

Milieu, geostatistiek en GIS.

A. Stein¹

Abstract

Geostatistics are applied within a Geographical Information System (GIS) to study the extent and the degree of soil contamination. Environmental variables show often a strong spatial correlation. Spatial patterns, including their spatial uncertainty can be visualized by means of spatial interpolation from points to areas of land. Moreover, conditional simulation of spatial patterns allow a sensitivity analysis of models that describe the movement of pesticides and fertilizers within soils to the ground water. Finally, comparison of the effects of different land use scenarios allows an efficient use of spatial variability procedures. For all these activities, a GIS is an indispensable tool, e.g. combine different sources of information. In this paper, an overview of geostatistical interpolation procedures is provided, based upon linear regression theory. To determine exceedance probabilities of critical threshold values, use is made of indicator variables, that take the values 0 and 1 only, for which the spatial interpolation crystallizes into indicator kriging. The study is illustrated with three different examples: the first study focuses on the concentration of heavy metals in *natural* circumstances, the second study focuses on a soil stress analysis to determine how much the soil may be burdened with nitrate, the third study focuses on an inventarisatie of a 4 ha urban area which is polluted with heavy metals.

Keywords: geostatistics, GIS, environment, indicator kriging, exceedance probabilities.

¹A. Stein, Landbouw Universiteit, Postbus 37, 6700 AA Wageningen, tel. 08370-82420. Deze studie is mede mogelijk geweest dankzij samenwerking met de ingenieursburo's Oranjewoud en IWACO en met de provincie Noord-Brabant. De steun van de studenten Harry-Jan Hendrickx Fransen en Chris Varekamp en van A.C. van Eijnsbergen van de vakgroep Wiskunde van de LUW heb ik zeer op prijs gesteld. Kritische kanttekeningen van L.C.A. Corsten hebben geleid tot een grondige herziening, inkorting en verbetering van het manuscript.

1. Inleiding

Bij veel milieuproblemen worden gegevens verzameld die gebonden zijn aan de plaats waar ze zijn waargenomen. Voorbeelden hiervan zijn gehalten aan zware metalen en organische koolwaterstoffen, zoals PAK's en BAP-equivalenten, maar ook metingen aan het nitraatgehalte dat naar het grondwater uitspoelt. Statistische vragen bij het verzamelen en het bewerken van de gegevens spelen een steeds belangrijker rol. In de eerste plaats is men vaak geïnteresseerd in het computermatig vervaardigen van kaarten, inclusief de onzekerheden, bij voorkeur in een 3-dimensionaal perspectief (Staritsky *et al.*, 1992). In de tweede plaats zijn er voor veel verontreinigende stoffen in het milieu kwantitatieve normen, en is het relevant om de kansen te bepalen dat deze normen worden overschreden (Stein, 1993). In de derde plaats is het voor veel beleidsinstanties van belang om effecten van scenario's voor landgebruik te visualiseren en met elkaar te vergelijken (Finke en Stein, 1994). Tenslotte is een belangrijk punt dat waarnemingen zo verzameld worden dat de bovenstaande punten zo efficiënt mogelijk beantwoord kunnen worden (Corsten en Stein, 1994). Hier kan met geostatistische methoden zinvol gewerkt worden.

Een belangrijk aspect bij milieuproblemen is dat er een mogelijke relatie bestaat met andere informatie, zoals historische informatie en informatie weergegeven op bodemkaarten, hydrologische kaarten en topografische kaarten (Stein, 1994). Het gedrag van verontreinigingen in de bodem is tevens vaak redelijk tot goed te modelleren met behulp van kwantitatieve modellen (Bregt, 1993). Gedurende de afgelopen jaren hebben we een snelle opkomst van geografische informatie systemen (GIS) kunnen constateren. Met deze systemen kan betrekkelijk eenvoudig een combinatie gemaakt worden tussen enerzijds kaarten die met geostatistische hulpmiddelen vervaardigd zijn en anderzijds beschikbare voorinformatie (Burrough, 1986; Heuvelink en Burrough, 1993). Deze systemen zijn bij uitstek geschikt om grote hoeveelheden gegevens te verwerken en grafisch weer te geven.

In deze studie wordt ingegaan op een aantal aspecten van geostatistisch onderzoek dat binnen een GIS wordt uitgevoerd in het kader van het milieu. Allereerst wordt een overzicht gegeven van geostatistische methoden. Vervolgens worden drie praktische cases geanalyseerd op hun mogelijkheden. De studie zal

worden afgesloten met enkele conclusies en aanbevelingen voor toekomstig statistisch en geostatistisch onderzoek in het kader van het milieu.

2. Van OLS-regressie tot kriging

OLS-regressie

Een veel gebruikt model voor de verwachting van een variabele y is het lineaire regressiemodel. De verwachting van y wordt geschreven als functie van k onafhankelijke variabelen $x_1, x_2, \dots, x_k$. In het bijzonder kunnen we denken aan functies van de coördinaten van waarnemingslocaties, zoals machten van de coördinaten. Als voor een studie in twee dimensies, naar bijvoorbeeld de verspreiding van het cyanide-gehalte, een element van de waarnemingsvector correspondeert met een plaats die coördinaten s_1 en s_2 heeft en een kwadratisch model gekozen wordt, dan is k gelijk aan 6: $x_1 = 1$, $x_2 = s_1$, $x_3 = s_2$, $x_4 = s_1^2$, $x_5 = s_1 s_2$ en $x_6 = s_2^2$. Voor de vector y van n waarnemingen, gegeven door $y = [y_1, y_2, \dots, y_n]^T$ wordt een lineair regressiemodel $y = \mu + \sigma \epsilon$ gebruikt, waarin μ de verwachtingsvector is en ϵ een vector van onderling onafhankelijke storingstermen met verwachting nul en variantie gelijk aan 1. De verwachtingsvector wordt gegeven door $\mu = X\beta$, waarbij $\beta = [\beta_1, \dots, \beta_k]$, een vector van k parameters ($k < n$) is en $X = [x_1, \dots, x_k]$ een $n \times k$ matrix van de waarden van k verklarende variabelen of regressoren, waaronder het niveau, de vector met alle elementen gelijk aan 1. De regressieparameters worden geschat op basis van de normaalvergelijkingen door:

$$\underline{b} = (X'X)^{-1} X'y, \quad (1)$$

waarin $\underline{b}$ de nauwkeurigste lineaire zuivere schatter van β is. Dit is de bekende OLS-schatter (Ordinary Least Squares).

