

Kwaliteit van geo-informatie

L. Heres (redactie)

Nederlandse Commissie voor Geodesie

Delft, juli 1998

Colofon

Kwaliteit van geo-informatie

L. Heres (redactie)

ISBN 90 6132 263 4

Uitgegeven door: Nederlandse Commissie voor Geodesie, Delft

Vormgeving en productie: Bureau Nederlandse Commissie voor Geodesie, Delft

Druk en bindwerk: Meinema Drukkerij, Delft

Omslag: Grondgebruiksgrenzen en foutief geclassificeerde pixels, H.A.M. Thunnissen

Bureau van de Nederlandse Commissie voor Geodesie

Bezoekadres: Thijsseweg 11, 2629 JA Delft

Postadres: Postbus 5030, 2600 GA Delft

Tel.: 015-278 28 19

Fax: 015-278 17 75

E-mail: ncg@geo.tudelft.nl

WWW: www.knaw.nl

De Nederlandse Commissie voor Geodesie (NCG) is een instituut van de Koninklijke Nederlandse Akademie van Wetenschappen (KNAW).

Inhoudsopgave

L. Heres
Voorwoord v

L. Heres
De kwaliteit van geo-informatie 1

H.J.G.L. Aalders
Standaardisatie van de geo-kwaliteit 4

F.J. Ormeling
Visualisatie van kwaliteitsinformatie 14

W.A. de Rooij
Kwaliteit in de CBS-praktijk 25

H.A.M. Thunnissen
Bepaling classificatienauwkeurigheid en kwaliteitsbeheersing van het Landelijk
Grondgebruiksbestand van Nederland 29

J. IJsselstein
Kwaliteit(en) kadastrale kaart 39

B. Kolk
TOP10kwaliteit (kwaliteitsaspecten van TOP10vector) 50

N. Schmorak
Een methodiek voor kwaliteitscontrole van vaarweggegevens 60

Voorwoord

L. Heres

Het gaat beslist te ver om te zeggen dat 'kwaliteit' een geheel nieuw onderwerp is in de geo-informatica. Al sinds mensenheugenis zijn geodeten immers gewoon om de kwaliteit van hun metingen te bewaken en te evalueren. Al die jaren hebben ze ook continu gewerkt aan een bijbehorend theoretisch model, wat inmiddels een imposant bouwwerk is geworden.

Het gaat daarbij echter louter om de *geometrische* informatie, i.e.: informatie over de positie, oriëntatie en vorm van objecten. Hoe belangrijk ook, geometrische informatie is slechts een beperkt onderdeel van de geografische informatie in haar totaliteit. Geografische informatie is immers ook: objectclassificatie, -identificatie en -kenmerken. Ook daarover is in het verleden met een zekere regelmaat gezegd dat die evenveel aandacht zou dienen te krijgen als de geometrische aspecten. Tot nu toe heeft dit echter nog niet geleid tot een wijd en zijd verbreide en toegepaste methodologie. Aan de andere kant heeft men de afgelopen ook niet stilgezeten en is er beslist het één tot stand gekomen, zowel op theoretisch als praktisch gebied.

Daarom leek het de Subcommissie Geo-Informatie Modellen van de Nederlandse Commissie van Geodesie een goed idee een studiedag te organiseren met het doel een momentopname van de huidige stand van zaken te geven. De centrale vragen daarbij waren:

- Wat zijn de internationale ontwikkelingen op het terrein van de theorievorming?
- Informatie over kwaliteit, hoe breng je die in beeld?
- Hoe gaan de belangrijkste Nederlandse leveranciers van geo-informatie momenteel met kwaliteit om?

De voordrachten van de studiedag zijn in deze publicatie gebundeld.

Aalders, hoogleraar aan de Katholieke Universiteit Leuven en hoofddocent aan de TU Delft, geeft een overzicht van de Europese en mondiale activiteiten op het gebied van de theorievorming rond de kwaliteit van geografische informatie. Hij belicht tevens de standaards die er op dit gebied de laatste jaren in CEN- en ISO-verband tot stand zijn gekomen.

De kwaliteit van geografische informatie hoeft niet homogeen te zijn, en kan ook zelf weer een ruimtelijke spreiding vertonen. Ormeling, hoogleraar kartografie aan de Universiteit Utrecht, gaat in op de methoden om deze ruimtelijke spreiding in kaart te brengen en te visualiseren.

De Rooij, wetenschappelijk medewerker bij het Centraal Bureau voor de Statistiek zet uiteen welke methoden het CBS gebruikt om de gegevenskwaliteit te bewaken. Hij signaleert daarbij een aantal knelpunten en schetst hoe daar in de toekomst.

Thunnissen, wetenschappelijk medewerker van het DLO-Staring Centrum, belicht de kwaliteitsbeheersing van het Landelijk Grondgebruiksbestand van Nederland (LGN). Hij gaat in het bijzonder in op de classificatienauwkeurigheid.

Het Kadaster heeft een lange traditie in het bewaken van de kwaliteit van haar producten. Met het digitaal beschikbaar komen van het kadastrale materiaal, treden er echter een aantal nieuwe aspecten aan het licht. IJsselstein van het Kadaster in Groningen, gaat hier dieper op in.

Ook de Topografische Dienst heeft een soortgelijke lange traditie en ook hier wordt men in het digitale tijdperk met een aantal nieuwe kwaliteitsaspecten geconfronteerd. Kolk, hoofd Topografie van de Topografische Dienst, zet uiteen welke.

Schmorak, wetenschappelijk medewerkerster van de Adviesdienst Verkeer en Vervoer, doet uit de doeken hoe men bij deze dienst het AQL (Acceptance Quality Level) inzet om ingekochte gegevens voor het Vaarwegenbestand te controleren.

De studiedag is gehouden op 14 november 1997 bij de Faculteit Ruimtelijke Wetenschappen, Universiteit Utrecht. De organisatoren bedanken de Universiteit Utrecht voor het beschikbaar stellen van de faciliteiten en de sprekers voor hun lezingen.

De kwaliteit van geo-informatie

L. Heres

Adviesdienst Verkeer en Vervoer
Rijkswaterstaat
E-mail: l.heres@avv.rws.minvenw.nl

Inleiding

Enkele jaren geleden deed bij Philips de volgende anekdote de ronde: Men had bij een Japanse toeleverancier 100.000 transistors besteld en daarbij een AQL¹ van 5 promille bedongen. Dit betekent dat er in een steekproef van 1000 stuks er maximaal 5 defecte exemplaren mogen zijn. Dit was in een begeleidende brief aan het Japanse bedrijf ook zo uitgelegd, omdat men wist dat de Japanners met deze wijze van kwaliteitsmeting nog niet erg vertrouwd waren.

Enkele weken later werd de partij afgeleverd met de volgende begeleidende brief: "Hierbij sturen wij de door u bestelde transistors; de 5 defecte exemplaren hebben we in een aparte enveloppe bijgevoegd".

Deze anekdote illustreert dat men verschillend over kwaliteit kan denken en dat het dus heel belangrijk is om naar een gemeenschappelijk *conceptueel model* toe te werken.

Wat is kwaliteit?

Van zo'n term kunnen uiteraard een heleboel definities gegeven worden. Er bestaan enkele heel ingewikkelde zoals die te vinden is in ISO 8402: *Geheel van kenmerken van een entiteit dat betrekking heeft op het vermogen van die entiteit om kenbaar gemaakte en vanzelfsprekende behoeften en te bevredigen.*

Voor de kwaliteit van geografische gegevens is deze definitie te ruim. Daarom zou ik de volgende operationele definitie willen hanteren: *De mate waarin iets aan kenbaar gemaakte behoeften voldoet.*

Ik gebruik hier met opzet het woord 'mate' omdat dat verwant is met *meten*. Dit geeft meteen aan dat als men een zinvolle uitspraak over gegevenskwaliteit wil doen, er eerst altijd iets gemeten moet worden. Een meting leidt meestal tot één of meerdere getallen, die als een maat van de kwaliteit beschouwd kunnen worden.

Een kwaliteitsmaat alleen is niet genoeg om een uitspraak te doen over de vraag of een product wel of niet voldoet. Er zijn daarnaast '*kenbaar gemaakte behoeften*' nodig, in de praktijk meestal 'specificaties' genoemd. Kwaliteitsmaat en specificaties samen kunnen leiden tot een zinvolle uitspraak over de kwaliteit van een product. Om naar het voorbeeld in de inleiding terug te keren: de specificatie is het AQL van 5, de steekproef is de

¹ AQL is een afkorting van Acceptance Quality Level, en is een speciale door de ISO genormaliseerde methode voor het beoordelen van kwaliteit van massaproducten door middel van steekproeven.

meting waaruit een kwaliteitsmaat uit voortkomt, zeg 7 defecte exemplaren. De uitspraak kan daarna luiden: afgekeurd. Dit illustreert ook dat kwaliteitsuitspraken altijd relatief zijn, namelijk gerelateerd aan de specificaties. Als het AQL geen 5 zou zijn geweest, maar 10, dan zou dezelfde partij zijn goedgekeurd.

Belangrijk is ook dat kwaliteit in principe een neutrale term is. De groenteboer die het heeft over '*kwaliteitje meneer*', bedoelt uiteraard goede kwaliteit, maar laat het goede weg. Kwaliteit is niet per definitie goed. Er bestaat ook slechte kwaliteit.

Een duidelijke relatie bestaat er tussen kwaliteit en prijs: met een wasmachine die 50 jaar mee gaat, maar die f 10.000 kost, ben je duurder uit dan met een machine die 10 jaar meegaat voor f 1000.

Waarom is kwaliteit zo belangrijk geworden in de geo-informatie?

Hiervoor zijn een aantal redenen aan te voeren. Toename van de complexiteit is een eerste: in plaats van enkelvoudige gedrukte kaarten leveren we nu geïntegreerde bestanden die uit diverse bronbestanden afkomstig zijn. Inzicht in de kwaliteit van deze brondata is daarom belangrijk.

Productaansprakelijkheid is een tweede. In de financiële wereld lopen al verscheidene processen van mensen die zich gedupeerd voelen door verkeerde informatie. Bij geografische gegevens is dit nog niet zo aan de orde, maar dit kan komen. Er is in ieder geval al iemand geweest die een navigatiesysteem aansprakelijk heeft gesteld voor het veroorzaken van een ongeluk vanwege het verstrekken van verkeerde informatie.

Het bestaan van certificaten is een derde reden: voor veel bedrijven is dit een aanleiding om iets aan kwaliteit te doen, omdat men zich op deze manier van andere bedrijven kan onderscheiden.

De verkoop van data aan derden is de vierde maar beslist niet de onbelangrijkste reden. Bij verkoop van gegevens loopt men het risico op ontevreden klanten en dan is het toch erg handig om aan de hand van cijfers aan te kunnen tonen of een bepaalde kwaliteit wel of niet gehaald is.

Vele aspecten van kwaliteit

Het begrip kwaliteit kan van veel kanten belicht worden. Lang niet alle, maar toch een aantal belangrijke daarvan komen in deze publicatie ter sprake.

Europese standaardisatie

Er bestaat een Europese normcommissie, de CEN TC 287, die de taak heeft een aantal zaken op het gebied van geografische informatie te standaardiseren. Onder andere de kwaliteitsaspecten. Het artikel van Aalders van de TU Delft zal in de eerste inleiding uitgebreid in op beide onderdelen ingaan.

Visualisatie van kwaliteit

De kwaliteit van geografische gegevens is in feite ook een stukje informatie. In jargon heet dat *meta-informatie*: informatie over informatie. Nu kan kwaliteit op zich ook weer een geografische component hebben: bijv. in het geval dat deze niet homogeen gespreid is. De behoefte kan dan ontstaan om deze geografische spreiding in kwaliteit te visualiseren. Metakartografie zou ik dit vakgebied willen noemen: het in kaart brengen van geografische metagegevens. Het artikel Ormeling gaat verder op dit onderwerp in.

Kwaliteitsbeheersing

De andere artikelen gaan alle over kwaliteitsbeheersing. Reden om daar wat dieper op in te gaan.

Metten is weten

Dit is het centrale credo van de kwaliteitsbeheersing: zonder meten geen informatie, en zonder informatie geen beheersing. Er zijn in principe twee hoofdmethodes om de kwaliteit te beheersen: *procesbeheersing* en *productbeheersing*; zij hoeven elkaar niet uit te sluiten en kunnen elkaar versterken.

Product versus proces

Procesbeheersing is gebaseerd op het idee dat men het hele productieproces opsplitst in zo klein mogelijke stappen. Bij iedere stap doet men een kwaliteitsmeting en op basis daarvan stuurt men het proces bij. De kwaliteit van het uiteindelijke product is in feite onbekend, maar doordat men het proces beheerst kan men garanderen dat de kwaliteit constant is. Iets dat voor veel afnemers voldoende is om te weten. Deze methode leent zich vooral voor productieprocessen,

- die overduidelijk in een aantal simpele fasen te verdelen zijn,
- die in de tijd niet veel veranderen;
- met een complex eindproduct, waaraan moeilijk te meten is.

Bij *productbeheersing* beschouwd men het proces als black box en meet men alleen de kwaliteit van het eindproduct. Deze methode leent zich vooral voor moeilijke te ontleden of sterk veranderlijke productieprocessen die echter resulteren in tamelijke eenvoudige en onveranderlijke producten.

De bijdragen van Thunnissen, IJsselstein en Schmorak zijn productgericht. De bijdragen van De Rooij en Kolk zijn meer procesgericht. Alle bijdragen geven een indruk hoe er in de praktijk met de kwaliteit van databestanden wordt omgegaan. De bijdrage van Schmorak is bovendien een goed voorbeeld van een toepassing van de AQL-methode (ISO 2859).

Standaardisatie van de geo-kwaliteit

H.J.G.L. Aalders

Subfaculteit Geodesie
Faculteit Civiele techniek en Geo-wetenschappen
Technische Universiteit Delft
Faculteit Toegepaste Wetenschappen
Katholieke Universiteit Leuven
E-mail: h.aalders@geo.tudelft.nl

Samenvatting

In [1] and [5] is een basis gelegd voor een kwaliteitsmodel dat in de praktijk kan worden toegepast, waarbij niet alleen de kwaliteit van de geometrische gegevens is beschouwd, maar ook van de niet-geometrische gegevens. De hiervoor ontwikkelde modellen passen dezelfde regels toe als in de informatie technologie worden gebruikt. Dit is een vereiste indien de kwaliteitsgegevens worden opgeslagen in bestanden zoals momenteel gebeurt in de informatie technologie (IT). In deze presentatie worden deze aspecten eerst behandeld alvorens in te gaan op de specifieke kwaliteitsaspecten in de geo-informatie

Introductie

Invoering van de snel veranderende informatietechnologie in toepassingsgebieden zoals kartografie fotogrammetrie en geodesie, technieken die veelvuldig worden gebruikt bij de invoering van ruimtelijke gegevens, veroorzaakt een overgang van de kaart een historisch met name in de laatste 300 jaar) gegroeid eindproduct naar een nieuw type eindproduct: een digitaal bestand die de representatie is van de object die verbonden zijn aan het aardoppervlak. In de digitale situatie is het mogelijk dat de gegevensverzamelaars de gegevens structuren, topologische en andere relaties en waardevolle thematische gegevens toevoegen, alsmede kwaliteitsgegevens.

Deze digitale bestanden worden vaak ook aan meerdere gebruikers ter beschikking gesteld, die zelf ook relaties en thematische gegevens toevoegen voor eigen gebruik. Daardoor wordt elk bestand een combinatie van de gegevens van verschillende gebruikers, waarop de oorspronkelijke gegevensverzamelaar geen enkele invloed meer kan uitoefenen. Ook de kwaliteit is dan van verschillende oorsprong en waarde voor elke nieuwe gebruiker. De volgende vragen ontstaan daarom direct: "Hoe wordt de kwaliteit door de gegevensverzamelaar gedefinieerd, zodat het voldoet aan de eisen van de gebruiker?" en "Welke criteria stellen de gebruikers bij de beoordeling van de gegevens voor een bepaald gebruik; is dat kwaliteit, beschikbaarheid of het gemak van de toegankelijkheid?" Daarnaast is het van belang voor een weloverwogen gebruikt de visualisatie van de kwaliteit van de dataset noodzakelijk.

"When taking the concept of Quality away from anything in life, the world would function but life would be so boring, it would not be worth living: life would continue without any value or goal"

"Zen and the art of Motorcycle Maintenance"
Robert M. Pirsig

Concluderend kan men stellen dat:

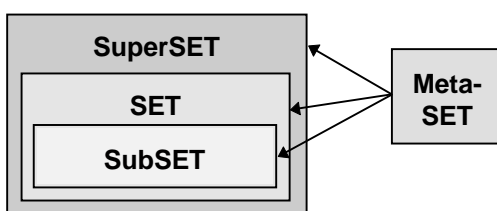
- kwaliteit van ruimtelijke gegevens kan worden omschreven met *de combinatie van eigenschappen en karakteristieken van een ruimtelijke dataset die voldoet aan de impliciete en expliciete, geformuleerde en bedoelde eisen van de gebruiker voor toepassing met een speciaal doel.*
Kwaliteitsparameters worden gedefinieerd voor het verschil tussen de werkelijkheid en de nominale grondslag (semantische kwaliteit) en tussen de nominale grondslag en het bestand (meetprecisie, idealisatie en generalisatie). De nominale grondslag vormt de basis voor de kwaliteitsdefinitie van de gegevens in het bestand;
- vastgesteld moet worden welke kwaliteitsparameters theoretisch gebruikt kunnen worden;
- de gebruiker en de leverancier moeten weten welke van die kwaliteitsgegevens nodig zijn om voldoende en afdoende informatie over de kwaliteit te kunnen leveren en
- welke methoden kunnen worden toegepast om die gegevens te visualiseren.

In dit artikel worden de kwaliteitskenmerken behandeld voor praktische toepassing, geplaatst in een theoretische achtergrond. Daarbij worden niet alleen de kwaliteit van de geometrische gegevens, maar ook van de niet-geometrische gegevens behandeld.

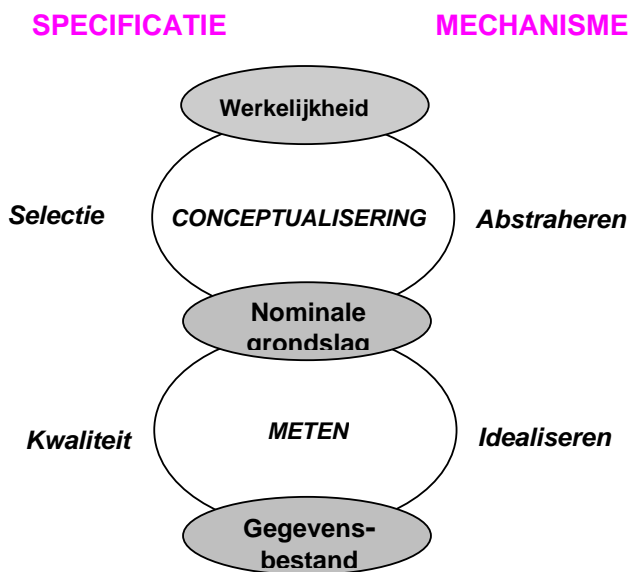
Datasets

Kwaliteit kan worden gedefinieerd voor zowel een - statische - verzameling gegevens (het bestand) - ook wel een set of dataset genoemd -, als voor de processen die op die verzamelingen gegevens worden toegepast.

Wanneer we de dataset nader beschouwen zoals in de informatie technologie wordt gebruikt, dan valt op dat de bijbehorende set kwaliteitsgegevens geen (onderdeel zijn van de dataset, maar dat zij een onderdeel vormen van een aparte dataset: de *METASET*, waarmee wordt aangeduid dat het gegevens betreffen over gegevens. De kwaliteitsgegevens uit de metadataset beschrijven een deel van de dataset (een subset) die bestaat uit één of meerdere exemplaren van een attribuut, entiteit of relatie, een gehele dataset, of meerdere datasets tegelijk. Daarmee ontstaat een hiërarchie voor de datasets zoals aangegeven in figuur 1. Deze datasets worden gebruikt voor het vastleggen van gegevens in de metadataset, inclusief de daartoe behorende kwaliteitsgegevens.



Figuur 1. Het gebruik van een metaset in een gegevensbestand.



Figuur 2. Het gebruik van de nominale grondslag bij de opbouw van geografische gegevensbestanden.

IT in GIS

Geografische gegevens worden verkregen door objecten in de werkelijkheid te selecteren en abstraheren volgens vooraf gestelde regels die de objecten specificeren en hoe ze moeten worden voorgesteld in het bestand. Deze subjectieve interpretatie van de werkelijkheid creëert een geïdealiseerde voorstelling, die de *nominale grondslag* wordt genoemd (Eng.: 'universe' of 'discourse', Am.: 'abstract view of the universe'). Deze interpretatie is subjectief om twee redenen:

- beschrijving van het objecttype. Hier wordt verondersteld dat de werkelijkheid uit een verzameling onderling gerelateerde objecten. De nominale grondslag beschrijft de abstractie van die objecten met hun eigenschappen en hun onderlinge relaties;
- toepassingsafhankelijkheid door de specificatie voor de nominale grondslag worden alleen die objecten in het bestand opgenomen die voor de betreffende toepassing van belang zijn.

Attributen

Door de aard van de geografische objecttypen komen de volgende attributen voor:

- *geometrische*, bestaande uit:
 - *positie* van het object gedefinieerd in een geodetisch referentie systeem;
 - *vorm* van het object, gedefinieerd door mathematische interpolatie functies;
 - *topologie* van de objecten, die de onderlinge nabijheidsrelaties van de objecten weergeeft.
 - Geometrie kan worden uitgedrukt in een raster of vector representatie, waarbij alleen de expliciete definitie van topologie verschilt en de impliciete definitie van de onderline grenzen. De aard van de geometrie is kwantitatief.
- *semantische*, die
 - *kwantitatief* van aard zijn, uitgedrukt in een bepaalde eenheid, zoals grootte, breedte, temperatuur, tijd, enz.;
 - *kwalitatief* van aard zijn, en refereren naar een vooraf gedefinieerde set waarden, zoals (klasse-)namen, adressen, kleuren, enz. Een kwalitatief attribuut kan *nominaal* zijn (waarmee wordt aangeduid dat er geen rangorde is in de waarden)

(bijvoorbeeld de klassenamen), of *ordinaal*, waarmee de rangorde in de attributen wordt aangeduid, bijvoorbeeld adressen en wegnummers.

Naast de semantiek van deze attributen heeft elk attribuut ook een *waarde*, waarvan het *domein* vooraf is vastgesteld. Het voorkomen van een objecttype in een bestand wordt een *exemplaar* genoemd en elk exemplaar heeft een unieke *identificatie* in het bestand.

Kwaliteit wordt ook in de vorm van attributen toegevoegd aan de objecten, waarbij de kwaliteit in aparte metasets wordt opgeslagen, waarnaar meerdere exemplaren in het bestand kunnen verwijzen.

Metakwaliteit

De kwaliteit van de kwaliteitsdefinitie wordt metakwaliteit genoemd. M.a.w. voor iedere kwaliteitsparameter zou een kwaliteitslabel moeten worden vastgesteld en opgeslagen, die de betrouwbaarheid van de kwaliteitsuitspraak formuleert, bestaande uit:

- betrouwbaarheidsparameter die de kwaliteit van de kwaliteitsparameters definieert;
- beschrijving van de methoden en technieken die hebben geleid tot de vaststelling van de kwaliteit van de gegeven in het bestand;
- semantische kwaliteitsparameters.

Structurering van kwaliteit in GIS

In 1982 werd in de VS de National Committee on Digital Cartographic Data Standards (NCDCDS) opgericht. De rapporten van deze commissie [9,10] bevatten secties over de gegevenskwaliteit met vijf elementen: oorsprong, metrische nauwkeurigheid, thematische nauwkeurigheid, volledigheid en logische consistentie (zie tabel 1).

The Internationale Kartografische Associatie (ICA) heeft in een gepubliceerd onderzoek [6] daar twee elementen aan toegevoegd, nl.: semantische kwaliteit en temporele kwaliteit.

De CEN/TC 287 heeft voor de ontwikkeling van een standaard op het gebied van kwaliteit in GIS de volgende elementen gebruikt: oorsprong, gebruik, homogeniteit en kwaliteitsparameters waarvan de primaire dezelfde zijn als in het SDTS voorstel en de secundaire tekstuele betrouwbaarheid. Dit voorstel is in september 1997 besproken naar aanleiding van het commentaar van de leden van de CEN.

SDTS (1992)	ICA (1995)	CEN/TC287 (1997)	ISO/TC211 (1997)
oorsprong	oorsprong	oorsprong, gebruik	overzicht (doel, gebruik, ontstaan)
metrische nauwkeurigheid	metrische nauwkeurigheid	metrische nauwkeurigheid	metrische nauwkeurigheid
thematische nauwkeurigheid	thematische nauwkeurigheid	thematische nauwkeurigheid	thematische nauwkeurigheid
volledigheid	volledigheid	volledigheid	volledigheid
logische consistentie	logische consistentie	logische consistentie	logische consistentie
	semantische nauwkeurigheid	metakwaliteit	
	temporele nauwkeurigheid	temporele nauwkeurigheid	temporele nauwkeurigheid
		homogeniteit	
			beoordelingstests en resultaten

Tabel 1. Overzicht van kwaliteitselementen gebruikt voor verschillende standaards.

