STAtOR

Periodiek van de VVSOR jaargang 24, nummer 3-4, december 2023

STAtOR
Jaargang 24, nummer 3-4, december 2023

STAtOR is een uitgave van de Vereniging voor Statistiek en Operations Research (VVSOR). STAtOR wil leden, bedrijven en overige geïnteresseerden op de hoogte houden van ontwikkelingen en nieuws over toepassingen van statistiek en operations research. Verschijnt 3 of 4 keer per jaar.

Redactie

Joaquim Gromicho (hoofdredacteur), Annelieke Baller, Caroline Jagtenberg, Guus Luijben (eindredacteur), Kerry Malone, Richard Starmans, Gerrit Stemerdink (eindredacteur), Vanessa Torres van Grinsven en Laura Zwep. Vaste medewerkers: Jelke Bethlehem, John Poppelaars, Gerard Sierksma en Henk Tijms.

Kopij en reacties richten aan

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

Bestuur van de VVSOR

Voorzitter: prof. dr. Casper Albers, db@vvsor.nl; Secretaris: Pieter Jongsma, MSc, secretaris@vvsor.nl; Penningmeester: dr. Rebecca Kuiper, penningmeester@vvsor.nl; Algemeen bestuurslid: dr. Marianne Jonker, db@vvsor.nl.

Voorzitters van de secties: dr. Marianne Jonker (Biometrical Section); prof. dr. Albert Wagelmans (Section for Operations Research); prof. dr. ir. Frank van der Meulen (Section Mathematical Statistics); dr. Rebecca Kuiper (Social Sciences Section); dr. Michel van de Velden (Economics Section); dr. Iris Yocarini (Section Data Science); Luise Gummi, BSc (Young Statisticians); dr. Sanne Willems (Section Statistics Communication); dr. Stéphanie van den Berg (Section Statistics Education).

Leden- en abonnementenadministratie van de VVSOR

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

Voor advertenties

Prof. dr. J.A.S. Gromicho, stator@vvsor.nl STAtOR verschijnt in maart, juli en december.

Uitgever

ISSN 1567-3383

Omslagen

We leven in een tijd van omslagen. Niet alleen de jaarlijkse omslag in het weer, maar als dit nummer uitkomt weten we ook welke omslag de Tweede Kamer-verkiezing heeft gebracht. En kijken we verder dan ons kleine land dan kan het haast niet anders dan dat de gebeurtenissen in Oekraïne en het Midden-Oosten voor omslagen zullen zorgen.

Het is dan ook niet vreemd als we overal omslagen menen te ontwaren, ook in ons vakgebied. De ontwikkelingen rond AI zullen ongetwijfeld invloed hebben, al was het alleen maar omdat het grote publiek anders tegen voorspellende modellen zal gaan aankijken. En het zou ons niet verbazen als we binnen niet al te lange tijd een artikel aangeboden krijgen dat bij nauwkeurige bestudering niet een lijfelijke auteur blijkt te hebben, maar geschreven is door ChatGPT.

Dit nummer van STAtOR is, zoals gewoonlijk, weer zeer gevarieerd. Doordat we overal omslagen denken te zien, valt ons op dat enkele artikelen zich bezighouden met omslagpunten: waar ligt de omslag tussen wel of niet spuien van een waterreservoir, het wel of niet vervangen van een filament in een röntgenbuis, het wel of niet aanpassen van bestellingen?

Over betrekkelijk nieuwe toepassingen hebben we twee artikelen: hoe kan de Marine drones optimaal inzetten voor verkenningen en hoe kunnen we met behulp van luchtfoto's een beter zicht krijgen op het aantal recreatiewoningen?

De veelzijdigheid van dit nummer wordt compleet gemaakt door een artikel over JASP, een open-source statistisch pakket dat kan dienen als alternatief voor prijzige commerciële software,

Daarnaast is er verenigingsnieuws, waaronder helaas een In Memoriam waarin Stef Tijs wordt herdacht. En een nummer van STAtOR is ondenkbaar zonder columns, ditmaal vier in totaal, variërend van leerzaam tot verstrooiend,

Eén opmerking tot slot: de omslag van STAtOR is wél 24 jaar hetzelfde gebleven! De keus voor indeling en kleur die Monique van Hootegem begin 2000 voor het proefnummer maakte is een zeer bestendige gebleken!

Wij wensen u veel leesplezier!

De STAtOR-redactie

INHOUD

Gradient descent optimization supports the generation of reservoir operation plans

Indra Marth and Bernhard Becker

The operating of large multi-purpose reservoirs means balancing conflicting use functions: during the dry season the reservoir must be filled to a level such that there is always enough water to supply the water users. During the flood season the filling level must be low enough such that the reservoir can catch a flood wave in order to prevent flood damage. How do reservoir operators know how much water to release when? Usually, there is an operational protocol that tells the operators how much water to release during the different seasons throughout a year. A reservoir optimization model can support the generation of such a release plan. But an optimization model can only provide solutions that are specific to a certain situation given with the input data. To transform such period-of-record solutions into a generic solution, we apply statistical methods and additional post-processing steps. The result is an operational plan that is easy to use, robust and self-correcting as a day-to-day decision support for reservoir operators. This article presents a method to generate a reservoir operation plan by means of gradient descent optimization.

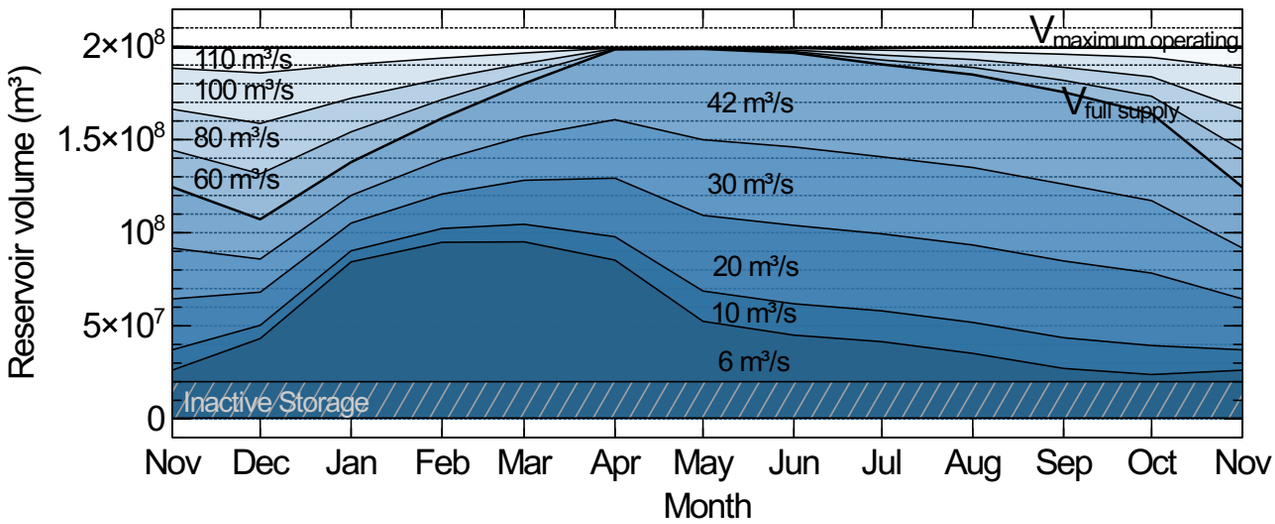


Figure 1: Volume-release plan based on gradient descent optimization (Marth, 2021).

Reduce complexity in reservoir operation by operational protocols

Reservoir functions can be conflicting with each other: operators must maintain sufficient flood space during the wet (flooding) season and at the same time make sure that there will be enough water available for the water supply during the upcoming dry season. In this paper we use the term "reservoir" to describe an artificial lake behind a dam. The dam closes off a valley and impounds the stream that flows through the valley to create a storage. Typical use functions of a reservoir are water supply for domestic, industrial or agricultural use, hydro power generation and flood management. The volume of water stored in a reservoir depends on the hydrological inflow and the release from the reservoir. The reservoir release should at least serve the water demand downstream at all times, but it is often higher than the water demand (Maniak, 2016).

Operators usually decide on the reservoir release by following an operational protocol like for example rule curves. Upper and lower rule curves define the operational range over the year by giving maximum and minimum volume limits for the reservoir operations, respectively. Rainfall and evaporation vary with the season throughout the year, so the volume stored in the reservoir varies with the season. The rule curves ensure that the reservoir is filled during wet period for sufficient water resources before the dry period begins. Furthermore, the rule curves ensure that there is sufficient space left in the reservoir for flood management – during the wet season usually more than during the dry season.

Rule curves alone do not yet give a clear decision support on the amount of water to release, whereas volume-release plans give such detailed information. A volume-release plan shows the reservoir volume between the upper and the lower rule curve divided in different zones (see Figure 1), where each zone contains a respective release. The underlying principle is to release more if there is more water in the reservoir. The boundaries between the zones follow the rule curves throughout the year. The volume-release plan can include a section for flood storage, too.

Volume-release plans have a self-correcting function. Operators query the release for the day from the plan based on the day of the year (x-axis) and the current reservoir volume (y-axis) (see Figure 1). If more water is released than the plan suggests – this can happen if less water flows into the reservoir than expected – the volume will drop to a lower volume segment in the coming days. Here, the assigned release is lower. This should correct for the higher release during the earlier day and stabilize the storage in the reservoir.

Different countries know the principle of volume-release plans. In German literature, volume-release plan is known as *"Lamellenplan"* (see e. g. Lohr, 2001; Maniak, 2016). Numerous reservoirs in Europe are operated with a volume-release plan. In the U.S., a similar concept, the volume release curves (see Fig. 2) is common. Volume release curves and volume-release plans follow the same dependency between volume and release, but display the information in a different way.

Volume-release plans can be used in daily operation, reservoir simulation and hydrological fore-

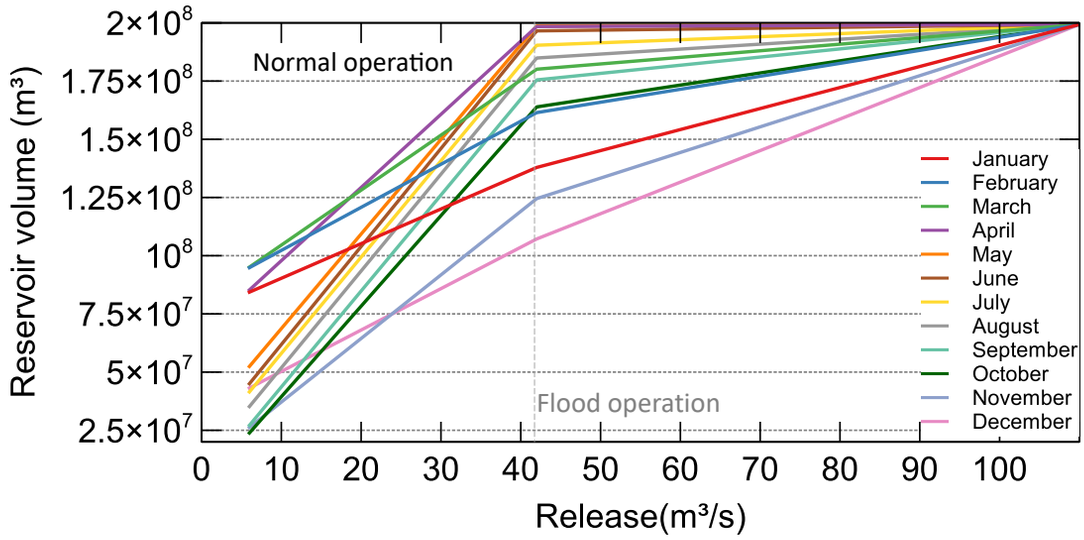


Figure 2: Underlying monthly volume release curves of the volume-release plan (Marth, 2021).

casting. A volume-release plan is self-correcting, intuitive ("release more when you have more"), easy to understand and can even be printed on paper. This makes a volume-based release plan very suitable for reservoirs where complex decision support systems for reservoir operations are not present. The plan can also easily be coded with lookup table logic, so that the concept can also be used in reservoir simulation models that run within advanced decision support systems or in planning studies. Another application for volume-release plans is hydrological forecasting for areas where the reservoir operations are unknown. It has been shown that a volume-release plan can be a method to roughly mimic the reservoir behaviour even if the reservoir in question is operated with a different control method (Marth, 2021).

The volume-release plan as an extension of rule curves

The generation of volume-release plans involves mathematical optimization, post-processing and user choices

The method to determine volume-release plans comprises two basic steps:

- Generation of upper and lower rule curves
- Division of volume zones and assignment of a release value.

For the generation of rule-curves different optimisation techniques can be found in literature (see e. g. Cuvelier et al., 2018). We use a linear gradient descent optimization including goal programming with the interior point optimization solver IPOPT. To resolve conflicting objectives, we use a so-called goal programming approach for the resolution of conflicting goals. This method is implemented in a generic software package, RTC-Tools (Schwanenberg, Becker, and Xu, 2015). RTC-Tools is a free and open source software developed by Deltares designed for control, modelling, and optimization of water systems.

The division of volume zones is not an exact method. It involves several user choices. We have followed Maniak (2016) and Lohr (2001).

Optimzation determines the upper and lower rule curve

The generation of rule curves subdivides again into two steps:

- Optimization of reservoir operations under multiple different hydrological scenarios. The scenarios are either historic or synthetic time series (e. g. based on climate projections) of hydrological inflow into the reservoir. They typically cover one year.
- Post-processing of the results with simple statistical methods.

The reservoir equation

$$V_t = V_{t-1} + I_t + O_t$$

forms an equality constraint of the optimization problem. The volume of stored water V at time step t is the storage volume at the previous time step $t-1$ with hydrological inflow to the reservoir I added and reservoir outflow O subtracted. The precipitation on the reservoir surface and the evaporation are neglected.

Table 1: Operational constraints used as optimization goals

Priority	Goal	Explanation
1	Minimum volume	Operational minimum volume
1	Maximum volume	Operational maximum volume
2	Minimum release	Accounts for water demand (ecology, water supply, industry, dilution for water quality)
2	Maximum release	Acceptable discharge (flood control)
3	Minimize volume / maximize volume	Primary goals for lower / upper rule curve

The optimization task is to find an outflow scheme that meets predefined operational goals as well as possible and obeys the structural limits. The hydrological inflow is an external time series usually provided by a hydrological rainfall-runoff model. The reservoir outflow is the optimization variable. The reservoir's structural limits are stated as constraints, the operational goals are formulated as goals (also referred to as soft constraints) and sorted by priority (see Table 1). The values that correspond to the operational goals presented in Table 1 must be known beforehand and are provided as input to the optimization model.

Not all goals can be met at all times. A goal programming approach allocates resources to the goals with higher priority first. For each priority in Table 1 an optimization problem is solved, where the results of the previous optimization apply as constraints. This way, meeting lower-priority goals is only allowed when it does not worsen the compliance to higher-priority goals. Each priority group narrows the solution space.

Each rule curve is determined in a different run with consecutive optimizations for all priorities. While the goals with priorities 1 and 2 originate from the reservoir's operational protocol, the minimize volume and the maximize volume goals (priority 3) are related to the rule curves (Cuvelier et al., 2018). For the lower rule curve, the optimization gives us the minimal possible volume that still ensures the fulfilment of all goals with higher priority as well as possible under the given hydrological inflow. The maximal possible volume in the reservoir under given hydrological inflow corresponds to the upper rule curve.

The post-processing of the optimization results from multiple scenarios follows Cuvelier et al. (2018):

1. Drop data with extreme conditions, i. e. where the operational limits for the reservoir volume or reservoir release are exceeded. The rule curves should not cover too extreme conditions. For situations outside the rule curves usually special operational protocols, e. g. "release whatever flows in" (pass inflow), apply.
2. Determine the maximum envelope of the reservoir storage from the minimize volume optimization to get minimum rule curve.
3. Determine minimum envelope of the reservoir storage from the maximize volume optimization to get maximum rule curve.
4. If the envelope function is too conservative, quantiles (e. g. the 90 % quantile) can be used instead (user choice).

The volume-release sections involve user choices

Now the operational volume of the reservoir over the year is given with the upper and the lower rule curve, the volume must be divided into sections. This is done as follows (Marth, 2021):

1. Decide how many volume sections the plan shall include (user choice, see also Maniak (2016)).
2. Decide which release to show in the volume-release plan for each section (user choice, usually round numbers).
3. Determine the relation between volume and release for each month in a year. The type of relation is a user choice, a linear relation is the most straightforward.
4. Read the corresponding volume for each release interval (step 2) from the monthly volume-release curves
5. Plot the volume related to the respective release and month.

Figure 3 illustrates the working steps for the generation of a volume-release plan from reservoir rule curves.

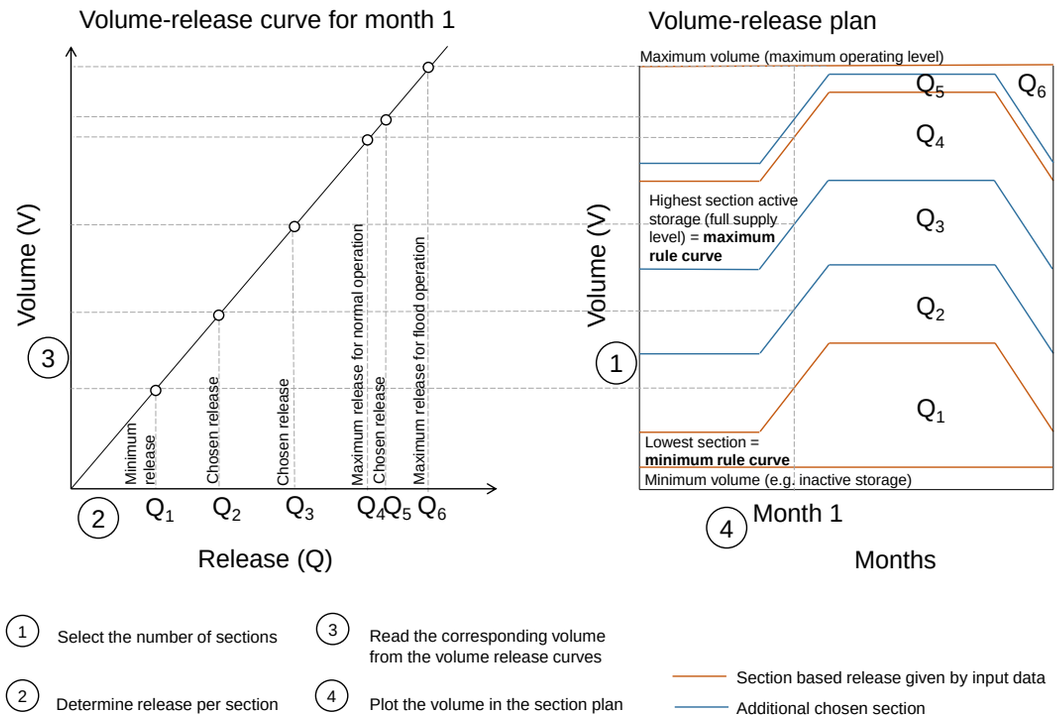


Figure 3: Method to expand rule curves to a volume-release plan (Marth, 2021)

Optimization results require additional post-processing to imitate the correlation between volume and release which is needed for the volume-release plan. Due to hydrological variations, the direct optimized reservoir volume and the corresponding reservoir release are usually not correlated (see also Labadie (2004)) without additional post-processing steps. In the optimization for the same volume, the corresponding release can be different in the same month of different years. A fundamental principle of the volume-release plan presented here, however, is such a correlation. More in general, operational protocols assume that operations are similar for similar situations, whereas the optimization is not restricted to this assumption; in other words: the operational protocol works for all situations, but it is in principle probably never optimal. The self-correction mechanism of the operational plan accounts for this.

To determine a relation between reservoir volume and reservoir release (step 4) we use points on the upper and lower rule curve as well as structural reservoir parameters. That gives us three points and two linear relations, one for normal operation and one for flood operation. The points are (min. rule curve | min. release), (max. rule curve | max. release for normal operation) and (max. volume | max. release for flood operation).

Concluding remarks

This article presents a method for the development of a generic operational protocol for reservoir operations, the volume-release plan. A reservoir optimization model provides a period-of-record solution. This means that the resulting release over time is only an optimal control scheme for the hydrological inflow conditions that have been applied, but for other scenarios it is not optimal. The method involves a reservoir optimization model and the post-processesing of the optimization results from multiple scenarios to go from a period-of-record to a generic solution.

The gradient-descent optimization technique is a core element of this method. A couple of off-the-shelve optimization software solutions are available that support this method; for this study the open source software package RTC-Tools has been used. The data need is fairly managable: some basic reservoir properties and inflow time series for multiple years are needed.

In the light of climate change, we consider it relevant for practical reservoir management to have methods available to generate operational plans

for reservoir management. The operational plans that are currently used will not work anymore in the future, because on the one hand the hydrological conditions change, on the other hand a growing population with a growing need for water puts pressure on water systems. An update of operational protocols can be an adaption measure to climate change and changes in water usage and land use.

References

T. Cuvelier et al. "Comparison between robust and stochastic optimisation for long-term reservoir management under uncertainty." In: *Water resources management* 32 (2018), pp. 1599–1614. DOI: https://doi.org/10.1007/s11269-017-1893-1.

J. W. Labadie. "Optimal operation of multireservoir systems: State-of-the-art review." In: *Journal of water resources planning and management* 130.2 (2004), pp. 93–111. DOI: https://doi.org/10.1061/(ASCE)0733-9496(2004)130:2(93).

H. Lohr. *Simulation, Bewertung und Optimierung von Betriebsregeln für wasserwirtschaftliche Speichersysteme*. Vol. Heft 118. Mitteilungen. Darmstadt: Institut für Wasserbau und Wasserwertschaft, Technische Universität Darmstadt, 2001.

U. Maniak. *Hydrologie und Wasserwirtschaft: Eine Einführung für Ingenieure*. 7. Auflage. Berlin: Springer Vieweg, 2016. DOI: https://doi.org/10.1007/978-3-662-49087-7.

