

Steekproefopzet regionale nitraatmonitoring

In opdracht van het Ministerie van Landbouw, Natuur en Voedselkwaliteit, Directie
Wetenschap en Kennis

Steekproefopzet regionale nitraatmonitoring

M. Kotters, J.J. de Gruijter, D.J. Brus

Alterra-rapport 1154
Reeks Sturen Op Nitraat 15

Alterra, Wageningen, 2005

REFERAAT

Martin Knotters, Jaap de Gruijter, Dick Brus, 2005. *Steekproefopzet regionale nitraatmonitoring*. Wageningen, Alterra-rapport 1154. Sturen op Nitraat 15. 35 blz.; 1 fig.; 3 tab.; 4 ref.

Om te kunnen beoordelen of het mestbeleid effectief is, is voor regio's jaarlijks inzicht nodig in de oppervlaktefracties waarbinnen de Europese nitraatnorm van 50 mg/l wordt overschreden. Op landelijke schaal is behoefte aan deze informatie per combinatie van bodemtype, grondwatertrap en gewas (clusters). Bovendien is een kwantificering van de nauwkeurigheid van deze informatie gewenst, om te kunnen toetsen of de oppervlaktefracties in de loop van de tijd veranderen. Uit waarnemingen van het nitraatdeel van Nmineraal kan met behulp van regressiemodellen de kans worden berekend dat op een bepaalde locatie de Europese nitraatnorm wordt overschreden. Met een gestratificeerde aselechte steekproef, met de clusters als strata, kan voor elke regio de oppervlaktefractie worden geschat met het ruimtelijk gemiddelde van de kansen dat op locaties de Europese norm wordt overschreden. Op landelijke schaal kunnen de oppervlaktefracties per cluster worden geschat. Met een *t*-toets kan worden beoordeeld of de oppervlaktefracties tussen jaren van elkaar verschillen. Gerichte waarnemingen van de nitraatconcentratie van het ondiepe grondwater in het voorjaar kunnen worden gebruikt om de regressiemodellen te verbeteren. De steekproefopzet kan in de loop van de tijd worden aangepast aan nieuwe informatie en eisen.

Trefwoorden: *design-based sampling strategy*, *stratified random sampling*, Student's *t*-toets, minerale-stikstofgehalte

ISSN 1566-7197

Dit rapport kunt u bestellen door Euro 20,- over te maken op banknummer 36 70 54612 ten name van Alterra, Wageningen, onder vermelding van Alterra-rapport 1154. Dit bedrag is inclusief BTW en verzendkosten.

© 2005 Alterra

Postbus 47; 6700 AA Wageningen; Nederland

Tel.: (0317) 474700; fax: (0317) 419000; e-mail: info.alterra@wur.nl

Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op 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.

Inhoudsopgave

Woord vooraf	7
Samenvatting	9
1 Inleiding	11
1.1 Achtergrond en probleemstelling	11
1.2 Doel van het onderzoek	11
1.3 Opbouw van het rapport	12
2 Ontwerp van een monitoringstrategie	13
2.1 Het <i>sampling scheme</i>	13
2.2 De keuze van een <i>design-based sampling strategy</i>	15
2.2.1 Ontwerp voor de steekproef per jaar	15
2.2.2 Verandering in de tijd	15
2.3 Vereiste steekproefomvang	16
2.4 Relatie tussen nitraatconcentratie en $N_{\min_{\text{nitraat}}}$	18
2.5 Benodigde informatie	18
2.6 Regressievoorspellingen van de nitraatconcentratie	18
2.7 Monitoring van de relatie nitraatconcentratie- $N_{\min_{\text{nitraat}}}$	20
3 Statistische verwerking	21
3.1 Oppervlaktefractie in een regio	21
3.2 Oppervlaktefractie in een cluster	22
3.3 Toets voor het verschil tussen twee jaren	23
4 Veldprotocol	25
Bibliografie	27
Bijlagen	29
A Sampling Scheme	29
B Beslissingsboom	33
C Voorspellingen en varianties voor clusters	35

Woord vooraf

Voor u ligt het rapport ‘Steekproefontwerp regionale nitraatmonitoring’, waarin op basis van de kennis uit het project ‘Sturen op Nitraat’ een strategie voor de monitoring van het mestbeleid is beschreven. Wij zijn Dethmer Boels, Wim Chardon, Mirjam Hack-ten Broeke en Annemieke Smit erkentelijk voor de bereidwilligheid waarmee zij hun kennis beschikbaar stelden. Het project werd gefinancierd door het DWK-programma 396, ‘Bodemkwaliteit: Risicobeoordeling en Risicobeheersing in het Landelijk Gebied’, van het Ministerie van Landbouw, Natuur en Voedselkwaliteit.

Samenvatting

Het project heeft de volgende doelen, in volgorde van prioriteit:

1. Op hoeveel locaties en hoe vaak, per combinatie van bodemtype, grondwatertrap en gewas (cluster) en per regio, en volgens welk steekproefontwerp, moeten $N_{\min_{\text{nitraat}}}$ en de nitraatconcentratie worden gemeten om met een gegeven budget en een vereiste nauwkeurigheid de fractie van de oppervlakte zand- en lössgronden van een regio te kunnen schatten waar de Europese nitraatnorm wordt overschreden?
2. Het volgen en bijstellen van de relatie tussen nitraatconcentratie en $N_{\min_{\text{nitraat}}}$ is onderdeel van het steekproefontwerp.
3. Welk tijdsinterval is minimaal vereist om veranderingen in de oppervlaktefractie nauwkeurig vast te kunnen stellen?
4. Met welke nauwkeurigheid levert dit steekproefontwerp oppervlaktefracties op per cluster op landelijke schaal voor zand- en lössgronden?

De monitoring richt zich op de landbouwgrond in zand- en lössgebieden in het voorjaar in een reeks jaren vanaf 2005, in elk van een aantal nog nader te definiëren regio's met oppervlakten van 10 000-20 000 ha. De doelgrootheid waarop het steekproefontwerp is gebaseerd zijn de ruimtelijke gemiddelden van de kansen dat de Europese nitraatnorm op een locatie wordt overschreden, in elk van de nog nader te definiëren regio's en in de clusters op landelijke schaal. Deze ruimtelijke gemiddelden zijn te interpreteren als oppervlaktefracties waarin de Europese nitraatnorm wordt overschreden. Een gestratificeerde steekproef met clusters als strata en tenminste twee punten per stratum is geselecteerd als meest geschikte strategie om jaarlijks de oppervlaktefracties te schatten. Op de steekproefpunten wordt $N_{\min_{\text{nitraat}}}$ gemeten. De steekproef wordt jaarlijks in het najaar uitgevoerd, waarbij de locaties telkens opnieuw worden geloot. In het voorjaar wordt de nitraatconcentratie van het ondiepe grondwater gemeten op een gerichte selectie van locaties waar in het najaar $N_{\min_{\text{nitraat}}}$ is gemeten. Deze waarnemingen kunnen worden gebruikt om het regressiemodel dat het verband beschrijft tussen $N_{\min_{\text{nitraat}}}$ en de nitraatconcentratie van het ondiepe grondwater bij te stellen.

