

Over de betrekkelijkheid van verificatie

Huug van den Dool

In mijn jaren op het KNMI, 1975-82, kwam ik sterk onder de indruk van het belang dat aan 'verificatie' van de verwachtingen werd toegekend. Als student in Utrecht had ik daar maar weinig over gehoord. Maar op het KNMI ging er geen dag voorbij of je hoorde wel over de PEE IE, de zgn Prestatie Index (PI). Het lag eigenlijk wel voor de hand, je doet uitspraken, en je wilt dan, als hele of halve wetenschapper en mede namens het publiek, wel weten hoe goed die uitkomen.

Het KNMI was door nijvere Haagse ambtenaren ook gevraagd een kosten-baten analyse te doen en daarbij helpt formele verificatie. Helemaal logisch was verificatie toch blijkbaar niet. Uit een WMO enquête bleek dat zeker de helft van de nationale weersinstituten in die tijd geen enkele verificatie van hun verwachtingen deden; daar bepaalde blijkbaar de klant of de weersverwachting een tevredenheidscertificaat verdiende. Dat was makkelijk zeggen want er was toen nog geen concurrentie, dus de klant had geen alternatief en was allerminst koning. Maar zelfs zonder mededingers was het KNMI uiterst serieus bezig en had diverse experts wier naam ook nu nog aan verificatie verbonden zijn: Hansen, Kuipers als ontwerpers van verificatiemethodes, Daan en Kruijzinga als liefhebber pur sang. (Onder dankzegging ook aan de energieke Allen Murphy die dankzij een kort bezoek (6 maanden?) de faam van deze Nederlanders vooruit hielp.) Misschien lag de verificatiedrift aan de volksaard, die calvinistische mythologie. Wat ik vooral leerde was dat een verwachting eenduidig en verifieerbaar moet zijn, dwz dat er na afloop geen ruimte voor gezeur is en er altijd objectief gezegd moet kunnen worden of de verwachting wel of niet uitkwam en niet iets vaags er tussenin. Het ergste vergrijp dat ik in Nederland mocht aanschouwen was als een assistent werd afgevaardigd om van de waarnemer te vergen dat hij (vrouwen waren er toen nog niet) de zonneshijnmeter met een daartoe bestemde doek van rijp, dauw en stof zou ontdoen zodat dat ene kwartiertje zon, dat we voor een 'treffer' nog nodig hadden, ook werkelijk zou inbranden. Dit incident bewijst hoe eerlijk we waren. We oefenden op meerdaagse en weekverwachtingen met een PI formulier, waarop strepen getrokken werden zodat er een precieze 'inzet' kon worden bepaald en, achteraf, al dan niet een 'treffer' met waarnemingen waarmee niet te sjoemelen viel. Geen halve treffer. Geen 'eigenlijk hadden we gelijk'.

Aldus voorbereid en geïnstrueerd voor een carrière buitengaats was ik verbaasd in de VS te leren over het begrip 'forecast of no regret', daar had ik op het KNMI niet van gehoord. Vertalen van dit begrip gaat eigenlijk niet. De voorbeelden maken het hopelijk duidelijk. Ik leerde hierover van Don Gilman als iets dat 'eigenlijk' niet mag, maar niet altijd te vermijden is. In alle gevallen gaat het om 'slecht' weer met grote negatieve invloeden op de maatschappij. Ik geef twee voorbeelden uit de lange termijnverwachting. Je voorspelt een koude winter in het Noordoosten van de VS, waar vele grote steden zijn. Komt de verwachting uit dan heb je gelijk, komt het niet uit dat is iedereen blij want niemand wil hoge brandstofprijzen. Dat is dus altijd dik voor elkaar. Ander voorbeeld: Je voorspelt een droogte voor het midden westen in de zomer. Komt het uit dan heb je gelijk, maar als er geen droogte komt is iedereen blij want droogte is niet goed voor de graanprijzen. Wederom alles in orde. Zo kun je wel een tijdje doorgaan, maar deze twee zijn de beste voorbeelden. Ik kan me niet een precies equivalent in NL bedenken. Het mag dus niet, want we zijn allemaal onbevooroordeeld en wetenschappelijk bezig. De heren Hansen en Kuipers staan klaar om je verwachting

