



ALTERRA

WAGENINGEN UR

Validatie van modelsystemen voor het voorspellen van de oppervlaktewaterkwaliteit en -kwantiteit in het stroomgebied 'de Drentse Aa'

Fases 1, 2 en 3

D.J.J. Walvoort, J. Roelsma en T.P. van Tol-Leenders

Alterra-rapport 1951, ISSN 1566-7197
Reeks Monitoring Stroomgebieden 18-I

Validatie van modelsystemen voor het voorspellen van de oppervlaktewaterkwaliteit
en -kwantiteit in het stroomgebied 'de Drentse Aa'

Validatie van modelsystemen voor het voorspellen van de oppervlaktewaterkwaliteit en -kwantiteit in het stroomgebied 'de Drentse Aa'

Fases 1, 2 en 3

D.J.J. Walvoort, J. Roelsma, T.P. van Tol-Leenders

Alterra-Rapport 1951

Reeks Monitoring Stroomgebieden nr. 18-I

Alterra, Wageningen, 2009

REFERAAT

Walvoort, D.J.J., J. Roelsma, van Tol-Leenders, T.P., 2009. *Validatie van modelsystemen voor het voorspellen van de oppervlaktewaterkwaliteit en -kwantiteit in het stroomgebied 'de Drentse Aa'*. Wageningen, Alterra-Rapport 1951, Reeks Monitoring Stroomgebieden nr. 18-I.

In het kader van het project 'Monitoring Stroomgebieden' wordt een instrument ontwikkeld dat tot doel heeft de bijdrage van de landbouw aan de belasting van het oppervlaktewater met stikstof en fosfor te kwantificeren en de veranderingen daarin als gevolg van het mestbeleid. Hiervoor is gekozen voor een aanpak waarbij naast metingen gebruik wordt gemaakt van modelsystemen. In dit rapport wordt de modelprestatie van de ontwikkelde modelsystemen gekwantificeerd op basis van validatie. Daarbij wordt gekeken hoe goed de ontwikkelde modelsystemen in staat zijn om temporele patronen in en absolute waarden van debieten en stikstof- en fosforconcentraties te beschrijven en te voorspellen.

Trefwoorden: waterafvoer, stikstof, fosfor, oppervlaktewater, validatie, statistiek

ISSN 1566-7197

Dit rapport is gratis te downloaden van www.alterra.wur.nl (ga naar 'Alterra-rapporten'). Alterra verstrekt geen gedrukte exemplaren van rapporten. Gedrukte exemplaren zijn verkrijgbaar via een externe leverancier. Kijk hiervoor op www.boomblad.nl/rapportenservice.

©2009 Alterra

Postbus 47; 6700 AA Wageningen; Nederland

Tel.: (0317) 480700; fax: (0317) 419000; e-mail: info.alterra@wur.nl

Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welk ander wijze ook zonder voorafgaande schriftelijke toestemming van Alterra.

Alterra aanvaardt geen aansprakelijkheid voor eventuele schade voortvloeiend uit het gebruik van de resultaten van dit onderzoek of de toepassing van de adviezen.

Inhoudsopgave

1	Inleiding	13
1.1	Achtergrond	13
1.2	Doelstelling	14
1.3	Leeswijzer	14
2	Werkwijze	15
2.1	Fasering van modelsimulaties	15
2.2	Verwerken van meetgegevens en modelresultaten	15
2.2.1	Meetgegevens	17
2.2.2	Modelresultaten	17
2.2.3	Koppelen van modelresultaten aan meetgegevens	17
2.3	Kwantificeren van modelprestaties	19
2.3.1	Inleiding	19
2.3.2	Doel van de validatie	20
2.3.3	Aanpak	21
2.3.4	Modelprestatie-maten	21
2.3.5	Interpretatie van afzonderlijke modelprestatie-maten	22
2.3.6	Interpretatie van modelprestatie-maten in onderlinge samenhang	25
3	Resultaten en discussie	31
3.1	Waterafvoer	32
3.2	Stikstof	39
3.2.1	Anorganisch stikstof	39
3.2.2	Organisch stikstof	44
3.2.3	Totaal stikstof	47
3.3	Fosfor	49
3.3.1	Anorganisch fosfor	49
3.3.2	Organisch fosfor	52
3.3.3	Totaal fosfor	54
4	Conclusies	57
4.1	Waterafvoer	57
4.2	Stikstof	58
4.3	Fosfor	58
4.4	Algemeen	59

5	Aanbevelingen	61
6	Dankbetuiging	63
A	Modelprestatie­maten	67
A.1	Algemeen foutenmodel	67
A.2	Variantie	67
A.3	Correlatie­coëfficiënt	68
A.4	Systematische of gemiddelde fout	68
A.5	Willekeurige fout	69
A.6	Totale fout (gemiddelde gekwadraterde fout)	69
A.7	Gemiddelde absolute fout	69
A.8	Model­efficiëntie	70
B	Taylor-diagram	71
C	Zonneplot	75
D	Datamodel modelresultaten	79

Woord vooraf

Deze rapportage vormt een onderdeel van het project ‘Monitoring Stroomgebieden’. Het primaire doel van dit project is het leveren van een bijdrage aan de evaluatie van het mestbeleid door het kwantificeren van het aandeel van de landbouw in de belasting van het oppervlaktewater en de verandering van dit aandeel van de landbouw als gevolg van (mest)beleid in een aantal representatieve stroomgebieden in karakteristieke landschappelijke regio’s. Het secundaire doel is om een methodiek te ontwikkelen die het mogelijk maakt en perspectieven biedt om deze methodiek ook in andere stroomgebieden in te voeren.

Voor dit project zijn vier pilotgebieden geselecteerd: Drentse Aa, Schuitenbeek, Krimpenerwaard en Quarles van Ufford. De waterbeheerders, Hoogheemraadschap van Schieland en de Krimpenerwaard, Waterschap Veluwe, Waterschap Rivierenland, Waterschap Hunze en Aa’s en Waterlaboratorium Noord participeren actief in dit project.

Het project wordt aangestuurd door een stuurgroep. In de stuurgroep hebben de Ministeries LNV, VROM en V&W als opdrachtgevers en een vertegenwoordiger van de Unie van Waterschappen namens de betrokken waterbeheerders zitting. Het project wordt uitgevoerd door Alterra Research Instituut voor de Groene Ruimte onderdeel van Wageningen Universiteit en Researchcentrum.

In de reeks rapportages van het project ‘Monitoring Stroomgebieden’ is per gebied een systeemverkenning verschenen. Op basis van deze uitgevoerde systeemverkenning wordt er aanvullend in de pilotgebieden gemeten en is het gefaseerd opzetten van een modelinstrumentarium per pilotgebied gestart. Fase 1 van het modelinstrumentarium is een gebiedsuitsnede uit het nationale instrumentarium STONE. STONE is ontwikkeld voor beleidsvoorbereiding en ex-ante evaluatie op nationale schaal. Vervolgens is in fase 2 een regionale schematisatie (bodem, water, gewas) uitgevoerd en is een module voor de beschrijving van de processen in het oppervlaktewater toegevoegd. In 2007 en 2008 is fase 3 uitgevoerd waarbij regiospecifieke data in de modellering zijn toegepast. In fase 3 zijn ook deelfases onderscheiden. Deze deelfases geven inzicht in het effect van specifiek regionale data.

Het doel van dit onderzoek is om de bruikbaarheid van al deze gefaseerde modelsystemen die in het kader van het project Monitoring Stroomgebied zijn ontwikkeld te evalueren ten behoeve van monitoring.

Voor informatie over het project ‘Monitoring Stroomgebieden’ kunt u terecht op www.monitoringstroomgebieden.nl of bij:

Dorothee van Tol-Leenders
0317 – 48 42 79
dorothee.vantol-leenders@wur.nl

Frank van der Bolt
0317 – 48 64 44
frank.vanderbolt@wur.nl

Samenvatting

Inleiding

In het kader van het project ‘Monitoring Stroomgebieden’ wordt een instrument ontwikkeld dat tot doel heeft de bijdrage van de landbouw aan de belasting van het oppervlaktewater met stikstof en fosfor te kwantificeren en de veranderingen daarin als gevolg van het mestbeleid. Hiervoor is gekozen voor een aanpak waarbij naast metingen gebruik wordt gemaakt van modelsystemen. Modellen maken het mogelijk om debieten, vrachten en concentraties te voorspellen op locaties waar en momenten waarop geen metingen beschikbaar zijn. Ook kunnen toekomstige situaties worden doorgerekend zodat effecten van potentiële maatregelen op voorhand kunnen worden geëvalueerd. Daarnaast kan modelanalyse leiden tot een beter begrip van de belangrijkste processen van het land- en oppervlaktewatersysteem en biedt zij de mogelijkheid om de metingen in het oppervlaktewater te relateren aan bronnen.

Validatie

Voorwaarde is wel dat de modellen de processen in het landsysteem en het oppervlaktewatersysteem in voldoende mate beschrijven. In dit rapport wordt dit nagegaan door middel van validatie. Bij validatie worden modelresultaten vergeleken met metingen. Het is van belang dat deze metingen niet eerder bij de modellering zijn gebruikt om te voorkomen dat de modelleurs (bewust of onbewust) naar de metingen toewerken. Dit kan leiden tot een veel te optimistisch beeld van de modelprestatie wat ten koste gaat van de inzetbaarheid van het model.

Bij de validatie is in eerste instantie gekeken of de modellen in staat zijn om temporele patronen (denk daarbij aan trends, de seizoensdynamiek, e.d.) in waterafvoeren en nutriëntenconcentraties te voorspellen. De modelresultaten hoeven daarbij niet exact overeen te komen met de metingen, het is voldoende als ze het temporele patroon in de metingen volgen. Een model dat patronen goed kan beschrijven, kan worden gebruikt om trends te voorspellen op basis waarvan tijdig maatregelen kunnen worden getroffen.

Vervolgens is gekeken of de modelsystemen de meetwaarden zelf kunnen voorspellen. Het gaat nu niet alleen meer om overeenkomsten in het temporele patroon, maar ook om het accuraat voorspellen van de meetwaarden. Dit is bijvoorbeeld van belang om te kunnen voorspellen of in de toekomst de con-

concentraties beneden de norm zullen blijven.

De resultaten komen grotendeels overeen met de bevindingen in de systeem-analyses. Echter, een belangrijk verschil met de systeemanalyse is dat de modelleurs alleen uitspraken konden doen over de modelprestatie tot 2000. De modelprestatie voor de periode daarna konden ze alleen op kwalitatieve wijze beoordelen (analyse van stof- en waterbalansen, en tijdreeksen) omdat ze niet de beschikking hadden over de meetgegevens. Die waren immers achtergehouden om een validatie mogelijk te maken.

Waterafvoer

De modelsystemen zijn in staat om het temporele patroon van de afvoermetingen te reproduceren en te voorspellen. Het zomerhalfjaar wordt vrijwel even goed gereproduceerd en voorspeld als het winterhalfjaar. De deelfase waarbij de hydrologie van het landsysteem wordt geregionaliseerd heeft een positief effect op het reproduceren en voorspellen van het temporele patroon. Toevoeging van de kleinste waterlopen reduceert de modelprestatie enigszins. De deelfase ‘neerslag/verdamping oppervlaktewater’ heeft vervolgens weer een gunstig effect op het reproduceren en voorspellen van temporele patronen.

De modelsystemen hebben voor de meeste meetlocaties ook een meerwaarde voor wat betreft het reproduceren en voorspellen van absolute waterafvoeren.

Het corrigeren van de leggergegevens heeft alleen een negatief effect op de modelprestatie voor de periode tot 2000. Waarschijnlijk mogen de gecorrigeerde leggergegevens alleen gebruikt worden voor het modelleren van recente perioden. Voor perioden verder terug in de tijd moet gebruik worden gemaakt van historische leggerinformatie.

Ook de deelfase waarin de kleinste waterlopen worden gemodelleerd vermindert de modelprestatie. Waarschijnlijk worden de afvoerpieken te veel gedempt.

Stikstof

De modellen zijn in staat om het temporele patroon in anorganisch stikstof te reproduceren en te voorspellen. De regionalisering van de hydrologie laat in eerste instantie een verslechtering zien waarna met name het toevoegen van de diffuse detailontwatering om de retentie in de kleinste waterlopen te simuleren tot een verbetering leidt. De laatste deelfase van fase 3, waarin de parameters van het oppervlaktewaterkwaliteitsmodel zijn geregionaliseerd, leidt echter tot een afname van de modelprestatie.

Als wordt gekeken naar absolute concentraties in plaats van naar temporele patronen dan blijkt dat alleen het fase 2 model in staat is om anorganisch stikstof te reproduceren en te voorspellen. De fase 1 en 3 systemen overschatten

de metingen. De achteruitgang in modelprestatie tussen fase 2 en fase 3 wordt veroorzaakt door deelfase ‘regionale kwaliteitsdata land’. Met name de afname van de denitrificatie is hiervoor verantwoordelijk waardoor meer anorganisch stikstof beschikbaar komt voor uitspoeling naar het oppervlaktewater.

De modelsystemen hebben meer moeite om de temporele patronen in en absolute concentraties van organisch stikstof te reproduceren en te voorspellen. De metingen worden over het algemeen onderschat. De modelresultaten hebben geen meerwaarde ten opzichte van de gemiddelde meetwaarde.

Fosfor

Het blijkt lastig om het temporele patroon in anorganische fosforgehalten te modelleren. Absolute concentraties zijn nog moeilijker te voorspellen. Over het algemeen hebben de modelsystemen geen meerwaarde boven het gemiddelde van de metingen. Dat geldt voor zowel puntmetingen (momentopnames) als weekgemiddelde waarnemingen op basis van debietproportionele metingen. Soortgelijke conclusies kunnen worden getrokken voor organisch fosfor en totaal fosfor.

Conclusies

In het kader van het project ‘Monitoring Stroomgebieden’ wordt een modelsysteem ontwikkeld dat uiteindelijk de ruggengraat moet gaan vormen van een monitoringsysteem. Het modelsysteem moet informatie verschaffen over waterafvoeren en nutriëntenconcentraties op locaties waarvoor en tijdstippen waarop geen metingen beschikbaar zijn. Het modelsysteem wordt gefaseerd ontwikkeld. Elke fase bouwt voort op vorige fases.

Uit de validatieresultaten blijkt dat de modelsystemen van fases 1 tot en met 3 nog niet altijd in staat zijn om overal in het beheersgebied de *absolute* nutriëntenconcentraties te reproduceren en te voorspellen. Het is dan nog niet goed mogelijk om uitspraken te doen over of kritische concentraties al dan niet worden overschreden. Uitzondering is het fase 2 model dat een meerwaarde heeft met betrekking tot het voorspellen van anorganisch en totaal stikstof. Met de fase 3 modellen is het op een aantal locaties al wel mogelijk om het temporele patroon te reproduceren en te voorspellen. Het modelsysteem zegt dan niets over de absolute grootte van de stikstofconcentratie, maar wel of er sprake is van een *verandering* van de concentratie. Voor het monitoren van trends is dit waardevolle informatie. Voor het monitoren of een kritische waarde wordt overschreden is het goed voorspellen van enkel het temporele patroon onvoldoende. Dan dienen de modellen ook in staat te zijn om absolute concentraties te voorspellen.

1 Inleiding

1.1 Achtergrond

In het kader van het project ‘Monitoring Stroomgebieden’ wordt een instrument ontwikkeld dat tot doel heeft de bijdrage van de landbouw aan de belasting van het oppervlaktewater met stikstof en fosfor te kwantificeren en de veranderingen daarin als gevolg van het mestbeleid. Hiervoor is gekozen voor een aanpak waarbij naast metingen gebruik wordt gemaakt van modelsystemen. Modellen maken het mogelijk om debieten, vrachten en concentraties te voorspellen op locaties waar en momenten waarop geen metingen beschikbaar zijn. Ook kunnen toekomstige situaties worden doorgerekend zodat effecten van potentiële maatregelen op voorhand kunnen worden geëvalueerd. Daarnaast kan modelanalyse leiden tot een beter begrip van de belangrijkste processen van het land- en oppervlaktewatersysteem en biedt zij de mogelijkheid om de metingen in het oppervlaktewater te relateren aan bronnen. Dat laatste is niet mogelijk gebleken op basis van metingen alleen (Portielje et al., 2002). Het belang van modellen ten behoeve van monitoring wordt ook in de internationale literatuur steeds meer onderkend (Jørgensen et al., 2007).

Voorwaarde is wel dat de modellen de processen in het landsysteem en het oppervlaktewatersysteem in voldoende mate beschrijven. Dit kan worden nagegaan door middel van validatie¹. Bij validatie worden de resultaten van een model vergeleken met metingen die niet eerder bij de modellering zijn gebruikt. Dergelijke metingen worden daarom ook wel *onafhankelijk* genoemd (van Waveren et al., 1999).

¹ Strikt genomen is het niet mogelijk om een model te valideren. Een model dat een set onafhankelijke metingen correct beschrijft schept vertrouwen, maar is niet noodzakelijkerwijs valide. Er is immers geen garantie dat het model elke willekeurige nieuwe set onafhankelijke metingen ook goed zal beschrijven. Mocht keer op keer blijken dat een model nieuwe metingen correct beschrijft dan geeft dat extra vertrouwen in het model. Faalt het model daarentegen, dan is het model ongeschikt in zijn huidige vorm en zal het moeten worden aangepast. Het verwerpen van een model wordt ook wel invalidatie of falsificatie genoemd. Een model is pas valide als het de werkelijkheid onder alle omstandigheden kan beschrijven. De werkelijkheid is echter onbekend. We beschikken meestal alleen over een afspiegeling van de werkelijkheid in de vorm van een beperkt aantal metingen dat behept is met fouten. De validiteit van een model kan dan ook strict genomen niet worden aangetoond. In plaats van modelvalidatie is het dan ook beter om te spreken van modelconfirmatie. Voor meer achtergronden wordt verwezen naar Konikow and Bredehoeft (1992) en Oreskes et al. (1994). Desalniettemin zal in dit rapport de term ‘(model)validatie’ worden gebruikt omdat deze term algemeen gangbaar is in het waterbeheer (van Waveren et al., 1999)

1.2 Doelstelling

Het doel van dit rapport is om de bruikbaarheid van de modellen die voor de Drentse Aa zijn ontwikkeld te evalueren ten behoeve van monitoring. Dit zal worden gedaan op basis van validatie. De modelsystemen en modelresultaten zijn uitgebreid beschreven in Roelsma et al. (2006a,b, 2008).

1.3 Leeswijzer

De opbouw van het rapport is als volgt. Eerst wordt de werkwijze beschreven en toegelicht aan de hand van een aantal eenvoudige voorbeelden. Vervolgens worden de resultaten voor de Drentse Aa gepresenteerd, bediscussieerd, en de conclusies kort samengevat. Tot slot wordt een aantal aanbevelingen geformuleerd. Inhoudelijke details zijn opgenomen in de bijlagen.

2 Werkwijze

2.1 Fasering van modelsimulaties

Bij het project ‘Monitoring Stroomgebieden’ is een gefaseerde aanpak gevolgd. Dit rapport heeft betrekking op de eerste drie hoofdfases en bijbehorende deelfases. De gebruikte modelsystemen in elke deelfase en de modelresultaten zijn uitvoerig beschreven in Roelsma et al. (2006a,b, 2008). In elke deelfase is een model ontwikkeld rond een bepaald thema zoals gegeven in tabel 2.1. Elke deelfase is een verfijning van voorgaande deelfases. Bijvoorbeeld, deelfase ‘diffuse detailontwatering’ voegt de invloed van kleine waterlopen toe aan deelfase ‘RWZI’, waarin rioolwaterzuiveringsinstallaties zijn toegevoegd aan het modelsysteem. Door deze deelfases met elkaar te vergelijken kan het effect van het meemodelleren van het haarvatensysteem worden geëvalueerd.

De meerwaarde van een bepaalde deelfase is sterk afhankelijk van de modelverfijningen in voorgaande deelfases. Met andere woorden, de meerwaarde van elke deelfase is afhankelijk van de gehanteerde volgorde in tabel 2.1. Er is een groot aantal volgordes mogelijk. Het zou te veel menskracht en rekentijd vergen om alle relevante volgordes door te rekenen en te evalueren. Daarom zijn door een groep deskundigen op het gebied van regionale modellering keuzes gemaakt op basis waarvan de volgorde in tabel 2.1 is vastgesteld.

Het voordeel van een gefaseerde aanpak is dat de meerwaarde van elke deelfase kan worden gekwantificeerd. Hiermee wordt kennis opgedaan op basis waarvan voor nieuwe stroomgebieden kan worden ingeschat of het zinvol is om al dan niet een bepaalde fase uit te voeren. Deze kennis zal worden aangewend om de blauwdruk ‘Monitoring Stroomgebieden’ vorm te geven.

2.2 Verwerken van meetgegevens en modelresultaten

De metingen en de modelresultaten zijn opgeslagen in relationele database management systemen (RDBMS). In een RDBMS worden niet alleen gegevens opgeslagen maar ook informatie over hoe die gegevens met elkaar samenhangen (volgens het zogenaamde relationele model (Codd, 1970)). Het voordeel van een dergelijk systeem is dat de gegevens gestructureerd zijn opgeslagen en snel kunnen worden opgevraagd.

Tabel 2.1: Overzicht van de belangrijkste hoofd- en deelfases.

	hoofdfase	deelfase	omschrijving
1	fase 1	nationaal instrumentarium	gebiedsuitsnede nationaal instrumentarium
2	fase 2	+ regionale schematisatie + opp.water	fase 1 met regionale schematisatie en opp.watersysteem
3		+ regionale kwantiteitsdata land	fase 2 met regionale data landsysteem kwantiteit
4		+ regionale kwaliteitsdata land	toevoeging regionale data landsysteem kwaliteit
5		+ correctie legger	correctie leggegevens
6		+ inlaat (kwantiteit)	toevoeging inlaat (kwantiteit)
7		+ stedelijk gebied	toevoeging stedelijk gebied
8		+ RWZI	toevoeging RWZI (kwantiteit en kwaliteit)
9		+ diffuse detailontwatering	toevoeging diffuse detailontwatering
10		+ neerslag/verdamping opp.water	toevoeging neerslag en verdamping hoofdwaterlopen
11		+ atmosferische depositie opp.water	toevoeging atmosferische depositie hoofdwaterlopen
12		+ inlaat (kwaliteit)	toevoeging inlaat (kwaliteit)
13	fase 3	+ parameters opp.kwaliteitsmodel	toepassing regionale parameters opp.kwaliteitsmodel

2.2.1 Meetgegevens

De meetgegevens worden aangeleverd door Waterschap Hunze en Aa's en Waterlaboratorium Noord en vervolgens opgeslagen in een RDBMS. Alle meetgegevens die naar de database worden geüpload passeren eerst een filter dat een uitgebreide kwaliteitscontrole (QC) uitvoert. Hierbij worden de meetgegevens automatisch gecontroleerd op een groot aantal punten variërend van potentiële typfouten tot fouten in de nutriëntenbalansen. Aangetroffen fouten worden gerapporteerd als webpagina en gecommuniceerd met Waterschap Hunze en Aa's. Op deze wijze wordt de kans verkleind dat foutieve metingen in de database terechtkomen. Tevens bevat de database procedures die de concentraties van ontbrekende stoffen kunnen schatten op basis van aangeleverde concentraties. Denk daarbij aan het schatten van mineraal stikstof op basis van ammonium, nitraat en nitriet. Zie Walvoort and van Tol-Leenders (2009) voor meer informatie.

