

RÄUMLICHE PUNKTPROZESSE
UND IHRE ANWENDUNG
IN BIOLOGIE UND ÖKOLOGIE ¹⁾

D. Pfeifer, Fachbereich Mathematik ²⁾
 Carl von Ossietzky Universität Oldenburg

Inhalt

1. Einleitung und Motivation
2. Was ist ein Punktprozeß?
3. Räumliche Poisson–Prozesse
4. Allgemeine Punktprozesse
5. Boole–Modelle
6. Mosaike
7. Zeitdynamische Punktprozesse

1. Einleitung und Motivation

Zufällige räumliche oder flächige Punktmuster findet man in allen möglichen Bereichen von Biologie und Ökologie: sei es bei der Ausbreitung von Insektenlarven auf Tabakpflanzen (RICHTER UND SÖNDGERATH (1990), Example 3.1), bei der Verteilung bestimmter Baumarten in Waldgebieten (CRESSIE (1991), Chapter 8.2: Spatial Data Analysis of Longleaf Pines; MECKE, SCHNEIDER, STOYAN UND WEIL (1990), S. 136ff.; RIPLEY (1981), Chapter 8.6: Examples), der Verteilung von Vogelnestern (STOYAN, KENDALL UND MECKE (1989), Example 5.1; RIPLEY (1981), S. 180ff.), oder der Wiederbesiedlung verödeter Wattflächen durch Würmer und Muscheln (PFEIFER, BÄUMER UND ALBRECHT (1992)), um exemplarisch einige Beispiele aufzuzählen.

Obgleich die mathematische Theorie der zufälligen Punktmuster und weitergehender, daraus abgeleiteter Modelle inzwischen recht weit entwickelt ist und es – wie die obigen Zitate zeigen – auch ausreichend anwendungsorientierte Darstellungen in der Literatur gibt, scheint doch die Bedeutung dieser Theorie für die mathematische Modellierung in den Bereichen Biologie und Ökologie allgemein noch nicht genügend anerkannt zu sein. In diesem Beitrag sollen deshalb einige der wichtigsten mathematischen Grundlagen der statistischen Punktprozeßtheorie sowie einiger daraus abgeleiteter Modelle vorgestellt werden.

¹⁾ Kurzfassung; eine ausführlichere englische Version ist in *Modeling of Geo– Biosphere Processes* erschienen, vgl. die Literaturliste am Ende.

²⁾ jetzt Institut für Mathematische Stochastik, Universität Hamburg.

2. Was ist ein Punktprozeß?

Anschaulich versteht man unter einem statistischen Punktprozeß im wesentlichen ein flächig oder räumlich zufällig angeordnetes Punktmuster, bei dem i.a. auch die Gesamtzahl der Punkte selbst noch zufällig ist. Will man eine solche Punktconfiguration – nennen wir sie ξ – mathematisch beschreiben, so bietet sich zunächst an, formal eine Darstellung als Zufallsvektor mit einer zufälligen Anzahl von Komponenten zu wählen, d.h.

$$\xi = (X_1, X_2, \dots, X_N) \quad (1)$$

mit Zufallsvektoren X_1, X_2, \dots in \mathbb{R}^2 bzw. \mathbb{R}^3 und einer weiteren ganzzahligen Zufallsvariablen N . Hierbei kommt den Zufallsvektoren X_1, X_2, \dots die Bedeutung der zufälligen *Position* der Punkte und der Zufallsvariablen N die Bedeutung der zufälligen *Anzahl* der Punkte zu. Allerdings läßt sich zeigen, daß dieser Formalismus sehr schnell zu Schwierigkeiten führt, wenn man z.B. die Verteilung der Punkte nur in gewissen Teilmengen $A \subset \mathbb{R}^2$ bzw. $A \subset \mathbb{R}^3$ betrachten will, was typischerweise der Fall ist, wenn man sich in Feldversuchen auf bestimmte festgelegte Untersuchungsgebiete konzentrieren muß. Dazu wären nämlich aus dem Zufallsvektor ξ nach (1) gerade diejenigen Komponenten X_{i_1}, X_{i_2}, \dots auszuwählen, welche in der Menge A liegen; da die Indices i_1, i_2, \dots damit selbst wieder zufällig sind und darüberhinaus von der Menge A abhängen, ergeben sich sofort Probleme in Bezug auf eine einigermaßen übersichtliche formale Darstellung. Aus diesem Grund wählt man zweckmäßigerweise einen maßtheoretisch orientierten Zugang in der Form