Jaarlijks wordt voor elke regio de oppervlaktefractie geschat waarin de Europese nitraatnorm wordt overschreden uit het ruimtelijke gemiddelde van de overschrijdingskansen van de nitraatnorm op de steekproefpunten. Deze overschrijdingskansen worden berekend met behulp van de waarnemingen van $N_{\min_{\text{nitraat}}}$ en regressiemodellen die de relatie tussen $N_{\min_{\text{nitraat}}}$ en de nitraatconcentratie van het ondiepe grondwater beschrijven. Daarnaast wordt de variantie van de schattingsfout van de

oppervlaktefracties geschat. De verandering in oppervlaktefractie tussen twee jaren wordt beoordeeld met behulp van Student's twee-steekproeven-toets voor het verschil van twee gemiddelden. Uit de waarnemingen die in de regio's zijn verricht kan ook voor elk jaar per cluster op landelijke schaal een oppervlaktefractie worden berekend waarin de Europese nitraatnorm wordt overschreden.

Het is belangrijk dat de monsternamen van $N_{min_{nitraat}}$ en de nitraatconcentratie van het ondiepe grondwater op dezelfde wijze plaatsvindt als in het project 'Sturen op Nitraat', omdat de regressiemodellen die in de monitoringstrategie worden gebruikt hierop gebaseerd zijn.

Hoofdstuk 1

Inleiding

1.1 Achtergrond en probleemstelling

Om te kunnen beoordelen of het mestbeleid in Nederland effectief is en aan de Europese nitraatrichtlijn voldoet, wordt de nitraatconcentratie van het ondiepe grondwater gemeten. Onderzocht is of gebruik van een indicator voor de nitraatconcentratie bij zand- en lössgronden een alternatief is voor directe meting en dan ook nog tegen lagere kosten. Uit onderzoek van Hack-ten Broeke et al. (2004) komt het nitraatdeel van N_{mineraal} in de bovenste 90 cm van de bodem ($N_{\text{min}_{\text{nitraat}}}$) in combinatie met een clusterindeling naar bodemtype, grondwatertrap en gewas naar voren als indicator voor de nitraatconcentratie. De relatie tussen de nitraatconcentratie van het grondwater en $N_{\text{min}_{\text{nitraat}}}$ is voor elk cluster beschreven met een enkelvoudig lineair regressiemodel.

De monitoring van de nitraatconcentratie richt zich op regio's met zand- of lössgrond en met een oppervlakte van minstens 10 000-20 000 ha. Voor elke regio moet worden geschat hoeveel procent van de oppervlakte niet voldoet aan de Europese nitraatnorm. Op landelijke schaal is echter behoefte aan deze informatie per cluster. Naast oppervlaktefracties is ook behoefte aan informatie over de veranderingen van deze fracties in de tijd, met als doel te kunnen beoordelen of het beleid effect heeft.

Er kan niet van worden uitgegaan dat de relatie tussen de nitraatconcentratie van het grondwater en $N_{\text{min}_{\text{nitraat}}}$ onveranderlijk is in de tijd. Daarom dient het volgen en bijstellen van deze relatie onderdeel te zijn van het monitoringontwerp.

1.2 Doel van het onderzoek

1. Op hoeveel locaties en hoe vaak, per cluster en per regio, en volgens welk steekproefontwerp, moeten $N_{\text{min}_{\text{nitraat}}}$ en de nitraatconcentratie worden gemeten om met een gegeven budget en een vereiste nauwkeurigheid de fractie van de oppervlakte van zand- en lössgronden in een regio te kunnen schatten waar de Europese nitraatnorm wordt overschreden?
2. Het volgen en bijstellen van de relatie tussen nitraatconcentratie en $N_{\text{min}_{\text{nitraat}}}$ is onderdeel van het steekproefontwerp.

3. Welk tijdsinterval is minimaal vereist om veranderingen in de oppervlaktefractie nauwkeurig vast te kunnen stellen?
4. Met welke nauwkeurigheid levert dit steekproefontwerp oppervlaktefracties op per cluster op landelijke schaal voor zand- en lössgronden?

De doelen zijn gerangschikt naar prioriteit. Eerst moeten we dus de oppervlaktefracties van regio's goed kunnen schatten, dan het regressiemodel kunnen bijstellen, vervolgens kunnen beoordelen of er een ontwikkeling in de tijd is en ten slotte oppervlaktefracties voor clusters op landelijke schaal kunnen schatten.

1.3 Opbouw van het rapport

Hoofdstuk 2 beschrijft het ontwerp van de monitoringstrategie. Een belangrijke rol speelt hierbij het *Scheme* uit De Gruijter et al. (prep). In dit hoofdstuk wordt ook geanalyseerd hoe kosten en nauwkeurigheid zich tot elkaar verhouden. In hoofdstuk 3 wordt de statistische verwerking uiteengezet. Tenslotte volgen in hoofdstuk 4 aandachtspunten voor het veldprotocol.

Hoofdstuk 2

Ontwerp van een monitoringstrategie

2.1 Het *sampling scheme*

In deze paragraaf zal aan de hand van De Gruijter et al. (prep) het *sampling scheme* worden uitgewerkt. Bijlage A geeft een toelichting bij het *sampling scheme*.

1. Gedetailleerde analyse en specificatie van het doel:

- (a) *Target universe*: De landbouwgrond in zand- en lössgebieden in het voorjaar in een reeks jaren vanaf 2005, in elk van een aantal nog nader te definiëren regio's met oppervlakten van 10 000-20 000 ha.
- (b) *Domains of interest*: De clusters, die bestaan uit combinaties van bodem, Gt en gewas c.q. bodemgebruik (paragraaf 2.6), en de afzonderlijke jaren.
- (c) *Target variables*: Doelvariabele is de kans dat de Europese nitraatnorm op een bepaalde locatie wordt overschreden. Deze kans volgt voor een proefplek waar $N_{\min_{\text{nitraat}}}$ is gemeten uit het regressiemodel voor het cluster waarbinnen de proefplek valt. Doelvariabele is de overschrijdingskans, en niet de nitraatconcentratie zelf of de 0/1-variabele met waarde 1 als de Europese norm wordt overschreden, omdat het ruimtelijk gemiddelde van deze overschrijdingskans kan worden geïnterpreteerd als de verwachte oppervlaktefractie waarin de nitraatnorm wordt overschreden.
- (d) *Target parameters*: Ruimtelijke gemiddelden van de overschrijdingskansen in de regio's. Daarnaast de ruimtelijke gemiddelden van de overschrijdingskansen in clusters, op landelijke schaal.
- (e) *Target quantities*: De ruimtelijke gemiddelden van de kansen dat de Europese nitraatnorm op een locatie wordt overschreden, in elk van de nog nader te definiëren regio's en in de clusters op landelijke schaal. Deze ruimtelijke gemiddelden zijn te interpreteren als oppervlaktefracties waarin de Europese nitraatnorm wordt overschreden.
- (f) *Type of result*: In eerste instantie kwantitatief, namelijk predictieintervallen van oppervlaktefracties. Bij herhaling in de tijd ook kwalitatief, wanneer getoetst wordt of de oppervlaktefracties op verschillende

tijdstippen van elkaar verschillen, om te kunnen beoordelen of het beleid effect heeft.

