

STAtOR

periodiek van de VVS jaargang 13, nummer 1, maart 2012

Programma van de Dag voor Statistiek en OR 2012
'Statistics for the millions: presenting statistics and operations research to the public'

Polymetrisch metselen

Het complexe netwerk van onze hersenen

Over regenkansen en paraplu's

Optimale experimentele studies voor functionele MRI

Veiligheid in de zorg. Draagt operations research bij aan betere zorg? (Deel 2)

Veel medische studies kunnen en moeten kleiner

Statistiek + OR is sexy !

In het augustus 2011 issue van het tijdschrift voor de leden van de American Statistical Association schreef de toenmalige President van de ASA Nancy Geller een prachtige column, met de titel: “Don’t shun de ‘S’ Word”. Dat zou met dezelfde alliteratie in het Nederlands worden: Schuw het ‘S’ woord niet! Het ‘S’ woord is oorspronkelijk een woord dat kinderen in de US niet van hun ouders mochten gebruiken. Maar Nancy Geller wil het gebruik van het ‘S’ woord juist bevorderen, alleen wel in de betekenis van Statistics. Haar column is mij uit het hart gegrepen.

Het lijkt wel of allerlei andere gebieden de statistiek over hebben genomen (ik gebruik hier de term ‘statistiek’ als pars pro toto voor ‘statistiek en operations research’, zoals in de naam van onze vereniging). Aan de ene kant hebben we machine learning en data mining waar N (het aantal objecten) heel erg groot is, en ook hebben we omics (zoals genomics en metabolomics) en GWAS (genomewide association studies) waar het aantal variabelen $P \gg N$. Er gebeurt heel veel, maar men noemt het geen statistiek. Het helpt ook niet als grote softwarepakketten statistiek en operations research samenvoegen onder de als sexy bedoelde term ‘Business Analytics’.

Het is schrijnend: terwijl de toegepaste statistiek in allerlei onderzoek een vernieuwende rol zou moeten spelen, lijkt het of men de term statistiek of statistische analyse niet meer wil gebruiken. Wellicht vindt

men het ouderwets en saai klinken, of misschien ook wel duidend op leugens en bedrog. Terwijl Hal Varian (de topeconoom van Google) al meer dan drie jaar geleden in een interview (in The McKinsey Quarterly) de inmiddels beroemde uitspraak deed: “I keep saying the sexy job in the next ten years will be statisticians... The ability to take data - to be able to understand it, to process it, to extract value from it, to visualize it, to communicate - it’s going to be a hugely important skill in the next decades...”.

Wij statistici hebben misschien een communicatieprobleem. En dit zal dan ook het thema zijn van onze komende jaarlijkse vergadering (op 29 maart 2012 in Utrecht): ‘Statistics for the Millions. How to present Statistics and Operations Research to the Public.’ De lezingen hebben allemaal op de een of andere manier als onderwerp hoe wij ons vak beter over het voetlicht kunnen brengen. En dat zal gebeuren door drie heel uitzonderlijke sprekers.

Sam Savage, Nick Fisher en Ionica Smeets zijn niet alleen experts, maar vooral ook heel erg onderhoudend. Het belooft dus niet alleen een interessante, maar vooral ook een vermakelijke dag te worden. Hoezo statistiek saai? Statistiek+OR is sexy!

Jacqueline Meulman

Thursday, March 29th 2012

Annual meeting of the Netherlands Society of Statistics
and Operations Research (VvS+OR)



The annual meeting of the
Netherlands Society of
Statistics and Operations
Research is the place to
meet other people with
an interest in statistics
and operations research
and people working in this
field.

STATISTICS FOR THE MILLIONS: presenting statistics and operations research to the public

LOCATION

Jaarbeurs Utrecht, Room 417,
Jaarbeursplein 6, 3521 AL Utrecht (near Central Station)

KEYNOTE SPEAKERS

Sam Savage (Stanford University)
Nicholas Fisher (ValueMetrics Australia, University of Sydney)
Ionica Smeets (Science journalist, LIACS Leiden University)

PROGRAM

9.30 - 10.00 Registration and coffee/tea
10.00 - 10.15 Opening and introduction first speaker
10.15 - 11.15 Sam Savage, Stanford University
11.15 - 11.30 coffee/tea
11.30 - 12.30 Annual General Meeting (ALV)
12.30 - 13.30 Lunch break (lunch not included)
13.30 - 13.45 Introduction next speaker
13.45 - 14.45 Nicholas Fisher, ValueMetrics Australia, University of Sydney
14.45 - 15.15 VvS+OR thesis award presentation
15.15 - 15.30 coffee/tea
15.30 - 16.00 VvS+OR thesis award presentation
16.00 - 16.15 Introduction next speaker
16.15 - 17.15 Ionica Smeets, Science journalist, LIACS Leiden University
17.15 Drinks

About Nicholas Fisher

Dr Nicholas Fisher is the founder of ValueMetrics Australia, an enterprise that carries out research and consulting primarily in the area of Performance Measurement. After 30 years as a statistical consultant and researcher in CSIRO (Commonwealth Scientific and Industrial Research Organisation), he left his position as a chief research scientist in May 2001. Whilst in CSIRO, he led the development of CSIRO's Organisational Performance Measurement (OPM®) system, which has been applied successfully in a number of private and public enterprises, and has been part of graduate programs at Wollongong University and Mount Eliza Business School. ValueMetrics Australia is one of 20 Core Participants in the Invasive Animals Cooperative Research Centre, with primary responsibility for monitoring national community attitudes using a proprietary community value approach.

Nick carries out research and consulting in performance measurement, with particular emphasis on improving quantitative reports to boards and top management, focusing on stakeholder satisfaction (customers, employees, partners...). He is currently completing a book on performance measurement. Recent and current clients include the Basque Statistical Office, CSIRO, the Australian Bureau of Statistics, John Holland Group, the Lane Cove Tunnel Project, Purdue University, Imperial College (London), La Trobe University and Life Technologies (Glasgow). Formal qualifications include a PhD from the University

Statistics for the Millions: How to Communicate Statistics to the Public. WHAT? WHEN? WHERE? WHO? WHY?

Nicholas Fisher

ValueMetrics Australia and University of Sydney

ABSTRACT

The subtext for this talk is:

“O wad some Power the giftie gie us
To see oursels as ithers see us!”

The topic of this annual VvS meeting begs some fundamental questions. Why should we bother communicating Statistics to the public? What might they want to know? Who, in fact, cares? Where are these people who might care? When should we communicate with them? (And there are others, not beginning with the letter ‘W’.) I shall address these in a circumlocutory fashion, in the hope that some potentially interesting ways forward may emerge from the gloom.

The subtext comes from ‘To a Louse. On Seeing one on a Lady’s bonnet at Church’, a poem by Robert Burns:

The poem begins

Ha! whare ye gaun, ye crawlin ferlie!
Your impudence protects you sairlic
I canna say but ye strunt rarely
Owre gauze and lace
Tho’ faith, I fear, ye dine but sparely
On sic a place

Ye ugly, creepin, blastit wonner
Detested, shunn’d, by saunt an’ sinner
How daur ye set your fit upon her
Sae fine a Lady!
Gae somewhere else and seek your dinner
On some poor body

and so on, and concludes with

O wad some Power the giftie gie us
To see oursels as ithers see us!
It wad frae mony a blunder free us,
An’ foolish notion:
What airs in dress and gait would leave us,
And even devotion!

of North Carolina, and a DSc from the University of Sydney. He is professionally accredited by the Statistical Society of Australia and by the American Statistical Association and is a visiting professor of statistics at the University of Sydney.

which has been translated (www.robertburns.plus.com/louse.htm) as
Oh, that God would give us the very smallest of gifts
To be able to see ourselves as others see us
It would save us from many mistakes
and foolish thoughts
We would change the way we look and gesture
and to how and what we apply our time and attention.

2

About Sam Savage

Sam L. Savage is Consulting Professor, Stanford University and Chief Scientist, Vector Economics.

He is the author of *The Flaw of Averages, Why we Underestimate Risk in the Face of Uncertainty* (John Wiley 2009). In 2008, Sam invented a new data type, the Distribution String, which, according to Nobel Laureate Harry Markowitz, 'represents a major breakthrough in the communication of risk and uncertainty. It significantly widens the practical applicability of sound theory in these areas.' In 2010, Sam co-founded Vector Economics, Inc., which provides Probability Management solutions by integrating the latest simulation, data, and cloud computing technologies.

Sam received a PhD in the area of computational complexity from Yale University, and holds the positions of Consulting Professor of Management Science and Engineering at Stanford University, and Fellow at Cambridge University's Judge Business School. He has consulted and lectured extensively to industry and has been called 'The Edward Tufte of Risk'.

Probability Management: A cure for the Flaw of Averages, or Doing for Distributions what Tesla did for Electricity

Sam L. Savage
Stanford University

ABSTRACT

In spite of the widespread teaching of probability and statistics, many managers have no intuitive feeling for the subject, and are oblivious to the implications of such pervasive concepts as the central limit theorem and Jensen's Inequality. I see three hurdles which must be overcome to increase the benefits and audience for statistics and stochastic operations research models. One is technical, one informational, and one managerial.

From a technical perspective, ultra high speed simulations driving graphical user interfaces will connect the seat of the intellect to the seat of the pants. These may be viewed as new light bulbs for illuminating uncertainty.

From an informational perspective, new data types can collect representations of probability distributions from statisticians and operations researchers (producers), and deliver them to managers (consumers) for use in the interfaces described above, potentially creating the new industry of Distribution Distribution. This is the electricity to light the bulbs.

From a managerial perspective, managers within organizations must begin to take responsibility for the probabilistic assumptions that they are distributing. The ultimate such responsibility must lie with what I call the Chief Probability Officer (CPO). This is the electric power commissioner who tries to prevent accidental electrocution.

3

About Ionica Smeets

Ionica Smeets has a PhD in mathematics and writes columns for *de Volkskrant* about everyday maths under the name Wiskundemeisjes (Math girls). She also writes for several other Dutch newspapers and magazines, published two popular-scientific books, presents tv-shows and gives many talks. Mainly about mathematics, because 'someone has to do it'.

Against all odds

Ionica Smeets

Science journalist, LIACS Leiden University

ABSTRACT

How can you tell the general public about the difference between correlation and causality? Or learn them how to spot a fake random pattern? And how do you show that averaging data can ruin everything?

Ionica Smeets explains how she does this in talks for high school students, managers and other people who know zilch about statistics. She uses examples from every day life, small experiments and clips from popular tv-shows.

Registration (before March 19th)

Members VvS+OR

The annual meeting is free for members but registration through the society's website is required: <www.vvs-or.nl>. Register before March 19th and bring the confirmation email and show it at the registration desk at the entrance of the lecture room.

Non-members

Non-members can either become a member (go to www.vvs-or.nl and click "Become a Member" in the top menu) and then immediately also have free access after registration (see above), or pay 50 euro (bank account: 202091, VvS+OR, 'registration annual meeting') before March 19th and bring proof of payment (copy of bank statement) to show at the registration desk at the entrance of the lecture room.

Organizing committee

The annual meeting is organized by the board of the VvS+OR. For questions, contact: Irene Klugkist (secretary) by e-mail at <info@vvs-or.nl>.

General Information

The annual meeting will be English spoken, except for the Annual General Meeting (ALV).

The Annual General meeting (ALV) is scheduled at the end of the morning. The relevant documents will be provided on the website two weeks before the meeting. You can also get them by email if you send a request to <info@vvs-or.nl>.

Coffee/tea during the breaks and drinks afterwards are offered by the Society.

Lunch is not included. The restaurant of the Jaarbeurs may be busy, but at a walking distance of just a few minutes (e.g. in direction central station) you will find several alternatives.

More information: <www.vvs-or.nl>

The annual meeting is sponsored by Genstat, Cosinus Computing BV, Ortec and Paragon.

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

Redactie

Joaquim Gromicho (hoofdredacteur), Ana Isabel Barros, Johan van Leeuwen, Mirjam Moerbeek, Gerrit Stermerdink (eindredacteur), Hilde Tobi. Vaste medewerker: Fred Steutel

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 VVS

Voorzitter:

prof. dr. Jacqueline Meulman <president@vvs-or.nl>

Secretaris:

dr. Irene Klugkist <bestuur@vvs-or.nl>

Penningmeester:

dr. Ad Ridder <penningmeester@vvs-or.nl>

Studentlid:

Maarten Kampert (Bsc) <student@vvs-or.nl>

Overige bestuursleden:

prof. dr. Fred van Eeuwijk (BMS), prof. dr. ir. Stan van Hoesel & dr. John Poppelaars (NGB), dr. Eric Cator (SMS), dr. Michel van de Velden (ECS), dr. Andries van der Ark (SWS),

Leden- en abonnementenadministratie van de VVS

VVS, Postbus 244, 6700 AE Wageningen, telefoon 0317 - 419572, fax 0317 - 421364, <admin@vvs-or.nl>.

Raadpleeg onze website over hoe u lid kunt worden van de VVS of een abonnement kunt nemen op STAtOR of op een van de andere periodieken.

VVS-website

www.vvs-or.nl

Advertentieacquisitie

Nikki Bisschop & Joren Brunekreef, Lange Nieuwstraat 6, 3512 PH Utrecht, 06-55874175, <adverteren.stator@vvs-or.nl>. STAtOR verschijnt in maart, juni, september en december.

Ontwerp en opmaak

Pharos | M. van Hootegem, Nijmegen

Druk

Drukkerij Zoetewij, Yerseke

Uitgever

© Vereniging voor Statistiek en Operationele Research
ISSN 1567-3383

Inhoud

- 2** Van de president: Statistiek + OR is sexy!
- 3** Programma van de Dag voor Statistiek en OR 2012
'Statistics for the millions: presenting statistics and operations research to the public'
- 8** Redactioneel: Kansen voor Statistiek en OR
- 9** Polymetrisch metselen
Jasper de Jong, Hana Vinduška, Erwin Hans & Gerhard Post
- 13** Het complexe netwerk van onze hersenen
Piet Van Mieghem
- 16** Over regenkansen en paraplu's
Maurice Schmeits
- 22** Young statisticians in de VVS+OR
- 23** Optimale experimentele studies voor functionele MRI
Bärbel Maus & Gerard J.P. van Breukelen
- 27** STAtOR-vleugel – column
Fred Steutel
- 28** Veiligheid in de zorg. Draagt operations research bij aan betere zorg? (Deel 2)
Joris van de Klundert
- 32** Veel medische studies kunnen en moeten kleiner
George Borm & Joanna in 't Hout
- 35** IM Jan Engel (1952-2011)
Bert Schriever
- 36** Wat bedoel je daarmee? – column
Gerrit Stermerdink



Dit nummer van *STATOR* bevat tevens het programma van de Dag voor Statistiek en OR. Het thema van de dag is *Statistics for the Millions* en de drie sprekers zullen zich richten op die éne vraag: hoe kunnen we ons vak op de juiste wijze onder de aandacht brengen van het publiek? Centraal staat daarbij het besef dat er vele kansen zijn voor toepassing van Statistiek en OR op alle mogelijke terreinen. Wij beoefenen een vak met een grote toekomst!

STATOR is al vanaf haar eerste nummer bezig ons vak voor een breder publiek dan alleen onszelf te presenteren. Inmiddels zijn er enkele honderden artikelen in ons blad verschenen waarin er steeds naar is gestreefd een duidelijke uitleg voor de niet-specialist te geven. De redactie denkt dat deze opzet redelijk is geslaagd, gezien de positieve reacties die we herhaaldelijk krijgen.

Polymetrisch metselen is een term die maar weinigen zullen kennen, maar Jasper de Jong en zijn mede-auteurs presenteren hiermee een geheel nieuwe en verrassende toepassing van OR.

Piet van Mieghem zet zijn specialisme telecommunicatienetwerken in voor het analyseren van het netwerk dat onze hersenen vormen. Het is dit soort toepassingen dat duidelijk maakt hoezeer onderzoek op één terrein vruchtbaar kan zijn voor een ogenschijnlijk geheel andere toepassing. Uiteindelijk profiteren beide terreinen van elkaar, en dat past weer uitstekend bij het thema van de Dag voor Statistiek en OR.