$$\xi = \sum_{k=1}^N \varepsilon_{X_k}, \quad (2)$$

bei dem ε_x das *Dirac-Maß* (auch Einpunkt-Verteilung genannt) bezeichnet, d.h. es gilt für $x \in \mathbb{R}^2$ bzw. $x \in \mathbb{R}^3$:

$$\varepsilon_x(A) = \begin{cases} 1, & x \in A \\ 0, & \text{sonst.} \end{cases} \quad (3)$$

Bei dieser Darstellung beschreibt dann die Zufallsvariable

$$\xi(A) = \sum_{k=1}^N \varepsilon_{X_k}(A) \quad (4)$$

die sich aus der Gesamtkonfiguration ergebende zufällige Anzahl der Punkte innerhalb einer geeigneten Menge $A \subset \mathbb{R}^2$ bzw. $A \subset \mathbb{R}^3$. Mathematisch handelt es sich bei einem Punktprozeß ξ also um ein *zufälliges diskretes Maß* mit den – zufälligen – Trägerpunkten X_1, X_2, \dots

Der Nachweis der Existenz und Wohldefiniertheit solcher Objekte ist mathematisch aufwendig und erfordert i.a. den Einsatz von Methoden der *topologischen Maßtheorie* (vgl. etwa DALEY UND VERE-JONES (1988) oder STOYAN, KENDALL UND MECKE (1989)). Elementarere Darstellungen geben z.B. COX UND ISHAM (1980), MATHAR UND PFEIFER (1990), Kapitel 3.4 oder TOPSØE (1990).

3. Räumliche Poisson-Prozesse

Das wichtigste Basismodell eines Punktprozesses ist sicherlich das eines flächigen oder räumlichen *Poisson-Prozesses*. Man erhält dieses aus (2), wenn dort die Zufallsvektoren X_1, X_2, \dots als stochastisch unabhängig mit derselben räumlichen Verteilung – sagen wir P – gewählt werden und die Anzahlvariable N hiervon ebenfalls unabhängig ist mit einer Poisson-Verteilung $\mathfrak{P}(\lambda)$, $\lambda > 0$. Die Zufallsvariablen $\xi(A)$ sind dann alle selbst wieder Poisson-verteilt mit dem Parameter (Erwartungswert) $E[\xi(A)] = \lambda P(A)$. Man nennt diesen Erwartungswert auch das *Intensitätsmaß* des Poisson-Prozesses, i.Z.: $E\xi(A)$. Betrachtet man die Menge A hierin als Variable, so erhält man in der Tat durch $E\xi$ ein *Maß* im Sinne der mathematischen Maßtheorie. $E\xi(A)$ beschreibt anschaulich gerade die *mittlere* Anzahl von Punkten in A ; in einem Bereich A , in dem diese mittlere Punktdichte hoch ist, werden also sehr wahrscheinlich mehr Punkte der Gesamtkonfiguration liegen als in solchen Bereichen, in denen die mittlere Punktdichte niedrig ist – hieraus erklärt sich auch die Bezeichnung *Intensitätsmaß*.

Eine weitere charakteristische Eigenschaft von Poisson-Prozessen ist die Unabhängigkeit der Punktzahlen in flächig bzw. räumlich disjunkten Bereichen, d.h. die Zufallsvariablen $\xi(A_1), \xi(A_2), \dots$ sind stets stochastisch unabhängig, wenn die Mengen A_1, A_2, \dots paarweise disjunkt sind, sich also nicht überlappen.

