

STATOR

Optimalisatie van productieplanning in de
mengvoederindustrie

Tekstanalysemethoden; Toepassingen in
de officiële statistiek

Een eureka-moment dat nog een beetje natrilt;
Over zwaartekrachtgolven en Bayesiaans rekenen

Planning van multidisciplinaire kankerzorg

Nederlandse Statistiek in *The Wall Street Journal*

Expeditie Robinson; Van ecologische correlatie
naar multi level analyse

Is gelijkheid altijd gewenst? De afweging tussen
gelijkheid en kwaliteit in personeelsroostering

KISS: Keep It Simple, Statistician!

In Memoriam Ivo W. Molenaar (1935 – 2018)

Young Statisticians



STATOR

Jaargang 19, nummer 2, juni 2018

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, Ana Isabel Barros, Joep Burger, Kristiaan Glorie, Caroline Jagtenberg, Guus Luijben (eindredacteur), Richard Starmans, Gerrit Stemerding (eindredacteur) en Vanessa Torres van Grinsven. Vaste medewerkers: Johan van Leeuwen, John Poppelaars, Gerard Sierksma en Henk Tijms.

Kopij en reacties richten aan

Prof. dr. J.A.S. Gromicho (hoofdredacteur), Faculteit der Economische Wetenschappen en Bedrijfskunde, afdeling Econometrie, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, telefoon 020 5986010, mobiel 06 55886747, j.a.dossantos.gromicho@vu.nl

Bestuur van de VVSOR

Voorzitter: prof. dr. Fred van Eeuwijk, db@vvsor.nl
Secretaris a.i.: dr. Laurence Frank, db@vvsor.nl
Penningmeester: dr. Ad Ridder, db@vvsor.nl
Overige bestuursleden: dr. Eric Cator (SMS), prof. dr. Ernst Wit (BMS), Maarten Kampert MSC., prof. dr. Albert Wagelmans (NGB), dr. Michel van de Velden (ECS), prof. dr. Jelte Wicherts (SWS), Elian Griffioen (Young Statisticians).

Leden- en abonnementenadministratie van de VVSOR

VVSOR, Postbus 1058, 3860 BB Nijkerk, telefoon 033 2473408, admin@vvsor.nl
Raadpleeg onze website www.vvsor.nl over hoe u lid kunt worden van de VVSOR of een abonnement kunt nemen op STATOR.

Advertentieacquisitie

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

Ontwerp en opmaak

Pharos, Nijmegen

Uitgever

© Vereniging voor Statistiek en Operations Research
ISSN 1567-3383

The Amazing Technicolor World of Statistics and OR

Met deze knipoog naar een musical van Andrew Lloyd Webber willen we aangeven dat we een geweldig veelzijdig vakgebied hebben! In werkelijk alle hoeken en gaten van de samenleving vinden we toepassingen van OR en statistiek. Ook op de recente Annual Meeting bleek dat bij een onderwerp als Climate Change overall ons vakgebied opduikt.

Een wel heel bijzondere toepassing is een Bayesiaanse schattingsmethode die Cajo ter Braak heeft ontwikkeld en die werd gebruikt bij de analyse van de gegevens van de LIGO-detector. Hiermee werd het bestaan van zwaartekrachtgolven aangetoond. De Nobelprijs voor de Natuurkunde 2017 die hiervoor werd toegekend straalt, wat ons betreft, dan ook een beetje op Cajo ter Braak af!

Een andere spraakmakende Nederlandse toepassing is die van het samenstellen van de schaatsploegen voor grote toernooien. Gerard Sierksma ontwikkelde een methode om dat op zo'n manier te doen dat het aantal te verwachten medailles maximaal is. *The Wall Street Journal* berichtte met bewondering over deze unieke aanpak en Sierksma verwerkte dat in zijn column.

In een artikel uit 1950 liet W.S. Robinson zien dat verbanden tussen individuele personen heel anders kunnen zijn dan verbanden tussen groepen van personen. Richard Starmans vertelt hier uitgebreid over en behandelt de impact die dit artikel heeft gehad.

Verder vindt u in dit nummer de boeiende rede van Jan Karel Lenstra ter gelegenheid van het 30-jarig bestaan van het Landelijk Netwerk Mathematische Besliskunde, artikelen over optimale productieplanning, tekstanalysemethoden, het afwegen van gelijkheid en kwaliteit bij personeelsroosting, planning van multidisciplinaire zorg en de beide juryrapporten voor de Willem R. van Zwet en de Jan Hemelrijk Awards.

Helaas is een grote en zeer aimabele statisticus ons ontvallen: ons erelid Ivo Molenaar overleed op 26 februari 2018. Anne Boomsma schreef een In Memoriam waarin zijn vele verdiensten worden belicht.

Het nieuws van de Young Statisticians, een mededeling over de Jan Tinbergen Awards en een column over simpele aanpak zorgen er voor dat dit nummer boordevol wordt. Deze STATOR is daarmee een van de meest afwisselende die we ooit hebben uitgebracht; prima geschikt om mee te nemen als vakantielectuur.

Wij wensen u een zeer veelzijdig leesplezier en een goede vakantie.



4



8



14



18



24



30

INHOUD

- 2 Redactioneel
- 4 Optimalisatie van productieplanning in de mengvoederindustrie | JOOST BERKHOUT, ERIC PAUWELS, ROB VAN DER MEI, JAN STOLZE & SIEM BROERSEN
- 8 Tekstanalysemethoden; Toepassingen in de officiële statistiek | ARNOU VAN DELDEN, PIET DAAS, OLAV TEN BOSCH & DICK WINDMEIJER
- 13 Jury report Willem R. van Zwet Award 2017
Jury rapport voor de Jan Hemelrijk Award 2017
- 14 Een eureka-moment dat nog een beetje natrilit; Over zwaartekrachtgolven en Bayesiaans rekenen | CAJO TER BRAAK
- 18 Planning van multidisciplinaire kankerzorg | GRÉANNE LEEFTINK & ERWIN HANS
- 22 Nederlandse Statistiek in *The Wall Street Journal* – column | GERARD SIERKSMA
- 24 Expeditie Robinson; Van ecologische correlatie naar multi level analyse | RICHARD STARMANS
- 30 Is gelijkheid altijd gewenst? De afweging tussen gelijkheid en kwaliteit in personeelsroosting | THOMAS BREUGEM
- 34 KISS: Keep It Simple Statistician! – column | GERRIT STEMERDINK
- 35 Jan Tinbergen Awards nu beschikbaar voor alle jonge statistici
- 36 In Memoriam Ivo W. Molenaar (1935 – 2018) | ANNE BOOMSMA
- 40 Thirty years LNMB | JAN KAREL LENSTRA
- 42 Euro working group on OR in practice; A report from the first workshop | RUTH KAUFMAN
- 43 Young Statisticians



OPTIMALISATIE VAN PRODUCTIEPLANNING IN DE MENGVOEDERINDUSTRIE

Een klassiek operationeel onderzoeksprobleem, zo zou je het productieplanningsvraagstuk kunnen typeren waar zo'n 120 mengvoederfabrieken in Nederland dagelijks mee te maken hebben. Echter, met een aantal processpecifieke restricties en de toenemende beschikbaarheid van procesdata, biedt het optimalisatieprobleem ook op academisch gebied veel uitdagingen. Dit artikel beschrijft het publiek-private samenwerkingsverband tussen het CWI en ENGIE Services* die samen deze uitdagingen zijn aangegaan. Na het introduceren van het productieproces en belichting van het planningsprobleem wordt onze optimalisatie van de productieplanning in de mengvoederindustrie beschreven.

JOOST BERKHOUT, ERIC PAUWELS, ROB VAN DER MEI, JAN STOLZE & SIEM BROERSEN

In het onderzoek staan de moderne mengvoederfabrieken die een groot aantal samengestelde biograndstoffen combineren en verwerken tot verschillende soorten diervoer (voornamelijk veevoer) centraal. Het productieproces bestaat uit grofweg twee fasen. In de eerste fase worden de biograndstoffen (bijvoorbeeld tarwe of gerst) *gemalen* en *gemengd* tot een homogeen geheel in de zogenaamde maalmenglijn. Tijdens de tweede fase in de perserij wordt het gemengde halffabricaat *geperst* tot korrels (zie figuur 1). Figuur 2 toont een schematische weergave van het hele productieproces, inclusief de grondstofinname en de verlading van eindproducten in vrachtwagens.

Planningsprobleem

Klanten (voornamelijk boeren) plaatsen bestellingen en het is aan de planners om de productievolgorde te bepalen, zodat klanten op tijd geleverd worden. Daarnaast is er een voorraad van standaardproducten die op peil moet worden gehouden. Wanneer de klanten op tijd geleverd kunnen worden, is een andere doelstelling om de benuttingsgraad van de fabriek te maximaliseren.

Verscheidene aspecten maken het opstellen van een toegestane en efficiënte planning lastig. Zo is er een beperkte silocapaciteit voor halffabricaten en eindproducten, wat kan leiden tot productiestilstand. Verder kan niet elk product geproduceerd worden op alle productielijnen



Figuur 1. Een mengvoederkorrel

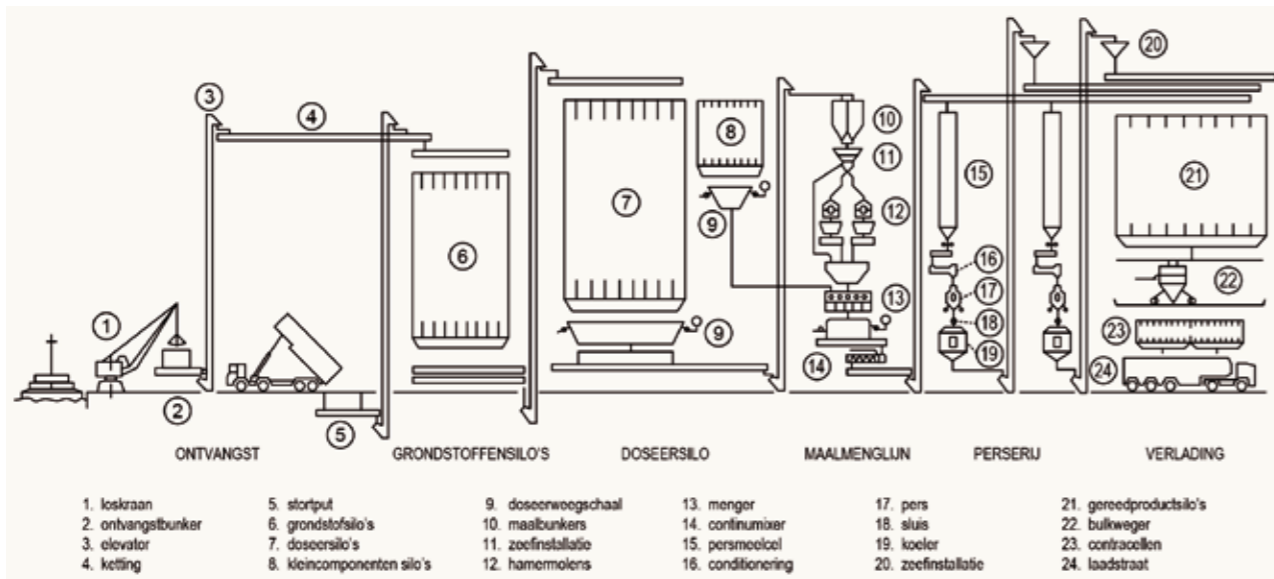
en ook de productietijden variëren. Daarbij komt het regelmatig voor dat een productielijn in onderhoud is.

Een ander planningsaspect is dat niet alle productieplanningen zijn toegestaan wegens vervuilingrestricties. Sommige producten vervuilen het productieproces, waardoor andere producten niet direct erachter gepland kunnen worden. Ter illustratie, koper is essentieel voor varkens terwijl een kleine dosis al fataal kan zijn voor schapen. In het geval van vervuiling moet de productielijn eerst worden schoongemaakt. Dit kan efficiënt gebeuren door de tussenproductie van een niet-vervuilend product waarvoor de bestaande vervuiling geen probleem is. Op deze manier wordt de productiecapaciteit benut, terwijl tegelijkertijd de productielijn wordt schoongemaakt. Handig omspringen met deze vervuilingrestricties, verhoogt de benuttingsgraad van de fabriek.

Naast de vervuilingrestricties is een ander volgordeafhankelijk planningsaspect dat producten verschillende machineconfiguraties vereisen. Tijdens een configuratiewijziging kan er niet geproduceerd worden wat ten koste gaat van de machinebenuttingsgraad. Zo kunnen producten verschillende zeefmaten vereisen bij het malen, waardoor er een productiebelemmerende zeefwissel moet plaatsvinden.

Een trend van de laatste decennia is dat klanten steeds meer specifieke producten bestellen (zogenaamde *mass customization*). Boeren zijn continu op zoek naar de beste voersamenstelling voor hun vee. Een boer kan bijvoorbeeld een product met extra eiwitten bestellen, omdat hij merkt dat het de melkproductie van zijn koeien doet toenemen. In sommige fabrieken was voorheen 80% standaardproductie, terwijl nu 80% van de productie maatwerk is. Gevolg is dat het tot complexere planningsituaties leidt.

Naast de genoemde deterministische aspecten is er ook nog een aantal stochastische aspecten die het plannen bemoeilijken. Zo zijn er regelmatig storingen, waardoor bedachte planningsherzieningen moeten worden met als doel om zoveel mogelijk klanten alsnog op tijd te kunnen leveren. Tevens zijn de productietijden stochastisch



Figuur 2. Schematische weergave van een mengvoeder productieproces (gebaseerd op Kenniscentrum InfoMil van Rijkswaterstaat, 2018)

met seizoentrends. De doorlooptijd in de maler, bijvoorbeeld, varieert per seizoen door verschillende taatheid van oogsten.

In veel fabrieken wordt het complexe productieplanningsprobleem nu met de hand opgelost door ervaren planningsexperts, maar dat is arbeidsintensief en inflexibel. Als er na veel arbeid een planning is opgesteld, zijn planners terughoudend met het aanpassen van de planning bij verstoringen of bij nieuwe kansen, bijvoorbeeld in het geval van een winstgevende spoedorder.

Het doel van dit onderzoeksproject is om algoritmes te ontwikkelen die het dynamische productieplanningsprobleem binnen de mengvoederindustrie (zo efficiënt mogelijk) oplossen. Een belangrijke focus daarbij is het ontwikkelen van robuuste plannings die zo min mogelijk wijzigingen vereisen bij onvoorziene omstandigheden, zoals een kapotte machine of een spoedorder. Het specifieke planningsprobleem kan in de grotere context van *smart industry* worden beschouwd. In *smart industry*, ook wel bekend als de vierde industriële revolutie, draait het om de benutting van de grote hoeveelheden aan (*real-time*) data, die alles kwantificeert van het productie en logistieke proces, voor een efficiëntere industrie.

Uitdaging

De vervuilsrestricties en de omschakelingen van machineconfiguraties zorgen voor volgordeafhankelijke opstarttijden. Specifiek, de vervuilsrestricties zorgen ervoor dat er sprake is van volgordeafhankelijke opstart-

tijden waarvoor de driehoeksongelijkheid *niet* geldt. Als voorbeeld, de totale opstarttijd voor planning $A \rightarrow B \rightarrow C$ kan kleiner zijn dan voor planning $A \rightarrow C$ in het geval productie A tot vervuiling leidt voor productie C en productie B als schoonmaak kan dienen. Het is interessant om de theoretische implicaties hiervan te onderzoeken en te benutten.

Zowel productiefase 1 (op een maalmenglijn) als fase 2 (bij de perserij) bestaan over het algemeen uit meerdere parallelle productielijnen. Naast de productievolgorde, moet de planner dus ook bepalen op welke maalmenglijn en perslijn een product te produceren. Bovendien is elke productielijn opgebouwd uit in serie geschakelde units waarin specifieke productiestappen plaatsvinden. Een maalmenglijn heeft bijvoorbeeld een maalunit en een mengunit. Iedere unit biedt ruimte aan een specifiek product zodat meerdere producten gelijktijdig geproduceerd kunnen worden in de verschillende units van een productielijn. Elk product heeft verschillende productietijden bij verschillende units en hun bottlenecks kunnen variëren. Tussen de units is meestal geen opslagruimte en inhalen is geen optie. Gevolg is dat een snel product evengoed langzaam wordt geproduceerd als deze achter een traag product wordt gepland. Dit resulteert in zogenaamde *verschuivende bottlenecks*, dat wil zeggen de bottleneck van een productielijn verschuift over de tijd (afhankelijk van de te plannen producten en de planning). Handig om-springen met deze verschuivende bottlenecks kan de productiecapaciteit verhogen.

Bekend is dat alle benoemde facetten een planningsprobleem over het algemeen moeilijker maken,

zie Pinedo (2016). Anders gezegd, het mengvoederplanningsprobleem is NP-hard, wat grofweg inhoudt dat er weinig hoop is voor een 'efficiënt' algoritme om het planningsprobleem exact op te lossen. Praktisch betekent dit dat de rekentijd exponentieel toeneemt met het aantal in te plannen productieorders. De grens van wat oplosbaar is binnen afzienbare rekentijd, zeg binnen een uur, wordt daarom snel bereikt voor een groeiend aantal productieorders.

Een eerste aanpak

Onze eerste aanpak is om de maalmenglijnen en de perserij van het planningsprobleem te modelleren als een *mixed integer linear programming* (MILP). Een MILP is een wiskundig optimalisatieprogramma van lineaire aard, waarin sommige optimalisatievariabelen zich beperken tot de gehele getallen. Evenals het mengvoederplanningsprobleem is het oplossen van een MILP over het algemeen complex (NP-hard) en dus wordt er met deze aanpak in essentie niet toegegeven aan de complexiteit. De verdere voornaamste redenen voor een MILP-aanpak zijn:

- De modeleerexercitie resulteert in diepgaand inzicht in de kern van het probleem.
- Een MILP-aanpak geeft veel flexibiliteit. Het maakt het mogelijk om met relatief weinig moeite verschillende fabrieksoptellingen (bijvoorbeeld een extra perslijn) te modelleren met verschillende doelstellingen. Ook bij een kapotte machine kan de MILP aangepast worden en opnieuw worden doorgerekend. Deze flexibiliteit is ook handig om het potentieel te onderzoeken van een extra productielijn en zodoende inzicht te krijgen of het de extra investering waard is.
- Ondanks dat het planningsprobleem theoretisch gezien een complex probleem is, maakt de enorme toename in computerrekenkracht in combinatie met de ontwikkelingen van commerciële MILP-oplossoftware het tegenwoordig mogelijk om aanzienlijke instanties op te lossen (waarvoor volledige enumeratie praktisch nog steeds onmogelijk is). Optimale oplossingen kunnen tevens benut worden bij het testen van heuristieken (stelregels) geschikt voor nog grotere instanties.
- Door de zoekruimte voor plannings te beperken met extra restricties, kan de MILP als basis dienen van een heuristische benadering voor grotere planningsinstanties.
- Het onderzoeksgebied van robuuste optimalisering neemt een MILP als uitgangspunt en past restricties aan zodat er meer speling ontstaat. De gedachte is dat

er niet meer op het scherpst van de snede wordt geoptimaliseerd maar op robuustheid, zodat een planning niet helemaal op de schop hoeft als een coëfficiënt in de programmering ook maar iets verandert.

Huidige stand en conclusie

Een algemeen MILP-model is reeds ontwikkeld, zie Berkhout et al. (2017), en dit wordt nu toegespitst op een daadwerkelijke fabrieksoptelling met historische productiedata. Het *fine tuning* van het model aan de werkelijkheid is een essentiële maar ook uitdagende en tijdrovende klus. Als deze stap is genomen, kan specifiekere gekeken worden naar hoe het probleem zo efficiënt mogelijk op te lossen. De eerste resultaten zijn hoopvol: initiële experimenten demonstreren dat met een efficiëntere planning de fabriek 8% minder tijd nodig heeft om hetzelfde te produceren als een eerder gerealiseerde planning terwijl de klantdeadlines gerespecteerd blijven. Het project is in 2017 met 2 jaar verlengd om de eerste inzichten verder uit te bouwen. Wordt dus vervolgd.

* ENGIE Services is de grootste technologische dienstverlener in Nederland die gespecialiseerd is in de automatisering van mengvoederfabrieken.

LITERATUUR

- Kenniscentrum InfoMil van Rijkswaterstaat. (2018) Geraadpleegd op 30-3-2018 via <https://www.infomil.nl/onderwerpen/lucht-water/lucht/activiteiten/diervoederindustrie/procesbeschrijving/>
- Berkhout, J., Pauwels, E., van der Mei, R. D., Stolze, J., & Broersen, S. (2017). Short-term Production Scheduling with Non-triangular Sequence-dependent Setup Times and Shifting Production Bottlenecks. *Under revision*.
- Pinedo, M. L. (2016). *Scheduling: Theory, Algorithms, and Systems*. Springer International Publishing.

Joost Berkhout is als postdoc Intelligent and Autonomous Systems werkzaam bij het Centrum Wiskunde & Informatica. E-mail: joost.berkhout@cwi.nl

Eric Pauwels is groepsleider Intelligent and Autonomous Systems bij het Centrum Wiskunde & Informatica.

Rob van der Mei is manager Research en Strategie bij het Centrum Wiskunde & Informatica.

Jan Stolze is als consultant MES Expert Center werkzaam bij ENGIE Services.

Siem Broersen is manager IA en MES Expert Center bij ENGIE Services.

TEKSTANALYSEMETHODEN

Toepassingen in de officiële statistiek

Tekstanalyse, ook wel *text mining* genoemd, is het proces waarmee geautomatiseerd waardevolle informatie wordt afgeleid uit teksten. Een voorbeeld is het automatisch classificeren van e-mails naar spam en niet-spam. Tekstanalysemethoden zijn in de officiële statistiek tot nu toe vooral gebruikt om antwoorden van respondenten op open enquêtevragen in te delen naar een vooraf vastgestelde classificatie. Bijvoorbeeld het toekennen van een tekstsomschrijving voor 'beroep' naar een categorie van de beroepenclassificatie. Momenteel worden binnen de officiële statistiek ook nieuwe toepassingen onderzocht. Aan welke toepassingen kun je dan denken? En hoe werkt dat? In het artikel gaan we op deze vragen in.

