

“Significantie is niet significant”

Over de stikstofuitspraken van de Raad van State en de eco-statistiek van stikstofdepositie

Door: Ronald Meester^{a)}, Jaap Hanekamp^{b)}, William M Briggs^{c)}, Geesje Rotgers^{d)}

^{a)} Afdeling Wiskunde, Vrije Universiteit Amsterdam, r.w.j.meester@vu.nl. ^{b)} University College Roosevelt, Middelburg, Environmental Health Sciences, University of Massachusetts, Amherst, MA, USA. ^{c)} Independent researcher, Michigan, U.S.A. ^{d)} Redactie Agrifacts.

Dit schrijven is gemotiveerd door de manier waarop de Raad van State (RvS) het woord “significant” gebruikt in belangrijke stikstof-uitspraken. Dit woord wordt tientallen malen gebruikt in drie uitspraken die de onderzoekers van dit artikel hebben bestudeerd.

Middels een uitgebreide analyse van de auteurs blijkt dat de RvS in haar stikstofuitspraken taalkundige, wetenschappelijke en juridische wanorde creëert door de statistische en taalkundige betekenissen van de term “significantie” onterecht en onoordeelkundig te vermengen.

Er is namelijk geen enkele overlap tussen het domein van de statistiek en het domein van de taal bij het overwegen van de term “significantie”.

Door twee onvergelykbare termen aangeduid met hetzelfde woord onkundig te verhaspelen, ontdoet de RvS de onderwerpen die door haar als “significant” worden aangeduid van enig belang en betekenis.

Inleiding

Dit schrijven is gemotiveerd door de manier waarop het woord “significant” in uitspraken van de Raad van State met betrekking tot stikstofdepositie (verwijzend naar stikstofverbindingen zoals ammoniak/ammonium en stikstofoxiden) gebruikt wordt, in combinatie met bepaalde beweringen van ecologen over de “significantie” van hun experimentele resultaten met betrekking tot diezelfde depositie. We zullen uitleggen dat de betekenis van dit woord (“significant”) op een aantal essentiële manieren misverstaan wordt. Dit heeft tot gevolg dat de uitspraken van de Raad van State op specifieke punten ten onrechte de indruk wekken wetenschappelijk onderbouwd te zijn.

Om te zien waar we het precies over hebben beginnen we deze overweging met een aantal voorbeelden, zodat de lezer de context van onze overwegingen beter kan begrijpen. In uitspraak ECLI:NL:RVS:2019:1603 van de Raad van State komen we bijvoorbeeld de volgende frases tegen:

- “[...] een verslechterend of significant verstoerend effect kan hebben [...]”
- “[...] maatregelen gericht op het voorkomen van verslechtingen en verstoringen die een significant effect kunnen hebben op de soorten en typen waarvoor een Natura 2000-gebied is aangewezen.”
- “[...] projecten met mogelijk significant negatieve gevolgen [...]”
- “[...] geen significant verslechterende effecten [...]”

In ECLI:NL:RVS:2022:3159 lezen we bijvoorbeeld dit:

- “Kritische depositiewaarde: dit is de grens waarboven het risico bestaat dat de kwaliteit van een habitat significant wordt aangetast door de invloed van stikstofdepositie.”
- “[...] “waarmee volgens het rapport is gewaarborgd dat significant negatieve effecten op de instandhoudingsdoelstellingen van de betrok-

ken Natura 2000-gebieden kunnen worden uitgesloten.”

- “De lidstaten treffen passende maatregelen om ervoor te zorgen dat de kwaliteit van de natuurlijke habitats en de habitats van soorten in de speciale beschermingszones niet verslechtert en er geen storende factoren optreden voor de soorten waarvoor de zones zijn aangewezen voor zover die factoren, gelet op de doelstellingen van deze richtlijn een significant effect zouden kunnen hebben.”
- “Dat is het geval als niet is uitgesloten dat een project, individueel gezien, significante gevolgen kan hebben voor een Natura 2000-gebied vanwege de stikstofdepositie die in de bouwfase wordt veroorzaakt.”

