Ontwikkeling van een indicator om te Sturen Op Nitraat

Sturen Op Nitraat wordt uitgevoerd in opdracht van de ministeries LNV en VROM

Ontwikkeling van een indicator om te Sturen Op Nitraat

Gegevens en regressie-analyse op basis van twee meetseizoenen (2000-2001 en 2001-2002)

S.L.G.E. Burgers[1]
H.F.M. ten Berge[2]
J.J. de Gruijter[3]
M.J.D. Hack-ten Broeke[3]
I.E. Hoving[4]
S. Radersma[5]
A. Smit[3]
G.L. Velthof[3]

Alterra-rapport 866
Reeks Sturen Op Nitraat 9

Alterra, Wageningen, 2004

REFERAAT

S.L.G.E. Burgers, H.F.M. ten Berge, J.J. de Gruijter, M.J.D. Hack-ten Broeke, I.E. Hoving, S. Radersma, A. Smit en G.L. Velthof, 2004. *Ontwikkeling van een indicator om te Sturen Op Nitraat; gegevens en regressie-analyse op basis van twee meetseizoenen (2000-2001 en 2001-2002)*. Wageningen, Alterra, Alterra-rapport 866. Sturen Op Nitraat 9; 73 blz. 4 fig.; 17 tab.; 10 ref.

In het project STuren OP NITraat wordt gezocht naar verbanden tussen zogenaamde kandidaat-indicatoren voor de nitraatbelasting van het grondwater. Na analyse van de gegevens van twee meetseizoenen (2000/2001 en 2001/2002) komen in ieder geval de indelingen in drie Gt-groepen en vier bodemgroepen naar voren als verklarende factoren voor het niveauverschil in de gemeten nitraatconcentraties. Het nitraatgedeelte van de hoeveelheid Nmin in het najaar, gesommeerd over de drie bodemlagen (i.e. 0-90 cm), komt naar voren als het meest geschikt als indicator voor nitraatuitspoeling. Opschaling naar bedrijfsniveau resulteert in een voorspelling op basis van rekenregels van de bedrijfsgemiddelde nitraatconcentratie in het grondwater met een veel lagere voorspelfout dan wanneer op puntniveau voorspellingen worden gedaan. Dit biedt zicht op een praktisch toepasbare indicator op bedrijfsniveau.

Trefwoorden: Nmineraal, nitraatconcentratie, N-overschot, MINAS, Gt

ISSN 1566-7197

Dit rapport kunt u bestellen door € 18,- over te maken op banknummer 36 70 54 612 ten name van Alterra, Wageningen, onder vermelding van Alterra-rapport 866. Dit bedrag is inclusief BTW en verzendkosten.

© 2004 Alterra
Postbus 47; 6700 AA Wageningen; Nederland
Tel.: (0317) 474700; fax: (0317) 419000; e-mail: info@alterra.nl

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

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

[Alterra-rapport 866/januari/2004]

Inhoud

Woord vooraf 7

Samenvatting 9

1 Inleiding 13
1.1 Werkwijze 14
2 Gebruikte gegevens 15
2.1 Korte samenvatting gegevensverzameling 15
2.1.1 Begrippenlijst 16
2.2 Beschikbare gegevens van de kandidaat-indicatoren 17
2.2.1 Nitraatconcentratie grondwater 17
2.2.2 N-mineraal 20
2.2.3 Overschotten 22
2.2.4 Opmerkingen 23
3 Methoden en technieken 25
3.1 Aspecten met betrekking tot de data 25
3.2 Regressie : aannames en gebruikte technieken 25
3.3 Neuraal netwerk 27
4 Resultaten van de analyse op basis van twee meetseizoenen 29
4.1 Akkerbouw 29
4.1.1 De beste regressie modellen 29
4.1.2 Effect van andere variabelen 33
4.2 Veehouderij 35
4.2.1 Gras, de beste regressie modellen 35
4.2.2 Gras, effect van andere variabelen 37
4.2.3 Maïs, de beste regressie modellen 39
4.2.4 Maïs, effect van andere variabelen 42
4.3 Resultaten van Neuraal netwerk 43
5 Opschaling van proefplek naar bedrijf 45
5.1 Inleiding 45
5.2 Voorspelling van de bedrijfsgemiddelde nitraatconcentratie met nauwkeurigheid 46
5.3 Berekening van de voorspelling en de nauwkeurigheid van een cluster- en een bedrijfsgemiddelde nitraatconcentratie 49
5.3.1 Akkerbouw 49
5.3.2 Veehouderij 50
6 Conclusies 53
6.1 Regressie-analyse met proefplekgegevens 53
6.2 Opschaling naar bedrijfsniveau 54

Literatuurlijst 57

Bijlagen

1	Beschikbare gegevens van de overige gemeten variabelen	59
2	Artificiële neutrale netten	61
3a	Modellen akkerbouw	67
3b	Modellen gras en maïs	69
4a	Model Akkerbouw met en zonder DOC	71
4b	Model Gras met DOC	73

Woord vooraf

De serie ‘Sturen op Nitraat’ bundelt de onderzoeksresultaten behaald in het kader van het gelijknamig project. Het project wordt uitgevoerd in opdracht van het Ministerie van Landbouw, Natuurbeheer en Visserij (sinds kort Landbouw, Natuurbeheer en Voedselkwaliteit) en het Ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieubeheer. Doel is een handzame indicator voor de nitraatbelasting van grondwater te ontwikkelen, ten behoeve van zowel monitoring doeleinden als voor sturing in de landbouwpraktijk.

Het project wordt uitgevoerd door een aantal onderzoekspartners binnen Wageningen UR: Alterra, Praktijkonderzoek Plant en Omgeving (PPO); Praktijkonderzoek Veehouderij (PV; onderdeel van Animal Sciences Group) en Plant Research International B.V. (PRI) en buiten Wageningen UR: Centrum voor Landbouw en Milieu (CLM)

Het project Sturen Op Nitraat is opgedeeld in deelprojecten. De projectleider van het totale project is Dethmer Boels (Alterra). Dit rapport is een produkt van het deelproject ‘integrale analyse van de bedrijfsgegevens’ (Noij et al., 2001). Aan dit deelproject werkten de volgende personen (tevens auteurs van dit rapport) mee:

Saskia Burgers (Biometris, statistiek)
Hein ten Berge (PRI, agrosysteemkunde)
Jaap de Gruijter (Alterra, ruimtelijke statistiek)
Mirjam Hack-ten Broeke (Alterra, deelprojectleider)
Idse Hoving (PV, melkveehouderij)
Simone Radersma (PPO, akkerbouw)
Annemieke Smit (Alterra, organische stof en nutriënten)
Gerard Velthof (Alterra, N-processen)

Dit rapport bevat de resultaten van de analyse op basis van twee meetseizoenen (2000-2002). Het nog uit te brengen eindrapport van dit deelproject zal een analyse bevatten van de gegevens van drie meetseizoenen. Tot dan toe zijn alle resultaten te beschouwen als voorlopig.

Naast de bovengenoemde auteurs is er nog een grote groep mensen die, achter de schermen, een bijdrage hebben geleverd aan dit rapport. Bij deze willen wij Anna Zwijnenburg, Ivonne Kok en Jouke Oenema bedanken voor het aanleveren van de data, Jessica Stoker voor haar bijdrage aan het databeheer en Wies Akkermans voor het fitten van de neurale netten.

Samenvatting

Het doel van het Europese nitraatbeleid is schoon grond- en oppervlaktewater. Omdat MINAS zo zijn beperkingen kent is een indicator gewenst, die dichter staat bij het milieudoel. Voor het stikstofbeleid op (droge) zand- en lössgronden is behoefte aan een mogelijkheid om gerichter te sturen op nitraat en om daarmee het milieurendement van maatregelen te verhogen. Er is behoefte aan een indicator voor nitraatuitspoeling die praktisch hanteerbaar, goed controleerbaar en handhaafbaar is en daarmee geschikt kan zijn als grondslag voor aanvullend N-beleid. Een indicator die geschikt is voor het bedrijfsniveau legt vooraf het verband tussen (gewenste) milieukwaliteit en (gewenste) bedrijfsvoering.

In het project 'STuren OP NITraat' (kortweg ook wel STOPNIT genoemd) wordt zodoende gezocht naar verbanden tussen zogenaamde kandidaat-indicatoren voor de nitraatbelasting van het grondwater. Middels regressie-analyse wordt onderzocht of de gegevens zoals het N-bedrijfsoverschot, het N-perceelsoverschot, N-mineraalgehalten in de bodem, weersgegevens en locatiespecifieke factoren zoals grondsoort en grondwatertrap (Gt) kunnen worden gebruikt voor een voorspelling van nitraatconcentraties. Deze nitraatconcentraties worden daarmee beschouwd als de meest directe maat om de stikstofbelasting te kwantificeren. Daartoe wordt op 34 bedrijven, verspreid over zand- en lössgronden van Nederland, informatie verzameld. De lokaties zijn zo goed mogelijk verdeeld over de verschillende voorkomende grondsoorten, Gt's en gewassoorten. De bijbehorende gegevensverzameling is uitgebreid beschreven door Smit et al. (2003).

Na analyse van de gegevens van twee meetseizoenen (2000/2001 en 2001/2002) voor het project Sturen Op Nitraat kunnen alleen voorlopige conclusies worden getrokken. Pas als alle gegevens van drie meetseizoenen verwerkt zijn, zullen de conclusies definitief zijn.

De regressie-analyses zijn uitgevoerd voor akkerbouwgewassen, gras en maïs afzonderlijk omdat er duidelijke verschillen waren tussen akkerbouw en melkveehouderij. In alle gevallen komt de indeling in drie Gt-groepen (respectievelijk met GHG ondieper dan 40 cm – mv., met GHG tussen 40 en 80 cm – mv. en met GHG dieper dan 80 cm – mv.) naar voren als een verklarende factor voor het niveauverschil in de gemeten nitraatconcentraties in het voorjaar van 2001 en 2002. Ook de indeling in bodemgroep (respectievelijk de lössgronden, zandgronden met veel organische stof of een dikke bovengrond, zandgronden met relatief veel organisch stof en een hoog leemgehalte en de overige zandgronden) komt naar voren als een verklarende variabele, dit in tegenstelling tot de resultaten van de analyse na één meetseizoen. Tenslotte speelde het al dan niet aanwezig zijn van een veenlaagje een rol bij zowel de akkerbouwgewassen als gras en maïs. Het aanwezig zijn van een veenlaagje verlaagt de verwachte nitraatconcentratie.

Uit de verzamelde Nmin-gegevens komt de hoeveelheid $Nmin_{nitraat}$ (het nitraatdeel van Nmin) in de meetperiode oktober-december, gesommeerd over de drie bodemlagen (i.e. 0-90 cm) als meest verklarende variabele naar voren uit de groep aan Nmin-gegevens. $Nmin_{nitraat}$ speelt in alle gevallen een belangrijke rol in de verklaring van de gemeten nitraatconcentratie. Het is gebleken dat de relatie tussen het nitraatdeel van Nmin met de nitraatconcentratie beter is dan die met de som van ammonium en nitraat. Ook uit de berekende bedrijfsgemiddelde van zowel de nitraatconcentratie als Nmin en $Nmin_{nitraat}$ blijkt dat het verband duidelijk beter is met $Nmin_{nitraat}$.

Bij akkerbouw is in de analyse een nieuwe gewasgroep-indeling toegevoegd. Dit is een indeling op basis van de verwachte Nmin in het najaar waarbij uiteindelijk de groepen 'hoog' en 'midden en laag' zijn onderscheiden. Deze nieuwe gewasgroepen-indeling komt naar voren als een verklarende factor voor het niveauverschil in de gemeten nitraatconcentratie, dit in tegenstelling tot de oorspronkelijke gewasgroepindeling. Bij akkerbouw speelt ook de neerslagsom in de zomerperiode (1 april tot 1 oktober) in alle drie de modellen een rol. Daarnaast zijn er een aantal andere variabelen met een significante parameter die een rol kunnen spelen in de voorspelling van de nitraatconcentratie. Dit zijn de C:N-verhouding, het neerslagoverschot in de zomerperiode en er is een kwadratisch verband met de neerslagsom in de winterperiode, gemeten tussen de datum van de Nmin- en de nitraat-meting. Het opnemen van deze variabelen in het regressie-model geeft echter maar een zeer beperkte verbetering in de nauwkeurigheid van de voorspelling en van de fractie verklaarde variantie.

Voor grasland zijn naast Gt-groep, bodemgroep, $Nmin_{nitraat}$ en veen, de volgende verklarende variabelen naar voren gekomen: C:N-verhouding, GHG en de potentiële mineralisatie die gezien moet worden als een maat voor afbreekbaarheid van organische stikstof. Voor maïs levert de regressie-analyse naast Gt-groep, bodemgroep, $Nmin_{nitraat}$ en veen de volgende extra verklarende variabelen op: GHG en de potentiële mineralisatie (als maat voor afbreekbaarheid van organisch stikstof) of GLG en de neerslagsom in de zomer . Ook voor gras en maïs geldt dat de modellen slechts in geringe mate verbeterd worden door toevoeging van deze verklarende factoren

Opvallend is dat, evenmin als in de analyse over het eerste meetseizoen, géén van de 'bemestingsvariabelen' (totale N-gift; werkzame N-gift, kunstmestgift, dierlijke-mestgift, N-overschot bedrijf; N-overschot perceel, MINAS N-overschot) als significant verklarende factor naar voren komt.

Als de regressie-vergelijkingen worden gebruikt om nitraatconcentraties te voorspellen op puntniveau, zijn de betrouwbaarheidsintervallen nogal groot. Grofweg zijn deze intervallen ± 100 tot 135 mg/l. Dit betekent dat voorspellingen of puntniveau nauwelijks zinvol zijn. De bedoeling van het project is echter om voorspellingen te doen voor cluster- of bedrijfsniveau. De opschaling naar bedrijfsniveau vindt plaats via de zogenaamde clusterbenadering. Dat wil zeggen dat de ontwikkelde regressie-vergelijkingen worden toegepast voor bodem-Gt-

gewascombinaties (clusters) en dat de voorspelling van de bedrijfsgemiddelde nitraatconcentratie niets anders is dan het oppervlaktegewogen gemiddelde van de aldus geschatte clustergemiddelden van een bedrijf. Van belang is daarbij dat vooral de voorspelfout van zo'n bedrijfsgemiddelde vele malen lager is dan de voorspelfout op puntniveau. In een aantal rekenvoorbeelden komt naar voren dat de 95%-betrouwbaar-heidsintervallen voor de voorspelde nitraatconcentratie dan nog slechts + of – 25 tot 37 mg/l zijn. Als de regressie-vergelijkingen gebaseerd zijn op meerdere groeiseizoenen zal deze voorspelfout naar verwachting verder afnemen.

Samengevat betekent dit dat op basis van twee meetseizoenen de voorlopige conclusie kan worden getrokken dat de grondwatertrap, het bodemtype, het soort gewas, het al dan niet aanwezig zijn van een veenlaagje en het nitraatdeel van N-mineraal in het najaar de belangrijkste verklarende variabelen zijn voor de gemeten nitraatconcentratie in het voorjaar. Daarmee lijkt N-mineraal ook na twee meetseizoenen nog steeds de meest perspectiefvolle indicator om te sturen op nitraat. Echter, de nauwkeurigheid van een voorspelling op bedrijfsniveau laat nog veel te wensen over. Pas als alle gegevens (van drie meetseizoenen) zijn geanalyseerd zal blijken of de conclusie overeind blijft en de nauwkeurigheid acceptabel wordt.

1 Inleiding

Oorspronkelijk, namelijk voorafgaand aan de uitspraak van het Europese Hof op 2 oktober 2003 over het Nederlandse mestbeleid, luidde de in 2000 geformuleerde doelstelling voor Sturen Op Nitraat als volgt:

Het realiseren van een lagere nitraatuitspoeling, zoals ook de Europese Nitraatrichtlijn aan de lidstaten voorschrijft, heeft voor de Nederlandse landbouw zware consequenties. De doelstelling voor een concentratie in het grondwater van (maximaal) 50 mg/l nitraat is vertaald in een stevig traject van N-verliesnormen tot 2003. Binnen de droge zandgronden en lössgronden liggen zo'n 50.000 ha prioritaire waterwingebieden, waar het belang van de milieukwaliteit extra groot is en aanleiding kan zijn om de doelstelling van 50 mg/l verder aan te scherpen. (Noij et al., 2001)

De samenhang tussen nitraatbelasting en MINAS-normen is weliswaar onmiskenbaar maar is ook omgeven met onduidelijkheden. Zo worden in MINAS enkele relevante N-aanvoerposten niet meegenomen (mineralisatie, vlinderbloemigen, e.a.). Bovendien is de vertaling van de toegestane nitraatbelasting naar een bepaald N-verlies niet één-op-één te maken. Tenslotte valt met MINAS het milieu-rendement van afzonderlijke maatregelen niet gemakkelijk vast te stellen. Doel van het nitraatbeleid is schoon grond- en oppervlaktewater. Daarom is een indicator gewenst, die dichter staat bij het milieudoel dan de MINAS-normen én die dichter staat bij de bedrijfsvoering. Feitelijke meting van nitraatconcentraties is weliswaar de meest precieze indicator voor de aanwezigheid van nitraat, maar is praktisch moeilijk uitvoerbaar en kostbaar. Bovendien is het geen directe indicator voor nitraatuitspoeling als vracht naar het grondwater.

Voor het stikstofbeleid op (droge) zand- en lössgronden is behoefte aan een mogelijkheid om gerichter te sturen op nitraat en om daarmee het milieurendement van maatregelen te verhogen. Er is behoefte aan een indicator voor nitraatuitspoeling die praktisch hanteerbaar, goed controleerbaar en handhaafbaar is en daarmee geschikt kan zijn als grondslag voor aanvullend N-beleid. Boeren willen gericht kunnen sturen op vermindering van de nitraatuitspoeling. Dit geldt in het bijzonder voor voorlopers en deelnemers aan experimenten in waterintrekgebieden. Hiervoor is een geschikte indicator nodig voor nitraatuitspoeling. Een indicator die geschikt is voor het bedrijfsniveau legt vooraf het verband tussen (gewenste) milieukwaliteit en (gewenste) bedrijfsvoering.

Het doel van dit onderzoek is meervoudig:

1. *De ontwikkeling van een indicator[1] voor nitraatuitspoeling die geschikt is:*
   - *als grondslag voor aanvullend stikstofbeleid,*
   - *voor management op bedrijfsniveau,*
   - *als instrument voor gebiedsgericht beheer en*
   - *voor de monitoring van gebiedsgericht beleid.*

[1] In het projectplan wordt consequent het enkelvoud van het woord indicator gebruikt maar uit het onderzoek kan ook een set indicatoren komen of een 'slimme indicator', die rekening houdt met maatregelen of processen die niet gemeten worden (zie ook 2.1).

2. *De toetsing van de indicator op onafhankelijke praktijkbedrijven en in een regionaal nitraatexperiment aan de criteria doelgerichtheid, meetbaarheid en beïnvloedbaarheid.*

Deze wens voor een indicator is niet veranderd door de uitspraak van het Europese Hof, die grosso modo betekent dat het MINAS systeem vervangen zal worden door een systeem van gebruiksnormen, zoals momenteel wordt voorbereid ten behoeve van het Derde Actieprogramma Nitraatrichtlijn, en dat in concept per 5 december 2003 werd voorgelegd aan de Tweede Kamer. Het heeft wel consequenties voor het vervolg van het Sturen Op Nitraat-project, omdat N-gebruik expliciet in de analyse als mogelijke indicator zal worden betrokken. Ook in de nieuwe context zal er blijvende behoefte zijn aan indicatoren, die voor boeren op uitspoelinggevoelige zandgronden kunnen dienen als hulpmiddel om de effectiviteit van maatregelen te beoordelen, dan wel om te gebruiken ter handhaving van de regelgeving.

In dit rapport, wordt verslag gedaan van het 'zoeken naar een indicator' op basis van twee meetseizoenen conform het oorspronkelijke projectplan. Bij de analyse van de gegevens van drie meetseizoenen, te verslaan in een vervolgrapport, zal N-gebruik wel worden betrokken.

1.1 Werkwijze

In het project 'STuren OP NITraat' (kortweg ook wel STOPNIT genoemd) wordt zodoende gezocht naar verbanden tussen zogenaamde kandidaat-indicatoren voor de nitraatbelasting van het grondwater. Middels regressie-analyse wordt onderzocht of de gegevens zoals het N-bedrijfsoverschot, het N-perceelsoverschot, N-mineraalgehalten in de bodem, weersgegevens en locatiespecifieke factoren zoals grondsoort en grondwatertrap (Gt) kunnen worden gebruikt voor een voorspelling van nitraatconcentraties. Deze nitraatconcentraties worden daarmee beschouwd als de meest directe maat om de stikstofbelasting te kwantificeren. Daartoe wordt op 34 bedrijven, verspreid over zand- en lössgronden van Nederland, informatie verzameld. De locaties zijn zo goed mogelijk verdeeld over de verschillende voorkomende grondsoorten, Gt's en gewassoorten. De bijbehorende gegevensverzameling is uitgebreid beschreven door Smit et al. (2003).

