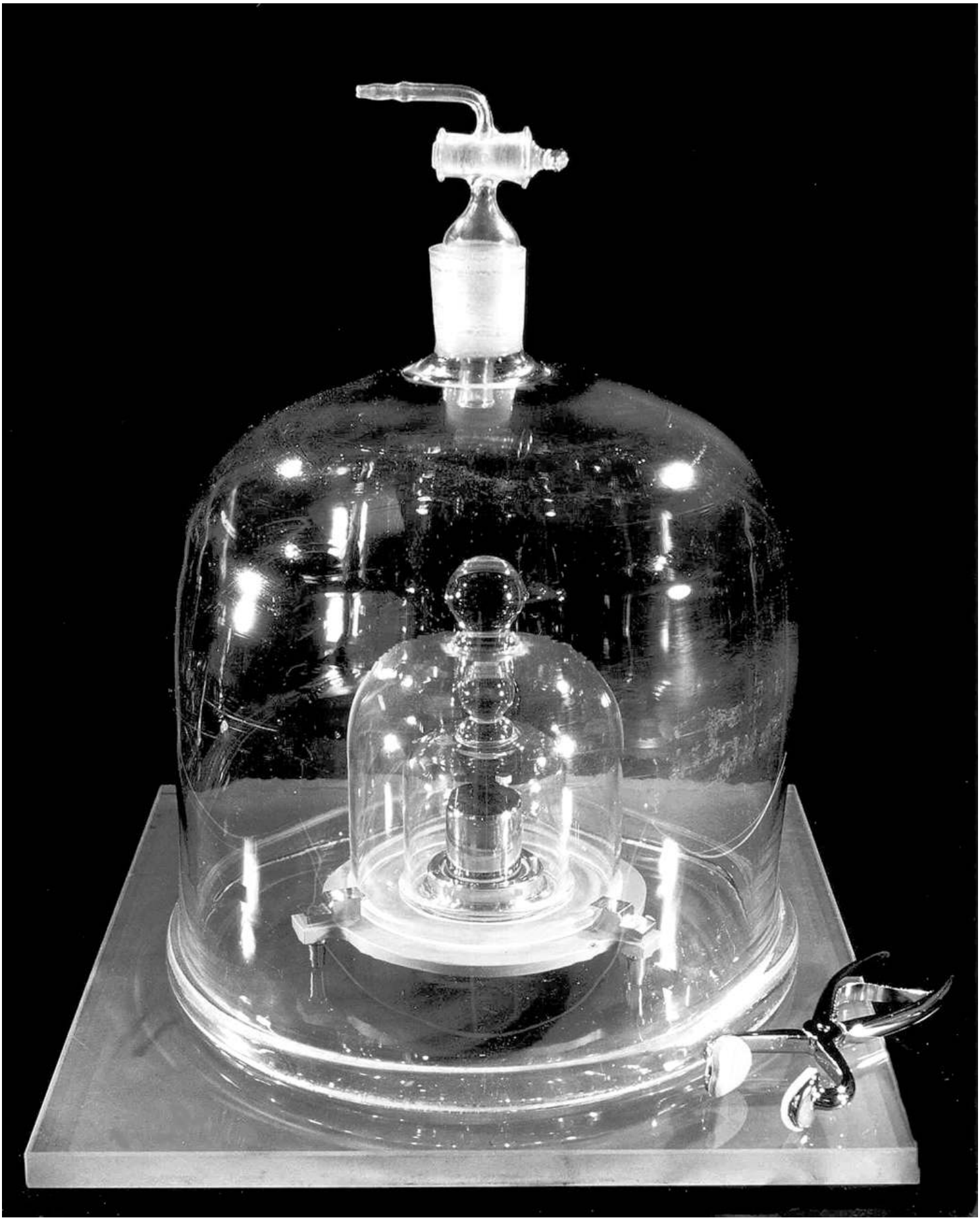


Sandjai Bhulai
Corresponderend auteur 'Weegschema's'
Faculteit der Exacte Wetenschappen
Vrije Universiteit
De Boelelaan 1081a
1081 HV Amsterdam
sbhulai@few.vu.nl

Marco Bijvank
Corresponderend auteur 'Drinkwater'
Faculteit der Exacte Wetenschappen
Vrije Universiteit
De Boelelaan 1081a
1081 HV Amsterdam
mbijvank@few.vu.nl

Misja Nuyens
Corresponderend auteur 'Selectie-effecten'
Faculteit der Exacte Wetenschappen
Vrije Universiteit
De Boelelaan 1081
1081 HV Amsterdam
mnuyens@few.vu.nl

Studiegroep wiskunde met de industrie

Brainstormen voor bruikbare wiskunde II

De studiegroep Wiskunde met de Industrie laat zien hoe bruikbaar wiskunde al dan niet is voor de samenleving. Hoe kunnen wiskundige disciplines commercieel worden ingezet, wat hebben we aan de nieuwste ontwikkelingen in de statistiek, wat is de betekenis van de steeds maar groeiende kennis van het modelleren met differentiaalvergelijkingen? Voor deze betekenis van wiskunde moet steeds opnieuw worden gestreden. Het kan zelfs voorkomen dat een oplossing wel bestaat, maar domweg te ingewikkeld is om door een bedrijf geaccepteerd te worden. In ieder geval levert wiskundige common sense een waardevolle alternatieve invalshoek aan bedrijven.

In dit tweede deel van het verslag van de studiegroep wiskunde met de industrie 2005 worden problemen aangeleverd door het Nederlands Meetinstituut, de KLM en het Forensisch Instituut besproken. Het eerste deel verscheen in het maartnummer.

Optimale weegschema's (Sandjai Bhulai)

De kilogram is de laatste fysische grootheid die nog gedefinieerd is in termen van een tastbaar object: het is de massa van een platinum-iridium cilinder, vervaardigd in 1889, die bij het Bureau des Poids et des Mesures in Frankrijk bewaard wordt. Om precies te zijn is de kilogram gedefinieerd als de massa van dit object net nadat het gewassen is. Platinum-iridium absorbeert koolhydraten uit de atmosfeer, waardoor de cilinder regelmatig gewassen moet worden om het gewicht te verwijderen dat erop neergeslagen is.

