

Grondwater opnieuw op de kaart

In opdracht van het ministerie van LNV

Grondwater opnieuw op de kaart

Methodiek voor de actualisering van grondwaterstands informatie en perceelsclassificatie naar uitspoelingsgevoeligheid voor nitraat

**J.J. de Gruijter, J.B.F. van der Horst, G.B.M. Heuvelink, M. Knotters,
T. Hoogland**

Alterra-rapport 915

Alterra, Wageningen, 2004

REFERAAT

Jaap de Gruijter, Jack van der Horst, Gerard Heuvelink, Martin Knotters, Tom Hoogland, 2004. *Grondwater opnieuw op de kaart; Methodiek voor de actualisering van grondwaterstandsinformatie en perceelsclassificatie naar uitspoelingsgevoeligheid voor nitraat* Wageningen, Alterra-rapport 915. 70 blz. 10 fig.; 2 tab.; 29 ref.

Dit rapport beschrijft en motiveert de methodiek voor actualisatie van grondwaterstandsinformatie en perceelsclassificatie naar uitspoelingsgevoeligheid zoals die vanaf eind 2003 door Alterra wordt toegepast. Eerst wordt via ‘gerichte opname’, tijdreeksanalyse en regressieanalyse de klimaatsrepresentatieve GxG bepaald op de locaties van een verdicht meetnet, en wordt gebiedsdekkende hulpinformatie verzameld vanuit het AHN en andere bronnen. Vervolgens vindt, gebruik makend van deze gegevens en hun onderlinge correlaties, geostatistische simulatie plaats van een groot aantal (300) gebiedsdekkende GxG velden. Tenslotte worden de percelen op basis van deze simulaties geclassificeerd m.b.v. een door de gebruiker te kiezen GxG criterium, een oppervlaktecriterium, en een kanscriterium. Dit laatste bepaalt de kans op misclassificaties.

Trefwoorden: grondwater, zandgronden, uitspoelingsgevoeligheid, nitraat, tijdreeksanalyse, geostatistiek, Universal Kriging, Full Gaussian co-simulation.

ISSN 1566-7197

Eerste auteur's e-mail: jaap.degruijter@wur.nl



© 2004 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 Doelstelling	12
1.3 Opbouw van het rapport en leeswijzer	12
1.4 Afkortingen en begrippen	13
2 De methodiek in hoofdlijnen	15
2.1 Schematische weergave	15
2.2 Gegevens over grondwaterstand, topografie en waterhuishouding . .	16
2.3 Gebiedsdekkende interpolatie van de GxG	18
2.4 Gebiedsdekkende simulatie van de GxG	18
2.5 Perceelsclassificatie naar uitspoelingsgevoeligheid	18
3 Gegevens over grondwaterstand, topografie en waterhuishouding	21
3.1 Inleiding	21
3.2 Indeling van het gebied in homogene deelgebieden	21
3.2.1 Aard en doel van de gebiedsindeling	21
3.2.2 Gegevens voor de stratificatie	22
3.2.3 Werkwijze bij de stratificatie	22
3.3 GxG's op peilbuislocaties	23
3.3.1 Selectie van grondwaterstandsreeksen	23
OLGA/DINO	23
Overige grondwaterstandsreeksen	24
3.3.2 Berekening van klimaatrepresentatieve GxG's op peilbuislocaties	25
3.4 GxG's op tijdelijke meetpunten	26
3.4.1 Gerichte opnames	26
Additionele opnames	26
3.4.2 Berekening van GxG's op gerichte-opnamelocaties	27
Stap 1: bepaling van het regressiemodel	27
Stap 2: toepassing van het regressiemodel	28
3.5 Gebiedsdekkende hulpinformatie over topografie en waterhuishouding	28
3.5.1 Voorbewerking van het AHN	28
3.5.2 Afleiding van hulpinformatie	29
Groep 1: relatieve maaiveldhoogten	29
Groep 2: drainagedichtheid	29
Groep 3: drooglegging ten opzichte van maaiveld	29
Groep 4: maaiveld ten opzichte van NAP	31

Groep 5: de GHG en GLG volgens de Gt-kaart 1 : 50 000 en de geschatte bergingscapaciteit	32
4 Gebiedsdekkende simulatie en interpolatie van de GxG	33
4.1 Inleiding	33
4.2 Regressieanalyse	34
4.3 Ruimtelijke structuur van de residuen	36
4.4 Verrekening van de onnauwkeurigheid van GxG-schattingen	37
4.5 Simulatie van GHG- en GLG-velden	38
4.6 Gebiedsdekkende interpolatie van de GxG	38
5 Perceelsclassificatie naar uitspoelingsgevoeligheid	41
6 Kwaliteit van de methodiek	43
Bibliografie	45
Bijlagen	47
A Tijdreeksmodellering	47
B Sequentiële Gaussische co-simulatie	51
B.1 Introduction	52
B.2 Theory	52
B.2.1 Sequential Simple Gaussian Simulation	52
B.2.2 Sequential Simple Gaussian Cosimulation	53
B.2.3 Sequential Full Gaussian cosimulation	55
B.3 Manual for gstat	56
B.4 Generality and efficiency issues in the gstat implementation	57
B.4.1 Generality	57
B.4.2 Efficiency	57
B.5 Tests	58
B.6 References	59
C Doorsnijding van percelen door stratumgrenzen	61
D Deel-onderzoek naar het benodigd aantal simulaties	63
E Deel-onderzoek naar een mogelijk perceelsgrootte-effect	69

Woord vooraf

Voor u ligt het rapport “Grondwater opnieuw op de kaart” dat door Alterra vervaardigd is in opdracht van het Ministerie van LNV. Het onderzoek is verricht als uitvoeris van en in samenhang met het project ‘Vaststellen verspreiding uitspoelingsgevoelige gronden’, dat wordt gefinancierd uit Nitraatgelden en DWK-programma 409 ‘Uitspoelingsgevoelige gronden’. In dit project werd als probleemstelling geformuleerd (projectplan 19-7-2000):

“Ten gevolge van het aanscherpen van de verliesnormen in het Nederlandse mestbeleid dient vanaf 1 januari het onderscheid klei/veen en zand/löss bekend te zijn, en vanaf 1 januari 2002 het onderscheid tussen uitspoelingsgevoelige zand/löss en niet-uitspoelingsgevoelige zand/löss. Er is reeds het nodige kaartmateriaal, maar dat dient aangepast op 2 punten:

1. Veengronden die tgv oxidatie etc. zijn veranderd in zandgronden volgens de BGDM-definitie dienen te worden uitgekarteerd en verwerkt in de bestaande kaarten 1:50.000
2. De actuele grondwatersituatie (grondwatertrap) dient in detail te worden vastgesteld en verwerkt in kaarten 1:50.000 van uitspoelingsgevoelige percelen.”

Een herkartering van de Gt op de klassieke wijze zou vele tientallen mensjaren vergen. Met het oog op de noodzaak tot actualisering op korte termijn is daarom door Finke e.a. een methodiek ontwikkeld waarbij, behalve van grondwaterstandsmetingen, gebruik wordt gemaakt van gebiedsdekkende hulpinformatie en een geostatistische interpolatietechniek. Deze statistische methodiek maakt het mogelijk de onzekerheid in de voorspelde GHG en GLG te kwantificeren. Hierop voortbouwend is het doel van dit onderzoek een methodiek te ontwikkelen voor perceelsclassificatie naar uitspoelingsgevoeligheid, waarbij rekening wordt gehouden met de onzekerheid in de GHG en GLG, en waarmee de kans op misclassificatie kan worden beperkt.

Het onderzoek, begonnen in 2001, is uitgevoerd als project ‘Classificatie uitspoelingsgevoelige percelen’, met financiering vanuit het DWK programma 395 ‘Bodem/grondwatergegevens’. Het project werd begeleid door de Begeleidingscommissie Uitspoelingsgevoelige Gronden, bestaande uit de volgende personen:

Ir. E.E. Biewinga (vz., LNV-DL)
J.T.M. Huinink MSc(secr., LNV-EC)

Ing. G.P. Beugelink (RIVM)
Ir. J. Bodegraven (LNV-DN)
Dr.ir. T. Breimer (LNV-DWK)
Ing. G.J. Grotentraast (LNV-DLG)
Dr.ir. G.B.M. Heuvelink (UvA, vanaf 1-6-2003 Alterra)
Drs. D.A. Jonkers (VROM, vanaf 31-12-2003)
Ing. N.J. Molenaar (VROM, tot 31-12-2003)
Dr. E.J. Pebesma (UU)
Drs. P.J.J. Torfs (WUR, tot 1-7-2003)
Dr.ir. F.C. van Geer (NITG-TNO)

Projectleiders van het project ‘Vaststellen verspreiding uitspoelingsgevoelige gronden’:

Dr. P.A. Finke (Alterra, tot 8-10-2001)
Ir. A.J. van Kekem (Alterra, vanaf 8-10-2001)

Wij danken de leden van de begeleidingscommissie uitdrukkelijk voor hun inzet voor dit project en de goede samenwerking. Vanaf het begin is er een intensief inhoudelijk overleg met hen geweest, en dat heeft sterk bijgedragen aan de kwaliteit van het hier gepresenteerde resultaat. Zonder de initiatieven en hulp van dr. P.A. Finke en ir. A.J. van Kekem was dit rapport niet tot stand gekomen. Verder danken wij ook ing. F. de Vries (Alterra) voor zijn bijdragen op GIS-gebied.

A handwritten signature in black ink, appearing to read 'W. van Vierssen', is written over a horizontal line. A vertical line descends from the top left of the signature.

Prof.dr. W. van Vierssen (Algemeen Directeur)

Wageningen, mei 2004

Samenvatting

De landsdekkende grondwatertrappenkaarten (Gt-kaarten), schaal 1 : 50 000, zijn sinds de inventarisaties in de periode 1961-1992 verouderd als gevolg van ingrepen in de waterhuishouding. Het is echter van belang om over Gt-informatie te beschikken die representatief is voor de actuele, heersende hydrologische condities. Bijvoorbeeld voor de aanwijzing van gronden die gevoelig zijn voor de uitspoeling van meststoffen naar het grondwater is actuele informatie nodig over de gemiddeld hoogste en gemiddeld laagste grondwaterstand (GHG en GLG, samen GxG). Daarom heeft dit onderzoek als doel een methodiek te ontwikkelen voor 1) kosteneffectieve en snelle actualisatie van de informatie over de grondwaterdynamiek (Gd), gegeven het heersende waterbeheer en het heersende klimaat; 2) kartering van kwantitatieve Gd-kenmerken zoals de GxG, met gekwantificeerde nauwkeurigheid, en 3) classificatie van percelen naar uitspoelingsgevoeligheid, rekening houdend met de nauwkeurigheid van de gekarteerde GxG's.

In de ontwikkelde methodiek worden de volgende gegevens en informatiebronnen gebruikt: grondwaterstandsreeksen uit het OLGA-bestand van TNO-NITG, neerslagen en verdampingsgegevens van het KNMI, het Actueel Hoogtebestand Nederland (AHN) van de Topografische dienst, de bodem- en Gt-kaart schaal 1 : 50 000, de geologische kaart schaal 1 : 50 000, de geomorfologische kaart schaal 1 : 50 000, de topografische kaart schaal 1 : 10 000 en het Top10-Vectorbestand, het landelijke grondgebruiksbestand Nederland (LGN4) en de ligging van stroomgebieden, waterlopen en vernattingsprojecten. Daarnaast worden er in het veld grondwaterstandswaarnemingen in een groot aantal boorgaten verricht op tijdstippen waarop de grondwaterstand zich rond GHG- of GLG-niveau bevindt (gerichte opnames).

De methodiek start met de indeling in hydrologisch homogene deelgebieden, de zogenaamde stratificatie. Binnen deze strata worden de locaties van de gerichte opnames geloot. Voor de locaties waar tijdreeksen van grondwaterstanden zijn waargenomen kan de GxG worden geschat met behulp van een tijdreeksmodel. Met behulp van deze schattingen, gerichte opnames en regressieanalyse wordt de GxG geschat voor de locaties van de gerichte opnames. Vervolgens kan de GxG gebiedsdekkend worden geïnterpoleerd of gesimuleerd, waarbij gebruik wordt gemaakt van zowel de indeling in strata als van de diverse bronnen van gebiedsdekkende hulpinformatie over topografie en waterhuishouding. *Interpolatie* vindt plaats als er een GxG-kaart moet worden gemaakt. *Simulaties* van GxG-velden worden berekend als percelen moeten worden geclassificeerd naar uitspoelingsgevoeligheid.

Classificatie van percelen naar de gevoeligheid voor uitspoeling van nitraat vindt plaats op basis van drie criteria: 1) een puntcriterium; 2) een oppervlaktecrite-

rium, en 3) een kanscriterium. Het puntcriterium houdt in dat een punt in een perceel uitspoelingsgevoelig is als de bodem bestaat uit zand, de GHG dieper is dan een kritische waarde en de GLG dieper is dan 120 cm beneden maaiveld. Het oppervlaktecriterium wil zeggen dat een perceel uitspoelingsgevoelig is als meer dan een bepaald kritisch percentage van de oppervlakte voldoet aan het puntcriterium voor uitspoelingsgevoeligheid. Het kanscriterium houdt in dat een perceel als uitspoelingsgevoelig wordt geclassificeerd als de kans dat aan het oppervlaktecriterium wordt voldaan groter dan een bepaald kritisch percentage is. Deze kans wordt berekend uit de simulaties van de GxG.

In kwaliteitsborging is voorzien door de methodiek te ontwikkelen als een combinatie van ‘bewezen’ methoden, alleen grondig uitgeteste programmatuur te gebruiken, tussenresultaten intensief te controleren, en aparte deel-onderzoeken te wijden aan potentiële zwakke punten.

Hoofdstuk 1

Inleiding

1.1 Achtergrond en probleemstelling

De grondwatertrappenkaart (Gt-kaart), schaal 1 : 50 000, is de enige *landsdekkende* beschrijving van de seizoensfluctuatie van freatische grondwaterstanden in Nederland. De Gt's werden in de periode 1961–1992 simultaan met de bodem in kaart gebracht en opgeslagen in het Bodemkundig Informatie Systeem (BIS). De Gt heeft betrekking op de diepte van het grondwater. Door de geringe diepte waarop zich in Nederland het grondwater bevindt is de Gt-informatie van belang bij allerlei vraagstukken met betrekking tot de inrichting, het beheer en de kwaliteit van het landelijk gebied. Een voorbeeld hiervan is de aanwijzing van gronden die gevoelig zijn voor de uitspoeling van nitraat naar het grondwater.

In de loop van de tijd veranderden de niveaus en de fluctuaties van de grondwaterstand als gevolg van aanpassingen van de waterhuishouding aan landbouwkundige wensen, drainage, drinkwaterwinning e.d. (Braat et al., 1989). Hierdoor veranderde de informatie op de Gt-kaarten. De behoefte aan Gt-informatie die de heersende, actuele hydrologische omstandigheden representeert bleef echter onverminderd groot. Daarom werd de actualisatie van de Gt-kaarten een speerpunt bij de verzameling van bodemdata (Finke, 2000). Inmiddels bleek ook dat de Gt-informatie de heersende klimatologische omstandigheden beter zou representeren wanneer niet alleen grondwaterstandswaarnemingen gedurende de ca. acht jaar voor afgaand aan de kartering zouden worden gebruikt, zoals tot dan toe gebruikelijk was bij de kartering van Gt's, maar ook gebruik zou worden gemaakt van meteorologische data en de samenhang tussen deze data en de grondwaterstand. Uit een onderzoek onder de gebruikers van de bodem- en Gt-kaart (Finke et al., 1999, tabel 1) bleek bovendien dat een uitgebreidere beschrijving van de grondwaterstandsfluctuatie gewenst was. Er was onder andere behoefte aan ruimtelijke informatie over de gemiddeld hoogste en gemiddeld laagste grondwaterstanden (resp. GHG en GLG, samengevat GxG) die ten grondslag liggen aan de indeling in Gt-klassen. Daarnaast wensten de gebruikers een indicatie van de nauwkeurigheid van de GxG's.

Kwantitatieve informatie over de nauwkeurigheid van GxG's is nodig voor de ondersteuning van beleidsbeslissingen, zoals de aanwijzing van gronden die gevoelig zijn voor uitspoeling van meststoffen. Bij het nemen van deze beslissingen is het

belangrijk inzicht te hebben in het risico dat een perceel ten onrechte als uitspoelingsgevoelig wordt aangewezen. Dit risico moet zo laag mogelijk zijn. Bovendien is het bij de aanwijzing van uitspoelingsgevoelige gronden van belang dat deze is gebaseerd op GxG's die het heersende grondwaterregime representeren, en niet op verouderde informatie.

1.2 Doelstelling

Het onderzoeksdoel is een methodiek te ontwerpen voor:

1. kosteneffectieve en snelle actualisatie van informatie over de dynamiek van de grondwaterstand, gegeven het heersende waterbeheer en de weersvariatie in de klimaatperiode (30 jaar);
2. kartering van kwantitatieve kenmerken voor de dynamiek van de grondwaterstand, zoals de GxG, met gekwantificeerde nauwkeurigheid;
3. classificatie van percelen naar de gevoeligheid voor uitspoeling van nitraat naar het grondwater, rekening houdend met de nauwkeurigheid van de informatie over de GxG's.

Met de eerste twee doelen in het vizier ontwierpen Finke et al. (2002) een methode om een groot aantal kenmerken in kaart te brengen die de grondwaterdynamiek (Gd) karakteriseren. Elk Gd-kenmerk moet zodanig worden gekarteerd dat weergave op schaal 1 : 50 000 verantwoord is, teneinde aan te sluiten bij de bestaande landsdekkende bodem- en Gt-kaart. Gezien de ruime beschikbaarheid van landsdekkende geografische datasets met hoge resolutie werd het volgende verondersteld:

- een aantal datasets bevat informatie die zodanig samenhangt met de Gd-kenmerken, dat deze kunnen worden voorspeld met behulp van een lineair regressiemodel;
- omdat de grondwaterstandsfluctuatie kan afhangen van bodemtype, landgebruik en landschappelijke ligging, kunnen hieruit deelgebieden (strata) worden gevormd waarbinnen de regressierelaties worden gezocht.

De methodiek die Finke et al. (2002) volgden is in deze studie aangepast met betrekking tot de ruimtelijke interpolatie van de GxG. Hier wordt nu de geostatistische methode van *universal cokriging* voor gebruikt. Tevens is de methodiek uitgebreid met een geostatistische simulatie techniek. Deze aanpassing en uitbreiding zijn noodzakelijk om GxG-informatie te kunnen leveren waarmee de aanwijzing van uitspoelingsgevoelige percelen kan worden ondersteund.

1.3 Opbouw van het rapport en leeswijzer

Hoofdstuk 2 schetst de methodiek in hoofdlijnen. Hoofdstuk 2 kan worden gelezen om snel inzicht te krijgen in de kern van de methodiek. Een gedetailleerde beschrijving en verantwoording van de methodiek wordt gegeven in hoofdstuk 3 tot en met

6. Hoofdstuk 3 beschrijft de wijze waarop uit verschillende informatiebronnen de gegevens worden afgeleid die worden gebruikt bij de gebiedsdekkende interpolatie en simulatie van de GxG. Onderdelen hierbij zijn de indeling in homogene deelgebieden (stratificatie), de grondwaterstandswaarnemingen en de gebiedsdekkende hulpinformatie over topografie en waterhuishouding. Hoofdstuk 4 beschrijft hoe de GxG gebiedsdekkend wordt gesimuleerd of geïnterpoleerd. Hoofdstuk 5 behandelt de wijze waarop percelen worden geclassificeerd naar gevoeligheid voor de uitspoeling van nitraat naar het grondwater. Hoofdstuk 6 bediscussieert de betrouwbaarheid van de ontwikkelde methodiek. Theoretische achtergronden van de methodiek worden in de bijlagen gegeven.

