

Kapitel 12 NØJAGTIGHEDSKRITERIER

Hans Jørgen Henriksen
Hydrologisk afdeling, GEUS

Torben Obel Sonnenborg
Geologisk Institut, Københavns Universitet

Nøglebegreber: Modelcyklus, performancekrav, testkørsel, kvalitative og kvantitative performance-mål, observations-datasæt, dokumentation af performance

ABSTRACT: Gennem modelcyklus i arbejdet med udbudsbetingelser, konceptuel model, modelopsætning, kalibrering og validering, udgør definerede krav til modellens nøjagtighed, dvs. performancekrav, et vigtigt instrument for design og analyse af en models nøjagtighed, herunder om modellen vurderes at kunne opfylde de formål, der er stillet for opgaven. Følgende spørgsmål er centrale for design og arbejde med formulering af nøjagtighedskriterier: Hvad skal modellen bruges til? Hvilket kompleksitetsniveau skal modellen designes ud fra? Hvilke ressourcer er der til rådighed og hvilke tidsplaner skal overholdes? I forbindelse med udbudsbetingelser skitseres kravene til modellens performance indledningsvist i form af såvel kvalitative som kvantitative mål. Forskellige typer rådata indgår i modelopstillingen og bidrager til tolkning af konceptuel model og valg af observations-datasæt til kalibrering og validering. Det er karakteristisk, at endelige krav til modellens performance først kan formuleres, når der foreligger en første testkørsel med modellen, hvor udvalgte observations-datasæt er inkorporerede. Herefter er opgaven i kalibreringsprocessen at tilpasse modellen (parametre, struktur, randbetingelser mv.), så slutmodellen overholder de opstillede performancekrav. Efterfølgende valideringstests kan så dokumentere, hvor god modellens performance er i forhold til uafhængige observations-datasæt. Gyldighed og relevans af dokumenteret performance og honorering (eller manglende honorering) skal afslutningsvis vurderes kritisk med henblik på den endelige evaluering af modellens gyldighedsområde og usikkerhed.

12.1 INDLEDNING

Fastlæggelse af nøjagtighedskriterier for et modelprojekt er en vigtig, men også vanskelig opgave. Modellen kan beskrive et antal variable f.eks. trykniveau, flow, koncentration osv., og man skal derfor tage stilling til, hvilken nøjagtighed hver enkelt variabel skal kunne beskrives med. Ofte fastlægger hydrogeologer et sæt kvantitative numeriske nøjagtighedskriterier (RMS, R^2 osv.) med henblik på at klassificere 'godheden' af en given model (Henriksen et al., 2003; Scholten and Van der Tol, 1998). Sådanne bestræbelser er nyttige med henblik på at gøre modellens performance gennemskuelig og kvantitativ, men de siger på den anden side ikke meget om, hvor god modellen egentlig er til at beskrive de formål, der er opstillet, f.eks. evnen til at beskrive grundvandsdannelsen eller vandbalanceforhold (Refsgaard et al., 2004). Undersøgelser med studerende der opstillede grundvandsmodeller fra USA har vist, at der ikke behøver at være nogen direkte sammenhæng mellem succes med hensyn til at matche trykniveauer i kalibreringen og efterfølgende evne til at simulere trykniveau ved ændrede stresspåvirkninger, Freyberg (1988). En god og nøjagtig kalibrering førte dermed ikke nødvendigvis til en god slutmodel eller til pålidelige simuleringer. Ikke fordi modellen ikke kunne matche trykniveauet, men fordi modellens troværdighed (goodness), var noget andet og mere end blot at kunne matche trykniveauet i en kalibreringssituation. Et godt fit mellem observeret og simuleret trykniveau og flow kan derfor højst siges at være et nødvendigt

krav, men ikke et tilstrækkeligt krav, til at bedømme en models kvalitet og nøjagtighed til simuleringsformål. Hvis den konceptuelle model er forkert, så er den samlede numeriske model nok også rimelig forkert, uanset hvor god en performance den så måtte opnå bedømt ud fra kvantitative kriterier. Det er altid relevant at skitsere mulige alternative konceptuelle modeller, og i arbejdet søge at eliminere disse med henblik på at indsnævre den mest troværdige model.

Det er heller ikke muligt at opstille et enkelt sæt universelle performancekrav, idet acceptkriterier for, hvornår en model er god nok, vil variere fra område til område. I takt med at der foretages en mere omfattende dataindsamling og tolkning i forbindelse med udarbejdelse af konceptuel model og modelopsætning opbygges efterhånden en nærmere forståelse for dels en nedre grænse for, hvad der vil kunne opnås med hensyn til nøjagtighed, som bl.a. hænger sammen med kvaliteten af observationsdata og usikkerheden på disse oversat til modelskala. Desuden kan der gives et bud på, hvilke nøjagtighedskrav der er nødvendige for at modellen bliver brugbar til den anvendelse, der er formuleret i udbudsbetingelserne og kravspecifikationer. Den øvre realistiske grænse kan herefter fastlægges som et mål for, hvad modellen skal kunne honorere i kalibrerings- og valideringskørsler. De realistiske performancekrav vil således være stærkt afhængige af de givne hydrogeologiske og hydrologiske forhold for området, heterogeniteten af de geologiske aflejringer, modellens diskretisering, og de processer der modelleres. Samtidig forudsættes en lidt dybere forståelse af kvaliteten af de indsamlede data, betydning af randbetingelser, inputdata, begrænsninger i valgte modelkode osv. før endelige krav til performance kan defineres.

Desuden vil kravene til nøjagtighed afhænge af den socio-økonomiske kontekst, altså hvad der er på spil i relation til et givent performance-niveau for modellen. Økonomi, interessenters præferencer, og risikovillighed spiller alt sammen en rolle for formuleringen af et sæt nøjagtighedskriterier for en grundvandsmodel (Refsgaard og Henriksen, 2004). Vores evne til på baggrund af data at foretage en modelbaseret nøjagtig beskrivelse af et system vokser som funktion af tid, penge, og den tilgængelige tekniske ekspertise (Middlemis, 2004). Derfor er det nødvendigt at moderere kravene til modellens nøjagtighed i forhold til de økonomiske rammer, tid og data der er til rådighed. Der kan ikke gives nogen endegyldig opskrift på, hvordan forskellige kvantitative og kvalitative krav bedst muligt kan balanceres og det afhænger desuden af, hvad modellen skal bruges til. Multiple kriterier, der fokuserer på både vandbalanceforhold, trykniveau og afstrømningsforhold, kan generelt anbefales, idet der herved er større chance for at opnå en troværdig model, som er i stand til at reproducere forskellige aspekter af systemets dynamik.

Det betyder at arbejdet med nøjagtighedskriterier i stedet for at kunne opfattes som et på forhånd givet universelt og objektivt kriterium, må opfattes mere som et resultat af refleksioner og faglige drøftelser undervejs f.eks. ved reviews, dvs. et produkt af en proces, hvor der gradvist i forløbet fastlægges et sæt endelige kriterier, som ikke kun er kvantitative (vandstand, flow, koncentrationer mm.), men også defineres med et kvalitativt indhold (realistiske parametre, dynamik, fordeling af residualer osv.). Som supplement til de kvantitative krav til nøjagtighed er det vigtigt at sikre, at variationer i rum og tid indenfor modelområdet har rimelig nøjagtighed. Dette sker i praksis i form af en formulering af et sæt kvalitative krav til modellen. Eksempler herpå kan være, at der fastsættes krav til de kalibrerede parameterverdier, så det sikres at parameterestimerne er realistiske, at modelstrukturen udgør en plausibel tolkning af geologien og at afvigelser mellem simulerede og målte variable er jævnt fordelt areal- eller dybdemæssigt. Det samme kan gælde tidsserier, hvor en kritisk vurdering af simulerede tidsserier i forhold til observerede altid er et krav. Hvis modellen skal bruges til partikelbanesimuleringer og afgrænsning af indvindingsområder er det særligt vigtigt at have øje for modellens evne til at simulere strømningsretninger, og kvalitative krav herom kunne derfor være naturlige.

Det er vigtigt at være opmærksom på, at der er nogle psykologiske mekanismer i aktion i rollespillet mellem vandressourceforvalter og hydrogeologer, og internt i hydrogeologgruppen. Tidspres og knaphed på ressourcer kan derfor forhindre nøgterne vurderinger af modellens performance. Den kultur (det rollespil) der i praksis styrer et modelprojekt, har derfor betydning for muligheden for kritisk at vurdere den konceptuelle model og designet af modelopsætningen og nøjagtighedskriterier. Eksterne reviews er her særligt nyttige, fordi udefrakommende eksperter vil kunne bidrage med forfriskende kritiske kommentarer til modellens resultater, performance osv., og dermed

forhindre alvorlige fejlfortolkninger af modellens gyldighedsområde og validitet. Der kan være en tendens til at hydrogeologer fokuserer for ensidigt på de kvantitative 'objektive facts' og dermed kommer til at ignorere det, som de ikke ved så meget om.

12.2 INDLEDENDE FASTLÆGGELSE AF NØJAGTIGHEDSKRITERIER I UDBUDSFASE

Den indledende fastlæggelse af nøjagtighedskriterier i udbudsfasen afhænger dels af (1) hvad modellen skal anvendes til (se kapitel 2 Formålet med opstilling af model) og dels af (2) den specificerede jobkompleksitet, altså hvilken detaljeringsgrad, der arbejdes med i modelprojektet. Der kan dels vælges et lavt niveau, dvs. en 'basic model'/screeningsmodel, et mellem niveau dvs. en 'intermediær model'/overslagsberegninger eller højt niveau dvs. en 'comprehensive model'/akvifer simulering.

Disse kravspecifikationer vil normalt være definerede som et led i en ydre planlægningsproces. Modelprojektet kan blot opfattes som et indre loop i denne cyklus. Hvor meget tid og hvor mange penge, der er til rådighed for modelprojektets udførelse, vil afhænge af nogle ydre socio-økonomiske faktorer. Fastlæggelsen af nøjagtighedskriterier for modellen bliver dermed primært et spørgsmål om at finde en rimelig balance mellem på den ene side et ønske om en så detaljeret og nøjagtig model som muligt, og på den anden side, det begrænsede antal ressourcer til at opnå dette mål, der i praksis er til rådighed. Det er særligt vigtigt at prioritere de spørgsmål, som modellen skal give svar på, så det er muligt at styre projektet hen mod de væsentligste behov for stor nøjagtighed. Man kan sige, at modellen 'vinkles' i forhold til hovedopgaven.

Forhold såsom konceptualiseringen af strømningssystemet, som i mange sammenhænge har særdeles stor betydning for modellens egnethed til at kunne belyse et opstillet hovedformål (USGS, 2004), kan måske overses, som følge af for ensidig fokusering på især de kvantitative nøjagtighedskriterier. Problemet er, at man her udelukkende fokuserer på den statistiske usikkerhed i stedet for at rette søgelyset mod det, som er sværere at kvantificere (scenariuusikkerhed, erkendt uvidenhed, se kapitel 19). I visse tilfælde vil det måske være mere optimalt at formulere en alternativ konceptuel model frem for at tvinge parametrene i den aktuelle konceptuelle model til at antage urealistiske værdier for at være i stand til at opfylde de kvantitative kalibreringskriterier. Pointen er derfor, at man ikke må lægge for meget i opfyldelsen af kvantitative nøjagtighedskriterier alene, hvis de kvalitative forhold ikke samtidig ser ud til at kunne opfyldes.

