



ALTERRA

WAGENINGEN UR

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

Fases 1, 2 en 3

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

Alterra-rapport 1952, ISSN 1566-7197
Reeks Monitoring Stroomgebieden 18-II



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

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

Fases 1, 2 en 3

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

Alterra-Rapport 1952

Reeks Monitoring Stroomgebieden nr. 18-II

Alterra, Wageningen, 2009

REFERAAT

Walvoort, D.J.J., J. Roelsma, R.J. Löschner-Wolleswinkel, T.P. van Tol-Leenders, 2009. *Validatie van modelsystemen voor het voorspellen van de oppervlaktewaterkwaliteit en -kwantiteit in het stroomgebied 'de Schuitenbeek'*. Wageningen, Alterra–Rapport 1952, Reeks Monitoring Stroomgebieden nr. 18-II.

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	20
2.3.4	Modelprestatie-maten	21
2.3.5	Interpretatie van afzonderlijke modelprestatie-maten	21
2.3.6	Interpretatie van modelprestatie-maten in onderlinge samenhang	25
3	Resultaten en discussie	31
3.1	Waterafvoer	32
3.2	Stikstof	40
3.2.1	Anorganisch stikstof	40
3.2.2	Organisch stikstof	46
3.2.3	Totaal stikstof	50
3.3	Fosfor	52
3.3.1	Anorganisch fosfor	52
3.3.2	Organisch fosfor	58
3.3.3	Totaal fosfor	61
4	Conclusies	65
4.1	Waterafvoer	65
4.2	Stikstof	66
4.3	Fosfor	67
4.4	Algemeen	67

5	Aanbevelingen	69
6	Dankbetuiging	71
A	Modelprestatiematen	75
A.1	Algemeen foutenmodel	75
A.2	Variantie	75
A.3	Correlatiecoëfficiënt	76
A.4	Systematische of gemiddelde fout	76
A.5	Willekeurige fout	77
A.6	Totale fout (gemiddelde gekwadrateerde fout)	77
A.7	Gemiddelde absolute fout	77
A.8	Modefficiëntie	78
B	Taylor-diagram	79
C	Zonneplot	83
D	Datamodel modelresultaten	87

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 modelleers (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 op 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 modelleers 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. Dat gaat beter voor het winterhalfjaar dan voor het zomerhalfjaar. Het zomerhalfjaar wordt namelijk gekenmerkt door extreme afvoerpieken afgewisseld met droogval hetgeen lastiger te modelleren is.

Hoewel het temporele patroon redelijk wordt gemodelleerd, overschatten de modelsystemen de gemeten afvoer. Dit geldt met name voor de meetpunten waar de afvoer is bepaald met een OTT-molenmeter. Op het continue meetpunt, waar de afvoermetingen nauwkeuriger zijn, is de systematische fout (systematische overschatting) veel kleiner en voor de validatieperiode zelfs vrijwel verdwenen. Het is bij validatie dan ook van belang dat men zich realiseert dat afwijkingen tussen gemeten en berekende waarden, niet alleen worden veroorzaakt door onvolkomenheden in het model, maar dat ook de metingen behept zijn met fouten.

Op basis van de modefficiëntie kan worden geconcludeerd dat de waterafvoer in fase 3 over het algemeen beter wordt voorspeld dan in voorgaande fases. Dit komt met name door de deelfase waarin de hydrologie van het landsysteem is geregionaliseerd. De modelsystemen hebben echter moeite om piekafvoeren te voorspellen in het zomerhalfjaar. Dit kan deels worden verbeterd door gebruik te maken van een hogere resolutie aan neerslaggegevens (en verdampingsgegevens) en/of aanpassen van de bodemschematisatie- en parameterisatie.

Over het algemeen presteren de modelsystemen beter op benedenstroomse dan op bovenstroomse meetlocaties ten gevolge van uitmiddelingseffecten. Daarnaast presteren de modelsystemen beter op het continue meetpunt dan op meetlocaties waar met OTT-molenmetingen wordt gemeten. Hiervoor zijn twee oorzaken aan te wijzen. Enerzijds zijn OTT-molenmetingen minder nauwkeurig, anderzijds is een OTT-molenmeting een momentopname en geen daggemiddelde zoals een continue meting. Momentopnames zijn voor snel reagerende systemen met een geringe hydraulische verblijftijd zoals de Schuitenbeek minder geschikt omdat piekafvoeren kunnen worden gemist. De modelresultaten sluiten dan ook veel beter aan bij de continue metingen, omdat die

op hetzelfde temporele aggregatieniveau betrekking hebben.

Stikstof

Het temporele patroon in anorganisch stikstof wordt het beste gereproduceerd en voorspeld op de benedenstroomse locaties. Dat geldt met name voor het continue meetpunt. Het patroon wordt over het algemeen beter voorspeld in het winterhalfjaar dan in het zomerhalfjaar. De locaties waar het patroon minder goed wordt voorspeld liggen met name meer bovenstrooms. Het patroon in organisch stikstof is daarentegen veel lastiger te voorspellen. Slechts op een beperkt aantal meetlocaties is sprake van een significant zwak verband.

De absolute stikstofconcentraties worden nog niet goed gereproduceerd en voorspeld. Voor fase 3 is sprake van een systematische overschatting van de anorganische en totale stikstofconcentraties. Als naar de afzonderlijke deelfases wordt gekeken dan blijkt dat dit vooral wordt veroorzaakt door een vermindering van de denitrificatie waardoor meer anorganisch stikstof beschikbaar is voor uitspoeling naar het oppervlaktewater.

Hoewel opeenvolgende fases vooruitgang laten zien in het modelleren van fosfor, geldt over het algemeen dat het temporele patroon en de absolute concentraties nog niet goed door de modellen worden beschreven. Voor het temporele patroon geldt dat de meeste verbanden niet significant zijn. Als sprake is van een significant verband dan is dat verband zwak.

Fosfor

De modelsystemen zijn momenteel ook nog niet in staat om absolute fosforconcentraties in de Schuitenbeek te beschrijven en te voorspellen. Dat geldt voor zowel puntmetingen als weekgemiddelde debietproportionele metingen. De fase 1 en fase 2 systemen laten een systematische overschatting zien, het fase 3 systeem een systematische onderschatting.

Uit dit rapport blijkt dat een verbetering van de modellering van de hydrologie (waterafvoer) niet noodzakelijkerwijs hoeft te leiden tot een verbetering van de modellering van de nutriëntenconcentraties. Hydrologie en chemie verdienen elk afzonderlijk aandacht, maar zeker ook in relatie tot elkaar. Wat voor de hydrologie en chemie geldt, geldt ook voor het landsysteem en het oppervlaktewatersysteem. Het oppervlaktewatermodel kan pas goed worden gedraaid als er geen grote onvolkomenheden meer in het landsysteem zitten. Pas als de modelleur daar vertrouwen in heeft kan op basis van een gevoeligheidsanalyse en expertkennis de modelprestatie verder worden verbeterd door middel van kalibratie.

Conclusies

In het kader van het project ‘Monitoring Stroomgebieden’ wordt een model-systeem 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 ook nog niet goed mogelijk om uitspraken te doen of kritische concentraties al dan niet worden overschreden. Voor stikstof 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 moeten de modellen ook in staat zijn om absolute concentraties te modelleren. Het modelsysteem zegt dan niets over de absolute grootte van de stikstofconcentratie, maar wel of er sprake is van een verandering van de concentratie. De modelsystemen presteren beter in het reproduceren en voorspellen van de waterafvoer. Dat geldt voor zowel het temporele patroon als voor de absolute waarden.

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 Schuitenbeek zijn ontwikkeld te evalueren ten behoeve van monitoring. Dit zal worden gedaan op basis van validatie. De modelsystemen en modelresultaten zijn uitgebreid beschreven in Jansen et al. (2006a,b, 2009).

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 Schuitenbeek 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 Jansen et al. (2006a,b, 2009). 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 meemodeleren 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.*

	hoofdphase	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 leggergegevens
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 Veluwe 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 typefouten tot fouten in de nutriëntenbalansen. Aangetroffen fouten worden gerapporteerd als webpagina en gecommuniceerd met Waterschap Veluwe. 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 noodzakelijk om eerst de hoeveelheid anorganisch stikstof uit te rekenen op basis van de gemeten stikstofcomponenten.

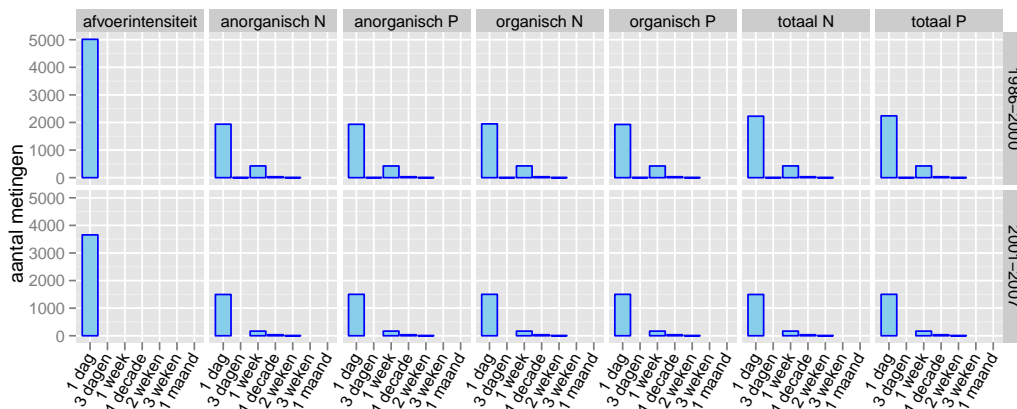
¹Voor de opslag van een terabyte aan gegevens zijn circa 1500 compact discs nodig.

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

Tabel 2.2: Aggregatie van modelresultaten die betrekking hebben op het oppervlaktewaterstelsel (- = 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)	-	+	+	+	+	+	+

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 deel van de meetgegevens is door de modelleers 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

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.

modelsysteem de meetgegevens kan *beschrijven*. Ze is echter ongeschikt om het *voorspellend* vermogen van het model te kwantificeren. Immers, omdat de modelleers 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 modelstelsel 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).

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 nitraatconcentratie in een beek). De grijze lijn stelt de werkelijke waarde voor. Deze

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
modefficiëntie	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 modefficiëntie (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

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 geparameteriseerd. 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 kwelflux, 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.

Modefficiëntie

De modefficiëntie (bijlage A.8) is gerelateerd aan de totale fout (zie vergelijking A.14). De modefficiëntie 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 modefficiënties 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 modefficiëntie zegt niets over het patroon in de metingen. Ook kan de modefficiëntie geen onderscheid maken tussen de systematische en willekeurige fout. Een lage modefficiëntie 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 zonnepplot.

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.

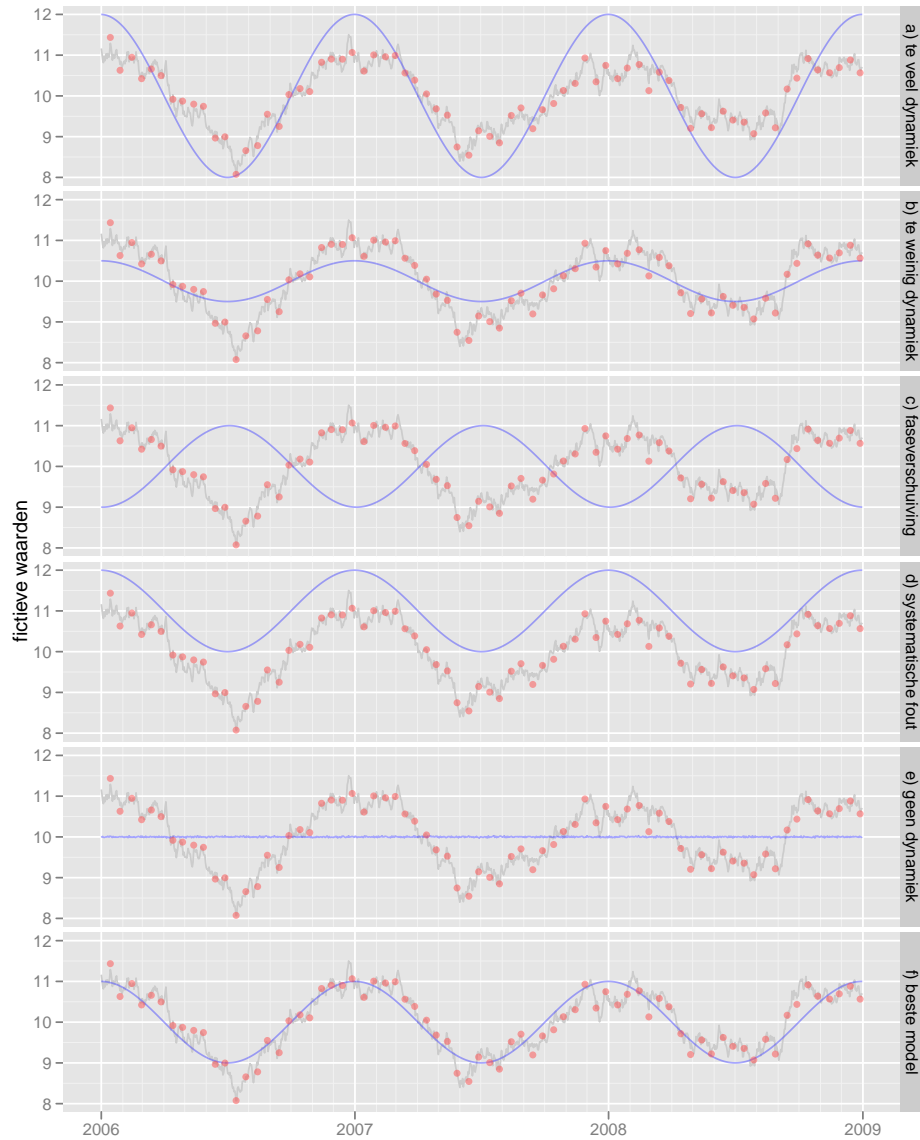
De modelresultaten kunnen worden geïnterpreteerd op basis van hun ligging ten opzichte van de kern van de zon.

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.

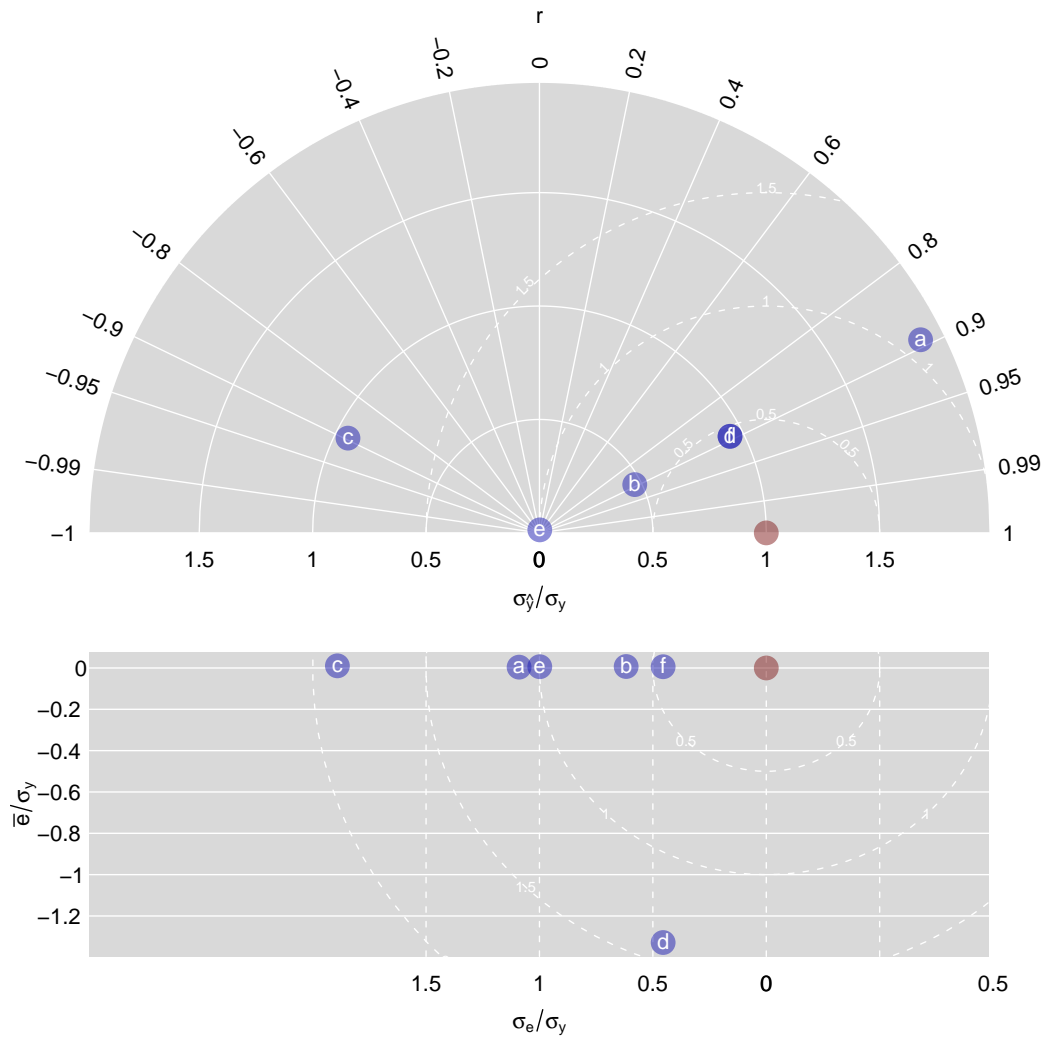
Drie modellen liggen buiten de zon. Dat zijn modellen **a**, **c** en **d**. Modellen die buiten de zon liggen hebben geen meerwaarde boven het gebruik van de gemiddelde meetwaarde als predictor. De oorzaak hiervan voor model **d** was al duidelijk. Hier is sprake van een systematische overschatting 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 (uitgebreide) 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

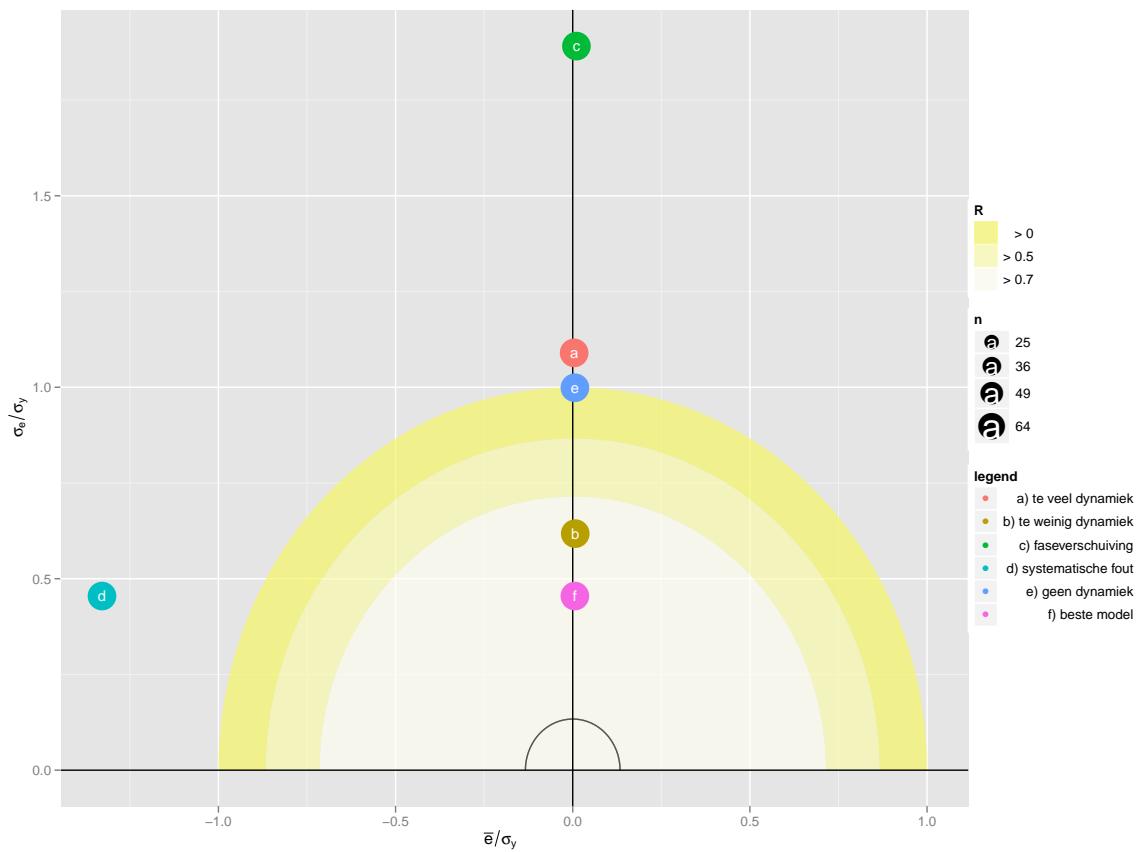
variatie in de meetgegevens te reproduceren.