Tenslotte lezen we in ECLI:NL:RVS:2023:3129 bijvoorbeeld het volgende:

- “Voor zo'n project geldt een vergunningplicht als dat afzonderlijk of in combinatie met andere plannen of projecten significante gevolgen kan hebben voor een Natura 2000-gebied.”
- “Naar het oordeel van de Afdeling is in het rapport van 15 maart 2022 en door de ecooloog van Arcadis op de zitting, overtuigend onderbouwd dat in dit geval de toename van de stikstofdepositie in de bodem die wordt veroorzaakt door het Porthos-project, omdat deze depositie tijdelijk en beperkt is, zich niet zal vertalen in verandering in de vegetatie die aanwezig is in de stikstofgevoelige onderzochte Natura 2000-gebieden en dat daarom op voorhand kan worden uitgesloten dat significante gevolgen optreden voor de habitatkwaliteit.”

We benadrukken dat deze opsomming slechts een selectie is; het woord “significant” wordt veel vaker gebruikt, merendeels in dezelfde context.

Wat is nu het probleem dat we in deze overweging willen adresseren? Welnu, het woord “significant” heeft twee totaal verschillende betekenissen die, helaas, vaak door elkaar gebruikt worden, met grote en serieuze gevolgen. We leggen dit eerst in grote lijnen uit, waarna we in meer detail kunnen treden.

In de Nederlandse taal betekent “significant” zoiets als “veelbetekenend”¹⁾, “belangrijk”, “groot”, “rijk aan implicaties”, “samenhangend”, “betekenisvol”, en vergelijkbare ‘succeswoorden’. Wanneer iemand in gewone spreek- of schrijftaal zegt: “Dit is significant”, dan kondigt hij of zij aan dat er iets heeft plaatsgevonden dat de moeite van onze aandacht waard is. We noemen deze invulling van het begrip “belangrijk-significantie”. Zo was de aanval op de Twin Towers zeker belangrijk-significant, want deze had een zeer grote impact op de geopolitieke dynamiek van de decennia erna.

Ook de Europese Commissie stelt, in een antwoord op een vraag van ons hierover, dat significantie primair moet worden verstaan als “belangrijk-significantie”. In een nadere toelichting, en onder verwijzing naar twee documenten²⁾, komt naar voren dat de Europese Commissie “significant” op een “objectieve manier” poogt te verduidelijken³⁾. Deze benaderingswijze lijkt in lijn met “belangrijk-significantie”, temeer ook omdat de Europese Commissie geen uitspraken doet over eventueel bijbehorende wetenschappelijk methoden.

Dat brengt ons bij de tweede context van het woord “significant”, namelijk zoals gebruikt in de wetenschap en in het bijzonder in de statistiek. In de Dikke van Dale wordt deze tweede betekenis ook expliciet genoemd, en omschreven als “verantwoorde conclusies toelatend”. Binnen de statistiek bestaat er een methode waarmee, zoals algemeen gedacht, bepaald zou kunnen worden welke wetenschappelijk verantwoorde conclusies uit data getrokken kunnen

worden: als de data verantwoorde conclusies toelaten, dan heten deze conclusies “significant”. Wij noemen dit, naar analogie van belangrijk-significant, dan ook “statistisch-significant”.

Deze twee verschillende contexten van het woord “significant” lijken wellicht op het eerste gezicht in elkaars verlengde te liggen. Immers, als iets kennelijk statistisch-significant wordt bevonden, en dus verantwoorde conclusies impliceert, waarom zou dat iets dan ook niet automatisch belangrijk-significant zijn? Andersom, als iets belangrijk-significant is, waarom zou de statistiek dat eveneens niet laten zien?

We zullen hieronder proberen uit te leggen dat statistisch-significant, zoals deze doorgaans wordt gebruikt, op geen enkele manier iets te maken heeft met “verantwoorde conclusies toelaten” zoals Van Dale definieert, en al helemaal niet met belangrijk-significant. Dit wetende zijn de problemen duidelijk: een statisticus concludeert dat een bepaald effect significant is. Hij bedoelt dat als statistisch-significant, maar de leek die het leest kan niet anders dan denken dat dit over belangrijk-significant gaat en het misverstand is geboren.

Onze stelling is dat deze verwarring op grote schaal in de uitspraken van de Raad van State aanwezig is. Als iemand bovenstaande citaten leest, dan zal het woord “significant” daarin gelezen worden als belangrijk-significant, want dat is hoe het woord gebruikt wordt in normaal taalgebruik. Echter, zoals we zullen zien bedoelt de Raad van State vaak wel degelijk statistisch-significant, waarmee het misverstand een feit is.