12.3 MODELOPSTILLING OG OBSERVATIONSDATA

Der er i princippet fire forhold som skal være opfyldte før den ofte tidskrævende kalibreringsproces kan sættes i værk:

1. Der skal være udvalgt et sæt kvalificerede (se afsnit 12.3.3) observationsdata for f.eks. trykniveau og afstrømning. Usikkerheden på observationsdata skal vurderes herunder hvilken maksimal nøjagtighed af værdier svarende til performancekriterier der kan opnås under hensyn til usikkerhed på observationsdata oversat til modelskala.
2. Der skal være udvalgt et sæt kvantitative performancekriterier for overensstemmelsen mellem f.eks. observeret og simuleret hydrauliske trykniveau
3. Som minimum bør et kvalitativt kriterium omkring realistiske parameterværdier defineres, det vil sige der bør produceres en tabel over startgæt på de enkelte frie parametre der kalibreres på og en forventet variationsramme (se tabel 12.6)
4. Der skal være foretaget en testkørsel (f.eks. stationær og dynamisk) for kalibreringsperioden med udtræk af værdier for performancekriterier. Det skal samtidig godtgøres, at der ikke er for meget numerisk støj på modelresultaterne. Det kan i visse tilfælde være nødvendigt at tjekke dette ved at anvende forskellige startgæt eller forskellige inputdata.

I forbindelse med opstillingen af den numeriske model er der en række forhold, der har betydning for den endelige fastlæggelse af nøjagtighedskriterier. Dels kan der fastsættes en nedre grænse for den performance, der vil kunne opnås (den kan jo ikke blive bedre end usikkerheden på data i forhold til modelskala), og dels skal det optimale mål for performance defineres ud fra modelfor-

mål, datakvalitet, skala osv. Indgangen til denne vurdering, som skal give input til kvantitative og kvalitative nøjagtighedskriterier, afhænger derfor af følgende forhold:

1. Kvaliteten af observationsdata (nøjagtighed og præcision af data)
2. Hydrogeologiske forhold (geologisk heterogenitet, usikkerhed på grundvandsdannelse, etc.)
3. Konceptuel model – eller hydrogeologisk tolkningsmodel (randbetingelser, geologisk model, anvendte simplificeringer, etc.)
4. Numerisk kode og modelopsætning (procesbeskrivelser og antagelser, fysiske data for vandløbssystem, grundvandsoppumpning, rumlig og tidlig opløsning af modelområde, m.m.)
5. Kalibrering (hvor godt er de optimale parameter, givet ovenstående ramme, estimeret via kalibreringen)

1) og 2) bestemmer den nedre grænse for, hvor god en overensstemmelse der kan opnås mellem observerede og simulerede variable. 3), 4) og 5) kan (vil) introducere fejl, der kan minimeres og er bl.a. en funktion af de økonomiske rammer for projektet. I det følgende vil der blive fokuseret på kvantificering af 1) kvaliteten af observationsdata. Effekten af hydrogeologisk heterogenitet er behandlet i kapitel 11 Skalaforhold og heterogenitet, mens forhold vedrørende den konceptuelle model, modelopsætning og kalibrering er beskrevet andre steder i håndbogen.

12.3.1 *Typer af observationsdata*

Observationsdata, som betegner målinger af systemets tilstand, f.eks. trykniveauet, strømningsretning eller koncentration, kan typisk karakteriseres ved geografi, frekvens og kvalitet. Geografi kan her betyde målinger i en boring (f.eks. trykniveau repræsenterende koordinater og dybdeinterval), et vandløbspunkt (f.eks. vandføring som egentlig er processeret ud fra vingeprofiler af målt hastighed og registrering af vandstand, der så er omregnet til vandføring ved hjælp af en simpel model, dvs. en QH-kurve) eller vandprøve (f.eks. oppumpet vandmængde fra boringsfilter), som så er analyseret på et laboratorium. Frekvens antyder, at observationsdata enten udgøres af tidsserier eller enkeltobservationer. Der er med andre ord sjældent tale om rådata, men oftest tale om processerede data, som tilmed er sekundære, fordi det ikke er data man selv har været med til at indsamle (primære data). Datakvaliteten og omstændigheder ved dataindsamlingen er derfor kun i begrænset omfang kendte. Kvalitet af data afhænger både af observationsnøjagtighed og i hvor høj grad målingen repræsenterer resultaterne fra modellen.

12.3.2 *Målinger af grundvandets trykniveau*

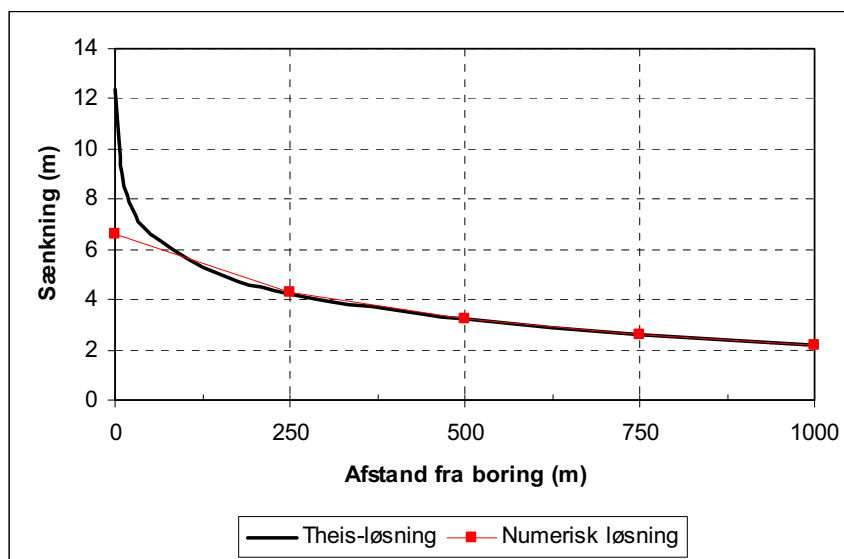
I arbejdet med strømningsmodeller vil det primært være målinger af hydraulisk trykniveau og vandløbsvandføring, der kan inkluderes i kalibreringsprocessen. I sjældne tilfælde vil der også være mulighed for at inddrage målinger af grundvandets alder samt grundvandets strømningsretning og -hastighed. I det følgende behandles udvælgelsen af kalibreringsdata, dvs. hvilke observationer af trykniveau og vandføring, der kan inkluderes i estimationen. Både den tidslige og rumlige fordeling af de tilgængelige observationsdata vil blive behandlet.

Overordnet kan observationer af grundvandets trykniveau inddeles i tidsserier og enkeltmålinger. Tidsserier har generelt højere værdi som kalibreringsgrundlag end enkeltmålinger, selv når der arbejdes med en stationær grundvandsmodel. Det skyldes dels, at den enkelte måling kan relateres til de øvrige målinger i tidsserien, hvilket gør det forholdsvis let at identificere fejlbehæftede målinger. Det kan være svært at udføre et lignende kvalitetscheck på en enkeltmåling, da man ikke har noget at sammenligne målingen med. Desuden kan der på basis af en tidsserie med stor nøjagtighed beregnes en karakteristisk værdi (middelværdi, 25%-fraktil, etc.), som den stationære model kan kalibreres mod. For en enkeltmåling kan det være svært at vurdere, hvor målingen ligger i forhold til f.eks. middeltrykniveauet, hvilket kan introducere fejl ved sammenligning mellem modelresultat og måling. Endelig kan det på baggrund af en tidsserie vurderes, om trykniveauet det pågældende sted har vist tegn på en trend, f.eks. faldende trykniveau pga. generelt stigende grundvandsindvinding. I givet fald udvælges kun målinger fra den periode, som modellen repræsenterer, som kalibreringsgrundlag.

Der foretages løbende målinger af grundvandets trykniveau, grundvandspejlinger, fra en lang række grundvandsboringer landet over, som hver især udgør en tidsserie af trykniveauets variation. GEUS drev fra 1950'erne et nationalt pejlestationsnet bestående af over 100 boringer. Nettet er nu overtaget af amterne. I forbindelse med Vandmiljøplanens grundvandsovervågningsprogram er der siden 1989 foretaget regelmæssige grundvandspejlinger i GRUMO og LOOP-områderne, som tilsammen udgør NOVANA pejlestationsnettet. Desuden er der flere amter, som driver eget pejlestationsnet. Der arbejdes i skrivende stund på at få alle de nævnte tidsserier lagt ind i Jupiter-databasen. I Jupiter findes også kortere eller længere grundvandspejleserier fra større hydrogeologiske projekter som f.eks. Karup-, Suså- og Nordvand-undersøgelserne.

Tidsserierne fra pejlestationsnettene udgør generelt den bedste kvalitet af trykniveaupejlinger, der findes i Danmark. Målingerne foretages som reglen i boringer med en veldefineret filterplacering, og i mange boringer er der installeret flere filtre, så vertikale strømningsforhold kan belyses. Det er tilstræbt at benytte observationsboringer, der er lokaliseret relativt langt fra større vandindvindinger, og som kan derfor antages at være forholdsvist upåvirket af sænkningstragten fra indvindingsboringer. Den tidlige opløsning vil typisk være på månedsniveau, men i nogle tilfælde kan der findes pejleserier, hvor der er målt hver 14. dag, mens der i andre tilfælde vil være udført færre pejlinger.

De større vandforsyninger driver ofte deres eget pejlestationsnet, hvor der foretages løbende observationer af trykniveauet i og omkring deres kildepladser. Men netop den korte afstand til indvindingsboringer gør, at disse målinger tit kan være svære at anvende i en numerisk model. Modellen simulerer sænkningen forårsaget af indvindingen i celler med et relativt stort volumen (f.eks. $250 \times 250 \times 10 \text{ m}^3$), mens målingen foretaget i en boring tilnærmelsesvis repræsenterer et punkt. Problemet er illustreret i figur 12.1, hvor en Theis-løsning er sammenlignet med løsningen fra en numerisk model med horisontal diskretisering på $250 \times 250 \text{ m}^2$. Begge modeller løser for sænkningen omkring en indvindingsboring i et 50 m tykt magasin med en hydraulisk ledningsevne på $1 \cdot 10^{-4} \text{ m/s}$ og et specifikt magasintal på $1 \cdot 10^{-4} \text{ m}^{-1}$. Indvindingsraten er på $4000 \text{ m}^3/\text{dag}$, og der ses på løsningen efter 100 dages oppumpning.