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.



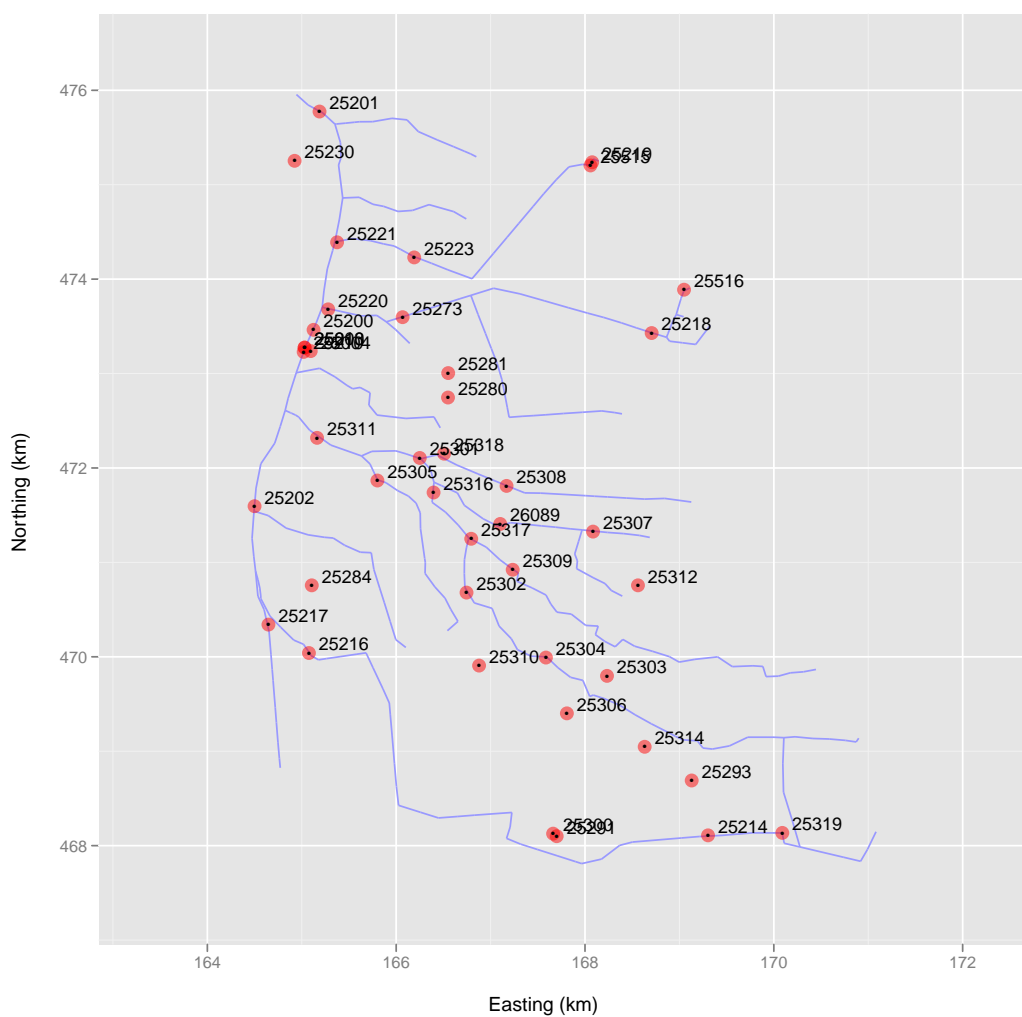
Figuur 2.3: Taylor diagram (boven) gekoppeld aan $\bar{e} - \sigma_e$ plot (onder). De letters a tot en met f verwijzen naar de tijdreeksen in figuur 2.2. Merk op dat d en f samenvallen in de bovenste helft van de figuur. Voor de wijze waarop deze figuur is geconstrueerd en moet worden geïnterpreteerd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

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

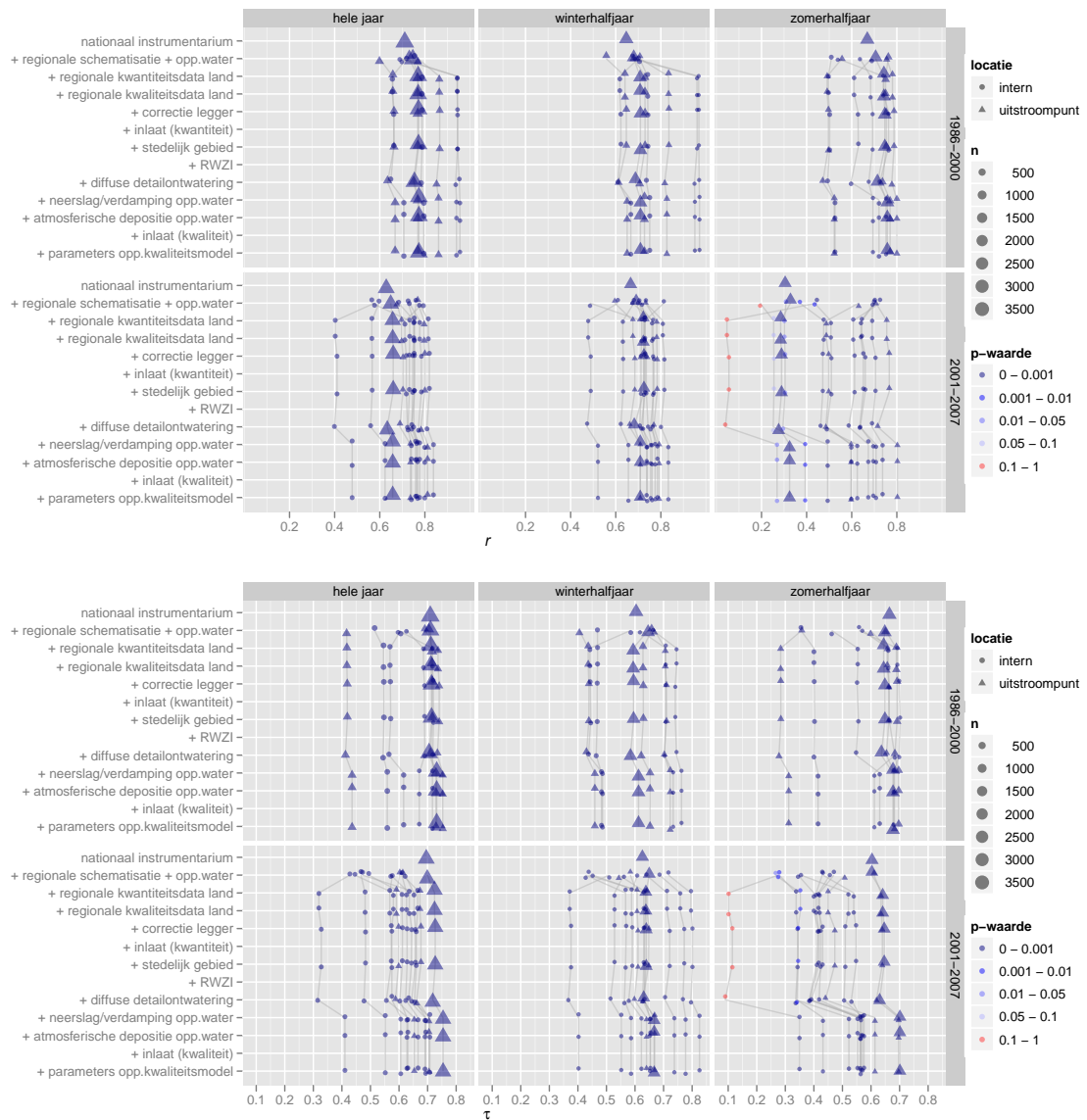
In figuur 3.2 is Pearsons correlatiecoëfficiënt (r) weergegeven (horizontale as) voor de opeenvolgende deelfases (verticale as). Pearsons correlatiecoëfficiënt geeft aan hoe goed de modelsystemen in staat zijn om temporele patronen te reproduceren. De bovenste drie panelen slaan op de periode 1986-2000, 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. Dit wordt deels veroorzaakt door de aanwezigheid van extreme waarden in de eerstgenoemde periode¹. Om het effect van extreme waarden te reduceren is ook Kendalls correlatiecoëfficiënt (τ) berekend (figuur 3.2, onder, zie ook paragraaf 2.3.5). Het verschil tussen de beide perioden is nu minder geworden. Verder blijkt dat in de validatieperiode (2001-2007) het patroon in de metingen beter wordt gemodelleerd in het winterhalfjaar dan in het zomerhalfjaar (met name voor het continue meetpunt (meetpunt 25210, figuur 3.1)). Dat is ook te verwachten omdat in de zomer extreme afvoerpieken worden afgewisseld door periodes met droogval. Dergelijke dynamische schommelingen zijn over het algemeen lastig te modelleren. Het grootste deel van de meetlocaties heeft een Pearsons correlatiecoëfficiënt (r) tussen de 0.60 en 0.95. Dat wil zeggen dat 36% tot 90% van de variatie in de metingen door het model kan worden verklaard. Verder valt op dat met name de deelfase waarin het effect van neerslag en verdamping op het oppervlaktewater wordt meegenomen een positieve bijdrage heeft op de modellering van het temporele patroon.

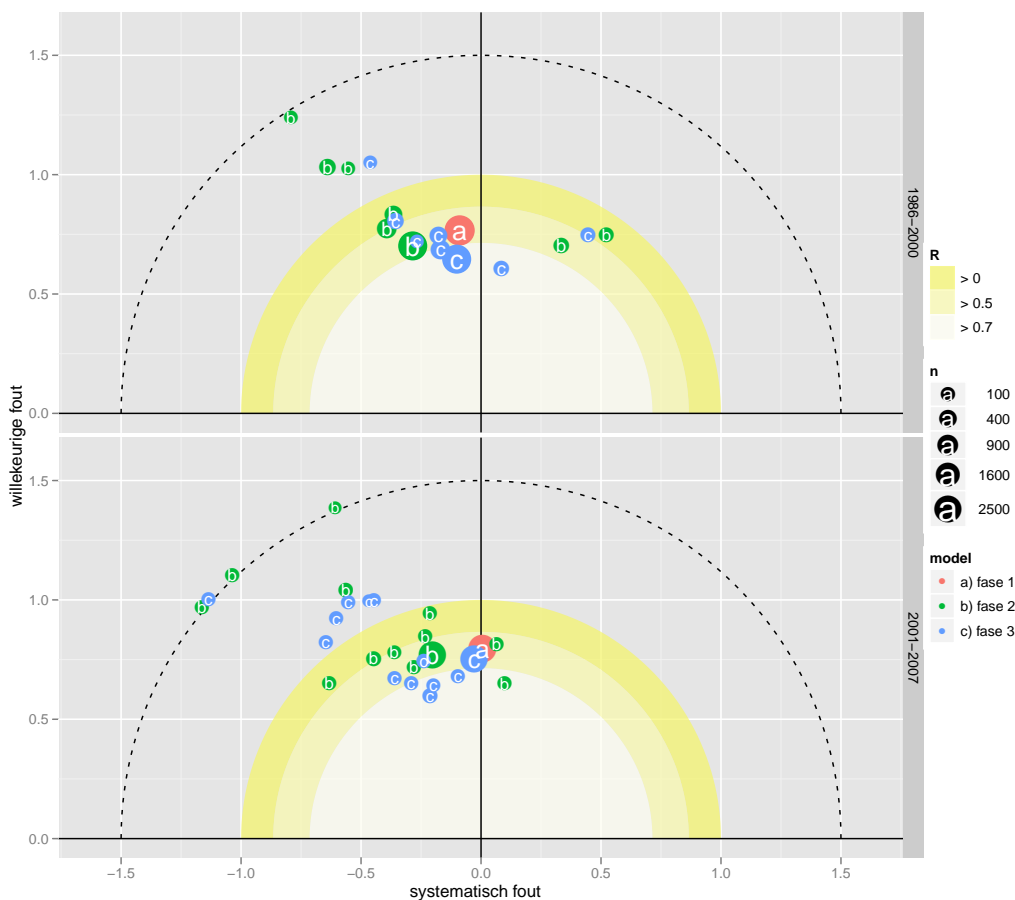
In figuur 3.3 is de modelprestatie gevisualiseerd met behulp van een zonnepplot (paragraaf 2.3.6, bijlage C). Uit de figuur blijkt dat voor zowel de periode 1986-2000 als 2001-2007 de waterafvoer voor de meeste locaties wordt overschat (zie ook Jansen et al. (2009, p. 105)). Fase 3 lijkt het over het algemeen iets beter te doen dan voorgaande fases. Opvallend is dat in de validatieperiode (2001-2007) de systematische fout kleiner is geworden en voor fase 1 en 3 zelfs nagenoeg is verdwenen (zie ook figuur 3.4). Dit komt doordat het gemodelleerde netto neerslagoverschot lager is in de tweede periode². Hierdoor zal minder water

¹De zogenaamde ‘hefboomwerking’ (leverage) van extreme waarden leidt tot hogere correlaties.

²bron: variantanalyse 2008-11-18



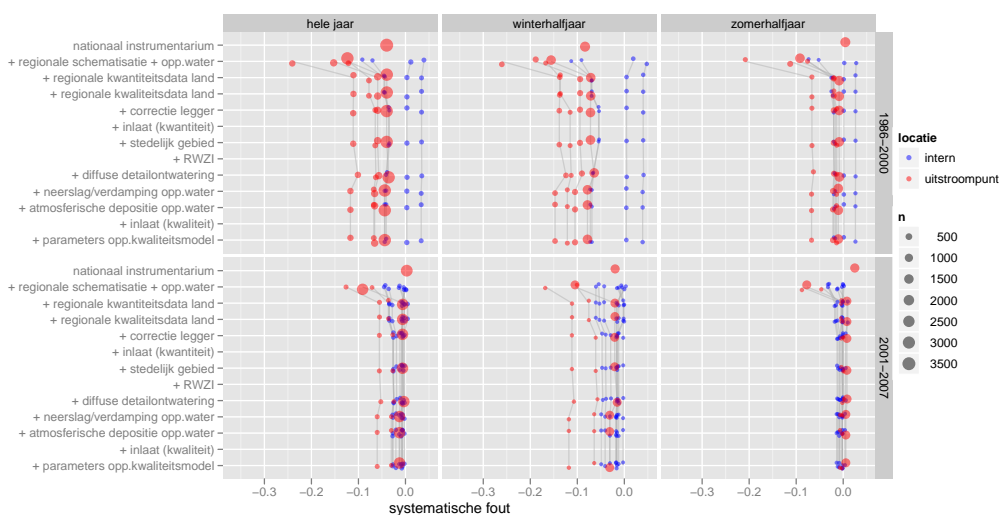
Figuur 3.2: Pearsons (r) en Kendalls correlatiecoëfficiënt (τ) tussen de gemeten en gesimuleerde waterafvoer ($\text{m}^3 \text{s}^{-1}$) op alle locaties in het stroomgebied van 'de Schuitenbeek' voor het hele jaar (kolom 1), het winterhalfjaar (kolom 2) het zomerhalfjaar (kolom 3) en twee perioden (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).

worden afgevoerd van het landsysteem naar het oppervlaktewatersysteem.

