

# SPATIAL CLUSTER MODELLING

Edited by  
**Andrew B. Lawson**  
**David G.T. Denison**



CHAPMAN & HALL/CRC

0212.4  
S738

---

# SPATIAL CLUSTER MODELLING

---

Edited by

**Andrew B. Lawson**

*Department of Mathematical Sciences  
University of Aberdeen  
Aberdeen, UK*



**David G.T. Denison**

*Department of Mathematics  
Imperial College of Science, Technology and Medicine  
London, UK*



E200301645



**CHAPMAN & HALL/CRC**

---

A CRC Press Company

Boca Raton London New York Washington, D.C.

## Library of Congress Cataloging-in-Publication Data

Lawson, Andrew (Andrew B.)

Spatial cluster modeling / edited by Andrew B. Lawson, David G.T. Denison.  
p. cm.

Includes bibliographical references and index.

ISBN 1-58488-266-2 (alk. paper)

1. Cluster analysis. 2. Spatial analysis. I. Denison, David G. T. II. Title.

QA278 .L39 2002

519.5'3—dc21

2002019297

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the authors and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

Neither this book nor any part may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, microfilming, and recording, or by any information storage or retrieval system, without prior permission in writing from the publisher.

All rights reserved. Authorization to photocopy items for internal or personal use, or the personal or internal use of specific clients, may be granted by CRC Press LLC, provided that \$1.50 per page photocopied is paid directly to Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923 USA. The fee code for users of the Transactional Reporting Service is ISBN 1-58488-266-2/02/\$0.00+\$1.50. The fee is subject to change without notice. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

The consent of CRC Press LLC does not extend to copying for general distribution, for promotion, for creating new works, or for resale. Specific permission must be obtained in writing from CRC Press LLC for such copying.

Direct all inquiries to CRC Press LLC, 2000 N.W. Corporate Blvd., Boca Raton, Florida 33431.

**Trademark Notice:** Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation, without intent to infringe.

**Visit the CRC Press Web site at [www.crcpress.com](http://www.crcpress.com)**

© 2002 by Chapman & Hall/CRC

No claim to original U.S. Government works

International Standard Book Number 1-58488-266-2

Library of Congress Card Number 2002019297

Printed in the United States of America 1 2 3 4 5 6 7 8 9 0

Printed on acid-free paper

---

# SPATIAL CLUSTER MODELLING

---

---

## List of Contributors

---

<b>Adrian Baddeley</b>	University of Western Australia
<b>Dankmar Böhning</b>	Free University of Berlin
<b>Simon D. Byers</b>	AT&T Labs, New Jersey
<b>Allan B. Clark</b>	University of Aberdeen
<b>Murray K. Clayton</b>	University of Wisconsin, Madison
<b>David G.T. Denison</b>	Imperial College, London
<b>José Tomé A.S. Ferreira</b>	Imperial College, London
<b>Jürgen Gallinat</b>	Free University of Berlin
<b>Ronald B. Gangnon</b>	University of Wisconsin, Madison
<b>Fred Godtliebsen</b>	University of Tromsø
<b>Christopher M. Hans</b>	Duke University
<b>Christopher C. Holmes</b>	Imperial College, London
<b>John T. Kent</b>	University of Leeds
<b>Hyoung Moon Kim</b>	Texas A&M University
<b>Andrew B. Lawson</b>	University of Aberdeen
<b>Marc Loizeaux</b>	Florida State University, Tallahassee
<b>Ian W. McKeague</b>	Florida State University, Tallahassee

<b>Bani K. Mallick</b>	Texas A&M University
<b>Kanti V. Mardia</b>	University of Leeds
<b>J. Steve Marron</b>	University of North Carolina, Chapel Hill
<b>Jesper Møller</b>	Aalborg University
<b>Stephen M. Pizer</b>	University of North Carolina, Chapel Hill
<b>Adrian E. Raftery</b>	University of Washington, Seattle
<b>Peter Schlattmann</b>	Free University of Berlin
<b>David A. van Dyk</b>	Harvard University
<b>Marie-Colette van Lieshout</b>	CWI, Amsterdam
<b>Rasmus P. Waagepetersen</b>	Aalborg University
<b>Christopher K. Wikle</b>	University of Missouri–Columbia

---

# Preface

---

The development of statistical methodology for the analysis of spatial data has seen considerable advances since the publication of the seminal work of Cressie (1993). In particular, the development of fast computational algorithms for sampling of complex Bayesian models (most notably Markov chain Monte Carlo algorithms) has allowed a wide range of problems to be addressed which hitherto could not be directly analysed. Many spatial problems can be considered within the paradigm of hierarchical Bayesian modelling and so the emphasis within this volume will lie within that area.

