

Utvikling av metodikk for valg av riktig beregningsmodell

Av Tor Haakon Bakken og Tuomo M. Saloranta

Begge forfattere er ansatt som forskere ved Norsk institutt for vannforskning (NIVA)

Sammendrag

Beregningsmodeller kan være svært nyttige verktøy i forbindelse med analyse av tekniske, økonomiske og miljømessige forhold ved forvaltning og bruk av vann. Samtidig finnes en myriade av tilgjengelige kommersielle og ikke-kommersielle modeller som kan være aktuelle redskap når en bestemt forvaltningsoppgave skal analyseres. Denne artikkelen beskriver en metode for å velge riktig matematisk modell, gitt et bestemt problem. Det har vært et krav at man skal benytte metoden til å velge ut modell uten at modellen først må settes opp. Bruk av metoden vil forhåpentligvis sikre at den valgte modell kan benyttes til å besvare nettopp de spørsmål forvaltningen ønsker belyst. Utvelgelsesmetodikken bygger på et sett av "benchmark"-kriterier som til sammen skal omhandle alle vesentlige aspekt vedrørende valg av modell, slik som hvor godt modellens utgangsdata relateres til forvaltningsoppgaven, modellens oppløsning i tid og rom og balansen mellom modellens krav til inngangsdata og tilgjengelige data. Det blir poengtert at benchmark-kriteriene skal anvendes i en dialog mellom forvalter og modellør.

1. Rammedirektivet for vann og behovet for modellverktøy

EUs rammedirektivet for vann er et ambisiøst direktiv med tanke på å analysere og iverksette de riktige miljøforbedrende tiltak i norske nedbørfelt. I mange nedbørfelt og vannforekomster vil man oppleve at det er avvik mellom observert tilstand og hva som kan defineres som "god økologisk status". Man blir da raskt konfrontert med spørsmålet om hvilke tiltak som må settes inn, og med hvilken sannsynlighet man vil nå det ønskede mål om "god økologisk status" innen 2015.

For å kunne analysere og besvare dette spørsmålet vil matematiske beregningsmodeller, implementert i form av et dataprogram (i det følgende bare omtalt som "modell"), kunne være svært nyttige verktøy. Valget av verktøy eller modell er ikke nødvendigvis bestandig like enkelt, blant hundrevis av kommersielle og ikke-kommersielle tilgjengelige modeller. Det er en rekke vurderinger av teknisk og naturvitenskapelig art, som for eksempel at de ønskede grupper av stoffer håndteres på en forsvarlig måte i modellen, at modellen kan settes opp

på den ønskede romlige og tidsmessige skala, at tilgjengelige inngangsdata passer modellens krav, at modellen forventes å produsere resultater nøyaktige nok til det gitte formål, at modellen kan eksekveres på den dataplattformen modelløren har tilgjengelig, osv. I tillegg er det også erkjent (for eksempel i Dale, 2003) at det finnes mange problemer knyttet til bruk av modeller innen miljøforvaltningen som ikke kan relateres direkte til naturvitenskapelige og tekniske forhold, men som helt klart reduserer nytten av bruk av modeller til å analysere miljøproblemer. Det tenkes her på generelt manglende eller dårlig kommunikasjon mellom modelløren og forvalteren, de ofte forskjellige oppfatninger om hva slags utsagnskraft modellen har og hva som forventes av svar for å kunne bruke dette i en beslutningsprosess.

Et steg i retning av både å redusere disse problemene og å unngå mulige fallgruver er å gi en større oppmerksomhet til nettopp prosessen hvor modellverktøy velges etter at studiens mål og omfang er definert. Det EU-finansierte prosjektet Benchmark Models for the Water Framework Directive (BMW-prosjektet), hvor NIVA er en sentral aktør, arbeider nettopp med å utvikle en metodikk for å kunne velge riktig modell, gitt et bestemt formål, *uten at modellen settes opp* for det aktuelle området og problemstilling.