De schatter van de verwachting van y voor gegeven waarden voor de verklarende variabelen is:

$$\hat{\mu}_0 = x_0'(X'X)^{-1} X'y, \quad (2)$$

waarin x_0 de vector is van verklarende variabelen behorende bij de te schatten

waarde. De onnauwkeurigheid van het kwadraat van de schatter van de verwachting wordt gegeven door $E(\underline{\hat{y}}_0 - \underline{\hat{y}}_0)^2$ de Mean Squared Error (MSE). Bij een zuivere schatter is MSE gelijk aan de variantie:

$$\text{Var}(\underline{\hat{y}}_0) = \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0\sigma^2. \quad (3)$$

In tegenstelling tot $\underline{\hat{y}}_0$ zijn we soms geïnteresseerd in een schatter van een nieuwe waarde voor $\mathbf{y}_0$ zelf, dat wil zeggen in een individuele waarde. Deze wordt gegeven door $\underline{\hat{y}}_0 = E(\mathbf{y}_0) + \sigma\epsilon_0$, waarbij essentieel de residuafwijkingen betrokken worden. Als de verwachting van de residuafwijking gelijk is aan nul, zoals in dit geval (omdat de storingstermen onderling ongecorrleerd zijn en voor elke storingsterm geldt dat de verwachting nul is) is de verwachting van deze schatter gelijk aan de verwachting van $\underline{\hat{y}}_0$. Om ook onderscheid in terminologie te maken heet $\underline{\hat{y}}_0$ ook wel de voorspeller van $\mathbf{y}_0$. De MSE van $\underline{\hat{y}}_0$ wordt gegeven door $E(\mathbf{y}_0 - \underline{\hat{y}}_0)^2$ en is gelijk aan $\text{Var}(\mathbf{y}_0 - \underline{\hat{y}}_0)$ omdat $E(\mathbf{y}_0 - \underline{\hat{y}}_0) = 0$. Dit is te herleiden tot

$$\begin{aligned} \text{Var}(\mathbf{y} - \underline{\hat{y}}_0) &= \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0\sigma^2 + \sigma^2 \\ &= \sigma^2[1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0]. \end{aligned} \quad (4)$$

De variantie van de voorspelfout behorende bij de beste zuivere voorspeller van $\mathbf{y}_0$ is groter dan de variantie van de schattingsfout van de verwachting: het verschil van de twee is precies gelijk aan de variantie van de storingstermen.

Verandering van de vooronderstellingen

In veel ruimtelijke studies is de vooronderstelling dat $\text{cov}(\mathbf{y}_i, \mathbf{y}_j) = 0$ ($i \neq j$) niet geldig. De covariantiematrix heeft nu op de niet-diagonaalposities op één of meerdere plaatsen waarden ongelijk aan nul. Het kan ook zijn dat de $\mathbf{y}$'s ongelijke variantie hebben zodat de covariantiematrix op de diagonaalposities getallen van verschillende grootte heeft. In dit geval geldt:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \underline{\epsilon}^* \quad (5)$$

met $\text{Var}(\underline{\epsilon}^*) = \mathbf{C}$. De gewogen kleinste kwadraten (GLS) schatter van $\boldsymbol{\beta}$ kan worden verkregen door de waarnemingsvector $\mathbf{y}$ te transformeren volgens $\underline{\mathbf{z}} = \mathbf{H}\mathbf{y}$, waarbij

H de matrix uit de Choleski-decompositie van C is, d.w.z. $C=H^{-1}(H')^{-1}$ en dus dat $C^{-1}=H'H$, waarin H de Choleski-transformatiematrix is. De getransformeerde waarnemingen $\underline{z}$ zijn ongecorrleerd en hebben variantie gelijk aan 1. Het regressiemodel wordt nu:

$$H\underline{y} = H\underline{X}\beta + H\underline{e}; \quad \text{Var} (H\underline{y}) = I \quad (6)$$

zodat

$$\underline{b} = ((H\underline{X})'(H\underline{X}))^{-1}((H\underline{X})'H\underline{y}); \quad \text{Var } \underline{b} = ((H\underline{X})'(H\underline{X}))^{-1}. \quad (7)$$

En na herschrijven:

$$\underline{b} = (X'C^{-1}X)^{-1}X'C^{-1}\underline{y}; \quad \text{Var } \underline{b} = (X'C^{-1}X)^{-1}. \quad (8)$$

Op dezelfde manier als bij de OLS-situatie wordt de verwachting in een bepaald punt met waarnemingen aan de onafhankelijke variabelen in x_0 , geschat door:

$$\hat{\mu}_0 = x_0'\underline{b} \quad (9)$$

met als variantie:

$$\begin{aligned} \text{Var}(\hat{\mu}_0) &= x_0'((H\underline{X})'(H\underline{X}))^{-1}x_0 \\ &= x_0'(X'C^{-1}X)^{-1}x_0. \end{aligned} \quad (10)$$

Dit is de zuivere schatter voor de verwachting van y_0 met minimale variantie. Merk op dat de OLS-schatter wordt verkregen door $C = \sigma^2 I$ te substitueren. De OLS-schatter toegepast op deze situatie is ook zuiver maar heeft geen minimale variantie.

Kriging

Bij GLS-regressie werd de schatter voor de verwachting μ_0 van de waarde van y_0 geformuleerd in een punt met regressoren bevat in x_0 . De zuivere voorspeller $\hat{y}_0$ met minimale variantie voor de waarde van y_0 zelf, waarbij de storingstermen betrokken worden, en die lineair is in de vector van variabelen $\underline{y}$, wordt gegeven door $\lambda'\underline{y}$ met $E(\lambda'\underline{y}-y_0)=0$ en $\text{Var}(\lambda'\underline{y}-y_0)$ minimaal. Er geldt:

$$\hat{y}_0 = x_0' \hat{b} + c_0' C^{-1} (y - X \hat{b}), \quad (11)$$

waarin c_0 de covarianties tussen het te voorspellen punt en de meetpunten bevat. Deze voorspeller heet ook wel de krigingvoorspeller. De waarnemingen zijn ruimtelijke punten en de covariantiematrix beschrijft de samenhang tussen de punten. Met behulp van de krigingvoorspeller kan op ieder willekeurig punt een voorspelling uitgevoerd worden voor een milieuvariabele. Door dit op meerdere plaatsen, bijvoorbeeld op een fijnmazig grid, uit te voeren en de voorspellingen met elkaar te verbinden, ontstaat een kaart. Deze activiteit heet interpolatie. De krigingvoorspeller wordt in bodemkunde en milieu uitgebreid toegepast. Vooral in intensief bodemkundig onderzoek met waarnemingspunten op korte afstand van elkaar kan er effectief gebruik van gemaakt worden (Stein en Corsten, 1991). Andere toepassingsgebieden zijn onder andere hydrologie, geologie, meteorologie, mijnbouw en oceanografie. Vaak wordt de term *krigingschatter* (estimator) in plaats van *krigingvoorspeller* (predictor) gehanteerd. Om het onderscheid tussen het schatten van de verwachting en van een individuele waarneming te accentueren, zal in deze tekst de term *voorspeller* gehanteerd worden.

Bij de voorspelling in een willekeurig punt worden ook de residuafwijkingen meegenomen. In deze algemene situatie verschilt de krigingvoorspeller van de schatter van de verwachting, een verschil dat bij OLS-regressie niet werd gevonden. De oorzaak is de correlatie tussen de residuafwijkingen: punten die ruimtelijk gezien dicht bij elkaar liggen hebben vaak een positieve correlatie in de meetwaarden van de variabele. Vooral de residuafwijkingen van nabijgelegen punten en geïsoleerde punten krijgen een groot gewicht omdat door middel van de term $c_0' C^{-1}$ de ruimtelijke correlatiestructuur meegenomen wordt bij de voorspelling van y_0 . Het tweede gedeelte in de krigingvoorspeller wordt dan ook wel de residuele bijdrage genoemd.

De MSE van $\hat{y}_0$ is opnieuw gelijk aan $\text{Var}(y_0 - \hat{y}_0)$ omdat $E(y_0 - \hat{y}_0) = 0$. Dit is in dit geval te herleiden tot

$$\text{Var}(y_0 - \hat{y}_0) = x_a' (X' C^{-1} X) x_a + (c_{00} - c_0' C^{-1} c_0), \quad (12)$$

waarin $x_a = x_0 - c_0' C^{-1} X$ en c_{00} de covariantie van de variabele op het te voorspellen punt is.