De ISO/TC 211 heeft in augustus 1997 een CD (Committee Draft) laten verschijnen, dat bedoeld is om commentaar te verzamelen op het voorstel. Het zal nog wel enige tijd duren voordat dit voorstel concrete vormen gaat aannemen.

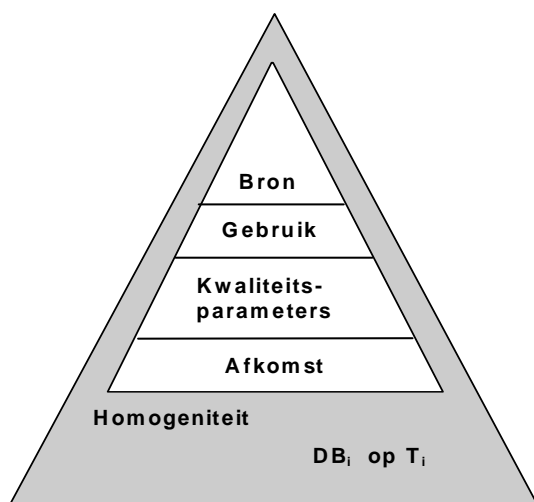
Om de kwaliteit van een gegevensbestand weer te geven wordt hier gebruik gemaakt van de indeling van de CEN, waarbij een soort hiërarchie wordt toegepast. (De ISO maakt gebruik van overzichtsgegevens en gegevens die betrekking hebben op specifieke datasets. Daarbij wordt de settheorie uit het oog verloren: kwaliteit is een metagegeven en heeft altijd betrekking op een dataset; dat geldt ook voor overzichtsgegevens als doel, gebruik of ontstaan).

De hiërarchie in het CEN-model ontstaat omdat de hoeveelheid kwaliteitsgegevens steeds verder uitbreiden naar mate men dieper in de hiërarchie komt terwijl de kwaliteitsgegevens betrekking hebben op steeds kleinere datasets. Vooraf dient echter te worden vermeld dat wordt verondersteld dat een gegevensbestand voortdurend aan wijzigingen onderhevig is. Dat wil zeggen de kwaliteitsgegevens betrekking hebben op de gegevens in een bestand zoals die op één bepaald moment voorkomen en dat tijdens dit moment geen veranderingen in het bestand plaatsvinden; een volgend moment kunnen de gegevens in het bestand zijn gewijzigd of andere kwaliteitsgegevens bevatten.

Kwaliteitsuitspraken worden ook in entiteiten opgeslagen en hebben betrekking op exemplaren van andere entiteiten, die op deze manier hun kwaliteitseigenschappen hebben. Dat veroorzaakt een nogal complexe logische structuur in een gegevensbestand.

De door de CEN gebruikte kwaliteitsindicatoren zijn:

- *bronvermelding* van het gegevensbestand. Dit geeft de naam van het bestand aan, het doel en de aanmaakdatum van het oorspronkelijke bestand en de verantwoordelijke organisatie. Natuurlijk kan deze informatie ook worden gezien als een onderdeel van de metadata van het gegevensbestand. Maar men moet zich realiseren dat alle kwaliteitsparameters metadata zijn. Bronvermelding kan ook worden gezien als de algemene informatie over de ontstaansgeschiedenis van het bestand (zoals het NCDCCDS rapport doet). Echter bronvermelding is het minimum aan gegevens dat voor een bestand aanwezig moet zijn. De attributen die behoren bij de bronvermelding beschrijven de juistheid van de gegevens in algemene termen; daarbij spelen een rol de tijdsperiode vanaf de laatste bijhouding, het oorspronkelijke doel van



Figuur 3. De kwaliteitspiramide.

- de gegevens, de indruk van de betrouwbaarheid van de leverancier en wie uiteindelijk verantwoordelijkheid draagt voor fouten in de gegevens;
- de oorspronkelijke *doelstelling* en het voorafgaand *gebruik* van de gegevens in een bestand door verschillende gebruikers voor verscheidene toepassingen kan een goede indicatie zijn voor de bruikbaarheid van het bestand voor de beoogde nieuwe toepassing en de betrouwbaarheid onder verschillende omstandigheden. Daarom zou elke organisatie die het bestand heeft gebruikt een kort relaas met bevindingen voor die dataset moeten weergeven, waarin onder andere wordt vermeld: de organisatie het soort gebruik de bruikbaarheid en ondervonden beperkingen die werden ondervonden bij het gebruik. Deze gegevens kunnen belangrijk zijn voor toekomstige gebruiker om te beoordelen of het bestand aan hun verwachtingen zou kunnen voldoen. Voor de leverancier verhoogt een gebruiksbeschrijving de potentiële toepassingen en daarmee het aantal potentiële klanten. Er kleeft echter ook een bezwaar aan de attributen bij gebruik: hoe waarheidsgetrouw zijn de gegevens ingevuld?
 - *kwaliteitsparameters* zullen in een volgende paragraaf worden beschreven;
 - de *afkomst* (ontstaansgeschiedenis) beschrijven alle processen die de dataset heeft ondergaan sinds het oorspronkelijke ontstaan van het object in het bestand. In [6] wordt gesteld dat bronvermelding gewoonlijk allereerst in de kwaliteitsdefinitie wordt gegeven, omdat daarmee de beschrijving van alle andere kwaliteitscomponenten wordt bepaald (en ook andersom). Bovendien verschaft de bronvermelding waardevolle historische gegevens voor toekomstige generaties. Voor elk proces in de ontstaansgeschiedenis van een bestand wordt een beschrijving gegeven die onder andere de gegevens zelf en de daarop toegepaste methoden weergeven met mogelijke referenties naar andere documenten die nog meer details bevatten over het proces; voorts wordt eveneens vermeld wie, wanneer, welke operaties heeft uitgevoerd ¹.

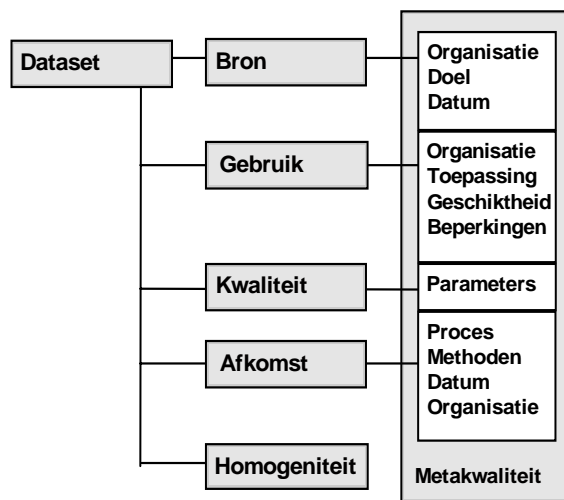
Parameterset voor GIS-kwaliteit

Een kwaliteitsparameter heeft een kwantitatieve eigenschap waarin de kwaliteit wordt uitgedrukt. Deze kwaliteitsgrootheden kunnen statistische grootheden zijn (uitgedrukt in getallen, met of zonder eenheid), maar ook een binaire of andere soort beschrijving. Aan deze kwaliteitsgrootheden moet ook de een tijdstip hangen om de validiteit ervan te kunnen vaststellen. De beschrijving van de algoritmen die zijn gebruikt om de kwaliteitsgrootheden vast te stellen kunnen worden opgeslagen als een onderdeel van de metakwaliteit.

De set kwaliteitsparameter bestaat uit verschillende parameters voor de drie verschillende aspecten van kwaliteit in GIS:

- *metrische nauwkeurigheid*. Volgens het NDCS rapport moet de metrische nauwkeurigheid de relatie naar het coördinatensysteem bevatten. Metrische nauwkeurigheid omvat zowel de planimetrische als de verticale aspecten.
- Bovendien moet het alle effecten van transformaties die op de gegevens zijn toegepast en ook de testresultaten van de bepaling van de metrische nauwkeurigheid voor de gegevens. Daarom kunnen we spreken van:

¹ Gedurende het bestaan van een geografische bestand kunnen er vele, complexe processen hebben plaatsgevonden, zoals invoer, transformaties, generalisaties, samenvoegingen, aggregaties, zonerings, conversies, bijhoudingen, validaties, consistentiecontroles, enz. Allen dragen ze bij tot de huidige kwaliteit van de gegevens in het bestand. Daarom is het noodzakelijk voor de gebruikers om over voldoende gegevens te beschikken om een juiste beoordeling te kunnen uitvoeren voor het beoogde nieuwe gebruik.



Figuur 4. De set kwaliteitsparameters met attributen.

- nauwkeurigheid in relatie tot de positie ten opzichte van het (horizontale en verticale) ruimtelijke referentiesysteem;
relatieve metrische nauwkeurigheid waarbij de nauwkeurigheid wordt bedoeld van objecten in het bestand ten opzichte van elkaar;
- resultaten van tests.

De mogelijke grootheden voor metrische nauwkeurigheid kunnen zijn:

- voor enkelvoudige attributen: RSME (Root Mean Square Error), range (minimum en maximum waarde), standaardafwijking, histogram met afwijkingen van de gemiddelde waarde, of betrouwbaarheid niveau of -interval;
- voor meervoudige attributen: nauwkeurigheid (voor ieder kwantitatief attribuut), standaardellips, correlatiematrix of -functie, range, eigenwaarde van een correlatiematrix.

In het ISO/TC211 kwaliteitsmodel wordt onderscheidt gemaakt tussen absolute (ten opzichte van het coördinaten referentie systeem) of relatieve nauwkeurigheid (ten opzichte van omliggende objecten). Bovendien onderscheidt men ook de planimetrische en hoogte nauwkeurigheid van de geometrische gegevens en wordt een parameter voor de vorm-betrouwbaarheid opgegeven (waarmee de afwijking van een kromme tussen twee verschillende afbeeldingen wordt bedoeld).

In [6] geeft Drummond enkele praktische voorbeelden van toepassingen door landmeters voor metrische nauwkeurigheid en de voortplanting ervan. Zij stelt dat de 'mogelijk maximale fout' een concept is dat veel gebruikers kunnen hanteren en dat het voorzien van homogene data(sub)sets van kwaliteitsattributen op een theoretische verantwoorde wijze voor hen vaak te moeilijk is.

- *thematische nauwkeurigheid*, refereert naar de nauwkeurigheid van kwantitatieve en kwalitatieve attributen van objecten of relaties in het gegevensbestand. Kwantitatieve attributen gedragen zich hetzelfde als de metrische nauwkeurigheid en kunnen daarom op dezelfde manier worden beschreven. Voor kwalitatieve attributen geldt de correcte of foutieve classificatie, spelfouten (refererend naar de tekstuele betrouwbaarheid), of een juiste toekenning van de attribuut waarde. De thematische nauwkeurigheid kan worden uitgedrukt in een percentage correct geclassificeerde attributen, door een misclassificatie matrix, een lijst van mogelijke alternatieve classificaties, de waarschijnlijkheid $\{\omega_i, P(X = \omega_i)\}$ van juiste classificaties, de maximale waarschijnlijkheid of de m grootste waarschijnlijkheden in ω_i .

- *temporele nauwkeurigheid* beschrijft de juistheid van tijdstippen opgeslagen in het gegevensbestand en van de bijhouding (actualiteit) door:
 - weergave van *fouten in de tijd*;
 - tijdstip van *laatste bijhouding*, met onderscheid tussen het oorspronkelijke ontstaan van de dataset, een verandering, verwijdering of een onveranderd gebruik;
 - *snellheid van wijzigingen* van objecten of attributen gedurende een bepaalde periode;
 - *geldigheid*, die aangeeft of de gegevens in het bestand nog geldig zijn, niet meer geldig zijn of nog geldig moeten worden.

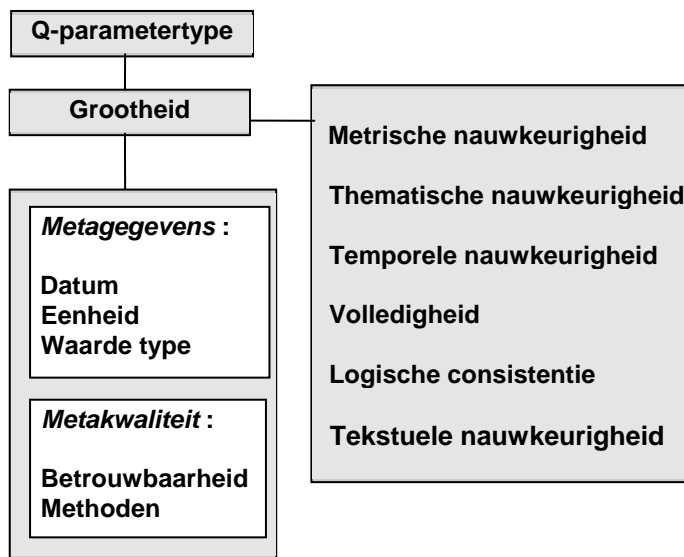
Attribuuttype (van objecten van relaties)	Thematisch	Metrisch	Temporeel	Logische consistentie
	Bron	Nauwkeurigheid	Volledigheid	
Enkelvoudige attributen	Beschrijving en data (wie, hoe, wanneer en waarvoor)	RSME, σ Fouten ellips, enz.	Volledigheid Overcompleteet (Evaluatie: verwacht, berekend, gecontroleerd) Veranderingssnellheid, bijhouding, enz.	. topologie . validiteit . integriteit . formaat . enz.
Meervoudige attributen		Misclassificatie matrix, enz.		

Tabel 2. Ontwerp matrix voor kwaliteitsrapportage.

- *completeheid* geeft een schatting aan van:
 - *omissie* en *overcompleteheid*, uitgedrukt in een percentage van missende en overcomplete gegevens in het bestand, in vergelijking met de nominale grondslag;
 - aantal wijzigingen die reeds zijn geregistreerd in de werkelijkheid maar nog niet zijn ingevoerd in het bestand *sinds de laatste bijhouding*, of de *verwachte veranderingen* die kunnen hebben plaatsgevonden tussen het tijdstip van de levering en de laatste bijhouding;

Bij completeheid kan onderscheid worden gemaakt tussen de completeheid voor de gegevens (zoals boven omschreven) en de completeheid voor het model (refererend naar de semantische kwaliteit: de juistheid van de definitie van de nominale grondslag);
- *logische consistentie* geeft de kwaliteit aan van de relaties die in het bestand voorkomen en die door de structuur aan het bestand worden opgelegd [6, 9, 10]. Voor ruimtelijke informatie zijn daarbij de topologische, grafische en visuele controles van belang, waarbij
 - *1-D topologische consistentie* bijvoorbeeld een indicatie is voor het percentage knooppunten dat aanwezig zou moeten zijn in het bestand maar er niet is;
 - *2-D topologische consistentie* bijvoorbeeld het percentage incorrect gevormde polygoenen aangeeft;
 - controle op *juistheid* van relaties en *bestaanbaarheid van relaties* zoals vastgelegd in de bestandsstructuur;
 - *attribuut- en formaat validiteit*.

Voor deze controle moet een binaire waarde (wel of niet -succesvol- uitgevoerd) worden opgegeven.



Figuur 5. Overzicht kwaliteitsparameters.

- *semantische kwaliteit* kan worden omschreven met de 'kwaliteit waarmee de geografische objecten worden beschreven in overeenstemming met het geselecteerde model' [6]. Semantische kwaliteit heeft betrekking op de betekenis van geografische objecten (zoals geselecteerd, maar vooral geabstraheerd in de nominale grondslag) en niet op de geometrische representatie. Daarmee wordt de semantische kwaliteit een element voor de beoordeling van de bruikbaarheid van een dataset;
- *tekstuele nauwkeurigheid* geeft:
 - het percentage spelfouten;
 - het gebruik van alternatieve spelling;
 - consistentie in afkortingen.
- *evaluatie procedure en resultaten* geven een beschrijving of een referentie naar andere documentatie die de methodologie beschrijft die is toegepast om een kwaliteitsparameter vast te stellen resulterend in een waarde of een verzameling waarden of die een evaluatie omvat van de verkregen waarde ten opzichte van de gestelde waarden in de nominale grondslag. Voor elke kwaliteitsparameter moet een methode worden vastgesteld waarop de waarde van de kwaliteitsparameter is bepaald.

Conclusies

De implementatie van het weergegeven kwaliteitsmodel om de mogelijkheden van geografische gegevens te beschrijven omvat veel verschillende kwaliteitsparameters die bovendien betrekking hebben op verschillende elkaar overlappende datasets. De elementaire aanname dat gegevens kunnen worden verzameld in homogene datasets en deze als nieuwe entiteiten kunnen worden beschouwd, maakt het mogelijk om de verschillende kwaliteitsaspecten te beschrijven in termen van kwaliteitsparameters en hun waarden op te slaan en te gebruiken bij de beoordeling van geografische gegevens voorafgaand aan het gebruik. Hoewel theoretisch correct, maakt dit systeem de bestandsstructuur complex; ook zal de bestandsgrootte extreem toenemen ten opzichte van zijn oorspronkelijke grootte.

Momenteel is nog geen ervaring opgedaan met de implementatie van een dergelijk complex, en omvangrijk model. Toch zijn dergelijke onderzoeken noodzakelijk te leren welke minimale set van kwaliteitsparameters door gebruikers noodzakelijk wordt geacht

om beoordeling van een gegevensbestand mogelijk te maken. Daarbij kan gebruik worden gemaakt van de evaluatiematrix van kwaliteitsbeoordeling:

		Data domein		
		Semantiek	Ruimte	Tijd
Kwaliteits- component	Nauwkeurigheid			
	Resolutie			
	Volledigheid			
	Consistentie			

Daarnaast is nog onderzoek noodzakelijk om ook de grafische voorstelling van de kwaliteit van gegevensbestanden weer te geven en te interpreteren.

Literatuur

1. Beard, K., and W. Mackaness, *Visual access to data quality in geographic information systems*, Cartographica 1993, V.30, pp 37 - 45.
2. Buttenfield, B, *Representing data quality*, Cartographica 1993, V.30, pp 1 - 7.
3. CEN TC 287 WG 2, *Data description: Quality*, Working paper N15, August 1994.
4. CEN TC 287 PT05, *Draft quality model for Geographic Information*, Working paper D3, January 1995.
5. Fisher, P., *Visualization uncertainty in soil maps by animation*, Cartographica 1993, V.30, pp. 20- 27.
6. Guptill, S and J. Morrison., (eds), *The elements of spatial data quality*, Elsevier 1995.
7. ISO, *Quality. Vocabulary*. Standard ISO 8402.
8. ISO, *Guide to expression of uncertainty in measurement*. Guide established by a working group supported by seven organizations, including BIPM and ISO, 1993.
9. Moellering, H., (ed), *A draft proposed standard for digital cartographic data*, Report #8, ACSM 1987 Bethesda, Md.
10. Moellering, H., (ed), *Digital cartographic data standards: An interim proposed standard*, Report #6, ACSM 1985 Bethesda, Md.
11. Naisbitt, J., *Global Paradox*, Avon books, New York 1994.
12. Pirsig, Robert M., *Zen and the Art of Motorcycle Maintenance. An Inquiry into Values*. ISBN 90-351-1095-1.
13. Tobler, W., *Frame independent spatial analysis*, in Accuracy of spatial databases, pp. 115 - 122, Taylor and Francis 1989, New York.
14. Veregin, H., *Error modeling for the mapoverlay operation*, in Accuracy of spatial databases, pp. 3 - 18, Taylor and Francis 1989, New York.

Visualisatie van kwaliteitsinformatie

F.J. Ormeling

Faculteit Ruimtelijke Wetenschappen
Universiteit Utrecht
E-mail: f.ormeling@frw.ruu.nl

Wat is kwaliteit?

Kartografie zit tussen een aantal wetenschapsgebieden in die zich met nauwkeurigheid bezig houden. Bij de geodesie gaat het om de nauwkeurigheid van de plaatsbepaling, bij de geografie gaat het om de nauwkeurigheid van de attribuutsinformatie, en bij de waarnemingspsychologie gaat het om de nauwkeurigheid van onze observaties. Hun vragen zijn respectievelijk: met welke nauwkeurigheid het verschijnsel daàr voorkomt, het tot een bepaalde categorie wordt gerekend en of we dat ook als zodanig waarnemen. Nemen we als voorbeeld de kartering van bevolkingsaantallen van agglomeraties, dan gaat het om de positionele nauwkeurigheid van de plaatsstippen, de nauwkeurigheid van de vaststelling van het aantal inwoners (wie reken je wel of niet mee?) en tenslotte of je de verschillen tussen de op een aantal locaties weergegeven proportionele symbolen goed kunt aflezen. Uit psychofysisch onderzoek weten we inmiddels dat er een wijdverbreide onderschatting is van grotere proportionele symbolen ten opzichte van kleinere. Teken je het ene symbool als tweemaal zo groot als het andere, dan wordt het als slechts anderhalf maal zo groot waargenomen.

De termen kwaliteit en onzekerheid zullen hier door elkaar heen worden gebruikt, hoewel onzekerheid een iets ruimere inhoud heeft, waar ook een aantal aspecten onder vallen die niet bij kwaliteit horen. In de karteringswetenschappen wordt gegevenskwaliteit nu algemeen opgevat als geschiktheid voor gebruik - fitness for use - van de aangeleverde, bewerkte gegevens. Die geschiktheid kan beïnvloed zijn door opnamemethode en door bewerkingen. Maar de geschiktheid van de uiteindelijke gegevens kan evengoed beïnvloed zijn door de keuze voor een bepaald model (we spreken dan van model uncertainty; Karimi and Hwang 1996) of van andere, ruimtelijk werkende factoren. Men kan zich bijvoorbeeld baseren op een lineaire toename van neerslagwaarden tussen waarnemingsstations, in plaats van op een exponentiële toename zoals in feite het geval lijkt te zijn. Er kan ook sprake zijn van een bepaalde vooringenomenheid - bias - bij de opnametechnieken of apparatuurfout, zodat alle gegevens eenzelfde vertekening vertonen. Ik wil echter hier niet ingaan op de oorzaken van een betere of mindere kwaliteit van de gegevens, maar als uitgangspunt nemen dat er, om welke reden dan ook, sprake is van verschillen in gegevenskwaliteit over het bestudeerde gebied - bijvoorbeeld omdat het model niet even goed op het hele gebied past, of omdat de gegevensverzameling in een deel van het gebied onder de maat was. Uitgangspunt is dus een bepaalde

kwaliteit van het gegevensbestand als geheel, en daarnaast een ruimtelijke variatie in de kwaliteit over het bestudeerde gebied.

Waarom zouden we die ruimtelijke variatie in de kwaliteit willen kennen en vervolgens visualiseren? Van het eerste hangt af of het zinvol is verschillende bestanden met elkaar te combineren of niet. Combineren we een bodemkaart met grenzen die een nauwkeurigheid hebben van 100 m met een topografische kaart 1:50 000 waarop de terreinobjecten een geometrische nauwkeurigheid hebben van 10 m, dan schep je daardoor bij gecombineerde presentatie de suggestie dat de bodemgrenzen ook die hogere nauwkeurigheid hebben gekregen. En een dergelijke combinatie is weinig zinvol, en daarom is kennis van de kwaliteit belangrijk.

We willen de verschillen in kwaliteit visualiseren omdat we op basis van de waargenomen waarden beslissingen willen nemen, en we er zeker van willen zijn dat die beslissingen vallen op basis van betrouwbare gegevens. Wanneer er nu sprake is van onderzoek op het gebied van de gegevenskwaliteit dan betreft dat naast de visualisatie op het ogenblik vooral de attribuutnauwkeurigheid, dus de nauwkeurigheid van toekenning van objecten aan klassen.

Waarschijnlijkheden en mogelijkheden (probabilities en possibilities) bij het vaststellen van de attribuutnauwkeurigheid

Daarbij spelen ook de opvattingen over het ontstaan van onzekerheid een belangrijke rol, en dat kan toegelicht aan de hand van het tegenover elkaar zetten van waarschijnlijkheden en mogelijkheden (probabilities en possibilities).

