


Deltares

**Neuraal netwerk PCLake
ten behoeve van KRW-verkenner**

Witteveen+Bos
van Twickelostraat 2
postbus 233
7400 AE Deventer
telefoon 0570 69 79 11
telefax 0570 69 73 44

Neuraal netwerk PCLake ten behoeve van KRW-verkenner

referentie	projectcode	status
UT565-2/posm/003	UT565-2-1	definitief
projectleider	projectdirecteur	datum
drs.ing. S.A. Schep	drs. M. Klinge	22 december 2010

autorisatie	naam	paraaf
goedgekeurd	drs. M. Klinge	

Witteveen+Bos
van Twickelostraat 2
postbus 233
7400 AE Deventer
telefoon 0570 69 79 11
telefax 0570 69 73 44

Het kwaliteitsmanagementsysteem van Witteveen+Bos is gecertificeerd op basis van ISO 9001.

© Witteveen+Bos

Niets uit dit bestek/drukwerk mag worden veelevoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze dan ook zonder voorafgaande toestemming van Witteveen+Bos Raadgevende ingenieurs B.V., noch mag het zonder een dergelijke toestemming worden gebruikt voor enig ander werk dan waarvoor het is vervaardigd.

INHOUDSOPGAVE	blz.
SAMENVATTING	
1. INLEIDING	1
2. DATASET VOOR SIMULATIES PCLAKE	3
2.1. Inleiding	3
2.2. Methode	3
2.3. Resultaten	4
3. DATASET REGRESSIEANALYSES	6
3.1. Inleiding	6
3.2. Methode	6
3.3. Resultaten	9
4. METAMODEL	10
4.1. Inleiding	10
4.2. Methode	10
4.3. Resultaten	11
5. CONCLUSIES	15
6. LITERATUURLIJST	16
laatste bladzijde	16
bijlagen	aantal bladzijden
I Karakteristieken Nederlandse en Europese meren en plassen	2
II Formules (gesnoeide) PUN's	6

SAMENVATTING

Recent is met de publicatie 'van helder naar troebel... en weer terug' (STOWA, 2008) een diagnostisch kader gepresenteerd voor ondiepe meren. Dit diagnostisch kader wordt geïmplementeerd in de KRW-verkenner. Centraal in de methode staat het ecologisch model PCLake (Janse, 2005). Omdat de tot nu toe ontwikkelde metamodellen van PCLake tekort schieten is een aanvullende analyse nodig. In deze studie zijn verschillende regressietechnieken vergeleken om tot een eenvoudig, transparant, betrouwbaar en bruikbaar metamodel van het ecologisch model PCLake te komen.

Een belangrijk aspect is het genereren van de dataset die als input voor het metamodel dient. Deze dataset is generiek opgezet aan de hand van relevante kenmerken zoals deze voorkomen in Nederlandse meren. Deze kenmerken zijn vervolgens gebruikt als invoer voor simulaties met PCLake om kritische grenzen voor de P-belasting te berekenen. Deze gegenereerde dataset van 41492 kritische grenzen is vervolgens gebruikt voor het ontwikkelen van het metamodel met behulp van multiple regressie, regressiebomen, black-box en white-box-neurale netwerken.

Van de gebruikte technieken presteren de neurale netwerken als enige voldoende goed om betrouwbare resultaten te leveren als metamodel. Multiple regressie ($R^2=0.77$) en regressiebomen ($R^2=0.84$) hebben gemiddelde relatieve fouten die groter zijn dan 0.4 (meer dan 40 % afwijking, gemiddeld respectievelijk 1.57 en 0.64). De neurale netwerken presteren ongeveer even goed (respectievelijk $R^2=0.97$; $R^2=0.96$; gemiddelde relatieve fouten 0.18 en 0.19), waarbij de white-box-variant in vergelijking met de black-box-versie eenvoudiger en beter interpreteerbaar is. Het vereenvoudigen (prunen) van de white-box neurale netwerken levert relatief weinig meerwaarde op voor de interpretatie.

De white-box neurale netwerken worden geïmplementeerd in de KRW-verkenner. De gemiddelde relatieve fouten van deze neurale netwerken zijn kleiner dan de modelonzekerheden en dus goed toepasbaar in de KRW-verkenner.

1. INLEIDING

achtergrond project

Met de ontwikkeling STOWA-publicatie van 'van helder naar troebel... en weer terug' (2008) is er een methode ontwikkeld waarmee waterbeheerders een goede diagnose kunnen maken van het ecologisch functioneren van watersystemen. Om ondersteuning te bieden voor deze methode is het nodig om dit diagnostische kader voor ondiepe meren verder te ontwikkelen binnen de KRW-verkenner. Binnen de methode is een centrale rol weggelegd voor het vergelijken van de actuele P-belasting met de kritische P-belasting van het systeem, zoals gemodelleerd met het ecologische model PCLake.

Het gebruik van PCLake direct binnen de KRW-verkenner is om technische redenen niet haalbaar vanwege de complexiteit en tijdrovendheid van het uitvoeren van berekeningen met het model. Ook is het model PCLake ingebed in het onconventionele pakket ACSL. Voor toepassing van dit concept binnen de KRW-verkenner is dus een goed betrouwbare en bruikbare ontsluiting van PCLake nodig. Als vereenvoudiging zijn al eerder metamodellen ontwikkeld voor PCLake (Vleeshouwers *et al.* 2004, Witteveen+Bos 2006), maar deze zijn onvoldoende betrouwbaar gebleken en daarom niet geschikt voor implementatie binnen de KRW-verkenner. Om toch tot een goede implementatie van PCLake in de KRW-verkenner te komen is in deze analyse opnieuw een metamodel gebouwd met de huidige nieuwe inzichten, verbeterde rekencapaciteit en analysetechnieken.

doel

Het doel van dit onderzoek is het ontwikkelen van een metamodel van het ecologisch model PCLake ten behoeve van de KRW-Verkenner.

Het metamodel moet voldoen aan een aantal eisen in aflopende prioriteit, te weten:

- betrouwbaar: het metamodel moet met het model PCLake berekende waarden benaderen;
- bruikbaar: het metamodel moet ingebouwd kunnen worden in de KRW-Verkenner;
- eenvoud: het metamodel moet zo eenvoudig mogelijk zijn;
- transparant: het metamodel moet interpreteerbaar zijn.

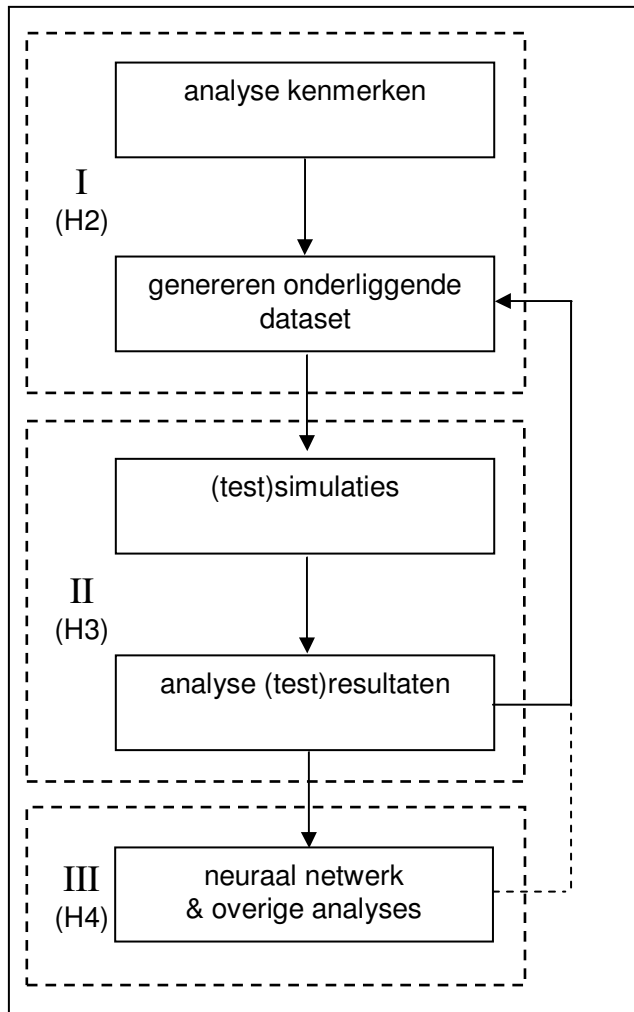
leeswijzer

De kapstok van dit onderzoek wordt gevormd door het hieronder weergegeven stappenplan (afbeelding 1.1). Er is onderscheid gemaakt in drie onderdelen. De verschillende onderdelen zijn beschreven in afzonderlijke hoofdstukken met elk een beschrijving van methode en resultaten:

- hoofdstuk 2: dataset voor simulaties PCLake (I);
- hoofdstuk 3: dataset regressieanalyses (II);
- hoofdstuk 4: metamodel (III).

Het eerste deel (I) geeft een eenvoudige analyse van de variatie van kenmerken in Nederlandse en Europese plassen en meren. Op basis van deze analyse is de bandbreedte vastgesteld van de voor het ecologisch model relevante kenmerken. Met deze informatie is een representatieve dataset samengesteld die gebruikt is als input voor de simulaties met PCLake. In het tweede deel (II) is beschreven welke aanpassingen in de aansturing van PCLake zijn gemaakt en hoe de kritische belasting wordt berekend. De resultaten van de simulaties zijn vervolgens eenvoudig geanalyseerd op consistentie en inhoud. Deze stappen zijn herhaald tot de resultaten voldeden aan gestelde eisen (zie hoofdstuk 3). In deel III is een aantal regressieanalyses uitgevoerd, waaronder neurale netwerk analyses. De resultaten hiervan zijn vergeleken op basis van de in de doelstelling genoemde criteria. Dit heeft geleid tot een zo goed mogelijk model voor het voorspellen van kritische grenzen voor de P-belasting in ondiepe meren.

afbeelding 1.1. Stappenplan, tevens leeswijzer neuraal netwerk PCLake



2. DATASET VOOR SIMULATIES PCLAKE

2.1. Inleiding

Het doel van dit onderdeel is het ontwikkelen van een dataset, waarmee simulaties met PCLake worden uitgevoerd. Deze simulaties vormen de input voor de analyses die ten grondslag liggen aan het metamodel PCLake. De kwaliteit van het metamodel is voor een belangrijk deel afhankelijk van de kwaliteit van de dataset.

