

Ontwikkeling van een indicator om te Sturen Op Nitraat

Ontwikkeling van een indicator om te Sturen Op Nitraat

Gegevens en regressie-analyse voor het eerste meetseizoen (2000-2001)

M.J.D. Hack-ten Broeke¹

S.L.G.E. Burgers²

H.F.M. ten Berge³

P.L.A. van Enckevort⁴

J.J. de Gruijter¹

I.E. Hoving⁵

A. Smit¹

G.L. Velthof¹

¹ **Alterra**

² **Biometris**

³ **Plant Research International**

⁴ **Praktijkonderzoek Plant en Omgeving**

⁵ **Praktijkonderzoek Veehouderij**

Alterra-rapport 772

Reeks Sturen Op Nitraat 4

Alterra, Research Instituut voor de Groene Ruimte, Wageningen, 2003

REFERAAT

Hack-ten Broeke, M.J.D., S.L.G.E. Burgers, H.F.M. ten Berge, P.L.A. van Enkevort, J.J. de Gruijter, I.E. Hoving, A. Smit 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, Research Instituut voor de Groene Ruimte. Alterra-rapport 772; Sturen op Nitraat 4; 65 blz.; 4 fig.; 24 tab.; 7 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 één meetseizoen (2000/2001) komt in ieder geval de indeling in drie Gt-groepen naar voren als een verklarende factor voor het niveauverschil in de gemeten nitraatconcentraties in het voorjaar van 2001. De hoeveelheid Nmin in oktober-december 2000, 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 772. Dit bedrag is inclusief BTW en verzendkosten.

© 2003 Alterra, Research Instituut voor de Groene Ruimte,
Postbus 47, NL-6700 AA Wageningen.
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.

Inhoud

Woord vooraf	7
Samenvatting	9
1 Inleiding	13
1.1 Werkwijze	14
2 Gebruikte gegevens	15
2.1 Gegevensverzameling	15
2.1.1 steekproefopzet	15
2.1.2 bemonsteringsprotocol	16
2.1.3 bedrijfsgegevens	17
2.2 Gebruikte data per proefplek	17
2.3 Beschikbare gegevens van de kandidaat-indicatoren	18
2.4 Beschikbare gegevens van de overige gemeten variabelen	23
3 Regressie-analyse, aannames en gebruikte technieken	27
3.1 Model-based versus design-based	27
3.2 Onderscheid naar de verschillende bronnen van variatie	28
3.3 Gebruikte selectie methoden	29
3.4 onderscheid naar akkerbouw en veehouderij	30
4 Resultaten van de regressie-analyse	31
4.1 Akkerbouw	31
4.1.1 Gewasgroep a, b en r	31
4.1.2 Gewasgroep t	34
4.2 Veeteelt	35
4.2.1 Gras	35
4.2.2 Maïs	38
5 Nmineraal in de tijd en de diepte	43
6 Opschaling van proefplek naar bedrijf	49
6.1 Inleiding	49
6.2 Voorspelling van de bedrijfsgemiddelde nitraatconcentratie	50
6.3 De nauwkeurigheid van voorspellingen van bedrijfsgemiddelde nitraatconcentraties	52
6.4 Berekening van de voorspelling en de nauwkeurigheid van een cluster- en een bedrijfsgemiddelde nitraatconcentratie	53
6.4.1 Akkerbouw	54
6.4.2 Veehouderij	54
7 Conclusies	57
7.1 Regressie-analyse met proefplekgegevens	57
7.2 Opschaling naar bedrijfsniveau	58
Literatuurlijst	59
Bijlage 1	61
Bijlage 2	65

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 onderzoekspartners Alterra, Research Instituut voor de Groene Ruimte, Praktijkonderzoek Plant en Omgeving (PPO); Praktijkonderzoek Veehouderij (PV); Centrum voor Landbouw en Milieu (CLM); en Plant Research International B.V. (PRI)

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:

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

Dit rapport bevat de eerste resultaten van de analyse op basis van slechts één meetseizoen. Als de gegevens van het tweede meetseizoen verwerkt zijn, zal opnieuw een rapport worden uitgebracht. Dit zal waarschijnlijk in het najaar van 2003 worden gerealiseerd. De analyse-resultaten kunnen dan verschillen van de resultaten in het rapport dat nu voor u ligt. Tenslotte zal een eindrapport worden uitgebracht van dit deelproject op basis van de gegevens van drie meetseizoenen. Tot dan toe zijn alle resultaten te beschouwen als voorlopig.

Samenvatting

Het doel van het Europese nitraatbeleid is schoon grond- en oppervlaktewater. Omdat het MINAS-systeem zo zijn beperkingen kent is een indicator gewenst, die dichterbij het milieudoel voor nitraat in het grondwater. Voor het stikstofbeleid op (droge) zand- en lössgronden is behoefte aan een mogelijkheid om gericht te sturen op nitraat in het grondwater 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-perceeloverschot, N-mineraalgehalten in de bodem, weersgegevens en locatiespecifieke factoren zoals grondsoort, grondwatertrap (Gt) en gewas 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 één meetseizoen (2000/2001) voor het project Sturen Op Nitraat kunnen alleen voorlopige conclusies worden getrokken. Pas als alle gegevens van drie meetseizoenen verwerkt zijn, zijn de conclusies niet voorlopig meer.

Op basis van waarnemingen op 466 proefplekken komt in ieder geval de indeling in drie Gt-groepen (respectievelijk Gt-groep 1 met GHG (Gemiddelde Hoogste Grondwaterstand) ondieper dan 40 cm beneden maaiveld, Gt-groep 2 met GHG tussen 40 en 80 cm beneden maaiveld en Gt-groep 3 met GHG dieper dan 80 cm beneden maaiveld) naar voren als een verklarende factor voor het niveauverschil in de gemeten nitraatconcentraties in het voorjaar van 2001.

Uit de verzamelde Nmin-gegevens op 334 proefplekken komt de hoeveelheid Nmin in de eerste meetperiode, namelijk oktober-december 2000, gesommeerd over de drie bodemlagen (i.e. 0-90 cm) als meest verklarende variabele naar voren uit de groep aan Nmin-gegevens. Daarmee is duidelijk dat met name Nmin, gemeten in die eerste meetperiode, het meest geschikt lijkt als indicator voor nitraatuitspoeling.

Tevens bleek dat deze hoeveelheid N_{min} van de kandidaatindicatoren de hoogste correlatie had met nitraat in het voorjaar, gevolgd door het MINAS-overschot dat vooral op akkerbouwbedrijven een hoge correlatie met nitraat bleek te hebben. Bij de regressie-analyse zijn een groot aantal aanvullende variabelen meegenomen, zoals potentiële mineralisatie en denitrificatie, bemesting en neerslag, waarvoor de veronderstelling geldt dat allen een zekere relatie met nitraatuitspoeling vertonen.

Omdat er duidelijke verschillen waren tussen akkerbouw en melkveehouderij zijn de regressie-analyses uitgevoerd voor akkerbouwgewassen, gras en maïs afzonderlijk. In alle gevallen komen Gt-groep en N_{min} als belangrijke verklarende variabelen naar voren. Bij akkerbouw kan vervolgens uit verschillende extra variabelen worden gekozen, die allen een min of meer vergelijkbare verbetering van de regressie-vergelijking bewerkstelligen. Deze variabelen zijn neerslagoverschot in de winterperiode, bedrijfs- of MINAS-overschot of de neerslagsom in de winter in combinatie met potentiële mineralisatie. Met name de relevantie van neerslag is verrassend, aangezien de gegevens slechts betrekking hebben op één meetperiode en dus nog niet op weersvariatie tussen verschillende jaren.

Voor grasland levert de regressie-analyse naast Gt-groep en N_{min} de volgende extra verklarende variabelen op: DOC (dissolved organic carbon) in het grondwater, aangevuld met de werkzame of totale N-gift via bemesting. Voor maïs is naast N_{min} en Gt-groep alleen DOC in het grondwater naar voren gekomen als verklarende variabele voor de nitraatconcentratie in het grondwater.

Als de regressie-vergelijkingen worden gebruikt om nitraatconcentraties te voorspellen op puntniveau, zijn de betrouwbaarheidsintervallen nogal groot. Grofweg zijn deze intervallen de voorspelde nitraatconcentratie plus of min 92 tot 160 mg/l. 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 betrouwbaarheidsintervallen voor de voorspelde nitraatconcentratie dan nog slechts de voorspelde gemiddelde nitraatconcentratie plus of min 32 tot 38 mg/l zijn. Als de regressie-vergelijkingen gebaseerd zijn op meerdere groeiseizoenen zal deze voorspelfout verder afnemen.

Samengevat betekent dit dat op basis van één meetseizoen de voorlopige conclusie kan worden getrokken dat de grondwatertrap en N_{mineraal} in het najaar de belangrijkste verklarende variabelen zijn voor de gemeten nitraatconcentratie in het voorjaar. Daarmee lijkt N_{mineraal} de meest perspectievolle indicator om te sturen op nitraat. Pas als alle gegevens (drie meetseizoenen in plaats van één) zijn geanalyseerd zal blijken of deze conclusie overeind blijft. Als N_{min} uiteindelijk de beste indicator voor nitraatuitspoeling zal blijken te zijn, wordt het belangrijk om te

streven naar een voldoende laag Nmin-gehalte. Het exploratief onderzoek, dat eveneens deel uitmaakte van STuren OP NITraat (Ten Berge, 2002), heeft aangetoond dat het in veel gevallen mogelijk is om het Nmin-gehalte in het najaar te voorspellen uit met name managementgegevens (gewas, bemesting etc.).

1 Inleiding

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, een 'achteraf-systeem', geen goede inschatting te maken van het milieurendement van maatregelen. 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 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 gericht 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 indicator¹ 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.

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-perceeloverschot, 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).

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 het eerste meetseizoen. In hoofdstuk 5 wordt apart aandacht besteed aan de analyse van N-min metingen op verschillende diepten en tijdstippen. Tenslotte komt in hoofdstuk 6 de opschaling naar bedrijfsniveau aan bod.

2 Gebruikte gegevens

2.1 Gegevensverzameling

Voor de gegevensverzameling zijn 35 bedrijven geselecteerd. Deze bedrijven liggen verspreid over de zand- en lössgronden ('uitspoelingsgevoelige gronden') van Nederland (zie bijlage 1). Er zijn 15 akkerbouw- of vollegrondsgroentenbedrijven, 19 melkveehouderijbedrijven en één gemengd bedrijf. Eén van de bedrijven is gedurende de opzet van het onderzoek afgevallen, omdat bij veldbezoek bleek dat op alle percelen van dit melkveehouderijbedrijf akkerbouwgewassen werden geteeld, dus de data-analyse betreft een totaal van 34 bedrijven. Dit aantal bestaat uit 8 proefbedrijven, 18 voorloperbedrijven van de projecten Koeien en Kansen, Telen met Toekomst, BIOM en BIOVEEM en tenslotte 8 praktijkbedrijven. Dit hoofdstuk beschrijft de gegevensverzameling globaal. Meer achtergronden zijn te vinden in Smit et al. (2003).

2.1.1 steekproefopzet

Bij het begin van het onderzoek is gekozen voor een gestratificeerde aselechte steekproef met een voor het onderzoek zo gunstig mogelijke verdeling van steekproefplekken over de bedrijven, de voorkomende grondsoorten, grondwatertrappen (Gt's) en geteelde gewassen. Stratificatie betekent dat er een indeling is gemaakt in zogenaamde strata van het voor bemonstering in aanmerking komende gebied. Aselect betekent dat steekproefplekken binnen de strata zijn geloot. De strata zijn gedefinieerd als combinaties van vier factoren, die belangrijk kunnen zijn voor nitraatuitspoeling: bedrijf, grondsoort, Gt-groep en gewasgroep. Voor de loting van steekproefplekken hield dit in dat van elk bedrijf een geografisch bestand werd gemaakt van cellen (potentiële steekproefplekken) van 5 X 5 m met steeds één grondsoort, Gt-groep en gewasgroep. Binnen een bedrijf werden deze combinaties van grondsoort, Gt-groep en gewasgroep clusters genoemd. De grondsoort en Gt-groep waren afkomstig van de bodemkaart 1 : 50 000 en de gewasgroep uit bedrijfsgegevens zoals die ten tijde van de loting beschikbaar waren. Vervolgens werden van alle op het bedrijf voorkomende combinaties van grondsoort, Gt-groep en gewasgroep een aantal cellen geloot als steekproefplekken. Er werd naar gestreefd deze aantallen zoveel mogelijk evenredig te kiezen met het oppervlak van de voorkomende combinaties. In het vervolg van deze tekst zal steeds gesproken worden over proefplekken in plaats van steekproefplekken. Voor de stratificatiefactoren grondsoort, Gt en gewas werden de volgende indelingen gekozen, met als leidraad het verwachte effect op nitraatuitspoeling.

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) GHG (Gemiddelde Hoogste Grondwaterstand) ondieper dan 40 cm (Gt I, II, II*, IIb, III, III*, V, V*)
- 2) GHG tussen 40 en 80 cm (Gt IIc, IV, VI)
- 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 de zomer van 2000 waren 27 van de 35 bedrijven bekend. Toen later dat najaar bekend was welke grondsoorten, grondwatertrappen (Gt's) en gewassen op die bedrijven voorkwamen, kon gericht gezocht worden naar aanvullende bedrijven met vooral die bodem-Gt-gewascombinaties waar nog proefplekken op nodig waren. Op die aanvullende bedrijven kon pas begin 2001 gestart worden met veldwerkzaamheden. In totaal zijn er 478 proefplekken geloot, verspreid over 47 bodem-Gt-gewascombinaties. Een van de proefplekken is afgefallen, toen na bemonstering en analyse bleek dat het hier een kleibodem betrof.

2.1.2 bemonsteringsprotocol

Voor het veldwerk is een protocol opgesteld om de gegevensverzameling volgens vooraf vastgestelde procedures te laten verlopen. Het betreft hier de volgende gegevens per proefplek:

- bodemprofielbeschrijving en grondwatertrap
- denitrificatiecapaciteit
- mineralisatiecapaciteit
- gehalte aan minerale N in de bodem (N_{min}) in het najaar
- nitraat in het bovenste grondwater of in het bodemvocht in het voorjaar

De eerste drie in dit rijtje zijn eenmalige bepalingen. De Nmin-metingen en nitraatbemonsteringen worden elk jaar herhaald tot in 2003. Het protocol schreef ook voor in welke gevallen een beoogde steekproefplek moest worden vervangen door een gelote reserve plek. Dit gebeurde bijvoorbeeld als ter plekke een ander gewas werd aangetroffen dan verwacht.

2.1.3 bedrijfsgegevens

De kandidaatindicatoren perceeloverschot en bedrijfsoverschot vallen onder de noemer bedrijfsgegevens. Vooral bij de veehouderijbedrijven is voor de 'vertaling' van het bedrijfsoverschot naar een perceeloverschot eerst een bodembalans nodig. Voor de data-analyse voor de periode zomer 2000-voorjaar 2001 zijn deze bedrijfsgegevens nodig voor het groeiseizoen van 2000. Ook is het bij de data-analyse interessant om te onderzoeken of factoren die in belangrijke mate bijdragen aan de uiteindelijke hoogte van het perceels- en/of bedrijfsoverschot ook variatie in de gemeten nitraatconcentraties kunnen verklaren. Daarom zijn voor alle proefplekken gegevens in de database opgenomen over bijvoorbeeld bemesting en bij grasland beweiding.