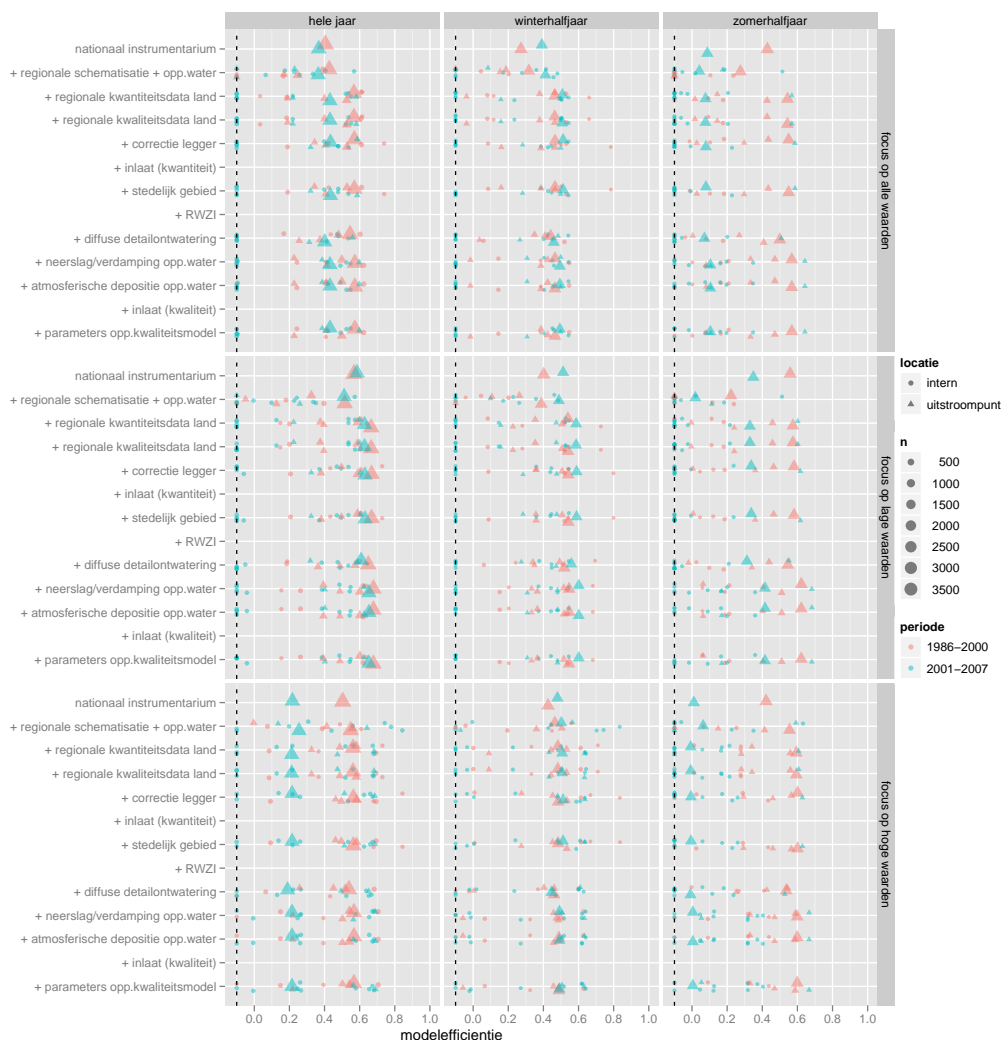
De afstand van een punt tot de oorsprong van de zonneplot is gelijk aan de modefficiëntie (vergelijking A.13). Deze is gegeven als functie van de deelfase, de periode, het locatietype, het aantal waarnemingen en het seizoen in de bovenste rij van figuur 3.5. De waterafvoer wordt in fase 3 over het algemeen beter voorspeld dan in voorgaande fases. Dit is met name het geval voor de periode 1986-2000. Uit figuur 3.4 en 3.5 blijkt dat dit vooral wordt veroorzaakt door de eerste deelfase van fase 3, te weten deelfase ‘regionale kwantiteitsdata land’. Deze deelfase omvat het gebiedspecifiek maken van de neerslagverdeling, de onderrand (kwel/wegzijging), de drainagekarakteristieken en de pakketdikte van het topsysteem (Jansen et al., 2009, pag. 30–33). Voor de validatieperiode zijn de verschillen minder groot. Ook blijkt uit deze figuur dat het model beter presteert in de periode tot 2001 dan in de periode daarna. Dit verschil lijkt te worden veroorzaakt door het zomerhalfjaar (figuur 3.5, derde kolom). Dat wordt duidelijk wanneer ook speciale varianten van de modefficiëntie worden berekend die kwantificeren in hoeverre een model in staat is om lage



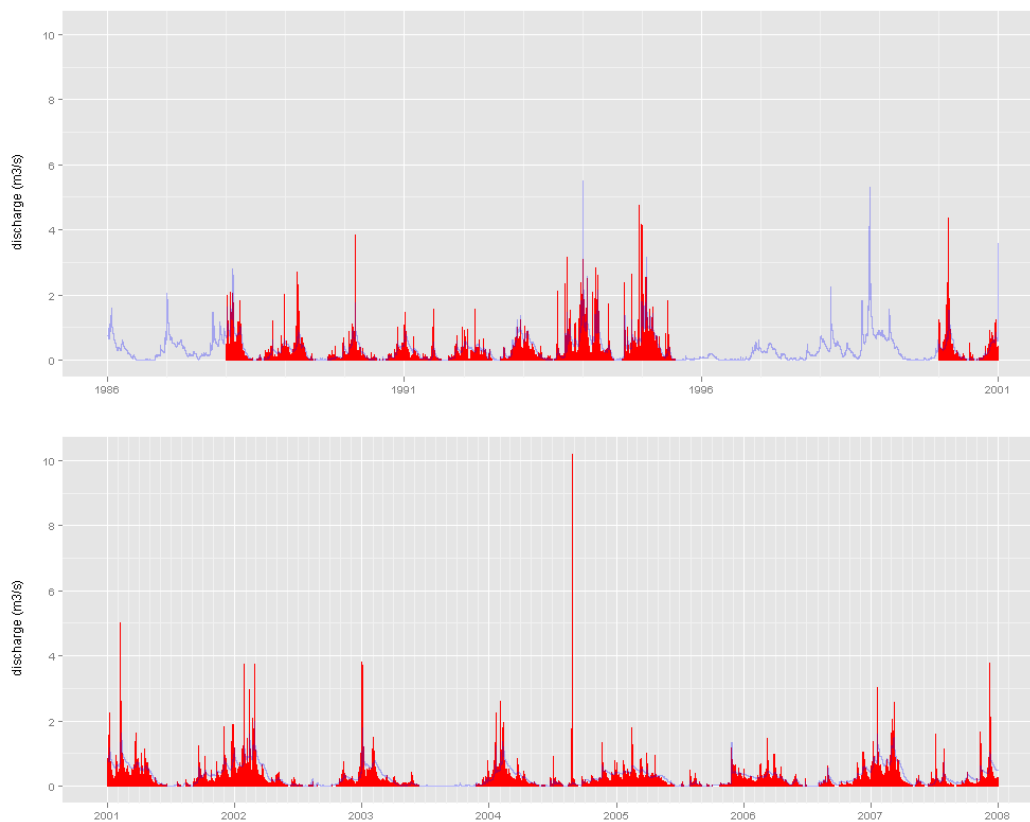
Figuur 3.4: Systematische fout ($\text{m}^3 \text{s}^{-1}$) m.b.t. de waterafvoer voor de afzonderlijke deelfases.

afvoeren te voorspellen (middelste rij) en hoge afvoeren (onderste rij). Dit zijn respectievelijk de modefficiëntie van Perrin en Guex (Gebremeskel et al., 2005). Uit de figuur blijkt dat het model moeite heeft om de afvoerpieken in de zomers na 2001 goed te voorspellen (zie subfiguur in de derde kolom en de onderste rij). In figuur 3.6 zijn de gemeten en gesimuleerde tijdreeksen voor beide perioden gegeven bij het continue meetpunt 25210. Opvallend is de hoge afvoerpiek in de zomer van 2004 die door het model niet goed wordt gesimuleerd. Verwijdering van deze piek levert echter maar een geringe verbetering van de modefficiëntie in de zomer op (niet weergegeven). Dit impliceert dat het model ook moeite heeft om de andere, minder extreme, zomerafvoerpieken goed te simuleren. Het zomerhalfjaar, dat veelal gekenmerkt wordt door droogval en extreme afvoeren blijkt lastig te modelleren. De simulatieresultaten zijn beter in het winterhalfjaar.

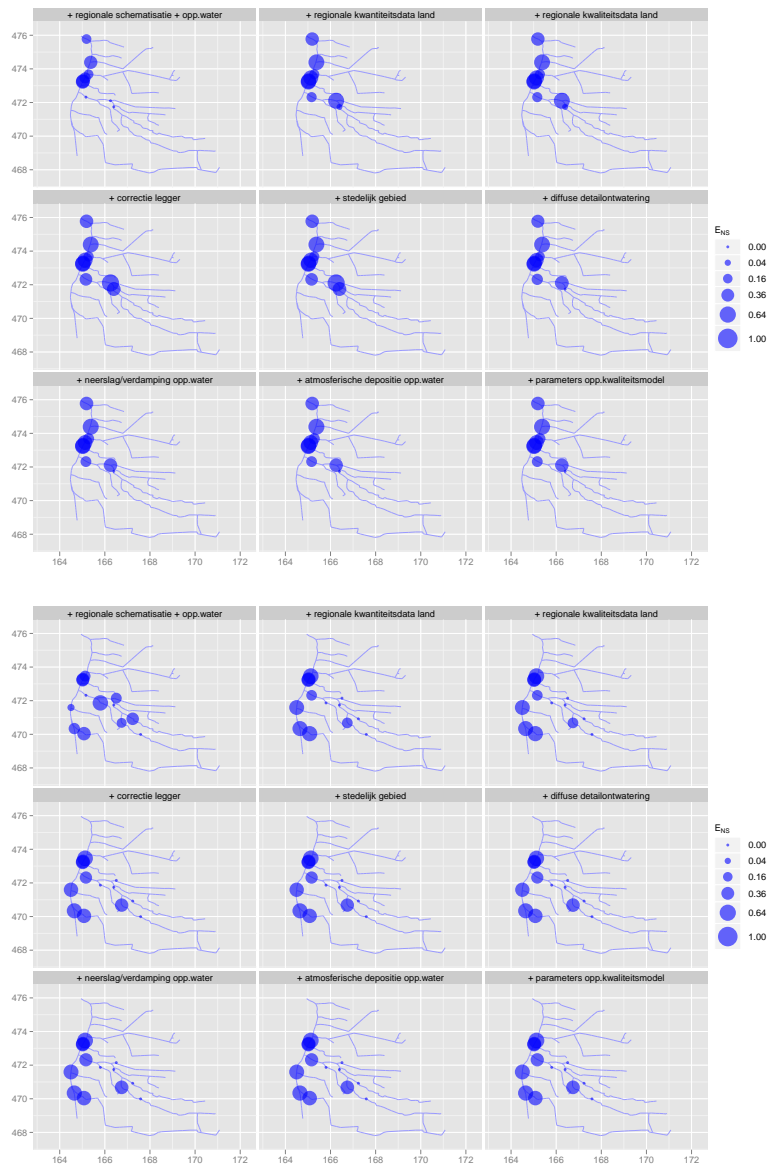
De ruimtelijke verdeling van de modefficiëntie is gegeven in figuur 3.7. Benedenstreamse meetpunten hebben over het algemeen een hogere modefficiëntie dan meer bovenstrooms gelegen meetpunten ten gevolge van uitmiddeling. Opvallende uitzondering is meetlocatie 25201, gelegen bij het uitstroompunt van de Schuitenbeek (figuur 3.1). De afvoerintensiteit op deze locatie is bepaald met een OTT-molen. Als de afvoerintensiteit op locatie 25201 wordt uitgezet tegen de afvoerintensiteit op het continue meetpunt 25210, dat iets verder stroomopwaarts ligt (figuur 3.1), dan valt op dat bij geringe afvoerintensiteiten de OTT-molen de afvoerintensiteit sterk onderschat (figuur 3.8, links). Dit wordt veroorzaakt door de metingen die in het winterhalfjaar van 1988 en 1989 zijn uitgevoerd (figuur 3.9). De reden van deze systematische onderschatting is onbekend. Na 1989 treedt dit effect vrijwel niet meer op en is het verband beter.



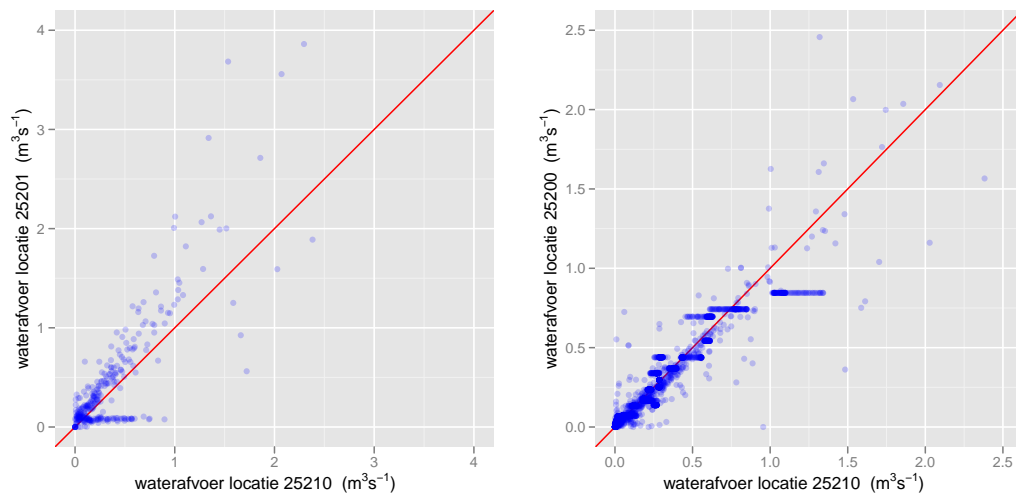
Figuur 3.5: Modefficiëntie m.b.t. de waterafvoer ($\text{m}^3 \text{s}^{-1}$) voor de afzonderlijke deelfases (zie ook tabel 2.1). Bovenste rij: modefficiëntie met focus op alle waarden; middelste rij: modefficiëntie met focus op lage waarden; onderste rij: modefficiëntie met focus op hoge waarden. Terwille van de leesbaarheid zijn alle modefficiënties kleiner dan -0.1 op de zwarte stippellijn geplaatst.



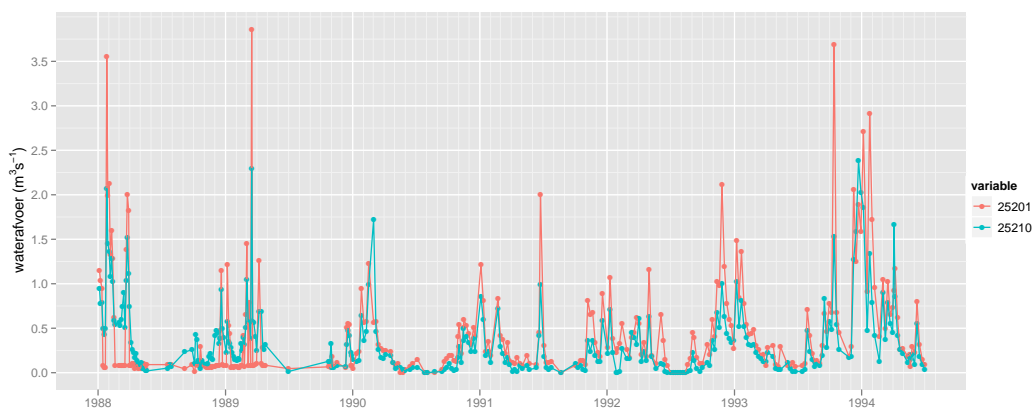
Figuur 3.6: Gemeten (rood) en berekende (blauw) afvoeren bij het continue meetpunt 25210 voor fase 3 voor de periode 1986-2000 en 2001-2007



Figuur 3.7: Modefficiëntie (-) 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 3.8: Waterafvoer ($\text{m}^3 \text{s}^{-1}$) op locaties 25201 en 25200 (OTT-molenmetingen) versus die op locatie 25210 (continue meting).



Figuur 3.9: Tijdreeksen van de waterafvoer ($\text{m}^3 \text{s}^{-1}$) op locaties 25201 (OTT-molenmetingen) en 25210 (continue meting).

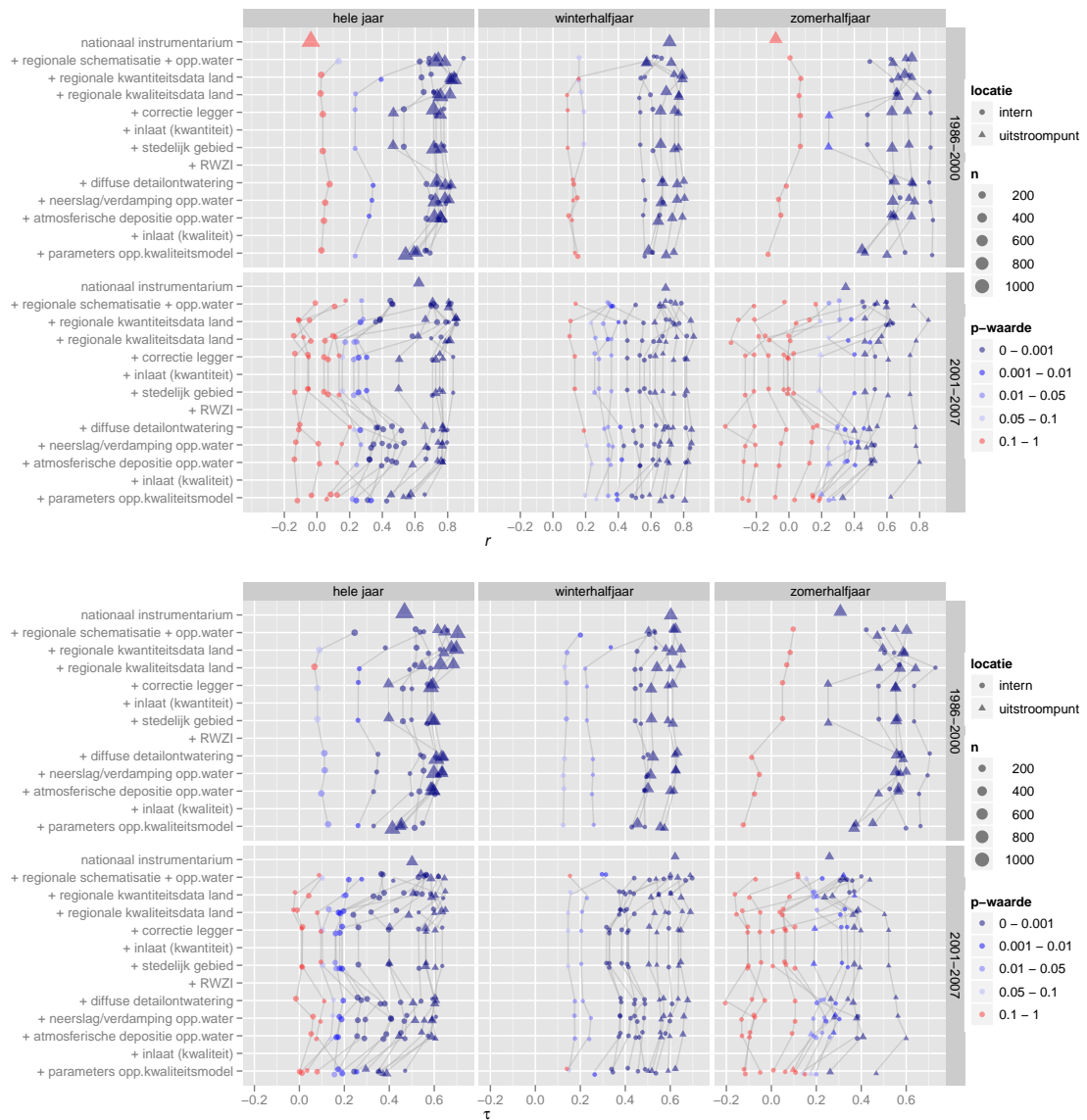
Benadrukt moet daarom weer worden dat de validatiematen niet alleen de betrouwbaarheid van het model kwantificeert, maar dat daar tevens de betrouwbaarheid van de metingen in zit verdisconteerd. Slechte validatieresultaten worden dus niet alleen bepaald door de kwaliteit van het model, maar tevens door die van de metingen. In de vorige paragraaf werd hiervan al een voorbeeld gegeven. Een ander voorbeeld is meetlocatie 25200, dat ruim 200 m stroomafwaarts ligt van het continue meetpunt 25210. Figuur 3.8 (rechts) geeft het verband weer tussen dit punt waar de afvoerintensiteit is bepaald met OTT-molenmetingen en het continue meetpunt. Opvallend is het discrete karakter van de OTT-metingen (zie horizontale structuren), terwijl de continue metingen en de modelresultaten continu van aard zijn. Het discrete karakter duidt op beperkingen in de resolutie waarmee de waarnemer de stroomsnelheid kan bepalen. Deze meetfout komt ook tot uiting in de validatiematen.