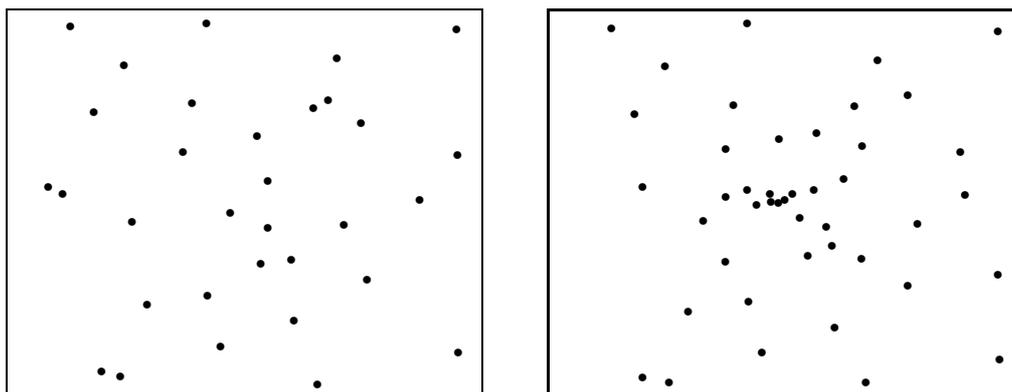
Das Intensitätsmaß $E\xi$ ist hier insofern charakteristisch für einen Poisson-Prozeß ξ , als es bereits die statistische Verteilung der Gesamtkonfiguration (als zufälliges Maß) eindeutig festlegt.

Ist $\mathcal{X} \subset \mathbb{R}^2$ bzw. $\mathcal{X} \subset \mathbb{R}^3$ eine abgeschlossene, beschränkte Menge – typischerweise ein Rechteck oder Quader – und P die *Gleichverteilung* über \mathcal{X} , so spricht man auch von einem *homogenen* Poisson-Prozeß über \mathcal{X} . Das Intensitätsmaß ist dann gerade ein Vielfaches des üblichen Flächeninhalts bzw. Volumens für alle geeigneten Teilmengen A von \mathcal{X} , weswegen man in diesem Fall auch von "rein zufälligen" Punktmustern spricht, da die mittlere Punktdichte dann für alle kongruenten Teilmengen A gleich ist. (Solche Punktmuster treten z.B. in sehr guter Näherung im Watt bei den Pierwürmern *Arenicola marina* auf.)

Statistische Tests zur Überprüfung der Annahme eines homogenen Poissonprozesses lassen sich ebenfalls durchführen. Bei dem "klassischen" Dispersionsindex-Test (vgl. etwa RICHTER UND SÖNDGERATH (1990), Kapitel 3.1.5) wird das Untersuchungsgebiet in n gleich große, paarweise disjunkte Fenster A_1, \dots, A_n unterteilt; unter der Annahme eines homogenen Poissonprozesses sind dann die Zufallsvariablen $Z_i = \xi(A_i)$, $1 \leq i \leq n$, stochastisch unabhängig und identisch Poisson-verteilt. Für die Poisson-Verteilung $\mathfrak{P}(\mu)$ mit dem Parameter μ stimmen aber der Erwartungswert μ und die Varianz überein; nach dem Gesetz der großen Zahlen kann man also davon ausgehen, daß dieser Sachverhalt sich näherungsweise auch für die aus den Daten Z_1, \dots, Z_n abgeleiteten empirischen Größen Mittelwert \bar{Z} und Streuung σ^2 ergibt, wobei

$$\bar{Z} = \frac{1}{n} \sum_{i=1}^n Z_i, \quad \sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (Z_i - \bar{Z})^2 \quad (5)$$

ist. Als Testgröße verwendet man hier zweckmäßigerweise den sogenannten *Dispersionssindex* $D_n = (n-1)\sigma^2/\bar{Z}$. Da der Dispersionsindex-Test wegen seiner lediglich asymptotisch χ^2 -verteilten Testgröße D_n nur bei großen beobachteten Punktzahlen verlässlich ist, stehen heute mehr die vor allem durch Ripley entwickelten *Abstandsmethoden* im Vordergrund (vgl. etwa RIPLEY (1981), (1988) oder STOYAN, KENDALL UND MECKE (1989)). Hier ist insbesondere eine einfache Konstruktion von Konfidenzbändern für die Testfunktion mit Hilfe von Monte-Carlo-Methoden möglich.



