

32/446(516.2)2^eex

Een nieuwe steekproefstrategie voor de inventarisatie van de fosfaattoestand van percelen

Validatie van het nauwkeurighedsmodel

D.J. Brus
W.J.M. te Riele
J.J. de Gruijter

BIBLIOTHEEK DE HAAFF
Droevendaalsesteeg 3a
6708 PB Wageningen

Rapport 516.2

Staring Centrum, Wageningen, 1999

977 945
977 946

20 APR 2000

REFERAAT

Brus, D.J., W.J.M. te Riele en J.J. de Gruijter, 1999. *Een nieuwe steekproefstrategie voor de inventarisatie van de fosfaattoestand van percelen; validatie van het nauwkeurigheidmodel*. Wageningen, Staring Centrum. Rapport 516.2. 30 blz. 4 fig.; 8 tab.; 3 ref.

Het model van de nauwkeurigheid van het gemiddelde fosfaatgehalte geschat met de nieuwe steekproefstrategie is gevalideerd door bemonstering van 14 percelen. Van elk perceel zijn zes mengmonsters genomen, twee van 20 steken, twee van 30 steken en twee van 40 steken. De chi-kwadraat toets laat zien dat, wanneer de twee maïspcelen op dekzand buiten beschouwing worden gelaten, de werkelijke steekproefvarianties niet significant afwijken van de voorspelde steekproefvarianties. Het gemiddelde verschil tussen de voorspelde en de geschatte standaardafwijking van de steekproeffout, berekend met alle percelen inclusief maïspcelen, is klein en over het algemeen niet significant bij een onbetrouwbaarheidsdrempel van 0.05. De voorspelde varianties met het gepoolde relatieve variogram zijn even goed dan die met de variogrammen behorend bij een bepaalde combinatie van landgebruik, grondsoort en fosfaattoestand. De bijdrage van de schattingsfout van het perceelsgemiddelde in de fout van de voorspelde varianties is te verwaarlozen.

De gemiddelde systematische fout van de bestaande BLGG methode bij 40 steken is geschat door van elk perceel ook een mengmonster volgens deze methode te nemen. Het aandeel van de systematische fout in de gemiddelde gekwadrateerde fout (*MSE*) van de bestaande methode is 66%. Het gemiddelde van de geschatte relatieve systematische fout bedraagt 15%. De nieuwe methode is veel nauwkeuriger dan de bestaande methode door het ontbreken van een systematische fout en een kleinere steekproefvariantie.

Trefwoorden: fosfaatgehalte, kanssteekproef, percelen, steekproeffout, toetsen, validatie

ISSN 0927-4499



© 1999 Staring Centrum, Instituut voor Onderzoek van het Landelijk Gebied (SC),
Postbus 125, NL-6700 AC Wageningen.
Tel.: (0317) 474200; fax: (0317) 424812; e-mail: postkamer@sc.dlo.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 Staring Centrum.

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

ALTERRA is de fusie tussen het Instituut voor Bos- en Natuuronderzoek (IBN) en het Staring Centrum, Instituut voor Onderzoek van het Landelijk Gebied (SC). De fusie gaat in op 1 januari 2000.

Inhoud

Samenvatting	7
1 Inleiding	9
2 Onderzoeksmethode	11
2.1 Dataverzameling	11
2.2 Validatie van het nauwkeurighedsmodel	15
2.2.1 Voorspelling van de steekproefvariantie	15
2.2.2 Schatting van de steekproefvariantie met de steekproefgegevens	15
2.2.3 Toetsen van hypothesen over de steekproefvariantie	17
2.3 Schatten van systematische steekproeffout van bestaande bemonsteringsmethode	18
2.4 Toetsen van hypothesen over duplogemiddelde van graslandpercelen bij bemonstering tot 10 cm	20
3 Resultaten	21
3.1 Kwaliteit van de voorspelde nauwkeurigheid	21
3.2 Systematische steekproeffout van de bestaande bemonsteringsstrategie	22
3.3 Duplogemiddeldes bij bemonstering van grasland tot 10 cm.	24
4 Conclusies	27
Literatuur	29

Samenvatting

Het doel van dit onderzoek is de validatie van het model voor de nauwkeurigheid van het gemiddelde fosfaatgehalte van percelen geschat met de nieuwe, door SC-DLO ontwikkelde steekproefstrategie. Hiervoor zijn 14 percelen bemonsterd. Van elk perceel zijn twee mengmonsters van 20 steken, twee mengmonsters van 30 steken en twee mengmonsters van 40 steken genomen. Een tweede doel van dit onderzoek is de bepaling van de systematische schattingsfout van de bestaande BLGG-bemonsteringsmethode. Daarom is van elk perceel ook een mengmonster volgens deze bestaande methode genomen. Voor graslandpercelen wordt door BLGG overwogen om in de toekomst steken tot 10 cm te nemen in plaats van tot 5 cm. De vraag is hoe groot de verschillen tussen de duplogemiddeldes van de mengmonsters tot 5 cm en tot 10 cm zullen zijn. In verband hiermee is van de 8 graslandpercelen ook een mengmonster tot 10 cm genomen. Deze steken zijn op dezelfde plaatsen genomen als die van het mengmonster volgens de BLGG-methode tot 5 cm.

Van elk perceel is de steekproefvariantie op drie manieren voorspeld, (i) met het variogram van het perceel uit het onderzoek van Brus en Spätjens (1997) dat qua landgebruik, grondsoort en fosfaattoestand het beste lijkt op dat van het nieuwe perceel; (ii) met het gepoolde relatieve variogram en een duplogemiddelde; en (iii) met het gepoolde relatieve variogram en het gemiddelde van de zes duplogemiddeldes. De steekproefvariantie voor een bepaald perceel en een bepaalde steekproefomvang is geschat door de geschatte variantie van de laboratoriumfout af te trekken van de geschatte variantie van de twee duplogemiddeldes. De laboratoriumfout is op twee manieren geschat: (i) per perceel met de zeven mengmonsters die in duplo zijn geanalyseerd; (ii) met een regressiemodel voor de variantie van de laboratoriumfout met als voorspellende variabele het duplogemiddelde. Negatieve geschatte steekproefvarianties zijn vervangen door de waarde 0. Hypotheses over de werkelijke steekproefvariantie zijn getoetst met een χ^2 toets, en een hypothese over het verschil tussen de voorspelde en de werkelijke standaardafwijking van de steekproeffout is getoetst met de t -toets voor gepaarde waarnemingen.