2. *Quality measure*: standaardfout van de schattingen.
3. *Constraints*: niet gegeven.
4. *Prior information*:
 - (a) *Sampling frame*: zie paragraaf 2.5;
 - (b) *Miscellaneous information*: zie Hack-ten Broeke et al. (2004);
 - (c) *Model of the variation* of the target variable: niet bekend.
5. Dimensies van de *sample elements*: dit is de omvang van de steken op de proefplekken. Deze moet het zelfde zijn als in het project ‘Sturen op Nitraat’ (Hack-ten Broeke et al., 2004), omdat de regressiemodellen uit dit onderzoek worden gebruikt. Smit et al. (2003) geven een beschrijving van de monsternamen, zie hoofdstuk 4.
6. *Assessment method*: Met het regressiemodel voor het betreffende cluster wordt uit het waargenomen $N_{\text{min}_{\text{nitraat}}}$ -gehalte de nitraatconcentratie van het grondwater voorspeld en een variantie van de voorspelfout geschat, waarmee vervolgens de kans wordt berekend dat op de betreffende proefplek de Europese nitraatnorm wordt overschreden, zie paragraaf 2.6.
7. *Composite sampling*: Uitsluitend binnen proefplekken worden mengmonsters genomen, overeenkomstig Smit et al. (2003) en Hack-ten Broeke et al. (2004), zodat de regressiemodellen uit Hack-ten Broeke et al. (2004) kunnen worden toegepast. Monsters van meerdere proefplekken kunnen niet worden gemengd, omdat er een niet-lineaire transformatie plaatsvindt naar een 0/1-variabele, die 1 is als de norm wordt overschreden en 0 als dit niet het geval is. Deze 0/1-variabele wordt een indicatorvariabele genoemd; dit moet niet worden verward met de indicator $N_{\text{min}_{\text{nitraat}}}$.
8. Keuze tussen *design-based* en *model-based inference*: Er wordt gekozen voor een *design-based inference*, omdat we antwoord willen op een ‘hoeveel’-vraag en niet op een ‘waar’-vraag, en omdat een model-onafhankelijke kwantificering van de nauwkeurigheid gewenst is.
9. Keuze voor een *design type*. Zie paragraaf 2.2.1.
10. Protocol voor veldwerk: zie hoofdstuk 4. Belangrijk is dat de bemonstering van $N_{\text{min}_{\text{nitraat}}}$ plaatsvindt overeenkomstig ‘Sturen op Nitraat’, zoals beschreven is door Smit et al. (2003), zie punt 5.
11. Methode voor *statistical inference*: zie hoofdstuk 3.
12. Voorspelling van de operationele kosten en kwaliteit van de resultaten: zie paragraaf 2.3.

2.2 De keuze van een *design-based sampling strategy*

2.2.1 Ontwerp voor de steekproef per jaar

Aan de hand van De Gruijter et al. (prep) wordt een keuze gemaakt voor een steekproefstrategie. Hierbij wordt de beslissingsboom toegepast, die is weergegeven in bijlage B. Wij beantwoorden nu de vragen van de deze beslissingsboom.

- (vraag 1) Is er een hulpvariabele die tegen lage kosten kan worden gemeten en die sterk gecorreleerd is met de doelvariabele? Nee, er is geen hulpvariabele.
- (vraag 5) Is het praktisch uitvoerbaar en niet te duur om bemonstering uit te stellen, te hervatten en herhaald waar te nemen, i.e., *sequential sampling*? Nee, *sequential sampling* is geen optie omdat we een regressiemodel gebruiken dat gebaseerd is op waarnemingen in het najaar.
- (vraag 6) Is de ruimtelijke bedekking erg belangrijk (verwacht je meer groot-schalige dan kleinschalige ruimtelijke variatie)? Nee (gezien de grote variatie op korte afstand), maar clustering moet wel worden vermeden, om een zo goed mogelijke schatting van oppervlaktefracties te krijgen.
- (vraag 10) Is reistijd (maar niet navigatie) een belangrijk aspect? Nee.
- (vraag 11) Is navigatie (maar niet reistijd) een belangrijk aspect? Nee.
- (vraag 12) Zijn er significante gradiënten in het gebied, met een bekende en min of meer constante richting? Nee.
- (vraag 13) Zijn zowel reistijd als navigatie belangrijke aspecten, en kun je je tenminste een gemiddelde steekproefomvang veroorloven ($n > 30$)? Nee.
- (vraag 14) Kan op basis van voorinformatie het gebied in deelgebieden worden verdeeld die duidelijk verschillen in gemiddelde, variatie of kosten van bemonstering of meting? Ja: Gestratificeerde steekproef met domeinen (clusters) als strata en tenminste twee punten per stratum. De clusters zijn de combinaties van bodemtype, grondwatertrap en gewas, zie paragraaf 2.6.

2.2.2 Verandering in de tijd

Het is niet wenselijk om op een volgend tijdstip op dezelfde locaties terug te keren. Je verliest weliswaar aan efficiëntie bij het schatten, maar je hebt wel volledige vrijheid om het steekproefontwerp tussentijds aan te passen. Dat kan nodig zijn als bijvoorbeeld regelgeving wijzigt, de resultaten van de steekproef er aanleiding toe geven, of nieuwe kennis gebruikt kan worden over nitraat in bodem en grondwater. Het interval tussen herhaalde steekproeven is een of meerdere jaren, omdat het regressiemodel is gebaseerd op waarnemingen van $N_{\text{min}_{\text{nitraat}}}$ in het najaar.

Het is uit praktisch oogpunt wèl wenselijk om elk jaar bij dezelfde boer terug te keren. Een tweetrapssteekproef met stratificatie in de tweede stap is dan een optie. Bij volgende jaren wordt dan teruggekeerd bij de bedrijven die in het eerste jaar hun medewerking verleenden. Dit heeft echter twee nadelen. In de eerste plaats ontstaan er statistische problemen bij het toetsen van verschillen tussen jaren, omdat er een

afhankelijkheid wordt gecreëerd die niet eenvoudig te verdisconteren is. Daarom kan het beste gekozen worden voor een steekproefopzet waarbij de insluitingskansen van steekproefpunten onveranderlijk blijven in de tijd. In de tweede plaats leidt een tweetrapssteekproef tot ruimtelijke clustering van de steekproefpunten, terwijl zoveel mogelijk ruimtelijke spreiding van de steekproefpunten moet worden nagestreefd omdat het doel immers het schatten van een oppervlaktefractie is. Weliswaar kan het effect van clustering worden ‘gecompenseerd’ door meer steekproefpunten te loten, maar het is de vraag of kosten en nauwkeurigheid dan nog wel goed in balans zijn.

