Beslissen over vrij grondverzet op basis van Bayesiaanse toets op oppervlaktefractie ernstig verontreinigde grond

Beslissen over vrij grondverzet op basis van Bayesiaanse toets op oppervlaktefractie ernstig verontreinigde grond

D.J. Brus

Alterra-rapport 867

Alterra, Wageningen, 2004

REFERAAT

Brus, D.J., 2004. *Beslissen over vrij grondverzet op basis van Bayesiaanse toets op oppervlaktefractie ernstig verontreinigde grond.* Wageningen, Alterra, Alterra-rapport 867. 40 blz.; 9 fig.; 3 tab.; 9 ref.

De vrijstellingsregeling grondverzet biedt de mogelijkheid tot vrijstelling van partijkeuringen wanneer aangetoond wordt dat de oppervlaktefractie ernstig verontreinigde grond kleiner is dan 5%. In dit onderzoek wordt de beslissing gebaseerd op een statistische toets omdat dan rekening wordt gehouden met het aantal monsters dat in sterke mate de mate van onzekerheid over de oppervlaktefractie bepaald. Bayesiaans toetsen biedt interessante mogelijkheden in vergelijking met klassiek toetsen omdat, gegeven het aantal monsters en het aantal ernstig verontreinigde monsters in de steekproef, de betrouwbaarheid van de conclusie 'oppervlaktefractie < 0.05' door het gebruik van de *a priori* verdeling over de oppervlaktefractie toeneemt. Ook het verlies van foute beslissingen (schade tengevolge van als bodem hergebruikte ernstig verontreinigde partij grond, en kosten van keuren van niet ernstig verontreinigde partijen) kan hierdoor gereduceerd worden.

Trefwoorden: Bayes, binomiale verdeling, bodemkwaliteit, grondverzet, onzekerheid, partijkeuring, toetsen.

ISSN 1566-7197

Dit rapport kunt u bestellen door € 13,- over te maken op banknummer 36 70 54 612 ten name van Alterra, Wageningen, onder vermelding van Alterra-rapport 867. Dit bedrag is inclusief BTW en verzendkosten.

Postbus 47; 6700 AA Wageningen; Nederland
Tel.: (0317) 474700; fax: (0317) 419000; e-mail: info@alterra.wur.nl

[Alterra-rapport 867/01/2004]

Inhoud

Woord vooraf

Dit onderzoek is uitgevoerd in het kader van het DWK-onderzoeksprogramma 'Basis- en kerngegevens bovengrond'. In dit programma wordt onderzoek gedaan naar hoe rekening kan worden gehouden met onzekerheid bij het nemen van beslissingen op basis van kaarten van de bovengrond, en wat de gevolgen hiervan zijn voor de praktijk. In 2002 is een studie gedaan waarin kaarten van het grondwaterstandsverloop centraal stonden. Voor 2003 is, in overleg met de begeleidingscommissie, gekozen voor een bodemkundig onderwerp. Gezien de actualiteit van het onderwerp bodemkwaliteit, en het hoge tempo waarin op dit moment in opdracht van gemeenten bodemkwaliteitskaarten worden vervaardigd, is gekozen voor een studie naar hoe onzekerheid over de milieuhygiënische bodemkwaliteit gebruikt kan worden bij het nemen van beslissingen over wel of geen verplichte keuring van partijen grond.

Graag wil ik de begeleidingscommissie van DWK-programma 395, in het bijzonder Jan Huinink, en de toenmalige programmaleider, Peter Finke, bedanken voor het in mij gestelde vertrouwen. Ik ben me bewust van de grote kloof tussen de huidige praktijk van beslissen over vrij grondverzet en de in dit rapport voorgestelde methode. Maar ik ben er ook van overtuigd dat in de toekomst, o.a. tengevolge van Europese regelgeving en het steeds mondiger worden van burgers, de Nederlandse overheid genoodzaakt zal zijn om, bij het ontwerpen van normen en richtlijnen ten behoeve van milieuonderzoek, meer gebruik te maken van wetenschappelijke methoden. Ik hoop dat dit rapport een bijdrage levert aan deze ontwikkeling.

Samenvatting

Wanneer van een gebied een bodemkwaliteitskaart is gemaakt volgens de hiervoor geldende Interim richtlijn bodemkwaliteitskaarten, kan vrijstelling van partijkeuring worden verkregen wanneer aangetoond wordt dat de kans dat de partij grond ernstig verontreinigd is kleiner is dan 0.05. Hieraan wordt volgens de richtlijn voldaan wanneer voor alle stoffen het 95-percentiel (P95) binnen de kaarteenheid waaruit de partij grond afkomstig is kleiner is dan de Tussenwaarde die ligt halverwege Interventiewaarde en Streefwaarde. Door de P95 te vergelijken met de Tussenwaarde in plaats van de Interventiewaarde wordt een vaste veiligheidsmarge gehanteerd, die nodig is in verband met onzekerheid over de kans op ernstig verontreinigde grond. In dit rapport wordt voorgesteld om de beslissing wel of geen verplichte partijkeuring (wel of geen vrijstelling) te baseren op een statistische toets op oppervlaktefractie ernstig verontreinigde grond (π). Een belangrijk voordeel van statistisch toetsen is dat de breedte van de veiligheidsmarge gerelateerd wordt aan de mate van onzekerheid over de oppervlaktefractie ernstig verontreinigde grond. Deze onzekerheid is afhankelijk is van de kwaliteit en kwantiteit van de gegevens, en is dus van geval tot geval verschillend. Een voor de hand liggende toets is de klassieke binomiale toets die gebruik maakt van het aantal ernstig verontreinigde monsters in de steekproef. Een nadeel van deze toets is het geringe onderscheidende vermogen, d.w.z. er zijn veel monsters nodig om, in geval $\pi < \pi_0$ (bijv. $\pi < 0.05$), dit met een bepaalde mate van betrouwbaarheid ook te kunnen vaststellen op basis van het steekproefresultaat. Het onderscheidende vermogen kan worden vergroot door het gebruik van een *a priori* inschatting van π in een Bayesiaanse aanpak. Doel van dit onderzoek is het vergelijken van deze Bayesiaanse toets met de klassieke toets in termen van kansen op foute beslissingen, en op basis van de kosten (het verlies) van foute beslissingen.

In de Bayesiaanse aanpak wordt de voorkennis over π beschreven met een *a priori* kansverdeling. Vervolgens worden gegevens verzameld waarmee de *likelihood*-functie van π bepaald wordt, d.w.z. de kans dat π gelijk is aan 0…1, gegeven het aantal ernstig verontreinigde monsters in de steekproef. Deze *likelihood*-functie wordt gebruikt om de *a priori* verdeling van π te *updaten*, wat resulteert in een a *posteriori* verdeling van π. Voor de *a priori* verdeling is gekozen voor een Beta-verdeling omdat de *a posteriori* verdeling dan ook een Beta-verdeling is. Er zijn 6 Beta-verdelingen gebruikt, drie referentieverdelingen, te weten de Beta(1,0), Beta(1,1) (uniforme verdeling), Beta(0.5,0.5) (Jeffrey's verdeling), en drie informatieve verdelingen, te weten Beta(0.8,7.2), Beta(0.425,8.075) en Beta(0.089,8.811). De Beta(1,0)-verdeling is de meest pessimistische verdeling. Voor $\pi < 1$ is de kans 0, voor $\pi = 1$ is de kans 1. De verwachtingswaarde van π voor de laatste drie verdelingen is resp. 0.1, 0.05 en 0.01.

Voor de Bayesiaanse aanpak is de *confidence of compliance* bepaald. d.w.z. de cumulatieve *a posteriori* kans dat $\pi < 0.05$, bij een totaal aantal monsters van 1 tot 200, waarvan er 0, 1, 2 enz. ernstig verontreinigd zijn. Voor de Beta(1,0)-verdeling blijkt het resultaat

identiek te zijn aan dat van de klassieke toets. De toename in *confidence of compliance* tengevolge van het gebruik van één van de drie informatieve *a priori* verdelingen ten opzichte van de Beta(1,0) en uniforme verdeling is voor minder dan 100 monsters aanzienlijk, zeker bij 2 of meer overschrijdingen. Bij een verwachte *a priori* oppervlaktefractie π_{prior} van 0.1 (en variantie van 0.01) is de betrouwbaarheid van de conclusie $\pi < 0.05$ bij 19 monsters met 0 overschrijdingen > 0.80, voor $\pi_{\text{prior}} = 0.05$ (variantie 0.005) geldt dit voor 6 monsters met 0 overschrijdingen, voor $\pi_{\text{prior}} = 0.01$ (variantie 0.001) voor 1 monster dat niet ernstig verontreinigd is. Vergelijk dit met 32 monsters met 0 overschrijdingen voor klassiek toetsen.

Het verlies (Eng. *loss*) van foute beslissingen op basis van statistisch toetsen is berekend als de som van het verlies tengevolge van onterechte vrijstelling (partij wordt niet gekeurd, maar is wel ernstig verontreinigd), het verlies tengevolge van onterechte verplichting tot partijkeuring (partij wordt gekeurd, maar is niet ernstig verontreinigd), en de kosten van bemonstering. Het verlies van beslissen over vrij grondverzet (hergebruik van partijen grond als bodem zonder partijkeuringen) hangt af van een groot aantal factoren. Een kleine oppervlaktefractie ernstig verontreinigd, een groot verwacht aantal partijen grond, en hoge kosten van saneren van hergebruikte, ernstig verontreinigde grond (relatief ten opzichte van partijkeuringen) werken in het voordeel van statistisch toetsen. Ook de onbetrouwbaarheidsdrempel is hierbij van invloed. In geval van 100 partijen, $c_{saneren}$ = 100 000 €, c_{keuren} = 1 500 €, $c_{bemonsteren}$ = 1 500 €, en $\pi = 0.001$, is beslissen op basis van klassiek statistisch toetsen naar verwachting goedkoper dan alle partijen keuren wanneer voldoende, maar niet teveel monsters worden genomen. Het optimale aantal monsters hangt af van de onbetrouwbaarheidsdrempel, en is 32 voor $\alpha = 0.2$, en 45 voor $\alpha = 0.1$. Voor $\pi = 0.01$ is alle partijen keuren naar verwachting goedkoper. Voor Bayesiaans toetsen hangt de het verlies ook af van de *a priori* verdeling. Des te kleiner de verwachting van de *a priori* verdeling van π, des te groter de *confidence of compliance* gegeven het totaal aantal monsters en het aantal overschrijdingen, des te groter de kans op vrijstelling, des te groter de verwachte schade tengevolge van hergebruik van ernstig verontreinigde grond, en des te lager de onnodig gemaakte kosten van partijkeuring. Algemene conclusies zijn moeilijk te trekken, en daarom zijn twee excel-programma's bijgevoegd waarmee door de gebruiker zelf bepaald kan worden of in zijn of haar situatie beslissen op basis van statistisch toetsen naar verwachting goedkoper is dan alle partijen keuren.

De gevoeligheid van de *confidence of compliance* en het onderscheidend vermogen van de Bayesiaanse toets voor de *a priori* verdeling maakt deze toets mogelijk minder geschikt als bewijsmiddel in juridische procedures, omdat de keuze van deze *a priori* verdeling in meer of mindere mate subjectief is. Daar staat tegenover dat de resultaten van klassiek toetsen overeen komen met die van Bayesiaans toetsen gebruikmakend van de Beta(1,0)-verdeling, d.w.z *a priori* wordt verondersteld dat 100% van de oppervlakte ernstig verontreinigd is. Klassiek toetsen geeft daardoor wel een erg pessimistisch beeld van de *confidence of compliance*.