De nulhypothese 'de werkelijke steekproefvariantie is gelijk aan de voorspelde steekproefvariantie' en 'de werkelijke steekproefvariantie is gelijk aan de gemiddelde voorspelde steekproefvariantie' werden verworpen. Wanneer de twee percelen met mais op dekzand buiten beschouwing werden gelaten, werden de nulhypothese bij een onbetrouwbaarheidsdrempel van 0.10 niet verworpen. Voor deze twee percelen was de voorspelde variantie veel kleiner dan de geschatte steekproefvariantie. De gemiddelde verschillen tussen de voorspelde en geschatte standaardafwijking zijn klein en over het algemeen niet significant bij een onbetrouwbaarheidsdrempel van 0.05. De voorspellingen met het gepoolde relatieve variogram zijn even goed tot iets beter dan die met de niet-relatieve, variogrammen behorend bij een bepaalde combinatie van landgebruik, grondsoort en fosfaattoestand. Wanneer het duplogemiddelde van het mengmonster gebruikt wordt als schatting van het perceelsgemiddelde dan is het aandeel van deze schattingsfout in de fout van de voorspelde steekproefvariantie te verwaarlozen.

De nieuwe SC-DLO methode is veel nauwkeuriger dan de bestaande methode ten gevolge van het ontbreken van een systematische fout en een kleinere steekproefvariantie.

Het duplogemiddelde van negen bodemkenmerken bij bemonstering tot 10 cm in grasland wijkt slechts voor twee kenmerken (K en Na) significant af van het duplogemiddelde bij bemonstering tot 5 cm.

1 Inleiding

In het kader van het Mineralen Aangifte Systeem (MINAS) is een nieuwe steekproefstrategie ontwikkeld voor de bepaling van de fosfaattoestand van percelen (Brus en Spätjens, 1997; Brus et al., 1998). In tegenstelling tot de huidige door het Bedrijfslaboratorium voor Grond- en Gewasonderzoek (BLGG) toegepaste methode worden in de nieuwe strategie de steekproefpunten geloot volgens een goed gedefinieerde steekproefopzet. Een dergelijke steekproef wordt een kanssteekproef genoemd. Bij een kanssteekproef zijn de kansen dat een bepaalde steekproef (configuratie van steekproefpunten) wordt getrokken bekend voor alle mogelijke steekproeven die uit het perceel kunnen worden getrokken. Zoals we hierna zullen zien biedt dit voordelen bij de bepaling van de nauwkeurigheid van de steekproefstrategie.

De nieuwe steekproefstrategie is zuiver, d.w.z. wanneer een perceel een zeer groot aantal keren zou worden bemonsterd, is het gemiddelde van de geschatte fosfaattoestanden gelijk aan de werkelijke fosfaattoestand. Dit geldt niet voor de oude strategie omdat in deze strategie de kans dat op een punt een steek wordt genomen niet voor alle punten in het perceel exact gelijk is.

Verder kan in het algemeen in geval van een kanssteekproef de steekproefvariantie, d.w.z. de variantie van het geschatte gemiddelde wanneer de steekproef een zeer groot aantal keren herhaald zou worden in hetzelfde perceel, zuiver geschat worden met de steekproefgegevens zonder dat hiervoor aannames gedaan hoeven te worden, bijvoorbeeld over de ruimtelijke variatie. Probleem bij de nieuwe strategie is dat, om de laboratoriumkosten zoveel mogelijk te beperken, per stratum slechts 1 steek wordt genomen. Deze steken worden samengevoegd tot een mengmonster. In dit geval kan de steekproefvariantie niet uit de steekproefgegevens geschat worden. Het minimum aantal steken per stratum (aantal mengmonsters) om de steekproefvariantie te kunnen schatten is twee. Wel blijft overeind dat, dankzij de bekende trekkingskansen van de steekproeven in de nieuwe steekproefstrategie, met behulp van een model van de ruimtelijke variatie (variogram) de steekproefvariantie van de geschatte fosfaattoestand voorspeld kan worden voor een reeks van steekproefomvangen (aantal steken). Bij de oude strategie kan deze steekproefvariantie niet met behulp van modellen voorspeld worden, maar slechts empirisch worden bepaald door een perceel een groot aantal keren te bemonsteren.

Het doel van dit onderzoek is (i) validatie van het model voor de nauwkeurigheid (steekproefvariantie, standaardafwijking van steekproeffout) van de nieuwe SC-DLO methode, en (ii) de bepaling van de systematische fout van de oude BLGG methode.

Voor grasland wordt overwogen om in de toekomst monsters te nemen tot 10 cm diepte in plaats van 5 cm zoals tot nu toe gebeurt. Nevendoel van dit onderzoek was na te gaan of de geschatte perceelsgemiddeldes bij bemonstering tot 10 cm systematisch afwijken van de geschatte gemiddeldes bij bemonstering tot 5 cm.