De modellen lijken beter te presteren op de continue meetlocatie (grootste symbolen in eerder genoemde figuren) dan op de overige meetlocaties waar met een OTT-molen wordt gemeten. Eén oorzaak daarvan is hierboven al genoemd, namelijk het voorkomen van meetfouten. Een andere oorzaak is het verschil in temporele aggregatie tussen de metingen en de modelresultaten. De OTT-molenmetingen zijn wekelijkse momentopnames. Gezien de geringe hydraulische verblijftijd van het Schuivenbeekstelsel kunnen afvoerpieken daarom (deels) worden gemist, met name in het zomerhalfjaar. De continue metingen daarentegen zijn daggemiddelden (berekend op basis van kwartiermetingen) en hebben hetzelfde temporele aggregatieniveau als de modelresultaten, waardoor de validatie een stuk netter is (zie paragraaf 2.2.3).

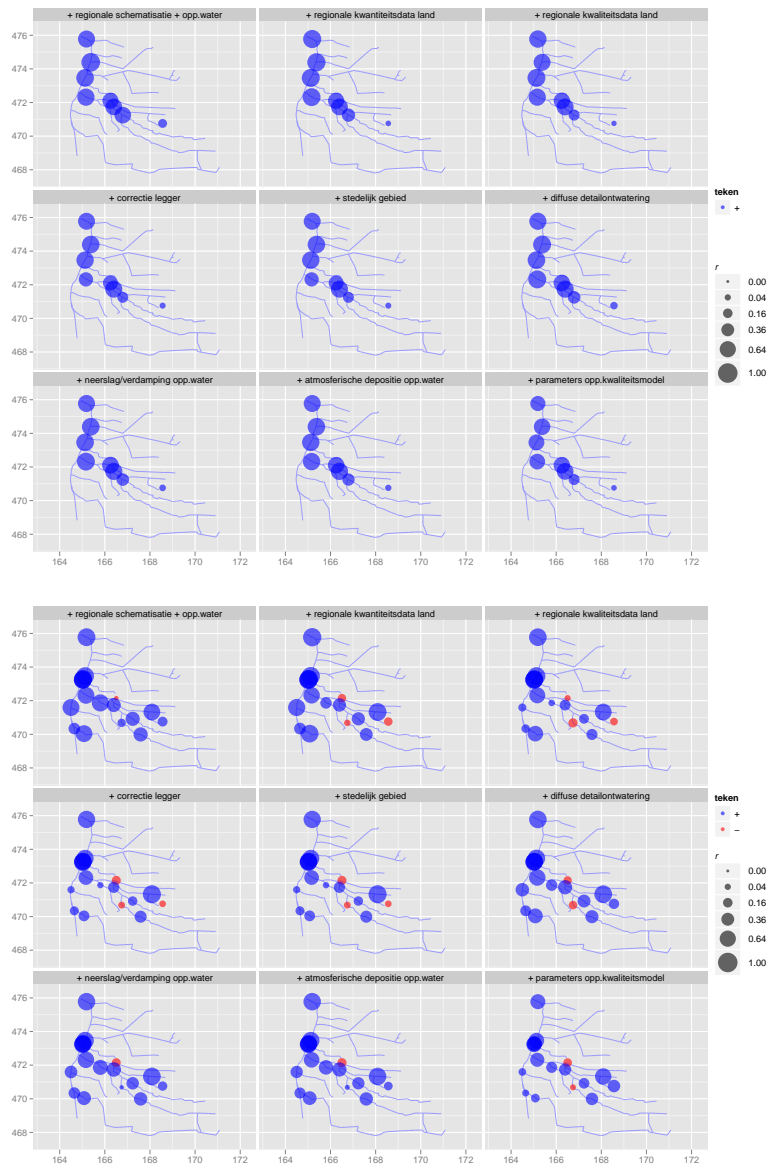
3.2 Stikstof

3.2.1 Anorganisch stikstof

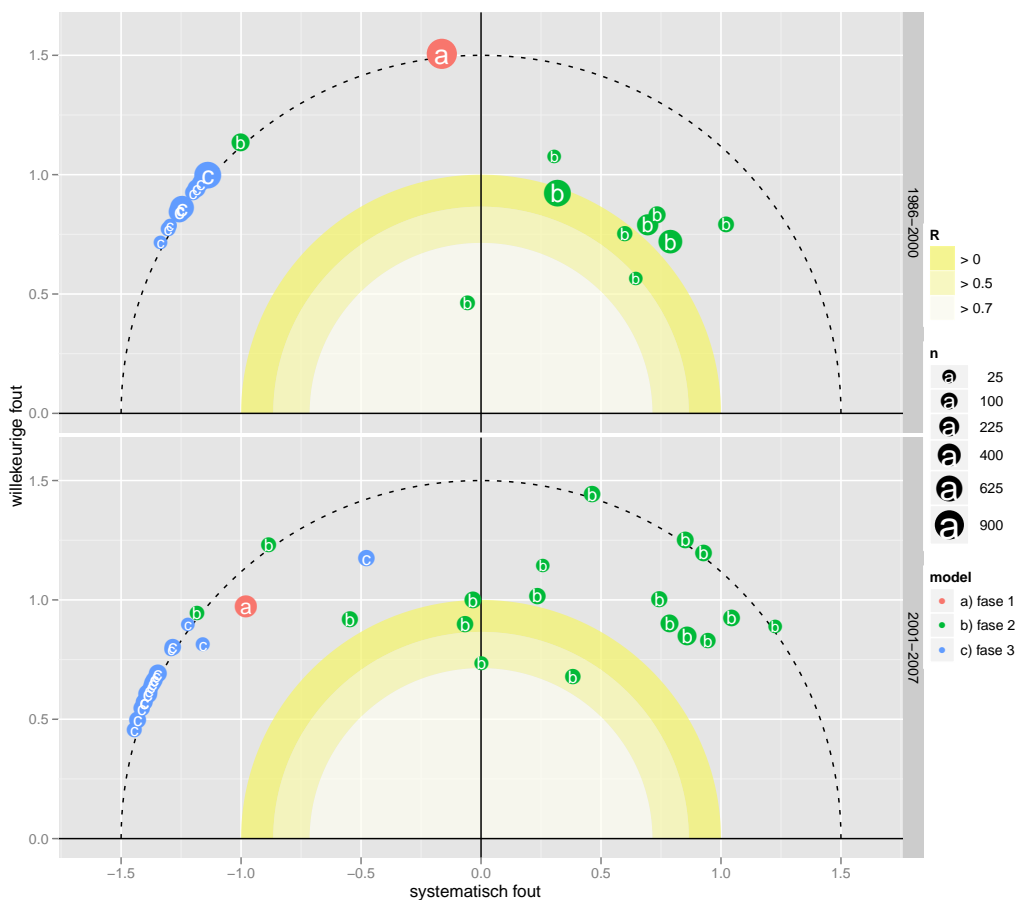
Hoe goed de modellen in staat zijn om het temporele patroon van de gemeten anorganische stikstofconcentraties te reproduceren en te voorspellen, blijkt uit figuur 3.10. Het temporele patroon wordt het beste beschreven op de benedenstroomse locaties (uitstroompunten) en dan met name in het winterhalfjaar. Voor een aantal locaties geldt dat in het zomerhalfjaar geen significant verband is in temporeel patroon tussen de modelresultaten en metingen. Deze locaties liggen bovenstrooms (figuur 3.11). Opvallend is dat fase 1 (nationale modelinstrumentarium) geen significant verband vertoont met de metingen. Dit wordt voornamelijk veroorzaakt door twee extreme waarden in de modelresultaten voor het zomerhalfjaar (niet weergegeven) die het verband sterk reduceren. Daarom is in figuur 3.10 (onder) ook Kendalls correlatiecoëfficiënt (τ) gegeven. Deze is minder gevoelig voor extreme waarden en geeft een genuanceerder beeld. Het verschil in modelprestatie met de overige fases is nu veel kleiner.



Figuur 3.10: Pearsons (r) en Kendalls correlatiecoëfficiënt (τ) tussen de gemeten en gesimuleerde anorganische stikstofconcentratie (mg l^{-1}) op alle locaties in het stroomgebied van 'de Schuitenbeek' 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.



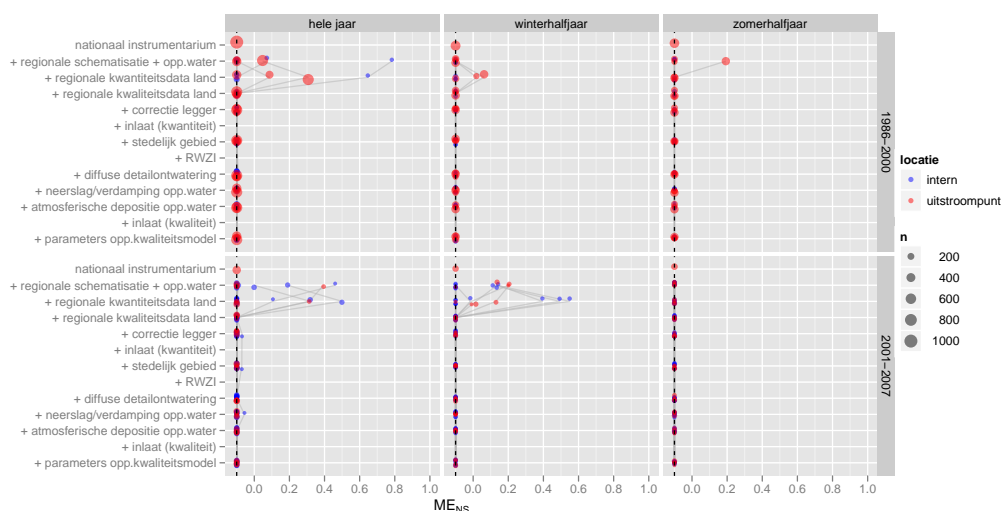
Figuur 3.11: Pearsons correlatiecoëfficiënt m.b.t. de anorganisch stikstof (mg l^{-1}) op dagbasis voor de periode 1986-2000 (boven) en 2001-2007 (onder) voor opeenvolgende deelfases.



Figuur 3.12: Anorganisch stikstof (mg l^{-1}) op dagbasis voor de periode 1986-2000 (boven) en 2001-2007 (onder).

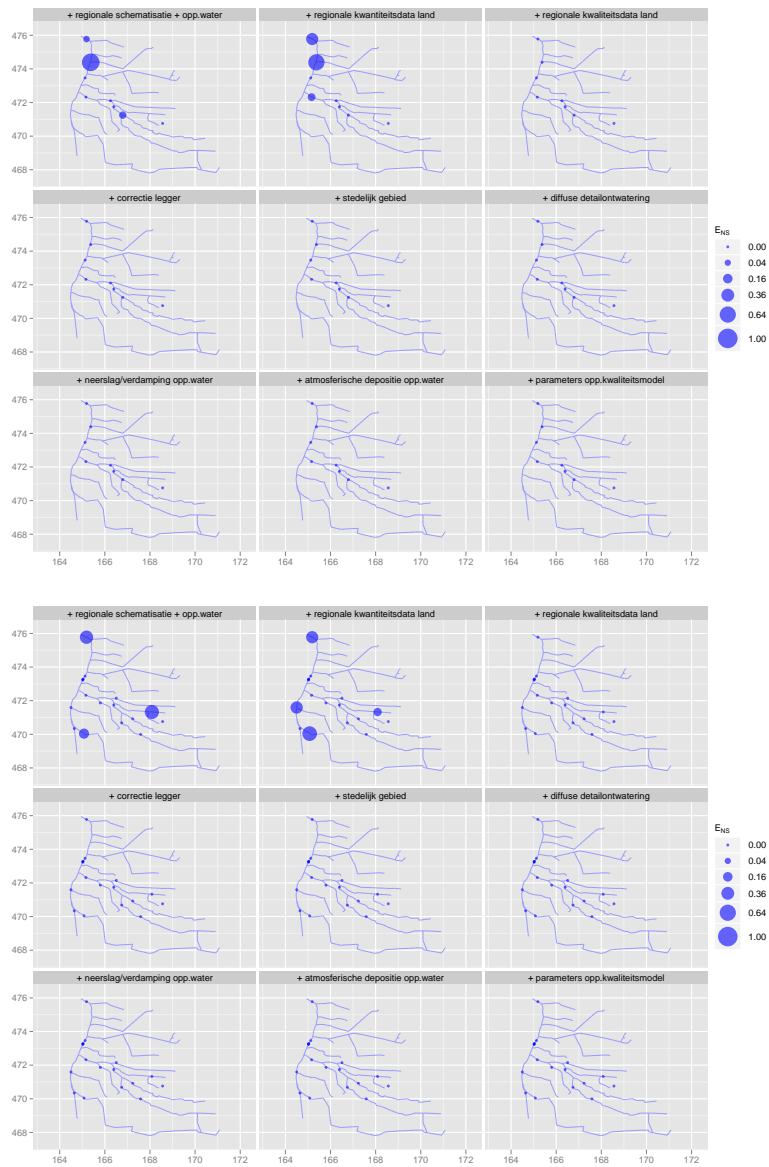
Figuur 3.12 geeft de zonneplot voor de hoofdfases met betrekking tot anorganisch stikstof weer. Deze figuur geeft aan hoe goed de modellen die voor de hoofdfases zijn opgesteld de metingen reproduceren en voorspellen. Alleen in fase 2 lijkt het modelinstrumentarium op een aantal locaties meerwaarde te hebben t.o.v. het gemiddelde van de metingen. Fase 3 en in minder mate fase 1 overschatten de anorganische stikstofconcentraties, terwijl fase 2 over het algemeen een systematische onderschatting te zien geeft. De reden voor het prestatieverlies wordt duidelijk als naar de afzonderlijke deelfases van fase 3 wordt gekeken (figuur 3.13). In de tweede deelfase van fase 3, te weten ‘regionale kwaliteitsdata land’, is de bodemchemische parameterisering van het landsysteem 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 kritische waarde van de vochtverzadigingsgraad waarboven denitrificatie³ optreedt. Verhoging leidt tot minder denitrificatie, waardoor meer anorganisch

³Denitrificatie 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.

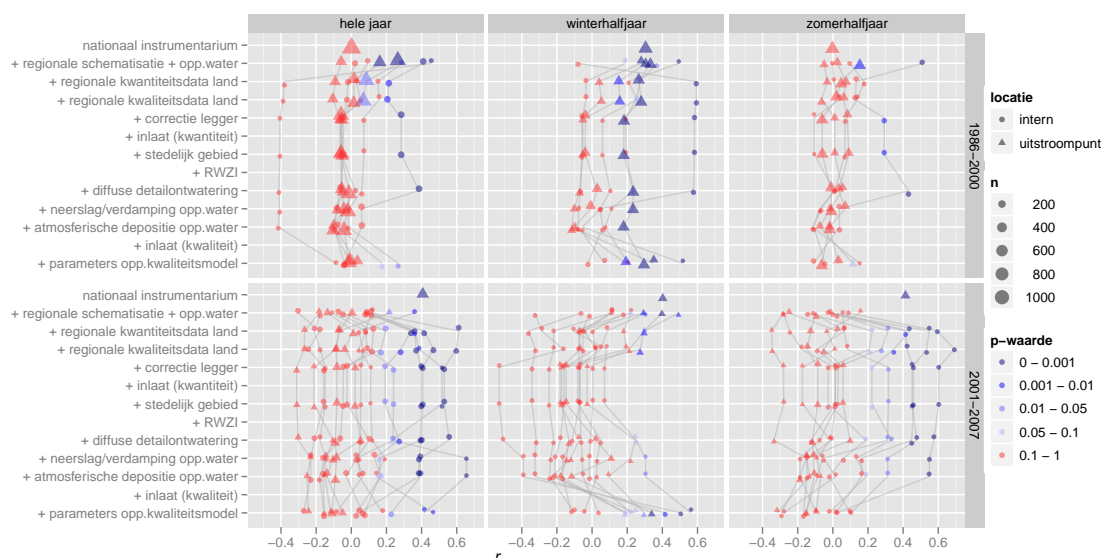


Figuur 3.13: Modelefficiëntie (-) m.b.t. anorganisch stikstof (mg l^{-1}) op dagbasis voor de afzonderlijke deelfases.

stikstof beschikbaar komt voor uitspoeling naar grond- en oppervlaktewater en de systematische fout toeneemt (figuur 3.13). 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. Dit komt ook duidelijk tot uiting in het ruimtelijke patroon (figuur 3.14): vanaf fase ‘regionale kwaliteitsdata land’ heeft het model op geen enkel meetpunt een meerwaarde ten opzichte van het gemiddelde van de metingen.