ARNOUT VAN DELDEN, PIET DAAS, OLAV TEN BOSCH & DICK WINDMEIJER

Tekstanalyse kan op meerdere manieren nuttig zijn voor de officiële statistiek. We kunnen het gebruiken om bedrijfsgegevens af te leiden uit teksten van bedrijfswebsites, zoals contactgegevens van bedrijven. Website-gebaseerde tekstanalyse kunnen we ook gebruiken voor het afbakenen van populaties, bijvoorbeeld voor het vinden van duurzame bedrijven. We kunnen het ook gebruiken om objecten automatisch, in plaats van handmatig in te delen naar standaardclassificaties. Zo gebruikt het CBS websites om prijs en andere informatie van kleding te verzamelen voor het berekenen van de prijsontwikkeling van kleding. Een ander soort toepassing is informatie

extractie, bijvoorbeeld uit een bouwvergunningstekst automatisch het soort object, de locatie, het bedrag en de start van de bouwperiode afleiden. Ten slotte kunnen we het toepassen voor het afleiden van sentimenten in teksten. Deze techniek heeft het CBS gebruikt om een sociale spanningen-indicator te ontwikkelen, zie < <https://www.cbs.nl/nl-nl/onze-diensten/innovatie/> > .

Classificatiemethoden

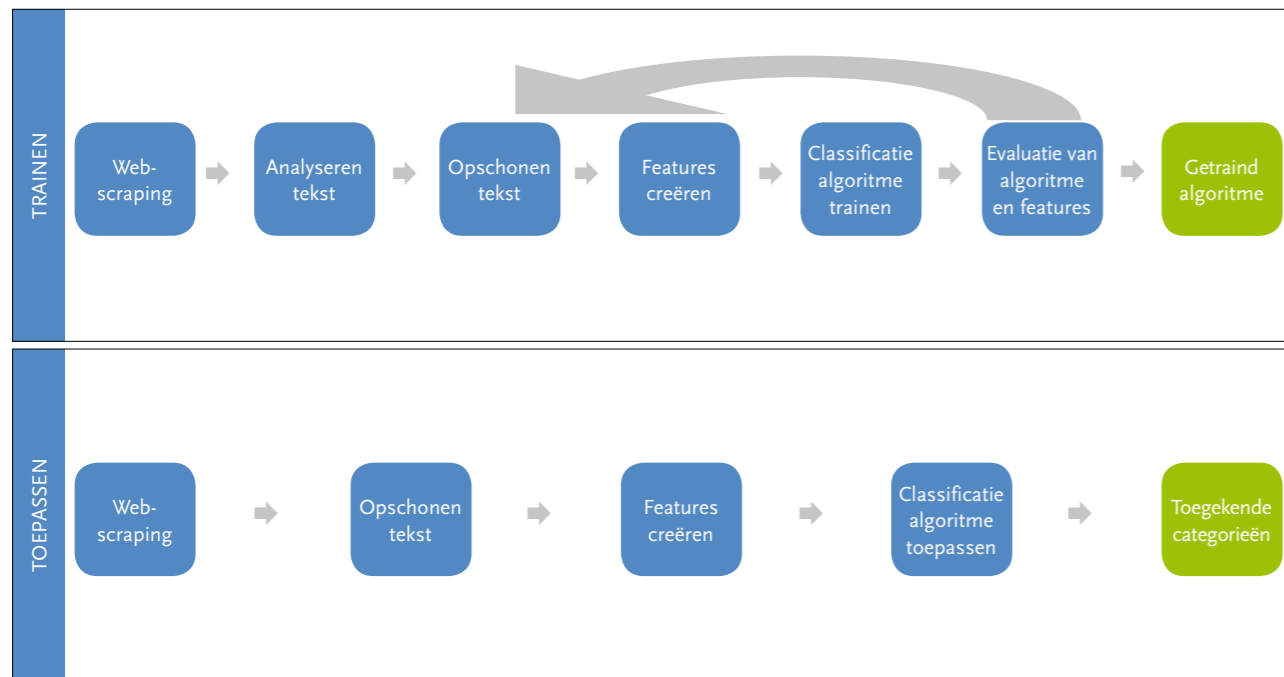
We beperken ons hier tot tekstanalysemethoden die zich



richten op automatisch classificeren. Daarvan zijn er drie soorten. De eerste zijn regel-gebaseerde methoden. Deze werken met ALS-DAN regels, bijvoorbeeld ALS 'meester' DAN 'leerkracht basisonderwijs'. Deze methode gebruikt het CBS bijvoorbeeld om beroepen uit een open enquêtevraag te classificeren en om doodsoorzaken op basis van informatie van artsen te classificeren. Hoewel er met ALS-DAN regels een behoorlijke precisie bereikt kan worden is het nadeel dat er vaak veel regels nodig zijn, die bovendien nogal onderhoudsintensief zijn. De andere twee methoden werken met statistische modellen. Hierbij onderscheiden we *supervised* en *unsupervised*

learning. Bij *supervised learning* worden modellen getraind op basis van een flink aantal voorbeelden die vooraf met de hand zijn geclassificeerd. Het getrainde model wordt vervolgens gebruikt om voor nieuwe teksten de categorie te kunnen voorspellen. Veel gebruikte methoden voor *supervised learning* zijn Naive Bayes, logistische regressiemodellen, Support Vector Machines, beslisbomen en neurale netwerken (Hastie et al., 2016).

Bij de methoden voor *unsupervised learning* wordt niet gewerkt met voorbeelden, maar worden de teksten in clusters ingedeeld. Hierbij wordt vaak *k-means* gebruikt maar ook zogenaamde 'topic modellen' (Blei, 2012) en



Figuur 1. *Supervised learning* bij tekstanalyse op basis van websites

word embeddings (Ruder, 2016) om (semantische) structuren in teksten te vinden.

In het vervolg van de tekst gaan we nader in op *supervised learning* methoden (zie ook figuur 1).

Features

Voor veel nieuwe toepassingen is het belangrijk om de juiste, meest voorspellende, eigenschappen, *features* genoemd, uit de tekst te halen voordat een *machine learning* model kan worden ingezet. Het begint er mee, dat de teksten van het internet gehaald moeten worden, bijvoorbeeld via *web scraping*. Vervolgens worden de teksten geanalyseerd: welke woorden en tekens komen voor en hoe vaak? Een grafische weergave van de teksten kan daarbij heel nuttig zijn. Uit de analyse kan naar voren komen dat bepaalde woorden vaak voorkomen, maar dat deze niet bijdragen aan de informatie die je uit de tekst wilt halen. De volgende stap is dan dat een tekst wordt opgeschoond. Daarbij worden bijvoorbeeld spaties, leestekens en emoticons verwijderd en woorden (of delen van woorden) geïdentificeerd. Vervolgens wordt vaak een automatische taalherkenning toegepast, mede omdat de vervolgstappen taalgevoelig zijn. Daarna kan het nuttig zijn de woorden te standaardiseren naar de

'stam' van het woord en stopwoorden te verwijderen. Ook worden heel vaak voorkomende woorden soms ook nog verwijderd. De woorden die overblijven vormen de basis voor de *features* die in de modellen ingezet worden. Daarbij worden ook varianten gebruikt, bijvoorbeeld 'n-grammen' waarbij groepen van woorden die vaak in combinatie voorkomen als één *feature* worden gekozen. Ten slotte worden de *features* omgezet naar getallen. De twee bekendste methoden hiervoor zijn het bepalen van de relatieve frequentie van een woord in een document, rekening houdend met het voorkomen ervan in alle andere documenten (TF-IDF) en het omzetten naar vectoren, de zogenaamde *word embeddings*. De laatste methode houdt rekening met de context van een woord in de tekst (Ruder, 2016).

Modelselectie

In praktijk worden meerdere modellen uitgetest, worden per model meerdere waarden van de modelparameters getest en wordt een *feature*-selectie uitgevoerd. De beste manier om een model te ontwikkelen bij *supervised learning* is om de dataset in tweeën te verdelen. Eén deel wordt gebruikt voor eenvoudige kruisvalidatie waarbij modelparameters worden getuned.



Figuur 2. Relatief aantal innovatieve bedrijven in Limburg per gemeente, gecorrigeerd voor de bevolkingsdichtheid

Het tweede deel van de data wordt gebruikt om een model waarvan alle parameters getuned zijn te testen. Bij het valideren en testen kan de kwaliteit van de voorspelde categorieën op verschillende manieren gemeten worden. De precisie van een voorspelde klasse meet welk aandeel daarvan correct is. De *recall* meet welk deel van een werkelijke categorie goed voorspeld is. De nauwkeurigheid meet de totale fractie goed voorspelde categorieën. Afhankelijk van het doel van het classificeren wordt bepaald welke maat het best kan worden gekozen. Als eenmaal een getraind model is verkregen, kan dat gebruikt worden om voor nieuwe teksten de categorie te voorspellen.

Innoverende bedrijven

Ten slotte geven we drie voorbeelden van recent onderzoek op het CBS waarbij tekstanalysemethoden zijn gebruikt. Het eerste voorbeeld betreft het vinden van innovatieve bedrijven. Het CBS houdt tweejaarlijks een innovatie enquête onder bedrijven met 10 of meer werknemers. De vraag was of innoverende bedrijven ook op een andere manier kunnen worden geïdentificeerd. Besloten werd hiervoor naar de tekst op de hoofdpagina van de website van bedrijven te kijken. De websites van 3000 innovatieve en 3000 niet-innovatieve bedrijven volgens de innovatie-enquête uit 2014 en die van bedrijven in de innovatietop 100 van het midden- en kleinbedrijf van 2009-2017 waren hierbij de training- en testset. Na opschoning van de tekst en trainen van het model, werden met een logistisch regressiemodel op een onafhankelijke testset de volgende kwaliteitsscores gehaald: precisie 100%, recall 87% en nauwkeurigheid 92% (van der Doef et al., 2018). Met behulp van dit model zijn veel meer bedrijven-websites geclassificeerd en zijn kaarten gemaakt waarbij een schatting is gemaakt van de dichtheid aan innovatieve bedrijven per provincie en gemeente, gecorrigeerd voor de bevolkingsdichtheid (zie figuur 2). Op basis van de enquêtegegevens kunnen we dit niet per gemeente schatten.

Verhuiswens

Een tweede voorbeeld is de vraag van een externe klant of verhuiscansen van personen afgeleid kunnen worden uit sociale media berichten. Daarbij zijn uit publieke sociale media berichten tussen 2014 en 2017 eerst alle berichten geselecteerd met de woorden verhuis* of verhuiz*. Als training- en testset is vervolgens uit deze berichten een random steekproef van 1000 berichten getrokken die door meerdere personen, onafhankelijk van elkaar, handmatig zijn getypeerd. Hierbij werd gekeken of het bericht



van een persoon afkomstig was die duidelijk aangaf te willen verhuizen. Uiteindelijk waren er bijna 120 berichten met een verhuiscens gevonden. Bij het opschonen bleek dat leestekens en emoticons verwijderd moeten worden, maar dat het standaardiseren van de woorden naar stam juist niet verstandig is. Vervolgens is een aantal modellen getest, waarvan er drie een vergelijkbare nauwkeurigheid bleken op te leveren: Logistische regressie, *Gradient Boosting* en een neurale netwerk (allen 85%). *Gradient Boosting* had daarnaast een precisie van 90% en een recall van 89%. In het vervolgonderzoek wil het CBS nagaan in hoeverre deze berichten te gebruiken zijn als voorspeller van verhuizingen, bijvoorbeeld door dit te vergelijken met de verdeling van verhuizingen in de populatie.

Economische activiteit

Een laatste voorbeeld betreft het afleiden van de economische activiteit van een bedrijf op basis van bedrijfswebsites (Roelands et al., 2017). Dit gegeven wordt geregistreerd als een bedrijf zich bij de Kamer van Koophandel inschrijft, maar wijzigingen in activiteit worden zelden doorgegeven. Met website-informatie kan mogelijk een actuelere code verkregen worden. Het onderzoek is beperkt tot het voorspellen van de negen zogeheten topsectoren van de economie, die weer onderverdeeld zijn in 29 subsectoren. Uit elke subsector is een steekproef van 70 bedrijven met URL getrokken. Van elke URL is de hoofdpagina plus één onderliggende pagina *gescraped*, en met taalherkenning zijn pagina's met de Nederlandse taal geselecteerd. De teksten zijn geschoond, en een trefwoordenlijst voor economische activiteit is gebruikt als de basis voor de voorspellende *features*. Met het best passende model is uiteindelijk op topsectorniveau, op een onafhankelijke testset, een kwaliteit behaald van 80% precisie, 58% recall en 51% nauwkeurigheid. Op subsector niveau waren de kwaliteitsscores een stuk lager. Hier zijn nog talloze punten waarop de tekstanalyse verbeterd kan worden, maar met name het voorspellen van een grote range verschillende categorieën lijkt lastig te zijn.

Toekomst

In de toekomst wil het CBS steeds meer gebruik maken van informatie die in de vorm van teksten, zoals op websites, beschikbaar is. Er is een aantal onderwerpen waar

we verder onderzoek naar willen doen. Ten eerste willen we graag populaties van personen en bedrijven flexibel en snel kunnen indelen naar allerlei kenmerken. Behalve het toekennen van de categorie op basis van de tekst, speelt daarbij ook de vraag hoe nauwkeurig we die populaties kunnen toekennen, en hoe we kunnen corrigeren voor selectiviteit in de verschillende categorieën waarvoor informatie beschikbaar is. Een tweede onderwerp vormt het combineren van informatie uit teksten met de al beschikbare bronnen op het CBS. Bij koppelen op eenheidenniveau is het probleem dat de direct identificerende kenmerken ontbreken (twitterberichten), dat ze nog allerlei fouten bevatten, of dat ze niet uniek zijn (websites). Een laatste onderwerp heeft te maken het schatten van de kwaliteit van CBS-publicaties. Stel we hebben van een categoriale variabele twee reeksen: één op basis van een steekproef of een administratieve bron en één op basis van tekstanalyse. Dan kunnen we vervolgens die twee reeksen combineren in een zogeheten latente variabele model om vervolgens de kwaliteit van beide bronnen, dus ook die van de CBS-publicatie, te schatten.

LITERATUUR

- Blei, D. (2012) Introduction to Probabilistic Topic Models. *Comm. ACM*, 55(4), 77–84.
- Hastie, T., Tibshirani, R., & Friedman, J. (2016). *The elements of statistical learning: Data mining, inference, and prediction* (second edition). New York: Springer VerlagSpringer.
- Roelands, M., Delden, A. van, & Windmeijer, D. (2017). Classifying businesses by economic activity using web-based text mining. CBS discussion paper 2017-18.
- Ruder, S. (2016) On word embeddings – part 1. Blog: <http://ruder.io/word-embeddings-1/>
- Van der Doef, S., Daas, P., & Windmeijer, W. (2018). Identifying innovative companies from their website. Abstract for BigSurv18 conference (ingediend).

Arnout van Delden is senior methodoloog bij het Centraal Bureau voor de Statistiek in Den Haag
E-mail: a.vandelden@cbs.nl

Piet Daas is senior methodoloog en lead data scientist bij het Centraal Bureau voor de Statistiek in Heerlen
E-mail: pjh.daas@cbs.nl

Olav ten Bosch is project manager bij het Centraal Bureau voor de Statistiek in Den Haag
E-mail: o.tenbosch@cbs.nl

Dick Windmeijer is data scientist bij het Centraal Bureau voor de Statistiek in Den Haag
E-mail: hjm.windmeijer@cbs.nl



Stephanie van der Pas

Jury report Willem R. van Zwet Award 2017

This year we received five nominations for the Van Zwet award. All five were high quality theses, and together they form a nice profile of our society by their variety of topics on Statistics and Operations Research. It was going to be a difficult choice for the jury. Nevertheless, the jury saw two theses standing out.

One of them is awarded as the runner-up: Krzysztof Postek, title of the thesis is *Distributionally and Integer Adjustable Robust Optimization*, supervised by professor Dick den Hertog from the Universiteit Tilburg. He defended his thesis in February 2017. Krzysztof's thesis is about new optimization models and algorithms that are robust with respect to all kinds of uncertainty in the data. The thesis is based on five papers, most of these are now published in high-ranked Operations Research journals. The jury was impressed by the scientific quality both theoretically and practically. And specifically the chapter where robust optimization is applied to stochastic programming problems is remarkable because it combines two major but rather different fields in OR.

The winner is Stephanie van der Pas, title of the thesis is *Topics in Mathematical and Applied Statistics*, supervised by professor Aad van der Vaart from the Universiteit Leiden. She defended her thesis also in February 2017. Stephanie's thesis is based on 7 papers that are spread over 4 topics in statistics: frequentist properties of Bayesian horseshoe prior; network analysis; sequentially collected data; and the last part is the application to survival analysis of hip replacements. All these papers found their way to international scientific journals. Specifically, the jury would like to mention chapter 3 which is based on a paper in which Stephanie shows that Bayesian methods do not correct automatically for multiplicity. This paper together with discussions of experts is published in *Bayesian Analysis*, and has already hundred plus citations. It will likely become an influential paper.

Jelle Goeman, chairman of the jury
e-mail: J.J.Goeman@lumc.nl

Jury rapport voor de Jan Hemelrijk Award 2017

Er waren dit jaar 4 aanmeldingen voor de Hemelrijk Award, gericht op de beste masterscriptie in Statistiek en Operation Research. We zien al een aantal jaar eenzelfde patroon hierbij: het gemiddelde cijfer van de 4 scripties was een 9,5, dus de eisen die de begeleiders stellen voordat ze een scriptie insturen zijn zeer hoog. Dit maakt het werk van de jury zowel lichter als plezieriger: hiervoor dank!

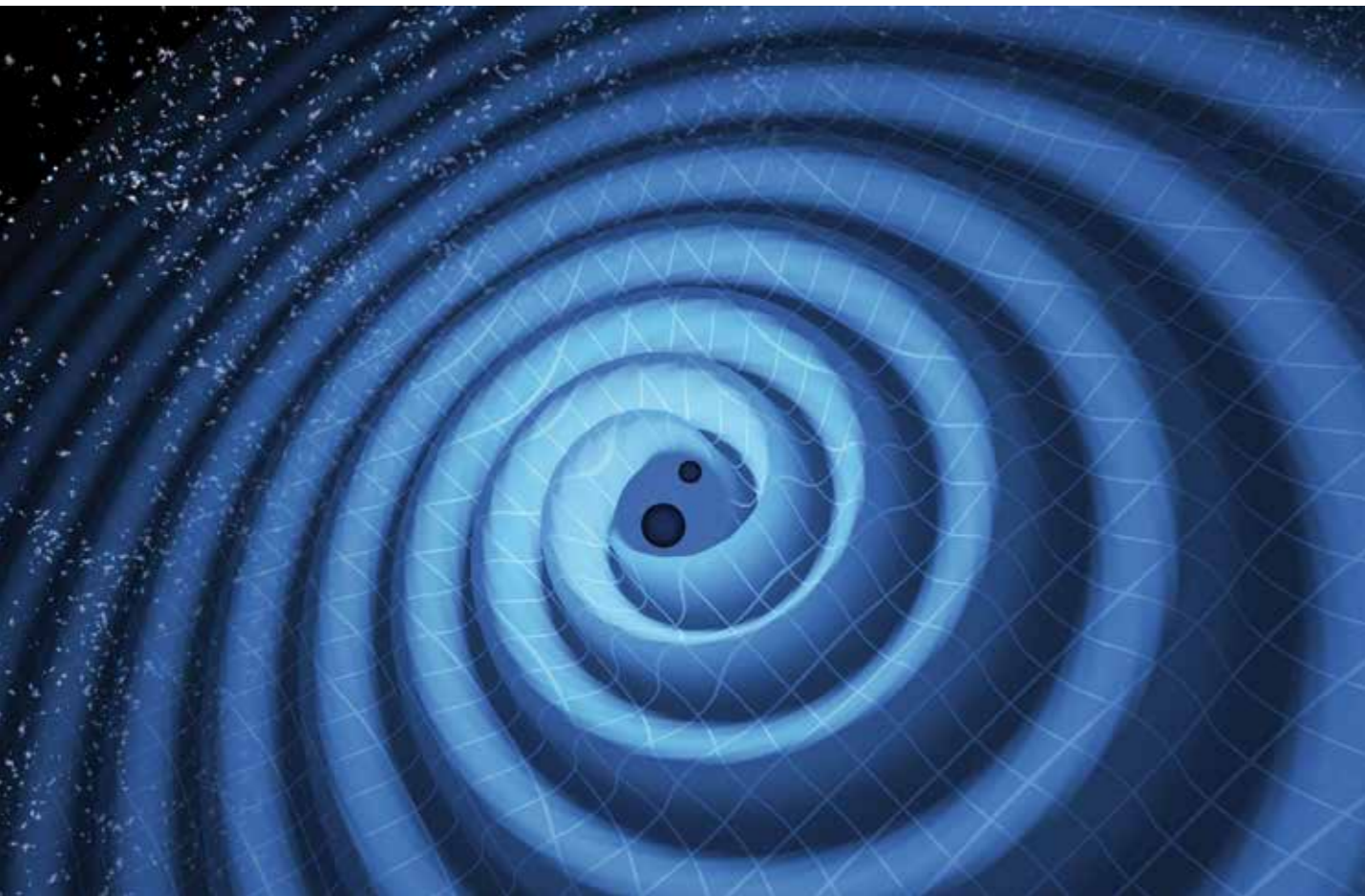
Een ander opvallend patroon is het volgende: hoewel de inzendingen allemaal van zeer hoog niveau zijn, is er ook dit jaar weer duidelijke overeenstemming binnen de jury over wie de Award verdient: een scriptie die op bijna alle punten beter scoort. Dit jaar is dat de scriptie van de winnaar van de Hemelrijk Award 2017: Stefan ten Eikelder!