Er wordt wel eens gezegd dat er maar één onderwerp is waarover iedereen altijd meepraat: de weersverwachtingen. Maurice Schmeits van het KNMI vertelt ons nu wat we ons daarbij moeten voorstellen en wat een 60% regenkans eigenlijk is.

Een tweede artikel op medisch gebied is dat van Bärbel Maus en Gerard van Breukelen. Zij laten zien hoe men bij hersenonderzoek aanzienlijk kan besparen, zowel in termen van geld als van belasting voor proefpersonen.

De column van Fred Steutel ademt enige nostalgie. Maar zijn zorg is terecht, te vaak verdwijnen op ondoordachte wijze zaken waarvan we het verlies later zullen betreuren. Wat hij beschrijft is weliswaar van een geringere orde dan de bedreiging die het museum Boerhaave ten deel viel, maar toch...

Eveneens medisch gericht is het artikel van Joris van de Klundert, een vervolg op zijn in ons themanummer *Veiligheid* (*STATOR* 2011/3-4). Het is boeiend hoe ook hier door een beter begrip voor de toepassingen van ons vak een verbetering in de gezondheidszorg tot stand kan komen.

Ook George Borm en Joanna in 't Hout tonen ons dat met een juiste toepassing van statistiek de hoeveelheid proefpersonen in medisch onderzoek beduidend kan worden verminderd. Omgekeerd kan men ook zeggen dat bij gelijk blijven van dat aantal de voorspellende kracht kan worden vergroot. Hoe dan ook: een goede toepassing brengt winst. Hoewel dit artikel geschreven is vanuit een medisch perspectief zijn de resultaten bijzonder nuttig op vrijwel ieder onderzoeksterrein.

In dit nummer wordt Jan Engel herdacht, iemand die zich enorm heeft ingezet voor het juiste begrip van statistiek. Ook zijn artikelen in *STATOR* hadden dat doel. Een citaat uit het In Memoriam is kenmerkend voor zijn aanpak: 'Hij zag in dat je researchers geen statistiek moest leren, maar het in hun taal moest uitleggen'. Jan zou zonder meer in het rijtje sprekers van deze Dag voor Statistiek en OR hebben gepast.

Als reserve-columnist besluit Gerrit Stemerding dit nummer met een pleidooi voor precisie in definities en vraagstellingen.

De redactie wenst u, zoals gewoonlijk, weer veel leesplezier. Wij hopen dat u door dit nummer aangespoord wordt de Dag voor Statistiek en OR te bezoeken. Wellicht tot ziens daar!



POLYMETRISCH METSELEN

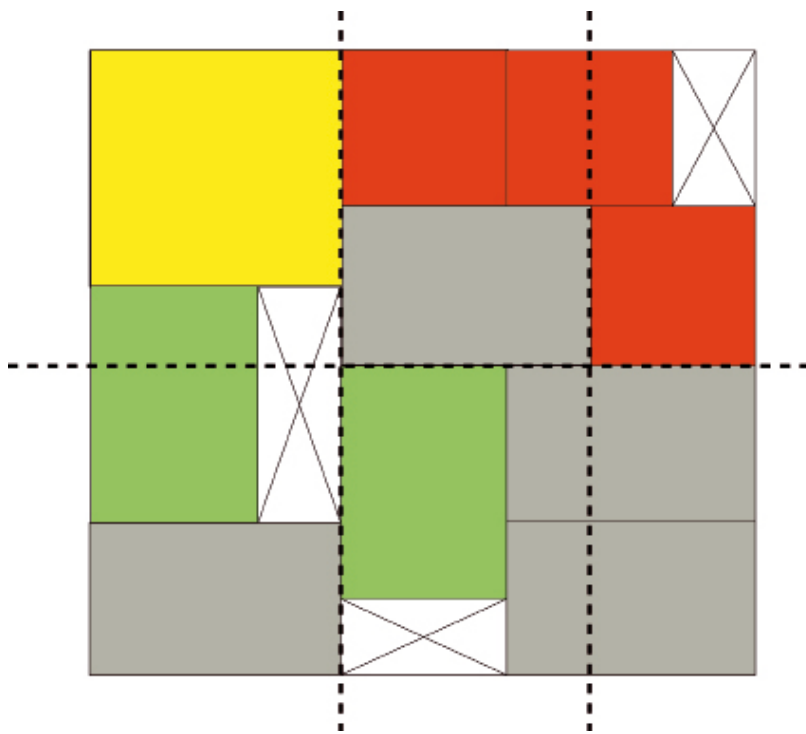
JASPER DE JONG, HANA VINDUŠKA, ERWIN HANS & GERHARD POST

Polymetrisch metselen? Dat is niet een onderwerp waar veel over is geschreven in de Operations Research literatuur. Vandaar de volgende korte uitleg. Soms worden muren gemetseld met stenen van verschillende afmetingen. Vroeger gebeurde dit met natuursteen, maar tegenwoordig wordt soms gekozen voor het metselen met stenen van (een beperkt aantal) verschillende afmetingen. De redenen zijn dan vaak esthetisch: de architect wil dan bijvoorbeeld een levendiger en gevarieerd gevelbeeld.

Bij polymetrisch metselwerk is het juist de bedoeling dat het metselwerk willekeurig lijkt. Toch is het mogelijk dat er wel degelijk een structuur aanwezig is. Zo kan het metselwerk een opeenstapeling zijn van zich herhalende *patronen* van hooguit enkele tientallen stenen. Het handmatig opbouwen van een dergelijk patroon is voor een architect of een bouwkundig tekenaar zeer tijdrovend werk. Het is een proces waarbij aan verschillende voorwaarden moet worden voldaan en dat wiskundig een interessant probleem oplevert. Wij zijn de uitdaging aangegaan dit probleem te onderzoeken.

In figuur 1 is een voorbeeld gegeven met 4 steenformaten, weergegeven in de kleuren geel, rood,

groen en grijs. Daarnaast bevat het patroon gesneden stenen, die weergegeven zijn als witte stenen met een kruis; dit zijn stenen die ontstaan door een steen van één van de gegeven steenformaten te verzagen. Er zijn verschillende voordelen bij het gebruik van patronen: de architect kan sneller een volledige muur ontwerpen (dit kan anders enkele dagen in beslag nemen), de verhouding in aantallen steenformaten is vast, en de metselaar kent op een gegeven moment het patroon, waardoor hij sneller kan werken. Het breken (snijden) van stenen kost de metselaar tijd, gesneden stenen zijn minder mooi (de natuurlijke bakvorm heeft een oneffen oppervlakte terwijl het gesneden oppervlak strak is) en vanuit economisch oogpunt is het snijden niet wenselijk omdat er dan ook reststukken als onnodig afval ontstaan. De belangrijkste doelstelling bij het maken van patronen is daarom het vermijden van gesneden stenen. Daarnaast moet het patroon mooi ogen. Wat dit precies betekent is deels een kwestie van smaak. Een voorkeur zal zijn dat de stenen van hetzelfde type mooi verdeeld zijn over het patroon; vooral grote stenen zullen niet naast elkaar geplaatst mogen worden. Daarnaast willen we graag dat voegen binnen het patroon onderbroken worden. Zo wordt



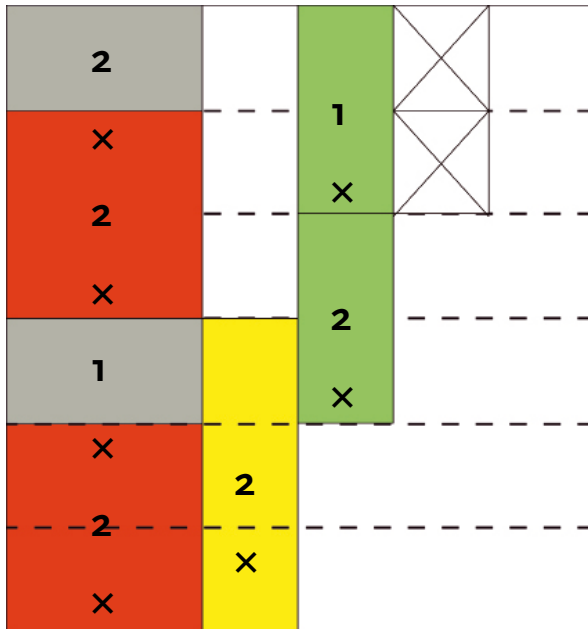
Figuur 1. Een patroon met 10 hele stenen

de horizontale voeg, weergegeven door de stippellijn, onderbroken, net als de tweede gestippelde verticale voeg. De eerste verticale gestippelde voeg echter niet. In het metselwerk als geheel moeten doorlopende verticale voegen vermeden worden voor de stevigheid van het metselwerk. Bovendien past de nadruk op horizontale gelaagdheid bij het gestapelde karakter van metselwerk. De patronen zullen daarom laagsgewijs horizontaal verschoven gebruikt worden.

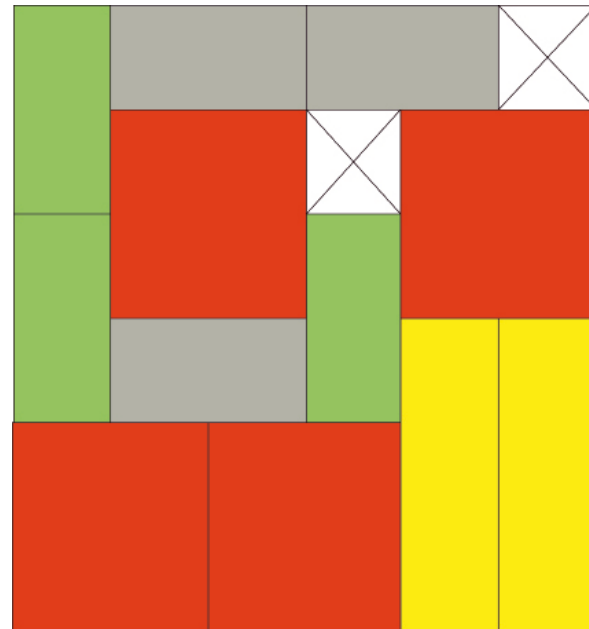
Het vinden van een mooi patroon met zo weinig mogelijk gesneden stenen is wiskundig gezien een interessant probleem. Dit probleem heeft veel gemeen met het Pallet Loading Problem, een 2-dimensionale variant van het Bin Packing Problem. Bij het Pallet Loading Problem moeten rechthoekige dozen op een rechthoekige pallet worden geplaatst, waarbij het doel is om een zo groot mogelijke oppervlakte te benutten. Hierbij lopen de randen van de dozen evenwijdig aan de randen van de pallet. De overeenkomsten tussen ons probleem en het Pallet Loading Problem zijn duidelijk als we de stenen zien als dozen, het patroon als de pallet en de gebroken stenen als lege ruimtes. De verschillen zitten vooral in de doelfunctie. Zo is er bij het plaatsen van dozen op een pallet natuurlijk geen enkele reden om gelijksoortige dozen te verspreiden,

integendeel. Daarnaast zal een pallet altijd rechthoekig zijn, terwijl bij patronen ook andere vormen mogelijk en zelfs nuttig zijn; dit zullen we verderop bespreken. Voorlopig gaan we uit van rechthoekige patronen.

Ons probleem kan als een Mixed Integer Linear Program (MILP) geformuleerd worden. We kunnen het MILP globaal als volgt beschrijven: we zoeken voor elke van de te plaatsen stenen een horizontale en verticale coördinaat, zó dat de stenen onderling niet overlappen. Twee stenen a en b overlappen niet als steen a zich links, rechts, onder, of boven steen b bevindt. Omdat we voor elke combinatie van twee stenen moeten voorkomen dat ze overlappen, neemt zowel het aantal (binaire) variabelen als het aantal voorwaarden kwadratisch toe met het aantal stenen. Hierdoor kan het dagen duren om grotere instanties op te lossen. Een manier om het aantal variabelen en voorwaarden te beperken is de 2-fasen aanpak, eerder bedacht door Wu & Ting (2007). In de eerste fase wordt voor elke steen een verticale coördinaat of laag gekozen. Hierbij wordt nog niet rekening gehouden met het voorkomen van overlap, maar wordt slechts gekeken of er per laag ruimte is voor alle stenen. Wel kan al rekening worden gehouden met het spreiden van stenen. Het voorkomen van overlap gebeurt in



Figuur 2a. Een oplossing van fase 1



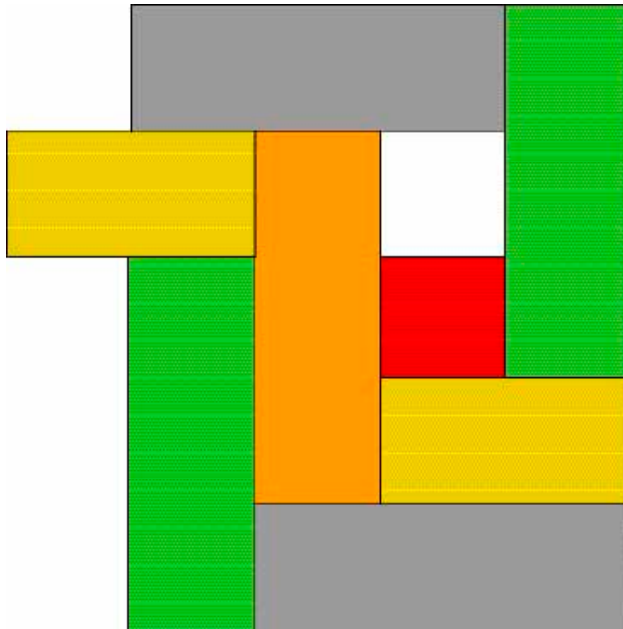
Figuur 2b. Een oplossing van fase 2

de tweede fase, waar op basis van de gekozen lagen voor de stenen, de horizontale coördinaten worden bepaald. Dit proces is afgebeeld in figuur 2.

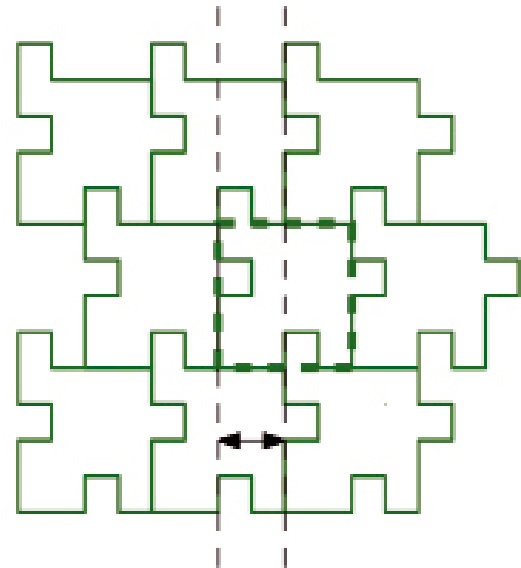
Het voordeel van de 2-fasen aanpak is dat stenen die geen laag gemeenschappelijk hebben, zeker niet overlappen. Daardoor hebben we aanzienlijk minder variabelen en voorwaarden nodig. Dit is een groot voordeel, maar hoe zit het met de nadelen? De kritische lezer zal zich hopelijk het volgende afvragen: 'Is het mogelijk het kiezen van de lagen in de 1e fase zo te doen, dat in de 2e fase de stenen onmogelijk kunnen worden geplaatst zonder te overlappen?' Figuur 3 bevat een voorbeeld waarbij dit het geval is. Ten eerste merken we op dat dit wel een oplossing is voor de 1e fase: in geen enkele laag is de som van de breedtes van de stenen breder dan het patroon. We kunnen zien dat de bovenste gele steen niet meer in het patroon past zonder met andere stenen te overlappen (probeer maar!). Het antwoord op de vraag of het in de 1e fase kan worden verprutst is dus 'Ja'. Aan de andere kant geven uitgebreide tests van Wu en Ting aan dat dit uitzonderlijk is: zij onderzochten een standaard verzameling van 7827 *pallet loading* instanties, en vonden bij slechts twee instanties dat de 2e fase geen oplossing had (*private communication*). Onder beperkende

voorwaarden kunnen we de oplossing in de 2e fase wel garanderen, bijvoorbeeld als de stenen maximaal twee lagen hoog zijn. Merk op dat in figuur 3 stenen van drie lagen hoog voorkomen.

