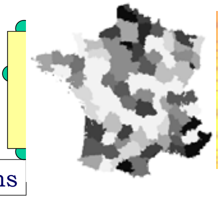
Spatial processes and statistical modelling

Peter Green
University of Bristol, UK
IMS/ISBA, San Juan, 24 July 2003

Spatial indexing

- Continuous space
- Discrete space
  - lattice
  - irregular - general graphs
  - areally aggregated
- Point processes
  - other object processes



Purpose of overview

- setting the scene for 8 invited talks on spatial statistics
- particularly for specialists in the other 2 areas

Perspective of overview

- someone interested in the development of methodology
  - for the analysis of spatially-indexed data
  - probably Bayesian
- models and frameworks, not applications
- personal, selective, eclectic

Genesis of spatial statistics

- adaptation of time series ideas
- ‘applied probability’ modelling
- geostatistics
- application-led

Space vs. time

- apparently slight difference
- profound implications for mathematical formulation and computational tractability

Requirements of particular application domains

- agriculture (design)
- ecology (sparse point pattern, poor data?)
- environmetrics (space/time)
- climatology (huge physical models)
- epidemiology (multiple indexing)
- image analysis (huge size)

Key themes

- conditional independence
  - graphical/hierarchical modelling
- aggregation
  - analysing dependence between differently indexed data
  - opportunities and obstacles
- literal credibility of models
- Bayes/non-Bayes distinction blurred

A big subject….

Noel Cressie:

“This may be the last time spatial statistics is squeezed between two covers”

(Preface to *Statistics for Spatial Data*, 900pp., Wiley, 1991)

Why build spatial dependence into a model?

- No more reason to suppose independence in spatially-indexed data than in a time-series
- However, substantive basis for form of spatial dependent sometimes slight - very often space is a surrogate for missing covariates that are correlated with location

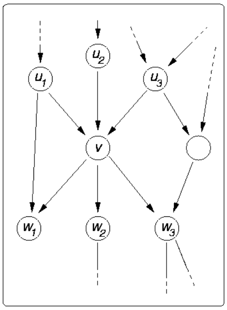
Discretely indexed data

Modelling spatial dependence in discretely-indexed fields

- Direct
- Indirect
  - Hidden Markov models
  - Hierarchical models

Hierarchical models, using DAGs

Variables at several levels - allows modelling of complex systems, borrowing strength, etc.



Modelling with undirected graphs

Directed acyclic graphs are a natural representation of the way we usually *specify* a statistical model - directionally:

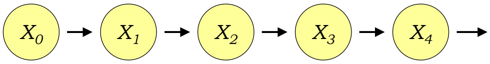
- disease → symptom
- past → future
- parameters → data ……

whether or not causality is understood.

But sometimes (e.g. spatial models) there is *no natural direction*

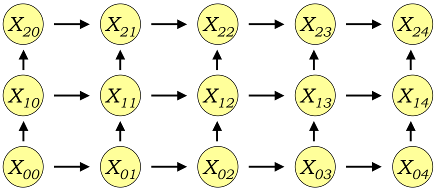
Conditional independence

In *model specification*, spatial context often rules out directional dependence (that would have been acceptable in time series context)



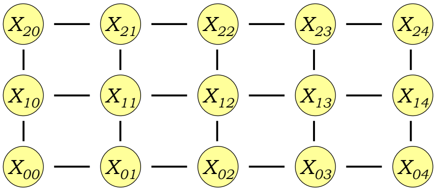
Conditional independence

In *model specification*, spatial context often rules out directional dependence

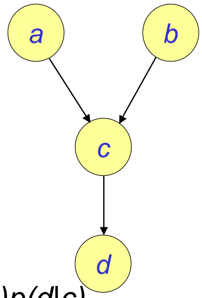


Conditional independence

In *model specification*, spatial context often rules out directional dependence



Directed acyclic graph



in general:

$$p(x)=\prod_{v\in V} p(x_v \mid x_{\mathrm{pa}(v)})$$

for example:

$$p(a,b,c,d)=p(a)p(b)p(c|a,b)p(d|c)$$

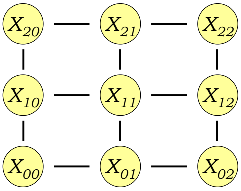
In the RHS, *any* distributions are legal, and uniquely define joint distribution

Undirected (CI) graph

Regular lattice, irregular graph, areal data...

Absence of edge denotes conditional independence given all other variables

But now there are non-trivial constraints on conditional distributions



Undirected (CI) graph

$$p(X) \propto \exp\left\{\sum_C V_C(X_C))\right\} \quad (♣)$$

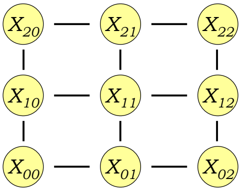
clique

$$\Rightarrow$$

$$p(X_i \mid X_{-i}) \propto \exp\left\{\sum_{C \ni i} V_C(X_C))\right\}$$

$$\Rightarrow$$

$$p(X_i \mid X_{-i}) = p(X_i \mid X_{\partial i})$$



The Hammersley-Clifford theorem says essentially that the converse is also true - the only sure way to get a valid joint distribution is to use (♣)