2 Onderzoeksmethode

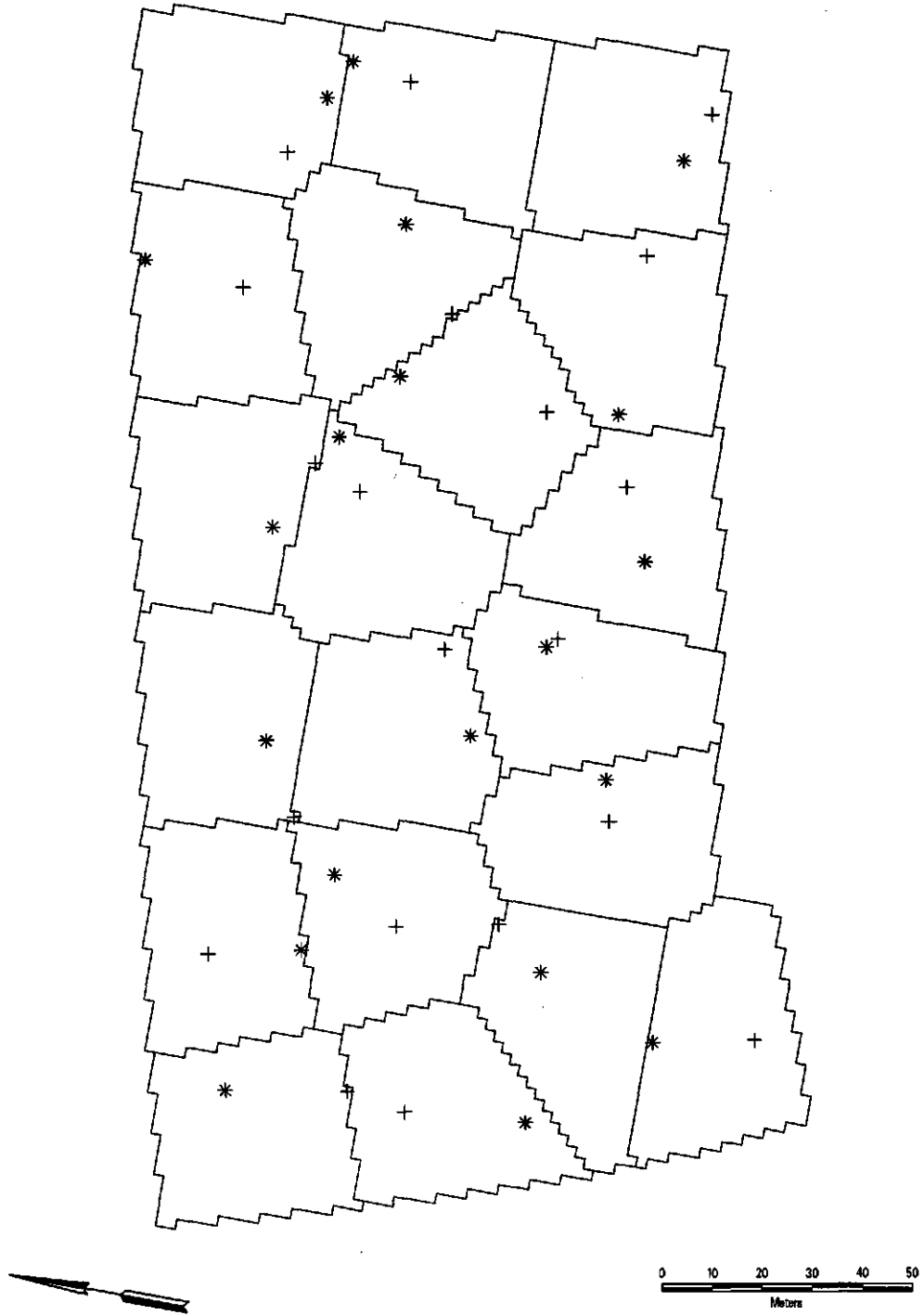
2.1 Dataverzameling

Voor de validatie zijn 14 percelen geselecteerd. Deze percelen vormen zeven combinaties van grondsoort en landgebruik (tabel 1). Van elk perceel zijn zeven mengmonsters genomen, zes volgens de nieuwe methode, en één volgens de bestaande BLGG-methode. De zes mengmonsters volgens de nieuwe methode bestaan uit twee mengmonsters van 20 steken, twee mengmonsters van 30 steken en twee mengmonsters van 40 steken (fig. 1). Van graslandpercelen is tevens een mengmonster volgens de BLGG- methode tot een diepte van 10 cm genomen. Deze steken zijn op dezelfde plaats genomen als die waarbij tot 5 cm is bemonsterd. Alle mengmonsters zijn in duplo geanalyseerd. Alle monsters van een perceel zijn op dezelfde dag genomen zodat verschillen in duplogemiddeldes tussen mengmonsters niet het gevolg kunnen zijn van veranderingen in de tijd.

Tabel 1. Landgebruik, grondsoort, gemiddeld fosfaatgehalte en oppervlakte van de bemonsterde percelen. Laatste kolom: perceel uit Brus et al. (1998) waarvan het variogram is gebruikt voor het voorspellen van de steekproefvarianti

	Landgebruik	Grondsoort	Fosfaat-toestand	Oppervlakte (ha)	Variogram gebruikt van
Est	grasland	rivierklei	27	2,04	Dreumel 13
Zetten	grasland	rivierklei	64	2,73	Dreumel 13
Beuningen	grasland	dekzand	14	1,77	Haaren Achter Broek
Weerselo	grasland	dekzand	63	3,29	Haaren Bremakker
Scherpenzeel	grasland	veen	13	4,08	Oldelamer 5
Rohel	grasland	veen	77	2,12	Oldelamer 11
Grouw	grasland	zware zeeklei	10	3,91	Oostrum Rotteveel
Farmsum	grasland	zware zeeklei	18	2,83	Oostrum Rotteveel
Dronten	akkerbouw	zeeklei	32	6,09	Den Bommel
Swifterbant	akkerbouwi	zeekei	39	1,57	Den Bommel
Onstwedde	akkerbouw	dalgrond	19	1,58	Exloermond 2
Exloo	akkerbouw	dalgrond	51	4,55	Exloermond 1
Schijndel (b)	maïs	dekzand	35	1,03	Wesepe
Schijndel (a)	maïs	dekzand	46	1,25	Wesepe

zetten



*Fig. 1a Geografische stratificatie van het perceel Zetten voor het nemen van een monster van 20 steken
+: locatie van steken van eerste mengmonster: *: idem van tweede mengmonster*

zetten

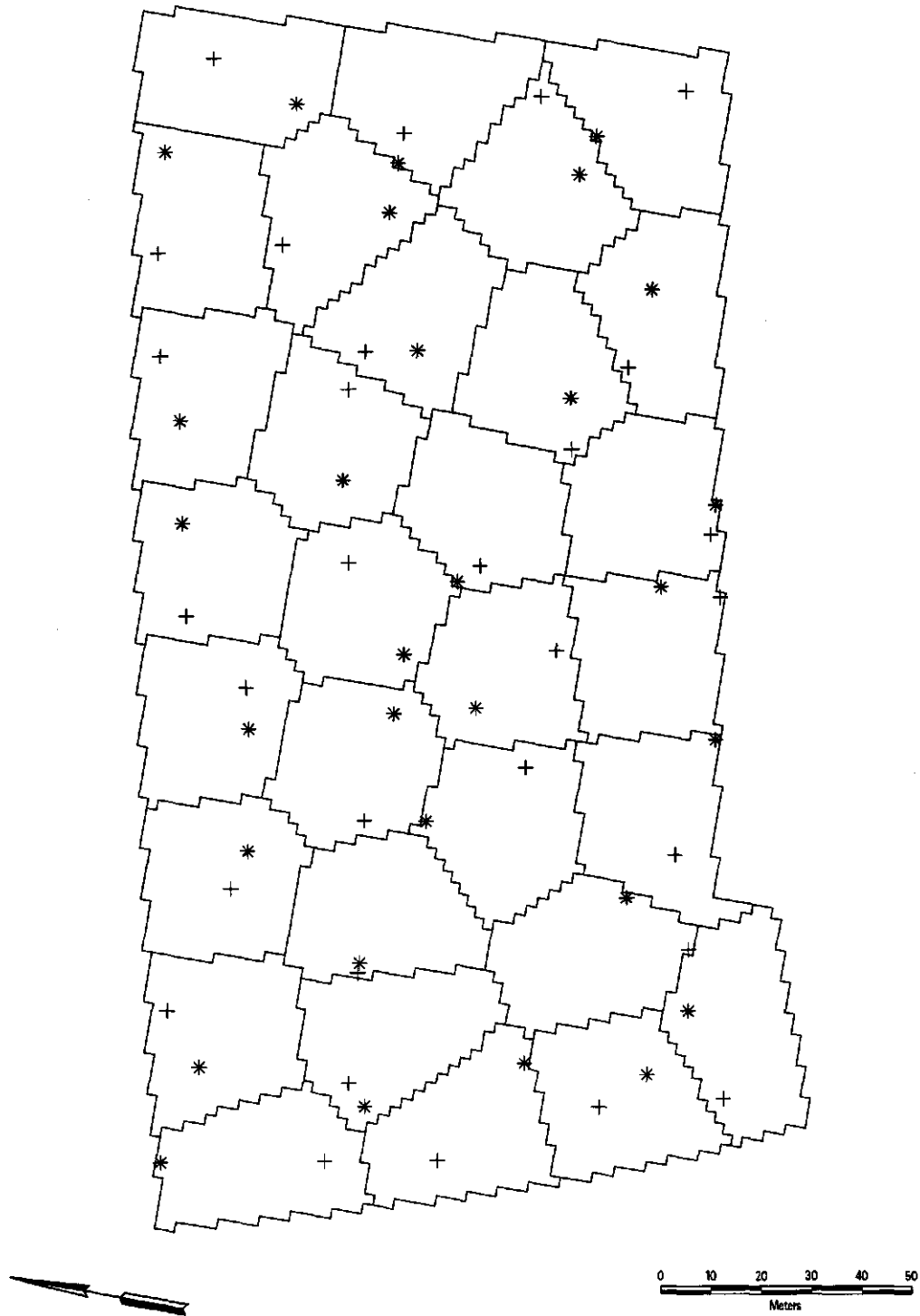


Fig. 1b Geografische stratificatie van het perceel Zetten voor het nemen van een monster van 30 steken
+: locatie van steken van eerste mengmonster; *: idem van tweede mengmonster