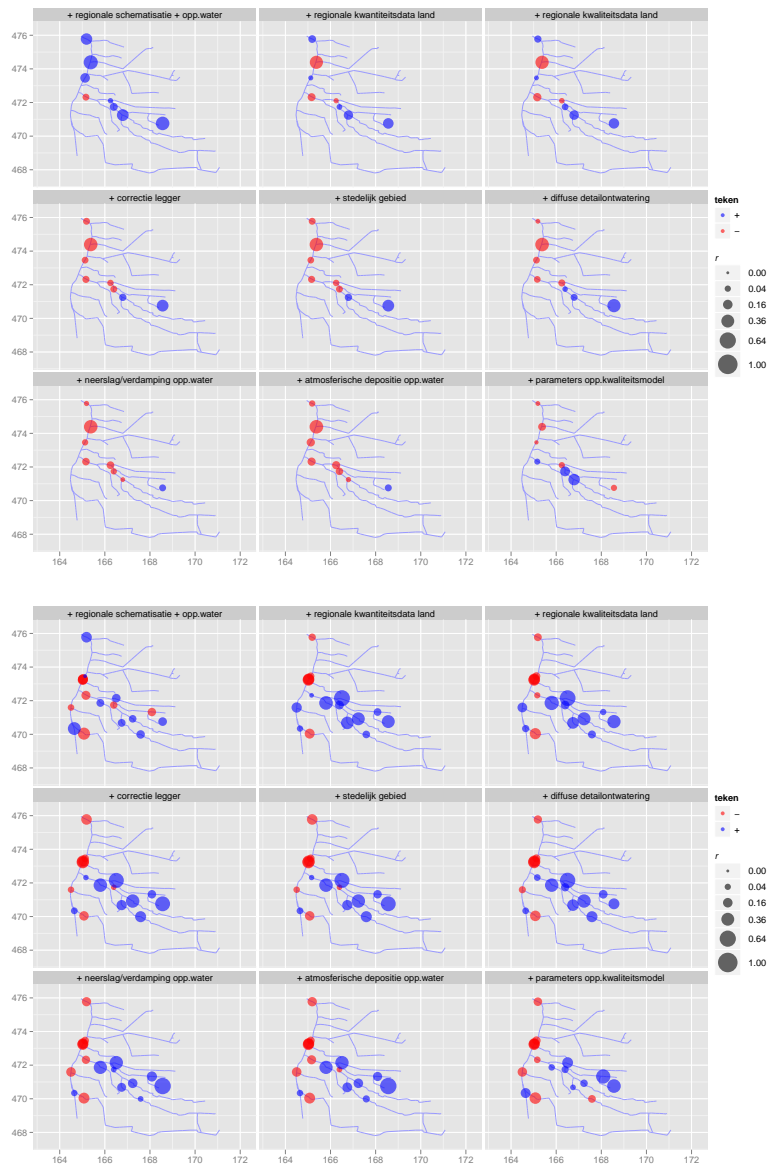
Figuur 3.14: Modefficiëntie (-) m.b.t. de anorganisch stikstof (mg l^{-1}) op dagbasis voor de periode 1986-2000 (boven) en 2001-2007 (onder) voor opeenvolgende (deel)fases.



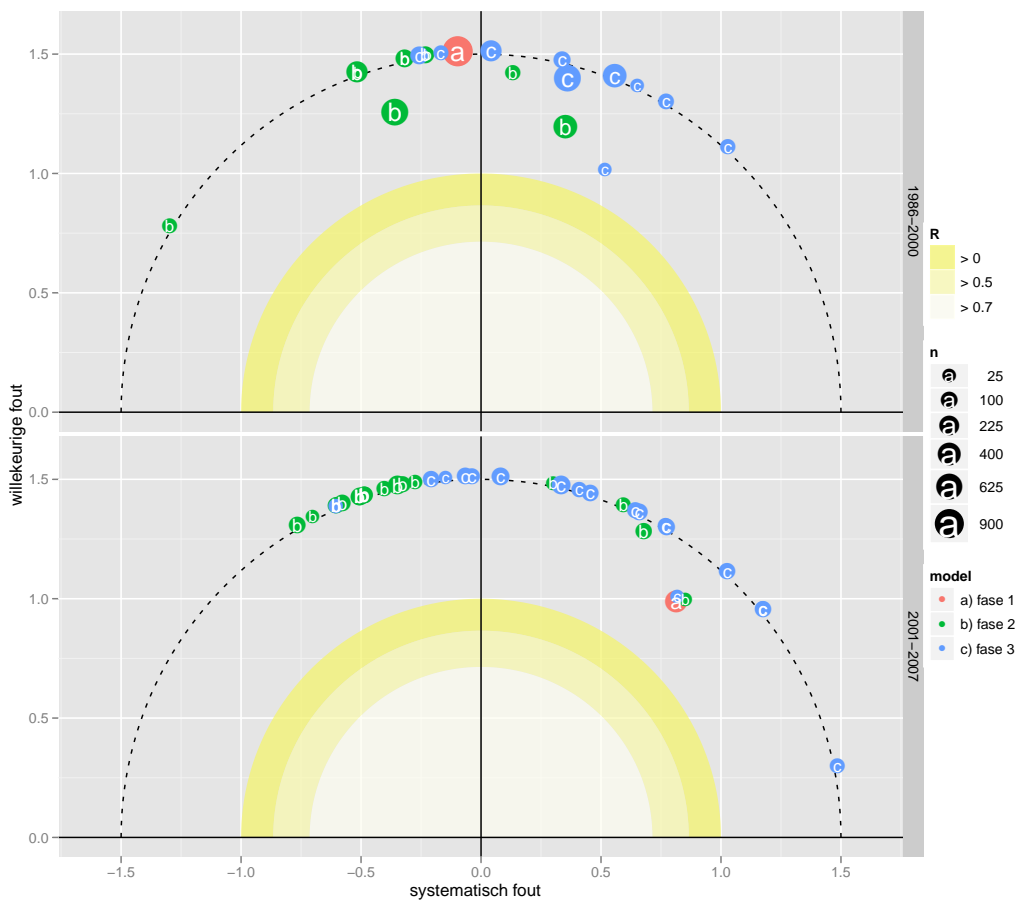
Figuur 3.15: Pearsons (r) correlatiecoëfficiënt tussen de gemeten en gesimuleerde organische stikstofconcentratie (mg l^{-1}) op alle locaties in het stroomgebied van ‘de Schuitenbeek’ voor het hele jaar (kolom 1), het winterhalfjaar (kolom 2) het zomerhalfjaar (kolom 3) en twee perioden (rijen). De 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

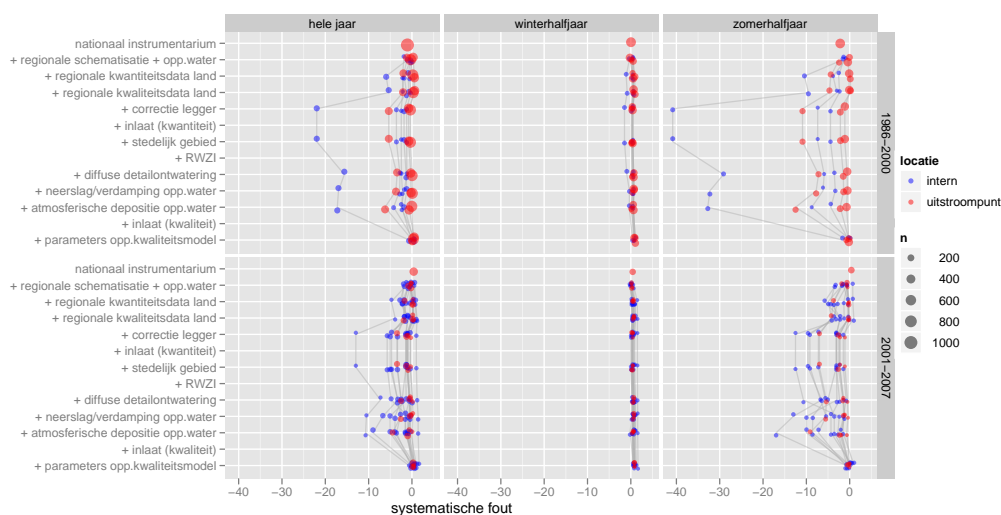
De modellen zijn niet in staat om het temporele patroon in organisch stikstof te reproduceren en te voorspellen (figuur 3.15). Op de uitstroompunten kan geen significant verband tussen de modelresultaten en de metingen worden aangetoond. Alleen op een aantal bovenstroomse punten is sprake van een zwak verband (figuur 3.16). Omdat het temporele patroon niet goed kan worden gemodelleerd zal dat ook gelden voor concentraties op dagbasis (figuur 3.17). Zowel de systematische fout als de willekeurige fout (en daardoor de totale fout) zijn over het algemeen groot. De modellen hebben geen meerwaarde boven het gemiddelde van de metingen. Jansen et al. (2009) verwachten dat de modelresultaten van fase 3 kunnen worden verbeterd door de verblijftijd van organisch stikstof in de bodem te verlagen door verbetering van de hydrologische module van het landsysteem. Er zal daardoor meer organisch stikstof uitspoelen naar het oppervlaktewatersysteem waardoor de systematische fout wordt verkleind. In figuur 3.17 zal dit tot uiting komen door het verschuiven van punten die betrekking hebben op het fase 3 modelsysteem naar links. Echter, uit figuur 3.18 blijkt dat voor de meeste (deel)fases geldt dat sprake is van een aanzienlijke overschatting van de hoeveelheid organisch stikstof. Alleen de laatste deelfase van fase 3, te weten, ‘parameters oppervlaktewaterkwaliteitsmodel’, laat voor een deel van de locaties ook een systematische onderschatting zien. De afname



Figuur 3.16: 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.

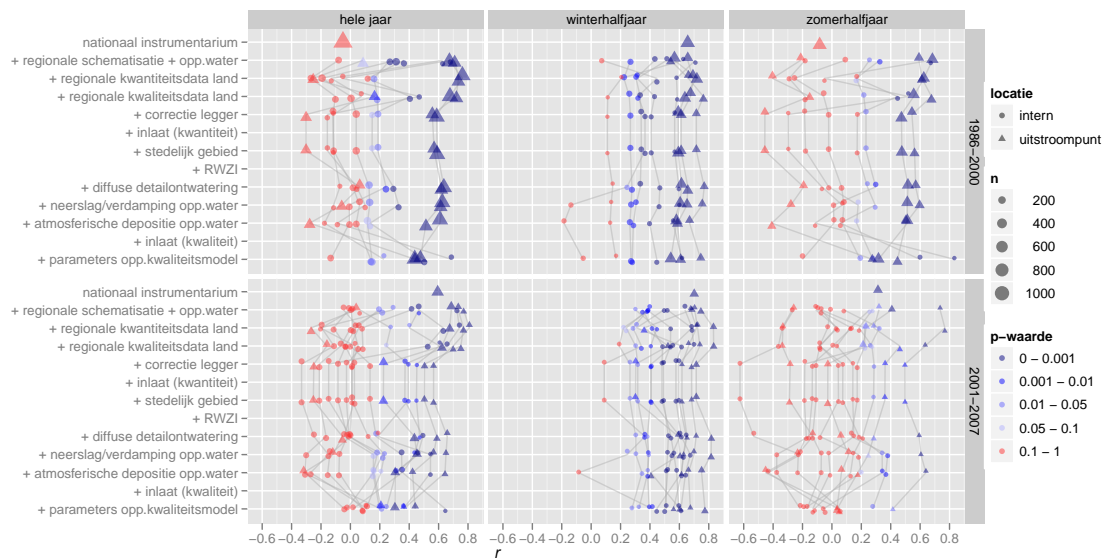


Figuur 3.17: Organisch stikstof (mg l^{-1}) op dagbasis voor de periode 1986-2000 (boven) en 2001-2007 (onder).



Figuur 3.18: Systematische fout (mg l^{-1}) m.b.t. organisch stikstof op dagbasis voor de afzonderlijke deelfases.

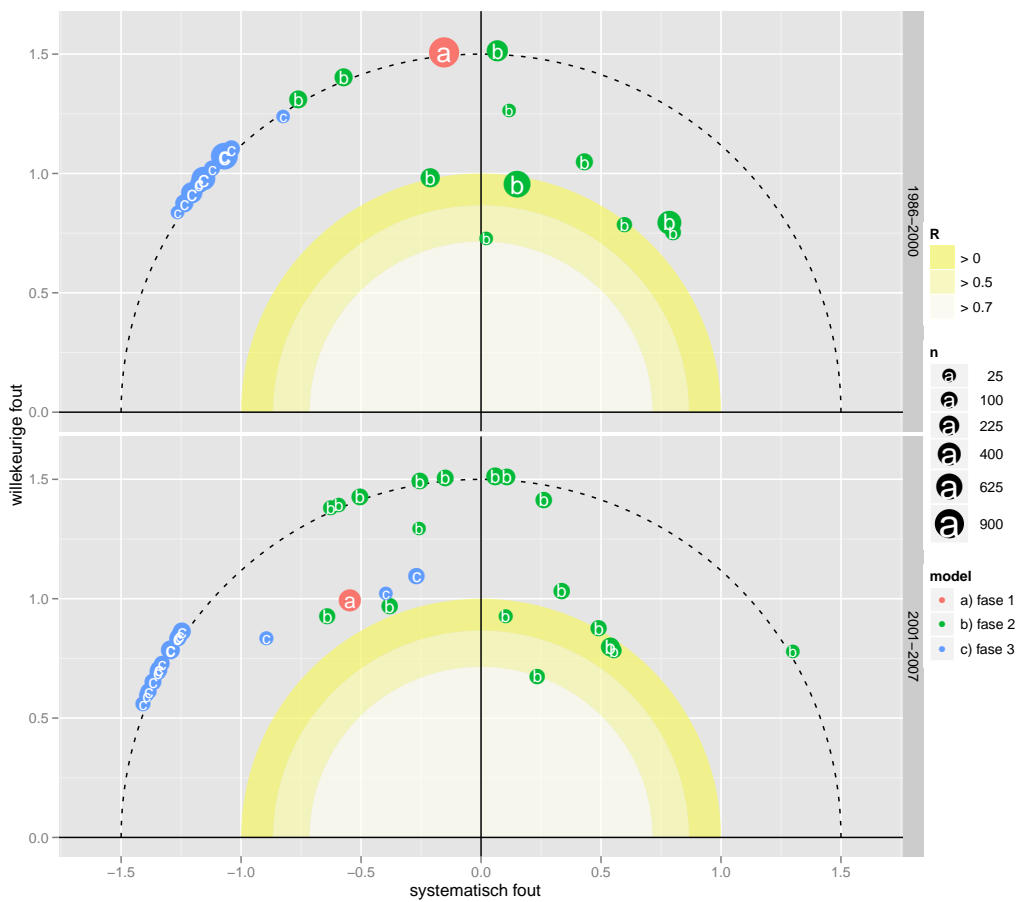
van de systematische fout voor de laatste deelfase van fase 3 is dan ook het gevolg van de aanpassing van de parameters van het oppervlaktewatermodel. De organische stikstofvracht vanuit het landsysteem naar het oppervlaktewaterstelsel is waarschijnlijk nog steeds aan de hoge kant, met name omdat ook de waterafvoer uit het landsysteem al te hoog was berekend (paragraaf 3.1). Om die organische stikstofvracht te reduceren zou ondermeer naar de mineralisatiesnelheid van organischestof in het landsysteem moeten worden gekeken. Die staat mogelijk iets te laag afgesteld.



Figuur 3.19: Pearsons (r) correlatiecoëfficiënt tussen de gemeten en gesimuleerde totale stikstofconcentratie (mg l^{-1}) op alle locaties in het stroomgebied van 'de Schuitenbeek' 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 3.19) als voor de absolute concentraties (figuur 3.20).



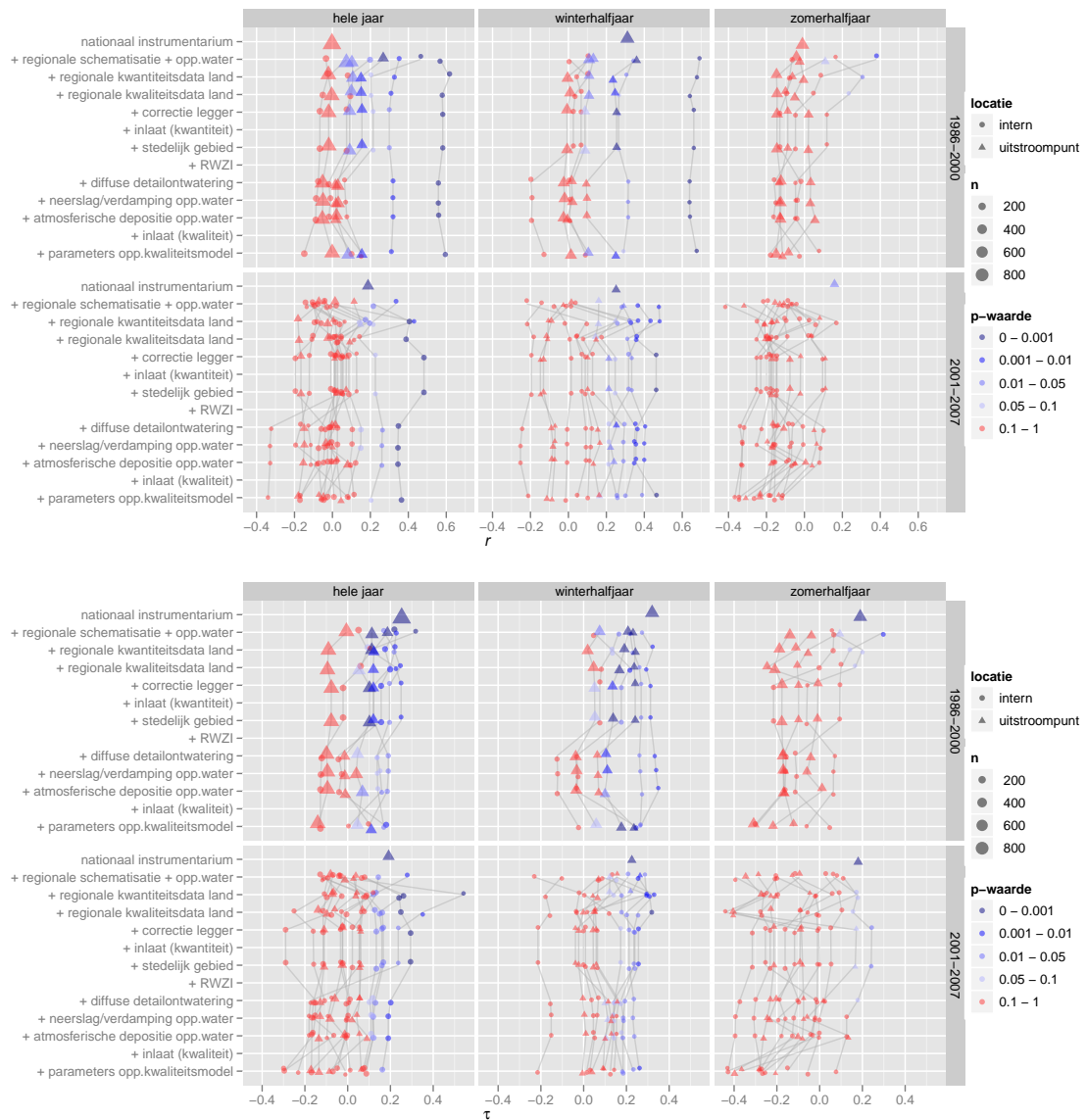
Figuur 3.20: Totaal stikstof (mg l^{-1}) op dagbasis voor de periode 1986-2000 (boven) en 2001-2007 (onder).

