

STATOR

thema SPORT, STATISTIEK EN OR

Statistiek in de sport: onderzoekstrends, nu en in de toekomst

De kans om een tenniswedstrijd te winnen: Federer-Nadal in de finale van Wimbledon 2007

Oost, west, thuis best?

Het plannen van de Belgische voetbalcompetitie

Sven Kramer passeert Eric Heiden; de beste schaatser en schaatsster aller tijden

Sportende statistici

Tienkamp: een kwestie van balans

Sport en toeval

STaTOR

Jaargang 9, nummer 2, juli 2008

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

Redactie

Goos Kant (hoofredacteur), Ana Isabel Barros, Mirjam Moerbeek, Gerrit Stemerding (eindredacteur), Fred Steutel, Hilde Tobi, Marnix Zoutenbier, Ruud H. Koning (gastredacteur), Gerard Sierksma (gastredacteur).

Kopij en reacties richten aan

Prof. dr. G. Kant (hoofredacteur), Faculteit der Economische Wetenschappen van de Universiteit van Tilburg, Postbus 90153, 5000 LE Tilburg, telefoon 013 - 4668234, mobiel 06-11045089, <G.Kant@uvt.nl>.

Bestuur van de VVS

Voorzitter: prof. dr. R. Gill <voorzitter@vvs-or.nl>

Secretaris: dr. C.G.H. Diks <c.g.h.diks@uva.nl>

Penningmeester: prof. dr. ir. C.A.G.M. van Montfort <kvmontfort@feweb.vu.nl>

Statistische dag: prof. dr. A.W. van der Vaart <aad@cs.vu.nl>

Namens de Bedrijfssectie (BDS):

prof. dr. R.J.M.M. Does <R.J.M.M.Does@uva.nl>

Namens de Biometrische Sectie (BMS):

prof. dr. A.H. Zwinderman <a.h.zwinderman@amc.uva.nl>

Namens de Economische Sectie (ECS):

dr. P.H.F.M. van Casteren <casteren@fee.uva.nl>

Namens het Ned. Genootschap voor Besliskunde (NGB):

prof. dr. J.J. van de Klundert <j.vandeklundert@math.uni-maastricht.nl>

Namens de Sectie Mathematische Statistiek (SMS):

dr. P.J.C. Spreij <spreij@science.uva.nl>

Namens de Sociaal Wetenschappelijke Sectie (SWS):

prof. dr. J.K. Vermunt <j.k.vermunt@uvt.nl>

Leden- en abonnementenadministratie van de VVS

VVS, Postbus 2095, 2990 DB Barendrecht, telefoon 0180 - 623796, fax 0180 - 623670, e-mail <admin@vvs-or.nl>.

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

VVS-website

<http://www.vvs-or.nl>

Advertentieacquisitie

Marieke Klein, p/a Vrije Universiteit, faculteit economie, De Boelelaan 1105, 1085 HV Amsterdam, e-mail: adverteren.stator@vvs-or.nl. STaTOR verschijnt in maart, juni, september en december.

Ontwerp en opmaak

Pharos / M. van Hootegem, Nijmegen

Uitgever

© Vereniging voor Statistiek en Operationele Research
ISSN 1567-3383

Inhoud

3 Sport in statistiek en OR

4 Statistiek in de sport: onderzoekstrends, nu en in de toekomst

Jerome P. Reiter

8 De kans om een tenniswedstrijd te winnen: Federer-Nadal in de finale van Wimbledon 2007

Franc J.G.M. Klaassen en Jan R. Magnus

12 Oost, west, thuis best?

Ruud H. Koning

16 Het plannen van de Belgische voetbalcompetitie

Dries Goossens en Frits C.R. Spijksma

23 Sven Kramer passeert Eric Heiden; de beste schaatser en schaatsster aller tijden

Gerard Sierksma en Bertus Talsma

27 Sportende statistici - column

Fred Steutel

29 Tienkamp: een kwestie van balans

Gerard Sierksma en Yori Zwols

35 Sport en toeval - column

Onno Boxma

37 Fred Steutel benoemd tot erelid van de VVS

37 Haastige spoed ...

38 In memoriam Theo Runnenburg (1932-2008)

38 Agenda



SPORT, STATISTIEK EN OR

De zomer van 2008 zit boordevol grote sportevenementen. Naast Wimbledon en de Tour de France wordt de zomer dit jaar opgefleurd door de Europese Voetbalkampioenschappen en de Olympische Spelen. Van alle minder belangrijke zaken is sport even de belangrijkste. En de sportzomer is reeds van start. Zeven miljoen Nederlanders keken op 9 juni naar de voetbalwedstrijd Nederland-Italië en de herenfinale van Roland Garros, een dag eerder, was goed voor zeshonderdduizend kijkers. Dat laatste is maar liefst 4% van de Nederlandse bevolking ouder dan zes jaar. Sport mag zich ook verheugen in een toenevende wetenschappelijke belangstelling en traditioneel spelen statistiek en operations research daarin een belangrijke rol. Hoog tijd derhalve voor een themanummer van *STATOR* op dit gebied!

Jerome Reiter, voorzitter van de sectie 'Statistics in Sports' van de American Statistical Association, trapt af met een overzicht van trends in statistisch sportonderzoek. Franc Klaassen en Jan Magnus analyseren de Wimbledonfinale van vorig jaar en Ruud Koning probeert het thuisvoordeel te meten in voetbalcompetities. Het vakgebied OR is vertegenwoordigd met een artikel van Frits Spieksma over wedstrijdroostering voor de Belgische voetbalcompetitie, een schaatsartikel van de hand van Bertus Talsma en Gerard Sierksma over de mogelijkheden van Sven Kramer om de beste schaatser

aller tijden te worden en tenslotte een artikel over optimale tienkamptraining van Yori Zwols en Gerard Sierksma.

Dit themanummer wordt gecompleteerd door twee bijdragen van James J. Cochran, voorzitter van de sectie OR in *SpORts'* van *INFORMS*: 'Operations Research and Sports: A Brief Overview' en 'The Optimal Value and Potential Alternatives of Bill James' Pythagorean Method of Baseball'. Beide artikelen zijn te vinden op de website van de VVS: www.vvs-or.nl.

Zoals u zult zien, operations research en statistiek bieden interessante inzichten in diverse sporten. Deze inzichten kunnen ook leiden tot betere prestaties en tot competitieve voordelen. Daarnaast is sport een prachtige metafoor om onze vakgebieden verder onder de aandacht van het brede publiek te brengen en studenten te motiveren om econometrie te kiezen.

Wij wensen u veel leesplezier!

Ruud H. Koning en Gerard Sierksma,
gastredacteurs

STATISTIEK IN DE SPORT

onderzoekstrends, nu en in de toekomst

Wij sportenthousiastelingen weten allemaal dat statistische gegevens het plezier in sportevenementen aanzienlijk vergroten. We kijken naar de prestaties van teams en spelers in het verleden en maken schattingen van de kans op resultaten die zijn behaald of resultaten waar we op hopen. We beroepen ons op statistieken in onze hartstochtelijke discussies over wie de beste speler is. We houden de loopbaan van onze favoriete spelers nauwkeurig bij om te zien of ze zich in de loop van de tijd positief ontwikkelen of juist niet. Het belang van sportstatistieken blijkt overduidelijk uit de eindeloze getallenreeksen die worden geciteerd door degenen die belang hebben bij sport als amusement, zoals kranten, websites, televisiestations en de sportclubs zelf.

JEROME P. REITER

De meest gebruikelijke statistische activiteit betreft het verzamelen en publiceren van samenvattende gegevens, zoals totalen, gemiddelden en percentages, maar de laatste tijd zijn onderzoekers van sportstatistiek zich gaan bezighouden met meer geavanceerde statistische modellen voor het analyseren van sportvraagstukken. Dit wordt deels veroorzaakt door de beschikbaarheid van steeds meer gedetailleerde sportgegevens. Ter illustratie van deze trend beschrijft dit artikel een aantal recente toepassingen op sportgegevens uit de literatuur. De voorbeelden zijn ingedeeld aan de hand van een grove classificatie van het soort vragen waarmee sportonderzoekers zich bezig houden. Voorbeelden hiervan zijn het karakteriseren van spelers en teams, de analyse van sportstrategieën en modeleigenschappen van sportwedstrijden. Het artikel besluit met een paar opmerkingen over de toekomst van het onderzoek over sportstatistiek.

Karakterisering van spelers en teams

Een groot deel van de sportstatistiek is gewijd aan het karakteriseren van spelers en teams. Voorbeelden hiervan zijn het rangschikken van spelers en teams, de bepaling van niveaus voor spelers of teams die succes voorspellen, en het beschrijven van loopbaantrends. Deze zaken worden door niet-onderzoekers vaak geanalyseerd met behulp van eenvoudige technieken zoals grafische voorstellingen, maar met deze eenvoudige technieken kunnen de ingewikkelde sportgegevens doorgaans niet goed worden gerepresenteerd. Een op modellen gebaseerde aanpak kan dat wel, zoals we nu aan de hand van twee voorbeelden zullen laten zien.

Het is algemeen geaccepteerd dat de vaardigheid van de meeste atleten afneemt aan het eind van hun loopbaan. Anderzijds is het redelijk om aan te nemen dat de vaardigheid van atleten toe-

neemt met hun ervaring. Daarom is te verwachten dat het verband tussen leeftijd en prestatie van een speler niet lineair verloopt en zal afhangen van de kenmerkende eigenschappen van zijn sport en zijn plaats daarin. Om dit verband te onderzoeken gebruikt Fair (2008) een steekproef van Major League Baseball spelers om grafieken te schatten die het verband aangeven tussen leeftijd en prestaties bij slagmannen en werpers. Hij schat de parameters in zijn model met behulp van niet-lineaire regressie. Hij vindt dat werpers hun top bereiken bij een leeftijd van 26 jaar en slagmannen bij 28 jaar. Hij vindt verder dat na hun top de prestaties van werpers sneller afnemen dan die van slagmannen. Het model van Fair kan ook gebruikt worden om een voor leeftijd gecorrigeerde rangschikking van spelers te bepalen.

Voor een andere toepassing van een geavanceerd model voor statistische problemen in de sport kijken we naar een analyse van tienkampprestaties door Woolf et al (2007). Zij gebruiken clusteranalyse om de tienkampers in vijf groepen te verdelen. Zij interpreteren de clusters als: een hardlopgroep, een werpgroep en drie afzonderlijke groepen voor succes in het polsstokhoogspringen, het hoogspringen en de 1500 meterloop. Verder onderzoeken Woolf et al de rangorde van de tienkampers in elke groep en vinden dat atleten die het goed doen op de sprintnummers mogelijk een voordeel hebben in de tienkamp.

Deze twee voorbeelden illustreren het nut van geavanceerde statistische modellen voor het karakteriseren van spelers en teams. Overig onderzoek op dit terrein maakt gebruik van hiërarchieke Bayesiaanse modellen en computerleertechnieken zoals 'support vector machines'.

Analyse van strategieën

Spelers, coaches en teams moeten voortdurend beslissingen nemen, zowel binnen de wedstrijd,

zoals het al of niet begaan van een opzettelijke overtreding, als erbuiten, bijvoorbeeld de beslissing welke spelers gewisseld of verkocht moeten worden. Wij geven twee voorbeelden van modelleringstechnieken die door sportstatistici gebruikt worden om dit soort problemen te analyseren. Barry Bonds heeft ongetwijfeld de beste aanvalscijfers in de geschiedenis van het Major League honkbal: hij heeft van alle spelers de meeste homeruns geslagen. In de seizoenen 2001 tot en met 2004 was zijn slagkracht zo gevreesd dat tegenstanders hem vaak liever een 'vrije loop' gaven dan op hem te werpen. Was dat een verstandige strategie? Om deze vraag te beantwoorden vergelijkt Reiter (2004) het aantal runs dat Bonds team heeft gescoord bij vrije lopen voor Bond met het aantal runs bij werpen op Bond. Met 'matching' technieken corrigeert hij voor variabelen die de vergelijking zouden kunnen verstoren, zoals de kwaliteit van de 'vijandelijke' werper en de situatie in de wedstrijd. Hij constateert dat de tegenstander in veel gevallen meer runs tegen krijgt door Bond een vrije loop te geven dan door op hem te werpen, in tegenspraak met de gebruikelijke keuze van honkbalmanagers. Dit voorbeeld illustreert hoe 'causale' methoden, ontwikkeld voor toepassingen in de medische- en sociale wetenschappen, gebruikt kunnen worden voor de beoordeling van strategieën in de sport.

Bij ijshockey wordt vaak bij achterstand van één doelpunt aan het eind van de wedstrijd - in de hoop op de gelijkmaker - de keeper vervangen door een aanvalser. De meeste teams wachten met het vervangen van de doelman tot de laatste minuut. Zouden ze dat eerder moeten doen? Om deze zaak te onderzoeken beschouwt Zannan (2001) een Markov-keten met zeven toestanden, vastgelegd door de positie van de puck, die varieert van "doelpunt gemaakt", via "in het neutrale vak" tot "doelpunt tegen gekregen". Uit gegevens van gespeelde wedstrijden schat hij de overgangskansen tussen de toestanden en simuleert hiermee wedstrijden

waarbij de doelman op verschillende momenten in de derde periode wordt gewisseld. Hij vindt als optimale tijd voor de wissel ongeveer zes minuten voor het eind van de wedstrijd. Zamans analyse illustreert een steeds meer gebruikte techniek voor de analyse van sporttactiek: beschrijf de wedstrijd met een Markov-model en simuleer wedstrijden onder verschillende strategieën. Deze techniek wordt ook gebruikt om voetbalwedstrijden te evalueren.

Het modelleren van sportwedstrijden

Veel onderzoekers zijn geavanceerde modellen gaan gebruiken om heel specifieke eigenschappen van sportwedstrijden en reeksen van wedstrijden te beschrijven. Zo gebruiken onderzoekers regressiemodellen om het effect van 'thuisvoordeel' te kwantificeren. Zij ontwikkelen Markov-modellen voor wedstrijden om de kansen te schatten dat de wedstrijd van een gegeven huidige toestand overgaat naar een belangrijke andere toestand. Ook gebruiken zij 'random effects' modellen om de wispelturigheid te kwantificeren van slagmensen bij honkbal en schutters bij basketbal. In deze sectie geven we twee voorbeelden van statistische modellen om kenmerken van sportwedstrijden te beschrijven.

Basketballspelers kunnen verschillende vaardigheden hebben, zoals 'passen' van de bal naar een medespeler, veroveren van de 'rebound' en het scoren van punten. Het ligt voor de hand om te veronderstellen dat voor spelers op bepaalde posities ('center', 'point guard', 'shooting guard', etc.) sommige vaardigheden belangrijker zijn dan andere. Kunnen deze verschillen gekwantificeerd worden? Om dit te onderzoeken bekijken Page et al (2007) de gedetailleerde resultaten van wedstrijden in de National Basketball Association. Zij gebruiken een hiërarchiek Bayes-model om verschillen in de einduitslag van wedstrijden te

voorspellen uit de verschillen in prestaties van spelers van de tegenpartij (bijv. de verschillen in gescoorde punten, 'rebounds' en 'steals' tussen de point guards van de beide teams). Zij nemen random effects mee voor individuele spelers en gebruiken verdelingen van de random effects met parameters voor de positieniveaus. Met behulp van Monte Carlo Markov technieken worden posterior verdelingen geschat en worden allebei conclusies getrokken, waarvan sommige de typische basketbalenthousiast niet zullen verbazen (meer aanvallende rebounds en minder balverlies gaan samen met positieve verschillen in de score), maar sommige andere wél (defensieve rebounds, zowel door de point guard als door de shooting guard, zijn van belang).

Statistische modellen kunnen niet alleen gebruikt worden om ingewikkelde vragen over sport te beantwoorden, maar ook om ingewikkelde gebeurtenissen binnen een wedstrijd te beschrijven. Brillinger (2007) bijvoorbeeld gebruikt een statistisch model voor één enkele spelsituatie: het doelpunt over 25 passes van Argentinië tegen Servië-Montenegro in the Wereldbeker van 2006. Brillinger gebruikt potentiaalfuncties, een begrip uit de statistische fysica, om de baan van het 25-passes doelpunt te beschrijven. Deze aanpak zou behalve voor de gedetailleerde beschrijving van spelsituaties ook gebruikt kunnen worden om hele voetbalwedstrijden te beschrijven, wat dan weer gebruikt zou kunnen worden voor spelerstraining met computers of voor het evalueren van strategieën (Brillinger (2007)). Andere onderzoekers hebben voetbaldata geanalyseerd om de kansverdelingen te beschrijven van het aantal doelpunten en de wachttijd tot een doelpunt.

De toekomst van statistisch sportonderzoek

De huidige trend in de sportstatistiek is blijkbaar de toepassing van geavanceerde statistische

methoden om ingewikkelde vragen te beantwoorden. Hoe zal het verder gaan? Zoals in veel andere wetenschapsgebieden beschikken sportwetenschappers over zeer grote hoeveelheden heel gedetailleerde gegevens. Je kunt je voorstellen dat deze trend zal doorzetten, mogelijk tot de situatie dat onderzoekers de acties van iedere speler op elk moment in de wedstrijd kunnen 'downloaden'. Het uitbuiten van dergelijke grote gegevensbestanden is een uitdaging. Onderzoekers hebben nog geen effectieve meetmethoden om bijvoorbeeld de waarde van het wegstompen bij basketbal te bepalen, van het 'stelen' van een pass bij Amerikaans voetbal of van het 'checken' van een tegenstander bij ijshockey. Het ontwikkelen van meetmethoden voor het van-seconde-tot-seconde volgen van gebeurtenissen zal een belangrijke uitdaging worden. Het analyseren van dergelijke gegevens zou ook een uitdaging voor computers kunnen worden. Vorderingen op het gebied van machine-leertechnieken zoals gebruikt in de bioinformatica en bij 'data mining' zouden een belangrijke rol kunnen spelen bij de analyse van deze enorme gegevensbestanden in de sport.

