

TOEGEPASTE STATISTIEK IN HET WATERBEHEER (9)

Trendvolger of trendsetter?

Naast het toetsen van normen worden veel milieugegevens verzameld om te zien of maatregelen effect hebben, of de goede trend is ingezet. Nu kun je een trend zien door alle meetwaarden in een grafiek te zetten tegen de tijd en daar op het oog een lijn door te trekken. Maar objectieve methoden om een trend vast te stellen zijn beter, omdat ze eenduidiger zijn en een kwantitatieve maat voor de trend geven die ook toetsbaar is. Deze (voorlopig laatste) bijdrage in een reeks over toegepaste statistiek in het waterbeheer geeft een overzicht van gangbare methoden voor het bepalen van een relatie tussen twee variabelen. De trend is een bijzonder geval hiervan, namelijk de relatie tussen meetwaarden en de tijd.

Een goed gebruik alvorens een analyse op de gegevens uit te voeren is het weergeven van de meetgegevens in een figuur. Hiermee wordt een indruk verkregen van het verband tussen meetwaarden en de tijd. We kwantificeren die samenhang met een correlatiecoëfficiënt. Een veel gebruikte is de Pearson correlatiecoëfficiënt (r). De waarde varieert tussen -1 (negatieve correlatie) via 0 (geen correlatie) tot 1 (positieve correlatie). Als deze coëfficiënt 0 is, bestaat geen lineair verband.

Een andere correlatiemaat is die van Spearman. Deze is gebaseerd op de rangorde van de waarden van de variabele (van laag naar hoog) en toetst op een stijgend of dalend verloop van de gegevens. Door de rangorde is deze maat minder gevoelig voor uitschieters. Voor het toetsen of een correlatiecoëfficiënt nul is, wordt bij Pearson verondersteld dat de gegevens van de variabelen normaal verdeeld zijn in tegenstelling tot de laatstgenoemde. In de statistiek maakt men dan ook onderscheid in parametrische (verdeling van de gegevens bekend) en parametervrije methoden, zoals Spearman.

In bijgaande grafiek zijn drie meetreeksen weergegeven waarvan de correlatiecoëfficiënt 0,75 ($r^2 = 0,87$) is. Het verloop is echter wel heel verschillend. De één heeft relatief veel spreiding, een ander heeft mogelijk een uitschieter en de derde lijkt meer gebaat bij een kwadratische verband.

Modelleren van de samenhang tussen twee variabelen

De meest bekende trendlijn is de lineaire die berekend wordt met behulp van de lineaire regressiemethode. Deze methode is gebaseerd op de kleinste kwadraten methode en wordt daarom relatief sterk beïnvloed door uitschieters. Belangrijk bij een regres-

sieanalyse is het beoordelen van de residuen (de waarneming - schatting vanuit de regressielijn). Vaak wordt aan de hand van het patroon van de residuen duidelijk of de aanname van een lineair verband wel de juiste is. De verdere verwerking van het resultaat (bijvoorbeeld het toetsen of de helling significant van nul verschilt), gaat uit van een normale verdeling van de residuen. Vandaar dat de lineaire regressie valt onder de parametrische methode.

Een parametervrije methode om de relatie tussen een variabele en de tijd te analyseren is de rangcorrelatiecoëfficiënt van Kendall, ook wel aangeduid met tau (τ).

De waarde van Kendall's tau is dan:

$$\tau = (K^+ - K^-) / K_{\text{totaal}}$$

τ varieert tussen -1 en 1 evenals de correlatiecoëfficiënt van Pearson.

Voor het berekenen van de helling van een lineaire trend, wordt dan de helling van

Theil gebruikt. Deze komt neer op het berekenen van de mediaan van de hellingen van elk puntenpaar uit een meetreeks en is helder beschreven in het artikel van Boukes, Baggelaar en Dries in *H₂O* nr 5 van 7 maart j.l.. In de Engelse literatuur wordt deze methode ook wel de Sen's slope genoemd.

Na deze voorbeelden willen we nog enkele zaken noemen die het ingewikkeld kunnen maken.

Lastig voor trendberekeningen zijn waarnemingen die in de tijd verband met elkaar houden. Doordat water in een rivier goed gemengd is, veranderen concentraties maar langzaam. Een aantal monsters chronologisch genomen bevat dus eigenlijk informatie die met elkaar samenhangt. We noemen dit autocorrelatie of omdat het zich in de tijd afspeelt, seriële correlatie. Bij de traditionele regressietechnieken gaan we er vanuit dat gegevens niet gecorreleerd zijn.

Tot slot bestaan ook cyclische patronen, zoals een seizoenseffect veroorzaakt door de temperatuur of het groeiseizoen (gewasbeschermingsmiddelen).

Voor het goed kwantificeren van de trend moet voor alle storende invloed gecorrigeerd worden. Al met al kan het goed vaststellen van het effect van het beleid op de waterkwaliteit ingewikkeld worden*.

* Waarschijnlijk na de zomer zal in *H₂O* in deze serie wat dieper ingegaan worden op de methoden voor het vaststellen van een trend.

Voor meer informatie:

r.n.m.duin@rikz.rws.minvenw.nl of
j.steenwijk@riza.rws.minvenw.nl

Vershillende patronen met dezelfde Pearson correlatiecoëfficiënt van 0,75. Die geeft een indruk van het verband, maar niet over het soort verband. De uitschieter treedt op bij tijdstip 10.

