

Alfred Stein

Earth Observation Science

International Institute for Geo-Information Science

and Earth Observation

Postbus 6, 7500 AA Enschede

stein@itc.nl



Alfred Stein

Oratie

Geef me de ruimte

Als we met Google Earth een stukje van onze aarde bekijken, dan vullen we de beelden die we zien onmiddellijk in met onze eigen interpretatie. Hoe doet een computer dit? Bij een dergelijk probleem komt heel wat spatiële statistiek kijken. In zijn oratie, uitgesproken op 20 september bij de aanvaarding van het ambt van bijzonder hoogleraar aan de Universiteit Twente, geeft Alfred Stein een overzicht van de statistische beeldanalyse.

Voordat ik met de wetenschap begin is het goed om het woord aan de kunst te laten. In dit geval richten we ons op sterk door wiskunde geïnspireerde kunstenaar M.C. Escher, met zijn beroemde litho Dag en nacht (zie figuur 1). We zien links de dag en rechts de nacht en de vogel die het allemaal waarneemt. Maar waar het mij omgaat is de verticale lijn in deze litho weiland wordt vogel. Hiermee geeft de kunstenaar een impressie waar we in ons vakgebied mee bezig zijn. De vogel die over het land vliegt probeert zijn objecten te kennen, hij gebruikt zijn kennis van het land en ziet met een enkele oogopslag wat voor zijn overleven van belang is. Wat hij is en wat hij ziet worden een. Hij kiest de ruimte, de ruimte om te vliegen en de ruimte om zijn kennis te vergaren.

Laten we overstappen naar de wetenschap en als eerste de naam van de leerstoel bekijken: stochastische beeldanalyse. Deze term bevat een aantal verschillende begrippen. In de eerste plaats komen we het toch wat lastige woord 'stochastisch' tegen. Daarnaast gaat het over beelden en tenslotte over analyses. In de stochastische beeldanalyse richten we ons op het verkrijgen van maximale informa-

tie uit beelden met behulp van stochastische methoden. De stochastiek (van het Griekse woord *στοχασμοι*: raken, gissen of mikken) houdt zich bezig met het modelleren van onzekerheden. Stochastiek komen we bijvoorbeeld tegen bij het gooien van een dobbelsteen, waarbij we de waarden 1 tot en met 6 kunnen krijgen, onder eerlijke condities allemaal met een kans 1/6. En bij het voorspellen van uw aankomsttijd bij deze lezing. En bij de lengte ervan. De stochastiek kom je overal tegen!

Beeldanalyse

Na deze ouverture kijken we nu als eerste thema naar een satellietbeeld. En we beginnen in Twente[5]. Figuur A laat een landsat-beeld zien waarbij we informatie in zes banden hebben. We mengen de kleuren en laten de kleuren rood, groen en blauw corresponderen met drie van de zes banden. Het is dus niet zo dat de kleur op het beeld overeenkomt met de kleur op de grond. Vaak zullen we de segmenten die bijvoorbeeld overeenkomen met een weiland, een stad, een kanaal, een gebouw op deze beelden automatisch willen onderscheiden. Daarvoor zijn verschillende me-

thoden ter beschikking. Globaal kunnen we een onderscheid maken in 'segmentatiemethoden en classificatie', in 'filters' en in 'wiskundige morfologie' [17]. Ik zal me vandaag beperken tot segmentatie en classificatie. Laten we zien hoe dat in zijn werk gaat.

Iedere pixel correspondeert met een locatie in het terrein. Segmentatie betekent dat we pixels willen samenvatten in segmenten. Zo kunnen we de pixels die allemaal de waarde 37 hebben een segment laten vormen. Deze segmenten kunnen groot of klein zijn en ze kunnen overal op het beeld voorkomen. Maar we kunnen ook wat variatie toestaan, door de pixels met waarden tussen 35 en 40 een segment laten vormen. Ook kunnen we kijken naar een regelmatig patroon van waarden. In figuur A zien we een segmentatie die met de 'region merging' procedure is uitgevoerd. We zien dan dat er een ruis-arm beeld ontstaat. Hiermee hebben we segmenten, maar nog geen objecten met een betekenis.

Deze objecten willen we graag identificeren uit al deze gegevens. Daartoe wordt als regel een classificatie uitgevoerd. Dat kan zowel op het oorspronkelijke beeld als op het gesegmenteerde beeld. De classificatie van het oorspronkelijke satellietbeeld laat grote en kleine objecten zien. De kleine objecten kunnen aan een bepaalde werkelijkheid voldoen, maar er is ook nogal wat misclassificatie, bijvoorbeeld veroorzaakt door pixels die sterk op elkaar lijken, of door atmosfe-

rische verstoring, onzuivere pixels, etcetera. Classificatie van het gesegmenteerde beeld laat grotere objecten zien, met minder ruis. Maar dus ook met minder kleine objecten. Dit kan wat verder van de werkelijkheid afstaan, maar kan voor interpretatie en beleidsondersteuning nuttig en waardevol zijn.

Toch zijn we er hier niet mee. Zolang de objecten scherp en helder zijn, zijn er niet zoveel problemen. We weten wat een huis is, we weten, althans hier in Nederland, wat een bos en wat een weiland is. Maar er bestaan ook minder scherpe objecten, objecten die inherent onzeker zijn. Hier komen we terecht op een punt dat wel enige aandacht vergt. We willen vaak zoveel moois uit een beeld halen, maar: zit het er allemaal wel in? In wezen zijn hier twee wegen om verder te gaan: een fuzzy benadering en een stochastische benadering.

Veel objecten, of het nu om een stad gaat, een weg, zijn moeilijk precies te onderscheiden, of hebben een vage definitie. Soms kunnen we de grenzen wel scherp krijgen, maar soms ook niet. Een 'verontreinigde rivier' is een vaag concept dat vaak geleidelijke overgaat in een 'schone rivier' (ook een vaag concept). Milieu-indicatoren zijn vage concepten – iedereen heeft wel een idee over 'global change', maar een uniforme definitie is lastig, misschien wel onmogelijk. De stochastische benadering modelleert dergelijke objecten vanuit de kansrekening, door bijvoorbeeld uit te gaan van random sets. Een fuzzy benadering neemt onzekerheid mee bij de inventarisatie en bij de behandeling. Dat geeft wel problemen, bijvoorbeeld bij de opslag van objecten in databases, bij het definiëren van operatoren, waar echter wel oplossingen voor zijn te bedenken. Ik zie dit als complementaire mogelijkheden. De fuzzy benadering heeft het voordeel dat we in de taal van alledag kunnen blijven praten. Een stochastische benadering is echter beter gefundeerd en daarmee helderder en eenvoudiger.