We zullen nu eerst uitleggen wat statistisch-significant is en in iets meer detail de verwarring bespreken. Daarna zullen we het probleem illustreren met een recente relevante studie.

Statistisch-significantie

Stel je bent als onderzoeker geïnteresseerd in de vraag of een bepaalde stikstofdepositie veroorzaakt wordt door een bepaalde bron, zeg een stal met vee. Hoe gaat een statisticus dan te werk? Meestal gaat dat als volgt. De onderzoeker begint met de zogenaamde “nulhypothese”: de bedoelde mogelijke oorzaak is *niét* de oorzaak van de depositie. Als nu deze aanname zeer moeilijk verenigbaar lijkt met de waargenomen data, dan trekt men veelal de conclusie dat de oorspronkelijke aanname – de nulhypothese – onjuist is én dat de bedoelde oorzaak dus juist *wél* de bron is.

Deze procedure gaat formeel ongeveer als volgt. De waargenomen gegevens worden meestal eerst samengevat in een bepaald getal, dat de “toetsingsgrootte” genoemd wordt. Meestal is de data namelijk te gecompliceerd om helemaal meegenomen te worden, en kies je voor een zinvolle samenvatting ervan. De mate waarin de uitkomst van de toetsingsgrootte opmerkelijk is onder de nulhypothese, wordt vervolgens numeriek uitgedrukt met een zogenaamde ‘p-waarde’. Een kleine p-waarde betekent dat de uitkomst opmerkelijk is onder de nulhypothese. Anders gezegd: Een kleine p-waarde betekent dat de feitelijke uitkomst van ons experiment ‘zeer onverwacht’ is onder de nulhypothese. Omdat de uitkomst toch is opgetreden, ‘geloven’ we niet meer in die nulhypothese.⁴⁾

Echter, de logica van de p-waarde faalt. Dat is ook niet moeilijk te bevatten, omdat een kleine p-waarde, of een p-waarde van welke waarde dan ook, niet logisch impliceert dat de nulhypothese onwaar is of zelfs onwaarschijnlijk. Of waar. Of iets. De grootte van de p-waarde heeft namelijk geen enkele invloed op, of relatie tot, de waarschijnlijkheid dat de nulhypothese waar of onwaar is. Immers, de p-waarde zegt iets *onder de aanname* dat de nulhypothese waar is.

De reden dat dit kennelijk moeilijk in te zien is, heeft wellicht te maken met het volgende feit uit de logica: als ik aanneem dat iets waar is, en vervolgens concludeer ik, gebruikmakend van alleen pure logica, dat deze aanname iets impliceert dat zeker onwaar is, dan *kán* het niet anders dan dat mijn oorspronkelijk aanname onjuist is. Als ik (uiteraard foutief) aanneem dat de oppervlakte van een rechthoek niet lengte maal breedte is, dan kom je uiteindelijk op een tegenspraak uit, waardoor je dus *weet* dat je aanname incorrect moet zijn geweest. Maar deze strikte logica gaat mank zodra er onzekerheid en kansen in het spel zijn. Dat een resultaat een kleine kans heeft onder een bepaalde hypothese zegt helemaal niets. Het zou bijvoorbeeld best zo kunnen zijn dat het resultaat onder elke redelijke hypothese een kleine kans heeft.

Beslissingen die worden genomen op basis van p-waarden zijn daarom pure wilsbeslissingen, en niets anders. Deze kritiek is uiteraard bekend in de statistiek⁵⁾, maar om de een of andere reden blijft de legende van het belang van statistisch-significantie voortleven. Wij denken dat Gerd Gigerenzer gelijk heeft als hij stelt dat p-waarden een simplistisch ritueel bieden voor wetenschappelijke oordelen. In feite neemt de grootte van de p-waarde de beslissing over van de onderzoekers, waardoor ze worden verlost van de last van het nadenken.⁶⁾

Zelfs als de nulhypothese feitelijk onwaar is, is de redenering dat de nulhypothese onwaar is *omdat* de p-waarde klein is, nog steeds ongeldig. Neem bijvoorbeeld dit incorrecte en onzinnige argument: “Socrates is een mens, en alle mensen zijn sterfelijk; *daarom* geldt “1 + 1 = 2”. De conclusie “1 + 1 = 2” is correct, maar dat kunnen we helemaal niet afleiden uit de sterfelijkheid van Socrates als mens. De premisse heeft niets te maken met de conclusie. Het zou dus verkeerd zijn om, zoals sommigen met

p-waarden op een vergelijkbare manier doen, te zeggen dat omdat de conclusie waar is (om andere redenen), uitspraken over Socrates “enig nut” hebben in statistische vraagstukken.