2.2. Methode

De dataset is afgeleid op basis van:

1. een eenvoudige analyse van de variatie van kenmerken in Nederlandse en Europese plassen en meren;
2. de gevoeligheid van PCLake voor de verschillende kenmerken;
3. de door de waterbeheerders vastgestelde maatregelen voor de KRW. De dataset is aangepast op basis van de resultaten in het vervolg van de analyse (simulatie met PCLake, regressieanalyses). De iteratieve aanpak heeft het mogelijk gemaakt de samenstelling van de dataset te optimaliseren op basis van zowel de gevoeligheid van PCLake als de eisen die gesteld worden vanuit de regressieanalyses.

kenmerken

De voor dit onderzoek relevante kenmerken zijn weergegeven in tabel 2.1. Het gaat om eenvoudig meetbare kenmerken die ook bij eerdere studies zijn opgenomen in metamodellen (Janse 2005, Witteveen+Bos 2006).

Er is tevens een analyse gemaakt van de gevoeligheid van het model voor andere kenmerken: de fractie opgelost fosfaat in de P-belasting, de fractie vis die per dag wordt verwijderd, het peilverschil tussen zomer en winter, de achtergrondextinctie, de NP-ratio van de nutriëntenbelasting en de concentratie SiO₂ in het inlaatwater. Op basis van de resultaten van een multi-pele regressie analyse met deze kenmerken is de achtergrondextinctie aanvullend opgenomen in de dataset. De achtergrondextinctie heeft een belangrijke invloed op de kritische belasting, is goed meetbaar en is van groot belang voor het ecologisch functioneren van plassen en meren met een veenbodem.

tabel 2.1. Relevante kenmerken van meren

parameter	eenheid	belang
diepte	m	De diepte is van direct belang voor het lichtklimaat en daarmee voor de potentie voor plantengroei. Verder bepaalt de diepte de sedimentatie en resuspensie.
strijklengte	m	De strijklengte (lengte waarover de wind invloed uitoefent op het water) bepaalt voornamelijk de sedimentatie en resuspensie.
aandeel moeras	m ² /m ²	Het aandeel moeras is onder andere van belang voor de verwijdering van nutriënten.
sedimenttype	-	De samenstelling van de waterbodem bepaalt onder andere de sedimentatie en resuspensie, binding van nutriënten en afbraak van organisch materiaal.
ingaaend debiet	mm/d	Het debiet bepaalt onder andere de opnamesnelheid van nutriënten door algen.
achtergrondextinctie	-	De achtergrondextinctie is van direct belang voor het lichtklimaat en daarmee voor de potentie voor plantengroei.

verdeling

Het uitgangspunt is dat de dataset representatief is voor Nederlandse ondiepe meren en plassen. De verdeling van waarden voor de verschillende kenmerken is daarom gebaseerd op een eenvoudige analyse van de variatie van kenmerken in 37 Nederlandse en 15 Europese ondiepe meren en plassen. Dit overzicht is overgenomen uit het proefschrift van Janse (2005) en weergegeven in bijlage I. De kwaliteit van het metamodel in het voor waterbeheerders relevante bereik is vergroot door veel voorkomende waarden zwaarder mee te laten wegen. Daarnaast is ingespeeld op de door waterbeheerders vastge-

stelde maatregelen voor de KRW en recente ontwikkelingen, zoals aanleg van nieuwe watersystemen waarbij soms grote arealen moeras worden aangelegd.

Het uitgangspunt is een normale verdeling. Er is uitgegaan van een specifieke bandbreedte en een specifieke verdeling per kenmerk op basis van statistische maten (mediaan, gemiddelde, percentielen) en een histogram. Het nadeel van een normale verdeling is dat het niet passend te maken is voor de volledige bandbreedte (de waarden in de 'staart' komen te weinig voor in verhouding met de waargenomen/gewenste verdeling). Om dit te verbeteren zijn er testen met een zogenaamde driehoeksverdeling gedaan. Het nadeel van de driehoeksverdeling is dat er juist teveel waarnemingen buiten het interessegebied voorkomen. Uiteindelijk is er voor gekozen om het beste van beide methoden te combineren door een geconstrueerde verdeling met een (log)normale verdeling als basis en een 'uitloopverdeling' (blokverdeling) voor de hogere waarden (zie de afbeeldingen onder resultaten).

Voor het aandeel moeras is bijvoorbeeld uitgegaan van een verdeling met een laag gemiddelde overeenkomstig de kenmerken, maar met een additionele uitloopverdeling tot en met 2 (dus 2x zoveel moeras als open water). Op dit moment is dit een weinig/niet voorkomende waarde voor het aandeel moeras, maar in de nabije toekomst zal dit wel worden toegepast in bijvoorbeeld nieuw te creëren meren/kreeksystemen.

2.3. Resultaten

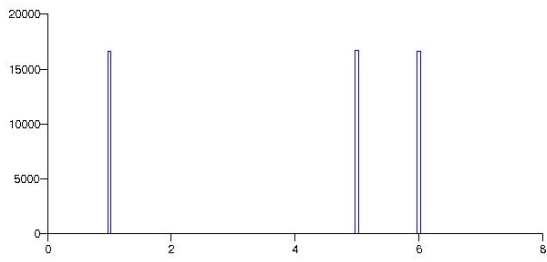
Voor het berekenen van de kritische grenzen met PCLake zijn uiteindelijk twee basisdatasets gegenereerd met elk 50.000 random combinaties. In totaal zijn deze datasets goed voor het in potentie berekenen van de kritische grenzen van maximaal 100.000 verschillende combinaties van eigenschappen van meren. De verdelingen staan in afbeelding 2.1 per kenmerk weergegeven.

Van de uiteindelijke dataset zijn een aantal basale beschrijvende statistieken berekend (tabel 2.2). Hiermee wordt een indruk gegeven van de verdeling van de verschillende parameters.

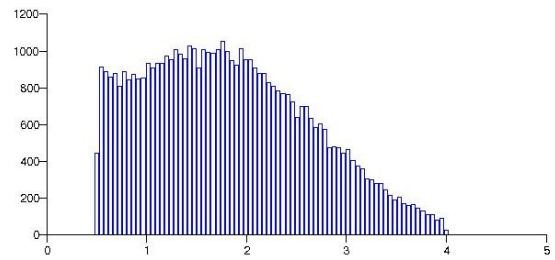
tabel 2.2. Beschrijvende statistiek definitieve dataset

	invoerparameters				kritische grens		
	diepte (m)	aandeel moeras (-)	strijklengte (m)	debiet (mm/d)	extinctie (-)	vanuit helder (mg/m ² /d)	vanuit troebel (mg/m ² /d)
gemiddelde	1.79	0.11	1464.13	42.10	0.66	6.49	5.20
mediaan	1.71	0.02	688.50	27.45	0.54	2.93	1.32
st.deviation	0.80	0.33	2135.48	44.25	0.38	11.99	12.13

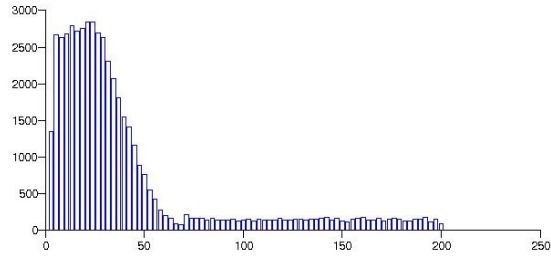
afbeelding 2.1. De uiteindelijke histogrammen van de gevarieerde kenmerken



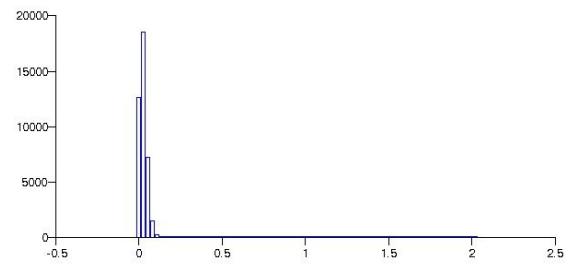
sedimenttype (1=zand; 5=veen; 6=klei)



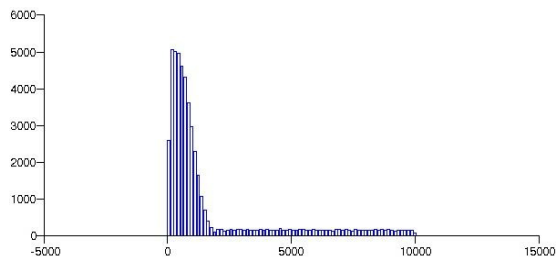
diepte (m)



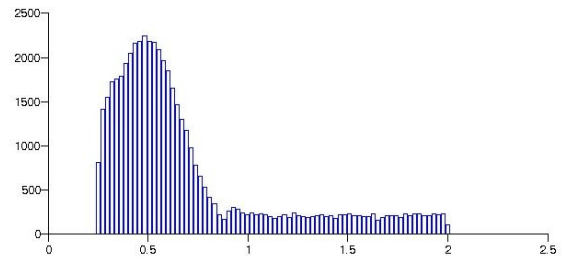
ingand debiet (mm/d)



moerasaandeed (m^2/m^2)



strijklengte (m)



achtergrondextinctie (-)

3. DATASET REGRESSIEANALYSES

3.1. Inleiding

Het doel van dit onderdeel is het afleiden van een dataset op basis van een groot aantal simulaties met PCLake voor de regressieanalyses, welke de basis vormen voor het metamodel PCLake. De basis voor de simulaties met PCLake vormt de dataset uit hoofdstuk 2. In dit hoofdstuk staan de keuzes die gemaakt zijn voor de simulaties met PCLake centraal:

1. bruikbaar criterium voor het afleiden van de kritische belasting;
2. methode van afleiding;
3. benodigde aanpassingen in het model PCLake.

3.2. Methode

Over het algemeen zijn de uitgangspunten voor het berekenen van de kritische belastingen in PCLake conform het proefschrift van Janse (2005): het model is toegepast met de gekalibreerde parameterwaarden over een tijdsperiode van 20 jaar en zowel vanuit een heldere als een troebele begintoestand. Afwijkingen op de door Janse gehanteerde methode zijn hieronder toegelicht.

criterium kritische belasting

De keuze van een criterium waarop kritische belasting wordt afgeleid is onderwerp van discussie: er zijn meerdere mogelijkheden (Janse, 2005, Witteveen+Bos 2006). Voor deze studie worden twee eisen gesteld aan een criterium voor het afleiden van de kritische belasting: het moet een generiek geldend criterium zijn en het moet geautomatiseerd kunnen worden bepaald. Beide eisen zijn gebaseerd op het feit dat er een aansturend programma geschreven wordt, waarmee de kritische grenzen geautomatiseerd bepaald worden.