zetten

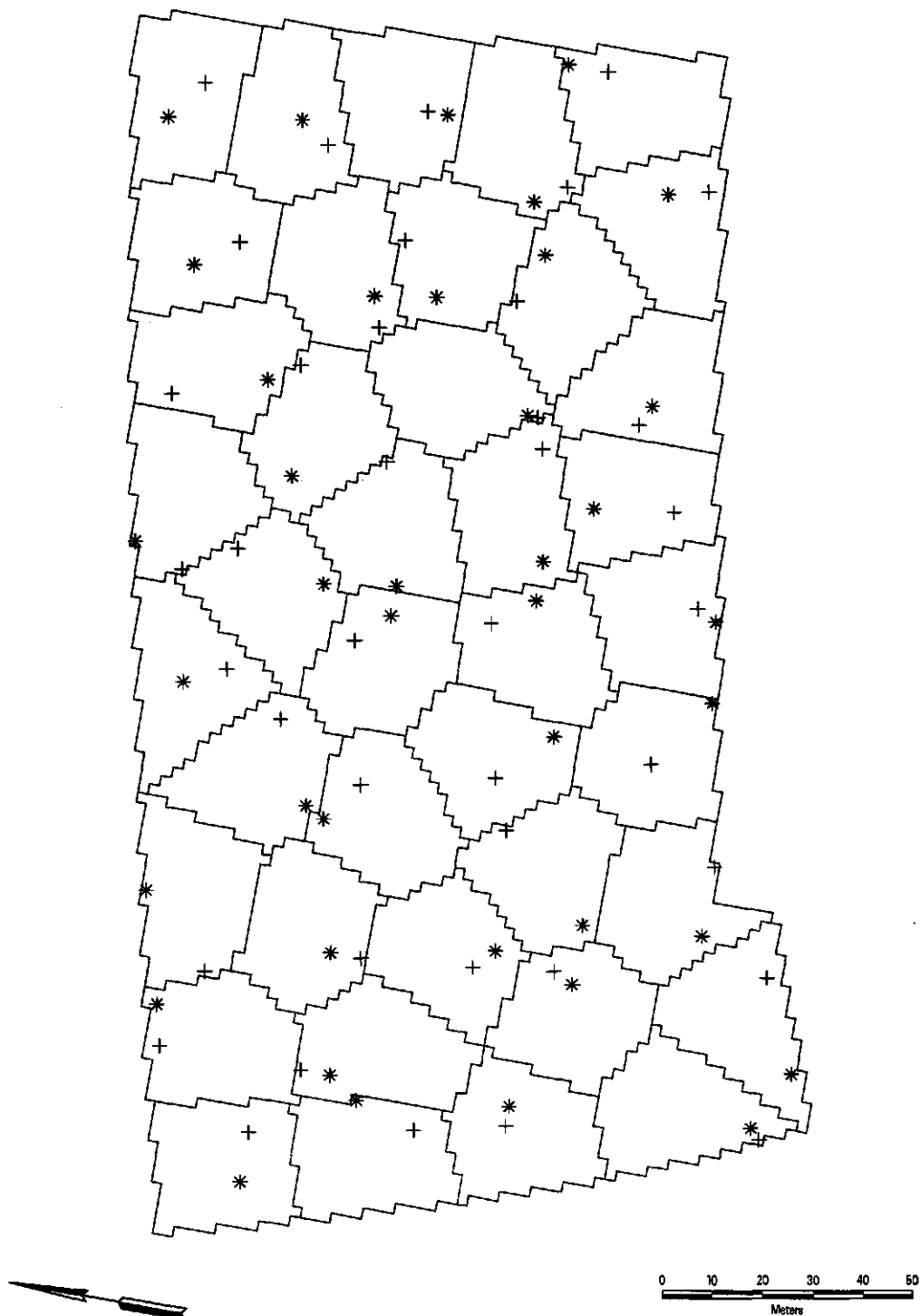


Fig. 1c Geografische stratificatie van het perceel Zetten voor het nemen van een monster van 40 steken
+: locatie van steken van eerste mengmonster: *: idem van tweede mengmonster

2.2 Validatie van het nauwkeurighedsmodel

2.2.1 Voorspelling van de steekproefvariantie

De steekproefvariantie is voorspeld volgens de methode en met de variogrammen beschreven in Brus en Spätjens (1997) en Brus et al. (1999). Gegeven het aantal steken, n , is voor elk perceel de steekproefvariantie voorspeld met het variogram van het perceel dat qua landgebruik, grondsoort en gemiddeld fosfaatgehalte het beste overeenkomt met dat van het nieuwe perceel, en met het gepoolde relatieve variogram. In het laatste geval is het duplogemiddelde van het eerste mengmonster samengesteld uit n steken gebruikt als schatting van de fosfaattoestand van het perceel. Om inzicht te krijgen in het aandeel van de fout in de schatting van de fosfaattoestand in de kwaliteit van de voorspelde steekproefvariantie, is de steekproefvariantie tevens voorspeld met behulp van het ongewogen gemiddelde van de zes duplo's van een perceel genomen met de nieuwe steekproefopzet. Dit gemiddelde van zes duplo's is een veel nauwkeurigere schatting van de fosfaattoestand van een perceel dan het gemiddelde van één duplo, zodat naar verwachting het aandeel van de fout in de geschatte fosfaattoestand in de fout van de voorspelde steekproefvariantie kleiner is. Gegeven een perceel en de steekproefomvang n , is de steekproefvariantie dus drie keer voorspeld, één keer met het gewone variogram, en twee keer met het gepoolde relatieve variogram. Hierna wordt de voorspeller die gebruik maakt van het gewone variogram aangeduid met voorspeller 1, de voorspeller die gebruik maakt van het gepoolde variogram en één duplogemiddelde met voorspeller 2, en de voorspeller die gebruik maakt van het gepoolde variogram en zes duplogemiddeldes als voorspeller 3.

