Marie-Colette van Lieshout
Centrum Wiskunde & Informatica, Amsterdam, en
Faculteit EWI, Universiteit Twente
m.n.m.vanlieshout@utwente.nl

Marie-Colette van Lieshout

Oratie

Kansen in ruimte en tijd

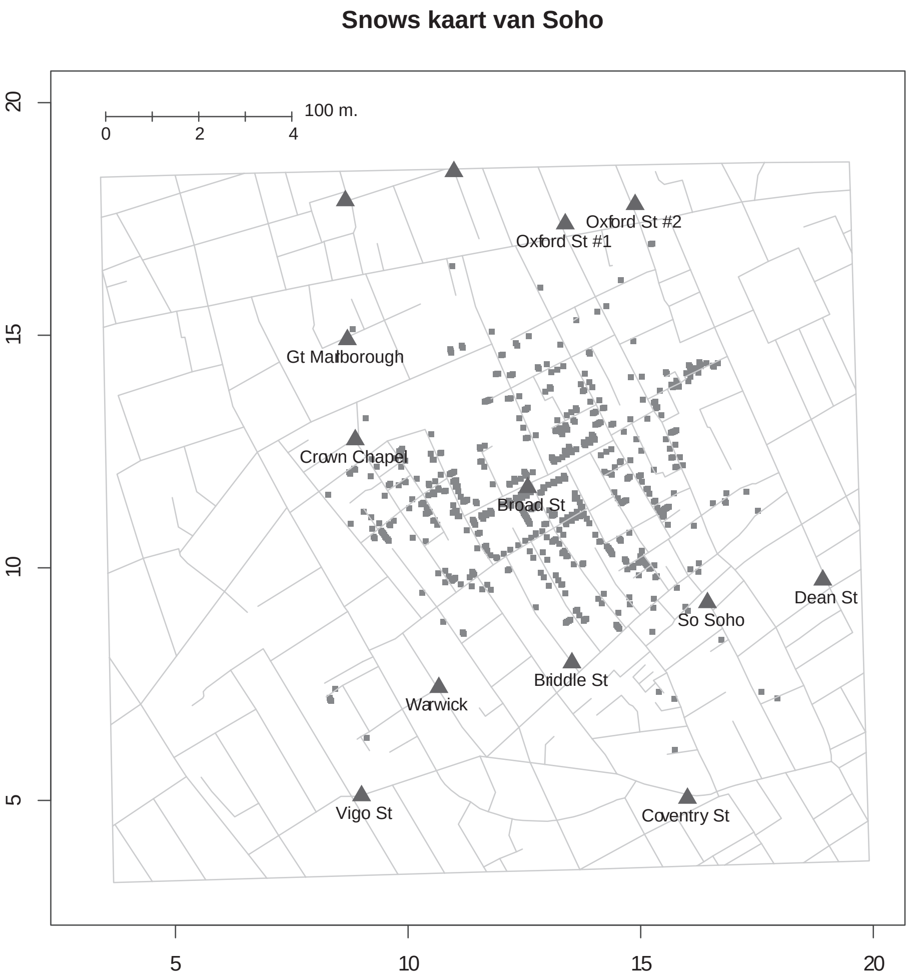
In de ruimtelijke stochastiek wordt de invloed onderzocht van verklarende variabelen (ook wel 'covariaten' genoemd) op een ruimtelijk patroon. Denk bijvoorbeeld, bij de gaswinning in Groningen, aan het verband tussen een waargenomen patroon (bevingen) en covariaten (gaswinning), en aan de structuur van het ruimtelijk patroon (naschokken) en patronen in de tijd (zwaardere bevingen). Sinds 1 februari 2015 is Marie-Colette van Lieshout, onderzoeker aan het Centrum Wiskunde & Informatica, als deeltijdhoogleraar verbonden aan de vakgroep Stochastische Operations Research van de Universiteit Twente. Op donderdag 3 december 2015 hield zij haar intreerede.