Dit rapport bevat een beschrijving van de verkregen data en de analyse van de gegevens voor de ontwikkeling van 'de' indicator. Dat betekent dat hier verder geen aandacht wordt besteed aan het regionale monitoringsconcept en evenmin aan de toetsbedrijven. In hoofdstuk 2 wordt kort ingegaan op de gegevensverzameling en worden de verzamelde gegevens gepresenteerd. In hoofdstuk 3 wordt uitgebreid ingegaan op de toegepaste analysemethoden. Hoofdstuk 4 bevat de voorlopige regressie-modellen voor de voorspelling van nitraatconcentraties in het grondwater op basis van twee meetseizoenen. Tenslotte komt in hoofdstuk 5 de opschaling naar bedrijfsniveau aan bod.

2 Gebruikte gegevens

2.1 Korte samenvatting gegevensverzameling

De data-analyse is gebaseerd op gegevens van een totaal van 34 bedrijven. Deze bedrijven liggen verspreid over de zand- en lössgronden van Nederland. Er zijn 15 akkerbouw- of vollegrondsgroentenbedrijven, 18 melkveehouderijbedrijven en één gemengd bedrijf. Deze paragraaf beschrijft de gegevensverzameling globaal. Meer achtergronden zijn te vinden in Smit et al. (2003).

Bij het begin van het onderzoek is gekozen voor een gestratificeerde aselecte steekproef met een voor het onderzoek zo gunstig mogelijke verdeling van steekproefplekken over de bedrijven, de voorkomende grondsoorten, grondwatertrappen (Gt's) en geteelde gewassen. De strata zijn gedefinieerd als combinaties van vier factoren, die belangrijk kunnen zijn voor nitraatuitspoeling: bedrijf, grondsoort, Gt-groep en gewasgroep. Voor de stratificatiefactoren grondsoort, Gt en gewas werden de volgende indelingen gekozen:

Indeling in 4 grondsoorten:
1) L : Lössgronden
2) Z1 : Zandgronden met veel organische stof of een dikke bovengrond (zoals enkeerdgronden, moerige gronden)
3) Z2 : Zandgronden met relatief veel organische stof en een hoog leemgehalte (zoals de meeste beekeerdgronden, sommige gooreerdgronden, zandgronden met een kleidek, keileemgronden)
4) Z3 : Overige zandgronden (sommige beekeerdgronden, meeste gooreerdgronden, podzolgronden)

Indeling in 3 Gt-groepen:
1) 1 : GHG (Gemiddelde Hoogste Grondwaterstand) ondieper dan 40 cm (Gt I, II, II*, IIb, III, III*, V, V*)
2) 2 : GHG tussen 40 en 80 cm (Gt IIc, IV, VI)
3) 3 : GHG dieper dan 80 cm (Gt IVc, VII, VII* en VIII)

Indeling in 6 gewasgroepen:
1) g : grasland
2) m : snijmaïs op melkveehouderijbedrijven,
3) t : andijvie, boerenkool, bloemkool, chinese kool, knolselderij, korrelmaïs, spitskool, ijsbergsla, CCM en MKS
4) a : aardappel, koolraap, koolrabi, kropsla, prei, radijs, snijmaïs, spinazie en ui
5) b : broccoli, knolvenkel, luzerne, peulvruchten, rode kool, spruitkool, suikerbiet, voederbiet en witte kool
6) r : aardbei, asperge, bospeen, gerst, haver, rode biet, rogge, schorseneer, tarwe, witlof en wortel

Dit leidt tot maximaal 60 combinaties van bodem-Gt-gewas, maar niet alle combinaties komen daadwerkelijk voor. In totaal zijn er 478 proefplekken geloot, verspreid over 47 bodem-Gt-gewascombinaties.

2.1.1 Begrippenlijst

In dit rapport komen een aantal begrippen steeds terug. Hieronder volgt een overzicht van de gebruikte begrippen met een omschrijving van wat ermee bedoeld wordt. Een uitgebreide beschrijving van deze begrippen is te vinden in het rapport over de gegevensverzameling (Smit et al., 2003)

- Grondsoort (bodemgroep), Gt-groep en gewasgroep
  De gebruikte indeling (met code) staat gegeven in paragraaf 2.1.1.
- Nitraat
  De nitraatconcentraties zijn gemeten in grondwater of bodemvocht in het 'voorjaar van 2001' (19 maart t/m 5 juni, na de MKZ-crisis) en in het 'voorjaar van 2002' (25 maart t/m 17 mei).
- Nmin
  Dit betreft Nmineraal-waarnemingen (nitraat en ammonium) gesommeerd over de bodemlagen 0-30, 30-60 en 60-90 cm – mv. voor de meetperiode oktober-december 2000 of oktober-november 2001.
  $Nmin_{nitraat}$: nitraatdeel van Nmineraal
  $Nmin_{ammonium}$: ammoniumdeel van Nmineraal
- Perceelsoverschot
  Behalve perceelsoverschot zijn belangrijke balansposten die mede bepalend zijn voor het perceelsoverschot bij de analyse betrokken. Het betreft: kunstmestgift en dierlijke mestgift (zowel als werkzame N en als de totale N); alle plekken op één perceel hebben dezelfde gift en overschot.
- Bedrijfsoverschot en MINAS-overschot; alle proefplekken van één bedrijf hebben hetzelfde overschot.
- Aantal dierweidedagen op grasland
- Groenbemester gezaaid in 2000 en 2001 (ja/nee) op akkerbouwland of maïs
- N-afvoer met het gewas (gemeten of geschat)
- Weersgegevens :
  Neerslagsom1 is de neerslagsom voor het groeiseizoen (respectievelijk 1 april 2000-1 oktober 2000 en 1 april 2001-1 oktober 2001) en Neerslagoverschot1 is het overschot voor dezelfde perioden.
  Neerslagsom2 is de neerslagsom voor het uitspoelingsseizoen (respectievelijk 1 oktober 2000-1april 2001 en 1 oktober2001-1 april 2002) en Neerslagoverschot2 is het overschot voor dezelfde periode. Neerslag2precies is de precieze neerslagsom (in de winterperiode) tussen de meetdatum van Nmineraal en de nitraatmeting.
  De neerslaggegevens zijn verzameld op de bedrijven zelf. Indien er gegevens missen zijn deze aangevuld met behulp van de dichtstbijzijnde weerstations. De referentie-gewasverdamping en gemiddelde correctiefactoren per

gewasgroep zijn altijd berekend op basis van de gegevens van de weerstations.

- Denitrificatie
  Denitrificatiecapaciteit (potentiële denitrificatie) voor 6 bodemlagen, namelijk 2-7 cm, 13-18 cm, 23-28 cm, 33-38 cm, 50-55 cm en 70-75 cm.
- Mineralisatie
  Mineralisatiecapaciteit en alle extra gegevens die als indicatief voor potentiële mineralisatie kunnen worden beschouwd : Ntotaal, Ctotaal (g per kg), C:N-verhouding en hot-KCl extraheerbaar ammonium
- Organische stof en organisch N
- $DOC_{grondwater}$: opgelost organisch koolstof in het grondwater (mg/l)
- Veen : het al dan niet aanwezig zijn van een veenlaagje in de bodem
- GHG en GLG : de gemiddeld hoogste respectievelijk laagste grondwaterstand bepaald tijdens veldbezoek/profielbeschrijving.

2.2 Beschikbare gegevens van de kandidaat-indicatoren

Op basis van de dataset zijn overzichten gemaakt die inzicht geven in de beschikbare gegevens van het eerste (2000-2001) en het tweede meetseizoen (2001-2002). In de gegevens van het eerste meetseizoen zijn kleine wijzingen opgetreden ten opzichte de eerder gerapporteerde gegevens over het eerste meetseizoen (Hack-ten Broeke et. al.; 2003), vanwege verbeteringen in de dataset.

2.2.1 Nitraatconcentratie grondwater

In de Tabellen 1, 2 en 3 worden de gemiddelden van de gemeten nitraatconcentraties gegeven evenals de minimum, maximum en mediaan waarde per waargenomen bodemgroep, Gtgroep en gewasgroep.

Tabel 1a. Overzicht van de Nitraatconcentraties (mg/l) per bodemgroep, meetseizoen 2000-2001

Bodem-Groep	*Aantal Proefplekken*	*Gemiddelde*	*Minimum*	*Maximum*	*Mediaan*
L	30	60.4	0.3	269.6	40.7
Z1	78	81.5	0.0	468.5	54.9
Z2	147	90.7	0.0	519.0	57.1
Z3	217	86.0	0.0	377.7	68.1

Tabel 1b. Overzicht van de Nitraatconcentraties (mg/l) per bodemgroep, meetseizoen 2001-2002

Bodem-Groep	*Aantal Proefplekken*	*Gemiddelde*	*Minimum*	*Maximum*	*Mediaan*
L	29	53.6	0.4	153.8	52.2
Z1	74	46.0	0.0	227.4	28.5
Z2	147	68.8	0.0	296.6	51.4
Z3	215	66.0	0.0	410.2	56.9

Tabel 2a. Overzicht van de Nitraatconcentraties (mg/l) per Gt-groep, meetseizoen 2000-2001.

Gt-groep	*Aantal proefplekken*	*Gemiddelde*	*Minimum*	*Maximum*	*Mediaan*
1	70	51.2	0.0	382.9	27.0
2	185	79.1	0.0	365.0	61.3
3	217	101.1	0.0	519.0	73.5

Tabel 2b. Overzicht van de Nitraatconcentraties (mg/l) per Gt-groep, meetseizoen 2001-2002

Gt-groep	*Aantal proefplekken*	*Gemiddelde*	*Minimum*	*Maximum*	*Mediaan*
1	70	42.6	0.0	296.6	22.8
2	183	60.3	0.0	308.2	44.7
3	212	72.0	0.3	410.2	59.5

Tabel 3a. Overzicht van de Nitraatconcentraties (mg/l) per gewasgroep, meetseizoen 2000-2001

Gewas-Groep	*Aantal Proefplekken*	*Gemiddelde*	*Minimum*	*Maximum*	*Mediaan*
a	90	116.5	0.0	405.7	106.2
b	60	84.2	0.0	237.9	75.4
g	174	65.6	0.0	452.0	34.2
m	73	74.6	0.2	468.5	64.6
r	63	80.0	0.0	519.0	63.6
t	12	226.0	0.4	377.7	279.7

Tabel 3b. Overzicht van de Nitraatconcentraties (mg/l) per gewasgroep, meetseizoen 2001-2002

Gewas-Groep	*Aantal Proefplekken*	*Gemiddelde*	*Minimum*	*Maximum*	*Mediaan*
a	76	83.3	0.0	257.6	72.9
b	42	61.6	0.0	258.7	59.9
g	182	50.0	0.0	308.2	39.1
m	76	72.6	0.0	410.2	53.7
r	59	52.0	0.0	263.3	38.8
t	10	136.7	0.0	263.5	139.6

De gemiddelde nitraatconcentraties liggen in het tweede meetseizoen lager dan in het eerste meetseizoen. Dat geldt voor alle bodemgroepen, Gt-groepen en voor alle gewasgroepen. Het verschil tussen het gemiddelde en de mediaan is in het tweede meetseizoen kleiner dan in het eerste meetseizoen. Dit betekent dat er in het tweede meetseizoen minder extreem hoge waarden gemeten zijn. Voor de bodemgroepen löss en Z1 geldt in the tweede meetjaar dat het gemiddelde rond de nitraatnorm van 50 mg/l ligt. Op basis van de mediaan kan gesteld worden dat ook voor de Z2 en Z3 zo'n 50 % van de waarnemingen in het tweede meetjaar onder de norm valt.

Net als in het eerste meetseizoen heeft Gtgroep 1 conform de verwachting gemiddeld de laagste nitraatconcentratie. Opvallend is dat van de gewasgroepen gras steeds de laagste nitraatconcentratie te zien geeft.

Naast de gewasgroep-indeling van Tabel 3 is een tweede gewasgroep-indeling gemaakt voor de akkerbouwgewassen. Deze indeling bestaat uit vier niveaus te weten 'extra hoog', 'hoog', 'midden' of 'laag' en deze omschrijving heeft betrekking op de

verwachte Nmin in het najaar (Enckevort et.al.; 2002). De oorspronkelijke indeling voorzag in een indeling in vier groepen (Smit et al., 2003) op basis van verwacht overschot én de verwachte mineralisatie als gevolg van bijvoorbeeld gewasresten. Vooral de combinaties van enerzijds een laag overschot met een hoge mineralisatie en anderzijds een hoog overschot met een lage mineralisatie zouden wel eens tot dezelfde uitspoeling kunnen leiden. Omdat het doel is om de nitraatconcentratie te voorspellen is een indeling op basis van verwachte Nmin in de bodem logischer. Daarom is de nieuwe indeling geïntroduceerd.

De nieuwe indeling ziet er als volgt uit:

1) laag: verwachte Nmin < 60 kg/ha N; dit geldt voor de gewassen cichorei, fijne peen, gras, groene braak, suikerbiet op zandgrond, waspeen, winterpeen, witlof, witte kool
2) midden: verwachte Nmin tussen 60-120 kg/ha N; dit geldt voor de gewassen bospeen (herfst), graszaad, haver, knolvenkel, luzerne, schorseneer, stam/stok slaboon, stam/stok snijboon, triticale, winterrogge, wintertarwe op zandgronden, zetmeelaardappelen op zandgronden, zomergerst, zomertarwe
3) hoog: verwachte Nmin tussen 120-180 kg/ha N; dit geldt voor de gewassen andijvie, Chinese kool, consumptieaardappelen, knolvenkel, korrelmaïs, kropsla, prei, radicchio rosso, snijmaïs
4) extra hoog: verwachte Nmin > 180 kg/ha N; dit geldt voor spinazie

Consumptie-aardappelen vallen op de grens tussen midden en hoog (namelijk verwachte Nmin = 120 kg/ha N) en zijn uiteindelijk toch bij de andere aardappelen in de categorie ‘midden’ ingedeeld. De categoriën ‘extra hoog’ en ‘hoog’ zijn samengevoegd. De oude gewasgroep t is daarmee geheel in ‘hoog’ terecht gekomen. Daarnaast zijn twee proefplekken uit gewasgroep b met dubbelteelt knolvenkel en prei plus de dubbelteelten uit gewasgroep a in ‘hoog’ terecht gekomen. Alle overige gewassen vallen in de categorieën ‘midden’ of ‘laag’. Snijmaïs en gras blijven aparte gewasgroepen. Hiermee is de verdeling van de proefplekken over de gewasgroepen een stuk beter en valt gewasgroep t niet meer zo uit de toon.

Tabel 4a. Overzicht van de Nitraatconcentraties (mg/l) per gewasgroep, volgens de nieuwe indeling, 2000-2001

Gewas-Groep	*Aantal Proefplekken*	*Gemiddelde*	*Minimum*	*Maximum*	*Mediaan*
Laag	74	94.5	0.0	519.0	89.7
Midden	117	85.4	0.0	273.1	71.4
Hoog	31	196.5	0.4	405.7	219.3
Gras	174	65.6	0.0	452.0	34.2
Maïs	73	74.6	0.2	468.5	64.6

Tabel 4b. Overzicht van de Nitraatconcentraties (mg/l) per gewasgroep, volgens de nieuwe indeling, 2001-2002

Gewas-Groep	*Aantal Proefplekken*	*Gemiddelde*	*Minimum*	*Maximum*	*Mediaan*
Laag	41	59.4	0.0	157.0	61.0
Midden	104	62.4	0.0	263.3	55.9
Hoog	42	105.3	0.0	263.5	87.9
Gras	182	50.0	0.0	308.2	39.1
Maïs	76	72.6	0.0	410.2	53.7

Er zijn in beide seizoenen geen duidelijke verschillen tussen de gewasgroepen 'laag' en 'midden'. Gewasgroep 'laag' bestaat grotendeels uit proefplekken met suikerbieten, een gewas wat (qua verwachte Nmin) tegen het grensgebied aanzit.

2.2.2 N-mineraal

In Tabel 5 wordt voor de kandidaat-indicator N-mineraal een overzicht gegeven volgens de nieuwe gewasgroep-indeling. De metingen zijn gedaan in het najaar, in de periode oktober tot december. Er is gekozen voor een overzicht per gewasgroep omdat de regressie-analyses per gewasgroep zijn gedaan en er binnen de gewasgroepen verschillen tussen bodem- en Gtgroepen zijn. Uit een tussentijdse analyse is gebleken dat een opsplitsing van Nmin naar het nitraat- en ammoniumdeel een beter resultaat geeft voor wat betreft de voorspelling van de nitraatconcentratie. Daarom staat in Tabel 5 ook Nmin uitgesplitst naar $Nmin_{nitraat}$ (5c en 5d) en $Nmin_{ammonium}$ (5e en 5f).

Tabel 5a. Overzicht van de Nmin (kg N/ha) gesommeerd voor laag 0-90, per gewasgroep, volgens de nieuwe indeling, meetseizoen 2000-2001

Gewas-Groep	*Aantal Proefplekken*	*Gemiddelde*	*Minimum*	*Maximum*	*Mediaan*
Laag	36	48.3	4.3	231.4	35.9
Midden	69	44.0	4.2	341.0	33.8
Hoog	27	138.9	13.7	580.4	121.4
Gras	154	49.9	3.7	206.8	44.4
Maïs	47	62.5	5.2	131.0	59.2

Tabel 5b. Overzicht van de Nmin (kg N/ha) gesommeerd voor laag 0-90, per gewasgroep, volgens de nieuwe indeling, meetseizoen 2001-2002

Gewas-Groep	*Aantal Proefplekken*	*Gemiddelde*	*Minimum*	*Maximum*	*Mediaan*
Laag	42	28.0	0.00	164.7	22.6
Midden	107	52.0	4.2	179.1	48.8
Hoog	43	60.5	12.2	132.2	59.0
Gras	182	53.9	0.0	174.5	44.6
Maïs	77	64.8	5.2	156.9	59.1

Tabel 5c. Overzicht van het nitraatdeel van Nmin (kg N/ha) gesommeerd voor laag 0-90, per gewasgroep, volgens de nieuwe indeling, meetseizoen 2000-2001

Gewas-Groep	*Aantal Proefplekken*	*Gemiddelde*	*Minimum*	*Maximum*	*Mediaan*
Laag	36	36.9	4.0	215.3	19.2
Midden	69	25.5	3.5	114.8	18.6
Hoog	27	121.3	5.2	570.3	103.3
Gras	154	33.5	0.0	206.8	22.3
Maïs	47	49.2	5.2	131.0	41.3

Tabel 5d. Overzicht van het nitraatdeel van Nmin (kg N/ha) gesommeerd voor laag 0-90, per gewasgroep, volgens de nieuwe indeling, meetseizoen 2001-2002

Gewas-Groep	*Aantal Proefplekken*	*Gemiddelde*	*Minimum*	*Maximum*	*Mediaan*
Laag	42	18.0	0.0	87.4	13.3
Midden	107	33.8	0.0	92.0	31.8
Hoog	43	47.2	0.0	107.1	41.7
Gras	182	32.2	0.0	103.8	26.9
Maïs	77	47.4	0.0	108.5	43.0

Tabel 5e. Overzicht van het ammoniumdeel van Nmin (kg N/ha) gesommeerd voor laag 0-90, per gewasgroep, volgens de nieuwe indeling, meetseizoen 2000-2001

Gewas-Groep	*Aantal Proefplekken*	*Gemiddelde*	*Minimum*	*Maximum*	*Mediaan*
Laag	36	11.4	0.0	43.5	7.4
Midden	69	18.5	0.0	259.8	9.5
Hoog	27	17.6	0.0	211.2	4.2
Gras	154	16.5	0.0	103.1	12.5
Maïs	47	13.3	0.0	59.9	9.4

Tabel 5f. Overzicht van het ammoniumdeel van Nmin (kg N/ha) gesommeerd voor laag 0-90, per gewasgroep, volgens de nieuwe indeling, meetseizoen 2001-2002

Gewas-Groep	*Aantal Proefplekken*	*Gemiddelde*	*Minimum*	*Maximum*	*Mediaan*
Laag	42	10.0	0.0	77.2	4.3
Midden	107	18.2	0.0	159.2	10.7
Hoog	43	13.3	0.0	47.5	9.4
Gras	182	21.7	0.0	123.2	18.4
Maïs	77	17.4	0.0	140.1	14.3

Het aantal waarnemingen voor Nmin is in het tweede meetseizoen toegenomen omdat op een aantal bedrijven de Nmin in het najaar van 2001 voor het eerst gemeten is. Ten opzichte van het eerste meetseizoen is Gtgroep 1 nu veel beter vertegenwoordigd.

