

32/446(651)
2^e ex

Bodemschematisatie

Opschaling door aggregatie van bodeminformatie voor modelsimulaties op nationale en regionale schaal

**W.J.M. de Groot
P. A. Finke
J. Oude Voshaar
M.J.D. Hack -ten Broeke
F. de Vries
Y. van Randen**

BIBLIOTHEEK "DE HAAFF"
Droevendaalsesteeg 3a
6708 PB Wageningen

Rapport 651

DLO-Staring Centrum, Wageningen, 1998

68 ng 57212

REFERAAT

W.J.M. de Groot, P. A. Finke, J. Oude Voshaar, M.J.D. Hack -ten Broeke, F. de Vries en Y. van Randen.,1988 *Bodemschematisatie: Opschaling door aggregatie van bodeminformatie voor modelsimulaties op nationale en regionale schaal*. Wageningen, DLO-Staring Centrum. Rapport 651. 65 blz.; 9 fig.; 12 tab.; 22 ref.

Dit rapport beschrijft de ontwikkeling van geautomatiseerde methoden voor opschaling van bodeminformatie door ruimtelijke en data aggregatie voor modeltoepassingen; bodemschematisatie genoemd. Ze bestaan uit de selectie van bodemgegevens, ruimtelijke aggregatie van kaarteenheden van de bodemkaart, data aggregatie door definiering van rekenprofielen met procesparameters en het genereren van bodemkundige modelinvoer. Middels statistische analyse van 12 verschillende modeluitvoervariabelen is nagegaan welke van de ontwikkelde methoden voor bodemschematisatie het geringste informatieverlies gaf uitgaand van de ruimtelijke structuur op de Bodemkaart van Nederland, schaal 1 : 50 000. De voorspelfout binnen kaarteenheden bedroeg minimaal 45%. Het algoritme voor de ruimtelijke aggregatiemethode waarbij het studiegebied Beerse Reuzel van 203 naar 23 ruimtelijke eenheden werd teruggebracht gaf het meest representatieve beeld van de variabiliteit tussen de oorspronkelijke kaarteenheden. De toename van de voorspelfout door data aggregatie bleef door twee van de vier hiervoor ontwikkelde methoden het kleinst.

Trefwoorden: bodemkaart, data aggregatie, landhoedanigheid, modelresultaten, profielbeschrijvingen, ruimtelijke aggregatie, statistische analyse, vertaalfuncties, voorspelfout

ISSN 0927-4499

© 1998 DLO Staring Centrum, Instituut voor Onderzoek van het Landelijk Gebied (SC-DLO),
Postbus 125, NL-6700 AC Wageningen.
Tel.: (0317) 474200; fax: (0317) 424812; e-mail: postkamer@sc.dlo.nl

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

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

Projectnummer 81009

[rapport 651/wdg/12-98]

Inhoud

Woord vooraf	7
Samenvatting	9
1 Inleiding	13
2 Methoden voor bodemschematisatie	15
2.1 Een overzicht	15
2.2 Selectie van bodemkundige gegevens	17
2.2.1 Proces overlay	18
2.2.2 Bodemgegevens voor basis-rekenprofielen (Proces PUNTINFO)	19
2.2.2.1 Data	19
2.2.2.2 Beschrijving van het selectieproces	20
2.2.3 Bodemgegevens voor kaarteenheden (Proces KAARTINFO)	22
2.2.3.1 Data	22
2.2.3.2 Het selectieproces	23
2.3 Clusteren van kaarteenheden (ruimtelijke aggregatie)	24
2.3.1 Landhoedanigheden	24
2.3.2 Clustermethoden	28
2.4 Genereren van rekenprofielen (data-aggregatie)	29
2.4.1 Inleiding	29
2.4.2 Schatten van procesvariabelen uit bodemkenmerken	30
2.4.2.1 Inleiding	30
2.4.2.2 Functionaliteit	30
2.4.2.3 Toepassing	33
2.4.3 Middeling van procesparameters	33
2.4.3.1 Inleiding	33
2.4.3.2 Functionaliteit	34
2.4.3.3 Toepassing	35
2.5 Invoer voor de simulatiemodellen SWAP en ANIMO	36
3 Statistische analyse van de methoden voor bodemschematisatie	39
3.1 Inleiding	39
3.2 Het basale niveau van onzekerheid	39
3.3 Het informatieverlies ten gevolge van aggregatie	40
3.4 Oorzaak van het informatieverlies	41
4 Toepassing van bodemschematisatie in studiegebied Beerse Reuzel	43
4.1 Gebiedsbeschrijving	43
4.2 Basisgegevens	43
4.3 Randvoorwaarden modelberekeningen	44

4.3.1	Weersgegevens	44
4.3.2	Gewas en bemesting	44
4.3.3	Hydrologie	44
4.3.4	Bodemfysica	45
4.3.5	Rekensessies	45
5	Resultaten	47
5.1	Ruimtelijke aggregatie	47
5.2	Modelresultaten	47
5.3	Vergelijking van aggregatie-alternatieven	48
5.3.1	Inleiding	48
5.3.2	Voorspelfouten ten gevolge van schematisatie	48
5.3.3	Relatieve voorspelfouten ten gevolge van schematisatie	49
5.3.4	Keuze uit aggregatiemethoden	50
5.3.5	Bronnen van onzekerheid	51
5.4	De beschikbare data	52
6	Conclusies en aanbevelingen	53
	Literatuur	55
	Aanhangsels	
	Aanhangsel 1 Te selecteren bodemgegevens voor de basis-rekenprofielen uit BIS	57
	Aanhangsel 2 Format basis-rekenprofiel	58
	Aanhangsel 3 Datamodel voor de selectie van bodemkundige gegevens	60
	Aanhangsel 4 Lijst met bodemkundige gegevens voor beschrijving van kaartenheden	62

Woord vooraf

Dit rapport bevat een beschrijving van de ontwikkelde methoden voor schematisatie van de bodem en het resultaat van onderzoek naar de methode die het minst informatieverlies oplevert vergeleken met het basale niveau van schematisatie: de kaarteenheden van de bodemkaart van Nederland, schaal 1 : 50 000.

Het onderzoek maakte deel uit van een programma voor verbetering van het instrumentarium voor milieubeleidevaluatie, zoals dat begin jaren negentig is ingezet. Het is afgerond binnen het huidige programma 'Bodem- en grondwaterinformatie voor de groene ruimte'.

Bij dit onderzoek zijn behalve de auteurs veel onderzoekmedewerkers van het SC-DLO betrokken geweest bij de discussie over de koers van het onderzoek. Allen dank voor hun bijdrage.

Samenvatting

Bodem- en grondwaterkwaliteitstudies ontwikkelen zich steeds meer in de richting van scenariostudies gebaseerd op modelsimulaties. Opschalen vormt een belangrijke stap om ook nationale en regionale schaal resultaten ten kunnen verkrijgen.

In dit onderzoeksrapport worden de resultaten van de ontwikkeling van een methode, waarmee beschikbare bodemkundige informatie optimaal wordt geaggregeerd, beschreven. Deze methode wordt geacht flexibel, snel, reproduceerbaar, gebiedsonafhankelijk te zijn en rekening te houden met de gevoeligheid van de te gebruiken modellen. Als secundair doel is gesteld dat er een methode beschikbaar komt om het informatieverlies ten gevolge van aggregatie te kwantificeren. Hierdoor wordt het mogelijk de meest geschikte methode aan te geven.

Deze studie is beperkt tot de aggregatie van bodemgerelateerde modelinvoer, en het proces wordt daarom bodemschematisatie genoemd. De ontwikkelde methode gaat uit van de Bodemkaart van Nederland, schaal 1 : 50 000 en de daarbij horende profielinformatie uit het Bodemkundig Informatiesysteem (BIS). De methode genereert bodemkundige modelinvoer voor N en P-studies met SWAP93 en ANIMO versie 3.5. Bij de methode voor bodemschematisatie kunnen twee stappen worden onderscheiden:

1. Ruimtelijke aggregatie dat zich richt op de definitie van de ruimtelijke begrenzingen van rekeneenheden op basis van de bodemkaart. Het houdt rekening met de functionele verschillen tussen kaarteenheden.
2. Data-aggregatie dat zich richt op de definitie van rekeneenheden in termen van procesparameters, gebruikmakend van profielgegevens uit het BIS.

Het ontwikkelde bodemschematisatieproces beschrijft alle handelingen die nodig zijn om de bodemkundige gegevens van een bepaald studiegebied te stroomlijnen richting invoer voor modellen. Het bestaat in hoofdzaak uit vier processen:

1. Selectie. Alle bodemgegevens van het studiegebied worden uit het BIS en van de digitale bodemkaart geselecteerd.
2. Ruimtelijke aggregatie. Kaarteenheden van de bodemkaart worden gegroepeerd tot clusters op basis van functionele landhoedanigheden van de bodem als vochtleverend vermogen, fosfaatbindend vermogen etc. De waarden voor landhoedanigheden worden berekend op basis van bodemkenmerken uit profielschetsen die gelden voor kaarteenheden. Meerdere clustermethoden kunnen daarbij worden onderscheiden door verschillen in classificatie en door de keuze van landhoedanigheden.
3. Data-aggregatie. Clusters uit de ruimtelijke aggregatie worden inhoudelijk beschreven door rekenprofielen te genereren. Op basis van geselecteerde

profielen met per laag basisgegevens uit BIS worden rekenprofielen gegenereerd door eerst met vertaalfuncties extra procesparameters toe te voegen en daarna eigenschappen te middelen. Methoden voor het genereren voor rekenprofielen onderscheiden zich door de wijze waarop gemiddeld wordt.

4. Genereren van modelinvoer. Data uit rekenprofielen worden opgeslagen in invoerbestanden voor de modellen SWAP93 en ANIMO versie 3.5.

Voor elk van de genoemde processen zijn programma's geschreven om het geheel geautomatiseerd te kunnen uitvoeren.

Conclusie is dat met de ontwikkeling van methoden voor bodemschematisatie een instrument gereed gekomen, dat landsdekkend op grotendeels geautomatiseerde wijze beschikbare ruimtelijke bodembestanden en -data aggregeert voor modelstudies naar nitraat- en fosfaattransport in de bodem.

Er zijn uiteindelijk 12 verschillende schematisatiemethoden (3 methoden om te clusteren én 4 methoden om profielen te genereren) in het studiegebied Beerse Reuzel (Noord-Brabant) toegepast. Met de resulterende rekenprofielen zijn modelberekeningen uitgevoerd. Daarnaast zijn ook complete profielbeschrijvingen van locaties in het studiegebied omgezet in rekenprofielen en doorgerekend met de modellen.

De modeluitkomsten zijn geconcentreerd op vier modelparameters (N-opname door het gewas, N-uitspoeling, opslag van P in het profiel en uitspoeling van P) in drie verschillende weerjaren (1990, 1991 en 1993).

Middels een statistische analyse is vervolgens op basis van deze modeluitkomsten nagegaan bij welke van de onderzochte methoden het kleinste informatieverlies optrad. Hiervoor is gebruik gemaakt van een aantal statistische grootheden.

Het basale (onvermijdelijke) niveau van onzekerheid in de modeluitkomsten wordt weergegeven door de variabiliteit binnen kaartenheden. De voorspelfout wordt bepaald door het gemiddelde verschil tussen de gemiddelde modeluitkomst van de kaartenheid met de erin liggende puntvoorspellingen. Op dezelfde wijze kan ook de voorspelfout op het niveau van clusters worden berekend. De toename van de voorspelfout geeft aan in welke mate de door ruimtelijke aggregatie verkregen clusters van kaartenheden met de door data-aggregatie verkregen rekenprofielen het informatieverlies vergroten. Deze dimensieloze grootheid beschrijft het effect van de aggregatievariant op alle modeluitkomsten. Arbitrair is gesteld dat deze grootheid niet groter dan 1,25 mag zijn oftewel dat de voorspelfouten met niet meer dan 25% mogen toenemen.

Vervolgens is nagegaan waardoor de voorspelfouten of het informatieverlies het sterkt toeneemt. Daarvoor kan de totale variantie die door de aggregatie is veroorzaakt worden verdeeld over een drietal componenten: de onvermijdelijke variantie, de variantie als gevolg van ruimtelijke aggregatie door samenvoeging van

weinig vergelijkbare kaarteenheden en de variantie als gevolg van de constructie van een onvoldoende representatief rekenprofiel.

Het basale (onvermijdelijke) niveau van bodemschematisatie op basis van kaarteenheden leidde tot een voorspelfout ten opzichte van de gemiddelde simulatie-uitkomst met profielen op gemeten punten van minimaal 45%.

Als gevolg van bodemschematisatie neemt de voorspelfout afhankelijk van de gekozen methode en modeluitkomst toe met 0 tot 80%. De relatieve toename van de voorspelfout ten opzichte van het basale (onvermijdelijke) niveau komt gemiddeld voor alle modeluitkomsten alleen bij clustermethode 2 nooit boven 25%. Dit betekent dat het algoritme voor ruimtelijke aggregatie leidend tot 23 clusters in het studiegebied Beerse Reusel het meest representatieve beeld van de variabiliteit tussen kaarteenheden oplevert. Aangezien dit tevens de methode is die de grootste ruimtelijke aggregatie (90% reductie van het aantal ruimtelijke eenheden) veroorzaakt, is hiermee ook de meest effectieve methode gekozen. De data-aggregatiemethoden 2 en 3 leveren in het algemeen een lagere toename van de gemiddelde voorspelfout voor alle modeluitkomsten dan methoden 2 en 4. Dit betekent dat methoden waarbij vanwege dataschaarste als alternatief veel gebruik wordt gemaakt van landelijke profielschetsen of waarbij alle profieldata gemiddeld worden leiden tot relatief sterke toename van de voorspelfout.

De bijdrage aan de totale variantie van de voorspelfout door ruimtelijke aggregatie van kaarteenheden naar cluster blijkt in 96% van de berekeningen kleiner dan die van data-aggregatie bij de constructie van het rekenprofiel. Hieruit kan worden geconcludeerd dat de oorzaak van het geleden informatieverlies ligt bij de data en het dus meer nut lijkt te hebben gegevens te verzamelen, data-aggregatie algoritmen bij te stellen en/ of de vertaalfuncties te verbeteren dan betere methoden van ruimtelijke aggregatie te ontwerpen. Deze conclusie geldt met name wanneer men uitgaat van de bodemkaart als het basale (onvermijdelijke) niveau voor ruimtelijke aggregatie. Vanwege de minimale voorspelfout uitgaand van de bodemkaart van 45% lijkt het verstandig ook de ontwikkeling van aggregatietechnieken vanaf puntniveau nader te bezien.

Het verdient aanbeveling bij de verdere verfijning van aggregatietechnieken voor de gebruikte modellen aandacht te houden voor de beschikbaarheid voor voldoende chemische data en verbetering van de kwaliteit van de vertaalfuncties voor zuurstofdiffusie.

1 Inleiding

Bodem- en grondwaterkwaliteitstudies ontwikkelen zich steeds meer in de richting van scenariostudies gebaseerd op modelsimulaties. Op nationale en regionale schaal betekent dit het toenemend gebruik van simulatiemodellen om bijvoorbeeld het transport van N en P naar grond- en oppervlaktewater te berekenen. De N- en P-stromingsmodellen zijn allereerst op veldschaal ontwikkeld. Opschalen vormt een noodzakelijke stap om op nationale en regionale schaal resultaten te verkrijgen. Hierop richt zich deze studie.

Opschaling wordt meestal uitgevoerd door aggregatie van modelinvoergegevens naar een hoger schaalniveau. Het aggregeren van bodemkundige gegevens is dan ook een belangrijke fase voordat berekeningen worden uitgevoerd. Het aggregeren naar een hoger schaalniveau is vaak ook ingegeven door het gebrek aan benodigde data. Kritiek op recht toe recht aan aggregatie van bodemkundige gegevens is dat informatie die juist belangrijke invloed kan hebben op de modeluitkomsten verloren gaat. Belangrijk aandachtspunten zijn verder dat tot nu toe gebruikte methoden nauwelijks reproduceerbaar waren en dat niet altijd optimaal gebruik gemaakt werd van de beschikbare bodemkundige informatie.

Het doel van onderzoek was dan ook een methode te ontwikkelen waarbij de beschikbare bodemkundige informatie optimaal wordt geaggregeerd. Deze methode zou flexibel, snel, reproduceerbaar en gebiedsonafhankelijk moeten zijn en rekening moeten houden met de gevoeligheid voor de te gebruiken modellen. De eerste criteria duiden op een zoveel mogelijk geautomatiseerde methode. Door rekening te houden met modelgevoeligheid zou deze methode zich moeten onderscheiden van generalisatie, en wordt daarmee een gedetailleerde vorm van database-generalisatie (Bregt en Bulens, 1995). Als secundair doel van dit onderzoek volgt dan ook, dat een manier moet worden ontwikkeld om het informatieverlies ten gevolge van de invoer-aggregatie te kwantificeren.

In deze studie beperken we ons tot de aggregatie van de bodem-gerelateerde modelinvoer, en noemen dit *bodemschematisatie*. Er is geen rekening gehouden met andere gebiedsindelingen die voor regionaal modelonderzoek van belang kunnen zijn zoals regionaal waterbeheer, landgebruik en gebruik van mest. Dit kan middels geografische koppeling in een volgend stadium van de voorbereiding van modelonderzoek worden gedaan.