I. F. Marth. "Application of reservoir operation models in hydrological forecasts for regulated catchments." MA thesis. Aachen: RWTH Aachen, Institute of Hydraulic Engineering and Water Resources Management, 2021.

D. Schwanenberg, B. P. J. Becker, and M. Xu. "The Open RTC-Tools Software framework for Modeling Real-Time Control in Water Resources Systems." In: *Journal of Hydroinformatics* 17.1 (2015), pp. 130–148. DOI: 10.2166/hydro.2014.046.

Indra Marth is a researcher/consultant at the operational water management department at Deltares. Indra develops decision support systems for operational water management. She holds a MSc. degree from RWTH Aachen University.
E-mail: indra.marth@deltares.nl

Bernhard Becker works as a researcher/consultant at Deltares and is affiliated as external lecturer with RWTH Aachen University and the University of Duisburg-Essen. At Deltares he is product manager of the RTC-Tools optimization software. His research interests are the optimization of control in water systems, in particular for operational reservoir management and hydropower scheduling.
E-mail: bernhard.becker@deltares.nl

Statistics & Operations Research for seeing the invisible

VVSOR Annual meeting March 21, 2024

Call for abstracts for the Annual Meeting

We are excited to announce the Call for Abstracts for the Annual Meeting of the Dutch Society for Statistics and Operations Research (VVSOR), taking place at In de Driehoek in Utrecht, the Netherlands on March 21, 2024.

This year's theme is *Statistics and operations research for seeing the invisible*. We invite you to **submit your proposals** by **December 12, 2023**.

Submitted research abstracts can be considered for oral presentations – as 'Short Talks' in a parallel session – and for poster presentations. Accepted abstracts will be displayed on the VVSOR website and in STAtOR. Proposals should include:

- Preferred format (individual talk/poster)
- Title of your presentation or session
- An abstract of up to 250 words, describing the content and objectives of your presentation or poster
- A brief biography of the presenter(s), including name(s), affiliation(s), and contact information
- Keywords to describe the main theme, to be chosen from a given list

Please submit your proposal to annualmeeting@vvsor.nl by Dec 12, 2023.

All submissions will be reviewed by our organizing committee, and we will notify you of the outcome by Jan 10, 2024.

We look forward to receiving your proposals and to your participation. Should you have any questions or require further information, please do not hesitate to contact us at annualmeeting@vvsor.nl.

The VVSOR Annual Meeting Committee & the VVSOR Board

Henk Tijms

Column

Illustratie: Pixabay

Pepsi-Cola 349 Fiasco

In de jaren vijftig van de vorige eeuw was het drinken van een flesje Coca Cola een bijzonderheid. Je dronk toen gewoonlijk kraanwater of gazeuse. Goedkoop en gezond. Uit mijn jeugd in die tijd herinner ik me nog goed een promotionele actie van Coca Cola om de consumptie van hun frisdrank te bevorderen. Onder de dop van elk Cola flesje werd één van de 26 letters van het alfabet gedrukt, waarbij elke letter gemiddeld genomen even vaak voorkwam. Als je doppen met alle acht letters van het woord Coca Cola had verzameld, dan won je een prijs.

Naar ik meen, bestond de prijs uit een VIP-arrangement met onder meer een bezoek aan de markante Coca Cola-fabriek die in het toenmalige dorp Sloterdijk was gevestigd bij het oude traject van de spoorlijn Haarlem-Amsterdam. Een traject waarop de treinreiziger ter hoogte van Halfweg niet ontkwam aan een onvergetelijke reclameslogan met de vraag ”Al een vat, kist of krat van de Phoenix gehad?”. De fabriek Phoenix in Halfweg evenals de Coca Cola-fabriek in Sloterdijk verdwenen eind jaren zestig in het kader van de vooruitgang. Hoewel ik destijds graag de Coca Cola-fabriek in Sloterdijk had bezocht, leek het mij niet verstandig om mijn zakgeld te investeren in het verzamelen van de acht benodigde Cola doppen voor een actie die maar beperkte tijd liep. Achteraf inderdaad verstandig, enkele niet-triviale kansberekeningen laten zien dat gemiddeld 98,93 flesjes Cola zouden moeten worden gekocht om acht doppen met de benodigde letters te verzamelen, aannemende dat ruilen met andere verzamelaars niet mogelijk was.

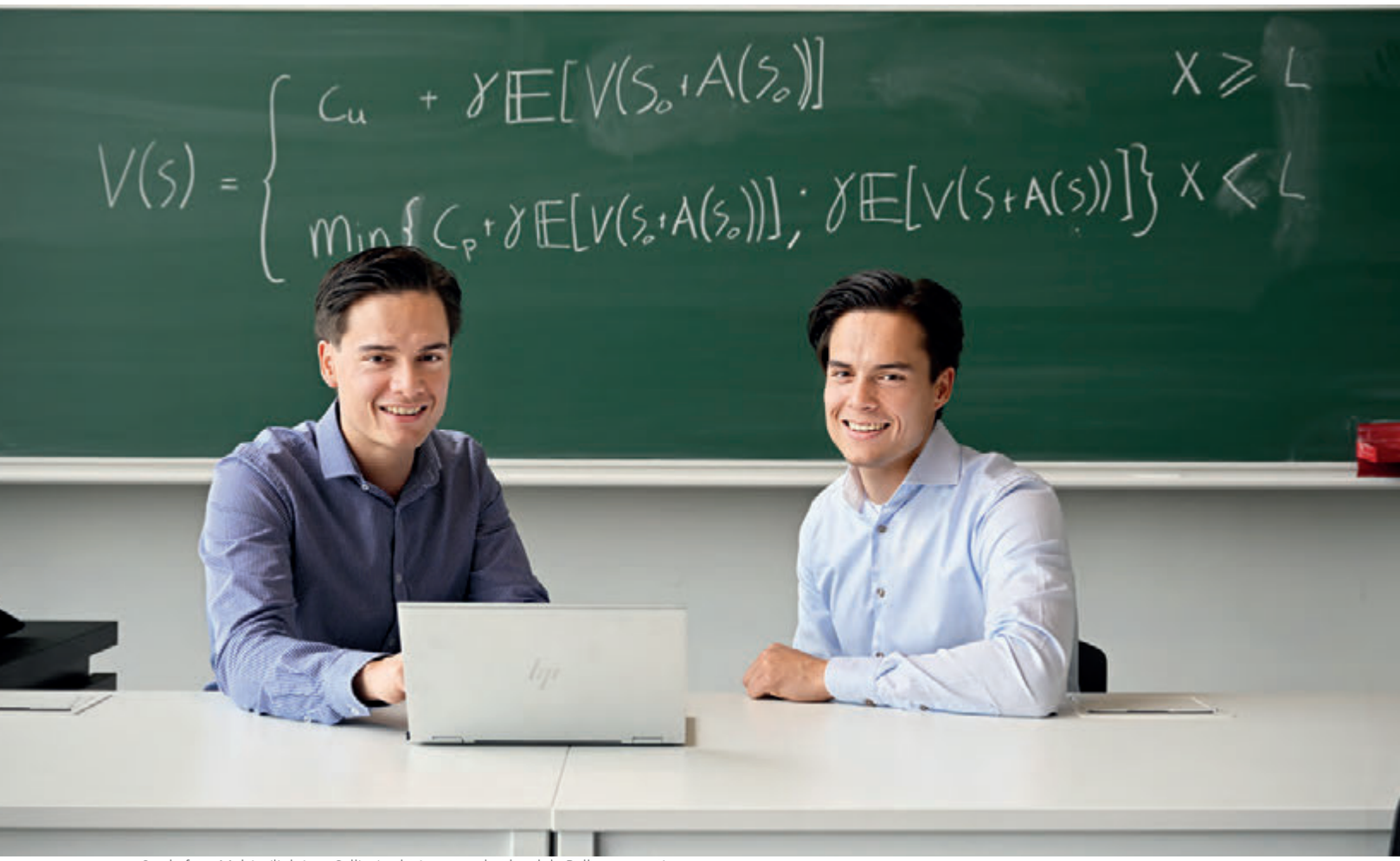
Promotionele activiteiten met winnende getallen of letters zijn niet geheel zonder risico en kunnen vergaande gevolgen hebben. In de jaren negentig probeerde Pepsi-Cola door te breken op de Filippijnse markt waar Coca-Cola dominant was. De Filippijnen vertegenwoordigde een grote frisdrankmarkt, met een snel groeiende bevolking van 62 miljoen mensen. Het management van Pepsi besloot begin jaren negentig een promotiecampagne op te zetten met prijzengeld. Onder de doppen van Pepsi-flesjes

zouden prijswinnende nummers komen te staan met als uitgangspunt veel winnaars van kleine prijzen en twee winnaars van elk een heel grote prijs. Pepsi rekende er op dat de aantrekkingskracht van het prijzengeld veel coladrinkers met een laag inkomen zou doen overstappen van Coca-Cola naar Pepsi-Cola. De Filippijnen was een land dat worstelde met een bescheiden economie en wijdverbreide armoede, en de hoofdprijs werd gezien als een levensveranderende hoeveelheid geld.

Op 5 februari 1992 kondigde Pepsi Philippines aan dat ze nummers, variërend van 001 tot 999, aan de binnenkant van de doppen van Pepsi flessen zouden drukken. Bepaalde nummers konden worden ingewisseld voor prijzen, die varieerden van 100 pesos (ongeveer $4) tot 1 miljoen pesos (ongeveer $40.000) voor elk van twee hoofdprijzen. In 1992 was de hoofdprijs gelijk aan 611 keer het gemiddelde maandsalaris in de Filippijnen. De doppen van de twee flessen met het winnende nummer voor de hoofdprijs werden streng gecontroleerd door Pepsi en kregen een beveiligingscode ter bevestiging mee. De Pepsi aktie was aanvankelijk een groot succes en iedereen van jong tot oud verzamelde de doppen van de Pepsi flesjes. De maandelijkse omzet van Pepsi ging van $10 miljoen naar $14 miljoen en het marktaandeel van 19,4% naar 24,9%. Het winnende nummer werd elke avond op de televisie bekend gemaakt en ruim 70% van de Filippijnse bevolking was dan aan de buis gekluisterd. Begin mei 1992 waren er 51 duizend prijzen verzilverd, waaronder 17 hoofdprijzen, en de campagne werd verlengd tot na de oorspronkelijk geplande einddatum van 8 mei met nog eens 5 weken. Op 25 mei 1992 kondigde het avondnieuwsprogramma van Channel-2 News in Manilla aan dat het hoofdprijsnummer voor die dag 349 was. Toen die avond het winnende nummer 349 op het televisiescherm flitste, konden tienduizenden Filippino's hun geluk niet op. Wat was er gebeurd? Door een computerfout in het productieproces, waren er 800 duizend reguliere flessen uitgebracht met het nummer 349 onder de dop, maar zonder de beveiligingscode. Deze flessendoppen waren theoretisch samen $32 miljard waard. Duizenden Filippino's haastten zich naar de bottelarijen van Pepsi om hun prijzen te claimen. Pepsi reageerde aanvankelijk met de mededeling dat de foutief gedrukte flessendoppen niet de bevestigingscode hadden en daarom niet konden worden ingewisseld. Na een spoedbijeenkomst van directieleden van Pepsi in de Filippijnen en USA op 27 mei, bood het bedrijf 500 pesos ($18) aan houders van de per ongeluk bedrukte flessendoppen aan, als een "gebaar van goede wil". Dit aanbod werd door 486 duizend mensen geaccepteerd, wat Pepsi 8,9 miljoen dollar kostte. Veel woedende 349 flessendop houders weigerden het aanbod van Pepsi te accepteren. Ze vormden een consumentengroep, de 349 Alliantie, die een boycot van Pepsi-producten organiseerde en rally's hield voor de kantoren van Pepsi en de Filippijnse overheid. De meeste protesten verliepen vreedzaam, maar op 13 februari 1993 werden een onderwijzeres en een 5-jarig kind in Manilla gedood door een zelfgemaakte bom die naar een Pepsi-truck werd gegooid en in mei werden drie Pepsi werknemers in Davao gedood door een granaat die in een magazijn werd gegooid. Directieleden van Pepsi ontvingen doodsbedreigingen en maar liefst 37 vrachtwagens van het bedrijf werden omver geworpen, gestenigd of verbrand. De Filippijnse Senaatscommissie voor Handel beschuldigde Pepsi van "grove nalatigheid" en merkte op dat het bedrijf ook betrokken was bij een soortgelijk fiasco in Chili slechts een maand voor het 349 incident.

Ongeveer 22 duizend mensen ondernamen juridische stappen tegen Pepsi; er werden minstens 689 civiele rechtszaken en 5200 strafrechtelijke klachten wegens fraude en misleiding ingediend. In januari 1993 betaalde Pepsi een boete van 150 duizend pesos (ongeveer $5700) aan het ministerie van Handel en Industrie voor het schenden van de goedgekeurde voorwaarden van de promotie. In juni 1996 kende een procesrechtbank de eisers in één van de rechtszaken elk 10 duizend pesos (ongeveer $380) toe als "morele schadevergoeding." Drie ontevreden eisers gingen in beroep en in juli 2001 kende het hof van beroep deze drie eisers elk 30 duizend pesos toe. Pepsi ging tegen deze beslissing in beroep. Het proces zou uiteindelijk het Hooggerechtshof van de Filippijnen bereiken, dat in 2006 oordeelde dat Pepsi niet aansprakelijk is om de bedragen die op de kronen gedrukt staan aan hun houders te betalen en verklaarde de zaak rond het 349 incident als gesloten. Het 349 incident komt ook aan de orde in de 2022 Netflix-documentaire "Pepsi, Where's My Jet?" over een rechtszaak waarin de 20-jarige student John Leonard aanspraak maakt op de ogenschijnlijk onmogelijk te winnen hoofdprijs- een straaljager- in een Pepsi-kansspel nadat hij een maas in de voorwaarden van dit kansspel had ontdekt.

Henk Tijms is emeritus hoogleraar operations research aan de Vrije Universiteit en auteur van diverse leerboeken over operations research en kansrekening. Zijn nieuwste boek is "A First Course in Probability for Computer and Data Science", World Scientific Press, 2023.
E-mail: h.c.tijms@xs4all.nl



Op de foto: Melvin (links) en Collin (rechts), met op het bord de Bellman equation.

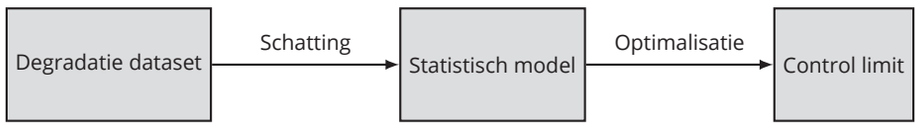
Leren van het optimale onderhoudsmoment

Collin Drent en Melvin Drent

Kapitaalgoederen degraderen wanneer zij in gebruik zijn en zullen op een bepaald moment onderhoud vereisen om nog te kunnen voldoen aan de eisen die hun gebruikers stellen. De degradatieprocessen van zulke kapitaalgoederen hangen af van vele individuele covariaten en zijn daarom dus niet identiek; het optimale moment waarop je onderhoud zou moeten doen is dat dus ook niet. Gelukkig zijn machines in toenemende mate uitgerust met sensoren die allerlei data over de conditie van de machine genereren in real time. In dit artikel, dat gebaseerd is op Drent e.a. (2023) , beschrijven we een methode om aan de hand van die data te leren wanneer het voor iedere specifieke machine optimaal is om onderhoud te plegen.

Introductie

Kapitaalgoederen – zoals de röntgenscanners van Philips of de chipmachines van ASML – zijn machines die een product of dienst leveren. Het is daarom van cruciaal belang dat zulke machines werken wanneer je ze gebruikt, en dat ze niet ongepland stil komen te staan. Dit ongepland stilstaan kan ontzettend veel geld kosten, vooral wanneer het te leveren product of dienst veel waarde heeft. Het meest schrijnende voorbeeld is de chipmachine van ASML waar de kosten van één uur ongeplande stilstand van een chipmachine kunnen oplopen tot 72 duizend euro. Het goede nieuws van deze ongeplande stilstand is dat bijna de helft veroorzaakt wordt door het falen van kritische componenten. Dit zijn componenten die cruciaal zijn voor het functioneren van de machine; als zo een component faalt, dan staat de machine stil. De implicatie hier van is dus dat we deze kritische componenten preventief kunnen vervangen vóórdat ze falen, om zo de ongeplan-



Figuur 1: De standaardmethode om optimale control limits te bepalen.

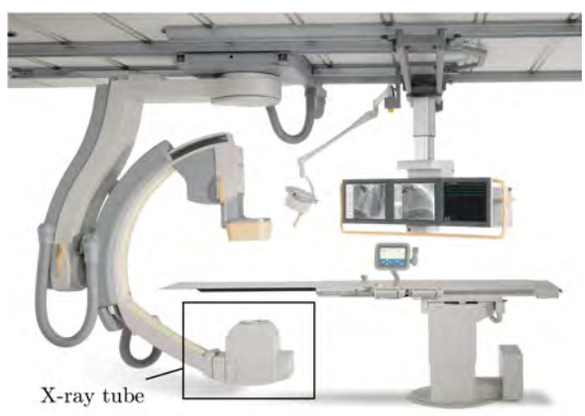
de stilstand te voorkomen. Echter, aangezien deze kritische componenten vaak zelf ook duur zijn, wil je ze niet te vaak preventief vervangen want dat leidt ook tot veel onnodige kosten. Er is dus een trade-off en de vraag is wanneer het optimale onderhoudsmoment is, waarbij je het risico dat je later vervangt dan vereist is, optimaal balanceert met het risico dat je eerder vervangt dan werkelijk nodig is. Componenten degraderen over tijd totdat zij een bepaalde grens bereiken en kapotgaan. Moderne kapitaalgoederen zijn in toenemende mate uitgerust met sensoren die in real time de degradatie kunnen meten. We kunnen dus dit onderhoudsmoment bepalen aan de hand van de degradatie van een component – dit wordt *condition-based maintenance* genoemd – waarbij preventief onderhoud gepleegd wordt zodra de degradatie boven een bepaalde *control limit* komt. De vraag is nu: wat is de optimale control limit gegeven de bovengeschetste trade-off?

De conventionele route naar optimale onderhoudsmomenten

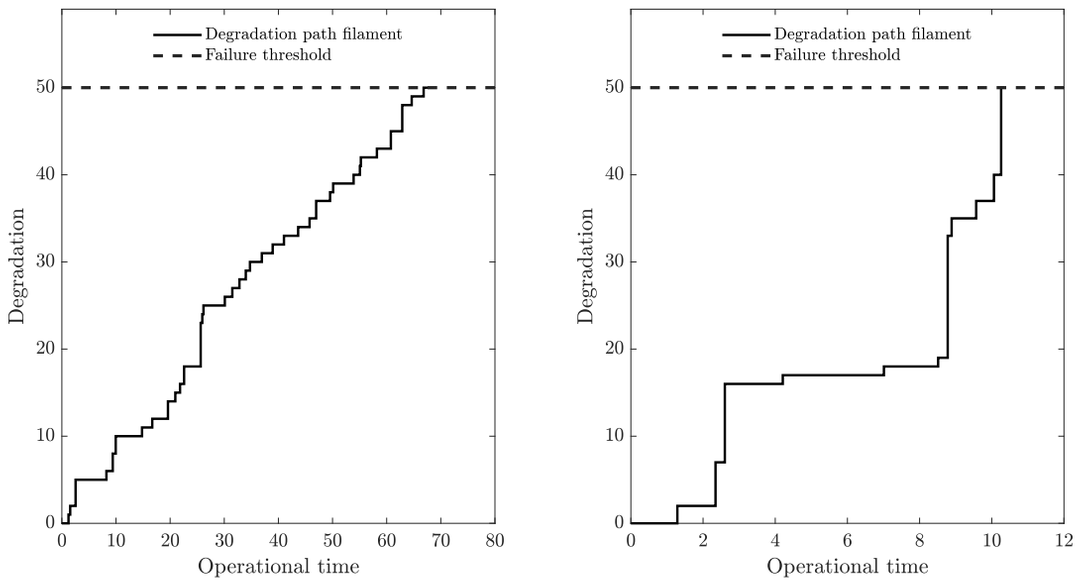
Zowel in de praktijk als in de literatuur is de standaardmethode om de optimale control limit te bepalen een sequentiële procedure bestaande uit twee stappen (zie Figuur 1 voor een schematische illustratie).

In de eerste stap schat men met behulp van een dataset van historische degradatiepaden van een bepaald type component, een statistisch model (meestal een stochastisch proces zoals een *Brownian motion* of een *compound Poisson process*) inclusief de relevante parameters dat het degradatieproces het meest accuraat modelleert. Vervolgens, in de tweede stap, optimaliseert men de control limit met als input het statistisch model en relevante kosten. Deze control limit wordt dan gebruikt om toekomstige onderhoudsmomenten van alle componenten van dat type te plannen. De cruciale aanname die ten grondslag ligt aan deze sequentiële methode is dat alle componenten van hetzelfde type altijd hetzelfde degradatiegedrag vertonen. Met andere woorden: er wordt verondersteld dat componenten van hetzelfde type statistisch niet van elkaar te onderscheiden te zijn. Een direct gevolg hiervan is dat alle componenten dezelfde control limit zullen hebben.

Voor dit onderzoek hebben wij samengewerkt met Philips – een producent van onder andere medische apparatuur. Een van Philips' producten is een zogenaamd iXR systeem. Dit is een röntgenapparaat dat gebruikt kan worden tijdens operaties, waarbij de chirurg in real time röntgenbeelden van de patiënt kan raadplegen (zie Figuur 2 voor een illustratie). Deze röntgenbeelden worden gegenereerd in een röntgenbuis (vierkant in Figuur 2) door een proces waarbij stroom op een filament gezet wordt, welke vervolgens de benodigde röntgenstraling afgeeft. Dit filament is dus de kritische component van het iXR systeem: wanneer het filament kapot is, staat het iXR systeem stil. Filamenten worden steeds dunner naarmate er vaker stroom op gezet wordt volgens een fysisch proces, totdat het zo dun is dat het breekt. Een intern onderzoek bij Philips liet zien dat de degradatie van dit filament gemeten kon worden aan de hand van de weerstand: Hoe meer weerstand het filament geeft bij een bepaalde stroomsterkte, des te meer degradatie het filament heeft.