De voorspeller $\hat{y}_0$ is een gewogen lineaire voorspeller, waarbij de gewichten λ' van de verschillende waarnemingen y worden gegeven door:

$$\lambda' = x_0' \Lambda_b' + c_0' \Lambda_f' \quad (13)$$

Hierin is elk van de k rijen van $\Lambda_b' = (X'C^{-1}X)^{-1}X'C^{-1}$ gelijk aan de gewichten van y van een GLS regressie coëfficiënt, terwijl $\Lambda_f = C^{-1}C^{-1}X(X'C^{-1}X)^{-1}X'C^{-1}$ een symmetrische $n \times n$ matrix is, die een correctie van de afhankelijkheid der verstoringen beheerst. Deze matrix heeft geen invloed als de waarnemingen onafhankelijk zijn, dus als c_0 een nulvector is. De gewichten in λ zijn zo bepaald dat de variantie van de voorspelfout kleiner is dan voor iedere andere lineaire combinatie (Cressie, 1991). In de praktijk van kriging zijn alleen de gewichten interessant van de gemeten punten die zo dicht bij het te voorspellen punt liggen dat een niet-verwaarloosbare positieve bijdrage aan $\Lambda_f c_0$ verwacht kan worden. Dit betekent dat bij kriging een beperkt aantal punten in de omgeving van het te voorspellen punt (vaak 12-16 punten van in totaal bijvoorbeeld een bestand van tenminste 100 waarnemingen) meegenomen wordt bij het uitvoeren van de voorspelling (Burgess en Webster, 1980). De naam 'gewicht' doet ten onrechte vermoeden dat de coëfficiëntenvector λ alleen positieve waarden kent; echter ook negatieve waarden komen voor.

Als er geen trend is en de covariantie structuur bekend is, dan wordt de krigingvoorspelling simple kriging genoemd. Als bij afwezigheid van trend met een geschatte covariantiestructuur wordt gerekend is er sprake van ordinary kriging. Als er een trend is in een gebied en de verwachte waarde voor een variabele een functie is van één of meer verklarende variabelen (zoals bijvoorbeeld de ruimtelijke coördinaten) komt dit tot uitdrukking in de X -matrix. De methode die dan gebruikt wordt wordt wel universal kriging genoemd (Deutsch en Journel, 1992).

3. Het covariogram en het semivariogram

De C -matrix van covarianties tussen de waarnemingen kan worden opgesteld op grond van een functie die de covariantie beschrijft als functie van de afstand tussen twee punten, het covariogram. Deze functie moet worden vastgesteld aan de hand van de waarnemingen. Aan de volgende voorwaarden moet voldaan zijn,

voordat een schatting van de parameters die deze functie karakteriseren kan worden uitgevoerd:

1. Elke $\underline{y}$ heeft verwachting $\mu(x)$ die constant is of een veelterm van graad v is in de coördinaten van x .
2. Bij elk paar punten x_i en x_j bestaat de bijbehorende covariantie $\text{cov}(\underline{y}_i, \underline{y}_j)$; in het bijzonder bestaat dan ook $\text{var}(\underline{y})$ voor alle x .
3. De covariantie tussen twee variabelen, $\text{cov}(\underline{y}_i, \underline{y}_j)$, hangt slechts af van de verschilvector $h = x_i - x_j$ en *niet* van de locaties zelf. Dit maakt het mogelijk iets over de covariantiestructuur te weten te komen uit één realisatie van een stochastische functie, die op een eindig aantal punten wordt waargenomen.

Het uitvoeren van de schatting wordt tevens vergemakkelijkt als voldaan is aan de voorwaarde:

4. De verschilvector heeft slechts invloed op $\text{cov}(\underline{y}_i, \underline{y}_j)$ via zijn lengte $|h| = r$; dit wordt de eigenschap van isotropie genoemd. Het covariogram, dat dus de covariantie tussen $\underline{y}_i$ en $\underline{y}_j$ uitdrukt in $r = |h|$ wordt hier aangegeven met $c(r)$.

Als er geen isotropie is kunnen de coördinaten soms zodanig lineair getransformeerd worden dat isotropie ontstaat. Met name kan het ellipsvormige patroon van ruimtelijke samenhang behorende bij anisotropie getransformeerd worden tot een cirkelvormig patroon door in de hoofdrichting van ruimtelijke samenhang de middellijn te verkleinen. Een ellipsvormig patroon van ruimtelijke samenhang is daarbij een bijzonder geval van anisotropie. Er zijn ook andere, niet ellipsvormige, gevallen bekend, die lastiger te beschrijven zijn.

Vaak worden slechts de parameters van een gekozen model voor $c(r)$ geschat. Voor zo'n model moet de voorwaarde gelden:

5. De variantie van elke lineaire combinatie $\lambda' \underline{y}$ van $\underline{y}_1, \underline{y}_2, \dots, \underline{y}_n$ is groter dan nul, tenzij λ gelijk is aan de nulvector ($\lambda' C \lambda = 0$). Enkele covariogrammen die bruikbaar zijn voor een drie-dimensionale situatie, en daarmee ook voor één- en een twee-dimensionale situatie, zijn:

$c_e(r) = A \cdot \exp(-r/a)$ voor $r \geq 0$, het zg. exponentiële model, waarin A de variantie, r de afstand en a een parameter die een maat is voor de uitgestrektheid van de ruimtelijke samenhang.

$c_G(r) = A \cdot \exp(-r^2/a^2)$, het Gaussische model.

$c_s(r) = A/2 \cdot (2 - 3(r/a) + (r/a)^3)$ voor $r < a$ en $c(r) = 0$ voor $r > a$, het sferische model.

$c_n(r) = A \cdot \delta(r)$, met $\delta(r) = 1$ als $r = 0$ en $\delta(r) = 0$ elders, het nugget model.

Deze functies die de covariantie tussen twee punten op een afstand r in een bepaald vlak of een bepaalde ruimte definiëren zijn alle monotoon dalend met maximum A in $r=0$ (A is de variantie) en met minimum 0 voor $r \rightarrow \infty$ (voor het sferische model voor $r \geq a$). De parameter a bepaalt hoe snel de functie $c(r)$ daalt. Ook positieve lineaire combinaties van bovengenoemde modellen leveren steeds een toegelaten model op.

De ruimtelijke samenhang kan ook op een andere, veel toegepaste, manier gemodelleerd worden, namelijk via het semivariogram, gedefinieerd als

$$\gamma(r_{ij}) = \frac{1}{2} \text{Var}(y_i - y_j), \quad (14)$$

waarin $\gamma(r_{ij})$ het semivariogram is, $(y_i - y_j)$ het verschil tussen de i^e en de j^e variabele en r_{ij} de afstand tussen de plaatsen waar y_i en y_j bepaald zijn of kunnen worden. Als het covariogram $c(r)$ gedefinieerd is, dan geldt dat $\gamma(r) = c(0) - c(r)$, waarbij $c(0)$ en $c(r)$ het covariogram voorstellen voor afstanden 0 en r , respectievelijk. De functies $\gamma(r)$ en $c(r)$ zijn elkaars spiegelbeeld ten opzichte van de horizontale lijn ter hoogte $\frac{1}{2}c(0)$. De vier eerder genoemde covariogrammen kunnen zo eenvoudig getransformeerd worden tot semivariogrammen. Daarnaast kunnen ook onbegrensde functies als semivariogrammen gebruikt worden. Het bekendste voorbeeld is het zg. power model $\gamma_p(r) = Ar^\ell$, waarbij $A > 0$ en $0 \leq \ell < 2$ (voor $r > 0$), die geen equivalent als covariogram kent.