2.2 Gebruikte data per proefplek

- Grondsoort, Gt-groep en gewasgroep
Voor de indeling in strata zijn gegevens ontleend aan de 1 : 50 000 Bodemkaart van Nederland voor bodem en Gt en aan een grondgebruiksinventarisatie voor de gewasgroep. De indeling is ter plekke gecontroleerd, tevens is de Gt ter plekke vastgesteld.
- Nitraat
De nitraatconcentraties zijn gemeten in grondwater of bodemvocht in het 'voorjaar van 2001' (na de MKZ-crisis). De gegevens ontbreken voor 5 van de 478 proefplekken.
- Nmineraal
Dit betreft Nmin-waarnemingen voor de bodemlagen 0-30, 30-60 en 60-90 cm – mv. voor de meetperiode oktober-december 2000. Voor een aantal proefplekken zijn ook Nmin-bepalingen beschikbaar op latere tijdstippen, namelijk voor de periode december/januari 2000/2001 (tweede meetperiode) en de periode na 15 januari 2001. Voor de bedrijven die pas bekend werden na de jaarwisseling van 2000 (i.c. de Praktijkcijfersbedrijven) zijn geen Nmin-bepalingen uitgevoerd.
- 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 bovendien als extra gegevens die als indicatief voor potentiële mineralisatie kunnen worden beschouwd: Ntotaal, Ctotaal (g per kg) en hot-KCl extraheerbaar ammonium
- Organische stof en organisch N
- Dissolved organic carbon (DOC) in grondwater
- Zwavel in grondwater
- Perceeloverschot
Behalve perceeloverschot zijn belangrijke balansposten die mede bepalend zijn voor het perceeloverschot bij de analyse betrokken. Het betreft: kunstmestgift en dierlijke mestgift (zowel als werkzame N en als de totale N).
- Bedrijfsoverschot en MINAS-overschot; alle proefplekken van één bedrijf hebben hetzelfde overschot.
- Aantal dierweidedagen op grasland
- Aanwezigheid urineplek (ja/nee) op grasland
- Groenbemester gezaaid in 2000 (ja/nee) op akkerbouwland of maïs
- N-afvoer met het gewas
- Weersgegevens:
Dit betreft de neerslagsommen voor groeiseizoen en uitspoelingsseizoen (respectievelijk 1 april 2000-1 oktober 2000 en 1 oktober 2000-1 april 2001) en de neerslagoverschotten voor dezelfde perioden. Deze gegevens zijn vooralsnog berekend voor de dichtstbijzijnde weerstations en met behulp van referentie-gewasverdamping en gemiddelde correctiefactoren per gewasgroep.

2.3 Beschikbare gegevens van de kandidaat-indicatoren

Op basis van de dataset zijn overzichten gemaakt die inzicht geven in de beschikbare gegevens. In de tabellen 1 en 2 worden de gemiddelden van de gemeten nitraatconcentraties gegeven evenals de minimum, maximum en mediaan waarden per waargenomen bodemgroep en gewasgroep. De som van het aantal proefplekken is niet 478, maar 466, omdat 5 nitraatwaarnemingen ontbreken en een aantal lokaties om andere redenen afvallen. Deze komen verderop in de tekst aan bod.

Tabel 1. Overzicht van de Nitraatconcentraties (mg/l) per bodemgroep

<i>Bodem- groep</i>	<i>Aantal proefplekken</i>	<i>Gemiddelde</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Mediaan</i>
L	23	66,6	0,32	269,6	33,4
Z1	78	81,5	0,00	468,5	54,9
Z2	148	90,1	0,00	519,0	56,5
Z3	217	86,0	0,00	377,7	68,1

Tabel 2. Overzicht van de Nitraatconcentraties (mg/l) per gewasgroep

Gewas-groep	Aantal proefplekken	Gemiddelde	Minimum	Maximum	Mediaan
a	84	114,2	0,00	304,2	106,2
b	56	100,0	0,03	405,7	90,4
g	172	60,3	0,00	452,0	33,1
m	79	83,8	0,18	468,5	65,4
r	63	79,1	0,00	519,0	60,8
t	12	226,0	0,41	377,7	279,7

Het valt op dat voor alle bodemgroepen en voor de meeste gewasgroepen de mediaan behoorlijk lager uitvalt dan het gemiddelde. Dit betekent dat een paar hoge nitraatconcentraties het gemiddelde omhoog trekken. Per bodemgroep ligt het gemiddelde overal ruim boven de nitraatnorm van 50 mg/l. Op basis van de mediaan kan gesteld worden dat voor zandgronden 1 en 2 toch zo'n 50% van de waarnemingen onder die norm valt. De waarnemingen op de löss-gronden liggen grotendeels onder de norm.

Gewasgroep t valt op door zijn hoge gemiddelde en bovendien is hier de mediaan ook hoger dan het gemiddelde. Deze gewasgroep wijkt daarmee zodanig af van de rest dat hij steeds apart wordt bekeken. Het aantal waarnemingen binnen gewasgroep t is bovendien beperkt. Voor gewasgroep g (gras) geldt dat zowel het gemiddelde als de mediaan van de nitraatconcentraties duidelijk lager liggen dan voor de andere gewasgroepen.

In tabel 3 wordt het gemiddelde, minimum, maximum en mediaan gegeven van de nitraatconcentratie per vastgestelde Gt-groep. Hierbij is onderscheid gemaakt tussen gewassen op melkveehouderijbedrijven (gras en maïs) en de akkerbouwgewassen (waarbij gewasgroep t zoals gemeld apart wordt gegeven).

Tabel 3. Overzicht van de Nitraatconcentraties (mg/l) per Gt-groep en gewasgroep

Gewas groep	Gt-groep	Aantal proefplekken	Gemiddelde	Minimum	Maximum	Mediaan
g	1	22	32,1	0,00	178,9	6,7
	2	81	52,1	0,18	299,9	25,6
	3	69	78,9	0,32	452,0	50,5
m	1	23	65,4	0,40	382,9	50,6
	2	29	63,4	0,18	284,0	64,6
	3	27	121,4	25,74	468,5	92,9
a+b+r	1	22	47,7	0,00	219,3	17,5
	2	69	105,3	0,00	240,8	107,8
	3	112	106,0	0,04	519,0	82,7
t	1	2	155,3	2,92	307,8	155,3
	2	5	254,3	0,41	365,0	291,1
	3	5	225,9	28,98	377,7	268,4

Bij gras en maïs loopt zowel het gemiddelde als de mediaan op bij een lagere GHG (is hogere Gt-groep). Voor de akkerbouwgewassen is er geen verschil tussen Gt-groep 2 en Gt-groep 3. Bij maïs en bij de akkerbouwgewassen is de mediaan gelijk aan het gemiddelde bij Gt-groep 2. Voor de andere twee Gt-groepen en bij gras is de

mediaan steeds behoorlijk lager dan het gemiddelde, hetgeen betekent dat een paar hoge waarnemingen het gemiddelde behoorlijk omhoog trekken. De Gt-groepen laten onderling behoorlijke verschillen zien in de nitraatconcentratie. Deze verschillen hangen wel af van de gewasgroep. Dit kan betekenen dat de Gt-groep-indeling een medeverklarende variabele is voor de niveauverschillen in de gemeten nitraatconcentraties.

In tabel 4, 5, 6 en 7 wordt voor de andere kandidaat-indicatoren, Nmin, perceeloverschot, bedrijfsoverschot en MINAS-overschot, een overzicht gegeven per Gt-groep. Ook hierbij is steeds onderscheid gemaakt tussen de veehouderijgewassen (gras en maïs) en de akkerbouwgewassen.

Tabel 4. Overzicht van de Nmin (kg N/ha) gesommeerd voor de laag 0-90 cm en gemeten in de eerste periode, per Gt-groep

Gewas groep	Gt-groep	Aantal proefplekken	Gemiddelde	Minimum	Maximum	Mediaan
g	1	18	44,7	470	214,2	36,7
	2	70	51,9	10,34	389,5	43,6
	3	69	49,0	3,73	140,3	44,8
m	1	4	60,6	27,73	92,2	61,1
	2	20	77,9	24,79	213,6	63,1
	3	26	61,6	5,20	111,8	64,3
a+b+r	1	10	76,2	24,83	283,3	56,4
	2	9	65,6	19,08	175,6	42,9
	3	96	54,3	4,25	396,5	38,0
t	1	2	120,2	16,88	223,6	120,2
	2	5	201,0	79,44	592,3	117,7
	3	5	87,0	14,14	179,7	63,9

Tabel 5. Overzicht van het perceeloverschot(kg N/ha) per Gt-groep

Gewas groep	Gt-groep	Aantal proefplekken	Gemiddelde	Minimum	Maximum	Mediaan
g	1	20	139,7	19,7	336,3	131,4
	2	75	182,3	-104,4	409,3	180,9
	3	68	191,4	10,9	546,6	189,1
m	1	21	113,8	26,6	202,7	111,4
	2	26	123,0	-48,5	634,1	111,4
	3	29	53,3	-103,7	344,4	63,6
a+b+r	1	22	124,6	19,0	273,0	99,4
	2	70	179,1	14,0	328,0	215,0
	3	112	106,9	-74,0	420,0	95,0
t	1	2	190,5	152,0	229,0	190,5
	2	5	204,2	183,0	260,0	183,0
	3	5	176,4	152,0	274,0	152,0

Tabel 6. Overzicht van het (werkelijke) bedrijfsoverschot (kg N/ ha) per Gt-groep

Gewas groep	Gt-groep	Aantal proefplekken	Gemiddelde	Minimum	Maximum	Mediaan
g	1	12	158,5	-15,0	251,7	180,7
	2	58	210,9	-15,0	563,0	228,2
	3	48	188,6	69,0	563,0	170,5
m	1	8	108,5	95,0	175,3	102,0
	2	16	249,2	-15,0	563,0	245,3
	3	25	185,9	69,0	563,0	170,5
a+b+r	1	15	119,5	55,0	221,0	102,0
	2	34	123,2	55,0	355,0	102,0
	3	115	110,3	55,0	355,0	102,0
t	1	2	190,5	160,0	221,0	190,5
	2	5	247,8	221,0	355,0	221,0
	3	5	208,8	160,0	221,0	221,0

Het overgrote deel van de waarnemingen van de akkerbouwgewassen heeft Gt-groep 3. Voor maïs is Gt-groep 1 zeer minimaal vertegenwoordigd. Dit komt onder meer omdat een deel van de bedrijven pas laat (in 2001) is geselecteerd, zodat er geen Nmin-bepalingen meer konden worden gedaan in het najaar van 2000, maar ook omdat maïs op natte gronden niet zoveel voorkomt. Opvallend is dat noch het gemiddelde noch de mediaan in deze tabellen 4 t/m 7 een duidelijke lijn laten zien over de Gt-groepen heen. Dit is wellicht te verklaren doordat het effect van Gt-groep ofwel tot uiting komt in gewaskeuze danwel pas tot uiting komt in de nitraatuitspoeling gedurende de winter.

Tabel 7. Overzicht van het MINAS-overschot (kg N/ha) per Gt-groep

Gewas groep	Gt-groep	Aantal proefplekken	Gemiddelde	Minimum	Maximum	Mediaan
g	1	15	112,9	63,00	188,0	123,0
	2	60	114,5	63,00	188,0	123,0
	3	67	108,0	63,00	203,0	105,0
m	1	24	122,3	58,00	162,0	149,0
	2	26	130,4	72,00	203,0	123,0
	3	29	119,0	63,00	188,0	122,0
a+b+r	1	19	63,6	-36,00	149,0	50,0
	2	37	60,5	-36,00	176,0	58,0
	3	117	15,8	-36,00	176,0	6,0
t	1	2	77,5	6,00	149,0	77,5
	2	5	154,4	149,00	176,0	149,0
	3	5	120,4	6,00	149,0	149,0

De mediaan en het gemiddelde van de overschotten komen grotendeels overeen dus de verdeling van de data is bij de overschotten redelijk symmetrisch. Er is geen duidelijke lijn te zien in de grootte van de overschotten over de Gt-groepen heen.

Het bereik van het berekende 'werkelijke' bedrijfsoverschot is groter dan dat van het MINAS-overschot. De minima zijn lager en de maxima zijn hoger. Voor gras en de akkerbouwgewassen is het gemiddelde MINAS-overschot een stuk lager dan het

gemiddelde bedrijfsoverschot (het aantal data waarop het gemiddelde gebaseerd is, is ongeveer gelijk), zoals te verwachten was.

De variatie in nitraatconcentraties tussen proefplekken binnen een bedrijf is groot en ook de variatie in nitraatconcentraties tussen proefplekken binnen één perceel is groot. De variatie tussen de proefplekken binnen één cluster is ook groot met een standaard deviatie van 74.4, 72.9 en 91.2 voor resp. akkerbouw (a+b+r), gras en maïs. Dit betekent dat bij proefplekken die dezelfde bodemgroep, Gt-groep en gewasgroep hebben er toch nog grote verschillen in nitraatconcentratie worden gevonden.

Er is gekeken of er een directe relatie te zien is 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 de eerste periode) 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 perceelsoverschot 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 8. Correlatie-coëfficiënt (r) tussen nitraatconcentratie en de vier andere kandidaat-indicatoren voor gras, maïs en akkerbouwgewassen. Correlatie-coëfficiënt is berekend op basis van de eenheid (plek, perceel, bedrijf) waarop de kandidaat-indicator is gemeten

	Gras	Maïs	Akkerbouw
Nmin	0,24	0,39	0,42
Perceelsoverschot	0,11	-0,04	0,24
Bedrijfsoverschot	-0,31	0,04	0,61*
MINAS-overschot	0,10	0,11	0,35

*) deze hoge correlatie leunt zwaar op 1 bedrijf

Met name de hoge negatieve correlatie tussen bedrijfsoverschot en nitraatconcentratie bij gras is bijzonder, omdat dit zou betekenen dat een hoger overschot tot een lagere concentratie leidt. Uit de tabel blijkt met name dat Nmin en nitraatconcentraties de hoogste correlatie vertonen en dat bij akkerbouw de relatie tussen de kandidaat-indicatoren en nitraatconcentratie het duidelijkst is.

2.4 Beschikbare gegevens van de overige gemeten variabelen

In tabel 9 wordt een overzicht gegeven van alle andere variabelen die gemeten zijn en die een rol hebben gespeeld in de analyse van de gegevens.

