

STAtOR

Van Olympisch Goud naar Gronings Gas — Interview met Volko de Jong

De waterweg als oplossing voor de infrastructuurcrisis in Nederland

De vergissing van Prometheus

Instrumentele variabelen, uitschieters en de robuuste statistiek

De magische oplossing voor alle roosterproblemen?

De casus van het Alphense kunstwerk

TMLE, Deep Learning en Transformers

Tussen snelheid en begrip: AI in OR



STAToR

Jaargang 27, nummer 2, juli 2026

STAToR is een uitgave van de Vereniging voor Statistiek en Operations Research (WSOR). STAToR wil leden, bedrijven en overige geïnteresseerden op de hoogte houden van ontwikkelingen en nieuws over toepassingen van statistiek en operations research. STAToR verschijnt in maart, juli en december.

Redactie

Ruud Koning (hoofredacteur), Caroline Jagtenberg, Miriam Loois, Guus Luijben (eindredacteur), Kerry Malone, Gerard Sierksma, Richard Starmans, Gerrit Stermerdink (eindredacteur), Nicky van Foreest, Vanessa Torres van Grinsven en Inez Zwetsloot. Vaste medewerkers: Eric-Jan Wagenmakers, Jelke Bethlehem en John Poppelaars.

Kopij en reacties richten aan

Prof. dr. R.H. Koning (hoofredacteur), Rijksuniversiteit Groningen, Faculteit Economie en Bedrijfskunde, Vakgroep Economics, Econometrics & Finance, Nettelbosje 2, 9747 AE Groningen, stator@wsor.nl

Voor advertenties

Prof. dr. R.H. Koning, stator@wsor.nl

VWSOR

Bestuur

Voorzitter: prof. dr. ir. Geurt Jongbloed, db@wsor.nl; Secretaris: secretaris@wsor.nl; Penningmeester: dr. Rebecca Kuiper, penningmeester@wsor.nl; Algemeen bestuurslid: dr. Marianne Jonker en dr. Eni Musta, db@wsor.nl.

Secties

Voorzitters van de secties: dr. Marianne Jonker (Biometrical Section); dr. ir. Marjan van den Akker (Section for Operations Research); prof. dr. ir. Frank van der Meulen (Section Mathematical Statistics); dr. Rebecca Kuiper (Social Sciences Section); Elena Petridou, BSc (Young Statisticians); dr. Sanne Willems (Section Statistics Communication); dr. Stéphanie van den Berg (Section Statistics Education).

Leden- en abonnementenadministratie

WSOR, Maarsbergseweg 20, 3956 KW Leersum, admin@wsor.nl. Raadpleeg onze website www.wsor.nl over hoe u lid kunt worden van de WSOR of een abonnement kunt nemen op STAToR.

Uitgever

© Vereniging voor Statistiek en Operations Research

ISSN 1567-3383

De redactie heeft ernaar gestreefd de rechten van de illustraties te regelen. Degenen die desondanks menen rechten te kunnen doen gelden, kunnen zich alsnog tot de redactie wenden.

Toegepast!

In dit nummer van STAToR wordt weer eens duidelijk hoezeer ons vakgebied gericht kan zijn op praktische, soms alledaagse, problemen.

Nicky van Foreest en Gerard Sierksma interviewen Volko de Jong. In zijn diverse loopbaan heeft hij zich bezig gehouden met de energietransitie van kolen naar gas die door de Groningse gaswinning in een versnelling kwam. Ook werkte hij aan een typisch sportprobleem: bij volleybal schuiven de spelers een plaats in het veld op als de opslag van partij wisselt. Een slimme opstelling van de spelers kan daarbij van invloed zijn op de uitslag.

Ook het artikel van Jesse Nagel, Joris Slootweg en Rob van der Mei gaat over toegepaste statistiek en OR: de aloude waterwegen kunnen een oplossing zijn voor problemen in de Nederlandse wegen-infrastructuur.

Het maken van een afspraak met een arts, hoe moeilijk kan het zijn? Dat is minder eenvoudig dan men op het eerste gezicht zou denken. Belangrijk is dat de schaarse tijd van artsen optimaal wordt benut. Nienke Hamster heeft hier haar afstudeerscriptie aan gewijd en vertelt erover.

Verder heeft Richard Starmans een gedegen overzicht geschreven naar aanleiding van 20 jaar Targeted Maximum Likelihood Estimation.

Tijdens de Annual Meeting van 5 maart 2026 werden de winnaars van zowel de Van Zwet Award als de Jan Hemelrijk Award bekendgemaakt. In dit nummer leest u meer over alle winnaars van beide prijzen, met daarbij een ruime selectie uit de juryrapporten.

Jens Klooster is een van de winnaars van de Van Zwet Award. Hij legt uit hoe hij met behulp van robuuste statistiek problemen oplost die ontstaan bij de analyse van data met uitschieters.

Op de Annual Meeting werd ook de winnaar van de Jan Hemelrijk Award bekendgemaakt. Meer daarover leest u in dit nummer, inclusief een ruime selectie uit de juryrapporten.

Een nummer van STAToR zonder columnisten is ondenkbaar. Eric-Jan Wagenmakers schrijft over verrassende uitkomsten bij het gooien van een munt en John Poppelaars over de toepassing van AI in de OR en de afweging daarbij van snelheid versus begrip. Jelke Bethlehem keek weer naar een peiling. De gemeente Alphen a/d Rijn wilde een bestaand kunstwerk vervangen door een meer eigentijdse uiting. Zowel de gemeente als het Algemeen Dagblad raadpleegden hierover de bevolking, met uiteenlopende resultaten. Uiteindelijk bleef het kunstwerk staan—hebben de peilingen daarbij een rol gespeeld?

Tenslotte ook verenigingsnieuws, de datum voor de Annual Meeting 2027 is bekend (8 april 2027) en de Young Statisticians vertellen over hun activiteiten, met onder andere een zeer geslaagde pubquiz.

Mocht u onverhoopt tijdens de zomervakantie een praktisch probleem tegenkomen, probeer het dan eens te benaderen vanuit ons vakgebied, het is er veelzijdig genoeg voor.

Wij wensen u veel leesplezier!

De STAToR-redactie



4



8

0.7853981633974483096156608458

12



14



16



20



23



26



30

INHOUD

2 Toegepast!

4 **Van Olympisch Goud naar Gronings Gas — Interview met Volko de Jong** | Nicky van Foreest en Gerard Sierksma

8 **De waterweg als oplossing voor de infrastructuurcrisis in Nederland** | Jesse Nagel, Joris Slootweg en Rob van der Mei

12 **De vergissing van Prometheus** – column | Eric-Jan Wagenmakers

14 De winnaars van de 2025 Awards

16 **Instrumentele variabelen, uitschieters en de robuuste statistiek** | Jens Klooster

20 **De magische oplossing voor alle roosterproblemen?** | Nienke Hamster

23 **De casus van het Alphenese kunstwerk** – column | Jelke Bethlehem

26 **TMLE, Deep Learning en Transformers** | Richard Starmans

29 From the Young Statisticians

30 **Tussen snelheid en begrip: AI in OR** – column | John Poppelaars

31 Save the date: April 8, 2027 – VVSOR Annual Meeting



Van Olympisch Goud naar Gronings Gas — Interview met Volko de Jong

Nicky van Foreest en Gerard Sierksma

Kort nadat we Volko de Jong vroegen voor het interview overleed zijn promotor Tom Wansbeek. Volko was in de jaren tachtig zijn eerste promovendus. De reden Volko te interviewen was het bijzondere verloop van zijn loopbaan en vooral zijn bijdrage aan twee belangwekkende gebeurtenissen. Dat zijn de bijdrage aan het spectaculaire Olympisch goud van de heren volleyballers in Atlanta 1996 en zijn rol in de wereldwijde energietransitie van kolen naar gas in de decennia omstreeks 1990 waarbij ook het Groningse gas een rol speelde. Vanzelfsprekend hebben we in het interview gezocht naar verbindingen tussen deze twee en de 'lesson learned'.

Promotie bij Wansbeek

Interviewers Waar ging je proefschrift over?

Volko Experts in statistiek, informatica en het toepassingsveld opereerden te onafhankelijk van elkaar, elk met hun eigen jargon. Dit stond effectieve ontwikkeling van statistische software in de weg. Ik heb onderzocht hoe je kennis uit die drie vakgebieden kunt samenvoegen zodanig dat experts hun kennis kunnen toevoegen zonder de details van collega-experts te begrijpen. Met Paul Klint van het CWI heb ik een systeemconcept ontwikkeld met formele algebra en dat vervolgens met Larry Menninga van

Western Washington University U.S.A. uitgewerkt tot een prototype genaamd *Multi Level Specificatie Systemen*.

De energietransitie van kolen naar gas

Interviewers Hoe ben je in de internationale energiewereld beland?

Volko Na mijn promotie heb ik zo'n tien jaar in de IT-wereld gewerkt, namelijk bij BSO/Origin, aan projecten voor Aegon, Gasunie en KPN. Bij Aegon werkten we toen al met neurale netwerken om patronen in schadeclaims te ontdekken. Via projecten voor Gasunie kwam ik in contact met de energiewereld. De transitieproblematiek vroeg om het samenbrengen van technische, economische en bedrijfsculturele kennis — precies het driedimensionale thema van mijn proefschrift.

Interviewers Wat voor soort energietransitie was dat en wat was jouw rol daarin?

Volko Gas was toen de transitiebrandstof om kolen te vervangen. Het Groningse gasveld werd in 1962 ontdekt en Nederland was in 1969 vrijwel geheel overgestapt op gas. Ik werkte als eerste Managing Director van het Energy Delta Institute (EDI), opgericht in 2002 door Gasunie, Shell, RWE en de Universiteit Groningen, feitelijk een business school voor de energietransitie.

Interviewers Vertel eens iets meer over Gasunie.

Volko In 2000 maakte 99% van de huishoudens gebruik van Gronings gas. Het was de motor van de

tuinbouwontwikkeling in het Westland en de drijfveer voor vestiging van grote industriële bedrijven. Gas heeft de Staat honderden miljarden opgeleverd; ook de Deltawerken zijn ermee medegefinancierd. George Verberg, CEO van Gasunie (1992–2004), was zich ervan bewust dat de voorraad gas eindig was. Russisch gas zou nodig zijn als vervanging, en een goede relatie met Gazprom was een kerndoel van zijn beleid.

Interviewers Wie waren de cursisten?

Volko Samen met Shell, Gazprom, en topinstituten als Oxford, ESMT Berlijn en Skolkovo Moskou bouwden we masterprogramma's als 'Master of Gas Business' en 'Master of Petroleum Business Engineering'. De deelnemers kwamen uit de top van Europese energiebedrijven, waaronder ook Gazprom. Onze band met Gazprom was uniek; we hebben twee keer de RusPrix Award gekregen, die tot 2019 werd uitgereikt aan personen of organisaties voor hun bijdrage aan de zakelijke of culturele banden tussen Rusland en Nederland. Ik doceerde tot in de boardroom van Gazprom. Bij het 15-jarig bestaan heb ik Tina Turner en Deep Purple zien optreden in het Kremlin, wat een fascinerende fusie opleverde tussen de westerse muziek en voormalige Sovjet cultuur. Medvedev bleek bijvoorbeeld een groot fan van Deep Purple te zijn.

Het goud ging vooraf

In 1996 behaalde het Nederlands mannenvolleybalteam onder leiding van coach Joop Alberda op spectaculaire wijze de gouden medaille tijdens de Olympische Spelen in Atlanta. In de finale werd Italië in vijf sets verslagen (15-12, 9-15, 16-14, 9-15, 17-15) in een wedstrijd van bijna drie uur. In de allesbeslissende vijfde set stond Nederland op matchpoint bij 16-15 toen een Italiaanse speler de bal buiten de antenne over het net sloeg: foutbal, en daarmee goud voor Nederland. In 1999 kozen de kijkers van NOS Studio Sport dit als het sportmoment van de twintigste eeuw.

Interviewers Maar in jouw cv staat niets over jouw betrokkenheid bij dat team, terwijl die rol wel belangrijk is geweest. Hoe zit dit?

Volko Eerst even het volgende voor de niet-ingewijden. Volleybal is een teamspel waarbij de zes spelers van het team één veldpositie roteren zodra het team de serve 'terughaaft' doordat de tegenstander een fout maakt. Aldus zijn er tijdens de wedstrijd $6 \times 6 = 36$ verschillende veldbezettingen mogelijk, waarin de beide teams tegenover elkaar staan. Daarbij kan de coach gedurende de wedstrijd spelers wisselen, wat leidt tot extra rotaties.



Volko de Jong smasht in 1984, met set-up van Joop Alberda

In mijn volleybaltijd werden acties van spelers tijdens de wedstrijd met de hand 'geturfd'. Wat ontbrak was het real time vastleggen van kwaliteiten van spelersacties, zoals het aantal goede en foute acties en wel in combinatie met de beide teamopstellingen op het moment van die actie.

Joop Alberda, mijn nog altijd goede volleybalvriend met wie ik jaren samengespeeld heb in 'het eerste' van het Groningse Lycurgus, snapte als coach dat, als je realtime kon zien in welke opstelling je sterk bent en in welke opstelling niet, je dan de strategie en de opstelling direct kunt aanpassen om zo je winstkansen te vergroten. Joop had begin jaren 90 een statistieksysteem van Doug Beal, de toenmalige Amerikaanse succescoach, onder ogen gekregen. In dit systeem ontbrak de bovengenoemde rotatie-informatie, en toen dacht ik onmiddellijk: "dat kan met het toevoegen van rotatie echt veel beter".

Interviewers Wat wilde Alberda precies kunnen zien?

Volko Drie dingen: op welke rotatie winnen of verliezen we de bal; hoe speelt de stand een rol; en zijn spelers goed of slecht aan het begin of eind van de set? Mijn systeem legde acties vast in combinatie met de stand en de rotatie-opstelling. De assistentcoach van Alberda, Erik Jansen, zat langs het veld en registreerde in real time alle acties — geslaagde aanvallen, bloks, services — per speler,

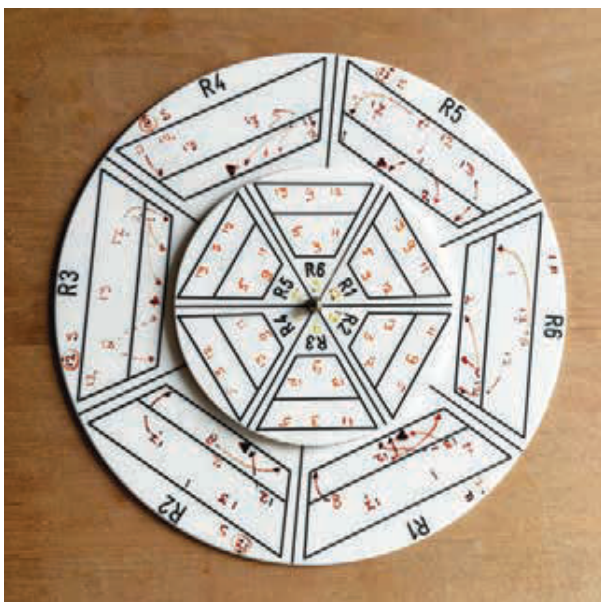
per rotatie, per stand, voor beide teams. Dat was nog een heel gedoe, met kabels en al. In aanloop naar Atlanta hebben we veel wedstrijden geanalyseerd en het systeem verbeterd. Uiteindelijk hadden we het 'succes'-systeem klaar en kon Joop tijdens de wedstrijden de beslissingen nemen die hij voor ogen had. Na afloop van de wedstrijd werden de resultaten met spelers en staf besproken.