Het verschil in oppervlaktefractie dat tussen twee opeenvolgende steekproeven wordt gevonden kan op ‘toeval’ berusten, omdat we immers onzeker zijn over de werkelijke oppervlaktefracties. Met een toets kan worden beoordeeld of het verschil significant is. Student’s twee-steekproeventoets is hiervoor geschikt.

2.3 Vereiste steekproefomvang

De steekproefomvang wordt bepaald door een kostencriterium en/of een kwaliteitscriterium. Omdat deze criteria nog niet zijn vastgesteld, wordt nu een analyse gemaakt van de verhouding tussen kosten en kwaliteit, op basis van de kennis die nu aanwezig is ten aanzien van de kosten voor bemonstering en de ruimtelijke variatie van $N_{\min_{\text{nitraat}}}$. Als uitgangspunt zijn de gegevens genomen die zijn gerapporteerd door Hack-ten Broeke et al. (2004). De 34 ontwikkelbedrijven zijn tezamen representatief geacht voor één regio. Uitsluitend de bodemtypen ‘Z1’, ‘Z2’ en ‘Z3’ (zandgronden) zijn beschouwd; lössgronden zijn in deze analyse buiten beschouwing gebleven gezien hun geringe oppervlakte. Uit de gegevens van de ontwikkelbedrijven werd dus een denkbeeldige regio in het zandgebied gevormd.

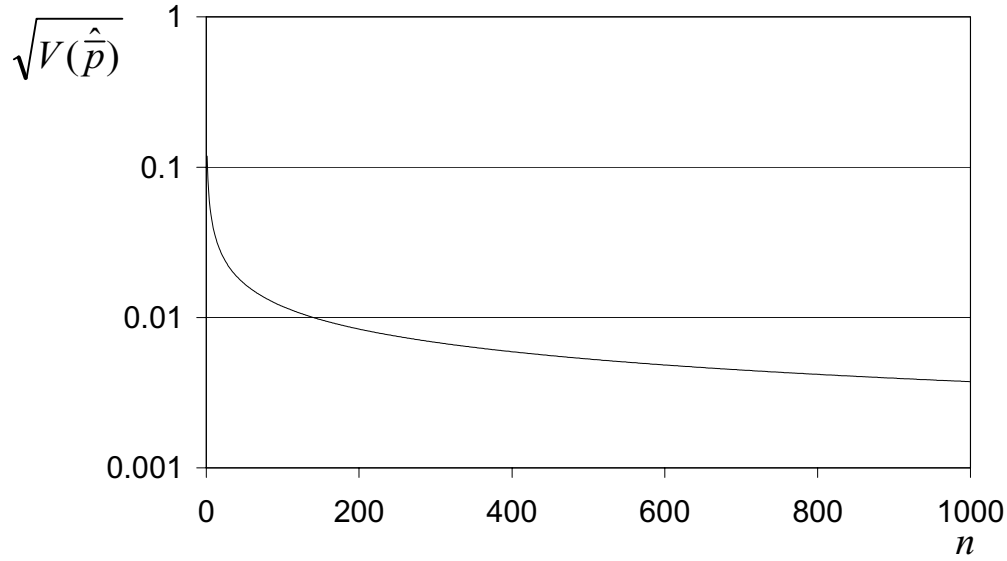
Met de modellen uit tabel 2.1 tot en met 2.3 zijn voorspellingen gedaan voor de nitraatconcentraties in het jaar 2001, op basis van de $N_{\min_{\text{nitraat}}}$ -concentraties die in het najaar van 2000 zijn gemeten. In 2000/2001 lagen de concentraties tussen de relatief lage niveaus in 2001/2002 en de relatief hoge niveaus in 2002/2003. De variantie van de voorspelfout voor een locatie 0 wordt geschat met

$$V(\hat{y}_0) = S^2(x) + \frac{S^2(x)}{n} + S^2 \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2}, \quad (2.1)$$

waarin $\hat{y}$ de voorspelde nitraatconcentratie is, x $N_{\min_{\text{nitraat}}}$, S^2 de restvariantie van het model voor akkerbouw, grasland of maïs, $\bar{x}$ het gemiddelde van $N_{\min_{\text{nitraat}}}$ in akkerbouw, grasland of maïs, en $x_i, i = 1 \dots n$ alle waarnemingen van $N_{\min_{\text{nitraat}}}$ in akkerbouw, grasland of maïs. Alle grootheden in vergelijking (2.1) zijn geschat op basis van de dataset voor drie meetseizoenen uit de studie van Hack-ten Broeke et al. (2004). Vervolgens wordt voor elke proefplek de overschrijdingskans van de Europese norm van 50 mg/l, p , berekend. Hierbij wordt verondersteld dat de voorspelfout van de nitraatconcentratie normaal verdeeld is, met op locatie 0 gemiddelde $\hat{y}_0$ en variantie $V(\hat{y}_0)$.

Voor elk stratum wordt de oppervlaktefractie waarin de Europese nitraatnorm wordt overschreden geschat met

$$\hat{p} = \frac{1}{n_h} \sum_{i=1}^{n_h} p_i,$$



Figuur 2.1. Nauwkeurigheid van de voorspelde oppervlaktefractie waarin de Europese nitraatnorm is overschreden, uitgezet tegen de steekproefomvang, voor een regio met zandgronden.

en de ruimtelijke variantie van p met

$$\widehat{S}^2(p) = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (p_i - \hat{p})^2 .$$

Vervolgens wordt de samenhang tussen het totale aantal steekproefpunten en de nauwkeurigheid van de voorspelling berekend met behulp van de volgende formule:

$$n = \frac{1}{V(\hat{p})} \cdot \sum_{h=1}^H \left(a_h \widehat{S}^2(p) \sqrt{c_h} \cdot \sum_{h=1}^H \frac{a_h \widehat{S}^2(p)}{\sqrt{c_h}} \right) , \quad (2.2)$$

(De Gruijter et al., prep). Hierin zijn a_h de stratumgewichten, die zijn geschat uit het aantal proefplekken in een bepaald cluster in 2000, volgens de studie van Hack-ten Broeke et al. (2004). De varianties $\widehat{S}^2(p)$ zijn geschat uit de gegevens van Hack-ten Broeke et al. (2004) voor clusters met twee of meer proefplekken; bijlage C geeft een overzicht van de resultaten. De kosten c_h voor het nemen van een monster in stratum h zijn niet bekend. In vergelijking 2.2 zijn echter alleen de verhoudingen tussen de kosten in verschillende strata van belang. Verondersteld is dat de verhouding in kosten van het nemen van een monster in grasland of bouwland gelijk is aan 1.5:1 (i.c. 45:30 Euro). In grasland worden zes steken per proefplek genomen en in bouwland drie; omdat de kosten voor het opsporen van de locatie en de kosten voor de chemische analyse van het mengmonster gelijk zijn, zal de kostenverhouding minder bedragen dan 2:1. Figuur 2.1 geeft de samenhang tussen voorspelnauwkeurigheid en steekproefomvang grafisch weer.

