

WOT-paper 11, December 2011

Achtergronden van deze paper zijn te vinden in WOT-werkdocument 226: Bogaart et al. (2011), en WOT-werkdocument 272: Van Voorn & Walvoort (2011).

Project WOT-04-002-223

ISSN 1879-4688

© 2011

PRI Biometris, Wageningen UR
Alterra Wageningen UR
Planbureau voor de Leefomgeving
WOT Natuur & Milieu, Wageningen UR

Een beoordelingslijst voor de complexiteit van modellen en bestanden

G.A.K. van Voorn	PRI Biometris, Wageningen UR
D.J.J. Walvoort	Alterra Wageningen UR
M. Knotters	Alterra Wageningen UR
P.W. Bogaart	Alterra Wageningen UR
H. Houweling	Wettelijke Onderzoekstaken Natuur & Milieu, Wageningen UR
P.H.M. Janssen	Planbureau voor de Leefomgeving

Bij beleidsevaluaties en om verwachtingen op te stellen, gebruiken organisaties zoals het Planbureau voor de Leefomgeving (PBL) numerieke rekenmodellen en ruimtelijke gegevensbestanden. Eenvoudige modellen en bestanden bevatten maar weinig processen en variabelen en hebben daardoor een beperkte voorspellingswaarde. Een voorbeeld. De bruikbaarheid van een exponentiële groeifunctie om de groei van de wereldbevolking te beschrijven is beperkt. Dit komt omdat in een dergelijk model geen verdere mechanismes zijn verwerkt die de groei kunnen doen veranderen, zoals technische innovatie of onverwachte epidemieën. Het idee is dat het toevoegen van meer processen en variabelen (het maken van een 'complexer' model of bestand) leidt tot meer nauwkeurige uitkomsten, en dus ook tot meer betrouwbare voorspellingen.

Bij dit idee zijn twee kanttekeningen te plaatsen. Ten eerste is het niet per se het geval dat het toevoegen van meer processen en variabelen leidt tot nauwkeuriger uitkomsten of betrouwbaardere voorspellingen. Dit ligt onder andere aan de beschikbaarheid van voldoende gegevens van voldoende kwaliteit. Ten tweede is het de vraag of het wel noodzakelijk is om een model of bestand complexer te maken. Dit hangt vooral af van de gewenste nauwkeurigheid met het oog op de toepassing van het model of bestand.

In deze paper gaan we op zoek naar een manier om te bepalen of de complexiteit van een model of bestand bijdraagt aan het vermogen om voorspellingen te doen met de vereiste nauwkeurigheid voor een bepaalde toepassing. Verder gaan we na of deze complexiteit wordt ondersteund door voldoende gegevens van voldoende kwaliteit.

Complexiteit neemt toe

Door de continue toename van beschikbare en betaalbare rekenkracht en opslagcapaciteit en door verbeterde inzichten zijn modellen en bestanden steeds meer uitgebreid. Met andere woorden, hun complexiteit is toegenomen. Ironisch genoeg bestaat er echter het reële gevaar dat de complexiteit 'doorgeslagen' is. Hiervoor zijn diverse onderliggende factoren te identificeren.

- De 'omdat het kan'-factor: door de toename in computercapaciteit lijkt het (onterecht) dat simulatietijd niet langer limiterend is.
- De 'opschep'-factor: een manager of geldschietter zal mogelijk sneller onder de indruk zijn van een complex model dan van een eenvoudig model.
- Het 'alles erin'-syndroom: uit angst om essentiële punten te missen, worden er zoveel mogelijk details toegevoegd aan het model.
- De gedachte dat modellen als 'kennisbanken' dienen, wat inhoudt dat een model dat meer kennis bevat ook automatisch beter is.
- Paradoxaal genoeg lijken essentiële zaken vaak nog te missen, omdat daar geen of weinig informatie over beschikbaar is.
- Complexe modellen omvatten veel parameters en andere zaken waarvoor gegevens voor de onderbouwing nodig zijn. Deze data kan makkelijk ontbreken.
- Het 'concorde'-effect: investeren in een model zorgt er ook voor dat er een toename is in beschermgedrag om het model in de lucht te houden, ook al laat een afweging tussen kosten en baten zien dat het beter is om het model te vervangen voor een nieuw model (Van Nes & Scheffer, 2005).
- Een gebrek aan duidelijkheid over de eisen die de toekomstige toepassing stelt aan het model kan ertoe leiden dat de grenzen te ruim gekozen worden (Chwif et al., 2000).