Interviewers Wat voor modellen gebruikten jullie?

Volko Zeker geen geavanceerde statistische modellen. Het ging om slim zoeken in data om vragen te beantwoorden. We maakten een dashboard dat per rotatie liet zien wat er was gebeurd — vanuit een vraag van de coach: 'waar zit onze kracht, waar de fouten?' Daarna konden we in de video-beelden zoeken wat er precies aan de hand was. De zwakste schakel identificeren, want die bepaalt veel in volleybal.

Interviewers Hoe ontvingen de spelers de nieuwe 'data'-aanpak van jou en Alberda?

Volko Fascinerend was het cultuurverschil. Een Amerikaan die 90 procent van zijn passes goed deed, dacht: dat moet beter. Onze spelers reageerden defensief: 'die getallen kloppen niet', of ze vertoonden vermijdingsgedrag — opzij stappen zodat een ander de fout maakt. Het boek *De logica van het gevoel* van Cornelis heeft me geholpen dit te begrijpen: data raken aan emotie. Afhankelijk van de onderliggende cultuur ga je vluchten of aanvallen. In de volleybalwereld was toentertijd een cultuuromslag nodig voordat het systeem zou worden omarmd en dezelfde rol kon gaan spelen die heeft geleid tot het succes van onze "Lange Mannen". Dat inzicht heb ik later ook in de energiewereld enorm kunnen gebruiken.



'Gouden' rotatiebord van Alberda

Interviewers Wisten de tegenstanders van jullie data-aanpak? Hoe ging de cultuuromslag?

Volko Onze aanpak was zeker niet breed bekend. We waren pioniers — *pioneers get the arrows, settlers get the land*. Tegenwoordig is data-analyse in topsport gemeengoed. Wij waren zeker pioniers en hadden te maken met vragen als: Hoe 'bewijs' je dat zo'n aanpak werkt? Zo'n vraag is lastig te beantwoorden. Wel kun je laten zien wat de data betekenen en daarbij ook accepteren dat toeval altijd een rol speelt. Zeker voor mij is dat 'op den duur' een systematische aanpak ondersteund door data voordeel oplevert. Maar dat bewijs je niet met een enkele realisatie.

Interviewers Welke speciale kwaliteiten moet een coach hebben om met data-driven decision support-systemen te kunnen werken? Had de opvolger van Alberda die kwaliteiten?

Volko Joop was als coach fascinerend. Hij combineerde intuïtie met de bereidheid om naar data te kijken. Omdat Joop mij vertrouwde en ik hem, konden wij samen zoiets ontwerpen. Dat vertrouwen is de basis. Ik was vooral zijn enabler en verbinder — iemand die Joop effectief maakte door de juiste informatie op het juiste moment te leveren. Je kunt niet zeggen 'hij moet eruit, want mijn statistieken zeggen het'. Nee, het is *decision support*. Maar als de coach die informatie wil gebruiken, dan heb je ineens een enorm krachtig instrument.

Joops opvolger had niet zoveel op met data en statistieken. Die zei: 'ik zie toch wel wat er gebeurt'. Dat is een veelvoorkomende bottleneck voor data-geschoolde mensen: aan de andere kant zitten volslagen intuïtieve mensen die zeggen 'ik doe dat al zoveel jaar'. Na 'Joop' is het behoorlijk bergaf gegaan met het Nederlandse volleybal. Na de Spelen van 2004 in Athene is zelfs de kwalificatie voor de Olympische Spelen niet meer gehaald. Ik weet niet goed of die weerstand tegen data vandaag nog steeds een rol speelt in de sport. Wel weet ik uit mijn EDI-tijd dat weerstand in bedrijven bij managementinformatie niet ongewoon is.

Terug naar de energietransitie

Interviewers Je bent een goede volleyballer geweest en op hoog niveau daarin actief. De tijd van jouw succesvolle samenwerking met Joop Alberda en je ervaringen moeten een rol hebben gespeeld in je latere carrière. Hoe heeft de sport jou gevormd als internationaal transitie-adviseur?

Volko In mijn studietijd heeft de volleybalsport mij inderdaad mede gevormd, kan ik nu wel zeggen. Het is vormend om in een team te presteren onder de

druk van te willen/moeten winnen in een omgeving van constante onderlinge spanningen. Het vereist de discipline om gedurende langere periode de focus te houden op het moment en op het eindresultaat. Als econometrist heb ik geleerd om gestructureerd naar het spel te kijken, terugkerende situaties te herkennen waardoor je tijdwinst in je gedachten kunt boeken op de spannendste momenten.

Mijn denken is fundamenteel stochastisch dynamisch. Ik accepteer dat wat voor mij ligt onzeker is. In de IT-wereld wist ik niet welke kwaliteiten de mensen in mijn team hadden, wat de klant precies wilde, wat er zou gebeuren met de wetgeving. En omdat ik zo denk, wil ik dat ook vormgeven — met buffers, met opties, met scenario's. Dat botst met mensen die alles het liefst vasttimmeren. Ik zeg vaak: de toekomst bestaat uit *business as usual*, *surprise* en *fatal surprise*. Als je geluk hebt, mag je een heel leven *business as usual* doen. Maar van de ene dag op de andere kan alles weg zijn.

Als laatste ben ik me bewust geworden dat je te maken krijgt met mensen met totaal verschillende achtergronden en culturen. Dat heeft me altijd geïnspireerd. En dit speelde ook in de relatie tussen ons en Gazprom — je moet de ander begrijpen om samen verder te komen.

Interviewers Tot slot. Wetenschap bedrijven is tegenwoordig, zeker voor het grote publiek, vooral heel veel data verzamelen en AI. Zijn er, net zoals bij het gebruik van fossiele brandstoffen, grenzen aan de groei van dat verzamelen?

Volko Een fascinerende vraag. In de jaren 70, mijn studententijd, rekende je met een mainframe-computer. Bij BSO/Origin werkten we al met neurale netwerken. Het is fascinerend om te zien wat er sindsdien tot stand is gebracht. Maar of daar grenzen aan zijn? We zullen moeten leren de risico's van te veel energiegebruik te managen. Of dat lukt, is per definitie onzeker. Je moet je afvragen of AI over enkele jaren nog een "license to operate" heeft: willen alle noodzakelijke stakeholders nog mee? In een periode van vijf jaar zullen allerlei paradigma's veranderen en onverwachte gebeurtenissen plaatsvinden. De toekomst is vol verrassingen. Zolang je dat kunt managen gaat het ok. Maar altijd is een "fatal surprise" mogelijk.

Die stomme antenne aan het net heeft ons met volleybal eeuwige roem gebracht. Toeval of wetenschappelijke analyse? Ik denk beide. En dat ene punt was dan wel het verschil tussen winnen en verliezen.

En daar waren we het mee eens.

Curriculum Vitae

1972 – 1979	Master Econometrie en Actuariële Wetenschappen.
1979 – 1989	UD Econometrie. Lid GRECON-Team: macro-economische voorspellingen als tegenhanger van het CPB.
1984 – 1985	Assistant Professor Computer Sciences: Western Washington University, U.S.A.
1985 – 1988	Proefschrift A Multilingual Specification System for Statistical Software. Promotoren Paul Klint (CWI) en Tom Wansbeek (RUG).
1972 – 1984	Topvolleyballer: 3 jaar Nederlands Jeugdteam, 10 jaar eredivisie bij Raak/SFC en Lycurgus, enkele wedstrijden in het Nederlands Team.
1988 – 1996	Volleybal Support System ontwikkeld voor Joop Alberda, coach van het gouden olympische volleybalteam in Atlanta 1996, en voor live-ondersteuning op tv.
1989 – 1999	Diverse functies bij BSO/Origin (nu onderdeel van ATOS): ICT-architect en directeur vestiging Groningen.
1999 – 2025	Directeur/eigenaar Pareto Movere: business development en participatie maatschappij. Deelname Team Support Systems (later: ORTEC-Sports).
1999 – 2002	Directeur/eigenaar van Mount Everest: grote ICT-omscholingstrajecten voor KPN Telecom, Gasunie, Aegon en BSO/Origin.
2002 – 2012	Directeur, initiator en kerndocent Energy Delta Instituut (EDI): internationale business school met focus op energietransitie van kolen naar gas.
2012 – 2025	Directeur/eigenaar GGNi: internationale kennis- en businessdevelopmenttrajecten op basis van in Nederland aanwezige kennis van gas en energietransitie.
2016 – 2024	Business Coach Energy VentureLab: coachen van startups in de energiesector, zoals Envitron, Carbon Collectors en SeaQurrent.
2025 – heden	Senior Advisor City Capital Group: financiering van internationale energietransitieprojecten.

Nicky van Foreest is universitair docent aan de Rijksuniversiteit Groningen.

Gerard Sierksma is emeritus hoogleraar aan de Rijksuniversiteit Groningen.



De waterweg als oplossing voor de infrastructuurcrisis in Nederland

Jesse Nagel, Joris Slootweg en Rob van der Mei

In het voorjaar van 2024 haalde de snelweg A7 meermaals het landelijk nieuws. De maandenlange renovatie van de brug over het Noordhollandsch Kanaal bij Purmerend zou volgens voorspellingen leiden tot een ongekende opstopping op de wegen van Noord-Holland. In eerste instantie bleef de ergste drukte uit: de weggebruiker had blijkbaar goed geluisterd naar het advies van Rijkswaterstaat om thuis te werken. Maar toen in mei een rijstrook werd afgesloten, volgde de beloofde file alsnog.

Het probleem van onze wegen

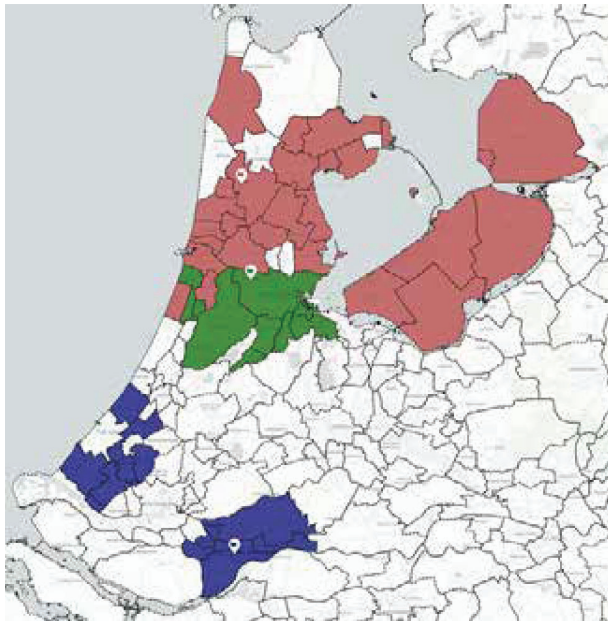
De problemen van de A7 zijn helaas niet uitzonderlijk. Grote delen van onze nationale infrastructuur zijn gebouwd gedurende de jaren 50 en 60, en bereiken nu gelijktijdig het einde van hun levensduur. De steeds grotere hoeveelheden voertuigen op onze wegen en het zwaarder worden van deze voertuigen, maakt de situatie alleen maar erger. Rijkswaterstaat heeft de komende vijf jaar meerdere grote nationa-

le projecten op de planning. Daarnaast zijn er op lokaal niveau nog veel meer bruggen, viaducten en wegen die aan vervanging of renovatie toe zijn. Zulke projecten halen zelden het nationale nieuws, maar vormen wel een hoofdpijndossier voor provincies en gemeenten. Genoeg redenen voor de provincie Noord-Holland om na te denken over alternatieven voor het traditionele vrachtverkeer.

Zo kwam de provincie Noord-Holland bij ons met een simpele observatie, namelijk: de wegen in de provincie zijn vol, maar de waterwegen zijn vrij. Daarna volgde een logische vraag:

Waar zijn mogelijkheden om zwaar verkeer van de weg te halen?

In samenwerking met studenten van de VU onderzochten we voor meerdere logistieke stromen in hoeverre vervoer over het water haalbaar was. Al snel bleek dat niet elke stroom even geschikt is: sommige stromen vragen om hoge snelheid of strakke levertijden, andere hebben een versnipperde herkomst of missen goede overslagpunten aan het water. Uiteindelijk kwamen we uit bij afval, een stroom die aan de meeste voorwaarden voor vervoer over het water voldoet.



Figuur 1: De in het onderzoek meegenomen gemeenten. De klantgemeenten van de verschillende verbrandingscentrales zijn kleurcodeerd.

Afval als vracht

Afval is een bijzondere logistieke stroom, waar de meesten van ons waarschijnlijk niet al te lang over nadenken nadat de kliko is geleegd of we het in de ondergrondse container hebben gegooid. Het meeste huishoudelijk restafval in Nederland wordt verbrand, waar we dan weer energie voor terugkrijgen. Dat gebeurt op een twintigtal locaties verspreid door het land, waar het afval vanuit de gemeentes naartoe gebracht moet worden. Inzamelvoertuigen rijden door de straten en legen de kliko's of de ondergrondse containers. Daarna wordt het vuilnis naar een overslagpunt in de gemeente gebracht. Daar wordt het verzameld en weer door andere vrachtwagens naar de verbrandingscentrales gebracht.

Specifieke eigenschappen van afval maken het bijzonder geschikt voor vervoer over het water. Zo is afval bijvoorbeeld *weinig bederfgevoelig*. Nou wil niemand dat afval ergens weken blijft liggen, maar een dag meer of minder maakt niet zoveel uit, in tegenstelling tot bijvoorbeeld de bevoorrading van winkels. Ook is de stroom huishoudelijk restafval die we produceren redelijk constant, deze verandert weliswaar langzaam van jaar op jaar maar de wekelijkse hoeveelheden zijn vrijwel gelijk. Er hoeft dus relatief weinig geïnvesteerd te worden in extra capaciteit voor piekbelasting. En misschien nog wel het belangrijkste, een groot deel van deze verbrandingscentrales ligt aan het water.