Figuur 2: Een iXR systeem van Philips. De locatie van de röntgenbuis is weergegeven met een vierkant.



Figuur 3: Twee degradatiepaden van twee filamenten van hetzelfde type die gebruikt zijn onder identieke omstandigheden. Op de x-as staat de tijd sinds het filament geïnstalleerd is, en op de y-as staat de degradatie. Een filament faalt wanneer de degradatie 50 bereikt. Het filament in de linker figuur degradeert minder snel dan het filament in de rechter figuur.

Wij hadden de beschikking over een dataset met 52 degradatiepaden (i.e., het historische degradatieverloop van installatie tot falen) van filamenten van hetzelfde type die gebruikt zijn onder identieke omstandigheden. Figuur 3 laat twee voorbeelden van zulke historische degradatiepaden zien. Iedere keer wanneer een röntgenfoto gemaakt wordt kan de degradatie van een filament toenemen met een stochastische hoeveelheid. Aangezien tussen zulke momenten de degradatie gelijk blijft, vertoont het degradatieproces een trappatroon. Op basis van visuele inspectie – en statistische testen bevestigden dit – kunnen we concluderen dat beiden paden gegenereerd zijn door waarschijnlijk hetzelfde stochastische proces (beiden hebben een vergelijkbaar trappatroon), maar met andere parameterwaardes (de lengte van de traptreden en de hoogtes tussen de traptreden zijn overduidelijk verschillend).

Deze dataset laat dus zien dat de aanname dat alle componenten van een bepaald type statistisch hetzelfde zijn ongegrond is. Het gebruiken van deze methode in het geval dat componenten daadwerkelijk statistisch onderscheidbaar zijn – bijvoorbeeld in het geval van deze filamenten – zal dus leiden tot suboptimale onderhoudsmomenten. In de volgende sectie introduceren wij een nieuwe methode die juist veronderstelt dat het degradatieproces van iedere component uniek is en waarvoor op basis van realtime sensor data het optimale onderhoudsmoment geleerd wordt (zie Figuur 4 voor een schematische illustratie).

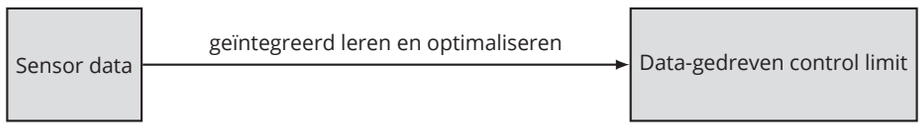
Een nieuwe route naar optimale data-gedreven onderhoudsmomenten

In deze sectie beschrijven we een wiskundig model dat in staat is om voor een specifieke component het leren van onbekende parameters van het degradatieproces met behulp realtime sensor data te integreren met de optimalisatie van het onderhoudsmoment.

We modelleren het degradatieproces $\{X_t\}_{t\geq 0}$ van een component als een compound Poisson process. Dit betekent dat de degradatie op tijdstip t sinds een component geïnstalleerd is geschreven kan worden als:

$$X_t = \sum_{i=1}^{N_t} Y_i, \qquad (1)$$

waar N_t een Poisson process is met parameter λ, en de Y_i's hebben een geometrische verdeling met parameter p. Het mooie aan een compound Poisson process is dat de paden ook – net zoals in Figuur 3 – het typische trappatroon vertonen. In de context van Philips is de interpretatie dus dat $\{N_t\}_{t\geq 0}$ de momenten voorstellen waarop we stroom op een fi-



Figuur 4: Nieuwe methode om optimale data-gedreven control limits te *leren*.

lament zetten, terwijl de Y_i's de toenames in schade voorstellen als gevolg daar van. We nemen aan dat we het degradatieproces continu kunnen observeren door middel van sensoren, maar dat we alleen onderhoud kunnen plegen op geplande momenten $t \in \{0, 1, 2, ...\}$. Dit is in lijn met wat er in de praktijk gebeurt. Bijvoorbeeld, onderhoudsmonteurs bezoeken ziekenhuizen op geplande momenten, bijvoorbeeld eens per half jaar, terwijl men wel van afstand het systeem continu kan monitoren. Wanneer de degradatie op zo een gepland moment groter of gelijk is aan een grens, genoteerd met L, dan is de component kapot en moet er correctief onderhoud plaatsvinden wat c_u kost. Als op zo een gepland moment de degradatie kleiner is dan L, dan kan er gekozen worden om de component preventief te vervangen, wat $c_p (< c_u)$ kost, of de component kan in operatie gelaten worden en opnieuw bekeken worden op het volgende geplande moment. De waarde c_u is in het algemeen veel groter dan c_p omdat c_u de kosten van ongeplande stilstand meeneemt, terwijl preventief onderhoud de machine niet ongepland laat stilstaan.

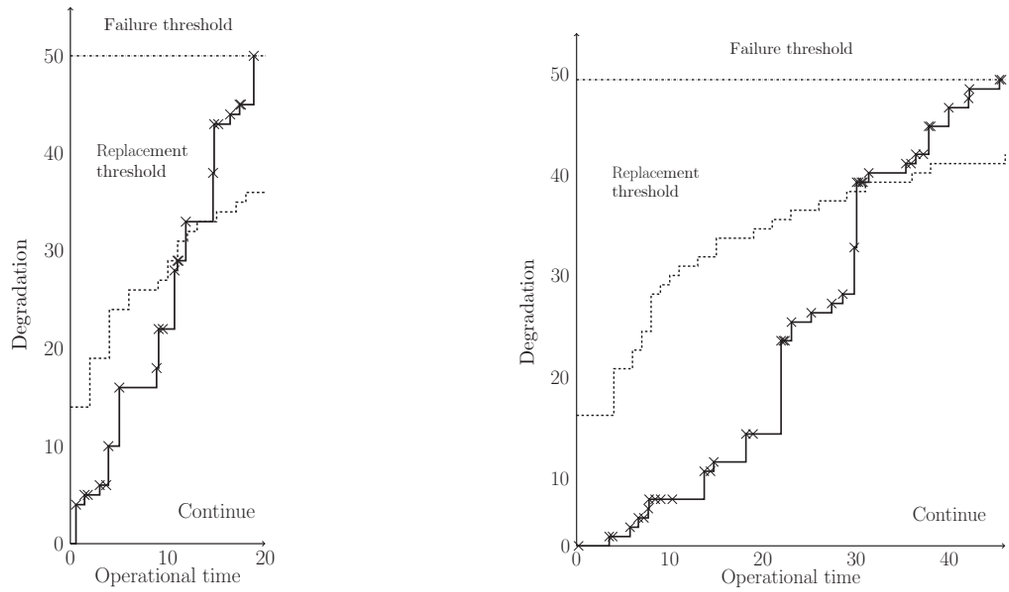
Iedere component is uniek. Dit wil zeggen dat iedere geïnstalleerde component een andere waarde voor λ en p heeft. Deze waardes zijn a-priori onbekend en dienen geleerd te worden van het specifieke degradatiepad dat de component aflegt. We leren deze waardes volgens een Bayesiaanse procedure. Onze kennis van λ en p wordt weerspiegeld door prior verdelingen die we sequentieel updaten met de degradatiedata op basis van Bayes' stelling tot posterior verdelingen, waarbij deze posterior verdelingen een verbeterd beeld geven van onze kennis over λ en p (zie De Heide (2022) voor een excellente uitleg over Bayesiaanse statistiek). We kiezen een conjugate prior verdeling voor zowel λ als p, omdat bij zogenaamde conjugate paren van verdelingen zowel de prior als de posterior verdeling tot dezelfde familie behoren, maar met ge"update parameters. Dit werkt als volgt. Op tijdstip $t = 0$, wanneer een nieuw component geïnstalleerd wordt, dan is λ *gamma* verdeeld met parameters α_0 en β_0, en p is *beta* verdeeld met parameters a_0 en b_0 (dit zijn de prior verdelingen). Stel dat de component nu op tijdstip t sinds de installatie, x degradatie heeft, die veroorzaakt is door n schade momenten (x is de realisatie van X_t en n is de realisatie van N_t), dan is λ gamma verdeeld met parameters $\alpha_t = \alpha_0 + n$ en $\beta_t = \beta_0 + t$, en p is beta verdeeld met parameters $a_t = a_0 + x$ en $b_t = b_0 + n$ (dit zijn de posterior verdelingen op basis van de verkregen data).

Met deze procedure kunnen we in real time, op basis van x, n, en t, de parameters λ en p leren in toenemende mate van nauwkeurigheid. Bovendien kunnen we op ieder gepland onderhoudsmoment de geüpdatete posterior verdelingen combineren met het compound Poisson process (zie vergelijking (1)), om voorspellingen te doen over het verloop van het degradatieproces van deze specifieke component. Door deze procedure te integreren in een Markov beslismodel, kun je voor iedere staat *(x,n,t)* bepalen of het optimaal is om preventief onderhoud te plegen op een gepland onderhoudsmoment. Deze beslissing zal dus afhangen van de sensordata die de individuele component gegenereerd heeft.

Figuur 5 illustreert de optimale onderhoudsmomenten voor twee verschillende filamenten. In deze figuur illustreert de stijgende stippellijn de control limit, zodanig dat preventief onderhoud optimaal is wanneer de degradatie boven deze stippellijn komt. Wanneer we de twee degradatiepaden vergelijken zie je dat het linker filament gemiddeld sneller degradeert ten opzichte van het rechter filament. Je wilt daarom dus ook eerder ingrijpen bij het linker filament op een onderhoudsmoment omdat je een groter risico loopt dat het filament kapot is op het volgende onderhoudsmoment, dan bij het rechter filament. Merk op dat wanneer we volgens de conventionele methode één control limit gekozen hadden voor beide filamenten, je hoogstwaarschijnlijk *te laat* ingegrepen had bij het linker filament en *te vroeg* bij het rechter filament.

De voordelen van geïntegreerd leren en optimaliseren

We sluiten dit artikel af met het kwantificeren van de kostenvoordelen die behaald kunnen worden met ons geïntegreerde model. We hebben een numerieke studie uitgevoerd op basis van de Philips dataset met 52 degradatiepaden waarin we onze methode vergelijken met de conventionele methode. In deze studie selecteerden we telkens x willekeurige degra-



Figuur 5: Twee degradatiepaden (de kruizen markeren de schademomenten) van twee filamenten en de optimale onderhoudsmomenten. Het linker filament wordt vervangen op $t = 12$ met degradatie $x = 29$ en het rechter filament wordt vervangen op $t = 31$ met degradatie $x = 39$.

datiepaden om de prior verdelingen te schatten die noodzakelijk zijn voor de initialisatie van ons model. Daarna berekenden we de onderhoudskosten van ons model zoals die in de praktijk gerealiseerd zouden zijn voor de resterende $52 - x$ degradatiepaden. We herhaalden deze procedure 150 keer om een betrouwbare schatting te krijgen van de gemiddelde onderhoudskosten, met een acceptabel betrouwbaarheidsinterval. Vervolgens vergeleken we deze onderhoudskosten met de onderhoudskosten zoals die onder de conventionele sequentiële methode gerealiseerd zouden zijn. Hierbij bepaalden we eerst puntschattingen voor λ en p op basis van de dezelfde x willekeurige degradatiepaden en berekenden we vervolgens de onderhoudskosten op basis van de resterende $52 - x$ degradatiepaden. In tegenstelling tot ons model, waarbij de control limit wordt bepaald op basis van het degradatieverloop, hanteert de sequentiële methode dezelfde control limit voor alle $52 - x$ degradatiepaden.

Uit de resultaten van onze numerieke studie blijkt overtuigend dat het integreren van leren en optimaliseren aanzienlijke kostenvoordelen oplevert vergeleken met de conventionele sequentiële aanpak. Gemiddeld zouden we met onze methode meer dan 10 procent besparen op de onderhoudskosten van de iXR-systemen van Philips. Deze besparingen zijn te danken aan enerzijds het langer inzetten van de filamenten wanneer we op basis van het realtime degradatie signaal leren dat dit mogelijk is, en anderzijds het proactief, eerder vervangen van filamenten die volgens datzelfde realtime degradatiesignaal sneller lijken te degraderen.

Literatuur

R. De Heide. „Optional stopping met Bayes factors; mogelijkheden en beperkingen". In: *STAtOR* 23.3-4 (2022), p. 22–24.

C. Drent e.a. „Real-time integrated learning and decision making for cumulative shock degradation". In: *Manufacturing & Service Operations Management* 25.1 (2023), p. 235–253.

Collin Drent is als universitair docent verbonden aan de groep Technische Bedrijfskunde van de Technische Universiteit in Eindhoven. Hij promoveerde in 2022 cum laude op zijn proefschrift *Structured Learning and Decision Making for Maintenance*. Voor dit proefschrift ontving hij in 2023 de Willem R. van Zwet Award van de VVSOR. Het onderzoek van Collin richt zich voornamelijk op het modelleren en optimaliseren van data-gedreven beslissingsprocessen in de praktijk.
Website: www.collindrent.com
E-mail: c.drent@tue.nl

Melvin Drent werkt als universitair docent in de groep Technische Bedrijfskunde aan de Technische Universiteit in Eindhoven. Hij promoveerde in 2021 op zijn proefschrift *Stochastic Models of Critical Operations* aan de Universteit van Luxemburg. Momenteel richt Melvin zich voornamelijk op het bestuderen van beslissingsproblemen in diverse domeinen, waaronder voorraadbeheer, onderhoud en gezondheidszorg.
Website: www.melvindrent.com
E-mail: m.drent@tue.nl

Jelke Bethlehem

Column

Adobe Stock

Over verwarrende symbolen in grafieken

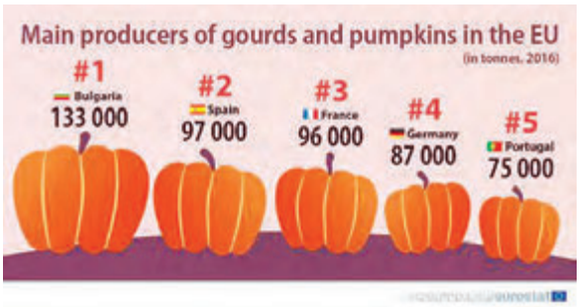
Grafieken kom je overal tegen. Je ziet ze in de media, in wetenschappelijke publicaties en in rapporten van de overheid. En ook het grote publiek ziet regelmatig grafieken langskomen, bijvoorbeeld in de uitkomsten van opiniepeilingen.

Hoe goed zijn de grafieken waarmee journalisten, overheidsfunctionarissen, politici, wetenschappers en het grote publiek te maken krijgen? Een goede grafiek is een krachtig instrument om de verborgen boodschap in de gegevens te onthullen. Maar er zijn ook slecht ontworpen grafieken die je makkelijk op het verkeerde been kunnen zetten. Zelfs een gerenommeerd instituut als Eurostat, het statistisch bureau van de Europese Unie, bezondigt zich soms aan slecht ontworpen grafieken die de toets der kritiek niet kunnen doorstaan. We geven twee voorbeelden van slechte grafieken van Eurostat. In beide gevallen ontstaan de problemen doordat de ontwerper van de grafiek de staven in een staafdiagram vervangt door andere, wat meer speelser, symbolen, zoals pompoenen en bierglazen.

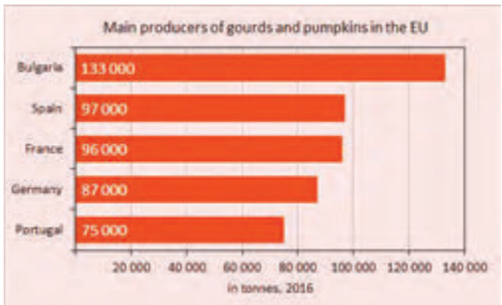
De productie van pompoenen

Op 31 oktober was het Halloween. Dit feest wordt traditioneel gevierd in Angelsaksische landen als Amerika, Engeland, Canada en Ierland. Halloween is echter ook steeds populairder in Nederland en andere Europese landen. Eurostat vond het zelfs nodig om vlak voor Halloween een statistiek over de pompoenproductie in Europa te publiceren. Helaas bevatte het bericht een slechte grafiek. Zie figuur 1a voor deze grafiek.

De grafiek is in feite een staafdiagram. Alleen zijn er geen staven getekend, maar pompoenen. Die pompoenen zijn zo getekend dat de breedte ervan overeenkomt met de pompoenproductie in het bijbehorende land. Zo is de pompoen van Bulgarije ongeveer 1,7 keer zo breed als de pompoen van



(a) Zoals bij Eurostat.



(b) Als staafdiagram.

Figuur 1: De belangrijkste producenten van pompoenen in de EU in 2016.

Portugal. Dat komt overeen met de verhouding van de producties in beide landen: 133.000 is ongeveer 1,7 keer zo groot als 75.000.

Door met symbolen te werken in plaats van met staven, maak je echter een essentiële fout. De ontwerper zag namelijk over het hoofd dat als een pompoen breder wordt, hij ook hoger wordt. En bij het bekijken van de pompoenen kijkt de lezer naar het oppervlak en niet alleen naar de breedte. Het oppervlak van de Bulgaarse pompoen is 2,9 keer zo groot als het oppervlak van de Portugese pompoen. Het verschil in grootte tussen beide pompoenen komt dus niet overeen met het verschil in productie.

Hoe groot is de liegfactor?

Je ziet deze fout wel vaker langskomen bij grafieken waarin symbolen worden gebruikt. Al in 1983 wees de statisticus Edward Tufte in zijn bekende boek 'The Visual Display of Quantitative Information' op de gevaren hiervan. Hij introduceerde zelfs een maat voor de mate waarin een grafiek de waarde van bepaalde verschijnselen verkeerd weergeeft. Dat is de 'liegfactor'. Die krijg je door de waargenomen waarde in de grafiek te delen door de werkelijk waarde. Bij een goede grafiek is de liegfactor gelijk aan 1. In de pompoenengrafiek is de verhouding tussen de Bulgaarse en Portugese pompoen gelijk aan 2,9, terwijl de verhouding tussen de producties 1,7 is. De liegfactor is hier dus gelijk aan 2,9 / 1,7 = 1,7. We kunnen concluderen dat de grafiek de verschillen in productie niet goed weergeeft. Het verschil tussen beeld en werkelijkheid is te groot.

Gewoon een staafdiagram

De eenvoudigste manier om problemen met de omvang van symbolen te vermijden, is helemaal geen symbolen gebruiken, maar gewoon een, simpel staafdiagram. In figuur 1b is de grafiek van Eurostat omgebouwd tot zo'n staafdiagram. Alle (horizontale) staven zijn even dik. Daardoor komen de oppervlaktes van de staven nu wel overeen met de bijbehorende producties. Misschien is deze grafiek wel wat saaier dan de pompoenengrafiek in figuur 1a, maar het is wel een duidelijker en eerlijker grafiek.

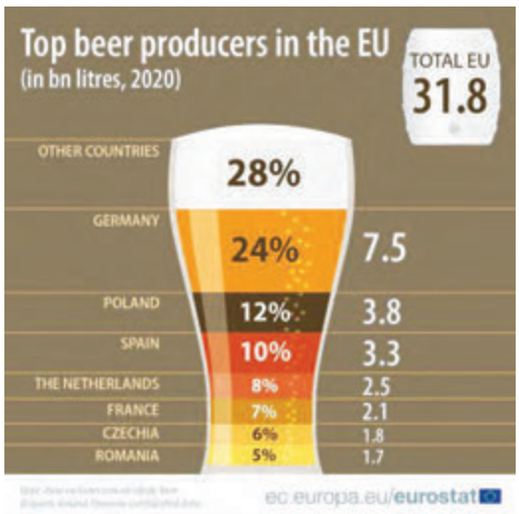
De productie van bier

Op 6 augustus 2021 was het weer 'International Beer Day'. Ter gelegenheid daarvan publiceerde Eurostat een statistiek over de bierproductie in Europa in 2020. Uit deze statistiek bleek dat Duitsland de grootste producent van bier was. De statistiek bevat ook een grafiek met de bierproductie per land. Je zou een staafdiagram verwachten, maar het werd een bierglas. Dat bemoeilijkte de correcte interpretatie van de grafiek behoorlijk. De grafiek is gereproduceerd in figuur 2a.

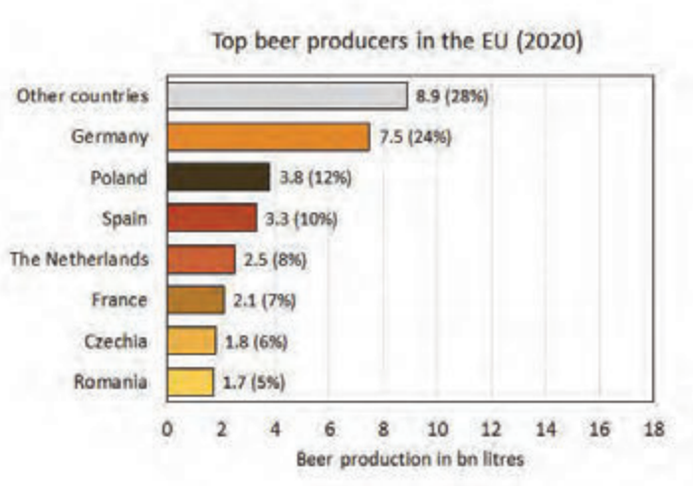
De linker kolom bevat de namen van de landen. 'Other countries' is met 28% de grootste categorie. Kennelijk komt 28% van de bierproductie uit 'Other countries'. Het is niet duidelijk welke landen er in die categorie zitten. Daarom gaat in de grafiek veel informatie verloren. Het lijkt niet logisch deze 'Other countries' prominent bovenaan in de grafiek te zetten. Misschien is het wel beter deze categorie onderaan te zetten. Daar staat de categorie 'Overige' meestal.