Onderzoekers beginnen deze uitdagingen op te pakken. Als voorbeeld kijken we naar het werk van Carruth en Jensen (2007), die het effect kwantificeren van de werpvaardigheid van de 'verrevelders' op het vermogen van het team om te winnen. Zij baseren hun analyse op de resultaten per wedstrijd voor alle wedstrijden in de Major Baseball League van 2002 tot 2005. Zij bepalen voor elke verrevelder het aantal gelegenheden om de bal te werpen in elk van de vele 'cellen', gedefinieerd door spelsituatie en de plaats in het veld. Het resultaat is een zeer groot gegevensbestand om te analyseren. Zij introduceren een meetmethode voor het evalueren van verrevelders gebaseerd op het verwachte aantal runs dat de tegenstander zal scoren bij een gegeven spelsituatie. Het verwachte aantal runs wordt bepaald

met een Markov-model voor honkbal. Met behulp van de verwachte aantallen runs per cel wordt een hiërarchiek Bayes-model met random effects aangepast voor verrevelders. Het accent ligt op de verdeling van deze random effects, die aangeven of de speler significant boven, vlakbij of significant onder gemiddeld presteert in het voorkomen van runs gebaseerd op zijn werpkwaliteit. Het blijkt dat heel weinig verrevelders significant verschillend van het gemiddelde presteren.

De onderzoekers in dit overzicht verzamelen en analyseren grote en gedetailleerde gegevensbestanden, ontwikkelen meetmethoden voor moeilijk kwantificeerbare uitkomsten en passen complexe statistische modellen aan om alle bronnen van onzekerheid mee te nemen. Ik denk dat de toekomst van het onderzoek in de sportstatistiek is gelegen in analyses die dit soort uitdagingen aangaan.

LITERATUUR

- Brillinger, David R. (2007). A potential function approach to the flow of play in soccer, *Journal of Quantitative Analysis in Sports*, 3:1, Article 3.
- Carruth, Matthew & Jensen, Shane (2007). Evaluating throwing ability in baseball, *Journal of Quantitative Analysis in Sports*, 3:3, Article 2.
- Fair, Ray C. (2008). Estimated age effects in baseball, *Journal of Quantitative Analysis in Sports*, 4:1, Article 1.
- Page, Garritt L., Fellingham, Gilbert W., & Reese, C. Shane (2007). Using box-scores to determine a position's contribution to winning basketball games, *Journal of Quantitative Analysis in Sports*, 3:4, Article 1.
- Reiter, Jerome P. (2004). Should teams walk or pitch to Barry Bonds? *Baseball Research Journal*, 32, 63 - 69.
- Woolf, Anne, Ansley, Les & Bidgood, Penelope (2007). Grouping of Decathlon Disciplines, *Journal of Quantitative Analysis in Sports*, 3:4, Article 5.
- Zaman, Zia (2001). Coach Markov Pulls Goalie Poisson, *Chance*, 14:2, 31-35.

JEROME REITER is hoogleraar in het Department of Statistical Science at Duke University. Hij is ook hoofd van de Section on Statistics in Sports of the American Statistical Association in 2008.
E-mail: jerry@stat.duke.edu



Roger Federer

DE KANS OM EEN TENNISWEDSTRIJD TE WINNEN Federer-Nadal in de finale van Wimbledon 2007

De simpelste methode om de winnaar te voorspellen, is te kijken naar eerdere prestaties. Dit artikel beschrijft een methode om de winnaar van een tenniswedstrijd te voorspellen, niet alleen bij aanvang van de wedstrijd, maar (juist) ook gedurende de wedstrijd. De kans dat een speler de wedstrijd zal winnen, wordt na elk gespeeld punt bijgesteld en dit leidt tot een kansgrafiek die zich ontrolt tijdens de wedstrijd. De methode is gebaseerd op een snel en flexibel computerprogramma, en op een statistische analyse van een grote dataset van Wimbledon, op wedstrijd- en op puntniveau. We zetten de methode kort uiteen en illustreren hem aan de hand van de Wimbledonfinale tussen Roger Federer en Rafael Nadal in 2007.

FRANC J.G.M. KLAASSEN & JAN R. MAGNUS

Tijdens een televisie-uitzending van een tenniswedstrijd krijgen de kijkers een aantal statistieken te zien. Uiteraard de stand, maar ook het percentage eerste services die in worden geslagen, het aan-

tal aces, en andere statistieken worden regelmatig getoond op het scherm. De commentatoren bediscussiëren deze statistieken om de kijkers meer inzicht te geven in verschillende aspecten van

de wedstrijd. Een statistiek over het belangrijkste aspect van de wedstrijd, namelijk wie zal winnen, wordt echter niet gegeven. We beschrijven nu een methode om die kans te schatten.

Er bestaat al een aantal methoden om de kans te schatten dat een speler de wedstrijd wint, bij *aanvang* van de wedstrijd. Men kan bijvoorbeeld kijken naar de inleg bij *bookmakers*. Of men kan een statistisch model gebruiken, zoals het model van Clarke en Dyte (2000) dat het aantal punten op de officiële (ATP en WTA) wereldranglijst gebruikt. Stel dat in een wedstrijd tussen speler A en speler B de initiële winkans 70% voor speler A is (en dus 30% voor speler B). Gedurende de wedstrijd komen nieuwe data beschikbaar en die kunnen worden gebruikt om de initiële kans bij te stellen. Als bijvoorbeeld A de eerste set heeft verloren, zal de kans dat A wint afnemen, maar de vraag is met hoeveel. Onze methode (Klaassen en Magnus, 2003) berekent niet alleen de kans dat A wint bij aanvang van de wedstrijd, maar (juist) ook gedurende het verloop van de wedstrijd, bij elk punt. Dit resulteert in een grafiek met elkaar opvolgende kansen dat een speler zal winnen, die langzamerhand zichtbaar worden gedurende de wedstrijd. Als de berekening uitkomt op meer dan 50% kans voor een speler, dan wordt voorspeld dat die speler de wedstrijd zal winnen. Dus de grafiek voorspelt ook de winnaar van de wedstrijd.

De grafiek en de onderliggende kansen zijn informatief voor de televisiekijkers. De score geeft aan wie er momenteel vóór staat in de wedstrijd, maar geeft geen goede indicatie van de mogelijke winnaar van de wedstrijd: een topspeler kan nog steeds de favoriet zijn ook nadat de eerste set is verloren. De score geeft ook maar gedeeltelijke informatie over het verloop van de wedstrijd: een score van 5-5 kan voorafgegaan zijn door 4-4, maar ook door 5-0. De wedstrijd Federer-Nadal die hieronder wordt beschreven is hiervan een goed voorbeeld. Samenvattende statistische gegevens, zoals het percentage eerste services in en het aantal

aces dragen hier weinig aan bij. Echter, een schatting van de kans dat A de wedstrijd zal winnen, geeft een directe aanwijzing van de mogelijke winnaar. En de grafiek met hoe groot de kans is dat een speler de wedstrijd wint na elk punt dat gespeeld is, geeft een overzicht van de ontwikkeling van de wedstrijd tot dan toe. Het geeft de informatie in één oogopslag, zodat het nuttig lijkt om de grafiek op televisie te laten zien ter ondersteuning van het commentaar.

Methode

Om de berekening van de kans dat een speler wint, en daarmee de hele grafiek, te bespreken maken we onderscheid tussen de kans vóór de wedstrijd begint (het eerste punt op de grafiek), en de kansen tijdens de wedstrijd (de rest van de grafiek). Om de eerste kans te schatten, gebruiken we een transformatie van de officiële *ranking* (positie op de wereldranglijst) van de spelers. Dit resulteert (bijvoorbeeld) in een kans van 60% dat speler A zal winnen van speler B. Natuurlijk is de *ranking* slechts één van de indicatoren om het onderlinge verschil tussen spelers aan te geven. Als er andere informatie beschikbaar is, zoals het feit dat een speler speciaal goed presteert op gras of het feit dat hij/zij last heeft van een blessure, kunnen aanpassingen plaatsvinden en kan de berekening op basis van de *ranking* worden verfijnd. Uiteindelijk zal er een schatting aan het begin van de wedstrijd zijn van bv. 70%. Klaassen en Magnus (2003) laten zien dat dergelijke aanpassingen uiteraard de grafiek iets doen verschuiven, maar dat het verloop van de grafiek niet veel verandert. Het nut van de grafiek hangt dus niet af van het exacte uitgangspunt.

Om de kans te schatten gedurende de wedstrijd, hebben wij een computerprogramma geschreven genaamd Tennisprob. We voeren eerst de specifieke regels van het toernooi in: een wedstrijd om

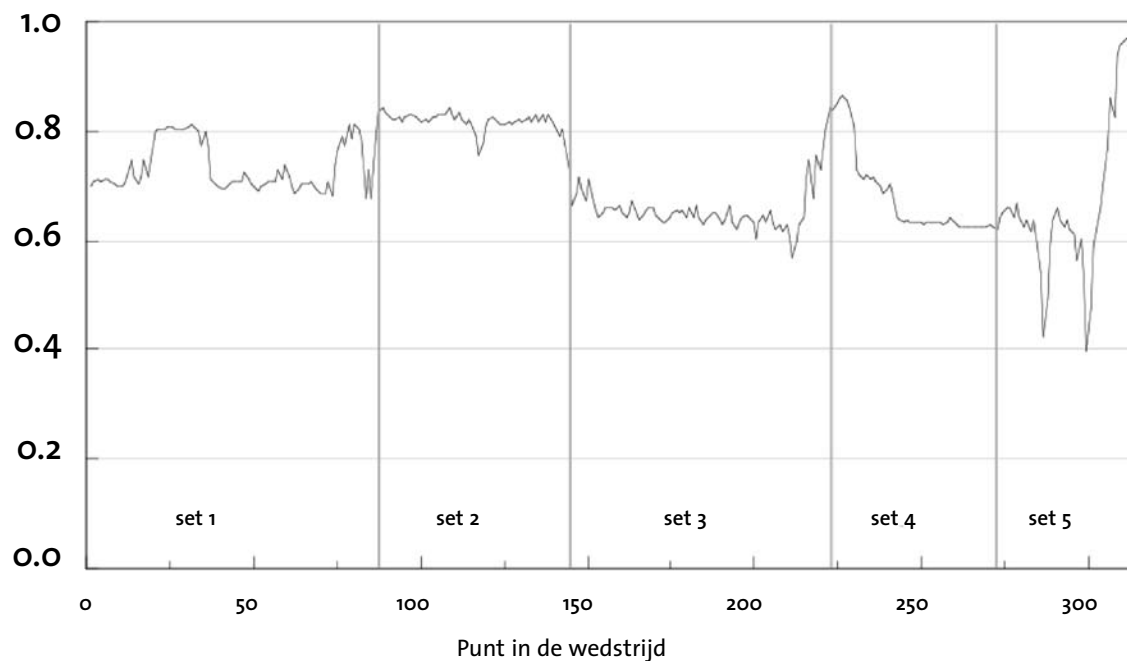
twee of drie gewonnen sets, wel of geen tiebreak in de laatste set. Gegeven de huidige score, wie er op dat moment serveert, gegeven ook de aanname dat het winnen van een servicepunt een identiek en onafhankelijk verdeeld proces is (zie Klaassen en Magnus, 2001, voor een rechtvaardiging van deze aanname), en gegeven twee in te voeren kansen, berekent Tennisprob de kans dat A wint, en wel op elk gewenst moment in de wedstrijd. Deze kans wordt exact en snel berekend (niet door simulatie, en binnen één seconde). De eerste in te voeren kans is de kans voor de wedstrijd begint, zoals hierboven geschat. De tweede in te voeren kans bestaat uit de som van twee kansen: de kans dat A een punt wint op service en de kans dat B een punt wint op service. Onze methode gebruikt de *ranking* om deze som te schatten. Dit zou bijvoorbeeld 130% kunnen zijn. Deze schatting hoeft niet erg precies te worden geschat omdat de kans die ons interesseert (dat A de wedstrijd wint) hier nauwelijks van afhangt. Deze twee stukjes informatie is alles wat we nodig hebben. In het bijzonder hoeven we niet de onderliggende punt-kansen te schatten. Alleen hun som én de initiële kans om de wedstrijd te winnen zijn voldoende. Dit is belangrijk omdat de laatste kansen veel robuuster geschat kunnen worden. Merk ook op dat er geen informatie nodig is over ontwikkelingen in de wedstrijd die na het huidig punt zullen plaatsvinden.

Om te laten zien hoe het voorspellen in de praktijk werkt, analyseren we de Wimbledon 2007 mannenfinale tussen Roger Federer en Rafael Nadal. Voordat de wedstrijd begint, moeten wij twee kansen invoeren zoals hierboven beschreven. De eerste is de kans dat Federer (speler A) de wedstrijd wint voordat de wedstrijd begint. Omdat hij nummer 1 is op de relevante ATP-wereldranglijst en Nadal nummer 2, krijgen we een kans van 60%. We weten echter dat Federer het toernooi de laatste vier achtereenvolgende jaren heeft gewonnen, en dit suggereert dat de

schatting op basis van de *ranking* alleen te laag is. Aan de andere kant, in de onderlinge ontmoetingen met Nadal was op dat moment de stand 4-9 (voornamelijk door Nadals overwinningen op gravel, maar Nadals optredens op snelle ondergronden was ook sterk verbeterd). Met deze informatie in het achterhoofd vonden wij een initiële kans van 70% een redelijk uitgangspunt. De tweede ingevoerde kans in 'Tennisprob' is de som van de kansen van beide spelers om een punt te winnen op service. Onze schatting op basis van *ranking* is 136%, daarmee aangevend dat er gemiddeld 68% kans is om een punt te winnen op de service. Dit lijkt redelijk.

Met deze ingevoerde kansen kunnen we de kans berekenen dat Federer de wedstrijd zal winnen, op elk punt in de wedstrijd. Sterker nog, als het punt is gespeeld en de nieuwe stand bekend is, wordt de nieuwe kans binnen één seconde weergegeven. Zo ontstaat een overzicht van de opeenvolgende kansen tijdens de wedstrijd. Omdat de wedstrijd die we als voorbeeld gebruiken al is gespeeld, kunnen we het complete overzicht weergeven, zie Grafiek 1.

Deze grafiek geeft een duidelijk inzicht in de wedstrijd. De eerste set werd door Federer gewonnen, maar de tweede set door Nadal (die Federer brak bij een stand van 4-5). De derde set was weer voor Federer, de vierde set voor Nadal na twee vroege *breaks*. De grafiek laat zien dat aan het begin van de vijfde set (er zijn 271 punten gespeeld) de kans dat Federer wint is gedaald van 70% naar 60%. De stand van de laatste set, 6-2, suggereert dat het een gemakkelijke overwinning was voor Federer. Maar uit de grafiek blijkt dat dit zeker niet het geval was. Sterker nog, op twee momenten was Federer in grote moeilijkheden. In beide gevallen stond hij achter met twee *breakpoints* (15-40) en werd verwacht dat hij de wedstrijd zou gaan verliezen. Maar, hij werkte de *breakpoints* weg en versloeg Nadal uiteindelijk met 7-6 / 4-6 / 7-6 / 2-6 / 6-2.



Figuur 1: Kans dat Federer de wedstrijd wint

Conclusie

De wedstrijd tussen Federer en Nadal is slechts een voorbeeld. Een grafiek kan worden gemaakt voor elke wedstrijd waar we twee variabelen invoeren voor aanvang van de wedstrijd en informatie over elk punt tijdens de wedstrijd. Daarom is de methode zoals hierboven beschreven een algemeen toepasbare voorspellingsmethode. Door informatie te geven over de mogelijke winnaar en het verloop van de wedstrijd geeft de grafiek (Figuur 1) extra informatie, naast de score en de samenvattende statistieken die men normaliter laat zien op de televisie. De informatie wordt ook in één oogopslag zichtbaar en kan direct worden opgeroepen. Daarom zou het interessant zijn om dit op televisie te laten zien, bijvoorbeeld om de twee games als de spelers van kant wisselen. Commentatoren kunnen de grafiek gebruiken om

de wedstrijd te becommentariëren, maar ook om deze na afloop te evalueren.

LITERATUUR

- Clarke S.R. en D. Dyte (2000). Using official rating to simulate major tennis tournaments. *International Transactions in Operational Research*; 7: 585-594.
- Klaassen F.J.G.M. en J.R. Magnus (2001). Are points in tennis independent and identically distributed? Evidence from a dynamic binary panel data model. *Journal of the American Statistical Association*; 96: 500-509.
- Klaassen F.J.G.M. en J.R. Magnus (2003). Forecasting the winner of a tennis match. *European Journal of Operational Research*; 148: 257-267.

FRANC J.G.M. KLAASSEN is als universitair hoofddocent verbonden aan de Faculteit Economische Wetenschappen en Econometrie van de Universiteit van Amsterdam en het Tinbergen Instituut.
E-mail: f.klaassen@uva.nl

JAN R. MAGNUS is als hoogleraar verbonden aan het Departement Econometrie & OR en het CentER van de Universiteit van Tilburg. E-mail: magnus@uvt.nl



OOST, WEST, THUIS BEST?

RUUD H. KONING

Thuisvoordeel is een zeer bekend fenomeen in de sport. Sporters die thuis spelen halen in het algemeen meer succes, en dat heeft er toe geleid dat oud-minister Winsemius er zelfs een boekje voor managers over heeft geschreven: 'Speel nooit een uitwedstrijd'. Dit roept wel de vraag op hoe groot thuisvoordeel nu eigenlijk is. Als je verder thuisvoordeel echt wilt gebruiken, is het ook nuttig te weten wat nu de oorzaken zijn van dat voordeel. In deze bijdrage gaan we nader in op meting van het thuisvoordeel. Het leidende voorbeeld is meting van thuisvoordeel in de Nederlandse eredivisie in het seizoen 2006/2007, maar dat voorbeeld kan ook voor andere sporten (en natuurlijk andere seizoenen) worden gebruikt.

Thuisvoordeel wordt in het algemeen toegerekend aan vier verschillende factoren: steun van het (thuis)publiek; bekendheid met de lokale omstandigheden; reistijd; spelregels.