Wiskundigen noemen mijn vakgebied de ruimtelijke statistiek of stochastiek. Dit laatste woord heeft Griekse wortels die zoiets als 'gissen' betekenen. Ruimtelijke stochastiek is dus de studie van fenomenen in ruimte en tijd waar onzekerheid een rol speelt. Om u een indruk te geven, stel ik u graag voor aan de Britse arts en anesthesioloog John Snow (1813–1858). Tijdens de cholera-uitbraak van 1854 in de Londense wijk Soho keek dokter Snow, die sceptisch stond tegenover de in zijn tijd gangbare opvatting dat cholera werd veroorzaakt door 'slechte lucht', naar de adressen van 578 slachtoffers [26]. Hij maakte hier — eerst in zijn hoofd, later ook op papier — een plaatje van door de plaats van overlijden middels een vierkantje op een plattegrond van de wijk aan te geven. Zie Figuur 1.

Zodoende valt op dat er in de buurt van de waterpomp in Broad Street veel meer sterfgevallen zijn dan elders. In statistische termen: de loopafstand tot de put is een verklarende variabele voor de sterfte-intensiteit. Uit gesprekken die Snow voerde met nabestaanden en buren bleek tevens dat slachtoffers uit andere buurten vaak nabij Broad Street werkten of naar school gingen. Bovendien valt de afwezigheid van choleragevallen in het blok net ten noorden Broad Street te verklaren uit de aanwezigheid van een werkhuis aldaar met een eigen watervoorziening. Hetzelfde geldt voor de brouwerij in het kleinere blok ten oosten van de pomp, hoewel de werknemers daar voornamelijk bier dronken. Snow zag aldus zijn hypothese dat cholera wordt verspreid door besmet water bevestigd en haalde de gemeente over om de bewuste pomp buiten werking te stellen. Later bleek dat de put in Broad Street nog geen meter verwijderd lag van een besmette beerput. Zie ook het afscheidscollege van Professor Tegelaers voor een vergelijkbaar verhaal uit onze tijd [28].

In het zojuist behandelde voorbeeld is het doel om de invloed van verklarende variabelen (ook wel 'covariaten' genoemd) op een ruimtelijk patroon te onderzoeken. Daarnaast kan er ook afhankelijkheid zijn tussen de punten onderling en kunnen deze tevens dragers zijn van additionele informatie. Wellicht herkent u het plaatje in Figuur 2 dat de afgelopen jaren veel in het nieuws is geweest. Het grillig gevormde gebied is het gasveld in de provincie Groningen. De middelpunten van de cirkels stellen de epicentra voor van aardbevingen, hun diameters zijn recht evenredig aan de magnitude. Ook wordt van elke beving bijgehouden wanneer hij optreedt. Dit soort extra informatie bovenop de locatie wordt 'markering' genoemd. Uiteraard zijn er ook covariaten beschikbaar betreffende gaswinning, breuklijnen en geologie van de bodem [18].

Net als in het cholera-voorbeeld is ook hier het verband tussen het waarge-



Figuur 1 Adressen van ten gevolge van cholera overleden inwoners van Soho gedurende de uitbraak van 1854 (vierkantjes) en locatie van waterputten (driehoekjes). De kaart is door R. Dodson (NCGIA, Santa Barbara) gedigitaliseerd en opgenomen in het R-pakket HistData. Zie [8] voor nadere details.

nomen patroon (bevingen) en covariaten (gaswinning) van belang. Daarnaast is men geïnteresseerd in de structuur van het ruimtelijk patroon, bijvoorbeeld in de vorm van naschokken, en in patronen in de tijd, zoals de vraag of er een neiging bestaat tot zwaardere bevingen. In Twente hebben drie studenten het afgelopen semester naar deze data gekeken. Zij vonden een significant verband tussen de jaarlijkse omvang van de gaswinning en het aantal bevingen; een toename van de magnitude bleek niet significant. Daarnaast hebben zij een model opgesteld voor de bevingsintensiteit in termen van de hierboven genoemde covariaten en binnen dit model verschillende gaswinningsscenario's doorgerekend [10].