Denne artikkelen beskriver selve metodikken for utvalgelse av riktig modell, testing av metodikken på Vansjø-Hobølvassdraget, hvilke forbedringer i metodikken som er under

utvikling og hvordan både forvaltere og modellører vil kunne anvende resultatene av prosjektet i fremtiden. Selve den naturfaglige modelleringsstudien er ikke avsluttet ennå, og vil bli rapportert i form av en standard NIVA-rapport (Saloranta, 2004) og forhåpentligvis også i VANN.

Det bør til slutt nevnes at utvikling av tiltaksplaner er kanskje det mest åpenbare anvendelsesområdet for modeller. Tiltaksplaner med dose-responsanalyser krever at utslippene/stoffavrenningen (kilder) kobles sammen med effektmodeller (elv, innsjø, marint område) ettersom målene knyttes til vannforekomstene, mens problemene skapes og tiltakene befinner seg på eller ved kildene. Ved siden av tiltaksplaner vil modeller også kunne spille en rolle ved fastsettelse av referansetilstand (analysere situasjonen før menneskelig påvirkning, kobling av vannkjemi og økologi), karakterisering (romlig ekstrapolering/interpolering, fyller "hull" i dataserier) og ved gjennomgang og etablering av overvåkingsprogram.

2. Utvikling av kriterier for valg av riktig modell

Metodikken utviklet av Saloranta et al. (2003) for utvalgelse av riktig modell baserer seg på 14 så kalte "benchmark"-kriterier, som er formulert som enkle spørsmål, og som skal berøre de fleste viktige praktiske aspekter ved modeller og modellering. Ved hvert spørsmål er det gitt korte beskrivelser som kan hjelpe en til å velge den mest riktige karakteren fra tre alternativer, "god (good)", "tilfredsstillende (adequate)", og "ikke tilfredsstillende

(inadequate)". I metodikken er "benchmark" (dvs. standarden man måler modellens kvalitet mot) satt slik at hvis modellen får én eller flere "ikke tilfredsstillende" karakterer, vil den ikke oppfylle "benchmark"-kriteriene for den spesifikke modelleringsoppgaven, som alltid må defineres i forkant av selve evalueringen.

Den generelle filosofien bak benchmark-kriterier gjenspeiler tankemåten i den såkalte post-normale vitenskaps-teorien (Ravetz, 1999). Den hevder at i den komplekse naturen og med nåtidens komplekse miljøproblemer, hvor betydelige usikkerheter alltid vil være til stede og konsekvensene av gal beslutning kan være store, er det viktigere å fokusere på kvaliteten av informasjonen enn bare å fokusere på leting etter en (muligens ikke-eksisterende) kvantitativ sannhet. Med andre ord, det er viktigere å gi forvaltningen en bred og gjennomiktig beskrivelse av usikkerhetene ved resultatene fra modellen enn å prøve å fortrenge eksistensen av disse usikkerhetene og late som om modellen gir et korrekt svar med stor nøyaktighet. Innholdet i benchmark-kriterier reflekterer også meninger fra forskjellige vitenskapsmiljøer om hva som anses å være gode modellegenskaper når modellen skal brukes i beslutnings- eller forvaltningsrelevante sammenheng.

Benchmark-kriterier kan brukes på flere forskjellige måter for å øke kvaliteten ved en forvaltningsrelevant modelleringsstudie: 1) man kan bruke kriterier til å evaluere modeller og for å hjelpe en å velge den mest riktige blant flere modellkandidater; 2) man

kan evaluere en allerede valgt modell og bruke evalueringen som et slags faktaark om modellens kvalitative egenskaper; 3) man kan bruke kriterier for seg selv som en sjekkliste for å øke bevisstheten om de forskjellige sider av modeller og modellering. Ofte er det dessuten slik at når man utvikler og bruker en bestemt modell, får man snart et sterkt eierskap til denne og det kan være vanskelig å distansere seg fra sin "egen" modell og se dens virkelige styrker og svakheter, gitt en bestemt anvendelse.

