

Evenwichtsanalyse modelcomplexiteit

Een verkennende studie

P.W. Bogaart
G.A.K. van Voorn
L.M.W. Akkermans

werkdocumenten



wot

Wettelijke Onderzoekstaken Natuur & Milieu



WAGENINGENUR
For quality of life

Evenwichtsanalyse modelcomplexiteit

De reeks 'Werkdocumenten' bevat tussenresultaten van het onderzoek van de uitvoerende instellingen voor de unit Wettelijke Onderzoekstaken Natuur & Milieu (WOT Natuur & Milieu). De reeks is een intern communicatiemedium en wordt niet buiten de context van de WOT Natuur & Milieu verspreid. De inhoud van dit document is vooral bedoeld als referentiemateriaal voor collega-onderzoekers die onderzoek uitvoeren in opdracht van de WOT Natuur & Milieu. Zodra eindresultaten zijn bereikt, worden deze ook buiten deze reeks gepubliceerd.

Dit werkdocument is gemaakt conform het Kwaliteitshandboek van de WOT Natuur & Milieu en is goedgekeurd door Harm Houweling (deel)programmameider WOT Natuur & Milieu.

WOT-werkdocument **226** is het resultaat van een onderzoeksopdracht van het Planbureau voor de Leefomgeving (PBL), gefinancierd door het Ministerie van Economische Zaken, Landbouw & Innovatie (EL&I). Dit onderzoeksrapport draagt bij aan de kennis die verwerkt wordt in meer beleidsgerichte publicaties zoals Balans van de Leefomgeving en thematische verkenningen.

Evenwichtsanalyse modelcomplexiteit

Een verkennende studie

P.W. Bogaart

G.A.K. van Voorn

L.M.W. Akkermans

Werkdocument 226

Wettelijke Onderzoekstaken Natuur & Milieu

Wageningen, mei 2011

Referaat

Bogaart, P.W., G.A.K. van Voorn & L.M.W. Akkermans, 2011. *Evenwichtsanalyse modelcomplexiteit; een verkennende studie*. Wageningen, Wettelijke Onderzoekstaken Natuur & Milieu, WOt-werkdocument 226. 49 blz.; 40 ref.; 1 bijl.

Dit werkdocument omvat een verkennende studie om de mogelijkheden te evalueren van de verhouding tussen de complexiteit van een model of bestand, de ondersteunende data, en de toepassing. Dit verhoudingsconcept wordt als 'evenwicht' aangeduid. Doel is een reductie in de onzekerheid in voorspellingen door modellen en bestanden. In dit document wordt een prototype van een evaluatielijst voorgesteld. Deze lijst is gebaseerd op de relevante punten die een rol spelen bij evenwicht. De lijst bestaat uit twee sublijsten, die elk uit onderdelen met vragen bestaan. De ene sublijst is gebaseerd op de bestaande evaluatielijst voor 'status A', die gebruikt wordt in de kwaliteitsborging bij de WOT Natuur & Milieu onderdeel van Wageningen UR. De andere sublijst omvat relevante kernvragen voor 'evenwicht' die nog niet operationeel zijn. Elk onderdeel is gekoppeld aan een fase in de modelleercyclus, behalve het overkoepelende onderdeel 'schaal'. De discussie beschrijft het projectvervolg, waarin wordt beoogd om de evaluatielijst te toetsen aan casussen en de kennis van experts.

Trefwoorden: Modelcomplexiteit, optimaal model, kwaliteit, modelonzekerheid

Auteurs

P.W. Bogaart – Centrum Water & Klimaat, Alterra

G.A.K. van Voorn & L.M.W. Akkermans – Biometris, PRI

©2011 **Alterra Wageningen UR**

Postbus 47, 6700 AA Wageningen

Tel: (0317) 48 07 00; fax: (0317) 41 90 00; e-mail: info.alterra@wur.nl

Biometris, PRI Wageningen UR

Postbus 100, 6700 AC Wageningen

Tel: (0317) 48 07 98; fax: (0317) 48 35 54; e-mail: biometris@wur.nl

De reeks WOt-werkdocumenten is een uitgave van de unit Wettelijke Onderzoekstaken Natuur & Milieu, onderdeel van Wageningen UR. Dit werkdocument is verkrijgbaar bij het secretariaat. **Het document is ook te downloaden via www.wotnatuurenmilieu.wur.nl.**

Wettelijke Onderzoekstaken Natuur & Milieu, Postbus 47, 6700 AA Wageningen

Tel: (0317) 48 54 71; Fax: (0317) 41 90 00; e-mail: info.wnm@wur.nl; Internet: 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 andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever. De uitgever aanvaardt geen aansprakelijkheid voor eventuele schade voortvloeiend uit het gebruik van de resultaten van dit onderzoek of de toepassing van de adviezen.

Woord vooraf

Dit werkdocument bevat de resultaten van het eerste deel van het onderzoek in het project "Evaluatie modelcomplexiteit". Een deel van het werk is tot stand gekomen na discussies met diverse mensen, die we hier graag willen bedanken. Dank aan Peter Janssen, Harm Houweling, Piet Groenendijk, Hans Stigter, Karel Keesman, Bob Kooi, Frank van der Bolt, Peter Leffelaar en Bas Kooijman voor hun opmerkingen in het kader van dit project.

*Patrick Bogaart
George van Voorn
Wies Akkermans*

Inhoud

Woord vooraf	5
Samenvatting	9
1 Inleiding	11
2 Evenwichtsanalyse	13
2.1 Definitie evenwicht	13
2.2 Ontwikkeling hulpmiddel: Evaluatielijst	15
2.2.1 Systeemanalyse	15
2.2.2 Schaal	16
2.2.3 Conceptueel model	17
2.2.4 Formeel model	19
2.2.5 Numeriek model	19
2.2.6 Schematisering	20
2.2.7 Gevoeligheidsanalyse	21
2.2.8 Kalibratie	22
2.2.9 Validatie en kruisvalidatie	24
2.3 Gestructureerde evenwichtsanalyse	25
2.3.1 Gevoeligheidsanalyse	25
2.3.2 Foutenanalyse	26
2.3.3 Stapsgewijze verfijning	26
2.4 Evenwicht en metamodellen	26
3 Evenwichtsanalyse en status A	27
3.1 Uitgangspunten	27
3.2 Vergelijking met status A	27
4 Samenstelling van de evaluatielijst	31
4.1 Uitgangspunten	31
4.2 Prototype evaluatielijst deel 1	31
4.2.1 Algemene vragen ter inleiding	32
4.2.2 Model	32
4.2.3 Data, invoer en uitvoer	33
4.2.4 Toepassing	35
4.2.5 Synthese	35
4.3 Prototype evaluatielijst deel 2	36
5 Discussie en conclusies	39
5.1 Selectie casus voor toetsing prototype	39
5.2 Aanpak toetsen prototype evaluatielijst	40
5.3 Discussiepunten voor verder onderzoek	41
5.4 Conclusie	42
Literatuur	43
Bijlage 1 Overeenstemmingsmaten	45

Samenvatting

Dit werkdocument is een verslag van het werk dat gedaan is voor het WOt-project “Evaluatie modelcomplexiteit 2009” (PRI nr. 3320010209, Alterra pr.nr. 5235784-01), in samenwerking met het Planbureau voor de Leefomgeving (PBL). PBL maakt veel gebruik van modellen en ruimtelijke databestanden die bij Wageningen UR zijn ontwikkeld. Deze modellen en bestanden worden over het algemeen steeds complexer, omdat de meeste modellen voor steeds meer toepassingen worden gebruikt. Hierdoor nemen onzekerheden in de voorspellingen van modellen en de bestanden toe, alsmede praktische zaken als rekentijd, terwijl de mogelijkheid voor gebruikers vermindert om zinnig met deze modellen en bestanden om te gaan.

Het onderzoek in dit rapport richt zich op de vraag hoe de complexiteit van deze modellen en bestanden zich verhoudt tot de beschikbaarheid van gegevens, en tot de toepassingen van deze modellen en bestanden. Er wordt aangenomen dat er voor elk model of bestand een soort van ‘evenwicht’ bestaat tussen de complexiteit van een model of bestand, de ondersteuning door gegevens, en de toepassingen van het model of bestand. Het doel van dit project is het ontwikkelen van een hulpmiddel, waarmee dit evenwicht per model of bestand zo goed mogelijk kan worden benaderd. De gedachte daarbij is dat een model of bestand dat in evenwicht is, zo eenvoudig als mogelijk is voor de gebruiker, de minst mogelijke rekentijd beslaat, maar wel de toepassing zo goed mogelijk dient.

In dit werkdocument wordt daarom verkend welke aspecten er aan modelcomplexiteit zijn, en hoe de toepassing die modelcomplexiteit kan beïnvloeden. Vervolgens wordt er een evaluatielijst voorgesteld, die zou kunnen dienen als hulpmiddel voor het beoordelen van modellen en bestanden in het licht van het bovenbedoelde evenwicht. Daarna wordt kort de evaluatielijst uit de kwaliteitscontrole van de WOT Natuur & Milieu, het zogenoemde status A, besproken. De bedoeling daarachter is dat de evaluatielijst voor modelcomplexiteit uiteindelijk een soortgelijke vorm moet krijgen. Ten slotte wordt er een plan voorgesteld voor het projectvervolg in 2010 en 2011, onder meer met de bedoeling om de voorgestelde evaluatielijst modelcomplexiteit te toetsen en eventueel te verbeteren.

1 Inleiding

Het Planbureau voor de Leefomgeving (PBL) doet een toenemend beroep op numerieke rekenmodellen en ruimtelijke gegevensbestanden om kennisvragen vanuit ministeries te beantwoorden, en om de effecten van beleidsimplementaties te onderzoeken.

Aanleiding

Hoewel er veel verschillen zijn tussen deze modellen en bestanden, hebben ze met elkaar gemeen dat het primaire doel voor PBL-doeleinden het doen van bruikbare kwantitatieve uitspraken/projecties is. Hierin verschillen deze modellen principieel van 'toy models', die vooral bedoeld zijn om inzicht te verkrijgen in de effecten van processen. Evenmin is het primaire belang dat deze modellen mechanistisch volledig correct onderbouwd zijn. De nadruk ligt er vooral op het model zodanig te kalibreren, dat het in elk geval binnen het beoogde toepassingsgebied betrouwbare resultaten geeft. Veelal houdt dit in dat een model op beperkte tijdschaal aannemelijke extrapolaties kan maken, of dat een bestand een betrouwbaar en bruikbaar kaartbeeld op een bepaalde schaal oplevert.

Voor de kalibratie van een model zijn data nodig. Deze data moeten niet alleen talrijk genoeg zijn, maar ook voldoende informatief om de verschillende modelparameters van betrouwbare waarden te voorzien. Verder moet er (liefst) onafhankelijke data zijn waarmee het gekalibreerde model gevalideerd kan worden binnen zijn toepassingsgebied. In veel gevallen is het echter onduidelijk of er voldoende data zijn voor de modeltoepassing, en wat de kwaliteit van deze data is.

Een ander punt voor de numerieke rekenmodellen die hierboven beoogd zijn, is dat veel van deze modellen over veel detail beschikken, in een poging om zo goed mogelijk inzetbaar te zijn voor steeds meer PBL-toepassingen. Behalve conceptueel (vanuit een wetenschappelijk oogpunt), is de grote mate van detaillering ook vanuit een praktisch standpunt een ongewenste situatie. Over het algemeen neemt de benodigde rekenkracht alsmaar toe met een toename in modeldetails en modules om data te verwerken, en dit resulteert vooral in zeer lange rekentijden. Dit betekent ook dat het onaantrekkelijk wordt om nieuwe berekeningen te maken na elke nieuwe modelaanpassing, wat eigenlijk wel nodig is om de kwaliteit van het model en zijn uitkomsten te garanderen.

In de praktijk is er nu een aantal modellen in het assortiment aan modellen en bestanden, gebruikt door PBL, dat als 'te groot' of 'te complex' kan worden geclassificeerd. Om aan deze bezwaren tegemoet te komen worden er nu ook zogenaamde 'metamodellen' ontwikkeld, kleinere varianten van de numerieke rekenmodellen, waarbij het doel is om tijd en rekenwerk te besparen. Dit zou ten goede kunnen komen aan de kwaliteit van de modellen, en dus ook de betrouwbaarheid van de gemaakte voorspellingen. Daar staat tegenover dat men deze metamodellen goed aan de originele modellen zou moeten toetsen, om te kijken of ze inderdaad, in elk geval voor de beoogde toepassing, nagenoeg dezelfde uitkomsten geven als de originele modellen.

Onderzoeksvraag

In dit werkdocument wordt nu geïnventariseerd hoe deze concepten, de kwaliteit van de data en de hoeveelheid data, de complexiteit van het model, en het beoogde toepassingsgebied van het model, zich tot elkaar verhouden. Het onderzoek richt zich allereerst op de vraag, of dat er een 'evenwicht' kan bestaan tussen deze drie concepten, en zo ja, of dat dan voor alle

typen modellen hetzelfde is. Daarna is de vraag hoe een dergelijk evenwicht bereikt zou kunnen worden, en wat er nodig is om een model of bestand in evenwicht te krijgen. De hier gevolgde aanpak is om een evaluatielijst te ontwikkelen, waarmee men in staat is om te beoordelen of een model in de buurt van het evenwicht komt, en zo nee, op welke punten verbeteringen zouden kunnen plaatsvinden. Deze lijst moet uiteindelijk in te vullen zijn door een ieder die kennis van modelleren heeft, of het nu een gebruiker van het model, een reviewer, of de modelontwikkelaar zelf betreft. De lijst moet ook in staat zijn om voor een gebruiker van de lijst duidelijk te maken of er nog eventueel veel verbeterd moet worden aan het model om wel in evenwicht te zijn.

Doelstelling

Het doel in dit werkdocument is het verkrijgen van inzicht in de problematiek wat betreft evenwicht in modelcomplexiteit, modeltoepassing, en databeschikbaarheid en datakwaliteit, en dit inzicht gebruiken om een evaluatielijst te ontwikkelen, die het evenwicht van PBL-modellen kan beoordelen. Het eerste onderzoekspunt is verkennend onderzoek naar de aandachtspunten binnen het beoogde gebied. Men moet hierbij denken aan welke begrippen vaak gehanteerd worden, en welke problematiek er is. Een tweede doel is de ontwikkeling van een evaluatielijst, op basis van de uitkomsten van dit verkennend onderzoek, waarmee redelijkerwijs kan worden beoordeeld of een model of bestand in evenwicht is. Het laatste doel is de ontwikkeling van een onderzoekslijn voor verder onderzoek, waaronder het toetsen en aanpassen van het prototype van de evaluatielijst. De ontworpen evaluatielijst zal getoetst moeten worden op bruikbaarheid en functionaliteit, twee aspecten die elkaar overigens deels overlappen. Daarom wordt een casus geselecteerd uit het assortiment van door het PBL gebruikte modellen. Dit zal meer inzicht verschaffen in de betekenis en kwaliteit van metamodellen in het instrumentarium. De geselecteerde casus bestaat uit een model, waarvan ook een metamodel beschikbaar is. Ook zal hier een connectie gemaakt worden met de kwaliteitsborging van modellen die bij Wageningen UR voor PBL ontworpen en beheerd worden (de zgn. status A).

2 Evenwichtsanalyse

2.1 Definitie evenwicht

Zoals al aangegeven, wordt met het 'evenwicht' van een model of bestand het resultaat bedoeld van een afweging tussen de complexiteit van een model of bestand, de beschikbaarheid van gegevens, en de beoogde toepassing(en). Het idee van evenwicht kan het beste worden geïllustreerd met een eenvoudig voorbeeld. Men kan zich voorstellen dat in het opzetten van een rekenmodel of bestand keuzes gemaakt moeten worden over welke processen en/of details wel of niet moeten worden meegenomen. In veel gevallen zal een uitgebreider model beter in staat zijn om het gedrag van de werkelijkheid te beschrijven. Echter, een uitgebreider model heeft doorgaans een grotere behoefte aan data, om het model aan te sturen, parameterwaarden te kunnen bepalen, en/of om ruimtelijke eenheden te kunnen definiëren. Die data zijn niet altijd voorhanden, of bezit niet de kwaliteit die noodzakelijk is om de genoemde functies te vervullen. Van de kant van de data bezien, zal een eenvoudiger model dus te prefereren zijn om optimaal gebruik te maken van de informatie die uit de data gehaald kan worden, terwijl vanuit de kant van de toepassing gezien een complexer model de voorkeur heeft.

Klassiek wordt de toepassing van een model niet heel expliciet overwogen, en wordt de complexiteit van het model vooral afgezet tegen de data, terwijl de complexiteit van een bestand nauwelijks een punt van aandacht lijkt te zijn. Het meest in het oog springende voorbeeld is hier het probleem van overparametrisatie bij statistische regressiemodellen (Gershenfeld, 1999). We moeten hier even opmerken, dat een model het model met specifieke parameterwaarden kan betekenen, of het model zonder specifieke waarden. Dus bijvoorbeeld, de lijn $y = 2x + 3$ is een model, maar $y = ax + b$ ook. Laten we voor hier even de laatste vorm aanduiden als 'modelstructuur'. Bij statistische regressiemodellen wordt de complexiteit eenvoudig uitgedrukt als het aantal parameters. Een complexere modelstructuur, dat wil zeggen een modelstructuur met meer parameters, heeft een groter vermogen om sets van data te beschrijven. Daarmee neemt de fout af waarmee de data gerepresenteerd worden door een model. Aan de andere kant zijn er al snel meer mogelijke modellen (meerdere sets van parameterwaarden) die dezelfde set van data kunnen beschrijven. Het model is niet langer uniek. De onzekerheid over elk van die mogelijke modellen, en dus de modelstructuur in z'n geheel, neemt daarmee weer toe. Immers, het is onduidelijk welk model nu de correcte weergave is. De complexiteit van een modelstructuur wordt als ideaal beschouwd wanneer de minimale modelcomplexiteit is bereikt die de data nog adequaat kunnen beschrijven.

Er zijn meerdere evaluatiematen ontwikkeld, die bovenbedoeld minimum zoeken (o.a. Akaike, 1974). Deze evaluatiematen zijn echter niet zonder meer toepasbaar op modellen en bestanden, anders dan statistische regressiemodellen. In regressiemodellen zijn alle parameters gelijkwaardig, in die zin dat hun invloed op het model vergelijkbaar is, bijvoorbeeld het verhogen van de orde van een polynoom met één. Dynamische modellen kunnen veranderingen vertonen in de uitvoer als functie van de tijd, en bevatten vaak diverse non-lineaire termen en terugkoppelingen. Parameters kunnen dan zeer verschillende zaken inhouden: Sommige zijn een drempelwaarde, andere een vermenigvuldigingsfactor, weer andere een verdelingsfactor, enzovoorts. Het ligt daarmee voor de hand om een andere definitie van complexiteit voor deze modellen aan te houden. Voor ruimtelijke gegevensbestanden is weer een andere definitie van complexiteit nodig. Deze verschillen zullen in een evaluatielijst voor complexiteit overwogen moeten worden.