We gaan nu over naar het volgende voorbeeld. Figuur B laat een opname van een perceel met suikerbieten van een boerenbedrijf in Zuid-West Nederland zien. Het terrein is zo'n 7 ha groot en de opnames zijn gemaakt vanuit een onbemand vliegtuigje. De gegevens zijn enkele jaren geleden gepubliceerd in een artikel van Virginie Epinat [4]. De resolutie is hier in de orde van grootte van 1 m². Tegenwoordig is dat een resolutie die we ook met satellietopnames wel halen. Met de verschillende banden wordt gerekend, zodat de meest relevante informatie verschijnt. En hier wordt het interessant: *hoe komen we van de getallen naar de relevante informatie*. Daar-

naast is er veel detailinformatie zichtbaar: er zijn lichte en donkere strepen, er is een witkleuring, er is een patroon zichtbaar dat zijn oorzaak moet hebben, kortom we zien al zaken, maar we kunnen slechts speculeren over de oorzaken. Wat een boer belangrijke objecten vindt, misschien wel het voorkomen van gewas-stress, is voor een historisch geograaf misschien niet zo interessant: die ziet de lijnen in de lengterichting van het veld en misschien nog wel een oude verdeling tussen vroegere percelen.

Het is de door mij zo gewaardeerde literator Fernando Pessoa die in zijn *Boek der rusteloosheid* wijst op onze mogelijkheden en verantwoordelijkheden:

“De waarachtige landschappen zijn diegene die we zelf creëren, want zo [...] zien we hen zoals ze werkelijk zijn, dat wil zeggen hoe ze geschapen zijn.”

De stochastiek van ruimtelijke gegevens

In een stochastisch kader maken we een aanname over de onzekerheden in de waarnemingen. We maken dan bij voorkeur een onderscheid tussen wat we stochastisch noemen en wat we als vast beschouwen. Laten we eens enkele benaderingen nalopen.

Het eerste onderdeel heet met een mooi woord geostatistiek. Er is daarbij sprake van een meting, van toeval in die meting, van een vaste positie in het land en van ruimtelijke samenhang. Ruimtelijke gegevens bestaan uit een variabele $Z(\underline{s})$ en waarnemingen aan die variabele. Voor deze Z kunnen we de reflectie nemen zoals die wordt geregistreerd door een satellietensor, maar ook een neerslaghoeveelheid, de grondwaterdiepte, etcetera. De \underline{s} geeft dan plaats aan, bijvoorbeeld in een coördinatenstelsel. We gaan ervan uit dat we deze variabele kunnen meten, althans in principe, op een eindig aantal locaties. Het is kenmerkend dat de variabele gekoppeld is aan de ruimte. Dat levert in eerste instantie problemen op bij de analyse, want statistici willen graag een gemiddelde bepalen, een spreidingsmaat, kansen kunnen berekenen. Daartoe is het belangrijk dat we de gegevens kunnen randomiseren. Maar: een neerslagmeting op een punt is daar gemeten en niet elders. Ook is het niet zo eenvoudig om herhalingen te krijgen: de metingen zijn gedaan op een bepaald moment (of over een bepaald interval) in de tijd. Er zijn, kortom, verdere stappen nodig. We maken daartoe de stap naar *ergodiciteit*: de verwachting van het eerste en het tweede moment van $Z(\underline{s})$ kunnen worden verkregen door ruimtelijke gemiddeldes te nemen.

We zullen voorts aannames moeten doen over de *stationariteit*. Er zijn bepaalde eigenschappen van de gegevens die niet veranderen als we alle locaties \underline{s} over een zelfde afstand \underline{h} en in dezelfde richting, verplaatsen. Zo zal de verwachte waarde op een niet-bezocht punt niet veranderen, als we tenminste ook de coördinaten van dat punt over dezelfde vector \underline{h} verschuiven. We kennen drie vormen van stationariteit, in afnemende sterkte. Strikte stationariteit vereist dat de multivariate kansverdeling van de n variabelen op even zoveel locaties niet verandert als alle punten opschuiven over een vector \underline{h} :

$$P(Z(\underline{s}_1) < z_1, \dots, Z(\underline{s}_n) < z_n) = \\ P(Z(\underline{s}_1 + \underline{h}) < z_1, \dots, Z(\underline{s}_n + \underline{h}) < z_n).$$

Tweede orde stationariteit vereist een constant eerste moment en een waarbij de covarianties enkel afhangen van de afstand tussen waarnemingslocaties en niet van de lokatie zelf:

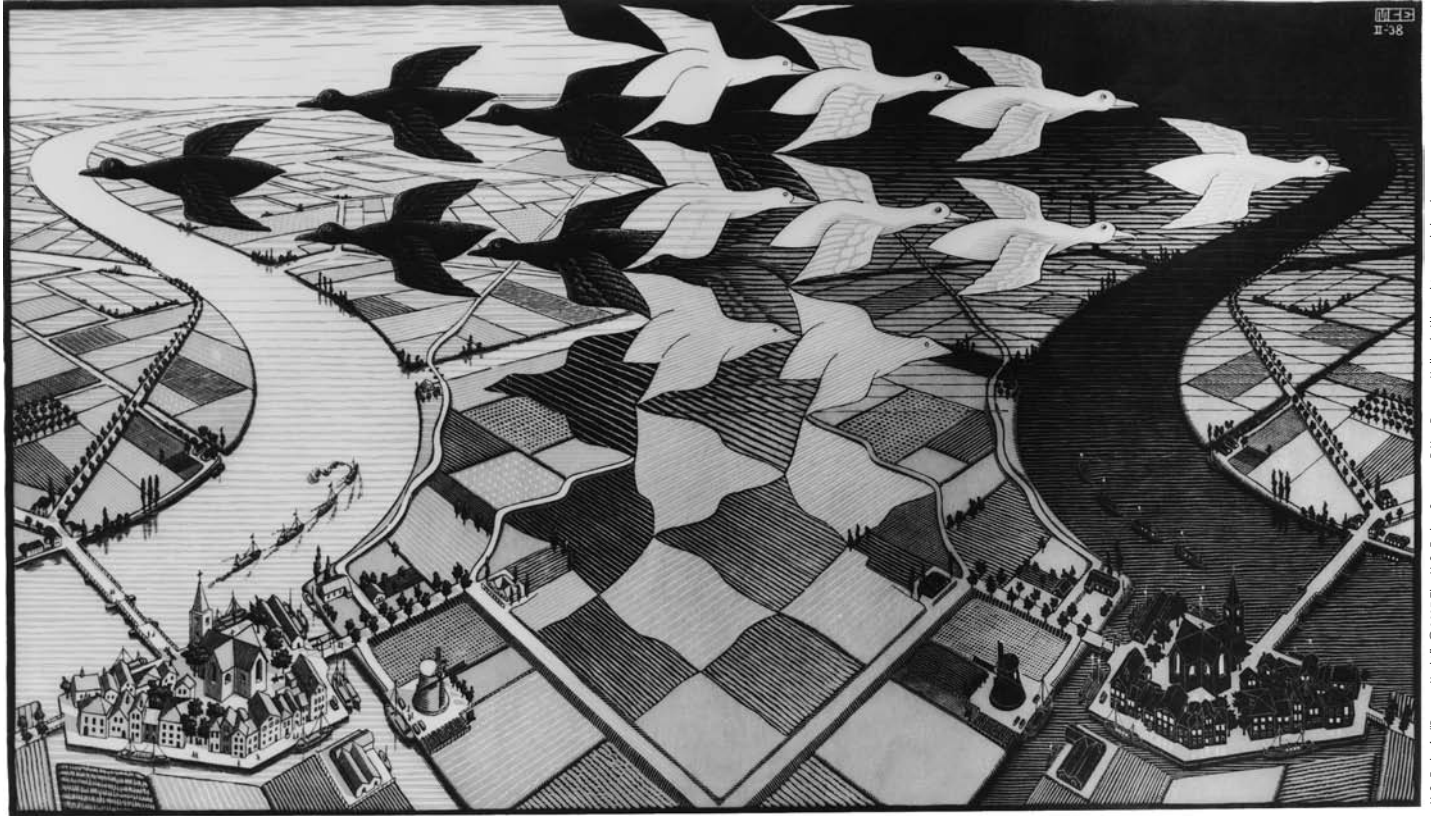
$$E(Z(\underline{s})) = \mu \\ E(Z(\underline{s}_1), Z(\underline{s}_2)) = C(\underline{s}_1 - \underline{s}_2)$$