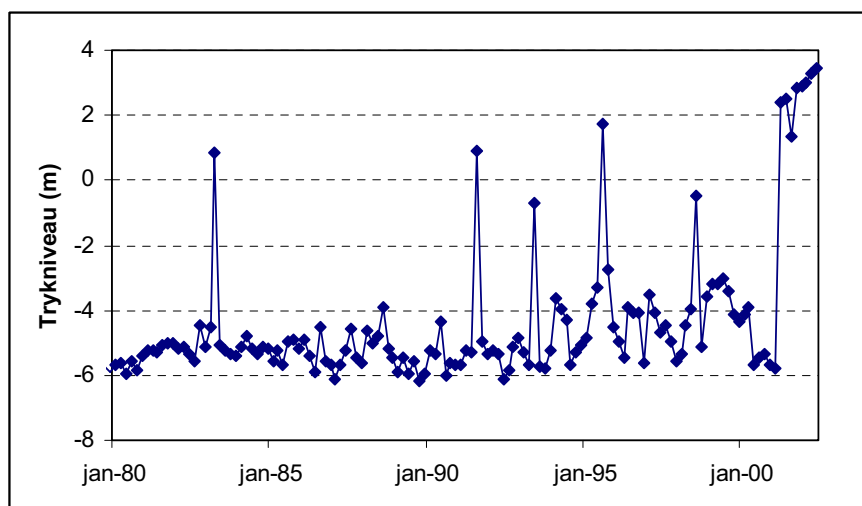
Figur 12.1 Sammenligning af sænkningstragt omkring en indvindingsboring fundet vha. hhv. en Theis-løsning og en numerisk løsning (MODFLOW), hvor der er anvendt en horisontal diskretisering på 250 m.

Fejlen på den numeriske løsning er på under én centimeter 750 m og længere væk fra indvindingsboringen, men stiger til hhv. 3 og 5 cm hhv. 500 m og 250 m fra boringen. I cellen, hvor boringen er placeret, simuleres en gennemsnitligt sænkning, som kan afvige signifikant fra en punktværdi. Theis-løsningen giver en sænkning på 12.4 m i en afstand af 1 m fra indvindingsboringen, hvilket er næsten en dobbelt så stor sænkning som den numeriske model simulerer.

Ovenstående eksempel viser, at observationer af trykniveauet tæt på indvindingsboringer skal anvendes med varsomhed. Det anbefales derfor, at afstanden til den nærmeste indvindingsboring altid bestemmes, og at det vurderes, hvor stor en sænkning der kan forventes i den givne afstand. Hvor egnet de pågældende målinger er som kalibreringsgrundlag for modellen afhænger bl.a. af, hvilken diskretisering der benyttes og dermed hvor godt høje gradienter i nærheden af indvindingsboringen kan reproducere.

Anvendelse af pejlinger fra selve pumpeboringen er endnu mere kritisk. Foruden diskretiseringsfejlen vil et eventuelt filtertab i boringen normalt ikke blive reproduceret i den numeriske model, hvilket vil forværre situationen yderligere. Det må derfor anbefales, at pejlinger fra pumpeboringer i drift anvendes med stor forsigtighed i forbindelse med modelkalibrering og helst helt undgås.

Observationer, der er foretaget tæt på indvindingsboringer, vil også være usikre pga. tidsvarierende indvindingsrater, som typisk ikke er repræsenteret i modellen, hvor der ofte benyttes årlige værdier for indvinding. En pejling udført på et tidspunkt med lav pumpe-rate vil derfor afvige kraftigt fra en pejling udført med høj pumpe-rate. Hvis opløseligheden af tidsserien er tilstrækkelig god, vil dette problem imidlertid kunne identificeres visuelt ved plotning af tidsserien, se eksemplet i figur 12.2.



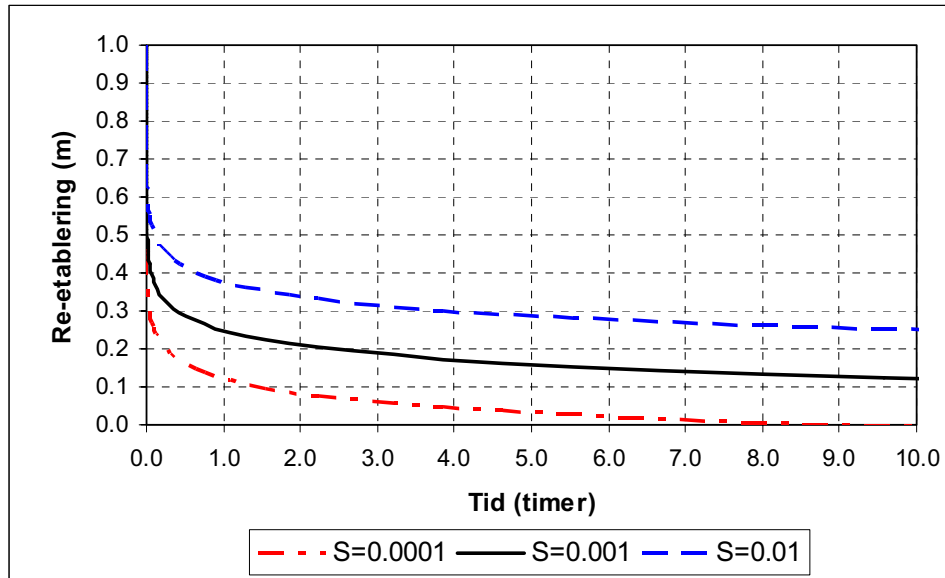
Figur 12.2 Pejlinger fra observationsboring placeret tæt på indvindingsboring med varierende indvindingsrate. I perioder, hvor der ikke pumpes, stiger trykniveauet med flere meter.

Enkeltpejlinger af trykniveauet foretages rutinemæssigt i forbindelse med udførslen af vandforsyningsboringer, geotekniske boringer, undersøgelsesboringer m.v. Dvs. for en stor del af de boringer, der er registreret i Jupiter, eksisterer der mindst en grundvandspejling. Ofte foretages en ekstra pejling i forbindelse med amtets lokalisering af boringen. Kvaliteten af målingerne afhænger i høj grad af, om boringen er nivelleret.

Da mange boringer i et givet område vil være af ældre dato, kan observationerne fra boringernes etablering sprede sig over en tidsmæssig lang periode, og det er derfor vigtigt at vurdere, om observationerne repræsenterer den situation, som modellen simulerer. Desuden vil det altid være vanskeligt at vurdere, hvor godt en enkeltmåling repræsenterer den sæsonmæssige tilstand (sommer, middel), som der simuleres.

Enkeltpejlinger af trykniveauet opnås typisk også ved gennemførelse af synkronpejlerunder, hvor tilgængelige boringer pejles indenfor et relativt kort tidsrum. I indvindingsboringer er det

praksis at stoppe pumpen i nogle timer, hvorefter pejling foretages. Det antages implicit ved denne fremgangsmåde, at det pejlede trykniveau repræsenterer rovandspejl. Hvor godt denne forudsætning er opfyldt afhænger af sænkning før pumpestop, transmissivitet T , og magasintal S , for det pågældende magasin. Figur 12.3 viser re-etableringen af trykniveauet i et 10 m tykt artesisisk magasin med $T = 1 \cdot 10^{-3} \text{ m}^2/\text{s}$, hvor der oppumpes $60 \text{ m}^3/\text{dag}$. Afsænkningen før pumpestop var på 1 m.



Figur 12.3 Sænkning som funktion af tiden efter pumpestop i et magasin, hvor sænkningen før pumpestop var på 1 m (beregnet vha. en Theis-løsning).

Det ses, at forløbet af re-etableringen er stærkt afhængig af magasintallet. Store magasintal resulterer i et system, der reagerer langsomt på ændringer i randbetingelser, mens små magasintal giver hurtige responstider. Figuren viser, at der i det viste eksempel let kan begås en signifikant fejl, hvis der pejles indenfor nogle få timer efter pumpestop. Der er derfor risiko for, at trykniveauet, der er fundet via pejlinger af denne type, vil være konsekvent for lavt i forhold til det "sande" trykniveau.

Hvis det antages, at indvindingsboringerne, hvor der er udført pejling under pumpestop, er indlagt i grundvandsmodellen, vil det introducere store fejl at anvende disse data i kalibreringen, da modellen i så fald vil simulere en sænkning svarende til den specificerede oppumpningsrate, men pejlingen tilnærmelsesvis repræsenterer et rovandspejl. Det må derfor anbefales ikke at anvende sådanne pejlinger til kalibrering af modellen.

I tabel 12.1 er pejlinger foretaget i forskellige borer og under forskellige betingelser kategoriseret og deres egnethed som kalibreringsgrundlag er vurderet.

Tabel 12.1 Kategorisering af grundvandspejlinger.

Pejlingstype	Egnethed som kalibreringsgrundlag
Tidsserie fra observationsboring	Velegnet til både stationær og dynamisk model. Indflydelse fra evt. nærliggende indvindingsboring bør vurderes
Tidsserie fra pumpeboring	Uegnet som kalibreringsgrundlag i numerisk model, da sænkningstragten omkring boringen vanskeligt lader sig reproducere tæt på boringen
Enkeltpejling ved etablering	Egnet som kalibreringsgrundlag for stationær model. Svært at vurdere eventuelle fejl. Svært at kvantificere ikke-stationære effekter præcist
Synkronmåling	Måling foretaget i pumpeboring med midlertidig indvindingsstop svær at anvende som kalibreringsgrundlag. Måling foretaget i observationsboring egnet, dog med samme forbehold som anført ovenfor (fejl og ikke-stationaritet)

12.3.3 Datakontrol og processering

Der bør foretages en kontrol af data med henblik på at vurdere konsistens. Herefter vurderes det, om udvalgte data skal forkastes og sådanne fravalg begrundes og dokumenteres. Der laves en oversigt over data, som har passeret denne indledende kvalitetssikring af processerede inddata. Disse data kan betegnes 'kvalificerede data'.

Kvaliteten og kvantiteten af kvalificerede data, der er egnede til modelkalibrering og validering skal vurderes. Det er specielt vigtigt, at der foreligger en rimelig rumlig fordeling af data. Hvis det vurderes, at der ikke er data nok, så bør behov for yderligere data beskrives og kommunikeres. For dynamisk modellering er det vigtigt, at der foreligger tilstrækkeligt lange tidsserier til både kalibrering og validering, typisk som minimum ca. 10 års data, og at disse data dækker samme periode.

Kontrollen af observationsdata vil i første omgang bestå i en analyse af det foreliggende datasæt for systematiske fejl. Trykniveaumålinger kan være påvirket af effekter, som ikke er inkluderet i den opstillede model. Eksempelvis kan visse observationer være influeret af prøvepumpningsforsøg, af indvindingsboringer som er nedlagt, m.m.

Når der anvendes en stationær grundvandsmodel, er udvælgelse af repræsentative data langt fra trivial. Hvis tidsserier af det hydrauliske trykniveau er til rådighed, skal der foretages en midling af trykniveauet, der afspejler, hvilken slags stationær model der er opstillet. Stationære tilstande vil kun i meget sjældne tilfælde optræde i virkelige grundvandsmagasiner, og det skal derfor vælges, hvilken situation, der ønskes en model for. Hvilke typer data der skal inddrages, og hvilken slags midling der skal udføres, afhænger af formålet med modelleringsarbejdet.

