

Springer-Verlag Berlin Heidelberg GmbH

Marc-Thorsten Hütt

Datenanalyse in der Biologie

Eine Einführung in Methoden
der nichtlinearen Dynamik, fraktalen
Geometrie und Informationstheorie

Mit 195 Abbildungen und 17 Tabellen



Springer

Dr. MARC-THORSTEN HÜTT
Institut für Botanik
Technische Universität Darmstadt
Schnittspahnstraße 3-5
64287 Darmstadt

ISBN 978-3-540-42311-9

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Hütt, Marc-Thorsten:

Datenanalyse in der Biologie : eine Einführung in Methoden der nichtlinearen
Dynamik, fraktalen Geometrie und Informationstheorie / Marc Thorsten Hütt. -

Berlin ; Heidelberg ; New York ; Barcelona ; Hongkong ; London ; Mailand ;
Paris ; Singapur ; Tokio : Springer, 2001

(Springer-Lehrbuch)

ISBN 978-3-540-42311-9 ISBN 978-3-642-56631-8 (eBook)

DOI 10.1007/978-3-642-56631-8

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, des Vortrags, der Entnahme von Abbildungen und Tabellen, der Funksendung, der Mikroverfilmung oder der Vervielfältigung auf anderen Wegen und der Speicherung in Datenverarbeitungsanlagen, bleiben auch bei nur auszugsweiser Verwertung vorbehalten. Eine Vervielfältigung dieses Werkes oder von Teilen dieses Werkes ist auch im Einzelfall nur in den Grenzen der gesetzlichen Bestimmungen des Urheberrechtsgesetzes der Bundesrepublik Deutschland vom 9. September 1965 in der jeweils geltenden Fassung zulässig. Sie ist grundsätzlich vergütungspflichtig. Zuwiderhandlungen unterliegen den Strafbestimmungen des Urheberrechtsgesetzes.

<http://www.springer.de>

© Springer-Verlag Berlin Heidelberg 2001

Ursprünglich erschienen bei Springer-Verlag Berlin Heidelberg New York 2001

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, daß solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Produkthaftung: Für Angaben über Dosierungsanweisungen und Applikationsformen kann vom Verlag keine Gewähr übernommen werden. Derartige Angaben müssen vom jeweiligen Anwender im Einzelfall anhand anderer Literaturstellen auf ihre Richtigkeit überprüft werden.

Satz: Druckfertige Vorlagen des Autors

Einbandgestaltung: de'blik Graphische Gestaltung, Berlin

Einbandabbildungen: Verteilung von Chlorophyllfluoreszenzaktivität auf einem Pflanzenblatt
(Erläuterungen s. Abb. 5.1); Strichzeichnung Abb. 2.11

SPIN 10842365 29/3130 - 5 4 3 2 1 0 - Gedruckt auf säurefreiem Papier

*R. Weiershausen für einen jahrelangen,
fortdauernden interdisziplinären Dialog*

Vorwort

Die Entwicklung neuer experimenteller Techniken und die Beschäftigung mit immer komplexeren Fragestellungen hat zur Folge, daß oft herkömmliche Methoden der Datenauswertung (Betrachtung von Mittelwert und Varianz, Signifikanztests, etc.) bei weitem nicht mehr ausreichen, um eine angemessene Interpretation der Daten zu erreichen. Immer stärker werden daher zur Analyse und Diskussion experimenteller biologischer Daten Methoden aus Nachbardisziplinen herangezogen, vor allem aus der theoretischen Physik, der Informationstheorie und der Theorie der Bildverarbeitung.

Das vorliegende Buch ist eine pragmatisch formulierte Einführung, die keine umfassende und mathematisch vollständige Darstellung in die behandelten Themen anstrebt. Vielmehr sollen hier die für Experimentatoren wissenswerten und brauchbaren Aspekte zusammengetragen werden. Es wurde dennoch versucht, soviel von dem theoretischen Hintergrund der verschiedenen Methoden darzustellen, daß sich für den Anwender eine gewisse thematische Sicherheit, eine Vorstellung der grundlegenden Zusammenhänge und eine klare Kenntnis der methodischen Grenzen ergibt.

Das wesentliche Ziel ist, die von theoretischer Seite zur Verfügung stehenden Methoden der Datenanalyse zu präsentieren und dem Benutzer zugänglich zu machen. Im Idealfall gelangt der Leser zu einer neuen Sicht auf seine Daten und damit vielleicht zu einem neuen Verständnis der in ihnen codierten Funktionsweise des betrachteten Systems. Viele technische Aspekte, die etwa bei der Computer-Implementierung einer Analysemethode von Bedeutung sind, konnten hier nur beispielhaft behandelt werden. Entsprechende Literaturhinweise und die Angabe von existierenden Softwarepaketen sollen diese Unterlassung auffangen.