De functie $C(\underline{h})$ wordt ook wel de covariantiefunctie genoemd. Intrinsieke stationariteit, tenslotte, is vergelijkbaar met tweede orde stationariteit, maar er wordt nu vereist dat de variantie van eenvoudige verschillen enkel afhangt van de afstand tussen waarnemingslocaties:

$$E(Z(\underline{s}_1) - Z(\underline{s}_2)) = 0 \\ E((Z(\underline{s}_1) - Z(\underline{s}_2))^2) = 2\gamma(\underline{s}_1 - \underline{s}_2)$$

waarin de functie $\gamma(\underline{h})$ het semivariogram wordt genoemd. Nog zwakkere vormen van stationariteit komen voor. Het is het concept *afstand* dat de ruimtelijke stochastiek zijn specifieke eigenschappen geeft.

Op de achtergrond speelt de theorie van stochastische toevalsfuncties. Een voorbeeld van enkele toevalsvelden zien we in figuur C. Er is daarbij sprake van een meting, van toeval in die meting, van een vaste positie in het land en van ruimtelijke samenhang. Deze benadering is gebaseerd op het baanbrekende werk van Georges Matheron in Fontainebleau in de jaren 70 en 80 [1] en voert terug op de kansmodellen van Kolmogoroff [20]. Wat we in deze ruimtelijke variatie steeds aantreffen is de variatie als functie van de afstand tussen plaatsen: waarnemingen dicht bij elkaar lijken sterker op elkaar dan waarnemingen



Figuur 1 De houtsnede Dag en nacht van M.C. Escher uit 1938

verder van elkaar af. De mate waarin deze afhankelijkheid optreedt kan verschillen per variabele.

Daarnaast zullen we een aanname moeten doen over de stationariteit. Er zijn bepaalde eigenschappen van de gegevens die niet veranderen als we alle locaties \underline{s} over een zelfde afstand \underline{h} en in dezelfde richting, verplaatsen. Zo zal de verwachte waarde op een niet-bezocht punt niet veranderen, als we tenminste ook de coördinaten van dat punt over dezelfde vector \underline{h} verschuiven. Onder bepaalde verdere aannames zal ook de kans dat een drempelwaarde wordt overschreden op dat punt niet veranderen. Er kan daarnaast sprake zijn van een globale trend, die we uiteraard in de analyses willen meenemen.

Ter illustratie kijken we naar een van de banden van een Spot5-beeld. Spot is een van oorsprong Franse satelliet. Het systeem is operationeel sinds 1986; inmiddels zijn we aangeland bij Spot5, die in mei 2002 is gelanceerd. De pixels hebben een resolutie van ongeveer 10 m. Het beeld betreft een opname uit de omgeving van Enschede (figuur D).

Ik heb nu een willekeurige rij, de 37e, geselecteerd en laat vervolgens enkele correlatiecoëfficiënten zien. De correlatiecoëfficiënt is bekende maat die relaties tussen reeksen getallen aangeeft. Een correlatiecoëfficiënt ter grootte 1 betreft een perfecte positieve line-

aire relatie (een identiteit), als deze 0 is, is er geen relatie, als deze -1 is is er een perfecte negatieve relatie. Door de waarden steeds één stapje op te laten schuiven, kan de correlatiecoëfficiënt bepaald worden als functie van de afstand tussen pixel-locaties. Ik laat het hier dus zien als een functie van de veelvouden van 10m (figuur 3).

In figuur 2 zien we steeds horizontaal de pixelwaarde aan rij 37 en verticaal de pixelwaarden aan diezelfde rij, maar dan 1, 2, 3 of 4 pixels opgeschoven. De correlatiecoëfficiënt is aangegeven als 'r'. De correlatie voor een afstand van 10m is hoog (0.948). Deze neemt dan vervolgens snel af tot waarden van 0.855, 0.780 en 0.718 voor afstanden 20, 30 en 40 m tussen de pixellocaties. Er is sprake van een ruimtelijke samenhang, die we een waarde kunnen geven. Overigens: als we dit herhalen voor een andere rij of een andere kolom zien we steeds hetzelfde patroon.

Waar zit hier de stochastiek? Deze bevindt zich in de waarde van de waarnemingen. De uitkomst van een meting wordt mede bepaald door de uitkomsten in de omgeving. De spreiding van de uitkomsten is altijd kleiner als er waarnemingen vlakbij zijn. Stelt u zich voor dat rondom u, om wat voor reden dan ook (een valse dobbelsteen, een magneet onder de tafel, aardstralen, you name it) met een dobbelsteen steeds een drie of een vier wordt

gegooid. Dan nemen we aan dat de kans dat uw worp ook 3 of 4 wordt groter is dan wat je bij onafhankelijke waarnemingen zou verwachten. Daarmee is de spreiding kleiner en kunnen we een betere verwachting uitvoeren van uw worp.

Er kan daarnaast een verband bestaan met andere variabelen [11]. In mijn voorbeeld waar ik het had over $Z(\underline{s})$ kunnen we voor de Z ook een *aantal* variabelen noteren: maximum temperatuur, minimum temperatuur, neerslag, etc. We moeten dan wel gaan kijken hoe de relaties tussen deze variabelen zijn. Vroeger, vóór de ontwikkelingen in ruimtelijke variabelen, gebruikten we de correlatie. Tegenwoordig modelleren we ook deze afhankelijkheid op basis van de afstand tussen waarnemingen. Het is het concept *afstand* dat de ruimtelijke stochastiek zijn specifieke eigenschappen geeft.

Laten nu eens kijken naar een ander belangrijk voorbeeld: methaanemissies op het eiland Java. Op Java wordt in grote hoeveelheden rijst verbouwd. Rijst is de motor van de groene revolutie geweest in de jaren '60 en vormt nog steeds het belangrijkste voedsel in veel ontwikkelingslanden. Een van de gasen die in rijstvelden ontsnappen is methaan (CH_4). Methaan is een belangrijk actor binnen global change. Het verdwijnt vanuit de bodem en het gewas in de atmosfeer. Met de groei van de wereldbevolking is de consumptie van

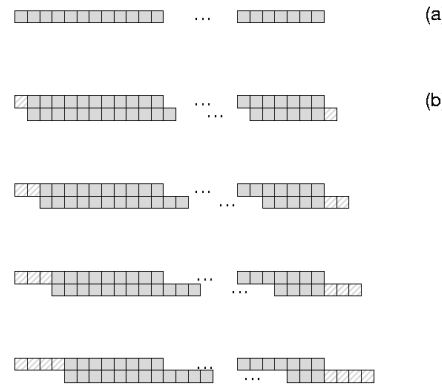
rijst en daarmee de uitstoot van methaan toe- genomen. Het gedrag is complex, de rol van verschillende bronnen is onzeker. De bijdrage van ruimtelijke stochastiek ligt erin om de ef- fecten van methaanemissie kwantitatief vast te stellen, inclusief de onzekerheden.

