

# STAtOR

---

periodiek van de VVSOR jaargang 22, nummer 3, september 2021

Hoe vaak is vaak?

Nieuwe paradigma's in officiële statistiek

Riddler Solitaire

The Lord of the Rings; De impact van sentiment op netwerkvorming

Combinatorische optimalisatie en Quantum Computing

Robuust klimaatbeleid

Monitoring the Future

---



# STATOR

Jaargang 22, nummer 3, september 2021

STATOR is een uitgave van de Vereniging voor Statistiek en Operations Research (VVSOR). STATOR wil leden, bedrijven en overige geïnteresseerden op de hoogte houden van ontwikkelingen en nieuws over toepassingen van statistiek en operations research. Verschijnt 4 keer per jaar.

### Redactie

Joaquim Gromicho (hoofdredacteur), Annelieke Baller, Joep Burger, Caroline Jagtenberg, Guus Luijben (eindredacteur), Kerry Malone, Richard Starmans, Gerrit Stermerdink (eindredacteur), Vanessa Torres van Grinsven en Sanne Willems. Vaste medewerkers: John Poppelaars, Gerard Sierksma en Henk Tijms.

### Kopij en reacties richten aan

Prof. dr. J.A.S. Gromicho (hoofdredacteur), Universiteit van Amsterdam Faculteit Economie en Bedrijfskunde, Sectie Operations Management | Amsterdam Business School, Plantage Muidergracht 12, 1018 TV Amsterdam, j.a.s.gromicho@uva.nl

### Bestuur van de VVSOR

Voorzitter: prof. dr. Casper Albers, db@vvsor.nl; Secretaris: Pieter Jongsma MSc, secretaris@vvsor.nl; Penningmeester: Judith ter Schure MSc, penningmeester@vvsor.nl; Algemeen bestuurslid: Thomas Wise MSc, db@vvsor.nl; Webmaster: Eugenio Traini MSc: webmaster@vvsor.nl.

Voorzitters van de secties: prof. dr. ir. Mark van de Wiel (Biometrical Section); prof. dr. Albert Wagelmans (Section for Operations Research); dr. Eduard Belitser (Section Mathematical Statistics); prof. dr. Casper Albers (Social Sciences Section); dr. Michel van de Velden (Economics Section); dr. Daniel Oberski (Section Data Science); Marije Sluiskes MSc (Young Statisticians) dr. Sanne Willems (Section Statistics Communication).

### Leden- en abonnementenadministratie van de VVSOR

VVSOR, Maarsbergseweg 20, 3956 KW Leersum, admin@vvsor.nl. Raadpleeg onze website [www.vvsor.nl](http://www.vvsor.nl) over hoe u lid kunt worden van de VVSOR of een abonnement kunt nemen op STATOR.

### Voor advertenties

M. van Hootegem, hootegem@xs4all.nl  
STATOR verschijnt in maart, juni, september en december.

### Ontwerp en opmaak

Pharos, Nijmegen

### Uitgever

© Vereniging voor Statistiek en Operations Research  
ISSN 1567-3383

## We kunnen het niet vaak genoeg zeggen!

Hoe vaak verzuchten we wel niet dat er meer aandacht voor ons vak moet komen. Regelmatig komen we berichten en uitspraken tegen die ons tot zo'n verzuchting brengen. Maar wat betekenen woorden als 'vaak' en 'regelmatig' eigenlijk? Dikwijls verrichten we precieze analyses op gegevens die verkregen zijn met dit soort niet-precieze frequentie-aanduidingen. Daarom zijn we erg blij met het artikel van Sanne Willems over het onderzoek dat ze samen met Casper Albers en Ionica Smeets heeft gedaan naar de interpretatie van dit soort woorden.

Naast dit openingsartikel bevat deze STATOR nog een aantal interessante bijdragen. Zo schrijft Frank Pijpers over paradigma's in de officiële statistiek en illustreren Jadé Dieteren en Demi de Kort de dynamiek van een sociaal netwerk en de sentimentscore van de actoren in dat netwerk aan de hand van het filmscript van *The Lord of the Rings*. Verder gaat Frank Phillipson in op de impact die Quantum Computing zal hebben op combinatorische optimalisering en schrijft Leo Huberts over voorspellingen op basis van huidige waarnemingen.

Natuurlijk hebben we ook enkele columns: Henk Tijms over een bedrieglijk eenvoudig kansspel, John Poppelaars over robuust klimaatbeleid en Gerrit Stermerdink die in nostalgisch gemijmer vervalt. Met groot plezier kunnen we u meedelen dat vanaf het volgende nummer ook regelmatig columns van Jelke Bethlehem zullen verschijnen, lezers die de blog van Jelke kennen zullen het met ons eens zijn dat dit een aanwinst is.

Dit nummer wordt gecompliceerd door een bericht over een belangrijke prijs voor Van Dantzig-laureaat Marloes Maathuis, nieuws van de Young Statisticians en – een unicum voor ons blad – een koninklijke bijdrage over het belang van statistiek. Tijdens de opening van het virtuele World Statistics Congress op 11 juli van dit jaar werd een boodschap van koningin Máxima voorgelezen, wij kregen toestemming die over te nemen.

Ten slotte nemen we afscheid van onze mede-redacteur Kristiaan Glorie. Kristiaan, bedankt voor de vele en interessante inbreng gedurende lange tijd!

Wij wensen u veel leesplezier, maar ja, wat bedoelen we met veel?



## INHOUD

- 2 We kunnen het niet vaak genoeg zeggen
- 4 Hoe vaak is vaak? | SANNE WILLEMS
- 8 Nieuwe paradigma's in officiële statistiek | FRANK PIJPERS
- 11 Ethel Newbold Prize 2021 voor Marloes Maathuis
- 12 Riddler Solitaire – column | HENK TIJMS
- 14 The Lord of the Rings; De impact van sentiment op netwerkvorming | JADÉ DIETEREN & DEMI DE KORT
- 19 Combinatorische optimalisatie en Quantum Computing | FRANK PHILLIPSON
- 24 Robuust klimaatbeleid – column | JOHN POPPELAARS
- 26 Monitoring the Future | LEO HUBERTS
- 30 Opa vertelt – column | GERRIT STEMERDINK
- 31 World Statistics Congress 2021; Message from H.M. Queen Máxima of The Netherlands
- 31 Young Statisticians



Waarschijnlijk baseer jij veel beslissingen op statistisch geschatte kansen. Als komend weekend de kans op mooi weer groot is, besluit je misschien wel om naar het strand te gaan. Wanneer je een vakantie boekt, doe je er wellicht een annuleringsverzekering bij. En de keuze voor of tegen vaccinatie kan ook te maken hebben met wat jij onder een kleine kans op bijwerkingen verstaat.

SANNE WILLEMS

Waarschijnlijk, wellicht, misschien, grote kans. Dit zijn allemaal manieren om een kans in woorden uit te leggen. Ook professionals doen dat, vooral in gesprekken. Je arts vertelt bijvoorbeeld dat je een grote kans hebt op genezing, maar dat mensen zoals jij vaak last hebben van

bijwerkingen van het medicijn.

Maar wat bedoelt de arts dan precies met deze woorden 'grote kans' en 'vaak'? Welke kans heeft deze arts in gedachten? En vooral: komt deze kans overeen met hoe jij hem interpreteert? En zou een andere arts dezelfde

woorden kiezen voor die kansen? Om inzicht te krijgen in de verschillen onderzocht ik samen met mijn collega's Ionica Smeets en Casper Albers de variatie in de interpretatie van kans- en frequentiewoorden.

#### Waarom in woorden?

De toekomst is onzeker. Als we toch voorspellingen maken, bijvoorbeeld over de kans op regen, dan doen we dat met statistische modellen. Deze modellen gebruiken gegevens uit het verleden, maar die bieden natuurlijk geen garantie voor de toekomst. Een voorspelling blijft altijd een schatting en die is nooit exact.

Deze onzekerheid is een reden om terughoudend te zijn in het geven van exacte getallen. Dat zou namelijk de verwachting kunnen scheppen dat de kans precies bekend is, in plaats van geschat. Vooral in gesprekken gebruiken we daarom vaak kanswoorden (onwaarschijnlijk, grote kans, misschien) of frequentiewoorden (vaak, nooit, meestal). Deze uitspraken geven indirect al de statistische onzekerheid aan.

Maar juist door die indirecte onzekerheid kan er verwarring ontstaan. Want wat als je arts met grote kans 90 procent kans op genezing bedoelt, en jij denkt dat de kans 'maar' 60 procent is? Zo'n verschil in interpretatie kan ervoor zorgen dat jij sterk twijfelt over

een behandeling die eigenlijk een erg grote kans van slagen heeft.

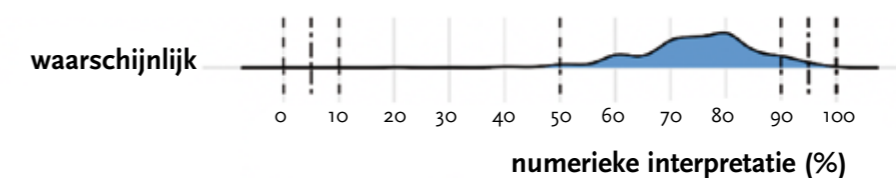
#### Grote variatie in interpretatie

Stel je leest de zin 'Waarschijnlijk wint het team de wedstrijd', hoe groot denk jij dat dan de kans is dat het team de wedstrijd wint? In ons onderzoek legden wij dat soort vragen voor aan onze deelnemers, elke zin met een ander kans- of frequentiewoord.

In figuur 1 zie je de resultaten voor waarschijnlijk. Op de horizontale as staan alle percentages van 0 tot 100 procent en de hoogte van de grafiek geeft aan hoeveel deelnemers een bepaald percentage kozen. Voorbeeld: de grafiek is het hoogst bij 80 procent, dit betekent dat dat het meest gekozen percentage is voor de interpretatie van waarschijnlijk.

De grafiek is erg breed. Hij begint bij ongeveer 45 procent en heeft een lange staart die doorloopt tot 100 procent. De interpretatie van het woord waarschijnlijk is dus erg gevarieerd.

Tussen 60 procent en 90 procent is de grafiek vrij hoog. Dus als jij waarschijnlijk gebruikt om een kans van 90 procent aan te duiden, is het niet ondenkbaar dat iemand anders dit interpreteert als een kans van nog geen 70 procent. Dat kan dus veel verwarring opleveren.



Figuur 1. De verdeling (dichtheidfuncties) van de interpretatie van kanswoord 'waarschijnlijk'; dit is een soort uitgesmeerd histogram, hierdoor vloeit de grafiek uit buiten de grenzen 0 en 100 procent

Figuur 2 geeft de resultaten van alle onderzochte kans- en frequentiewoorden weer. Bij de meeste uitspraken zie je hetzelfde verschijnsel: de grafieken zijn heel breed. Er zitten dus grote verschillen in de interpretatie van veel van deze woorden.

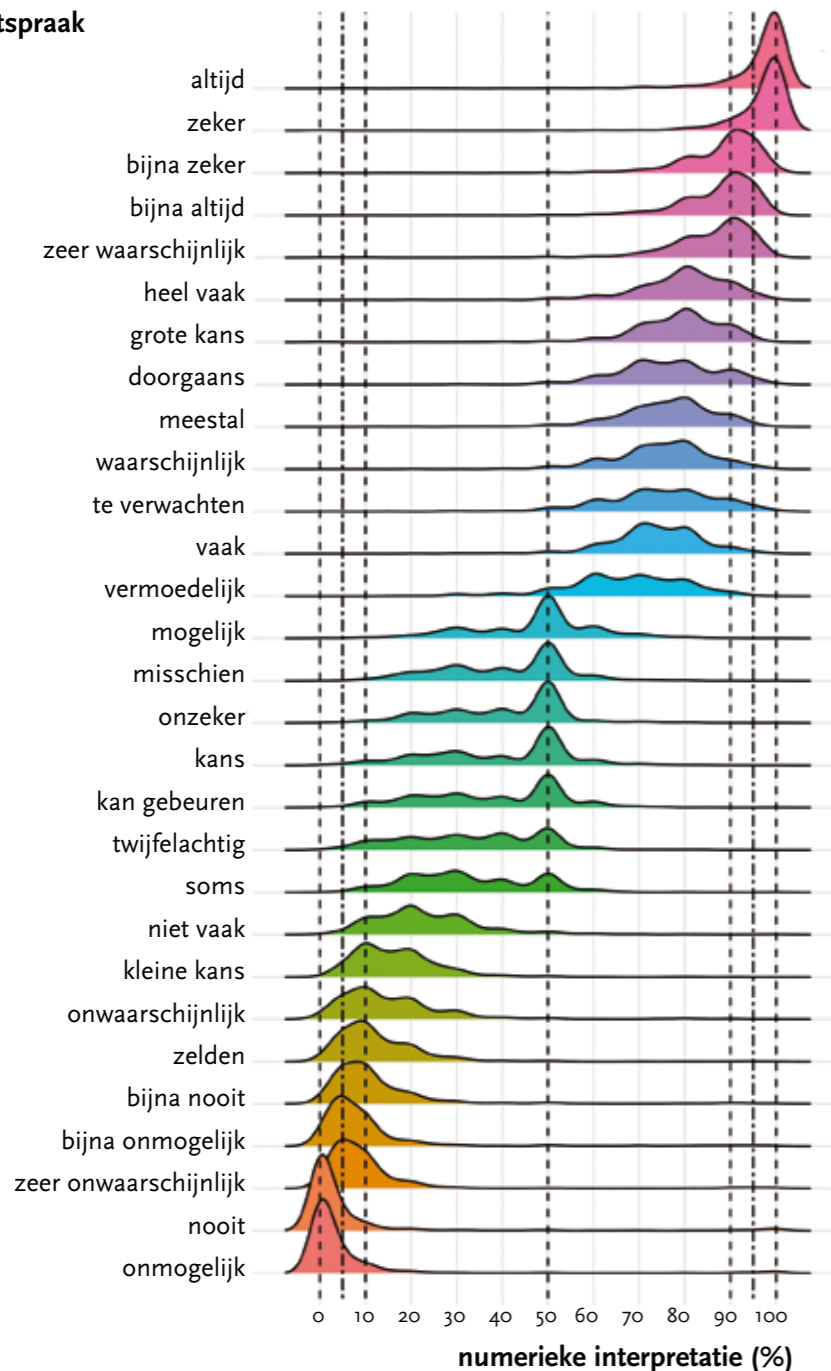
Alleen bij de extreme uitspraken, zoals altijd, zeker, nooit en onmogelijk, heeft iedereen bijna dezelfde interpretatie. De grafieken zijn minder breed en er zijn hoge pieken rond 100 en 0 procent. Dit is niet onverwacht; bij deze woorden is er weinig ruimte voor eigen interpreta-

tie. Wat me wel verbaast is dat sommige deelnemers een onmogelijke gebeurtenis toch nog een kans van 10 procent gaven. Het onmogelijke is dus toch mogelijk?