Tabel 9. Overzicht van de variabelen, gebruikt in de regressie-analyse, ingedeeld naar de akkerbouwgewassen (gewasgroep a+b+r+t) en melkveehouderijgewassen (gewasgroep g+m)

Variabele	Aantal waarnemingen	Gemiddelde	Minimum	Maximum	Mediaan
potentiële denitrificatie, laag 1					
akkerbouw (mg N/kg/dag)	218	3,116	0,09266	14,76	2,404
veeteelt (idem)	251	4,630	0,00821	38,45	3,952
potentiële denitrificatie, laag 2					
akkerbouw (idem)	218	2,410	0,0403	11,59	1,881
veeteelt (idem)	252	3,362	-0,2114	14,40	2,855
potentiële denitrificatie, laag 3					
akkerbouw (idem)	218	1,800	0,02857	8,37	1,484
veeteelt (idem)	249	2,487	0,04977	13,05	2,011
potentiële denitrificatie, laag 4					
akkerbouw (idem)	218	1,152	-0,00028	12,62	0,5510
veeteelt (idem)	237	1,355	0,02189	9,73	0,8892
potentiële denitrificatie, laag 5					
akkerbouw (idem)	211	0,6346	-0,03966	3,560	0,2823
veeteelt (idem)	232	0,5341	-0,00305	5,396	0,2277
potentiële denitrificatie, laag 6					
akkerbouw (idem)	180	0,7025	0,00646	15,755	0,2232
veeteelt (idem)	199	0,4645	-0,03581	3,850	0,1814
Ntotaal in bouwvoor (g per kg)					
akkerbouw	222	1,692	0,5664	7,079	1,409
veeteelt	254	1,623	0,1852	6,254	1,472
Ctotaal in bouwvoor (g per kg)					
akkerbouw	222	35,21	10,584	136,6	28,18
veeteelt	254	26,92	9,324	101,9	25,05
C:Nverhouding					
akkerbouw	220	20,45	10,91	30,60	20,05
veeteelt	239	16,86	10,08	44,46	16,64
hot-KCl extraheerbaar NH ₄ -N					
akkerbouw (mg/kg)	220	18,63	0,100	64,17	16,40
veeteelt (idem)	239	25,71	4,333	76,85	24,20
potentiële mineralisatie bouwv.					
akkerbouw (mg N/ha)	220	1,008	-0,4982	2,691	0,954
veeteelt	239	2,107	0,0268	5,705	1,893
denitrificatiecapac. bouwvoor					
akkerbouw	220	3,373	0	10,70	3,100
veeteelt	239	6,468	0	110,30	4,200
kunstmest (km-)gift (kg N/ha)					
akkerbouw	218	70,91	0	320,0	66,00
veeteelt	245	100,41	0	337,5	89,10
N-werkzaam uit dierlijke mest					
akkerbouw (kg N/ha)	218	90,3	0	265,0	101,0
veeteelt (idem)	245	116,2	0	308,2	112,6
N-totaal uit dierlijke mest (dm)					
akkerbouw (kg N/ha)	215	138,1	0	300,0	147,0
veeteelt (idem)	245	201,3	0	405,6	199,2

<i>Variabele</i>	<i>Aantal waarnemingen</i>	<i>Gemiddelde</i>	<i>Minimum</i>	<i>Maximum</i>	<i>Mediaan</i>
N-werkzaam uit km + dm					
akkerbouw	218	161,2	0	320,0	159,0
veeteelt	245	216,6	0	499,9	220,5
N-totaal uit km + dm					
akkerbouw	215	208,7	0	483,5	198,0
veeteelt	245	301,7	0	652,5	282,0
Dierweidedagen					
akkerbouw	218	0,00	0	0	0,000
veeteelt	218	112,86	0	2238	18,000
% organische stof in bouwvoor					
akkerbouw	223	5,565	2,000	25,00	4,000
veeteelt	255	4,380	1,500	20,00	4,000
N-organisch bouwvoor (g/kg)					
akkerbouw	220	5,739	0,8500	13,25	5,575
veeteelt	239	9,162	1,0000	32,00	8,400
Afvoer N via gewas (kg N/ha)					
akkerbouw	216	121,0	0,00	313,0	118,5
veeteelt	237	244,4	32,97	710,8	234,0
Zwavelgehalte grondwater					
akkerbouw (mg/l)	140	24,15	1,918	192,9	18,79
veeteelt (idem)	216	14,54	1,875	89,7	11,08
DOC-gehalte grondwater					
akkerbouw	140	29,06	3,135	123,6	25,22
veeteelt	215	35,51	2,701	370,8	27,08
Neerslagsom groeiseizoen					
akkerbouw (mm)	223	449,7	381,2	540,7	447,5
veeteelt	255	451,0	390,8	525,6	451,9
Neerslagsom winterseizoen					
akkerbouw	223	449,0	363,1	522,5	452,6
veeteelt	255	463,7	363,1	513,5	454,0
Neerslagoverschot groeiseizoen					
akkerbouw	223	65,17	-28,37	148,9	72,52
veeteelt	255	3,14	-74,98	74,2	5,79
Neerslagoverschot winter					
akkerbouw	223	412,8	328,8	485,3	411,7
veeteelt	255	394,2	328,8	476,4	400,7

Tevens is van een groot aantal proefplekken (voornamelijk op de akkerbouwbedrijven) genoteerd of er een groenbemester heeft gestaan in 2000. Het overzicht van deze data staat in tabel 10. Tenslotte is er voor de graslanden genoteerd of er urineplekken gedetecteerd zijn ter hoogte van de proefplekken of niet.

Tabel 10. Overzicht van de aantal proefplekken met en zonder groenbemester in 2000 en het gemiddelde nitraatgehalte (mg/l) van deze plekken

<i>Gewasgroep</i>	<i>nee</i>		<i>ja</i>	
	<i>aantal waarnemingen</i>	<i>gemiddelde nitraatconc.</i>	<i>aantal waarnemingen</i>	<i>gemiddelde nitraatconc.</i>
m	21	101,5	25	44,8
a	73	115,5	8	117,7
b	43	82,7	13	157,4
r	55	83,2	8	51,5
t	2	16,0	10	268,0

Er is hiermee geen duidelijk beeld te verkrijgen van het effect van een groenbemester op de nitraatconcentratie. Voor maïs (m) en gewasgroep r lijkt de nitraatconcentratie te dalen als gevolg van de groenbemester. Bij gewasgroep b echter ligt de nitraatconcentratie beduidend hoger indien er een groenbemester wordt geteeld. Dit zou door specifieke gewassen binnen de gewasgroep veroorzaakt kunnen worden, maar is niet verder nagegaan, omdat de gegevens niet compleet zijn.

In tabel 11 wordt de lineaire relatie tussen de nitraatconcentratie en alle gemeten variabelen gegeven in de vorm van een correlatie-coëfficiënt. Alle variabelen zijn per proefplek gemeten (zie ook tabel 9). De correlatie-coëfficiënt is gebaseerd op alle proefplekken waarop de betreffende variabele gemeten is.

Tabel 11. Correlatie van de nitraatconcentratie met alle gemeten variabelen (gebaseerd op alle proefplekken waarop een variabele gemeten is)

variabele	correlatie overall	correlatie akkerbouw	correlatie veehouderij
Nmin in eerste meetperiode	0,39	0,44	0,31
potentiële denitrificatie, laag 1	-0,11	-0,10	-0,04
potentiële denitrificatie, laag 2	-0,03	0,05	-0,01
potentiële denitrificatie, laag 3	0,00	0,06	0,04
potentiële denitrificatie, laag 4	-0,04	-0,04	-0,01
potentiële denitrificatie, laag 5	-0,03	-0,09	0,01
potentiële denitrificatie, laag 6	-0,06	-0,12	-0,02
Denitrificatiecapaciteit bouwvoor	-0,14	-0,19	-0,11
Potentiële mineralisatie	-0,27	-0,18	-0,22
Ntotaal in bouwvoor	-0,03	-0,02	-0,07
Ctotaal in bouwvoor	-0,02	-0,06	-0,10
C:Nverhouding	-0,02	-0,16	-0,06
Hot-KCl extraheerbaar NH ₄ -N	-0,16	-0,13	-0,07
N-werkzaam uit dierlijke en kunstmest	0,07	0,26	0,05
N-totaal uit dierlijke en kunstmest	0,09	0,33	0,11
Dierweidedagen	-0,07	*	-0,01
perceeloverschot	0,10	0,24	0,03
bedrijfsoverschot	0,05	0,40	-0,04
MINAS-overschot	0,10	0,35	0,08
Neerslagsom groeiseizoen	-0,24	-0,34	-0,16
Neerslagsom winterseizoen	0,13	0,35	-0,04
Neerslagoverschot groeiseizoen	-0,04	-0,31	-0,14
Neerslagoverschot winterseizoen	0,25	0,35	0,06
% organische stof in bouwvoor	0,06	-0,01	0,08
Afvoer N via gewas	-0,07	0,11	0,05
N-organisch in bouwvoor	-0,15	-0,01	-0,09
Zwavelgehalte grondwater	0,15	0,12	0,02

Een uitgebreide analyse van de verbanden tussen enerzijds de potentiële mineralisatie en potentiële denitrificatie en anderzijds mogelijke indicatoren voor bodemkwaliteit, namelijk het gehalte aan totaal N en C, oplosbaar N en C, de onderlingen verhoudingen C/N en hot KCl extra heerbaar ammonium is beschreven door Velthof (2003).

3 Regressie-analyse, aannames en gebruikte technieken

Op basis van de beschikbare gegevens is onderzocht of er een relatie bestaat tussen verschillende zogenoemde kandidaatindicatoren. Tevens wordt onderzocht of er nog andere variabelen zijn die variatie in de metingen mede kunnen verklaren. Regressie-analyse is daarvoor de aangewezen methodiek. In regressie-analyse wordt onderzocht hoe goed één bepaalde variabele (ook wel responsvariabele genoemd) voorspeld kan worden uit één of meer verklarende variabelen (ook wel predictorvariabelen genoemd). In de uitgevoerde analyse is de nitraatconcentratie in grondwater of bodemvocht als responsvariabele genomen. Er is onderzocht in hoeverre de nitraatconcentratie, zoals gemeten op een proefplek, voorspeld kan worden uit alle andere gemeten variabelen.

Om te komen tot een goed regressie-model moeten een aantal keuzes en aannames worden gedaan die betrekking hebben op de steekproefopzet, de manier waarop de data verzameld zijn en op het uiteindelijke doel van het model. In de volgende paragrafen worden de belangrijkste aspecten die een rol hebben gespeeld bij de analyse-methode en de keuzes die zijn gemaakt uitgelegd en verantwoord.

3.1 Model-based versus design-based

De proefplekken in de dataset zijn het resultaat van een gestratificeerde steekproef waarbij gestratificeerd is naar “cluster”, te weten een combinatie van bodemtype, Gt-groep, gewasgroep en bedrijf. Een complete verantwoording is te vinden in het rapport over gegevensverzameling (Smit et al., 2003).

In de statistische methodologie wordt onderscheid gemaakt naar een model-based versus een design-based benadering. De klassieke methoden en modellen van statistische inferentie (gevolgtrekkingen), zoals regressie- en variantie-analyse, zijn gebaseerd op de aanname dat de data verkregen zijn via een eenvoudige random steekproef. Indien de data worden verkregen via een complex design (zoals stratificatie), zijn standaard methoden beschikbaar voor puntschattingen die betrekking hebben op eigenschappen van de onderzochte populatie. De dualiteit van complexe steekproef-designs enerzijds en de hypothetische modellen nodig voor voorspellingen anderzijds betekent dat er gezocht moet worden naar een combinatie van de ‘design-based’ en ‘model-based’ benadering. Het kan bewezen worden (Nathan, 1988) dat de parameterschattingen van een gewogen regressie-analyse, waarbij ieder steekproeftrekking gewogen wordt met de inverse van zijn eigen insluitkans, model-zuiver en design-consistent zijn.

In een steekproef hebben alle proefplekken een bepaalde kans om geloot te worden, de zogenoemde insluitkans. Indien de beschreven stratificatie en allocatie (steekproefpunten evenredig verdeeld over oppervlakte per stratum) zou zijn opgevolgd en er geen vervanging van proefplekken zou zijn geweest, zijn de

insluitkansen van alle proefplekken gelijk. Het al dan niet wegen van deze kansen in een gewogen regressie-analyse heeft dan geen enkele invloed op het resultaat. Echter, bij het trekken van de proefplekken zijn er aanpassingen gedaan met betrekking tot de evenredige allocatie. Er is voor gekozen minimaal één punt te loten in elk cluster dat op een bedrijf voorkomt. Verder zijn er reserve punten geloot en gebruikt indien op het veld een andere gewas stond dan was voorzien. Daarom zijn de uiteindelijk gerealiseerde insluitkansen van alle proefplekken berekend. Er bleken grote verschillen in insluitkansen voor te komen. Daarmee ligt het voor de hand een gewogen regressie-analyse uit te voeren.

Het probleem bij een gewogen regressie-analyse zit vervolgens niet in het schatten van de modelparameters maar in het verkrijgen van de juiste varianties en covarianties. De varianties die gegeven worden door standaard statistische pakketten zijn in deze context incorrect. Daarmee zijn ook de standaard toetsen (t-toets en F-toets) niet valide en kunnen geen betrouwbaarheidsintervallen rond voorspellingen worden opgesteld. Het verkrijgen van goede varianties is theoretisch en praktisch niet eenvoudig.

Bovendien zijn er nog een aantal andere goede argumenten die pleiten voor een ongewogen regressie-analyse. Dit zijn de argumenten die horen bij een 'model-based' benadering. Het doel van het onderzoek is niet zozeer het beschrijven van de situatie rond stikstof op de populatie van bedrijven die nu meedoen in de steekproef, maar het opstellen van een model voor het voorspellen van nitraatuitspoeling voor alle bedrijven in Nederland. Daarnaast is met reden afgeweken van de evenredige allocatie. Door tenminste één punt te loten in ieder cluster dat op een perceel voorkomt, wordt meer inzicht verkregen in de mogelijke variatie in dat cluster dan wanneer alle punten van een cluster op een paar grote percelen van één bedrijf terecht zouden zijn gekomen. Deze punten in kleine percelen hebben nu een relatief grote insluitkans en zouden in een gewogen regressie weer een zeer ondergeschikte rol spelen.

De gegeven modellen in het vervolg van dit rapport zijn zodoende toch gebaseerd op een ongewogen regressie-analyse. Voor de gegeven modellen is steeds gecontroleerd of de parameterschattingen wezenlijk zouden veranderen indien toch een gewogen regressie uitgevoerd zou worden. Zolang dit niet het geval is, is er in de tekst verder geen aandacht aan besteed.

3.2 Onderscheid naar de verschillende bronnen van variatie

Naast het punt van de 'model-based' of 'design-based' benadering is er in deze dataset nog een ander aspect dat statistisch gezien speciale aandacht vraagt.

Een groot aantal variabelen is gemeten op het niveau van de proefplekken zodat iedere proefplek zijn eigen meting heeft. Voor een aantal variabelen is dit niet het geval. Zo hebben alle proefplekken op hetzelfde perceel per definitie hetzelfde gewas, perceeloverschot, kunstmestgift, dierlijke mestgift, aantal dierweidedagen,

groenbemester, N-afvoer via het gewas en neerslagoverschot. Verder zijn er variabelen die op bedrijfsniveau worden bepaald. Alle proefplekken van één bedrijf hebben hetzelfde bedrijfsoverschot, MINAS-overschot en dezelfde neerslagsom.

Het zal duidelijk zijn dat een variabele die op bedrijfsniveau wordt gemeten zoals bijvoorbeeld het bedrijfsoverschot, geen variatie kan verklaren tussen de proefplekken binnen een bedrijf. Dit betekent ook dat een mogelijk effect van deze variabele niet getoetst kan worden tegen variatie tussen proefplekken maar getoetst moet worden tegen variatie tussen bedrijven. In de modelformulering dient dus rekening gehouden te worden met de verschillende niveaus van variatie en welke variabele op welk niveau variatie kan verklaren en dus getoetst moet worden. In een REstricted Maximum Likelihood (REML) model kan men opgeven welke random termen er zijn (bedrijven, percelen en proefplekken) en welke variabelen (fixed termen) er zijn.

Van de gegeven modellen is gecontroleerd of de variantiecomponenten van de bedrijven en percelen voldoende klein zijn, en daarmee dus verwaarloosbaar, ten opzichte van de restvariantie (onverklaarde variantie tussen proefplekken). Indien dit zo is, is de variatie tussen bedrijven en tussen percelen binnen bedrijven van een vergelijkbare grootte als de variatie tussen de proefplekken en kan het effect van alle variabelen worden getoetst tegen deze variantie.

3.3 Gebruikte selectie methoden

Het doel van selectie van variabelen is te komen tot enerzijds een model dat zo goed mogelijk voorspelt (dus met een zo laag mogelijke restvariantie) met anderzijds zo min mogelijk parameters. Het volledige model (met alle verklarende variabelen) heeft per definitie de laagste restvariantie. De selectie vindt plaats op basis van de volgende criteria: percentage verklaarde variantie, *Mallow's C_p*, significantie van parameters en voldoende ongecorrleerd zijn van de verklarende variabelen. *Mallow's C_p* is een maat om te bekijken in hoeverre de restvariantie van het geselecteerde model groter is dan de restvariantie van het volledige model.