Ik heb voor u een kaartje gemaakt, met daarop in verschillende kleuren de hoeveel- heid methaan (figuur E). Het betreft hier zo'n vijfhonderd waarnemingen en gemodelleerde waarden op kleine veldjes. Ieder veldje meet ongeveer 30 bij 30 meter, terwijl Java 126000 km² groot is. Een eenvoudig sommetje leert dat de oppervlakte van de waarnemingen dus een 140-miljoenste van de oppervlakte be- slaat — het is alsof we op basis van 48 per- sonen een uitspraak over de wereldpopulatie willen doen. En toch is het anders: we weten al vaak wat meer van ruimtelijke processen die de methaanconcentraties verklaren. Op ba- sis van ruimtelijke afhankelijkheid kunnen we dan een kaart maken, waarbij we deze afhan- kelijkheid essentieel gebruiken. Omdat de af- hankelijkheid specifiek is voor iedere ruimte- lijke variabele, is het karteren met het mee- nemen van de ruimtelijke variatie variabele- specifiek geworden. In de tachtiger en negen- tiger jaren van de twintigste eeuw is hier al veel aan ontwikkeld, maar nieuwe uitdagin- gen zijn er voldoende.

Een nieuwe uitdaging betreft de kwaliteit van gegevens. Alle gegevens zijn onzeker, het ene wat meer dan het andere. Zo kennen we onzekerheid *in de meetlocaties*, een *beperkt aantal* meetpunten en onzekerheid bij het modelleren van de *ruimtelijke samenhang*. In ons voorbeeld kunnen we de onzekerheid ver- kleinen door gebruik te maken van een koppe- ling tussen methaan gehalte aan landgebruik, enkele bodemeigenschappen en ook aan de tijd van het jaar.

Recent is de promovendus Iswar Das be- gonnen met het vervaardigen van risicokaar- ten. Op zich is dat niet nieuw, want we kun- nen al lang kaarten produceren waarop een percentiel, of de voorwaardelijke verwachting wordt weergegeven dat een limiet wordt over- schreden. Maar we realiseren ons steeds be- ter dat 'risico' een lastig concept is. Iswar richt zich op 'landslides': het risico betreft dan bij- voorbeeld de kans op voorkomen van zo'n modderstroom en de effecten daarvan op een gebouw, waar al dan niet mensen verblijven.

Een derde uitdaging betreft validatie. Van oudsher is een kaart zo goed als het de werke- lijkheid weerspiegelt — 'the proof of the pud- ding is the eating'. De gebruiker speelt hier- bij een belangrijke rol. In een recente studie uitgevoerd samen met het Wageningse be-



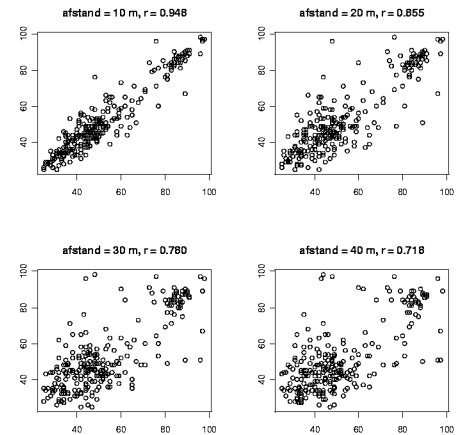
Figuur 2 De oorspronkelijke pixels uit de 37^e rij (a) en de pixels met een kopie ervan opgeschoven over 1, 2, 3 en 4 pixels (b)

drijf Synoptics, inmiddels overgenomen door Vexcel, dat op zijn beurt weer door Microsoft is overgenomen, hebben we de validatie on- derzocht door verschillende gebruikers van een hoogtekkaart te definiëren [9]. We kwa- men daarbij tot een aantal verrassende in- zichten. Maar fundamenteel is wellicht een inherent onvermogen om te valideren — hoe kunnen we ooit onze global change modellen valideren? En ook is het laatste woord over het valideren van risicokaarten nog lang niet gesproken.

Punt patronen

In de tweede vorm van ruimtelijke analyse be- kijken we punten, waarbij de eigenschap die willen analyseren, bekend is, maar de *locatie* aan toeval onderhevig is. De stochastische onderbouwing komen we tegen in de litera- tuur van met name Dietrich Stoyan in Freiberg [21], Adrian Baddeley in Perth, Australië [7] en Peter Diggle in Lancaster [15]. We zullen vaak spreken over een ruimtelijk proces dat een patroon genereert.

We maken hierbij van oudsher een onder- scheid tussen drie vormen: een volledig toe- vallig patroon, een regelmatig patroon en een geclusterd patroon (zie figuur 4). Van te voren weten we vaak niet wat we aan zullen treffen. Inzicht in het type proces kan verhelderend zijn om de achterliggende fenomenen beter te begrijpen. Opnieuw is hier sprake van een afstandsverdeling, maar nu tussen puntob- jecten. De stochastiek in dergelijke analyses bevindt zich in de verdeling van de afstan- den tussen de objecten. In een volledig toe- vallig patroon volgen de afstanden een the- oretische functie, in een regelmatig patroon komen sommige afstanden vaak voor, terwijl andere afstanden helemaal niet voorkomen, terwijl in een geclusterd patroon betrekkelijk korte afstanden veelvuldig voorkomen, terwijl langere afstanden misschien wat minder vaak



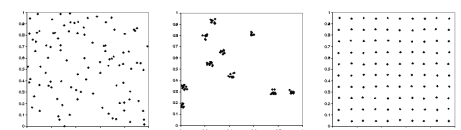
Figuur 3 Correlaties tussen pixelwaarden voor afstanden van 10 m (linksboven), 20 m (rechtsboven), 30 m (linkson- der) en 40 m (rechtsonder) tussen locaties. Duidelijk is de afname te zien van de correlaties bij toenemende afstand.

voorkomen. Dat modelleren we bijvoorbeeld met een G-functie.