De eerste factor behoeft weinig toelichting. Een voorbeeld van de tweede factor is de bekendheid van een sporter met de belijning in zijn eigen sporthal, terwijl die in een andere sporthal anders kan zijn. Reizen kan vermoeiend zijn, dus het bezoekende team of de bezoekende sporter heeft op dit gebied

een nadeel. In sommige sporten bieden de spelregels bepaalde voordelen aan het team dat thuis speelt, zo moeten uitspelende honkbal teams in de Verenigde Staten aan slag beginnen, zodat het thuisteam steeds weet hoeveel punten het moet maken om de wedstrijd te winnen.

Een recent overzicht van thuisvoordeel in verschillende sporten wordt gegeven in Stefani (2007). Hij analyseert alleen teamsporten, en definieert thuisvoordeel als de mate waarin een team thuis vaker wint dan in uitwedstrijden. De sport met het grootste thuisvoordeel is rugby (25.1%), gevolgd door voetbal (21.7%), NBA basketbal (21.0%), American Football (17.5%), NHL ijshockey (9.7%), en MLB honkbal (7.5%). De verschillen tussen de sporten zijn groot te noemen, waarbij het opvalt dat de sporten met het grootste thuisvoordeel continu zijn: spelers beginnen en zijn in het algemeen pas uitgespeeld bij het einde signaal. Vermoeidheid ten gevolge van reizen kan een verklaring zijn voor dat grote thuisvoordeel.

Deze aanpak om thuisvoordeel te meten heeft één groot voordeel – eenvoud – en twee nadelen: de maatstaf zegt niets over mogelijke variatie van thuisvoordeel tussen verschillende teams, en de maatstaf is ook niet goed toepasbaar in individuele sporten.

Voetbal

Thuisvoordeel in voetbal is van alle tijden, en van alle landen. Ruwweg de helft van alle wedstrijden wordt gewonnen door het thuisspelende team, een kwart eindigt in een gelijkspel, en een kwart wordt gewonnen door het team dat uit speelt. Echter, dit zijn gegevens op geaggregeerd niveau, voor een hele competitie over langere tijd. Voetballiefhebbers weten echter dat het soms spookt in de Euroborg, terwijl elders wordt geschamperd over 'Ajax publiek'. Thuisvoordeel zal niet hetzelfde zijn voor elke club, dus een iets fijnzinniger maat is nodig.

Een praktische procedure om thuisvoordeel te meten voor individuele teamsporten in een volledige competitie is voorgesteld door Clark en Norman (1995). Zij gaan er van uit dat het verwachte resultaat in een wedstrijd is toe te dichten aan twee factoren: thuisvoordeel en kwaliteitsverschil. Beide factoren worden voor elk team geschat. Nu zit er wel een addertje onder het gras als je thuisvoordeel voor individuele teams gaat meten. Stel dat de competitie uit drie teams zou bestaan, A, B, en C. Team A is het sterkste team, en wint zowel thuis als uit van B met 2-1, en van C met 3-1. Team B wint thuis en uit van C met 2-1. Er is geen thuisvoordeel, en alle teams scoren thuis even veel als uit, en krijgen thuis even veel goals tegen als uit. De doelsaldo's van de drie teams staan vermeld in de eerste twee kolommen van tabel 1. Er is geen thuisvoordeel, het doelsaldo van alle teams in thuis- en uitwedstrijden is even goed. Nu krijgt team C een thuisvoordeel van twee doelpunten, dus het speelt thuis gelijk tegen team A, en wint met 3-2 van team B. Het doelsaldo van team A in thuiswedstrijden is nu +3, en in uitwedstrijden is het +1, zoals ook in de laatste twee kolommen van tabel 1 staat. Echter, het lijkt nu alsof team A ook een thuisvoordeel heeft, terwijl dit niet het geval is. Het verschil in doelsaldo tussen thuis- en uitwedstrijden is geen goede maatstaf voor *individueel* thuisvoordeel,

aangezien het ook het gemiddelde thuisvoordeel in de gehele competitie bevat.

Clarke en Norman stellen dus een andere maatstaf van thuisvoordeel voor. Allereerst modelleren zij het doelpuntenverschil in een wedstrijd tussen teams i (thuis) en j (uit) als volgt:

$$w_{ij} = u_i - u_j + h_i + \varepsilon_{ij} \quad (1)$$

In deze vergelijking is w_{ij} het doelpuntenverschil in een wedstrijd. Dit is positief als het thuisteam wint, nul als de wedstrijd in een gelijkspel eindigt, en negatief als het uitteam wint. Dit verschil hangt af van drie factoren: het verschil in kwaliteit tussen beide teams ($u_i - u_j$), het thuisvoordeel van team i (h_i), en toevalsfactoren (het weer, een gemiste strafschoep) die worden gemodelleerd met een storingsterm (ε_{ij}), die verwachting 0 heeft en gelijke variantie voor elke waarneming. De verwachting van het resultaat is dus $Ew_{ij} = u_i - u_j + h_i$, ofwel, het verwachte resultaat hangt af van het kwaliteitsverschil ($u_i - u_j$) en het thuisvoordeel van het thuisspelende team h_i . Het resultaat van een individuele wedstrijd wordt uiteraard ook nog bepaald door het toeval, ε_{ij} , maar uiteindelijk zijn we meer geïnteresseerd in de parameters u en h , die kwaliteit en thuisvoordeel gedurende een heel seizoen weergeven, dan in toeval dat soms een

	GEEN THUISVOORDEEL		THUISVOORDEEL C	
	HGD	AGD	HGD	AGD
TEAM A	3	3	3	1
TEAM B	0	0	0	-2
TEAM C	-3	-3	1	-3

Tabel 1. Doelsaldo in thuis- en uitwedstrijden (zonder en met thuisvoordeel voor team C, HGD = doelsaldo thuiswedstrijden, AGD = doelsaldo uitwedstrijden).

belangrijke rol speelt bij de bepaling van de uitslag van een individuele wedstrijd. In zekere zin is het ontwikkelen en gebruiken van thuisvoordeel ook een kwaliteit, maar dat bedoelen we niet met de parameter u_i . u_i meet de kwaliteit van een team, als op een neutraal terrein zou worden gespeeld. Het gemiddelde van alle u 's is 0, zodat een team met een positieve u beter is dan het gemiddelde team, en een team met een negatieve u slechter. Thuisvoordeel heeft nu ook een natuurlijke interpretatie: het is het verwachte doelpuntenverschil als beide teams even goed zijn (dus als $u_i - u_j = 0$). De restrictie dat de som van alle kwaliteitsparameters 0 is, is noodzakelijk, anders is kan het model niet worden geschat. Als in vergelijking (1) alle u 's met een constante worden verhoogd, verandert het verschil niet, en dus wordt de kansverdeling van de geobserveerde grootheid (het doelpunten-

verschil) dan niet uniek bepaald door de parameters. De identificerende restrictie $\sum_i u_i = 0$ lost dit probleem op. Een andere identificerende restrictie zou kunnen zijn $u_1 = 0$, dus de de kwaliteit van het eerste team is 0, en de kwaliteitsparameters van alle andere teams worden dus gemeten ten opzichte van team 1. Echter, het is informatiever om te weten dat een bepaald team een positieve u heeft, en dus beter dan gemiddeld is, dan dat dit team beter is dan team 1.

De parameters u_i en h_i kunnen geschat worden met de methode der kleinste kwadraten, maar het kan ook op een equivalente, andere manier. We gebruiken tabel 2, waarin de gegevens van de Nederlandse eredivisie van het seizoen 2006/2007 staan vermeld. Er namen 18 teams deel aan deze competitie, dus elk team speelt 17 thuiswedstrijden. In die tabel staan gegevens over thuiswed-

team	HW	HD	HL	H.f	H.a	HGD	AW	AD	AL	A.f	A.a	AGD	GD	p	h	u
1 Ajax	12	3	2	44	12	32	11	3	3	40	23	17	49	75	0.257	1.535
2 AZ	10	6	1	44	13	31	11	3	3	39	18	21	52	72	-0.055	1.774
3 ADO Den Haag	2	4	11	19	36	-17	1	4	12	21	36	-15	-32	17	-0.805	-0.184
4 Excelsior	6	3	8	27	28	-1	2	3	12	16	37	-21	-22	30	0.570	-0.594
5 Feyenoord	10	5	2	29	24	5	5	3	9	27	42	-15	-10	53	0.570	-0.260
6 FC Groningen	8	4	5	32	26	6	7	2	8	22	28	-6	0	51	0.070	0.267
7 sc Heerenveen	10	4	3	35	14	21	6	3	8	25	29	-4	17	55	0.882	0.333
8 Heracles Almelo	7	6	4	27	19	8	0	5	12	5	45	-40	-32	32	2.320	-1.747
9 NAC Breda	6	7	4	22	21	1	6	0	11	21	33	-12	-11	43	0.132	-0.069
10 N.E.C.	8	3	6	22	20	2	4	5	8	14	24	-10	-8	44	0.070	0.045
11 PSV	15	0	2	53	14	39	8	6	3	22	11	11	50	75	1.070	1.156
12 Roda JC	11	2	4	29	14	15	4	7	6	18	22	-4	11	54	0.507	0.354
13 Sparta Rotterdam	6	5	6	20	24	-4	4	2	11	20	42	-22	-26	37	0.445	-0.642
14 FC Twente	13	3	1	47	15	32	6	6	5	20	22	-2	30	66	1.445	0.413
15 FC Utrecht	11	4	2	30	11	19	2	5	10	11	33	-22	-3	48	1.882	-0.722
16 Vitesse	7	4	6	30	23	7	3	4	10	20	32	-12	-5	38	0.507	-0.090
17 RKC Waalwijk	5	5	7	19	24	-5	1	4	12	14	36	-22	-27	27	0.382	-0.639
18 Willem II	7	2	8	21	27	-6	1	5	11	10	37	-27	-33	31	0.632	-0.931

Tabel 2. Berekening individueel thuisvoordeel.

strijden (in de kolommen met een H), en gegevens over uitwedstrijden (kolommen met een A). De eerste kolom is het aantal gewonnen thuiswedstrijden (HW), die wordt gevolgd door het aantal gelijk geëindigde thuiswedstrijden (HD) en het aantal verloren thuiswedstrijden (HL). Vervolgens wordt voor de thuiswedstrijden het aantal doelpunten voor ($H.f$), tegen ($H.a$) en het doelsaldo in thuiswedstrijden (HGD) gegeven. De laatste twee kolommen geven de schattingen voor thuisvoordeel h en kwaliteit u . Die zijn als volgt berekend:

1. H is het gemiddelde thuisvoordeel voor de gehele competitie, $H = \sum_i HGD_i/17$. In dit geval vinden we $H = 11$.
2. Het thuisvoordeel voor elk team is $h_i = (HGD_i - AGD_i - H)/16$, dus het verschil van het doelsaldo in thuis- en uitwedstrijden, verminderd met het gemiddelde thuisvoordeel, gedeeld door 16. Er moet door 16 worden gedeeld, omdat de waarnemingen $HGD_i - AGD_i - H$ aan twee restricties voldoen: $H = \sum_i HGD_i/17$ en $\sum_i HGD_i + \sum_i AGD_i = 0$.
3. Kwaliteit tenslotte is dat deel van het doelsaldo in thuiswedstrijden dat niet te danken is aan thuisvoordeel: $u_i = (HGD_i - (18-1) \times h_i)/18$.

De resultaten van deze rekenpartij staan in de laatste twee kolommen van tabel 2. Het team met de hoogste kwaliteit was AZ Alkmaar, maar hun thuisvoordeel was laag. Als elke competitiewedstrijd op neutraal terrein zou zijn gespeeld, was AZ misschien wel kampioen geworden. Aan de andere kant zien we ook dat de kwaliteit van PSV lager is dan die van Ajax, maar dat het thuisvoordeel in Eindhoven de doorslag heeft gegeven. Uit de tabel is in elk geval duidelijk dat het ene team een veel groter thuisvoordeel heeft dan het andere team. Het model is in staat om ruim 40% van de variatie in wedstrijdresultaten te verklaren. De geschatte residuen $\hat{\epsilon}_{ij}$ volgen inderdaad bij benadering een normale verdeling.

De Clarke-Norman methode voor het schatten

van thuisvoordeel is bruikbaar voor alle sporten die in een volledige competitie worden gespeeld, zoals voetbal, hockey, waterpolo en basketbal. Een nadeel van deze methode is dat die niet bruikbaar is om thuisvoordeel in individuele sporten te schatten.

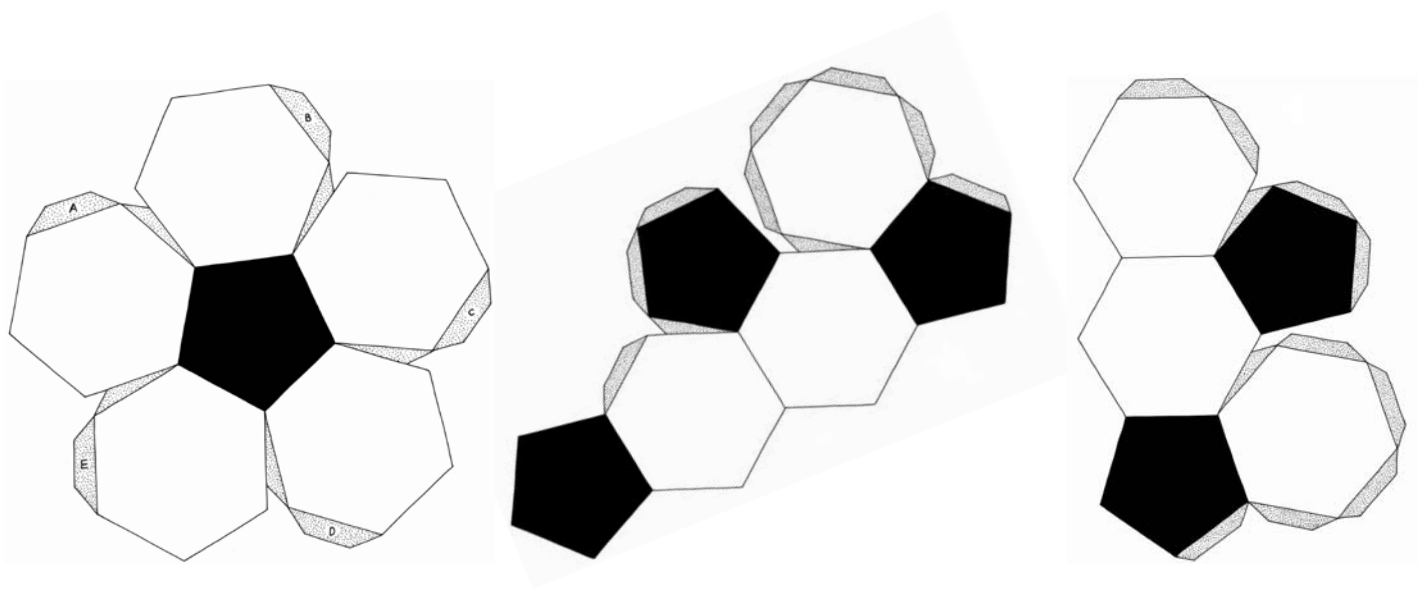
Conclusie

Thuisvoordeel is belangrijk in sport. Goede meting van thuisvoordeel is echter moeilijk, omdat sportwedstrijden niet als experiment worden uitgevoerd. De vorm van de competitie (volledige competitie met uit- en thuiswedstrijden, of een toernooi vorm waarbij niet elke speler tegen elkaar speelt) bepaalt in grote mate in hoeverre thuisvoordeel meetbaar is, en welke maatstaf handig is. In het voorbeeld is gebleken dat thuisvoordeel niet gelijk is voor elk team in de Nederlandse eredivisie in het seizoen 2006-2007. De methode die is besproken om thuisvoordeel te schatten is goed toepasbaar in sporten, waarin een volledige competitie met uit- en thuiswedstrijden wordt gespeeld. Thuisvoordeel in individuele sporten, die vaak in toernooivorm worden gespeeld, is moeilijker (zie bijvoorbeeld voor meting van thuisvoordeel in tennis Koning (2008)). Toch is ook voor verschillende individuele sporten het belang van thuisvoordeel aangetoond.

LITERATUUR

- Clarke, S.R. and J.M. Norman (1995). Home ground advantage of individual clubs in English soccer. *The Statistician* 44(4), 509-521.
- Koning, R.H. (2008). Home advantage in professional tennis. Manuscript.
- Stefani, R. (2007). Measurement and interpretation of home advantage. In: J. Albert and R.H. Koning (eds), *Statistical Thinking in Sports*, Boca Raton: Chapman & Hall/CRC, pp. 203-216.

RUUD H. KONING is als hoogleraar sporteconomie verbonden aan de Faculteit Economie en Bedrijfskunde van de Rijksuniversiteit Groningen. E-mail: r.h.koning@rug.nl



HET PLANNEN VAN DE BELGISCHE VOETBALCOMPETITIE

Elke voetbalcompetitie heeft een planning nodig. Zo'n planning maakt duidelijk wie tegen wie speelt, en wanneer en waar dat moet gebeuren. Die planning kan natuurlijk invloed hebben op de belangen van de betrokken partijen, en zelfs op de uitkomst van de competitie. Dit artikel beschrijft onze ervaringen bij het plannen van de hoogste klasse van het Belgisch voetbal, de Jupiler league. We verduidelijken de manier waarop we het opstellen van de voetbalkalender hebben geautomatiseerd en verbeterd. De resulterende kalenders zijn gebruikt in de seizoenen 2006-2007 en 2007-2008, en de verwachting is dat ook het komend seizoen de hieronder beschreven aanpak weer gebruikt gaat worden om een kalender samen te stellen. Voor een meer gedetailleerde beschrijving van onze aanpak, inclusief referenties, verwijzen we naar Goossens en Spieksma (2007).