Tot nu toe hebben we de situatie beschreven waarin het patroon rechthoekig van vorm is. Dit heeft als belangrijk nadeel dat het patroon eenvoudig te herkennen is. Daarom zullen we vormen toestaan waarbij de stenen buiten het patroon kunnen vallen. In figuur 4 is een mogelijke vorm te zien. Omdat het patroon zich herhaalt, zal een steen die het rechthoekige patroon aan de rechterzijde doorbreekt, aan de linkerkant een ander patroon weer binnen komen. Omdat alle patronen gelijk zijn, zal dus ook een steen het oorspronkelijke patroon binnenkomen. We kunnen het patroon dus eigenlijk zien als een cyclisch patroon, of patroon op een torus: stenen die het patroon aan één kant doorbreken, komen het patroon aan de andere kant weer binnen. Doordat we de patronen per laag horizontaal verplaatsen, zullen stenen die de bovenkant van het patroon doorbreken zelfs in een andere kolom het patroon weer binnenkomen. Merk op dat nu ook de horizontale voeg in het metselwerk onderbroken wordt. Ook in dit geval kunnen we een MILP opstellen, dat behoorlijk gecompliceerd wordt door de



Figuur 3. Geen oplossing in de 2e fase



Figuur 4. Een cyclisch patroon

boekhouding van stenen die wel of niet het patroon doorbreken. Ook moet precies geformuleerd worden, wat links en rechts, dan wel boven en onder is. De 2-fasen aanpak is hier ook mogelijk, en heeft hetzelfde manco: de oplossing van de 1e fase garandeert geen oplossing voor de 2e fase. In het cyclische geval is echter een constructie mogelijk als de stenen maximaal drie lagen hoog zijn. Zoals we in figuur 3 zagen, is dat bij niet-cyclische patronen niet altijd mogelijk. Helaas zal deze constructie meestal niet leiden tot mooie patronen.

Deze eerste studie naar ontwikkelen van patronen voor polymetrisch metselen heeft verschillende wiskundige vraagstukken blootgelegd. Voor een aantal van deze vraagstukken hebben wij een oplossing gevonden en verwerkt in een software tool. Zo hebben wij patronen kunnen genereren van tien lagen hoog, bestaande uit ongeveer 40 stenen. Zo'n patroon is zo groot dat het niet gemakkelijk te herkennen is in het muurvlak en toch zo klein dat de metselaar het kan onthouden. Jorissen-Simonetti architecten gaat graag de uitdaging aan om samen met wiskundigen tot een tool te komen, waarmee snel mogelijke patronen zijn te genereren voor het gewenste gevelbeeld.

LITERATUUR

- De Jong, J. (2011). *Het ontwerpen van patronen voor polymetrische metselwerken*. Master thesis Universiteit Twente (augustus 2011).
- Wu, K. C. & Ting, C. J. (2007). *A Two-phase algorithm for the manufacturer's pallet loading problem*. *Industrial Engineering and Engineering Management*, 2007 IEEE International Conference, pp. 1574–1578.
- Voorbeeld polymetrisch metselwerk: <<http://tinyurl.com/6wrdvcj>>.

JASPER DE JONG rondde zijn MSc-opleiding Applied Mathematics af met een onderzoek naar het genereren van patronen voor polymetrisch metselen. Hij is AiO bij de leerstoel Discrete Wiskunde en Mathematisch Programmeren aan de Universiteit Twente.

E-mail: <jasperov@hotmail.com>

HANA VINDUŠKA was als architecte werkzaam bij Jorissen-Simonetti architecten in Rijssen. Zij kwam met het idee voor dit onderzoek bij de Universiteit Twente.

E-mail: <youcreates@gmail.com>

ERWIN HANS is universitair hoofddocent en opleidingsdirecteur bij de faculteit Management en Bestuur van de Universiteit Twente.

E-mail: <E.W.Hans@utwente.nl>

GERHARD POST is parttime universitair docent bij de afdeling Toegepaste Wiskunde van de Universiteit Twente. Tevens is hij werkzaam als OR-specialist bij ORTEC bv.

E-mail: <G.F.Post@utwente.nl>

HET COMPLEXE NETWERK VAN ONZE HERSENEN

PIET VAN MIEGHEM

Een relatief nieuwe stroming is de 'netwerkwetenschap' (*network science*), die de functies in een complex netwerk met zijn onderliggende structuur bestudeert. De netwerkwetenschap is ontstaan uit vragen omtrent de structuur en het functioneren van het internet. Bijvoorbeeld: Wat is de invloed van de verdeling van de burens van een knoop (*router*) in het internet op de snelheid waarmee een bericht van zender naar ontvanger reist? Hoe ziet de internet-topologie er uit als het aantal knopen verdubbelt? Wat is de structuur van de in de tijd veranderende 'webgraaf', alle webpagina's die door hyperlinks worden verbonden? In dit artikel zal ik beschrijven hoe we deze theorie van complexe netwerken op onze hersenen kunnen toepassen.

De wetenschap zwoegt al decennia om kennis te verwerven over de werking van onze hersenen. We weten al veel, maar eigenlijk nog steeds te weinig. Onze boordcomputer is verrassend veelzijdig, maar tegelijkertijd verschrikkelijk complex. Elk nieuw inzicht kan leiden tot een dubbel voordeel: het bestrijden van mogelijke hersenafwijkingen en ziektebeelden (zoals tumoren, Alzheimer, epilepsie, schizofrenie) enerzijds en het ontwerpen van betere computers of intelligente systemen anderzijds. Bedenk maar hoe goed en snel wij iemand herkennen op een foto en hoe moeilijk dat nog steeds is voor een computer.



Hersenactiviteiten-netwerk

Op de foto van figuur 1 wordt een magneto-encephalogram (MEG) geregistreerd: rondom het hoofd van een patiënt zijn 151 sensoren geplaatst die de magnetische hersenactiviteit meten gedurende een bepaalde tijd. De foto en data komen van het VU medisch centrum te Amsterdam, uit de groep van collega Kees Stam, waarmee mijn groep aan de TUDelft samenwerkt.¹ Het middelste plaatje van figuur 1 toont een typische meting van de hersensignalen per sensor. Wanneer deze signalen paarsgewijs worden gecorreleerd, vinden we de correlatiematrix in de rechterfiguur. Ons verhaal begint bij deze correlatiematrix.

Elk netwerk kan beschreven worden door een topologie of *adjacency*-matrix A , waarin een element a_{ij} gelijk is aan 1 als knoop i met knoop j is verbonden; anders is het element 0. De hersencorrelatiematrix beschrijft dus een netwerk van 151 knopen, waarbij de waarde van de correlatie een soort sterkte aangeeft tussen twee knopen (sensoren rond het hoofd).

Als de correlatie beneden een drempelwaarde ligt, die een functie is van ondermeer de meetnauwkeurigheid,

heid en achtergrondruis, zetten we deze waarde op 0 en in andere gevallen wordt de waarde 1. Op deze manier hebben we de hersencorrelatiematrix herleid tot een adjacency-matrix, die ons toelaat om het overeenkomstig netwerk van hersenactiviteiten te tekenen. Verschillende opdrachten resulteren in verschillende netwerken. Bijvoorbeeld, wanneer een persoon slaapt meten de sensoren een andere hersenactiviteit dan wanneer een persoon een tekst leest of een puzzel oplost. Daarom spreekt men van *functional brain networks*.

Welke topologische eigenschappen heeft een dergelijk hersenactiviteiten-netwerk? Hoe verandert de structuur van het brein met de leeftijd? Kan men uit het netwerk afleiden of iemand slim is (of een hoog IQ heeft)? Dergelijke vragen proberen we via netwerkwetenschap te beantwoorden.

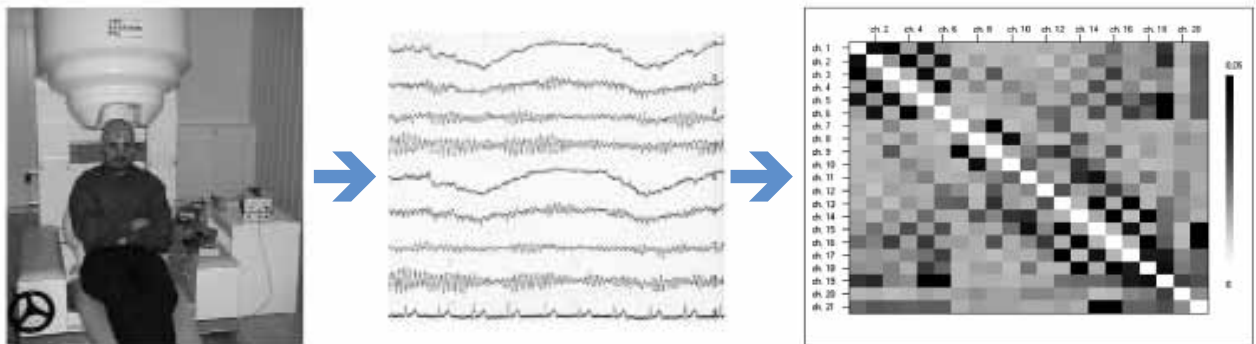
Clusteringscoëfficiënt

Ons functioneel breinnetwerk is te vergelijken met een 'Watts-Strogatz small-world network'. Duncan Watts, een promovendus van professor Steven Strogatz, heeft aan de Cornell University (USA) een vrij eenvoudig model voorgesteld dat verrassende dingen kon verklaren die de beroemde Erdős-Rényi graaf niet kon. Het wiskundig genie Paul Erdős en zijn vriend Alfréd Rényi hebben rond 1960 de tot nu toe meest bestudeerde stochastische graaf gelanceerd. De Erdős-Rényi graaf bestaat uit N knopen en de kans dat er een verbinding is tussen twee knopen is gelijk aan p en onafhankelijk van andere verbindingen. Vele eigenschappen zoals de

verdeling van het aantal burens (binomiaal met parameters $N-1$ en p), de clusteringscoëfficiënt (gelijk aan p), de gemiddelde padlengte etc. zijn analytisch te berekenen. Erdős en Rényi hebben bewezen dat de structuur van hun netwerk plots verandert met de kans p wanneer het aantal knopen N groot genoeg is. Als p kleiner is dan een kritische verbindingsdichtheid $p_c = O(\log N/N)$, dan is de graaf bijna zeker niet verbonden, terwijl voor $p > p_c$ hij met overweldigende kans wel verbonden is.

Net zoals een Erdős-Rényi graaf heeft een Watts-Strogatz graaf een korte gemiddelde padlengte tussen twee willekeurige knopen. Als informatie langs een kortste pad stroomt dan bestaat dit pad slechts uit een klein aantal verbindingen van de orde $O(\log N)$. Vele realistische netwerken bezitten deze 'efficiëntie' – eigenschap ook – het rooster is een tegenvoorbeeld. Een dergelijke korte gemiddelde padlengte is ook bekend als 'het 6 hops verwijderd zijn van de president van de USA'. De psycholoog Stanley Milgram heeft een beroemd brievenexperiment uitgevoerd, waarin een persoon die een brief ontving deze slechts kon doorgeven aan een vriend van wie hij vermoedde dat die vriend een andere vriend had, die dichterbij de ontvanger zou staan. Het brievenexperiment toonde ook de '6 degrees of separation' aan tussen een willekeurige zender- en ontvangerknoop.

Echter, in tegenstelling tot de Erdős-Rényi graaf, bezit de Watts-Strogatz graaf een hoge clusteringscoëfficiënt, die weergeeft hoeveel burens van een knoop onderling verbonden zijn. Functionele breinnetwerken bezitten een hoge clusteringscoëfficiënt die niet door een Erdős-Rényi graaf verklaard kan worden; echter wel door een Watts-Strogatz graaf.



Figuur 1

Assortativiteit

Maar helaas is de werkelijkheid weerbarstiger: het hersenactiviteitennetwerk bezit ook eigenschappen, zoals modulariteit en assortativiteit, die noch door de Erdős-Rényi noch door de Watts-Strogatz graaf verklaard kunnen worden. Assortativiteit is een reflectie van het gezegde 'soort zoekt soort' of 'de rijken worden steeds rijker', ofwel dat een knoop met veel burens met grotere kans verbonden is met burens die ook veel burens hebben; en omgekeerd dat knopen met weinig burens verbonden zijn met gelijkaardige knopen. In veel complexe netwerken zien we assortatief of disassortatief gedrag: een belangrijke knoop kan vele kleinere knopen verbinden (disassortatief), terwijl in een sociaal netwerk, mensen met gelijke interesses elkaar opzoeken (assortatief). Het merkwaardige van de hersenen is dat ze op macroscopisch functioneel niveau assortatief zijn, terwijl op neurale niveau – we bezitten zo'n 10^{11} neuronen (knoten) en 10^{14} verbindingen – het netwerk disassortatief is. Hoe komt dat? We weten het niet. De observatie betekent dat de topologische structuur van de onderliggende bouwstenen (neuronen) anders blijkt te zijn dan van hersen-subnetten (die uit miljarden neuronen bestaan), waarvan we de activiteit meten via een MEG.

De verbindingen in de hersenen zijn niet vast, maar kunnen veranderen in de tijd en met de functie die we uitvoeren. Als een baby wordt geboren, kunnen we het functionele breinnetwerk nog redelijk goed beschrijven met een Erdős-Rényi graaf. Er zijn ontzettend veel neuronen met veel verbindingen ertussen, die een baby het enorme leervermogen bieden dat volwassenen niet meer hebben. Het blijkt dat de veelheid aan verbindingen door het leerproces naar de volwassenheid toe continue verandert: verbindingen met zwakke neurale koppelingen sterven af en de verbindingen die bovengemiddeld sterk zijn, worden nog sterker. Hierdoor verandert het functionele breinnetwerk met de leeftijd van een Erdős-Rényi graaf naar een Watts-Strogatz graaf met assortatief gedrag en modulaire structuur. Een mooi recent artikel van Van

den Heuvel en Stam beschrijft het verband tussen IQ en gemiddelde padlengte. Het vermoeden dat hoogbegaafde personen een effectievere hersenwerking hebben (en dus sneller kunnen denken en werken), wordt bevestigd door metingen: het IQ is omgekeerd evenredig met de gemiddelde padlengte. De informatieverwerking in de hersenen vergt minder tussenliggende stappen of anders gezegd, de doorstroming van informatie is sneller. Het groeiproces van baby naar volwassenheid en de externe factoren die de verbindingen tussen neuronen al dan niet versterken blijkt nu nog onvoldoende begrepen te zijn. Andere metingen tonen aan dat zevenjarigen al een andere hersenstructuur hebben dan vijfjarigen: het overeenkomstige functionele netwerk is minder 'willekeurig' (dus minder een Erdős-Rényi graaf).

Tergend traag vooruit

Hebben we dan geen beter model dat de evolutie van het functionele breinnetwerk beschrijft? Nog niet. Hieraan wordt momenteel gesleuteld: Kees Stam, hoogleraar klinische neurofysiologie aan het VU Medisch Centrum in Amsterdam heeft interessante ideeën voor toekomstig onderzoek. Echter, de complexiteit van de wiskunde zal van een andere orde zijn dan de relatief eenvoudige Erdős-Rényi random en de Watts-Strogatz kleine-wereld graaf. De kunst ligt in het vinden van een hersenmodel dat we nog analytisch kunnen behappen, en dat een stap dichterbij de werkelijkheid staat. Maar dit laatste geldt voor elk fundament der wetenschap: het gaat meestal tergend traag vooruit en als het eenmaal verworven kennis is, staan we versteld waarom dat toch zolang had moeten duren. Wetenschap is als een breinnetwerk.