2.2.2 Schatting van de steekproefvariantie met de steekproefgegevens

Om van ieder perceel de steekproefvariantie bij n steken nauwkeurig te schatten zou van elk perceel een groot aantal mengmonsters bestaande uit n steken genomen moeten worden volgens de nieuwe methode. Dit levert een nauwkeurige schatting op van de steekproefvariantie voor dit perceel, en dus ook van de fout in de voorspelde steekproefvariantie, maar deze onderzoeks aanpak geeft geen inzicht in de steekproefvariantie en kwaliteit van de variantievoorspellingen van andere percelen. Daarom hebben wij gekozen voor een zeer beperkt aantal herhalingen binnen een perceel, zodat meerdere percelen bemonsterd kunnen worden. Dit betekent dat met dit onderzoek uitspraken gedaan worden over de gemiddelde kwaliteit van de variantievoorspellingen.

De totale fout van het geschatte gemiddelde fosfaatgehalte van een perceel is gelijk aan de som van de steekproeffout en de laboratoriumfout. De steekproeffout en laboratoriumfout zijn onafhankelijk, zodat de variantie van de steekproeffout (steekproefvariantie) gelijk is aan de variantie van de totale fout min de variantie van de laboratoriumfout:

$$V(e_{veld}) = V(e_{tot}) - V(e_{lab}) \quad (1)$$

De steekproefvariantie kan geschat worden door de twee variantiecomponenten van vergelijking (1) te schatten:

$$\hat{V}(e_{veld}) = \hat{V}(e_{tot}) - \hat{V}(e_{lab}) \quad (2)$$

Het fosfaatgehalte van een mengmonster is bepaald door twee keer een submonster te nemen en beide submonsters een keer te analyseren (duplo-bepaling), en vervolgens het ongewogen gemiddelde van deze twee analyseuitslagen te nemen. De variantie van de totale fout kan in dit geval geschat worden met:

$$\hat{V}(e_{tot}) = \frac{1}{I-1} \sum_{i=1}^I (\bar{m}_i - \bar{m})^2 \quad (3)$$

waarin:

\bar{m}_i = duplogemiddelde van *i*-de mengmonster,

\bar{m} = gemiddelde van *I* duplogemiddeldes.

In dit onderzoek zijn per perceel twee mengmonsters van een zelfde aantal steken genomen zodat *I* gelijk is aan twee.

Omdat elk mengmonster twee keer, soms drie keer is geanalyseerd kan met elk mengmonster de variantie van de laboratoriumfout geschat worden met

$$\hat{V}(e_{lab}) = \frac{1}{J-1} \sum_{j=1}^J (m_{ij} - \bar{m}_i)^2 \quad (4)$$

waarin:

m_{ij} = *j*-de bepaling van fosfaatgehalte van *i*-de mengmonster.

Voor het schatten van de laboratoriumfout zijn ook de mengmonsters genomen volgens de bestaande BLGG-methode gebruikt, echter niet het mengmonster van graslandpercelen tot een diepte van 10 cm. Per perceel zijn er dus zeven mengmonsters waarmee de variantie van de laboratoriumfout is geschat. Vervolgens is per perceel het ongewogen gemiddelde van de zeven geschatte varianties van de laboratoriumfout berekend. Deze gemiddelde variantie is afgetrokken van de geschatte totale variantie (Vergelijking 3) wat resulteert in een geschatte steekproefvariantie voor een gegeven perceel en een bepaald aantal steken. Merk op dat de geschatte steekproefvariantie negatief is als de geschatte variantie van de laboratoriumfout groter is dan de geschatte totale variantie. Deze negatieve waarden zijn vervangen door 0.

De laboratoriumfout is tevens bepaald door middel van een regressiemodel. Zetten we de geschatte variantie van de laboratoriumfout uit tegen het duplogemiddelde, dan blijkt dat de variantie van de laboratoriumfout toeneemt met het duplogemiddelde (fig. 2). Tevens blijkt dat de restvariantie sterk toeneemt in deze richting. In verband

hiermee is door middel van gewogen lineaire regressie een model aangepast, met als gewichten de inverse van de gefitte waardes. Hiervoor is de regressie een aantal keren herhaald. De eerste keer zijn gelijke gewichten gebruikt, de tweede keer zijn de gewichten gelijk aan de inverse van de gefitte waarde in de eerste stap, enz. Deze iteratieve regressie is herhaald totdat de regressiecoëfficiënten niet meer veranderen. In tabel 2 staan de resultaten van de lineaire regressie. De uiteindelijke gefitte waarden voor de twee mengmonsters van een perceel van hetzelfde aantal steken zijn gemiddeld.

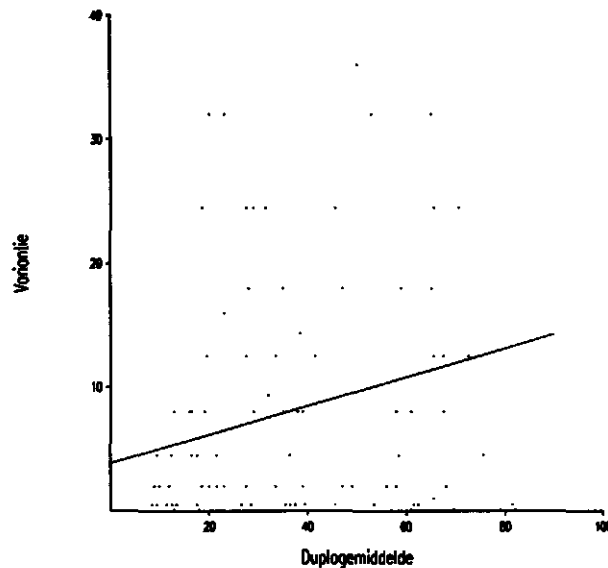


Fig. 2. Spreidingsdiagram van de variantie van de laboratoriumfout bij enkelvoudige bepaling versus het duplogemiddelde. De lijn is aangepast met iteratieve, gewogen regressie met als gewicht de inverse van de gefitte waarde

Tabel 2 Geschatte coëfficiënten van het lineaire regressiemodel voor de variantie van de laboratoriumfout (bij enkelvoudige bepaling) met als voorspellende variabele het duplogemiddelde

		Standaardfout	t(96)
B_0	3.83	1.76	2.17
B_1	0.1160	0.048	2.42

2.2.3 Toetsen van hypothesen over de steekproefvariantie

De steekproefgegevens zijn gebruikt om hypothesen over de steekproefvariantie te toetsen. Voor alle drie steekproefomvangs zijn drie nulhypotesen getoetst. In de eerste nulhypothese wordt verondersteld dat voor alle 14 percelen de werkelijke steekproefvariantie bij n steken gelijk is aan de voorspelde steekproefvariantie bij n steken van dat perceel:

$$H_0 : V_p(e_{veld}) = \tilde{V}_p(e_{veld}), p = 1 \dots P \quad (5)$$

waarin:

$V_p(e_{veld})$ = de werkelijke steekproefvariantie van perceel p bij n steken;

$\tilde{V}_p(e_{veld})$ = de voorspelde steekproefvariantie van perceel p bij n steken.