Ondanks die verschillen in definities van complexiteit is de vraagstelling niet wezenlijk anders: Welk niveau van modelcomplexiteit wordt ondersteund door de gebruikte data zonder dat er sprake is van overparametrisatie? Dit punt is in de literatuur door verschillende mensen op verschillende wijzen aangepakt. Een helder raamwerk waarin de onderlinge relaties tussen modelcomplexiteit, toepassing(en), en beschikbare data worden geanalyseerd, is besproken door Wagener *et al.* (2001). Naarmate modellen structureel complexer worden, waarbij complexiteit gedefinieerd is als het aantal parameters en variabelen, neemt de 'performance', het vermogen van het model om data te beschrijven, toe. De parameteronzekerheid zal echter ook toenemen wanneer er meer parameters in het model zitten dan er kunnen worden geïdentificeerd op basis van de data. De toepassing bepaalt voor een groot deel wat de minimale performance, en dus modelcomplexiteit, moet zijn. De gebruikte (kalibratie)data bepalen welke parameters identificeerbaar zijn, en dus welke modelcomplexiteit ondersteund wordt door de data. Er is vervolgens sprake van evenwicht als de modelcomplexiteit afdoende is voor de toepassing, maar niet groter dan dat.

Hjalmarsson (2009) volgt een iets ander spoor. Hij bekijkt de 'experimentele kosten' om een model te kalibreren, waarbij de toepassing expliciet in ogenschouw wordt genomen. Deze kosten kunnen enorm naar beneden, wanneer voor een toepassing een model niet voor zijn gehele bereik goed hoeft te presteren. Een voorbeeld: Een model dat waterstanden voorspelt in de tijd moet gekalibreerd worden. Men kan grote kosten maken om data te verzamelen waarmee alle parameters van dit model van 'juiste' waarden worden voorzien. Deze kosten zijn groot, vanwege het bijna altijd aanwezig zijn van het 'waterbedeffect': Als men probeert het model op het ene punt zo goed mogelijk te krijgen, dan gaat dat veelal ten koste van de modelprestatie op een ander punt. Vergelijk dit met het wegduwen van een bobbel in een waterbed. Pas met veel kosten is het model op meerdere punten goed te krijgen. In de vergelijking met het waterbed zijn er nu meerdere mensen op verschillende plaatsen aan het duwen. Echter, voor de toepassing is dit wellicht niet relevant. Stel, het model uit het voorbeeld voor waterstanden wordt toegepast om de waterniveaus bij hoog water te voorspellen, bijvoorbeeld in het kader van mogelijke overstromingen. Het is dan niet zinvol voor de toepassing om moeite te doen om het model de lage waterstanden goed te laten voorspellen. Deze kosten kunnen dan bespaard worden. In dit licht bezien is het watermodel uit het voorbeeld wellicht in evenwicht wanneer het zodanig gekalibreerd is, dat het de hoge waterstanden voorspelt met een onzekerheid die binnen geaccepteerde grenzen ligt, en dus voor de toepassing 'ideaal' is.

Er bestaan nog meer ideeën over, en definities van evenwicht. Bij de beoordeling van de modellen en bestanden in het PBL-instrumentarium moeten we de toepassing van het model of bestand expliciet meenemen. Een model of bestand in deze context moet projecties geven binnen een geaccepteerde onzekerheid. Daarvoor is voldoende begrip van het systeem, en daaraan gekoppeld, een bepaalde complexiteit van het model of bestand nodig, die weer een bepaalde behoefte aan hoeveelheid en kwaliteit van data met zich meeneemt. Een complexer model of bestand kan nog steeds projecties met de gewenste zekerheid geven, maar zal meestal wel een grotere databehoeftte hebben, een langere doorlooptijd hebben, lastiger te kalibreren zijn, lastiger te analyseren zijn, etc., en is daarom minder gewenst. Modelcomplexiteit, gekoppeld aan begrip van het systeem, datakwaliteit en -hoeveelheid, en toepassing zijn in deze context in 'evenwicht', wanneer projecties gemaakt kunnen worden binnen geaccepteerde marges van onzekerheid.

2.2 Ontwikkeling hulpmiddel: Evaluatielijst

Na het geven van bovenstaande definitie van evenwicht is de vraag, hoe men dit evenwicht kan vinden. In dit werkdocument wordt een evaluatielijst voorgesteld, die bedoeld is om bij het ontwikkelen van een model of bestand als een leidraad te dienen teneinde dit evenwicht na te streven. Verder kan deze informatie als evaluatiecriterium dienen om voor bestaande modellen en bestanden te bepalen of deze in evenwicht zijn. Deze lijst volgt de modelleercyclus. Bij de verschillende stappen uit de cyclus behandelen we ideeën over hoe het concept van evenwicht een rol kan spelen bij de keuzes die gemaakt moeten worden in elk van de stappen. De idee hierachter is dat als bij elke stap van de modelleercyclus het evenwicht nagestreefd wordt, het eindproduct waarschijnlijk ook in evenwicht zal zijn. Gezien het grote belang van 'schaal' in vrijwel alle toepassingen en gebruikte datasets wordt dit thema afzonderlijk van de modelleersteps behandeld. In het laatste deel van het hoofdstuk worden vervolgens een aantal voorbeelden gegeven hoe in de modelleerliteratuur vorm is gegeven aan evenwichts-analyse. Hierbij komt de nadruk sterk te liggen op de identificeerbaarheid van parameters.

Het modelleerproces is in een aantal fasen opgedeeld. Bij elke fase moeten keuzen gemaakt worden. Bij elke fase worden vragen gesteld in het licht van het evenwicht tussen modelcomplexiteit, data en toepassing(en). Deze vragen dienen als leidraad bij het maken van de keuzen. In de praktijk zal er overigens vaak sprake van zijn dat niet dit hele traject wordt doorlopen. Vaak zal een al bestaand model worden toegepast. Toch is het ook in die situaties nuttig om te beseffen welke evenwichtsaspecten zijn meegenomen tijdens de eerdere stadia van de modelontwikkeling en -toepassing. In principe zal per toepassing moeten worden bekeken of de onderliggende keuzes die hier zijn gemaakt hebben geleid tot een model of bestand dat in evenwicht is.

2.2.1 Systemanalyse

De eerste fase van het modelleerproces omvat de analyse van het te modelleren systeem. Een model is altijd noodzakelijkerwijs een vereenvoudigde versie van de perceptie van de werkelijkheid. De kernvraag is wel hoe we haar kunnen vereenvoudigen. Bij de keuzes in aannamen voor het modelleren komt direct de toepassing om de hoek kijken. Een algemene analyse is erop gericht te bepalen welke processen dominant zijn, maar de toepassing bepaalt welke onderdelen van de werkelijkheid relevant zijn voor ons probleem. Dit geldt zowel voor de attributen (de materiële systeemcomponenten), alsmede de processen die de dynamica van de attributen bepalen. Per toepassing kan deze analyse dus ook verschillend uitpakken.

We bekijken wat voorbeelden. In principe heeft in de milieuwetenschappen 'alles' met 'alles' te maken, maar we maken met het oog op de toepassing toch een onderscheid tussen het direct en indirect belang van attributen en processen. Attributen en processen, waarin men per definitie is geïnteresseerd, zijn van direct belang. Men zou kunnen zeggen dat deze altijd meegenomen moeten worden in het model, ongeacht de toepassing. Voor een klimaatmodel is dat bijvoorbeeld temperatuur en neerslag, of voor een waterkwaliteitsmodel de concentratie nutriënten in het oppervlaktewater. Van indirect belang zijn die attributen en processen, die een invloed uitoefenen op de attributen en processen van direct belang. Echter, voor de toepassing is het niet per se nodig ze expliciet in het model mee te nemen. Een veelgebruikte pragmatische benadering is om uit te gaan van Ockham's Scheermes, dus het aannemen van de meest eenvoudige verklaring voor bepaalde observaties. Concreet toegepast op de analyse van het systeem impliceert dit dat indien attributen en processen probleemloos vervangen kunnen worden door constanten en/of randvoorwaarden, dit ook gedaan zou moeten worden. In dat geval kunnen deze attributen en processen buiten beschouwing worden gelaten, of als randvoorwaarden worden opgenomen.

Welke attributen en processen nu precies van direct belang en welke van indirect belang zijn, is minder evident. De keuze tussen direct en indirect is nog niet zomaar bepaald. Twee elementen spelen een rol in die keuze. Het eerste element is de beschikbaarheid van data. Zaken die als belangrijk worden aangemerkt, kunnen toch zelden worden meegenomen als er echt geen gegevens over beschikbaar zijn. In de praktijk zullen deze zaken soms ook niet als belangrijk worden herkend, juist omdat gegevens erover ontbreken. Maar soms zal men wel vermoedens hebben dat iets belangrijk is, zonder dat deze vermoedens ondersteund worden door beschikbare gegevens. Het tweede element is het bestaan van sterk niet-lineaire relaties en terugkoppelingen in bijna alle natuurlijke systemen. Wanneer de dynamica van directe attributen via terugkoppelingsprocessen significant wordt beïnvloed door indirecte attributen, dan is het raadzaam om de betrokken indirecte attributen expliciet in het model mee te nemen.

De significantie van genoemde effecten wordt deels bepaald door de toepassing. Dit kan worden geïllustreerd met een voorbeeld. In de werkelijkheid is er, met name in continentale klimaten, een terugkoppeling tussen neerslag en verdamping. Neerslag zal voor een deel worden afgevoerd via rivieren, en voor een deel bijdragen aan bodemvocht, en vanuit dit reservoir verdampen. Bij afwezigheid van advectieve aanvoer van vocht zal dit de voornaamste bron van atmosferisch vocht zijn, en dus van neerslag. Indien men nu geïnteresseerd is in de klimaatdynamiek in zo'n regio, dan zal deze terugkoppeling een direct effect hebben op de neerslag en verdamping. Het is dan dus significant, en zal, net als neerslag en verdamping zelf, in het model meegenomen moeten worden. Is de toepassing echter een operationeel model voor waterbeheer, dan kan men vaak deze terugkoppeling negeren, omdat de neerslag uit meetreeksen wordt verkregen. De neerslag kan dan als randvoorwaarde in het model worden opgenomen. Verdamping wordt in dit geval wel dynamisch berekend.

Vragen voor de checklist: Welke attributen van de werkelijkheid en welke processen zijn direct relevant voor de uiteindelijke toepassing? Welke terugkoppelingen zijn aanwezig? In welk opzicht zijn deze terugkoppelingen significant voor de toepassing? Hoe zijn de grenzen van het systeem gedefinieerd? Hoe wordt met relevante attributen en processen buiten deze grenzen omgegaan (randvoorwaarden, constanten, genegeerd)? Zijn hier gegevens voor beschikbaar?

2.2.2 Schaal

Het concept 'schaal' speelt in alle latere fasen van de modelleercyclus een rol, en is verder dermate belangrijk, dat deze als apart element in de evaluatielijst wordt opgenomen. Indien er uit deze analyse blijkt dat er een schaalprobleem opduikt, dan is er geen sprake van evenwicht. Er kunnen drie verschillende aspecten van het verzamelbegrip 'schaal' worden onderscheiden (Bierkens *et al.*, 2000).

In beginsel varieert de waarde van attributen continu in ruimte en tijd, die beide oneindig zijn. Het begrip schaal duidt op wanneer deze onderliggende continuïteit discreet en/of eindig wordt gemaakt, wat in de praktijk altijd het geval is. Attributen, zowel de gemeten als de gemodelleerde, hebben een beperkt bereik ('extent') in ruimte en/of tijd. Dit 'bereik' is het eerste aspect. Denk hierbij aan de lengte van en meetreeks of de grootte van een gekarteerd gebied. Binnen dit bereik worden de attributen niet als continue functies van ruimte of tijd beschouwd, maar zijn deze opgedeeld in discontinue intervallen. Deze intervallen hebben vaak allemaal dezelfde grootte, bijv. de cellen in een ruimtelijk gegevensbestand hebben allemaal dezelfde grootte, en in meetreeksen in de tijd wordt vaak geprobeerd de metingen op een vast tijdstip te doen, bv. om de 5 minuten. Dit is het tweede aspect, het aannemen dat de waarde van een attribuut de gemiddelde waarde over het interval is. Er zijn echter vernuftiger methoden om de waarde van een attribuut in een interval te bepalen (bijv. 'kriging').

Het derde aspect volgt uit een analyse in hoeverre het bereik volledig wordt afgedekt door alle intervallen. De ratio tussen deze twee wordt de dekking ('coverage') genoemd. Als voorbeeld kan een tijdreeks van grondwaterstanden worden genomen, waarbij gedurende een jaar (dat is het bereik) om het uur waterstanden worden geregistreerd door middel van een druksensor, die zelf gedurende 1 seconde een meting doet ('support'). Hoewel het in de volksmond heet dat deze metingen op uurbasis zijn genomen, is het in werkelijkheid een meetreeks op basis van seconden, met daar tussenin gaten van vrijwel een uur. De bijbehorende dekking is dan ook slechts 1/3600.

Met schaling ('scaling') wordt aangegeven dat voor een dataset het bereik, de support, de dekking, of een combinatie hiervan worden aangepast. Extrapolatie is het vergroten van het bereik, terwijl interpolatie het vergroten van de dekking is. Sampling is het verkleinen van het bereik, terwijl het bereik wordt verkleind door middel van een uitsnede. Het vergroten van de support gebeurt door middel van opschaling, terwijl neerschalen het verkleinen hiervan is.

Het supportprobleem kan kort worden samengevat in de formule

$$f(\text{avg}(x)) \neq \text{avg}(f(x)),$$

waarbij x de data is, $f()$ een nabewerking, en $\text{avg}()$ een procedure voor opschaling, zoals uitmiddeling. De support van een dataset is relevant als aan twee voorwaarden wordt voldaan. Allereerst moet de dataset heterogeen zijn. De attribuutwaarden verschillen dan per interval. Ten tweede wordt de dataset gebruikt in een niet-lineaire analyse (model of andere nabewerking).

Vragen voor de checklist: Wat zijn de extent, de support en de dekking van de gebruikte dataset? En van het model? En van de toepassing? Wat zijn de heterogeniteiten? Welke transformaties worden toegepast? Genereert dit een schaalprobleem? Zo ja, wat is dit schaalprobleem?

2.2.3 Conceptueel model

In de volgende fase wordt het conceptuele model bepaald, dat op relatief informele wijze de relatie tussen de systeemcomponenten vastlegt. Keuzes worden gemaakt voor de dimensies en de schaal van het model, en het type model. Deze keuzes zijn eveneens afhankelijk van het modeldoel, en van de aanwezigheid en eigenschappen van data die gebruikt kunnen of moeten worden in latere fasen. Naast 'schaal', dat hiervoor al besproken is, en dat altijd een belangrijke rol speelt, zijn er twee belangrijke punten in deze fase: Dimensies en modeltype.

Dimensies

Met dimensies wordt bedoeld de ruimtelijke dimensies of het aggregatieniveau. Er kan onderscheid worden gemaakt tussen modellen van 0, 1, 2 en 3 dimensies. Nuldimensionale modellen behandelen een systeem in hun geheel, zonder expliciet onderscheid te maken naar ruimtelijke variabiliteit binnen het model. Eindimensionale modellen volgen meestal een dominante flux, bijvoorbeeld stroomafwaarts langs een rivier, of verticaal door de bodem. Tweedimensionale modellen en ruimtelijke gegevensbestanden delen meestal het oppervlak van de Aarde, Nederland, of iets anders in een aantal elementen in. Doorgaans zijn deze elementen gekoppeld, maar dat is niet noodzakelijk. In dat laatste geval kunnen we beter spreken van een verzameling van nul- of ééndimensionale modellen. Driedimensionale modellen voegen doorgaans de verticale dimensie aan een tweedimensionaal model toe. Daarnaast wordt gebruik gemaakt van quasi-driedimensionale modellen, waarbij een ruimtelijk tweedimensionaal model van het horizontale vlak is gekoppeld aan een ruimtelijk ééndimensionaal model voor de verticale component.

Welke dimensie het best geschikt is, hangt van de toepassing af. Wanneer men primair in de bruto respons van een object of regio geïnteresseerd is, dan kan een nuldimensionaal ('gelumped') model voldoen. Een voorwaarde is wel dat men de eventuele dynamiek en fluxen binnen het systeem ook geaggregeerd kan beschrijven. Wanneer uitvoer is gewenst die ruimtelijk expliciet is gedistribueerd, dan ligt een één-, twee- of driedimensionale benadering meer voor de hand, tenzij men over een post-processing methodiek beschikt om achteraf dimensies toe te voegen (dis-aggregatie). De ruimtelijke interacties bepalen in welk opzicht het noodzakelijk is om binnen één+-dimensionale modellen de dimensies expliciet te betrekken. Dit kan geïllustreerd worden met een hydrologisch voorbeeld. In de onverzadigde zone (kortweg de zone boven het grondwater) is de dominante flux de verticale. De laterale fluxen zijn vaak verwaarloosbaar. Een landsdekkend onverzadigde zonemodel kan dus bestaan uit een losse verzameling ééndimensionale modellen. Een expliciete tweedimensionale benadering is niet noodzakelijk. Voor het bovenste grondwater is dat anders. De dominante flux is daar lateraal. Uitwisselingen met de onverzadigde zone (boven) en dieper grondwater kunnen door middel van randvoorwaarden van parameterwaarden worden voorzien. Een tweedimensionale benadering kan dus adequaat zijn. Voor een compleet grondwatermodel is een driedimensionaal model vereist.

De gegevensvoorziening speelt hier uiteraard ook een rol. Meerdimensionale modellen hebben in principe meerdimensionale data-invoer nodig. Indien deze data niet voorhanden zijn, dan moet er dus een keuze gemaakt worden tussen twee opties. De eerste optie is om meerdimensionale modelinvoer te genereren vanuit lagerdimensionale gegevens, waarbij gebruik wordt gemaakt van een disaggregatietechniek. De tweede optie is een aanpassing van het modelconcept door het aantal dimensies te reduceren.

Vragen voor de checklist: Op welk aggregatieniveau zijn antwoorden gewenst? Wat zijn de dominante fluxen door het systeem? Wat zijn de ruimtelijke interacties? Welke data zijn voorhanden? Hebben de data de juiste dimensies en eenheden? Zo nee, is er een disaggregatiemethode voorhanden?

Modeltype

Er bestaan verschillende modeltypes, die bruikbaar zijn om een model vorm te geven. Deze worden grofweg in twee groepen ingedeeld. Wederom hangt de keuze af van de toepassing en van de databeschikbaarheid.

De eerste groep bestaat uit de statistische regressiemodellen. Deze modellen zijn vooral geschikt om datagebaseerde modellen te bouwen. Hierbij is het belangrijker om de parameters van waarden te voorzien via kalibratie, dan om de onderliggende fysische principes goed te modelleren, en een betrouwbare extrapoleerbaarheid te verkrijgen. Neurale netwerken zijn een goed voorbeeld uit deze groep. De tweede groep bestaat uit de 'proces-response' modellen. Deze modellen zijn opgebouwd volgens de principes van de 'General Systems Theory', en zijn het meest geschikt om fysische kennis over toestandsvariabelen, fluxen, terugkoppelingen en (pseudo)meetbare parameters in het model mee te nemen. In de praktijk hebben bijna alle modellen en ruimtelijke gegevensbestanden iets van beide in zich.