3.3 Fosfor

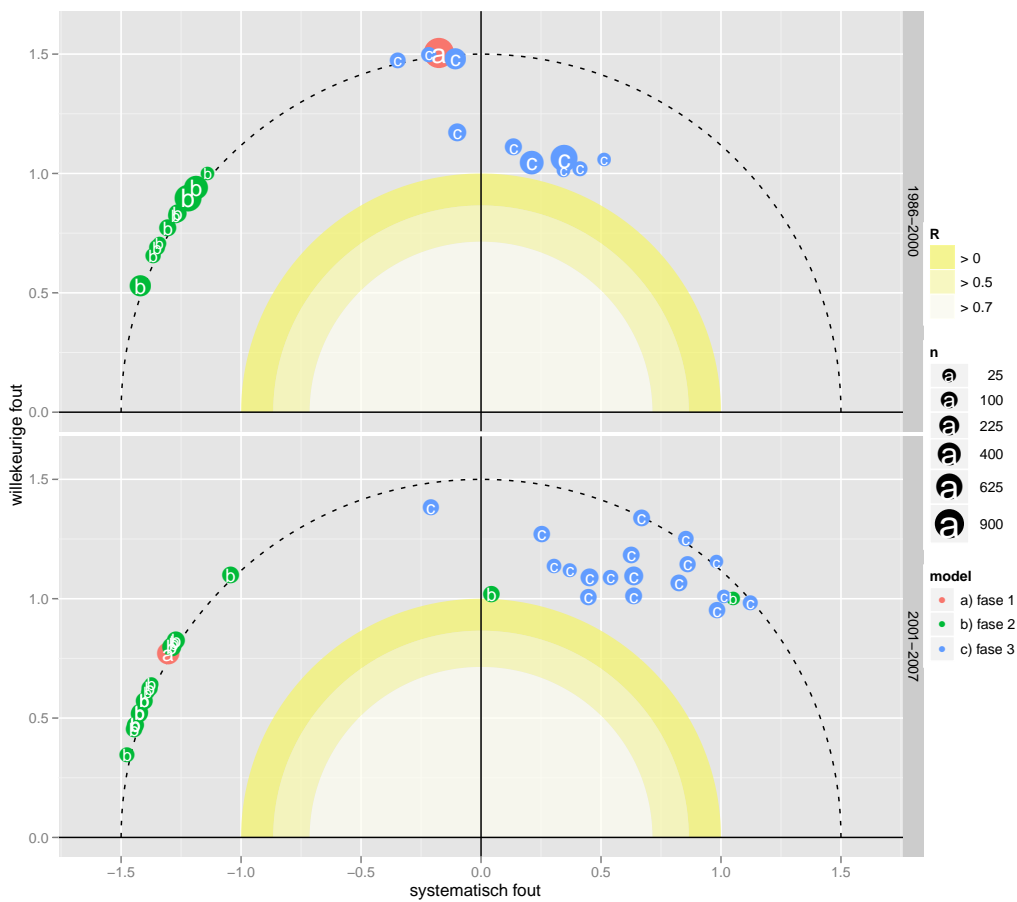
3.3.1 Anorganisch fosfor

Uit figuur 3.21 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.

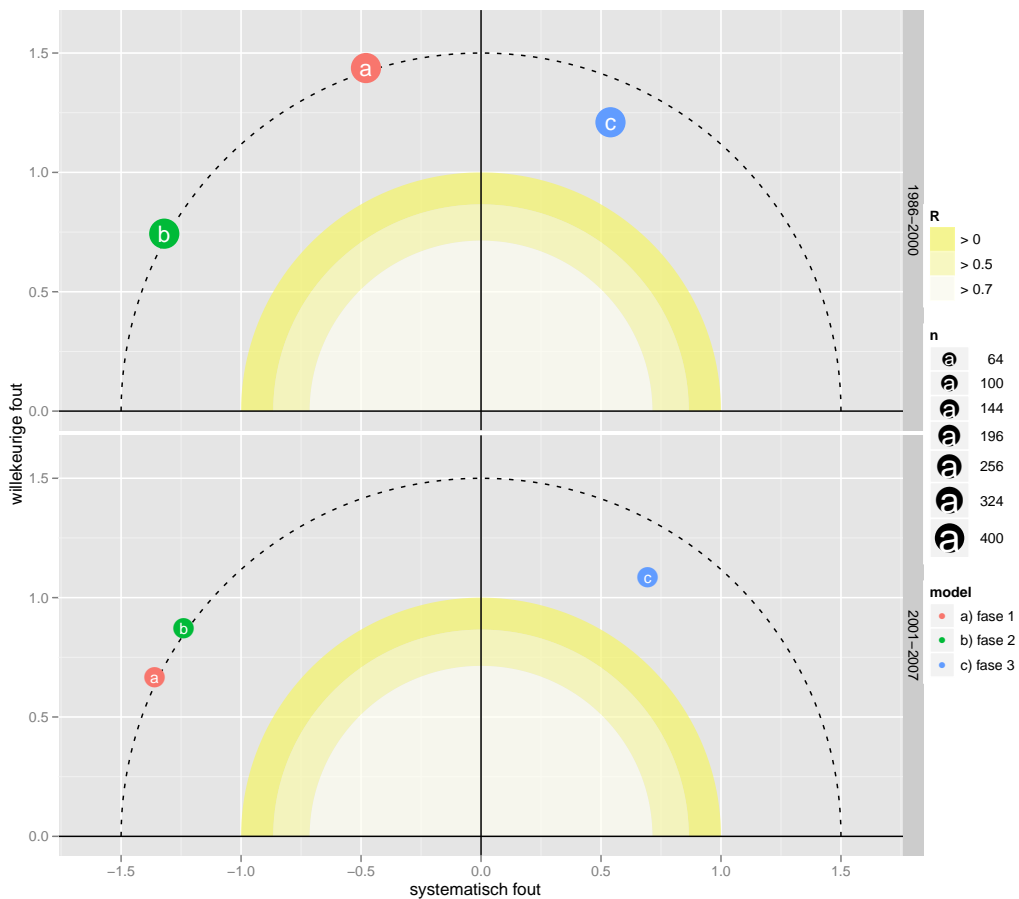
Wat voor het patroon geldt, geldt ook voor het modelleren van absolute anorganisch fosforconcentraties: deze worden nog niet goed door de modellen gesimuleerd (figuur 3.22). Dit geldt voor zowel voor puntmetingen (momentopnames, figuur 3.22) als voor weekgemiddelde concentraties op basis van debietproportionele metingen (figuur 3.23). Alle punten liggen buiten de zon. De gemiddelde absolute fout (figuur 3.24) laat echter zien dat er desalniettemin toch vooruitgang is geboekt. Immers, de absolute fout neemt van fase 1 naar fase 3 sterk af. Over het algemeen wordt in fase 1 en 2 het gehalte aan anorganisch fosfor sterk overschat en in fase 3 iets onderschat (figuur 3.25). Fase 1, d.w.z. het ‘nationaal modelinstrumentarium’, moet ook een systematische overschatting geven omdat dat modelsysteem geen fosforverwijderende processen (‘sinks’) in het oppervlaktewater bevat. De systematische onderschatting in fase 3 blijkt ook uit figuur 3.26. Deze figuur geeft de gemodelleerde en gemeten anorganische fosforconcentraties weer als tijdreeksen. De gemeten concentraties zijn debietsproportioneel en moeten worden geïnterpreteerd als weekgemiddelden. De gemodelleerde fosforconcentraties zijn gemiddeld lager dan de gemeten concentraties.



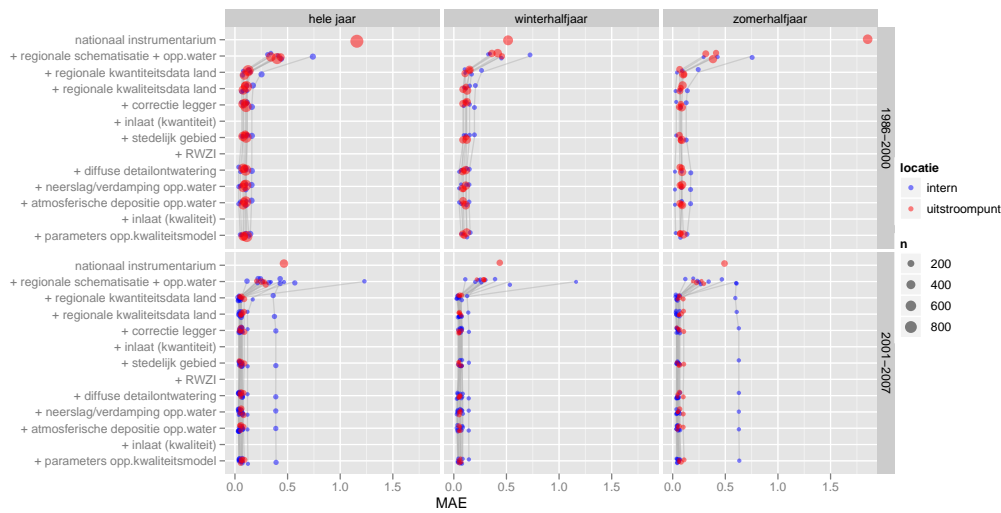
Figuur 3.21: Pearsons (r) en Kendalls correlatiecoëfficiënt (τ) tussen de gemeten en gesimuleerde anorganische fosforconcentratie (mg l^{-1}) op alle locaties in het stroomgebied van 'de Schuitenbeek' voor het hele jaar (kolom 1), het winterhalfjaar (kolom 2) het zomerhalfjaar (kolom 3) en twee perioden (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.



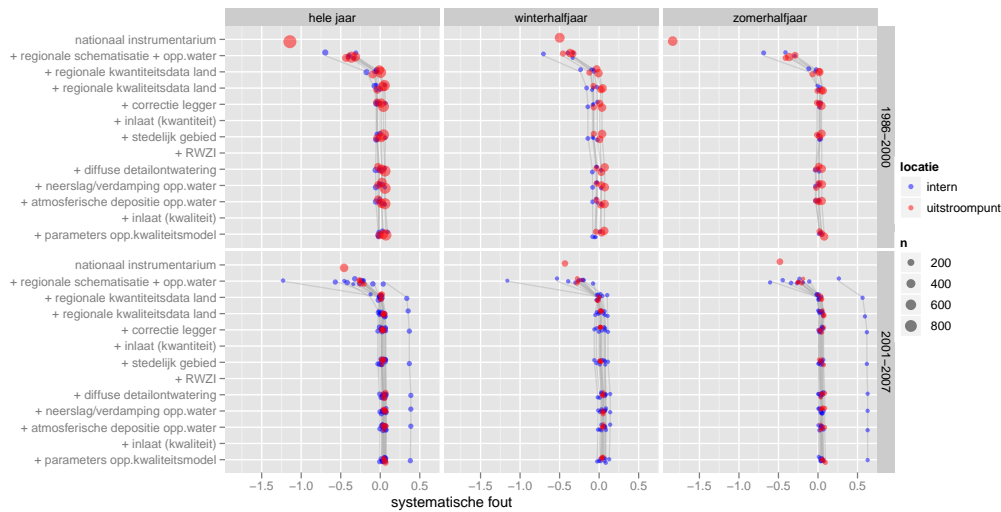
Figuur 3.22: Anorganisch fosfor (mg l^{-1}) op basis van puntmetingen (momentopnames) voor de periode 1986-2000 (boven) en 2001-2007 (onder).



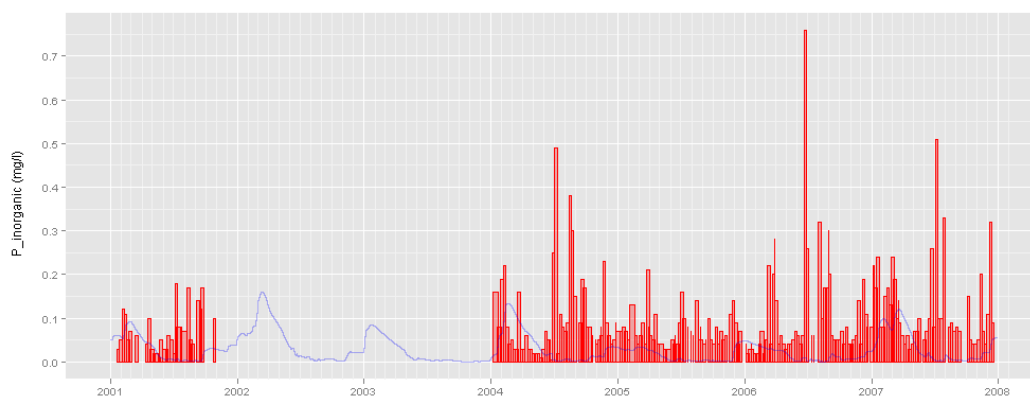
Figuur 3.23: Anorganisch fosfor (mg l^{-1}) op basis van debietproportionele metingen (weekgemiddelden) op het continue meetpunt (25210) voor de periode 1986-2000 (boven) en 2001-2007 (onder).



Figuur 3.24: Gemiddelde absolute fout (mg l^{-1}) voor anorganisch fosfor (mg l^{-1}) op dagbasis voor de hoofdfases.



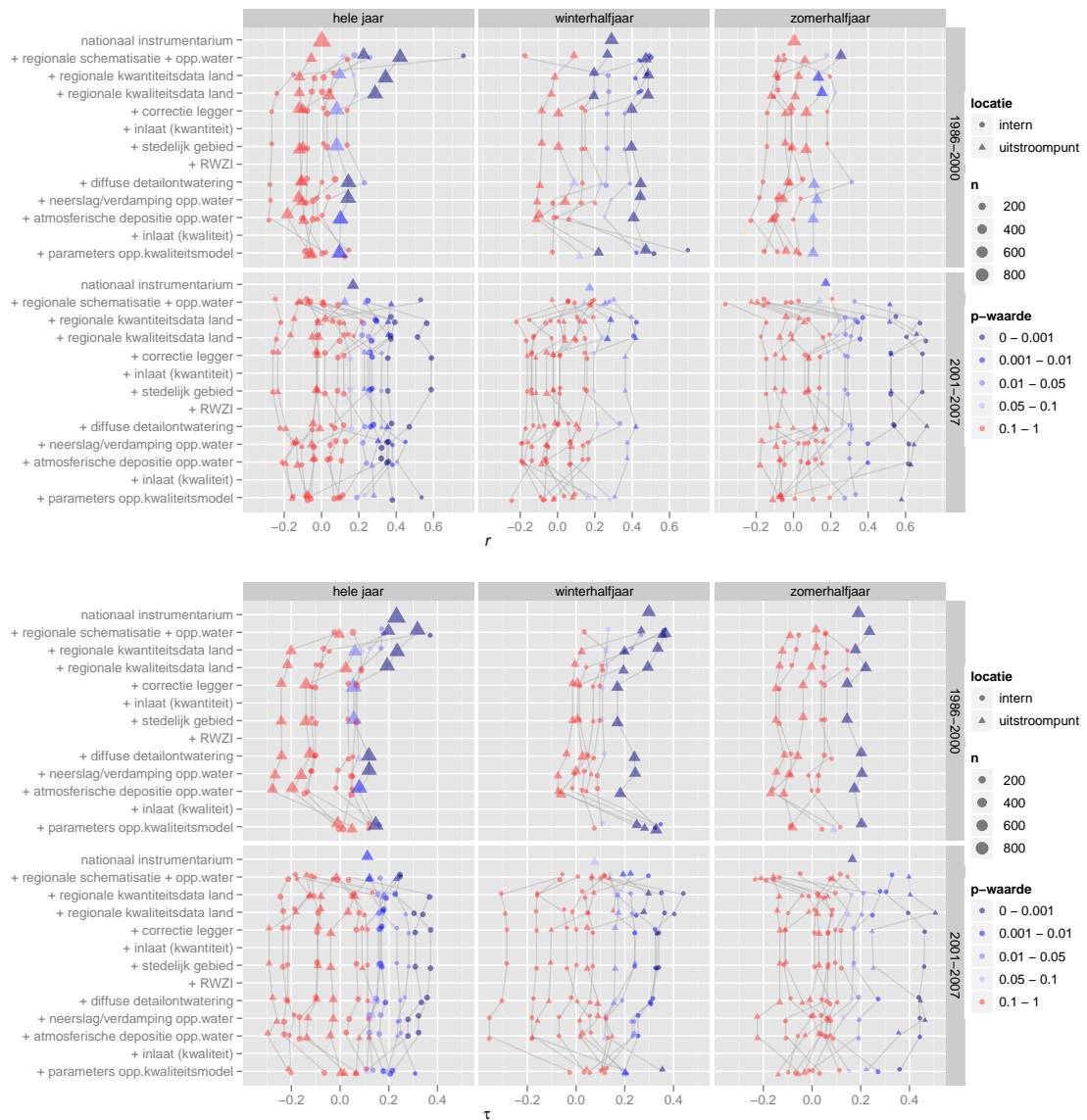
Figuur 3.25: Systematische fout (mg l^{-1}) voor anorganisch fosfor (mg l^{-1}) op dagbasis voor de hoofdfases.