- Het 'model op de plank'-fenomeen: in een project-gedreven omgeving is het verleidelijker om een bestaand model aan te passen aan een nieuwe toepassing, dan om een nieuw model te ontwerpen.
- Dit leidt op zijn beurt weer tot het 'Zwitsers zakmes'-fenomeen: een model is gemakkelijk te complex voor elke aparte toepassing, omdat het aan diverse, niet of maar deels overlappende toepassingen is aangepast.

Slak met een te groot huis

Een meer complex model of bestand bevat (als het goed is) meer inzichten in de werking van het gemodelleerde systeem. Echter, er ontstaat daarmee ook een veel grotere databehoeftte om deze processen te kwantificeren. Verder zijn er mogelijkwerwijs factoren geïntroduceerd waarvan maar beperkte kennis beschikbaar is. Daarnaast kleven er allerlei praktische bezwaren aan een toename in complexiteit. Zo is de rekentijd makkelijk groter dan wordt gecompenseerd door de groei in computercapaciteit. Weliswaar verdubbelt de computercapaciteit (de 'wet van Moore') grofweg elk jaar, maar in de praktijk neemt de complexiteit van veel modellen en bestanden toch nog meer toe. Dit verkleint de mogelijkheden om het model te testen en te valideren. Uiteindelijk neemt het vertrouwen in het model of bestand, en daarmee ook de zekerheid dat het gemodelleerde systeem begrepen wordt, weer af. Een weinig flatteuze maar naar ons idee aansprekende metafoor voor deze observaties is dat het model of bestand een soort van slak met een veel te groot geworden huis is geworden.



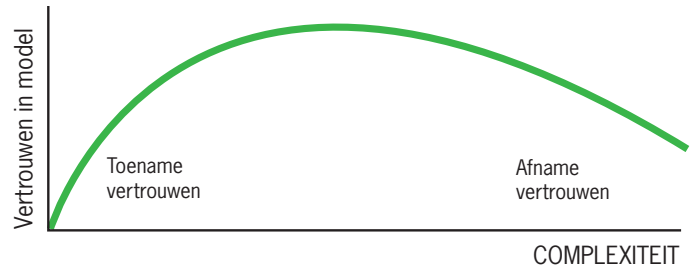
Enigszins gesimplificeerd kan de bovengenoemde ontwikkeling grafisch worden geschetst als een curve, waarbij 'complexiteit' op de ene as, en 'vertrouwen in het model/bestand' op de andere as staat uitgezet, zie figuur 1 (vrij naar Chwif et al., 2000). Deze curve kent een optimum in het vertrouwen in het model of bestand, waarbij vertrouwen – in de basis – de perceptie van een *trade off* is tussen beperkingen in het model of bestand en beperkingen in de data. Gaan we voorbij aan dit maximum door de complexiteit te vergroten, dan neemt het vertrouwen weer af.

Het concept 'evenwicht'

We zoeken naar een manier om voor een model of bestand de juiste balans te bepalen. Deze balans wordt gevormd door:

- voldoende complexiteit om voorspellingen te maken met de vereiste nauwkeurigheid voor een bepaalde toepassing;
- ondersteuning van de complexiteit door voldoende gegevens van voldoende kwaliteit, en voldoen aan praktische randvoorwaarden (zoals rekentijd).

Deze balans noemen we 'evenwicht'. Een model of bestand is dan in 'evenwicht' wanneer het aan alle bovenstaande eisen voldoet. Met andere woorden, het model of bestand is complex genoeg om voor een toepassing uitkomsten binnen een vereiste nauwkeurigheid te genereren. Bovendien is het ondersteund door voldoende gegevens van voldoende kwaliteit, en is de complexiteit geminimaliseerd om zoveel mogelijk aan praktische randvoorwaarden te voldoen. Daarbij moet bijvoorbeeld ook kritisch gekeken worden naar de daadwerkelijke meerwaarde van toegevoegde componenten.