In GenStat zijn een tweetal procedures beschikbaar waarbij alle deelmodellen worden aangepast, dat wil zeggen alle mogelijke combinaties van verklarende variabelen. De beste combinaties (i.o. hoogste percentage verklaarde variantie en/of laagste *Mallow's C_p*) worden gegeven per aantal verklarende variabelen in het model. In de procedure RSELECT kunnen alle gemeten variabelen in de dataset in één keer worden meegenomen in de selectie. Echter, de clusterindeling bodemgroep, Gt-groep en gewasgroep kan al dan niet worden opgenomen in het model, maar speelt geen rol in de selectie. Het resultaat van RSELECT, al dan niet met cluster-indeling, is vervolgens als start gebruikt in de procedure RSEARCH. In RSEARCH kunnen kwalitatieve variabelen zoals de clusterindeling wel mee doen in de selectie.

De gepresenteerde modellen zijn modellen met het hoogste percentage verklaarde variantie waarbij alle parameters in het model nog significant zijn. De gegeven

modellen hebben de laagste C_p maar voldoen niet per se aan het criterium $C_p < p+3$ (Oude Voshaar; 1994). Er is gecontroleerd of de verklarende variabelen in het model onderling voldoende ongecorrleerd zijn om een stabiel model te geven. Tenslotte is onderzocht of de interactie tussen bodemgroep, Gt-groep en/of gewasgroep significant is, of een interactie tussen de clusterindeling en één van de verklarende variabelen in het model significant is en of er kwadratische termen in het model moeten worden opgenomen. Dit was nergens het geval.

Voor ieder model hebben er een aantal datapunten een hoog residu. Dit betekent dat zij niet goed bij de regressielijn passen. Het al dan niet weglaten van deze punten heeft weinig invloed op de parameterschattingen en dus niet op het model en de modelvoorspellingen. Het weglaten van deze punten resulteert wel in een lagere standaardfout van het model en daarmee een kleiner betrouwbaarheidsinterval rond de voorspelling, maar dit is uiteraard geen goede reden om waarnemingen weg te laten. Er is een uitzondering gemaakt voor één proefplek met gras en twee proefplekken met maïs wegens extreem hoge nitraatconcentraties, die niet goed pasten bij de rest van de dataset. De invloed van deze punten op de selectie, de modelparameters en/of de betrouwbaarheid van het model kan daarmee te groot zijn.

3.4 onderscheid naar akkerbouw en veehouderij

In eerste instantie is gestart met een analyse waarbij alle proefplekken meedoen. Het resultaat van de regressie-analyse na selectie van variabelen leidde niet tot goede voorspellende modellen. Vervolgens is onderzocht of het resultaat te verbeteren is door onderscheid te maken naar de akkerbouwgewassen (gewasgroep a, b, r en t) en naar de gewassen op veehouderijbedrijven (gras en maïs). Dit leverde het inzicht op dat voor beide gewasgroepen andere verklarende variabelen een rol spelen die in het model worden opgenomen. Zodoende is er bij de verdere analyse voor gekozen deze gewasgroepen apart te onderzoeken.

4 Resultaten van de regressie-analyse

4.1 Akkerbouw

De gegevens van gewasgroep t zijn dusdanig afwijkend dat deze apart behandeld worden (zie bijvoorbeeld tabel 3 in hoofdstuk 2). Ze hebben dus niet meegedaan in het vaststellen van de regressie-modellen voor de akkerbouwgewassen.

4.1.1 Gewasgroep a, b en r

Op basis van de gewasgroepen a, b en r zijn er 203 proefplekken waar de nitraatconcentratie is gemeten in het voorjaar van 2001. Op 115 van deze plekken is ook Nmin in de eerste periode (oktober-december 2000) gemeten. Dit is een belangrijke verklarende variabele, zodat de regressiemodellen gebaseerd zijn op maximaal 115 proefplekken. Uiteindelijk zijn alle modellen gebaseerd op 109 proefplekken door het ontbreken van enkele gegevens voor de andere verklarende variabelen.

Uit de selectie komen 4 vergelijkbaar goede regressie-modellen naar voren. De modellen zien er als volgt uit :

$$\text{Model 1. : Nitraat} = C_i * Gt\text{-groep} + a * Nmin + b * \text{Neerslagoverschot2}$$

$$\text{Model 2. : Nitraat} = C_i * Gt\text{-groep} + a * Nmin + b * \text{bedrijfsoverschot}$$

$$\text{Model 3. : Nitraat} = C_i * Gt\text{-groep} + a * Nmin + b * \text{MINAS-overschot}$$

$$\text{Model 4. : Nitraat} = C_i * Gt\text{-groep} + a * Nmin + b * \text{Neerslagsom2} + e * \text{PotMin}$$

waarbij:

a, b, e en C_i (met $i=1,2,3$) = de te schatten regressie-coëfficiënten

$Nitraat$ = nitraatconcentratie in het voorjaar (mg/l)

$Nmin$ = Nmineraal in de eerste meetperiode voor de laag 0-90 cm (kg N per ha)

$Gt\text{-groep}$ = grondwatertrap-groep, 3 niveaus met constante C_i per niveau

$Neerslagoverschot2$ = Neerslagoverschot in winterperiode (mm)

$Bedrijfsoverschot$ = werkelijke bedrijfsoverschot (kg N per ha)

$MINAS\text{-overschot}$ = MINAS-bedrijfsoverschot (kg N per ha)

$PotMin$ = potentiële mineralisatie (mg N per ha per dag)

$Neerslagsom2$ = Neerslagsom in winterperiode (mm)

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

Tabel 12 Schatting van de regressie-coëfficiënten met standaardfout (se), het percentage verklaarde variantie (R^2_{adj}) en de standaardfout (sd) van model 1 t/m 4, akkerbouw

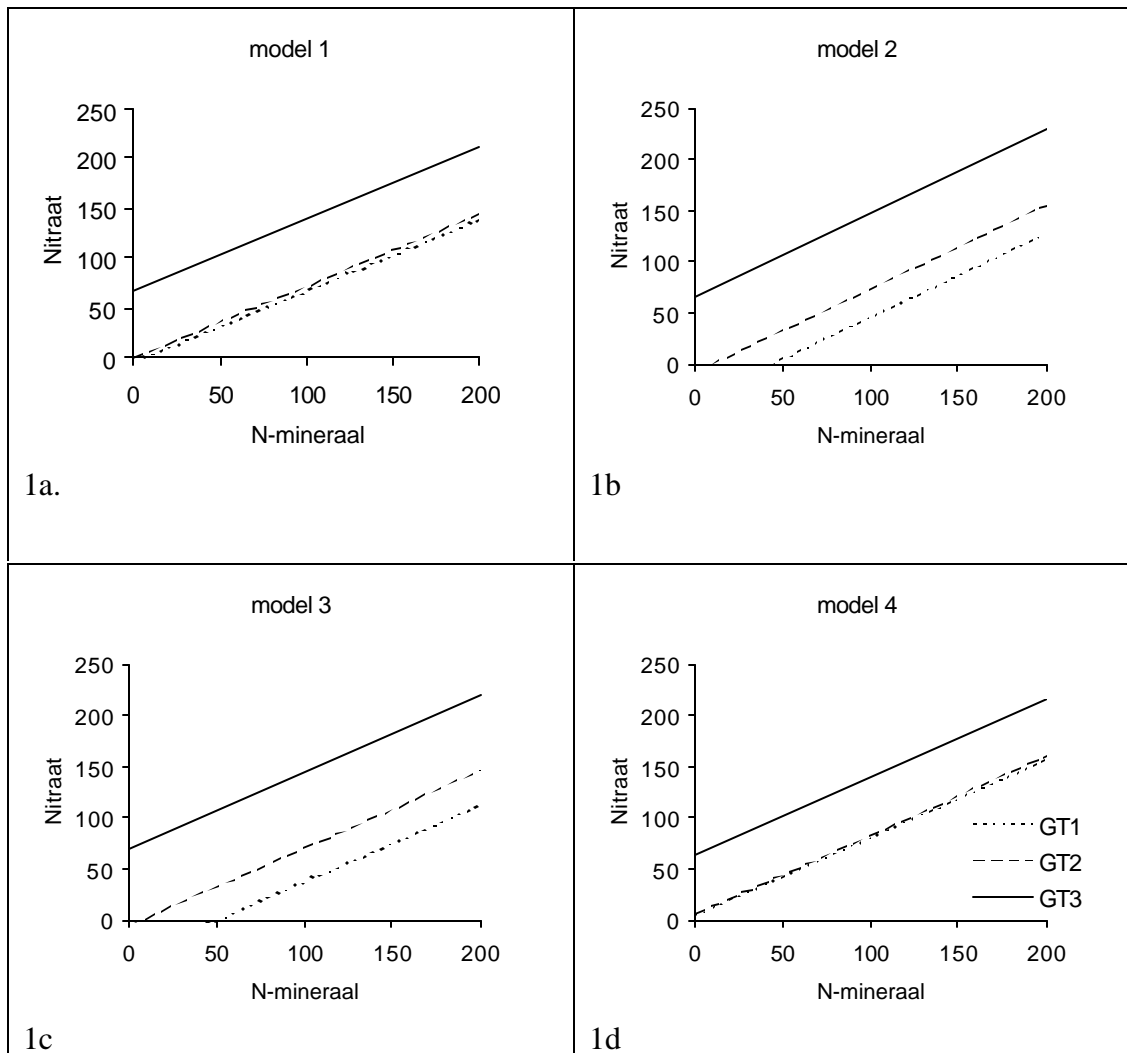
Model	Schatting						$R^2_{adjusted}$
	C ₁	C ₂	C ₃	a	b	e	
1	-651,5	-647,1	-579	0,72	1,51	-	42,1
2	-74,3	-46,1	28,6	0,82	0,33	-	35,7
3	-44,5	-11,7	63,8	0,76	0,45	-	36,0
4	-597	-595	-536	0,77	1,37	-31,7	43,2

Model	se						
	C ₁	C ₂	C ₃	a	b	e	sd
1	138	145	143	0,15	0,34	-	72,0
2	28,3	30,0	14,7	0,16	0,12	-	75,8
3	26,4	27,0	11,0	0,17	0,16	-	75,5
4	158	164	161	0,15	0,34	18,1	71,3

De relatie tussen de verklarende variabelen en nitraat lijkt heel redelijk maar de standaardfout van het model (sd) is groot. Voorspelling van nitraatconcentraties op basis van deze modellen zal daarom een groot betrouwbaarheidsinterval hebben.

Voor elk van de vier modellen zijn er enkele datapunten die sterk aan de regressielijn trekken (punten met een zogenoemd high leverage, dus grote hefboomwerking). Dit zijn waarnemingen van hoge nitraatwaarden bij relatief lage N_{min} waarden of vice versa. Het weglaten van deze zeven punten in de analyse resulteert bij de gegeven vier modellen in een steilere helling voor N_{min} (a wordt ongeveer 1.1), het effect van bedrijfsoverschot en MINAS-overschot wordt kleiner en het verschil tussen Gt-groep 2 en Gt-groep 3 wordt groter. Het weglaten van een zevental punten heeft dus een behoorlijke invloed op het resultaat. Daarmee zijn de gepresenteerde modellen dus niet stabiel.

De voorspellingen op basis van de vier regressie-modellen worden geïllustreerd in figuur 1. Voor de drie grondwatertrap-groepen wordt de voorspelde nitraatwaarde weergegeven bij een range van N_{min} en een vaste instelwaarde van de andere verklarende variabe(n) in het model. Die vaste waarde (zie figuuronderschrift) komt overeen met het gemiddelde van de data waarop het regressie-model is gefit.



Figuur 1a. t/m 1d. Voorspelling van de nitraatconcentratie op de y-as bij een range van Nmin (x-as) voor resp. model 1 t/m 4, akkerbouw :
 1a. Neerslagoverschot₂ = 429
 1b. bedrijfsoverschot = 113
 1c. MINAS-overschot = 12
 1d. PotMin = 1.11 en Neerslagsom₂ = 465

Opvallend is dat voor Gt-groep 3 alle vier modellen nagenoeg dezelfde lijn voorspellen. Voor de nattere Gt-groepen is er wat meer verschil tussen de vier modellen. De lijnen lopen parallel omdat in het regressie model geen interactie is opgenomen tussen Gt-groep en Nmineraal. Deze interactie is namelijk niet significant.

Naast de voorspelling is ook de betrouwbaarheid van die voorspelling van belang. In de onderstaande tabel worden voor Model 3 als voorbeeld de voorspellingen met *se*'s (voorspelfout) gegeven voor een aantal combinaties van ingestelde waarde van de variabelen. Een 95% betrouwbaarheidsinterval rond de voorspelling wordt verkregen door die voorspelling te nemen plus of min $1.96 * se$

Tabel 13 Voorspelling van nitraatconcentratie (mg/l) met voorspelfout *se* op basis van model 3. akkerbouw, bij gegeven instelling van Gt-groep, Nmin en MINAS(-overschot)

Model 3.	Nmin	MINAS = 40		MINAS = 80	
		Voorspelling	<i>se</i>	Voorspelling	<i>se</i>
Gt-groep 1	30	0 (-3,7)	79,8	14,3	80,4
	60	19,0	79,4	37,1	79,9
	90	41,7	79,4	59,8	79,6
Gt-groep 2	30	29,1	80,0	47,2	80,5
	60	51,9	79,8	69,9	80,1
	90	74,6	79,8	92,6	80,0
Gt-groep 3	30	104,6	76,4	122,7	77,2
	60	127,3	76,2	145,4	76,8
	90	150,0	76,3	168,1	76,8

De nitraatconcentratie stijgt bij een stijgende Nmin en er is een groot verschil tussen de drie Gt-groepen. De nitraatwaarde stijgt bij een dalende GHG ('drogere Gt-groep'). De verschillen in *se* zijn het gevolg van het verschillende aantal datapunten per Gt-groep.

In een REML-analyse, waarbij rekening wordt gehouden met de foutenstructuur in de data en met dezelfde verklarende variabelen in het model zijn de variantiecomponenten voor bedrijven en voor percelen binnen bedrijven heel erg klein (negatieve schattingen) ten opzichte van de schatting voor de variantie s^2 , het residu tussen de proefplekken. Dit betekent dat de verschillen in gemeten concentraties tussen bedrijven, en ook tussen percelen binnen een bedrijf, vergelijkbaar zijn met de variatie die men binnen een perceel aantreft. Het blijkt dat in deze dataset de variatie tussen bedrijven voor een groot deel wordt verklaard door het neerslagoverschot of de neerslagsom danwel de bedrijfsoverschotten.

Het opnemen van de factor 'bodemgroep' in het regressie-model geeft geen significante verbetering van het model. Het verschil tussen de drie parameterschattingen voor Z(and) 1, 2 en 3 is klein, het verschil met L(öss) is iets groter (-20). Het samenvoegen van de bodemgroepen Z2 en Z3 kan wel, maar ook dit resulteert niet in een significante bijdrage van de factor bodem in het model. Zelfs het samenvoegen van alle zandgronden resulteert toch niet in een significant verschil ten opzichte van löss.

4.1.2 Gewasgroep t

Gewasgroep t bestaat uit slechts 12 proefplekken verdeeld over 3 bedrijven. De gemiddelde nitraatconcentratie in deze groep is echter bijzonder hoog en afwijkend van die van de andere akkerbouwgewassen (tabel 2). Deze data hebben daarmee een grote en storende invloed indien zij samen met de andere akkerbouwgewassen in één analyse zouden worden opgenomen.