De nadelige gevolgen van het hanteren van de p-waarde en statistisch-significantie zijn nauwelijks te overschatten. In een belangrijke publicatie rekende epidemioloog John Ioannidis voor dat het niet an-

ders kan dan dat de meeste wetenschappelijke hypothesen die statistisch-significant zijn bevonden, helemaal niet waar zijn⁷⁾. Dat gegeven alleen al zou voldoende moeten zijn om p-waarden naar het rijk der fabelen te verwijzen, beseffend dat er meer argumenten zijn (van veelal complexere aard) die significantie naar de vuilnisbelt van de geschiedenis verwijzen.

Het verwarren van statistisch-significant met belangrijk-significant

Zoals de lezer kan zien, heeft het statistiekwoord geen enkele relatie met het Nederlandse succeswoord. Hooguit bij toeval kunnen beiden samenvallen. Omdat er geen verband is tussen statistische theorie en de Nederlandse taal, ontstaat, zoals nu duidelijk moet zijn, de drogreden van dubbelzinnigheid. Dit gebeurt wanneer het statistiek-woord wordt verward met het succeswoord, wanneer “statistisch-significant” wordt verward met “belangrijk-significant”.

Een onderzoeker kan schrijven dat een bepaald resultaat “significant” is, wat dus betekent dat hij een kleine p-waarde heeft berekend. Het probleem ontstaat wanneer iemand die de bevindingen van die onderzoeker leest (of misschien zelfs de onderzoeker zelf) de aankondiging opvat als implicerend dat iets belangrijks of cruciaals is ontdekt. Dit kan waar zijn, maar ook niet. Belangrijk-significantie kan namelijk helemaal niet worden begrepen via statistisch-significantie. Echt helemaal niet.

Het is betreurenswaardig dat het Nederlandse succeswoord ooit is ingelijfd en in dienst is gesteld van de statistiek. Het is moeilijk, zo niet onmogelijk, om “statistisch-significant” te lezen en niet “belangrijk-significant” te horen of te impliceren, hoewel er elke keer dat dit gebeurt een formele drogreden wordt begaan. Het gebruik van het woord heeft deze drog-

reden van dubbelzinnigheid tot een constante van wetenschappelijk onderzoek gemaakt en heeft als zodanig onnoemelijke schade toegebracht aan helder denken. We hopen de lezer zo met een voorbeeld een idee te geven van de ravage die hiermee is aangericht bij de wetenschappelijke studie naar stikstofdepositie en de eventuele impact daarvan op natuur.

De situatie wordt verergerd door wat we het toevals- of correlatiedrogreden zouden kunnen noemen. Alle wetenschappers weten, daar bestaat geen twijfel over, en het is inderdaad een elementair principe van de logica, dat correlatie niet logischerwijs oorzaak en gevolg impliceert. Wanneer een correlatie is ontdekt, kan het evengoed toeval zijn, en een bepaalde verdenking van een wetenschapper heeft niet, zoals vermoed, gewerkt: het resultaat is puur toeval. Toch is de verleiding nauwelijks te weerstaan om wanneer een kleine p-waarde is gevonden, de correlatie waarop de kleine p-waarde is berekend, een oorzaak te noemen. Deze drogreden komt zeer vaak voor.

De correlatiedrogreden wordt verergerd en zelfs giftig gemaakt wanneer deze wordt gekoppeld aan de drogreden van dubbelzinnigheid. Want niet alleen wordt een correlatie ten onrechte beschouwd als een oorzaak, het wordt beschouwd als een belangrijk-significante oorzaak.

Een voorbeeld

Als p-waarden de facto nooit zinvol zijn, waar moet je dan wel op letten? De echte wetenschappelijk-statistische vraag waar het om draait is de volgende: gegeven de uitkomst van een experiment of analyse, wat is daar dan de beste verklaring voor? Dit is geen vraag waar de p-waarde ook maar iets mee te maken heeft, zoals we hierboven duidelijk hebben gemaakt. Deze vraag dwingt de onderzoeker om ook naar alternatieven te kijken, en te onderzoeken welke hypothese het beste past bij de gevonden resultaten. Nogmaals: hier heeft de p-waarde echt helemaal niets mee te maken, en dat is de onderliggende reden dat statistisch-significant helemaal niet belangrijk-significant hoeft te impliceren.