1 Inleiding

Door middel van de Vrijstellingsregeling grondverzet wordt in bepaalde situaties vrijstelling verleend van bepaalde eisen van het Bouwstoffenbesluit bodem- en oppervlaktewateren bescherming (Bouwstoffenbesluit). De vrijstelling geldt voor het hergebruik van licht verontreinigde grond als bouwstof in grondwerken. In bepaalde gevallen mag deze ook als bodem worden hergebruikt. Voor die vrijstelling is het onder meer noodzakelijk dat de partij grond wordt hergebruikt in een gebied waarvan de bodemkwaliteit in kaart is gebracht. De kwaliteit van de toe te passen partij grond moet dan wel vergelijkbaar zijn met die van de ontvangende zone (stilstands-principe). Om dit vast te stellen kan een partijkeuring worden verricht. Echter wanneer ook de bodemkwaliteit van de bronzone in kaart is gebracht, kan vrijstelling van partijkeuring worden verkregen wanneer aangetoond wordt dat de kans dat een willekeurige parij grond uit de bronzone ernstig verontreinigd is, kleiner is dan 0.05.

In de 'Interim-richtlijn 'Opstellen en toepassen bodemkwaliteitskaarten in het kader van de Vrijstellingsregeling grondverzet' (Ministerie VROM et al., 1999) wordt een procedure beschreven voor het maken van bodemkwaliteitskaarten. Ook worden grootheden gedefinieerd voor het toetsen van hypotheses over de kans op een ernstig verontreinigde partij grond en over vergelijkbare kwaliteit, en wordt beschreven wanneer welke conclusie over de kans op ernstig verontreinigde grond en over vergelijkbare kwaliteit moet worden getrokken.

De beslissing 'wel of geen verplichte partijkeuring' wordt genomen op basis van de per stof geschatte 95-percentielwaarde (hierna aangeduid met P95). Deze moeten voor alle (kritische) stoffen kleiner zijn dan de tussenwaarde (T-waarde). Uit de nota 'Grond grondig bekeken' (Ministerie VROM et al., 1999) blijkt dat het in feite gaat het om een inschatting van de kans dat een partij grond afkomstig uit een kaarteenheid (homogene zone) ernstig verontreinigd is. Een partij grond is ernstig verontreinigd als tenminste 1 stof boven de interventiewaarde (I-waarde) ligt. Deze kans moet kleiner dan 0.05 zijn. Wanneer aangenomen wordt dat de kans dat een toekomstige partij grond genomen wordt op een bepaalde locatie voor alle locaties binnen de zone gelijk is, betekent dit dus dat het gaat om een schatting van de oppervlaktefractie 'ernstig verontreinigde grond'. Deze kans (dit oppervlaktefractie) wordt dus geschat door te kijken naar de stofspecifieke P95's. Deze geschatte P95's worden vergeleken met de T-waarden in plaats van de I-waarden om rekening te houden met onzekerheid tengevolge van steekproeffouten: als de geschatte P95 kleiner is dan de T-waarde, dan is de kans groot dat de werkelijke P95 kleiner is dan de I-waarde. Een punt van kritiek op deze methode voor het schatten van de kans op (de oppervlaktefractie) ernstig verontreinigde partij grond is dat wanneer voor alle stoffen de werkelijke P95 kleiner is dan de I-waarde (m.a.w. voor alle stoffen afzonderlijk geldt dat de oppervlaktefractie groter dan de I-waarde kleiner is dan 0.05), de oppervlaktefractie ernstig verontreinigde grond wel groter kan zijn dan 0.05. Dit is het geval wanneer voor meer dan één stof concentraties boven de I-waarde voorkomen waarbij ‘de vereniging’ van de stofspecifieke oppervlaktes groter is dan het maximum van de stofspecifieke oppervlaktes boven de I-waarde.

Een tweede kritiekpunt is dat, om rekening te houden met onzekerheid in het geschatte oppervlaktepercentage boven de I-waarde, een vaste veiligheidsmarge wordt gehanteerd (P95 wordt vergeleken met tussenwaarde in plaats van interventiewaarde). Deze veiligheidsmarge is dus niet gerelateerd aan de kwantiteit en kwaliteit van de gegevens (monsters). Een gemeente wordt hierdoor niet aangemoedigd om meer monsters te nemen om zodoende het onderscheidend vermogen van de toets te verhogen.

Een veel gebruikte methode voor het nemen van beslissingen, rekening houdend met onzekerheid over het steekproefresultaat, is het statistisch toetsen van hypothesen. In deze aanpak wordt de mate van onzekerheid gekwantificeerd, en omdat deze afhankelijk is van de steekproefopzet (aantal monsters en verdeling over gebied), is de veiligheidsmarge niet een vaste waarde, maar o.a. afhankelijk van het aantal monsters. Omdat het gaat om het schatten van oppervlaktefracties, ligt het voor de hand de nulhypothese te toetsen op basis van de aantallen ernstig verontreinigde monsters in de steekproef. Deze aantallen zijn, onder bepaalde aannames, binomiaal verdeeld. Een nadeel van deze toets is het geringe onderscheidende vermogen, d.w.z. er zijn veel monsters nodig om, in geval de $\pi < \pi_0$, dit met een bepaalde mate van betrouwbaarheid ook te kunnen vaststellen op basis van het steekproefresultaat. Een mogelijkheid om dit onderscheidende vermogen te vergroten is het meenemen van een *a priori* inschatting van de oppervlaktefractie ernstig verontreinigd op basis van expert kennis, of op basis van reeds aanwezige monsters. Bayesiaanse statistiek biedt goede mogelijkheden om dergelijke *a priori* inschattingen mee te nemen.

Doel van dit onderzoek is het uitwerken van een Bayesiaanse aanpak voor het toetsen van hypothesen over de kans op ernstig verontreinigde grond (oppervlaktefractie ernstig verontreinigde grond), en het vergelijken van deze methode met de klassieke toets in termen van kansen op foute beslissingen, en op basis van kosten van foute beslissingen.

2 Klassiek statistisch toetsen

2.1 Inleiding

Een beproefde methode voor het nemen van beslissingen rekening houdend met de onzekerheid over de onderzoeksresultaten is het statistisch toetsen van hypothesen. In een toetsingsprocedure kunnen een aantal stappen onderscheiden worden (EPA, 2000). In de eerste stap wordt een doelgrootheid gekozen, en wordt een nulhypothese over dit kenmerk opgesteld. Een doelgrootheid is een combinatie van doelvariabele, statistische parameter en gebied. Een voorbeeld van een doelgrootheid is het gemiddelde (statistische parameter) fosfaatgehalte (doelkenmerk) van perceel A (domein). De doelgrootheid heeft dus betrekking op de werkelijkheid. Voor de beslissing 'wel of geen verplichte partijkeuring' is de doelgrootheid 'oppervlaktefractie ernstig verontreinigde grond van een bodemkwaliteitszone' geschikt, omdat deze doelgrootheid immers de kans bepaalt dat een partij grond, genomen op een willekeurige plek in deze zone, ernstig verontreinigd is. Door hergebruik van een ernstig verontreinigde partij grond als bodem onstaat een nieuw geval van bodemverontreiniging, wat voorkomen dient te worden, en voorkomen kan worden door partijkeuring. In geval van het verzoek om vrijstelling of het aanvragen van een vergunning is het logisch om als nulhypothese te kiezen de toestand waarin geen vrijstelling of vergunning wordt verleend. Het is aan de initiatiefnemer van het grondverzet om te bewijzen dat de kans dat het onderzoeksgebied zich in deze toestand bevindt zeer onwaarschijnlijk is, en dus prima vrijstelling of vergunning kan worden verleend. Bijvoorbeeld, in geval van de vrijstellingsregeling grondverzet is een logische keuze van de nulhypothese 'de oppervlaktefractie ernstig verontreinigde grond is groter of gelijk aan 0.05'. Wanneer de nulhypothese juist is, vindt de overheid de kans dat een partij grond ernstig verontreinigd is te groot om zonder partijkeuring te kunnen worden hergebruikt. Wanneer de initiatiefnemer door middel van grondonderzoek aantoont dat de nulhypothese zeer onwaarschijnlijk is, is de overheid bereid vrijstelling van partijkeuring te verlenen. De nulhypothese wordt getoetst met een toetsingsgrootheid (*test statistic*). Deze toetsingsgrootheid is een functie van de steekproef, bijv. het steekproefgemiddelde, en is dus een stochastische variabele. De hypothese over de doelgrootheid 'oppervlaktefractie ernstig verontreinigde grond' wordt getoetst met de toetsingsgrootheid 'aantal ernstig verontreinigde monsters in de steekproef'.

In de tweede stap wordt een onbetrouwbaarheidsdrempel, α, gekozen, d.w.z. een maximale kans dat de nulhypothese ten onrechte wordt verworpen. In het hiervoor genoemde voorbeeld wordt dus een maximale kans gekozen dat ten onrechte wordt geconcludeerd dat de oppervlaktefractie ernstig verontreinigd kleiner is dan 0.05, en dus ten onrechte vrijstelling van partijkeuring wordt verleend. Het is aan de overheid om deze onbetrouwbaarheidsdrempel te kiezen. Veel gekozen onbetrouwbaarheidsdrempels zijn 0.05 en 0.1, en soms 0.2.

In de derde stap wordt de cumulatieve kans bepaald dat de toetsingsgrootheid de in de uitgevoerde steekproef gerealiseerde waarde aanneemt wanneer de nulhypothese waar zou zijn. In het geval van de toetsingsgrootheid 'aantal ernstig verontreinigde monsters in de steekproef' wordt bepaald hoe groot de kans is dat er e of minder van de n monsters in de steekproef ernstig verontreinigd zijn, wanneer de oppervlaktefractie ernstig verontreinigde grond gelijk is aan 0.05 (e is aantal overschrijdingen d.w.z. aantal ernstig verontreinigde monsters in steekproef, n is totaal aantal monsters). Is deze kans kleiner dan de gekozen onbetrouwbaarheidsdrempel, dan wordt de nulhypothese verworpen en wordt vrijstelling van partijkeuring verleend, is deze groter dan wordt de nulhypothese aangehouden en wordt dus geen vrijstelling verleend. Ook kan eerst op basis van n en α het kritieke aantal overschrijdingen worden bepaald, en vervolgens vergelijken we het werkelijke aantal overschrijdingen met dit kritieke aantal. Als $e < e_{krit}$ dan wordt de nulhypothese verworpen, als $e \geq e_{krit}$ dan wordt deze aangehouden. Voor het bepalen van de cumulatieve kans hebben we de kansverdeling van de toetsingsgrootheid nodig (zie par. 2.3).

2.2 Kans op foute beslissingen

Wanneer beslissingen worden genomen op basis van een statistische toets is het onvermijdelijk dat in een aantal gevallen de verkeerde beslissing wordt genomen. Hierboven is al de onterechte beslissing tot vrijstelling van partijkeuring genoemd. Deze foute beslissing wordt genomen wanneer de werkelijke oppervlaktefractie ernstig verontreinigde grond groter of gelijk is aan 0.05, maar het aantal overschrijdingen in de steekproef kleiner is dan het kritieke aantal. De kans op deze foute beslissing wordt aan een maximum gebonden, de onbetrouwbaarheidsdrempel α. Er is nog een tweede foute beslissing mogelijk, namelijk de onterechte verplichting tot partijkeuring. Deze foute beslissing wordt genomen wanneer de werkelijke oppervlaktefractie ernstig verontreinigde grond kleiner is dan 0.05, maar het aantal ernstig verontreinigde monsters in de steekproef groter is dan de kritieke waarde (dan wordt de nulhypothese immers niet verworpen). De kans op deze foute beslissing, β, ook wel kans op fout van de tweede soort genoemd, hangt af van de oppervlaktefractie ernstig verontreinigde grond. Naarmate de oppervlaktefractie meer afwijkt van 0.05 (dichter bij 0 ligt) is de kans dat de nulhypothese ten onrechte wordt aangehouden (ten onrechte geen vrijstelling wordt verleend) kleiner. Met andere woorden, de kans dat terecht wordt geconcludeerd dat de oppervlaktefractie < 0.05 wordt groter naarmate de oppervlaktefractie dichter bij 0 ligt. Het onderscheidend vermogen (Eng. *power*), $1 - \beta$, neemt dus toe naarmate de oppervlaktefractie dichter bij 0 ligt. Ook de gekozen onbetrouwbaarheidsdrempel α is van invloed op deze kans omdat deze immers de ligging van de kritieke waarde (het kritieke aantal overschrijdingen in de steekproef) bepaalt. Gegeven de oppervlaktefractie, komt een relatief kleine α (hoge betrouwbaarheid) overeen met een relatief grote kans op fout van tweede soort.