### Helpt ervaring?

Hoe zit dat bij de experts? Je zou misschien verwachten dat statistici het onderling wel eens zijn over de interpretatie van verbale kansuitspraken. Daarom vroegen wij

### Kansuitspraak



Figuur 2. De verdeling (dichtheidsfuncties) van de interpretatie van kans- en frequentie woorden; dit zijn een soort uitgesmeerde histogrammen, hierdoor vloeien de grafieken uit buiten de grenzen 0 en 100 procent

deelnemers of zij regelmatig (wekelijks of maandelijks) statistische analyses doen. Zo konden we de antwoorden van de twee groepen vergelijken en nagaan of ervaring met statistiek voor overeenstemming in de interpretatie zorgt.

In figuur 3 vergelijken we de resultaten van de twee groepen voor vijf kansuitspraken. De rode grafieken geven de antwoorden van de statistici weer, en die van de andere deelnemers zijn blauw. De paarse gebieden laten de overlap tussen de twee groepen zien. Er is veel overlap tussen de twee groepen, het grootste gedeelte van de grafiek is namelijk paars. En statistici zijn het onderling ook niet eens zijn over de interpretatie van de verschillende uitspraken, want ook de rode grafieken zijn erg breed.

Meer ervaring met kansen zorgt dus niet voor een eenduidige interpretatie. Misschien is dit niet zo vreemd. Als ik met mijn collega's de uitkomsten van modellen bespreek, dan doen we dat eigenlijk altijd in getallen. We gebruiken onderling dus zelden kanswoorden en bespreken daarom ook nooit de interpretatie daarvan.

### Hoe dan wel?

Niet iedereen interpreteert kanswoorden op dezelfde manier. Dit kan tot verwarring leiden. Hoe lossen we dat op? Een voor de hand liggende oplossing lijkt om de verbale kansuitspraken te verbannen en alleen nog met numerieke percentages te werken. Maar helaas is het niet zo mak-

kelijk. Veel mensen vinden percentages namelijk lastig te begrijpen. Het gebruik van percentages leidt dus óók tot misverstanden.

Wetenschappers aan verschillende universiteiten doen onderzoek naar de duidelijkste weergave van kansen. Zo bekijken ze bijvoorbeeld de combinatie van woorden en kansen en het effect van verschillende weergaven van percentages, bijvoorbeeld '25 procent' versus '25 van de 100'. Ook onderzoeken ze welke factoren van invloed zijn, zoals gecijferdheid.

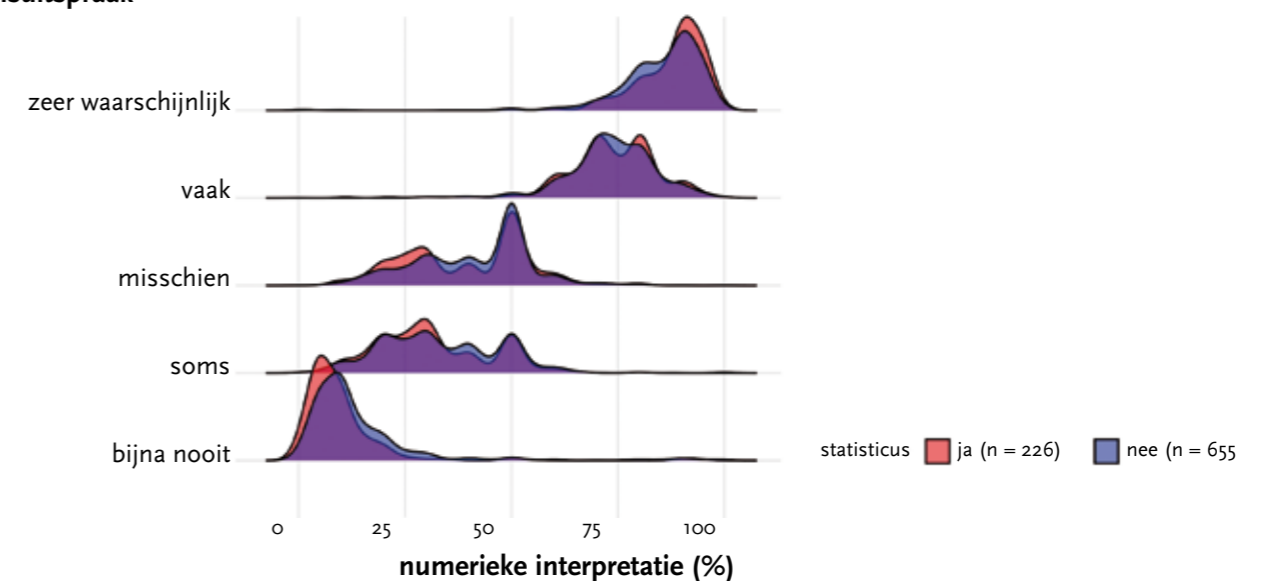
Dé oplossing is nog niet gevonden. Het is wel duidelijk dat woorden alleen niet genoeg zijn. Gebruikt jouw arts toch een keer een kanswoord? Vraag dan om verduidelijking.

### VERANTWOORDING

Dit artikel is eerder verschenen op het nieuwe online blog van de VVSOR (blog.vvsor.nl). Het is gebaseerd op de resultaten van een onderzoek dat de auteur heeft uitgevoerd in samenwerking met Prof.dr. Ionica Smeets en Prof.dr. Casper Albers. De volledige resultaten zijn gepubliceerd in het wetenschappelijke artikel van S. Willems, C. Albers & I. Smeets. (2020). Variability in the interpretation of probability phrases used in Dutch news articles; A risk for miscommunication. *JCOM: Journal of Science Communication*, 19(2). [https://jcom.sissa.it/archive/19/02/JCOM\\_1902\\_2020\\_A03](https://jcom.sissa.it/archive/19/02/JCOM_1902_2020_A03)

SANNE WILLEMS is universitair docent bij de sectie Methodologie en Statistiek van Universiteit Leiden. Verder is zij voorzitter van de VVSOR sectie Statistics Communication. E-mail: s.j.willems@fsw.leidenuniv.nl

### Kansuitspraak



Figuur 3. De verdeling (dichtheidsfuncties) van de interpretatie van kans- en frequentie woorden; dit zijn een soort uitgesmeerde histogrammen. Let op, hierdoor vloeien de grafieken uit buiten de grenzen 0 en 100 procent



# NIEUWE PARADIGMA'S IN OFFICIËLE STATISTIEK

FRANK PIJPERS

Naast het uitgebreide tabellenmateriaal en de bijbehorende berichtgeving die het CBS uitbrengt ter vervulling van het kernprogramma voor de officiële statistiek van Nederland, voert het CBS ook veel aanvullend statistisch onderzoek uit, waaronder bijvoorbeeld voor diverse ministeries, en voor andere organen van openbaar bestuur zoals de planbureaus. In verreweg de meeste gevallen vereist een antwoord op de vraagstukken die binnen komen meer dan puur het presenteren van gegevens. Zelfs de beoordeling van significantie van een verschil in een indicator tussen groepen, vereist een onderliggend statistisch model en dus ook een zekere mate van technische

duiding. Soms is het nodig om te corrigeren voor bepaalde achtergrondkenmerken, bij het beantwoorden van gedetailleerde vragen over subgroepen in de populatie. Ook daar zijn statistische modellen en modelleringstechnieken onontbeerlijk.

Het is dus vanzelfsprekend dat het CBS veel expertise heeft opgebouwd over statistische modelleringstechnieken en de rol van modellen bij het toetsen van hypothesen. Het is een centrale rol van de methodologen binnen het CBS om innovaties in dergelijke technieken bij te houden zodanig dat het voor de statistiek van het CBS van meerwaarde is. Het CBS vindt het in voorkomende

gevallen ook belangrijk om zelf ook bij te dragen aan de ontwikkeling daarvan en aan de toepassing van deze technieken binnen het CBS maar ook bij andere (inter-) nationale statistische instituten. Een visie op dit onderzoek voor de komende vijf jaar is ook te vinden op de CBS website (Pijpers, 2020).

## Complexe samenleving

Een relatief jonge wetenschappelijke discipline houdt zich bezig met systemen die bestaan uit heel veel onderdelen die zelf een zekere autonomie hebben, maar wel onderling op elkaar inwerken. Deze wordt wel complexiteitswetenschap genoemd, of complex-systeemanalyse. Die onderlinge wisselwerking tussen deeltjes (die ook wel actoren of agents genoemd worden, afhankelijk van de context) kan een ogenschijnlijk heel simpele vorm hebben, zoals de meest basale  $1/r^2$  waar  $r$  de afstand is tussen de onderdelen. Toch leidt dat vaak tot een enorme rijkheid aan verschijnselen van het systeem als geheel.

Dat geldt niet alleen voor de vorming van grote-schaal-structuren van melkwegstelsels in het heelal. Je kunt net zo goed denken aan personen in de samenleving die gezamenlijk spontaan trends opleveren in bijvoorbeeld economie, of mode, of maatschappelijke *grassroots* bewegingen. Complexe interacties kunnen dus zowel tot solidariteit als polarisatie tussen groepen leiden. In de economie speelt het ook een rol bij beoordelingen van systeemrisico, en veerkracht of kwetsbaarheid, van economische waardeketens in de industrie. Natuurlijk zijn de interactieregels bij personen of bedrijven niet een simpele  $1/r^2$  wet maar dat betekent des te meer dat de paradigma's van complexiteitswetenschap van toepassing zijn op economie en samenleving.

Er is een grote behoefte aan het identificeren en kwantificeren van oorzaken en gevolgen; van causale verbanden. Een voorbeeld is het kwantificeren van hoe centraal de economische rol is van de bedrijfssector van luchtvaart, of van de chemisch-technologische bedrijven in de Nederlandse economie; die economie is een netwerk van handel en diensten die worden uitgewisseld tussen bedrijven en daarbinnen is 'centraliteit' wiskundig te definiëren en objectief te meten en te aggregeren naar bedrijfssectoren. Een dergelijke objectieve meting is cruciale input voor planbureaus. Een tweede voorbeeld is de wisselwerking tussen fysieke en sociale leefomgeving en gezondheid, tussen gezondheid en opleidingsniveau, en tussen opleidingsniveau en draagvlak voor maatschappelijke investeringen in verbeteringen van leefomgeving, opleiding, en gezondheid. Het is belangrijk te kwantifi-

ceren hoe groot en hoe snel een investering in ieder van deze gebieden echt effectief wordt. Die tijdschalen en effectgroottes zijn objectief te meten uit trends over recente jaren. Ook dit is weer cruciale input voor planbureaus. Zeker nu de Verenigde Naties, en ook Nederland, de zogenaamde Sustainable Development Goals hebben omarmd en die doelen ook worden gemonitord in bijvoorbeeld de monitor brede welvaart van het CBS, staat de complexe verwevenheid van maatschappelijke activiteiten bovenaan beleidsagenda's.

In het verleden is het CBS terughoudend geweest in uitspraken over (causale) mechanismen. Reden daarvoor is enerzijds het besef dat terwijl correlaties tussen meetbare grootheden relatief eenvoudig aan te tonen of juist te verwerpen zijn, causatie veel moeilijker te beoordelen is. Het CBS-werkterrein laat geïsoleerde of gecontroleerde experimenten niet toe en daarom valt met klassieke methoden alleen in heel uitzonderlijke omstandigheden uit te sluiten dat een correlatie tussen grootheden veroorzaakt is door een gemeenschappelijke, al dan niet waargenomen, oorzaak. Het wordt steeds gebruikelijker om een expliciet causaal model aan te nemen en dan de causale bijdrage te kwantificeren. Het is daarom noodzakelijk om moderne, geaccepteerde wetenschappelijke toetsingstechnieken te gebruiken die ontworpen zijn om hypothetische causaties te kunnen verwerpen, ook wanneer het niet mogelijk is om een volledig gecontroleerd en afgesloten experiment uit te voeren.

Anderzijds is terughoudendheid over causatie gemotiveerd door een groeiend besef dat een enkelvoudig verband tussen één of enkele oorzakelijke effecten voor een gegeven waargenomen fenomeen in de samenleving een oversimplificatie is van de werkelijkheid. De samenleving bestaat uit heel veel en heel diverse actoren en, in het bijzonder, uit de relaties en interacties daartussen. Daardoor zijn trends in de samenleving in feite collectieve veranderingen in het systeem van die relaties. Positieve en negatieve terugkoppeling en niet-lineaire wederzijdse beïnvloeding tussen actoren werken op een microscopisch niveau, en komen tot uitdrukking op macroscopisch niveau als een set van emergente fenomenen of trends. Dit betekent ook een verschuiving van inzicht omtrent de toestand van de economie of de samenleving. Waar in het verleden bij modellen uitgegaan werd van evenwichten, wordt duidelijk uit de complexiteitswetenschap dat ogenschijnlijke verbanden tussen variabelen helemaal geen uiting van evenwicht hoeven te zijn, en dat op een fundamenteel niveau het systeem van economie of samenleving wellicht in het geheel geen stabiel evenwicht heeft.