Met dank aan C. J. Stam en J. S. H. van Leeuwen.
Zie ook: Connected brains <<http://home.kpn.nl/stam7883>>

PIET VAN MIEGHEM is hoogleraar telecommunicatienetwerken aan de TU Delft.
E-mail: <P.F.A.VanMieghem@tudelft.nl>



OVER REGENKANSSEN EN PARAPLU'S

MAURICE SCHMEITS

Van vrijwel alle weersverwachtingen is de 24-uurskans op neerslag in de komende 5 of 6 dagen een vast onderdeel. Maar wat betekent het nu eigenlijk als voor de volgende dag de kans op regen 60% is? Dit artikel gaat in op die vraag, maar ook op de manier waarop die kansverwachting tot stand komt, hoe je die kansverwachting kunt gebruiken om te beslissen of je wel of geen paraplu meeneemt en de wijze waarop kansverwachtingen geïnterpreteerd worden.

Het KNMI definieert de kans op neerslag als de kans dat iemand die zich op een willekeurige vaste plek in Nederland bevindt, op een dag (00-24 uur) neerslag

krijgt. Die neerslag is meestal regen, maar kan ook sneeuw of hagel zijn. Het getal zegt niets over de duur van de regen: dat kan 10 minuten zijn of 24 uur. Wel is vereist dat er tenminste 0,3 mm neerslag valt, wat overeenkomt met 0,3 liter per vierkante meter. Bedenk dat een neerslagkans van 50% zowel kan duiden op een zekere verwachting dat het in het noorden van Nederland droog is en in het zuiden regent, als op een 50% kans dat het in heel Nederland regent.

Hoe wordt de neerslagkans bepaald?

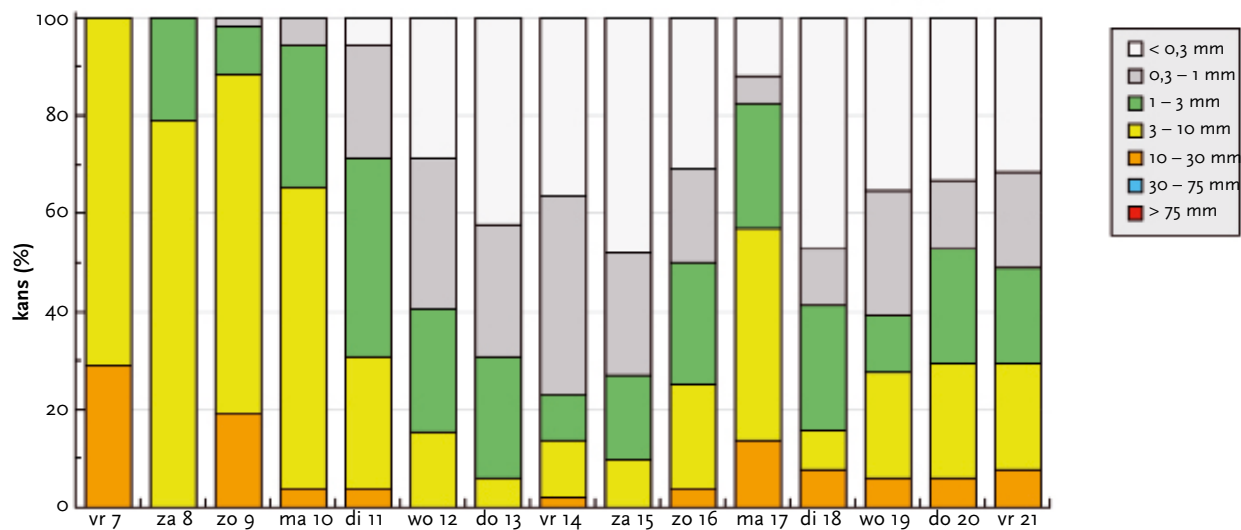
Er zijn tegenwoordig verschillende manieren om de neerslagkans te bepalen. Voor de middellange termijn

maakt het KNMI gebruik van de modelberekeningen van het European Centre for Medium-range Weather Forecasts (ECMWF). Dit centrum is een samenwerkingsverband tussen een groot aantal Europese landen en heeft het beste mondiale weermodel ontwikkeld. Dit model gebruikt een groot aantal waarnemingen van over de hele wereld om de beginconditie zo goed mogelijk te bepalen. Maar ook al zou je de begintoestand van de atmosfeer voor elke m³ op aarde weten, dan nog zal de verwachting na typisch enkele dagen gaan afwijken van het werkelijke weer. Er zit een inherente onzekerheid in de weersverwachting ten gevolge van wat het Vlindereffect (zie <http://nl.wikipedia.org/wiki/Vlindereffect>) is gaan heten. Door het chaotische karakter van de atmosfeer kunnen kleine verstoringen – veroorzaakt door een vlinder als metafoor – uitgroeien tot grotere verstoringen die een zogenoemde deterministische weersverwachting – dat wil zeggen zonder onzekerheidsinformatie – na een bepaalde tijd nutteloos maakt.

Hoe kunnen we die onzekerheid nu kwantificeren? De manier waarop het ECMWF dat doet, en ook een

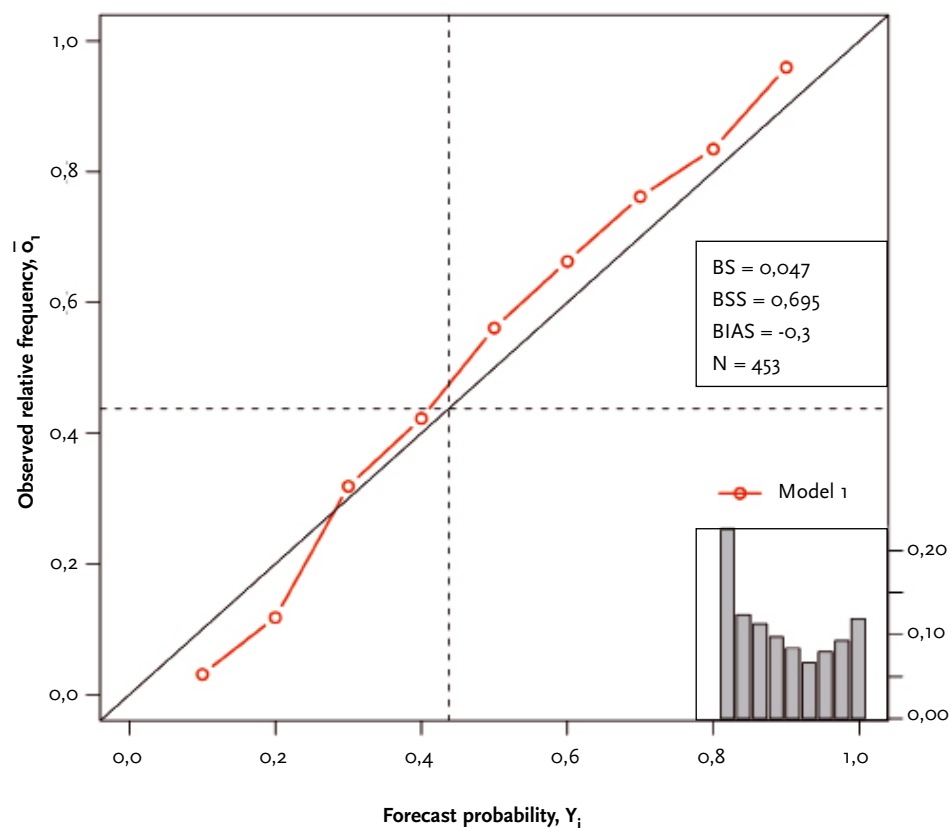
aantal andere centra over de hele wereld, is door niet alleen één modelrun te doen vanuit de ‘beste’ beginconditie, maar daarnaast een groot aantal modelruns (50), elk met een kleine realistische verstoring van de beginconditie. Dit wordt een ensemblepredictiesysteem (EPS) genoemd. Door vervolgens het aantal ensembleleden te tellen dat bijvoorbeeld neerslag geeft en dit te delen door het totaal aantal ensembleleden, kan een schatting van de kans op neerslag verkregen worden. Figuur 1 geeft een voorbeeld van de uitvoer van het EPS van het ECMWF voor 24-h neerslagsommen in De Bilt.

Niettemin kon vier decennia geleden, toen de computers nog niet zo krachtig waren en er met moeite één modelrun gedaan kon worden, de kans op neerslag ook al worden geschat. Daarvoor werd een andere techniek gebruikt die Model Output Statistics (MOS; Glahn and Lowry, 1972) wordt genoemd. Om MOS toe te kunnen passen heb je een archief nodig van uitvoer van modelruns en van neerslagwaarnemingen in dit geval. Door statistische verbanden te zoeken tussen modelvariabelen en neerslagwaarnemingen, bijvoor-



Figuur 1. Voorbeeld van kansverdelingen voor 24-h neerslagsommen in de Bilt op basis van het EPS van het ECMWF; de verwachtingstermijn is 1-14 dagen (run: vrijdag 7 oktober 2011 00 UTC)

Figuren 2a en 2b. Reliability diagram voor de 24-h neerslagkansen van het KNMI voor (a) 1 dag en (b) 5 dagen vooruit. In het diagram staat de waargenomen frequentie als functie van de verwachte kans (rode lijn). Het histogram geeft de verdeling van de neerslagkansen (tussen 0,1 en 0,9) weer, de diagonaal perfecte reliability en de gearceerde lijnen de sample klimatologie. Het verificatiegebied is Nederland en de verificatieperiode is juni 2010 t/m augustus 2011. BS is een afkorting voor de Brier score (1) en BSS voor de Brier skill score (3) met de sample klimatologie als referentie.



Figuur 2a. Reliability diagram voor de 24-h neerslagkansen van het KNMI voor 1 dag vooruit

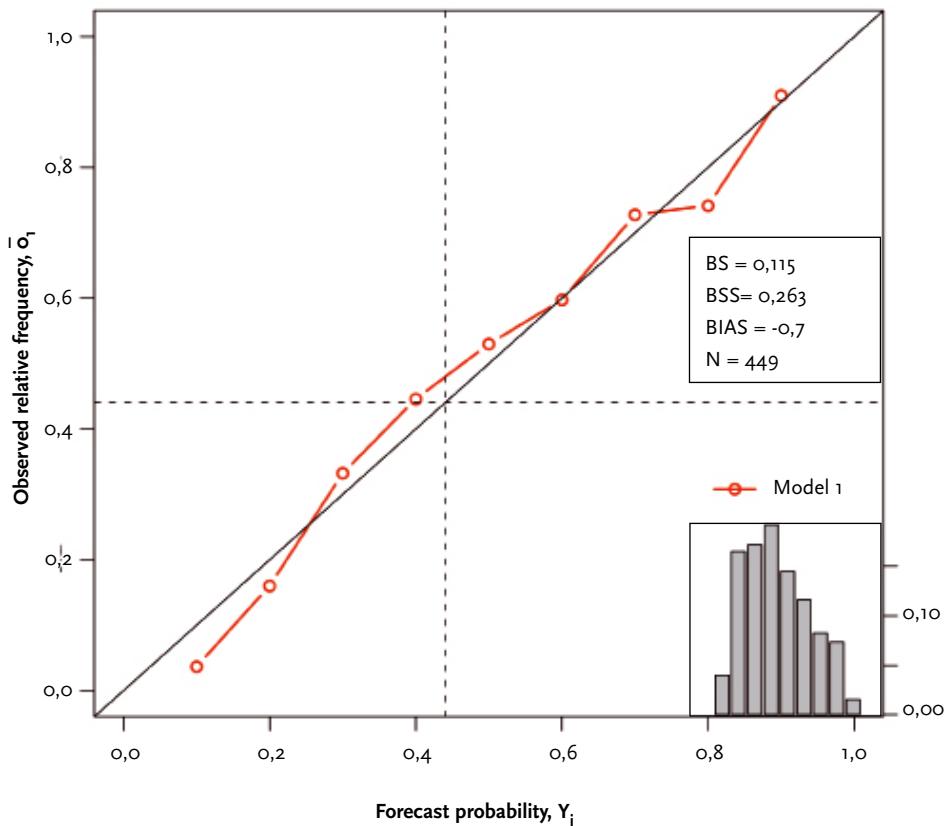
beeld met behulp van logistische regressie (Wilks, 2006), kan dan een kansverwachtingsmodel afgeleid worden. Dit model wordt vervolgens op actuele modeluitvoer toegepast, teneinde kansverwachtingen te maken. Ook tegenwoordig wordt MOS nog vaak gebruikt, onder andere om de kans te bepalen op het optreden van gebeurtenissen die niet expliciet door het weermodel worden weergegeven zoals onweer, maar die wel samenhangen met verschijnselen die het model kan representeren zoals zware neerslag. De neerslagkansen die het KNMI uitgeeft, zijn ook gebaseerd op een MOS-systeem. Hierbij wordt ondermeer informatie van het ECMWF EPS gebruikt.

Hoe worden kansverwachtingen geverifieerd?

Om te bepalen hoe goed een kansverwachtingssysteem is, dien je het te verifiëren. Kansverwachtingen

kunnen echter pas geverifieerd worden als je een groot aantal verwachtingen uitgegeven hebt. Indien je verwacht dat de kans op het optreden van een gebeurtenis $x\%$ is, waarbij $0 \leq x \leq 100$, en de waargenomen frequentie van die gebeurtenis gelijk is aan $x\%$, dan is je kansverwachting *reliable*. Om te bepalen of een kansverwachtingssysteem *reliable* is, wordt vaak gebruik gemaakt van een *reliability diagram* (zie figuur 2). Hierin is de waargenomen frequentie afgezet tegen de verwachte kans. Als deze twee gelijk zijn, met andere woorden als alle punten in het diagram op de diagonaal liggen, dan zijn je kansverwachtingen dus *reliable*. Uit figuur 2 blijkt dat zowel de 24-h neerslagkansen van het KNMI voor 1 dag vooruit (figuur 2a) als voor 5 dagen vooruit (figuur 2b) *reliable* zijn.

Een klimatologische kans is ook *reliable*, mits de klimatologie niet verandert. Als je echter steeds een klimatologische kans uitgeeft, discrimineer je niet tus-



Figuur 2b. Reliability diagram voor de 24-h neerslagkansen van het KNMI voor 5 dagen vooruit

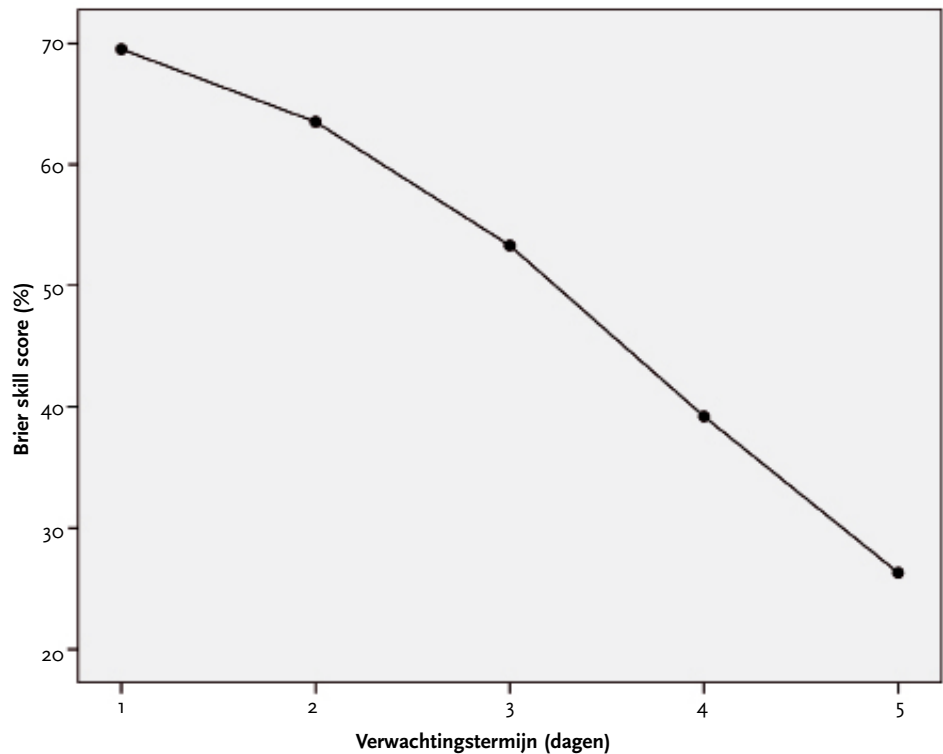
sen gevallen waarbij een grote kans op een gebeurtenis is en gevallen waarbij die kans klein is. We zeggen dan dat het kansverwachtingssysteem geen resolutie heeft. Als een systeem zoveel mogelijk richting 0% verwacht als het fenomeen niet optreedt en richting 100% als het wel optreedt, dan heeft het systeem een hoge resolutie. De 24-h neerslagkansen van het KNMI voor 1 dag vooruit (figuur 2a) hebben een hogere resolutie dan voor 5 dagen vooruit (figuur 2b). In de figuren 2a en 2b is dit te zien aan de verschillende vormen van de histogrammen, die de verdeling van de neerslagkansen weergeven, in combinatie met het reliable zijn van de kansen. Zowel reliability als resolutie zijn dus belangrijk voor een *skilful* kansverwachtingssysteem. Een kansverwachtingssysteem dat niet reliable is, kan gecalibreerd worden. Veel ensemblesystemen laten een te kleine spreiding zien, met name voor de korte verwachtingstijden en kunnen gecalibreerd wor-

den met behulp van statistische nabewerking van de modeluitvoer.