2.2.2 Modelresultaten

De modelsystemen schrijven de resultaten weg naar gestructureerde gegevensbestanden. De reden om de modelresultaten op te slaan in een RDBMS is dan ook niet zozeer de behoefte aan gestructureerde opslag als wel datacompressie en de mogelijkheid om op efficiënte wijze gegevens te selecteren uit de bijna terabyte aan modelgegevens¹.

2.2.3 Koppelen van modelresultaten aan meetgegevens

Bij validatie worden modelresultaten vergeleken met meetgegevens. Hiervoor is het van belang dat de juiste modelresultaten aan de juiste meetgegevens worden gekoppeld. Er kan een onderscheid worden gemaakt tussen koppelen op basis van chemische speciatie en op basis van meetperiode.

Koppelen op basis van chemische speciatie

De chemische speciatie van de nutriënten die wordt gemeten komt niet altijd overeen met de chemische speciatie die door het model wordt berekend. Het oppervlaktewaterkwaliteitsmodel (Siderius et al., 2008) berekent bijvoorbeeld niet de hoeveelheid nitraat (NO_3^-), nitriet (NO_2^-) en ammonium (NH_4^+) afzonderlijk, maar enkel de som daarvan, namelijk de hoeveelheid anorganisch stikstof. De laboratoria daarentegen rapporteren meestal wel de afzonderlijke stikstofcomponenten, maar niet de som als anorganisch stikstof. Voor het koppelen van de meetgegevens aan de modelresultaten is het dan ook noodza-

¹Voor de opslag van een terabyte aan gegevens zijn circa 1500 compact discs nodig.

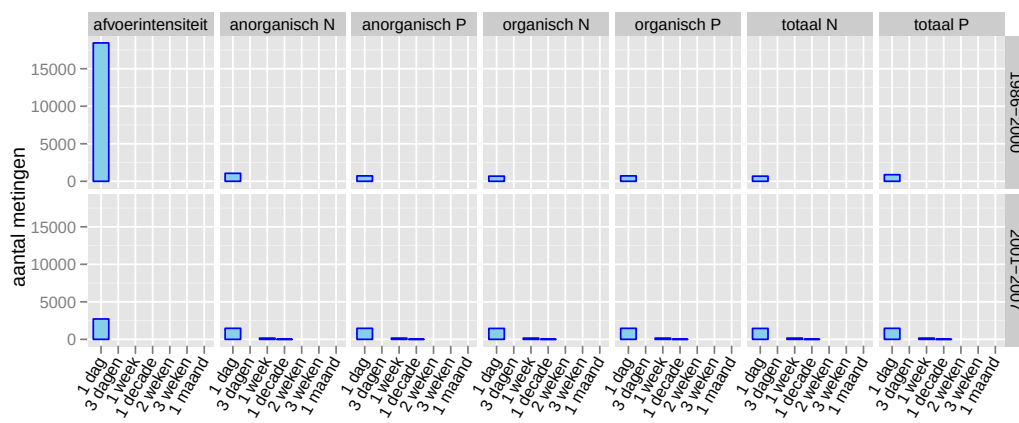
kelijk om eerst de hoeveelheid anorganisch stikstof uit te rekenen op basis van de gemeten stikstofcomponenten.

De database die voor het project ‘Monitoring Stroomgebieden’ wordt gebruikt (paragraaf 2.2.1) bevat rekenregels die deze conversies automatisch uitvoeren. Een overzicht van deze rekenregels is gegeven in Walvoort and van Tol-Leenders (2009).

Een bijkomende complicatie is dat de chemische speciatie die het model berekent kan afwijken van de chemische speciatie die wordt gemeten. Bijvoorbeeld, totaal fosfor zoals door het oppervlaktewaterkwaliteitsmodel wordt berekend bevat *geen* particulier fosfor. Ook is niet altijd bekend of de gemeten totale fosforconcentratie particulier fosfor bevat. Als de metingen wel particulier fosfor bevatten dan introduceert deze mismatch een schijnbare ‘systematische fout’ (par. 2.3.5), ook in het geval van een perfect (foutloos) model en foutloze metingen.

Koppelen op basis van meetperiode

Voor validatie is het van belang dat het temporele aggregatieniveau van de modelresultaten goed aansluit bij dat van de meetgegevens. De meetgegevens zijn beschikbaar op verschillende temporele aggregatieniveaus variërend van puntmetingen (momentopnames) tot meerdaagse gemiddelden op basis van tijd- en debietproportionele metingen. Zie figuur 2.1 voor het aantal metingen dat beschikbaar is voor elk aggregatieniveau.



Figuur 2.1: het aantal metingen dat per aggregatieniveaus beschikbaar is voor de afvoerintensiteit ($\text{m}^3 \text{s}^{-1}$), en de stikstof- en fosforconcentraties (mg l^{-1}) voor twee periodes.

De modelresultaten voor het landsysteem van fases 1 en 2 zijn beschikbaar als decadegemiddelden, die van fase 3 als daggemiddelden. Alleen voor fases 2

Tabel 2.2: Aggregatie van modelresultaten die betrekking hebben op het oppervlaktewatersysteem (- = geen aggregatie, + = aggregatie, d = dag, w = week, m = maand)

modelresultaten		meetresultaten						
		$\leq 1d$	3d	1w	10d	2w	3w	1m
fase 1	10d (bodem)	-	-	-	-	+	+	+
fase 2	1d (opp.water)	-	+	+	+	+	+	+
fase 3	1d (opp.water)	-	+	+	+	+	+	+

en 3 is het oppervlaktewatersysteem gemodelleerd. De resultaten daarvan zijn voor beide fasen beschikbaar als daggemiddelden.

Om modelresultaten te koppelen aan meetgegevens is het temporele aggregatieniveau van de meetgegevens als uitgangspunt genomen. Het is immers altijd mogelijk om modelresultaten te genereren op hetzelfde aggregatieniveau als de meetgegevens. Het omgekeerde is meestal niet mogelijk. Daarom zullen alleen modelresultaten worden geaggregeerd en worden meetresultaten als gegeven beschouwd. Modelresultaten worden in de regel alleen geaggregeerd als de metingen op een hoger aggregatieniveau beschikbaar zijn dan de modelresultaten (tabel 2.2). De modelresultaten van fase 1 zijn bijvoorbeeld beschikbaar als decadegemiddelden. Deze worden alleen geaggregeerd indien de meetresultaten betrekking hebben op hogere aggregatieniveaus dan 10 dagen (d.w.z. 2 weken, 3 weken of 1 maand)². De resultaten van de fase 2 en 3 modellen zijn daarentegen beschikbaar op dagbasis. Deze worden geaggregeerd indien de meetresultaten beschikbaar zijn als meerdaagsegemiddelden (3 dagen, 1 week, 10 dagen, 2 weken, 3 weken of 1 maand).

Bij het aggregeren van concentraties (mg l^{-1}) moet rekening worden gehouden met het watervolume. Rekenkundig middelen van daggemiddelde concentraties is niet correct. Daarom wordt de gemiddelde concentratie voor een bepaalde periode berekend als het quotiënt van de totale hoeveelheid stof (kg) en het totale watervolume (m^3) gedurende die periode.

2.3 Kwantificeren van modelprestaties

2.3.1 Inleiding

Onder modelprestatie (Eng: *model performance*) wordt verstaan de mate waarin het model de metingen kan beschrijven. De modelprestatie zal worden gekwantificeerd door de modelresultaten te vergelijken met meetgegevens. Een

²Merk op dat het hogere aggregatieniveau geen veelvoud hoeft te zijn van het lagere aggregatieniveau. Het gebruikte aggregatiealgoritme kan namelijk ook overweg met fracties van het lagere aggregatieniveau.

deel van de meetgegevens is door de modelleurs gebruikt om het modelsysteem mee te bouwen en te controleren. De modelprestatie die op basis van deze meetgegevens wordt berekend, geeft slechts aan hoe goed het betreffende modelsysteem de meetgegevens kan *beschrijven*. Ze is echter ongeschikt om het *voorspellend* vermogen van het model te kwantificeren. Immers, omdat de modelleurs de beschikking hadden over de meetgegevens bestaat het risico dat er bewust of onbewust naar de meetgegevens is toegewerkt. Het voorspellend vermogen kan alleen worden vastgesteld op basis van onafhankelijke meetgegevens. Dit zijn meetgegevens die niet bij het modelleren zijn gebruikt. Het kwantificeren van de modelprestatie op basis van onafhankelijke meetgegevens wordt ook wel validatie genoemd (van Waveren et al., 1999).

In het onderstaande zal eerst het doel van de validatie nader worden uitgewerkt. Vervolgens zal de gehanteerde methodiek worden beschreven en de maten die worden gebruikt om de modelprestatie te kwantificeren.

2.3.2 Doel van de validatie

In dit rapport zullen de volgende vragen worden beantwoord:

1. Zijn de modellen geschikt om veranderingen in debieten en concentraties te voorspellen? Het gaat daarbij om het voorspellen van temporele patronen zoals trends en de seizoensdynamiek;
2. Zijn de modellen geschikt om absolute debieten en concentraties te voorspellen?

Bij vraag 1 gaat het om het temporele patroon van debieten en concentraties. Een model dat in staat is temporele patronen te voorspellen kan worden gebruikt om trends te kwantificeren. Het gaat bij vraag 1 om de *richting* en niet om de *grootte* van de trend. Vraag 1 is relevant voor veel beleidsvragen. Immers, vaak wil men in eerste instantie alleen weten *of* de oppervlaktewaterkwaliteit de komende periode zal veranderen gegeven het huidige beleid.

Vraag 2 is ambitieuzer. Hier gaat het niet alleen om het temporele patroon, maar ook om de mate waarin gemodelleerde waarden overeenkomen met metingen. Een model dat absolute debieten en concentraties kan voorspellen kan bijvoorbeeld worden gebruikt om na te gaan of aan de normen wordt voldaan. Merk op dat vraag 2 vraag 1 omvat, immers wanneer de debieten en concentraties goed kunnen worden voorspeld, zal dat ook gelden voor het temporele patroon. Het omgekeerde hoeft echter niet op te gaan.

2.3.3 Aanpak

Voor validatie is het noodzakelijk dat de meetgegevens onafhankelijk zijn. Daarom is de totale rekenperiode opgedeeld in twee delen:

De modelbouwperiode 1986-2000: De meetgegevens van deze periode zijn (deels) gebruikt om de modelsystemen op te zetten. De modelsystemen zijn gecontroleerd op fouten door de modelresultaten te vergelijken met deze meetgegevens. Bij het opzetten van de modelsystemen is vooral gekeken naar uitstroompunten van (deel-)stroomgebieden, en niet (of minder) naar tussenliggende meetlocaties. Voor die uitstroompunten kan het beschrijvend vermogen van de modelsystemen worden geëvalueerd. Voor de tussenliggende locaties kan de periode 1986-2000 ook worden gebruikt om te beoordelen in hoeverre de modelsystemen geschikt zijn als ruimtelijke interpolator. Dat wil zeggen, het kwantificeren van de water- en nutriëntentoestand op locaties waar geen metingen beschikbaar zijn.

De validatieperiode 2001-2007: De meetgegevens van deze periode zijn niet gebruikt door de modelleers en zijn daarom geschikt voor validatie. Deze periode kan worden gebruikt om te beoordelen in hoeverre de modelsystemen in staat zijn om toekomstige situaties te voorspellen. Een modelsysteem dat daartoe in staat is, is een waardevolle aanvulling op het monitoringmeetnet, omdat *pro-actief* effecten van maatregelen kunnen worden berekend.

2.3.4 Modelprestatie-maten

De vragen in paragraaf 2.3.2 zullen worden beantwoord door modelresultaten te vergelijken met metingen. De discrepantie tussen metingen en modelresultaten zullen worden gekwantificeerd op basis van een aantal modelprestatie-maten (Eng.: *model performance measures*). Een overzicht van veelgebruikte modelprestatie-maten is gegeven in bijlage A. Er bestaat echter geen universele modelprestatie-maat waarmee elke vraag kan worden beantwoord (Stow et al., 2009). Zo is bijvoorbeeld Pearsons correlatiecoëfficiënt (verg. A.7) een veelgebruikte maat om de overeenkomst tussen het patroon van de metingen en de modelresultaten te kwantificeren (Taylor, 2001; Jolliff et al., 2009). Echter, daar deze maat ongevoelig is voor systematische en proportionele verschillen tussen waarnemingen en modelresultaten (Legates and McCabe, 1999) is ze bijvoorbeeld ongeschikt om systematische fouten te identificeren. De modelprestatie-maten zijn veelal complementair aan elkaar en kunnen elkaar deels overlappen. Ze moeten daarom ook in onderlinge samenhang worden bestudeerd. Grafische hulpmiddelen zijn daarbij onmisbaar (zie bijlage B en C).

Tabel 2.3: Overzicht van enkele veelgebruikte modelprestatie-maten, het gehanteerde symbool, het bereik (minimum, maximum), de optimale waarde en een verwijzing naar een vergelijking in bijlage A.

modelprestatie-maat	symbool	bereik	optimum	vergelijking
genormaliseerde standaardafwijking	$\tilde{\sigma}_{\hat{y}}$	$[0, +\infty)$	1	A.6
Pearsons correlatiecoëfficiënt	r	$[-1, 1]$	1	A.7
systematische fout	$\bar{e}$	$(-\infty, +\infty)$	0	A.8
willekeurige fout	σ_e	$[0, +\infty)$	0	A.9
totale fout	σ_t	$[0, +\infty)$	0	A.10
modelefficiëntie	E_{NS}	$(-\infty, 1]$	1	A.13

Tabel 2.4: Modelprestatie-maten behorend bij figuur 2.2 met van links naar rechts de genormaliseerde standaarddeviatie, d.w.z., de verhouding tussen de standaarddeviatie van de gemodelleerde ($\sigma_{\hat{y}}$) en waargenomen (σ_y) tijdreeksen; Pearsons correlatiecoëfficiënt (r); de systematische fout ($\bar{e}$); de willekeurige fout (σ_e); de gemiddelde gekwadrateerde fout (σ_t^2), en de modelefficiëntie (E_{NS}).

symbool	paneel in figuur 2.2	$\tilde{\sigma}_{\hat{y}}$	r	$\bar{e}$	σ_e	σ_t	E_{NS}
a	a) te veel dynamiek	1.90	0.90	0.03	0.65	0.81	-0.17
b	b) te weinig dynamiek	0.47	0.90	0.03	0.20	0.45	0.63
c	c) faseverschuiving	0.95	-0.91	0.03	2.01	1.42	-2.62
d	d) systematische fout	0.95	0.90	-0.97	0.10	1.03	-0.89
e	e) geen dynamiek	0.01	-0.02	0.03	0.56	0.75	-0.00
f	f) beste model	0.95	0.90	0.03	0.10	0.32	0.81
	optimum:	1.00	1.00	0.00	0.00	0.00	1.00

2.3.5 Interpretatie van afzonderlijke modelprestatie-maten

Een modelprestatie-maat levert meestal een enkele waarde op. Om een gevoel te krijgen voor de betekenis van deze waarde is in figuur 2.2 een aantal (fictieve) tijdreeksen gegeven. Op de horizontale as staat de tijd uitgezet en op de verticale as de meetwaarde en een modelvoorspelling (bijvoorbeeld de nitraat-concentratie in een beek). De grijze lijn stelt de werkelijke waarde voor. Deze is onbekend, maar kan op basis van metingen worden geschat. Op een beperkt aantal tijdstippen zijn metingen verricht. Deze zijn weergegeven als rode stippen. Merk op dat de rode stippen niet noodzakelijkerwijs samenvallen met de grijze lijn. Dit komt doordat metingen behept zijn met fouten. Vervolgens is op grond van de metingen getracht om de werkelijkheid te voorspellen met zes modellen. Deze voorspellingen zijn gegeven als blauwe lijnen.

Tabel 2.3 geeft een overzicht van zes veelgebruikte modelprestatie-maten. Deze modelprestatie-maten zijn toegepast op de tijdreeksen in figuur 2.2. De resultaten zijn gegeven in tabel 2.4. In het onderstaande worden deze resultaten eerst voor elke modelprestatie-maat afzonderlijk besproken en vervolgens gecombineerd.

Genormaliseerde standaardafwijking

De genormaliseerde standaardafwijking $\tilde{\sigma}_{\hat{y}}$ geeft de verhouding weer tussen de standaardafwijking van de modelresultaten ten opzichte van die van de metingen (bijlage A.2). Idealiter ligt deze waarde dicht bij één. Als er meer variatie zit in de gemodelleerde reeks dan is deze modelprestatie maat groter dan één. Dat is alleen het geval bij model **a** en kan bijvoorbeeld duiden op het feit dat een procesparameter (zoals een reactiecoëfficiënt) niet goed is ingesteld waardoor te veel dynamiek wordt gesimuleerd.

De genormaliseerde standaardafwijking is kleiner dan één als er te weinig variatie wordt gesimuleerd met het model. Dat is het geval bij model **b** en model **e**. Dergelijke gesimuleerde reeksen duiden op het ontbreken van één of meerdere relevante processen in het model. Dit geldt met name voor model **e** waar de variatie nagenoeg afwezig is.

De genormaliseerde standaardafwijking wijkt ook af van één als er een trend is in de waargenomen reeks, maar niet in de gesimuleerde reeks of omgekeerd.

Modellen **c**, **d**, en **f** genereren tijdreeksen met ongeveer dezelfde variatie als de metingen. De genormaliseerde standaardafwijking is dan ook vrijwel gelijk aan één.

Pearsons correlatiecoëfficiënt

Pearsons correlatiecoëfficiënt (bijlage A.3) geeft aan in hoeverre het temporele patroon in de metingen door het model wordt gereproduceerd. Dat gaat goed voor modellen **a**, **b**, **d**, en **f** in figuur 2.2 (r is bijna 1). Bij model **c** is sprake van een faseverschuiving waardoor r sterk negatief is. Dit kan worden veroorzaakt doordat een proces ontbreekt in het model, een proces ten onrechte wordt gemodelleerd of een proces verkeerd is geparаметeriseerd. Ook kan dit wijzen op een fout (faseverschuiving) in de modelinvoer. Bij model **e** wijkt het gemodelleerde patroon sterk af van het gemeten patroon door het ontbreken van variatie in de gemodelleerde tijdreeks. Dit resulteert in een correlatiecoëfficiënt van nagenoeg nul. Naast Pearsons correlatiecoëfficiënt wordt in dit rapport ook gebruik gemaakt van Kendalls correlatiecoëfficiënt (Kendall, 1938; Siegel and Castellan, 1988). Deze laatste is minder gevoelig voor extreme waarden in de metingen en modelresultaten en kan een waardevolle aanvulling zijn op Pearsons correlatiecoëfficiënt.

Systematische fout

Er is sprake van een systematische fout (bijlage A.4, Eng.: *bias*) als de gemodelleerde tijdreeks de gemeten tijdreeks systematisch onder- danwel overschat. Model **d** in figuur 2.2 geeft hiervan een voorbeeld. De gemodelleerde reeks

is systematisch te hoog ingeschat. Modelprestatie maat $\bar{e}$ is hierdoor negatief. Een systematische fout wordt vaak veroorzaakt doordat een randvoorwaarde van het model systematisch te hoog of te laag wordt ingeschat. Denk bijvoorbeeld aan de kwel flux, de gewasverdamping of de initiële fosforophoping in de bodem. Ook kunnen fouten/simplificaties in de modelopzet leiden tot systematische fouten. Denk bijvoorbeeld aan het negeren van slecht doorlatende lagen in de bodem waardoor in werkelijkheid meer nutriënten op geringere diepte zullen uitspoelen. Bij de overige modellen in figuur 2.2 worden grote fouten gecompenseerd door kleine fouten zodat de systematische fout nagenoeg gelijk is aan nul.

Willekeurige fout

De willekeurige fout (bijlage A.5) is de fout die overblijft als de systematische fout wordt geëlimineerd. De willekeurige fout kan tal van oorzaken hebben, zoals het ontbreken van een proces, oversimplificatie van processen, of een mismatch tussen model- en meetschaal. Ook dragen meetfouten bij aan de willekeurige fout. De willekeurige fout is een maat voor de precisie van het model. Modellen **d** en **f** hebben de kleinste willekeurige fout en zijn het meest precies. Echter, door een systematische fout is model **d** minder nauwkeurig dan model **f**.

Totale fout

De totale fout (bijlage A.6) bestaat uit twee componenten: de systematische fout en de willekeurige fout (zie vergelijking C.1) en kan daarom vele oorzaken hebben. De faseverschuiving in de derde tijdreeks (model **c**) van figuur 2.2 heeft de grootste totale fout. Opvallend is dat model **a** ongeveer even goed presteert als model **e** in termen van de totale fout. Dit benadrukt dat modelprestatie maten altijd in relatie tot elkaar moeten worden beschouwd (paragraaf 2.3.6). Model **a** volgt immers veel beter het patroon in de metingen en zal daarom de voorkeur genieten. De totale fout is een maat voor de nauwkeurigheid van het model.

Modelefficiëntie

De modelefficiëntie (bijlage A.8) is gerelateerd aan de totale fout (zie vergelijking A.14). De modelefficiëntie kan worden gezien als een gestandaardiseerde versie van de totale fout. Het optimum is één. Een waarde van nul geeft aan dat het gemiddelde van de waarnemingen een even goede predictor is als de modelvoorspellingen. Het model biedt dan geen meerwaarde boven het gemiddelde van de metingen. Dit is het geval bij model **e** in figuur 2.2. Negatieve modelefficiënties geven aan dat het gemiddelde van de meetgevens een betere

predictor is dan de modelvoorspellingen. Dit is het geval bij model **a**, model **c** en model **d**. Wordt model **a** vergeleken met model **e**, dan blijkt weer dat de modelprestatie-maten in onderlinge samenhang moeten worden beschouwd (paragraaf 2.3.6). Immers, de modelefficiëntie zegt niets over het patroon in de metingen. Ook kan de modelefficiëntie geen onderscheid maken tussen de systematische en willekeurige fout. Een lage modelefficiëntie kan namelijk het gevolg zijn van een grote systematische fout en/of een grote willekeurige fout. Vergelijk bijvoorbeeld model **c** met model **d** in figuur 2.2. Dit benadrukt wederom het belang om de modelprestatie-maten in onderlinge samenhang te beschouwen.