Per definitie is een semivariogram voor $r = 0$ gelijk aan 0. In de praktijk blijkt echter dat een semivariogram vaak een discontinuïteit vertoont bij $r = 0$. Dit heet het **nugget effect**, dat gelijk is aan de sprong van het semivariogram bij de oorsprong. Het is een maat voor niet-ruimtelijke fluctuaties, veroorzaakt door o.a. meetfouten. Het zorgt ook voor discontinuïteiten bij interpolatie bij de waarnemingen en draagt

in belangrijke mate bij aan de variantie van de krigingschatter (Burgess en Webster, 1980a). De **range** is de afstand r waarop de functie $\gamma(r)$ een bepaalde plafondwaarde bereikt heeft: het is de maximale afstand waarvoor ruimtelijke correlatie tussen twee punten optreedt. De **sill** is de hoogte van dit plafond.

Als aan de voorwaarde voldaan is dat $E(y_i - \hat{y}_i) = 0$, dan kunnen zuivere schatters van $\gamma(r)$ verkregen worden door b.v. kwadraten van paarverschillen op afstand r te sommeren en te delen door twee maal hun aantal. In de uitdrukking $\frac{1}{2}E(y_i - \hat{y}_i)^2 = \gamma(r_{ij})$ wordt de verwachting vervangen door het gemiddelde. Ook andere, robuuste, schattingsmethoden voor het semivariogram zijn bekend (Cressie, 1991). De functie $\gamma(r)$ kan in een grafiek uitgezet worden tegen r voor verschillende afstanden (ingedeeld in klassen). Interpretatie van dergelijke grafieken kan vaak bruikbare informatie opleveren.

Als er sprake is van een trend is het schatten van de ruimtelijke samenhang ingewikkelder. Dit kan volgens de methode der incrementen (Corsten, 1989). Deze incrementenmethode geeft uitgebreidere toepassingsmogelijkheden voor kriging in het geval van niet-stationariteit. Als er sprake is van een polynomiale verwachtingsfunctie kan door toepassing van incrementen deze trend eruit gefilterd worden zodat alleen covarianties van incrementen een rol spelen. Het gevolg is dat algemenere covariantiefuncties mogelijk zijn.

4. Een parametervrije vorm van kriging: indicator kriging

In paragraaf 2 is kriging gepresenteerd als een methode om op basis van een beperkt aantal metingen aan variabele y op andere plaatsen in het gebied ook de waarde van variabele y te voorspellen. De kriging voorspeller is zuiver, heeft minimale variantie en is daarom onder de lineaire voorspellers de beste. Een typisch verschijnsel bij kriging is dat 'smoothing' optreedt: het voorspelde beeld is gladder dan de werkelijkheid. Als de waarnemingen scheef verdeeld zijn en er grote uitschieters zijn in de meetgegevens kan de lineaire schatter het daarom niet meer zo goed doen. In het geval van verontreinigingen kunnen uitschieters frequent voorkomen. De variatiecoëfficiënt in de gemeten variabelen is als regel hoog en de zeer grote heterogeniteit maakt semivariogrammen gevoelig voor uitschieters. In zo'n

geval loont het wellicht de moeite een niet-lineaire voorspeller te gebruiken, de zg. indicator voorspeller (Journel, 1983). Indicator-kriging wordt o.a. toegepast om de kans te bepalen dat een bepaalde kritieke waarde overschreden wordt.

Stel er is de beschikking over n metingen aan variabele y op n verschillende plaatsen $\{y(x_i), i=1, \dots, n\}$ in een gebied D . In de geostatistiek wordt dit geïnterpreteerd als een realisatie van de stochastische functie $\{y(x), x \in D\}$. Er wordt verondersteld dat de stochastische functie $y(x)$ stationair binnen D is, met name dat er geen trend is in het gebied. Bovendien wordt verondersteld dat alle waarnemingen uit dezelfde verdeling $F(y) = \text{Prob}(y \leq y)$ afkomstig zijn. Er wordt nu een aantal indicatorvariabelen ingevoerd. Een indicatorvariabele op het punt $x \in D$ voor een grenswaarde y (Eng.: threshold value, critical value) is gedefinieerd als:

$$\underline{i}(x; y) = \begin{cases} 0 & \text{als } y(x) > y \\ 1 & \text{als } y(x) \leq y \end{cases} \quad (15)$$

De verwachting van $\underline{i}(x; y)$ wordt gegeven door $E(\underline{i}(x; y)) = F(y)$. Deze verwachting is de waarde van de verdelingsfunctie die hoort bij de grenswaarde y en is dus onafhankelijk van de plaats x . Evenzo wordt de variantie gegeven door $\text{Var}(\underline{i}(x; y)) = F(y) \cdot (1 - F(y))$. Voor een indicator-variabele kan een indicator-semivariogram gedefinieerd worden. Dat is gelijk aan $\gamma_i(r; y) = \frac{1}{2} E[\underline{i}(x+r; y) - \underline{i}(x; y)]^2 = \frac{1}{2} \text{Var}(\underline{i}(x+r; y)) + \frac{1}{2} \text{Var}(\underline{i}(x; y)) - \text{Cov}(\underline{i}(x+r; y), \underline{i}(x; y)) = F(y)(1 - F(y)) - \text{Cov}(\underline{i}(x+r; y), \underline{i}(x; y))$. Dit kan ook geschreven worden als:

$$\frac{\gamma_i(r; y)}{F(y)(1 - F(y))} = 1 - \rho_i(\underline{i}(x+r; y), \underline{i}(x; y)) \quad (16)$$

waarin $\rho_i(\underline{i}(x+r; y), \underline{i}(x; y))$ het indicator-correlogram is voor grenswaarde y . Dit is gelijk aan de correlatie coëfficiënt tussen indicator-variabelen op een afstand r van elkaar. Het indicator-semivariogram voor afstand r and grenswaarde y wordt dus ook wel gegeven door:

$$\gamma_i(r; y) = F(y)(1 - F(y))(1 - \rho_i(r; y)). \quad (17)$$

Voor het bepalen van de variantie, covariantie en dus ook het semivariogram is een aantal parameters nodig. Dezelfde procedures als beschreven in paragraaf 3 kunnen

hiervoor gebruikt worden.

Een schatting voor $F(y)$ kan worden verkregen op grond van de experimentele verdelingsfunctie van de waarnemingen. Hiertoe wordt deze ingedeeld in een aantal, zeg K , klassen, bijvoorbeeld in kwantielen. Vervolgens wordt de experimentele verdelingsfunctiewaarde voor alle grenswaarden die als klassegrenzen gaan fungeren bepaald. Voor elke grenswaarde y is een schatting voor de verdelingsfunctie:

$$\hat{F}(y) = \frac{1}{K} \sum_{k=1}^K \sum_{i(k)=1}^{n_k} \frac{1}{n_k} i(x_{i(k)}; y) \quad (18)$$

waarin n_k aantal waarnemingen binnen klasse k is en $i(x_{i(k)}; y)$ de indicatorwaarde van waarneming y_k op plaats $x_{i(k)}$. Door middel van eenvoudige (lineaire) interpolatie en, bijvoorbeeld, kwadratische extrapolatie naar opgegeven minimum en maximum waarde, kan vervolgens $F(y)$ bepaald worden voor iedere willekeurige waarde y .

Interpolatie is betrekkelijk stabiel met betrekking tot uitschieters omdat de schatting slechts de plaatsing in een klasse gebruikt in plaats van de absolute waarde. De best bruikbare experimentele indicator-semivariogrammen corresponderen met grenswaarden dichtbij de mediaanwaarde, waarbij de helft van de indicator-waarden gelijk is aan 0 en de andere helft gelijk is aan 1. Switzer (1977) heeft laten zien dat het indicator-correlogram vrij ongevoelig is voor grenswaarden y die enigszins afwijken van de mediaanwaarde. Preciezer gezegd: het indicator-correlogram is stabiel voor grenswaarden y die in het interval $[0.33, 0.67]$ van de verdelingsfunctie liggen.