Bij het vaststellen van de attribuutnauwkeurigheid gaat het om de zekerheid of onzekerheid dat een bepaald object tot een bepaalde attribuutklasse hoort. De onzekerheid daarover kan onder andere afhangen van het feit of er sprake is van een:

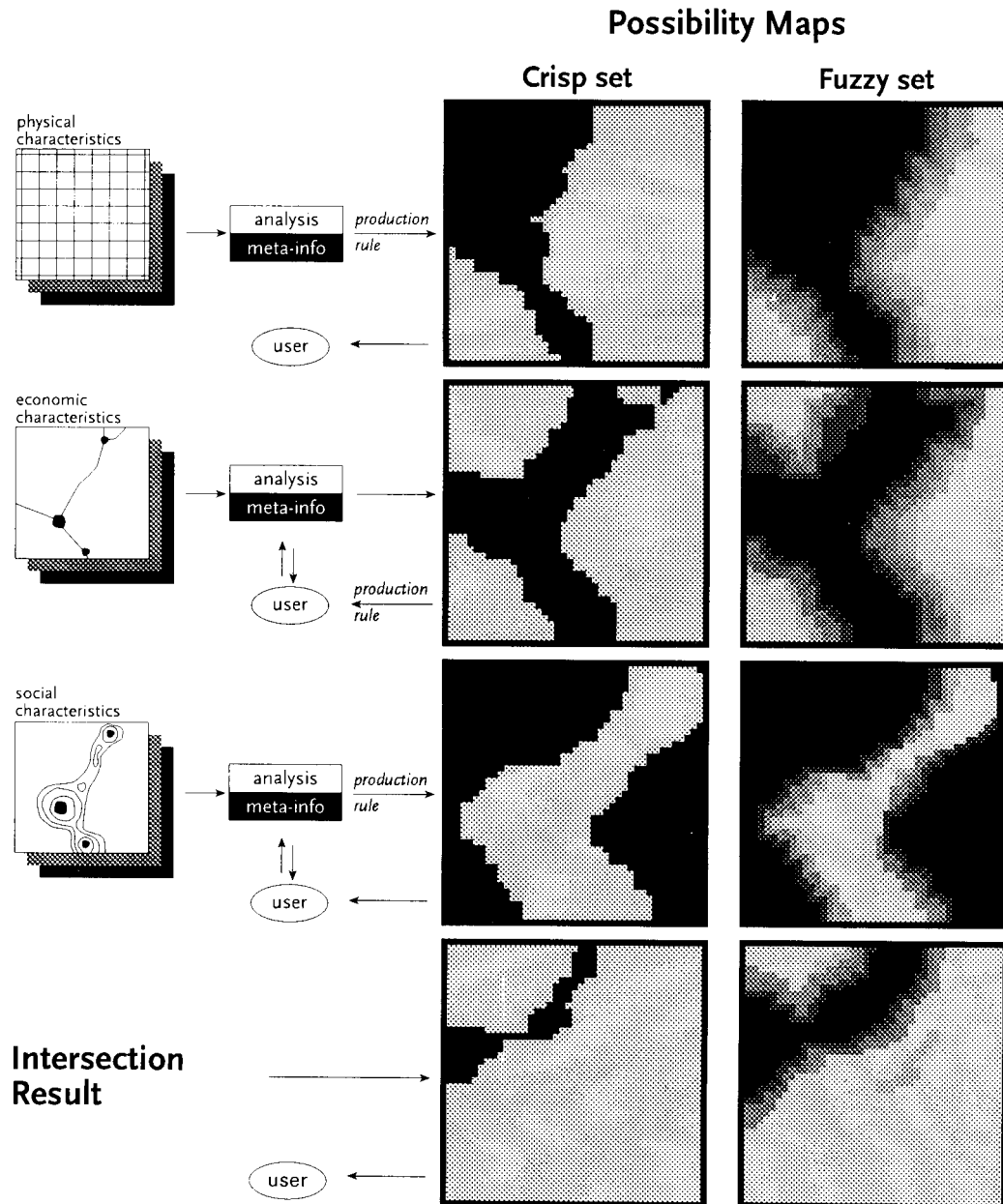
- *al of niet correcte interpretatie (of correct toepassen van definities);*
- *al of niet correcte taxonomische classificatie (bijvoorbeeld bij indeling in bodemsoorten);*
- *al of niet correcte numerieke classificatie (bijvoorbeeld in bevolkingsdichtheidsklassen).*

Bij het vaststellen van de attribuutnauwkeurigheid spelen er verschillende opvattingen over de oorzaak van het ontstaan van afwijkingen in de nauwkeurigheid. Er kan sprake zijn van de opvatting dat het feit of een object tot een bepaalde klasse behoort door het toeval wordt bepaald. In dat geval worden statistische waarschijnlijkheidstheorieën gehanteerd om vast te stellen of het object wel of niet tot de betreffende klasse hoort. Voor elk object geldt dat het maar tot één klasse tegelijk kan horen. De kans dat een object tot een klasse hoort wordt aangegeven met waarschijnlijkheidswaarden of probabilities.

Bij de fuzzy set benadering daarentegen gaat men ervan uit dat de werkelijkheid vaag is en niet met de klassieke wiskunde kan worden beschreven, omdat er teveel aspecten van nauwkeurigheid zijn die niet aan het toeval kunnen worden toegeschreven. Volgens de fuzzy set theorie bestaat er een overgang tussen het volledig wèl en het volledig niet lid zijn van een klasse voor elk object. Objecten kunnen volgens deze theorie lid zijn van meerdere klassen tegelijk. De mate waarin objecten tot een klasse horen wordt uitgedrukt met mogelijkheidswaarden of possibilities. De fuzzy set theorie wordt toegepast

wanneer er sprake is van vage omschrijvingen, die elkaar nooit helemaal kunnen uitsluiten.

Zowel bij kwantificeren van waarschijnlijkheden als van mogelijkheden kunnen de waarden variëren tussen 0 en 1. Bij de waarschijnlijkheidstheorie geeft een waarde van 0,7 aan dat er 70% kans is dat het object tot klasse A hoort en niet tot klasse B, terwijl er 30% kans is dat het tot klasse B hoort en niet tot A. Bij de fuzzy set theorie betekent de waarde 0,7 dat het object voor 70% aan klasse A valt toe te kennen en voor 30% aan klasse B.



Figuur 1. Vergelijking van resultaten van een zeefanalyse met resp. 'crisp' en 'fuzzy' grenzen (uit Van der Wel en Hootsmans 1993).

Het effect van een en ander op de kaart is het volgende: Wanneer we een spoorlijn moeten inplannen die op een stevige ondergrond moet lopen en ten bate van de constructiewerkzaamheden niet te ver van het bestaande wegennet verwijderd moet zijn en waar de bevolking geen last van moet hebben, kan je de betreffende factoren opgeven in een GIS, dat bepaalt voor elke factor of een gebiedseenheid wel of niet aan de eisen beantwoordt en via overlay-analyse komt dan het geschikte gebied er uit. Maar uit ervaring weten we dat er geen scherpe ('crisp') bodemgrenzen zijn, en dat bevolkingsdichtheid een functie is van de grenzen van telgebieden; met andere woorden in plaats van scherpe grenzen moet je de mate waarin gebieden aan eisen beantwoorden karteren - en als je dan een GIS-pakket zou hebben met overlay-analyse functionaliteit die met al die modaliteiten rekening kan houden, dan komt er een heel ander beeld uit, op basis waarvan we onze beslissingen moeten nemen (zie figuur 1).

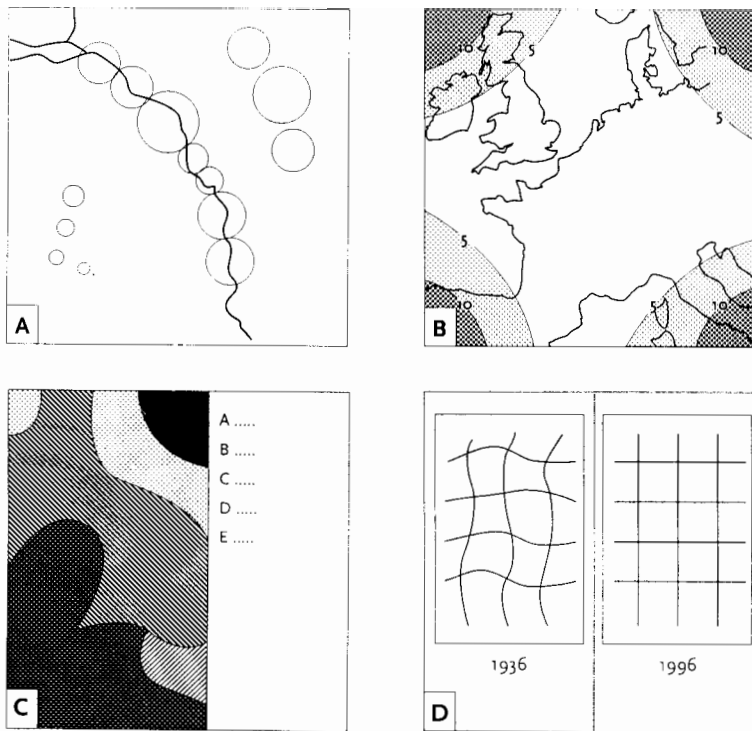
Bij de huidige GIS-pakketten worden objecten nog aan klassen toegekend op basis van de waarschijnlijkheden. We weten nog niet precies hoe, bijvoorbeeld bij overlay-operaties, de onzekerheden in elk van de sets van gegevens met elkaar combineren. Jane Drummond heeft voor haar proefschrift (1991) nagegaan hoe, door combinatie van gegevens over bodemsoorten, temperatuur, hellingspercentages en regenval, op het eiland Dominica de geschiktheid voor de uitbreiding van bananenteelt kon worden aangegeven. Op basis van steekproeven kon vervolgens worden vastgesteld dat maar voor 26% van het op basis van de overlay-operatie als geschikt aangemerkte oppervlak ook inderdaad van die geschiktheid sprake was. Bodemsoorten, temperatuur en regenval worden op basis van steekproeven vastgesteld, de nauwkeurigheid van de hellingspercentages zal een functie zijn van het aantal hoogtemetingen per oppervlakte-eenheid en de grootte van de gridcel waarbinnen de hellingen uit het verschil tussen hoogste en laagste hoogtewaarden bepaald worden. Al de bestudeerde factoren hebben een bepaalde onzekerheid in zich, en het zal dus relevant zijn die te visualiseren.

Traditionele kwaliteitsvisualisatie en recente trends

Traditioneel zijn in de kartografie een aantal visualisatiemethoden ontwikkeld voor weergave van de geometrische nauwkeurigheid (Van der Wel e.a. 1994). We kennen stip-, streep- en vlek-weergavemethoden (zie figuur 2).

Voor de eerste zijn de indicatrices van Tissot een mooi voorbeeld, omdat ze zowel de hoek- als oppervlaktevertekening aangeven bij projecties. Een ander voorbeeld zijn de cirkels van Mekenkamp die er zijn betrouwbaarheidsonderzoek van oude kaarten mee visualiseert. Voorbeelden van lineaire visualisatiemethoden zijn de Perkalbanden en de isolijnen van gelijke vertekening zoals die ook weer bij projecties worden gebruikt; maakt men het lezen van de isolijnen eenvoudiger door de tussenliggende vlakken met grijswaarden op te vullen, dan gaat de methode over in een areale. Een andere vorm van het gebruik van vlakken om vertekening aan te geven zijn de grids die ook in de historische kartografie worden gebruikt en uitgaande van het moderne kaartbeeld aangeven hoe groot de vertekening in de plaatsbepaling in het verleden was. En ten slotte zijn er de uit vlakken bestaande betrouwbaarheidsdiagrammen of reliability diagrams.

Eigenaardig genoeg zijn al die traditionele vormen van kwaliteitsvisualisatie ontwikkeld voor het aangeven van de geometrische kwaliteit of nauwkeurigheid (dus de mate waarin een plaatsbepaling overeenkomt met de 'correcte' coördinaatgegevens), met uitzondering misschien van Halley's isogonenkaart. Dat ligt achteraf wel voor de hand omdat de topografische kartering ouder is dan de thematische. Maar er zijn, zoals Aalders in



Figuur 2. Voorbeelden van stip-, streep- en vlekweergavemethoden van de geometrische nauwkeurigheid: a) cirkelmethode van Mekenkamp; b) isolijnen van gelijke vertekening; c) betrouwbaarheidsdiagrammen (Van der Wel e.a. 1996); en d) vertekeningsgrijs (Imhof 1964).

deze publicatie aangaf, nog andere aspecten aan de gegevenskwaliteit te onderscheiden dan de geometrische. Bij het NCGIA, het consortium van Amerikaanse universiteiten dat voor de National Science Foundation het GIS-onderzoek uitvoert, onderscheidt men naast de geometrische kwaliteit ook de attribuutnauwkeurigheid (de mate dus waarin de benoemde aspecten juist geklassificeerd zijn), verder de volledigheid, de logische consistentie of topologische integriteit en de afkomst of lineage van de gegevens. Onder dat laatste kan ook worden verstaan de actualiteit of mate van veroudering van de gegevens. Voor het bepalen van de mate van actualiteit en volledigheid zijn nog geen maten ontwikkeld; er bestaat derhalve ook nog geen adequate visualisatiemethode voor de weergave van deze aspecten. Dat geldt al helemaal voor de logische consistentie of topologische integriteit.

Volledigheid is wel al een issue geweest in een aantal onderzoeken van het Staring Centrum waarin werd nagegaan in hoeverre karteringsvoorschriften voor de topografische kaart consequent werden opgevolgd dan wel ook beïnvloed werden door overwegingen betreffende de grafische dichtheid en leesbaarheid van de kaarten. In dat kader kunnen de onderzoeken van De Veer (1985) naar alleenstaande bomen en van Bakermans (1986) naar de volledigheid van de weergave van perceelsscheidingen worden genoemd.

De huidige stand van het onderzoek naar de gebruiksmogelijkheden van grafische variabelen voor de weergave van de kwaliteitsaspecten wordt gegeven door Van der Wel en co-auteurs in zijn klassieke studie uit 1994 (visualisation of data quality). Daarin wordt aangegeven welke variabelen geschikt zijn om de op verschillende meetschalen bepaalde kwaliteitswaarden weer te geven (figuur 3).

level of measurement	nominal	ordinal	interval	ratio	
graphic variable	L	L A P C	A P C	A P	
colour hue	+	- ± ± -	+ + ±	- -	
orientation	+	- - - -	- - -	- -	
shape	±	- - - -	- - -	- -	
size	-	- + + -	+ + -	+ +	- not suitable
texture	+	± ± ± ±	- - -	- -	± suitable under predefined conditions
value	-	+ + - +	+ + ±	- -	+ suitable
abstraction	-	- - + +	- - -	- -	L lineage
pattern	+	- - - -	- - -	- -	A attribute accuracy
colour saturation ¹	-	+ + + +	+ + -	- -	P positional accuracy
					C completeness

Figuur 3. Geschiktheid van de grafische variabelen voor weergave van op verschillende meetschalen bepaalde kwaliteitswaarden (uit Van der Wel e.a. 1996).

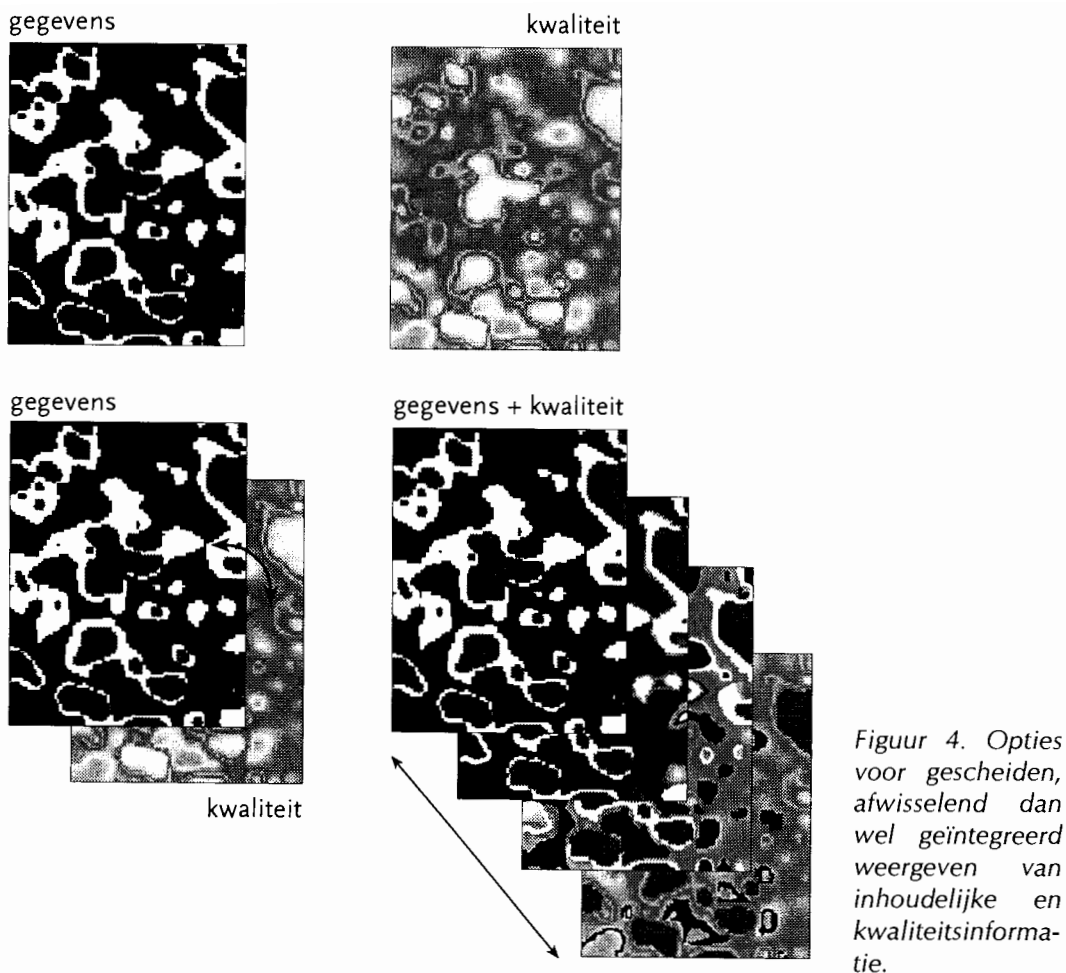
Uit het betreffende schema kunnen we aflezen dat wanneer er sprake is van ratioverschillen in de geometrische nauwkeurigheid we bij weergave alleen gebruik kunnen maken van verschillen in grootte (zoals bij Mekenkamp), en dat grijswaarde en verzadiging over het algemeen beide even goed scoren bij ordinale en intervallschaal.

Nieuwe ontwikkelingen in de visualisatie van kwaliteitsmaten

Na die inventarisatie van Van der Wel c.s. is onder andere nagegaan hoe de kwaliteitsinformatie het beste met de inhoudelijke informatie kan worden gecombineerd. We hebben als opties a) het naast elkaar weergeven van het inhoudelijke beeld en het kwaliteitsplaatje van de inhoudelijke gegevens; b) het afwisselend presenteren (toggle) van beide beelden en c) de geïntegreerde presentatie van zowel inhoudelijke als kwaliteitsgegevens (zie figuur 4). Bij de laatste mogelijkheid kunnen we dan nog differentiëren tussen statische en dynamische presentatie. Hootsmans (1996) heeft met name met het gebruik van animatie van de kwaliteitsbeelden, gecombineerd met het inhoudelijke beeld, een overtuigende exploratiemogelijkheid geboden. In het Diffuse programma dat hij ook op de internet-site van de disciplinegroep Kartografie van de Universiteit Utrecht heeft gezet, is het mogelijk, door het manipuleren van de slider bar die de dispersion value (de breedte van de overgangszone tussen verschillende klassen) aangeeft, na te gaan in een animatie welk deel van het inhoudelijke beeld stabiel is, en welk deel direct beïnvloed wordt door de manipulatie. Hij heeft het gecombineerd gebruik van grijswaarde voor gebieden met een te hoge onzekerheid en kleur voor inhoudelijke gegevens voor gebieden met een aanvaardbare onzekerheid geïntroduceerd. Deze methode verhindert direct conclusies te trekken voor de grijze gebieden.

Ook het speciale nummer van Computers and Geosciences dat in 1997 naar aanleiding van het ICA congres in Stockholm uitkwam, bevat bijdragen over de methodologie van het gebruik van animaties voor weergave van onzekerheid (Ehlschlaeger e.a. 1997).

Bij die visualisatie is het zinvol uit te gaan van metaforen die onzekerheid in zich dragen, zoals mist, ruis, opflikkeren van een lichtbron, een willekeurig verdeeld patroon of de onmogelijkheid ergens in detail naar te kijken. Daarnaast moet de mogelijkheid geboden worden om de verschillende waarden voor onzekerheid te visualiseren en daarvoor komt (naast grootte) grijswaarde het meest in aanmerking, gevolgd door verzadiging. Hoe meer een kleur verzadigd is, hoe zuiverder hij is, en dat geeft in veel gevallen



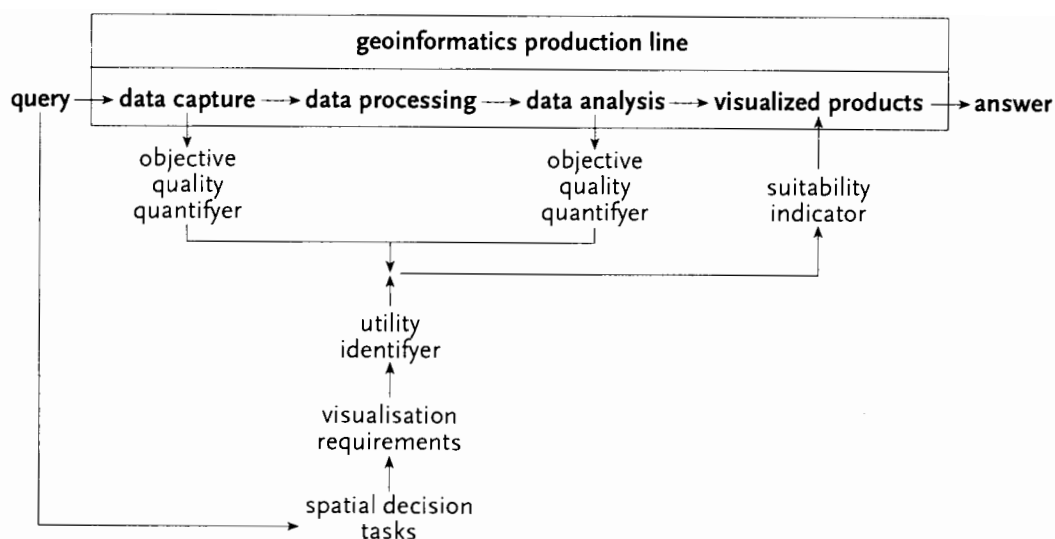
Figuur 4. Opties voor gescheiden, afwisselend dan wel geïntegreerd weergeven van inhoudelijke en kwaliteitsinformatie.

associaties met zekerheid. Jiang heeft in zijn proefschrift (1996) nagegaan bij welke van deze grafische variabelen gebruikers de sterkste associaties optreden met onzekerheid, en in zijn onderzoek scoort het willekeurige puntenpatroon (random dot) hoog.

Hootsmans heeft nog nagegaan in welke gevallen animatie met een slide bar - of interactieve visualisatie zoals hij het noemt - toegepast kan worden, zonder daarbij al uitspraken te doen over de optimale toepassing van de animaties.

Geschiktheid voor gebruik vs nuttigheid (kwaliteitsinformatie per deeltaak)

Er is op dit moment nog geen eendoordeel over de meest geschikte weergavemethode van de onzekerheid, en dat hangt waarschijnlijk samen met het feit dat de toepassingen te divers zijn en de soorten van ruimtelijke besluitvorming nog onvoldoende in kaart zijn gebracht. Want wat hebben we gedaan tot nu toe: We hebben in de geoinformatische productielijn een paar objectieve kwaliteitscriteria ingebouwd. We weten daardoor dat de antwoorden op de gestelde vragen een bepaalde betrouwbaarheid hebben. Maar in dit hele proces - en dat komt voor mij een beetje neer op de typische benadering door geodeten - kijkt men niet verder dan het antwoord. Men kijkt niet naar wat er met dat antwoord gedaan wordt, voor welke ruimtelijke besluitvormingstaken het gebruikt wordt. Besluitvormingstaken zijn complex, en bestaan meestal uit een aantal deeltaken die elk hun eigen eisen stellen aan het beeld waarop de beslissing gebaseerd zal worden



Figuur 5. Geoinformatieproductielijn waarin objectieve kwaliteitscriteria zijn ingebouwd, maar waarin ook op basis van de besluitvormingstaken gekeken is naar de nuttigheid en de geschiktheid van deze maten.

(Ormeling, 1998). Neem als voorbeeld de navigatie aan de hand van kaarten. Op basis van de kaart wordt georiënteerd, de bestemming gevonden, een route geselecteerd, een aantal herkenningspunten gezocht en tijdens de route geïdentificeerd en de bestemming uiteindelijk geverifieerd. Elk van die deeltaken vereist zijn eigen nauwkeurigheidsbeeld: de aard van de deeltaak bepaalt de eisen aan de visualisatie, en men zal na moeten gaan wat de nuttigheid en geschiktheid van de objectieve kwaliteitsmaten is voor deze beslissingsdeeltaken (zie figuur 5). Ik denk dat we pas in zo'n situatie na kunnen gaan welke manier van visualisatie van de kwaliteitsgegevens het beste is, voor welk soort deeltaak.

Daar komt nog bij dat we niet alleen de visualisatiemethode kunnen variëren en de specifieke deeltaak waarvoor hij gebruikt wordt, maar ook de kwaliteitsindicatoren. Van der Wel en Hootsmans (1993) hebben de bestaande kwaliteitsmaten uitgebreid met een aantal nieuwe begrippen die bijvoorbeeld het verschil tussen de maximum en second likelihood of de maximum en second possibility aangeven. Tot de door hen ontwikkelde maten horen ook fuzziness, ambiguity, credibility en boundary certainty. Met fuzziness wordt de mate aangegeven waarin een object aan een bepaalde klasse kan worden toegekend; ambiguity geeft aan hoe goed een object dat aan de ene klasse is toegekend van andere klassen onderscheiden kan worden. Met credibility en boundary certainty worden maten voor de stabiliteit van klassegrenzen op de betreffende afbeelding aangegeven. En wanneer de besluitvormingstaak bijv. bestaat uit het afgrenzen van gebieden ten opzichte van elkaar is zo'n laatste maat zinnig (zie figuur 6).

We kunnen dus specifieke deeltaken onderscheiden bij de ruimtelijke besluitvorming op basis van kaarten, daaraan aangepaste kwaliteitsmaten en adequate visualisatiemethoden.