De kritische belasting wordt traditioneel bepaald aan de hand van de afname van vegetatiebedekking (Janse 2005, Witteveen+Bos 2006): Janse heeft gebruik gemaakt van het criterium 20 % vegetatiebedekking, in het onderzoek van Witteveen+Bos is 5 % vegetatiebedekking gehanteerd. Omdat de bedekkingafname rond het omslagpunt groot is, is de invloed van het gekozen percentage op het te bepalen omslagpunt beperkt. Het criterium vegetatiebedekking is voor deze studie niet goed bruikbaar, omdat deze een niet-lineair (rond het omslagpunt) afnemend systeemspecifiek verband vertoont met toenemende belasting en deze afname niet eenduidig is. Het is (zoals ook in het onderzoek van Witteveen+Bos is geconstateerd) niet goed mogelijk een vaste waarde voor vegetatiebedekking te geven, waarbij in alle gevallen sprake is van een omslag. Een alternatief criterium voor het bepalen van de kritische belasting is de chlorofylconcentratie. De relatie tussen belasting en chlorofylconcentratie is vergelijkbaar met de relatie tussen belasting en vegetatiebedekking. Het verband is niet-lineair (toenemend rond het omslagpunt) en systeemspecifiek. De toename is niet eenduidig en daarom niet bruikbaar voor het bepalen van de kritische belasting.

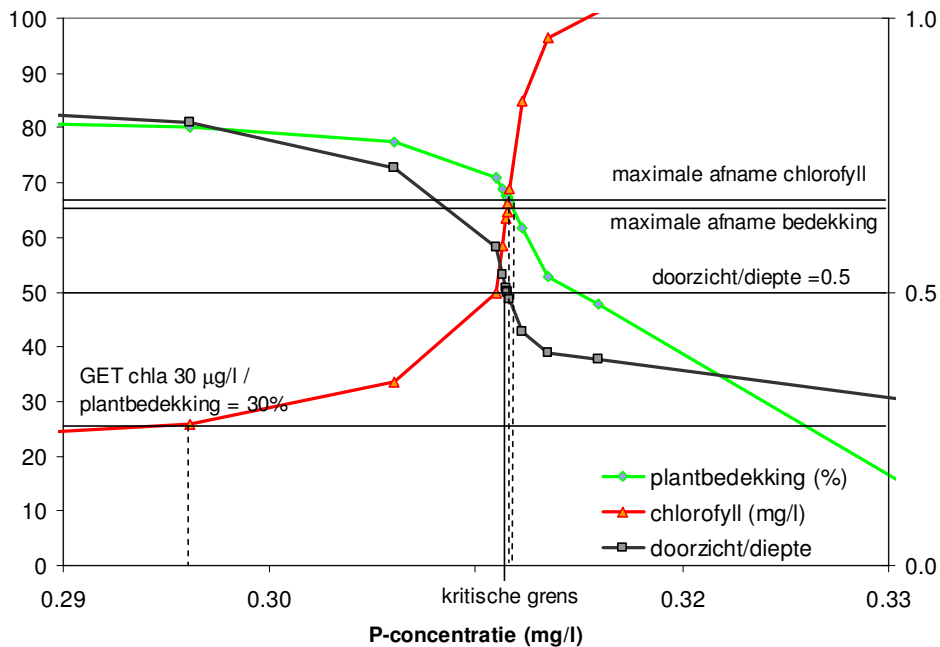
Tenslotte is doorzicht een mogelijk criterium voor het afleiden van de kritische belasting. Doorzicht (secchi-diepte) is een uitvoerparameter in PCLake. Ook de relatie tussen belasting en doorzicht is systeemspecifiek. Wanneer echter naar de fysisch/biologische achtergrond gekeken wordt kan doorzicht worden omgevormd tot een generiekere parameter. Ondergedoken waterplanten kunnen groeien als het licht de bodem bereikt: ofwel als de euphotische diepte (de diepte waarop 1 % van het invallende licht doordringt) gelijk is aan de waterdiepte. Empirisch is afgeleid dat de euphotische diepte ruwweg twee maal de secchi-diepte is. Hieruit volgt dat het omslagpunt kan worden bepaald bij:

$$\text{secchi-diepte/waterdiepte} = 0.5$$

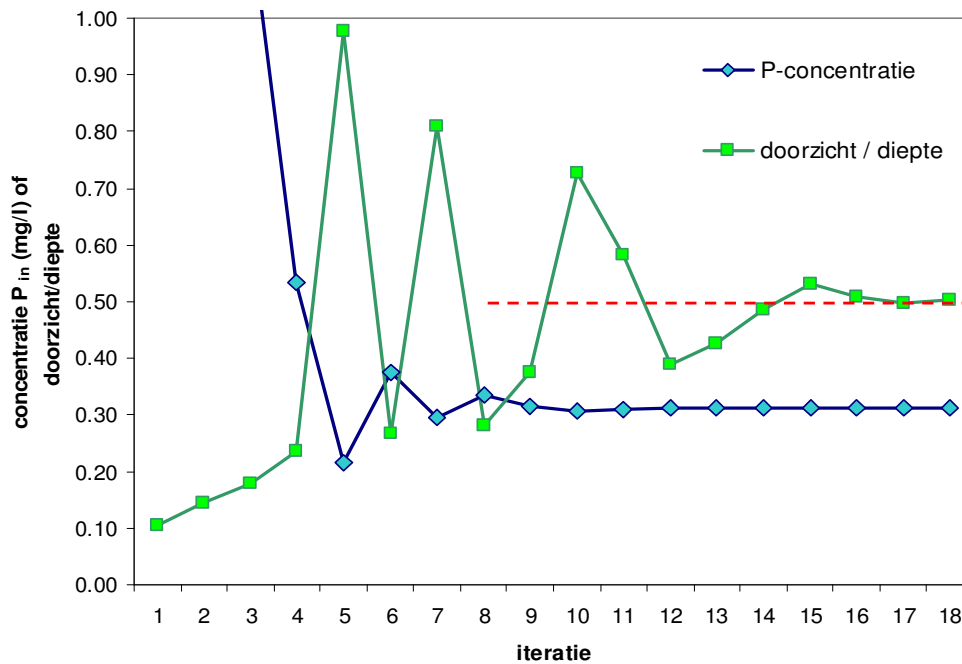
Modellsimulaties bevestigen dat dit een robuust generiek criterium is voor het afleiden van de kritische belasting. De op basis van dit criterium berekende kritische belasting komt overeen met de maximale afname van de vegetatiebedekking en met de maximale toename van de chlorofylconcentraties (afbeelding 3.1).

Het blijkt dat bij zeer kleine veranderingen in de belasting rond het omslagpunt de doorzicht/diepte-ratio nog duidelijk verandert (afbeelding 3.2). Dit betekent dat op dit punt zich een bistabiele situatie bevindt waarin beide toestanden in elkaar overgaan: kritische grens.

afbeelding 3.1. Voorbeeld van relatie tussen fosfaatconcentratie (inlaatwater) op x-as en plantbedekking, chlorofylconcentratie en doorzicht/diepte op y-as met criteria voor kritische grens



afbeelding 3.2. Verloop fosfaatconcentratie (inlaatwater) en doorzicht/diepte per iteratie: bij kleine concentratieveranderingen verandert ratio doorzicht/diepte nog flink



afleiden van kritische belasting

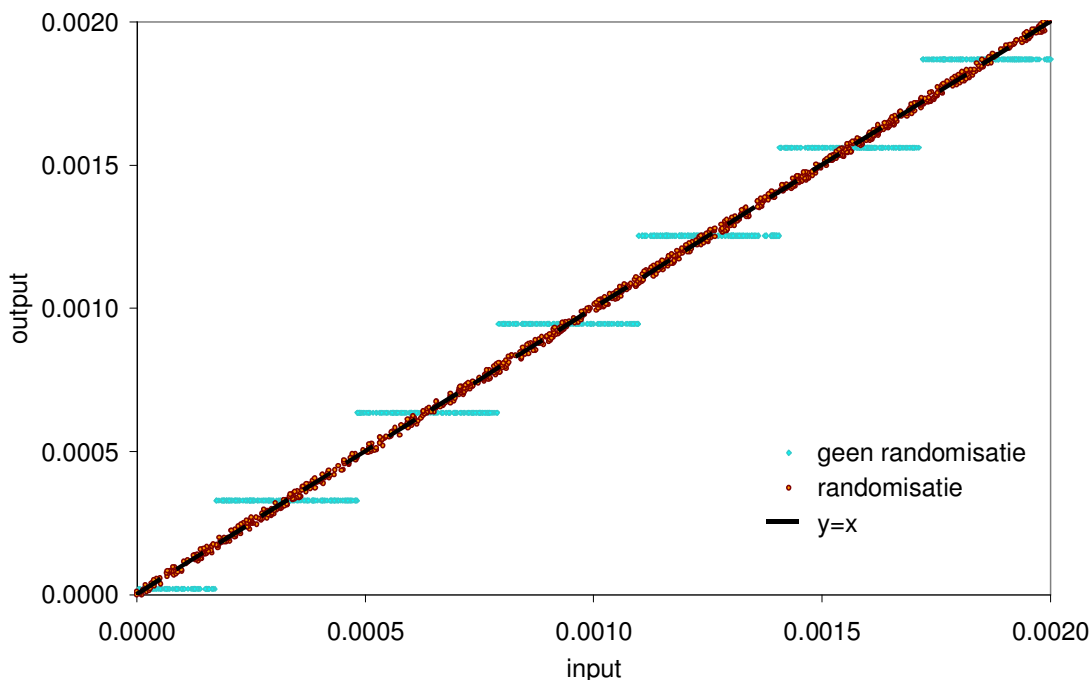
De kritische grenzen zijn geautomatiseerd bepaald aan de hand van het criterium doorzicht/diepte, waarbij uitgegaan is van een optimale combinatie van maximale nauwkeurigheid (precies mogelijk bepaling van de kritische belasting) en minimale inspanning (zo min mogelijk runs met PCLake). Dit heeft geresulteerd in een benadering, waarbij de kritische belasting zeer nauwkeurig wordt bepaald. De kritische belasting wordt bepaald in 18 stappen van grof naar fijn, waarbij steeds verder wordt ingezoomd op het gebied waar de omslag plaatsvindt.