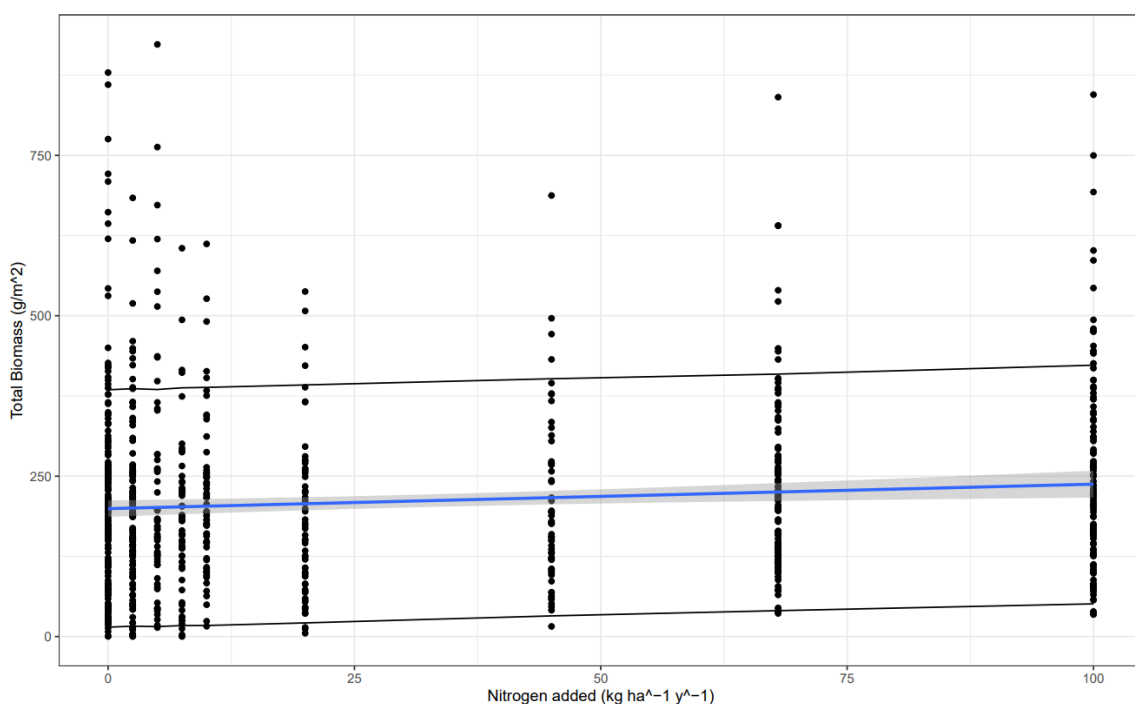
Om ons punt te illustreren geven we één voorbeeld waarin statistisch-significant duidelijk niet belangrijk-significant blijkt te zijn. Symstad et al.⁸⁾ voerden stikstofexperimenten uit in kleine percelen, gedurende vier jaar (2010-2013) in het noordelijke Great Plainsgebied van de Verenigde Staten. De hoeveelheden stikstof die aan verschillende percelen werden toe-

gevoegd, varieerden van 0 (de controlegroepen), 5, 10, 45 en 100 kilo per hectare per jaar. Aan het einde van elk jaar werden verschillende metingen gedaan, zoals de totale biomassa (in gram per vierkante meter).

De achtergronddepositie van stikstof op de locatie werd door de auteurs geschat als ongeveer 3 tot 9 kilo per hectare per jaar.

Als de relatie tussen toegevoegde stikstof en de biologische maatregel “significant” was, werd het model gebruikt om een kritische stikstofbelasting te definiëren. Er zijn verschillende manieren waarop deze gegevens kunnen worden geanalyseerd; hier kiezen we voor de eenvoudigste, namelijk een lineaire regressie van stikstof en totale biomassa met gebruik van alle door Symstad et al. gepubliceerde gegevens. Het model en de resultaten worden weergegeven in deze figuur die we zelf hebben gemaakt op basis van hun data:

Figuur: De data van Symstad et al: relatie tussen stikstofdepositie en biomassa.