1.4 Afkortingen en begrippen

In dit rapport worden een aantal afkortingen en begrippen gebruikt:

Grondwaterstand: De stijghoogte van het freatische grondwater ten opzichte van het maaiveld, gemeten in een boorgat of een peilbuis met ondiepe filterdiepte (in het algemeen minder dan 5 meter onder het maaiveld).

HG3 en LG3: Het gemiddelde van de drie hoogste, resp. laagste grondwaterstanden die in een hydrologisch jaar (1 april t/m 31 maart) worden gemeten, uitgaande van een halfmaandelijke meetfrequentie.

VG3: De gemiddelde grondwaterstand op de meetdata 14 maart, 28 maart en 14 april in een bepaald kalenderjaar.

Gemiddeld Hoogste Grondwaterstand, GHG: Het gemiddelde van de HG3 over een aaneengesloten periode van tenminste acht jaar waarin geen ingrepen hebben plaatsgevonden. In dit rapport zijn alle gepresenteerde GHG's berekend over 30 jaar (de klimaatperiode).

Gemiddeld Laagste Grondwaterstand, GLG: Het gemiddelde van de LG3 over een aaneengesloten periode van tenminste acht jaar waarin geen ingrepen hebben plaatsgevonden. In dit rapport zijn alle gepresenteerde GLG's berekend over 30 jaar (de klimaatperiode).

Gemiddelde VoorjaarsGrondwaterstand, GVG: Het gemiddelde van de VG3 over een aaneengesloten periode van tenminste acht jaar waarin geen ingrepen hebben plaatsgevonden. In dit rapport zijn alle gepresenteerde GVG's berekend over 30 jaar (de klimaatperiode).

GxG: Staat in dit rapport voor het drietal GHG, GVG en GLG.

Klimaatrepresentatieve GxG: De GxG zoals die berekend zou kunnen worden uit metingen in de volgende situatie:

- Vanaf -bijvoorbeeld- 1 april 2001 wordt op de 14^{ste} en 28^{ste} van elke maand de freatische grondwaterstand gemeten gedurende een periode van 30 jaar (tot en met 1 april 2031 dus).
- Gedurende deze periode verandert er niets aan het huidige peilbeheer, de inrichting van het watersysteem, het debiet van grondwateronttrekkingen et cetera (geen nieuwe menselijke ingrepen dus).

- De GxG wordt op basis van deze gegevens berekend (eerst per hydrologisch jaar de HG3 en LG3, daarna het 30-jaars gemiddelde van de HG3 en LG3, leidend tot GHG en GLG).

De aldus verkregen GxG representeert het effect van de gehele weersvariatie binnen de klimaatperiode van 30 jaar, gegeven de huidige ontwateringsituatie.

Grondwatertrap, Gt: Een typerende combinatie van GHG- en GLG-klassen welke op thematische kaarten kan worden weergegeven.

Duurlijn: Geeft aan welke totale tijdsduur binnen het jaar een bepaalde grondwaterstand wordt overschreden.

Regimecurve: Geeft aan wat de verwachte grondwaterstand is op een bepaalde datum in een toekomstig jaar.

Drooglegging: Het hoogteverschil tussen de waterspiegel in een waterloop en het grondoppervlak.

Kwel: Opwaartse grondwaterflux. In dit onderzoek wordt de kwel niet berekend uit het verschil tussen stijghoogten in ondiepe en diepe filters maar uit tijdreeksanalyse. De kwel of infiltratie wordt gepresenteerd in klassen om schijn-nauwkeurigheid te vermijden.

GrondwaterDynamiek, Gd: Een verzamelterm voor GxG, Gt, duurlijn, regime-curve en kwelklasse.

Stratificatie: Het onderverdelen van een gebied in hydrologisch homogene deelgebieden, ook wel strata genoemd. Met ‘hydrologisch homogeen’ wordt bedoeld: een vergelijkbare hydro-geologische en bodemkundige ondergrond. Soms wordt de landschappelijke ligging (beekdalen), het peilbeheer (grote polders) en het landgebruik (grote natuurgebieden met karakteristiek peilbeheer) bij de stratificatie betrokken.

Hoofdstuk 2

De methodiek in hoofdlijnen

2.1 Schematische weergave

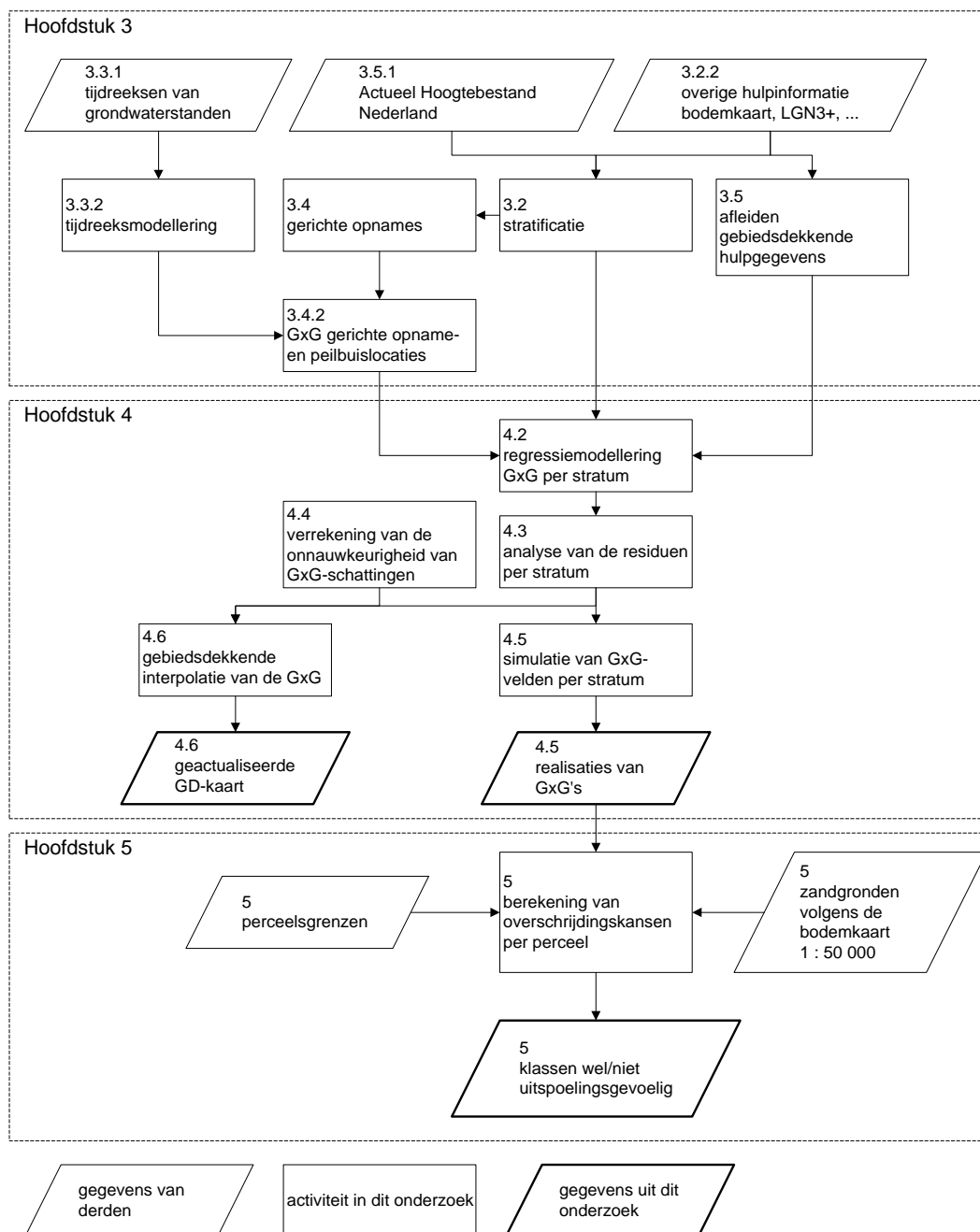
De ontwikkelde methodiek bestaat globaal gezien uit de volgende drie fasen;

Verzamelen van benodigde gegevens In deze eerste fase worden klimaatsrepresentatieve GxG gegevens verzameld op de locaties van een verdicht meetnet (via ‘gerichte opname’, tijdreeksanalyse en regressieanalyse), en wordt gebiedsdekkende hulp-informatie over topografie en waterhuishouding verzameld.

Geostatistische verwerking van de gegevens Als het doel is de GxG in kaart te brengen, dan wordt in deze fase de GxG ruimtelijk geïnterpoleerd tussen de meetpunten, gebruik makend van de gebiedsdekkende hulp-informatie. Als het doel is percelen te classificeren naar uitspoelingsgevoeligheid, dan worden in deze fase een groot aantal (hier 300) gebiedsdekkende GxG-vlakken gesimuleerd. De variatie tussen deze vlakken representeert de onzekerheid over de werkelijke GxG in het gebied.

Nabewerking van de geostatistische resultaten Als het doel is de GxG in kaart te brengen dan kan in deze laatste fase volstaan worden met het geïnterpoleerde GHG- en GLG-vlak om te zetten in een Gt-vlak en dit kartografisch weer te geven. Als het doel is percelen te classificeren, dan worden per perceel een drietal classificatie criteria toegepast op het betreffende gedeelte van de gesimuleerde GxG-vlakken.

De methodiek voor de actualisering van de grondwaterstands-informatie en de classificatie van uitspoelingsgevoelige percelen is globaal weergegeven in Figuur 2.1. De nummers verwijzen naar de paragrafen in hoofdstuk 3 tot en met 5, waarin een gedetailleerde beschrijving van de methoden wordt gegeven. De methodiek wordt hieronder globaal samengevat, waarbij de figuur van boven naar beneden wordt doorlopen.



Figuur 2.1. Globale werkwijze bij de classificatie van uitspoelingsgevoelige percelen

2.2 Gegevens over grondwaterstand, topografie en waterhuishouding

De volgende gegevens worden gebruikt:

- tijdreeksen van grondwaterstanden uit het OLGA- en OLGA/SUN-bestand (TNO-NITG);

- neerslag- en verdampingsgegevens (KNMI);
- het Actueel Hoogtebestand Nederland, AHN (Topografische Dienst);
- de bodem- en Gt-kaart, schaal 1 : 50 000 (Alterra);
- de bodem- en Gt-kaarten van detailkarteringen, schaal 1 : 10 000 en 1 : 25 000 (Alterra);
- de geologische kaart, schaal 1 : 50 000 (TNO-NITG);
- de geomorfologische kaart, schaal 1 : 50 000 (Alterra);
- de topografische kaart, schaal 1 : 10 000, het Top10-Vectorbestand (Topografische Dienst);
- het Landelijk Grondgebruiksbestand Nederland LGN4 (Alterra);
- de ligging van stroomgebieden (waterschappen);
- de ligging van gebieden met vernattingsprojecten (provincies of waterschappen).

Een belangrijk onderdeel van de methodiek is de verdeling van het gebied in hydrologisch homogene deelgebieden, de zogenaamde strata (paragraaf 3.2). Deze indeling in strata wordt zowel gebruikt bij het selecteren van tijdelijke meetpunten als bij de gebiedsdekkende berekening van GxG's en daarmee de GxG-kaart en de uiteindelijke classificatie van uitspoelingsgevoelige percelen.

Voor locaties waar grondwaterstandsreeksen zijn waargenomen in peilbuizen, kan de GxG worden geschat met behulp van tijdreeksanalyse (paragraaf 3.3.2). De dichtheid van het peilbuizennet is echter te laag voor het maken van nauwkeurige ruimtelijke voorspellingen van de GxG. Het net kan worden verdicht door het verrichten van zogenaamde gerichte opnames (paragraaf 3.4). Hierbij wordt op een groot aantal locaties tegelijkertijd de grondwaterstand gemeten, op tijdstippen dat deze zich rond het GHG- of GLG-niveau bevindt. Op het zelfde moment wordt ook in de peilbuizen de grondwaterstand gemeten. Er wordt vervolgens een regressiemodel bepaald dat de samenhang beschrijft tussen de GxG's die geschat zijn uit tijdreeksen en de metingen ten tijde van de gerichte opnames. Met dit regressiemodel wordt de GxG voorspeld voor de locaties van de gerichte opnames. De locaties van peilbuizen en gerichte opnames vormen tezamen het 'verdichte meetnet' van de GxG. De locaties van de gerichte opnames worden zodanig gekozen dat er in elk stratum tenminste twintig waarnemingen worden gedaan, gelijkmatig verdeeld over de *range* van droogleggingen (zie paragraaf 3.5.2). Naast een voorspelling van de GxG wordt voor elke locatie ook de nauwkeurigheid van deze voorspelling berekend. De dichtheid van het verdichte meetnet bedroeg in Gd-karteringen in de provincie Noord-Brabant één GxG-waarde per 100 à 110 hectare (zie bijvoorbeeld Finke et al. (2002)).

Het verdichte meetnet van de GxG geeft nog niet een gedetailleerd beeld van de ruimtelijke patronen. Daarom worden er gebiedsdekkende bestanden gemaakt met hulpinformatie die uit onder meer het AHN en de topografische kaart is afgeleid

(paragraaf 3.5.2). Voorbeelden hiervan zijn de drooglegging, de afstand tot waterlopen en de relatieve maaiveldhoogte. Ook het grondgebruik en de ‘oude’ Gt zijn als hulpinformatie gebruikt. Deze hulpinformatie is voor alle gridcellen van het AHN (25×25 m) bekend, en heeft dus een veel grotere dichtheid dan de op punten bepaalde GxG. Alle hulpbestanden samen worden verder in dit rapport aangeduid als de ‘hulpinformatie’.

De GxG-waarden op punten, hulpinformatie en strata vormen samen de gegevens waarmee de GxG gebiedsdekkend wordt berekend. Afhankelijk van het doel kunnen twee soorten berekening worden toegepast: *interpolatie* tussen meetpunten als een GxG-kaart moet worden gemaakt, of *simulatie* als percelen moeten worden geclassificeerd naar uitspoelingsgevoeligheid.

2.3 Gebiedsdekkende interpolatie van de GxG

Voor elk stratum wordt met behulp van regressieanalyse de samenhang onderzocht tussen de GxG's van het verdichte meetnet en de hulpinformatie. Dit levert per stratum een vergelijking op waarmee de GHG wordt voorspeld uit een combinatie van verschillende hulpgegevens, en een soortgelijke vergelijking voor de GLG. Omdat de hulpinformatie gebiedsdekkend is kan de GxG ook gebiedsdekkend worden voorspeld. Hierbij wordt gebruik gemaakt van een geostatistische techniek ('Universal Cokriging'), die in paragraaf 4.6 meer gedetailleerd wordt beschreven. Het resultaat wordt weergegeven als een GxG kaart.

2.4 Gebiedsdekkende simulatie van de GxG

Voor het maken van een GxG-kaart is één gebiedsdekkende voorspelling van de GxG voldoende. Als echter percelen moeten worden geclassificeerd naar een kenmerk (zoals wel of niet uitspoelingsgevoelig) dient, om op de juiste wijze rekening te houden met de onzekerheid in de GxG, een groot aantal mogelijke waarden (d.w.z. aselechte realisaties) van de GxG gebiedsdekkend berekend te worden door middel van geostatistische simulatie. Simulatie en interpolatie resultaten zijn onderling consistent in die zin dat, als een groot aantal gebiedsdekkende GxG's worden gesimuleerd, het gemiddelde daarvan gelijk is aan de geïnterpoleerde gebiedsdekkende GxG. Hoofdstuk 4 geeft een gedetailleerde beschrijving van het hoe en waarom van de simulatie.

2.5 Perceelsclassificatie naar uitspoelingsgevoeligheid

Zodra een groot aantal gebiedsdekkende GxG's zijn gesimuleerd, kunnen deze worden nabewerkt om percelen te classificeren naar uitspoelingsgevoeligheid. Het uitgangspunt is daarbij dat een perceel als uitspoelingsgevoelig wordt geclassificeerd indien er voldoende zekerheid bestaat dat tenminste een zeker percentage van het oppervlakte van het perceel uitspoelingsgevoelig is. Daartoe worden de volgende criteria gehanteerd:

- Puntcriteria: een punt in het perceel is gedefinieerd als uitspoelingsgevoelig indien op dat punt:
 - de bodem bestaat uit zand, en
 - de GHG dieper is dan GHG_{crit} cm beneden maaiveld (GHG_{crit} is bijvoorbeeld 40, 50, 60, 70 of 80), en
 - de GLG dieper is dan 120 cm beneden maaiveld.
- Oppervlaktecriterium: het perceel als geheel is gedefinieerd als uitspoelingsgevoelig, indien meer dan O_{crit} procent van zijn oppervlak voldoet aan het puntcriterium voor uitspoelingsgevoeligheid. (O_{crit} is bijvoorbeeld 50, 67 of 80.)
- Kanscriterium: het perceel wordt als uitspoelingsgevoelig geclassificeerd, indien de uit de simulaties te berekenen kans dat aan het oppervlakte-criterium is voldaan, groter dan P_{crit} is. (P_{crit} is bijvoorbeeld 0.90 of 0.95.)

De perceelsclassificatie vindt plaats door per perceel voor elke gesimuleerde GxG het percentage van het oppervlak te berekenen waar voldaan is aan de punt-criteria voor uitspoelingsgevoeligheid, en vervolgens te bepalen of dit percentage O_{crit} overschrijdt. Tenslotte wordt de fractie GxG's berekend waarin O_{crit} wordt overschreden. Als deze fractie groter is dan P_{crit} , dan wordt het perceel als uitspoelingsgevoelig geclassificeerd, anders als niet-uitspoelingsgevoelig.

Hoofdstuk 3

Gegevens over grondwaterstand, topografie en waterhuishouding

3.1 Inleiding

Dit hoofdstuk beschrijft de wijze waarop uit verschillende informatiebronnen de gegevens worden afgeleid die als basis dienen voor de gebiedsdekkende interpolatie en/of simulatie van de GxG. Een belangrijk onderdeel van het onderzoek is de indeling in homogene deelgebieden. Deze stratificatie speelt bij de verzameling van grondwaterstandswaarnemingen al een rol. Daarom volgt eerst een beschrijving van de stratificatie in paragraaf 3.2. Vervolgens komt in paragraaf 3.3 de selectie van peilbuizen aan de orde, en de berekening van de GxG uit de tijdreeksen die in deze peilbuizen zijn waargenomen. Paragraaf 3.4 gaat in op de selectie van de locaties waar de zogenaamde gerichte opnames worden verricht, en geeft een beschrijving van de wijze waarop voor deze locaties GxG's worden berekend. In paragraaf 3.5 volgt tenslotte een beschrijving van de gebiedsdekkende hulpinformatie over de topografie en de waterhuishouding.

3.2 Indeling van het gebied in homogene deelgebieden

3.2.1 Aard en doel van de gebiedsindeling

Het onderscheiden van homogene deelgebieden wordt stratificatie genoemd. Het doel van de stratificatie is om de onzekerheid over de gebiedsdekkende voorspellingen van de GxG te reduceren. Er mag namelijk worden verondersteld dat binnen homogene deelgebieden de samenhang tussen de GxG en allerlei hulpinformatie sterker zal zijn dan in het gebied als geheel. Er worden deelgebieden (strata) onderscheiden op basis van eigenschappen die gerelateerd zijn aan hydrologie, maaiveldshoogte en bodemopbouw. Voor elk van de strata wordt de samenhang tussen de GxG enerzijds en variabelen die aan de maaiveldshoogte zijn gerelateerd anderzijds beschreven met een regressiemodel. Dit regressiemodel wordt gebruikt bij de gebiedsdekkende voorspelling van de GxG. De regressiemodellen kunnen tussen de afzonderlijke strata aanzienlijk verschillen.