Vor allem in den Kapiteln 2.2 und 2.3 scheint die thematische Auswahl recht willkürlich. Ich habe dabei versucht, Aspekte in den Vordergrund zu stellen, die einem experimentell arbeitenden Wissenschaftler einen Einblick in die Denkweise von Theoretikern und Modellierern geben

können und ein Gefühl zu vermitteln vermögen, wie komplexe dynamische Phänomene aus relativ elementaren Gleichungen entstehen können. Schon aus diesem Grund liegt ein deutlicher Schwerpunkt auf Differenzgleichungen und weniger auf einer umfassenden Diskussion von (mehrdimensionalen) Differentialgleichungen.

Eine weitere Hervorhebung ist begründenswert: Zelluläre Automaten, die hier als eine Verallgemeinerung von Differenzgleichungen dargestellt werden (Kapitel 3), gehören heute nicht mehr notwendigerweise zum Unterrichtskanon von Kursen über Musterbildung und Modellierung. Eine häufige Begründung ist, daß die Regelwerke zellulärer Automaten im wesentlichen eine Approximation an partielle Differentialgleichungen sind und daher auch in der mathematisch fortgeschritteneren Weise behandelt werden können. Wenn das Ziel jedoch ein praxisorientierter Umgang mit den mathematischen Werkzeugen ist, läßt sich gegen dieses Argument einiges einwenden. Gerade in der Biologie ist es von Vorteil, die lokalen Regeln der Selbstorganisation in einer unverschlüsselten Form zu kennen und ohne größeren Aufwand die resultierenden globalen Dynamiken ermitteln zu können. Bei meinen eigenen Arbeiten habe ich feststellen können, daß die Akzeptanz dieser Darstellung sehr viel größer ist als etwa die eines Systems von nichtlinearen Differentialgleichungen, was sich einem fachübergreifenden wissenschaftlichen Dialog als äußerst zuträglich erweist. Eine Strategie dieses Buches liegt darin, bestimmte Formalismen und Rechentechniken erst dann einzuführen, wenn sie inhaltlich benötigt werden (z.B. das Konzept komplexer Zahlen in Kapitel 4.1, das Ising-Modell in Kapitel 5.2 oder die Matrixmultiplikation in Kapitel 6.1). Auf diese Weise sieht man die allgemeine Idee in unmittelbarer Verbindung mit der Anwendung, und die in mathematische Formalismen einführenden Kapitel bleiben frei für die Schilderung der übergreifenden Aspekte dynamischer Systeme. Um den Zugang zu (hauptsächlich englischsprachigen) weiterführenden Texten zu erleichtern, habe ich zu einigen Schlüsselbegriffen im Text auch die englischen Bezeichnungen angegeben, was gerade bei einer Suche in Bibliothekskatalogen und Internet-Suchmaschinen eine Hilfe darstellen kann.

Ich danke Andreas Bohn für seine inhaltliche Unterstützung bei der Ausarbeitung der Passagen über Fourieranalyse (in Kapitel 4.2) und Bernd Blasius für die Überlassung der Abbildungen 4.28 und 4.30-4.33 aus seiner Dissertation (Blasius 1997). Die experimentellen Daten zum Crassulaceen-Säurestoffwechsel wurden mir von Uwe Rascher zur

Verfügung gestellt und durch lange und weitreichende Diskussionen nahegebracht. Beim Abfassen des Kapitels über fraktale Geometrie (speziell der Kapitel 6.2 und 6.3) habe ich von intensiven Diskussionen mit Larry Liebovitch profitiert, der mir auch das Material für die Abbildungen 6.31 und 6.33 zugänglich gemacht hat. An vielen Stellen haben mir die Diskussionen mit Roland Neff geholfen, der mir auch bei der Analyse des Forest-Fire-Modells und des Ising-Modells behilflich war. Die Abbildungen 2.32, 2.31 und 2.30, sowie die Darstellung der Beschreibung von Insektenplagen in Kapitel 2.2 sind in Anlehnung an das Buch von Steve Strogatz (Strogatz 1994) abgefaßt. Die Einführung zellulärer Automaten über allgemeine boolesche Netzwerke orientiert sich an dem Buch von Daniel Kaplan und Leon Glass (Kaplan u. Glass 1995). Weitere Bücher und Übersichtsartikel waren mir als Hintergrundmaterial ein unverzichtbares Hilfsmittel. Eine Auswahl ist in Anhang B angegeben.

Alle mathematisch orientierten Abbildungen wurden mit dem Softwarepaket *Mathematica*® von Wolfram Research Inc. erstellt und in *Adobe Illustrator*® nachbearbeitet. Die schematischen Darstellungen sind direkt in *Adobe Illustrator*® entworfen worden. Aus externen Quellen übernommene Abbildungen sind in der Bildunterschrift durch weitere Angaben kenntlich gemacht. Als Satzprogramm wurde L^AT_EX verwendet.