In Nederland is het Nederlands Meetinstituut (NMi) in letterlijke zin verantwoordelijk voor de kilogram: de Nederlandse standaardkilogram, de platinum-iridium cilinder nummer 53, wordt op NMi terrein onder zorgvuldig gereguleerde omstandigheden bewaard.

De cilinder reist af en toe naar Parijs om daar vergeleken te worden met de internationale standaard. Uiteraard draagt het NMi zorg voor meer dan alleen de veilige opslag van de kilogram: de afdeling 'massameting' is, onder meer, ook verantwoordelijk voor zeer nauwgezette kalibratie van gewichten.

Figuur 1 De platinum-iridium cilinder die precies 1 kilogram definieert. De kilogram is opgeslagen bij het Bureau des Poids et des Mesures in Frankrijk.

Deze gewichten worden gebruikt om de massa's van gewichten van lagere standaarden te bepalen, die bijvoorbeeld door supermarkten gebruikt worden om hun weegschalen te kalibreren, of door doping laboratoria, waar zeer kleine massa's met grote nauwkeurigheid vastgesteld moeten worden.

Voor de kalibratie van gewichten gebruikt het NMi gewichten van roestvrij staal, waarvan de massa's met zeer grote nauwkeurigheid bepaald moeten worden. Om resultaten met voldoende precisie te verkrijgen, moeten de metingen gecorrigeerd worden voor effecten als de opwaartse luchtdruk en de verschillen in de positie van het massamiddelpunt voor verschillende gewichten!

Het probleem

De Nederlandse kilogram wordt gebruikt als startpunt in het bepalen van de massa's van de gewichten van de hoogste kwaliteit roestvrij staal: eerst wordt de massa van een roestvrij stalen kilogram bepaald door directe vergelijking met de nationale standaard. In de tweede stap wordt de twee roestvrij stalen kilogram gebruikt en niet de nationale platinum-iridium kilogram, opdat externe invloeden minimaal effect hebben op de nationale standaard. De roestvrij stalen kilogram wordt vervolgens gebruikt om een roestvrij stalen set van gewichten te kalibreren bestaande uit een gewicht met nominale massa van 500 gram, twee van 200 gram en twee van 100 gram, zodat de honderdvouden van 1000 gram naar 100 gram gedekt zijn. Het verschil tussen de nominale massa en de echte massa van de gewichten is extreem klein. De gewichten van deze set worden vervolgens gebruikt om massa's van andere sets met andere ranges te bepalen.

Om de massa van een individueel gewicht te bepalen, kunnen we het vergelijken met het standaardgewicht met gelijke nominale massa waarvan de feitelijke massa bekend is met voldoende nauwkeurigheid – tenzij we natuurlijk de massa op het hoogste niveau van precisie willen bepalen. Massa-metrologieinstituten gebruiken zogenaamde *weegschema's* om dit probleem op te lossen. Een weegschema voor de set van gewichten van 1000g naar 100g van het NMi bestaat uit paren van combinaties van gewichten uit de collectie van de 6 gewichten. Een schema kan bijvoorbeeld een vergelijking van een van de 200g gewichten met de twee

100g gewichten in zich hebben. Voor elk paar in het schema worden de verschillen in massa tussen de twee paren bepaald door middel van de STS-procedure, die hieronder beschreven zal worden. Om voldoende precisie te garanderen, worden de paren in een weegschema zo gekozen dat ze gelijke nominale massa hebben.

Het is in theorie mogelijk om de vijf onbekende massa's te bepalen door vijf geschikte metingen van massaverschillen uit te voeren. In de praktijk worden er echter meetfouten gemaakt en moeten er meer metingen —niet noodzakelijk allemaal met verschillende combinaties van gewichten— verricht worden om tot nauwkeurigere schattingen van de ware massa's te komen. De schattingen van de massa's van de gewichten en de onzekerheid in deze massa's kunnen verkregen worden uit een overbepaald stelsel met behulp van de kleinste-kwadratenanalyse.

Gedurende de 'Studiegroep Wiskunde met de Industrie', werd de volgende vraag door het NMi gesteld:

Wat is een optimaal weegschema voor de set van gewichten van het NMi, dat wil zeggen, een weegschema dat de onzekerheid in de geschatte massa's minimaliseert onder de voorwaarde dat het aantal metingen kleiner is dan een gegeven getal?

De STS-procedure

Zoals eerder vermeld is, wordt voor elk paar in een weegschema het verschil in massa bepaald door middel van de STS-procedure. Neem, om de STS-procedure te illustreren, twee sets van gewichten met ongeveer dezelfde massa. De balans om de massaverschillen te meten is een balans met enkelvoudige arm, waarmee het massaverschil tussen twee sets gewichten als volgt bepaald wordt. De eerste set, die we de standaardset (S) noemen, wordt geplaatst op de balans, die daarna op 0 wordt gesteld. De set S wordt verwijderd, en daarna nog eens op de balans geplaatst, waarna de eerste meting, x_0, afgelezen wordt.

Helaas is x_0 in het algemeen niet gelijk aan nul; in de praktijk wordt er een drift geobserveerd tussen opeenvolgende metingen, ook worden er meetfouten gemaakt. Vervolgens wordt de set S verwijderd en de tweede set, die we de testset (T) noemen, wordt op de balans geplaatst, waarna de tweede meting, x_1, afgelezen wordt. De metingen worden nu door S en T te alterneren voorgezet; dit verklaart de naam *STS-procedure*. Als de drift niet al te wild fluctueert tussen opeenvolgende metingen, dan kan dit geëlimineerd worden door gebruik te maken van de STS-procedure.

De metingen worden zo veel mogelijk herhaald om tot betrouwbare resultaten te komen. In de huidige opstelling die het NMi nu gebruikt, worden de gewichten handmatig op elkaar gestapeld, wat een tijdrovende procedure is. In de praktijk is het daarom zelden mogelijk om meer dan twintig metingen te verrichten zonder onderbrekingen.

