

Het artikel van Paul Goedhart is een reactie op de heranalyse van Hanekamp et al. van een aantal WUR-experimenten om de ammoniakemissie van verschillende bemestingstechnieken te schatten. De discussie over de wetenschappelijke onderbouwing van het ammoniakbeleid is eerder gevoerd in het vakblad *Soil Use and Management*. Het Rathenau Instituut spreekt van een hardnekkige controverse. Het conflict heeft ook een politieke kant, doordat sommige veehouders zich beperkt voelen in hun ondernemerschap door milieuregels van de overheid. Goedharts artikel gaat met name over het berekenen van de onzekerheidsmarges en de zijns inziens incorrecte statistische analyse van Hanekamp et al. Om te voorkomen dat het een preek voor eigen parochie wordt heeft de *STAtOR*-redactie Hanekamp om een reactie gevraagd, zodat de lezer zelf een mening kan vormen.

HOE EEN STATISTISCHE FOUT HET AMMONIAKDEBAT OP EEN DWAALSPoor ZET

PAUL GOEDHART

Ammoniak (NH_3) is slecht voor de natuur, dat is onbetwist. In Nederland wordt door de agrarische sector veel ammoniak uitgestoten vooral door het bemesten van landbouwgrond en door verliezen uit de stal. Sinds 1985 is er overheidsbeleid om deze uitstoot te reduceren. Zo is het verplicht om emissiearme bemestingstechnieken te gebruiken zoals bijvoorbeeld de zodenbemester in plaats van mest bovengronds uit te rijden. Alle overheidsmaatregelen tezamen hebben volgens het rekenmodel NEMA geleid tot een reductie van de ammoniakemissie met 64% (zie www.rivm.nl/ammoniak/emissies). Het meetmodel OPS van het Rijksinstituut voor Volksgezondheid en Milieu (RIVM), dat gebruik maakt van gemeten ammoniakconcentraties in de lucht, geeft echter aan dat de reductie kleiner is. Het verschil in de uitkomst van het rekenmodel NEMA en het meetmodel OPS wordt het ammoniakgat genoemd. Sommigen claimen op grond hiervan dat het overheidsbeleid niet werkt en dat het bijvoorbeeld onzin is om bovengronds uitrijden van mest te verbieden.

Het rapport *Ammoniak in Nederland* van chemicus Jaap Hanekamp (University College Roosevelt, Middel-

burg), wetenschapsjournalist Marcel Crok en de Amerikaanse wiskundige Matt Briggs wordt op 24 januari 2017 aangeboden aan de Tweede Kamerleden Elbert Dijkgraaf (SGP) en Jaco Geurts (CDA). In dit rapport (Hanekamp et al., 2017a) constateren zij: 'een opeenstapeling van rekenkundige, modelmatige, en argumentatieve tekortkomingen. Men kan om die redenen niet concluderen dat het ammoniakbeleid wetenschappelijk goed onderbouwd zou zijn, zoals de overheid beweert. Integendeel.' Naast kritiek op de door het RIVM uitgevoerde trendanalyse op gemeten ammoniakconcentraties in de lucht, zijn de auteurs zeer kritisch over het emissieonderzoek van Wageningen Universiteit en Research (WUR). Dit onderzoek richt zich mede op de vergelijking van verschillende bemestingstechnieken zoals bovengronds uitrijden en de zodenbemester. Uit het WUR-onderzoek blijkt dat emissiearme technieken de uitstoot van ammoniak flink verminderen. Hanekamp et al. heranalyseren acht WUR-experimenten om de ammoniakemissie van een specifieke techniek in te schatten. Op basis van hun heranalyse concluderen zij dat de onzekerheidsmarges van de geschatte emissies 'fors



Bovengronds uitrijden van mest

Zodenbemester aan het werk

bleken' en tevens dat 'het niet onwaarschijnlijk (is) dat de verschillende bemestingstechnieken dichter bij elkaar liggen wat betreft ammoniakemissies dan tot op heden is gerapporteerd en toegepast in het beleid'.

Samen met de kritiek op de trendanalyse van het RIVM plaatsen ze grote vraagtekens bij het door de overheid gevoerde ammoniakbeleid. Nu zijn onzekerheden de *core business* van elke statisticus, en daarom ben ik nagegaan of de kritiek van Hanekamp et al. op het WUR-onderzoek hout snijdt.