De zwarte punten zijn de data, de blauwe lijn is de regressie, en het grijze gebied correspondeert met het 95% betrouwbaarheidsinterval van de regressie zelf. De dunne zwarte lijnen onder en boven vormen de grens van het 90% predictie-interval: gegeven de data (en de regressie) is er 90% kans dat de totale biomassa tussen die waardes ligt. De p-waarde was 0,0052, wat klein is en daarom “statistisch significant” werd genoemd. Hoe die p-waarde in een geval als dit precies wordt bepaald is niet zo belangrijk voor dit artikel, maar het komt erop neer dat gekeken wordt naar de kans op het gevonden verband (die natuurlijk van de data afhangen) als er in werkelijkheid geen verband bestaat tussen depositie en biomassa. De auteurs schrijven nu dat er een “kritische belasting van 6 – 10 kilo per hectare per jaar” werd geïdentificeerd “door aan te nemen dat elk meetbaar effect van de antropogene invloed op de vorm of functie van het ecosysteem schadelijk kan zijn”.

Maar is hun conclusie wel aanvaardbaar? Laten we nog eens goed naar de figuur kijken. Stel dat de kritische belastingwaarde 10 kilo per hectare per jaar is, en onthoud dat de regressie “statistisch significant” is. Als er géén causaal verband is tussen toegevoegde stikstof en de totale biomassa, dan zou de onzekerheid die we hebben over de totale biomassa op elk punt (bij elke waarde van toegevoegde stikstof)

ongeveer hetzelfde zijn. Dat wil zeggen, de waarschijnlijkheid dat de totale biomassa een bepaalde waarde heeft op elk punt, gegeven elke waarde van toegevoegde stikstof, zou ongeveer hetzelfde zijn.

Als toenemende stikstofdepositie leidt tot toenemende biomassa, zoals deze regressie lijkt te suggereren, dan zou de waarschijnlijkheid dat de totale biomassa op een hoog niveau van stikstofdepositie groter is dan de totale biomassa op een laag niveau van stikstof groter moeten zijn dan 50%. Immers, de relatie tussen beiden – stikstofdepositie en totale biomassa – tekent zich steeds scherper af naarmate de stikstofdepositie toeneemt.

In dit geval zien we dat de situaties bij 0 en 10 nauwelijks van elkaar verschillen: als je precies kijkt dan blijkt er een kans van 50,7% te zijn op een hogere totale biomassa bij de “kritische belasting” dan bij 0. Dat is feitelijk niet of nauwelijks te onderscheiden van een muntworp. “Statistische-significantie” heeft zich dus niet ‘vertaald’ in daadwerkelijke belangrijk-significantie; en dat mag nu geen verbazing meer wekken. Bedenk dat we de altijd aanwezige onzekerheid in metingen van zowel biomassa, toegevoegde stikstofdepositie én de achtergronddepositie hier niet eens meerekenen! Als we dat wel zouden doen dan verdwijnt het ‘verschil’ tussen 0 en 10 helemaal.

Kosten en baten

Er zijn nog meer redenen om helemaal niet naar significantie te kijken. Naast vragen over de waarheid en onzekerheid in elke stelling of voorgestelde oplossing, moeten er namelijk ook altijd kosten en verwachte baten van eventueel handelen worden overwogen. Daarbij moeten we minimaal met de volgende dingen rekening houden:

- Wat is de schade en hoeveel onzekerheid hebben we daar eigenlijk over? Deze onzekerheid is nodig om te kunnen beoordelen of acties eigenlijk wel gewenst en doelgeschikt zijn.

- Wat is de mogelijke schade of gevolg van de oplossing? De oplossing kan variëren van “niets doen” tot een hele reeks aan activiteiten. De exacte voorgestelde oplossing of oplossingen moeten duidelijk worden gespecificeerd. Vage oplossingen zijn wetenschappelijk nutteloos, hoe politiek vruchtbaar ze ook mogen zijn (in de politiek loont het vaak om vaag en onduidelijk te zijn). Bijvoorbeeld, laten we veronderstellen dat een oplossing is om een grote boerderij, in de buurt van een locatie waar vermoed wordt dat groei van ‘onkruid’ is toegenomen vanwege hoge stikstofemissies, te sluiten. Deze oplossing kan al dan niet zeker zijn in haar beweerde effecten. De onzekerheid van dit beoogde effect moet worden vermeld. In de meeste gevallen is het zelden zo dat de oplossing zekere effecten sorteert.
- Wat zijn de kosten? Het sluiten van een boerderij zal grote kosten met zich meebrengen. Deze moeten nauwkeurig worden gespecificeerd, niet alleen in de directe kosten, zoals het verlies van levensonderhoud voor velen en het daaropvolgende verdwijnen van voedselproducten die de boerderij produceerde, maar ook in de indirecte kosten, zoals de kansen op

