i$  er en vektcoeffisient og  $obs^i$  og  $sim^i$  er henholdsvis observerte og simulerte variable. I kalibreringen av de hydrologiske modellene brukes vannføringen ved alle tidsskritt i kalibreringsperioden i objektfunksjonen, i tillegg til avviket mellom observert og simulert volum i kalibreringsperioden. Vektcoeffisienten er satt lik 1.0 for alle elementer. Objektfunksjonen kan i vår anvendelse da skrives som

$$\Phi = \sum_{t=1}^T [(Q_{obs}^t - Q_{sim}^t)]^2 + (V_{obs} - V_{sim})^2 = T \cdot MSE + (V_{obs} - V_{sim})^2$$

der  $Q_{obs}^t$  og  $Q_{sim}^t$  er henholdsvis observert og simulert vannføring ved tiden  $t$ ,  $MSE$  er midlere kvadratavvik,  $T$  er antall tidsskritt og  $V_{obs}$  og  $V_{sim}$  er henholdsvis observert og simulert volum.

Metoden er ikke anvendt med en global tilnærming. Kalibreringen tar utgangspunkt i ett punkt i parameterrommet, og det er godt mulig at det optimale parametersettet som kalibreringen ender opp med ikke er et globalt optimum. Det antas likevel at for dette formålet så er metodikken tilfredsstillende til å gi et representativt resultat.

Det er 12 modellparametere som tilpasses gjennom kalibreringen. For mer informasjon om modellen, modellfeltene og parameterne henvises til Ruan og Langsholt (2017).

Da testen ble innledet var inndataene basert på versjon 1 av seNorge-griddet. Modellene var kalibrert med data fra dette griddet og data med alternative døgninndelinger ble generert med dette griddet som utgangspunkt. Da del 2 av testen (kalibrering på hvert datasett) skulle utføres, var versjon 2 av seNorge-griddet blitt operativt, og testen ble basert på det. Dette griddet antas å gi nedbør- og temperaturverdier som er bedre fordelt over terrenget enn det gamle. Derfor leses og tilordnes inndata fra versjon 2 høydesonevis til nedbørfeltet uten høydekorrigerings. Tidligere ble inndataene lest inn for hele feltet under ett, og et gjennomsnitt for feltet ble brukt som inndata, med høydegradienter for nedbør og temperatur som ble tilpasset i kalibreringen.

## 2.4 Evalueringkriterier

Modellresultatene med hensyn til de ulike datasettene er vurdert og sammenliknet ved bruk av fire forskjellige evalueringkriterier, Nash-Sutcliffekoeffisienten, Kling-Guptakriteriet, taylordiagrammet og treffrate/falsk alarm/suksessindeks-kriteriet.

### 2.4.1 Nash-Sutcliffekoeffisienten

Nash-Sutcliffekoeffisienten (NS) er det tradisjonelt mest brukte kvalitetskriteriet når det gjelder vurdering av hydrologiske modellsimuleringer i NVE, og er et mål på hvor godt

sammenfall det er mellom observert og simulert vannføring. NS er ekvivalent med determinasjonskoeffisienten,  $R^2$ , og uttrykkes ved

$$NS = 1 - \frac{\sum_{t=1}^T (Q_{obs}^t - Q_{sim}^t)^2}{\sum_{t=1}^T (Q_{obs}^t - \overline{Q_{obs}})^2} = 1 - \frac{MSE}{\sigma^2_{Q_{obs}}}$$

Her er  $Q_{obs}^t$  observert vannføring ved tiden  $t$ ,  $Q_{sim}^t$  er simulert vannføring ved tiden  $t$ ,  $\overline{Q_{obs}}$  er gjennomsnittet av de observerte vannføringene,  $T$  er antall tidsskritt,  $MSE$  er midlere kvadratavvik og  $\sigma^2_{Q_{obs}}$  er variansen til vannføringsobservasjonene. NS kan ha verdier fra  $-\infty$  til 1.0. Ved  $NS = 1.0$  er det perfekt sammenfall mellom simulert og observert vannføring, og jo nærmere NS-verdien er 1.0, dess bedre er simuleringen. Når  $NS = 0.0$ , dvs. når variansen i modellavviket er like stor som variansen til observasjonene, er gjennomsnittet av observasjonene et like bra estimat på vannføringen som modellsimuleringen, og hvis NS er negativ, er gjennomsnittet bedre. På grunn av kvadreringen er NS sensitiv for høye avvik. Det vil generelt medføre at tilpasning til flomvannføringer vektlegges tyngre enn tilpasning til moderate og lave vannføringer.

#### 2.4.2 Kling-Guptakriteriet (KGE)

Gupta et al. (2009) viste hvordan NS kunne dekomponeres, slik at den kan fungere som en multiobjektiv funksjon, et kvalitetskriterium som kvantifiserer hvor god modellen er med hensyn til ulike aspekter ved tilpasningen til observerte data. Denne omformuleringen av NS kalles KGE, og har tre komponenter som korresponderer med volumavvik, korrelasjon og variabilitet. To kalibreringer med samme NS- eller KGE-verdi kan ha forskjellige egenskaper, og verdiene av de tre komponentene viser styrke og svakheter mht. volumavvik, korrelasjon og variabilitet. Man kan bruke kriteriet slik at de tre komponentene veies ulikt etter hva man vil legge mest vekt på, men i denne testen veier kriteriene likt. En variant av KGE ble videreutviklet i Kling et al. (2012), der komponentene for volumavviket og variabiliteten ikke er krysskorrelert. Det er den varianten som benyttes her, og den kan uttrykkes

$$KGE = 1 - \sqrt{(r - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2}.$$

De tre komponentene av  $KGE$  er korrelasjonskoeffisienten

$$r = \frac{cov(Q_{obs}, Q_{sim})}{\sigma_{Q_{obs}} \sigma_{Q_{sim}}},$$

der  $cov(Q_{obs}, Q_{sim})$  er kovariansen mellom observert og simulert vannføring og  $\sigma_{Q_{obs}}$  og  $\sigma_{Q_{sim}}$  er standardavviket til henholdsvis de observerte og simulerte vannføringsdataene, volumavviksforholdet

$$\beta = \frac{\mu_{Q_{sim}}}{\mu_{Q_{obs}}},$$

der  $\mu_{Q_{obs}}$  og  $\mu_{Q_{sim}}$  er middelveien av henholdsvis de observerte og simulerte vannføringsdataene, og variabilitetsforholdet

$$\gamma = \frac{CV_{sim}}{CV_{obs}} = \frac{\sigma_{Q_{sim}}/\mu_{Q_{sim}}}{\sigma_{Q_{obs}}/\mu_{Q_{obs}}},$$

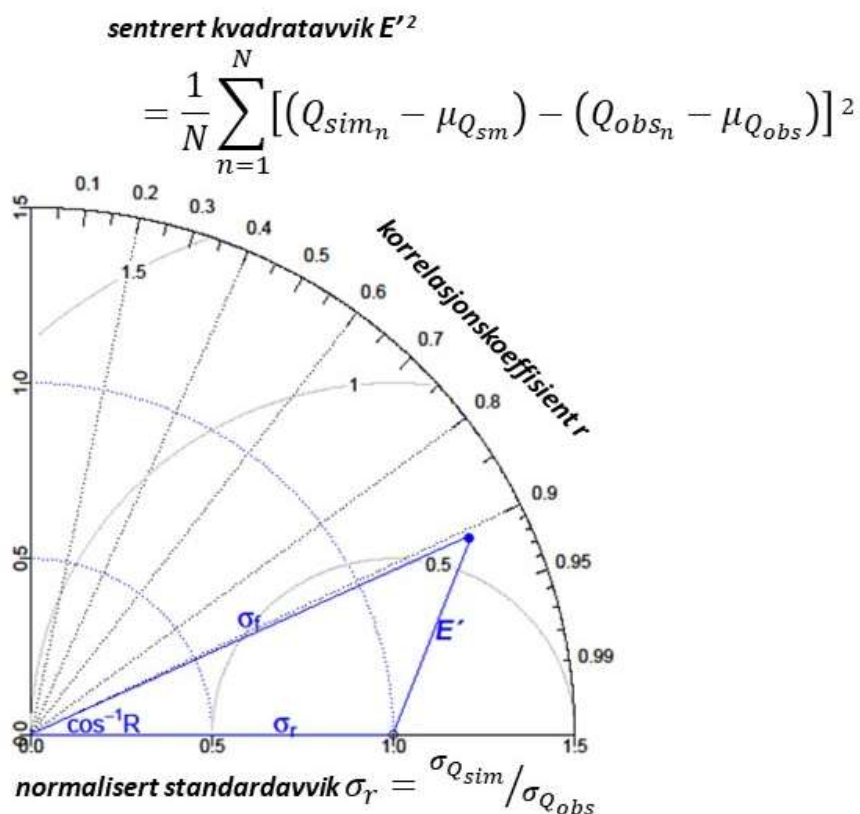
Der  $CV_{sim}$  og  $CV_{obs}$  er variasjonskoeffisienten til henholdsvis simuleringene og observasjonene.  $KGE$ ,  $r$ ,  $\beta$  og  $\gamma$  er optimale når verdien er 1.0. Graden av samvariasjon mellom simuleringene og observasjonene, som blant annet inkluderer hvordan simuleringen treffer observasjonene i tid, kvantifiseres av  $r$ , fordelingen av vannføringer, eller variasjonen i datasettet, gjenspeiles av verdien til  $\gamma$ , og  $\beta$  er et mål på volumtilpasningen.