## Onderscheid tussen modelvalidatie en -toepassing

Het is belangrijk om op te merken dat er een scheiding is tussen de activiteiten van CBS en de planbureau's. Het CBS verzamelt en publiceert meetgegevens en schattingen die zijn gebaseerd op dergelijke gegevens aangevuld door statistische modellen. De planbureau's (het centraal planbureau (CPB), het sociaal en cultureel planbureau (SCP) en het planbureau voor de leefomgeving (PBL)) gebruiken voorspelmodellen, gecalibreerd op die gegevens, voor het doorrekenen van de (potentiële) consequenties van beleidskeuzes. Precies die activiteit er tussenin, calibreren of valideren van voorspelmodellen, moet gedaan worden op basis van verschijnselen en trends in het recente verleden. Voordat een model door bijvoorbeeld planbureaus of andere beleidsondersteuners wordt ingezet, is het nodig dat we ons ervan vergewissen dat ze minimaal in staat waren om die recente verschijnselen en trends te reproduceren. Het is juist hier dat er een rol is voor het CBS: gebruik maken van de enorme schat aan gegevens als tijdreeksen om in ieder geval onjuiste modellen of modelparameters te helpen uit te sluiten van verder gebruik.

Er valt over te debatteren waarom de scheiding tussen bijvoorbeeld CBS en CPB ooit is afgesproken (Van den Bogaard, 2000). Er is in ieder geval een inhoudelijke reden om die scheiding aan te brengen, zoals er ook een onderscheid wordt gemaakt tussen theoretische en experimentele fysica. In het geval van het onderscheid tussen CBS en CPB voorkomt die scheiding expliciet dat de indruk zou kunnen worden gewekt dat er een politieke kleuring is van de meetgegevens. De politieke keuzes die op Prinsjesdag in de landsbegroting worden gepresenteerd, worden door het CPB in modellen doorgerekend. Het materiaal van het CBS behoudt daarentegen expliciet zichtbaar een hogere mate van objectiviteit en onafhankelijkheid.

Dat betekent echter niet dat er binnen het CBS geen kennis is, of zou moeten zijn, over de economische of sociale modellen die door economen en wetenschappers zijn opgesteld. Om die nieuwe modelvalidatie-taak grondig uit te kunnen voeren, is het vanzelfsprekend dat het CBS ook zelf die modellen kan implementeren en bekend is met alle details en aannames.

Daarnaast is het niet alleen van belang dat het CBS deze taak invult, ten behoeve van bestuur en samenleving. Het is ook van belang om te onderzoeken welke

'metingen' aan de samenleving nodig zijn in de toekomst, die het meest gevoelig en nauwkeurig zijn om de cruciale parameters van die complexe modellen vast te kunnen leggen. Het is namelijk niet zeker dat nu al alle meest toepasselijke gegevens worden verzameld. De eerste resultaten van dergelijk werk zijn ook al gepubliceerd. Zie bijvoorbeeld het onderzoek aan personen-netwerken, in het bijzonder dat van ouderen (Das & De Jonge, 2020), en onderzoek aan productienetwerken (Mattson et al., 2021).

De officiële statistiek zou, naast de huidige rol, ook gezien kunnen worden als de empirische poot van complexiteitswetenschap op sociaal en economisch vlak. De Sustainable Development Goals, klimaatverandering, en vergrijzing zijn allemaal emergente fenomenen: uitdrukkingen van complexe systemen. Deze verschuiving van een paradigma van evenwicht naar dynamiek moet ook bij het CBS terugkomen in de vorm van nieuwe statistieken en nieuwe rapportages over de toestand van Nederland. Alleen dan blijft het CBS een steunpilaar voor onderbouwd beleid.

De standpunten in dit artikel zijn die van de auteur en weerspiegelen niet noodzakelijk het beleid van het CBS.

### LITERATUUR

Bogaard, A. van den. (2000). Het CPB, wiskunde en praktijk in wording. *Nieuw Archief voor Wiskunde (NAW)*, 5(1), 294–300. <http://www.nieuwarchief.nl/series/pdf/naw5-2000-01-3-294.pdf>.

Das, M., & Jonge, E. de. (2020). *Zelfredzaamheid van ouderen en gebruik van Wmo*. CBS. <https://www.cbs.nl/nl-nl/long-read/statistische-trends/2020/zelfredzaamheid-van-ouderen-en-gebruik-van-wmo>.

Mattson, C.E.S., Takes, F.W., Heemskerk, E.M., Diks, C., Buiten, G., Faber, A., & Sloot, P.M.A. (2021). Functional structure in production networks. *Frontiers in Big Data*, 23(4), <https://doi.org/10.3389/fdata.2021.666712>.

Pijpers, F.P. (2020). Methodology Research Vision 2020-2025. <https://www.cbs.nl/nl-nl/achtergrond/2020/25/visie-methodologie-onderzoek-2020-2025>.

FRANK P. PIJPERS is in 1991 gepromoveerd als sterrenkundige, en is enkele decennia als wetenschappelijk onderzoeker op dat gebied aan diverse internationale instituten verbonden geweest. Hij is sinds 2010 werkzaam bij het CBS, en heeft sinds begin 2021 een aanstelling als bijzonder hoogleraar op het gebied van complexiteit voor officiële statistiek bij het Korteweg-de Vries instituut voor wiskunde van de UvA. Hij is de auteur van de monografie *Methods in helio- and asteroisemology* (2006).

E-mail: [f.pijpers@cbs.nl](mailto:f.pijpers@cbs.nl)

## ETHEL NEWBOLD PRIZE 2021 VOOR MARLOES MAATHUIS



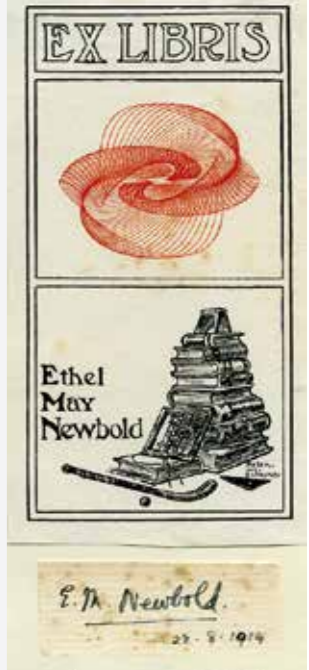
Marloes Maathuis, die de Van Dantzig Prijs 2020 kreeg, heeft weer een belangrijke onderscheiding gekregen. Tijdens het ISI World Statistics Congress in juli 2021 ontving ze de tweejaarlijkse Ethel Newbold Prize.

Deze door Wiley gesponsorde prijs wordt toegekend namens de Bernoulli Society for Mathematical Statistics, een van de Associations van het ISI. Hoewel de naam de historisch belangrijke rol van vrouwen in de statistiek benadrukt, wordt de prijs niet beperkt tot vrouwen, maar staat open voor iedereen.

Twee citaten uit de toelichting:

“The Prize is awarded biennially to an outstanding statistical scientist for a body of work that represents excellence in research in mathematical statistics, and/or excellence in research that links developments in a substantive field to new advances in statistics.”

“Marloes Maathuis is an outstanding mid-career scientist with a superb record of excellent research work in mathematical statistics. She has contributed pioneering work in the areas of causal inference, graphical models, and machine learning, with significant applications to genetics, HIV-AIDS, and imaging data in psychiatry. Her work has greatly advanced the field.”



## Ethel Newbold

Ethel May Newbold (1882 – 1933) was an English statistician and the first woman to be awarded the Guy Medal in Silver by the Royal Statistical Society, in 1928. During her short academic career (1921 – 1930) she published 17 papers in statistics and subject matter journals.

After obtaining her undergraduate degree from Cambridge University, she taught school for two years, and then worked for the Ministry of Munitions from 1919 – 1929, which is where her interest in statistics developed. She obtained her MSc and PhD from the University of London in 1926 and 1929, respectively.

Most of her published work was undertaken when she was a member of the National Institute of Medical Research, as the member of a committee appointed by the Medical Research Council to co-ordinate and supervise medical and industrial statistical inquiries. The Guy Medal was awarded for her paper “Practical applications of statistics of repeated events, particularly to industrial accidents” (Newbold, 1927), which was the first to give a theoretical treatment of compound Poisson distributions, for the analysis of accident data in industry.



## Riddler Solitaire

Riddler is een puzzelcolumn in het befaamde blog *FiveThirtyEight* van Nate Silver. Elke week worden in deze column problemen gegeven die liggen op het terrein van de wiskunde, logica en kansrekening. Twee puzzels worden elke week gepresenteerd: de meer eenvoudige Riddler Express en de meer uitdagende Riddler Classic. Overmoedig geworden door het feit dat eerdere puzzels van mij – overigens ontleend aan mijn columns in *STATOR* – waren verschenen in Riddler 538, stuurde ik onlangs een puzzel op die onverwacht een groot succes werd met verrassende oplossingen van adepten van Riddler 538. Dit wil ik de trouwe lezers van mijn *STATOR*-column niet onthouden. De puzzel die verscheen als Riddler Classic luidde als volgt.

Riddler solitaire is played with 11 cards: an ace, a two, a three, a four, a five, a six, a seven, an eight, a nine, a 10 and a joker. Each card is worth its face value in points, while the ace counts for 1 point. To play a game, you shuffle the cards so they are randomly ordered, and then turn them over one by one. You start with 0 points, and as you flip over each card your score increases by that card's points, as long as the joker hasn't shown up. The moment the joker appears, the game is over and your score is 0. The key is that you can stop any moment and walk away with a nonzero score.

What strategy maximizes your expected number of points?

Extra credit: With an optimal strategy, how many points would you earn on average in a game of Riddler solitaire?

De bepaling van de optimale stopregel is niet erg lastig, dit in tegenstelling tot de bepaling van de maximale gemiddelde score per spel. Veel inzenders kwamen op het idee om te bekijken wat het effect is op de huidige score zonder joker wanneer niet gestopt wordt maar nog één kaart wordt gepakt. De redenering gaat ruwweg als volgt. Stel dat je huidige score  $p$  punten is en dat je tot nu  $c$  kaarten hebt gepakt. In deze situatie zijn er nog  $10-c$  niet-joker kaarten in het spel die een totale waarde van  $55-p$  punten hebben, waarbij elk van deze kaarten een gemiddelde waarde van  $(55-p)/(10-c)$  punten heeft. De volgende kaart die je pakt is dan de joker met kans  $1/(11-c)$  en is een niet-joker kaart met kans  $(10-c)/(11-c)$ . Dus de verwachte toename van je huidige score van  $p$  punten is dan

$$T = \frac{10-c}{11-c} \times \frac{55-p}{10-c} = \frac{55-p}{11-c},$$

terwijl de verwachte afname van je huidige score gelijk is aan

$$A = \frac{p}{11-c}.$$

De verwachte toename  $T$  is alleen dan groter dan de verwachte afname  $A$  als  $55-p > p$ , oftewel als  $p < 27,5$ . Dit leidt tot de conclusie om alleen door te gaan als je huidige score minder dan 28 punten is en te stoppen zodra je score 28 of meer punten is. Dit is inderdaad de optimale stopregel zoals met Markov beslissingstheorie voor optimale stopproblemen kan worden bewezen (de zogeheten *one-stage-look-ahead rule* is optimaal voor stopproblemen met de eigenschap dat het proces in de verzameling van ongunstige toestanden blijft als het daar eenmaal in beeld is). Meer algemeen, voor het geval van  $N$  niet-joker kaarten met respectievelijke waarden 1 tot en met  $N$  geeft

bovenstaande redenering dat het optimaal is om te stoppen zodra je score  $\frac{1}{4} N(N+1)$  of meer is en anders door te gaan. Opmerkelijk is dat de optimale strategie alleen afhangt van het aantal punten dat al vergaard is en niet van het aantal kaarten dat al getrokken is.

Veel inzenders van een oplossing voor de puzzel gebruikten Monte Carlo simulatie om de maximale gemiddelde score per spel te vinden en sommigen vonden daarbij het verrassende en opmerkelijke resultaat dat onder de optimale stopregel de kans om de joker te pakken gelijk is aan 50%. Eén inzender toonde aan dat deze kans inderdaad precies gelijk is aan 50% door op slimme wijze de computer alle mogelijke permutaties van de 11 kaarten te laten doorlopen en zo de kansverdeling van de eindscore onder de optimale stopregel te berekenen. Waar ik zelf dacht aan een dynamische programmering recursie met een multi-dimensionale toestand als alternatief voor simulatie, was er ook een inzending met een andere aanpak om zonder gebruik van simulatie de maximale gemiddelde score per spel te bepalen. Daartoe werd voor het geval van 11 kaarten eerst opgemerkt dat voor de optimale stopregel met een stopwaarde van 28 punten tenminste 4 kaarten en ten hoogste 7 kaarten nodig zijn, waarbij de eindscore niet meer dan 37 (=27+10) punten kan zijn. De verwachtingswaarde kan dan vervolgens berekend worden met de formule

$$\sum_{c=4}^7 \sum_{p=28}^{37} p \frac{a_c(p)}{\binom{10}{c}} \frac{11-c}{11},$$

waarbij  $a_c(p)$  wordt gegeven door het aantal combinaties van  $c$  verschillende niet-joker kaarten zodat de som van de waarden van de kaarten gelijk aan  $p$  is en minder dan 28 is als één van deze  $c$  kaarten verwijderd zou worden. De waarden van de  $a_c(p)$  werden door aftelling bepaald. Dit leidde tot de verwachtingswaarde 15,453 voor de eindscore onder de optimale stopregel (de standaarddeviatie van de eindscore is dan 15,547).