2.3 Binomiale toets

De nulhypothese 'de oppervlaktefractie ernstig verontreinigde grond is groter dan 0.05' wordt dus getoetst op basis van het aantal ernstig verontreinigde monsters in de steekproef. Wanneer aangenomen wordt dat de kans op overschrijding (kans dat monster ernstig verontreinigd is) voor alle monsters gelijk is, en de monsters als onafhankelijke waarnemingen kunnen worden beschouwd, is dit aantal ernstig verontreinigde monsters in de steekproef binomiaal verdeeld met als parameters het aantal monsters en de (onbekende) kans op overschrijding (oppervlaktefractie ernstig verontreinigde grond). In dit geval kan de kans op de fout van de eerste soort berekend worden met

$$p_{fout1} = \Pr(e < e_{krit} \mid n, \pi = \pi_0) = \sum_{e=0}^{e_{krit}} \binom{n}{e} \pi_0^{\,e} (1-\pi_0)^{n-e} \leq \alpha, \tag{1}$$

waarin e_{krit} het grootste aantal overschrijdingen (ernstig verontreinigde monsters) is dat voldoet aan $p_{fout1} \leq \alpha$, en π_0 de in de nulhypothese genoemde oppervlaktefractie is (0.05).

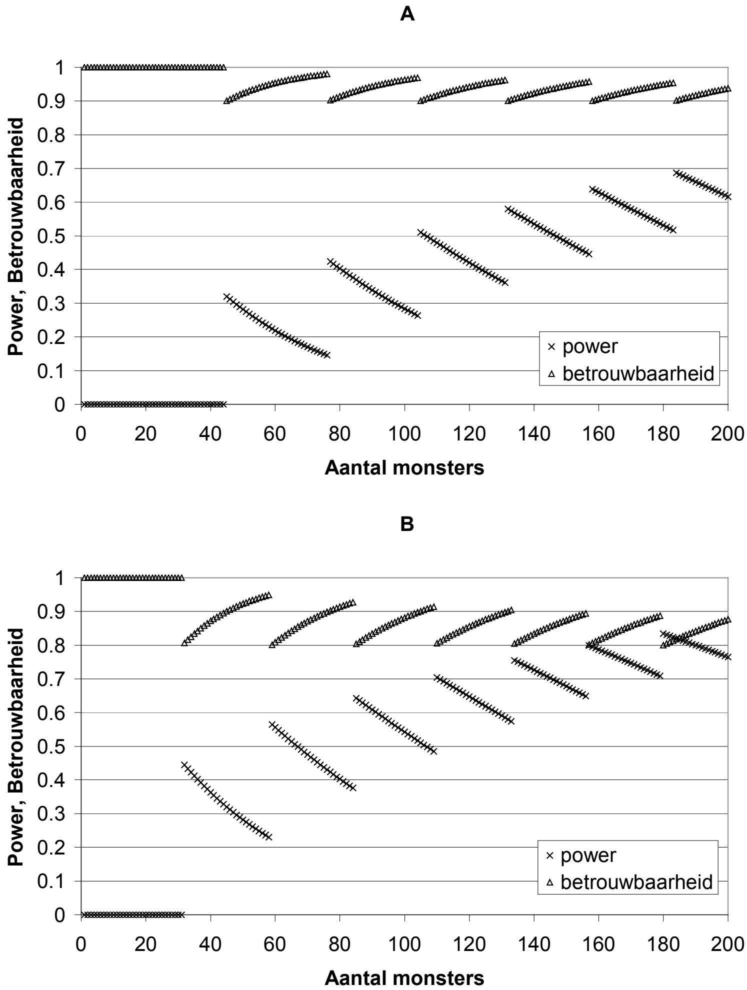
Het onderscheidend vermogen, 1- de kans op onterechte verplichting tot partijkeuring, bij een werkelijk oppervlaktefractie ernstig verontreinigd van $\pi_i < \pi_0$ kan worden berekend met

$$power(\pi_i) = \Pr(e < e_{krit} \mid n, \pi = \pi_i) = \sum_{e=0}^{e_{krit}} \binom{n}{e} \pi_i^{\,e} (1-\pi_i)^{n-e}. \tag{2}$$

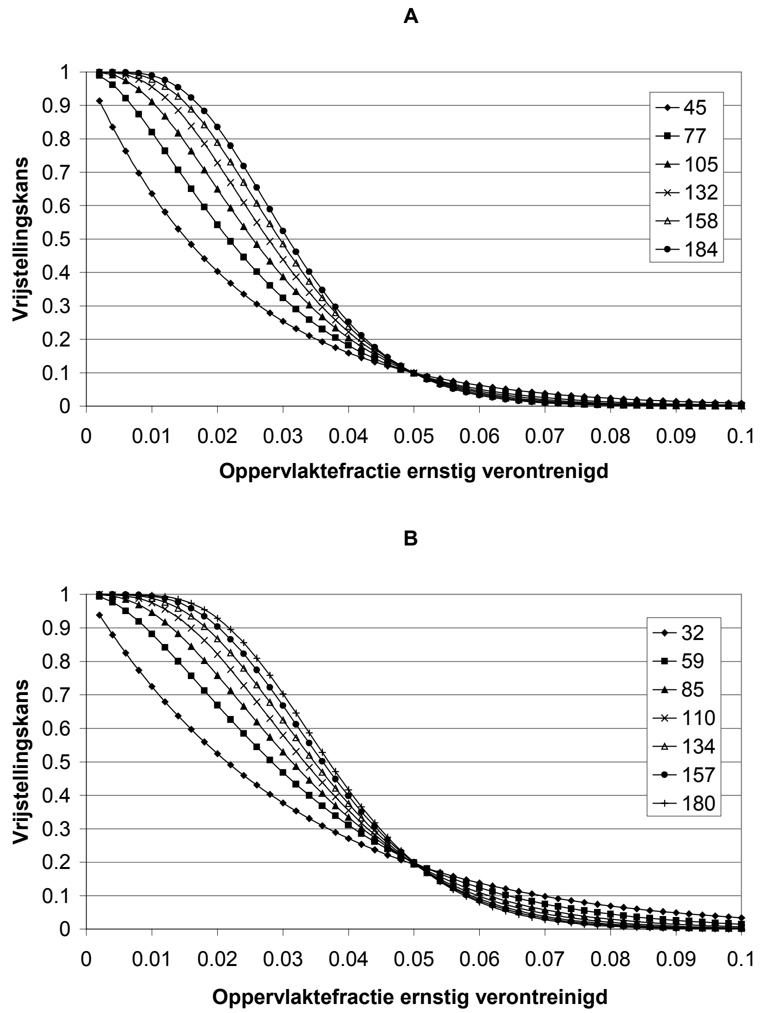
Voor waardes van $\pi_i > \pi_0$ is de met vergelijking 2 berekende kans gelijk aan de kans op onterechte vrijstelling.

In figuur 1A is weergegeven het onderscheidend vermogen voor een oppervlaktefractie ernstig verontreinigde grond π van 0.025, wanneer de nulhypothese '$\pi \geq 0.05$' binomiaal getoetst wordt bij een minimale betrouwbaarheid van 0.9. Uit deze figuur blijkt dat bij 45 monsters het onderscheidend vermogen 0.32 is, m.a.w. de kans dat ten onrechte wordt geconcludeerd dat $\pi \geq 0.05$ (en dus onterecht geen vrijstelling van partijkeuring wordt verleend) is 1 – 0.32 = 0.68. Op het eerste gezicht merkwaardig is dat het onderscheidend vermogen tussen 45 en 76 monsters geleidelijk afneemt tot 0.15. Dit komt omdat voor 45 tot 76 monsters in de steekproef het kritieke aantal ernstig verontreinigde monsters e_{krit} in de steekproef gelijk blijft, namelijk 0. Als 0 van de 45 monsters ernstig verontreinigd zijn wordt de nulhypothese verworpen. Hetzelfde geldt voor 0 van de 76 monsters. Pas bij 77 monsters wordt ook bij 1 ernstig verontreinigd monster in de steekproef de nulhypothese verworpen. De sprongen in de *power*-curve vallen dus samen met sprongen in e_{krit} in de steekproef. Merk op dat de afname van het onderscheidend vermogen tussen 45 en 76 monsters gepaard gaat met een toename van de betrouwbaarheid, m.a.w. de kans dat de nulhypothese ten onrechte wordt verworpen (kans dat ten onrechte vrijstelling tot partijkeuring wordt verleend) neemt geleidelijk af van 0.1 tot 0.2. Merk verder op dat tot 44 monsters de nulhypothese nooit wordt verworpen, ook niet bij 0 ernstig verontreinigde monsters in de steekproef, omdat anders de kans dat de nulhypothese ten onrechte wordt verworpen groter is dan 0.1, de gekozen onbetrouwbaarheidsdrempel.

Wanneer voor een andere α wordt gekozen, dan komen de sprongen in de kritieke aantallen ernstig verontreinigde monsters in de steekproef op een andere plaats (bij andere aantallen monsters) te liggen. Figuur 1B geeft het onderscheidend vermogen en de betrouwbaarheid weer voor α = 0.2 (minimale betrouwbaarheid van 0.8). In vergelijking met figuur 1A zijn de lijnstukken verder naar links doorgetrokken, en zijn de lijnstukken aan de rechterkant ingekort. De lijnstukken zijn afgekapt waar de betrouwbaarheid onder de gekozen minimale betrouwbaarheid zakt (0.9 in figuur 1A, 0.8 in figuur 1B).



Figuur 1 Onderscheidend vermogen (power) en betrouwbaarheid van klassieke toets van hypothese π > 0.05, voor π = 0.025 bij een onbetrouwbaarheidsdrempel van 0.1 (A) en 0.2 (B)



Figuur 2 Kans op vrijstelling bij klassieke toets van de hypothese $\pi > 0.05$ met een onbetrouwbaarheidsdrempel van 0.1 (A) en 0.2 (B).

Tot slot is voor een groot aantal waarden voor de oppervlaktefractie ernstig verontreinigde grond de kans berekend dat de nulhypothese wordt verworpen en dus vrijstelling wordt verleend. Voor oppervlaktefracties < 0.05 is deze verwerping terecht en wordt dus terecht vrijstelling verleend, voor oppervlaktefracties ≥ 0.05 is deze verwerping onterecht, en wordt onterecht vrijstelling verleend. Figuur 2A toont deze kans op vrijstelling wanneer de nulhypothese getoetst wordt bij $\alpha = 0.10$, figuur 2B bij $\alpha = 0.20$. Voor het aantal monsters zijn de lokale optima genomen uit figuur 1. Uit figuur 2 blijkt dat voor grotere monsteraantallen de vrijstellingscurve steiler verloopt. Merk verder op dat de kans op vrijstelling bij een oppervlaktefractie van 0.05 voor deze monsteraantallen gelijk is aan α. Het gebied onder de vrijstellingscurve rechts van 0.05 is het gebied waar tenonrechte vrijstelling wordt verleend, het gebied boven de curve links van 0.05 is het gebied waar tenonrechte geen vrijstelling wordt verleend.