Stefan heeft onderzocht hoe bij *radiation therapy* gekozen moet worden tussen fotonen- en protonenstraling. Deze worden op verschillende manieren opgenomen door het lichaam, waardoor ze ook op verschillende manieren schade aan gezond weefsel veroorzaken. U kunt zich voorstellen dat dit zeer relevant onderzoek is, en Stefans werk wordt dan ook nu al gebruikt in sommige ziekenhuizen om stralingstherapie te optimaliseren. Maar deze scriptie is niet alleen zeer relevant: om de gebruikte stralingsmethoden te kunnen optimaliseren binnen redelijke tijd, moest Stefan een nieuwe op maat gemaakte optimalisatiemethode ontwikkelen, gebaseerd op slimme heuristiek, maar ook getest op kwaliteit. Zo draagt Stefan ook bij aan de wiskunde in het algemeen. Al met al een zeer indrukwekkend resultaat en hij is dan ook een zeer verdiende winnaar van de Hemelrijk Award 2017!

Eric Cator, voorzitter van de jury
e-mail: e.cator@science.ru.nl



Stefan Eikelder



EEN EUREKA-MOMENT DAT NOG EEN BEETJE NATRILT

over zwaartekrachtgolven en Bayesiaans rekenen

CAJO TER BRAAK

Dit artikel gaat over een eureka-moment, mijn eureka-moment, dat leidde tot een Bayesiaanse rekenmethode die gebruikt is bij het analyseren van zwaartekrachtgolven. Deze zijn in 2015 voor het eerst gedetecteerd in het LIGO-experiment (Laser Interferometer Gravitational Wave Observatory), ongeveer 100 jaar nadat Albert Einstein ze voorspeld had (Ligo & Virgo, 2016). De Nobelprijs 2017 voor de Natuurkunde¹ ging naar de bedenkers en leider van dit experiment. Dat straalt toch een beetje af op de statistiek die hierbij is gebruikt en mijn indirecte bijdrage daaraan.

Zwaartekrachtgolven ontstaan door de botsing van twee zwarte gaten (of neutronensterren) en kan op aarde gemeten worden als een extreem kleine verandering van de lengte van een lange meetlat waar zo'n golf langs loopt. De meting is dus opnieuw dus een golf (trilling) maar nu gemeten als een relatieve lengte. Er is fysieke theorie hoe de parameters van de botsing – zoals de ligging en massa van de zwarte gaten, de massa na de botsing en de vrijgekomen energie – zich vertalen in de gemeten trillingen. En daarbij komen ook nog parameters over het meetproces (bijvoorbeeld over de dampkring van de aarde), in totaal 13 parameters. Kortom, er zijn data, een ingewikkeld model voor de data en, gelukkig, wat voorkennis over mogelijke waarden van parameters van het model. Bayesiaanse statistiek is dan bij uitstek geschikt om de initiële onzekerheid over de parameters (de a-priori-distributie) te integreren met de informatie uit de data (de aannemelijkheidsfunctie), om zo te komen tot nieuwe kennis over de parameters (de a-posteriori-distributie). De Bayesiaanse rekenmethoden zoals Markov Chain Monte Carlo (MCMC) representeren de a-posteriori-distributie in de vorm van een steekproef. Het schatten van parameters wordt daarbij eenvoudigweg het samenvatten van de verkregen steekproef. Afgeleide parameters en voorspellingen kunnen dan voor elk element uit de steekproef worden berekend, en de verkregen waarden samengevat in gemiddelde, standaardfout of credible interval. Eerst rekenen en dan samenvatten dus (Savage, 2012), zoals vereist door Jensen's ongelijkheid.² Dit stuk gaat over het ontstaan van een eenvoudige maar efficiënte methode (Differential Evolution Markov Chain, DE-MC) om een steekproef te trekken uit de a-posteriori-distributie.

Zoektocht

Een statistisch consultant moet zijn kennis over statistiek bijhouden. Zo had ik met collega's rond 2000 de eerste editie doorgenomen van Gelman's *Bayesian Data Analysis*. Dus ik wist iets van Bayesiaanse statistiek en Markov Chain Monte Carlo, en in het bijzonder het Metropolis algoritme (zie kader 1). Zo gingen we ook op 26 november 2003 naar een AIO-cursus Genetic algorithms, georganiseerd door het Wageningen Institute of Animal Sciences

Laat $\pi(\theta)$ de a-posteriori verdeling zijn die voorkennis en data y combineert met θ , een vector van parameters. Het Metropolis algoritme is een wandeling door de parameterruimte vanuit een gekozen startpunt voor θ , een gekozen stapgrootte c en een gekozen covariantiematrix Σ^* , vaak $\Sigma^* = I$. De initiële steekproef bevat alleen het startpunt θ .

Stap 1. Vanuit het huidige punt θ , genereer een voorstel voor een sprong naar een nieuw punt, $\theta^* = \theta + e$, met e een trekking uit een symmetrische verdeling, bijvoorbeeld $e \sim N(0, c^2 \Sigma^*)$.

Stap 2. Bereken het Metropolis quotiënt $r = \pi(\theta^*) / \pi(\theta)$.

Stap 3. Accepteer het voorgestelde punt θ^* ($\theta \leftarrow \theta^*$) met kans $\min(1, r)$, anders verwerp het (spring niet). Als $r \geq 1$, wordt het voorstel dus altijd geaccepteerd.

Stap 4. Voeg het huidige punt θ toe aan de steekproef.

Notitie:

Als $\pi(\theta)$ bij benadering multivariaat normaal verdeeld is, dan is de optimale optimale covariantiematrix $\Sigma^* = \text{cov}(\theta)$ en de optimale stapgrootte $c = 2,38 / \sqrt{p}$ met p het aantal parameters (Roberts en Rosenthal, 2001).

Kader 1. Het random walk Metropolis algoritme

en gegeven door een Australische Nederlander, Julius van der Werf. Hij maakte ons enthousiast voor de eenvoud en schoonheid van Differential Evolution (DE)³, een optimaliseringsmethode die bijvoorbeeld ook in Matlab zit. Waar standaardmethoden, zoals Gauss-Newton en varianten, werken met één oplossingsvector die steeds verbeterd wordt op basis van de eerste en tweede afgeleiden, werkt DE met een populatie van oplossingen die via recombinitie, mutatie en selectie evolueert naar de beste oplossing. In het practicum van de cursus werden we geacht een niet-lineaire regressie uit te voeren met Excel macro's. Mijn collega's vertrokken omdat we hiervoor al goede methoden kenden, maar ik bleef en verzond mijn

DE-MC werkt met een initiële populatie van N punten, $\theta_1, \theta_2, \dots, \theta_N$. Om beurten wordt een voorstel voor een vervanging van een punt gedaan en al dan niet geaccepteerd:

Stap 1. Voor punt θ_i ($i = 1, \dots, N$) is het voorstel

$$\theta_i^* = \theta_i + \gamma (\theta_{R_1} - \theta_{R_2}),$$

waarbij θ_{R_1} en θ_{R_2} twee aselekt gekozen punten zijn uit de huidige populatie en γ de stapgrootte.

Stap 2. Bereken het Metropolis quotiënt $r = \pi(\theta_i^*) / \pi(\theta_i)$.

Stap 3. Vervang θ_i door het voorstelde punt θ_i^* ($\theta_i \leftarrow \theta_i^*$) met kans $\min(1, r)$, anders verwerp θ_i^* (vervang niet).

Als $r \geq 1$, wordt het voorstel dus altijd geaccepteerd.

Stap 4. Voeg het huidige punt θ_i toe aan de steekproef.

Notities:

1. In DE wordt een voorstel alleen aangenomen als $r \geq 1$.

Dat maakt DE een optimalisatiemethode. Vaak wordt ook het startpunt voor een update random gekozen, maar dat kan niet in DE-MC omdat de sprongen omkeerbaar moeten zijn.

2. Het voorstel kan geschreven worden als $\theta_i^* = \theta_i + \gamma \varepsilon^*$ met $\varepsilon^* = \theta_{R_1} - \theta_{R_2}$. Het voorstel is symmetrisch want de random punten kunnen ook in omgekeerde volgorde gekozen worden. Na convergentie geldt $\text{var}(\gamma \varepsilon^*) = 2\gamma^2 \Sigma$ waarbij $\Sigma = \text{cov}(\theta)$. Als $\pi(\theta)$ bij benadering multivariaat normaal verdeeld is, dan is, gebruik makend van het stapgrootte resultaat in kader 1, de optimale stapgrootte $\gamma = 2,38 / \sqrt{(2p)}$. Deze stapgrootte is ook goed als de staarten van de verdeling zwaarder zijn. De reden is dat de staarten van het verschil ε^* dan ook zwaarder zijn.

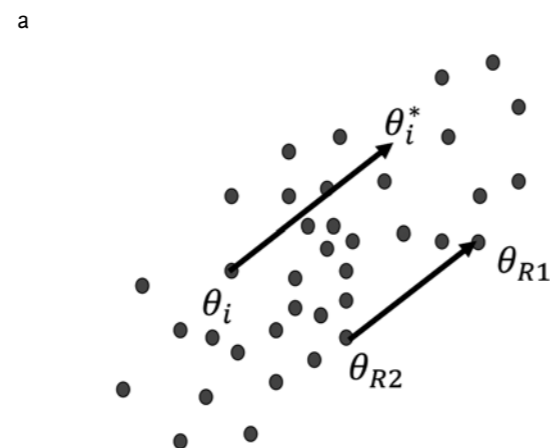
3. Met $\gamma \approx 1$ is het mogelijk om tussen modi van multimodale verdelingen te springen en zo het relatieve belang van het gebied rond elke modus in de steekproef tot uitdrukking te brengen (figuur 1b). In implementaties, wordt $\gamma \approx 1$ met kans 0,1 genomen en $\gamma = 2,38 / \sqrt{(2p)}$ met kans 0,9.

4. In tegenstelling tot de meeste adaptieve MCMC algoritmes, vormt DE-MC nog steeds een Markovketen. DE-MC is namelijk te zien als *Metropolis-within-Gibbs sampling* in S^N met S^N de parameterruimte: elk punt van p parameters wordt geupdate conditioneel op alle overige huidige punten (ter Braak en Vrugt, 2008). Net als Metropolis, is DE-MC is daarom moeilijk echt te paralleliseren; op een bescheiden wijze kan het overigens wel (ongepubliceerd werk).

eigen opgave: wat is de oplossing van $x^2=9$, oftewel, wat is het minimum van $(x^2-9)^2$? Een methode die met één oplossing werkt, geeft daarbij één antwoord, $x=3$ of $x=-3$. Ik had de hoop dat DE, met zijn populatie van oplossingen, in één run beide antwoorden zou kunnen geven. En dat werkte voor de meeste start-populaties.

Op de bank thuis die avond bedacht ik: als het algoritme twee waarden kan retourneren, dan moet het ook mogelijk zijn om het een hele verdeling van waarden te laten geven door een stap toe te voegen voor acceptatie/rejectie van voorstellen, precies zoals in het Metropolis algoritme. Eureka! In twee uur tijd ontwikkelde ik de basale theorie waarom de DE-voorstellen efficiënt zouden kunnen zijn (figuur 1a en kader 2). In DE is, naast de startpopulatie, de stapgrootte een belangrijke parameter die de gebruiker moet instellen. Altijd lastig. Maar in de MCMC-versie, DE-MC, heb ik die parameter afgeleid uit de theorie van de optimaal geschaalde random walk Metropolis van Roberts en Rosenthal (2001). Van die optimale stapgrootte moet overigens met een kleine kans worden afgeweken om ook multimodale distributies aan te kunnen (figuur 1b), iets waar gewone random walk Metropolis heel slecht in is. Daarna volgde een periode van hard werken: klopt het, werkt het en ..., bestaat het al?

Zoekend in de literatuur leerde ik dat er zoiets bestond als populatie MCMC en evolutionaire MCMC. Wat later zag ik adaptive direction sampling (Gilks et al. 1994) dat er heel dichtbij komt, maar moeilijker is dan DE-MC en ook nog een extra vereiste heeft die moeilijk te verifiëren is. Het idee om genetische of evolutionaire algoritmes te koppelen aan MCMC was dus al eerder geopperd, maar

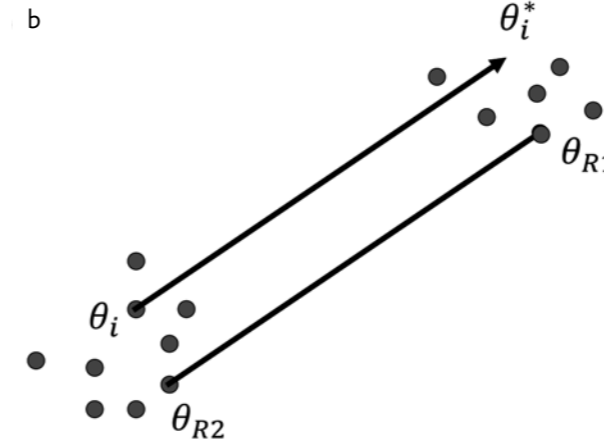


Figuur 1. DE-MC in twee dimensies ($p = 2$) met $\gamma = 2,38 / \sqrt{(2 \times 2)} = 1,2$ (a)

de eenvoudige combinatie die ik had gevonden niet. Na een presentatie op de ISBA-workshop over adaptieve MCMC in Bormio (Italië) in 2005, is het eerste artikel over DE-MC gepubliceerd in 2006 in *Statistics & Computing* (Ter Braak, 2006). In alle enthousiasme voor de methode was ik ook wat blind voor de nadelen. Gelukkig was ik in staat die in een tweede artikel in 2008 in *Statistics & Computing* grotendeels te ondervangen (Ter Braak & Vrugt, 2008). De methode is beschikbaar in het R-pakket BayesianTools.

Verzilveren

Hoe is DE-MC terecht gekomen in de astronomie en de analyse van zwaartekrachtgolven? Dat weet ik niet, maar ik kwam er via een omweg wel achter dat het gebruikt werd. Ik las in het tijdschrift *Significance* van de Amerikaanse en Britse verenigingen voor de statistiek, het verhaal 'Gravitational waves: a statistical autopsy of a black hole merger' van Renate Meyer en Nelson Christensen (2016). In dat prachtige verhaal staat de rol van adaptieve MCMC beschreven bij de analyse van de zwaartekrachtgolven met een verwijzing naar een software bibliotheek hiervoor, beschreven in een artikel in *Physical Review D* (Veitch et al., 2015). Was ik die naam, Veitch, niet eerder tegengekomen in de citatie-attending van Scopus? Dat vermoeden was aanleiding genoeg het artikel te bestuderen. Het is een dik artikel, maar waar MCMC-voorstellen besproken worden is het eerste voorstel mijn DE-MC voorstel uit 2006. In hun aanpak wordt het gemengd met



en $\gamma = 1$ (b)

een aantal andere waardevolle voorstellen. Bij navraag, schreef John Veitch dat ze DE-MC kozen vanwege zijn 'natuurlijke' staprichting en stapgrootte, tezamen met de mogelijkheid om multimodale verdelingen te kunnen verkennen.

Via John Veitch kwam ik in contact met Jonathan Goodman en Jonathan Weare die minstens net zo enthousiast als ik bleken, toen ze methoden vonden die invariant waren onder affine transformatie van de parameter-ruimte (Goodman & Weare, 2010). Hun paradepaard is de 'stretch move', die nu ook populair is in de astronomie. Het aardige is dat DE-MC ook tot die invariante familie behoort. Gek genoeg zijn de methoden tot op heden nog niet systematisch met elkaar vergeleken. Dat vergt meer dan een eureka-moment.

NOTEN

1. http://www.nobelprize.org/nobel_prizes/physics/laureates/2017/press.html
2. https://nl.wikipedia.org/wiki/Ongelijkheid_van_Jensen
3. https://en.wikipedia.org/wiki/Differential_evolution

LITERATUUR

- Gilks, W.R., Roberts, G.O., & George, E.I. 1994. Adaptive Direction Sampling. *The Statistician*, 43, 179–189.
- Goodman, J., & Weare, J. 2010. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science*, 5, 65–80.
- Ligo & Virgo Collaborations. 2016. Observation of Gravitational Waves from a Binary Black Hole Merger. *Physical Review Letters*, 116, 061102.
- Meyer, R., & Christensen, N. 2016. Gravitational waves: A statistical autopsy of a black hole merger. *Significance*, 13, 20–25.
- Roberts, G.O., & Rosenthal, J.S. 2001. Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science*, 16, 351–367.
- Savage, S.L. 2012. *The flaw of averages: Why we underestimate risk in the face of uncertainty*. Hoboken, NJ: Wiley.
- Ter Braak, C.J.F. 2006. A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. *Statistics and Computing*, 16, 239–249.
- Ter Braak, C.J.F., & Vrugt, J.A. 2008. Differential Evolution Markov Chain with snooker updater and fewer chains. *Statistics and Computing*, 18, 435–446.
- Veitch, J., et al. 2015. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. *Physical Review D*, 91, 042003.

CAJO TER BRAAK is persoonlijk hoogleraar Multivariate Analyse voor de levenswetenschappen bij Biometris, Wageningen University & Research.
E-mail: <cajo.terbraak@wur.nl>

Kader 2. Differential Evolution Markov Chain (DE-MC)



PLANNING VAN MULTIDISCIPLINAIRE KANKERZORG

Multidisciplinaire zorgprocessen zijn aan de orde van de dag in de kankerzorg. De organisatie van deze zorg is complex, en veroorzaakt vaak lange wachttijden voor patiënten en hoge fluctuaties in werkdruk voor de betrokken zorgprofessionals. Dit onderzoek laat zien dat door de agenda-inrichtingen van artsen op elkaar af te stemmen en slim rekening te houden met de variatie en onzekerheid aan diagnoses en behandelmogelijkheden voor patiënten een multidisciplinaire planning kan worden bereikt die voor patiënt én zorgverlener beter is.

GRÉANNE LEEFTINK & ERWIN HANS

In de oncologie is multidisciplinaire zorg aan de orde van de dag. Door vergaande specialisatie van zorgverleners en geavanceerde nieuwe technologieën en onderzoek, is het mogelijk om patiënten zorg op maat te bieden. Dit betekent dat meerdere zorgprofessionals nauw met elkaar en de patiënt moeten samenwerken om de beste keuzes te maken voor de gezondheid van de patiënt, maar dat het vaak niet meteen duidelijk is welke zorgverleners dit precies moeten zijn voor een patiënt.

In dit onderzoek staat een multidisciplinaire polikli-

niek centraal, waar de patiënt zijn/haar uitslaggesprek heeft en een behandelplan wordt opgesteld (vaak weet de patiënt al dat hij/zij kanker heeft, en gaan de afspraken met name in op de staging en behandelmogelijkheden). Dit gebeurt naar aanleiding van het multidisciplinair overleg (MDO), waar de betrokken artsen gezamenlijk de diagnose en een mogelijk behandeltraject bepalen. Om de patiënt zo snel mogelijk op de hoogte te brengen van de uitslag en de mogelijkheden, vinden de afspraken van de patiënt bij de polikliniek direct volgend op het MDO

plaats. Kortom, het MDO valt in de lunchpauze en in de middag spreekt de patiënt de betrokken artsen.

Allereerst heeft de patiënt een uitslaggesprek waarin de diagnose en de behandelmogelijkheden worden toegelicht. Na dit uitslaggesprek, vindt een tweede consult plaats met de gekozen behandelend arts. Bijvoorbeeld, stel dat op een zekere dinsdag tijdens het MDO is bepaald dat chemotherapie de beste optie is voor een patiënt, dan krijgt deze patiënt diezelfde dinsdagmiddag eerst een gesprek met een arts waarin hem/haar de uitslag van het MDO wordt verteld, waarna diezelfde middag nog een volgend gesprek plaatsvindt met de oncoloog over de behandeling. Zo is de patiënt direct op de hoogte van het vervolgtraject, en zijn de slapeloze nachten in onwetendheid voor de patiënt minimaal.

In dit systeem is de mogelijke behandeling, en dus de behandelend arts bij wie de patiënt moet worden ingepland, nog onbekend tot op het MDO. Dit introduceert een onvoorspelbaarheid in de planning voor de artsen, en maakt het lastig voor de planners om goed in te schatten in welke volgorde de patiënten voor hun uitslaggesprek moeten worden uitgenodigd.

Naast de multidisciplinaire patiënten zien de artsen ook nog andere patiënten, afhankelijk van de verwachte drukte en de mogelijk zorgpaden van de multidisciplinaire patiënten. Een consult met een radiotherapeut is bijvoorbeeld maar voor een klein aantal van de multidisciplinaire patiënten relevant, terwijl een groot aantal patiënten langs de chirurg gaat. Dat betekent dat de radiotherapeut meer ruimte heeft in zijn agenda om andere patiënten in te plannen. De uitdaging is om deze andere patiënten in de juiste aantallen op de juiste momenten uit te nodigen, aangezien zij graag minimaal een week van tevoren uitgenodigd willen worden.

Het doel van het onderzoek is om een agenda-inrichting voor de betrokken zorgprofessionals te creëren, waarin de wachttijd van de multidisciplinaire patiënten wordt geminimaliseerd, de overtijd in de polikliniek wordt geminimaliseerd, en tegelijkertijd de benutting van de zorgprofessionals wordt gemaximaliseerd. De agenda-inrichting moet aangeven welke tijdsintervallen vrijgehouden moeten worden voor mogelijke multidiscipli-

naire patiënten en in welke tijdsintervallen de arts een andere patiënt kan plannen.