Benchmark-kriterier skiller seg fra andre metoder for modellevaluering idet de er kvalitative, det vil si at de ikke kvantitativt måler modellenes tekniske yteevne. Samtidig tilbys en noenlunde fastsatt standard ("benchmark") man kan evaluere forskjellige modeller mot. Kriteriene er dessuten formulert i et språk som også ikke-spesialister kan forstå. En annen styrke ved benchmark-kriteriene er at de også kan brukes uten å måtte gjennomføre den ofte tidkrevende oppsettningen og kalibreringen av modellen som trengs for eksempel ved en mer kvantitativ sammenligning av modellytelse. Det er nettopp dette ressursbehovet som oftest hindrer en kvantitativ testing av flere modeller, selv om dette ideelt sett skulle vært med i en helhetlig modellevaluering. Uansett hvordan man evaluerer modeller vil en slik vurdering alltid være gyldig bare for den spesifikke modellapplikasjonen (eller for svært lignende applikasjoner). Ingen modeller kan være universelt gode til alle tenkelige anvendelser.

Man kan kritisere kriteriene for å

være for ”myke” og kvalitative og for å ikke direkte evaluere modellens tekniske ytelse, det vil si hvor godt modellen er i stand til å gjenskape observasjoner, for eksempel gjennom statistiske analyser. På den annen side får man aldri noen garanti for at en modell som beskriver nåtidens eller historisk tilstand tilfredsstillende også vil beskrive fremtiden på en god måte. Vi tror også at beslutningstagere ønsker å få mer innsikt i selve modellen og modelleringsprosessen enn hva enkle ytelseskriterier gir når viktige samfunnsmessige beslutninger baseres på modellresultatene. Det kan være nødvendig å klargjøre en bestemt modells troverdighet på bakgrunn av dens historie, struktur, hvilke prosesser som er inkludert og hvor omfattende testing den har gjennomgått. Mulighetene benchmark-kriteriene gir er nettopp å kunne evaluere modellens evne til å produsere troverdig informasjon av god kvalitet til beslutningstagere og forvaltere, utover hva de tradisjonelle teknikker for ytelse kan gjøre.

3. Testing av kriteriene

For å vurdere brukbarheten og deretter justere og forbedre metodikken for utvelgelse av riktig modell ble kriteriene testet ut i Vansjø-Hobøl-vassdraget. Selve modelleringen er (dokumentert i saloranta, 2004) og vil bli omtalt mer utdypende i senere artikler, slik at her er kun testingen av selve settet med kriterier beskrevet.

Det første som måtte gjøres i henhold til Saloranta et al. (2003) var å tydelig definere hvilke forvaltningsoppgaver som skulle analyseres. For

Vansjø's vedkommende var dette å kvantifisere nødvendige reduksjoner i tilførsler av fosfor til Storefjorden (det største bassenget i Vansjø) for å nå miljømålet om en maksimal fosforkonsentrasjon i innsjøen på 11 µg/l. I tillegg ønsket man å analysere hvordan fordelingen av tilførslene fra de ulike delnedbørfeltene påvirket situasjonen i Storefjorden, hvilke faktorer som styrer eutrofiproblemet (tilførsler, intern gjødsling, klima, osv), samt beskrive langtidsvariasjonen. For en mer fyllestgjørende beskrivelse av problemkomplekset henvises det til Lyche Solheim et al. (2001) og Saloranta et al. (2004).

Etter at modelleringsproblemet er definert må man ved hjelp av et beslutningsdiagram (se Appendix A i Saloranta et al., 2003) forsikre seg om at anvendelsen har et forvaltningsperspektiv ved seg (i motsetning til for eksempel mer forskningspregede anvendelser), at modellen eksisterer i form av programvare og at det er tilstrekkelige ressurser tilgjengelig for å forsvare anvendelsen av en modell. For å kunne besvare de nevnte forvaltningsspørsmål valgte man å sette opp en elvemodell som håndterer transport og omsetning av næringssalter fra de enkelte delnedbørfeltene gjennom bekke-/elvesystemet ned til Storefjorden, og en innsjømodell som beregner fosforkonsentrasjon og algebiomasse (gitt som kl-a). Den første kandidaten til elvemodell var SOBEK, utviklet av WL | Delft Hydraulics (www.wldefl.nl). Anvendelse av benchmark kriterier (kortversjon av disse gitt i annekset A til denne artikkel) viste at SOBEK fikk karak-