Voor de gemeten Nmin geldt dat de gemiddelde hoeveelheid in het tweede meetseizoen is gedaald voor de gewasgroepen ‘laag’ en ‘hoog’ ten opzichte van het eerste meetseizoen. De gewasgroep ‘midden’ is licht gestegen en heeft in het tweede meetseizoen veel meer waarnemingen. De gemiddelden voor gras en maïs zijn vrijwel gelijk met die van het eerste meetseizoen. De verschillen tussen de meetseizoenen

zijn terug te zien in het $Nmin_{nitraat}$ en niet in $Nmin_{ammonium}$.Vooral de maximum ammonium-waarden zijn onverwacht hoog, dus het effect van splitsing zal niet verwaarloosbaar zijn.

2.2.3 Overschotten

In tabel 6, 7 en 8 wordt voor de kandidaat-indicatoren perceelsoverschot, bedrijfsoverschot en MINAS-overschot, een overzicht gegeven volgens de nieuwe gewasgroep-indeling. Tabel 7 en 8, met bedrijfs- en MINAS-overschot, zijn wat artificiëel in die zin dat er gemiddeld is over de proefplekken van bedrijven die tot een bepaalde gewasgroep behoren. Alle proefplekken op één bedrijf hebben uiteraard hetzelfde bedrijfsgemiddelde, wat duidelijk naar voren komt bij de minimum en maximum waarden in de tabellen.

Tabel 6a. Overzicht van het perceelsoverschot (kg N/ha) per gewasgroep, volgens de nieuwe indeling, meetseizoen 2000-2001

Gewas-Groep	*Aantal Proefplekken*	*Gemiddelde*	*Minimum*	*Maximum*	*Mediaan*
Laag	75	150.7	-74.0	279.0	153.0
Midden	110	118.4	-25.0	328.0	96.0
Hoog	30	167.4	-37.0	382.0	155.5
Gras	144	179.6	-184.3	619.5	164.8
Maïs	71	78.1	-103.7	344.4	86.9

Tabel 6b. Overzicht van het perceelsoverschot (kg N/ha) per gewasgroep, volgens de nieuwe indeling, meetseizoen 2001-2002

Gewas-Groep	*Aantal Proefplekken*	*Gemiddelde*	*Minimum*	*Maximum*	*Mediaan*
Laag	44	113.0	-6.0	186.0	105.0
Midden	105	90.1	-29.3	335.0	74.0
Hoog	43	208.6	35.0	501.0	197.0
Gras	172	200.7	-34.0	522.0	207.4
Maïs	78	133.6	-59.0	312.7	102.4

Tabel 7a. Overzicht van het (werkelijk) bedrijfsoverschot (kg N/ha) per gewasgroep, volgens de nieuwe indeling, meetseizoen 2000-2001

Gewas-Groep	*Aantal Proefplekken*	*Gemiddelde*	*Minimum*	*Maximum*	*Mediaan*
Laag	53	110.7	55.0	234.0	102.0
Midden	90	108.0	55.0	355.0	102.0
Hoog	31	192.2	93.0	355.0	221.0
Gras	145	184.1	118.0	322.0	170.0
Maïs	62	204.2	118.0	322.0	215.0

Tabel 7b. Overzicht van het (werkelijk) bedrijfsoverschot (kg N/ha) per gewasgroep, volgens de nieuwe indeling, meetseizoen 2001-2002

Gewas-Groep	*Aantal Proefplekken*	*Gemiddelde*	*Minimum*	*Maximum*	*Mediaan*
Laag	44	132.9	77.0	176.0	141.0
Midden	105	128.0	77.0	311.0	129.0
Hoog	43	151.5	69.0	308.0	136.0
Gras	172	188.4	110.0	343.0	176.0
Maïs	78	228.3	110.0	343.0	224.5

Tabel 8a. Overzicht van het MINAS-overschot (kg N/ha) per gewasgroep, volgens de nieuwe indeling, meetseizoen 2000-2001

Gewas-Groep	*Aantal Proefplekken*	*Gemiddelde*	*Minimum*	*Maximum*	*Mediaan*
Laag	57	27.7	-36.0	114.0	40.0
Midden	97	16.6	-36.0	176.0	6.0
Hoog	31	95.3	6.0	176.0	149.0
Gras	145	116.5	-8.0	246.0	114.0
Maïs	67	129.0	68.0	246.0	114.0

Tabel 8b. Overzicht van het MINAS-overschot (kg N/ha) per gewasgroep, volgens de nieuwe indeling, meetseizoen 2001-2002

Gewas-Groep	*Aantal Proefplekken*	*Gemiddelde*	*Minimum*	*Maximum*	*Mediaan*
Laag	44	39.9	-19.0	108.0	33.0
Midden	107	30.1	-19.0	159.0	20.0
Hoog	43	46.8	-5.0	122.0	38.0
Gras	183	79.7	1.7	262.0	80.0
Maïs	80	134.0	1.7	262.0	159.0

Voor de perceelsoverschotten geldt dat deze voor gras een beetje en voor maïs behoorlijk zijn gestegen. Voor de akkerbouwgewassen is het gemiddelde perceelsoverschot gestegen voor de categorie ‘hoog’ en gedaald voor ‘midden’en ‘laag’. Het bedrijfsoverschot is voor de alle gewasgroepen gemiddeld wat hoger in het tweede meetseizoen. Het gemiddelde MINAS-overschot is voor de akkerbouwgewassen gedaald voor de categorie ‘hoog’en gestegen voor ‘midden en ‘laag’. Voor gras is het MINAS-overschot omlaag gegaan en voor maïs is het vrijwel gelijk gebleven. Als we vervolgens weten dat de nitraatconcentratie in het tweede meetseizoen allemaal gemiddeld lager zijn, valt te verwachten dat de relatie tussen de kandidaat-indicatoren en de nitraatuitspoeling niet verbeterd is ten opzichte van het eerste jaar.

2.2.4 Opmerkingen

Een overzicht van alle andere variabelen die gemeten zijn in het tweede meetseizoen en die een rol spelen in de analyse van de gegevens maar geen kandidaat-indicator zijn is te vinden in bijlage 1.

De duidelijke lagere nitraatconcentraties in het tweede meetseizoen voor de categoriën ‘laag’ en ‘hoog’ van de akkerbouwgewasssen komt overeen met een lagere Nmin. Voor maïs blijven zowel de nitraatconcentratie als Nmin gelijk. Voor gras en de categorie ‘midden’ daalt de nitraatconcentratie wel maar voor Nmin geldt dat niet. De soms toegenomen bedrijfsoverschotten sluiten wel bij de stijgende Nmin aan maar niet bij de dalende nitraatconcentraties.

Bij de nieuwe gewasindeling valt op dat er geen duidelijke verschillen tussen de groepen ‘laag’ en ‘midden’ zijn in nitraatconcentratie (Tabel 4a en 4b) en Nmin in het jaar 2000-2001. De groepen ‘laag’ en ‘midden’ zijn dus niet op een eenduidige logische manier te onderscheiden. De groep ‘hoog’ heeft daarentegen inderdaad hogere waarden voor de nitraatconcentratie, Nmin en de overschotten. Tenslotte valt op dat de gewassen die in de categorie ‘laag’ vallen gemiddeld een hoger perceelsoverschot hebben dan de gewassen in de gewasgroep ‘midden’.

Er is gekeken of er een directe relatie bestaat tussen nitraatconcentraties en de andere kandidaat-indicatoren. Die relatie wordt weergegeven als correlatie-coëfficiënt. De correlatie-coëfficiënt tussen nitraat en Nmin (gesommeerd over de laag 0-90 cm – mv. in oktober-november) is berekend op het niveau van de proefplekken. Beide variabelen worden immers per proefplek gemeten. Voor het perceelsoverschot geldt dat alle proefplekken op één perceel hetzelfde perceels-overschot hebben. Voor de berekening van de correlatie-coëfficiënt is eerst het gemiddelde van de nitraatconcentraties per perceel berekend. Alle proefplekken op één bedrijf hebben hetzelfde bedrijfsoverschot en hetzelfde MINAS-overschot. Correlatie-coëfficiënten met deze variabelen zijn gebaseerd op bedrijfsgemiddelden.

Tabel 9. Correlatie-coëfficiënt (r) tussen nitraatconcentratie en de vier andere kandidaat-indicatoren voor gras, maïs en akkerbouwgewassen. Correlatie-coëfficiënt is berekend per meetseizoen (2000-2001 en 2001-2002) op basis van de eenheid (plek, perceel, bedrijf) waarop de kandidaat-indicator is gemeten.

	2000-2001			2001-2002		
	Gras	Maïs	Akkerbouw	Gras	Maïs	Akkerbouw
Nmin	0.30	0.34	0.42	0.22	-0.04	0.18
Perceelsoverschot	0.14	-0.12	0.19	0.28	0.16	0.29
Bedrijfsoverschot	-0.04	-0.13	0.50	0.28	-0.62	0.14
MINAS-overschot	0.08	-0.07	0.56	0.04	-0.33	-0.02

De directe correlatie tussen Nmin en de nitraatconcentratie ligt in het tweede meetseizoen een stuk lager dan in het eerste meetseizoen. Gegeven de eerdere constatering dat de nitraatconcentratie in het tweede meetseizoen een stuk lager ligt terwijl dat voor de Nmin niet geldt, is dit niet onverwacht. De correlaties in het eerste meetseizoen zijn veranderd ten opzichte van de eerder door Hack-ten Broeke et al. (2003) gepresenteerde gegevens door verbeteringen in de data.

3 Methoden en technieken

3.1 Aspecten met betrekking tot de data

Op basis van de beschikbare gegevens is onderzocht of er een relatie bestaat tussen verschillende zogenoemde kandidaat-indicatoren. Hierbij is gebruik gemaakt van de gegevens van twee meetseizoenen, namelijk 2000-2001 en 2001-2002. Met een meetseizoen wordt bedoeld de N-giften, gewas ect. in de zomer, de Nmin-meting in het najaar en de nitraatconcentratie in het volgend voorjaar. Voor een deel van de proefplekken is Nmin alleen in het tweede seizoen gemeten. Op het grootste deel van de proefplekken zijn metingen van twee jaar beschikbaar. Van deze proefplekken is ongeveer 50 % van gewasgroep veranderd in het tweede meetseizoen ten opzichte van het eerste meetseizoen. Bovendien is binnen de akkerbouwgewassen vaak een ander gewas geteeld dat wel in dezelfde gewasgroep valt. Een ander gewas betekent een andere bemesting, andere gewasresten, ander oogsttijdstip etc. Op basis daarvan is in de analyse verondersteld dat de metingen over de twee jaar onafhankelijk zijn. Deze aanname is achteraf gecontroleerd door de correlatie tussen de residuen te berekenen en aannemelijk gebleken.

Op basis van het resultaat dat verkregen is na het eerste meetseizoen, is er ook in de analyse van de twee meetseizoenen samen voor gekozen om de akkerbouwgewassen, gras en maïs apart te analyseren. Bij de akkerbouwgewassen is ook de nieuwe indeling, met onderscheid naar laag, gemiddeld en hoog (te verwachten Nmin op 1 oktober in de bodem) meegenomen in de analyse. Daarnaast zijn er nog een aantal nieuwe mogelijk verklarende variabelen bij gekomen. Dit betreft de opsplitsing van Nmin in het nitraat- en het ammoniumdeel ($Nmin_{nitraat}$ en $Nmin_{ammonium}$), de actuele grondwaterstand op het tijdstip van de nitraatmeting, de GHG (gemiddeld hoogste grondwaterstand) en GLG (gemiddeld laagste grondwaterstand) van de profielbeschrijving, het al dan niet voorkomen van een veenlaagje in de bodem en de precieze neerslagsom- en -overschot, berekend voor de periode tussen de datum van de Nmin-meting en de datum van de nitraatmeting.

3.2 Regressie : aannames en gebruikte technieken

Een volledige beschrijving met betrekking tot het model-based versus design-based aspect van de dataset is te vinden in paragraaf 3.1 van de rapportage over één meetseizoen door Hack-ten Broeke et al. (2003). In de overall analyse van de twee meetseizoenen gezamenlijk is wederom gekozen voor een ongewogen regressie-analyse. Voor de gegeven modellen is steeds gecontroleerd of de parameterschattingen wezenlijk zouden veranderen indien toch een gewogen regressie (gewogen met 1/insluitkans) zou worden uitgevoerd. Dit is nergens het geval.

Een groot aantal variabelen heeft betrekking op de individuele proefplekken. Een ander deel heeft betrekking op perceels- of bedrijfsniveau. In de statistische analyse dient hiermee rekening te worden gehouden (zie paragraaf 3.2, Hack-tenBroeke et. al., 2003). Dit kan door middel van een REstricted Maximum Likelihood (REML) model. Voor de gegeven modellen is gecontroleerd in een REML-model of de variantiecomponenten van de bedrijven en/of percelen klein zijn ten opzichte van de restvariantie (onverklaarde variantie tussen de proefplekken). Indien dit zo is, is de variatie tussen bedrijven en tussen percelen binnen bedrijven van een vergelijkbare grootte als de variatie tussen proefplekken. Het effect van alle variabelen kan dan getoetst worden tegen deze variantie waarmee dus kan worden volstaan met een gewoon regressie-model. Voor alle modellen die gegeven worden in hoofdstuk 4 van dit rapport waren de variantiecomponenten van bedrijven en percelen heel klein ten opzichte van de restvariantie. Daarom is gekozen voor een gewoon regressie-model met één restterm. In de tekst van hoofdstuk 4 wordt verder dan ook geen aandacht meer besteed aan de REML-modellen.

Om te komen tot een selectie van verklarende variabelen is gestart met een voorwaartse, achterwaartse en stapsgewijze selectie van variabelen. Bij voorwaartse selectie wordt steeds een variabele aan het model toegevoegd, bij achterwaartse selectie wordt gestart met het volledige model (alle mogelijke verklarende variabelen erin) en valt steeds een variabele af. Stapsgewijze selectie is een combinatie van voorwaarts en achterwaarts. Het resultaat van deze selectie methoden is vervolgens gebruikt als startpunt voor de GenStat procedure RSEARCH. Hierin worden alle deelmodellen, dat wil zeggen alle mogelijke combinaties van verklarende variabelen, aangepast en de beste combinaties (met het hoogste percentage verklaarde variantie en/of laagste Mallow's C_p) worden gegeven. Vervolgens is uitgegaan van de modellen met het hoogste percentage verklaarde variantie waarbij alle parameters in het model nog significant zijn. Voor deze modellen is gecontroleerd of één van de andere variabelen nog een relevante bijdrage had m.b.t. de nitraatconcentratie. We spreken van een relevante bijdrage als het effect van een variabele op de nitraatconcentratie 10 mg/l of meer is (over 50% van de range van die variabele; het verschil tussen de upper en lower quartiel). Een parameter voor een verklarende variabele is misschien niet significant maar kan door zijn grootte (in combinatie met het bereik van de variabele) toch een relevant effect laten zien op de nitraatconcentratie. Dergelijke variabelen zijn dan alsnog wel opgenomen in de gepresenteerde modellen. De grens van 10 mg/l is een arbitraire keuze. Er is dus bij de selectie van variabelen in tegenstelling tot de analyse voor het eerste meetseizoen (Hack-ten Broeke et.al., 2003) niet uitsluitend naar significantie gekeken.

Voor alle modellen is de invloed gecontroleerd van waarnemingen met een groot residu en/of een grote hefboomwerking (high leverage). Indien deze punten een grote invloed hadden op de parameters van het regressie model en daarmee op de ligging van de lijn, zijn ze verwijderd en hebben ze uiteindelijk niet mee gedaan in de analyse. De gepresenteerde resultaten in hoofdstuk 4 zijn daarmee stabiel en hangen niet af van een paar proefplekken.

De enige variabelen waarvoor gecontroleerd is of er sprake is van een niet-lineair verband met de nitraatconcentratie zijn de neerslagsom en het neerslagoverschot. Een grotere neerslag kan er in eerste instantie voor zorgen dat Nmin terecht komt in het grondwater waarmee de nitraatconcentratie oploopt. Heel veel neerslag zou tot een verdunning en dus een lagere nitraatconcentratie kunnen leiden. Voor de andere variabelen wordt een min of meer lineair stijgend danwel dalend verband verwacht met de nitraatconcentratie.

Tenslotte is voor alle mogelijke verklarende variabelen gecontroleerd of de interactie met bodemgroep en/of Gtgroep significant is. Dit zou betekenen dat het verband tussen die verklarende variabele en de nitraatconcentratie verschilt per bodemgroep of per Gtgroep. Aangezien het aantal beschikbare proefplekken per gewas-Gtgroep of gewas-bodemgroep niet groot meer is, zijn interacties alleen in het model opgenomen indien zij significant zijn.

3.3 Neuraal netwerk

Bij het fitten van lineaire regressie modellen om de nitraatconcentratie te verklaren op de data van het eerste en tweede meetseizoen blijft het percentage verklaarde variantie laag. Er is onderzocht of het toepassen van een artificieel neuraal net tot een hogere verklaarde variantie leidt. Indien dit zo zal zijn betekent het dat toch belangrijke interactie of niet-lineaire termen gemist zijn.

Een uitgebreide beschrijving van de methode en de resultaten is te vinden in bijlage 2.

Er bestaan verschillende soorten neurale netten (Bishop, 1995; Ripley, 1996; Haykin, 1999), o.a. voor classificatie, voor clustering, voor feedbackprocessen. In dit onderzoek is gebruik gemaakt van het zogenaamde MeerLaags Perceptron (MLP) met Backpropagation en Bayesian Regularization (BR). Het MLP is in feite een niet-lineair regressiemodel. De term 'backpropagation' slaat op het algorithme waarmee de afgeleide van de errorfunctie wordt berekend (Rumelhart et al., 1986). De term Bayesian Regularization verwijst naar een manier om de grootte van de geschatte parameters in de hand te houden en daarmee overfitting te voorkomen. De berekeningen zijn uitgevoerd met behulp van MATLAB.

Er zijn afzonderlijke modellen gefit voor het eerste en het tweede meetjaar. Aangezien het de bedoeling is een zo groot mogelijk percentage verklaarde variantie te bereiken, heeft geen speciale modelselectie plaatsgevonden, maar zijn eenvoudigweg alle beschikbare variabelen in het model opgenomen. Wel zijn sommige verklarende variabelen weggelaten in verband met het grote aantal ontbrekende waarnemingen.
Bij het indelen van de data in trainingset, stopset en testset is een verhouding 60:20:20 gehanteerd. De records zijn random aan deze drie categorieën toegewezen.

Bij het fitten van neurale netten is het handig als de verklarende variabelen ongeveer dezelfde range hebben. Alle variabelen zijn daarom gestandaardiseerd tot variabelen met gemiddelde 0 en standaardafwijking 1.

Onderzocht is of een logtransformatie van de nitraatconcentratie tot een hoger percentage verklaarde variantie leidde. Ook is geprobeerd het aantal inputparameters te beperken door eerst een principale componentenanalyse op de verklarende variabelen uit te voeren, en alleen die componenten te behouden die meer dan 1 % van de variantie voor hun rekening nemen. Deze inputparameters zijn dan ongecorreleerd.

4 Resultaten van de analyse op basis van twee meetseizoenen

4.1 Akkerbouw

4.1.1 De beste regressie modellen

Op basis van de gewasgroepen a, b, r en t zijn er (gesommeerd over de twee seizoenen) 318 proefplekken in de akkerbouw waarop Nitraat en Nmin zijn gemeten. Nmin is op te delen in een nitraat- en een ammoniumdeel ($Nmin_{nitraat}$ en $Nmin_{ammonium}$). Deze twee variabelen hebben deze keer mee gedaan in de selectie van variabelen. Ook de nieuwe gewasgroep-indeling heeft mee gedaan in de selectie. De selectie is uiteindelijk gebaseerd op ongeveer 300 datapunten.

De combinatie löss met Gtgroep 1 en 2 is zwaar ondervertegenwoordigd in de dataset. Voor een aantal proefplekken geldt dat het eigenlijk een kleigrond betreft. Deze proefplekken vielen op (door een hoog residu of een sterke hefboomwerking) en doen daarom uiteindelijk niet meer mee in de analyse. Daarnaast zijn er twee proefplekken met een extreem hoge nitraatwaarde. Deze punten vallen ver buiten het bereik van de overige waarden en kunnen daarom van te grote invloed zijn op het resultaat. Deze punten doen daarom niet mee in de analyse. Tenslotte bleken er twee proefplekken sterk aan de regressielijn te trekken (high leverage) en hun invloed op de ligging van de lijn was groot. Ook deze punten doen dus niet mee in de analyse.