5. Simpele indicatorkriging voor schattingen van de verdelingsfuncties op punten

In een punt x_0 waar geen waarneming werd uitgevoerd kan de onbekende indicatorwaarde $\hat{i}(x_0; y)$ geschat worden door simpele indicatorkriging. De schatting van $F(y)$ is bekend op basis van vgl. (18) en de overige ontbrekende parameters zijn verkregen uit de experimentele bepaling van het indicator-semivariogram. De indicator-voorspelling wordt nu verkregen analoog aan de voorspelling met behulp van ordinary kriging:

$$\hat{i}(x_0; y) = \hat{F}(y) + c_0' C^{-1} (i(x_0; y) - \hat{F}(y)) \quad (19)$$

waarin $\hat{i}(x_0; y)$ de verwachte indicatorwaarde in het punt x_0 en $i(x_0; y)$ de tot indicatorwaarden getransformeerde vector van waarnemingen is.

Als de elementen van $c_0' C^{-1}$ worden vervangen door gewichten α , dan ontstaat de volgende uitdrukking:

$$\hat{i}(x_0; y) = \hat{F}(y) + \sum_{j=1}^n \alpha_j(y) (i(x_j; y) - \hat{F}(y)) \quad (20)$$

waarin $\alpha_j(y)$ het gewicht is behorende bij waarneming j en grenswaarde y . Dit kan geschreven worden als:

$$\hat{i}(x_0; y) = \sum_{i=1}^n \lambda_i(y) i(x_i; y) + (1 - \sum_{i=1}^n \lambda_i(y)) \hat{F}(y). \quad (21)$$

Het corresponderende krigingsstelsel is dan:

$$\sum_{j=1}^n \lambda_j(y) \rho_f(x_i - x_j; y) = \rho_f(x_i - x_0; y) \quad i=1, 2, \dots, n \quad (22)$$

waarin $\rho_f(x_i - x_j; y)$ het correlogram is tussen de stochastische variabelen op locaties x_i en x_j bij grenswaarde y en $\lambda_j(y)$ de gewichten zijn die behoren bij de locatie j . De geschatte variantie is:

$$\sigma^2(y) = E(\hat{i}(x; y) - \hat{i}(x; y))^2 = s^2(y) (1 - \sum_{i=1}^n \lambda_i(y) \rho_f(x_i - x; y)) \quad (23)$$

waarbij $s^2(y)$ de variantie van $\hat{i}(x; y)$ is, die gelijk is aan de sillwaarde behorende bij grenswaarde y .

De schatting van $\hat{i}(x_0; y)$ geeft de geschatte waarde van de verdelingsfunctie op punt x_0 voor de grenswaarde y . De kans dat op dat punt de waarde voor de variabele $y(x)$ boven deze grenswaarde ligt wordt dan geschat door $1 - \hat{i}(x; z)$. Om de kans te bepalen dat de waarde tussen twee grenswaarden a en b ($a < b$) ligt, kan de procedure twee keer worden toegepast:

$$\begin{aligned}
 \hat{i}(x; y_a) &= a \\
 \hat{i}(x; y_b) &= b \\
 Prob(y(x) \in [y_a, y_b]) &= b - a
 \end{aligned}
 \tag{24}$$

waarin y_a de grenswaarde behorende bij a en y_b grenswaarde behorende bij b is. Door de procedure te herhalen voor een aantal drempelwaarden is het mogelijk de verdelingsfunctie van de indicatorwaarde voor iedere willekeurige waarde y in een punt x_0 te bepalen.

6. Toepassingen

De toepassingen hebben alle betrekking op het gebruik van geostatistiek binnen een geografisch informatie systeem. Het begrip 'milieu' is uiteraard buitengewoon divers. Het heeft te maken met zowel de levende natuur als de levenloze natuur. Wat de levenloze natuur betreft, heeft het betrekking op de kwaliteit van bodem, water en atmosfeer. We kunnen daarbij denken aan milieu op verschillende schalen, van een verontreinigde slootkant tot beleidsmaatregelen op internationaal niveau. Het milieu kan schoon zijn, terwijl verontreiniging leidt tot kostbare maatregelen. Voor de statistiek begint het milieu als regel als er een probleem is. Dan zijn er voldoende waarnemingen verricht of moeten er voldoende waarnemingen verricht worden, om statistisch onderbouwde uitspraken te krijgen.

Achtereenvolgens richt ik me op het milieu in natuurlijke omstandigheden, en zal ik kijken naar interpolatie op verschillende schalen. Het gebruik van Geografische Informatie Systemen zal hierbij geïllustreerd worden. Vervolgens richt ik me op de effecten van scenario studies. Opnieuw is dit een studie waarbij de rol van GIS een cruciale is. We denken dan aan mogelijke belasting van het milieu en stellen ons telkens de vraag wat de consequenties van zo'n belasting zullen zijn. Tenslotte richt ik me op het milieu in verontreinigde omstandigheden en besteed aandacht aan de risico's die daaraan verbonden zijn voor het 'natuurlijke' milieu en voor de volksgezondheid.

Een algemeen kader voor het gebruik van geostatistiek in het milieu begint met een probleem: ergens is gesignaleerd dat er een terrein verontreinigd is of zou

kunnen zijn of dat analyse op het gebied van de milieubescherming moet worden uitgevoerd. De eerste vragen die beantwoord worden betreffen de ruimtelijk bemonstering: hoeveel en waar moet er bemonsterd worden. Veel onderzoek is daar gedurende de afgelopen tijd aan uitgevoerd, dat onder andere in expert systemen haar beslag krijgt (Domburg et al., 1994). Nadat bemonstering is uitgevoerd en waarnemingen zijn verkregen, wordt interpolatie naar gebieden uitgevoerd: een perceel, een stroomgebied, een waterschap, etc. Bij de interpolatie dient informatie op het gebied van de bodemkunde, de hydrologie, en de geologie betrokken te worden. Vaak wordt daarbij van een rekenmodel gebruik gemaakt. De resultaten worden vervolgens gekoppeld aan een risico-analyse. Hierbij wordt gebruik gemaakt van kwantitatieve normen, gebaseerd op gezondheidswetenschappen en toxicologie. Ook speelt de aard van het terrein soms een rol: voor volkstuintjes kunnen andere normen gelden dan voor natuurgebieden of fabrieksterreinen. Op basis van de normen en van met name ook van financiële afwegingen worden dan beslissingen genomen: sanering of isolatie van een terrein, nader onderzoek naar deelgebieden, het geheel laten zoals het is, omdat er geen aantoonbare schade voor de volksgezondheid optreedt, of, in een extreem geval, aanpassing van de norm. Deze lijst is zonder veel problemen verder uit te breiden. Geostatistiek richt zich op het ontwerpen van een bemonsteringsschema, ruimtelijke interpolatie en aspecten van validatie. Ook talloze andere statistische vragen kunnen in milieustudies naar voren komen. Denk aan het schatten van model parameters, het uitvoeren van een verantwoorde risico-analyse, het vaststellen van normen, etc.

Studie 1: grondwater in Noord-Brabant.