teren "good" på mange av spørsmålene og "adequate" på enkelte. Modellen får full uttelling på de kriteriene hvor kommersielle modeller tradisjonelt er veldig sterke, det vil si for eksempel hvor godt den er testet, dokumentasjonens beskaffenhet, versjonskontroll, brukergrensesnitt/-vennlighet og "aksept" i markedet. Av spørsmålene SOBEK viste seg svakere på, men fortsatt fikk karakteren "adequate", var hvorvidt modellen er for komplisert i forhold til de forvaltningsoppgaver den skal analysere (1.4) og særlig hvordan forholdet mellom tilgjengelige data og modellens datakrav er (1.5). Til den hydrodynamiske delen trenger SOBEK detaljerte topografiske data som ikke er tilgjengelig for bekkene/elvene i Vansjø, samt vannstand og vannhastighet for å kalibrere denne modulen. I tillegg er det ønskelig med bedre data på tilførselene, særlig variasjonen over tid, og mer vannkjemidata for kalibrering av modulen for vannkvalitet. Det er imidlertid på spørsmålet om hvor egnet modellen er til å kjøre sensitivitets-/usikkerhetsanalyse (2) at modellen får karakteren "inadequate". Dette er begrunnet utfra det faktum at særlig den hydrodynamiske delen krever for lang beregningstid (opp mot ett døgn på en standard bærbar PC), samt at det ikke er innebygde rutiner som enkelt støtter gjentagende modellkjøringer med varierende parametersett.

På tross av påpekte svakheter for vår anvendelse, noen av svakhetene så alvorlige at den ikke burde vært benyttet hvis man skulle konkludere utelukkende basert på kriteriene, ble det

valgt å sette den opp. Dette var også en måte å teste kriteriene på, hvorvidt en anerkjent modell (av enkelte omtalt som en "god" modell uten å referere til en bestemt anvendelse) som ikke passerer kriteriene, allikevel vil gi de svarene vi er ute etter.

Til å simulere forholdene i Storefjorden i Vansjø ble det først valgt ut 4 forskjellige innsjømodellkandidater. Disse var DYRESM linket med CAEDYM, MINLAKE, CE-QUAL-W2 og MyLake (se komplette referanser i Saloranta et al. (2004)). Det ble bestemt å benytte en tilnærming hvor man ønsket å bruke enklere 1-dimensjonale modeller i første omgang, og heller senere anvende mer kompliserte modeller, hvis gode grunner for det skulle komme frem. Av denne grunn ble den 2-dimensjonale modellen CE-QUAL-W2 valgt bort allerede før selve benchmark-kriteriene ble anvendt. MINLAKE måtte også droppes da det viste seg praktisk umulig å få tilgang til modellen. Dermed var det 2 aktuelle kandidater igjen. Det viste seg at ingen av disse kandidatene heller fikk karakteren "good" eller "adequate" på alle 14 kriteriene. DYRESM/CAEDYM feilet på spørsmål 1.2 da den ikke kunne simulere isproduksjon og følgelig ikke kunne kjøres over flere år. MyLake ble bedømt til å være "inadequate" på spørsmålet om testing og dokumentasjon av tidligere anvendelser (1.3), naturlig nok ettersom det var en nokså nyutviklet modell. Etersom MyLake ellers fikk mange gode karakterer, og siden modellkoden var lett tilgjengelig og kjent, ble det bestemt å gå videre med

MyLake i stedet for å forsøke å lete etter andre kandidater. Dette illustrerer det faktum at kriteriene er bare ment til å hjelpe i modellutvelgelsen, og at man selvfølgelig også kan bruke andre sterke argumenter, slik som tidligere kjennskap til modellen, pris, etc. i modellutvelgelsesprosessen.

