4.2 Landelijke en provinciale bodemkwaliteitsmeetnetten: vlees noch vis?

D.J. Brus,
Alterra, Wageningen UR

In een recente evaluatie van de landelijke en provinciale bodemkwaliteit- en grondwatermeetnetten is geconcludeerd dat het fundament van de landelijke en provinciale bodemkwaliteitsmeetnetten goed is. In deze evaluatie zijn echter de steekproeftechnische aspecten van de meetnetten buiten beschouwing gebleven. Vanuit steekproeftechnisch oogpunt kunnen vraagtekens bij deze conclusie worden geplaatst. De huidige meetnetten zijn niet zo geschikt voor het doen van statistische uitspraken over (veranderingen in) de kwaliteit van de Nederlandse bodem. Kanssteekproeven zijn hiervoor veel geschikter, zoals onder andere blijkt uit het milieumonitoring-programma in de Verenigde Staten van Amerika.

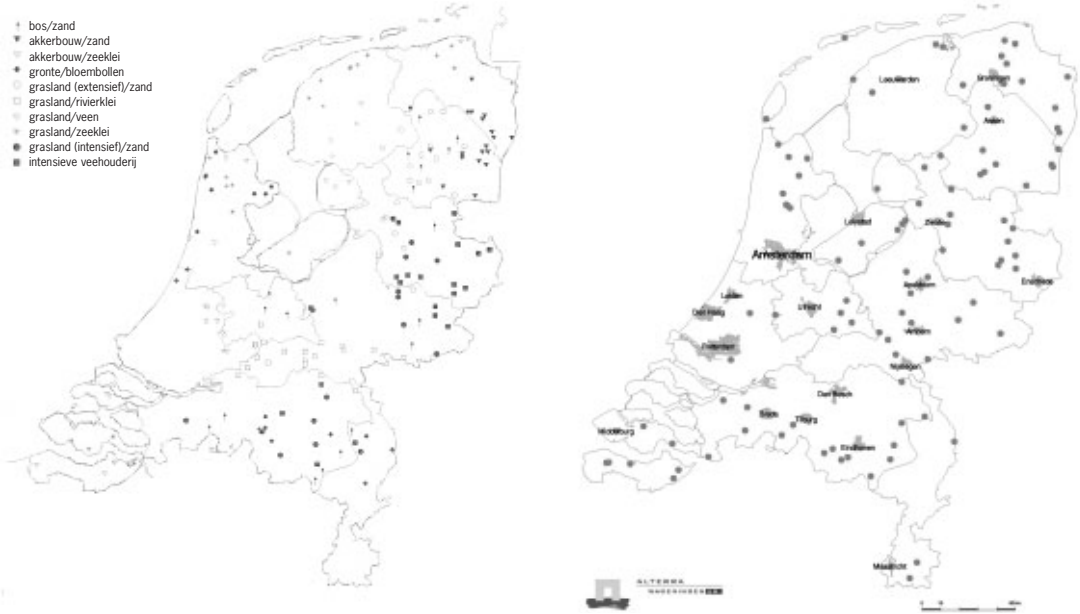
TCB-advies Monitoren

In 2007 adviseerde de Technische Commissie Bodembescherming (TCB) de minister van VROM over monitoren[1]. Dit advies, onder andere gebaseerd op het TNO-rapport 'Slim Monitoren' [2], pleit voor harmonisatie van de bodem- en grondwatermeetnetten. Bodem en grondwater worden niet alleen gemonitord in een landelijk meetnet maar ook in provinciale meetnetten. Bij de inrichting van het meetnet maakten de betreffende overheden hun eigen keuzes, waardoor bodem en grondwater op verschillende manieren worden gemonitord. Zo zijn er verschillen in medium waarin gemeten wordt (bodemvocht of vaste fase van bodem), ontsluitings- en bepalingsmethode, bemonsteringsdiepte, vorm en grootte van de monitoringplots, tijdstip van bemonstering etc. Hierdoor zijn de monitoringgegevens niet zonder meer vergelijkbaar, waardoor bijvoorbeeld de gegevens van de provinciale meetnetten niet zonder meer gebruikt kunnen worden om uitspraken te doen voor heel Nederland.