DRIES GOOSSENS EN FRITS C.R. SPIEKSMASMA

Geen sinecure

De bekendmaking van een nieuwe voetbalkalender heeft in het verleden al vaak stof doen opwaaien. Naar aanleiding van het verschijnen van de kalender voor het seizoen 2005-2006, verscheen een artikel in De Standaard met als titel

'Genk boos' (Lambaerts 2005), waarin toenmalig algemeen directeur van KRC Genk Paul Heylen zijn ongenoegen uit: 'Voor onze club is dit een heel onevenwichtige kalender. Voor de winterstop ontvangen we de clubs uit de top vijf van vorig seizoen. Bovendien krijgen we ook onze burens van Sint-Truiden op bezoek. Dat betekent dat we vijf

cruciale matches in de terugronde op verplaatsing moeten afwerken. Hoe gaan we dit seizoen onze thuiswedstrijden in de tweede competitiehelpt commercieel interessant kunnen houden? ... Dit is niet ernstig.'

Een paar maand later, na de Champions League wedstrijd Bayern München – Club Brugge, tekennen we een boze reactie op bij Antoine Vanhove, algemeen directeur bij Club Brugge (Reunes 2005): 'Dit is nu al de vijfde of de zesde keer dat we na een Europese verplaatsing naar Charleroi moeten. Zij hopen op die manier te profiteren van onze eventuele vermoeidheid. Als ik zie wie er voorzitter is van de kalendercommissie, concludeer ik dat het allemaal geen toeval is.' De kalendercommissie is het orgaan binnen de Belgische voetbalbond dat bevoegd is voor het opstellen van de kalender voor de eerste klasse en heeft als voorzitter Charleroi-secretaris Pierre-Yves Hendrickx. In de kalendercommissie zijn echter nog 5 ploegen vertegenwoordigd, waaronder ook Club Brugge.

Uit deze reacties blijkt dus niet alleen dat de kalender door een aantal ploegen als oneerlijk gezien wordt, maar ook dat het tot stand komen van die kalender een mistig proces is, wat tot allerlei verdachtmakingen leidt en waar zelfs ploegen die in de kalendercommissie vertegenwoordigd zijn blijkbaar weinig zicht op hebben.

De Jupiler league

De hoogste klasse van het Belgisch voetbal, genaamd de 'Jupiler league', telt 18 ploegen, die elk twee keer tegen elk ander team spelen (1 keer in eigen stadion, 1 keer in dat van de tegenstander). Dit leidt tot 34 speeldagen, waarvan de eerste 17 de heenronde worden genoemd, en de volgende 17 de terugronde. Rekening houdend met wedstrijden van de nationale ploegen en een winterstop, kiest de voetbalbond exact 34 weekends waarop gevoetbald wordt. Traditioneel

is de terugronde het spiegelbeeld van de heenronde: dezelfde opeenvolging van wedstrijden, maar met het thuisvoordeel omgedraaid. Verder mogen ploegen ook niet meer dan twee opeenvolgende thuis- of uitwedstrijden spelen, wat men een 'break' noemt. Het totaal aantal breaks in de competitie moet minimaal zijn, en het is niet wenselijk dat een ploeg het seizoen start of eindigt met een break.

Voor wij in het verhaal betrokken werden, werd de kalender opgesteld door Robert Sterckx, secretaris-generaal van de Profliga. Naast de bovenstaande vereisten krijgt hij te maken met een groot aantal uiteenlopende wensen van alle betrokken partijen. Zo wil bijvoorbeeld de politie van Sint-Truiden half februari geen thuiswedstrijd tegen een topploeg, wegens carnaval, en wil Charleroi het seizoen openen met een thuismatch. In Gent speelt men dan weer liever niet samen thuis als Club Brugge ook een thuiswedstrijd heeft. Belgacom TV is het meest gebaat met een evenwichtige spreiding van de interessante wedstrijden over het seizoen, en natuurlijk met een competitie die spannend blijft tot de laatste speeldag. Bovendien wordt er een kalender verwacht die niemand financieel of sportief benadeelt.

Een monnikenwerk

Om deze taak tot een goed einde te brengen vertrekt Robert Sterckx van een zogenaamd basisrooster. Een basisrooster geeft voor elke ploeg, voorgesteld door een getal, aan wat de opeenvolgende tegenstanders zijn, en of de wedstrijd thuis of op verplaatsing dient te worden afgewerkt. Tabel 1 toont de eerste 7 speeldagen van een basisrooster, waaruit bijvoorbeeld kan afgeleid worden dat ploeg 6 het seizoen thuis opent tegen ploeg 15, om dan naar ploeg 17 te trekken en op de derde speeldag terug thuis te spelen tegen ploeg 2, enz. Om een kalender te bekomen moet elk getal aan

1	2	3	4	5	6	7
1-3	2-4	1-7	2-8	1-11	2-12	1-15
4-17	3-18	3-5	4-6	3-9	4-10	3-13
6-15	5-1	6-2	5-18	5-7	6-8	5-11
8-13	7-16	8-17	7-3	8-4	7-18	7-9
10-11	9-14	10-15	9-1	10-2	9-5	10-6
12-9	11-12	12-13	11-16	12-17	11-3	12-4
14-7	13-10	14-11	13-14	14-15	13-1	14-2
16-5	15-8	16-9	15-12	16-13	15-16	16-17
18-2	17-6	18-4	17-10	18-6	17-14	18-8

Tabel 1 Basisrooster (speeldag 1-7)

een ploeg toegewezen worden; de wedstrijden volgen dan direct uit het basisrooster. Deze oefening wordt met de hand uitgevoerd, gebruikmakend van Excel om de toewijzingen te visualiseren. Hierbij wordt eerst gestreefd naar een gunstige toewijzing voor de topploegen (Anderlecht, Club Brugge, Standard Luik, en Racing Genk), om dan de andere ploegen één voor één een nummer toe te wijzen, rekening houdend met zoveel mogelijk beperkingen. Nadien wordt er geprobeerd de toewijzing van een paar ploegen om te wisselen, om op die manier de kalender verder te verbeteren.

Het zal niemand verwonderen dat met deze aanpak slechts aan een beperkt aantal wensen voldaan kon worden. Aangezien de lokale overheden de bevoegdheid hebben om een wedstrijd te verbieden werd er dan ook vooral met hun wensen rekening gehouden. De wensen van de televisie en de clubs werden daarentegen voor een groot deel terzijde geschoven, met de hierboven geschetste uitingen van onvrede in de media tot gevolg. Dit was voor ons de aanleiding om onze diensten aan te bieden. Nadat Sylvie Demasure in haar master thesis (Demasure 2005) kon aantonen dat er een veel betere kalender voor het toemalige seizoen 2005-2006 mogelijk geweest was mits een modelmatige aanpak, werden we door

de Belgische voetbalbond uitgenodigd om te werken aan een kalender voor het nieuwe seizoen.

Wat is van belang voor een goede kalender?

In overleg met de kalendercommissie en de betrokken partijen hebben wij geprobeerd duidelijkheid te scheppen in de wirwar van wensen omtrent de voetbalkalender. De eisen van politie en lokale overheid waren al snel terug te leiden tot drie types:

- ploeg x mag niet thuis spelen op een speeldag s
- ploeg x mag niet thuis spelen als ploeg y ook thuis speelt
- ploeg x mag thuis geen risicowedstrijd spelen op een speeldag s

De achterliggende reden voor deze eisen was meestal een gebrek aan manschappen. Zo kan er op speeldag s al een ander evenement gepland staan dat politieaandacht vereist, of kunnen ploegen x en y hun stadion in dezelfde regio hebben, zodat de belasting bij de politie te groot zou zijn bij twee gelijktijdige thuiswedstrijden. Een risicowedstrijd is een wedstrijd waar onregelmatigheden tussen supportersclans te verwachten zijn. Het is aan de politieafdelingen in elke regio om in te schatten welke wedstrijden dat zijn voor hun club(s).

Ook Belgacom TV slaagde erin haar wens voor een aantrekkelijke competitie concreet te formuleren. Er werd gevraagd om ten hoogste één topwedstrijd (een topwedstrijd is een wedstrijd tussen twee topploegen) per speeldag te hebben, en bij voorkeur gespreid over het seizoen (maar niet in het begin). Verder moesten wij streven om op zo veel mogelijk speeldagen een uitwedstrijd te hebben voor twee van de vier topclubs. De achterliggende motivatie is dat een thuiswedstrijd van een topploeg met betrekking tot kijkcijfers veel minder interessant is, aangezien de topploeg die wedstrijden meestal zonder veel spanning wint.

De clubs zelf waren verantwoordelijk voor de grootste diversiteit aan wensen, gebaseerd op zowel economische als sportieve redenen. Na overleg met de kalendercommissie bleken de clubs bereid te zijn zich te houden aan de volgende typen wensen:

- geen thuis-/uitmatch op speeldag s
 - geen gelijktijdige thuismatch met ploeg x
 - thuiswedstrijd tegen ploeg x in heen-/terugronde
 - geen wedstrijd tegen een topploeg op speeldag s
- Verder namen wij voor elke club de wens op om minimaal één topploeg thuis te mogen ontvangen in zowel heen- als terugronde en zorgden we voor ten hoogste twee confrontaties met topploegen per reeks van vier opeenvolgende wedstrijden. Om conflicten met de Champions League en de UEFA Cup kalender te vermijden, legde de kalendercommissie een lijstje voor met speeldagen waarop ze liever geen wedstrijd zagen tussen topploegen of ploegen die ook op het Europese toneel actief waren (zie ook Demasure (2005)).

Tijdens het inventariseren van de wensen werd al snel duidelijk dat het onmogelijk zou zijn om aan alle wensen te voldoen, aangezien ze conflicterend waren. Daarop heeft de kalendercommissie elke wens ingedeeld in één van vijf prioriteitsniveaus. De hoogste prioriteit werd gegeven aan dwingende beperkingen met betrekking tot de beschikbaarheid van het stadion. Het spreekt bijvoorbeeld vanzelf dat twee ploegen die eenzelfde stadion delen niet op het zelfde moment thuis kunnen spelen. Prioriteit twee werd voorbehouden voor vereisten van politie en lokale overheid, aangezien zij het recht hebben om een wedstrijd te verbieden als de veiligheid niet gegarandeerd kan worden. De overige wensen, waaronder die van de clubs en de televisie, werden verdeeld over de resterende drie prioriteitsniveaus, waarbij de kalendercommissie een afweging maakte tussen de onderliggende financiële of sportieve redenen en de *fairness* van de kalender als geheel.

Een kalender gebruikmakend van het basisrooster: geheeltallig programmeren

In onze zoektocht naar een betere kalender hebben we ons - in eerste instantie - op verzoek van de kalendercommissie aan het basisrooster gehouden. Het belang van elk prioriteitsniveau hebben we kwantitatief uitgedrukt, in de vorm van een aantal strafpunten dat opgelopen wordt bij schending van een wens van dat niveau. Tenslotte hebben we een geheeltallig programmeringsmodel opgesteld dat het toewijzen van nummers aan teams, zoals tot voorheen met de hand werd gedaan, optimaliseert. Een optimale kalender is er dan natuurlijk een met een minimaal aantal strafpunten. Het resultaat werd voorgesteld aan de kalendercommissie, die prompt vaststelde dat een aantal wensen klaarblijkelijk een te gering belang gekregen had, terwijl het belang van andere wensen was overschat. Na een aantal aanpassingen en vergaderingen kwamen wij tot een kalender waarin de leden van de kalendercommissie zich konden vinden, en die officieel werd vastgesteld als de kalender voor het seizoen 2006-2007.

Naar een aanpak in twee fasen

Ondanks het feit dat het basisrooster een aantal heel interessante eigenschappen heeft (o.a. een minimaal aantal breaks), is het uiteraard ook erg beperkend. Afstappen van het basisrooster vergroot dan ook de zoekruimte en geeft zo extra mogelijkheden om een betere kalender te vinden. Om efficiënt te kunnen zoeken, werkten we volgens een methode in twee fasen, waar de eerste fase inhoudt dat voor elke ploeg wordt vastgelegd op welke speeldagen die ploeg thuis speelt en op welke speeldagen er een uitmatch is. In deze fase moet alleen rekening gehouden worden met beperkingen omtrent de beschikbaarheid van een

stadion, beperkingen die stellen dat een paar ploegen niet samen thuis mag spelen, en beperkingen van Belgacom TV omtrent het aantal top-ploegen dat op verplaatsing speelt op elke speeldag. In een tweede fase wordt dan vastgelegd wat de eigenlijke tegenstanders zijn op elke speeldag en worden alle andere beperkingen in rekening genomen, met als resultaat een kalender. Deze gefaseerde aanpak werd gebruikt om de kalender voor het seizoen 2007-2008 op te stellen.

Resultaten

Het is niet eenvoudig om de drie besproken methodes (manuele planning, planning volgens een basisrooster, twee fasen aanpak) te vergelijken, aangezien ze werden toegepast op drie verschillende seizoenen. Bovendien werden er in de loop van de jaren telkens meer wensen en beperkingen geopperd werden. Desalniettemin bleef het aantal ploegen en het type wensen constant. Tabel 2 geeft een greep uit de belangrijkste resultaten.

Eerst en vooral blijkt dat de kwaliteit van de gevonden roosters sterk verbeterd is. Uit Tabel 2 blijkt bijvoorbeeld dat er in het seizoen 2005-2006 aan 70% van de wensen van de politie werd vol-

daan, terwijl dit steeg tot 95% voor het seizoen 2006-2007 en zelfs tot 100% met de gefaseerde aanpak. Sinds de manuele aanpak werd verlaten kan aan twee derden van de clubwensen worden voldaan. Dit betekent een verdubbeling, en dit ondanks een steeds stijgend aantal clubwensen. Door de wensen zorgvuldig over de prioriteitsniveaus te verdelen, kon bovendien voor elke club aan minstens één wens worden voldaan. In de manuele kalender moest Belgacom TV het stellen met 10 speeldagen met twee topploegen die op verplaatsing aantreden. In het seizoen 2006-2007 konden we Belgacom TV al 26 wedstrijden aanbieden met uitwedstrijden voor twee topteam. Deze resultaten zijn zelfs nog verder verbeterd met de twee fasen aanpak. Tenslotte ontvangen ook alle 18 ploegen nu in zowel heen- als terugronde ten minste één topploeg thuis, waar dat met de manuele aanpak slechts voor 13 ploegen het geval was.

Verder kost de geautomatiseerde aanpak een stuk minder rekentijd dan de manuele aanpak. Aangezien er niet veel tijd is tussen het moment waarop bekend is welke ploegen in de competitie zullen aantreden, en het moment waarop de kalender af moet zijn, is die geringe rekentijd een niet te onderschatten voordeel.

	Manuele planning (2005-2006)	Basisrooster (2006-2007)	2 fasen aanpak (2007-2008)
Rekentijd	1 week	4-5 uur	5-10 minuten
Doelfunctiewaarde	> 75,000	11,698	2,144
Politiewensen	70%	95%	100%
Clubwensen	32%	68%	66%
Aantal speeldagen met uitwedstrijden voor topploegen (2/1/0)	10 / 24 / 0	26 / 8 / 0	28 / 6 / 0
Aantal ploegen met topwedstrijden thuis in beide seizoenshelften	13	18	18

Tabel 2: Vergelijking kalenders volgens 3 methodes

Tot slot

De reacties op de nieuwe aanpak en de kalender die daar uit volgde waren over het algemeen zeer positief (Cuvelier 2006). Zowel voor de pers als voor de clubs werd immers meer duidelijkheid geschaapen over het tot stand komen van de kalender. Net zoals in de vorige kalenders kon niet aan elke wens van elke betrokken partij worden voldaan. Via onze aanpak konden wij echter motiveren dat het niet mogelijk was om aan al die wensen te voldoen, zonder daarvoor minstens evenveel andere gelijkwaardige wensen te moeten opgeven. Dat de ideale kalender evenwel niet bestaat, mag blijken uit een reactie van Guy Mangelschots, voormalig trainer van Sint-Truiden: 'Als je na vijf matches de balans opmaakt van de punten die je hebt, kan je pas echt zeggen of de kalender goed of slecht was' (Martens 2007).

REFERENTIES

- Cuvelier M. (2006). Voetbalkalender voor volgend seizoen bekend. *Het Nieuwsblad*, 12 juni 2006, VUM.
- Demasure, S. (2006) *Wedstrijdplanning van de nationale voetbalcompetitie*. Master thesis: K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen.
- Goossens, D. en F.C.R. Spieksma. (2007) *Scheduling the Belgian Soccer League*. Research Report KBI_0732, K.U.Leuven, Faculteit Economie en Bedrijfswetenschappen.
- Lambaerts, G. (2005). Voetbalkalender eerste klasse 2005-2006: Genk boos. *De Standaard*, 15 juni 2005, VUM.
- Martens, L. (2007). Wij zijn niet ontevreden. *Het belang van Limburg*, 20 juni 2007, Concentra.
- Reunes, M. (2005). Club boos op Charleroi. *Het Nieuwsblad*, 29 september 2005, VUM.

DRIES GOOSSENS is verbonden aan de onderzoeksgroep Operationeel Onderzoek en Bedrijfsstatistiek (ORSTAT) van de Katholieke Universiteit Leuven.
E-mail: Dries.Goossens@econ.kuleuven.be

FRITS C.R. SPIEK SMA is verbonden aan de onderzoeksgroep Operationeel Onderzoek en Bedrijfsstatistiek (ORSTAT) van de Katholieke Universiteit Leuven.
E-mail: Frits.Spieksma@econ.kuleuven.be

column



Piz Bernina

FRED STEUTEL

Er is heel veel geschreven over statistiek in de sport, maar veel minder over sportende statistici. Sporten statistici wel, en zo ja, wat doen ze dan allemaal?