De te ontwikkelen methode richt zich dus op het vinden van de optimale aggregatie van de Bodemkaart van Nederland, schaal 1 : 50 000 en de daarbij behorende profielinformatie in het Bodemkundig Informatie Systeem (BIS) ten behoeve van N- en P-transportstudies met de simulatiemodellen SWAP93 en ANIMO versie 3.5. De berekeningen richten zich op het bodemsysteem tot 2 m - mv. Bij het vinden van een optimale aggregatiemethodiek is het noodzakelijk twee stappen te onderscheiden:

1. *Ruimtelijke aggregatie.* Ruimtelijke aggregatie richt zich op de definitie van de ruimtelijke begrenzingen van de rekeneenheden uit de geografische eenheden (de kaarteenheden uit de Bodem- en Gt-kaart 1 : 50 000) en houdt daarmee rekening met de functionele verschillen tussen kaarteenheden
2. *Data-aggregatie.* Data-aggregatie richt zich op de definitie van de rekeneenheden zelf in termen van procesparameters, gebruikmakend van profielgegevens uit het BIS.

Bodemschematisatie staat dus in dit rapport voor *ruimtelijke aggregatie* gevolgd door *data-aggregatie*.

In hoofdstuk 2 worden de ontwikkelde methoden voor bodemschematisatie uitgebreid toegelicht. In hoofdstuk 3 is beschreven op welke wijze de verschillende methoden onderling met elkaar worden vergeleken om het informatieverlies ten gevolge van de verschillende manieren van aggregatie te kunnen vergelijken. In hoofdstuk 4 worden de resultaten van een toepassing in het gebied Beerse reuzel beschreven. Hoofdstuk 5 geeft de conclusies en aanbevelingen.

2 Methoden voor bodemschematisatie

Voor bodemschematisatie zijn verschillende methoden ontwikkeld. In dit hoofdstuk worden deze methoden toegelicht. Het principe van het bodemschematisatieproces is steeds gelijk, maar voor de methoden waarop kaarteenheden van de bodemkaart geclusterd worden en waarmee profielen voor de clusters worden gegenereerd, zijn verschillende alternatieven ontwikkeld. Uiteindelijk zal middels een toetsing worden bepaald welke van de ontwikkelde methoden het meest bruikbaar is.

2.1 Een overzicht

Het bodemschematisatieproces zoals dat ontwikkeld is, beschrijft in logische volgorde alle handelingen die nodig zijn om de bodemkundige gegevens van een bepaald studiegebied te stroomlijnen naar invoer voor modellen. Met een gegevensstroom-diagram zijn in figuur 1 de hoofdlijnen aangegeven. Het Bodemkundig Informatie Systeem (BIS) en de bodemkaart van Nederland, schaal 1 : 50 000 vormen de gegevensbronnen. Het BIS is een ORACLE database met o.a. een grote hoeveelheid opgeslagen profielbeschrijvingen en profielschetsen. Profielbeschrijvingen zijn beschrijvingen van puntwaarnemingen. Profielschetsen zijn representatieve beschrijvingen van eenheden op de bodemkaart. De bodemkaart van Nederland schaal 1 : 50 000 is een Geografisch Informatie Systeem (GIS) digitaal beschikbaar met het programma ARC INFO.

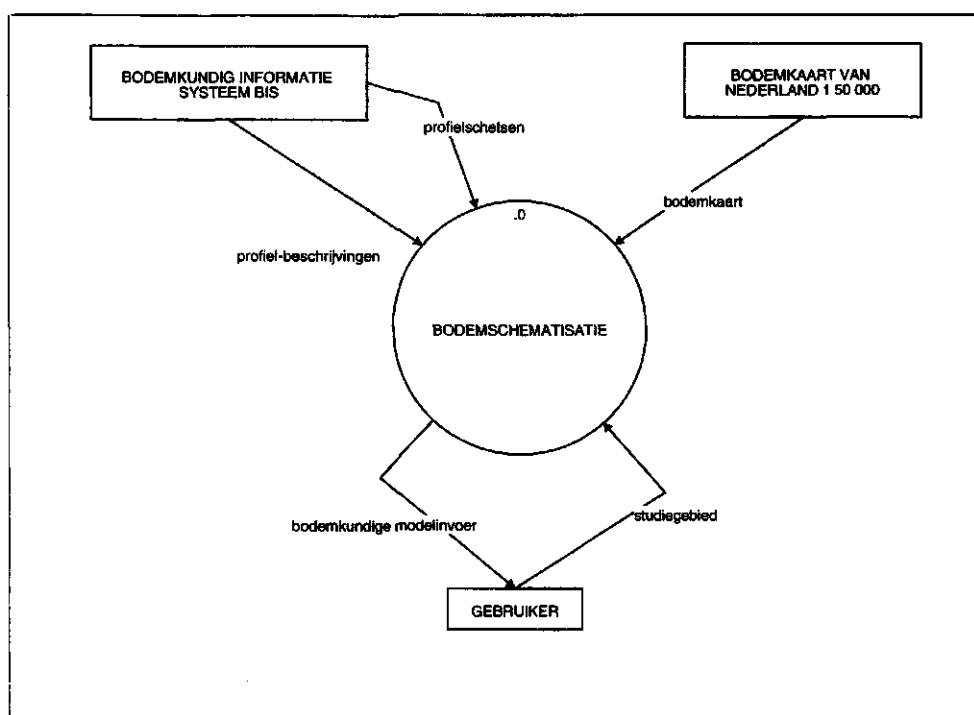


Fig. 1 Context diagram voor het bodemschematisatieproces.

Een gebruiker van het bodemschematisatieproces biedt een begrenzing van een studiegebied aan en krijgt bodemkundige modelinvoer in geschematiseerde vorm aangeboden.

Het Bodemschematisatieproces bestaat in hoofdlijnen uit vier deelprocessen.

1. *Selectie.* Allereerst worden bodemkundige gegevens van het studiegebied uit BIS en van de digitale bodemkaart geselecteerd. Alle voorkomende profielbeschrijvingen en profielschetsen worden opgeslagen in een bestand met een eigen format; de 'basis-rekenprofielen'. Daarnaast worden van elke in het studiegebied voorkomende kaarteenheden representatieve beschrijvingen opgeslagen als 'ales-profielen'.
2. *Ruimtelijke aggregatie.* Het volgende proces dat uitgevoerd wordt is de *ruimtelijke aggregatie* (clustering van kaarteenheden). Kaarteenheden worden gegroepeerd tot clusters op basis van functionele landhoedanigheden. Functionele landhoedanigheden zijn eigenschappen van de bodem die relevant zijn bij de modelberekeningen die na de bodemschematisatie volgen. Een voorbeeld is het fosfaatbindend vermogen van de grond wanneer met modelberekeningen fosfaattransport moet worden gesimuleerd. De waarden voor de landhoedanigheden worden vervolgens geclassificeerd in gradaties. Elke voorkomende combinatie van gradaties voor de functionele landhoedanigheden leidt tot een nieuw cluster. Verschillende clustermethoden onderscheiden zich door de wijze van classificatie en door de keuze van de landhoedanigheden.
3. *Data-aggregatie.* Vervolgens moeten clusters op enigerlei wijze inhoudelijk beschreven worden. Dit gebeurt in het derde proces, data-aggregatie, waarin rekenprofielen worden gegenereerd. Op basis van alle in een cluster voorkomende 'basis-rekenprofielen' wordt één 'cluster-rekenprofiel' gevormd. Van de basisrekenprofielen zijn vaak niet alle eigenschappen bekend die voor modelberekeningen nodig zijn. Voor elke laag in deze profielen worden met vertaalfuncties op basis van wel bekende bodemkenmerken extra modelparameters berekend. Vervolgens kunnen de eigenschappen van de verschillende in een cluster voorkomende profielen worden gemiddeld. Er zijn een aantal methoden om 'cluster-rekenprofielen' te genereren ontwikkeld, die zich met name onderscheiden naar de wijze waarop gemiddeld wordt.
4. *Genereren van modelinvoer.* Het format waarin 'cluster-rekenprofielen' worden opgeslagen is niet direct geschikt als invoer voor verschillende rekenmodellen. Voor het watertransportmodel SWAP93 en het nutriëntenmodel ANIMO versie 3.5 zijn specifieke invoerbestanden geformuleerd. Het proces dat modelinvoer genereert stroomlijnt deze datastroom. In feite is dit proces geen onderdeel meer van het kernproces bodemschematisatie, maar legt het een link met concrete modelberekeningen.

In figuur 2 zijn de onderscheiden deelprocessen met hun onderlinge koppeling door datastromen in beeld gebracht.

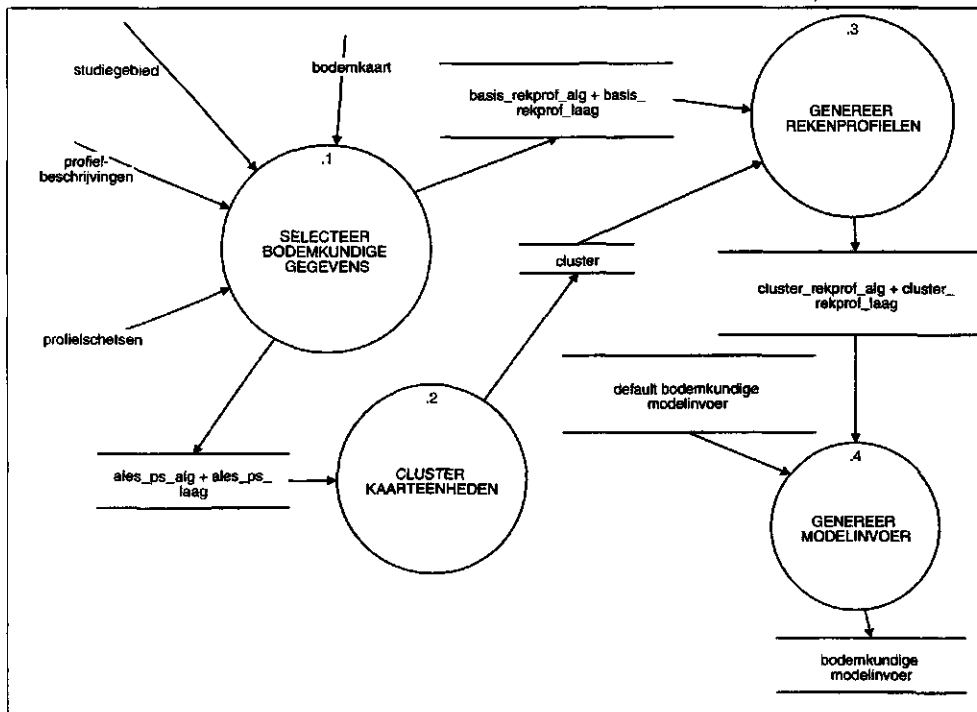


Fig. 2 Het bodemschematisatieproces

2.2 Selectie van bodemkundige gegevens

De selectie van bodemkundige gegevens uit het BIS kan verdeeld worden in drie processen: (a) een overlay van het studiegebied met de bodemkaart van Nederland 1 : 50 000; (b) De selectie van profielbeschrijvingen met relevante gegevens. Deze gegevens worden vastgelegd als 'basis-rekenprofielen'; (c) De selectie van bodemkundige gegevens voor kaartenheden van de Bodemkaart van Nederland, schaal 1 : 50 000 gelegen in het studiegebied. Deze gegevens worden opgeslagen als 'ALES-profielen'. In figuur 3 staat een gegevensstroomdiagram van de selectie van bodemkundige gegevens. In deze figuur worden de drie processen **OVERLAY**, **SELECTEER PUNTINFO** en **SELECTEER KAARTINFO** genoemd.

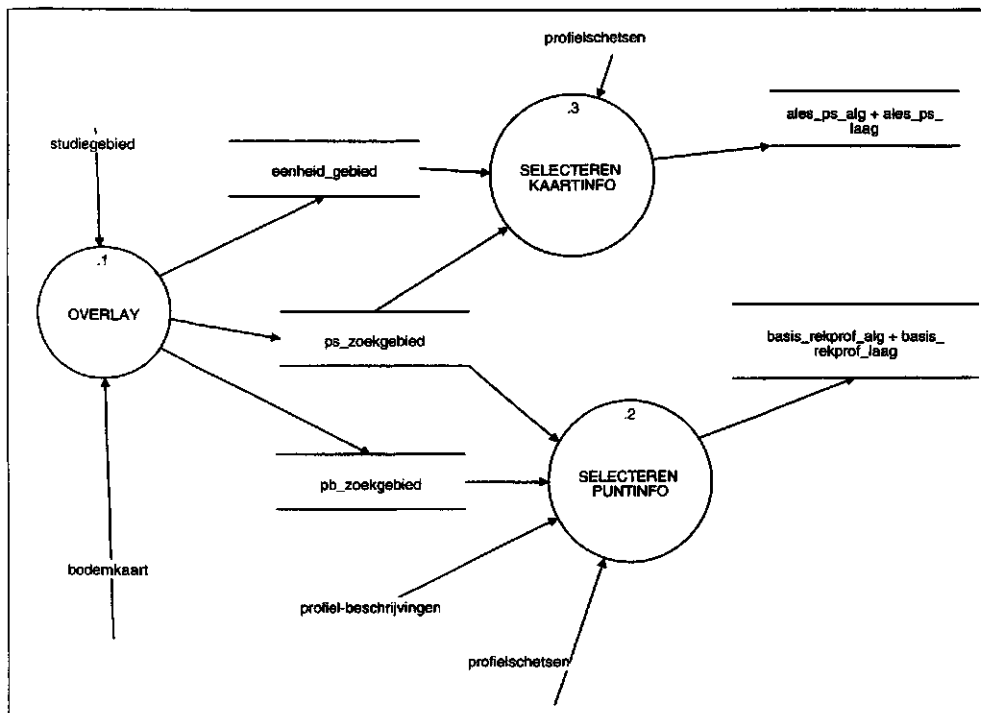


Fig. 3 Gegevensstroomdiagram van de selectie van bodemkundige gegevens.

2.2.1 Proces overlay

Het eerste proces voordat tot selectie van profielgegevens kan worden overgaan is een **OVERLAY** van studiegebied over de bodemkaart. Middels deze overlay wordt bekend welke kaarteenheden binnen het studiegebied voorkomen en welke profielen in het studiegebied liggen. In de praktijk bestaat de overlay uit twee onderdelen. De eerste overlay gebeurt met het ARC INFO programma CLIP_BOD.AML. Hiervoor is (een coverage met) de begrenzing van het studiegebied nodig. De tweede overlay wordt uitgevoerd met de commandoprocedure PB_CLIP.COM. (bestaat grotendeels uit een AML-programma). Ook hiervoor wordt de (coverage met) de begrenzing gebruikt. Er vindt dan een overlay plaats met een landelijk bestand met de lokaties van de profielbeschrijvingen (punten-coverage PB). De lokaties in het studiegebied worden in ORACLE tabel PB_zoekgebied opgeslagen. Voor het overzicht is dit voldoende. In werkelijkheid worden er twee soorten profielbeschrijvingen onderscheiden: profielbeschrijvingen (PB) en profielbeschrijvingen van de Landelijke Steekproef kaarteenheden (LSK) (zie paragraaf 2.2.1.1) ieder met hun eigen tabelstructuur.

Wanneer er onvoldoende profielbeschrijvingen binnen het studiegebied voorkomen, kan een ruimere begrenzing van het gebied worden gedefinieerd.

De selectie van de 'basis-rekenprofielen' (*PUNTINFO*) en 'ALES'-profielen (*KAARTINFO*) worden in de volgende twee paragrafen uitvoerig beschreven. Eerst volgt steeds een subparagraaf over de gebruikte data. Daarna wordt het selectieproces beschreven. Het selectieproces bestaat steeds uit een beschrijving van de geraadpleegde tabellen. Daarna wordt uitgelegd hoe de verschillende tabellen worden geraadpleegd in SQL-procedures om de juiste informatie te selecteren. De tabelnamen komen overeen met de objecten in het datamodel voor de selectie van bodemkundige gegevens dat is weergegeven in aanhangsel 3. In het datamodel zijn niet alle tabellen en views vermeld. Het is meer bedoeld om de algemene structuur van BIS en daarmee verwante gegevensbestanden aan te duiden.

2.2.2 Bodemgegevens voor basis-rekenprofielen (Proces PUNTINFO)

Deze selectie heeft tot doel zo veel mogelijk bruikbare profielgegevens te verzamelen. De geselecteerde profielen worden gebruikt om in later stadium basis-rekenprofielen mee te genereren (data-aggregatie).