Ammoniakconcentratie en windprofiel

Een veel gebruikte experimentele methode om de ammoniakemissie van een bemestingsmethode te bepalen

is de zogenaamde *Integrated Horizontal Flux* methode. Aan het begin van een experiment wordt eenmalig mest uitgereden op een veelal cirkelvormige plot van ongeveer 40 m diameter. In het centrum van de plot wordt op diverse hoogtes zowel de ammoniakconcentratie als de windsnelheid gemeten (figuur 1). Deze metingen worden gebruikt om een ammoniakconcentratieprofiel $c(h)$ en een windprofiel $u(h)$ te schatten, beide als functie van de hoogte h . De ammoniakemissie wordt vervolgens gebaseerd op de hoeveelheid ammoniak die de wind wegblaast: $\int_{h_0}^{h_p} u(h) c(h) dh$. Het integratie-interval loopt van de hoogte h_0 waarop de windsnelheid nul is, tot de hoogte h_p waarop de ammoniakconcentratie gelijk is aan de achtergrondconcentratie (die ook wordt gemeten). Na een correctie voor de achtergrondconcentratie en een normalisatie



Figuur 1. Links een meetmast met windsnelheidsmeters op verschillende hoogtes, en rechts een zogenaamde Impinger oftewel een flesje met zuur om ammoniak af te vangen

wordt uiteindelijk een percentage emissie van de oorspronkelijke hoeveelheid ammonium in de mest verkregen. Deze procedure wordt gevolgd voor verschillende opeenvolgende perioden na toediening van de mest. De som van de berekende percentages over de opeenvolgende perioden is de zogenaamde emissiefactor. Deze geeft aan welk aandeel van de ammonium in de mest verloren gaat via vervluchtiging.

Het is gebruikelijk om voor het concentratie- en wind-profiel een lineaire relatie te nemen met als verklarende variabele de logaritme van de hoogte h : $c(h) = -A \ln(h) + B$ en $u(h) = D \ln(h) + E$. De parameters A , B , D en E worden uit de metingen geschat met lineaire regressie. Met deze profielen kan de integraal expliciet berekend worden:

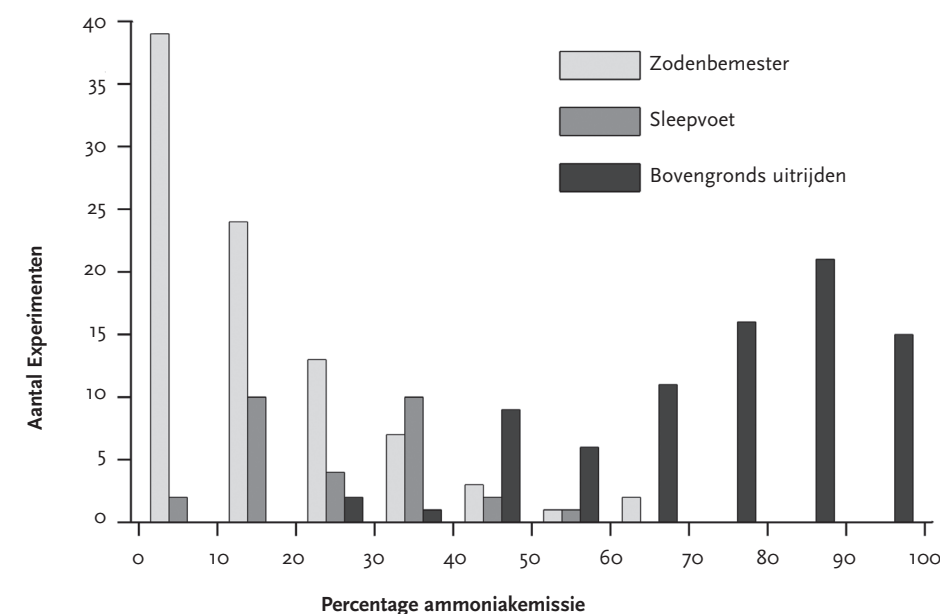
$$\int_{h_0}^{h_p} u(h) c(h) dh = [-AD(h \ln^2 h - 2h \ln h + 2h) + (BD - AE)(h \ln h - h) + BE h]_{h_0}^{h_p}$$

De integraal, en dus de emissie, is een functie $F(A, B, D, E)$ van de vier regressieparameters. Hanekamp et al. berekenen een 95%-betrouwbaarheidsinterval voor de uitkomst door het invullen van 95%-betrouwbaarheidsgrenzen voor de individuele parameters zelf in $F(A, B, D, E)$. Gebruikmakend van het superscript + voor een rechter betrouwbaarheidsgrens berekenen zij $F(A^+, B^+, D^+, E^+)$ voor de bovengrens in plaats van $F^+(A, B, D, E)$. Deze fout verergert zij door het sommeren van de zo verkregen betrouwbaarheidsgrenzen over de opeenvolgende meetperioden. Oftewel, met bijvoorbeeld drie meetperioden, berekenen zij $(F_1^+ + F_2^+ + F_3^+)$ in plaats van het correcte $(F_1 + F_2 + F_3)^+$.