3.2.2 Gegevens voor de stratificatie

Bij de stratificatie wordt de volgende gebiedsdekkende informatie gebruikt:

- de geologische kaart van Nederland, schaal 1 : 50 000 (TNO-NITG);
- de geomorfologische kaart van Nederland, schaal 1 : 50 000 (Alterra);
- het Actueel Hoogtebestand Nederland, AHN 25 × 25 meter (Topografische Dienst);
- de ligging van waterlopen volgens de topografische kaart van Nederland, schaal 1 : 10 000, Top10-Vector (Topografische Dienst);
- de Landelijke Grondgebruikskaat Nederland, LGN4, 25 × 25 meter (Alterra);
- de bodem- en Gt-kaart van Nederland, schaal 1 : 50 000 (Alterra).

Verder worden er gegevens gebruikt die niet altijd gebiedsdekkend beschikbaar zijn:

- de indeling in stroomgebieden en peilgebieden volgens de waterschappen;
- gedetailleerde bodemkaarten, schaal 1 : 10 000 of schaal 1 : 25 000, die zijn vervaardigd voor bijvoorbeeld landinrichtingsprojecten (Alterra);
- de locaties van droogvallende waterlopen (waterschappen);
- de ligging van gebieden met vernattingsprojecten (waterschappen of proveniencies).

3.2.3 Werkwijze bij de stratificatie

De bodem- en Gt-kaart van Nederland, schaal 1 : 50 000, vormt de basis voor de stratificatie. Indien nodig worden de kaartvlakken aan de hand van additionele informatie gesplitst. De stratificatie verloopt in drie stappen:

1. Er worden geohydrologische hoofdeenheden onderscheiden op basis van dagzomende geologische formaties en een aantal breuklijnen die op de geologische kaart zijn weergegeven;
2. De geohydrologische hoofdeenheden worden onderverdeeld in bodemkundig-hydrologische eenheden, op basis van de bodem- en Gt-kaart, en gegevens over het afwateringspatroon en de maaiveldshoogte. Hierbij wordt gelet op:
 - verschillen in droog en nat volgens de Gt-kaart;
 - de aanwezigheid van leemlagen in de ondergrond;
 - de textuur van de bodem;
 - de aanwezigheid van grof zand in de ondergrond;
 - de intensiteit van afwateringspatroon.

Voorbeelden van deze bodemkundig-hydrologische eenheden zijn:

- beekdalen;
- droge zandgronden;
- lemige gronden met stagnatie van grondwater door leemlagen.

Eén eenheid kan uit meerdere kaartvlakken bestaan en de oppervlakte van de vlakken loopt in de praktijk uiteen van minder dan honderd tot bijna 10 000 ha (Finke et al., 2002). De bodemkundige informatie die gebruikt wordt bij stap 2 is voornamelijk ontleend aan de bodem- en Gt-kaart, schaal 1 : 50 000. Als er echter digitaal kaartmateriaal van detailkarteringen (1 : 10 000 of 1 : 25 000) beschikbaar is, dan wordt dit bij de stratificatie gebruikt.

3. In deze stap worden de uiteindelijke strata gevormd. Vlakken van bodemkundig-hydrologische eenheden worden samengevoegd tot strata, om tot meer aaneengesloten strata te komen en versnippering tegen te gaan en daarmee het risico van kunstmatige discontinuïteiten op stratumgrenzen beperkt te houden. De strata zijn gemiddeld 3 000 hectare groot (Finke et al., 2002). In gebieden met grote bodemkundig-hydrologische verschillen kunnen strata echter kleiner zijn, terwijl in homogene gebieden de strata groter zijn.

Als een waterschap de opdracht heeft gegeven wordt het resultaat van stap 3 eventueel met deskundigen van het betreffende waterschap besproken en eventueel bijgesteld. Ook de tweede versie van de stratificatie wordt aan de deskundigen van het waterschap voorgelegd. Na een eventuele tweede bijstelling wordt de stratificatie vastgelegd.

De strata zijn bodemkundig-hydrologisch uniforme deelgebieden en stratumgrenzen behoeven niet samen te vallen met perceelsgrenzen. Een perceel kan dus uit delen van meerdere strata bestaan. De implicaties hiervan worden besproken in Bijlage C.

3.3 GxG's op peilbuislocaties

3.3.1 Selectie van grondwaterstandsreeksen

De grondwaterstandsreeksen komen voor het overgrote deel uit de OLGA-databank van TNO-NITG (Van Bracht, 1988). Vanaf 1 januari 2001 heet deze databank *DINO*. Daarnaast worden korte meetreeksen van grondwaterstanden aangeleverd door de waterschappen en provincies. Hierna wordt beschreven hoe grondwaterstandsreeksen worden geselecteerd die geschikt zijn voor de berekening van GxG's.

OLGA/DINO

De benodigde reeksen worden geselecteerd uit alle reeksen van peilbuislocaties op een topografisch kaartblad, schaal 1 : 25 000, uit het OLGA/DINO-bestand, inclusief de reeksen uit het OLGA-SUN-deel van het archief. Dit OLGA-SUN-bestand bevat grondwaterstanden die zijn verzameld in de terreinen van het Staatsbosbeheer, de Unie van Landschappen en Natuurmonumenten. In deze terreinen komen relatief veel extreem natte en extreem droge situaties voor. In het OLGA/DINO-bestand

wordt onderscheid gemaakt tussen landbouwbuizen (L-buizen) en peilbuizen (P-buizen). Daar waar in dit rapport wordt gesproken over peilbuizen worden zowel de landbouw- als de peilbuizen bedoeld.

De selectiecriteria zijn:

1. de reeksen moeten zijn waargenomen in peilbuizen met een filter dat maximaal vijf meter beneden maaiveld eindigt;
2. de grondwaterstanden moet gemeten zijn t.o.v. maaiveld;
3. de plaatscoördinaten van de buislocatie moeten bekend zijn;
4. er moeten minimaal negen waarnemingen per hydrologisch halfjaar zijn verricht;
5. de meetreeks moet minimaal drie jaar lang zijn;
6. de meetreeks moet doorlopen tot minimaal een half jaar voor het begin van de kartering;
7. de afstand van de peilbuis tot een rivier of kanaal moet tenminste 40 m zijn;
8. de afstand van de peilbuis tot waterlopen en beken moet tenminste 20 m zijn;
9. de afstand van de peilbuis tot een watervoerende perceelssloot moet tenminste 8 m zijn;
10. het meetpunt mag niet worden beïnvloed door afstromend oppervlaktewater (bijvoorbeeld nabij verharding), zich bevinden in een kuil, op een dijk, een oprit, en dergelijke.

De criteria 7 tot en met 10 worden getoetst in het veld.

Overige grondwaterstandsreeksen

Waterschappen en provincies exploiteren een aantal grondwaterstandsmeetnetten. Over het algemeen zijn deze meetnetten er tijdelijk, voor een specifiek doel. De gegevens die door waterschappen en provincies worden aangeleverd moeten tenminste voldoen aan criteria 1, 2 en 3 hierboven. Afhankelijk van de waarnemingsfrequentie (criterium 4), de reekslengte (criterium 5) en de waarnemingsperiode (criterium 6) wordt besloten op welke wijze uit de gegevens de GxG wordt berekend. Als de gegevens voldoen aan de criteria die hierboven zijn gesteld dan wordt de GxG berekend met behulp van tijdreeksmodellering (zie paragraaf 3.3.2). Als de reeksen te kort zijn voor tijdreeksmodellering, dan wordt de methode van regressiemodellering gevolgd die is beschreven door Oude Voshaar en Stolp (1997). Als ook daarvoor het aantal waarnemingen te klein is worden de gegevens gebruikt als gerichte opnames van de grondwaterstand en wordt de methode zoals beschreven in paragraaf 3.4 gevolgd. (Deze laatste variant komt in de praktijk het meeste voor.)

In tegenstelling tot de peilbuizen uit het OLGA/DINO-bestand worden de buizen uit de lokale meetnetten niet in het veld bezocht. Dit is een gevolg van de korte tijd tussen het beschikbaar komen van de gegevens en de kartering.

3.3.2 Berekening van klimaatrepresentatieve GxG's op peilbuislocaties

Voor de berekening van klimaatrepresentatieve GxG's voor peilbuislocaties worden tijdreeksmodellen gebruikt die de samenhang tussen het neerslagoverschot en de grondwaterstand beschrijven. Dit is nodig omdat maar zelden gedurende dertig jaar grondwaterstanden worden gemeten op één locatie, zonder dat het hydrologische regime in die periode wijzigt door ingrepen in de waterhuishouding. Het is gebleken dat op basis van uitsluitend korte reeksen een vertekend (systematisch te nat of te droog) beeld van de grondwatersituatie kan ontstaan ten opzichte van de klimaatperiode van dertig jaar. Daarom worden er tijdreeksmodellen gekalibreerd die de samenhang beschrijven tussen het potentiële neerslagoverschot en de grondwaterstand. Deze modellen worden vervolgens gebruikt om op basis van reeksen van neerslagoverschotten die door het KNMI gedurende tenminste dertig jaar zijn verzameld grondwaterstandsreeksen te simuleren. Uit deze gesimuleerde grondwaterstandsreeksen van dertig jaar kunnen klimaatsrepresentatieve kenmerken zoals de GxG en Gt worden berekend, en duurlijnen en regimecurves worden geconstrueerd (Knotters en van Walsum, 1994). Uit een analyse van Knotters en Bierkens (1999) blijkt dat reeksen van vier tot acht jaar meestal lang genoeg zijn om de samenhang tussen het neerslagoverschot en de grondwaterstand te kunnen modelleren, en dat ook reeksen met lengte van drie jaar in veel situaties nog zullen voldoen. Het potentiële neerslagoverschot wordt berekend uit de verschil tussen de etmaalsommen neerslag en de potentiële etmaalverdamping voor een referentiegewas volgens Makink (De Bruin, 1987) (zie A). Voor de periode dat de etmaalverdamping volgens Makink nog niet werd gepubliceerd, is deze gereconstrueerd door het KNMI.

De klimaatsrepresentatieve GxG's op peilbuislocaties worden berekend met behulp van een tijdreeksmodel en een reeks neerslagoverschotten van tenminste dertig jaar. Veronderstel nu dat het neerslagoverschot over de laatste dertig jaar in gemiddelde en temporele variatie gelijk is aan het neerslagoverschot over de komende dertig jaar. De klimaatsrepresentatieve GxG is dan voor te stellen als de GxG zoals die over dertig jaar zou worden berekend als vanaf heden gedurende 30 jaren de grondwaterstanden zouden worden gemeten op de 14^{de} en de 28^{ste} van elke maand, en als er geen ingrepen in de grondwatersituatie zouden plaatsvinden die buiten het huidige peilbeheer vallen.

Met het gekalibreerde tijdreeksmodel en de neerslagoverschotreeks van de dertig jaar voorafgaand aan de kartering wordt voor elke peilbuislocatie een groot aantal (100) grondwaterstandsreeksen gesimuleerd van dertig jaar lang. Uit deze 100 gesimuleerde reeksen wordt de gemiddelde GxG en zijn standaardfout berekend.

Bijlage A geeft een gedetailleerde beschrijving van het transfer-ruismodel en de berekening daarmee van klimaatrepresentatieve GxG's.

3.4 GxG's op tijdelijke meetpunten

3.4.1 Gerichte opnames

Tijdens de kartering worden op twee momenten grondwaterstanden gemeten in een groot aantal boorgaten, op vooraf vastgestelde locaties. Omdat zowel de tijdstippen als de locaties van de opnames vooraf worden bepaald, wordt gesproken van twee 'gerichte opnames'. Het doel van deze gerichte opnames is om het net van puntgegevens te verdichten, zodat er uiteindelijk voldoende GxG-data zijn om de statistische relatie met maaiveldhoogten en daarvan afgeleide eigenschappen te kunnen bepalen (paragraaf 4.2). Hiertoe is het nodig dat in elk stratum van tenminste twintig locaties GxG-waarden bekend zijn. In de grotere strata is dat aantal groter; gemiddeld zijn voor dertig locaties per stratum GxG-waarden bekend. Een aanvullende eis is dat deze locaties gelijkmatig over het stratum en de droogleggingen die in het stratum voorkomen zijn verspreid. In de praktijk leidt het bovenstaande tot een gemiddelde van één locatie per 110 hectare waarvoor GxG-waarden bekend zijn (Finke et al., 2002). De locaties worden als volgt gekozen:

1. voor elk stratum wordt bepaald hoeveel waarnemingen er moeten worden gedaan: gemiddeld 30 maar tenminste 20;
2. met het Top10-Vectorbestand en het AHN wordt een kaart van de drooglegging gemaakt met voor elk 25×25 m-pixel een drooglegging, zie paragraaf 3.5 voor een beschrijving van de werkwijze;
3. de droogleggingen worden per stratum gesorteerd van nat (ondiep) naar droog (diep);
4. de lijst met gesorteerde droogleggingen wordt in een aantal klassen gesplitst, evenveel als het aantal waarnemingen dat zal worden verricht in het stratum. Deze klassen zijn kleiner bij 'natte' dan bij 'droge' droogleggingen. Zo wordt bewerkstelligd dat er voldoende waarnemingen in de relatief natte terreindelen van het stratum worden gedaan, dit met het oog op de voor de meeste toepassingen gewenste nauwkeurigheid voor die terreindelen;
5. per klasse wordt aselekt één waarnemingslocatie gekozen, en twee reservelocaties op tenminste 250 m daarvan verwijderd.

Er wordt tweemaal gemeten: éénmaal in de winter, als de grondwaterstand zich rond het GHG-niveau bevindt, en éénmaal in de zomer, als de grondwaterstand zich rond het GLG-niveau bevindt. In regenperiodes wordt niet gemeten, omdat dan de diepte van het freatisch vlak te veel varieert. Tijdens het veldwerk worden boorgaten gemaakt tot 10 cm onder het grondwaterniveau (maar nooit dieper dan 2,50 m). Na een instelperiode (1-2 dagen) wordt de grondwaterstand gemeten.

Additionele opnames

Gegevens van gerichte opnames die in het recente verleden ten behoeve van andere projecten hebben plaatsgevonden worden bij het onderzoek betrokken. De gegevens

zijn afkomstig uit detailkarteringen (schaal 1 : 10 000 en schaal 1 : 25 000) van na 1990. Het gaat hierbij om korte meetreeksen (in het algemeen circa één jaar) en om gerichte opnames. De locaties van de meetreeksen en opnames worden, voor zover nodig, vanaf gepubliceerde kaarten gedigitaliseerd en de gemeten grondwaterstanden worden ingevoerd. De gegevens worden op dezelfde wijze verwerkt als de grondwaterstandsmetingen die in het kader van de Gd-actualisatie worden gedaan.

3.4.2 Berekening van GxG's op gerichte-opnamelocaties

Stap 1: bepaling van het regressiemodel

Op de tijdstippen van de gerichte opnames worden ook grondwaterstanden gemeten in de peilbuizen waarvan de GxG is berekend (paragraaf 3.3.2). Deze peilbuizen liggen idealiter in de directe omgeving van de meetpunten, en vertegenwoordigen samen alle grondwatertrappen. In de praktijk worden er op één meetdag in uitgestrekte gebieden metingen verricht, en liggen de peilbuizen die samen alle grondwatertrappen vertegenwoordigen ook over een groot gebied verspreid. Het aantal peilbuizen waarin per meetdag wordt gemeten varieert tussen de vijftien en de vijfentwintig. Dit is voldoende om voor elke meetdag een regressiemodel op te stellen dat de relatie beschrijft tussen de waargenomen grondwaterstanden en de berekende GxG. Het regressiemodel kan twee vormen hebben:

1. een lineair model dat de relatie beschrijft tussen de GxG en de grondwaterstand die is waargenomen tijdens de gerichte opname die plaatsvond rond GxG-niveau:

$$GxG = \beta_0 + \beta_1 x_1 + \epsilon,$$

waarin x_1 de grondwaterstand rond GHG-niveau of de grondwaterstand rond GLG-niveau is, β_0 en β_1 regressiecoëfficiënten en ϵ de foutenterm;

2. een lineair model dat de relatie beschrijft tussen enerzijds de GxG en anderzijds de grondwaterstand die is waargenomen rond GHG- en GLG-niveau:

$$GxG = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon,$$

waarin x_1 de grondwaterstand rond GHG-niveau, x_2 de grondwaterstand rond GLG-niveau is, β_0 , β_1 en β_2 regressiecoëfficiënten zijn en ϵ de foutenterm.

Het model dat het meest geschikt is voor de voorspelling van de GxG wordt geselecteerd op basis van het kleinste Mallow's C_p -criterium (Thompson, 1978; Miller, 1990):

$$C_p = SS_{\text{res}}/s^2 - n + 2p, \quad (3.1)$$

waarin SS_{res} de som is van de gekwadrateerde verschillen tussen de waargenomen GxG-waarden en de regressievoorspellingen, s^2 de residuele variantie is van het volledige model (met alle parameters, dus in dit geval met β_0 , β_1 en β_2), en p het aantal modelparameters is (in dit geval $p = 2$ of $p = 3$).

Stap 2: toepassing van het regressiemodel

Met het geselecteerde model wordt de GxG voorspeld uit grondwaterstandswaarnemingen die tijdens de gerichte opname zijn verricht. Tevens worden de standaardfouten van deze voorspellingen berekend.

Bij de regressievoorspellingen kan zich de situatie voordoen dat de grondwaterstand niet wordt waargenomen binnen de maximale boordiepte van 2,50 meter. Dit is een zogenaamde ‘gecensureerde waarneming’: de precieze waarde is onbekend, maar wel is bekend dat de waarneming ‘dieper dan’ een grenswaarde is. Wij kiezen voor de volgende benadering (Cohen, 1991): eerst wordt een *maximum likelihood*-schatting gemaakt van de grondwaterstand. Daarbij wordt de code ‘> 2,50’ vervangen door de meest waarschijnlijke diepte groter dan 2,50 m. Dit getal wordt vervolgens ingevoerd in de regressievergelijking. De check op gecensureerde waarnemingen en de vervanging met het meest waarschijnlijke getal is standaard ingebouwd in het computerprogramma waarmee GxG’s worden voorspeld.

3.5 Gebiedsdekkende hulpinformatie over topografie en waterhuishouding

3.5.1 Voorbewerking van het AHN

Er wordt gebruik gemaakt van de meest recente versie van het Actueel Hoogtebestand Nederland van de Topografische Dienst, met een resolutie van 25×25 m. Dit bestand wordt getoetst voordat de hulpinformatie ervan wordt afgeleid. In het bestand blijken soms hoogten voor te komen die toe te schrijven zijn aan bijvoorbeeld schuren, huizen, wegen en viaducten. Deze ‘onnatuurlijke’ hoogten worden verwijderd. Allereerst worden alle cellen die behoren tot de LGN4-klassen ‘zoet water’, ‘zout water’, ‘stedelijk bebouwd gebied’ en ‘hoofdwegen en spoorwegen’ verwijderd. Vervolgens worden de standaardafwijkingen berekend van de hoogten in vensters van 3×3 cellen. Cellen worden verwijderd als deze standaardafwijking groter is dan 100 cm, en de cellen behoren tot één van de volgende klassen van het LGN4-bestand: ‘bebouwing in agrarisch en buitengebied’, ‘loof- en naaldbos in bebouwd gebied’, ‘bos met dichte bebouwing’, ‘gras in bebouwd gebied’, ‘kale grond in bebouwd buitengebied’. Als de standaardafwijking groter is dan 100 cm en de cellen liggen minder dan 50 m verwijderd van ‘hoofdwegen en spoorwegen’ dan worden deze ook verwijderd.