Casus

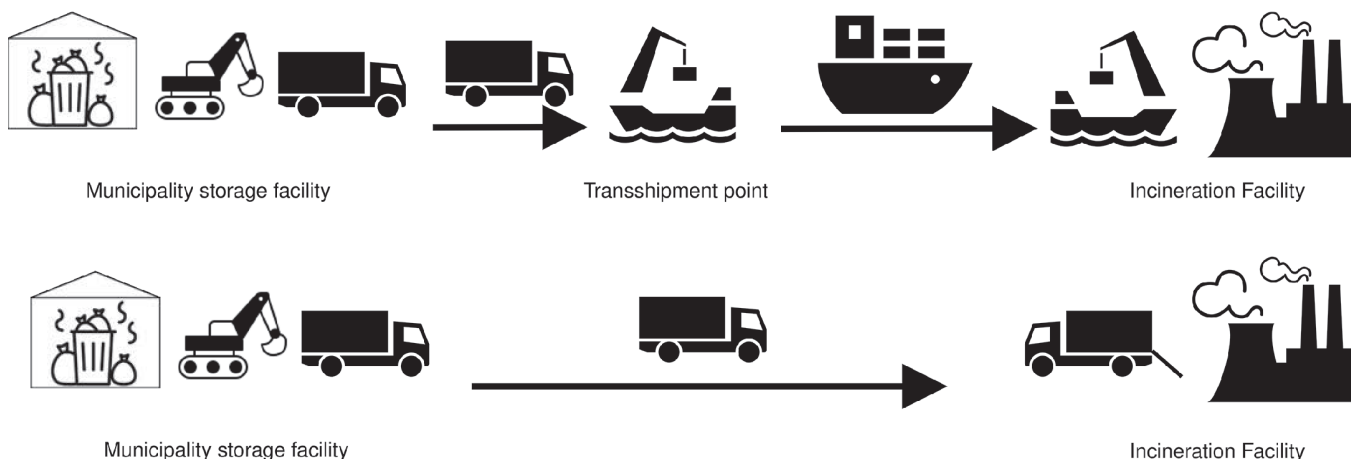
Tot nu toe hebben we niks verteld wat nog niet bekend was. Zo publiceerde Topsector Logistiek in 2022 al een wegwijzer 'Afvaltransport over het water' met suggesties voor een *modal shift* naar het water (Topsector Logistiek, 2022). Echte cijfers blijven echter uit, terwijl hier vanuit het ambtelijk apparaat wel behoefte aan is. Door een systeem voor vervoer van huishoudelijk restafval over het water door te rekenen proberen wij hier verandering in te brengen.

In het model gaan we uit van het huishoudelijk restafval van 55 gemeenten verspreid over de provincies Noord-Holland, Zuid-Holland en Flevoland. Het huishoudelijk restafval van deze gemeenten wordt op drie locaties verwerkt: HVC Alkmaar, HVC Dordrecht, AEB Amsterdam. Hierbij vormen de klantgemeenten van HVC Dordrecht een los cluster in Zuid-Holland, terwijl de klantgemeenten van HVC Alkmaar en AEB Amsterdam een gecombineerd gebied in Noord-Holland en Flevoland vormen. Zie figuur 1 voor een overzicht van de in het onderzoek meegenomen gemeenten.

In figuur 2 visualiseren we hoe het beoogde systeem met vervoer over het water moet werken. Gemeenten verzamelen het afval lokaal. Vanuit de gemeente kan het afval vervolgens op twee manieren naar de verbrandingscentrale worden gebracht. Via een directe verbinding per vrachtwagen, of via het water. In dat laatste geval moet het afval wel eerst per vrachtwagen naar een overslagpunt gebracht worden. Zo'n overslagpunt is een kade met kraan, waar het afval vanuit de vrachtwagen op de schepen kan worden geladen. In ons model gaan we uit van nieuwe locaties die speciaal voor het afvalvervoer geschikt gemaakt moeten worden.

We bekijken twee mogelijke opties voor het vervoer over het water. De eerste optie is om dieselschepen met duwbakken te gebruiken, het alternatief is om moderne elektrische duwschepen te gebruiken. De dieselschepen zijn makkelijk te huren, maar hebben een hoge uitstoot. De elektrische duwschepen hebben een significant lagere uitstoot, maar zijn een nieuwe technologie. De enige bij ons bekende inzet van elektrische duwschepen van groot formaat is bij de cacaoafabriek van Cargill in Zaandam, waar het dagelijks vervoer van cacao naar de haven van Amsterdam met E-Pushers wordt uitgevoerd. We gaan er in deze casus vanuit dat we de dieselschepen kunnen charteren, maar dat de elektrische schepen aangeschaft moeten worden.

De casus is volledig op openbaar toegankelijke data gebaseerd. Het doel is om voor een representa-



Figuur 2: Schematisch overzicht van het beoogde systeem.

tieve set gemeenten duidelijk te krijgen of afvalvervoer over het water een kansrijke strategie is voor provincies en gemeenten.

Model

Ons model moet meerdere systeemkeuzes maken:

1. Welke overslagpunten moeten geopend worden?
2. Hoeveel afval gaat er per gemeente over de weg naar de verbrandingscentrale?
3. Hoeveel afval gaat er per gemeente naar het overslagpunt?
4. Welke routes varen de schepen en hoeveel afval halen ze op bij de overslagpunten die ze aandoen?

Om deze vragen te beantwoorden gebruiken we een Integer Programming (IP) formulering, die we oplossen met een externe solver. Deze formulering is een combinatie van standaardformuleringen voor het *facility location problem* (FLP) en het *capacitated vehicle routing problem* (CVRP).

Doordat we tegelijk beslissingsvariabelen hebben voor zowel de locaties van de overslagpunten als de routing van de schepen, ontstaat een complex probleem. Een observatie uit de praktijk maakt alles gelukkig een stuk simpeler. Door de schepen maximaal één overslagpunt aan te laten doen per route kunnen we gebruik maken van twee duwbakken per route. Terwijl het duwschip een volle bak van het overslagpunt naar de verbrandingscentrale brengt, kan op het overslagpunt de leeg achtergelaten duwbak gevuld worden. Zo wordt de wachttijd van het duwschip geminimaliseerd en kan het relatief dure duwschip efficiënter gebruikt worden. Hierdoor verdwijnt de routing-combinatoriek grotendeels en blijft een toewijzingskeuze over.

In een gevoeligheidsanalyse keken we ook naar routes met meer dan één stop. Hieruit blijkt dat de extra efficiëntie van routes met meerdere stops (ongeveer 1% minder kosten) niet opweegt tegen de praktijkvoordelen van het systeem met meerdere duwbakken.

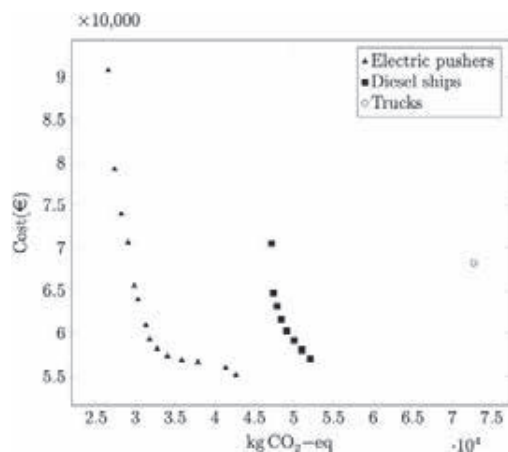
Hiermee reduceren we het probleem in feite tot een toewijzingsprobleem, waarbij we schepen toewijzen aan overslagpunten. Dit probleem is een stuk simpeler dan de eerder genoemde FLP en CVRP, en de IP solver heeft er dan ook weinig moeite mee.

De doelfunctie houden we simpel: we minimaliseren de wekelijkse kosten van het totale systeem. Hierbij nemen we naast de operationele kosten ook de investeringen in elektrische duwbotten en overslagpunten mee. Vervolgens kijken we naar twee andere metriekeken: de uitstoot in kg CO₂-eq en het aantal tonkilometers vervoerd per vrachtwagen. Deze metriekeken worden echter niet meegenomen in de doelfunctie.

Resultaten

Door het Integer Program op te lossen voor de verschillende netwerken (dieselschepen, elektrische schepen, geen schepen) krijgen we een beeld van kosten en baten van de verschillende systemen.

De resultaten zijn voor ons als onderzoekers verrassend: het systeem met elektrische duwbotten zorgt niet alleen voor de minste uitstoot, maar is ook het goedkoopste. We zien een kostenreductie van 19% ten opzichte van het systeem met alleen vrachtwagens voor het systeem met elektrische duwbotten en 16,5% voor het systeem met dieselboten.



Figuur 3: De afweging tussen uitstoot en kosten.

Daarnaast zien we een reductie van de uitstoot van broeikasgassen van 41% en 28,5% voor respectievelijk het elektrische en het dieselsysteem.

Als we kijken naar het aantal tonkilometers afval per vrachtwagen, zien we dat het dieselschepensysteem net iets meer vrachtwagens van de weg haalt. Tegelijk realiseren beide systemen een reductie van ongeveer 50% in afvalgerelateerd vrachtwagenverkeer. Dat verschil komt vooral door een van de modelaannames: voor dieselschepen staan we charteren toe. Daardoor kunnen we bij dieselschepen flexibel extra schepen inzetten wanneer dat loont. Elektrische schepen moeten in ons model echter worden aangekocht; die hogere vaste investering zorgt voor een andere afweging tussen wel of niet een schip inzetten ten opzichte van het dieselsysteem.

In figuur 3 laten we de afweging tussen uitstoot en kosten zien. Voor het systeem met enkel vrachtwagens laten we maar één oplossing zien, we hebben hierbij alleen op kosten geoptimaliseerd en niet op uitstoot. Een iets duurzamer systeem met alleen vrachtwagens is dus eventueel mogelijk.

De kernboodschap is dat elektrische duwbotten ruimte bieden voor een veel grotere uitstootreductie dan tot nu toe gezien in het meest kostenefficiënte systeem. Met een relatief beperkte extra investering in elektrische duwbotten is een substantiële verdere daling van de uitstoot haalbaar. Wáár precies de optimale balans ligt tussen extra kosten en extra reductie, is uiteindelijk een politieke keuze.

Vooruitzichten

Dat een overstap naar huisafvalvervoer over het water een duurzamere keuze was dan vervoer over de weg voelde iedereen die we spraken binnen het ambtelijk apparaat eigenlijk wel aan. Onze conclusie, dat het systeem niet alleen voor minder uitstoot zorgt maar ook met een verlaging van kosten gepaard

gaat, was een verrassing. Logische vervolgstappen zijn om het onderzoek in de praktijk te brengen, bijvoorbeeld door een pilot-project bij een van de verbrandingscentrales.

Daarnaast is het ook interessant om te kijken naar welke andere logistieke stromen geschikt zijn voor vervoer over het water. Huishoudelijk restafval is immers een grote logistieke stroom, maar zeker niet de enige.

De druk op onze infrastructuur zal in de komende jaren alleen maar groter worden, hopelijk kan het water hierbij een ontlastende rol spelen. De politiek moet hierbij een verbindende rol spelen. Individuele partijen (vervoerders, eindgebruikers, etc.) weten niet waar te beginnen, ondanks dat de intentie er vaak wel is. De benodigde investeringen zijn vaak te hoog en de gevolgen te onzeker. De provincie zou hierbij kunnen coördineren, bijvoorbeeld op het gebied van pilots. Zo zouden zorgen van ondernemers en andere belanghebbenden kunnen worden weggenomen.

Voor meer details over de wiskundige inhoud van dit artikel, zie J. Nagel, J. Slootweg, R.D. van der Mei et al., 2026.

Literatuur

J. Nagel, J. Slootweg, R.D. van der Mei et al. "Transporting household waste over water can reduce costs and emissions: A case study in the Netherlands". In: *Waste Management* 210 (2026), p. 115241.

Topsector Logistiek. *Wegwijzer Afvaltransport over Water*. 2022. URL: <https://topsectorlogistiek.nl/kennisbank/wegwijzer-voor-afvalvervoer-over-water..>

Jesse Nagel is PhD student bij het CWI en de VU. Zijn onderzoek is deel van het COCONUT project dat zich tot doel stelt knelpunten in de infrastructuur op te lossen of te verbeteren door slimme algoritmes op basis van coördinatie en personalisatie. Infrastructuur wordt in dit onderzoek breed opgevat en bevat bijvoorbeeld terreinen als energie, toerisme, distributiecentra, luchthavens, (water)wegen en openbaar vervoer. E-mail: jesse.nagel@cw.nl

Joris Slootweg is PhD student bij het CWI en lecturer aan de VU. Zijn promotieonderzoek is ook deel van het bovengenoemde COCONUT project.

Rob van der Mei is als hoogleraar Toegepaste Wiskunde en Manager Research & Strategy verbonden aan de VU en het CWI. Hij richt zich met name op het modelleren en optimaliseren van logistieke processen in de praktijk, en slaat daarmee een belangrijke brug tussen theorie en toepassing. In 2021 ontving hij voor zijn werk samen met VU-collega Sandjai Bhulai de prestigieuze Hui-brechtsenprijs. De laatste jaren is hij vooral werkzaam op het gebied van ambulance- en brandweerplanning, predictive policing, het reduceren van wachtlijsten in de zorg, revenue management, en het optimaliseren van personenmobiliteit en vrachtlogistiek. Hij is tevens de initiator en leider van het COCONUT project.

0.785398163397448309615660845819875721049292349843776455343726480744051122...

De vergissing van Prometheus

Als Prometheus een wiskundige zou zijn geweest, dan had hij de mensheid twijfelloos niet het vuur geschonken, maar het getal π . Oorspronkelijk gedefinieerd als de verhouding tussen omtrek en diameter van een cirkel, duikt π overal in de wiskunde op, zelfs als er nergens een cirkel te bekennen valt. In de statistiek bijvoorbeeld is $1/\sqrt{2\pi}$ de normalisatieconstante voor de standaardnormale verdeling. Deze constante volgt uit de Gauss-Euler-Poisson-integraal, die eigenlijk vooral toegeschreven moet worden aan Laplace (en aan de Moivre):

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}.$$

Deze vergelijking (en het latere inzicht m.b.t. haar universele toepasbaarheid) behoort tot de grootste triomfen van het menselijk intellect. Het is spijtig dat dit voor de student vaak angstvallig verborgen wordt gehouden.

Hier wil ik het hebben over een ander scenario waar π plotsklaps om de hoek komt kijken (zie ook Propp, 2026). Werp een eerlijke munt net zolang totdat er één keer vaker 'kop' dan 'munt' is gevallen, en noteer vervolgens de proportie 'kop'. Dus voor de reeks 'MKMKK' noteer je '3/5', voor 'MKK' noteer je '2/3', en voor 'K' noteer je '1'. Speel dit spel opnieuw en opnieuw. De verwachtingswaarde van de proporties is... $\pi/4$ (!)