In de jaren tachtig heeft de provincie Noord-Brabant een analyse laten uitvoeren naar de natuurlijke gehalten aan zware metalen in het grondwater. Op basis van een enquête is een database gecreëerd met daarin van 25 gemeentes het gehalte aan zware metalen, zoals gemeten in ca. 1000 peilbuizen. De reden was dat vaak onverwacht hoge waarden voorkomen, die op geen enkele reden verklaard konden worden. Bovendien zou een beter inzicht in de natuurlijke achtergrondwaarden kunnen leiden tot een meer genuanceerde normstelling. Het betrof hier geaggregeerde

gegevens: de coördinaten van de peilbuizen zijn op gehele kilometers afgerond. Aan blokken van 1 bij 1 km waar zich meer dan één peilbuis bevond, is een gemiddelde waarde toegekend. Met name ging het om het zinkgehalte, waarvan vrij hoge waarden werden gemeten; bovendien is de zinkbepaling vrij nauwkeurig. De grenswaarden voor zink zijn 150 µg/l (A-waarde), 200 µg/l (B-waarde) en 800 µg/l (C-waarde). Beneden de A-waarde is er sprake van geen verontreiniging, terwijl boven de C-waarde een gevaar voor de volksgezondheid optreedt. Om een verklaring te vinden voor de ruimtelijke variatie is vervolgens ingezoomd op de gemeente Oss, waar ca. 80 peilbuizen waren, en binnen de gemeente Oss is een verzameling percelen geselecteerd, waar zich een hoge concentratie peilbuizen bevond. De relevante vraag is hoe een ruimtelijke interpolatie op de verschillende schalen kan worden uitgevoerd. Op het provinciale niveau ligt het voor de hand dat een relatie tussen zink en landschappelijke gegevens, zoals verschillende geologische eenheden en verschillende vormen van kwel bestaat. Om dit te onderzoeken zijn in het GIS systeem ArcInfo de geologische kaart van Noord-Brabant en de kwelkaart gedigitaliseerd. Voor deze studie zijn de eenheden gegeneraliseerd tot drie eenheden met een duidelijk verschillende signatuur. We konden constateren dat er significante verschillen tussen eenheden bestaan, zowel in de gemiddelde zinkconcentraties als in de spreiding binnen de eenheden. We concludeerden derhalve, dat de eenheden als drager voor de informatie konden fungeren: de gehalten aan zink volgen min of meer de grenzen zoals die door de geologische kaart worden weergegeven. Een soortgelijk verhaal gaat op voor de kwelkaart. We onderscheiden hier 5 eenheden, van diepe kwel tot infiltratiegebieden. Ook hier constateerden we dat er sprake is van significante verschillen tussen de eenheden, waarbij met name de eenheden met kwel van grote rivieren, ondiepe kwel en infiltratie een hogere waarde laten zien dan de eenheden die diepe kwel kennen, dan wel tot het tussenliggend gebied horen. Dit lijkt erop te wijzen dat hoge zinkwaarden vooral veroorzaakt worden door factoren van buitenaf.

Op een lager niveau constateren we geen significantie meer bij een onderscheid in verschillende eenheden. Op dit niveau lijkt het gerechtvaardigd om geostatistiek toe te passen, waarbij immers een zekere mate van stationariteit verondersteld moet worden. Op deze schaal waren de coördinaten gegeven met een nauwkeurigheid van

ca. 10 m. Verschillende waarnemingen zijn boven de C-waarde en vormen dus een bedreiging voor de volksgezondheid. Tevens is er sprake van een grote ruimtelijke variatie: hoge en lage waarnemingen kunnen op korte afstand van elkaar voorkomen. Met deze waarnemingen is kriging uitgevoerd, die een kaart heeft opgeleverd van de zinkgehaltes (fig. 1). Een paar kleine gebieden zijn aan te wijzen waar het gehalte aan zink in het grondwater hoog is. Het vermoeden bestaat dat de lokale industrie hier niet helemaal onschuldig aan is.

Uit de betrouwbaarheidskaart blijkt dat de onzekerheid substantieel is: deze ligt in de grootte-orde van 200 tot 400 $\mu\text{g}/\ell$. Alleen in het gedeelte van het terrein waar meerdere waarnemingen zijn uitgevoerd is sprake van een geringere ruimtelijke onzekerheid. Op het perceelniveau, tenslotte, is het gebruik van geostatistiek niet meer goed mogelijk, gezien het beperkte aantal waarnemingen, en moeten we ons beperken tot een sub-optimale methode. Hier is gebruik gemaakt van laplacian smoothing splines, terwijl inverse afstands interpolatie of Thiessen polygonen eveneens gebruikt hadden kunnen worden (voor beschrijving zie Cressie, 1991).

Uit deze studie concluderen we dat ruimtelijke variatie in natuurlijke achtergrondwaarden erg groot is. Bij sterk verhoogde waarden lijkt een vingerwijzing naar verontreiniging wel erg voor de hand liggend. Ruimtelijke interpolatie is afhankelijk van de schaal en beschikbare voorinformatie. Geostatistiek is bruikbaar op met name het stadsniveau.

Studie 2: effecten van beleidsscenario's.

De tweede studie betreft een analyse van de effecten van beleids-scenario's. We richten ons op de omgeving van Uden en Veghel, waar gekeken is naar de effecten van nitraat uitspoeling. Hiertoe worden modellen gebruikt die ontwikkeld worden bij o.a. het Staring Centrum in Wageningen en aan de Landbouw Universiteit Wageningen. Dit zijn vaak complexe modellen waarin veel gegevens worden gebruikt die alle te maken hebben met de beweging van nitraathoudende stoffen in de bodem, op weg naar het grondwater. Zolang nitraat in de bodem zit is het nog relatief ongevaarlijk, maar zodra het grondwater bereikt wordt nemen de risico's voor de volksgezondheid toe.