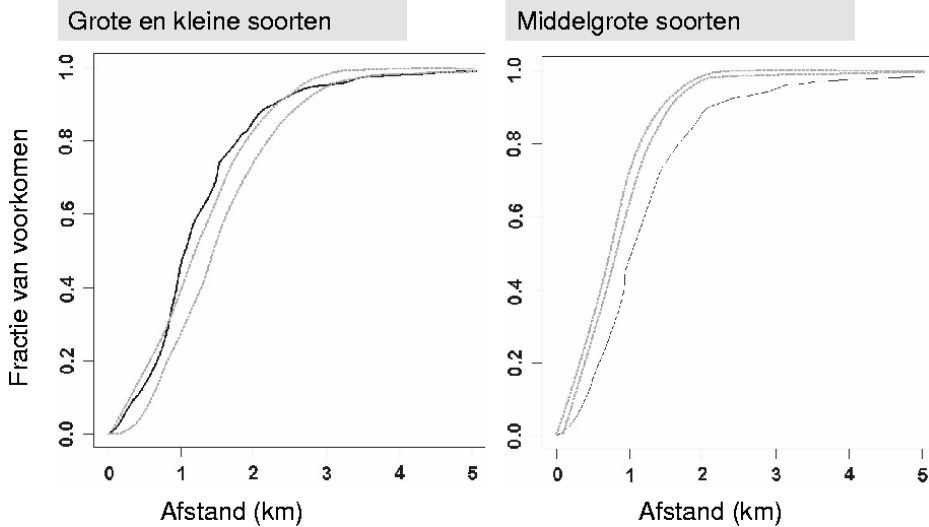
We bekijken een patroon X dat onder an- dere punten x en y zou kunnen bevatten. $G(r)$ is de verdelingsfunctie van de afstanden van een punt tot zijn dichtstbijzijnde buur, $G(r) = Pr(\rho(x, y) \geq r)$ voor $r \geq 0$. Hier- bij is ρ de Euclidische afstand tussen twee punten. $G(r)$ kunnen we interpreteren als de voorwaardelijke verwachting van de overige punten in X gegeven een punt op een wille- keurige lokatie. Een heuristische interpretatie is $1 - G(r)$ de kans is dat een schijf met straal r op een willekeurig punt geen ander punt van het patroon X bevat.

Een voorbeeld dat we recent geanalyseerd hebben betreft de verspreiding van kudde grote grazers in een Afrikaans savanneland- schap [10]. Het Laikipiapark is een natuurlijk landschap dat niet de status van een natio- naal park heeft. Doel van dit onderzoek was de vraag of we de patronen van verspreiding van kudde wilde dieren konden begrijpen, bijvoorbeeld op basis van ecologische gron- den.

In figuur 5 zien we links een G-functie gegenereerd door een patroon van herbi- voren waarop niet door predatoren gejaagd wordt en rechts een die van een patroon van kudde waarop predatoren wel jagen. De gejaagde soorten laten een regelmatig pa- troon zien, terwijl de niet-gejaagde soorten



Figuur 4 Drie gesimuleerde puntpatronen. Een volledig random patroon (links), een geclusterd patroon (midden) en een regelmatig patroon (rechts)



Figuur 5 G-functie voor grote en kleine dieren (links) en voor prooidieren (rechts) in het Laikipia park, Kenya

een clustering laten zien. De ecologische verklaring kan dan zijn dat de eerste groep preferentie van voedsel centraal stelt, terwijl de tweede groep een soort ‘risicominimalisatie’ laat zien.

Bij het modelleren van dergelijke processen zijn we steeds beter in staat om ook niet-stationariteit mee te nemen. Zo kan de aanwezigheid van dorpen aan een kant van het landschap invloed hebben op de verdeling van de afstanden, net zoals aanwezigheid van voldoende voedsel en water. Het nieuwe element zit er dan in dat we de afstandsverdeling koppelen aan co-variabelen.

Aspecten van gegevenskwaliteit zijn er volop: ondanks GPS zijn de plaatsen van voorkomen soms onzeker, met daarbij onzekere afstanden tussen de objecten. Het is vaak maar de vraag of alle objecten helder gedefinieerd zijn, of we alle objecten wel te pakken hebben, tellen we sommige objecten niet twee keer, er kan sprake zijn van een weinig uniforme waarneming, enzovoort.

Rooster data

De derde vorm van ruimtelijke stochastiek betreft de kwantitatieve analyse van gebiedsdelen. Hier zijn we geïnteresseerd in verklarende kenmerken, we hebben te maken met vastliggende eenheden, zoals gemeentes, provincies, landen, biotopen, catchments of bodemeenheden. Het concept van een afstand is lastiger — immers: wat is de afstand tussen twee provincies die aan elkaar grenzen? We werken in dergelijke studies daarom liever met een ‘buurt’-relatie. Buurten grenzen aan elkaar of niet. Dat kun je aangeven met een 1 of een 0, maar we kunnen dat ook aangeven met een fractie. Met dit laatste creëren we wel asymmetrie: de fractie van de grens van Ne-

derland met Duitsland is veel groter dan de fractie van de grens van Duitsland met Nederland. Het gaat bij dergelijke analyses om het vinden van een goed verklarend model, een regressiemodel in de aanwezigheid van ruimtelijke relaties. Autoregressieve processen spelen dan een rol. In de jaren tachtig hebben we belangrijke bijdragen gezien, met name het beroemde artikel van Geman en Geman [2] en een reeks publicaties van Julian Besag (onder andere [3]).

Een voorbeeld van een dergelijke studie betreft de sterfte aan Buruli Ulcer, onderzocht door Alfred Duker in zijn promotiewerk [14]. Buruli Ulcer is een zwerenziekte die wereldwijd veel voorkomt. Het is een aan lepra gerelateerde ziekte die gekenmerkt wordt door de vorming van grote zweren op de huid en het uitblijven van een immunologische reactie. Het komt vooral voor in tropische gebieden in de wereld, zoals in verschillende landen in West Afrika. Het komt steeds meer voor en is na lepra en tbc inmiddels de derde oorzaak van mycobacteriele infecties bij gezonde mensen. De ziekte lijkt vooral voor te komen als er veranderingen in het milieu hebben plaatsgevonden, zoals de ontwikkeling van wateropslag, het winnen van zand of irrigatie. In een aantal dorpen (settlements) in Ghana hebben we de sterfte aan de zwerenziekte Buruli Ulcer gerelateerd aan het voorkomen van arsenicum in het milieu. Er was mogelijk een relatie met de afstand tot een mijn in de buurt, met de hoogte in het landschap en de kwaliteit van bodem en grondwater. In een bodemmodel vonden we significantie tussen de sterfte aan deze ziekte en het arsenicum in de bodem. In een watermodel vonden we significantie tussen de sterfte

aan deze ziekte en de afstand tot de mijn en de relatieve hoogte.

De stochastiek bevindt zich op dezelfde plaats als in het geostatistische voorbeeld: in de waarneming zelf. De locatie ligt vast. Buurtrelaties worden in het voorbeeld uit Afrika gegeven door de grenzen van de nederzetting — andere administratieve grenzen zijn in deze landen vaak niet zo helder.

Schaal

Een belangrijk aspect binnen de ruimtelijke stochastiek betreft ‘schaal’. In eerste instantie lijkt dat vrij triviaal — maar er zijn momenten dat het toch even anders wordt. Neerslaggegevens op de nationale schaal laten een totaal ander beeld zien dan op de schaal van een individueel perceel. Veranderingen in schaal hebben onze aandacht. In een stochastische benadering bekijken we gegevens met een bepaalde resolutie. Dat kan een puntwaarneming zijn, maar ook een veldopname van 900 m². We willen dan opschalen of neerschalen naar een andere resolutie.