3 Bayesiaans toetsen

Een alternatief voor het klassiek statistisch toetsen van hypothesen is Bayesiaans toetsen. Een belangrijk voordeel van een Bayesiaans aanpak is dat voorkennis over de nulhypothese en alternatieve hypothese (of over de te schatten parameter) gemakkelijk meegenomen kan worden in de analyse, wat de efficiëntie van de steekproef ten goede kan komen. Deze voorkennis wordt beschreven in termen van kansen, bijv. de *a priori* kans dat de nulhypothese juist is en de *a priori* kans dat de alternatieve hypothese juist is, of in geval van een parameter de *a priori* kansverdeling van de parameter. Voor een uiteenzetting over Bayesiaanse statistische methodes verwijzen we naar Congdon (2001). In de context van dit onderzoek is met name par. 2.5.1 uit dit boek en het artikel van McBride en Ellis (2001) relevant omdat het hierin gaat over het updaten van de *a priori* kansverdeling van parameter π, de kans op overschrijding, met binomiale data. Dit updaten van de *a priori* kansverdeling met de verzamelde gegevens gaat als volgt

$$h(\pi \mid e,n) = \left[\frac{L(\pi \mid e,n)}{\int_0^1 L(\pi \mid e,n) g(\pi) \mathrm{d}\pi} \right] g(\pi) \tag{3}$$

waarin e het aantal overschrijdingen in de n monsters van de steekproef is, $L(\pi \mid e,n)$ is de *likelihood* functie van π, gegeven e en n, $g(\pi)$ is de *a priori* kansverdeling van de oppervlaktefractie ernstig verontreinigde grond (kans op ernstig verontreinigde grond), en $\mathrm{h}(\pi \mid e,n)$ is de *a posteriori* kansverdeling. De *likelihood* functie voor een steekproef van n monsters is de kansdichtheid functie voor π, gegeven het aantal overschrijdingen e in de steekproef.

Congdon (2001) raadt aan voor de binomiale *likelihood* functie de beta verdeling Beta(a,b) als *a priori* verdeling te nemen:

$$p(\pi) = \frac{\pi^{a-1}(1-\pi)^{b-1}}{\int\limits_0^1 \pi^{a-1}(1-\pi)^{b-1} \mathrm{d}\pi.}, \tag{4}$$

In dit geval is de *a posteriori* verdeling namelijk ook een beta-verdeling met parameters $a+e$ en $b+n-e$ (Congdon, 2001. p.29):

$$p(\pi \mid e,n) = \frac{\pi^{a+e-1}(1-\pi)^{b+n-e-1}}{\int\limits_0^1 \pi^{a+e-1}(1-\pi)^{b+n-e-1} \mathrm{d}\pi.}, \tag{5}$$

en is de cumulative *a posteriori* verdeling (McBride en Ellis, 2001):

$$F(\pi < \pi_0 \mid e, n) = \frac{\int_0^{\pi_0} \pi^{a+e-1}(1-\pi)^{b+n-e-1}}{\int_0^1 \pi^{a+e-1}(1-\pi)^{b+n-e-1}} \tag{6}$$

Deze cumulatieve kans wordt wel de *confidence of compliance* (CC) genoemd. Het complement, 1-CC, wordt de *confidence of failure* (CF) genoemd, en het complement, 1-CC, de *confidence of compliance* (CF).

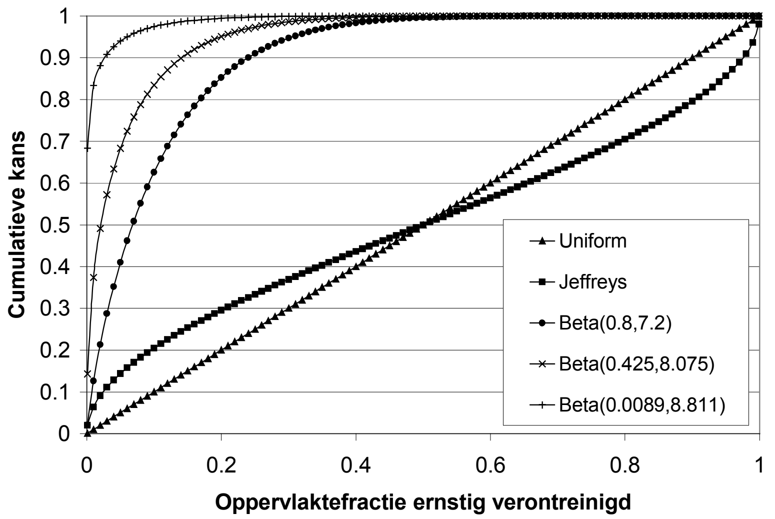
De *confidence of compliance* is berekend voor 6 *a priori* kansverdelingen van de kans op overschrijding. Als eerste *a priori* verdeling is gekozen voor een verdeling met kansdichtheid 0 voor oppervlaktefracties < 1, en kansdichtheid 1 voor oppervlaktefractie gelijk aan 1. Dit is de meest pessimistische inschatting, n.l. 100% van het oppervlak is ernstig verontreinigd. Deze verdeling komt overeen met de Beta(1,0)-verdeling. Een realistischer alternatief is om als *a priori* verdeling de uniforme kansverdeling te gebruiken. Dit impliceert dat we geen enkele idee hebben over de oppervlaktefractie ernstig verontreinigd. De uniforme verdeling komt overeen met de Beta(1,1) verdeling. Veel gebruikt wordt de Jeffreys verdeling, Beta(0.5,0.5), een U-vormige verdeling die dicht bij 0 en 1 steil omhoog loopt (zie figuur 3). Deze eerste drie verdelingen zijn referentie-verdelingen, verdelingen die gebruikt kunnen worden als je niets (zeer weinig) weet, of waarmee de resultaten van meer informatieve verdelingen vergeleken kunnen worden. In veel gevallen is het aannemelijk dat de kans op overschrijding (oppervlaktefractie ernstig verontreinigde grond) niet erg groot is. Als *a priori* schattingen van deze kans zijn gebruikt 0.1 (met een variantie s^2 van 0.01), 0.05 (variantie 0.005) en 0.01 (variantie 0.001). Met deze *a priori* schattingen van de kans en de variantie van de kans kunnen de parameters *a* en *b* van de beta-verdeling worden berekend (Congdon, 2001, p.31) volgens

$$a = \pi_{prior}\left(\frac{\pi_{prior}(1-\pi_{prior})}{s^2} - 1\right) \text{en } b = a\left(\frac{1}{\pi_{prior}} - 1\right). \tag{7}$$

Dit leidt tot de volgende 6 verdelingen

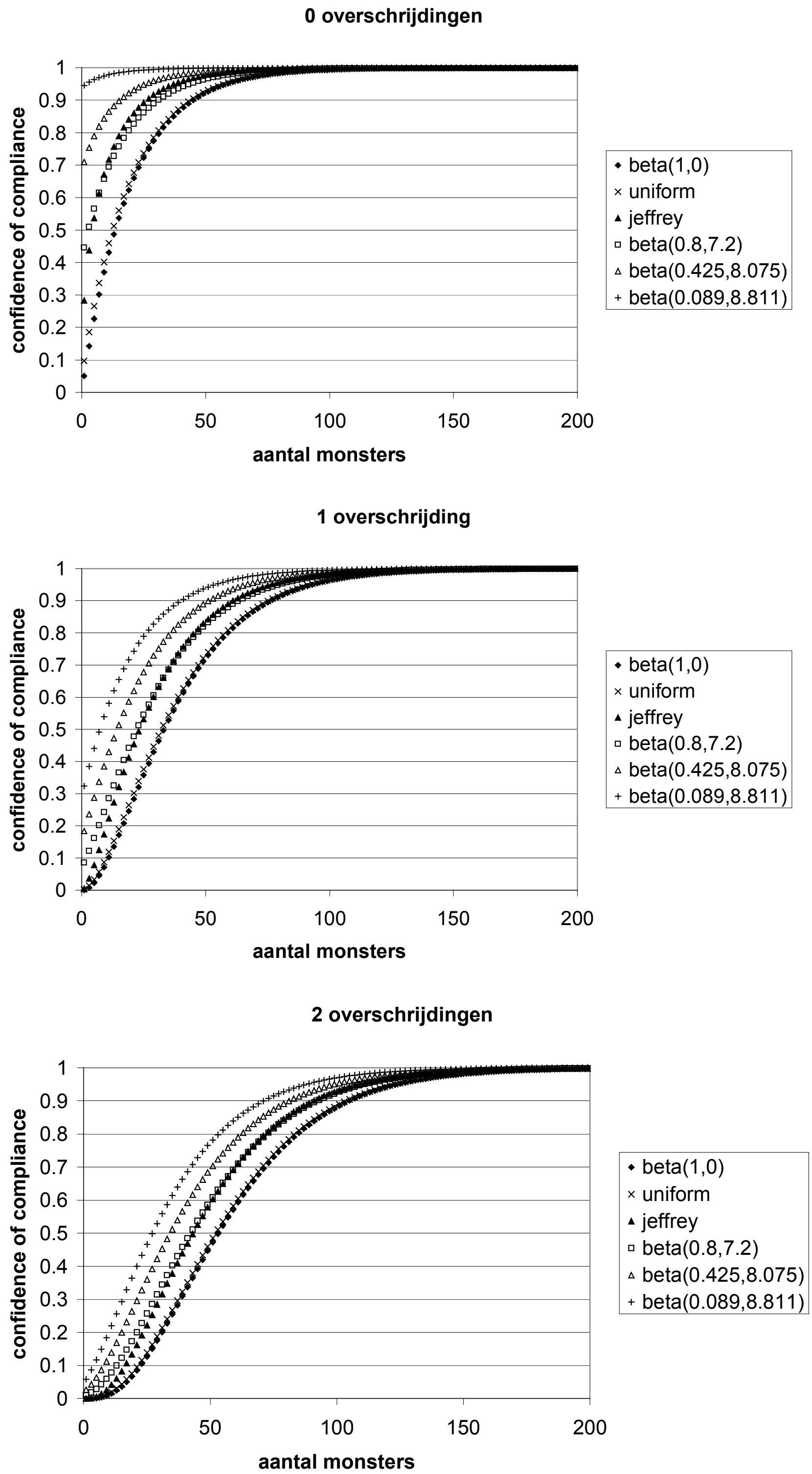
1. Beta(1,0)-verdeling
2. Beta(1,1) (Uniforme verdeling)
3. Beta (0.5,0.5) (Jeffreys verdeling)
4. Beta(0.8,7.2)-verdeling (π_{prior}=0.1, s^2=0.01)
5. Beta(0.425,8.075)-verdeling (π_{prior}=0.05, s^2=0.005)
6. Beta(0.089,8.811)-verdeling (π_{prior}=0.01, s^2=0.001)

In figuur 3 zijn de *a priori* verdelingen weergegeven.



Figuur 3 Vijf a priori kansverdelingen voor oppervlaktefractie ernstig verontreinigde grond

In figuur 4 is de *confidence of compliance* weergegeven voor 0, 1 en 2 overschrijdingen als functie van het aantal monsters (1-200) in geval van een 5%-norm (oppervlaktefractie ernstig verontreinigde grond is 0.05). Uit deze figuur blijkt dat de betrouwbaarheid tengevolge van het gebruik van de *a priori* verdelingen is toegenomen. In tabel 1 en 2 is het maximaal aantal overschrijdingen weergegeven als functie van het totaal aantal monsters teneinde een minimale betrouwbaarheid te waarborgen. Door de hogere betrouwbaarheid van de Bayesiaanse toets is voor bepaalde aantallen monsters het kritieke aantal overschrijdingen (e_{krit}) 1 groter zijn dan voor de klassieke toets. Hierdoor is de *power* van de Bayesiaanse toets groter (figuur 5). Deze *power* is op dezelfde manier berekend als bij de klassieke toets (vergelijking 2). Uit figuur 5 (en tabel 2) kan worden afgelezen dat voor de Beta (0.0089,8.811)-verdeling 1 monster dat niet ernstig verontreinigd is, voldoende is om met 80% betrouwbaarheid aan te tonen dat de oppervlaktefractie ernstig verontreinigde grond < 0.05, voor de Beta(0.425,8.075) zijn hiervoor minstens 6 monsters nodig waarvan 0 ernstig verontreinigd enz. Tot slot geeft figuur 6 de kans op vrijstelling weer als functie van π wanneer de Beta(0.087,8.811) als *a priori* verdeling wordt gebruikt. Vergelijken we deze figuur met figuur 2 dan zien we dat de kans op vrijstelling voor π's < 0.05 aanzienlijk groter zijn dan voor de klassieke toets. Bijv. bij 32 monsters is bij $\pi = 0.02$ de kans op vrijstelling bij de klassieke toets 0.52, en bij 27 monsters bij de Bayesiaanse toets 0.92. Echter, de kans op onterechte vrijstelling ($\pi > 0.05$) is voor de Bayesiaanse toets ook aanzienlijk groter dan voor de klassieke toets, met name voor kleine aantallen monsters en oppervlaktefracties net boven 0.05. Voor $\pi = 0.05$ is deze kans 0.61 voor 27 monsters en 0.35 voor 175 monsters (bij klassieke toets 0.2, ongeacht aantal monsters). Als toetje heb ik ook de *confidence of compliance* berekend in geval van een 1%-norm ($\pi_0 = 0.01$), zie figuur 7. Gegeven het aantal monsters is de *confidence of compliance* kleiner. Over het algemeen is de toename in *confidence* door het meenemen van een *a priori* verdeling bij de 1%-norm groter dan bij de 5%-norm.