Figuur 1 (vrij naar Chwif et al., 2000): Een klein model of databestand is beperkt toepasbaar omdat het zelden alle relevante processen beschrijft. Een beperkte toename in de complexiteit van model of bestand zal vrijwel zeker leiden tot meer vertrouwen in de kwantitatieve voorspellingen. Een verdere toename in complexiteit gaat gepaard met een nog grotere databehoeftte en heeft praktische bezwaren, wat weer leidt tot afname in het vertrouwen. Een 'optimaal' model of bestand heeft een complexiteit, die bij een gegeven databeschikbaarheid leidt tot de meest betrouwbare voorspellingen en het meeste vertrouwen.

Sleutelrol voor toepassing

De toepassing speelt een sleutelrol binnen het concept 'evenwicht'. Een voorbeeld dat voor Nederlanders zeer tot de verbeelding spreekt is het gebruik van een hydrologisch model om risicoanalyses te maken voor het overstromen van dijken. Wat is nu belangrijk bij het maken van dergelijke analyses? In dit geval moet het model in staat zijn om bij hoogwater de verwachte hoogte van het waterpeil met grote zekerheid te voorspellen, en op het juiste tijdstip. Daarentegen is het in deze analyses totaal niet van belang of de voorspelling van lage waterstanden betrouwbaar is. Het heeft dus weinig zin om veel moeite te doen om een model te maken dat betrouwbare schattingen geeft van lage waterstanden.

Bij het bovenstaande moeten we wel twee kanttekeningen plaatsen. Ten eerste, we herhalen hierbij dat veel modellen en bestanden voor meer toepassingen worden gebruikt. Er zouden dus meerdere 'evenwichten' mogelijk kunnen zijn voor hetzelfde model of bestand. In dat geval betogen wij impliciet dat het model of bestand wellicht door verschillende modellen of bestanden vervangen moet worden, waarbij elk afzonderlijk model of bestand in balans is met een aparte toepassing. Ten tweede, het concept is specifiek gericht op modellen en bestanden in relatie tot toepassingen. Het is dus niet per se relevant voor bijvoorbeeld modellen die bedoeld zijn voor onderzoek, waarin verschillende zaken worden 'uitgeprobeerd'.

Gebalanceerd modelleren

Het concept van gebalanceerd modelleren bestaat al langer, in elk geval in de context van statistische modellen. Met automatische criteria voor modelselectie, zoals het Akaike informatiecriterium (Akaike, 1974),

wordt bijvoorbeeld de nauwkeurigheid (uitgedrukt met een maat voor *goodness of fit*) afgewogen tegen complexiteit (uitgedrukt met een strafterm waarin het aantal parameters in rekening is gebracht). Een complexer model kan de data beter beschrijven (de '*goodness of fit*') dan een eenvoudiger model, maar is gevoeliger voor '*over-fitting*'. Dit laatste komt omdat data altijd fouten en witte (aselecte) ruis bevatten. Proberen om een model perfect te laten passen op data met fouten en ruis leidt tot een complex model met fouten erin. De voorspellingskracht van een dergelijk model is lager dan de voorspellingskracht van een minder complex model zonder die fouten.

Kanttekeningen bij bestaande methoden

Bij de criteria voor automatische modelselectie zijn drie belangrijke kanttekeningen te plaatsen. Ten eerste is de 'complexiteit' in deze context erg strikt gedefinieerd, namelijk als functie van het aantal parameters. Hierin zit een valkuil van 'schijnbare' complexiteit die ontstaat omdat parameters sterk gecorreleerd kunnen zijn. Maar bovenal zullen veel mensen zich bij complexiteit meer iets voorstellen in termen van het aantal processen en interacties dat is meegenomen in het model of bestand. Ten tweede zijn de criteria niet geschikt voor bestanden of alle typen modellen, die PBL en dergelijke instanties gebruiken. Maar nog veel belangrijker is het derde punt. De rol van de toepassing is bij deze criteria onvoldoende benadrukt. Er wordt vooral gekeken of de modelvoorspellingen goed passen bij de gegevens. Er wordt echter niet beoordeeld wat de relevantie van de gegevens voor de toepassing is. Laten we nog even het 'dijken'-voorbeeld aanhalen. Er kan veel moeite worden gedaan om een model passend te krijgen op gegevens van lage waterstanden. Maar voor de toepassing was dat niet nodig, en dus hoeven we die moeite ook niet te doen.