Vragen voor de checklist: Welke conceptualisatie van de werkelijkheid is gewenst, gegeven de toepassing? Welke conceptualisatie wordt in het model gebruikt? Hoe bruikbaar is die conceptualisatie voor de toepassing? Wat zijn de implicaties van de conceptualisatie voor het schaalniveau, in ruimte en tijd? Wat zijn de implicaties voor de datavoorziening?

2.2.4 Formeel model

Een conceptueel model wordt geformaliseerd met een set van wiskundige vergelijkingen die de processen beschrijven, of door regels wanneer het om een kennissysteem ('expert-judgement') gaat. Keuzes die hier moeten worden gemaakt betreffen onder andere de vorm en orde van de vergelijkingen (bv. discrete drempelwaarde of continue machtsrelatie). Hoewel in zekere mate alle vergelijkingen empirisch zijn, wordt ook hier grofweg een tweedeling gemaakt. De ene groep, de fysisch gebaseerde vergelijkingen, bestaan uit wetmatigheden die onafhankelijk gevalideerd zijn, en uit grootheden die met SI-eenheden worden beschreven. Voorbeelden hiervan zijn de vergelijking van Darcy, de zwaartekrachtconstante, chemische reacties, etc. De tweede groep, de empirische modellen, zijn feitelijk regressiemodellen op het niveau van een afzonderlijke vergelijking. Voorbeelden hiervan zijn de diverse deelprocessen in modellen van fotosynthese, of de hydraulische geometrie relaties in fluviale geomorfologie.

Voor de toepassing wordt de keuze voornamelijk bepaald door de betrouwbaarheid en toepasbaarheid van het model. Fysisch gebaseerde vergelijkingen zijn ironisch genoeg minder betrouwbaar in diverse gevallen door problemen met de identificeerbaarheid van parameters. Deze vergelijkingen kunnen andere data nodig hebben dan beschikbaar zijn, bijvoorbeeld om hun begin en/of randvoorwaarden, of hun parameters goed vast te stellen. Empirische modellen kunnen eenvoudiger met minder data worden opgezet. Verder kunnen er problemen zijn met het inherente schaalniveau, dat vaak aan een fysisch gebaseerd model is gekoppeld. Bijvoorbeeld, de Richards' vergelijking, die waterstroming door de onverzadigde zone beschrijft, kan eigenlijk alleen maar worden toegepast met een ruimtelijke resolutie van ca. 5-10 cm. Bodemeigenschappen zouden dus ook op deze resolutie beschikbaar moeten zijn, tenzij er een adequate disaggregatietechniek voorhanden is. Voor homogene substraten is dit minder een probleem dan voor heterogene substraten. Verder kunnen fysisch gebaseerde modellen parameters nodig hebben die überhaupt niet beschikbaar zijn. Dit probleem speelt veel minder bij empirische modellen.

Vragen voor de checklist: Welke types vergelijkingen zijn beschikbaar? Is er een a-priori voorkeur, gegeven de toepassing? Wat is de databehoeft? Wordt er een probleem verwacht met het identificeren van parameters? Welke mate van onzekerheid kan worden getolereerd? Hoe heterogeen is het domein?

2.2.5 Numeriek model

Het numerieke model is de implementatie van het formele model in een computersysteem. Hiervoor kan gebruik worden gemaakt van generieke programmeertalen als C++ of Fortran, een dedicated modellersysteem, of generieke (statistische) software. Keuzes die hier moeten worden gemaakt, slaan bijvoorbeeld op methodologie (eindige differenties, eindige elementen), numerieke convergentiecriteria, enzovoort.

Naar verwachting zouden in deze fase data en toepassing de kleinste rol spelen, aangezien numerieke methoden erop gericht zijn om het formele model op correcte wijze door te rekenen, zonder relatie tot data of toepassing. De voorkeur voor numerieke methoden en de optimale discretisering van ruimte en tijd worden echter voor een deel door de toepassing en beschikbaarheid van data bepaald. Veel numerieke methoden zijn gelimiteerd door aspecten als verlies van stabiliteit onder bepaalde condities. Verder worden randvoorwaarden opgelegd door externe factoren als rekencapaciteit. Een keuze voor discretisering heeft dus een direct effect op de doorlooptijd van de modellen. In de dagelijkse praktijk kan dit er voor zorgen dat gewenste modelleerstappen als uitgebreide kalibratie of onzekerheidsanalyse niet haalbaar

zijn. Aan de andere kant, positiever beschouwd kan het voor een toepassing juist een prima keuze zijn om genoeg te nemen met een minder nauwkeurige rekenmethode, die wel sneller is of niet wordt beperkt door de rekencapaciteit.

Hoewel de toepassing er zelf niets mee te maken heeft, is het voor de toepassing verder wel van belang, dat er een verificatie heeft plaatsgevonden om de omzetting van formeel model naar numeriek model te controleren. Een niet geverifieerd rekenmodel kan namelijk ook niet op een zinnige manier gevalideerd worden. Verder zou het in principe kunnen gebeuren, dat een formeel model zeer geschikt is voor een toepassing, maar dat deze winst verloren gaat door een verkeerde omzetting naar numeriek model.

Vragen voor de checklist: Is het formele model op adequate wijze omgezet in een numeriek model? Heeft er verificatie van de code plaatsgevonden? Kan de benodigde en beschikbare invoerdata worden gebruikt in het operationele modelleringproces? Wat zijn de rektijden van het model? Valt dit binnen de beschikbare operationele capaciteit? Is er afdoende capaciteit voor een adequate kalibratie, validatie, gevoeligheidsanalyse en/of gestructureerde onzekerheidsanalyse?

2.2.6 Schematisering

Schematisering behelst het bepalen van de ruimtelijke eenheden waarmee gedistribueerde modellen moeten werken. Dit is uiteraard alleen van toepassing op één-, twee- en driedimensionale modellen. De schematisering staat los van de ruimtelijke discretisering die bij de vorige fase heeft plaatsgevonden. Dit kan worden geïllustreerd met een eenvoudig voorbeeld. Waterstroming door de onverzadigde zone wordt beschreven met behulp van de Richards' vergelijking, wat een (meestal ééndimensionale) partiële differentiaalvergelijking oplevert. De ruimtelijke discretisering gebeurt in de stapjes die bepaald zijn door de eigenschappen van de numerieke methode. Dit hoeft zeker niet samen te vallen met de ruimtelijke schematisering, dat het aantal bodemlagen bepaalt dat wordt onderscheiden. De schematisering wordt dus primair bepaald door de heterogeniteit van de bodem, de beschikbaarheid van data over de bodem, en de wenselijkheid vanuit de toepassing om deze heterogeniteit te representeren in het model. Alleen in het extreme geval dat de bodemheterogeniteit als continue functie van de ruimte gekend of verondersteld is, kunnen schematisering en discretisering samenvallen.

De schematisering is erg afhankelijk van de data die beschikbaar zijn om ruimtelijke eenheden te definiëren, en van nut of noodzaak om dit te doen. Dit zal per toepassing verschillen. Een paar concrete voorbeelden volgen hieronder.

STONE

Het modelinstrumentarium STONE is een landsdekkend, driedimensionaal emissiemodel voor nutriënten, werkend op een ruimtelijk grid van 250 bij 250 m. Ruimtelijke terugkoppelingen zijn er in de verticale dimensie. Om die reden wordt alleen de verticale dimensie expliciet doorgerekend, en worden de ruimtelijke dimensies gerepresenteerd door een verzameling, in essentie onafhankelijke, rekeneenheden, die 'plots' worden genoemd. Ruimtelijke variabiliteit is er alleen tussen de plots. Met variabiliteit binnen de plot wordt in principe geen rekening gehouden. De plots zijn dusdanig gedefinieerd, dat met een redelijk aantal plots (6405 in geval van STONE) zoveel mogelijk unieke combinaties van geohydrologie, landgebruik en bodemeigenschappen kunnen worden beschreven. Eén van de consequenties van deze benadering is dat de 'virtuele' plots bestaan uit 'reële' gridcellen die niet aaneensluitend zijn. Koopveengronden in verschillende polders kunnen bijvoorbeeld in één en dezelfde plot eindigen. Zolang er wordt voldaan aan de aanname dat er geen significante laterale processen

zijn, is dit geen probleem. Dit is echter wel een beperking voor het toepassingsgebied. Voor een bepaalde toepassing zou dit model dus als 'te weinig complex', en dus als niet in evenwicht zijnde kunnen worden omschreven.

Heterogeniteit bovenste bodemlagen

Op detailniveau zijn de Nederlandse bodems vrij heterogeen. Ook ogenschijnlijk homogene bodems, zoals de zandgronden in Zuid- en Oost-Nederland, blijken te bestaan uit een complexe stratigrafie van meer fijnkorrelige en grofkorrelige dekzandpakketten, afgewisseld met dunne leem- of juist grindlaagjes. Traditioneel worden deze details van de bodemopbouw niet meegenomen in hydrologische modellen. Voor klassieke toepassingen, zoals grondwaterhydrologie, is dit ook niet echt een probleem, aangezien het domein van interesse vooral het dieper gelegen grondwater is, en men daar sowieso geen informatie met deze precisie heeft. Hydraulische eigenschappen zijn voornamelijk 'effectieve' parameters, die worden verkregen door kalibratie aan de hand van pompproeven. Echter, voor een andere toepassing, bijvoorbeeld de waterbeschikbaarheid voor droogtegevoelige vegetatie, kunnen deze details in de bodemopbouw wel relevant zijn. Dit komt, omdat in de onverzadigde zone dunne bandjes met een afwijkende textuur (korrelgrootteverdeling) het vochtgehalte in de bodemhorizonten hier direct boven of onder sterk kunnen beïnvloeden. Voor dit soort toepassingen is dus een meer gedetailleerde bodemschematisering vereist dan wanneer de toepassing in de grondwaterhydrologie ligt.

De gekozen schematisering zal in veel gevallen ook afhankelijk zijn van de beschikbaarheid van data. Het heeft immers weinig zin om ruimtelijke eenheden te formuleren als er geen data voorhanden zijn om elke eenheid van unieke eigenschappen te voorzien. Bepaalde uitzonderingen zijn echter denkbaar. Ten eerste kan men disaggregatietechnieken gebruiken om kunstmatig de resolutie van de onderliggende data te verfijnen. Ten tweede kunnen a priori aannames worden gedaan aangaande de ruimtelijke variabiliteit van systeemeigenschappen of parameters, al dan niet in de vorm van een onderliggend model. Een voorbeeld is de veronderstelling dat in verweringsbodems de hydraulische doorlaatbaarheid exponentieel afneemt met de diepte. Hierbij dient te worden opgemerkt dat beide situaties niet onafhankelijk zijn. Disaggregatietechnieken maken vaak gebruik van aannames voor de ruimtelijke variabiliteit op kleine schaal.

Vragen voor de checklist: Welke schematisering is gewenst vanuit de toepassing? Is deze schematisering haalbaar, gegeven de doorlooptijd van het model? Is het wenselijk en/of mogelijk om eenheden samen te voegen tot 'plots'? Zijn er afdoende data beschikbaar om de schematisering uit te kunnen voeren? Zijn deze data op het juiste schaalniveau voorhanden? Zo nee, is er de beschikking over een schaaltechniek? Kan deze worden toegepast?

2.2.7 Gevoeligheidsanalyse

Bij een gevoeligheidsanalyse wordt onderzocht of er in de waarden van een parameter, een begrenzingsconditie, beginconditie of andere invoer, een interval is, waarbinnen de modeluitvoer sterk verandert (Saltelli, 2000, 2004). Om de gevoeligheid te bepalen, wordt een aantal 'runs' van het model uitgevoerd waarbinnen de parameters systematisch een (groot) deel van hun hele bereik doorlopen. Het doel van een gevoeligheidsanalyse is om inzicht te krijgen in welke parameters nauwkeurig geschat zouden moeten worden en welke niet. Deze fase wordt vaak als voorbereiding op de kalibratie doorlopen.

Voor sommige parameters kunnen kleine verschillen in de waarde tot grote veranderingen in de modeluitvoer leiden, terwijl dit voor andere parameters niet het geval is. Zelfs grote verschillen doen er dan niet veel toe. Het hangt er vaak echter ook weer vanaf in welke

toestand het model zich bevindt. Om die reden wordt er nog onderscheid gemaakt tussen lokale en globale gevoeligheidsanalyse. Een lokale gevoeligheidsanalyse richt zich alleen op de nominale toestand van het model, en bepaalt de 'gevoeligheid' van de verschillende parameters in deze toestand. Bij een globale analyse daarentegen worden de gevoeligheden van de verschillende parameters voor het hele bereik van toestanden beschouwd. Een globale gevoeligheidsanalyse is uiteraard veel arbeidsintensiever. Een lokale gevoeligheidsanalyse kan afdoende zijn met het oog op een specifieke toepassing.

Het nauwkeurig kunnen schatten van parameters hangt deels van de data af, maar dus ook deels van het model zelf. Echter, het is prima mogelijk dat het voor een toepassing niet relevant is om een bepaalde parameter goed te bepalen, ook al is dat wel mogelijk. Anderzijds kan een nauwkeurige schatting van een parameter heel relevant zijn voor een toepassing, maar blijkt bij de gevoeligheidsanalyse dat dit niet goed mogelijk is. In dit laatste geval zou de conclusie kunnen luiden dat het evenwicht tussen model, data en toepassing toch niet goed benaderd wordt.

Gevoeligheidsanalyses worden overigens vaak verward met onzekerheidsanalyses. Bij een onzekerheidsanalyse gaat het erom te achterhalen welke parameters het meest bijdragen aan de variabiliteit van de uitvoer. Ook hier zal in het algemeen een reeks modelruns worden uitgevoerd, maar nu wordt niet systematisch het bereik van de parameters doorlopen. In plaats daarvan wordt er voor elke parameter een schatting gemaakt van de meest waarschijnlijke waarde, en daaromheen wordt een kansverdeling aangenomen. Ook kan er rekening gehouden worden met mogelijke correlaties/afhankelijkheden tussen de parameters. Vervolgens wordt een groot aantal trekkingen gedaan uit de verdelingen voor alle parameters. Voor elke getrokken set wordt het model gerund. Als de trekkingen op een handige manier zijn gedaan (bijvoorbeeld met een 'winding stairs sample'), dan is het mogelijk om de variatie in de modeluitvoer toe te schrijven aan bepaalde (groepen van) parameters (Jansen, 1996).

Vragen voor de checklist: Is er een gevoeligheidsanalyse uitgevoerd? En zo ja, waar is die beschreven? Is er een onzekerheidsanalyse uitgevoerd? En zo ja, waar is die beschreven?

2.2.8 Kalibratie

Een model omvat praktisch altijd meerdere parameters, die nog van waarden voorzien moeten worden, om kwantitatieve voorspellingen te kunnen doen. Het bepalen van de waarden van deze parameters wordt aangeduid als 'kalibratie'. Manieren om de parameters van een model van waarden te voorzien zijn, onder meer, het gebruik van waarden gegeven in de literatuur, expertkennis, het gebruik van waarden verkregen in aparte proeven, en het vergelijken van modelvoorspellingen met gegevens. Diverse auteurs bedoelen met de term 'kalibratie' vooral dat laatste.

Er kunnen zich diverse problemen voordoen bij het gebruik van literatuur en expertkennis bij het bepalen van de parameterwaarden. Het is bijvoorbeeld altijd de vraag in hoeverre de data uit andere bronnen geschikt zijn voor een adequate kalibratie. Schaal is altijd een punt, maar ook de 'commensurabiliteit'. Dit laatste houdt in, dat de informatiebron en de parameter weliswaar dezelfde naam hebben, maar er verder toch sprake is van een andere conceptmatige invulling. Dit zal weinig een probleem zijn bij parameters die evident zijn, zoals de valsnelheid, maar wel een probleem zijn bij relatief slecht gedefinieerde parameters. Schaalproblemen doen zich vooral voor als de schaal waarop parameters worden gemeten niet overeenkomt met de schaal waarop de parameters in het model voorkomen. Deze laatste hangt vaak samen met de gebruikte schematisering of numerieke discretisering.

Bij kalibratie in nauwere zin worden modeluitkomsten vergeleken met gegevens. Deze modeluitkomsten zijn vooral variabelen, zoals temperatuur, waterniveau, of het aantal dieren per vierkante meter, waarvan men ook metingen in de tijd heeft gemaakt. Het model wordt verschillende malen gedraaid, elke keer met een verschillende set parameterwaarden. Na elke keer draaien met een andere set van parameterwaarden worden de uitkomsten van het model vergeleken met dezelfde gegevens. Hierbij wordt een van tevoren gekozen maat (een 'objective function') aangehouden, die bij elke keer draaien evalueert hoe goed de uitkomsten passen bij de data. De set parameterwaarden, die volgens deze maat het beste passen bij de gegevens, worden als de 'correcte' parameterwaarden aangenomen. Dit proces kan handmatig worden uitgevoerd, maar er bestaan diverse technieken waarbij via een iteratief schema parameterwaarden dusdanig worden aangepast, dat de kwaliteit van het model geoptimaliseerd wordt. We noemen hier alleen het klassieke Simplex algoritme (Nelder & Mead, 1965), Markov Chain Monte Carlo (Karssenbergh *et al.*, 2010), en PEST (Doherty, 2003).

Toepassing en data spelen in deze fase een significante rol. Hieronder bespreken we een aantal aspecten. Allereerst de kwaliteit van een model, die men op verschillende wijzen kan kwantificeren. Klassiek worden modelvariabelen vergeleken met een meetreeks met corresponderende variabelen, al kan de wijze waarop dit gebeurt variëren. Hiervoor zijn diverse statistieken beschikbaar, die steeds de nadruk leggen op een ander aspect. Een groot aantal van deze statistieken zijn gegeven in bijlage 1. Het is echter ook afhankelijk van de toepassing welk deel van het dynamisch bereik van het systeem relevant is. Zoals eerder al genoemd, zijn voor een overstromingsmodel hoge waterniveaus relevant, terwijl voor een studie naar de effecten van droogte vooral de lage waterniveaus relevant zijn. In academische zin streeft men vaak naar een 'perfect' model, dat voor zowel hoge als lage waterstanden een goede voorspelling geeft, maar in de praktijk is een dergelijk model moeilijk te verkrijgen. De toepassing bepaalt dus de keuze op welke situatie de nadruk wordt gelegd tijdens de kalibratie.