Het zo berekende interval wordt door hen gepresenteerd als een 95%-interval. Een basale fout.

Het volgende voorbeeld illustreert dat het optellen van betrouwbaarheidsgrenzen een veel hogere overdekingskans geeft. Veronderstel dat we twee steekproefgemiddelden $\bar{x}_1$ en $\bar{x}_2$ hebben, beide gebaseerd op n waarnemingen, uit normale verdelingen met gemiddelden respectievelijk μ_1 en μ_2 en bekende variantie σ^2 . De éézijdige α -rechter betrouwbaarheidsgrens voor μ_1 is dan $\bar{x}_1 + z_\alpha \sigma / \sqrt{n}$, met $\Phi(z_\alpha) = \alpha$, en de som van de bovengrenzen is $\bar{x}_1 + \bar{x}_2 + 2 z_\alpha \sigma / \sqrt{n}$. Dit is echter géén betrouwbaarheidsgrens voor $\mu_1 + \mu_2$. Immers, uit $\bar{x}_1 + \bar{x}_2 \sim N(\mu_1 + \mu_2, 2 \sigma^2/n)$ volgt onmiddellijk dat de correcte, en lagere, bovengrens gelijk is aan $\bar{x}_1 + \bar{x}_2 + \sqrt{2} z_\alpha \sigma / \sqrt{n}$, nu met een factor $\sqrt{2}$ in plaats van 2. De consequentie is dat de intervallen van Hanekamp et al. een veel grotere overdekingskans hebben dan de door hen geclaimde 95%. De intervallen zijn veel te breed en dit verklaart de door hen geclaimde forse onzekerheden. In bovenstaand voorbeeld heeft het foutieve interval met $\alpha = 95\%$ overigens een overdekking van 99%. In het algemene geval, met K steekproeven, heeft het foutieve éézijdige interval een overdekking van $\Phi(z_\alpha \sqrt{K})$.

De volgende analogie kan wellicht nog duidelijker maken dat het optellen van betrouwbaarheidsgrenzen niet een gelijke maar een grotere overdekingskans geeft. Als ik met één dobbelsteen gooi, dan is de kans $5/6 = 0,83$ dat ik een getal ≤ 5 gooi. Hanekamp et al. zouden dan



Figuur 2. Histogram van de geschatte percentages ammoniakemissie voor drie technieken om mest op landbouwgrond aan te brengen; de zodenbemester en de sleepvoet zijn emissie-arme technieken

redeneren dat gooien met twee dobbelstenen een gelijkblijvende kans $5/6$ geeft op een uitkomst ≤ 10 ($5 + 5$), waarbij dus de predictiegrens 5 die hoort bij één dobbelsteen, verdubbeld is. Het is eenvoudig in te zien dat deze kans groter is, en wel gelijk aan $33/36 = 0,92$. Het probleem wordt pregnanter als we met meer dobbelstenen gooien. Bijvoorbeeld met 6 dobbelstenen is de kans niet 0,83 maar 0,99 op een uitkomst ≤ 30 (6×5).

De forse onzekerheden van Hanekamp et al. zijn dus gebaseerd op een statistische fout. Daarmee vervalt een groot deel van hun kritiek op het WUR-onderzoek.

Hoe het verder ging?