The aim of this volume is not to present a general review of spatial statistical modelling but rather to focus on the area of spatial *cluster* modelling. Hence the theme of this work is the highlighting of the diverse approaches to the definition of clusters and clustering in space (and its adjunct space-time), and to present state-of-the-art coverage of the diverse modelling approaches which are currently available. In Chapter 1 we provide a brief historical introduction to the subject area and, in particular, compare conventional and spatial clustering. In addition this chapter introduces the notation and different areas of study explored. After this initial chapter the volume is split into 3 parts, each relating to a specific area of cluster modelling. Part I deals with point and object process modelling, Part II involves spatial process modelling, while Part III contains papers relating to spatio-temporal models.

One of the features of modelling spatial data is the need to use fast computational algorithms to be able to evaluate the complex posterior distributions or likelihood surfaces which arise in spatial applications. The 1980s saw the development of Markov chain Monte Carlo algorithms based on the Gibbs and Metropolis-Hastings samplers, and witnessed rapid development of models for complex spatial problems. Not only could existing models be sampled from but newer more sophisticated models could also be developed and applied. Often these models are of a hierarchical form so this naturally leads to the Bayesian paradigm being of importance in a great deal of the work.

As the potential fields of application for spatial methods are so wide we cannot hope to cover all of them. Nevertheless the chapters here do make reference to data in astrophysics (Chapter 10), spatial epidemiology (Chapters 5,7,8,14), ecology (Chapters 4,11), imaging (Chapter 13), ge-

ology and the geosciences (Chapters 6,4,7,9,12). In addition, the volume provides a useful insight into the current issues and methodology used for spatial cluster modelling. We have specifically included the burgeoning area of spatio-temporal modelling as an important extension to standard spatial data analysis and Chapters 12,13,14 specifically deal with this topic.

Finally we would like to thank all the contributors for their timely and thoughtful articles. In addition, we acknowledge the help of the staff at CRC Press, in particular Kirsty Stroud, Jasmin Naim and Helena Redshaw for their continued support and encouragement in the production of this volume.



---

# Contents

---

<b>List of Contributors</b>	<b>xi</b>
<b>Preface</b>	<b>xiii</b>
<b>1 Spatial Cluster Modelling: An Overview</b>	<b>1</b>
1.1 Introduction	1
1.2 Historical Development	3
1.2.1 Conventional Clustering	6
1.2.2 Spatial Clustering	8
1.3 Notation and Model Development	12
1.3.1 Nonparametric Approaches	13
1.3.2 Point or Object Process Modelling	15
1.3.3 Random Effect Modelling	16
1.3.4 Partition Modelling	18
1.3.5 Spatio-Temporal Process Modelling	18
<b>I Point process cluster modelling</b>	<b>21</b>
<b>2 Significance in Scale-Space for Clustering</b>	<b>23</b>
2.1 Introduction	23
2.2 Overview	24
2.3 New Method	29
2.4 Future Directions	35
<b>3 Statistical Inference for Cox Processes</b>	<b>37</b>
3.1 Introduction	37
3.2 Poisson Processes	39
3.3 Cox Processes	41
3.4 Summary Statistics	43
3.5 Parametric Models of Cox Processes	45
3.5.1 Neyman-Scott Processes as Cox Processes	45
3.5.2 Log Gaussian Cox Processes	48
3.5.3 Shot Noise G Cox Processes	49
3.6 Estimation for Parametric Models of Cox Processes	51