2.3.6 Interpretatie van modelprestatie-maten in onderlinge samenhang

In het bovenstaande is benadrukt dat modelprestatie-maten niet enkel afzonderlijk maar ook in onderlinge samenhang moeten worden bestudeerd. Dat kan op basis van een tabel zoals tabel 2.4. Echter, een dergelijke tabel wordt al gauw onoverzichtelijk als meerdere modellen moeten worden beoordeeld. Meer overzicht wordt verkregen door de modelprestatie-maten grafisch te presenteren. In bijlagen B en C worden twee diagrammen besproken die hiervoor geschikt zijn, te weten het (uitgebreide) Taylor diagram en de zonneplot.

Taylor diagram

Het Taylordiagram is ontwikkeld om een aantal modelprestatie-maten te presenteren in onderlinge samenhang (Taylor, 2000, 2001), te weten de genormaliseerde standaardafwijking (bijlage A.2), Pearsons correlatiecoëfficiënt (bijlage A.3) en de willekeurige fout (bijlage A.5). In dit rapport is het Taylordiagram uitgebreid met een extra dimensie, zodat ook de systematische fout (bijlage A.8) en de totale fout (bijlage A.10) kunnen worden weergegeven. Voor meer informatie over het Taylordiagram wordt verwezen naar Taylor (2000, 2001) en voor het uitgebreide Taylordiagram naar bijlage B.

Figuur 2.3 geeft het uitgebreide Taylordiagram weer voor de tijdreeksen in figuur 2.2. De rode stip representeert de metingen en de blauwe stippen de modelresultaten. De modelprestatie kan worden afgeleid uit de relatieve ligging van de blauwe punten ten opzichte van het rode punt. De afstand van een blauw punt tot het rode punt representeert de willekeurige fout in de bovenste figuur en de totale fout in de onderste figuur.

De horizontale as in de bovenste figuur geeft de genormaliseerde standaardafwijking $\tilde{\sigma}_{\hat{y}}$ (vergelijking A.6) weer. Hoe verder een punt van de oorsprong ligt, hoe groter de genormaliseerde standaardafwijking. De metingen hebben (per definitie) een genormaliseerde standaardafwijking van één. Modellen **c**, **d** en **f** liggen ongeveer op de halve doorgetrokken cirkel door $\tilde{\sigma}_{\hat{y}} = 1$. Dat

wil zeggen dat deze modellen tijdreeksen simuleren die bij benadering dezelfde variatie hebben als de metingen. Model **a** overschat de variatie in de metingen, en modellen **b** en **e** onderschatten de variatie. Model **e** laat zelfs helemaal geen variatie zien. Hoewel modellen **a** en **e** ongeveer dezelfde willekeurige, systematische en daardoor totale fout hebben, beschrijft model **a** veel beter het patroon in de metingen ($r = 0.9$).

Modellen **d** en **f** blijken het patroon in de metingen goed te beschrijven ($r = 0.9$). Dit geldt niet voor model **c** ($r = -0.9$). Deze ligt aan de linkerkant van de figuur wat impliceert dat de modelresultaten in tegenfase zijn met de metingen. De bovenste figuur suggereert dat modellen **d** en **f** even goed presteren (ze vallen samen). Niets is echter minder waar. Uit de onderste figuur blijkt dat model **d** de metingen systematisch overschat. Model **f** geeft dan ook de beste prestatie.

Zonneplot

Figuur 2.4 presenteert de modelprestatie-maten in een zonneplot (bijlage C). In een zonneplot wordt de (genormaliseerde) willekeurige fout uitgezet tegen de (genormaliseerde) systematische fout. Een model presteert beter als het dichterbij de oorsprong (kern van de zon) ligt. De binnenste halve cirkel (zwarte doortrokken lijn) geeft de meetfout weer. Modellen die binnen deze halve cirkel liggen presteren ongeveer even goed. Verdere aanpassingen aan het model zullen dan niet meer tot betere prestaties leiden.

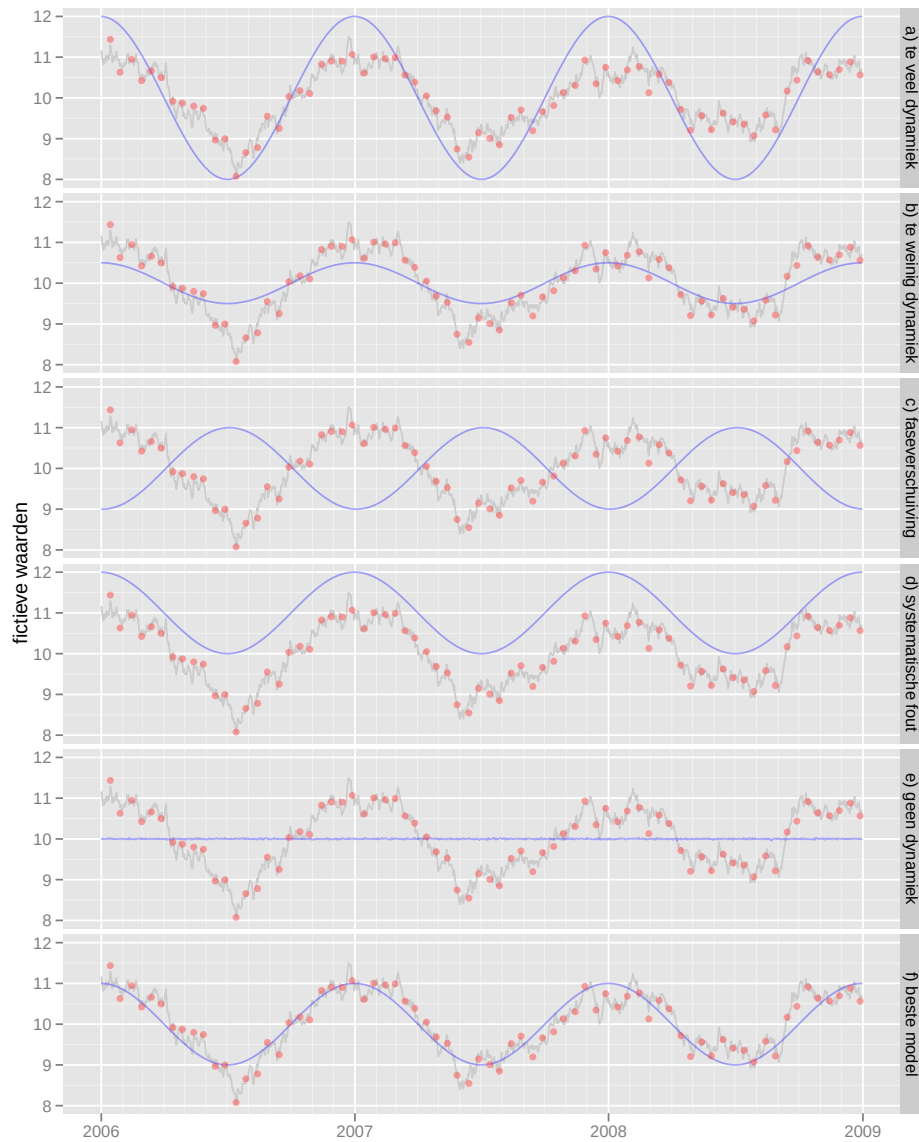
De modelresultaten kunnen worden geïnterpreteerd op basis van hun ligging ten opzichte van de kern van de zon.

Model **d** ligt ver verwijderd van de overige punten links van de zon. Dit impliceert dat model **d** de waarnemingen systematisch overschat. De overige modellen liggen op de verticale as wat impliceert dat geen sprake is van een systematische fout. Dit strookt met de bevindingen op basis van de onderste figuur van het uitgebreide Taylor diagram.

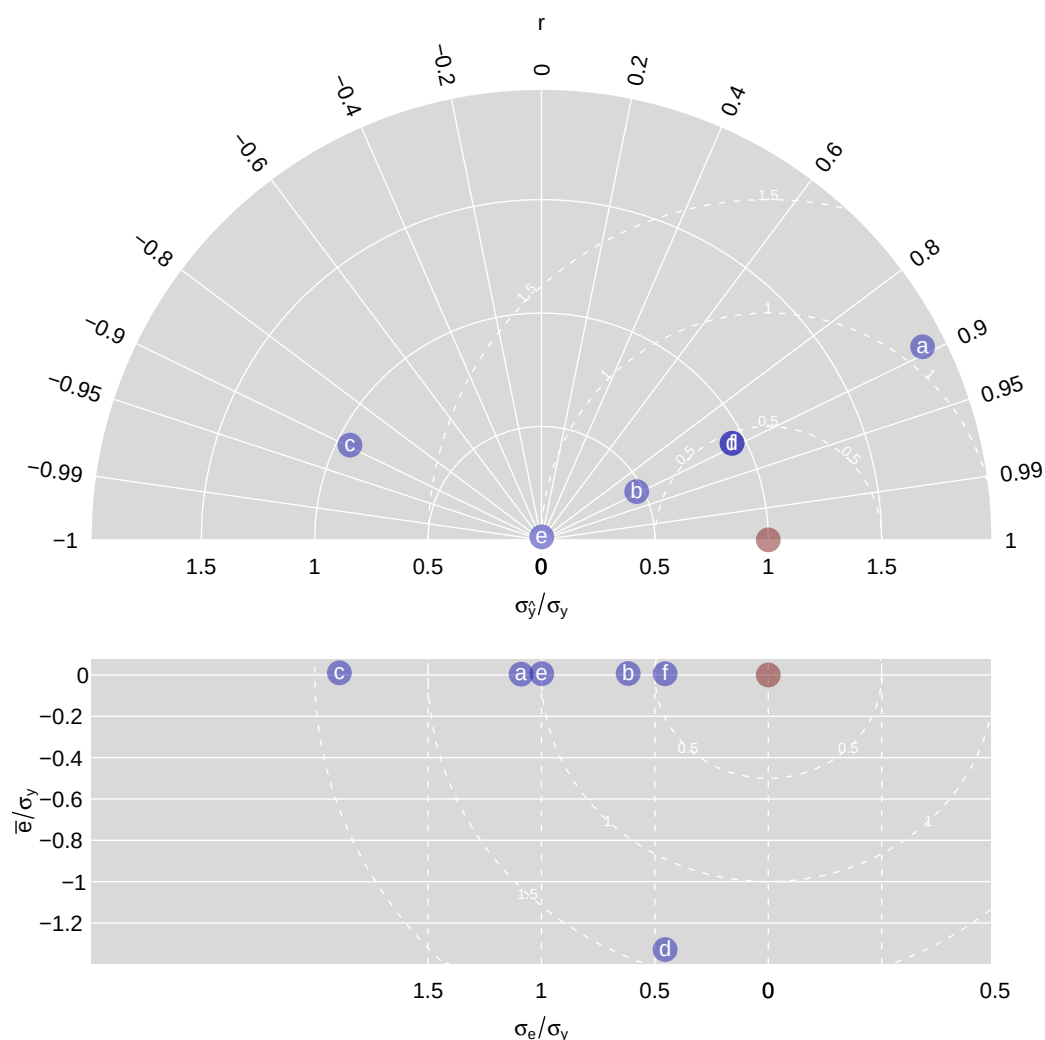
Drie modellen liggen buiten de zon. Dat zijn modellen **a**, **c** en **d**. Modellen die buiten de zon liggen hebben geen meerwaarde boven het gebruik van de gemiddelde meetwaarde als predictor. De oorzaak hiervan voor model **d** was al duidelijk. Hier is sprake van een systematische overschatting van de meetgegevens. Systematische fouten duiden op verkeerde inschatting van randvoorwaarden, of onjuiste parameterisatie van de processen en is door de modelleur te corrigeren (bijvoorbeeld door middel van kalibratie). De slechte prestatie van modellen **a** en **c** zijn terug te voeren op een grote willekeurige fout. Die kan worden veroorzaakt door bijvoorbeeld een systematische overschatting van de variatie in de metingen (model **a**) of doordat het patroon in de metingen niet goed wordt gesimuleerd (model **c**). Op basis van het zonneplot alleen is nadere specificering van deze fout niet mogelijk. Het (uit-

gebreide) Taylor diagram kan dan uitsluitsel geven.

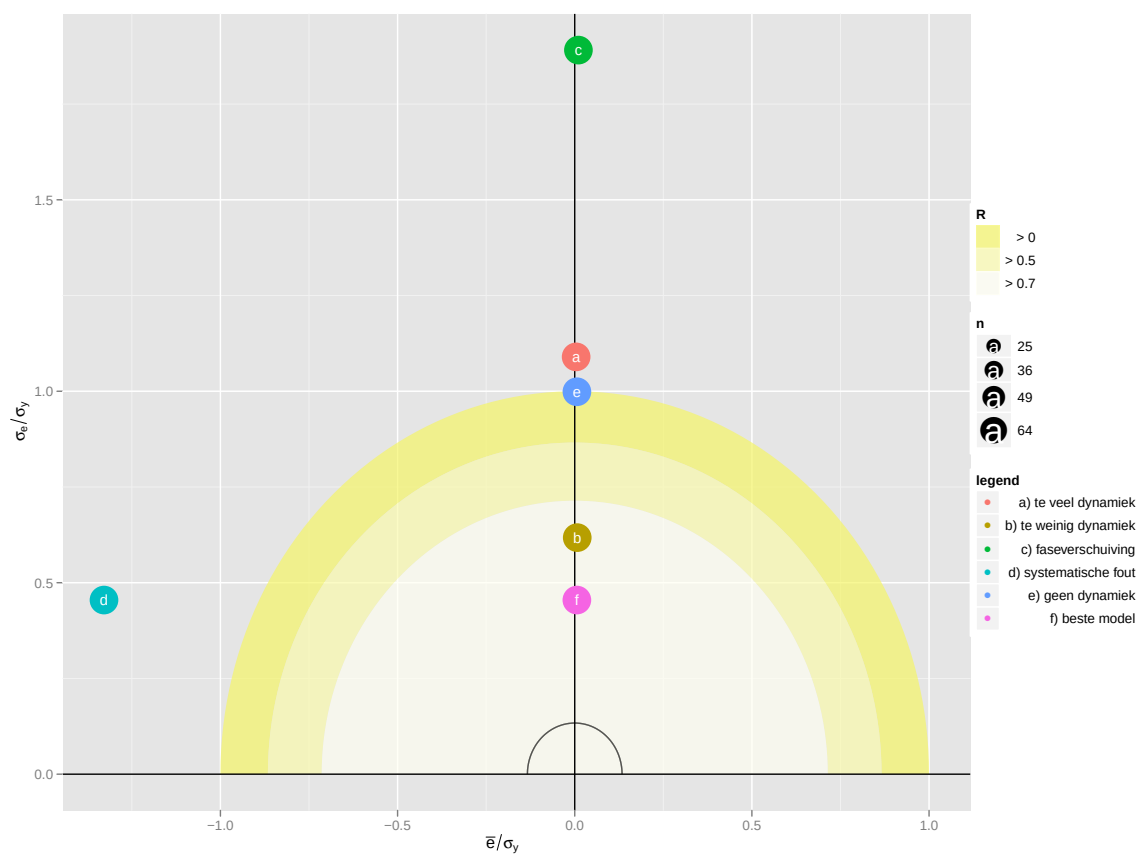
Modelresultaten die binnen de zon vallen hebben een meerwaarde boven het gebruik van de gemiddelde meetwaarde als predictor. Modellen **b** en **f** hebben dus meerwaarde, waarbij model **f** beter presteert (nauwkeuriger is) dan model **b**. De reden waarom model **f** beter presteert dan model **b** kan uit het uitgebreide Taylordiagram worden afgeleid: model **f** is beter in staat om de variatie in de meetgegevens te reproduceren.



Figuur 2.2: Synthetische voorbeelden van modelresultaten (blauwe lijn) en de te modelleren werkelijkheid (grijze lijn) waarvan de waarde (inclusief meetfout) slechts op de meetmomenten (rode stippen) bekend is. De modellen worden aangeduid met de letters a tot en met f.



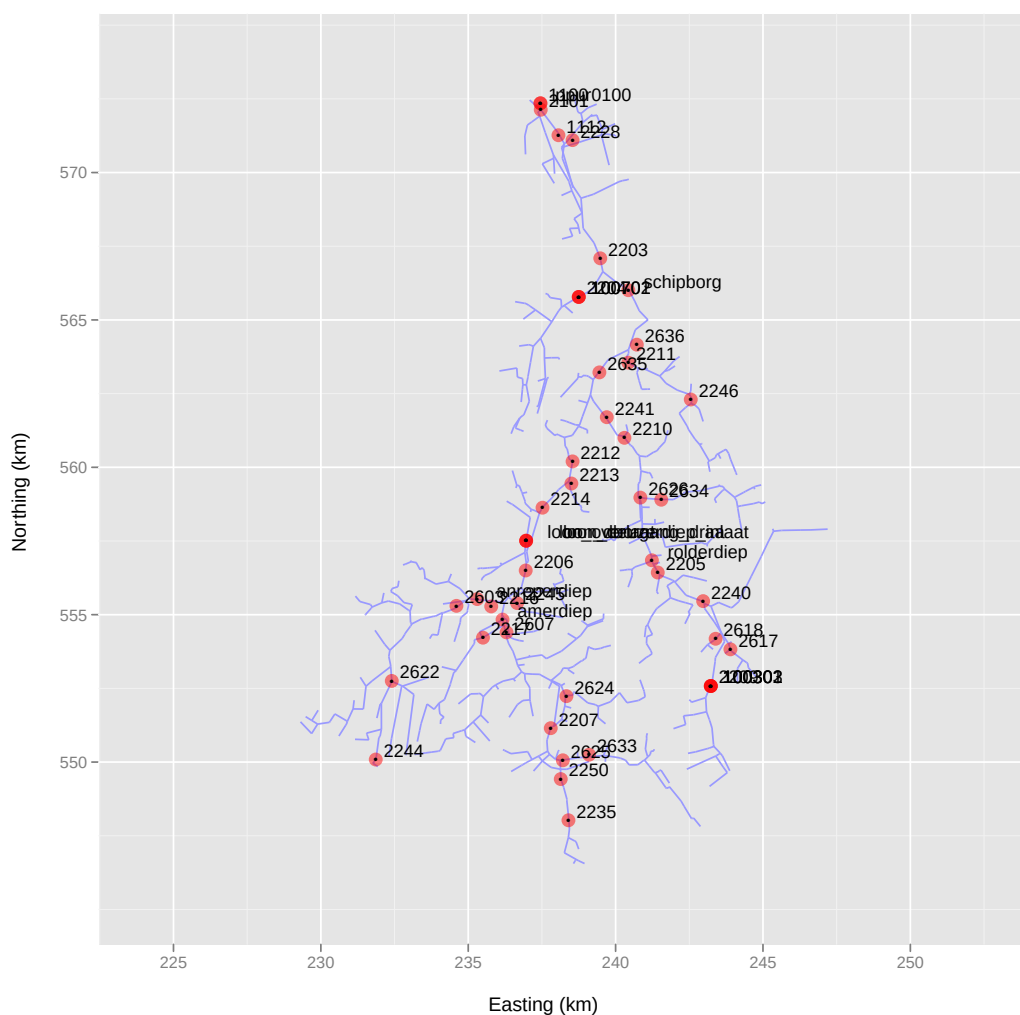
Figuur 2.3: Taylor diagram (boven) gekoppeld aan $\bar{e} - \sigma_e$ plot (onder). De letters a tot en met f verwijzen naar de tijdreeksen in figuur 2.2. Merk op dat d en f samenvallen in de bovenste helft van de figuur. Voor de wijze waarop deze figuur is geconstrueerd en moet worden geïnterpreteerd wordt verwezen naar bijlage B.



Figuur 2.4: Voorbeeld van een zonneplot. Op de horizontale as is de genormaliseerde systematische fout uitgezet, op de verticale as de genormaliseerde willekeurige fout. De letters a tot en met f verwijzen naar de tijdreeksen in figuur 2.2. Voor de wijze waarop deze figuur moet worden geïnterpreteerd wordt verwezen naar bijlage C.

3 Resultaten en discussie

In dit hoofdstuk wordt de modelprestatie t.a.v. het voorspellen van de waterafvoer en de stikstof- en fosforconcentraties op een aantal meetlocaties besproken. De meetlocaties zijn gegeven in figuur 3.1.



Figuur 3.1: Ligging van de meetlocaties in het stroomgebied van ‘de Drentse Aa’.

In onderstaande paragrafen wordt eerst nagegaan hoe goed waargenomen temporele patronen door de modellen kunnen worden gereproduceerd en voorspeld. Daarna wordt nagegaan in hoeverre de modelsystemen in staat zijn om abso-

lute meetwaarden te reproduceren en te voorspellen.

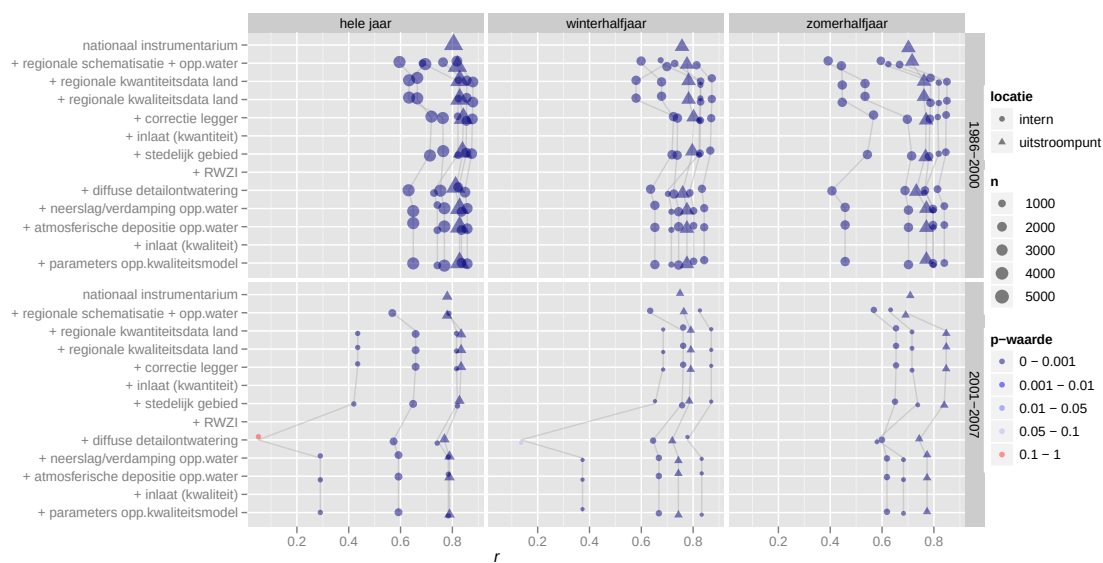
3.1 Waterafvoer

Voordat de validatieresultaten die betrekking hebben op de waterafvoer worden besproken moet wederom worden benadrukt dat de modelprestatiematen niet alleen onvolkomenheden van de modelsystemen bevatten, maar ook die van de metingen. Dat is met name relevant voor de afvoermetingen van de Drentse Aa, omdat ten tijde van de analyse van de validatieresultaten nog onderzoek gaande was bij waterschap Waterschap Hunze en Aa's om de afvoergegevens nauwkeuriger in te schatten op basis van debietformules.