Een steekproef ter grootte één lijkt de eerste vraag met 'ja' te beantwoorden: ikzelf heb door de jaren heen vrij veel gesport, van alles. Van mijn negende tot mijn elfde was ik lid van gymnastiekvereniging Vlugheid en Kracht, met elke zomer een openbare uitvoering; ik kijk daar met weinig vreugde op terug. In die tijd heb ik ook leren schaatsen (niet zo erg goed). Ik heb een beetje gevoetbald (kon ik niet), gehockeyd (ietsje beter; had in een competitiewedstrijd ooit een gelukkige *hattrick*) en getennist (nog iets minder slecht). Op de middelbare school heb ik korte tijd gebokst; de sportleraar had een partij goedkope oefenhandschoenen op de kop getikt, maar ik hield niet van tikken op mijn kop. In die tijd ook wat hard gelopen op schoolwedstrijden (kon ik tamelijk goed,

zelfs zonder trainen; heb ik niet doorgezet). In mijn studententijd even gejudood (één diploma-tje gehaald). Daarna lange tijd niets. In Twente (THT) vrij fanatiek gebadmintond, beetje geten-nist en weer wat hard gelopen. In Eindhoven – was toen begin veertig – tijdje vrij intensief gesquasht en wat serieuzer gaan hardlopen. Was lange tijd lid (later zelfs erelid) van de Karpendonk Road Runners, onder leiding van de Eindhovense statis-ticus en Nederlandse seniorenkampioen 100 km hardlopen. Ik heb vijf maal een marathon gelopen (snelste tijd 3h,16m, 01 s - bruto). Ben lid geworden van een schaatsclub; heb beter leren schaatsen; ‘zilveren schaatsje’ verdiend met 62 rondjes in een uur. Mijn snelste rondje (400 meter) ooit ging in 50 seconden; Jan Bos doet dat in de helft van die tijd. In die periode ook met een collega kansreke-naar van de VU verschillende tochten gereden, langste tocht 80 km.

Maar, ik heb ook een iets grotere steekproef: de bijeenkomsten van ‘stochastici’ (kansrekenaars en mathematisch statistici) in Lunteren. Daar werd door de deelnemers niet alleen gebiljart, getafeltennist, geschaakt en gewandeld, maar ook hard gelopen. Er werden door een vrij grote groep statistici in de lange middagpauze (van 13.30 tot 15.30) grote afstanden afgelegd, deels door het mulle zand. De aanvoerder was ook hier de Eindhovense honderd-kilometerkampioen. De grootste afstand die ik in Lunteren heb meegelopen was 26 kilometer; sommigen van ons, waar-onder ikzelf hadden pas een marathon achter de rug en waren dus goed getraind. Bij die gelegenheid bleef een minder goed getrainde collega – tot veler sadistisch genoeg – kokhalzend in de berm achter. Echte mannen!

Op nog wat grotere schaal vind ik op internet de Association of Road Running Statisticians. Een internationale club van mensen die hardlooppres-taties bijhouden voor afstanden van 3000 meter en meer. Ze houden niet alleen de cijfers bij, het zijn meestal ook zelf hardlopers.

Nog een aantal individuele voorbeelden. Een statisticus in Tilburg kan behalve vrij goed biljar-ten ook heel hard fietsen, Luik-Bastenaken-Luik en zo. Ik kende een Amerikaanse OR-man die lang geleden schaakkampioen was van de Bronx. Hij was opgetogen over de winst van Fischer op Spasky in 1972: ‘*He didn’t beat him, he clobbered him!*’ Statisticus Jan Hemelrijk kon goed tafel-tennissen; hij was, meen ik, ooit kampioen van Alkmaar.

Het is niet alleen maar vrolijkheid bij spor-tende statistici. Vrij veel wiskundigen, onder wie statistici/kansrekenaars, waren bergbeklim-mers en verschillende daarvan zijn bij het klim-men omgekomen. De bekendste kansrekenaar onder hen is *Oxford prodigy* Rollo Davidson, die in juni 1970 bij beklimming van de Piz Bernina in de Alpen om het leven kwam. Te zijner her-innering zijn door Kendall en Harding twee prachtige boeken uitgegeven: *Stochastic Algebra en Stochastic Analysis*. Er is ook een presti-gieuze prijs naar hem genoemd. Afgelopen jaar is de Rollo Davidsonprijs prijs toegekend aan de Eindhovense kansrekenaar Remco van der Hofstad.

Er zijn natuurlijk ook statistici die nooit spor-ten. Zo heb ik van David Van Dantzig, Stan van Eeden, Theo Runnenburg of Harry Kesten nooit enige sportprestatie vernomen maar volgens Theo – mij in zijn laatste dagen verteld – ging zijn vriend Harry soms trimmen.”

Voor statistici heeft sporten iets paradoxaals: mensen die sporten worden misschien niet ouder, maar ze worden wel gezonder oud. Daar staat tegenover dat zij tijdens het sporten (zeg vijf of tien procent van de dag) aan een verhoogd risi-co blootstaan. Hoe deze risico’s moeten worden gewogen is een vraag die buiten het bestek van dit verhaal valt.

FRED STEUTEL is emeritus hoogleraar kansrekening aan de TU Eindhoven. Hij is redacteur van STATOR.
E-mail: f.w.steutel@tue.nl



SVEN KRAMER PASSEERT ERIC HEIDEN

De beste schaatser en schaatsster aller tijden

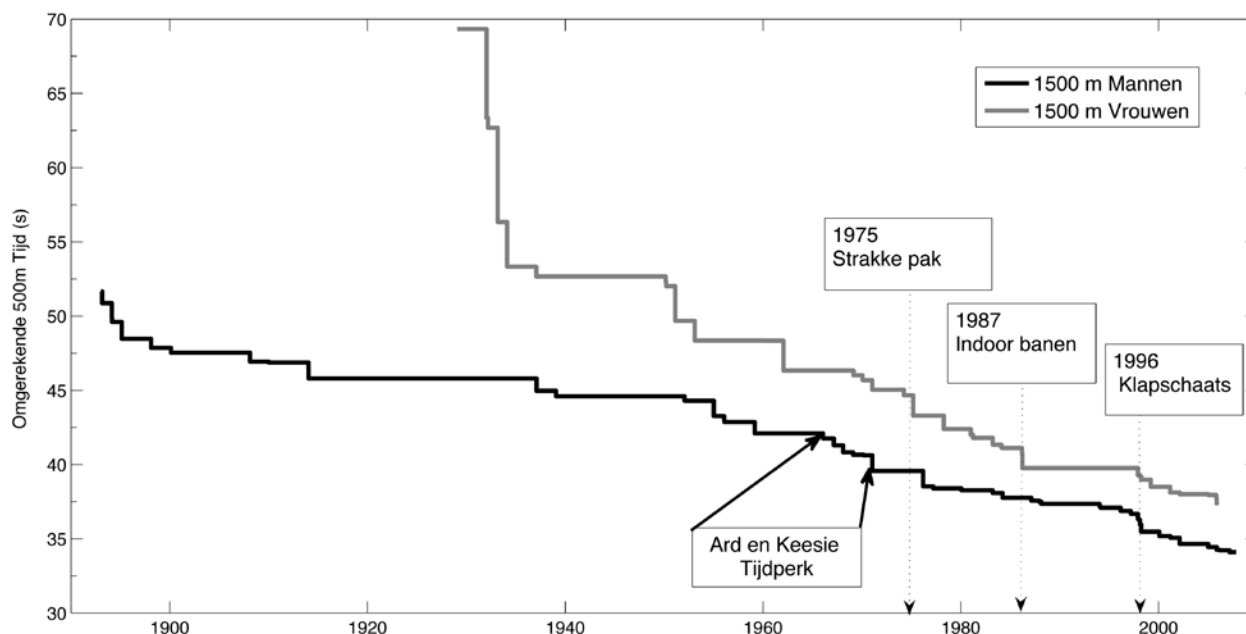
Wie zijn de beste schaatsers en schaatssters ooit? Ard Schenk of Sven Kramer, Yvonne van Gennip of Ireen Wüst? Uit de meer dan 3000 schaatsers en schaatssters, die sinds 1893 om de prijzen en de eer strijden en uit meer dan 90.000 schaatstijden is het 'Universeel Schaatsklassement' berekend voor zowel allrounders als sprinters. Inderdaad is Sven Kramer hard op weg de beste allrounder aller tijden te worden, maar dan zal hij in Vancouver zeker de beide gouden plakken moeten verzilveren.

GERARD SIERKSMA EN BERTUS TALSMA

De tijden veranderen

Het is 23 februari 1980. Eric Heiden staat op het punt een nooit eerder vertoond Grand Slam

te behalen. In een grauwe sneeuwjacht op het openlucht-schuurpapier-ijs van Lake Placid (zeeniveau!) verpulvert Heiden in een tijd van 14 minuten en 28,13 seconden het oude record van



Figuur 1. Ontwikkeling wereldrecord 1500m, mannen en vrouwen

Viktor Ljoskin met maar liefst 6,20 seconden, terwijl Lotskin reed op het olie-ijs van de Medeobaan op 1691 meter boven zeeniveau. Amper twee maanden later rijdt Dmitri Ogloblin op Medeo naar een nieuw wereldrecord in 14.26,71. Maar wie van de twee levert nu eigenlijk de grootste prestatie? Heiden schaatst zijn gouden tijd in een sneeuwstorm, terwijl Ogloblin de wonderbaan van Medeo gebruikt. Niemand kent inmiddels meer Ogloblin en de wonderbaan van Medeo is intussen gesloten.

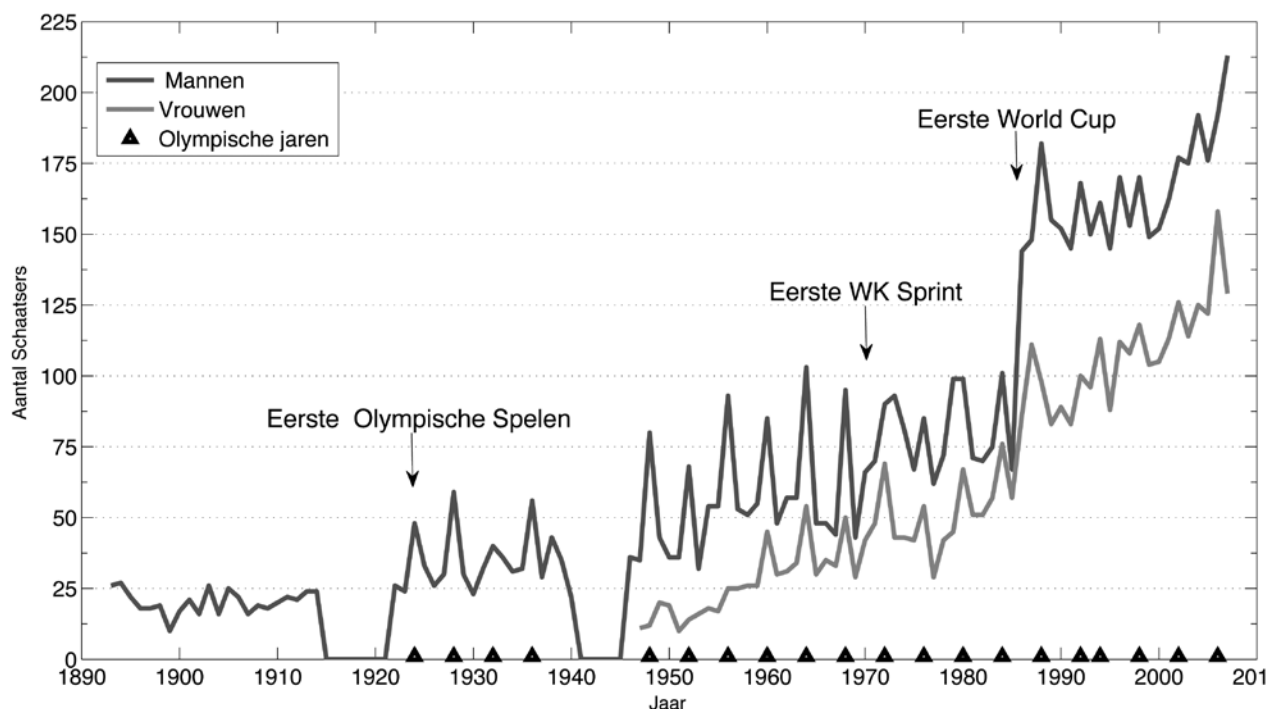
Calgary, 28 februari 1988. De eerste indoor(!) Olympische schaatswedstrijden houden Nederland uit de slaap. Waar niemand op heeft durven hopen is eerder die week toch gebeurd: Yvonne van Gennip heeft de Oostduitse 'meiden' twee keer geklopt. Op de 5000 meter haalt Van Gennip haar derde gouden plak in een direct duel met de dan nog onbekende Gunda Niemann, die na een val huilend over de finish komt. Is Yvonne van Gennip de beste schaatsster aller tijden? In Calgary zal niemand daaraan getwijfeld hebben,

maar drie weken later op het oeroude zonovergoten natuurijsbaantje in het Noorse Skien herstelt Karin Kania de aloude Oostduitse hegemonie en wordt zij wereldkampioen. Fenomenen als Jaap Eden, Oscar Mathisen, Ard Schenk en Eric Heiden kunnen de cijfers een paar jaar in hun greep houden, maar altijd komt het moment dat het doek valt. Niet omdat ze verslagen worden, want dan zijn de helden van aleer allang gestopt.

In het blad *SchaatsSport* (2004) van de KNSB stelt Ottavio Cinquanta, de president van de International Speed Skating Union, de vraag of er een correcte methode is om de toppers van nu en vroeger te vergelijken en te rangschikken.

Van strak pak tot klapschaats

Tallose ontwikkelingen en innovaties hebben ervoor gezorgd dat het alsmaar harder gaat. Wie herinnert zich nog de Zwitserse schaatsveteraan Franz Krienbühl, die bij het Europees kampi-



Figuur 2. Totaal aantal wedstrijdschaatsers en schaatsters

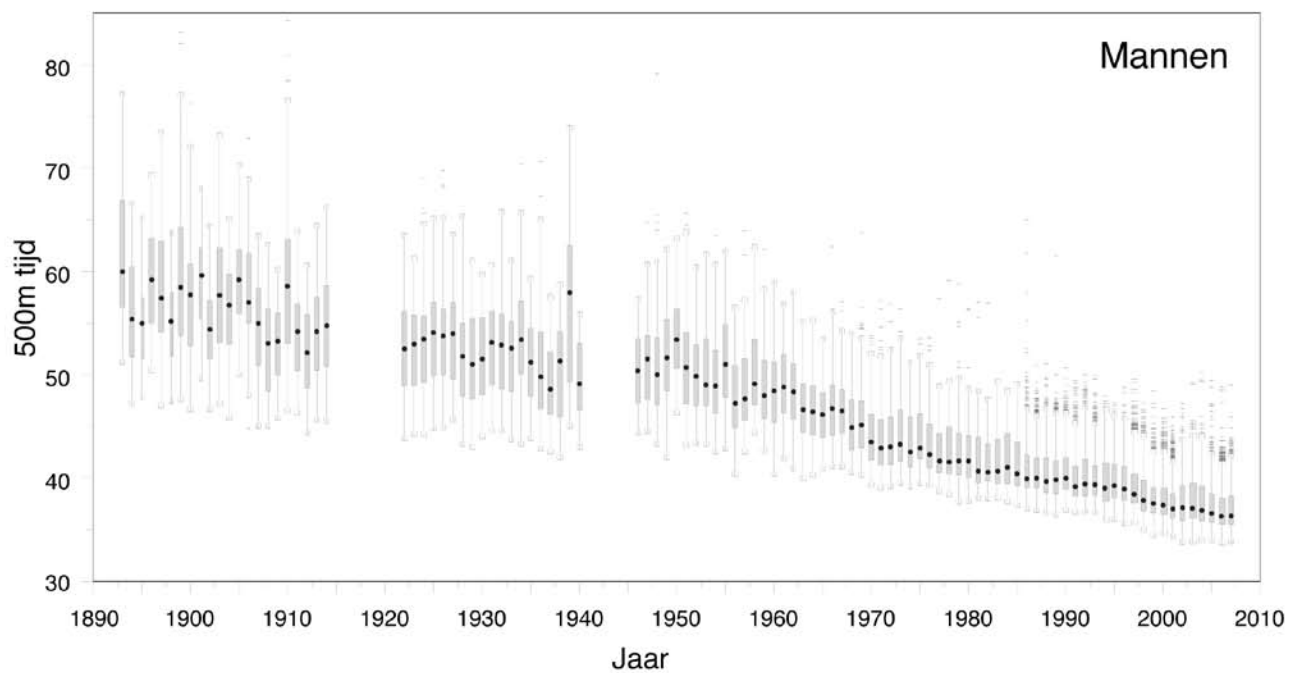
oenschap in 1974 aan de start verschijnt in een naadloos sluitend 'skinpak'? De introductie van kunstijsbanen (1958: Gothenburg), indoorbanen (1987: Thialf) en klapschaatsen (1995: Tonny de Jong) zorgen voor vergelijkbare sprongen voorwaarts.

De gevolgen van al die ontwikkelingen laten zich raden. Het gaat steeds sneller, soms met sprongen tegelijk. In Figuur 1 is de ontwikkeling te zien van de 1500m-wereldrecords en de beste seizoentijden op de grote toernooien vanaf 1893. Hoewel de vrouwen zich met name in de beginperiode (de jaren dertig) als het ware 'op de mannen stortten', lijkt zich nu een constante afstand tussen de beide sexen af te tekenen. Met al die nieuwigheden lijken de prestaties van Jaap Eden, Kees Broekman, Henk van der Grift, Stien Kaiser en Carry Geijzen onvergelijkbaar geworden met die van Marianne Timmer, Gerard van Velde en de andere toppers van de moderne tijd. Tenzij op een andere manier naar de schaatstijden wordt gekeken.