Behalve het reliability diagram wordt ook een groot aantal verificatiescores gebruikt. Een veel gebruikte score in de meteorologie is de Brier Score (BS). De BS is gedefinieerd als de *mean-square error* van kansverwachtingen:

$$BS = \frac{1}{N} \sum_{j=1}^N (p_j^f - p_j^o)^2, \quad (1)$$

waarbij N het totaal aantal gevallen is, p_j^f de verwachte kans en p_j^o het wel/niet optreden van een gebeurtenis; voor een gebeurtenis die niet of wel optreedt kan p^o dus uitsluitend de waarde 0 of 1 hebben. Volgens de definitie van de neerslagkans die aan het begin van het artikel is gegeven, kan p^o echter elke waarde tussen 0 en 1 hebben en wordt dan bepaald door het percentage van de Nederlandse neerslagstations die tenminste



Figuur 3. Brier skill score (met de sample klimatologie als referentie) voor de 24-h neerslagkansen van het KNMI als functie van de verwachtingstermijn voor 1 tot 5 dagen vooruit; het verificatiegebied is Nederland en de verificatieperiode is juni 2010 t/m augustus 2011

0,3 mm neerslag hebben gemeten in een etmaal. De BS is gelijk aan 0 voor een perfect systeem en 1 voor een compleet waardeloos systeem.

Als de kansen in l klassen ingedeeld worden, komen de hierboven genoemde termen reliability en resolutie ook voor in de volgende decompositie van de Brier score (Wilks, 2006):

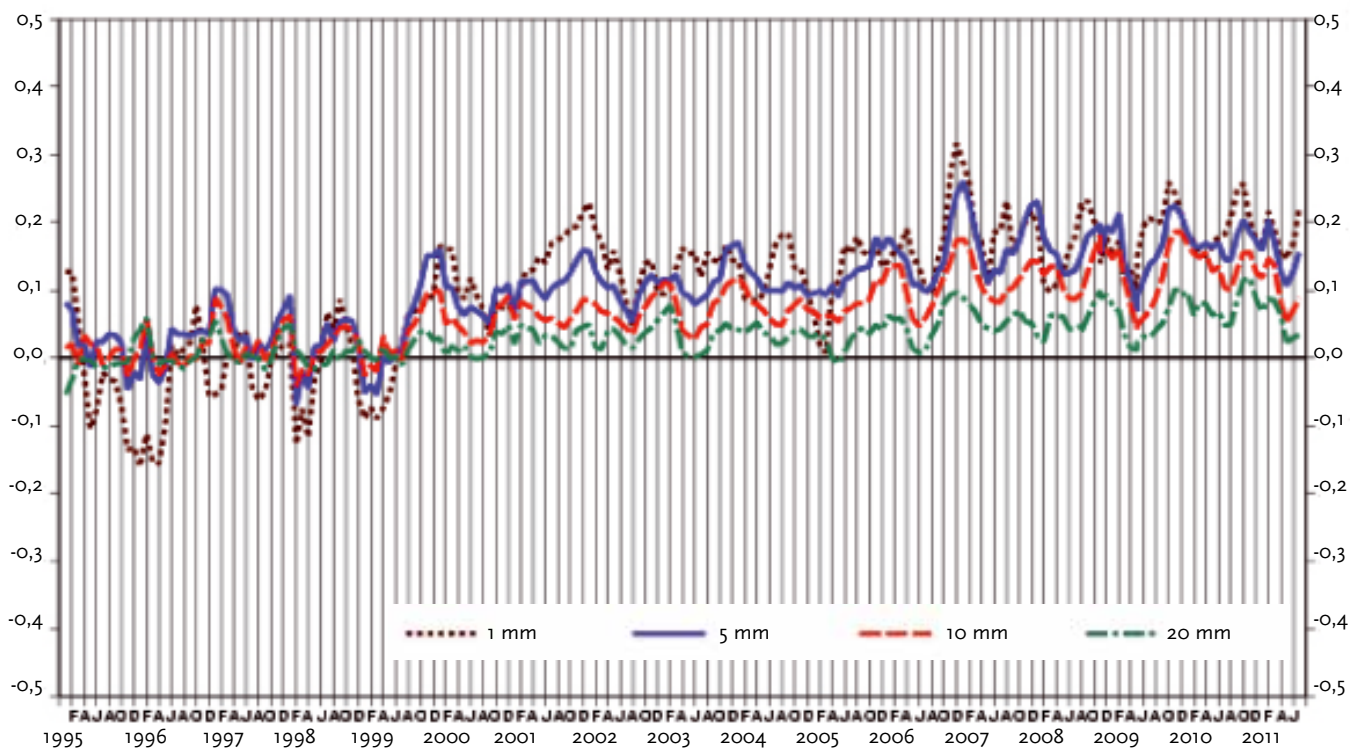
$$BS = \frac{1}{N} \sum_{i=1}^l n_i (p_i^f - \overline{p_i^o})^2 - \frac{1}{N} \sum_{i=1}^l n_i (\overline{p_i^o} - \overline{p^o})^2 + \overline{p^o} (1 - \overline{p^o}), \quad (2)$$

waarbij $\overline{p_i^o}$ de waargenomen frequentie is over alle N gevallen, p_i^o de waargenomen frequentie per klasse en p_i^f de verwachte kans per klasse. De eerste term in (2) representeert de reliability, de tweede de resolutie en de derde de *uncertainty*. Deze laatste term hangt uitsluitend van de sample klimatologie af en wordt dus niet beïnvloed door de verwachtingen.

Om een idee te krijgen hoe goed een systeem is, wordt het vaak vergeleken met een referentie. Deze referentie kan bijvoorbeeld klimatologie of persistentie zijn. Dit leidt dan tot een zogenoemde *skill score* en voor de Brier score leidt dit tot de Brier skill score (BSS):

$$BSS = \frac{BS - BS_{ref}}{BS_{perf} - BS_{ref}} = 1 - \frac{BS}{BS_{ref}}, \quad (3)$$

omdat de Brier score van perfecte verwachtingen (BS_{perf}) gelijk aan 0 is. In deze vergelijking is BS_{ref} de Brier score van het referentiesysteem. De BSS is gelijk aan 0 als je systeem niet beter is dan het referentiesysteem en 1 als je een perfect systeem hebt. Als de BSS positief is, is je systeem beter dan het referentiesysteem en vice versa als de BSS negatief is.



Figuur 4. Evolutie van de Brier skill scores voor de ECMWF EPS kanssen op 24-h neerslagsommen > 1, >5, >10 en >20 mm voor 6 dagen vooruit. Het verificatiegebied is Europa en de verificatieperiode loopt van 1995 tot heden (bron: ECMWF)

Hoe groot is de skill van neerslagverwachtingen?

Figuur 3 geeft een beeld van de skill van de KNMI MOS neerslagkansverwachtingen in de vorm van Brier skill scores voor 1 tot 5 dagen vooruit. De skill neemt af met de verwachtingstermijn, hetgeen logisch is omdat de onzekerheid toeneemt met die termijn.

Men kan zich afvragen hoe het zit met de skill van overschrijdingskanssen voor diverse neerslagdrempels en met de skill voor nog langere verwachtingstijden. En of de verwachtingen door de jaren heen beter zijn geworden.

In figuur 4 is de evolutie van de Brier skill score te zien van de ECMWF EPS neerslagkanssen voor een viertal drempels en voor 6 dagen vooruit. Allereerst valt op dat de skill van de ECMWF EPS neerslagkanssen duidelijk is toegenomen: de skill vanaf 2000 is duidelijk beter dan ervoor. Dit hangt samen met een aantal

verbeteringen in het model, in de waarnemingen en in de constructie van het ensemble. Daarnaast is de skill voor de lagere drempels beter dan voor de hogere drempels. Dit is ook wat je verwacht: een weermodel heeft namelijk moeite met het verwachten van extremen. Soortgelijke figuren voor 4 en 10 dagen vooruit (niet getoond) laten zien dat de skill voor 4 dagen vooruit duidelijk beter is dan voor 6 dagen vooruit. Voor 10 dagen vooruit hebben de neerslagverwachtingen geen skill meer.

Hoe neem je een beslissing op basis van een kansverwachting?

Om op een objectieve wijze een beslissing te nemen op basis van een kansverwachting, wordt vaak gebruik gemaakt van een zogenaamde *cost-loss* analyse. Hierbij

dien je inzicht te hebben in de kosten die je maakt als je voorzorgsmaatregelen neemt om schade te voorkomen en in de potentiële schade die je oploopt als je geen voorzorgsmaatregelen neemt. Het quotiënt van deze twee, de cost-loss ratio genaamd, bepaalt vervolgens de kansdrempel, waarboven je actie zou moeten ondernemen.

Als voorbeeld kun je denken aan een fruitteler die de bloesem wil beschermen tegen nachtvorst. Stel dat zijn cost-loss ratio 0,1 is en de kans op nachtvorst 20%, dan zal hij de bloesem willen beschermen tegen schade door nachtvorst, terwijl hij dat niet zou doen als de kans op nachtvorst 5% is. Hierbij gaan we er wel van uit dat de kansverwachtingen reliable zijn.

Als daarentegen de cost-loss ratio van een gebruiker heel hoog is, bijvoorbeeld 0,9, dan zal hij of zij pas actie ondernemen als de kans > 90% is. Door kansverwachtingen uit te geven, kan elke gebruiker zelf bepalen, op basis van de cost-loss ratio, wanneer er actie moet worden ondernomen. Dit is een groot voordeel ten opzichte van deterministische verwachtingen, waarbij die onzekerheidsinformatie niet voorhanden is.

Kwalitatief speelt cost-loss ook een rol bij de beslissing of je een paraplu meeneemt. De ene persoon zal een paraplu meenemen als de kans op neerslag >50% is, terwijl de andere dat pas zal doen als de kans >80% is.

Met dank aan Gerrit Burgers, Martin Stam en Kees Kok van het KNMI. Meer informatie over het KNMI is te vinden op <www.knmi.nl> en over het ECMWF op <www.ecmwf.int>.

LITERATUUR

- Glahn, H. R., & Lowry, D. A. (1972). The use of model output statistics (MOS) in objective weather forecasting. *Journal of Applied Meteorology*, (11)8, 1203–1211.
- Wilks, D.S. (2006). *Statistical methods in the atmospheric sciences*, 2nd ed. International Geophysics Series, Vol. 59, Academic Press.

MAURICE SCHMEITS studeerde meteorologie en fysische oceanografie aan de Universiteit Utrecht en promoveerde op een onderwerp in de fysische oceanografie aan diezelfde universiteit. Sinds 2001 werkt hij als onderzoeker op het KNMI. E-mail: <schmeits@knmi.nl>

YOUNG STATISTICIANS IN DE VVS+OR

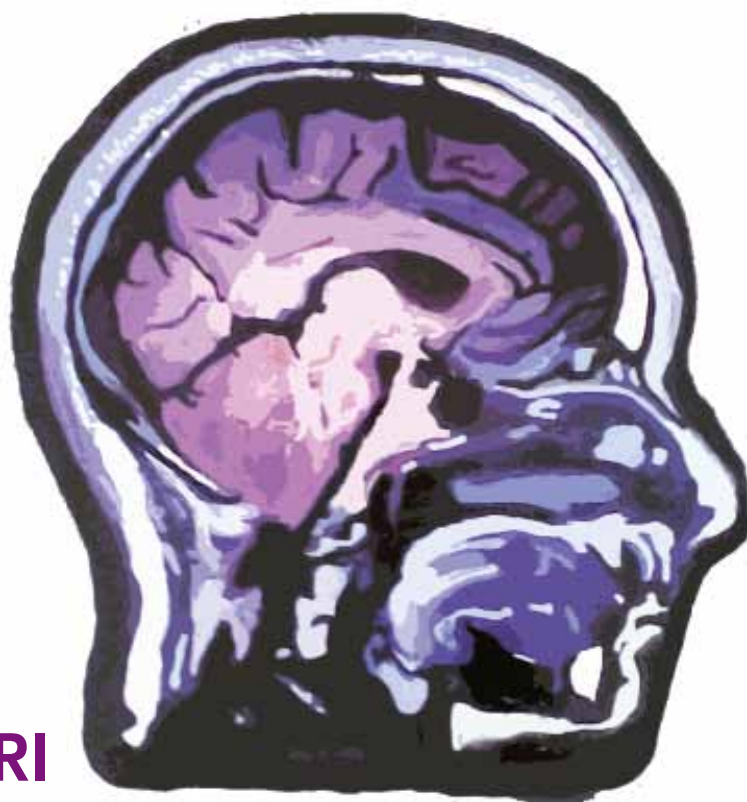
Op initiatief van student-bestuurslid Maarten Kampert is er sinds de Dag voor Statistiek en OR van 2011 een groep jonge statistici actief. We zijn nu een jaar verder en het blijkt dat er veel belangstelling is voor het idee om jonge mensen die zich aangetrokken voelen tot dit brede vak op regelmatige tijden met elkaar in contact te brengen.

De deelnemers zijn velerlei: Master en PhD studenten, maar ook Post Docs en anderen. Op de druk bezochte eerste bijeenkomst werden veel ideeën geopperd, die op een volgende avond meer concreet werden uitgewerkt. Het resultaat was onder andere een Pub Meeting, een Science Cafe en een Workshop over Chemometrics. Zelfs sponsoring bleek al mogelijk in dit eerste jaar!

Over het Science Cafe van november vorig jaar vertelt Corine Baayen, een van de organisatoren: ‘We wilden een informele avond met veel jonge statistici. Harry van Zanten en Herbert Hoijsink waren gevraagd respectievelijk het Frequentistische en het Bayesiaanse standpunt te verdedigen. Daarna was er een korte discussie onder leiding van Richard Gill. Gedurende de avond veranderde ik van “ik heb geen voorkeur” via een Bayesiaan in een Frequentist om weer te eindigen als een “ik heb geen voorkeur”. Het bijzondere van de avond was het zeer levendige debat met een goed glas bier in de hand.’ Het volledige verhaal van Corine staat op de VVS+OR website.

De Young Statisticians, zoals ze zich noemen, gaan dan ook zeker door met hun activiteiten. Op de VVS+OR website zijn de bijeenkomsten vermeld. Iedereen die jong is, of zich jong voelt, is van harte welkom. Voor meer informatie kun je een mailtje sturen naar <young@vvs-or>.