Binnen gebieden waarvan de cellen bij de filtering zijn verwijderd, kunnen losse cellen met een hoogtecijfer, of clustertjes van deze cellen voorkomen. Als het aantal cellen in zo’n cluster kleiner dan twintig is, dan wordt het uit het bestand verwijderd.

In het AHN-bestand blijken clusters van cellen met hoogtecijfers te ontbreken. Voor zover deze leemten in het bestand door de filtering ontstaan die hierboven is beschreven, worden zij uiteraard niet opgevuld. Om het aantal ontbrekende waarnemingen te beperken, worden de clusters van minder dan twintig cellen met ontbrekende informatie opgevuld met de gemiddelde hoogte van de 25 omliggende cellen. Gerekend

over het totaal van de landbouwpercelen beslaan de resterende leemten minder dan 1 % van de cellen.

3.5.2 Afleiding van hulpinformatie

Van het AHN bestand, de 1 : 50 000 Bodem- en Gt-kaart en het Top10-Vector bestand worden vijf groepen hulpinformatie afgeleid. Elk van deze groepen hulpinformatie bevat één of meer kaarten met daarop variabelen die hydrologisch relevant zijn. Variabelen die onderling een sterke correlatie vertonen zijn samengebracht in één groep. De inhoud en de afleiding van de gebiedsdekkende hulpinformatie wordt hieronder voor elk van de vijf groepen beschreven. Het AHN bestand tezamen met de hulpbestanden wordt hier aangeduid als ‘de hulpinformatie’.

Groep 1: relatieve maaiveldhoogten

Uit onderzoek van Te Riele en Brus (1992) en Te Riele et al. (1995) is gebleken, dat grondwaterstanden kunnen samenhangen met de maaiveldhoogte ten opzichte van NAP en met de relatieve maaiveldhoogte. Met relatieve maaiveldhoogte wordt de hoogte van een punt ten opzichte van de gemiddelde hoogte in een gebied binnen een bepaalde straal rond dat punt bedoeld (zie Figuur 3.1). Voor elk punt in het AHN wordt voor omgevingen met een straal van 100, 200, 300, 400 en 500 meter de relatieve maaiveldhoogte bepaald.

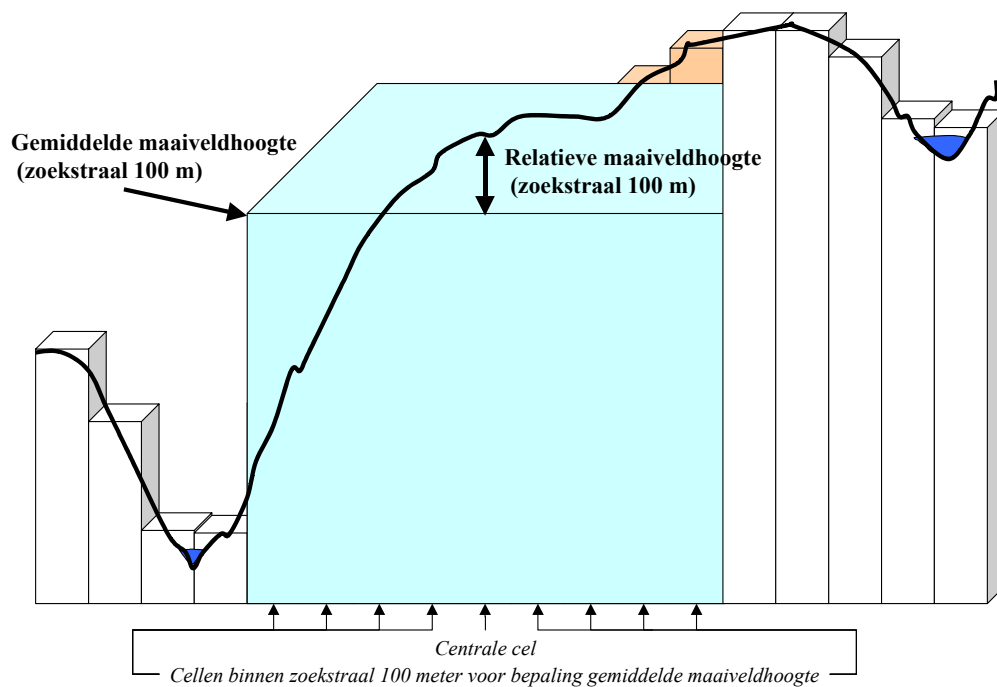
Groep 2: drainagedichtheid

De dichtheid waarmee een gebied is ontwaterd beïnvloedt de grondwaterstand (met name de GHG, maar indien er sprake is van waterinlaat ook de GLG). Daarom wordt de drainagedichtheid gebiedsdekkend geschat en gebruikt als hulpinformatie. Uit het Top10-Vectorbestand worden alle watergangen geselecteerd. Hiervan worden twee bestanden gemaakt: een lijnenbestand met alle watergangen (‘sloot en greppel’) en een bestand waar de detailontwatering uit is verwijderd (‘alleen sloot’). Voor elke 25×25 meter cel wordt bepaald hoeveel naburige cellen met een waterloop er voorkomen binnen een zoekstraal van 100 meter (Figuur 3.2). Hieruit volgt een indicatie van de drainagedichtheid voor ‘alleen sloot’ en ‘sloot en greppel’.

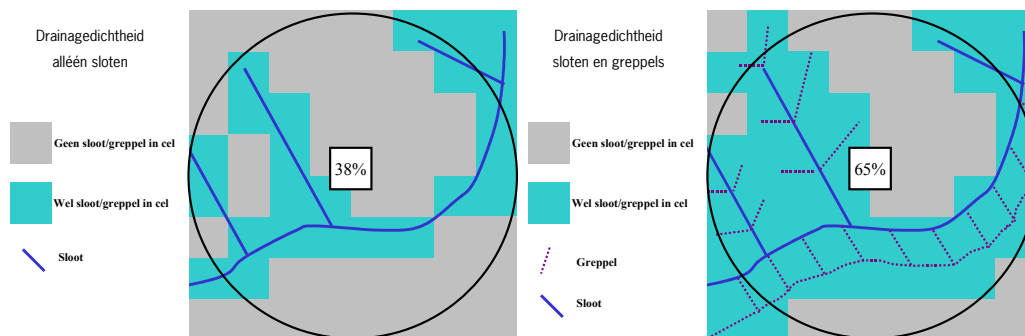
Groep 3: drooglegging ten opzichte van maaiveld

De drooglegging is het hoogteverschil tussen de waterspiegel in een waterloop en het grondoppervlak. Het kan worden geïnterpreteerd als het verwachte effect van het peilbeheer en de waterlopen-infrastructuur op de grondwaterstand. Een gebiedsdekkende schatting van de drooglegging zal daarom naar verwachting nuttige hulpinformatie opleveren bij een grondwaterkartering.

Uit een combinatie van de gedigitaliseerde waterlopen uit het Top10-Vectorbestand en het AHN wordt een bestand afgeleid waar per 25×25 meter pixel de drooglegging wordt bepaald (Figuur 3.3). Dit gebeurt in drie stappen:

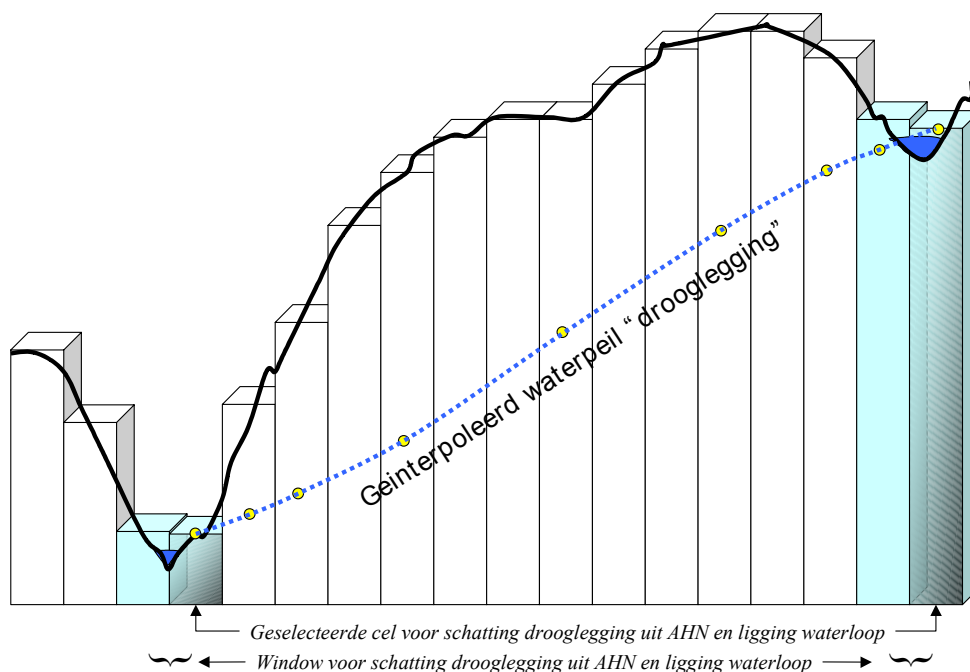


Figuur 3.1. Bepaling relatieve maaiveldhoogte uit het AHN



Figuur 3.2. Bepaling van de drainagedichtheid

1. in elk segment van een watergang wordt een peil ten opzichte van NAP geschat met de waarde van het laagst gelegen punt van het AHN in de directe omgeving van de watergang;
2. dit peil ten opzichte van NAP wordt gebiedsdekkend geïnterpoleerd, gewogen naar de inverse afstand tot de waterloop;



Figuur 3.3. Schatting van de drooglegging

3. door deze geïnterpoleerde peilen af te trekken van de maaiveldhoogte uit het AHN wordt de drooglegging ten opzichte van het maaiveld gebiedsdekkend berekend.

Waarschijnlijk geeft de drooglegging die op bovenstaande wijze wordt geschat een systematisch te nat beeld van de drooglegging, omdat het peil geschat in stap 1 over het algemeen op hoogtecijfers van de oevers zal zijn gebaseerd, en niet op het waterpeil zelf.

Gedurende het veldwerk wordt op een groot aantal locaties het slootpeil ten opzichte van de lokale maaiveldhoogte gemeten (de feitelijke drooglegging). Het verschil tussen de feitelijke en de geschatte drooglegging wordt gebiedsdekkend geïnterpoleerd en opgeteld bij de kaart van geschatte droogleggingen. Zowel deze gecorrigeerde droogleggingkaart als de kaart met oorspronkelijke schattingen van de drooglegging worden als hulpinformatie gebruikt in het vervolg van het onderzoek.

Groep 4: maaiveld ten opzichte van NAP

Dit is de absolute maaiveldhoogte volgens het 25×25 meter AHN bestand.

Tabel 3.1. *Omzetting van de Gt op de bodemkaart 1 : 50 000 naar GHG en GLG*

Gt op kaart	GHG _{boven}	GHG _{onder}	GLG _{boven}	GLG _{onder}
I	-17	40	26	50
II	-9	40	50	80
II*	25	50	50	80
III	0	40	80	120
III*	25	40	80	120
IV	40	120	80	120
V	0	40	120	150
V*	25	40	120	160
VI	40	80	120	190
VII	80	120	160	260
VII*, VIII	140	220	160	400

Groep 5: de GHG en GLG volgens de Gt-kaart 1 : 50 000 en de geschatte bergingscapaciteit

De Gt-kaart 1 : 50 000 is weliswaar verouderd wat betreft de absolute niveaus van de GHG en GLG, maar geeft de ruimtelijke variatie mogelijk nog wel goed weer. Daarom zullen mogelijk de ‘GHG-oud’ en ‘GLG-oud’, afgeleid uit de Gt-kaart 1 : 50 000, de actuele GxG goed kunnen voorspellen. Om die reden wordt de 1 : 50 000 Gt-kaart omgezet in kaarten van de GHGoud en GLGoud, waarbij gebruik wordt gemaakt van de karakterisatie van Gt’s door Van der Sluijs (1982, 1990), zie Tabel 3.1.

De Gt-klassen van de 1 : 50 000 bodem- en Gt-kaart en het AHN hoogtebestand worden gebruikt om een neergeschaalde kaart van de GHG en GLG te maken. Voor elk Gt-vlak van de 1 : 50 000 kaart wordt het 15^{de} en 85^{ste} percentiel van de AHN-hoogten in dat vlak bepaald. Voor de hoogten tussen het 15^{de} en 85^{ste} percentiel wordt verondersteld dat deze lineair samenhangen met de GHG’s en GLG’s tussen de onder- en bovengrens voor de betreffende Gt-klasse. Vervolgens worden de hoogten uit het AHN-bestand met deze lineaire relatie getransformeerd in GHG’s en GLG’s. Voor de locaties waarvan de hoogte niet in het interval van het 15^{de} tot het 85^{ste} percentiel ligt, wordt een GHG en GLG geïnterpoleerd t.o.v. NAP, vanuit de overige punten die wel in het interval vallen. Vervolgens zijn deze GHG’s en GLG’s teruggerekend naar diepten t.o.v. maaiveld, met behulp van de hoogten uit het AHN-bestand.

Met behulp van de bodemfysische vertaling van de bodemkaart (Wösten et al., 1988), het programma CAPSEV (Wesseling, 1991) en de grondwaterstanden in Tabel 3.1 is de bergingscapaciteit bij GHGoud en GLGoud berekend. Deze bergingscapaciteiten zijn eveneens als gebiedsdekkende hulpinformatie gebruikt.

Hoofdstuk 4

Gebiedsdekkende simulatie en interpolatie van de GxG

4.1 Inleiding

Voor elk stratum wordt met behulp van regressieanalyse de samenhang onderzocht tussen de GxG's van het verdichte meetnet en de hulpinformatie. Dit levert per stratum één vergelijking op waarmee de GHG wordt voorspeld uit een combinatie van verschillende hulpgegevens, en een soortgelijke vergelijking waarmee de GLG wordt voorspeld. Met behulp van deze vergelijkingen, de hulpinformatie en de GxG's op puntlocaties, kan de GxG gebiedsdekkend worden voorspeld via geostatistische *interpolatie*. Dit is voldoende als het doel is een GxG kaart van het gebied te maken. Voor perceelsclassificatie naar uitspoelingsgevoeligheid is één voorspelling echter niet voldoende. De reden daarvoor is dat deze classificatie wordt gebaseerd op de kans dat het perceel voor méér dan een bepaald percentage van het oppervlak uitspoelingsgevoelig is (zie Hoofdstuk 5). Om deze kans correct te bepalen dient een groot aantal mogelijke waarden van de GxG gebiedsdekkend te worden berekend door middel van geostatistische *simulatie*. De methode van Gd-kartering zoals deze is beschreven door Finke et al. (2002) is voor dit doel aangepast.

De gebiedsdekkende simulatie en interpolatie van de GxG vindt plaats in een aantal stappen, die weergegeven zijn in Figuur 4.1. Allereerst wordt de samenhang tussen de GxG op puntlocaties en de acht groepen gegevens uit de hulpinformatie onderzocht met behulp van regressieanalyse. De analyses worden per stratum uitgevoerd. Dit levert voor elk stratum een selectie van verklarende variabelen op, een gekalibreerd regressiemodel en een aantal residuen (verschilwaarden tussen de GxG's op puntlocaties en de gefitte waarden). De regressieanalyse wordt beschreven in paragraaf 4.2.

Vervolgens wordt, afzonderlijk voor de GHG, GVG en GLG, de ruimtelijke structuur van de residuen bepaald. Dit gebeurt niet per stratum, maar voor alle strata tegelijk, teneinde over voldoende gegevens te beschikken om de ruimtelijke structuur te kunnen modelleren. Daarom worden de residuen eerst gestandaardiseerd, zodat de variantie van de residuen voor alle strata even groot is (en gelijk aan één).

Op basis van de gestandaardiseerde residuen wordt een model voor de ruimtelijke structuur (variogram) geschat (zie paragraaf 4.3). Dit variogram wordt vervolgens gestandaardiseerd, dat wil zeggen dat het wordt geschaald naar de oorspronkelijke varianties van de residuen in de strata. Dit levert voor elk stratum een variogram op voor de residuen van GHG, GVG en GLG. Om rekening te houden met onderlinge correlaties worden tevens z.g. kruisvariogrammen berekend van de residuen van GHG en GVG, GHG en GLG, en GVG en GLG.

Zodra de (kruis)variogrammen beschikbaar zijn kan interpolatie en/of simulatie worden uitgevoerd. De interpolatie wordt besproken in paragraaf 4.6. De feitelijke simulatie van GHG- en GLG-velden wordt beschreven in paragraaf 4.5. Eerst wordt uit de kansverdeling van de geschatte regressieparameters een set parameterwaarden geloot. Vervolgens wordt er een punt uit het 25×25 -meter grid geloot, waarvoor *simultaan* een GHG en een GLG wordt gesimuleerd, op basis van de gelote parameterwaarden, de (kruis)variogrammen en de posities en waarden van de GHG en GLG op meetlocaties. Hierbij wordt rekening gehouden met de correlatie tussen de GHG en de GLG. De gesimuleerde waarden worden toegevoegd aan de dataset en in het vervolg van de simulatie meegenomen als extra (pseude-)meetwaarden. Vervolgens wordt voor een volgend geloot punt op dezelfde wijze een GHG en een GLG gesimuleerd. Dit wordt net zolang herhaald tot alle punten een keer geloot zijn; denkbeeldig neemt de set GxG's op meetlocaties dus toe. Het resultaat is het eerste gesimuleerde veld van GHG's en een daarmee gecorreleerd veld van GLG's. Met telkens een nieuwe set gelote parameterwaarden wordt de procedure net zolang herhaald, tot er een groot aantal gesimuleerde velden van GHG's en GLG's zijn (hier 300; zie paragraaf 4.5).

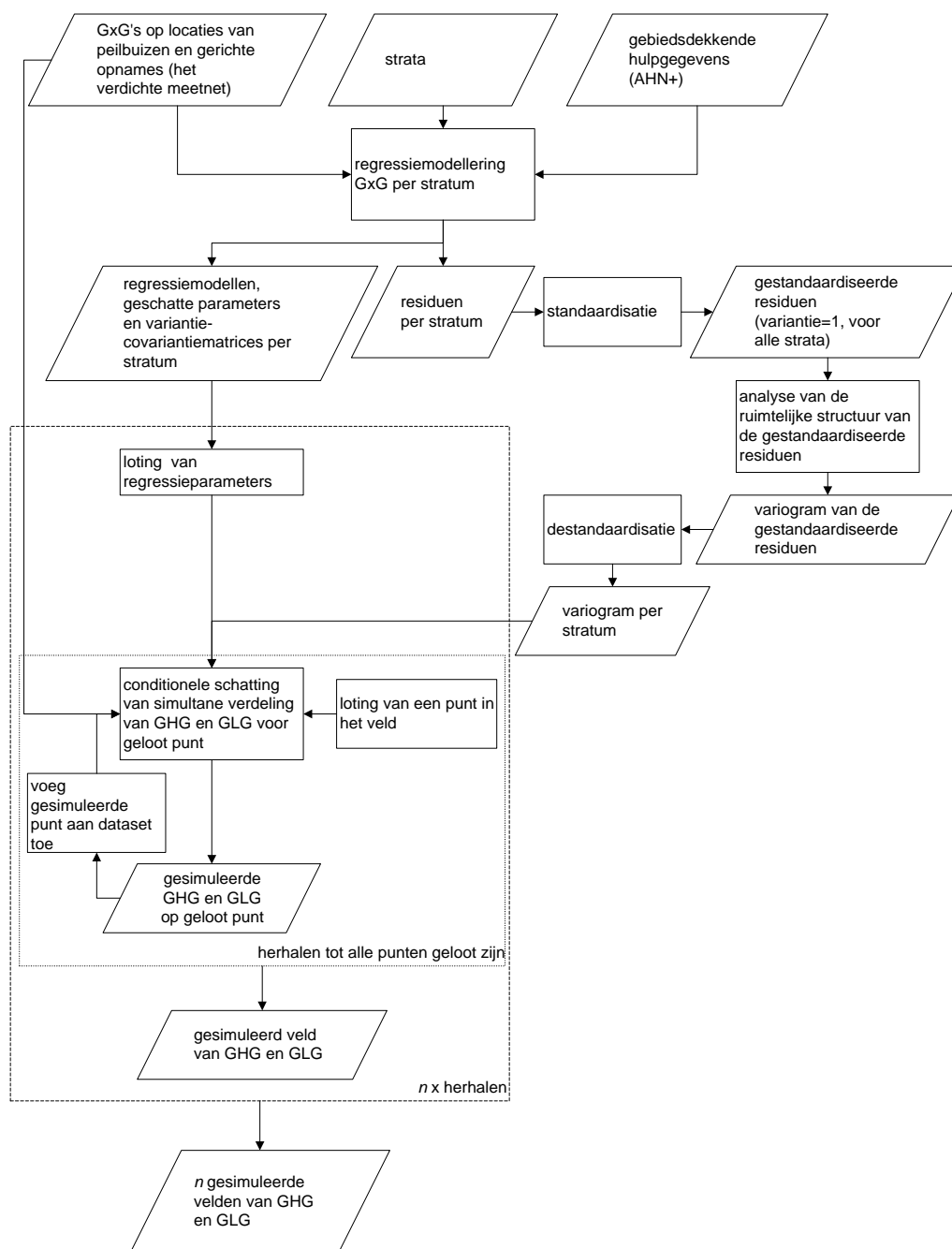
Zowel bij de simulatie als bij de interpolatie wordt rekening gehouden met de onnauwkeurigheid van de GxG's op peilbuislocaties en locaties van de gerichte opnames; dit zijn immers geen foutloze metingen, maar schattingen en voorspellingen, zie paragraaf 3.3.2 en 3.4. Paragraaf 4.4 beschrijft hoe de onnauwkeurigheid van de GxG-schattingen wordt verdisconteerd in de simulaties.