Zo worden in het Landelijk Meetnet Bodemkwaliteit (LMB) en in het bodemkwaliteitsmeetnet van de provincie Noord-Brabant hele bedrijven bemonsterd (een plot bestaat dus uit alle percelen van een bedrijf), terwijl in de bodemkwaliteitsmeetnetten van Drenthe en Groningen individuele percelen van ongeveer 1 ha worden bemonsterd. Wanneer we de gegevens van deze meetnetten willen gebruiken om bijvoorbeeld het areaal fosfaatverzadigde landbouwgronden te schatten, dan is het duidelijk dat we deze gegevens niet zonder meer op een hoop kunnen gooien. Het moge ook duidelijk zijn dat deze situatie 'suboptimaal' is, en daarom heeft de TCB de minister geadviseerd om de meetnetten te harmoniseren. Om dit te bereiken stelt de TCB dat er behoefte is aan meer 'centrale regie'. Centrale regisseurs zouden onder andere moeten zorgen voor een betere coördinatie van de monitoringactiviteiten, ze zouden de kwaliteit van de monitoring moeten borgen, en de meetnetten moeten optimaliseren: adviezen die ik van harte ondersteun.

Hoe goed is het fundament van de meetnetten?

Opvallend is dat in het TNO-rapport 'Slim Monitoren' steekproeftechnische aspecten van de meetnetten buiten beschouwing blijven. Dat is jammer, want de basis van een monitoringmeetnet wordt gevormd door een steekproef in ruimte en tijd. Het had dan ook voor de hand gelegen dat het rapport ruime aandacht zou besteden aan de steekproefopzet, dat wil zeggen de methode volgens welke de meetnetlocaties



Figuur 1 Het Landelijk Meetnet Bodemkwaliteit (links) is een gerichte steekproef van bedrijven. Dit meetnet is minder geschikt voor het doen van valide uitspraken over de bodemkwaliteit in Nederland. Percentielwaarden, waaronder de mediaan en het 90e percentiel, van contaminanten in de Nederlandse bodem zijn geschat met een kanssteekproef (rechts). De plots zijn geloot volgens een gestratificeerde enkelvoudig aselecte steekproefopzet (Bron: Bronswijk et al.[4])

(plots) en de tijdstippen zijn (worden) geselecteerd. Ik ervaar dit als een gemis, omdat hierdoor geen goed beeld is verkregen van de mogelijkheden en onmogelijkheden van de huidige meetnetten. Ik wil nu kort ingaan op de streekproefopzet van de huidige meetnetten bodem- en grondwaterkwaliteit (zie figuur 1).

We hebben inmiddels al vastgesteld dat de meetnetten in een groot aantal aspecten verschillen. In één opzicht zijn ze echter allemaal gelijk: in alle meetnetten zijn de meetnetlocaties (plots) gericht geselecteerd. Bij de selectie van de meetnetlocaties is gebruik gemaakt van kennis van experts op het gebied van bodemgeografie en bodemkundige en hydrologische processen. Zo is het te monitoren gebied onderverdeeld in zogeheten 'homogene gebieden', gebieden met een bepaald bodemtype (grondsoort) en landgebruik. De (ruimtelijke) gemiddeldes van een of meerdere bodem- en grondwaterkenmerken verschillen mogelijk tussen deze deelgebieden. Ook kan de verandering in de tijd (temporele trend) voor een aantal bodem- en grondwaterkenmerken verschillen tussen de deelgebieden. Beide verschillen zijn goede redenen om bij de selectie van de meetnetlocaties rekening te houden met de bodem- en hydrologische gesteldheid. Binnen de 'homogene deelgebieden' zijn de locaties gericht geselecteerd, of beter gezegd: niet geselecteerd door middel van loting. Hierna zal blijken dat dit consequenties voor de bruikbaarheid van de monitoringresultaten.

Ik vermijd liever het begrip aselecte steekproef, waarmee immers in de praktijk vaak, wat de Belgen zo treffend noemen, een lukrake steekproef wordt bedoeld