In de eerste stap wordt de ratio doorzicht/diepte bepaald voor drie initiële belastingen (op basis van drie concentraties: 0,0001, 1 en 10 mg/l vermenigvuldigd met het ingaand debiet). In de tweede stap wordt voor één aanvullende belasting de ratio doorzicht/diepte bepaald in het midden van de laagste en middelste belasting of in het midden van de middelste en hoogste belasting, afhankelijk van in welk traject de omslag plaats vindt. Vervolgens wordt:

- in elke volgende stap een belasting gekozen tussen deze aanvullende belasting en de eerstvolgende eerder berekende (hogere) belasting als de ratio doorzicht/diepte de grens van 0,5 onderschrijft;
- in elke volgende stap een belasting gekozen tussen deze aanvullende belasting en de laatste eerder berekende (lagere) belasting als de ratio doorzicht/diepte de grens van 0,5 overschrijft.

De resultaten laten zien dat het trainen van de neurale netwerken gevoelig is voor een discrete verdeling in de dataset. In de gekozen methode wordt de kritische belasting nauwkeurig berekend door de schaal op te delen in steeds kleinere stukken. Ondanks de nauwkeurigheid blijft de verdeling van de berekende waarden discreet: een groot aantal uitkomsten, maar tussenliggende waarden kunnen niet worden bereikt. Ten behoeve van een continue verdeling is de methode aangepast: in de eerste stappen is een random factor geïntroduceerd, zodat niet precies het midden wordt genomen van twee voorgaande belastingen. Op basis van een aantal testen (waarbij initiële belastingen, het totaal aantal stappen, het aantal stappen met random factor en de random factor is gevarieerd) is de methode voor het berekenen van de kritische belasting voor de definitieve dataset verder geoptimaliseerd. In afbeelding 3.3 is het verschil weergegeven van een bepaling van de kritische belasting zonder aanpassingen en met gebruik van een random factor.

afbeelding 3.3. Effect van introductie van een random factor op berekende kritische belasting



aanpassingen model PCLake

Normaal gesproken worden de simulaties met PCLake uitgevoerd met een bepaalde (standaard) integratiemethode. Tijdens de simulaties op basis van deze methode zijn knelpunten naar voren gekomen. Na een specifiekere analyse van deze knelpunten blijkt dat:

- het model vast loopt bij specifieke sets van kenmerken;
- de rekentijd sterk oploopt bij specifieke sets van kenmerken (soms tot in de range van 10-25 minuten in plaats van ongeveer 7 seconden per iteratie).

De modeltechnische knelpunten zijn niet uitvoerig geanalyseerd, maar zeer waarschijnlijk zijn de knelpunten een gevolg van discontinuïteiten in de modelbenadering. Deze problemen zijn ondervangen door toepassing van een nauwkeurigere integratie algoritme (vijfde orde Runge-Kutta-Fehlberg in plaats van tweede orde Runge-Kutta-Fehlberg). Na toepassing van deze integratiemethode zijn de rektijden sterk verbeterd.

kader instabiele veenmeren

Het model wordt instabiel bij systemen met een zeer korte verblijftijd in combinatie met een veenbodem en (zeer) grote strijklengte. Nader onderzoek wijst uit dat deze instabiliteit ontstaat door de extreme afvoer van bodemmateriaal als gevolg van 1) de windwerking die voor veel resuspensie zorgt en 2) de afvoer van dit materiaal door het hoge debiet, waardoor het geresuspendeerde materiaal niet sedimenteert. Besloten is om meren met deze kenmerken uit de dataset te halen. In de praktijk kunnen deze systemen ook niet (langdurig) bestaan. In Nederland komen zulke systemen niet voor.

3.3. Resultaten

De kritische grenzen zijn berekend aan de hand van het doorzicht/diepte criterium met de halveringsmethode (zie hierboven). Er zijn 18 iteraties gebruikt, waarvan de eerste drie met een kleine random bijdrage van 1 %.

Er zijn simultaan 4 PCLake simulaties berekend (2 datasets x 2 begintoestanden) op een rekencomputer met 4 rekeneenheden. Na steeds 5000 berekeningen is getest of de aanvullende berekeningen van toegevoegde waarde zijn voor het neurale netwerk of niet.

Uiteindelijk zijn er 22295 sets van kritische grenzen berekend; c.q. de kritische belasting vanuit een heldere en troebele toestand voor een meer met unieke eigenschappen. Voordat de definitieve dataset is vastgesteld zijn de resultaten geanalyseerd. Voor 1549 gesimuleerde meren is gebleken dat de systemen na 20 jaar zelfs zonder belasting troebel worden of blijven, onafhankelijk van de initiële toestand. In deze systemen is dus nooit sprake van een door de nutriëntenbelasting geforceerde omslag. Dit was gezien de kenmerken verklaarbaar. De 1549 sets zijn uit de resultaten verwijderd. De definitieve dataset bestaat uit 20746 sets van kritische grenzen; c.q. de kritische belasting vanuit een heldere en troebele toestand (in totaal dus 41492 kritische grenzen op basis van 746.856 PCLake berekeningen). Deze complete dataset vereiste in totaal een rekentijd van 29 dagen (24 uur per dag).

De berekende dataset is gecontroleerd op consistentie door het originele regressiemodel van Janse (2005) te vergelijken met de gegenereerde dataset. Voor de kritische grens vanuit heldere toestand is de correlatie $R = 0.75$, voor de kritische grens vanuit troebel is de correlatie $R = 0.52$. De correlaties zijn vrij hoog, maar voornamelijk in het hogere bereik van kritische grenzen zit een grote discrepantie; hier zijn de originele regressiemodellen ook niet op geijkt. Ook is de extra variabele extinctie meegenomen in de huidige dataset; dit zorgt voor extra ruis ten opzichte van de originele regressiemodellen. De hier berekende dataset is dus consistent met het eerdere werk, maar bevat meer informatie vanwege het grotere bereik en volume (aantal datapunten) en de extra variabele extinctie.

4. METAMODEL

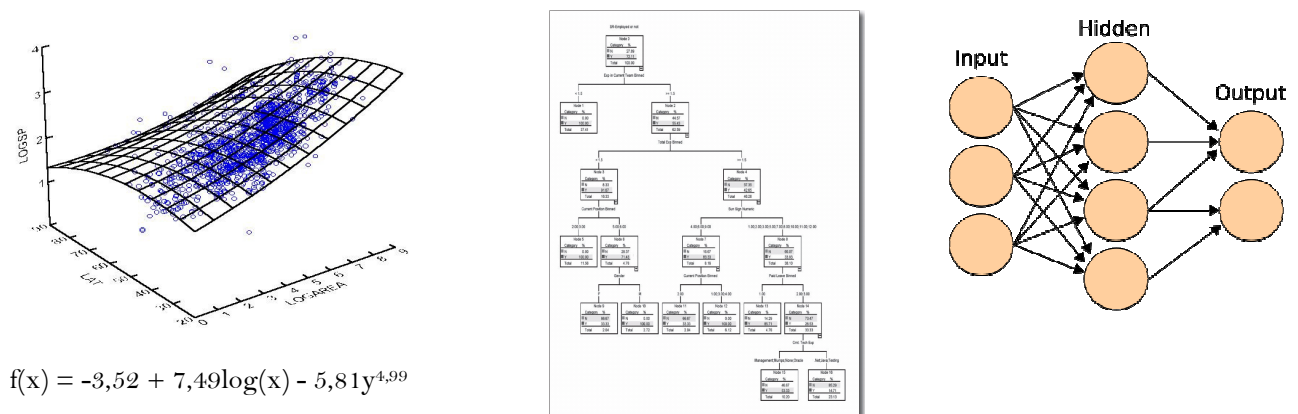
4.1. Inleiding

Het doel van dit onderdeel (en sluitstuk van het onderzoek) is een vergelijking van regressieanalyses, waarbij een zo goed mogelijk metamodel wordt verkregen voor het voorspellen van kritische grenzen van ondiepe meren. De resultaten van de regressieanalyses zijn vergeleken op basis van de criteria eenvoud, transparantie, betrouwbaarheid en bruikbaarheid.

4.2. Methode

Er is gebruik gemaakt van een multi-analyse aanpak, waarbij verschillende statistische technieken voor modelbouw zijn vergeleken, zowel veelgebruikte technieken als multiple regressie, regressiebomen, en black box neurale netwerken als de nieuwe techniek white box neurale netwerk. De technieken zijn hieronder kort toegelicht (zie ook afbeelding 4.1). Naast het white box netwerk is ook het "vereenvoudigde white-box netwerk" beschouwd. Hierbij is een snoei-algoritme toegepast op het netwerk. Elke methode wordt afzonderlijk toegepast voor de sedimenttypes en voor beide kritische grenzen. Dat resulteert in 6 verschillende modellen per methode.

afbeelding 4.1. Van links naar rechts: multi-pele regressie, regressieboom, neurale netwerk



Bij elke techniek worden de verklarende-, onafhankelijke- of invoervariabelen (onder andere diepte, strijklengte) gerelateerd aan de afhankelijke- of uitvoervariabele (kritische belasting). Voor de vergelijking van technieken is de dataset (uit hoofdstuk 3) opgedeeld in 2 delen: 5/6 deel is gebruikt om het model op te fitten (calibratie), het resterende 1/6 deel is gebruikt om de modellen op te testen (validatie). De invoer- en uitvoervariabelen zijn genormaliseerd (getransformeerd naar een verdeling met gemiddelde 1 en standaarddeviatie 0.2). Dit is een vereiste voor gebruik van deze technieken.

multiple regressie

Bij multi-pele regressie oefenen verscheidene onafhankelijke variabelen tegelijkertijd invloed uit op de afhankelijke variabele. Bij de berekening van de parameters wordt het effect van één variabele nagegaan terwijl dat van de andere variabelen constant wordt gehouden. Door het toevoegen en verwijderen van verklarende variabelen in het regressiemodel wordt een zo goed mogelijk verklarend (lineair) model gemaakt van de kritische belasting. Met behulp van de kleinste kwadratenmethode worden de parameters van het model bepaald. De voordelen van deze techniek zijn de simpliciteit en toepasbaarheid. Nadelen zijn de lineariteit van het model (invoergegevens hebben een sterk niet-lineair karakter) en het ontbreken van interactie-effecten. Beide nadelen zijn min of meer oplosbaar door bijvoorbeeld niet lineaire regressie te gebruiken en het van te voren definiëren van interactie-effecten, maar een combinatie van beide is hoogst onconventioneel.