Optimale weegschema's

Na het modelleren van de STS-metingen kunnen de optimale weegschema's voor de set van gewichten van het NMi bepaald worden. De balans in de STS-procedure kan gebruikt worden om nauwkeurig hele kleine verschillen in massa's te meten, maar kan niet met voldoende precisie voor grote massaverschillen gebruikt worden. Dit impliceert dat de test en de standaardmassa's in een STS-meting dezelfde nominale massa moeten hebben, wat een sterkte restrictie in het aantal mogelijke combinaties met zich meebrengt. In feite zijn er voor de set van gewichten van het NMi slechts tien mogelijke combinaties (of: *weegvergelijkingen*) mogelijk. Deze worden in tabel 1 weergegeven.

Bij een gegeven weegschema kan er een reeks van STS-metingen worden gedaan voor elke weegvergelijking in het schema. Zodoende krijgen we een reeks van $\Delta\mu_i$, zoals gedefinieerd in (3), voor elk van deze combinaties. De $\Delta\mu_i$ kunnen gecombineerd worden in een standaard lineair model. Met behulp van statistische technieken kan er een schatting gegeven worden van de ware massa's van de gewichten uitgedrukt in de gemeten waarden van de $\Delta\mu_i$, alsmede een schatting van de onzekerheid. In deze procedure worden de $\Delta\mu_i$ gebruikt, en niet het (ongewogen) gemiddelde, zoals in de huidige procedure bij het NMi gedaan wordt. Deze aanpassing leidt tot betere schattingen.

Veronderstel nu dat we een optimaal schema willen bereken-

Het modelleren van STS-metingen

Stel dat we $k+1$ STS-metingen verrichten met de sets S, met massa m_S, en T, met massa m_T. Laat x_i de i-de meting zijn, voor $0 \leq i \leq k$. We veronderstellen dat

$$x_i = 1_{\{i \text{ oneven}\}}(m_T - m_S) + D(i) + V_i,$$

waar V_i een meetfout is, die we proportioneel aan de totale massa op de balans veronderstellen. Om precies te zijn nemen we aan dat V_i normaal verdeeld is met parameters

$$V_i \sim N(0, \alpha^2 m_S^2),$$

waarbij $\alpha \in \mathbf{R}$ een onbekende constante is. De term $D(i)$ beschrijft de drift van de balans. We gaan er vanuit dat

$$\frac{D(i+1) + D(i-1)}{2} - D(i) \approx 0,$$

voor alle $1 \leq i \leq k-1$, wat consistent is met aannamen die doorgaans door het NMi gemaakt worden. Aan deze eis wordt natuurlijk voldaan wanneer D lineair is.
Definieer voor $1 \leq i \leq k-1$

$$\Delta\mu_i = \frac{(-1)^{i+1}}{m_S}\left(x_i - \frac{x_{i+1} + x_{i-1}}{2}\right) \approx \frac{m_T - m_S}{m_S} + E_i,$$

waarbij

$$E_i = \frac{1}{m_S}\left(V_i - \frac{V_{i+1} + V_{i-1}}{2}\right) \sim N(0, \alpha^2),$$

onafhankelijk van m_S. Observeer dat, gebruikmakend van (2), de drift geëlimineerd wordt.
In de procedure die nu door het NMi gebruikt wordt, wordt het gemiddelde van de $\Delta\mu_i$ gebruikt als een benadering voor $\frac{m_T - m_S}{m_S}$. We stellen een procedure voor die deze tussenliggende stap overbodig maakt.

Weegvergelijking	Standaard	Test
1	1000	500, 200, 200•, 100
2	1000	500, 200, 200•, 100•
3	500	200, 200•, 100
4	500	200, 200•, 100•
5	200, 100	200•, 100•
6	200, 100•	200•, 100
7	200	200•
8	200	100, 100•
9	200•	100, 100•
10	100	100•

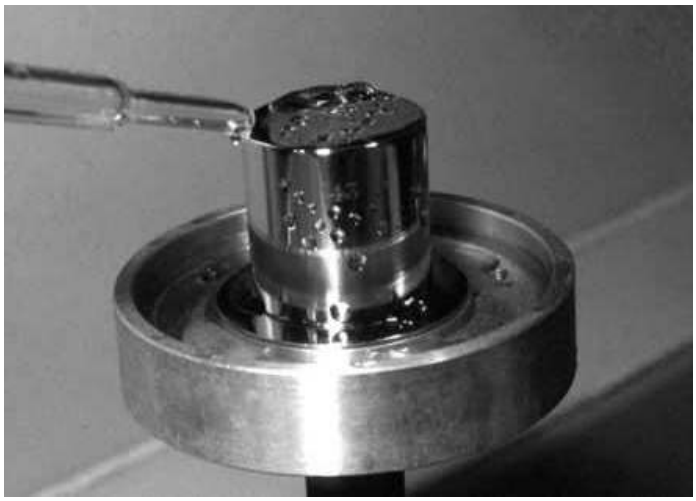
Tabel 1 Mogelijke combinaties van gewichten in nominale massa's genoteerd. De notatie 200 en 200• wordt gebruikt om het onderscheid tussen de twee gewichten met nominale massa 200g te maken. Analoog geldt dit ook voor de twee gewichten met nominale massa 100g.

en met een gegeven aantal weegvergelijkingen, die niet allemaal noodzakelijkerwijs verschillend zijn: een meting herhalen van dezelfde vergelijking in een weegschema kan nuttig zijn, omdat herhalingen extra, onafhankelijke informatie verschaffen. Als deze vergelijkingen uit tabel 1 moeten komen, dan is het aantal weegschema's bestaande uit N vergelijkingen begrensd door 10^N (deze grens is uiteraard niet scherp). Het gevolg hiervan is dat het mogelijk is om de onzekerheid in de massa's van de gewichten door te rekenen voor alle mogelijke weegschema's bestaande uit maximaal 14 vergelijkingen —het maximum aantal opgelegd door het NMi— met behulp van Matlab op een gewone computer.