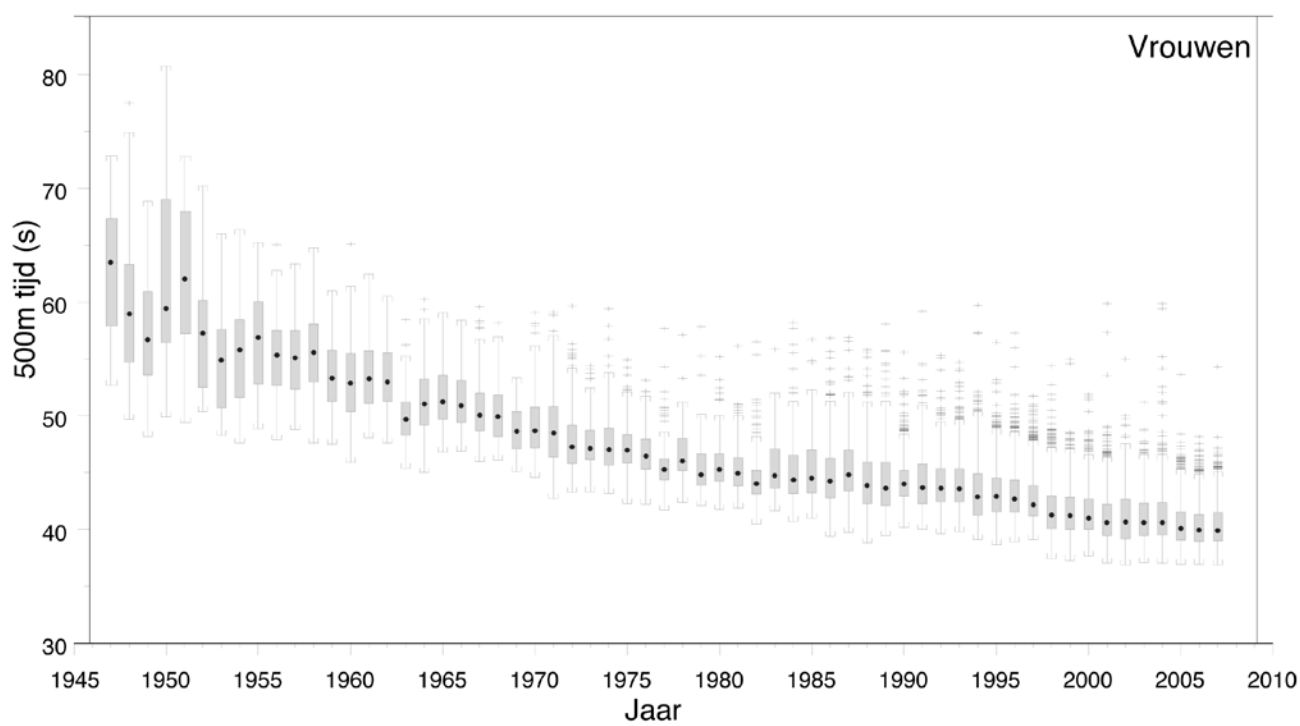
Van absolute tijden naar tijdsverschillen

Sinds 1893 hebben zo'n 2200 mannen en vanaf 1947 zo'n 1100 vrouwen gestreden om de titels en de eer. In Figuur 2 hebben we het verloop van deze aantallen weergegeven. Opvallend is de enorme toename bij zowel de mannen als de vrouwen, waarbij de invoering van de World Cup tot een extra toename bij de mannen heeft geleid. De pieken in beide grafieken worden veroorzaakt door de Olympische Spelen, omdat dan ook de kleine schaatslanden sporters afvaardigen.

Het aantal toernooien voor mannen en vrouwen tezamen ligt op ongeveer 800. Het totale aantal wedstrijden ligt een flink stuk boven de 3000. Dat maakt dat de dataset van Jeroen Heijmans (zie www.skateresults.com) maar liefst zo'n 90000 schaatstijden bevat. In Figuur 3 hebben we alle(!) schaatstijden uit de database van Heijmans in Box Plots weergegeven. Langs de horizontale as staan de jaartallen; bijvoorbeeld 2006 betekent het seizoen 2005/2006. Langs de verticale as staan de, naar 500 meter omgerekende,



Figuur 3(a). Alle schaatstijden omgerekend naar 500 meter-tijden; mannen



Figuur 3(b). Alle schaatstijden omgerekend naar 500 meter-tijden; vrouwen

schaatstijden. Duidelijk is te zien dat de gemiddelde seizoentijden een dalend karakter vertonen en dat de boxen steeds kleiner worden. Met andere woorden de schaatstijden, vooral de hele snelle, komen steeds dichter bij elkaar te liggen.

Voor een eerlijke vergelijking moeten alle schaatstijden worden gecorrigeerd en wel zó dat de voor- en nadelen van de omstandigheden worden geneutraliseerd. Omdat absolute tijden ongeschikt zijn, lijkt het slimmer naar *onderlinge tijdsverschillen* te kijken. Tijdsverschillen in de uitslag van een wedstrijd weerspiegelen immers de onderlinge krachtsverschillen. Regelmatig winnen met grote voorsprong, zoals Gianni Romme deed in de periode 1998-2001, tekent het grote talent en de suprematie van de allergrootsten.

In eerste instantie was het de bedoeling om de verschillen met de nummer één te nemen. Het nadeel hiervan is dat dan alle nummers één gelijkwaardig worden. Immers vanaf de nummer twee krijgt elke schaatser een verschiltijd toegewezen. Het feit dat Romme en Kramer hun 10 kilometers met veel grotere voorsprong wonnen dan bijvoorbeeld Bob de Jong (Olympisch goud in Turijn 2006!) wordt dan niet verdisconteerd. Daarom is gekozen om per gereden afstand de verschillen te nemen met de gemiddelde tijd van de 'eerste vijf'. Het feit dat voor die 'vijf' is gekozen is in zekere zin willekeurig. Een voordeel van deze keuze is dat de nummers één en twee per toernooi zich nu ook kunnen onderscheiden, waarmee recht wordt gedaan aan de suprematie van mensen als Romme en Kramer op de lange afstanden. De aldus gecorrigeerde onderlinge tijdsverschillen worden nog gewogen naar de belangrijkheid van de toernooien. Zo zijn de Olympische Spelen als viermaal zo belangrijk genomen als de wereldkampioenschappen. De andere toernooien zijn mutatis mutandis gewaardeerd op basis van hun specifieke status. Rest nog een laatste probleem. Zouden we het gemiddelde nemen over alle schaatsjaren van een schaatser, dan wordt bijvoorbeeld Rintje Ritsma 'gestraft'

voor het zeer lang rekken van z'n carrière. Vandaar dat gekozen is de schaatsters te rangschikken op basis van hun vier allerbeste jaren. Eis is wel dat tenminste één van die vier jaren een Olympisch jaar is, ook al is dat geen topjaar voor de betreffende schaatser. In het geval een schaatser nooit aan de Spelen heeft meegedaan wordt een 'straf-tijd' toegekend, die lager is naarmate men actiever is geweest op andere toernooien. Schaatsters met minder dan vier actieve jaren worden niet meegenomen in de rangschikking.

Heiden en Niemann: kanjers onder kanjers

In Tabel 3 staat de top 15 van beide allroundklassementen en in Tabel 4 de top 15 sprinters. In deze tabellen betekent een score de gemiddelde voorsprong (-) of achterstand (+) op de gemiddelde tijd gereden tijdens zijn/haar vier beste jaren, omgerend naar 500 meter-tijden. In Tabel 4 bijvoorbeeld, betekent de score -0,255 van Heiden en de +0,063 van Jan Bos dat het verwachte verschil tussen Bos en Heiden tijdens een sprinttoernooi (vier afstanden!) maar liefst 1,272 punten zou zijn in het voordeel van Heiden.

Op 13 januari 2008 wordt Sven Kramer in Kolomna 'met twee vingers in de neus' Europees kampioen. Op de 10 kilometer speelt hij een kat-en-muis spelletje met zijn directe tegenstander de Noor Håvard Bøkko. Pas in de laatste drie ronden rijdt Kramer bij hem weg en wint met 'slechts' drie seconden voorsprong. Computerberekeningen laten zien dat in 2010 het mannen-allroundschaatsen wel eens een nieuwe leider zou kunnen krijgen. Kramer zal dan de komende twee seizoenen ook op de 10 kilometer gas moeten blijven geven en op de Olympische Spelen in 2010 goud moeten pakken op de '5' en de '10'. De kans dat Gunda Niemann van de troon wordt gestoten in het Universeel Klassement is beduidend minder groot. Hoe dan ook, de dagen van Eric Heiden lijken geteld.

ALLROUNDLIJST MANNEN

1	Eric Heiden	USA	-0.272	1979, 1978, 1980, 1976
2	Ard Schenk	NED	-0.231	1972, 1973, 1971, 1967
3	Johann Olav Koss	NOR	-0.193	1994, 1991, 1993, 1990
4	Oscar Mathisen	NOR	-0.160	1912, 1914, 1913, 1908
5	Gianni Romme	NED	-0.138	1998, 2000, 2003, 2002
6	Jaap Eden	NED	-0.134	1896, 1895, 1893, 1894
7	Hjalmar Andersen	NOR	-0.125	1951, 1952, 1950, 1954
8	Clas Thunberg	FIN	-0.102	1925, 1924, 1931, 1929
9	Ivar Ballangrud	NOR	-0.100	1930, 1936, 1926, 1938
10	Rintje Ritsma	NED	-0.047	1995, 1998, 1996, 1993
11	Ids Postma	NED	-0.007	1998, 1996, 1997, 2001
12	Kees Verkerk	NED	0.007	1967, 1969, 1966, 1968
13	Oleg Goncharenko	URS	0.013	1953, 1958, 1954, 1956
14	Sven Kramer	NED	0.020	2008, 2007, 2005, 2006
15	Bernt Evensen	NOR	0.026	1927, 1934, 1928, 1931

ALLROUNDLIJST VROUWEN

1	Gunda Kleemann	GER	-0.391	1995, 1991, 1996, 1998
2	Karin Enke	GDR	-0.332	1986, 1987, 1984, 1980
3	Andrea Mitscherlich	GDR	-0.156	1984, 1985, 1987, 1983
4	Lidia Skoblikova	URS	-0.152	1963, 1964, 1960, 1962
5	Claudia Pechstein	GER	-0.149	2000, 1998, 1994, 2003
6	Anni Friesinger	GER	-0.108	2005, 2008, 2007, 2002
7	Inga Artamonova	URS	-0.076	1965, 1958, 1962, 1957
8	Natalya Petrusyova	URS	-0.062	1982, 1981, 1983, 1980
9	Cindy Klassen	CAN	-0.055	2006, 2003, 2005, 2007
10	Stien Kaiser	NED	-0.037	1967, 1972, 1971, 1965
11	Bonnie Blair	USA	-0.018	1994, 1992, 1988, 1986
12	Valentina Stenina	URS	-0.002	1961, 1965, 1966, 1960
13	Atje Keulen-Deelstra	NED	0.006	1973, 1974, 1972, 1970
14	Sheila Young	USA	0.009	1976, 1973, 1975, 1974
15	Yvonne van Gennip	NED	0.026	1988, 1989, 1985, 1992

Tabel 3. Universeel Klasseмент Top 15 Allround, mannen en vrouwen

SPRINTLIJST MANNEN

1	Eric Heiden	USA	-0.255	1979, 1978, 1980, 1977
2	Igor Zhelezovski	URS	-0.127	1985, 1989, 1986, 1992
3	Uwe-Jens Mey	GER	-0.092	1990, 1991, 1989, 1988
4	Gaetan Boucher	CAN	-0.080	1984, 1985, 1982, 1979
5	Dan Jansen	USA	-0.042	1986, 1994, 1988, 1989
6	Jeremy Wotherspoon	CAN	-0.030	2008, 2003, 2000, 1998
7	Hiroyasu Shimizu	JPN	-0.006	1996, 2000, 1999, 1998
8	Sergey Klevchenya	URS	0.016	1996, 1994, 1997, 1995
9	Frode Rønning	NOR	0.030	1981, 1982, 1979, 1980
10	Valeri Muratov	URS	0.034	1976, 1973, 1972, 1970
11	Peter Mueller	USA	0.061	1976, 1977, 1979, 1974
12	Jan Bos	NED	0.063	1999, 2000, 1998, 2008
13	Akira Kuroiwa	JPN	0.084	1987, 1983, 1986, 1988
14	Sergey Khlebnikov	URS	0.086	1982, 1984, 1981, 1980
15	Erben Wennemars	NED	0.091	2003, 2004, 1998, 2002

SPRINTLIJST VROUWEN

1	Karin Enke	GDR	-0.238	1986, 1987, 1984, 1980
2	Bonnie Blair	USA	-0.214	1994, 1987, 1989, 1990
3	Natalya Petrusyova	URS	-0.154	1982, 1980, 1983, 1981
4	Christa Rothenburger	GDR	-0.143	1989, 1988, 1986, 1984
5	Catriona LeMay	CAN	-0.106	1998, 2002, 2001, 1999
6	Sheila Young	USA	-0.102	1973, 1976, 1975, 1981
7	Monique Garbrecht	GER	-0.069	2001, 2003, 2000, 2002
8	Leah Poulos	USA	-0.024	1976, 1979, 1980, 1977
9	Qiaobo Ye	CHN	0.015	1992, 1993, 1991, 1994
10	Angela Stahnke	GDR	0.034	1990, 1989, 1985, 1994
11	Sabine Volker	GER	0.041	2001, 2002, 1997, 1999
12	Franziska Schenk	GER	0.064	1997, 1994, 1998, 1995
13	Atje Keulen-Deelstra	NED	0.097	1973, 1970, 1972, 1974
14	Sylvia Burka	CAN	0.099	1977, 1979, 1976, 1973
15	Monika Pflug	FRG	0.120	1972, 1973, 1981, 1982

Tabel 4. Universeel Klasseмент Top 15 Sprint, mannen en vrouwen

LITERATUUR

S.M. Berry, S.C. Reese, P.D. Larkey (1999), Bridging Different Eras in Sports, *Journal of the American Statistical Association* 94, pp. 661-687.
D. Heuvelman, F. van Schoonderwalt, G. Sierksma (2007), *Tour de France top 100*, Tirion Sport
R.H. Koning (2005), Home Advantage in Speed Skating: Evidence from Individual Data, *Journal of Sports Sciences* 23, pp. 417-428.
G.H. Kuper, E. Sterken (2002), Endurance in Speed Skating: The Development of World Records, *European Journal of Operational Research* 148, pp. 293-301.

R.T. Stefani (1997), Survey of the Major World Sports Rating Systems, *Journal of Applied Statistics* 24, pp. 635-647.
G. Sierksma, H. Snoep (2008), *Schaatsen top 100*, Tirion Sport.

GERARD SIERKSMA is hoogleraar kwantitatieve logistiek aan de Rijksuniversiteit Groningen.

E-mail: g.sierksma@rug.nl

BERTUS G. TALSMA is Ph.D. student aan de Rijksuniversiteit Groningen.

E-mail: b.g.talsma@rug.nl



TIENKAMP: EEN KWESTIE VAN BALANS

Atletiekcoaches hebben vaak veel tijd nodig voor het maken van de jaarlijkse trainingsschema's voor hun atleten. In dit artikel beschrijven we de resultaten van een case-study over de periodisering van het trainingsjaar voor tienkampers. De studie is uitgevoerd in samenwerking met voormalig KNAU-bondscoach Vince de Lange.

GERARD SIERKSMA EN YORI ZWOLS

De tienkamp, of de decathlon, is een atletiekwedstrijd over tien wedstrijdonderdelen, verdeeld over twee dagen. Hiervan bestaat de eerste dag uit 100 meter hardlopen, verspringen, kogelstoten, hoogspringen en 400 meter hardlopen. Op de tweede dag wordt de wedstrijd voortgezet met 110 meter hordelopen, discuswerpen, polsstokhoogspringen, speerwerpen en 1500 meter hardlopen. Voor elk onderdeel worden punten toegekend en degene met het hoogst totaal aantal punten is de winnaar.

Het evenement vindt zijn oorsprong in de Griekse oudheid. In die tijd is de pentatlon het hoogtepunt van de Olympische Spelen en staat de vijfkampatleet symbool voor het Griekse ideaalbeeld van de gebalanceerde man die op alle vlakken uitblinkt. De pentatlon bestaat uit de onderdelen verspringen, discuswerpen, speerwerpen, hardlopen en worstelen, die in de Griekse tijd naakt worden uitgevoerd. In 1912 wordt de atletiekmeerkamp in de vorm van de tienkamp tijdens de Spelen van Stockholm opnieuw een

Olympische sport. Door een uitbreiding van het aantal onderdelen wordt niet alleen, zoals bij de Grieken, de snelheid en kracht van de atleten gemeten, maar spelen ook techniek en uithoudingsvermogen nu een grote rol.

Schaarse trainingstijd

Omdat tienkampers te maken hebben met uiteenlopende sporten is het plannen van de trainingen een ingewikkelde bezigheid. Hoe verdeel je de beschikbare tijd over de verschillende onderdelen? Aan de ene kant zou een atleet zijn sterke punten hard moeten trainen, omdat daar zijn concurrentievoordeel ligt, maar aan de andere kant weet iedereen dat hoe beter je bent, hoe moeilijker het is nog verder te verbeteren. Daarnaast zijn er meerdere soorten trainingsoefeningen, namelijk technische die vooral de motoriek van de atleet verbeteren en conditionele die de spiermassa vergroten en de conditie verbeteren. Ook dan rijst de vraag: hoe verdeel je de beschikbare tijd over de twee groepen oefeningen?

In [5] geven we een wiskundig model dat ondersteuning biedt bij het beantwoorden van deze vragen. In het huidige artikel houden we ons bezig met een ander aspect van de trainingsplanning. Naast het verdelen van de beperkte trainingstijd over de tien onderdelen, is het namelijk ook van belang die tijd over het jaar te verdelen. Het is bijvoorbeeld niet effectief alle trainingen gelijkmatig over het jaar uit te smeren. Het is beter het jaar in te delen in perioden en binnen die perioden de trainingen te variëren.

De variatie betreft dan met name de *omvang* van de trainingen en de *intensiteit* waarop getraind wordt. De trainingsomvang is een maat voor de duur van de training en het aantal malen dat een oefening herhaald wordt. Een voorbeeld van een trainingsonderdeel met een kleine omvang is vijf maal 60 meter sprinten, terwijl tien maal 60

meter sprinten een grote omvang heeft. Bij intensiteit gaat het om het energieverbruik. Vijf maal 60 meter sprinten kun je bijvoorbeeld op 80% of op 100% afwerken. Vanzelfsprekend zijn de waarden van omvang en intensiteit sterk atleetgebonden.