regressieboom

Een regressieboom is een vertakkend model waarbij een (groot) aantal dichotome scheidingen in de invoerparameters leidt tot discrete waarden aan het uiteinden (takken) voor de te modelleren variabele (kritische belasting). Aan de hand van een regressie-analyse wordt de boom opgesteld. Dit is een veel gebruikt model vanwege het gemak in het gebruik en de inzichtelijkheid van de onderliggende relaties. Een groot nadeel is het uitkomen op discrete waarden, hierdoor is het minder geschikt voor het modelleren van continue verdelingen. Een bijkomstigheid hierbij is dat bij het bouwen van het model op een continue verdeling er vaak een zeer groot aantal vertakkingen gebruikt wordt, zodat iedere in de dataset voorkomende waarde een eigen eindtakje krijgt. Hierdoor kan het model erg goed passen op de dataset waarop getraind wordt, maar weinig voorspellende waarde hebben bij het aanbieden van nieuwe datapunten. Bij een dergelijk groot aantal vertakking gaat het aanvankelijke voordeel van inzichtelijkheid eveneens verloren.

neurale netwerken

Neurale netwerken zijn (kunstmatige) netwerken gebaseerd op de werking van zenuwcellen, waarbij punten (neuronen) onderling verbonden zijn. De connecties tussen de verschillende neuronen kunnen worden aangepast door het netwerk te trainen op een dataset. Er is gebruik gemaakt van twee technieken voor het bouwen van neurale netwerken: een 'black-box'-neuraal netwerk en een 'white-box'-neuraal netwerk (zie hieronder). Voor de neurale netwerken is de dataset verder opgesplitst in een trainingsset, een validatieset en een testset. De testset (1/6 deel van dataset) is dezelfde als gebruikt voor de andere technieken en wordt alleen achteraf gebruikt voor het bepalen van de kwaliteit van het model. Verder wordt 2/3 van de totale dataset gebruikt voor het trainen en het overige 1/6 deel voor het valideren van het neurale netwerk tijdens het trainen van de netwerken. Het tussentijds valideren bepaalt de keuzes in het trainen van de neurale netwerken en voorkomt het overtrainen (dit is een conditie waarbij het netwerk de trainingsset heel goed voorspelt, maar erg slecht presteert op nieuwe data).

Bij neurale netwerken moet gekozen worden voor een aantal verborgen knopen. Voor zowel black- als white-box netwerken is dit aantal 7. Dit is een vrij willekeurig aantal. De resultaten veranderen niet wezenlijk als er voor meer of minder verborgen knopen wordt gekozen. Voor het snoeien van het white-box netwerk geldt het getrainde white-box netwerk als uitgangspunt en wordt een beperkte toename van de fout geaccepteerd.

black-box neuraal netwerk

Een black-box neuraal netwerk is een techniek waarbij een neuraal netwerk wordt gebouwd op basis van gewogen sommen van inputs en transferfuncties. Het is een 'black-box' aangezien de uitkomst dusdanig complex is dat interpretatie onmogelijk is. Ook is het niet mogelijk dit soort netwerken te vereenvoudigen. Wel geeft deze techniek over het algemeen de beste prestaties.

white-box neuraal netwerk

Product-unit networks (PUN, white-box neuraal netwerk) kunnen beter relaties voorspellen die machten en delingen bevatten. Aangezien de verwachting is dat bij een sterk niet-lineair model als PCLake de achterliggende relaties soortgelijke verbanden vertonen ligt het voor de hand om deze methode te gebruiken. Een bijkomend voordeel is dat in tegenstelling tot een black-box neuraal netwerk de uitkomsten relatief simpel kunnen zijn, wat de interpretatie van de uitkomsten mogelijk kan maken. Een ander voordeel is dat de geproduceerde netwerken achteraf vereenvoudigd (gesnoeid) kunnen worden door de minst significante parameters weg te strepen, waardoor een beter interpreteerbaar resultaat overblijft.

4.3. Resultaten

De modellen zijn op basis van vier indicatoren beoordeeld: informatie inhoud, verklarende variantie, relatieve fout en standaard deviatie van de fout. Alle waarden in de tabellen in dit hoofdstuk hebben betrekking op de testset.

informatie inhoud

De informatie inhoud van het model is het aantal parameters dat het model vastlegt. Het aantal parameters per model is terug te vinden in tabel 4.1. De lineaire regressie modellen zijn het eenvoudigst, ze bevatten één parameter voor elke verklarende variabele en één constante: in totaal 6. Neurale netwerken bevatten veel meer parameters en regressiebomen zijn nog weer uitgebreider. Het effect van snoeien (vereenvoudigen) op het white-box netwerk is duidelijk zichtbaar.

tabel 4.1. Informatie inhoud

initiële conditie	sediment type	regressieboom	lineaire regressie	neuraal netwerk [5 7 1]	white-box netwerk [5 7 1]	vereenvoudigd white-box netwerk [5 7 1]
helder	klei	287	6	50	43	16
troebel	klei	309	6	50	43	25
helder	veen	201	6	50	43	26
troebel	veen	227	6	50	43	27
helder	zand	461	6	50	43	27
troebel	zand	481	6	50	43	31

verklarende variantie

De verklarende variantie is ook wel bekend als R-kwadraat. In tabel 4.2 is de verklarende variantie van alle modellen weergegeven. Dit getal geeft de correlatie weer tussen de werkelijke waarde en de waarde die het model berekend. Een hoge R-kwadraat zegt niet alles, omdat de waarde mede afhangt van de grootte van de testset. De neurale netwerken scoren duidelijk beter dan de regressieboom en lineaire regressie. De black-box netwerken halen met meer dan 96 % de hoogste R-kwadraat, maar het verschil met de white-box netwerken is zeer beperkt.

tabel 4.2. Verklarende variantie of R-kwadraat

initiële conditie	sediment type	regressieboom	lineaire regressie	neuraal netwerk [5 7 1]	white-box netwerk [5 7 1]	vereenvoudigd white-box netwerk [5 7 1]
helder	klei	0,83	0,76	0,96	0,95	0,94
troebel	klei	0,85	0,78	0,98	0,97	0,96
helder	veen	0,76	0,72	0,96	0,95	0,91
troebel	veen	0,80	0,70	0,96	0,95	0,93
helder	zand	0,90	0,81	0,99	0,99	0,99
troebel	zand	0,91	0,84	0,98	0,97	0,97

relatieve fout

Als het meta-model in de KRW-verkenner een kritische grens berekend, zal de gebruiker geïnteresseerd zijn in de relatieve nauwkeurigheid van die waarde. Hoe groot is de onzekerheid in de uitkomst van het model? Om dat inzichtelijk te maken is voor elk model de gemiddelde relatieve fout berekend. De absolute fout (het verschil tussen de voorspelling en de waarde van PCLake) is daarbij gedeeld door de waarde die het model voorspelt. De som van deze fouten is gemiddeld. De resulterende factor geeft een indicatie van de onzekerheid in de uitkomst. Een waarde van 0,2 geeft aan dat de werkelijk waarde 20 % kan afwijken van de voorspelde waarde.

In tabel 4.3 is de relatieve fout voor alle modellen terug te vinden. Wat direct opvalt is dat de gemiddelde fout bij de regressieboom voor sedimenttype klei en veen onacceptabel groot is (60-380 %). De prestaties voor zand komen dichterbij de gewenste range. Ook de fouten van de lineaire regressiemodellen zijn erg groot (40-85 %). Alleen de neurale netwerken, zowel black-box als (vereenvoudigd) white-box, zijn acceptabel. Met een fout in de orde van 20 % is de relatieve fout van deze neurale netwerken kleiner dan de modelonzekerheid in PCLake zelf (Janse gebruikt een onzekerheid van ongeveer 40 %). Binnen de prestaties van de neurale netwerken valt op dat de kritische belasting van helder naar

troebel bij zand het best voorspeld wordt: de fout is slechts 8 %. Black- en white-box ontlopen elkaar niet veel qua prestaties.

Over het algemeen neemt de fout bij vereenvoudiging van het white-box netwerk licht toe. Dat komt deels doordat dat in het snoei-algoritme is toegestaan ten gunste van vereenvoudiging van de formule.

tabel 4.3. Relatieve fout

initiële conditie	sediment type	regressieboom	lineaire regressie	neuraal netwerk [5 7 1]	white-box netwerk [5 7 1]	vereenvoudigd white- box netwerk [5 7 1]
helder	klei	0,64	0,50	0,17	0,19	0,22
troebel	klei	2,15	0,79	0,19	0,22	0,24
helder	veen	3,84	0,69	0,21	0,21	0,28
troebel	veen	2,15	0,85	0,24	0,24	0,32
helder	zand	0,28	0,41	0,08	0,07	0,08
troebel	zand	0,41	0,62	0,16	0,21	0,22

standaarddeviatie van de fout

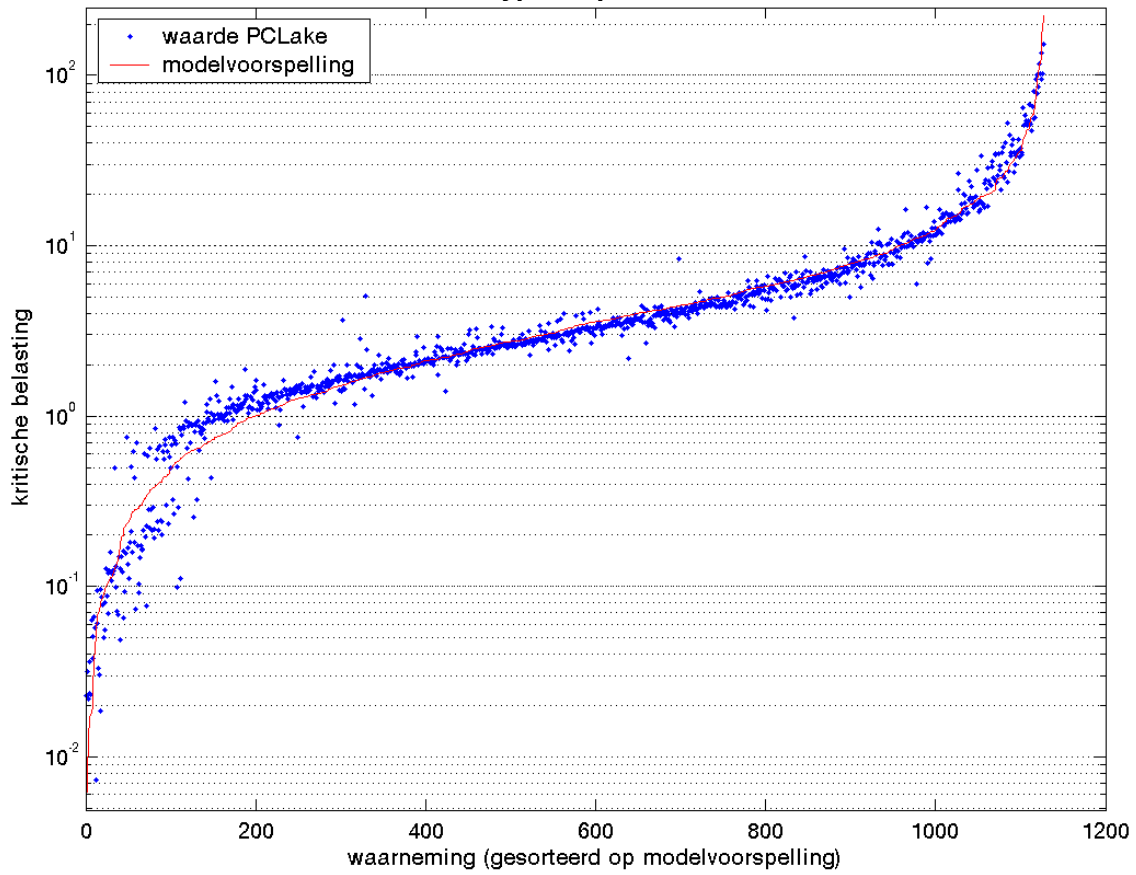
De relatieve fout zegt niet alles over de prestaties van het model. Het betreft een gemiddelde en daarin wegen grote uitschieters (missers) niet zwaar. Daarom is het van belang ook te kijken naar de spreiding van de fout. Een maat voor die spreiding is de standaarddeviatie: het kwadratisch gemiddelde van de fout. Een grote standaard deviatie betekent dat er veel grote afwijkingen voorkomen. In tabel 4.4 zijn de standaarddeviaties gegeven van de relatieve fout.