Geschiedenis

De geboortedatum van het vakgebied ruimtelijke statistiek is moeilijk exact te bepalen. We zagen al het werk van Snow in de negentiende eeuw maar kunnen verder terug in de tijd, naar het jaar 1767 en het plaatsje Thornhill bij Leeds, waar John Michell (1724–1793) rector van de Heilige Michaëlskerk was. Deze veelzijdige man was al hoogleraar geologie geweest in Cambridge, en zou later de eerste wetenschapper worden die het bestaan van zwarte gaten postuleerde. Ons interesseert nu zijn artikel over dubbelsterren.

Michell wilde de zwaartekrachttheorie van Newton ondersteunen door te laten zien dat er meer sterren in paren voorkomen dan door toeval kan worden verklaard [14]. Daartoe beschouwt hij de hemelsfeer, de rand van een denkbeeldige bol met straal $R = 60 \times (180/\pi)$ boogminuten die concentrisch is met de aarde en waartegen de hemellichamen worden waargenomen. Een kleine 'cirkel' met straal r op de hemelsfeer is bij goede benadering twee-dimensionaal met oppervlakte πr^2. Omdat de boogoppervlakte van de hemelsfeer $4\pi R^2$ is, heeft een willekeurig geplaatste ster kans

$$p(r) = \frac{r^2}{4R^2}$$

om binnen zulk een cirkel te liggen. Evenzo is de kans dat geen enkele van n willekeurig en onafhankelijk van elkaar geplaatste sterren binnen zulk een cirkel valt gelijk aan $(1-p(r))^n$.

Michell was geïnteresseerd in de kans dat geen enkele ster in de buurt van welke andere ster dan ook ligt, en benaderde deze kans door $(1-p(r))^{n^2}$. Er zijn immers n^2 paren te vormen uit n sterren. Op deze manier rekenend voor de dubbelster β-Capricorni krijgt hij (met $n = 230$, het aantal hem bekende sterren met een helderheid die tenminste even groot is als die van β-Capricorni, en $r = 3{,}33$, de afstand tussen de twee leden van de dubbelster) een kans van $0{,}988$. Met andere woorden, wanneer sterren geheel willekeurig en onafhankelijk van elkaar over de hemelsfeer zouden zijn verstrooid, zou een stelsel als β-Capricorni kans $1-0{,}988$, minder dan twee procent, hebben om voor te komen. In moderne termen: de nulhypothese van een willekeurige verstrooiing op de hemelsfeer wordt bij een drempelwaarde van twee procent verworpen. Michell verklaart het tóch voorkomen van onwaarschijnlijke configuraties als β-Capricorni door aantrekking tussen de hemellichamen, met andere woorden, door de zwaartekracht. Nadere details kunt u vinden in Hughes en Cartwright [11].

Hoewel in Michells berekeningen een fout zit (n mag niet zomaar worden vervangen door n^2 aangezien de cirkels om de n sterren elkaar kunnen overlappen), zien we hier een voorbeeld van een Poissonproces ruim voor de geboorte van Siméon-Denis Poisson in 1781. De naam is waarschijnlijk in zwang geraakt op de Universiteit van Stockholm aan het eind van de dertiger jaren. Hij duikt op in een artikel uit 1940 van de bekende kanstheoreticus William Feller die van 1934 tot 1939 in Stockholm werkte [7]. De algemenere term 'puntproces' werd geïntroduceerd door Connie Palm in zijn proefschrift uit 1943 [19]. Het zou nog twee decennia duren voordat José Enriques Moyal in 1962 een stevige wiskundige basis zou leggen onder de theorie van puntprocessen in algemene ruimten [15]. Voor een uitgebreider historisch overzicht verwijs ik graag naar Guttorp and Thorarinsdottir [9].

De moderne geschiedenis neemt een aanvang in de zeventiger jaren van de vorige eeuw. Een buitengewoon invloedrijk man is Brian Ripley, inmiddels emeritus hoogleraar te Oxford. Hij heeft een aantal boeken op zijn naam, waaronder *Spatial Statistics* uit 1981 [24], en daarnaast een grote bijdrage geleverd aan software ontwikkeling [30].