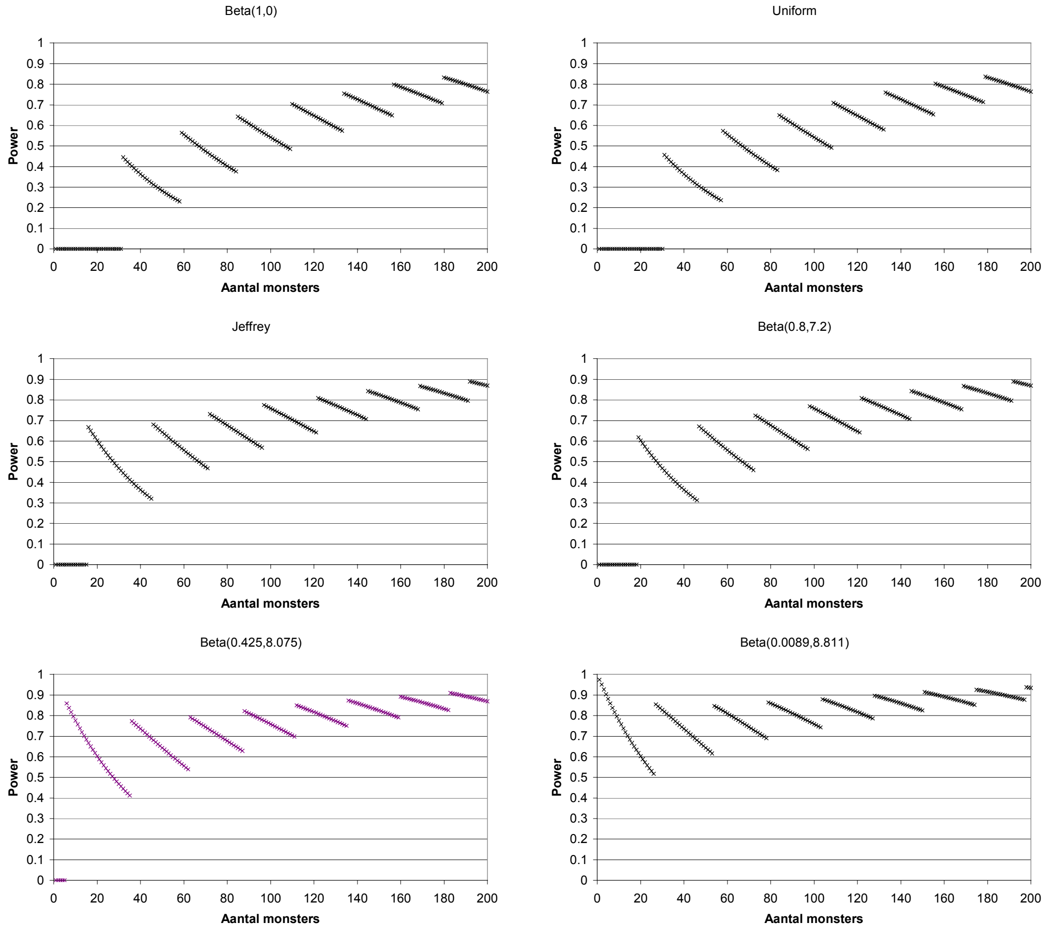
Figuur 4 Confidence of compliance voor 0, 1 en 2 overschrijdingen in geval van $\pi_0 = 0.05$

Tabel 1 Aantal monsters en maximaal toegestaan aantal overschrijdingen voor een confidence of failure < 0.1 voor π_0 = 0.05

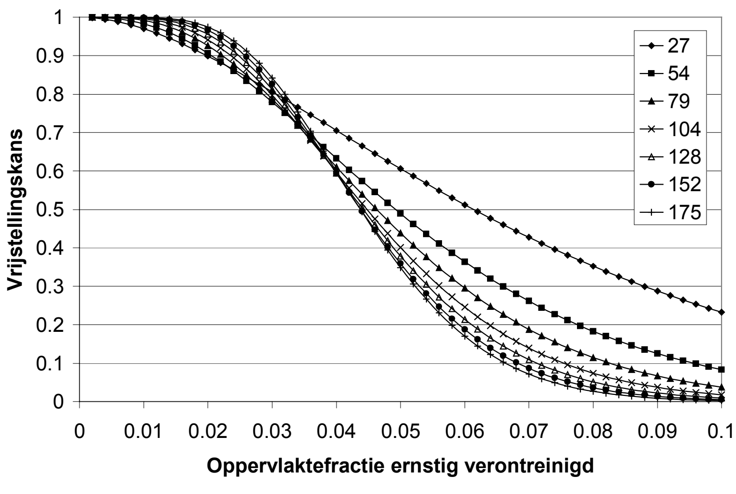
Maximaal aantal overschrij-dingen	Klassiek binomiaal toetsen	Bayesiaans toetsen					
		Beta (1,0)	Uniform	Jeffreys	Beta (0.8,7.2)	Beta (0.425,8.075)	Beta (0.0089,8.811)
0	45-76	45-76	44-75	27-61	31-63	16-51	1-40
1	77-104	77-104	76-103	62-90	64-92	52-81	41-71
2	105-131	105-131	104-130	91-118	93-119	82-108	72-99
3	132-157	132-157	131-156	119-144	120-145	109-135	100-126
4	158-183	158-183	157-182	145-170	146-171	136-161	127-152
5	184-208	184-208	183-207	171-195	172-196	162-186	153-177

Tabel 2 Aantal monsters en maximaal toegestaan aantal overschrijdingen voor een confidence of failure < 0.2 voor π_0 = 0.05

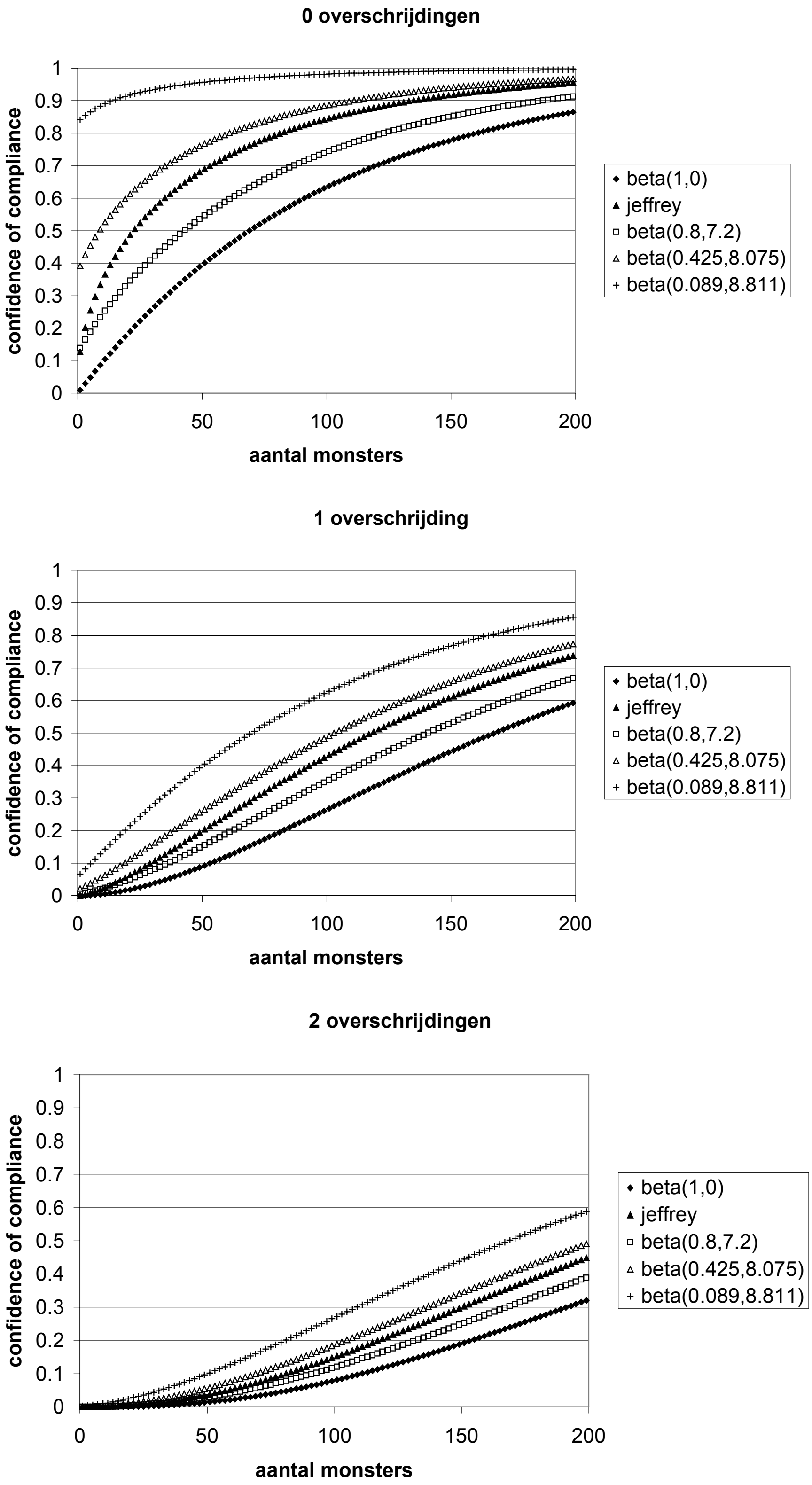
Maximaal aantal overschrij-dingen	Klassiek toetsen	Bayesiaans toetsen					
		Beta (1,0)	Uniform	Jeffreys	Beta (0.8,7.2)	Beta (0.425,8.075)	Beta (0.0089,8.811)
0	32-58	32-58	31-57	16-45	19-46	6-35	1-26
1	59-84	59-84	58-83	46-71	47-72	36-62	27-53
2	85-109	85-109	84-108	72-96	73-97	63-87	54-78
3	110-133	110-133	109-132	97-121	98-121	88-111	79-103
4	134-156	134-156	133-155	122-144	122-144	112-135	104-127
5	157-179	157-179	156-178	145-168	145-168	136-159	128-150



Figuur 5 Onderscheidend vermogen van Bayesiaanse toets van hypothese $\pi > 0.05$ bij $\pi = 0.025$ en een confidence of failure < 0.2, voor 6 a priori kansverdelingen voor π.



Figuur 6 Kans op vrijstelling bij Bayesiaanse toets van de hypothese $\pi > 0.05$ bij een confidence of failure < 0.2, met Beta(0.089,8.811) als a priori verdeling.