Opvallend is dat opnieuw de regressieboom opnieuw slecht presteert voor systemen met een klei- en veenbodem. Ook de spreiding bij lineaire regressie is onacceptabel groot. Voor beide methoden geldt overigens dat de relatieve fout en de standaard deviatie op de trainingsset aanzienlijk kleiner zijn. Dat betekent dat beide methoden moeilijk met nieuwe data kan omgaan. Uit de tabel blijkt verder dat de spreiding bij white-box netwerken kleiner is dan bij black-box netwerken. Na vereenvoudiging is de spreiding vergelijkbaar. Dat betekent dat de white-box netwerken het niet-lineaire gedrag over de gehele parameterruimte iets beter simuleert.

tabel 4.4. Standaard deviatie van de fout

initiële conditie	sediment type	regressieboom	lineaire regressie	neuraal netwerk [5 7 1]	white-box netwerk [5 7 1]	vereenvoudigd white- box netwerk [5 7 1]
helder	klei	8,15	0,74	0,37	0,39	0,38
troebel	klei	47,54	1,09	0,41	0,32	0,39
helder	veen	102,84	1,18	0,58	0,38	0,48
troebel	veen	35,29	1,22	0,63	0,42	0,63
helder	zand	0,38	0,60	0,12	0,11	0,15
troebel	zand	0,78	0,89	0,25	0,27	0,29

afbeelding 4.2. Vergelijking PCLake en vereenvoudigd white-box netwerk voor helder initiële conditie en sediment type 1 op basis van testset



5. CONCLUSIES

Het doel van dit onderzoek is het ontwikkelen van een eenvoudig, transparant, betrouwbaar en bruikbaar metamodel van het ecologisch model PCLake.

De regressiebomen doen het op de trainingset heel aardig. Echter zoals uit de tabellen blijkt, vallen de prestaties op de testset erg tegen. In iets mindere mate geldt dat voor de lineaire regressiemodellen. De neurale netwerken komen als beste uit de bus. Het verschil tussen white- en black-box is daarbij beperkt. White-box netwerken verdienen de voorkeur vanwege een betere spreiding van de fout, de eenvoudige implementatie en de mogelijkheid de netwerken te interpreteren.

Het vereenvoudigen van het white-box netwerk levert niet heel veel op. Door de transformaties vooraf en achteraf wordt de interpretatie bemoeilijkt. Bovendien blijkt de vorm van vereenvoudigde formules sterk afhankelijk van de initiële schatting. Bij herhaald uitvoeren van het experiment kan de formule behoorlijk verschillen, terwijl de prestaties vergelijkbaar zijn. Het is dan ook het meest praktisch het volledige white-box netwerk te gebruiken.

De conclusie is dat het volledige white-box neurale netwerk het beste voldoet aan de eisen en daarom ingebouwd wordt in de KRW-verkenner.

6. LITERATUURLIJST

1. Jaarsma, N.G., M. Klinge & L. Lamers, 2008. Van helder naar troebel... en weer terug. Een ecologische systeemanalyse en diagnose van ondiepe meren en plassen voor de Kaderrichtlijn Water. STOWA-rapport 2008-04.
2. Janse, J.H. 2005. Model studies on the eutrophication of shallow lakes and ditches. Dissertatie. Universiteit Wageningen.
3. Schep, S. 2006. Interacties tussen stuurvariabelen voor ecologische doelen in meren, fase 2: analyse van simulaties. Witteveen+Bos, rapport nr. BHV24-1-1.
4. Vleeshouwers LM, Janse JH, Aldenberg T, Knoop JM. 2004. A metamodel for PCLake. RIVM rapport 703715007. Bilthoven.

BIJLAGE I Karakteristieken Nederlandse en Europese meren en plassen

KARAKTERISTIEKEN NEDERLANDSE EN EUROPESE MEREN EN PLASSEN

**tabel I.1. Overzicht kenmerken Europese meren en plassen (uit proefschrift Jan Janse (2005):
Model studies on the eutrophication of shallow lakes and ditches, tabel 4.3)**

nr.	naam	land	diepte (m)	strijklengte (m)	fractie moeras (-)	sedimenttype (-)	debiet (mm/d)
1.	Hollands-Ankeveen	NL	1.3	922	0.1	5	3.5
2.	Bergse Achterplas	NL	2	640	0.001	1	19.3
3.	Bergse Voorplas	NL	2	775	0.001	1	21.1
4.	Beulakerwijde	NL	1.8	3606	0.001	5	10
5.	Botshol	NL	2	731	0.1	5	12.9
6.	Kaag	NL	2.4	1789	0.001	1	109.6
7.	Langeraar	NL	1.9	529	0.001	5	4.4
8.	Loosdrecht	NL	2.2	3065	0.001	5	8.4
9.	Naarden	NL	1.3	1304	0.1	6	10
10.	Naarden	NL	1	612	0.1	6	5.2
11.	Reeuwijk	NL	1.8	1020	0.001	5	5.5
12.	Nieuwkoop	NL	3	1225	0.001	5	6.3
13.	Nieuwkoop	NL	3	1000	0.001	5	27.4
14.	Reeuwijk	NL	2	1044	0.001	5	16.4
15.	Westeinderplassen	NL	2.8	2919	0.001	6	20.7
16.	Zuidlaardermeer	NL	1	2449	0.001	6	43.5
17.	Nanneveld	NL	1	1000	0.001	5	6.8
18.	Blankaart	B	1	548	0.001	1	71.4
19.	Maten	B	0.8	171	0.001	6	92
20.	Visvijver	B	0.8	86	0.001	1	3
21.	Gara	IR	1	1425	0.001	5	137
22.	Maumwee	IR	2	520	0.001	5	54.8
23.	Mullagh	IR	2.3	592	0.001	5	4.7
24.	Pollaphuca	IR	6.8	4443	0.001	5	33.3
25.	Amor	IR	3	2722	0.001	5	48.3
26.	Luknajno	PL	1.8	2510	0.1	6	3.2
27.	Waalboezem	NL	3.8	883	0.001	1	20
28.	Binnenbedijkte	NL	4	1257	0.001	1	14.6
29.	Brielse	NL	5.5	2216	0.001	1	40.7
30.	Volkerak	NL	5	6760	0.001	6	45.9
31.	Zoommeer	NL	6	3975	0.001	6	105.6
32.	Oude	NL	1.3	316	0.001	5	17.1
33.	Veluwemeer	NL	1.6	5692	0.05	2	26.7
34.	Wolderwijd	NL	1.9	4243	0.001	6	16.5
35.	Nuldernauw	NL	2.1	2950	0.001	6	47
36.	Drontermeer	NL	1.3	2324	0.05	6	69
37.	Braassemmermeer	NL	3.9	2126	0.001	1	62.9
38.	Langeraar	NL	1.9	866	0.001	5	3.4
39.	Mooie	NL	4.5	1077	0.001	6	88.1
40.	Het	NL	1	548	0.1	5	6.4
41.	Kortenhoef	NL	1.2	1389	0.001	5	5.9
42.	Stichts-Ankeveen	NL	1	1000	0.1	5	3.3
43.	Frisian	NL	1.4	2276	0.001	5	9.3
44.	Enso	DK	1.9	326	0.001	3	27.1
45.	Nederso	DK	1.6	374	0.001	3	320

nr.	naam	land	diepte (m)	strijklengte (m)	fractie moeras (-)	sedimenttype (-)	debiet (mm/d)
46.	Soby	DK	3	849	0.001	6	10.1
47.	Hinge	DK	1.2	954	0.001	3	60
48.	Kvind	DK	1.9	387	0.001	6	126.7
49.	Lading	DK	1	671	0.001	3	14.7
50.	Silkeborg-Langso	DK	2	678	0.001	6	250
51.	Honda	SP	1	307	0.001	3	7.8
52.	Nueva	SP	1.5	521	0.001	3	1.8