Huidige situatie

Bij het inrichten van de agenda van een dergelijke polikliniek is men in de praktijk gewend om te werken op basis van gemiddelden, gecombineerd met de ervaringen en wensen van de artsen. Dit heeft meerdere consequenties. Ten eerste wordt elke arts individueel gevraagd naar zijn/haar geprefereerde agenda-indeling. Dit heeft tot gevolg dat agenda's niet op elkaar zijn afgestemd, en dat artsen veelal in het begin van de middag een aantal tijdsintervallen vullen met andere patiënten, zodat de benutting niet in het gedrang komt. Dit leidt tot langere wachttijden voor multidisciplinaire patiënten en tot uitloop bij artsen die achteraan het zorgpad zitten. Ten tweede wijkt in de praktijk elke dag af van de gemiddelde dag. De agenda is alleen niet ingericht om op deze niet-gemiddelde dagen goed te presteren. Dit betekent dat er niet alleen op drukke dagen een hoge werkdruk en lange wachttijden ontstaan, maar ook regelmatig op dagen die minder druk zijn dan gemiddeld.

Om deze problemen op voorhand te voorkomen, hebben wij samen met UMC Utrecht onderzocht hoe de agenda's van de betrokken zorgprofessionals van twee nieuw op te zetten poliklinieken moeten worden ingericht.

Het model

Bij het zoeken naar een geschikte manier om dit probleem te modelleren, is het belangrijk om de toepassing in het oog te houden, zodanig dat de uitkomsten eenvoudig kunnen worden verklaard aan de medewerkers op de werkvloer. Een van de redenen hiervoor is dat het voor een optimale agenda-inrichting belangrijk is om mee te nemen dat de agenda van de ene zorgverlener invloed heeft op de prestaties bij de volgende zorgverlener. Dit vraagt niet alleen om een solide wiskundige aanpak, maar ook

om een goede uitleg in de zorgpraktijk, aangezien artsen over het algemeen niet gewend zijn om rekening te houden met de agenda's van hun directe collega's.

Door het opstellen van een ILP, waarbij de verschillende aankomstscenario's van de multidisciplinaire patiënten meegenomen worden, zoeken we de optimale agenda-inrichting voor alle artsen. Hiertoe hebben we een gewogen doelfunctie van drie prestatie-indicatoren meegenomen: de wachttijd voor patiënten tussen hun uitslag- en behandelconsult, de leegstand bij zorgprofessionals, en de overtijd per zorgprofessional.

In het stochastische ILP worden twee keuzes gemaakt, één op operationeel niveau, en één op tactisch niveau, die met hetzelfde model worden opgelost. Allereerst wordt op operationeel niveau de volgorde waarin de patiënten gezien worden bij het uitslaggesprek geoptimaliseerd. Deze beslissing kan voorafgaand aan de polikliniek (maar na het MDO) worden gemaakt, waardoor een optimale toewijzing kan worden gemaakt voor elk gerealiseerd aankomstscenario van multidisciplinaire patiënten.

Ten tweede wordt op tactisch niveau de agenda-inrichting van de betrokken zorgprofessionals ingericht. Deze agenda's moeten tijdig (zeg 6 week van tevoren) worden vastgezet, waardoor een optimale inrichting moet worden bepaald op basis van de mogelijke aankomstscenario's.

Omdat het resulterende model niet binnen afzienbare tijd oplosbaar is, hebben we gebruik gemaakt van intuïtieve approximatiemethode om tot een goede oplossing te komen, de Sample Average Approximation approach. In deze sampling-methode wordt een selectie van de aankomstscenario's meegenomen om een agenda-inrichting te bepalen.

Resultaten

Om tot een realistische agenda-inrichting voor de poliklinieken te komen, hebben we op basis van historische data de aankomstverdeling van de verschillende typen patiënten en de doorverwijspercentages naar de verschillende behandelend artsen bepaald. Samen met de capaciteit van de polikliniek geeft dit de inputparameters voor het model.

Om de resultaten van het model te kunnen vergelijken met de verwachte resultaten in de praktijk wanneer er geen planningsoplossing was geboden, vergelijken we onze resultaten voor de polikliniek met de optimale agenda-inrichting op basis van het gemiddelde aankomstscenario (wat naar verwachting een overschatting

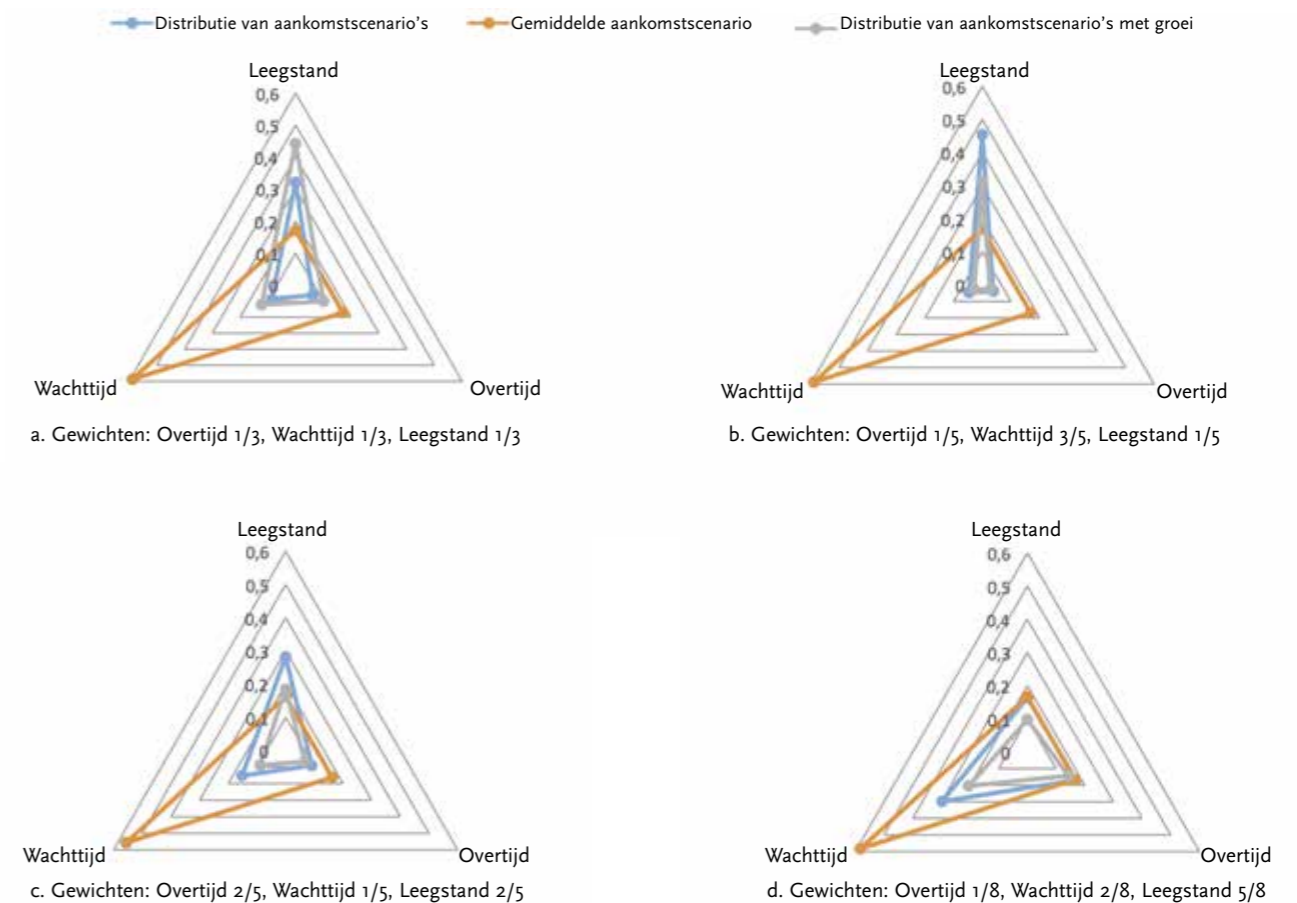
is van de prestatie in de praktijk, aangezien we er hierbij wel vanuit gaan dat de patiënten op operationeel niveau altijd optimaal gepland worden). Hierbij zien we een significante verbetering van 21% in de doelfunctie. Zoals gezegd is deze doelfunctie een gewogen gemiddelde van wachttijd, overtijd en leegstand, die we willen minimaliseren. De prestatie op elk van deze indicatoren voor vier verschillende wegingen is weergegeven in figuur 1. Hierin is de gemodelleerde situatie weergegeven (voor een polikliniek met capaciteit voor 16 multidisciplinaire patiënten per middag), een groeiscenario (waarin het aantal multidisciplinaire patiënten verdubbelt, net als de hoeveelheid medewerkers, in verband met regionale samenwerking), en de huidige situatie op basis van het gemiddelde. De prestatie is genormaliseerd naar de capaciteit van de polikliniek.

De figuren laten zien dat onafhankelijk van de toewijzing van gewichten aan een prestatie maat, het gemiddelde scenario altijd leidt tot een slechte prestatie op het gebied van wachttijd voor patiënten, en dat in het gemiddelde scenario de leegstand (onbewust) wordt geprioriteerd. Dit komt door het hogere aantal reguliere patiënten dat ingepland wordt, want het verschil in overtijd en wachttijd is groter dan in het door het model voorgestelde alternatief. In het geval de leegstand daadwerkelijk een belangrijke prestatie maat is, zoals weergegeven in Figuur 1d, dan kun je zien dat het model een betere agenda-inrichting voorstelt, terwijl hetzelfde aantal patiënten gezien wordt in de polikliniek.

Naast een verbetering van de huidige planningspraktijk, geven de resultaten ook inzicht in de gevolgen van schaalvergroting door een mogelijke regionale samenwerking. Hierin kun je zien dat ondanks dat de grootte van de polikliniek verdubbelt (zowel in patiënten als medewerkers), de overtijd, wachttijd en leegstand niet verdubbelen. Door de schaalvergroting treedt het pooling-effect op, waardoor efficiënter gebruik kan worden gemaakt van de beschikbare capaciteit.

En nu?

Dit onderzoek laat zien dat ziekenhuizen grote winst kunnen behalen bij het inrichten van multidisciplinaire poliklinieken door het gebruik van slimme plannings technieken. Het opstellen van een stochastische ILP met twee fasen (een operationele en een tactische fase) heeft als voordeel dat de operationele fase elke dag in de praktijk kan worden opgelost. Implementatie hiervan in de zorgpraktijk vraagt een integratie van het model in het



Figuur 1. Prestatie voor vier mogelijke wegingsscenario's

gebruikte informatiesysteem, wat voor nu nog niet mogelijk is in het UMC Utrecht. Wel is op basis van de resultaten van dit onderzoek een agenda-inrichting gemaakt voor de betrokken artsen, die recent geïmplementeerd is in de praktijk.

In de zorg bestaan allerlei variaties van multidisciplinaire poliklinieken. In vervolgonderzoek willen we het model uitbreiden met zorgpaden met meer dan één behandelconsult. Dit komt regelmatig voor, bijvoorbeeld wanneer een patiënt een gecombineerd behandeltraject van chirurgie en chemotherapie of radiotherapie wordt aangeboden. In dat geval wil je de patiënt zowel een gesprek met de chirurg als de medisch oncoloog of radiotherapeut aanbieden. Ook kan het zorgpad nog uitgebreid worden met paramedische zorg, zoals een diëtist of fysiotherapeut. Een eerste analyse met behulp van computersimulatie laat zien dat eenzelfde exercitie voor zorgpaden met 3 of meer afspraken kan resulteren in grote verbeteringen in de balans tussen werkdruk en wachttijd. Meer hierover kunt u lezen in Hoofdstuk 6 en 7 van het proefschrift van Leeftink (2017).

Over dit onderzoek is eerder gepubliceerd in *Health Care Management Science*.

LITERATUUR

- Leeftink, A.G. (2017). *Why wait? Organizing integrated processes in cancer care. PhD thesis*. Enschede: University of Twente.
 Leeftink, A.G., Vliegen, I.M.H., & Hans, E.W. (2017). Stochastic integer programming for multi-disciplinary outpatient clinic planning. *Health Care Management Science*, 1–15.

GRÉANNE LEEFTINK is universitair docent aan de Universiteit Twente binnen onderzoekscentrum CHOIR (Center for Healthcare Operations Improvement & Research) en de vakgroep Industrial Engineering and Business Information Systems. Ze is in december 2017 gepromoveerd bij CHOIR aan dezelfde universiteit. In haar onderzoek richt ze zich op de inrichting en optimalisatie van processen in de kankerzorg, waarvoor ze tijdens haar promotieonderzoek parttime in het UMC Utrecht was aangesteld.
 E-mail: a.g.leeftink@utwente.nl

ERWIN HANS is hoogleraar Operations Management in de Zorg aan de Universiteit Twente binnen de vakgroep Industrial Engineering and Business Information Systems. Daarnaast is hij mede oprichter van onderzoekscentrum CHOIR.
 E-mail: e.w.hans@utwente.nl



Annouk van der Weijden. Foto: Stichting LOOT

NEDERLANDSE STATISTIEK IN THE WALL STREET JOURNAL

Sommige columnisten zeggen wel eens dat hun column zich 'vanzelf schrijft'. Zo gaat dat bij mij niet, maar ditmaal hoefde ik er toch weinig aan te doen, *The Wall Street Journal* schreef hem al voor mij.

Dit gerenommeerde blad toonde zich onlangs opgetogen over het eclatante succes van de Nederlandse schaatsers en schaatssters tijdens de Olympische Winterspelen in februari dit jaar.

'The Netherlands cemented their place as the greatest speed skating nation on earth by leaning on university statisticians. It's Winter Olympics Moneyball (...) for the team with the most money and talent. (...) the Dutch took 23 of 36 possible medals in Sochi and, as of Friday morning, grabbed 13 of 33 here, including

six of the 11 available gold. Leaning on statisticians at the University of Groningen, the team developed an algorithm to tell the coaches how best to deploy their squad of 10 skaters across the 14 Olympic medal events to maximize their chances of winning gold.'

Aldus Joshua Robinson in *The Wall Street Journal* van 25 februari 2018. Het werd uiteindelijk acht keer goud, zes keer zilver en zes keer brons voor Nederland met een vijfde plaats in de medaillering voor landen. Zelf ben ik een van de Nederlandse statistici die de KNSB hebben ondersteund bij het opstellen van de Olympische selectieprocedure. De andere is Bertus Talsma van ORTEC/Sports. Hieronder het artikel van Robinson in *The Wall Street Journal*.

OLYMPICS

The Secret of Dutch Speed skating—It's Not What You Think

By Joshua Robinson
Febr 23, 2018

"We needed a kind of order," said Arie Koops, the technical director for the Dutch national speed skating program. "So how would we do that?"

In classic Dutch fashion, they chose to devise the most rational solution they could imagine. Leaning on statisticians at the University of Groningen, the team developed an algorithm to tell the coaches how best to deploy their squad of 10 skaters across the 14 Olympic medal events to maximize their chances of winning gold. It bakes in two years' worth of results, calibrates them for location, since ice conditions and altitude affect speed skating times, and compares them to other results around the world.

The answers that the computer spits out form the basis of the Dutch strategy. It's Winter Olympics Moneyball, only for the team with the most money and talent.

Over and over, the skaters have proved the model right. Since refining it ahead of the 2014 Olympics, the Dutch took 23 of 36 possible medals in Sochi and, as of Friday morning, grabbed 13 of 33 here, including six of the 11 available golds.

"We thought it's never possible to do something like that again" after Sochi, Dutch speed skater Annouk van der Weijden said. "And it's happening again. That's really weird."

The Dutch might be the only ones who don't think a repeat performance makes perfect sense. No other country is so successful at just one Olympic sport as the Netherlands. South Koreans are the undisputed kings of short track, but their history with the sport isn't as long—it only became an Olympic event in 1992. The Norwegians blow away the competition in cross-country events, but they are also good at other things. But the Dutch? They know their Olympic gold mine is inside the speed skating oval.

As of Friday morning, 118 of their 127 all-time Winter Olympics medals have been in long-track speed skating. Only one Dutch person, a snowboarder in Sochi four years ago, has ever won the country a Winter Games medal without wearing skates.

And, because speed skating also puts more medals on offer than any other sport, the Netherlands inevitably soar up the medal standings. Sitting on 17, they were ahead of winter powerhouses like Austria and Switzerland. "Now the competition is harder than in Sochi," Koops said. "But still, we're doing our job."

The heaviest lifting occurred before anyone even got on the plane to South Korea, back at the Netherlands' Olympic trials. Most countries would simply take the top one or two skaters at a given distance, but the Dutch see that as hopelessly inefficient. Using the matrix, they handicap every Olympic race to figure how where they can best trounce the competition.

This wouldn't be necessary if the Dutch didn't already have the deepest well of talent in world speed skating. With elite performers to select from at every event, they consider themselves to be the Winter Olympics' equivalent of USA Track and Field, the self-styled "Hardest Team to Make."

"There are a lot of speed skaters from the Netherlands competing here right now," van der Weijden said. "But there are a lot of speed skaters at home, who are just as good as some of them here."

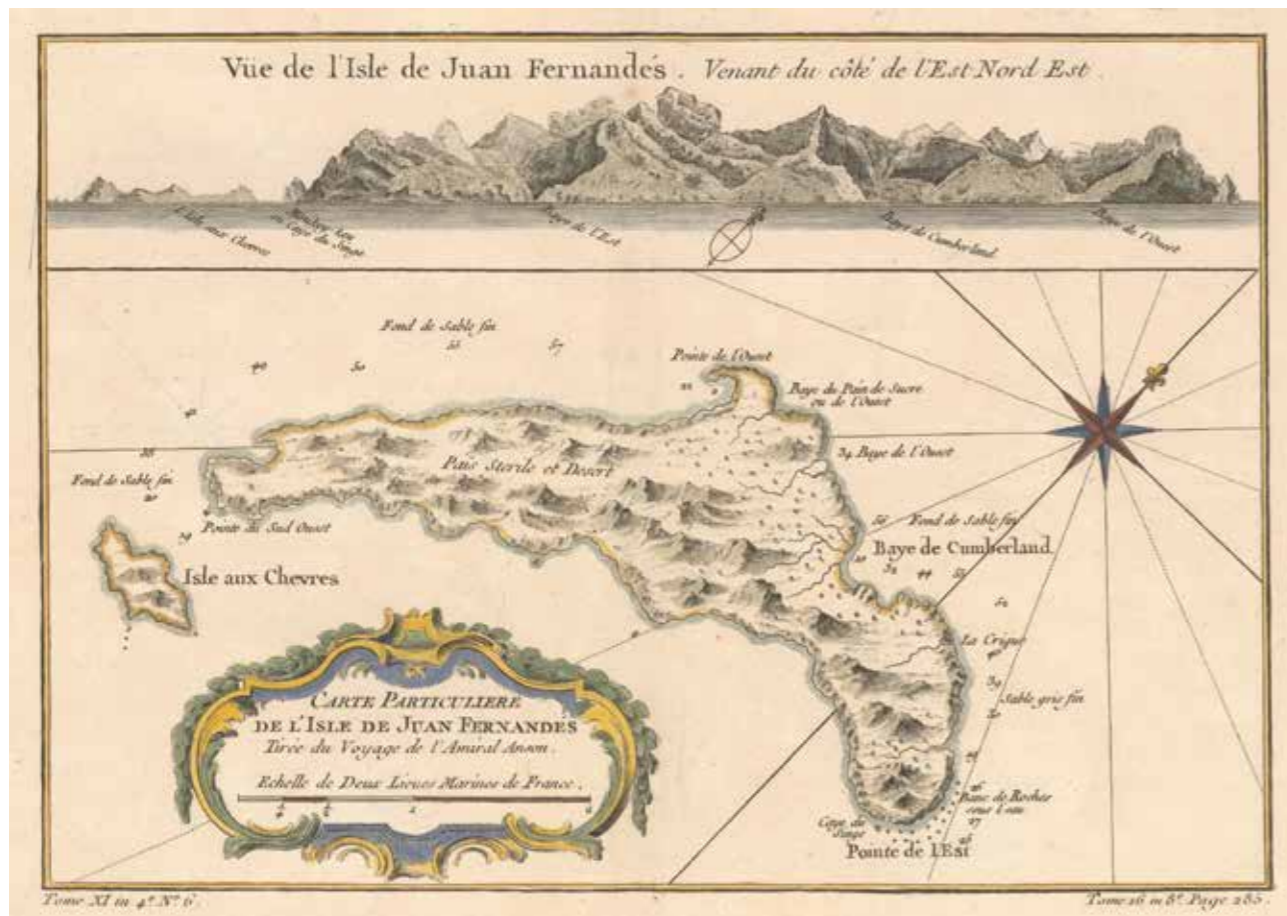
That's why their Olympic trials, held every four years in the week between Christmas and New Year's, become the Netherlands' most significant speed skating battle. The five days of cutthroat racing inevitably culminate in a 10,000-fan sell-out for the final session, all to see who makes the cut for what the Dutch call the Order of Selection. Plenty of Olympic medallists don't.

"If you're not good at that moment, the end of December, then you're out," said van der Weijden, a specialist of the longer distances. "All the training, four years, is already done even before you can go to the Olympics. It's even more important than this competition."

LITERATUUR

Zie ook <https://www.thiscomplexworld.com/ortec-helps-dutch-gold-rush-sotsji>
Sierksma, G. (2017). Bhaiparthait en Pyeongchang; Olympische schaatsstatistiek met prijsvraag. *STATOR*, 18(4), 32–33.