Uit de analyse blijkt verder dat bij ca. 58 % van de zandgronden binnen de 34 ontwikkelbedrijven de Europese nitraatnorm werd overschreden in het voorjaar van 2001. Overigens blijkt uit bijlage C een grote spreiding tussen de clusters, van 23 tot 99 % overschrijding.

2.4 Relatie tussen nitraatconcentratie en $N_{\min_{\text{nitraat}}}$

De selectie van punten voor het volgen en bijstellen van de relatie tussen nitraatconcentratie in het grondwater en $N_{\min_{\text{nitraat}}}$ kan het beste gericht plaatsvinden, om een zo groot mogelijke spreiding te krijgen over het bereik van de concentraties en gehalten. Aangezien $N_{\min_{\text{nitraat}}}$ in het najaar wordt gemeten, kan uit deze metingen gericht een subset worden geselecteerd met een grote spreiding over de clusters en de range van $N_{\min_{\text{nitraat}}}$. In het voorjaar kan dan op de geselecteerde locaties de nitraatconcentratie van het grondwater worden gemeten.

2.5 Benodigde informatie

Voorafgaand aan de steekproef moeten de volgende gegevens beschikbaar zijn voor een regio:

1. bodemtype volgens de bodemkaart, schaal 1 : 50 000;
2. aanwezigheid van veenlaagjes;
3. grondwatertrap volgens de bodemkaart, schaal 1 : 50 000;
4. grondgebruik: gras, maïs, akkerbouw. Bij akkerbouw moet bovendien het gewas bekend zijn.

Voor elk steekproefpunt moet $N_{\min_{\text{nitraat}}}$, het nitraatdeel van $N_{\min}$ gemeten in de bodem in het najaar (oktober-december) voor de laag 0-90 cm (kg/ha), worden bepaald.

2.6 Regressievoorspellingen van de nitraatconcentratie

Op de steekproefpunten wordt $N_{\min_{\text{nitraat}}}$ in het najaar waargenomen. Uit deze waarnemingen wordt voorspeld wat de nitraatconcentratie in het grondwater zal zijn in het voorjaar. Hack-ten Broeke et al. (2004) geven regressiemodellen die voor clusters (combinaties van bodemtype, grondwatertrap en gewas) de relatie tussen $N_{\min_{\text{nitraat}}}$ en de nitraatconcentratie in het grondwater beschrijven, op het niveau van proefplekken, percelen en bedrijven. Voor de regionale monitoring zijn de modellen op proefplekniveau van belang, omdat voor proefplekken (steekproefpunten) de nitraatconcentratie wordt voorspeld.

Het regressiemodel dat Hack-ten Broeke et al. (2004) selecteerde voor akkerbouw, grasland en maïs heeft de volgende structuur:

$$Y = b_0 + b_1 X + \varepsilon ,$$

waarin:

Y de nitraatconcentratie van het grondwater is (mg/l);

$X = N_{\min_{\text{nitraat}}}$, het nitraatdeel van $N_{\min}$ gemeten in de bodem in het najaar (oktober-december) voor de laag 0-90 cm (kg/ha);

Tabel 2.1. Regressiemodellen op proefplekniveau voor akkerbouw (Hack-ten Broeke et al., 2004).

parameter	schatting			standaardfout			
b_1	0.69			0.06			
$R^2_{\text{adj.}}$	36%						
Sd	59.6 mg/l						
parameter	schatting						
	gewas	a+b+r			t		
	Gt-groep bodem	1	2	3	1	2	3
b_0	Löss			28.5			
	Z1		27.4	42.5		94.1	109.2
	Z2	0.0	47.8	62.9	66.7	114.5	
	Z3	-4.3	43.5	48.6	62.4	110.2	125.3

Tabel 2.2. Regressiemodellen op proefplekniveau voor grasland (Hack-ten Broeke et al., 2004).

parameter	schatting			standaardfout		
b_1	0.65			0.06		
$R^2_{\text{adj.}}$	21%					
Sd	49.8 mg/l					
parameter	schatting					
	Gt-groep bodem	1	2	3		
b_0	Z1			32.3		37.9
	Z2		11.0	22.9		28.4
	Z3		19.0	30.9		36.4

b_0 en b_1 regressiecoëfficiënten zijn;
 ε het residu is.

De geschatte parameterwaarden voor akkerbouw zijn samengevat in tabel 2.1, voor grasland in tabel 2.2 en voor maïs in tabel 2.3.

Voor een locatie waar $N_{\text{min}_{\text{nitraat}}}$ is gemeten kan de nitraatconcentratie als volgt worden voorspeld:

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x .$$

De waarden van $\hat{b}_0$ en $\hat{b}_1$ zijn gegeven in tabel 2.1 tot en met tabel 2.3. De variantie van de voorspelfout op locatie 0 kan worden berekend met

$$V(\hat{y}_0) = S^2(x) + \frac{S^2(x)}{n} + S^2 \frac{(x_0 - \bar{x})^2}{\sum (x_i - \bar{x})^2} ,$$

waarin $S^2(x)$ de restvariantie van het model voor akkerbouw, grasland of maïs is (zie tabel 2.1-2.3), $\bar{x}$ het gemiddelde van $N_{\text{min}_{\text{nitraat}}}$ voor akkerbouw, grasland of maïs en $x_i, i = 1 \dots n$ de waarden voor $N_{\text{min}_{\text{nitraat}}}$ op alle locaties in akkerbouw,

Tabel 2.3. Regressiemodellen op proefplekniveau voor maïs (Hack-ten Broeke et al., 2004).

parameter	schatting	standaardfout		
b_1	0.76	0.13		
$R^2_{\text{adj.}}$	22%			
Sd	65.6 mg/l			
parameter	schatting			
	Gt-groep bodem	1	2	3
b_0	Z1	24.1	21.8	66.3
	Z2	22.9	20.6	65.1
	Z3	17.4	15.1	59.6

grasland of maïs. Verondersteld wordt dat de voorspelling normaal verdeeld is met gemiddelde $\hat{y}_0$ en variantie $V(\hat{y}_0)$. De kans dat de Europese norm van 50 mg/l wordt overschreden kan nu worden berekend:

$$p = \Pr(Y > r) ,$$

waarbij r de Europese nitraatnorm van 50 mg/l is en Y normaal verdeeld is met gemiddelde $\hat{y}_0$ en variantie $V(\hat{y}_0)$.