2.2.2.1 Data

Van een aantal verschillende soorten profielbeschrijvingen die in het BIS beschikbaar zijn worden relevante bodemkundige gegevens geselecteerd:

- Data van profielbeschrijvingen (PB): dit zijn vrij complete puntbeschrijvingen met van bemonsterde lagen analyses van chemische en fysische parameters. Ze liggen op select gekozen punten.
- Data van profielbeschrijvingen uit de Landelijke Steekproef Kaarteenheden (LSK): dit zijn complete profielbeschrijvingen van aselect gekozen punten bedoeld om van kaarteenheden van de Bodemkaart van Nederland, schaal 1 : 50 000 statistische waarden voor bodemkenmerken vast te stellen (Visschers, 1993). Aan de profielen van vrijwel alle steekproeven zijn ook chemische analyses uitgevoerd.
- Data van representatieve profielbeschrijvingen van kaarteenheden per kaartblad ook wel profielschetsen (PS) genoemd. Dit zijn dus geen profielbeschrijvingen van punten, maar van vlakken. Ze bevatten uitsluitend schattingen van bodemkenmerken. Er zijn geen chemische bepalingen.
- Data van representatieve profielbeschrijvingen (profielschetsen) van kaarteenheden op landelijke schaal (PSL): voor elke kaarteenheid van de Bodemkaart van Nederland, schaal 1 : 50 000 is via een vertaaltabel een representatieve beschrijving bekend. Het bevat ook chemische waarden. Deze beschrijvingen zijn speciaal in het kader van dit project gemaakt. Als basis dienden een bestand met profielschetsen van de belangrijkste bodemeenheden van de Bodemkaart van Nederland 1 : 50 000 over de eerste 30 cm (De Vries, 1994), een bestand van de profielschetsen voor de Bodemkaart van Nederland, schaal 1 :

250 000 tot 120 cm - mv. (De Vries, 1993) en een bestand met profielschetsen van een groot aantal bodemeenheden van de bodemkaart van Nederland, schaal 1 : 50 000 met Fe- en Al-oxalaat gehalten van alle onderscheiden lagen per 5 cm. Het eerstgenoemde bestand is aangevuld met bodemgegevens van de ondergrond uit het als tweede genoemde bestand. Fe- en Al-oxalaatgehalten per 5 cm uit het derde bestand zijn daarna aan elke laag toegevoegd. Middels een bestand met vertalingen van de belangrijkste bodemeenheden van de Bodemkaart van Nederland, schaal 1 : 50 000 naar alle voorkomende eenheden kan vervolgens voor elke eenheid een landelijke profielschets (PSL) worden vastgesteld.

In aanhangsel 1 zijn alle bodemgegevens die uit de opgeslagen profielbeschrijvingen van het BIS worden geselecteerd, beschreven. De geselecteerde bodemgegevens worden weggeschreven in een bestand met een vast format. Dit format bevat naast de geselecteerde bodemgegevens ook ruimte voor bodemparameters die later via vertaalfuncties worden toegevoegd. In aanhangsel 2 is het format van het (basis)-rekenprofiel weergegeven.

2.2.2.2 Beschrijving van het selectieproces

De selectie van bodemkundige gegevens uit het BIS, een ORACLE database, gebeurt met selectieprogramma's in SQL (Standaard Query Language) taal. In SQL worden steeds verschillende ORACLE-tabellen aangeroepen en relevante gegevens geselecteerd. Speciaal voor de selectie van bodemgegevens voor 'basis-rekenprofielen' zijn extra tabellen aangemaakt:

- Tabel PB_VLAK geeft per PB_ALG aan in welk vlak van de bodemkaart een profielbeschrijving ligt. Dit vormt dus een link tussen bodemkaart en profielbeschrijving (zie paragraaf 2.2).
- Tabel PB_ST_VLAK geeft per boorpunt uit LSK (Landelijke Steekproef) aan in welk vlak van de bodemkaart het ligt. Dit vormt dus een link tussen bodemkaart en LSK- boorpunt.
- Tabel PB_ZOEKGEBIED geeft aan welke profielbeschrijvingen er in een door de gebruiker opgegeven zoekgebied liggen. Deze tabel bepaalt dus welke profielbeschrijvingen geselecteerd worden voor een studiegebied. Voor de LSK-boorstaten heet deze tabel PB_ST_ZOEKGEBIED. Er is ook een zoekgebied gedefinieerd dat geldt voor de profielschetsen:
- PS_ZOEKGEBIED_PSA. Het bevat een lijst van projecten met een volgnummer dat de volgorde aangeeft waarin naar profielschetsen kan worden gezocht.
- Tabel GEBIEDEN geeft de definitie van gebieden. Er zijn in principe twee gebieden mogelijk namelijk het studiegebied en het zoekgebied. Het zoekgebied is vaak groter om voldoende basisgegevens te kunnen selecteren.

- De tabel EENHEID_GEBIED geeft aan welke kaarteenheden uniek voorkomen in het studiegebied. Ook de oppervlakte per kaarteenheid is hierin geregistreerd.
- De tabel ASSOCIATIE_VERTALING geeft voor alle associaties aan uit welke kaarteenheden ze bestaan.
- View (tussentijdse tabel) PBCODZOEK en PBSTCODZOEK bevat selecties van punten die voorkomen in de tabellen PB_VLAK respectievelijk PB_ST_VLAK en ook in de tabel PB_ZOEKGEBIED. Daarnaast moeten de bodemcodes van het vlak waarin de punten liggen (zonder de Gt) ook voorkomen in de tabel EENHEID_GEBIED. In PB_VLAK en PB_ST_VLAK staat in het veld 'code' soms een associatiecode (U...). In de definitie van de views worden die vertaald met behulp van de tabel ASSOCIATIE_VERTALING.
- De tabellen BASIS_REKPROF_ALG en BASIS_REKPROF_LAAG definiëren de tabellen waarin de geselecteerde bodemgegevens worden opgeslagen. Hieraan gekoppeld is ook een BASIS_REKPROF_SWAP en BASIS_REKPROF_ANIMO en BASIS_REKPROF_VV met respectievelijk gegevens voor het waterhuishoudkundig model SWAP, het nutriëntenstromenmodel ANIMO en de verklaarde variantie (VV) van de gebruikte variabelen.

Al deze tabellen worden aangeropen in een aantal opeenvolgende procedures (in SQL packages genoemd). Voor elke soort te selecteren profielbeschrijving (PB, ST, LSK, PS) is een aparte procedure InsertBasisRekprof_* geschreven.

1. De selectie voor profielbeschrijvingen (PB) verloopt als volgt:
 - a. Eerst wordt algemene profielinformatie geselecteerd via GEBIEDEN en PBCODZOEK gecombineerd met PB_ALG en PROJ_KE en weggeschreven in BASIS_REKPROF_ALG (voorbeelden van gegevens zijn BODEM_C, X, Y en BEW).
 - b. Vervolgens wordt uit PB_LAAG, PB_LA_CHT, PB_LA_GRA en PB_LA_FYS relevante bodemgegevens per laag geselecteerd en opgeslagen in BASIS_REKPROF_LAAG (bijvoorbeeld HOR_CODE, RHO_D_M, PH_KCL, GEO_FOR_C, ORG, LUTUM_S, LEEM_S, M50_S).
 - c. Daarna worden voor de tabellen BASIS_REKPROF_SWAP, BASIS_REKPROF_ANIMO en BASIS_REKPROF_VV tabellen op dezelfde wijze laaginformatie uit BIS geselecteerd. Een groot aantal velden blijven voorlopig leeg, omdat er bij de volgende stap procesparameters worden ingevuld. Enkele bodemparameters die wel worden geselecteerd zijn CEC, FE_OX, AL_OX en P_TOT.
2. Bij de procedure voor profielbeschrijvingen van de landelijke steekproef (LSK) worden data uit de tabellen PB_ST_ALG en PB_ST_LAAG geselecteerd. De

view PBSTCODZOEK wordt daarbij gebruikt om te bepalen welke profielbeschrijvingen uit PB_ST_ALG geselecteerd moeten worden.

3. Bij de procedure voor profielschetsen (PS) worden data uit de tabellen PS_ALG en PS_LAAG geselecteerd.
4. Bij de landelijke profielschetsen (PSL) waarbij voor elke eenheid in het gebied (in tabel EENHEID_GEBIED) data uit de tabel DEFAULT_BOD50_PS_LAAG worden geselecteerd, worden uit de tabel ALES_PS_FE_AL via de tabel ALES_EENHEID_VERTALING Fe- en Al-oxalaatgehalten toegevoegd.

Bij de opslag van elke profielbeschrijving worden alle geselecteerde profielen gelabeld met hun herkomst. In het proces van data-aggregatie worden deze labels gebruikt om de kwaliteit van de data te kunnen onderscheiden. Profielbeschrijvingen (PB, ST) hebben een hogere kwaliteit dan profielschetsen (PS, PSL) én landelijke profielschetsen (PSL) hebben een lagere kwaliteit dan profielschetsen van de kaartbladen (PS).

2.2.3 Bodemgegevens voor kaarteenheden (Proces KAARTINFO)

Deze procedure heeft tot doel om voor elke kaarteenheid een representatieve beschrijving te selecteren. De geselecteerde profielen worden gebruikt om kaarteenheden op basis van landhoedanigheden te clusteren (ruimtelijke aggregatie).

2.2.3.1 Data

De meest direct bruikbare beschrijving voor de kaarteenheid in een studiegebied vormt de in vorige paragraaf genoemde profielschets (PS). Deze profielschetsen zijn echter maar voor een beperkt aantal kaartbladen beschikbaar (kaartblad 12W, 14W, 14O, 19W, 21O, 24O, 25W, 27O, 28O, 29W, 42W, 42O, 44O, 45W, 47O en 48W). Daarnaast ontbreken er voor het vaststellen van de waardering voor een aantal landhoedanigheden belangrijke bodemgegevens. Vandaar dat voor het opbouwen van een bestand met zogenaamde 'ALES-profielen' ook andere gegevensbestanden worden geraadpleegd. Ook een bestand met profielschetsen van de belangrijkste bodemeenheden van de Bodemkaart van Nederland, schaal 1 : 50 000 voor de eerste 30 cm (inclusief pH, Al- en Fe-oxalaatgehalten) en een bestand van Fe- en Al-oxalaatgehalten van alle lagen voor een aantal belangrijke kaarteenheden van de bodemkaart worden daar voor gebruikt.

In aanhangsel 4 is aangegeven welke bodemgegevens worden geselecteerd om de kaarteenheden te beschrijven.

2.2.3.2 Het selectieproces

Speciaal voor de selectie van bodemgegevens voor 'ALES-profielen' zijn extra tabellen aangemaakt. Hierna worden te selecteren tabellen beschreven in hun onderlinge relatie.

- De tabellen BASIS_PS_ALG en BASIS_PS_LAAG bevatten de data voor deze 'ALES'-profielen. Ze worden gevuld met gegevens vanuit PS_ALG en PS_LAAG van profielschetsen of van de tabellen PROJ_KE , ALES_EENHEID_VERTALING, en ALES_DEFAULT_PS_LAAG.
- TABEL ALES_EENHEID_VERTALING bevat per bodemcode een vertaling naar een hoofd_bodemcode en een vertaling naar een fe_al_bodemcode. De fe_al_bodemcode kan weer worden gebruikt om in ALES_PS_FE_AL bodeminformatie op te zoeken.
- ALES_PS_FE_AL bevat voor elke fe_al_bodemcode per laag Al- en Fe-oxalaat waarden en dichtheid.

De hoofd_bodemcode kan worden gebruikt om wanneer profielschetsen of bodemkenmerken ontbreken 'default' informatie te verzamelen door een koppeling met de tabel ALES_DEFAULT_PS_LAAG en/ of met ALES_DEFAULT_PS_PH te leggen. De default informatie bestaat dus bodemgegevens uit bestanden die via een vertaling voor alle bodemeenheden van de Bodemkaart van Nederland, schaal 1 : 50 000 beschikbaar worden gemaakt.

Deze tabellen worden aangeroepen in een procedure. De procedure InsertBasisPS creëert Basisprofielschetsen in drie stappen:

1. Eerst worden profielschetsen van PS_ALG uit het zoekgebied geselecteerd en in de tabel BASIS_PS_ALG opgeslagen.
2. Vervolgens wordt informatie geselecteerd voor kaartenheden zonder profielschetsen in het zoekgebied uit de tabel ALES_DEFAULT_PS_LAAG. Aan deze selectie wordt de pH van de eerste horizont uit ALES_DEFAULT_PS_PH via de koppeling in tabel ALES_EENHEID_VERTALING toegevoegd. Grondsoort wordt daarna vastgesteld op basis van letter- en cijferdeel van de code voor de eenheid.
3. De laaginformatie voor de profielschetsen wordt vervolgens geselecteerd uit PS_LAAG en in de tabel BASIS_PS_LAAG opgeslagen. Van eenheden waarvoor geen profielschetsen in het zoekgebied voorkomen, wordt analoog aan de algemene informatie uit ALES_DEFAULT_PS_LAAG via ALES_EENHEID_VERTALING laaginformatie geselecteerd.

2.3 Clusteren van kaarteenheden (ruimtelijke aggregatie)

Alle kaarteenheden moeten op een of andere manier onderling vergelijkbaar worden gemaakt om ze daarna te kunnen clusteren. Het onderling vergelijkbaar maken gebeurt door toepassing van methoden uit de kwalitatieve landevaluatie (Van Soesbergen et al., 1986). Het samenvoegen van kaarteenheden kan daarna op verschillende wijze gebeuren. Hiertoe zijn een aantal verschillende clustermethoden ontwikkeld. In figuur 4 is een gegevensstroomdiagram weergegeven dat de belangrijkste processen in onderlinge samenhang weergeeft.

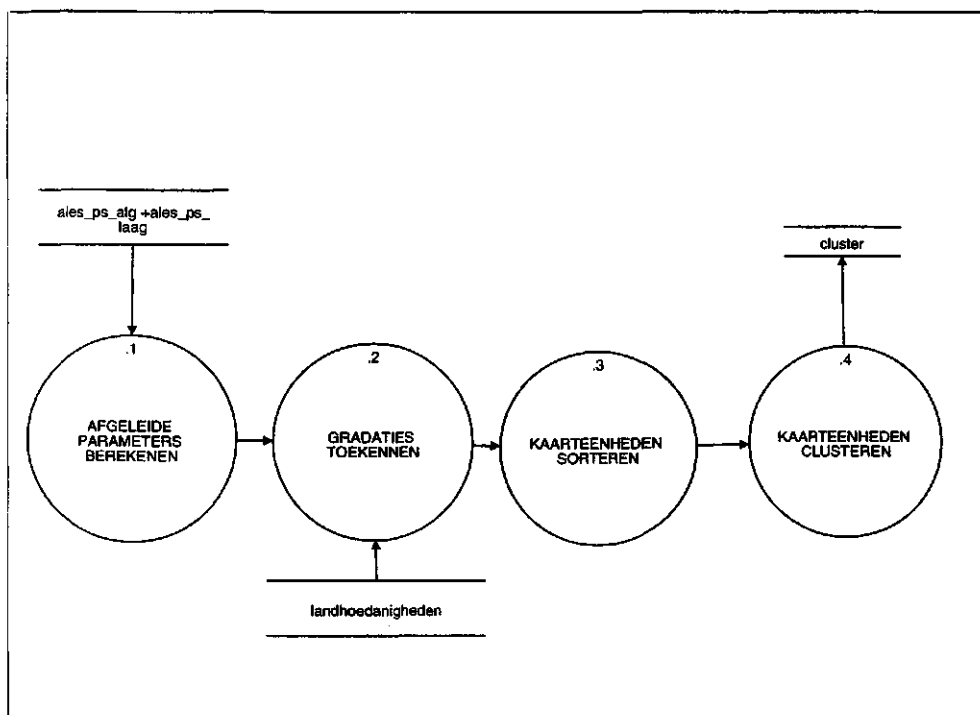


Fig. 4 Gegevensstroomdiagram voor het clusteren van kaarteenheden

2.3.1 Landhoedanigheden

Van alle kaarteenheden van de bodemkaart in een studiegebied worden aan een aantal landhoedanigheden (eigenschappen van de bodem die voor de toepassing relevant zijn) zijn waarderingen toegekend. Deze semi-kwantitatieve landevaluatieprocedure gebeurt op basis van de informatie uit de 'ALES-profielen'. De volgende landhoedanigheden zijn bij modelberekeningen over het transport van nitraat en fosfaat relevant:

1. vochtleverend vermogen per landgebruik
2. ontwateringstoestand
3. mineralisatiecapaciteit
4. fosfaatbindend vermogen
5. pH per landgebruik
6. kationenadsorptiecapaciteit

Deze landhoedanigheden worden hieronder kort omschreven:

1. Voor het *vochtleverend vermogen per landgebruik* zijn de volgende parameters van belang: bewortelbare diepte per landgebruik (akker-, weide- en bosbouw), kritieke z-afstand (maximale hoogte vanaf grondwater tot onderkant wortelzone waarbij nog voldoende capillaire opstijging plaatsvindt), GHG, GLG, grondsoort, leem-, lutum- en organischestofgehalte van de bewortelbare zone en de ondergrond. De kritieke z-afstand komt niet direct uit de selectie van bodemgegevens uit BIS beschikbaar. Het wordt op basis van de andere genoemde bodemkenmerken berekend (Hendriks et al., in voorb.).
2. Voor de *ontwateringstoestand* wordt gebruik gemaakt van de GHG en de grondsoort (Van Soesbergen et al., 1986).
3. De *mineralisatiecapaciteit* wordt grotendeels bepaald door de organische stofvoorraad (Ruitenbergh et al., 1991). De volgende bodemkenmerken zijn van belang: grondsoort, veensoort en organischestofgehalte van de bovengrond.
4. Het *fosfaatbindend vermogen* wordt berekend volgens het protocol voor bepaling van de fosfaatverzadiging (Breeuwsma et al., 1986); daarvoor zijn benodigd: Fe- en Al-oxalaatgehalte, GHG en dichtheid met de formule: $FBV = 0,5 * (Fe_{ox} + Al_{ox}) * laagdikte * dichtheid * 7,1$, waarbij voor de laagdikte de GHG of max. 1 m wordt ingevuld.
5. De *pH* komt rechtstreeks uit de selectie van bodemgegevens. Er wordt onderscheid gemaakt naar landgebruik in akker- weide- en bosbouw.
6. De *kationenadsorptiecapaciteit* (of CEC) wordt berekend volgens Breeuwsma (1986) op basis van dichtheid, organischestofgehalte en lutumgehalte. In formule $CEC = 0,1 * laagdikte * dichtheid * (p * organischestofgehalte + 0,006 * lutumgehalte)$ met: $p = 0,025$ voor B-horizonten en $p = 0,015$ voor andere horizonten

Van de meeste landhoedanigheden worden waarden uitgerekend op basis van genoemde formules. Aan sommige landhoedanigheden kunnen rechtstreeks via vertaling waarderingen worden toegekend. Daarna worden er middels classificatie gradaties toegekend aan alle landhoedanigheden van de kaarten van de bodemkaart van Nederland, schaal 1 : 50 000 gelegen in het studiegebied.