Ten slotte nog enkele opmerkingen over de uitdagende

puzzel waarop zo veel interessante reacties binnenkwamen. Sommige inzenders pasten een combinatie van Monte Carlo simulatie en regressieanalyse toe. Zo werd voor het geval van 11 kaarten de kans op een eindscore nul als functie van het stoppunt  $s$  benaderd door de regressielijn

$$0,01587s + 0,0556$$

en werd voor het algemene geval van  $N$  niet-joker kaarten met respectievelijke waarden 1 tot en met  $N$  de maximale verwachtingswaarde van de eindscore benaderd door de regressiecurve

$$0,1251N^2 + 0,2879N + 0,061.$$

met de respectievelijke benaderingswaarden 0,49996 voor  $s=28$  en 15,45 voor  $N=10$ .

De verrassende *trade-off* tussen de kans van 50% op een eindscore nul en de verwachtingswaarde van de eindscore onder de optimale stopregel lijkt ook van toepassing te zijn voor het algemene geval van  $N$  kaarten (voor  $N=2$  en  $N=3$  is de bewering simpel te verifiëren). Het zou interessant zijn dit verder te onderzoeken. Tot nu toe is verondersteld dat er één joker onder de kaarten is. Wat zijn de resultaten bij twee jokers? Het is dan simpel na te gaan dat het gemiddeld aantal punten per spel maximaal is door te stoppen zodra het aantal vergaarde punten  $1/6 N(N+1)$  of meer is als  $N$  het aantal niet-joker kaarten is met waarden 1 tot en met  $N$ . Maar hoe zit het met de andere resultaten? Wat gebeurt als er onder de kaarten meerdere kaarten zijn met dezelfde waarde? Vragen te over. Leuke werkstukken voor studenten.

HENK TIJMS is emeritus-hoogleraar operations research aan de Vrije Universiteit en auteur van diverse leerboeken over operations research en kansrekening.  
E-mail: h.c.tijms@xs4all.nl



# THE LORD OF THE RINGS

## De impact van sentiment op netwerkvorming

In dit artikel willen we laten zien dat er verband is tussen de dynamiek van een sociaal netwerk en de sentimentscore van de actoren in dat netwerk. Hiervoor gebruiken we de gegevens uit het filmscript van *The Lord of the Rings: The Fellowship of the Ring* die we vergelijken met wat er bekend is over het verhaal. De sentimentscore van het filmscript wordt verkregen met behulp van *machine learning*. Deze sentimentscores dienen vervolgens als input voor een sociale netwerkanalyse met behulp van het stochastische actor-gebaseerde model voor netwerkdynamiek, genaamd SIENA. Dit model schat de waarden van de parameters door een iteratief stochastisch simulatiealgoritme toe te passen. De uitkomst laat zien dat het effect van sentiment positief en significant is.

JADÉ DIETEREN & DEMI DE KORT

De wereld bestaat uit talloze complexe netwerksystemen die een belangrijk doel dienen in onze samenleving. Aan de hand van het sociale netwerk uit de film *The Lord of the Rings: The Fellowship of the Ring* tonen we aan dat er een verband is tussen de dynamiek van een sociaal netwerk en de sentimentscore van de actoren in dat netwerk.<sup>1,2</sup> We maken gebruik van een stochastisch actor-gebaseerd model voor netwerkdynamiek, genaamd SIENA. Dit model maakt de statistische evaluatie van de netwerkeffecten mogelijk. Uit het filmscript kan ook het sentiment van de personages worden gehaald door gebruik te maken van *machine learning*-technieken. We onderzoeken vervolgens of de sentimentscore van een actor in een sociaal netwerk de dynamiek van de netwerkstructuur significant beïnvloedt.

Hoewel de dynamiek van sociale netwerken en sentimentanalyse zeer inzichtelijk kan zijn, zijn er nog steeds uitdagingen. Menselijke taal kan bijvoorbeeld moeilijk te begrijpen en zeer contextueel zijn. Het unieke aan het gebruik van een filmscript is dat elke beweging van elk personage bekend is. *The Lord of the Rings* is daarom een ideale testcasus voor netwerkanalyse, omdat we de mogelijkheid hebben om de modelresultaten op basis van data te vergelijken met de kennis die we hebben over het verhaal.

### Sentimentscore

Sentimentanalyse is een toepassing van Natural Language Processing (NLP) en wordt gedefinieerd als de inter-

pretatie van teksten om de emotionele toon achter een reeks woorden te vinden. We vergeleken de prestaties van twee *supervised learning*-modellen en één *unsupervised learning*-model dat gebruikt maakt van een lexicon om de scènes te analyseren.

Het eerste model maakt gebruik van gegevens die op Twitter zijn verzameld, het tweede model gebruikt gegevens uit filmrecensies om zichzelf te trainen. Het laatste en best presterende model op basis van handmatige klassificering, is het AFINN Lexicon Based Model. De huidige versie van het lexicon bevat 2.477 woorden die elk een polariteitsscore tussen -5 en 5 hebben. Hoe hoger de score, hoe meer positiviteit aan dat woord wordt geassocieerd. De cumulatieve score van alle woorden is de score voor een zin of voor een hele scène. De output van dit model is een numerieke score. Hierdoor kan onderscheid worden gemaakt tussen zeer positieve scènes en licht positieve scènes.

### Simulatie voor sociale netwerken

Een sociaal netwerk is een netwerk van sociale interacties en persoonlijke relaties. Stel dat er een sociaal netwerk is dat bestaat uit  $n$  actoren. Dit netwerk kan dan worden weergegeven door een  $n$  bij  $n$  matrix  $X = (x_{ij})$  waarbij  $x_{ij} = 1$  als er een verbinding tussen actor  $i$  en  $j$  aanwezig is en anders  $x_{ij} = 0$ .

*Network Dynamics* is een onderzoeksgebied dat zich richt op netwerken die zich in de loop van de tijd ontwikkelen. Netwerkevolutie wordt gedreven door bepaal-

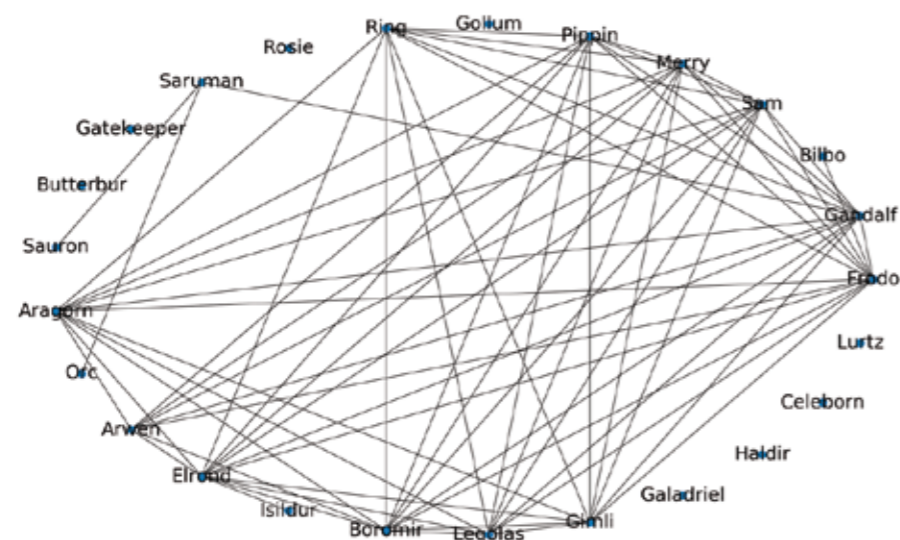


de netwerkeffecten. Ripley et al. (2011) geeft een gedetailleerde uitleg over hoe verschillende effecten worden gedefinieerd. Enkele prominente effecten zijn:

- Transitiviteit ontstaat als er een band bestaat tussen persoon A en persoon B, en weer een band tussen persoon B en persoon C, en dan ontstaat een band tussen persoon A en persoon C;
- Homofilie is de neiging om zich te verhouden tot mensen met vergelijkbare kenmerken;
- *Outdegree* is het aantal uitgaande banden dat een persoon heeft.

Voor onze analyse werken we met SIENA (Simulation Investigation for Empirical Network Analysis), een programma voor de statistische analyse van sociale netwerkgegevens. Het doel van het model is om de netwerkdynamiek weer te geven op basis van waargenomen longitudinale gegevens, en deze te evalueren volgens het paradigma van statistische inferentie (Snijders et al., 2010). Het model gaat ervan uit dat de onderliggende tijdparameter continu is en dat het veranderende netwerk het resultaat is van een Markov-proces. Daarnaast wordt aangenomen dat de actoren hun uitgaande verbindingen beheersen en dat op een gegeven moment één actor de kans krijgt om een uitgaande verbinding te veranderen.

Homofilie wordt gedreven door covariaten, die onafhankelijke variabelen in het model zijn. We zullen het sentiment van de actoren koppelen aan de sociale netwerkanalyse via een in de tijd variërende covariaat.



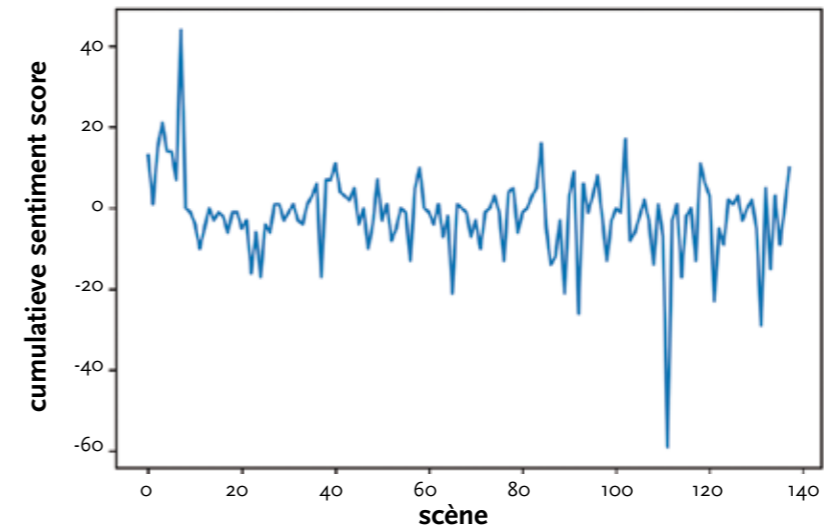
Figuur 1. Grafiek van een scènegroep; het Netwerk bevat de belangrijkste personages van het verhaal, die met elkaar verbonden zijn als ze samen in een scène voorkomen

De actor  $i$  die de kans krijgt om een verbinding te veranderen in netwerk  $x$ , heeft de mogelijkheid om niets te doen of één uitgaande connectie te veranderen. De kans voor alle mogelijkheden hangt af van de *objective function* (Snijders et al, 2010), gedefinieerd als:

$$\frac{\exp(f_i(\beta, x))}{\sum_{x' \in C} \exp(f_i(\beta, x'))}$$

waarbij  $f_i(\beta, x) = \sum_k \beta_k S_{ik}(x)$ .  $S_{ik}(x)$  zijn de bovengenoemde effecten en de statistische parameters. Om de parameterwaarden  $\beta_k$  voor elk van de  $k$  effecten te vinden, wordt een iteratief stochastisch simulatiealgoritme toegepast. Een groot aantal simulaties wordt herhaald om de netwerkevolutie na te bootsen tussen twee opeenvolgende observatiemomenten van het netwerk. De schattingsprocedure die in SIENA wordt gebruikt is de Method of Moments (MoM).

Het algoritme kent drie fasen waarin verschillende stappen worden doorlopen om de hyperparameters van het model te bepalen. In de eerste fase wordt een eerste ruwe schatting van de matrix van derivaten gemaakt (Ripley et al., 2011). Het doel van de tweede fase is om de parameterwaarden te vinden waarvoor de afwijkingen tussen gegenereerde en waargenomen waarden gemiddeld nul zijn. Dit resulteert in een uiteindelijke waarde voor de parametervector  $\beta_k$ . De derde fase produceert schattingen voor de covariantiematrix en de matrix van derivaten. Deze kunnen worden gebruikt om de standaardfouten van de parameterwaarden te berekenen.



Figuur 2. Sentimentscore van het AFINN lexicon model

### Case study: Lord of the Rings

Om de netwerkanalyse uit te voeren, zijn de scènes in het filmscript van *The Lord of the Rings: The Fellowship of the Ring* verdeeld in vijf groepen (zie figuur 1 voor scènegroep 3). De belangrijkste locaties van de vijf sociale netwerken zijn de Shire, Bree, Rivendell, Moria en Lothlórien. In deze netwerken werd aangenomen dat personages als ze samen in een scène voorkomen een connectie hebben.

De sentimentscore van het AFINN lexicon model voor elke scène is te zien in figuur 2. Scène 8 krijgt een score van 43 en is de enige scène met een sentimentscore boven de 20. In deze scène vindt Bilbo's verjaardagsfeest plaats. Het is een vrolijke gebeurtenis, waar vuurwerk wordt afgestoken en mensen dansen. De laagste senti-

mentscore wordt behaald in scène 112. In deze scène, die zich afspeelt in de Mines of Moria, stoot Pippin een scheidel om die naar beneden valt en veel lawaai maakt. Door dit te doen, weten de Orks waar de groep zich in de mijn bevindt en komen ze hen aanvalen. De scène is gevuld met veel angst, gevolgd door een felle en gewelddadige strijd. Gemiddeld is het sentiment van de film negatief, namelijk -1,73. Dit past bij de sombere sfeer van de film.