OPTIMALE EXPERIMENTELE STUDIES VOOR FUNCTIONELE MRI



BÄRBEL MAUS & GERARD J.P. VAN BREUKELLEN

Functionele MRI (*magnetic resonance imaging*, beeldvorming met magnetische resonantie) wordt steeds meer toegepast in de cognitieve neurowetenschappen en de psychiatrie. Door het effect van gepresenteerde stimuli, bijvoorbeeld afbeeldingen of geluiden, op de hersenenactiviteit van de proefpersoon te bestuderen, worden functionele hersengebieden bepaald. Een optimale planning van fMRI-studies helpt onderzoekers om een precieze schatting van het stimuluseffect te verkrijgen tegen minimale kosten. Het proefschrift van Maus (2011) bestudeert wat de optimale volgorde van de gepresenteerde stimuli is, en hoeveel proefpersonen en hoeveel metingen per proefpersoon optimaal zijn voor efficiënte effectschattingen binnen vaste experimentele kosten en voor een gegeven stimulusvolgorde. Hier geven we een overzicht van de belangrijkste resultaten en aanbevelingen.

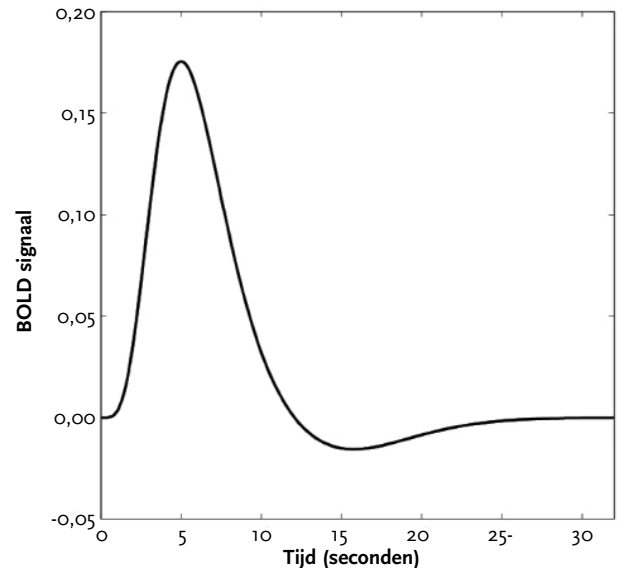
Functionele MRI (fMRI) is een beeldgevende methode voor de studie van hersenactiviteit. Functionele hersengebieden kunnen worden geïdentificeerd door een fMRI-experiment waarbij de proefpersonen geconfronteerd worden met stimuli, bijvoorbeeld foto's van gezichten of huizen. De proefpersoon bevindt zich in een MRI-scanner die een sterk magnetisch veld creëert. Vanwege de verschillen in magnetische eigenschappen van zuurstofrijk en zuurstofarm bloed en de veranderingen in de verhouding van zuurstofrijk tot zuurstofarm bloed in de geactiveerde hersengebieden kan het *blood-oxygen-level-dependent* (BOLD) fMRI-sigitaal worden gemeten. Tijdens het scannen van een proefpersoon wordt het fMRI-sigitaal gemeten met een vaste bemonsteringsfrequentie, bijvoorbeeld elke twee seconden. Het BOLD-sigitaal over de tijd na een korte presentatie van een stimulus wordt beschreven door de hemodynamische responsfunctie (HRF) (zie figuur 1). Kenmerkend is dat de HRF stijgt na stimuluspresentatie, haar maximum bereikt na ongeveer vijf

seconden, naar beneden valt en terug naar nul gaat na circa 30 seconden.

Experimenteel fMRI-onderzoek is zeer prijzig. Een uur scannertijd kan tussen de 200 tot 800 euro kosten (Culham, 2006). Typische fMRI-studies omvatten 10 tot 20 personen en elk proefpersoon heeft een scannertijd van vijf minuten tot een uur. Dit leidt tot een tijdreeks van honderden herhaalde metingen per hersengebied per proefpersoon. De signaal tot ruis verhouding in zo'n tijdreeks is klein, wat betekent dat het BOLD-signaal met ruis, die los staat van de functionele activiteit, wordt verward. Het is daarom belangrijk om via optimalisatie van het fMRI-experiment de signaal-ruis verhouding te verhogen, dat wil zeggen om efficiënte effectschattingen te verkrijgen tegen minimale kosten. Voor de optimalisatie moeten we eerst beslissen of het doel van het fMRI-experiment de detectie van actieve hersengebieden (detectie van activering) is of de schatting van de HRF (schatting), omdat deze twee experimentele doeleinden niet door hetzelfde type experiment worden geoptimaliseerd (Liu & Frank, 2004).

Types van fMRI experimenten

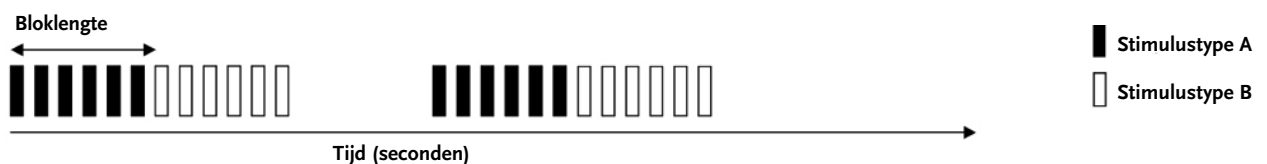
Er zijn in hoofdzaak twee types van fMRI-experimenten. Een daarvan is een *blocked design* waar presentaties (*trials*) van een stimulustype, bijvoorbeeld foto's van gezichten of huizen, in blokken worden aangeboden met een kort interstimulus interval (ISI) tussen opeenvolgende presentaties (zie figuur 2). Blokken kunnen taakblokken zijn die bestaan uit presentaties van een bepaald stimulustype, of rustblokken waarin de proefpersonen rusten of een eenvoudige controletaak



Figuur 1. Typisch tijdsverloop van de hemodynamische responsfunctie na stimuluspresentatie op tijdstip nul

uitvoeren zoals het bekijken van een fixatiekruis. Drie factoren kunnen voor een blocked design met vaste scannertijd en Q stimulustypes worden gevarieerd: ISI, bloklengthe en blokvolgorde, bijvoorbeeld $A_1A_2 \dots A_Q R$, of $A_1RA_2R \dots A_Q R$, of $A_1A_2 \dots A_Q$. A_1 tot en met A_Q staan hier voor de blokken van de verschillende stimulustypes A_1 tot en met A_Q en R staat voor rustblok. De blokvolgorde wordt een aantal malen herhaald tijdens het experiment. De bloklengthe wordt gegeven door de tijd in seconden tussen de start van een blok en het begin van het volgende blok (zie figuur 2).

De andere belangrijke vorm van experiment is een *event-related design* (zie figuur 3). In tegenstelling tot blocked designs hebben event-related designs een willekeurige volgorde van de verschillende stimulustypes en vaak ook een willekeurig ISI. Het is bekend dat



Figuur 2. Blocked design met blokvolgorde ABR



Figuur 3. Event-related design met een willekeurig ISI en willekeurige volgorde van de stimulustypes

blocked designs optimaal zijn voor de detectie van activering terwijl event-related designs optimaal zijn voor HRF schatting (Liu & Frank, 2004). Een nadeel van blocked designs is hun voorspelbaarheid, wat kan leiden tot verveling, gewenning of anticipatie van de proefpersonen. Daarom worden event-related designs ook toegepast voor de detectie van activering.

Optimalisatie voorbeeld 1 (blocked design)

We beschouwen een blocked design met twee stimulustypes, foto's van gezichten of huizen. Het experimentele doel is de detectie van het hersengebied verantwoordelijk voor de waarneming van gezichten. Door in elk kandidaat-hersengebied het contrast in fMRI-sigitaal tussen de presentatie van huizen en gezichten te bestuderen, kan het geactiveerde hersengebied worden geïdentificeerd. We willen nu weten: wat is de optimale blok lengte om dit contrast efficiënt te schatten? In welke volgorde moeten we de blokken presenteren? Welke ISI is optimaal?

In Maus (2011) gebruiken we optimale designmethoden om deze vragen te beantwoorden en om het optimale blocked design voor een proefpersoon binnen een bepaalde totale scannertijd te vinden. Meestal verrichten onderzoekers hun experimenten echter met verschillende proefpersonen, om de resultaten te generaliseren naar een populatie van proefpersonen of om groepen personen te vergelijken. Meer proefpersonen en langere scannertijd per proefpersoon zorgen voor meer gegevens en informatie, maar leiden ook tot hogere experimentele kosten. Gegeven een beperkt studiebudget worden we geconfronteerd met

de vraag: Is het efficiënter om minder proefpersonen langer te scannen, of om meer proefpersonen korter te scannen?

Richtlijnen voor blocked designs en detectie van activering

Maus (2011) concludeert dat de optimale blok lengte voor detectie tussen de 10 tot 20 seconden is. Verder is voor de detectie van paarsgewijze contrasten tussen Q verschillende stimulustypes de optimale blokvolgorde $A_1 \dots A_Q$, waar A_1 tot A_Q de verschillende bloksoorten aanduiden. Hiervoor zijn geen rustblokken nodig. Voor de opsporing van de individuele effecten van de verschillende stimulustypes, is de optimale blokvolgorde $A_1 \dots A_Q R$. Een rustblok is nodig per cyclus $A_1 \dots A_Q$ van de taakblokken. Het optimale ISI binnen een blok moet zo klein zijn als praktische overwegingen, zoals de stimulusduur, toelaten.

Het optimale aantal proefpersonen en de optimale scannertijd per proefpersoon zijn afhankelijk van verschillende factoren:

- de kosten per proefpersoon, de kosten voor de scannertijd en de totale kostenbegroting
 - de sterkte van de relatie tussen opeenvolgende metingen van een proefpersoon (dat wil zeggen autocorrelatie)
 - de variantie tussen de proefpersonen (tussen-proefpersoon variantie) en tussen de metingen binnen een proefpersoon (binnen-proefpersoon variantie), namelijk de verhouding van tussen-proefpersoon variantie tot binnen-proefpersoon variantie
- Wanneer de kosten per proefpersoon hoog zijn in ver-

houding tot de kosten per scannertijd, is het efficiënter om proefpersonen met een langere scannertijd te meten. Voor de relatie tussen de metingen van een proefpersoon hebben wij rekening gehouden met twee mogelijkheden:

1. de metingen binnen een proefpersoon zijn ongecorrigeerd en, meer realistisch,
2. de metingen binnen een proefpersoon zijn positief gecorrigeerd. Het optimale aantal proefpersonen neemt af, terwijl de optimale scannertijd toeneemt, met de correlatiesterkte tussen de metingen.

Ten slotte, als de verhouding van tussen-proefpersoon tot binnen-proefpersoon variantie toeneemt, neemt het optimale aantal proefpersonen toe terwijl de optimale scannertijd afneemt. In Maus (2011) presenteren we wiskundige formules voor het optimale aantal proefpersonen en de optimale scannertijd voor ongecorrigeerde metingen, en een numerieke methode om het optimale design voor gecorrigeerde metingen te vinden.

Optimalisatie voorbeeld 2 (event-related design)

We beschouwen nu een event-related design, bijvoorbeeld met foto's van gezichten en huizen als stimulus-types zoals in voorbeeld 1. In Maus (2011) bepalen we de optimale stimulusfrequentie p voor event-related designs met Q stimulus-types voor detectie en schatting indien elk stimulus-type even vaak gepresenteerd wordt. De stimulusfrequentie geeft dan het aantal keren dat een stimulus-type wordt gepresenteerd tijdens het experiment, dat wil zeggen

$$p = N_A / N_T = N_A / (Q \cdot N_A + N_R),$$

waarbij N_A het aantal metingen is waarbij een bepaalde stimulus A gepresenteerd is, N_T het totale aantal metingen per proefpersoon, en N_R het aantal metingen zonder stimulus (rust trial). We gebruiken net als in voorbeeld 1 onder andere een bekend optimaliteitscriterium, namelijk het D- of determinantcriterium, dat de efficiëntie van een design meet (Atkinson et al., 2007). In Maus (2011) wordt getoond dat de D-optimale stimulusfrequentie p_D door

$$p_D = 1 / (Q + 1)$$

is gegeven, wat betekent dat het aantal metingen N_R met een rust trial ook gelijk aan N_A is. Deze optimale stimulusfrequentie wordt verkregen voor ongecorrigeerde metingen. Uit berekeningen bleek echter dat de optimale stimulusfrequentie voor gecorrigeerde metingen bijna gelijk is aan de stimulusfrequentie voor ongecorrigeerde metingen.

Conclusie

Het proefschrift van Maus (2011) bevat meer resultaten en aanbevelingen dan hierboven weergegeven zijn. Met de verdere ontwikkeling van fMRI-onderzoeksvragen en -methoden, bijvoorbeeld voor het bestuderen van netwerken van hersengebieden, zal de behoefte aan meer geavanceerde en tegelijkertijd optimale designs alleen maar toenemen en resulteren in nieuw werk voor geïnteresseerde statistici.

LITERATUUR

- Atkinson, A. C., Donev, A. N., & Tobias, R. D. (2007). *Optimum experimental designs with SAS*. Oxford: Oxford University Press.
- Culham, J. C. (2006). Functional neuroimaging: Experimental design and analysis. In R. Cabeza & A. Kingstone (Eds.), *Handbook of functional neuroimaging of cognition* (2nd ed., pp. 53–82). Cambridge, MA: MIT Press.
- Liu, T.T. & Frank, L.R. (2004). Efficiency, power, and entropy in event-related fMRI with multiple trial types, Part I: theory. *NeuroImage*, 21, 387–400.
- Maus, B. (2011). *Optimal experimental designs for functional magnetic resonance imaging*. PhD thesis. Maastricht: Universitaire Pers Maastricht/Datawyse.

BÄRBEL MAUS werkt als Post Doc bij de onderzoeksgroep Bioinformatics – Statistical Genetics, University of Liège, België. Haar proefschrift aan de Universiteit Maastricht ging over optimale experimentele designs voor fMRI. E-mail: <bmaus@ulg.ac.be>

GERARD VAN BREUKELN is universitair hoofddocent bij de afdeling Methodologie & Statistiek, Universiteit Maastricht, en gespecialiseerd in optimal design van interventiestudies in de psychologie, geneeskunde en gezondheidswetenschappen. E-mail: <gerard.vbreukelen@maastrichtuniversity.nl>



STATOR-VLEUGEL

Vloer 9 van de zuidvleugel van het Hoofdgebouw van de TU Eindhoven biedt onderdak aan STATOR-vakken zoals kansrekening, statistiek en operations research. Gaande van noord naar zuid over deze verdieping zie je eerst twee plaquettes ter ere van de verdiensten van professor H.C. Hamaker op het terrein van de kwaliteitscontrole: de Shewart Medal van 1979 en de Deming Award van 1982. Hamaker gaf in de jaren zeventig een dag in de week op de THE (TU/e) toegepaste statistiek. Hij was de enige die in de fietsenstalling een gereserveerde plaats had. De overige vier dagen was hij in dienst van Philips. Ik heb moeten zoeken naar Hamakers voornaam. In die tijd wist je niet dat hoogleraren voornamen hadden, in ieder geval gebruikte je die nooit.

Zes hoogleraren kijken je vanaf de muur aan: Jacques Benders – nog onlangs (2009) bekroond met de European Gold Medal for Operations Research voor een artikel uit 1966 –, Hugo Hamaker, Jaap Wessels, Paul van der Laan, Roel Doornbos en ondergetekende. Verschillende van hen waaronder ikzelf zijn (waren) erelid van de Vereniging voor Statistiek, maar van ‘versierselen’ geen spoor. Hoewel meer dan de helft van de geportretteerden nog in leven is zijn de lijsten diepzwart; de portretten ogen als groot formaat ‘bidprentjes’ – geen tekst op de achterkant.

Iets verder hangen er ansichtkaarten, door medewerkers met vakantie naar het thuisfront gestuurd, dat wil zeggen naar COSOR-secretaresse Harma Koops, die ze ophing. De meeste zijn van lang geleden. Er hangt er een van mij, die de Oostendorper watermolen in Haaksbergen laat zien, in de sneeuw.

