

## TOEGEPASTE STATISTIEK IN HET WATERBEHEER (1)

# Statistiek voor waterkwaliteitsgegevens

“Bah! een rubriek over statistiek”, zult u denken, “wat moet ik daar nu mee?”. Tja, statistiek kan taai zijn, maar we kunnen niet zonder. Bij meetnetontwerp, interpretatie en de presentatie van gegevens is statistiek onmisbaar. Soms moeten arbitraire keuzes worden gemaakt. Welke betrouwbaarheid is gewenst? Hoe om te gaan met waarden onder de detectiegrens bij berekeningen? Binnen het Rijksinstituut voor Integraal Waterbeheer en Afvalwaterbehandeling (RIZA) en het Rijksinstituut voor Kust en Zee (RIKZ) liep een project waarbij afstemming is gezocht voor dit soort problemen. In deze rubriek presenteren Jaap van Steenwijk (RIZA) en Richard Duin (RIKZ) de resultaten en de aanbevelingen van dit onderzoek naar toegepaste statistiek in het waterbeheer. Maar ze slingeren af en toe ook een discussie aan.

Veel mensen hebben moeite met statistiek, zoniet een regelrechte aversie. Mogelijk heeft dit diepe wortels, want van oudsher worden bevolkingen geteld voor het aantal te leveren soldaten en de te heffen belastingen. De naam statistiek wordt in de moderne tijd dan ook voor het eerst gebruikt door de Duitser Gottfried Achenwald (in 1749), die de staat beschreef zoals het CBS dat thans nog doet.

In het waterkwaliteitsbeheer is naast beschrijvende statistiek het toetsen en analyseren van meetgegevens even belangrijk. Veel statistische methoden veronderstellen echter ‘een model’. Als we ons daar niet bewust van zijn kunnen we de mist ingaan. Zo is de aanname dat gegevens normaal verdeeld zijn bij meetgegevens over waterkwaliteit vaak onjuist, omdat ze aan de ene kant door de detectiegrens worden ‘afgekapt’ en aan de andere kant uitschieters voorkomen. Toch mag vaak bij gegevens die samengevat worden tot bijvoorbeeld een gemiddelde, worden uitgegaan van een normale verdeling mits de steekproef, het aantal monsters, groot genoeg is.

Een volgend punt is dat veel methoden zijn ontworpen voor het toetsen van experimenten. Milieugegevens die een beschrijving geven van processen in de natuur hebben weinig gemeen met een gecontroleerd experiment! Kan ik dan al die toetsen nog wel gebruiken?

Soms kom je een publicatie tegen die weet te melden dat uitkomst zus of zo ‘statistisch significant’ is, echter zonder te vermelden wat men daar mee bedoeld. Dat kan niet en het is dan ook zeer dubieus of er iets significant in die publicatie staat. Want bij

het toetsen van veronderstellingen is het nog altijd zo dat je vooraf moet vastleggen (en dat is een arbitraire keuze) wat jij als significant beschouwd. Bij toetsen heb je te maken met onzekerheid die we in de statistiek vastleggen met  $\alpha$ , waarvan de gebruiker de grootte bepaald. Indien er veel op het spel staat en je wilt een kleine kans dat er iets mis gaat dan kun je een kans van 1 op honderd ( $\alpha = 0,01$ ) of zelfs  $\alpha$  is 0,001 kiezen! Belangrijk is dat je vooraf éénduidig je uitgangspunt (dat subjectief is), eventueel voorzien van een motivatie, vastlegt. Is statistiek dan geen exact vak? Ja, de procedure is wel degelijk wiskundig onderbouwd, maar het uitgangspunt van de onderzoeker of rapporteur is arbitrair.

De detectiegrenzen zijn al genoemd. Een probleem daarbij is hoe om te gaan met de

‘kleiner dan de detectiegrens’-waarden als je een gemiddelde (of een ander kengetal) berekent? Het blijkt dat mensen (soms zelfs binnen een instituut) op verschillende manieren omgaan met detectiegrenzen en uitschieters. Alleen al door de keuze van hoe men omgaat met de detectiegrens en hoe vaak zo’n ‘kleiner dan’ voorkomt kun je al aardig wat verschillende gemiddelde waarden krijgen.

In de komende maanden zullen we een aantal onderwerpen behandelen. Na een korte inleiding rond enkele begrippen in de statistiek in de eerstkomende aflevering, komen achtereenvolgens aan de orde:

- Hoe ontstaan detectiegrenzen, met een inleiding hoe een statistische toets werkt;
- Hoe ga je om met detectiegrenzen bij verwerking van meetgegevens?
- de betrouwbaarheid van de kengetallen,
- Hoeveel monsters zijn nodig om een trend te kunnen detecteren?
- berekening van het 90-percentiel als toetswaarde,
- variatiebronnen in gegevens,
- inleiding op trendanalyses.

Na deze eerste reeks korte artikelen over praktische en toegepaste statistiek volgt mogelijk, afhankelijk van reacties, na de zomer een nieuwe reeks met onderwerpen als het berekenen van een trend en vrachten. ◀

Voor meer informatie:

Jaap van Steenwijk,  
e-mail: [j.steenwijk@riza.rws.minvenw.nl](mailto:j.steenwijk@riza.rws.minvenw.nl)  
of Richard Duin,  
e-mail: [r.n.m.duin@rikz.rws.minvenw.nl](mailto:r.n.m.duin@rikz.rws.minvenw.nl)

Het milieuschip de IJsselmeer van de regionale directie IJsselmeergebied van Rijkswaterstaat.