De toekenning van de gradaties de uiteindelijke groepering van kaarten in clusters op basis van deze gradaties van verschillende landhoedanigheden (de

landevaluatie) wordt uitgevoerd met het programma ALES (Automated Landevaluation System; Rossiter, 1990). Dit programma is een raamwerk waarin beslisbomen voor het toekennen van waarderingen aan landhoedanigheden kunnen worden opgebouwd. Recent is door Hendriks et al. (in voorb.) op deze wijze het kennissysteem KLASSE opgezet dat gebaseerd is op de kennis uit de WIB-C (Werkgroep Interpretatie Bodemkaarten, stadium C) (Van Soesbergen et al., 1986). De beslisbomen voor de landhoedanigheden vochtleverend vermogen en ontwateringstoestand zijn hiervan overgenomen. De andere landhoedanigheden zijn toegevoegd. ALES is niet geschikt om ingewikkelde rekenprocedures mee uit te voeren. Bij enkele landhoedanigheden moeten uit de bodemgegevens van 'ALES-profielen' meer complexe bodemeigenschappen worden berekend. Dit geldt bijvoorbeeld voor het berekenen van de kritieke z-afstand die nodig is om het vochtleverend vermogen te berekenen. Hetzelfde geldt voor de berekening van de CEC en FBV (zie vermelde formules).

In de hier volgende tabellen staat voor elk van de landhoedanigheden aangegeven hoe op basis van een berekende waarde gradaties kunnen worden toegekend. Er zijn twee indelingen. De oorspronkelijke gradatie-indeling en een indeling in drie gradaties. In de volgende paragraaf wordt uitgelegd waarom twee indelingen zijn gemaakt.

Tabel 1 Indeling in gradaties voor het vochtleverend vermogen

Oorspronkelijke indeling in gradaties	Vereenvoudigde indeling in 3 gradaties	Beschikbaar vocht (mm)
1	1	>200
2	1	150 -200
3	2	100 -150
4	3	50 -100
5	3	<50

Tabel 2 Indeling in gradaties voor de ontwateringstoestand

Oorspronkelijke indeling in gradaties	Vereenvoudigde indeling in 3 gradaties	GHG(cm - mv.)	Grondsoort
1	1	>80	
2	2	40 -80	
3	2	25 -40	
4	3	15 -25	
4	3	5 -15	Zand, leem en klei
5	3	5 -15	Veen, moerig
5	3	≤5	

Tabel 3 Indeling in gradaties voor de mineralisatiecapaciteit ingedeeld naar organischestofgehalte

Oorspronkelijke indeling in gradaties	Vereenvoudigde indeling in 3 gradaties	Grondsoort	Veensoort	Lutumgehalte (%)	Organischestofgehalte (%)
1	1	Veen	Eutroof		
2	1	Veen	Overig		
3	2	Moerig			>15
4	3	Zand, leem, klei		≤12	>9
4	3			12 -25	>10
4	3			25 -35	>11
4	3			35 -50	>12
4	3			>50	>14
5	3	Zand, leem, klei			3-8
5	3			≤12	2,5-9
5	3			12 -25	2,5-10
5	3			25 -35	3-11
5	3			35 -50	3-12
5	3		>50		4-14

Tabel 4 Indeling in gradaties voor het fosfaatbindend vermogen (FBV)

Oorspronkelijke indeling in gradaties	Vereenvoudigde indeling in 3 gradaties	FBV (ton/ha P ₂ O ₅)
1	1	≤ 5
2	1	5-10
3	2	10-15
4	2	15-20
5	3	20-30
6	3	30-40
7	3	40-50
8	3	50-60
9	3	60-70
10	3	>70

Tabel 5 Indeling in gradaties voor de pH

Oorspronkelijke indeling in gradaties	Vereenvoudigde indeling in 3 gradaties	pH
1	1	≤ 4
2	1	4-5
3	2	5-6
4	2	6-7
5	3	7-7,5
6	3	>7,5

Tabel 6 Indeling in gradaties voor de kationen adsorptiecapaciteit (CEC).

Oorspronkelijke indeling in gradaties	Vereenvoudigde indeling in 3 gradaties	CEC (mol/kg)
1	1	$\leq 0,2$
2	1	0,2-0,5
3	2	0,5-0,75
4	2	0,75-1,0
5	3	1,0-1,25
6	3	$>1,25$

Het resultaat van de toekenning van gradaties voor kaartenheden in ALES wordt gevormd door een uitvoerbestand met voor elke kaartenheid een lijst van gradaties.

2.3.2 Clustermethoden

Er zijn een aantal verschillende clustermethoden ontwikkeld. De clustermethoden onderscheiden zich in:

1. Het weglaten van minder differentiërende (= belangrijke) landhoedanigheden waardoor minder clusters resteren.
2. Het veranderen van het aantal gradaties waarin een landhoedanigheid wordt geclassificeerd. Wanneer minder gradaties worden gebruikt neemt het aantal clusters af.

Er zijn zes clustermethoden onderscheiden (tabel 7). De methoden zijn ontstaan door na te gaan wat het effect van weglaten van een landhoedanigheid op het aantal overgebleven clusters had. Landhoedanigheden zijn belangrijker wanneer ze bij weglaten een grote vermindering van het aantal clusters veroorzaken. De pH en vochtleverend vermogen onderscheiden naar landgebruik vormden minder differentiërende factoren. Het terugbrengen van het aantal gradaties per landhoedanigheid tot drie leverde op soortgelijke wijze andere clustermethoden op. Vochtleverend vermogen is als eerste landhoedanigheid geschrapt gekozen, omdat deze landhoedanigheid samen met CEC de grootste afhankelijkheid met andere landhoedanigheden vertoonde. Het weglaten van de landhoedanigheid vochtleverend vermogen leverde vervolgens de grootste reductie van het aantal clusters op. Eventueel kunnen, wanneer een beeld is ontstaan over de afwijkingen van de modeluitkomsten ten opzichte van de oorspronkelijke bodemkaart, ook andere landhoedanigheden dan vochtleverend vermogen nog geschrapt worden. Bij de keuze is verder meegenomen dat er uiteenlopende aantallen clusters ontstaan. Hierdoor wordt onderlinge vergelijking interessanter omdat het een behoorlijke invloed op het aantal door te rekenen clusters heeft.

Tabel 7 Keuze van verschillende clustermethoden op basis van de invloed van landgebruik en gradatie-indeling

Invloed bodemgebruik	Uitgebreid (incl. weide- en bosbouw)	Beperkt (alleen akkerbouw)	
		Met vochtleverend vermogen	Zonder vochtleverend vermogen
Oorspronkelijk	U_o	B_o	B_o_v
Drie gradaties	U_3	B_3	B_3_v

Met behulp van een sorteerprogramma wordt afhankelijk van de clustermethode het bestand met kaartenheden en gradaties in een volgorde gezet. Aan elke unieke combinatie van gradaties wordt een clusternummer toegekend. Per clustermethode wordt een uitvoerbestand geschreven (extensie 'cls').

2.4 Genereren van rekenprofielen (data-aggregatie)

2.4.1 Inleiding

Het genereren van rekenprofielen voor ruimtelijk geaggregeerde clusters gebeurt in twee stappen. Het schatten van procesvariabelen met vertaalfuncties en het middelen van procesvariabelen van alle beschikbare profielen. In figuur 5 staat een gegevensstroomdiagram dat deze processen en de datastromen weergeeft.

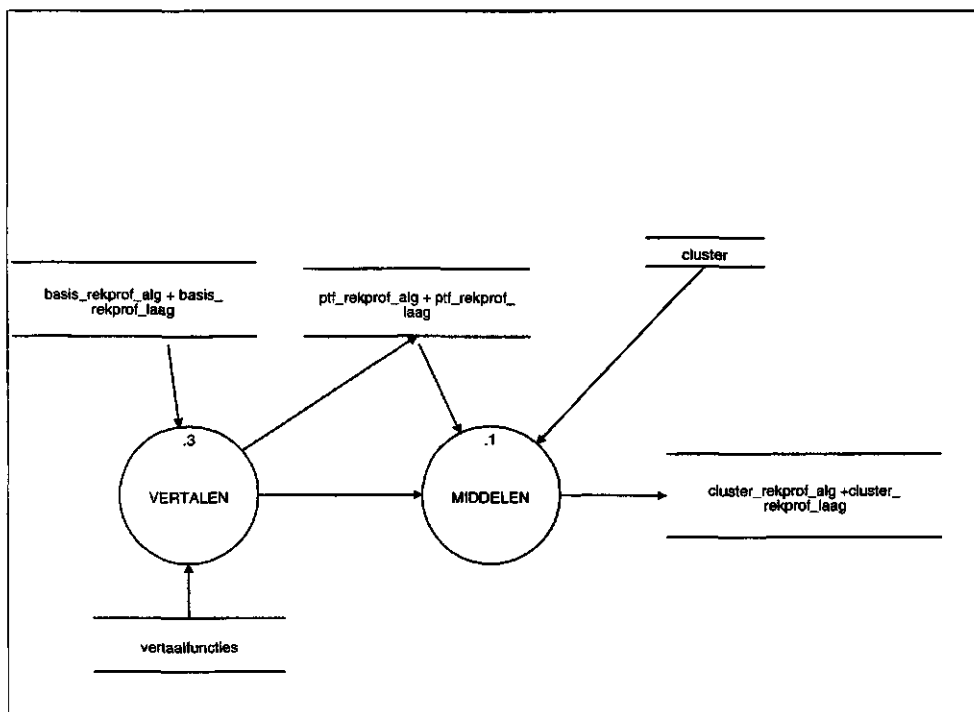


Fig. 5 Gegevensstroomdiagram van het genereren van rekenprofielen (data-aggregatie).

2.4.2 Schatten van procesvariabelen uit bodemkenmerken

2.4.2.1 Inleiding

Een onderdeel van de invoergegevens voor modellen bestaat uit procesvariabelen. Omdat de waarden van deze variabelen in het algemeen niet in voldoende mate zijn gemeten, moeten ze worden geschat. Hiertoe zijn in het verleden vertaalfuncties bepaald. De vertaalfuncties die in deze paragraaf worden beschreven leveren procesvariabelen op voor de modellen SWAP en ANIMO. Binnen het bodemschematisatie-instrumentarium is het schatten van procesvariabelen geïmplementeerd in het MS-DOS-programma PTF.EXE. Hieronder wordt kort ingegaan op de functionaliteit en de toepassing van dit programma. Voor een uitgebreide beschrijving wordt verwezen naar Finke (1996a).

2.4.2.2 Functionaliteit

De variabelen waarvoor in PTF.EXE een waarde wordt geschat, alsmede de wijze van schatting worden hieronder toegelicht:

1. Bulkdichtheid (kg/m^3).

Deze variabele wordt geschat per bodemlaag. Er zijn 3 mogelijke functies:

- Indien er sprake is van een zandlaag, wordt de bulkdichtheid geschat uit het organischestofgehalte (%), het leemgehalte (%), de mediaan van de zandfractie (micrometer) en de variabele 'wel of niet bovengrond'. Deze vertaalfunctie is afkomstig uit Wösten (1997);
- Indien er sprake is van een kleilaag, wordt de bulkdichtheid geschat uit het organischestofgehalte (%) en het kleigehalte (%). Deze vertaalfunctie is afkomstig uit Wösten (1997);
- Indien er sprake is van een veenlaag, wordt de bulkdichtheid geschat met de formule van Boekel, met de variabelen 'verzadigd vochtgehalte' uit de Staringreeks en het organischestofgehalte (%).

2. pH-H₂O

De pH-H₂O wordt geschat per bodemlaag. Er zijn drie mogelijke functies, alle gemaakt binnen dit project:

- Zandlaag: geschat uit het organischestofgehalte (%) en de pH-KCl
- Kleilaag: geschat uit de pH-KCl
- Veenlaag: geschat uit de pH-KCl

3. CEC (mol/kg)

De CEC wordt geschat per bodemlaag met de vertaalfunctie van Breeuwsma et al. (1986):

- B-horizont: geschat uit organischestofgehalte (%) en kleigehalte (%);
- overige horizonten: geschat uit organischestofgehalte (%) en kleigehalte (%).

4. Van Genuchten parameter 'residuaire vochtgehalte' (cm^3/cm^3).

Deze wordt altijd gelijk gesteld aan 0.

5. Van Genuchten parameter 'verzadigd vochtgehalte' (cm^3/cm^3).

Deze variabele wordt geschat per bodemlaag. Er zijn 3 mogelijkheden:

- Zandlaag: geschat uit leemgehalte (%), bulkdichtheid (kg/m^3), mediaan van de zandfractie (micrometer) en organischestofgehalte (%). Deze vertaalfunctie is afkomstig uit Wösten et al. (1995);
- Kleilaag: geschat uit bulkdichtheid (kg/m^3), organischestofgehalte (%), kleigehalte (%) en de variabele 'wel of niet bovengrond'. Deze vertaalfunctie is afkomstig uit Wösten et al. (1995);
- Lagen met veen: de Staringreeks (Wösten et al., 1994) wordt toegepast.

6. Van Genuchten parameter 'verzadigde doorlatendheid' (cm/dag)

Deze variabele wordt geschat per bodemlaag. Er zijn 3 mogelijkheden:

- Zandlaag: geschat uit leemgehalte (%), bulkdichtheid (kg/m^3) en organischestofgehalte (%). Deze vertaalfunctie is afkomstig uit Wösten et al. (1995);
- Kleilaag: geschat uit bulkdichtheid (kg/m^3), organischestofgehalte (%) en kleigehalte (%). Deze vertaalfunctie is afkomstig uit Wösten et al. (1994);
- Lagen met veen: de Staringreeks (Wösten et al., 1994) wordt toegepast.

7. Van Genuchten parameter 'alfa' (1/cm)

Deze variabele wordt geschat per bodemlaag. Er zijn 3 mogelijkheden:

- Zandlaag: geschat uit leemgehalte (%), bulkdichtheid (kg/m^3), de variabele 'wel of niet bovengrond' en het organischestofgehalte (%). Deze vertaalfunctie is afkomstig uit Wösten et al. (1995);

- Kleilaag: geschat uit bulkdichtheid (kg/m^3) en organischestofgehalte (%). Deze vertaalfunctie is afkomstig uit Wösten et al. (1995);
- Lagen met veen: de Staringreeks (Wösten et al., 1994) wordt toegepast.

8. Van Genuchten parameter 'n' (-)

Deze variabele wordt geschat per bodemlaag. Er zijn 3 mogelijkheden:

- Zandlaag: geschat uit leemgehalte (%), de mediaan van de zandfractie (micrometer), bulkdichtheid (kg/m^3) en het organischestofgehalte (%). Deze vertaalfunctie is afkomstig uit Wösten et al. (1995);
- Kleilaag: geschat uit bulkdichtheid (kg/m^3), kleigehalte (%) en organischestofgehalte (%). Deze vertaalfunctie is afkomstig uit Wösten et al. (1995);
- Lagen met veen: de Staringreeks (Wösten et al., 1994) wordt toegepast.

9. Van Genuchten parameter 'l' (-)

Deze variabele wordt geschat per bodemlaag. Er zijn 3 mogelijkheden:

- Zandlaag: geschat uit organischestofgehalte (%) en de variabele 'wel of niet ondergrond'. Deze vertaalfunctie is afkomstig uit Wösten et al. (1995);
- Kleilaag: geschat uit bulkdichtheid (kg/m^3), kleigehalte (%) en organischestofgehalte (%). Deze vertaalfunctie is afkomstig uit Wösten et al. (1995);
- Lagen met veen: de Staringreeks (Wösten et al., 1994) wordt toegepast.