Dit is op zichzelf al een hoogst opmerkelijk resultaat (er is immers geen cirkel in zicht, en ook geen normaalverdeling), maar het wordt nog gekker – veel gekker. In een recente preprint hebben Bruss en

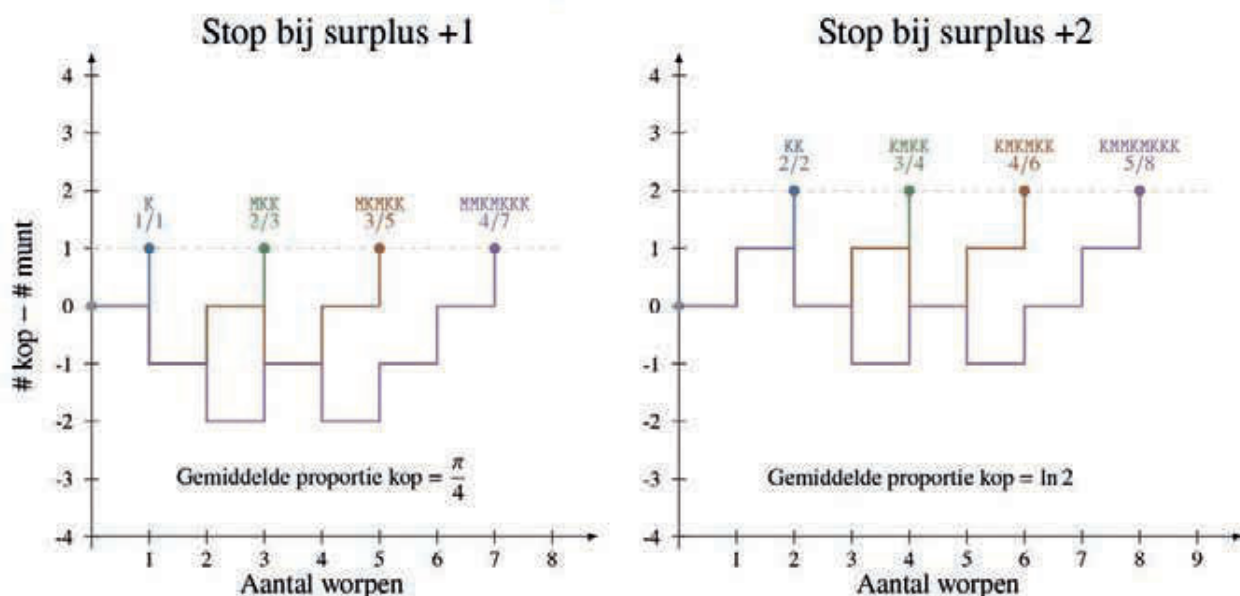
Paindaveine (2025) bestudeerd wat er gebeurt als je stopt bij een surplus 'kop' van $m = 2$ (dus bv. 'KK', of 'MMKKMKKK'). In dat geval is de verwachtingswaarde van de proportie... $\ln 2$ (!). Figuur 1 visualiseert de twee scenario's als een random walk met een enkele absorberende bovengrens.

Tabel 1 laat de verwachtingswaarden zien voor $m = 1, \dots, 8$; het is duidelijk dat een *oneven* surplus leidt tot oplossingen van het type $\pi/4$, terwijl een *even* surplus leidt tot oplossingen van het type $\ln 2$. De twee transcendenten wisselen elkaar af als in een wiskundige tango.

Tabel 1: Verwachtingswaarde voor de proportie 'kop' wanneer een eerlijke munt net zolang wordt geworpen totdat het verschil in het voordeel van 'kop' precies m groot is. De simulatiewaarden zijn gebaseerd op 10^4 – 10^5 onafhankelijke runs. Zie de blogpost <https://www.bayesianspectacles.org/a-transcendental-dance/>.

m	Type	Verwachte Proportie		
		Exact	Numeriek	Simulatie
1	$\pi/4$	$\pi/4$	0,7854 ...	0,7858
2	$\ln 2$	$\ln 2$	0,6931 ...	0,6939
3	$\pi/4$	$3 - \frac{3}{4}\pi$	0,6438 ...	0,6442
4	$\ln 2$	$2 - 2 \ln 2$	0,6137 ...	0,6144
5	$\pi/4$	$-\frac{10}{3} + \frac{5}{4}\pi$	0,5937 ...	0,5946
6	$\ln 2$	$-\frac{3}{2} + 3 \ln 2$	0,5794 ...	0,5817
7	$\pi/4$	$\frac{91}{15} - \frac{7}{4}\pi$	0,5689 ...	0,5696
8	$\ln 2$	$\frac{10}{3} - 4 \ln 2$	0,5607 ...	0,5617

Muntworpen als Random Walk



Figuur 1: De muntworp als wandeling. Bij kop gaat het pad omhoog, bij munt omlaag; de wandeling stopt zodra de gestippelde horizontale grens wordt bereikt. Links stopt de wandeling zodra een surplus 'kop' van +1 is bereikt; rechts zodra een surplus van +2 is bereikt. De labels tonen per pad de worpreeks en de bijbehorende proportie kop.

Merk op dat de verwachtingswaarde tendeert naar 0,5 wanneer m toeneemt. Voor grote m stop je weliswaar bij een hogere absolute verschillscore, maar om deze te bereiken moet je veel vaker gooien, zodat de verwachtingswaarde voor de proportie dichter bij 0,5 ligt.

Hoe komt het nu dat een oneven surplus leidt tot een $\pi/4$ -term, en een even surplus tot een $\ln 2$ -term? Intuïtief gezien moet eerst worden opgemerkt dat een reeks die stopt bij een (on)even surplus enkel kan eindigen na een (on)even aantal worpen (bv. voor $m = 1$ na 1, 3, 5, etc. worpen, en voor $m = 2$ na 2, 4, 6, etc. worpen). In de uitkomst ruimte zijn de twee klassen van reeksen dan ook strikt gescheiden. Wiskundig gezien zit het grofweg als volgt (zie Wagenmakers en Waldorp, 2026 voor details). Een *oneven* surplus m brengt uiteindelijk de volgende integraal met zich mee:

$$c \int_0^1 \frac{1}{1+u^2} du = c \cdot \arctan(1) = c \cdot \frac{\pi}{4},$$

en de π komt dus in het spel vanwege de arctangensfunctie. Een *even* surplus m daarentegen brengt een andere integraal met zich mee, namelijk

$$c \int_0^1 \frac{u}{1+u^2} du = c \cdot \frac{1}{2} \ln(1+u^2) \Big|_0^1 = c \cdot \frac{1}{2} \ln 2,$$

en dit verklaart de $\ln 2$ -term.

Of dit resultaat van enig praktisch belang is valt nog te bezien, want het is onduidelijk waarom je een ongewogen gemiddelde zou nemen van proporties die gebaseerd zijn op verschillende aantallen worpen. Desondanks maakt deze transcendente tango een verbluffende eerste indruk. Ik wil niet ondankebaarder lijken, maar wellicht had Prometheus toch het verkeerde geschenk gekozen...

Literatuur

- F. T. Bruss en D. Paindaveine. "Win rates at first-passage times for biased simple random walks". In: *ArXiv* (2025). URL: <https://arxiv.org/abs/2512.21254>.
- J. Propp. "Estimating π with a coin". In: *ArXiv* (2026). URL: <https://www.arxiv.org/abs/2602.14487>.
- E.-J. Wagenmakers en L. Waldorp. "A note on the $\pi/4$ coin-toss formula and its extension". In: *Technical Report* (2026). URL: <https://osf.io/nwv2v/files/cvr8x>.

Eric-Jan Wagenmakers is hoogleraar Bayesiaanse Methodologie bij de afdeling Psychologische Methodenleer aan de Universiteit van Amsterdam. Sinds 2012 ontwikkelt zijn team het open-source statistiekprogramma JASP (<https://jasp-stats.org/>), en sinds 2025 is hij CEO van JASP Services B.V. (<https://www.jasp-services.com/>). Het is derhalve aan te raden om positieve opmerkingen over JASP onafhankelijk te verifiëren. E-mail: EJ.Wagenmakers@gmail.com



De winnaars van de 2025 Awards

Als onderdeel van de Annual Meeting werden op 5 maart 2026 de winnaars van twee van de drie VWSOR Awards bekendgemaakt. De derde Award is de prestigieuze Van Dantzig Award die eens in de vijf jaar wordt toegekend en in 2029 weer aan de beurt is.

Jan Hemelrijk Award

Voor de Jan Hemelrijk Award voor de beste MSc Thesis waren er vier nominaties: twee op het gebied van de statistiek en twee over een onderwerp uit de Operations Research. Volgens het juryrapport werd gekeken naar:

- a critical approach and an original contribution to the subject;
- ability to state and apply theoretical concepts from statistics and/or operations research in an inventive way;
- the practical applicability of the findings and results;
- a clear line of reasoning and communication skills;
- in the thesis the candidate's own research contribution should be clearly noticeable.

Joshua Rooijakkers

De scriptie die het best aan deze criteria voldeed was "Robust and Conjugate Multi-Output Gaussian Processes", geschreven door Joshua Rooijakkers, onder supervisie van professor Dick van Dijk van de Erasmus School of Economics van de EUR. Het onderzoek is deels uitgevoerd aan het University College London.

De jury oordeelt: "The thesis extends the framework of robust and conjugate Gaussian processes to multiple dimensions, so to handle multiple outputs simultaneously. It combines scientific innovation with relevant applications such as cancer drug-response. Joshua's contribution is clearly visible, there is excellent and advanced theoretical analysis, and the empirical validation makes it a practical convincing work."

Willem R. van Zwet Award

Dit jaar waren er veel sterke inzendingen voor de Willem R. van Zwet Award, waaronder veel proefschriften over toegepaste statistiek. De voorzitter is blij dat zijn oproepen hiervoor effect hebben gehad. Door de hoge kwaliteit heeft de jury gekozen voor twee winnaars en een eervolle vermelding.

Lasse Vuursteen

De eerste winnaar is Lasse Vuursteen met zijn proefschrift "Statistical Inference in Distributed and Privacy-Constrained Settings", verdedigd bij de Delft University of Technology onder supervisie van Aad van der Vaart, Harry van Zanten en Botond Szabó. De jury oordeelt: "Lasse's thesis addresses the important problem of statistical inference in which two problems come together: the data are distributed across multiple locations and communication or privacy constraints limit the flow of information between these locations. This setting arises in applications ranging from hospital networks to autonomous vehicle systems. Within this framework, Lasse derives rigorous minimax lower bounds for signal detection and hypothesis testing, and shows that these bounds are sharp by constructing procedures that attain them. A particularly striking technical contribution is his novel application of the Brascamp-Lieb inequality from probability theory to derive lower bounds in distributed testing, an innovative tool that proved effective across both the communication-constrained and privacy-constrained settings. The thesis is grounded in Le Cam's theory of statistical experiments, which provides a unifying framework for the diverse problems considered. Based on six papers — including publications in the Annals of Statistics, IEEE Transactions on Information Theory, and NeurIPS — the thesis is written not as a mere collection of articles but as a coherent and carefully structured monograph. The jury was deeply impressed by the mathematical depth, originality, and independence of the work, as well as by the breadth of Lasse's research interests and his already remarkable standing in the international community."

Jens Klooster

De tweede winnaar is Jens Klooster met het proefschrift "Robust Inference in Instrumental Variable Models", verdedigd aan de Erasmus University Rotterdam onder supervisie van Chen Zhou. Hierover zegt de jury: "The thesis addresses a long-standing problem in econometrics and applied statistics: how to conduct reliable inference in instrumental variable models when data are contaminated by outliers and instruments may be weak. These two complications had previously been treated separately, but Jens's thesis develops a coherent framework that handles both simultaneously. The first chapter introduces a robust version of the classical Anderson-Rubin test, which is known to handle weak instruments but is sensitive to outliers; the robustified version retains validity under weak identification

while resisting the influence of data contamination. The thesis goes on to develop a robust conditional likelihood ratio test for linear Instrumental Variables models, extending the framework considerably, and concludes with a chapter on the theoretical properties of M-estimators in settings where maximum likelihood and least trimmed squares coincide, grounding the applied contributions in statistical theory. The thesis revisits influential empirical studies in economics, showing how conclusions can shift once robust inference is applied. This compellingly demonstrates the practical relevance of this thesis. The jury appreciated the combination of theoretical rigour and clear applicability at the interface of statistics and econometrics."

Max Welz

De eervolle vermelding was voor Max Welz, met zijn proefschrift "Robust Categorical Data Analysis", verdedigd aan de University Rotterdam onder supervisie van Patrick Groenen en Andreas Alfons. De jury oordeelt "Max's thesis addresses the highly relevant problem of model misspecification in statistical analyses of ordinal and categorical data, a situation in which outliers are not so easy to define and think about. For this problem, Max developed a new class of estimators, C-estimators, which are provably robust against partial misspecification of virtually any model for categorical data, while being efficient and computationally no more costly than standard methods. This striking result required nonstandard proof techniques and a high level of mathematical skill and originality. The thesis further has a clear applied motivation and contains compelling applied work, including a robust estimator of polychoric correlation and a novel changepoint detection method for identifying the onset of careless responding at the person level. The jury was impressed by the breadth and depth of the thesis, as well as by Max's dedication to open science through publicly available R packages and replication materials."

Namens de VVSOR,

- Dr. Ad Ridder, juryvoorzitter Jan Hemelrijk Award
- Prof. dr. Jelle Goeman, juryvoorzitter Willem R. van Zwet Award
- Dr. Sander Scholtus, Secretaris der beide jury's tot 2026
- Dr. Rosanne Turner, Secretaris der beide jury's vanaf 2026
- Prof. dr. ir. Geurt Jongbloed, voorzitter VVSOR

NB In dit nummer van STAtOR vindt u een bijdrage van Jens Klooster over zijn winnende onderwerp. Wij verwachten in de komende nummers artikelen te mogen plaatsen van de andere winnaars.



Instrumentele variabelen, uitschieters en de robuuste statistiek

Jens Klooster

*In dit artikel geef ik een korte samenvatting van mijn proefschrift *Robust Inference in Instrumental Variable Models*. Ik leg uit wat (zwakke) instrumentele variabelen zijn, hoe uitschieters voor problemen kunnen zorgen en hoe deze problemen in mijn proefschrift worden aangepakt met behulp van robuuste statistiek.*

Introductie

Onderzoekers in de sociale wetenschappen zijn vaak geïnteresseerd in causale vraagstukken. Een bekend voorbeeld is het onderzoek van Angrist en Krueger (1991), waarin wordt gekeken of een langere (verplichte) schoolduur leidt tot hogere toekomstige inkomens. Het analyseren van zo'n verband is lastiger dan op het eerste gezicht lijkt. Als je willekeurig mensen op straat vraagt hoe lang zij naar school zijn gegaan en wat hun inkomen is, zul je waarschijnlijk een positief verband vinden. Dit betekent echter niet per se dat een langere schoolduur de oorzaak is van een hoger inkomen. Het is namelijk mogelijk dat dit verband wordt gedreven door andere

factoren, zoals intrinsieke vaardigheden, die zowel de schoolduur als het inkomen beïnvloeden. Dit soort factoren, die vaak lastig te meten, zijn worden daarom ook wel *verstorende variabelen* genoemd. Idealiter zou je dit onderzoeken met een gerandomiseerd experiment, waarbij er ook een controlegroep is van studenten die minder lang naar school gaan. In de praktijk is zo'n experiment echter niet uitvoerbaar, omdat je studenten niet zomaar kunt dwingen om eerder te stoppen met school.

Een manier om dit probleem op te lossen is door gebruik te maken van *instrumentele variabelen*. In de context van de studie van Angrist en Krueger is een instrumentele variabele een variabele die invloed heeft op de schoolduur, maar geen direct effect heeft op het inkomen en niet wordt beïnvloed door verstorende variabelen. Dit zijn sterke aannames en het construeren van zo'n instrumentele variabele is daarom niet eenvoudig.