De kernvraag luidt: hoe kunnen we het trainingsjaar van een tienkamper zodanig periodiseren dat de trainingstijd optimaal benut wordt, en tegelijkertijd de kans op blessures minimaal blijft. Hierbij verstaan we onder een periodisering een plan van aanpak dat bestaat uit een wekelijks voorgeschreven trainingsomvang en -intensiteit. De modellen in dit artikel zijn gebaseerd op die uit [2].

Trainingsdoelstellingen

Voordat we kunnen spreken van een optimale voorbereiding moeten de doelen die atleet en coach nastreven worden gedefinieerd. Hoewel we er van uitgaan dat dit jaarlijkse doelstellingen zijn, kunnen zij ook de langere termijn betreffen, zoals het bereiken van de komende Olympische Spelen.

Een trainingsjaar loopt doorgaans van oktober tot september, waarbij tijdens het eerste deel, van oktober tot ongeveer maart, de trainingen en wedstrijden indoor plaatsvinden. Omstreeks maart begint het buitenseizoen, waarin ook de belangrijke wedstrijden plaatsvinden. Aan het begin van het seizoen stelt de coach een lijst samen met wedstrijden waaraan de atleet gaat deelnemen. Atleten die zich nog moeten kwalificeren voor de Olympische Spelen doen altijd mee aan de wereldkampioenschappen tienkamp. Daarnaast zijn er wedstrijden die bijvoorbeeld vanwege sponsorverplichtingen ingepland zijn. Gedurende het jaar werkt de atleet toe naar de ingeroosterde wedstrijden. De jaarplanning wordt op die wedstrijden afgestemd.

Gezien de doelstellingen van de atleet zijn niet alle wedstrijden even belangrijk en sommige wedstrijden zijn niets anders dan trainingen. In Tabel 1 hebben we het wedstrijdschema van een Olympische atleet en de bijbehorende wegingsfactoren weergegeven. De wegingsfactoren lopen uiteen van een A voor een zeer belangrijke wedstrijd tot een E voor een relatief onbelangrijke wedstrijd.

Periodisering: omvang, intensiteit en blessures

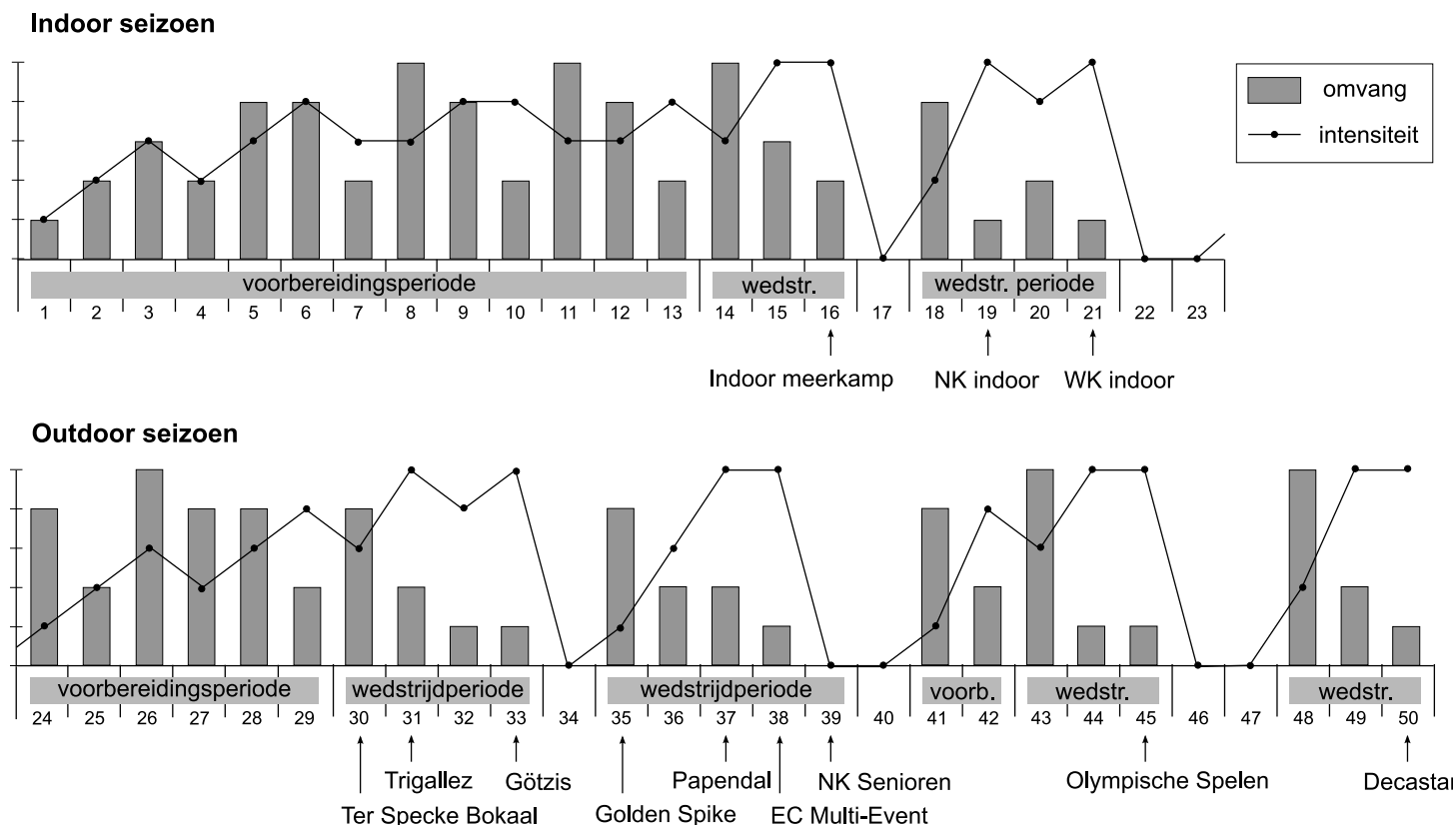
In de standaard trainingsleer (zie [1], [3], [4]) is het gebruikelijk het trainingsjaar in te delen in één of meer 'grote' cycli, die zelf ook weer bestaan uit kleinere cycli. Een grote cyclus begint met een *voorbereidingsperiode*, waarin de atleet conditie opbouwt en de techniek traint. De nadruk tijdens zo'n periode ligt op de

omvang en niet zo zeer op de intensiteit. Pas aan het einde van de voorbereidingsperiode wordt de intensiteit opgeschroefd. In de laatste week van de voorbereidingsperiode, de zogenaamde *voorwedstrijdweek*, wordt de intensiteit tijdelijk verminderd en wordt er op maximale omvang getraind.

De voorwedstrijdweek luidt de *wedstrijdperiode* in, waarin er gepresteerd moet worden. De atleet traint tijdens deze periode met een hoge intensiteit. Hoeveel er precies getraind wordt hangt af van het belang van de wedstrijden, waarbij, zoals gezegd, minder belangrijke wedstrijden als training kunnen fungeren. Aan het eind van de wedstrijdperiode wordt er doorgaans een *overgangsperiode* ingelast van één of twee weken, waarin rust en herstel centraal staan. Deze periode komt bij voorkeur meteen na een zware wedstrijd. Als er twee zware wedstrijden dicht na elkaar gepland staan valt de overgangsperiode na de tweede wedstrijd.

WEEK	DATUM	OMSCHRIJVING	WEGINGSFACTOR
INDOOR			
16	31 jan	Indoor meerkamp interland, Zuidbroek	B
19	21 feb	NK indoor senioren/junioren B, Gent	C
21	7 mrt	WK indoor, Boedapest	B
OUTDOOR			
30	8 mei	Ter Specke Bokaal, Lisse	E
31	15 mei	Trigallez Recordwedstrijden, Hoorn	E
33	29 mei	Hypo-Meeting Götzis	A
35	12 jun	Golden Spike, Leiden	E
37	23 jun	Papendal Games	E
38	3 jul	EC Multi-Event Super League/1st League, Hengelo	B
39	10 jul	NK senioren, Utrecht	C
45	29 aug	Olympische Spelen, Athene	A
50	18 sep	Decastar, Talence	B

Tabel 1: Wedstrijdschema met wegingsfactoren. (De weken worden geteld vanaf het begin van het seizoen, d.w.z. week 1 is de eerste van het trainingsseizoen).



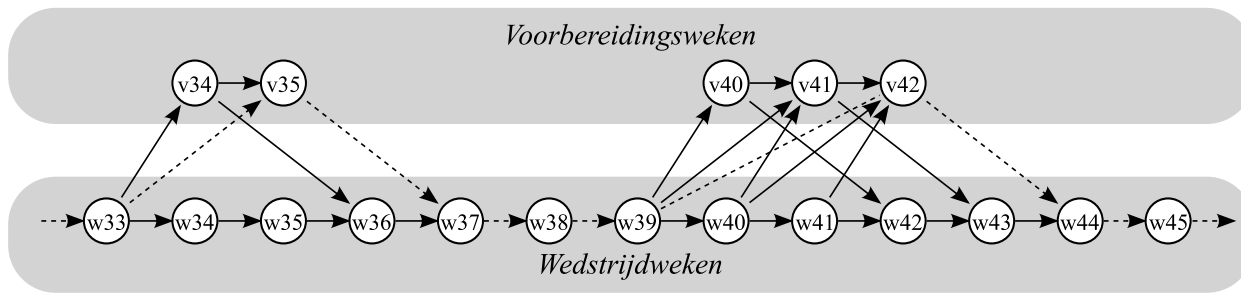
Figuur 1: Een mogelijke periodisering

Om optimaal te presteren worden de grote cycli opgedeeld in kleinere cycli van twee tot drie weken. Elke kleine cyclus begint met een grote training en in de twee tot drie weken die volgen wordt de omvang langzaam afgebouwd. In verband met wedstrijden is het echter niet altijd mogelijk precies aan deze regel te voldoen. Een andere uitzondering op de regel is het begin van het jaar, waarin omvang en intensiteit geleidelijk opgebouwd moeten worden. Om het risico op blessures zo klein mogelijk te houden dient met een aantal specifieke beperkingen rekening te worden gehouden. Het is bijvoorbeeld niet verstandig om in één week zowel de omvang als de intensiteit op een hoog niveau in te plannen. Daarnaast is het evenmin verstandig om de intensiteit van de training snel te laten toenemen.

Bovendien kan drie weken achtereenvolgend trainen op een hoge intensiteit tot overtraining leiden met alle nare gevolgen van dien.

Modelleren en oplossen

Het periodiseringsmodel bestaat uit twee fasen. De eerste fase verdeelt het jaar in voorbereidingsperioden en wedstrijdperioden. De tweede fase neemt de indeling uit de eerste fase over en voegt hier de trainingsomvang en -intensiteit aan toe. Het trainingsjaar is in 50 weken verdeeld. Voor elk van deze weken stellen we de omvang en de intensiteit vast, beide uitgedrukt op een schaal van 1 tot 5. Om alvast een idee te geven hoe zo'n periodisering eruit ziet verwijzen we naar Figuur 1.



Figuur 2: Periodisering als kortste-pad probleem. Deze figuur beeldt een deel (weken 30 t/m 45) van de gerichte graaf af die gebruikt wordt in de eerste fase.

Eerste Fase

We gebruiken de voorbereidingsperiode als uitgangssituatie. Omdat er wedstrijden op de planning staan, moet de atleet af en toe vanuit die uitgangssituatie door een *wedstrijdblok*. Zo'n blok bestaat uit een voorwedstrijdweek, een wedstrijdperiode en een overgangperiode. We stellen de 'harde' eis dat elke wedstrijd in een wedstrijd-blok valt. Daarnaast kennen we 'zachtere' eisen, zoals de mogelijkheid dat een overgangperiode korter is dan twee weken. Overschrijding van de twee-wekeneis levert in het model 'strafpunten' op. Omdat de voorbereidingsperiode het grootste trainingsrendement oplevert, proberen we zoveel mogelijk weken als voorbereidingsweken te gebruiken. We willen daarom de wedstrijd-blokken zodanig over het jaar verdelen dat het aantal voorbereidingsweken zo groot mogelijk is, maar tegelijkertijd het aantal strafpunten minimaal is. Dit probleem is als een kortste-pad probleem gemodelleerd. In Figuur 2 is een deel van de onderliggende graaf weergegeven corresponderend met het wedstrijdschema van Tabel 1. De gestippelde pijlen vormen een (suboptimaal) pad. Hierin is week 33 een wedstrijdweek. De pijl $w_{33} \rightarrow v_{35}$ betekent dat 34 overgangswEEK is en 35 voorbereidingsweek. De pijl $v_{35} \rightarrow w_{37}$ betekent dat 36 voorwedstrijdweek is en 37 wedstrijdweek. Merk op dat de pijl $w_{33} \rightarrow v_{35}$ strafpunten oplevert, maar de pijl $w_{39} \rightarrow v_{42}$ niet. Het vinden van een kortste pad van het meest linkse punt naar

het meeste rechtse is nu equivalent met het vinden van een periodisering waarin zowel het aantal wedstrijdweken als het aantal strafpunten zo klein mogelijk is. Na het oplossen van het model kan de coach knelpunten in de periodisering vaststellen door te kijken naar de pijlen waarvoor strafpunten gerekend worden. Indien gewenst kan dan het wedstrijdschema aangepast worden om zo tot een betere periodisering te komen.

Tweede Fase

Op basis van de periodisering van fase 1 maken we een planning voor de trainingsomvang en -intensiteit. Omdat de nadruk tijdens een voorbereidingsperiode op de trainingsomvang ligt, proberen we deze tijdens de voorbereidingsperiode te maximaliseren, terwijl we tijdens wedstrijdperioden de intensiteit maximaliseren. Tegelijkertijd proberen we zoveel mogelijk de kleine cycli van afnemende omvang vast te houden en de blessurerisico's te beperken. Dit model wordt beschreven als een geheeltallig programmeringsprobleem. Hoewel geheeltallig programmeren behoort tot de klasse van moeilijke problemen (NP-hard), is ons model vanwege het relatief kleine aantal variabelen in de praktijk eenvoudig op te lossen met gebruikelijke computersoftware.

Resultaten

Zoals gezegd is in Figuur 1 de periodisering voor de Olympische atleet weergegeven. Hierin is in

de eerste dertien weken van het seizoen goed te zien dat de omvang in cycli van drie weken wordt gevarieerd. Waar dit mogelijk is wordt deze trend vastgehouden, maar in de weken 15 t/m 21 en in de weken 41 t/m 45 is gekozen voor een combinatie met kortere cycli van twee weken. In de lange voorbereidingsperioden wordt de intensiteit steeds geleidelijk opgevoerd. Ook is goed te zien hoe de intensiteit vlak voor belangrijke wedstrijden vanaf een laag niveau opgebouwd wordt tot het hoogste niveau, zodat de atleet piekt op het juiste moment.

Scenarioanalyses

Het hierboven beschreven model sluit aan bij ons eerdere model in [5], dat zich bezighoudt met de vraag hoe de beperkte beschikbare tijd van de atleet zo efficiënt mogelijk gebruikt kan worden voor de training van de uiteenlopende trainingsonderdelen. In [5] is de periodisering en de beschikbare trainingstijd als gegeven verondersteld. Omdat de rekentijden van de hierbeschreven (wiskundige) modellen zeer klein zijn, is het mogelijk om snel verschillende alternatieven door te rekenen. In Figuur 1 is bijvoorbeeld te zien dat zowel het Europees Kampioenschap in week 38 op de planning staat als het Nederlands Kampioenschap voor senioren in week 39. We hebben de planning doorgerekend zowel met als zonder het NK. Op basis van de resultaten en rekening houdend met de Olympische Spelen van zes weken later, heeft de coach besloten de atleet niet aan de Nederlandse Kampioenschappen mee te laten doen.

Conclusies

Het model biedt de tienkampcoach een gestructureerde methode om het trainingsjaar van zijn

pupillen te periodiseren. Periodiseren betekent altijd het balanceren van meerdere vuistregels. Aan de ene kant moet de atleet op het juiste moment pieken, aan de andere kant loert het gevaar van overtraining met als gevolg daarvan het ontstaan van blessures. In combinatie met het model beschreven in [5] biedt ons systeem de coach de mogelijkheid snel alternatieve trainingsschema's door te rekenen en het ondersteunt hem daardoor dus in het evalueren van keuzes aangaande de periodisering. Daarnaast wordt de coach, zoals in veel beslissingsondersteunende systemen, gedwongen zijn kennis en intuïtie zodanig te structureren dat deze geschikt is als input voor de computermodellen. Dit heeft als neveneffect dat de coach goed moet nadenken over zijn keuzes. Op de vraag aan coach De Lange of de hier gepresenteerde modellen hem helpen bij het plannen van zijn trainingen antwoordde hij: 'Het maken van trainingsplanningen kostte altijd veel tijd. Nu heb ik meer tijd beschikbaar voor de atleten'.

Dankwoord

Wij danken Vince de Lange voor zijn expertise en het leveren van de tienkampdata.

LITERATUUR

1. U. Jonath, E. Haag, en R. Krempel. *Atletiek: training, techniek, tactiek*. Elmar B.V, Rijswijk, 1977.
2. W.B. Lang. *Computer Support for Decathlon Training*. Afstudeerscriptie, Rijksuniversiteit Groningen, 2004.
3. L.P. Matwejew. *Die Periodisierung des sportlichen Trainings*. Leistungssport 2, pages 401–409, 1972.
4. L.P. Matwejew. *Grundlagen des sportlichen Trainings*. Sportverlag, Berlin, 1981.
5. Y. Zwols en G. Sierksma. *Training Optimization for the Decathlon*. Operations Research, 2009.

GERARD SIERKSMA is hoogleraar kwantitatieve logistiek aan de Rijksuniversiteit Groningen.

E-mail: g.sierksma@rug.nl

YORI ZWOLS is een Ph.D. student aan het Department of Industrial Engineering and Operations Research van Columbia University. E-mail: yz2198@columbia.edu



SPORT EN TOEVAL

ONNO BOXMA

De opdracht leek zo mooi voor iemand die van beide disciplines houdt: maak een column in een special issue over sport en stochastiek. Toch werd het flink laptobben, want sport laat zich juist niet vangen in getallen of kansen.