In 1977 hield Ripley een voordracht voor de Royal Statistical Society in Londen die met de erop volgende discussie is gepubliceerd en nu, bijna veertig jaar later, nog volop geciteerd wordt [23]. In deze rede brengt hij voor het eerst een aantal modellen bij elkaar die in gescheiden toepassingsdomeinen waren ontwikkeld, waaronder die van Quenouille–Cox [6, 22], Matérn [13], Neyman en Scott [17], Strauss [27] alsmede de Markov-modellen die hij zelf recent samen met Frank Kelly had gekarakteriseerd [25]. Voor deze laatste klasse ontwikkelt hij een simulatiemethode die kan worden gezien als een voorloper van de huidige Monte Carlomethoden [16].

Ten slotte bespreekt hij zijn filosofie omtrent toetsen voor aanpassing. Hij moet aannemen dat het onderliggende kansmodel stationair is, dat wil zeggen dat zijn eigenschappen niet veranderen als alle punten evenveel worden verschoven. Wanneer er geen interactie tussen de punten zou zijn, zou het verwachte aantal punten in een cirkel om een specifiek punt rechtevenredig zijn aan de oppervlakte van deze cirkel. Als functie van de straal krijgt men dus een kwadratische grafiek. Deze wordt vervolgens vergeleken met de daadwerkelijk waargenomen aantallen.

Deze aanpak lijkt op die van John Michell twee eeuwen eerder. In beide gevallen wordt een statistiek gekozen (in het geval van Michell een kans, in dat van Ripley een verwacht aantal) waarvan de waarde wordt berekend onder een verondersteld model zoals willekeurige verstrooiing en vervolgens vergeleken met de data. Het verschil is dat Ripley eventuele interactie op elke schaal meeneemt door de straal van zijn cirkels te laten variëren. De bovengenoemde filosofie zou de ruimtelijke stochastiek tot circa 1990 domineren en is nog steeds wijdverbreid.

Figuur 2 Cirkels rond epicentra van aardbevingen in het Groninger veld gedurende 2014 en met straal recht evenredig aan de magnitude. Bevingen met magnitude $1{,}5$ of zwaarder zijn in grijs weergegeven. Deze gegevens worden bijgehouden op het portaal [18].

Belangrijke doorbraken

Vanaf de tachtiger jaren heeft de ruimtelijke stochastiek een grote vlucht genomen. Hier zijn verschillende verklaringen voor aan te voeren. Een eerste belangrijke factor is de ontwikkeling van de computer. Eind jaren tachtig had lang niet iedere wetenschapper een pc op zijn kamer; zalen met op een universitair mainframe aangesloten terminals waren de norm. Op het CWI werden in die tijd de eerste SPARC work stations aangekocht. Ook netwerken kwamen langzaam op. Zo is SURFnet opgericht in 1986, terwijl op 17 november 1988 Piet Beertema de bevestiging ontving dat het CWI (als eerste in Europa) met het Amerikaanse computernetwerk NSFnet, een voorloper van het huidige web, was verbonden.

Op theoretisch gebied moet het pionierswerk van Julian Besag worden vermeld. Deze in 2010 overleden Britse statisticus is, onder andere, bekend geworden door zijn in 1974 gepresenteerde idee om complexe ruimtelijke processen te modelleren door middel van eenvoudige zogenaamde lokale specificaties [2]. In een voordracht voor de Royal Statistical Society legt hij zijn aanpak uit aan de hand van data over de aanwezigheid van *Plantago lanceolata* (smalle weegbree) in een voormalig mijngebied in Wales. Dit gebied is verdeeld in een 10×940-rooster met cellen van twee cm bij twee cm, zodat een simultane verdeling maar liefst 9.400-dimensionaal is. Een lokale specificatie daarentegen kan worden opgebouwd uit een collectie eendimensionale kansen, bijvoorbeeld een eerste orde auto-logistieke regressie

$$p_{ij}(1 \mid y_{ij}) = \frac{e^{\alpha+\beta y_{ij}}}{1+e^{\alpha+\beta y_{ij}}}.$$

Hier stelt $p_{ij}(1 \mid y_{ij})$ de voorwaardelijke kans voor dat cel (i, j) smalle weegbree bevat onder de aanname dat van de vier naastgelegen cellen er precies $y_{ij} \in \{0, \ldots, 4\}$ ook deze plant bevatten. In onderstaand voorbeeld is $y_{ij} = 3$.