*Realisation eines flächig
homogenen Poisson-Prozesses*

*Realisation eines inhomogenen
unimodalen Poisson-Prozesses*

Fig. 1

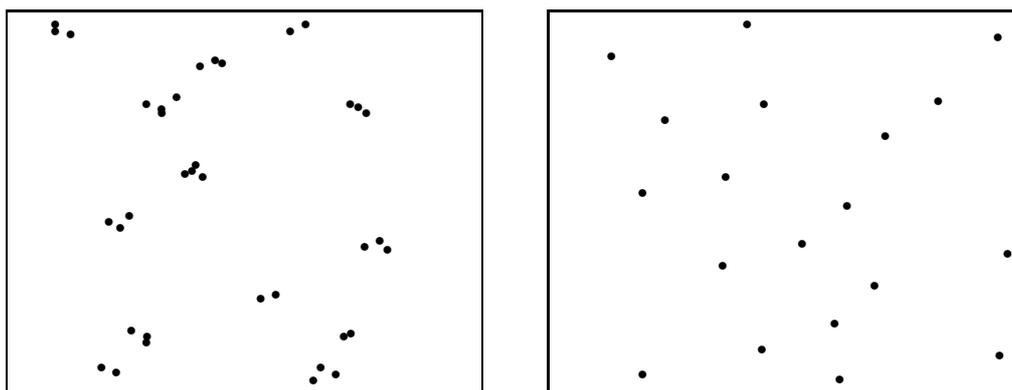
4. Allgemeine Punktprozesse

In der Natur treten zufällige Punktconfigurationen häufig nicht in homogener Form, sondern entweder "geklumpt" oder eher "gleichmäßig" auf. Man spricht dann im ersten Fall von Punktprozessen mit *Anziehung*, im zweiten von Punktprozessen mit *Abstoßung*. Ein Beispiel für die erste Situation ist etwa die eingangs erwähnte Ausbreitung von Insektenlarven auf Tabakpflanzen oder die Verteilung von Bäumen in Jungwäldern, während die zweite Situation z.B. bei älteren Baumbeständen oder der Verteilung von Vogelnestern zu beobachten ist. Zur Beschreibung solcher Phänomene durch allgemeine Punktprozesse (allgemeine zufällige diskrete Maße) lassen sich oft zweistufige Modelle heranziehen, die als Basis-Modell einen homogenen Poisson-Prozeß verwenden. Beispielsweise sind bei gewissen *Clusterprozessen* die sogenannten Elternpunkte homogen Poisson-verteilt, die Tochterpunkte entstammen dann weiteren Punktprozessen, die unabhängig von der Ausgangskonfiguration um die Elternpunkte herum angesiedelt sind. Bei gewissen *Neyman-Scott-Prozessen* sind die Tochterprozesse z.B. selbst wieder Poisson-Prozesse, beim *Thomas-Prozeß* entstammen die Tochterpunkte mehrdimensionalen Normalverteilungen (vgl. RICHTER UND SÖNDGERATH (1990), Kapitel 3, oder

STOYAN, KENDALL UND MECKE (1989), Kapitel 5). Bei den sogenannten SSI-Prozessen ("simple sequential inhibition") werden in geeigneter Weise um die Punkte eines homogenen Poisson-Prozesses sukzessiv Kreise oder andere Mengen gelegt; diejenigen Punkte der Ausgangskonfiguration, die auf diese Weise "in der Nähe" der ausgewählten Mittelpunkte liegen, werden eliminiert (vgl. etwa RICHTER UND SÖNDGERATH (1990), Kapitel 3.1.2 oder MECKE, SCHNEIDER, STOYAN UND WEIL (1990), Kapitel 4.5).

Naturgemäß sind für solche zusammengesetzten Modelle statistische Verfahren nur noch beschränkt exakt ableitbar; häufig muß man hier mit Monte-Carlo-Methoden Vorlieb nehmen (vgl. aber die Ausführungen in STOYAN, KENDALL UND MECKE (1989), Kapitel 4 und 5, und MECKE, SCHNEIDER, STOYAN UND WEIL (1990), Kapitel 4.3 bis 4.5).

Eine weitere Schwierigkeit für den Anwender besteht häufig darin, daß unterschiedliche Modelle durchaus ähnliche Punktkonfigurationen erzeugen können: z.B. läßt sich ein nicht-homogener Poisson-Prozeß mit einem multimodalen Intensitätsmaß nicht ohne weiteres von einem Clusterprozeß unterscheiden. Bei der Modellierung nicht-homogener Phänomene in Biologie und Ökologie ist daher immer besondere Vorsicht in Bezug auf die Auswahl möglicher Modelle geboten.