Een belangrijk aspect is ook het vermogen om parameters uniek te 'identificeren'. Problemen met het uniek bepalen kunnen door de structuur van het model zelf komen, maar ook door de beschikbare data. Het zogenaamde 'equifinalityprobleem' (Beven, 2006) is het probleem, dat verschillende modelstructuren dezelfde uitvoer kunnen genereren. Het is niet triviaal hoe onderscheid gemaakt zou moeten worden tussen de verschillende mogelijke modelstructuren. Verder zijn er problemen te verwachten met de identificeerbaarheid van parameters. In de praktijk zullen meerdere parameters in een model min of meer gekoppeld zijn. Een te lage waarde voor de ene parameter wordt dan gecompenseerd door een te hoge waarde voor een andere parameter. Dit levert dan wel problemen op, wanneer het model voor een andere toepassing wordt gebruikt. Intelligente kalibratieschema's zijn vaak wel in staat om situaties te detecteren, waarbij sprake is van een lage identificeerbaarheid van de betrokken parameters.

Klassiek wordt vaak slechts één modelvariabele tegelijk betrokken bij het kalibratieproces. De identificeerbaarheid van parameters kan sterk worden vergroot door meerdere variabelen tegelijkertijd te betrekken (Bergström *et al.*, 2002; Wagener *et al.*, 2003). Ook 'soft data' (kwalitatief, discontinue of geschat) kunnen een grote meerwaarde hebben (Seibert & McDonnell, 2002). Data, met een grote variantie of met veel 'meetfouten', zullen een duidelijke grens stellen aan het kalibratieproces. De kwaliteit van data kan getoetst worden door onder meer het gebruik van de Fisher-informatie. Een kanttekening bij dit alles is echter, dat als een model voor een toepassing is gekalibreerd, het voor die toepassing zelf eigenlijk niet meer heel relevant is of de parameters nu uniek geïdentificeerd zijn of niet. Wel is het belangrijk om een idee te hebben van hoe de kalibratie tot stand is gekomen, om inzicht te krijgen in de kwaliteit van het model.

Vragen voor de checklist: Hoe zijn de waarden van de parameters van het model bepaald? Via een numeriek algoritme, of direct? Indien direct, wat zijn de gebruikte informatiebronnen? Hoe onzeker zijn deze? Is er sprake van een schaalprobleem? Zo ja, hoe is dit opgelost? Is er sprake van incommensurabiliteit tussen bron en modelparameters? Welke maat of maten van overeenstemming wordt/worden gebruikt tijdens de kalibratie? Welke zouden wenselijk zijn gegeven de modeltoepassing? Welke delen van het dynamisch bereik van het model zijn relevant voor de toepassing? Zijn de modelparameters onafhankelijk van elkaar? Is de informatie-inhoud van de kalibratiedata afdoende om de parameters eenduidig te identificeren? Welke mogelijkheden tot multi-criteria kalibratie zijn aanwezig, gewenst en/of gebruikt? Welke mogelijkheden tot gebruik van soft data zijn aanwezig, gewenst en/of gebruikt? Per gekalibreerde variabele: op welke wijze is de overeenstemming tussen kalibratiedata en output van het kalibratiemodel vastgesteld? Welke overeenstemmingmaten zijn gebruikt, welke outliertests, en wat voor residuenplots zijn gemaakt? Worden er uitspraken gedaan over de betrouwbaarheid van de gekalibreerde parameterwaarden, en hoe is deze informatie verkregen?

2.2.9 Validatie en kruisvalidatie

De validatie van een model houdt in, dat wordt geëvalueerd of het model voldoet voor zijn doel ('fitness for function') en of het een zinnige representatie van de werkelijkheid is. Dit staat los van de verificatie, waarbij gekeken wordt of een conceptueel model goed omgezet is tot numeriek model. Dit conceptuele model kan daarmee nog steeds 'fout' zijn in de zin, dat het niet de werkelijkheid voldoende benadert. Deze beoordeling is uiteraard zeer afhankelijk van de toepassing. Om deze reden wordt er door meerdere auteurs nog onderscheid gemaakt tussen de validatie van een model, en de toepassingsgerichte validatie van een model.

Bij de validatie van een model dienen andere data gebruikt te worden, dan die gebruikt zijn voor de kalibratie van het model. Hoewel voor een goede validatie in principe nieuwe data zouden moeten worden gebruikt, is het in de praktijk vaak om diverse redenen te moeilijk om dit te doen. Een veelvoorkomende praktijk is dat men een tijdsreeks over meerdere jaren heeft ingedeeld in twee (of meerdere) delen. Een deel wordt dan gebruikt voor de kalibratie. Het model wordt dan gedraaid, en men evalueert of de modelvoorspellingen het andere deel van de data goed benaderen. Vanzelfsprekend is het de bedoeling dat er bij een gevalideerd model een zo klein mogelijke afwijking zit tussen modelvoorspellingen en de validatiedata. Dit wordt uitgedrukt als 'goodness of fit'. Voor deze evaluatie worden maten gebruikt, die in principe op maten lijken, die bij de kalibratie worden gebruikt. De meest bekende maat voor validatie is de 'R-squared'.

In veel gevallen zijn er zo weinig data beschikbaar dat het als jammer wordt ervaren om ze niet allemaal bij de kalibratie te mogen gebruiken. In dat geval wordt meestal gebruik gemaakt van wat verwarrenderwijs 'kruisvalidatie' wordt genoemd. Bij kruisvalidatie is meestal sprake van het 'leave one out'. Dit wil zeggen, dat de data in X sets is ingedeeld. Voor de kalibratie worden $X-1$ delen van de data gebruikt. Het resterende deel wordt voor de validatie gebruikt. Daarna wordt een ander deel van de data voor de validatie gereserveerd, en wordt het hele proces herhaald. Dit gebeurt in totaal X keer.

Het is mogelijk dat het kalibratiemodel zogenaamde metaparameters bevat. Dit zijn parameters die niet statistisch worden geschat, maar die door de gebruiker als het ware moeten worden 'ingesteld'. Een voorbeeld is het aantal componenten bij een model voor waterafvoer. Voor metaparameters is altijd een dubbele kruisvalidatie nodig. Er is dan een buitenste 'loop' waarin het aantal compartimenten wordt gevarieerd, en een binnenste 'loop', waarin de eigenlijke parameterschatting gebeurt (Stone, 1974; Varma & Simon, 2006).

Vragen voor de checklist: Is het model gevalideerd? Zo ja, wat is er dan precies gevalideerd? Het hele model? Onderdelen van het model? Welke onderdelen? En welke onderdelen zijn niet gevalideerd? Welke aannames zijn bij de validatie gemaakt? Is bij het schatten van de 'goodness of fit' gebruik gemaakt van kruisvalidatie? Voor welke parameters wel en voor welke parameters niet? Zijn er metaparameters? Zo ja, zijn die bepaald met een dubbele kruisvalidatieloop?

2.3 Gestructureerde evenwichtsanalyse

Klassieke methodes om optimale modelcomplexiteit te bereiken bouwen direct voort op de eerdergenoemde technieken die voor regressiemodellen worden gebruikt, waaronder het gebruik van het Akaike informatiecriterium. Deze voldoen niet voor onze te ontwikkelen evaluatielijst, omdat deze methodes geen of onvoldoende rekening houden met de toepassing van het model. Daarnaast zoeken we een bredere definitie van 'evenwicht', waarin alle modellen en bestanden van het PBL-instrumentarium vallen. De kernvraag is hoe een alternatieve maat voor overparameterisatie kan worden ontwikkeld. Een drietal benaderingen hiervoor, inmiddels veelgebruikt, bespreken we hieronder. Echter, ook deze methoden lijken niet geschikt om zomaar over te nemen, omdat er nog steeds weinig nadruk op de toepassing ligt.

2.3.1 Gevoeligheidsanalyse

De eerste alternatieve methode wordt onder meer besproken door Spear & Hornberger (1980), die voor een waterkwaliteitsmodel de meest kritische parameters identificeren door middel van een globale gevoeligheidsanalyse. Na een Monte Carlo-analyse worden de zgn. 'behavioural runs' gescheiden van de 'non-behavioural runs' door gebruik te maken van een set regels, en van de ratio tussen de posterior en prior parameterverdeling, die een maat voor de gevoeligheid is. De methode wordt verder uitgewerkt in de software DYNIA (Wagener *et al.*, 2001, 2003), waarin een compleet raamwerk voor de analyse van de identificeerbaarheid wordt voorgesteld, die gebaseerd is op het verschil tussen de uniforme prior parameterverdeling, en de posterior verdeling van de beste parameterwaarden. Met dit verschil wordt een maat voor de informatie-inhoud van de data bepaald. Door gebruik te maken van een zgn. glijdend venster kan deze informatie-inhoud variabel in de tijd zijn. Op een vergelijkbare wijze, door gebruik te maken van de gradiënt van de cumulatieve posterior parameterverdeling, wordt de identificeerbaarheid van de modelparameters bepaald. Een optimale modelstructuur is een keuze, een 'trade off', tussen het verminderen van de voorspellingsfout en het vergroten van de identificeerbaarheid van parameters. Echter, de toepassing wordt niet expliciet overwogen in deze methode.

Van der Perk (1997) gebruikt ook een gevoeligheidsanalyse om te komen tot een optimale modelcomplexiteit, al wordt er hiervoor wel een klassiek criterium gebruikt, namelijk de AIC. De optimale complexiteit komt dus overeen met de situatie waarin de AIC minimaal is. Van der Perk selecteert kandidaatmodellen op basis van een variantie. Modellen die een variantie hebben van minder dan minimaal +8% van het minimum van het beste model, worden geselecteerd. Door het klassieke gebruik van de AIC voldoet deze methode niet voor ons doel.

Momenteel de meest gebruikte en meest krachtige methode voor gevoeligheidsanalyse maakt gebruik van Sobol's variantie-gebaseerde globale gevoeligheidsanalyse (Sobol, 1993). In deze methode wordt de variantie in de modeluitvoer opgesplitst in twee delen, namelijk een variantie door individuele parameters, en een variantie door de interactie van meerdere parameters. Aldus kan overparameterisatie (equifinality) worden opgespoord. Van Werkhoven *et al.* (2008) passen deze methode met succes toe op een neerslagafvoermiddel.

2.3.2 Foutenanalyse

Een tweede benadering is om de modelcomplexiteit stapsgewijs te vergroten. In elke stap wordt het model opnieuw gekalibreerd en gevalideerd. De validatiefout wordt gescheiden in een datafout (zowel in metingen als andere modelinvoer) en een modelfout. De optimale modelcomplexiteit is bereikt bij minimale modelfout. De Wit & Pebesma (2001) passen dit principe toe op een nutriëntenmodel voor de Rijn en Elbe. Schoups & Hopmans (2006) onderzoeken de optimale modelcomplexiteit voor ééndimensionale modellen met verticaal water- en zouttransport. Hun aanpak is gebaseerd op de decompositie van totale modelleerfouten als gevolg van onzekerheid in modelstructuur, en fouten als gevolg van onzekerheid in modelinvoer.

2.3.3 Stapsgewijze verfijning

Een derde methode is ook gebaseerd op een stapsgewijze toename van de modelcomplexiteit. In tegenstelling tot de vorige methode wordt de modelcomplexiteit systematisch vergroot als reactie op gevonden deficiënties in de modelvoorspellingen, totdat de benodigde nauwkeurigheid is bereikt (Klemes, 1983). Atkinson *et al.* (2002) bouwen voort op deze aanpak, en richten zich op een afweging tussen modelcomplexiteit, nauwkeurigheid en voorspellingonzekerheid in een afvoermodel voor neerslag. Deze zaken worden bepaald door gebruik te maken van gevoeligheidsanalyse en Monte Carlo-foutenanalyse. Deze analyses identificeren de meest dominante fysieke controlemechanismen voor de variabiliteit van, in dit geval, de rivierstroming. Er worden schattingen gemaakt van de nauwkeurigheid en onzekerheid in de voorspellingen, en daarmee dus van de mate van benodigde en beschikbare modelcomplexiteit. De modelstructuur wordt in een aantal stappen verfijnd (en dus complexer gemaakt), terwijl na elke verfijning opnieuw de voorspellingskracht van het model wordt getoetst. Er worden meerdere criteria hiervoor gebruikt, waaronder correlatie (timing), runoff ratio (voor de omvang), en 'signature plots' (sensu Farmer *et al.*, 2002). Het model wordt onder verschillende condities getoetst, waaronder verschillen in tijdschaal en klimaat (droogte). Voor de toepassing vonden Atkinson *et al.* dat de modelcomplexiteit toeneemt met korte tijdschalen en toenemende droogte.

2.4 Evenwicht en metamodellen

Metamodellen zijn modellen, die van een complex referentiemodel afgeleid en vereenvoudigd zijn. Metamodellen zijn in het kader van evenwicht op meerdere manieren interessant. Ten eerste is er de kwestie in hoeverre de boven beschreven evenwichtsanalysemethoden ook toepasbaar zijn op metamodellen. Ten tweede de vraag welke rol metamodellen kunnen spelen om de evenwichtsanalysemethoden beter te kunnen uitvoeren. Bijvoorbeeld, Schoups & Hopmans (2006) voeren hun evenwichtsanalyse uit met een serie, in complexiteit toenemende metamodellen die steeds worden vergeleken met het bijbehorende fysisch-gebaseerde referentiemodel. De totale modelfout kan worden gekwantificeerd onder de aanname dat het referentiemodel de 'werkelijkheid' is. Ten derde is een praktisch nadeel van veel methoden voor het doen van onzekerheid, identificatie- en evenwichtsanalyses dat ze onder andere gebruik maken van Monte Carlo-benaderingen. Voor modellen met een lange doorlooptijd is dit niet altijd haalbaar. Metamodellen, gegeven dat ze een 'exacte' kopie zijn van het originele model, kunnen hier uitkomst bieden. Dat werpt ook meteen de interessante vraag op of dat metamodellen idealiter niet gewoon de evenwichtige variant van het volledige model zijn.

3 Evenwichtsanalyse en status A

3.1 Uitgangspunten

Zoals al besproken is het de bedoeling om tot een evaluatielijst te komen, waarmee vastgesteld kan worden of modellen en data in evenwicht zijn. In het vorige hoofdstuk zijn vragen voor deze lijst opgesteld per fase in de modelontwikkeling, en daarnaast nog vragen voor het aspect 'schaal'. Sommige vragen zijn al meteen redelijk concreet, zoals: "Hoe vindt de parameterisering van het model plaats?" Deze vraag is duidelijk te beantwoorden. Andere vragen zijn minder concreet. Bijvoorbeeld een vraag als "welke overeenstemmingmaten zouden wenselijk zijn voor de toepassing?" is niet zonder wetenschappelijke achtergrondkennis te beantwoorden.

In dit hoofdstuk bekijken we de al bestaande evaluatielijst voor de zogenoemde 'status A' voor modellen, die gebruikt wordt voor de kwaliteitscontrole bij de WOT Natuur & Milieu. Er is ook een status A-evaluatielijst voor bestanden, die een deel van dezelfde vragen omvat. De status A-lijst bevat ook een aantal meer abstracte onderwerpen, maar is inmiddels behoorlijk getoetst in de praktijk. In die zin kan het als inspiratie bieden voor hoe de evaluatielijst modelcomplexiteit eruit zou kunnen komen te zien. Bovendien zijn een aantal vragen van de status A-lijst van directe relevantie voor de evaluatie van de modelcomplexiteit. En uiteindelijk zal de evaluatielijst modelcomplexiteit een rol binnen dezelfde kwaliteitscontrole toebedeeld krijgen. Redenen genoeg dus om de lijst voor status A in meer detail te bekijken.

In dit hoofdstuk bespreken we de evaluatielijst van status A per onderdeel. Voor elk onderdeel geven we een kort overzicht van de vragen die vanuit status A gesteld worden, en wat de bedoeling van het onderdeel is. Daarnaast geven we aan wat de relevantie van het onderdeel voor de evaluatie van evenwicht zou kunnen zijn. Dit verschilt nogal per onderdeel, maar zeker een aantal onderdelen sluit probleemloos aan bij gewenste zaken voor evaluatie modelcomplexiteit.

3.2 Vergelijking met status A

Status A is een kwaliteitsniveau dat vooral gericht is op de documentatie en het beheer van de modellen en bestanden. Daarmee is er dus ook een garantie dat er van de modellen en bestanden die goedgekeurd zijn volgens status A, voor ons doel bruikbare documentatie aanwezig is. De status A-checklist is ingedeeld in acht onderdelen, namelijk theorie, technische documentatie, gebruikersdocumentatie, testen van de software, kalibratie, validatie, gevoeligheidsanalyse en beheer. Zoals al gemeld zijn niet alle punten even relevant voor de beoordeling van het evenwicht, maar desalniettemin gaan we alles langs. Daarbij geven we ook de in status A gestelde vragen.

Theorie

Binnen status A wordt gevraagd naar de onderbouwing van het conceptuele model, het expliciet maken van de aannamen en vereenvoudigingen, het mathematische model en het beschrijven van het doel en het beoogde toepassingsgebied van het model. Kort gezegd is dit de beschrijving van de stap van het model in het hoofd naar vergelijkingen e.d. op papier. Deze vragen zijn allemaal zeer relevant voor de bepaling van het evenwicht. De conceptuele onderbouwing is noodzakelijk om te begrijpen hoe het model in elkaar steekt, een essentieel

punt voor de evaluatie van evenwicht. De expliciete aannames en vereenvoudigingen moeten kunnen worden beoordeeld in termen van evenwicht. Het is bijvoorbeeld goed mogelijk dat er veel aannames zijn gemaakt omdat er weinig data beschikbaar waren, of omdat bepaalde aspecten op dat moment niet relevant werden geacht. Bij de beschrijving van het doel wordt aangegeven waarom het model gemaakt is, en voor het toepassingsgebied moet toegelicht worden wat de restricties op het gebruik van het model zijn.

Vragen in status A zijn: Is de theoretische onderbouwing van het model omschreven? Is het doel, waarvoor het model is ontworpen, beschreven? Is het toepassingsgebied van het model beschreven? Zijn de vereenvoudigingen en aannames over de gebruikte representatie van de werkelijkheid gemotiveerd en beschreven?

Technische en gebruikersdocumentatie

De volgende twee onderdelen, respectievelijk de technische en gebruikersdocumentatie, zijn zelf voor de beoordeling van het evenwicht niet zo van belang. Echter, om een model als casus te kunnen gebruiken, dient men de beschikking te hebben over dergelijke documentatie.

Vragen in status A zijn: Is er een document met meta-informatie van het model? Is er een globale beschrijving van de werking van het computerprogramma? Zijn alle modelparameters beschreven? Is alle invoer beschreven? Is alle uitvoer beschreven? Is het toepassingsgebied van het model beschreven en zijn er voorbeelden van uitgevoerde modelstudies gegeven? Is het benodigde kennisniveau van de gebruiker van het model beschreven? Zijn de beperkingen van het computerprogramma beschreven? Is het user interface beschreven? Is de invoer beschreven? Is de uitvoer beschreven? Is er een korte samenvatting van de validaties, de verificaties, het testen, de gevoeligheidsanalyses en de onzekerheidsanalyses van het computerprogramma?

Testen

Voor het verkrijgen van status A moet een model of bestand afdoende zijn getest om enigszins te kunnen verifiëren of de code een goede vertaling van het concept is. Een model moet geverifieerd worden om het 'geldigheidswaarde' te geven. Immers, een numeriek model hoeft niet een goede vertaling van een conceptueel model te zijn.