Angrist en Krueger (1991) realiseerden zich dat zij gebruik konden maken van de Amerikaanse leerplichtwetgeving om een instrumentele variabele te creëren. Leerlingen zijn verplicht om tot hun 18de naar school te gaan, maar starten allemaal in dezelfde maand van het jaar met school. Dit betekent dat sommige leerlingen een aantal maanden ouder zijn dan andere leerlingen in hetzelfde cohort. Oudere leerlingen bereiken eerder de leeftijd van 18 jaar en mogen daardoor eerder stoppen met school. Sommige leerlingen zullen hier gebruik van maken en hierdoor ontstaat variatie in schoolduur die wordt veroorzaakt door de geboortemaand. De geboortemaand, of het kwartaal waarin iemand is geboren, kan dan worden gebruikt als een instrumentele variabele. Deze heeft invloed op de schoolduur, maar wordt niet beïnvloed door verstorende variabelen zoals intrinsieke vaardigheden. Met behulp van deze instrumentele variabele en een daarop gebaseerde schattingsmethode laten Angrist en Krueger zien dat er een positief verband bestaat tussen schoolduur en toekomstige inkomsten.

Zwakke instrumenten

Een later onderzoek van Bound, Jaeger en Baker (1995) wijst op een probleem in de studie van Angrist en Krueger (1991). Zij laten zien dat de geboortemaand maar weinig variatie in de schoolduur kan verklaren. Het instrument is *zwak*. Bound, Jaeger en Baker (1995) laten zien dat zwakke instrumenten problematisch zijn, omdat de gebruikelijke asymptotische resultaten in dat geval niet opgaan. Hierdoor kunnen de schattingen vertekend zijn en tot verkeerde conclusies leiden.

Het zwakke-instrumentenprobleem heeft geleid tot verschillende ontwikkelingen op zowel theoretisch als praktisch gebied. Op theoretisch gebied heeft het geleid tot de ontwikkeling van statistische toetsen die betrouwbaar zijn, ongeacht de sterkte van het instrument, zoals de Anderson-Rubin-toets (Anderson en Rubin, 1949). Op praktisch gebied heeft het ertoe geleid dat empirische onderzoekers eerst toetsen of er sprake is van een zwak of sterk instrument. Dit wordt gedaan door met behulp van een *F*-toets te onderzoeken of er een lineair verband is tussen het instrument en de *endogene variabele* (schoolduur). Een standaard vuistregel is dat de *F*-statistiek groter moet zijn dan 10. Indien het instrument zwak blijkt te zijn, wordt gebruikgemaakt van toetsingsmethoden zoals de Anderson-Rubin-toets.

Uitschieters

Een ander probleem waar empirische onderzoekers vaak mee te maken hebben, zijn uitschieters. Young (2022) repliceerde de resultaten van 30 studies die gebruik maken van instrumentele variabelen en die zijn gepubliceerd in gerenommeerde economische academische tijdschriften. Hij laat zien dat significante resultaten in deze studies vaak gedreven worden door één of enkele uitschieters. Standaard instrumentele-variabele schatters en toetsen zijn niet robuust tegen uitschieters, en één uitschieter kan een arbitrair groot effect hebben op de resultaten.

Om dit te illustreren, bekijken we een studie van Ananat (2011), die het effect van segregatie op welzijn onderzoekt. Ananat merkt op dat een directe regressieanalyse van welzijn op segregatie onbetrouwbaar is vanwege verstorende variabelen. Daarom wordt er in deze studie ook gebruik gemaakt van een instrumentele variabele. Wanneer we echter in Figuur 1 een spreidingsdiagram van de data bekijken, zien we dat er één duidelijke uitschieter aanwezig is. Met deze uitschieter in de data is de *F*-statistiek rond de 20, wat wijst op een sterk instrument. Als we deze uitschieter echter uit de dataset verwijderen en de *F*-statistiek opnieuw berekenen, daalt deze tot rond de 2. Het instrument lijkt dus sterk, maar deze schijnbare sterkte wordt volledig gedreven door één uitschieter.

De wisselwerking tussen uitschieters en zwakke instrumenten maakt een empirische analyse vaak erg moeilijk. In het geval van Ananat (2011) is de uitschieter duidelijk zichtbaar, maar in veel gevallen is het onmogelijk om uitschieters te vinden door simpelweg naar de data te kijken. Idealiter gebruiken we een methode die zowel robuust is tegen uitschieters

De invloedsfunctie is geïntroduceerd door Frank Hampel (mijn academische overgrootvader!) en geeft aan hoeveel invloed een uitschieter kan hebben op een schatter of toets (Hampel, 1974). Een methode is *robuust* wanneer de invloedsfunctie begrensd is. In dat geval kan een uitschieter wel een

Ik ben natuurlijk niet de eerste die zich bezighoudt met de ontwikkeling van methoden die robuust zijn

Figuur 1: Spreidingsdiagram van de endogene variabele (x), de instrumentele variabele (z) en een controle variabele (w) die gebruikt worden in de studie van Ananat (2011).

effect hebben, maar dit effect kan niet onbegrensd zijn. Een eenvoudig voorbeeld van een schatter met een begrensde invloedsfunctie is de mediaan. Als we één datapunt veranderen en bijvoorbeeld heel groot maken, zal de mediaan slechts beperkt veranderen. Dit geldt niet voor het gemiddelde, waarvan de invloedsfunctie onbegrensd is.

De mediaan is een robuuste schatter die bestand is tegen veel uitschieters in de data. Deze robuustheid is vooral waardevol wanneer er daadwerkelijk veel uitschieters aanwezig zijn. In situaties met slechts één uitschieter, zoals in de studie van Ananat (2011), is die robuustheid tegen veel uitschieters minder noodzakelijk. In dat geval leidt het gebruik van de mediaan tot een verlies aan efficiëntie, omdat een groot deel van de 'goede' data niet wordt benut. M-schatters vormen een klasse van schatters die een balans proberen te vinden tussen robuustheid en efficiëntie. Deze schatters, ontwikkeld door Peter Huber (1964), hebben een begrensde invloedsfunctie en zijn daarom robuust, maar beschikken over betere efficiëntie-eigenschappen dan de mediaan.

In mijn proefschrift laat ik zien dat standaard instrumentele-variabele methoden, zoals de Anderson-Rubin-toets, een onbegrensde invloedsfunctie hebben. Dit geldt ook voor de *F*-toets die wordt gebruikt om te beoordelen of een instrument sterk of zwak is, waardoor de uitschieter in de data van Ananat (2011) een groot effect kan hebben op de *F*-statistiek. De toetsen die ik heb ontwikkeld hebben daarentegen een begrensde invloedsfunctie en zijn daarom minder gevoelig voor uitschieters. Deze toetsen zijn gebaseerd op M-schatters en combineren robuustheid met efficiëntie: ze presteren goed wanneer de data uitschieters bevatten, terwijl zij in het geval van schone data niet onderdoen voor standaardmethoden in termen van efficiëntie. Dit maakt de voorgestelde toetsen breed toepasbaar in empirisch onderzoek.

Conclusie

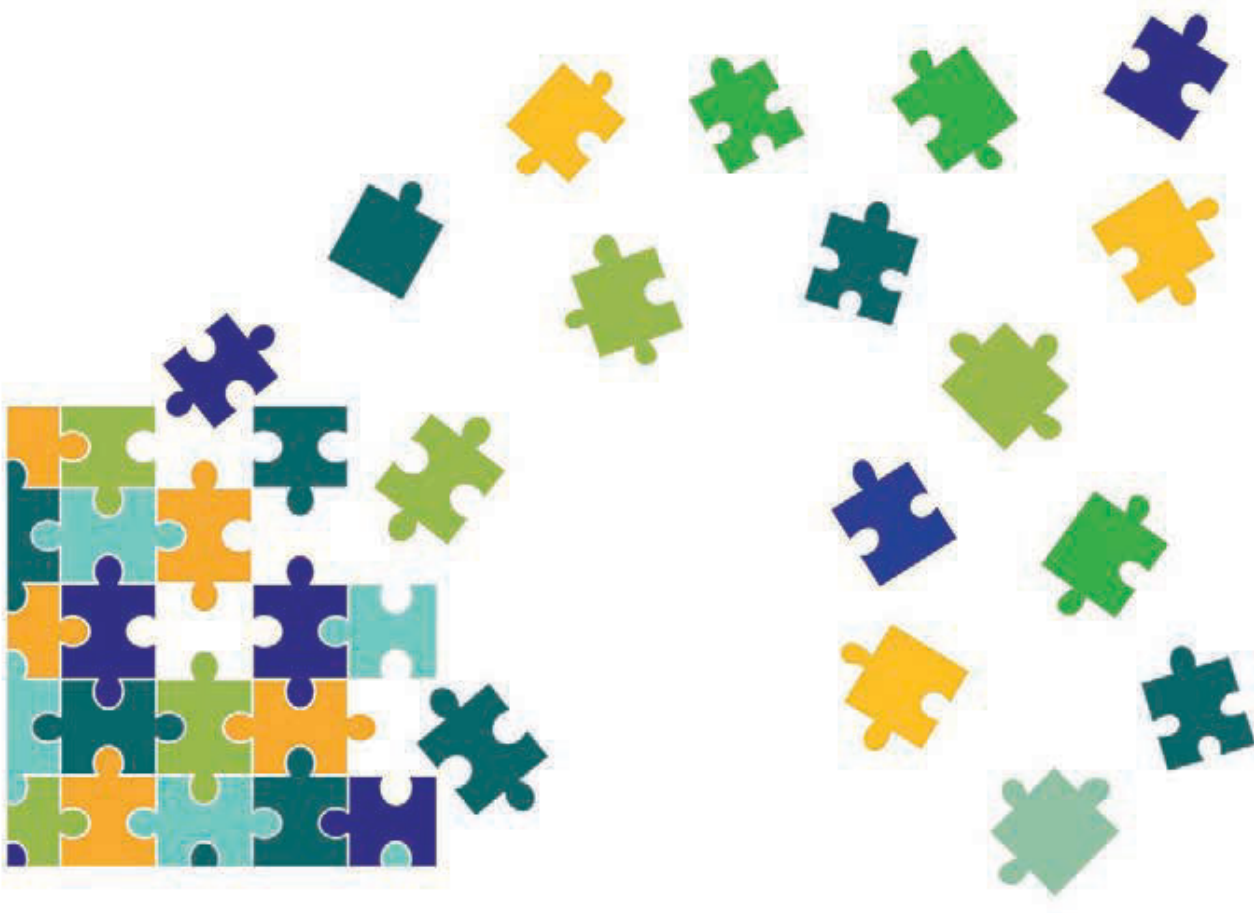
In mijn proefschrift heb ik mij vooral beziggehouden met het ontwikkelen van methoden die robuust zijn tegen zowel zwakke instrumenten als uitschieters. Ik hoop dat duidelijk is geworden hoe belangrijk het gebruik van robuuste methoden is, evenals de ontwikkeling daarvan. Hoewel het probleem van uitschieters in combinatie met zwakke instrumenten grotendeels is aangepakt, blijft er nog veel ruimte voor toekomstig onderzoek. Met name de beschikbaarheid van grote hoeveelheden data zorgt ervoor dat onderzoekers in de sociale wetenschappen steeds vaker gebruikmaken van nieuwe machine-

learningmethoden. Deze methoden kunnen veel grotere datasets verwerken dan traditionele methoden. Hoewel dit veel kansen biedt, brengt het ook nieuwe uitdagingen met zich mee. Zo is de kans groter dat er uitschieters in de data aanwezig zijn en is het, juist door de omvang van de data, moeilijker om deze te identificeren. Ook hier is behoefte aan methoden die automatisch robuust zijn tegen uitschieters. Kortom, er liggen volop kansen voor nieuw en interessant onderzoek.

Literatuur

- E. O. Ananat. "The wrong side (s) of the tracks: The causal effects of racial segregation on urban poverty and inequality". In: *American Economic Journal: Applied Economics* 3.2 (2011), p. 34–66.
- T. W. Anderson en H. Rubin. "Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations". In: *The Annals of Mathematical Statistics* 20.1 (1949), p. 46–63.
- J. D. Angrist en A. B. Krueger. "Does Compulsory School Attendance Affect Schooling and Earnings?" In: *The Quarterly Journal of Economics* 106.4 (1991), p. 979–1014.
- J. Bound, D. A. Jaeger en R. M. Baker. "Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak". In: *Journal of the American Statistical Association* 90 (1995), p. 443–450. DOI: 10.1080/01621459.1995.10476536.
- F. R. Hampel. "The Influence Curve and Its Role in Robust Estimation". In: *Journal of the American Statistical Association* 69 (1974), p. 383–393. DOI: 10.1080/01621459.1974.10482962.
- P. J. Huber. "Robust Estimation of a Location Parameter". In: *The Annals of Mathematical Statistics* 35 (1964), p. 73–101. DOI: 10.1214/aoms/1177703732.
- J. Klooster en M. Zhelonkin. "Outlier robust inference in the instrumental variable model with applications to causal effects". In: *Journal of Applied Econometrics* 39.1 (2024), p. 86–106.
- A. Young. "Consistency Without Inference: Instrumental Variables in Practical Application". In: *European Economic Review* 147 (2022), p. 104–112.

Jens Klooster is universitair docent Econometrie aan de Rijksuniversiteit Groningen. Daarvoor was hij promovendus aan het Econometrisch Instituut van de Erasmus Universiteit Rotterdam, onder begeleiding van Chen Zhou en Mikhail Zhelonkin. Voor zijn proefschrift ontving hij de Willem R. van Zwet Award (2025). E-mail: j.klooster@rug.nl



De magische oplossing voor alle roosterproblemen?

Een mathematisch optimalisatiemodel voor artsenroosters in de zorg

Nienke Hamster

Artsenroosters zijn complexe puzzels. In mijn BSc-afstudeeropdracht Technische Bedrijfskunde heb ik een mathematisch model ontwikkeld dat dit proces moet optimaliseren en automatiseren. Wat ik niet had verwacht, was dat mijn onderzoek zoveel aandacht zou trekken: van regionale en landelijke kranten tot NPO Radio 1 en 2. Daarna werd ik door meer dan 200 (vooral zorg-)organisaties benaderd die alles wilden weten over mijn roosteroplossing. Iedereen leek zich te herkennen in het probleem. Maar is mijn oplossing het ei van Columbus?

Zorg onder druk

We horen het steeds vaker: de zorg staat onder druk. In het nieuws gaat het vrijwel dagelijks over de personeelstekorten, hoge werkdruk en lange toegangstijden. Met de toenemende vergrijzing en zorgvraag en het dalende aantal mensen dat in de zorg werkt, zullen deze problemen alleen maar groter worden.