4.2 Regressieanalyse

Voor elk stratum wordt een model geselecteerd dat de samenhang beschrijft tussen de GxG en de gebiedsdekkende hulpinformatie uit de vijf groepen die beschreven zijn in paragraaf 3.5.2. Omdat verklarende variabelen die tot dezelfde groep behoren sterk gecorreleerd zijn wordt, om multicollineariteit te voorkomen, uit elke groep slechts één hulpvariabele in het regressiemodel opgenomen. Op deze wijze kan een set kandidaatmodellen worden samengesteld, met elk een regressieconstante en met maximaal vijf predictorvariabelen. De omvang van de set kandidaatmodellen M is dan gelijk aan

$$M = \prod_{i=1}^N (G_i + 1) + 1,$$

waarin N het aantal groepen is en G_i het aantal verklarende variabelen in de i^{de} groep. In dit geval is $N = 5$, $G_1 = 5$, $G_2 = 2$, $G_3 = 2$, $G_4 = 1$ en $G_5 = 4$, dus $M = 540$. Voor elk stratum wordt het beste model geselecteerd op basis van Mallows Cp-criterium, overeenkomstig de procedure in paragraaf 3.4.2, zie vergelijking



Figuur 4.1. Werkwijze bij de gebiedsdekkende simulatie van GxG's

(3.1). Als waarde voor s^2 in vergelijking (3.1) wordt de restvariantie genomen van het model met de beste fit.

Om de onnauwkeurigheid van de GxG-schattingen voor peilbuis- en gerichte-opnamelocaties te kunnen verdisconteren wordt een *gewogen* regressie uitgevoerd, waarbij de nauwkeurigheid als gewicht dient: hoe nauwkeuriger de GxG-schatting, hoe meer gewicht in de regressieanalyse. In formulevorm is het gewicht gelijk

aan $f/(s_i^2)$, waarin s_i^2 de variantie is van de fout in de berekende GxG op locatie i , $i = 1 \dots n$ van een peilbuis of een gerichte opname, en f een factor is die ervoor zorgt dat de gewichten sommeren tot het aantal waarnemingen n dat wordt gebruikt in de regressieanalyse.

De modellen die voor elk stratum als beste worden geselecteerd, worden gebruikt bij de gebiedsdekkende simulatie en interpolatie van de GxG, zie paragraaf 4.5.

4.3 Ruimtelijke structuur van de residuen

Het residu e is gedefinieerd als het verschil tussen de GxG-waarde zoals berekend voor een peilbuislocatie of een locatie van de gerichte opname (GxG), en de waarde die voor die locatie is geschat met het geselecteerde regressiemodel (\widehat{GxG}):

$$e_{k,i} = GxG_{k,i} - \widehat{GxG}_{k,i},$$

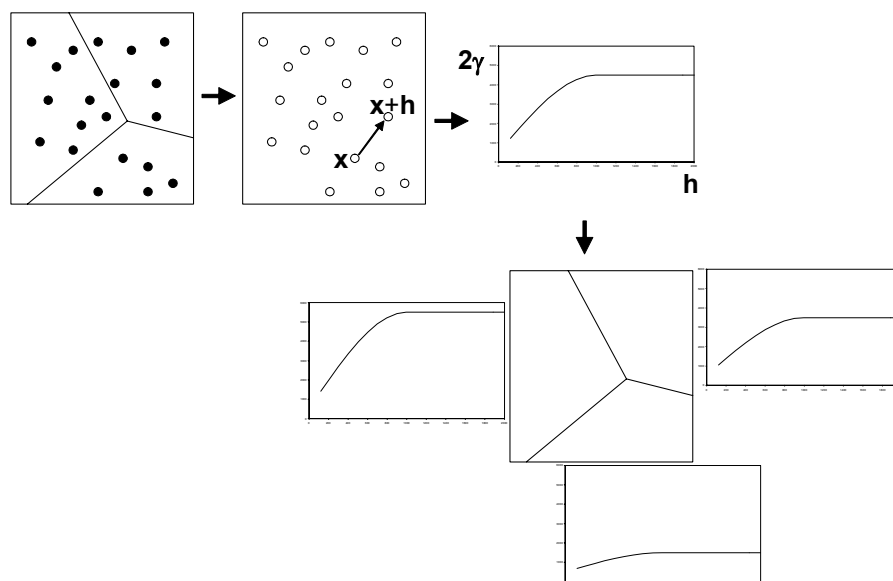
waarin i de i^{de} locatie aangeeft in stratum k . De residuen zijn gemiddeld nul en hebben een variantie s_k^2 . Waarschijnlijk vertonen de residuen een ruimtelijke correlatie of structuur. Het regressiemodel behoeft immers niet alle ruimtelijke structuur in de GxG verklaard te hebben uit de hulpinformatie. Voor de simulatie van de GxG-velden en uiteindelijk de berekening van overschrijdingskansen op basis waarvan percelen worden geclassificeerd als uitspoelingsgevoelig, is het van belang om rekening te houden met de ruimtelijke structuur van de residuen.

De ruimtelijke structuur wordt gemodelleerd met een variogram (Davis, 2002, blz. 254-264). En dergelijk variogram geeft de afhankelijkheid weer tussen een waarneming op locatie \mathbf{x} en een waarneming op locatie $\mathbf{x} + \mathbf{h}$, waarbij \mathbf{h} een vector is die zowel afstand als richting aangeeft. Hier wordt aangenomen dat de richting geen invloed heeft op de afhankelijkheid, en dat dus alleen de afstand een rol speelt. Het variogram geeft de varianties $\text{Var}[e(\mathbf{x}) - e(\mathbf{x} - \mathbf{h})]$ als functie van de afstandsvector \mathbf{h} . Voordat het variogram wordt gemodelleerd, worden eerst de residuen gestandaardiseerd, zodat de residuen in alle strata variantie één hebben. Dankzij de standaardisatie kunnen de residuen uit *alle* strata worden gebruikt bij de modellering van één, gestandaardiseerd variogram. Zonder standaardisatie zouden er variogrammen voor de strata afzonderlijk moeten worden gemodelleerd, wat gezien het geringe aantal waarnemingen in de strata (gemiddeld 30) onnauwkeurige modellen zou kunnen opleveren. Het gestandaardiseerde variogram wordt vervolgens gedestandaardiseerd met de varianties s_k^2 .

Om rekening te houden met onderlinge correlaties worden, behalve de variogrammen van de residuen van GHG, GVG en GHG, ook de z.g. kruisvariogrammen berekend voor de residuen van paren GHG en GVG, GHG en GLG, en GVG en GLG. Deze kruisvariogrammen worden als volgt van de gedestandaardiseerde variogrammen afgeleid:

$$\gamma_{1,2}(h) = 0.7 \sqrt{\gamma_1(h) \times \gamma_2(h)}$$

waarin $\gamma_{1,2}(h)$: het niveau van het kruisvariogram voor locaties met een onderlinge afstand h en variabelen 1 en 2, bijvoorbeeld de residuen van GHG en GLG; $\gamma_1(h)$, $\gamma_2(h)$: het niveau van het variogram van variabele 1, resp. 2, voor locaties met



Figuur 4.2. Analyse van de ruimtelijke structuur van de residuen. ●: residuen; ○: gestandaardiseerde residuen; $2\gamma(\mathbf{h})$ = variantie van de verschilwaarde $e(\mathbf{x}) - e(\mathbf{x} + \mathbf{h})$; destandaardisatie per stratum.

een onderlinge afstand h ; 0.7: geschat gemiddelde van de kruiscorrelaties tussen de residuen.

De gestandaardiseerde variogrammen en kruisvariogrammen worden gebruikt bij de gebiedsdekkende simulatie van de GxG (paragraaf 4.5). De modellering van de ruimtelijke structuur van de residuen is samengevat in Figuur 4.2.

4.4 Verrekening van de onnauwkeurigheid van GxG-schattingen

De GxG's op de peilbuislocaties en de locaties van de gerichte opnames zijn geen foutloze metingen, maar geschatte of voorspelde waarden, zie paragraaf 3.3.2 en 3.4. De onnauwkeurigheid van deze schattingen is gekwantificeerd met varianties van de schattings- of voorspelfout. Bij twee onderdelen van de simulatieprocedure wordt de onnauwkeurigheid van de GxG's in rekening gebracht: 1) bij het residuele variogram, en 2) bij het kriging stelsel dat gebruikt wordt bij de simulatie van GxG-velden.

Het residuele variogram moet worden gereduceerd met de onnauwkeurigheid van de GxG's, omdat het doel is de *werkelijke* GxG te voorspellen en niet de GxG zoals die voor een meetlocatie zou kunnen worden berekend. Daarom wordt per stratum de gemiddelde variantie van de schattings- en voorspellingsfouten geschat. Deze variantie wordt afgetrokken van het variogram, tot een niveau van uiterlijk nul om negatieve varianties te vermijden.

Ook het kriging stelsel dat is beschreven in Bijlage B wordt aangepast: de variantie van de schattings- en voorspellingsfouten van de GxG's op peilbuislocaties en locaties van de gerichte opnames wordt ingevuld op de corresponderende diagonaalelementen

van de matrix Z_+ . Dit is analoog aan kriging met onzekere gegevens (Delhomme, 1978).

4.5 Simulatie van GHG- en GLG-velden

De gebiedsdekkende simulatie van GHG en GLG vindt simultaan plaats, en wordt voor de strata afzonderlijk en onafhankelijk van elkaar uitgevoerd. De procedure is stapsgewijs als volgt:

1. uit de gezamenlijke kansverdeling van de parameters van de geselecteerde regressiemodellen (één voor de GHG en één voor de GLG) wordt een set parameterwaarden geloot;
2. er wordt een locatie geloot uit het 25×25 metergrid van de hulpinformatie (zonder waarde voor GxG), zeg \mathbf{x}_i ;
3. voor \mathbf{x}_i wordt de conditionele gezamenlijke kansverdeling van de GHG en GLG geschat, met behulp van de gelote regressieparameters, de GHG's en GLG's op de omliggende locaties van peilbuizen en gerichte opnames, de predictorvariabelen van de hulpinformatie, en het gestandaardiseerde variogram van de residuen;
4. uit de conditionele kansverdeling van de GHG en GLG op locatie \mathbf{x}_i worden waarden voor de GHG en GLG geloot. Deze worden beschouwd als 'pseudo-metingen'. Ze worden toegevoegd aan de set GHG's en GLG's op peilbuis- en gerichte-opnamelocaties, teneinde een samenhangend vlak te simuleren;
5. stap 2 tot en met 4 worden herhaald totdat alle punten van het 25×25 metergrid een keer zijn geloot. Dit levert het eerste gesimuleerde GHG- en GLG-veld;
6. stap 1 tot en met 5 worden herhaald tot het gewenste aantal gesimuleerde velden is bereikt.

Deze simulatieprocedure, die *sequential Gaussian cosimulation* wordt genoemd, wordt uitgevoerd met het programma `gstat` (Pebesma en Wesseling, 1998). De methode is gedetailleerd beschreven in Bijlage B. Het benodigde aantal gesimuleerde velden (stap 6) is vastgesteld op 300, zoals bediscussieerd in Bijlage D.

4.6 Gebiedsdekkende interpolatie van de GxG

Voor de gebiedsdekkende interpolatie ter vervaardiging van GxG kaarten worden dezelfde gegevens gebruikt als voor de simulatie beschreven in paragraaf 4.5: de berekende GxG's op de meetlocaties en de hulpinformatie. Tevens wordt van dezelfde regressiemodellen en dezelfde residuele variogrammen uitgegaan, en verder worden de berekeningen net als bij de simulatie simultaan uitgevoerd voor GHG, GVG en GLG, en afzonderlijk voor elk stratum.

Een verschil is dat, in tegenstelling tot bij simulatie, nieuw berekende punten niet aan de dataset worden toegevoegd, dus dat voorspellingen berekend worden met behulp van alleen de GxG's van stambuizen en gerichte opname. Het belangrijkste

verschil is dat bij simulatie voor een gegeven locatie een aantal GxG waarden worden geloot uit de voor die locatie berekende kansverdeling, terwijl bij interpolatie het gemiddelde van die kansverdeling wordt genomen als beste voorspelling. Dit waarborgt dat de simulatie- en interpolatie-resultaten onderling consistent zijn, in die zin dat het gemiddelde van een groot aantal gesimuleerde GxG velden gelijk is aan het geïnterpoleerde GxG veld. De benaming van deze interpolatie methode is ‘Universal Co-Kriging’.

Een en ander houdt in dat, vergeleken met de methode van Gd-kartering zoals beschreven door Finke et al. (2002), de volgende verbeteringen, cq. veranderingen in de interpolatie-methodiek zijn aangebracht:

- er wordt rekening gehouden met de onzekerheid in de berekende GxG’s op meetpunten;
- er wordt rekening gehouden met de correlatie tussen GHG, GVG en GLG;
- er worden, behalve voor de bepaling van het residuele variogram, bij de interpolatie binnen een stratum geen gegevens van buiten het stratum gebruikt.

Hoofdstuk 5

Perceelsclassificatie naar uitspoelingsgevoeligheid

Zodra een groot aantal gebiedsdekkende GxG-velden zijn gesimuleerd, kunnen deze worden nabewerkt om percelen te classificeren naar uitspoelingsgevoeligheid. Het uitgangspunt is daarbij dat een perceel als uitspoelingsgevoelig wordt geclassificeerd indien er voldoende zekerheid is dat tenminste een bepaald percentage van het oppervlakte van het perceel uitspoelingsgevoelig is. Daartoe worden de volgende criteria gehanteerd:

- Puntcriteria: een punt in het perceel is gedefinieerd als uitspoelingsgevoelig indien op dat punt:
 - de bodem bestaat uit zand, en
 - de GHG dieper is dan GHG_{crit} cm beneden maaiveld (GHG_{crit} is bijvoorbeeld 40, 50, 60, 70 of 80), en
 - de GLG dieper is dan 120 cm beneden maaiveld.
- Oppervlaktecriterium: het perceel als geheel is gedefinieerd als uitspoelingsgevoelig, indien in meer dan O_{crit} procent van zijn oppervlak is voldaan aan de punt-criteria voor uitspoelingsgevoeligheid. (O_{crit} is bijvoorbeeld 50, 67 of 80.)
- Kanscriterium: het perceel wordt als uitspoelingsgevoelig geclassificeerd, indien de uit de simulaties te berekenen kans dat aan het oppervlaktecriterium is voldaan, groter dan P_{crit} is. (P_{crit} is bijvoorbeeld 0.90 of 0.95.) Dit criterium garandeert dat de kans dat een perceel ten onrechte wordt geclassificeerd als uitspoelingsgevoelig in elk geval niet groter is dan $1 - P_{crit}$. Naarmate dus $1 - P_{crit}$ hoger wordt gekozen, wordt een grotere zekerheid verkregen dat een als uitspoelingsgevoelig geclassificeerd perceel inderdaad uitspoelingsgevoelig is. Bij de keuze $1 - P_{crit} = 0.5$ leveren al meer dan 50 % van de simulaties de uitslag op dat het perceel uitspoelingsgevoelig is.

De perceelsclassificatie vindt plaats door per perceel voor elk tweetal gesimuleerde GHG- en GLG-velden het percentage van het oppervlak te berekenen waar voldaan is aan de punt-criteria voor uitspoelingsgevoeligheid. Dit gebeurt door per gridpunt

van het 25×25 m grid te bepalen of het voldoet aan de puntcriteria, en het aantal gridpunten waar dit het geval is te delen door het totaal aantal gridpunten in het perceel. Vervolgens wordt bepaald of dit percentage O_{crit} overschrijdt. Tenslotte wordt de fractie gesimuleerde GxG's berekend waarin O_{crit} wordt overschreden. Als deze fractie groter is dan P_{crit} , dan wordt het perceel als uitspoelingsgevoelig geclassificeerd, anders als niet-uitspoelingsgevoelig.

Voor de afgrenzing van zandbodems t.o.v. andere bodems wordt gebruik gemaakt van de Systematische Bodemkaart, schaal 1 : 50 000. Voor de ligging van de perceelsgrenzen wordt het PIPO bestand versie 2003 van Laser gebruikt.

Om de betrouwbaarheid van de classificaties na te gaan is voor vijf representatief geachte gebieden (Dommel en Dongestroom, Zuid Limburg, Friese Zandgronden, Regge en Dinkel, Reest en Wieden) het verwachte percentage misclassificaties berekend, volgens de rekenwijze beschreven in Bijlage D. Daarbij is uitgegaan van de criteria $GHG_{\text{crit}}=60$, $O_{\text{crit}}=67$ en $P_{\text{crit}}=0.95$. Het percentage percelen dat ten onrechte als uitspoelingsgevoelig is geclassificeerd, varieert in deze gebieden van 0.1 % tot 1.5 %. Het percentage percelen dat ten onrechte *niet* als uitspoelingsgevoelig is geclassificeerd, varieert van 18.2 % tot 31.6 %.

Hoofdstuk 6

Kwaliteit van de methodiek

In kwaliteitsborging is voorzien door bij de methodiekontwikkeling de volgende maatregelen te nemen, cq. uitgangspunten te kiezen.