KGE er konstruert slik at den har to fortrinn framfor NS. Som vi ser av definisjonen på NS, kan den uttrykkes som en funksjon av forholdet mellom midlere kvadratavvik og variansen. Det betyr at kvadratavviket vektet mindre i felt der variansen er stor, altså i felt med store variasjoner i vannføringen, enn i felt med liten varians. Det viser seg dessuten at optimal NS-verdi oppnås når standardavviket til simuleringene er mindre enn standardavviket til observasjonene. Optimalisering av NS vil altså underestimere variabiliteten i simuleringen, som vil medvirke til at flomtoppene underestimeres og lavvannshendelser overestimeres. Forsøk viser (se Gupta et al. 2009) at KGE optimaliserer vannføringssimuleringer med mindre volumavvik og bedre standardavvik, men litt dårligere korrelasjonskoeffisient enn NS.

### 2.4.3 Taylordiagrammet

Taylordiagrammet er en grafisk framstilling av tre kvalitetskriterier som har mye til felles med de tre komponentene av  $KGE$ . Taylordiagrammet ble foreslått som evalueringskriterium for flomvarslingsmodellene av Haddeland (2014).

Taylordiagrammet plotter korrelasjonskoeffisienten, kvadratavviket og standardavviket i samme diagram. Korrelasjonskoeffisienten er et mål på samvariasjonen mellom to variabler. Kvadratavviket måler hvor stort avvik det er i sammenfallet til de to variablene, mens standardavviket gir et mål på variasjonen innen datasettene, og kan for eksempel si noe om gjennomsnittlige flom- og lavvannsutslag. Perfekt tilpasning til observasjonene representeres ved punktet langs x-aksen der  $x$  er lik 1.0. Der er forholdet mellom datasettenes standardavvik lik 1.0, kvadratavviket mellom dem, som plottes som avstand til punktet, er lik 0.0, og korrelasjonen, som måles ved vinkelen fra y-aksen, er 1.0. Kvaliteten til de ulike modellene vurderes grafisk, ved å se hvilket datasett som ligger nærmest det optimale punktet. Plasseringen av datasettene i forhold til hverandre viser datasettenes styrker og svakheter mht. samvariasjon, avvik og ekstremverdiutslag



Figur 4. Taylordiagrammet. Det geometriske forholdet mellom korrelasjonskoeffisienten,  $r$ , kvadratavviket,  $E'$ , og normalisert standardavvik for datasettene er markert.

#### 2.4.4 Treffrate/falsk alarm/suksessindeks-kriteriet

I varslingssammenheng er det interessant å evaluere hvordan modellene treffer i forhold til en varslingsgrense. Et viktig spørsmål for flomvarsleren er: «vil det bli flom eller ikke?». Bruk av dette kriteriet for overskridelse/ ikke overskridelse av terskel ble foreslått som evalueringskriterium for flomvarslingsmodellene av Haddeland (2014).

Flomvarslingen har et sett kriterier for definisjon av flom på ulike nivåer. Kriteriene er ikke rigide, men må ses i sammenheng med hvor mange og hvilke stasjoner som har en vannføring som overstiger det gitte nivået og skadeomfanget som følge av hendelsen.

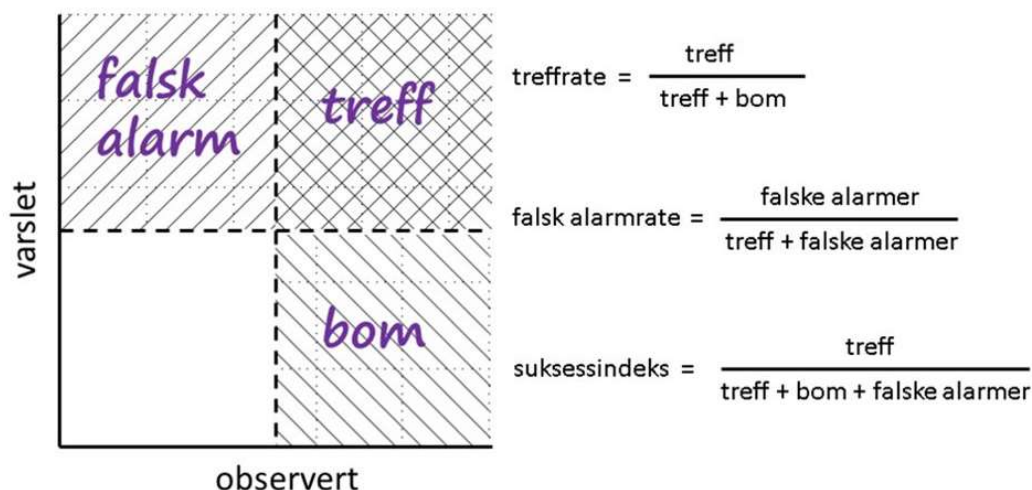
Kriteriene er:

- flom på gult nivå:  $\text{middelflom} < Q_{\text{obs}} < 5\text{-årsflom}$
- flom på oransje nivå:  $5\text{-årsflom} < Q_{\text{obs}} < 50\text{-årsflom}$
- flom på rødt nivå:  $50\text{-årsflom} < Q_{\text{obs}}$

I denne studien er det evaluert hvordan modellene treffer med hensyn til det laveste flomkriteriet, middelflom. Det er relativt få ganger på hver stasjon at vannføringen overstiger de høyere flomnivåene 5-årsflom og 50-årsflom, så når det gjelder disse nivåene blir det statistiske materialet for lite til å gjøre en meningsfull analyse.

Observasjonene testes mot nivået for observert middelflom, mens simuleringene testes mot modellenes middelflomnivå, som oftest er lavere enn det observerte nivået.

Hvordan treff, bom og falske alarmer telles opp kan illustreres med Figur 5.



Figur 5. Definisjonen av treff, bom, falsk alarm og suksessindeks for et varsel i forhold til observasjonen.

Når både observert og simulert/varslet vannføring overstiger den definerte terskelen, registreres et «treff». Når det observeres en vannføring over terskelen, men simulert vannføring ligger under, har vi en «bom» og når simulert vannføring overstiger terskelen, men observert ligger under, får vi en «falsk alarm». Treffraten er forholdet mellom antall treff og alle ganger det observeres en flom. Falsk alarmraten er forholdet mellom falske alarmer og antall ganger det simuleres en flom. Suksessindeksen sammenfatter modellens ytelse med hensyn til både treff, bom og falske alarmer, og beregnes som forholdet mellom antall treff og det totale antallet ganger enten observasjonen, simuleringen eller begge overstiger terskelen. Treffraten, falsk alarmraten og suksessindeksen varierer alle fra 0 (dårligst) til 1 (best).

I studien brukes det to forskjellige definisjoner av treff. Først defineres treff som treff samme dag, deretter inkluderes treff også den påfølgende dagen, dvs. modellen treffer dersom det er observert flom en dag og modellen simulerer flom enten samme eller påfølgende dag.