Een groot probleem van deze grafiek is dat de bierproductie is weergegeven in de vorm van een bierglas. Daarbij zien de makers over het hoofd dat het glas niet overal even breed is. Daardoor weerspiegelen de (in verschillende kleuren) getoonde oppervlaktes niet de bierproducties in de bijbehorende landen.

Zo leert enig meetwerk dat het grijze oppervlak (van 'Other countries') 10 procent kleiner is dan het oranje oppervlak van Duitsland. Maar de productie



(a) Zoals bij Eurostat.



(b) Als staafdiagram.

Figuur 2: De grootste bierproducenten in de EU in 2020.

van die 'Other countries' is in werkelijkheid ongeveer 1,2 keer zo groot als die van Duitsland. Dus in de grafiek is de bierproductie van 'Other countries' kleiner dan die van Duitsland, terwijl het in werkelijkheid juist andersom is.

Het is helaas niet duidelijk of de witte schuimkop op het glas deel uitmaakt van de oppervlakte van 'Other countries'. Als dit niet het geval is, dan kun je maar beter verwarring voorkomen en hem weglaten.

De vertekening is nog groter als je Nederland vergelijkt met 'Other countries'. Het oppervlak van de 'Other countries' is 5 keer zo groot als het oppervlak van Nederland. De productie van 'Other countries' is in werkelijkheid echter maar 3 keer zo groot als die van Nederland. De in de grafiek weergegeven verschillen zijn dus veel te groot.

Driedimensionale symbolen?

Het kan nog erger. Als de gebruiker de grafiek opvat als een driedimensionaal object (er zit immers diepte in het glas), dan is de gepercipieerde omvang van de onderdelen nog veel groter. En daarmee zullen de verschillen tussen de landen nog veel groter overkomen dan ze al zijn.

Figuur 2b bevat een betere versie van de grafiek. Het is een simpel staafdiagram. De staven zijn horizontaal getekend. Dan is er meer ruimte voor teksten. Dat maakt de grafiek beter leesbaar. Alle staven hebben een andere kleur. Die kleuren zijn overgenomen uit de grafiek van Eurostat. In het staafdiagram zijn die verschillen niet echt nodig. Je kunt de grafiek even goed interpreteren als alle staven dezelfde kleur hebben.

De lengte van de staven komt overeen met de bierproductie in de desbetreffende landen. Er kunnen geen misverstanden ontstaan door verkeerde interpretatie. Ook is het eenvoudiger om de producties van verschillende landen met elkaar te vergelijken. Dit simpele staafdiagram verdient dus duidelijk de voorkeur.

Literatuur

Meer over richtlijnen voor grafieken:

Bethlehem, Jelke (2022), *Het grafiekenboek*. Amsterdam University Press.

Het boek van Tufte: Tufte, Edward (1983), The Visual Display of Quantitative Information. Graphics Press, Cheshire, Connecticut.

Meer over de pompoenproductie in Europa in 2016 kun je vinden in een artikel van Eurostat. Zie:
https://seenews.com/news/bulgaria-biggest-eu-producer-of-gourds-pumpkins-in-2016-eurostat-589025.

Meer over de bierproductie in Europa in 2020 kun je vinden in een artikel van Eurostat. Zie:
https://ec.europa.eu/eurostat/web/products-eurostat-news/-/EDN-20210805-1.

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

In Memoriam Prof. dr. Stef Tijs (1937-2023)

Peter Borm en Hans Peters

De onderstaande tekst zal ook gepubliceerd worden in het komende nummer van het Nieuw Archief voor de Wiskunde.

Op 13 juni 2023 is Stef Tijs overleden. Hij was bijna 86 jaar oud en, behalve de laatste 10 jaren van zijn leven, actief in de wetenschap en vooral in de wiskundige speltheorie. Niet alleen laat hij een groot aantal van bijna 300 publicaties na: hij heeft vooral het vakgebied van de wiskundige speltheorie zowel in de Nederlandse academische gemeenschap als in internationale samenwerkingsverbanden geïntroduceerd en mede ontwikkeld. Hij had 35 promovendi, van wie er velen in de academische wereld werkzaam zijn gebleven en die op hun beurt het vakgebied verder hebben ontwikkeld en verspreid.

Stef Tijs heeft eerst scheikunde en later wis- en natuurkunde gestudeerd aan de Rijksuniversiteit te Utrecht, en colleges gevolgd bij o.a. F. van der Blij, H. Freudenthal, en T.A. Springer. In 1975 is hij gepromoveerd in de wiskunde aan de Katholieke Universiteit Nijmegen onder begeleiding van Arnoud van Rooij en Freddy Delbaen. Zijn proefschrift handelde over oneindige matrix- en bimatrix-spelen, dat wil zeggen twee-persoons niet-coöperatieve nulsomspelen en niet-nulsomspelen, waarbij minstens één speler (aftelbaar) oneindig veel strategieën heeft (Tijs, 1975), een onderwerp dat teruggrijpt op het baanbrekende artikel van Von Neumann (1928) over nulsomspelen en de minimaxstelling. Stefs eerste speltheoretische tijdschriftpublicatie (Tijs, 1976) ging over een ander onderwerp, te weten het verband tussen strategische equivalentie en dominantie binnen coöperatieve spelen, en ook dit onderwerp grijpt terug op

klassiek werk, namelijk het begrip dominantie en daarmee de theorie van stabiele verzamelingen van Von Neumann en Morgenstern (1944).

Gedurende de hierop volgende 40 jaar breidde Stefs wetenschappelijk productie zich gestaag uit over vrijwel alle onderwerpen binnen de speltheorie en over nauw verwante onderwerpen binnen de sociale keuzetheorie en de wiskundige economie. Niettemin kun je in de loop van deze jaren een zekere trend herkennen, namelijk de analyse van coöperatieve spelen, gebaseerd op specifieke economische en besliskundige situaties en problemen. Relatief vroege voorbeelden hiervan zijn de toepassing van speltheorie op problemen van kostenverdeling (Tijs en Driessen, 1986) en op volgordeproblemen (Curiel et al, 1989). Ook introduceerde Stef zijn eigen oplossingsconcept voor coöperatieve spelen, dat hij de τ-value noemde, volgens eigen zeggen naar de tweede letter van zijn voornaam (Tijs 1981, 1987).

Daarnaast was Stef Tijs een echte onderwijsman. Net als zijn wetenschappelijke presentaties bruisten zijn colleges van een aanstekelijk enthousiasme. Hij schreef de borden en zijn overhead slides vol en sprong van het ene onderwerp en detail naar het andere, maar zonder dat het warrig werd. Hij ging er prat op dat hij met beide handen tegelijk kon schrijven en zo in één keer twee samenhangende regels op het bord kon laten verschijnen. Stef heeft 16 collegedictaten geschreven, niet alleen over speltheorie, maar ook bijvoorbeeld over wiskunde voor scheikundigen, en over statistiek (Berendts et al, 1973). Vaak lagen deze collegedictaten ten grondslag aan nieuw onderzoek. Zo was zijn collegedictaat over sociale keuzetheorie tevens de introductie van dit vakgebied in Nederland en vormde het de basis van onderzoek door Stef en een aantal van zijn promovendi. Ook heeft Stef bijgedragen aan de ontwikkeling van de zogenaamde Wageningse Methode voor wiskunde in het voortgezet onderwijs, samen met zijn vriend Leon van den Broek.

Stefs enthousiaste onderwijs en zijn doenermentaliteit hebben in belangrijke mate eraan bijgedragen dat veel van zijn promovendi in de academische wereld werkzaam zijn gebleven, en het vakgebied van de speltheorie in Nederland maar ook in internationale samenwerkingsverbanden op de kaart hebben gezet. Naar aanleiding van het bezoek aan Nijmegen van de latere Nobelprijswinnaar Lloyd Shapley in 1983 startte Stef met een seminarreeks, genaamd Speltheoriedag, later aangevuld met een Sociale Keuzetheoriedag. Vanaf 2005 werd in samenwerking met Italiaanse en Spaanse collega's de jaarlijkse SING (Spain Italy Netherlands meeting on Game theory) gestart, welke later overging in de European Meeting on Game Theory SING.

Het grote belang van Stef Tijs voor de ontwikkeling van de speltheorie in Nederland blijkt ook uit het feit dat er zowel ter ere van zijn 50ste als zijn 65ste verjaardag bundels verschenen met speltheoretisch werk van zijn promovendi (Peters en Vrieze, 1987; Borm en Peters, 2002; zie ook Borm et al, 2005).

Van 1985 tot 1991 was Stef Tijs hoogleraar aan de Katholieke Universiteit Nijmegen, en sinds 1991 aan de Universiteit Tilburg. Daarnaast had hij deeltijdbenoemingen aan de Universiteit Maastricht en aan de Universiteit van Genua, en ontving hij in 2000 een eredoctoraat van de Universiteit van Elche (Spanje). In 2003 werd hij benoemd tot Ridder in de Orde van de Nederlandse Leeuw.

Literatuur

Berendts B., H. Blaauw, B. Harmsen, and S. Tijs (1973) Foutenleer en Statistiek. Agon Elsevier, Amsterdam

Peter Borm and Hans Peters eds (2002) Chapters in game theory in honor of Stef Tijs. Kluwer Acadamic Publishers, Dordrecht

Peter Borm, Hans Peters and Jos Potters (2005) On Stef Tijs. International Journal of Game Theory 33:379-380

Imma Curiel, Giorgio Pederzoli and Stef Tijs (1989) Sequencing Games. European Journal of Operational Research 40:344–351

Hans Peters and Koos Vrieze, eds (1987) Surveys in game theory and related topics. CWI Tract, Amsterdam

Stef Tijs (1975) Semi-infinite and infinite matrix games and bimatrix games. PhD dissertation, University of Nijmegen

Stef Tijs (1976) On S-equivalence and isomorphisms of games in characteristic function form. International Journal of Game Theory 5:209-210

Stef Tijs (1981) Bounds for the core of a game and the τ-value. In: O.Moeschlin, D.Pallaschke (eds.) Game Theory and Mathematical Economics, North-Holland, Amsterdam, 123-132

Stef Tijs (1987) An axiomatization of the τ-value. Mathematical Social Sciences 13:177-181

Stef Tijs and Theo Driessen (1986) Game theory and cost allocation problems. Management Science 32:1015-1028

John von Neumann (1928) Zur Theorie der Gesellschaftsspiele. Mathematische Annalen 100: 295-320

John von Neumann and Oskar Morgenstern (1944) Theory of Games and Economic Behavior. Princeton University Press, Princeton

Peter Borm Universiteit Tilburg

Hans Peters Universiteit Maastricht

Illustratie: Pixabay

Menselijke aanpassingen bij het voorspellen van demand: motivaties en succesfactoren

Robert Fildes, Paul Goodwin en Shari De Baets

Het oordeelkundig aanpassen van de voorspellingen van computersystemen is heel gebruikelijk. Uit enquêtes blijkt dat managers in organisaties gemiddeld ongeveer een derde van de voorspellingen van hun computersoftware veranderen (Fildes en Petropoulos, 2015). Al deze aanpassingsactiviteiten kunnen tijdrovend en duur zijn, maar waarom vindt het plaats? En leidt het tot meer accurate voorspellingen? Om dit na te gaan, hebben we zes datasets verkregen (met 16 *bedrijven/eenheden) waarmee we meer dan* 146.000 *computer-gebaseerde voorspellingen konden bestuderen. Elke voorspelling ging vergezeld van de werkelijke uitkomst en een oordeelkundig aangepaste voorspelling. Het volledige verhaal is te vinden in Fildes, Goodwin en De Baets, 2023.*

Dat er vaak aanpassingen gebeuren aan statistische voorspellingen, is niets nieuws. Het proces is bekend: *demand planners* hebben een keuze tussen een computer-gebaseerde voorspelling accepteren, of deze naar boven of beneden bijstellen. De demand planners zijn echter onderhevig aan een breed scala aan factoren die van invloed zijn op het forecasting resultaat. Zo leidt de complexiteit van het voorspellingsproces tot *biases* (Karelse, 2021). Zo is er bijvoorbeeld overoptimisme (Fildes, Goodwin, Lawrence e.a., 2009), het zien van patronen waar er geen zijn (Harvey, Ewart en West, 1997), of politieke motivaties zoals lagere targets stellen (Mello, 2009).

Een manier om te beoordelen of aanpassingen de voorspellingsprestaties verbeteren, is het gebruik van *Forecast Value Added Analysis* of FVA (Gilliland, 2013). Hiermee worden de prestaties van aangepaste voorspellingen vergeleken met de oorspronkelijke op modellen gebaseerde voorspellingen. Met prestaties bedoelen we niet alleen nauwkeurigheid. Bias – een aanhoudende neiging om te hoge of te

lage voorspellingen te doen – kan net zo schadelijk zijn, zo niet meer, als onnauwkeurigheid. Samen zijn deze twee metingen van cruciaal belang bij het bepalen van de prestaties van de toeleveringsketen. Het is dus belangrijk om te weten onder welke omstandigheden aanpassingen leiden tot betere, of juist slechtere prestaties. Verder is het ook van belang te weten wat juist díe aanpassingen stimuleert.

Hoe effectief zijn oordeelkundige aanpassingen?

We gebruikten een maatstaf genaamd de *Average Relative Mean Absolute Error* (ARMAE) om de nauwkeurigheid van aangepaste voorspellingen te vergelijken met de oorspronkelijke voorspellingen. Details van deze maatstaf zijn te vinden in Davydenko en Fildes, 2013, maar in wezen wordt de verhouding van de absolute fouten (d.w.z. negeren of ze positief of negatief zijn) van de aangepaste en originele voorspellingen genomen en deze over perioden gemiddeld. Tabel 1 toont, voor elke dataset, het percentage SKU's (Stock Keeping Units) waarbij de nauwkeurigheid en bias zijn verbeterd door aanpassing. Over het algemeen had slechts ongeveer de helft van de SKU's een grotere nauwkeurigheid en minder bias als gevolg van manuele tussenkomst. In veel gevallen was de prestatieverbetering niet veel beter dan willekeurig. Er is echter variatie tussen de datasets. Vooral de aanpassingen in het retail-distributiecentrum waren schadelijk voor de nauwkeurigheid en de bias van de voorspellingen.

Spelen de richting en grootte van de aanpassing een rol?

In onze analyses maakten we een opsplitsing tussen opwaartse (verhogen van de cijfer-forecast) en neerwaartse aanpassingen (verlagen van de cijfer-forecast; Perera e.a., 2019). Een opwaartse aanpassing houdt in dat het outputcijfer van de software-forecast omhoog wordt gehaald: dit verhoogde cijfer kan vervolgens leiden tot het bestellen van meer stock. Als dit onnodig is, krijgen we overstock en bij producten met beperkte bewaartijd, afval en bijhorende verwerkingskosten. Als de cijfer-forecast van de software onnodig wordt verlaagd, lopen we het risico op stock-out en verlies van potentiële omzet en lagere service levels omdat producten niet voorradig zijn. Uit onze analyses blijkt dat, hoewel in de meeste gevallen opwaartse aanpassingen iets vaker voorkwamen dan neerwaartse, de prestaties varieerden tussen de datasets. Over het algemeen waren neerwaartse aanpassingen het meest succesvol, waarbij de bias doorgaans met 38% werd verminderd en de nauwkeurigheid met 15% werd verbeterd. Daarentegen hadden opwaartse aanpassingen doorgaans geen effect op de bias, maar verminderde de nauwkeurigheid met 9%. Vervolgens hebben we onderzocht of de grootte van de aanpassingen van invloed was op de prestaties van de voorspellingen. Nogmaals, neerwaartse aanpassingen waren effectiever. Verrassend genoeg was de grootte van positieve aanpassingen niet geassocieerd met de kans op verbeterde prestaties. Voor negatieve aanpassingen verbeterde de nauwkeurigheid, afgezien van die aanpassingen waarbij de aangepaste voorspelling op nul was gezet – deze waren ernstig schadelijk. Als we focussen op die aanpassingsgrootte an sich, weten we dat kleine aanpassingen per definitie alleen kleine wijzigingen in de voorspellingsprestaties kunnen aanbrengen. Grote aanpassingen zijn echter geen garantie voor betere performantie. Zo veroorzaakten opwaartse aanpassingen over het algemeen meer schade aan de prestaties naarmate ze groter waren, terwijl het tegenovergestelde gold voor neerwaartse aanpassingen (met uitzondering van gevallen waarin de uiteindelijke voorspelling op nul was gezet). Over het algemeen is het verhaal consistent in alle datasets, met als enige uitzondering de distributie in de detailhandel.

Waarom beschadigen of verbeteren menselijke aanpassingen de nauwkeurigheid van voorspellingen?

Er zijn twee omstandigheden waarin een menselijke aanpassing de nauwkeurigheid van een voorspelling kan schaden. De eerste is wanneer de aanpassing in de verkeerde richting is, bijvoorbeeld als u een neerwaartse aanpassing maakt terwijl de voorspelling had moeten worden verhoogd. De tweede is wanneer u een aanpassing in de goede richting maakt, maar deze is overdreven. Uit onze analyse bleek dat in totaal 23% van de opwaartse bijstelling en 14% van de neerwaartse bijstelling in de verkeerde richting ging. Doorgaans zorgden deze aanpassingen voor meer dan een verdubbeling van de voorspellingsfouten en het niveau van de bias. Negen procent van de opwaartse bijstellingen was overdreven groot terwijl dit voor slechts 4% van de neerwaartse bijstellingen gold. De juiste richting kiezen en de aanpassingsmaat niet overdrijven, is zeker de sleutel. Maar toch even een positieve noot: sommige bedrijven slagen er wel in effectieve aanpassingen

Tabel 1: Percentage SKU's dat verbeteringen ziet in nauwkeurigheid en bias door aanpassing.

Datasets	Aantal SKUs	% SKUs met verbeterde bias	% SKUs met verbeterde accuraatheid
1. 3 bedrijven: pharma, voeding, huishoudelijke producten	582	40,4	62,4
2. Retail distributie	759	77,5	16,5
3. Pharma	1100	58,3	49,8
4. 3 bedrijven: industrial coating & voeding	1217	44,4	55,1
5. Pharma	45	48,9	53,3
6. Drank (bier)	32	37,5	84,4
Totaal	3735	53,0	52,6

door te voeren: twee van de 16 slaagden erin de accuraatheid, bias en FVA te verbeteren.

Waarom werden aanpassingen gemaakt?

We gebruikten logistische regressie om te beoordelen in hoeverre voorspellers aan het aanpassen waren omdat ze over nuttige extra informatie beschikten die niet beschikbaar was voor het computeralgoritme, of om andere redenen. De beschikbaarheid van extra informatie is ingeschat door de laatste modelcomputervoorspelling te vergelijken met wat er feitelijk is gebeurd. Als de computer-gebaseerde voorspelling bijvoorbeeld 100 eenheden was en de werkelijke 120 eenheden bleek te zijn, suggereert dit dat er mogelijk extra informatie beschikbaar was voor de voorspeller, maar niet voor het model, om tenminste een deel van de extra informatie te verklaren. Bedenk dat het gebruik van dergelijke informatie de belangrijkste rechtvaardiging is voor het maken van aanpassingen. Dus hoe effectief waren de voorspellers bij het benutten van dit potentieel? We ontdekten dat de niet-geobserveerde, maar zeer relevante informatie, een relatief zwakke associatie had met een beslissing om aan te passen. In plaats daarvan pasten voorspellers zich aan op basis van irrelevante of onbetrouwbare informatie. We stelden vast dat de waarschijnlijkheid van aanpassing toenam wanneer de vorige voorspelling werd aangepast en wanneer de computer-gebaseerde voorspelling een neergang voorspelde. De kans op bijstelling naar boven werd vergroot wanneer de eerdere voorspelling naar boven was bijgesteld. Neerwaartse bijstelling werd waarschijnlijker toen de vorige bijgestelde voorspelling beduidend te hoog was geweest. Hoe verhielden deze factoren zich tot de omvang van de aanpassingen? Hoewel er enige variatie was tussen de datasets, waren het opnieuw de 'niet-markt'-factoren, zoals de grootte van de computer-gebaseerde voorspelling of de grootte van de vorige aanpassing die een sterker verband hadden met de grootte van de aanpassing. Waar de ongeziene informatie een duidelijke, zij het zwakke, invloed had op de omvang van de aanpassing, had negatieve informatie (indicatie voor neerwaartse aanpassing) de grotere impact. Er is dus nog veel werk aan de winkel voor voorspellers om zich te oriënteren op de relevante informatie over toekomstige *demand*.

Kunnen we de aanpassingen *debiassen*?

Als voorspellers consistente fouten maken wanneer ze voorspellingen aanpassen, kan het mogelijk zijn om deze fouten te *de-biassen* en zo de nauwkeurigheid van voorspellingen te verbeteren. Als we bijvoorbeeld de fout van een voorspeller kunnen voorspellen met behulp van een regressiemodel, kunnen we deze fout uit hun voorspelling verwijderen. We hebben met verschillende benaderingen

geëxperimenteerd om te zien welke het meest succesvol zou zijn geweest in het verbeteren van de nauwkeurigheid. De grootste nauwkeurigheid werd bereikt door regressiemodellen te gebruiken die de fout in de aangepaste voorspelling voorspelden op basis van vier factoren: de vorige uitkomst, de computer-gebaseerde voorspelling, de vorige aangepaste voorspellingsfout en de vorige aanpassing. Over het algemeen zorgde het verwijderen van de voorspelde fout uit de voorspellingen voor een verbetering van 19% ten opzichte van de computer-gebaseerde voorspelling voor opwaartse aanpassingen en van 29% voor neerwaartse aanpassingen.