Op het rapport van Hanekamp et al. volgen vele berichten op sociale media. Dit culmineert in een rondetafelgesprek met de vaste Kamercommissie voor Economische Zaken op 22 februari 2017, waarbij Hanekamp cum suis én vertegenwoordigers van het RIVM en de WUR aanwezig zijn. Hierin worden harde verwijten gemaakt. Hanekamp spreekt over 'de afwezigheid van ... wetenschappelijke tegenspraak ... en de ondermijnende consequenties daarvan'. Vogelaar, voorzitter van het Mesdag Zuivelfonds, beticht de onderzoekers van 'misleiding van de Kamer' en 'een NH_3 -anarchie in onze democratie'. De statistische fout is vervolgens tijdens een bijeenkomst met Hanekamp en Crok in Wageningen op 6 maart 2017 door mij naar vo-

ren gebracht. Dat weerhoudt de auteurs van het rapport er niet van om hun bevindingen in ongewijzigde vorm wetenschappelijk te publiceren (Hanekamp et al., 2017b). In een reactie daarop kwantificeer ik de onzekerheid met behulp van de parametrische bootstrap (Goedhart & Huijsmans, 2017). De betrouwbaarheidsintervallen zijn dan een factor drie minder breed. Hanekamp et al. zijn niet onder de indruk. Zij schrijven dat hun en onze methode '... only differ in the manner of estimating this uncertainty. Both approaches have strengths and shortcomings' (Briggs et al., 2017). Op 8 februari 2019 informeert Hanekamp leden van de Tweede Kamer over zijn bezwaren tegen het Landelijk Meetnet Luchtkwaliteit van het RIVM (Rotgers & Hanekamp, 2018). Tijdens deze briefing zegt Hanekamp over nieuwe data die hij van de WUR heeft ontvangen: '... de onzekerheden ... die zijn zo groot dat de vraag rijst of al de experimenten die er ooit voor gedaan zijn überhaupt wel datgene hebben opgeleverd wat we hadden willen weten. ... De eerste bevindingen lijken toch heel anders dan wat de WUR altijd heeft gezegd daarover.' Mogelijk is opnieuw de gewraakte fout gemaakt.

Hanekamp et al. richten hun pijlen vooral op de onzekerheid in de meetmethode voor ammoniakemissie. Waar ze aan voorbij gaan, is dat er 199 emissie-experimenten zijn gedaan, geheel conform de heilige graal van de statistiek: herhalen, herhalen en nog eens herhalen. Een eenvoudige statistische analyse op de emissie percentages van figuur 2 geeft aan dat er zeer significante

én grote verschillen zijn tussen de bemestingstechnieken (Goedhart & Huijsmans, 2017) en dat de emissiearme technieken tot een veel lagere ammoniakemissie leiden. Dit rechtvaardigt de conclusie dat de door Hanekamp et al. gezaaide twijfel over het emissiebeleid van de overheid gebaseerd is op een totaal verkeerd gebruik van statistische methoden.

Als statisticus krijg je niet vaak de kans om rechtstreeks bij te dragen aan het publieke en politieke debat. Het is evident dat statistische inbreng hier cruciaal is.

LITERATUUR

Briggs, W. M., Hanekamp, J. C., & Crok M. (2017). Comment on Goedhart and Huijsmans (2017). *Soil Use and Management*, 33, 603–604. <https://doi.org/10.1111/sum.12393>
Goedhart, P. W., & Huijsmans, J. F. M. (2017). Accounting for

uncertainties in ammonia emission from manure applied to grassland. *Soil Use and Management*, 33, 595–602.

<https://doi.org/10.1111/sum.12381>

Hanekamp, J. C., Crok, M., & Briggs, M. (2017a). *Ammoniak in Nederland; Enkele kritische wetenschappelijke kanttekeningen. V-focusrapport*. Wageningen. <https://bit.ly/2KoHRSg>

Hanekamp, J. C., Briggs, W. M., & Crok, M. (2017b). A volatile discourse; Reviewing aspects of ammonia emissions, models and atmospheric concentrations in The Netherlands. *Soil Use and Management*, 33, 276–287. <https://doi.org/10.1111/sum.12354>

Rotgers, G. R., & Hanekamp, J. C. (2018). *Ammoniak in Nederland. Een noordoostelijke spelbreker*. Leeuwarden: Stichting Mesdag Zuivelfonds. <https://bit.ly/2HXXOGp>

PAUL W. GOEDHART is sinds 1983 statisticus bij Biometris, Wageningen Universiteit en Research.
E-mail: paul.goedhart@wur.nl

Response to Paul Goedhart

We are grateful to Goedhart for helping to highlight the over-certainty that was ubiquitous in the discourse of atmospheric ammonia before 2017. It is good to be part of the change that woke people up to the idea that we were too sure of what was happening! Even now it is difficult to imagine estimates taken directly from the Ryden and McNeil (R&M) model were assumed to be, or were used as if they were, without error or uncertainty. We provided illustrations thereof in our report and paper Goedhart refers to. There are three main sources of the over-certainty and error in the R&M model.