Uit de selectie komen 3 vergelijkbaar goede regressie-modellen naar voren. Alle modellen hebben als verklarende variabelen in het model :
Gt-groep
Bodemgroep
Gewasgroep (indeling op basis van verwachte Nmin in het najaar)
$Nmin_{Nitraat}$

De indeling van de bodem-, Gt- en gewasgroepen is besproken in paragraaf 2.1.1 en 2.2.1. De nieuwe gewasgroep-indeling blijkt in de regressie-analyse beter de verschillen in de nitraatconcentratie te kunnen verklaren dan de oorspronkelijke indeling. Er is geen duidelijk verschil tussen de gewasgroep ‘midden’ en ‘laag’. Bovendien hebben beide groepen dezelfde range voor Nmin. Daarom zijn bij de selectie de gewasgroepen ‘midden’ en ‘laag’ samengevoegd, resulterend in 2 groepen : ‘hoog’ en ‘midden+laag’. Door deze gewasgroepen samen te voegen worden 9 parameters bespaard zonder dat het model inboet aan voorspellingskracht. De drie variabelen Gt-, bodem- en gewasgroep resulteren in een verschillende constante C_i per bodem-Gt-gewasgroep-combinatie. De modellen zien er als volgt uit :

Model 1 : $Nitraat = C_i + a*Nmin_{nitraat} + b*Neerslagsom1 + d*Veen$
Model 2 : $Nitraat = C_i + a*Nmin_{nitraat} + b*Neerslagsom1 + d*Veen + e*C_Nverh$
Model 3 : $Nitraat = C_i + a*Nmin_{nitraat} + b*Neerslagsom1 + d*Veen + e*C_Nverh + f*Neerslagoverschot1 + g*Neerslag2precies + h*(Neerslag2precies)^2$

waarbij :

C_i = te schatten regressie-coefficiënt per bodem-Gt-gewascombinatie (i=1…24)

a,b,d,e,f,g en *h* = te schatten regressie-coëfficiënten

Nitraat = nitraatconcentratie in het voorjaar (mg/l)

$Nmin_{nitraat}$ = nitraatdeel van Nmin gemeten in het najaar (okt-dec) voor de laag 0-90 cm (kg N/ha)

Neerslagsom1 = Neerslagsom in de zomerperiode (mm), 1 april-1 oktober

Veen = het al dan niet aanwezig zijn van een veenlaagje (0 = niet aanwezig, 1 = wel aanwezig)

C_Nverh = C:N-verhouding voor de bouwvoor

Neerslagoverschot1 = Neerslagoverschot in de zomerperiode (mm), 1 april-1 oktober

Neerslag2precies = de exacte neerslagsom in de winterperiode (mm) tussen datum van de Nmin- en nitraat-meting.

In tabel 10 worden per model de parameterschattingen met standaardfout gegeven, het percentage verklaarde variantie (R^2_{adj}) en de standaardfout van het model (*sd*).

Tabel 10 Schatting van de regressie-coëffeciënten met standaardfout (se), het percentage verklaarde variantie (R^2_{adj}) en de standaardfout (sd) van model 1 t/m 3, akkerbouw, en het aantal datapunten waarop de modellen zijn gebaseerd..

Model	1	se	2	se	3	se
a	1.06	0.12	1.00	0.12	1.04	0.12
b	-0.203	0.053	-0.171	0.055	-0.453	0.142
d	-13.9	13.0	-12.1	13.0	-12.8	12.9
e			-2.01	0.91	-2.14	0.97
f					0.314	0.135
g					0.971	0.421
h					-0.0011	0.0005
R^2_{adj}	46.4 %		47.0 %		48.1 %	
Sd	55.7		55.5		54.9	
aantal data	304		301		301	

Tabel 11. Schatting van de constante C_i (in mg/l) per bodem-Gt-gewasgroep combinatie van model 1, akkerbouw.

	Gewas	*Hoog*			*Midden+laag*		
	Gt-groep	*1*	*2*	*3*	*1*	*2*	*3*
Model	*Bodem*						
1	Löss	97.4	139.2	158.4	69.0	110.8	130.0
	Z1	109.7	151.5	170.7	81.3	123.1	142.3
	Z2	126.3	168.1	187.4	97.9	139.7	159.0
	Z3	119.6	161.4	180.6	91.2	133.0	152.2
2	Löss	107.4	147.2	170.1	79.9	119.7	142.6
	Z1	135.7	175.5	198.4	108.2	148.0	170.9
	Z2	153.9	193.7	216.6	126.4	166.2	189.2
	Z3	148.9	188.7	211.6	121.4	161.2	184.2
3	Löss	42.6	77.5	100.7	21.1	56.0	79.2
	Z1	49.3	84.1	107.3	27.8	62.6	85.8
	Z2	64.4	99.2	122.4	42.9	77.8	101.0
	Z3	60.4	95.3	118.5	38.9	73.8	97.0

De constante is gegeven voor iedere gewas-bodem-Gtgroep combinatie die voorkomt in de dataset. De constante is de schatting voor de nitraatconcentratie gegeven dat alle andere verklarende variabelen in het model nul zijn. Om een idee te krijgen van de nitraatconcentratie bij een nulwaarde voor $Nmin_{nitraat}$ kan in Model 1 het beste gekeken worden bij de gemiddelde Neerslagsom1 van de dataset, die gelijk is aan 503. Voor de drie zandgronden (Z1, Z2 en Z3) bij Gtgroep=3 komt de nitraatconcentratie bij de gewasgroep 'hoog' dan neer op resp. 68.5, 85.2 en 78.4. Dit betekent dat zelfs wanneer de $Nmin_{nitraat}$ nul is de norm voor nitraat niet wordt gehaald. Voor de gewasgroep 'midden+laag' ligt het intercept bij 40.1, 56.8 en 50.0 voor resp. Z1, Z2 en Z3 bij Gtgroep=3. Voor de twee andere Gtgroepen ligt deze waarde een stuk lager.

De gemiddelde C:N-verhouding voor deze dataset is 21.0, het gemiddelde Neerslagoverschot in de zomerperiode is 101 mm. en de gemiddelde exacte neerslagsom in de winterperiode is is 409 mm.

De helling van nitraat versus $Nmin_{nitraat}$ is voor alle drie de modellen gelijk. Het effect van veen is in geen van de modellen significant maar de grootte is wel relevant, het verlaagt de voorspelde nitraatconcentratie namelijk met ongeveer 13 mg/l, zodat het al dan niet aanwezig zijn van veenlaagjes wel in het model is opgenomen. Het percentage verklaarde variantie en de standaardfout (van het model) is voor de drie modellen vrijwel gelijk. Dit betekent dat een voorspelling met het derde model niet veel nauwkeuriger zal zijn dan met het eerste model. Dit model is toch gegeven omdat het wel voldoet aan de eisen zoals die gesteld zijn in paragraaf 3.2.

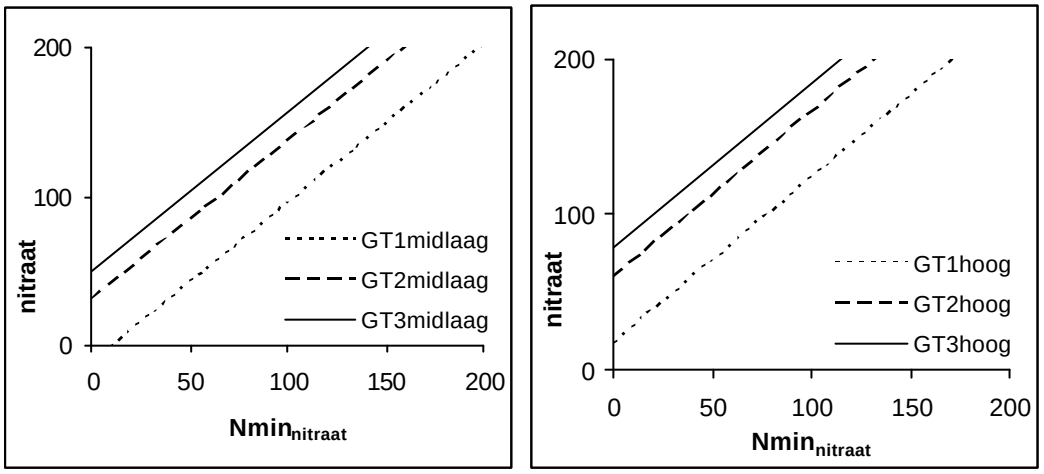
Het effect van de Gtgroep is in alle drie de modellen significant en relevant. Voor Gt-groep 2 ligt de nitraatconcentratie gemiddeld 20 mg/l lager dan bij Gt-groep 3 en voor Gt-groep 1 zo'n 60 mg/l lager. Ook het effect van bodemgroep is significant en relevant. Een groot deel van dit effect wordt veroorzaakt door löss die gemiddeld een veel lagere nitraatconcentratie heeft dan de zandgronden.

Het effect van de neerslagsom in de zomerperiode is significant en relevant. Het teken van de parameter is negatief. Dat wil zeggen dat de nitraatconcentratie in het voorjaar afneemt bij een toenemende neerslag in de zomer van het jaar er voor (20 mg/l bij 100 mm neerslag). Dit kan het resultaat zijn van een beduidend hogere N-opname door goede vochtvoorziening of juist van uitspoeling van N in de zomer, die je in het volgende voorjaar niet meer kunt meten. Dit effect lijkt groter in Model 3 maar wordt daar gecompenseerd door een positieve parameter voor het neerslagoverschot in de zomer. Het effect van de C:N-verhouding is significant en relevant en ongeveer gelijk in Model 2 en 3. Een verhoging van de C:N-verhouding van 5 punten (bijv. van 10 naar 15) geeft een afname in de nitraatconcentratie van 10 mg/l. Een mogelijke verklaring hiervoor is dat naarmate de C:N-verhouding hoger wordt, de netto mineralisatie van N (= verschil tussen mineralisatie en immoblilisatie) kleiner wordt. De neerslagsom in de winterperiode laat een kwadratisch verband zien met de nitraatconcentratie. Een toenemende neerslag geeft in eerste instantie een verhoging van de concentratie maar vanaf 423 mm neerslag neemt de

nitraatconcentratie weer af en is het effect van verdunning blijkbaar groter dan het effect van uitspoeling.

De modellen hebben geen of slechts één punt met een sterke hefboomwerking en zijn dus zeer stabiel. Het weglaten van een zestal punten met een groot residue bij model 1 resulteert in een stijging van het percentage verklaarde variantie (52.2%) en een daling van de standaardfout (sd=48.6). Er zijn echter geen valide reden te bedenken om deze punten weg te laten dus wordt uitgegaan van de gepresenteerde modellen.

De voorspellingen op basis van regressie-model 1 worden geïllustreerd in Figuur 1. Voor de drie grondwatertrap-groepen en de twee gewasgroepen wordt de voorspelde nitraatwaarde weergegeven voor bodemgroep Z3 bij een range van $Nmin_{nitraat}$ .Voor Neerslagsom in de winter is een vaste instelwaarde van 503 mm. gekozen, het gemiddelde van de data waarop het regressie-model is gefit.



Figuur 1a en 1b. Voorspelling van de nitraatconcentratie op de y-as bij een range van $Nmin_{nitraat}$ op basis van model 1, akkerbouw, bodemgroep Z3, geen veen en Neerslagsom1 = 503 : 1a. gewasgroep 'midden+laag',1b. gewasgroep 'hoog' . GT geeft de Gt-groep aan, zoals bescgreven in hoofdstuk 2.

Model 1 en 3 zijn tevens gefit op de data van de twee meetjaren apart. De resultaten worden gegeven in Bijlage 3a. In grote lijnen is het resultaat in Tabel 10 zoals te verwachten is het gemiddelde van de modellen voor de twee jaren apart. De helling van $Nmin_{nitraat}$ ligt in het eerste meetjaar wat hoger en in het tweede meetjaar juist wat lager dan de helling die gegeven is in Tabel 10. De constanten zijn ook behoorlijk wat hoger in het eerste meetjaar ten opzichte van het tweede jaar. Het percentage verklaarde variantie is in het eerste meetjaar wat hoger maar dat geldt ook voor de standaardfout van het model zodat niet gesteld kan worden dat het ene jaar een veel beter resultaat geeft dan het andere jaar. Merkwaardig echter zijn het positieve effect van löss en veen in het eerste meetseizoen. Nader onderzoek wijst uit dat dit het gevolg is van een verstrengeling met de neerslagsom in de zomer. De löss-proefplekken hebben toevallig een veel hogere neerslagsom in de zomerperiode. Het effect van neerslag is verdubbeld (van –0.20 naar –0.46) waardoor de proefplekken,

in combinatie met ook een lage $Nmin_{nitraat}$, veel te laag uitkomen. Dit wordt vervolgens gecompenseerd door een positief effect van löss. Hetzelfde speelt voor de proefplekken met veen. Een model zonder neerslagsom in de winter en $Nmin_{nitraat}$ geeft een effect voor löss van –21 mg/l en een effect voor veen van –22 mg/l.

4.1.2 Effect van andere variabelen

In deze paragraaf wordt verslag gedaan van alle andere variabelen en factoren die expliciet zijn onderzocht. De resultaten worden gegeven met de argumenten waarom zij uiteindelijk niet terecht zijn gekomen in de beste modellen.

Indien Nmin in het model wordt opgenomen in plaats van het nitraatdeel van Nmin ($Nmin_{nitraat}$) daalt het percentage verklaarde variantie van Model 1 van 46.4 naar 41.0% en wordt de standaardfout van het model 58.5 (mg/l). De helling voor nitraat (mg/l) versus Nmin (kg/ha) wordt 0.61 (met se=0.09). Het verschil tussen de gewasgroepen wordt een stuk groter en ook het verschil tussen Gtgroep1 en Gtgroep3 wordt groter. Nader onderzoek wijst uit dat dit wordt veroorzaakt door verschillen in de $Nmin_{nitraat}$/Nmin-verhouding. Deze verhouding is gemiddeld het laagst bij gewasgroep 'midden+laag' op Gtgroep 1 (0.32) en het hoogste bij gewasgroep 'hoog' en Gtgroep 3 (0.87). Voor alle andere combinaties ligt de gemiddelde verhouding tussen $Nmin_{nitraat}$ en Nmin rond 0.70. Voor de overige parameters (bodemgroep, Neerslagsom1 en veen) blijft het model ongeveer gelijk. Een aanvullende reden om $Nmin_{nitraat}$ als verklarende variabele te verkiezen boven Nmin, is dat het ammonium-deel van Nmin minder uitspoelingsgevoelig is dan nitraat omdat ammonium beter aan de bodem adsorbeert (vnl. aan klei en organische stof). Bovendien is de nitrificatie (omzetting van ammonium in nitraat) in de winterperiode waarschijnlijk beperkt vanwege de lage bodemtemperatuur en het hoge bodemvochtgehalte.

Een mogelijk alternatief vierde model is Model 1 met daaraan toegevoegd het MINAS-bedrijfsoverschot. De parameter voor het MINAS-bedrijfsoverschot is niet significant maar wel positief (p=0.096) en zou qua grootte relevant kunnen zijn. De overige parameters en modelkenmerken van model 1 veranderen nauwelijks. Bij nader onderzoek blijkt het verband tussen het MINAS-bedrijfsoverschot en de nitraatconcentratie wel aanwezig bij de gewasgroep 'hoog' (p=0.309) maar niet bij de gewasgroep 'midden+laag' (p=0.005). Er is dus sprake van een interactie tussen gewasgroep en MINAS-bedrijfsoverschot. Door het geringe aantal data in de gewasgroep 'hoog' en het nog geringer aantal bedrijven waarop deze data zijn gebaseerd is de basis om deze interactie in het model op te nemen erg smal. Het lijkt beter dit punt opnieuw te onderzoeken wanneer de gegevens van alle drie meetseizoenen beschikbaar zijn.

$DOC_{grondwater}$ (opgelost organisch koolstof in het grondwater (mg/l)) heeft wel een significante en relevante bijdrage in de verklaring van de variatie van de nitraatconcentratie. $DOC_{grondwater}$ opgenomen in Model 1 geeft een percentage verklaarde variantie van 54.4 % met sd=58.9. Echter de $DOC_{grondwater}$-meting

ontbreekt op 136 datapunten. Bovendien is $DOC_{grondwater}$ geen potentiële indicator omdat voor deze meting een grondwatermonster moet worden genomen waarin dan ook de nitraatconcentratie zelf gemeten kan worden. Ter informatie is het model met $DOC_{grondwater}$ gegeven in Bijlage 4a.

Interacties tussen de clusterindelingen gewas-, Gt- en bodemgroep is voor geen van de combinaties significant, met andere woorden het effect van Gt op nitraat verschilt niet tussen de gewasgroepen. Vervolgens is voor alle gemeten variabelen in de dataset gecontroleerd of zij een significante interactie met gewas-, bodem- dan wel Gtgroep hebben. Er is geen interactie gevonden tussen de $Nmin_{nitraat}$ en één van de clusterindelingen. Dit betekent dat het gevonden verband (de helling) tussen $Nmin_{nitraat}$ en de nitraatconcentratie gelijk is voor alle combinaties en dat er alleen sprake is van een verschillend intercept. De interactie tussen gewasgroep en de C:N-verhouding is net niet significant. Opnemen van deze interactie zou resulteren in de gegeven helling voor gewasgroep 'midden + laag' maar in een steilere helling voor 'hoog'. Dit wijst er weer op dat een hoge Nmin tot relatief hoge nitraatconcentraties in het grondwater leidt.

Een aantal gemeten variabelen (zoals Ntotaal, Ctotaal, organisch stof en potentiële mineralisatie) hebben een significante interactie met bodemgroep, hun effect op nitraat hangt dus af van de bodemgroep. Bij nader onderzoek blijkt dit vooral te komen door de lössgronden die in combinatie met zo'n variabele sterk afwijken van de zandgronden. Löss is echter zo'n beperkt deel van de dataset dat het opnemen van dergelijke interacties niet leidt tot stabiele modellen.

De neerslaggegevens resulteren af en toe in significante interacties met bodem-, Gt- of gewasgroep. Op het niveau van de individuele parameters zijn de verschillen veelal niet meer significant. Bovendien zijn de neerslagsommen gebaseerd op één gemiddelde per bedrijf waarmee de basis voor een interactie heel smal wordt.

Bij ongeveer 20% van de proefplekken is een groenbemester toegepast na de teelt van het akkerbouwgewas. Onderzocht is of dit mogelijk een effect heeft op de nitraat-concentratie. In eerste instantie is een zeer significante interactie tussen de gewasindeling en de groenbemester gevonden. Deze interactie bleek veroorzaakt te worden door 4 proefplekken (2 uit iedere gewasgroep). Zonder deze proefplekken is de interactie niet meer significant maar zijn de interactie-effecten nog steeds wel relevant. De interactie bestaat eruit dat het toepassen van een groenbemesters na een gewas uit de gewasgroep 'hoog' een verhoging van de nitraatconcentratie tot gevolg heeft terwijl dat bij de gewasgroep 'midden+laag' tot een verlaging van de nitraatconcentratie leidt. De grootte van de effecten (zonder de vier extreme proefplekken) is resp. +13 en –13 mg/l. Dit resulteert er logischerwijs in dat wanneer alleen gekeken wordt naar het hoofdeffect van groenbemester er geen enkel effect (significant noch relevant) wordt gevonden. Uit nader onderzoek blijkt dat de een groot deel van de proefplekken met een gewas uit gewasgroep 'hoog' met groenbemester laat in het seizoen nog een mestgift heeft gehad. Dit geldt niet voor de proefplekken met een gewas uit gewasgroep 'hoog' zonder groenbemester.

Daarmee is de interactie verstrengeld met een mestgift (en geen reële interactie) en wordt niet in het model opgenomen.

4.2 Veehouderij

De twee gewassen, gras en snijmaïs, die voorkomen op veehouderij-bedrijven in de dataset zijn apart geanalyseerd en worden hieronder apart besproken.

4.2.1 Gras, de beste regressie modellen

Er zijn (gesommeerd over de twee seizoenen) 334 proefplekken met gras waarop Nitraat en Nmin is gemeten. Nmin is op te delen in een nitraat- en een ammoniumdeel ($Nmin_{nitraat}$ en $Nmin_{ammonium}$). Deze twee variabelen hebben deze keer mee gedaan in de selectie van variabelen. De selectie van variabelen is uiteindelijk gebaseerd op ongeveer 270 proefplekken als gevolg van missende waarden voor een groot aantal, mogelijk belangrijke, variabelen.

De combinatie löss met Gtgroep 1 en 2 komt niet voor in de dataset. Voor een drietal proefplekken geldt dat het eigenlijk een kleigrond betreft. Deze proefplekken vielen op door een heel groot residu en doen daarom uiteindelijk niet meer mee in de analyse. Daarnaast is er één proefplek met een extreem hoge nitraatwaarde. Dit punt valt ver buiten het bereik van de overige waarden en kan daarom van grote invloed zijn op het resultaat. Dit punt doet niet mee in de analyse.