GERARD SIERKSMA is emeritus hoogleraar Kwantitatieve Logistiek en Sportstatistiek aan de Rijksuniversiteit Groningen. E-mail: g.sierksma@rug.nl



Kaart uit 1753 van Isla Más a Tierra – beter bekend als het eiland Robinson Crusoe – is het grootste eiland binnen de Juan Fernández-archipel en ligt in de Zuidelijke Grote Oceaan op 674 kilometer van Chili

EXPEDITIE ROBINSON

VAN ECOLOGISCHE CORRELATIE NAAR MULTI LEVEL ANALYSE

RICHARD STARMANS

Structuren in een gedataficeerde wereld

In tijden van data science en big data is het bon ton krachtige hyperbolen niet te schuwen. Zo heeft zich volgens ingewijden een data-revolutie voltrokken, zijn we getuige van een data-explosie of – beter nog – een data-tsunami en leven we onmiskenbaar in een gedataficeerde wereld. De alomtegenwoordigheid en diversiteit van data en de snelheid waarmee zij beschikbaar komen, geven voeding aan de gedachte dat de werkelijkheid in al haar complexiteit nagenoeg geheel wordt gevat, gecodeerd en gecodificeerd in data. Deze voorstelling gaat deels terug tot 19e eeuwse opvattingen van onder meer Ernst Mach (1838-1916) en vooral Karl Pearson (1857-1937). Laatstgenoemde deed recht aan variatie en verandering in de natuur door deze niet te identificeren in errors, maar in de ver-

schijnselen zelf (gecodeerd in data), en terug te voeren tot verschillende (klassen van) kansverdelingen. Hij zag in dat vele verschijnselen niet normaal, maar scheef waren verdeeld en met behulp van vier parameters (gemiddelde, variantie, scheefheid en welving) konden worden beschreven. Kansverdelingen met bijbehorende parameters werden de objecten van de wetenschap. Als eerste gaf Pearson daarmee kansverdelingen een volwaardige plaats in de wetenschap en zag de wereld op een niveau van abstractie waarbij data, variatie in data, data genererende mechanismen en parameters van de kansverdelingen de werkelijkheid veeleer coderen en opbouwen en niet zozeer een (vermeende) fysische werkelijkheid representeren of afbeelden. Die laatste voorstelling krijgt enige plausibiliteit tegen de achtergrond van het feit dat de zeer

empiristische en anti-causalistische Pearson opmerkelijk genoeg geen materialistische, maar een idealistische denker was. De werkelijkheid geldt primair als een constructie van de menselijke geest en het doel van wetenschap is een inventaris te maken van de inhoud daarvan (Pearson, 1892). Het oude Kantiaanse idee dat het kenvermogen zelf een structuur oplegt aan de werkelijkheid maakt het idee van een volledige datafificatie minder buitenissig.

Een en ander werpt een nieuw licht op het traditionele filosofische *dualisme* van taal en werkelijkheid en de daarbij gekoesterde correspondentietheorie van de waarheid, maar ook op de methodologische queeste naar het ultieme valide meetinstrument en het wiskundige streven naar een adequate *representatiestelling*. Aan de ene kant zijn er de complexiteit en dynamiek van de werkelijkheid in al haar verschijningsvormen, processen, variatie en verandering, onderliggende structuren, inclusief de *wisselwerking van entiteiten met de omgeving* waarin zij zich bevinden c.q. de groepen waartoe zij behoren. En uiteraard de voorstellingen die wij van die werkelijkheid maken, de gepostuleerde theoretische/abstracte entiteiten en de verhouding tot de al dan niet vermeende waarneembare entiteiten, en de taal waarin we daarover spreken. Aan de andere kant zijn er de data, die aan dit alles recht moeten doen en waaraan dus hoge eisen moeten worden gesteld: betrouwbaarheid, validiteit, kennis van het data genererende mechanisme en vooral de structuur in de data. Met het oog op dit laatste zij hier opgemerkt dat statistici in de loop der tijden uitermate bedreven en volhardend zijn gebleken in het zoeken naar en creëren en analyseren van nieuwe *bronnen van varia(n)tie* door het ontwikkelen van complexere methoden en technieken. We beperken ons hier tot één aspect van dit alles, de voornoemde *wisselwerking van individuen met de omgeving*. Statistici herkennen hierin data die doorgaans als *hiërarchisch* worden opgevat; zij zijn genest, ingebed, of gelaagd, dat wil zeggen zij kennen verschillende niveaus. Een traditionele, maar geenszins exclusieve *interpretatie* (of gepostuleerde correspondentie met de werkelijkheid) hiervan, betreft de opvatting dat deze data soms eigenschappen beschrijven van *individuen*, min of meer 'ondeelbare', singuliere entiteiten, die deels invariant zijn, op zichzelf staan en een eigen persoonlijke identiteit krijgen toebedeeld. Soms beschrijven de data eigenschappen van groepen of clusters, waartoe deze entiteiten behoren, waarin zij zijn ondergebracht of georganiseerd, of waaruit zij voortkomen. Deze groepen 'bepalen' hun afkomst, soort, status,

omgeving, hetzij langs natuurlijke, hetzij langs artificiële of institutionele weg. Zo kan de inbedding plaatsvinden op basis van fysische, biologische, sociale of demografische criteria. De groepeigenschappen worden dikwijls beschreven met *geaggregeerde* data, dat wil zeggen data die ontleend zijn aan – en dus op hun beurt afhangen van – metingen verricht aan voornoemde singuliere entiteiten; gemiddelden, proporties, indexcijfers of correlaties. Omgekeerd kunnen eigenschappen van voornoemde entiteiten weer afhankelijk zijn van karakteristieken van de bewuste groepen. Zo beschouwd spelen de wisselwerking, interdependenties en interacties tussen de verschillende *levels* in vrijwel alle wetenschapsgebieden een rol, variërend van biologie, ecologie en epidemiologie tot sociologie, economie en robotica. De statisticus die recht wil doen aan de voornoemde wisselwerkingen en verschillende 'niveaus' simultaan in een analyse wil betrekken, beschikt vandaag de dag over een groot aantal technieken die hier gemakshalve als *multilevel* analyse zullen worden getypeerd. Of het nu gaat om multilevel models in strikte zin, hierarchical (linear) models, nested data, mixed models, klassieke split-plot designs, random coefficient of random-effect models, repeated measures, et cetera. Strikte definitie- en demarcatiekwesties blijven hier buiten beschouwing. Het gaat erom dat het zoeken naar en creëren van nieuwe *bronnen van varia(n)tie* leidt tot modellen met meer random effects, minder fixed effects en een meer gesofisticeerde analyse van residuele error, dit alles binnen en tussen de verschillende niveaus.

De anatomie van de ecologische correlatie

In handboeken en overzichtswerken op het gebied van multilevel analyse wordt doorgaans benadrukt dat de problematiek een respectabele genealogie kent. Dikwijls betonen de auteurs zich daarbij op de een of andere wijze schatplichtig aan het werk van de Amerikaanse statisticus William S. Robinson (1913-1996). In 1950 verscheen van zijn hand het amper vijf bladzijden tellende artikel 'Ecological Correlations and the Behaviour of Individuals' in de *American Sociological Review*. Het behoort vandaag de dag tot de meest geciteerde papers in de methodologie van de sociale wetenschappen en is uiterst relevant in het licht van de hier geschetste problematiek. Een aantal merkwaardige aspecten van het artikel bleef lang onopgemerkt ondanks het grote aantal citaties en de nagenoeg

universele bijval die de publicatie ten deel viel. Deze betreffen zowel kleine fouten als inhoudelijke inconsequenties van het artikel. Enkele facetten ervan passeren hier kort de revue.

Robinson licht de thematiek toe aan de hand van een tweetal eenvoudige maar saillante voorbeelden, ontleend aan data van de volkstelling in de VS uit 1930. Allereerst onderzoekt hij het verband tussen etniciteit (blank versus zwart) en analfabetisme. Robinson berekent om te beginnen een zogenaamde ecologische correlatie: *'The statistical object is a group of persons. The thing described is the population of a state, and not a single individual. The variables are percentages, descriptive properties of groups, and not descriptive properties of individuals.'* (Robinson, 1950) Hij visualiseert deze met behulp van een scatterplot. De onderzoekseenheden worden gevormd door negen kiesdistricten, die elk getypeerd worden met het percentage zwarten en het percentage analfabeten aldaar woonachtig. De Pearson-correlatiecoëfficiënt tussen beide groepskenmerken blijkt zeer hoog: 0,97 (geval I). Vervangt men de kiesdistricten door de 48 verschillende staten, dan wordt de ecologische relatie 0,77. Beide ecologische correlaties suggereren ruwweg: in gebieden met veel zwarten is veel analfabetisme. Vervolgens berekent Robinson een individuele correlatie, een *'correlation in which the statistical object or thing described is indivisible'* op basis van ruim negentig miljoen onderzoekseenheden, weergegeven in een in een 2×2 kruistabel. De individuele *fourfold-point* correlatie-coëfficiënt is weliswaar eveneens positief, maar bedraagt slechts 0,2 (geval II). Alsof dit contrast tussen I en II nog niet genoeg is, komt Robinson met een tweede voorbeeld, waarbij ook het teken en daarmee dus de richting van de correlatie veranderen. Dit voorbeeld betreft het verband tussen geboorteland (*native born* versus *foreign born*) en analfabetisme. De ecologische relaties op basis van districten (-0,62) en staten (-0,53) zijn negatief (I), de individuele correlatie blijkt positief: 0,12 (II). Daarmee sluit zijn voorbeeld nog beter aan bij de vele klassieke paradoxen en anomalieën (confounding, 'spurious correlation', Simpson-paradox) die traditiegetrouw in inleidende cursussen statistiek worden besproken.

Ogenscheinlijk is Robinsons doelstelling bescheiden. Hij beoogt niet een nieuwe analysetechniek of eigen paradigma te ontwikkelen, maar geeft in eerste instantie slechts een caveat, een profylactische waarschuwing. Naar eigen zeggen verschaft hij *'a definite answer as to whether ecological correlations can validly be used as sub-*

stitutes for individual correlations. They cannot.' De auteur stelt dat *'the purpose of this paper will have been accomplished, however, if it prevents the future computation of meaningless correlations and stimulates the study of similar problems with the use of meaningful correlations between the properties of individuals.'* Hoe dan ook, vandaag de dag wordt deze problematiek dikwijls getypeerd als de *ecologische drogreden*, complementair aan de *atomistische drogreden*, waarbij juist op basis van individuele data conclusies worden getrokken over (eigenschappen van) groepen.

Robinson besluit de kwestie te verhelderen door recht-toe rechtaan de 'anatomie van de ecologische correlatie' te analyseren, waarbij hij een minimum aan wiskundige voorkennis vooronderstelt. Zo laat hij onder meer zien dat de verbindende schakel tussen analyse I en II wordt gevormd door de *within-areas individual correlations*, oftewel de individuele *fourfold-point* correlaties tussen etniciteit en analfabetisme, berekend voor elk van de negen districten (III). Analyses I en II hangen beide hiervan af, maar op een verschillende manier; voor I geldt dat deze wordt bepaald *'only upon their marginal totals'*, II daarentegen *'upon the internal frequencies'*. Uiteraard bepalen in een 2×2 kruistabel de rij- en kolomtotalen (de 'afzonderlijke' verdelingen) niet de inhoud van de vier cellen (de 'gemeenschappelijke' verdeling). Andere aspecten, waaronder de wegingen van de data die Robinson hanteert en de formele wiskundige relaties tussen I, II en III blijven hier buiten beschouwing. Zie hiervoor (Robinson, 1950) en bij voorbeeld (Kreft, 1987) en (Te Grotenhuis, 2011). Hier beperken we ons tot een tweetal opmerkelijke gezichtspunten van Robinson, die van cruciaal belang zijn, zowel binnen de context van de sociologie, als in het licht van de ontwikkelingen in de statistiek.

De erfenis van Durkheim

Het eerste opmerkelijke gezichtspunt van Robinson wordt duidelijk wanneer hij onomwonden stelt dat gebruik van ecologische correlaties niet geschiedt omdat de auteurs geïnteresseerd zouden zijn in (correlaties tussen de) eigenschappen van gebieden, groepen et cetera. Bij alle ecologische studies is *'the obvious purpose'* iets te leren over het gedrag van individuen. *'Ecological correlations are used simply because correlations between the properties of individuals are not available. In each instance however the substitution is made tacitly rather than explicitly'*. Cu-

rius is dat de uitspraak wordt gedaan in een tijdschrift voor sociologie, omdat de wisselwerking tussen individu en groep juist daar op bijzondere, eigen manier gestalte krijgt. Centraal staan hier zowel de vraag naar de ontologische status van groepen (instituties, collectieve representaties) als gepostuleerde 'echte' bouwstenen van de samenleving, als de vraag naar de rol van de statistiek bij dit alles. Een en ander werd reeds actueel sedert de opkomst van de sociologie in de 19e eeuw toen pioniers als Karl Marx, Adolphe Quetelet, Auguste Comte en later vooral Max Weber (1864-1920) en Emile Durkheim (1858-1917) trachtten het vak een eigen wetenschappelijke status te geven door het af te splitsen van de wijsbegeerte en de daarmee verbonden speculatieve traditie. Daarmee werd de jonge discipline direct in een krachtenveld van enkele klassieke positiebepalingen of tegenstellingen getrokken: empirisme versus rationalisme, *Erklären* versus *Verstehen* (Dilthey), het nomothetische versus het idiografische (Windelband) en vooral een deterministisch versus een probabilistisch wereldbeeld.

Tegen deze achtergrond moesten object en methode worden bepaald. Zo lag het in de rede aansluiting te zoeken bij het dominante empirisme, dat onder meer culmineerde in Comte's positivism. Dat empirisme vroeg om een sterke empirische basis, hetgeen in 'de met cijfers bedekte 19e eeuw' (Stamhuis, 1992), waarin een *'avalanche of numbers'* (Hacking, 1989) waarneembaar was, niet problematisch leek. Bovendien was er naast deze -door de net opgerichte statistische bureaus in Europa vergaarde- data nog een tweede bron van gegevens, de etnografische gegevens van uitheemse culturen, verzameld in overzeese gebieden. Volgens (Oberschall, 1989) vormen beide databronnen de *'empirical roots of social theory'*. Hoe dan ook, om de wetenschappelijke status te verhogen lag het volgens velen meer voor de hand aansluiting te zoeken bij de op *Erklären* gerichte nomothetische natuurwetenschappen, dan eigen spelregels te formuleren of zich te richten op *Verstehen* en een idiografische benadering (waarvoor de etnografische data natuurlijk meer geschikt leken). En als men dan voor de natuurwetenschappelijke basis koos (Comte, Quetelet), moest dan een klassieke deterministische benadering (die in de 19e eeuw op zijn retour was) het uitgangspunt zijn of moest juist gebruik worden gemaakt van de opkomende 'nieuwe' statistiek in een eeuw waarin variatie en verandering in geschiedenis en natuur als inherent werden beschouwd en een probabilistisch wereldbeeld zich begon af te tekenen?

Eén probleem bij dit alles was dat het empirisme naast een empirische basis ook een sterk anti-metafysische, anti-speculatieve strekking vertoonde. Wetenschappers mochten niet teruggrijpen op abstracties waarbij onwaarneembare entiteiten werden gepostuleerd. Gedenkwaardig was het verzet van fysici als Mach en Ostwald tegen een realistische interpretatie van de zeer succesvolle atoomtheorie. Pas na het werk van Einstein en vooral Perrins studie *Les Atomes* uit 1913 kanteelde de discussie (Starmans, 2013). Dat alles pleit op het eerste gezicht natuurlijk ook in de sociologie niet voor het postuleren van instituties en collectieve representaties als *causaal werkzame grootheden* of *krachten*, die de samenleving - al dan niet beschouwd als mechanisme of organisme- opbouwen. Ook leek het eenvoudiger de nieuwe statistiek toe de passen op individuen en hun eigenschappen, dan op voornoemde abstracte entiteiten. Mede door dit alles worstelden velen met de statistiek en de toepassing ervan in de 19e eeuwse sociologie was allerm minst een klare zege-tocht. Veeleer vertoonde deze toepassing kenmerken van de Processie van Echternach, die sedert de Middeleeuwen jaarlijks in de gelijknamige Luxemburgse stad wordt gehouden en waarbij door de gelovigen altemeerend drie stappen vooruit en twee stappen achteruit worden gezet. Vele tijdgenoten aarzelden, anderen bleken spijtoptanten en keerden op hun schreden terug van het statistische avontuur (Oberschall, 1989; Porter, 1986).

De centrale figuur in dit alles rond de eeuwwisseling was Emile Durkheim, die trachtte de verschillende stromingen met elkaar te verzoenen. Van een reductionisme van instituties naar individuen was bij hem geen sprake. Voorop stonden 'sociale feiten', die niet van individuele preferenties, overtuigingen of intenties afhangen, maar alleen van andere sociale feiten en feiten die gezamenlijk de structuren in de wereld constitueren. Beroemd werden zijn studies over religie als bindend element in moderne samenlevingen. In het klassieke *Le Suicide* (Durkheim, 1897) gebruikte Durkheim vele statistische gegevens, maar gaandeweg evolueerde van hij van aanhanger van statistiek naar scepticus en liet hij etnografische data prevaleren boven de Westerse numerieke data. Steeds meer richtte hij zich op het functionalisme dat in de biologie opgang maakte en paste hij evolutionaire principes toe op groepen en instituties. Anderen, zoals Max Weber, kozen wel de individuele benadering die mede zou leiden tot het methodologische individualisme in de sociale wetenschappen. Een en ander moet in dit korte essay

verder buiten beschouwing blijven, maar de problematiek van de status van theoretische concepten binnen een *niet-reductionistische* sociologie, de micro-macroloof en de vraag hoe statistiek op dit geaggregeerde of abstracte niveau kon worden toegepast blijft actueel. De ecologische correlatie is daarvan hoe dan ook een manifestatie en kan dus geenszins als overbodig of gratuit worden aangemerkt.

Fishers handboek

Een ander opmerkelijk gezichtspunt van Robinson sluit hierbij aan en is om twee redenen historisch-filosofisch relevant. Allereerst constateert hij dat *'ecological correlations are used in an impressive number of quantitative sociological studies, some of which by now have attained the status of classics'*. Hij beseft dat zijn conclusie *'...has serious consequences, and that its effect appears wholly negative because it throws serious doubt upon the validity of a number of important studies made in recent years'*. Vervolgens noemt Robinson een groot aantal artikelen met naam en toenaam. De daarin gehanteerde technieken betreffen onder meer factoranalyse, canonische correlatie-analyse en andere multivariate technieken waaronder de door Sewall Wright ontwikkelde pad-analyse die later tot structurele vergelijkingsmodellen zou uitgroeien. In al deze technieken wordt getracht op basis van gemeten variabelen op individueel niveau, latente variabelen te creëren die corresponderen met theoretische termen, gepostuleerde entiteiten, et cetera. De vraag of deze ook een causale interpretatie toelaten is tot op heden een kernvraag binnen zowel methodologie als epistemologie. (Denis, 2006; Subramanian, 2007) Het teruggrijpen op en omarmen van deze methoden door sociologen kan uiteraard worden beschouwd als pogingen de (multivariate) statistiek te combineren met Durkheimiaanse inzichten.

Maar er is nog een ander, wezenlijk verschillend aspect. Het artikel van Robinson verschijnt precies 25 jaar na de poging van Ronald Fisher (1890-1962) met zijn *Statistical methods for research workers* (1925) statistische leken een handboek te verschaffen dat beoogde een state-of-the-art overzicht te geven op het gebied van *inferentiële statistiek*; onproblematische, 'veilige' technieken en inzichten, die de kloof tussen steekproef en populatie kunnen overbruggen. In weerwil van talloze herdrukken,

was dit streven deels illusoir. Dat bleek onder meer uit het feit dat onmiddellijk na het verschijnen van het boek de oude vete tussen Fisher enerzijds en Jerzy Neyman (1894-1981) en de familie Pearson anderzijds weer opblaide (Pearson, 1926). Veel technieken werden weliswaar grif toegepast in biologie en psychologie, maar waren nog allerminst uitgekristalliseerd. Gert Gigerenzer laat in zijn essay 'The inference experts' (Gigerenzer, 1989) zien hoe zelfs over de meest elementaire begrippen geen overeenstemming bestond: p-waarden, parameterschatting, betrouwbaarheidsintervallen, hypothese toetsing, et cetera. Hybride methoden en onversneden eclecticisme vierden hoogtij. Waar op methodologisch gebied convergentie, standaardisering en consensus soms mogelijk bleken (Blalock, A.D. de Groot) bleek op statistisch gebied anno 1950 zo'n handboek – waaraan nu juist binnen de kwantitatieve sociologie behoefte bestond – een brug te ver. De vraag of dit ook heden ten dage de status quo is, blijft hier buiten beschouwing (Starmans, 2018).

Epiloog

Tot slot enige kanttekeningen.

Al is de relevantie van Robinsons paper voor multilevel analyse evident, het is de vraag of hij als pionier en wegbereider ervan moet worden beschouwd. Hij kon latere ontwikkelingen op het gebied vanzelfsprekend niet voorzien en het was evenmin zijn oogmerk de weg te plaveien voor een nieuwe methode. De studie van contingentietabellen, categoriale data-analyse en multilevel methoden zouden pas in de decennia daarna een hoge vlucht nemen. Omdat daarmee ook complexere hiërarchische structuren mogelijk worden, is de traditionele *interpretatie* van hiërarchische data als 'indivisible' individu versus groep, die wij hanteerden om aan te sluiten bij Robinsons definitie, uiteraard rigide. Het onderscheid wordt relatief als het individu zelf 'groep' wordt: herhaalde metingen genest binnen een proefpersoon, studies genest in een meta-analyse, geïnterviewden genest in een interviewer, et cetera. Dat geldt a fortiori in ketens van inbeddingen: genen, cellen, organen, individuen, families, wijken, steden, et cetera, waarin het niveau van beschouwing al helemaal geen a priori gegeven kan zijn. Hoe dan ook, in het licht van de beschikbare technieken en de status quo binnen de kwantitatieve sociologie rond

1950 was Robinsons caveat zeker terecht.