Het aantal waarnemingen is te klein om meer dan één verklarende variabele in een model op te nemen dus er kan alleen gekeken worden naar enkelvoudige lineaire

verbanden. Het is niet mogelijk om een volledige selectie van mogelijk verklarende variabelen uit te voeren.

De drie Gt-groepen hebben een verschillende gemiddelde nitraatconcentratie maar deze verschillen zijn niet significant. Ook is geen duidelijk verband met Nmin vast te stellen. De gevonden helling (0.44) is wel stijgend maar niet significant ($p=0.13$). Op het niveau van de perceelwaarnemingen is er wel een significante maar negatieve relatie met perceeloverschot en een positieve relatie met de totale N-gift (som van kunstmest-N en N-totaal in dierlijke mest). Op het niveau van bedrijfswaarnemingen kunnen geen uitspraken worden gedaan aangezien de data zijn gebaseerd op waarnemingen van slechts drie bedrijven.

4.2 Veeteelt

De twee gewassen, gras en maïs, die voorkomen op de veehouderijbedrijven in de dataset worden apart besproken.

4.2.1 Gras

Er zijn 172 proefplekken met gras waarop de nitraatconcentratie is gemeten. Op 157 van deze plekken is ook Nmin in de eerste periode (oktober-december 2000) gemeten. Eén datapunt doet niet mee in de analyse vanwege de extreem hoge nitraatconcentratie van dit punt (>450) in combinatie met een lage Nmin-waarde. Dit punt veroorzaakt veel variatie en heeft een missende waarde voor een aantal verklarende variabelen. In de selectie van variabelen blijkt ook DOC (opgelost organisch koolstof) in het grondwater van belang en deze variabele is gemeten op 129 proefplekken met gras.

Uit de selectie komen 2 vergelijkbaar goede regressie-modellen naar voren. Echter door het opnemen van $DOC_{\text{grondwater}}$ in het model daalt het aantal waarnemingen met een hoog nitraatgehalte. Dit is van invloed op de totale variatie in de data en verkleint daarmee de standaardfout van het model op een discutabele manier. Daarom wordt ook het model zonder $DOC_{\text{grondwater}}$ gegeven.

$$\text{Model 1. : Nitraat} = C_i * \text{Gt-groep} + a * \text{Nmin}$$

$$\text{Model 2. : Nitraat} = C_i * \text{Gt-groep} + a * \text{Nmin} + b * DOC_{\text{grondwater}}$$

$$\text{Model 3. : Nitraat} = C_i * \text{Gt-groep} + a * \text{Nmin} + b * DOC_{\text{grondwater}} + e * N_{\text{kunst_werk}}$$

$$\text{Model 4. : Nitraat} = C_i * \text{Gt-groep} + a * \text{Nmin} + b * DOC_{\text{grondwater}} + e * N_{\text{kunst_toddier}}$$

waarbij:

a, b, e en C_i (met $i=1,2,3$) = de te schatten regressie-coëfficiënten

Nitraat = nitraatconcentratie in het voorjaar (mg/l)

Gt-groep = grondwatertrap-groep, 3 niveaus met constante C_i per niveau

- N_{min} = Nmineraal in de eerste meetperiode voor de laag 0-90 cm (kg N per ha)
- $DOC_{\text{grondwater}}$ = opgelost organisch koolstof in het grondwater (mg/l)
- N_{kunst_werk} = som van kunstmest-N en de werkzame N uit dierlijke mest (kg N per ha)
- $N_{kunst_toddier}$ = som van kunstmest-N en de totale N uit dierlijke mest (kg N per ha)

De laatste twee variabelen zijn logischerwijs onderling sterk gecorreleerd en daarmee inwisselbaar in een regressie-model. N-afvoer met het gewas is positief gecorreleerd met de gift en zou een alternatief kunnen zijn (maar dan met een positieve regressie-coëfficiënt). In het model 1. is de toevoeging van de variabele 'potentiële mineralisatie' wel significant maar heeft de regressie-coëfficiënt een negatief teken. Er vallen dan 10 datapunten uit vanwege missende waarden en het percentage verklaarde variantie komt op 17.1 %.

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).

Tabel 14 Schatting van de regressie-coëfficiënten met standaardfout, het percentage verklaarde variantie en de standaardfout (sd) van model 1 t/m 4, grasland

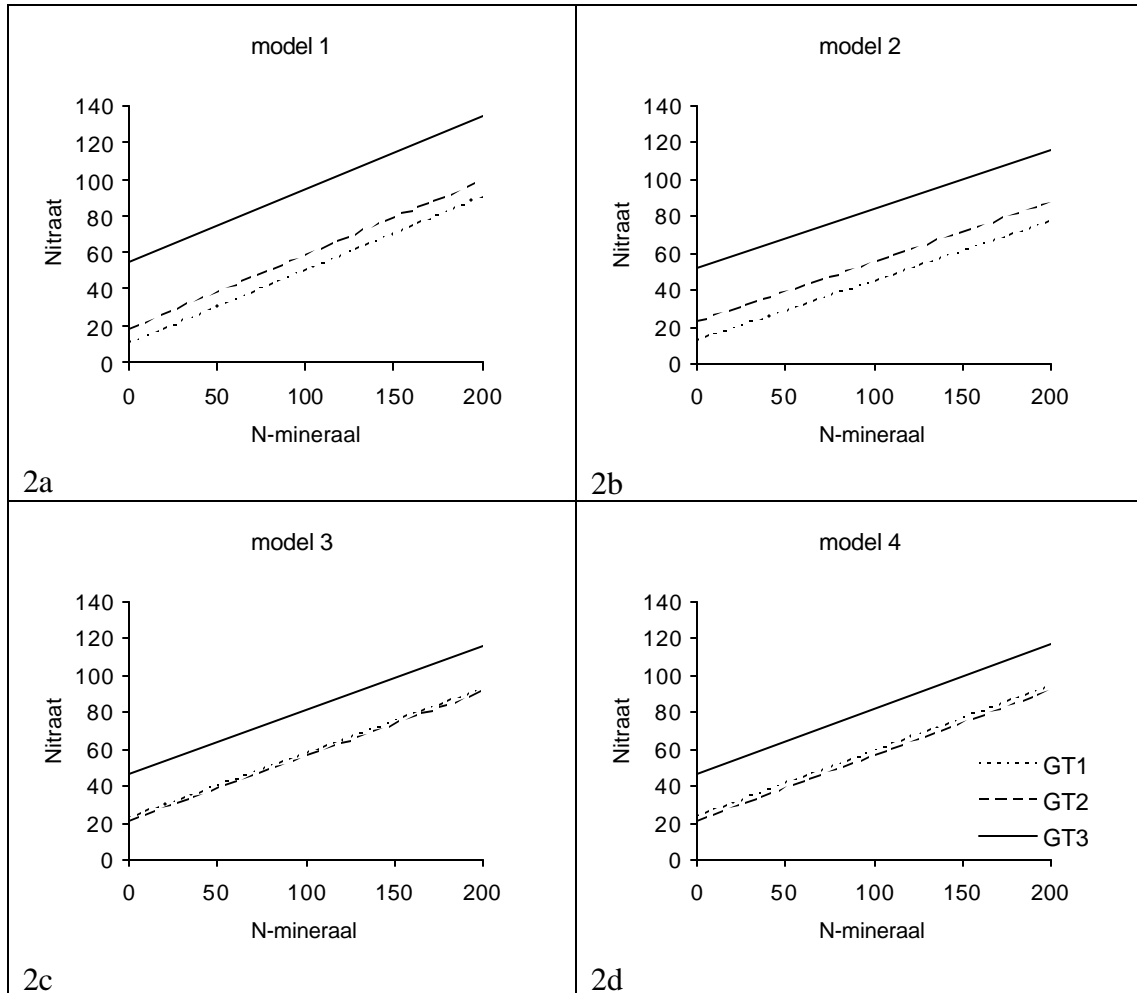
Model	Schatting						Aantal waarn.	R^2_{adjusted}
	C_1	C_2	C_3	a	b	e		
1	10,3	18,4	54,8	0,40	-	-	155	15,7
2	31,2	41,3	70,8	0,32	-0,61	-	129	21,7
3	16,2	14,4	39,7	0,35	-0,63	0,10	122	27,1
4	11,2	8,7	34,2	0,35	-0,65	0,10	122	27,0

Model	se						
	C_1	C_2	C_3	a	b	e	sd
1	14,0	8,7	8,6	0,11	-	-	55,6
2	13,6	10,3	10,3	0,09	0,19	-	46,7
3	14,3	13,4	14,8	0,10	0,19	0,03	45,8
4	14,9	14,7	16,2	0,10	0,20	0,03	45,9

Geen van de modellen geeft een goede relatie tussen de verklarende variabelen en de nitraatconcentratie. In alle gevallen zal daarom de voorspelling van nitraat op basis van het model onnauwkeurig zijn en een groot betrouwbaarheidsinterval hebben.

Alle vier de modellen laten een paar datapunten zien die sterk aan de regressielijn trekken. Het weglaten van deze punten in de analyse geeft een iets steilere helling voor N_{min} maar heeft verder geen grote invloed op de resultaten.

De voorspellingen op basis van de vier regressie-modellen worden geïllustreerd in figuur 2a. t/m 2d. Voor de drie grondwatertrapgroepen wordt de voorspelde nitraatwaarde weergegeven bij een range van N_{min} en een vaste instelwaarde van de andere verklarende variabele(n) in het model. Die vaste waarde wordt gegeven in het figuuronderschrift en komt overeen met het gemiddelde van de data waarop het regressie-model is gefit.



2a. t/m 2d. Voorspelling van de nitraatconcentratie op de y-as bij een range van Nmin (x-as) voor resp. model 1 t/m 4, grasland :

2b. DOC = 31.8

2c. DOC = 31.7 en Nkunst_werk=255

2d. DOC= 31.7 en Nkunst_totdier=319

Naast de voorspelling is ook de betrouwbaarheid van die voorspelling van belang. In de onderstaande tabel worden voor model 3 de voorspellingen met se 's gegeven voor een aantal combinaties van ingestelde waarde van de variabelen. Een 95% betrouwbaarheidsinterval rond de voorspelling wordt verkregen door die voorspelling te nemen plus of min $1.96 * se$.

Tabel 15 Voorspelling van nitraatconcentratie (mg/l) met se van model 3 voor grasland, bij gegeven instelling van Gt-groep, Nmin, Nkunst_werk en DOC=31.7

Model 3.	Nmineraal	Nkunst-werk=200		Nkunst-werk=300	
		Voorspelling	se	Voorspelling	se
Gt 1	20	23,8	47,3	34,2	47,5
	50	34,3	47,2	44,7	47,5
	80	44,8	47,3	55,2	47,6
Gt 2	20	22,1	46,4	32,4	46,3
	50	32,6	46,2	42,9	46,2
	80	43,1	46,3	53,4	46,3
Gt 3	20	47,4	46,6	57,7	46,5
	50	57,8	46,5	68,2	46,4
	80	68,3	46,6	78,7	46,5

De nitraatconcentratie stijgt bij een stijgende Nmin maar minder sterk dan bij de akkerbouwgewassen. Ook het verschil tussen de Gt-groepen is kleiner. Bij grasland is geen verschil te zien tussen Gt-groep 1 en Gt-groep 2 en de nitraatconcentratie is het hoogst bij de diepste GHG.

In een REML-analyse waarbij rekening wordt gehouden met de foutenstructuur in de data zijn de variantiecomponenten voor bedrijven en percelen binnen bedrijven heel erg klein ten opzichte van de schatting voor s^2 , het residu tussen de units. Dit geldt zelfs voor model 1 waar alleen de Gt-groep en Nmin in het model zitten. Dit betekent dat er in dit deel van de dataset (graslandplekken) geen grote verschillen zijn tussen de bedrijven.

Het opnemen van de factor 'bodem' in het regressie-model geeft geen significante verbetering van het model. Zodra DOC in het model zit zijn er voor grasland geen data op löss meer beschikbaar. Het verschil tussen zand 2 en zand 3 is klein. Samenvoegen van deze twee grondsoorten kan wel maar resulteert niet in een significante bijdrage van bodem.

4.2.2 Maïs

Er zijn 79 proefplekken met maïs waarop de nitraatconcentratie is gemeten. Op 50 van deze plekken is ook Nmin in de eerste periode (oktober-december 2000) gemeten. Twee datapunten doen niet mee in de analyse vanwege de extreem hoge nitraatconcentraties van deze punten (>350). In de selectie van variabelen blijkt ook DOC (opgelost organisch koolstof) in het grondwater van belang en die is gemeten op 42 proefplekken met maïs.

Uit de selectie komt één regressie-model naar voren. Echter, door het opnemen van $DOC_{\text{grondwater}}$ in het model daalt het aantal waarnemingen met een hoog nitraatgehalte. Dit is van invloed op de totale variatie in de data en verkleint daarmee de standaardfout van het model op een discutabele manier. Daarom wordt ook het model zonder $DOC_{\text{grondwater}}$ gegeven.

$$\text{Model 1. : Nitraat} = C_i * Gt + a * Nmin$$

$$\text{Model 2. : Nitraat} = C_i * Gt + a * Nmin + b * DOC_{\text{grondwater}}$$

waarbij:

a, b en C_i (met $i=1,2,3$) de te schatten regressie-coëfficiënten zijn.

$Nitraat$ = nitraatconcentratie in het voorjaar (mg/l)

Gt -groep = grondwatertrapgroep met 3 niveaus met constante C_i per niveau

$Nmin$ = N_{mineraal} in de eerste meetperiode voor de laag 0-90 cm (kg N per ha)

$DOC_{\text{grondwater}}$ = opgelost organisch koolstof in het grondwater (mg/l)

In tabel 16 worden per model de parameterschattingen met standaardfout gegeven, het percentage verklaarde variantie (R^2_{adj}) en de standaardfout van het model (sd).

Tabel 16 Schatting van de regressie-coëfficiënten met standaardfout, het percentage verklaarde variantie en de standaardfout (sd) van model 1 en 2, maïs

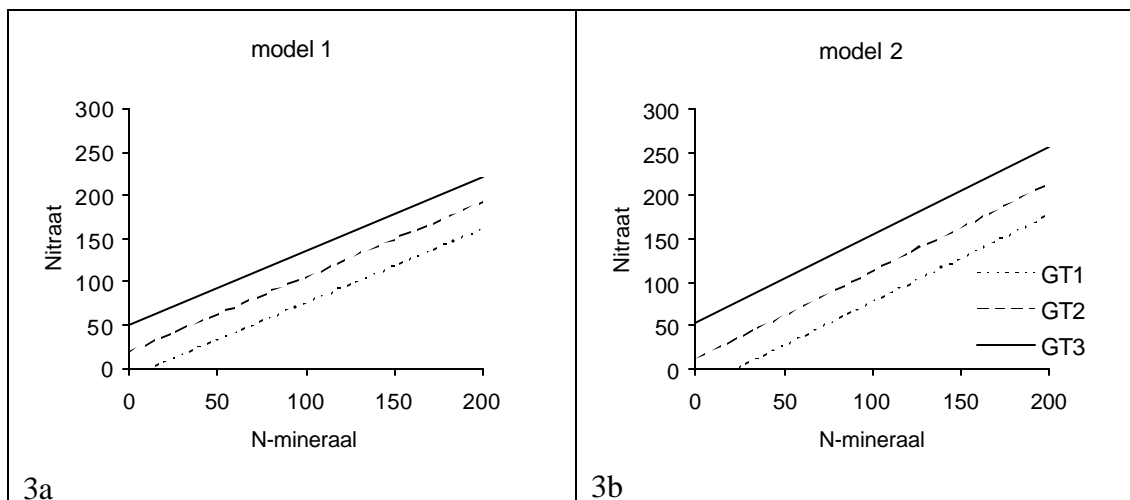
Model	Schatting					Aantal waarn.	R^2_{adjusted}
	C_1	C_2	C_3	a	b		
1	-11,3	18,6	48,7	0,86	-	47	23,4
2	-8,9	26,1	70,0	1,01	-0,55	42	38,3

Model	se					
	C_1	C_2	C_3	a	b	sd
1	36,5	21,6	17,6	0,22	-	56,8
2	36,5	27,5	20,1	0,22	0,45	52,6

De parameterschatting voor $DOC_{\text{grondwater}}$ is niet significant maar in de REML-analyse is deze wel significant. Daarom wordt het model met $DOC_{\text{grondwater}}$ toch gezien als een goede kandidaat. Het hoge percentage verklaarde variantie komt mede door het wegvallen van 5 datapunten met een hoge nitraatconcentratie. Geen van de modellen geeft een goede relatie tussen de verklarende variabelen en nitraat. In alle gevallen zal daarom de voorspelling van de nitraatconcentratie op basis van het model onnauwkeurig zijn en een groot betrouwbaarheidsinterval hebben.

