Antwoord op de reactie van Wim de Lange op het artikel "Hoezo significant?"

MARTIN KNOTTERS, PAUL K. BAGGELAAR EN EIT VAN DER MEULEN

Met interesse lazen wij de reactie van Wim de Lange op ons artikel "Hoezo significant? Over het effect van een ingreep op de grondwaterstand" (Knotters e.a., 2020). Fijn dat ons artikel met gezond boerenverstand te volgen is. Dat beschouwen we als een groot compliment. Statistiek is immers niets anders dan geformaliseerd boerenverstand en onnodig ingewikkeld doen laten we graag aan anderen over. De uitdaging om een aantal volgende stappen te zetten gaan wij graag aan.

De eerste uitdaging: "*Dus aan het begin van het beoordelingsproces kan het alle kanten op gaan. Dan zou je toch een toetsing ontwerpen die beide kanten op kan beoordelen, verhoging en verlaging, maar hoe doe je dat dan?*". Dit is een belangrijk punt. Je kunt één- en tweezijdig toetsen. De 'zijdigheid' van de toets leid je vooraf af uit de onderzoeksvraag en niet achteraf uit de toetsresultaten. In de casus van ons artikel is de vraag of de grondwaterstand is gestegen als gevolg van een ingreep. Dit vertaalt zich in een eenzijdige toets, in dit geval rechtszijdig. Ook speelt hier het verschil in paradigma:

- wil je een beslissing nemen over vergoeding voor natschade? Dat geval behandelen wij in ons artikel met een hypothesetoets (Neyman-Pearson);
- of ben je op zoek naar bewijs voor een effect? Dan kun je een significantietoets (Fisher) uitvoeren. Heb je nog geen idee van de richting van het effect dan toets je tweezijdig, met bijvoorbeeld een *z*-toets bij voldoende vrijheidsgraden, of anders een t-toets. Hoe je dat dan doet? Dat lichten we als volgt toe:
  1. Formuleer de onderzoeksvraag: is na 14 februari 2013 het gemiddelde niveau van de grondwaterstand veranderd?
  2. Definieer de interesseparameter: δ, de staptrendparameter.
  3. Formuleer de nul- en alternatieve hypothese:
     - H_0: $\delta = 0$
     - H_a: $\delta \neq 0$
  4. Kies een significantieniveau, bijvoorbeeld 0,05. Als de kans op de uitkomst van de toetsingsgrootheid, of extremere waarden, gegeven de nulhypothese kleiner is dan dit niveau, dan verwerpen we de nulhypothese. Deze overschrijdingskans noemen we de *p*-waarde.
  5. Definieer de toetsingsgrootheid:

$$z = \frac{\hat{\delta}}{S.E.(\delta)}$$

  waarin $S.E.(\delta)$ de standaardfout van δ. Onder de nulhypothese heeft de toetsingsgrootheid een standaardnormale verdeling: $z \sim \mathrm{N}(0,1)$

6. Bereken de uitkomst van de toetsingsgrootheid en de tweezijdige *p*-waarde:

$$z = \frac{-7{,}20}{3{,}48} = -2{,}069$$

De bijbehorende tweezijdige *p*-waarde is $P(|z| > 2.069) = 0{,}039$. Dit betekent dat de kans op de uitkomst van de toetsingsgrootheid, of extremere uitkomsten, 0,039 is, als de nulhypothese waar zou zijn. We hebben hier als voorbeeld de resultaten van een model uit tabel 2 van ons artikel genomen: TRG, Box-Jenkins, PNO, einddatum 28-9-2017, halfmaandelijkse frequentie. Hiervan is immers aangetoond dat aan de modelveronderstellingen wordt voldaan.

7. De *p*-waarde van 0,039 is kleiner dan het gekozen significantieniveau van 0,05. We concluderen daarom dat na 14 februari het gemiddelde niveau van de grondwaterstand is veranderd, en wel gedaald.

Vervolgens kan deze conclusie aanleiding zijn voor interessant vervolgonderzoek: hoezo gedaald?

"*En als je wel naar de verhoging sec kijkt, waarom dan niet alleen de natte voorjaarsperiode, die er juist voor de landbouwschade toe doet, in de berekening selecteren.*" Wij zijn het met Wim de Lange eens dat je bij landbouwschade kunt kijken naar bijvoorbeeld alleen het voorjaar. Ook zien wij de voordelen van het gebruik van hoogfrequente data, bijvoorbeeld om vast te stellen hoe vaak een aaneengesloten periode van overschrijding van een grondwaterstand voorkomt, met een duur die plantenfysiologisch van belang is. Voor een schatting van de verandering in de gemiddelde grondwaterstand sinds de datum van de ingreep gebruiken we echter liever een reeks met een constante waarnemingsfrequentie, die in dit geval alleen te verkrijgen was door het laatste, hoogfrequente, deel uit te dunnen tot een halfmaandelijkse frequentie.

De tweede uitdaging: "*Laat eens zien die tweezijdige test: hoe doe je dat, hoe reken je dat uit, wat zijn dan de resultaten die je gaat beoordelen?*" Dat hebben we hierboven laten zien. Maar zoals gezegd leidt de onderzoeksvraag in het geval van ons artikel tot een eenzijdige hypothesetoets (Neyman-Pearson) die als doel heeft een beslissing te nemen op basis van de data. Tweezijdig toetsen is gezien de vraagstelling dus niet aan de orde. We pasten die eenzijdige hypothesetoets toe op een werkelijk geval, namelijk een beslissing over schade-uitkering in verband met hogere grondwaterstanden na een ingreep. Daarbij lieten we zien hoe een evenwichtige verdeling van risico's kan worden bereikt bij die beslissing, namelijk door instelling van de *error rates* en α en β een kleinste relevant geachte afwijking, en trokken ter illustratie een parallel met kwaliteitskeuring.

De derde uitdaging: "*Als derde uitdaging zie ik de mogelijkheid om deze aanpak te combineren met de resultaten van een numeriek grondwatermodel*", en "*Ik denk dat de besluitvorming aanmerkelijk meer evenwichtig en gelijkwaardig wordt doordat de resultaten worden bepaald met de marge die de veroorzakende en ontvangende partij kiezen. Graag lees ik hoe de schrijvers tegen deze toepassing aankijken en wat daarvoor nodig zal zijn*". Om een hypothesetoets toe te kunnen passen op de resultaten van een numeriek grondwatermodel is het nodig dat de nauwkeurigheid van die resultaten is gekwantificeerd. Dat kan dus uitbreiding van het model met een stochastische component betekenen, of een onzekerheidsanalyse. Inderdaad, het is een uitdaging om hydrologie, gericht op het beschrijven van fysische processen, en statistiek, gericht op het kwantificeren van onzekerheid, te combineren en het resultaat hiervan toe te passen bij besluitvorming. Dit geeft de betrokken partijen inzicht in de risico's die zij lopen en maakt doelmatiger en evenwichtiger besluitvorming mogelijk.

Referentie

Knotters, M., P.K. Baggelaar en E. van der Meulen (2020) Hoezo significant? Over het effect van een ingreep op de grondwaterstand; in: Stromingen, volume 26(3), pag 37-49.

Auteurs

MARTIN KNOTTERS
Wageningen Environmental Research
Martin.knotters@wur.nl

PAUL K. BAGGELAAR
PB Icastat
Paul.baggelaar@planet.nl

EIT VAN DER MEULEN
AMO
amo@home.nl