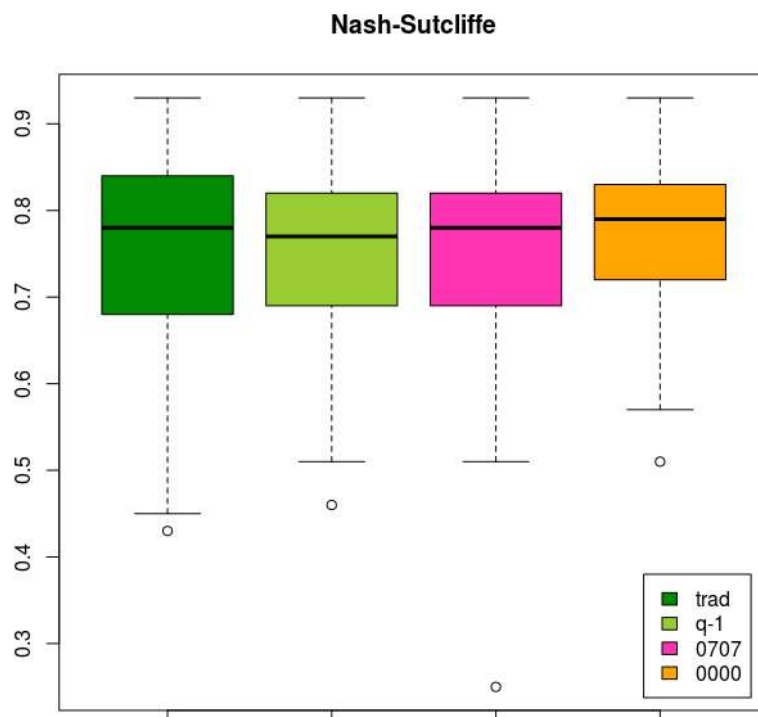
## 3 Resultater

### 3.1 Første runde: modellparametere tilpasset tradisjonell døgninndeling

#### 3.1.1 Nash-Sutcliffkoeffisienten

Figur 6 viser et boksplokk av NS-koeffisienten slik utfallet ble med de fire ulike datasettene anvendt i modellene som er kalibrert på data med tradisjonell døgninndeling. Det er ikke store forskjeller i hvordan de ulike datasettene presterer. Datasettet der vannføringsdøgnnet er en dag forskjøvet (q-1) skårer dårligst mht. medianen, mens det samkjørte døgnnet som går fra midnatt til midnatt (00-00) gir best median og minst spredning i verdiene, det vil si det beste resultatet. Det samkjørte døgnnet som går fra kl. 07 til kl. 07 har flere utslag i negativ retning, trolig på grunn av dårlige vannføringsdata.

Datasettet med tradisjonell døgninndeling, som modellene er kalibrert på, gir resultater midt på treet.



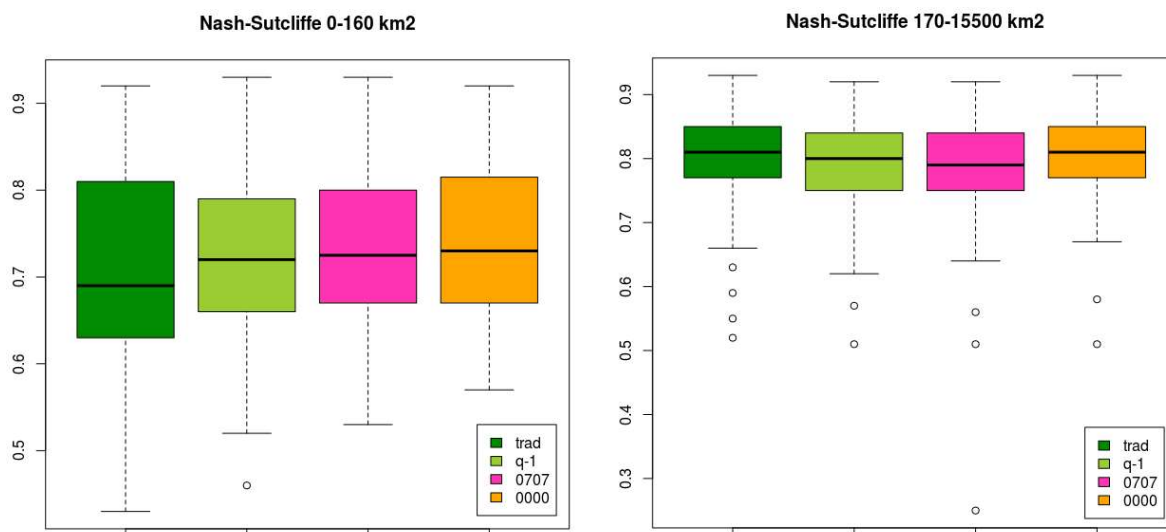
Figur 6. Boksploott som viser fordelingen av NS-verdier over de fire ulike datasettene.

	min	25%	median	75%	maks
trad	0.43	0.68	0.78	<b>0.84</b>	<b>0.93</b>
q-1	0.46	0.69	0.77	0.83	<b>0.93</b>
07-07	0.25	0.69	0.78	0.82	<b>0.93</b>
00-00	<b>0.51</b>	<b>0.72</b>	<b>0.79</b>	0.83	<b>0.93</b>

Tabell 3. Tabellen viser verdiene fra boksplottet i figur 6. Beste verdi i hver kolonne er uthevet.

I Figur 7 er resultatene for de små og de store modellfeltene presentert hver for seg. Generelt er det slik at NS-koeffisienten ligger høyere i store enn i små felt, noe som gjelder alle datasettene. Men det er en tydelig tendens at samkjøring av vær- og vannføringsdøgnet slår positivt ut i små felt, der vannføringen responderer raskere på nedbøren og timingen av nedbørhendelsen har større effekt på sammenfallet av observert og simulert vannføring. I de større modellfeltene blir effekten av å samkjøre døgnet mindre, betydningen av data og kalibrering med god kvalitet veier tyngre. Det tradisjonelle datasettet og samkjørt 00-00-døgn gir omtrent like bra resultater.





Figur 7. Boksploott som viser fordelingen av NS-verdier over de 4 ulike datasettene. Resultatene for den halvdel av feltene med minst areal (2 – 160 km<sup>2</sup>) er vist til venstre og halvdel med størst areal (170 – 15500 km<sup>2</sup>) er vist til høyre.

	min	25%	median	75%	maks
<i>små felt: 2 – 160 km<sup>2</sup></i>					
trad	0.43	0.63	0.69	0.81	0.92
q-1	0.46	0.66	0.72	0.79	<b>0.93</b>
07-07	0.53	<b>0.67</b>	0.725	0.80	<b>0.93</b>
00-00	<b>0.57</b>	<b>0.67</b>	<b>0.73</b>	<b>0.815</b>	0.92
<i>store felt: 170 – 15500 km<sup>2</sup></i>					
trad	<b>0.52</b>	<b>0.77</b>	<b>0.81</b>	<b>0.85</b>	<b>0.93</b>
q-1	0.51	0.75	0.80	0.84	0.92
07-07	0.25	0.75	0.79	0.84	0.92
00-00	0.51	<b>0.77</b>	<b>0.81</b>	<b>0.85</b>	<b>0.93</b>

Tabell 4. Tabellen viser verdiene fra boksplottet i figur 7. Beste verdi i hver kolonne er uthevet.

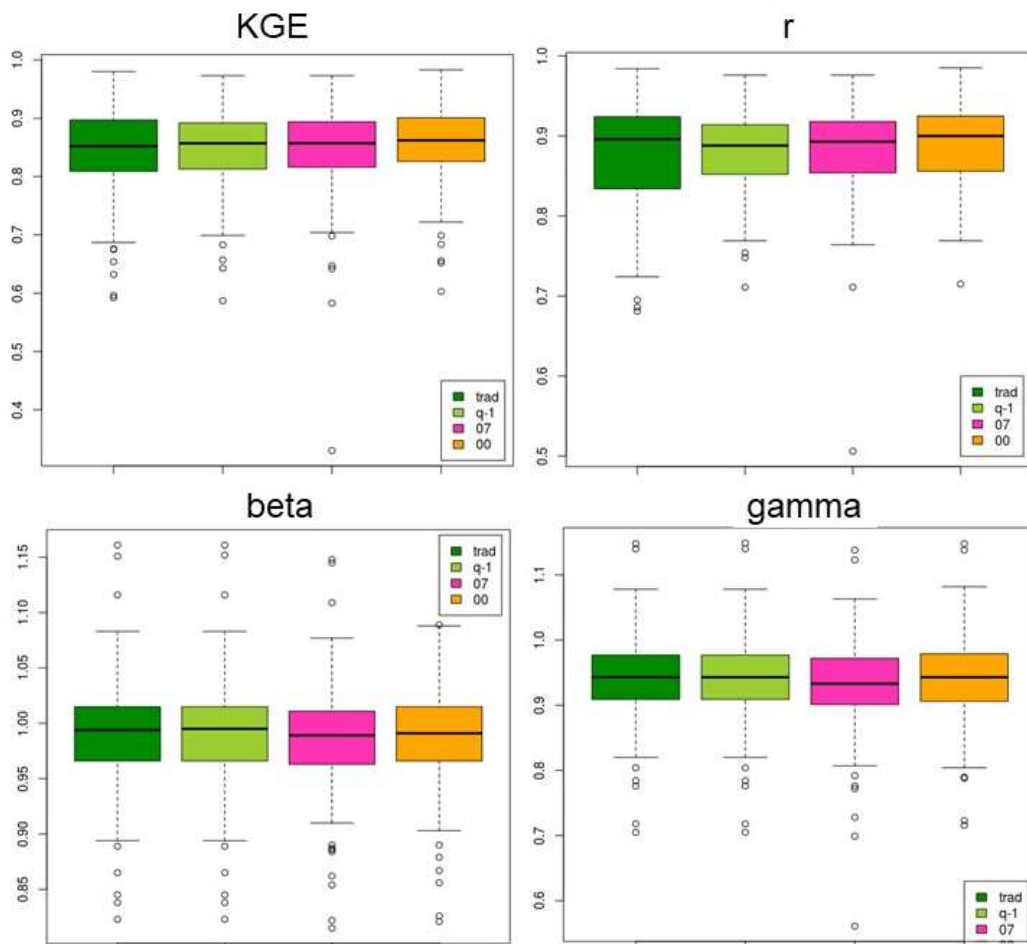
### 3.1.2 Kling-Guptakriteriet

Figur 8 og tabell 5 viser hvordan KGE og de tre komponentene til KGE fordeler seg over de fire datasettene. Som for NS-verdien, er det ikke veldig markerte forskjeller mellom datasettene, og også her har datasettet med samkjørte døgn 00-00 hårfint høyest median og minst spredning. Det samme gjelder samvariasjon og timing, som uttrykkes i r-verdiene. Datasettet der vannføringsdøgnet er en dag forskjøvet har en bedre fordeling av den nedre halvdel av dataene, som trolig kan forklares ved at de mindre feltene får et