De ideale kwaliteitsindicatiemodule

Het is een van de doelstellingen van het kartografische onderzoeksprogramma op de Universiteit Utrecht om een aan een GIS te hangen module te bouwen (zie figuur 7) die berekent en aangeeft op basis van de metagegevens, wat de kwaliteit is van de gevisualiseerde inhoudelijke gegevens, gericht op hun toepassing. Daarvoor zullen soorten

	uncertainty measure			
	fuzziness	ambiguity	credibility	boundary certainty
static display environment				
• achromatic value				
separate	+	+	+	+
combined	+	+	+	+
• chromatic value/saturation				
lightness (-)	-	-	-	-
lightness (+)	-	-	-	-
saturation	-	-	-	-
• size	-	-	+	+
interactive display environment				
• achromatic value				
separate	+	+	+	+
combined	+	+	+	+
• chromatic value/saturation				
lightness (-)	+	+	-	-
lightness (+)	+	+	-	-
saturation	+	+	-	-
• size	-	-	+	+

Figuur 6. Nauwkeurigheidsmaten gerelateerd aan grafische variabelen (naar Hootsmans 1996).

toepassingen gerubriceerd moeten worden en moet worden uitgewerkt voor welke toepassing welke kwaliteitsmaat het meest relevant is, en vervolgens hoe die kwaliteitsmaat het beste kan worden afgeleid en gevisualiseerd. En voor dat geheel moet dan nog een interface worden gebouwd die de gebruiker in staat stelt bij het intypen van een bepaalde kaartgebruikstoepassing meteen het goede kwaliteitsbeeld voorgeschoteld te krijgen. Het is nog niet zeker of we dat onderzoek kunnen voltooien omdat het zelfstandige GIS-Kartografie onderzoeksprogramma mogelijk binnen de Nethur onderzoeksschool wordt gebracht, maar wellicht kan het in het kader van een stedelijke toepassing alsnog worden gerealiseerd.

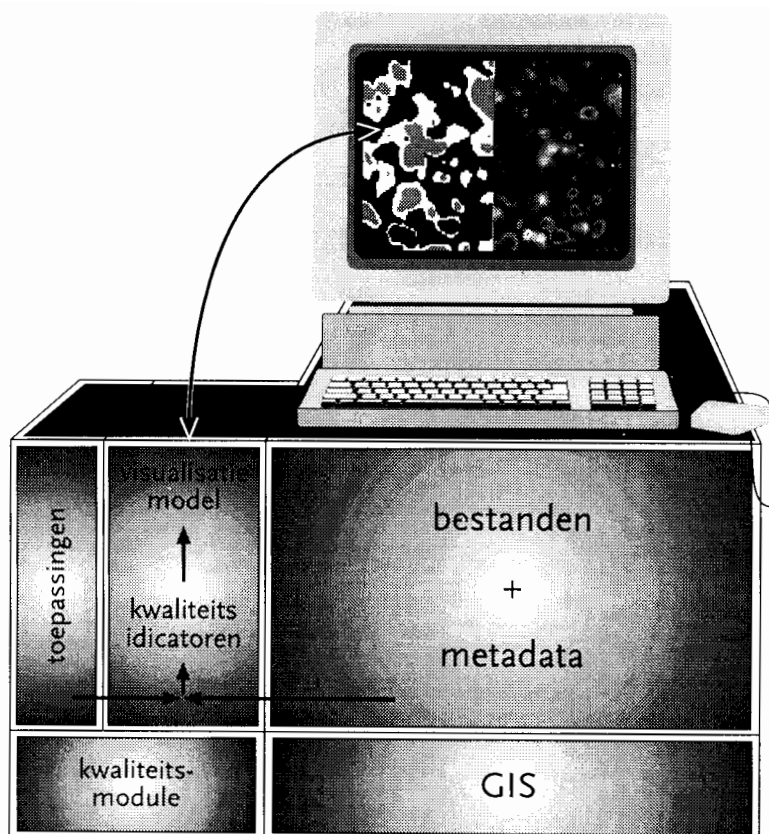
GIS-kartografen moeten op de hoogte zijn van de gegevenskwaliteit om de juiste visualisatiemethoden te kiezen om gegevens te communiceren, of om de juiste omgeving te scheppen om de gegevens op een relevante manier te kunnen exploreren. Objectieve kwaliteitsmaten vormen daar de basis voor, maar als we de verschillende kaartgebruiksstappen bezien die nodig zijn voor de besluitvorming, dan realiseren we ons dat er voor elke stap weer andere visualisatiemethoden nodig kunnen zijn. We moeten dus uiteindelijk op de hoogte zijn van de nuttigheid van kwaliteitsinformatie voor elk soort ruimtelijke beslissingstaak, en we moeten de geschiktheid van de kwaliteitsindicatoren baseren op hun gebleken nut bij zo'n taak.

Literatuur

M.M.G.J.Bakermans - Gebruiksbeperkingen van de moderne topografische kaart bij onderzoek in het cultuurlandschap. Reeks landschapsstudies 7. Wageningen: Pudoc 1986.

J.Drummond - Determining and processing quality parameters in geographic information systems. Unpubl. PhD thesis. University of Newcastle upon Tyne 1991.

C.R.Ehlschlaeger, A.M.Shortridge and M.F.Goodchild - Visualizing spatial data uncertainty using animation. <http://geo.swf.uc.edu/chuck/CGdraft/paper/html> (Computers and Geosciences. Exploratory cartographic visualization Special issue. 1997).



Figuur 7. De geplande module voor visualisatie van gegevenskwaliteit.

R.M.Hootsmans - Fuzzy sets and series analysis for visual decision support in spatial data exploration. Proefschrift Universiteit Utrecht, 1996.

E.Imhof - Beiträge zur geschichte der topographischen Kartographie. Internationales Jahrbuch für Kartographie vol 4, 1964, pp 129-154.

Bin Jiang - Fuzzy overlay analysis and visualization in geographic information systems. Proefschrift, Universiteit Utrecht 1996.

H.A.Karimi and D.Hwang - Towards managing model uncertainty in GISs: an algorithm for uncertainty analysis of air quality advection models. pp 251-259, Geomatica, vol.50-3, 1996.

P.G.M.Mekenkamp - Die Entwicklung einer neuen Methode für die Bestimmung der Genauigkeit von alten Karten. In: Proceedings Kartographiehistorisches Kolloquium Oldenburg 1990. Oldenburg: Deutsche Gesellschaft für Kartographie 1990.

F.J.Ormeling - Map use steps and their data quality requirements. Cartographic Perspectives. forthcoming winter 1998.

A.A. de Veer - Geografie van de opvallende boom in het agrarisch landschap van Nederland. Proefschrift Rijksuniversiteit Groningen 1985.

F.J.M. van der Wel, R.J.Hootsmans and F.J.Ormeling - Visualization of data quality. pp 313-333 in: A.M.MacEachren and D.R.F.Taylor (eds) - Visualization in modern cartography. Modern Cartography vol.2. Oxford: Pergamon 1994.

F.J.M. van der Wel en R.J.Hootsmans - Visualization of quality information as an indispensable part of optimal information extraction from a GIS. In P.Mesenburg (ed) Proceedings 16th International Cartographic Conference. Deutsche Gesellschaft für Kartographie. Keulen (1993) pp 881-897.

Kwaliteit in de CBS-praktijk

W.A. de Rooij

Centraal Bureau voor de Statistiek
E-mail: wroy@cbs.nl

Inleiding

De doelstelling van het Centraal Bureau voor de Statistiek (CBS) is om gegevens te verzamelen, te bewerken en te publiceren voor zowel beleid, (wetenschappelijk) onderzoek als de praktijk. Bij de kwaliteit van de statistieken werd traditioneel de meeste aandacht besteed aan de betrouwbaarheid en de continuïteit van de gepubliceerde gegevens. Aan andere kwaliteitsaspecten, met name tijdigheid, werd tot voor enkele jaren minder aandacht besteed. Inmiddels is er een traject gestart om kwaliteitszorg CBS-breed in te voeren.

In dit artikel zullen de wensen van de gebruikers van statistieken ter sprake komen en de wijze waarop het CBS daar op in speelt. Daarna wordt ingegaan op de gesignaleerde knelpunten op het gebied van kwaliteitszorg, de huidige maatregelen om die te verbeteren en de plannen voor de nabije toekomst. Vervolgens worden de specifieke problemen op het gebied van regionale en geometrische gegevens toegelicht. Afgesloten wordt met een kort overzicht.

Wensen gebruikers

De kwaliteit van statistieken wordt mede bepaald door het doel waarvoor ze gebruikt worden. Het achterhalen van de doelen waarvoor specifieke statistieken worden gebruikt, is haast ondoenlijk vanwege de omvang. Om de globale sterke en zwakke punten op te sporen zijn echter door het CBS een aantal onderzoeken uitgevoerd onder diverse soorten gebruikers.

Uit de onderzoeken bleek dat gebruikers het CBS sterk vinden overkomen in het leveren van onafhankelijke, betrouwbare gegevens over een breed scala van onderwerpen. Daarnaast werden ook de continuïteit van de gegevens (tijdreeksen) en het gebruik van de statistische methoden genoemd als sterke punten. Minder sterk vond men de tijdigheid, waarbij voor sommige gebruikers actualiteit zelfs voor accuratesse gaat. Minder sterk vond men ook de toegankelijkheid van de gegevens; publicaties zijn veelal op een bepaalde statistiek gericht in plaats van op een bepaald thema.

Internationaal gezien heeft het CBS zich een goede positie verworven. Onder andere door de kwaliteit van de nationale rekeningen en het gebruik van moderne informatie-

technologie eindigde het CBS als derde in een poll gehouden door The Economist (in 1993). Maar ook hier werd tijdigheid als een minder sterk punt aangemerkt.

Statistisch proces

Tot medio jaren tachtig verzamelde het CBS zijn gegevens voornamelijk door het houden van een groot aantal (vooral schriftelijke) enquêtes. Het verwerken bestond uit het intoetsen van de gegevens, gevolgd door een controle op fouten en het corrigeren, vaak na telefonisch contact met de berichtgever. Daarna werden de gegevens verder geanalyseerd en verwerkt tot publicaties met vele tabellen. Teneinde het intoetsen en corrigeren van gegevens te beperken, worden voor een aantal enquêtes de gegevens direct elektronisch ingevoerd. Op deze wijze kunnen er gedurende het invoeren al controles uitgevoerd worden en kan het overtoetsen achterwege blijven. Dit heeft geleid tot een grotere efficiëntie en een kortere doorlooptijd.

Een aan het CBS opgelegde doelstelling is het terugdringen van de enquêtedruk voor zowel bedrijven als individuen. In de eerste plaats worden hiervoor enquêtes gecombineerd, zodat gegevens, die voor meer dan één statistiek nodig zijn slechts eenmaal geleverd behoeven te worden. Verder is er het streven om zoveel mogelijk gebruik te maken van elektronische gegevensinwinning, via EDI (Electronic Data Interchange) of diskettes, en het gebruikmaken van externe registraties, zoals de GBA (Gemeentelijke Basisadministratie). Het inzetten van EDI kan betekenen dat er een CBS-module opgenomen wordt in het administratief systeem van een bedrijf of een administratiekantoor, waarmee de voor bepaalde statistieken relevante gegevens verzameld wordt. Deze gegevens kunnen daarna via een modem doorgestuurd worden naar het CBS.

Publicaties

Bij het publiceren van gegevens wordt er gestreefd naar een 1-op-1-norm, d.w.z. dat de doorlooptijd van de peildatum van de inwinning tot en met het uitbrengen van de publicatie hoogstens gelijk is aan de periode waarover gepubliceerd wordt. Maandstatistieken bijvoorbeeld worden volgens deze norm binnen een maand na de peildatum gepubliceerd.

Om de samenhang tussen de statistieken te vergroten, wordt er onder andere druk gewerkt om alle gegevens die gepubliceerd worden op te nemen in een centrale database genaamd 'StatLine'. De huidige versie van StatLine (die nog niet geheel compleet) is inmiddels beschikbaar op CD-ROM, terwijl er ook een versie van StatLine gratis is raadplegen op de Internet-site van het CBS (<http://www.cbs.nl>).

De laatste jaren is er ook veel aandacht besteed aan het samenstellen van thematische publicaties, waarin samenhangende statistische gegevens worden gepresenteerd. Voorbeelden hiervan zijn het Assurantie Jaarboek, het Jaarboek Wonen en het Jaarboek Welvaartsverdeling. Deze publicaties worden in samenwerking met een intermediair, die het marktsegment goed kent en de marketing en verkoop zijn rekening neemt, op de markt gebracht.

Kwaliteitszorg

Wat betreft de kwaliteitszorg zijn er aantal knelpunten. Zo kunnen de eisen waaraan een statistiek moet voldoen onduidelijk zijn, zodat er mogelijk te veel aandacht wordt besteed aan micro-editing (het 'gaafmaken' van alle individuele records), terwijl macro-editing (het gebruik van statistische technieken bij het gaaf maken) zou kunnen volstaan. Omdat er onvoldoende inzicht is in de wensen en behoeften van gebruikers, zijn de

statistische producten niet goed te spitsen op specifieke gebruikers(groepen). Verder ontbreekt er een richtlijn voor het documenteren van de statistische processen, zodat de mate van detail bij het vastleggen van processen sterk wisselt tussen de verschillende statistieken.

Om de kwaliteitszorg te verbeteren is in 1996 een voorlopige richtlijn kwaliteitszorg opgesteld die als raamwerk zal fungeren. Het is de bedoeling dat de sectoren die de statistieken produceren zelf toegespitste richtlijnen voor de kwaliteitszorg zullen opstellen (uiteraard binnen een gemeenschappelijk kader). Sinds juni 1996 worden er audits (kwaliteitsonderzoeken) uitgevoerd. Hierbij worden de statistische processen van een sector door een audit-team van ongeveer vier CBS-medewerkers nader onderzocht. Uiteindelijk zal iedere statistiek eens in de vijf jaar aan een audit onderworpen worden. Aan het eind dient er voor iedere statistiek een uitgewerkte richtlijn kwaliteitszorg gereed te zijn. Volgens het businessplan dienen in 2000 meer dan de helft van de statistieken aan de richtlijn voor kwaliteitszorg te voldoen.

Regionale gegevens

Bij het publiceren van regionale gegevens dient er gestreefd te worden naar een complete dekking; er zouden geen regionale eenheden met ontbrekende gegevens dienen voor te komen. In de regionale statistieken ontbreken gegevens als er te weinig waarnemingen per regionale eenheid zijn om een zekere nauwkeurigheid te garanderen (betrouwbaarheid) of omdat er sprake is van een zeker ontsluitingsrisico (statistische beveiliging). Het merendeel van de statistieken wordt geproduceerd zonder of met een beperkt regionaal doel. Het resultaat kan een aantal regionale gebieden met ontbrekende gegevens (op de kaart herkenbaar als witte vlek) zijn. Het aantal witte vlekken zou verminderd kunnen worden door bij het ontwerp van de steekproef rekening te houden met regionaal gebruik door een minimum aantal waarnemingen per regionale eenheid te eisen. Voor de steekproef van het Regionaal Inkomensonderzoek is dit al het geval.

Geometrische gegevens

Verkeer en vervoer

Bij een aantal statistieken van het CBS is een duidelijke geometrische component aanwezig. Er valt hierbij te denken aan de statistieken van verkeer en vervoer. Op dit moment worden deze samengesteld op basis van schriftelijk enquêtes aan onder andere gemeenten. Als het Nationaal Wegenbestand (NWB) als gegevensbron kan worden gebruikt, zal de enquêtedruk voor de gemeenten worden verlaagd; het NWB fungeert hierbij als externe (integrale) registratie. Inzet van het NWB zal de kwaliteit van de gegevens verbeteren, omdat enerzijds de nauwkeurigheid zal toenemen en er anderzijds aangesloten wordt bij een mogelijke nationale standaard.

Bodemgebruik

De statistiek van het bodemgebruik worden samengesteld op basis van luchtfoto's die op het CBS worden gedigitaliseerd. De digitale kaart van het bodemgebruik, die inmiddels als zelfstandig product een plaats op de markt heeft verworven, fungeert hierbij als tussenproduct in het statistisch proces. Bij het samenstellen van deze statistiek wordt de inzet van TOP10vector van de Topografische Dienst als een externe registratie overwogen. Naar schatting 95% van de benodigde gegevens uit het TOP10vector is te herleiden. Afhankelijk van de prijsstelling zou dit kunnen leiden tot een verhoging van het rendement van het productieproces. Anderzijds kan de harmonisatie met de geometrie van TOP10vector bijgedragen aan een bepaalde mate van standaardisatie. Aansluiting bij een standaard vergroot de kwaliteit een bestand voor de gebruikers, omdat integra-

tieproblemen worden verminderd. Als het draagvlak groot genoeg is, zal TOP10vector een de facto standaard worden. Het tempo waarin dit gebeurt zal onder andere afhangen van de selecties die geleverd kunnen worden en de prijsstelling.

Wijken en buurten

Het CBS coördineert de landelijke wijk- en buurtindeling, waarbij de gemeenten zelf verantwoordelijk zijn voor de indeling van de gemeente in wijken en buurten. Digitale bestanden met de gemeente-, wijk- en buurtgrenzen vanaf 1993 zijn beschikbaar. De kwaliteit van deze grenzenbestanden wordt bepaald door de kwaliteit van het door de gemeenten aangeleverde materiaal, dat varieert van (grootschalige) digitaal grenzenbestanden tot stadsplattegronden waarop met een forse viltstift de grenzen zijn ingetekend. In het laatste geval wordt getracht aan de hand van topografische kaarten zo goed mogelijk te achterhalen wat de bedoeling is geweest. Het CBS is wat betreft de aanlevering van wijk- en buurtgrenzen volledig afhankelijk van wat de gemeenten aanleveren. Eisen kunnen niet worden gesteld; wel wordt er vaak rekening gehouden met de wensen als het de aanlevering van digitale bestanden betreft. Sinds 1994 worden de bestanden geharmoniseerd met het TOPgrenzen-bestand van de Topografische Dienst, dat de 'officiële' grenzen van de Nederlandse gemeenten bevat, zodat een bijdrage wordt geleverd aan de standaardisatie van de gemeentegrenzen.

Tot besluit

De nauwkeurigheid van de statistieken van het CBS wordt grotendeels bepaald door de hoeveelheid basisgegevens (de grootte van steekproeven is in het algemeen beperkt) en de kwaliteit van de gegevens die door de berichtgevers worden aangeleverd. Het controleren en corrigeren van de binnenkomende gegevens blijft uiteraard een essentieel onderdeel van het statistisch proces. De efficiëntie wordt vergroot door controles tijdens het inwinnen van gegevens en door de macro-editing van deze gegevens. De enquêtedruk wordt verlaagd door het combineren van enquêtes, elektronische gegevensinwinning (EDI) en het gebruik van externe registraties.

De kwaliteit van CBS-producten wordt verhoogd door gegevens sneller te publiceren, de samenhang tussen statistieken te vergroten en aan te sluiten bij standards. De invoering van kwaliteitszorg is in volle gang: er is een voorlopige richtlijn kwaliteitszorg en periodiek worden de statistieken aan kwaliteitsonderzoeken onderworpen. Uiterlijk eind 1998 is er een voor duurzaam gebruik uitgewerkte richtlijn kwaliteitszorg voor iedere statistiek. In 2000 zal meer dan de helft van de statistieken aan hun specifieke richtlijn moeten voldoen.

Het doel van bovengenoemde lopende en geplande acties is het verhogen van de kwaliteit van de CBS-producten met gelijktijdige verhoging van de efficiëntie van het statistisch proces vanwege budget-beperkingen) en verlaging van de enquêtedruk (opgelegde doelstelling).

Bepaling classificatienauwkeurigheid en kwaliteitsbeheersing van het Landelijke Grondgebruiksbestand van Nederland

H.A.M. Thunnissen

DLO-Staring Centrum
Instituut voor Onderzoek van het Landelijk Gebied (SC-DLO)
E-mail: h.a.m.thunnissen@sc.dlo.nl

Samenvatting

Het Landelijke Grondgebruiksbestand van Nederland (LGN-bestand) bestaat uit rasters van 25 x 25 m over heel Nederland. Van iedere rastercel is het grondgebruik bepaald door interpretatie van satellietbeelden en andere ondersteunende bestanden. Voor toepassingen van het bestand is het van groot belang te weten wat de classificatienauwkeurigheid van het bestand is. In theorie zijn voldoende methoden beschikbaar om de classificatienauwkeurigheid op statistisch betrouwbare wijze vast te stellen. In de praktijk blijken deze methoden niet altijd haalbaar en zal een evenwicht moeten worden gevonden tussen wat statistisch correct en praktisch haalbaar is. Een niet-representatieve bemonstering kan echter tot relatief grote fouten in de schatting van de classificatienauwkeurigheid leiden. Gezien de benodigde inspanning is vooral bij landbouwgewassen een representatieve bemonstering moeilijk haalbaar. Naast een pixelgewijze validatie worden de landbouwgewassen ook per stratum gevalideerd door vergelijking van het classificatieresultaat met de CBS-landbouwstatistieken.

Doordat het productieproces en de classificatienauwkeurigheid van het LGN-bestand sterk afhankelijk zijn van de kwaliteit van de beschikbare beelden en bestanden en van de expertise van degene die de classificatie uitvoert, kan het productieproces niet volledig worden vastgelegd, wat de kwaliteitsbeheersing aanzienlijk bemoeilijkt.

In de praktijk blijkt beschikbare informatie over de classificatienauwkeurigheid nog nauwelijks te worden gebruikt. Meer aandacht voor de kwaliteit van bestanden en nader onderzoek naar de voortplanting van fouten bij combinatie van verschillende bestanden en bij combinatie van bestanden met modellen is gewenst.

1. Inleiding

Het Landelijke Grondgebruiksbestand van Nederland (LGN-bestand) wordt vervaardigd door geïntegreerde toepassing van satellietbeelden en andere digitale, geografische gegevens. De kwaliteit van het LGN-bestand is van grote invloed op de resultaten van toepassingen van het bestand. In dit artikel zal vooral worden ingegaan op de bepaling van de classificatienauwkeurigheid van het LGN-bestand. Na een korte beschrijving van het LGN-bestand wordt nader ingegaan op de (theoretische) methoden voor de bepaling van de classificatienauwkeurigheid en de bruikbaarheid daarvan voor het LGN-bestand.

Het belang van kwaliteitsbeheersing van het LGN-bestand wordt nader toegelicht en er wordt aangegeven hoe de kwaliteit van het bestand zo goed mogelijk wordt gewaarborgd.

Het LGN-bestand is reeds in vele projecten toegepast. Aan de hand van ervaringen wordt aangegeven hoe de gebruiker met de kwaliteit van het LGN-bestand omgaat en welke eisen de gebruiker aan de kwaliteit stelt.

2. LGN-bestand

Het LGN-bestand wordt vervaardigd door geïntegreerde toepassing van satellietbeelden en andere digitale, geografische informatie. Vanwege de sterke vraag naar actuele grondgebruiks-informatie is in 1990 in Nederland voor het eerst een landsdekkend bestand van het grondgebruik (het zgn. LGN-bestand) beschikbaar gekomen. Hierbij is gebruik gemaakt van satellietbeelden uit 1986. Het LGN-bestand wordt regelmatig geactualiseerd. Momenteel wordt de derde versie van het LGN-bestand vervaardigd met satellietbeelden uit 1995 en 1997. Door verbetering van de classificatiemethodiek is de nauwkeurigheid van het bestand toegenomen (Thunnissen en Noordman, 1996).

Het LGN-bestand bestaat uit een raster met cellen van 25 x 25 m. Van iedere rastercel is het grondgebruik bepaald. In het LGN-bestand worden 5 hoofdklassen en 26 subklassen onderscheiden (tabel 1). Er wordt voor de hoofd- en subklassen gestreefd naar een minimale nauwkeurigheid van respectievelijk 90 en 70%. Voor de vervaardiging van het LGN-bestand worden voornamelijk beelden gebruikt van de Thematic Mapper (TM) aan boord van Amerikaanse LANDSAT satelliet. Wanneer de beschikbare satellietbeelden een classificatienauwkeurigheid van de landbouwgewassen van minimaal 70% niet toelaten zijn in het tweede LGN-bestand mengklassen gedefinieerd.

3. Remote sensing en landgebruik

De wijze waarop een voorwerp aan het aardoppervlak zonnestraling reflecteert, wordt spectrale signatuur genoemd. Wanneer verschillende oppervlakken een verschillende spectrale signatuur hebben, kunnen ze in principe met behulp van satellietbeelden van elkaar worden onderscheiden. Verschillende (landbouw)gewassen, die in een bepaalde periode dezelfde spectrale signatuur vertonen, kunnen in een andere periode vaak wel van elkaar worden onderscheiden (bijv. akkerbouwgewassen met uiteenlopende ontwikkelingsstadia). In zo'n situatie is het voor een grondgebruiksclassificatie van belang te beschikken over satellietbeelden uit verschillende perioden. We spreken dan over een multitemporele classificatie.