2.7 Monitoring van de relatie nitraatconcentratie- Nmin_{nitraat}

Nadat in het najaar Nmin_{nitraat} is gemeten, kan in het daaropvolgende voorjaar de nitraatconcentratie worden gemeten op een aantal geselecteerde locaties die de *range* van concentraties zo volledig mogelijk bestrijken, zie paragraaf 2.4. De modellen uit tabel 2.1 tot en met 2.3 kunnen vervolgens worden verbeterd door de dataset uit Hack-ten Broeke et al. (2004) uit te breiden met de nieuwe data en de modellen opnieuw te kalibreren. Als de modellen veel veranderen kan dit aanleiding zijn om de analyses uit paragraaf 2.3 opnieuw uit te voeren met de nieuwe gegevens en de steekproefomvang voor het volgende jaar aan te passen.

Hoofdstuk 3

Statistische verwerking

3.1 Oppervlaktefractie in een regio

We gaan in eerste instantie uit van een enkelvoudige aselechte steekproef. Op steekprofeenheid i is de variabele $I_i = 1$ als de nitraatnorm wordt overschreden en $I_i = 0$ als dat niet het geval is. De doelgrootheid is $\bar{I}$, namelijk de oppervlaktefractie waarin de Europese nitraatnorm wordt overschreden.

Op de steekproefpunten is I niet foutloos gemeten, maar voorspeld met een regressiemodel. We kunnen voor elk steekproefpunt de kans berekenen dat de nitraatnorm op dat punt wordt overschreden. De kans dat de nitraatnorm wordt overschreden op de locatie van steekproefpunt i , p_i , is de verwachte waarde van I_i , $E(I_i)$. De kans p_i kan worden berekend met het regressiemodel voor de relatie tussen $N_{\text{min}_{\text{nitraat}}}$ en de nitraatconcentratie in het grondwater. We veronderstellen dat de voorspelfouten van het regressiemodel ruimtelijk onafhankelijk zijn. Dit is aannemelijk, omdat uit de studie van Hack-ten Broeke et al. (2004) blijkt dat $N_{\text{min}_{\text{nitraat}}}$ op korte afstand grote variatie vertoont. De schatter voor onze doelgrootheid $\bar{I}$ is

$$\hat{p} = \frac{1}{n} \sum_{i=1}^n p_i, \quad (3.1)$$

waarin p_i de kans is dat op een locatie i de nitraatnorm wordt overschreden. $\hat{p}$ is dus de geschatte fractie van de oppervlakte waarbinnen de nitraatnorm wordt overschreden. De vraag is nu wat de variantie van de schattingsfout is: wat is $V(\hat{p})$? Voor de variantie van de schattingsfout geldt:

$$V(\hat{p}) = \frac{S^2(p)}{n}, \quad (3.2)$$

waarin $S^2(p)$ de ruimtelijke variantie van p is. De variantie $V(\hat{p})$ kan worden geschat met

$$\widehat{V}(\hat{p}) = \frac{1}{n(n-1)} \sum_{i=1}^n (p_i - \hat{p})^2. \quad (3.3)$$

We breiden nu de schatting van de oppervlaktefractie uit voor de situatie van een gestratificeerde aselechte steekproef. De verwachte oppervlaktefractie waarin de norm

wordt overschreden wordt dan geschat met:

$$\hat{\bar{p}} = \sum_{h=1}^L W_h \hat{p}_h, \quad (3.4)$$

waarin $h, h = 1, \dots, L$ de strata zijn, W_h de stratungewichten, en $\bar{p}_h$ de gemiddelden per stratum, die worden berekend met

$$\hat{p}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} p_i,$$

waarin n_h het aantal steekproefpunten is in stratum h . De variantie van $\hat{\bar{p}}$ wordt geschat met

$$\widehat{V}(\hat{\bar{p}}) = \sum_{h=1}^L W_h^2 \widehat{V}(\hat{p}_h), \quad (3.5)$$

waarbij $\widehat{V}(\hat{p}_h)$ voor elk stratum wordt berekend met

$$\widehat{V}(\hat{p}_h) = \frac{1}{n_h(n_h - 1)} \sum_{i=1}^{n_h} (p_i - \hat{p})^2. \quad (3.6)$$

Vergelijking (3.4) levert een schatting van de doelgrootheid ‘oppervlaktefractie van een regio waarin de Europese nitraatnorm wordt overschreden’. De schattingsfout hiervan wordt gekwantificeerd met de variantie die met vergelijking (3.5) wordt geschat.

3.2 Oppervlaktefractie in een cluster

Naast de oppervlaktefractie van een regio, is ook de oppervlaktefractie landbouwgrond waarin de nitraatnorm wordt overschreden, in een cluster op landelijke schaal voor zand- en lössgronden, een doelgrootheid van de monitoring. Deze fractie wordt geschat uit de waarden voor $\hat{p}_h$ in elk van de regio’s, gewogen naar de oppervlakten die de clusters in de regio’s innemen. Als we voor cluster h de verwachte oppervlaktefractie op landelijke schaal $\hat{p}_H$ noemen, dan kan deze uit k regio’s die tezamen landsdekkend zijn als volgt worden berekend:

$$\hat{p}_H = \frac{1}{A_H} \sum_{i=1}^k A_{hi} \hat{p}_{hi}, \quad (3.7)$$

waarin $A_H = \sum_{i=1}^k A_{hi}$. De variantie van de schattingsfout wordt geschat met

$$\widehat{V}_p(\hat{p}_H) = \sum_{i=1}^k \left(\frac{A_{hi}}{A_H} \right)^2 \widehat{V}_p(\hat{p}_{hi}), \quad (3.8)$$

waarin $\widehat{V}_p(\hat{p}_{hi})$ berekend is met vergelijking 3.6 voor de $i = 1, \dots, k$ regio’s.