bedre sammenfall mellom observert og simulert vannføring. Datasettet for 07-07-døgnet, der findata for vannføring er brukt, medfører en slenger i negativ retning som går igjen i flere av plottene. Denne kan trolig tilskrives at kvalitetskontrollen av findataene generelt er dårligere enn kontrollen av døgndata.

Komponenten  $\beta$  kvantifiserer tilpasningen av simuleringen med hensyn til volum. Vi ser at det tradisjonelle datasettet og datasettet der vannføringen er en dag forskjøvet har identiske bokplott. Det er rimelig siden datasettene er helt like, bortsett fra vannføringsforskyvningen. Findataene for vannføring i 07-07-døgnet gir en noe større middelværdi enn det tradisjonelle døgnet, noe som gir en noe mindre median og spredning i overkant for  $\beta$ . Nedbør- og temperaturdata for 00-00-døgnet gir noe mindre median i den simulerte vannføringen enn det tradisjonelle døgnet og mindre utslag i overkant.

Også når det gjelder  $\gamma$ , forholdet mellom variasjonskoeffisientene til henholdsvis simuleringer og observasjoner gir de to første datasettene igjen sammenfallende resultater. Resultatene for 00-00-døgnet sammenfaller også godt med disse. Findataene for vannføring i 07-07-døgnet gir noe større variasjonskoeffisient for de observerte dataene, trolig på grunn av generelt dårligere kvalitetskontroll, og vi ser en forskyvning i negativ retning for dette datasettet.



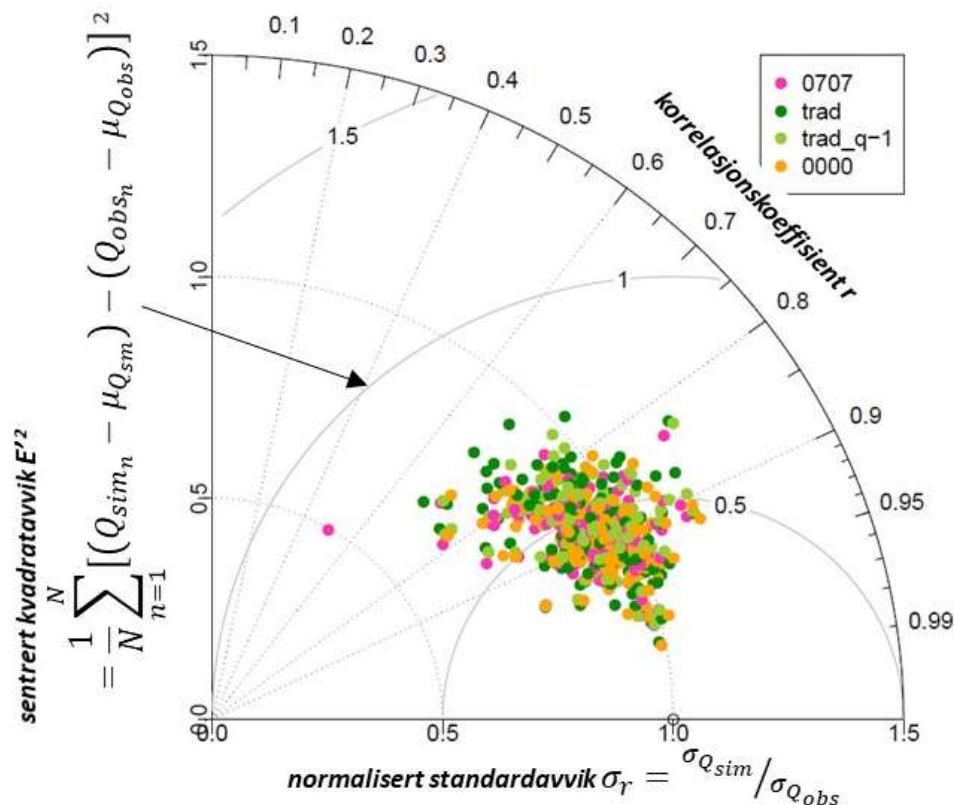
Figur 8. Bokplott som viser fordelingen av KGE og KGEs komponenter over de fire datasettene.

	min	25%	median	75%	maks
	<i>KGE (1.0 er optimalt)</i>				
trad	0.592	0.809	0.852	0.897	0.980
q-1	0.587	0.813	0.857	0.892	0.973
07-07	0.330	0.816	0.857	0.894	0.973
00-00	<b>0.603</b>	<b>0.826</b>	<b>0.862</b>	<b>0.901</b>	<b>0.983</b>
	<i>r (1.0 er optimalt)</i>				
trad	0.681	0.834	0.896	0.924	0.984
q-1	0.711	0.852	0.888	0.914	0.976
07-07	0.506	0.854	0.893	0.918	0.976
00-00	<b>0.715</b>	<b>0.856</b>	<b>0.900</b>	<b>0.925</b>	<b>0.985</b>
	<i>beta (1.0 er optimalt)</i>				
trad	<b>0.823</b>	<b>0.966</b>	<b>0.994</b>	1.015	1.161
q-1	<b>0.823</b>	<b>0.966</b>	<b>0.994</b>	1.015	1.161
07-07	0.815	0.963	0.989	<b>1.011</b>	1.148
00-00	0.821	<b>0.966</b>	0.991	1.015	<b>1.089</b>
	<i>gamma (1.0 er optimalt)</i>				
trad	0.705	<b>0.909</b>	<b>0.943</b>	0.977	1.148
q-1	0.705	<b>0.909</b>	<b>0.943</b>	0.977	1.149
07-07	0.561	0.901	0.933	0.972	<b>1.138</b>
00-00	<b>0.716</b>	0.906	<b>0.943</b>	<b>0.979</b>	1.148

Tabell 5. Tabellen viser verdiene fra boksplokkene i figur 8. Beste verdi i hver kolonne er uthevet.

### 3.1.3 Taylordiagrammet

I Figur 9 ser vi hvordan simuleringene med testens fire datasettene plasserer seg i et taylordiagram. Også i taylordiagrammet er tendensen at datasettet for 00-00-døgnet skiller seg ut i positiv retning, i og med at de oransje markørene er noe mer konsentrert rundt tyngdepunktet til punktsvermen, og dermed ser ut til å ha en noe mindre spredning, enn de andre markørene. Det betyr at dette datasettet samlet sett gir simuleringer som er relativt gode med hensyn til ekstremverdiutslag, kvadratavvik og samvariasjon/timing. Det tradisjonelle datasettet ser ut til å være dårligst på samvariasjon, og det er trolig småfeltene som gir de store utslagene i negativ retning. Forskyvningen av vannføringen gir en bedre plassering med hensyn til samvariasjon (lysegrønne markører). Slengeren fra datasettet til 07-07-døgnet dukker opp også i Taylordiagrammet, se rosa markør i periferien.



Figur 9. Taylordiagram som viser kvadratavviket, korrelasjonskoeffisienten og standardavviket for vannføring simulert med testens fire datasett i forhold til observert vannføring.