Het SIENA-model moest in totaal twee keer worden uitgevoerd voordat de gewenste convergentie werd verkregen. De parameterschattingen voor de effecten zijn te vinden in tabel 1.

De parameters in tabel 1 kunnen worden geïnterpreteerd met de evaluatiefunctie:

$$f_i = \sum_{j=1}^6 \beta_k s_{ik}(x) ,$$

		$\beta$	S.E.	$t$	$p$
1	Outdegree	-8,59	1,86	0,10	0,0000
2	Evenwicht	0,44	0,15	-0,01	0,0037
3	Transitiviteit	5,96	1,70	0,11	0,0005
4	4-cycles	-0,26	0,11	0,11	0,0147
5	Populariteit	0,59	0,22	0,11	0,0061
6	Sentimentovereenkomst	1,88	0,86	0,08	0,0292

Tabel 1. Selectie van geschatte waarden

waarbij  $k$  overeenkomt met de effecten. Dus via  $f_i$  hebben we een proxy van de verdeling over de mogelijke acties die actor  $i$  kan ondernemen. Een directe vergelijking van de waarden van parameterschattingen is mogelijk, maar houd er rekening mee dat dit niet-gestandaardiseerde coëfficiënten zijn (Ripley et al., 2011).

De parameterschatting heeft de grootste absolute waarde voor het *outdegree*-effect. De negatieve schatting kan worden verklaard door het feit dat het een inspanning kost om vriendschappen en connecties te onderhouden. Het effect is significant, zoals blijkt uit de  $p$ -waarde. Uit het resultaat blijkt ook dat er sprake is van een hoge mate van transitiviteit. Dit geeft aan dat het principe 'vrienden van mijn vrienden zijn mijn vrienden' in dit netwerk op zijn plaats is.

De parameter voor sentimentovereenkomst is positief en significant. Dit betekent dat actoren met sentiment scores die vergelijkbaar zijn, meer kans hebben om een verbinding met elkaar aan te gaan. Aangezien dit sociale netwerk is geconstrueerd op basis van personages die samen in een filmscène zitten, zouden ze bovendien de positieve of negatieve ervaring delen en daarom waarschijnlijk een vergelijkbare sentiment score hebben.

De parameterschattingen voor evenwicht en populariteit zijn laag, maar significant. Voor evenwicht betekent dit dat er een lichte voorkeur is voor de personages om een connectie te hebben met andere personages die veel van dezelfde connecties hebben als ze al hebben. De positieve schatting van de populariteit wijst op een lichte voorkeur om een connectie te maken met populaire personages. De reden hiervoor kan zijn dat de hoofdpersonen in de film het meest voorkomen in scènes en daarom als populair kunnen worden beschouwd. Omdat het verhaal om hen draait, creëren ze ook meer nieuwe banden. De schatting voor 4-cycli is negatief, wat wijst op een neiging om niet uit clusters van grootte vier te komen.

## Conclusie

In dit artikel zijn gegevens uit de film *The Lord of the Rings: The Fellowship of the Ring* gebruikt om het sociale netwerk en sentiment scores van de personages te extraheren. Voor de extractie van het karaktersentiment werd

gebruik gemaakt van het AFINN Lexicon Based Model. Vervolgens werd een sociale netwerkanalyse uitgevoerd met behulp van SIENA. We vonden een positieve parameterwaarde voor sentiment die significant verschilt van nul.

Meestal kwamen onze bevindingen met behulp van de modellen overeen met wat we weten over het verhaal. Er zijn echter nog steeds uitdagingen bij het toepassen van sentimentanalyse en sociale-netwerkanalyse. Bij sociologisch empirisch onderzoek is het daarom noodzakelijk om de resultaten van de analyse te interpreteren en na te denken over wat de data-analyses je wel en niet kunnen laten zien. We weten bijvoorbeeld dat de film om de ring draait. Het is echter niet eenvoudig om hiermee rekening te houden bij het uitvoeren van de analyse.

Voor toekomstig onderzoek kan dezelfde analyse worden uitgevoerd met andere filmscripts uit verschillende genres. Bovendien zou dit onderzoek kunnen worden uitgevoerd met behulp van echt bestaande sociale netwerken. In het echte leven zal het echter moeilijk zijn om gegevens te verzamelen. Hoewel de filmgegevens erg waardevol zijn omdat er geen ontbrekende gegevens zijn en het mogelijk is om resultaten te valideren, weerspiegelt het niet helemaal een situatie in het echte leven.

### NOOT

1. Dit artikel is gebaseerd op de masterscriptie van Demi de Kort voor haar studie Econometrie & Operations.
2. Het volledige filmscript is te vinden op <https://www.imsdb.com/scripts/Lord-of-the-Rings-Fellowship-of-the-Ring-The.html>.

### LITERATUUR

- Snijders, T. A. B., van de Bunt, G. G., & Steglich, C. E. G. (2010). Introduction to stochastic actor-based models for network dynamics. *Social Networks*, 32, 44–60.
- Ripley, R. M., Snijders, T. A. B., Boda, Z., Vörös, A., & Preciado, P. (2020). *Manual for SIENA version 4.0*. University of Oxford, Department of Statistics; Nuffield College.

JADÉ DIETEREN is forecasting business analyst bij Amazon en PhD student aan de Vrije Universiteit.  
E-mail: [jjdd@amazon.cz](mailto:jjdd@amazon.cz)

DEMI DE KORT is data scientist bij izi. Met hun analyses willen de data scientists van izi bijdragen aan een betaalbare en doelmatige zorg in Nederland.  
E-mail: [demidekort@hotmail.com](mailto:demidekort@hotmail.com)



# COMBINATORISCHE OPTIMALISATIE EN QUANTUM COMPUTING

De Quantum Computer komt met rase schreden dichterbij. Eén van de veelbelovende toepassingsgebieden is optimalisatie. Voor Operations Research practitioners is het van belang deze ontwikkeling te volgen, aangezien ze binnen een aantal jaren quantum algoritmes aan hun gereedschapskist zullen kunnen toevoegen. In dit artikel bespreken we twee manieren waarop optimalisatieproblemen kunnen worden aangepakt met de Quantum Computer.

FRANK PHILLIPSON

Een nieuwe computer die in slechts enkele seconden het antwoord op vragen kan geven waar de huidige generatie computers vele jaren mee bezig zou zijn. Dat is wat er beloofd wordt door de wetenschappers die, onder anderen

in Delft, werken aan de Quantum Computer (QC). Deze QC maakt gebruik van kwantummechanische verschijnselen, zoals superpositie en verstrengeling, om bewerkingen uit te voeren.

Superpositie is de bekendste eigenschap die gebruikt wordt door de QC. Waar klassieke computers vereisen dat de gegevens worden gecodeerd in binaire cijfers (bits), die elk altijd in een van de twee toestanden (0 of 1) zijn, gebruikt de QC quantum bits, qubits genaamd, die zich schijnbaar in beide toestanden tegelijkertijd kunnen bevinden. Deze staat van een qubit heet superpositie. Op het moment dat de positie van de qubit wordt waargenomen zal hij, schijnbaar willekeurig, zich echter voordoen als een 0 of 1. De andere eigenschap die gebruikt wordt is verstrengeling. Hierbij zijn de toestanden van twee deeltjes, die elk op een andere plaats kunnen zijn, aan elkaar gerelateerd.

De QC zal niet het antwoord zijn op alle problemen. In complexiteitstermen zal er een nieuwe groep problemen ontstaan, BQP, Bounded-error Quantum Polynomial time, een uitbreiding van P, met problemen die door de Quantum Computer kunnen worden opgelost, met ten hoogste 1/3 foutkans. De kans van 1/3 is in de definitie arbitrair gekozen. Eén van de bekendste problemen die in BQP zit is integer factorisatie, de reden dat verschillende soorten cryptografie niet meer de beloofde veiligheid zullen bieden.

Qua ontwikkeling staan we met de QC nog maar aan het begin. De huidige generatie heeft nog maar een beperkt aantal qubits die ook nog eens een beperkte levensduur hebben en beperkt stabiel zijn. Voor deze generatie, de Noisy Intermediate-Scale Quantum (NISQ) Computers, wordt wel al hard aan toepassingen gewerkt. De, naar verwachting, belangrijkste toepassingsgebieden waar de QC al meerwaarde kan hebben boven klassieke computing zijn chemische simulaties, machine learning en optimalisatievraagstukken.

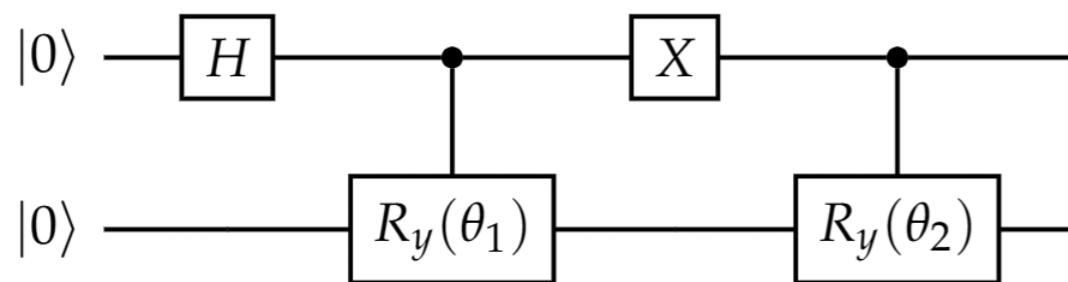
## Twee soorten

We kunnen in quantum computing een tweetal paradigma's onderscheiden: Gate-Based Computers (GBC) en Quantum Annealers (QA). GBC lijken qua werking het meest op de huidige generatie van computers; ze zijn in staat bewerkingen (gate-operaties, zoals AND, OR) uit te voeren op specifieke qubits of op meerdere qubits tegelijkertijd. Hiermee kan dus daadwerkelijk geprogrammeerd worden, wat vaak wordt gevisualiseerd via circuitdiagrammen, zoals in figuur 1. Hier staat elke lijn voor een qubit en loopt de tijd van links naar rechts, waarbij de operaties, gates, als blokken worden weergegeven.

De QA daarentegen zijn single-purpose machines. Quantum Annealing begon met het werk van Kadowaki en Nishimori (Kadowaki et al. 1998). QA kunnen in principe maar één ding: het vinden van een minimumwaarde van een specifieke functie. Deze functie wordt gecodeerd in de qubits, waarna een kwantummechanische evolutie zal leiden tot een oplossing die de energie minimaliseert.

## De QUBO

Deze specifieke functie wordt gerepresenteerd door een Ising of QUBO probleem. Een Ising probleem definieert en minimaliseert de energie van elektron-spins via een Hamiltonian. Een QUBO, Quadratic Unconstrained Binary Optimisation, probleem is de binaire representatie hiervan en wordt veelal voor



Figuur 1. Voorbeeld van quantum circuit

combinatorische optimalisatie gebruikt:

QUBO: minimaliseer  $x^T Q x$

waarbij  $x$  een vector is met binaire beslisvariabelen en  $Q$  een vierkante matrix met constante waarden. Veel (constrained) combinatorische optimalisatieproblemen kunnen eenvoudig worden omgeschreven naar een QUBO formulering (Glover et al., 2018a). Neem bijvoorbeeld het probleem

$$\min c^T x, \text{ onder de voorwaarden } Ax = b,$$

dan kunnen we voorwaarden in de doelfunctie onderbrengen met gebruik van een penalty  $\lambda$ :

$$\min c^T x + \lambda (Ax - b)^T (Ax - b).$$

Als we nu  $P=Ic$  definiëren als de matrix met de waarden van  $c$  op de diagonaal, dan is dit identiek aan:

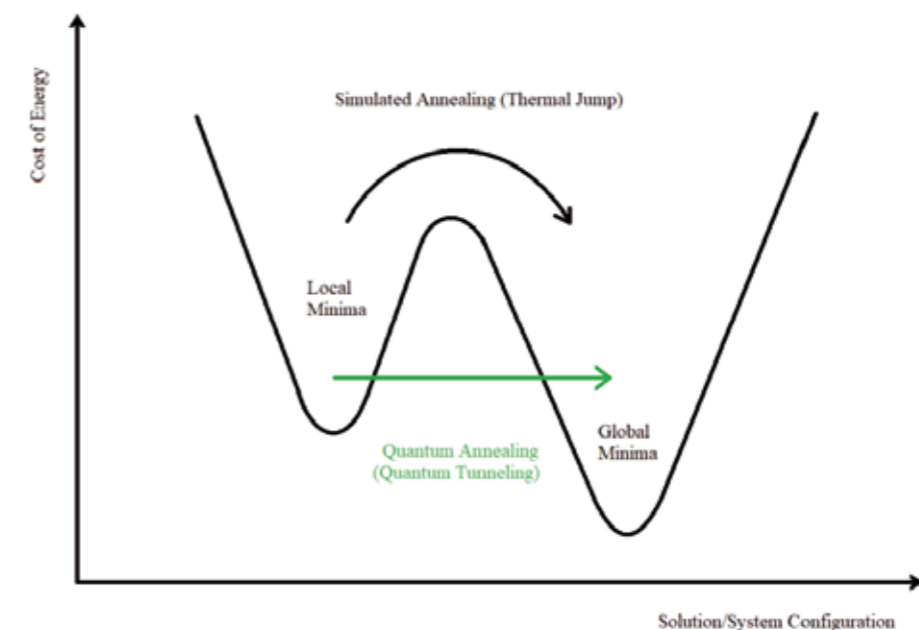
$$\min x^T P x + \lambda (Ax - b)^T (Ax - b) = x^T P x + x^T R x + d = x^T Q x,$$

waar de matrix  $Q$  en de constante  $d$  een gevolg zijn van de vermenigvuldiging en de constante  $d$  verwaarloosd kan worden. De matrix  $Q$  is nu afhankelijk van de penalty  $\lambda$ . De QUBO kan vervolgens in een pre-processing stap worden gereduceerd (Glover et al., 2018b).