Einen großen Anteil am Zustandekommen dieses Buches haben die Teilnehmenden meiner Veranstaltungen zur Datenanalyse im Wintersemester 1999/2000 an der TU Darmstadt, insbesondere Harald Bierbaum, Jasmin Grünig, Sven Hinderlich, Klaus Höhne, Thomas Mohr, Isabelle Philipp, Maike Rothermel und Christina Tritt. Bei der Gestaltung, der Erstellung zahlreicher Abbildungen und bei vielen Aspekten der EDV-Arbeit war mir Jasmin Grünig eine enorme Hilfe. Ihr sei an dieser Stelle herzlich gedankt.

Dem Team vom Springer-Verlag, vor allem Iris Lasch-Petersmann und Stefanie Wolf, möchte ich für das Engagement danken, mit dem die Publikation vorangetrieben wurde.

Das Buch hat sehr gewonnen durch die engagierte Mitarbeit von Christiane Hilgardt beim Erstellen der Endfassung. Für ihre Unterstützung und ihr Interesse an den Inhalten möchte ich ihr ganz herzlich danken.

Friedrich Beck und Friedemann Kaiser danke ich für die gemeinsame spannende Forschung zur Theorie und Anwendung der nichtlinea-

ren Dynamik. Ihre Sichtweisen der theoretischen Physik hatten einen großen Einfluß auf die vorliegende Darstellung.

In hohem Maße profitiert habe ich von Diskussionen mit Ulrich Lüttge, der mein Verständnis von interdisziplinärer Arbeit geprägt hat.

Das Projekt wurde finanziell unterstützt vom DFG-Graduiertenkolleg 340 "Kommunikation in biologischen Systemen" an der TU Darmstadt.

Darmstadt, im Mai 2001

Marc-Thorsten Hütt

Inhaltsverzeichnis

1. Mathematische Grundlagen	1
1.1 Terminologie und Ziele	1
1.2 Einige mathematische Werkzeuge	9
1.2.1 Kurvendiskussion und Eigenschaften spezieller Funktionen	10
1.2.2 Wahrscheinlichkeiten und Elemente der Statistik	22
1.2.3 Anpassung von mathematischen Kurven an ex- perimentelle Daten	26
1.3 Gewöhnliche Differentialgleichungen	33
2. Grundbegriffe der nichtlinearen Dynamik	43
2.1 Differenzgleichungen und Iterationsmethoden	43
2.1.1 Konzept der Differenzgleichung	43
2.1.2 Nichtlineare Differenzgleichungen	48
2.1.3 Fixpunkte und ihre Stabilität	49
2.1.4 Zyklen und ihre Stabilität	52
2.1.5 Chaos in Differenzgleichungen	57
2.2 Eindimensionale Differentialgleichungen und Bifurka- tionen	69
2.3 Mehrdimensionale Differentialgleichungen	87
3. Zelluläre Automaten	103
3.1 Grundidee zellulärer Automaten und Begriff des Netz- werks	103
3.2 Methoden zum Erstellen von Regelwerken	111
3.3 Wichtige Anwendungen	116
3.4 Exkurs: Membranmodellierung mit Hilfe zellulärer Au- tomaten	130

4. Elemente der nichtlinearen Zeitreihenanalyse	137
4.1 Vorbemerkung	137
4.2 Lineare Zeitreihenanalyse und ihre Grenzen	142
4.3 Nichtlineare Zeitreihenanalyse und Vorhersagbarkeit ei- ner Zeitreihe	158
4.4 Anwendungsbeispiele	177
4.5 Methodische Schwierigkeiten	184
4.6 Exkurs: Kontrolle von Analysemethoden durch Surro- gatdaten	188
5. Analyse raumzeitlicher Strukturen	191
5.1 Problemstellung	191
5.2 Raumzeitliche Verallgemeinerungen zeitlicher Methoden	194
5.3 Methoden auf der Grundlage zellulärer Automaten	204
5.4 Anwendungsbeispiele	211
6. Selbstähnlichkeit und fraktale Geometrie	219
6.1 Selbstähnlichkeit als Ordnungsprinzip	219
6.2 Fraktale Datenanalyse	240
6.3 Anwendungsbeispiele	252
6.4 Exkurs: Fraktalität in der Dynamik von Ökosystemen . .	255
7. Methoden aus der Informationstheorie	267
7.1 Grundbegriffe: Entropie und Transinformation	267
7.2 Anwendungsbeispiele	272
A. Software-Pakete und Internet-Datenbanken	279
B. Weiterführende Literatur zu ausgewählten Themen .	285
C. Übungsaufgaben	291
Literaturverzeichnis	301
Sachwortverzeichnis	307