Den viktigste erfaringen fra modelleringsøvelsen i Vansjø-Hobølvasdraget er at det er vanskelig å kvantifisere når en modell er "god nok". Dette er i utgangspunktet vanskelig selv når modelleringen er gjennomført og statistiske kriterier kan anvendes. I forkant av selve modelleringen er dette følgelig enda vanskeligere. På bakgrunn av dette må derfor anvendelse av kriteriene mer sees på som en bevisstgjøring omkring utvelgelse av modell og gjennomføring av en modelleringsprosess. Kriteriene skal berøre alle sentrale aspekt omkring valg av modell. Erfaringen fra det norske testvassdraget faller derfor sammen med erfaringene fra de andre vassdragene i BMW-prosjektet hvor *kriteriene bør fungere som en stimulan til en dialog mellom modelløren og forvalteren* i forkant av valget og oppsett av modell. Kriteriene er heller ingen garanti for at selve modelleringen blir gjennomført på en god måte. Denne må helst gjennomføres i henhold til god modelleringspraksis, for eksempel som anbefalt av van Waveren et al. (1999) i "Handbook of good modelling practise". Når det gjelder modelleringen med SOBEK og MyLake så gir anvendelsen av disse modellene øket innsikt i prosessene som foregår i nedbørfelt. Det erkjennes imidlertid at anvendelsen av

SOBEK neppe forsvarer medgåtte ressurser ettersom både problemene med å kjøre sensitivitets-/usikkerhetsanalyse og manglende tidsoppløsningen i inngangsdata reduserer nytten av modellen.

Anvendelsen av kriteriene viser seg også å være personavhengig. Vurdering av en og samme modell på nøyaktig den samme problemstillingen i det samme området vil kunne gi forskjellig resultat. Dette er i for seg ikke oppsiktsvekkende, både sett i lys av erfaringen nevnt i forrige avsnitt hvor det er rom for en personlig vurdering omkring hva som er "godt nok", men også at forståelsen av spørsmålene og de alternative svarene ikke nødvendigvis er entydige. Begge disse problemene kan imidlertid reduseres ved hjelp av for eksempel hjelpetekster som følger med spørsmålene.

Det kan også være slik at enkelte spørsmål framstår som mindre viktige eller til og med irrelevante. Et eksempel på dette kan være spørsmålet omkring dokumentasjon av modellen hvis selve modelleringen skal gjennomføres av modellutvikleren selv. Det kan også være at forvalterne kan ha en spesiell og begrunnet tillit til en bestemt modellør, noe som kan redusere viktigheten av dokumentasjon og testing av en modell hvis vedkommende modellør "går god" for den aktuelle modellen.

Til slutt bør det igjen poengteres at for både elve- og innsjømodellen så kan man ikke konkludere med at de på generelt grunnlag er gode eller dårlige (egnet/uegnet), kun at de er blitt evaluert i lys av én bestemt anven-

delse i ett bestemt nedbørfelt eller vannforekomst.

4. Videreutvikling av kriteriesettet

Erfaringer fra testingen av kriteriene innenfor BMW-prosjektet tyder på at kriteriene, slik de er presentert i Saloranta et al. (2003), er godt dekkende og allsidige, men at det er nødvendig med ekstra forklaringer til spørsmålene (hjelpetekst), bedre definisjon av modelleringsterminologi og rollene til de involverte aktørene (modellør, forvalter, legfolk). Det poenget at både modellutvelgelsen og selve modelleringstudiet bør skje gjennom en tett dialog mellom modelløren (produsenten av resultater), forvaltningen og andre involverte (for eksempel de som bli påvirket av beslutningene), bør nok også understrekes sterkere. Et utkast til slike hjelpetekster og en utvidet bruksanvisning til kriteriene foreligger allerede.

Idéene bak benchmark-kriteriene skal videreutvikles i BMW-prosjektet til et interaktivt evalueringsverktøy som etter planen skal implementeres og være tilgjengelig over Internett fra adressen www.rbm-toolbox.net. Fra denne adressen vil det også bli gjort tilgjengelig lister over modeller med detaljerte beskrivelser, man vil kunne evaluere disse modellene gjennom implementeringen av benchmark-kriteriene, og få mer innsikt i de forskjellige sidene av modelleringsverdenen. Prosjektet er inne i sitt siste år, så vi forventer at dette vil være operativt innen begynnelsen av 2005.