koel en zakelijk te beoordelen. Geen echte oplichterij hier, zoals bij de waarzegger die een zwangere vrouw vertelt dat het een meisje wordt, maar 'jongetje' in z'n grootvoorspellingenboek schrijft; wordt het een meisje dan heb je een tevreden klant, wordt het een jongetje blijkt de klant niet goed geluisterd te hebben zodat de ondernemer altijd gelijk heeft. Maar toch is de forecast of no regret het begin van het slipperige pad. Bij de 6-10daagse verwachting in de VS doorzien de forecasters het verificatiesysteem en spelen er op in. Zij weten met name waar de ruwweg 100 verificatiestations liggen. Als ze b.v. Much Below (MB; in een 5 klassensysteem) mogelijk achten, maar er niet voor gestraft willen worden als het niet uitkomt, dan trekken ze de MB contour tussen de stations door in een gebied dat overigens alleen maar Below is. Het PI systeem bespelen was natuurlijk ook de KNMI-ers niet vreemd. Er is, de menselijk psychologie begrijpende, bovendien een neiging jezelf te rechtvaardigen, of om je in te dekken voor diverse mogelijkheden, vooral die mogelijkheden waar je het meeste last mee kan krijgen. Eerlijke verwachtingen dus, maar wel in combinatie met beroepsmatige overlevingstrategie om het zwarte pieten voor te blijven. Probleem is dat de klant (in de VS 'user' genaamd, geen customer in het Engels en geen gebruiker in het Nederlands) dat al half verwacht en daar een cynische houding door krijgt. Dat is een groot probleem met waarschijnlijkheidsverwachtingen. Afgezien van de liefhebbers (ikzelf, energy traders, hydrologen, weather derivatives fans en een enkele verwarde academicus), worden waarschijnlijkheidsverwachtingen alom opgevat als je indekken om nooit echt fout te zijn. Lafheid. Dat cynisme wordt versterkt door het feit dat de werkelijkheid maar één keer gebeurt en dus een deterministische indruk maakt. Morgen is er 80% kans op (meetbare) regen in een tijdvak van 12 uur. Prachtig. Deze verwachting is niet verifieerbaar!!, en dus verboden volgens de vroeger geldende regels. Of het nu wel of niet regent....het kan best waar zijn dat de kans 80% was, maar we zullen het nooit weten. Hoewel hij veel met Allan Murphy werkte was Lev Gandin de, voor mij, meest verrassende tegenstander van waarschijnlijkheidsverwachtingen met de opmerking: 'Wat wil je nu eigenlijk...dat ik 80% van m'n paraplu meebreng??'

Verificatie, hoe belangrijk ook om vooruitgang te meten, is voornamelijk werk in de ivoren toren, want het kan alleen worden toegepast op eenduidig materiaal, zoals de strepen op een PI formulier, of numerieke modeluitkomsten. Met de tekstverwachtingen zijn we, als verificateur, natuurlijk helemaal en totaal in troebel water. En 99% van de klanten hoort niets anders. En er is geen enkele verbetering t.o.v. vroeger. Op 16 Februari 2005 was de lange termijn verwachting van het KNMI: "Kwakkelweer met in de nacht overwegend lichte vorst en overdag temperatuur enkele graden boven nul. De kans op regen of sneeuw neemt af van 80 naar 40% gedurende de periode". Geen geringe zaak. 'Mijn' local weather service office in de VS zei op 18 Februari over a.s. Maandag (vrije dag; 21 Feb): "PRESIDENTS DAY CLOUDY. SNOW LIKELY WITH A CHANCE OF RAIN IN THE MORNING...THEN A CHANCE OF RAIN IN THE AFTERNOON. HIGHS AROUND 50. CHANCE OF PRECIPITATION 60 PERCENT." Misschien is de klant tevreden met deze tekstverwachtingen, maar ga dat maar eens verifiëren! Terzijde: In dit land werken we dus met hoofdletters voor nimmer aflatende nadruk en Fahrenheit, want Celsius is een communistisch complot. Als we hier in de VS een sneeuwstorm missen (we voorspelden 0 inches), en er liggen al 6 inches op de grond dan heet het opeens dat we een total accumulation van 8 inches verwachten. Met terugwerkende kracht kletsen we die 6 inches er alsnog bij. En als de westkust zware regen heeft gehad en er veel geleden wordt in overstroomde woonwagenkampen dan zegt het Journaal standaard:

“Forecasters say ‘more is on the way’”. Een forecast of no regret. Vaak is de neerslag dan al bijna over.

Ik heb de lezer de cumulative ranked probability score en andere hoogstandjes uit de moderne verificatie bespaard. Een bijzonder voorval in begin 1998 wil ik U niet onthouden. Dat was net na afloop van de succesvolle seizoensverwachting voor DJF 1997/98. De formele verificatie wordt eerst met stationswaarden gedaan (score 1; die viel hoog uit), en een tijdje later met voorlopige climate divisions (score 2). Score 2 was nog hoger dan score 1. Na verloop van maanden komen de zgn finale climate division waarnemingen en wederom ging de score omhoog. Onze verwachting werden beter en beter. Helemaal finaal waren de waarnemingen niet want bij de 4de herziening ging onze verificatie er wederom op vooruit. Zouden de verwachtingen beter zijn dan de waarnemingen?. Verificatie is een symmetrisch werkje, en soms, heel soms verifieer je de waarnemingen tegen de verwachting.