3.7	Prediction	54
3.7.1	Conditional Simulation for Neyman-Scott Processes	55
3.7.2	Conditional Simulation for LGCPs	55
3.7.3	Conditional Simulation for Shot-noise G Cox Processes	56
3.8	Discussion	58
<b>4</b>	<b>Extrapolating and Interpolating Spatial Patterns</b>	<b>61</b>
4.1	Introduction	61
4.2	Formulation and Notation	62
4.2.1	Germ-grain Models	63
4.2.2	Problem Statement	63
4.2.3	Edge Effects and Sampling Bias	64
4.2.4	Extrapolation	65
4.3	Spatial Cluster Processes	66
4.3.1	Independent Cluster Processes	67
4.3.2	Cox Cluster Processes	68
4.3.3	Cluster Formation Densities	69
4.4	Bayesian Cluster Analysis	72
4.4.1	Markov Point Processes	72
4.4.2	Sampling Bias for Independent Cluster Processes	74
4.4.3	Spatial Birth-and-Death Processes	75
	Example: Redwood Seedlings	76
4.4.4	Parameter Estimation	78
	Example: Cox-Matérn Cluster Process	80
4.4.5	Adaptive Coupling from the Past	80
4.4.6	Example: Cox-Matérn Cluster Process	84
4.5	Summary and Conclusion	86
<b>5</b>	<b>Perfect Sampling for Point Process Cluster Modelling</b>	<b>87</b>
5.1	Introduction	87
5.2	Bayesian Cluster Model	89
5.2.1	Preliminaries	89
5.2.2	Model Specification	90
5.3	Sampling from the Posterior	93
5.4	Specialized Examples	95
5.4.1	Neyman-Scott Model	95
5.4.2	Pure Silhouette Models	98
5.5	Leukemia Incidence in Upstate New York	100
5.6	Redwood Seedlings Data	106
<b>6</b>	<b>Bayesian Estimation and Segmentation of Spatial Point Processes Using Voronoi Tilings</b>	<b>109</b>
6.1	Introduction	109

6.2	Proposed Solution Framework	110
6.2.1	Formulation	110
6.2.2	Voronoi Tilings	111
6.2.3	Markov Chain Monte Carlo Using Dynamic Voronoi Tilings	111
6.3	Intensity Estimation	112
6.3.1	Formulation	112
6.3.2	MCMC Implementation	112
	Fixed Number of Tiles	113
	Variable Number of Tiles	114
6.4	Intensity Segmentation	115
6.4.1	Formulation	115
	Fixed Number of Tiles	115
	Variable Number of Tiles	117
6.5	Examples	117
6.5.1	Simulated Examples	117
	A Sine Wave	117
	A Linear Feature	118
6.5.2	New Madrid Seismic Region	118
6.6	Discussion	119
<b>II</b>	<b>Spatial process cluster modelling</b>	<b>123</b>
<b>7</b>	<b>Partition Modelling</b>	<b>125</b>
7.1	Introduction	125
7.2	Partition Models	126
7.2.1	Partitioning for Spatial Data	127
7.2.2	Bayesian Inference	128
7.2.3	Predictive Inference	131
7.2.4	Markov Chain Monte Carlo Simulation	132
7.2.5	Partition Model Prior	133
7.3	Piazza Road Dataset	135
7.4	Spatial Count Data	135
7.4.1	The Poisson-Gamma Model for Disease Mapping	137
7.4.2	Disease Mapping with Covariates	138
7.4.3	East German Lip Cancer Dataset	141
7.5	Discussion	144
7.6	Further Reading	144
<b>8</b>	<b>Cluster Modelling for Disease Rate Mapping</b>	<b>147</b>
8.1	Introduction	147
8.2	Statistical Model	148
8.3	Posterior Calculation	150
8.4	Example: U.S. Cancer Mortality Atlas	153

8.4.1	Breast Cancer	154
8.4.2	Cervical Cancer	155
8.4.3	Colorectal Cancer	155
8.4.4	Lung Cancer	159
8.4.5	Stomach Cancer	159
8.5	Conclusions	159
<b>9</b>	<b>Analyzing Spatial Data Using Skew-Gaussian Processes</b>	<b>163</b>
9.1	Introduction	163
9.2	Skew-Gaussian Processes	164
9.2.1	The Model	165
9.2.2	Bayesian Analysis	166
9.2.3	Computational Strategy	167
9.3	Real Data Illustration: Spatial Potential Data Prediction	168
9.4	Discussion	171
<b>10</b>	<b>Accounting for Absorption Lines in Images Obtained with the Chandra X-ray Observatory</b>	<b>175</b>
10.1	Statistical Challenges of the Chandra X-ray Observatory	175
10.2	Modeling the Image	179
10.2.1	Model-Based Spatial Analysis	179
10.2.2	Model-Based Spectral Analysis	183
10.3	Absorption Lines	184
10.3.1	Scientific Background	184
10.3.2	Statistical Models	185
10.3.3	Model Fitting	188
10.3.4	A Simulation-Based Example	190
10.4	Spectral Models with Absorption Lines	192
10.4.1	Combining Models and Algorithms	192
10.4.2	An Example	195
10.5	Discussion	196
<b>11</b>	<b>Spatial Modelling of Count Data: A Case Study in Modelling Breeding Bird Survey Data on Large Spatial Domains</b>	<b>199</b>
11.1	Introduction	199
11.2	The Poisson Random Effects Model	200
11.2.1	Spectral Formulation	202
11.2.2	Model Implementation and Prediction	204
	Selected Full-Conditional Distributions	205
	Prediction	206
	Implementation	206
11.3	Results	206
11.4	Conclusion	207