Op twee leestafels liggen in het totaal zeven exemplaren van STATOR, waarvan vijf verschillende; allemaal uit de tijd dat ik als columnist nog twee extra

exemplaren kreeg – om ter inzage te leggen. Vroeger was dat.

In een vitrine aan de rechterkant van de gang liggen curiositeiten: oorkonden voor eerste- of tweedejaars onderwijsprijzen voor STATOR-vakken, zowel voor onderwijs aan wiskundestudenten als voor onderwijs aan ‘andersdenkenden’. Verder staat er allerhande demonstratiemateriaal voor toegepaste statistiek, meestal afkomstig van Hamaker. Een ‘tienzijdige dobbelsteen’ (een cilinder met een genummerde regelmatige tienhoek als basis) en een glazen potje met twee kleuren kralen, voorzien van een schepje om steekproeven te nemen, en een voor mij onduidelijk bouwwerk dat ook met steekproeftrekking te maken heeft.

Iets lager liggen de boekwerken. Een proefschrift van Doornbos, een nummer van *Statistica Neerlandica* helemaal gevuld door Jaap Wessels, het afscheidscolleage van Benders, congresverslagen over optimalisering geredigeerd door Rinnooy Kan, Schrijver, J.K. Lenstra en/of Wessels, een cursusboek over toegepaste statistiek, en natuurlijk mijn meesterwerk over oneindige deelbaarheid met Klaas van Harn, eigenlijk het enige echte boek.

Wie de STATOR-hoek nog wil bezoeken, zal zich moeten haasten. De faculteit Wiskunde&Informatica verhuist naar een nieuw gebouw, waar ik misschien nog een plaatsje krijg, maar misschien ook niet. Ik denk dat doctorandus Halbe – Wim Kan noemde doctorandussen dertig jaar geleden al ‘half-afgestudeerden’, een profetische blik – mijn plekje wel zal hebben wegbezuinigd. Voor de hier besproken relikwieën zal daar ook geen ruimte zijn. Als u ze nog wilt zien, haast u dus.

FRED STEUTEL is emeritus hoogleraar kansrekening aan de TU Eindhoven.

E-mail: <f.w.steutel@tue.nl>

VEILIGHEID IN DE ZORG DRAAGT OPERATIONS RESEARCH BIJ AAN BETERE ZORG? (DEEL 2)

JORIS VAN DE KLUNDERT

In deel 1 (*STAtOR*, 12(2011)3/4) hebben we gezien hoe geneeskunde en gezondheidswetenschappen zich steeds meer bedienen van, of mogen we zeggen bekeken tot, *evidence based practice*. Daarbij wordt vooral empirisch statistisch bewijs als *evidence* gezien voor de geschiktheid van een verandering in de zorg, voor de beoordeling van een interventie. Dit zien we niet alleen bij medicijnen, medische technologie, of medische behandelingen, maar ook bij veranderingen in management en organisatie van zorg. Net als menselijke lichamen, zijn zorgorganisaties complexe entiteiten waarbinnen de effecten van interventies zich moeilijk op basis van modelmatige abstracties laten voorspellen. Dit laat zich bijvoorbeeld illustreren aan de hand van de invoering van medicatieveiligheidssystemen die niet altijd leiden tot een betere veiligheid, maar soms tot een verslechtering, met dodelijke consequenties (Yong e.a., 2005). De invoering van nieuwe software kan leiden tot het niet meer plaatsvinden van mondelinge communicatie die letterlijk van vitaal belang bleek voor de pasgeborenen. De ΠΟΙΟΤΕΣ, de eigenschappen die de software en de organisatie waarin zij werd toegepast bleken te hebben, bleken anders dan tevoren verondersteld.

Empirisch bewijs

Een belangrijk verschil tussen *evidence based medicine* en *evidence based health services* is gelegen in het ver-

zamelen van empirisch bewijs. Laten we ter illustratie een organisatorische interventie beschouwen vanuit het research design van de Randomized Controlled Trial (RCT) zoals besproken in deel 1. De te beschouwen organisatorische interventie is de invoering van personeelsroostering software die gebruik maakt van combinatorische optimalisering en wachtrijmodellen.

Hoe zouden we het effect daarvan kunnen meten in een *double blind* RCT? We zouden dan een grote populatie van organisaties moeten hebben, verdeeld over een interventiegroep en een controlegroep, waarvan noch degenen die de interventie doorvoeren, noch de ontvangende organisaties weten of zij in de interventiegroep of in de controlegroep zitten.

Het moge duidelijk zijn dat dit researchdesign niet mogelijk is. De interventie kan onmogelijk *double blind* zijn. Bovendien is het in de praktijk niet mogelijk om een steekproefomvang van voldoende grootte te realiseren. In health services research bedient men zich daarom van zwakkere researchdesigns. Het gebruik van een controlegroep wordt bijvoorbeeld vaak vervangen door een *before-after-design*, waarin de prestaties voor de interventie worden vergeleken met de prestaties na de interventie. De kwaliteit van de empirische bewijsvoering speelt echter onverminderd een belangrijke rol bij de beoordeling van onderzoek en de waarde die aan de resultaten ervan wordt gehecht.

Een belangrijk aandachtspunt is de externe validiteit: zijn de resultaten ook van toepassing voor andere

De zuil van Hammurabi uit het antieke Mesopotamië. Collectie: Louvre Parijs.

Zoals Joris van de Klundert in deel schreef is de aandacht voor kwaliteit en veiligheid van zorg geen nieuwlichterij. De Codex Hammurabi uit ca. 1780 v.Chr. is een van de oudste tot dusver gevonden wetboeken. In de codex van Hammurabi wordt al uitgebreid aandacht besteed aan staaroperaties. De codex van Hammurabi regelt de veiligheid op juridische wijze door middel van het straffen van operateurs wanneer de operatie lijdt tot vervolgschade.

organisaties dan die waarbinnen het onderzoek heeft plaatsgevonden? Kunnen we bijvoorbeeld op basis van het onderzoek concluderen dat de invoering van de personeelsroostering software in andere organisaties dan die welke in het onderzoek geëvalueerd zijn dezelfde effecten zal hebben? Een *single case study* waarbij de interventie binnen één organisatie wordt toegepast, geldt daarbij als zwakste vorm van empirische evidentie. Laat dit nu net het favoriete design zijn voor toegepaste operations research.

Menselijke fouten voorkomen

Het probleem van operations researchers ten aanzien van het leveren van een bijdrage ligt daarmee vooral in de gebruikte methoden en de daarmee opgebouwde kennis. Operations researchers bedienen zich vooral van onderzoeksmethoden en onderzoeksopzetten die tot zwakke vormen van evidence leiden vanuit het gezichtspunt van hedendaagse *health service-researchers*. Uit eigen onderzoek weten we bovendien dat de operations research- en operations managementliteratuur veelal nalaat tot evidentie te komen. Van ruim honderd gereviewde artikelen die beogen algemeen geldige operations research/managementmodellen te ontwikkelen, bleken er minder dan twintig tot implementatie te hebben geleid. Van slechts zeven bleken de empirische resultaten geëvalueerd. Daarnaast bleek dat veel studies geen



of weinig gebruik hebben gemaakt van gerelateerde resultaten van andere onderzoekers, zodat maar zeer beperkt sprake lijkt van gezamenlijke constructie van wetenschappelijke kennis. Veelzeggend in dit kader is de vraag van Brailsford (2003): wat kan een 1019e simulatiestudie naar spoedeisende hulp afdelingen toevoegen aan de 1018 die al in de literatuur verschenen zijn? Meer algemeen, wat gaat ons volgende onderzoeksproject bijdragen aan de kwaliteit en veiligheid van zorg of de wetenschappelijke kennis daarover?

Kan operations research bijdragen aan de kwaliteit en veiligheid van zorg? Hoewel ik deze vraag vol overtuiging, en ondersteund met empirisch bewijs, positief zal beantwoorden, roep ik ook op tot bescheidenheid. Ten eerste is er een lange weg te gaan. Ten tweede zijn er belangrijke determinanten van kwaliteit en veiligheid waarvoor operations research in directe zin weinig te bieden heeft. Zo is bekend dat sociale vaardigheden en communicatie van groot belang zijn voor veiligheid. Zowel bij de onveilige zorg in in het UMC St. Radboud in Nijmegen als bij een dodelijk ongeval in het Westfriesgasthuis dat de nationale pers haalde, bleek slechte communicatie een cruciale oorzaak van onveiligheid.

Anderzijds zijn er tal van determinanten van veiligheid die operations research in principe positief kan beïnvloeden. *To err is human*, en mensen die langer werken maken meer fouten. Chirurgen maken meer fouten wanneer ze daags na een nachtdienst waarin ze hard hebben gewerkt opereren, dan wanneer ze een goede nachtrust hebben genoten. Orgaantransplantatieteams die vaak responsief in actie komen wanneer een orgaan beschikbaar komt, presteren ondoelmatiger bij transplantaties in weken waarin ze al een groot aantal transplantaties hebben verricht. Verpleegkundigen die grote werkdruk ervaren als gevolg van slechte personeelsbezetting maken eerder medicatiefouten. Deze menselijke fouten kunnen voorkomen worden door het systeem waarbinnen ze plaatsvinden, of ze er zelfs toe uitnodigt, te veran-

deren. Operations research kan een rol spelen bij het verbeteren, het optimaliseren van het systeem. Uit bovenstaande voorbeelden blijkt dat personeelsplanning en roostering van groot belang is voor de kwaliteit en veiligheid van de zorg.

Personeelsplanning en zorgpaden

Bij de huidige toepassingen van operations research technieken voor personeelsplanning en roostering ligt de nadruk vaak op de dimensie efficiëntie: het rooster vullen tegen de laagst mogelijke salarislasten. De impliciete ceteris paribus gedachte lijkt dat de andere kwaliteitsdimensies, in het bijzonder veiligheid, niet beïnvloed worden, terwijl er evidentie is dat het tegendeel waar is. Bovendien leidt onveilige zorg vervolgens tot extra kosten, zodat een initiële besparing op personeelslasten uiteindelijk tot duurdere zorg kan leiden. Bij toepassing van operations research technieken in personeelsplanning en roostering kan ook een expliciete bijdrage worden geleverd aan andere dimensies zoals doelmatigheid, veiligheid en tijdigheid van zorg. De operations research kennis kan juist van grote waarde zijn bij de multi-criteria optimaliseringsproblemen die zich voordoen bij het verbeteren van het multi-dimensionale begrip kwaliteit van zorg. Daartoe is noodzakelijk dat bekende empirische evidentie met betrekking tot de effecten van personele inzet op de verschillende dimensies van kwaliteit worden gebruikt en dat implementaties worden goed worden geëvalueerd teneinde verdere evidentie te verzamelen.

Personeelsplanning en roostering zijn slechts één voorbeeld. Planning en scheduling van het primair zorgverleningsproces biedt zelfs directere mogelijkheden om de kwaliteit van zorg positief te beïnvloeden. Dit kan bijvoorbeeld vorm krijgen in het plannen van zogeheten klinische paden die vaak een uitwerking vormen van evidence based medische behandelrichtlijnen. Uit een recente systematische

review (Rotter e.a., 2010) blijkt dat implementatie van zorgpaden echter zeker niet vanzelfsprekend resulteert in lagere kosten en/of verhoogde doelmatigheid. Net als scheduling, of personeelsroostering, is implementatie van zorgpaden een complexe interventie, die in een complexe organisatie zoals een ziekenhuis tal van onverwachte effecten kan hebben. Dit leidt tot de constatering dat operations research kennis alleen waardevol is wanneer die gepaard gaat met kennis over onder welke voorwaarden, in welke organisaties, en op welke wijze, operations research technieken kunnen bijdragen aan de verbetering van de kwaliteit en veiligheid van zorg. Het opbouwen van die kennis vraagt andere research designs dan theoretische modellen en single case studies.

Waardevol

Operations researchers blijken hun eigen *quality chasm* te dichten te hebben. Daarbij zij echter opgemerkt dat verscheidene onderzoekers daarmee al begonnen zijn. Een voorbeeld waarin toepassingen van operations management binnen health services research worden besproken, met daartoe geëigende researchdesigns, is het recente proefschrift van Ellen van Vliet (2011). Ook recente publicaties van de Universiteit Twente (2011) gaan in deze richting. Ander recent werk zijn de prachtige toepassingen op het gebied van toelevering van medicijnen in sub-Sahara Afrika, waarbinnen ook research designs met intelligent gekozen controle groepen worden toegepast zijn. Ook recent onderzoek naar de doelmatigheid van vaccinatie tegen HIV is vermeldenswaard. Er vindt prachtig en inspirerend onderzoek plaats. Daarbij horen ook simulatiemodellen, waarin waarschijnlijke effecten van interventies in complexe werkelijkheden die de menselijke ratio te boven gaan worden geanalyseerd. Op basis van modelstudies kunnen vervolgens randomized controled trials of andere grootschalige

en/of kostbare empirische onderzoeken worden uitgevoerd waarin verdere evidentie wordt verkregen. Operations research kan er zo voor zorgen dat de effectiviteit van evidence based health services research wordt vergroot.

Bovenstaande voorbeelden pretenderen niet volledig of representatief te zijn. Ze vormen illustraties, en wellicht inspiratie, voor de inzet van het krachtige instrumentarium van operations research: het modelleren van de essentie van de complexe werkelijkheid van zorgverlening. Operations researchers kunnen daarmee potentiële kwaliteitsverbeteringen vinden die anderen niet zullen ontdekken. Om te kijken of ze ook echt werken vereist dat ze in de praktijk worden getest en systematisch worden geëvalueerd. Zo kan operations research van grote waarde zijn voor de kwaliteit en veiligheid van zorg, en daarmee voor de gezondheid en kwaliteit van leven van miljarden mensen.

LITERATUUR

- Yong, Y. H., Carcillo, J. A., Venkataraman, S. T., Clark, R. S. B., Watson, R. S., Nguyen, T. C., Bayir, H., & Orr, R.A. (2005). Unexpected increased mortality after implementation of a commercially sold computerized physician order entry system. *Pediatrics* 116, 1506–1512
- Brailsford, S. (2003). *Overcoming the barriers to implementation of operations research simulation models in health-care*. Southampton UK: University of Southampton.
- Rotter, T., Kinsman, L., James, E. L., Machotta, A., Gothe, H., Willis, J., Snow, P. & Kugler, J. (2010). *Clinical pathways: effects on professional practice, patient outcomes, length of stay and hospital costs, accessible via <http://onlinelibrary.wiley.com/doi/10.1002/1471-2369.1111>*
- Vliet, E. van (2011). *Care pathways and operations management in het organisation of ophthalmic care*. Doctoral thesis, KU Leuven.
- Lent, W. A. M. van, Berkel, P. T. van, & Harten, W. H. van (2011). *A review on the relation between simulation and improvement*. Working paper, Universiteit Twente.

JORIS VAN DE KLUNDERT is hoogleraar Management & Organisatie van Zorgverlening bij het Instituut Beleid & Management van de Gezondheidszorg van de Erasmus Universiteit Rotterdam
E-mail: <vandeklundert@bmg.eur.nl>



VEEL MEDISCHE STUDIES KUNNEN EN MOETEN KLEINER

GEORGE BORM, JOANNA IN 'T HOUT

Een veelgebruikte methode in medisch onderzoek is de dubbelblinde gerandomiseerde klinische studie. Hierbij wordt de werkzaamheid van een of meer behandelingen geëvalueerd, bij voorbeeld de werking van een medicatie. Die behandelingen kunnen onderling vergeleken worden, of met een – in principe – niet-werkzame placebo behandeling. Het lot bepaalt welke behandeling iedere deelnemer krijgt, en de deelnemers, behandelaars en onderzoekers weten niet wie welke behandeling krijgt.

Onderzoekers die de ambitie hebben om hun onderzoek in een hoog-scorend tijdschrift te publiceren, moeten hun studie vooraf registreren op internet databases, bij voorbeeld bij <clinicaltrials.gov>. Uit deze registratie blijkt dat er ieder jaar wereldwijd minimaal 8.000 klinische studies worden gestart, met

3 miljoen patiënten. Hoewel de gemiddelde studie-grootte 340 is, kan de omvang van een studie aanzienlijk meer zijn en hebben sommige studies meer dan 10.000 deelnemers.