Realisation eines Clusterprozesses

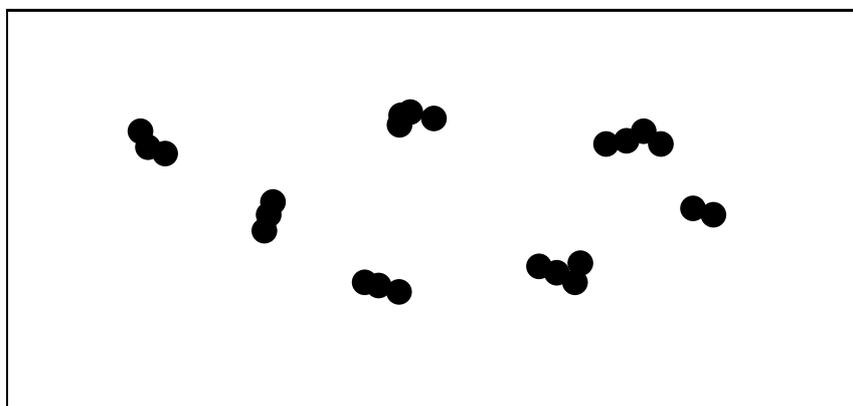
Realisation eines SSI-Prozesses

Fig. 2

5. Boole-Modelle

Neben reinen Punktkonfigurationen treten in der Natur sehr häufig fleckenförmige Besiedlungsformen oder vergleichbare Muster auf (z.B. Algenteppiche, anoxische Flecken im Watt, Muschelbänke usw.). Die im vorigen Abschnitt angedeutete Konstruktion von Clusterprozessen läßt sich für eine mathematische Modellierung solcher Phänomene sofort dahingehend verallgemeinern, daß man um die Elternpunkte z.B. eines homogenen Poisson-Prozesses anstatt von Punkten andere geeignete geometrische Objekte (z.B.

Kreise, Rechtecke, Ellipsen, Kugeln u.a.) legt (sog. *Boole-Modelle*). Die so entstehenden "zufälligen Mengen" sind Untersuchungsgegenstand der *mathematischen Morphologie*; eine Übersicht über die zugehörige Theorie gibt etwa RIPLEY (1981), Kapitel 9, RIPLEY (1988), Kapitel 6, STOYAN, KENDALL UND MECKE (1989), Kapitel 1.4, 3 und 6, oder CRESSIE (1991), Kapitel 9. Statistisch von Interesse sind hier vor allem Schätzungen für die Flächen- bzw. Raumdichte, d.h. die mittlere Fläche bzw. das mittlere Volumen des Boole-Modells bezogen auf das Einheitsquadrat bzw. den Einheitswürfel; ferner Schätzungen von Kovarianzen oder der sogenannten sphärischen *Kontaktverteilungsfunktion*, d.h. im wesentlichen der Wahrscheinlichkeit dafür, daß das Boole-Modell einen festen Kreis bzw. eine feste Kugel schneidet, unter der Bedingung, daß der zugehörige Mittelpunkt nicht vom Boole-Modell überdeckt wird (vgl. etwa STOYAN, KENDALL UND MECKE (1989), Kapitel 6.2 und 6.3 oder MECKE, SCHNEIDER, STOYAN UND WEIL (1990)).



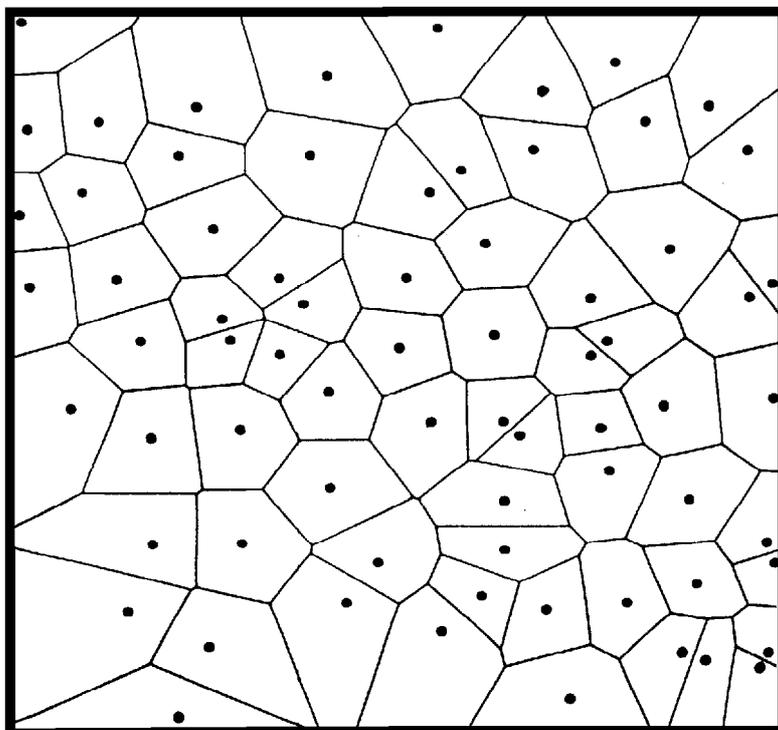
Realisation eines Boole-Modells mit Kreisen