Regelmatig kom ik in rapportages tegen dat de steekproef als aselect kan worden beschouwd, of dat de steekproeflocaties 'vrijwel aselect' zijn gekozen [3]. Blijkbaar is er verwarring rond het begrip aselect. Belangrijker dan de vraag of de steekproeflocaties wel of niet of min of meer 'aselect' gekozen zijn, is de vraag of de kansen bekend zijn waarmee de locaties zijn geselecteerd. Dit is voor geen van de bodem- en grondwaterkwaliteitsmeetnetten het geval. Bij selectie van de locaties door middel van loting volgens een goed gedefinieerde steekproefopzet, bijvoorbeeld gestratificeerd, enkelvoudig aselect, inclusief een specificatie van het aantal steekproefpunten per stratum, zijn deze kansen wel bekend. Een steekproef waarbij de selectiekansen bekend zijn, en bovendien alle locaties die tot het te monitoren gebied behoren, een selectiekans groter dan 0 hebben, wordt een kanssteekproef genoemd. Ik vermijd liever het begrip aselecte steekproef, waarmee immers in de praktijk vaak, wat de Belgen zo treffend noemen, een lukrake steekproef wordt bedoeld. De eis dat alle locaties een selectiekans groter dan 0 hebben is nodig om een uitspraak te kunnen doen over het hele onderzoeksgebied. Bij de selectie van bedrijven van het LMB zijn bedrijven < 10 ha buiten beschouwing gelaten. Alleen om deze reden al is het LMB dan ook geen kanssteekproef. Het is dan ook niet mogelijk om met behulp van klassieke steekproeftheorie op basis van de LMB data statistische uitspraken te doen over alle bedrijven in Nederland.

Waarom is zo belangrijk dat we deze 'selectiekansen' kennen? De reden hiervoor is dat we de gegevens graag willen gebruiken om statistische uitspraken te doen over het hele gebied waaruit de locaties zijn geselecteerd, en niet alleen over de meetnetlocaties in de steekproef. Als bijvoorbeeld uit de eerste meetronde van het LMB blijkt dat in 29% van de bedrijven de streefwaarde van Cd in grondwater wordt overschreden [4], is dit leuk om te weten, maar hoe zit het nu met dit percentage voor heel Nederland? Met andere woorden, we zijn niet zozeer geïnteresseerd in het percentage van de bedrijven in de steekproef waarin de streefwaarde wordt overschreden, maar in het percentage over alle bedrijven in Nederland, of misschien nog meer in het oppervlaktepercentage van Nederland waarbinnen de streefwaarde wordt overschreden. We zijn, afgezien van meetfouten, zeker van het percentage bedrijven in de steekproef waarin de streefwaarde wordt overschreden.

Over het percentage bedrijven *in Nederland* waarin de streefwaarde wordt overschreden zijn we echter allerminst zeker. Als we andere bedrijven geselecteerd zouden hebben, dan hadden we immers een andere schatting van het percentage bedrijven in Nederland verkregen. Door het percentage bedrijven in de steekproef te gebruiken als schatting voor het percentage over alle bedrijven in Nederland, introduceren we een extra fout, de zogeheten steekproeffout. Deze fout kennen we niet, maar we willen wel graag iets kunnen zeggen over de kansverdeling van deze fout. Dat is goed mogelijk als de steekproef een kanssteekproef is. Daarentegen is dit vaak erg problematisch bij niet-kanssteekproeven.

Kwantificering van de nauwkeurigheid van het resultaat is van groot belang bij monitoring. Stel dat uit de volgende meetronde van het LMB

blijkt dat het percentage bedrijven waarbinnen de streefwaarde voor Cd in grondwater wordt overschreden is afgenomen van 29% tot 25%. De vraag zou dan gesteld kunnen worden of deze afname van het percentage bedrijven statistisch gezien significant is. Met andere woorden, als we andere bedrijven hadden geselecteerd, zouden we dan ook een afname hebben gezien, of zou op deze bedrijven het Cd-gehalte veel sterker afgenomen zijn, gelijk gebleven of zelfs toegenomen zijn? Wanneer de nauwkeurigheid van het geschatte verschil in bedrijfspercentages netjes wordt gekwantificeerd, kan statistisch getoetst worden of het verschil significant afwijkt van 0. Als het verschil niet significant blijkt te zijn, dan kunnen ook geen conclusies getrokken worden over bijv. het effect van genomen maatregelen. Ook voor toetsing aan normen (*compliance-monitoring*) is een objectieve kwantificering van de nauwkeurigheid van het resultaat van groot belang, om discussies over de geldigheid van de conclusie (gemiddelde ligt wel/niet boven de norm) te vermijden.