Satellietbeelden verschaffen alleen informatie over de bodembedekking (bijv. gras) en niet over het functioneel gebruik van de bodem (gras kan bijvoorbeeld worden gebruikt als cultuurgrasland, maar ook als golfterrein of stadspark). Voor veel toepassingen is een nadere onderverdeling naar gebruik echter wel gewenst. Daarnaast blijken veel grondgebruiksklassen elkaar geheel of gedeeltelijk spectraal te overlappen. Bovenstaande problemen kunnen grotendeels worden opgelost door naast satellietbeelden gebruik te maken van andere (digitale) geografische informatie. Het te classificeren gebied wordt dan onderverdeeld in kleinere en qua grondgebruik meer homogene gebieden of strata (stratificatie), waarbij ieder stratum afzonderlijk wordt geclassificeerd. Bij de vervaardiging van het LGN-bestand zijn 4 hoofdstrata onderscheiden: landbouw, bos en natuur, stedelijk bebouwd gebied en bebouwing buitengebied. Het landbouwstratum is nader onderverdeeld met behulp van de 66 CBS landbouwgebieden. Van deze gebieden zijn landbouwstatistieken beschikbaar. Naast automatische classificatietechnieken wordt steeds meer gebruik gemaakt van visuele interpretatie van de satellietbeelden.

<i>Hoofdklassen</i>	<i>Subklassen</i>
Landbouw	gras maïs aardappelen bieten granen overige landbouwgewassen kale (landbouw)grond glastuinbouw boomgaard bollen
Bos	loofbos naaldbos
(Open) natuurgebied	droge heide overig open begroeid natuurgebied kale grond in natuurgebied
Water	open (binnen)water (zoet) open (buiten)water (zout)
Bebouwd gebied	stedelijk bebouwd gebied bebouwing in buitengebied loofbos in bebouwd gebied naaldbos in bebouwd gebied bos met dichte bebouwing gras in bebouwd gebied kale grond in bebouwd buitengebied hoofdwegen en spoorwegen bebouwing in agrarisch gebied

Tabel 1. Onderscheiden grondgebruiksklassen in de derde versie van het LGN-bestand.

Sommige LGN-klassen, zoals kassen, boomgaarden, wegen en verspreide bebouwing in landelijk gebied zijn vaak moeilijk te classificeren met satellietbeelden. Vaak zijn deze klassen echter met een hoge nauwkeurigheid in andere bestanden aanwezig. Zo mogelijk worden deze klassen dan overgenomen in het LGN-bestand. Satellietbeelden blijken soms wel van nut te kunnen zijn bij een tussentijdse actualisatie van deze klassen. Ook financiële redenen kunnen het gebruik van satellietbeelden voor genoemde klassen soms aantrekkelijk maken.

Het LGN-bestand wordt op dit moment nog vooral gebruikt door landelijke en regionale overheden voor toepassingen op het gebied van milieu en ruimtelijke ordening (Thunnissen en Noordman, 1996).

4. Kwaliteit van het LGN-bestand

Kwaliteit is een ruim begrip. Onder de kwaliteit van geo-informatie wordt onder andere verstaan geometrische en thematische kwaliteit, volledigheid, actualiteit en consistentie. Dit laatste betekent onder andere dat gegevens binnen één bestand niet tot strijdige uitspraken mogen leiden (Molenaar, 1997). In dit artikel beperk ik me tot de thematische kwaliteit of classificatienauwkeurigheid.

De classificatienauwkeurigheid kan worden onderscheiden in plaatsgebonden en niet-plaats-gebonden nauwkeurigheid. Bij de plaatsgebonden validatie wordt van iedere pixel afzonderlijk zowel de nauwkeurigheid als de ligging beschouwd, terwijl bij de niet-plaatsgebonden validatie alleen de statistische verdeling van de klassen van de pixels in een gebied wordt beschouwd.

4.1 Vaststellen classificatienauwkeurigheid in theorie

De meest gebruikelijke wijze van plaatsgebonden validatie bestaat uit de selectie van representatieve pixels in het geclassificeerde beeld en vergelijking van deze pixels met referentiegegevens. Er zijn verschillende bemonsteringsschema's mogelijk:

- Een *willekeurige bemonstering* van de pixels over het gehele stratum garandeert een goede schatting van het classificatieresultaat in het betreffende stratum. Een probleem dat zich voordoet bij willekeurige bemonstering is dat grote klassen een groter aantal bemonsteringspunten bevatten dan kleine klassen. Om dit te voorkomen wordt vaak een gestratificeerde, willekeurige bemonstering uitgevoerd.
- Bij een *gestratificeerde, willekeurige bemonstering* wordt een minimum aantal pixels per 'stratum', i.c. klasse geselecteerd. Doordat de strata worden gevormd door de afzonderlijke thematische klassen van het classificatieresultaat zelf, kunnen de referentiegegevens pas worden verzameld na afloop van de classificatie. Hierdoor kan de verzameling van referentiegegevens die snel in de tijd veranderen (bijv. landbouwgewassen) problematisch worden.
- Om de bemonstering meer efficiënt te maken kan een *systematische bemonstering* worden uitgevoerd. Bij een systematische bemonstering worden de sample pixels op regelmatige afstanden in de ruimte geplaatst (bijv. snijpunten van een grid). Een nadeel van systematische bemonstering is dat wanneer de populatie enige periodiciteit bevat, de regelmatige afstand tussen de bemonsteringspunten kan resulteren in een niet-representatieve bemonstering (Congalton, 1988). Bij een onregelmatig grondgebruikspatroon zal deze situatie zich echter weinig voordoen.
- Als de geselecteerde bemonsteringspunten in het veld moeilijk bereikbaar zijn, waardoor de bemonstering duur en tijdrovend wordt, wordt tenslotte vaak de (*willekeurige*) *clusterbemonstering* gebruikt. Hierbij worden clusters van aan elkaar grenzende pixels in één keer bemonsterd. Zo kan snel over veel pixels informatie worden verkregen. Een probleem bij clusterbemonstering ligt bij de ruimtelijke correlatie tussen buurpixels. Omdat de reflectiewaarden van aangrenzende pixels niet onafhankelijk van elkaar zijn, kan de classificatienauwkeurigheid worden overschat. Wanneer de clusters echter niet te groot worden gekozen (10-25 pixels) dan blijkt de invloed van de onderlinge afhankelijkheid beperkt (Congalton, 1988).

Als de bemonstering correct is uitgevoerd kunnen de betrouwbaarheidsintervallen van de geschatte classificatienauwkeurigheid worden bepaald, er van uitgaande dat de classificatie-nauwkeurigheid kan worden beschreven met de binomiale kansverdeling (de classificatie van een pixel is goed of fout!). Hord en Brooner (1976) en Richards (1986) geven een methode waarmee het 95% betrouwbaarheidsinterval kan worden berekend, gegeven de bemonsteringsgrootte en het aantal correct geclassificeerde pixels in een monster.

Om er zeker van te zijn dat de validatie een nauwkeurige schatting van de classificatienauwkeurigheid geeft, is naast een representatieve bemonstering een minimum aantal pixels per klasse vereist. Van Genderen et al. (1978) en Rosenfield et al. (1982) hebben zich met dit vraagstuk bezig gehouden, uitgaande van de binomiale kansverdeling. Uitgaande van verschillende gezichtspunten kwamen zij tot verschillende waarden voor de

minimale bemonsteringsgrootte. Beide gezichtspunten hebben hun voordelen en in de praktijk wordt vaak een compromis gekozen tussen 30 en 60 te bemonsteren pixels per klasse (Richards, 1986). Congalton (1991) stelt echter dat vanwege het grote aantal pixels in een remote sensing beeld traditioneel denken over monsternames niet meer geoorloofd is. De keuze van de bemonsteringsgrootte wordt daarom steeds vaker bepaald door praktische overwegingen. Aan de hand van ervaring heeft Congalton een aantal vuistregels opgesteld. Hij beveelt een bemonsteringsgrootte aan van minimaal 50 pixels per klasse. Voor grote gebieden (meer dan 4000 km²) of bij een groot aantal klassen (meer dan 12) beveelt hij aan de minimale bemonsteringsgrootte per klasse te verhogen naar 75 tot 100 pixels per klasse.

4.2 Vaststellen classificatienauwkeurigheid LGN-bestand

Voor de validatie van het LGN-bestand is een onderscheid gemaakt tussen landbouwklassen en niet-landbouwklassen. Voor de validatie van de meeste niet-landbouwklassen wordt een gestratificeerde, systematische, willekeurige bemonstering uitgevoerd. De bemonstering wordt uitgevoerd per (groep van) afzonderlijke klasse(n) en de bemonsteringspunten (i.c. pixels) vallen samen met de snijpunten van de km-lijnen op de topografische kaart. De bemonsteringsgrootte bedraagt 100 per klasse en de selectie van de betreffende pixels gebeurt willekeurig.

Alleen de 'belangrijkste' klassen worden volgens bovenstaande wijze bemonsterd, te weten: loofbos, naaldbos, open natuurgebied, stedelijk bebouwd gebied en stedelijk groen (i.e. aggregatie van de 'groene' klassen uit hoofdklasse 'Bebouwd gebied', tabel 1). De referentieklassen van de geselecteerde pixels worden afgeleid van topografische kaarten en/of luchtfoto's. Redelijk homogene spectrale klassen, zoals water en kale grond, en klassen die (gedeeltelijk) zijn overgenomen uit andere bestanden, zoals kassen, hoofdwegen, spoorwegen, hoofdwaterlopen en bebouwing in buitengebied worden alleen gevalideerd op een kwalitatieve (d.w.z. globale vergelijking met topografische kaarten of luchtfoto's) of niet-plaatsgebonden wijze (d.w.z. vergelijking met statistieken).

Voor de bepaling van de classificatienauwkeurigheid van de landbouwklassen moeten gegevens worden verzameld in het veld. In de praktijk brengen een willekeurige bemonstering, een gestratificeerde, willekeurige bemonstering of een systematische bemonstering (3.1) van landbouwgewassen grote problemen met zich mee, omdat de te bemonsteren pixels vaak moeilijk bereikbaar zijn en het aantal pixels groot is. Dit maakt de bemonstering duur en tijdrovend. Bovendien is vaak slechts een beperkte tijd beschikbaar voor de bemonstering: de tijd tussen de keuze van de te gebruiken satellietbeelden en het einde van het groeiseizoen. Eventueel kunnen na afloop van het groeiseizoen nog referentiegegevens worden verkregen door ondervraging van boeren. De benodigde inspanning voor een gestratificeerde, willekeurige bemonstering van de landbouwklassen in het LGN-bestand wordt geschat op 150 á 200 mensdagen. Een willekeurige clusterbemonstering kan deze inspanning slechts in beperkte mate verminderen. Bij deze schatting is er van uitgegaan dat alleen in 'heterogene' strata een volledige bemonstering wordt uitgevoerd.

We zien daarom dat in de praktijk vaak overgegaan wordt tot het gebruik van referentiegebieden: aaneengesloten gebieden die geheel of gedeeltelijk worden bemonsterd. De bemonsteringseenheden bestaan niet uit (groepjes) pixels maar uit gehele percelen. Deze bemonsteringsstrategie is ook toegepast bij de validatie van de eerste versie van het LGN-bestand (Thunnissen et al., 1992). De classificatienauwkeurigheid wordt dan bepaald door het classificatieresultaat pixel voor pixel te vergelijken met een verrasterd

referentiebestand. Statistisch gezien is deze wijze van validatie niet juist, omdat door onderlinge afhankelijkheid tussen buurpixels de classificatienauwkeurigheid kan worden overschat. Bovendien kan de geometrische nauwkeurigheid van het classificatieresultaat en het referentiebestand grote invloed hebben op schatting van de classificatienauwkeurigheid. Bij deze wijze van bepaling van de classificatienauwkeurigheid wordt er namelijk van uitgegaan dat het classificatieresultaat en het referentiebestand perfect samenvallen. In de praktijk is dat meestal niet zo. Uit onderzoek van Verbyla en Hammond (1995) bleek dat bij een pixel per pixel vergelijking van het classificatieresultaat en het referentiebestand door een verschuiving van beide bestanden ten opzichte van elkaar met één pixel het classificatieresultaat werd onderschat met 15 tot 36% afhankelijk van het aantal klassen.

Figuur 1 toont een foutenkaart (i.e. ligging van de foutief geclassificeerde pixels) en de ligging van de gewasgrenzen van het referentiegebied Biddinghuizen (Thunnissen et al., 1992). Dit referentiegebied is gebruikt voor de validatie van de eerste versie van het LGN-bestand. Veel foutief geclassificeerde pixels liggen langs gewas- en andere grondgebruiksgrenzen. De randfouten worden behalve door de geometrische nauwkeurigheid ook bepaald door fouten bij het intekenen van gewasgrenzen tijdens veldwerk, digitaliseerfouten, omzetting van vectorbestanden naar rasterbestanden en foutief geclassificeerde randpixels. Het aantal randpixels kan tientallen procenten van het totale aantal pixels bedragen. Door de randpixels niet te gebruiken bij de validatie kunnen bovengenoemde problemen gedeeltelijk worden vermeden (Thunnissen et al., 1992).

Voor de validatie van de landbouwgewassen in de tweede en derde versie van het LGN-bestand zijn we daarom overgegaan op de zogenaamde 'gerichte' clusterbemonstering. Deze bemonstering vormt min of meer een combinatie van bemonstering van referentiegebieden en clusterbemonstering. Aan de hand van de satellietbeelden, CBS-landbouwstatistieken en topografische kaart worden 'representatieve' referentiegebieden geselecteerd. In deze gebieden worden gewassen gekarteerd van aan de weg grenzende percelen. Vervolgens worden van alle gekarteerde percelen clusters bemonsterd van 9 pixels uit het midden van het perceel. Daarnaast worden ad hoc ook clusters uit door derden beschikbaar gestelde referentiegebieden gebruikt. Vergelijking van de resultaten van de 'gerichte' clusterbemonstering met die van een systematische bemonstering in een studiegebied in Drenthe toonde een redelijke overeenkomst tussen beide methoden. We moeten ons bij gebruik van het LGN-bestand echter blijven realiseren dat er geen sprake is van een echte representatieve bemonstering. Bovendien blijkt dat ook bij deze wijze van bemonsteren in de praktijk nagenoeg altijd onvoldoende referentiegegevens beschikbaar komen. Momenteel wordt onderzocht of het mogelijk is om in de toekomst bij de vervaardiging van het LGN-bestand gebruik te maken van de gewasgegevens uit het perceelsregistratiesysteem PIPO (PerceelsIdentificatie en ProductieOmvang) van het Ministerie van LNV. Indien dat mogelijk is, zullen altijd voldoende referentiegegevens beschikbaar zijn.

De niet-plaatsgebonden validatie van de landbouwgewassen in het LGN-bestand vindt plaats door vergelijking van het classificatieresultaat per stratum met de CBS-landbouwstatistieken. De CBS-landbouwstatistieken bevatten echter netto betaalde oppervlakken, terwijl de landbouwklasse in het LGN-bestand nagenoeg het gehele landbouwgebied beslaat, inclusief smalle sloten, wegen en houtsingels, erven en bebouwing (in de derde versie van het LGN-bestand is agrarische bebouwing wel opgenomen). Bovendien zijn de, in economische zin, kleinere landbouwbedrijven niet in de CBS-landbouwstatistieken opgenomen. Dit 'extra' oppervlak in het LGN-bestand wordt grotendeels geclassificeerd als het aangrenzende grondgebruik, grasland of kale grond.

5. Kwaliteitsbeheersing LGN-bestand

Een hoge classificatienauwkeurigheid van het LGN-bestand is van groot belang voor de kwaliteit van toepassingen van het bestand. Om een hoge en constante classificatienauwkeurigheid te kunnen waarborgen is een goede beheersing van de kwaliteit van het productieproces van groot belang. De classificatienauwkeurigheid van het LGN-bestand is echter sterk afhankelijk van de kwaliteit van de beschikbare beelden en bestanden en



Figuur 1. Grondgebruiksgrenzen (A) en foutief geclassificeerde pixels (B) in referentiegebied Biddinghuizen.

van de expertise van degene die de classificatie uitvoert. Hierdoor kan het productieproces niet volledig worden vastgelegd, wat de kwaliteitsbeheersing aanzienlijk bemoeilijkt. De kwaliteitsbeheersing van het LGN-bestand bestaat uit het stellen van een aantal voorwaarden waaraan zoveel mogelijk moet worden tegemoet gekomen:

- Het LGN-bestand wordt alleen geactualiseerd wanneer voldoende beelden van goede kwaliteit beschikbaar zijn, die zijn opgenomen in geschikte perioden. Hierdoor is het niet meer nodig mengklassen te definiëren, zoals bij de tweede versie van het LGN-bestand op grote schaal gebeurd is, en kan voor iedere afzonderlijke landbouwklasse een minimale classificatienauwkeurigheid van 70% worden gegarandeerd. Gebruik van optimale beelden leidt niet alleen tot een kwaliteitsverbetering van het bestand maar ook tot een verlaging van de productiekosten.
- Klassen worden zoveel mogelijk overgenomen uit andere bestanden, indien de nauwkeurigheid van deze bestanden voldoende is. Uiteraard dient hierbij rekening te worden gehouden met eventuele auteursrechten.
- Met inachtneming van de kosten-kwaliteitsverhouding wordt de classificatiemethodiek regelmatig verbeterd. Zo is bij de vervaardiging van de derde versie van het LGN-bestand de rol van visuele interpretatie toegenomen ten koste van automatische classificatie zonder dat de kosten zijn toegenomen. Bij visuele interpretatie wordt naast kleur (spectrale waarden) gebruik gemaakt van verschillende karakteristieken van objecten aan het aardoppervlak, zoals grootte, textuur, vorm, patronen en context. Kenmerkend voor visuele interpretatie is dat allerlei informatie uit de naaste omgeving van de betreffende pixels bij de interpretatie wordt betrokken. De huidige hard- en software maken het mogelijk meerdere beelden tegelijk bij de classificatie te betrekken. Visuele interpretatie van satellietbeelden resulteert over het algemeen in een aanzienlijke toename van de classificatienauwkeurigheid vergeleken met automatische classificatie.
- Bij het vertrek van medewerkers wordt de opgedane expertise zo goed mogelijk overgedragen. Dit gebeurt onder andere door een intensieve inwerkperiode onder begeleiding van de vertrekkende medewerker.

6. Gewenste nauwkeurigheid LGN-bestand door gebruikers

Kwaliteit is wat de gebruiker wil. De vraag is echter of de meeste gebruikers wel weten wat ze willen. Wanneer expliciet naar de gewenste nauwkeurigheid wordt gevraagd, wordt veelal geantwoord: "zo nauwkeurig mogelijk". De leverancier van gegevensbestanden mag hier geen misbruik van maken en zou moeten meedenken over de gewenste nauwkeurigheid van gegevens voor een bepaalde toepassing. De toepassing (schaal, gewenste klassen, combinatie met en nauwkeurigheid van andere bestanden, ...) bepaalt uiteindelijk of een bestand nauwkeurig genoeg is.

De ervaring leert dat de gebruiker eerder vertrouwen heeft in een bestand als het bestand er 'netjes' uitziet, bijvoorbeeld weinig ruis en herkenbare vormen. Ook moeten geen grote inconsistenties in het bestand voorkomen. De gebruiker valt vaak over één 'heidepixel' in een landbouwgebied, terwijl een 'bietenpixel' in een aardappelperceel niet als storend wordt ervaren. De gebruiker stelt hoge eisen aan de bewerkbaarheid/toepasbaarheid van de gegevens. Mengklassen worden vaak niet geaccepteerd en gehercoördineerd naar zuivere klassen, ondanks de geringere nauwkeurigheid (< 70%).

Veel gebruikers werken met de digitale bestanden zonder ooit een afbeelding te hebben gezien, omdat men niet beschikt over de daarvoor benodigde apparatuur. Een afbeelding geeft over het algemeen een goede eerste indruk van (de kwaliteit van) het bestand. Momenteel worden daarom hard copies van (een deel van) het LGN-bestand meegeleverd. Eventuele beschikbare informatie over de nauwkeurigheid van het bestand wordt in de praktijk overigens nauwelijks gebruikt. Dit wordt mede veroorzaakt door de beperkte kennis van foutenvoortplanting bij de combinatie van meerdere bestanden en bij de combinatie van bestanden met modellen. Bovendien is van veel digitale geografische bestanden geen informatie over de nauwkeurigheid beschikbaar.

In het digitale tijdperk, waarin steeds meer mogelijkheden ontstaan voor geïntegreerd gebruik van geodata, wordt de gebruiker zich veel sneller bewust van de (geometrische) kwaliteit van bestanden. Een gemeenschappelijke ondergrond (Top10vector?) zou een efficiënt gebruik van bestanden sterk bevorderen.

7. Discussie en conclusies

In theorie zijn voldoende methoden beschikbaar om de classificatienauwkeurigheid van het LGN-bestand op statistisch betrouwbare wijze vast te stellen. In de praktijk blijken deze methoden echter vaak niet haalbaar en zal een evenwicht moeten worden gevonden tussen wat statistisch correct en praktisch haalbaar is. Gezien de benodigde inspanning is vooral bij landbouwgewassen een representatieve bemonstering moeilijk haalbaar. Niettemin kan met de zogenaamde 'gerichte' clusterbemonstering een redelijk inzicht in de classificatienauwkeurigheid worden verkregen. We moeten ons bij gebruik van het LGN-bestand echter blijven realiseren dat er geen sprake is van een echte representatieve bemonstering. Bovendien blijkt dat ook bij deze wijze van bemonsteren in de praktijk nagenoeg altijd onvoldoende referentiegegevens beschikbaar komen. Momenteel wordt onderzocht of het mogelijk is om in de toekomst bij de vervaardiging van het LGN-bestand gebruik te maken van de gewasgegevens uit het perceelsregistratiesysteem PIPO (Perceelsidentificatie en ProductieOmvang) van het Ministerie van LNV. Indien dat mogelijk is, zullen altijd voldoende referentiegegevens beschikbaar zijn. De CBS-landbouwstatistieken lenen zich over het algemeen goed voor een niet-plaatsgebonden validatie van de landbouwgewassen.

Om een hoge en constante classificatienauwkeurigheid te kunnen waarborgen is een goede beheersing van de kwaliteit van het productieproces van groot belang. Doordat het productieproces en de classificatienauwkeurigheid van het LGN-bestand sterk afhankelijk zijn van de kwaliteit van de beschikbare beelden en bestanden en van de expertise van degene die de classificatie uitvoert, kan het productieproces niet volledig worden vastgelegd, wat de kwaliteitsbeheersing bemoeilijkt. Niettemin is de kwaliteit van het LGN-bestand in de loop van de tijd aanzienlijk toegenomen en redelijk gewaarborgd. Naast het gebruik van optimale satellietbeelden speelt hierbij vooral de toename van visuele interpretatie van de satellietbeelden een belangrijke rol.

Bij de gebruikers bestaat vaak nog weinig besef over de gewenste kwaliteit van bestanden. Eventuele beschikbare informatie over de classificatienauwkeurigheid wordt in de praktijk nog nauwelijks gebruikt. Meer aandacht voor de kwaliteit van bestanden en nader onderzoek naar de voortplanting van fouten bij combinatie van verschillende bestanden en bij combinatie van bestanden met modellen is gewenst. Een gemeenschappelijke ondergrond van geografische bestanden zou een efficiënt gebruik van bestanden sterk kunnen bevorderen.

Literatuur

- Congalton, R.G., 1988. A comparison of sampling schemes used in generating error matrices for assessing the accuracy of maps generated from remotely sensed data. *Photogrammetric Engineering and Remote Sensing* 54: 593-600.
- Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment* 37: 35-46.
- Genderen, J.L. van, B.F. Lock and P.A. Vass, 1978. Remote sensing: statistical testing of thematic map accuracy. *Remote Sensing of Environment* 7: 3-14.
- Hord, R.M. and W. Brooner, 1976. Land use map accuracy criteria. *Photogrammetric Engineering and Remote Sensing* 42: 671-677.
- Molenaar, M. Data: veelzijdigheid van processing en integratie. Lezingenbundel Geodesia-congres '97. Geodesie, professioneel en veelzijdig. 21-24 oktober 1997. Jaarbeurs Congressentrum Utrecht. Delft University Press. Delft. The Netherlands.
- Richards, J.A., 1986. *Remote sensing digital image analysis; An introduction*. Berlin, Springer-Verlag.
- Rosenfield, G.H., K. Fitzpatrick-Lins and H.S. Ling, 1982. Sampling for the thematic map accuracy testing. *Photogrammetric Engineering and Remote Sensing* 48: 131-137.
- Thunnissen, H.A.M. en E. Noordman, 1996. Classification methodology and operational implementation of the land cover database of the Netherlands. Wageningen, DLO-Staring Centrum. Report 124.
- Thunnissen, H.A.M., R. Olthof, P. Getz en L. Vels, 1992. Grondgebruiksdata van Nederland vervaardigd met behulp van Landsat Thematic Mapper opnamen. Wageningen, DLO Staring Centrum. Rapport 168.
- Verbyla, D.L. en T.O. Hammond, 1995. Conservative bias in classification accuracy assessment due to pixel-by-pixel comparison of classified images with reference grids. *Int. J. Remote Sensing*, 16: 581-587.

Kwaliteit(en) kadastrale kaart

J. IJsselstein

Kadaster Groningen

Het Kadaster heeft een lange traditie waar het gaat om het besteden van zorg aan de kwaliteit van zijn producten en processen. Meten we die kwaliteit af aan bijvoorbeeld het zeer geringe aantal ontvangen klachten of door rechthebbenden aangespannen gerechtelijke procedures in verhouding tot het grote aantal transacties, dan is de kwalificatie 'goed' voor het kadastrale werk en product zeker terecht.