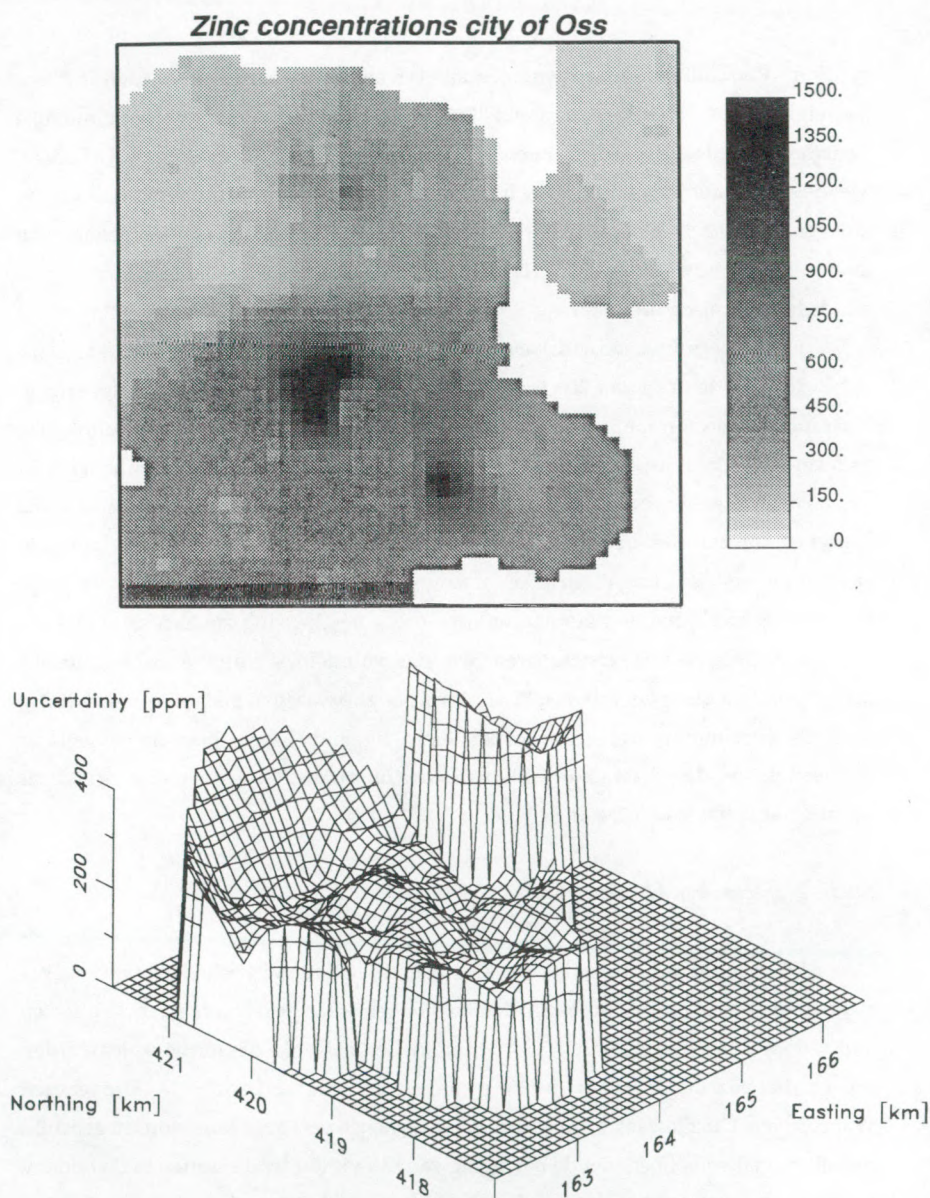


Fig. 1. Zinkconcentratie in het grondwater in de gemeente Oss. Concentraties [in $\mu\text{g l}^{-1}$] staan in de bovenste figuur, de onzekerheden [in $\mu\text{g l}^{-1}$] in de onderste figuur.

In deze studie zijn de effecten van 3 scenario's met elkaar vergeleken die door het ministerie van VROM zijn geformuleerd: i) wat gebeurt er als we onder de huidige omstandigheden doorgaan; ii) wat gebeurt er als de toevoeging van nitraat drastisch gereduceerd wordt en iii) wat gebeurt er als de uitstoot van nitraat gereduceerd wordt met 50%. Alle maatregelen voeren ver, maar zijn in principe goed denkbaar.

In het GIS pakket IDRISI is gebruik gemaakt van een digitale bodemkaart, gegeneraliseerd naar 8 eenheden, een grondwatertrappenkaart, die aangeeft hoe diep het grondwater zich onder de oppervlakte van de bodem bevindt (het zgn. maaiveld) en een gedigitaliseerd 1:25,000 satellietbeeld, geïnterpreteerd naar landgebruik, gegeneraliseerd naar 8 eenheden. Generalisatie was nodig om de bewerkingen mogelijk te maken. Voor iedere combinatie van bodemkaart, hydrologische eenheid en landgebruikseenheid is een nitraatuitspoelingsmodel gehanteerd, dat de uitspoeling naar het grondwater kwantificeert. Bij de huidige belasting (scenario 1) blijkt dat er nogal wat variatie optreedt: overheersend relatief hoge waarden met een paar geïsoleerde eenheden die wat lagere uitspoeling kennen, etc. Als er een forse reductie plaatsvindt in de nitraatbelasting (scenario 2), dan vermindert de uitspoeling. Het terugbrengen van de uitstoot van nitraat (scenario 3) leidt echter niet tot een belangrijke reductie. Al met al kunnen we uit deze studie constateren dat beleidsmaatregelen op hun effecten uitstekend vergeleken kunnen worden. De drager van informatie bij het uitvoeren van ruimtelijke interpolatie is in alle gevallen dus steeds de combinatie bodemkaart - landgebruik - hydrologie geweest. Het is de vraag of voor het nemen van beleidsmaatregelen een dergelijk gedetailleerd beeld noodzakelijk is.

In deze studie is geen geostatistiek gebruikt. De mogelijkheden zijn daar wel voor aanwezig:

- het interpoleren van berekende nitraatuitspoeling kan met behulp van geostatistiek worden uitgevoerd; dit levert bijvoorbeeld kansenkaarten op dat de kritieke waarden bij bepaalde scenario's worden overschreden.
- het interpoleren van invoergegevens die niet op alle punten bekend zijn (of betrokken zijn van de kaarten) kan met geostatistiek uitgevoerd worden.
- conditionele simulaties kunnen als een Monte-Carlo-achtige benadering gebruikt

worden om de gevoeligheid van het model op invoer parameters te onderzoeken (Deutsch en Journal, 1992).

Studie 3: bodemverontreiniging en overschrijdingskansen.

De derde studie betreft een onderzoek naar verontreinigingen met zware metalen, o.a. lood, op een 4 ha groot terrein. Lood is op het terrein met name terechtgekomen via afvallozingen. In deze studie is gekeken naar de overschrijding van kritieke drempelwaarden op verschillende dieptes. De waarnemingen zijn uitgevoerd op de dieptes 0-0.5m, 0.5-1m en 1-1.5m. In totaal zijn er ca. 500 waarnemingen beschikbaar. Voor beleidsmaatregelen was het nodig om vooral aandacht te besteden aan de kansen dat de kritieke waarden worden overschreden. De relevante A-, B- en C-waarden voor lood zijn gelijk aan 85, 150 en 600 mg Pb per kg droge stof. De A-waarde is specifiek voor het gebied, omdat deze mede bepaald wordt door het klei en het organische-stofgehalte in de bodem. Allereerst is een kaart gemaakt met daarop de contouren van overschrijding van de A-, B- en C-waarden (fig. 2). We constateren dat goed bruikbare kaarten worden geproduceerd, ondanks het feit dat stationariteit in deze gegevens betrekkelijk ver te zoeken is. Er zijn verschillende deelgebieden met hoge waarden te ontdekken, die vermoedelijk een aparte behandeling zullen vergen.

Op grond van historische informatie bleek dat deze deelgebieden gekoppeld zijn aan activiteiten die een bijzonder verontreinigend karakter dragen. Aan de andere kant is het gebied zodanig verontreinigd, dat een integrale behandeling noodzakelijk lijkt. Aan de randen van het gebied is er sprake van duidelijk minder verontreiniging. Vervolgens zijn kaarten geconstrueerd met de kansen op overschrijding van de A-, B- en C-waarde binnen de bovenste laag (fig. 3). Duidelijk is dat de globale structuur van deze kaarten overeen komt met het patroon dat we eerder zagen. De verschillen in kansen op overschrijding van de A- en B-waarde zijn gering, omdat de A- en B-waarden voor lood dicht bij elkaar liggen.

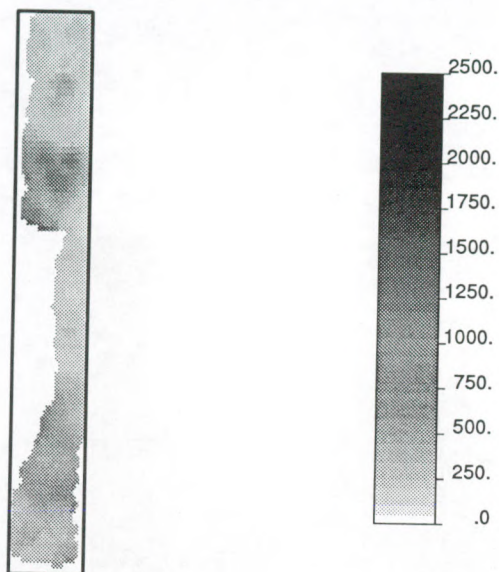


Fig. 2. Loodconcentraties uitgedrukt in mg lood per kg droge grond in een gebied van 250 m bij 2 km.