Hulpmiddel: beoordelingslijst

Uit het voorgaande concluderen we dat er een methode moet worden ontwikkeld, die zowel met alle facetten van complexiteit rekening houdt, als met de toepassingen van het model. Wij hebben een beoordelingslijst opgezet (genaamd EMC= Evaluatie Modelcomplexiteit), die hierbij kan dienen als hulpmiddel. Met een dergelijke lijst is een flexibeler beoordeling van modelcomplexiteit mogelijk dan met automatische criteria voor modelselectie. Weliswaar zijn deze laatste criteria veel kwantitatiever, maar ze zijn weinig flexibel en slechts toepasbaar op specifieke modeltypen. Voor statistische modellen gelden bijvoorbeeld andere criteria dan voor fysisch-gebaseerde proces-responsmodellen. Daarnaast zou het gebruik van de lijst ook kunnen leiden tot adviezen om een model of bestand te verbeteren. Hierbij moet dan wel worden beoordeeld hoe ver een model of bestand uit evenwicht is, en hoe kwalijk dat is.

Opbouw vragenlijst

Hoe zou deze vragenlijst er uit moeten zien? In elk geval zijn er een aantal criteria waarop het model of bestand wordt beoordeeld. Verder is besloten om de beoordelingslijst in elk geval grofweg de modelleercyclus te laten volgen. Alle belangrijke stappen in het modelleerproces komen dan aan bod.

Het aantal vragen op de lijst is ook van belang. Er zit eveneens een zeker 'optimum' in de 'complexiteit' van de vragenlijst. Een lijst die uit zeer veel vragen bestaat, nodigt niet uit tot invullen en suggereert kennis over het onderwerp 'modelcomplexiteit' die er (nog) niet is. Een lijst met weinig vragen is waarschijnlijk veel te algemeen, waardoor er evenmin voldoende inzicht wordt verkregen in het gewenste 'evenwicht'. De lijst moet dus uit een gebalanceerd aantal vragen bestaan die voldoende begrensd zijn. Duidelijk moet zijn welke informatie moet worden geleverd. Maar de lijst moet ook afdoende flexibel zijn om ruimte te geven om uit te wijden over details.

De vragen moeten daarnaast een enigszins confronterend karakter hebben, bv. vraag A: 'Wat heb je nodig voor X?', en dan vraag B: 'Wat is er gedaan aan X?'. Dit zorgt ervoor dat zaken expliciet gemaakt worden, en dat onderzoekers worden gestimuleerd om na te denken hoe ze hun model of bestand kunnen verbeteren in relatie tot de beoogde toepassing(en).

Appendix

Het prototype van de beoordelingslijst is weergegeven in de appendix bij deze paper. Deze lijst is een geactualiseerde versie van de lijst die Bogaart *et al.* (2011) hebben ontwikkeld. De oudere versie van de lijst is getest aan de hand van casussen die zijn gekozen uit de wetenschappelijke literatuur, en is daarnaast ter beoordeling voorgelegd aan reviewers. De resultaten hiervan zijn opgenomen in Van Voorn *et al.* (2011), en zijn gebruikt bij het opstellen van de nieuwe versie van de lijst EMC v1.0 die in deze paper wordt gepresenteerd.

Discussie

EMC v1.0 wordt verder getest aan de hand van casussen. Naast de oude casussen gebruiken we ook nieuwe casussen, die voor PBL relevant zijn. Een dergelijke casus is MetaSWAP, die is geselecteerd in verband met zijn rol in het Nationaal Hydrologisch Instrumentarium (NHI). Bij het testen wordt onder andere beoordeeld of vragen duidelijk, eenduidig én nuttig zijn, en of er nog zinnige vragen ontbreken. Verder moet ook het nut en de toegevoegde waarde van de lijst in zijn geheel blijken. We willen graag weten of de lijst werkelijk inzicht oplevert in het 'evenwicht' van modellen en bestanden, en nuttige informatie aandraagt voor potentiële verbeteringen. Afhankelijk van de uitkomsten van deze testen zal de lijst nogmaals worden geactualiseerd.