Het huidige weegschema van het NMi kent acht weegvergelijkingen gelezen uit de tien mogelijkheden in tabel 1. De door het NMI gebruikte vergelijkingen zijn

$$1, 2, 3, 4, 5, 6, 8, 9.$$

Het is onduidelijk hoe dit weegschema, dat dateert uit een tijd waarin alle mogelijke schema's doorrekenen niet mogelijk was, tot stand is gekomen. Met behulp van de hedendaagse moderne computers is het mogelijk om aan te tonen dat dit schema niet optimaal is. Dus het is mogelijk een ander weegschema met acht weegvergelijkingen te construeren zodanig dat de onzekerheid in de geschatte massa's van de gewichten kleiner is. Tabel 2 geeft de optimale schema's bestaande uit N combinaties ($8 \leq N \leq 14$) weer, waarbij aangenomen wordt dat voor elke weegvergelijking er twintig STS-metingen verricht worden.



Figuur 2 Het wassen van de platinum-iridium cilinder (uit: Davis R.S., Possible new definitions of the kilogram, Phil. Trans. Royal Soc. A, 2005, 363, 2249-2264)

N	optimaal weegschema
8	1, 1, 2, 4, 7, 8, 9, 10
9	1, 1, 2, 3, 4, 7, 8, 9, 10
10	1, 1, 2, 2, 3, 4, 7, 8, 9, 10
11	1, 1, 1, 2, 2, 3, 4, 7, 8, 9, 10
12	1, 1, 1, 2, 2, 3, 4, 7, 8, 9, 10, 10
13	1, 1, 1, 2, 2, 3, 4, 7, 8, 9, 9, 10, 10
14	1, 1, 1, 2, 2, 3, 4, 4, 7, 8, 8, 9, 10, 10

Tabel 2 Optimaal weegschema bestaande uit vergelijkingen uit tabel 1

De onzekerheid in het weegschema van het NMi is $1.1812\,\alpha^2$, waarbij α de constante is in (1), terwijl de onzekerheid in het optimale schema met 8 vergelijkingen $0.8468\,\alpha^2$ is. Dit betekent dus dat er een reductie van ongeveer 28% in onzekerheid bewerkstelligd kan worden zonder extra metingen te verrichten. Als het aantal weegvergelijkingen in het schema verhoogd wordt tot 14, wat overigens extra werk met zich meebrengt, dan kan de onzekerheid tot $0.4655\,\alpha^2$ teruggebracht worden, een reductie van ongeveer 63%. Zoals wel te verwachten is, neemt de onzekerheid af naarmate het aantal weegvergelijkingen in het schema toeneemt.

Het is opvallend dat in geen van de optimale schema's de vergelijkingen 5 en 6 voorkomen, de enige vergelijkingen waarbij de standaardset uit meer dan één gewicht bestaat. Het schema van het NMi bevat deze vergelijkingen wel. In feite bevat geen enkele oplossing binnen 1% van de optimale de vergelijkingen 5 en 6, wat impliceert dat dit geen effect van afrondingsfouten is. In plaats van de vergelijkingen 5 en 6, bevat het optimale schema vergelijkingen 7 en 10. Het is eenvoudig na te gaan dat deze samen dezelfde informatie verschaffen als de vergelijkingen 5 en 6 samen. In praktische situaties leidt dit tot een reductie in onzekerheid, vanwege het feit dat vergelijkingen 7 en 10 minder opstapelingen van gewichten vereisen.

Het verhogen van het aantal weegvergelijkingen is niet de enige manier om de onzekerheid te reduceren in de schattingen van de ware massa's. In de nabije toekomst stapt het NMi over op automatische balansen die de STS-metingen minder afhankelijk maken van handmatige procedures. Bovendien kunnen er zo ook meer metingen verricht worden in een STS-serie.

Het verhogen van het aantal STS-metingen per weegvergelijking kan effectiever zijn dan het verhogen van het aantal weegvergelijkingen: het optimale schema bestaande uit tien vergelijkingen met dertig STS-metingen per vergelijking heeft een kleinere onzekerheid ($0.4358\,\alpha^2$) dan het schema met twaalf vergelijkingen en 25 STS-metingen ($0.4458\,\alpha^2$). In beide gevallen is het aantal metingen in totaal driehonderd. Het verhogen van het aantal STS-metingen per weegvergelijking is echter niet altijd de beste aanpak: 10×28 STS-metingen leiden tot betere schattingen dan 8×35 metingen.

Andere sets van gewichten

Verschillende massa-metrologieinstituten gebruiken verschillende sets van gewichten. Het resultaat voor de set van ge-

wichten van het NMi, dat bestaat uit gewichten met nominale massa 1000g, 500g, 200g (tweemaal) en 100g (tweemaal), kan gegeneraliseerd worden naar sets die door andere massametrologieinstituten gebruikt worden. Het Duitse metrologieinstituut gebruikt bijvoorbeeld een set bestaande uit acht gewichten met 104 mogelijke combinaties. Het berekenen van de onzekerheid voor alle mogelijke schema's was te tijdrovend om binnen de studiegroep door te rekenen. Echter, gebaseerd op de inzichten met de Nederlandse weegschema's werd er besloten om alleen de weegvergelijkingen mee te nemen waarbij de testset uit een enkel gewicht bestond. Het optimale schema in deze vereenvoudigde setting had een onzekerheid die ongeveer 28% kleiner was dan die van het corresponderende schema in gebruik.

Conclusie

Het weegschema dat door het NMi gebruikt wordt is suboptimaal. Door over te gaan naar een ander weegschema, kan de onzekerheid in de massa's van de nationale standaardgewichten gereduceerd worden met ongeveer 63%. Dit weegschema gebruikt weliswaar meer metingen dan het huidige schema. Indien de hoeveelheid werk gelijk gehouden wordt, dan kan er een reductie van 28% gerealiseerd worden.

Het onderzoek waarover dit artikel rapporteert, is uitgevoerd door Sandjai Bhulai, Thomas Breuer, Eric Cator en Fieke Dekkers. Onze dank gaat uit naar Inge van Andel van het NMi voor de door haar geleverde informatie.