verhoogde kosten van voedsel, de kosten van het geven van meer macht aan de staat, en dergelijke. Al deze kosten zullen onzekerheden met zich meebrengen en die moeten allemaal worden vermeld.

- Wat zijn de voordelen? Het voordeel kan bijvoorbeeld zijn dat de kans op ongewenste groei van ‘onkruid’ wordt verminderd. Of het voordeel kan zijn goede wil van de superieuren van de politicus die het verbod heeft voorgesteld. Wat de voordelen ook zijn, net als de kosten moeten ze in precieze bewoordingen worden uiteengezet. Vage handgebaren zijn nutteloos en allesbehalve wetenschappelijk. Er zal, net als bij de kosten, slechts een kans zijn dat de voordelen worden gerealiseerd. De onzekerheden moeten daarbij ook worden vermeld.

Dit alles betekent dat we sowieso met een complexe wetenschappelijke situatie te maken hebben, die niet zinvol samengevat kan worden in een p-waarde of welk ander getal dan ook. Alle genoemde vragen dienen bij de beoordeling meegenomen te worden.

Discussie en conclusies

Het is nu belangrijk om terug te kijken naar de significantie-citaten van de Raad van State. Ogenschijnlijk is het zo, we schreven het al aan het begin van dit artikel, dat de Raad van State het over belangrijk-significant heeft. Maar dát kan niet zo zijn, en we onderbouwen dit met drie opmerkingen.

Allereerst, wanneer bijvoorbeeld significantie in combinatie met de kritische depositiewaarde ter sprake komt, dan is het *per definitie* zo dat het over statistisch-significant gaat, want de kritische deposi-

tiewaarde is alleen maar in die statistische context gedefinieerd.⁹⁾ In ons laatste citaat wordt verder een ecooloog genoemd die het beroepshalve, en dus wetenschappelijk, over significantie heeft. Tenslotte wordt uit de wetenschappelijke literatuur duidelijk dat voor de onderbouwing van het beleid slechts statistisch-significant bedoeld kan zijn. Zo lezen we in een rapport van Wageningen University¹⁰⁾: “Anders dan in de studie naar de door struikheide gedomineerde heiden, waren er in deze studie wél voldoende waarnemingen in de laagste depositieklasse om

statistisch significante uitspraken te doen bij lage depositieniveaus.” en “De totale soortenrijkdom nam significant af met toenemende stikstofdepositie. De sterkste afname lijkt op te treden tussen de klassen 15-20 kg N/ha/j. en 20-25 kg N/ha/j. Deze conclusie hangt wel af van de precieze lijn die wordt getrokken door de punten: een lineaire regressielijn is significant, maar een exponentiële afname is significanter [...]”.

Dit zijn slechts twee van de vele voorbeelden die in dit rapport en vele andere rapporten over stikstof te vinden zijn. Er kan dus geen twijfel zijn over het soort significantie waar de Raad van State op doelt,

dat wil zeggen statistisch-significant. Daarmee is er ironisch genoeg geen enkele reden om aan te nemen dat er daadwerkelijk iets belangrijk-significant wordt gemeld door de Raad van State. We hebben namelijk uitgelegd waarom de definitie van een p-waarde geen relevante vraag kan beantwoorden over enige hypothese van belang: een p-waarde behandelt eenvoudigweg niets van wetenschappelijke of enige andere waarde. Het ligt, bij nader inzien, voor de hand dat dit juridische orgaan in zijn uitspraken statistiek en taal verwart, met alle nadelige consequenties van dien.

Samenvattend creëert de Raad van State in haar stikstofuitspraken taalkundige, wetenschappelijke en juridische wanorde. Met haar onoordeelkundig vermenging van twee onvergelykbare termen aangeduid met hetzelfde woord, ontdoet de Raad van State de onderwerpen die door haar als “significant” worden aangeduid van enig belang en betekenis. In de noodzakelijke verwijzingen naar ecologisch-wetenschappelijk werk in de uitspraken, ontkracht de Raad van State onbedoeld de benodigde wetenschappelijkheid om haar uitspraken te staven.