Aanpak testen lijst

Omdat het testen 'dubbelzijdig' werkt (de lijst én de modellen of bestanden worden getest), kan het invullen van de lijst op dit moment het beste gebeuren door een klein team van mensen tegelijkertijd. Daaronder valt in elk geval een modelleur die betrokken is bij de ontwikkeling van het model of bestand, een modelgebruiker die inzicht in de toepassing heeft, en één van de auteurs van deze paper als samenstellers van de lijst om eventueel vragen te verhelderen, en om feedback van de andere teamleden te verwerken. Verder is dit teamverband van belang omdat het invullen is gebaseerd op *expert judgement*. Het is goed mogelijk dat verschillende betrokkenen verschillende visies hebben op bepaalde punten, die op deze manier aan het licht komen. Het invullen zal voorlopig nog *ad hoc* gebeuren, dus na de ontwikkeling van een model.

Evaluatie casussen

Na het testen zal de uiteindelijke versie van de lijst worden gebruikt om de Natuurplanner en andere modellen en bestanden in het Wageningen UR/PBL-instrumentarium te evalueren. De bedoeling is dan dat de lijst door een klein team van modelleers en modelgebruikers kan worden ingevuld, en dat dat gebeurt ten tijde van de (verdere) modelontwikkeling. De resultaten kunnen dan iteratief worden gebruikt voor de modelontwikkeling.

Verder vooruitzicht

We verwachten dat het mogelijk en nuttig is om de functionaliteit van de lijst uit te breiden. Niet alleen wordt dan geëvalueerd of het model of bestand in evenwicht is, maar ook hoe de uitkomsten van de evaluatie kunnen leiden tot adviezen om een model of bestand te verbeteren. Verder zien wij voor de beoordelingslijst een rol in de ontwikkeling van een nog nader te definiëren kwaliteitsniveau binnen de kwaliteitsborging van modellen en bestanden aan Wageningen UR. Voorwaarde bij dit alles is wel dat de lijst zelf niet te complex wordt en daardoor in een 'trage slak' verandert. Een toekomstig punt van onderzoek is om te kijken of dat het aantal vragen op de lijst niet gereduceerd kan worden.

Bewustzijn scheppen

Het eerste doel van de lijst is evenwel het scheppen van bewustzijn. Wij verwachten niet dat per se het juiste evenwicht wordt gevonden voor alle toepassingsgerichte modellen en bestanden, of dat er *überhaupt* maar één dergelijk evenwicht bestaat. Wel zal hopelijk de wat confronterende aard van de lijst het onderwerp 'modelcomplexiteit' meer op de kaart zetten bij de verschillende betrokkenen, zoals modelleers, modelgebruikers, programmaleiders, en beleidsmakers. Verder hopen wij dat de lijst hulp kan bieden bij de ontwikkeling van breed gedragen en goed toepasbare modellen en bestanden.

Literatuur

- Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, AC-19, 716-723.
- Bierkens, M. F. P., P.A. Finke & P. de Willigen (2000). Upscaling and Downscaling Methods for Environmental Research. Kluwer Academic Publishers, Dordrecht.
- Bogaart, P.W., G.A.K. van Voorn & W. Akkermans (2011). Evenwichtsanalyse modelcomplexiteit – een verkennende studie. WOt-werkdocument 226. WOt Natuur & Milieu, Wageningen UR, Wageningen
- Chwif, L., M.R.P. Barretto & R.J. Paul (2000). On simulation model complexity. Proceedings of the 2000 Winter Simulation Conference, 449-455, J.A. Joines, R.R. Barton, K. Kang & P.A. Fishwick, eds.
- Nes, E.H. van & M. Scheffer (2005). A strategy to improve the contribution of complex simulation models to ecological theory. Ecological Modelling 185, 153-164.
- Voorn, G.A.K. van & D.J.J. Walvoort (2011). Evaluation of an evaluation list for model complexity. WOt-werkdocument 272. WOt Natuur & Milieu, Wageningen UR, Wageningen.

Colofon

De reeks 'WOt-papers' is een uitgave van de Wettelijke Onderzoekstaken (WOT) Natuur & Milieu, onderdeel van Wageningen UR. Een WOt-paper bevat resultaten van afgerond onderzoek op een voor de doelgroep zo toegankelijk mogelijke wijze. De maatschappelijke discussie waarbinnen en waarom het onderzoek is uitgevoerd, komt daarbij nadrukkelijk aan de orde, evenals de beleidsrelevantie en mogelijk de wetenschappelijke relevantie van de resultaten.