Fig. 3

6. Mosaik

Eine weitere Möglichkeit, aus Punktprozessen zufällige flächige oder räumliche geometrische Objekte zu konstruieren, ergibt sich durch sogenannte *Tesselationen* oder Mosaik. Hierunter versteht man die Aufteilung einer Ebene oder des Raumes in Polygone bzw. Polyhedra. Beispielsweise lassen sich so Bodenrisse in ausgetrockneten Gebieten oder Zellstrukturen modellieren; Hinweise auf eine große Zahl von weiteren Anwendungen in Biologie und Ökologie findet man z.B. in STOYAN, KENDALL UND MECKE (1989), S. 260f. Ein *Poisson-Voronoi-Mosaik* in der Ebene erhält man etwa ausgehend von einem homogenen Poisson-Prozeß, indem man jedem Punkt x des Prozesses ein Polygon zuordnet, das aus denjenigen Punkten des \mathbb{R}^2 besteht, die zu x eine nicht größere

Entfernung haben als zu den anderen Punkten des Prozesses (MECKE, SCHNEIDER, STOYAN UND WEIL (1990), Kapitel 3). Die Kanten des Mosaiks bestehen also aus Abschnitten der Mittelsenkrechten zu den Verbindungsstrecken zwischen benachbarten Punkten des Prozesses.



Realisation eines Poisson-Voronoi-Mosaiks

Fig. 4

Für zufällige Mosaik gibt es inzwischen eine Reihe von Ergebnissen bezüglich der sogenannten *typischen Objekte*, z.B. der Kanten- oder Knotenzahl, des Umfangs oder der Fläche einer typischen Zelle usw. Für Poisson-Voronoi-Mosaik ergibt sich etwa die folgende Verteilung der typischen Kantenzahl K (mit Erwartungswert $E(K) = 6$):

n	3	4	5	6	7	8
$P(K = n)$	0.011	0.107	0.259	0.294	0.199	0.090

nach STOYAN, KENDALL UND MECKE (1989), Table 10.4

Fig. 5

7. Zeitdynamische Punktprozesse

Die bisher betrachteten Modelle sind sämtlich *statisch*, d.h. sie beschreiben "Momentaufnahmen" von zufälligen Punkt- bzw. Objektmustern, ohne eine eventuelle Veränderungsmöglichkeit der Muster in der Zeit zu berücksichtigen. In der Biologie und Ökologie ist aber gerade das Studium von zeitlichen Abhängigkeiten solcher Strukturen von enormer Bedeutung, insbesondere wenn man Prognosen über zukünftige Entwicklungen machen möchte. Ansatzweise sind entsprechende Modelle in der Theorie bei den räumlichen *Geburts- und Todesprozessen* gegeben (vgl. STOYAN, KENDALL UND MECKE (1989), Kapitel 5.5.5, MECKE, SCHNEIDER, STOYAN UND WEIL (1990), Kapitel 4.5 c) oder CRESSIE (1991), S. 678f.). Eine einfache Möglichkeit, Geburten, Tode und Bewegungen von Teilchen in zeitlicher Abhängigkeit simultan zu erfassen, ohne dabei den Rahmen der Poisson-Prozesse zu verlassen, wird in PFEIFER, BÄUMER UND ALBRECHT (1992) vorgeschlagen, z.B. zur Modellierung der Wiederbesiedlung verödeter Wattflächen. Formal läßt sich die Zeitabhängigkeit dabei berücksichtigen, indem man in (2) zusätzlich einen weiteren Parameter t einführt, also Punktprozesse der Form