Het concrete voorbeeld dat wij gaven, waarin statistische significantie wordt geclaimd in een stikstofdepositie studie maar waar geen belangrijk effect werd gemeten, is hiervan een kenmerkend voorbeeld.

De meest voor de hand liggende conclusie, die keer op keer door statistische vakmensen wordt benadrukt¹¹⁾, is dat men nooit p-waarden moet gebruiken, onder geen enkele conditie. Dat geldt dus ook voor ecologische statistische vraagstukken.

Daarnaast heeft de statistische term “significantie” niets te maken met de betekenis van het woord “significantie” in de Nederlandse taal, waar het zoiets als “belangrijk” of “relevant” betekent. Er is dus geen overlap tussen het domein van de statistiek en het domein van de taal bij het overwegen van de term “significantie”. Statistische significantie heeft geen feitelijke relatie tot belangrijkheid, bruikbaarheid of waarheid. Statistische significantie mag daarom niet worden gebruikt of vertrouwd als basis voor enige beslissing, zeker in het licht van de kosten en baten die daaruit voortvloeien.

Bronnen

- 1) Dit is de beschrijving in de Dikke van Dale.
- 2) MEDEDELING VAN DE COMMISSIE. Beoordeling van plannen en projecten met betrekking tot Natura 2000-gebieden — Methodologische richtsnoeren inzake de bepalingen van artikel 6, leden 3 en 4, van de habitatrichtlijn (92/43/EEG) (2021/C 437/01).
- 3) EUROPESE COMMISSIE. Beheer van Natura 2000-gebieden De bepalingen van artikel 6 van de habitatrichtlijn (92/43/EEG) (2019/C 33/01).
- 4) Objectiviteit is in hoofdzaak de mogelijkheid om kennis van zaken te hebben over iets. Die kennis van zaken geeft inzicht in goede en slechte argumenten die bepaalde overtuigingen over het onderzochte wel of niet ondersteunen. We zullen zien, maar dat niet verder uitwerken, dat statistisch-significant geen inhoudelijke bijdrage kan leveren aan objectiviteit zoals beschreven in de drie vonnissen van de RvS.
- 5) Meester, R., Hanekamp, J.C. *Wetenschap op het snijvlak van recht en onrecht*. Zie bijvoorbeeld <https://denieuwewereld.substack.com/p/ter-verdieping> (11-01-2024).
- 6) De technische details doen er hier niet toe, en laten we achterwege.
- 7) Zie bijvoorbeeld Richard Royal, *Statistical Evidence*, CRC Press 1996; Ronald Meester en Klaas Slooten, *Kan dat geen toeval zijn? – een kritische blik op statistisch bewijs*, AUP 2021; William M. Briggs, *Uncertainty: the soul of probability, modeling & statistics*, Springer 2016.
- 8) Gerd Gigerenzer, *Mindless Statistics*, *The Journal of Socio-Economics* 33: 587 – 606, 2004.
- 9) J. P. A. Ioannidis, *Why most published research findings are false*, *Plos Medicine* 2005.
- 10) Amy J. Symstad, Anine T. Smith, Wesley E. Newton, and Alan K. Knapp, *Experimentally derived nitrogen critical loads for northern great plains vegetation*. *Ecological Applications*, 29(5):e01915, 2019.
- 11) Dit is de oorspronkelijke definitie: "A quantitative estimate of an exposure to one or more pollutants below which significant harmful effects on specified sensitive elements of the environment do not occur according to present knowledge". (Nilsson, S.I. en P. Grenfeldt (eds.), 1988. *Critical Loads for Sulphur and Nitrogen*. NORD 1988: 97. Kopenhagen: Nordic counsel of ministers.)
- 12) G.W.W. Wamelink et al., *Relaties tussen de hoeveelheid stikstofdepositie en de kwaliteit van habitattypen*, Rapport 3089, 2021, blz. 21 respectievelijk blz. 25.
- 13) Zie bijvoorbeeld S.H. Hurlbert, R.A. Levine, J. Utts 2019. *Coup de Grâce for a Tough Old Bull: "Statistically Significant" Expires*. *The American Statistician*, 73:sup1, 352-357, DOI: 10.1080/00031305.2018.1543616.