Het artikel kan dan ook zonder overdrijving een klassieker worden genoemd, temeer omdat het vrijwel onmiddellijk leidde tot reacties van statistici om het probleem te repareren, (Duncan, 1953; Goodman, 1853, 1959) en Robinsons werk een vernietigende kritiek gaf, die in de sociologie nog steeds nagalmt. In 2011 wijdde de *Journal of Epidemiology* een themanummer aan Robinsons paper en gaf het opnieuw uit. De socioloog Te Grotenhuis ontdekte verschillende nog niet eerder opgemerkte (kleine) fouten en onnauwkeurigheden in het artikel, voerde een replicatiestudie en een multilevel analyse uit op de oorspronkelijke data (Te Grotenhuis, 2011). Deze fouten van Robinson en zijn wellicht enigszins geringe empathisch vermogen betreffende de oogmerken en noden van de sociologie doen aan dit alles niets af.

Ecologische analyses vonden reed plaats in vroege 17e-eeuwse statistische studies van John Graunt en William Petty en blijven omwille van ethiek en privacy dikwijls ook onmisbaar in het tijdperk van data science. In weerwil van de volop beschikbare geavanceerde multilevel software maakt de hier beknopt geschetste problematiek duidelijk dat Fishers goedwillende 'research workers', die zich een weg moeten banen in de jungle van vele data-analytische technieken en statistische methoden, nog steeds voor een ongewis en hachelijk avontuur kunnen komen te staan. Zo beschouwd is de expliciete verwijzing naar de populaire *reality soap*, waar goed bedoelende avonturiers in onherbergzame gebieden trachten te overleven ten aanschouwen van een miljoenen publiek, wellicht niet geheel misplaatst, zeker tegen de achtergrond van de eerder gesignaleerde hyperbolen in tijden van data science en big data.

LITERATUUR

- Denis, D.J., & Legerski, J. (2006). Causal Modelling and the origins of Path Analysis. *Theory and Science*, 7(1).
- Duncan, O.D., & Davis, B. (1953). An alternative to ecological correlation. *American Sociological Review*, 18, 665–66.
- Durkheim, E. (1897). *Le suicide*. Paris: F. Alcan; English translation by J A Spalding, 1951. Toronto: Free Press, Collier-MacMillan, Toronto.
- Gigerenzer, G. (1989). *The Empire of Chance; how probability changed science and everyday life*. Cambridge: Cambridge University Press.
- Goodman, L. (1953). Ecological regression and the behavior of individuals. *American Sociological Review*, 18, 663–64.
- Goodman, L. (1959). Some alternatives to ecological correla-

tion. *American Journal of Sociology*, 64, 610–625.

- Grotenhuis, M. Te, Eisinga, R., & Subramanian, S.V. (2011). Robinson's Ecological Correlations and the Behavior of Individuals: methodological corrections. *The International Journal of Epidemiology*, 40(4).
- Hacking, I. (1989). *The Taming of Chance*. Oxford: Oxford Publishers.
- Oberschall, A. (1989). The two empirical roots of Social Theory and the Probabilistic Revolution. In: Lorenz Kruger (ed.) *The Scientific Revolution, Part II*. Cambridge: MIT Press, Cambridge.
- Selvin, H.C., (1958). Durkheim's suicide and problems of empirical research. *American Journal of Sociology*, 63, 607–619.
- King, G. (1997). *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton: Princeton University Press.
- Kreft, I., & De Leeuw, E.D. (1987). The see-saw-effect: a multi-level problem? A reanalysis of some findings of Hox and De Leeuw. *Quality and Quantity*, 22, 127–137.
- Pearson, K. (1892) *The Grammar of Science*. London.
- Pearson, E., (1926). Review (1926) of Statistical Methods for Research Workers (R. A. Fisher). *Science Progress*, 20, 733–734.
- Porter, T.M. (1986). *The Rise of Statistical Thinking, 1829–1900*. Princeton, New Jersey: Princeton University Press.
- Robinson, W.S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15, 351–357.
- Stamhuis, J.H., & Knecht-van Eekelen, A. de (1992). *De met cijfers bedekte negentiende eeuw: Toepassing van statistiek en waarschijnlijkheidsrekening in Nederland en Vlaanderen 1840–1920*. Rotterdam: Erasmus Publishing.
- Starmans, R.J.C.M. (2013). Perrin and the Partisans of the Void (in Dutch). *Nederlands Tijdschrift voor Natuurkunde*, 79(12).
- Starmans, R.J.C.M. (2018). The Predicament of Truth: on Statistics, Causality, Physics and the Philosophy of Science. In: Mark J. Van der Laan & Sherri Rose (Eds.), *Targeted Learning in Data Science: Causal Inference for Complex Longitudinal Studies*. Springer Series in Statistics. New York: Springer.
- Subramanian, S.V. (2004) The relevance of multilevel statistical methods for identifying causal neighborhood effects. *Social Science & Medicine*, 58, 1961–1967.
- Subramanian, S.V., Glymour, M.M., & Kawachi, I. (2007) Identifying causal ecologic effects on health: a methodologic assessment. In: Galea S, editor. *Macrosocial Determinants of Population Health*. New York: Springer Media. pp. 301–331.

RICHARD STARMANS is verbonden aan de Faculteit Bètawetenschappen (Department of Information and Computing Sciences) van de Universiteit Utrecht. Hij doet onderzoek op het snijvlak van filosofie, statistiek en informatica. E-mail: starmans@cs.uu.nl



Een trein in Madurodam, een gemoderniseerde ICM4, 2014. Foto: Roel Hemkes (CC BY 2.0)

IS GELIJKHEID ALTIJD GEWENST? De afweging tussen gelijkheid en kwaliteit in personeelsroostering

THOMAS BREUGEM

De toewijzing van werk aan personeel, beter bekend als personeelsplanning, is een belangrijk onderdeel van de planning voor een openbaarvervoerbedrijf als de Nederlandse Spoorwegen (NS).

In de praktijk wordt het personeelsplanningprobleem in twee fases opgelost. In de eerste fase worden de diensten samengesteld om de geplande ritten (bijvoorbeeld een rit van Rotterdam naar Den Haag) uit te voeren. De diensten specificeren een werkdag voor een personeelslid en moeten voldoen aan allerlei regels (zo geldt er een maximale dienstlengte en moet de dienst een pauze bevatten). In de tweede fase worden de roosters gemaakt. Hierbij worden de gemaakte dien-

sten toegewezen aan het personeel. Ook hier dienen verschillende regels in acht te worden genomen. Zo geldt er, bijvoorbeeld, een maximale werklust per week en moet er voldoende rusttijd tussen opeenvolgende diensten worden gepland. Tevens dient het leuke (en niet leuke) werk eerlijk verdeeld te worden over de verschillende personeelsleden.

Personeelsplanning is een zeer complex probleem, voornamelijk vanwege de enorme omvang en de twee tegenstrijdige planningsdoelen: het minimaliseren van de operationele kosten aan de ene kant en het maximaliseren van de kwaliteit van het werk aan de andere kant. Planningsgerelateerde kwaliteitsaspecten die in het recente

verleden naar voren zijn gekomen, zijn, bijvoorbeeld, te weinig variatie in het werk (met als gevolg de inmiddels bekende uitdrukking 'rondje om de kerk'), of te weinig conducteurs op treinen van bepaalde materieeltypes. Onvrede onder het personeel leidde tot de ontwikkeling van de 'Lusten-en-Lasten-Delen' regels voor de diensten, een nieuwe lijst van planningsregels met als doel de verdeling van het werk te verbeteren (zie Abbink et al., 2005). De 'Lusten-en-Lasten-Delen' regels worden ook, al dan niet altijd strikt, toegepast bij het roosteren.

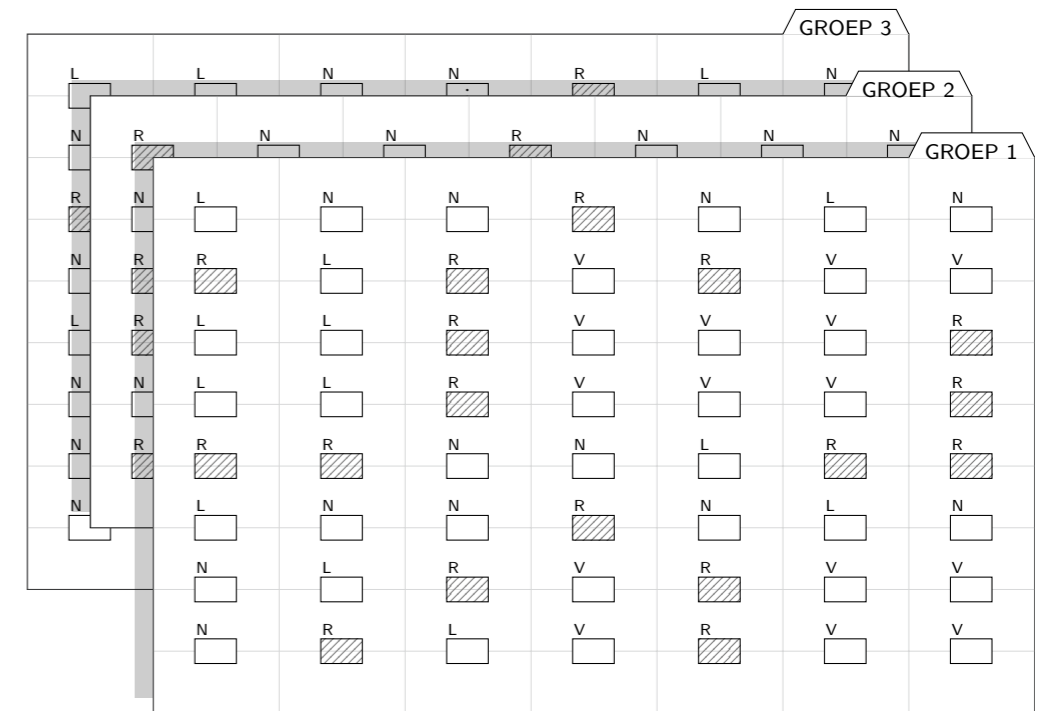
In het huidige roosterproces wordt allereerst gezorgd dat het werk gelijk verdeeld is over de personeelsleden, daarna wordt gekeken naar de kwaliteit van de gemaakte roosters. We laten zien dat een dergelijke sequentiële aanpak, waarbij de focus ligt op gelijkheid, over het algemeen suboptimaal is en zelfs tot een 'ongelijke' oplossing kan leiden.

In ons onderzoek ontwikkelen we een geïntegreerde aanpak voor het roosterproces, waarbij de gelijkheid en de kwaliteit van de roosters gelijktijdig geoptimaliseerd worden. In deze aanpak wordt gebruik gemaakt van 'basisroosters' om een bepaalde minimum kwaliteit te garanderen.

Personeelsroostering

Nadat de diensten zijn gemaakt, worden deze toegewezen aan het personeel in de vorm van roosters. In dit onderzoek kijken we naar cyclische roosters, welke vaak worden gebruikt binnen openbaarvervoerbedrijven. De personeelsleden worden opgedeeld in verschillende *roostergroepen*, gebaseerd op bepaalde voorkeuren en kenmerken (zoals bijvoorbeeld leeftijd), en elke groep opereert vervolgens in een cyclisch rooster. In een cyclisch rooster 'rouleren' alle personeelsleden, wat ervoor zorgt dat, op de lange termijn, elk personeelslid in de groep exact hetzelfde werk doet. Een voorbeeld van een cyclisch roosteringsprobleem is gegeven in figuur 1, waarbij het basisrooster te zien is voor drie roostergroepen. De cyclische rooster impliciteert dat de eerste werknemer van de eerste groep (vooraan in figuur 1) start met het werk in de eerste week, vervolgens het werk in de tweede week uitvoert en na acht weken weer opnieuw begint met de eerste week. Op eenzelfde manier begint de tweede werknemer in de tweede week en schuift daarna telkens een week door.

In tegenstelling tot veel operationele planningspro-



Figuur 1. Voorbeeld van een roosteringsprobleem. De diensten dienen toegewezen te worden aan drie roostergroepen. Voor elke groep is het basisrooster (het type werk per dag) gespecificeerd: een vroege (V), late (L), of nacht (N) dienst; of een rustdag (R). Elke kolom van het rooster is gerelateerd aan een dag in de week (bijvoorbeeld maandag) en elke rij in het rooster specificeert een werkweek voor een van de personeelsleden

blemen, is het hoofddoel bij het maken van roosters niet het minimaliseren van kosten. In plaats daarvan draait het om het optimaliseren van *gelijkheid* en *kwaliteit*. De gelijkheid van de roosters relateert aan de manier waarop het werk verdeeld is over de verschillende personeelsleden. Alhoewel *binnen* elke groep de personeelsleden hetzelfde werk doen, is dit niet noodzakelijk het geval *tussen* de verschillende groepen. De gelijkheidsmaatstaf van de roosters wordt daarom gebaseerd op de balans tussen de verschillende groepen met als doel dat verschillende aspecten (zoals de gemiddelde dienstlengte, of het percentage werk op dubbeldekkertreinen binnen de diensten) voor elke groep (zo goed als) hetzelfde zijn. De kwaliteit van het rooster is, daarentegen, gebaseerd op een individuele basis. Dit kan bijvoorbeeld gebaseerd zijn op de totale werklust of de tijd tussen opeenvolgende diensten. Zo wordt een rusttijd van minder dan 16 uur tussen twee diensten als onwenselijk gezien.

De gelijkheid en de kwaliteit van de roosters zijn twee verschillende concepten. Om gelijkheid te meten wordt enkel gekeken naar de toewijzing van de diensten aan de groepen, zonder daarbij in acht te nemen hoe het resulterende rooster eruit ziet. Om kwaliteit te meten wordt, aan de andere kant, enkel gekeken naar de individuele roosters, zonder daarbij het grotere geheel te beschouwen. Het is daarom goed mogelijk dat roosters gelijk, doch van lage kwaliteit zijn, of vice versa. Dit betekent dat roosters een afweging is tussen gelijkheid en kwaliteit. Aan de ene kant dient de verdeling van werk zo gelijk mogelijk te zijn, maar aan de andere kant moet ook de kwaliteit van de resulterende roosters zo hoog mogelijk zijn.

Oplosmethode

We formuleren het roosteringsprobleem als een optimalisatieprobleem met twee doelfuncties: gelijkheid en kwaliteit. Het probleem is geformuleerd als een toewijzingsprobleem met extra restricties. Gelijkheid wordt gemodelleerd aan de hand van een aantal aspecten, zoals bijvoorbeeld de gemiddelde dienstlengte per groep. Aan elk aspect wordt een waarde toegekend, en het doel is de gemiddelde waarde per groep (ongeveer) gelijk te hebben voor elk aspect. Als dit niet het geval is, en er dus een te groot verschil is tussen de groep met de hoogste waarde en de groep met de laagste waarde, leidt dit tot ongelijke roosters. Deze verschillen worden daarom geminimaliseerd als een gewogen som over de verschillende aspecten. Verder zijn er afgesproken onder- en bovengrenzen waaraan moet worden voldaan. Kwaliteit wordt gemodel-

leerd door middel van harde en zachte restricties. De harde restricties zorgen dat de gemaakte roosters voldoen aan de afgesproken regels en de zachte restricties beboeten onwenselijke eigenschappen van het rooster (zoals een rusttijd van minder dan 16 uur).

In het algemeen zal er geen oplossing bestaan die beide doelstellingsfuncties gelijktijdig optimaliseert. We beschouwen daarom een collectie Pareto-efficiënte oplossingen, waarbij een oplossing Pareto-efficiënt genoemd wordt als er geen oplossing bestaat die beter is in beide doelstellingsfuncties. Voor het roosteringsprobleem betekent dit dat er geen oplossing mag bestaan die zowel gelijkjer als van hogere kwaliteit is. Alle efficiënte oplossingen kunnen worden gevonden door verschillende niveaus van gelijkheid te vereisen en vervolgens de kwaliteit te maximaliseren.

Omdat restricties met betrekking tot de weken van een rooster veel voorkomen (denk bijvoorbeeld aan een maximale werklust over de week), ontwikkelen we een wiskundig model gebaseerd op weken. Dit betekent dat we direct meerdere diensten toewijzen aan een gehele week, in plaats van apart een dienst toe te wijzen aan elke dag in de desbetreffende week. Deze manier van modelleren leidt tot een sterke formulering, omdat veel regels impliciet kunnen worden gemodelleerd. Een nadeel van deze aanpak is echter dat het aantal mogelijke beslissingsvariabelen sterk toeneemt. Hierom wordt het model opgelost aan de hand van een Branch-and-Price-aanpak, waarin mogelijk gunstige beslissingsvariabelen 'on the fly' worden gegenereerd. Voor de details van deze methode refereren we naar Breugem et al., 2017.

Van theorie naar praktijk: Roosteren bij NS

Om de voordelen van de integrale aanpak te bepalen, is het model toegepast op instanties gebaseerd op data van NS. Dit leidde tot de volgende belangrijke inzichten.

Ten eerste dienen planners er op te letten de gelijkheid niet te 'over-optimaliseren'. In onze experimenten kwam naar voren dat een sterke focus op gelijkheid leidt tot een lage kwaliteit van de roosters. In een aantal gevallen was een verwaarloosbare toename van de gelijkheid al voldoende om de kwaliteit fors te laten afnemen (zie figuur 2). Dit impliceert dat een sequentiële aanpak, een aanpak waarin gelijkheid de hoogste prioriteit krijgt, enkel kan worden geadviseerd als gelijkheid bijzonder hoog gewaardeerd wordt.

Ten tweede is het twijfelachtig of gelijkheid, zoals dit nu is gedefinieerd, wenselijk is vanuit het oogpunt van

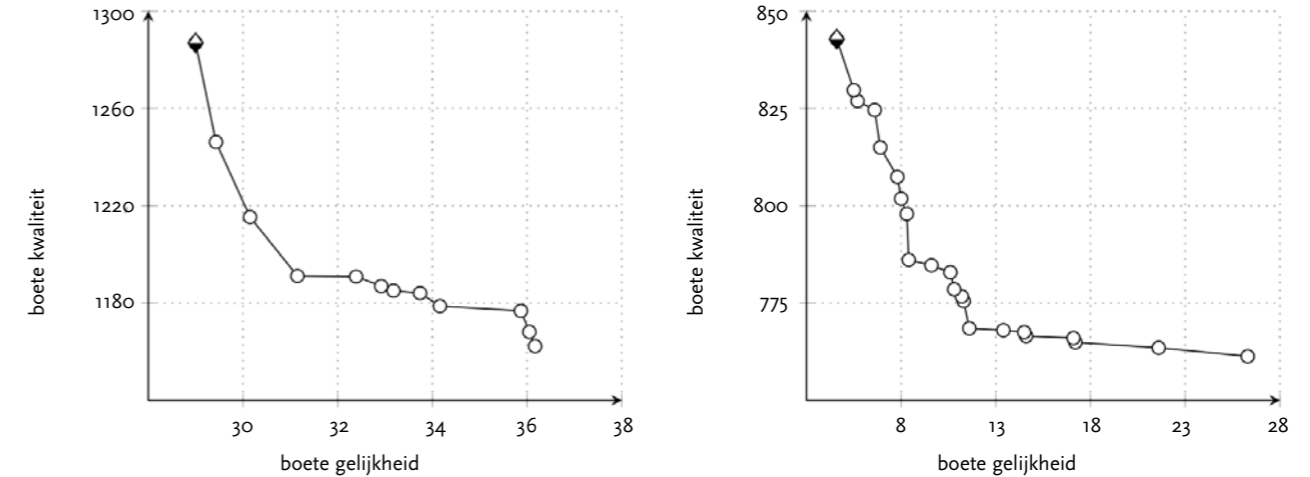


Fig 2. Curve met efficiënte oplossingen voor twee instanties. De kwaliteit en gelijkheid worden gemaximaliseerd door de boete te minimaliseren. De boete voor kwaliteit wordt bepaald aan de hand van de zachte restricties, de boete voor gelijkheid is de waarde van de gewogen som van de spreidingen. Het benadrukte punt representeert de (best mogelijke) oplossing gevonden met de sequentiële aanpak voor de betreffende instantie

de werknemer. Onze analyse laat zien dat de lagere kwaliteit van de roosters, die ontstaat door de hogere mate van gelijkheid, scheef verdeeld is over de verschillende roostergroepen. Dit betekent dat, wanneer de gelijkheid toeneemt, de kwaliteit van het rooster voor sommige groepen snel achteruit kan gaan, terwijl andere groepen er soms zelfs op vooruit kunnen gaan. De afname in kwaliteit was het grootst voor groepen die opereren in een onregelmatig rooster (een rooster met veel verschillende type diensten). Juist voor deze groepen is dit onwenselijk, omdat deze personeelsleden al een rooster van mindere kwaliteit 'accepteren' door in een onregelmatig rooster te werken.

De experimenten tonen aan dat de huidige aanpak niet adequaat omgaat met de complexiteit van het roosteringsprobleem. De expliciete afweging tussen gelijkheid en kwaliteit dient te worden meegenomen om een goed beeld te krijgen van het probleem. Tevens mist de huidige meetwijze van gelijkheid een aantal belangrijke aspecten. Zo dient, bijvoorbeeld, de verdeling van de kwaliteit meegenomen te worden wanneer de gelijkheid van de roosters wordt bepaald.