Vandføringsdata for grundvandszonen er ikke direkte tilgængelig, men der kan opnås et indirekte mål for grundvandets udveksling med vandløb, her benævnt grundvandstilstrømning, ved at inddrage værdier af vandløbenes sommervandføringer, hvor den mindste vandføring i vandløbet optræder. Hvis det ikke har regnet i længere tid forud for målingen, kan det for visse grundvands-systemer forsvareres at tolke denne vandføring udelukkende som et resultat af vandstrømning gennem den mættede zone til vandløbet (samt evt. spildevandsbidrag eller andre tilledninger). Dvs. bidrag fra overfladeafstrømning, interflow, drænvandsafstrømning samt udstrømning fra søer og vådområder forudsættes at være negligible i forhold til grundvandstilstrømningen.

Grundvandstilstrømningen varierer over året, og det er derfor vigtigt at gøre sig klart, at den målte minimumsvandføring repræsenterer den lave ende af grundvandstilstrømningens variationsområde. Hvis der arbejdes med en stationær model, der skal simulere sommerperioden med lille nettonedbør, vil det være hensigtsmæssigt at anvende medianminimum som er et udtryk for det sandsynlige minimumsflow. Derimod vil medianminimum ikke repræsentere den gennemsnitlige grundvandstilstrømning, og medianminimumsværdier kan derfor være farlige at anvende direkte i

en stationær model, hvor reservoirets gennemsnitlige forhold ønskes belyst (Sonnenborg et al., 2003).

Hvis det skal være muligt at estimere rumligt distribuerede parametre, er det vigtigt, at der er observationsdata til rådighed i hele det modellerede område. Ideelt set er det mest fordelagtigt, hvis kalibreringsdata fordeler sig jævnt indenfor området. I realiteten vil der altid være flere data til rådighed i nogle områder end andre. Det kan derfor være hensigtsmæssigt at eliminere nogle observationer i områder, hvor densiteten af data er stor. Herved undgås, at visse områder tillægges meget større vægt end andre, når kalibreringen gennemføres. Desuden vil informationsmængden pr. observation typisk være mindre i områder med høj koncentration af data.

Jo flere typer data, der anvendes, des større chance er der for at undgå problemer med manglende identificerbarhed og entydighed (non-uniqueness). Desuden kan inddragelsen af flere datatyper resultere i en reduktion af usikkerheden på de estimerede parametre (Hill, 1998).

12.4 USIKKERHED PÅ OBSERVATIONSDATA

Kvaliteten (usikkerheden) af kalibreringsdata skal vurderes, inden parameterestimationen påbegyndes. Formålet med at kvantificere usikkerheden på kalibreringsdata er for det første at opnå et mål for, hvor præcist modellen i bedste fald kan forventes at reproducere data, dels at opnå et objektivi kriterium for, hvorledes data af samme type skal vægtes indbyrdes og endnu vigtigere, hvordan data af forskellig type skal vægtes i forhold til hinanden.

Observationsdata vil altid være behæftet med usikkerhed (se kapitel 19). I nærværende sammenhæng vil faktorer, som resulterer i ikke systematiske afvigelser mellem målinger og modelresultater blive fortolket som usikkerhed på observationsdata. Målefejl er en af årsagerne til en del af denne usikkerhed, mens uoverensstemmelse mellem naturens kontinuerte variable og modellens diskrete variable er en anden årsag til afvigelser mellem observationer og modelprediktioner. Effekterne, som beskrives i dette afsnit, vil i middel være nul og introducerer derfor ikke nogen systematisk fejl i modelarbejdet. I det følgende vil usikkerheden på trykniveau- og vandføringsobservationer blive kvantificeret. Der sigtes på at bestemme en størrelsesorden for usikkerheden, idet en meget nøjagtig kvantificering kun undtagelsesvist kan lade sig gøre. Desuden er det ikke nødvendigt at fastlægge usikkerheden med speciel høj præcision for at kunne vurdere kvaliteten af observationerne.

12.4.1 Observationer af hydraulisk trykniveau

Trykniveauobservationer kan være behæftet med *målefejl*. Selve pejlingen i boringen kan være fejlbehæftet, idet både tryktransducere og vandspejlsmåleudstyr har en endelig præcision. Denne fejl vil under normale omstændigheder være relativ beskeden (få centimeter). Større fejl kan opstå pga. manuelle aflæsningsfejl eller tastefejl. Desuden kan det målte vandniveau i boringen afvige fra den sande værdi i magasinet pga. delvis tilstopning af eller omkring filtersætningen. Da barometertrykket sjældent indgår eksplicit i grundvandsmodeller, og atmosfæriske lufttryksændringer derfor ikke beskrives af modellen, kan barometereffekter også introducere en fejl på det måle trykniveau. Samlet vil målefejlene typisk resultere i en standardafvigelse på observationsværdien på 5 – 30 cm.

Koten, hvortil dybden til vandspejlet i boringen relateres (*målepunktskoten*), kan være behæftet med betydelig usikkerhed. Hvis terrænkoten er bestemt ud fra et topografisk kort med skalaen 1:25.000, vil intervallerne mellem konturlinierne være 2.5 m med en præcision på 0.5 m. Medtages boringens usikre placering på kortet, kan den samlede standardafvigelse på målepunktskoten være på 1 – 2 m. Hvis terrænkoten er bestemt vha. GPS vil usikkerheden typisk være af størrelsesorden centimeter.

Skalaeffekter medfører en yderligere usikkerhed på data. Skalaeffekter opstår, fordi der anvendes numeriske celler af endelig størrelse til beskrivelse af den kontinuerte fysiske virkelighed. Eksempelvis kan der være uoverensstemmelse mellem boringens filtersatte interval og den *vertikale diskretisering* i modellen. Både hvis filterintervallet er mindre end lagtykkelsen men specielt, hvis det er større end lagtykkelsen eller den vertikale cellestørrelse, vil der opstå fejl ved sammenligningen af observeret og simuleret trykniveau. Hvis boringen eksempelvis er filtersat over to

sandmagasiner adskilt af et lavpermeabelt lerlag, vil det være vanskeligt at relatere det målte trykniveau til resultater fra en model, der opløser den geologiske lagserie. Normalt vil filterets midtpunkt blive anvendt til at afgøre, hvilken celled trykniveau der skal repræsentere det observerede. Dette valg introducerer en fejl, som er svær at kvantificere i det generelle tilfælde. Fejlen vil bl.a. afhænge af faktorer som filterlængden, den vertikale diskretisering og den geologiske opbygning af reservoiret.

Observationsboringens horisontale placering kan være behæftet med usikkerhed pga. unøjagtig opmåling. Hvis placeringen er vurderet ud fra kort, kan denne fejl være betydelig. Desuden vil boringens placering kun sjældent være sammenfaldende med midtpunktet af en numerisk celle. Det kan derfor være nødvendigt at interpolere mellem nærliggende celler, for at opnå det bedst mulige estimat af trykniveauet ved boringen. Derved introduceres en *interpolationsfejl*, som vil afhænge af den anvendte horisontale diskretisering og gradienten på trykniveauet i området.

Topografiens variation indenfor de numeriske celler kan indirekte give anledning til afvigelser mellem observeret og simuleret trykniveau, da topografisk variation bl.a. kan påvirke fordelingen af nettonedbør og grundvandsdannelse indenfor en numerisk celle. Trykniveauet i specielt de terrænnære frie magasiner vil være følsomt overfor variationer i topografien. Standardafvigelsen på det hydrauliske trykniveau vurderes generelt at afhænge af topografiens variation, den anvendte diskretisering og det frie magasinets ledningsevne, men det er svært at kvantificere fejlen i det generelle tilfælde. Ved opsætning af DK-model Fyn (Henriksen et al., 1997), hvor der benyttes cellestørrelser på 1×1 km, blev der beregnet en standardafvigelse på 4.9 m på cellernes middellag. Der var derfor en betydelig usikkerhed på simuleret trykniveau i de øverste lag af modellen.

Den sidste skalafejl skyldes den *geologiske heterogenitet* indenfor de numeriske celler, som det i en deterministisk grundvandsmodel er umuligt at beskrive eksplicit, da hver celle skal tilskrives ét sæt hydrauliske parametre. Ifølge Gelhar (1986) er usikkerheden på trykniveauet, som skyldes ikke modelleret heterogenitet, en funktion af gradienten på det hydrauliske trykniveau, variansen på logtransformeret hydraulisk ledningsevne samt korrelationslængden for samme størrelse, se kapitel 11. Det kræver derfor kendskab til den rumlige variabilitet af den hydrauliske ledningsevne, som sjældent er tilgængelig i praksis, at kvantificere denne type fejl. Imidlertid vil det undertiden være muligt at give et kvalificeret skøn på den hydrauliske ledningsevnes geostatistiske egenskaber ved at inddrage erfaringsværdier fra sammenlignelige områder, se kapitel 11. Alternativt kan fejlen vurderes, hvis trykniveaudata fra tætstående boringer er til rådighed.

Tidsskalaeffekter kan være en fejlkilde, hvis der anvendes en stationær grundvandsmodel. Anvendelse af observationsdata, som repræsenterer ikke-stationære tilstande, vil ved brug af en stationær model resultere i afvigelser mellem observeret og simuleret trykniveau, som ikke kan elimineres. Hvis tidsserier af trykniveaumålinger er til rådighed, kan data analyseres, og en værdi, der repræsenterer den stationære tilstand, som ønskes beskrevet, kan beregnes. Herved kan den ikke-stationære fejl minimeres til et niveau, som er bestemt af tidsseriens længde og den anvendte analysemetode. I mange tilfælde vil der imidlertid kun være enkelte eller få målinger til rådighed fra de fleste af de installerede boringer, og i dette tilfælde vil det være svært at filtrere den ikke-stationære effekt fra. Da målingerne kan indeholde værdifuld information om trykniveaubilledet i området, vil de ofte blive anvendt i kalibreringen alligevel. Hvis trykniveauerne i de tilgængelige boringer er fremkommet ved måling på tilfældige tidspunkter af året, vil det imidlertid være rimeligt at antage, at målingerne i gennemsnit repræsenterer et middeltrykniveau. Pga. sæsonvariationerne vil de pågældende data være behæftet med en betydelig usikkerhed, som kan kvantificeres vha. tidsserier af trykniveaumålinger fra det aktuelle område. Herved kan et estimat for årstidsvariationerne og dermed usikkerheden på datapunkterne skønnes. For Esbjergmodellen, se appendiks A, blev trykniveaudata fra Jupiterdatabasen, som indeholder mange boringer med kun én eller få målinger, anvendt. Ud fra tidsserier fra området blev trykniveauet vurderet at variere med ca. ± 1 m, svarende til en standardafvigelse på omkring en halv meter.

I nedenstående tabel ses en oversigt over de ovenfor beskrevne usikkerheder, hvor standardafvigelsen er anvendt til kvantificeringen. Yderligere er det forsøgt at beregne de enkelte usikkerhedsbidrag for hhv. Esbjergmodellen og DK-model Fyn, i begge tilfælde for beregningslag 3 som er

domineret af sandformationer, samt for DK-model Sønderjylland og DK-model Sjælland. Variansen på log K er sat til hhv. 1 og 2 for Esbjerg og Fyn, og korrelationslængden, α , er for begge områder sat til 500 m.

Hvis fejlkilderne antages at være uafhængige, kan de enkelte bidrags varians (kvadratet på de tabulerede standardafvigelse) summeres, og den samlede standardafvigelse på observationsdata kan beregnes som kvadratroden af denne sum (vist i kolonnen længst til højre i tabel 12.2).