We gaan uit van een waar patroon x en we bekijken $Z(\underline{s}_i)$, voor $i = 1, 2, \dots, n$. We veronderstellen dat de gegevens uit een Gaussische verdeling komen en onafhankelijk zijn, gegeven x , met a priori variantie $\kappa - 1$ en een verwachting die lineair is in x :

$$Z \sim N(Hx, (\kappa - 1)I).$$

De $H_i(\underline{s})$ vormen de support van de gegevens: $H_i(\underline{s})$ is bijvoorbeeld constant rond \underline{s}_i en elders. Voor opschaling van de gegevens hebben we interesse in x over een groot gebied S . Op een andere resolutie bekijken we dan lineaire functionalen A op x :

$$Ax = \left(\int_{\underline{s} \in S} A_j(\underline{s})x(\underline{s})d\underline{s} \right)_{j=1}^J$$

Het model dat u hier ziet koppelt de resolutie op een schaal van waarnemen met die op een andere. Dat levert een stochastisch model op waarin we op basis van normaliteit kunnen schalen en uitspraken over onzekerheid kunnen kwantificeren. Hiermee kunnen we van de gemeentelijke schaal naar de provinciale schaal gaan, maar ook uitspraken de andere kant uit doen: disaggregatie van de provinciale naar de gemeentelijke schaal.

Ruimte-tijd

We zullen nu aandacht besteden aan de stochastiek in ruimte en tijd. De Franse filosoof Bergson vroeg zich of ruimte en tijd aan elkaar

gelijk gesteld kunnen worden [18]. Hij maakt een onderscheid tussen de verleden tijd en de huidige tijd. Hij geeft daarbij aan de tijd niet zozeer te zien als een soort ruimte waarin zich de processen afspelen, maar het zich meer als een scheppende, dynamische kracht voor te stellen en de ruimte als het podium waarop zich dit afspeelt. Dat is wat mij betreft wat te simplistisch, want ik zie de ruimte wel degelijk als een scheppende kracht.

Een goed uitgangspunt in de tijd is dat wat morgen gebeurt vermoedelijk meer lijkt op wat vandaag gebeurt dan op wat gisteren gebeurde. Processen positioneren we toch liefste in de tijd, met een duidelijk oorzaak en gevolg. In de ruimte, we hebben dat al gezien gaat het vooral om de afstand. Bij het analyseren van ruimte-tijdprocessen modelleren we zowel afhankelijkheid in de tijd als afhankelijkheid in de ruimte. Deze combinatie levert daarmee toch wel wat lastige problemen op – er is bijvoorbeeld geen natuurlijke manier om de eenheden (meter en dag) te verbinden. Bij een stochastische beeldanalyse kunnen we een reeks van beelden vergelijken die allemaal hetzelfde proces of fenomeen laten zien op verschillende momenten. De beelden moeten precies over elkaar heen vallen en de beelden moeten onder vergelijkbare omstandigheden zijn opgenomen. Figuur B, het eerdere voorbeeld van het perceel met suikerbieten, laat zo'n reeks van beelden zien. Binnen de stochastiek is het van belang objecten op deze beelden te koppelen en verschillen aan te geven en te interpreteren.

In ruimte en tijd analyses proberen we stochastische beeldanalyse te koppelen aan de dynamiek in de tijd. We hebben bij het International Institute for Geo-Information Science and Earth Observation (ITC) de beschikking over Meteosat beelden die iedere 15 minuten binnenkomen. Deze hoge temporele resolutie gaat ten koste van de ruimtelijke resolutie. We hebben deze beelden gebruikt bij het monitoren van bosbranden, hun bewegingen en het voorspellen ervan op een dag in Spanje [6]. In dit stuk van de wereld is de ruimtelijke resolutie ca. 3 km, dat wil zeggen iedere pixel is de gemiddelde reflectiewaarde over een eenheid van 9 km². Figuur F laat een ruimte-tijd kubus zien van bosbranden. Horizontaal is de ruimte geprojecteerd, verticaal de tijd. Met een wolkenvrije lucht kun je dan een redelijk model vaststellen.

Verdere voorbeelden zien we uiteraard op het epidemiologisch vlak waar we de verspreiding van ziektes in ruimte en tijd bestuderen. Ik zou nog even door kunnen gaan, maar ik denk niet dat dat zo relevant is. Ik wil eigen-

lijk overgaan naar een wat meer prangende vraag: waar doen we het allemaal voor. Met andere woorden: hoe kunnen we op een verstandige en creatieve manier omgaan met deze onzekerheden en er misschien zelfs wel ons voordeel mee doen.

Beslissingsondersteuning

Een belangrijk element in al dit werk bestaat uit het ondersteunen van beslissingen. Dat is vaak het speuren om het beste resultaat te halen. In het verleden, in het promotiewerk van Van Groenigen [13], hebben we aspecten van simulated annealing gebruikt om bijvoorbeeld optimale bemonsteringsschema's te vinden. Maar beslissingsondersteuning gaat verder. Recent heeft Van de Vlag in zijn promotiewerk aangetoond hoe vage eenheden (duinen en strand) en beslissingen voor zand suppletie gecombineerd kunnen worden [12]. Hij heeft laten zien hoe dat tot beter onderbouwde beslissingen kan leiden, door beter om te gaan met de natuurlijke, onzekere, eigenschappen van systemen en daar een goed gefundeerde analyse aan te koppelen. Verder wegen op het gebied van Bayesiaanse netwerken proberen we op dit moment nog verder richting te geven.

Recente ontwikkelingen

Ontwikkelingen in het opnemen en verspreiden van digitale beelden gaan steeds sneller. De mate van detaillering wordt alleen maar groter, evenals de frequentie van waarnemen. De snelle opmars van internet met een snelle en efficiënte toegang tot een veelheid van beelden en (ruimtelijke) informatie en de combinatie van digitale camera, de mobiele telefoon en navigatiesystemen zijn de belangrijkste oorzaken. Daarmee groeien de mogelijkheden en worden de vragen die we kunnen stellen en beantwoorden steeds gevarieerder. Het bekendste voorbeeld is wellicht Google Earth, waarbij we op verschillende niveaus kunnen inzoomen op – in principe – ieder stukje van de wereld. Het zal niet lang meer duren of Google Earth is eenvoudig toegankelijk op de mobiele telefoon. Maar dit zijn gegevens, getalletjes. Hiermee hebben we nog niet de informatie waar we echt wat mee kunnen doen.

Een bijzonder aspect betreft de verandering van de vragen die we nu kunnen stellen. Vonden we het vroeger prachtig als we een enkel beeld hadden van een gebied, nu gaat het om meer dan alleen maar mooie plaatjes. De informatie die we weergeven moet een duidelijk doel dienen. En naarmate de informatie toeneemt wordt het doel specifieker.

Een stochastisch element dat we telkens langs zien komen betreft de kwaliteit van gegevens. Kwaliteit is een breed begrip, maar is in de stochastische literatuur wel beschreven. Kwaliteit hangt nauw samen met het gebruik van de gegevens. Recent hebben we bij het ITC een congres georganiseerd waar we aan deze aspecten uitvoerig aandacht hebben besteed.

Het blijkt steeds vaker te lonen om voorinformatie te gebruiken. In tegenstelling tot financiële transacties is dat in de stochastiek een heel gangbare en geaccepteerde werkwijze. We noemen dat een Bayesiaanse analyse. Voorinformatie kan gebruikt worden om een eerste indruk te verkrijgen, om subjectieve kennis objectief te maken en te gebruiken, etc. Gemaakte waarnemingen passen dan deze voorinformatie aan.