- De methodiek is grotendeels opgebouwd als een combinatie van reeds bestaande methoden waarvan de bruikbaarheid is bewezen. Dit betreft met name de onderdelen:
 - de methode van ‘gerichte’ opname van de GxG: reeds in vele projecten toegepast en in een aantal gevallen gevalideerd (Finke et al., 1999; Hoogland et al., 2003; Pleijter et al., 2003);
 - tijdreeksanalyse: een reeds bestaande methode van tijdreeksanalyse welke reeds in vele Gd-karteringen is toegepast;
 - lineaire multiple regressieanalyse: een standaard statistische techniek;
 - interpolatie: de hier gevolgde geostatistische interpolatie methode ‘Universal Kriging’ bestaat reeds lang in de literatuur, en is in velerlei onderzoeken met succes toegepast;
 - simulatie: de hier gevolgde geostatistische simulatie methode ‘Sequential Full Gaussian simulation’ (Abrahamsen en Benth, 2001) is relatief nieuw, maar Pebesma en Heuvelink hebben de methode reeds met succes in de bodemkunde toegepast (Pebesma en Heuvelink, 2001).
- Voor de berekeningen worden voornamelijk reeds lang en algemeen gebruikte programma’s toegepast: Genstat voor tijdreeksanalyse en regressieanalyse, **gstat** voor geostatistische interpolatie en simulatie, Arc-Info voor GIS bewerkingen.
- Intensieve controle van tussenproducten, in de vorm van kritische beoordeling van tabellen en kaartbeelden van simulatieresultaten door materiedeskundigen van de begeleidingscommissie en Alterra. De begeleidingscommissie heeft 11 maal vergaderd, en heeft daarbij vooral gecontroleerd of de diverse geproduceerde testresultaten compatibel waren met bestaande kennis en informatie. Onafhankelijk daarvan hebben deskundigen van Alterra soortgelijke controles uitgevoerd.
- Aparte deel-onderzoeken zijn gewijd aan enkele potentieel zwakke plekken in de methodiek:

Stratumgrenzen door percelen: in Bijlage C is een aparte beschouwing gewijd aan het feit dat sommige percelen worden doorsneden door stratumgrenzen. De conclusie is dat deze manier van stratificeren acceptabel is.

Benodigd aantal simulaties: het aantal simulaties dat nodig is om tot voldoende betrouwbare perceelsclassificaties te komen is onderzocht, en vastgesteld op 300 (zie Bijlage D).

Mogelijk effect van perceelsgrootte: aan de hand van een ‘case study’ is nagegaan of de grootte van een perceel van invloed is op de kans dat het wordt geclassificeerd als uitspoelingsgevoelig (zie Bijlage E). Een dergelijk effect bleek niet op te treden.

Ondanks bovengenoemde kwaliteitsborging zijn in de methodiek een aantal onvolkomenheden aan te wijzen. De belangrijkste zijn in onze visie:

- De onzekerheid in de GxG wordt wel verdisconteerd, maar die in de afgrenzing van het zandgebied niet.
- Weliswaar wordt rekening gehouden met meetfouten in de GxG bepalingen op de meetlocaties, maar daarbij wordt aangenomen dat deze ruimtelijk onafhankelijk zijn. Deze aanname is vermoedelijk niet realistisch, maar wordt noodgedwongen geïntroduceerd omdat geen bruikbare informatie over de ruimtelijke afhankelijkheid van de meetfouten voorhanden is.
- De methodiek maakt geen gebruik van informatie over lokale afwijkingen in de hydrologie, bijvoorbeeld bij grondwateronttrekkingen. Deze zouden een versturende invloed kunnen hebben.
- De berekening van de residuele variantie van de maaiveldregressie geeft waarschijnlijk een onderschatting van de werkelijke variantie omdat geen rekening is gehouden met het feit dat de keuze van de predictoren via ‘subset selection’ (stepwise regression) tot stand is gekomen (Chatfield, 1995).

Bibliografie

- Abrahamsen, P. en Benth, F. (2001). Kriging with inequality constraints. *Mathematical Geology*, 33:719–744.
- Bierkens, M., Knotters, M., en van Geer, F. (1999). Tijdreeksanalyse nu ook toepasbaar bij onregelmatige meetfrequenties. *Stromingen*, 5(2):43–54.
- Braat, L., van Amstel, A., Gerritsen, A., van Gool, C., Gremmen, N., Groen, C., Rolf, H., Runhaar, J., en Wiertz, J. (1989). *Verdroging van natuur en landschap in Nederland. Beschrijving en analyse*. Ministerie van Verkeer en Waterstaat, 's-Gravenhage.
- Chatfield, C. (1995). Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society, Series A*, 158:419–466.
- Cohen, A. (1991). *Truncated and censored samples: theory and applications*. Dekker Inc., New York.
- Davis, J. (2002). *Statistics and data analysis in geology*. Wiley, New York, third edition.
- De Bruin, H. (1987). Van Penman naar Makkink. In Hooghart, J., redactie, *Neerslag en Verdamping*, CHO-TNO Mededeling 39. CHO-TNO, Den Haag.
- Delhomme, J. (1978). Kriging in the hydrosiences. *Advances in water resources*, 1:251–266.
- Finke, P. (2000). Updating the (1:50,000) Dutch groundwater table class map by statistical methods: an analysis of quality versus cost. *Geoderma*, 97:329–350.
- Finke, P., Bierkens, M., Brus, D., van der Gaast, J., Hoogland, T., Knotters, M., en de Vries, F. (2002). Klimaatsrepresentatieve grondwaterdynamiek in Waterschap Mark en Weerij. Rapport 387, Alterra.
- Finke, P., Hoogland, T., Bierkens, M., Brus, D., en de Vries, F. (1999). Pilot naar grondwaterkaarten in het Weerijgebied. , Staring Centrum.
- Hoogland, T., Finke, P., en de Vries, F. (2003). Actualisatie grondwatertrappenkaart Waterschap Rijn en IJssel. Rapport 126, Alterra.
- Knotters, M. (2001). *Regionalised time series models for water table depths*. PhD thesis, Wageningen Universiteit.
- Knotters, M. en Bierkens, M. (1999). Hoe lang moet je de grondwaterstand meten om iets over de dynamiek te weten? *Stromingen*, 5(4):5–12.

- Knotters, M. en van Walsum, P. (1994). Uitschakeling van weersinvloeden bij de karakterisering van het grondwaterstandsverloop. Rapport 350, Staring Centrum.
- Miller, A. (1990). *Subset selection in regression*. Chapman and Hall, London.
- Oude Voshaar, J. en Stolp, J. (1997). Schatting van GHG en GLG van tijdelijke peilbuizen met korte meetreeksen. Technisch Document 30, Staring Centrum.
- Pebesma, E. en Heuvelink, G. (2001). Sequential simulation of Gaussian random fields with unknown mean function: an application to heavy metal pollution data. In *Fourth Conference of the Working on Pedometrics of the International Union of Soil Sciences, September 19-21*, pages Abstract book, p. 84, Gent, Belgium.
- Pebesma, E. en Wesseling, C. (1998). Gstat, a program for geostatistical modelling, prediction and simulation. *Computers and Geosciences*, 24(1):17–31; <http://www.gstat.org/>.
- Pleijter, M., Brouwer, F., en Brus, D. (2003). Kaarten met grondwaterstandsverloop nader bekeken. Kan de kwaliteit van grondwaterstandkaarten die gemaakt zijn met ruimtelijke modellen verbeterd worden door aanvullend veldwerk? Rapport 736, Alterra.
- Te Riele, W. en Brus, D. (1992). Het gebruik van fysisch-geografische voorinformatie bij de ruimtelijke voorspelling van grondwaterstanden en grondwaterkarakteristieken (GHG en GLG). Rapport 209, Staring Centrum.
- Te Riele, W., Querner, E., Knotters, M., en Pomper, A. (1995). Geostatistische interpolatie van grondwaterstandsdiepten met behulp van fysisch-geografische informatie en de resultaten van een regionaal stromingsmodel. Rapport 414, Staring Centrum.
- Thompson, M. (1978). Selection of variables in multiple regression: Part I. A review and evaluation. *International Statistical Review*, 46:1–19.
- Van Bracht, M. (1988). OLGA: On Line Grondwater Archief. Rapport PN88-11, DGV-TNO.
- Van der Sluijs, P. (1982). De grondwatertrap als karakteristiek van het grondwaterstandsverloop. *H₂O*, 15:42–46.
- Van der Sluijs, P. (1990). Hoofdstuk 11: Grondwatertrappen. In Locher, W. en de Bakker, H., redactie, *Bodemkunde van Nederland, deel 1: Algemene Bodemkunde*. Malmberg, Den Bosch.
- Wesseling, J. (1991). CAPSEV; steady state moisture flow theory; program description; user manual. Report 37, Staring Centrum.
- Wösten, J., de Vries, F., Denneboom, J., en van Holst, A. (1988). Generalisatie en bodemfysische vertaling van de bodemkaart van Nederland, 1 : 250 000, ten behoeve van de PAWN-studie. Rapport 2055, Stichting voor Bodemkartering.

Bijlage A

Tijdreeksmodellering

In het tijdreeksmodel dat de dynamische relatie tussen neerslagoverschot p en de grondwaterstand h beschrijft, bestaat de grondwaterstand op tijdstip t (h_t) uit een som van twee componenten: een deterministische of transfercomponent (h_t^*) en een ruiscomponent n_t : $h_t = h_t^* + n_t$. De relatie wordt transfer-ruismodel genoemd. De deterministische component h_t^* is het deel van de grondwaterstand dat kan worden verklaard uit een lineaire samenhang met het neerslagoverschot, en de ruiscomponent n_t bevat de resterende invloeden.

De algemene vorm van het transfer-ruismodel is:

$$h_t^* = \sum_{i=1}^r \delta_i h_{t-i}^* + \omega_0 p_{t-b} - \sum_{j=1}^s \omega_j p_{t-j-b}, \quad (\text{A.1})$$

$$(n_t - \mu) = \sum_{k=1}^p \phi_k (n_{t-k} - \mu) + \epsilon_t - \sum_{l=1}^q \theta_l \epsilon_{t-l}, \quad (\text{A.2})$$

$$h_t = h_t^* + n_t, \quad (\text{A.3})$$

waarin:

h_t de grondwaterstand op tijdstip t is [L];

p_t het gemiddelde potentiële neerslagoverschot is tussen $t - 1$ en t [LT^{-1}];

h_t^* de deterministische component van de grondwaterdiepte op tijdstip t is, welke wordt verklaard door het neerslagoverschot p_t [L];

n_t de ruiscomponent op tijdstip t is, welke de resultante is van alle invloeden behalve p_t [L];

μ een constante is, het gemiddelde van n_t [L];

ϵ_t de foutenterm is met discrete tijdindex t , waarvan wordt aangenomen dat het een witte-ruisproces is, met gemiddelde nul en eindige en constante variantie σ_ϵ^2 [L];

$\delta_1, \dots, \delta_r$ de autoregressieve parameters van het transfermodel zijn tot aan orde r [–];

$\omega_0, \dots, \omega_s$ de *moving average*-parameters van het transfermodel zijn tot aan orde s [T];

ϕ_1, \dots, ϕ_p de autoregressieve parameters van het ruismodel zijn tot aan orde p [-]; $\theta_1, \dots, \theta_q$ de *moving average*-parameters van het ruismodel zijn tot aan orde q [-]; b de vertraging vertegenwoordigt tussen *input* p en *output* h (de tijd die er overheen gaat voordat een verandering in p resulteert in een verandering in h ; in deze studie geldt $b = 0$, omdat p_t de gemiddelde neerslagoverschotintensiteit is tussen $t - 1$ en t , die altijd met de grondwaterstand op t blijkt samen te hangen) [T].

Het dagelijkse potentiële neerslagoverschot p wordt berekend uit het verschil tussen dagneerslagsom (p_N) en de etmaalverdamping voor een referentiegewas volgens Makkink (e_M), zoals deze door het KNMI is gegeven (De Bruin, 1987):

$$p_t = p_{N,t} - e_{M,t}.$$

In dit onderzoek wordt een eenvoudige vorm van het transfer-ruismodel gebruikt, met de indices $r = 1$, $s = 0$, $p = 1$, $q = 0$ en $b = 0$. Knotters (2001, hoofdstuk 4) toonde aan dat dit model eenvoudig fysisch is te verklaren en in veel situaties een goede beschrijving van de grondwaterdynamiek geeft. De deterministische component h_t^* wordt geschat met:

1. de vorige grondwaterstandmeting uit de tijdreeks en
 2. het neerslagoverschot tussen de huidige en de vorige meting.
- De vergelijkingen (A.1) tot en met (A.3) reduceren in dit geval tot:

$$h_t^* = \delta_1 h_{t-1}^* + \omega_0 p_t, \quad (\text{A.4})$$

$$(n_t - \mu) = \phi_1 (n_{t-1} - \mu) + \epsilon_t, \quad (\text{A.5})$$

$$h_t = h_t^* + n_t, \quad (\text{A.6})$$

Het transfer-ruis model in vergelijking (A.4) tot en met (A.6) is ingebed in een Kalman-filter en gebaseerd op de dagfrequentie van de neerslagoverschotreeks, waardoor kalibratie op de minder frequent en onregelmatig waargenomen grondwaterstandsreeksen mogelijk is (Bierkens et al., 1999).

De coëfficiënten van het transfer-ruismodel worden gekalibreerd met het programma KALTFN. De neerslaggegevens zijn afkomstig van het KNMI-neerslagstation dat zich het dichtst bij de peilbuis bevindt. De gegevens betreffende de referentiegewasverdamping zijn eveneens afkomstig van het dichtsbijzijnde weerstation waar deze worden verzameld.

De berekening van de GxG uit de gesimuleerde reeksen van dertig jaar lang verloopt in twee stappen:

1. per hydrologisch jaar wordt het gemiddelde van de drie hoogste en laagste grondwaterstanden genomen (resp. HG3 en LG3);
2. de dertig HG3's en LG3's worden gemiddeld tot een GHG en een GLG.

Het gebruik van alleen de deterministische component bij de simulaties zou leiden tot een onderschatting van de temporele variatie. Met stochastische simulatie kan

dit worden voorkomen, doordat dan de ruiscomponent in rekening wordt gebracht. Omdat zowel de GHG als de GLG extreme grondwaterstanden beschrijven, is het van belang dat de temporele variatie juist wordt gesimuleerd. Daarom wordt gebruik gemaakt van stochastische simulatie, (zie Knotters, 2001, hoofdstuk 2); deterministische simulatie zou tot te diepe schattingen van de GHG en te ondiepe schattingen van de GLG leiden. Omdat de grondwaterstandsreeksen van dertig jaar lang worden gesimuleerd door middel van stochastische simulatie, kan ook de onnauwkeurigheid als gevolg van het veronderstelde verband tussen neerslagoverschot en grondwaterstand worden gekwantificeerd.

Bijlage B

Sequentiële Gaussische co-simulatie

Sequential simulation of Gaussian random
fields with unknown mean coefficients:
theory and implementation

Edzer J. Pebesma
Dept. of Physical Geography
Utrecht University
January 2003

abstract

This document describes how Gaussian random fields can be simulated in the case where the mean coefficients are not assumed to be known. In traditional stochastic simulation of random fields (e.g. Deutsch and Journel, 1998; Goovaerts, 1997) only the residual part of the spatial variation is subject to simulation, leading to realisations that jointly follow the Gaussian distribution with the simple kriging mean and variance. Here, we address multiple correlated variables and include the uncertainty of estimated trend coefficients to the simulation process, leading to realisations that jointly follow the Gaussian distribution with the universal cokriging mean and variance. Details of the implementation in `gstat` are given, as well as some preliminary test results.

B.1 Introduction

Traditionally, geostatistical simulation of Gaussian random fields (e.g. Deutsch and Journel, 1998; Goovaerts, 1997) only address the residual part of the spatial variation. When the estimation error of trend coefficient plays more than a negligible role in the prediction errors, ignoring this error may lead to simulations that underestimate the prediction error. In this paper, we will present a simulation algorithm for multiple correlated Gaussian random fields that does address the error in trend coefficients, we describe an implementation, and we show some preliminary test results.

The algorithm used here extends the sequential Gaussian simulation algorithm (sGs, as found in Deutsch and Journel, 1998 and Goovaerts 1997). The extension was described in Abrahamsen and Benth (2001) and an environmental application was presented by Pebesma and Heuvelink (2001). Here, we will call the extended algorithm sFGs, for sequential Full Gaussian simulation, as contrasted to sequential simple Gaussian simulation that only addresses residual variability.

This research was supported by Alterra Green World Research, under supervision of dr. J.J. de Gruijter.

B.2 Theory

B.2.1 Sequential Simple Gaussian Simulation

Suppose we have a spatially correlated variable $Z(s)$, observed at n spatial locations $s = (s_1, \dots, s_n)$. We assume that the spatial variability of $Z(s)$ can be decomposed in a trend and a residual:

$$Z(s) = \mu(s) + e(s),$$

where the trend is a linear function of p known regressors $X_j(s)$:

$$\mu(s_i) = \beta_0 + \sum_{j=1}^p X_j(s_i)\beta_j = X(s_i)\beta, \quad i = 1, \dots, n,$$

with regressor vector $X(s_i) = (1, X_1(s_i), \dots, X_p(s_i))$ and coefficient vector $\beta = (\beta_0, \dots, \beta_p)^T$, and where T denotes matrix or vector transpose. Furthermore, we assume that β is known. Standard sequential Gaussian simulation (Johnson, 1987; Deutsch and Journel 1998) proceeds as follows:

1. define a random path through the q simulation locations s_{n+1}, \dots, s_{n+q} , and let $i = 1$
2. at the unvisited location s_{n+i} , calculate the conditional distribution of $Z(s_{n+i})$ from the Gaussian distribution with mean value the simple kriging mean

$$\hat{z}_{\text{sk}}(s_{n+i}) = X(s_{n+i})\beta + c^T C^{-1}(Z_+ - X_+\beta)$$

and variance the simple kriging variance:

$$\sigma_{\text{sk}}^2(s_{n+i}) = \text{Var}(Z(s_{n+i})) - c^T C^{-1}c$$

where $c = (\text{Cov}(Z(s_1), Z(s_{n+i})), \dots, \text{Cov}(Z(s_{n+i-1}), Z(s_{n+i})))^T$, and where C is the covariance matrix of the vector $Z_+ = (Z(s_1), \dots, Z(s_{n+i-1}))$ containing data and previously simulated values, and where X_+ is the $(n+i-1) \times (p+1)$ matrix having $X(s_i)$ as its i -th row.

3. draw a value $\tilde{z}(s_{n+i})$ randomly from $\mathcal{N}(\hat{z}_{\text{sk}}(s_{n+i}), \sigma_{\text{sk}}^2(s_{n+i}))$, the normal distribution with mean $\hat{z}_{\text{sk}}(s_{n+i})$ and variance $\sigma_{\text{sk}}^2(s_{n+i})$
4. add this value to the data set; call it $Z(s_{n+i})$
5. if $i < q$, let $i = i + 1$ and go to step 2.

Typically, the number of simulation locations is much larger than the number of observed data ($q \gg n$). At each step, the data set Z_+ is augmented with one simulated value. When many data are available, or when simulating large fields (say, $n + q > 200$), in the end the computation becomes extremely slow because $C^{-1}c$ has to be solved each step. Solving $C^{-1}c$ has complexity $O(N^2)$, which means that doubling $n + q$ roughly quadruples the computing time needed. This is bad if we want to simulate large fields.

To avoid this quadratic behaviour, usually a subset of say r measurements that are closest (in euclidian distance, or in correlation) to $Z(s_{n+i})$ are selected, and Z_+ is replaced by $Z_{[r]}(s_{n+i})$. This involves an approximation of the true conditional expectation. For that reason (i) the random path is chosen through the simulation locations, and (ii) r should be chosen as large as computing time permits: the larger r , the better the approximation. Restricting the neighbourhood size to a fixed number of nearest points makes the behaviour of the sequential simulation algorithm much closer to $O(N)$, i.e. close to linear in $n + q$. The “expensive” results at each simulation location are $C^{-1}c$ and the spatial neighbourhood selection (especially when $n + q$ becomes large).