Wat sport zo mooi maakt is voor mij: de totaal onverwachte wending vlak voor het einde, de psychische druk die vaak tot zulke verrassingen leidt, de emotie, en de tactische meesterzet waarmee de tegenstander wordt verrast. Niet voor niets is *Winning ugly* van de tenniscoach Brad Gilbert één van mijn favoriete sportboeken. Hij beschrijft er onder meer in hoe hij in zijn actieve periode een keer van John McEnroe – toen nagenoeg onverslaanbaar – won door hem tactisch te ontregelen en psychisch te breken (door bijvoorbeeld de service een meter naast de baan staande te gaan ontvangen). McEnroe was zo ontdaan over zijn nederlaag tegen een qua slagen zo veel minder begaafde tegenstan-

der, dat hij enige maanden niet gespeeld heeft. Gilbert wist niets van mathematische statistiek en besliskunde, maar ontwierp wel altijd een op percentages gebaseerd speelplan, waarbij alles in het teken stond van het optimaliseren van de winstkans.

Wat maakt sport leuk als kijkspel?

Wat iemand aantrekt in sport is heel persoonlijk. Laat ik het hier niet hebben over het bedrijven van sport (dat is waarschijnlijk ook niet verstandig voor iemand die een bal met moeite twee keer kan hooghouden, althans als het volledig windstil is), maar me in deze column beperken tot het kijken naar sport. Ik zal proberen de vraag te beantwoorden wat voetbal en tennis tot goede kijksporten maakt, en welke rol stochastiek en toeval daarin spelen.

Waarom is voetbal zo'n goede kijk- en praatsport?

De schoonheid van voetbal speelt maar een ondergeschikte rol. Voetbal heeft echter de perfecte combinatie van toeval en de wet van de grote aantallen: 'uiteindelijk wordt de sterkste meestal kampioen'.

Je maakt als fan van PSV ieder jaar een grote kans dat je ploeg een prijs wint, maar het is bijna elke wedstrijd toch weer spannend. Ter illustratie heb ik uitgerekend hoeveel wedstrijden uit de eredivisie op het moment van schrijven (32 ronden) eindigden in een gelijkspel (65 van de 288), bij hoeveel er een doelpunt verschil was (101), twee doelpunten verschil (59) en meer dan twee (63). In 4 op de 7 wedstrijden is het verschil dus 0 of 1 goal.

Een sterk punt bij voetbal is ook dat er zo weinig hoogtepunten zijn; gemiddeld wordt iets meer dan drie keer per wedstrijd gescoord.

Een andere geweldige vondst van het voetbalspel is buitenspel. Dit is natuurlijk een volslagen belachelijke regel, die enorm heeft bijgedragen aan de populariteit van het spel. Neem alleen al zo'n vage toevoeging als 'hinderlijk buitenspel' die volop ruimte voor meerdere interpretaties laat, en bedenk dat het fysiek onmogelijk is dat scheids- en grensrechters zonder hulpmiddelen feilloos constateren of het net wel of net niet buitenspel is. Een relatief groot aantal doelpunten valt uit zulke al-dan-niet-buitenspel situaties. De rest komt trouwens grotendeels uit penalties en vrije trappen, over het toekennen waarvan meestal ook veel te doen is, of via van richting veranderde ballen. Kortom, volop ruimte voor toeval, en veel gespreksstof na afloop. En laat ik het dan maar niet eens hebben over de mogelijkheid dat één der grensrechters of doelmannen een kleine tegemoetkoming wordt aangeboden in ruil voor het beïnvloeden van bovengenoemde kansen. Succes lijkt dan nagenoeg verzekerd; maar goed dat het allemaal zulke eerlijke mensen zijn ...

Helaas onderkent de sportpers nauwelijks dat het toeval een grote rol speelt in de uitkomst van een individuele wedstrijd. In wedstrijdverslagen lijkt de sportjournalist vaak te denken dat het zijn of haar taak is, een uitslag te verklaren. Als de geluksfactor heel erg nadrukkelijk aanwezig was, volgt een dooddooener als 'PSV dwong het geluk af'. Overigens had Van Gaal vroeger bij Ajax een bordje in zijn kantoor met de tekst 'Kwaliteit is het uitsluiten van toeval'. Het uitsluiten van toeval bij voetbal lijkt mij echter een illusie.

Een simpel model voor de kansen in voetbal

Om voetbalploegen en -competities beter te analyseren, zou mij als leek het volgende simpele model als startmodel redelijk lijken (maar ongetwijfeld hebben de experts hier onderzoek naar gedaan, en veel verfijnder modellen ontwikkeld). Kies bij elke wedstrijd voor beide clubs een binomiale verdeling $\text{bin}(n,p)$, met n het aantal te creëren kansen en p de kans dat een kans in een doelpunt wordt omgezet. Stel dat bij een wedstrijd tussen clubs 1 en 2 geldt: club 1 heeft $n_1 = 10$, $p_1 = 0.2$, en club 2 heeft $n_2 = 4$, $p_2 = 0.4$. Club 1 heeft dus het sterkere veldspel, en het creëert meer kansen; maar club 2 heeft de effectievere doelpuntenmakers, en misschien ook de betere doelman. Nu is de kans op een gelijkspel ongeveer een kwart, en de kans op één doelpunt verschil is ongeveer 0.4. 0-1 winst voor de (waarschijnlijk Italiaanse) club is nu iets waarschijnlijker dan 1-0.

De tennistelling

Ook tennis is een heel populair kijkspel, niet op de laatste plaats dankzij de bizarre telling (15-0, 30-0, 40-0, game; en in een tweede set begin je helemaal opnieuw). Een -vanuit het oogpunt van spanning- zwak punt van de tennistelling is, dat

zij niveauvergroterend werkt. Als een speler telkens een kans 0.6 heeft om een rally te winnen, dan is de kans op gamewinst ongeveer 0.78; dit volgt uit absorptiekansen voor random walks met absorberende grenzen. Dat de sterkere speler vervolgens toch niet altijd de partij wint maakt tennis als kijksport aantrekkelijker, en vindt zijn oorzaak opnieuw in lastig kwantificeerbare factoren als slimme aanpassingen in de tactiek en mentale kracht op de momenten dat het er echt op aan komt.

Een heel sterk punt van de tennistelling en van tellingen bij sommige andere niet aan tijd gebonden sporten is dat een speler op matchpoint kan komen en vervolgens twee uur later alsnog verliest: door de spanning daalde (simplistisch gesteld) zijn kans per punt van 0.6 naar 0.45, en dat was voldoende om de partij te laten kantelen.

Mijn stelling is dan ook dat, anders dan bij voetbal, toeval een vrij kleine rol speelt bij een tenniswedstrijd - maar kansen, die de spelers zelf kunnen beïnvloeden, een des te grotere.

Tot slot een waargebeurde anecdote over sport en toeval.

Een Utrechts statisticus merkte dat ik, jonge Utrechtse promovendus in de jaren zeventig, een tennisfanaat was en gaf me enige tweedehands tennisboeken uit de jaren twintig. Een daarvan was *The quest for the Davis Cup* uit 1928. Mijn hospita zag dat boek liggen en viel bijna van haar stoel van verbazing: voor in het boek stond de naam van de oorspronkelijke eigenaar, en deze was ook de eerste bewoner van haar huis geweest.

Na bijna 50 jaar was het boek weer thuis gekomen ...

ONNO BOXMA is hoogleraar Stochastische Besliskunde bij de Faculteit Wiskunde en Informatica van de Technische Universiteit Eindhoven en wetenschappelijk directeur van EURANDOM. E-mail: boxma@win.tue.nl.



FRED STEUTEL BENOEMD TOT ERELID VAN DE VVS

Tijdens de dag voor Statistiek en Besliskunde, 27 maart 2008, is Fred Steutel benoemd tot erelid van de VVS. De redactie feliciteert haar mederedactielid graag met deze openlijke waardering voor het zeer vele en belangrijke werk dat Fred voor onze vereniging heeft gedaan. De redactie.

Haastige spoed ...

We kennen allemaal het gevaar van haastwerk. Maar soms ontkomen we er niet aan, en prompt slaat dan het noodlot toe. Zo ook bij het eerste nummer van Sator voor dit jaar. Om verzendkosten te besparen, door gelijktijdig het programma van de Dag voor Statistiek en Besliskunde mee te zenden, was er een spijkerharde deadline voor het drukken van dat nummer. Voeg daarbij het feit dat enkele auteurs op voorjaarsvakantie waren en dat de eindredacteur op familiebezoek in het buitenland was en daar geen goede computerfaciliteiten had. Samen genoeg om enkele storende fouten in het artikel van Don en Van den Eijnden te veroorzaken. In de tekst en de formules is enkele malen een α verwisseld met een λ en omgekeerd. We bieden de beide auteurs onze verontschuldiging aan voor deze fout. Direct na ontdekking is een correcte versie op de website geplaatst, ook is het via de elektronische VVS nieuwsbrief aan de leden meegedeeld. De redactie.

IN MEMORIAM THEO RUNNENBURG (1932-2008)

Johannes Theodorus (Theo) Runnenburg werd op 19 februari 1932 geboren in Amsterdam, waar zijn ouders op de Overtoom een wasserij hadden. Hij haalde zijn HBS-diploma aan de 'Tweede Vijfjarige' aan het Roelof Hartplein. Daarna studeerde hij wis- en natuurkunde aan de Universiteit van Amsterdam. Zijn leermeesters in de wiskunde waren, naast Heijting en De Groot, vooral Van Dantzig en De Bruijn. Bij deze laatste haalde hij (informatie van De Bruijn) als enige student ooit, een tien voor het vak Maattheorie. Na zijn afstuderen in 1954 ging hij onder Van Dantzig werken aan het Mathematisch Centrum, nu het Centrum voor Wiskunde en Informatica. De eerste jaren werkte hij nauw samen met Harry Kesten. Zij schreven samen twee baanbrekende artikelen onder de titel *Priority in waiting line problems I en II*. Harry emigreerde kort daarna naar de VS.

Ik leerde Theo in september 1956 op het Mathematisch Centrum kennen; ik had net mijn kandidaatsexamen gedaan. Mijn eerste artikeltje schreef ik met Theo's hulp, over prioriteiten in een wachtrijmodel. Wij vertaalden samen één van de eerste boeken over wachttijdtheorie: *Problèmes stochastiques posés par le phénomène de formation d'une queue d'attente à un guichet*, van F. Pollaczek, in het Nederlands.

Hij was tussen allerlei andere werkzaamheden door – met name veel werk voor het rapport van de Deltacommissie in verband met de waters-

noodramp in 1953 – aan een proefschrift bezig, met Van Dantzig als promotor. Toen in 1959 het werk, *On the Use of Markov Processes in One-server Waiting-time Problems and Renewal Theory*, praktisch klaar was, overleed Van Dantzig. De promotie is toen in 1960 afgehandeld door N.G. de Bruijn. Kort daarna werd Runnenburg benoemd tot lector en weinig later tot hoogleraar kansrekening. Hij gaf uiterst consciëntieus onderwijs, hij werkte zich letterlijk in het zweet en moest zich na een college verkleden.

Hij was een heel scherpzinnige wiskundige. Hij heeft aan heel diverse onderwerpen gewerkt, maar toch betrekkelijk weinig gepubliceerd. Zijn belangrijkste verdienste voor de wetenschap ligt in het begeleiden van een aantal getalenteerde promovendi. Elf mensen, onder wie ikzelf, zijn bij hem gepromoveerd; vijf daarvan zijn hoogleraar geworden.

Zijn laatste jaren werden gekenmerkt door eenzaamheid. Zijn vrouw was opgenomen in een verzorgingshuis en hij had weinig contact met zijn kinderen. Hij had weinig blijvende vrienden gemaakt, ook onder zijn promovendi. Hij bleef wel tot het laatst in veel dingen geïnteresseerd, binnen en buiten de wiskundewereld. Op 17 april is hij overleden, drie dagen na zijn vrouw. Ik heb hem de laatste weken van zijn leven verschillende malen bezocht. Ik zal met waardering aan hem denken.

Fred Steutel

AGENDA

27 - 28 augustus 2008

Oprichters van EURANDOM, voormalige postdocs en vooraanstaande sprekers vanuit industrie en wetenschap blikken terug op 10 jaar EURANDOM en kijken vooruit naar nieuwe ontwikkelingen. Voor inlichtingen, zie <www.eurandom.nl>.

12-14 november 2008

International Workshop 'Flexible Modelling: Smoothing and Robustness' (FMSR 2008). Algemeen thema: semi-en niet-parametrische analyses en robuuste statistische methodes. Er zijn invited talks, contributed talks en poster sessions. De workshop wordt gevolgd door een intensieve cursus voor doctoraatsstudenten. Zie ook: <http://wis.kuleuven.be/stat/fmsr2008.php>.

Dagelijks internationaal bezig zijn bij een oer-Hollands bedrijf?



TNT is een internationaal opererende onderneming met een actief wereldwijd netwerk. Zakelijk en maatschappelijk zoeken wij continu naar nieuwe kansen op zowel de Nederlandse als de internationale markt. Daarom bieden wij een afwisselende loopbaan aan medewerkers die grenzen verleggen, kansen grijpen en hun visie delen, mét de mogelijkheid om door te groeien.

Wij zoeken een gedreven adviseur met veel affiniteit voor het ontwikkelen van statistische modellen.

Statistisch Consultant

De functie

Als consultant ondersteun je het management vanuit je statistische of econometrische achtergrond. Je verdiept je in de problemen en wensen van je klanten, je adviseert op maat over mogelijke oplossingen en ondersteunt hen door het maken van analyses en het ontwikkelen van voorspelmodellen en steekproefsystemen. Zo kom je in aanraking met vragen van operationeel tot strategisch niveau in Nederland en in de omliggende landen. De afdeling Kwantitatieve ondersteuning in Den Haag speelt een essentiële adviserende rol bij het handhaven en uitbouwen van de strategische positie van TNT Post. We gebruiken onze expertise bij het modelleren, simuleren, optimaliseren en monitoren van netwerken. Daarnaast adviseren wij managers op basis van onze kwantitatieve expertise op logistiek, commercieel of financieel/ economisch gebied.

Gewenst profiel

- WO econometrie (richting econometrie) of wiskunde (statistiek);
- enige jaren ervaring als adviseur;
- goede contactuele vaardigheden en een servicegerichte instelling;
- doorgroeipotentieel.

Tevens zijn wij op zoek naar een pas afgestudeerd econometrist of wiskundige (masters in operationele research).

Wij bieden

- maximaal € 55.000,-;
- een intellectueel uitdagende functie met veel diversiteit;
- een informele werksfeer op een intern adviesbureau;
- een organisatie die op effectieve manier investeert in medewerkers;
- aanvullende voorwaarden volgens de CAO van TNT zoals premie-vrijpensioen en collectieve regelingen;
- de mogelijkheid om parttime te werken.

Interesse?

Surf naar www.werkenbijtnt.nl en kijk bij het werkveld logistiek. Voor meer informatie kun je bellen met Henk van der Brug, Manager Kwantitatieve Ondersteuning op telefoonnummer 070-3347371.



BINNEN PEARSON IS DE VACATURE ONTSTAAN VAN

Pearson (voorheen Harcourt Test Publishers) maakt deel uit van een internationaal netwerk van testuitgeverijen. Als grootste in haar branche voor Nederland en Vlaanderen staat Pearson voor 60 jaar ervaring in het uitgeven en distribueren van wetenschappelijk onderbouwde psychologische meetinstrumenten. Van oudsher zijn onze producten gericht op Gezondheidszorg, Onderwijs en Human Resource Management. Momenteel omvat ons aanbod ruim 300 verschillende tests over onder meer persoonlijkheid, vaardigheden en interesses.

Testontwikkelaar (m/v, 1.0 fte)

Functie-inhoud

Als testontwikkelaar draag je bij in de ontwikkeling van tests voor gebruik in o.a. de klinische, neuropsychologische, logopedische of HRM markt. Je adviseert collega's of werkt zelf intensief mee in revisie-, normerings- en testontwikkelingsprojecten. Je verzamelt en analyseert testgegevens (o.a. validerings- en betrouwbaarheidsgegevens) en je bouwt contacten t.b.v. testontwikkeling in het externe netwerk uit.

Kerngebieden

1. Psychometrische / statistische ondersteuning en uitvoering
2. Kwaliteitscontrole van instrumenten op het gebied van testconstructie, normering en psychometrische achtergrond
3. Contacten en overleg binnen expertisegebied
4. Ontwikkelen van vakkennis op het gebied van de testontwikkeling en psychometrie
5. Schrijven van artikelen, presenteren van informatie over de ontwikkeling van onze instrumenten
6. Bijdrage leveren aan het implementeren van interne en internationale kwaliteitsprocedures voor testontwikkeling

Wie zoekt Pearson?

- Een ervaren psychometrist / statisticus met sterke affiniteit in testontwikkeling.
- Iemand met de volgende eigenschappen: doelgerichtheid, goede communicatieve eigenschappen, stressbestendigheid en gewend te werken met deadlines.
- Iemand met goede beheersing van Engels en kennis van en affiniteit met (psychologische, logopedische, HRM, etc.) tests en testontwikkeling.
- Iemand die informatie over testontwikkeling en psychometrie goed kan overbrengen.
- Iemand met ruime ervaring in de psychometrie, de statistiek en de ontwikkeling van psychologische instrumenten.

Informatie/procedure

Voor meer informatie over deze functie kunt u contact opnemen met Barbara Schoonhoven, Project Coördinator (op maandag, woensdag en donderdag), e-mail: barbara.schoonhoven@pearson.com of Andress Kooij, Manager Product Development, e-mail: andress.kooij@pearson.com

Schriftelijke sollicitaties kunt u richten aan:
Pearson
T.a.v. Mw. B. Schoonhoven / Mw. T. Van Meurs
Radarweg 60-A1
1043 NT Amsterdam
barbara.schoonhoven@pearson.com

Dé partner in professioneel testgebruik

PEARSON