Toen ik aan mijn afstudeeropdracht begon bij Gelre Ziekenhuizen op de afdeling cardiologie, kwamen diezelfde vraagstukken voorbij. Mijn eerste stap was om te begrijpen hoe de afdeling eigenlijk in elkaar zit. Al snel werd duidelijk dat het artsenrooster de ruggengraat is van de procesorganisatie. Maar dit rooster werd volledig met de hand gemaakt. Dat klinkt misschien niet zo lastig, maar in de praktijk is het een tijdrovend en arbeidsintensief proces,

			Arts 1		Arts 2		Arts 3		Arts 4		Arts 5		Arts 6		Legenda	
Datum	Dag	Nacht	Ochtend	Middag	Ochtend	Middag	Ochtend	Middag	Ochtend	Middag	Ochtend	Middag	Ochtend	Middag	Pa	Pz
Week 14 · 2025																
31-3	Maandag	2	Ka	Ka	Pa	HCK	HCK	Pa	Pz	Pz	V	V	Ka	Pa	Pa	Poli Apeldoorn
1-4	Dinsdag	3	Ka	Ka	Pz	Pz	O	Pa	Pa	HCK	V	V	Pz	O	Pz	Poli Zutphen
2-4	Woensdag	4	Ka	Ka	HCK	Pa	Pz	Pz	Pa	A	V	V	HCK	Ka	Ka	Kliniek Apeldoorn
3-4	Donderdag	5	Ka	Ka	Pz	FB	HCK	Pa	Pa	HCK	V	V	Pa	Pz	HCK	Hartkatheterisatiekamer
4-4	Vrijdag	4	Ka	Ka	MRI	HCK	Pa	Pa	HCK	Pz	V	V	FB	HCK	MRI	MRI Beeldvorming
5-4	Zaterdag	3					Ka	Ka			V	V			A	Administratie
6-4	Zondag	4					Ka	Ka			V	V			O	Onderzoek
Week 15 · 2025																
7-4	Maandag	2	Pz	Pz	Ka	Ka	Pa	HCK	HCK	Pa	Pz	Pa	V	V	FB	Functie Pleper
8-4	Dinsdag	3	O	Pz	Ka	Ka	Pz	Pa	Pa	HCK	HCK	O	V	V	V	Vakantie
9-4	Woensdag	4	Pa	A	Ka	Ka	Pz	Pa	HCK	Pz	Pa	HCK	V	V		
10-4	Donderdag	5	FB	Pz	Ka	Ka	HCK	Pa	Pa	HCK	Pz	FB	V	V		
11-4	Vrijdag	4	HCK	MRI	Ka	Ka	Pa	HCK	Pz	Pa	FB	Pz	V	V		
12-4	Zaterdag	3					Ka	Ka			V	V				
13-4	Zondag	4					Ka	Ka			V	V				

Figuur 1: Voorbeeld van een artsenrooster

en vrijwel altijd met een suboptimaal resultaat. Een vereenvoudigd voorbeeld van zo'n artsenrooster is weergegeven in Figuur 1. Een rooster moet voldoen aan tientallen randvoorwaarden, van wettelijke eisen en ziekenhuisspecifieke regels tot persoonlijke voorkeuren. Elke aanpassing hierin heeft complexe gevolgen voor de rest, en een fout is zo gemaakt. Hier valt dus veel winst te behalen. Dit werd de focus van mijn onderzoek.

Het model

Om het rooster te optimaliseren heb ik een MILP-model (Mixed Integer Linear Programming) opgesteld met een doelfunctie van gewogen genormaliseerde penalties. Alle eisen voor het rooster zijn als randvoorwaarden opgenomen in het model. Hierbij heb ik onderscheid gemaakt tussen harde en zachte randvoorwaarden. Harde randvoorwaarden zijn voorwaarden waar de oplossing aan moet voldoen. Een voorbeeld is het specialisme van een arts: een arts kan alleen taken uitvoeren binnen zijn of haar eigen vakgebied. Het model sluit automatisch taken uit die buiten het specialisme van een arts vallen.

Aan zachte randvoorwaarden willen we graag voldoen, maar oplossingen die hier niet aan voldoen zijn nog steeds toegelaten. Een voorbeeld is reistijd: Gelre Ziekenhuizen heeft meerdere locaties en je wilt liever niet dat een arts op één dag heen en weer rijdt. Om dit te modelleren wordt bijgehouden op hoeveel locaties een arts op een bepaalde dag werkt; idealiter is dit er maar één. Als een arts toch op meerdere locaties werkt, vangt een

slackvariabele de overtreding op. Een slackvariabele is een variabele die de mate van overtreding van een zachte randvoorwaarde bijhoudt: bij geen overtreding is de waarde nul, en bij een overtreding neemt de variabele een positieve waarde aan die de ernst ervan weergeeft. Door deze slackvariabelen op te nemen in de doelfunctie zorgen we ervoor dat overtredingen zoveel mogelijk worden geminimaliseerd. Naast reistijd heeft het model nog meer zachte randvoorwaarden. Zo zijn er taken die artsen liever niet opeenvolgend uitvoeren, zoals meerdere nachtdiensten of weekenddiensten achter elkaar. Daarnaast is er een zachte randvoorwaarde voor taken die niet verplicht zijn, maar wel de voorkeur hebben.

De doelfunctie bestaat uit alle slackvariabelen van de zachte randvoorwaarden. De beste oplossing is dus die met de minste overtredingen. Omdat het schenden van de ene zachte randvoorwaarde minder impactvol is dan voor de ander, hebben we de bijdragen van de zachte randvoorwaarden aan de doelfunctie eerst genormeerd, en daarna weegfactoren gegeven. Voor de normering heb ik de methode van Cardoen (2009) gebruikt, waarbij elke zachte randvoorwaarde afzonderlijk geoptimaliseerd is om de beste en slechtste mogelijke waarden te bepalen, waarna de werkelijke waarde ten opzichte van deze grenzen wordt geschaald en vermenigvuldigd met een weegfactor.

Toen ik de eerste versie van mijn model klaar had, dacht ik er bijna te zijn. Echter, toen ik de gegenereerde oplossing liet zien, bleek het nog niet aan alle randvoorwaarden te voldoen. Die waren bij de

eerdere gesprekken met de betrokkenen niet naar voren gekomen. Uiteindelijk had ik meerdere van deze validatierondes nodig om het model helemaal kloppend te maken voor de situatie.

Het model is geïmplementeerd in AIMMS en opgelost met zowel Gurobi als CPLEX. Het bleek heel goed oplosbaar: in 15 seconden wordt een optimaal rooster gevonden voor twee weken, voor een kwartaal in 5 minuten en voor een half jaar in ongeveer 2 uur. Het model levert niet alleen snel een rooster dat aan alle voorwaarden voldoet, maar het brengt ook inzicht. Doordat het model in seconden een rooster genereert, ontstaat er ruimte om analyses uit te voeren. Zo heb ik met een sensitiviteitsanalyse onderzocht wat het effect is van het versoepelen of volledig loslaten van een bepaalde voorwaarde en of het verschil in uitkomst die voorwaarde daadwerkelijk rechtvaardigt. Ook heb ik specifieke situaties doorgerekend, zoals reductieweken, vakantieperiodes en scenario's waarbij extra artsen beschikbaar zijn. Ten slotte heb ik de uitkomsten van het model vergeleken met de productiebegroting. Doordat je ook een half jaar (of jaar) vooruit kunt plannen, wordt al vroeg inzichtelijk of de productie achterloopt, voorligt of goed op schema zit. Dit alles stelt de cardiologen in staat om niet alleen verder vooruit te plannen, maar ook om datagedreven beslissingen te nemen.

Van het model naar de praktijk

Maar is dit dan de oplossing voor al die roosterproblemen die na de media-aandacht op me af kwamen? Roosters worden vrijwel overal nog met de hand gemaakt. Roostersoftware bestaat wel, maar lijkt in de praktijk nauwelijks (goed) te worden gebruikt. Dat heeft vooral twee oorzaken: een standaardpakket sluit zelden aan op de specifieke voorwaarden van een afdeling, en er is weerstand om de controle uit handen te geven. Het specifiek aanpassen van roostersoftware voor de situatie is kostbaar, en vraagt veel onderhoud - wanneer randvoorwaarden of beleid veranderen, moet het roostermodel mogelijk worden aangepast. Zorginstellingen hebben zelf vaak niet de mensen in huis die dit zelf kunnen.

Mijn aanpak probeerde beide obstakels aan te pakken. Door veel in gesprek te gaan met artsen en mensen op de werkvloer zijn de specifieke randvoorwaarden toegevoegd aan het model. Die gesprekken, het samen doorlopen van de uitkomsten en het bijsturen totdat het klopte, gaven vertrouwen in de aanpak. En het resultaat is dat Gelre Ziekenhuizen het model daadwerkelijk wil gaan gebruiken middels een Python-implementatie. Het ziekenhuis

heeft een Integraal Capaciteitsmanagement team waarin technisch goed onderlegde mensen dit proces kunnen ondersteunen.

Maar is dit dan ook iets voor andere zorginstellingen? Mijn model werkt specifiek voor de cardiologie van Gelre Ziekenhuizen, en dat is precies de kracht van het model, maar ook de beperking. Schaalbaarheid is lastig. Voor elke andere afdeling of zorginstelling moeten eerst alle specifieke randvoorwaarden opnieuw in kaart worden gebracht en in het model worden verwerkt. Daar heb je mensen voor nodig met de juiste kennis en kunde. En als je kijkt naar hoeveel roosterproblemen er in ziekenhuizen spelen, en hoeveel mensen met die kennis we hebben, hebben we er daar lang niet genoeg van.

Daar ligt dus een grote uitdaging: hoe ontwikkel je een betaalbare, klantspecifieke roosteroplossing – in aanschaf, implementatie en onderhoud – die goede resultaten geeft en waarbij de gebruiker het gevoel heeft aan het stuurwiel te zitten? Op het gebied van AI zijn er al mooie ontwikkelingen die hierbij kunnen helpen, maar dat staat nog in de kinderschoenen. Ik verwacht dat die ontwikkelingen snel zullen gaan, en dat over niet al te lange tijd roosteraars met behulp van AI hun ingewikkelde roostervraagstukken om kunnen zetten in een wiskundig model. Daartoe zal nog wel een tool gebouwd moeten worden met een gebruikersvriendelijke interface die door niet-wiskundig onderlegde roosteraars gebruikt kan worden. De tool moet de roosteraar ondersteunen door met verschillende validatiestappen inzicht te geven in hoeverre het roostermodel al doet wat het moet doen.

Is dit dan de magische oplossing voor alle roosterproblemen? Niet helemaal — maar het is wel een mooie stap in de goede richting. Er liggen nog uitdagingen, en de ontwikkelingen op het gebied van AI zijn veelbelovend, maar dat heeft nog tijd nodig. Eerst maar even mijn master afronden.

Literatuur

B. Cardoen. "Operating Room Planning and Scheduling: Solving a Surgical Case Sequencing Problem". Section 5.2.1.2: Towards a multi-objective function, pp. 90–92. Proefschrift. Katholieke Universiteit Leuven, 2009.

Nienke Hamster is student aan de Universiteit Twente. Ze studeert de master Industrial Engineering & Management (ook wel Technische Bedrijfskunde genoemd) en heeft voor haar bachelorafstudeeropdracht onderzoek gedaan bij Gelre Ziekenhuizen. Ze is hierbij begeleid door dr.ir. Gréanne Leeftink en prof.dr.ir. Erwin Hans, die beide werkzaam zijn in het welbekende zorglogistiek-onderzoekscentrum CHOIR van Universiteit Twente. Haar volledige scriptie is te downloaden via <https://essay.utwente.nl/#/108598>.



Het bestaande kunstwerk: 'A Dark Limousine Exit' van Ernst Hazenbroek. Bron: Jelke Bethlehem.

De casus van het Alphen kunstwerk

Op het Raoul Wallenbergplein in Alphen a/d Rijn staat een kunstwerk. Het heet 'A Dark Limousine Exit'. Het is in 1992 gemaakt door de kunstenaar Ernst Hazenbroek. In de zomer van 2018 besloot het gemeentebestuur van Alphen a/d Rijn dat dit kunstwerk moest worden vervangen door een ander, markanter kunstwerk dat de identiteit van de gemeente beter weerspiegelde. De gemeente wilde de inwoners laten beslissen wat er moest gebeuren. Maar hoe doe je dat? De gemeente besloot tot een online peiling die wel wat leek op een online referendum. Alle inwoners van de gemeente kregen de gelegenheid hun mening te geven. Ook de regionale editie van het dagblad AD voerde een online peiling uit. En dan was er nog de peiling op Facebook van de lokale omroep. De drie aanpakken boden een mooie gelegenheid om verschillende manieren van peilen met elkaar te vergelijken.

Drie kunstwerken

Het huidige kunstwerk (dat afgebeeld is op de foto boven deze column) moest dus weg. De gemeente Alphen a/d Rijn kwam met twee voorstellen voor een nieuw kunstwerk. De eerste kandidaat van de gemeente was 'De Poort' van Luc Ticheler. Zie figuur 1a. En de tweede kandidaat was 'Het Drieluik' van Linda Verkaaik. Zie figuur 1b.

De peiling van de gemeente

Uit de drie kunstwerken moesten de inwoners een keuze maken. De gemeente voerde daarvoor een online peiling uit onder alle inwoners van Alphen a/d Rijn. Er werd gepeild van 20 september 2018 tot en met 7 oktober 2018. Om te kunnen stemmen moest je op de website van de gemeente inloggen met je DigiD. Deze aanpak heeft het voordeel dat alleen inwoners van Alphen a/d Rijn (de doelgroep) konden stemmen. De gemeente kreeg via de DigiD, het Burgerservicenummer en de eigen bevolkingsadministratie de beschikking over de adressen van de stemmers. Zo kon worden vastgesteld of iemand wel of niet in de gemeente woonde.

Het bleek mogelijk om met een DigiD meer keren je stem uit te brengen. Raadslid Henk Goes stelde



(a) 'De Poort' van Luc Ticheler.



(b) 'Het Drieluik' van Linda Verkaaik.

Figuur 1: De voorstellen voor een nieuw kunstwerk. Bron: gemeente Alphen a/d Rijn.

hierover een vraag in de gemeenteraad van Alphen. Wethouder Leo Maat gaf toe dat dit kon gebeuren, maar voegde eraan toe dat alleen de eerste stem meetelde. De overige stemmen werden handmatig verwijderd. Je kunt je afvragen of dit in de praktijk allemaal wel goed heeft gewerkt. Een (bewuste of onbewuste) fout is immers zo gemaakt.

Dus alleen inwoners van Alphen a/d Rijn konden stemmen. En ze konden maar één keer stemmen. Deze aanpak lijkt wel wat op een online referendum. Een nadeel van deze aanpak was dat stemmen niet anoniem was. De gemeente kreeg immers de beschikking over het Burgerservicenummer van de stemmers. Het was daarmee technisch mogelijk iemands stem te koppelen aan zijn of haar gegevens in de bevolkingsadministratie. Dus de anonimiteit van de stemmers was niet gegarandeerd.