Selectie-effecten in de forensische wetenschap (Misja Nuyens)

Op het slachtoffer van een misdrijf treft men een rode vezel aan. Nadat de politie een verdachte heeft aangehouden, wordt er een rode trui in zijn klerenkast aangetroffen. De trui en de vezel worden vervolgens naar het forensisch laboratorium gebracht. Daar wordt gecontroleerd of ze van hetzelfde materiaal zijn en zo ja, hoe sterk dit bewijsmateriaal is. Een heel zeldzame trui zou namelijk als sterker bewijsmateriaal moeten gelden dan een die bij een grote kledingzaak is gekocht. Maar hoe hangt de sterkte van het bewijsmateriaal af van de omstandigheden waarin de trui is gevonden? Is het bewijsmateriaal bijvoorbeeld sterker als de verdachte geen andere truien in zijn kast had hangen, of maakt dat niet uit?

Dit is een voorbeeld van de volgende vraag die door het Nederlands Forensisch Instituut (NFI) werd gesteld aan de Studiegroep Wiskunde met de Industrie 2005: moet de forensisch expert weten op wat voor manier het bewijsmateriaal is geselecteerd? Deze vraag is ook relevant wanneer iemand verdacht wordt wanneer zijn of haar DNA in een DNA-database zit en overeenkomt met een stukje DNA dat op de plaats van een misdrijf is gevonden. Zulke DNA-databases bevatten natuurlijk niet het DNA van de hele bevolking, maar maakt dat wat uit?

Op dit moment wordt in zaken waar vezels als bewijsmateriaal worden gebruikt geen rekening gehouden met de manier waarop het bewijsmateriaal is verzameld. Dit is verontrustend, omdat het intuïtief duidelijk lijkt dat een match tussen de vezel en een verdachte die weinig truien heeft sterker bewijsmateriaal is dan een match met een 'truienverzamelaar'. Als dit inderdaad waar is, dan zouden statistische correcties gemaakt moeten worden wanneer dit soort bewijsmateriaal in de rechtzaal wordt gepresenteerd. De NFI-expert die de zaak behandelt, is echter in het algemeen geen statisticus. Hij heeft de mogelijkheid om de hulp van een statisticus in te roepen, maar doet dat alleen als hij dat noodzakelijk vindt. De vraag die aan de Studiegroep werd gesteld kan daarom gezien worden als een vraag om hulp in situaties die eenvoudig lijken, maar eigenlijk statistische correcties verlangen.

In dit stukje illustreren we hoe kansrekening gebruikt kan worden om de sterkte van bewijsmateriaal te beoordelen. We maken een heel simpel model voor een vezel-match, en trekken daar een aantal conclusies uit. Uiteraard is elk model ver verwijderd van de realiteit, en moet men erg oppassen met het toepassen van theoretische resultaten op situaties in de echte wereld. Het is helemaal gevaarlijk wanneer statistiek wordt gebruikt om te bewijzen dat een misdaad heeft plaatsgevonden, zie [3]. Desondanks geloven we dat ons simpele model ons iets kan vertellen over hoe de selectie van bewijsmateriaal de sterkte van dit bewijsmateriaal kan beïnvloeden.

De waarde van bewijsmateriaal beoordelen met kansrekening

Tijdens een rechtzaak moet de forensisch expert de sterkte van het bewijsmateriaal beoordelen. Een gedeelte van het bewijsmateriaal kan met de verdachte in verband worden gebracht (dit geven we aan met E). De vraag is of de verdachte dit materiaal zelf heeft achtergelaten (deze gebeurtenis geven we aan met H), of dat de relatie tussen de verdachte en het bewijsmateriaal toevallig is. Deze gebeurtenis is het complement van H: H^c. De forensisch expert wordt verzocht om een zogenaamd aannemelijkheidquotiënt (Engels: *likelihood ratio*) te rapporteren: de kans op het bewijsmateriaal gegeven dat de verdachte schuldig is, gedeeld door de kans op het bewijsmateriaal gegeven dat de verdachte onschuldig is:

$$LR = \frac{P(E|H)}{P(E|H^c)}.$$

Het is verleidelijk om te denken dat een groot aannemelijkheidquotiënt impliceert dat de verdachte schuldig is, maar dit is in het algemeen niet waar. We zijn dan ook niet geïnteresseerd in het aannemelijkheidquotiënt zelf, maar in de kans dat de verdachte schuldig is gegeven dat het bewijsmateriaal overeenkomt, $P(H|E)$. Natuurlijk hebben deze twee met elkaar te maken: met regel van Bayes kunnen we $P(H|E)$ als volgt uitrekenen in termen van het aannemelijkheidquotiënt :

$$P(H|E) = \frac{P(H \cap E)}{P(E)} = \frac{P(E|H)P(H)}{P(E|H)P(H) + P(E|H^c)P(H^c)}.$$

Wanneer we de teller en noemer delen door $P(E|H)P(H)$, krijgen we

$$P(H|E) = \left(1 + \frac{P(H^c)}{P(H)} \frac{P(E|H^c)}{P(E|H)}\right)^{-1} = \left(1 + \frac{P(H^c)}{P(H)} \frac{1}{LR}\right)^{-1}.$$

De kans $P(H|E)$ heet een *posteriori kans*, $P(H)$ en $P(H^c)$ heten *a priori kansen*. De rechter heeft schattingen van deze laatste nodig om een idee te krijgen van $P(H|E)$.

Het truimodel

We gaan nu een heel simpel model bouwen voor de in de inleiding beschreven situatie van een gevonden vezel. Veronderstel

dat we weten dat er een gebeurtenis heeft plaatsgevonden waarbij een onbekend persoon (de *donor*) een vezel van zijn trui heeft achtergelaten op een andere persoon (het *slachtoffer*). Het moment waarop de vezel werd overgedragen heet het transfermoment. Het type vezel noemen we Y en er zijn geen andere vezels op het slachtoffer gevonden. Om de donor te vinden, onderzoeken we de truien in de kast van een willekeurig persoon (de *verdachte*). Daar vinden we een trui gemaakt van hetzelfde type vezel.