### 3.1.4 Treffrate/falsk alarm/suksessindeks-kriteriet

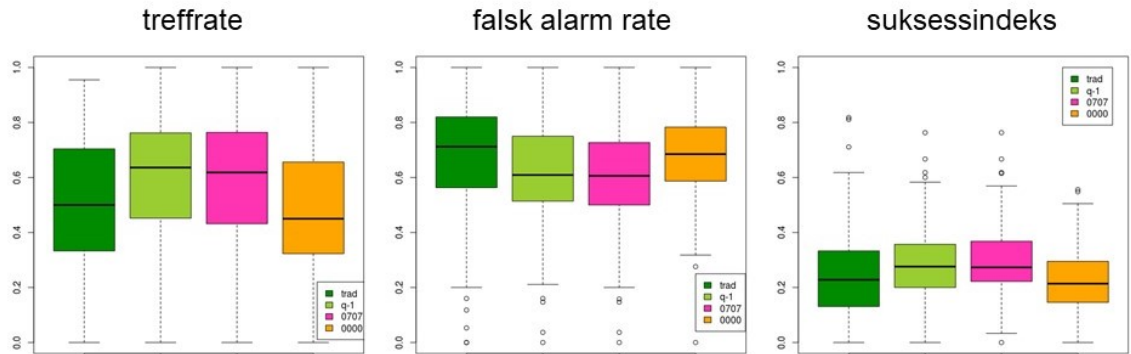
I figur 10 ser vi resultatene med hensyn til treffrate, falsk alarmrate og suksessindeks for de fire datasettene når modellene er kalibrert på data med tradisjonell døgninndeling. Selv om terskelen som er benyttet er laveste varslingsnivå, middelflom, så er resultatene preget av at det er få hendelser å evaluere resultatene på. Det kan gi store utslag, som f.eks. når det bare har vært én hendelse og modellen har bommet på denne, noe som gir en treffrate på 0. Dersom modellen traff, blir treffraten 1. Tilsvarende kan fort falsk alarmraten bli 1 eller i beste fall 0. Dette vil til en viss grad jevne seg ut og slå omtrent likt ut for de forskjellige datasettene.

I første omgang er det testet på at modell og simulering overlapper samme dag, uten slingringsmonn. Da kommer det tradisjonelle datasettet relativt dårlig ut, på grunn av forskyvningen i tid mellom vær- og vannføringsdata, noe som gjennomgående fører til at vannføringssimuleringene ligger en dag etter observasjonene. Flomtopper vil dermed ofte komme en dag forsinket, og kan gi en bom eller en falsk alarm.

Det enkle forskyvningen av døgnet som er gjort i datasettet med grønne bokser gir en stor forbedring i treffrate, falsk alarm og suksessindeks. Dette datasettet gir sammen med det samkjørte datasettet for 07-07-døgnet de beste resultatene. Responsfunksjonen i HBV-modellen har en tendens til å beregne en forsinket respons fra modellen i forhold til observasjonene, og denne ekstra forsinkelsen gjør at et litt forsinket vannføringsdøgn er gunstig for om den simulerte flomhendelsen treffer eller ei. En høy treffrate henger gjerne sammen med en lav falsk alarmrate, og datasettene med vannføringen en dag forskyvnet

og 07-07-datasettet kommer gunstigst ut også når det gjelder falske alarmer. Dermed får disse datasettene også de beste suksessindeksene.

Det er litt overraskende at 07-07-døgnet gir såpass mye bedre resultater enn 00-00-døgnet. Det kan være en forklaring at disaggregering og aggregering gir en viss utflating av værdataene som slår negativt ut på antall hendelser, noe som kan gi store utslag i noen felt, da dette er et datasett med få hendelser.



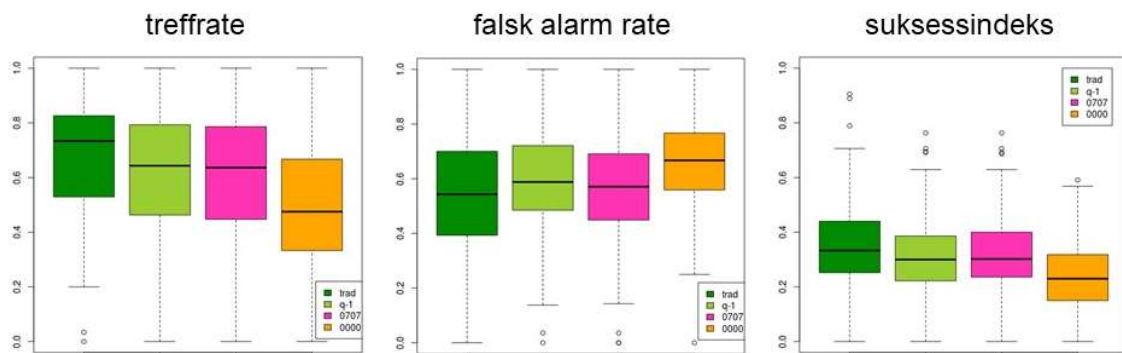
Figur 10. Bokplott som viser resultatene for treff- og falsk alarmrate og suksessindeks for de fire datasettene.

	min	25%	median	75%	maks
<i>treffrate (1 er optimalt)</i>					
trad	0.000	0.333	0.500	0.704	0.955
q-1	0.000	<b>0.452</b>	<b>0.636</b>	0.762	<b>1.000</b>
07-07	0.000	0.432	0.618	<b>0.764</b>	<b>1.000</b>
00-00	0.000	0.323	0.450	0.656	<b>1.000</b>
<i>falsk alarm rate (0.0 er optimalt)</i>					
trad	0.000	0.563	0.712	0.820	1.000
q-1	0.000	0.514	0.609	0.750	1.000
07-07	0.000	<b>0.500</b>	<b>0.606</b>	<b>0.727</b>	1.000
00-00	0.000	0.587	0.685	0.783	1.000
<i>suksessindeks (1.0 er optimalt)</i>					
trad	0.000	0.131	0.228	0.333	<b>0.818</b>
q-1	0.000	0.200	<b>0.276</b>	0.357	0.763
07-07	0.000	<b>0.222</b>	0.273	<b>0.368</b>	0.763
00-00	0.000	0.146	0.214	0.295	0.556

Tabell 6. Tabellen viser verdiene fra bokplottene i figur 10. Beste verdi i hver kolonne er uthevet.

Treff-, falsk alarm- og suksessindekskriteriet er også beregnet med innlagt slingringsmonn. Det vil si at en simulert flomtopp en dag etter en flomobservasjon

aksepteres som treff. Det tas altså høyde for tendensen modellen har til å ligge en dag på etterskudd med simuleringen. I varslingsammenheng gir en simulering som ligger litt etter i tid også en god indikasjon på om det vil bli flom eller ikke. En flomvarsler som er kjent med modellens tendens til forsinket responssimulering vil ta høyde for dette i modelltolkningen. Tilsvarende er det gitt et slingringsmonn når det gjelder falske alarmer: ved simulert flomnivå regnes det ikke som falsk alarm dersom observasjonen er oppe i flomnivå dagen før, selv om observasjonen er under flomnivå samme dag. Figur 11 viser resultatene av beregningene. Resultatene forbedres for alle datasett, men forbedringen er mye større for det tradisjonelle datasettet enn for de andre datasettene. Det er fordi det er spesielt mye å vinne på å legge inn et slingringsmonn for dette datasettet, som har minst sammenfall mellom vær og vannføring. Det tradisjonelle datasettet får best resultater av de fire datasettene når en dags slingringsmonn aksepteres. Slingringsmonnet tar bare høyde for forsinkelse av simuleringen i forhold til observasjonen. For de samkjørte datasettene kan det hende at simuleringen er raskere ute enn observasjonen, ettersom når på døgnet nedbøren faller. Det tas ikke høyde for manglende sammenfall den veien, noe som kan være med å gi en fordel for det tradisjonelle døgnet. I tillegg har det tradisjonelle døgnet den fordel at modellene er kalibrert på dette datasettet.



Figur 11. Boksplott som viser resultatene for treff- og falsk alarmrate og suksessindeks for de fire datasettene. En dags slingringsmonn aksepteres ved beregning av kriteriene.

	min	25%	median	75%	maks
<i>treffrate (1 er optimalt)</i>					
trad	0.000	0.529	<b>0.733</b>	0.826	1.000
q-1	0.000	0.463	0.643	0.793	1.000
07-07	0.000	0.447	0.636	0.786	1.000
00-00	0.000	0.333	0.475	0.667	1.000
<i>falsk alarm rate (0.0 er optimalt)</i>					
trad	0.000	0.393	<b>0.543</b>	0.700	1.000
q-1	0.000	0.485	0.588	0.721	1.000
07-07	0.000	0.449	0.571	0.691	1.000
00-00	0.000	0.559	0.667	0.767	1.000

	<i>suksessindeks (1.0 er optimalt)</i>				
trad	0.000	0.252	<b>0.333</b>	0.440	0.905
q-1	0.000	0.222	0.300	0.386	0.763
07-07	0.000	0.236	0.302	0.400	0.763
00-00	0.000	0.150	0.230	0.318	0.591

Tabell 7. Tabellen viser verdiene fra boksplokkene i figur 11. Beste verdi i hver kolonne er uthevet.