Onderzoeksopdrachten van de WOT Natuur & Milieu worden gefinancierd door het Ministerie Economische Zaken, Landbouw & Innovatie (EL&I).

Deze paper is gemaakt conform het Kwaliteitshandboek van de unit WOT Natuur & Milieu.

Contactgegevens auteurs

G.A.K. van Voorn
PRI Biometris, Wageningen UR
Postbus 100, 6700 AC Wageningen
Tel: (0317) 48 60 01; e-mail: info.pri@wur.nl

D.J.J. Walvoort, M. Knotters & P.W. Bogaart
Alterra Wageningen UR
Postbus 47, 6700 AA Wageningen
Tel: (0317) 48 07 00; e-mail: info.alterra@wur.nl

H. Houweling
Wettelijke Onderzoekstaken Natuur & Milieu Wageningen UR
Postbus 47, 6700 AA Wageningen
Tel: (0317) 48 54 71; e-mail: info.wnm@wur.nl

P.H.M. Janssen
Planbureau voor de Leefomgeving (PBL)
Postbus 303, 3720 AH Bilthoven
Tel : (030) 274 27 45; e-mail: info@pbl.nl

Wettelijke Onderzoekstaken Natuur & Milieu

Postbus 47
6700 AA Wageningen
t (0317) 48 54 71
f (0317) 41 90 00
info.wnm@wur.nl
www.wotnatuurenmilieu.wur.nl

Alle rechten voorbehouden. Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke wijze dan ook, zonder voorafgaande schriftelijke toestemming van de uitgever.

Appendix bij WOt-paper 11: Beoordelingslijst Modelcomplexiteit (EMC) v. 1.0

Vraag 1: BASALE GEGEVENS

- Geef aan: model/bestand, naam, versienummer, revisienummer, uitgiftedatum.
- Onder model verstaan we de broncode + invoergegevens.

Vraag 2: DOEL & TOEPASSINGEN

- A. Wat was het doel van het bestand/model?
- B. Wat was het beoogde toepassingsgebied, en waaruit blijkt dit?
- C. Wat zijn de toepassingen nu? Heeft het model/bestand een rol in een modelketen? Zo ja, welke?
- D. Overlapt het beoogde toepassingsgebied alle daadwerkelijke toepassingen (ja/deels/nee)?
- E. Indien relevant, in hoeverre overlappen de verschillende toepassingen met elkaar (goed/matig/slecht)?

Vraag 3: SYSTEEMANALYSE

- A. Geef een systeemanalyse. Dit kan bv. een verwijzing naar een rapport zijn. Aspecten die daarin aan bod komen zouden moeten zijn: Wat zijn de belangrijkste attributen en processen van het systeem? Welke terugkoppelingen zijn aanwezig? Hoe zijn de grenzen van het systeem bepaald of gedefinieerd? Hoe wordt met relevante attributen en processen buiten deze grenzen omgegaan, bv. randvoorwaarden, constanten, of simpelweg genegeerd? Over welke processen, attributen, terugkoppelingen, etc. bestaat er onzekerheid? Hoe is die onzekerheid bepaald?

Beschouw nu per toepassing:

- B. Welke systeem-analytische aspecten (attributen, processen, terugkoppelingen, etc.) zijn (direct) relevant voor de toepassing? En welke in mindere mate of niet? En hoe is die relevantie bepaald?
- C. Zijn alle voor de toepassing belangrijke aspecten als gegeven in B meegenomen in het model/bestand? En welke niet?
- D. Zijn er minder of niet belangrijke aspecten meegenomen in het model/bestand (ja/weinig/geen)?
- E. Beoordeel de mate van 'evenwicht' m.b.t. de systeemanalyse, gebaseerd op de antwoorden op 3C en 3D (goed/matig/slecht).

Vraag 4: CONCEPTUEEL MODEL of ONTWERPMODEL

- A. Geef het conceptuele model (voor een model) of ontwerpmodel (voor een bestand). Dit kan bv. een verwijzing naar een rapport zijn. Het conceptuele model legt op relatief informele wijze de relaties tussen componenten vast. Aspecten die aan bod kunnen komen zijn:

Welke aannamen en vereenvoudigingen zijn er gemaakt? Wat zijn de ruimtelijke dimensies? Welke ruimtelijk-temporele aggregatieniveaus worden gebruikt? Welke zaken spelen een rol voor het bestand? Wat wordt er voor (des)aggregatie van data gebruikt? Voor modellen: Welk modeltype is gebruikt?