Vragen bij status A: Is er een set testgegevens waarmee de vertaling van de modelvergelijkingen naar de programmacode is geverifieerd? Zijn de meest basale tests op het computerprogramma uitgevoerd? Is het rekenhart geheel getest? Zijn de testgegevens reproduceerbaar opgeslagen? Zijn de uitgevoerde tests beschreven?

Kalibratie

Voor status A moet het model voor een toepassing gekalibreerd zijn, en deze kalibratie moet reproduceerbaar zijn gedocumenteerd. De informatie uit dit onderdeel is zeker noodzakelijk, maar hoogstwaarschijnlijk niet afdoende om evenwicht te kunnen bepalen. Het is niet waarschijnlijk dat bijvoorbeeld de informatie-inhoud van de gegevens beschreven is, of dat er afwegingen staan vermeld in de documentatie over welke overeenstemmingmaat is gebruikt. In de evaluatielijst zal daarom het onderdeel 'kalibratie' uitgebreid moeten worden. Deze vraag is overigens zelden van toepassing op ruimtelijke gegevensbestanden.

De vragen bij status A zijn: Is het model voor een toepassing gekalibreerd? Is de kalibratie beschreven?

Validatie

Om status A te verkrijgen, moet het bestand of model voor één of meerdere toepassingen gevalideerd zijn. Deze validatie moet reproduceerbaar zijn gedocumenteerd, en liefst publiekelijk toegankelijk. Tevens moet er een kritische analyse zijn die volgt uit deze validatie. Men beschrijft de mogelijke tekortkomingen van het model, en welke stappen nog mogelijk zijn om met deze tekortkomingen om te gaan. In de praktijk blijkt het voor veel modellen en bestanden echter moeilijk haalbaar om tot een goede validatie te komen.

Dit onderdeel is zeer relevant voor een evaluatielijst voor evenwicht. Een validatie vindt altijd plaats in het kader van een toepassing. Deze toepassing moet hetzelfde zijn als de beoogde toepassing in de evaluatie van het evenwicht. Indien er een validatie is gedaan die niet de toepassing dekt die beoogd is, dan is een evaluatie van het evenwicht niet mogelijk in deze fase. Verder is het, net als bij de kalibratie, niet onwaarschijnlijk dat er vele details in de documentatie ontbreken die wel relevant zijn voor de evaluatie van evenwicht. Het is aannemelijk dat dit onderdeel in de evaluatielijst voor evenwicht uitgebreid moet worden ten opzichte van de evaluatielijst voor status A.

Status A-vragen zijn: Zijn de uitgevoerde validaties beschreven? Is in deze beschrijving opgenomen wat nog niet is gevalideerd? Is er een kritische analyse van mogelijke tekortkomingen?

Gevoeligheidsanalyse

Om status A te verkrijgen, moet een gevoeligheidsanalyse gemaakt zijn van het model, en die analyse moet beschreven zijn. Voor bestanden is dit punt vaak niet van toepassing, ook al bestaan er mogelijkheden voor het doen van gevoeligheidsanalyses. Deze gevoeligheidsanalyse hoeft voor het verkrijgen van status A niet uitputtend te zijn, maar dient wel tenminste voor een aantal onderdelen, zoals parameterwaarden, begrenzingcondities, begincondities, en/of andere invoer, een indruk te geven van hun invloed op de uitvoer.

Dit onderdeel is zeer relevant voor de evaluatie van het evenwicht. In het vorige hoofdstuk zijn een aantal methoden besproken, die gebruik maken van gevoeligheidsanalyses om tot een 'juiste' mate van modelcomplexiteit te komen. Omdat voor status A de eisen niet zeer hoog zijn, zal voor dit onderdeel vaak een aanvullende gevoeligheidsanalyse moeten worden uitgevoerd om toepasbare resultaten te hebben voor de evaluatie van evenwicht. Ook hangt het van de beoogde toepassing af, of dat de resultaten van een gevoeligheidsanalyse alle informatie levert voor de evaluatie van het evenwicht of niet.

Vragen bij status A: Zijn voor het toepassingsgebied van het model gevoeligheidsanalyses uitgevoerd? Zijn deze gevoeligheidsanalyses beschreven?

Beheer

Beheer is een onderwerp dat geen directe relevantie heeft voor het onderwerp van dit rapport, maar desalniettemin is het zinvol om hier zicht op te hebben. Beheersplannen zullen erop gericht zijn om de kwaliteit van een model of bestand te verhogen. Als bijvoorbeeld uit de analyse van een validatie volgt dat een model bepaalde aspecten onder- of overbelicht, of als uit een studie blijkt dat er te weinig steun van data is voor bepaalde onderdelen, dan zal een beheersplan erop gericht zijn deze punten aan te pakken. Dat specifieke aspect houdt wel verband met evenwicht.

Vragen bij status A: Is er een beheersplan? Is het inhoudelijk beheer geregeld? Is het technisch beheer geregeld? Is de ondersteuning naar de gebruikers geregeld? Zijn de uitgevoerde verbeteringen gerapporteerd? Zijn de geplande verbeteringen voor het model beschreven?

4 Samenstelling van de evaluatielijst

4.1 Uitgangspunten

In dit hoofdstuk bespreken we het prototype van de evaluatielijst voor evenwicht. Dit prototype omvat zowel vragen, die gebaseerd zijn op vragen van de evaluatielijst voor status A, als de vragen die zijn opgesteld in hoofdstuk 2. In zekere zin stellen we hier dus twee aparte deellijsten op. De evaluatielijst voor evenwicht moet in de toekomst een coherent geheel worden, maar voor nu is het doel om alle aspecten die van belang zijn voor de evaluatie van evenwicht mee te nemen. De evaluatielijst moet de balans kunnen beoordelen tussen modelcomplexiteit, het toepassingsgebied van het model, en de databeschikbaarheid en -kwaliteit. Bij het opstellen van deze lijst is het relevant, om zowel te kijken naar de invoer, met andere woorden wat is redelijkerwijs vraagbaar en bruikbaar, als naar de uitvoer, ofwel kijken naar wat voor type(n) uitkomsten we willen hebben van deze lijst. Het is verder relevant, maar in eerste instantie nog niet haalbaar, om te kijken wie de evaluatielijst uiteindelijk gaan gebruiken, en verder hoe die lijst wordt afgestemd op het specifieke doel dat de gebruiker ermee voor ogen heeft. Ook laten we nog even achterwege hoe de bevindingen van het gebruik van de lijst zouden moeten worden gecommuniceerd, en naar wie.

Dit hoofdstuk is in twee secties ingedeeld. In het eerste deel (par. 4.2) behandelen we de vragen die gebaseerd op of geïnspireerd zijn door vragen van de evaluatielijst voor status A, waarvan redelijk zeker is dat deze zinvol en bruikbaar zijn. De verwachting is dat deze vragen bij de praktijktoetsing zonder veel problemen kunnen worden ingevuld, en daarnaast relevante informatie opleveren voor de bepaling van het evenwicht (dus een vorm van 'no regret' vragen). Daarnaast is het opnemen van deze vragen uit de evaluatielijst voor status A ook ingegeven door het feit dat de evaluatielijst voor evenwicht los moet staan van de evaluatielijst voor status A. In het tweede deel (par. 4.3) bespreken we vragen, die zeer relevant worden geacht voor de evaluatie van evenwicht, maar waarvan nu minder duidelijk is, hoe ze geïmplementeerd kunnen worden in de evaluatielijst voor evenwicht.

In het projectvervolg zullen de twee lijsten samengevoegd worden tot één geheel. De vragen zullen bij latere toetsingen tegen casussen ook soms nog toegespitst moeten worden op gebruikers. Dat houdt in, dat alle punten eenduidig geformuleerd moeten zijn, en dat het de gemiddelde gebruiker snel duidelijk moet zijn wat van hem wordt verwacht. Deze aspecten worden voor nu naar de achtergrond geschoven.

4.2 Prototype evaluatielijst deel 1

Voor het opstellen van een prototype van de evaluatielijst van evenwicht is de voorgestelde aanpak om eerst naar de meta-informatie van het model te vragen, zodat bv. helder is over welk model en eventueel metamodel het gaat. De vragen bij dit onderdeel zijn bijna gelijk aan de vragen A5 en A16 uit de evaluatielijst voor status A. Daarna worden de drie onderwerpen die een rol spelen in het evenwicht, namelijk modelcomplexiteit, data, en toepassing, eerst afzonderlijk geëvalueerd. Als laatste wordt gekeken naar de verhouding tussen deze drie onderwerpen in het onderdeel 'synthese'. Dit deel van de lijst is dus vooral praktijkgericht.

4.2.1 Algemene vragen ter inleiding

In deze sectie worden algemene vragen gesteld over de status van het model. Bij modellen en bestanden, die aan status A voldoen, zouden deze vragen direct beantwoordbaar moeten zijn.

Is er een document met meta-informatie van het model, en met een kort overzicht van testen, kalibratie, validatie, en gevoeligheidsanalyse?

Onder meta-informatie verstaan we naam (acronym), versienummer en uitgiftedatum van het model, een zeer korte beschrijving van de werking, doel, het toepassingsgebied, schaalniveau (temporeel en ruimtelijk), benodigde in- en uitvoer, programmeertaal en user interface, platform. Indien van toepassing, geef ook de meta-informatie van het metamodel, en geef daarbij heel kort de essentiële verschillen tussen model en metamodel (bv. de versie 5.6 van het model 5.6 is Fortran, het metamodel versie 1.0 in C; het metamodel clustert alle variabelen X op basis van die en die relatie). Verder moet er een kort overzicht zijn, waaruit duidelijk wordt, dat het model afdoende getest, geverifieerd, gekalibreerd, en gevalideerd is, en er een gevoeligheidsanalyse is gedaan.

4.2.2 Model

Bij dit onderdeel wordt gevraagd naar het conceptuele model, naar de omzetting van concept naar vergelijkingen, en van daaruit naar numeriek rekenmodel. Deze omzettingen moeten ook geverifieerd zijn.

Wat is het doel van het model?

Beschrijf waarom het model is ontworpen. Een model kan ingewikkelder of minder ingewikkeld zijn, dan nodig is voor het doel. Het doel van het model dient ook overeen te komen met de toepassing(en) van het model.

Is er een overzichtelijke beschrijving van het conceptuele model?

Met het conceptuele model bedoelen we hier het grofmazige model, zonder het precieze raamwerk te beschrijven. Er hoeven geen formules gegeven te worden, maar wel de belangrijkste concepten, en hoe die gekoppeld zijn. Let hierbij op aggregatie, schalen in ruimte en tijd, terugkoppelingen. Grafische representaties zijn veelal zeer behulpzaam. Voor de evaluatie van evenwicht moet duidelijk zijn hoe het model eruit ziet.

Is er een overzicht van de gemaakte aannamen en vereenvoudigingen?

Voor alle bestanden en modellen worden aannamen en vereenvoudigingen gemaakt. Beschrijf deze, en motiveer waarom ze gemaakt zijn. De aannamen en vereenvoudigingen moeten duidelijk te achterhalen zijn. Verscheidene experts beschouwen dit punt als de kern van modelleren. Het maken van veel aannames zal het toepassingsgebied van het model snel beïnvloeden.

Is er een beschrijving van het rekenmodel?

Het raamwerk waarin het conceptuele model is beschreven bestaat meestal uit vergelijkingen als lineaire vergelijkingen of differentiaalvergelijkingen. Geef de vergelijkingen, en voorzie de documentatie van tabellen met de gebruikte variabelen, parameters en constanten met een korte betekenisomschrijving, dimensies en eenheden. Voor de evaluatie van evenwicht is het nodig om eerst te controleren of het concept goed is omgezet naar vergelijkingen. Een vraag zou hier bijvoorbeeld kunnen zijn of het ook eenvoudiger zou kunnen.

Is er een beschrijving van het rekenhart/gecodeerde model?

De rekenmodellen zijn in de praktijk bijna altijd te complex om algebraïsch op te kunnen lossen, en worden daarom geïmplementeerd als computerprogramma's (numerieke rekenmodellen). De berekeningen vinden plaats in het rekenhart van zulke programma's. Er moet gekeken worden, of een model niet te ingewikkeld of te simpel omgezet is van model naar code, en of adequate algoritmen zijn toegepast. Ook dient deze vraag als opzet voor de verificatie.

Is er een verificatie, en een beschrijving van de testresultaten van het rekenhart?

Het rekenhart dient getest te zijn, en er moet een verificatie zijn van de omzetting van model naar code. Ook moet gekeken worden naar de geschiktheid van de gebruikte hardware. Deze tests kunnen ook relevant zijn voor de metamodellering van het conceptuele model. Als er veel goede testresultaten zijn, dan geeft dit vertrouwen in de omzetting van mathematisch model naar computerprogramma.

Is er een beschrijving van de werking van het programma?

Apart van het rekenhart bestaat een programma ook uit andere onderdelen, waaronder de onderdelen die in- en uitvoer behandelen. Dit punt is van direct belang voor de evaluatie van evenwicht, bijvoorbeeld doordat vaak (des)aggregatiemethoden worden gebruikt. Verder is het relevant omdat voor de evenwichtsanalyse het programma begrijpelijk moet zijn.

Is er een beschrijving van het benodigde kennisniveau van de gebruiker?

Geef aan, wat de niet-triviale benodigde kennis is voor een gebruiker om met het model om te gaan, zowel vakinhoudelijk als computertechisch. Een programma dat bijvoorbeeld veel kennis vereist maar weinig 'doet', zou als niet gebalanceerd kunnen worden beschouwd. Voor de evaluatie van het evenwicht is het verder van belang, dat het op de achtergrondkennis van de gebruiker aansluit met het oog op de toepassing van het model.

Is er een evaluatie van de prestaties en beperkingen van het programma?

Het complete programma, rekenhart, behandeling van in- en uitvoer, etc. heeft een bepaald vermogen in termen van tijdsduur van een run, benodigde capaciteit, etc. Het programma moet afdoende getest zijn, zowel op uitkomsten als op functionaliteiten. Een programma kan bugs bevatten die resultaten onbetrouwbaar maken. Deze bugs zouden echter ook kunnen zitten in onderdelen die overbodig zijn. Een programma kan te complex zijn voor het beoogde doel. Een programma dat veel tijd kost, kan wellicht gereduceerd worden tot een efficiënter programma of tot metamodel.

4.2.3 Data, invoer en uitvoer

Deze punten zijn gericht op data en invoer voor het model en de geleverde uitvoer van het model. Data kunnen voor allerlei zaken gebruikt zijn, bijvoorbeeld voor het opstellen van het model, en hoeft dus niet per se invoer te zijn voor het rekenprogramma. Aan de andere kant is niet alle invoer data, bijvoorbeeld invoer om het programma aan te sturen.

Is er een beschrijving van de oorsprong van data en invoer?

Geef een kort overzicht van de gebruikte bronnen van parameterwaarden, gebruikte data, expert judgment, initiële condities, enzovoorts. Mogelijke bronnen zijn wetenschappelijke artikelen, eigen experimenten, ander model in modellentrein, enzovoorts. Dit overzicht kan het beste gegeven worden in tabelvorm. De kwaliteit van de data kan deels beoordeeld worden op basis van de oorsprong. Het is wellicht mogelijk om invoer of data te vervangen om de

betrouwbaarheid in z'n geheel te verhogen. In de evaluatie van het evenwicht is het verder van belang om te kijken of bepaalde invoer of data overbodig zijn, en of essentiële invoer of data nog ontbreken voor de toepassing.

Waar zijn data en invoer voor gebruikt?

Beschrijf waar data en invoer in het model voor worden gebruikt. Dit kan zijn gebruik voor het geven van waarden aan parameters, de algehele kalibratie van het model, of voor validatie. Misschien is de invoer overbodig, of juist niet afdoende. Als data voor validatie gebruikt worden, dan moet deze validatie wel goed passen bij de toepassing.

Is er een beschrijving van de eventuele bewerkingen op data en invoer?

Voor de verschillende toepassingen voor data en invoer (kalibratie, validatie, etc.) kunnen er manipulaties zijn uitgevoerd. Denk hierbij bijvoorbeeld aan het fitten op data om waarden voor parameters te verkrijgen, het omzetten van getallen naar percentages, logaritmische bewerkingen, schattingen, betrouwbaarheidsintervallen, enzovoorts. Het is mogelijk dat er fouten en onbetrouwbaarheden zitten in deze bewerkingen. Sommige bewerkingen zouden niet relevant kunnen zijn voor de toepassing.

Is er een overzicht van welke testen op kwaliteit van data en invoer zijn gedaan?

Beschrijf hoe de invoer en data zijn getest op betrouwbaarheid, onzekerheden, informatie-inhoud, etc. Als de invoer afkomstig is van andere modellen of bestanden, verwijst dan naar de relevante documentatie van die bronnen. Data kunnen een slechte informatie-inhoud hebben, een grote bron van onbetrouwbaarheid blijken, etc. De data of invoer kunnen goed gebruikt zijn, maar niet voldoende getest. Voor sommige toepassingen is een hogere kwaliteit van bepaalde data vereist dan voor andere toepassingen. De kwaliteit van data heeft invloed op het vermogen om parameters te identificeren.

Is duidelijk voor welke onderdelen nog data nodig zijn?

Beschrijf of er nog onderdelen zijn waar data of invoer voor nodig zijn. Bijvoorbeeld, het programma bestaat uit drie modules, waarvan er één nog niet gekalibreerd is omdat de benodigde gegevens ontbreken. De onzekerheid over missende onderdelen kan groot zijn, of juist klein. Wellicht blijkt bijvoorbeeld uit gevoeligheidsanalyses dat deze onderdelen eigenlijk niet gekalibreerd hoeven te worden.

Is er een beschrijving van de uitvoer?

Beschrijf duidelijk de door het programma geproduceerde uitvoer. Het is behulpzaam als dit tabellen met korte omschrijvingen, variabelen, eenheden, etc. zijn. Het moet voor de toepassing duidelijk zijn wat het programma aan uitvoer produceert. Een numeriek rekenmodel of bestand kan getest en geverifieerd zijn, maar dat hoeft niet te betekenen dat de uitvoer bruikbaar is voor een toepassing.

Is duidelijk wat de onzekerheid op de uitvoer is?

Beschrijf welke testen er zijn gedaan voor de uitvoer. Is duidelijk wat de betrouwbaarheid is van de uitvoer? Het moet vast te stellen zijn of de uitvoer bruikbaar is voor andere modellen of bestanden, en wat de onzekerheid voor de toepassing is, zoals voorspellingen of beleidsadviezen.

4.2.4 Toepassing

Bij dit onderdeel wordt er gevraagd naar de toepassingen van het model, en voor welke toepassingen het model gekalibreerd en/of gevalideerd is.

Is er een beschrijving van het toepassingsgebied van het model of bestand?