$$\begin{array}{ccc} & 1 & \\ 1 & ? & 0 \\ & 1 & \end{array}$$

Natuurlijk moeten de lokale specificaties wel consistent zijn. Een belangrijk bijkomend voordeel van Besags aanpak is dat de eendimensionale verdelingen eenvoudig van vorm zijn en geen normalisering behoeven. In ons voorbeeld is de

Illustratie: Ryu Tajiri

simultane kans $p(C)$ op een configuratie $C=(C_{ij})_{ij}$ van nullen (`geen planten in de cel') en enen gelijk aan

$$p(C)=\frac{1}{Z}\exp\left[\alpha\sum C_{ij}+\beta\sum_{\{ij\text{ ligt naast }kl\}}C_{ij}C_{kl}\right].$$

De constante Z is een som over het astronomische aantal van $2^{9.400}$mogelijke configuraties, een getal met meer dan 2829 nullen! Wanneer ook de vier diagonale buren van de centrale cel worden meegenomen, wordt het totale aantal configuraties zelfs nog groter.

Besag paste de zojuist beschreven aanpak toe in de beeldanalyse [3], in agrarisch veldonderzoek en voor het maken van risicokaarten [5]. Ook puntprocessen kunnen worden beschreven door een lokale specificatie omdat, wanneer de cellen klein zijn, er ten hoogste één punt in elke cel valt [4].

Naast een elegante manier om complexe, gestructureerde modellen te bouwen uit overzichtelijke bouwstenen, leent Besags idee zich ook uitstekend voor iteratieve berekeningen. Hij besefte dat de zogenaamde Monte Carlo-methoden die gedurende de Tweede Wereldoorlog door fysici waren ontwikkeld, konden worden aangepast voor het genereren van steekproeven, voor het schatten van modelparameters, het toetsen van hypotheses en het construeren van betrouwbaarheidsintervallen. Dit samenspel tussen modelbouw, de daarop gebaseerde algoritmen en de beschikbaarheid van snelle computers heeft tot een revolutie in de (Bayesiaanse) statistiek geleid.

Omdat Monte Carlo-methoden gebaseerd zijn op het heel vaak uitvoeren van simpele berekeningen, hebben zij als nadeel dat het niet eenvoudig is om te bepalen wanneer ‘heel vaak’ ook genoeg is. Rond het millennium was er — in navolging van Propp en Wilson [20] — veel enthousiasme over een klasse van Monte Carlo-methoden die zelf aangeven wanneer te stoppen. Dit enthousiasme lijkt wat bekoeld, voornamelijk omdat de constructie ingewikkeld en delicaat is. Bovendien kan het lang duren voor men mag stoppen.

Een volgende belangrijke factor is dat er, mede door het internet, het laatste decennium veel grootschalige dataverzamelingen beschikbaar zijn gekomen, waardoor het niet langer noodzakelijk is om bij gebrek aan data simplificerende maar niet-realistische modelaannames te maken. Opensourcesoftware [1] tenslotte, heeft er voor gezorgd dat grote groepen gebruikers zelf hun gegevens kunnen analyseren.

Ruimtelijke stochastiek in het derde millennium

De onderzoeksgroepen waar ik deel van heb mogen uitmaken, hebben aan alle in het vorige hoofdstuk genoemde aspecten bijgedragen.

Veel aandacht is besteed aan de ontwikkeling van voorwaardelijk gespecificeerde modellen voor puntprocessen en, meer in het algemeen, stochastische verzamelingen, alsmede aan een juist begrip van de bijbehorende interactiestructuren. Dit onderzoek is uitgemond in een monografie [12].

Wat de beeldanalyse betreft, hebben wij Besags modellen gebruikt voor de analyse van vlakverdelingen in de Indiase miniatuurkunst, stonden wij aan de basis van het gebruik van gemarkeerde punt- en objectprocessen voor het herkennen van objecten in beelden en van sequentiële puntprocessen voor het volgen van zulke objecten in video’s, en hebben wij laten zien dat zowel lijnsegmentprocessen als polygonale Markovvelden kunnen worden gebruikt voor segmentatie en netwerkextractie. En passant werden nieuwe, op maat gemaakte Monte Carlo-methoden ontwikkeld, waaronder exacte simulatiemethoden.