We willen nu de kans uitrekenen dat de verdachte daadwerkelijk de donor van de vezel op het slachtoffer is, gegeven het bewijsmateriaal. De vraag is nu hoeveel we moeten weten van het gevonden bewijsmateriaal: volstaat het om te weten dat een van de truien in de kast van de verdachte overeenkwam met de vezel op het slachtoffer? Of moeten we bijvoorbeeld ook weten hoeveel truien de verdachte in zijn kast had?

Om de kans uit te rekenen dat de verdachte de donor was, nemen we aan dat de relatieve frequentie waarmee de hele bevolking truien draagt die bestaan uit vezels van type Y geschat kan worden op g_Y; de kans dat de verdachte een trui aanhad van dit type noemen we f_Y. Tenslotte nemen we aan dat niemand truien heeft verborgen of weggegooid en dat alle truien uit één vezeltype bestaan.

We schrijven E_1 voor de gebeurtenis dat de vezel op het slachtoffer van type Y is, en E_2 voor de gebeurtenis dat we een vezel van type Y in de kast van de verdachte vinden. De gebeurtenis dat de verdachte de donor van de vezel op het slachtoffer is noemen we H. We nemen ook aan dat E_2 en H onafhankelijk zijn. Zoals uitgelegd in de vorige paragraaf, kunnen we nu $P(H|E_1 \cap E_2)$ uitrekenen:

$$\begin{aligned} P(H|E_1 \cap E_2) &= \left(1 + \frac{P(H^c)}{P(H)} \frac{P(E_1 \cap E_2|H^c)}{P(E_1 \cap E_2|H)}\right)^{-1} \\ &= \left(1 + \frac{P(H^c)}{P(H)} \frac{P(E_2|H^c)P(E_1|H^c \cap E_2)}{P(E_2|H)P(E_1|H \cap E_2)}\right)^{-1} \\ &= \left(1 + \frac{P(H^c)}{P(H)} \frac{P(E_2)g_Y}{P(E_2)f_Y}\right)^{-1} \\ &= \left(1 + \frac{P(H^c)}{P(H)} \frac{g_Y}{f_Y}\right)^{-1}. \end{aligned}$$

We zien dat als g_Y klein is (type Y vezels zijn zeldzaam), dan is $P(H|E_1 \cap E_2)$ relatief groot. Verder, als f_Y klein is (de verdachte draagt zijn trui met type Y vezels niet vaak), dan is $P(H|E_1 \cap E_2)$ klein. In het speciale geval dat de verdachte k truien heeft en die even vaak draagt, krijgen we $f_Y = 1/k$. We zien dan dat hoe meer truien de verdachte heeft, hoe kleiner de kans is dat hij de donor is. Dat lijkt redelijk: een persoon die duizend truien heeft, loopt een grote kans dat een van zijn truien overeenkomt met de vezel op het slachtoffer, maar de sterkte van die match is natuurlijk niet erg groot.

Conclusies

In ons basale truimodel hebben we laten zien dat veel dingen gerapporteerd zouden moeten worden om het bewijsmateriaal goed te kunnen interpreteren. Niet alleen dat er een trui in de kast van de verdachte is gevonden die overeenkomt met een vezel op de plaats van de misdaad, maar bijvoorbeeld ook hoeveel truien hij in zijn kast had, en eigenlijk ook hoe vaak de verdachte die bepaalde trui draagt.

Hoewel ons model ver van realistisch was, is het duidelijk dat informatie over de manier waarop bewijsmateriaal is vergaard zo compleet mogelijk moet zijn. Alle informatie moet in een groot model gestopt worden, net als al het bewijsmateriaal dat gunstig is voor de verdachte. We realiseren ons dat we zelfs in dat geval slechts een model hebben, en dat elk model tot veel discussies aanleiding zal geven.

We concluderen dat selectie-effecten in de forensische wetenschap een belangrijke rol spelen en dat de moeite gedaan zou moeten worden om de statistische interpretatie van bewijsmateriaal in de rechtzaal te verbeteren.

Wiskundigen die aan dit onderzoek werkten: Geert Jan Franx, Yves van Gennip, Peter Hochs, Misja Nuyens, Luigi Palla, Corrie Quant en Pieter Trapman.

Planning van drinkwater voor vliegtuigen (Marco Bijvank, Maarten Soomer)

Tijdens een vlucht wordt er drinkwater gebruikt voor consumpties en voor het doorspoelen van de toiletten. Het management van de Nederlandse luchtvaartmaatschappij KLM wil dat er genoeg drinkwater aan boord is op hun vluchten. Daarentegen willen ze ook niet te veel water meenemen omdat dit, vanwege het gewicht, hogere kerosine kosten met zich meebrengt.

Voor een vlucht wordt de watertank van het vliegtuig gevuld. De apparatuur waar dit mee gebeurt, heeft alleen de mogelijkheid de tank te vullen tot een veelvoud van achtsten van de tank. De hoeveelheid water wordt van te voren bepaald aan de hand van de bestemming en het aantal passagiers van de vlucht. Overtollig water wordt na de vlucht uit de watertank geloosd. Het doel van dit onderzoek is om een optimale hoeveelheid water te bepalen waarmee het vliegtuig moet vertrekken. Er moet aan het eind van de vlucht zo weinig mogelijk water over zijn en tegelijkertijd moet de kans op een watertekort tijdens de vlucht erg klein zijn. Om het laatste vast te stellen moet er een geschikte definitie voor het serviceniveau worden gevonden.

Om een optimale hoeveelheid drinkwater te bepalen, maken we gebruik van historische data. Deze gegevens hebben helaas een aantal nadelen. Voordat het vliegtuig opstijgt, leest de purser het waterniveau van een display in de cabine. In de meeste vliegtuigen heeft deze display maar acht markeringen, waardoor het moeilijk is de exacte hoeveelheid water af te lezen. Daarom wordt deze waarde door de purser afgerond naar de dichtstbijzijnde markering. Omdat de tank ook gevuld wordt in achtsten beschouwen we deze waarde als exact. Na de landing wordt het watervolume weer genoteerd. Ook deze waarde is weer afgerond (en kan niet als exact worden beschouwd). Verder bevatten de datarecords gegevens over het type van het vliegtuig, het aantal passagiers, het vertrekpunt en de bestemming, de vertrektijd en vliegduur van de vlucht. De vraag is nu wat voor invloed de afgeronde data hebben en hoe dit verwerkt moet worden in het model.