3.3 Toets voor het verschil tussen twee jaren

Om te kunnen beoordelen of het beleid in een regio effect heeft wordt het verschil in $\hat{p}$ tussen twee opeenvolgende jaren getoetst met de twee-steekproeventoets van Student. De nulhypothese H_0 luidt:

$$H_0 : \bar{p}_{(1)} = \bar{p}_{(2)} ,$$

waarbij (1) en (2) de twee jaren aangeven. Uit de steekproeven die in beide jaren hebben plaatsgevonden kunnen we $\hat{p}_{(1)}$ en $\hat{p}_{(2)}$ berekenen en varianties $\hat{V}(\hat{p}_{(1)})$ en $\hat{V}(\hat{p}_{(2)})$. Het verschil $\bar{p}_{(1)} - \bar{p}_{(2)}$ is te schatten met $\hat{p}_{(1)} - \hat{p}_{(2)}$. Deze grootheid is normaal verdeeld met verwachting $\bar{p}_{(1)} - \bar{p}_{(2)}$ en variantie $V(p_{(1)}) + V(p_{(2)})$, die kan worden geschat met $\hat{V}(\hat{p}_{(1)}) + \hat{V}(\hat{p}_{(2)})$. Als toetsingsgrootheid nemen we:

$$t = \frac{\hat{p}_{(1)} - \hat{p}_{(2)}}{sed} ,$$

waarin *sed* de *standard error of differences* is, de standaardfout van de verschillen $\hat{p}_{(1)} - \hat{p}_{(2)}$. Deze wordt benaderd met

$$sed = \hat{V}(\hat{p}_{(1)}) + \hat{V}(\hat{p}_{(2)}) .$$

Het effectieve aantal vrijheidsgraden te schatten voor de steekproeven in jaar (1) en (2). Dit kan als volgt:

$$n_e = \frac{\left(\sum_{h=1}^L g_h \widehat{S}_h^2 \right)^2}{\sum_{h=1}^L \frac{g_h^2 \widehat{S}_h^4}{n_h - 1}} , \quad (3.9)$$

(Cochran, 1977), met

$$g_h = \frac{N_h(N_h - n_h)}{n_h} ,$$

waarin N_h de totale populatieomvang in stratum h is. Voor oneindige populaties kan g_h in vergelijking (3.9) worden benaderd met een heel groot getal. Het aantal vrijheidsgraden voor toetsingsgrootheid t kan worden benaderd met $n_{e,(1)} + n_{e,(2)}$.

Hoofdstuk 4

Veldprotocol

Dit hoofdstuk bevat de aandachtspunten voor het veldprotocol. Omdat er nog geen keuze is gemaakt ten aanzien van het aantal steekproefpunten, de ligging van de regio's nog niet bekend is en de punten nog niet geloot zijn, kan het veldprotocol niet volledig zijn. In elk geval moet een veldprotocol het volgende bevatten:

1. een lijst met voor elke bodem-gt-gewascombinatie (cluster) de locaties van de steekproefpunten. Het is veilig om per cluster een aantal reservepunten te loten, waar kan worden bemonsterd als een punt vervalt omdat er bijvoorbeeld geen toestemming door de grondgebruiker wordt verleend;
2. met behulp van GPS wordt de locatie in het veld opgezocht. Er mag niet worden afgeweken van de locatie die de GPS aangeeft. Indien er op de aangegeven locatie geen monster kan worden genomen omdat deze in een sloot of op een verharding valt, dan dient te worden uitgeweken naar een reservepunt;
3. op de gelote locatie worden steekmonsters genomen binnen een oppervlakte van 5×5 m, volgens het circeldesign dat is beschreven door Smit et al. (2003). In akkers wordt driemaal per locatie gestoken, in grasland zesmaal. In bouwland wordt gestoken tot 90 cm met een ronde gutsboor, in drie stappen van 30 cm, waarbij de diameter van de guts afneemt met de diepte: 0-30 cm 30 mm, 30-60 cm 25 mm, en 60-90 cm 20 mm. Dit is om vermenging te voorkomen. Vervolgens worden de drie steken per diepte gemengd, zodat er drie mengmonsters per locatie worden gevormd. Het gehalte $N_{\min_{\text{nitraat}}}$ in kg/ha volgt uit de som van de gehalten die voor de drie dieptes zijn bepaald. In grasland wordt met een mechanische, hoekige guts gestoken met binnenafmetingen van 12×15 mm. Er wordt in één keer tot 90 cm diepte gestoken. Uit de zes steken wordt één mengmonster samengesteld;
4. de locaties van de steken ligt vast door het cirkeldesign van Smit et al. (2003). Hier mag niet van worden afgeweken;
5. op geselecteerde locaties wordt een bepaling van de nitraatconcentratie van het ondiepe grondwater verricht, op de wijze die is beschreven door Smit et al. (2003).

Bibliografie

Cochran, W. (1977). *Sampling Techniques*. Wiley, New York.

de Gruijter, J., Brus, D., Bierkens, M., en Knotters, M. (in prep.). *Sampling for Natural Resource Monitoring*. Springer, Heidelberg.

Hack-ten Broeke, M., Burgers, S., Smit, A., ten Berge, H., de Gruijter, J., Hoving, I., Knotters, M., Radersma, S., en Velthof, G. (2004). Ontwikkeling van een indicator om te Sturen Op Nitraat; Gegevens en regressieanalyse op basis van drie meetseizoenen (2000-2001, 2001-2002 en 2002-2003). 1053, Alterra.

Smit, A., Hack-ten Broeke, M., ten Berge, H., Burgers, S., Chardon, W., van Enkevort, P., de Gruijter, J., Hoving, I., en Velthof, G. (2003). Gegevensverzameling Sturen Op Nitraat. Op zoek naar een indicator. , Alterra.

Bijlage A

Sampling Scheme

Bron: de Gruijter et al. (prep)

1. Detailed analysis and specification of the *objective*.
 - (a) *Target universe*: a precise definition of the universe of interest, with boundaries in space and/or time, and possibly a specification of exclusions. (Note that, for various reasons, the target universe may differ from the actually *sampled universe*.)
 - (b) *Domain(s)* of interest: a specification of the part(s) of the universe for which a separate result is needed. At one extreme (in terms of extent), this is the entire universe, at the other extreme it is a point or a set of points in the universe (e.g., grid points used to prepare a contour map). In between these extremes, there may be a number of domains with smaller or larger extents in space and/or time. Examples include sub-areas within a spatial universe, or spatial cross-sections through a space–time universe (i.e., the space at given points in time).
 - (c) *Target variable(s)*: a precise definition of the variable(s) to be determined for each of the sample elements. (Note that target variables are generally not identical with actually *measured variables*, because of measurement errors or transformations of measured variables prior to statistical inference.)
 - (d) *Target parameter*: the type of statistic for which a result is needed, given the target variable(s) and the domain(s). Examples include total, mean, fraction, median, standard deviation or trend parameter.
 - (e) *Target quantity*: the combination of a domain, target variable and target parameter is referred to in this book as a target quantity. An example is the mean (parameter) phosphate content in the topsoil (target variable) of the agricultural soils in the Netherlands (domain). A target quantity that is related to the entire universe is referred to as a ‘global quantity’; in all other cases it is referred to as a ‘local quantity’.
 - (f) *Type of result*: quantitative or qualitative. If a quantitative result is desired, then the mode of inference has to be *estimation* or *prediction*. If a qualitative result is needed, e.g., an answer to the question whether or not the target quantity exceeds a given level, then the mode of inference has to be *testing* or *classification*.