10. Zuurstofdiffusieparameter 'a' (-)

Deze variabele wordt geschat per bodemlaag. Vertaalfuncties zijn binnen het project bepaald. Er zijn 4 mogelijkheden:

- Zandlaag: geschat uit leemgehalte (%) en organischestofgehalte (%);
- Kleilaag: geschat uit kleigehalte (%);
- Venig zand en venige klei: 2
- Zandig veen, kleilig veen en veen: 0.3

11. Zuurstofdiffusieparameter 'b' (-)

Deze variabele wordt geschat per bodemlaag. Vertaalfuncties zijn binnen het project bepaald. Er zijn 4 mogelijkheden:

- Zandlaag: geschat uit leemgehalte (%) en organischestofgehalte (%);
- Kleilaag: geschat uit kleigehalte (%);
- Venig zand en venige klei: 2.5
- Zandig veen, kleilig veen en veen: 1.5

2.4.2.3 Toepassing

Invoer en uitvoerbestanden

Het programma leest één invoerbestand. Het formaat is dat van de STORE basis_rekprof_alg+basis_rekprof_laag. Een uitvoerbestand van hetzelfde formaat wordt geproduceerd, waarin de 'missing values' zijn ingevuld. Bovendien wordt een uitvoerbestand PTF.LOG geproduceerd met voortgangs- en foutmeldingen.

Programma-executie

Het programma is vanuit MS-DOS aan te roepen met nul, één of twee commandline parameters. Bij nul commandline parameters:

PTF

gaat PTF uit van standaardnamen van de in- en uitvoerbestanden: PTFIN.DAT en PTFOUT.DAT. Indien als enige commandline parameter 'interactief' wordt ingegeven:

PTF INTERACTIEF

dan gaat PTF in de interactieve modus, en kunnen interactief procesvariabelen worden gekozen en daarvoor waarden worden geschat uit bodemkenmerken. Indien twee commandline parameters worden ingegeven, bijvoorbeeld:

PTF REUSEL.BRP REUSEL.PTF

interpreteert PTF deze achtereenvolgens als de namen van het invoerbestand en het uitvoerbestand.

2.4.3 Middeling van procesparameters

2.4.3.1 Inleiding

Een onderdeel van het instrumentarium voor bodemschematisatie is het schatten van procesvariabelen voor clusters van kaarteenheden. Hiertoe worden de procesvariabelen zoals die met PTF.EXE op het niveau van punten (bodemprofielen) zijn geschat, geaggregeerd naar verondersteld representatieve waarden op het niveau van vlakken (clusters van kaarteenheden). Er zijn verschillende aggregatiemethoden geïmplementeerd in het MS-DOS programma CLUSPROF.EXE. Hieronder wordt kort

ingegaan op de functionaliteit en de toepassing van dit programma. Voor een uitgebreide beschrijving wordt verwezen naar Finke (1996b).

2.4.3.2 Functionaliteit

De aggregatie van individuele bodemprofielen naar rekeneenheden voor SWAP en ANIMO verloopt in twee opeenvolgende stappen:

- a. Kaarteenheden worden gegroepeerd naar clusters van kaarteenheden ('ruimtelijke aggregatie'). Hiervoor zijn verschillende methoden ontworpen. Het programma CLUSPROF maakt gebruik van de resultaten van deze aggregatie door een of meer bestanden met clusters in te lezen.
- b. Per cluster kaarteenheden wordt per bodemlaag van 5 cm een representatieve set procesvariabelen bepaald. Dit gebeurt binnen CLUSPROF.

Er zijn 5 verschillende methoden geïmplementeerd, die verschillen in de gegevens die bij het vaststellen van representatieve variabelen worden gebruikt:

1. Methode 1 gebruikt alleen die profielgegevens met een bodemtype dat het meest frequent binnen de rekeneenheid voorkomt (de 'dominante' kaarteenheid). Bovendien worden bij de middeling van laaggegevens alleen gegevens gebruikt die overeenstemmen met de meest frequent gevonden combinatie van horizontcode en moedermateriaal
2. Methode 2 gebruikt eveneens alleen die profielgegevens met een bodemtype dat het meest frequent binnen de rekeneenheid voorkomt (de 'dominante' kaarteenheid). Bij de middeling van laaggegevens worden nu echter alle beschikbare gegevens gebruikt.
3. Methode 3 gebruikt alle profielgegevens met een bodemtype dat binnen de rekeneenheid voorkomt. Bij de middeling van de laaggegevens worden alleen de gegevens gebruikt die overeenstemmen met de meeste frequent gevonden combinatie van horizontcode en moedermateriaal.
4. Methode 4 gebruikt eveneens alle profielgegevens met een bodemtype dat binnen de rekeneenheid voorkomt. Bij de middeling van de laaggegevens worden alle beschikbare gegevens gebruikt.
5. Methode 5 is opgezet vanuit de filosofie dat het beter is met behulp van simulatieresultaten representatieve profielinformatie voor een rekeneenheid te kiezen, dan deze informatie uit middeling te verkrijgen. Elk individueel profiel wordt dus omgezet in informatie die als invoer voor SWAP en ANIMO kan dienen.

Voor verschillende procesvariabelen worden verschillende manieren van middeling toegepast. Deze worden beschreven in tabel 8.

Tabel 8 Door CLUSPROF te middelen profielinformatie en wijze van middeling.

Relatie	Variabele	Object	Wijze van middeling *
Q(h)	A B	Rekeneenheid	Gewogen geometrische middeling
-	Bewortelbare diepte	Rekeneenheid	gewogen middeling
-	Bulkdichtheid	Laag/rekeneenheid	gewogen middeling
-	pH-H ₂ O	Laag/rekeneenheid	gewogen middeling
-	Organischestof gehalte	Laag/rekeneenheid	gewogen middeling
-	CEC	Laag/rekeneenheid	gewogen middeling
-	P-totaal	Laag/rekeneenheid	gewogen middeling
-	Al-ox	Laag/rekeneenheid	gewogen middeling
-	Fe-ox	Laag/rekeneenheid	gewogen middeling
h(theta-K)	theta-res theta-sat Ksat Alfa N L	Laag/rekeneenheid	simultane fit met RETC.EXE door tabel met gemiddelde theta en geometrisch gemiddelde K bij verschillende h; startwaarden zijn gemiddelde of geometrisch gemiddelde waarden van de Van Genuchten parameters
Zuurstof- diffusie	A B	laag/rekeneenheid	log-lineaire regressie

* De gewichten bij de middeling worden bepaald door het oppervlaktegewicht van elke waarneming, dus het oppervlak waarvoor de waarneming representatief wordt geacht, en ook door de betrouwbaarheid van de waarneming. Geschatte getallen zijn minder nauwkeurig dan gemeten getallen. Zie Finke (1996b) voor een nadere uitwerking.

2.4.3.3 Toepassing

Invoer en uitvoerbestanden

Het programma leest één invoerbestand met profielgegevens volgens het formaat van de STORE basis_rekprof_alg+basis_rekprof_laag. Dit bestand is aangemaakt met PTF.EXE . Bovendien worden één of meer invoerbestanden (STORE cluster, filenaam: *.CLS) gezocht en gelezen met daarin varianten van clustering van kaartenheden. Een aantal uitvoerbestanden van hetzelfde formaat als het invoerbestand met profielgegevens wordt geproduceerd, één voor elke combinatie van <cluster> en (5) middelingsmethoden. Deze bestanden hebben de naam CLU?PRO?.DAT, waar op de eerste ? het nummer aan de clustermethode wordt ingevuld, en op de tweede ? een nummer tussen 1 en 5 voor de middelingsmethode. Bovendien wordt een uitvoerbestand CLUSPROF.LOG geproduceerd met voortgangsen foutmeldingen.

Programma-executie

Het programma is vanuit MS-DOS aan te roepen met nul of één commandline parameters. Bij nul commandline parameters:

CLUSPROF

gaat CLUSPROF uit van een standaardnaam van het invoerbestanden: PTFOUT.DAT.

Indien één commandline parameter wordt ingegeven, bijvoorbeeld:

CLUSPROF REUSEL.PTF

interpreteert CLUSPROF deze als de naam van het invoerbestand.

2.5 Invoer voor de simulatiemodellen SWAP en ANIMO

De bodemkundige invoergegevens die per cluster in een rekenprofiel staan, zijn verzameld om modelberekeningen mee uit te voeren. SWAP93 is een versie uit 1993 van het rekenmodel SWACROP (Feddes et al., 1988) dat de waterhuishouding en gewasgroei kan simuleren. ANIMO3.5 is een recente versie van een model dat de nutriëntenhuishouding berekent met name wat betreft stikstof en fosfaat. De te gebruiken modellen SWAP93 en ANIMO versie 3.5 kennen ieder hun eigen invoerbestanden. Deze invoerbestanden bevatten verschillende plaatsen waar bodemkundige parameters dienen te worden ingevuld. In figuur 6 wordt dit proces met een gegevensstroomdiagram weergegeven. De figuur geeft tevens het onderlinge verband met ander processen in het bodemschematisatiesysteem aan.

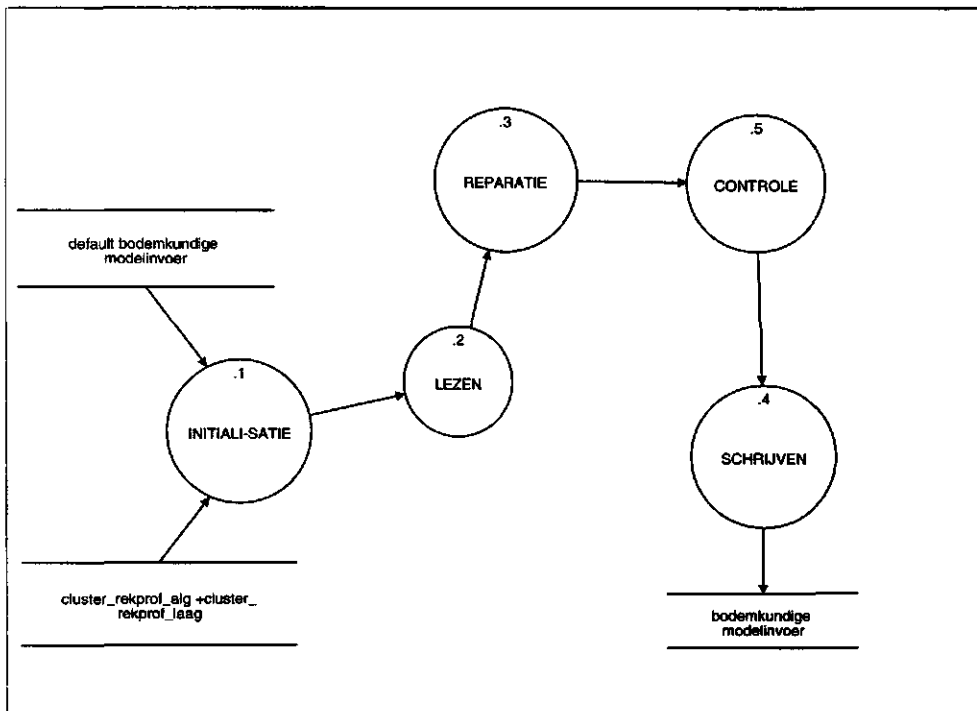


Fig. 6 Gegevensstroomdiagram voor het genereren van modelinvoer

Om de link tussen bestanden met rekenprofielen en invoerbestanden voor de twee modellen te leggen is een FORTRAN programma geschreven genaamd SWAANIIN (SWAp ANImo INvoer) (De Groot, 1997). Het bestand met clusterprofielen (CLU*PRO*.ok) wordt in dit programma ingelezen. Vervolgens worden (in steeds een andere subroutine) invoerbestanden met default-waarden ingelezen: SWATRE.INP voor SWAP93 en SWATRE.INP, INITIAL.INP, SOIL.INP voor ANIMO versie 3.5. Deze bestanden worden vervolgens weggeschreven met data van de individuele clusterprofielen. Er wordt daarbij rekening gehouden met verschillen in landgebruik. Per landgebruik zijn andere default-invoerbestanden beschikbaar.

Standaard wordt voor 40 lagen van 5 cm invoer gegenereerd. Bij sommige profielen kunnen problemen ontstaan, wanneer het grondwaterniveau beneden 200 cm – mv. komt. Vandaar dat de onderste laag standaard 200 cm dik is. Hierdoor worden profielen van bijna 4 m doorgerekend.

Daarnaast wordt op basis van het landgebruik een wortelontwikkeling gegenereerd. Aan het invoerbestand van SWAP93 worden de volgende bodemparameters toegevoegd:

- Q(h)parameters per Gt
- Van Genuchten parameters per laag.
- Beworteling op basis van bewortelbare diepte.

Aan de invoerbestanden van ANIMO 3.5 worden de volgende bodemparameters toegevoegd:

SWATRE.INP

- aantal bewortelde lagen

-

INITIAL.INP

- organische stof uitgesplitst per fractie

-

- eventueel P-totaal

SOIL.INP:

- zuurstofdiffusieparameters
- dichtheid
- verzadigde doorlatendheid
- pH
- Al_oxalaatgehalte
- Fe-oxalaatgehalte

Het P-totaal in INITIAL.INP is facultatief. Aangezien ANIMO nogal eens problemen geeft bij de P-verdeling over het profiel is gekozen voor een standaardinstelling. Wanneer die instelling ook problemen geeft, wordt uitgaand van het profiel (dikte A-horizont) de P-toestand aangepast.

Het komt regelmatig voor dat van één of enkele lagen (van 5 cm) bodemparameters ontbreken. Om te er voor te zorgen dat SWAANIIN in alle gevallen invoerbestanden kan aanmaken voor zo'n clusterprofiel is er een subroutine aan het programma toegevoegd genaamd PROFREPAIR dat bij ontbrekende laaggegevens, waarden van een andere laag toekent. Eerst wordt in de bovenliggende laag gekeken naar een waarde voor de ontbrekende parameter, vervolgens eventueel naar een diepere laag.

3 Statistische analyse van de methoden voor bodemschematisatie

3.1 Inleiding

In hoofdstuk 2 worden verschillende methoden beschreven voor ruimtelijke aggregatie (clusteren van kaarteenheden) en data-aggregatie (genereren van representatieve profielen). Om na te gaan welke methode het meest geschikt is, wordt het informatieverlies ten gevolge van de aggregatie statistisch geanalyseerd.

Het succes van een methode van ruimtelijke en data-aggregatie moet blijken uit de kwaliteit van de modelvoorspellingen met SWAP en ANIMO. Succesvolle methoden beperken het aantal door te rekenen eenheden zoveel mogelijk, waarbij het informatieverlies ten gevolge van het samenvoegen van kaarteenheden zoveel mogelijk blijft beperkt. Het informatieverlies moet worden gekwantificeerd voor verschillende modeluitvoervariabelen met betrekking tot fosfaat en nitraatuitspoeling in een studiegebied. Daarnaast is het gewenst dat het informatieverlies voor de totale set uitvoervariabelen ook wordt gekwantificeerd, zodat uitspraken over de generiek beste methode kunnen worden gedaan.

Bovenstaande overwegingen hebben geleid tot de definitie van een aantal kwantitatieve statistische grootheden. Deze worden in de volgende paragrafen beschreven.

3.2 Het basale niveau van onzekerheid

Het basale (onvermijdelijke) niveau van onzekerheid hangt samen met de variabiliteit die modeluitkomsten vertonen binnen de kaarteenheden. Kaarteenheden zijn de kleinste ruimtelijke eenheden die uit bodemschematisatie kunnen worden verkregen, namelijk als de ruimtelijke aggregatie in het geheel niet plaatsvindt. Deze onzekerheid wordt hier uitgedrukt als een voorspelfout, met als dimensie die van de modeluitvoervariabele:

$$\sqrt{S_{ke}^2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_{ke})^2},$$

waar n is het aantal punten met modeluitkomsten in kaarteenheden ke , y_i is de modeluitkomst in punt i , gebruikmakend van de profielgegevens van dat punt en \bar{y}_{ke} is het gemiddelde van de y_i in de kaarteenheden. De onvermijdelijke onzekerheid is hierboven voor één kaarteenheden uitgedrukt, maar kan voor een gebied worden berekend ($\sqrt{S_{gebied}^2}$) door eerst het oppervlaktegewogen gemiddelde van alle S_{ke}^2 te berekenen.

Dit onvermijdelijke niveau van onzekerheid wordt alleen bereikt, als het gekozen representatieve profiel voor elke kaarteenheid in precies de gemiddelde modeluitkomst binnen die kaarteenheid resulteert. Meestal is dit niet het geval, en wordt de onzekerheid op het niveau van de kaarteenheid berekend met:

$$RMSE_{ke} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_{ke})^2},$$

met y_{ke} is de modeluitkomst voor het representatieve profiel bij kaarteenheid ke (dit is dus meestal de profielschets). Er geldt dus: $RMSE_{ke} \geq \sqrt{S_{ke}^2}$.

3.3 Het informatieverlies ten gevolge van aggregatie

De onzekerheid ten gevolge van een ruimtelijke en data-aggregatie kan worden gekwantificeerd in absolute zin door opnieuw een voorspelfout te berekenen:

$$RMSE_{cluster} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_{cluster,i})^2},$$

waar $y_{cluster,i}$ is de modeluitkomst in punt i zoals die volgt uit het door data-aggregatie verkregen rekenprofiel voor de uit ruimtelijke aggregatie verkregen *cluster*, en y_i is de modeluitkomst in punt i gebruikmakend van de profielgegevens van dit punt zelf. De voorspelfout $RMSE_{gebied}$ voor een gebied wordt bepaald door eerst het oppervlaktegewogen gemiddelde van $MSE_{cluster}$ te berekenen.