### 3.2 Andre runde: modellparametere tilpasset hvert enkelt datasett

I andre runde ble modellene kalibrert på hvert enkelt datasett før analysen ble gjort og resultatene sammenliknet. Her ble inndata fra seNorge-v2 brukt, og inndataene ble lest inn høydesonevis. Resultatene er beregnet for valideringsperioden 01.09.1985 – 31.08.2000. For 07-07-datasettet gir dette en kortere valideringsperiode for mange felt enn i første runde av testen, der resultatene ble beregnet over perioden 01.01.1995 – 31.12.2014.

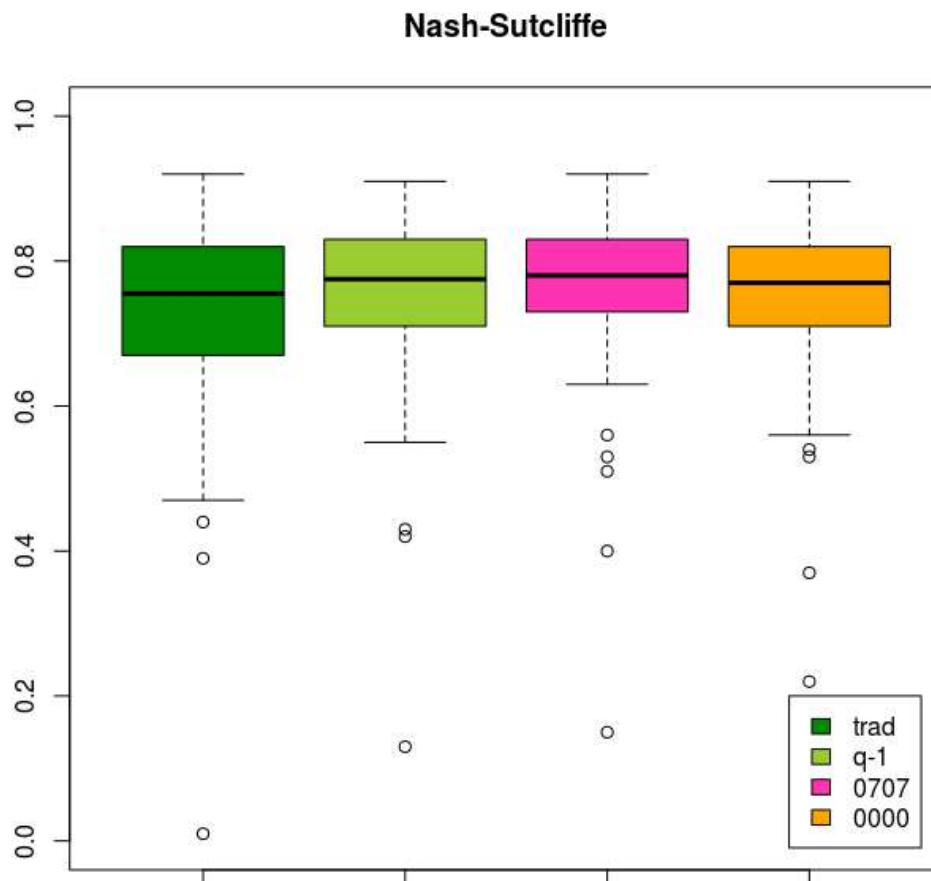
Nash-Sutcliffkoeffisienten

Figur 12 viser NS-koeffisienten for de fire datasettene. Kalibrering ga en betydelig relativ forbedring i resultatet for 07-07-døgnet, som i denne runden er det datasettet som gir best

	<b>min</b>	<b>25%</b>	<b>median</b>	<b>75%</b>	<b>Maks</b>
trad	-0.55	0.67	0.76	0.82	<b>0.92</b>
q-1	-0.45	0.71	<b>0.78</b>	<b>0.83</b>	0.91
07-07	-0.37	<b>0.73</b>	<b>0.78</b>	0.83	0.92
00-00	-0.23	0.71	0.77	0.82	0.91

resultat. Medianen for valideringsperioden er den samme som i første runde, men tyngdepunktet i datasettet er hevet. Valideringsperioden er ikke overlappende med perioden som er brukt i første runde, og vannføringsdatasettene er ikke fullstendige for alle felt. Korte serier kan gi utslag i positiv retning i enkeltfelt, men en tilsvarende tendens ses i kalibreringsperioden (ikke vist her), noe som støtter opp under resultatet. Alle datasett har en del slengere i negativ retning, noe som er utslag av dårlige data i enkeltfelt.

Alt i alt er forskjellene mellom datasettene små. Nå som alle datasett kjøres i modeller som er kalibrert til det gjeldende datasettet skårer det tradisjonelle datasettet hacket dårligere enn de mer samkjørte dataene, trolig på grunn av ikke-synkrone vær- og vannføringsdata.



Figur 12. Boksploott som viser fordelingen av NS-verdier over de fire ulike datasettene. Modellene er kalibrert på hvert enkelt datasett. Figuren viser resultatene for valideringsperioden.

	min	25%	median	75%	Maks
trad	-0.55	0.67	0.76	0.82	<b>0.92</b>
q-1	-0.45	0.71	<b>0.78</b>	<b>0.83</b>	0.91
07-07	-0.37	<b>0.73</b>	<b>0.78</b>	0.83	0.92
00-00	-0.23	0.71	0.77	0.82	0.91

Tabell 8. Tabellen viser verdiene fra boksplottet i figur 12. Beste verdi i hver kolonne er uthevet.

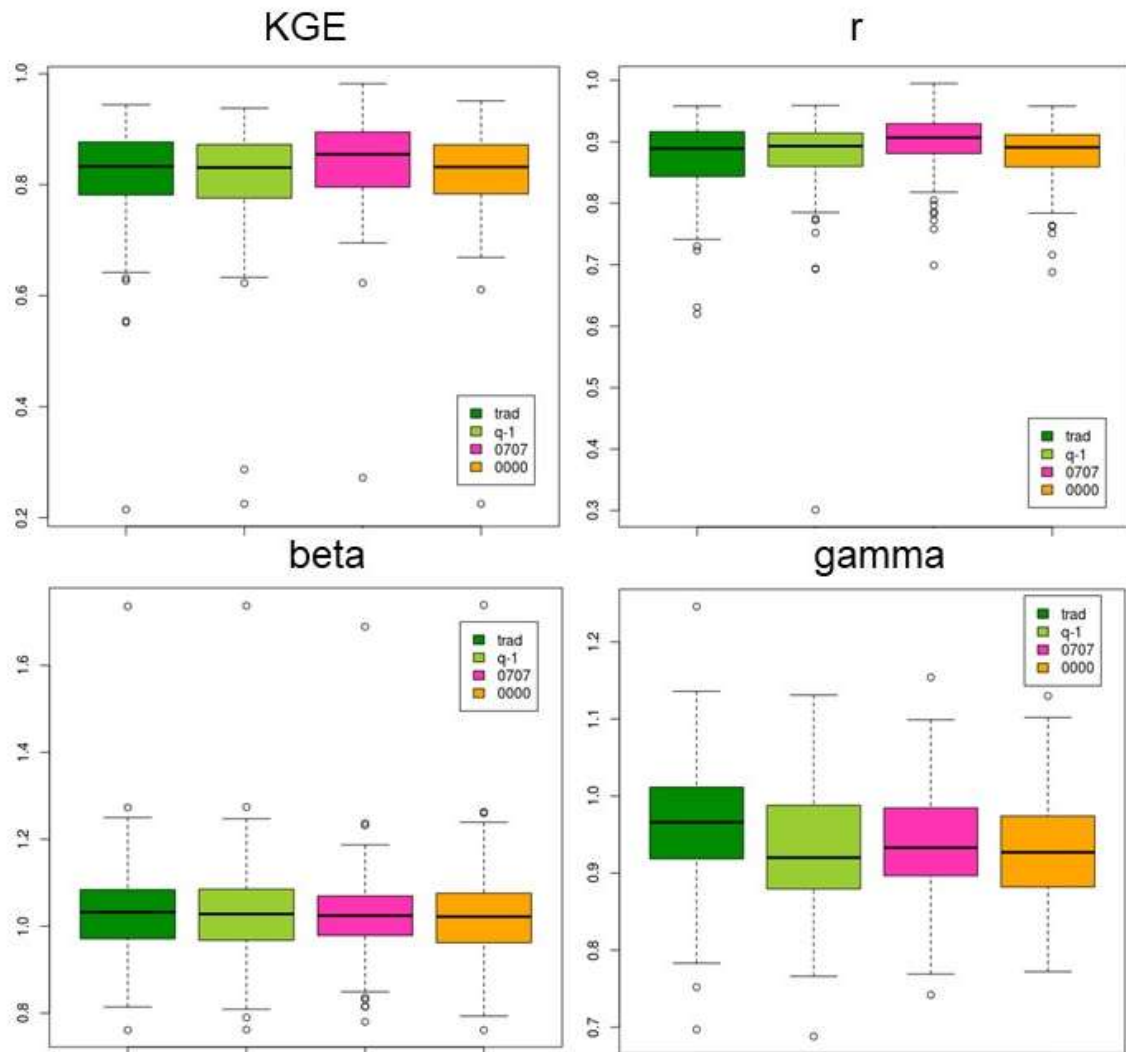
### 3.2.1 Kling-Guptakriteriet

Også når det gjelder KGE kommer de samkjørte datasettene best ut i andre runde, og det er 07-07-datasettet som skårer høyest. Det er også her en del negative slengere, som skyldes dårlige data i enkeltfelt.