Een ander belangrijk onderzoeksterrein betrof zogenaamde momentmaten en de daarop gebaseerde statistieken. Een specifiek geval kwamen we al tegen bij Ripley die keek naar het verwachte aantal punten in de buurt van een specifiek punt. Ook zagen we al de intensiteitsfunctie die weergeeft hoe waarschijnlijk het is om punten aan te treffen op gegeven locaties. De combinatie van zulke momentmaten levert krachtige statistieken op voor exploratieve data-analyse en modelvalidatie die ook voor niet-homogene data kunnen worden gebruikt. Samen met professor Stein van ITC zijn deze statistieken toegepast op data met betrekking tot bodemgesteldheid en aardbevingen.

Binnen de vakgroep Stochastische Operations Research aan de Universiteit van Twente bestudeert men het optimaliseren van processen waarbij onzekerheid een rol speelt. Vaak moeten kosten en baten tegen elkaar worden afgezet en tegelijkertijd rekening worden gehouden met allerlei nevenvoorwaarden. U kunt denken aan het voorraadbeheer van een winkelier die besluit hoeveel artikelen hij moet bestellen op basis van de verwachte vraag van klanten, de bestel- en opslagkosten en rekening houdend met de grootte van zijn magazijn [21, 29].

Dergelijke vragen kunnen ook gesteld worden in een ruimtelijke context. Verplaatst u zich, om de gedachten te bepalen, eens in een boswachter die een ongelijkjarig gemengd perceel beheert. De bomen op zijn perceel vormen een puntpatroon; als markering gebruiken we de diameter en het houtvolume. Onze boswachter kan elk jaar besluiten om alle bomen die dikker zijn dan een zekere drempelwaarde te kappen. Hoe deze drempelwaarde te kiezen? Een te kleine keuze betekent dat erg veel bos wordt gekapt zodat er weliswaar op korte termijn veel inkomsten zijn maar het ook lang duurt voordat het bos weer aangroeit; een te hoge waarde leidt tot weinig houtopbrengst en dus weinig inkomsten.

Om de beste drempelwaarde te kunnen bepalen, zijn een aantal stappen nodig. Allereerst moet men de onzekerheden kwantificeren, dat wil zeggen, een dynamisch model bouwen voor de ontwikkeling van het bos met inachtneming van groei, de wedijver om zonlicht en voedingsstoffen, natuurlijke sterfte alsmede de ontwikkeling van de houtprijs. Dan moet de verwachte netto opbrengst over een geschikte tijdshorizon worden bepaald als functie van de drempelwaarde, en tenslotte moet deze functie worden gemaximaliseerd. Uiteraard zijn legio vergelijkbare scenario’s te bedenken.

Een geïntegreerde theorie voor ruimtetijd-puntprocessen staat nog in de kinderschoenen. Vaak wordt de ruimte dan wel de tijd als een markering opgevat. In die gevallen waarin ruimte en tijd wel op gelijke voet staan, zijn de gekozen modellen meestal separabel, dat wil zeggen dat de momentmaten factoriseren in ruimte en tijd. Zulk een aanpak is handig, maar lijkt minder realistisch. Ook predictie, zowel tijdelijk in de zin van het voorspellen van toekomstig gedrag, als ruimtelijk in de zin van het invullen van een onvolledig waargenomen patroon, verdient nadere aandacht en theorievorming. Hetzelfde kan worden gezegd van optimaliserings- en besturingsproblemen voor zich dynamisch ontwikkelende patronen.

Kansen en uitdagingen genoeg om de komende tijd mee aan de slag te gaan!

Referenties

1 A. Baddeley, E. Rubak en R. Turner, *Spatial Point Patterns: Methodology and Applications with R*, CRC Press, 2015.

2 J. Besag, Spatial interaction and the analysis of lattice systems (with discussion), *Journal of the Royal Statistical Society, Series B (Methodological)* 36 (1974), 192–236.