B.2.2 Sequential Simple Gaussian Cosimulation

Suppose we have m correlated variables $Z_k(s_k)$, $k = 1, \dots, m$, observed at m sets of n_k spatial locations $s_k = (s_{1,k}, \dots, s_{n_k,k})$, and that correlation includes both spatial

correlation and (spatial) correlation between distinct variables. Only for ease of notation we will further assume $s_k = s$, (different variables are observed at identical locations) and $p_k = p$ (all variables have the same number of regressors). Let $Z(s) = (Z_1(s_1), \dots, Z_1(s_n), \dots, Z_m(s_1), \dots, Z_m(s_n))^T$

We assume that the spatial variability of each of the $Z_k(s)$ can be decomposed in a trend and a residual:

$$Z_k(s) = \mu_k(s) + e_k(s), \quad k = 1, \dots, m$$

where the trend is a linear function of p known regressors $X_k(s)$:

$$\mu_k(s_i) = \beta_{0,k} + \sum_{j=1}^p X_{j,k}(s_{i,k})\beta_{j,k} = X_k(s_i)\beta_k, \quad i = 1, \dots, n; k = 1, \dots, m$$

with regressor vector $X_k(s_i) = (1, X_{1,k}(s_i), \dots, X_{p,k}(s_i))$ and coefficient vector $\beta_k = (\beta_{0,k}, \dots, \beta_{p,k})^T$. Let $\beta = (\beta_1^T, \dots, \beta_m^T)^T$, and let $X_k(s_i)$ be the i -th row of $X_k(s)$. Furthermore, we assume that β is known. Standard sequential Gaussian cosimulation proceeds as follows:

1. define a random path through the q simulation locations s_{n+1}, \dots, s_{n+q} , and let $i = 1$
2. at the unvisited location s_{n+i} , calculate the conditional distribution of the vector $Z(s_{n+i})$ from the m -variate Gaussian distribution with mean vector the $m \times 1$ simple cokriging prediction vector

$$\hat{z}_{\text{sk}}^*(s_{n+i}) = X_0(s_{n+i})\beta + c^T C^{-1}(Z_+ - X_+\beta)$$

and prediction error covariance matrix the $m \times m$ simple kriging prediction error covariance matrix:

$$\Sigma_{\text{sk}}(s_{n+i}) = \text{Cov}(Z(s_{n+i})) - c^T C^{-1}c$$

with the $m \times m(p+1)$ matrix $X_0(s_{n+i})$ formed by

$$X_0(s_{n+i}) = \begin{bmatrix} X_1(s_{n+i}) & 0 & \dots & 0 \\ 0 & X_2(s_{n+i}) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_m(s_{n+i}) \end{bmatrix}$$

with 0 conforming vectors of 0, where $\text{Cov}(Z(s_{n+i}))$ is the covariance matrix of $Z(s_{n+i}) = (Z_1(s_{n+i}), \dots, Z_m(s_{n+i}))^T$, where c is an $m(n+i-1) \times m$ matrix with sub-vectors

$$c = \begin{bmatrix} c_{1,1} & c_{1,2} & \dots & c_{1,m} \\ c_{2,1} & c_{2,2} & \dots & c_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m,1} & c_{m,2} & \dots & c_{m,m} \end{bmatrix}$$

with $c_{i,j} = (\text{Cov}(Z_i(s_1), Z_j(s_{n+i})), \dots, \text{Cov}(Z_i(s_{n+i-1}), Z_j(s_{n+i})))^T$, and where C is the covariance matrix of $Z_+ = (Z_1(s_1), \dots, Z_1(s_{n+i-1}), \dots, Z_m(s_1), \dots, Z_m(s_{n+i-1}))$, and where X_+ is the $m(n+i-1) \times m(p+1)$ matrix with regressor values corresponding to Z_+ (see also Ver Hoef and Cressie,

1993), having m distinct non-zero $(n+i-1) \times (p+1)$ submatrices $X_{k,+}(s)$ at their appropriate place:

$$X_+ = \begin{bmatrix} X_{1,+}(s) & 0 & \dots & 0 \\ 0 & X_{2,+}(s) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X_{m,+}(s) \end{bmatrix}$$

where 0 represent conformant submatrices filled with zeroes.

3. draw a m -vector $(\tilde{z}_1(s_{n+i}), \dots, \tilde{z}_m(s_{n+i}))$ randomly from $\mathcal{N}(\hat{z}_{\text{sk}}^*(s_{n+i}), \Sigma_{\text{sk}}(s_{n+i}))$, the multivariate normal distribution with mean vector $\hat{z}_{\text{sk}}^*(s_{n+i})$ and covariance matrix $\Sigma_{\text{sk}}(s_{n+i})$
4. insert this vector in the appropriate places of the data set; calling them $Z_1(s_{n+i}), \dots, Z_m(s_{n+i})$
5. if $i < q$, let $i = i + 1$ and go to step 2.

The same remarks that were made at the end of section *Sequential Simple Gaussian Simulation* apply here as well. If the size of a local kriging neighbourhood is held constant for each variable, simulating m variables independently would take m times the computing time, whereas simulating m dependent variables would take m^2 times the computing time, because the size of $C^{-1}c$ is now m times as large.

B.2.3 Sequential Full Gaussian cosimulation

Both simulation algorithms described above assume β to be a vector with known coefficients. In most real life problems however, β is not known. We will now describe the simulation algorithm that applies to the case where β is not known, but estimated from the data.

Under the multivariable general linear model $Z(s) = X\beta + e(s)$, with symbols as defined in the previous section, the best linear unbiased estimate of β is the generalised least squares estimate $\hat{\beta} = (X^T C^{-1} X)^{-1} X^T C^{-1} Z(s)$, and the corresponding estimation error covariance matrix is $\Sigma_{\beta} = (X^T C^{-1} X)^{-1}$.

The sequential full Gaussian cosimulation algorithm is very similar to the sequential simple Gaussian cosimulation algorithm, the only difference is that in step 2 and 3 the simple cokriging mean vector \hat{z}_{sk}^* is replaced by

$$\hat{z}_{\text{sk}, \tilde{\beta}^l}^*(s_{n+i}) = X_0(s_{n+i})\tilde{\beta}^l + c^T C^{-1}(Z_+ - X_+ \tilde{\beta}^l) \quad (\text{B.1})$$

where $\tilde{\beta}^l$ is a single $m(p+1) \times 1$ vector drawn randomly from $\mathcal{N}(\hat{\beta}, \Sigma_{\beta})$, the multivariate normal distribution with mean and covariance equal to the generalised least squares estimate and estimation covariance of β . For each realisation of the random field, a single value $\tilde{\beta}^l$ is drawn.

The residual fields with respect to the simulated values $\tilde{\beta}^l$ have the correct spatial and cross covariance structure because simple kriging is applied to obtain conditional distributions.

The marginal variance of a simulated value using the sequential simple Gaussian cosimulation algorithm is equal to the simple kriging variance, given the observed data alone. To see this, consider the first simulated location ($i = 1$): the value has variance Σ_{sk} , based only on observed data ($Z_+ \equiv Z(s)$ only if $i = 1$), which is the simple kriging variance from observed data. Because a random path is defined, any simulation can be the first, and therefore this argument is valid for all simulation locations.

Let $\tilde{\beta}$ be $\mathcal{N}(\hat{\beta}, \Sigma_{\tilde{\beta}})$ distributed. Under the sequential full Gaussian cosimulation algorithm, a random component due to $\tilde{\beta}$ is added to the simulations. To see how the marginal variance at a simulation location increases, consider the variability of $\hat{z}_{sk, \tilde{\beta}}^*$ given $Z(s)$ at the first (random) simulation location:

$$\begin{aligned} \text{Var}(\hat{z}_{sk, \tilde{\beta}}^* | Z(s)) &= \text{Var}((X_0(s_{n+1}) - c^T C^{-1} X) \tilde{\beta}) \\ &= (X_0(s_{n+1}) - c^T C^{-1} X) \text{Var}(\tilde{\beta}) (X_0(s_{n+1}) - c^T C^{-1} X)^T \\ &= (X_0(s_{n+1}) - c^T C^{-1} X) (X^T C^{-1} X)^{-1} (X_0(s_{n+1}) - c^T C^{-1} X)^T \end{aligned}$$

which equals the difference between the universal cokriging variance and the simple cokriging variance. It takes into account (i) the estimation error covariance of $\tilde{\beta}$, (ii) the regression extrapolation effect: to which extent should X_0 be considered an extrapolation relative to the average of the regressors of the data, $c^T C^{-1} X$, and (iii) the correlation between the two components on the right hand side of (B.1). With respect to (iii), note that as $\tilde{\beta}^t$ appears twice in (B.1), a value simulated at an observed data location equals the observed value.

B.3 Manual for gstat

An implementation of the three simulation algorithms described above is given in **gstat**, a program for geostatistical modelling, prediction and simulation (Pebesma and Wesseling, 1998).

Here, we will describe how a bivariate random field with unknown mean coefficients can be simulated with **gstat**. Suppose the two variables are called GLG and GHG, and that the trend of GLG is modelled as a linear function of P, Q and R, and the trend of GHG is modelled as a linear function of R and S. We assume an ascii, column data file called **gxg.dat** with the following columns:

column	contents
1	x-coordinate
2	y-coordinate
3	GLG
4	GHG
5	P
6	Q
7	R
8	S

and we assume that the values of P, Q, R and S are given at simulation locations in the grid maps **P.map**, **Q.map**, **R.map** and **S.map** respectively. The **gstat** command file


```

# comment lines start with a #
data(GLG): 'gxg.dat', x=1, y=2, v=3, X=5&6&7, max=20;
data(GHG): 'gxg.dat', x=1, y=2, v=4, X=7&8, max=20;
# residual variograms:
variogram(GLG): 1 exp(400); # replace with something realistic
variogram(GHG): 1.5 exp(400);
# residual cross variogram:
variogram(GLG,GHG): 0.8 exp(400);
method: gs;
masks: 'P.map', 'Q.map', 'R.map', 'R.map', 'S.map';
# mask map contain the entries for X0; their order should match the
# order of the regressor variables defined in consecutive data statements
predictions(GLG): 'GLGsim';
predictions(GHG): 'GHGsim';
# to create 10 simulations, uncomment the following line:
#set nsim = 10;

```

will create `GLGsim` and `GHGsim` as grid maps with simulated values. When more than one simulation is created during a single program run, numbers are appended to the file names defined in `predictions`.

B.4 Generality and efficiency issues in the `gstat` implementation

B.4.1 Generality

In the `gstat` implementation there is no limit to the data size n , the number of regressors p , or the number of variables m . In addition, for each variable, n and/or p may be different.

The simulation algorithm in `gstat` is modified to work for the case when a data location coincides with a simulation location, in which case $\Sigma_{\text{sk}}(s_{n+i})$ is singular.

B.4.2 Efficiency

The following implementation details improved the speed of simulation heavily:

- Multiple realisations can be generated following a single random path. This means that the expensive results, $C^{-1}c$ and the neighbourhood selection have to be calculated only once at each simulation location, and are re-used for each of s simulations that are generated in one program run (see example: e.g. `set nsim = 1000` generates 1000 realisations). Note that all simulations have to be kept in memory, requiring memory usage of $m \times s \times q \times 4$ bytes, in addition to the observed data and mask maps.
- The selection of the r nearest data locations from the augmented data set Z_+ can be an expensive operation, especially when q is large. `Gstat` uses a

fast neighbourhood selection that only requires one parameter r , based on a quadtree search index structure.

- When multiple variables share the same spatial configuration (i.e., $s_k \equiv s$), the neighbourhood selections for the first variable is copied to subsequent variables.

B.5 Tests

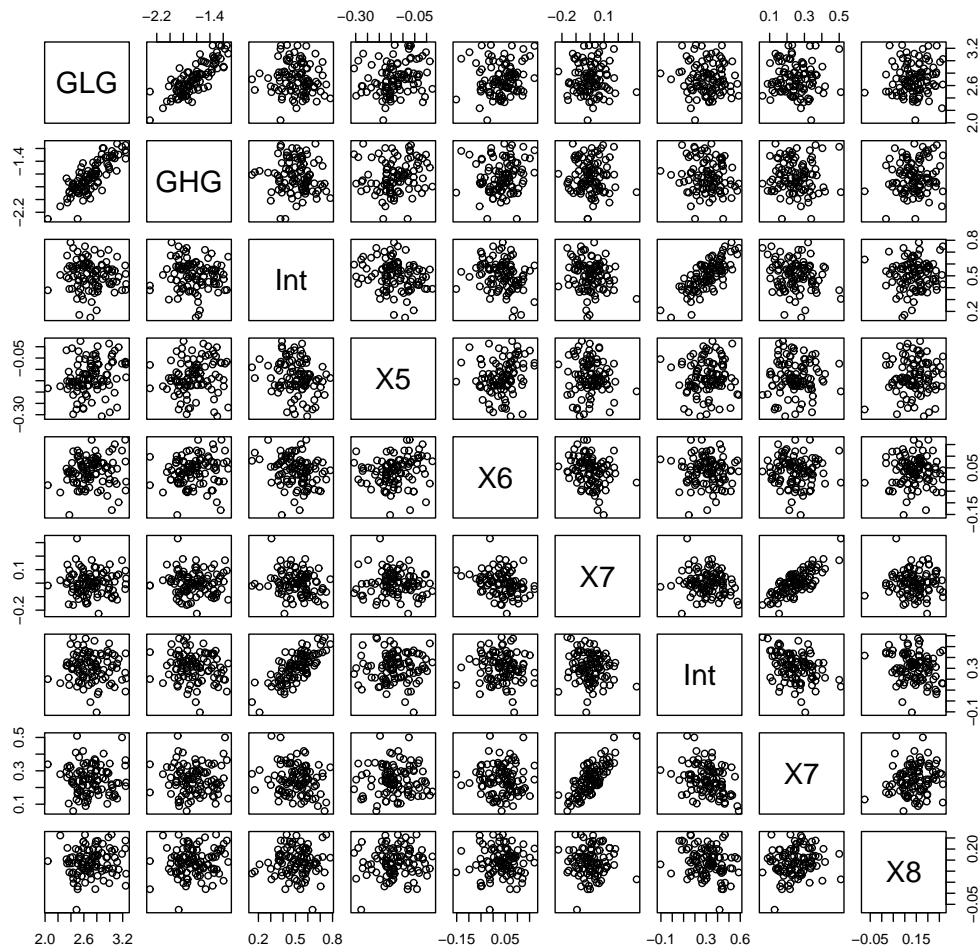


Figure B.1. scatter plot matrix of simulated GLG and GHG and simulated regression coefficients for GLG (rows/columns 3-6) and GHG (rows/columns 7-9)

In a simple test, randomly generated data were simulated in the $[0, 1] \times [0, 1]$ unit square, and simulations were done on a single location in the centre, having coordinates (0.5, 0.5). The following command file was used:

```
data(GLG): 'gxg.dat', x=1, y=2, v=3, X=5&6&7, max=20;
data(GHG): 'gxg.dat', x=1, y=2, v=4, X=7&8, max=20;
variogram(GLG): .25 exp(.1); # replace with something realistic
```

```

variogram(GHG): .25 exp(.1);
variogram(GLG,GHG): 0.2 exp(.1); # correlation is 0.8
method: gs;
data(): 'sim.loc', x=1, y=2, X=3&4&5&5&6;
# with X the columns with X0 regressors, corresponding to those
# define in consecutive data statements
output: 'sim.out';
set nsim = 100;

```

with `gxx.dat` containing 8 variables and 100 cases, filled with random uniform numbers between 0 and 1; `sim.loc` only contains the following line:

```
0.5 0.5 0.4736944 0.3986703 0.5646855 0.9540066 0.5504883
```

Figure B.1 shows simulation results using these dummy data. It shows all bivariate scatter plots between the simulated values and simulated regression coefficients.

Another test, involving a univariate simulation with a single regressor was presented in Pebesma and Heuvelink (2001).

B.6 References

- Abrahamsen, P., F. Espen Benth: Kriging with inequality constraints. *Mathematical Geology* 33 (6), 719-744, 2001.
- Deutsch, C.V., A.G. Journel: *GSLIB geostatistical software library and user's guide*, second edition. Oxford University Press, 1998.
- Goovaerts, P.: *Geostatistics for Natural Resources Evaluation*. Oxford University Press, 1997.
- Johnson, M.E., 1987, *Multivariate Statistical Simulation*. Wiley, New York, 230 pp.
- Pebesma, E.J. and Wesseling, C.G., 1998. Gstat, a program for geostatistical modelling, prediction and simulation. *Computers & Geosciences*, 24 (1), pp. 17-31; <http://www.gstat.org/>
- Pebesma, E.J., G.B.M. Heuvelink, 1999. Latin hypercube sampling of Gaussian random fields. *Technometrics* 41 (4), pp. 303-312.
- Pebesma, E.J. and G.B.M. Heuvelink, 2001. Sequential simulation of Gaussian random fields with unknown mean function: an application to heavy metal pollution data. Abstract book 4th conference of the Working Group on Pedometrics of the International Union of Soil Science (Ed. M. van Meirvenne). Ghent University, Ghent (pp. 84-84), 2001.
- Ver Hoef, Jay M., Noel Cressie, 1993. Multivariable Spatial Prediction. *Mathematical Geology*, 25 (2), pp. 219-240.

Bijlage C

Doorsnijding van percelen door stratumgrenzen

De strata zijn bodemkundig-hydrologisch uniforme deelgebieden en stratumgrenzen behoeven daarom niet samen te vallen met perceelsgrenzen. Een perceel kan dus uit delen van meerdere strata bestaan. De GxG-velden worden echter per stratum gesimuleerd, onafhankelijk van die in de andere strata (zie paragraaf 4.5). Deze onafhankelijkheid hoeft in werkelijkheid echter niet binnen het perceel aanwezig te zijn. Door deze onafhankelijkheid wel te veronderstellen bij het opschalen naar perceelsniveau, kan er teveel korte-afstandsvariatie worden weggemiddeld. Daardoor kan de berekende onzekerheid over het oppervlaktepercentage ‘uitspoelingsgevoelig’ voor een dergelijk perceel kleiner zijn dan wanneer het in zijn geheel binnen één stratum met een vergelijkbare variatie zou zijn gevallen.

De praktische consequentie van het onderschatten van de onzekerheid over het oppervlaktepercentage ‘uitspoelingsgevoelig’ is dat er een zekere tendens zal zijn om doorsneden percelen eerder als uitspoelingsgevoelig te classificeren dan niet-doorsneden percelen. De reden hiervan is dat percelen pas als uitspoelingsgevoelig worden geclassificeerd als men voldoende zeker is (zeg 95%) dat het perceel ook daadwerkelijk uitspoelingsgevoelig is. Reductie van onzekerheid leidt daarom tot een toename van als uitspoelingsgevoelig geclassificeerde percelen. Hoe sterk deze tendens is hangt van vele factoren af, en is zonder nader onderzoek niet te kwantificeren.

Dit effect van perceelsdoorsnijding is een gevolg van het feit dat de ‘beste voorhanden zijnde’ methode is gebruikt en de ‘best voorhanden zijnde’ gegevens; de bodemkundig/hydrologische stratificering is immers gericht op het bereiken van een zo groot mogelijke nauwkeurigheid. Of dit effect acceptabel is, is geen statistische maar een beleidsmatige vraag. Het alternatief zou zijn om de stratumgrenzen te laten samenvallen met perceelsgrenzen. Daaraan kleven echter de volgende bezwaren:

1. de resultaten worden minder nauwkeurig (omdat de strata heterogener zullen zijn);
2. het stratificeren wordt een omslachtiger proces;
3. perceelsgrenzen kunnen veranderen in de loop van de tijd. Dientengevolge zou de classificatie verouderen, waarbij actualisatie een complete herberekening (inclusief regressie en simulatie) zou vergen. Volgens de huidige methodiek

met vaste strata zou voor actualisatie na verandering van de percelering alleen de nabewerking van de simulatieresultaten (lokaal) moeten worden overgedaan;

4. De huidige methodiek voor classificatie maakt gebruik van de gegevens en de modellering van de Gd-kartering. Bij dit laatste is stratificatie volgens perceelsgrenzen ongewenst omdat de resultaten ervan ook voor allerlei niet-perceelsgerichte toepassingen worden gebruikt. Door met twee verschillende stratificaties te werken zou inconsistentie tussen de resultaten ontstaan.