Beide modellen hebben een zestal datapunten die sterk aan de regressielijn trekken (punten met een zogenoemde high leverage, dus grote hefboomwerking). Het weglaten van deze punten in de analyse betekent dat er geen data meer in Gt -groep 1 mee doen en dat de parameterschatting van $DOC_{\text{grondwater}}$ heel klein wordt. Zes datapunten is gegeven het totale aantal in deze dataset groot, maar betekent toch dat een relatief klein aantal punten een grote invloed heeft op het gepresenteerde resultaat.

De voorspellingen op basis van de twee regressie-modellen worden geïllustreerd in figuur 3a en 3b. Voor de drie grondwatertrapgroepen wordt de voorspelde nitraatconcentratie weergegeven bij een range van $Nmin$ en een vaste instelwaarde voor $DOC_{\text{grondwater}}$ bij Model 2. Deze vaste waarde, die wordt gegeven in het figuuronderschrift, komt overeen met het gemiddelde van de data waarop het regressie-model is gefit.



Figuur 3a en 3b. Voorspelling van de nitraatconcentratie op de y-as bij een range van Nmin (x-as) voor resp. model 1 en 2, maïs :
2b. DOC = 28.0

Naast de voorspelling is ook de betrouwbaarheid van die voorspelling van belang. In de onderstaande tabel worden voor model 2 de voorspellingen met *se*'s gegeven voor een aantal combinaties van Gt-groep, Nmin en DOC_{grondwater}. Een 95% betrouwbaarheidsinterval rond de voorspelling wordt verkregen door die voorspelling te nemen plus of min 1.96**se*.

Tabel 17 Voorspelling van de nitraatconcentratie (mg/l) met *se* van model 2. maïs bij gegeven instelling van Gt-groep, Nmin en DOC_{grondwater}

Model 2.	Nmineraal	DOC=20		DOC=40	
		Voorspelling	<i>se</i>	Voorspelling	<i>se</i>
Gt 1	20	0,3	61,8	0 (-10,6)	62,0
	50	30,5	61,0	19,6	61,3
	80	60,8	60,8	49,8	61,2
Gt 2	20	35,4	56,0	24,4	55,3
	50	65,6	54,8	54,7	54,2
	80	95,8	54,3	84,9	53,9
Gt 3	20	79,2	54,4	68,3	54,9
	50	109,5	54,0	98,5	54,6
	80	139,7	54,3	128,8	55,1

De nitraatconcentratie stijgt bij een stijgende Nmin, vergelijkbaar met de stijging bij de akkerbouwgewassen. De drie Gt-groepen laten onderling een groot verschil zien waarbij de nitraatwaarde ook hier het hoogste is bij de diepste GHG. De grotere *se* bij Gt-groep 1 is het gevolg van het kleine aantal datapunten in deze Gt-groep.

Een REML-analyse waarbij rekening wordt gehouden met de foutenstructuur in de data, geeft voor model 1 variantiecomponenten die niet verwaarloosbaar klein zijn

ten opzichte van de schatting voor s^2 . Dit betekent dat de variatie tussen de bedrijven niet wordt verklaard door Nmineraal en/of Gt-groep (de twee verklarende variabelen in dit model). In het opstellen van betrouwbaarheids-intervallen rond voorspellingen op basis van dit model, zou rekening gehouden moeten worden met de foutenstructuur in de data.

Het opnemen van de factor 'bodem' in het regressie-model geeft geen significante verbetering van het model. Zodra $\text{DOC}_{\text{grondwater}}$ in het model zit, zijn er ook voor maïs geen data op löss meer beschikbaar. Het verschil tussen zand 1 en zand 2 is klein zodat samenvoegen van zand 2 en zand 3 geen zin heeft.

5 Nmineraal in de tijd en de diepte

De Nmin-metingen zijn uitgevoerd voor 3 bodemlagen, namelijk 0-30, 30-60 en 60-90 cm. Voor 169 proefplekken zijn naast de Nmin-bepalingen in de periode oktober-december 2000 ook metingen gedaan op latere tijdstippen, namelijk voor de periode december-januari 2000/2001 en de periode na 15 januari. Dit betekent dat er voor deze proefplekken in feite negen Nmin-waarnemingen beschikbaar zijn, maar deze zijn niet onafhankelijk. Ook de datum van iedere meting is bekend. In dit hoofdstuk wordt onderzocht of er een verband bestaat tussen nitraatconcentratie in het grondwater en het Nmin-gehalte op verschillende diepten en tijdstippen en of er een relatie bestaat tussen de Nmin-waarnemingen op de verschillende tijdstippen onderling.

In tabel 18 wordt een overzicht gegeven van de gemiddelde nitraatconcentratie per bodemgroep, Gt-groep en gewasgroep voor de 169 proefplekken waarvan op drie tijdstippen Nmin is bepaald. De gemiddelde nitraatconcentraties van de gewasgroepen wijken behoorlijk af van gemiddelde concentraties van de gehele dataset (in tabel 2).

Tabel 18 Overzicht van de nitraatconcentraties (mg/l) per Gt-groep, bodemgroep en gewasgroep voor de proefplekken waarop drie keer in de tijd een Nmin-meting is uitgevoerd

	<i>Aantal waarnemingen</i>	<i>Gemiddelde nitraatconcentratie</i>
Gtnieuw		
1	17	48,2
2	40	83,4
3	112	95,3
bodemnieuw		
L	17	60,7
Z1	37	91,8
Z2	43	85,4
Z3	72	93,4
gewasnieuw		
a	18	86,4
b	16	63,2
g	74	71,6
m	26	116,0
r	26	48,4
t	9	297,7

In tabel 19 wordt een overzicht gegeven van de gemiddelde Nmin-gehalten (in kg N per ha) in de drie lagen op de drie tijdstippen.

Tabel 19 Overzicht van de gemiddelde Nmin-gehalten (in kg N per ha) in de drie lagen op de drie tijdstippen per Gt-groep, bodem- en gewasgroep (op basis van de plekken die op alle drie de tijdstippen zijn gemeten)

Tijdstip	Laag 0-30 cm			30-60 cm			60-90 cm		
	1	2	3	1	2	3	1	2	3
Gt-groep									
1	28,49	13,63	19,64	19,49	15,72	16,92	28,98	20,69	24,37
2	20,01	13,28	19,79	28,60	14,36	14,67	33,69	22,72	20,25
3	14,83	10,20	14,77	17,00	11,90	14,06	21,03	15,47	15,03
Bodemgroep									
L	16,99	13,60	15,37	22,67	18,47	19,02	30,03	29,07	26,43
Z1	19,05	11,78	19,27	19,24	13,41	16,80	16,56	16,11	18,66
Z2	20,68	12,09	18,89	21,58	12,64	13,97	26,78	15,70	16,14
Z3	14,72	9,76	13,76	18,41	11,07	12,24	26,16	16,19	14,08
Gewasgroep									
a	27,56	16,39	24,88	18,10	12,20	15,24	22,62	13,90	14,76
b	19,21	15,56	16,15	22,81	20,15	24,52	37,97	38,29	33,51
g	17,28	11,26	16,63	17,80	11,41	14,32	17,73	12,54	14,24
m	17,85	7,79	14,29	27,11	17,36	11,31	25,29	22,37	20,48
r	9,80	9,48	14,85	8,21	6,79	9,52	7,55	5,98	7,00
t	14,25	6,34	8,40	47,94	14,83	16,01	106,02	41,62	29,28

Uit deze tabel blijkt dat het verloop met de diepte afhangt van het tijdstip en vooral ook van de gewasgroep. Tussen Gt-groepen en bodemgroepen zijn op het oog geen opvallende verschillen te zien. Voor het verloop met de diepte bij de indeling per gewasgroep lijkt er over het algemeen sprake van een stijging met de diepte, behalve op tijdstip 3, want dan is er sprake van lagere Nmin-gehalten in de laag 30-60 in vergelijking met de laag erboven én met de laag eronder. Als de bodemgroepindeling wordt gehanteerd levert dit hetzelfde beeld op, dus opnieuw is bij de derde meetperiode sprake van lagere Nmin-gehalten in de tweede meetlaag. Bij de gewasgroepindeling komt weer een ander beeld naar voren. Bij gewasgroep r bijvoorbeeld is in alle gevallen sprake van een daling van Nmin-gehalte met de diepte, terwijl bij gewasgroepen b en t op alle tijdstippen sprake is van een stijging met de diepte. Het gewas lijkt dus voor Nmin-verdeling binnen het bodemprofiel van belang.

Voor het verloop van Nmin in de tijd is het opvallend dat in de tweede meetperiode de waarnemingen per bodemlaag altijd lager uitvallen dan in de eerste meetperiode, behalve voor gewasgroep b in de laag 60-90 cm. Verder is het zo dat in de derde meetperiode in de laag 0-30 cm, maar grotendeels ook in de bodemlagen daaronder, de Nmin weer hoger is dan in de tweede meetperiode. De verwachting is dat de oorzaak hiervoor tussen periode 1 en 2 voornamelijk uitspoeling is en dat er tussen periode 2 en 3 alweer sprake is van (beperkte) mineralisatie. Op basis van de potentiële mineralisatie en denitrificatie metingen wordt verwacht dat mineralisatie in de bouwvoor in de winter hoger is dan de denitrificatie, waardoor er dus een netto stikstofaanvoer in de bouwvoor plaatsvindt gedurende de winter (Velthof, 2003).

Er is een regressie-analyse uitgevoerd met de negen Nmin-metingen (drie lagen op drie tijdstippen) en de drie tijdstippen waarop gemeten is als verklarende variabelen in het model tegen de nitraatconcentratie. Daarnaast zijn de clusterindeling (Gt-groep, bodemgroep en gewasgroep) en de neerslagsom per periode (periode tussen eerste en tweede bemonstering en tussen tweede en derde bemonstering) in het model opgenomen. De resultaten van de regressie-analyse laten zien dat alleen het Nmin-gehalte in de 0-30 cm laag van het eerste tijdstip en het Nmin-gehalte in de 60-90 cm laag van het tweede tijdstip een significante en positieve relatie met de nitraatconcentratie vertonen. Dit zou kunnen worden verklaard door transport van nitraat; nitraat uit de 0-30 cm laag op tijdstip 1 bevindt zich op tijdstip 2 in de 60-90 cm en later in het grondwater.

Er is wel een verschil in de gemiddelde nitraatconcentratie tussen de Gt-groepen maar niet tussen de bodemgroepen. Gewasgroep t heeft een veel hogere nitraatconcentratie dan de andere gewasgroepen. Het percentage verklaarde variantie van het model zonder gewasgroep t is 22,4%. De voorspelfout van het model is groot (74,2 mg per l) zodat voorspellingen op basis van dit model onnauwkeurig zijn.

De neerslagsom tussen de derde Nmin-meting en de nitraat-concentratie geeft een significant negatieve parameterschatting, dus in deze dataset resulteert een hogere neerslaghoeveelheid in deze periode in een lagere nitraatconcentratie. De hoeveelheid neerslag tijdens de winter kan de nitraatconcentratie in het grondwater op drie manieren beïnvloeden:

- I. toename in de neerslag leidt tot meer uitspoeling en hogere nitraatconcentraties,
- II. toename in de neerslag leidt tot meer water in het bodemprofiel en tot een verdunning (en lagere concentraties) van het nitraat en
- III. toename in neerslag leidt tot meer denitrificatie en daardoor tot lagere nitraatconcentraties.

Deze effecten spelen door elkaar en de resultaten van de statistische analyse suggeren dat in de winter van 2000/2001 de tweede en derde factor een grotere rol hebben gespeeld dan het eerste effect.

In de beschreven regressie-analyse wordt alleen gezocht naar een directe relatie tussen de Nmin-gehalten in de lagen op de verschillende tijdstippen en de nitraatconcentratie. Een andere manier om naar de data te kijken is te onderzoeken of het verloop in de tijd (per laag of gesommeerd over de lagen) een relatie heeft met de nitraatconcentratie. Tijdens de winter zijn er verschillende processen die tot veranderingen in de hoeveelheid Nmin in bodemlagen kunnen leiden. Aanvoer van minerale N kan optreden door mineralisatie, atmosferische depositie (bovenste laag) en transport uit bovenliggende lagen naar eronder gelegen lagen (aannemende dat er geen bemesting of beweiding plaats vindt). Afvoer van minerale N kan optreden door uitspoeling naar diepere lagen, denitrificatie, immobilisatie en indien er een

gewas aanwezig is door gewasopname (dit geldt met name voor grasland en percelen met wintergewassen).

Het verloop van het Nmin-gehalte per laag in de tijd is te beschrijven met een eenvoudige lineaire regressielijn. De helling van deze lijn en bijvoorbeeld het intercept kunnen worden gezien als twee karakteristieken om het verloop in de tijd te beschrijven. De helling geeft hierbij de snelheid aan waarmee de hoeveelheid Nmin in een laag verandert en het intercept de hoeveelheid Nmin die in de herfst aanwezig was in de betreffende laag.

Per proefplek is per laag een eenvoudige lineaire regressie uitgevoerd op de drie Nmin-bepalingen tegen de meetdatum. De helling van de regressielijn en het intercept van de drie datapunten is berekend. Per proefplek worden zo 6 nieuwe variabelen verkregen, namelijk een helling en een intercept per laag, die het verloop in de tijd beschrijven. Vervolgens is met multi-pele lineaire regressie-analyse onderzocht of er een verband bestaat tussen deze karakteristieken en de bodem, Gt-, gewascombinaties enerzijds en de nitraatconcentratie anderzijds.

q

In tabel 20 wordt voor de laag 0-30 cm de gemiddelde helling in kg N per ha per dag gegeven per Gt-gewas-combinatie.