Tabel 12.2 Angivelse af standardafvigelse, s_{obs} (i m), på observationer af hydraulisk trykniveau. Δx betegner den horisontale diskretisering, og J er den hydrauliske gradient.

	Pejlefejl		Skalafejl		Ikke-stationaritet	Andre effekter ³⁾	Samlet usikkerhed
	Målefejl	Kote	Interpol.	Heterogen.			
Generelt	0.05–0.3	0 – 2	0.5 Δx J	$C^{1/2} \lambda s_{\ln K} J$ ¹⁾	$\Delta h_t/2$ ²⁾	0 – 1	$\sqrt{\Sigma s^2}$
Esbjerg ⁴⁾	0.1	1.5	0.5	1.0	0.5	0.25	2.0
Fyn	0.1	1.5	1.5	2.1	0.5	0.25	3.0
Sønderjylland ⁵⁾	0.5 - 1		1 - 2		<=		3.0
Sjælland ⁶⁾	0.1	1.5	1.5	2.1	0.5	0.5	3.1

- 1) $s_{\ln K}$ er standardafvigelsen på log K. λ er korrelationslængden for log K. C er en konstant hvis værdi afhænger af strømningssystemet, se tabel 11.7. Se i øvrigt kapitel 11 for yderligere forklaring.
- 2) Δh_t angiver amplituden på sæsonvariationerne i hydraulisk trykniveau.
- 3) Inkludere effekter som vertikal skalafejl, variationer i topografi og usikkerhed på grundvandsdannelsen..
- 4) Kilde: Harrar et al. (2003)
- 5) Kilde: Sonnenborg et al. (2003)
- 6) Kilde: Henriksen et al. (2003)

Usikkerheden beregnet i kolonnen længst til højre i tabel 12.2 repræsenterer den nedre grænse for, hvor præcist en model med horisontal diskretisering Δx kan forventes at reproducere de tilgængelige trykniveauobservationer. Eventuelle fejl i f.eks. den konceptuelle model, modelopsætningen, og parameterværdier vil forringe præcisionen af modelresultaterne udover dette niveau.

Boks 12.1 *Eksempel Lollandsmodel (Storstrøms Amt, 2003a).*

Det er vurderet at usikkerheden på pejlinger foretaget i nivellerede borer er omkring 1,25 m og ro-pejlinger foretaget i ikke nivellerede pejleboringer er omkring 2 meter. Disse værdier skønnes at være gældende på Lolland – hvor den gennemsnitlige fejl på observationsdata skønnes til 1,5 meter. Ideelt set er det mest fordelagtigt, hvis pejlesteder anvendt i kalibreringen har ens kvalitet og fordeler sig jævnt ud over området. En gennemgang af pejleboringer afslører, at kvaliteten med henblik på kalibrering ikke er ensartet – forskellige længder af tidsserier og påvirkning af indvinding præger pejlingerne. Der er endog tale om få perioder, hvor der foreligger en relativ stor mængde pejlinger i forskellige borer – der er bedst dækning efter år 2000, hvor Amtet etablerede 10 pejleboringer med logger. Før den tid er pejlegrundlaget primært enkeltstående pejlerunder foretaget på mindre områder – eller tidsserier fra pejlinger foretaget af vandværkerne i deres borer.

12.4.2 *Observationer af vandløbsvandføring*

Tidsserier af afstrømningsdata er behæftet med usikkerhed fra kilder såsom registrering af vandstand, vingemålinger af strømningshastighed og efterfølgende transformation til vandføring, fastlæggelse af Q-H relationen (sammenhængen mellem vandføringen Q og vandstanden H) for det grødefri vandløb samt bestemmelse af variationer i vandføringsevnen pga. grødevækst/skæring og sedimenttransport (ændringer i Q-H relationen med tiden).

Vandføringsmålinger, der er udført vha. målinger af strømningshastigheder i veldefinerede tværsnit, er behæftet med en relativ lille *målefejl* (5%). Større fejl vil der være på vandføringsdata, der er bestemt ved måling af vandstand i vandløbet og efterfølgende konverteret til vandføring vha. *Q-H-kurver*. Usikkerheden på Q-H relationen vil blive overført til usikkerhed på beregnet vandføringsværdi, typisk af størrelsesorden 10% (Blicher, 1991).

Usikkerheden på *medianminimum* ved faste målestationer vil afhænge af længden af tidsserien, der er til rådighed til beregningen af værdien, samt af spredningen på data. Baseret på analyse af årsminima fra 10 danske stationer (tidsserier af 65 års varighed) finder Bjarnov (1987) frem til følgende sammenhæng mellem medianminimum, q , og standardafvigelsen på årsminimum, s_q ,

$$s_q = 0.495 + 0.088q \quad (12.1)$$

hvor både s_q og q er i enheden l/s/km² (specifik afstrømning). Hvis årsminimum antages at være normalfordelt vil standardafvigelsen på medianminimum, $s_{\bar{q}}$, approksimativt kunne udtrykkes som

$$s_{\bar{q}} = \frac{s_q}{\sqrt{n}} = \frac{0.495 + 0.088q}{\sqrt{n}} \quad (12.2)$$

hvor n er antallet af årsminima, hvorpå analysen bygges. Hvis anbefalingerne fra Jensen (1993) følges, er $n = 20$. Ved anvendelse af ovenstående ligning (12.2) findes den relative usikkerhed på medianminimum ($s_{\bar{q}}/\bar{q}$) til hhv. 13% og 3% for medianminimumsværdier på 1 og 10 l/s/km², hvilket giver en ide om størrelsesordenen af usikkerheden på medianminimum ved referencestationer.

Synkronmålinger, der repræsenterer medianminimum, vil være behæftet med usikkerhed pga. korrektion ud fra en referencestation, dvs. en fast målestation med en tidsserie på 20 år eller mere. Analysen antager, at vandføringen, hvor synkronmålingen er foretaget, varierer på samme måde som vandføringen ved referencestationen. Pålideligheden af denne fremgangsmåde vil naturligvis afhænge af, hvor godt denne antagelse er opfyldt, og usikkerheden på medianminimumsværdier baseret på synkronmålinger vil derfor være betydelig større end usikkerheden på medianminimum ved referencestationer.

Skalaeffekter har også indflydelse på vandføringsdata. Detaljeringsgraden, hvormed et vandløb kan repræsenteres i en numerisk model, vil afhænge af den anvendte diskretisering. Eksempelvis vil det ikke være muligt at repræsentere vandløb med et opland, der har mindre udstrækning end cellestørrelsen i det numeriske net. Desuden kan den geometriske repræsentation af vandløbet afhænge af den numeriske cellestørrelse. De hydrauliske parametre, der beskriver vandudvekslingen mellem vandløb og hhv. grundvandszonen, overfladekomponenten og umættet zone (eksempelvis vandløbslækagekoefficient, overfladeruhed, infiltrationskapacitet) vil være påvirket af skalaeffekter. Skalaeffekten vurderes at resultere i en usikkerhed på vandføringsdata, der afhænger af forholdet mellem arealet af de numeriske celler og oplandsarealet til vandføringsstationen.

I boks 12.2 ses et eksempel på overvejelser omkring usikkerheden på målinger af vandløbsafstrømning for en model for Lolland.

Boks 12.2 Eksempel Lollandsmodel (Storstrøms Amt, 2003a).

Målinger af vandføringsdata er behæftet med usikkerhed hidrørende fra forskellige kilder. Usikkerheden på vandløbsafstrømningen er typisk relativ lille på målinger af strømningshastighed i veldefinerede tværsnit. Vandføringsdata, som er bestemt ved brug af Q-H relation, vil typisk have en væsentlig større usikkerhed. Usikkerheden vil afhænge af de data som Q-H relationen er opbygget af. Da der oftest vil være flere målinger i normalområdet, vil ekstreme hændelser være behæftet med en større usikkerhed. Forskellige undersøgelser påpeger, at der ikke kan påvises nogen sæsonvariation i usikkerheden på vandløbsafstrømninger. Relativt observeres dog den største fejl på vandføringen i grødesæsonen. Betydelige afvigelser observeres ligeledes uden for sommermånederne, sammenfaldende med stor afstrømning. Den absolutte usikkerhed ses at være klart størst ved store afstrømninger. Der er ligeledes en tydelig sammenhæng mellem måleintensitet og usikkerhed, idet en større måleintensitet medfører en sikrere bestemmelse af afstrømningen. En kvantificering af usikkerheden er vanskeligere end på grundvandspejlingerne. Ved benyttelse af vandføringsdata i en numerisk model vil skalaeffekter have en indflydelse på vandføringsdata. Detaljeringsgraden samt den geometriske repræsentation vil afhænge af den valgte diskretisering.

12.4.3 Anvendelse af usikkerhedsestimater

Summen af ovenstående fejlkilder udgør teoretisk set et mål for, hvor stor en afvigelse der kan forventes mellem observeret og simuleret hydraulisk trykniveau (eller grundvandstilstrømning til vandløb). Det kræver imidlertid, at procesbeskrivelser, den numeriske formulering, randbetingelser, m.m. er fejlfrie. Desuden kræver det, at de hydrauliske parametre kendes i hver enkel celle i det numeriske net, hvilket aldrig er tilfældet for grundvandsmagasiner. Foruden observationsfejlen introduceres der dermed en modelfejl, som skyldes unøjagtigheder i det anvendte ligningssystem, diskretiseringsfejl, de anvendte randbetingelser, de anvendte hydrauliske parametre, etc. Det er denne modelfejl, der kan minimeres ved passende valg af bl.a. modelparametre. Observationsfejlen angiver en nedre grænse for, hvor lille afvigelse der i gennemsnit kan opnås mellem observeret og simuleret tilstandsvariabel (her normalt trykniveau). De såkaldte kalibreringsmål (på engelsk: *target*), som er et udtryk for, hvor præcist det kan forventes, at modelresultaterne beskriver observationsværdierne, er derfor ofte givet ved følgende relation (eksemplificeret ved trykniveauet): $h_{\text{obs}} \pm a \cdot s_{\text{obs}}$, hvor h_{obs} er det observerede trykniveau, a er en konstant, $a \in [1, 2, \dots]$, og s_{obs} er standardafvigelsen på trykniveauobservationen. Hvis simuleringsværdien falder indenfor kalibreringsmålet, må det betragtes som værende tilfredsstillende.

Desuden kan estimatorne af usikkerheden på observationerne anvendes til vægtning af forskellige residualer af forskellige type, hvilket specielt er anvendeligt i forbindelse med invers modellering, se kapitel 14.

12.5 PERFORMANCE KRITERIER

12.5.1 Typer af kriterier

Nærværende afsnit er baseret på en let revision af udsnit fra kapitel 10 i den gamle STÅBI, jf. Sonnenborg (2001).

Kvantitative performancekriterier baseres hyppigt på et mål for den gennemsnitlige afvigelse mellem observeret og simuleret værdi, også benævnt en norm. Nedenfor er angivet nogle normer, som vægter de enkelte residualer på forskellig måde.