The first arises in the internal parts of the R&M model, which uses a statistical technique to approximate atmospheric ammonia by height. That technique assumed certain mathematical parameters (from a regression) were fixed. We initially suggested one way of adding uncertainty to these parameters, and Goedhart later suggested another. The differences between these techniques, given the prior over-certainties, and the other difficulties of the R&M model, were not especially large or important. Indeed, as we will soon show in upcoming work, the differences are trivial when other problems with the model are accounted for.

Now since all probability, and therefore all probability models, is conditional on the assumptions made, the goodness of the uncertainty method in the mathematical parameters depends on how well the assumptions made accord with reality. Goedhart used a common technique, found in all statistical toolboxes, tied to the idea that the R&M model parameters “had” correlations. Most researchers would use this technique as a kind of default, and Goedhart’s tutorial on this is adequately done. We used a different technique that did not put much interest in the correlations between parameters; our method instead turned an eye towards the predictive ability of the R&M model to better reproduce atmospheric ammonia. That is, we were not trying to adhere to frequentist theory, but instead to physical reality. We wanted predictions, and predictive uncertainties, matching with observations. In the limited samples of data we had at the time, the method we used did the job asked of it.

Second came the other problems with the R&M model, which show our initial attempts did not go far enough. We will leave a full discussion of these shortcomings to our new paper. Here we can say that the R&M model often ‘breaks’ when fed observations that are not perfectly in line with the crude statistical assumption R&M made (about ammonia by height). Absurd, i.e. completely non-physical, values are the result. And this is so regardless which uncertainty-in-the-parameters technique is used. That this has been missed so far in the literature is rather stunning.

Goedhart was right on the money when he said we ought to ‘herhalen, herhalen en nog eens herhalen’ (repeat, repeat, and repeat) our analysis. We did, obviously, take his advice, and subsequently discovered these much larger flaws.

The third and last problem are the observations themselves, produced by experiments that were, as we will show, not the cleanest, to say the least. This often leads to over-estimates of atmospheric ammonia.

Goedhart is right once again that ‘statistische inbreng hier cruciaal is’ to the ammonia debate. But since everybody knows how easy it is to manipulate statistics, or how confusing they can be to interpret, what’s far more important is physics. We want a model that makes skillful and accurate predictions of reality. Everybody can understand good forecasts. With the fixes we suggest in our upcoming paper, we are well on our way to delivering these.

MATT BRIGGS AND JAAP C. HANEKAMP

Matt Briggs is an independent researcher.
E-mail: matt@wmbriggs.com

Jaap C. Hanekamp is a faculty member of the University College Roosevelt, Middelburg, the Netherlands and an adjunct faculty member of the University of Massachusetts Amherst School of Public Health and Health Sciences, Amherst, MA, USA. E-mail: j.hanekamp@ucr.nl

Vrouwen hebben invloed op de omvang van STAtOR!