Figuur 7 Confidence of compliance voor 0, 1 en 2 overschrijdingen in geval van $\pi_0 = 0.01$

4 Statistisch toetsen of keuren?

Dit hoofdstuk gaat over de financiële aspecten van beslissen over vrij grondverzet op basis van statistisch toetsen op oppervlaktefractie ernstig verontreinigde grond. De kosten, of beter gezegd het verlies (Eng. *loss*), van foute beslissingen op basis van statistisch toetsen worden vergeleken met de kosten van keuren van alle vrijkomende partijen grond. We gaan er van uit dat in het laatste geval geen foute beslissingen worden genomen, d.w.z. een ernstig verontreinigde partij grond wordt nooit hergebruikt als bodem. Wanneer op basis van statistisch toetsen vrijstelling wordt verleend, bestaat er een kans dat de als bodem hergebruikte partij grond toch ernstig verontreinigd is, waardoor een nieuw geval van bodemverontreiniging wordt gecreëerd. Merk op dat, wanneer $\pi < \pi_0$, statistisch gezien weliswaar een juiste beslissing wordt genomen (terechte vrijstelling), maar toch schade kan ontstaan door hergebruik van ongekeurde partijen grond. De kosten van het wegnemen van dit nieuwe geval van bodemverontreiniging zullen vaak hoger zijn dan die van het reinigen van de partij grond omdat

1. onderzoek nodig is om het nieuwe geval af te bakenen (begrenzen);
2. door vermenging met onderliggende bodem meer grond gereinigd moet worden dan is toegepast;
3. inmiddels ook grondwater verontreinigd is;
4. lokatie inmiddels bebouwd is;
5. schade is aangericht aan de gezondheid van mens of levende natuur.

Het verwachte verlies kan worden berekend met

$$L_{hergebruik} = \Pr_{vrijstelling} * \pi * (c_{saneren} - c_{keuren}) * m, \tag{8}$$

waarin:

$\Pr_{vrijstelling}$ = de kans op vrijstelling;
π = oppervlaktefractie ernstig verontreinigde grond;
$c_{saneren}$ = de kosten van saneren van als bodem hergebruikte, ernstig verontreinigde partij grond;
c_{keuren} = de kosten van ex-situ partijkeuringen per partij
m = verwacht aantal partijen grond (max. 2000 ton) waarvoor een beslissing moet worden genomen.

Wanneer op basis van statistisch toetsen geen vrijstelling wordt verleend, is er een kans dat de partij toch niet ernstig verontreinigd is (schoon of licht verontreinigd dus) en de partij grond dus in principe zonder partijkeuring had kunnen worden hergebruikt als bodem. In dit geval worden dus onnodige kosten gemaakt voor het (*ex situ*) keuren van partijen grond. Deze verwachte kosten kunnen worden berekend met:

$$L_{partijkeuring} = (1 - \Pr_{vrijstelling}) * (1 - \pi) * c_{keuren} * m, \tag{9}$$

Tot slot zijn er de kosten van bemonstering, d.w.z. de kosten van de *in situ* partijkeuringen waarmee de nulhypothese getoetst wordt. De totale *loss* is gelijk aan

$$Loss = n * c_{bemonsteren} + L_{vrijstelling} + L_{partijkeuring}, \quad (10)$$

waarin:
n = aantal monsters;
$c_{bemonsteren}$ = de kosten van in-situ partijkeuringen per partij (monster).

Met het excel-programma *Loss(n)-klassiek.xls* op de bijgevoegde CD-Rom kan de *loss* worden berekend als functie van het aantal monsters, voor gegeven instellingen van de onbetrouwbaarheidsdrempel α, de werkelijke oppervlaktefractie ernstig verontreinigd, het verwachte aantal partijen grond, de kosten van *in situ* en van *ex situ* partijkeuringen, en de kosten van saneren na hergebruik. In dit programma wordt uitgegaan van een klassieke toets op oppervlaktefractie ernstig verontreinigde grond. Voor een Bayesiaanse toets kan de *loss* berekend worden met het programma *Loss(n)_Bayes.xls*. In dit programma moet ook een *a priori* inschatting van π, en een nauwkeurigheid (variantie) van deze schatting ingevoerd worden.

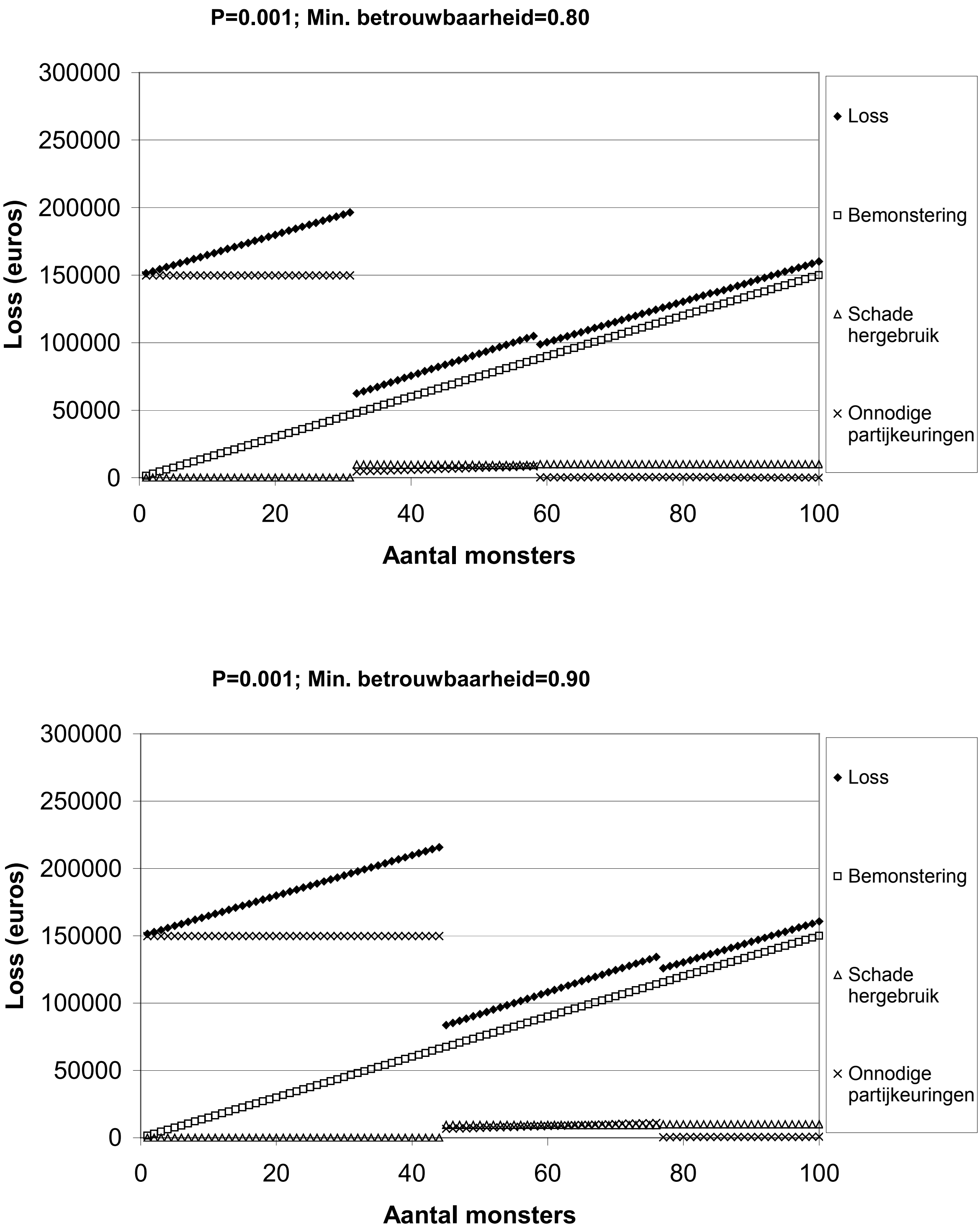
Met deze programma's kan worden bepaald of de aanpak van statistisch toetsen op oppervlaktefractie ernstig verontreinigd naar verwachting tot kostenbesparingen leidt ten opzichte van alle in de toekomst vrijkomende partijen grond keuren. Een kleine oppervlaktefractie ernstig verontreinigd, een groot verwacht aantal partijen grond, en hoge kosten van saneren van hergebruikte, ernstig verontreinigde grond (relatief ten opzichte van partijkeuringen) werken in het voordeel van statistisch toetsen. Ook α is hierbij van invloed. Figuur 8 geeft weer de *loss* in geval van m = 100, $c_{saneren}$ = 100 000 €, c_{keuren} = 1 500 €, en $c_{bemonsteren}$ = 1 500 €, voor de klassieke toets. Uit de figuur blijkt dat voor π = 0.001 beslissen op basis van statistisch toetsen goedkoper is dan alle partijen keuren (kosten 100 * 1 500 = 150 000 €) wanneer voldoende, maar niet teveel monsters worden genomen. Het optimale aantal monsters hangt af van α, en is 32 voor een drempel van 0.2, en 45 voor een drempel van 0.1 (tabel 3). Het is van belang op te merken dat de kosten van het maken van de bodemkwaliteitskaart niet zijn meegenomen, alleen de kosten van bemonstering (in situ partijkeuringen) binnen de kaarteenheden. Bij α = 0.2 en het optimale aantal monsters van 32 mogen deze kosten niet hoger zijn dan 150 000 – 62 406 = 87 594 €. Zijn de kosten hoger dan kunnen beter alle 100 partijen gekeurd worden. Voor π = 0.01 is het bij deze instellingen van verwacht aantal partijen grond, saneringskosten enz. goedkoper om alle partijen te keuren.

Wanneer de beslissing wel of geen vrijstelling wordt gebaseerd op een Bayesiaanse toets, hangt het verlies tevens af van de *a priori* verdeling van π. In geval van π = 0.001 en een maximale *confidence of failure* van 0.2 is het optimale aantal monsters voor de Beta(0.8,7.2) verdeling 19 monsters, voor de Beta(0.425,8.075) 6 monsters, en voor de Beta(0.089,8.811)-verdeling 1 monster (figuur 9, tabel 3). Dit zijn dus 13, 26 en 31 monsters minder dan bij de klassieke toets. Het optimale aantal monsters om met betrouwbaarheid van tenminste 0.80 aan te tonen dat π < 0.05, neemt dus af naarmate de verwachting van π_{prior} kleiner wordt. In tegenstelling tot klassiek binomiaal toetsen is ook bij $\pi = 0.01$ Bayesiaans toetsen goedkoper dan alle partijen keuren.

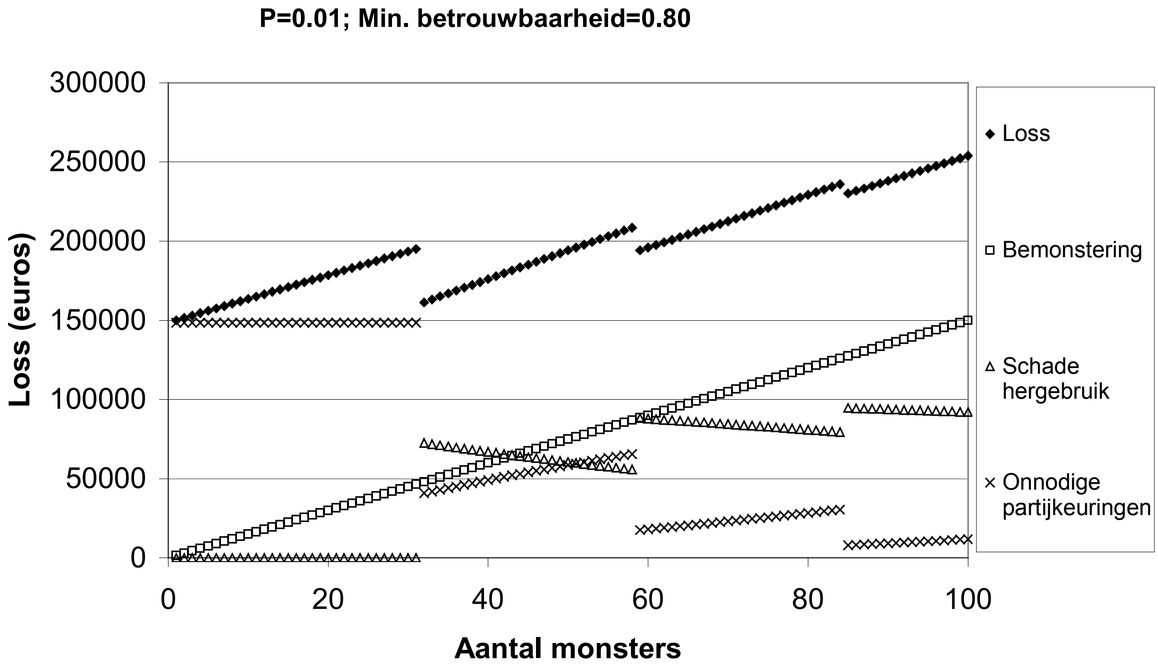
In het bovenstaande ben ik er van uit gegaan dat er geen bestaande monsters zijn die kunnen worden gebruikt bij het toetsen. Wanneer dit wel het geval is heeft dit invloed op de *loss* van de statistische toets. Wanneer alle bestaande monsters niet ernstig verontreinigd zijn, wordt de *loss* kleiner, en daarmee de winst ten opzichte van 'alle partijen keuren' groter. In de programma's *Loss(n)_klassiek* en *Loss(n)_Bayes* is er de mogelijkheid rekening te houden met bestaande monsters (zie ook par. 5.1).