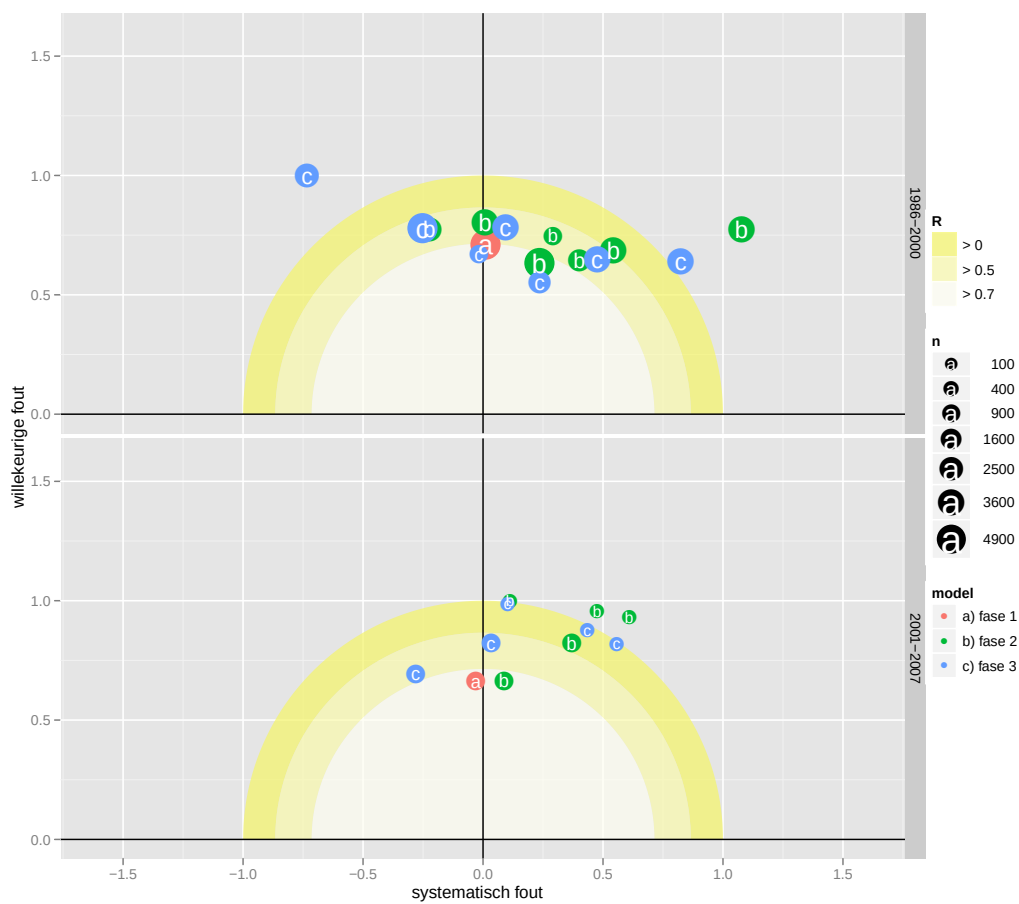
In figuur 3.2 is Pearsons correlatiecoëfficiënt (r) weergegeven (horizontale as) voor de opeenvolgende deelfases (verticale as). Pearsons correlatiecoëfficiënt geeft aan hoe goed de modelsystemen in staat zijn om temporele patronen te reproduceren. De bovenste drie panelen slaan op de periode 1986-2000, de onderste drie panelen op de periode 2001-2007 (validatieperiode). Tevens is een onderscheid gemaakt naar seizoenen (kolommen). De kleuren geven de mate van significantie weer. Blauwe symbolen geven aan dat er een significant verband bestaat tussen de gemeten en berekende afvoeren, rode symbolen geven aan dat op basis van de beschikbare metingen geen overeenkomst in patroon kon worden aangetoond. Hoe groter r , hoe beter het gesimuleerde patroon overeenkomt met het waargenomen patroon.

Uit de figuur blijkt dat op vrijwel alle meetlocaties een significant verband bestaat tussen de metingen en de modelvoorspellingen. De correlatiecoëfficiënt (r) voor de periode 1986-2000 is over het algemeen iets hoger dan voor de periode 2001-2007. Verder blijkt dat het patroon in het zomerhalfjaar vrijwel even goed wordt gereproduceerd en voorspeld als in het winterhalfjaar. Het grootste deel van de meetlocaties heeft een Pearsons correlatiecoëfficiënt (r) tussen de 0.6 en 0.9. Dat wil zeggen dat 36% tot 81% van de variatie in de metingen door het model kan worden verklaard. Verder valt op dat met name de deelfase 'regionale kwantiteitsdata land' tot een geringe verbetering in de modelprestatie leidt. Deze deelfase omvat het toevoegen van regionale meteorologische gegevens, aanpassing van de onderrand (kwel/wegzijging), drainagekarakteristieken volgens van der Gaast et al. (2006) en aanpassing van de pakketdikte van het topsysteem (Roelsma et al., 2008, pag. 31-34). De toevoeging van de kleinste waterlopen (diffuse detailontwatering) heeft een geringe terugslag tot gevolg, waarna door toevoeging van neerslag en verdamping op het oppervlaktewater het resultaat weer enigszins verbetert.

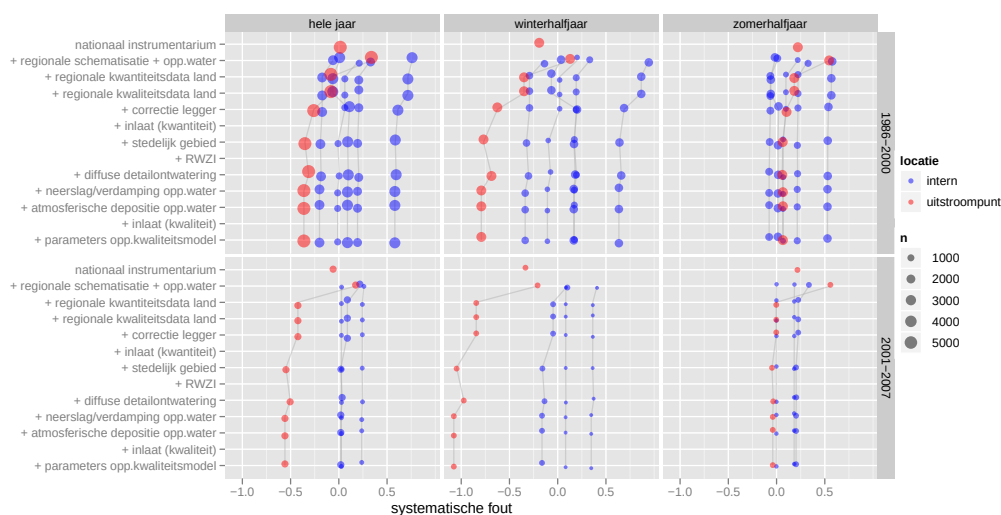
In figuur 3.3 is de modelprestatie gevisualiseerd met behulp van een zonneplot (paragraaf 2.3.6, bijlage C). Uit de zonneplot blijkt dat de meeste locaties binnen de zon liggen. Voor deze locaties heeft het model meerwaarde boven het gemiddelde van de metingen.



Figuur 3.2: Pearsons correlatiecoëfficiënt (r) tussen de gemeten en gesimuleerde waterafvoer ($\text{m}^3 \text{s}^{-1}$) op alle locaties in het stroomgebied van 'de Drentse Aa' voor het hele jaar (kolom 1), het winterhalfjaar (kolom 2) het zomerhalfjaar (kolom 3) en twee perioden (rijen). De significantie van r is gegeven als p -waarde: $0-0.001$ = uiterst significant, $0.001-0.01$ = zeer significant, $0.01 - 0.05$ = significant, $0.05 - 0.1$ = zwak significant, $0.1 - 1$ = niet significant.



Figuur 3.3: Afvoer ($\text{m}^3 \text{s}^{-1}$) op dagbasis voor de periode 1986-2000 (boven) en 2001-2007 (onder).

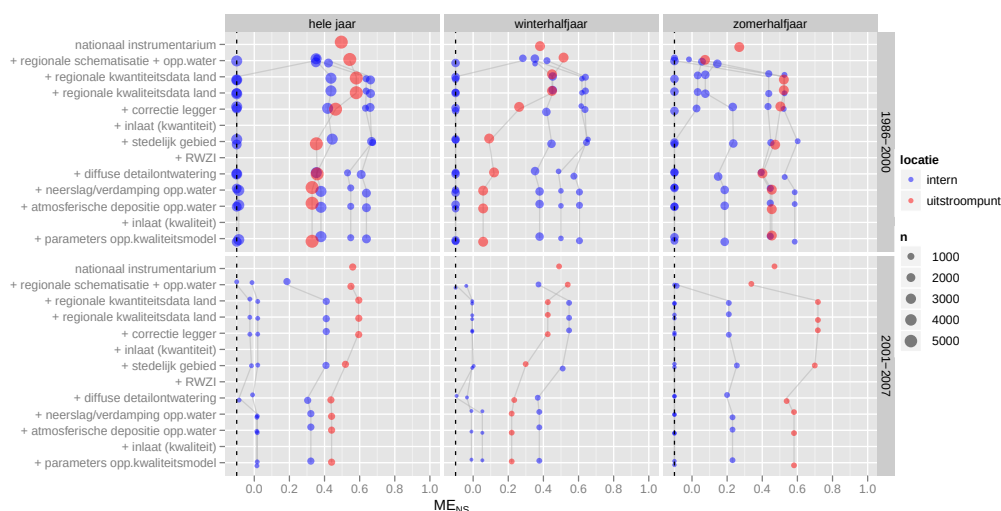


Figuur 3.4: Systematische fout ($\text{m}^3 \text{s}^{-1}$) m.b.t. de waterafvoer voor de afzonderlijke deelfases.

Uit de figuur blijkt dat fase 1 voor zowel de periode 1986-2000 als de validatieperiode 2001-2007 vrijwel geen systematische fout heeft. Het nationaal modelinstrumentarium lijkt gemiddeld gesproken goede resultaten te geven. Daarbij dient echter wel een kanttekening te worden geplaatst. Het nationaal model doet namelijk voorspellingen voor het *hele* stroomgebied van de Drentse Aa terwijl de metingen waarop gevalideerd wordt betrekking hebben op meetstation Schipborg, dat op circa 15 km stroomopwaarts van het echte uitstroompunt ligt (zie figuur 3.1). Het nationaal modelinstrumentarium zou dus een *overschatting* moeten laten zien als de modelresultaten worden vergeleken met de metingen bij Schipborg.

De resultaten van het ‘uitstroompunt’ van fase 2 en 3 hebben wel betrekking op Schipborg. Fase 2 geeft een systematische onderschatting te zien en fase 3 een systematische overschatting van de afvoer bij Schipborg (figuur 3.4). De systematische overschatting in fase 3 wordt veroorzaakt door deelfase ‘regionale kwantiteitsdata land’ waarbij de hydrologie van het landsysteem is geregionaliseerd (Roelsma et al., 2008, pag. 31–34). De systematische overschatting wordt versterkt door deelfases ‘correctie legger’ en ‘stedelijk gebied’, en in mindere mate, door deelfase ‘neerslag en verdamping op het oppervlaktewater’ (figuur 3.4). Dit geldt met name voor het winterhalfjaar. In het zomerhalfjaar is de systematische fout zeer gering. Dat geldt voor zowel de validatieperiode als de periode 1986-2000 (figuur 3.4). Daarbij dient opgemerkt te worden dat de zomerafvoeren ook gering zijn.

In fase 2 en fase 3 wordt niet alleen naar het uitstroompunt gekeken, maar ook naar meer bovenstroomsgelegen punten. Uit figuur 3.4 blijkt dat zowel locaties voorkomen die metingen systematisch overschatten als onderschatten. Ruimtelijk gezien kunnen twee clusters worden onderscheiden: ‘Anreperdiep’



Figuur 3.5: Modelefficiëntie m.b.t. de waterafvoer voor de afzonderlijke deelfases. Terwille van de leesbaarheid zijn waarden kleiner dan -0.1 op de stippellijn geplaatst.

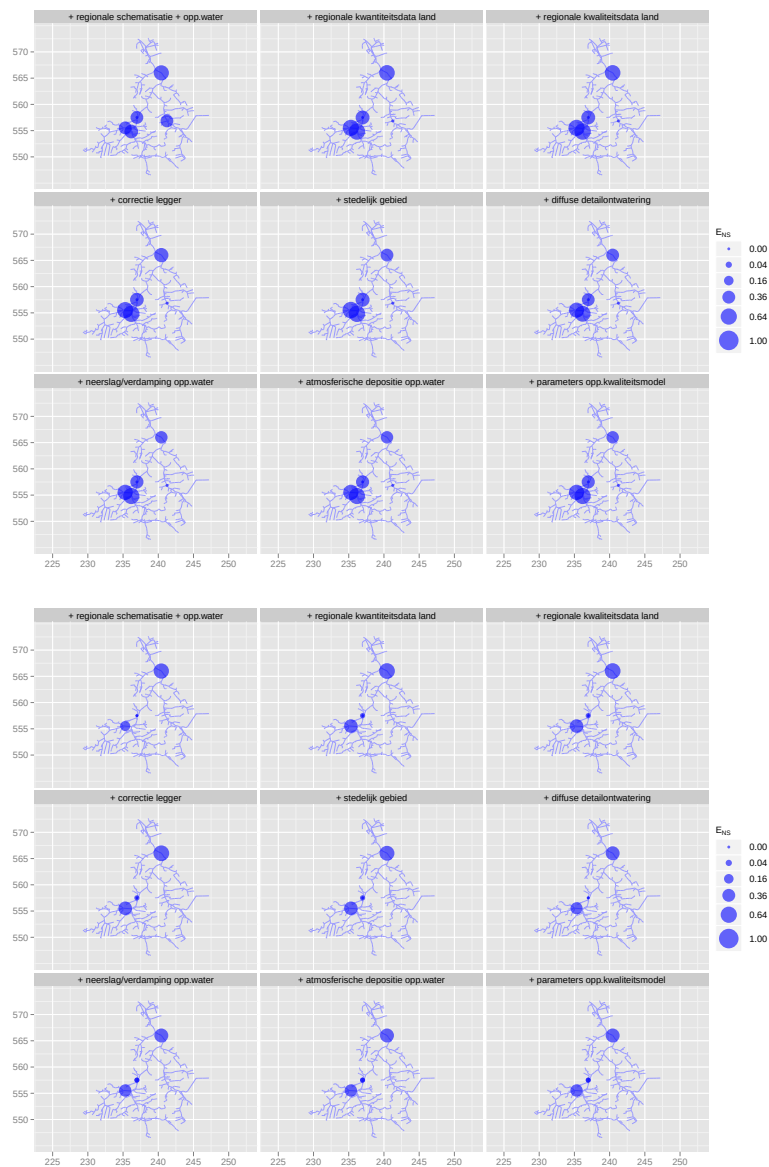
en ‘Amerdiep’ in het westen van de Drentse Aa laten een systematische onderschatting zien, ‘Rolderdiep’ en ‘Schipborg’ in het oosten van de Drentse Aa een systematische overschatting.

De afstand van een punt tot de oorsprong van de zonneplot is gelijk aan de modelefficiëntie (vergelijking A.13). Deze is gegeven als functie van de deelfase, de periode, het locatietype, het aantal waarnemingen en het seizoen in de bovenste rij van figuur 3.5. Uit de figuur blijkt dat de modelprestatie op het uitstroompunt van fase 1 naar fase 2 verbetert. De verbetering zet door tot en met verfijningsstap ‘regionale kwaliteitsdata land’. Echter, daarna leidt een aantal verfijningsstappen van fase 3 tot een verslechtering. Voor de periode 1986-2000 zijn dat deelfasen ‘correctie legger’ en ‘toevoeging stedelijk gebied’. Zoals in een vorige paragraaf al naar voren kwam wordt dit deels veroorzaakt doordat de systematische fout toeneemt (grotere systematische overschatting van de gemeten afvoeren). Voor de validatieperiode 2001-2007 echter heeft deelfase ‘correctie legger’ nauwelijks effect. Dit komt waarschijnlijk doordat de aangepaste leggergegevens betrekking hebben op de huidige situatie, en niet op de periode tot 2000. Een oplossing is om bij het verbeteren van de dimensionering van het oppervlaktewatermodel gebruik te maken van zowel historische als actuele leggergegevens en informatie over het beheer. Dit compliceert het modelleringsproces, maar levert mogelijk betere resultaten op.

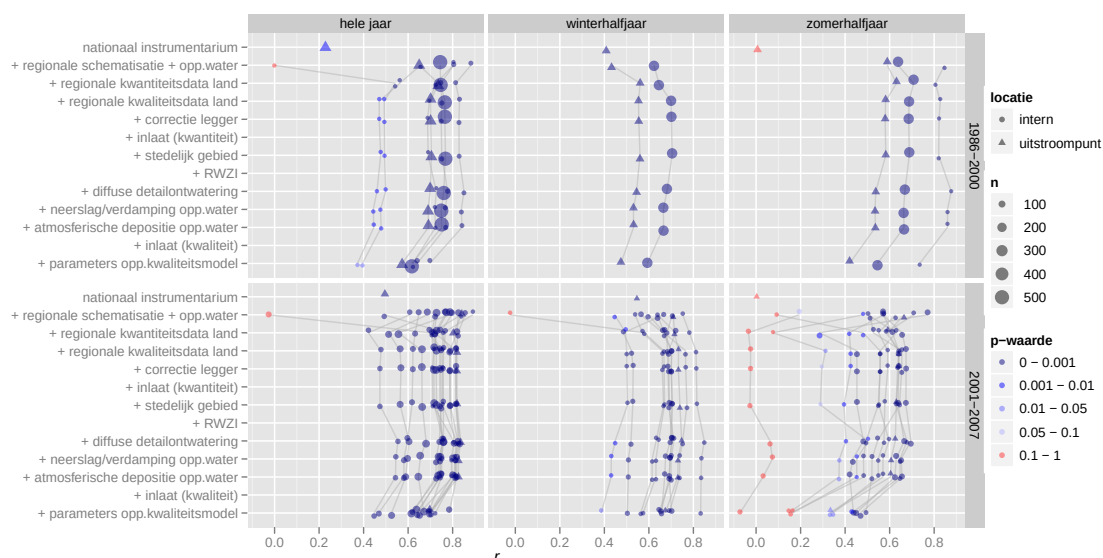
Ook in de validatieperiode (2001-2007) leidt deelfase ‘stedelijk gebied’ tot een verslechtering van de modelprestatie. Mogelijk is de aanname dat 60% van de oppervlakte van het stedelijk gebied onverhard is aan de hoge kant (Roelsma et al., 2008, p.36). Een verdere verslechtering volgt na toevoeging van de kleinste waterlopen in deelfase ‘diffuse detailontwatering’. Deze stap heeft ook

effect op de bovenstroomse meetlocaties. Zoals Roelsma et al. (2008, p.89) al opmerken wordt hierdoor de dynamiek van de waterafvoer mogelijk te veel gedempt waardoor pieken mogelijk minder goed worden gesimuleerd.

De ruimtelijke verdeling van de modefficiëntie is gegeven in figuur 3.6. Hieruit blijkt dat met name de modellering van het Rolderdiep niet goed gaat. Dit wordt veroorzaakt door de eerste deelfase van fase 3 ‘waterkwantiteitsdata land’, waarbij de hydrologie van het landsysteem is geregionaliseerd. Merk op dat deze deelfase wel tot een verbetering leidt bij de overige gebiedsdelen.



Figuur 3.6: Modelefficiëntie (-) m.b.t. de waterafvoer ($\text{m}^3 \text{s}^{-1}$) op dagbasis voor de periode 1986-2000 (boven) en 2001-2007 (onder) voor opeenvolgende (deel)fases.

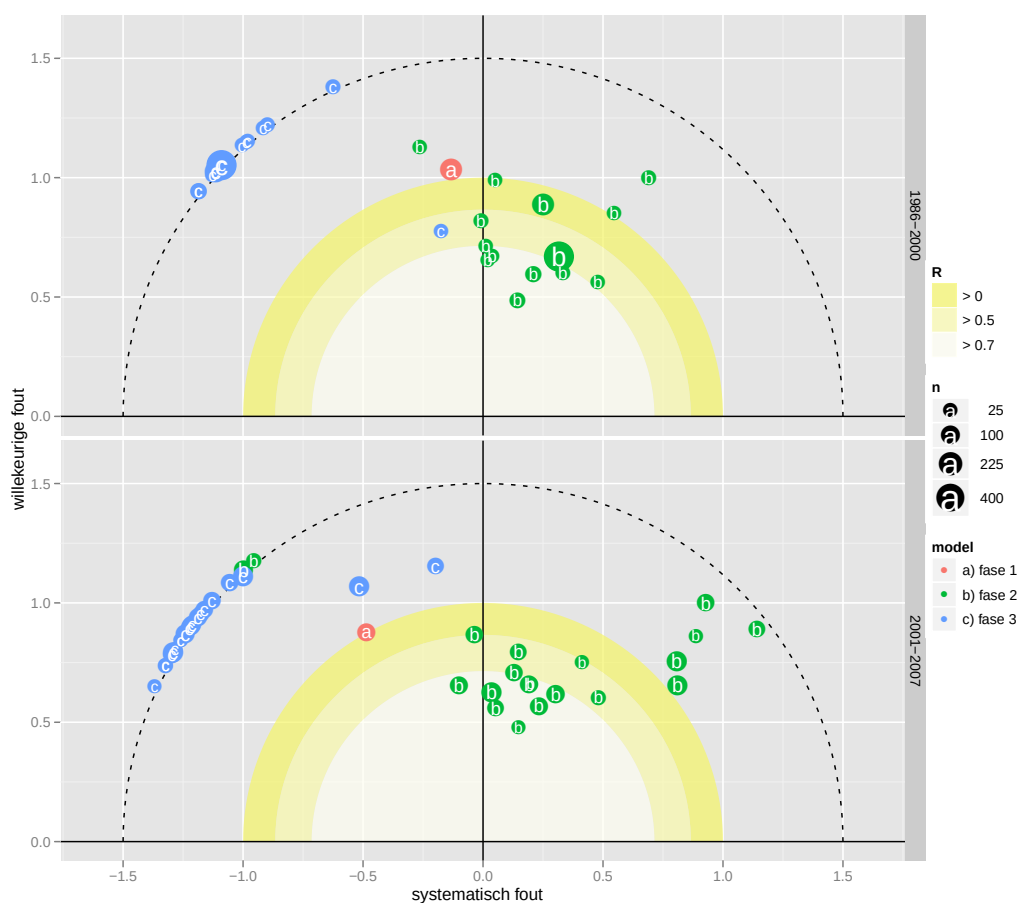


Figuur 3.7: Pearsons correlatiecoëfficiënt (r) tussen de gemeten en gesimuleerde anorganische stikstofconcentratie (mg l^{-1}) op alle locaties in het stroomgebied van 'de Drentse Aa' voor het hele jaar (kolom 1), het winterhalfjaar (kolom 2) het zomerhalfjaar (kolom 3) en twee perioden (rijen). De significantie van r is gegeven als p -waarde: $0-0.001$ = uiterst significant, $0.001-0.01$ = zeer significant, $0.01 - 0.05$ = significant, $0.05 - 0.1$ = zwak significant, $0.1 - 1$ = niet significant.