Uit de selectie komen 2 vergelijkbaar goede regressie-modellen naar voren. Beide modellen hebben als verklarende variabelen in het model:
- Gt-groep
- Bodemgroep
- $Nmin_{nitraat}$

De indeling van de bodem- en Gtgroepen is besproken in paragraaf 2.1.1. Deze twee variabelen resulteren in een verschillende constante C_i per bodem-Gt-combinatie. De modellen zien er als volgt uit:

Model 1: $Nitraat = C_i + a*Nmin_{nitraat} + b*Veen$
Model 2: $Nitraat = C_i + a*Nmin_{nitraat} + b*Veen + d*C_Nverh + e*GHG + f*PotMin$

waarbij :

C_i	= te schatten regressie-coefficiënt per bodem-Gt-combinatie (i=1…10)
a,b,d,e en f	= te schatten regressie-coëfficiënten
$Nitraat$	= nitraatconcentratie in het voorjaar (mg/l)
$Nmin_{nitraat}$	= nitraatdeel van Nmin gemeten in het najaar (okt-dec) voor de laag 0-90 cm (kg N/ha)

Veen	= het al dan niet aanwezig zijn van een veenlaagje (0 = niet aanwezig, 1 = wel aanwezig)
C_Nverh	= C:N-verhouding voor de bouwvoor
GHG	= gemiddeld hoogste grondwaterstand (cm onder maaiveld)
PotMin	= potentiële mineralisatie die gezien kan worden als een maat voor de afbreekbaarheid van organische stikstof (mg N per kg grond per dag)

In Tabel 12 worden per model de parameterschattingen met standaardfout gegeven, het percentage verklaarde variantie (R^2_{adj}) en de standaardfout van het model (*sd*).

Tabel 12. Schatting van de regressie-coëffeciënten a t/m f met standaardfout (se), het percentage verklaarde variantie (R^2_{adj}) en de standaardfout (sd) van model 1 en 2, grasland, en het aantal datapunten waarop de modellen zijn gebaseerd.

	Model 1	*se*	*Model 2*	*se*
a	0.76	0.10	0.76	0.10
b	-37.9	15.1	-13.9	16.3
d			-3.4	1.4
e			0.4	0.2
f			-10.4	3.3
R^2_{adj}	17.3		23.0	
Sd	50.9		49.9	
aantal data	330		312	

Tabel 13. Schatting van de constante C_i (in mg/l) per bodem-Gt-combinatie van model 1 en 2, grasland.

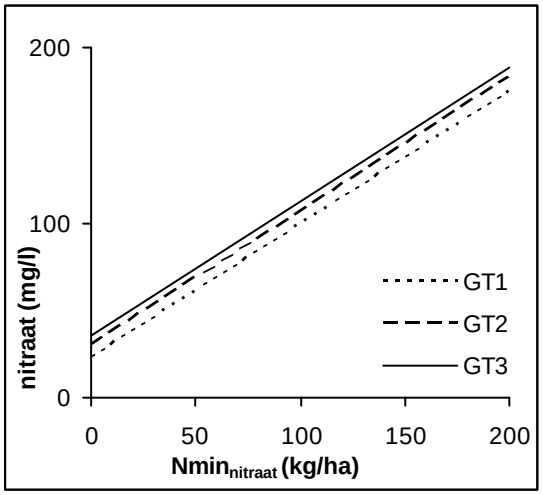
	Model1			*Model 2*		
Gt-groep	*1*	*2*	*3*	*1*	*2*	*3*
Bodem						
Löss	*	*	3.0	*	*	1.5
Z1	19.5	27.0	32.4	84.1	78.8	60.9
Z2	17.9	25.4	30.7	93.6	88.3	70.4
Z3	22.5	30.0	35.4	94.0	88.7	70.8

De constante is gegeven voor iedere bodem-Gt-groep combinatie die voorkomt in de dataset. De constante is de schatting voor de nitraatconcentratie gegeven dat alle andere verklarende variabelen in het model nul zijn. In Model 1 hoort de constante bij $Nmin_{nitraat}$=0 en veen=nee (en is dus het snijpunt met de Y-as als $Nmin_{nitraat}$ op de X-as staat). De gemiddelde waarde van C:N-verhouding in deze dataset is 17.0, de gemiddelde GHG is 84.0 cm -mv en de gemiddelde *PotMin* is 2.5 mg N per kg grond per dag.

De helling van de nitraatconcentratie versus $Nmin_{nitraat}$ is in beide modellen gelijk maar een stuk lager dan bij de modellen voor akkerbouw. Het effect van veen is in Model 2 een stuk kleiner en wordt blijkbaar opgevangen door verschillen in de C:N-verhouding of de GHG. Het percentage verklaarde variantie van het tweede model is wat hoger maar de standaardfout is niet veel kleiner. Daarmee zullen voorspellingen met het tweede model niet veel nauwkeuriger zijn. De verklaarde variantie is teleurstellend laag in vergelijking met akkerbouw.

Het effect van Gtgroep is over het geheel niet significant maar de verschillen tussen de Gt-groepen zijn wel groot en daarmee relevant zodat Gtgroep wel in het model is opgenomen. De gemiddelde nitraatconcentratie van Gtgroep 1 ligt 12.9 mg/l lager dan bij Gtgroep 3. Het effect van bodemgroep bestaat vooral uit het significante verschil tussen löss en de zandgronden. Het effect van de aanwezigheid van veenlaagjes in het profiel is in Model 2 niet meer significant maar een verschil van 15 mg/l in de nitraatconcentratie is wel een relevant verschil. De parameters van GHG, C:N-verhouding en *PotMin* zijn significant. Een diepere GHG heeft gemiddeld een hogere nitraatconcentratie. Een hogere *PotMin* en/of een hogere C:N-verhouding leidt tot een afname van de nitraatconcentratie.

De voorspellingen op basis van regressie-model 1 worden geïllustreerd in Figuur 2. Voor de drie grondwatertrap-groepen wordt de voorspelde nitraatwaarde weergegeven voor bodemgroep Z3 bij een range van $Nmin_{nitraat}$.



Figuur 2. Voorspelling van de nitraatconcentratie op de y-as bij een range van $Nmin_{nitraat}$ op basis van model 1, grasland, bodemgroep Z3, geen veen:

De modellen gefit op de data van de twee meetjaren apart worden gegeven in Bijlage 3b. In grote lijnen is het resultaat in Tabel 11 het gemiddelde van de modellen voor de twee jaren apart. De helling van $Nmin_{nitraat}$ is in beide meetjaren vrijwel gelijk. De constanten zijn behoorlijk wat hoger in het tweede meetjaar ten opzichte van het eerste jaar. Dit is in tegenstelling met wat is gevonden bij de akkerbouw. Het percentage verklaarde variantie is in het eerste meetjaar wat hoger maar dat geldt ook voor de standaardfout van het model zodat niet gesteld kan worden dat het ene jaar een veel beter resultaat geeft dan het andere jaar.

4.2.2 Gras, effect van andere variabelen

In deze paragraaf wordt verslag gedaan van alle andere variabelen en factoren die expliciet zijn onderzocht. De resultaten worden gegeven met de argumenten waarom zij uiteindelijk niet terecht zijn gekomen in de beste modellen.

Indien Nmin in het model wordt opgenomen in plaats van het nitraatdeel van Nmin daalt het percentage verklaarde variantie Model 1 van 17 naar 12 %. De helling voor Nmin is 0.53 (se=0.09). Het verschil tussen löss en de zandgronden wordt groter maar verder blijft het model ongeveer gelijk.

De neerslagsom in de winterperiode is exact berekend op basis van de meetdatum van zowel Nmin als de nitraatconcentratie. Deze neerslagsom samen met zijn kwadratische term levert significante parameters in het model op (maar met een tegengesteld teken ten opzichte van Model 3 bij akkerbouw). Het percentage verklaarde variantie en de sd van Model 2 veranderen niet bij het opnemen van deze variabelen. Het effect van deze parameters op nitraatconcentratie is zo klein (2 mg/l over het totale bereik) dat besloten is deze variabelen niet in het model op te nemen.

$DOC_{grondwater}$ (opgelost organisch koolstof in het grondwater (mg/l)) heeft wel een significante en relevante bijdrage in de verklaring van de variatie van de nitraatconcentratie. $DOC_{grondwater}$ opgenomen in model 1 geeft een percentage verklaarde variantie van 25 % met sd=47.3. Echter de $DOC_{grondwater}$-meting ontbreekt op 60 proefplekken. Bovendien is $DOC_{grondwater}$ geen potentiële indicator omdat de meting vergelijkbaar moeilijk en duur is met de nitraatmeting in het grondwater. Ter informatie is het model met $DOC_{grondwater}$ gegeven in bijlage 4b.

De interactie tussen Gtgroep en bodemgroep is niet significant (p=0.33 bij model 1).

Vervolgens is voor alle verklarende variabelen in de dataset gecontroleerd of zij een significante interactie met bodemgroep dan wel Gtgroep hebben. De interactie tussen $Nmin_{nitraat}$ en Gtgroep is niet significant. Opname van deze interactie in het model zou resulteren in een minder steile helling voor Gtgroep1 en een iets steilere helling voor Gtgroep2. In Model 1 is de interactie tussen $Nmin_{nitrtaat}$ en bodemgroep wel significant. Opname van de interactie in het model geeft veel steilere helling voor Z3 (a=1.05) en een minder steile helling voor de overige gronden. De steile helling voor Z3 blijkt te worden veroorzaakt door een vijftal proefplekken. Zonder deze proefplekken wordt de helling 0.83. De helling voor löss is gebaseerd op heel weinig data, daarmee onbetrouwbaar en niet significant verschillend van Z3. Ook de helling van Z2 verschilt niet significant van Z3. De veel minder steile helling voor Z1 wordt veroorzaakt door één proefplek. Zonder deze proefplek in de analyse verschilt ook Z1 niet meer significant van Z3. Uiteindelijk blijkt de significante interactie tussen $Nmin_{nitrtaat}$ en bodemgroep dus af te hangen van zes proefplekken. Daarmee zijn de modellen met de interactie-termen niet stabiel en worden niet als mogelijk model gepresenteerd. Het lijkt wel zinvol dit punt opnieuw te onderzoeken als de data van drie meetseizoenen beschikbaar zijn.

Ook alle andere significante interacties zijn verder onderzocht maar leveren geen bruikbare modellen op (bijv. de interactie Nafvoergewas en Gtgroep is significant maar resulteert in een positief verband tussen de Nafvoergewas en de nitraatconcentratie voor Gtgroep 3 en een negatief verband voor Gtgroep 1 en 2).

4.2.3 Maïs, de beste regressie modellen

Er zijn (gesommeerd over de twee seizoenen) 120 proefplekken met maïs waarop Nitraat en Nmin is gemeten. Nmin is op te delen in een nitraat- en een ammoniumdeel ($Nmin_{nitraat}$ en $Nmin_{ammonium}$). Deze twee variabelen hebben deze keer mee gedaan in de selectie van variabelen. De selectie van variabelen is uiteindelijk gebaseerd op ongeveer 100 proefplekken als gevolg van missende waarden voor een groot aantal mogelijk belangrijke variabelen.

De combinatie löss met Gtgroep 1 en 2 komt niet voor in de dataset. De combinatie löss met Gtgroep 3 heeft maar drie proefplekken waarvan er twee eigenlijk kleigrond blijken te zijn. Deze twee proefplekken vielen op door een groot residue en doen daarom uiteindelijk niet meer mee in de analyse. Vervolgens is besloten ook de derde proefplek met löss te laten vervallen. Er is één proefplek met een extreem hoge nitraatwaarde. Dit punt valt ver buiten het bereik van de overige waarden en kan daarom van grote invloed zijn op het resultaat. Dit punt doet niet mee in de analyse. Tenslotte doen er drie proefplekken (in het tweede seizoen) niet mee omdat zij het vorige seizoen gras hadden en nu opvielen door hun extreem hoge residu. Het effect van voorvrucht zal nader worden bestudeerd als de gegevens van 3 meetseizoenen worden geanalyseerd.

Uit de selectie komen 3 vergelijkbaar goede regressie-modellen naar voren. Alle modellen hebben als verklarende variabelen in het model :
- Gt-groep
- Bodemgroep
- $Nmin_{nitraat}$

De indeling van de bodem- en Gtgroepen is besproken in paragraaf 2.1.1. Deze twee variabelen resulteren in een verschillende constante C_i per bodem-Gt-combinatie. De modellen zien er als volgt uit :

Model 1 : $Nitraat = C_i + a*Nmin_{nitraat} + b*Veen$
Model 2 : $Nitraat = C_i + a*Nmin_{nitraat} + b*Veen + e*GHG + f*PotMin$
Model 3 : $Nitraat = C_i + a*Nmin_{nitraat} + b*Veen + e*GLG + f*Neerslagsom1$

waarbij :

C_i	= te schatten regressie-coefficiënt per bodem-Gt-combinatie (i=1…9)
a,b,d,e en f	= te schatten regressie-coëfficiënten
$Nitraat$	= nitraatconcentratie in het voorjaar (mg/l)
$Nmin_{nitraat}$	= nitraatdeel van Nmin gemeten in het najaar (okt-dec) voor de laag 0-90 cm (kg N/ha)
$Veen$	= het al dan niet aanwezig zijn van een veenlaagje (0 = niet aanwezig, 1 = wel aanwezig)
GHG	= gemiddeld hoogste grondwaterstand (cm -mv)
$PotMin$	= potentiële mineralisatie die gezien kan worden als een maat voor afbreekbaarheid van organische stikstof (mg per kg grond per dag

GLG = gemiddeld laagste grondwaterstand (cm -mv)
Neerslagsom1 = Neerslagsom in de zomerperiode (mm)

In tabel 14 worden per model de parameterschattingen met standaardfout gegeven, het percentage verklaarde variantie (R^2_{adj}) en de standaardfout van het model (*sd*). Voor ieder model is sprake van één regressielijn waarbij iedere combinatie van Gtgroep en bodemgroep een andere constante heeft.

Tabel 14. Schatting van de regressie-coëffeciënten a t/m f met standaardfout (se), het percentage verklaarde variantie (R^2_{adj}) en de standaardfout (sd) van model 1, 2 en 3, snijmaïs, en het aantal datapunten waarop de modellen zijn gebaseerd.

	Model 1	*se*	*Model 2*	*se*	*Model 3*	*se*
a	0.69	0.25	0.65	0.25	0.52	0.27
b	-23.0	31.0	-8.5	33.9	-21.5	31.0
e			0.55	0.34	0.80	0.43
f			-13.0	10.6	-0.32	0.17
R^2_{adj}	20.8 %		23.5 %		22.8 %	
Sd	64.4		60.1		63.6	
aantal	113		103		103	

Tabel 15. Schatting van de constante C_i (in mg/l) per bodem-Gt-combinatie van model 1, 2 en 3, snijmaïs.

Model	*1*			*2*			*3*		
Gt-groep	*1*	*2*	*3*	*1*	*2*	*3*	*1*	*2*	*3*
Bodem									
Z1	-3.6	8.1	67.7	-9.9	-7.8	21.4	38.1	31.4	78.9
Z2	9.2	20.8	80.4	8.5	10.7	39.8	56.3	49.7	97.1
Z3	15.6	27.2	86.8	4.3	6.5	35.6	65.4	58.7	106.2

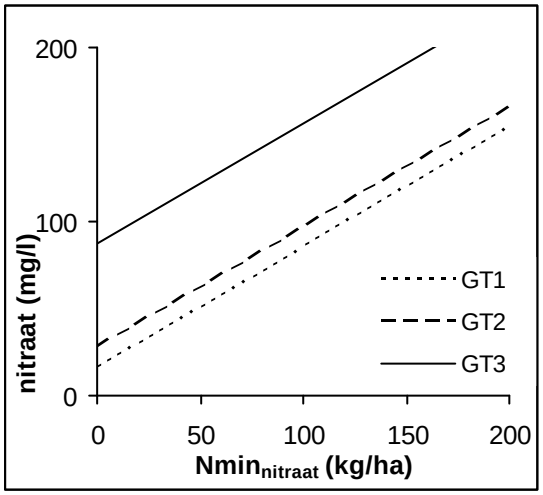
De constante is gegeven voor iedere bodem-Gt-groep combinatie (die voorkomt in de dataset). De constante is de schatting voor de nitraatconcentratie gegeven dat alle andere verklarende variabelen in het model nul zijn. In model 1 hoort de constante bij $Nmin_{nitraat}$=0 en veen=nee (en is dus het snijpunt met de Y-as als $Nmin_{nitraat}$ op de X-as staat). De gemiddelde waarde van de GHG voor deze dataset is 74.8, de gemiddelde *PotMin* is 1.2, de gemiddelde GLG is 158.4 en de gemiddelde *Neerslagsom1* is 458 mm.

De helling van $Nmin_{nitraat}$ komt voor Model 1 en 2 goed overeen en ligt iets lager voor model 3. Het effect van veen is in Model 2 een stuk kleiner en wordt net als bij gras blijkbaar opgevangen door verschillen in de GHG en/of PotMin. Het percentage verklaarde variantie van het tweede model is wat hoger en de standaardfout is kleiner dan de overige modellen. Daarmee zullen voorspellingen met het tweede model iets nauwkeuriger zijn. Model 3 is vergelijkbaar met model 1 maar heeft meer informatie nodig. De verklaarde variantie van deze modellen is iets hoger dan bij gras, maar in vergelijking met akkerbouw nog steeds erg laag.

Het effect van Gtgroep is groot en significant in de Modellen 1 en 3. Gtgroep 3 heeft een veel hogere nitraatconcentratie dan de Gtgroepen 1 en 2. Het effect van Gtgroep is nog steeds relevant maar niet meer significant in Model 2. Het effect van

bodemgroep is voor geen van de modellen significant maar de verschillen tussen de drie zandgronden zijn wel groot (groter dan 10 mg/l) en daarmee relevant zodat bodemgroep wel in het model is opgenomen. Ook de variabelen GHG en PotMin hebben geen significante parameter maar het effect van deze variabelen is qua grootte wel relevant zodat ze wel in Model 2 zijn opgenomen. De parameters van GLG en *Neerslagsom1* zijn bijna significant (p=0.06)

De voorspellingen op basis van regressie-model 1 worden geïllustreerd in Figuur 3. Voor de drie grondwatertrap-groepen wordt de voorspelde nitraatwaarde weergegeven voor bodemgroep Z3 bij een range van $Nmin_{nitraat}$.



Figuur 3. Voorspelling van de nitraatconcentratie op de y-as bij een range van $Nmin_{nitraat}$ op basis van model 1, maïs, bodemgroep Z3, geen veen.

De Modellen 1 en 2 zijn gefit op de data van de twee meetjaren apart en de resultaten worden gegeven in Bijlage 3c. In grote lijnen is het resultaat in Tabel 12 wel het gemiddelde van de modellen voor de twee jaren apart maar de verschillen tussen de jaren zijn groot. De helling van $Nmin_{nitraat}$ is in het eerste meetjaar 1.36 en in het tweede meetjaar 0.28 (bij Model 1). De constanten zijn net als bij gras behoorlijk wat hoger in het tweede meetjaar en dus in tegenstelling met hetgeen is gevonden bij de akkerbouw. Het percentage verklaarde variantie is in het eerste meetjaar iets hoger maar dat geldt ook voor de standaardfout van het model zodat niet gesteld kan worden dat het ene jaar een veel beter resultaat geeft dan het andere jaar. Het effect van veen is in het eerste meetjaar gebaseerd op één proefplek met een lage $Nmin_{nitraat}$ (35) en een hoge nitraatconcentratie (102) en geeft daarom een zeer onbetrouwbare maar positieve schatting van het effect. Bij maïs is al opgemerkt dat het verstandig lijkt om in de vervolgstudie de punten met veen buiten beschouwing te laten.

4.2.4 Maïs, effect van andere variabelen

In deze paragraaf wordt verslag gedaan van alle andere variabelen en factoren die expliciet zijn onderzocht. De resultaten worden gegeven met de argumenten waarom zij uiteindelijk niet terecht zijn gekomen in de beste modellen.

Indien Nmin in het model wordt opgenomen in plaats van het nitraatdeel van Nmin dan is de helling voor Nmin niet meer significant (a=0.34, se=0.22) verschillend van nul. Bovendien zijn de constantes (de nitraatconcentratie bij een Nmin van nul) gemiddeld rond de 10 mg/l hoger.

Bij alle drie de modellen waren er een zestal datapunten met een sterke hefboomwerking. De vijf proefplekken met een veenlaagje maakten onderdeel uit van deze zes punten (ondanks de opname van veen in het model). Het weglaten van deze punten geeft een model zonder veen en de helling voor de nitraatconcentratie versus $Nmin_{nitraat}$ wordt 0.75. Dit is iets steiler en meer conform de helling van grasland. De invloed op de overige kenmerken van het model was zeer gering. In de analyse over drie meetseizoenen zou overwogen kunnen worden de punten met veen bij de maïs niet meer mee te nemen.