Betrouwbaarheid

Terwijl men in de natuurkunde vaak uitgaat van een 'vijf sigma'-criterium voor statistische evidentie, corresponderend met een p-waarde van ongeveer 0,00003%, is men in de medische wetenschap vaak tevreden met twee sigma, dat wil zeggen een p-waarde van 5%. Ondanks dit tamelijk milde criterium zijn er vaak grote studies nodig om tot eenduidige conclusies te komen. Hiervoor zijn diverse redenen aan te geven:

- Het resultaat van medische behandelingen varieert vaak erg. De ene patiënt heeft baat, de andere niet, het komt zelfs vaak voor dat een patiënt gedurende een bepaalde periode wel baat heeft, maar op een later tijdstip niet, of omgekeerd.
- Ook een placebo behandeling is soms erg effectief. Bij een studie naar antidepressiva kan het gebeuren dat 25% van de patiënten met placebobehandeling positief reageert, terwijl in de medicatiegroep een bescheiden 40% vooruitgaat.
- Resultaten zijn soms moeilijk te meten. Uiteindelijk is het doel meestal dat de patiënt betere 'kwaliteit van leven' heeft. Kwaliteit van leven is echter moeilijk te meten en hangt sterk af van zaken die niets met de behandeling te maken (lijken te) hebben, bijvoorbeeld sociale contacten of gebeurtenissen in de familie of nabije omgeving. Maar ook de ernst van de depressie, de ernst van de ziekte van Parkinson of de ernst van reuma zijn lastig eenduidig te meten en zelfs schijnbaar ondubbelzinnige gebeurtenissen als een hartinfarct blijken bij nader inzien niet altijd zo eenduidig vast te stellen.
- Vaak wordt er gepoogd om morbiditeit te voorkomen, bijvoorbeeld bij hartinfarcten. Omdat het percentage mensen dat een hartinfarct krijgt – statistisch gezien – laag is, zijn er relatief grote studies nodig om dit te onderzoeken.

Studies kunnen kleiner en transparanter

Het grote aantal proefpersonen dat nodig is in medisch onderzoek is een serieus probleem. Het leidt tot kostbare en langdurige studies. Soms zou een studie zo groot moeten zijn dat het onmogelijk is hem uit te voeren. Bij zeldzame aandoeningen bij voorbeeld zijn er soms gewoonweg niet genoeg patiënten om de studie te kunnen doen, zelfs als alle patiënten mee zouden doen.

Het is daarom van groot belang om onderzoek zo efficiënt mogelijk op te zetten en te zoeken naar methoden die leiden tot een zo betrouwbaar en nauw-

keurig mogelijk resultaat bij zo weinig mogelijk deelnemers. Helaas gebeurt dat vaak niet. Afhankelijk van het type onderzoek zouden 5% tot 25% van de studies zoals ze nu worden uitgevoerd, 25% tot 50% kleiner kunnen zijn, zonder verlies van betrouwbaarheid of precisie. De aanpassingen die hiervoor nodig zijn, zijn tamelijk eenvoudig:

Op de eerste plaats is efficiëntiewinst mogelijk door kleine wijzigingen in de studieopzet. Vaak kunnen er bijvoorbeeld gemakkelijk één of twee extra metingen aan het eind van de studie gedaan worden. Als deze meegenomen worden in de analyse neemt de precisie substantieel toe of, omgekeerd, kan er met een kleinere studie worden volstaan. Ook worden belangrijke baseline variabelen vaak niet als covariaten in de analyse meegenomen. Gebruik van covariantie-analyse kan in zo'n geval de precisie aanzienlijk verhogen (Borm, 2007).

Van dit soort eenvoudige verfijningen van de studieopzet en de analyses wordt te weinig gebruikt gemaakt, omdat de voordelen qua efficiëntie weinig bekend zijn bij de onderzoekers. Ze worden niet vermeld in de standaard statistiekboeken en maken ook niet standaard deel uit van het onderwijs. Het overgrote deel van de medisch onderzoekers maakt ook geen gebruik van ondersteuning door een statisticus en kopieert de methoden van eerdere studies. Daardoor blijven deze en vele andere efficiënte methoden onbekend en ongebruikt.

Daarnaast kan er nog winst geboekt worden door bestaande statistische methoden verder te optimaliseren. Een voorbeeld hiervan is de veelgebruikte 'responder analyse'. Hierbij wordt een behandeling als geslaagd beschouwd als het resultaat boven een bepaalde grens ligt. Bij voorbeeld een gewichtsverlies van 3 kilo voor een afslankmiddel. In de analyse worden dan het percentage proefpersonen met gewichtsverlies geëvalueerd. Een dergelijke studie zou aanzienlijk efficiënter worden als in plaats van een harde grenswaarde voor response een *fuzzy* benadering wordt gekozen. Hierbij wordt er rekening gehouden met de onzekerheid of er wel echt

een 'response' of een 'non-response' is, in het geval dat de waarde dicht bij de grenswaarde zit (Borm, 2007).

Drie kleine studies is beter dan één grote

Vaak worden er meerdere studies uitgevoerd om de werkzaamheid van een bepaalde behandeling te onderzoeken. Om uiteindelijk tot een conclusie te komen, wordt er dan een meta-analyse uitgevoerd die de resultaten van de studies combineert. Daarbij blijkt in veel gevallen dat de resultaten van de studies onderling onverwacht en onverklaarbaar sterk verschillen (heterogeniteit). Waarschijnlijk wordt die heterogeniteit veroorzaakt door factoren zoals culturele verschillen tussen landen en streken, eetgewoonten, genetische verschillen, verschillen in de gezondheidszorg, verschillen tussen ziekenhuizen, artsen, verpleegkundigen, etc. Om deze heterogeniteit te vermijden, wordt er dikwijls aanbevolen om maar één grote studie uit te voeren.

Dit is echter struisvogelpolitiek, want als er maar één studie gedaan is, kan er geen heterogeniteit worden vastgesteld, er zijn immers geen vergelijkbare studies. Maar die (extra) variabiliteit in de uitkomst van een studie is er nu eenmaal en daarmee moet dus rekening worden gehouden (Borm, 2009). De enige manier om dat te doen is door meerdere studies uit te voeren, liefst door verschillende onderzoekers, in verschillende landen, onder verschillende omstandigheden, etc., zodat die mogelijke heterogeniteit ten volle duidelijk kan worden. Omdat de informatie van deze studies gepoold kan worden, kunnen ze elk op zich kleiner zijn.

Ethische commissies

Medisch onderzoek moet vooraf worden goedgekeurd door ethische commissies. Zij letten erop dat de lasten en risico's voor de deelnemers aan het onderzoek

niet te groot worden en opwegen tegen de mogelijke informatie die de studie zal opleveren. Daarbij beoordelen zij ook of de studie niet te groot en vooral niet te klein is: de studie moet op zichzelf een duidelijk resultaat kunnen opleveren. Ze kijken echter niet of de studie efficiënt is opgezet, en of hetzelfde met minder deelnemers bereikt zou kunnen worden. Eigenlijk is dit vreemd, want het betekent dus dat soms onnodig veel proefpersonen risico en ongemak ondervinden. Ethische commissies, maar ook tijdschriften en subsidiegevers zouden daarom moeten aandringen op het gebruik van efficiënte methoden. Omdat de informatie die is gebaseerd op verschillende kleine studies betrouwbaarder is dan informatie verkregen uit één grote studie, zouden zij ook de mogelijkheid moeten bieden om kleinere studies uit te voeren, zelfs al zullen die elk op zich dan niet tot een definitieve conclusie leiden (Borm, 2009).

LITERATUUR

- Borm, G. F., Fransen, J. & Lemmens, W. A. J. G. (2007). A simple sample size formula for analysis of covariance in randomised clinical trials. *Journal of Clinical Epidemiology*, (60)12, 1234–1238.
- Borm, G. F., Munneke, M., Lemmers, O. & van Zuijlen, M. (2007). An efficient test for the analysis of dichotomized variables when the reliability is known. *Statistics in Medicine*, (26)18, 3498–3510.
- Borm, G. F., Lemmers, O., Fransen, J. & Donders, R. (2009). The evidence provided by a single trial is less reliable than its statistical analysis suggests. *Journal of Clinical Epidemiology*, (62), 711–715.
- Borm, G. F., Den Heijer, M. & Zielhuis, G. A. (2009). Publication bias was not a good reason to discourage trials with low power. *Journal of Clinical Epidemiology*, (62), 47–53.
- Borm, G. F. & Donders, R. A. (2009). Treatment should be evaluated by small trials. *Journal of Clinical Epidemiology*, (62) 879–887.

GEORGE BORM is hoogleraar biostatistiek aan de Faculteit der Medische Wetenschappen van de Radboud Universiteit Nijmegen en verbonden aan de Afdeling Epidemiologie, Biostatistiek en HTA van de UMC St. Radboud.
E-mail: <g.borm@ebh.umcn.nl>

JOANNA IN 'T HOUT is als universitair docent verbonden aan de Afdeling Epidemiologie, Biostatistiek en HTA van de UMC St. Radboud Nijmegen.

IN MEMORIAM

JAN ENGEL (1952-2011)



'Bridge the gap between statistics and the users'

Veel te jong en veel te vroeg is op 19 november 2011 na een kort ziekbed Dr. Jan Engel overleden. Jan was principal consultant bij adviesbureau CQM, waar hij ruim 34 jaar werkte. Hij stond nog volop in het leven, maar werd slechts 59 jaar.

Aansluitend op zijn hbs-diploma in 1969, studeerde Jan wis-, natuur- en scheikunde aan de Universiteit van Amsterdam. In 1976 ontving hij cum laude zijn doctoraal mathematische fysica met als bijvakken natuurkunde en mechanica. Geen statistiek dus. Dat heeft hij zich na zijn indiensttreding bij CQM min of meer zelf aangeleerd. In 1987 promoveerde Jan met lof bij prof. dr. Paul van der Laan in Wageningen en sinds 2003 is hij geregistreerd Biostatisticus VVS.

Het grootste gedeelte van zijn consultancywerk bestond uit ondersteuning aan R&D-projecten bij Philips Research Laboratories. Kenmerkend aan Jan was dat hij zich verdiepte in de research-vraagstellingen tot hij begreep hoe het onderliggende mechanisme in elkaar stak en het verantwoord was conclusies te trekken. Zijn scherpe analytisch vermogen zorgde ervoor dat je niet te snel naar conclusies sprong. Hij maakte duidelijk dat je een statisticus niet achteraf moest betrekken, maar aan het begin van het onderzoek. Jan sprak de taal van de researchers. Hij zag in dat je researchers geen statistiek moest leren, maar het in hun taal moest uitleggen.

Jan had, naast muziek en wandelen in de bergen, een grote passie voor ons vak. Soms was er voor een research-vraagstelling geen wiskundige oplossing voorhanden. Dat was een uitdaging voor hem. Zo ontstond eind jaren 80 zijn proefschrift. In de avond-

uren geschreven. De jaren erna zette hij het wiskundig onderzoek voort. Gemiddeld één publicatie per jaar en vele bijdragen aan workshops en conferenties. Zijn onderzoek was er steeds op gericht een brug te slaan tussen statistiek en de toepassers. Op een grondige en gedegen manier stelde hij deze kloof aan de orde in zijn bijdragen aan de ENBIS-conferenties, zijn beschouwingen in STAtOR en in zijn publicaties in vakbladen samen met researchers van Philips. Er vielen Jan twee zaken op:

In de industrie en het bedrijfsleven wordt veel empirisch onderzoek gedaan zonder dat hier statistici bij betrokken zijn. Soms resulterend in verkeerde analyses en conclusies.

Statistische methoden hebben een grote potentie voor toepassing in de industrie, mits er aan twee criteria wordt voldaan: 'functioneel' en 'bruikbaar' voor de toepasser. Statistische methoden worden veelal ontwikkeld door universiteiten en andere onderzoeksinstellingen. Sommige methoden voldoen aan deze twee criteria, vele ook niet. Het gevolg is dat potentieel relevante technieken in de praktijk niet zo vaak toegepast worden als zou kunnen.

Hij was nog niet klaar en had graag verder gewerkt, het heeft helaas niet zo mogen zijn. Met zijn bezieling en gedrevenheid heeft Jan een grote bijdrage geleverd aan ons vak, CQM en onze klanten. Wij wensen zijn vrouw Saapke, dochters Annet en Martine en alle anderen die hem een warm hart toedragen veel sterkte bij het verwerken van dit verlies. We blijven aan Jan denken met veel respect en grote bewondering.

Jaap Praagman & Bert Schriever

Wat bedoel je daarmee?

Als kind moet ik een ramp zijn geweest voor mijn ouders en onderwijzers. Ik was vrijwel nooit tevreden met een uitleg en wilde altijd meer weten.

Het begon al rond mijn zesde toen ik weigerde rode kool zo te noemen: iedereen kon zien dat die kool blauw of paars was en beslist niet rood. Tegen het einde van mijn lagere-school tijd speelde de Koreaanse oorlog. Omdat daar ook Nederlandse troepen bij betrokken waren was er steeds wel iets daarover te lezen in kranten en te horen op de radio (tv was er toen nog niet in ons land). Op een dag begon ik na te denken over het woord schiereiland dat ik in dit verband vaak tegenkwam. Mijn onderwijzer legde mij uit dat dit 'bijna'-eiland betekende, daarmee was ik op dat moment tevreden. Maar een dag later stond ik met een atlas voor hem, om te vragen waarom Italië geen schiereiland genoemd werd en Spanje/Portugal wel. Ook vroeg ik me af of Noord-Holland zonder Afsluitdijk misschien ook een schiereiland was enzovoorts. Helaas moest hij me teleurstellen: soms was het niet goed afgesproken wat men bedoelde en dan moest je maar kijken naar wat gebruikelijk was geworden.

Die fascinatie met goede definities heeft me nooit losgelaten en zal ook wel een mede-oorzaak voor mijn exacte studiekeuze zijn geweest. Ik herinner me het gevoel van opwinding bij de eerste wiskunde-colleges, ruim 50 jaar geleden. Veel van wat je op de middelbare school had geleerd kon je vergeten, er werd opnieuw begonnen en alles werd goed gedefinieerd. Heerlijk was dat!

Toen ik later in de statistiek verzeild was geraakt en voornamelijk sociale wetenschappers adviseerde bij hun onderzoek kwam die instelling uitstekend van pas. Vele malen waren er uitgebreide discussies als gevolg van mijn constante dóórvragen. Je wilt weten hoelang iemand aan zijn lesvoorbereiding besteedt? Prima, maar wat bedoel je met lesvoorbereiding, wil je daar ook jaarlijkse bijscholing bij rekenen, of het inwerken in een nieuwe lesmethode? Dat soort vragen stelde ik steeds in een poging om de definities helder te krijgen. Nu ik wat ouder en een klein beetje milder ben geworden kan ik achteraf best begrijpen dat ik wel eens ben beschuldigd van geslachtsgemeenschap met mieren...

Toch steekt die neiging naar goede definities en het willen weten hoe gegevens tot stand zijn gekomen nog steeds de kop op. Toen de afgelopen weken in het nieuws kwam dat Grieken in grote getale 'vergeten' overledenen aan te melden om zodoende de pensioenen te blijven innen, was dan ook mijn eerste gedachte dat hierdoor cijfers over de relatief hoge Griekse leeftijdsverwachting dus wel onjuist zullen zijn. Zou het zo maar kunnen dat het veelgeroemde heilzame effect van een mediterrane dieet niet bestaat? Misschien leef je helemaal wel niet langer door al die tomaten, olijfolie en knoflook! Toch maar eens een Griekse relatie vragen hoe ze hun leeftijds-statistieken berekenen: 'overleden zijn, wat bedoel je daarmee?'

GERRIT STEMERDINK is eindredacteur van Stator
E-mail: <gjstemerding@hotmail.com>