3.2 Stikstof

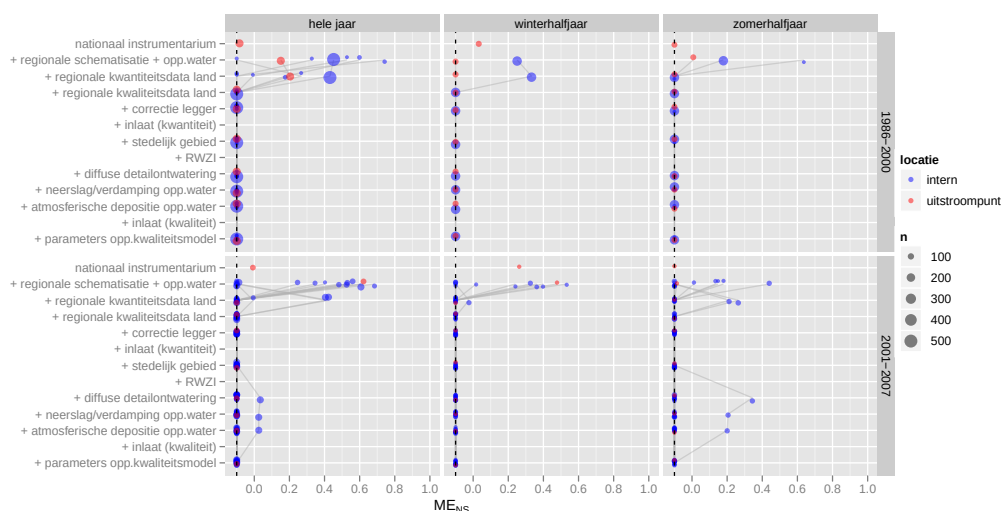
3.2.1 Anorganisch stikstof

Hoe goed de modellen in staat zijn om het temporele patroon van de gemeten anorganische stikstofconcentraties te reproduceren en te voorspellen, blijkt uit figuur 3.7. Fase 2 laat een forse verbetering zien t.o.v. fase 1 wat betreft het reproduceren en voorspellen van het temporele patroon. De eerste deelfase van fase 3 (regionalisering van de hydrologie) laat vervolgens een verslechtering zien waarna de modelprestatie min of meer op hetzelfde niveau blijft. Alleen de laatste deelfase van fase 3 laat vervolgens een verslechtering zien voor vrijwel alle meetlocaties. In deze deelfase is de parameterisering van het oppervlaktewatermodel geregionaliseerd (Roelsma et al., 2008, p.38). Opvallend is verder dat de verschillen in modelprestatie tussen de meetlocaties groter is in de zomer dan in de winter. Dit komt doordat in de zomer ook de afvoeren meer variabel zijn.



Figuur 3.8: Anorganisch stikstof (mg l^{-1}) op dagbasis voor de periode 1986-2000 (boven) en 2001-2007 (onder).

Figuur 3.8 geeft de zonneplot voor de hoofdfases met betrekking tot anorganisch stikstof. Deze figuur geeft aan hoe goed de modellen die voor de hoofdfases zijn opgesteld de metingen reproduceren en voorspellen. Uit de figuur blijkt dat alleen het fase 2 model hier voor een groot aantal meetlocaties goed in slaagt. Een groot aantal meetlocaties van deze fase ligt binnen de zon, en een aanzienlijk deel daarvan zelfs binnen het lichtgele vlak. Dat geldt met name voor de validatieperiode (2001-2007). Het fase 3 systeem, en in mindere mate het fase 1 systeem, overschatten de metingen. Figuur 3.8 geeft slechts de resultaten van de laatste deelfase van fase 3. Op basis van figuur 3.9 kan worden nagegaan welke verfijningsstappen (deelfases) debet zijn aan de achteruitgang in modelprestatie tussen fase 2 en de laatste deelfase van fase 3. Dat blijkt de tweede deelfase van fase 3 te zijn, te weten ‘regionale kwaliteitsdata land’. In deze deelfase is de bodemchemische parameterisering van het landstelsel aangepast op basis van de meest recente inzichten met betrekking tot nutriëntenmodellering op nationale schaal. In dit kader is ondermeer het ‘watergetal’ (Groenendijk et al., 2005, p.86) verhoogd van 0.5 naar 0.8. Dit is de

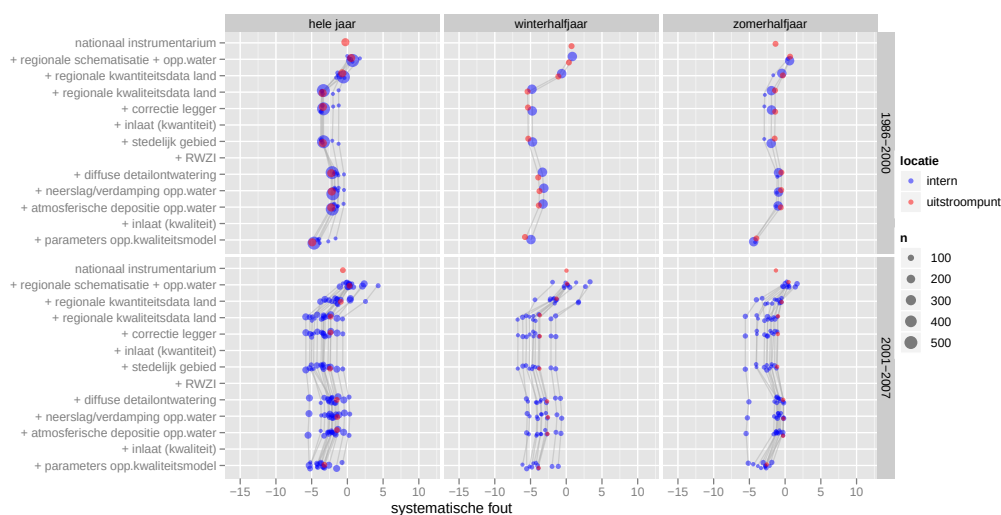


Figuur 3.9: Modelefficiëntie (-) m.b.t. anorganisch stikstof (mg l^{-1}) op dagbasis voor de afzonderlijke deelfases.

kritische waarde van de vochtverzadigingsgraad waarboven denitrificatie¹ optreedt. Verhoging leidt tot minder denitrificatie, waardoor meer anorganisch stikstof beschikbaar komt voor uitspoeling naar grond- en oppervlaktewater en de systematische fout toeneemt. Dat laatste blijkt uit figuur 3.10. Deze aanpassing overschaduwde mogelijk het effect van latere deelfases. Alleen de toevoeging van de kleinste waterlopen in deelfase 'diffuse detailontwatering' heeft tot gevolg dat de systematische fout iets wordt gereduceerd door meer retentie van anorganisch stikstof in het oppervlaktewatersysteem.

Wordt naar het ruimtelijke patroon van de modelefficiëntie gekeken (figuur 3.11) dan valt op dat voor de periode 1986-2000 de modelprestatie over het algemeen fors afneemt ten gevolge van deelfase 'regionale kwaliteitsdata land', behalve bij meetlocatie 2211. Dit is het uitstroompunt van het 'Anloërdiepje'. Deze meetlocatie wordt gekenmerkt door een toename van de modelprestatie, die enkel bij de laatste deelfase van fase 3 weer afneemt. Uit figuur 3.10 blijkt dat de toename van de modelprestatie vooral te danken is aan een afname van de systematische fout. Het bovenstroomse gebied van meetlocatie 2211 wordt gedomineerd door natuur. Echter, Roelsma et al. (2006b) hebben hier relatief hoge stikstofconcentraties waargenomen die niet door het model konden worden verklaard. Door aanpassing van het 'watergetal' in fase 3 neemt in het hele stroomgebied van de Drentse Aa de stikstofuitspoeling fors toe waardoor de systematische fout op locatie 2211 wordt verkleind en de modelprestatie toeneemt. Met andere woorden, hier is sprake van een 'compensatieeffect'. Het gebrek aan informatie in het fase 2 model waardoor de stikstofconcentratie op locatie 2211 te laag werd gemodelleerd, wordt door de fase 3 modellen

¹Denitrificatie is het proces waarbij bacteriën nitraat (NO_3 , een vorm van anorganisch stikstof) omzetten in stikstofgas (N_2). Circa 78% van de lucht bestaat uit stikstofgas.



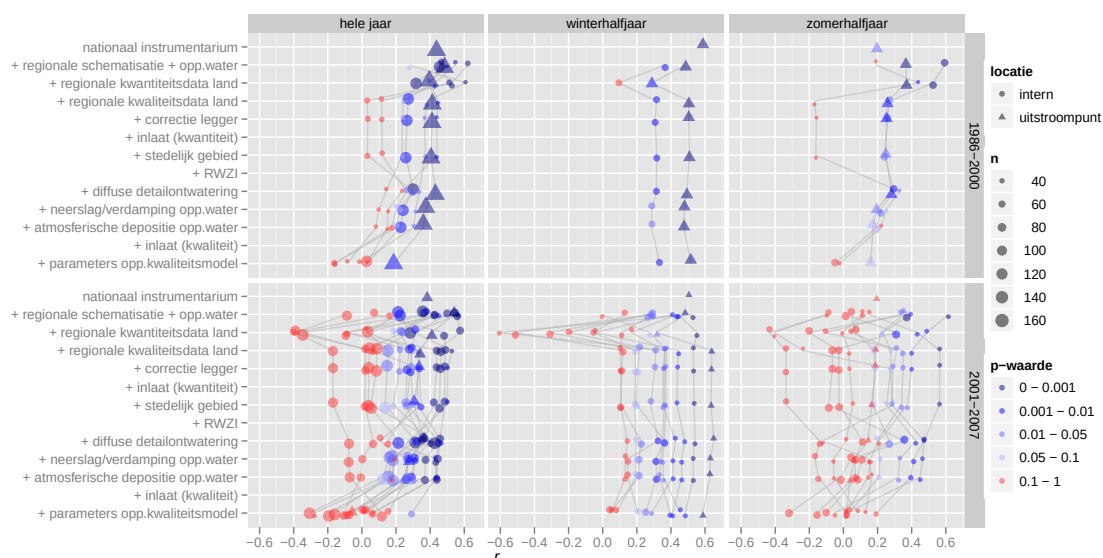
Figuur 3.10: Systematische fout (mg l^{-1}) m.b.t. anorganisch stikstof op dagbasis voor de afzonderlijke deelfases.

gecompenseerd door een te hoge stikstofuitspoeling ten gevolge van een te lage denitrificatie.

Een dergelijk effect treedt ook op in de periode 2001-2007 (figuur 3.10). Echter, de modefficiëntie wordt niet beter (figuur 3.9 en 3.11) omdat de willekeurige fout groot is.



Figuur 3.11: Modelefficiëntie (-) m.b.t. de anorganisch stikstof (mg l^{-1}) op dagbasis voor de periode 1986-2000 (boven) en 2001-2007 (onder) voor opeenvolgende (deel)fases.

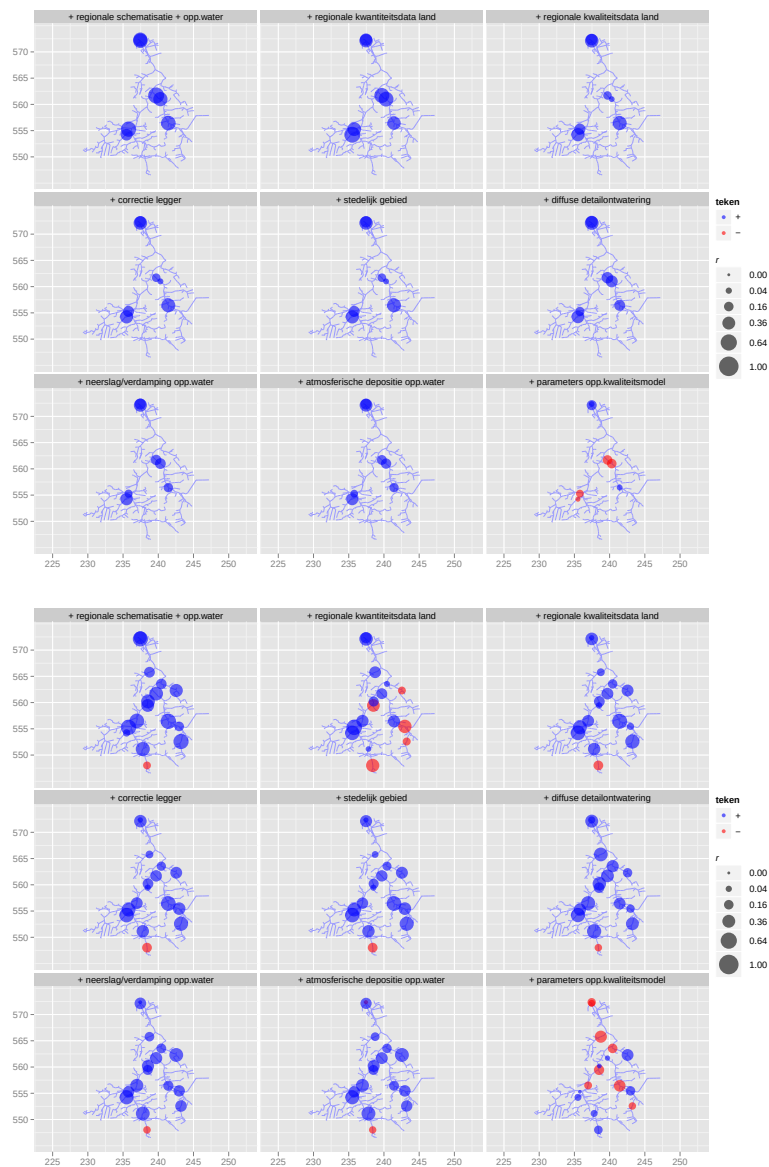


Figuur 3.12: Pearsons (r) correlatiecoëfficiënt tussen de gemeten en gesimuleerde organische stikstofconcentratie (mg l^{-1}) op alle locaties in het stroomgebied van ‘de Drentse Aa’ voor het hele jaar (kolom 1), het winterhalfjaar (kolom 2) het zomerhalfjaar (kolom 3) en twee perioden (rijen). De significantie van r is gegeven als p -waarde: $0-0.001$ = uiterst significant, $0.001-0.01$ = zeer significant, $0.01 - 0.05$ = significant, $0.05 - 0.1$ = zwak significant, $0.1 - 1$ = niet significant.

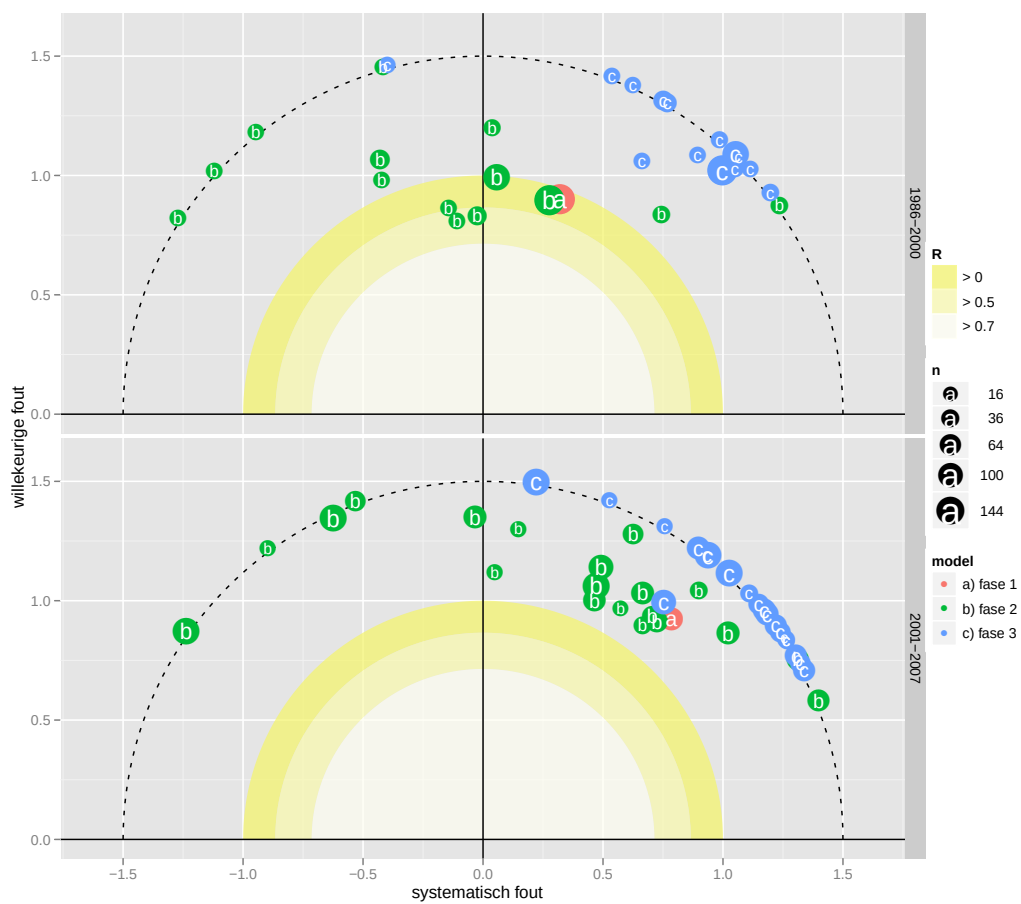
3.2.2 Organisch stikstof

De modellen zijn minder goed in staat om het temporele patroon in *organisch* stikstof te reproduceren en te voorspellen (figuur 3.12). Op een groot aantal locaties kan geen significant verband worden gevonden tussen de modelresultaten en de metingen. Significante verbanden zijn over het algemeen zwak. Opvallend is dat deelfase ‘regionale kwantiteitsdata land’, waarbij de hydrologie van het landsysteem is geregionaliseerd, negatieve correlaties laat zien. Dit duidt er op dat de modelresultaten enigszins in tegenfase zijn met de metingen (paragraaf 2.3.5). Wordt gekeken naar het ruimtelijke beeld, dan blijken de negatieve correlaties met name op te treden in het zuiden en het midden van het stroomgebied en bij de laatste deelfase van fase 3 (figuur 3.13).

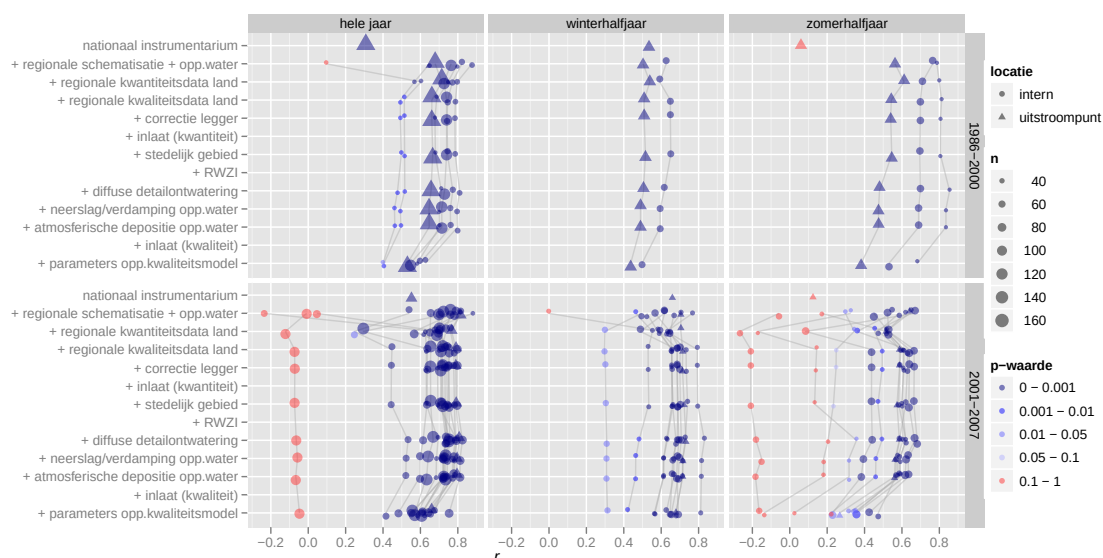
Als het gaat om absolute concentraties, dan blijkt uit figuur 3.14 dat de modelsystemen voor de meeste locaties geen meerwaarde hebben boven het gemiddelde van de metingen. Alleen voor de periode 1986-2000 ligt een aantal locaties binnen de zon. De fase 3 modelsystemen onderschatten overwegend het organisch stikstofgehalte.



Figuur 3.13: Pearsons correlatiecoëfficiënt m.b.t. organisch stikstof (mg l^{-1}) op dagbasis voor de periode 1986-2000 (boven) en 2001-2007 (onder) voor opeenvolgende deelfases.