Conclusies

Hoewel er tussen bedrijven onvermijdelijk verschillen waren in hun aanpassingsgedrag en het succes van hun interventies, waren er ook veel consistente patronen. We vonden weinig bewijs dat voorspellers effectief gebruik maakten van informatie die niet beschikbaar was voor de computer-gebaseerde voorspelling. In plaats daarvan hadden factoren zoals recente voorspellingsfouten of het feit of er in de voorgaande periode een aanpassing was doorgevoerd een significant effect op de motivatie om bij te sturen. Over het algemeen hadden beoordelingsaanpassingen slechts een kans van ongeveer 50% om de nauwkeurigheid en vertekening van voorspellingen te verbeteren, hoewel we, net als eerdere onderzoeken, ontdekten dat neerwaartse aanpassingen meer kans van slagen hadden. We ontdekten echter dat sommige verbeteringen in nauwkeurigheid kunnen worden bereikt door patronen te identificeren in de manier waarop voorspellers informatie gebruiken en hun voorspellingen te wijzigen om deze voorspelbare fouten te verwijderen. Ondanks onze bevindingen twijfelen we er niet aan dat oordeelsaanpassingen een waardevolle bijdrage kunnen leveren aan voorspellingsprocessen. Maar als ze waarde willen toevoegen aan voorspellingen, moeten voorspellers worden ondersteund door verbeterde organisatorische processen en softwaresystemen. Organisaties zullen waarschijnlijk profiteren van een grotere uitwisseling van relevante informatie tussen afdelingen en het maken van beoordelingen van de waarde en betrouwbaarheid van die informatie. Het ontwerp van ondersteuningssystemen voor voorspellingen moet voorzieningen omvatten die helpen bij het beoordelen, zoals feedback, het verstrekken van informatie over de effecten van analoge gebeurtenissen en ontleding, zodat complexe beoordelingen kunnen worden opgesplitst in eenvoudigere delen. Verbeteringen als deze zouden voorspellers in staat moeten stellen effectiever gebruik te maken van zowel hun tijd als van beschikbare informatie. Als gevolg hiervan zouden hun interventies beter gericht moeten zijn en beter afgestemd op de tekortkomingen van computer-gebaseerde voorspellingen.

Literatuur

A. Davydenko en R. Fildes. „Measuring Forecasting Accuracy: The Case of Judgmental Adjustments to SKU-level Demand Forecasts". In: *International Journal of Forecasting* 29.3 (2013), p. 510–522.

R. Fildes, P. Goodwin en S. De Baets. „Forecast Value Added in Demand Planning". Working paper. 2023.

R. Fildes, P. Goodwin, M. Lawrence e.a. „Effective Forecasting and Judgmental Adjustments: An Empirical Evaluation and Strategies for Improvement in Supply-Chain Planning". In: *International Journal of Forecasting* 25.1 (2009), p. 3–23.

R. Fildes en F. Petropoulos. „Improving Forecast Quality in Practice". In: *Foresight* 36 (2015), p. 5–12.

M. Gilliland. „FVA: A Reality Check on Forecasting Practices". In: *Foresight* 29 (2013), p. 14–18.

N. Harvey, T. Ewart en R. West. „Effects of data noise on statistical judgement". In: *Thinking & Reasoning* 3.2 (1997), p. 111–132.

J. Karelse. „Mitigating Unconscious Bias in Forecasting". In: *Foresight* 61 (2021), p. 5–14.

J. Mello. „The Impact of Sales Forecast Game Playing on Supply Chains". In: *Foresight* 13 (2009), p. 13–22.

H. N. Perera e.a. „The Human Factor in Supply Chain Forecasting: A Systematic Review". In: *European Journal of Operational Research* 274.2 (2019), p. 574–600.

Robert Fildes is oprichter en directeur van het Lancaster University Centre for Marketing Analytics and Forecasting en van het International Institute of Forecasters. Hij is lid van de adviesraad van Foresight.
E-mail: r.fildes@lancaster.ac.uk

Paul Goodwin is emeritus hoogleraar aan de Universiteit van Bath, VK, auteur van talloze boeken en artikelen over het gebruik van beoordelingsvermogen bij voorspellingen, en redacteur van Foresight voor Hot New Research.
E-mail: p.goodwin@bath.ac.uk

Correspondentie: Shari De Baets is doctor in de toegepaste economische wetenschappen aan de Universiteit Gent en is universitair Docent aan de Open Universiteit Nederland en gastprofessor aan de KULeuven, België. Ze is redacteur van Foresight voor de Judgment column.
E-mail: shari.debaets@ou.nl

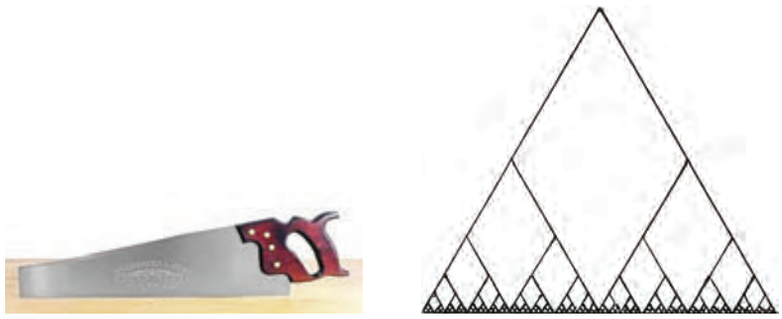
1 + 2 gratis ≠ 3 gratis

Om te beginnen: 1+2≠3

In onze treincoupé hing naast de toegangsdeur een NS-geel plakkaatje met de tekst '1 + 2 kinderen gratis'. Vanuit mijn eersteklaszitplaats kon ik onmogelijk lezen of de kleine lettertjes onderaan dat gele plakkaatje de uitleg verschaften. Tussen mijn hersens schuurde echter iets. Iets, ja iets, drong aan om het triviale rekensommetje 1+2 per omgaande te vertalen in NS-terminologie, namelijk in "drie kinderen gratis met de trein", en "als je gratis wilt reizen, doe dat dan in groepjes van drie". 1+2=3, nietwaar! De trein maakte een tussenstop op een plek waar mijn oog viel op, als ik me goed herinner, een supermarktreclame voor donuts besmeurd met een kleurrijk suikerlaagje. Daarbij het rekensommetje '3+1 gratis'. En ja hoor, weer die suffe conclusie: "Wil je een donut, dan krijg je er vier voor nop. Wel even scannen natuurlijk, maar in de 'bonus' zie je de aftrek". Ik vroeg mijn reisgenote of zij ook van die stomme associaties heeft bij dergelijke reclameteksten. Die had ze niet.

Iets anders: ATAC= Bi1

Niets smaakt natuurlijk beter dan mijn home made jam van bramen-uit-eigen-tuin. Ik doe er weinig suiker bij, maximaal zo'n 20%. Zo weinig betekent, zoals kenners weten, gezonder en helaas ook een kortere houdbaarheid van max (eigen ervaring!) een week, zelfs in de koelkast. In de zomer 'eigen' jam meenemen naar Zuid-Frankrijk slaat derhalve nergens op zonder koelkast. De reeds thuis uitgeknobbelde locatie van de supermarché le plus proche du camping wordt door mijn geliefde en mij onderweg naar die camping steevast als een soortement rituele vakantie-ouverture bezocht voor de broodnodige volgende-ochtend-ontbijtingrediënten, inclusief de ontbijtcroissantjam plus de drie croissants: 1 + 2 voor mij. Tot voor kort, maar zeker ruim voor 'corona', heette die supermarché in mijn campingbuurt ATAC, wat wij heel vroeger thuis riepen als het eten ter tafel kwam (vijf van de uiteindelijk tien kinderen deelden de aanval, de andere vijf lagen in wieg of verschiet). Misschien al eerder vanwege onze verplichte coronapauze, maar dit jaar werd mij (klemtoon op mij) duidelijk dat de ATAC was omgedoopt tot Bi1. Hoe zo Bi!? Tot onze grote verbazing had de Bi1 een schap met een tamelijk rijke sortering aan biologische producten, inclusief biologische jam aux multiple saveurs, van fruit de bois tot oranges. Ik koos de abricots. Inhoud 250 grammes. Ook nog maar even de achterkanttekst van de pot geraadpleegd: Ingrédients: abricots 60%, sucre 50%..... Excusez moi: 60+50 = 110 toch?! Ik toonde de vermeende rekenfout aan mijn geliefde.... en herkende onmiddellijk dat minzame lachje: "Zouden de abrikozen misschien ook suiker kunnen bevatten?" Zelf vindt ze zelfs de bio-jam nog vaak te zoet. Terug op de parkeerplaats stond, de aanvankelijk in de schaduw geparkeerde auto, inmiddels in de volle zon. Een ware aanslag op mijn confiture de la maison zou dat zijn. De volgende ochtend aten we de '1+2'-croissants met de abricozenjam, "Jouw bramenjam is lekkerder', zei ze zonder dat lachje. Ondanks de 50+10 gramme sucre per 100 grammes toch maar wel de jampot in de koeling gedaan. Maar waarom eigenlijk Bi!? 2-1, twee-in-één? Be one, wees een(sgezind)? Zeg het maar. Ik had aucune idée.



Van zaagtanden tot mes, of te wel 2=1

Nog één: 2=1

Het zou kunnen zijn dat u hem kent. Dan rest alleen mijn verhaaltje erbij voor u als mogelijk interessant. Het volgende, hm... niet-triviale, bewijs van 1=2 gaat als volgt. À propos, ik ken 'het bewijs' sinds m'n studententijd. Voor zover ik me herinner van een mede-wiskundestudent. "Waar zit de fout" was toen en is nog steeds de eindvraag na de presentatie van 'het bewijs'.

Het bewijs dus. Als volgt. Beschouw een gelijkzijdige driehoek met zijdelengtes 1. Neem één der zijden als basis; zie de figuur hierboven. Ter benadrukking van de trivialiteit van de redenering zeg ik er vaak zelf bij: "maakt niet uit welke!". De som der beide opstaande zijden is, ook triviaal, gelijk aan 2. Op de basislijn staat dus nu één driehoek; er volgen meer. Vervolgens worden op die basislijn twee nieuwe driehoeken geconstrueerd, waarvan de hoekpunten precies op de helft van de zijden van de grote driehoek liggen. De vier opstaande zijden van deze twee driehoeken hebben samen de lengte, inderdaad weer, 2. Met die twee driehoeken wordt dezelfde grap uitgehaald. Namelijk, op de helft van de zijden van die twee driehoeken construeren we nu vier nieuwe driehoeken op de oorspronkelijke basislijn; de lengte van de opstaande zijden van die vier driehoeken is -yes!- weer 2. En nu even het gas erop (excusez le mot). Weer halverwege de zijden van de vier driehoeken construeren we 16 driehoekjes op de oorspronkelijke basislijn. Die 16 driehoekjes zou je kunnen zien als 16 tanden van een zaag. Als je goed kijkt naar de figuur hierboven, zie je een zaagtandlijn met 32 tandjes. Het aantal, aldus achtereenvolgens geconstrueerde en te construeren driehoeken, vormt de exponentiële rij 1,2,4,8,16,32,64, 128,..... Op plek 10 van deze rij staat het getal 1.048.576, wat –zo zou je dat kunnen zien- overeenkomt met een zaag met meer dan een miljoen tanden. Dichter en dichter kruipen de zaagtanden dus naar de oorspronkelijke basislijn en is in de limiet gelijk aan die basislijn. Jusqu'ici tout va bien. Of niet soms? De rij van getallen, zijnde de lengtes van de opstaande zaagtanden, is 2,2,2,2,2,2..... Dus de tanden van de zaag worden steeds kleiner en fijner, maar de lengte van de zaag -gemeten over de tanden- verandert niet en blijft 2. Met een beetje wiskunde in de rugzak weet je dat de limiet van deze rij ook 2 is.

Ondertussen heeft zich in het oneindige een mirakel voltrokken. De zaag is getransformeerd in een scherp mes. In de limiet is namelijk de zaagtandlijn volledig overgegaan in de oorspronkelijke basislijn. En had die niet de lengte 1? Kortom, de zaagtandlengte 2 is getransformeerd in de basislengte 1. En dus, 2=1. Quod erat demonstrandum.

Weinigen geloven nog in de roomse (biblebelts voor katholiek) transsubstantiatie van vlees in brood (of andersom, dat ben ik kwijt), en niemand van 2 naar 1, hoop ik. Mijn vraag "Waar zit de fout?" wordt nogal eens beantwoord met de opmerking "Het is fout, want 1 is geen 2!" En 2 geen 1, mag ik hopen. Beide antwoorden zijn geen antwoord op de vraag. Ooit kreeg ik als antwoord: "10 centimeter is wel gelijk aan 1 decimeter, maar daarom is 10 nog niet gelijk aan 1." Bijna raak, maar niet helemaal. "Waar zit de fout?" is kennelijk een goeie vraag.

P.S.

Stel ik hang een soepellopende maar volstrekt, voor u op dat moment, overtuigende redenering op met een 'verborgen' redeneerfout(!) en concludeer uit die redenering dat -laten we zeggen- 1+2=3. Dat laatste lijkt mij een waarheid als een koe in een megastal. Echter er zijn welbespraakte lieden op onze fraaie aarde die -alzo plausibel redenerend- hele volksstammen misleiden omdat de eindconclusie van de demagoog hun welgevallig is. Een ontmaskering van de redeneerfout is aan de volgelingen van de nepnieuwsprofeet niet besteed. Maar dit ter zijde. En dat die Franse marketeers de ATAC-supermarché hebben getransformeerd in een Bi1 is voor mij nog steeds net zo'n mysterie als dat mijn maison-made bramenjam zou zijn getransformeerd in biologische confiture d'abricot met 60+50=100 grammes sucre. Onlangs constateerde ik dat het gele NS-affiche met '1+2 gratis' verdwenen is. Jammer.

Gerard Sierksma is emeritus hoogleraar operations research aan de Rijksuniversiteit Groningen.
E-mail: g.sierksma@rug.nl

De Nederdrone: een nieuw tijdperk voor verkenningsmissies op zee

Tess Zijlstra

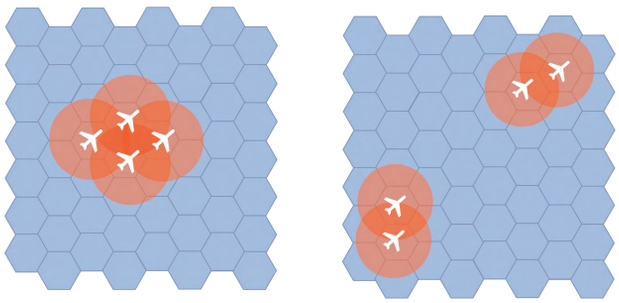
De opkomst van de Nederdrone, ontwikkeld door de TU Delft, Kustwacht Nederland en de Nederlandse Koninklijke Marine, betekent de start van een nieuw tijdperk voor verkenningsmissies op zee. De onbemande luchtvaartuigen gaan een cruciale rol vervullen in zoek- en detectietaken, zoals bij het lokaliseren van vluchtelingen in noodsituaties en het opsporen van drugskoeriers op open water.

Figuur 1: Eerste versie van de Nederdrone tijdens het testen op het dek van Zr. Ms. Johan de Witt. (Marchand, 2020).

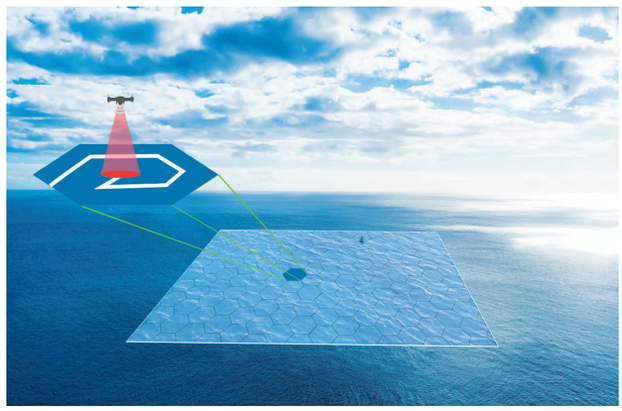
De toekomstvisie is om drones als *autonoom multi-agentsysteem* in te zetten om *maritime situational awareness* te creëren vóór het bemande schip uit. Dit betekent dat naast aandacht voor gebruikelijke eigenschappen van de drone, zoals energiecapaciteit, draaicirkels en sensoren, ook aandacht vereist is voor de communicatie en autonomie van het systeem. Deze twee componenten zijn essentieel voor een nauwkeurige beeldvorming van de omgeving en een effectieve besluitvorming tijdens de missie.

Een autonoom detectiesysteem

Autonoom betekent letterlijk "zelfstandig handelen zonder menselijke tussenkomst". In het kader van zoek- en detectietaken betekent dit dat het team drones zelf zal moeten beslissen hoe en waar ze zoeken op basis van de beschikbare informatie. Zo een autonoom systeem voor detectietaken kan opgedeeld worden in drie delen (Saadaoui en Bouanani, 2018). Als eerste observeren de drones hun omgeving en



Figuur 2: PI: continue communicatie (links) en PII: periodieke communicatie (rechts).



Figuur 3: Zee gemodelleerd door cellen, elke drone zoekt in één cel gedurende een vaste tijd; één zoekperiode.

slaan hierbij informatie op. Ten tweede delen ze de informatie met elkaar voor een zo nauwkeurig en compleet mogelijke beeldvorming. Ten derde bepalen ze waar te zoeken op basis van de beschikbare informatie. Dit artikel beschrijft een deel van het onderzoek dat zich richt op het tweede aspect: de communicatie tussen drones en het effect ervan op het voltooien van de missie (Zijlstra, 2022). Vanwege gerubriceerde informatie is de masterscriptie alleen beschikbaar binnen het Ministerie van Defensie. In het onderzoek zijn twee communicatieprotocollen bestudeerd: PI continue communicatie, waarbij drones voortdurend met elkaar communiceren via directe of indirecte verbindingen, en PII periodieke communicatie, waarbij de drones opsplitsen en elkaar na een bepaalde tijd ontmoeten om informatie uit te wisselen (Figuur 2). In de praktijk zal een commandant moeten beslissen welk protocol de drones zullen uitvoeren tijdens de missie. Daarom is het cruciaal om inzicht te krijgen in welk protocol geschikt is in welke situatie.

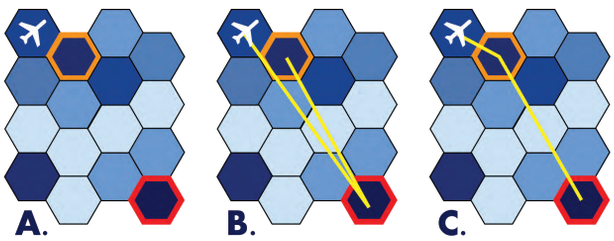
Missie

Om de situatie concreet te maken, wordt een team van N_U drones ingezet bij een zoek- en detectie missie op zee, waar N_T stilliggende objecten, ook wel 'targets' genoemd, gesitueerd zijn in een begrensde zoekregio. Het doel is om de targets zo snel mogelijk te vinden en het schip te voorzien van hun coördinaten. Wanneer alle targets zijn gedetecteerd en het hele team hiervan op de hoogte is, vliegen ze terug naar hun startpositie en is de missie voltooid. Er zijn twee prestatiematen waaraan de uitvoering wordt getoetst: 1) de missietijd; de tijd die verstrijkt tussen de start en het voltooien van de missie, en 2) de informeertijd (per target); die aangeeft op welk exact tijdstip na start van de missie het schip is geïnformeerd over de locatie van het target.

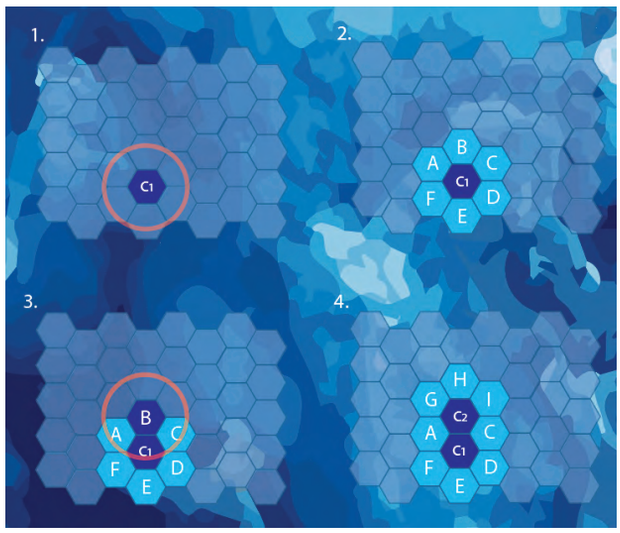
Model

Voor het waarnemen van de omgeving wordt de zee gemodelleerd als een plattegrond van cellen waarin zich wel of geen targets bevinden. Informatie over de zoekregio wordt vertaald naar kansen voor targetaanwezigheid binnen die cel, afgekort met de TEP (Target Existence Probability). Elke drone zoekt in een toegewezen cel met behulp van camera's of infrarood sensoren. De drones scannen de zoekregio cel voor cel gedurende 'zoekperiodes', waarin per zoekperiode één cel per drone wordt gescand (Figuur 3). De TEPs worden geüpdatet na elke zoekperiode wanneer meer informatie over het gebied beschikbaar wordt. Om te bepalen in welke cel de drone moet zoeken, wordt een 'greedy' heuristiek gebruikt. Hierbij wordt bij het kiezen de voorkeur gegeven aan cellen met de hoogste TEP. Doordat het vliegen naar een cel tijd kost en te 'greedy' zoeken kan leiden tot zeer inefficiënte routes, zie Figuur 4, wordt naast de kans ook de afstand tot de cel meegenomen in de beslissing waar te zoeken. Naar deze methode wordt verwezen als de Greedy-Search methode.

De selectie van zoekcellen wordt naast de Greedy-Search methode ook beïnvloed door de communicatierange van de drones en het communicatieprotocol. Beschouw als voorbeeld de situatie in Figuur 5 waarbij de drones een communicatierange hebben van één cel. De eerste zoekcel wordt gekozen door Greedy-Search toe te passen, dit blijkt cel c_1 (1). Vanuit c_1 worden de cellen geselecteerd die



Figuur 4: Hoe donkerder de cel, hoe hoger de TEP, rood gemarkeerde cel hoogste TEP, oranje gemarkeerde cel op één na hoogste TEP. Zonder het integreren van afstand zal de route in B. worden gekozen (inefficiënt), met het integreren van afstand zal de route in C. worden gekozen (wel efficiënt).