Een goede kwaliteit is echter meer dan een zeer laag klachtenpercentage. Tevens is de betekenis van kwaliteit aan wijziging onderhevig als gevolg van veranderingen in de maatschappij of van veranderingen binnen bedrijven. Een belangrijk voorbeeld is de overgang van analoog naar digitaal werken. Wellicht zelfs als gevolg daarvan kan geconstateerd worden dat men geneigd is veel meer aandacht te (gaan) schenken aan kwaliteit en aan een verdergaande verzakelijking.

Gerelateerd aan de kwaliteit(en) van de digitale kadastrale kaart wordt in deze bijdrage ingegaan op verschillende kwaliteitsaspecten die met die constatering te maken hebben.

Inhoud

De kadastrale kaart is een kaart waarop de eigendomssituatie is afgebeeld. De inhoudselementen van de kaart (figuur 1) zijn: eigendomsgrenzen (perceelsgrenzen), perceelnummers, gebouwen inclusief huisnummers en namen van wegen en straten. Kortom een qua inhoud eenvoudige kaart. Een belangrijk gegeven van de kaart is dat die inhoud in geheel Nederland hetzelfde is én dat het Kadaster medio 1998, in reactie op nadrukkelijk geuite wensen van de afnemers, voor geheel Nederland de kadastrale kaart digitaal beschikbaar zal hebben.

Het Kadaster heeft met die digitale kadastrale kaart de beschikking over een geometrisch en thematisch sterk gestructureerd basisproduct, grootschalig én objectgericht. Via het unieke perceelnummer zijn met de kadastrale kaart, met name dankzij de automatisering met bijbehorende koppelings- en analysemogelijkheden, op eenvoudige wijze vele gegevens te visualiseren en te analyseren.

Ontstaan er door het digitaal beschikbaar zijn dus veel meer toepassingsmogelijkheden voor de kadastrale kaart, tegelijkertijd vraagt het kwaliteitsaspect veel nadrukkelijker en indringender de aandacht dan in het analoge tijdperk. Logisch overigens, want met een druk op de knop kan men in een digitale omgeving allerlei bewerkingen uitvoeren, die



Figuur 1. Kadastrale kaart.

met analoge kaarten niet of slechts met zeer veel inspanning kunnen worden uitgevoerd. Voortdurend moet men zich echter afvragen of die bewerkingen wel zijn toegestaan en de resultaten wel een juiste analyse mogelijk maken?

Toch wil het Kadaster, met die vragen in het achterhoofd, de vele toepassingsmogelijkheden van de digitale kadastrale (basis)kaart breed ondersteunen. Daarom moet terdege aandacht worden geschonken aan aspecten als nauwkeurigheid, actualiteit, toegankelijkheid en toepassingsmogelijkheden.

Nauwkeurigheid

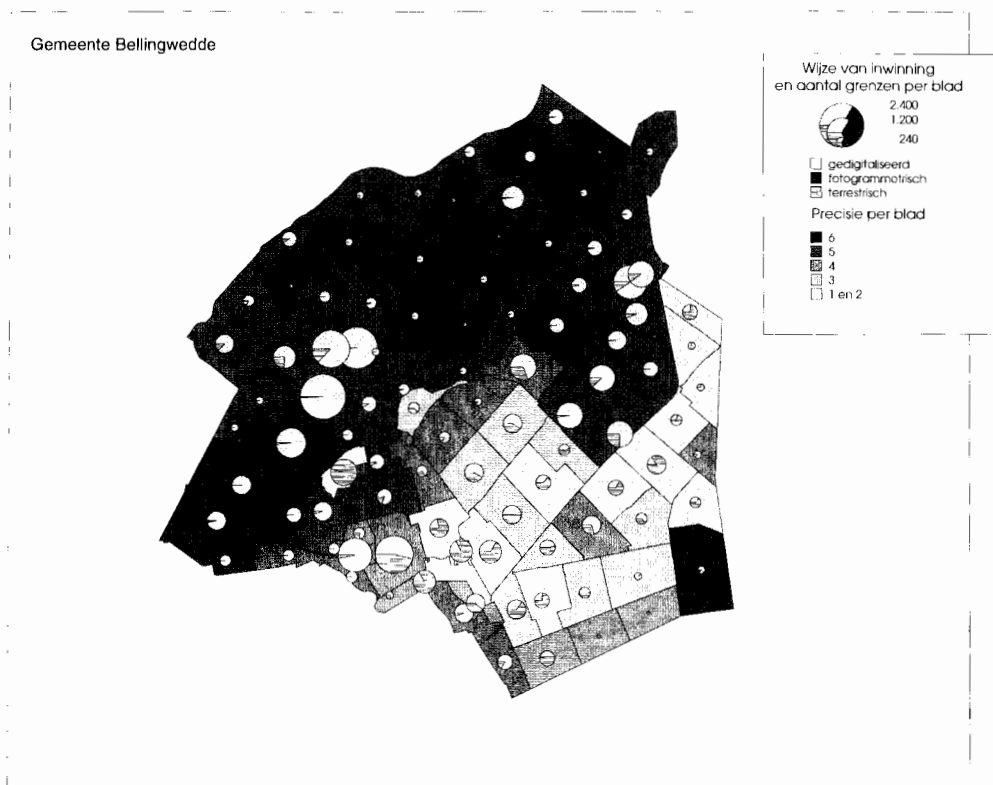
Op grond van wensen van afnemers is het Kadaster bezig om er voor te zorgen dat er medio 1998 een landsdekkende digitale kadastrale kaart aanwezig is. Ten opzichte van eerdere plannen waarbij het compleet digitaal zijn van de kadastrale kaart pas een aantal jaren na 2000 zou worden gerealiseerd betekent dit een behoorlijke versnelling. Een versnelling met consequenties, immers door de vele manieren waarop de digitalisering is uitgevoerd is de kwaliteit van dit digitale kadastrale bestand niet homogeen. In de komende 5 jaren zal door het Kadaster dan ook hard aan de verbetering van dat bestand worden gewerkt. Die verbetering behelst naast een betere aansluiting aan het RD-net ook het beter afstemmen van de kadastrale grenzen op de in de GBKN voorkomende topografie. Richtinggevend en handleiding voor dat zogenoemde kaartverbeteringsproces is sinds 1996 de nieuwe Handleiding voor de Technische Werkzaamheden van het Kadaster, veel vaker kortweg de HTW genoemd. Die HTW richt zich met name op de beheersing van de geometrische kwaliteit, maar staat in een afzonderlijk hoofdstuk ook stil bij het kaartverbeteringsproces. Veel aandacht gaat binnen het Kadaster momenteel uit naar het ontwikkelen van het benodigde instrumentarium voor dat verbeteringspro-

ces én naar mogelijkheden om de gebruikers van de digitale kadastrale kaart zo goed en direct mogelijk te informeren over de kwaliteit van die kaart.

Het realiseren van een homogene kwaliteit van het digitale kadastrale kaartproduct is één kant van de zaak. Daarnaast moet vanzelfsprekend ook het bijhoudingsproces dusdanig gestructureerd zijn dat de instandhouding van de gewenste en beschreven kwaliteit gegarandeerd is. De opdracht die de HTW daarvoor geeft is er voor te zorgen dat een meting in zijn omgeving wordt ingemeten, waarbij zoveel mogelijk correlatie met vorige en eventueel latere metingen wordt gewaarborgd. De HTW propageert in dat opzicht, vanuit met name het betrouwbaarheidsaspect, het gebruik van bovengrondse geselecteerde natuurlijke grondslagpunten. Overigens vormt dit punt in veel intern gegeven HTW-opleidingen nog vaak een stevig punt van discussie.

Vanzelfsprekend dient tenslotte voordat intern of extern vervaardigde digitale mutatiebestanden worden verwerkt in het digitale geometrische basisbestand, fiattering, door inschakeling van binnen het Kadaster beschikbare toetsmethodieken, van die bestanden plaats te vinden.

Om nu de medewerkers van het Kadaster zo goed mogelijk uit te rusten met de HTW-filosofie (een systematische beheersing van de geometrische kwaliteit) hebben in 1997 reeds veel medewerkers een als goed beoordeelde HTW-cursus gevolgd. Aan het management is het nu de taak en de zorg om er op toe te zien dat die HTW-filosofie ook in de praktijk van alledag op de juiste manier wordt uitgevoerd. Overigens wordt het ex-



Figuur 2. Wijze van inwinning en gemiddelde precisieaanduiding per blad voor de gemeente Bellingwedde.

terne belang van de HTW, gelet op de opleidingsverzoeken van gemeenten, nutsbedrijven en ingenieursbureaus, ook buiten het Kadaster steeds meer onderkend.

Bij dit alles heeft het Kadaster met een redelijke uitgangspositie te maken, omdat al jaren geleden is onderkend dat niet alleen het registreren van coördinaten, maar ook van de kwaliteit van die coördinaten van belang is. In LKI (het Landmeetkundig en Kartografisch Informatiesysteem van het Kadaster) zijn daartoe attributen gedefinieerd voor het beschrijven van de metrische kwaliteit in de vorm van 'precisie' en 'idealisatie'. Daarnaast kunnen ook de attributen 'wijze van inwinning' en 'bronvermelding' waardevolle informatie geven over de digitale totstandkoming van de kadastrale grenzen.

Indien op een juiste wijze gehanteerd kan in het kaartverbeteringsproces, wellicht een dankbaar gebruik worden gemaakt van deze kwaliteitsinformatie. Momenteel wordt daar nog het nodige onderzoek naar gedaan.

Zo is als voorbeeld door het Kadaster Groningen op basis van de ter beschikking staande attribuutgegevens op kaartbladniveau in figuur 2 informatie over de precisie en over de wijze van inwinning gevisualiseerd.

Om een kwalitatief goede digitale kadastrale kaart aan de klanten te kunnen leveren heeft het Kadaster dus de opdracht om te zorgen voor een homogeen product met een juiste en consequente invulling van de reeds genoemde attributen.

Een heel ander, maar zeker zo belangrijk aspect is dat een juiste en consequente invulling van de attributen mogelijkheden biedt om de geometrische kwaliteit van de bestanden op een, voor de gebruikers, meer begrijpelijke wijze in beeld te brengen. Dat dit zeer wenselijk is blijkt vaak in contacten met de klanten. Die klanten hebben veelal wel een aardig idee wat de kwaliteit is van een analoge kaart, maar zodra er gewerkt gaat worden met vectorbestanden heeft men toch heel vaak de neiging de onnauwkeurigheid van de gegevens te vergeten.



Figuur 3. Visualisatie precisie per grens met behulp van strookbreedtes.

Met in het achterhoofd het visualiseren van die onnauwkeurigheid is in Groningen_de volgende kwaliteitsinformatie voor een kadastraal kaartblad per grens in beeld gebracht.

In figuur 1 is een normale kadastrale kaart gepresenteerd in zijn analoge verschijningsvorm. Hoe nauwkeurig of onnauwkeurig een grens is, is op grond van deze presentatie niet te bepalen. In figuur 3 krijgt de gebruiker door het gebruik van aan de precisie gerelateerde strookbreedtes wel een duidelijk nauwkeurighedsbeeld gepresenteerd.

Met name deze wijze van presenteren spreekt in klantcontacten zeer tot de verbeelding. Regelmatig kan worden vernomen dat men zich op grond van dit eenvoudige plaatje pas realiseert dat het niet geoorloofd is zomaar maten te ontlenen aan een vectorbestand, maar dat men zich daarbij goed bewust moet zijn van de onnauwkeurigheid van de betreffende gegevens. Een dataleverancier dient zich dan ook goed te realiseren dat met betrekking tot die onnauwkeurigheid, anderen spreken ook wel van schijnnauwkeurigheid, er een informatieopdracht is naar de afnemers.

Om dat goed te kunnen doen is het niet alleen noodzaak om te zorgen dat het te leveren product aan vastgelegde eisen voldoet, maar ook om een goed inzicht te verkrijgen in waarvoor de gegevens zoal worden gebruikt. Om hier, naast uiteraard andere aandachtsgebieden, inzake de digitale kadastrale kaart meer inzicht in te verkrijgen zijn er vanaf begin 1998 binnen het Kadaster accountmanagers werkzaam.

Actualiteit

Een heel ander belangrijk gegeven van de kadastrale kaart is dat de bijhouding gegarandeerd is. Echter het Kadaster kan niet volstaan met een gegarandeerde bijhouding alleen. Afnemers c.q. klanten wijzen ons wat dat betreft met nadruk op het belang van twee zaken.

Allereerst de noodzaak om de actualiteit van de kadastrale geometrie in overeenstemming te brengen met de kadastrale administratie, hiermee wordt bedoeld dat ook in LKI direct zichtbaar moet zijn dat er mutaties in de eigendomsbegrenzing hebben plaatsgevonden.

Ten tweede hebben onze klanten moeite met de situatie dat ze als vervreemders en verkrijgers van nieuwe percelen vooruit moeten betalen voor meetwerk dat gemiddeld pas na zo'n 15 maanden wordt afgerond.

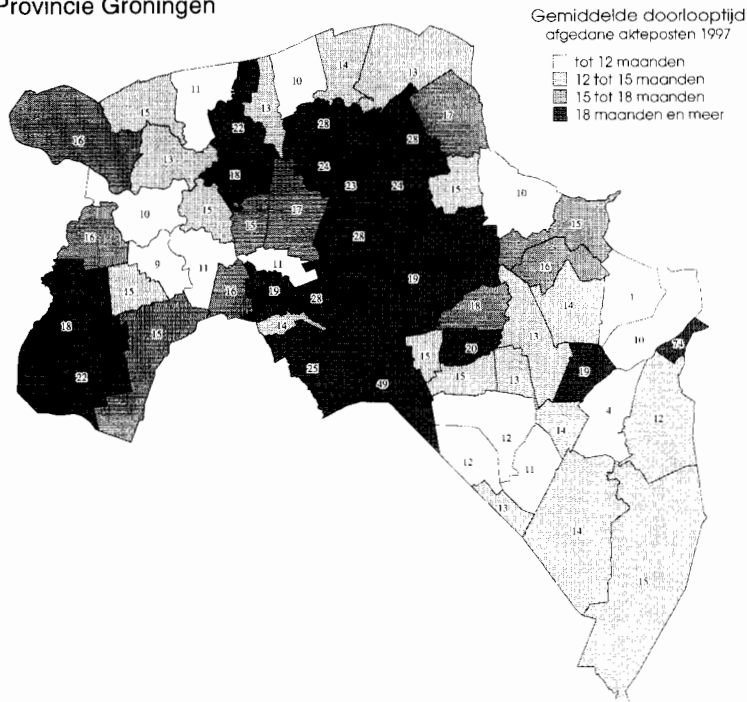
Het eerste probleem, het met elkaar in overeenstemming brengen, c.q. het consistent maken van geometrie en administratie, is momenteel in zoverre opgelost dat in LKI de mogelijkheid aanwezig is om door middel van een automatische onderstreping van het perceelnummer aan te geven dat er sprake is van deelpercelen (zie perceel 2642 in figuur 1), ofwel dat er wijzigingen in de eigendomsbegrenzing zijn opgetreden.

Bij raadpleging wordt men er dus op gewezen dat er wijzigingen hebben plaatsgevonden. Uiteraard wil het Kadaster nog graag een stap verder gaan, maar dat betekent dat het notariaat reeds bij de overdrachts-akte een kaart dient bij te voegen, waarop de nieuwe toekomstige grenzen zijn aangegeven. Om dit zo goed mogelijk ingang te doen vinden is ook weer een opdracht voor de al eerder genoemde accountmanagers.

Voor het gesignaleerde consistentieprobleem is dus vooralsnog een afdoende oplossing gerealiseerd.

Ook het tweede probleem, binnen het Kadaster ook wel de doorlooptijd van de akteposten genoemd, heeft volop de aandacht. Voor een goed begrip, onder een aktepost wordt

Stand van zaken Akteposten per kadastrale gemeente
Provincie Groningen



Figuur 4. Gemiddelde doorlooptijd akteposten per gemeente in Groningen.

verstaan; de meting van een nieuwe grens die is ontstaan ten gevolge van de overdracht van een gedeeltelijk perceel.

Ook deze informatie is door koppeling van de betreffende administratieve gegevens aan de kadastrale kaart eenvoudig te visualiseren.

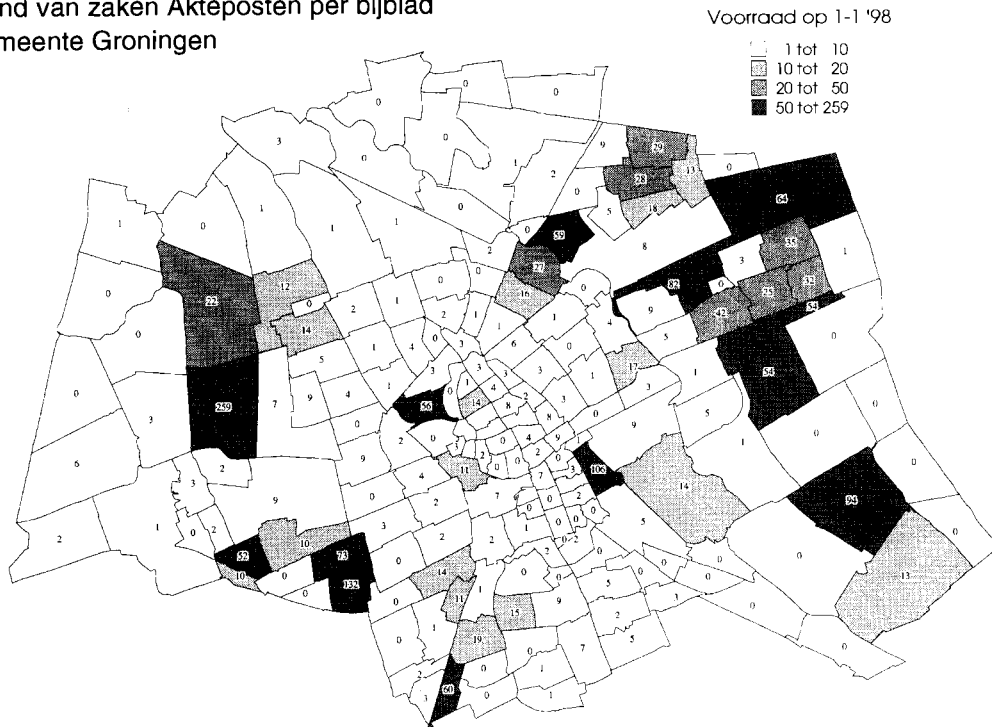
Figuur 4 geeft duidelijk aan dat er in Groningen voor bepaalde gebieden een gemiddeld hogere doorlooptijd voor de akteposten was dan 15 maanden. Overigens is dat voor het overgrote deel te verklaren doordat vanuit efficiencyoverwegingen is besloten voor bepaalde gebieden de aktepostenproductie op te houden, vanwege de in die gebieden lopende GBKN-vervaardiging. Na afronding van de GBKN-vervaardiging is het vervolgens mogelijk met minder meetwerk en dus goedkoper de akteposten af te doen. Op het overzicht is duidelijk te zien dat het betreffende gebied in de middenstreek van Groningen ligt. Het betrof een GBKN-project van ruim 50.000 ha.

Overigens zijn overzichten zoals gepresenteerd in figuur 4 ook zeer informatief voor de uitvoerende medewerkers, die allen de verantwoordelijkheid voor één of meer gemeenten toegewezen hebben gekregen. Vanzelfsprekend is het ook nuttige sturingsinformatie voor het management. Duidelijk is te zien waar de doorlooptijd niet aan de eisen voldoet en er dus bij afwezigheid van afdoende argumenten, conform stand beleid, steviger aan getrokken dient te worden.

Met de beschikbaar staande gegevens kan de informatie ook op een andere wijze (figuur 5) gepresenteerd worden.

De gepresenteerde informatie in de figuren 4 en 5 is gebaseerd op een doorlooptijd voor de akteposten van gemiddeld 15 maanden. Om aan de wensen van de klanten zo goed mogelijk tegemoet te komen wordt binnen het Kadaster echter zeer nadrukkelijk nagedacht over een halvering van die doorlooptijd naar 7,5 maand.

Stand van zaken Akteposten per bijblad Gemeente Groningen



Figuur 5. Werkvoorraad akteposten per kadastrale kaart in de gemeente Groningen.

Een belangrijk aspect tenslotte met betrekking tot zówel de inhoud áls de nauwkeurigheid én de actualiteit van de kadastrale kaart is dat belanghebbenden na verwerking van een uitgevoerde aktepostenmeting een kennisgeving krijgen toegezonden waarop zij al dan niet kunnen reageren. Hetzij met een vraag om nadere informatie, hetzij in de vorm van een klacht. Óók deze vragen en klachten zijn zeer waardevol en helpen het Kadaster om de kwaliteit van de kadastrale registratie op het gewenste niveau te houden.

Toegankelijkheid

Bij het kwaliteitsonderdeel toegankelijkheid zijn een aantal bepalende onderdelen te onderscheiden, nl. de aanschafkosten van het product, de leveringsmogelijkheden en de leesbaarheid van het product.

Allereerst de aanschafkosten van de digitale kadastrale kaart. Deze zijn in de afgelopen jaren spectaculair gedaald. Uitgaande van het jaar van verzelfstandiging van het Kadaster, 1994, zijn de kosten voor de eerste aanschaf met ruim 60% gedaald en de daarop volgende jaarlijkse abonnementskosten met zo'n 30%. Toch moet worden geconstateerd dat m.n. voor potentiële afnemers met een uitgebreider verzorgingsgebied er signalen worden ontvangen dat de abonnementskosten toch nog als een hoge kostenpost worden ervaren. Kortom, een blijvende opdracht om daar waar mogelijk de tarieven in neerwaartse zin bij te stellen.

De leveringsmogelijkheden van de digitale kadastrale kaart zijn velerlei. Uiteraard is levering van bestanden op verschillende media mogelijk, waarbij zowel flexibiliteit in de tijd, in de omvang en in de inhoud mogelijk is. Maatwerk is dan ook steeds vaker aan de orde.



Figuur 6. Combinatie luchtfoto met kadastrale kaart.

Ook is raadpleging van het digitale kadastrale kaartbestand bij bijvoorbeeld gemeenten met de zogenoemde (kadastrale) gemeentebalie mogelijk. Uiteraard blijft tevens de analoge afname een mogelijkheid.

Heeft men een digitaal product afgenomen dan is het vervolgens nog mogelijk om met behulp van mutatiebestanden, inhoudende de was/wordt-informatie, een bepaalde actualiteit te verzekeren.

Als laatste onderdeel van het aspect toegankelijkheid blijft de leesbaarheid over. Door zijn eenvoudige inhoud en het duidelijke thema is de leesbaarheid in feite eenduidig. Men weet waarvoor men de kadastrale kaart raadpleegt. Dat neemt niet weg dat tegelijkertijd moet worden geconstateerd dat toch behoorlijk wat niet kadastraal geschoolden moeite hebben om zich te oriënteren. In dat opzicht is het dan ook heel informatief wanneer de kadastrale kaart gecombineerd wordt afgebeeld (figuur 6) met bijvoorbeeld een luchtfoto.

Meteen is duidelijk waar men is, maar ook andere waardevolle informatie komt door de combinatie beschikbaar. Zo kan met bijvoorbeeld de luchtfoto de volledigheid van de bebouwing op de kadastrale kaart gecontroleerd worden. Duidelijk is in het midden van figuur 6 te zien dat op de kadastrale kaart nog een huis staat afgebeeld dat in het terrein allang tot een parkeerplaats is gebombardeerd.

Toepassingsmogelijkheden

Een zeer belangrijk kwaliteitsaspect van de digitale kadastrale kaart is te vinden in de vele toepassingsmogelijkheden van de digitale kadastrale kaart. Immers, uit de vele toepassingsmogelijkheden van de digitale kadastrale kaart spreekt de grote basiskracht van het product. Enkele voorbeelden:



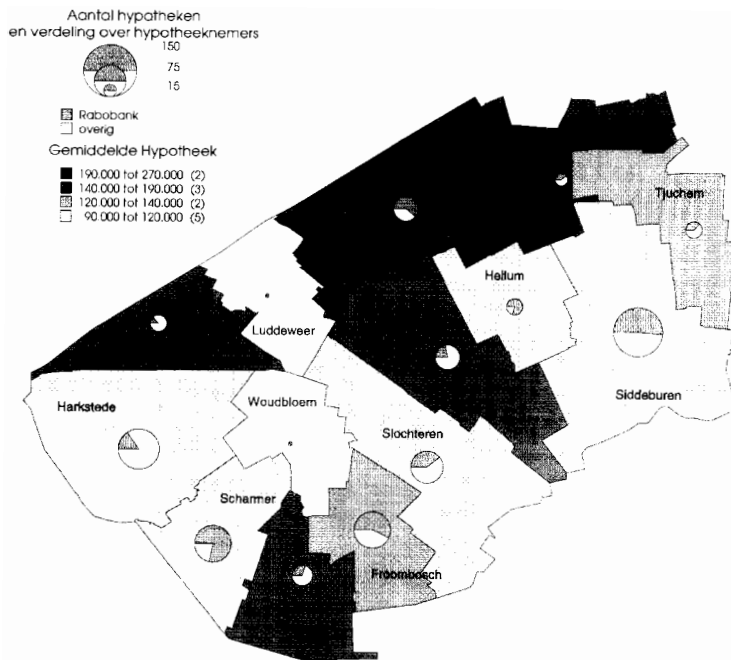
Figuur 7. Koopsommen in Slochteren.