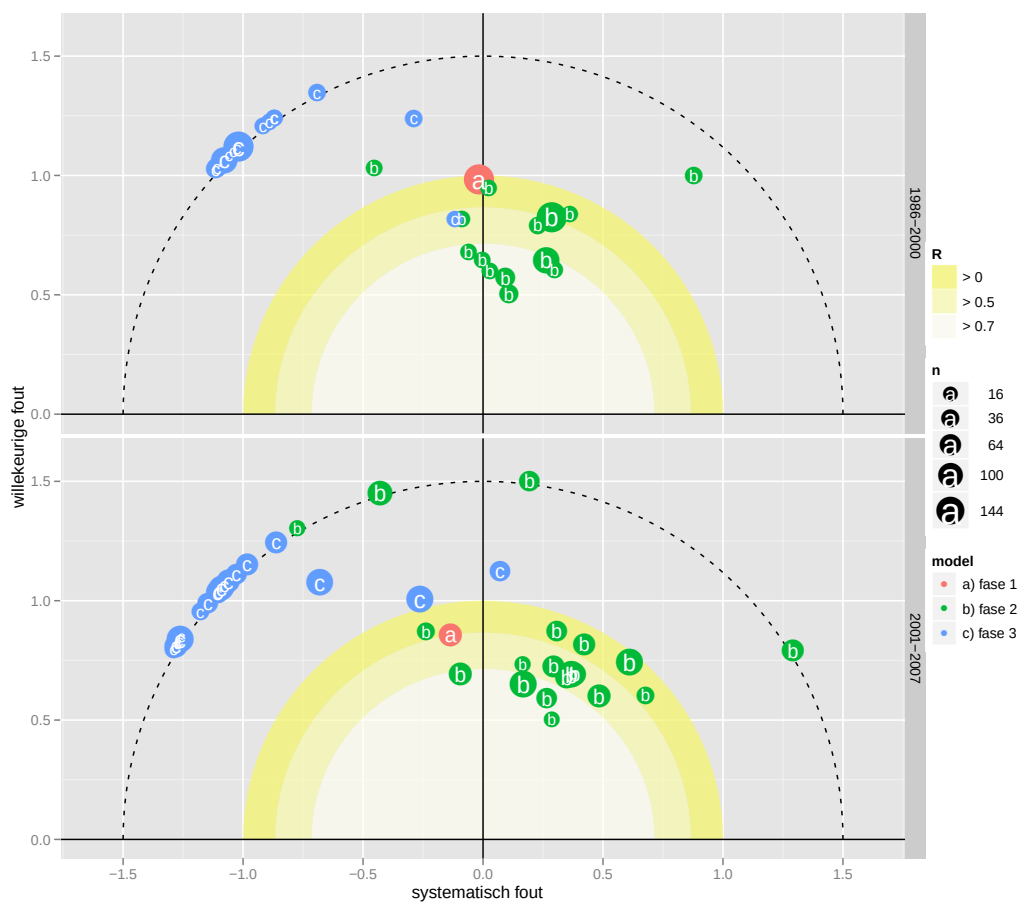
Figuur 3.14: Organisch stikstof (mg l^{-1}) op dagbasis voor de periode 1986-2000 (boven) en 2001-2007 (onder).



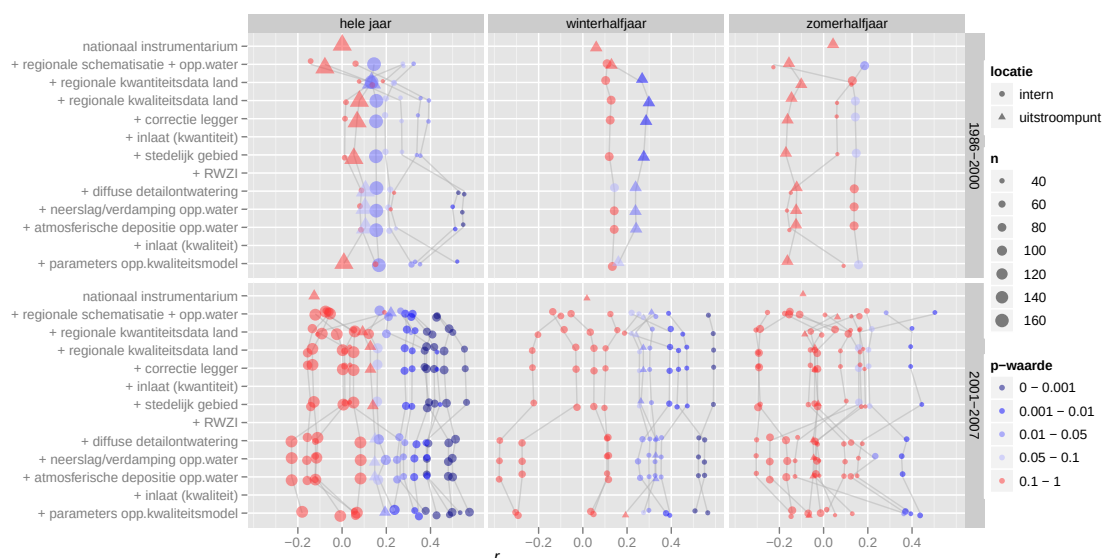
Figuur 3.15: Pearsons (r) correlatiecoëfficiënt tussen de gemeten en gesimuleerde totale stikstofconcentratie (mg l^{-1}) op alle locaties in het stroomgebied van 'de Drentse Aa' voor het hele jaar (kolom 1), het winterhalfjaar (kolom 2) het zomerhalfjaar (kolom 3) en twee perioden (rijen). De significantie van r is gegeven als p -waarde: $0-0.001$ = uiterst significant, $0.001-0.01$ = zeer significant, $0.01 - 0.05$ = significant, $0.05 - 0.1$ = zwak significant, $0.1 - 1$ = niet significant.

3.2.3 Totaal stikstof

Omdat totaal stikstof voor een groot deel bestaat uit anorganisch stikstof, zijn de resultaten voor totaal stikstof vergelijkbaar met die voor anorganisch stikstof. Dat geldt voor zowel het patroon (figuur 3.15) als voor de absolute concentraties (figuur 3.16).



Figuur 3.16: Totaal stikstof (mg l^{-1}) op dagbasis voor de periode 1986-2000 (boven) en 2001-2007 (onder).



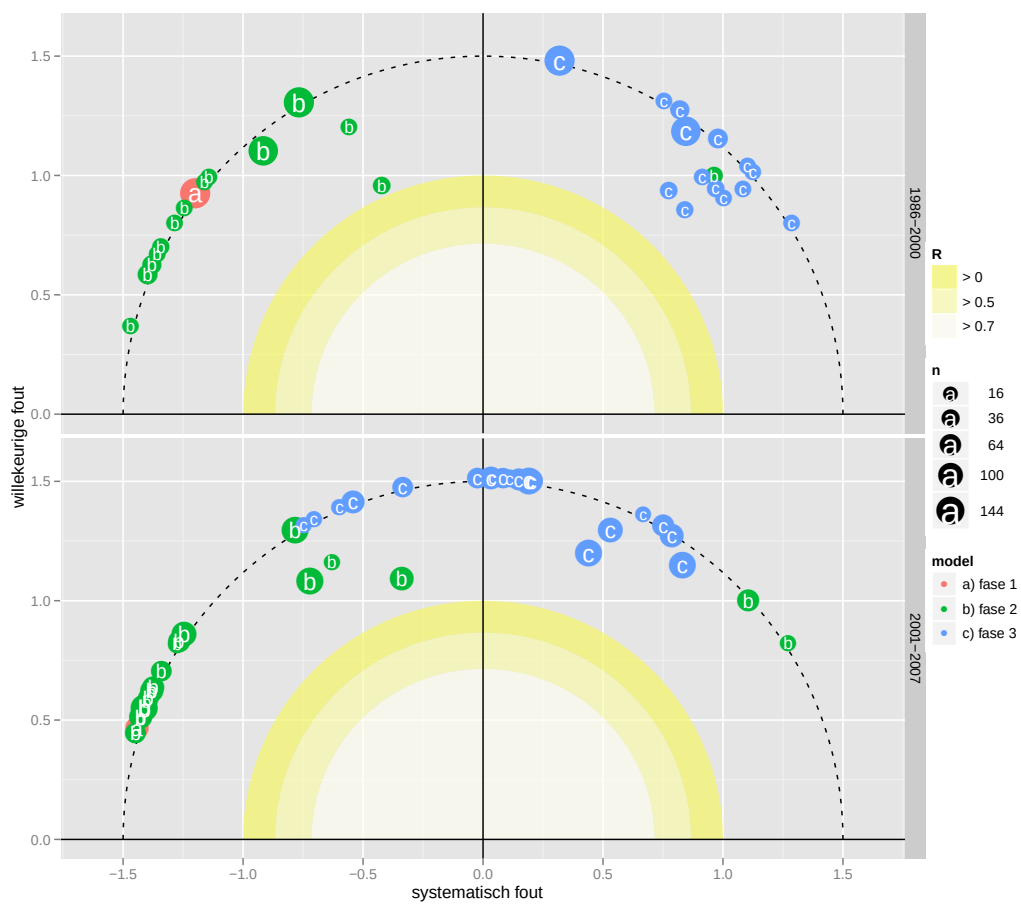
Figuur 3.17: Pearsons (r) correlatiecoëfficiënt (r) tussen de gemeten en gesimuleerde anorganische fosforconcentratie (mg l^{-1}) op alle locaties in het stroomgebied van ‘de Drentse Aa’ voor het hele jaar (kolom 1), het winterhalfjaar (kolom 2) het zomerhalfjaar (kolom 3) en twee perioden (rijen). De significantie van r is gegeven als p -waarde: $0-0.001$ = uiterst significant, $0.001-0.01$ = zeer significant, $0.01 - 0.05$ = significant, $0.05 - 0.1$ = zwak significant, $0.1 - 1$ = niet significant.

3.3 Fosfor

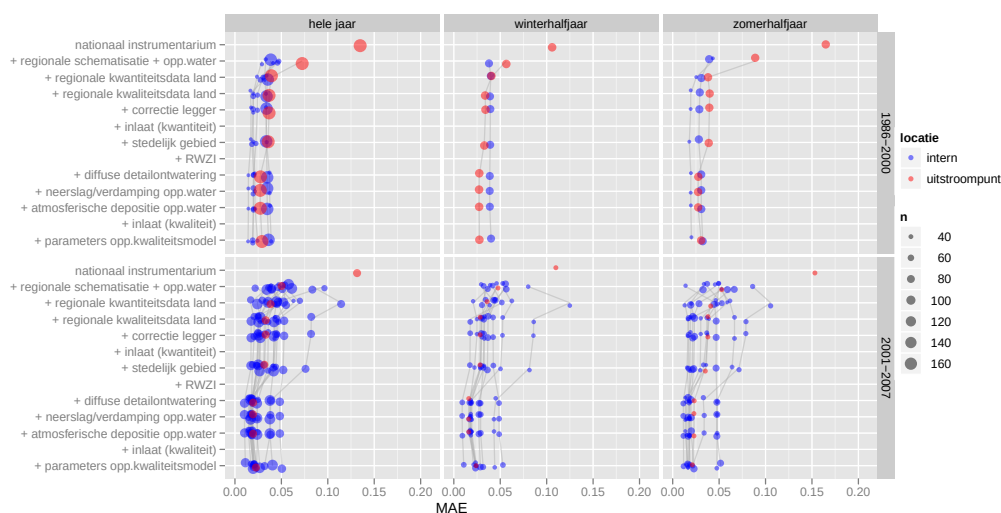
3.3.1 Anorganisch fosfor

Uit figuur 3.17 blijkt dat het patroon in anorganisch fosfor lastig is te modelleren. De meeste verbanden zijn niet significant, en als ze significant zijn dan is het verband zwak. Opvallend is dat ook de anorganische fosforconcentraties op de uitstroompunten slecht worden gereproduceerd en voorspeld.

Wat voor het patroon geldt, geldt ook voor het modelleren van absolute anorganisch fosforconcentraties: deze worden nog niet goed door de modellen gesimuleerd. Dit geldt voor zowel voor puntmetingen (momentopnames, figuur 3.18) als voor weekgemiddelde concentraties op basis van debietproportionale metingen (niet weergegeven). Voor zowel de periode 1986-2000 als de validatieperiode 2001-2007 geldt dat vrijwel alle punten buiten de zon liggen. De gemiddelde absolute fout (figuur 3.19) laat echter zien dat er desalniettemin enige vooruitgang is geboekt. Immers, de absolute fout neemt van fase 1 naar fase 3 af, met name in de validatieperiode. Over het algemeen wordt in fase 1 en 2 het gehalte aan anorganisch fosfor overschat, en in fase 3 meestal onderschat (figuur 3.18). Fase 1, d.w.z. het ‘nationaal modelinstrumentarium’,

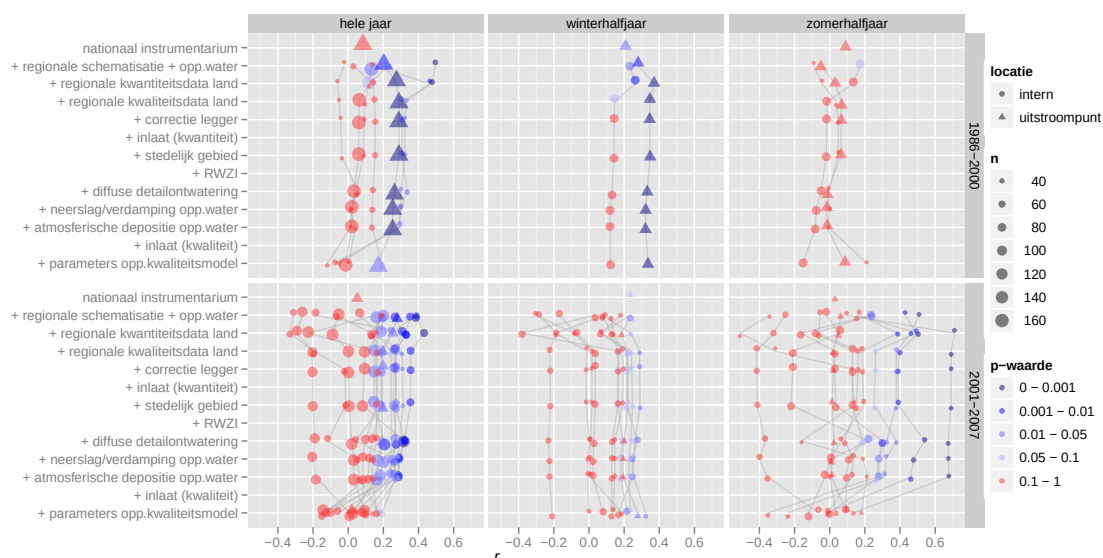


Figuur 3.18: Anorganisch fosfor (mg l^{-1}) op basis van puntmetingen voor de periode 1986-2000 (boven) en 2001-2007 (onder).



Figuur 3.19: Gemiddelde absolute fout (mg l^{-1}) voor anorganisch fosfor (mg l^{-1}) op dagbasis voor de hoofdfases.

moet ook een systematische overschatting geven, omdat dat modelsysteem geen fosforverwijderende processen ('sinks') in het oppervlaktewater bevat.

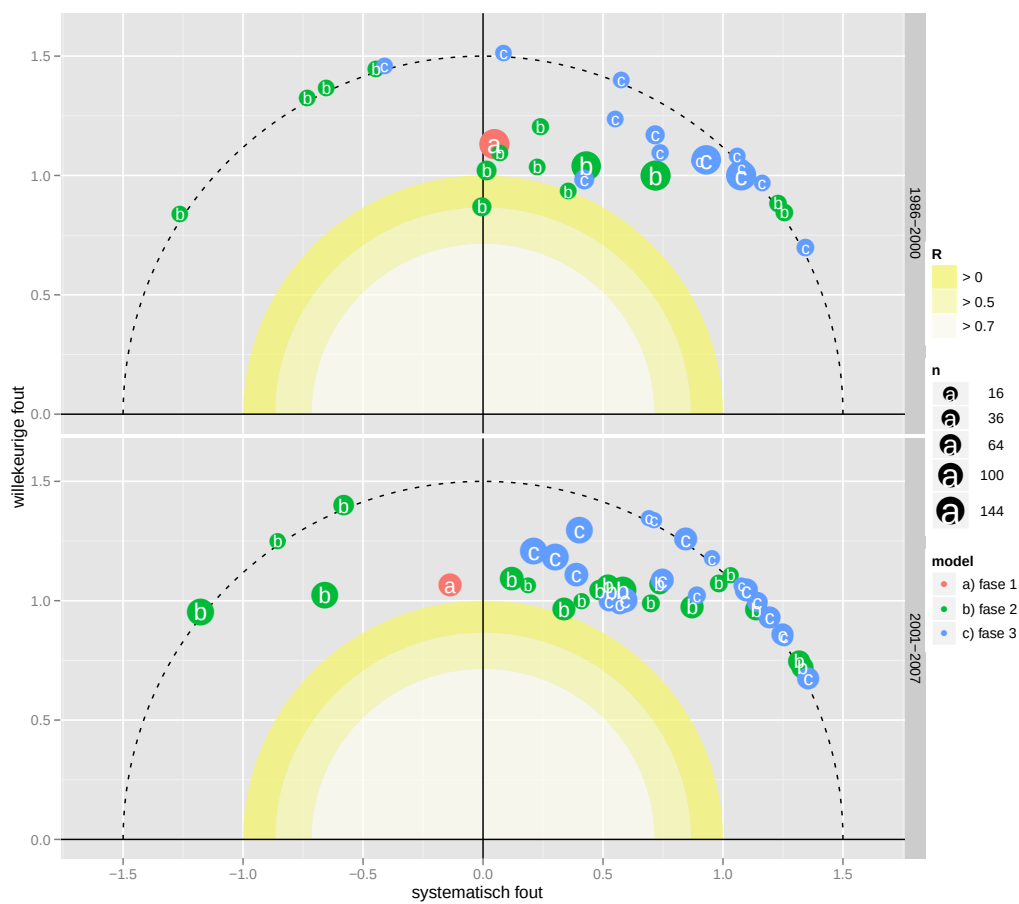


Figuur 3.20: Pearsons correlatiecoëfficiënt (r) tussen de gemeten en gesimuleerde organische fosforconcentratie (mg l^{-1}) op alle locaties in het stroomgebied van 'de Drentse Aa' voor het hele jaar (kolom 1), het winterhalfjaar (kolom 2) het zomerhalfjaar (kolom 3) en twee perioden (rijen). De significantie van r is gegeven als p -waarde: $0-0.001$ = uiterst significant, $0.001-0.01$ = zeer significant, $0.01 - 0.05$ = significant, $0.05 - 0.1$ = zwak significant, $0.1 - 1$ = niet significant.

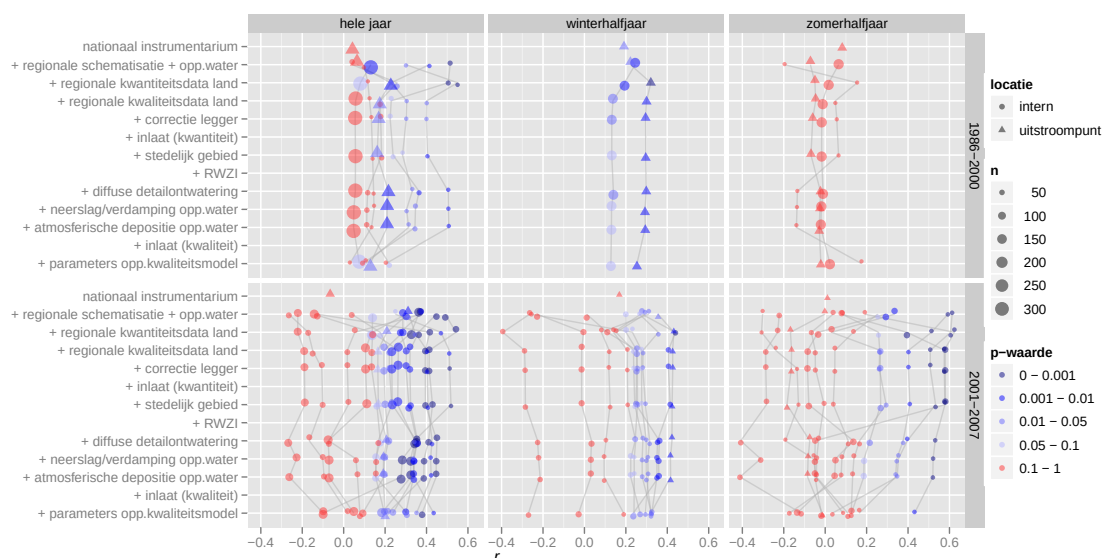
3.3.2 Organisch fosfor

Uit figuur 3.20 blijkt dat de temporele patronen in organisch fosfor lastig zijn te modelleren. Niet op alle meetlocaties wordt een significant verband gevonden. Als dat wel het geval is dan is het verband meestal zwak.

Uit figuur 3.21 blijkt dat ook absolute organische fosforconcentraties lastig zijn te modelleren. Vrijwel geen enkele locatie ligt binnen de zon.



Figuur 3.21: Organisch fosfor (mg l^{-1}) op basis van puntmetingen voor de periode 1986-2000 (boven) en 2001-2007 (onder).