Beschrijf voor welke toepassingen het model of bestand gebruikt wordt, of zou kunnen worden gebruikt. Geef ook aan voor welke verwante toepassingen het niet bruikbaar is. Vermeld ook expliciet wanneer een bestand of uitvoer als invoer voor andere bestanden of modellen wordt gebruikt, zoals in modellentreinen. Het toepassingsgebied van een model of bestand moet bij het doel passen. Veel bestanden en modeluitkomsten worden gebruikt als invoer voor andere modellen in een modellentrein. Een model kan te eenvoudig of complex zijn voor het toepassingsgebied. Bij meerdere toepassingen is het de vraag of het model even geschikt is voor elk van die toepassingen. Het is mogelijk dat de evaluatie van het evenwicht goed uitvalt voor de ene toepassing, maar niet goed voor de andere.

Is er een beschrijving van de kalibratie van het model?

Beschrijf voor welke toepassingen het model gekalibreerd is, bijvoorbeeld een hydrologisch model is gekalibreerd voor toepassingen voor de Rijn, maar niet voor de Waal. Het moet duidelijk zijn hoe de kalibratie is gedaan (bv. hydrologische modellen worden vaak met PEST gekalibreerd). De methode voor kalibratie van een model voor de ene toepassing is niet per se een bruikbare methode voor een andere toepassing. Als een model eenvoudig te kalibreren is voor andere toepassingen, dan kan dit betekenen dat het model algemeen geldig is en goed opgebouwd is. Evenzo kan de noodzaak voor een volledig nieuwe kalibratie bij elke toepassing erop duiden dat het model mechanistisch slecht onderbouwd is. Dit is ook een interessant punt in het kader van evaluatie van evenwicht.

Is er een beschrijving van de validatie van het model?

Beschrijf voor welke toepassingen het model gevalideerd is. Let erop dat voor een validatie het beste andere data(sets) gebruikt kunnen worden dan voor een kalibratie. Een programma kan geverifieerd zijn, maar dat houdt niet automatisch in dat het programma ook een goede of bruikbare afspiegeling is van de 'werkelijkheid' met het oog op de toepassing. Verder is het mogelijk dat de programma-uitvoer niet relevant of bruikbaar is voor de toepassing. Anderzijds is het mogelijk dat bepaalde ongevalideerde uitvoer niet van belang is voor de toepassing, bijvoorbeeld het is niet nodig als invoer voor een ander model. Deze laatste observatie is dan ook een punt dat van belang is in de evaluatie van evenwicht.

4.2.5 Synthese

Bij dit onderdeel wordt er gekeken naar de interacties tussen modelvorm, data en invoer, en toepassingsgebied van het model.

Is er een overzicht en een beschrijving van de gedane gevoeligheidsanalyses?

Beschrijf de gevoeligheidsanalyses die voor onderdelen of toepassingen van het model of bestand zijn uitgevoerd. Bij een gevoeligheidsanalyse worden parameterwaarden, drempelwaarden, etc. gevarieerd en wordt de gevoeligheid van de modeluitkomst voor deze factoren bepaald. Gevoeligheidsanalyses zijn bruikbaar om de relatieve rol van parameters, etc. te evalueren, en worden veel gebruikt om de modelcomplexiteit te analyseren. De resultaten zijn ook van belang voor de identificatie van parameters. Voor parameters die niet gevoelig blijken te zijn voor een bepaalde toepassing, hoeft ook niet veel moeite te worden gedaan om tot juiste identificatie te komen (m.a.w. ze van een betrouwbare waarde te voorzien).

Is duidelijk welke parameters en variabelen (niet/slecht) identificeerbaar zijn?

Beschrijf welke parameters en variabelen in het model uit data of invoer af te leiden zijn, en welke niet. Het is mogelijk dat de data verantwoordelijk zijn voor het gebrek aan identificeerbaarheid. In dat geval moet er andere of aanvullende data gevonden worden. Het kan ook aan de modelstructuur zelf liggen, of aan de toepassing. Voor de toepassing is het niet per se problematisch dat bepaalde parameters niet identificeerbaar zijn. Dergelijke observaties zijn echter wel interessant voor de evaluatie van evenwicht. Parameters, die niet identificeerbaar zijn maar er ook niet toe lijken te doen, dragen wellicht onnodig bij aan de complexiteit van het model.

Is er een overzicht en beschrijving van de gedane onzekerheidsanalyses?

Beschrijf de onzekerheidsanalyses die gedaan zijn op onderdelen van het model, dan wel het gehele model. Bij een onzekerheidsanalyse wordt nagegaan welke factoren (parameters, drempelwaarden, typen invoer, etc.) de meeste onzekerheid in de modeluitvoer genereren. De onzekerheid kan door onderdelen veroorzaakt worden die geen belangrijke rol spelen. Omgekeerd kunnen deze onderdelen juist heel belangrijk zijn in het model, waardoor het model zijn waarde verliest. De reductie van overbodige onzekerheid in de voorspellingen door modellen en bestanden is het ultieme doel achter de evaluatie van evenwicht, en in die zin is dit onderdeel het meest cruciaal van alle.

4.3 Prototype evaluatielijst deel 2

Dit tweede deel van de lijst is dus in de opzet zoals gewenst, maar is nog onvoldoende operationeel voor gebruik door derden. In dit deel kijken we naar de aspecten, die in hoofdstuk 2 naar voren zijn gekomen, waarvan in dit stadium niet geheel duidelijk is hoe deze in een operationele evaluatielijst voor evenwicht het best tot hun recht kunnen komen. Omdat deze aspecten wel raken aan de kern van wat 'evenwicht' inhoudt, worden ze hier nog eens samengevat. In het projectvervolg zal de operationalisering van deze aspecten aan bod komen. Nog ten overvloede melden we dat het uiteindelijk de bedoeling is dat de vragen uit beide losse lijsten worden samengevoegd en omgezet tot vragen, die zowel praktisch zijn als de modelcyclus volgen. Het doel hiervan is om gedurende de gehele modelontwikkeling het evenwicht te monitoren, en om eventueel bij te sturen als in een fase teveel van het evenwicht wordt afgeweken. Gezien de grote hoeveelheid vragen, is het van belang dat de gebruiker per onderdeel aangeeft welke specifieke aspecten het meeste aandacht verdienen voor de beoogde toepassing, en waarom.

Systeemanalyse

Welke attributen van de werkelijkheid en welke processen zijn direct relevant voor de uiteindelijke toepassing? Welke terugkoppelingen zijn aanwezig? In welk opzicht zijn deze terugkoppelingen significant voor de toepassing? Hoe zijn de systeemgrenzen gedefinieerd? Hoe wordt met relevante attributen en processen buiten het modelsysteem omgegaan? Zijn hier gegevens voor beschikbaar? Een aantal van deze punten zouden beantwoord moeten zijn met de vragen over de aannamen, vereenvoudigingen, en de modelbeschrijving, zoals gesteld in het eerste deel van de lijst.

Schaal

Wat zijn de extent, de supportdekking van de gebruikte dataset? Van het model? Van de toepassing? Wat zijn de heterogeniteiten? Welke transformaties worden toegepast? Genereert dit een schaalprobleem? Indien er uit deze analyse blijkt dat er een schaalprobleem opduikt, dan is er geen sprake van evenwicht. Het 'aspect' schaal speelt een rol in elke fase van de modelleercyclus. Deze vragen moeten dus constant herhaald worden bij elk nieuw onderdeel.

Conceptueel model

Op welk aggregatieniveau zijn antwoorden gewenst? Wat zijn de dominante fluxen door het systeem (niet relevant voor modellen die niet ruimtelijk zijn)? Wat zijn de ruimtelijke interacties? Welke data zijn voorhanden? Hebben de data de juiste dimensies? Zo nee, is er een methode voor disaggregatie voorhanden? Welke conceptualisatie van de werkelijkheid is gewenst, gegeven de toepassing? Welke conceptualisatie wordt in het model gebruikt? Hoe bruikbaar is die conceptualisatie voor de toepassing? Wat zijn de implicaties van de conceptualisatie voor het schaalniveau, in ruimte en tijd? Wat zijn de implicaties voor de datavoorziening?

Formeel model

Welke types vergelijkingen zijn beschikbaar? Is er een a-priori voorkeur gegeven de toepassing? Wat is de databehoeftte? Wordt er een parameteridentificatieprobleem verwacht? Welke mate van onzekerheid kan worden getolereerd? Hoe heterogeen is het domein?

Numeriek model

Is het formele model op adequate wijze omgezet in een numeriek model? Heeft er verificatie van de code plaatsgevonden? Kunnen de benodigde en beschikbare invoerdata worden gebruikt in het operationele modeleringsproces? Wat zijn de rektijden van het model? Valt dit binnen de beschikbare operationele capaciteit? Is er afdoende capaciteit voor een adequate kalibratie? Validatie? Gevoeligheidsanalyse? Onzekerheidsanalyse? Gestructureerde onzekerheidsanalyse?

Schematisering

Welke schematisering is gewenst vanuit de toepassing? Is deze schematisering haalbaar, gegeven de doorlooptijd van het model (relevante vraag voor metamodellen)? Is het wenselijk en/of mogelijk om eenheden samen te voegen tot 'plots'? Is er afdoende data beschikbaar om de schematisering uit te kunnen voeren? Zijn deze data op het juiste schaalniveau voorhanden? Zo nee, heeft men de beschikking over een schalingstechniek? Kan deze worden toegepast?

Gevoeligheidsanalyse

Is er een gevoeligheidsanalyse uitgevoerd? En zo ja, waar is die beschreven? Is er een onzekerheidsanalyse uitgevoerd? En zo ja, waar is die beschreven? Deze vragen zouden goeddeels al beantwoord moeten zijn met de vragen in het eerste deel van de evaluatielijst over gevoeligheids- en onzekerheidsanalyses.

Kalibratie

Hoe worden de parameters van waarden voorzien? Via een kalibratie, of direct? Indien direct, wat zijn de gebruikte informatiebronnen? Dit punt zou duidelijk moeten worden uit de vragen, gesteld in het eerste deel van de lijst. Hoe onzeker zijn de gebruikte informatiebronnen? Is er sprake van een schaalprobleem? Zo ja, hoe is dit opgelost? Is er verder nog sprake van incommensurabiliteit tussen bron en modelparameter? Welke criteria voor 'goodness of fit' worden gebruikt tijdens de kalibratie? Welke criteria zouden wenselijk kunnen zijn, gegeven de modeltoepassing? Welke delen van het dynamisch bereik van het model zijn relevant voor de toepassing? Zijn de modelparameters onafhankelijk van elkaar? Is de informatie-inhoud van de kalibratiedata afdoende om de parameters eenduidig te identificeren? Welke mogelijkheden tot multi-objective kalibratie zijn aanwezig, gewenst, gebruikt? Welke mogelijkheden tot gebruik van soft data zijn aanwezig, gewenst, gebruikt? Op een aantal van deze vragen is geen eenduidig antwoord te geven. In het projectvervolg wordt hier nog dieper op ingegaan.

Validatie

Is het model gevalideerd? Zo ja, wat is er dan precies gevalideerd? Het hele model? Onderdelen van het model? Welke onderdelen? En welke onderdelen zijn niet gevalideerd? Welke aannames zijn bij de validatie gemaakt? Is bij het schatten van de 'goodness of fit' gebruik gemaakt van kruisvalidatie? Voor welke parameters wel en voor welke parameters niet? Zijn er metaparameters? Zo ja, zijn die bepaald door middel van een dubbele kruisvalidatieloop?

5 Discussie en conclusies

In dit hoofdstuk richten we ons op een aantal zaken. Eerst bespreken we in enig detail een model dat zeer bruikbaar wordt geacht als casus om het prototype van de evaluatielijst tegen te toetsen. Daarna bespreken we nog enkele discussiepunten die relevant zijn voor verder onderzoek. Ten slotte geven we enkele voorlopige conclusies en aanbevelingen voor het onderwerp 'modelevenwicht'.

5.1 Selectie casus voor toetsing prototype

Het is zeer zinvol om het hier ontwikkelde prototype van de evaluatielijst te toetsen door het toe te passen in een aantal geselecteerde casussen. Dit dient meerdere doelen. Allereerst, het toetst het vermogen van de evaluatielijst om daadwerkelijk het evenwicht tussen modelcomplexiteit, data en toepassing te vinden. Daarnaast is het te verwachten dat het stellen van de vragen om het evenwicht te evalueren ook leidt tot het beter stellen van de vragen. De vragen, die nu nog niet operationeel zijn, zullen verbeteren. Verder zullen de getoetste modellen en bestanden zelf waarschijnlijk profiteren van dit experiment. Er kunnen punten naar boven komen die voorheen over- of onderbelicht zijn gebleven. De verwachting is daarmee ook dat de onzekerheden in de voorspellingen door de modellen en bestanden in de casussen zullen verminderen, en dat het vertrouwen in de voorspellingen zal toenemen.

Alvorens dit te doen, is het noodzakelijk om eerst een aantal representatieve en bruikbare casussen te selecteren binnen en buiten het PBL-modellenbestand. Eén beoogde casus is in elk geval het hydrologisch model SWAP. Er zijn meer redenen voor de keuze van SWAP. Allereerst heeft SWAP een centrale rol in modellentreinen. Het model krijgt (een deel van) zijn uitvoer van andere modellen, en levert op zijn beurt ook weer invoer voor andere modellen. Kennis over de onzekerheid en kwaliteit van dit model is hiermee van grote waarde. Het model SWAP heeft al vroeg aan de normen voor status A voldaan, wat inhoudt dat de documentatie aanwezig is om snel inzicht in het model te kunnen krijgen. Verder wordt er op dit moment een metamodelversie van het model ontwikkeld (MetaSWAP). Daarmee kan ook meteen gekeken worden naar de rol van het metamodel voor de evenwichtsevaluatie.

Het numerieke rekenmodel SWAP (Kroes & Van Dam, 2003) is een eendimensionaal niet-lineair deterministisch en dynamisch model. Voor de beschrijving van het transport van water in de bodem wordt gebruik gemaakt van de Richards' vergelijking (Richards, 1931). Verder bevat het model een versimpeld gewasgroeimodel, omdat vegetatie een grote invloed kan uitoefenen op de waterbalans van de bodem. Er zijn meer randvoorwaarden beschreven, waaronder neerslag en verdamping op dagbasis, een opgelegde grondwaterstand of een opgelegde kwel/infiltratieflux.

Hoewel SWAP oorspronkelijk is ontwikkeld voor de lokale schaal (Pawn-districten, WSV-plots), is het geografisch toepassingsgebied uitgebreid tot heel Nederland. Er is verder geen duidelijke begrenzing in de tijd. De tijdsresolutie is een dag of korter. SWAP is gebruikt in de Milieuverkenning 5 en Natuurbalans 2000 voor onderzoek in landbouw, ecosystemen en natuur, in de thema's vermessing, verzuring, verdroging en verspreiding. Het model is gekoppeld aan andere modellen. Het wordt met name gebruikt voor de berekening van grondwaterstanden, kwel/infiltratiefluxen, etc. voor het Landelijk GrondwaterModel (LGM), en het levert invoer voor een bestrijdingsmiddelenmodel (PEARL) en een nutriëntenmodel (STONE).

De validatie van SWAP heeft vooral plaatsgevonden op lokale schaal, en minder op regionale schaal. Verder is de uitvoer vergeleken met de uitvoer van soortgelijke modellen. De onzekerheid in de modeluitspraken is sterk afhankelijk van het beschouwde probleem, een punt dat zeker van belang is in de context van modelevenwicht. Het model is goedgekeurd voor status A, het al besproken kwaliteitssysteem bij de WOT Natuur & Milieu. Er is bruikbare documentatie van onder meer de code, de kalibratie en validatie, betrouwbaarheids-analyse, wetenschappelijke review, het proces van archivering van toepassingen, en evaluatie van die toepassingen.

Het model MetaSWAP is de beoogde metamodelformulering van SWAP, en is sinds 2010 beschikbaar. Het bestaat uit een snel onverzadigd zonemodel, dat gebruikt wordt samen met een regionaal grondwatermodel. Het is eveneens gebaseerd op de Richards' vergelijking, maar heeft als doel om de benodigde rekencapaciteit te reduceren zonder gebruik te maken van te gesimplificeerde methoden gebaseerd op 'lumping' (Van Walsum & Groenendijk, 2008). Het metamodel is verdeeld in drie fasen. Bij de eerste fase wordt gebruik gemaakt van relaties tussen bodemfysische eigenschappen voor aggregatie in wortelzone en de bodem hieronder, afgeleid uit computereperimenten met SWAP. In de laatste fase worden de uitvoervariabelen weer uiteengevlochten tot fysische variabelen op gedetailleerde schaal. Hiermee kunnen de kernberekeningen in de middelste fase gereduceerd worden.

Voor de validatie zijn de modeluitkomsten weer vergeleken met de uitkomsten van SWAP. Op dit moment presteert het snelle onverzadigde zonemodel redelijk in ca. 75% van de testgevallen. Op dit moment is de validatie echter beperkt. De verwachting is dat het percentage toeneemt na koppeling met een regionaal grondwatermodel, en verbetering van de modelleringmethode voor bepaalde grondsoorten.

5.2 Aanpak toetsen prototype evaluatielijst

Het in dit werkdocument ontwikkelde prototype van de evaluatielijst zal moeten worden getoetst in het vervolgonderzoek. Hiervoor stellen we een aantal methoden voor. Allereerst zal het prototype worden overlegd aan enkele deskundigen die werkzaam zijn op het gebied van modelleren. Hiermee kan vrij snel een verwachting van het functioneren van het prototype worden bepaald. Daar staat tegenover dat de toetsing theoretisch blijft. Verder stellen we voor om de concepten achter de evaluatielijst in een peer-reviewed artikel in een internationaal tijdschrift te publiceren, om het geheel zo in een internationale achtergrond te plaatsen.

Het grootste deel van de toetsing zal echter praktijkervaring moeten zijn. De evaluatielijst zal worden getoetst aan een aantal nog te selecteren kleine casussen, en aan de geselecteerde casus SWAP/MetaSWAP. Uit deze toetsing zal blijken of het prototype van de evaluatielijst redelijk presteert in termen van bruikbaarheid, duidelijkheid, en het bepalen van het evenwicht. Deze eerste toetsing zal vermoedelijk punten opleveren die zullen leiden tot aanpassing van de evaluatielijst.

De aangepaste evaluatielijst moet vervolgens opnieuw getoetst worden, ditmaal aan andere modellen om een idee te krijgen van de toepasbaarheid van de lijst. Het advies is daarbij om minimaal twee andere casussen te gebruiken, die beide significant verschillen van het hydrologische model SWAP. Bovendien moeten deze modellen aan status A voldoen om er zeker van te zijn dat er afdoende documentatie beschikbaar is. Deze casussen zijn nog niet geselecteerd, maar te denken valt aan een ruimtelijk gegevensbestand en een niet-hydrologisch model.