Resultatene for r viser at de samkjørte datasettene har best samvariasjon. Når det gjelder beta ligger de fire datasettene veldig jevnt. Det tradisjonelle datasettet og datasettet der vannføringen er forskjøvet en dag gir ikke lenger helt identiske resultater for beta og



gamma, siden modellene nå kjører på forskjellige kalibreringer. Det tradisjonelle datasettet gir resultater med best sammenfall i variasjonskoeffisienten.



Figur 13. Bokplott som viser fordelingen av KGE-koeffisienten over de fire datasettene. Modellene er kalibrert på hvert enkelt datasett. Figuren viser resultatet for valideringsperioden.

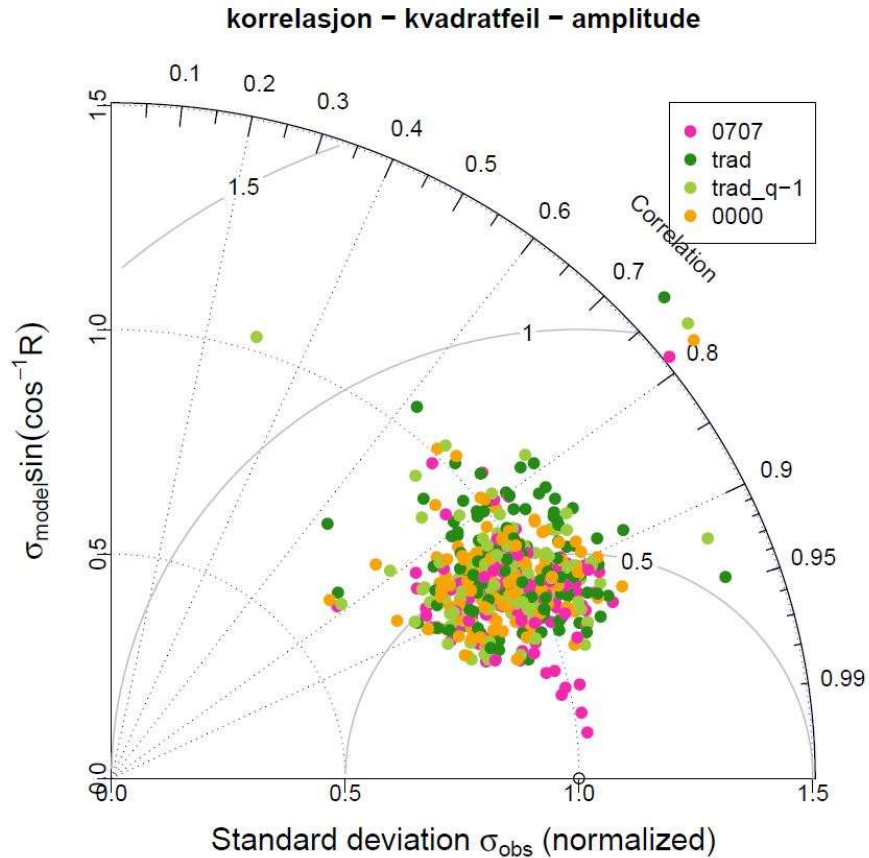
	min	25%	median	75%	maks
	<i>KGE (1.0 er optimalt)</i>				
trad	0.215	0.785	0.833	0.877	0.944
q-1	0.225	0.776	0.831	0.873	0.938
07-07	<b>0.272</b>	<b>0.796</b>	<b>0.855</b>	<b>0.895</b>	<b>0.982</b>
00-00	0.225	0.784	0.832	0.872	0.951
	<i>r (1.0 er optimalt)</i>				
trad	0.620	0.844	0.889	0.916	0.958

q-1	0.301	0.861	0.895	0.914	0.959
07-07	<b>0.699</b>	<b>0.881</b>	<b>0.907</b>	<b>0.932</b>	1.00
00-00	0.688	0.859	0.892	0.912	0.958
	<i>beta (1.0 er optimalt)</i>				
trad	0.761	0.971	1.032	1.084	1.736
q-1	0.762	0.966	1.028	1.083	1.737
07-07	0.780	0.979	1.024	1.069	1.689
00-00	0.761	0.961	1.022	1.074	1.739
	<i>gamma (1.0 er optimalt)</i>				
trad	0.697	0.919	0.967	1.012	1.246
q-1	0.688	0.881	0.920	0.989	1.131
07-07	0.742	0.898	0.935	0.988	1.154
00-00	0.772	0.882	0.928	0.977	1.130

Tabell 9. Tabellen viser verdiene fra boksplokkene i figur 13. Beste verdi i hver kolonne er uthevet.

### 3.2.2 Taylordiagrammet

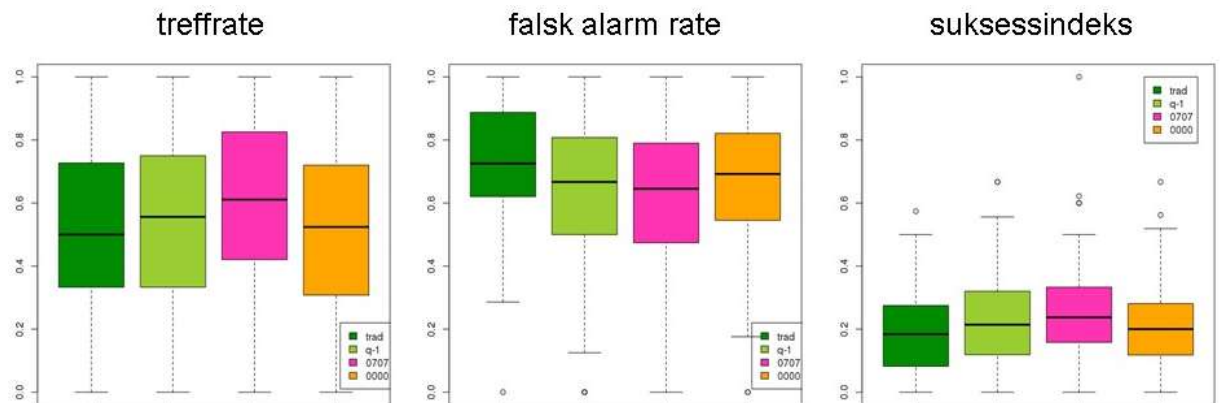
Taylordiagrammet viser liknende egenskaper som KGE i en grafisk form. Vi ser hvordan de rosa prikkene, 07-07-dataene ser ut til å ha en noe bedre plassering sentralt i datasvermen enn de andre datasettene. Det er også noen av de rosa punktene som ligger markert nærmere det optimale punktet enn de øvrige punktene. Dette sammenfaller godt med 07-07-settets gode skår når det gjelder KGE. 07-07-settet har trolig gitt relativt gode modeller, men det kan også spille inn at dette datasettet har noen relativt korte dataserier i valideringsperioden. Dette kan gi positive utslag i enkelte felt. Taylordiagrammet viser også at noen få felt gir svært dårlige simuleringer, med markeringer langt fra det optimale punktet.



Figur 14. Taylordiagram som viser kvadratavviket, korrelasjonskoeffisienten og standardavviket for vannføringen simulert med testens fire datasett i runde 2 i forhold til observert vannføring

### 3.2.3 Treff/falsk alarm/suksessindeks-kriteriet

Hvordan simulering og observasjoner sammenfaller i tid er viktig for utfallet mht. treff og falske alarmer. Derfor ser vi en tydelig effekt av å samkjøre døgnet når det gjelder disse kriteriene, se Figur 15. 07-07-datasettet kommer bra ut også når det gjelder dette kriteriet, både med hensyn til treffrate, falsk alarm og suksessindeks. Datasettet der vannføringen er forskjøvet en hel dag gjør det omtrent jamgodt med 00-00-satasettet.