BIJLAGE II Formules (gesnoeide) PUN's

FORMULES (GESNOEIDE) PUN'S

grens = helder, sedtype = klei

transformatie vooraf

$$\text{SDEPTHw0_t} = (\text{diepte} - 1.786928) * (0.200000 / 0.801227) + 1.000000$$

$$\text{FMARSH_t} = (\text{aandeel moeras} - 0.112454) * (0.200000 / 0.331317) + 1.000000$$

$$\text{CFETCH_t} = (\text{strijklengte} - 1464.130344) * (0.200000 / 2135.478675) + 1.000000$$

$$\text{CQIN_t} = (\text{debiet} - 42.101943) * (0.200000 / 44.251998) + 1.000000$$

$$\text{CEXTWAT_t} = (\text{extinctie} - 0.657166) * (0.200000 / 0.380499) + 1.000000$$

volledig netwerk

$$\text{kP_gm2j_transformed} =$$

$$+ 0.9067$$

$$- 0.0061812 * [\text{SDEPTHw0_t}]^8.3927 * [\text{FMARSH_t}]^0.21586 * [\text{CFETCH_t}]^3.5231 * [\text{CQIN_t}]^{-1.1053} * [\text{CEXTWAT_t}]^8.9234$$

$$+ 0.0050354 * [\text{SDEPTHw0_t}]^2.1925 * [\text{FMARSH_t}]^{-0.21099} * [\text{CFETCH_t}]^{-19.457} * [\text{CQIN_t}]^{-2.1913} * [\text{CEXTWAT_t}]^2.3951$$

$$- 0.59266 * [\text{SDEPTHw0_t}]^2.0814 * [\text{FMARSH_t}]^{-5.5461} * [\text{CFETCH_t}]^1.9944 * [\text{CQIN_t}]^{-1.9833} * [\text{CEXTWAT_t}]^0.69827$$

$$- 0.023766 * [\text{SDEPTHw0_t}]^2.4015 * [\text{FMARSH_t}]^{-0.39138} * [\text{CFETCH_t}]^1.6849 * [\text{CQIN_t}]^{-10.484} * [\text{CEXTWAT_t}]^1.1343$$

$$+ 0.082983 * [\text{SDEPTHw0_t}]^{-0.31908} * [\text{FMARSH_t}]^{-2.0768} * [\text{CFETCH_t}]^2.5077 * [\text{CQIN_t}]^{-1.4805} * [\text{CEXTWAT_t}]^1.7717$$

$$+ 0.16538 * [\text{SDEPTHw0_t}]^{-2.0853} * [\text{FMARSH_t}]^0.26822 * [\text{CFETCH_t}]^{-0.53038} * [\text{CQIN_t}]^1.0786 * [\text{CEXTWAT_t}]^{-0.29245}$$

$$+ 0.49623 * [\text{SDEPTHw0_t}]^2.3944 * [\text{FMARSH_t}]^{-5.4067} * [\text{CFETCH_t}]^1.9596 * [\text{CQIN_t}]^{-1.7684} * [\text{CEXTWAT_t}]^0.51266$$

vereenvoudigd netwerk

$$\text{kP_gm2j_transformed} =$$

$$+ 0.57192$$

$$+ 0.0057305 * [\text{CFETCH_t}]^{-21.751}$$

$$- 0.065134 * [\text{SDEPTHw0_t}]^2.1193 * [\text{FMARSH_t}]^{-2.1569} * [\text{CFETCH_t}]^1.7598 * [\text{CQIN_t}]^{-6.6734}$$

$$- 0.021184 * [\text{SDEPTHw0_t}]^6.2776 * [\text{CEXTWAT_t}]^5.6054$$

$$+ 0.33652 * [\text{SDEPTHw0_t}]^{-1.7635} * [\text{CQIN_t}]^0.61266$$

$$+ 0.18043 * [\text{SDEPTHw0_t}]^2.1091$$

transformatie achteraf

$$\text{kP_gm2j} = 10^{[(\text{kP_gm2j_transformed} - 1.000000) * (0.627249 / 0.200000) + 0.434686]}$$

grens = troebel, sedtype = klei

transformatie vooraf

$SDEPTHw0_t = (SDEPTHw0 - 1.786928) * (0.200000 / 0.801227) + 1.000000$
 $FMARSH_t = (FMARSH - 0.112454) * (0.200000 / 0.331317) + 1.000000$
 $CFETCH_t = (CFETCH - 1464.130344) * (0.200000 / 2135.478675) + 1.000000$
 $CQIN_t = (CQIN - 42.101943) * (0.200000 / 44.251998) + 1.000000$
 $CEXTWAT_t = (CEXTWAT - 0.657166) * (0.200000 / 0.380499) + 1.000000$

volledig netwerk

$kP_gm2j_transformed =$
- 5.4226
- $0.00052376 * [SDEPTHw0_t]^{12.825} * [FMARSH_t]^{0.44296} * [CFETCH_t]^{3.8726} * [CQIN_t]^{-1.7617} * [CEXTWAT_t]^{14.18}$
+ $0.23997 * [SDEPTHw0_t]^{1.7157} * [FMARSH_t]^{-0.22844} * [CFETCH_t]^{-0.14677} * [CQIN_t]^{0.18964} * [CEXTWAT_t]^{-0.64702}$
+ $0.00022222 * [SDEPTHw0_t]^{11.603} * [FMARSH_t]^{0.99325} * [CFETCH_t]^{-6.9781} * [CQIN_t]^{-2.0934} * [CEXTWAT_t]^{12.55}$
+ $0.71822 * [SDEPTHw0_t]^{-2.0857} * [FMARSH_t]^{-5.821} * [CFETCH_t]^{1.4495} * [CQIN_t]^{-10.326} * [CEXTWAT_t]^{0.50542}$
+ $6.1895 * [SDEPTHw0_t]^{-0.17262} * [FMARSH_t]^{0.02238} * [CFETCH_t]^{-0.014108} * [CQIN_t]^{0.058737} * [CEXTWAT_t]^{0.014885}$
- $0.73309 * [SDEPTHw0_t]^{-1.9933} * [FMARSH_t]^{-5.8968} * [CFETCH_t]^{1.4433} * [CQIN_t]^{-10.431} * [CEXTWAT_t]^{0.50493}$
+ $0.0029928 * [SDEPTHw0_t]^{-0.43129} * [FMARSH_t]^{-0.31691} * [CFETCH_t]^{-25.124} * [CQIN_t]^{-2.2178} * [CEXTWAT_t]^{-0.61153}$

vereenvoudigd netwerk

$kP_gm2j_transformed =$
- 16.734
- $0.0020376 * [SDEPTHw0_t]^8.8766 * [CEXTWAT_t]^9.0082$
+ $0.24203 * [SDEPTHw0_t]^6.0532$
- $0.17457 * [SDEPTHw0_t]^6.5809$
+ $1.6051 * [SDEPTHw0_t]^{-1.6059} * [FMARSH_t]^{-4.9553} * [CFETCH_t]^{1.8152} * [CQIN_t]^{-8.4341} * [CEXTWAT_t]^{0.96146}$
+ $17.691 * [SDEPTHw0_t]^{-0.054558} * [CQIN_t]^{0.016408}$
- $1.6466 * [SDEPTHw0_t]^{-1.5458} * [FMARSH_t]^{-4.9843} * [CFETCH_t]^{1.8078} * [CQIN_t]^{-8.4632} * [CEXTWAT_t]^{0.95538}$
+ $0.0095071 * [CFETCH_t]^{-19.543}$

transformatie achteraf

$kP_gm2j = 10^{[(kP_gm2j_transformed - 1.000000) * (0.776714 / 0.200000) + 0.102223]}$

grens = helder, sedtype = veen

transformatie vooraf

$SDEPTHw0_t = (SDEPTHw0 - 1.786928) * (0.200000 / 0.801227) + 1.000000$
 $FMARSH_t = (FMARSH - 0.112454) * (0.200000 / 0.331317) + 1.000000$
 $CFETCH_t = (CFETCH - 1464.130344) * (0.200000 / 2135.478675) + 1.000000$
 $CQIN_t = (CQIN - 42.101943) * (0.200000 / 44.251998) + 1.000000$
 $CEXTWAT_t = (CEXTWAT - 0.657166) * (0.200000 / 0.380499) + 1.000000$

volledig netwerk:

$kP_gm2j_transformed =$
 $+ 0.95763$
 $- 4.5185 * [SDEPTHw0_t]^1.2181 * [FMARSH_t]^0.1908 * [CFETCH_t]^1.0106 * [CQIN_t]^{-0.73621} * [CEXTWAT_t]^{-0.41899}$
 $+ 0.019613 * [SDEPTHw0_t]^1.1749 * [FMARSH_t]^{-0.20873} * [CFETCH_t]^{-15.131} * [CQIN_t]^{-1.7956} * [CEXTWAT_t]^0.93784$
 $+ 1.9774 * [SDEPTHw0_t]^{-0.1252} * [FMARSH_t]^0.19852 * [CFETCH_t]^1.1211 * [CQIN_t]^{-0.28654} * [CEXTWAT_t]^{-0.15168}$
 $+ 3.6284 * [SDEPTHw0_t]^2.4372 * [FMARSH_t]^0.2607 * [CFETCH_t]^0.586 * [CQIN_t]^{-0.82723} * [CEXTWAT_t]^{-0.90324}$
 $- 1.0785 * [SDEPTHw0_t]^3.3716 * [FMARSH_t]^0.37095 * [CFETCH_t]^{-0.0089113} * [CQIN_t]^{-0.8009} * [CEXTWAT_t]^{-1.3627}$
 $- 0.041754 * [SDEPTHw0_t]^2.0865 * [FMARSH_t]^{-7.6877} * [CFETCH_t]^1.9182 * [CQIN_t]^{-7.6234} * [CEXTWAT_t]^0.71715$
 $- 0.011163 * [SDEPTHw0_t]^7.0714 * [FMARSH_t]^0.32725 * [CFETCH_t]^2.6787 * [CQIN_t]^{-0.42661} * [CEXTWAT_t]^7.4437$

vereenvoudigd netwerk

$kP_gm2j_transformed =$
 $+ 0.21748$
 $- 4.7225 * [SDEPTHw0_t]^1.3826 * [CFETCH_t]^1.0561 * [CQIN_t]^{-0.68117} * [CEXTWAT_t]^{-0.43504}$
 $+ 0.025867 * [CFETCH_t]^{-14.806}$
 $+ 1.8331 * [SDEPTHw0_t]^{-0.3667} * [CFETCH_t]^0.81256 * [CQIN_t]^{-0.25287}$
 $+ 4.1516 * [SDEPTHw0_t]^1.9303 * [FMARSH_t]^0.13479 * [CFETCH_t]^0.77732 * [CQIN_t]^{-0.48323} * [CEXTWAT_t]^{-0.94691}$
 $- 0.50027 * [SDEPTHw0_t]^2.217 * [FMARSH_t]^0.5507 * [CEXTWAT_t]^{-2.8569}$
 $- 0.02202 * [CQIN_t]^{-9.6384}$
 $- 0.069383 * [SDEPTHw0_t]^5.7835$

transformatie achteraf

$kP_gm2j = 10^{[(kP_gm2_transformed - 1.000000) * (0.627249 / 0.200000) + 0.434686]}$

grens = troebel, sedtype = veen

transformatie vooraf

$SDEPTHw0_t = (SDEPTHw0 - 1.786928) * (0.200000 / 0.801227) + 1.000000$
 $FMARSH_t = (FMARSH - 0.112454) * (0.200000 / 0.331317) + 1.000000$
 $CFETCH_t = (CFETCH - 1464.130344) * (0.200000 / 2135.478675) + 1.000000$
 $CQIN_t = (CQIN - 42.101943) * (0.200000 / 44.251998) + 1.000000$
 $CEXTWAT_t = (CEXTWAT - 0.657166) * (0.200000 / 0.380499) + 1.000000$

volledig netwerk:

$kP_gm2j_transformed =$
 $+ 1.1021$
 $+ 1.673 * [SDEPTHw0_t]^2 * 1.25 * [FMARSH_t]^{-0.13121} * [CFETCH_t]^2 * 1.878 * [CQIN_t]^{-1.9584} * [CEXTWAT_t]^{-0.58745}$
 $- 0.0013414 * [SDEPTHw0_t]^{11.306} * [FMARSH_t]^{-0.18649} * [CFETCH_t]^2 * 0.0914 * [CQIN_t]^{-1.1453} * [CEXTWAT_t]^{12.86}$
 $+ 0.12408 * [SDEPTHw0_t]^{-2.9962} * [FMARSH_t]^{0.24861} * [CFETCH_t]^{1.3157} * [CQIN_t]^{0.76494} * [CEXTWAT_t]^{0.13236}$
 $- 1.8881 * [SDEPTHw0_t]^{1.9885} * [FMARSH_t]^{-0.16189} * [CFETCH_t]^2 * 1.722 * [CQIN_t]^{-1.9987} * [CEXTWAT_t]^{-0.44454}$
 $+ 0.0063297 * [SDEPTHw0_t]^{-1.7807} * [FMARSH_t]^{-0.96351} * [CFETCH_t]^{-15.96} * [CQIN_t]^{-5.203} * [CEXTWAT_t]^{-0.26179}$
 $+ 0.00042493 * [SDEPTHw0_t]^{10.687} * [FMARSH_t]^{-0.35468} * [CFETCH_t]^{-8.2694} * [CQIN_t]^{-1.2535} * [CEXTWAT_t]^{11.967}$
 $- 0.010862 * [SDEPTHw0_t]^{0.050467} * [FMARSH_t]^{-8.9521} * [CFETCH_t]^{-0.70947} * [CQIN_t]^{-14.884} * [CEXTWAT_t]^{0.21327}$

vereenvoudigd netwerk

$kP_gm2j_transformed =$
 $+ 1.0706$
 $+ 1.673 * [SDEPTHw0_t]^2 * 1.25 * [FMARSH_t]^{-0.13121} * [CFETCH_t]^2 * 1.878 * [CQIN_t]^{-1.9584} * [CEXTWAT_t]^{-0.58745}$
 $- 0.0010105 * [SDEPTHw0_t]^{11.306} * [CEXTWAT_t]^{12.86}$
 $+ 0.13578 * [SDEPTHw0_t]^{-2.9962} * [CFETCH_t]^{1.3157}$
 $- 1.8881 * [SDEPTHw0_t]^{1.9885} * [FMARSH_t]^{-0.16189} * [CFETCH_t]^2 * 1.722 * [CQIN_t]^{-1.9987} * [CEXTWAT_t]^{-0.44454}$
 $+ 0.012733 * [CFETCH_t]^{-15.96}$
 $+ 0.0011218 * [SDEPTHw0_t]^{10.687} * [CEXTWAT_t]^{11.967}$
 $- 0.011124 * [FMARSH_t]^{-8.9521} * [CQIN_t]^{-14.884}$

transformatie achteraf

$kP_gm2j = 10^{[(kP_gm2j_transformed - 1.000000) * (0.776714 / 0.200000) + 0.102223]}$

grens = helder, sedtype = zand

transformatie vooraf

$SDEPTHw0_t = (SDEPTHw0 - 1.786928) * (0.200000 / 0.801227) + 1.000000$
 $FMARSH_t = (FMARSH - 0.112454) * (0.200000 / 0.331317) + 1.000000$
 $CFETCH_t = (CFETCH - 1464.130344) * (0.200000 / 2135.478675) + 1.000000$
 $CQIN_t = (CQIN - 42.101943) * (0.200000 / 44.251998) + 1.000000$
 $CEXTWAT_t = (CEXTWAT - 0.657166) * (0.200000 / 0.380499) + 1.000000$

volledig netwerk

$kP_gm2j_transformed =$
 $+ 3.2348$
 $- 0.0035789 * [SDEPTHw0_t]^3.7972 * [FMARSH_t]^{-0.084542} * [CFETCH_t]^{0.66818} * [CQIN_t]^{-13.496} * [CEXTWAT_t]^3.5394$
 $- 0.022063 * [SDEPTHw0_t]^{-0.14405} * [FMARSH_t]^{-9.5834} * [CFETCH_t]^{0.61151} * [CQIN_t]^{-6.6821} * [CEXTWAT_t]^{-0.19825}$
 $- 2.2647 * [SDEPTHw0_t]^{0.0672} * [FMARSH_t]^{-0.016901} * [CFETCH_t]^{0.031564} * [CQIN_t]^{-0.13599} * [CEXTWAT_t]^{0.053158}$
 $+ 0.0030835 * [SDEPTHw0_t]^{0.58834} * [FMARSH_t]^{-0.085605} * [CFETCH_t]^{-24.253} * [CQIN_t]^{-1.4976} * [CEXTWAT_t]^{0.33702}$
 $- 0.0014915 * [SDEPTHw0_t]^{8.7327} * [FMARSH_t]^{-0.083755} * [CFETCH_t]^{0.35584} * [CQIN_t]^{-0.57509} * [CEXTWAT_t]^{9.5285}$
 $+ 0.08962 * [SDEPTHw0_t]^{-3.3642} * [FMARSH_t]^{0.1064} * [CFETCH_t]^{0.05113} * [CQIN_t]^{0.32548} * [CEXTWAT_t]^{0.15831}$
 $+ 0.0022701 * [SDEPTHw0_t]^{2.442} * [FMARSH_t]^{0.0039825} * [CFETCH_t]^{0.22365} * [CQIN_t]^{-6.9813} * [CEXTWAT_t]^{6.2808}$

vereenvoudigd netwerk

$kP_gm2j_transformed =$
 $+ 3.2392$
 $- 0.0034906 * [SDEPTHw0_t]^3.7972 * [CQIN_t]^{-13.496} * [CEXTWAT_t]^3.5394$
 $- 0.022535 * [FMARSH_t]^{-9.5834} * [CFETCH_t]^{0.61151} * [CQIN_t]^{-6.6821}$
 $- 2.2647 * [SDEPTHw0_t]^{0.0672} * [FMARSH_t]^{-0.016901} * [CFETCH_t]^{0.031564} * [CQIN_t]^{-0.13599} * [CEXTWAT_t]^{0.053158}$
 $+ 0.0030263 * [CFETCH_t]^{-24.253} * [CQIN_t]^{-1.4976}$
 $- 0.0014138 * [SDEPTHw0_t]^{8.7327} * [CEXTWAT_t]^{9.5285}$
 $+ 0.08594 * [SDEPTHw0_t]^{-3.3642} * [CQIN_t]^{0.32548}$
 $+ 0.0021897 * [CQIN_t]^{-6.9813} * [CEXTWAT_t]^{6.2808}$

transformatie achteraf

$kP_gm2j = 10^{[(kP_gm2j_transformed - 1.000000) * (0.627249 / 0.200000) + 0.434686]}$

grens = troebel, sedtype = veen

transformatie vooraf

$SDEPTHw0_t = (SDEPTHw0 - 1.786928) * (0.200000 / 0.801227) + 1.000000$
 $FMARSH_t = (FMARSH - 0.112454) * (0.200000 / 0.331317) + 1.000000$
 $CFETCH_t = (CFETCH - 1464.130344) * (0.200000 / 2135.478675) + 1.000000$
 $CQIN_t = (CQIN - 42.101943) * (0.200000 / 44.251998) + 1.000000$
 $CEXTWAT_t = (CEXTWAT - 0.657166) * (0.200000 / 0.380499) + 1.000000$

volledig netwerk

$kP_gm2j_transformed =$
- 2.3755
+ 3.5052*[SDEPTHw0_t]^0.23836*[FMARSH_t]^0.023097*[CFETCH_t]^0.0094448*[CQIN_t]^0.065622*[CEXTWAT_t]^0.020056
- 6.9116e-005*[SDEPTHw0_t]^12.592*[FMARSH_t]^0.051989*[CFETCH_t]^0.21119*[CQIN_t]^0.30926*[CEXTWAT_t]^13.239
+ 0.01154*[SDEPTHw0_t]^0.20006*[FMARSH_t]^0.13491*[CFETCH_t]^0.16.936*[CQIN_t]^0.5379*[CEXTWAT_t]^0.25015
+ 0.7624*[SDEPTHw0_t]^1.0614*[FMARSH_t]^0.3.3114*[CFETCH_t]^0.40683*[CQIN_t]^4.8134*[CEXTWAT_t]^0.4616
- 1.8635*[SDEPTHw0_t]^1.8258*[FMARSH_t]^0.3.7581*[CFETCH_t]^0.36072*[CQIN_t]^2.3643*[CEXTWAT_t]^0.15309
- 0.17177*[SDEPTHw0_t]^0.35618*[FMARSH_t]^0.2.7844*[CFETCH_t]^0.46462*[CQIN_t]^8.7291*[CEXTWAT_t]^0.51102
+ 1.1318*[SDEPTHw0_t]^2.3415*[FMARSH_t]^0.3.8328*[CFETCH_t]^0.3733*[CQIN_t]^1.8526*[CEXTWAT_t]^0.050215

vereenvoudigd netwerk

$kP_gm2j_transformed =$
- 2.1742
+ 3.24*[SDEPTHw0_t]^0.25762*[CQIN_t]^0.09404
- 4.0886e-005*[SDEPTHw0_t]^13.315*[CEXTWAT_t]^14.1
+ 0.010393*[CFETCH_t]^17.964
+ 0.9151*[SDEPTHw0_t]^0.67607*[FMARSH_t]^0.3.263*[CFETCH_t]^0.29215*[CQIN_t]^4.7086*[CEXTWAT_t]^0.38753
- 2.3875*[SDEPTHw0_t]^1.565*[FMARSH_t]^0.3.7619*[CFETCH_t]^0.33298*[CQIN_t]^2.6908*[CEXTWAT_t]^0.06587
- 0.16342*[FMARSH_t]^2.816*[CFETCH_t]^0.34066*[CQIN_t]^8.8347*[CEXTWAT_t]^0.53945
+ 1.5568*[SDEPTHw0_t]^2.006*[FMARSH_t]^0.3.8172*[CFETCH_t]^0.35241*[CQIN_t]^2.272

transformatie achteraf

$kP_gm2j = 10^{[(kP_gm2j_transformed - 1.000000) * (0.776714 / 0.200000) + 0.102223]}$