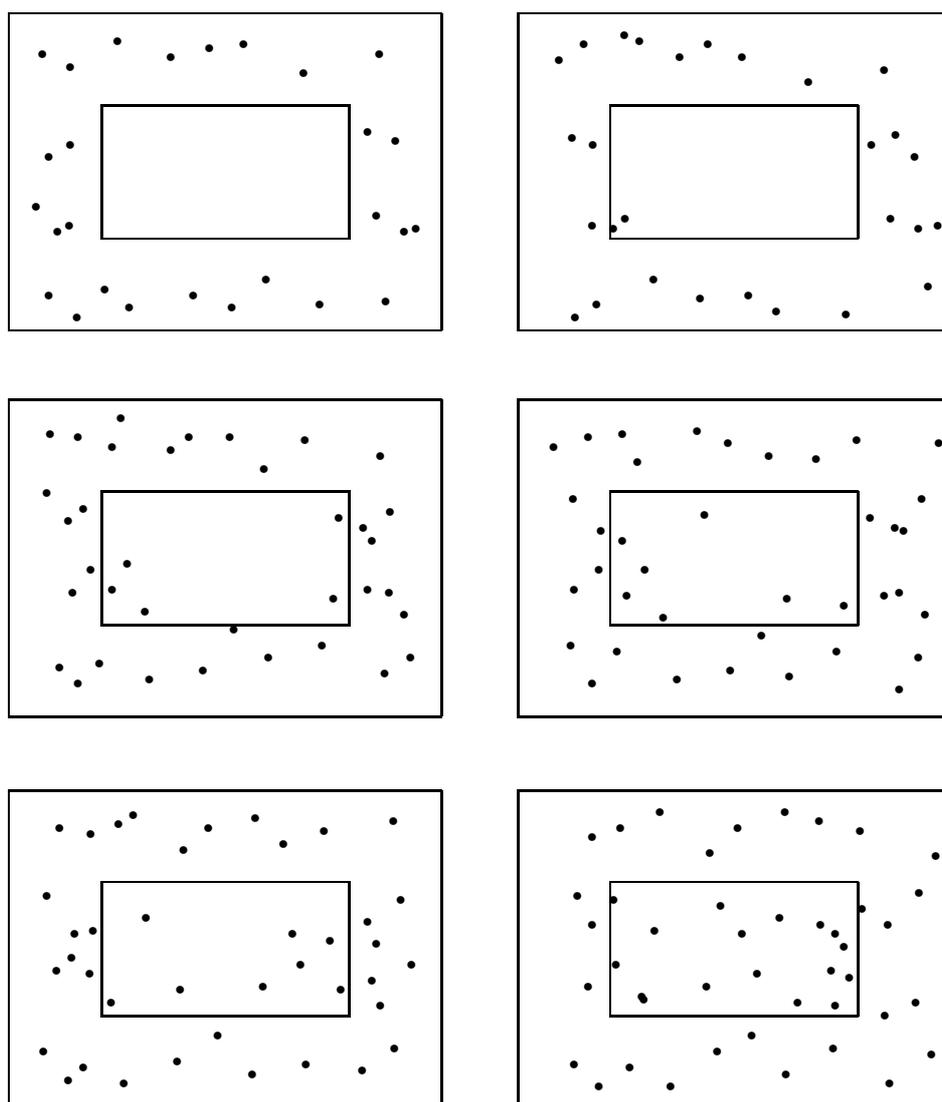
$$\xi_t = \sum_{k=1}^{N(t)} \mathbb{1}_{\{T_k > t\}} \varepsilon_{X_k(t)}, \quad t \geq 0, \quad (6)$$

betrachtet, wobei $\{N(t)\}_{t \geq 0}$ ein eindimensionaler Poisson-Prozeß ist, der die Geburt neuer Teilchen steuert, und $\{T_k\}_{k \in \mathbb{N}}$ eine Familie davon stochastisch unabhängiger Zufallsvariablen, die die Lebensdauern der Teilchen modellieren. Die sogenannte *Indikatorfunktion* $\mathbb{1}_A$ gibt dabei über die Beziehung $\mathbb{1}_A(x) = \varepsilon_x(A)$ die Zugehörigkeit eines Punktes x zu einer Menge A an. Die Zufallsvariablen $X_1(t), X_2(t), \dots$ beschreiben schließlich wie zuvor die Position der Teilchen in der Ebene bzw. im Raum, in Abhängigkeit von der Zeit t .