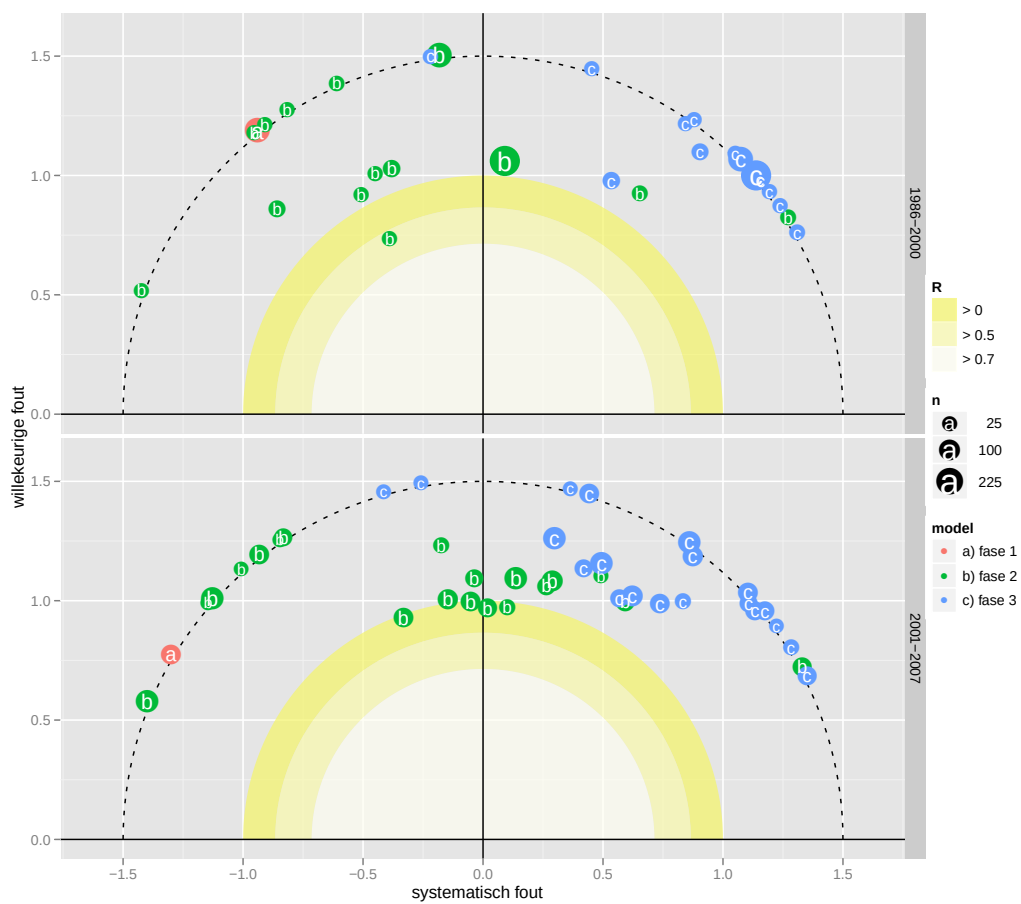


Figuur 3.22: Pearsons (r) correlatiecoëfficiënt tussen de gemeten en gesimuleerde totale fosforconcentratie (mg l^{-1}) op alle locaties in het stroomgebied van 'de Drentse Aa' voor het hele jaar (kolom 1), het winterhalfjaar (kolom 2) het zomerhalfjaar (kolom 3) en twee perioden (rijen). De significantie van r is gegeven als p -waarde: $0-0.001$ = uiterst significant, $0.001-0.01$ = zeer significant, $0.01 - 0.05$ = significant, $0.05 - 0.1$ = zwak significant, $0.1 - 1$ = niet significant.

3.3.3 Totaal fosfor

Wat voor anorganisch en organisch fosfor geldt, geldt ook voor totaal fosfor: het temporele patroon wordt niet goed gemodelleerd. Op veel locaties wordt geen significant verband gevonden, en als er een significant verband bestaat dan is die zwak (figuur 3.22).

Ook het modelleren van de absolute totale fosforconcentraties gaat nog niet goed. In fase 1 vindt een systematische overschatting plaats van totaal fosfor. Dat is ook te verwachten omdat retentieprocessen in het oppervlaktewater bij fase 1 niet zijn meegenomen. In fase 3 wordt de totale fosforconcentratie voornamelijk onderschat (zie ook Roelsma et al. (2008)).



Figuur 3.23: Totaal fosfor (mg l^{-1}) op basis van puntmetingen voor de periode 1986-2000 (boven) en 2001-2007 (onder).

4 Conclusies

In dit hoofdstuk worden de conclusies voor de waterafvoer, de stikstofconcentraties en de fosforconcentraties kort samengevat. De conclusies zijn tevens in kwalitatieve termen opgenomen in tabel 4.1.

4.1 Waterafvoer

De modelsystemen zijn in staat om het temporele patroon van de afvoermetingen te reproduceren en te voorspellen. Het zomerhalfjaar wordt vrijwel even goed gereproduceerd en voorspeld als het winterhalfjaar. De deelfase waarbij de hydrologie van het landsysteem wordt geregionaliseerd heeft een positief effect op het beschrijven en voorspellen van het temporele patroon. Toevoeging van de kleinste waterlopen reduceert de modelprestatie enigszins. De deelfase ‘neerslag/verdamping oppervlaktewater’ heeft vervolgens weer een gunstig effect op het reproduceren en voorspellen van temporele patronen.

Het modelsysteem heeft voor de meeste meetlocaties een meerwaarde wat betreft het reproduceren en voorspellen van absolute waterafvoeren.

Het corrigeren van de leggergegevens heeft alleen een negatief effect op de modelprestatie voor de periode tot 2000. Waarschijnlijk mogen de gecorrigeerde leggergegevens alleen gebruikt worden voor het modelleren van recente perioden. Voor perioden verder terug in de tijd moet gebruik worden gemaakt van historische leggerinformatie.

Ook de deelfase waarin de kleinste waterlopen worden gemodelleerd vermin-

Tabel 4.1: Kwalitatieve samenvatting van de resultaten voor ‘de Drentse Aa’

grootheid	temporeel patroon	absolute waarden	opmerkingen
water	redelijk	matig-redelijk	
anorganisch N	matig-redelijk	slecht-redelijk	gaat redelijk voor fase 2
organisch N	slecht	slecht	
totaal N	matig-redelijk	slecht-redelijk	vergelijkbaar met anorganisch N
anorganisch P	slecht-(matig)	slecht	
organisch P	slecht	slecht	
totaal P	slecht	slecht	

deren de modelprestatie wat betreft de waterafvoer. Waarschijnlijk worden de afvoerpieken te veel gedempt.

4.2 Stikstof

De modellen zijn in staat om het temporele patroon in anorganisch stikstof en totaal stikstof te reproduceren en te voorspellen. De regionalisering van de hydrologie laat in eerste instantie een verslechtering zien waarna met name het toevoegen van de diffuse detailontwatering om de retentie in de kleinste waterlopen te simuleren tot een verbetering leidt. De laatste deelfase van fase 3, waarin gebruik wordt gemaakt van een regionalisering van de parameters van het oppervlaktewaterkwaliteitsmodel leidt echter weer tot een afname van de modelprestatie.

Als wordt gekeken naar absolute concentraties in plaats van naar temporele patronen dan blijkt dat alleen het fase 2 model in staat is om anorganisch stikstof en totaal stikstof te reproduceren en te voorspellen. De fase 1 en 3 systemen overschatten de metingen. De reden voor de achteruitgang in modelprestatie tussen fase 2 en fase 3 wordt veroorzaakt door deelfase 'regionale kwaliteitsdata land'. Met name de afname van de denitrificatie is hiervoor verantwoordelijk, waardoor meer anorganisch stikstof beschikbaar komt voor uitspoeling naar het oppervlaktewater.

De modelsystemen hebben meer moeite om de temporele patronen in en absolute concentraties van organisch stikstof te reproduceren en te voorspellen. De metingen worden over het algemeen onderschat. De modelresultaten hebben geen meerwaarde ten opzichte van de gemiddelde meetwaarde.

4.3 Fosfor

Het blijkt lastig om het temporele patroon in anorganische fosforgehalten te modelleren. Absolute concentraties zijn nog moeilijker te voorspellen. Over het algemeen hebben de modelsystemen geen meerwaarde boven het gemiddelde van de metingen. Dit geldt voor zowel puntmetingen (momentopnames) als weekgemiddelde waarnemingen op basis van debietproportionele metingen. Hoewel het momenteel nog niet mogelijk is om anorganisch fosfor goed te modelleren neemt de totale fout wel af gaande van fase 1 naar fase 3. Soortgelijke conclusies kunnen worden getrokken voor organisch fosfor en totaal fosfor.

4.4 Algemeen

In het kader van het project ‘Monitoring Stroomgebieden’ wordt een modelsysteem ontwikkeld dat uiteindelijk de ruggengraat moet gaan vormen van een monitoringsysteem. Het modelsysteem moet informatie verschaffen over waterafvoeren en nutriëntenconcentraties op locaties waarvoor en tijdstippen waarop geen metingen beschikbaar zijn. Het modelsysteem wordt gefaseerd ontwikkeld. Elke fase bouwt voort op vorige fases.

Uit de resultaten blijkt dat de modelsystemen van fases 1 tot en met 3 nog niet in staat zijn om absolute nutriëntenconcentraties te reproduceren en te voorspellen. Uitzondering is het fase 2 model dat een meerwaarde heeft met betrekking tot het voorspellen van anorganisch en totaal stikstof. Met de fase 3 modellen is het op een aantal locaties al wel mogelijk om het temporele patroon te reproduceren en te voorspellen. Het modelsysteem zegt dan niets over de absolute grootte van de stikstofconcentratie, maar wel of er sprake is van een *verandering* van de concentratie. Voor het monitoren van trends is dit waardevolle informatie. Voor het monitoren of een kritische waarde wordt overschreden is het goed voorspellen van enkel het temporele patroon onvoldoende. Dan dienen de modellen ook in staat te zijn om absolute concentraties te kunnen voorspellen.

5 Aanbevelingen

1. Aangezien er nog veel onzekerheden bestaan omtrent de parameterisering van het landsysteem verdient het aanbeveling om de parameterisering en ruimtelijke schematisering van het landsysteem te verbeteren. Met name het ‘watergetal’ en daarmee de denitrificatie verdient extra aandacht. Evenals het oppervlaktewaterstelsel dient het landsysteem te worden getoetst aan de hand van meetgegevens. Databestanden die daarvoor in aanmerking komen zijn met name grondwaterstanden en grondwaterkwaliteitgegevens (bovenste grondwater).
2. Uit de validatie van de afvoerintensiteit is naar voren gekomen dat verbetering van het knooppuntennetwerk van het oppervlaktewatermodel op basis van gecorrigeerde/geactualiseerde leggerinformatie een negatief effect kan hebben op de modelprestatie. Aanbevolen wordt dan ook om de gecorrigeerde/geactualiseerde leggerinformatie alleen te gebruiken voor het modelleren van de huidige toestand, en voor het modelleren van historische tijdreeksen gebruik te maken van historische leggerinformatie.
3. De regionalisering van de parameterisatie van het oppervlaktewaterkwaliteitsmodel in de laatste deelfase van fase 3 dient beter te worden onderbouwd. Deze deelfase is voor een aanzienlijk deel verantwoordelijk voor de slechte resultaten van fase 3.

6 Dankbetuiging

We willen de volgende personen hartelijk bedanken voor het kritisch doorlezen en becommentariëren van conceptversies van één of meerdere validatierapporten (rapportnummers 1951 tot en met 1954):

Dick Brus (Alterra), Patrick Bogaart (Alterra), Peter van Boheemen (Ministerie van LNV), Frank van der Bolt (Alterra), Gé van den Eertwegh (Waterschap Rivierenland), Ariënné Oortwijn (Alterra), Hella Pomarius (Waterschap Rivierenland), Jan Roelsma (Alterra), Christian Siderius (Alterra), Dorothée van Tol - Leenders (Alterra) en Wim Twisk (Hoogheemraadschap van Schieland en de Krimpenerwaard).

Literatuur

- Codd, E. F., 1970. A relational model of data for large shared data banks. Communications of the ACM 13 (6), 377–387.
- Euclid, 2008. Elements book 1. In: Fitzpatrick, R. (Ed.), Euclids Elements Of Geometry (Euclidis Elementa), 1st Edition. The University of Texas, Austin. URL <http://farside.ph.utexas.edu/euclid/Elements.pdf>
- Groenendijk, P., Renaud, L. V., Roelsma, J., 2005. Prediction of nitrogen and phosphorus leaching to groundwater and surface waters. Process descriptions of the animo4.0 model. Tech. Rep. 983, Alterra, Wageningen. URL http://library.wur.nl/wasp/bestanden/LUWPUBRD_00340937_A502_001.pdf
- Jolliff, J. K., Kindle, J. C., Shulman, I., Penta, B., Friedrichs, M. A. M., Helber, R., Arnone, R. A., 2009. Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. Journal of Marine Systems 76, 64–82.
- Jørgensen, L. F., Refsgaard, J. C., Højberg, A. L., 2007. The inadequacy of monitoring without modelling support. Journal of Environmental Modelling 9, 931–942.
- Kendall, M. G., 1938. A new measure of rank correlation. Biometrika 30 (1/2), 81–93.
- Konikow, L. F., Bredehoeft, J. D., 1992. Ground-water models cannot be validated. Advances in Water Resources 15, 75–83.
- Legates, D. R., McCabe, G. J., 1999. Evaluating the use of ‘goodness-of-fit’ measures in hydrologic and hydroclimatic model validation. Water Resources Research 35, 233–241.
- Nash, J. E., Sutcliffe, J. V., 1970. River flow forecasting through conceptual models part I – a discussion of principles. Journal of Hydrology 10 (3), 282–290.
- Oreskes, N., Shrader-Frechette, K., Belitz, K., 1994. Verification, validation, and confirmation of numerical models in the earth sciences. Science 263, 641–646.
- Portielje, R., van der Gaast, J. W. J., van der Kolk, J. W. H., Schoumans, O. F., Boers, P. C. M., 2002. Nutriëntenconcentraties en -trends in kleine landbouwbeïnvloede wateren, 1985-2000. Tech. Rep. 2002.008, Rijksinstituut voor Integraal Zoetwaterbeheer en Afvalwaterbehandeling (RIZA), Alterra.

- Rode, M., Suhr, U., 2007. Uncertainties in selected river water quality data. *Hydrology and Earth System Sciences* 11, 863–874.
- Roelsma, J., van der Bolt, F. J. E., Leenders, T. P., Renaud, L. V., 2006a. Systeemanalyse voor het stroomgebied van de Drentse Aa fase 1. Reeks Monitoring Stroomgebieden 5-I 1271, Alterra, Wageningen.
- Roelsma, J., van der Bolt, F. J. E., Leenders, T. P., Renaud, L. V., de Vries, I., van der Molen, K., 2006b. Systeemanalyse voor het stroomgebied van de Drentse Aa fase 2. Reeks Monitoring Stroomgebieden 8-I 1386, Alterra, Wageningen.
- Roelsma, J., van Tol-Leenders, T. P., van der Bolt, F. J. E., Löschner-Wolleswinkel, R. J., Renaud, L. V., Schaap, J. D., Schoumans, O. F., Siderius, C., van der Heide, H., van der Molen, K., 2008. Systeemanalyse voor het stroomgebied van de Drentse Aa fase 3. Tech. Rep. 1764, Alterra, Wageningen.
- Siderius, C., Groenendijk, P., Gerven, L. P. A. v., Jeuken, H. H. J. L., Smit, A. A. M. F. R., 2008. Process description of NUSWALITE. A simplified model for the fate of nutrients in surface waters. Alterra report 1226.2, Alterra, Wageningen, The Netherlands.
- Siegel, S., Castellan, N. J., 1988. Nonparametric statistics for the behavioral sciences. McGraw-Hill, New York.
- Stow, C. A., Jolliff, J., McGillicuddy, D. J., Doney, S. C., Allen, J. I., Friedrichs, M. A. M., Rose, K. A., Wallhead, P., 2009. Skill assessment for coupled biological/physical models of marine systems. *Journal of Marine Systems* 76, 4–15.
- Taylor, K. E., April 2000. Summarizing multiple aspects of model performance in a single diagram. Program For Climate Model Diagnosis And Intercomparison (PCMDI) 55, University Of California, Lawrence Livermore National Laboratory, Livermore, CA 94550.
- Taylor, K. E., April 2001. Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research* 106 (D7), 7183–7192.
- van der Gaast, J. W. J., Massop, H. T. L., Vroon, H. R. J., Staritsky, I. G., 2006. Hydrologie op basis van karteerbare kenmerken. Tech. Rep. 1339, Alterra, Wageningen.
- van Waveren, R. H., Groot, S., Scholten, H., van Geer, F., Wösten, H., Koeze, R., Noort, J., 1999. Vloeiend modelleren in het waterbeheer: Handboek Good Modelling Practice. Tech. Rep. 99-05, STOWA/RWS-RIZA, Utrecht/Lelystad.
- Walvoort, D. J. J., van Tol-Leenders, T. P., 2009. Database ‘monitoring stroomgebieden’: een slimme database voor het beheren van monitoringsgegevens. Tech. Rep. 1955, Alterra Wageningen, Wageningen.

A Modelprestatiematen

A.1 Algemeen foutenmodel

De gebruikte validatiematen zijn een functie van

- de meetwaarden (y_i) voor $i = 1 \dots n$;
- de daarmee corresponderende modelvoorspellingen $(\hat{y}_i)$ voor $i = 1 \dots n$;
- of van het verschil tussen y_i en $\hat{y}_i$ (e_i) voor $i = 1 \dots n$.

Hier is n het aantal meetwaarden. De relatie tussen deze drie karakteristieken wordt gegeven door:

$$y_i = \hat{y}_i + e_i \quad (\text{A.1})$$

De restterm $e_i = y_i - \hat{y}_i$ is het gevolg van onvolkomenheden in het model (onjuiste procesformulering, ontbreken van processen, fouten in de parameterisering, fouten in de gebiedschematisering) en/of meetfouten.

A.2 Variantie

De variantie (σ_y^2) kwantificeert de variatie van de waarnemingen rond het rekenkundig gemiddelde en kan gezien worden als een maat voor de dynamiek:

$$\sigma_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \quad (\text{A.2})$$

$$\sigma_{\hat{y}}^2 = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2 \quad (\text{A.3})$$

waarin:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (\text{A.4})$$

het rekenkundig gemiddelde van de n waarnemingen y_i en

$$\bar{\hat{y}} = \frac{1}{n} \sum_{i=1}^n \hat{y}_i \quad (\text{A.5})$$

het rekenkundig gemiddelde van de n modelvoorspellingen $\hat{y}_i$.

De variantie ligt in $[0, \infty)$. Een variantie van 0 geeft aan dat er geen dynamiek is, d.w.z. dat y of $\hat{y}$ niet veranderen in de tijd. Hoe groter de variantie, hoe meer dynamiek.

Een interessante maat is de verhouding tussen de dynamiek van de berekende reeks ten opzichte van die van de waargenomen reeks:

$$\tilde{\sigma}_y^2 = \frac{\sigma_{\hat{y}}^2}{\sigma_y^2} \quad (\text{A.6})$$

Deze ratio geeft de genormaliseerde variantie van de berekende tijdreeks weer en ligt in $[0, \infty)$. Een waarde gelijk aan 1 geeft aan dat de berekende en waargenomen dynamiek gelijk zijn. Een waarde groter dan 1 geeft aan dat de waargenomen dynamiek wordt overschat, een waarde kleiner dan 1 duidt op onderschatting van de waargenomen dynamiek.

A.3 Correlatiecoëfficiënt

Pearsons (product-moment) correlatiecoëfficiënt is een maat voor de (lineaire) overeenkomst in patroon tussen de gemeten en gemodelleerde tijdreeksen en wordt gegeven door:

$$r = \frac{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sigma_y \sigma_{\hat{y}}} \quad (\text{A.7})$$

De correlatiecoëfficiënt ligt in $[-1, 1]$. De correlatiecoëfficiënt is gelijk aan 1 als de gemodelleerde reeks lineair evenredig is met de waargenomen reeks. Dat wil zeggen dat

$$\frac{(y_i - \bar{y})}{(\hat{y}_i - \bar{\hat{y}})} = c \quad \text{voor } c > 0, \text{ en } i = 1 \dots n$$

Een correlatie van 1 wil zeggen dat de waargenomen en berekende tijdreeksen perfect in fase zijn. Een correlatie van -1 geeft aan dat de tijdreeksen in tegenfase zijn. Er is dan sprake van een faseverschuiving.

A.4 Systematische of gemiddelde fout

De gemiddelde fout of systematische fout (Eng. *bias*), wordt gegeven door:

$$\bar{e} = \frac{1}{n} \sum_{i=1}^n e_i = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i) = \bar{y} - \bar{\hat{y}} \quad (\text{A.8})$$

Een zuiver (Eng. *unbiased*) model heeft een systematische fout van 0. Veel statistische modellen zijn zuivere schatters en hebben geen systematische fout.

Een procesmodel daarentegen kan systematisch te hoge waarden ($\bar{e} < 0$) of te lage waarden ($\bar{e} > 0$) geven. Een systematische fout kan bijvoorbeeld worden veroorzaakt doordat randvoorwaarden van het model (bijvoorbeeld de hoeveelheid neerslag, de fluxdichtheid van de kwel, of de belasting vanuit de industrie) systematisch te hoog of te laag worden ingeschat.

A.5 Willekeurige fout

De willekeurige fout is een maat voor de precisie en kan worden berekend als de variatie van de fout rond de gemiddelde fout:

$$\sigma_e^2 = \frac{1}{n} \sum_{i=1}^n (e_i - \bar{e})^2 \quad (\text{A.9})$$

Deze fout wordt veroorzaakt door ondermeer:

1. meetfouten (fouten meetapparatuur, bemonsteringsfouten, laboratoriumfouten, e.d.). Zie ook bijvoorbeeld Rode and Suhr (2007);
2. modelfouten (ontbreken van processen, fouten c.q. vereenvoudigingen in de gebiedsschematisatie, modelparameterisatie en modelrandvoorwaarden (kwel, neerslag, verdamping, inlaatconcentraties, e.d.);
3. verschil in ruimtelijke en temporele schaal van meetwaarden en modelresultaten;

A.6 Totale fout (gemiddelde gekwadrateerde fout)

$$\sigma_t^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \frac{1}{n} \sum_{i=1}^n e_i^2 \quad (\text{A.10})$$

De totale fout (gemiddelde gekwadrateerde fout, Eng. *Mean Squared Error*, MSE) is een maat voor de nauwkeurigheid. Deze fout kan worden opgesplitst in twee componenten, namelijk, de gekwadrateerde systematische fout en de willekeurige fout:

$$\sigma_t^2 = \bar{e}^2 + \sigma_e^2 \quad (\text{A.11})$$

A.7 Gemiddelde absolute fout

$$\bar{e}_a = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (\text{A.12})$$

De gemiddelde absolute fout (Eng. *Mean Absolute Error*, MAE) lijkt op de gemiddelde gekwadrateerde fout, maar is minder gevoelig voor extreme fouten. Immers, de fout wordt niet gekwadrateerd, maar absoluut genomen.

A.8 Modefficiëntie

De modefficiëntie volgens Nash and Sutcliffe (1970) wordt berekend met:

$$E_{\text{NS}} = \frac{F_0^2 - F^2}{F_0^2} \quad (\text{A.13})$$

waarin $F_0^2 = \sum_{i=1}^n (y_i - \bar{y})^2$ en $F^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$. F^2 is de resterende kwadraatsom die niet door het model kan worden verklaard. F_0^2 is de kwadraatsom van de metingen. De modefficiëntie ligt in $(-\infty, 1]$. $E_{\text{NS}} = 1$ komt overeen met een perfect model, $E_{\text{NS}} = 0$ met een model die het net zo goed doet als wanneer het gemiddelde van de meetgegevens als predictor wordt gebruikt. Waarden kleiner dan 0 geven aan dat het gemiddelde van de meetgegevens een betere predictor is dan het model.