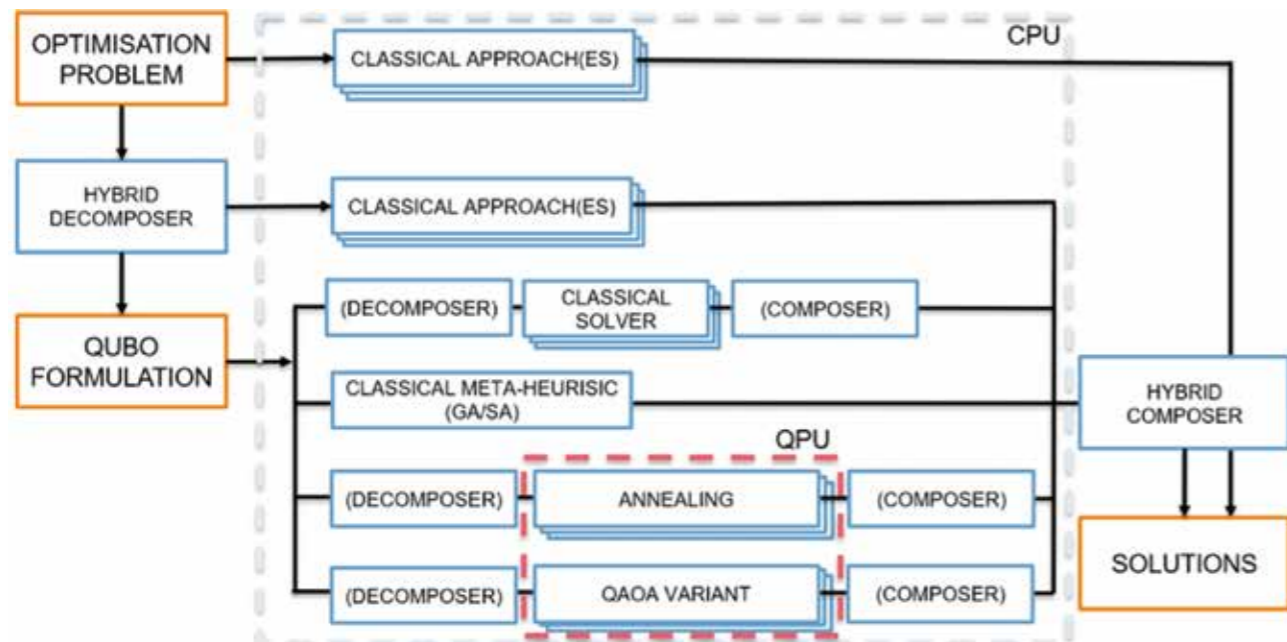
## Quantum Annealing

De QUBO wordt als input meegegeven aan de QA. Het uitgangspunt hier is om een gelijke superpositie te creëren over alle mogelijke toestanden. Vervolgens wordt een probleemspecifiek magnetisch veld ingeschakeld, waardoor de qubits met elkaar interacteren. Nu bewegen de qubits naar de staat met de laagste energie, waaruit de optimale oplossing afgeleid kan worden van het oorspronkelijke probleem. Quantum annealing heeft overeenkomsten met het bekende Simulated Annealing. Waar Simulated Annealing echter alleen thermische sprongen kan maken, gebruikt quantum annealing ook het quantum tunneling effect (zie figuur 2). De QA die al het verst gevorderd is, is de versie van D-Wave. Hun devices zijn praktische implementaties van adiabatische evolutie (Farhi et al. 2000). De evolutie van een quantumtoestand op de quantum processor van D-Wave wordt beschreven door een tijdsafhankelijke Hamiltonian ( $H(t)$ ), bestaande uit de oorspronkelijk Hamiltonian ( $H_0$ ), waarvan de grondtoestand gemakkelijk te creëren is, de gelijke superpositie, en de uiteindelijke Hamiltonian ( $H_1$ ), waarvan de grondtoestand codeert voor de oplossing van het huidige probleem, via de QUBO:

$$H(t) = \left(1 - \frac{t}{T}\right) H_0 + \frac{t}{T} H_1.$$



Figuur 2. Verschil Simulated Annealing en Quantum Annealing (Sharma et al. 2018)



Figuur 3. Voorbeeld optimalisatie pijplijn framework (TNO)

Dit systeem wordt geïnitieerd in de grondtoestand van de oorspronkelijke Hamiltonian, dat wil zeggen  $H(0) = H_0$ . De adiabatische stelling stelt dat als het systeem evolueert volgens de Schrödingervergelijking, en de minimale spectrale opening van  $H(t)$  niet nul is,  $H(T)$ , voor  $T$  groot genoeg, zal convergeren naar de grondtoestand van  $H_T$ . Hoewel we ons hier niet bezighouden met de technische details, is het goed te weten dat het meestal niet mogelijk is om een tijd  $T$  in te schatten om ervoor te zorgen dat het systeem altijd evolueert naar de gewenste staat. Voor enkele klassen aan problemen is de optimale annealing tijd experimenteel bepaald (Albash en Lidar 2018). Er is dan ook geen garantie van optimaliteit. Verder is het niet triviaal een goede waarde te vinden voor de penalty  $\lambda$  en de zogenaamde chainstrength; de QA is erg gevoelig voor deze parameters. Chainstrength is een parameter die een boete oplevert indien meerdere qubits die één variabele representeren niet dezelfde waarde hebben. Daarnaast geldt dat de state-of-the-art QA van D-Wave nu 5640 qubits heeft die een connectiviteit hebben van maximaal 15. Hiermee kan een clique probleem met  $n=70$  worden opgelost (McGeoch 2020). Voor problemen met een lagere afhankelijkheid tussen de variabelen, kunnen grotere problemen opgelost worden. Problemen die niet op de chip van deze grootte kunnen worden geplaatst, zullen in stukken moeten worden verdeeld, wat ten koste gaat van de kwaliteit van de oplossing. D-Wave biedt wel een standaard functie aan die deze decompositie automatisch verzorgt.

### QAOA

Het algoritme dat quantum annealing vertaalt naar de gate-based quantum computer is QAOA, Quantum Approximate Optimization Algorithm. Hierbij wordt de adiabatische evolutie benaderd door een gediscretiseerde functie die een benadering oplevert van de grondtoestand van de gewenste Hamiltonian. Het doel van het QAOA algoritme is om een zo laag mogelijke bovengrens te vinden voor de grondtoestand. Ook hier wordt gestart met een gelijke superpositie over alle oplossingen  $|s\rangle$ , een toestand in de zogenaamde bra-ket notatie. Het algoritme past vervolgens afwisselend de volgende twee bewerkingen, operators, toe op deze toestand:

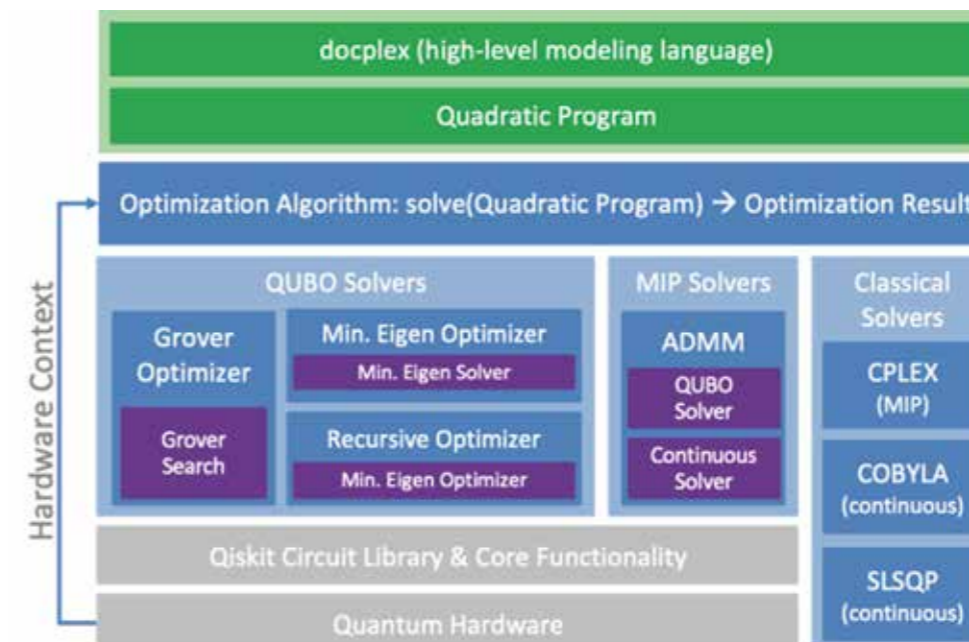
$$U(H_0, \beta) = e^{-i\beta H_0} \text{ en } U(H_1, \gamma) = e^{-i\gamma H_1},$$

waarbij  $\beta$  in  $[0, 2\pi]$  en  $\gamma$  in  $[0, 4\pi]$ . Als we deze operators  $p$  keer toepassen, dan produceert de QAOA de volgende quantum state:

$$|\gamma, \beta\rangle = U(H_0, \beta_p) U(H_1, \gamma_p) \dots U(H_0, \beta_1) U(H_1, \gamma_1) |s\rangle.$$

Met behulp van een QC kan vervolgens de verwachte waarde  $F_p = \langle \gamma, \beta | H_1 | \gamma, \beta \rangle$  berekend worden wat een bovengrens geeft voor de grondtoestand. Als  $p \rightarrow \infty$ , zal deze benadering, voor de juiste keus van hoeken  $\beta_p$  en  $\gamma_p$ , convergeren naar de optimale oplossing – de exacte grondtoestand van het onderliggende probleem. Het vinden van de juiste hoeken is echter niet triviaal en als  $p$  groot moet zijn voor een goede benadering is de effectiviteit van het algoritme nog een onzekere factor.

De QAOA heeft hierbij wat meer vrijheid dan QA.



Figuur 4. Voorbeeld IBM Qiskit's Optimization Module

Een andere keuze voor de zogenoemde *mixing operator*  $U(H_0, \beta)$ , de Quantum Alternating Operator Ansatz (Hadfield et al., 2019), kan zorgen voor een beperking van de zoekruimte, bijvoorbeeld bij harde *constraints*. De QAOA is een voorbeeld van een *variational of hybrid* algoritme. Hierbij worden klassieke parameters geoptimaliseerd om zo een quantum circuit te creëren dat een probleem oplost. Deze algoritmes worden gezien als het soort dat de eerste praktische toepassing zal hebben op de gated quantum computers. Het heeft immers maar een beperkte hoeveelheid qubits nodig en ook de diepte van het circuit (de tijdsfactor) is beperkt als  $p$  niet te groot wordt gekozen. Daarnaast zijn variational algoritmes redelijk resistent tegen de ruis, de beperkte stabiliteit, van de qubits.

### Conclusie

De gepresenteerde quantum algoritmes zijn feitelijk metaheuristische gecombineerd met specialistische hardware. De quantum computer is hier niet de enige mogelijkheid, ook photonic gebaseerde Ising machines zijn in opkomst (Van Vaerenbergh et al. 2021). Voor OR-practitioners geldt dat dit extra tools zullen zijn in hun gereedschapskist, zie bijvoorbeeld figuur 3 voor een voorbeeld hoe Quantum Computing in zo'n gereedschapskist kan passen. Ook zullen commerciële aanbieders de komende jaren hun aanbod gaan verrijken met Quantum algoritmes, zoals IBM, die naast CPLEX ook quantum tools aanbiedt (figuur 4) zonder dat gebruikers precies hoeven te weten wat er onder de motorkap gebeurt.

### LITERATUUR

Albash, T., & Lidar, D. A. (2018). Demonstration of a scaling advantage for a quantum annealer over simulated annealing. *Physical Review X*, 8(3), 031016.

Farhi, E., Goldstone, J., Gutmann, S., & Sipser, M. (2000). Quantum computation by adiabatic evolution. *arXiv:quant-ph/0001106*.

Glover, F., Kochenberger, G., & Du, Y. (2018a). A tutorial on formulating and using qubo models. *arXiv:1811.11538 [cs.DS]*.

Glover, F., Lewis, M., & Kochenberger, G. (2018b). Logical and inequality implications for reducing the size and difficulty of quadratic unconstrained binary optimization problems. *European Journal of Operational Research*, 265(3), 829–842.

Hadfield, S., Wang, Z., O’Gorman, B., Rieffel, E. G., Venturelli, D., & Biswas, R. (2019). From the quantum approximate optimization algorithm to a quantum alternating operator ansatz. *Algorithms*, 12(2), 34.

Kadowaki, T., & Nishimori, H. (1998). Quantum annealing in the transverse Ising model. *Physical Review E*, 58(5), 5355.

McGeoch, C., & Farré, P. (2020). *The D-Wave Advantage System: An Overview. Technical Report*. D-Wave.

Sharma, A. K., & Maharjan, P. (2018). *Quantum Annealing as an Optimized Simulated Annealing: A Case Study*. Conference paper 4th International IT Conference on ICT with smart computing.

Van Vaerenbergh, T. et al. (2021). Integrated coherent Ising machines for next-generation optimization accelerators. *Proceedings Volume 11703, AI and Optical Data Sciences II*. doi.org/10.1117/12.2585839.