Monitoring op gericht-geselecteerde locaties is geen typisch Nederlandse gewoonte. Uit een analyse van Parr *et al.*[5] blijkt dat vrijwel geen enkel Europese meetnet is gebaseerd op een kanssteekproef: *'Instead the ensemble of monitoring sites in Europe seems to form a series of case studies selected by judgment sampling, the only exception being the Level 1 network for forest condition monitoring'. Verderop concluderen ze dan ook dat 'When the sample is not probabilistic, statistically- based inference is not possible. Model-based inferences can be investigated, but they need a number of assumptions.'* Precies om deze reden heeft indertijd de begeleidingscommissie van het project Achtergrondwaarden 2000 (AW2000) besloten om voor het schatten van percentielwaarden van contaminanten in de bodem geen gebruik te maken van het Landelijk Meetnet Bodemkwaliteit, maar een nieuwe kanssteekproef te ontwerpen [6]. Op basis van het bovenstaande zou ik niet zonder meer de conclusie van de deelnemers van de monitoringsconferentie Kwali-Tijd in november 2006 willen ondersteunen dat het fundament van de meetnetten goed is.

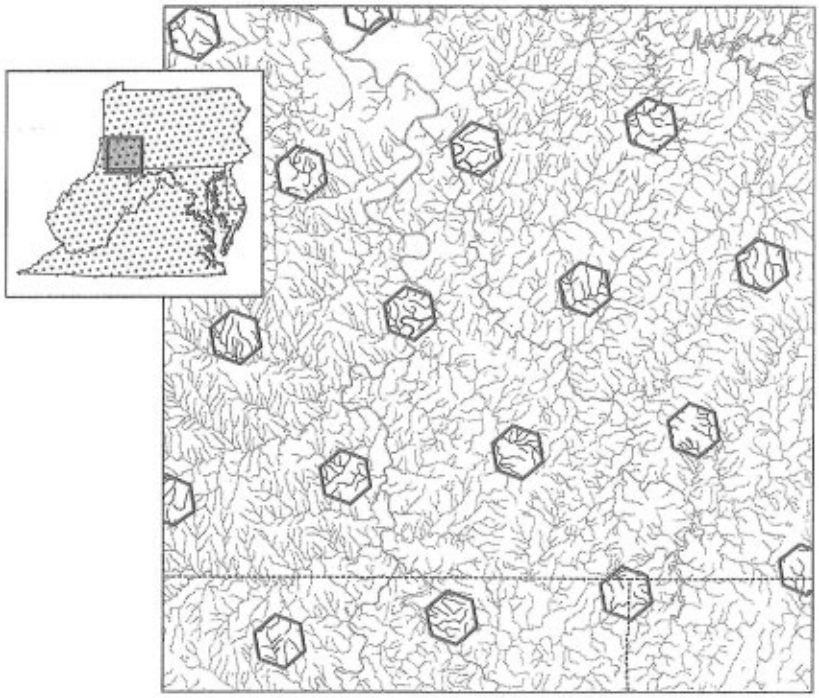
Environmental Monitoring and Assessment Programme (EMAP)

De TCB heeft de minister van VROM geadviseerd de gewenste harmonisatie te realiseren door de monitoringactiviteiten in Nederland meer centraal te regisseren. Dit advies ondersteun ik van harte, onder meer gezien de zegeningen in de USA van een deskundige centrale regie. In de USA is binnen het Environmental Protection Agency (EPA) de National Health and Environmental Effects Research Laboratory (NHEERL) 'belast' met de nationale monitoring van de natuurlijke hulpbronnen. Het NHEERL heeft voor het monitoren van de aquatische ecosystemen een raamwerk voor monitoring ontworpen, het Environmental Monitoring and Assessment Programme, kortweg EMAP [7,8]. Het doel van EMAP is het beantwoorden van vragen over de status, veranderingen en trends in indicatoren van de ecologische conditie van de nationale natuurlijke hulpbronnen. De aquatische hulpbronnen die gemonitord worden zijn estuariën, rivieren (streams) (met een totale lengte van ongeveer 5.6 miljoen km) en meren (zo'n 44,000).

Uit een analyse van Parr et al. blijkt dat vrijwel geen enkel Europese meetnet is gebaseerd op een kanssteekproef

Figuur 2 (Figure 1 uit Herlihy et al.12) EMAP-kanssteekproef van rivieren en beken in de vijf mid-Atlantische staten van de USA. De figuur toont alleen de in de eerste trap gelote hexagonen van 40 km2. Deze hexagonen vormen een driehoekig grid dat de hele USA bedekt. In de tweede trap worden locaties geloot in de rivieren en beken die binnen de 40 km2 vallen.