<b>III Spatio-temporal cluster modelling</b>	<b>211</b>
<b>12 Modelling Strategies for Spatial-Temporal Data</b>	<b>213</b>
12.1 Introduction	213
12.2 Modelling Strategy	214
12.3 D-D (Drift-Drift) Models	215
12.4 D-C (Drift-Correlation) Models	220
12.5 C-C (Correlation-Correlation) Models	222
12.6 A Unified Analysis on the Circle	224
12.7 Discussion	225
<b>13 Spatio-Temporal Partition Modelling: An Example from Neurophysiology</b>	<b>227</b>
13.1 Introduction	227
13.2 The Neurophysiological Experiment	227
13.3 The Linear Inverse Solution	228
13.4 The Mixture Model	229
13.4.1 Initial Preparation of the Data	229
13.4.2 Formulation of the Mixture Model	230
13.5 Classification of the Inverse Solution	232
13.6 Discussion	234
<b>14 Spatio-Temporal Cluster Modelling of Small Area Health Data</b>	<b>235</b>
14.1 Introduction	235
14.2 Basic Cluster Modelling approaches	235
14.2.1 Case Event Data Models	236
Inter-Event versus Hidden Process Modelling	236
14.2.2 Small Area Count Data Models	239
14.2.3 Spatio-Temporal Extensions to Cluster Models	239
14.3 A Spatio-Temporal Hidden Process Model	240
14.4 Model Development	240
14.4.1 Estimation of $\mathbf{g}(\mathbf{s}, t)$	243
The Prior Distribution for Cluster Centres	244
Choice of Cluster Distribution Function	245
Other Prior Distributions and the Posterior Distribution	245
14.4.2 Region Counts	246
Integrated Intensity	247
14.5 The Posterior Sampling Algorithm	248
14.5.1 Goodness-of-Fit Measures for the Model	249
14.6 Data Example: Scottish Birth Abnormalities	249
14.6.1 Introduction	249
14.6.2 Exploratory Analysis	250

14.6.3 Model Fitting Results	252
14.7 Discussion	256
<b>References</b>	<b>259</b>
<b>Index</b>	<b>277</b>
<b>Author Index</b>	<b>281</b>

---

## CHAPTER 1

# Spatial Cluster Modelling: An Overview

---

*A.B. Lawson     D.G.T. Denison*

### 1.1 Introduction

When analysing spatial data one is often interested in detecting deviations from the expected. For instance, we may be interested in the answers to questions like, “Is there an unusual aggregation of leukemia cases around a nuclear power station ?” or, “Where is it likely that the air pollution level is above the legally allowed limit ?”. In both cases the focus is on finding regions in (usually two-dimensional) space in which higher than expected counts, or readings, are observed. We shall call such areas *clusters* and determining their nature forms the focus of this work.

This volume brings together a collection of papers on the topic of spatial cluster modelling and gives descriptions of various approaches which begin to solve the problem of detecting clusters. The papers are statistical in nature but draw on results in other fields as diverse as astrophysics, medical imaging, ecology and environmental engineering.

Two examples of the sort of spatial processes that we shall consider here are displayed in Figs. 1.1-1.2. Fig. 1.1 is an example of a point process, where each dot is an “event” (in this case the occurrence of a cancer). Here it is of interest to determine whether the cases are more aggregated, or *clustered*, than expected and whether the clustering relates to the locations of any possible pollution sources. To assess this a background control disease map (which is not shown) is often used to represent the expected variation in the distribution of cases; this is often a function of the relative population density.

Fig. 1.2 is an example of a dataset that consists of observations of an underlying spatial process at a number of locations. The usual aim of an analysis of this type of data is to determine the value of the spatial process at all the locations in the domain of interest, assuming that each measurement is only observed in the presence of a random error component. However, we may also be interested in determining areas where the process is above some predefined limit or even, in some sense, above average.