Bei diesem Modell bildet die zur Zeit t vorhandene Teilchenzahl $M_t = \xi_t(\mathbb{R}^2)$ bzw. $M_t = \xi_t(\mathbb{R}^3)$ einen gewöhnlichen, i.a. nicht-homogenen Geburts- und Todesprozeß mit den Geburts- und Todesraten

$$\beta_n(t) = \lambda(t)(1 - F(t)), \quad \delta_n(t) = \frac{nf(t)}{1 - F(t)}, \quad t \geq 0, \quad n \in \mathbb{Z}^+, \quad (7)$$

wenn $\lambda(t) = \frac{d}{dt} E[N(t)]$ die Intensität des Poisson-Prozesses $\{N(t)\}$ und $f(t) = \frac{d}{dt} F(t)$, $t \geq 0$, die Dichte der Lebensdauerverteilung der Teilchen bezeichnet. Je nach Wahl der Parameter kann in diesem Modell z.B. Auslöschung der Population, Einstellung eines Gleichgewichts oder explosionsartiges Wachstum simuliert werden (PFEIFER, BÄUMER UND ALBRECHT (1992), Abschnitt 3).



Simulation eines Gleichgewichts nach Verödung

Fig. 6

Schlußbemerkung

Ein Teil der hier angesprochenen Modelle wurde im Rahmen des vom *Bundesminister für Forschung und Technologie*, Förderkennzeichen FZ 03F0023F, bewilligten Vorhabens *Ökosystemforschung Niedersächsisches Wattenmeer — Pilotphase — : Beiträge der Angewandten Statistik zur Bearbeitung von Maßstabsfragen und zur Versuchsplanung für die Analyse räumlicher Strukturen und dynamischer Vorgänge im Watt* angewendet und weiterentwickelt.

Literatur

- COX, D.R. UND ISHAM, V. (1980): *Point Processes*. Chapman and Hall, London.
- CRESSIE, N. (1991): *Statistics for Spatial Data*. Wiley, N.Y.
- DALEY, D.J. UND VERE-JONES, D. (1988): *An Introduction to the Theory of Point Processes*. Springer Series in Statistics. Springer, N.Y.
- MATHAR, R. UND PFEIFER, D. (1990): *Stochastik für Informatiker*. Leitfäden und Monographien der Informatik. Teubner, Stuttgart.
- MECKE, J., SCHNEIDER, R.G., STOYAN, D. UND WEIL, R.R. (1990): *Stochastische Geometrie*. DMV-Seminar, Bd. 16. Birkhäuser, Basel.
- PFEIFER, D., BÄUMER, H.-P. UND ALBRECHT, M. (1992): Point processes and their applications to biology and ecology. *Modeling Geo-Biosphere Processes* 1, 145 – 161.
- PFEIFER, D., BÄUMER, H.-P. UND ALBRECHT, M. (1993): Moving point patterns — the Poisson case. In: O. Opitz and B. Lausen (eds.): *Information and Classification: Concepts, Methods and Applications*. Studies in Classification, Data Analysis, and Knowledge Organization, vol. 3, 1993, 248 – 257. Springer, N.Y.
- RICHTER, O. UND SÖNDGERATH, D. (1990): *Parameter Estimation in Ecology*. The Link between Data and Models. VCH Verlagsgesellschaft, Weinheim.
- RIPLEY, B.D. (1981): *Spatial Statistics*. Wiley, N.Y.
- RIPLEY, B.D. (1988): *Statistical Inference for Spatial Processes*. Cambridge Univ. Press, Cambridge.
- STOYAN, D., KENDALL, W.S. UND MECKE, J. (1989): *Stochastic Geometry and Its Applications*. Wiley, N.Y.
- TOPSØE, F. (1990): *Spontane Phänomene*. Stochastische Modelle und ihre Anwendungen. Vieweg, Braunschweig u. Wiesbaden.