- Door de koppeling van geometrische en administratieve perceelsgegevens kunnen door het aanwijzen van een perceel in de kadastrale kaart direct gegevens over de rechthebbende worden getoond (zie figuur 1, perceel 2653);
- De in figuur 6 gepresenteerde combinatie van luchtfoto en kadastrale kaart geeft bijvoorbeeld ook direct de mogelijkheid om zonder veldbezoek de eigenaar van een stuk grond op te sporen waarop onrechtmatig is gebouwd;
- Een vergelijkbare combinatie is die van bijvoorbeeld een satellietbeeld met de kadastrale kaart, waardoor de mogelijkheid bestaat om de informatie uit het satellietbeeld te relateren aan een kadastraal perceel en de eigenaar;
- Ook allerlei andere, niet kadastraal, administratieve informatie is, indien er bijvoorbeeld een relatie is te leggen naar het perceelnummer of adres, met behulp van de kadastrale kaart eenvoudig te visualiseren;

Een kadastraal voorbeeld daarvan is figuur 7 waarin door koppeling van administratieve gegevens en kadastrale kaart door een bepaalde bevraging in een aantal klassen de koopsommen in het dorp Slochteren in beeld zijn gebracht. Het zal duidelijk zijn dat het presenteren van dergelijke informatie met het oog op de privacy gevoelig ligt.

- Door de vele koppelingsmogelijkheden van het kadastraal perceelnummer is het zeer eenvoudig via aggregatie informatie op een niet tot individuen herleidbaar niveau te presenteren.

Een voorbeeld daarvan is de weergave in figuur 8 van de gemiddelde hypotheeksom voor de dorpen binnen de gemeente Slochteren. Eveneens is in figuur 8 weergegeven wat het marktdeel is van een grote hypotheeknemer.



Figuur 8. Aggregatie koopsommen/hypotheek.

Aggregeren dient, met het oog op de kwaliteit van de te presenteren informatie, echter wel doordacht te gebeuren, waarbij niet alleen rekening moet worden gehouden met extremen, aantallen, registratiedata enz., maar - heel belangrijk - met name ook met de vraagstelling van

de gebruiker. In ieder geval is het, gelet op de betrouwbaarheid van de informatie, belangrijk om te weten hoe de informatie is geaggregeerd, welke de basisgegevens zijn en wat de kwaliteit is van die basisgegevens. Informatie over gegevens die gelet op de kadastrale kaart steeds meer en steeds beter beschikbaar komt.

- Een belangrijke toepassingsmogelijkheid van de digitale kadastrale kaart, met name in kwalitatief opzicht, is dat de digitale kadastrale kaart zeer goed bruikbaar is voor de kwaliteitscontrole van andere gegevensverzamelingen.

Door aan het perceelnummer gerelateerde gegevens te visualiseren in de kadastrale kaart is het mogelijk door het bedrijf heen te kijken en tot nu toe niet gesignaleerde fouten op te sporen. Immers gepresenteerd in een kaartbeeld zie je plotseling wat je anders niet ziet!

Tenslotte

Door allerlei maatschappelijke en technologische ontwikkelingen, maar ook door ontwikkelingen op het gebied van de regelgeving neemt de behoefte aan goede geografische data toe. Een vastgoedinformatiedienst moet daarom zorgen voor een duidelijke toelichting op de kwaliteit van het geleverde product, moet snel en flexibel op klantwensen kunnen inspelen, moet leveren conform gemaakte afspraken, moet een goede communicatie met de klanten onderhouden en moet de klanten tijdig informeren over voor die klanten van belang zijnde mogelijkheden en ontwikkelingen.

Betrekkend op de kadastrale kaart kan gesteld worden dat deze voor vele klanten een belangrijke, zo niet onmisbare informatiebron is voor diverse werkzaamheden. Het Ka-

daster wil zich daarom ook blijven inspannen om als een betrouwbaar, open en degelijk bedrijf in de maatschappij te staan en streeft vanuit die optiek langdurige, continue en vruchtbare relaties met zijn klanten na.

Literatuur

Koen, L.A., Kwaliteitszorg bij het Kadaster. NGT Geodesia 1991 no. 11.

Ijsselstein, J.A. en Kap A.P., Het kadastraal perceel: een stevig fundament! NGT Geodesia 1995 no. 7/8.

Veen van der, J.B., Beheer en levering van GEO-informatie door het Kadaster. Lezing uitgesproken op het congres 'De veelzeggende kaart' in mei 1996.

TOP10kwaliteit (kwaliteitsaspecten van TOP10vector)

B. Kolk

Topografische Dienst Nederland
E-mail: info@tdn.nl

Samenvatting

De kwaliteit van het product TOP10vector wordt toegelicht aan de hand van het productieproces. De wensen van de klant worden vertaald in eisen aan de leveringsomgeving en aan het product. Ingegaan wordt op zaken als prijsstelling, procesbewaking, productbeschrijving en uitwisselingsformaten. De relatie tussen de Topografische Dienst Nederland (TDN) en de gebruikers, vertegenwoordigd door Ravi en Gebruikersgroep, worden belicht en vertaald naar klantondersteunende activiteiten. Internationale ontwikkelingen, met name in Cerco-verband zijn veelbelovend. Het kwaliteitsmanagement van TDN beweegt zich in de richting van de ISO-9000-normen.

Inleiding

De Topografische Dienst Nederland heeft in ruim 180 jaar een reputatie opgebouwd als producent van gedetailleerde en nauwkeurige topografische kaarten. In het laatste decennium hebben zich een drietal ontwikkelingen voorgedaan die deze reputatie op de proef stellen. In de eerste plaats betekent de overgang van analoog naar digitaal een nieuw en ingewikkelder productieproces dan voorheen. In de tweede plaats is de opdrachtgever, het Ministerie van Defensie, zich meer bewust geworden van zijn rol. De opdrachtgever stelt nu eisen en stuurt middels jaarlijkse convenanten de TDN aan. Last but not least is er, in de vorm van digitale bestanden, een nieuw productassortiment op de markt gekomen dat zich mag verheugen op een intense belangstelling van de gebruikers. Deze gebruikers stellen eisen aan de producent.

Deze ontwikkelingen leiden tot hogere eisen aan het management en tot strengere eisen aan productieprocessen en producten. Kwaliteitsmanagement is een middel om de eerder genoemde reputatie opnieuw inhoud te geven. In dit artikel worden de kwaliteitsaspecten van TOP10vector besproken. De volgende gezichtspunten komen aan de orde:

1. productie;
2. product;
3. klant;
4. kwaliteitsmanagement, hoe verder?

1. Productie

Het productieproces wordt geanalyseerd op de volgende aspecten:

- a. luchtfotografie;
- b. fotogrammetrie;
- c. inwinning;
- d. digitalisatie;
- e. ISO 2859 (AQL).

Per onderdeel worden de kwaliteitsaspecten behandeld. Onder e) wordt ingegaan op (de niet toegepaste) toetsing op basis van Accepted Quality Level (SQL).

1a. Luchtfotografie

De vervaardiging van luchtfoto's is in een land als Nederland een delicaat proces. Zeker nu TDN een vaste vierjarencyclus realiseert, is het noodzakelijk dat een fotovlucht in het geplande jaar wordt uitgevoerd. Hier is sprake van een tegenstrijdigheid tussen de beschikbaarheid van de luchtfoto en de kwaliteit ervan. Het risico van de beschikbaarheid wordt binnen de perken gehouden door met meerdere bedrijven in zee te gaan. De kwaliteit van de luchtopname is van invloed op de interpretatiemogelijkheden en daarmee direct van invloed op de kwaliteit van het product. Daarnaast beïnvloedt de luchtfotokwaliteit de productiekosten. Goede luchtfoto's hebben twee kenmerken. Ze zijn opgenomen onder de juiste weersomstandigheden en het gebladerte aan de bomen belemmert het doorzicht naar de grond nog niet. Wanneer de bladeren beginnen te ontluiken worden deze eisen strijdig. Het kan dus voorkomen dat de beeldkwaliteit moet wijken voor het hogere doel recentheid.

1b. Fotogrammetrie

De fotogrammetrie wordt hoofdzakelijk toegepast voor het vervaardigen van orthofoto's. Paspuntsbepaling, triangulatie en blokvereffening zijn processen met overbepaalde waarnemingen. De geodetische waarnemingstheorie zorgt voor een goede kwaliteitsbewaking van het resultaat.

Voor het vervaardigen van orthofoto's is een hoogtemodel onontbeerlijk. TDN bevindt zich in de gelukkige positie te beschikken over twee hoogtebestanden van Nederland: TOPhoogte MD, de digitale kloon van de Hoogtekaart Nederland, en een bestand van gevectoriseerde hoogtelijnen van de kaartserie 1:50.000 dat eveneens is afgeleid van de Hoogtekaart. Uit TOPhoogteMD zijn hoogtelijnen gegenereerd. Beide sets hoogtelijnen zijn met elkaar vergeleken. Het aantal foutief gedigitaliseerde punten van TOPhoogte-



Figuur 1. Orthofotoproductie.



Figuur 2. De topograaf ter plekke.

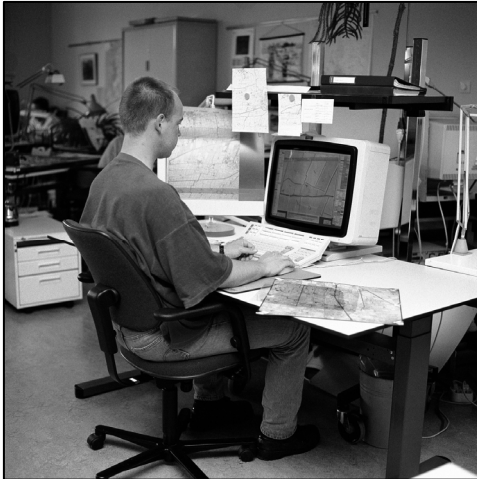
MD kon worden teruggebracht van 12 punten per kaartblad tot nagenoeg 0. Aangezien de Hoogtekaart Nederland 20 à 30 jaar oud is, worden in gebieden waar topografische wijzigingen hebben plaatsgevonden langs fotogrammetrische weg aanvullende metingen verricht. Locale verzakkingen spelen nauwelijks een rol omdat ze relatief klein zijn. Ook de omgeving is verzakt waardoor de doorwerking op de orthofoto te verwaarlozen is.

1c. Inwinning

De inwinning is, in tegenstelling tot de fotogrammetrische puntsbepaling, geen proces met overbepaalde waarnemingen. De topograaf bepaalt enkelvoudig de inhoudelijke kwaliteit van TOP10vector tijdens de terreinverkenning. De "Verkenningregels", een boekwerkje dat voorschrijft hoe de verkenning moet worden uitgevoerd, dient als leidraad. Aangezien 35 topografen met 35 percepties van het terrein de verkenning bepalen, zijn de verkenningvoorschriften ook gericht op het uniform interpreteren van de terreinsituatie. Jaarlijks worden nieuwe topografische situaties ontdekt welke, na enige discussie, in de verkenningvoorschriften worden vastgelegd. Topografen worden tijdens de verkenning regelmatig bezocht door Sectiehoofden waarbij het hoofd de werkzaamheden doorspreekt met de topograaf en tevens een *selecte* controle uitvoert.

1d. Digitalisatie

Het digitaliseerproces vindt interactief plaats waarbij het vectorbestand wordt aangevuld met de verkenninggegevens van de topograaf. De geannoteerde verkenningfoto, waarop de mutaties zijn aangegeven, wordt gescand en als referentiebestand onder TOP10vector afgebeeld op het beeldscherm. De digitalisatie wordt uitgevoerd door een ander dan degene die verkend heeft. Zo komen de topografen met elkaars werk in aanraking. Dit heeft een positieve uitwerking op de uniforme toepassing van de verkenningregels. Het digitaliseerproces is gebaseerd op de 'digitaliseervoorschriften', een set regels welke de digitaliseerder in acht dient te nemen. De bestandsstructuur wordt softwarematig volledig gecontroleerd op consistentie. Andere typen automatische controles, zoals het testen op heidevelden in het IJsselmeer, zijn nog nauwelijks ontwikkeld. De inhoudelijke controle vindt volledig plaats middels controleplots. Er zijn meerdere



Figuur 3. Interactief digitaliseren.

controleplots per bestand nodig om alle attributwaarden te kunnen weergeven. De plots worden nagezien door de topograaf die de verkenning heeft uitgevoerd. Hij/zij weet het beste wat de annotaties op de luchtfoto betekent en verkrijgt door de confrontatie met de controleplot tevens een terugkoppeling op het eigen werk.

1e. ISO 2859 (AQL-methode)

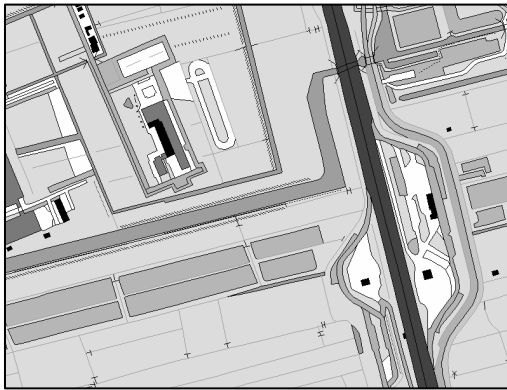
Wanneer het product gereed is, kan een algehele toetsing van de productkwaliteit plaats vinden. Een voorbeeld van een dergelijke controle is het testen van gloeilampen. Op basis van een representatieve steekproef kan bepaald worden of een partij aan de criteria voldoet. In [1] wordt deze methode toegepast op geografische informatie. Een gebied wordt als steekproef geselecteerd en voor de tweede keer gekarteerd. Het resulterende bestand wordt vergeleken met het eerder geproduceerde bestand. De analyse geeft een prima vergelijking van de kwaliteit van het steekproefgebied. Zo simpel als het ligt bij de lamp, waar het toetsingscriterium volstrekt helder is, ligt het bij geografische informatie niet. Twee aspecten wil ik nader belichten.

In de eerste plaats, de verschillen die geconstateerd worden in het steekproefgebied, doen geen uitspraak over de kwaliteit van de kartering doch uitsluitend over het verschil tussen de twee karteringen van het steekproefgebied. Misschien is de steekproef wel fout. Er is geen duidelijke eis waar de kartering aan moet voldoen.

Ten tweede, indien de steekproef toch wordt verworpen, dient zich de vraag aan voor welk gebied het steekproefgebied representatief is? Indien het naastliggende gebied is gekarteerd door een collega, wat zegt de steekproef dan over dat gebied? Niets, want karteren is een individuele bezigheid. Het ligt dus voor de hand het werk van de door de steekproef 'getroffen' medewerker af te keuren. Welk werk keur je dan af? Dat van die bewuste maandagochtend, of het werk van de hele week, maand, of jaar? Of keur je het werk van de maandagochtenden af? Is de vrijdagmiddag dan ook niet verdacht?

Mijn conclusie is, dat het slecht gesteld is met de representiviteit van de steekproef.

Ik denk dat het de voorkeur verdient om tijdens het productieproces het product te testen. Karteringen verwerpen en nogmaals het proces van a tot z doorlopen is dweilen met de kraan open. Medewerkers moeten voldoende geschoold zijn en tijdens hun werk voldoende geconcentreerd zijn. Dit vergt een op de persoon toegesneden kwaliteitsbegeleiding die voorafgaande aan en tijdens het productieproces moet plaatsvinden.



Figuur 4. Afbeelding van TOP10vector.

2. Product

Het product wordt belicht vanuit de invalshoeken:

- a. kosten;
- b. opbrengsten;
- c. conversie productie-leveringsbestand;
- d. productbeschrijving;
- e. uitwisseling van bestanden.

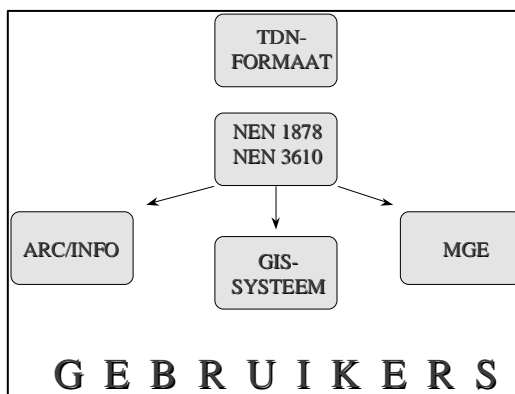
De kwaliteitsbewaking van het product zelf - de inhoud - is in de paragraaf over het productieproces behandeld.

2a. Kosten

TDN kent al tientallen jaren een systeem van kostprijsbepaling. De gewerkte uren, materialen en ingekochte diensten worden geregistreerd. De urenregistratie van TOP10vector is onderverdeeld in zo'n 10 categorieën. Op deze wijze wordt per onderdeel van het productieproces inzicht verkregen in de kosten. De registraties uit het verleden maken het mogelijk een lange termijnplanning te maken, inclusief een kostenschatting. De planningsgegevens worden jaarlijks gebruikt voor het aansturen van het productieproces.

2b. Opbrengsten

De prijs van TOP10vector wordt volledig bepaald door de kosten. Het Ministerie van Defensie betaalt een deel, de civiele gebruiker betaalt het ontbrekende gedeelte. In 1994 hebben het Ministerie van Defensie, waar TDN onder ressorteert, en de Ravi afspraken gemaakt over het gebruik van TOP10vector als Kernbestand voor de civiele gebruiker. Een van de afspraken betreft de opbrengsten van TOP10vector [2]. De klacht dat de prijs te hoog is, klinkt regelmatig. Het gebruik van het bestand door financieel



Figuur 5. Uitwisseling.

minder draagkrachtige afnemers wordt er door geremd en voor sommige groepen zoals particulieren zelfs onmogelijk gemaakt.

De discussie over de prijs/kwaliteit van het product is een lastige: enerzijds is de prijs gebonden aan de kosten van het product terwijl anderzijds de prijs die de klant wil betalen, wordt bepaald door het nut van het bestand voor de gebruiker. Dit is een principiële dilemma dat slechts op twee manieren kan worden opgelost: de productiekosten drastisch verlagen, of een deel van de inkomsten uit bijvoorbeeld de algemene Rijksmiddelen bekostigen zodat een maatschappelijk acceptabele prijs mogelijk wordt. De productiekosten gaan zeker omlaag. Het tempo waarin dat mogelijk is, biedt de komende jaren echter geen soelaas tenzij ernstige concessies aan de kwaliteit worden gedaan. Een politieke oplossing lijkt daarom meer voor de hand te liggen.

2c. Conversie productie - leveringsbestand

In 1990 heeft TDN de structuur van TOP10vector bepaald op basis van twee criteria: het bestand moest geschikt zijn om er topografische kaarten op de schalen 1:10.000 en 1:25.000 uit af te leiden en het bestand moest voldoen in een GIS-omgeving [3]. Dat laatste betekende o.a. dat de structuur eenvoudig diende te zijn opdat conversie naar een willekeurig gebruikerssysteem relatief eenvoudig zou zijn te realiseren. De eerste doelstelling is gehaald, de tweede verdient een herbezinning. Sommige systemen kunnen prima met de data uit de voeten, voor andere systemen is de conversie een crime. Daarnaast is de soort toepassing bepalend voor de conversie-inspanning.

De toepassingen van gebruikers worden steeds volwassener hetgeen de roep naar een meer volwassen bestand met zich brengt. Het samenspel tussen producent en gebruiker moet leiden tot een aan de eisen des tijds aangepast product. Voor TDN betekent dit concreet dat intern een onderscheid gemaakt wordt tussen het productiebestand en het daaruit afgeleide leveringsbestand. Het ene bestand is geoptimaliseerd voor het productieproces, terwijl het andere op de gebruikerseisen is toegesneden.

2d. Productbeschrijving

De 'Objectcatalogus' geeft een beschrijving van TOP10vector. Dit betreft alle aspecten van het bestand zoals de structuur, attribuering en objectbeschrijving. Deze productbeschrijving maakt een soortgelijke ontwikkeling door als de TOP10-bestanden. Er bestaat een elementaire bestandsbeschrijving die nu echter niet meer voldoet. Begin 1998 is een nieuwe productbeschrijving met meer informatie beschikbaar.

2e. Uitwisseling van bestanden

Top10vector wordt aangeboden in de formaten NEN 1878, DXF, (Intergraph-)DGN en Arc-Info. NEN 1878 is het uitwisselingsformaat in Nederland. Bij de conversie van TOP10vector van NEN 1878 naar een GIS-systeem, staat de gebruiker voor de taak de conversie uit te voeren. In het geval van Arc-Info bleek de benodigde inspanning van de gebruiker dermate hoog dat een beroep werd gedaan op TDN deze conversie te realiseren. In samenwerking met het Ministerie van LNV en Logisterion, de Nederlandse importeur van Arc-Info, is dat gelukt.

Wat we zien is, dat de afspraken omtrent de uitwisseling van bestanden, gemaakt in 1981 -toen SUF, de voorloper van NEN 1878, het levenslicht zag - niet meer voldoende op de praktijk zijn toegesneden. Het behoort niet de bedoeling te zijn dat elke gebruiker individueel een moeizaam conversiepad moet afleggen alhoewel dit wel de consequentie is van de afspraken over NEN1878. Van de andere kant lijkt het mij niet reëel van TDN te verwachten om conversies naar een veertigtal GIS- en CAD-pakketten te onderhouden. In overleg is uiteraard veel mogelijk maar ik denk dat de oplossing gevonden

moet worden in het zodanig structureren van TOP10vector in NEN1878 dat er een relatief eenvoudige conversie naar gebruikerssystemen mogelijk wordt. De gebruikersgroep TOPbestanden en de gebruikersgroepen van de diverse systemen zijn hier de natuurlijke gesprekspartners voor TDN. Een goede opslag van TOP10vector binnen NEN 1878 kan dan gepaard gaan met de implementatie van NEN 3610.

3. De klant

De volgende aspecten worden behandeld:

- a. Ravi;
- b. Overlegplatform gebruikers TOP-bestanden (OGT);
- c. Klant-ondersteunende activiteiten.

3a. Ravi

De belangrijkste overheidsgebruikers van geografische informatie zijn verenigd in de Ravi. De Ravi heeft rond 1993 het idee van het Kernbestand gelanceerd, een bestand dat voor de gebruikers een basis moet vormen voor (gemeenschappelijke) geografische toepassingen. Dit heeft in 1994 geleid tot de keuze van TOP10vector als basis voor het Kernbestand. De Ravi houdt de ontwikkeling van TOP10vector scherp in de gaten ten einde haar doelstellingen met het Kernbestand te realiseren. (De rol van de Ravi kwam al eerder ter sprake in verband met de prijsstelling van TOP10vector.) De Ravi timmert op dit moment aan de weg met de projectgroep "gebruikerseisen Kernbestand". Zij vervult daarmee de rol van kwaliteitsbewaker op nationaal geografisch beleidsniveau.

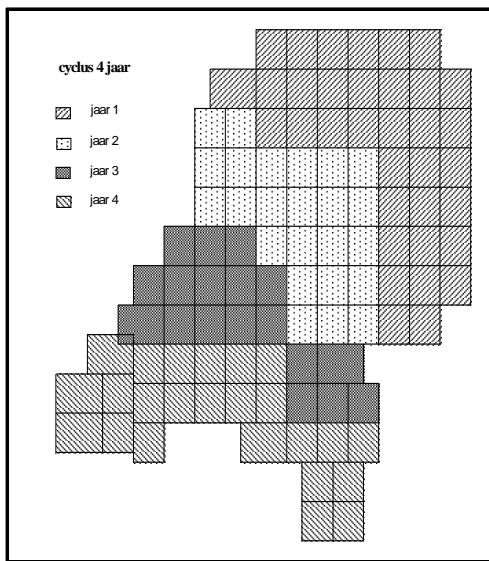
TDN heeft haar, in 1989 gelanceerde, slogan "Nederland in 1997 digitaal" waargemaakt. Ook is in 1997 de overstap gemaakt naar de vierjarencyclus. Beide mijlpalen heeft TDN mede door de inzet van de Ravi kunnen realiseren.

Kernbestand = TOP10vector + gemeentegrenzen + wegassen + hogere bijhoudingsfrequentie
--

Figuur 6.

3b. Overlegplatform Gebruikers TOP-bestanden (OGT)

In 1995 vond in Emmen de eerste bijeenkomst plaats tussen afnemers van TOPbestanden en TDN. 60 gebruikers waren aanwezig. Het heeft tot 1997 geduurd voor er een vervolg aan het overleg gegeven kon worden in de vorm van het OGT. Het OGT kan rekenen op de warme steun van Ravi en TDN. Het OGT is de belangrijkste partner van TDN in het formuleren van de eisen aan TOP10vector. Het OGT vervult een prominente rol in de voornoemde Ravi-projectgroep Gebruikerseisen. Dit lijkt mij voor het OGT een aardig aangrijpingspunt om de prille doelstellingen in daden om te zetten. De vertaling van de veelheid aan vormen van "fitness for use" naar de productie van TOP10vector is voor beide partijen nieuw. De link tussen productie en klant, de afweging tussen wensen en mogelijkheden, geeft een welkome en stimulerende dimensie aan het werk van de medewerkers van TDN.



Figuur 7.

