

Spatial Analysis in Epidemiology

Spatial Analysis in Epidemiology

Dirk U. Pfeiffer

Epidemiology Division, Royal Veterinary College, University of London,
United Kingdom

Timothy P. Robinson

Food and Agricultural Organization of the United Nations, Italy

Mark Stevenson

Epicentre, Institute of Veterinary, Animal and Biomedical Sciences,
Massey University, New Zealand

Kim B. Stevens

Epidemiology Division, Royal Veterinary College, University of London,
United Kingdom

David J. Rogers

Department of Zoology, Oxford University, United Kingdom

Archie C. A. Clements

Division of Epidemiology and Social Medicine, School of Population
Health, University of Queensland, Australia

OXFORD
UNIVERSITY PRESS

OXFORD

UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi

Kuala Lumpur Madrid Melbourne Mexico City Nairobi

New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece

Guatemala Hungary Italy Japan Poland Portugal Singapore

South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States

by Oxford University Press Inc., New York

© Oxford University Press 2008

The moral rights of the authors have been asserted

Database right Oxford University Press (maker)

First published 2008

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
without the prior permission in writing of Oxford University Press,
or as expressly permitted by law, or under terms agreed with the appropriate
reprographics rights organization. Enquiries concerning reproduction
outside the scope of the above should be sent to the Rights Department,
Oxford University Press, at the address above

You must not circulate this book in any other binding or cover
and you must impose the same condition on any acquirer

British Library Cataloguing in Publication Data

Data available

Library of Congress Cataloging in Publication Data

Data available

Typeset by Newgen Imaging Systems (P) Ltd., Chennai, India

Printed in Great Britain

on acid-free paper by

Antony Rowe Ltd., Chippenham

ISBN 978-0-19-850988-2 (Hbk.) 978-0-19-850989-9 (Pbk.)

10 9 8 7 6 5 4 3 2 1

Contents

Contents	v
Abbreviations	ix
Preface	xi
1 Introduction	1
1.1 Framework for spatial analysis	2
1.2 Scientific literature and conferences	3
1.3 Software	4
1.4 Spatial data	5
1.5 Book content and structure	6
1.5.1 Datasets used	6
1.5.1.1 Bovine tuberculosis data	6
1.5.1.2 Environmental data	6
2 Spatial data	9
2.1 Introduction	9
2.2 Spatial data and GIS	9
2.2.1 Data types	9
2.2.2 Data storage and interchange	11
2.2.3 Data collection and management	12
2.2.4 Data quality	13
2.3 Spatial effects	14
2.3.1 Spatial heterogeneity and dependence	14
2.3.2 Edge effects	14
2.3.3 Representing neighbourhood relationships	15
2.3.4 Statistical significance testing with spatial data	15
2.4 Conclusion	16
3 Spatial visualization	17
3.1 Introduction	17
3.2 Point data	17
3.3 Aggregated data	17
3.4 Continuous data	23

3.5	Effective data display	23
3.5.1	Media, scale, and area	23
3.5.2	Dynamic display	24
3.5.3	Cartography	26
3.5.3.1	Distance or scale	26
3.5.3.2	Projection	26
3.5.3.3	Direction	27
3.5.3.4	Legends	27
3.5.3.5	Neatlines, and locator and inset maps	27
3.5.3.6	Symbology	27
3.5.3.7	Dealing with statistical generalization	28
3.6	Conclusion	31
4	Spatial clustering of disease and global estimates of spatial clustering	32
4.1	Introduction	32
4.2	Disease cluster alarms and cluster investigation	32
4.3	Statistical concepts relevant to cluster analysis	33
4.3.1	Stationarity, isotropy, and first- and second-order effects	33
4.3.2	Monte Carlo simulation	33
4.3.3	Statistical power of clustering methods	34
4.4	Methods for aggregated data	34
4.4.1	Moran's I	35
4.4.2	Geary's c	37
4.4.3	Tango's excess events test (EET) and maximized excess events test (MEET)	37
4.5	Methods for point data	37
4.5.1	Cuzick and Edwards' k -nearest neighbour test	37
4.5.2	Ripley's K -function	39
4.5.3	Rogerson's cumulative sum (CUSUM) method	41
4.6	Investigating space–time clustering	41
4.6.1	The Knox test	42
4.6.2	The space–time k -function	42
4.6.3	The Ederer–Myers–Mantel (EMM) test	43
4.6.4	Mantel's test	43
4.6.5	Barton's test	43
4.6.6	Jacquez's k nearest neighbours test	44
4.7	Conclusion	44
5	Local estimates of spatial clustering	45
5.1	Introduction	45
5.2	Methods for aggregated data	46
5.2.1	Getis and Ord's local $G_i^*(d)$ statistic	46
5.2.2	Local Moran test	47
5.3	Methods for point data	49
5.3.1	Openshaw's Geographical Analysis Machine (GAM)	49
5.3.2	Turnbull's Cluster Evaluation Permutation Procedure (CEPP)	49
5.3.3	Besag and Newell's method	50