Data-analyse

Op dit moment gebruikt de KLM de historische data van vluchten met hetzelfde beginpunt en dezelfde bestemming, om het waterverbruik op een bepaalde vlucht te schatten. Het is interessant om te onderzoeken of vluchten met een ander beginpunt of bestemming, toch een vergelijkbaar patroon in waterverbruik heb-

ben. Als dit het geval is, kunnen de data van vluchten geclusterd worden en kan er dus een betrouwbaardere schatting worden gegeven.

Voor de meeste bestemmingen is er een duidelijke correlatie tussen de bestemming en het waterverbruik per passagier. Dit betekent dat deze vluchten niet hetzelfde patroon in waterverbruik hebben. Vluchten met dezelfde vliegduur kunnen ook niet samengenomen worden. Alleen tussen een aantal vluchten naar dezelfde regio, zoals vluchten van Amsterdam naar Aruba en van Amsterdam naar Bonaire, kon geen onderscheid worden gemaakt. De data van deze vluchten kunnen dus gecombineerd worden om de waterhoeveelheid te bepalen.

Service Level

De huidige definitie van service level, die de KLM hanteert, is het percentage vluchten (met hetzelfde beginpunt en dezelfde bestemming) dat genoeg drinkwater aan boord heeft. Een service level van 95% betekent dus dat er op ten hoogste 5% van die vluchten een watertekort is. Dit betekent echter niet dat de kans voor een passagier om geconfronteerd te worden met een watertekort ook maar 5% is. Als een watertekort een structureel probleem is op vluchten met veel passagiers, dan heeft een passagier in een druk vliegtuig een grotere kans om met een watertekort geconfronteerd te worden. Dit is de reden dat wij een ander service level hebben gedefinieerd. Deze wordt hieronder besproken.

De totale waterconsumptie tijdens een vlucht met n passagiers, zal aangeduid worden met S_n. Op basis van de historische data hebben we alleen gegevens over afgeronde waardes van S_n. Om het bovenvermelde probleem te omzeilen, definiëren we nu het *service level* als de kans dat er genoeg water aanwezig is, gegeven dat er n passagiers aan boord zijn. Dit service level moet dan groter of gelijk zijn aan een vooraf (door het management van de KLM) gedefinieerde waarde α. Dit betekent dat aan de volgende voorwaarde voldaan moet worden:

$$\forall n \quad \mathbf{P}\left(S_n \leq \frac{j}{8}T\right) \geq \alpha, \qquad j \in \{0,1,\ldots,8\},$$

waarbij $j/8$, het percentage van de tank dat gevuld wordt is en T staat voor de capaciteit van de tank (in liters). Deze definitie van service level wordt ook wel *Quality of Service* (QoS) genoemd, omdat het geldt voor elk passagiersaantal. Het is gedefinieerd vanuit het perspectief van de consument, die heeft nu onafhankelijk van het aantal medepassagiers hetzelfde risico om met een watertekort te worden geconfronteerd. Deze definitie was een openbaring voor de KLM en bracht nieuw inzicht in het probleem.

Uiteraard is de volgende stap om een zo klein mogelijke hoeveelheid water te bepalen dat aan boord aanwezig moet zijn, zodanig dat aan het service level voldaan wordt (dat wil zeggen de kleinste waarde voor j in de vergelijking). Om dit te bepalen hebben we een kansverdeling voor de totale waterconsumptie S_n nodig. We hebben drie verschillende methoden ontwikkeld om deze kansverdeling te bepalen. De methoden gebruiken de historische data en moeten daarom rekening houden met de afrondingen. De methoden worden in de volgende drie paragrafen besproken.

Empirische Benadering

De kans dat j-achtste van de watertank wordt gebruikt als er n passagiers aan boord zijn, wordt bij deze aanpak afgeleid uit de frequentie dat dit voorkomt in de data. Deze kansen kunnen worden gebruikt als een kansdichtheidsfunctie voor het watergebruik gedurende de vlucht. Om echter tot betrouwbare frequenties te komen, moeten er genoeg vluchten in de data zijn met precies n passagiers. Helaas is dat niet het geval. Dit kan worden opgelost door ook waarnemingen mee te nemen waarbij het aantal passagiers in de buurt van n ligt.

We kunnen een kansverdeling schatten door een continue functie door deze negen punten te construeren en dan de oppervlakte onder de functie te normaliseren. Dit betekent dat we een interpolatiemethode nodig hebben. De meest gebruikte methode is een benadering door polynomen. In dit geval lijkt *cubic spline interpolatie* (Press et al. [1]) beter geschikt, omdat dit stabieler is dan het gebruik van polynomen. Het doel van cubic spline is een interpolatie formule te krijgen die continu is in de tweede afgeleide. Als dit gebeurd is, moet de verkregen functie genormaliseerd worden. Hierna kan het minimum waterniveau, waarvoor het service level op zijn minst α is, worden bepaald. Voor praktische toepassing zorgen we ervoor dat dit waterniveau een monotoon stijgende functie is van het aantal passagiers. Omdat er weinig data beschikbaar zijn van vluchten met weinig passagiers, nemen we ook aan dat het gemiddelde waterverbruik per passagier per uur maximaal één liter is.