Conclusie

Ons onderzoek laat het belang zien van een geïntegreerde optimalisatie van gelijkheid en kwaliteit voor perso-

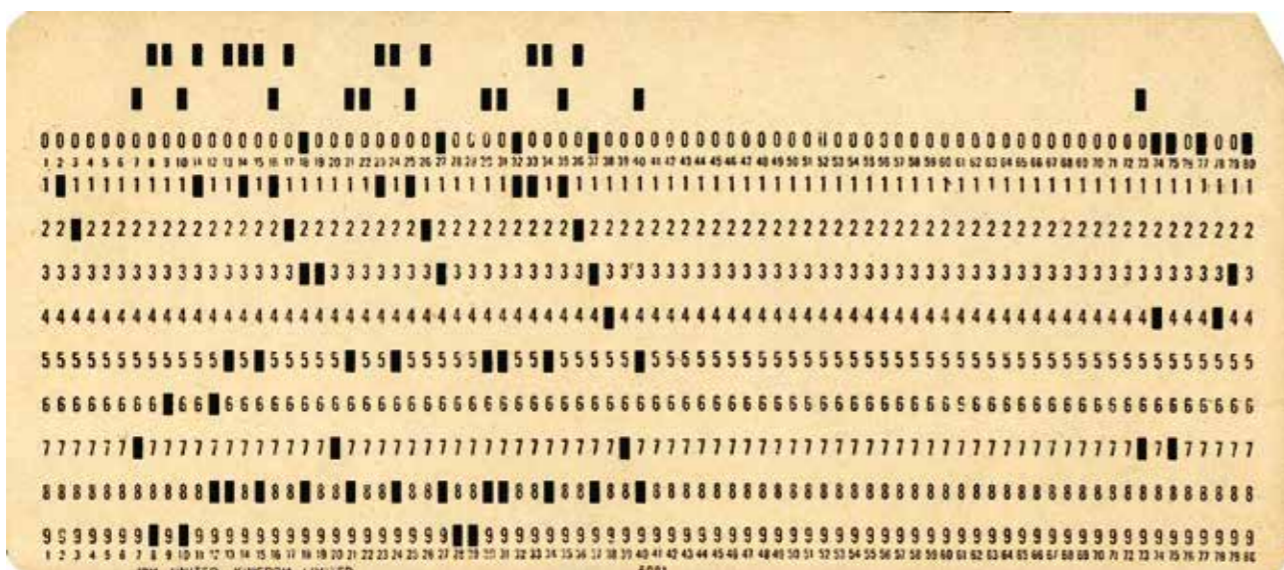
neelsroosteren. Het expliciet afwegen van deze twee concepten geeft een beter beeld van het roosteringsprobleem als geheel, wat ervoor zorgt dat de planner een betere beslissing kan nemen. Door verschillende aspecten gelijktijdig te optimaliseren vinden we een collectie van oplossingen, elk optimaal voor bepaalde voorkeuren, in plaats van één enkele oplossing. Hierdoor kan de afweging tussen de verschillende planningsdoelen expliciet worden geanalyseerd, wat ervoor zorgt dat alle partijen uiteindelijk beter af zullen zijn.

LITERATUUR

- Abbink, E., Fischetti, M., Kroon, L., Timmer, G., & Vromans, M. (2005). Reinventing crew scheduling at Netherlands Railways. *Interfaces*, 35(5), 393-401.
- Breugem, T., Dollevoet, T., & Huisman, D. (2017). Is equality always desirable? Analyzing the trade-off between fairness and attractiveness in crew rostering. *Econometric Institute Research Papers*, (EI2017-30).

Dit onderzoek heeft de eerste prijs gewonnen bij de 2017 INFORMS RAS student paper award contest on analytics and fact-based decision making in railway applications.

THOMAS BREUGEM is promovendus aan de Erasmus Universiteit Rotterdam. Zijn onderzoek richt zich op personeelsplanning binnen NS, met als hoofddoel het integreren van verschillende planningsfasen.
Email: breugem@ese.eur.nl



KISS

Keep It Simple, Statistician!

Deze uitdrukking komt u bekend voor, maar u twijfelt toch enigszins? Dat is goed mogelijk, oorspronkelijk is het namelijk Stupid! Maar ik ben terughoudend in het negatief kwalificeren van anderen, vandaar de Statistician. Waarom doe ik deze oproep? Omdat ik denk dat het hard nodig is.

De lezing door Sander van der Linden op onze Annual Meeting van 28 maart dit jaar benadrukte dat statistici meer aandacht moeten besteden aan de presentatie van hun resultaten'. Bovenal dienen ze zich te realiseren dat het 'grote publiek' in een klein verschil in overeenstemming tussen deskundigen al snel aanleiding ziet alle deskundigen te wantrouwen. Sander noemde als voorbeeld dat meer dan 95% van de deskundigen het er over eens is dat er Global Warming optreedt, maar dat de nog geen 5% twijfelaars/ontkenners genoeg zijn om in een aanzienlijk deel van de samenleving ontkenning teweeg te brengen.

Zelf denk ik dat er een dieperliggende oorzaak is: mensen zijn van jongs af aan bang voor cijfers en zetten dat om in vijandigheid. Al in het basisonderwijs lopen kinderen die goed kunnen rekenen heel snel het risico het

'pisaaltje' van de klas te worden. Goed kunnen voetballen is een veel belangrijker eigenschap. En die angst voor cijfers zet zich versterkt voort in de rest van het onderwijs met een fundamentele achterdocht jegens getallen als blijvend gevolg.

Ook kwam in Sanders lezing naar voren dat het verstandig is uitkomsten van een onderzoek op een niet te complexe manier te presenteren. Inderdaad, Keep It Simple, een oproep waar ik volledig achter sta.

Begin jaren '70 werkte ik mee aan een onderzoek naar het beeld dat bestaat over het gedrag van docenten. Zowel hun eigen idee daarover als dat van enkele honderden leerlingen per docent werd gemeten met een vragenlijst met zo'n tachtig items op een 7-punts schaal. De term was toen nog niet bekend, maar achteraf gezien waren dit Big Data, zeker gezien de toenmalige technische mogelijkheden. We maakten gebruik van optisch leesbare formulieren die op het kantoor van een grote zuivelfabriek werden ingelezen en in ponskaarten omgezet. Dat had ik voor een zacht prijsje kunnen regelen, die apparatuur werd verder toch alleen maar aan het eind van de dag een uurtje gebruikt om de bestellingen van afnemers te

verwerken. Na inlezen en controleren van de vele tientallen dozen ponskaarten werd onder andere factoranalyse ingezet om een onderliggende structuur te ontdekken. Er werd een aantal duidelijke componenten gevonden en de scores daarop zouden worden gebruikt om de docenten feedback te geven over hun gedrag. Het enige toen beschikbare programma voor factoranalyse berekende die scores met een gemiddelde van 50 en een standaarddeviatie 10. Dat was geen probleem voor het gebruik in het vervolg van het onderzoek, dat is dan ook gebeurd.

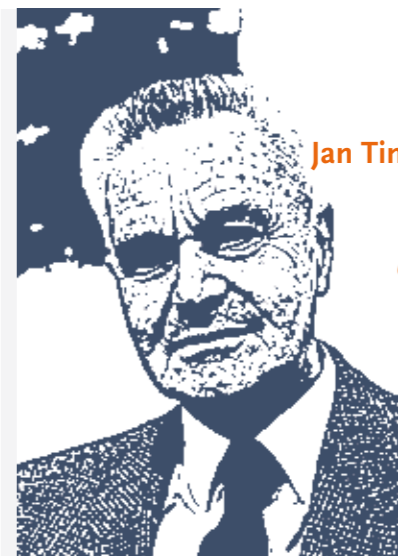
Maar een belangrijk aspect van het onderzoek was het terugkoppelen van deze scores naar de docenten om vervolgens na enige tijd te meten of er hierdoor verandering in hun gedrag was opgetreden. En het leek mij onbegonnen werk dit soort scores terug te koppelen, daar zou een heleboel uitleg bij nodig zijn. Gelukkig werd elk van de gevonden componenten gekenmerkt door een aantal items met hoge factorladingen. Ik heb toen voor iedere component het gemiddelde berekend van de scores op die items die minimaal een lading van 0,50 hadden en de andere items buiten beschouwing gelaten. Vervolgens heb ik de correlatie berekend van deze 'jan-boerenfluitjes' factorscores met de officieel berekende factorscores en die bleek ruim 0,90 te zijn. Met enige moeite heb ik de onderzoekers kunnen overtuigen dat het verstandiger zou zijn de terugkoppeling naar de docenten in de vorm van deze gemiddeldes te doen. De uitleg werd dan iets als "Wij hebben over alle docenten heen een dimensie in het gedrag gevonden die duidt op veel interactie met de klas (bijvoorbeeld) en die wordt gekenmerkt door de volgende items Uw gemiddelde score op deze items is...". Dat was begrijpelijke taal, de score werd namelijk gepresenteerd in precies dezelfde 7-punts schaal als de oorspronkelijke items en dat bleek achteraf zeer gewaardeerd.

This Statistician kept it simple.

Noot

- 1 Tijdens de Annual Meeting werd bekend dat een groep leden zich bezig gaat houden met de manier waarop statistisch onderzoek kan worden gepresenteerd. Dat belooft een heel interessante en nuttige nieuwe Sectie van de VVSOR te worden.

GERRIT STEMERDINK is eindredacteur van STA+OR.
E-mail: gjstemerding@hotmail.com



Jan Tinbergen Awards nu beschikbaar voor alle jonge statistici

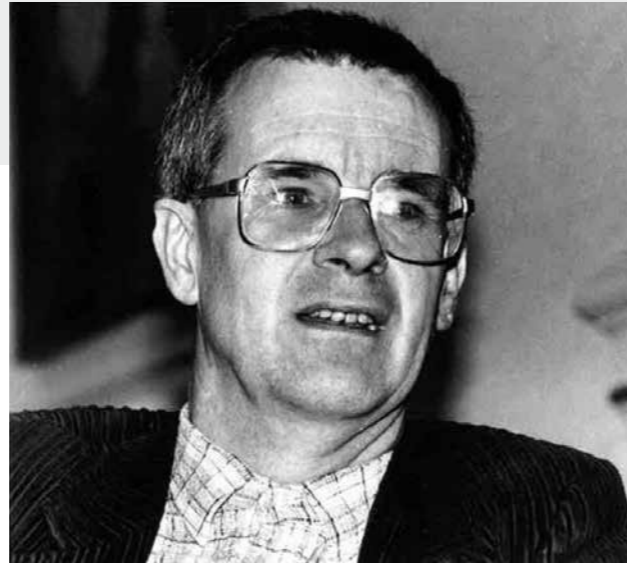
Het International Statistical Institute ISI reikt op haar tweejaarlijkse World Statistics Congresses al tientallen jaren de Jan Tinbergen Awards uit aan jonge statistici (niet ouder dan 30 jaar). Tot vorig jaar was de prijs bedoeld voor excellent werk van jonge statistici uit ontwikkelingslanden. Maar met ingang van de huidige competitie wordt de prijs uitgereikt aan werk dat gericht is op de verbetering van statistiek in ontwikkelingslanden. De eis dat de auteur uit zo'n land komt is daarmee vervallen. De nadruk ligt niet meer op de nationaliteit van de auteur maar op het onderwerp van het paper. Dat betekent dat deze prestigieuze prijs nu ook voor Nederlandse statistici beschikbaar komt.

De Awards zijn vernoemd naar de beroemde Nederlandse econometrist en winnaar van de Nobelprijs voor de Economie Jan Tinbergen (1903-1994). Ze worden mogelijk gemaakt door de Nederlandse Stichting Internationaal Statistisch Studiefonds. Het doel van deze stichting is het ondersteunen en bevorderen van statistiek in ontwikkelingslanden. De Awards worden daarom toegekend aan papers die deze doelstelling ondersteunen. In de woorden van de oproep tot deelname: 'Papers should address applied statistical problems of real relevance in building statistical capacity in countries with limited statistical infrastructure (developing countries).'

De Awards zullen worden uitgereikt tijdens het WSC 2019 in de Maleisische hoofdstad Kuala Lumpur. De prijs bestaat uit een bedrag van € 2.500 met daarnaast een retourticket economy class naar Kuala Lumpur, de inschrijfkosten voor het WSC en een bedrag voor hotel- en verblijfkosten. Meer bijzonderheden zijn binnenkort te vinden op de website van het ISI: isi-web.org.

Het ISI vraagt met nadruk ook vrouwelijke kandidaten zich aan te melden.

Ivo W. Molenaar (1935–2018)



ANNE BOOMSMA

Het overlijden van Ivo Molenaar op 26 februari 2018, voormalig voorzitter en erelid van de VVSOR, geeft aanleiding tot een korte terugblik op zijn persoon en zijn werk.* Drie aspecten staan daarin centraal: de baas, de meester en de onderzoeker, met vooraf enige aandacht voor zijn opleiding en eerste baan.

De beginjaren

Op 12 september 1935 werd Ivo in Den Haag geboren, jongste zoon in een welgesteld gezin met drie kinderen. Zijn vader was leraar Duits aan een hogereburgerschool. Na de middelbare schooltijd op het Gymnasium Hagnum (β en α) studeerde Ivo wiskunde aan de Universiteit van Amsterdam, waar hij onder meer les kreeg van D. van Dantzig en J. Hemelrijk. Hij was assistent bij de hoogleraren A. Heyting (meetkunde en grondslagen van de wiskunde) en T. Runnenburg (kansrekening). In 1962 haalde hij cum laude zijn doctoraal.

Met die opleiding vond hij gemakkelijk een baan als junior onderzoeker bij de statistiekafdeling van het Mathematisch Centrum (MC) in Amsterdam (het huidige Centrum voor Wiskunde & Informatica). Naast onderzoek was zijn belangrijkste taak: advies en consultatie voor Nederlandse onderzoekers en kennisverspreiding in de vorm van postacademische cursussen. Hij was medeauteur van twee deeltjes uit de bekende MC-serie *Leergang besliskunde*. Zijn werk op het MC vond hij plezierig en stimulerend.

Aan het MC voltooide Ivo in 1970 zijn dissertatie over benaderingen van discrete kansverdelingen (de Poisson,

de binomiale en de hypergeometrische) met Hemelrijk als promotor en Van Zwet als co-promotor. Na zijn promotie kreeg hij van G.P. Patil, expert op het gebied van discrete verdelingen, een aanbod om gedurende het academisch jaar 1970–1971 als ‘visiting assistant professor’ aan de Pennsylvania State University te komen werken. Met zijn vrouw Ans de Roos en drie dochters ging hij een jaar lang op avontuur.

In december 1970 kreeg hij een onverwacht telefoontje uit Groningen: of hij belangstelling had voor de nieuwe leerstoel Statistiek en Meettheorie voor de Sociale Wetenschappen. Dat had hij zeker. Op weg naar Groningen voor oriënterende gesprekken las hij snel het recent voltooide proefschrift van Rob Mokken, die hij goed kende uit zijn studietijd in Amsterdam. De benoeming tot hoogleraar aan de Rijksuniversiteit Groningen (RUG) zag Ivo als een ‘Herausforderung’. Zijn oratie was dan ook getiteld: *Dit is een uitdaging*. Minder bekend is de maatschappelijk getinte toevoeging: ‘Het is een uitdaging om te leven in 1972’ (Molenaar, 1972, p. 16).

De baas

Ivo heeft bijna dertig jaar leiding gegeven aan de kleine vakgroep Statistiek en Meettheorie binnen de faculteit Gedrags- en Maatschappijwetenschappen, zoals ze nu heet. Eerst vormde hij in zijn eentje de vakgroep, maar in de loop der jaren kwamen daar drie vaste medewerkers bij, waaronder één zeer duurzame. Hij heeft het belang van dat kleine maar open bastion stevast door dik en

dun verdedigd, altijd met verve, soms met een al even bijzondere boosheid, niet zelden uit teleurstelling ontkend door collega’s en bestuurders. Het benadrukken en veiligstellen van de onmisbare rol van het vak statistiek bij sociaal-wetenschappelijk onderzoek binnen de faculteit stond daarbij steeds voorop. Dat hij enkele jaren minder gemotiveerd was geraakt toen hij zich in dat streven (systematisch) bedreigd voelde en dat hij daar persoonlijk veel last van had gehad, verwoordde hij haarfijn in zijn afscheidscollege (Molenaar, 2000).

Het statistisch bolwerk aan de Oude Boteringestraat had ook een tijdlang (1985–1991) een dependance aan de Grote Markt. Daar hebben in alle vrijheid zes promovendi met succes aan hun dissertatie gewerkt. Met zoveel enthousiast en jeugdig volk in de vakgroep was het een hoogst plezierige en vruchtbare tijd. Dat kwam mede doordat Ivo niet op een autoritaire, dwingende manier leiding gaf. Van spanning, overbelasting of hiërarchische druk was naar eigen bevinden en observatie zelden sprake. Er was permanente aandacht vanuit het monumentale hoofdkwartier aan de Oude Boteringestraat en altijd directe bereidheid tot hulp en advies. Het resultaat: werkgeuk in kleine, open kring. Dat de vakgroep Statistiek en Meettheorie vlak na Ivo’s emeritaat werd opgeheven was navrant en voer voor psychologen.

De meester

Waar ik persoonlijk als student, direct al na zijn benoeming in Groningen, bijzonder veel waardering voor kreeg waren Ivo’s didactische kwaliteiten. Zijn colleges waren een verademing: Hogg en Craig (1970) voor een paar studenten in een zaaltje bij Psychologie. Dat was nog eens statistiek en de meester deed alles uit zijn hoofd! Ik heb in meerdere opzichten veel van hem geleerd, zeker ook op het gebied van statistisch advies en consultatie, een van de primaire taken van de vakgroep.

Als weinig anderen doorzag Ivo ingewikkelde problemen snel en kon hij de essentie ervan helder uitleggen – zijn taalgebruik was onberispelijk. Hij had bovendien een voorkeur voor eenvoud: een simpele aanpak van statistische problemen vereenvoudigde het statistisch ontwerp en de interpretatie van resultaten. Dat gold bijvoorbeeld ook voor modelselectie: ‘Een eenvoudig model is beter uit te leggen aan anderen, en een eenvoudig model presteert doorgaans beter bij replicatie’ (Molenaar, 2000, p. 8). Talloze studenten, onderzoekers en promovendi hebben van zijn meesterlijke kwaliteiten in colleges, begeleidingsgesprekken en consultaties kunnen profiteren.

Ivo’s invalshoek bij de uitleg en analyses was altijd praktisch: hoe kan ik goed gebruik van statistiek bevorderen om specifieke onderzoeksproblemen adequaat op te lossen. Een dienstverlenende benadering, waarbij hij de inhoudelijke merites van een probleem nooit uit het oog verloor. Vanuit dat principe van hulpvaardigheid was het ook vanzelfsprekend dat hij in 1975 volop steun gaf aan de oprichting van de Methodologiewinkel, een initiatief van enkele gemotiveerde sociologiestudenten: statistische en methodologische eerstelijns hulp voor studenten en onderzoekers binnen de faculteit. Het is verheugend dat deze bijstandswinkel voor statistische problemen, die onder de paraplu van zijn vakgroep floreerde en landelijk navolging vond, nog steeds bestaat.

Ivo had zonder mankeren zichtbaar plezier in het hoofdmeesterschap, dat hij van huis uit had meegekregen. Zijn vader gaf hem tijdens de oorlog lange tijd zelf les: iedere dag lagere school aan huis in Putten. Bij zijn vader ligt wellicht ook de oorsprong van de woorden die jarenlang de voorkant van zijn syllabus *Voortgezette Statistiek* sierden, een citaat van de Duitse tekenaar en dichter Wilhelm Busch (1832–1908), dat luidt: *Wenn alles schläft und einer spricht, den Zustand nennt man Unterricht*. Zo hield ook hij, de meester, zijn leerlingen wakker.

Als emeritus professor liet hij zich over de leraarsrol in een interview het volgende ontvallen: ‘Een leraar moet een combinatie zijn van een zendeling en een toneelspeeler’ (Oud & Stemerink, 2002, p. 5). Zijn uitlegkunsten en liefde voor het vak statistiek zijn blijvend tastbaar in heldere en tot nadenken stemmende publicaties, die op een breed publiek waren gericht (Molenaar, 1985, 2004). Zo ook zijn artikel uit 1977: *Ik word ziek van de statistiek, of: er van weten zonder er naar te handelen*. Wie wilde daar nu niet het zijne van weten? Zijn denksandwich, smakelijke statistiek tussen vóórdenken en nádenken (Molenaar, 1990), staat niet alleen bij oraties op het menu, maar is in Groningen ook verplichte kost voor sociologen die voortgezette statistiek volgen.

De onderzoeker

Wat het wetenschappelijk onderzoek betreft moeten boven alles Ivo’s verdiensten voor de psychometrie worden genoemd. In de decennia na het verschijnen van klassieke studieboeken als *Statistical theories of mental test scores* van Lord en Novick in 1968 en *Einführung in die Theorie psychologischer Tests* van Fischer in 1974, ontwikkelde de item-response-theorie (IRT) zich snel. In die opbloei van de IRT heeft Ivo een belangrijke rol gespeeld.

Ivo Molenaar neemt na zijn afscheidsrede op 12 september 2000 de voor hem samengestelde afscheidsbundel in ontvangst



Zijn visie op de psychometrie kan worden gekarakteriseerd door de definitie die hij in 1999 op de Europese bijeenkomst van de Psychometric Society in Lüneburg gaf: 'Psychometrics is mathematical statistics in the service of substantive psychology.' Deze uitspraak weerspiegelt zijn wetenschappelijke inzet en carrière als geheel: na het voltooien van een proefschrift in de wiskundige statistiek en het bekleden van de voor Nederland unieke, nieuwe Groningse leerstoel voor statistische analyse en meettheorie, kreeg hij in toenemende mate belangstelling voor de psychometrie. Wat dat betreft lijkt zijn loopbaan op die van Georg Rasch (1901–1980). Hoewel Ivo op vele gebieden werkzaam is geweest, heeft hij een groot deel van zijn kennis en vaardigheden ten dienste van de psychometrie gesteld. Van aanvankelijk belangstellende ontwikkelde hij zich tot een invloedrijke onderzoeker binnen het vakgebied van de IRT, niet alleen in Nederland maar ook internationaal.