E_{NS} is gerelateerd aan σ_t^2 via:

$$E_{\text{NS}} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sigma_t^2}{\sigma_y^2} \quad (\text{A.14})$$

E_{NS} is dus gelijk aan 1 minus de genormaliseerde gemiddelde gekwadrateerde fout.

B Taylor-diagram

De basis voor het Taylor-diagram is de volgende vergelijking:

$$\sigma_e^2 = \sigma_y^2 + \sigma_{\hat{y}}^2 - 2\sigma_y\sigma_{\hat{y}}\cos r \quad (\text{B.1})$$

waarin σ_e^2 de willekeurige fout, σ_y^2 de variantie van de waarnemingen, $\sigma_{\hat{y}}^2$ de variantie van de modelvoorspellingen, en r de correlatiecoëfficiënt tussen de waarnemingen en de modelvoorspellingen (zie bijlage A voor meer informatie over deze modelprestatiematen).

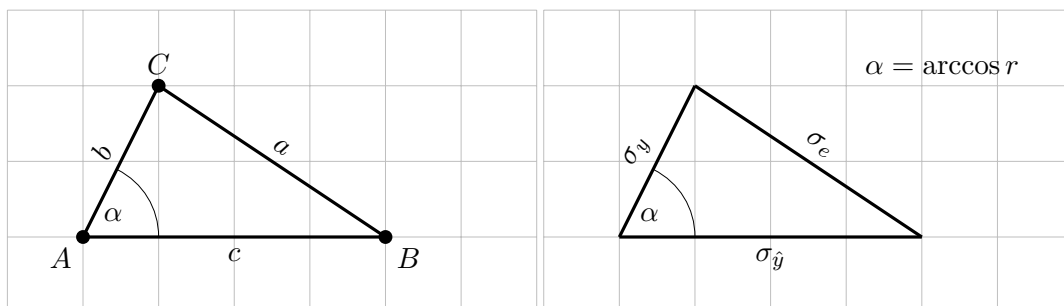
(Taylor, 2000, 2001) gebruikte de gelijkenis tussen vergelijking B.1 en de cosinusregel

$$a^2 = b^2 + c^2 - 2bc\cos\alpha \quad (\text{B.2})$$

om een nieuw type diagram af te leiden waarin de modelprestaties in relatie tot elkaar kunnen worden weergegeven. Een geometrische weergave van deze gelijkenis is gegeven in figuur B.1.

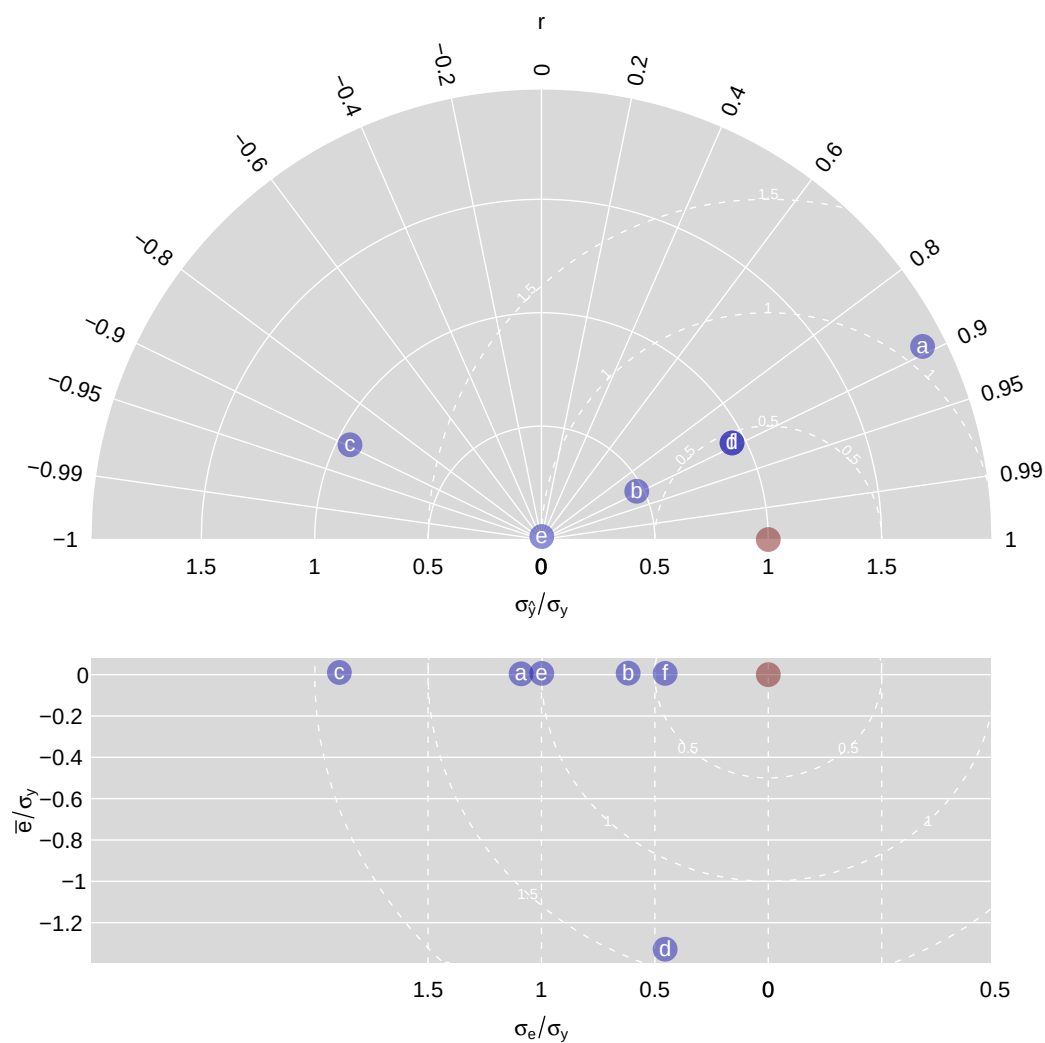
Het Taylor-diagram is gegeven in de bovenste helft van figuur B.2. Langs de horizontale as is de genormaliseerde standaarddeviatie weergegeven. Dat is de vierkantswortel uit de genormaliseerde variantie (A.6). Langs de cirkelboog is de correlatiecoëfficiënt r uitgezet. De rode stip stelt de waarnemingen voor. Omdat de waarnemingen perfect gecorreleerd zijn met zichzelf ($r = 1$) en per definitie een genormaliseerde standaardafwijking hebben van 1 heeft de rode stip als coördinaten $(\sigma_{\hat{y}}/\sigma_y = 1, r = 1)$. De blauwe stippen stellen de modelresultaten voor uit figuur 2.2. De afstand tussen een blauwe stip en de rode stip is de genormaliseerde willekeurige fout σ_e^2/σ_y . Hoe dichter een blauwe stip bij de rode stip ligt, hoe preciezer het model.

Een nadeel van het Taylor-diagram is dat het geen onderscheid kan maken tussen modelprestatie op basis van systematische fouten (vergelijking A.4). Modellen **d** en **f** lijken het daarom even goed te doen. Het feit dat model **d** de waarnemingen systematisch overschat blijkt niet uit het Taylor-diagram. Daarom is in dit rapport het Taylor-diagram uitgebreid met een extra as die de genormaliseerde systematische fout weergeeft. Deze as staat loodrecht op het Taylor-diagram en gaat door punt $(\sigma_{\hat{y}}/\sigma_y = 1, r = 1)$. De onderste helft van figuur B.2 kan worden gezien als een projectie van het Taylor-diagram op het grondvlak door rotatie van alle blauwe punten rond deze as. De gestreepte cirkels van het Taylor-diagram gaan als verticale gestreepte lijnen verder in de onderste helft van figuur B.2. De afstand van een blauwe stip tot de rode stip in de onderste helft van figuur B.2 is de genormaliseerde totale fout σ_t^2/σ_y . Hoe kleiner deze afstand, hoe groter de nauwkeurigheid van het model. Model

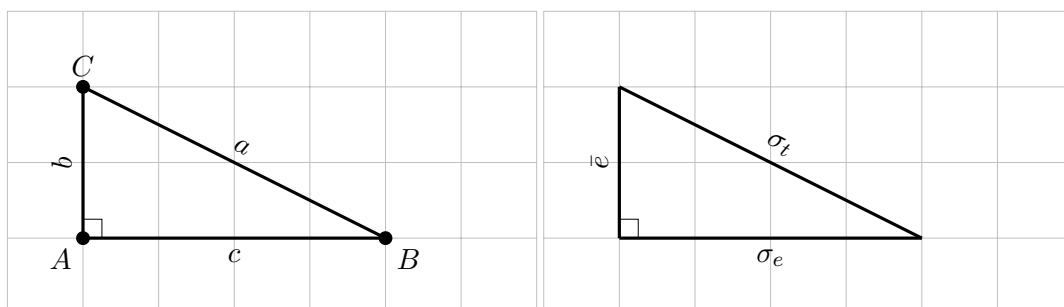


Figuur B.1: Grafische weergave van de relatie tussen de cosinusregel en enige validatiematen.

Ⓕ is daarom nauwkeuriger dan model Ⓓ. Dit komt doordat Ⓓ een grote systematische fout heeft. De basis voor het Taylor diagram is de cosinusregel, de basis voor het onderste deel van figuur B.2 is de stelling van Pythagoras (figuur B.3). Voor meer informatie over het Taylor-diagram en een aantal interessante toepassingen wordt verwezen naar (Taylor, 2001).



Figuur B.2: Taylor diagram (boven) gekoppeld aan $\bar{e} - \sigma_e$ plot (onder). De letters a tot en met f verwijzen naar de subfiguren in figuur 2.2.



Figuur B.3: Grafische weergave van de relatie tussen de stelling van Pythagoras en enige validatiematen.

C Zonneplot

De zonneplot is complementair aan het Taylordiagram. Het Taylordiagram geeft informatie over mogelijk faseverschuivingen (verg. A.7), de dynamiek van de gemodelleerde versus de waargenomen tijdreeks (vergelijking A.6) en de willekeurige fout (verg. A.9).

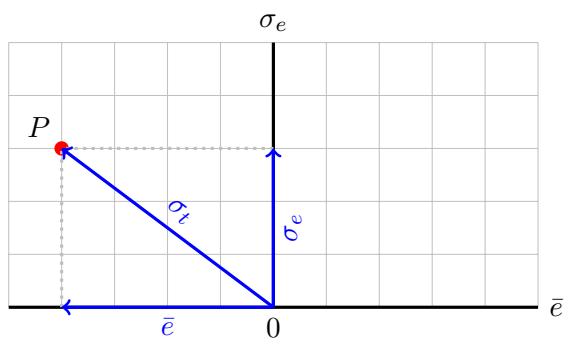
De zonneplot geeft de relatie weer tussen de willekeurige fout (verg. A.9, de systematische of gemiddelde fout (verg. A.8) en de totale fout (verg. A.9). Een geometrische voorstelling van deze relatie is gegeven in figuur C.1. Op de x -as staat de systematische of gemiddelde fout gegeven ($\bar{e}$). Negatieve waarden duiden op een systematische overschatting van de meetgegevens door het model, positieve waarden op een systematische onderschatting (zie ook vergelijking A.8). Op de y -as staat de willekeurige fout gegeven (σ_e). Hoe groter de willekeurige fout, hoe lager de precisie van het model. De totale fout (σ_t^2) kan worden berekend met behulp van de stelling van Pythagoras (Euclid, 2008, boek 1, propositie 47)

$$\sigma_t^2 = \bar{e}^2 + \sigma_e^2 \quad (\text{C.1})$$

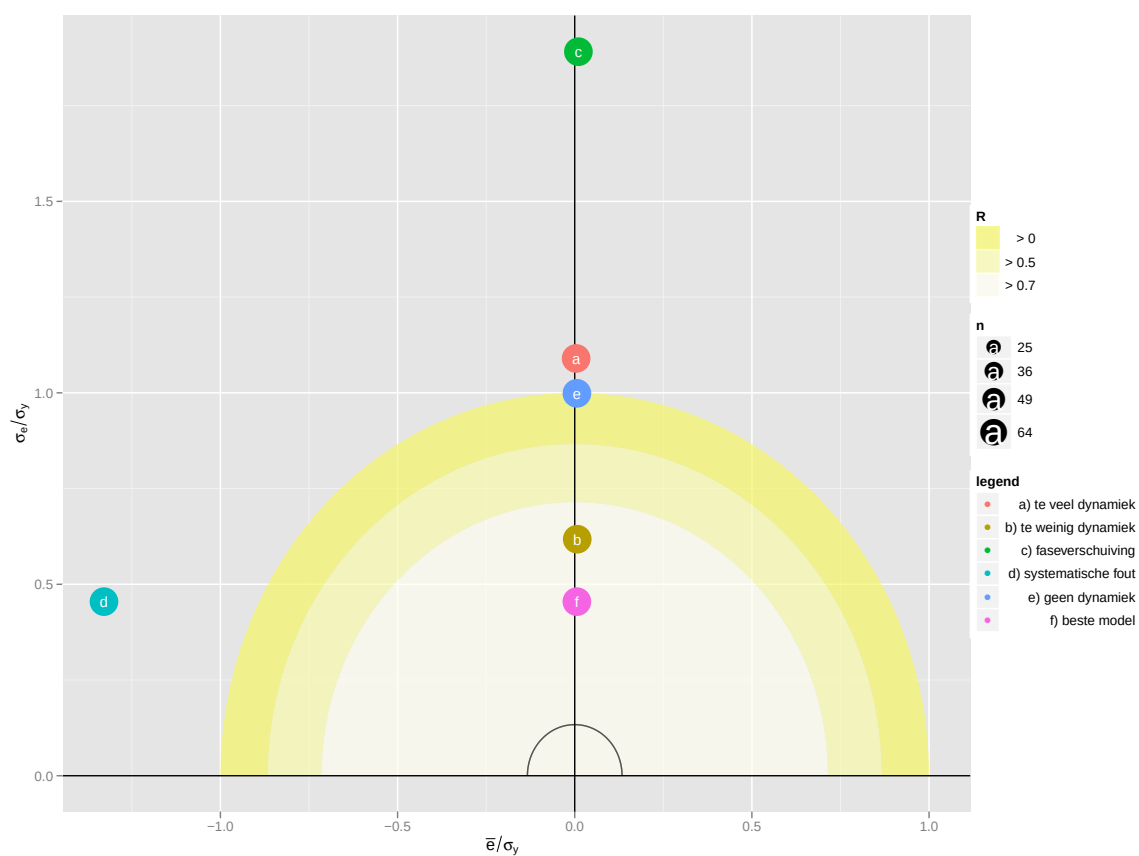
Op deze stelling berust de zonneplot. De zonneplot is gebaseerd op het ‘schietschijfdiagram’ (Engels: ‘target diagram’) van Jolliff et al. (2009). Een belangrijk verschil met het schietschijfdiagram is dat slechts twee in plaats van vier kwadranten worden gebruikt. Jolliff et al. (2009) gebruiken de resterende twee kwadranten om aan te geven *of* sprake is van een systematische onderdanwel overschatting van de variantie van de gemeten tijdreeks. De *grootte* van die systematische onderdanwel overschatting blijkt echter niet uit het schietschijfdiagram. Dit kan leiden tot verkeerde conclusies, doordat modelresultaten die marginaal van elkaar verschillen toch ver uit elkaar kunnen liggen in het schietschijfdiagram. In de zonneplot liggen deze resultaten wel dicht bij elkaar.

Een voorbeeld van een zonneplot is gegeven in figuur C.2. De assen zijn genormaliseerd door te delen door de standaarddeviatie van de metingen. Net zoals in figuur C.1 geldt dat de afstand van een punt tot de oorsprong gelijk is aan de totale fout σ_t .

Het diagram ontleent haar naam aan de halve cirkels rond de oorsprong. De buitenste cirkel correspondeert met een *genormaliseerde* totale fout van 1. Uit vergelijking A.14 blijkt dat dan de modefficiëntie gelijk is aan nul ($E_{\text{NS}} = 0$). Voor punten die buiten deze cirkel vallen geldt dan ook dat het gemiddelde van de waarnemingen een betere predictor is dan het model. Omgekeerd geldt dat voor punten die binnen deze cirkel vallen het model een meerwaarde heeft t.o.v. het gebruik van de gemiddelde meetwaarde als predictor. Ter wille van de



Figuur C.1: Geometrische weergave van de relatie tussen de willekeurige fout (σ_e^2), de systematische fout ($\bar{e}$) en de totale fout (σ_t^2) voor punt P.



Figuur C.2: Voorbeeld van een zonneplot.

leesbaarheid worden alle punten die verder dan een bepaalde kritische afstand d_c van het centrum van de zon zijn verwijderd, op een zwarte gestippelde cirkel geplaatst met straal $r = d_c$ waarvan het middelpunt samenvalt met het centrum van de zon¹.

Jolliff et al. (2009) tonen aan dat deze cirkel tevens een gebied afgrenst waarbinnen de correlatiecoëfficiënt (verg. A.7) groter is dan 0. Punten buiten deze cirkel duiden op modelresultaten die mogelijk (maar *niet* noodzakelijk) in tegenfase zijn met de waargenomen tijdreeksen. Faseverschuivingen kunnen echter beter worden geïdentificeerd met behulp van een Taylor diagram (bijlage B). De twee binnenste cirkels omgrenzen gebieden waar de lineaire correlatie tussen de gemodelleerde tijdreeksen en de gemeten tijdreeksen groter is dan respectievelijk 0.5 en 0.7.

Zoals eerder opgemerkt zijn niet alleen modelresultaten behept met fouten, maar ook de metingen. De meetfout wordt weergegeven door de zwarte open cirkel. Punten die binnen deze cirkel vallen hebben de maximale nauwkeurigheid bereikt. Verdere verkleining van het verschil tussen de gemodelleerde en gemeten waarden is dan weinig zinvol.

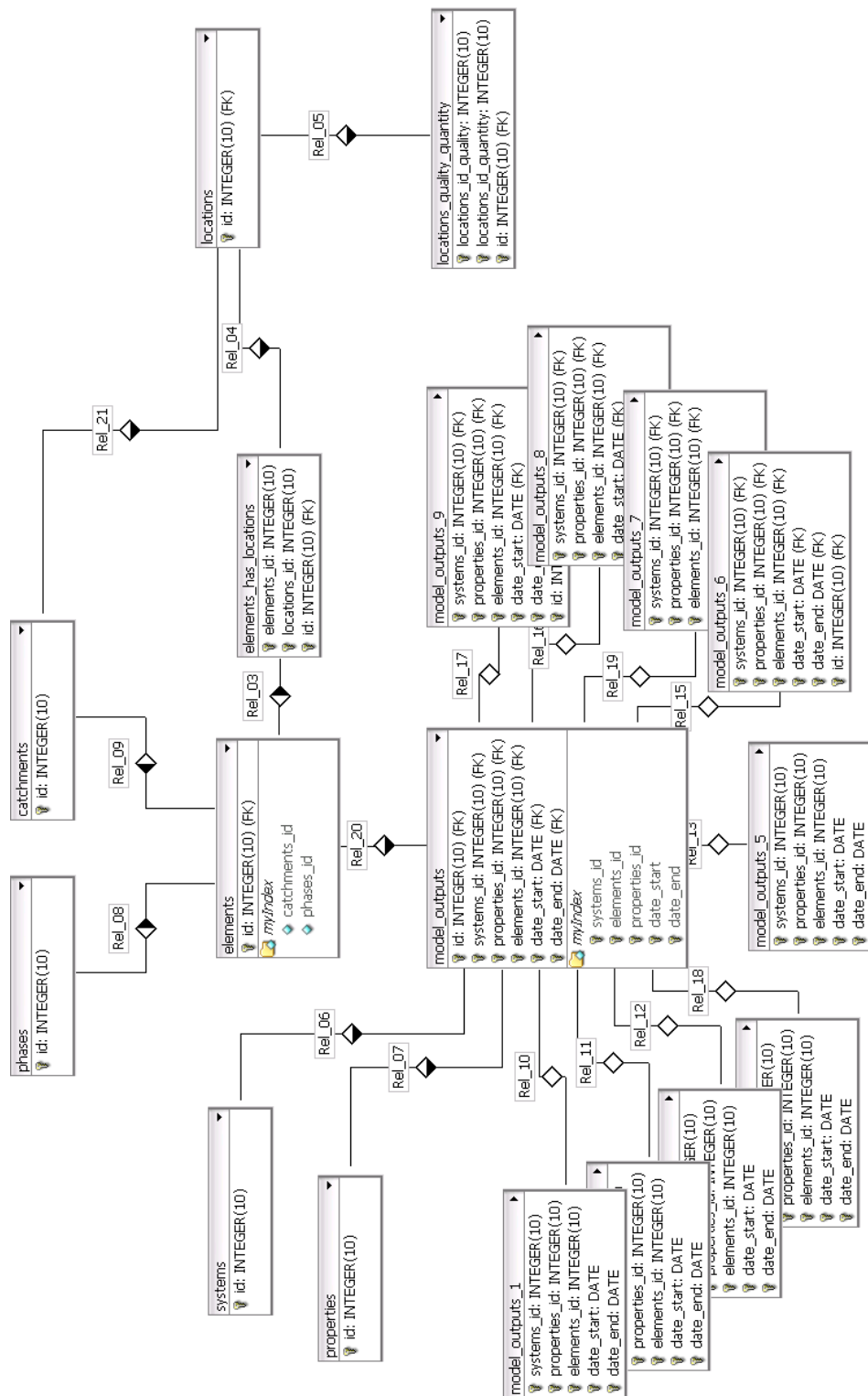
In figuur C.2 is de totale fout σ_t die correspondeert met de meetfout berekend door de variatiecoëfficiënt van de meetfout te vermenigvuldigen met het gemiddelde van de waarnemingen. Voor een aantal stoffen in het oppervlaktewater is de variatiecoëfficiënt van de meetfout getabelleerd in Rode and Suhr (2007). De genormaliseerde totale fout ($\frac{\sigma_{t,me}}{\sigma_y}$) die correspondeert met de meetfout kan op basis van de variatiecoëfficiënt $\frac{\sigma_{me}}{\bar{y}}$ als volgt worden berekend:

$$\frac{\sigma_{t,me}}{\sigma_y} = \frac{\bar{y} \frac{\sigma_{me}}{\bar{y}}}{\sigma_y}$$

¹In wiskundige bewoording: Stel dat een punt P op een afstand d van het middelpunt M van de zon ligt, en dat $d > d_c$, dan wordt P verplaatst naar het punt P' dat op een afstand van d_c van M op lijnstuk PM ligt.

D Datamodel modelresultaten

Figuur D.1 geeft het datamodel weer dat gebruikt is voor de opslag van modelresultaten.



Figuur D.1: Datamodel van de database waarin de modelresultaten zijn opgeslagen.