We berekenen voor elk perceel de toetsingsgrootheid

$$\frac{(n-1)\hat{V}_p(e_{veld})}{\tilde{V}_p(e_{veld})} \quad (6)$$

waarin:

$\hat{V}_p(e_{veld})$ = de geschatte steekproefvariantie van perceel p bij n steken;

$\tilde{V}_p(e_{veld})$ = de voorspelde steekproefvariantie van perceel p bij n steken.

Wanneer we aannemen dat het bodemkenmerk normaal verdeeld is, is deze toetsingsgrootheid bij benadering χ^2 verdeeld met $n-1 = 1$ vrijheidsgraad. De som van deze toetsingsgrootheden per perceel is ook bij benadering χ^2 verdeeld, maar met P vrijheidsgraden, waarin P het aantal percelen is.

Wanneer in tenminste 1 perceel de geschatte steekproefvariantie significant afwijkt van de voorspelde steekproefvariantie, wordt de nulhypothese verworpen. Deze toets zegt niets over de gemiddelde kwaliteit van de variantievoorspellingen. Daarom is ook de nulhypothese 'de werkelijke steekproefvariantie bij n steken is gelijk aan de gemiddelde (over de 14 percelen) voorspelde steekproefvarianties bij n steken' getoetst:

$$H_0 : V_p(e_{veld}) = \tilde{V}_{gem}(e_{veld}), p = 1 \dots P \quad (7)$$

waarin:

$V_p(e_{veld})$ = de werkelijke steekproefvariantie van perceel p bij n steken;

$\tilde{V}_{gem}(e_{veld})$ = het gemiddelde van de voorspelde steekproefvarianties bij n steken.

Tot slot is getoetst of het verschil tussen de voorspelde en de geschatte standaardafwijking van de steekproeffout significant afwijkt van 0 met een t -toets voor gepaarde waarnemingen. Merk op dat in deze toets gebruik wordt gemaakt van standaardafwijkingen in plaats van varianties. Het voordeel hiervan is dat de geschatte standaardafwijkingen minder scheef verdeeld zijn. Bovendien is de standaardafwijking relevanter voor de praktijk dan de variantie. In deze toets wordt aangenomen dat het gemiddelde verschil gedeeld door de standaardafwijking van het gemiddelde t -verdeeld is met $P-1$ vrijheidsgraden.

2.3 Schatten van systematische steekproeffout van bestaande bemonsteringsmethode

In de bestaande BLGG-bemonsteringsmethode is de kans dat op een punt een steek wordt genomen niet voor alle punten in het perceel gelijk, met als gevolg een

systematische fout van het geschatte perceelsgemiddelde. Deze systematische fout is geschat door op alle percelen ook een mengmonster bestaande uit 40 stekken volgens de bestaande bemonsteringsmethode te nemen. Net als bij de steekproefvariantie zou, om van ieder perceel de systematische fout nauwkeurig te schatten, zowel de bemonstering volgens de bestaande methode als die volgens de nieuwe methode een groot aantal keren moeten worden herhaald. Dit levert voor dit perceel een nauwkeurige schatting op van de systematische fout, maar deze onderzoeks aanpak geeft geen inzicht in de systematische fout bij bemonstering van andere percelen. Daarom is van elk perceel slechts 1 mengmonster volgens de bestaande methode genomen. Weliswaar levert dit slechts een ruwe (onnauwkeurige) schatting op van de systematische fout, maar door deze te middelen over alle 14 percelen krijgen we een redelijk nauwkeurige schatting van de gemiddelde systematische fout. Om te voorkomen dat systematische fouten van verschillende percelen tegen elkaar wegvallen, zijn de geschatte systematische fouten van de percelen gekwadrateerd voordat deze zijn gemiddeld:

$$\text{Systematische fout} = \sqrt{\frac{1}{P} \sum_p^P E^2(\hat{m}_p^{BLGG} - m_p)} \quad (8)$$

waarin:

P = het aantal percelen;

$E(.)$ = de statistische verwachtingswaarde bij herhaling van de bemonstering volgens de BLGG-methode;

\hat{m}_p^{BLGG} = het geschatte gemiddelde van perceel p bij bemonstering volgens BLGG-methode;

m_p = het werkelijke gemiddelde van perceel p .

Omdat de schatter van het perceelsgemiddelde van de nieuwe strategie zuiver is, kan de systematische fout geschat worden met

$$\text{Systematische fout} = \sqrt{\frac{1}{P} \sum_p^P E^2(\hat{m}_p^{BLGG} - \hat{m}_p^{SC-DLO})} \quad (9)$$

waarin:

\hat{m}_p^{SC-DLO} = het geschatte gemiddelde van perceel p bij bemonstering volgens de nieuwe SC-DLO methode.

De nauwkeurigheid van de geschatte systematische fout kan worden vergroot door voor een gegeven perceel niet het fosfaatgehalte van 1 mengmonster genomen met de nieuwe strategie te nemen, maar het gemiddelde van de zes mengmonsters genomen volgens de nieuwe strategie.

Tot slot hebben we de relatieve systematische fout geschat door het verschil tussen het 'BLGG-duplogemiddelde' en het gemiddelde van de zes 'SC-DLO-duplogemiddeldes' te delen door het gemiddelde van de zes SC-DLO duplogemiddeldes.

Het gemiddelde over de 14 percelen van de relatieve systematische fout is berekend door eerst de absolute waarde van het bovengenoemde verschil te nemen.

2.4 Toetsen van hypothesen over duplogemiddelde van graslandpercelen bij bemonstering tot 10 cm

Voor alle grasland percelen is het verschil berekend tussen de duplogemiddelden bij bemonstering tot 5 cm en bij bemonstering tot 10 cm. Vervolgens is getoets of het gemiddelde verschil significant afwijkt van 0 door middel van de *t*-toets voor gepaarde waarnemingen. Voor die kenmerken waarvoor het verschil significant afwijkt van 0 is een regressiemodel aangepast voor het duplogemiddelde tot 10 cm met als verklarende variabele het duplogemiddelde tot 5 cm.

3 Resultaten

3.1 Kwaliteit van de voorspelde nauwkeurigheid

Hierna worden alleen de resultaten gepresenteerd waarbij perceelsspecifieke schattingen van de variantie van de laboratoriumfout zijn gebruikt voor het schatten van de steekproefvariantie. Gebruik van het regressiemodel van tabel 2 voor het schatten van de variantie van de laboratoriumfout leidde tot vergelijkbare resultaten. In tabel 3 en 4 staan de resultaten van de twee χ^2 -toetsen.