Figuur 5: De selectieprocedure, het selecteren van zoekcellen door Greedy-Search en communicatierestricties.

zich binnen de communicatierange (oranje cirkel) bevinden (2), dus A, B, C, D, E en F. Vervolgens wordt de cel met de hoogste TEP gekozen binnen deze selectie (3), in dit geval cel B. Daarna worden de cellen opnieuw geselecteerd die zich binnen de communicatierange van de eerder gekozen cellen bevinden (4), waaruit de cel met hoogste TEP wordt geselecteerd. Dit proces herhaalt zich totdat het aantal geselecteerde zoekcellen (donkerblauw) gelijk is aan het aantal inzetbare drones. De 'selectieprocedure' verwijst naar de combinatie van Greedy-Search en het proces in Figuur 5.
Deze methode wordt zowel toegepast voor PI: continue communicatie, als voor PII: periodieke communicatie. Het verschil voor de protocollen zit in het zoekgebied waarin dit wordt toegepast.

Voor PI is het gebied voor het selecteren van zoekcellen de gehele zoekregio. Nadat de zoekcellen zijn toegewezen aan de drones, vliegen ze naar deze cellen en beginnen ze met scannen. Na één zoekperiode delen de drones informatie waarna de TEPs worden geüpdatet. Vervolgens worden nieuwe zoekcellen geselecteerd met de selectieprocedure en starten de drones weer met zoeken.

Voor PII wordt de selectieprocedure toegepast in kleinere gebieden binnen de zoekregio; de subregio's. Per subregio wordt een deel van het team; een subteam, toegewezen voor een vast aantal ψ zoekperiodes. Het bepalen van de subregio's wordt gedaan door middel van een clusteringsalgoritme (K-means++) toe te passen op cellen met een hoge TEP, wat gedetailleerd omschreven is in het onderzoeksrapport (Zijlstra, 2022). Het aantal drones dat wordt toegewezen aan een subregio hangt af van het relatieve aantal cellen met een hoge TEP binnen die subregio en vormt een subteam. Deze subteams zoeken in hun toegewezen subregio zoals eerder uitgelegd voor PI. Na ψ zoekperiodes komen de drones samen op een vooraf bepaalde *rendez-vous* locatie om nieuwe informatie te delen. Op basis hiervan worden nieuwe subregio's en subteams gemaakt, en beginnen de drones weer met zoeken gedurende ψ zoekperiodes.

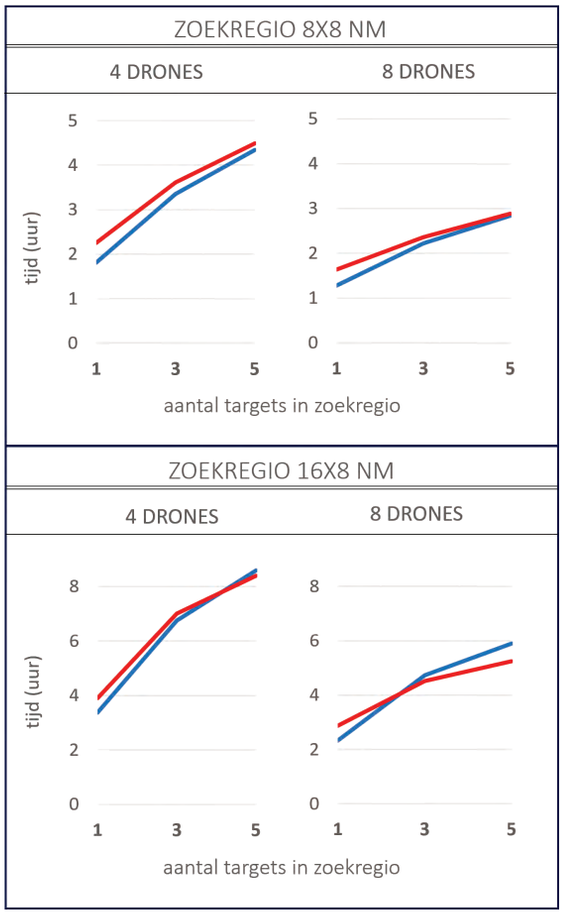
Simulatie & Resultaten

Omdat het onderzoek specifiek voor de Nederdrone is uitgevoerd, zijn de missies gesimuleerd in een computeromgeving met reële parameters voor situaties waarin deze drone zal worden ingezet. Twaalf verschillende missietypes zijn gesimuleerd, variërend in de volgende eigenschappen: grootte van zoekregio, grootte van het team en in het aantal targets. Figuur 6 geeft deze missietypes weer.

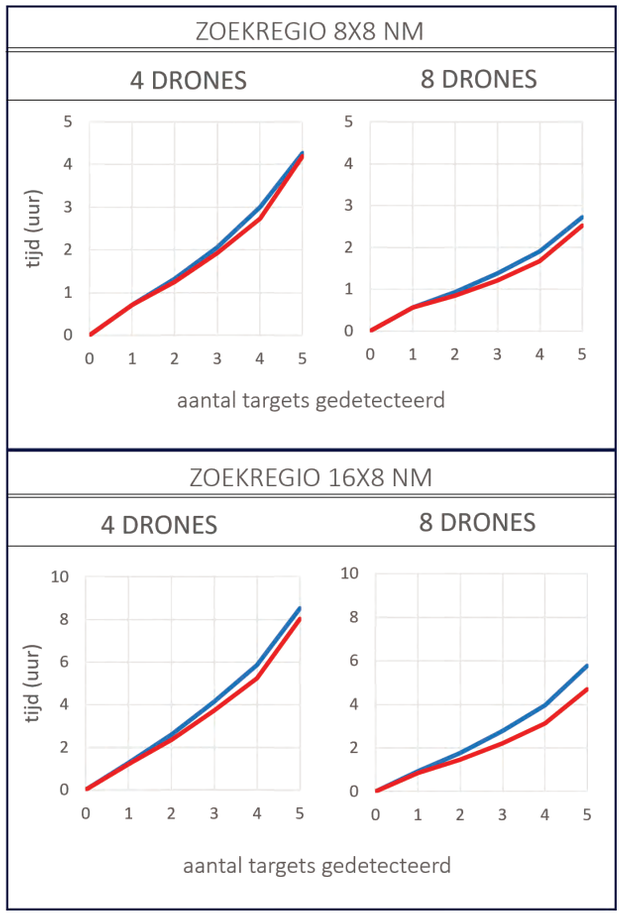
ZOEKREGIO 8x8 NM		ZOEKREGIO 16x8 NM	
4 DRONES	8 DRONES	4 DRONES	8 DRONES
1 TARGET	1 TARGET	1 TARGET	1 TARGET
3 TARGETS	3 TARGETS	3 TARGETS	3 TARGETS
5 TARGETS	5 TARGETS	5 TARGETS	5 TARGETS

Figuur 6: Twaalf type missies, variërend in grootte regio (NM = nautical miles), team en het aantal targets.

Van elk missietype zijn 300 verschillende startscenario's uitgevoerd waarna de gemiddelde prestaties



Figuur 7: Gemiddelde **missietijden** voor de twaalf verschillende missietypes.



Figuur 8: Gemiddelde **informeertijden** voor de missies met vijf geplaatste targets.

van PI en PII zijn vergeleken. Zoals eerder benoemd, wordt de effectiviteit van PI en PII beoordeeld volgens de prestatiematen missie- en informeertijd.

Uit de resultaten blijkt dat de prestatie per protocol afhangt van zowel de eigenschappen als het doel van de missie (missie- of informeertijd). Een snelle missietijd heeft de voorkeur wanneer er alleen inzicht moet worden verkregen in de zoekregio, maar niet direct actie nodig is na detectie. Een korte informeertijd is cruciaal bij bijvoorbeeld het opsporen van drenkelingen. Bovendien blijkt continue communicatie (PI) een betere keuze in kleinere zoekregio's met minder targets en minder drones. Periodieke communicatie (PII) blijkt juist beter in relatief grotere zoekregio's, met meer targets en meer drones.

De grafieken in Figuur 7 laten de gemiddelde missietijden van elke type missie zien, waarbij de missietijd op de verticale as staat en het aantal targets per missie op de horizontale as. PI (blauw) voltooit de missie sneller dan PII (rood) voor de zoekregio van 8x8 cellen. Het verschil neemt echter af bij meer targets en meer drones. Bij de grote zoekregio gaat PII beter presteren dan PI (rode lijn snijdt de blauwe lijn), waarbij het omslagpunt bij een team van vier drones ligt bij vijf targets, en bij een team van acht bij drie targets.

De resultaten van de informeertijden tonen een vergelijkbaar verband, maar op een eerder moment. In Figuur 8 worden de gemiddelde informeertijden per target weergegeven voor missies met vijf targets. Hier presteert PII altijd iets beter dan PI. Dit is anders dan Figuur 7 aangeeft voor de totale missietijd, waarbij voor de zoekregio van 8x8 cellen (linkerkant) PI nog beter presteerde dan PII met vijf targets. In deze situatie wordt het schip gemiddeld sneller op de hoogte gebracht van de targets met periodieke communicatie, maar voltooit het team de missie langzamer. Ook blijkt uit Figuur 8 dat het verschil tussen PI en PII toeneemt naarmate er meer drones worden ingezet en de zoekregio groter wordt.

Voor de bevindingen zijn er verschillende verklaringen. Het creëren van subregio's resulteert in efficiëntere routes, maar het kost wel extra tijd om weer samen te komen om elkaar te voorzien van nieuwe informatie. Dit resulteert in een trade-off tussen route-efficiëntie en informatieoverdracht.

De bevindingen tonen aan dat bij grotere zoekregio's efficiënte routes belangrijker worden. Ook het teamformaat speelt een rol, waarbij meer drones leiden tot meer subregio's en efficiëntere zoekroutes. Daarnaast gaat PII ook beter presteren vergeleken PI bij meer targets. Een verklaring hiervoor is dat met één target de drones relatief langer onnodig doorzoeken dan wanneer er meer targets te vinden zijn. Dit suggereert dat naarmate de missie langer duurt, het ontbreken van informatie gedurende dezelfde periode minder belangrijk wordt.

Conclusie

Het onderzoek laat zien dat de juiste keuze van continue of periodieke communicatie tussen de drones sterk afhangt van de situatie. Factoren zoals het aantal beschikbare drones, zoekregiogrootte en het aantal targets spelen een belangrijke rol. Daarnaast is het essentieel om te bepalen of korte missie- of informeertijd prioriteit heeft. Snelheden, vlieghoogte en sensorkwaliteit zijn tevens bepalende factoren voor het meest geschikte communicatieprotocol, maar variatie hierin valt buiten het kader van dit artikel. De integratie van een simulatiesysteem dat op basis van de eigenschappen van de missie het optimale protocol kiest, is wellicht de sleutel tot het effectief ondersteunen bij zoek- en detectiemissies in de toekomst.

Literatuur

A. Marchand. *Explosieve krachtbron voor drone*. Jul 2020. URL: https://magazines.defensie.nl/allehens/2020/06/08_drones-op-waterstof.

H. Saadaoui en F. E. Bouanani. „Information sharing based on local PSO for UAVs cooperative search of unmoved targets". In: *2018 International Conference on Advanced Communication Technologies and Networking (CommNet)* (2018), p. 1–6. DOI: 10.1109/COMMNET.2018.8360276.

T. Zijlstra. *The influence of communication protocols on the performance of a Greedy Search strategy for maritime S&D missions*. 2022.

Tess Zijlstra behaalde haar master in Quantitative Logistics and Operations Research, Econometrics and Management Science in december 2022 aan de Erasmus Universiteit Rotterdam. In maart startte ze als assistent-docent aan de Nederlandse Defensie Academie waar ze werkzaam is tot eind 2023. Ze is momenteel op zoek naar een nieuwe uitdaging.
E-mail: tess_zijlstra@hotmail.com

De auteur bedankt Albert Wagelmans, Herman Monsuur en Martijn van Ee voor hun hulp en toewijding aan het onderzoek en daarnaast Kerry Malone voor haar aandacht tijdens het schrijven van het artikel.

VVSOR
OPERATIONS RESEARCH

Joint LNMB/NGB seminar: Urban Logistics and Mobility

On Wednesday January 17, 2024, the The Dutch Network on the Mathematics of Operations Research (LNMB) and the Dutch OR Society (NGB) are organizing a one-day seminar on *Urban Logistics and Mobility*.

Speakers include:

- Claudia Archetti (ESSEC Business School, Paris) - Freight on Transit problem: Strategic, Tactical and Operational Planning
- Jan Fransoo (Tilburg University) - Parking for delivery: Quantifying, modeling, and improving urban logistics
- Shadi Sharif Azadeh (Delft University of Technology)
- Niels Agatz (Erasmus University Rotterdam) and Iris Vis (University of Groningen)
- Hans Quak (TNO) - Developments influencing city logistics and their impact on neighborhoods
- Jaap Hatenboer (Axira) - Medical drone services

NGB provides an additional arrangement for MSc students who are members of the NGB (student membership is free). Up to a maximum of twenty students can participate for a reduced rate of €50. Students who want to register for the reduced rate should not use the link on the website, but send an email to secretariaat@ngb-online.nl.

The seminar will be held at Conference center Kontakt der Kontinenten in Soesterberg. More information can be found on

https://www.vvsor.nl/operations-research/news/seminar-urban-logistics/

Rangen

Je kunt zaken op een rij zetten: personen, landen, producten of wat dan ook. Voor het bepalen van de volgorde gebruik je een criterium, zoals grootte, inkomen of gewicht. Zoiets heet een rangorde en de statistiek kent toetsen om het verband van zo'n rangorde variabele met andere variabelen te beschrijven.

Bij mij komen bij het werken met rangordes echter veel jeugdherinneringen boven, rang en stand waren toen heel belangrijk. Kent u dat? De buurvrouw wordt aangesproken als vrouw Meerdink, de echtgenote van een middenstander als juffrouw Bekkers en die van een notabele als mevrouw Hoogland. In het Winterswijk van de jaren '50 waar ik

opgroeide (ik ben van 1941) moest men het niet wagen hier een vergissing in te maken! Gerrit Komrij, die ook uit dit dorp kwam, heeft enkele niet al te vleiende opmerkingen gemaakt over de geest die er heerste in het dorp dat in zijn woorden 'ligt in de verste uithoek van een streek die zelf weer de Achterhoek heet'.

Denk maar niet dat er een simpele indeling was: zelfs binnen de arbeidersklasse, wat men daar ook onder mag verstaan, waren er allerlei nuances. Zo maakte het uit of iemand los werk deed bij een boer of dat hij in dienst was bij een middenstander. Tijdens mijn studie in Amsterdam kwam ik een plaatsgenoot tegen die daar telefoonmonteur was. Ik kende hem nog van een jeugdvereniging en we hebben het contact vernieuwd en zijn heel goede vrienden geworden. Winterswijkse buren waarschuwden mijn ouders, ik zou later waarschijnlijk leraar op een HBS worden en dan zou ik die vriend nooit kunnen uitnodigen op feestjes waar mijn collega's kwamen, dat zou heel ongepast zijn!

Dat besef van rangen en standen werkte ver door. Mijn vader was schoenmakersknecht en ik had al vrij jong een bijbaantje in het bezorgen van de driemaandelijkse rekeningen bij klanten die op krediet kochten. Dat waren so-wie-so de betere standen, een arbeider werd geacht direct bij aankoop te betalen. Maar sommige van die klanten kregen hun rekening per post, een jongen als ik behoorde niet te weten dat zulke hoog boven mij verheven mensen op krediet kochten. Achteraf bezien kan ik hier alleen maar om lachen, iedereen wist alles van iedereen in die kleine besloten gemeenschap, hoe zeer men ook de illusie van privacy koesterde.

Vanaf mijn 14e heb ik iedere zomervakantie in het laboratorium van een textielfabriek gewerkt. Ook daar heersten rangen en standen. De arbeiders begonnen om 7.00 uur 's ochtends en moesten vijf minuten voor tijd ingeklokt hebben. Die vijf minuten had je nodig om van de prikklok naar je werkplek dieper in de fabriek te lopen. Was je ook maar één minuut te laat dan werd er een kwartier gekort op je loon van toen ca. 50 cent per uur. Voor kantoorpersoneel golden uiteraard andere regels.

Zelfs bij de wc's in die fabriek was er standsverschil. Die voor kantoorpersoneel stonden ruggelings tegen die voor de arbeiders, wel met een solide muur ertussen. Dat was handig bedacht, zo kwamen ze beide uit op dezelfde onderliggende afvoer. Maar de kantoor wc's waren betegeld en van boven dicht en hadden deuren van onder tot boven. Op die deuren stond 'dames' of 'heren'. De arbeiders wc's hadden kale cementen muren, waren van boven open en op de driekwart deuren stond 'mannen' of 'vrouwen'.

Ik zat op de HBS, het plaatselijk Lyceum was niet zo groot, daarom werden de klassen 4 en 5 HBS-B voor de exacte vakken gecombineerd met 5 en 6 Gymnasium-ß. Ik was extreem goed in scheikunde, net als de gymnasiast Jan. De docent zette ons daarom naast elkaar in de bovenste bank van het amfitheater, dan kon niemand bij die twee uitblinkers afkijken. Na ons eindexamen in 1958 deed ik mijn gebruikelijke vakantiewerk. Die Jan waar ik twee jaar lang enkele uren per week mee in dezelfde bank had gezeten was de zoon van de directeur van die fabriek. Op een dag bezocht hij het bedrijf en kwam mij even opzoeken in het laboratorium. We hebben kort gepraat over onze voorgenomen studies in Amsterdam en Delft en de problemen bij het vinden van een kamer. We wensten elkaar veel succes bij de studie en een fijne vakantie en ik ging verder met mijn werk. Niet lang daarna kreeg ik een standje: ik had zomaar met de toekomstige directeur staan praten en hem Jan genoemd in plaats van mijnheer Jan. Dat ik twee jaar samen met Jan in één bank had gezeten telde niet: je moest je plaats in de sociale rangorde kennen!

Overigens speelde dit standsbesef niet alleen in die verre Achterhoek. Ik heb ooit met wat studievrienden in Friesland gezeild. Drinkwater haalden we in melkbussen bij een grote verffabriek even voorbij onze ligplaats. Die fabriek was van de familie van een van mijn vrienden en jawel hoor, ook hier het vertrouwde beeld dat ik zo verfoeide. Van alle kanten klonk het 'goedemorgen mijnheer Bert' als we door de fabriek liepen en werden er eerbiedig petten afgenomen.

Het moge duidelijk zijn dat ik liever met het statistische begrip rangordes van doen heb dan met die licht traumatische jeugdherinneringen.

Gerrit Stemerdink is eindredacteur van STAtOR.
E-mail: gjstemerdink@hotmail.com

Het automatisch detecteren van recreatiewoningen met luchtfoto's in Zeeland

Lisa Mulder

Het Kadaster heeft verschillende basisregistraties in beheer, zoals de Basisregistratie Adressen en Gebouwen (BAG). De BAG bevat informatie over onder andere panden, verblijfsobjecten en standplaatsen, wat recreatiewoningen kunnen zijn. Het Kadaster wil de precieze aantallen en locaties van recreatiewoningen in Nederland weten, omdat deze vraag speelt bij onder andere de overheid en Regionale Informatie- en Expertise Centra. Zij willen meer zicht op vakantieparken krijgen, zodat onder andere ondermijning kan worden tegengegaan. Uit voorgaand onderzoek van het Kadaster blijkt dat niet alle recreatiewoningen geregistreerd staan in de BAG, maar dat deze wel te zien zijn vanaf luchtfoto's. Kunnen deze recreatiewoningen automatisch uit luchtfoto's gedetecteerd worden, zodat de aantallen en locaties gevonden kunnen worden?

Het Kadaster wil weten hoeveel recreatiewoningen er zijn in Nederland en waar ze zich bevinden. Daarom is onderzocht hoe recreatiewoningen automatisch gedetecteerd kunnen worden. Er is specifiek naar Zeeland gekeken, omdat er eerder door HZ Kenniscentrum Kusttoerisme is onderzocht hoeveel recreatiewoningen er in Zeeland zijn. Op 1 januari 2022 waren dit er afgerond 43.500 (Korteweg Maris, Soeters en Lemmers, 2022). De gevonden aantallen konden met elkaar vergeleken worden. Tijdens dit onderzoek is de volgende vraag onderzocht: 'Hoeveel recreatiewoningen, die niet zijn opgenomen in de Basisregistratie Adressen en Gebouwen, zijn er binnen vakantieparken en recreatiegebieden in de provincie Zeeland te vinden op basis van luchtfoto's?'

De data

Jaarlijks worden luchtfoto's gemaakt van heel Nederland en ter beschikking gesteld als open data door het samenwerkingsverband Beeldmateriaal. De luchtfoto's die gebruikt zijn in dit onderzoek, zijn gemaakt in de winter van 2022 (z.d., 2023). Met behulp van een raster zijn de luchtfoto's van Zeeland opgedeeld in afbeeldingen van 1024 bij 1024 pixels. Er is door het Kadaster een dataset met gelabelde luchtfoto's samengesteld vanuit de basisregistraties om het objectdetectiemodel op te trainen. Op deze luchtfoto's zijn 16.531 panden met een logies- of woonfunctie en standplaatsen binnen vakantieparken en recreatiegebieden in Zeeland te zien, die geregistreerd staan in de BAG. Voorbeelden van panden die voor recreatieve doeleinden gebruikt worden zijn vakantiehuizen en bungalows. Stacaravans en chalets zijn verplaatsbaar en mogen daarom niet geregistreerd worden als pand, maar wel als standplaats. Van de panden en standplaatsen zijn de geometrieën gebruikt voor het objectdetectiemodel. De geometrie geeft de coördinaten van de omtrek van het pand of de standplaats weer.