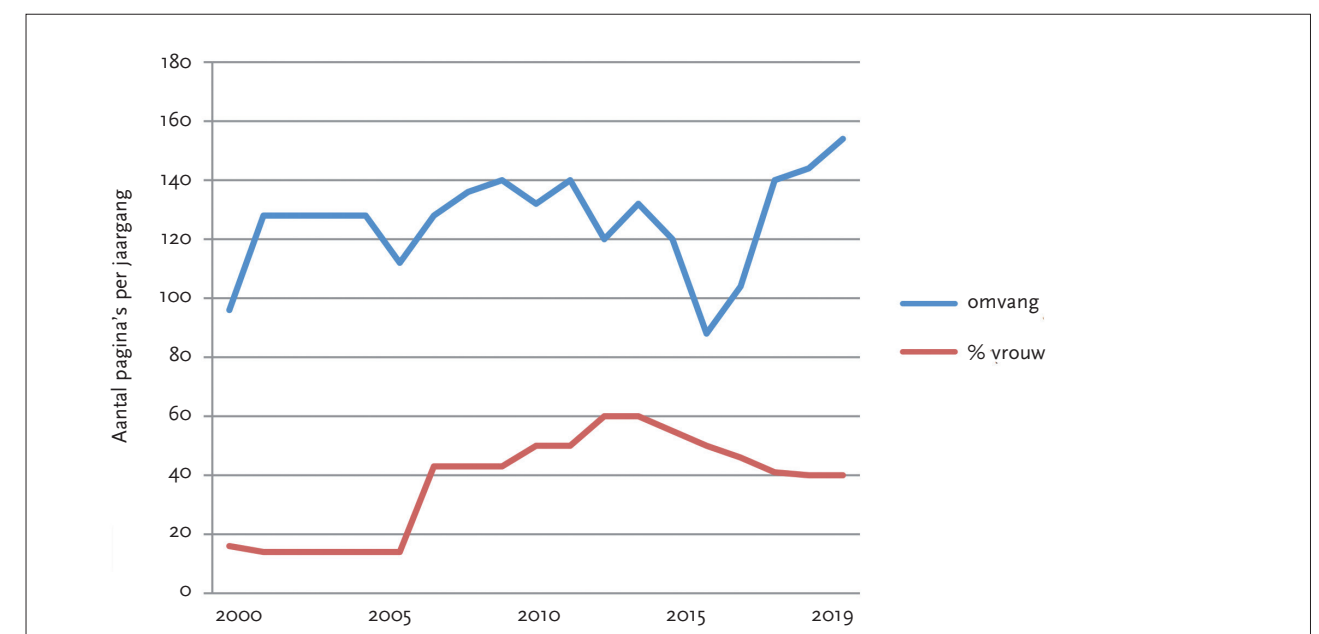
De vakantie staat voor de deur en dat betekent een periode van ontspannen. Ik heb daar steeds een beetje moeite mee gehad, bij mij was er nooit zo’n strikte scheiding tussen werk en vrije tijd. Mijn werk heb ik altijd als ontspannen ervaren en tijdens vakanties zag ik altijd wel iets dat mijn cijfergerichte geest bezig hield. Ik had een soort lijfspreuk, vrij naar Hieronymus van Alphen, ‘mijn werken is spelen, mijn spelen is werken’. Mijn vrouw had als commentaar daarop dat ze hoopte dat ik niet zou overwerken.

We kennen allemaal onderzoeken waarbij correlatie en causaliteit door elkaar worden gehaald. Soms uit onkunde, soms met opzet. In de jaren 80 van de vorige eeuw leverde Rothamsted in de UK bij het computerprogramma Genstat enkele setjes testdata waarmee men kon controleren of de installatie op het lokale computersysteem goed was verlopen. Een van die datasets bestond uit het aantal waargenomen ooievaars en het aantal geboortes in het Verenigd Koninkrijk gedurende een reeks van jaren. Uiteraard was er een sterke correlatie, anders zou de befaamde Britse humor gefaald hebben. Na een uitgebreide congresborrel waarbij dit ter sprake kwam beweerde een Australiër dat hij in zijn land eens op zoek zou gaan naar de correlatie tussen het aantal schapen en het aantal patiënten met een venerische ziekte ...

Met cijfers spelen is leuk. Daarom heb ik de 20e jaargang van STAtOR aangegrepen om alle afgelopen jaargangen eens in cijfers om te zetten. Het kost even tijd, maar dan heb je ook wat. In een spreadsheet heb ik het aantal pagina’s per nummer gezet, samen met het aantal mannelijke en vrouwelijke redactieleden. In de beste traditie van flutonderzoeken heb ik een grafiek gemaakt om te zien of er een verband is tussen deze gegevens. De verrassende conclusie is dat zodra het aandeel vrouwen in de redactie stijgt de omvang van ons blad sterk gaat variëren! (Zie figuur 1) Een boraal politicus zal hier ongetwijfeld iets van vinden, ik beperk me slechts tot de kale waarnemingen.

Terug naar het ontspannen, dat gun ik al onze lezers. Ik nodig u uit om verder te spelen met deze data (de dataset vindt u op de VVSOR-webpagina’s). U kunt er van alles aan toevoegen, van de CBS-prijsindex tot het aantal tropische dagen, het aantal bezoekers van dierentuinen of het aantal veroordelingen voor rijden onder invloed. Gebruik uw inzicht en uw fantasie en verras ons met opzienbarende correlaties! Uiteraard worden de fraaiste verklaringen voor de omvang van STAtOR in ons blad gepubliceerd.

GERRIT STEMERDINK is eindredacteur van STAtOR.
E-mail: gjstemerding@hotmail.com



Figuur 1. Het aandeel vrouwen in de STAtOR-redactie en de omvang van STAtOR van 2000 tot 2019