Tabel 3. χ^2 in geval van de nulhypothese 'de werkelijke steekproefvariantie is gelijk aan de voorspelde steekproefvariantie'. Tussen haakjes: χ^2 wanneer de twee mais percelen op dekzand buiten beschouwing worden gelaten

Aantal steken	Voorspeller 1	Voorspeller 2	Voorspeller 3
20	143.9 (18.4)	36.6 (12.6)	36.3 (13.5)
30	52.0 (18.8)	20.1 (14.6)	21.6 (16.1)
40	164.8 (44.0)	42.4 (17.9)	38.3 (17.0)

Tabel 4. χ^2 in geval van de nulhypothese 'de werkelijke steekproefvariantie is gelijk aan de gemiddelde voorspelde steekproefvariantie'. Tussen haakjes: χ^2 wanneer de twee mais percelen op dekzand buiten beschouwing worden gelaten

Aantal steken	Voorspeller 1	Voorspeller 2	Voorspeller 3
20	60.5 (24.1)	53.9 (21.5)	54.2 (21.6)
30	26.8 (20.7)	23.9 (18.4)	24.1 (18.5)
40	29.0 (12.8)	25.9 (11.4)	26.0 (11.5)

De kritische waarde van χ^2 bij een onbetrouwbaarheidsdrempel van 0.05 en 0.10 is respectievelijk 24.7 en 22.4. Dit betekent dat voor $n = 20$ en $n = 40$ beide nulhypotheseën in alle gevallen wordt verworpen. De nulhypotheseën worden (net) niet verworpen wanneer de steekproefvariantie voor $n = 30$ wordt voorspeld met het gepoolde variogram (voorspeller 2 en 3). Uit een nadere analyse bleek dat de bijdrage van het perceel Schijndel 1 (mais op dekzand) aan de toetsingsgrootte χ^2 onevenredig groot was. De geschatte steekproefvarianties voor dit perceel zijn extreem groot: 91.4, 25.4 en 53.9 voor respectievelijk $n = 20$, 30 en 40. Ook voor Schijndel 2 zijn de geschatte steekproefvarianties groot, zij het minder extreem. Wanneer deze twee percelen buiten beschouwing worden gelaten, dan worden voor voorspeller 2 en 3 beide nulhypothese bij een onbetrouwbaarheidsdrempel van 0.10 over het algemeen niet verworpen, met andere woorden er zijn dan geen significante verschillen tussen de voorspelde en de geschatte steekproefvarianties.

Tot slot presenteren we de resultaten van de t -toets. In tabel 5 staan de gemiddeldes en varianties van de verschillen tussen de voorspelde en de geschatte standaardafwijking van de steekproeffout.

Tabel 5. Gemiddelde en variantie van het verschil tussen voorspelde en geschatte standaardafwijking van de steekproeffout voor drie voorspellers en drie steekproefomvang. Tussen haakjes: *t*-waarde

		20	30	40
Voorspeller 1	Gemiddelde	-0.144 (-0.492)	0.104 (0.640)	-0.255 (-1.248)
	Variantie	16.7	5.2	8.2
Voorspeller 2	Gemiddelde	0.094 (0.373)	0.319 (2.161)	-0.115 (-0.671)
	Variantie	12.4	4.3	5.8
Voorspeller 3	Gemiddelde	0.092 (0.363)	0.289 (1.939)	-0.094 (-0.555)
	Variantie	12.6	4.4	5.6

Uit tabel 5 blijkt dat de gemiddelde verschillen klein zijn, zeker in relatie tot het gemiddelde van de 14 geschatte standaardafwijkingen. Deze gemiddeldes zijn 2.413, 1.684 en 1.764 voor respectievelijk $n = 20$, 30 en 40. Uit de *t*-toets blijkt dat bij een onbetrouwbaarheidsdrempel van 0.05 het verschil voor voorspeller 2 bij $n = 30$ significant is, en bij een onbetrouwbaarheidsdrempel van 0.10 het verschil voor voorspeller 2 en 3 bij $n = 30$ significant is. (tabel 5). De kritische waarden voor onbetrouwbaarheidsdrempels van 0.05 en 0.10 zijn respectievelijk 2.160 en 1.771.

Uit tabel 5 blijkt verder dat voorspeller 2 en 3 het iets beter doen dan voorspeller 1. Voor $n = 20$ en 40 is het gemiddelde verschil kleiner, en in alle gevallen is de variantie van het verschil kleiner. Opvallend is dat de kwaliteit van de standaardafwijking voorspeld met het gepoolde variogram niet beter wordt wanneer het gemiddelde van de zes duplogemiddeldes wordt gebruikt als schatting van het perceelsgemiddelde.

Voor voorspeller 2 hebben we tot slot een regressiemodel zonder intercept aangepast voor de geschatte standaardafwijking van de steekproeffout met als voorspellende variabele de voorspelde standaardafwijking van de steekproeffout (tabel 6). Alleen Schijndel 1 is hierbij buiten beschouwing gelaten. Het gestandaardiseerde residu van Schijndel 2 was niet extreem groot of klein. Uit tabel 6 blijkt dat er een significante relatie bestaat tussen de voorspelde standaardfout en de geschatte standaardfout.

Tabel 6 Geschatte coëfficiënten van de lineaire regressiemodellen zonder intercept voor de geschatte standaardafwijking van de steekproeffout met als voorspellende variabele de voorspelde standaardafwijking van de steekproeffout

Aantal steken	<i>B</i> 1	Standaardfout van <i>B</i> 1	<i>t</i> (13)
20	0.967	0.365	2.65
30	0.946	0.270	3.51
40	1.046	0.378	2.77

3.2 Systematische steekproeffout van de bestaande bemonsteringsstrategie

Figuur 3 is een histogram van het verschil tussen het duplogemiddelde van het mengmonster genomen volgens de bestaande BLGG-strategie en het geschatte gemiddelde volgens de nieuwe strategie. Deze verschillen zijn dus schattingen van de systematische fout per perceel. Merk op dat zowel negatieve als positieve verschillen voorkomen, d.w.z. soms wordt het gemiddelde onderschat, soms overschat.

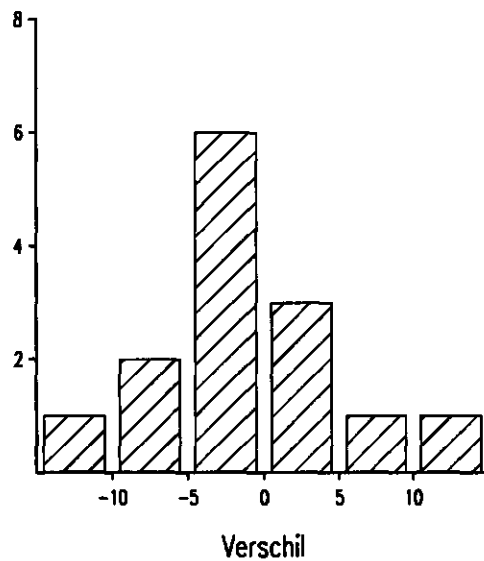


Fig. 3. Histogram van het verschil tussen het duplogemiddelde van het BLGG-mengmonster en het gemiddelde van de zes duplogemiddeldes van de SC-DLO mengmonsters