Op dit moment hebben meer dan dertig staten een meetnet volgens dit raamwerk ingericht. Een belangrijk voordeel van EMAP is dat de gegevens die op staat-niveau worden verzameld in de meetnetten gemakkelijk, en zonder gewaagde aannames, geaggregeerd kunnen worden naar grotere gebieden, zo ook naar het niveau van de hele USA. Lokale aanpassingen van het meetnet om tegemoet te komen aan lokale eisen, bijvoorbeeld een grotere of kleinere dichtheid van steekproeflocaties, zijn door de gekozen steekproefopzet heel goed mogelijk. Uitgangspunt voor EMAP is dat de resultaten een zuiver (*unbiased*) beeld moeten geven van de conditie van de natuurlijke hulpbronnen, met een bekende betrouwbaarheid (EPA, 2002, blz. II-1). Om deze redenen heeft het EPA gekozen voor een zogeheten ontwerpgebaseerde (*design-based*) benadering, wat inhoudt dat de locaties worden geselecteerd door middel van een kanssteekproef.



Figuur 2 toont hoe de rivieren in de vijf mid-Atlantische staten zijn geloot. Allereerst zijn hexagonen van 40 km^2 geselecteerd. Merk op dat de hexagonen een driehoeksgrid vormen. Dit driehoeksgrid bedekt de hele USA. Hierdoor wordt gezorgd voor een goede ruimtelijke spreiding van de steekproefpunten over het hele gebied. Vervolgens zijn locaties geloot in de rivieren en beken die binnen de hexagonen van 40 km^2 vallen. Dit is gedaan op een ingenieuze wijze die er voor zorgt dat het aantal locaties goed verdeeld worden over de verschillende netwerken van rivieren en beken en over de Strahler-ordes. Ook de meren worden met dezelfde steekproefopzet gemonitord[9].

Hoe nu verder?

EMAP is gebaseerd op een gestapelde monitoringbenadering (figuur 3). In deze benadering worden drie verdiepingen van monitoringactiviteiten onderscheiden. De laagste verdieping wordt gevormd door remote-sensingonderzoek. De ruimtelijke dichtheid van de waarnemingen (aantal metingen per oppervlakte-eenheid) in dit onderzoek is zeer hoog. De ruimtelijke resolutie van het resultaat, de kaarten, is dan ook veelal zeer groot, d.w.z. de kaart bestaat uit kleine pixels met (gemiddelde) waardes van de metingen. De correlatie tussen wat met de remote-sensingtechnieken gemeten wordt en wat we eigenlijk willen

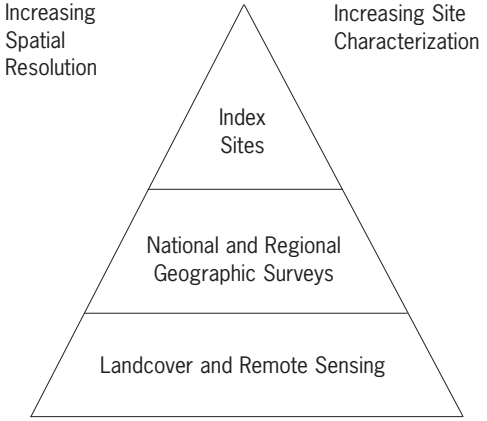
weten laat soms te wensen over, maar de kracht van deze monitoringactiviteiten zit in het aantal, de grote dichtheid.

De middelste verdieping wordt gevormd door nationale en regionale steekproefsgewijze inventarisaties (*surveys*). Op de steekproeflocaties worden metingen verricht die nauw gerelateerd zijn aan het doelkenmerk. Omdat de steekproeflocaties bezocht moeten worden voor het verrichten van de metingen zijn de kosten per meting in vergelijking met remote-sensingmetingen relatief hoog. Hierdoor is de ruimtelijke dichtheid van de metingen noodzakelijkerwijs een stuk kleiner, en levert het minder gedetailleerde kaarten op. De uitdaging is om de remote-sensingmetingen te combineren met de steekproefresulaten (*field-truth*), teneinde betere kaarten te maken dan op basis van de steekproefresultaten alleen, dat wil zeggen kaarten met meer detail of een groter nauwkeurigheid.

De bovenste verdieping wordt gevormd door zogeheten indexlocaties. Dit zijn gericht geselecteerde locaties waar een zeer groot aantal kenmerken wordt gemeten, in meerdere delen van het ecosysteem: vegetatie, bodem, grondwater, lucht. Het rijk geschakeerde palet aan metingen maakt het mogelijk om dynamische procesmodellen te kalibreren. Onderzoek op deze locaties levert hierdoor inzicht in de processen achter de metingen, waarmee vragen kunnen worden beantwoord over het waarom van de veranderingen in het ecosysteem die zijn waargenomen op de steekproeflocaties van de onderliggende verdieping. Door de hoge meetkosten is het aantal indexlocaties noodzakelijkerwijs zeer beperkt.