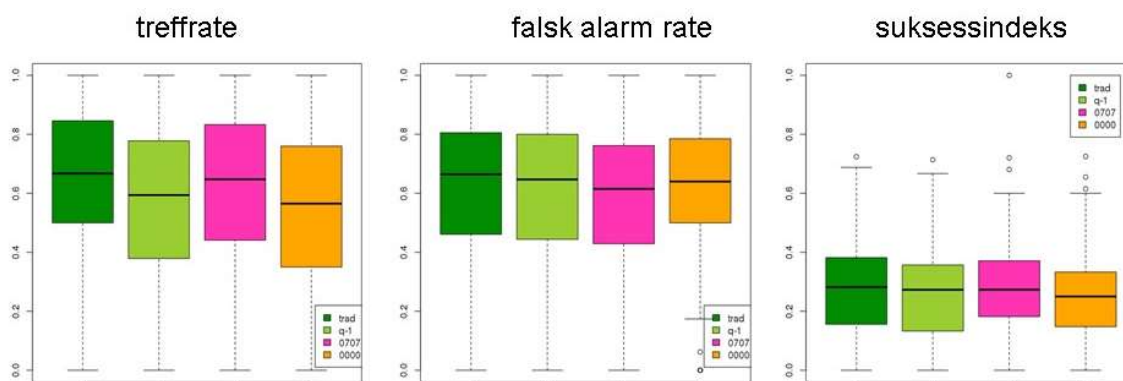


Figur 15. Boksplott som viser resultatene for treff- og falsk alarmrate og suksessindeks for de fire datasettene. Resultatene for valideringsperioden er vist.

	min	25%	median	75%	maks
<i>treffrate (1 er optimalt)</i>					
trad	0.000	0.333	0.500	0.727	1.000
q-1	0.000	0.333	0.556	0.750	1.000
07-07	0.000	<b>0.421</b>	<b>0.611</b>	<b>0.826</b>	1.000
00-00	0.000	0.323	0.475	0.667	1.000
<i>falsk alarm rate (0.0 er optimalt)</i>					
trad	0.000	0.621	0.726	0.888	1.000
q-1	0.000	0.500	0.667	0.808	1.000
07-07	0.000	<b>0.474</b>	<b>0.645</b>	<b>0.790</b>	1.000
00-00	0.000	0.559	0.668	0.810	1.000
<i>suksessindeks (1.0 er optimalt)</i>					
trad	0.000	0.082	0.184	0.275	0.574
q-1	0.000	0.119	0.214	0.320	0.667
07-07	0.000	<b>0.158</b>	<b>0.237</b>	<b>0.333</b>	1.000
00-00	0.000	0.110	0.200	0.280	0.630

Tabell 10. Tabellen viser verdiene fra boksploottene i figur 15. Beste verdi i hver kolonne er uthevet.

Det er også gjort en opptelling av treff og falske alarmer der det er gitt rom for en dags forsinkelse av den simulerte flommen i forhold til den observerte. Det vil si at det regnes som treff dersom det simuleres flom samme dag eller dagen etter. Figur 16 viser resultatene med dette slingringsmonnet. Det tradisjonelle datasettet forbedrer treffraten betraktelig, men ligger fortsatt litt høyere når det gjelder falske alarmer. De fire datasettene skårer ganske likt mht. suksessindeks.



Figur 16. Boksploott som viser resultatene for treff- og falsk alarmrate og suksessindeks for de fire datasettene. Det er gitt et slingringsmonn på en dag på definisjonen av sammenfall. Resultatene for valideringsperioden er vist.

	min	25%	median	75%	maks
	<i>treffrate (1 er optimalt)</i>				
trad	0.000	<b>0.500</b>	<b>0.667</b>	<b>0.846</b>	1.000
q-1	0.000	0.379	0.594	0.778	1.000
07-07	0.000	0.441	0.648	0.833	1.000
00-00	0.000	0.350	0.565	0.760	1.000
	<i>falsk alarm rate (0.0 er optimalt)</i>				
trad	0.000	0.461	0.665	0.806	1.000
q-1	0.000	0.444	0.647	0.800	1.000
07-07	0.000	<b>0.429</b>	<b>0.615</b>	<b>0.762</b>	1.000
00-00	0.000	0.500	0.640	0.785	1.000
	<i>suksessindeks (1.0 er optimalt)</i>				
trad	0.000	0.156	<b>0.282</b>	<b>0.382</b>	0.724
q-1	0.000	0.133	0.273	0.357	0.714
07-07	0.000	<b>0.182</b>	0.273	0.371	1.000
00-00	0.000	0.148	0.250	0.333	0.725

Tabell 11. Tabellen viser verdiene fra boksploene i figur 16. Beste verdi i hver kolonne er uthevet.

## 4 Drøfting og konklusjon

Datasettene der vær- og vannføringsdøgnet er sammenfallende fungerer bra, både som inndata til modeller som er kalibrert på data med den tradisjonelle døgninndelingen og som grunnlagsdata for kalibrering og simulering. Findataene som benyttes for å lage de samkjørte datasettene er beheftet med feil og mangler, noe som reduserer datakvaliteten på datasettene, men denne ulempen veier ikke tyngre enn fordelene ved å bruke data som er samkjørt i tid. Det er jamt over ikke veldig stor forskjell på hvordan modellene presterer med de ulike datasettene. Rangeringen av datasettene er litt forskjellig i runde 1 (modellene er kalibrert på data med tradisjonell døgninndeling) og runde 2 (modellene er kalibrert på de enkelte datasettene). I tillegg til at kalibreringen er forskjellig, skiller de to rundene seg fra hverandre ved at ulike tidsperioder og ulike versjoner av seNorge benyttes. De to rundene er dermed ikke helt sammenliknbare, men det synes å være en tydelig tendens til at det gir noe bedre resultater at modellparameterne er tilpasset datasettet det kjøres på.

00-00-datasettet skårer hårfint best med hensyn til NS, KGE og taylordiagrammet i første runde. Når det gjelder suksessindeksen, som måler treff og falske alarmer i forhold til en terskel, er det 07-07-datasettet som skårer høyest, med sterk konkurranse fra det tradisjonelle datasettet når en dags slingringsmonn tillates. Effekten av å samkjøre døgnet

er markert størst i små nedbørfelt. I andre runde scorer 07-07-datasettet best med hensyn til alle kvalitetskriterier. Begge datasettene viser at samkjørte data gir vannføringssimuleringer med best timing av flomtoppen. Det er ikke mulig å gi en helt entydig anbefaling når det gjelder hvilket av de samkjørte datasettene som bør brukes. Begge datasettene ventes å gjennomgå en kvalitetsheving med tiden, ettersom lengdene på findatatidsseriene for vannføring øker og kvaliteten på dataene forventes å øke med nye kvalitetsløft, og datagrunnlaget for nedbør- og temperaturdata blir bedre med flere automatstasjoner med timesoppløsning, noe som kan erstatte de disaggregerte dataene ved sammensetting av 00-00-døgnet.

Samkjørte data for vær og vannføring er intuitivt enklere å forholde seg til for de som skal kjøre og tolke vannføringsmodellresultater. Når dataene i tillegg gir minst like gode og til dels bedre resultater enn det tradisjonelle datasettet, spesielt i de mindre feltene, kan testen konkluderes med en klar anbefaling om å bruke samkjørte data, enten for 00-00- eller for 07-07-døgnet i den videre driften av flomvarslingsmodellene.

## 5 Referanser

Hoshin V. Gupta, H. Kling, K. K. Yilmaz and G. F. Martinez. 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, Volume 377, Issues 1-2, 20 October 2009, Pages 80-91. DOI: 10.1016/j.jhydrol.2009.08.003. ISSN 0022-1694

Gusong Ruan og Langsholt, E. 2017. Rekalibrering av flomvarslingas HBV-modeller med inndata fra seNorge, versjon 2.0. NVE Rapport 2017:71

Ingjerd Haddeland (red). 2014. Evaluering av flomvarslingas modellverktøy. NVE Rapport 2014:84

Harald Kling, M. Fuchs and M. Paulin. 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, Volumes 424-425, 6 March 2012, Pages 264-277, DOI:10.1016/j.jhydrol.2012.01.011

PEST. Model-Independent Parameter Estimation & Uncertainty Analysis ([www.pesthomepage.com](http://www.pesthomepage.com))

Klaus Vormoor and T. Skaugen. 2013. Temporal disaggregation of daily temperature and precipitation grid data for Norway. *Journal of Hydrometeorology* 14/2013. DOI: 10.1175/JHM-D-12-0139.1.



NVE

## Norges vassdrags- og energidirektorat

---

MIDDELTHUNSGATE 29  
POSTBOKS 5091 MAJORSTUEN  
0301 OSLO  
TELEFON: (+47) 22 95 95 95

[www.nve.no](http://www.nve.no)