Normale Verdeling

De vorige aanpak gebruikte maar een gedeelte van de historische data om het waterverbruik te schatten (namelijk alleen de vluchten met ongeveer gelijke passagiersaantallen). In deze aanpak zal de totale waterconsumptie S_n voor een vlucht met n passagiers weergegeven worden door

$$S_n = \sum_{k=1}^{n} Y_k,$$

waarin Y_k staat voor de waterconsumptie van de k-de passagier (in liters). De aanname die nu gemaakt wordt is dat het waterverbruik per passagier onafhankelijk en hetzelfde verdeeld is. Aangezien het aantal passagiers groot is, kan de centrale limietstelling (Ross [2]) toegepast worden. Als we deze stelling generaliseren door een constante toe te voegen aan het gemiddelde en de variantie van het waterverbuik, dan krijgen we

$$S_n \overset{d}{\sim} N(\mu_0 + n\mu, \sigma_0^2 + n\sigma^2),$$

met μ en σ^2 het gemiddelde en de variantie van het waterverbruik per passagier en μ_0 en σ_0^2 het gemiddelde en variantie van de minimale hoeveelheid water die altijd gebruikt wordt. De waarden van deze vier parameters moeten geschat worden. Deze schatting wordt uitgevoerd met de maximum-likelihoodmethode (Ross [2]).

Binomiale Verdeling

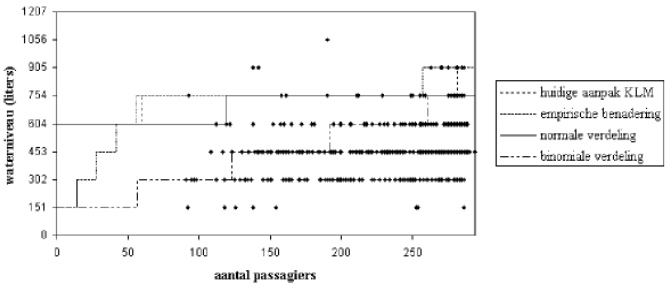
In de vorige aanpak werd geen aanname gemaakt over de exacte verdeling van het waterverbruik per passagier. Bij deze aanpak veronderstellen we dat een passagier een maximum hoeveelheid water (M) verbruikt met kans p of een minimum hoeveelheid wa-

waterniveau liters	Aantal passagiers (drempelwaardes voor het waterniveau)			
	huidige aanpak KLM	empirische benadering	normale verdeling	binomiale verdeling
151	-	0 - 13	-	0 - 56
302	-	14 - 27	-	57 - 122
453	-	28 - 41	-	123 - 191
604	0 - 59	42 - 55	0 - 118	192 - 260
754	60 - 281	56 - 256	119 - 294	261 - 294
905	282 - 294	257 - 294	-	-

Tabel 3 Aantal passagiers en aanbevolen waterhoeveelheid met een 95% *Quality of Service*, bepaald met de verschillende methodes, voor de vlucht AMS-BKK

ter (m) met kans $1 - p$. Dit betekent dat zowel het aantal passagiers dat de maximum hoeveelheid water verbruikt, als het aantal passagiers dat de minimale hoeveelheid verbruikt, binomiaal verdeeld is.

De waarden voor m, M en p moeten geschat worden om het gewenste waterniveau te bepalen. Dit kan op een intuïtieve manier gedaan worden, maar ook door meer geavanceerde methodes, zoals de maximum-likelihoodmethode.



Figuur 3 Aanbevolen waterhoeveelheid met een 95% *Quality of Service* met de verschillende methoden en de afgeronde data voor de vlucht AMS-BKK

Resultaten en Conclusies

In dit onderzoek hebben we een model ontwikkeld om de minimale hoeveelheid water te bepalen dat benodigd is tijdens een vlucht. Er moet echter wel aan een vooraf gesteld serviceniveau voldaan worden. Hierin is het serviceniveau gedefinieerd als de kans dat er genoeg water aanwezig is, gegeven het aantal passagiers aan boord. Het vaststellen van een kansverdeling, voor het totale watergebruik tijdens een vlucht, hebben we op drie verschillende manieren aangepakt. De grootste beperking waarmee rekening gehouden moest worden, was dat de data over het waterverbruik gegeven was in veelvouden van achtsten van de watertank. Het gebruik van de empirische benadering is lastig omdat er in de praktijk voor veel vluchten niet genoeg data beschikbaar is voor een gegeven passagiersaantal. Daarom wordt ook de data met ongeveer gelijke passagiersaantallen gebruikt. Dit heeft tot gevolg dat de verdeling voor het totale waterverbruik dikkere staarten krijgt, waardoor het verkregen waterniveau te hoog ligt. Deze methode geeft dus eigenlijk een bovengrens voor de benodigde waterhoeveelheid.

De methoden met de normale en binomiale verdeling gaan er vanuit dat het gemiddelde en de variantie van het totale waterverbruik lineair toenemen met het aantal passagiers. Het eerste klopt inderdaad met de data, het tweede helaas niet. Het verschil tussen deze twee methodes ligt in de geschatte parameters voor het gemiddelde en de variantie. Het binomiale model heeft als eigenschap dat het minimum en maximum waterverbruik begrensd is.

Voor de vlucht van Amsterdam (AMS) naar Bangkok (BKK) zullen we de drie methodes numeriek toelichten. De resultaten zijn samengevat in tabel 3 en figuur 3.

Onze voorkeur gaat uit naar de methode met de normale verdeling, waarbij de empirische benaderingsmethode als bovengrens gebruikt kan worden. De binomiale methode kan gebruikt worden om de invloed van het wijzigen van parameters te evalueren. De resultaten van de huidige methode van de KLM, lijken op die van de methode met de normale verdeling. Deze laatste heeft echter een betere onderbouwing en gebruikt een betere definitie van het service level, daarom heeft deze methode ook de voorkeur van de KLM.

Aan het drinkwaterprobleem werkten de wiskundigen Marco Bijvank, Menno Dobber, Quentin Botton, Eléonore de le Court, Jean-Christophe Van den Schrieck, Moïra de Viron, Maarten Soomer, Myriam Cisneros-Molina, Klaus Schmitz, Remco van der Hofstad, Ellen Jochemsz, Tim Mussche, Martin Summer, Maroescha Hoekstra, Jeroen Mulder en Mark Paelinck

Referenties

1 W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, *Numerical Recipes in C*, Cambridge University Press, New York, 1988.

2 S.M. Ross, *Introduction to Probability Models*, Eighth Edition, Academic Press, San Diego, 2003.

3 M. van Lambalgen and R. Meester, *On the (ab)use of statistics in the legal case against the nurse Lucia de B*, preprint, available from www.few.vu.nl/~rmeester/pre.html (2005).