De geschatte gemiddelde systematische fout (zie vergelijking 9), bedraagt 5.743. Om de gemiddelde gekwadraterde fout (*Mean Squared Error, MSE*) van de bestaande strategie te verkrijgen moet deze systematische fout gekwadrateerd worden en opgeteld worden bij de variantie van de bestaande strategie. Deze variantie is door Ferrari en Vermeulen (1955) experimenteel bepaald door 1000 percelen twee keer te bemonsteren. Voor P-citroen vonden zij een standaardafwijking van de totale fout (inclusief laboratoriumfout) van 11% van het gemiddelde. Wanneer we aannemen dat voor Pw en PAL de totale fout ook 11% is en de laboratoriumfout van een P-citroen bepaling gelijk is aan die van Pw en PAL, betekent dit voor de 14 percelen een gemiddelde steekproefvariantie van 16.9. De gemiddelde *MSE* van de bestaande methode komt hiermee op 49.8. Het aandeel van de systematische fout in de *MSE* is aanzienlijk, n.l. 66%. Voor de nieuwe strategie is de systematische fout 0, zodat de *MSE* gelijk is aan de steekproefvariantie. De gemiddelde geschatte steekproefvariantie van de nieuwe strategie bij 40 steken is 8.6 zodat we kunnen concluderen dat de nieuwe strategie veel nauwkeuriger is dan de bestaande strategie.

De geschatte relatieve systematische fout, d.w.z. de geschatte systematische fout gedeeld door het geschatte perceelsgemiddelde, varieert van -49% tot 33%. (fig. 4). Het gemiddelde van de absolute waarde van de relatieve systematische fout bedraagt 15%.

De nauwkeurigheid van de geschatte systematische fouten per perceel is laag omdat er slechts 1 mengmonster volgens de BLGG-methode is genomen. De nauwkeurigheid van het gemiddelde van de systematische fout is groter omdat dit gemiddelde is gebaseerd op 14 herhalingen. We hebben de standaardafwijking van de geschatte gemiddelde systematische fout (vergelijking 9) benaderd met de eerste orde Taylor methode. Dit resulteerde in een standaardafwijking van 1.10, zodat de geschatte gemiddelde systematische fout significant afwijkt van 0. Duidelijke oorzaken van de

systematische fout hebben wij niet kunnen aangeven. Nader onderzoek naar de grootte en de oorzaak van de systematische fout is daarom gewenst.

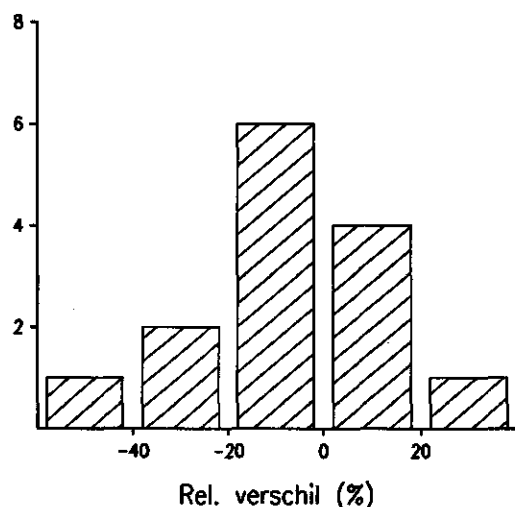


Fig. 4. Histogram van het relatieve verschil (%) tussen duplogemiddelde van het BLGG-mengmonster en het gemiddelde van de zes duplogemiddeldes van de SC-DLO mengmonsters

3.3 Duplogemiddeldes bij bemonstering van grasland tot 10 cm

In tabel 7 staan de gemiddelde verschillen tussen de duplogemiddelden bij bemonstering tot 5 cm en bij bemonstering tot 10 cm voor 9 bodemkenmerken. Tevens zijn de *t*-waarden weergegeven. Alleen voor K en Na is het verschil significant. Voor deze twee bodemkenmerken is een regressiemodel aangepast zonder intercept (tabel 8). Het intercept week niet significant af van 0.

Tabel 7 Het gemiddelde verschil tussen de duplogemiddeldes bij bemonstering tot 5 cm en bij bemonstering tot 10 cm van grasland; een positief verschil betekent dat gehalte tot 5 cm groter is dan tot 10 cm

	Gemiddeld verschil	<i>t</i> (7)
pH	-0.068	-1.165
Org	0.265	0.384
PAL	0.563	0.405
K	2.481	2.457
Mg	55.25	1.581
Cu	0.500	1.219
Cb	0.044	0.990
Na	0.375	1.932
Lutum	0.113	0.280

Tabel 8 Geschatte coëfficiënt van het lineaire regressiemodel zonder intercept voor het duplogemiddelde bij bemonstering tot 10 cm. met als voorspellende variabele het duplogemiddelde bij bemonstering tot 5 cm. voor Kalium en Natrium

	<i>B1</i>	<i>Standaardfout B1</i>	<i>Restvariantie</i>
K	0.886	0.030	4.903
Na	0.934	0.022	0.205

4 Conclusies

- Het gemiddelde verschil tussen de voorspelde en geschatte standaardafwijking van de steekproeffout bij bemonstering volgens de nieuwe SC-DLO methode, is klein en over het algemeen niet significant.
- Het model voorspelt voor de twee percelen met mais op dekzand een veel te lage steekproefvariantie. Worden deze twee percelen buiten beschouwing gelaten dan wijken de geschatte steekproefvarianties niet significant af van de voorspelde varianties.
- De voorspellingen met het gepoolde relatieve variogram zijn even goed of zelfs iets beter dan die met de niet-relatieve, variogrammen behorend bij een bepaalde combinatie van landgebruik, grondsoort en fosfaattoestand.
- Wanneer het duplogemiddelde van een mengmonster gebruikt wordt als schatting van het perceelsgemiddelde dan is het aandeel van deze schattingsfout in de fout van de voorspelde steekproefvariantie te verwaarlozen.
- De nieuwe bemonsteringsmethode is veel nauwkeurig dan de bestaande methode. De systematische fout van de nieuwe methode is 0 terwijl deze voor de bestaande methode aanzienlijk is, en de steekproefvariantie van de nieuwe methode is veel kleiner dan de door Ferrari en Vermeulen in 1955 bepaalde steekproefvariantie.
- Het duplogemiddelde van negen bodemkenmerken bij bemonstering tot 10 cm in grasland wijkt slechts voor twee kenmerken (K en Na) significant af van het duplogemiddelde bij bemonstering tot 5 cm.

Literatuur

Brus, D.J. en L.E.E.M. Spätjens, 1997. Een nieuwe steekproefstrategie voor de inventarisatie van de fosfaattoestand van percelen. Voorspelling van de nauwkeurigheid en kosten. Rapport 516.1 SC-DLO, Wageningen.

Brus, D.J., L.E.E.M. Spätjens, and J.J. de Gruijter, 1998. 'A sampling scheme for estimating the mean extractable phosphorus concentration of fields for environmental regulation'. *Geoderma* 89, 129-148.

Ferrari, Th. J. en F.H.B. Vermeulen, 1955. 'Soil heterogeneity and soil testing'. *Netherlands Journal of Agricultural Science* 3, 265-275.