De toename van de onzekerheid ten gevolge van de aggregatie kan worden gekwantificeerd met de grootheid.

$$relRMSE_{gebied} = \frac{RMSE_{gebied}}{\sqrt{S_{gebied}^2}}$$

Deze grootheid is dimensieloos, en dit heeft als voordeel dat de effecten van een aggregatievariant op verschillende modeluitkomsten goed vergelijkbaar worden. Voor elke aggregatievariant zijn 12 modeluitkomsten beschouwd:

1. de stikstofopname (kg/jaar N);
2. de stikstofuitspoeling (kg/jaar N);
3. de fosfaatuitspoeling (kg/jaar P);
4. de fosfaatopslag (kg/jaar P);

Voor een nat jaar, een droog jaar en een gemiddeld jaar. De $relRMSE_{gebied}$ kan voor elk van deze 12 modeluitkomsten worden berekend. Aan de gemiddelde waarde van $relRMSE_{gebied}$ over de 12 uitkomsten kunnen uitspraken over de generiek meest gunstige aggregatievariant worden ontleend.

3.4 Oorzaak van het informatieverlies

Het kan nuttig zijn om te weten in welke mate het opgetreden informatieverlies wordt veroorzaakt door de ruimtelijke aggregatie en in welke mate door de data- aggregatie. Als de primaire oorzaak van het informatieverlies ligt in de ruimtelijke aggregatie, dan impliceert dit dat er naar slimmere manieren moet worden gezocht om kaarteenheden te combineren. Als de oorzaak ligt bij de data- aggregatie, dan kan het nodig zijn meer gegevens te verzamelen, betere vertaalfuncties op te stellen of betere data-aggregatie algoritmen op te stellen. De onderzoeksaanpak leent zich in principe voor bovenstaande analyse, omdat de totale onzekerheid kan worden geschreven als de som van drie variantiecomponenten:

$$MSE_{cluster} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_{ke})^2 + \frac{1}{n} \sum_{i=1}^n (\bar{y}_{ke} - \bar{y}_{cluster})^2 + \frac{1}{n} \sum_{i=1}^n (\bar{y}_{cluster} - y_{cluster,i})^2$$

De eerste component is de bovenvermelde onvermijdelijke voorspelfout zonder clustering, de tweede component is groter naarmate de ruimtelijke aggregatie meer qua gedrag afwijkende kaarteenheden samenvoegt en de derde component is groter naarmate de bij de aggregatie geconstrueerde profielen minder representatief zijn. De ratio van de tweede en derde component is een dimensieloze grootte. Met deze ratio kunnen modeluitkomsten voor verschillende meteorologische jaren en verschillende modeluitvoervariabelen voor verschillende aggregatievarianten worden vergeleken. Uit de ratio blijkt welke aggregatieslag de meeste onzekerheid veroorzaakt.

4 Toepassing van bodemschematisatie in studiegebied Beerse Reuzel

In dit hoofdstuk is de toepassing van de bodemschematisatiemethode in het studiegebied Beerse Reuzel beschreven.

4.1 Gebiedsbeschrijving

Het studiegebied ligt in het midden van de provincie Noord-Brabant en is ongeveer 50 000 ha groot. In figuur 7 is de ligging van het gebied aangegeven. De bodemkaart van Nederland, schaal 1 : 50 000 bevat in het studiegebied 203 verschillende kaartenheden.



Fig. 7 Ligging van het studiegebied

4.2 Basisgegevens

Voor de selectie van bodemkundige gegevens uit het BIS is gebruik gemaakt van een zoekgebied rond het studiegebied Beerse Reuzel. Dit zoekgebied had de volgende x- en y-coördinaten van het RD systeem: linksonder 80 000, 365 000 en rechtsboven 200 000, 420 000. Voor de selectie van profielschetsen is gebruik gemaakt van de kaartbladen 44Oost en 45West.

Er zijn 122 profielschetsen (PS), 269 profielbeschrijvingen uit BIS(PB), 598 profielbeschrijvingen uit de LSK(ST) en 321 landelijke profielschetsen (PSL) geselecteerd. De landelijke profielschetsen zijn geselecteerd voor de kaarteenheden in het studiegebied. Het aantal PSL-profielschetsen is groter dan het aantal kaarteenheden in Beerse Reuzel omdat er vaak voor meerdere soorten landgebruik (akkerbouw, weidebouw en bosbouw) profielschetsen zijn.

4.3 Randvoorwaarden modelberekeningen

De modelberekeningen zijn uitgevoerd met SWAP93 (Van den Broek et al., 1994) en ANIMO versie 3.5 (Kroes en Roelsma, 1998). Bij beide modellen is het aantal functionele lagen waaruit een profiel maximaal bestaan uitgebreid tot 40. In SWAP93 is daarvoor de programmatuur enigszins aangepast.

4.3.1 Weersgegevens

Er is gebruik gemaakt van weersgegevens voor de jaren voorjaar 1988 tot en met voorjaar 1994 (zes jaren). De weersgegevens waren afkomstig van KNMI-station Twenthe met neerslag van neerslagstation Heino. De metadata zijn ook voor ander onderzoek gebruikt (Kroes et al., 1994). De gewasverdamping is berekend op basis van de referentiegewasverdamping volgens Makkink (Werkgroep Herziening Cultuurtechnisch vademecum, 1988). Van drie jaren zijn de resultaten verwerkt. Dit waren 1990 (gemiddeld met 761 mm neerslag), 1991 (droog met 546 mm neerslag) en 1993 (nat met 990 mm neerslag).

4.3.2 Gewas en bemesting

In de zomerperiode groeide er steeds snijmaïs. Met name de berekende gewasverdamping is belangrijk voor de opname van voedingsstoffen en de waterbalans. De gewasproductie is verder niet in het onderzoek betrokken.

De wortelontwikkeling is nog wel van belang. De wortelontwikkeling zoals die voor elk clusterprofiel is berekend, bepaalt vanaf welke diepte voedingsstoffen worden opgenomen.

Jaarlijks werd er in het voorjaar 75 m³ runderdrijfmest geïnjecteerd. De runderdrijfmest bestond uit 420 kg N (waarvan 188 kg N-NH₄ en 83 kg P (waarvan 60 kg P-PO₄). Daarnaast vond er een rijenbemesting met NP-kunstmest plaats (bestaand uit 20 kg/ha N en 20 kg/ha P₂O₅).

4.3.3 Hydrologie

In de balans voor watertransport door het bodemprofiel wordt rekening gehouden met grondwaterafvoer en afvoer naar het oppervlaktewater. De grondwaterafvoer is

direct gerelateerd aan de grondwaterstand via de $Q(h)$ -vergelijking van (Feddes et al, volgens Ernst (1988)). In formule:

$$Q = a * e^{-b*/h}$$

Waarin:

a en b de modelparameters

Q = afvoer naar het grondwater (cm)

h = de grondwaterstand beneden maaiveld (cm)

Deze vergelijking blijkt afhankelijk van de grondwatertrap. De Groot (in voorb.) heeft daarvoor op basis van een studie in de Achterhoek (Massop en De Wit, 1994) modelparameters per Gt afgeleid. Per Gt kan daarmee de grondwaterafvoer bij elke berekende grondwaterstand worden bepaald.

De initiële grondwaterstand bedroeg 1 m - mv. Voor de afvoer naar het oppervlaktewater is steeds uitgegaan van één drainageniveau. Het betrof een sloot met een diepte van 2 m en een radiale weerstand van 300 cm, een slootafstand van 600 m en een drainweerstand van 600 dagen. Er is speciaal met slootafvoer rekening gehouden om te zorgen dat met name bij de natte gronden de grondwaterafvoer niet overschat zou gaan worden.

4.3.4 Bodemfysica

Per laag van 5 cm zijn modelparameters voor de Van Genuchten vergelijkingen als invoer gekozen. Hiermee wordt de (on-)verzadigde doorlatendheid en waterretentie voor variërende drukhoogten gekarakteriseerd.

4.3.5 Rekensessies

De berekeningen met SWAP93 en ANIMO versie 3.5 zijn uitgevoerd op de VAX-computer. Voor elke combinatie van cluster- en profielmethode zijn commandfiles opgebouwd. Deze commandfiles lezen de juiste door het programma SWAANIIN opgebouwde invoerfiles en een aantal standaard-invoerfiles in per cluster. De invoerfiles die specifiek voor elk cluster zijn opgebouwd zijn beschreven in paragraaf 2.4. Verder zijn voor SWAP93 standaard nodig:

- HEMET**.dat (files per jaar met weersgegevens)

Voor ANIMO3.5 zijn nog nodig:

- general.inp (algemene gegevens)
- material.inp (gehalten meststoffen, oogstresten etc)
- boundary.inp (depositie en grondwatergehalten C, N en P)
- swatre.unf (uitvoerbestand per cluster van SWAP93)
- plant.inp (gewasgegevens)
- addition.inp (mestgiften, tijdstippen)

De uitvoer van ANIMO is eerst per dag berekend en in uitvoerbestanden per voorkomende N of P -bestanddeel opgeslagen (bestanden met extensie 'out').

Met het programma ANIBAL zijn balansen per jaar opgesteld. Uit deze balansfiles zijn met een selectieprogramma de juiste modeluitkomsten per jaar geselecteerd en opgeslagen in uitvoerbestanden per cluster- en profielmethode.

Van de modeluitkomsten zijn de volgende parameters gekozen om onderlinge vergelijking uit te voeren:

- N-afvoer naar het grondwater op 1 m - mv.
- N-opname door het gewas
- P-uitspoeling naar het grondwater op 1 m - mv.
- P-opslag (neerslag en adsorptie) in het profiel tot 1 m - mv.

5 Resultaten

5.1 Ruimtelijke aggregatie

Het studiegebied Beerse Reusel bestaat op de Bodemkaart van Nederland, schaal 1 : 50 000 uit 203 verschillende kaarteenheden. De 6 geformuleerde methoden van ruimtelijke aggregatie resulteerden in verschillende aantallen clusters (tabel 9).

Tabel 9 Aantal clusters per clustermethode voor het studiegebied Beerse Reusel

Clustermethode	Aantal clusters
U_o	111
U_3	40
B_o	77
B_3	31
B_o_v	54
B_3_v	23

Van deze methoden zijn met de programma's PTF en CLUSPROF voor de 4 verschillende data-aggregatiemethoden representatieve profielen vastgesteld. Er is uiteindelijk om operationele redenen voor gekozen om slechts drie van de zes onderscheiden methoden voor ruimtelijke aggregatie door te rekenen: B_o, B_3 en B_3_v. De methoden waarbij pH en vochtleverend vermogen voor drie soorten landgebruik werden onderscheiden, zijn daarmee afgefallen.

5.2 Modelresultaten

Voor vier data-aggregatiemethoden en drie methoden van ruimtelijke aggregatie zijn modelberekeningen met SWAP93 en ANIMO uitgevoerd. Van een uiteenlopend aantal clusters per combinatie van methoden zijn de berekeningen niet afgerond vanwege ernstige vertraging. In tabel 10 is het aantal wél doorgerekende clusters aangegeven.

Tabel 10 Aantal en percentage clusters per profiel- en clustermethode dat volledig is doorgerekend

Profielmethode	1	2	3	4
Clustermethode				
1 (B 3, 31 clusters)	26 (84%)	28 (90%)	26 (84%)	30 (97%)
2 (B 3 v, 23 clusters)	18 (78%)	21 (91%)	19 (83%)	22 (96%)
3 (B_o, 77 clusters)	66 (86%)	73 (95%)	71 (92%)	73 (95%)

Het blijkt dat het aantal door te rekenen clusters vooral bepaald wordt door de methode waarmee profielen zijn gegenereerd. Data-aggregatiemethode 1 en 3 leveren de grootste problemen bij het doorrekenen van de clusters. De problemen ontstonden met name bij de berekeningen met SWAP93. Veelal waren het

clusterprofielen met veenlagen (>20% organische stof) die onder extreem natte situaties problemen kregen met de waterhuishouding. Bij berekeningen met ANIMO3.5 werd steeds uitgegaan van een gemiddelde initiële fosfaattoestand. In een aantal gevallen moest deze P-toestand voor de bovengrond worden aangepast, om de berekeningen te kunnen laten voortgaan.

5.3 Vergelijking van aggregatie-alternatieven

5.3.1 Inleiding

De voorspelfouten van 3 manieren om bodemeenheden te clusteren zijn berekend voor 3 karakteristieke weerjaren (1990, 1991 en 1993) en 4 uitvoervariabelen (N-opname, N-uitspoeling, P-uitspoeling en P-opslag). Deze voorspelfouten geven weer in welke mate modeluitkomsten in punt-profielen een ander resultaat geven dan modeluitkomsten die voor dat punt zouden gelden bij de uitgevoerde bodemschematisatie. Deze voorspelfouten hebben de dimensie van de uitvoervariabele. Hierdoor wordt vergelijking van verschillende uitvoervariabelen problematisch. Bovendien geven de voorspelfouten niet aan, in welke mate de schematisatie de voorspelfout ten opzichte van de onvermijdelijke voorspelfout (bij niet-schematiseren) verslechtert. Bij het evalueren van alternatieven voor bodemschematisatie is het van belang op 3 aspecten te letten:

1. Hoe groot is de voorspelfout uitgedrukt in de eenheden van de verschillende variabelen? Met andere woorden, is het informatieverlies ten gevolge van schematisatie relevant.
 2. Hoe groot is het informatieverlies uitgedrukt ten opzichte van het onvermijdelijke informatieverlies door variabiliteit binnen kaarteenheden?
 3. Stemmen de conclusies overeen voor alle weerjaren en variabelen?
- Aspect 1 wordt behandeld in sectie 5.3.2; aspecten 2 en 3 in sectie 5.3.3.

5.3.2 Voorspelfouten ten gevolge van schematisatie

De gemiddelde simulatie-uitkomsten, de voorspelfouten ten gevolge van de variabiliteit binnen kaarteenheden en de ratio tussen deze twee getallen zijn voor alle combinaties van variabelen en weerjaren gegeven in tabel 11. Uit tabel 11 blijkt, dat de onvermijdelijke voorspelfout in het algemeen meer dan 45% van de gemiddelde simulatie-uitkomst betreft, en dat het, in de dimensie van de variabelen, om aanzienlijke hoeveelheden gaat. Bodemschematisatie zal leiden tot een toename van de ratio voorspelfout/gemiddelde, en op basis van de reeds aanzienlijke ratio's in tabel 11 is een minimalisatie van deze toename door een 'slimme' schematisatie zinvol en relevant. Enigszins arbitrair wordt daarom gesteld, dat een toename van de ratio met 25% of meer als een belangrijke verslechtering en onwenselijke situatie wordt beschouwd. Dit criterium wordt in de volgende secties toegepast om schematisatie-alternatieven te evalueren.

Tabel 11 Gemiddelde waarde, onvermijdelijke voorspelfout en de ratio voorspelfout/gemiddelde voor verschillende modeluitvoervariabelen en jaren

Weerjaar	Parameter	Dimensie	Modelvariabele			
			Nopn	Nuit	Puit	Pstor
1990	Gemiddelde	(variabele)	127,7	103,3	2,794	86,42
	Voorspelfout	(variabele)	98,82	63,02	2,375	42,81
	Fout/gemiddelde	-	0,77	0,61	0,85	0,50
1991	Gemiddelde	(variabele)	110,3	135,5	3,013	79,46
	Voorspelfout	(variabele)	89,22	85,63	2,439	35,93
	Fout/gemiddelde	-	0,81	0,63	0,81	0,45
1993	Gemiddelde	(variabele)	72,36	233,2	4,165	76,42
	Voorspelfout	(variabele)	65,37	160,9	2,443	45,91
	Fout/gemiddelde	-	0,90	0,69	0,59	0,60

Nopn = Stikstofopname (kg/ha N)
 Nuit = Stikstofuitspoeling (kg /ha N)
 Puit = Fosfaatuitspoeling (kg/ha P)
 Pstor = Fosfaatopslag (kg/ha P)

5.3.3 Relatieve voorspelfouten ten gevolge van schematisatie

Om vergelijking tussen verschillende variabelen mogelijk te maken, en tevens te kunnen aangeven in welke mate de voorspelfout ten gevolge van de aggregatie groter is dan de onvermijdelijke voorspelfout ten gevolge van de variabiliteit binnen individuele kaartenheden, wordt de relatieve voorspelfout $relRMSE_{gebied}$ berekend (hoofdstuk 3). De waarden van $relRMSE_{gebied}$ (tabel 12) laten zien dat ten gevolge van bodemschematisatie de voorspelfout tussen de 0 en 80% kan toenemen ten opzichte van de voorspelfout die wordt veroorzaakt door de variabiliteit binnen kaartenheden. Hierbij lijkt er geen groot effect te zijn van het jaar waarop de simulatie betrekking heeft, en lijken de aan fosfaat gerelateerde variabelen iets gevoeliger te zijn voor de manier van bodemschematisatie dan de aan stikstof gerelateerde variabelen. De relatief grote bijdrage van de variabiliteit binnen kaartenheden aan de totale voorspelfout (tussen 55 en 100%) geeft aan, dat het minimaliseren van informatieverlies bij ruimtelijke simulatiestudies met ANIMO en SWAP mogelijk beter zou kunnen dan met bodemschematisatie. Uitvoeren van simulaties op de puntschaal op alle locaties met voldoende gegevens, en het aggregeren van de simulatieresultaten zou mogelijk tot minder informatieverlies leiden (Finke et al., 1996). Deze aanpak is echter zeer bewerkelijk en gevoelig voor lokaal ontbrekende modelinvoergegevens.