5.3.4	Kulldorff's spatial scan statistic	51
5.3.5	Non-parametric spatial scan statistics	52
5.3.6	Example of local cluster detection	53
5.4	Detecting clusters around a source (focused tests)	56
5.4.1	Stone's test	60
5.4.2	The Lawson–Waller score test	61
5.4.3	Bithell's linear risk score tests	62
5.4.4	Diggle's test	62
5.4.5	Kulldorff's focused spatial scan statistic	62
5.5	Space–time cluster detection	63
5.5.1	Kulldorff's space–time scan statistic	63
5.5.2	Example of space–time cluster detection	64
5.6	Conclusion	64
6	Spatial variation in risk	67
6.1	Introduction	67
6.2	Smoothing based on kernel functions	67
6.3	Smoothing based on Bayesian models	70
6.4	Spatial interpolation	73
6.5	Conclusion	80
7	Identifying factors associated with the spatial distribution of disease	81
7.1	Introduction	81
7.2	Principles of regression modelling	81
7.2.1	Linear regression	81
7.2.2	Poisson regression	83
7.2.3	Logistic regression	86
7.2.4	Multilevel models	87
7.3	Accounting for spatial effects	90
7.4	Area data	92
7.4.1	Frequentist approaches	93
7.4.2	Bayesian approaches	94
7.5	Point data	97
7.5.1	Frequentist approaches	97
7.5.2	Bayesian approaches	99
7.6	Continuous data	100
7.6.1	Trend surface analysis	100
7.6.2	Generalized least squares models	102
7.7	Discriminant analysis	103
7.7.1	Variable selection within discriminant analysis	106
7.8	Conclusions	107
8	Spatial risk assessment and management of disease	110
8.1	Introduction	110
8.2	Spatial data in disease risk assessment	110
8.3	Spatial analysis in disease risk assessment	111

8.4 Data-driven models of disease risk	112
8.5 Knowledge-driven models of disease risk	113
8.5.1 Static knowledge-driven models	113
8.5.2 Dynamic knowledge-driven models	117
8.6 Conclusion	118
References	120
Index	137

Abbreviations

AIC	Akaike information criterion
ASF	African swine fever
AVHRR	Advanced Very High Resolution Radiometer
AUC	Area under the curve
BPA	Basic probability assignments
BSE	Bovine spongiform encephalopathy
CAR	Conditional autoregressive
CEPP	Cluster Evaluation Permutation Procedure
CJD	Creutzfeldt-Jakob disease
CUSUM	Cumulative sum
DEMP	Density equalized map projection
DBMS	Database management system
DST	Dempster-Shafer theory
EET	Excess events test
EMM	Ederer-Myers-Mantel
ESDA	Exploratory spatial data analysis
FAO	Food and Agriculture Organization of the United Nations
FMD	Foot-and-mouth disease
GAM	Geographical Analysis Machine
GIS	Geographic information systems
GPS	Global positioning system
HEPP	Heterogeneous Poisson process
HGE	Human granulocytic ehrlichiosis
ICC	Intraclass correlation coefficient
IDW	Inverse distance weighting
K-L	Kullback-Leibler
LISA	Local indicators of spatial association
MA	Moving average
MAUP	Modifiable areal unit problem
MCDA	Multicriteria decision analysis
MCDM	Multicriteria decision making
MCMC	Markov chain Monte Carlo
MEET	Maximized excess events test
MLR	Maximum likelihood ratio
NDVI	Normalized Difference Vegetation Index
NNA	Nearest neighbour areas
NOAA	National Oceanic and Atmospheric Administration
ODBC	Open database connectivity