Deze dataset is eerst gecontroleerd op kwaliteit. In figuur 1 zijn enkele voorbeelden te zien van wat opviel bij het analyseren van de data. Ten eerste viel op dat de geometrieën van de panden en stand-

Figuur 1: Geometrieën komen vaak niet overeen met de recreatiewoningen die te zien zijn op de luchtfoto's (a, b). Er zijn geometrieën bekend waar geen recreatiewoningen te zien zijn (c, d). Recreatiewoningen in bossen zijn slecht zichtbaar (e). Aan de buitenkant van een pand is niet te zien hoeveel recreatiewoningen erin zitten (f). De witte vlakken zijn geometrieën van panden en de rode randen zijn geometrieën van standplaatsen.

plaatsen, die recreatiewoningen kunnen zijn, vaak niet overeenkomen met de recreatiewoningen die te zien zijn op de luchtfoto's. In figuur 1a zijn kleine afwijkingen te zien tussen de vormen van de witte geometrieën van panden en de onderliggende recreatiewoningen. Daarnaast komt vaak de rood omlijnde geometrie van een standplaats niet overeen met de recreatiewoning, omdat de geometrie de grond van de standplaats weergeeft (figuur 1b). Ten tweede zijn er in figuur 1c geometrieën van standplaatsen te zien op luchtfoto's waar geen gebouwen te zien zijn, maar gewoon een stuk grond. Hier zou een tijdelijke recreatiewoning op geplaatst mogen worden, maar op het moment dat de luchtfoto gemaakt was, was dit niet het geval. Ook zijn er soms geometrieën van panden ingetekend die niet te zien zijn op de luchtfoto, doordat de bouw nog niet is gestart (figuur 1d). Verder is in figuur 1e te zien dat recreatiewoningen in bossen slecht of zelfs helemaal niet zichtbaar zijn. Daarnaast zijn er ook gebouwen die meerdere recreatiewoningen bevatten, maar vanaf de luchtfoto is niet te zien hoeveel dit er precies zijn (figuur 1f). De dataset zal aangepast moeten worden om te kunnen gebruiken voor het trainen van een objectdetectiemodel.

Bij het analyseren van de data viel ook op dat met name stacaravans en chalets niet geregistreerd staan. Om dit type recreatiewoningen zo goed mogelijk te kunnen detecteren, zijn uitsluitend stacaravans en chalets gebruikt voor het trainen van het model. De volgende recreatiewoningen zijn, naar aanleiding van de kwaliteitscontrole, niet meegenomen bij het trainen van het model:

- Geometrieën van panden en standplaatsen waarop geen recreatiewoningen staan.
- Recreatiewoningen die in bossen staan.
- Geometrieën van gebouwen die meerdere recreatiewoningen bevatten.

Deze bewerkingen zijn handmatig gedaan door alle recreatiewoningen in een vakantiepark tegelijk te bekijken. Van de geselecteerde recreatiewoningen zijn de geometrieën die niet overeenkomen met de recreatiewoningen op de luchtfoto's handmatig gecorrigeerd. Omdat er weinig stacaravans en chalets geregistreerd staan, is ervoor gekozen om handmatig extra stacaravans en chalets op luchtfoto's te labelen. Ook zijn er afbeeldingen met achtergrond, waarop geen gelabelde recreatiewoningen staan, gebruikt. Zo leert het model dat lege standplaatsen, tenten, campers en caravans geen recreatiewoningen zijn. Uiteindelijk zijn er 500 afbeeldingen met 6271 recreatiewoningen gebruikt voor het trainen van het model.

Objectdetectiemodellen

Er is gekozen voor de methode objectdetectie. Hierbij worden alle recreatiewoningen gelokaliseerd binnen een afbeelding en voorzien van een bounding box, die de objecten met een rechthoek omsluiten. Het algoritme dat is gebruikt voor objectdetectie is YOLOv8, dat beschikbaar is als applicatie van de programmeertaal Python. Dit algoritme heeft als invoer afbeeldingen en bijbehorende txt-bestanden nodig, waarin de bounding boxen van de gelabelde recreatiewoningen staan. Van de afbeeldingen en txt-bestanden is 80% afgesplitst in een train-dataset, 10% in een validatie-dataset en nog eens 10% in een test-dataset. Omdat er weinig afbeeldingen zijn om het model mee te trainen, is er gebruikgemaakt van data-augmentatie om de train-dataset verder uit te breiden. Hierbij worden de afbeeldingen steeds iets aangepast, door de kleur te veranderen, de helderheid aan te passen, de afbeelding te verschuiven, bij te snijden en te spiegelen. Ook is er gebruikgemaakt van mozaïek-augmentatie, waarbij vier afbeeldingen worden samengevoegd tot één afbeelding. YOLOv8 heeft verschillende modelgroottes, variërend van het nano model tot het extra large model. Hoe groter het model, des te beter het presteert, maar hoe langzamer het is in het trainen van het model en het maken van voorspellingen. Met de beste waarden voor de parameters van de data-augmentatie zijn de parameters voor de modelgroottes onderzocht. Het model dat het beste presteerde op de validatiedataset is uiteindelijk gekozen om voorspellingen mee te doen op de test-dataset en op de dataset waarin afbeeldingen zitten met recreatiewoningen die niet bekend zijn vanuit de BAG. Er is gekozen voor het small model. Dit model presteerde het beste op zowel de validatie-dataset als de test-dataset. Het model heeft een mean average Precision (mAP) van 0,99 op de test-dataset. De mAP is een veelgebruikte prestatiemaat voor objectdetectie, met als maximum een 1. Ook kan op basis van de precision en recall gezegd worden dat 95% van de voorspellingen correct is en 97% van de daadwerkelijke recreatiewoningen uit de test-dataset gedetecteerd is.

Analyse van de voorspellingen

Het model is toegepast op een nieuwe dataset met luchtfoto's waarop recreatiewoningen te zien zijn, die nog niet geregistreerd staan in de BAG, en heeft hierbij 22.306 bounding boxen voorspeld. Een recreatiewoning kan meerdere bounding boxen hebben,

doordat hij is afgesneden langs de randen van de afbeeldingen. Deze zijn samengevoegd, om te voorkomen dat een recreatiewoning meerdere keren wordt meegeteld. Uiteindelijk zijn er 18.802 recreatiewoningen in Zeeland gevonden, die niet zijn opgenomen in de BAG. Op basis van de precision en recall op de test-dataset zouden 940 voorspellingen eigenlijk geen recreatiewoningen zijn en 552 recreatiewoningen niet gedetecteerd zijn.

De voorspelde recreatiewoningen op de luchtfoto's zijn vervolgens handmatig gecontroleerd om te zien hoe goed de voorspellingen zijn en hoeveel recreatiewoningen gemist zijn. Er valt op dat het objectdetectiemodel over het algemeen erg goed is in het herkennen van recreatiewoningen. In figuur 2a is te zien dat op grote vakantieparken met veel recreatiewoningen vrijwel alle recreatiewoningen zijn herkend. Ook is het objectdetectiemodel over het algemeen goed in het onderscheid maken tussen recreatiewoningen en objecten die geen recreatiewoningen zijn. Tenten, campers en caravans worden terecht niet herkend (figuur 2b).

Toch zijn er ook nog enkele aandachtspunten waar het objectdetectiemodel iets tekortschiet. Recreatiewoningen die dicht op elkaar staan of aan elkaar vastzitten zijn soms niet herkend (figuur 2c). Dit zou verklaard kunnen worden doordat het model uitsluitend getraind is op vrijstaande recreatiewoningen. Daarnaast heeft het objectdetectiemodel soms moeite met het detecteren van witte stacaravans (figuur 2d). Het model heeft geleerd dat campers en caravans geen recreatiewoningen zijn. Aangezien campers en caravans in de meeste gevallen wit zijn, kunnen witte stacaravans hierop lijken. Dit zou kunnen verklaren waarom witte stacaravans vaker niet herkend worden dan andere kleuren stacaravans.

Verder zijn recreatiewoningen die onder bomen staan slecht zichtbaar vanuit luchtfoto's. Het objectdetectiemodel is hier niet op getraind en heeft dan ook moeite met het detecteren van deze recreatiewoningen (figuur 2e). Tot slot zijn er ook situaties waarin het model een object detecteert als recreatiewoning, terwijl het geen recreatiewoning is, maar een tent, caravan, parkeerplaats, luchttrampoline

Figuur 2: Correct gedetecteerde recreatiewoningen zijn weergegeven als blauwe rechthoeken (a, b, c, d, e, f), gemiste recreatiewoningen zijn rood omlijnd (a, c, d, e), correct niet-gedetecteerde campers en caravans zijn groen omlijnd (b) en incorrect gedetecteerde tenten zijn paars omlijnd (f).

of een ander rechthoekig object. In figuur 2f is te zien dat tenten onterecht zijn herkend als recreatiewoning.

Vergelijking met HZ Kenniscentrum Kusttoerisme

Wanneer de aantallen die gevonden zijn met objectdetectie die niet bekend zijn vanuit de BAG opgeteld worden bij de aantallen die wel bekend zijn vanuit de BAG, zijn er 35.333 recreatiewoningen in Zeeland binnen de geselecteerde gebieden gevonden. Het onderzoek van HZ Kenniscentrum Kusttoerisme heeft afgerond 43.500 recreatiewoningen opgeleverd. Er is dus een verschil tussen beide onderzoeken van ongeveer 8.000 recreatiewoningen. Een verklaring voor het verschil is dat HZ Kenniscentrum Kusttoerisme alle jaarplaatsen heeft meegeteld, terwijl het objectdetectiemodel alleen jaarplaatsen heeft kunnen detecteren waar recreatiewoningen op staan. Een andere verklaring is dat tijdens dit onderzoek uitsluitend is gekeken naar recreatiewoningen binnen vakantieparken en recreatiegebieden, terwijl HZ Kenniscentrum Kusttoerisme ook losse eenheden heeft meegeteld die zich niet op een vakantiepark bevinden. Gebaseerd op dit onderzoek kan een volgende stap zijn om met HZ Kenniscentrum Kusttoerisme het verschil van 8.000 recreatiewoningen verder te onderzoeken en te zien of dit verschil te verklaren is met informatie over jaarplaatsen en losse eenheden. Daarnaast kan het objectdetectiemodel door zijn goede prestaties breder ingezet worden, zodat de recreatiewoningen in de rest van Nederland gevonden kunnen worden.

Literatuur

D. Korteweg Maris, M. Soeters en E. Lemmers. *Aanboddatabase verblijfsrecreatie*. https://www.kenniscentrumtoerisme.nl/wiki/index.php/KCKT_Publication_PR_00010. HZ Kenniscentrum Kusttoerisme, 2022.

B. z.d. *Luchtfoto's downloaden*. Beeldmateriaal Nederland. 2023. URL: https://opendata.beeldmateriaal.nl/pages/downloaden.

Lisa Mulder is in 2023 afgestudeerd bij de studie Toegepaste Wiskunde op de Hogeschool van Amsterdam. Tijdens haar afstudeerstage deed ze, in opdracht van het Kadaster, onderzoek naar het automatisch detecteren van recreatiewoningen uit luchtfoto's in Zeeland. Dit artikel is een samenvatting van het afstudeeronderzoek. Het hele onderzoek is te vinden via https://bit.ly/STAtOR_LGM
E-mail: lisa.mulder99@hotmail.com

Toegankelijke en duurzame statistiek met JASP

Eric-Jan Wagenmakers, Johnny van Doorn en Don van den Bergh

JASP is open-source software die beoogt oude en nieuwe statistische methoden te ontsluiten voor een breed publiek. De gebruiksvriendelijke interface stelt studenten, docenten en onderzoekers in staat om zelfs ingewikkelde analyses in een oogwenk uit te voeren, zodat de nadruk kan liggen op de interpretatie van de uitkomsten in plaats van op de berekeningen of de computer code. JASP biedt een volwaardige en moderne vervanging voor commerciële programma's zoals SPSS en Minitab.

Sinds 2012 ontwikkelen we aan de Universiteit van Amsterdam het open-source statistiekprogramma JASP (jasp-stats.org). De oorspronkelijke doelstelling van JASP was om de Bayesiaanse statistiek net zo toegankelijk te maken als de frequentistische statistiek: het streven was dus naar een Bayesiaanse *aanvulling* op SPSS, maar dan open-source en vriendelijker in het gebruik. Naarmate er steeds meer frequentistische functionaliteit aan JASP werd toegevoegd verschoof deze doelstelling en werd het steeds realistischer om te streven naar een volledige *vervanging* van SPSS. We denken nu op dit punt te zijn aangeland.

Op dit moment gebruiken honderden universiteiten wereldwijd JASP ter ondersteuning van hun statistiekonderwijs (zie Figuur 1). Dit komt mede doordat JASP in verschillende talen beschikbaar is: Engels, Nederlands, Duits, Spaans, Portugees, Frans, Galicisch, Pools, Russisch, Chinees, Japans, en Indonesisch.

Figuur 1: De 284 universiteiten (in 66 landen) waarvan we weten dat ze JASP inzetten in hun statistiekonderwijs. JASP wordt per maand ongeveer 75.000 keer gedownload. Bron: JASP

Het is onmogelijk om in een kort bestek alle functionaliteit van JASP te bespreken of ook maar bondig samen te vatten – op dit moment gebruikt JASP 475 verschillende R pakketten (https://jasp-stats.org/r-package-list/) verspreid over zeven standaard analysetechnieken (beschrijvende statistiek, *t*-toetsen, ANOVA, gemengde modellen, regressie, frequenties en factor-analyse) en 25 modules die de gebruiker naar believen kan activeren en deactiveren (bv. voor power-analyse, meta-analyse, structurele vergelijkingsmodellen, etc.). Hieronder presenteren we vier voorbeeld-analyses, maar allereerst willen we de aandacht vestigen op een aantal belangrijke eigenschappen die JASP onderscheiden van de meeste andere statistiekprogramma's.

Kenmerkende Eigenschappen van JASP

De ontwikkeling van JASP is gedreven door desiderata van gebruiksgemak, toegankelijkheid, overzichtelijkheid, reproduceerbaarheid, duurzaamheid en statistische inclusiviteit. Deze desiderata vinden hun weerslag in de volgende kenmerkende eigenschappen:

- JASP heeft een *graphical user interface* (GUI). Dit betekent dat de gebruiker opties selecteert uit een menu, net zoals bij SPSS. Let wel: de JASP GUI is 100% reproduceerbaar – een opgeslagen `.jasp` file bevat niet alleen de data, de resultaten en eventuele annotaties, maar ook de aangevinkte analyse-opties.
- Sinds kort staat JASP de gebruiker toe om voor de standaard analyses de *corresponderende R code* op te slaan. In JASP is de relatie tussen R code en GUI *tweerichtingsverkeer*: aanpassingen aan de GUI veranderen de R code, maar aanpassingen aan de R code veranderen de opties die zijn aangevinkt in de GUI. Op deze manier kan JASP aangestuurd worden zowel door R code als door het menu. De volledige integratie met R hopen we nog dit jaar te bewerkstelligen.
- De JASP GUI geeft *onmiddellijk feedback*. Als de optie 'gemiddelde' in het input paneel wordt aangevinkt, dan wordt in het output paneel het gemiddelde direct toegevoegd aan de overeenkomstige tabel; als de optie 'gemiddelde' wordt uitgevinkt en de optie 'mediaan' wordt aangevinkt, dan is het effect van deze acties ook direct zichtbaar: het gemiddelde wordt vervangen door de mediaan. Op deze manier wordt de student uitgenodigd om te ontdekken wat bepaalde statistische procedures doen.
- JASP is gestoeld op het principe van *'progressive disclosure'*. In eerste instantie wordt de gebruiker geconfronteerd met een relatief eenvoudige output – vaak een enkele tabel met de belangrijkste resultaten. Aanvullende informatie is binnen handbereik, maar vereist dat de gebruiker deze eerst selecteert. Dit maakt het duidelijker wat het verband is tussen input en output; tevens voorkomt het dat de gebruiker wordt overspoeld met resultaten en door de bomen het bos niet meer ziet.
- Oudere statistiekprogramma's zijn vaak onoverzichtelijk – ze hebben gaandeweg steeds meer functionaliteit toegevoegd zonder voldoende aandacht te besteden aan de organisatie. In JASP streven we naar *overzichtelijkheid* door complexere functionaliteit aan te bieden in aparte tabbladen en in modules die apart geactiveerd moeten worden.
- In JASP kunnen alle output elementen –figuren, tabellen– voorzien worden van *annotaties*. Op deze manier kunnen docenten data-specifieke uitleg verschaffen, of studenten vragen stellen. De annotaties staan sinds kort ook LaTeX code toe (zie bv. de banner bovenaan dit artikel).
- JASP is *open-source*. Dit betekent dat iedereen JASP kan installeren en gebruiken, zonder enige verplichting, financieel of anderszins. We maken van de gelegenheid gebruik om het verdienmodel van SPSS aan de kaak te stellen. De jaarlijkse SPSS campus licenties zijn *relatief* laag geprijsd, zodat universiteiten het programma blijven gebruiken en aldus hun studenten statistisch afhankelijk maken van SPSS. Deze studenten gaan later bij bedrijven en zorginstellingen pleiten voor de aanschaf van SPSS – maar in die niet-academische context kost een enkele SPSS licentie jaarlijks al snel een klein fortuin.[1] Een open-source fanaat zou kunnen beweren dat universiteiten die SPSS inzetten onbewust medeplichtig zijn aan het verspillen van gemeenschapsgeld en het creëren van een dure levenslange verslaving voor hun alumni. Voor alle duidelijkheid: het gaat hier om astronomische bedragen. Op de SPSS website wordt vermeld dat 80% van de universiteiten en colleges in de VS gebruik maakt van SPSS. Dat komt neer op ongeveer 4000 instellingen; als iedere instelling 25.000 dollar per jaar betaalt dan maakt de academische wereld alleen in de VS al 100 miljoen dollar over naar IBM – ieder jaar opnieuw. Deze situatie wordt des te schrijnender in het besef dat SPSS vooral wordt ingezet voor triviale analyses (bv. standaard ANOVA en lineaire of logistische

regressie). Voor meer ingewikkelde analyses gebruiken data professionals andere programma's, zoals bijvoorbeeld `R`, `Python`, en `Julia`.

- JASP is *ontworpen door universiteiten, voor universiteiten*, en sluit derhalve nauw aan bij de huidige onderwijs- en onderzoekspraktijk. Een bescheiden voorbeeld is dat de tabellen in JASP zijn opgemaakt in 'APA' stijl, dus zonder verticale lijnen. Een ander voorbeeld is dat JASP specifieke modules heeft die zijn toegesneden op onderwijs, zoals de 'Learn Bayes' module en de recente 'Learn Stats' module. Verder heeft JASP een ingebouwde 'Bibliotheek' met meer dan 50 voorbeeld-datasets gebaseerd op populaire cursusboeken en wetenschappelijke artikelen.[2] Op de JASP website staat een uitgebreid overzicht van boeken, artikelen, YouTube videos, en ander instructiemateriaal over JASP (https://jasp-stats.org/resources/).
- JASP biedt standaard statistische procedures zoals ANOVA en regressie, maar *bevat ook state-of-the-art analyses*. Deze analyses zijn veelal beschikbaar in specifieke JASP modules (bv. machine-learning, quality control, netwerk-analyse, SEM, auditing, etc.).
- JASP biedt relatief veel Bayesiaanse methoden (Ly e.a., 2021), en zo kan de misvatting ontstaan dat JASP specifiek Bayesiaanse software is. Deze indruk is onjuist – JASP kent een grote verscheidenheid aan *zowel Bayesiaanse als frequentistische methoden*.
- In JASP is veel aandacht besteed aan de *kwaliteit van de figuren*. Het streven is om figuren aan te bieden die elegant, inzichtelijk, en 'publication-ready' zijn. De figuren kunnen intern worden bewerkt, maar ze kunnen ook worden bewaard (bv. als pdf of in Powerpoint formaat) voor verdere externe bewerking.
- JASP wordt ontwikkeld door een toegewijd team van software ontwikkelaars, postdocs, en promovendi. Dit team staat in *direct contact met de gebruikers* door middel van de JASP GitHub repository. Bug reports en feature requests worden op deze manier efficiënt verwerkt.
- Sinds kort is de *JASP Cooperative* opgezet, een consortium van kennisinstellingen die financieel de krachten bundelen zodat JASP ook in de toekomst actief kan worden doorontwikkeld. De huidige lijst van ondersteunende instellingen is te vinden op https://jasp-stats.org/cooperative-institutional-members/.

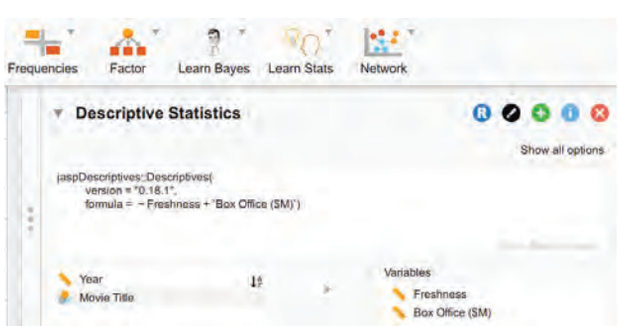
Deze lijst eigenschappen is niet uitputtend, en zoals alle software kan JASP pas op waarde worden geschat op het moment dat het in de praktijk wordt toegepast. Om een meer volledige indruk te geven van JASP presenteren we daarom vier voorbeeld-analyses m.b.t. de volgende onderwerpen: (1) de beschrijvende statistiek, (2) de Bayesiaanse *t*-test, (3) de Learn Stats module, en (4) netwerk-analyses.