Visie

Na het presenteren van de twee thema's beeldanalyse en ruimtelijke stochastiek is het tijd voor het contrapunt: mijn visie op het vakgebied. In mijn visie komen de beelden, de vragen die we aan beelden stellen en de mogelijkheden die moderne stochastische methoden bieden om tot bevredigende antwoorden te komen steeds dichter tot elkaar. Onzekerheid is inherent en het is aan ons om, in de woorden van Spinoza: zo veel mogelijk de baas over het toeval te worden en ons handelen op een redelijk plan te baseren.

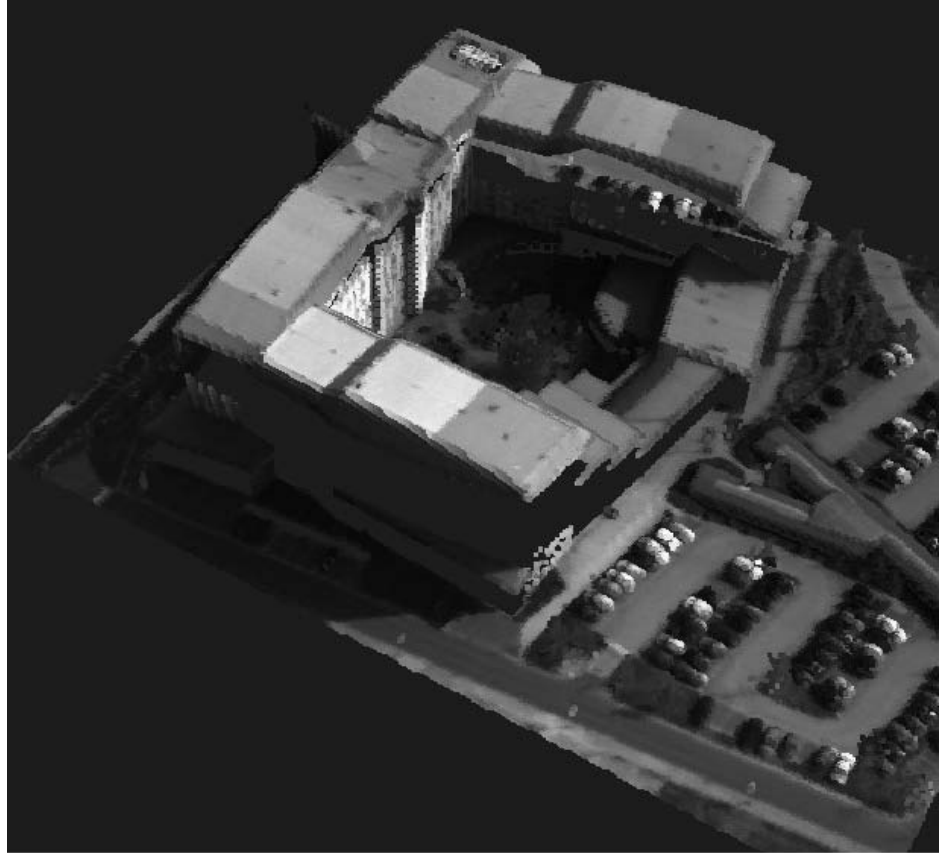
De belangrijkste vraag is wat mij betreft: hoe kunnen we niet alleen *waarnemen* in de ruimte maar hoe kunnen we procedures bedenken om de ruimte met al haar fenomenen *begrijpen*. De kennistheorieën van Kant en Locke vormen daarbij een interessant startpunt. Hebben we 'a priori' informatie, is er zoiets als een 'vaag object an sich', en zo ja: hoe kunnen we daarmee omgaan. Ik zie de stochastiek hierbij, naast bijvoorbeeld proceskennis en systematische inventarisatie, als een onafhankelijke richting in de wetenschap. Het is een veld met een eigen terminologie, maatschappelijke toepassingen en validatiemogelijkheden.

Ik zie veel mogelijkheden bij een verdere integratie van beeldanalyse met stochastische methoden. Stochastische meetkunde, stereologie, multivariate analyses van heel veel banden, het ontdekken van extremen (vooral in vegetatie en ondergrondstudies van belang), waarbij we niet alleen de onzekerheid vaststellen, maar het misschien ook wel als een groot goed zien en het kunnen gebruiken.

Integratie tussen modellen en beeldinformatie gaat steeds verder. Met modellen bedoel ik dan bijvoorbeeld transportmodellen, landbouwkundige modellen en milieumodellen. Dergelijke wiskundige modellen kunnen in principe van beelden gebruik maken. De verschillen tussen de beschikbaarheid van informatie en de vragen bij het gebruik ervan liggen nog steeds ver uit maar komen stilaan wel tot elkaar. Hier is echter nog steeds een lange weg te gaan.

Dat brengt ons aan de vraagkant van deze methoden en technieken: wie zit er op ons te wachten? Voorbeelden te over, los van de voorbeelden die ik al heb laten zien. In de dagelijkse omgeving hebben we het fileprobleem, dat we vermoedelijk niet kunnen oplossen maar wel kunnen verlichten als real-time beelden kunnen worden gekoppeld aan actuele weginformatie. In de preciselandbouw speelt het al langer een grote rol, door vroegtijdig aan te geven waar gewasproblemen optreden en waar de boer moet ingrijpen. In het milieu kunnen we beter ingrijpen als we weten welk schip welke verontreiniging heeft veroorzaakt, hoe de kwaliteit van lucht zich ontwikkelt en via een intelligent systeem is wellicht ook informatie over de kwaliteit van bodem en grondwater te verkrijgen. Als het gaat om uitbreidingen van steden, over beleidsondersteuning om wetlands te redden, landbouwgrond toe te wijzen, stedelijke ontwikkelingen te monitoren: het is het wachten op een verstandig gebruik van de juiste beelden om substantieel verder te kunnen komen.

Maar overal is er sprake van onzekerheid. Recent heeft het ITC een helikopter over Enschede laten vliegen. Deze heeft uiterst precieze lasergegevens opgenomen die op hun beurt kunnen worden vertaald in zeer gedetailleerde hoogte-informatie met een punt-dichtheid van 12 tot 25 punten per m^2 . Bij elkaar zijn er misschien wel meer gegevens dan Mozart noten heeft geschreven (en die had daar geen helikopter bij nodig!). U ziet in figuur 6 een opname van het ITC-gebouw. De auto's kun je onderscheiden, maar de takken van de bomen op de binnenplaats nog net niet. Soms is het zelfs mogelijk om *in* de auto te kijken. Als we als boef in een auto wensen in te breken is de informatie gelukkig onvoldoende. Dit is interessant voor juristen in het gehoor. Meer serieus: als we als stedelijke dienst de hoeveel hout aan een boom willen vaststellen dan blijft er een grote mate van onzekerheid. Kortom: we kunnen weer zoveel meer, vooral vragen beantwoorden die vroeger niet te beantwoorden waren. Zoals: vragen met een sterke tijdscomponent



Figuur 6 Een laserbeeld van het ITC-gebouw, opgenomen met een dichtheid van 12 tot 25 punten per m^2

of die een nog hogere resolutie nodig hebben. Het is hier dat ik de komende jaren een flinke vooruitgang verwacht.

Er zijn verschillende technieken om hier wat verder mee te gaan. Ik denk daarbij aan wavelets, Bayesiaanse netwerken, maar ook aan het modernere 'boosting', kortom aan veel aspecten die we tegenkomen binnen het gebied van machine learning.