Voor de neerslagsommen en overschotten in de zomer en de winter wordt een mogelijk niet-lineair effect verwacht. Onderzocht is of het opnemen van een kwadratische term een verbetering van de modellen oplevert. Dit is niet het geval.

$DOC_{grondwater}$ (opgelost organisch koolstof in het grondwater (mg/l)) heeft wel een significante en relevante bijdrage in de verklaring van de variatie van de nitraatconcentratie. $DOC_{grondwater}$ opgenomen in Model 1 geeft een percentage verklaarde variantie van 33.0 % met sd=49.8. Dit model is gebaseerd op 99 waarnemingen en laat een duidelijk verbetering zien ten opzicht van Model 1. Echter $DOC_{grondwater}$ is een vergelijkbaar moeilijke en dure meting als de nitraatmeting zelf en daarmee geen potentiële indicator. Ter informatie is het model met $DOC_{grondwater}$ gegeven in Bijlage 4b.

Een mogelijk effect van groenbemester bij maïs kon niet goed worden onderzocht. Voor minder dan de helft van de proefplekken is bekend of er al dan niet een groebemester heeft gestaan. Voor de beschikbare data is het effect van groenbemester als toevoeging aan model 1 klein en niet significant.

De interactie tussen Gtgroep en bodemgroep is niet significant (p= 0.56 in Model 1). Vervolgens is voor alle verklarende variabelen in de dataset gecontroleerd of zij een significante interactie met bodemgroep dan wel Gtgroep hebben. De interactie tussen $Nmin_{nitraat}$ en Gtgroep is niet significant (p=0.37). Ook de interactie tussen $Nmin_{nitraat}$ en bodemgroep is niet significant (p=0.73). Andere wel significante interacties met bodem- of Gtgroep zijn nader onderzocht maar leveren geen bruikbare modellen.

4.3 Resultaten van Neuraal netwerk

De percentages verklaarde variantie die met een neuraal net werden behaald, zijn vergelijkbaar met die bij lineaire regressie. Dat is een aanwijzing dat niet meer verband is tussen de nitraatconcentratie en alle mogelijke verklarende variabelen dan dat er met lineaire regressie al wordt gevonden.

Het aantal geselecteerde principale componenten van de verklarende variabelen bedraagt iets meer dan 20. Deze variabelen zijn ongecorreleerd, en ze zijn in staat bijna evenveel variantie te verklaren als de ongeveer 35 oorspronkelijke variabelen. Het aantal effectieve parameters bij gebruik van pca-scores blijkt maximaal 24. Ook dit lijkt er op te duiden dat er sprake kan zijn van enkele interacties, maar zeker niet van veel. Deze bevinding zou kunnen worden gezien als een tweede argument voor de conclusie dat het onwaarschijnlijk lijkt dat er bij het fitten van een lineair regressiemodel belangrijke interacties over het hoofd zijn gezien.

De analyse van de dataset over twee meetseizoenen met behulp van een neuraal netwerk heeft aangetoond dat met regressie-analyse de belangrijkste effecten worden gevonden en er geen grote effecten worden gemist.

5 Opschaling van proefplek naar bedrijf

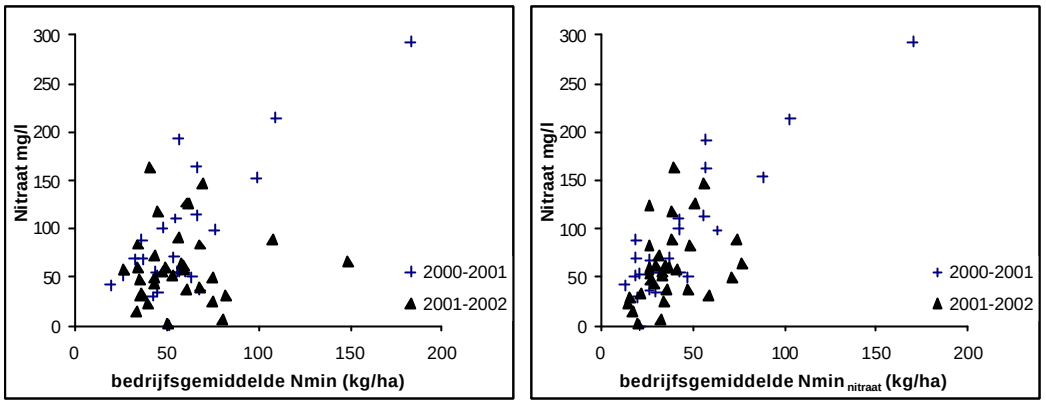
5.1 Inleiding

Bij het opzetten van het monitoringsysteem voor het project Sturen op Nitraat is er bewust voor gekozen om de te ontwikkelen regressie-modellen te baseren op het schaalniveau van proefplekken, en dus niet op dat van bijvoorbeeld monsters, percelen, bedrijven, clusters (bodem-Gt-gewascombinaties) of gebieden. Dat heeft twee belangrijke consequenties:

- Ruimtelijke variaties op het lagere schaalniveau, d.w.z. monsters binnen proefplekken, worden niet door de modellen voorspeld en komen als pseudo-meetfout terecht in de foutentermen van de modellen.
- De modellen zullen worden gebruikt om gemiddelde nitraatconcentraties te voorspellen op hogere schaalniveaus, met name op bedrijfs- en op gebiedsniveau. Dat kan wel, maar daartoe zullen de modellen moeten worden opgeschaald.

Een volledige beschrijving over het waarom en hoe van de opschaling, speciaal die van proefplek naar bedrijf, is te vinden in een eerdere rapportage op basis van één seizoen (Hack-ten Broeke et al.,2003). In dit hoofdstuk wordt een samenvatting gegeven en worden de resultaten van de opschaling na twee meetjaren gegeven. Om te komen tot een opschaling van proefplek naar bedrijf is het nodig eerst nader te bekijken hoe de bedrijfsgemiddelde nitraatconcentraties voorspeld zullen gaan worden.

De bedrijfsgemiddelde nitraatconcentratie per meetjaar kan berekend worden op basis van de waarnemingen waarbij rekening wordt gehouden met de clusteroppervlakten. Hetzelfde kan worden gedaan voor de bedrijfsgemiddelde Nmin en $Nmin_{nitraat}$. In onderstaande figuren zijn de bedrijfsgemiddelde Nmin en $Nmin_{nitraat}$ uitgezet tegen de bedrijfsgemiddelde nitraatconcentratie per meetjaar. Hieruit blijkt dat het verband tussen $Nmin_{nitraat}$ en de nitraatconcentratie duidelijk beter is dan het verband tussen Nmin en nitraatconcentratie. Naarmate de waarden hoger worden wordt het verband steeds duidelijker.



Figuur 4a en 4b. Berekende bedrijfsgemiddelde nitraatconcentratie als functie van de berekende bedrijfsgemiddele Nmin (4a) en $Nmin_{nitraat}$ (4b) per meetjaar

5.2 Voorspelling van de bedrijfsgemiddelde nitraatconcentratie met nauwkeurigheid

De resultaten van de analyse over de gegevens van twee meetjaren zijn lineaire modellen voor de akkerbouwgewassen, gras en maïs apart. De modellen hebben met elkaar gemeen dat ze verschillende intercepten hebben voor de Gt-groepen 1, 2 en 3 en voor de bodemgroepen Z1, Z2, Z3 en löss. Ook het al dan niet aanwezig zijn van een veenlaagje is van invloed op het intercept maar is bij de opschaling naar bedrijfsnivo verder buiten beschouwing gelaten. Voor akkerbouw is er ook nog een verschillend intercept voor de twee gewasgroepen (resp. 'hoog'en 'midden+laag'). De modellen verschillen wat betreft de kwantitatieve predictoren die erin zijn opgenomen.

Omdat we niet alleen verschillende modellen voor akkerbouw, gras en maïs hebben, maar binnen die modellen ook verschillende intercepten per Gt- en bodemgroep, kan het modellarium m.b.t. voorspelling óók beschouwd worden als te zijn opgebouwd uit 3x3x4=36 verschillende modellen (één voor elk cluster, d.w.z. gewas-Gt-bodem combinatie), van de volgende vorm:

$$\hat{y}_{ij} = \hat{a}_j + \hat{b}_{1j} x_{1ij} + \hat{b}_{2j} x_{2ij} \qquad [1]$$

waarin:

$\hat{y}_{ij}$: de voorspelde nitraatconcentratie op plek *i* binnen cluster *j* van het betreffende bedrijf;

$\hat{a}_j$: intercept van het model voor cluster *j*, bepaald door de gewas-Gt-bodemgroep;

$\hat{b}_{1j}, \hat{b}_{2j}$: de geschatte regressiecoefficienten van de kwantitatieve predictoren in het model voor cluster *j*;

x_{1ij}, x_{2ij} : de waarden van die kwantitatieve predictoren op de betreffende plek. Hierbij wordt opgemerkt dat welke predictoren dit betreft, wordt bepaald door het cluster, en dat de predictor x_2 kan ontbreken in het model.

In dit hoofdstuk beperken we ons voor de akkerbouw tot model 1 (zie hoofdstuk 4), het model met de gewas- Gt- en bodemgroep, $Nmin_{nitraat}$ en de neerslagsom in de zomer. Voor gras en maïs beperken we ons tot het model met Gt- en bodemgroep en $Nmin_{nitraat}$ (model 1). In al deze modellen is de helling voor $Nmin_{nitraat}$ (en voor neerslagsom) gelijk voor de Gt- en bodemgroepen en bij akkerbouw ook gelijk voor de twee gewasgroepen.

In het vervolg is er vanuit gegaan dat voor alle predictoren cluster-gemiddelden als invoer worden gebruikt, bepaald aan de hand van bedrijfsgegevens of via analyse van mengmonsters. Dit betekent dat voor het schatten van clustergemiddelden binnen het bedrijf modellen van het volgende type worden gebruikt:

$$\hat{\bar{y}}_j = \hat{a}_j + \hat{b}_{1j} \cdot \hat{\bar{x}}_{1j} + \hat{b}_{2j} \cdot \hat{\bar{x}}_{2j} \qquad [2]$$

waarin streepje en dakje boven de variabelen aangeven dat het gaat om een geschat clustergemiddelde. Het bedrijfsgemiddelde kan dan worden geschat als het gewogen gemiddelde van de clustergemiddelden:

$$\hat{\bar{y}} = \sum_{j=1}^{J} a_j \hat{\bar{y}}_j \qquad [3]$$

waarin:

a_j : oppervlaktefractie van cluster *j* binnen het bedrijf;

j : aantal clusters binnen het bedrijf.

Omdat de helling voor $Nmin_{nitraat}$ in het model gelijk is voor de verschillende Gt- en bodemgroepen (en bij akkerbouw ook voor de twee gewasgroepen), hoeft alleen het intercept gewogen te worden met de oppervlaktefractie van het cluster. Zo wordt de bedrijfsgemiddelde nitraatconcentratie bij akkerbouwbedrijven voorspeld met :

$$\hat{\bar{y}} = \sum_{j=1}^{3} A_j \hat{a}_j + \hat{b}_1 \hat{\bar{x}}_1 + \hat{b}_2 \hat{\bar{x}}_2 \qquad [4]$$

waarin:

A_j : relatief oppervlak van Gt-bodem-gwasgroep *j* in het bedrijf;

$\hat{a}_j$: geschat intercept voor Gt-bodem-gewasgroep *j* in het akkerbouwmodel;

$\hat{b}_1$: geschatte regressie-coefficient voor $Nmin_{nitraat}$;

$\hat{b}_2$: geschatte regressie-coefficient voor neerslagsom in de zomer;

$\hat{\bar{x}}_1$: geschat bedrijfsgemiddelde voor $Nmin_{nitraat}$;

$\hat{\bar{x}}_2$: aangegeven waarde voor neerslagsom.

In de huidige modellen is het dus niet noodzakelijk om per Gt-bodem-groep een schatting voor $Nmin_{nitraat}$ te hebben om te komen tot een goede voorspelling van de bedrijfsgemiddelde nitraatconcentratie. Een goede schatting van de bedrijfsgemiddelde $Nmin_{nitraat}$ is bij de akkerbouwbedrijven voldoende. Bij de veehouderijbedrijven moet wel per areaal gras en maïs apart een schatting van $Nmin_{nitraat}$ worden bepaald.

De voorspelling van de bedrijfsgemiddelde nitraatconcentratie is het oppervlaktegewogen gemiddelde van de geschatte clustergemiddelden in het bedrijf. De nauwkeurigheid van de voorspelling voor het bedrijf wordt dus bepaald door de

nauwkeurigheid van de geschatte oppervlakten en van de geschatte gemiddelden van de clusters. Als wordt aangenomen dat precies bekend is waar welke gewassen op het bedrijf worden geteeld, dan wordt de nauwkeurigheid van de clusteroppervlakten bepaald door de nauwkeurigheid van de Gt-kaart en bodemkaart van het bedrijf. De nauwkeurigheid van de clustergemiddelden van de kwantitatieve predictoren wordt bepaald door de nauwkeurigheid van bedrijfsgegevens en bemonstering.

Eerst wordt nagegaan hoe de nauwkeurigheid van de voorspellingen van de clustergemiddelde nitraatconcentraties gekwantificeerd kan worden. Vervolgens wordt ingegaan op de nauwkeurigheid van de bedrijfsgemiddelden.

Als maatstaf voor de nauwkeurigheid gebruiken we de Mean Squared Error (MSE, gemiddelde gekwadrateerde fout), welke gelijk is aan het kwadraat van de 'bias' (de systematische fout) plus de variantie (maat voor de toevallige fout):

$$MSE(\hat{\bar{y}}_j) = \left[bias(\hat{\bar{y}}_j)\right]^2 + Var(\hat{\bar{y}}_j) \qquad [5]$$

Deze vergelijking kan stapsgewijs worden uitgewerkt (zie Hack-ten Broeke et al. (2003)). Voor de modellen van gras en maïs geldt dat er maar één kwantitatieve predictor ($Nmin_{nitraat}$) is en dat de regressiecoëfficiënt van deze predictor in alle clusters gelijk is. In die eenvoudige situatie geldt dat de totale fout te schatten is m.b.v.:

$$E_m[MSE_b(\hat{\bar{y}}_j)] \cong Var_m(\hat{a}_j) + Var_m(\hat{b}) \cdot [\hat{\bar{x}}_j^{\,2} - Var_b(\hat{\bar{x}}_j)] + 2Cov_m(\hat{a}_j, \hat{b}) \cdot \hat{\bar{x}}_j + Var_b(\hat{\bar{x}}_j) \cdot \hat{b}^2 + E_m(\bar{e}_j^{\,2}) \qquad [6]$$

In formule [6] is alles bekend, behalve de laatste term: de verwachting van het kwadraat van het ruimtelijk gemiddelde van de model-residuen (als deze residuen binnen het cluster op het bedrijf steeds tegen elkaar zouden wegvallen, dan zou deze term nul zijn.) Voor de ontwikkelbedrijven kan deze residu-term worden geschat door de residuen van de proefplekken per meetjaar te middelen binnen de strata (clusters en bedrijven) en de bijbehorende variantie van dat gemiddelde residu uit te rekenen. Vervolgens wordt per bedrijf en per meetjaar een oppervlakte-gewogen gemiddelde over de clusters berekend van dit residu-gemiddelde en de-s variantie. Per bedrijf per meetjaar wordt nu de restterm berekend door het gewogen gemiddelde te kwadrateren en de variantie ervan af te trekken. Tenslotte worden al deze jaarlijkse resttermen per bedrijf weer gemiddeld (gewogen op basis van de relatieve oppervlakte van de bedrijven). Dit gemiddelde is gebruikt als de beste schatting voor de term $E_m(\bar{e}_j^{\,2})$.

Vervolgens wordt de nauwkeurigheid van voorspellingen van bedrijfsgemiddelde nitraatconcentraties berekend. Als de fouten in de voorspellingen van de cluster-gemiddelden van een bedrijf ongecorreleerd zijn, dan is formule [6] eenvoudig op te schalen naar bedrijfsniveau:

$$E_m[MSE_b(\hat{\bar{y}}_j)] \cong \sum_{j=1}^{J} a_j^{\,2} E_m[MSE_b(\hat{\bar{y}}_j)], \qquad [7]$$

d.w.z. de gewogen som van de fouten per cluster, met de kwadraten van de relatieve oppervlakten van de clusters als gewichten.

5.3 Berekening van de voorspelling en de nauwkeurigheid van een cluster- en een bedrijfsgemiddelde nitraatconcentratie

Vergelijking [6] geeft de totale fout van de voorspelling van de clustergemiddelde nitraatconcentraties. Vergelijking [6] bestaat in feite uit drie onderdelen :

- termen die betrekking hebben op de onnauwkeurigheid van de regressielijn.
- termen die betrekking hebben op de onnauwkeurigheid van $\text{Nmin}_{\text{nitraat}}$ (en mogelijke andere verklarende variabelen in het regressie model). Ook de waarde van $\text{Nmin}_{\text{nittraat}}$ zelf speelt hierin een rol. Door de $\text{Nmin}_{\text{nitraat}}$ anders te bepalen verandert de bijdrage van deze term aan de totale onnauwkeurigheid.
- de naar bedrijfsniveau opgeschaalde restspreiding $E_m(\overline{e}_j^{\,2})$

Deze drie onderdelen kunnen apart worden uitgerekend. Daarmee wordt inzichtelijk gemaakt hoe de totale onnauwkeurigheid is opgebouwd.

De variantie van $\text{Nmin}_{\text{nitraat}}$ is geschat op 1000. Deze schatting is gebaseerd op de $\text{Nmin}_{\text{nitraat}}$ gegevens in de dataset. Voor iedere combinatie van bedrijf, gewasgroep (gras, maïs, hoog en midden+laag) en seizoen is de variantie van $\text{Nmin}_{\text{nitraat}}$ tussen de proefplekken berekend en vervolgens gewogen gemiddeld. Bij de voorspelling is er vanuit gegaan dat er op basis van 40 steken op een bedrijf een mengmonster wordt gemaakt waaraan de $\text{Nmin}_{\text{nitraat}}$ wordt bepaald.

5.3.1 Akkerbouw

Voor model 1, het regressiemodel met gewasgroep, bodemgroep, Gtgroep, $\text{Nmin}_{\text{nitraat}}$ en neerslagsom in zomer (en veen), worden de drie termen berekend die gezamenlijk de nauwkeurigheid van een clustergemiddelde bepalen. Er is uitgegaan van een gemiddelde neerslagsom in de zomer van 503 mm (gemiddelde van de twee meetseizoenen). De fout rond de gemiddelde neerslag is nul verondersteld. De voorspellingen worden gegeven voor $\text{Nmin}_{\text{nitraat}}$-waarde van 30 kg/ha.

De bijdrage van de onnauwkeurigheid van de regressielijn aan de totale variantie van de voorspelling van een clustergemiddelde varieert van 35 tot 335. Deze bijdrage hangt onder andere af van de nauwkeurigheid van de schatting van het intercept en varieert dus voor de verschillende clusters. De bijdrage van $\text{Nmin}_{\text{nitraat}}$ aan de totale variantie komt op 27.7. De restspreiding op bedrijfsniveau is geschat op 300.

In de onderstaande tabel wordt per gewasgroep, bodem en Gt-combinatie de voorspelling van het clustergemiddelde met diens nauwkeurigheid (se) gegeven.

Tabel 16. De voorspelde nitraatconcentratie per cluster(bodem-Gt-gewascombinatie) met de bijbehorende se op basis van 40 steken voor de $Nmin_{nitraat}$ bepaling bij een gestelde $Nmin_{nitraat}$ =30 en een neerslagsom=503 (op basis van het model 1, hoofdstuk 4)

bodem	gewas Gt	hoog voorspelling	se	laag voorspelling	se
L	1	27.0	25.8	0.0	24.2
	2	68.8	24.1	40.4	22.4
	3	88.0	23.1	59.6	20.9
Z1	1	39.3	24.6	10.9	24.1
	2	81.1	22.4	52.7	21.8
	3	100.3	22.0	71.9	21.1
Z2	1	55.9	22.6	27.5	21.5
	2	97.7	21.4	69.3	20.1
	3	117.0	20.8	88.6	19.2
Z3	1	49.2	23.1	20.8	22.5
	2	90.9	20.7	62.6	19.9
	3	110.2	20.2	81.8	19.1

Aangezien in de dataset Gtgroep 3 beter vertegenwoordigd is als de andere twee Gtgroepen , zijn de schattingen voor deze Gtgroep nauwkeuriger (se is kleiner).

5.3.2 Veehouderij

Voor gras en maïs is uitgegaan van model 1, het regressiemodel met bodemgroep, Gtgroep en $Nmin_{nitraat}$ (en veen), en worden de drie termen berekend die gezamelijk de nauwkeurigheid van een clustergemiddelde bepalen. De voorspellingen worden gegeven voor $Nmin_{nitraat}$-waarde van 30 kg/ha.