Voorbeeld 1: Beschrijvende Statistiek

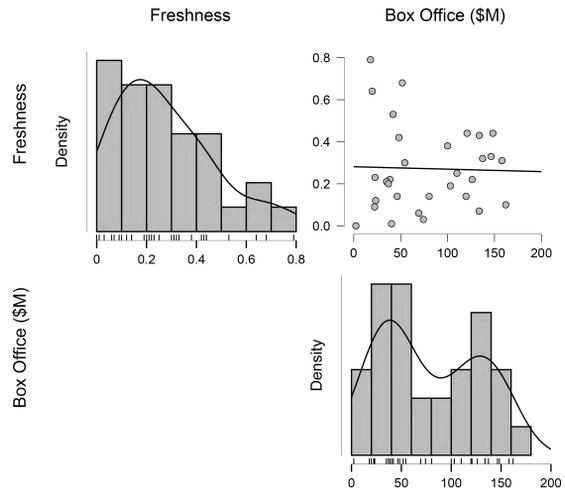
Bij aanvang kan de gebruiker kiezen voor het invoeren van data (middels een interne data editor) of het openen van een databestand. In dit voorbeeld openen we een databestand uit de JASP Data Library. Vanuit het hoofdmenu navigeren we middels `Open` → `Data Library` → `4. Regression` naar de 'Adam Sandler' dataset. Deze dataset bevat een lijst van 31 films met Adam Sandler uit de periode 2000-2015. Na openen worden de data getoond in een spreadsheet formaat, met een rij per film (zie Figuur 2). De relevante afhankelijke variabelen zijn 'Freshness' (d.w.z. een inschatting van de kwaliteit van de film zoals beschikbaar op de website *Rotten Tomatoes*, op een schaal van 0 tot 1) en 'Box Office ($M)' (d.w.z. hoeveel miljoen dollar de film opleverde aan de kassa).

Voor een eerste indruk van de data kiezen we

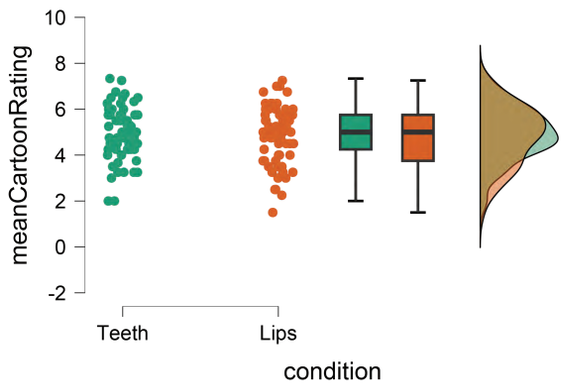
Figuur 2: Spreadsheet overzicht van de Adam Sandler data in JASP. Bron: JASP.



Figuur 3: Drag-en-drop selectie van de Adam Sandler variabelen 'Freshness' en 'Box Office ($M)' ten behoeve van beschrijvende statistiek. Bron: JASP.



Figuur 4: Correlatieplot van de Adam Sandler data in JASP. Het meest rechtse datapunt in het spreidingsdiagram is voor de film 'Grown Ups', die in 2010 maar liefst $162 miljoen opleverde en slechts 10% 'vers' scoorde. Bron: JASP.



Figuur 5: Gemiddelde "grappigheid-ratings" van vier cartoons, apart voor 65 proefpersonen die een pen tussen de tanden hielden ("Teeth") en 65 proefpersonen die een pen tussen de lippen hielden ("Lips"). Van links naar rechts: een regenwolk plot (met jitter), box plots, en niet-parametrische dichtheidsschatters. Bron: JASP.

`Descriptives` op het lint bovenaan. Een gedeelte van de bijbehorende interface wordt getoond in Figuur 3. De gebruiker kan met 'drag-en-drop' een selectie maken onder de beschikbare variabelen. Van de geselecteerde variabelen wordt dan direct een tabel gegenereerd met onder andere het gemiddelde en de standaard deviatie. De tabel kan naar believen worden aangepast.

We openen het tabblad `Basic plots` en vinken de optie `Correlation plots` aan. Het resultaat is weergegeven in Figuur 4. In deze figuur laten de histogrammen zien dat de variabelen waarschijnlijk niet normaal verdeeld zijn; het spreidingsdiagram suggereert dat relatief goede films van Adam Sandler *niet* meer publiek trekken dan relatief slechte films.

Voorbeeld 2: Een Bayesiaanse t-test van de Facial Feedback Hypothese

In 2016 waren we betrokken bij een 'Registered Replication Report' van de facial feedback hypothese (Wagenmakers e.a., 2016). Kort gezegd ging het erom of proefpersonen cartoons grappiger vinden wanneer ze een pen tussen hun tanden klemden (d.w.z. met de gezichtsspieren in de lachstand) in plaats van tussen hun lippen (d.w.z met de gezichtsspieren in de pruilstand). Het totale project omvatte 17 replicatie-experimenten met in totaal 1894 proefpersonen; hier analyseren we enkel de data van het experiment uitgevoerd aan de Universiteit van Amsterdam. Na toepassing van gepreregistreerde exclusie-criteria behelsde dit experiment 130 proefpersonen die ieder vier cartoons scoorden op grappigheid; 65 proefpersonen waren toegewezen aan de 'tanden' conditie, en de andere 65 waren toegewezen aan de 'lippen' conditie. Beschrijvende statistiek laat zien dat de gemiddelde grappigheidsrating 4,94 was in de tanden conditie (SD = 1,14), en 4,79 in de lippen conditie (SD = 1,30). Een frequentistische onafhankelijke *t*-toets geeft een *p*-waarde van 0,48 en een *t*-waarde lager dan 1. Dit geeft natuurlijk geen aanleiding om de nul-hypothese te verwerpen. Maar in hoeverre bieden deze data nu ondersteuning *voor* de nul-hypothese? In andere woorden: is er afwezigheid van evidentie, of is er evidentie voor afwezigheid? Om dit te bepalen kunnen we in JASP een Bayesiaanse *t*-toets uitvoeren.

Via `T-Tests` (op het lint bovenaan) → `Bayesian` → `Independent Samples T-Tests` activeren we de Bayesiaanse *t*-test interface. Met drag-en-drop selecteren we de relevante variabelen (d.w.z. 'condition' en 'meanCartoonRating'). Het aanvinken van de optie 'Raincloud plots' resulteert in Figuur 5.

Vervolgens voeren we de Bayesiaanse toets uit. Dit vereist een specificatie van de populatie effect sizes δ die plausibel geacht worden onder de alternatieve hypothese H_1. We kunnen eenvoudigweg de standaard opties gebruiken (d.w.z. *default* prior verdelingen voor het toetsen van hypotheses), maar het is interessanter om wat dieper na te denken over de beschikbare achtergrondkennis. Het gaat hier tenslotte om een replicatie-experiment, en het is derhalve verdedigbaar om de prior verdeling voor effect size te laten bepalen door de posterior verde-

ling van het oorspronkelijke experiment (Verhagen en Wagenmakers, 2014). Die posterior verdeling is bij benadering normaal verdeeld met gemiddelde van 0,4 en een standaard deviatie van 0,25.

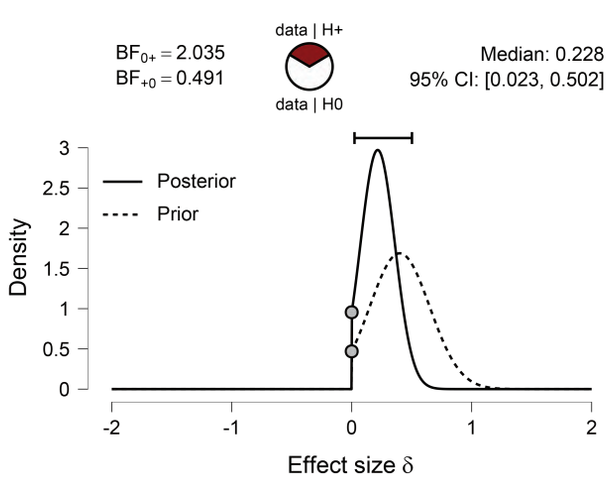
In de GUI openen we het tabblad `Prior` en definiëren we de gewenste normaalverdeling. Tenslotte geven we aan dat de alternatieve hypothese een richting heeft (d.w.z cartoons zouden grappiger zijn met een pen tussen de tanden, niet minder grappig). Het aanvinken van de optie `Plots` → `Prior and posterior` resulteert in Figuur 6. Deze figuur biedt een relatief compleet overzicht van de Bayesiaanse conclusies. We zien met name dat de *Bayes factor*, BF_{0+}, gelijk is aan 2,035. Dit betekent dat de null hypothese (d.w.z. $H_0 : \delta = 0$) de geobserveerde data ongeveer twee maal beter voorspelt dan de alternatieve hypothese (d.w.z. $H_+ : \delta \sim N(0, 4, 0, 25^2)I(0, \infty)$). Evidentie van een dergelijke impact zou de prior kans voor H_0 verhogen van 0,50 naar $2,035/3,035 \approx 0,67$ – nauwelijks de moeite van het vermelden waard, vond Jeffreys (1961, Appendix B). Voor het vergelijken van deze twee specifieke hypotheses is er dus eerder sprake van 'afwezigheid van evidentie' dan van 'evidentie voor afwezigheid'.

Het produceren van dergelijke analyses is in JASP letterlijk secondewerk – vaak volstaat het aanvinken van een optie.

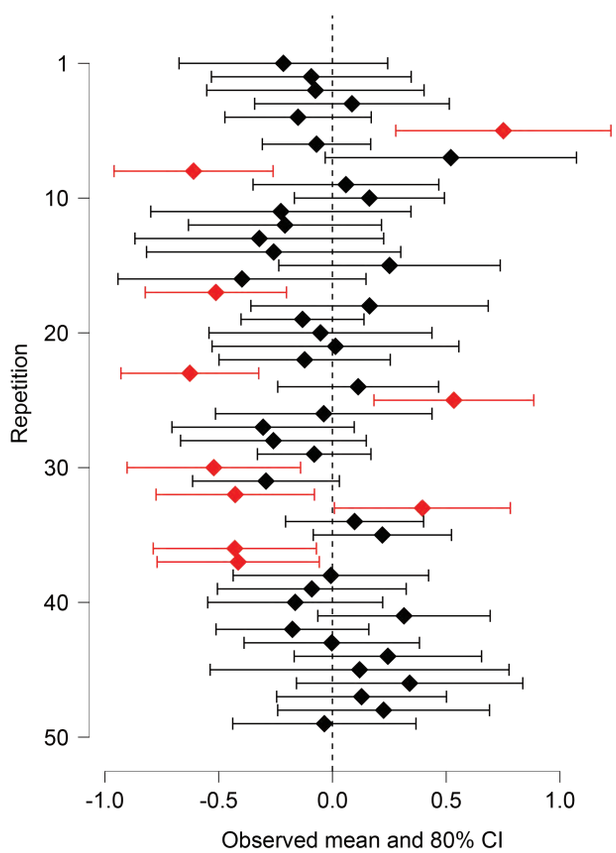
Voorbeeld 3: De Learn Stats Module

Recentelijk hebben we aan JASP een 'Learn Stats' module toegevoegd. Deze module bevat een coherente set demonstraties die docenten kunnen helpen bij het uitleggen van statistische sleutelbegrippen, te weten de normaalverdeling, de binomiaalverdeling, de centrale limietstelling, de standaardfout, beschrijvende statistiek, steekproef-variabiliteit, *p*-waardes, betrouwbaarheidsintervallen en effectgroottes. De Learn Stats module heeft ook een statistische beslisboom.

Het achterliggende idee is dat veel sleutelbegrippen geassocieerd zijn met een visuele representatie. De demonstraties in de Learn Stats module proberen die visuele representaties voor de student op



Figuur 6: Prior en posterior verdelingen van effect size voor een van de replicatie-experimenten gerapporteerd in Wagenmakers e.a. (2016). Bron: JASP.



Figuur 7: Een visualisatie uit de Learn Stats module om het concept van een betrouwbaarheidsinterval te verduidelijken. Bron: JASP.

de voorgrond te plaatsen. Een eerste voorbeeld: na activatie van de Learn Stats module kiezen we `Confidence Intervals`. We zetten het `Confidence level` op 80% en het aantal `Repetitions` op 50. Het resultaat is weergegeven in Figuur 7. De figuur onderstreept de correcte definitie van een 80% betrouwbaarheidsinterval, namelijk dat het gaat om een interval gegenereerd door een procedure die in 80% van de experimenten de ware waarde omsluit.

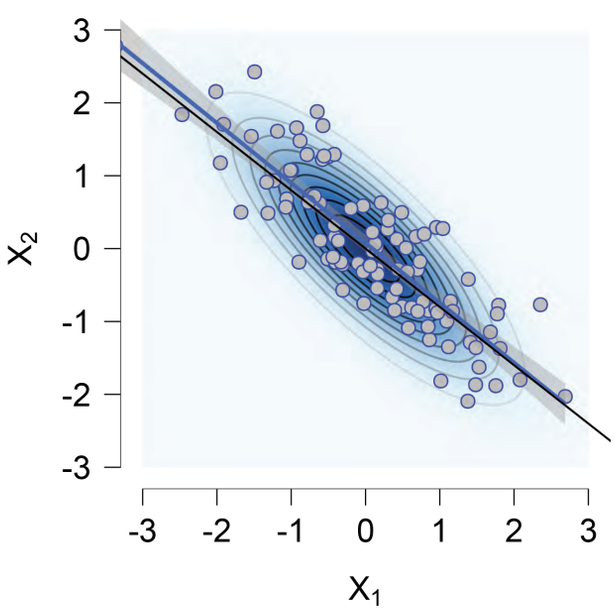
Een tweede voorbeeld: na activatie van de Learn Stats module kiezen we `Effect Sizes`, en klikken we op `Pearson correlation coefficient` ρ. Middels de GUI kan vervolgens een bivariate normaalverdeling worden gespecificeerd; in de bijbehorende figuur wordt deze normaalverdeling met ellipsen weergegeven, eventueel samen met een steekproef van willekeurige grootte en de bijbehorende regressielijn (zie Figuur 8).

In de nabije toekomst hopen we steeds meer onderwijs-specifiek materiaal toe te kunnen voegen aan JASP. Hier ligt ons inziens een grote kans om het statistiekonderwijs nog inzichtelijker en toegankelijker te maken.

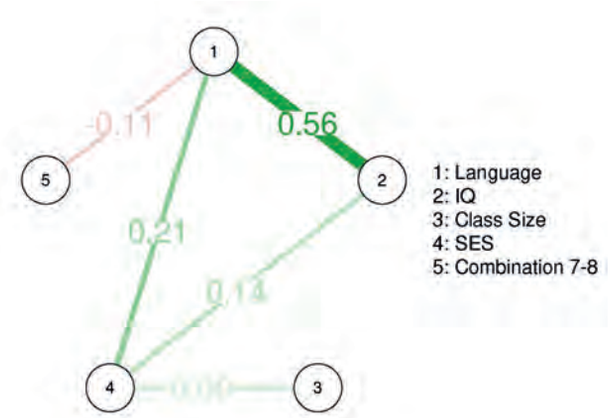
Voorbeeld 4: Een Netwerk-analyse

Netwerk-analyse wordt gebruikt om de unieke relaties tussen variabelen te ontdekken en te kwantificeren. Het belang hiervan kan worden onderstreept met een klassiek voorbeeld: de relatie tussen de verkoop van ijsjes en het aantal verdrinkingen. Deze twee variabelen zijn sterk gecorreleerd; wanneer er veel ijsjes verkocht worden verdrinken er ook meer mensen. De statistische associatie tussen deze twee variabelen komt natuurlijk geheel op conto van een derde variabele: temperatuur. Dit soort 'valse' relaties kunnen worden achterhaald door het berekenen van een netwerk met partiële correlaties. Een dergelijk netwerk toont louter relaties tussen variabelen die niet verklaard kunnen worden door de resterende variabelen in de dataset.

Het voorbeeld hieronder betreft data van 2287 achtste groepers (publiekelijk beschikbaar in het `R`-pakket MASS; Snijders en Bosker, 1999). Deze kinderen hebben een taaltoets gemaakt (Language), er is een IQ test afgenomen (IQ), de sociaal-economische-status van hun familie is gekwantificeerd (SES), er is bijgehouden hoe groot de klassen waren (Class Size), en of het al dan niet combinatieklassen waren (d.w.z. een klas met zowel kinderen in groep 7 als groep 8; Combination 7-8). Een statistisch probleem bij het schatten van de partiële correlaties tussen de variabelen is dat het a priori onzeker is welke netwerkstructuur ertoe doet. Hoe waarschijnlijk is een netwerk waarin alleen IQ en SES verbonden zijn? Hoeveel beter (of slechter) presteert een netwerk als de relatie tussen SES en Language wordt toegevoegd? Met de Bayesiaanse netwerk-analyse



Figuur 8: Een visualisatie uit de Learn Stats module om het concept van de Pearson correlatie coëfficiënt te verduidelijken. De populatie-coëfficiënt is hier –0,80; de ellipsen beschrijven de vorm van de corresponderende bivariate normaalverdeling; ieder van de 100 cirkels is een trekking uit de populatieverdeling. Bron: JASP.



Figuur 9: De sterkte van relaties tussen variabelen in een netwerkstructuur. De gewichten zijn bepaald door alle mogelijke netwerken tegelijkertijd in ogenschouw te nemen. Bron: JASP.

in JASP (Huth e.a., 2023) wordt achter de schermen de ruimte van *alle mogelijke netwerkmodellen* op een stochastische wijze verkend. Netwerkmodellen die de data relatief goed voorspellen winnen aan waarschijnlijkheid, terwijl netwerkmodellen die de data relatief slecht voorspellen aan waarschijnlijkheid inboeten. De uiteindelijke conclusies zijn gestoeld op een gewogen gemiddelde van alle netwerkmodellen, waar de gewichten bepaald worden door de waarschijnlijkheid van de modellen. De gewogen conclusies voor de voorbeelddata zijn samengevat in Figuur 9.

De analyse laat zien dat IQ en Language positief gerelateerd zijn. Daarnaast zijn er positieve relaties tussen SES en zowel Language, IQ, als Class Size. Voorts bestaat er nog een kleine negatieve relatie tussen Combination 7-8 en Language. Deze bevinding zou aanleiding kunnen geven om te onderzoeken of het zitten in een combinatieklas de taalontwikkeling negatief beïnvloedt. Let wel, deze analyse geeft geen evidentie voor een causaal verband en kan dus ook niet zo geïnterpreteerd worden.

Waarom Niet 'Gewoon' R Gebruiken?

De gemiddelde STAtOR lezer zal zich wellicht afvragen wat de toegevoegde waarde is van JASP ten opzichte van R. Het is zeker waar dat voor sommige opleidingen (met name wiskunde en informatica) de voorkeur uit zal gaan naar R. Daarbij dient wel de kanttekening geplaatst te worden dat wij voor onze eigen analyses JASP prefereren boven R. De belangrijkste reden is efficiëntie: de analyses in JASP zijn met een paar muisklikken voltooid, incluis figuren en tabellen, terwijl wij in R eerst naarstig moeten zoeken naar welk pakket ook al weer de gewenste functionaliteit biedt, wat de argumenten zijn voor de relevante functie, en hoe de resultaten het best in een figuur kunnen worden weergegeven. Voor ons kosten dezelfde analyses *seconden* in JASP en *uren* in R.

Voor andere opleidingen lijkt een tweetrapsmodel mogelijk: alle studenten beginnen met JASP, en zij met een methodologische interesse schakelen na een of twee jaar over naar R. In de toekomst gaan we JASP verder uitbouwen om deze transitie te vergemakkelijken. Voor andere opleidingen (bv. geneeskunde) is R geen voor de hand liggende optie, maar het is wel noodzakelijk dat studenten de literatuur kunnen begrijpen en eventueel zelf ook toetsen uit kunnen voeren. JASP is voor deze groep gebruikers ideaal.

JASP is op dit moment een volwaardige vervanging voor SPSS, en een nuttige aanvulling op R. We zijn optimistisch dat JASP in de toekomst uit kan groeien tot een een breed gedragen interuniversitair project, met aanzienlijke voordelen voor zowel onderwijs als onderzoek.

Voetnoten

1) De precieze bedragen staan niet vermeld op de SPSS website en zijn moeilijk te achterhalen.
2) De bibliotheek is tevens online beschikbaar op https://johnnydoorn.github.io/DataLibraryBookdown/.

Literatuur

K. Huth e.a. „Bayesian analysis of cross-sectional networks: A tutorial in R and JASP". In: *Advances in Methods and Practices in Psychological Science* (2023). In press.

H. Jeffreys. *Theory of Probability*. 3de ed. Oxford, UK: Oxford University Press, 1961.

A. Ly e.a. „Bayesian Inference With JASP". In: *The ISBA Bulletin* 28 (2021), p. 7–15.

T. A. B. Snijders en R. J. Bosker. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, UK: Sage, 1999.

A. J. Verhagen en E.-J. Wagenmakers. „Bayesian Tests to Quantify the Result of a Replication Attempt". In: *Journal of Experimental Psychology: General* 143 (2014), p. 1457–1475.

E.-J. Wagenmakers e.a. „Registered Replication Report: Strack, Martin, & Stepper (1988)". In: *Perspectives on Psychological Science* 11 (2016), p. 917–928.

Eric-Jan Wagenmakers is hoogleraar Bayesiaanse Methodologie bij de afdeling Psychologische Methodenleer op de Universiteit van Amsterdam.
E-mail: EJ.Wagenmakers@gmail.com

Johnny van Doorn is Universitair Docent bij de afdeling Psychologische Methodenleer op de Universiteit van Amsterdam.
E-mail: J.B.vanDoorn@uva.nl

Don van den Bergh is postdoc bij de afdeling Psychologische Methodenleer op de Universiteit van Amsterdam.
E-mail: d.vandenbergh@uva.nl

De auteurs zijn nauw betrokken bij de ontwikkeling en verdere verspreiding van JASP, en het is derhalve onvermijdelijk dat deze bijdrage herinneringen oproept aan de onsterfelijke reclame-slogan "Wij van Wc-eend adviseren: Wc-eend". Voor een eigen oordeel kan de sceptische lezer JASP gratis downloaden van jasp-stats.org. We zijn alle leden van het JASP team erkentelijk voor hun inzicht en inzet door de jaren heen.