Tabel 20 Gemiddelde helling van lineaire regressielijn per proefplek per laag door de tijd voor laag 0-30 per Gt-gewascombinatie (kg N per ha per dag)

gewasnieuw	Gtnieuw		
	1	2	3
a	-0,4983	1,0691	0,0474
b	-0,1100	-0,0524	-0,0451
g	-0,0064	-0,0055	0,0055
m	0,0434	-0,0597	-0,0219
r	0,1079	0,1887	0,0575
t	-0,0571	-0,0649	-0,0822

De variatie tussen de hellingen van proefplekken per Gt-, bodem- of gewasgroep is groot. Voor bijvoorbeeld de laag 0-30 cm op zandgroep 1, Gt-groep 3 en gras varieert de helling van -0.26 tot $+0.43$ kg N per ha per dag. Dit betekent dat het verloop van de hoeveelheid Nmin in de tijd zowel kan dalen als kan stijgen voor proefplekken met eenzelfde bodem-Gt-gewascombinatie. In tabel 20 resulteert dit in een hele lage gemiddelde helling. In de gegevens uit tabel 20 is geen duidelijke lijn te ontdekken. De hellingen voor de lagen 30-60 en 60-90 cm geven een vergelijkbaar resultaat. Dit betekent waarschijnlijk dat er binnen bodem-Gt-gewascombinaties grote verschillen bestaan in de bovengenoemde aan- en afvoerposten van Nmin tijdens de winter. Vaak wordt aangenomen dat Nmin-gehalten tijdens de winter in een profiel sterk afnemen door verliezen via uitspoeling en denitrificatie ("het profiel is leeg in het vroege voorjaar"). De resultaten voor de winter 2000/2001, een relatief natte winter, geven dus heel duidelijk aan dat dit niet algemeen geldt en dat er ook een groot aantal proefplekken waren waarbij de hoeveelheid Nmin tijdens de winter toenam. De enige oorzaak hiervoor lijkt mineralisatie te zijn (zie ook Velthof, 2003).

De grote variatie tussen hellingen (binnen een bodem-Gt-gewascombinatie) zorgt er tevens voor dat er geen significante relatie wordt gevonden voor de helling van de drie lagen met de nitraatconcentratie. Ook het gemiddelde per laag per tijdstip heeft geen significante relatie met de nitraatconcentratie. In de huidige dataset is het niet mogelijk een goede relatie te vinden tussen het verloop van het Nmin-gehalte in de tijd met de gemeten nitraatconcentratie.

Een verandering van het Nmin-gehalte in de tijd kan ook onderzocht worden door te kijken naar het verschil in Nmin tussen twee tijdstippen, een zogenaamde Δ -Nmin. Er is bekeken of Δ -Nmin een relatie laat zien met de neerslagsom in dezelfde periode. Geen enkel verband kan worden aangetoond. Tenslotte is onderzocht of Δ -Nmin samen met die neerslagsom de variatie in nitraatconcentraties kan verklaren. In die analyse bleek de verandering tussen de eerste en tweede meetperiode in de bovenste laag (0-30 cm) significant, samen met de neerslagsom tussen de derde meting en de nitraat-meting. Dit komt overeen met eerder gevonden resultaten en levert geen verbetering op ten opzichte van die resultaten (in termen van percentage verklaarde variatie of voorspelfout).

Op dezelfde manier als het verloop met de tijd kan ook het verloop in de diepte worden onderzocht. Het verloop in de diepte wordt beschreven door een lineaire regressielijn per plek per tijdstip. De drie Nmin-bepalingen in de diepte worden dan uitgezet tegen de gemiddelde diepte. Ook nu is de helling van de lijn (snelheid waarmee Nmin toe- of afneemt in de diepte) op te vatten als een karakteristiek die het verloop in de diepte op één tijdstip beschrijft. In plaats van het intercept (Nmin op diepte 0) is nu de gemiddelde Nmin als andere beschrijvende karakteristiek genomen. Er is een multiële lineaire regressie-analyse uitgevoerd om te onderzoeken of er een verband bestaat tussen de bodem-Gt-gewasindeling en verloop van Nmin met de nitraatconcentratie.

Zoals verwacht is er een positief verband tussen de gemiddelde Nmin in de 0-90 cm laag van de eerste meetperiode en de nitraatconcentratie. Dit komt overeen met de resultaten van hoofdstuk 4 waarin Nmin, gesommeerd over de drie lagen in de eerste periode, steeds als een verklarende variabele naar voren kwam. Zonder gewasgroep t wordt er geen significante relatie gevonden tussen de snelheid waarmee Nmin toe- of afneemt in de diepte en de nitraatconcentratie. In de huidige dataset is het dus niet mogelijk een relatie te vinden tussen het verloop in de diepte en de gemeten nitraatconcentratie.

6 Opschaling van proefplek naar bedrijf

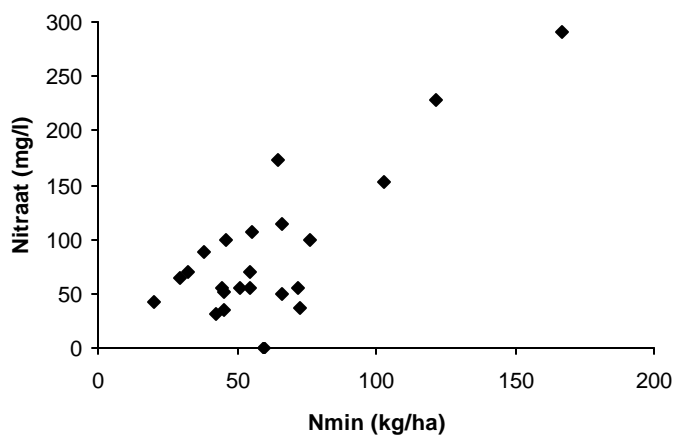
6.1 Inleiding

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:

- 1) 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.
- 2) 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.

Dit hoofdstuk gaat over het waarom en hoe van de opschaling, speciaal die van proefplek naar bedrijf. Daarvoor is het nodig eerst nader te bekijken hoe de bedrijfsgemiddelde nitraatconcentraties voorspeld zullen gaan worden.

De bedrijfsgemiddelde nitraatconcentratie kan berekend worden op basis van de waarnemingen waarbij rekening wordt gehouden met de clusteroppervlakten. Hetzelfde kan worden gedaan voor de bedrijfsgemiddelde N_{min} . In figuur 4 is de bedrijfsgemiddelde nitraatconcentratie uitgezet tegen de bedrijfsgemiddelde N_{min} . Naarmate de waarden hoger worden is het verband steeds duidelijker.



Figuur 4. Berekende bedrijfsgemiddelde nitraatconcentratie als functie van berekende bedrijfsgemiddelde N_{min}

6.2 Voorspelling van de bedrijfsgemiddelde nitraatconcentratie

De resultaten van de analyse van de gegevens uit de eerste meetronde zijn aparte lineaire modellen voor de akkerbouwgewassen (zonder gewasgroep t), gras en maïs. De eerste hebben betrekking op de akkerbouwbedrijven, de laatste twee op veehouderijbedrijven. De modellen hebben met elkaar gemeen dat ze verschillende intercepten hebben voor de Gt-groepen 1, 2 en 3. Ze verschillen wat betreft de kwantitatieve predictoren (verklarende variabelen) die erin zijn opgenomen.

Omdat we niet alleen verschillende modellen per gewasgroep hebben, maar binnen die modellen ook verschillende intercepten per Gt-groep, kan het modellarium met betrekking tot de voorspelling óók beschouwd worden als te zijn opgebouwd uit $3 \times 3 = 9$ verschillende modellen (één voor elk cluster, d.w.z. gewas-Gt-combinatie), van de volgende vorm:

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

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 Gt-groep;

$\hat{b}_{1j}, \hat{b}_{2j}$: de geschatte regressiecoëfficiënten 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 en 3 (zie hoofdstuk 4), namelijk het model met de Gt-groep, Nmin en resp. Neerslagoverschot in de winter danwel het Minas-bedrijfsoverschot. Voor gras en maïs beperken we ons tot het model met Gtgroep en Nmin (voor beiden model 1). In al deze modellen is de helling voor Nmin (en voor Neerslagoverschot danwel Minas-overschot) per gewasgroep gelijk voor de drie Gt-groepen.

Voor opschaling is het belangrijk te constateren dat een deel van de kwantitatieve predictoren direct op bedrijfsniveau worden bepaald (zoals bijvoorbeeld Minas-bedrijfsoverschot) en dat anderen worden bepaald via monsternamen (zoals Nmin). Analyse van afzonderlijke monsters zal in de praktijk niet haalbaar zijn, en is ook niet nodig gezien het feit dat de modellen lineair in de predictoren zijn, dat wil zeggen dat het niet uitmaakt of je met het model eerst nitraatconcentraties op monsterplekken schat en dan het gemiddelde daarvan berekent of juist eerst gemiddelden van de predictoren bepaalt en die dan in het model gebruikt.

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

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

waarin streepje en dakje boven de variabelen aangeven dat het gaat om een geschat (dakje) cluster-gemiddelde (streepje). Het bedrijfsgemiddelde kan dan worden geschat als het gewogen gemiddelde van de cluster-gemiddelden:

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

waarin:

a_j : oppervlaktefractie van cluster j binnen het bedrijf;

J : aantal clusters binnen het bedrijf.

Als we [2] in [3] substitueren krijgen we de meest algemene formule voor het schatten van het bedrijfsgemiddelde:

$$\hat{y} = \sum_{j=1}^J a_j \hat{y}_j = \sum_{j=1}^J a_j (\hat{a}_j + \hat{b}_{1j} \cdot \hat{x}_{1j} + \hat{b}_{2j} \cdot \hat{x}_{2j}) \quad [4]$$

In deze formulering is per cluster een schatting van de gemiddelde Nmin nodig. Echter, omdat de helling voor Nmin in het model gelijk is voor de verschillende Gt-groepen, hoeft alleen het intercept gewogen te worden met de oppervlaktefractie van het cluster. Zo wordt de bedrijfsgemiddelde nitraatconcentratie bij akkerbouwbedrijven volgens model 3 voorspeld met :

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

waarin:

A_j : relatief oppervlak van Gt-groep j in het bedrijf;

\hat{a}_j : geschat intercept voor Gt-groep j in het akkerbouwmodel;

\hat{b}_1 : geschatte regressie-coëfficiënt voor Nmin;

\hat{b}_2 : geschatte regressie-coëfficiënt voor Minas-overschot;

\hat{x}_1 : geschat bedrijfsgemiddelde voor Nmin;

\hat{x}_2 : aangegeven waarde voor Minas-overschot.

De (onbekende) werkelijke waarde van deze concentratie is gelijk aan:

$$\bar{y} = \sum_{j=1}^3 A_j \mathbf{a}_j + \mathbf{b}_1 \bar{x}_1 + \mathbf{b}_2 \bar{x}_2 + \bar{e}$$

waarin:

\bar{e} : het bedrijfsgemiddelde van de residuele afwijkingen op proefplekniveau.

In de huidige modellen is het dus niet noodzakelijk om per Gt-groep een schatting voor Nmin te hebben om te komen tot een goede voorspelling van de bedrijfsgemiddelde nitraatconcentratie. Een goede schatting van de bedrijfsgemiddelde Nmin is voldoende.

6.3 De nauwkeurigheid van voorspellingen van bedrijfsgemiddelde nitraatconcentraties

De nauwkeurigheid van de voorspellingen valt in feite samen met het begrip 'doelgerichtheid' van de indicatoren zoals gedefinieerd in de interne Sturen op Nitraat notitie 'Perspectieven voor indicatoren als hulpmiddel bij het realiseren van de doelstelling van de Europese Nitraat Richtlijn' (31-1-2002). Deze nota geeft aan dat doelgerichtheid, naast meetbaarheid, beïnvloedbaarheid en consistentie, het belangrijkste criterium is voor de bruikbaarheid van de indicatoren. Daarom dient aan de nauwkeurigheid van de voorspellingen, en dus ook de kwantificering daarvan, ruime aandacht te worden besteed.

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 van de clusters 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 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 gekwadraterde fout), welke gelijk is aan het kwadraat van de 'bias' (de systematische fout) plus de variantie (maat voor de toevallige fout):

$$MSE(\hat{y}_j) = [bias(\hat{y}_j)]^2 + Var(\hat{y}_j) \quad [5]$$

Deze vergelijking kan stapsgewijs worden uitgewerkt (zie bijlage 1). Voor de modellen van gras en maïs geldt dat er maar één kwantitatieve predictor (Nmin) 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 met behulp van:

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

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 te middelen binnen clusters en bedrijven. Deze residu-gemiddelden worden dan gekwadeerd en als bias-correctie worden daar de varianties van afgetrokken:

$$\hat{e}^2 = (\hat{\bar{e}})^2 - Var_m(\hat{\bar{e}})$$

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

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

(=de gewogen som van de fouten per cluster, met de kwadraten van de relatieve oppervlakten van de clusters als gewichten).

6.4 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 Nmin (en mogelijke andere verklarende variabelen in het regressie-model). Ook de waarde van Nmin zelf speelt hierin een rol. Door in de toekomst de Nmin anders te bepalen verandert de bijdrage van deze term aan de totale onnauwkeurigheid.
- de naar bedrijfsniveau opgeschaalde restspreiding $E_m(\bar{e}_j^2)$

Deze drie onderdelen kunnen apart worden uitgerekend (zie bijlage 2). Daarmee wordt inzichtelijk gemaakt hoe de totale onnauwkeurigheid is opgebouwd.

6.4.1 Akkerbouw

Voor model 1 en 3, dus voor de regressiemodellen met Gt-groep, Nmin en resp. neerslagoverschot2 of MINASoverschot, worden de drie termen berekend die gezamenlijk de nauwkeurigheid van een clustergemiddelde bepalen. Er is uitgegaan van Nmin=35. De variantie van Nmin is geschat op 3000. Er zijn twee scenario's doorgerekend: er wordt op basis van respectievelijk 10 of 40 steken een mengmonster gemaakt en Nmin bepaald. Er is verder uitgegaan van een gemiddelde neerslag (429 mm voor dit meetseizoen) of MINAS-bedrijfs-overschot (12). De fout rond de gemiddelde neerslag of MINAS is nul verondersteld.

Tabel 21. De drie onderdelen uit vergelijking [6] die gesommeerd de totale variantie van het voorspelde clustergemiddelde geven voor de clusters akkerbouw met Gt-groep = 1, 2 en 3, uitgaande van een Nmin=35 en één extra term in het regressie-model. De variantie van Nmin is bepaald uitgaande van resp. 10 (10Nmin) en 40 steken (40Nmin)

Extra term	Gt-groep	Var. lijn	10 Nmin	40 Nmin	E-kwadraat
Neerslag	1	602,9	209,6	52,4	191,9
	2	593,8	209,6	52,4	191,8
	3	61,4	209,6	52,4	191,8
BOMINAS	1	608,6	219,1	54,8	588,7
	2	657,9	219,1	54,8	588,7
	3	70,9	219,1	54,8	588,7

Tabel 22. De voorspelde nitraatconcentratie (mg/l) per cluster met de bijbehorende se op basis van resp. 10 (Se10) en 40 (Se40) steken voor Nmin-bepaling bij een gestelde Nmin=35 op basis van het regressie-model met neerslagoverschot in de winterperiode of model met MINAS-bedrijfsoverschot

Extra term	Gt-groep	Voorspelling	Se 10	Se 40
Neerslag	1	20,0	31,7	29,1
	2	24,4	31,6	29,0
	3	92,5	21,5	17,5
BOMINAS	1	-12,6	37,6	35,4
	2	20,3	38,3	36,1
	3	95,7	29,6	26,7

6.4.2 Veehouderij

Voor gras en maïs is uitgegaan van het model met alleen Nmin en de Gt-groepen in de regressie-vergelijking. Voor deze twee gewassen zijn de drie termen die gesommeerd de variantie rond de voorspelling bepalen berekend en gegeven in tabel 23. Er is uitgegaan van een Nmin=35. De variantie van Nmin is geschat op 3000. Er zijn twee scenario's doorgerekend: er wordt op basis van resp. 10 of 40 steken een mengmonster gemaakt en Nmin bepaald.