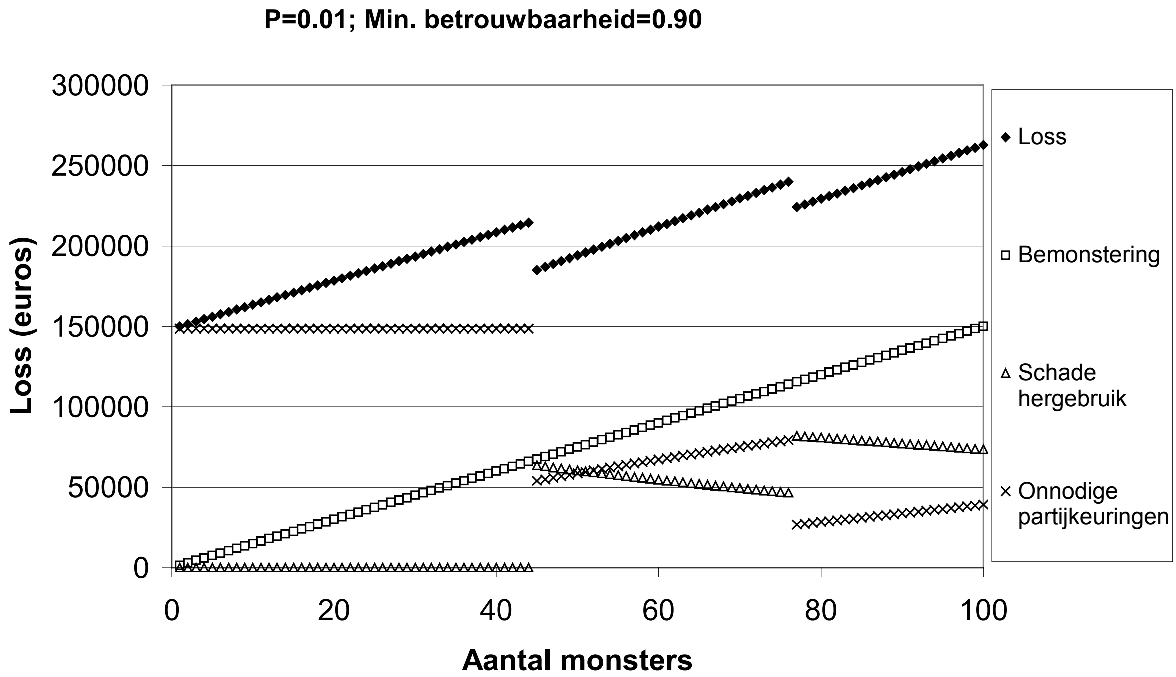
Tabel 3 Optimale aantallen monsters (aantallen met minimale loss) voor klassieke biniomale toets en Bayesiaanse toets bij $m = 100$, $c_{keuren} = 1\ 500$ €, $c_{bemonsteren} = 1\ 500$ €, $c_{saneren} = 100\ 000$ €; - loss groter dan wanneer alle 100 partijen gekeurd worden.

π_0	α	π	Klassiek binomiaal	Be(0.5,0.5)	Be(0.8,7.2)	Be(0.426,8.075)	Be(0.089,8.811)
0.05	0.2	0.01	32	16	19	6	1
0.05	0.1	0.01	45	27	31	16	1
0.05	0.2	0.001	32	16	19	6	1
0.05	0.1	0.001	45	-	31	16	1
0.01	0.2	0.01	-	-	-	-	1
0.01	0.1	0.01	-	-	-	-	14
0.01	0.2	0.001	-	82	-	61	1
0.01	0.1	0.001	-	-	-	-	14

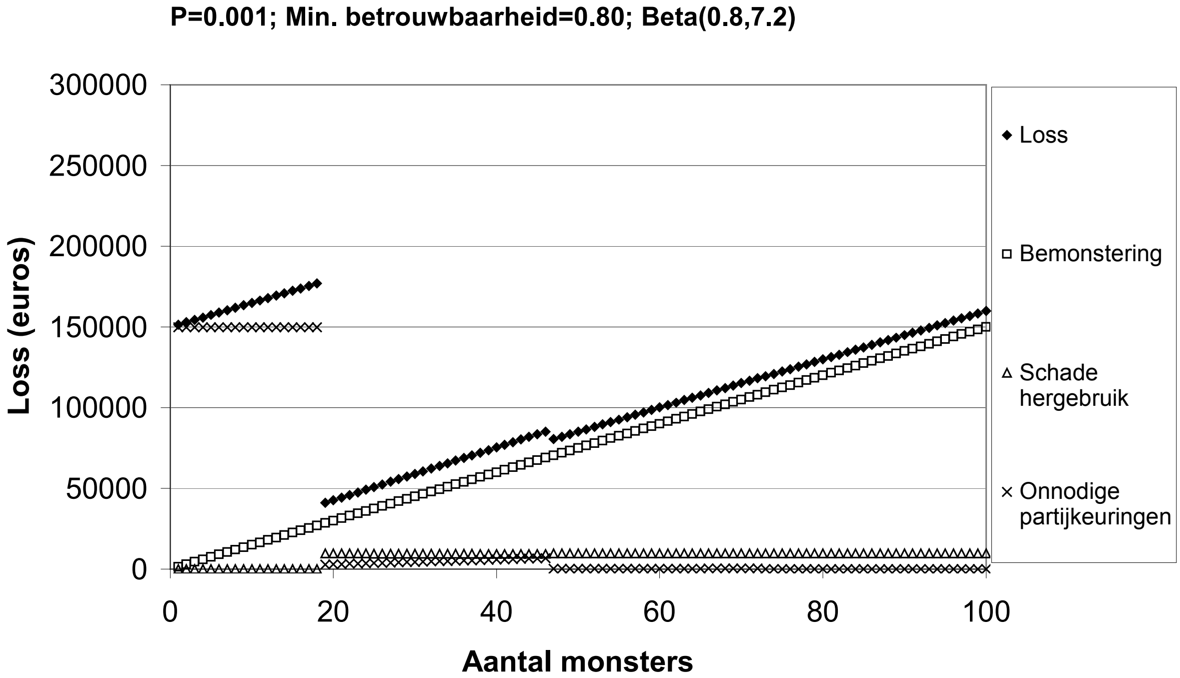


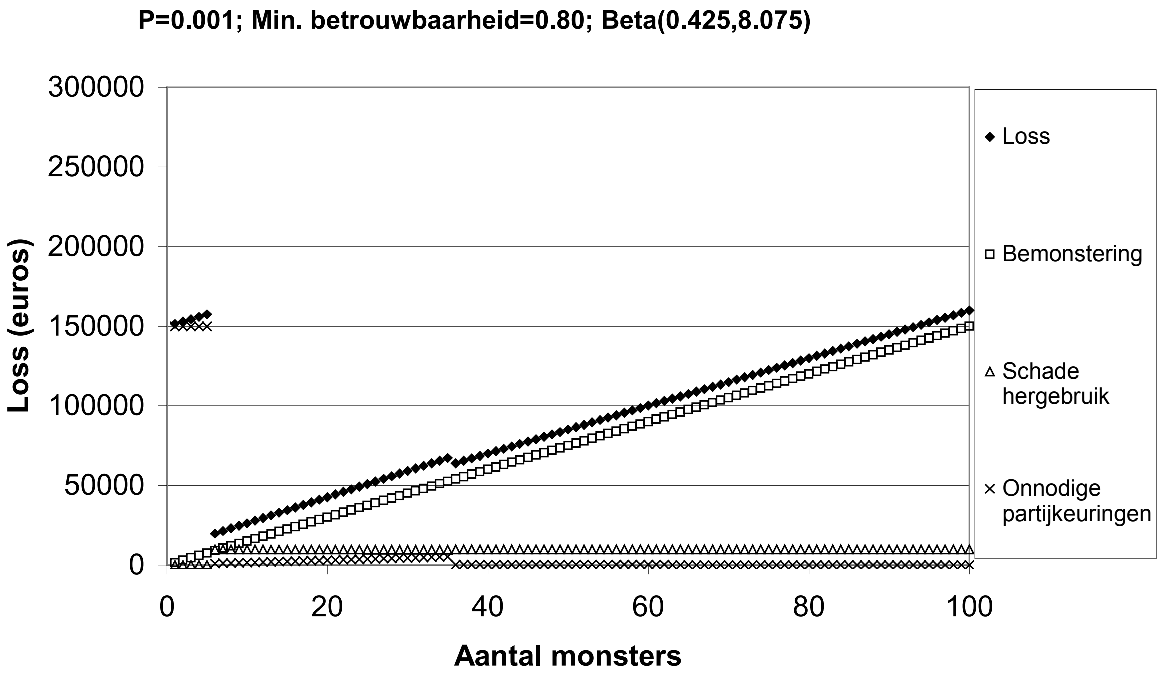
Figuur 8 Voor toelichting, zie volgende pagina