Beschouw nu per toepassing:

- B. Op welke (ruimtelijk-temporeel) aggregatieniveaus zijn antwoorden gewenst, gezien de toepassing? Welke ruimtelijke dimensies en welk modeltype/ontwerpmodel zijn gewenst?
- C. Komen de ruimtelijke dimensies, aggregatieniveaus, etc. overeen met die vanuit de toepassing zijn gewenst (ja/matig/niet)?

Vraag 5: DATA

- A. Welke gegevens zijn benodigd, gezien de toepassing, en motiveer? Wees specifiek m.b.t. resolutie, nauwkeurigheid, schaal, etc., en vermijd algemene en ongemotiveerde opmerkingen als "file X moet gebruikt worden".
- B. Welke gegevens zijn beschikbaar, en in welke mate?
- C. Wat is uiteindelijk de invoer van het model of bestand?
- D. Komt de invoer overeen met de door de toepassing gewenste gegevens (ja/deels/nee)? Vergelijk de antwoorden op 5A, 5B en 5C. Let hierbij sterk op eenheden en de drie elementen van schaal (coverage, support, extent; zie Bierkens *et al.*, 2000; Bogaart *et al.*, 2011). Heeft de data de juiste dimensies en eenheden? Zo nee, is er een (des)aggregatiemethode voorhanden? Schaalproblemen doen zich overal voor door de 'vertaling' van attributen naar metingen, die gediscretiseerd zijn in tijd en ruimte, en onderhevig zijn aan allerlei bronnen van meetruis.

Vraag 6: FORMEEL MODEL (niet voor databestanden)

Gebruik de antwoorden op vragen 5A tot 5D om vragen 6B tot 6D te antwoorden.

- A. Geef het formeel model, en motiveer de keuze of verwijs naar paginanummers uit een gepubliceerd rapport of artikel. Het formeel model is de omzetting van concept naar bijvoorbeeld wiskundige vergelijkingen of een set rekenregels. De keuze van formeel model hangt af van de beoogde toepassing en de data, en wordt mede bepaald door de gestelde onzekerheidsmarges.
- B. Welke gegevens zijn benodigd voor dit formele model?
- C. Welke gegevens zijn beschikbaar, en in welke mate?
- D. Zijn de gegevens die benodigd én beschikbaar zijn ook gebruikt? Waarom wel/niet?

Vraag 7: NUMERIEK REKENMODEL of BESTAND

- A. Geef het numeriek rekenmodel/gegevensbestand (de implementatie in software), en motiveer de keuze of verwijfs naar paginanummers uit een gepubliceerd rapport of artikel.
- B. Beschrijf de verificatie of verwijfs naar paginanummers uit een gepubliceerd rapport of artikel. Verificatie moet niet verward worden met validatie. De eerste controleert de omzetting van formeel model naar code, de tweede toetst de resultaten van de code aan de werkelijkheid.
- C. Beschrijf en motiveer de keuze voor numerieke rekenmethoden, en geef de gebruikte discretisatie of verwijfs naar paginanummers uit een gepubliceerd rapport of artikel. Veelal bestaat er bij numerieke rekenmethoden een afweging tussen rekensnelheid en –nauwkeurigheid.

Vraag 8: SCHEMATISERING

De schematisering is bv. het indelen van de bodem in verschillende lagen ('stratificatie'), of het clusteren van landgebruikstypen. De mate van detail die bij schematisering wordt gevolgd is afhankelijk van de toepassing, maar moet wel voldoende ondersteund worden door de data.

- A. Indien van toepassing, beschrijf de schematisering, of geef relevante en specifieke referenties?
- B. Wordt de schematisering voldoende ondersteund door de data (ja/matig/nee)? Motiveer.

Vraag 9: GEVOELIGHEDEN, ONZEKERHEDEN & OPBOUW CODE

Deze vraag is gericht op het bepalen welke parameters, forcing, invoer, delen van de code, etc. belangrijk zijn en welke overbodig zijn, en welke onderdelen een grote onzekerheid introduceren.