Tabel 23. De drie termen uit vergelijking [6] die gesommeerd de totale variantie van het voorspelde clustergemiddelde geven voor de clusters gras-Gt1, gras-Gt2 etc., uitgaande van een $N_{min}=35$. De variantie van N_{min} is bepaald uitgaande van resp. 10 en 40 steken

Gewas	Gt-groep	Var. lijn	10 N_{min}	40 N_{min}	E-kwadraat
Gras	1	172,9	116,7	29,2	193,6
	2	47,5	116,7	29,2	193,6
	3	48,5	116,7	29,2	193,6
Mais	1	1143	242,6	60,7	392,6
	2	253	242,6	60,7	392,6
	3	164	242,6	60,7	392,6

In tabel 24 staan de bijbehorende voorspellingen van de clustergemiddelde nitraatconcentraties met de berekende se (op basis van die drie termen).

Tabel 24. De voorspelde nitraatconcentratie (mg/l) per cluster met de bijbehorende se op basis van resp. 10 en 40 steken voor N_{min} -bepaling bij een gestelden $N_{min}=35$

Gewas	Gt-groep	Voorspelling	Se (10)	Se (40)
Gras	1	24,3	22,0	19,9
	2	32,4	18,9	16,4
	3	68,8	18,9	16,5
Mais	1	18,7	42,1	40,0
	2	48,7	29,8	26,6
	3	78,8	28,3	24,9

Om vervolgens tot bedrijfsgemiddelde voorspellingen te komen, kan gebruik gemaakt worden van de clustergegevens in de tabellen 21-24. Hieronder volgen enkele rekenvoorbeelden.

Rekenvoorbeeld 1

Een bedrijf teelt alleen gras en het bedrijfsareaal valt geheel binnen Gt-groep 2. Stel dat N_{min} wordt gemeten op basis van één mengmonster van 10 steken en de gemeten N_{min} -waarde is 35. De voorspelde nitraatwaarde is dan gelijk aan 32.4 mg/l NO_3 met een $se=18.9$ (zie tabel 24).

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

$$NO_3 = 32.4 \pm 2 \cdot 18.9 \text{ mg/l}$$

Rekenvoorbeeld 2

Een bedrijf teelt op de ene helft van zijn areaal gras en op de andere helft maïs. De graspercelen vallen allemaal binnen Gt-groep 3 en de maïspcelen voor de helft in Gt-groep 1 en voor de andere helft in Gt-groep 2. Stel er wordt N_{min} gemeten op basis van één mengmonster (van 10 steken) per cluster (voor de drie clusters: gras-Gt-3, mais-Gt-1 en mais-Gt-2). De N_{min} -waarde is voor dit rekenvoorbeeld overall 35.

De voorspelde bedrijfsgemiddelde nitraatwaarde in het grondwater in het voorjaar (zie tabel 24) is nu :

$$NO_3 = 0.5 \cdot 68.8 + 0.25 \cdot 18.7 + 0.25 \cdot 48.7 = 51.2 \text{ mg/l}$$

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

$$se = \sqrt{(0.5^2 \cdot 18.9^2 + 0.25^2 \cdot 42.1^2 + 0.25^2 \cdot 29.8^2)} = 16.0$$

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

$$NO_3 = 51.2 \pm 2 \cdot 16.0 \text{ mg/l}$$

7 Conclusies

Na analyse van de gegevens van één meetseizoen (2000/2001) voor het project Sturen Op Nitraat kunnen alleen voorlopige conclusies worden getrokken. Pas als alle gegevens van drie meetseizoenen verwerkt zijn, zijn de conclusies niet voorlopig meer.

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 regressievergelijkingen 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 N_{min} (gemeten in het najaar), perceeloverschot en bedrijfsoverschot (MINAS en werkelijk berekend) in ogenschouw genomen als mogelijk verklarende variabelen voor de nitraatconcentratie in het voorjaar. Voorafgaand aan de gegevensverzameling is verondersteld dat een aantal factoren hierbij van groot belang zijn, namelijk plaatsgebonden factoren zoals bodem en grondwatertrap, het weer (met name neerslag) en het gewas.

7.1 Regressie-analyse met proefplekgegevens

Op basis van waarnemingen op 466 proefplekken komt in ieder geval 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.

Uit de verzamelde N_{min} -gegevens op 334 proefplekken komt de hoeveelheid N_{min} in de meetperiode oktober-december 2000, gesommeerd over de drie bodemlagen (i.e. 0-90 cm) als meest verklarende variabele naar voren uit de groep aan N_{min} -gegevens. De groep aan N_{min} -gegevens bevat waarnemingen voor drie opeenvolgende bodemlagen en ook op drie verschillende meetperioden. Onderzocht is of er een verklaring kon worden gevonden voor het verloop van die N_{min} -waarnemingen in de tijd en ook in de diepte. Hier is echter middels regressievergelijking geen verklaring voor gevonden. Wel is duidelijk dat met name de metingen uit de eerste meetperiode geschikt lijken als indicator voor nitraatuitspoeling.

Tevens bleek dat deze hoeveelheid N_{min} van de kandidaatindicatoren de hoogste correlatie had met nitraat in het voorjaar, gevolgd door het MINAS-overschot dat vooral op akkerbouwbedrijven een hoge correlatie met nitraat bleek te hebben. Bij de

regressie-analyse zijn een groot aantal aanvullende variabelen meegenomen, waarvoor de veronderstelling geldt dat allen een zekere relatie met nitraatuitspoeling vertonen.

Omdat er duidelijke verschillen waren tussen akkerbouw en melkveehouderij zijn de regressie-analyses uitgevoerd voor akkerbouwgewassen, gras en maïs afzonderlijk. In alle gevallen komen Gt-groep en Nmin als belangrijke verklarende variabelen naar voren. Bij akkerbouw kan vervolgens uit verschillende extra variabelen worden gekozen, die allen een min of meer vergelijkbare verbetering van de regressie-vergelijking bewerkstelligen. Deze variabelen zijn neerslagoverschot in de winterperiode (model 1), bedrijfs- of MINAS-overschot (modellen 2 en 3) of de neerslagsom in de winter in combinatie met potentiële mineralisatie (model 4). Met name de relevantie van neerslag is verrassend, aangezien de gegevens slechts betrekking hebben op één meetperiode en dus nog niet op weersvariatie tussen verschillende jaren.

Voor grasland levert de regressie-analyse naast Gt-groep en Nmin (model 1) de volgende extra verklarende variabelen op: DOC in het grondwater (model 2), aangevuld met de werkzame of totale N-gift via bemesting (modellen 3 en 4). Voor maïs is naast Nmin en Gt-groep (model 1) alleen DOC in het grondwater (model 2) naar voren gekomen als verklarende variabele voor de nitraatconcentratie in het grondwater.

Als de regressie-vergelijkingen worden gebruikt om nitraatconcentraties te voorspellen op puntniveau, zijn de betrouwbaarheidsintervallen nogal groot. Grofweg zijn deze intervallen + of - 92 tot 160 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.

7.2 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 betrouwbaarheidsintervallen voor de voorspelde nitraatconcentratie dan nog slechts + of - 32 tot 38 mg/l zijn. Als de regressie-vergelijkingen gebaseerd zijn op meerdere groeiseizoenen zal deze voorspelfout verder afnemen.

Literatuurlijst

Berge, H.F.M. ten (ed.), 2002. A review of potential indicators for nitrate loss from cropping and farming systems in the Netherlands. Wageningen, Plant Research International Report 31. Reeks Sturen Op Nitraat 2.

Nathan, G., 1988. Inference based on data from complex sample designs. Handbook of Statistics, vol. 6, pp 247-266. Elsevier Science Publishers B.V.

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

Oude Voshaar, J.H., 1994. Statistiek voor onderzoekers. Wageningen, Wageningen Pers.

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

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.

Velthof, G.L., 2003 (in druk). Relaties tussen mineralisatie, denitrificatie en indicatoren voor bodemkwaliteit in landbouwgronden. Wageningen, Alterra. Alterra-rapport 769. Reeks Sturen Op Nitraat 6.

Bijlage 1

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

$$MSE(\hat{y}_j) = [bias(\hat{y}_j)]^2 + Var(\hat{y}_j) \quad [1]$$

We gaan eerst de bias na, daarna de variantie. De bias is gelijk aan:

$$bias(\hat{y}_j) = E_b(\hat{y}_j) - \bar{y}_j \quad [2]$$

De eerste term is de verwachtingswaarde (het gemiddelde) van de voorspelling, de tweede term is het werkelijke (onbekende) clustergemiddelde binnen het bedrijf. (Met het subscript b wordt aangegeven dat het gemiddelde wordt genomen over (denkbeeldige) herhalingen van de bemonstering op het bedrijf.)

Eenvoudigheidshalve gaan we er voor de volgende afleiding van uit dat er maar één kwantitatieve predictor is (bijv. Nmin), en dat de regressiecoëfficiënt daarvan in alle clusters gelijk is. De voorspeller van het clustergemiddelde is dan:

$$\hat{y}_j = \hat{a}_j + \hat{\mathbf{b}} \cdot \hat{x}_j \quad [3]$$

en voor de verwachtingswaarde daarvan geldt:

$$E_b(\hat{y}_j) = E_b(\hat{a}_j + \hat{\mathbf{b}} \cdot \hat{x}_j) = \hat{a}_j + \hat{\mathbf{b}} \cdot E_b(\hat{x}_j) \quad [4]$$

Als we verder aannemen dat we een zuivere schatting hebben van het clustergemiddelde van x , d.w.z. zonder een systematische fout (bijvoorbeeld door ervoor te zorgen dat er geen systematische monsternamen en analysefouten zijn), dan is de verwachtingswaarde:

$$E_b(\hat{y}_j) = \hat{a}_j + \hat{\mathbf{b}} \cdot \bar{x}_j$$

Het werkelijke clustergemiddelde is gelijk aan:

$$\bar{y}_j \equiv \frac{1}{N_j} \sum_{i=1}^{N_j} y_{ij} = \frac{1}{N_j} \sum_{i=1}^{N_j} (\mathbf{a}_j + \mathbf{b}x_{ij} + e_{ij}) = \mathbf{a}_j + \mathbf{b}\bar{x}_j + \bar{e}_j \quad [5]$$

en de bias is dus:

$$bias(\hat{y}_j) = E_b(\hat{y}_j) - \bar{y}_j = (\hat{a}_j + \hat{\mathbf{b}}\bar{x}_j) - (\mathbf{a}_j + \mathbf{b}\bar{x}_j + \bar{e}_j) = (\hat{a}_j - \mathbf{a}_j) + (\hat{\mathbf{b}} - \mathbf{b})\bar{x}_j - \bar{e}_j \quad [6]$$

De twee eerste termen hierin zijn de bijdragen van de schattingsfouten in de regressiecoëfficiënten (merk op dat de tweede mede afhangt van het clustergemiddelde \bar{x}_j). De derde term is het clustergemiddelde van de regressie-residuen, d.w.z. de bijdrage van alle onvolkomenheden in de structuur van het regressie-model.

Voor de variantie van het clustergemiddelde geldt:

$$Var_b(\hat{y}_j) = Var_b(\hat{\mathbf{a}}_j + \hat{\mathbf{b}} \cdot \hat{x}_j) = \hat{\mathbf{b}}^2 \cdot Var_b(\hat{x}_j) \quad [7]$$

(Het subscript b geeft weer aan dat de variantie de bemonstering op het bedrijf betreft.) Als het clustergemiddelde \bar{x}_j wordt geschat uit een mengmonster van n steken, genomen als een ‘enkelvoudige aselelect’ steekproef, dan is de variantie van die schatting:

$$Var_b(\hat{x}_j) = \frac{1}{n} \cdot S_j^2(x) \quad [8]$$

waarin $S_j^2(x)$ de ruimtelijke variantie is van de predictor (bijv. Nmin) tussen plekken in cluster j (binnen het bedrijf). In de praktijk zullen waarschijnlijk efficiëntere steekproefopzetten worden toegepast dan ‘enkelvoudig aselelect’, wat betekent dat de variantie kleiner wordt dan volgens [8].

Voor de MSE krijgen we dus (zie [1]):

$$MSE_b(\hat{y}_j) = [bias(\hat{y}_j)]^2 + Var_b(\hat{y}_j) = [(\hat{\mathbf{a}}_j - \mathbf{a}_j) + (\hat{\mathbf{b}} - \mathbf{b})\bar{x}_j - \bar{e}_j]^2 + \hat{\mathbf{b}}^2 \cdot Var_b(\hat{x}_j) \quad [9]$$

Als we dit uitwerken en daarvan de verwachting nemen over de bemonstering voor de modellering, E_m , (dus over denkbeeldige herhalingen van de bemonstering zoals uitgevoerd voor de regressie-analyse), dan krijgen we voor de totale fout:

$$E_m[MSE_b(\hat{y}_j)] = Var_m(\hat{\mathbf{a}}_j) + Var_m(\hat{\mathbf{b}}) \cdot \bar{x}_j^2 + 2Cov_m(\hat{\mathbf{a}}_j, \hat{\mathbf{b}}) \cdot \bar{x}_j + Var_b(\hat{x}_j) \cdot E_m(\hat{\mathbf{b}}^2) + E_m(\bar{e}_j^2) \quad [10]$$

Hierin zijn de $Var_m()$ en de $Cov_m()$ bekend via de regressie-analyse. Verder zijn hierin te schatten:

$$\bar{x}_j^2 \text{ als: } \hat{x}_j^2 - \text{Var}_b(\hat{x}_j)$$

$$\bar{x}_j \text{ als: } \hat{x}_j$$

$$E_m(\hat{\mathbf{b}}^2) \text{ als: } \hat{\mathbf{b}}^2$$

We kunnen dus de totale fout schatten m.b.v.:

$$E_m[MSE_b(\hat{y}_j)] \cong \text{Var}_m(\hat{\mathbf{a}}_j) + \text{Var}_m(\hat{\mathbf{b}}) \cdot [\hat{x}_j^2 - \text{Var}_b(\hat{x}_j)] + 2\text{Cov}_m(\hat{\mathbf{a}}_j, \hat{\mathbf{b}}) \cdot \hat{x}_j + \text{Var}_b(\hat{x}_j) \cdot \hat{\mathbf{b}}^2 + E_m(\bar{e}_j^2)$$

[11]

Bijlage 2

Voor akkerbouw kunnen de drie termen per Gt-groep als volgt worden berekend:

$$\begin{aligned} \text{term1} = & V_m(\hat{\mathbf{a}}_j) + \hat{x}_1^2 V_m(\hat{\mathbf{b}}_1) + \hat{x}_2^2 V_m(\hat{\mathbf{b}}_2) + 2\hat{x}_1\hat{x}_2 C_m(\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2) \\ & + 2\hat{x}_1 C_m(\hat{\mathbf{a}}_j, \hat{\mathbf{b}}_1) + 2\hat{x}_2 C_m(\hat{\mathbf{a}}_j, \hat{\mathbf{b}}_2) \end{aligned}$$

$$\text{term2} = [\hat{\mathbf{b}}_1^2 - V_m(\hat{\mathbf{b}}_1)]V_b(\hat{x}_1) + [\hat{\mathbf{b}}_2^2 - V_m(\hat{\mathbf{b}}_2)]V_b(\hat{x}_2)$$

$$\text{term3} = \hat{e}^2 - V_b(\hat{e})$$

Opmerking: hierbij is aangenomen dat de regressie-coëfficiënten enerzijds, en de residuele restterm op bedrijfsniveau anderzijds, (bij benadering) ongecorreleerd zijn.

Voor akkerbouw is verondersteld dat de fout rond de gemiddelde neerslag danwel rond het gemiddelde MINAS-bedrijfsoverschot nul is : $V_b(\hat{x}_2) = 0$

Voor gras en mais vallen alle termen met β_2 weg.