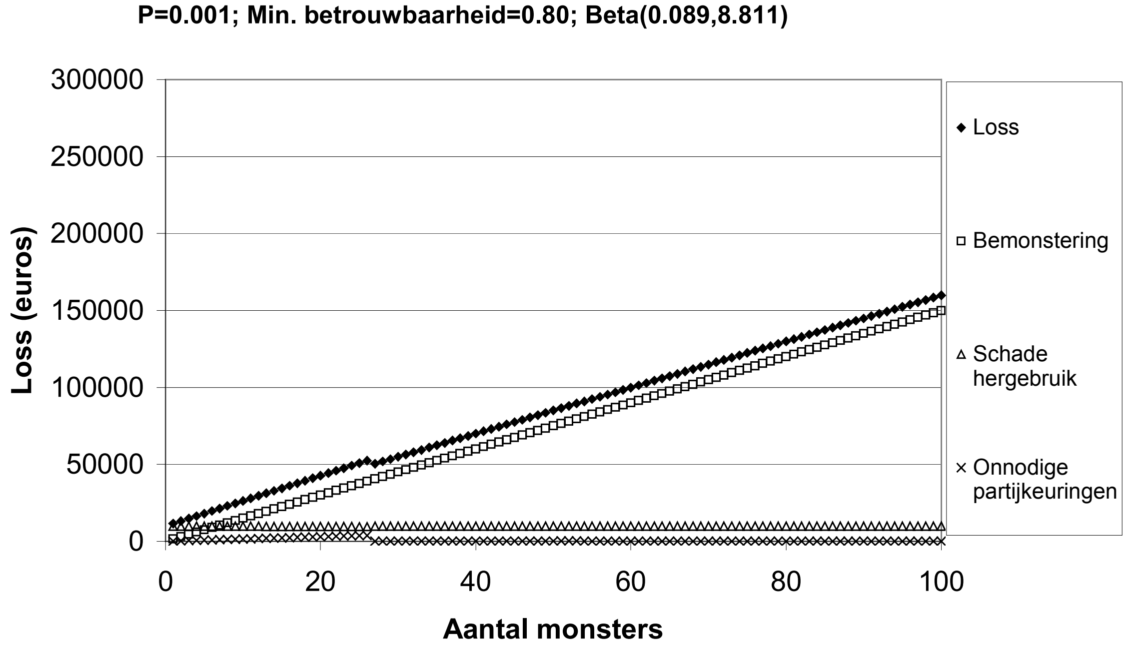


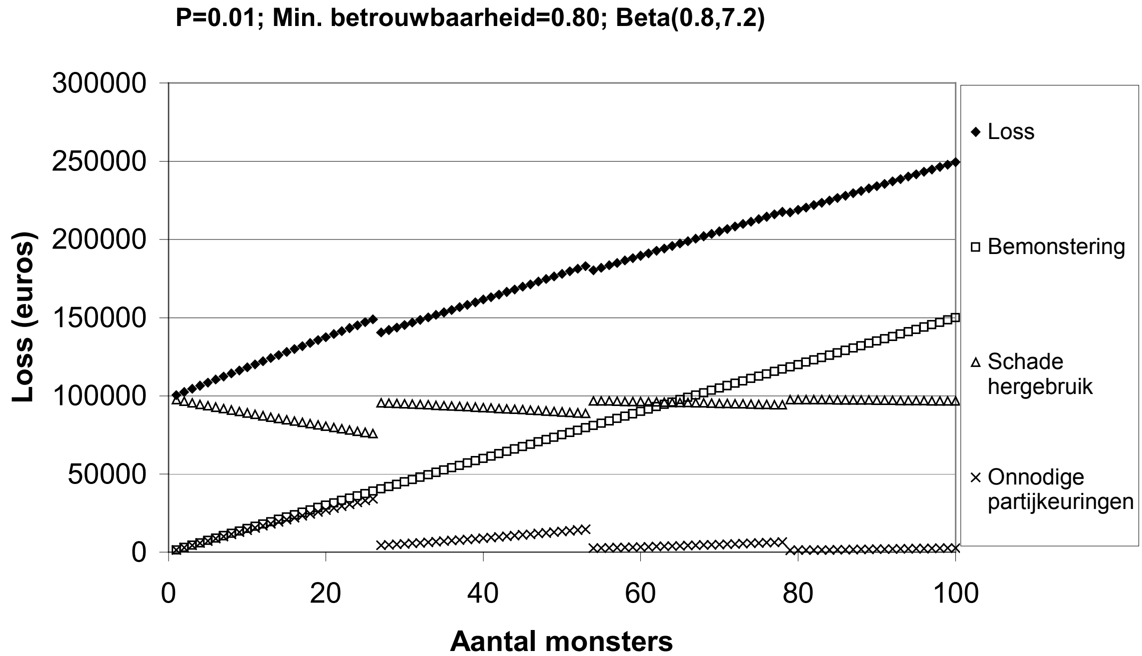
Figuur 8 Verlies (loss) als functie van aantal monsters voor klassieke toets van hypothese $\pi > 0.05$, *bij* $m = 100$, $c_{keuren} = 1\,500$ €, $c_{bemonsteren} = 1\,500$ €, $c_{saneren} = 100\,000$ €, $\pi = 0.001, 0.01$, *en minimale betrouwbaarheid van 0.80, 0.90.*





Figuur 9 Voor toelichting, zie volgende pagina.





Figuur 9 Verlies (loss) als functie van aantal monsters voor Bayesiaanse toets $\pi > 0.05$, bij $m = 100$, $c_{keuren} = 1\,500$ €, $c_{bemonsteren} = 1\,500$ €, $c_{saneren} = 100\,000$ €, $\pi = 0.001, 0.01$, minimale betrouwbaarheid van 0.80, en drie a priori verdelingen voor π

5 Discussie

5.1 Ruimtelijke afhankelijkheid

Bij beide toetsen (klassieke, binomiaal toets) wordt aangenomen dat de steekproef beschouwd kan worden als *n onafhankelijke* trekkingen, waarbij de kans op overschrijding (ernstig verontreinigde grond) voor alle trekkingen gelijk is. Bij de Bayesiaanse toets wordt immers gebruik gemaakt van de binomiale *likelihood*. Deze aanname is juist voor enkelvoudig aselecte steekproeven (volledig aselecte steekproeven). In veel gevallen zijn de gegevens waarmee de oppervlaktefractie ernstig verontreinigd wordt geschat niet verzameld door middel van kanssteekproeven (een kanssteekproef is een aselecte steekproef die voldoet aan bepaalde eisen), maar worden bestaande gegevens gebruikt, verzameld op gericht geselecteerde lokaties, bijv. de gegevens verzameld ten behoeve van aanvragen van bouwvergunningen. Omdat de steekproeflocaties niet door middel van loting zijn geselecteerd, kunnen we niet langer gebruik maken van de klassieke steekproeftheorie om onze onzekerheid over de oppervlaktefractie ernstig verontreinigd te kwantificeren, maar dienen we gebruik te maken van de geostatistiek. Alleen wanneer we aannemen dat de ruimtelijke variatie van de indicator i (i is 1 als locatie ernstig verontreinigd is, anders 0) kan worden beschreven met een model met een constante verwachtingswaarde en een covariantie gelijk aan 0 (ruimtelijke onafhankelijkheid; puur nugget variogram), is het aantal ernstig verontreinigde monsters in de steekproef binomiaal verdeeld. Als dit niet aannemelijk is, bijv. wanneer de lokaties ruimtelijk sterk geclusterd zijn, dan verdient het de voorkeur het gebiedsgemiddelde van de indicator te voorspellen met een geostatistische methode, bijv. block-indicator kriging (Goovaerts, 1997) of *model-based kriging* (Diggle en Tawn, 1998).

5.2 Parametrisch toetsen

Zowel de klassieke toets als de hierboven beschreven Bayesiaanse toets maken gebruik van aantallen monsters (totaal aantal monsters, aantal ernstig verontreinigde monsters). In beide toetsen wordt geen gebruik gemaakt van de metingen zelf, de concentraties van de onderzochte stoffen. Een monster waarin de concentratie van 1 stof zich net boven de interventiewaarde bevindt wordt op dezelfde manier behandeld als een monster waarin meerdere stoffen de interventiewaarde vele malen overschrijden. Een alternatief is om een aanname te doen over het type verdeling, en vervolgens de parameters van deze verdeling te schatten, dan wel hypotheses over deze parameters te toetsen. Voor methoden en toepassingen in milieuonderzoek verwijzen we naar Smith et al. (2003). In de bedrijfskundige statistiek (kwaliteitskunde) worden steekproeven ten behoeve van parametrisch toetsen van hypothesen 'steekproefsystemen op variabelen' genoemd (Dewaide, 1988). In geval van toetsen op oppervlaktefractie ernstig verontreinigde grond zijn de mogelijkheden van deze aanpak mogelijk beperkt, omdat we niet te maken hebben met 1 stof, maar met een groot aantal stoffen. We moeten dus een aanname doen over de multivariate verdeling, en hebben te maken met een zeer groot aantal parameters (o.a. correlaties tussen stoffen). Wanneer we ook nog rekening willen houden met ruimtelijke afhankelijkheid (zie par. 5.1) is er al gauw sprake van over-parameterisatie (veel parameters, weinig monsters).

6 Conclusies

Bayesiaans toetsen op oppervlaktefractie ernstig verontreinigde grond is een interessant alternatief voor klassiek toetsen. De betrouwbaarheid en het onderscheidend vermogen van de Bayesiaanse toets, bij een gegeven aantal monsters en aantal overschrijdingen, is groter dan van de klassieke toets. De winst hangt o.a. af van de a *priori* verdeling van de oppervlaktefractie. Des te kleiner de *a priori* verwachting van de oppervlaktefractie, des te groter de winst. Bij een verwachte *a priori* oppervlaktefractie π_{prior} van 0.1 (en variantie van 0.01) is de betrouwbaarheid van de conclusie $\pi > 0.05$ bij 19 monsters met 0 overschrijdingen > 0.80, voor $\pi_{\text{prior}} = 0.05$ (variantie 0.005) geldt dit voor 6 monsters met 0 overschrijdingen, voor $\pi_{\text{prior}} = 0.01$ (variantie 0.001) voor 1 monster dat niet ernstig verontreinigd is. Vergelijk dit met 32 mosnters met 0 overschrijdingen voor klassiek toetsen.

De gevoeligheid van de *confidence of compliance* en het onderscheidend vermogen van de Bayesiaanse toets voor de *a priori* verdeling maakt deze toets mogelijk minder geschikt als bewijsmiddel in juridische procedures, omdat de keuze van deze *a priori* verdeling in meer of mindere mate subjectief is. Daar staat tegenover dat de resultaten van klassiek toetsen overeen komen met die van Bayesiaans toetsen gebruikmakend van de Beta(1,0)-verdeling, d.w.z *a priori* wordt verondersteld dat 100% van de oppervlakte ernstig verontreinigd is. Klassiek toetsen geeft daardoor wel een erg pessimistisch beeld van de *confidence of compliance*.

Het verlies (*loss*) van beslissen over vrij grondverzet (hergebruik van partijen grond als bodem zonder partijkeuringen) hangt af van een groot aantal factoren. Een kleine oppervlaktefractie ernstig verontreinigd, een groot verwacht aantal partijen grond, en hoge kosten van saneren van hergebruikte, ernstig verontreinigde grond (relatief ten opzichte van partijkeuringen) werken in het voordeel van statistisch toetsen. Ook de onbetrouwbaarheidsdrempel is hierbij van invloed. In geval van 100 partijen, $c_{saneren}$ = 100 000 €, c_{keuren} = 1 500 €, $c_{bemonsteren}$ = 1 500 €, en oppervlaktefractie ernstig verontreinigde grond $\pi = 0.001$, is beslissen op basis van klassiek statistisch toetsen naar verwachting goedkoper dan alle partijen keuren wanneer voldoende, maar niet teveel monsters worden genomen. Het optimale aantal monsters hangt af van de onbetrouwbaarheidsdrempel, en is 32 voor $\alpha = 0.2$, en 45 voor $\alpha = 0.1$. Voor $\pi = 0.01$ is alle partijen keuren naar verwachting goedkoper. Voor Bayesiaans toetsen hangt de *loss* ook af van de *a priori* verdeling. Des te kleiner de verwachting van de *a priori* verdeling van π, des te groter de *confidence of compliance* gegeven het totaal aantal monsters en het aantal overschrijdingen, des te groter de kans op vrijstelling, des te groter de verwachte schade tengevolge van hergebruik van ernstig verontreinigde grond, en des te lager de onnodig gemaakte kosten van partijkeuring. Algemene conclusies zijn moeilijk te trekken, en daarom zijn twee excel-programma's bijgevoegd waarmee door de gebruikre zelf bepaald kan worden of in zijn of haar situatie beslissen op basis van statistisch toetsen naar verwachting goedkoper is dan alle partijen keuren, of niet.

Literatuur

Congdon, P., 2001. Bayesian statistical modelling. Wiley, Chichester.

Dewaide, H.T., 1988. Steekproeven en steekproefsystemen. In: B. Veen (red.). Bedrijfskundige statistiek . Kwaliteitskunde, III. Kluwer, Deventer.

Diggle, P.J. en J.A. Tawn, 1998. Model-based geostatistics. Appl. Statist. 47, p.299-350.

EPA, 2000. Guidance for the Data Quality Objectives Process (EPA QA/G-4). www.epa.gov/quality

Goovarts, P. 1997. Geostatistics for natural resources evaluation. Oxford Univ. Press, New York.

Ministerie VROM, Ministerie LNV, Interprovinciaal Overleg, Vereniging voor Nederlandse Gemeenten, 1999. Grond grondig bekeken. verantwoord omgaan met schone en verontreinigde grond.

Ministerie VROM, Ministerie LNV, Interprovinciaal Overleg, Vereniging voor Nederlandse Gemeenten, Organisatie van Nederlandse Raadgevende Ingenieur-bureaus, 1999. Interim-richtlijn. Opstellen en toepassen bodemkwaliteitskaarten in het kader van de vrijstellingsregeling grondverzet. Bijlage 1 van de nota 'Grond grondig bekeken'.

Smith, E.P., A. Zahran, M. Mahmoud and K. Ye, 2003. Evaluation of water quality using acceptance sampling by variables. Environmetrics 14: 373-386.

McBride, G.B. en J.C. Ellis, 2001. Confidence of compliance: a Bayesian approach for percentile standards. Wat. Res. Vol. 35(5), p. 1117-1124.