5.3 Discussiepunten voor verder onderzoek

Hoewel het prototype van de evaluatielijst in deze verkennende studie nog niet is toegepast, kunnen we al wel een verwachting opstellen op basis van de ervaring die er is met de modellen die voor PBL getoetst worden met status A voor kwaliteitsslag. De praktijk met status A leert dat veel van de getoetste modellen en bestanden snel gelimiteerd zijn in zowel kwantiteit als kwaliteit van data. Dit heeft vaak zijn oorsprong in de praktische onmogelijkheid van sommige data om vergaard te worden, bijvoorbeeld om redenen van privacy of om financiële redenen. Daarnaast is er echter ook een grote druk om modellen alsmaar uit te breiden, zonder elke stap opnieuw te testen en te valideren. Verder zijn veel modellen en bestanden onderdeel van een modellentrein waarin de uitvoer van het ene model of bestand (deel van) de invoer van het volgende model is. Bij veel invoer is het niet geheel duidelijk hoe groot de betrouwbaarheid is.

Een tweede les uit de praktijk met status A is dat sommige modellen zijn gebruikt voor toepassingen waar ze van origine niet voor bedoeld waren. Een reden hiervoor is bijvoorbeeld dat een model van groot belang is voor onderzoek naar de implementatie van bepaalde wetgeving. De validatie van zulke modellen is nog wel eens lastig vanwege het genoemde gebrek aan data, en dit maakt het op zijn beurt moeilijk om een inschatting te maken van de capaciteit van de modellen voor extrapolatie. Het is dus goed mogelijk dat na verandering van de wetgeving de toepassing, en dus ook de mogelijke geldigheid van het model, is veranderd. In een enkel geval is om dergelijke redenen het originele model verlaten en een nieuw model geformuleerd. Zulke gevallen zijn in het kader van evenwicht ook interessant. Aan de ene zijde zal het model weer meer aan het (nieuwe) toepassingsgebied zijn aangepast, maar aan de andere zijde houdt dat in dat waarschijnlijk alles opnieuw geverifieerd en gevalideerd moet worden.

Het vermoeden is nu, deels gebaseerd op de twee bovenstaande observaties, dat veel modellen en bestanden die nu voor het onderzoek door bijvoorbeeld het PBL worden ingezet, op dit moment waarschijnlijk eerder te gecompliceerd dan te eenvoudig zijn ten opzichte van de beschikbare data en het toepassingsgebied. Deze verwachting geldt ook voor de metamodellen, want bij metamodellen lijkt veel gebruik te worden gemaakt van uit data afgeleide relaties tussen modelvariabelen. Metamodellen moeten daarom nog meer zijn ingebed in data.

Een beoogde toepassing voor de evaluatielijst evenwicht op de wat langere termijn is in de inhoudelijke kwaliteitscontrole van modellen en bestanden voor PBL. Op dit moment zijn een groot aantal modellen geëvalueerd in het kader van status A, die vooral op documentatie en beheer gericht is. Verder loopt er ook een project voor de ontwikkeling van twee inhoudelijke kwaliteitsniveaus, aangeduid met status A+ en status AA. De evaluatielijst evenwicht zou onderdeel kunnen worden van kwaliteitstatus AA. De verwachting is namelijk dat een evenwichtig model ook een inhoudelijk 'goed' model is, met daarbij de kanttekening dat het perfecte model uiteraard niet bestaat.

In dit stadium is het nog te vroeg om methodes aan te reiken aan onderzoekers, die geschikt zijn om de inhoudelijke kwaliteit van de verschillende deelgebieden te bepalen en/of te vergroten (bv. de kwaliteit van de data), dan wel om 'evenwicht' te bepalen en/of vergroten. Daarom is er voor de evaluatielijst meer de opzet van een vragenlijst als status A beoogd. De verwachting is echter dat meer (literatuur)onderzoek ook een meer inhoudelijke opzet van de evaluatielijst mogelijk zal maken. De evaluatielijst zal dan uitgebreid kunnen worden van vragenlijst naar een lijst, die ook met adviezen komt die afhankelijk zijn van het antwoord dat op de vragen is gegeven. Bijvoorbeeld, er wordt dan niet alleen gevraagd of de kwaliteit van

de data is gecontroleerd, maar ook of dat volgens bepaalde geadviseerde methoden is gebeurd.

Een verder discussiepunt betreft de inbedding van metamodellen in dit onderzoek. Hoewel metamodellen hun bestaansrecht primair danken aan de wens om rekentijd en -capaciteit te reduceren, zouden ze ook conceptueel een belangrijke rol kunnen krijgen. Metamodellen zijn 'gedwongen' bedoeld als kleinere versies van de moedermodellen. De bovenstaande uitgesproken verwachting is dat de meeste modellen nu te complex zijn voor hun toepassing. Een interessant punt binnen dit onderzoek zou de vraag kunnen zijn, of men door de modelleur te dwingen het model te versimpelen tot een metamodel, automatisch het evenwicht, en wellicht dus ook de kwaliteit, van het model verbetert. Deze vraag zou echter niet de primaire onderzoeksvraag moeten zijn voor het vervolgonderzoek.

5.4 Conclusie

Er is onderzoek gedaan naar het concept van 'evenwicht' van een model, tussen de complexiteit van het model, kwaliteit en beschikbaarheid van gebruikte data, en het toepassingsgebied van het model. Belangrijke begrippen in dit concept zijn besproken, en diverse, nog redelijk abstracte, vragen zijn opgeworpen die een rol kunnen spelen bij de bepaling van evenwicht. Er is een prototype van een evaluatielijst voor evenwicht ontwikkeld. Deze lijst bestaat uit een aantal punten, ingedeeld in twee onderdelen. Het eerste onderdeel hiervan is sterk op het al bestaande status A geënt, en bestaat uit vragen die vermoedelijk zowel goed begrijpelijk en beantwoordbaar zijn, als ook significant inzicht inleveren in het evenwicht van het model. Het tweede onderdeel bestaat uit vragen die nog niet uitgekristalliseerd zijn, maar die waarschijnlijk wel van groot belang zijn voor het bepalen van evenwicht. De evaluatielijst moet nu getoetst worden. Er zijn een aantal methoden hiervoor voorgesteld, evenals een modelcasus die kan dienen voor de praktijktoetsing.

Literatuur

- Aitken, A. P. (1973). Assessing systematic errors in rainfall-runoff models. *Journal of Hydrology* 20:121-136.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* 19:716-723.
- Atkinson, S. E., R. A. Wood & M. Sivapalan (2002). Climate and landscape controls on water balance model complexity over changing timescales. *Water Resources Research* 38:1314, doi: 10.1029/2002WR001487.
- Bergström, S., G. Lindström & A. Pettersson (2002). Multi-variable parameter estimation to increase confidence in hydrological modeling. *Hydrological processes* 16:413-421.
- Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology* 320:18-36.
- Bierkens, M. F. P., P.A. Finke & P. de Willigen (2000). *Upscaling and Downscaling Methods for Environmental Research*. Kluwer Academic Publishers, Dordrecht.
- Clarke, R. T. (2008). A critique of present procedures used to compare performance of rainfall-runoff models. *Journal of Hydrology* 352:379-387.
- Doherty, J. (2003). Ground water model calibration using pilot points and regularization. *Ground water* 41(2): 170-177.
- Fila, A., G. Bellocchi, M. Acutis & M. Donatelli (2003). Irene: a software to evaluate model performance. *European Journal of Agronomy* 18:369-372.
- Gershenfeld, N. (1999). *The Nature of Mathematical Modelling*. Cambridge University Press, Cambridge.
- Greenwood, D. J., J. J. Neeteson & A. Draycott (1985). Response of potatoes to N fertilizer: Dynamic model. *Plant Soil* 85:185-203.
- Grubbs, F. (1969). Procedures for detecting outlying observations in samples. *Technometrics* 11:1-21.
- Hjalmarsson, H. (2009). System identification of complex and structured systems. *European journal of control* 15:275-310.
- Jansen, M. (1996). *Winding stairs sample analysis program windings 2.0*. Technical Report GLW note MJA-1996-2, Biometris, Wageningen.
- Karssenbergh, D., O. Schmitz, P. Salamon, K. de Jong & M. F. P. Bierkens (2010). A software framework for construction of process-based stochastic spatio-temporal models and data assimilation. *Environmental Modelling & Software* 25(4), 489-502.
- Kobayashi, K. & M. U. Salam (2000). Comparing simulated and measured values using mean squared deviation and its components. *Agron. J.* 92:345-352.
- Kroes, J. G. & J. C. van Dam (2003). *Swap 3.0.3 reference manual*. Technical Report 773, Alterra, Wageningen University and Research Center, Wageningen.
- Kumar, A. (2000). *Dispersion and risk modeling*. Technical report, University of Toledo.
- Legates, D. R. & G. J. McCabe Jr. (1999). Evaluating the use of 'goodness-of-fit' measures in hydrologic and hydroclimatic model validation. *Water resources research* 35:233-242.
- Loague, K. & R. E. Green (1991). Statistical and graphical methods for evaluating solute transport models: overview and application. *Journal of contaminant hydrology* 7:51-73.
- Montgomery, D. C., E. A. Peck, and G. G. Vining (2001). *Introduction to Linear Regression Analysis* (third edition). Wiley, New York.

- Nash, J. E. & J. V. Sutcliffe (1970). River flow forecasting through conceptual models. part i: A discussion of principles. *Journal of Hydrology* 10:282-290.
- Nelder, J. A. & R. Mead (1965). A simplex method for function minimization. *The computer journal* 7(4):308-313.
- Van der Perk, M. (1997). Effect of model structure on the accuracy and uncertainty of results from water quality models. *Hydrological processes* 11:227-239.
- Richards, L. A. (1931). Capillary conduction of liquids through porous mediums. *Physics* 1(5):318-333.
- Saltelli, A., K. Chan & E.M. Scott (2000). *Sensitivity Analysis*. Wiley Series in Probability and Statistics.
- Saltelli, A., S. Tarantola, F. Campolongo & M. Ratto (2004). *Sensitivity Analysis in Practice. A Guide to Assessing Scientific Models*. John Wiley & Sons.
- Schaei, B. & H. V. Gupta (2007). Do nash values have value? *Hydrological Processes* 21:2075-2080.
- Schoups, G. & J. W. Hopmans (2006). Evaluation of model complexity and input uncertainty of field-scale water flow and salt transport. *Vadose Zone Journal* 5:951-962.
- Seibert, J. & J. J. McDonnell (2002). On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. *Water Resources Research* 38(11):1241. doi:10.1029/2001WR000978.
- Sobol', I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Math. Model. Comput. Exp.* 1:407-417.
- Spear, R. C. & G. M. Hornberger (1980). Eutrophication in peel inlet II. Identification of critical uncertainties via generalized sensitivity analysis. *Water Research* 14(1):43-49. doi: 10.1016/0043-1354(80)90040-8.
- Stone, M. (1974). Cross-validators choice and assessment of statistical predictions. *Journal of the Royal Statistical Association, Series B* 36:111-147.
- Varma, S. & R. Simon (2006). Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics* 7:91. doi:10.1186/1471-2105-7-91.
- Van Walsum, P. E. V. & P. Groenendijk (2008). Quasi steady-state simulation of the unsaturated zone in groundwater modeling of lowland regions. *Vadose Zone Journal* 7(2):769-781.
- Wagener, T., D. P. Boyle, M. J. Lees, H. S. Wheater, H. V. Gupta & S. Sorooshian (2001). A framework for the development and application of hydrological models. *Hydrology and Earth System Sciences* 5:13-26.
- Wagener, T., N. McIntyre, M. J. Lees, H. S. Wheater & H.V. Gupta (2003). Towards reduced uncertainty in conceptual rainfall-runoff modeling: Dynamic identifiability analysis. *Hydrological Processes* 17:455-476.
- Van Werkhoven, K., T. Wagener, P. Reed & Y. Tang (2008). Characterization of watershed model behavior across a hydroclimatic gradient. *Water Resources Research* 44:W01429. doi: 10.1029/2007WR006271.
- Willmott, C. J. (1981). On the validation of models. *Phys. Geog.* 2:184-194.
- Willmott, C. J. (1982). Some comments on the evaluation of model performance. *Bull. Amer. Meteorol. Soc.* 63:1309-1313.
- De Wit, M. J. M. & E. J. Pebesma (2001). Nutrient fluxes at the river basin scale. II: the balance between data availability and model complexity. *Hydrological processes* 15:761-775.

Bijlage 1 Overeenstemmingsmaten

Er zijn verschillende manieren om te kijken hoe goed een gekalibreerde parameter bij de data past. We noemen residuenplots, uitbijtertests en overeenstemmingsmaten. In deze bijlage wordt de volgende notatie gebruikt:

- i : index over observaties $1 \dots N$
- y_i : geobserveerde waarde (data)
- f_i : door model voorspelde waarde
- $E(x)$: verwachtingswaarde van de variabele x
- gemiddelde van de variabele x ,
$$\bar{x} = \sum x / N$$

Wanneer hieronder bij een begrip geen referentie staat, dan houdt dat in, dat het om een standaardbegrip gaat. Deze worden in elk statistiekboek behandeld wordt, onder meer in bijvoorbeeld Montgomery *et al.* (2001).

Bias De 'bias' is het systematisch verschil tussen de modeluitvoer en de observaties. De meest eenvoudige schatter voor de bias is gegeven als

$$bias = \bar{y} - \bar{f}$$

Andere schatters worden gegeven door onder meer Willmott (1981, 1982). Men kan verder voor hydrologische modellen ook Aitken (1973) gebruiken.

Variantie De variantie $E(y - E(y))^2$ kan worden geschat door

$$Var = \sum_{i=1}^N (y_i - \bar{y})^2 / (N - 1)$$

Dit is de spreiding van de modeluitvoer rond zijn eigen gemiddelde. Dit is dus de toevallige afwijking.

MSE De MSE, de 'Mean squared error', is de som van variantie en de gekwadrateerde bias, en is gegeven als

$$MSE = \sum_{i=1}^N (y_i - f_i)^2 / (N - 1)$$

CD De CD, de 'coefficient of determination', is vergelijkbaar met het percentage verklaarde variantie R^2 uit een regressie-analyse. De CD is gegeven als

$$CD = \sum_{i=1}^N (\bar{y} - f_i)^2 / \sum_{i=1}^N (y_i - \bar{y})^2$$

CV De CV, de 'coefficient of variation', kan, losjes gezegd, gezien worden als de standaardafwijking, gedeeld door het gemiddelde. De CV is gegeven als

$$CV = 100 / \bar{f} \sqrt{\sum_{i=1}^N (y_i - f_i)^2 / (N - 1)}$$

MSV De MSV, de 'mean squared variation' (Kobayashi and Salam, 2000), is een index, waarbij een hoge waarde aangeeft dat de modeluitvoer minder variatie vertoont dan de data. De MSV is gegeven als

$$MSV = \sum_{i=1}^N [(y_i - \bar{y}) - (f_i - \bar{f})]^2 / (N - 1)$$

NMSE De NMSE, de 'normalized mean squared error' (Kumar, 2000), is een maat die wordt aangehaald in de documentatie bij het softwarepakket IRENE (Fila *et al.*, 2003). De NMSE is gegeven als

$$NMSE = \sum_{i=1}^N (y_i - f_i)^2 / \sum_{i=1}^N y_i f_i$$

ME De ME, de 'Modeling efficiency' (Nash and Sutcliffe, 1970; Greenwood *et al.*, 1985; Loague and Green, 1991) is gegeven als

$$ME = 1 - MSE / Var = [\sum_{i=1}^N (y_i - \bar{y})^2 - \sum_{i=1}^N (y_i - f_i)^2] / \sum_{i=1}^N (y_i - \bar{y})^2$$

r De Pearson's correlatiecoëfficiënt r geeft een indruk van de samenhang tussen twee variabelen. Deze coëfficiënt is ongevoelig voor eventuele bias, en daarnaast is zij alleen een maat voor lineaire samenhang. De Pearson's correlatiecoëfficiënt is gegeven als

$$r = \sum (y_i - \bar{y})(f_i - \bar{f}) / \sqrt{\sum (y_i - \bar{y})^2 \sum (f_i - \bar{f})^2}$$

Een aantal van de bovenstaande coëfficiënten wordt kritisch besproken door onder meer Legates & McCabe Jr. (1999), Clarke (2008), en Schaei & Gupta (2007). De maten hierboven maken vrijwel allemaal gebruik van gekwadraterde verschillen. Deze maten hebben handige statistische eigenschappen. Een mogelijk nadeel is dat grote verschillen hierdoor zwaarder meewegen dan kleine verschillen. Dat is niet altijd gewenst. In zo'n geval is het bijvoorbeeld mogelijk te kijken naar de gemiddelde absolute afwijking.

De verschillende overeenstemmingmaten zijn geïmplementeerd in diverse softwarepakketten. Een voorbeeld is het pakket IRENE (Fila *et al.*, 2003). Van de meeste genoemde en gebruikte coëfficiënten is in dit pakket direct duidelijk waar ze voor dienen.

Residuenplots

Met de bovenstaande maten wordt vooral de residuele variantie onderzocht. Als die klein is, dan liggen de observaties dicht bij de gefitte waarden. Zowel bij hoge als bij lage residuele varianties is het echter ook nuttig om ook te kijken naar de verdeling van de residuen. Daaraan kan bijvoorbeeld worden gezien, of er aan de aannames omtrent verdelingen wordt voldaan. Ten eerste zijn er de residuenplots. Daarin worden de gestandaardiseerde residuen geplot als functie van de gefitte waarde. Standaardiseren van de residuen is met name nodig wanneer een model wordt gefit met niet-normaal verdeelde errors. Als er in deze plot nog een trend zichtbaar is, dan is er ruimte voor verbetering van het model. Een tweede manier om de residuen grafisch te onderzoeken, zoals in de Q-Q plot. Hierin wordt de cumulatieve verdeling van de residuen uitgezet tegen de theoretische verwachte cumulatieve verdeling. Als aan de verdelingsaannamen voldaan is, dan liggen de punten in de grafiek op de 45 graden lijn. Beide plots kunnen worden gebruikt om uitbijters op het spoor te komen. Dit laatste kan ook door gebruik te maken van speciale uitbijtertests zoals bijvoorbeeld de Grubbs test (Grubbs, 1969).