ME (mean error eller middelfejl) udtrykker den gennemsnitlige afvigelse mellem observeret ψ_{obs} og simuleret ψ_{sim} tilstandsvariabel

$$ME = \frac{1}{n} \sum_{i=1}^n (\psi_{\text{obs},i} - \psi_{\text{sim},i}) \quad (12.3)$$

hvor n er antallet af observationer. *ME* kan give et indtryk af, om der introduceres nogen overordnet fejl i modelresultaterne, dvs. om f.eks. trykniveauet simuleres generelt for lavt eller højt. Hvis $ME \rightarrow 0$, vil der globalt set ikke optræde systematiske fejl i modellen.

MAE (mean absolute error eller gennemsnitlig absolut fejl) beregner et gennemsnit af de absolutte residualer

$$MAE = \frac{1}{n} \sum_{i=1}^n |\psi_{\text{obs},i} - \psi_{\text{sim},i}| \quad (12.4)$$

I tilfælde hvor observationerne konsekvent simuleres for højt i et område og for lavt i et andet område, kan *ME* godt være tæt på nul og dermed indikere et godt kalibreringsresultat. *MAE* vil afsløre fejl af denne type og kan derfor være et vigtigt supplement til *ME* i vurderingen af den gennemsnitlige fejl.

RMS (root mean squared error eller middelværdien af kvadratafgivelsessummen) er det kriterium, der oftest anvendes til at måle den opnåede overensstemmelse mellem data og model

$$RMS = \sqrt{\frac{1}{n} \sum_{i=1}^n (\psi_{obs,i} - \psi_{sim,i})^2} \quad (12.5)$$

Denne norm er et mål for spredningen på residualerne (lig standardafvigelsen, hvis $ME = 0$), og kan sammenlignes med den estimerede standardafvigelse på observationsdata.

SE (standard error, goodness of fit eller standardafvigelsen) er et direkte mål for modellens evne til at reproducere de observerede data

$$SE = \sqrt{\frac{1}{n-P} \sum_{i=1}^n w_i (\psi_{obs,i} - \psi_{sim,i})^2} \quad (12.6)$$

hvor w_i er vægtningen af observationsdata nr. i , og P er antallet af kalibreringsparametre. I en regressionsmæssig sammenhæng angiver $n - P$ antallet af frihedsgrader. Hvis vægterne w_i specificeres til den reciprokke værdi af variansen på observationerne ($w_i = 1/s_{obs,i}^2$) fås

$$SE = \sqrt{\frac{1}{n-P} \sum_{i=1}^n \left(\frac{\psi_{obs,i} - \psi_{sim,i}}{s_{obs,i}} \right)^2} \quad (12.7)$$

og normen vil dermed tage hensyn til, at der kan være forskellig usikkerhed knyttet til observationsværdierne. Når samtlige modelfejl er elimineret og kun observationsfejl resterer, vil $SE \rightarrow 1$. *SE* giver dermed et direkte mål for, hvor godt de observerede værdier simuleres i forhold til usikkerheden på observationerne.

De ovenstående normer er primært anvendelige til at karakterisere den rumlige fordeling af afvigelsen mellem observeret og simuleret trykniveau (index i henfører til boringen). Hvis der i stedet fokuseres på ikke-stationære tilstande, hvor det er interessant at sammenligne tidsserier af observeret og simuleret trykniveau, er de anførte normer ikke hensigtsmæssige. Følgende analyse kan anvendes til at vurdere den dynamiske fejl: Hvis n angiver antallet af boringer og m er antallet af observationer i tidsserien, vil den samlede fejl kunne kvantificeres som

$$MSE = \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{j=1}^m (\psi_{obs,i}^j - \psi_{sim,i}^j)^2 \quad (12.8)$$

hvor index i og j angiver hhv. stedet og tiden. *MSE* er "mean squared error" eller middelværdien af de kvadrerede afvigelser. Ovenstående sum kan skrives som:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\bar{\psi}_{obs,i} - \bar{\psi}_{sim,i})^2 + \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{j=1}^m [(\psi_{obs,i}^j - \bar{\psi}_{obs,i}) - (\psi_{sim,i}^j - \bar{\psi}_{sim,i})]^2 \quad (12.9)$$

hvor $\bar{\psi}$ angiver den tidsmidlede variabel. Den første sum kvantificerer forskellen mellem tidsmidlet observeret og simuleret trykniveau for de anvendte boringer, og svarer dermed til *RMS*, lign. (12.5). Den anden sum kvantificerer forskellen mellem hhv. det observerede og simulerede trykniveaus variation omkring deres respektive tidslige middelværdier. Den sidste sum giver dermed et

mål for, hvor godt dynamikken i det pågældende system simuleres. For en enkelt boring kan den tidlige fejl derfor kvantificeres vha. følgende udtryk:

$$RMST = \sqrt{\frac{1}{m} \sum_{j=1}^m [(\psi_{obs}^j - \bar{\psi}_{obs}) - (\psi_{sim}^j - \bar{\psi}_{sim})]^2} \quad (12.10)$$

De sidste normer, der gengives her, er R^2 , $Fbal$ og FL altså indikatorer der siger noget om afvigelse i forhold til daglig vandføring, middelfafstrømning og minimumsvandføring (R^2 benævnes også "model efficiency" eller "explained variance" – Nash and Sutcliffe (1970). Bemærk: I MIKE SHE terminologi svarer R^2 til størrelsen E og ikke til korrelationskoefficienten R^2 der produceres i statistikken):

$$R^2 = \frac{\sum (\psi_{obs} - \bar{\psi}_{obs})^2 - \sum (\psi_{obs} - \psi_{sim})^2}{\sum (\psi_{obs} - \bar{\psi}_{obs})^2} \quad (12.11)$$

hvor $\bar{\psi}_{obs}$ er middelværdien af de observerede data. R^2 udtrykker, hvor stor en del af den totale variation i observationsdata, som bliver forklaret af modellen. R^2 er med andre ord et mål for tilpasningsgraden af den optimerede model. R^2 kan maksimalt blive 1.0, hvilket er udtryk for en perfekt overensstemmelse mellem observeret og simuleret tilstandsvariabel, og er ubegrænset nedadtil. Hvis R^2 bliver mindre end 0.0, giver middelværdien af de observerede data en bedre beskrivelse af data end modellen gør, og der er i dette tilfælde grund til at reformulere modellen. R^2 anvendes ofte til kvantificering af overensstemmelsen mellem tidsserier af observeret og simuleret tilstandsvariabel, typisk vandløbsafstrømning.

Et mål for evnen til at simulere den gennemsnitlige afstrømning for en bestemt vandføringsstation kan vurderes ud fra $Fbal$ (i %), Henriksen et al. (2003):

$$Fbal = 100 \frac{\bar{Q}_o - \bar{Q}_s}{\bar{Q}_o} \quad (12.12)$$

Et mål for evnen til at simulere minimumsafstrømning kan evt. vurderes ud fra FL (Wood, 1974; Henriksen et al., 2003):

$$FL = \sum_i \left[\frac{(\bar{Q}_{o,i} - \bar{Q}_{s,i}) \bar{Q}_o}{\bar{Q}_{o,i}^2} \right]^2 \quad (12.13)$$

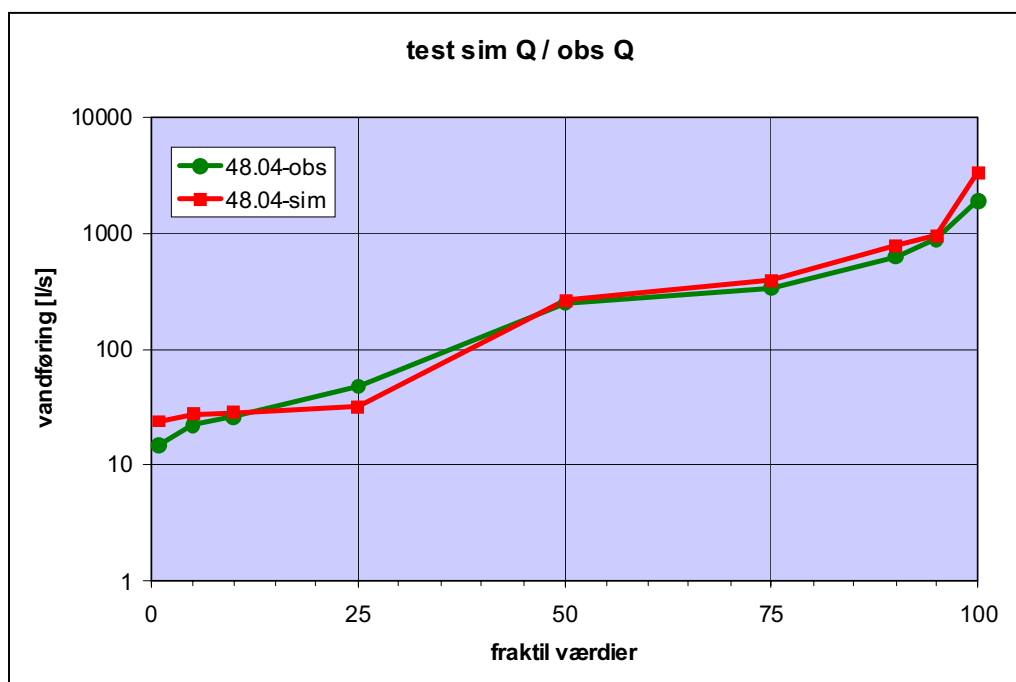
hvor $\bar{Q}_{s,i}$ er simuleret daglig vandføring for dag i (l/s), $\bar{Q}_{o,i}$ er observeret daglig vandføring for dag i (l/s), \bar{Q}_o er gennemsnitlig observeret afstrømning i test-perioden (l/s), og \bar{Q}_s er gennemsnitlig simuleret afstrømning i test-perioden (l/s).

Minimumsafstrømnings-indikatoren FL er meget følsom overfor observeret afstrømning tæt på nul og forudsætter omhyggelig udvælgelse af vandføringsstationer til kalibrering og validering. For at minimere støj bliver FL normaliseret i forhold til en valgt referencekørsel:

$$FL_{norm} = \frac{FL}{FL_{ref}} \quad (12.14)$$

Det understreges, at der ikke findes et godt mål til vurdering af afvigelser i forhold til minimumsafstrømningen. FL giver en statistisk størrelse, som vægter de små afstrømninger mest, men på den anden side er størrelsen pr. definition ustabil ved de meget små vandføringer.

Som alternativ til at sikre en god beskrivelse af afstrømningshydrografen, som ovenstående flow-indikatorer alle tilstræber, kan det i mange tilfælde være mere optimalt at kigge på udvalgte fraktilværdier baseret på en rangordning af de daglige vandføringer og beregning af f.eks. 5, 10, 25, 50, 75, 90 og 95 % fraktil værdier, se figur 12.4.



Figur 12.4 Sammenligning af fraktilværdier for station 48.04 Højbro å. Baseret på National vandressource model for Sjælland og daglig vandføring 1991-1996. Et robust performancekrav kunne være residualer på udvalgte fraktilværdier, f.eks. 5% fraktilen - Q95 i habitat-terminologi eller 10 % fraktilen - Q90). I så fald kunne man specificere et nøjagtighedskrav, hvor modellen max måtte afvige med en vis procentdel i forhold til udvalgte fraktiler.