OWA	Ordered weighted averaging
Pmax	Poisson maximum
SAM	Statistical Analysis module
ROC	Receiver operating characteristic
SAR	Simultaneous autoregressive
SARS	Severe acute respiratory syndrome
SD	Standard deviation
SIDS	Sudden infant death syndrome
SIR	Susceptible-infected-recovered
SLE	Systemic lupus erythematosus
SMR	Standardized mortality/morbidity ratio
SQL	Structured Query Language
TB	Tuberculosis
TIN	Triangulated irregular network
UMP	Uniformly most powerful
URISA	Urban and Regional Information Systems Association
VPD	Vapour pressure deficit
WLC	Weighted linear combination

Preface

Over the last 20 years, the application of spatial analysis in the context of epidemiological surveillance and research has increased in an exponential fashion. Having been involved in this field since 1988, first as researchers and then also as postgraduate teachers, we felt there was a need for a textbook that helps to guide epidemiologists and other biologists logically through the complexities of spatial analysis.

This book aims to provide a practical introduction to spatial analysis, by focusing on application rather than theory, and by drawing on a wide range of examples from both human and animal health, including vector-borne and infectious diseases and non-infectious conditions. We provide worked examples of the principal methodologies, using mainly the same disease dataset throughout, which allows for direct comparison of the various techniques and helps to demonstrate their comparative strengths and weaknesses.

The book is written primarily for postgraduate students and postdoctoral researchers embarking upon epidemiological studies that may require the use of spatial analytical methods. However, the methods described are also relevant to students and researchers dealing with spatial data in the fields of ecology, zoology, parasitology, environmental science, geography, and statistics. Whilst the book is written in plain language, avoiding jargon as much as possible, a basic understanding of epidemiology and statistics is assumed.

The sequence around which we have structured the book involves firstly visualizing spatial patterns in data, then describing these spatial patterns, and finally attempting to explain the observed patterns. This further enables us to predict changes in patterns and to use our explanations and predictions to inform decisions and to guide policy

formulation. Following an introductory chapter, Chapters 2 and 3 address spatial data and the different ways in which they can be observed and presented. Chapters 4, 5, and 6 elaborate on the methods used to describe and quantify spatial patterns, while Chapter 7 looks at some of the methods that can be used to help explain spatial patterns, mostly in terms of environmental variables. Finally, Chapter 8 looks into ways of assessing disease risk and informing decision-making.

We have tried to be consistent with notation, but where this would lead to clumsiness have not forced ourselves to be so. Where notations deviate from the norm, the context should make this clear. At the risk of becoming fairly quickly outdated, we have included references to specific software programmes and provided links to websites. Whilst these all worked at the time of publishing we cannot guarantee their future validity.

The majority of worked examples presented in the book are based on data collected as part of Great Britain's national bovine tuberculosis (TB) control programme. A subset of the national database, comprising cattle TB data from the period 1986 to 1999 was used with permission from the United Kingdom Department for Environment, Food and Rural Affairs (DEFRA) and was kindly provided by Mr. Andy Mitchell and Dr. Richard Clifton-Hadley of the Veterinary Laboratories Agency (VLA).

The Animal Production and Health Division of the Food and Agriculture Organization of the United Nations (FAO) has supported this work as part of its mandate to build national and international capacity for the formulation of evidence-based disease control policies and strategies. In co-publishing the book, FAO hopes to promote its use among member countries.

The motivation to write this book came from our experience with epidemiological spatial analysis as researchers, as teachers, and as practitioners in policy formulation and advice.

Over the years, we have published numerous reviews on spatial analysis and geographic information systems (GIS) in epidemiology (Sanson et al. 1991; Pfeiffer and Morris 1994; Pfeiffer 2000; Robinson 2000; Pfeiffer and Hugh-Jones 2002; Pfeiffer 2004), have run short courses and distance learning modules in spatial analysis, and have taught spatial analysis as part of the masters' courses at the Royal Veterinary College and the London School of Hygiene and Tropical Medicine.

Through discussions with colleagues and post-graduate students in spatial analysis, it became clear to us that there is no spatial epidemiology textbook that provides a comprehensive introduction to the subject area, yet at the same time is

accessible to the wider group of epidemiologists, covering issues from spatial data management to analytical decision support tools. We hope that this book will go at least some of the way towards redressing this shortfall.

We were very fortunate in being able to convince our co-authors Mark Stevenson, Kim Stevens, David Rogers, and Archie Clements to join us in this endeavour. Particular thanks are due to Kim Stevens who, apart from contributing her own material to the book, also took over the editing: without her we would not have been able to complete it.

Dirk U. Pfeiffer
Royal Veterinary College
University of London

Timothy P. Robinson
Food and Agriculture Organization
of the United Nations