- A. Zijn er gevoeligheids-/onzekerheidsanalyses uitgevoerd? Zo ja, beschrijf deze of verwijfs naar paginanummers uit een gepubliceerd rapport of artikel en analyseer de resultaten. Sommige factoren kunnen weinig van belang blijken voor de toepassing, zodat er weinig moeite hoeft te worden gedaan om deze goed te kalibreren, of gebruikt kunnen worden om het model te vereenvoudigen. Andere factoren kunnen juist essentieel blijken.
- B. (niet voor databestanden) Welke numerieke integratiemethode is gebruikt? Bespreek het prestatievermogen van de methode.
- C. Wat is de rol van elke module van het model/bestand, bv. een gewasgroei-deelmodel als onderdeel van een hydrologisch model, of stappen in het stroomschema of de scripts van een bestand? Bespreek dat in relatie met de toepassing. Zijn er onderdelen die overbodig zijn? Missen er nog onderdelen? Motiveer waarom.
- D. Beoordeel de complexiteit in termen van rekentijd en efficiëntie. Wat is de duur van een run? Hoeveel (tussentijdse) uitvoer is er? Hoe groot is de datastroom? Zou de code sneller/efficiënter kunnen, bv. door selectie, aggregatie, andere methode, andere modulaire opbouw, etc.?

Vraag 10: KALIBRATIE (niet voor databestanden)

De kalibratie is het proces waarbij de parameters, begincondities, etc. van het model of bestand van waarden worden voorzien. Kalibratie is vaak gebaseerd op gevoeligheidsanalyses (zie vorige vraag).

- A. Indien beschikbaar, beschrijf de uitgevoerde kalibraties en motiveer de keuze, of verwijfs naar paginanummers uit een gepubliceerd rapport of artikel. Geef ook aan welke doelfuncties (lokaal, globaal, deterministisch, stochastisch) zijn gekozen, of dat de betrouwbaarheid van de gekalibreerde parameters wordt gekwantificeerd, hoe, op welke data, of en hoe de betrouwbaarheidsintervallen zijn bepaald, etc. met eventuele verwijfsingen naar de literatuur. Let op over-fitting, meervoudige lokale optima, en identificeerbaarheidsproblemen.
- B. Welke eisen stelt de toepassing aan de nauwkeurigheid waarmee parameters, etc. wordt bepaald. Let ook op over-fitting.
- C. Welke mogelijkheden leveren de data om waarden van parameters, etc. te bepalen. Denk hierbij aan resolutie, nauwkeurigheid, aantal gegevens, het aggregatieniveau van de gegevens, en let op over-fitting.

Vraag 11: VALIDATIE

Wij definiëren hier een validatie als een toets of het model of bestand een redelijke representatie is van het werkelijke systeem dat wordt gemodelleerd. Dit geschiedt op basis van een set onafhankelijke gegevens (gegevens die nog niet in een eerder stadium bij het model leren gebruikt zijn).

- A. Indien beschikbaar, beschrijf de gedane validatiestudies of verwijfs naar paginanummers uit een gepubliceerd rapport of artikel.
- B. Onderbouw/beoordeel de waarde van de validatie(s). Hoe relevant zijn de verschillende validatiestudies voor de verschillende toepassing(en) van het model of bestand? Zijn alle toepassingen voldoende afgedekt door de studies? Is ook de betrouwbaarheid van de validatie gekwantificeerd? Welke toepassingen moeten opnieuw worden gevalideerd?

Vraag 12: SAMENVATTING en ALGEMEEN OORDEEL

Dit laatste onderdeel is gericht op het geaggregeerde resultaat van de antwoorden op de verschillende vragen en de meningen van de verschillende betrokkenen, en kan indirect ook tot adviezen leiden ter verbetering van het model of bestand.

- A. Beoordeel de mate van vertrouwen in het model/bestand in het algemeen, mede gebaseerd op wat er is gedaan aan testen, verificatie, kalibratie, validatie, gevoeligheids- en onzekerheidsanalyses, en motiveer deze.
- B. Zijn er veel missende zaken? Zijn er veel vragen waar de zaken als 'te eenvoudig' zijn beoordeeld? Motiveer.
- C. Zijn er veel overbodige componenten? Zijn er veel onderdelen waar het eenvoudiger zou kunnen? Motiveer.
- D. Is er voldoende datasteun, en waaruit blijkt dat? Motiveer
- E. Welke specifieke suggesties voor verbetering en data-aanlevering komen naar voren uit deze analyse?