Er du interessert i å se på eller teste benchmark-metodikken beskrevet i denne artikkelen, ta gjerne kontakt med en av artikkelforfatterne.

5. Referanser

Benchmark Models for the Water Framework Directive (BMW), (www.vyh.fi/eng/research/euproj/bmw/homepage.htm). EU-prosjekt med kontrakt nr: EVK1-CT-2001-00093.

Dale, V. H. (red.). Ecological modeling for resource management, Springer, New York, 328 sider. 2003.

Lyche Solheim, A., Vagstad, N., Kraft, P., Turtumøygard, S., Skoglund, S., Løvstad, Ø. og Selvik, J.R. Tiltaksanalyse for Morsa, Vansjø-Hobøl-vassdraget: sluttrapport. NIVA rapport nr. 4377-2001, 104 sider. 2001.

Ravetz, J. R. What is Post-Normal Science? Futures 31, 647-654. 1999.

River Basin Manager's Toolbox (www.rbm-toolbox.net).

Saloranta, T. M., Bakken, T.H. og Skiple Ibrek, A. Modelling long-term changes in water quality of Lake Vansjø, Norway: Case story of the application of the models SOBEK and MyLake to Vansjø River Basin. NIVA-rapport under utarbeidelse. 2004.

Saloranta, T. M., Kämäri, J., Rekolainen, S. og Malve, O. Benchmark criteria: A tool for selecting appropriate models in the field of water

management. Environmental Management 32, 322-333, DOI: 10.1007/s00267-003-0069-3. 2003.

van Waveren, R. H., Groot, S., Scholten, H., van Geer, F. C., Wösten, J. H. M., Koeze, R. D. og Noort, J. J. Handbook of good modelling practice. STOWA report 99-05, Dutch dept. of public works, Inst. for inland water management and waste water treatment, report 99.036. 1999.

Anneks A. Benchmark-kriterier (Saloranta et al, 2003)

Dette anneks gjengir benchmark-kriteriene slik de er publisert i Saloranta et. al. (2003). I dette anneks er imidlertid kun spørsmålene gitt, mens det i den opprinnelige publikasjonen også er foreslått kvantitativt og/eller kvalitativt hva som skal til for å oppnå karakterene "god (good)", "tilfredsstillende (adequate)", og "ikke tilfredsstillende (inadequate)".

Spørsmåls-nummer	Spørsmålstekst
1.1	Hvor godt kan modellens utgangsdata relateres til forvaltningsoppgaven?
1.2	Hvor godt samsvarer modellens oppløsning og utstrekning i tid og rom med kravene gitt av forvaltningsoppgaven?
1.3	Hvor godt er modellen blitt testet?
1.4	Hvor komplisert/kompleks er modellen i forhold til den forvaltningsoppgaven som skal analyseres?
1.5	Hvordan er balansen mellom modellens krav til inngangsdata og tilgjengelige data?
1.6	Hvor lett kan modellparameterne identifiseres?
1.7	Hvor lett kan modellens utgangsdata forstås og tolkes?
1.8	Hvor stor grad av aksept har modellen oppnådd i vitenskapelige miljøer og hvor konsistent er modellen med vitenskapelig teori?
2	Hvor godt egnet er modellen til sensitivitets- og usikkerhetsanalyser, og i hvilken grad er dette blitt utført og dokumentert?
3.1	Hvor god er modellens versjonskontroll?
3.2	Hvor god er modellens brukermanual og opplæringsprogram?
3.3	Hvor god er modellens tekniske dokumentasjon?
3.4	Hvor brukervennlig er modellen, hvor stor grad av brukerinteraksjon er mulig og hvor egnet er den for medvirkning/deltagelse av sluttbruker?
3.5	Hvor fleksibel er modellen med tanke på tilpasning og forbedringer/utvidelser?