En als je geen DigiD had? Aanvankelijk was niet duidelijk of die mensen ook konden stemmen. Op 26 september meldde de gemeente echter in de rubriek 'Week in Beeld' in de huis-aan-huis bladen dat die mensen op het gemeentehuis konden stemmen. Ze moesten dan wel een identiteitsbewijs meenemen. Zo kwam het Burgerservicenummer ook beschikbaar voor mensen zonder DigiD en kon de gemeente controleren of de stemmer in Alphen woonde, en niet al gestemd had.

Al met al kon iedere inwoner van Alphen a/d Rijn zijn of haar stem uitbrengen. Daarbij moet wel worden opgemerkt dat het voor inwoners zonder een DigiD een stuk lastiger was om te stemmen. Je moest er immers voor naar het gemeentehuis. Je

kunt vermoeden dat hierdoor de steekproef niet representatief was.

Je komt in aanmerking voor een DigiD als je een Burgerservicenummer hebt. Bij je geboorte krijg je een Burgerservicenummer. Dat zou betekenen dat zelfs jonge kinderen al een DigiD zouden kunnen hebben. Wil je dat die meedoen met de peiling over het kunstwerk? Of moet er een leeftijdsgrens zijn? In een poging meer informatie hierover te krijgen, nam ik contact op met de helpdesk van DigiD. Die beantwoordde mijn vragen niet en verwees me door naar de helpdesk van Logius. En Logius verwees me terug naar DigiD en naar *MijnOverheid*. MijnOverheid verwees mij door naar DigiD. Daarmee was de cirkel rond. Met een kluitje in het riet gestuurd!

Je kunt je afvragen in hoeverre mensen met een DigiD afwijken van mensen zonder DigiD. Om je belastingformulier in te vullen heb je een DigiD nodig. Dus mensen zonder DigiD vullen geen belastingformulier in. Dit zou de representativiteit van uitkomsten van de peiling wel eens kunnen aantasten.

Figuur 2a toont de vraag die werd gesteld in de peiling van de gemeente Alphen. Er waren drie mogelijke antwoorden (corresponderend met de drie kunstwerken). Je moest één kunstwerk kiezen. Het was dus niet mogelijk om aan te geven dat het je niet uitmaakte. En ook was het niet mogelijk om aan te geven dat je helemaal geen kunstwerk wilde. Er was dus sprake van een te beperkte vraagstelling. En de onvermijdelijke uitkomst zou altijd zijn dat er een kunstwerk komt.

Stemmen op een kunstwerk

Voorrang: 85%

Gegevens

Voor welk kunstwerk brengt u een stem uit?

- ☐ 1. De Poort
- ☐ 2. Het Drieluik
- ☐ 3. A Dark Luminous East (bestaand kunstwerk)

[Vorige pagina](#) [Volgende pagina](#)

(a) Alphen a/d Rijn

POLL

Welk kunstwerk vind jij het mooist?

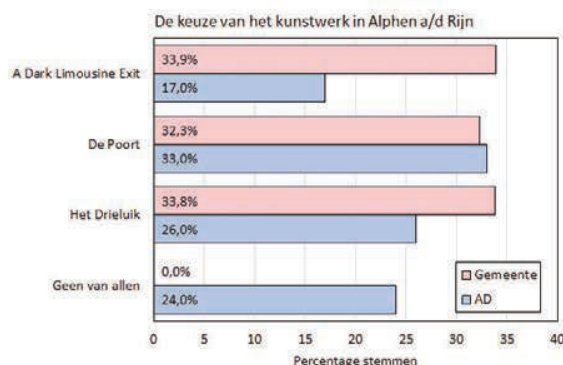
- ☐ De Poort
- ☐ Het Drieluik
- ☐ A Dark Luminous East
- ☐ Geen van allen

[Bekijk resultaten op](#) [Stem](#)

(b) AD

Figuur 2: De vraagstelling in de peiling van de gemeente en van het AD.

Kunstwerk	Gemeente Alphen	AD
A Dark Limousine Exit	33,9%	17%
De Poort	32,3%	33%
Het Drieluik	33,8%	26%
Geen van allen		24%
Totaal	100,0%	100%



Figuur 3: De uitkomsten van de peilingen van de gemeente en het AD.

De peiling van het AD

Op 18 september 2018 kwam de regionale editie van het AD ook met een peiling over het kunstwerk. De peiling stond op de website van het AD tot 20 september 2018. Deze peiling was een typisch voorbeeld van een 1-vraags peiling. Zie figuur 2b. De peiling was niet gebaseerd op een aselechte steekproef, maar op zelfselectie. Iedereen kon meedoen. Daardoor was de peiling niet representatief. In totaal deden 441 personen mee aan de peiling van het AD. Het ligt voor de hand dat dit vooral lezers van het AD waren. Die respondenten vormen vast geen correcte afspiegeling van de Alphenense bevolking.

Iedereen kon meedoen met de AD-peiling. Dus ook personen die niet in Alphen a/d Rijn woonden. Dit 'vervuilde' de steekproef. Het bleek ook mogelijk de vragenlijst meer keren in te vullen. Ik heb de vragenlijst zelf ingevuld op mijn PC thuis, op mijn laptop, op mijn smartphone en op de computer op kantoor. Daarmee loop je als onderzoeker een ernstig risico op manipulatie van de peiling. Het was dus onwaarschijnlijk dat die een representatief beeld opleverde. De journalisten concludeerden daarom dat dit geen journalistiek product was. Er was dan ook geen journalistieke berichtgeving over deze peiling.

De grote vraag was nu of de uitslag van de peilingen een representatief beeld gaf van de mening van de Alphenaren over de kunstwerken. Het is jammer dat het gemeentebestuur tijdens het stemproces probeerde de uitslag te beïnvloeden. Volgens wethouder Kees van Velzen was het bestaande kunstwerk 'foeilijk'. Volgens raadslid Robert-Jan Blom kon het zo naar de 'oudijzerboer'. En de lokale politieke partij Nieuw Elan riep de inwoners op te kiezen voor kunstwerk 'De Poort'.

Uitkomsten

Figuur 3 bevat de uitkomsten van de peilingen van de gemeente Alphen a/d Rijn en van het AD. In de

peiling van de gemeente Alphen kreeg het bestaande kunstwerk ('A Dark Limousine Exit') net de meeste stemmen. Dit kunstwerk kreeg één stem meer dan 'Het Drieluik' (358 tegen 357 stemmen). Het derde kunstwerk, 'De Poort', kreeg de minste stemmen (341), maar de verschillen waren erg klein.

Het is opvallend hoeveel de peiling van het AD afwijkt van de peiling van de gemeente. In de AD-peiling krijgt het bestaande kunstwerk juist de minste stemmen (17%). Dat is maar de helft van het percentage stemmen van de gemeente-peiling. De peiling van het AD was, zoals al eerder gezegd, geen goede peiling.

De peiling van het AD had als extra optie 'Geen van allen'. Respondenten konden dus aangeven dat ze geen van de drie kunstwerken wilden. En dat deed bijna een kwart van de respondenten. Bij de peiling van de gemeente was die mogelijkheid er niet. Je kunt je afvragen hoe de uitslag van de gemeentelijke peiling zou zijn geweest als die mogelijkheid er wel was geweest.

Gelukkig kunnen we vaststellen dat de beïnvloeding van de gemeente geen effect heeft gehad. De negatieve opmerkingen van wethouder Kees van Velzen en raadslid Robert-Jan Blom over het bestaande kunstwerk hebben niet geholpen. In tegendeel, 'A Dark Limousine Exit' kreeg juist de meeste stemmen. Ook de oproep van Nieuw Elan om op 'De Poort' te stemmen had geen effect. Dat kunstwerk kreeg in de peiling van de gemeente juist de minste stemmen. De burgers van Alphen zijn kennelijk mondige burgers die zich niet laten sturen.

Uiteindelijk besloot het gemeentebestuur van Alphen dat het bestaande kunstwerk 'A Dark Limousine Exit' kon blijven staan. Er kwam geen nieuw kunstwerk. Het oude kunstwerk staat er nu (in 2026) nog steeds.

Jelke Bethlehem is expert op het gebied van steekproeven, vragenlijsten en weergave van onderzoeksresultaten.
E-mail: mail@jelkebethlehem.nl



TMLE, Deep Learning en Transformers

Richard Starmans

Dit jaar publiceert The International Journal of Biostatistics (IJB) een themanummer, getiteld "Special issue on Targeted Maximum Likelihood Estimation; Achievements, Evolution and Future Directions", uitgegeven bij De Gruyter Brill. Deze uitgever is naar eigen zeggen al ruim 300 jaar actief, "since the dawn of academic publishing in the seventeenth century". Aanleiding voor dit initiatief is het feit dat precies twintig jaar geleden in hetzelfde tijdschrift het artikel "Targeted Maximum Likelihood Learning" van de hand van Mark van der Laan en Daniël Rubin verscheen, waarin een benadering van statistiek werd bepleit, die in de jaren daarna zou uitgroeien tot een nieuw paradigma voor statistiek, 'statistical learning' en 'causal inference' in observationele data. Inmiddels werden onder meer twee volumieuze tekstboeken over TMLE gepubliceerd, alsmede honderden wetenschappelijke artikelen en diverse softwarepakketten in R en Python. De methode kent ook volop industriële toepassingen, onder meer bij farmaceutische en biotechbedrijven en bij aanvragen betreffende goedkeuring van nieuwe medicijnen door de Amerikaanse Food and Drug Administration (FDA), die toezicht houdt op veiligheid en werkzaamheid van geneesmiddelen.

Inleiding

In STATOR is de afgelopen jaren diverse malen aandacht besteed aan ontwikkelingen binnen Targeted Maximum Likelihood Learning (TMLE). TMLE werd indertijd ontwikkeld tegen de achtergrond van diverse problemen en uitdagingen in de data-analytische praktijk. Men denke aan routinematig gebruik van foutief gespecificeerde parametrische modellen, de erosie van het modelbegrip, de ooit door Ioannides aangekondigde tsunami aan fout-positieve resultaten, 'p-hacking', de replicatiecrisis, de kloof tussen statistical learning and machine learning, de oude vete tussen frequentisten en bayesianen, 'the curse of causality'. Daaraan kan het in sommige kringen vigerende en historisch gewortelde wantrouwen jegens probabilistisch redeneren worden toegevoegd (Starmans, 2020). Vele problemen zijn zo oud als de in de 19e eeuw begonnen probabilistische revolutie zelf en het feit dat sommige tekstboeken incompatibele inzichten en scholen presenteren als een homogeen, verenigd veld lost dit niet op; pogingen rond de eeuwwisseling van onder meer Leo Breiman om tot een prudente boedelscheiding te komen evenmin (Breiman, 2001). De Amerikaanse filosofe Deborah Mayo sprak over de voortwoedende "statistics wars" en daagde statistici uit de oplossing te zoeken in haar "error-statistical method" (Mayo, 2018).

In 2012 in de hoogtijdagen van datascience wijdde STATOR een themanummer aan Big Data, waarin TMLE werd geïntroduceerd als een poging tot verzoening van statistical learning en machine learning, waarbij de teloorgang van statistisch redeneren wordt bestreden door terug te gaan naar de basisprincipes van de inferentiële statistiek. Zo is voor de statisticus kennis van het datagenererende proces essentieel bij de interpretatie van de data, die

worden beschouwd als realisatie van een toevalsvariabele met bijbehorende kansverdeling, de datagenererende verdeling. Om het aantal mogelijke verdelingen waardoor de data kunnen zijn gegenereerd te beperken, dient realistische kennis van het datagenererende proces te worden opgenomen in het statistische model, een verzameling kansverdelingen, die de (benadering van de) ware datagenererende verdeling moet bevatten. In dialoog met de onderzoeker wordt vervolgens een “targeted estimand” gekozen, die zoveel mogelijk aansluit bij de onderzoeksvraag, een mapping van het model naar de parameterruimte behelst en voor een bepaalde verdeling het kenmerk specificeert dat moet worden geleerd van de data. Nadat op deze manier het schattingsprobleem is gedefinieerd wordt een schatter (estimator) geconstrueerd, die leidt tot een schatting (estimate) van de targeted estimand. Deze schatter is zelf een toevalsvariabele, heeft uiteraard een kansverdeling, de sampling distribution, waarvan de spreiding de onzekerheid van de schatter aangeeft. Deze in theoretisch opzicht onomstreden inzichten worden in de data-analytische praktijk dikwijls naar de achtergrond verdrongen. Daarbij vindt dan misspecificatie van het model plaats, dat daardoor niet de ware data genererende functie bevat; fictieve kennis c.q. aannames over de vorm van de stochastische relaties tussen de covariaten leidt tot het arbitrair postuleren van een functioneel verband, waardoor de target estimand niet langer de onderzoeksvraag representeert en het gerapporteerde betrouwbaarheidsinterval waarschijnlijk niet langer het antwoord op die vraag verschaft. Dat dit alles dikwijls niet eens problematisch wordt geacht, vormt een belangrijke factor in de zogenaamde “erosie van het modelbegrip”.

TMLE behelst een secundaire “targeting” stap, die primair de bias-variance tradeoff van de parameter of interest optimaliseert en bezit enkele eigenschappen die de methode zeer geschikt maken voor causal inference in observationele data, waarbij overigens een methodische scheiding wordt gemaakt tussen de statistische target parameter en de causale estimand. We lichten dit toe aan de hand van een bekende causale estimand, ‘the average treatment effect’ (ATE), maar de methode werkt evenzeer bij ingewikkelder causale estimands met bij voorbeeld rechts-gecensureerde en tijdafhankelijke covariaten. Allereerst is TMLE dubbel robuust en garandeert zuivere schattingen indien hetzij het uitkomstmechanisme, hetzij het blootstellingsmechanisme correct is gespecificeerd in het geval van parametrische regressie. Voor ATE betekent dit dat deze zuiver wordt geschat wanneer het uitkomstmechanisme $E(Y|A,X)$

niet consistent is geschat, maar het blootstellingsmechanisme $P(A=1|X)$ wel. Omgekeerd zal bij enkel een consistente schatting van $E(Y|A,X)$, de targeting stap de zuivere schatting bewaren en eindige steekproef bias verwijderen. Zijn beide mechanismen consistent geschat dan is TMLE ook asymptotisch efficiënt. Daarmee integreert TMLE de bekende G-computation en propensity score methoden (zoals de bekende Inverse Probability Weighting techniek). Daarnaast behelst TMLE een substitutieschatter, die robuuster is bij outliers en sparsity. Voor verdere details, waaronder de met TMLE geassocieerde ensembletechniek Super Learner, verwijzen we naar de eerdergenoemde STATOR-publicatie (van der Laan, 2017).