Op grond van deze kaarten kunnen we constateren dat de overschrijdingskansen uiteraard afhangen van de betreffende drempelwaarde, maar ook van de ruimtelijke variatie (die in de middelste laag het grootste was, omdat deze nogal weinig van karakter was). Daarnaast hangen de kansen af van de configuratie van de waarnemingen: een toename van de onbetrouwbaarheid naar de grenzen toe is goed zichtbaar. Soortgelijke kaarten zijn ook vervaardigd voor de twee andere lagen. Dergelijke kaarten kunnen een geheel nieuwe optie openen op het gebied van het saneringsbeleid: enerzijds kan men besluiten tot sanering van terreinen waar de C-waarde met, b.v. tenminste 5%, 50% of 95% zekerheid overschreden wordt, anderzijds kan men besluiten tot het uitgeven van 'onverontreinigingscertificaten' waarop vermeld staat dat bepaalde bouwgrond met 95% zekerheid geen overschrijding van de A-waarde laat zien, een lijn die in sommige gemeentes al gevolgd wordt.

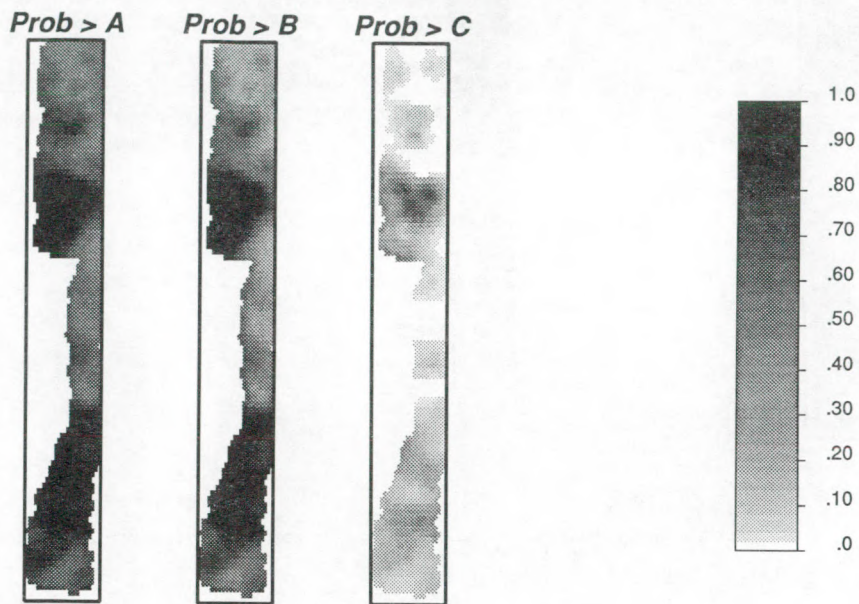


Fig. 3. De kansen dat drie kritieke grenswaarden, de A-, B- en C-waarden, gelijk aan 150, 200 en 800 ppm, respectievelijk, worden overschreden.

We concluderen uit deze studie dat geostatistiek inmiddels voorbij de fase van ruimtelijke interpolatie is gekomen. Nadrukkelijk blijkt het mogelijk te zijn vragen op het gebied van beheer en beleid op een goed onderbouwde, relevante wijze te beantwoorden.

7. Conclusies

Op basis van de uitgevoerde studies blijkt dat op statistisch gebied nog veel

te doen valt op het gebied van het milieu:

- het verder ontwikkelen van modellen voor risico-evaluatie; hierbij denk ik met name aan het vaststellen van de toxiciteit van risico-volle stoffen;
- het toe- en aanpassen van statistische bemonsterings procedures: hierbij is niet alleen het *aantal* waarnemingen van belang, maar ook de *plaats* waar deze genomen worden.
- het verder ontwikkelen van (o.a. fuzzy) classificatiemethoden, teneinde homogene deelgebieden te creëren die een uniforme behandeling mogelijk maken.

De bijdragen op het gebied van GIS en geostatistiek liggen dan met name bij

- het in kaart brengen van gehalten aan zware metalen, maar ook aan nitraatconcentraties, PAK's, PCB's, etc., inclusief de ruimtelijke onzekerheid.
- het uitvoeren en visualiseren van de effecten van risico-evaluaties; combinatie van geostatistiek met beschikbare voorinformatie en met modellen binnen een GIS is daarbij de primaire activiteit.

Literatuur

- Bregt, A.K., 1993, Integrating GIS and process models for global environmental assessment. In *Proceedings of the international workshop on global GIS, August 24-25, 1993*, edited by R. Tateishi (Tokyo) pp. 77-84.
- Burgess, T.M. en R. Webster. 1980. Optimal interpolation and isarithmic mapping of soil properties. I The semi-variogram and punctual kriging. *Journal of Soil Science*, 31: 315 - 331.
- Burrough P.A., 1986, *Principals of Geographical Information Systems for Land Resources Assessment* (Oxford: Clarendon Press).
- Corsten, L.C.A. 1989. Interpolation and optimal linear prediction. *Statistica Neerlandica* 43, nr 2: 29 - 44.
- Corsten, L.C.A., and Stein, A. 1994. Nested sampling for estimating spatial semivariogram compared to other designs. *Applied Stochastic Models and Data Analysis*, 10, 103-122.
- Cressie, N.A.C. 1991. *Statistics for spatial data*. Wiley, New York.
- Deutsch, C.V. en A.G. Journel. 1992. *GSLIB. Geostatistical Software Library and User's*

- Guide*. Oxford, Oxford University Press.
- Domburg, P., J.J. De Gruijter en D.J. Brus. 1994. A structured approach to designing soil survey schemes with prediction of sampling error from variograms. *Geoderma* 62, 151-164.
- Finke, P.A., and A. Stein. 1994. Application of disjunctive cokriging to compare fertilizer scenarios on a field scale. *Geoderma* 62, 247-263.
- Heuvelink, G.B.M., en P.A. Burrough. 1993. Error propagation in logical cartographic modelling using Boolean methods and continuous classification. *International Journal of Geographical Information systems*, 7, 231.
- Journal, A.G. 1983. Nonparametric estimation of spatial distributions. *Mathematical Geology*, vol. 15, no. 3: 445 - 468.
- McBratney, A.B. en R. Webster. 1986. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. *Journal of Soil Science*, 37: 617 - 639.
- Staritsky, I.G., P.H.M. Sloot en A. Stein (1992). Spatial variability and sampling of cyanide polluted soil on former galvanic factory premises. *Water, Air, and Soil Pollution*, 61, 1-16.
- Stein, A. and L.C.A. Corsten. 1991. Universal Kriging and cokriging as a Regression Procedure. *Biometrics*, 47, 575-587.
- Stein, A. 1993. Modelling spatial variability of soil pollution variables with geostatistics for environmental purposes. In: H.J.P. Eijsackers and T. Hamers (Eds.). *Integrated soil and sediment research: a basis for proper protection*. Kluwer, Dordrecht, 431-434.
- Stein, A. 1994. The use of prior information in spatial statistics. *Geoderma*, 62, 199-216.
- Switzer, P. 1977. Estimation of spatial distributions from point sources with application to air pollution measurement. *Bulletin of the International Statistical Institute*, 47, book 2, 123-137.