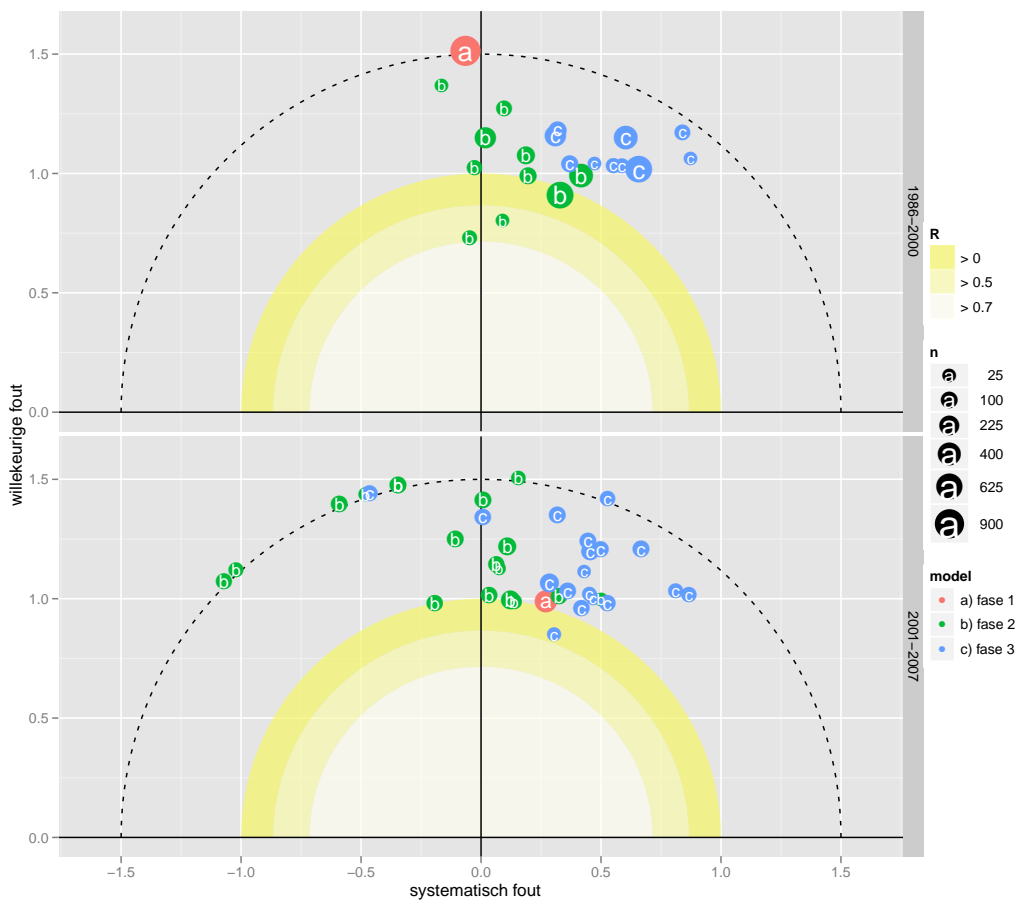
Figuur 3.26: Gemeten (rood) en berekende (blauw) anorganische fosforconcentraties (mg l^{-1}) bij het continue meetpunt 25210 voor fase 3 voor de periode 2001-2007.

3.3.2 Organisch fosfor

Uit figuur 3.27 blijkt dat de temporele patronen in organisch fosfor lastig zijn te modelleren. Organisch fosfor (figuur 3.28) wordt over het algemeen te laag gemodelleerd. Evenals bij anorganisch fosfor is de totale fout hoog.



Figuur 3.27: Pearsons (r) en Kendalls correlatiecoëfficiënt (τ) tussen de gemeten en gesimuleerde organische fosforconcentratie (mg l^{-1}) op alle locaties in het stroomgebied van 'de Schuivenbeek' 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.

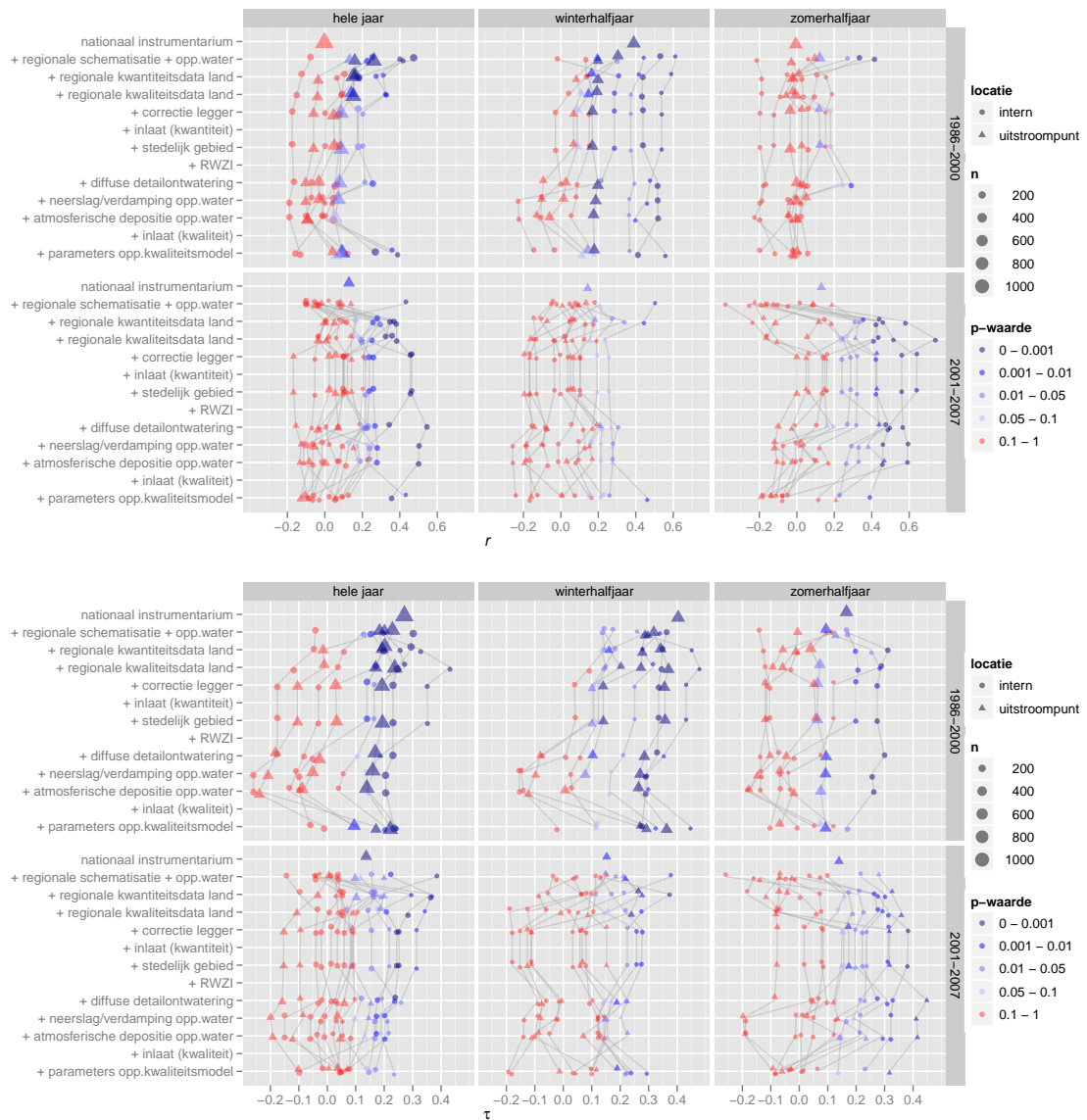


Figuur 3.28: Organisch fosfor (mg l^{-1}) op basis van puntmetingen (momentopnames) voor de periode 1986-2000 (boven) en 2001-2007 (onder).

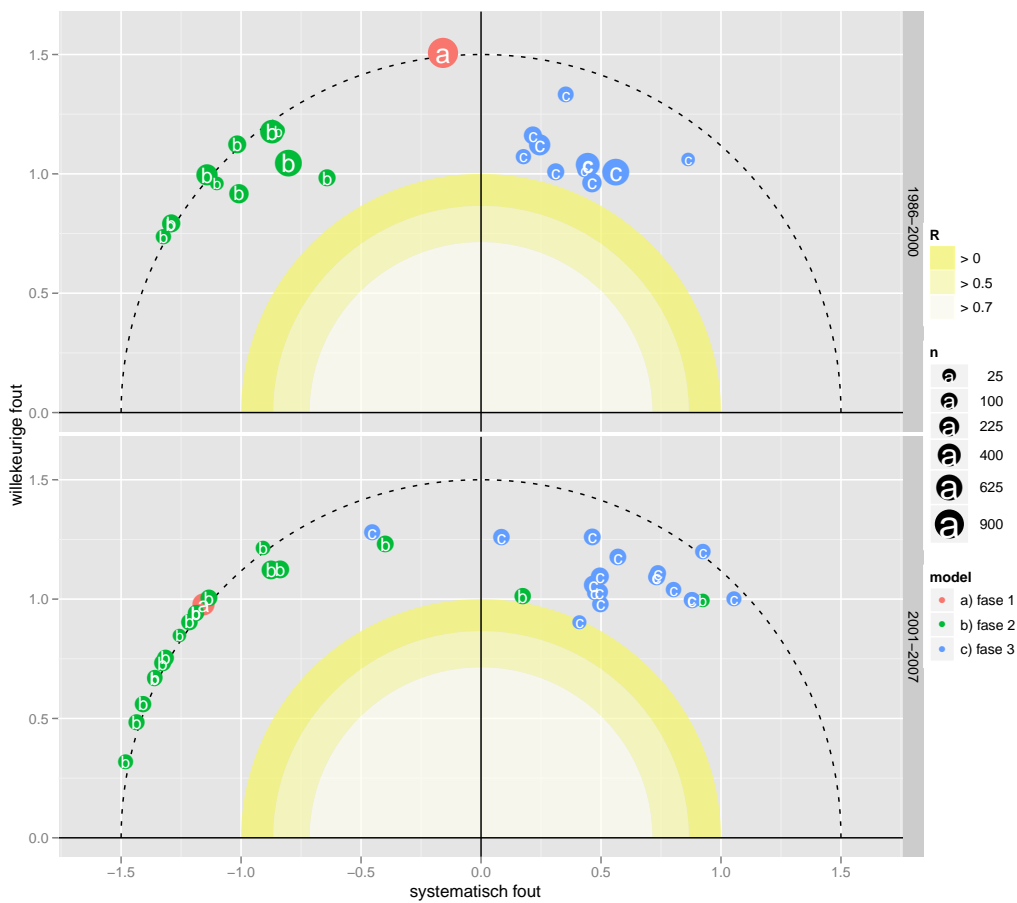
3.3.3 Totaal fosfor

Wat voor anorganisch en organisch fosfor geldt, geldt ook voor totaal fosfor: het temporele patroon wordt niet goed gemodelleerd (figuur 3.29). Op veel locaties wordt geen significant verband gevonden en als er een significant verband bestaat is de relatie zwak.

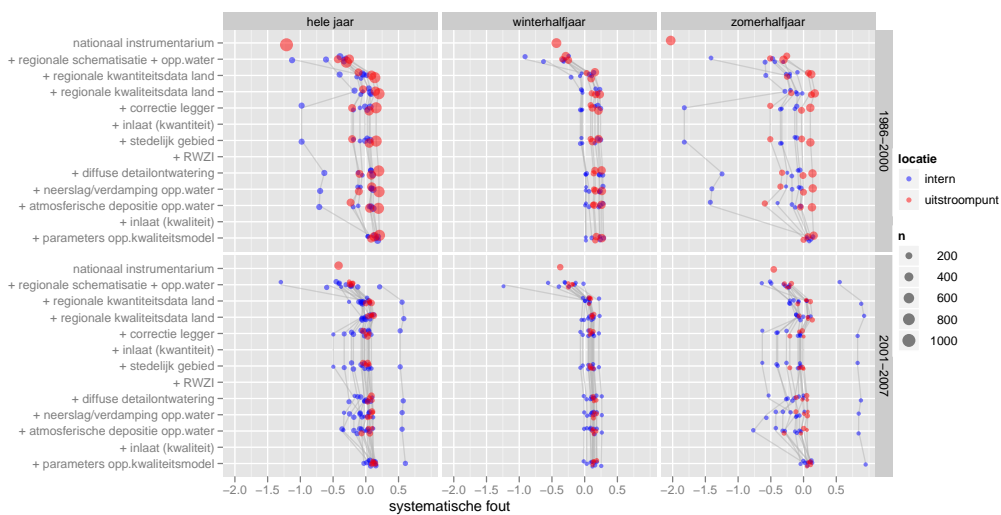
Ook het modelleren van de absolute totale fosforconcentraties gaat nog niet goed (figuur 3.30). In fase 1 en 2 werd totaal fosfor te hoog berekend. Het fase 1 model moet ook te hoge fosforconcentraties leveren omdat dit model geen oppervlaktewatermodule heeft. Het fase 1 model berekent derhalve enkel de nutriëntenvrachten die het landsysteem verlaten. Processen in het oppervlaktewater, zoals sedimentatie, die de fosforconcentratie verlagen zijn dus niet meegenomen. In fase 3 worden over het algemeen te lage fosforconcentraties berekend. De systematische fout is echter wel afgenomen ten opzichte van voorgaande fases (figuur 3.31). Wordt naar de absolute fout gekeken (figuur 3.32) dan blijkt dat gaande van fase 1 naar fase 3 het modelinstrumentarium iets is verbeterd.



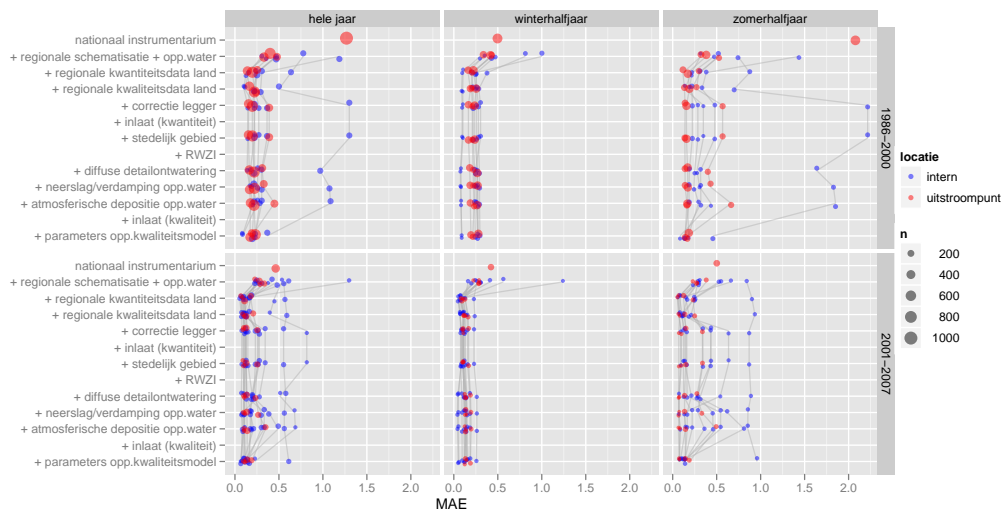
Figuur 3.29: Pearsons (r) en Kendalls correlatiecoëfficiënt (τ) tussen de gemeten en gesimuleerde totale fosforconcentratie (mg l^{-1}) op alle locaties in het stroomgebied van ‘de Schuitenbeek’ voor het hele jaar (kolom 1), het winterhalfjaar (kolom 2) het zomerhalfjaar (kolom 3) en twee perioden (rijen). 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 3.30: Totaal fosfor (mg l^{-1}) op basis van puntmetingen (momentopnames) voor de periode 1986-2000 (boven) en 2001-2007 (onder).



Figuur 3.31: Systematische fout (mg l^{-1}) voor totaal fosfor (mg l^{-1}) op dagbasis voor de afzonderlijke deelfases.



Figuur 3.32: Gemiddelde absolute fout (mg l^{-1}) totaal fosfor (mg l^{-1}) op dagbasis voor de afzonderlijke deelfases.

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. Dat gaat iets beter voor het winterhalfjaar dan voor het zomerhalfjaar. Het zomerhalfjaar wordt namelijk gekenmerkt door extreme afvoerpieken, afgewisseld met droogval, hetgeen lastiger te modelleren is.

Hoewel het temporele patroon redelijk wordt gemodelleerd, overschatten de modelsystemen de gemeten afvoer. Dit geldt met name voor de meetpunten waar de afvoer is bepaald met een OTT-molenmeter. Op het continue meetpunt, waar de afvoermetingen nauwkeuriger zijn, is de systematische fout (systematische overschatting) veel kleiner en voor de validatieperiode zelfs vrijwel verdwenen. Het is bij validatie dan ook van belang dat men zich realiseert dat afwijkingen tussen gemeten en berekende waarden niet alleen worden veroorzaakt door onvolkomenheden in het model, maar dat ook de metingen behept zijn met fouten. Systematische fouten kunnen worden verkleind door de randvoorwaarden van het model te verbeteren (bijvoorbeeld kwel, neerslagverdeling, verdamping) en door kalibratie van modelparameters waarvoor het model gevoelig is en waarvan de waarde onvoldoende bekend is.

Op basis van de modefficiëntie kan worden geconcludeerd dat de wateraf-

Tabel 4.1: Kwalitatieve samenvatting van de resultaten voor ‘de Schuitenbeek’

grootheid	temporeel patroon	pa- troon	absolute waar- den	waar- opmerkingen
water	matig-redelijk		matig-redelijk	
anorganisch N	matig-redelijk		slecht-matig	
organisch N	slecht		slecht	
totaal N	matig		slecht-(matig)	
anorganisch P	slecht		slecht	
organisch P	slecht		slecht	
totaal P	slecht		slecht	

voer in fase 3 over het algemeen beter wordt voorspeld dan in voorgaande fases. Dit komt met name door de deelfase waarin de neerslagverdeling, de onderrand (kwel/wegzijging), de drainagekarakteristieken en de dikte van het topsysteem gebiedspecifiek zijn gemaakt. De modelsystemen hebben echter moeite om piekafvoeren te voorspellen in het zomerhalfjaar. Een belangrijke reden waarom piekafvoeren momenteel niet goed worden voorspeld is dat deze informatie niet in de modelinvoer aanwezig is. Dit kan deels worden verbeterd door gebruik te maken van een hogere resolutie aan neerslaggegevens (en verdampingsgegevens).

Over het algemeen presteren de modelsystemen beter op benedenstroomse meetlocaties dan op bovenstroomse meetlocaties ten gevolge van uitmiddelingseffecten. Daarnaast presteren de modelsystemen beter op het continue meetpunt dan op meetlocaties waar met OTT-molenmetingen wordt gemeten. Hiervoor zijn twee oorzaken aan te wijzen. Enerzijds zijn OTT-molenmetingen minder nauwkeurig, anderzijds is een OTT-molenmetingen een momentopname, en geen daggemiddelde zoals een continue meting. Momentopnames zijn voor snel reagerende systemen met een geringe hydraulische verblijftijd zoals de Schuitenbeek minder geschikt omdat piekafvoeren kunnen worden gemist. De modelresultaten sluiten dan ook veel beter aan bij de continue metingen, omdat die op hetzelfde temporele aggregatieniveau betrekking hebben.

4.2 Stikstof

Het temporele patroon in anorganisch stikstof wordt het beste gereproduceerd en voorspeld op de benedenstroomse locaties. Dat geldt met name voor het continue meetpunt. Het patroon wordt over het algemeen beter voorspeld in het winterhalfjaar dan in het zomerhalfjaar. De locaties waar het patroon minder goed wordt voorspeld liggen met name meer bovenstrooms. Het patroon in organisch stikstof is daarentegen veel lastiger te voorspellen. Slechts op een beperkt aantal meetlocaties is sprake van een significant zwak verband.