12.5.2 Valg af nøjagtighedskriterier

Det er hensigtsmæssigt at opstille kriterier for, hvor præcist modellen skal kunne reproducere kalibreringsdata, før kalibreringsprocessen påbegyndes. Herved sikres det, at modellen opnår en kvalitet, som er i overensstemmelse med formålet med modelarbejdet. Samtidig sikrer klare kalibreringskriterier, at modelløren ved, hvad målet med estimationsarbejdet er, og hvornår modellen kan betragtes som være færdigkalibreret.

Både kvantitative og kvalitative kriterier kan med fordel specificeres. De kvantitative kriterier kan opdeles i en vurdering af den gennemsnitlige fejl og spredningen på residualerne. Den gennemsnitlige fejl skal generelt være tæt på nul og kan vurderes vha. *ME*. Det er sværere at opnå præcise resultater, hvis trykniveauet ændrer sig meget inden for korte afstande (dvs. i områder med store gradienter). Samtidig vil en model med et givent fejlniveau simulere strømningsmønstret bedre i områder, hvor der optræder store forskelle i trykniveau end i områder med små forskelle i trykniveau. Det er derfor rimeligt at vurdere den gennemsnitlige fejl i forhold til det totale trykniveaufald i modelområdet.

Spredningen på residualerne kan med fordel relateres til usikkerheden på observationsdata. Standardafvigelsen på det hydrauliske trykniveau inkluderer effekten af både unøjagtige målemetoder og hydrogeologiske karakteristika såsom den hydrauliske gradient og den geologiske heterogenitet. Det er derfor relevant at relatere forskellen mellem observeret og simuleret trykniveau til den estimerede standardafvigelse på observationerne. Dette gøres direkte ved brug af lign. (12.7) og kan alternativt gøres ved at skalere *RMS* med s_{obs} , hvis standardafvigelsen er ens for de indgående observationer. Hvis der ikke er foretaget en analyse af observationernes usikkerhed, kan spredningen på residualerne relateres til det totale trykniveaufald i området.

Hvis der simuleres ikke-stationært vil der ud over kriterier for fordelingen af den rumlige fejl være behov for vurdering af, hvor godt dynamikken reproduceres. Til dette formål kan størrelsen

RMST, lign. (12.10), benyttes, og det vil være rimeligt at relatere den beregnede RMST-værdi til amplituden på de sæsonmæssige variationer i observeret trykniveau.

I tabel 12.3 er der angivet alternative kriterier for overensstemmelsen mellem observeret og simuleret hydraulisk trykniveau.

Tabel 12.3 Kriterier for overensstemmelsen mellem observeret og simuleret hydraulisk trykniveau og vandløbsafstrømning. β_i ($i = 1,2,3,4$) er nøjagtighedskriterier.

Nr.	Kriterium	Kommentarer
1	$\frac{ME}{\Delta h_{\max}} \leq \beta_1$	Vurdering af middelfejl. Δh_{\max} er forskellen mellem maksimum og minimum hydraulisk trykniveau i området. Dette kriterium udtrykker, at den globale under- eller overprediktion i forhold til den globale trykniveauforskelle i modelområdet skal være mindre end β_1 .
2	$\frac{RMS}{s_{obs}} \leq \beta_2$	Vurdering af spredningen på residualerne i forhold til standardafvigelsen på observationsværdierne. Anvendes, hvis usikkerheden på observationsdata er kvantificeret (her angivet ved standardafvigelsen s_{obs}), hvis der kun indgår én datatype i kriteriet og alle data er behæftet med samme usikkerhed.
3	$SE \leq \beta_2$	Vurdering af spredningen på residualerne i forhold til standardafvigelsen på observationsværdierne. Hvis der er variabel usikkerhed på de inkluderede observationsdata kan dette kriterium anvendes.
4	$\frac{RMS}{\Delta h_{\max}} \leq \beta_3$	Vurdering af spredningen på residualerne i forhold til trykniveauvariationen i området. Hvis der ikke er foretaget en kvantificering af observationsusikkerheden, vil dette kriterium kunne anvendes til at vurdere, hvor godt observationsdata i gennemsnit simuleres.
5	$\frac{RMST}{\Delta h_t} \leq \beta_4$	Vurdering af tidlig fejl. Dette dynamiske kriterium benyttes til vurdering af, hvor godt de ikke-stationære variationer simuleres. Δh_t er amplituden på de sæsonmæssige variationer i observeret trykniveau.
6	$R^2 \geq \beta_5$	R^2 udtrykker, hvor stor en del af den totale variation i observationsdata, som bliver forklaret af modellen. R^2 er med andre ord et mål for tilpasningsgraden af den optimerede model.
7	$ F_{bal} \leq \beta_6$	Et mål for evnen til at simulere den gennemsnitlige afstrømning for en bestemt vandføringsstation
8	FL	Minimumsafstrømnings indikatoren FL er meget følsom overfor observeret afstrømning tæt på nul og forudsætter omhyggelig udvælgelse af vandføringsstationer til kalibrering og validering.

Foruden de kvantitative krav opstillet ovenfor, kan der angives kvalitative kriterier. Følgende tre kriterier vil være fornuftige: (1) De estimerede parametre skal have realistiske værdier. (2) Residualerne skal være fordelt rimeligt jævnt både i tid og sted. Selv om kriterium nr. 1 i tabel 12.1 er opfyldt for området under ét, kan der sagtens være områder eller tidsrum, hvor modellen konsekvent simulerer for store eller små værdier. Den type systematisk fejlmodellering skal så vidt muligt elimineres fra modellen. (3) Områdets hydrogeologiske karakteristika skal reproducere af modellen. Det kan eksempelvis kræves, at modellen er i stand til at simulere strømningsretning, retningen af gradienten over dybden eller beliggenheden af grundvandsskel korrekt.

12.6 TESTKØRSEL OG FASTLÆGGELSE AF KVANTITATIVE OG KVALITATIVE NØJAGTIGHEDSKRITERIER FORUD FOR KALIBRERING

Afhængigt af om usikkerheden på trykniveauobservationsdata er bestemt, kan både kriterium nr. 1 samt et eller flere af kriterierne 2 – 5 i tabel 12.3 specificeres. Hvor stringente krav, der skal opstilles til en given model, afhænger af formålet med undersøgelsen. I tabel 12.4 er der givet et eksempel på, hvordan nøjagtighedskriterierne bliver skærpet i takt med at kravene til modellens pålidelighed øges.

Tabel 12.4 Nøjagtighedskriterier for forskellige modeltyper. Procentsatsen i parentes angivet for kriterium 2 svarer til konfidensintervallet for det pågældende nøjagtighedskriterium.

	Overslagsberegning	Akvifer simulering	
		Konservativ	High fidelity
β_1	0.05	0.025	0.01
β_2	2.6 (99%)	2 (95%)	1.65 (90%)
β_3	0.15	0.1	0.05
β_5	0.4	0.6	0.75
β_6	0.3	0.2	0.1

Hvis det antages, at observationsfejlen er normalfordelt, vil eksempelvis kriteriet $\beta_2 = 2$ betyde, at simuleringstværdierne i gennemsnit skal falde indenfor observationsværdiernes 95%-konfidensinterval.

Det er vigtigt at pointere, at en models performance ikke kan karakteriseres alene ud fra et enkelt af kriterierne opskrevet i tabel 12.4. For at kategorisere en model som f.eks. ”hi-fi” skal både de højeste krav til ME, RMS og evt. andre størrelser være opfyldt samtidigt med at de opstillede kvalitative krav overholdes.

Tabel 12.5 Eksempel på opstillede performance kriterier for DK model Sjælland (Henriksen et al., 2003). I stedet for et enkelt krav til f.eks. RMS, så kan det være hensigtsmæssigt at arbejde med en karakterisering efter forskellige klasser f.eks. fremragende, meget god, god, ringe og meget ringe, og så sigte efter at opnå så mange stjerner som muligt.

Performance indikator	Fremragende (5 points)	Meget god (4 points)	God (3 points)	Ringe (2 points)	Meget ringe (1 point)
RMS (m)	< 4	4–6	6–8	8–10	> 10
R^2	> 0.85	0.65–0.85	0.50–0.65	0.20–0.50	< 0.20
F_{bal}	< 5%	5–10%	10–20%	20–40%	> 40%
Aggregeret score	☆☆☆☆☆	☆☆☆☆	☆☆☆	☆☆	☆

I tabel 12.5 ses et eksempel på kategorisering af overensstemmelsen mellem observeret og simuleret variabel. Intervallet fra <4 (fremragende) til >10 (meget ringe) på RMS er subjektivt fastlagt dog med en skalering på baggrund af skøn af usikkerheder på pejleobservationer i forhold til modelskala. Denne var her opgjort til 3,1 m (Henriksen et al., 2003) for DK-model Sjælland, men afspejler dels en modelskala på 1x1 km og dels at der er anvendt ret usikre observationsdata fra JUPITER.

Ved modellering på mindre skala og ud fra bedre datagrundlag f.eks. baseret på pejletidsserier eller nivellerede observationsboringer skal anvendes en mere restriktiv skala. Ofte vil man med modeller med maskevidder på 200-500 meter og god datakvalitet på observationer, kunne nå under 2,0 meter i RMS værdi. I andre tilfælde betyder heterogeniteter og andre forhold, at man må acceptere mindre restriktive performancekrav.

Erfaringer med R^2 værdier for afstrømning anbefaler værdier mellem 0.50 og 0.95 (Henriksen et al., 2003). Værdierne for F_{bal} er baseret på oplande af størrelsesordenen 100-1000 km². For små oplande kan det være nødvendigt at operere med mindre restriktive performancekrav.

Boks 12.4 Eksempel på indledende nøjagtighedskriterier og endelige nøjagtighedskriterier for Lollandmodel - Storstrøms Amt, (2003a).

I udbudsbetingelser for opstilling af en regional strømningsmodel for Lolland, som havde til formål at bidrage med viden om behov for særlig indsats angående geologisk detailkortlægning, definition af randbetingelser for numeriske lokalmodeller for fem indsatsområder og belysning af grundvandsressourcens størrelse, blev nøjagtighedskrav specificeret som følger:

Der forudsættes opstillet en 3D numerisk strømningsmodel med henblik på akvifersimulering, det vil sige med et højt krav til modellens beskrivelse af strømningsforhold i tre dimensioner, samt tidslige variationer, og udveksling mellem grundvand og vandløb...Tilbud skal indeholde en nærmere beskrivelse af princip for fastlæggelse af beregningslag. Den regionale model diskretiseres horisontalt i 500 x 500 meter kasser...I forbindelse med partikelbanesimuleringer skal indvindingsoplande til oppumpninger > 25.000 m³/år beskrives...Ud fra en første kørsel med den stationære model og ikke stationære model og bestemte kvantitative kalibreringsmål (fx ME, RMS, SE eller R², se Henriksen et al., 2001), skal der være udarbejdet nøjagtighedskrav til slutmodellen efter endt kalibrering.