FRANK PHILLIPSON is senior scientist bij TNO en bijzonder hoogleraar Operations Research aan Maastricht University. Naast het oplossen van optimalisatievraagstukken voor logistiek, telecommunicatie en militaire toepassingen is hij verantwoordelijk voor het onderzoek naar algoritmes voor en toepassingen van quantum computers op het gebied van optimalisatie en machine learning. E-mail: frank.phillipson@tno.nl



## ROBUUST KLIMAATBELEID

De invloed van klimaatverandering op de economie is nu al merkbaar en zal in de nabije toekomst steeds groter worden. Zelfs als we onze klimaatdoelen halen en de uitstoot van broeikasgassen tot netto nul reduceren, zal het effect van de ingezette klimaatverandering nog decennia lang aanhouden en tot een stijging van risico's leiden. Voor de overheid, verzekeraars, (centrale) banken en bedrijven is het belangrijk goed zicht op deze risico's te verkrijgen. Nog belangrijker is het verkrijgen van inzicht in de factoren die deze risico's positief of negatief beïnvloeden zodat robuust en adaptief beleid gevoerd kan worden. Echter de gebruikte modellen om deze factoren en risico's te analyseren schieten om meerdere redenen tekort en zijn daarmee van beperkte waarde bij het bepalen van klimaatrisico mitigerend beleid. Beter gebruik van data en empirische modellen, aangevuld met scenario denken levert een reëler beeld van de risico's en de bepalende factoren op, wat zal leiden tot een robuust en adaptief risico mitigerende klimaatbeleid.

### Geen 'echte' prijs

Om de ontwikkeling van de economie goed te kunnen

bestuderen en te beïnvloeden worden macro-economische modellen gebruikt. Centrale banken, het Centraal Planbureau, banken, verzekeraars en vermogensbeheerders gebruiken ze bij het vaststellen van hun beleid. Om de ontwikkeling van effectief economisch of financieel beleid goed te kunnen ondersteunen moeten deze modellen de meest relevante en te beïnvloeden aspecten van onze economie bevatten. Echter de gebruikte modellen missen essentiële componenten waardoor een verkeerd beeld van de oorzaken en invloed van klimaatrisico's ontstaat. Zo bevatten modellen als SAFFIER II van het CPB veelal alleen arbeid en kapitaal (=investeringen) als input voor productie en niet het verbruik van niet hernieuwbare grondstoffen zoals fossiele brandstoffen of schaarse metalen/mineralen. Ook de maatschappelijke impact van het vrijkomen van broeikasgassen en andere afvalstoffen is niet in deze modellen opgenomen. Als gevolg worden niet de echte kosten voor het verbruik van grondstoffen gebruikt en wordt de schade als gevolg van economische activiteit volkomen genegeerd. Deze omissie heeft geleid tot de klimaatverandering zoals we die nu al ervaren. Het leidt tevens tot een oneerlijke vergelijking met hernieuwbare alternatieven, waardoor investeringen in deze alternatieven achterblijven.

### Fundamenteel onzeker

Een tweede en misschien wel belangrijkere tekortkoming is dat de gebruikte modellen uitgaan van een stationaire, zich naar een evenwicht ontwikkelende, economie. Deze aanname vormt de basis van veel macro-economische modellen, zoals bijvoorbeeld SAFFIER II en commerciële macro-econometrische modellen. Echter empirisch onderzoek laat zien dat deze aanname onhoudbaar is. Als gevolg van continue beleidsinterventies, technologische veranderingen, crises en het veranderende gedrag van consumenten is geen sprake van een economie die steeds naar een evenwicht beweegt. De economie is niet stationair maar fundamenteel onzeker. Gevolg is dat het beleid gebaseerd op de veronderstelde stabiliteit compleet de plank zal misslaan, zoals pijnlijk duidelijk werd met de financiële crisis van 2007–2008. Klimaatverandering zal tot verdere verstoringen van de economie leiden als gevolg van veranderingen in consumentengedrag, de productie en het verbruik van energie, welke grondstoffen we gebruiken en hoe we dingen maken. Dogmatisch vasthouden aan een onvolledige economische theorie zal dan tot ineffectief beleid leiden om klimaatrisico's te mitigerend.

### Een gediversifieerde aanpak leidt tot robuuster beleid

Iedere economische theorie is onvolledig, onjuist en veranderlijk, zo is de aanname van stabiliteit een grove onderschatting van de complexiteit van de economie. Analyses met een theorie als enig uitgangspunt zullen daarom slechts van beperkte waarde zijn. Met data als empirische bewijsmateriaal kunnen we ons begrip van de economie beter onderbouwen. Deze manier van werken

levert ons de factoren die relevant zijn voor ons begrip van de impact van bijvoorbeeld klimaatverandering op de economie. Door data- én theorie-gedreven modellen te ontdekken, in plaats van de data te folteren totdat ze in het theoretische juiste maar incomplete hokje passen, komen we tot praktische bruikbare inzichten en houden we beter zicht op de op handen zijnde veranderingen.

Als gevolg van de fundamentele onzekerheid van de economie zal data ons alleen inzichten uit het verleden bieden en zullen modellen falen in het accuraat voorspellen van de toekomst. Modellen en data alleen zijn dus niet voldoende. Het zijn echter niet de enige instrumenten die ons ter beschikking staan. Goed risicomanagement ontstaat alleen als een rijk, gediversifieerd beeld van de onzekerheden beschikbaar is. Door data en modellen te gebruiken om toekomstscenario's te genereren kunnen we een deel van dat beeld invullen. Aanvullend kan scenario-denken ons helpen situaties voor te stellen, zoals sociale, technologische, ecologische of politieke verschuivingen, die geen onderdeel zijn van de data of de modelvoorspellingen. Als gevolg ontstaat een beter geïnformeerd beeld van mogelijke onvoorziene situaties en kunnen we, al dan niet geholpen door beslissingsondersteunende modellen, een effectiever klimaatbeleid formuleren, uitvoeren, evalueren en aanpassen.

Door scenario-denken te combineren met inzichten uit data en empirische modellen ontstaat een gediversifieerd beeld van de economische impact van klimaatrisico's en kan de effectiviteit van verschillende mitigerende maatregelen worden bepaald. We krijgen daardoor een reëler beeld van klimaatrisico's en zijn beter voorbereid om robuuste en adaptieve keuzes te maken in het aanpakken van de gevolgen van klimaatverandering.

JOHN POPPELAARS, Doing the Math  
E-mail: john@doingthemath.nl



# MONITORING THE FUTURE

LEO HUBERTS

When the heart stops pumping, the body is in trouble. Immediate action is required to reboot the heart and get the blood circulation flowing. Sometimes these actions are taken on time and a life is saved (e.g., football player Christian Eriksen at the European Championship in 2021), sometimes the actions are late or insufficient and have lasting consequences (e.g., football player Abdelhak Nouri of Ajax at an exhibition match in 2017). We cannot determine the problem by simply observing from the

outside. We do not see the heart rate and have only indirect hints towards blood flow (the color of the skin). Often there are no clear warning signs. The person seems fine, until they don't. Detecting physical problems such as a heart attack as quickly as possible requires monitoring of vital signs such as heart rate, blood pressure, and flow, oxygen levels, etc. The functioning of a body can be described as a complex process. The vital signs give us an indication of the overall state of that process.

## Monitoring the present

A large variety of processes are being monitored in industry and services. When we talk about monitoring a process, we generally mean closely watching the process in real time to find problems. The current state of the process is assessed to determine if intervening action is required. In factories the weight and dimensions of products are monitored, the temperatures of machines, and the production time. In banking, the location and amounts of transactions are monitored to detect fraud and theft, financial markets are monitored to mitigate risk and the duration of service inquiries is monitored to improve customer service. In healthcare we monitor the vital signs of patients, but also the length of stay to detect inefficiencies and medicine stock to optimize storage and drug availability. To support public policy, we monitor crime statistics, family incomes and Covid-19 cases and hospital admissions.

## Looking forward

The existing monitoring applications overwhelmingly focus on the current state of a process. Problems are detected when they occur. However, using the explosion in available data and computing power, it is becoming more and more attainable to monitor the future states of a process. Instead of monitoring the current state of a process, we can model future outcomes from the effects of current and

past patterns in process indicators. For example, instead of monitoring heart rate to detect heart failure, we model the risk of heart failure as a function of indicators such as blood pressure, heart rate, and oxygen levels. The patterns of these indicators can hint towards future problems before they occur, enabling preventive action. Such action can prevent defective batches of products and machine damage in factories, fraud in financial markets, and large-scale outbreaks of infectious diseases.

## Monitoring Processes in the Age of Big Data

In *Statistical and Predictive Process Monitoring; Monitoring Processes in the Age of Big Data* (Huberts, 2021) we outline the shift from monitoring the current state (statistical process monitoring) to the future state of a process (predictive process monitoring). Firstly, the increase in available data and processing power enables improvements when monitoring the current state. For example, using the central limit theorem, fewer assumptions are needed to monitor non-normal data when using subgroups of sufficient size. This means it becomes easier to monitor a process indicator that has a non-normal distribution. Furthermore, updating the process parameters during monitoring can greatly improve the performance of the monitoring procedure.

Let's illustrate this with an example. Suppose we want to monitor the time it takes to perform a common surgical procedure such as an Appendectomy. The durations are

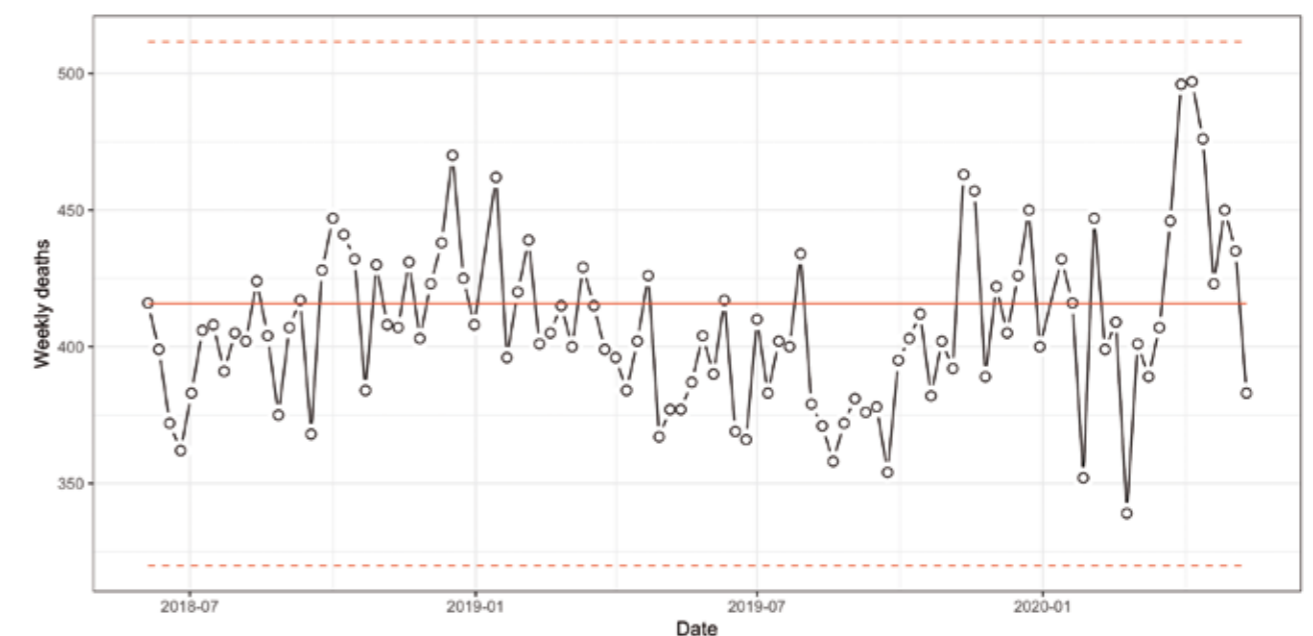


Figure 1. An example of a Shewhart control chart showing the weekly death rate among people of ages 0 to 65 in the Netherlands

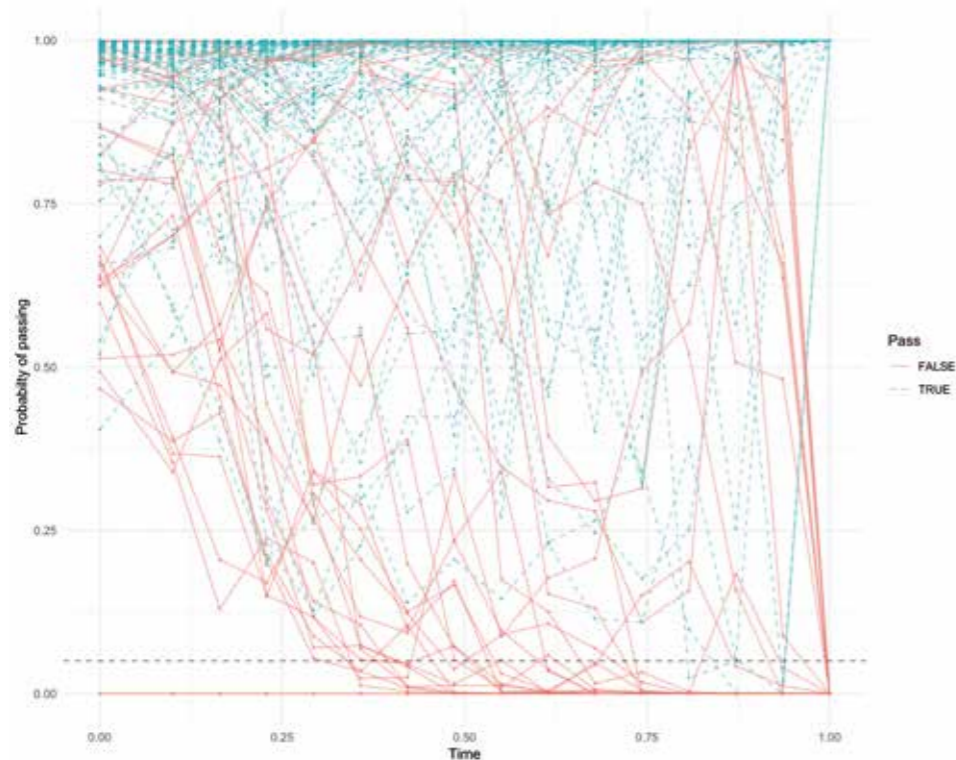


Figure 2. Graph showing the estimated probabilities of passing for a set of students throughout a school year

right-skewed, thus not normally distributed. Furthermore, we want to start monitoring as soon as possible. We can monitor this process using a control chart (see Figure 1. An example of a Shewhart control chart showing the weekly death rate among people of ages 0 to 65 in the Netherlands for an example of a Shewhart control chart). Much of the existing control chart theory uses the assumption of normality. By taking subgroups of sufficient size, we can circumvent the normality assumption (see chapter 2 of Huberts, 2021, for more details). To start monitoring, we estimate the in-control situation using the available data. These data should be for operations that are representative for the in-control process. The estimates are imperfect and updating these estimates during monitoring will improve the performance of the monitoring procedure (see chapter 3 of Huberts, 2021, for more details). Incorporating a delay in the updates can prevent problems (see chapter 4 of Huberts, 2021, for more details). In the long run, these adjustments to traditional process monitoring will reduce the number of false alarms, improve the detection rate, and increase usability.