Verschenen documenten in de reeks Werkdocumenten van de Wettelijke Onderzoekstaken Natuur & Milieu vanaf 2009

Werkdocumenten zijn verkrijgbaar bij het secretariaat van Unit Wettelijke Onderzoekstaken Natuur & Milieu, te Wageningen. T 0317 – 48 54 71; F 0317 – 41 90 00; E info.wnm@wur.nl

De werkdocumenten zijn ook te downloaden via de WOT-website www.wotnatuurenmilieu.wur.nl

2009

- 126** *Kamphorst, D.A.* Keuzes in het internationale biodiversiteitsbeleid; Verkenning van de beleidstheorie achter de internationale aspecten van het Beleidsprogramma Biodiversiteit (2008-2011)
- 127** *Dirkx, G.H.P. & F.J.P. van den Bosch.* Quick scan gebruik Catalogus groenblauwe diensten
- 128** *Loeb, R. & P.F.M. Verdonschot.* Complexiteit van nutriëntenlimitaties in oppervlaktewateren
- 129** *Kruit, J. & P.M. Veer.* Herfotografie van landschappen; Landschapsfoto's van de 'Collectie de Boer' als uitgangspunt voor het in beeld brengen van ontwikkelingen in het landschap in de periode 1976-2008
- 130** *Oenema, O., A. Smit & J.W.H. van der Kolk.* Indicatoren Landelijk Gebied; werkwijze en eerste resultaten
- 131** *Agricola, H.J.A.J. van Strien, J.A. Boone, M.A. Dolman, C.M. Goossen, S. de Vries, N.Y. van der Wulp, L.M.G. Groenemeijer, W.F. Lukey & R.J. van Til.* Achtergrond-document Nulmeting Effectindicatoren Monitor Agenda Vitaal Platteland
- 132** *Jaarrapportage 2008.* WOT-04-001 – Koepel
- 133** *Jaarrapportage 2008.* WOT-04-002 – Onderbouwend Onderzoek
- 134** *Jaarrapportage 2008.* WOT-04-003 – Advisering Natuur & Milieu
- 135** *Jaarrapportage 2008.* WOT-04-005 – M-AVP
- 136** *Jaarrapportage 2008.* WOT-04-006 – Natuurplanbureauafunctie
- 137** *Jaarrapportage 2008.* WOT-04-007 – Milieuplanbureauafunctie
- 138** *Jong de, J.J., J. van Os & R.A. Smidt.* Inventarisatie en beheerskosten van landschapselementen
- 139** *Dirkx, G.H.P., R.W. Verburg & P. van der Wielen.* Tegenkrachten Natuur. Korte verkenning van de weerstand tegen aankopen van landbouwgrond voor natuur
- 140** *Annual reports for 2008; Programme WOT-04*
- 141** *Vullings, L.A.E., C. Blok, G. Vonk, M. van Heusden, A. Huisman, J.M. van Linge, S. Keijzer, J. Oldengarm & J.D. Bulens.* Omgaan met digitale nationale beleidskaarten
- 142** *Vreke, J., A.L. Gerritsen, R.P. Kranendonk, M. Pleijte, P.H. Kersten & F.J.P. van den Bosch.* Maatlat Government – Governance
- 143** *Gerritsen, A.L., R.P. Kranendonk, J. Vreke, F.J.P. van den Bosch & M. Pleijte.* Verdrogingsbestrijding in het tijdperk van het Investeringsbudget Landelijk Gebied. Een verslag van casuonderzoek in de provincies Drenthe, Noord-Brabant en Noord-Holland.
- 144** *Luesink, H.H., P.W. Blokland, M.W. Hoogeveen & J.H. Wisman.* Ammoniakemissie uit de landbouw in 2006 en 2007
- 145** *Bakker de, H.C.M. & C.S.A. van Koppen.* Draagvlakonderzoek in de steigers. Een voorstudie naar indicatoren om maatschappelijk draagvlak voor natuur en landschap te meten
- 146** *Goossen, C.M.,* Monitoring recreatiegedrag van Nederlanders in landelijke gebieden. Jaar 2006/2007
- 147** *Hoefs, R.M.A., J. van Os & T.J.A. Gies.* Kavelruil en Landschap. Een korte verkenning naar ruimtelijke effecten van kavelruil.
- 148** *Klok, T.L., R. Hille Ris Lambers, P. de Vries, J.E. Tamis & J.W.M. Wijsman.* Quick scan model instruments for marine biodiversity policy.
- 149** *Spruijt, J., P. Spoorenberg & R. Schreuder.* Milieueffectiviteit en kosten van maatregelen gewasbescherming.
- 150** *Ehlert, P.A.I. (rapporteur).* Advies Bemonstering bodem voor differentiatie van fosfaatgebruiksnormen.
- 151** *Wulp van der, N.Y.* Storende elementen in het landschap: welke, waar en voor wie? Bijlage bij WOT-paper 1 – Krassen op het landschap
- 152** *Oltmer, K., K.H.M. van Bommel, J. Clement, J.J. de Jong, D.P. Rudrum & E.P.A.G. Schouwenberg.* Kosten voor habitattypen in Natura 2000-gebieden. Toepassing van de methode Kosteneffectiviteit natuurbeleid.
- 153** *Adrichem van, M.H.C., F.G. Wortelboer & G.W.W. Wamelink (2010).* MOVE. Model for terrestrial Vegetation. Version 4.0
- 154** *Wamelink, G.W.W., R.M. Winkler & F.G. Wortelboer.* User documentation MOVE4 v 1.0
- 155** *Gies de, T.J.A., L.J.J. Jeurissen, I. Staritsky & A. Bleeker.* Leefomgevingsindicatoren Landelijk gebied. Inventarisatie naar stand van zaken over geurhinder, lichthinder en fijn stof.
- 156** *Tamminga, S., A.W. Jongbloed, P. Bikker, L. Sebek, C. van Bruggen & O. Oenema.* Actualisatie excretiecijfers landbouwhuisdieren voor forfaits regeling Meststoffenwet
- 157** *Van der Salm, C., L. M. Boumans, G.B.M. Heuvelink & T.C. van Leeuwen.* Protocol voor validatie van het nutriëntenemissiemodel STONE op meetgegevens uit het Landelijk Meetnet effecten Mestbeleid
- 158** *Bouwma, I.M.* Quicksan Natura 2000 en Programma Beheer. Een vergelijking van Programma Beheer met de soorten en habitats van Natura 2000
- 159** *Gerritsen, A.L., D.A. Kamphorst, T.A. Selnes, M. van Veen, F.J.P. van den Bosch, L. van den Broek, M.E.A. Broekmeyer, J.L.M. Donders, R.J. Fontein, S. van Tol, G.W.W. Wamelink & P. van der Wielen.* Dilemma's en barrières in de praktijk van het natuur- en landschapsbeleid; Achtergronddocument bij Natuurbalans 2009.
- 160** *Fontein R.J., T.A. de Boer, B. Breman, C.M. Goossen, R.J.H.G. Henkens, J. Luttkik & S. de Vries.* Relatie recreatie en natuur; Achtergronddocument bij Natuurbalans 2009
- 161** *Deneer, J.W. & R. Kruijine. (2010).* Atmosferische depositie van gewasbeschermingsmiddelen. Een verkenning van de literatuur verschenen na 2003.
- 162** *Verburg, R.W., M.E. Sanders, G.H.P. Dirkx, B. de Knegt & J.W. Kuhlman.* Natuur, landschap en landelijk gebied. Achtergronddocument bij Natuurbalans 2009.
- 163** *Doorn van, A.M. & M.P.C.P. Paulissen.* Natuurgericht milieubeleid voor Natura 2000-gebieden in Europees perspectief: een verkenning.
- 164** *Smidt, R.A., J. van Os & I. Staritsky.* Samenstellen van landelijke kaarten met landschapselementen, grondeigendom en beheer. Technisch achtergronddocument bij de opgeleverde bestanden.
- 165** *Pouwels, R., R.P.B. Foppen, M.F. Wallis de Vries, R. Jochem, M.J.S.M. Reijnen & A. van Kleunen.* Verkenning LARCH: omgaan met kwaliteit binnen ecologische netwerken.

- 166 *Born van den, G.J., H.H. Luesink, H.A.C. Verkerk, H.J. Mulder, J.N. Bosma, M.J.C. de Bode & O. Oenema*. Protocol voor monitoring landelijke mestmarkt onder een stelsel van gebruiksnormen, versie 2009.
- 167 *Dijk, T.A. van, J.J.M. Driessen, P.A.I. Ehlert, P.H. Hotsma, M.H.M.M. Montforts, S.F. Plessius & O. Oenema*. Protocol beoordeling stoffen Meststoffenwet- Versie 2.1
- 168 *Smits, M.J., M.J. Bogaardt, D. Eaton, A. Karbauskas & P. Roza*. De vermaatschappelijking van het Gemeenschappelijk Landbouwbeleid. Een inventarisatie van visies in Brussel en diverse EU-lidstaten.
- 169 *Vreke, J. & I.E. Salverda*. Kwaliteit leefomgeving en stedelijk groen.
- 170 *Hengsdijk, H. & J.W.A. Langeveld*. Yield trends and yield gap analysis of major crops in the World.
- 171 *Horst, M.M.S. ter & J.G. Groenwold*. Tool to determine the coefficient of variation of DegT50 values of plant protection products in water-sediment systems for different values of the sorption coefficient
- 172 *Boons-Prins, E., P. Leffelaar, L. Bouman & E. Stehfest (2010)* Grassland simulation with the LPJmL model
- 173 *Smit, A., O. Oenema & J.W.H. van der Kolk*. Indicatoren Kwaliteit Landelijk Gebied
- 2010**
- 174 *Boer de, S., M.J. Bogaardt, P.H. Kersten, F.H. Kistenkas, M.G.G. Neven & M. van der Zouwen*. Zoektocht naar nationale beleidsruimte in de EU-richtlijnen voor het milieu- en natuurbeleid. Een vergelijking van de implementatie van de Vogel- en Habitatrichtlijn, de Kaderrichtlijn Water en de Nitraatrichtlijn in Nederland, Engeland en Noordrijn-Westfalen
- 175 *Jaarrapportage 2009*. WOT-04-001 – Koepel
- 176 *Jaarrapportage 2009*. WOT-04-002 – Onderbouwend Onderzoek
- 177 *Jaarrapportage 2009*. WOT-04-003 – Advisering Natuur & Milieu
- 178 *Jaarrapportage 2009*. WOT-04-005 – M-AVP
- 179 *Jaarrapportage 2009*. WOT-04-006 – Natuurplanbureauafunctie
- 180 *Jaarrapportage 2009*. WOT-04-007 – Milieuplanbureauafunctie
- 181 *Annual reports for 2009; Programme WOT-04*
- 182 *Oenema, O., P. Bikker, J. van Harn, E.A.A. Smolders, L.B. Sebek, M. van den Berg, E. Stehfest & H. Westhoek*. Quickscan opbrengsten en efficiëntie in de gangbare en biologische akkerbouw, melkveehouderij, varkenshouderij en pluimveehouderij. Deelstudie van project 'Duurzame Eiwitvoorziening'.
- 183 *Smits, M.J.W., N.B.P. Polman & J. Westerink*. Uitbreidingsmogelijkheden voor groene en blauwe diensten in Nederland; Ervaringen uit het buitenland
- 184 *Dirkx, G.H.P. (red.)*. Quick responsefunctie 2009. Verslag van de werkzaamheden.
- 185 *Kuhlman, J.W., J. Luijt, J. van Dijk, A.D. Schouten & M.J. Voskuilen*. Grondprijkskaarten 1998-2008
- 186 *Slangen, L.H.G., R.A. Jongeneel, N.B.P. Polman, E. Lianouridis, H. Leneman & M.P.W. Sonneveld*. Rol en betekenis van commissies voor gebiedsgericht beleid.
- 187 *Temme, A.J.A.M. & P.H. Verburg*. Modelling of intensive and extensive farming in CLUE
- 188 *Vreke, J.* Financieringsconstructies voor landschap
- 189 *Slangen, L.H.G.* Economische concepten voor beleidsanalyse van milieu, natuur en landschap
- 190 *Knotters, M., G.B.M. Heuvelink, T. Hoogland & D.J.J. Walvoort*. A disposition of interpolation techniques
- 191 *Hoogeveen, M.W., P.W. Blokland, H. van Kernebeek, H.H. Luesink & J.H. Wisman*. Ammoniakemissie uit de landbouw in 1990 en 2005-2008
- 192 *Beekman, V., A. Pronk & A. de Smet*. De consumptie van dierlijke producten. Ontwikkeling, determinanten, actoren en interventies.
- 193 *Polman, N.B.P., L.H.G. Slangen, A.T. de Blaeij, J. Vader & J. van Dijk*. Baten van de EHS; De locatie van recreatiebedrijven
- 194 *Veeneklaas, F.R. & J. Vader*. Demografie in de Natuurverkenning 2011; Bijlage bij WOT-paper 3
- 195 *Wascher, D.M., M. van Eupen, C.A. Mûcher & I.R. Geijzendorffer*. Biodiversity of European Agricultural landscapes. Enhancing a High Nature Value Farmland Indicator
- 196 *Apeldoorn van, R.C., I.M. Bouwma, A.M. van Doorn, H.S.D. Naeff, R.M.A. Hoefs, B.S. Elbersen & B.J.R. van Rooij*. Natuurgebieden in Europa: bescherming en financiering
- 197 *Brus, D.J., R. Vasat, G. B. M. Heuvelink, M. Knotters, F. de Vries & D. J. Walvoort*. Towards a Soil Information System with quantified accuracy; A prototype for mapping continuous soil properties
- 198 *Groot, A.M.E. & A.L. Gerritsen, m.m.v. M.H. Borgstein, E.J. Bos & P. van der Wielen*. Verantwoording van de methodiek Achtergronddocument bij 'Kwalitatieve monitor Systeeminnovaties verduurzaming landbouw'
- 199 *Bos, E.J. & M.H. Borgstein*. Monitoring Gesloten voer-mest kringlopen. Achtergronddocument bij 'Kwalitatieve monitor Systeeminnovaties verduurzaming landbouw'
- 200 *Kennismarkt 27 april 2010*; Van onderbouwend onderzoek Wageningen UR naar producten Planbureau voor de Leefomgeving.
- 201 *Wielen van der, P.* Monitoring Integrale duurzame stallen. Achtergronddocument bij 'Kwalitatieve monitor Systeeminnovaties verduurzaming landbouw'
- 202 *Groot, A.M.E. & A.L. Gerritsen*. Monitoring Functionele agrobiodiversiteit. Achtergrond-document bij 'Kwalitatieve monitor Systeeminnovaties verduurzaming landbouw'
- 203 *Jongeneel, R.A. & L. Ge*. Farmers' behavior and the provision of public goods: Towards an analytical framework.
- 204 *Vries, S. de, M.H.G. Custers & J. Boers*. Storende elementen in beeld; de impact van menselijke artefacten op de landschapsbeleving nader onderzocht.
- 205 *Vader, J. J.L.M. Donders & H.W.B. Bredenoord*. Zicht op natuur- en landschapsorganisaties; Achtergronddocument bij Natuurverkenning 2011.
- 206 *Jongeneel, R.A., L.H.G. Slangen & N.B.P. Polman*. Groene en Blauwe Diensten; Een raamwerk voor de analyse van doelen, maatregelen en instrumenten
- 207 *Letourneau, A.P., P.H. Verburg & E. Stehfest*. Global change of land use systems; IMAGE: a new land allocation module
- 208 *Heer, M. de*. Het Park van de Toekomst. Achtergronddocument bij Natuurverkenning 2011
- 209 *Knotters, M., J. Lahr, A.M. van Oosten-Siedlecka & P.F.M. Verdonschot*. Aggregation of ecological indicators for mapping aquatic nature quality. Overview of existing methods and case studies.
- 210 *Verdonschot, P.F.M. & A.M. van Oosten-Siedlecka*. *Graadmeters Aquatische natuur. Analyse gegevenskwaliteit Limnodata*
- 211 *Linderhof, V.G.M. & Hans Lenema*. *Quickscan kosteneffectiviteitsanalyse aquatische natuur*
- 212 *Leneman, H. V.G.M. Linderhof & R. Michels*. *Mogelijkheden voor het inbrengen van informatie uit de 'KRW database' in de 'KE database'*
- 213 *Schrijver, R.A.M., A. Corporaal, W.A. Ozinga & D. Rudrum*. *Kosteneffectieve natuur in landbouwgebieden; Methode om effecten van*

maatregelen voor de verhoging van biodiversiteit in landbouwgebieden te bepalen, een test in twee gebieden in Noordoost-Twente en West-Zeeuws-Vlaanderen

- 214** *Hoogland, T., R.H. Kemmers, D.G. Cirkel & J. Hunink.* Standplaatsfactoren afgeleid van hydrologische model uitkomsten; Methode-ontwikkeling en toetsing in het Drentse Aa-gebied.
- 215** *Agricola, H.J., R.M.A. Hoefs, A.M. van Doorn, R.A. Smidt & J. van Os.* Landschappelijke effecten van ontwikkelingen in de landbouw
- 216** *Kramer, H., J. Oldengarm en L.F.S. Roupioz.* Nederland is groener dan kaarten laten zien; Mogelijkheden om 'groen' beter te inventariseren en monitoren met de automatische classificatie van digitale luchtfoto's
- 218** *Hazeu, G.W., Kramer, H., J. Clement & W.P. Daamen.* Basiskaart Natuur 1990rev
- 219** *Boer, T.A. de.* Waardering en recreatief gebruik van Nationale Landschappen door haar bewoners
- 220** *Leneman, H., A.D. Schouten & R.W. Verburg.* Varianten van natuurbeleid: voorbereidende kostenberekeningen; Achtergronddocument bij Natuurverkenning 2011
- 221** *Knegt, B. de, J. Clement, P.W. Goedhart, H. Sierdsema, Chr. van Swaay & P. Wiersma.* Natuurkwaliteit van het agrarisch gebied
- 222** *Kamphorst, D.A. en M.M.P. van Oorschot.* Kansen en barrières voor verduurzaming van houtketens
- 224** *Bikker, P., M.M. van Krimpen & G.J. Remmelink.* Stikstofverteerbaarheid in voeders voor landbouwhuisdieren; Berekeningen voor de TAN-excretie
- 227** *Kleunen A. van, K. Koffijberg, P. de Boer, J. Nienhuis, C.J. Camphuysen, H. Schekkerman, K.H. Oosterbeek, M.L. de Jong, B. Ens & C.J. Smit.* Broedsucces van kustbroedvogels in de Waddenzee in 2007 en 2008
- 2011**
- 223** *Salm, C. van der en O.F. Schoumans.* Langetermijn effecten van verminderde fosfaatgiften
- 230** *Jaarrapportage 2010.* WOT-04-001 – Koepel
- 231** *Jaarrapportage 2010.* WOT-04-002 – Onderbouwend Onderzoek
- 232** *Jaarrapportage 2010.* WOT-04-003 – Advisering Natuur & Milieu
- 233** *Jaarrapportage 2010.* WOT-04-005 – M-AVP
- 234** *Jaarrapportage 2010.* WOT-04-006 – Natuurplanbureaufunctie
- 235** *Jaarrapportage 2010.* WOT-04-007 – Milieuplanbureaufunctie
- 236** *Arnouts, R.C.M. & F.H. Kistenkas.* Nederland op slot door Natura 2000: de discussie ontrafeld; Bijlage bij WOT-paper 7 – De deur klemt
- 237** *Harms, B. & M.M.M. Overbeek.* Bedrijven aan de slag met natuur en landschap; relaties tussen bedrijven en natuurorganisaties. Achtergrond-document bij Natuurverkenning 2011
- 238** *Agricola, H.J. & L.A.E. Vullings.* De stand van het platteland 2010. Monitor Agenda Vitaal Platteland; Rapportage Midterm meting Effectindicatoren
- 239** *Klijn, J.A.* Wisselend getij. Omgang met en beleid voor natuur en landschap in verleden en heden; een essayistische beschouwing. Achtergrond-document bij Natuurverkenning 2011