Tabel 12 Ratio van voorspelfout ten gevolge van schematisatie en de onvermijdelijke voorspelfout binnen kaarteenheden voor verschillende modeluitvoervariabelen en jaren. Ratio's groter dan 1,25 zijn grijs aangegeven.

Methode		1990				1991				1993			
Cluster	Profiel	Nopn	Nuit	Puit	Pstor	Nopn	Nuit	Puit	Pstor	Nopn	Nuit	Puit	Pstor
1	1	1,23	1,33	1,27	1,27	1,20	1,38	1,40	1,28	1,10	1,30	1,45	1,19
1	2	1,08	1,23	1,16	1,34	1,12	1,13	1,24	1,34	1,07	1,07	1,48	1,30
1	3	1,09	1,43	1,15	1,31	1,10	1,26	1,15	1,36	0,93	1,16	1,24	1,22
1	4	1,30	1,26	1,19	1,37	1,31	1,44	1,59	1,31	1,12	1,46	1,40	1,22
2	1	1,20	1,04	1,21	1,29	1,22	1,17	1,36	1,27	1,11	1,19	1,32	1,16
2	2	1,11	1,07	1,19	1,30	1,13	1,06	1,17	1,31	1,09	1,08	1,34	1,20
2	3	1,23	1,09	1,33	1,30	1,18	1,16	1,32	1,30	1,11	1,26	1,36	1,22
2	4	1,25	1,03	1,21	1,33	1,20	1,10	1,23	1,29	1,12	1,20	1,22	1,22
3	1	1,22	1,34	1,25	1,26	1,17	1,33	1,35	1,30	1,05	1,23	1,44	1,22
3	2	1,04	1,25	1,07	1,41	1,11	1,12	1,15	1,39	1,02	1,06	1,33	1,26
3	3	1,14	1,37	1,14	1,28	1,10	1,26	1,18	1,31	0,95	1,16	1,28	1,23
3	4	1,04	1,27	1,09	1,40	1,09	1,14	1,21	1,42	0,97	1,07	1,34	1,27

Nopn = Stikstofopname (kg/ha N)
 Nuit = Stikstofuitspoeling (kg/ha N)
 Puit = Fosfaatuitspoeling (kg/ha P)
 Pstor = Fosfaatopslag (kg/ha P)

5.3.4 Keuze uit aggregatiemethoden

Omdat de getallen uit tabel 12 dimensieloos zijn, kan per aggregatievariant de gemiddelde relatieve voorspelfout worden berekend. Figuur 8 geeft een overzicht van de verschillende dimensieloze voorspelfouten als functie van methode van ruimtelijke en data-aggregatie (cluster- resp. profielmethode). Alleen relatieve voorspelfouten groter dan 125% zijn in de figuur met balken aangegeven.

Uit figuur 8 blijkt, dat clustermethode 2 in het algemeen de laagste (gunstigste) relatieve voorspelfout heeft, en (over de 4 modelvariabelen en 3 simulatiejaren gemiddeld) onder de norm van 125% blijft. Het algoritme voor ruimtelijke aggregatie leidend tot 23 clusters in het studiegebied levert blijkbaar een representatief beeld van de bodemvariabiliteit tussen kaarteenheden op. De data-aggregatiemethoden 2 en 3 leveren in het algemeen een gunstiger voorspelfout dan 1 en 4. Waarschijnlijk komt dit omdat methode 1 te selectief is met het kiezen van data uit het BIS, en er daardoor veel 'missing data' velden gevuld moeten worden met landelijke gegevens. Profielmethode 4 is juist te globaal: veel variabiliteit tussen bodemclusters wordt alsnog uitgemiddeld doordat zoveel BIS-data bij de definitie van het representatieve profiel worden betrokken.

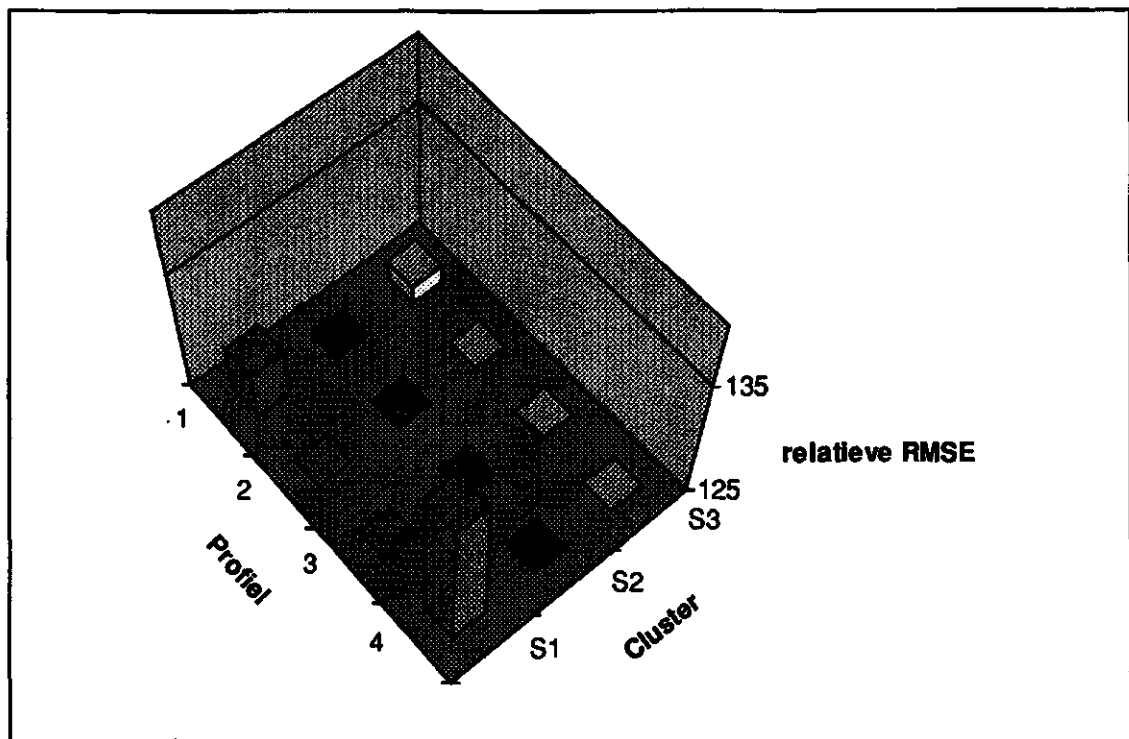


Fig. 8. Relatieve voorspelfouten groter dan 125%.

5.3.5 Bronnen van onzekerheid

Voor 4 variabelen, 3 weerjaren en 12 methoden van ruimtelijke en data-aggregatie is berekend wat de bijdrage aan de totale onzekerheid is van de variantiecomponent ten gevolge van de ruimtelijke aggregatie (*ra*) en die van de data-aggregatie (*da*). De ratio *ra/da* is dimensieloos en maakt vergelijking van de ($4 \cdot 3 \cdot 12 =$) 144 uitkomsten mogelijk. Figuur 9 geeft een frequentiediagram van de ratio *ra/da* over deze 144 uitkomsten. Opmerkelijk resultaat is, dat in 138 van de 144 uitkomsten de data-aggregatie de grootste bron van onzekerheid vormt. Dit betekent, dat verbeterde bodemschematisatie zich zou moeten richten op: (a) betere vertaalfuncties; (b) meer basisgegevens en (c) betere methoden van data-aggregatie. De gemiddelde ratio's van ruimtelijke aggregatie methoden 1, 2 en 3 zijn respectievelijk 0,45, 0,60 en 0,15. Dit betekent, dat de onzekerheid ten gevolge van data-aggregatie relatief het grootst is bij methode 3 en het kleinst bij methode 2. Dit is, naast de gunstigste relatieve voorspelfout, een extra argument voor keuze van aggregatiemethode 2

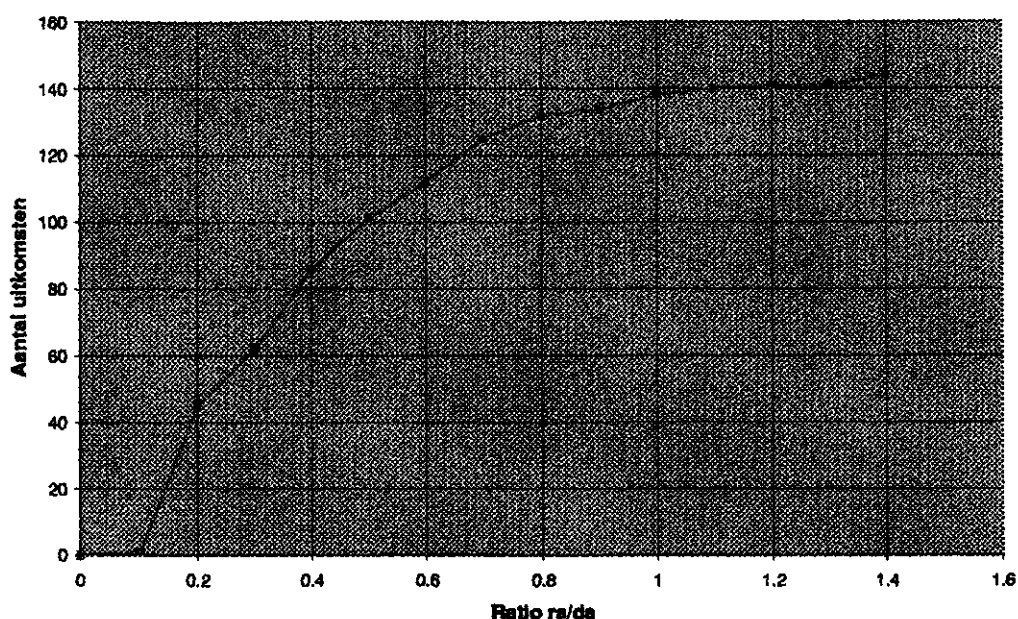


Fig. 9 Frequentiediagram van ratio ra/da

5.4 De beschikbare data

Er is voor het studiegebied Beerse Reuzel (50 000 ha) gebruik gemaakt van veel en diverse soorten profielbeschrijvingen en profielschetsen (zie paragraaf 4.2). Het aantal complete gemeten profielbeschrijvingen bedroeg slechts 78. Ongeveer 15% van de clusters werd uitsluitend beschreven met een landelijke profielschets (PSL). Van nog veel meer clusters werd voor specifieke bodemgegevens als Fe- en Al-oxalaatgehalten vrijwel uitsluitend gebruik gemaakt van die schetsen.

Zoals blijkt uit paragraaf 5.3 heeft het voor een verdere verbetering van de aggregatie meer zin energie te stoppen in de data. Data kunnen worden verzameld of middels kwalitatief goede vertaalfuncties worden verkregen.

Tijdens de modelberekeningen bleek het rekenmodel ANIMO versie 3.5 zeer gevoelig voor lage waarden voor de zuurstofdiffusieparameter 'a'. De modeluitkomsten N-opname en N-uitspoeling werden voor een zeer belangrijk deel daardoor bepaald. Deze parameter is berekend met een vertaalfunctie met een zeer lage verklaarde variantie (Wösten, 1997). Lage waarden voor de zuurstofdiffusieparameter 'a' worden vooral berekend wanneer bij leemarme of zwaklemige zandgronden het organischestofgehalte groter dan 4 is.

Op landelijke schaal zijn er voor een groot aantal grondsoorten als veen en klei weinig data beschikbaar. Met name specifieke bodemkenmerken als Fe- en Al-oxalaatgehalten blijken dan schaars te zijn.

6 Conclusies en aanbevelingen

Er is met de ontwikkeling van de methode voor bodemschematisatie een instrument gereed gekomen dat landsdekkend op grotendeels geautomatiseerde wijze beschikbare ruimtelijke bodembestanden en -data aggregereert voor modelstudies naar nitraat- en fosfaattransport in de bodem.

Het basale (onvermijdelijke) niveau van bodemschematisatie op basis van kaarteenheden leidt tot een voorspelfout ten opzichte van de gemiddelde simulatieuitkomst met profielen op gemeten punten van minimaal 45%. De ontwikkelde methoden voor bodemschematisatie leiden altijd tot een toename van deze fout.

Als gevolg van bodemschematisatie neemt de voorspelfout afhankelijk van de gekozen methode en modeluitkomst toe met 0 tot 80%. De relatieve toename van de voorspelfout ten opzichte van het basale (onvermijdelijke) niveau komt gemiddeld voor alle modeluitkomsten alleen bij clustermethode 2 nooit boven 25%. Dit betekent dat het algoritme voor ruimtelijke aggregatie leidend tot 23 clusters in het studiegebied het meest representatieve beeld van de variabiliteit tussen kaarteenheden oplevert. Aangezien dit tevens de methode is die de grootste ruimtelijke aggregatie (90% reductie van het aantal ruimtelijke eenheden) veroorzaakt, is hiermee ook de meest effectieve methode gekozen.

De data-aggregatiemethoden 2 en 3 leveren in het algemeen een lagere toename van de gemiddelde voorspelfout voor alle modeluitkomsten dan methoden 2 en 4. Dit betekent dat methoden waarbij vanwege dataschaarste als alternatief veel gebruik wordt gemaakt van landelijke profielschetsen of waarbij alle profieldata gemiddeld worden leiden tot relatief sterke toename van de voorspelfout.

De bijdrage aan de totale variantie van de voorspelfout door ruimtelijke aggregatie van kaarteenheden naar cluster blijkt in 96% van de berekeningen kleiner dan die van data-aggregatie bij de constructie van het rekenprofiel. Hieruit kan worden geconcludeerd dat de oorzaak van het geleden informatieverlies ligt bij het keizen van representatieve data en het dus meer nut lijkt te hebben gegevens te verzamelen, data-aggregatie algoritmen bij te stellen en/of de vertaalfuncties te verbeteren.

Deze conclusie geldt met name wanneer men uitgaat van de bodemkaart als het basale (onvermijdelijke) niveau voor ruimtelijke aggregatie. Vanwege de minimale voorspelfout uitgaand van de bodemkaart van 45% van de simulatieuitkomst lijkt het verstandig ook de ontwikkeling van aggregatietechnieken vanaf puntniveau nader te bezien.

Het verdient aanbeveling bij de verdere verfijning van aggregatietechnieken voor de gebruikte modellen aandacht te houden voor de beschikbaarheid voor voldoende chemische data en verbetering van de kwaliteit van de vertaalfuncties voor zuurstofdiffusie.

Literatuur

- Breeuwsma, A., J. H.M. Wösten, J.J. Vleeshouwer, A.M. van Slobbe en J. Bouma, 1986. Derivation of land qualities to assess environmental problems from soil survey. *Soil Sci. Soc. of Am. Journal* 50: 186-190.
- Bregt, A.K., en J.D. Bulens, 1995. Generaliseren in de praktijk; drie methoden voor het generaliseren van vlak-objecten, NGT GEODESIA, 1995-12.
- Feddes, R.A., Kabat, P., P.T.J. van Bakel, J.B.B. Bronswijk and J. Halbertsma, 1988. Modeling soil water dynamics in the unsaturated zone; state of the art. *J. of Hydrol.* 100 (1/3): 69-111.
- Finke, P.A., W.J.M. de Groot, M.J.D. Hack- ten Broeke, Y van Randen, J.H. Oude Voshaar and F. de Vries, 1996. Linking digital soil maps and databases to simulation models: functional soil map aggregation in the Netherlands. In: Heineke et al. (eds.) *Land-Information systems. Developments for planning the sustainable use of land resources.* European Soil Bureau Research Report 4, EUR 17729 EN, (1998), 546 pp. Office for Official Publications of the European Communities, Luxembourg.
- Hendriks, C.M.A., M.J.D. Hack-ten Broeke en G.A. van Soesbergen, in voorb. Een kennissysteem voor landevaluatie. DLO-Staring Centrum. Rapport 395.
- Kroes, J., and J. Roelsma, 1998. ANIMO3.5, User's guide for the ANIMO version 3.5 nutrient leaching model. SC-DLO, Wageningen. Technical document 46.
- Massop H.Th.L., en P.A.J.W. de Wit, 1994. Hydrologisch onderzoek naar de drainageweerstand van het tertiair ontwateringsstelsel in Oost-Gelderland. SC-DLO, Wageningen. Rapport 373.
- Rossiter, D.G., 1990. ALES, a framework for land evaluation using a microcomputer. *Soil Use and Management* 6: 7-20.
- Ruitenbergh, G.H., F.A. Wopereis en O. Oenema, 1991. Berekenende optimale stikstofbemesting voor grasland als functie van grondsoort. DLO-Staring Centrum. Rapport 173, Wageningen
- Soesbergen, G.A. van, C. van Wallenburg, K.R. van Lynden en H.A.J. van Lanen, 1986. Interpretatie van bodemkundige gegevens voor de akker- en weidebouw. Wageningen, Stiboka. Rapport 1967.
- Visschers, R., 1993. Upgrading van de bodemkaart van Nederland, schaal 1 : 50 000 door steekproeven in kaarteenheden van veldpodzolgronden. Wageningen, DLO-Staring Centrum. Rapport 183.

Vries, F. de, 1993. Een fysisch-chemische karakterisering van de eenheden van de Bodemkaart van Nederland, schaal 1 : 250 000. SC-DLO. Wageningen. Rapport 265.