2. *Quality measure*: the quantity used to express numerically the statistical quality of the survey or monitoring result. Examples include the half-width of a 95% confidence interval in estimation, the error variance in prediction, the power in hypothesis testing, the error rate in classification.
3. *Constraints*: the allocated budget and/or minimum quality of the result, field-work (optional), transport (optional) and laboratory capacity (optional).
4. *Prior information*.
 - (a) *Sampling frame*: the list, file or map identifying the elements of the target universe from which the sample elements are selected.
 - (b) *Miscellaneous information*: general knowledge, experience and information from comparable projects, existing sample data, maps or GIS files.
 - (c) *Model of the variation* of the target variable within the universe (eventually needed if model-based inference is chosen; see item 8). Examples include a variogram adopted from a comparable case or estimated from a preliminary sampling round.
5. Dimensions of the *sample elements*, in the case of a continuous universe; physical sampling devices for taking aliquots (optional).
6. *Assessment method*: field and/or laboratory measurement procedures (reference to existing protocols where possible); method of data pre-processing to calculate the target variable from measured variables (optional).
7. Whether and how to use *composite sampling*, i.e., bulking aliquots.
8. Choice between *design-based* and *model-based inference* from sample data.
9. For design-based inference: choice of probabilistic sampling *design type* and *attributes* of the chosen design type.
10. For model-based inference: choice of deterministic *sample type* and *optimization* algorithm and restrictions (optional).
11. Identification of the actually selected *sample*: a list of element labels, a map with sample locations, a table of sample times or coordinates of sample points in space-time.
12. *Protocols* on data recording and field work.
13. Method to be used for *statistical inference*.
14. *Prediction* of operational *costs* and *quality* of results.

The scheme item ‘target parameter’ deserves more attention before we continue with examples of a scheme. Parameters may be related to either a frequency distribution or a probability distribution. Frequency distributions are integrals over space and/or time, which can of course only be relevant in the case of non-point inference elements. Parameters usually related to frequency distributions are the total, mean, mode, standard deviation, percentiles (e.g., the median) and fractions. (A fraction is the relative size of that part of the domain where a given state is present, such as exceedance of a threshold value). Examples include an areal fraction, a temporal

mean and a spatio-temporal total. We refer to such parameters as ‘frequential parameters’. Probability distributions, on the other hand, are integrals of probabilities defined by a stochastic model. Parameters usually related to probability distributions are the expected value, mode, standard deviation, percentiles (e.g., the median) and probabilities of exceedance. Such parameters are referred to as ‘probabilistic parameters’. They may be defined for the target variable at a given point of the universe but also for frequential parameters, which unfortunately complicates matters. For instance, the chosen parameter may be the probability that the spatio-temporal total of the target variable over a given domain exceeds a given threshold value.

Bijlage B

Beslissingsboom

Bron: de Gruijter et al. (prep)

Question 1: Can you measure an ancillary variable at low cost that is strongly correlated with the target variable (say $|r| > 0.7$)?

Yes: Question 2

No: Question 5

Question 2: Is this ancillary variable qualitative?

Yes: Question 3

No: Question 4

Question 3: Are the sizes of the strata known?

Yes: Use post-stratification estimator

No: Two-phase sampling for stratification

Question 4: Is the mean of the ancillary variable known?

Yes: Use regression-estimator

No: Two-phase sampling for regression

Question 5: Is it practically feasible and not too expensive to adjourn and resume sampling and measuring repeatedly?

Yes: Sequential sampling

No: Question 6

Question 6: Is spatial coverage of the sample very important (do you expect more large scale than small scale spatial variation)?

Yes: Question 7

No: Question 10

Question 7: Is a model-free estimation of the accuracy of the estimate required?

Yes: Stratified Sampling with compact geographical strata and at least 2 points per stratum

No: Question 8

Question 8: Could there be a cyclic pattern of variation with which a regular grid sample might interfere?

Yes: Question 9

No: Grid Sampling

Question 9: Are there parallel linear structures (e.g., roads, ditches) at constant mutual distances?

Yes: Markov Chain Sampling

No: Systematic Unaligned Sampling

Question 10: Is travel time, not navigation in the terrain an important issue?

Yes: Two-stage Sampling

No: Question 11

Question 11: Is navigation, not travel time in the terrain an important issue?

Yes: Cluster Sampling with regular point configurations as clusters

No: Question 12

Question 12: Are there significant gradients in the area with known and more or less constant direction?

Yes: Cluster Sampling with line transects in the direction of the gradients as clusters

No: Question 13

Question 13: Are both travel time and navigation in the terrain important issues, and can you afford at least a medium sample size (say $n > 30$)?

Yes: compound design types

No: Question 14

Question 14: On the basis of prior information, can you divide the area into sub-areas that differ clearly in mean value, variability or cost of sampling or measurement?

Yes: Stratified Sampling with these sub-areas as strata

No: Question 15

Question 15: Apart from an estimate for the area as a whole, are there large sub-areas for which separate estimates are required?

Yes: Stratified Sampling with these sub-areas as strata

No: Simple Random Sampling.

Bijlage C

Voorspellingen en varianties voor clusters

Tabel C.1 geeft de gegevens die zijn gebruikt in de analyse van de benodigde steekproefomvang (paragraaf 2.3). De gegevens zijn het resultaat van voorspellingen met ‘model 1’ uit Hack-ten Broeke et al. (2004), voor de nitraatconcentratie in het voorjaar van 2001.

Tabel C.1. Voorspellingen $\hat{p}$ en varianties $\widehat{S^2}(p)$ per cluster. Codes: zie Hack-ten Broeke et al. (2004)

gewasgroep	bodemgroep	Gt-groep	$\hat{p}_h$	$\widehat{S^2}(p_h)$	n_h
abr	Z1	2	0.5235	0.04135	4
abr	Z1	3	0.6948	0.03111	9
abr	Z2	1	0.3251	0.04339	7
abr	Z2	2	0.6305	0.02096	2
abr	Z2	3	0.7078	0.007024	33
abr	Z3	1	0.2324	0.0002149	2
abr	Z3	2	0.6715	0.006159	5
abr	Z3	3	0.6934	0.01722	29
t	Z2	2	0.9981	0.000003033	3
t	Z3	3	0.9701	0.001270	4
g	Z1	2	0.5882	0.02537	10
g	Z1	3	0.5504	0.01135	26
g	Z2	1	0.3304	0.01396	12
g	Z2	2	0.3972	0.009333	23
g	Z2	3	0.498	0.014	11
g	Z3	1	0.3271	0.001672	5
g	Z3	2	0.5298	0.02852	32
g	Z3	3	0.5729	0.02317	28
m	Z1	3	0.8412	0.004357	8
m	Z2	1	0.4946	0.002027	3
m	Z2	2	0.5548	0.01575	7
m	Z2	3	0.8553	0.008578	2
m	Z3	2	0.516	0.009614	12
m	Z3	3	0.7302	0.005156	16