De absolute stikstofconcentraties worden nog niet goed gereproduceerd en voorspeld. Voor fase 3 is sprake van een systematische overschatting van de anorganische en totale stikstofconcentraties. Als naar de afzonderlijke deelfases wordt gekeken dan blijkt dat dit vooral wordt veroorzaakt door een vermindering van de denitrificatie waardoor meer anorganisch stikstof beschikbaar is voor uitspoeling naar het oppervlaktewater. De modelprestatie kan worden verbeterd door meer onderzoek te doen naar de parameterisatie van het 'watergetal'.

4.3 Fosfor

Hoewel opeenvolgende fases vooruitgang laten zien in het modelleren van fosfor, geldt over het algemeen dat het temporele patroon en de absolute concentraties nog niet goed door de modellen worden beschreven. Voor het temporele patroon geldt dat de meeste verbanden niet significant zijn. Als sprake is van een significant verband dan is dat verband zwak.

De modelsystemen zijn momenteel ook nog niet in staat om absolute fosforconcentraties in de Schuitenbeek te beschrijven en te voorspellen. Dat geldt voor zowel puntmetingen (momentopnames) als weekgemiddelde debietproportionele metingen. De fase 1 en fase 2 systemen laten een systematische overschatting zien, het fase 3 systeem een systematische onderschatting.

4.4 Algemeen

Uit dit rapport blijkt dat een verbetering van de modellering van de hydrologie (waterafvoer) niet noodzakelijkerwijs hoeft te leiden tot een verbetering van de modellering van de nutriëntenconcentraties. Hydrologie en chemie verdienen elk afzonderlijk aandacht, maar zeker ook in relatie tot elkaar. Wat voor de hydrologie en chemie geldt, geldt ook voor het landsysteem en het oppervlaktewatersysteem. Het oppervlaktewatermodel kan pas goed worden gedraaid als er geen grote onvolkomenheden meer in het landsysteem zitten. Pas als de modelleur daar vertrouwen in heeft kan op basis van een gevoeligheidsanalyse en expertkennis de modelprestatie verder worden verbeterd door middel van kalibratie. Bij kalibratie worden de waarden van modelparameters waarvoor het modelresultaat gevoelig is en die bovendien lastig zijn te kwantificeren op basis van statistische technieken en expertkennis verder ingeregeld.

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 ook nog niet goed mogelijk om uitspraken te doen of kritische concentraties al dan niet worden overschreden.

Voor stikstof is het op een aantal locaties al wel mogelijk om het temporele patroon te beschrijven. Het modelsysteem zegt dan niets over de absolute grootte van de stikstofconcentratie, maar wel of er sprake is van een verander-

ing van de concentratie (toe- of afname). De modelsystemen presteren beter in het reproduceren en voorspellen van de waterafvoer. Dat geldt voor zowel het temporele patroon, als voor de absolute waarden.

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 oppervlaktewatersysteem 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).

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>
- Gebremeskel, S., Liu, Y. B., De Smedt, F., Hoffmann, L., Pfister, L., 2005. Assessing the hydrological effects of landuse changes using distributed hydrological modelling and GIS. *International Journal of River Basin Management* 3 (4), 261–271.
- 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
- Jansen, H. C., Löschner-Wolleswinkel, R. J., Sicco Smit, M. E., van der Bolt, F. J. E., Roelsma, J., Schoumans, O. F., Siderius, C., van Tol-Leenders, T. P., 2009. Systeemanalyse voor het stroomgebied van de Schuitenbeek fase 3. Tech. Rep. 1765, Alterra.
- Jansen, H. C., Renaud, L. V., Leenders, T. P., van der Bolt, F. J. E., 2006a. Systeemanalyse voor het stroomgebied van de Schuitenbeek fase 1. Tech. Rep. 1272, Alterra, Wageningen.
- Jansen, H. C., Sicco Smit, M. E., Leenders, T. P., van der Bolt, F. J. E., Renaud, L. V., 2006b. Systeemanalyse voor het stroomgebied van de Schuitenbeek. Tech. Rep. 1387, Alterra, Wageningen.
- 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.
- 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 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:

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

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 modelprestatie-maten).

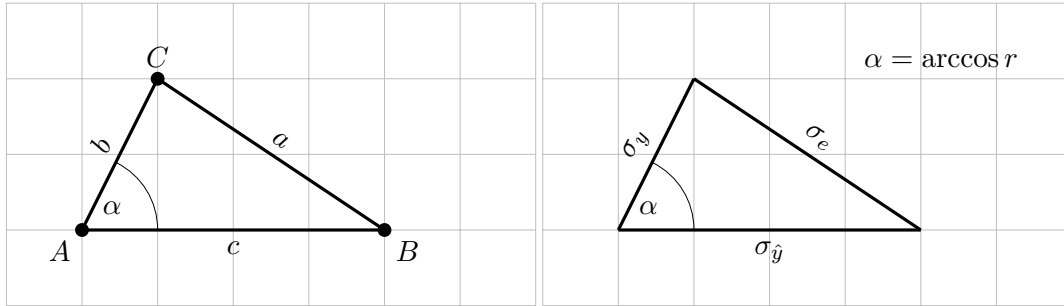
(Taylor, 2000, 2001) gebruikte de gelijkenis tussen vergelijking B.1 en de cosinusregel

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

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.

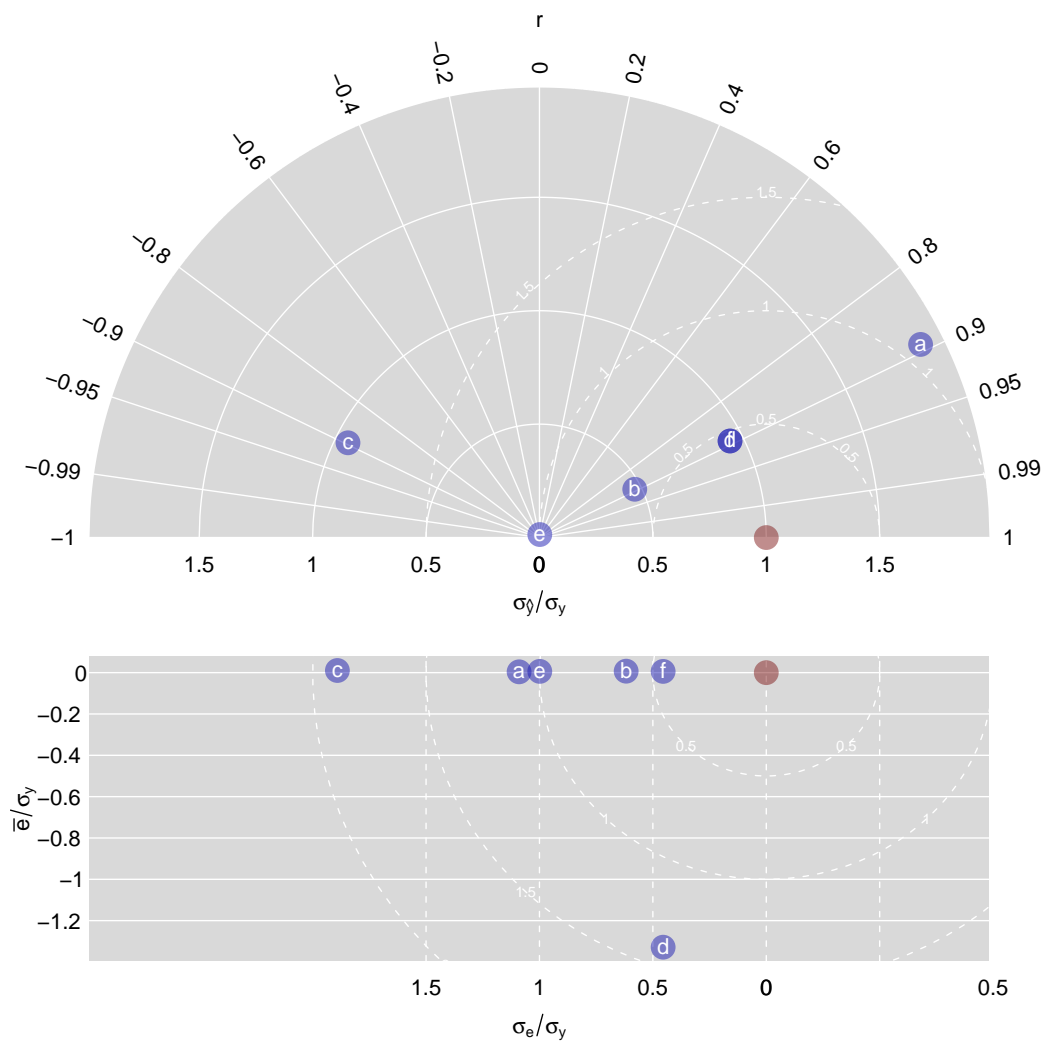
Het Taylor-diagram is gegeven in de bovenste helft van figuur B.2. Langs de horizontale as is de genormaliseerde standaarddeviatie weergegeven. Dat is de vierkantswortel uit de genormaliseerde variantie (A.6). Langs de cirkelboog is de correlatiecoëfficiënt r uitgezet. De rode stip stelt de waarnemingen 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

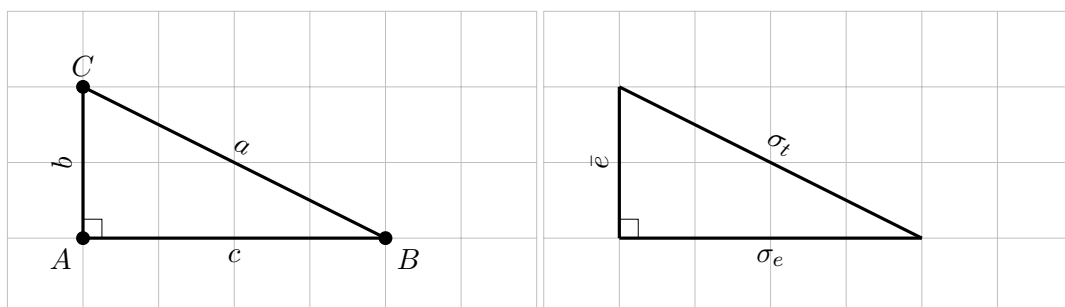


Figuur B.1: Grafische weergave van de relatie tussen de cosinusregel en enige validatiematen.

ⓕ 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).

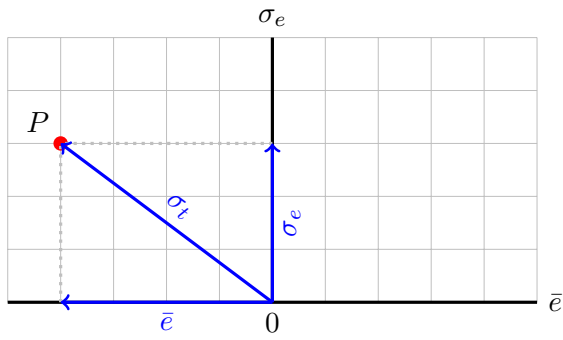
De zonneplot geeft de relatie weer tussen de willekeurige fout (verg. A.9, de systematische of gemiddelde fout (verg. A.8) en de totale fout (verg. A.9). Een geometrische voorstelling van deze relatie is gegeven in figuur C.1. Op de x -as staat de systematische of gemiddelde fout gegeven (\bar{e}). Negatieve 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)

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

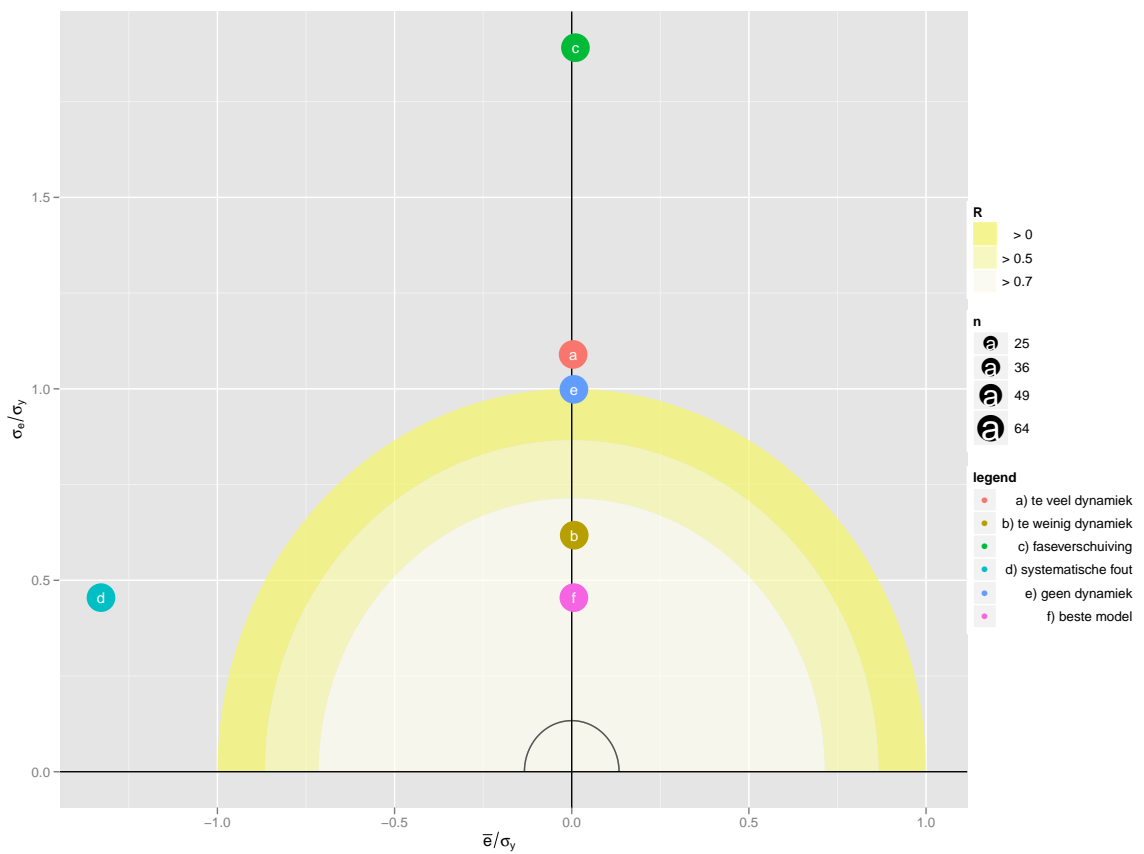
Op deze stelling berust de zonneplot. De zonneplot is gebaseerd op het ‘schietschijfdiagram’ (Engels: ‘target diagram’) van Jolliff et al. (2009). Een belangrijk verschil met het schietschijfdiagram is dat slechts twee in plaats van vier kwadranten worden gebruikt. Jolliff et al. (2009) gebruiken de 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 ($E_{NS} = 0$). 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-erwijs) 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.

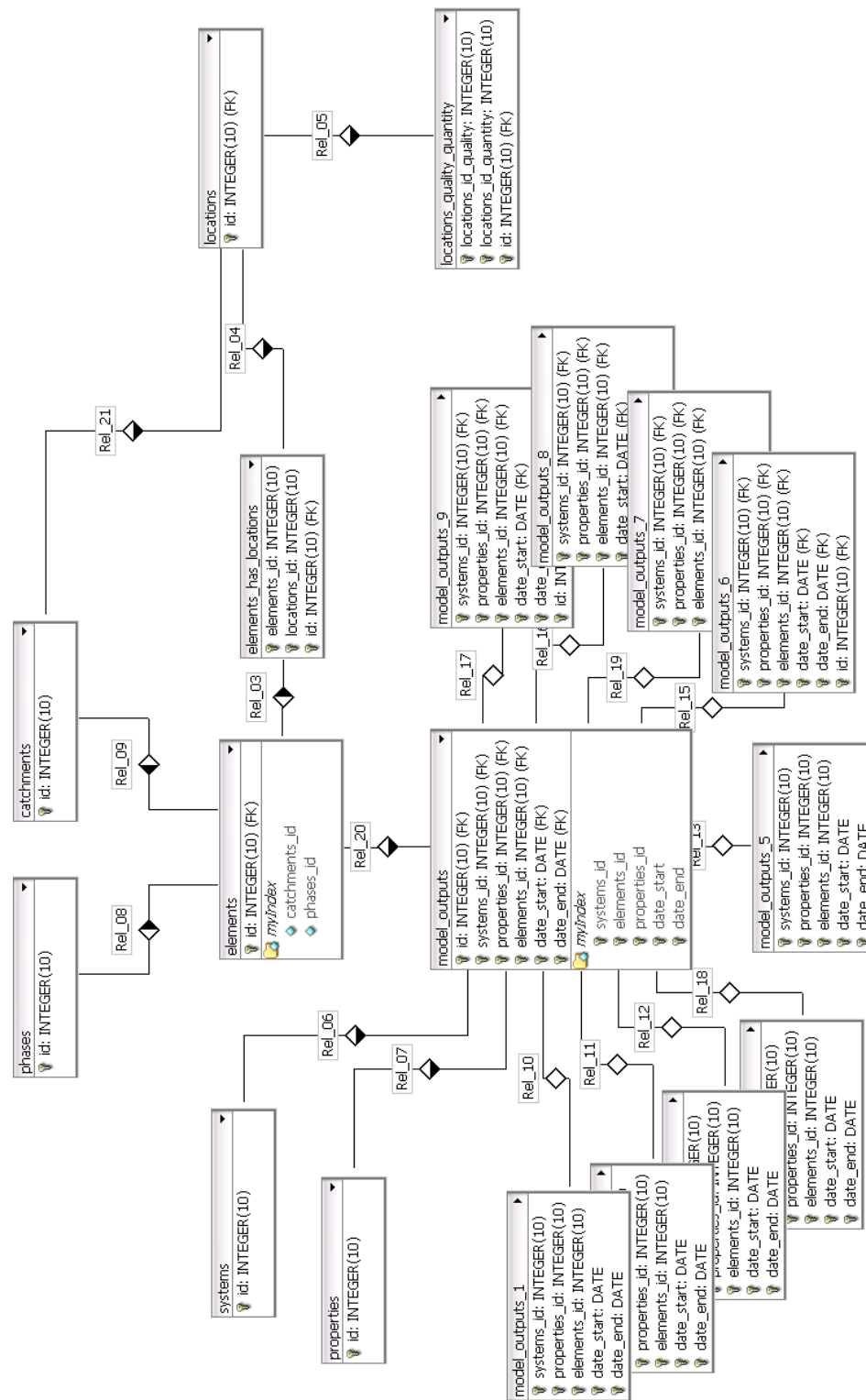
In figuur C.2 is de totale fout σ_t die correspondeert met de meetfout berekend door de variatiecoëfficiënt van de meetfout te vermenigvuldigen met het gemiddelde van de waarnemingen. Voor een aantal stoffen in het oppervlaktewater is de variatiecoëfficiënt van de meetfout getabelleerd in Rode and Suhr (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.