Vries, F. de, 1994. Een fysisch-chemische karakterisering van de bodemeenheden van de Bodemkaart van Nederland, schaal 1 : 50 000, met onderscheid naar landgebruik. SC-DLO. Wageningen. Rapport 286.

Werkgroep Herziening Cultuurtechnisch vademecum, 1988. Cultuurtechnische vereniging. Utrecht. Wösten, J.H.M., 1997. Bodemkundige vertaalfuncties bij SC-DLO, State of the art, SC-DLO, Wageningen. Rapport 563.

Wösten, J.H.M., G.J. Veerman en J. Stolte, 1994. Waterretentie- en doorlatendheidskarakteristieken van boven- en ondergronden in Nederland: de Staringreeks. vernieuwde uitgave 1994. Technisch Document 18. DLO-Staring Centrum, Wageningen.

Wösten, J.H.M., P.A. Finke and M.J.W. Jansen, 1995. Comparison of class- and continuous pedotransfer functions to generate soil hydraulic characteristics. *Geoderma* 66: 227-237.

Niet gepubliceerde bronnen

Broek, B.J. van den, J.A. Elbers, J. Huygen, P. Kabat, J.G. Wesseling, J.C. van Dam and R.A. Feddes, 1994. SWAP 1993 Input instructions manual, DLO-Staring Centrum. Wageningen. Interne mededeling, p. 291.

Finke, P.A. 1996a. Beschrijving PTF. DLO-Staring Centrum, Wageningen, Interne Mededeling 425.

Finke, P.A. 1996b. Beschrijving CLUSPROF. DLO-Staring Centrum, Wageningen, Interne Mededeling 426.

Groot, W.J.M. de, 1997. Beschrijving SWAANIIN, een programma voor de opbouw van invoerbestanden voor de rekenmodellen SWAP93 en ANIMO 3.5 met bodemkundige parameters uit de bodemschematisatie. DLO-Staring Centrum, Wageningen, Interne Mededeling 480.

Groot, W.J.M. de., in voorb. Vertaalfunctie van grondwaterafvoerrelaties voor grondwatertrappen, SC-DLO, Wageningen

Aanhangsel 1 Te selecteren bodemgegevens voor de basis-rekenprofielen uit BIS

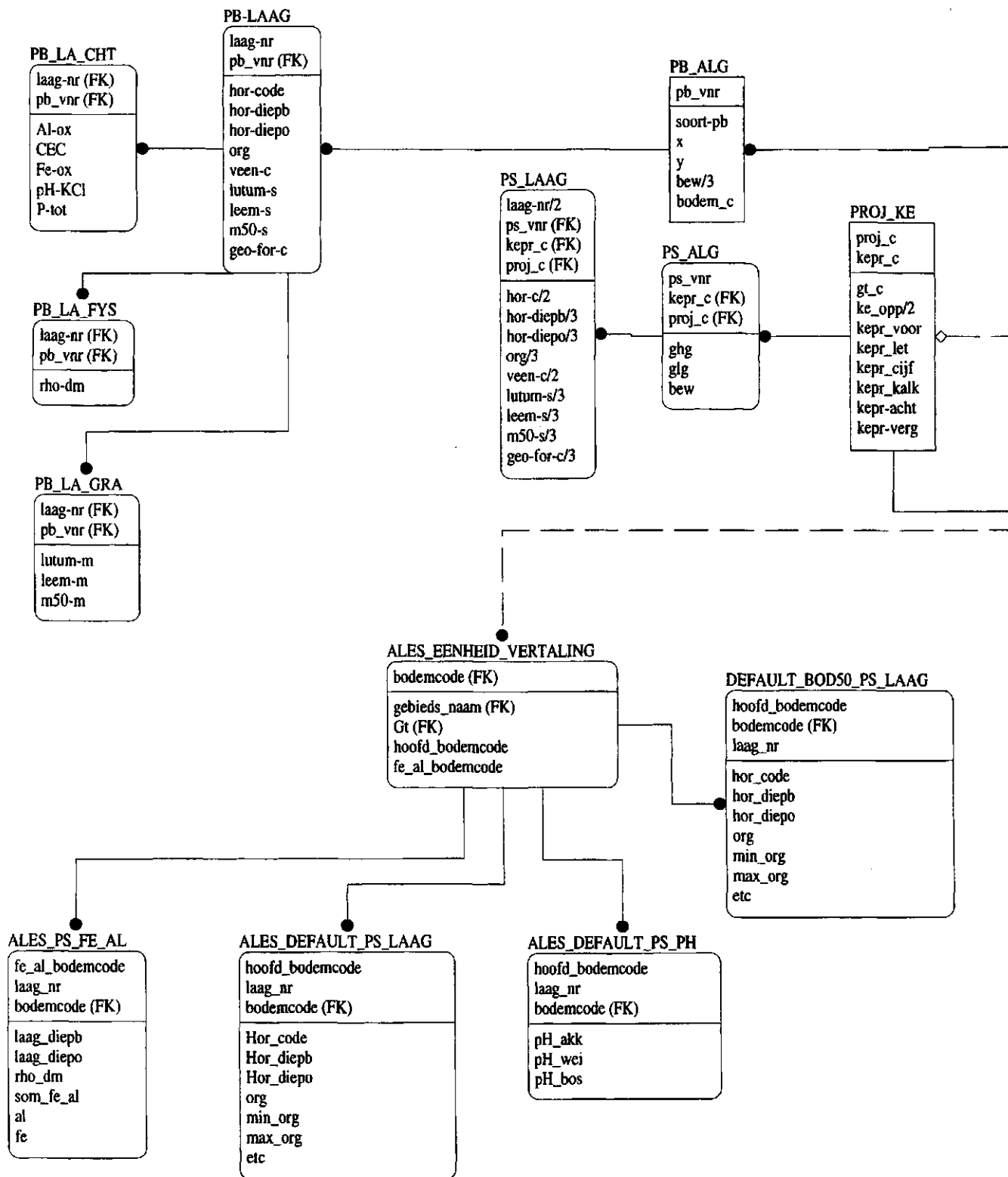
Variabele	Eenheid	Betekenis
<i>Algemene profielgegevens (elk gegeven op een aparte regel)</i>		
BODEMCODE	-	Code bodemeenheid
Gt	-	Grondwatertrap
BODGEBR	-	Landgebruikscode
X	m	X-coördinaat
Y	m	Y-coördinaat
BEW	cm	Bewortelbare diepte
<i>Laaggegevens 1 (per laag 1 regel)</i>		
HOR_CODE	-	Horizont- of laagcode
HOR_DIEPB	cm	Bovengrens horizont of laag
HOR_DIEPO	cm	Ondergrens horizont of laag
RHO_DM	kg/dm ³	Bulkdichtheid
PH_KCL	-	PH-KCl
PH_H2O	-	PH-H ₂ O
ORG	%	Organischestofgehalte (schatting)
HUMUS	%	Organischestofgehalte (meting)
LUTUM_S	%	Lutumgehalte vaste delen (schatting)
LEEM_S	%	Leemgehalte vaste delen (schatting)
M50_S	µm	Mediaan zandfractie (schatting)
LUTUM_M	%	Lutumgehalte vaste delen (meting)
LEEM_M	%	Leemgehalte vaste delen (meting)
M50_M	µm	Mediaan zandfractie (meting)
GEO_FOR_C	-	Geologische formatie
PTOT	mg P ₂ O ₅ /100g	Totaal Fosfaatgehalte
FEOX	mmol/kg	Oxalaat extraheerbaar IJzer
ALOX	mmol/kg	Oxalaat extraheerbaar Aluminium
CEC	meq/100g	Kationenomwisselcapaciteit

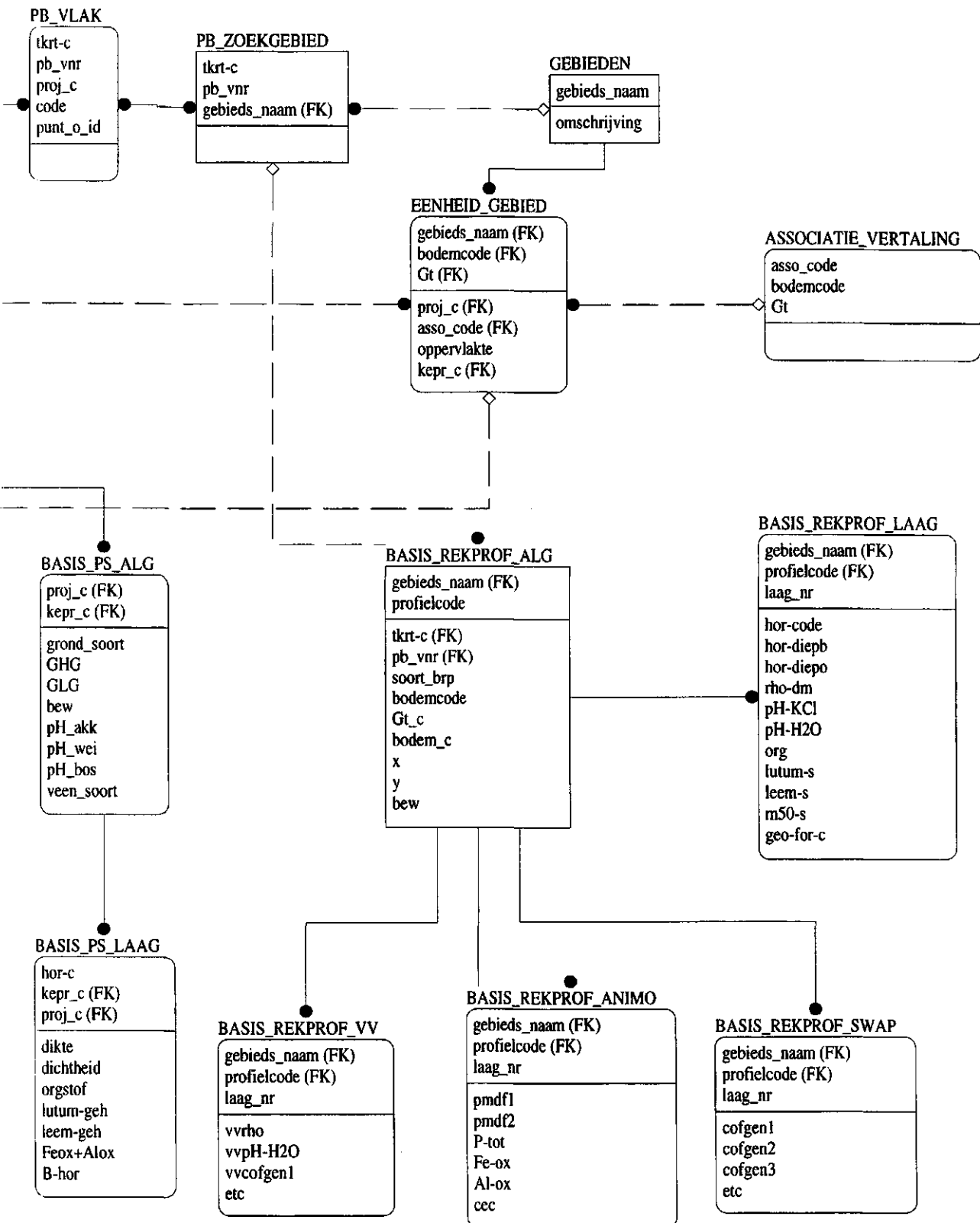
Aanhangsel 2 Format basis-rekenprofiel

Naam variabele	Eenheid	Fortran Format	Betekenis
<i>Algemene profielgegevens (elk gegeven op een aparte regel)</i>			
PROFIELCODE	-	A30	
KAARTEENH	-	A20	Code bodemeenheid
GT	-	A4	Grondwatertrap
BODGEBR	-	A4	Landgebruikscade
X	m	I7	X-coördinaat
Y	m	I7	Y-coördinaat
AANTALPS	-	I4	Aantal profielschetsen uit BIS
AANTALPB	-	I4	Aantal profielbeschrijvingen uit BIS
AANTALSPB	-	I4	Aantal profielbeschrijvingen uit LSK
KWLABEL	-	A4	Kwaliteitslabel clusterprofiel
CLUSTERMETH	-	I4	Code van toegepaste methode van clustering van kaarteenheden
CLUSTERNR	-	I4	Nummer van cluster waarbij PROFIELCODE is ingedeeld
PROFIELMETH	-	I4	Code van toegepaste methode van middeling van profieleeigenschappen
CLUSTBODGEBR	-	A4	Bodemgebruik binnen de cluster
COFQHA	cm/d	F8.3	A-parameter Q(h)-relatie
COFQHB	-	F9.4	B-parameter Q(h)-relatie
BEW	cm	I4	Bewortelbare diepte
NUMLAY	-	I4	Aantal unieke lagen
NUMNOD	-	I4	Aantal rekenlagen
<i>Laaggegevens 1 (per laag 1 regel)</i>			
HOR_CODE	-	A8,1X	Horizont- of laagcode
HOR_DIEPB	cm	I4,1X	Bovengrens horizont of laag
HOR_DIEPO	cm	I4,1X	Ondergrens horizont of laag
RHO_DM	kg/dm ³	F8.3,1X	Bulkdichtheid
PH_KCL	-	F6.1,1X	pH-KCl
PH_H2O	-	F6.1,1X	pH-H ₂ O
ORG	%	F6.1,1X	Organischestofgehalte
LUTUM_S	%	F6.1,1X	Lutumgehalte vaste delen
LEEM_S	%	F6.1,1X	Leemgehalte vaste delen
M50_S	mm	F6.1,1X	Mediaan zandfractie
GEO_FOR_C	-	I4,1X	Geologische formatie
COFGEN1	cm ³ /cm ³	F8.3,1X	Residuaire vochtgehalte
COFGEN2	cm ³ /cm ³	F8.3,1X	Verzadigd vochtgehalte
COFGEN3	cm/d	F8.2,1X	Verzadigde doorlatendheid
COFGEN4	1/cm	F9.4,1X	α-parameter
COFGEN5	-	F8.3,1X	L-parameter
COFGEN6	-	F8.3,1X	n parameter
PMDF1	-	F9.4,1X	Zuurstofdiffusie parameter A
PMDF2	-	F9.4,1X	Zuurstofdiffusie parameter B
PTOT	mg P ₂ O ₅ /100g	F8.3,1X	Totaal Fosfaatgehalte
FEOX	mmol/kg	F8.3,1X	Oxalaat extraheerbaar Ijzer
ALOX	mmol/kg	F8.3,1X	Oxalaat extraheerbaar Aluminium
CEC	meq/100g	F8.3	Kationenomwisselcapaciteit
VVPHH2O	%	F4.1,1X	Percentage verklaarde variantie bulkdichtheid
VVRHO	%	F4.1,1X	Percentage verklaarde variantie pH-H ₂ O
VVCOFGEN1	%	F4.1,1X	Percentage verklaarde variantie residuaire vochtgehalte

Naam variabele	Eenheid	Fortran Format	Betekenis
VVCOFGEN2	%	F4.1,1X	Percentage verklaarde variantie verzadigd vochtgehalte
VVCOFGEN3	%	F4.1,1X	Percentage verklaarde variantie verzadigde doorlatendheid
VVCOFGEN4	%	F4.1,1X	Percentage verklaarde variantie α parameter
VVCOFGEN5	%	F4.1,1X	Percentage verklaarde variantie L parameter
VVCOFGEN6	%	F4.1,1X	Percentage verklaarde variantie n parameter
VVPMDF1	%	F4.1,1X	Percentage verklaarde variantie zuurstofdiffusieparameter A
VVPMDF2	%	F4.1,1X	Percentage verklaarde variantie zuurstofdiffusieparameter B
VVCEC	%	F4.1,1X	Percentage verklaarde variantie kationomwisselcapaciteit

Aanhangsel 3 Datamodel voor de selectie van bodemkundige gegevens





Aanhangsel 4 Lijst met bodemkundige gegevens voor beschrijving van kaartenheden

Gegeven	Toelichting
Code kaartenheid	Legenda bodemkaart schaal 1 : 50 000
Grondsoort	kl=klei, zl=zavel, zd=zand, mr=moerig, lm=leem
GHG	cm
GLG	cm
bewortelbare diepte weidebouw	max. 40 cm
bewortelbare diepte akkerbouw	waarde uit profielschets (cm).
bewortelbare diepte bosbouw	Als bewortelbare diepte akkerbouw <50 dan +10 cm: max. GLG-10; als bewortelbare diepte akkerbouw tussen 50 en 100 dan + 30 cm: max. GLG-20; als bewortelbare diepte akkerbouw >100 dan +20 cm: max. GLG -20 (cm).
z-afstand weidebouw	bepaald met programma VERPF (cm)
z-afstand akkerbouw	idem
z-afstand bosbouw	idem
lutumgehalte bewortelbare zone	gewogen gemiddelde van horizonten
leemgehalte bewortelbare zone	idem
organischestofgehalte bewortelbare zone	idem
lutumgehalte ondergrond	idem
leemgehalte ondergrond	idem
organischestofgehalte ondergrond	idem
pH weidebouw	
pH akkerbouw	
pH bosbouw	
veensoort	eutroof=1 oligo-/mesotroof=2
Voor 4 horizonten	
Dikte (cm)	
dichtheid	
som van Fe-ox en Al-ox	
organischestofgehalte	
lutumgehalte	
B-horizont	0=nee, 1=ja