Geef me de ruimte

Met ruimte hebben we allemaal te maken. De ruimte wordt schaars en een zorgvuldig omgaan met deze schaarse ruimte is een essentiële opgave voor ons allen. Waar vinden we de rust en de stilte, waar is voldoende voedsel, waar is schone lucht waar onze kinderen en kleinkinderen en vele generaties daarna recht op hebben? Niet alleen over honderd of tienduizend jaar, maar zoals de hooggeleerde Kwee het ooit formuleerde: ook nog over een miljoen jaar [8]. Het omgaan met onzekerheden, zoals ik hier heb laten zien, vraagt een kwantitatieve, stochastische benadering. Maar het vraagt ook een kwaliteit van onze bestuurders — om te kunnen gaan met kansen en met onzekerheden. Mede om die reden is dit werkveld zo de moeite waard.

Circa één jaar geleden ben ik met veel enthousiasme begonnen aan mijn aanstelling bij de Universiteit Twente. Het is echter ook al weer bijna twaalf jaar geleden dat ik voor het eerst een oratie hield — in de tussentijdse periode was ik hoogleraar in Wageningen. Toen formuleerde ik het als de grootste uitdaging om tot een integrale benadering van alle ruimtelijke statistiek te komen. Hoewel we een heel eind zijn, ben ik geneigd dat nu toch wat te classificeren onder de noemer 'jeugdige overmoed'. Mijn ambitie met de huidige leerstoel is wat bescheidener: graag wil ik proberen te werken aan het uitbreiden van de mogelijkheden van een stochastische analyse van beelden. Theoretische kennis bestaat of ontwikkelen we bij de universiteit, de problemen en maatschappelijke uitdagingen vinden we in al hun variatie binnen het ITC en haar internationale netwerk. Met de hooggeleerde Bagchi en met verschillende medewerkers van de leerstoelgroep stochastische systeem- en signaaltheorie en van de leerstoelgroep kansrekening statistiek en verschillende collega's binnen de onderzoekschool Centrum voor Telematica en Informatie Technologie (CTIT) heb ik al een aantal inspirerende gesprekken kunnen voeren. Er is veel kennis voorhanden op het vlak van de tem-

porele stochastiek — deze richt zich vooral op patronen in de tijd. De waarde van aandelen, prijsontwikkelingen, temperatuurschommelingen, etcetera. De ruimte vormt daarbij een onmisbare component.

Mijn eigen groep aan het ITC, de staf en de promovendi, maar ook mijn collegae: zij allen dragen bij aan de verdere ontwikkeling van het vakgebied. Contacten met het bedrijfsleven zijn zeer vruchtbaar en inspirerend. Ook met de Universiteit Wageningen onderhoud ik verschillende levendige contacten. Binnen Nederland is met name de vereniging voor statistiek en operationele research belangrijk. Daarnaast kennen we het netwerk van stochastici met hun inmiddels legendarische bijeenkomsten onder de vallende blaadjes in Lunteren.

Internationaal liggen er grote mogelijkheden. Europees, uiteraard met universiteiten en instituten in vele landen en met mogelijkheden in oost-Europa. Verspreiding van kennis via een wetenschappelijk tijdschrift: het International Journal of Applied Earth Observation and Geoinformation (JAG). Interessante samenwerking hebben we al langer met India, met name met het Indian Institute of Remote Sensing. We bestuderen nu ook de ontwikkeling van theeplantages, en proberen patronen van bomen en bossen op verschillende schalen te modelleren met fractale methoden. Samenwerking met China gaat steeds verder en wordt concreter. Binnenkort gaan we met de universiteit in Wuhan op PhD-niveau aan de slag. We gaan kijken naar het monitoren van onzekere objecten. Verder

gaan we stochastische beelden analyseren gekoppeld aan het monitoren van de droogte in Oost-Afrika. Daarnaast hebben we lopende onderwijscontacten met instituten in Tanzania en Nigeria, terwijl samenwerking met Zuid-Afrika begint te lopen.

Ik blijf het fascinerend vinden om met buitenlandse collega's en studenten samen te werken, verschillende culturen en wetenschappelijke benaderingen te ervaren en gezamenlijk tot een mooi resultaat te komen. De grootste fascinatie is dan vooral de specifieke vraagstelling, ingegeven door lokale ontwikkeling, lokale voorwaarden en lokale mogelijkheden. Zo is in West-Europa de milieuregeling al ver, terwijl we in Nigeria nog moeten proberen om de eerste milieueffecten van de oliepijpleidingen in kaart te krijgen.

Referenties

- Matheron, G. 'The intrinsic random functions and their applications', *Advances in applied probability*, 5, 439 – 468, 1973
- Geman, S., and Geman, D. 'Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721 – 741, 1984
- Besag, J. 'On the statistical analysis of dirty pictures', *Journal of the Royal Statistical Society*, B-48, 259 – 302, 1984
- Epinat, V., and Stein, A., and De Jong, S.M., and Bouma, J. 'A wavelet characterization of high-resolution NDVI patterns for precision agriculture', *International Journal of Applied Earth Observation and Geoinformation*, 3, 121–132, 2001
- Stein, A., and De Beurs, K. 'Map indices to quantify semantic accuracy in segmented Landsat images', *International Journal of Remote Sensing*, 26, 2937–2951, 2005
- Umamaheshwaran, R., and Bijker, W., and Stein, A. 'Image mining for modeling of forest fires from Meteosat images', *IEEE Transactions on Geoscience and Remote Sensing*, 45, 246–253, 2006
- Baddeley, A.J. 'Combinatorial foundations of stochastic geometry', *Proc. London Math. Soc.*, 42, 154 – 177, 1981
- Kwee, S.L. 'Mens, techniek en toekomst', *Wijzerig perspectief op maatschappij en wetenschap*, 22, 46 – 53, 1981
- Kyaruzi, J.K. *Quality assessment of DEM from radargrammetry data: quality assessment from the user perspective*, ITC, 2005
- Khaemba, W.M. *Development and application of spatial and temporal statistical methods for improved sampling of wildlife*, Wageningen University, ITC, 2000
- Stein, A., *Spatial interpolation*, Wageningen University, 1991
- Van de Vlag, D.E. *Modeling and visualizing dynamic landscape objects and their qualities*, Wageningen University, ITC, 2006
- Van Groenigen, J.W. *Constrained optimisation of spatial sampling: a geostatistical approach*, Wageningen University, 1999
- Duker, A. *Spatial analysis of factors implicated in *Micobacterium ulcerans* infection in Ghana*, Wageningen University, 2005
- Diggle, P.J. *Statistical Analysis of Spatial Point Patterns – 2nd edition*, Arnold Publishers, London, 2003
- Cox, D. *Principles of Statistical Inference*, Cambridge University Press, 2006
- Glasbey, C.A., and Horgan, G.W. *Image analysis for the biological sciences*, Wiley, Chichester, 1995
- Bergson, H. *Time and Free Will*, Dover Publications, New York, 2001
- Bergson, H. *Durée et simultanéité*, Presses Universitaires de France, 1921
- Kolmogoroff, A.N. *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, Springer-Verlag Berlin, 1933
- Stoyan, D., and Kendall, W.S., and Mecke, J. *Stochastic geometry and its applications*, Wiley, New York, 1995