Voorbij regularisatie

Twee jaar geleden werd in STATOR aandacht gevraagd voor Hyper Adaptive LASSO (HAL). In een artikel “Voorbij regularisatie...Highly Adaptive LASSO” werd de merkwaardige en vertraagde opkomst van LASSO geschetst, maar ook de beperkingen ervan benoemd tegen de hier geschetste achtergrond (van der Laan en Starman, 2024). Zo is een pakket als SPSS pas sedert 2022 uitgebreid met een Regularisation-tab, waarachter een drietal technieken schuilgaat: Ridge regression (L2-regularisatie), LASSO (L1-regularisatie) en elastic net regression, dat feitelijk een combinatie van L1 en L2 is. De algoritmen hebben gemeen dat zij een uitbreiding van het aloude lineaire model vormen, waarbij aan de gebruikelijke lossfunction een zogenaamde penalty-term wordt toegevoegd. Deze beoogt de performance op nieuwe data te verbeteren door overfitting te verminderen en dus een betere bias-variance tradeoff te bewerkstelligen, en wil bovendien een vorm van feature selection mogelijk maken in hoogdimensionale datasets. Dit alles geschiedt door zogenaamde “shrinkage”, waarbij sommige β -coëfficiënten krimpen, heel klein worden (ridge regression) of soms ook tot nul worden gereduceerd (LASSO) en de bijbehorende minder relevante variabelen uit het model (de regressievergelijking) kunnen worden verwijderd. Een en ander moet dan leiden tot zuiniger en beter interpreteerbare modellen. Bij ridge regression is de penalty term de som van de gekwadeerde coëfficiënten, bij LASSO gaat het om de absolute waarde hiervan. De grootte van de penalty wordt beheerst door een hyperparameter λ , waarvan de waarde door cross-validation kan worden bepaald.

Juist omdat de inmiddels populaire LASSO geen enorme revisie van GLM betreft, erft de techniek

niet alleen de voordelen, maar ook de beperkingen. Binnen TMLE werd daartoe een non-parametrische least-square schatter LASSO voorgesteld, die de empirische MSE minimaliseert over alle cadlag-functies met begrenzing M op hun sectional variatie norm (van der Laan en Benkeser, 2017). Dit levert een zeer groot model en het is dan ook realistisch aan te nemen dat de ware data genererende functie hierin is bevat. Het minimaliseren van het empirische risico is uiteraard niet nieuw, maar bij HAL gaat het om deze specifiek gedefinieerde klasse van functies met de sectional variatie norm als voornaamste constraint. Bovendien hangt de schatter niet af van lokale smoothness-aannamen, maar respecteert globale smoothness. Het is feitelijk de meest non-parametrische schatter voor regressie en dezelfde principiële definitie kan worden toegepast op andere noties van empirisch risico, voorbij MSE. Neem bij voorbeeld een HAL-estimator van de dichtheid van de data, gegeven een non-parametrische MLE; we maximaliseren dan de empirische log-likelihood over alle dichtheden met een begrenzing op de “sectional” variatie norm. Daarmee construeren we ook HAL-schatters van (conditionele) dichtheden, gemiddelden en behandelingseffecten en feitelijk elke parameter van de verdeling waarin we geïnteresseerd zijn.

Inmiddels is HAL verder uitgebreid tot een krachtige klasse van machine learning-algoritmen voor het schatten van voorspellingsfuncties met opmerkelijke theoretische eigenschappen en prestaties. Een typische implementatie van HAL (bijvoorbeeld voor kleinste-kwadratenregressie) vereist echter het opbouwen van een $(n \times p)$ ontwerpmatrix met een groot aantal p kolommen (zoals $n2^d \times n$, waarbij d de dimensie van de covariantievector is) die de evaluaties van de basisfuncties bij de n waargenomen waarden vertegenwoordigen. Daarom moet een praktische implementatie van HAL altijd een ad-hoc screening uitvoeren, wat de prestaties van de resulterende HAL beïnvloedt. In een nieuwe ontwikkeling, PC-HAL, vervangen we de p basisfuncties door n hoofdcomponentenbasisfuncties, die lineaire combinaties zijn van de oorspronkelijke p basisfuncties die door de hoofdcomponenten worden geïmpliceerd. Bovendien kan, door alleen deze n hoofdcomponentenbasisfuncties bij de n waarnemingen te kennen, de (relatief kleine!) $n \times n$ ontwerpmatrix voor deze hoofdcomponentenbasisfuncties worden berekend zonder de enorme $p \times n$ ontwerpmatrix te hoeven berekenen. Bovendien blijkt dat de theoretische eigenschappen van HAL behouden blijven door deze pc-gebaseerde HAL. Dit vergroot de praktische toepasbaarheid van HAL aanzienlijk,

wat een cruciale rol speelt in TMLE en de super learner. De klassieke principale componenten analyse, die Karl Pearson reeds in 1901 introduceerde en die in machine learning de laatste decennia een sterke comeback heeft gemaakt, biedt dus eveneens uitkomst als praktisch hulpmiddel voor HAL.

Generatieve AI

Uiteraard kunnen we in tijden van generatieve AI niet voorbijgaan aan de betekenis ervan voor TMLE. Het Targeted Learning-framework kan belangwekkende ontwikkelingen in generatieve AI integreren of, als alternatief, kan deep learning in TMLE integreren. Zo is Deepltmle een project dat de krachtige Computational Transformer-architectuur gebruikt om longitudinale TMLE te implementeren voor het schatten van uitkomstverdelingen na interventies onder longitudinale interventies. De standaard ltmle() voor longitudinale causale inferentie gebruikt sequentiële uitkomstregressie om het conditionele behandelingsspecifieke gemiddelde van de klinische uitkomst op elk tijdstip te leren, conditioneel op de waargenomen geschiedenis, en gebruikt vervolgens TMLE om deze regressies te richten op een door de gebruiker gekozen marginaal, behandelingsspecifieke gemiddelde van de klinische uitkomst. ‘Deepltmle’ daarentegen gebruikt temporeel differentieel leren, wat overeenkomt met het gelijktijdig aanpassen van deze regimespecifieke conditionele regressies, dat superieur is aan het sequentieel aanpassen van deze regressies zoals in ltmle.

Als gevolg van deze temporele differentieel leerenadering kan deepltmle een willekeurig aantal tijdstippen verwerken, terwijl ltmle() vastloopt bij een groot aantal tijdstippen. Bovendien maakt de kracht van de Transformer het mogelijk om rekening te houden met rijke en lange (verstorende) geschiedenissen, terwijl standaard machine learning vaak afhankelijk is van specifieke feature-extracties om de dimensionaliteit te beheren. Daarnaast maakt deze computationele implementatie het nu mogelijk om rijke datasets met miljoenen proefpersonen te fitten, wat over het algemeen niet mogelijk is met de meer conventionele R-implementaties zoals ltmle(). De kwaliteit van een Transformer van deze regressies (en daarmee van deepltmle van de causale grootheid) ten opzichte van een meer standaard ML-fit kan echter sterk afhangen van de steekproefomvang, het type en de dimensie van covariaten (bijv. afbeeldingen, NLP, et cetera). Daarom is het plan om de Transformer-algoritmen voor het fitten van deze regressies te integreren in een ensemble-techniek zoals Super Learner, die een grote verza-

meling machine learning-benaderingen bevat. Op deze manier verbetert de Transformer de algehele fit van de Super Learner zonder volledig afhankelijk te hoeven zijn van de Transformer-benadering, en zoals gebruikelijk werkt de TMLE nu op deze door de Transformer verbeterde Super Learner. Een andere ontwikkeling is het uitvoeren van targeting binnen deep learning-algoritmen om hun fit voor causale parameters te targeten.

Hoe dan ook, de methodoloog die de kloof tussen machine learning en inferentiële statistiek wil overbruggen vindt met name in het tijdperk van generatieve AI nieuwe uitdagingen, juist omdat de onzekerheid van b.v. de LLM's gekwantificeerd moet worden en dit zonder statistiek niet goed mogelijk is. Temeer daar in de omvangrijke populaire literatuur over AI eigentijdse klaagzangen van onheilsprofeeten en doemdenkers de boventoon voeren en de angst voor AI, die immers 'slechts probabilistisch' is, onverminderd wordt gecultiveerd (Starmans, 2024; Starmans, 2026).

De auteur dankt Mark J. van der Laan voor commentaar op een eerdere versie van dit artikel.

Literatuur

- L. Breiman. "Statistical modeling: the two cultures". In: *Statistical Science* 16.3 (2001), p. 199–231.
- D. Mayo. *Statistical Inference as Severe Testing: how to get beyond the statistics wars*. Cambridge University Press, 2018.
- R. J. C. M. Starmans. "Eigentijdse klaagzangen en de angst voor AI". In: *Filosofie-Tijdschrift* 36.1 (2026).
- R. J. C. M. Starmans. "Prometheus unbound or Paradise regained: the concept of causality in the contemporary AI-data science debate". In: *Journal of the French Statistical Society* 161.1 (2020).
- R. J. C. M. Starmans. "Telescopen, experimenten en filosofen". In: *Nederlands Tijdschrift voor Natuurkunde* 90.10 (2024).
- M. J. van der Laan. "Targeted Learning. The link from statistics to data science". In: *STATOR* 18.4 (2017).
- M. J. van der Laan en D. Benkeser. "Generally Efficient Targeted Minimum Loss Based Estimator based on the Highly Adaptive Lasso". In: *International Journal of Biostatistics* 13(2) (2017).
- M. J. van der Laan en R. J. C. M. Starmans. "Voorbij regularisatie... Highly Adaptive LASSO". In: *STATOR* 25.2 (2024).

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



Foto:

From the Young Statisticians

Over the past few months, the Young Statisticians have organised two fun statistics-themed events! As part of the VVSOR's Annual Meeting on March 5, we hosted a Statistics Pub Quiz along with the Annual Meeting's dinner. Participants answered statistics and data science-themed trivia, or had to complete light-hearted challenges, such as write a funny on-theme joke, or pose as a certain distribution. We hope you all enjoyed the evening as much as we did.

We are also happy to share that on April 10 we organised a well-attended Statistics Cafe on Industry vs Academia. During the evening, we were joined by 45 participants who heard from five insightful speakers, whose careers have brought them to Academia, Industry or both. We would like to extend a special thank-you to our speakers for taking the time to attend the evening, talk to us about their career progression, and share insights and tips with young statisticians from our network.



Would you like to stay up to date with our future events? Then please sign up for our newsletter by scanning the QR code!



Bron: ChatGPT

Tussen snelheid en begrip: AI in OR

Wat dit betekent voor de OR-praktijk

AI verandert Operations Research fundamenteel, maar niet zoals je misschien denkt. Het vakmanschap van OR wordt belangrijker, niet minder. De vraag is niet óf AI OR-professionals vervangt, maar welk type professional zal floreren: degenen die AI-snelheid combineren met OR-vakmanschap.

Modellen zonder modelleur

Na een bezoek aan de recente NGB-conferentie bleven twee presentaties hangen. Niet omdat ze geruuststellend waren, maar omdat ze een vraag op tafel leggen die we in de OR-praktijk liever ontwijken: maakt AI ons vak beter, of alleen maar sneller?

Een van de presentaties ging over OptiMUS, een systeem van Madeleine Udell. Je beschrijft een optimalisatieprobleem in gewone taal en krijgt werkende MILP-code terug. De resultaten zijn indrukwekkend: een slagingspercentage van 67% op complexe problemen, tegenover zo'n 33% bij naïef gebruik van een taalmodel. Wat vroeger dagen kostte, past nu

in een middag.

Maar wat gebeurt hier eigenlijk? Het systeem *begrijpt* het probleem niet; het *voorspelt* een model dat erop lijkt. En dat verschil doet ertoe. Het verzint randvoorwaarden die er niet horen en, misschien wel gevaarlijker, laat randvoorwaarden weg die essentieel zijn. Het model runt, produceert een oplossing, en zwijgt. Variabelen zien er plausibel uit, maar modelleren potentieel de verkeerde grootte. We debuggen dan niet langer onze eigen modellen, maar die van een AI. Is dat echt een verbetering?

Algoritmen die zichzelf uitvinden

Hiverge presenteerde een andere richting. Via evolutionair zoeken genereert hun AI volledig nieuwe algoritmen. Geen model opstellen en met een bestaande solver oplossen dus, maar de solver genereren. Op het Airbus-Beluga-laadprobleem leverde dat een snelheidswinst op van een factor 10.000.

Indrukwekkend, maar niet gratis. Je krijgt prestaties waar klassieke methoden niet aan tippen, en

code die perfect op jouw probleem lijkt toegesneden. Wat je *niet* krijgt, zijn garanties op kwaliteit van de oplossing of uitleg over waarom de methode werkt. De aanpak van Hiverge levert een heuristiek, geen bewijs van optimaliteit. En “de AI zei het zo” houdt geen stand in een gesprek met een opdrachtgever. Dit kan voldoende zijn voor hoogvolume, repetitieve problemen zoals planning, routing, laden en lossen, maar onvoldoende wanneer een probleem uniek is of wanneer je je conclusies moet vertalen in beleid.

Het echte knelpunt

Beide benaderingen leggen dezelfde ongemakkelijke waarheid bloot: de solver was en is nooit het echte probleem. Het knelpunt zit in het rommelige, menselijke proces eromheen, werken met vage eisen, half begrepen aannames, iteratief gepruts met formuleringen, en eindeloze validaties.

AI versnelt precies het deel dat we al redelijk onder controle hadden: formaliseren en implementeren. Wat het niet oplost, is alles daarvoor en daarna, het verzamelen van relevante randvoorwaarden, validatie, beheer. Precies daar waar projecten falen of juist succesvol worden afgerond.

AI accepteert ambiguïteit zonder protest. Maar ambiguïteit is geen feature, het is een bron van fouten. De vakterm die daar bij hoort is nog altijd *garbage in, garbage out*. Een goed geformuleerd probleem blijft de bottleneck.

Sterker nog: als het model werkt maar jij het niet zelf hebt opgebouwd, wordt validatie geen bijzaak maar hoofdzaak. Want als jij het model al niet volledig begrijpt, wie dan wel?

Wat wordt schaars?

OptiMUS en Hiverge optimaliseren het probleem dat je *geeft*, niet het probleem dat je *hebt*. Als de eisen onvolledig zijn, is het model onvolledig. AI maakt slechte formuleringen sneller, niet beter.

Codegeneratie wordt overvloedig. Snelle heuristieken worden overvloedig. Maar oordeel over probleemformulering, validatie en het opbouwen van vertrouwen bij opdrachtgevers, dat wordt schaars. De OR-beoefenaar die deze tools combineert met inhoudelijk vakmanschap, weet wanneer hij een AI vertrouwt en wanneer niet, en kan een oplossing uitleggen aan een opdrachtgever die er zijn beslissing op baseert. Dat is, als je het zo bekijkt, niet anders dan altijd.

John Poppelaars is oprichter en CEO van Doing the Math
E-mail: john@doingthetmath.nl, [@oratwork.bsky.social](https://twitter.com/oratwork.bsky.social)



Save the date: April 8, 2027 – VVSOR Annual Meeting

De Annual Meeting (AM) van 2027 zal plaatsvinden op 8 april in Utrecht of in Amsterdam. Meer informatie over het programma en over het bijwonen van de Annual Meeting volgt in het volgende nummer van STaTOR en op de VVSOR-website.

Wanted: new member of the Annual Meeting Committee



Als je wilt helpen bij de organisatie van de Annual Meeting, of als je een goed idee hebt voor het thema of voor een onderdeel van het programma van dit jaar, stuur ons dan een e-mail.

annualmeeting@vvsor.nl