3 J. Besag, On the statistical analysis of dirty pictures (with discussion), *Journal of the Royal Statistical Society, Series B (Methodological)* 48 (1986), 259–302.

4 J. Besag, R. Milne en S. Zachary, Point process limits of lattice processes, *Journal of Applied Probability* 19 (1982), 210–216.

5 J. Besag, J. York en A. Mollié, Bayesian image restoration, with two applications in spatial statistics. With discussion and a reply by Besag, *Annals of the Institute of Statistical Mathematics* 43 (1991), 1–59.

6 D.R. Cox, Some statistical methods connected with series of events (with discussion), *Journal of the Royal Statistical Society, Series B (Methodological)* 17 (1955), 129–164.

7 W. Feller, On the integro-differential equations of purely discontinuous Markoff processes, *Transactions of the American Mathematical Society* 48 (1940), 488–515.

8 M. Friendly, HistData: Data sets from the history of statistics and data visualization, http://CRAN.R-project.org/package=HistData, 2014.

9 P. Guttorp en T.L. Thorarinsdottir, What happened to discrete chaos, the Quenouille process, and the sharp Markov property? Some history of stochastic point processes, *International Statistical Review* 80 (2012), 253–268.

10 E. van Hove, R. van Lingen en S. Riemens, *Geïnduceerde aardbevingen in gasveld Groningen. Een statistische analyse*, bachelorscriptie, Universiteit Twente, 2015.

11 D.W. Hughes en S. Cartwright, John Michell, the Pleiades, and odds of 496,000 to 1, *Journal of Astronomical History and Heritage* 10 (2007), 93–99.

12 M.N.M. van Lieshout, *Markov Point Processes and their Applications*, Imperial College Press, 2000.

13 B. Matérn, Spatial variation, *Meddelanden från Statens Skogsforskningsinstitut* 49 (1960).

14 J. Michell, An inquiry into the probable parallax and magnitude of the fixed stars from the quantity of light which they afford us, and the particular circumstances of their situation, *Philosophical Transaction of the Royal Society* 57 (1767), 234–254.

15 J.E. Moyal, The general theory of stochastic population processes, *Acta Mathematica* 108 (1962), 1–31.

16 J. Møller en R.P. Waagepetersen, *Statistical Inference and Simulation for Spatial Point Processes*, CRC Press, 2004.

17 J. Neyman en E.L. Scott, A theory of the spatial distribution of galaxies, *Astrophysical Journal* 116 (1952), 144–163.

18 NL Olie en Gasportaal, www.nlog.nl, Ministerie van Economische Zaken, Koninklijk Nederlands Meteorologisch Instituut en TNO Innovation for life, 2015.

19 C. Palm, Intensitätsschwankungen im Fernsprechverkehr, *Ericsson Technics* 44 (1943).

20 J.G. Propp en D.B. Wilson, Exact sampling with coupled Markov chains and applications to statistical mechanics, *Random Structures & Algorithms* 9 (1996), 223–252.

21 M.L. Puterman, *Markov Decision Processes*, Wiley, 1994.

22 M.H. Quenouille, Problems in plane sampling, *Annals of Mathematical Statistics* 20 (1949), 355–375.

23 B.D. Ripley, Modelling spatial patterns (with discussion), *Journal of the Royal Statistical Society, Series B (Methodological)* 39 (1977), 172–212.

24 B.D. Ripley, *Spatial Statistics*, Wiley, 1981.

25 B.D. Ripley en F.P. Kelly, Markov point processes, *Journal of the London Mathematical Society* 15 (1977), 188–192.

26 J. Snow, *On the Mode of Communication of Cholera*, John Churchill, 2nd ed., 1855.

27 D.J. Strauss, A model for clustering, *Biometrika* 62 (1975), 467–475.

28 W.H.H. Tegelaers, Het kind van de rekening, afscheidscollege, Universiteit van Amsterdam, 1984.

29 H.C. Tijms, *A First Course in Stochastic Models*, Wiley, 2003.

30 W.N. Venables en B.D. Ripley, *Modern Applied Statistics with S*, Springer, 1994.