Modelløren foreslog på dette grundlag såvel et sæt kvantitative og et sæt kvalitative kriterier for kalibreringen. Der blev foreslået krav til trykniveau formuleret som middelfejl (ME) < 0,75 m, root mean square afvigelse (RMS) < 1,5 m og standardafvigelse (STDres/delta H) < 0,3 m. For vandløb anvendtes krav til vandbalancefejlen Fbal (%). Sidstnævnte blev formuleret som afhængig af middelvandføringens størrelse. For vandføring > 500 l/s var kravet: Fbal < 10%, 200-500 l/s: Fbal < 15 %, 100-200 l/s: Fbal < 25 %, 50 – 100 l/s: Fbal < 50 % og for vandføringer under 50 l/s: Fbal < 100 %. De tilsvarende krav for E (R²) var fastsat til: For vandføring > 500 l/s var kravet: > 0,75; 200-500 l/s: > 0,65; 100-200 l/s: > 0,5; 50 – 100 l/s: > 0,2 og for vandføringer under 50 l/s: > 0,0.

Tabel 12.6 Eksempel på startgæt og variationsrammer for DK model Sjælland (Henriksen et al., 2003).

Parametre i mættet zone (SZ)	Startgæt (Range)	
Tidskonstant, dræn	$2 \times 10^{-7} \text{ s}^{-1} (1 \times 10^{-6} - 1 \times 10^{-8})$	
Hydraulisk ledningsevne	Horisontal	Vertikal
Lag 1 (sprækket moræneler)	$1 \times 10^{-5} \text{ m/s} (1 \times 10^{-6} - 1 \times 10^{-4})$	$1 \times 10^{-7} \text{ m/s} (1 \times 10^{-8} - 1 \times 10^{-6})$
Lag 2, 4, 6, 8 and 10 (moræneler og ler)	$1 \times 10^{-7} \text{ m/s} (1 \times 10^{-10} - 1 \times 10^{-6})$	$1 \times 10^{-9} \text{ m/s} (1 \times 10^{-11} - 1 \times 10^{-7})$
Lag 3, 5, 7 and 9 smeltevandssand og grus)	$1 \times 10^{-4} \text{ m/s} (1 \times 10^{-6} - 1 \times 10^{-2})$	$1 \times 10^{-5} \text{ m/s} (1 \times 10^{-7} - 1 \times 10^{-3})$
Lag 11 (kalk og limsten)	distribueret ($1 \times 10^{-7} - 1 \times 10^{-3}$)	distribueret (x 0.1) ($1 \times 10^{-8} - 1 \times 10^{-4}$)
Impermeabel bund	$1 \times 10^{-20} \text{ m/s}$	$1 \times 10^{-20} \text{ m/s}$
Sandlinser i ler	$1 \times 10^{-4} \text{ m/s} (1 \times 10^{-7} - 1 \times 10^{-3})$	$1 \times 10^{-5} \text{ m/s} (1 \times 10^{-8} - 1 \times 10^{-4})$
Glacialtektoniske linser i ler	$2 \times 10^{-6} \text{ m/s}$	$1 \times 10^{-7} \text{ m/s}$
Glacialtektoniske linser i sand	$1 \times 10^{-5} \text{ m/s}$	$1 \times 10^{-5} \text{ m/s}$
Magasintal	Frit magasintal	Artesisk magasinkoefficient
Lag 1 (sprækket moræneler)	0.25 m/m (0.05–0.20)	$0.0001 \text{ m}^{-1} (1 \times 10^{-4} - 1 \times 10^{-2})$
Lag 2, 4, 6, 8 and 10 (moræneler and ler)	0.25 m/m (0.01–0.18)	$0.0001 \text{ m}^{-1} (1 \times 10^{-4} - 1 \times 10^{-2})$
Lag 3, 5, 7 and 9 smeltevandssand og grus)	0.25 m/m (0.10–0.35)	$0.0001 \text{ m}^{-1} (1 \times 10^{-5} - 1 \times 10^{-3})$
Lag 11 (kalk)	0.25 m/m (0.01–0.35)	$0.0001 \text{ m}^{-1} (1 \times 10^{-6} - 1 \times 10^{-3})$
Udveksling mellem grundvand og vandløb		
Lækagekoefficient, vandløbsbund	$1 \times 10^{-7} \text{ s}^{-1} (1 \times 10^{-10} - 1 \times 10^{-6})$	
Interaktionstype	type 3 (kun vertikalt tryktab)	
Overlandflow (OC)		
Overfladeruhed, Manning (M)	$3 \text{ m}^{-1/3}/\text{s} (1-5)$	
Overflademagasiner	0.01 m (0–0.05)	
Vandløbsruhed, Manning (M)	$20 \text{ m}^{-1/3}/\text{s} (10-35)$	

Boks 12.4 Eksempel Stevns model (Storstrøms Amt, 2003b).

Som kvantitative performancekrav blev efter stationær og dynamisk testkørsel med MIKE SHE model med DAISY input fastlagt krav til *RMS* ($< 1,75$ m), R^2 (udvalgte stationer $> 0,65$ og kun $\frac{1}{4}$ af stationerne måtte være $< 0,6$) og *Fbal*.

Kravet til *RMS* blev fastlagt ud fra at der i formålet var specificeret akvifer simulering, og at 5 % af den største forskel mellem potentialeverdierne inden for modelområdet (forskul egentlig = 75 m, men kun to observationer lå i dette interval, derfor blev det mere restriktive interval hvor de fleste observationer var beliggende vurderet til = 35 m; 5 % af 35 meter = 1,75 m som krav til *RMS*). Testkørslen viste for den stationære model en *RMS* værdi på 4,85 m, en middelfejl *ME* på $-2,63$ m og en *MAE* på 3,62 m.

Testkørslen viste relativt dårlige R^2 værdier med en middelværdi på 0,05 for kalibreringsperioden. Dette problem viste sig dog til dels at skyldes problemer med DAISY inputtet som manglede noget dynamik.

Kravet til *Fbal* blev fastsat så der for stationer med stor afstrømning (> 15 mio. m^3 pr. år) blev sat et krav på 10 %, for intervallet 5 – 15 mio. m^3 pr. år var kravet 15 %, og under 5 mio. m^3 pr. år 25 %. Den samlede vandbalance for modelområdet (*Qsim total* / *Qobs total*) var fastsat til 5 %. Testkørslen viste for den stationære model at to ud af 8 stationer umiddelbart overholdt kriteriet med afvigelse på 6-13 %. Flere stationer havde afvigelse på indtil 50 %.

De kvalitative krav var fastsat som følgende:

- De estimerede parametre skal have realistiske værdier
- Residualerne skal være fordelt fornuftigt arealmæssigt. F.eks. kan de kvantitative kriterier godt være opfyldt for hele modellen under ét, mens modellen konsekvent simulerer for store eller små værdier
- Områdets hydrogeologiske karakteristika skal reproducere af modellen. F.eks. skal modellen være i stand til korrekt at simulere strømningsretning, retningen af gradienten over dybden eller beliggenheden af grundvandskel
- For vandløbene skal den rigtige dynamik afspejles. Dvs. en hændelse giver den rigtige respons både i tid og sted.
- Vandløbsmodellen giver de korrekte ekstremværdier (min og max). Dvs. at dynamikken er korrekt også i yderområdet af spektret.

12.7 REFERENCER

Beven, K. (2002) Towards a coherent philosophy for modelling the environment. Proc roy Soc Lond A 2002; 458(2026): 2465-84.

Bjarnov, S. (1987) Beregning af medianminimum på grundlag af årsminimumsserier. Hedeselskabets Forskningsvirksomhed. Beretning nr. 32.

Blicher, A.S. (1991) Usikkerhed på bearbejdning af data fra vandføringsstationer. Publikation nr. 1 (rev. udgave) fra Fagdatacenter for Hydrometriske Data, Hedeselskabet.

Feldman, S.P. (2004) The culture of objectivity: Quantification, uncertainty, and the evaluation of risk at NASA. Human Relations. Vol. 57(6): 691-718. The Tavistock Institute. SAGE Publications. London.

Freyberg, D. (1998) An exercise in ground-water model calibration and prediction. Ground Water, v. 26, no. 3, p. 350-360.

Gelhar (1986) Stochastic subsurface hydrology. Englewood Cliffs. New Jersey. Prentice Hall.

Harrar, W.G., Sonnenborg, T.O., and Henriksen, H.J. (2003) Capture zone, travel time, and solute-transport prediction using inverse modeling and different geological models. Hydrogeology Journal 11: 536-548.

Henriksen, H.J., Refsgaard, J.C., Sonnenborg, T.O., Gravesen, P., Brun, A., Refsgaard, A. and Jensen, K.H. (2001) STÅBI i grundvandsmodellering. Danmarks og Grønlands Geologiske Undersøgelse Rapport 2001/56.

Henriksen, H.J., Troldborg, L., Nyegaard, P., Sonnenborg, T.O., Refsgaard, J.C. and Madsen, B. (2003) Methodology for construction, calibration and validation of a national hydrological model for Denmark. *Journal of Hydrology* (280) 52-71.

Hill, M.C. (1998) Methods and guidelines for effective model calibration. U.S. Geological Survey. Water-Resources investigation report 98-4005. Denver, Colorado. 1998.

Jensen, J.L. (1993) Rererenceperiode for karakteristiske afstrømningsværdier. Publikation nr. 13 fra Fagdatacenter for Hydrometriske Data, Hedeselskabet.

Middlemis, H. (2004) Benchmarking best practice for groundwater flow modelling. The Winston Churchill Memorial Trust of Australia. 13. December 2004.

Nash, J.E. and Sutcliffe, J.V. (1970) Riverflow forecasting through conceptual models. Part 1: a discussion on principles. *Journal of Hydrology* 10, 282-290.

Refsgaard, J.C. and Henriksen, H.J. (2004) Modelling guidelines – terminology and guiding principles. *Advances in water resources*. (27) 71-82.

Refsgaard, J.C., Henriksen, H.J., Harrar, W.G., Scholten, H. and Kassahun, A. (2004) Quality assurance in model based water management –review of existing practice and outline of new approaches. *Environmental Modelling & Software* (in press).

Scholten, H., Van der Tol, M.W.M. (1998) Quantitative validation of deterministic models; when is a model acceptable? In: Obaidat, M.S., Davoli, F., DeMarinis, D. (Eds.), *The proceedings of the summer computer simulation conference*. SCS, The society for computer simulation international, San Diego, CA, USA, pp. 404-409.

Sonnenborg, T.O. (2001) Kalibrering af strømningssmodel. Kapitel 10 i Henriksen et al. (2001).

Sonnenborg, T.O., et al. (2003)

Storstrøms Amt (2003a) Regional strømningssmodel for Lolland – Hydrogeologisk model. DHI, Institut for vand og miljø.

Storstrøms Amt (2003b) Indsatområderne Stevns og Fakse, Fase 2: Detailkortlægning, Milepæl 2: Modelopstilling og fastlæggelse af nøjagtighedskrav, Rambøll, August 2003.

The council for regulatory environmental modellering (2003) Draft guidance on the development, evaluation, and application of regulatory environmental models. Principal authors: Pasky Pascual, Neil Stiber and Elsie Sunderland. Office of science policy. Office of research and development. Washington D.C.

USGS (2004) Guidelines for evaluating ground-water flow models. Scientific Investigation Report 2004-5038. U.S. Department of the Interior. U.S. Geological Survey. By Thomas E. Reilly and Arlen W. Harbaugh. Reston, Virginia.