Gezien de bezwaren die kleven aan het laten samenvallen van stratumgrenzen met perceelsgrenzen is besloten deze aanpassing van de methode achterwege te laten.

Bijlage D

Deel-onderzoek naar het benodigd aantal simulaties

Om te onderzoeken hoeveel simulaties er nodig zijn om met voldoende betrouwbaarheid te kunnen vaststellen of het voor de classificatie gekozen kanscriterium P_{crit} al dan niet wordt overschreden, is het nodig eerst enkele grootheden en bijbehorende symbolen te definiëren. Het oppervlaktepercentage van een perceel waar, gegeven de grondwater- en textuurcriteria, sprake is van uitspoelingsgevoeligheid geven we aan met O . Het is onzeker hoe groot dit percentage is, d.w.z. het kan verschillende waarden hebben tussen 0 en 100, waarbij de kans op die waarden wordt bepaald door de beschikbare gegevens en het model van de ruimtelijke variatie. Met andere woorden: O wordt als een stochastische variabele beschouwd.

Stel dat er n maal een tweetal GHG- en GLG-velden wordt gesimuleerd. Dan wordt uit elk tweetal het oppervlaktepercentage O_i , $i = 1 \dots n$, berekend (zie Hoofdstuk 5). Vervolgens worden deze waarden vergeleken met het gekozen oppervlakte-criterium O_{crit} (bijv. $O_{\text{crit}} = 67\%$), en wordt geteld hoe vaak O_{crit} wordt overschreden, d.w.z. in hoeveel van de n gevallen $O_i > O_{\text{crit}}$. We noemen dit aantal k .

De kans dat het perceel volgens de gekozen criteria uitspoelingsgevoelig is kan dan worden berekend volgens:

$$\hat{P} = k/n \tag{D.1}$$

Dit is slechts een schatting van de werkelijke kans (P), omdat er maar een beperkt aantal keren wordt gesimuleerd. Naarmate het aantal simulaties groter is, zal de geschatte kans \hat{P} de werkelijke kans P beter benaderen, maar (hoe klein ook) er zal steeds een fout in de schatting overblijven. De vraag is nu hoeveel simulaties er moeten worden uitgevoerd om ervoor te zorgen dat enerzijds de simulatie slechts een verwaarloosbare invloed heeft op de kans op misclassificaties, en anderzijds niet een onnodig grote aanslag wordt gedaan op reken- en geheugencapaciteit. Om deze vraag te beantwoorden moeten we de kans op misclassificatie nader beschouwen.

Om te beslissen of een perceel als uitspoelingsgevoelig moet worden geclassificeerd, wordt een kanscriterium gekozen: P_{crit} (bijv. 0.90 of 0.95). De classificatie-regel is dan:

”Als $\hat{P} > P_{\text{crit}}$ dan wordt het perceel als uitspoelingsgevoelig geclassificeerd; anders als niet-uitspoelingsgevoelig.”

Met het oog op de bepaling van het benodigde aantal simulaties is het handig om de vergelijking van kansen te herschrijven als een vergelijking van aantallen simulaties:

$$\hat{P} = k/n > P_{\text{crit}}, \quad (\text{D.2})$$

ofwel:

$$k > n \cdot P_{\text{crit}}. \quad (\text{D.3})$$

Hierin stelt $n \cdot P_{\text{crit}}$ het ‘kritieke aantal’ simulaties voor, verder aangeduid met k_{crit} . (Om ongewenst afrondingseffect te voorkomen nemen we aan dat n en P_{crit} zò worden gekozen dat k_{crit} een heel getal is.) De classificatie-regel luidt nu:

“Als $k > k_{\text{crit}}$ dan wordt het perceel geclassificeerd als ‘uitspoelingsgevoelig’, anders als ‘niet-uitspoelingsgevoelig’.”

De kans dat het perceel als ‘uitspoelingsgevoelig’ wordt geclassificeerd, aangeduid met P_{class} , kan worden berekend volgens:

$$P_{\text{class}} \equiv \text{Prob}(k > k_{\text{crit}}) = \sum_{k=k_{\text{crit}}+1}^n B(k, n, P) \quad (\text{D.4})$$

Hierin stelt $B(k, n, P)$ de binomiale kansverdeling voor, d.w.z. de kans op in totaal k maal een overschrijding van de oppervlakte-grens ($O_i > O_{\text{crit}}$), bij n simulaties met elk een kans P op zo’n overschrijding.

De kansen op juiste danwel onjuiste classificatie van het perceel kunnen nu worden berekend via de produktregel van kansen, als volgt:

$$\text{Prob}(\text{‘juist uitspoelingsgevoelig’}) = P \times P_{\text{class}} \quad (\text{D.5})$$

$$\text{Prob}(\text{‘onjuist uitspoelingsgevoelig’}) = (1 - P) \times P_{\text{class}} \quad (\text{D.6})$$

$$\text{Prob}(\text{‘juist niet-uitspoelingsgevoelig’}) = (1 - P) \times (1 - P_{\text{class}}) \quad (\text{D.7})$$

$$\text{Prob}(\text{‘onjuist niet-uitspoelingsgevoelig’}) = P \times (1 - P_{\text{class}}) \quad (\text{D.8})$$

We zien dus dat de kans op misclassificatie (onjuist ‘uitspoelingsgevoelig’ en onjuist ‘niet-uitspoelingsgevoelig’) afhangt van P en, via P_{class} , ook van k_{crit} (het gekozen kanscriterium) en n (het aantal simulaties).

Het effect van het aantal simulaties op de kans op misclassificatie bij $P_{\text{crit}} = 0.95$ is grafisch weergegeven in Figuur D.1, waarbij kortheidshalve ‘uitspoelingsgevoelig’ is aangeduid als ‘droog’ en ‘niet-uitspoelingsgevoelig’ als ‘nat’. De figuur linksboven geeft voor $n = 100, 200, 300$ en 1000 de kans dat het perceel als uitspoelingsgevoelig wordt geclassificeerd (P_{class}), als functie van de kans dat het perceel ook in werkelijkheid uitspoelingsgevoelig is (P). Hoe groter het aantal simulaties, hoe stijler de curve. Door vermenigvuldiging met de kans dat het perceel in werkelijkheid niet-uitspoelingsgevoelig is (weergegeven in de figuur linksmidden) vindt men de kans op onjuiste classificatie ‘uitspoelingsgevoelig’ zoals weergegeven in de figuur links onder (zie vergelijking D.5). Ook hier correspondeert de stijlste curve met de situatie van 1000 simulaties. Op dezelfde wijze ontstaat de figuur rechtsonder, welke de kans op onjuiste classificatie ‘niet-uitspoelingsgevoelig’ weergeeft. Ook hier geldt weer dat hoe groter het aantal simulaties is, hoe stijler de curve.

Uit de figuur blijkt dat bij kansen op uitspoelingsgevoeligheid juist boven P_{crit} (in dit geval dus boven 0.95), de kans op onjuist ‘uitspoelingsgevoelig’ groter wordt

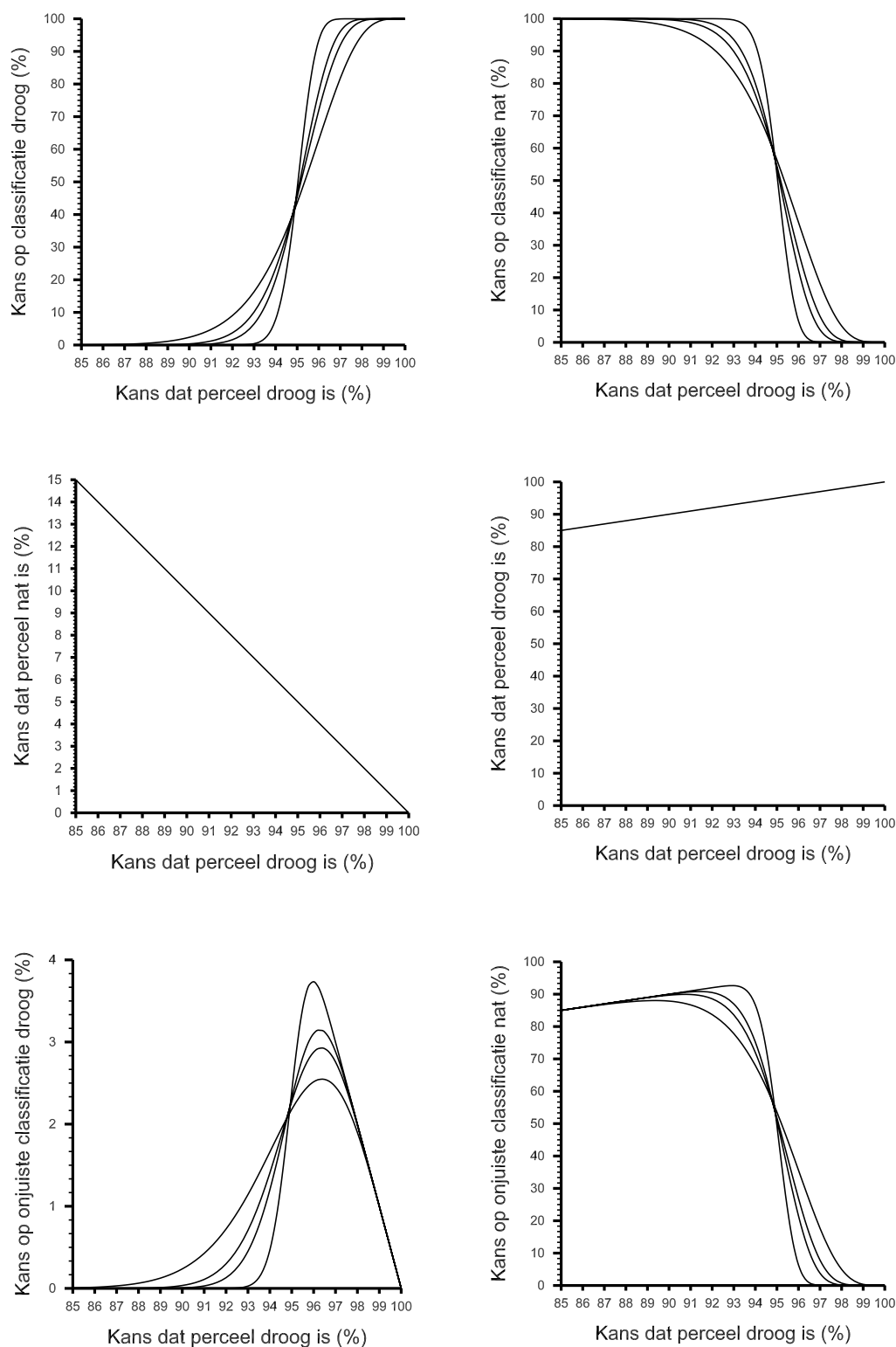
naarmate het aantal simulaties wordt verhoogd. Bij kansen op uitspoelingsgevoeligheid beneden P_{crit} gebeurt het omgekeerde. Voor de kans op onjuist ‘niet-uitspoelingsgevoelig’ geldt het tegenovergestelde. Merk op dat de kans op onjuist ‘uitspoelingsgevoelig’ bij toenemend aantal simulaties in geen geval groter is dan $(1 - P_{\text{crit}})$, in dit geval dus 0.05.

Behalve de kans op misclassificatie bij individuele percelen is ook het verwachte aantal misclassificaties in een gebied als geheel van belang. Aangezien in een gebied i.h.a. beide situaties zullen voorkomen (P boven en beneden P_{crit}), is een vorm van compensatie te verwachten door bovengenoemde tegengestelde effecten van het aantal simulaties op de kans op misclassificatie. Om dit na te gaan is voor vijf representatief geachte gebieden (Dommel en Dongestroom, Zuid Limburg, Friese Zandgronden, Regge en Dinkel, Reest en Wieden) het verwachte aantal misclassificaties berekend voor verschillende aantallen simulaties: 20, 40, 60, 80, 100, 200, 300 en 1000. Daarbij is uitgegaan van de criteria $GHG_{\text{crit}}=60$, $O_{\text{crit}}=67$ en $P_{\text{crit}}=0.95$. Het effect op het verwachte aantal onjuiste classificaties ‘uitspoelingsgevoelig’ en ‘niet-uitspoelingsgevoelig’, is weergegeven in Figuur D.2 en D.3. Het verwachte aantal misclassificaties is hierin uitgedrukt als percentage van het verwachte aantal percelen dat wordt geclassificeerd als ‘uitspoelingsgevoelig’, resp. ‘niet-uitspoelingsgevoelig’. Het verwachte aantal onjuiste classificaties ‘uitspoelingsgevoelig’ kan worden berekend door sommatie van de betreffende kansen van de percelen:

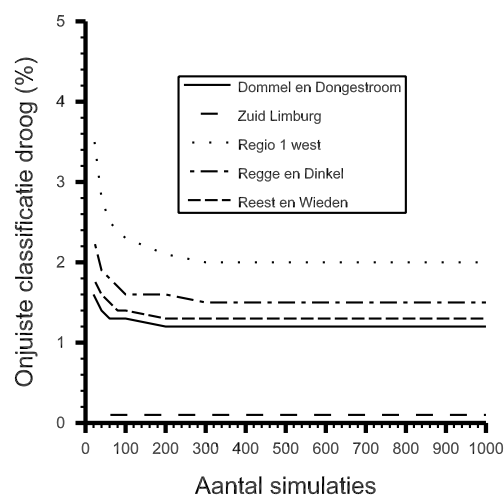
$$E(\text{aantal ‘onjuist uitspoelingsgevoelig’}) = \sum_{i=1}^N \text{Prob}_i(\text{‘onjuist uitspoelingsgevoelig’}) \quad (\text{D.9})$$

waarin $E(\cdot)$ de statistische verwachting aangeeft en N het totaal aantal percelen in het gebied voorstelt. Op dezelfde wijze is het verwachte aantal classificaties onjuist ‘niet-uitspoelingsgevoelig’ te berekenen.

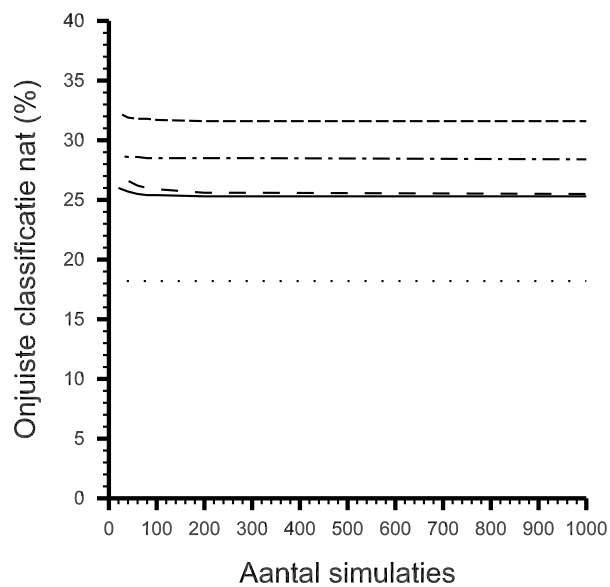
Uit Figuur D.2 en D.3 blijkt dat verhoging van het aantal simulaties van 20 tot 200 in alle vijf gebieden een vermindering van het percentage verwachte misclassificaties oplevert. Vanaf 300 blijft dit percentage echter nagenoeg constant. De conclusie is dan ook dat een aantal van 300 simulaties voldoende is om de invloed van het aantal simulaties op de kans op misclassificatie verwaarloosbaar klein te maken. Dit aantal van 300 simulaties legt daarnaast geen te grote aanslag op reken- en geheugen capaciteit. (Desgewenst is achteraf altijd te controleren of dit aantal voldoende is geweest, door dezelfde berekeningen uit te voeren als hier is gedaan voor de vijf test-gebieden.)



Figuur D.1. Effect van het aantal simulaties op de kans op misclassificatie bij $P_{\text{crit}}=0.95$. De vier curven in de grafieken gelden, in volgorde van toenemende steilte, voor 100, 200, 300 en 1000 simulaties.



Figuur D.2. Effect van het aantal simulaties op het verwachte aantal onjuiste classificaties ‘uitspoelingsgevoelig’ in vijf representatieve gebieden



Figuur D.3. Effect van het aantal simulaties op het verwachte aantal onjuiste classificaties ‘niet-uitspoelingsgevoelig’ in vijf representatieve gebieden

Bijlage E

Deel-onderzoek naar een mogelijk perceelsgrootte-effect

Op grond van de ontwikkelde methodiek valt niet bij voorbaat uit te sluiten dat het oppervlak van het perceel de kans op classificatie als uitspoelingsgevoelig mede beïnvloedt. Een eventueel grootte-effect zou veroorzaakt kunnen worden doordat door uitmiddeling de onzekerheid over de GHG en GLG voor een groot perceel geringer is dan voor een klein perceel, zodat over grote percelen minder onzekerheid resteert dan over kleine percelen. Dat zou dan het ongewenste gevolg hebben dat grote percelen een grotere kans hebben als uitspoelingsgevoelig te worden geclassificeerd dan kleine percelen.

Voor het gebied Reest en Wieden is aan de hand van de simulatie-resultaten nagegaan of er een duidelijk systematisch grootte-effect is opgetreden. In Tabel E.1 zijn, voor verschillende combinaties van grootte en (perceelsgemiddelde) GHG, de aantallen als uitspoelingsgevoelig geclassificeerde percelen weergegeven, uitgedrukt als percentage van het totaal aantal percelen met de betreffende grootte en GHG. De getabelleerde percentages betreffen het scenario met GHG criterium 60 cm, oppervlaktecriterium 66 %, en kanscriterium 95 %.

Om te beoordelen of er een duidelijk systematisch grootte-effect is opgetreden moeten de percentages worden vergeleken van percelen van verschillende grootte, maar in dezelfde GHG-klasse, d.w.z. binnen de kolommen van Tabel E.1. Er zijn binnen deze kolommen weliswaar verschillen te zien, maar deze vertonen geen duidelijke trend.

De conclusie is dan ook dat in het gebied Reest en Wieden geen duidelijk systematisch perceelsgrootte-effect is opgetreden. Een dergelijk effect is ook niet te verwachten voor andere gebieden, waar de Gd op dezelfde wijze wordt gekarteerd.

Tabel E.1. Percentages als uitspoelingsgevoelig geclassificeerde percelen per combinatie van oppervlak- en GHG-klasse

GHG (cm)														Alle GHG	Aantal percelen
Oppervlak (ha)	0-20	20-40	40-60	60-80	80-100	100-120	120-140	140-160	160-180	180-200	> 200				
0-1	0	0	0	1	10	38	67	85	96	99	100	16	8945		
1-2	0	0	0	0	7	35	68	91	95	100	100	7	4439		
2-3	0	0	0	0	8	40	80	78	100	100	100	5	3064		
3-4	0	0	0	0	5	39	82	100	92	100	100	5	2002		
4-5	0	0	0	0	6	49	84	89	83	100	100	6	1404		
5-6	0	0	0	0	8	66	89	92	100	100	100	9	919		
6-7	0	0	0	0	4	60	69	82	100	100	100	9	685		
7-8	0	0	0	0	6	36	81	100	100	100	100	9	485		
8-9	0	0	0	0	11	56	70	100	100	100	100	12	374		
9-10	*	0	0	0	4	42	78	100	75	100	100	11	304		
> 10	0	0	0	0	9	57	86	98	100	100	100	19	1537		
Alle percelen	0	0	0	0	8	42	72	89	96	99	100	11			
Aantal percelen	436	3128	8150	5764	3127	1576	806	415	283	158	316		24159		