Door het geringe aantal indexlocaties bestaat het risico dat de resultaten, waargenomen op de indexlocaties, geen goed beeld geven van wat er zich in het hele land afspeelt. De vraag is dus hoe de resultaten van de index-locaties 'opgeschaald' kunnen worden naar het hele land. De *surveys* van de onderliggende laag kunnen hierin een belangrijke rol spelen, als er aan een aantal voorwaarden wordt voldaan. In de eerste plaats moet een relatie gelegd kunnen worden tussen de uitkomsten van de procesmodellen en de waarnemingen op de steekproeflocaties. Idealiter moet met de kenmerken, gemeten op de steekproeflocaties, een metamodel (regressiemodel) voor de model-uitvoer gebouwd kunnen worden.
In de tweede plaats moeten de steekproeflocaties 'representatief' zijn voor het hele land.

Welke rol kunnen de landelijke en provinciale meetnetten in bovenstaande aanpak nu vervullen? De bedrijven van het LMB zijn geselecteerd uit de ca. 1500 bedrijven die deelnemen aan het



Figuur 3 Gestapelde monitoring benadering van EMAP (Figure 7 uit EPA [8])

Bedrijven-Informatie-Net. Dit is gedaan omdat voor deze bedrijven balansen van mineralen en zware metalen aanwezig zijn, wat balansstudies en het berekenen van fluxen mogelijk maakt. De selectie van de LMB-plots lijkt gestuurd te zijn door twee doelen, nl. het verkrijgen van inzicht in de processen en het verkrijgen van een 'representatief' beeld van wat er zich in heel Nederland afspeelt. De vraag die zich opdringt is of door dit tweeledige doel het LMB niet vlees noch vis is. Enerzijds is door het vrij grote aantal locaties het aantal kenmerken dat gemeten wordt beperkt, wat de mogelijkheden van procesonderzoek beperkt. Anderzijds is door aansluiting bij het BIN het LMB minder geschikt voor de bovengenoemde 'opschalingsfunctie' dan we wel zouden willen. Ditzelfde geldt voor de provinciale meetnetten. Voor de opschalingsfunctie verdient, zoals hierboven betoogd, een nette kanssteekproef de voorkeur.

Een nadeel is natuurlijk dat de resultaten van een nieuwe meetronde niet zonder meer te vergelijken zijn met die van al uitgevoerde meetrondes

Dit alles overwegende komt bij mij de volgende vraag op: wat zijn de voor- en nadelen van het vervangen van (een deel van) het bestaande LMB-meetnet door een nieuw meetnet van 'slim' gelote locaties? Een nadeel is natuurlijk dat de resultaten van een nieuwe meetronde niet zonder meer te vergelijken zijn met die van al uitgevoerde meetrondes. Dit nadeel is mijns inziens wel beperkt, aangezien er slechts één meetronde is geweest. Van groot belang is verder de vraag hoe de data verzameld op gerichte steekproeven en andere niet-kanssteekproeven het beste statistisch verwerkt kunnen worden. Wereldwijd zijn zeer veel bestanden aanwezig met data van niet-kanssteekproeven. Deze databestanden bevatten ongetwijfeld veel nuttige informatie. Het simpelweg negeren van deze databestanden is kapitaalvernietiging, maar de vraag is wel hoe we de gewenste informatie uit deze databestanden extraheren. In een eerste studie stellen Brus en De Gruijter [10] voor de niet-kanssteekproef aan te vullen met een kanssteekproef, en de twee steekproeven te combineren middels zogeheten regressieschatters. Ook een zogeheten Bayesiaanse aanpak biedt goede mogelijkheden, zoals o.a. blijkt uit de recente studie van Graves *et al.*[11].

We zullen, misschien wel knarsetandend, moeten toegeven dat we van EMAP veel kunnen leren in Nederland en Europa. Het is naar mijn mening niet te laat om lessen te trekken. Daarom pleit ik ervoor dat er snel een stevig statistisch fundament wordt gelegd onder de bodem- en grondwaterkwaliteitsmeetnetten, waarbij locaties voortaan worden geselecteerd met behulp van kanssteekproeven.