3c. Klant-ondersteunende activiteiten

De explosieve groei van de afzet van bestanden, heeft tot gevolg dat een service-organisatie ter ondersteuning van de gebruikers nodig is. Het aantal medewerkers op de afdeling Marketing en Verkoop is in vier jaar gestegen van 0,5 naar 7. Het aantal Gigabytes is in die periode gestegen van 0,5 naar een 'kwadratisch veelvoud' van 7. De technische ondersteuning van enerzijds de klant en anderzijds de data is tot een krachttoer uitgegroeid. De huidige, in de afgelopen jaren spontaan gegroeide opzet is uit zijn jasje gegroeid en wordt binnenkort vervangen. Zo zal er voor de klant een professionele Helpdesk beschikbaar zijn. Ook het data-beheer voor productie, levering en archivering van raster- en vectorbestanden wordt in een grotere jas gestoken.

Aangezien gebleken is dat het doorgronden van de TOP-bestanden voor de gebruiker een aanzienlijke belasting is, met name in tijd, heeft TDN het idee opgevat een cursus TOP(10)vector op te zetten. De voorbereidingen hiervoor zijn gestart. Deze cursus biedt tevens de mogelijkheid de klant-leverancier relatie aan te halen.

TOP10vector = KERNBESTAND:
 - vierjaarlijks: luchtfoto en verkenning
 >> vanaf 1997
 - tussentijds: GBKN-mutaties
 - belangrijke objecten
 >> vanaf 2000

Figuur 8. Beleid TOP10vector.

4. Kwaliteitsmanagement, hoe verder?

Internationaal zijn enkele belangrijke ontwikkelingen gaande op het gebied van kwaliteit. Dit betreft:

- a. Cerco en
- b. CEN TC 287 en ISO TC 211.

Tenslotte de hamvraag: hoe denkt TDN in de toekomst om te gaan met kwaliteit:

- c. TDN en ISO-9000.

4a. Cerco

Cerco is het overlegorgaan van de directeuren van Europese Topografische Diensten. Cerco heeft in 1997 besloten een werkgroep Kwaliteit op te richten. Het voorwerk was al verricht door een ad hoc groep. Na de goedkeuring door Cerco, kon de werkgroep op 15 november 1998 een vliegende start maken. De werkgroep concentreert zich op de volgende onderwerpen:

- ISO 9000 en Topografische Diensten. De Engelse Ordnance Survey heeft onlangs het certificatie-traject voor ISO-9000 succesvol doorlopen. De ervaringen daarbij opgedaan dienen als leidraad voor de collegadiensten. Voor 1998 staat er een workshop op het programma dat gewijd zal zijn aan de implementatie van het ISO-9000-traject. Het uitwisselen van ervaringen is het tweede thema.
- Data-kwaliteitsaspecten. Hier gaat het om zaken als dataset-specificatie, kwaliteits-evaluatie-procedures en bepaling van gebruikersbehoeften.
- Implementatie van de CEN/TC287-normen prEN 12656 (Kwaliteit) en prEN 12657 (Metadata). Het doel is te komen tot een gezamenlijke interpretatie van de TC287-normen. Daarnaast zullen aanbevelingen worden gedaan voor verbeteringen.

4b. CEN TC287 en ISO TC211

Elders in deze artikelenreeks wordt uitvoerig ingegaan op voornoemde werkgroepen. De normen resulterend uit CEN TC287 hebben een bindend karakter. Het is derhalve zaak dat de normen een hoog realiteitsniveau hebben zodat ze in de praktijk zinvol kunnen worden toegepast.

De gescheiden activiteiten van CEN en ISO vind ik bevreemdend. Te meer daar Europeanen in ISO TC211 goed vertegenwoordigd zijn. Het wekt niet de indruk dat de klant centraal staat, iets wat je gezien het onderwerp, wel zou mogen verwachten. Het is opvallend dat bijvoorbeeld op ICA-congressen de belangstelling voor deze materie minimaal is. Die belangstelling lijkt omgekeerd evenredig aan het enthousiasme van de selecte groep mensen die de werkgroepen bevolken. Ik heb daarom het gevoel dat er nog heel wat geschaafd zal moeten worden aan de normen alvorens ze tot acceptatie in de praktijk zullen leiden. Gezien de uitgebreidheid van de materie is het initiatief van de Cerco, de normen gezamenlijk te toetsen, zeer toe te juichen.

4c. TDN en ISO-9000

In de inleiding is al gesteld dat de druk op het management van de organisatie is toegenomen. Die druk zal eerder verder toenemen dan afnemen. Voor het goed managen van bedrijfsprocessen is kwaliteitsmanagement gebaseerd op ISO-9000 in mijn ogen een uitstekend middel. Succesvolle implementatie betekent naar mijn mening wel dat een klantgerichte bedrijfscultuur moet bestaan. Hier wringt bij overheidsorganisaties, waarvan TDN er ook een is, de schoen. TDN is op dit moment op bescheiden schaal bezig de beheersing van bedrijfsprocessen aan te scherpen. Het is noodzakelijk in de pas te blijven lopen met de (veranderende) bedrijfscultuur. In feite is er sprake van twee parallelle processen die alleen tezamen het gewenste doel kunnen bereiken. Het initiatief van Cerco om het ISO-9000-traject te verkennen is daarom een uitstekend initiatief dat voor TDN op het juiste moment komt.

Literatuur

[1] Pätynen, V., Kempainen, I. en Ronkainen, R., Testing for completeness and thematic accuracy of the national topographic data system in Finland. Proceedings 18th ICA Conference 1997.

- [2] Berends, J. en Janssen, H.C., Het 1:10.000 Kernbestand. NGT Geodesia/Kartografisch Tijdschrift 1995 no. 1.
- [3] Kolk, E., Op weg met TOP10vector. NGT Geodesia/Kartografisch Tijdschrift 1995 no. 1.

Een methodiek voor kwaliteitscontrole van vaarweggegevens

N. Schmorak

Adviesdienst Verkeer en Vervoer
Rijkswaterstaat
E-mail: n.s.schmorak@avv.rws.minvenw.nl

1. Inleiding

De Adviesdienst Verkeer en Vervoer (AVV) is een specialistische dienst van Rijkswaterstaat die als taak heeft het technisch ondersteunen van het beleid op gebied van verkeer en vervoer. De hoofdafdeling Basis Gegevens (BG) van AVV is producent, beheerder en verstrekker van gegevens en informatie, die in onderzoek, evaluatiestudies en voor het aansturen van beleid gebruikt worden.

De inwinning van grondstoffen voor het vervaardigen van het productenpakket wordt steeds meer aan derden overgelaten. Om voldoende kwaliteit in zijn product te kunnen garanderen moet BG duidelijke afspraken met de leveranciers van grondstoffen maken omtrent kwaliteitscontroles.

Een van de productgroepen van BG betreft het vaarwegennet. De BG-database WEGBIN (WEGwijzer voor de BINnenscheepvaart)¹ bevat informatie met betrekking tot kenmerken van de Nederlandse vaarwegen. Gebruikers van de database zijn de beleidsmakers die zich met goederenvervoer bezighouden, gebruikers van de natte infrastructuur (schippers en vervoersmaatschappijen), beheersinstanties van vaarwegen (rijk, provincies, waterschappen, gemeenten) en anderen. Tevens is de database bedoeld als invoer van andere informatiesystemen.

De kenmerken die de vaarweg karakteriseren moeten voortdurend worden bijgehouden. Hiervoor heeft BG een bedrijf opdracht gegeven. Jaarlijks krijgt BG gegevens van ca. 200 vaarwegen welke door de opdrachtnemer zijn waargenomen. Op deze manier wordt om de ca. 3 jaar een update van het hele vaarwegennet gerealiseerd. BG is momenteel in onderhandeling met de opdrachtnemer om een kwaliteitscontrole contractueel vast te leggen. De procedure die hiervoor is opgesteld en die in dit artikel wordt beschreven is op de ISO norm 2859 gebaseerd.

Dit artikel beoogt de manier te illustreren waarop de kwaliteit van geografische informatie gekwantificeerd kan worden. Ten eerste wordt de ISO norm 2859 in grote lijnen beschreven. Daarna komt aan de orde de toepassing van deze norm op de controle van gegevens. Beschreven zal worden hoe BG van plan is de kwaliteit van de informatie omtrent het vaarwegennet te bewaken.

¹ WEGBIN wordt medio 1998 door ViN (Vaarwegkenmerken in Nederland) vervangen. A.g.v. de invoer van dit nieuwe systeem zal de hier beschreven kwaliteitscontrolemethodiek in de toekomst enigszins maar niet op essentiële wijze moeten veranderen.

2. De ISO norm 2859

Deze ISO norm geeft richtlijnen voor het opzetten van kwaliteitscontroles d.m.v. inspecties op steekproeven. De norm is voor industriële productie ontwikkeld waar de productiestroom uit telbare eenheden bestaat.

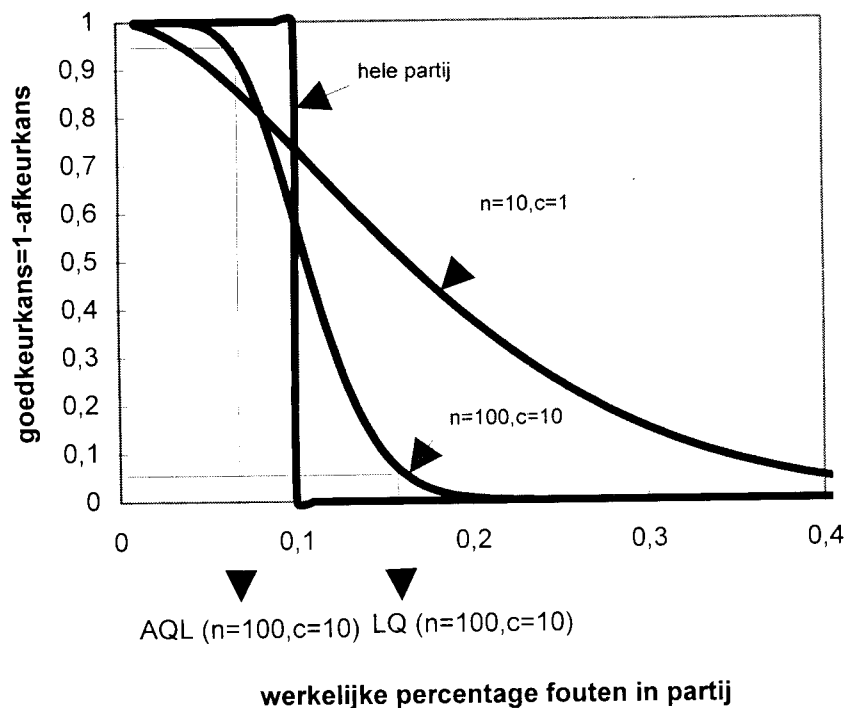
De productiestroom wordt voor de kwaliteitscontrole in partijen ingedeeld. Een partij bestaat uit een bepaald aantal (productie)eenheden of items waaruit een steekproef aselekt wordt uitgetrokken. Bij de inspectie worden de attributen van de items, die deel van de steekproef uitmaken, geïnspecteerd. Attributen zijn kenmerken c.q. eigenschappen van de items welke aan een kwaliteitsnorm moeten voldoen. B.v. als de items balpenen zijn, kunnen attributen zoals de lengte, de kleur en/of het schrijfvermogen worden gecontroleerd.

Een items is foutief als één of meer van zijn attributen niet aan de norm voldoen. Aan de hand van het aantal foutieve items in de steekproef beslist men over de acceptatie van de partij. De norm neemt aan dat de partijen qua kwaliteit homogeen zijn d.w.z. dat ze uit eenheden c.q. items bestaan die onder dezelfde omstandigheden en op dezelfde manier zijn vervaardigd.

Bij steekproefcontrole zijn er zowel voor producent als voor consument risico's gemeoid. Deze risico's moeten expliciet worden gemaakt bij het afspreken van een kwaliteitsprocedure.

Figuur 1 geeft de kans weer dat een steekproef door een bepaalde keuring komt als functie van de kwaliteit van de partij waaruit de steekproef getrokken is. Deze kans is afhankelijk van de steekproefgrootte en van het toegestane aantal fouten in de steekproef. Deze krommen worden in de ISO 2859 norm als OC-curve aangeduid, 'Operating Characteristic Curves'.

Stel dat men een partij met 10% fouten niet acceptabel vindt. Mocht men beslissen de hele partij te controleren dan is er zekerheid over de aan de hand van het experiment te



Figuur 1. De risico's bij controle op steekproef.

trekken conclusie. Als de partij minder dan 10% foutieve items bevat is 100% zeker dat de partij door de keuring komt.

Spreken consument en producent af om een steekproef van 100 items te nemen en goed te keuren bij 10 foutieve items of minder dan ontstaat er risico. Aan de ene kant is er kans dat de betreffende partij door de keuring komt terwijl hij meer dan 10% fouten bevat (consument risico). Aan de andere kant is er een kans dat de partij minder dan 10% foutieve items bevat maar toch afgekeurd wordt (producent risico). Naarmate de steekproefgrootte kleiner wordt neemt de onzekerheid toe hetgeen te zien is in het steeds minder steil lopen van de OC-krommen.

Het idee achter de ISO 2859 norm is om de steekproefgrootte (n) en het maximaal toegestane aantal fouten in de steekproef (c) zodanig te kiezen dat zowel het consument als het producent risico binnen de perken blijft. Het onacceptabel percentage fouten in de partij moet een kleine goedkeuringskans hebben (5 á 10%) en wordt in de ISO 2859 norm aangeduid als LQ, 'Limiting Quality'. Het acceptabel percentage fouten in de partij die een kleine afkeuringskans moet hebben (5%) wordt in de ISO 2859 norm aangeduid als AQL, 'Acceptable Quality Level'. Indien de producent zorgt dat de kwaliteit van de productie in de buurt van de AQL blijft dan lopen de partijen nauwelijks risico om afgekeurd te worden.

De ISO 2859 norm bestaat uit 3 delen:

ISO 2859-1: Dit eerste gedeelte is bedoeld voor de inspectie van een continue productiestromen. Uitgaande van de AQL en de partijgrootte kan een controleplan afgeleid worden. Hierbij is de steekproefgrootte afhankelijk van de resultaten van voorafgaande controles. Indien de kwaliteit van een aantal partijen heel goed blijkt te zijn en er geen verandering in de werkwijze wordt verwacht kan men naar kleinere steekproeven overgaan (reduced inspection). Blijkt de gewenste kwaliteit niet gehaald te worden dan worden er grotere steekproeven genomen (tightened inspection).

ISO 2859-2: Dit tweede deel van de norm is bedoeld voor de keuring van losse partijen en richt zich vnl. op de beperking van de consument risico. Uitgaande van LQ en de partijgrootte kan de steekproefgrootte worden bepaald.

ISO 2859-3: Hiermee kunnen skip-lot procedures worden opgezet. Als de kwaliteit van partijen heel goed blijkt te zijn, worden een aantal controles overgeslagen.

In de ISO 2859 norm wordt de kwaliteitscontrole op attributen uiteengezet. Een andere methode van controle is op variabelen. Hierbij worden dimensies c.q. eigenschappen van de items gemeten. Uit de controle resulteert niet alleen informatie over het wel of niet voldoende behalen van een norm maar ook een schatting van de waarde van de betreffende dimensie. De ISO normen die deze methode ondersteunen zijn de ISO 3951 norm en 8423 norm.

Het voordeel van de controle op variabelen t.o.v. de controle op attributen is het hogere steekproefrendement. D.w.z. de onzekerheid bij een beslissing aan de hand van een bepaalde steekproefgrootte is kleiner bij de variabele methode dan bij de attributen methode. Dit is waar mits de juiste aannames over de soort en vorm van de frequentieverdelingen worden gemaakt. De behoefte aan aannames maakt de controle op variabelen minder robuust en ingewikkelder dan de controle op attributen.

3. De toepassing op vaarweggegevens

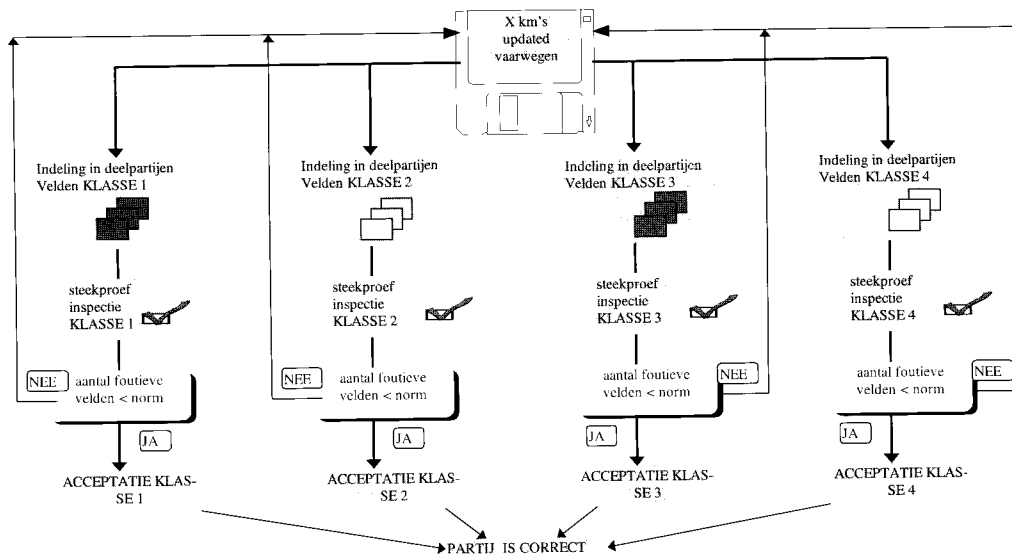
De vaarwegkenmerken zijn in punt- en lijngegevens in te delen welke aan vaarwegnummers en kilometreringen in de database worden gerelateerd. Puntgegevens zijn kenmerken van objecten die zich op de vaarwegen bevinden zoals bruggen, sluizen en ankerplaatsen. Lijngegevens zijn kenmerken van gedeelten vaarweg zoals toegestane vaar diepte, bijzondere bepalingen en vaarwegnaam.

De meeste van de kenmerken worden op kantoor nagezien door informatie uit kaarten te halen of door de beheerder van de betreffende vaarweg op te bellen. In situ wordt de aanwezigheid c.q. wijziging van objecten en/of kenmerken van deze objecten nagegaan.

Dimensies, zoals vaarwegdiepten, breedten en hoogten, worden nooit zelf gemeten maar aan beheerders nagevraagd. De beheerders zijn verantwoordelijk voor de juistheid van deze gegevens en zijn daarom de maatstaf. De controle moet dan vaststellen of de dimensies in de database wel of niet afwijken van wat de beheerders aangeven. Hiervoor, net als voor de overige kenmerken van de database, wordt de methode van controle op attributen gebruikt.

De opdrachtnemer is contractueel verplicht een update van het hele wegennet, d.i. ca. 6500 km vaarwegen, om de 3 jaar te realiseren. De leveringen geschieden in vaste termijnen van ca. 2 maanden zodat er sprake is van een permanent binnenstromen van 'pakketten' gegevens. Die 'pakketten' kunnen als partijen beschouwd worden. AVV is van plan controles op compleetheit en correctheid van deze leveringen contractueel vast te leggen.

Omdat het huidige contract met de opdrachtnemer geen controles specificiert worden momenteel kwaliteitsmetingen slechts incidenteel uitgevoerd. Deze metingen die verder geen financiële gevolgen voor de opdrachtnemer hebben, zijn bedoeld om gewenste en haalbare kwaliteit op elkaar af te stemmen. Hieruit zullen AQL-waarden afgeleid worden voor de op contractbasis geregelde keuring. De tweetal controles die AVV/BG voorstelt zijn:



Figuur 2. Procedure voor controle op correctheid.

Controle op correctheid

De partij is hierbij de verzameling kenmerken (punt- en lijn- gegevens) die bij één en dezelfde periodieke levering horen. Per kenmerk wordt een tolerantie vastgesteld zodat, als een waarneming de tolerantie overschrijdt, automatisch als fout aangemerkt kan worden. Deze toleranties samen met de beschrijving van de correcte waarnemingswijze zijn in een handleiding vastgesteld. Deze handleiding is in feite het instrument waarmee de kwaliteit van de waarnemingen wordt gedefinieerd.

Niet alle kenmerken van de database zijn even belangrijk. Bij sommige kenmerken kunnen fouten tot rampzalige gevolgen leiden, b.v. bij doorvaarthoogte hoogspanningslijn. Met andere kenmerken mag men flexibeler omgaan; of omdat het kenmerk weinig relevant is (b.v. kleur en soort licht) of omdat het een kenmerk betreft dat bij een lage orde vaarweg hoort.

De partij wordt daarom voor de controle op correctheid naar belang in 4 klassen ingedeeld. De kenmerken van klasse 1 zijn de belangrijkste en worden strenger gekeurd. D.w.z. dat een strengere AQL voor deze klasse gehanteerd wordt hetgeen leidt tot een grotere steekproef en tot een kleiner percentage toegestane fouten dan bij de overige klassen.

Figuur 2 illustreert de procedure voor controle op correctheid van kenmerken. Per klasse wordt uit de partij een steekproef van items aselekt getrokken. De steekproefgrootte en het maximaal toegestane aantal fouten per klasse wordt aan de hand van de ISO 2859 norm bepaald.

De kenmerken die binnen de steekproef vallen worden vervolgens door medewerkers van AVV/BC opnieuw waargenomen en op correctheid beoordeeld. De gevonden fouten worden met de opdrachtnemer besproken en indien nodig worden aanvullingen c.q. wijzigingen van de handleiding uitgevoerd. Mocht uit de inspectie blijken dat bij één of meerdere deelpartijen het aantal foutieve kenmerken groter is dan het maximaal toegestane aantal dan moet de hele partij kenmerken (zonder betaling van meerwerk) opnieuw worden waargenomen.

Controle op compleetheid

De partij is hierbij de verzameling kenmerken van puntgegevens die bij één en dezelfde periodieke levering horen. Een 'fout' in het kader van controle op compleetheid wordt gedefinieerd als: een in de werkelijkheid bestaand puntgegeven dat in de partij niet is opgenomen of omgekeerd een in de partij opgenomen puntgegeven dat in de werkelijkheid niet (meer) bestaat.

Figuur 3 illustreert de procedure. Uit de te controleren partij worden één voor één km vaarwegen aselekt getrokken. Bij iedere trekking wordt steeds de cumulatieve som berekend van het aantal puntgegevens dat aan die kilometer vaarweg staat gekoppeld. De steekproef is compleet op het moment dat die cumulatieve som gelijk is aan het door de ISO 2859 norm vereiste aantal items in de steekproef. Op deze manier wordt een lijst km's vaarwegen samengesteld. AVV/BC gaat vervolgens naar de betreffende locaties en noteert het aantal items dat in de geselecteerde km's aanwezig is.

Uit vergelijking tussen die controlewaarnemingen en de door de opdrachtnemer geleverde waarnemingen volgt het aantal 'fouten' in de steekproef. Slechts als dit aantal kleiner/gelijk blijkt te zijn dan/aan het toegestane aantal, wordt de partij goedgekeurd. Anders moet de opdrachtnemer de partij op eigen kosten completeren.

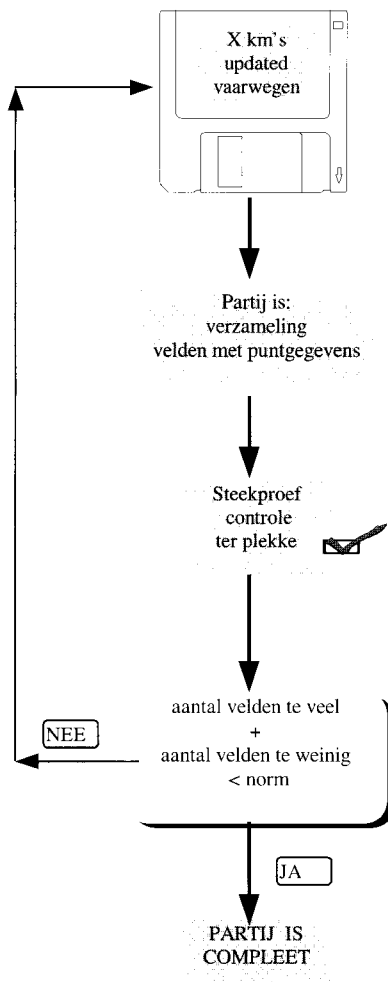
4. Slotopmerking

Hier is een methode gepresenteerd voor het meten van kwaliteit van (geografische) gegevens. Binnen AVV/BG zijn op diverse plaatsen van de organisatie soortgelijke kwaliteitsprocedures geïmplementeerd. Deze procedures zijn als instrument bedoeld om de productie van de verschillende BG- gegevensverzamelingen te bewaken.

Het behalen en handhaven van kwaliteit is een proces op zich dat voortdurend aandacht vergt. Daarin kan een kwaliteitscontrole een belangrijke rol spelen. Het objectief definiëren van kwaliteit in termen van meetbare parameters en het regelmatig bewaken van deze parameters kan heel bruikbaar zijn in optimalisatie van processen voor het inwinnen en verwerken van gegevens. Nog een opbrengst van een goed opgezette kwaliteitscontrole is dat het tot meer betrokkenheid van de medewerkers leidt die in een 'gezonde' situatie goed werk willen leveren.

Referenties

- [1] Draft International Standard ISO2859 (Part 0): "Introduction to the ISO 2859 attribute sampling system".
- [2] Acheson J. Duncan, Ph.D. "Quality Control and Industrial Statistics". ISBN 0-256-03535-0.



Figuur 3. Procedure voor controle op compleetheid.