Ivo's publicaties op het terrein van de IRT bestrijken een breed gebied, waaronder de *passing* van items en personen in IRT-modellen. Zijn grootste wetenschappelijke bijdrage was de ontwikkeling van een niet-parametrisch model voor polytome items, met daarbij behorende programmatuur (Sijtsma, Debets & Molenaar, 1990). Ivo bedacht een praktisch belangrijke uitbreiding van het schaalanalysemodel voor dichotome items van Mokken (1971), waarbij de wiskundige uitwerking van het begrip itemstap hem over ordinale drempels trok (Molenaar, 1982, 1983, 1997).

Voorts heeft hij met de Oostenrijker Gerhard Fischer een boek over Rasch-modellen geredigeerd (Fischer & Molenaar, 1995) en schreef hij met zijn leerling Klaas Sijtsma een inleidend studieboek over niet-parametri-

sche IRT (Sijtsma & Molenaar, 2002).

Ivo heeft zich altijd ingezet voor samenwerking met Nederlandse zusteruniversiteiten. Het beste voorbeeld daarvan is zijn betrokkenheid bij de oprichting in 1987 van het IOPS, de Interuniversitaire Onderzoeksschool voor Psychometrie en Sociometrie, en zijn zorg om haar latere wel en wee. Door zijn vasthoudende inzet slaagde Ivo er uiteindelijk in lokale facultaire en universitaire bestuurders te overtuigen van het belang van IOPS-deelname. Niet alleen was het IOPS als opleidingsinstituut voor jonge onderzoekers zijn tijd ver vooruit, maar het bood Ivo en zijn collega's in Nederland ook de mogelijkheid om met hun promovendi nieuwe terreinen van de IRT met succes te ontginnen.

Het was dan ook met vreugde en enige trots dat hij in het jaar 2000 bij zijn academisch afscheid het boek *Essays on item response theory* in ontvangst nam. Een door leden van zijn vakgroep geredigeerd boek met bijdragen van nationale en internationale IRT-onderzoekers (Boomsma, Van Duijn & Snijders, 2000).

Een ander onderwerp waar Ivo zich toe voelde aangehouden was de Bayesiaanse statistiek. Al tijdens zijn studie las hij Savage en Di Finetti, maar bij zijn Amsterdamse leermeesters kreeg de Bayesiaanse opvatting van het kansbegrip destijds weinig bijval: voor Van Dantzig was er sprake van afgoderij en Van Zwet had het over zwarte kunst, terwijl Hemelrijk van hem verlangde de persoonlijke kans van een Bayesiaan met de term 'pans' aan te duiden, zo memoreerde hij zelf (Molenaar, 2000, p. 6). Niettemin bleef hij zich voor een Bayesiaanse aanpak interesseren.

Eenmaal in Groningen werd hij geïnspireerd door een boek van Novick en Jackson (1974) over toegepaste Bayesiaanse statistiek voor de sociale wetenschappen.

In 1975 kwam Melvin Novick op voorspraak van Ivo's collega en vriend Willem Hofstee twee weken naar Groningen, onder meer voor het geven van een meerdaagse cursus Bayesiaanse statistiek. Daarmee was Novick de eerste docent in een lange reeks postacademische cursussen, die bijna jaarlijks door de vakgroep Statistiek en Meettheorie zouden worden georganiseerd. Ivo's affiniteit voor Bayesiaanse statistiek werd mede gestimuleerd door vele discussies met Willem Hofstee, bekend van diens weddenschapmodel (Hofstee, 1980), en door de komst van Charles Lewis die met Novick had samengewerkt en in 1978 als vaste medewerker tot de vakgroep toetrad. Zo ontwikkelde zich gaandeweg een Bayesiaans netwerk, waarbinnen Ivo inhoudelijk veel belangstelling had voor pogingen om subjectieve waarschijnlijkheden te meten, het eliciteren van a-priori-kansverdelingen bij personen.

Over de richtingenstrijd tussen frequentistische en Bayesiaanse statistiek zei hij later: '... ik ben een beetje een eclecticist geworden' (Oud & Stemerding, 2002, p. 9). Men zou zich na lezing van een eerder interview met Van Zwet (Oud, 2000) kunnen afvragen of Ivo daarin wel zoveel met Van Zwet van mening verschilde, maar voor de laatste bleef de Bayesiaanse aanpak absoluut een tweede keus. Ivo daarentegen had meerdere promovendi die zich expliciet op Bayesiaanse onderwerpen richtten. Zijn leerling Herbert Hoijtink – gepromoveerd in de IRT, maar snel daarna 'omgezwaaid' – werd in 2003 hoogleraar toegepaste Bayesiaanse statistiek aan de Universiteit Utrecht.

Internationaal speelde Ivo een rol als hoofdredacteur van *Psychometrika*, het lijfblad van de Psychometric Society, en wel van 1984 tot en met 1988, Vols. 48(3) – 53(2). Het was een positie waar hij zeer mee ingenomen was. Gedurende 1997–1998 was hij voorzitter van diezelfde vereniging. Wie uit die periode iets van zijn opvattingen over statistiek wil proeven leze nog eens zijn toespraak voor de Psychometric Society, getiteld: *Data, model, conclusion, doing it again* (Molenaar, 1998). Daarnaast is hij redacteur geweest van drie tijdschriften: *Statistica Neerlandica*, *Computational Statistics and Data Analysis* en *Journal of Educational Statistics*.

Bij al zijn activiteiten ging Ivo openbare discussies nooit uit de weg. Een goed voorbeeld daarvan is het in 1987 door de VVSOR georganiseerde debat over de formele statistische aanpak versus een informele data-analyse. De Groningse tegenover de Leidse benadering, zoals men destijds zei, respectievelijk vertegenwoordigd door Ivo Molenaar en Jan de Leeuw. In *Statistica Neerlandica* is van de krachtmeting gedetailleerd verslag gedaan (Mo-

lenaar, 1988; De Leeuw, 1988; De Gruijter, 1988). Ivo's bijdrage had een pakkende titel: *Formal statistics and informal data analysis, or why laziness should be discouraged*. De Leeuw sloot zijn artikel, *Models and techniques*, af met een prikkelende conclusie: 'Statistics is data analysis. ... statistics has always been data analysis' (p. 98).

Zulke openbare twistgesprekken, eerder met Eddy Roskam over de (on)bruikbaarheid van het Mokken-model, waren een uitdaging voor Ivo. Ze markeren hem als intellectueel, in de betekenis die Lolle Nauta er ook aan gaf: hartstocht voor het publieke debat (cf. Nauta et al., 1993).

Het eindpunt

Na het einde van zijn loopbaan werd Ivo in 2002 benoemd tot erelid van de VVSOR. Aanleiding voor Han Oud en Gerrit Stemerding om hem voor *STATOR* te interviewen (Oud & Stemerding, 2002). Eerder had hij in een lustrumboek van de Rijksuniversiteit Groningen al een verhelderende toelichting op zijn liefde voor het vak statistiek gegeven (Molenaar, 1994).

Voor de statistiek en de psychometrie heeft hij steeds zijn best gedaan, indachtig het motto van het Koninklijk Wiskundig Genootschap dat hem zeer aansprak: 'Een onvermoeibare arbeid komt alles te boven.'

Tot het voor hem, Ivo Wouter Molenaar, de markante en bescheiden leraar en onderzoeker, niet meer ging. Hij heeft ons geleerd dat we het statistisch landschap, dat hij zelf koesterde en heeft verrijkt, voor een breed publiek toegankelijk en aantrekkelijk moeten maken. Hij heeft onderzoekers geleerd hoe zij zich in het open veld, met een aantal vrijheidsgraden, naar behoren kunnen gedragen.

* Deze terugblik is een bewerking en aanvulling van mijn rede uitgesproken ter gelegenheid van Ivo's afscheid als hoogleraar op 12 september 2000 in de Aula der Rijksuniversiteit Groningen. Ik dank Marijtte van Duijn en Tom Snijders voor waardevolle kritische opmerkingen.

LITERATUUR

- Boomsma, A., Van Duijn, M.A.J., & Snijders, T.A.B. (Eds.). (2000). *Essays on item response theory* (Lecture Notes in Statistics, 157). New York: Springer.
- De Gruijter, D.N.M. (1988). Data analysis and statistics, report of a discussion. *Statistica Neerlandica*, 42(2), 99–102.
- De Leeuw, J. (1988). Models and techniques. *Statistica Neerlandica*, 42(2), 91–98.
- Fischer, G. (1974). *Einführung in die Theorie psychologischer Tests: Grundlagen und Anwendungen*. Bern: Huber.
- Fischer, G.H., & Molenaar, I.W. (Eds.). (1995). *Rasch models: Foundations, recent developments, and applications*. New York: Springer.

Hofstee, W.K.B. (1980). *De empirische discussie. Theorie van het sociaal-wetenschappelijk onderzoek*. Meppel: Boom.

Hogg, R.V., & Craig, A.T. (1970). *Introduction to mathematical statistics* (3rd ed.). London: Collier-Macmillan.

Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Mokken, R.J. (1971). *A theory and procedure of scale analysis: With applications in political research*. The Hague: Mouton.

Molenaar, W. (1970). *Approximations to the Poisson, binomial and hypergeometric distribution functions*. Amsterdam: Mathematisch Centrum.

Molenaar, W. (1972). *Dit is een uitdaging*. Oratie, Rijksuniversiteit Groningen.

Molenaar, W. (1977). Ik word ziek van de statistiek, of: er van weten zonder er naar te handelen. *Mens en Maatschappij*, 52(1), 58–71.

Molenaar, I.W. (1982). Mokken scaling revisited. *Kwantitatieve Methoden*, 3(8), 145–164.

Molenaar, I.W. (1983). *Item steps* (Heymans Bulletin, HB-83-630-EX). Rijksuniversiteit Groningen, Psychologische Instituten.

Molenaar, I.W. (1985). Statistics in the social and behavioral sciences. *Statistica Neerlandica*, 39(2), 169–179.

Molenaar, I.W. (1988). Formal statistics and informal data analysis, or why laziness should be discouraged. *Statistica Neerlandica*, 42(2), 83–90.

Molenaar, I.W. (1990). Statistiek smaakt beter in een denksandwich. *Psychologie en Maatschappij*, 53, 357–366.

Molenaar, I.W. (1994). Verliefd op statistiek. In G. Gritter (Red.). ... *waarvan akte. Groninger geleerden en hun fascinatie voor de wetenschap* (pp. 104–111). Groningen: Passage.

Molenaar, I.W. (1997). Nonparametric models for polytomous items. In W.J. van der Linden & R.K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 369–380). New York: Springer.

Molenaar, I.W. (1998). Data, model, conclusion, doing it again. *Psychometrika*, 63(4), 315–340.

Molenaar, I.W. (2000). *Vast versus variabel: de statistische splitsing*. Afscheidscollege, Rijksuniversiteit Groningen.

Molenaar, I.W. (2004). About handy, handmade and handsome models. *Statistica Neerlandica*, 58(1), 1–20.

Nauta, L., De Vries, G., Harbers, H., Koenis, S., Mol, A., Pels, D., & De Wilde, R. (1993). *De rol van de intellectueel. Een discussie over distantie en betrokkenheid*. Amsterdam: Van Gennep.

Novick, M.R., & Jackson, P.H. (1974). *Statistical methods for educational and psychological research*. New York: McGraw-Hill.

Oud, H. (2000). Interview met prof. dr. W.R. van Zwet, benoemd tot erelid van de VVS. *STATOR*, 1(1), 4–8.

Oud, H., & Stermerdink, G. (2002). Interview met prof. dr. W. Molenaar, benoemd tot erelid van de VVS. *STATOR*, 3(1), 4–9.

Sijtsma, K., Debets, P., & Molenaar, I.W. (1990). Mokken scale analysis for polychotomous items: Theory, a computer program and an empirical application. *Quality & Quantity*, 24(2), 173–188.

Sijtsma, K., & Molenaar, I.W. (2002). *Introduction to non-parametric item response theory*. Thousand Oaks, CA: Sage.

Thirty years LNMB

Ladies and gentlemen,

Thank you so much for inviting me and many congratulations! LNMB, the Dutch Graduate Network in the Mathematics of Operations Research, is thirty years of age. At thirty, energy is still present and wisdom is approaching. A fruitful combination, and you will need it.

The Network was born in 1987. I would like to argue that it was not a singularity, a one-of-a-kind animal of an unknown species, but a step in a long and natural evolution.

Let me go back further, to the sixties of the previous century. Dutch academia was faced by developments for which it was ill-prepared. The babyboom generation arrived: more young people, of which a much higher ratio went on into higher education. The country got many more students. Next to education, those students also wanted more democracy in academia. This does not bear on my story and it died out due to natural causes. But what did not die out: they also wanted more mathematics, applied mathematics, needed by a society of increasing complexity and facilitated by the advent of computing equipment. Numerical analysis, statistics, and operations research.

The first generation of OR faculty included some giants, who set the stage for Dutch OR. First, Wim Cohen, of the single-server queue. Then, Jacques Benders, a household name in decomposition methods for linear programming. Also Guus Zoutendijk, of the feasible directions method in nonlinear optimization. He moved into industry and politics, and you may not know his name. But at the Optimization Symposium in Atlanta in 2000, eight awards were given to the founders of the field, and Zoutendijk was the only European among them. And finally Gijs de Leve, who worked on Markov decision processes and became, in OR, the embodiment of the sense of nationwide cooperation that set out to rebuild our country after the war.

It was De Leve who, in the seventies, made the first step towards what you could call the Dutch University of

OR. He made a traveling salesman tour visiting all his colleagues and started the Lunteren meetings in 1976. Strangely enough, not everyone thought this was such a great idea, but most did. It worked, we met, we talked, we got to know each other.

The second step was the start of the Dutch Research Community in the Mathematics of OR and System Theory in 1979, chaired by Jaap Wessels. Wessels was a clever strategist, who operated in the background and had a big and true impact on Dutch OR. The Research Community solidified our cooperation, under the auspices of the Dutch Science Council. For some time, it even gave us a voice in evaluating research proposals on a national scale. Being in one bag with systems and control was more a political necessity than born out of love and mutual understanding, a marriage de convenance that came to a peaceful end when both disciplines had gotten enough momentum of themselves.

And then, as the third step, we got the Graduate Network in 1987, after several years of hard labor by Wim Klein Haneveld, who did all the heavy groundwork and became its first director. De Leve, Wessels and Klein Haneveld are the three unsung heroes of the unity of Dutch OR. I know, because I served De Leve and Wessels as secretary and Klein Haneveld as chair.

As I said, all these steps were phases of an evolution, carefully stage-managed by our government, with the grand purpose of increasing the quality of Dutch academic education and research after the explosive growth of the sixties and seventies. Research communities, graduate networks, in later years conditional funding, research schools, and what have you, they all served this single purpose. It worked, step by step.

Founding the Network was not easy. Our ally in the government was Roel in 't Veld, deputy secretary of education and science, and our example was Arie Kapteyn, then in Tilburg, the founder of the Network for General and Quantitative Economics, for which he obtained several millions of guilders from the government. I knew Arie, Alexander Rinnooy Kan knew Roel, both of us knew

Wim Klein Haneveld, the three of us managed to get 2.6 million guilders, and the rest is history.

This is a simple summary of a complex process. If you wish to know more, please read the interview with Wim and me in the *Nieuw Archief voor Wiskunde* of June 2016. Let me just say that economists and system and control engineers are pragmatic, while some mathematicians tend to prefer a small-scale and more individual approach. It took Wim quite some massaging before the mathematicians of OR reached a consensus.

But even mathematicians accept a reality once it exists and the Network was off to a flying start. From what I understand, it is still flying, with a broad program covering foundations and applications, taught by the very best people we have. Now, thirty years later, the academic landscape has changed. The Network has survived and even reinforced many developments in Dutch academia. It still provides high-level graduate and master classes to universities and research schools, acting as the OR department of the University of the Netherlands.

And the Network is more relevant than ever. In the new century, university education is going through another wave of expansion. This is especially true for the mathematical sciences, which have evolved from the language of the natural sciences to the language of society. Academia has to respond to an increased demand, as in the sixties, but with restricted funding, unlike the sixties. The matching of funds for education and research is less evident than it has been. On both sides, in education and research, we need, more than ever, national cooperation rather than competition.

Your chairman asked me to give you some advice for the future. I'll try. First, there are only three criteria: quality, quality, and quality. Please continue to invite the very best speakers for Lunteren and the very best teachers for the Network. My second advice is a corollary: don't listen to me. I am seventy. Listen to the young people. They have the brains and the future. Thank you very much.

JAN KAREL LENSTRA

Het Landelijk Netwerk van Mathematische Besliskunde (LNMB) bestaat 30 jaar. Tijdens de LNMB-conferentie die op 17 januari 2018 in Lunteren plaatsvond, blikte Jan Karel Lenstra terug op de ontwikkelingen en de succes van de afgelopen 30 jaar

EURO WORKING GROUP ON OR IN PRACTICE
A report from the first workshop*



RUTH KAUFMAN

Exchanging ideas, knowledge and experiences with other professionals is one of the most effective, reliable and enjoyable ways of developing your own professional practice. The EURO Working Group on OR in Practice (EWG-POR) was established to support such exchanges between OR/analytics practitioners on a European level, emulating EURO's enormous success in building networks across OR research.

The first event took place last February, with a mix of presentations and working sessions on the theme 'measuring the impact of OR projects'. An intensive set of 15-minute short talks and 3 longer keynotes covered an enormous variety of application areas.

The spirit of the event was 'The wise man learns from his own mistakes, the wiser man learns from the mistakes of others' and speakers and audience alike were remarkably open about failures, mistakes and challenges. Several common issues emerged:

- You forget at your peril the 'people being optimised' in routing, scheduling, and rostering, such as train drivers who don't want more than their fair share of routes with stroppy customers.
- Equally crucial is the layer above the 'optimisees': the planners, dispatchers and administrators whose work is transformed by OR's systems. These people are competent: at least two speakers described how their first attempts at optimisation performed worse than the manual systems, as the operators could work around a constraint or were familiar with local exceptions that the machine knew nothing of.
- Listen to what the customer actually needs (one speaker vividly described a disastrous project arising from a 'dialogue of the deaf'). In measuring impact, understand what actually matters to them.
- Few OR implementations justify the authoritative language of case studies: 'We saved £Xmillion'. Paper savings are certainly misleading, as field experience with actual people (eg subject to supplier delays) in actual physical settings (eg loading areas not large enough for the numbers of trucks scheduled) is different. 'Measuring impact is misleading and error-prone' one speaker concluded.

For some participants, the most exciting part was the three working sessions, on models for improving and standardising the way we measure impact; the critical factors for successful projects; and the opportunities and changes brought by machine learning. They valued the opportunity to discuss, with peers from other countries and industries, some factors that are key to doing OR better but that are rarely addressed in conferences or the literature.

For me as a UK practitioner, used to meeting other practitioners and 'practice-oriented academics' and discussing 'Making an Impact' at events like the OR Society conference, one thing stood out. Almost every speaker, and almost everyone in the audience, was mainly engaged in optimisation: routing, scheduling, location, and more. Whether this reflects this particular audience, or all European practitioners, I cannot say. However, it was clear that:

- OR optimisation methods are hugely influential and being widely and successfully applied;
- The practical challenges that in the UK have led to practitioners' equal interest in simulation, systems approaches, MCDA, problem-structuring methods and so forth, are universal and recognised, even if the methods are not.

The formal output from the working sessions is on the EWG-POR website. The next meeting of EWG-POR will take place as part of the EURO conference in Valencia, July 8-11, 2018.

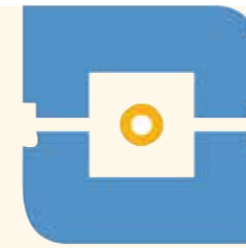
* This text is a slightly abridged version of what was originally prepared for the UK OR Society's newsletter *Inside OR*.

SOME USEFUL LINKS

EWG-POR website: <http://www.euro-online.org/websites/or-in-practice/>

EURO2018 website: <http://euro2018valencia.com/making-an-impact/#>

RUTH KAUFMAN is a former president of the OR Society and chairperson of the POR working group.
 E-mail: ruth.kaufman@btinternet.com



The past weeks have been filled with activities for Young Statisticians in the Netherlands. Over these weeks we have met a diverse group of people interested in statistical applications, and learned a lot from events and company visits. We visited three companies, went to talks and activities associated with the VVSOR day and gained a new board member.



Company visits to SciSports, Marktplaats and Algemene Rekenkamer

As many of our members are still students or taking their first career steps, we organize regular visits at companies that rely on statistical modeling for our members to get an idea of the diverse use of statistics outside of academia and a glimpse of how it feels to work at these companies.

In the beginning of March we visited SciSports, and learned about their ways of analyzing historical performance data of football players to estimate player qualities.

In the middle of April, 30 people frequented our visit to Marktplaats (eBay) in Amsterdam. During six presentations we learned what kind of data they have, how they use it to optimize their revenue and we got a good impression of how it would be like to work at Marktplaats.

In the end of April, we visited Algemene Rekenkamer (Netherlands court of audit) to learn about the way they audit governmental institutions. We experienced the work as a data analyst at Algemene Rekenkamer in a case study where we constructed research questions, analysed data and communicated the results to employees of the Algemene Rekenkamer.

VVSOR days

For the VVSOR Annual Meeting, Young Statisticians organized both a lunch for our members as well as a Pub Quiz for all of VVSOR. On the first day of the meeting, members of YS gathered together to enjoy food and drinks, get to know each other and talk about statistics and life. After the lunch, we proceeded to the talks

organized by the VVSOR where we learned among other things about factors affecting the climate. In the evening we hosted a few pub quiz rounds for all members of the conference to get together over questions about climate change, statistical software, R, music and more!

Young Statisticians Board

In the end of January we welcomed a new board member to our team: Hannah Rós Sigurdardóttir. Originally from Iceland, Hannah moved to the Netherlands in 2015 and is currently a PhD student at Tilburg University, working on time series data used in clinical settings. She will be in charge of the soon upcoming content on the YS website.



We are excited about our future company visits, statistics cafés and events where we get to network with other people sharing our love for statistics, and gather knowledge and insights into even more fields of study and application.