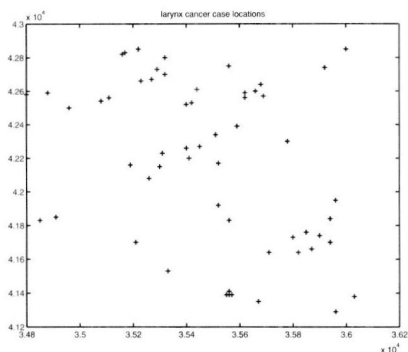


Figure 1.1 *An example of a point process: The Larynx cancer case event map (Diggle 1990) relating to cases in Lancashire, UK from 1974 to 1983.*

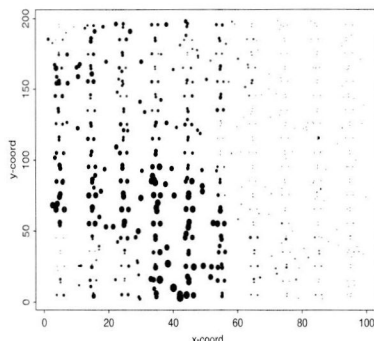


Figure 1.2 *An example of observations of a spatial process: The Piazza Road data (Higdon et al. 1999). Each dot represents a location where a measurement of the dioxin concentration in an area around the Piazza Road is taken. The size of the dots gives an indication of the observed measurements, with larger dots representing higher concentrations.*

Spatio-temporal data will also be looked at in this volume. In these examples either measurements or point processes are observed over time and similar questions arise but now clusters can occur in space, in time, or even in space and time jointly.

The analysis of spatial clustering has had a varied history with developments often resulting from particular applications, and so these developments have related to varying interest over time in different applications. For example, much work in the 1960s and 1970s on clustering developed from ecology applications (e.g. Diggle 1983, Ripley 1981), whereas an increased interest in image analysis in the 1980s led to associated advances in image segmentation and object recognition (e.g. Besag et al. 1991). Increased public and scientific interest in environmental hazards and health in the late 1980s and 1990s, has led to increased emphasis on cluster detection in small area health studies (Lawson et al. 1999, Elliott et al. 1999, Lawson 2001). The context of this historical development in relation to methodological progress is discussed more fully in the next section.



## 1.2 Historical Development

The analysis of clustering has a long history in statistical science. In one dimension the analysis of aggregation of data around a preferred location is at the heart of much statistical work, whether the focus is on the mean tendency of the aggregation of data or on its spread or variance. In addition in cluster studies the location of the maximum aggregation may also be of interest (modal property). In two dimensions, the natural extension of these ideas is to a two-dimensional aggregation, perhaps around a single point. In this case the centre of the aggregation may be defined either by mean or modal properties while the variance or spread of the aggregation can be defined around the putative centre which is now a two-dimensional location. In the case of spatial data these quantities have obvious interpretations as cluster centre location and cluster spread.

In the spatial domain, it is possible to view a clustered pattern in different ways depending on the focus of the analysis. First, it may be possible to conceive of the pattern as the realisation of a random process which produces aggregations as a result of the global structure of the process, whereby a small number of parameters control the scale and frequency of aggregations but the defined process does not parameterise the *locations* of the aggregations. This is akin to the geostatistical view of random processes, where the intensity or local density of events is defined by, for example, a spatial Gaussian process. The peaks of this process would correspond with local aggregations, but no parameterisation of the locations is made. Recent examples of this approach can be found in Cressie (1993) and Diggle et al. (1998). Essentially, this approach regards the aggregations as produced by random effects which are governed by global model parameters, i.e. the degree of aggregation and spread of the aggregations would be controlled by a small number of global parameters. This form of random effect modelling is at the heart of much hierarchical Bayesian modelling in this context, and in the literature of spatial applications the term *clustering* or *cluster modelling* is used to refer to such random effect modelling. An example from small area health studies is Clayton and Bernardinelli (1992).

The second approach to the modelling of clusters is to include within the modelled process specific elements which relate to cluster location and how these locations relate to the surrounding data. Much of the early work in stochastic geometry relating to point processes examined clustering of point processes and models such as the Neyman-Scott and Cox cluster processes. The fundamental feature of these processes was the definition of a set of cluster centres around which the *offspring* (data) lie. The term offspring comes from the idea that the clustering could arise from a multigenerational process. The variation in the local aggregation of data is thought to be summarised by the local density of events around the cluster centre set. In the case of a Neyman-Scott or Poisson cluster process the local density is