De bijdrage van de onnauwkeurigheid van de regressielijn aan de totale variantie van de voorspelling van een clustergemiddelde varieert van 34 tot 240 voor gras en van 128 tot 492 voor maïs. Deze bijdrage hangt onder andere af van de nauwkeurigheid van de schatting van het intercept en varieert dus voor de verschillende clusters. De bijdrage van $Nmin_{nitraat}$ aan de totale variantie komt voor gras op 14.3 en voor maïs op 10.3. De restspreiding op bedrijfsniveau is nog steeds geschat op 300.

In de onderstaande tabel wordt per bodem-Gt-combinatie de voorspelling van het clustergemiddelde met diens nauwkeurigheid (se) gegeven.

Tabel 17. De voorspelde nitraatconcentratie per cluster met de bijbehorende se op basis van 40 steken voor de $Nmin_{nitraat}$ bepaling bij een gestelde $Nmin_{nitraat}$ =30 (op basis van het model 1 voor resp. gras en maïs, hoofdstuk 4)

bodem	gewas Gt	gras voorspelling	se	maïs voorspelling	se
L	1	*	*	*	*
	2	*	*	*	*
	3	25.9	23.5	*	*
Z1	1	42.4	20.8	17.1	28.3
	2	49.9	19.4	28.8	26.8
	3	55.3	18.9	88.4	24.7
Z2	1	40.8	19.8	29.9	23.2
	2	48.3	18.6	41.5	23.6
	3	53.7	19.2	101.1	25.3
Z3	1	45.4	20.1	36.3	25.3
	2	52.9	18.5	47.9	20.9
	3	58.3	18.7	107.5	21.3

Om vervolgens tot bedrijfsgemiddelde voorspellingen te komen, kan gebruik worden gemaakt van de clustergegevens in de tabellen 10 t/m 15. Hieronder volgen enkele rekenvoorbeelden.

Rekenvoorbeeld 1

Een bedrijf teelt alleen gras en het bedrijfsareaal valt geheel binnen bodemgroep Z3 en Gt-groep 2. Stel dat $Nmin_{nitraat}$ wordt gemeten op basis van één mengmonster van 40 steken en de gemeten $Nmin_{nitraat}$-waarde is 30. De voorspelde nitraatwaarde is dan gelijk aan 52.9 mg/l NO_3 met een se=18.5 (zie Tabel 17).

Dit betekent dat we met 95% betrouwbaarheid kunnen zeggen dat de werkelijke bedrijfsgemiddelde nitraatconcentratie in het grondwater in het voorjaar voor zo'n bedrijf zal liggen tussen

$$NO_3 = 52.9 \pm 2 * 18.4 \text{ mg/l}$$

Rekenvoorbeeld 2

Een bedrijf teelt op de ene helft van zijn areaal gras en op de andere helft maïs. Het gehele bedrijf heeft bodemgroep Z3. De graspercelen vallen allemaal binnen Gt-groep 3 en de maïspercelen voor de helft in Gt-groep 1 en voor de andere helft in Gt-groep 2. Stel er wordt $Nmin_{nitraat}$ gemeten op basis van één mengmonster (van 40 steken) per cluster (voor de drie clusters: gras/Gtgroep-3/bodemgroep-Z3, maïs/Gtgroep-1/bodemgroep-Z3 en maïs/Gtgroep-2/bodemgroep-Z3). De $Nmin_{nitraat}$-waarde is voor dit rekenvoorbeeld overal 30.

De voorspelde bedrijfsgemiddelde nitraatwaarde in het grondwater in het voorjaar (zie tabel 17) is nu :

$$NO_3 = 0.5 * 58.3 + 0.25 * 36.3 + 0.25 * 47.9 = 50.2 \text{ mg/l}$$

De bijbehorende standaardafwijking se (zie tabel 24) wordt als volgt berekend :

$$se = \sqrt{\left(0.5^2 * 18.7^2 + 0.25^2 * 25.3^2 + 0.25^2 * 20.9^2\right)} = 12.4$$

Dit betekent dat we met 95% betrouwbaarheid kunnen zeggen dat de werkelijke bedrijfsgemiddelde nitraatconcentratie in het grondwater in het voorjaar voor het bedrijf van dit tweede rekenvoorbeeld zal liggen tussen

$$NO_3 = 50.2 \pm 2 * 12.4 \text{ mg/l}$$

Hieruit blijkt dat de variantie omlaag gaat (en de betrouwbaarheid omhoog) naarmate een bedrijf uit meerdere clusters bestaat. Op het eerste gezicht lijkt dit merkwaardig. Dit wordt veroorzaakt door twee componenten. In de bovenstaande voorbeelden is uitgegaan van een mengmonster van 40 steken per cluster. Dit betekent dat er in het eerste voorbeeld 40 steken en in het tweede voorbeeld 120 steken zijn genomen. Het is logisch dat daarmee de nauwkeurigheid van de voorspelling omhoog gaat. Een ander element dat een rol speelt is het feit dat een bedrijfsgemiddelde, berekend op basis van verschillende (min of meer onafhankelijke) clustergemiddelden, nauwkeuriger is dan een bedrijfsgemiddelde op basis van één clustergemiddelde.

6 Conclusies

Na analyse van de gegevens van twee meetseizoenen (2000/2001 en 2001/2002) voor het project Sturen Op Nitraat kunnen alleen voorlopige conclusies worden getrokken. Pas als alle gegevens van drie meetseizoenen verwerkt zijn, zullen de conclusies definitief zijn.

Deze rapportage heeft alleen betrekking op gegevens die zijn verzameld op de zogenaamde ontwikkelbedrijven op in totaal 478 proefplekken binnen een gestratificeerde steekproef (zie Smit et al., 2003). De voorlopige regressie-vergelijkingen die zijn opgesteld en weergegeven in hoofdstuk 4 van dit rapport zijn gebruikt in het deelproject 'regionaal monitoringsconcept' (Roelsma et al., in prep.)

Het doel van de uitgevoerde regressie-analyse is het identificeren van een indicator die geschikt is om op bedrijfsniveau de nitraatuitspoeling te voorspellen. Daartoe zijn met name kandidaatindicatoren Nmin (gemeten in het najaar), perceelsoverschot en bedrijfsoverschot (MINAS en werkelijk berekend) in ogenschouw genomen als mogelijk verklarende variabelen voor de nitraatconcentratie in het voorjaar. Voor Nmin is tevens gekeken naar het nitraatdeel van Nmin ($Nmin_{nitraat}$) en het ammoniumdeel ($Nmin_{ammonium}$). Voorafgaand aan de gegevensverzameling is verondersteld dat een aantal factoren van groot belang zijn voor de nitraatconcentratie, namelijk plaatsgebonden factoren zoals bodem en grondwatertrap, het weer (met name neerslag) en het gewas.

6.1 Regressie-analyse met proefplekgegevens

Van de kandidaat-indicatoren heeft de hoeveelheid Nmin de hoogste correlatie met nitraat in het voorjaar. Het beeld is echter niet erg consistent over de twee meetjaren, in het tweede meetjaar zijn de correlaties aanzienlijk lager dan in het eerste meetjaar. Bij de regressie-analyse zijn een groot aantal aanvullende variabelen meegenomen, waarvoor de veronderstelling geldt dat allen een zekere relatie met nitraatuitspoeling vertonen. Om er zeker van te zijn dat er geen belangrijke verbanden (niet-lineaire verbanden, interacties ect.) over het hoofd worden gezien is een neuraal net gefit op de data. De analyse over de twee meetseizoenen met behulp van een neuraal netwerk heeft aangetoond dat met regressie-analyse de belangrijkste effecten worden gevondenen er geen grote effecten onopgemerkt blijven.

De regressie-analyses zijn uitgevoerd voor akkerbouwgewassen, gras en maïs afzonderlijk, omdat er duidelijke verschillen waren tussen akkerbouw en melkveehouderij. In alle gevallen komt de indeling in drie Gt-groepen (respectievelijk met GHG ondieper dan 40 cm – mv., met GHG tussen 40 en 80 cm – mv. en met GHG dieper dan 80 cm – mv.) naar voren als een verklarende factor voor het niveauverschil in de gemeten nitraatconcentraties in het voorjaar van 2001 en 2002. Ook de indeling in bodemgroep (respectievelijk de lössgronden, zandgronden met

veel organische stof of een dikke bovengrond, zandgronden met relatief veel organisch stof en een hoog leemgehalte en de overige zandgronden) komt naar voren als een verklarende variabele, dit in tegenstelling tot de resultaten van de analyse na één meetseizoen. Tenslotte speelde het al dan niet aanwezig zijn van een veenlaagje een rol bij zowel de akkerbouwgewassen als gras en maïs. Het aanwezig zijn van een veenlaagje verlaagt de verwachte nitraatconcentratie.

Uit de verzamelde Nmin-gegevens komt de hoeveelheid $Nmin_{nitraat}$ in de meetperiode oktober-december, gesommeerd over de drie bodemlagen (i.e. 0-90 cm) als meest verklarende variabele naar voren uit de groep aan Nmin-gegevens. $Nmin_{nitraat}$ speelt in alle gevallen een belangrijke rol in de verklaring van de gemeten nitraatconcentratie. Het is gebleken dat de relatie tussen het nitraatdeel van Nmin met de nitraatconcentratie beter is dan de som van ammonium en nitraat. Ook uit het berekende bedrijfsgemiddelde van zowel de nitraatconcentratie als Nmin en $Nmin_{nitraat}$ blijkt dat het verband duidelijk beter is met $Nmin_{nitraat}$.

Bij akkerbouw is in de analyse een nieuwe gewasgroep-indeling meegenomen. Dit is een indeling op basis van de verwachte Nmin in het najaar waarbij uiteindelijk een groep 'hoog' en 'midden en laag' zijn onderscheiden. Deze nieuwe gewasgroep-indeling komt naar voren als een verklarende factor voor het niveauverschil in de gemeten nitraatconcentratie, dit in tegenstelling tot de oorspronkelijke gewasgroepindeling. Bij akkerbouw speelt ook de neerslagsom in de zomerperiode (1 april tot 1 oktober) in alle drie de modellen een rol.

Voor grasland zijn naast Gt-groep, bodemgroep, $Nmin_{nitraat}$ en veen (model 1) de volgende verklarende variabelen naar voren gekomen : C:N-verhouding, GHG en de potentiële mineralisatie die gezien moet worden als een maat voor afbreekbaarheid van organische stikstof (model 2). Voor maïs levert de regressie-analyse naast Gt-groep, bodemgroep, $Nmin_{nitraat}$ en veen (model 1) de volgende extra verklarende variabelen op: GHG en de potentiële mineralisatie (maat voor afbreekbaarheid organisch stikstof) (model 2) of GLG en de neerslagsom in de zomer (model 3).

Als de regressie-vergelijkingen worden gebruikt om nitraatconcentraties te voorspellen op puntniveau, zijn de betrouwbaarheidsintervallen nogal groot. Grofweg zijn deze intervallen ± 100 tot 135 mg/l. Dit betekent dat voorspellingen of puntniveau nauwelijks zinvol zijn. De bedoeling van het project is echter om voorspellingen te doen voor cluster- of bedrijfsniveau.

6.2 Opschaling naar bedrijfsniveau

De opschaling naar bedrijfsniveau vindt plaats via de zogenaamde clusterbenadering. Dat wil zeggen dat de ontwikkelde regressie-vergelijkingen worden toegepast voor bodem-Gt-gewascombinaties (clusters) en dat de voorspelling van de bedrijfsgemiddelde nitraatconcentratie niets anders is dan het oppervlaktegewogen gemiddelde van de aldus geschatte clustergemiddelden van een bedrijf. Van belang is daarbij dat vooral de voorspelfout van zo'n bedrijfsgemiddelde vele malen lager is

dan de voorspelfout op puntniveau. In een aantal rekenvoorbeelden komt naar voren dat de 95%-betrouwbaarheidsintervallen voor de voorspelde nitraatconcentratie dan nog slechts ± 25 tot 37 mg/l zijn. Als de regressie-vergelijkingen gebaseerd zijn op meerdere groeiseizoenen zal deze voorspelfout naar verwachting verder afnemen.

Literatuurlijst

Bishop, C. M., 1995. Neural networks for pattern recognition. Oxford University Press.

Enckevort P.L.A. van, J.R. van der Schoot en W. van den Berg, 2002. Relatie tussen N-overschot en N-uitspoeling, op gewasniveau voor de akkerbouw en vollegrondsgroente-teelt. Wageningen, Praktijkonderzoek Plant & Omgeving; PPO-rapport 1125234.

Hack-ten Broeke M.J.D., S.L.G.E. Burgers, H.F.M. ten Berge, P.L.A. van Enckevort, J.J. de Gruijter, I.E. Hoving Smit, A., en G.L. Velthof, 2003. Ontwikkeling van een indicator om te Sturen Op Nitraat; Gegevens en regressie-analyse voor het eerste meetseizoen (2000-2001). Wageningen, Alterra. Alterra-rapport 772. Reeks Sturen Op Nitraat 4.

Hagan, M.T., H.B.Demuth en M.Beale, 1996. Neural network design. Boston, PWS Publishing Company.

Haykin, S., 1999. Neural networks, a comprehensive foundation. New Jersey, Prentice Hall.

Noij, G.J., E. Hees, P. Dekker, R. Schils, J. Schröder en H. ten Berge, 2001. Onderzoeksvoorstel. Wageningen, Alterra. Reeks Sturen Op Nitraat 1.

Ripley, B. D., 1996. Pattern Recognition and Neural Networks. Cambridge University Press.

Roelsma, J., C. Rougoor en P. Dik, in prep., Regionaal Nitraatmonitoringsconcept. Wageningen, Alterra. Reeks Sturen Op Nitraat 7

Rumelhart, D.E., G.E. Hinton and R.J. Williams, 1986. Learning representations of back-propagation errors. Nature, London, vol.323, pp.533-536.

Smit, A., M.J.D. Hack-ten Broeke, H.F.M. ten Berge, S.L.G.E. Burgers, W. Chardon, P.L.A. van Enckevort, J.J. de Gruijter, I.E. Hoving en G.L. Velthof, 2003. Gegevensverzameling Sturen Op Nitraat; Op zoek naar een indicator. Wageningen, Alterra. Alterra-rapport 658. Reeks Sturen Op Nitraat 3.

Bijlage 1 Beschikbare gegevens van de overige gemeten variabelen

In de onderstaande tabel wordt een overzicht gegeven van alle variabelen die gemeten zijn in het tweede meetseizoen en die een rol hebben gespeeld in de analyse van de gegevens maar die geen kandidaat-indicator zijn.

Overzicht van de variabelen, gebruikt in de regressie-analyse, ingedeeld naar de akkerbouwgewassen (gewasgroep a+b+r+t ofwel laag+midden+hoog) en melkveehouderijgewassen (gewasgroep g+m)

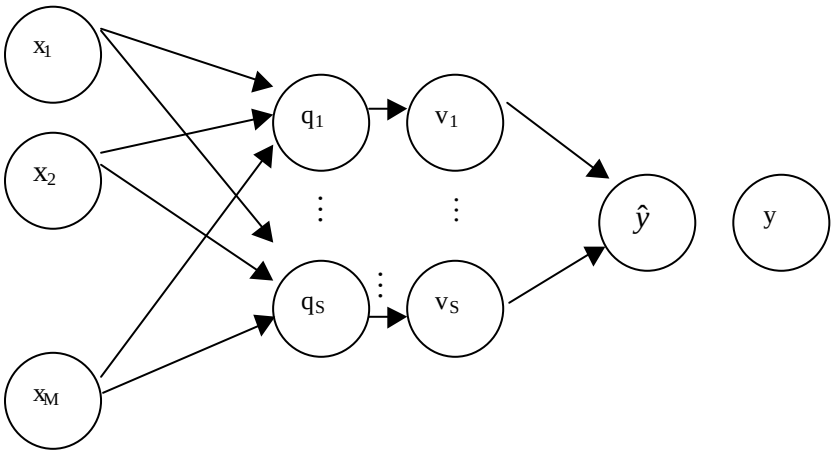
Variabele	*Aantal Waarnemingen*	*Gemiddelde*	*Minimum*	*Maximum*	*Mediaan*
Kunstmest (km-)gift (kg N/ha)					
Akkerbouw	194	61.6	0	310	49
Veeteelt	263	76.1	0	252	74
N-werkzaam uit dierlijke mest					
Akkerbouw (kg N/ha)	194	72.5	0	215	75
Veeteelt (idem)	263	138.1	0	345	126
N-totaal uit dierlijke mest (dm)					
Akkerbouw (kg N/ha)	194	109.7	0	245	128
Veeteelt (idem)	263	208.7	0	453	211
N-werkzaam uit km + dm					
Akkerbouw	194	134.1	0	398	148
Veeteelt	263	214.2	0	470	177
N-totaal uit km + dm					
Akkerbouw	194	171.3	0	440	196
Veeteelt	263	284.8	0	606	250
Dierweidedagen					
Akkerbouw	189	0.9	0	58	0
Veeteelt	247	402.6	0	6991	0
Afvoer N via gewas (kg N/ha)					
Akkerbouw	192	117.5	0	278	120
Veeteelt	250	214.1	0	843	195
Neerslagsom groeiseizoen					
Akkerbouw (mm)	194	531.6	398.6	632.4	513.5
Veeteelt	263	475.9	398.2	632.4	486.3
Neerslagsom winterseizoen					
Akkerbouw	194	485.2	430.7	567.4	477.3
Veeteelt	263	459.6	430.7	567.4	452.0
Neerslagoverschot groeiseizoen					
Akkerbouw	194	115.1	-81.7	273.6	106.3
Veeteelt	263	10.2	-86.6	220.8	24.3
Neerslagoverschot winter					
Akkerbouw	194	437.9	355.8	523.4	431.5
Veeteelt	263	372.9	322.6	523.4	370.4

De mestgift ligt voor de akkerbouwgewassen gemiddeld lager in het tweede meetseizoen. Voor de veehouderijgewassen is de gemiddelde mestgift gelijk. Het gemiddelde aantal dierweidedagen op grasland ligt in het tweede meetseizoen een stuk hoger. De gemiddelde N-afvoer met het gewas is voor de akkerbouwgewassen vrijwel gelijk gebleven terwijl dit voor gras en maïs is gedaald. Met betrekking tot de neerslag in het groeiseizoen is het in het tweede meetseizoen een stuk natter geweest. In de winterperiode is de neerslagsom en het overschot bij de akkerbouwbedrijven

gemiddeld hoger terwijl het overschot bij de veeteeltbedrijven een beetje lager is in het tweede meetjaar.

Bijlage 2 Artificiële neutrale netten

Er bestaan verschillende soorten neurale netten (Bishop, 1995; Ripley, 1996; Haykin, 1999), o.a. voor classificatie, voor clustering, voor feedbackprocessen. In dit onderzoek is gebruik gemaakt van het zogenaamde MeerLaags Perceptron (MLP) met Backpropagation en Bayesian Regularization (BR). Het MLP is in feite een niet-lineair regressiemodel. De term 'backpropagation' slaat op het algorithme waarmee de afgeleide van de errorfunctie wordt berekend (Rumelhart et al., 1986). De term Bayesian Regularization verwijst naar een manier om de grootte van de geschatte parameters in de hand te houden en daarmee overfitting te voorkomen (zie verderop). De berekeningen zijn uitgevoerd met behulp van MATLAB. Schematisch kan een MLP worden voorgesteld als in Figuur 1.



Figuur 1: Voorbeeld van een Meerlaags Perceptron met S verborgen neuronen

Dit neurale net heeft 3 lagen: links de invoer, rechts de uitvoer, en in het midden een zg. verborgen laag. De eerste of invoerlaag wordt gevormd door de data x_1 tot en met x_M. Er zijn dus M x-variabelen; en de kolom in laag 1 is in feite een rij uit de datamatrix. De tweede, verborgen, laag heeft, zoals in de figuur te zien is, een input (q_1 tot en met q_S) en een output (v_1 tot en met v_S). In een MLP bestaat de input van laag 2 uit een lineaire combinatie van de invoervariabelen x_1 t/m x_M:

$$q_s = \sum_{m=1}^{M} w_{sm} x_m \text{ , voor s=1…S.}$$

De uitvoer van laag 2 wordt gegeven door

$$v_s = \mathrm{f}(q_s),$$

waarin f een niet-lineaire functie is; vaak, en ook in dit onderzoek, wordt gekozen voor de hyperbolische tangensfunctie f=tanh:

$$v_s = \tanh(q_s) = \frac{\exp(q_s) - \exp(-q_s)}{\exp(q_s) + \exp(-q_s)}$$

Voor de derde laag geldt in het MLP dat die bestaat uit een lineaire combinatie van de uitvoer van laag 2, waarbij de gewichten in deze tweede laag worden aangeduid met het symbool p:

$$\hat{y} = \sum_{s=1}^{S} p_s v_{s.}$$

Met de formules voor v_s en q_s geeft dit:

$$\hat{y} = \sum_{s=1}^{S} p_s \tanh(q_s) = \sum_{s=1}^{S} p_s \frac{\exp(q_s) - \exp(-q_s)}{\exp(q_s) + \exp(-q_s)},$$

$$= \sum_{s=1}^{S} p_s \frac{\exp(\sum_{m=1}^{M} w_{sm} x_m) - \exp(-\sum_{m=1}^{M} w_{sm} x_m)}{\exp(\sum_{m=1}^{M} ws_{sm} x_m) + \exp(-\sum_{m=1}^{M} w_{sm} x_m)}.$$

Net als bij lineaire regressie kan aan de verklarende variabelen een intercept worden toegevoegd; deze heet in de neurale netten context de 'bias'.