As discussed, the increase in the amount of available data and computing power can facilitate predictive monitoring. In the second part of Huberts (2021), we discuss this transition to 'monitoring the future'. Two predictive process monitoring case studies are presented, one in mental healthcare and one in education.

### Monitoring mental health crises

In the first case study, we investigate monitoring mental health crises to assist healthcare workers to improve planning. Using a unique big data set on all Dutch citizens, we zoom in on people diagnosed with schizophrenia. Schizophrenia is a debilitating disease estimated to affect 1% of the population. One of the properties of schizophrenia is that people often relapse into crisis care. In Huberts et al. (2021) we set out to predict these crises, to enable preventive care and improved resource planning. Using a wide variety of variables on healthcare activities, diagnoses, income, and personal information over 250 variables were constructed based on input from practitioners. These variables were modeled to predict the probability of a crisis. Gradient boosting outperformed other methods (regression, support vector machines, random forest, etc.) in terms of predictive accuracy and efficiency. The next step was to set up a monitoring procedure based on the predictions. A threshold is needed to determine when a predicted probability is too high. It was unclear how to determine this threshold. Therefore, in Huberts et al. (2021), we propose a simple tuning procedure to find a threshold that will produce an acceptable false alarm rate. The results are promising, with performance depending on the time frame and chosen false alarm rate.

		C					
		0.05	0.1	0.25	0.5	0.75	0.999
	0.0	1 (0.07)	1 (0.07)	1 (0.07)	0.67 (0.13)	0.74 (0.47)	0.25 (0.93)
	0.1	1 (0.07)	1 (0.07)	1 (0.07)	1 (0.27)	0.71 (0.40)	0.71 (0.40)
	0.3	1 (0.10)	1 (0.20)	0.85 (0.37)	0.76 (0.53)	0.67 (0.67)	0.27 (1)
TIME	0.5	1 (0.33)	1 (0.43)	0.94 (0.53)	0.79 (0.63)	0.67 (0.67)	0.34 (0.97)
	0.7	1 (0.57)	1 (0.63)	0.88 (0.73)	0.77 (0.70)	0.70 (0.77)	0.40 (0.97)
	0.9	0.90 (0.63)	0.86 (0.63)	0.88 (0.70)	0.81 (0.70)	0.81 (0.73)	0.59 (0.90)
	1.0	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)	1 (1)

Table 1. Precision (Recall) for monitoring student failure for various values of threshold C and time t

### Monitoring student performance

In Huberts, Schoonhoven & Does (2020) we investigate predictive process monitoring for high school students. This challenge was proposed by high school managers and consists of identifying over- and underperforming students as early as possible. This enables targeted intervention by teachers and management. In Huberts, Schoonhoven & Does (2020) we propose a hierarchical Bayesian model for statistical process monitoring on the one hand and predictive process monitoring on the other. Using this model, schools can monitor at the individual course grade level, student level, and school level either signaling current problems (statistical process monitoring) or early warnings of future problems (predictive process monitoring). The model identifies much more structure in student performance than other investigated modeling techniques do. Figure 2. Graph showing the estimated probabilities of passing for a set of students throughout a school year shows a resulting 'control chart', monitoring the end-of-year performance for a set of students. The chart shows the probability of passing the academic year and, depending on a chosen threshold, signals when this probability is too low. The system works well, for example identifying a quarter of the students that will fail the year with 100% precision after just 10% of the year has passed (see Table 1. Precision (Recall) for monitoring student failure for various values of threshold C and time t).

In conclusion, the increase in computing power and the explosion of available data is enabling improvements in monitoring the current state of the process and

in monitoring the future state of that process. This predictive process monitoring has great potential to enable preventive action in services and industry, as demonstrated with examples in mental healthcare and education (see Huberts, 2021, for more details). At the Amsterdam Business School, we will continue working on these procedures, combining predictive modeling with process monitoring methods. Signaling as early as possible can be imperative in taking preventive measures in sectors such as healthcare, education, manufacturing, maintenance, and more. It can improve the quality of products and services.

#### REFERENCES

- Huberts, L.C.E. (2021). *Statistical and Predictive Process Monitoring: Monitoring Processes in the Age of Big Data (Doctoral Thesis)*. University of Amsterdam. <https://pure.uva.nl/ws/files/59618725/Thesis.pdf>.
- Huberts, L.C.E., Does, R.J.M.M., Ravesteijn, B., & Lokkerbol, J. (2021). Predictive monitoring using machine learning algorithms and a real-life example on schizophrenia. Early view in *Quality and Reliability Engineering International*.
- Huberts, L.C.E., Schoonhoven, M., & Does, R.J.M.M (2020). Monitoring student progress: A case study to predict student success or failure. Early view in the *Journal of Quality Technology*.

LEO C.E. HUBERTS is assistant professor in Business Analytics at the University of Amsterdam. He studies and teaches applied statistics and machine learning with an emphasis on process monitoring and analytics for a better world. This article is based on the work for his recent PhD thesis *Statistical and Predictive Process Monitoring: Monitoring Processes in the Age of Big Data* (2021).  
E-mail: L.C.E.Huberts@uva.nl

## Opa vertelt...

In de serie 'Opa vertelt' een herinnering aan de lang vervlogen tijd dat computers kolossale machines waren waar gebruikers, met de nodige eerbied, pakjes ponskaarten mochten aanleveren. Ze werden nog net niet door een stoommachine aangedreven, maar dat was het dan ook.

Iedereen maakte zijn eigen programma's. Standaardprogramma's bestonden nauwelijks, als je het trof gaf een collega je toestemming een door hem gemaakt programma te gebruiken. De meest gebruikte programmeertaal was Fortran en van virtueel geheugen en interactief werken was nog geen sprake.

Een beetje slimme programmeur maakte veel gebruik van zogenaamde subroutines. Op die manier hoefde je bijvoorbeeld niet op meerdere plekken in je programma de code voor het berekenen van een standaarddeviatie te herhalen. Je kon volstaan met een 'call' naar de subroutine die deze berekening uitvoerde en je het antwoord terugstuurde.

Een groot probleem in die beginjaren was de reken nauwkeurigheid. Een computer slaat getallen maar met een beperkte precisie op, alles voorbij een vastgesteld aantal decimalen wordt afgekapt. Nu komen in veel statistische formules zaken voor als het verschil van de som van de kwadraten van de gegevens en het kwadraat van de som van de gegevens. Dat kunnen enorme getallen worden waarvan de laatste decimalen worden afgekapt. Maar daarin zit soms het verschil dat we willen berekenen en dat wordt dan nul. Je kunt kan dit zelf gemakkelijk demonstreren met een eenvoudige rekenmachine. De getallen 1, 2, 3 hebben dezelfde standaarddeviatie als 11, 12, 13 of 101, 102, 103 of 10001, 10002, 1003 etc. Probeer het maar, en maak het aantal nullen steeds groter. Afhankelijk van de rekenmachine komt er dan op een gegeven ogenblik een foutmelding. Ik heb begin jaren '80 een artikel gepubliceerd in *Kwantitatieve Methoden* waarin ik de reken nauwkeurigheid van de statistische programma's op alle Nederlandse universitaire reken centra heb vergeleken. Daar waren grote verschillen, het meest dramatisch



Illustratie: Cornelis Jetses

was een zelfbouw programma dat een product-momentcorrelatie gaf in de buurt van 3. En dat terwijl die toch echt tussen -1 en 1 moet liggen. De oorzaak was dat programmeurs vaak niet meer deden dan formules uit een statistiekboek omzetten in Fortran-code zonder met dit soort zaken rekening te houden.

Gelukkig kwam er een pakket op de markt dat een grote hoeveelheid subroutines voor allerlei wiskundige en statistische functies bevatte die allemaal wél behoorlijk waren geprogrammeerd. Zoiets heette een subroutine-bibliotheek. Als dat pakket was geïnstalleerd kon je die subroutines als kant-en-klare bouwstenen direct 'linken' aan je eigen programma. Naast het verkrijgen van meer betrouwbare resultaten bespaarde je ook een hoop tijd en moeite.

Op een niet nader te noemen universitair rekencentrum was dit pakket aangeschaft en men had met veel publiciteit een presentatie georganiseerd. Daar kwam een groot aantal programmeurs van allerlei pluimage opdagen. De meesten kenden elkaar wel van contacten tijdens het aanleveren van de ponskaarten en het enkele uren later weer ophalen van de grote vellen kettingpapier met de uitkomsten. Maar er waren ook twee onbekenden aanwezig die wat bij de rest afstaken: duidelijk wat ouder en in een keurig pak gestoken. Dat bleken vertegenwoordigers van de Universiteitsbibliotheek te zijn. Zij begonnen vragen te stellen over de maximale uitleentijd, controle op het onbeschadigd terugbrengen, de officiële omschrijving van de routines volgens de standaardregels voor bibliotheken etc. Zij waren kennelijk naar de bijeenkomst gestuurd om die vrijgevochten cowboys van programmeurs eens goed uit te leggen hoe het écht in een bibliotheek behoort toe te gaan. Het heeft enige moeite gekost de heren te overtuigen dat de benaming subroutine-bibliotheek niets te maken had met het gebruikelijke werk van een klassieke bibliotheek...

Tja, dat waren nog eens pioniersdagen, mijmert deze opa...

GERRIT STEMERDINK is eindredacteur *STATOR*.  
E-mail: gjstemerding@hotmail.com

Van 11-16 juli 2021 vond het 63<sup>e</sup> World Statistics Congress van het International Statistical Institute ISI plaats. Tijdens de openingsceremonie werd een boodschap van koningin Máxima voorgelezen.



Foto: RVD - Z.M. de Koning

### MESSAGE FROM HER MAJESTY QUEEN MÁXIMA OF THE NETHERLANDS

*H.M. Queen Máxima of The Netherlands, UN SG's Special Advocate for Inclusive Finance for Development, Member of the Netherlands Committee for Enterprise*

Never has the International Statistical Institute's ambition to promote 'Statistical Science for a Better World' been more relevant than during this pandemic. Statistics provide evidence that guide decisions and it has been data and statistics that have informed health authorities, governments and citizens through this global crisis. I am pleased to see that in your programme you will be discussing the statistical lessons to be learnt from the pandemic.

Reliable statistics are important for all citizens, especially young people. Statistical and data literacy are indispensable skills that enable young people to assess the reliability of the data they use.

In my work as United Nations Secretary General's Special Advocate for Inclusive Finance for Development (UNSGSA), I have championed the importance of data to inform policies and action. For example, the World Bank Findex which provides a common set of indicators on financial inclusion consistently measured across countries has galvanized global support and informed many national strategies on financial inclusion. In addition, together with the International Monetary Fund we encouraged the reporting of gender-disaggregated data in the Financial Access Survey. Such granular data can look more closely at important issues and catalyze needed policy reform.

Please do take the opportunity of the virtual conference to meet with friends and colleagues. Your work is of great importance and value to us all. Exchanging information, experiences, research and acquired knowledge lies at the basis of that. I wish you all the best for a successful and productive Conference.



### Young Statisticians 10 years old

To start off with a fun fact, worthy of some celebration: Young Statisticians is 10 years old! We were founded in May 2011, with the mission to enrich the environment in which young statisticians in the Netherlands take the first steps of their career. Since then, different YS boards have organized numerous company visits, workshops, statistics/science cafés and pub quizzes. Hopefully at least as many events will follow in the decade to come.

### YS board

The YS board has recently undergone some changes: the current board members are Marije Sluiskes, Iris Hoekstra, Maxine Storm and Jonas Ammeling. We are looking forward to organizing many interesting and fun events this year, meant for anyone who considers him- or herself a 'young statistician' (or strives to be one). We encourage anyone with ideas for future YS events to contact us via [youngstatisticians@vvsor.nl](mailto:youngstatisticians@vvsor.nl).

### September 23 Statistics Café

We are excited to announce that the first YS event since the start of the pandemic is planned for Thursday evening September 23rd: a Statistics Café with the theme 'Statistics: playing in everyone's backyard'.

For those of you – understandably – wondering what a backyard has got to do with statistics: it's a reference to John Tukey's famous quote that 'the best thing about being a statistician is that you get to play in everyone's backyard' – which we couldn't agree with more.

### Newsletter

If you want to keep updated on the details (location + speakers) of this event, or any of our future events, then subscribe to our newsletter via [vvsor.nl/young-statisticians](http://vvsor.nl/young-statisticians).

You can also follow us on

Twitter: @YoungStatNL

Facebook.com/young.statisticians

LinkedIn.com/company/young-statisticians