De uitvoer wordt hier dus voorspeld met een niet-lineaire functie van de invoervariabelen X. De parameters in het model zijn de gewichten w en p in de eerste en tweede laag. Ook het aantal neuronen in de verborgen laag zou als parameter van het model kunnen worden gezien. Een hard criterium voor het 'beste' aantal neuronen is er niet. Als vuistregel wordt vaak gebruikt dat $S \approx \sqrt{M}$.

Het schatten van de parameters gebeurt in principe door voor alle invoeren de uitvoer $\hat{y}$ te vergelijken met de geobserveerde waarde van y; hieruit een kwadratensom $\sum_{i=1}^{N} (\hat{y}_i - y_i)^2$ te berekenen; en vervolgens deze errorfunctie te minimaliseren met betrekking tot alle parameters.

Neurale netten hebben in het algemeen veel parameters. Een lineair regressiemodel met 35 X-variabelen heeft 37 parameters (35 regressiecoefficienten, 1 intercept, en de residuele variantie); een neuraal net heeft er veel meer. Een net met S verborgen neuronen heeft S x (M+1) gewichten in de eerste laag, en S+1 in de tweede (de extra gewichten ontstaan door de biasterm). Samen zijn dat er SM+S+S+1=SM+2S+1. Bij 35 X-variabelen en 5 verborgen neuronen geeft dit 5(35)+2(5)+1=186; bij 10 verborgen neuronen zijn er 10(35)+20+1=371 parameters. Hiermee ontstaat het gevaar van overfitten: door het grote aantal parameters is het relatief eenvoudig om een zeer goed model op te stellen, maar bij validatie op nieuwe data kan dan blijken dat het model helemaal niet zo goed is als het leek.

Werkwijze
Er zijn afzonderlijke modellen gefit voor het eerste en het tweede meetjaar. Aangezien de bedoeling is een zo groot mogelijk percentage verklaarde variantie te bereiken, heeft geen speciale modelselectie plaatsgevonden, maar zijn eenvoudigweg alle beschikbare variabelen in het model opgenomen. Wel zijn sommige verklarende variabelen weggelaten in verband met het grote aantal ontbrekende waarnemingen.
Bij het indelen van de data in trainingset, stopset en testset is een verhouding 60:20:20 gehanteerd. De records zijn random aan deze drie categorieën toegewezen.

Bij het fitten van neurale netten is het handig als de verklarende variabelen zo'n beetje dezelfde range hebben. Alle variabelen zijn daarom gestandaardiseerd tot variabelen met gemiddelde 0 en standaardafwijking 1.

Onderzocht is of een logtransformatie van de afhankelijke variabele tot een hoger percentage verklaarde variantie leidde. Ook is geprobeerd het aantal inputs te beperken door eerst een principale componentenanalyse op de x-variabelen uit te voeren, en alleen die componenten te behouden die meer dan 1 % van de variantie voor hun rekening nemen. Deze inputs zijn dan ongecorreleerd.

Het aantal inputs M (inclusief dummies voor factoren) ligt zo rond de 30, daarom zijn modellen met S = 5 en met 10 verborgen neuronen onderzocht. Omdat de gevonden oplossing kan afhangen van de gekozen startwaarden (neurale netten hebben in het algemeen meerdere locale minima) zijn steeds 5 herstarts uitgevoerd.

Resultaten

Eerste meetjaar.
Van de beschikbare verklarende variabelen zijn een aantal variabelen opgenomen in het model, uit een aantal variabelen is een keuze gemaakt en enkele variabelen hebben niet meegedaan.

In model: Gtgroep, bodemgroep, veen, perceelsoverschot, Nafvoergewas, kunstmest, Nwerkzaam in dierlijke mest, Minas-bedrijfsoverschot, Neerslagsom1, Neerslagoverschot1, GHG, GLG, C:Nverhouding, Ctotaal, Ntotaal , Potmin, N_NH4, denit1 t/m denit5, decap, organisch_stof, Norg.

Keuze: gewasgroep of nieuwe gewasgroep-indeling, Nmin of $Nmin_{nitraat}$ en $Nmin_{ammonium}$.

Niet in model: Ntotaal in dierlijke mest, dierweidedagen, bedrijfsoverschot, grondwaterstand, $DOC_{grondwater}$, Neerslagsom2, Neerslagoverschot2, denit6

Er zijn 2 modellen gefit: Model 1a bevat gewasgroep en Nmin, en Model 1b bevat in plaats daarvan de nieuwe gewasgroepindeling, $Nmin_{nitraat}$ en $Nmin_{ammonium}$. De overige variabelen in beide modellen zijn gelijk. De aantallen inputs (inclusief dummies voor factorlevels) en overblijvende records in beide modellen staan in het onderstaande overzicht.

Model	Aantal inputs: x	Aantal inputs: pca	Aantal records (oorspr. 478)
1a: gewasgroep + Nmin	35	21	239
1b: gewasnieuw + $Nmin_{nitraat}$ + $Nmin_{ammonium}$	33	21	239

De volgende tabel bevat de percentages verklaarde variantie voor de Modellen 1 en 2. De resultaten bij de 5 herstarts lagen in het algemeen dicht bij elkaar; vermeld zijn hun gemiddelden. De aanduiding log=0 dan wel log=1 betekent dat er al (1) dan niet (0) een logtransformatie is uitgevoerd; en idem voor pca=0 of pca=1.

Tabel 1: Percentage verklaarde variantie bij de 2 modellen voor het eerste meetjaar:

R^2	*Neuraal net met 5 neuronen*				*Neuraal net met 10 neuronen*			
	Train	*Stop*	*Test*	*Eff. # pars*	*Train*	*Stop*	*Test*	*Eff. # pars*
Model 1a								
log=1, pca=0	35	15	10					
log=1, pca=1	38	13	11					
log=0, pca=0	45	31	20	30	46	30	22	28
log=0, pca=1	41	30	19	24	41	30	22	23
Model 1b								
log=0, pca=0	43	40	20	28	47	36	21	28
log=0, pca=1	44	39	19	23	44	38	21	20

Bij het fitten van 5 neuronen op Model 1a bleek de logtransformatie niet tot betere resultaten te leiden dan modelleren van de ongetransformeerde concentratie. Daarom is de logtransformatie bij het fitten van 10 neuronen en bij Model 1b verder buiten beschouwing gelaten. Het analyseren van pca-scores leidde bij Model 1a tot een iets lagere verklaarde variantie dan de analyse zonder pca, en bij Model 1b (gewasextra, somammop1 en somnitp1) was dat zelfs helemaal niet het geval. We kunnen dus concluderen dat er wel enige overlap zit in de verklarende variabelen. Model 1b geeft iets betere resultaten dan Model 1a, en 10 neuronen lijken het ook wat beter te doen dan 5. Maar de verschillen zijn allemaal niet dramatisch.
Het effectieve aantal parameters is laag. Hierboven was berekend dat het aantal parameters in een neuraal net met 35 inputs zo'n 370 bedraagt; hier zien we dat in principe met minder dan 30 parameters zou kunnen worden volstaan.

<u>Tweede meetjaar.</u>
Ook hier zijn twee modellen gefit: Model 2a met gewasgroep en Nmin, en Model 2b met de nieuwe gewasgroepindeling , $Nmin_{nitraat}$ en $Nmin_{ammonium}$. Het verschil met het eerste meetjaar is dat bedrijfsoverschot en de neerslaggegevens in de winter (Neerslagsom2 en Neerslagoverschot2) nu wel in het model zijn meegenomen.

De aantallen inputs en records staan in het onderstaande schema:

Model	Aantal inputs: x	Aantal inputs: pca	Aantal records (oorspr. 478)
2a: gewasgroep + Nmin	36	23	388
2b: gewasnieuw+ $Nmin_{nitraat}$ +$Nmin_{ammonium}$	34	22	391

Tabel 2 bevat de gevonden percentages verklaarde variantie in het tweede meetjaar:

Tabel 2: Percentage verklaarde variantie bij de 2 modellen voor het tweede meetjaar:

R^2	*Neuraal net met 5 neuronen*				*Neuraal net met 10 neuronen*			
	Train	*Stop*	*Test*	*Eff. # pars.*	*Train*	*Stop*	*Test*	*Eff . # pars.*
Model 2a								
log=0, pca=0	27	12	19	30	26	13	18	28
log=0, pca=1	24	11	17	23				
Model 2b								
log=0, pca=0	32	20	32	29	32	19	30	28
log=0, pca=1	31	20	32	23				

In grote lijnen zijn de resultaten vergelijkbaar met die van het eerste meetjaar: Model 2b lijkt iets beter dan Model 2a, het verschil tussen wel en geen pca is klein, en het verschil tussen 5 en 10 verborgen neuronen ook. De aantallen effectieve parameters zijn eveneens vergelijkbaar met die in het eerste jaar.

Wat wel opvalt, is dat de verklaarde varianties dit jaar lager lijken dan bij het eerste meetjaar. Om te onderzoeken of er misschien toevallig een onhandige verdeling van de data in training-, stop- en testset is gekozen, zijn de analyses voor het tweede meetjaar nog herhaald met verschillende andere partities van de data. De resultaten bleven vrijwel gelijk.

Bijlage 3a Modellen akkerbouw

Model 1 en 3 voor **akkerbouw** gefit op de data per meetseizoen:

Tabel 1 Schatting van de regressie-coëffeciënten a t/m f met standaardfout (se), het percentage verklaarde variantie (R^2_{adj}) en de standaardfout (sd) van model 1 en 3, akkerbouw per meetseizoen..

	Model 1, seizoen 1		*Model 1 seizoen 2*		*Model 3 seizoen 1*		*Model 3 seizoen 2*	
	Schatting	*se*	*Schatting*	*se*	*Schatting*	*se*	*Schatting*	*se*
a	1.11	0.19	0.73	0.18	1.12	0.20	0.62	0.18
b	-0.460	0.283	-0.320	0.075	-0.805	0.476	-0.992	0.230
d	15.1	24.4	-23.3	14.6	33.1	25.0	-27.3	14.6
e					-1.98	1.78	0.66	1.50
f					0.169	0.388	0.558	0.165
g					3.43	2.35	1.49	1.03
h					-0.0045	0.0034	-0.0016	0.0011
R^2_{adj}	53.3		37.1		55.7		40.1	
Sd	63.2		48.8		61.5		47.7	

Tabel 2. Schatting van de constante C_i (in mg/l) per bodem-Gt-combinatie van model 1 en 3 voor model op basis van eerste meetseizoen, akkerbouw.

Gt-groep		*Model 1*			*Model 3*		
Bodem	*gewas*	*1*	*2*	*3*	*1*	*2*	*3*
Löss	hoog	262.1	302.1	328.3	-146.0	-131.6	-99.8
Z1	hoog	200.6	240.6	266.8	-256.3	-241.9	-210.1
Z2	hoog	242.2	282.2	308.4	-218.6	-204.2	-172.3
Z3	hoog	239.3	279.2	305.4	-205.9	-191.5	-159.6
Löss	mid+laag	218.0	258.0	284.2	-174.6	-160.2	-128.4
Z1	mid+laag	156.5	196.5	222.7	-284.9	-270.5	-238.7
Z2	mid+laag	198.1	238.1	264.3	-247.1	-232.8	-200.9
Z3	mid+laag	195.2	235.1	261.3	-234.4	-220.1	-188.2

Tabel 3. Schatting van de constante C_i (in mg/l) per bodem-Gt-combinatie van model 1 en 3 voor model op basis van tweede meetseizoen, akkerbouw.

Gt-groep		*Model1*			*Model 3*		
Bodem	*gewas*	*1*	*2*	*3*	*1*	*2*	*3*
Löss	hoog	130.7	172.9	198.9	96.7	133.5	168.9
Z1	hoog	185.6	227.8	253.8	136.1	173.0	208.3
Z2	hoog	195.2	237.4	263.4	137.5	174.3	209.7
Z3	hoog	182.3	224.5	250.5	120.8	157.6	193.0
Löss	mid+laag	109.9	152.1	178.0	81.6	118.4	153.8
Z1	mid+laag	164.8	207.0	233.0	121.0	157.9	193.2
Z2	mid+laag	174.4	216.6	242.6	122.4	159.2	194.6
Z3	mid+laag	161.5	203.7	229.7	105.7	142.6	177.9

Meetseizoen 1 = 2000-2001, meetseizoen 2 = 2001-2002

Bijlage 3b Modellen gras en maïs

Model 1 en 2 voor **gras en maïs** gefit op de data per meetseizoen :

Tabel 1. Schatting van de regressie-coëffeciënten a t/m f met standaardfout (se), het percentage verklaarde variantie (R^2_{adj}) en de standaardfout (sd) van model 1 en 2, grasland per meetseizoen..

	Model 1, seizoen 1		*Model 1 seizoen 2*		*Model 2 seizoen 1*		*Model 2 seizoen 2*	
	Schatting	*se*	*Schatting*	*se*	*Schatting*	*se*	*Schatting*	*se*
a	0.76	0.13	0.81	0.16	0.79	0.15	0.75	0.18
b	-51.0	25.3	-32.3	18.6	-20.3	28.0	-16.6	20.5
d					-2.45	2.33	-3.21	1.75
e					0.61	0.23	0.18	0.20
f					-6.3	5.9	-11.7	3.9
R^2_{adj}	23.1 %		15.8 %		28.8 %		18.4 %	
Sd	53.6		47.2		52.6		46.9	

Tabel 2. Schatting van de constante C_i (in mg/l) per bodem-Gt-combinatie van model 1 en 2, grasland.

		Model1			*Model 2*		
Gt-groep		*1*	*2*	*3*	*1*	*2*	*3*
Bodem	*seizoen*						
Löss	1	*	*	15.2	*	*	-46.3
Z1		9.0	18.3	47.1	39.7	29.2	19.5
Z2		13.7	23.1	51.9	57.8	47.3	37.6
Z3		8.3	17.6	46.4	49.3	38.9	29.1
Löss	2	*	*	-10.2	*	*	33.1
Z1		25.9	31.1	15.9	100.5	99.6	76.8
Z2		20.9	26.1	10.9	103.1	102.2	79.4
Z3		33.3	38.5	23.3	111.5	110.6	87.8

Tabel 3. Schatting van de regressie-coëffeciënten a t/m f met standaardfout (se), het percentage verklaarde variantie (R^2_{adj}) en de standaardfout (sd) van model 1 en 2, snijmaïs per meetseizoen..

	Model 1, seizoen 1		*Model 1 seizoen 2*		*Model 2 seizoen 1*		*Model 2 seizoen 2*	
	Schatting	*se*	*Schatting*	*se*	*Schatting*	*se*	*Schatting*	*se*
a	1.36	0.45	0.28	0.32	1.34	0.43	0.30	0.33
b	33.5	72.6	-27.0	35.6	47.2	64.3	-1.9	45.0
e					0.01	0.45	1.23	0.53
f					4.2	20.6	-16.3	14.4
R^2_{adj}	20.2		15.4		25.3		21.5	
Sd	67.5		63.2		55.9		62.3	

Tabel 4. Schatting van de constante C_i (in mg/l) per bodem-Gt-combinatie van model 1 en 2, snijmaïs.

		Model1			*Model 2*		
Gt-groep		*1*	*2*	*3*	*1*	*2*	*3*
Bodem	*seizoen*						
Z1	1	0.1	-32.6	21.3	-9.7	-49.6	0.5
Z2		13.6	-19.1	34.8	13.4	-26.5	23.6
Z3		39.4	6.7	60.6	45.0	5.1	55.2
Z1	2	3.5	21.0	83.6	-35.4	-43.9	-54.2
Z2		28.7	46.1	108.8	8.6	0.1	-10.2
Z3		22.7	40.1	102.7	-12.3	-20.8	-31.0

meetseizoen 1 = 2000-2001, meetseizoen 2 = 2001-2002

Bijlage 4a Model Akkerbouw met en zonder DOC

Akkerbouw, Model 1 met en zonder $DOC_{grondwater}$ gebaseerd op alleen de proefplekken waar $DOC_{grondwater}$ gemeten is.

Tabel 1. Schatting van de regressie-coëffeciënten met standaardfout (se), het percentage verklaarde variantie (R^2_{adj}) en de standaardfout (sd) van model 1, akkerbouw, met en zonder $DOC_{grondwater}$ gebaseerd op 168 datapunten

Model	1		+ DOC	
a	1.02	0.16	1.04	0.16
b	-0.28	0.08	-0.25	0.08
c	-6.8	16.7	3.1	17.2
$DOC_{grondwater}$			-0.40	0.19
R^2_{adj}	53.4 %		54.4 %	
Sd	59.5		58.9	

Tabel 2. Schatting van de constante C_i (in mg/l) per bodem-Gt-gewasgroep combinatie van model 1, akkerbouw, met en zonder $DOC_{grondwater}$.

Model	*Gewas*	*Hoog*			*Midden+laag*		
	Gt-groep	*1*	*2*	*3*	*1*	*2*	*3*
	Bodem						
1	Löss	*	*	*	*	*	*
	Z1	148.8	194.2	212.8	112.8	158.2	176.8
	Z2	167.5	212.9	231.5	131.5	176.9	195.5
	Z3	159.6	205.0	223.6	123.6	169.0	187.6
+DOC	Löss	*	*	*	*	*	*
	Z1	150.8	191.3	205.1	116.1	156.6	170.4
	Z2	168.5	209.0	222.8	133.8	174.3	188.1
	Z3	163.7	204.2	218.0	129.0	169.5	183.3

Bijlage 4b Model Gras met DOC

Gras, model met $DOC_{grondwater}$ gebaseerd op alleen de proefplekken waar $DOC_{grondwater}$ is gemeten.

Tabel 1. Schatting van de regressie-coëffeciënten met standaardfout (se), het percentage verklaarde variantie (R^2_{adj}) en de standaardfout (sd) van model 1, grasland, met en zonder $DOC_{grondwater}$ gebaseerd op 273 datapunten

	Model 1		*+ DOC*	
	Schatting	*Se*	*Schatting*	*Se*
a	0.80	0.10	0.73	0.10
b	-35.5	14.9	-29.0	14.6
$DOC_{grondwater}$			-0.44	0.12
R^2_{adj}	20.9 %		24.6 %	
Sd	48.4		47.3	

Tabel 2. Schatting van de constante C_i (in mg/l) per bodem-Gt-combinatie van model 1, grasland , met en zonder $DOC_{grondwater}$.

	Model 1			*+ DOC*		
Gt-groep	*1*	*2*	*3*	*1*	*2*	*3*
Bodem						
Löss	*	*		*	*	
Z1	11.5	22.3	24.9	26.4	37.1	36.8
Z2	11.1	21.9	24.5	25.5	36.2	35.8
Z3	16.9	27.7	30.3	38.2	48.9	48.6

Maïs, model met $DOC_{grondwater}$ gebaseerd op alleen de proefplekken waar $DOC_{grondwater}$ is gemeten.

Tabel 3. Schatting van de regressie-coëffeciënten met standaardfout (se), het percentage verklaarde variantie (R^2_{adj}) en de standaardfout (sd) van model 1, maïs, met en zonder $DOC_{grondwater}$ gebaseerd op 99 datapunten

	Model 1		*+ DOC*	
	Schatting	*Se*	*Schatting*	*Se*
a	0.88	0.22	0.96	0.21
b	-9.5	29.4	-2.4	28.9
$DOC_{grondwater}$			-0.76	0.33
R^2_{adj}	29.9 %		33.0 %	
Sd	51.0		49.8	

Tabel 4. Schatting van de constante C_i (in mg/l) per bodem-Gt-combinatie van model 1, maïs, met en zonder $DOC_{grondwater}$.

	Model1			*+ DOC*		
Gt-groep	*1*	*2*	*3*	*1*	*2*	*3*
Bodem						
Löss	*	*	*	*	*	*
Z1	-8.3	6.3	61.1	7.4	20.6	66.2
Z2	1.1	15.7	70.5	15.1	28.4	74.0
Z3	1.2	15.8	70.6	24.9	38.1	83.7