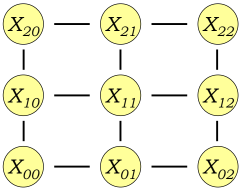
Hammersley-Clifford

A positive distribution *p(X)* is a Markov random field

$$p(X_i \mid X_{-i}) = p(X_i \mid X_{\partial i})$$

if and only if it is a Gibbs distribution

$$p(X) \propto \exp\left\{\sum_C V_C(X_C)\right\}$$



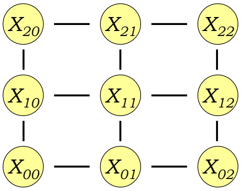
- Sum over cliques *C* (complete subgraphs)

Partition function

Almost always, the constant of proportionality in

$$p(X) \propto \exp\left\{\sum_C V_C(X_C)\right\}$$

is not available in tractable form: an obstacle to likelihood or Bayesian inference about parameters in the potential functions $\{V_C(X_C))\}$



Physicists call $Z = \int_X \exp\left\{\sum_C V_C(X_C)\right\}$ the partition function

---

Gaussian Markov random fields: spatial autoregression

If $V_C(X_C)$ is $-\beta_{ij}(x_i - x_j)^2/2$ for $C=\{i,j\}$ and 0 otherwise, then

$$p(X) \propto \exp\left\{\sum_C V_C(X_C)\right\}$$

is a multivariate Gaussian distribution, and

$$p(X_i \mid X_{-i}) = p(X_i \mid X_{\partial i})$$

is the univariate Gaussian distribution

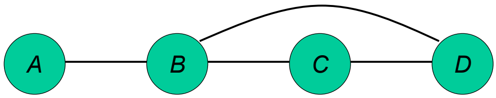
$$X_i \mid X_{-i} \sim N(\sigma_i^2 \sum_{j \in \partial i} \beta_{ij} X_j, \sigma_i^2) \quad \text{where} \quad \sigma_i^2 = 1/\sum_{j \in \partial i} \beta_{ij}$$

	A	*B*	*C*	*D*
A	2	1	0	0
B	1	2	−1	1
C	0	−1	4	2
D	0	1	2	3

Gaussian random fields

non-zero ⇔ cov(*B*,*D* | *A*,*C*) non-zero

Inverse of (co)variance matrix: dependent case



Gaussian Markov random fields: spatial autoregression

Distinguish these conditional autoregression (CAR) models from the corresponding simultaneous autoregression (SAR) models

$$X_i = \sum_j \tilde{\beta}_{ij} X_j + \varepsilon_i \qquad \{\varepsilon_i\} \text{ i.i.d. normal}$$

(cf time series case).

The latter are less compatible with hierarchical model structures.

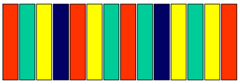
Non-Gaussian Markov random fields

Pairwise interaction random fields with less smooth realisations obtained by replacing squared differences by a term with smaller tails, e.g.

$$p(X) \propto \exp\left\{\sum_C V_C(X_C)\right\}$$

$$= \exp\left\{-\beta\delta(1+\delta)\sum_{i\sim j}\log\cosh(\frac{x_i - x_j}{\delta})\right\}$$

Agricultural field trials



- strong cultural constraints
- design, randomisation, cultivation effects
- 1-D analysis in 2-d fields
- relationships between IB designs, splines, covariance models, spatial autoregression…

Discrete Markov random fields

Besag (1974) introduced various cases of

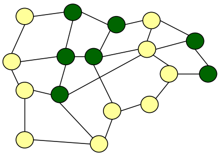
$$p(X) \propto \exp\left\{\sum_C V_C(X_C)\right\}$$

for discrete variables, e.g. auto-logistic (binary variables), auto-Poisson (local conditionals are Poisson), auto-binomial, etc.

Auto-logistic model (X_i = 0 or 1)

$$p(X) \propto \exp\left\{\sum_C V_C(X_C)\right\}$$

$$\propto \exp\left\{\sum_i \alpha_i x_i + \sum_{i\sim j} \theta_{ij} x_i x_j\right\}$$



$p(X_i \mid X_{-i}) = p(X_i \mid X_{\partial i})$ is Bernoulli(p_i) with

$$\log(p_i/(1-p_i)) = (\alpha_i + \sum_{j\in\partial i} \theta_{ij} x_j)$$

- a very useful model for dependent binary variables (*NB various parameterisations*)

Statistical mechanics models

$$p(X) \propto \exp\left\{\sum_C V_C(X_C)\right\}$$

The classic Ising model (for ferromagnetism) is the symmetric autologistic model on a square lattice in 2-D or 3-D. The Potts model is the generalisation to more than 2 ‘colours’

$$p(X) \propto \exp\left\{\sum_i \alpha_{x_i} + \beta\sum_{i\sim j} I[x_i = x_j]\right\}$$

and of course you can usefully un-symmetrise this.

Auto-Poisson model

$$p(X) \propto \exp\left\{\sum_C V_C(X_C)\right\}$$

$$\propto \exp\left\{\sum_i (\alpha_i x_i - \log x_i!) + \sum_{i\sim j} \theta_{ij} x_i x_j\right\}$$

$$p(X_i \mid X_{-i}) = p(X_i \mid X_{\partial i})$$

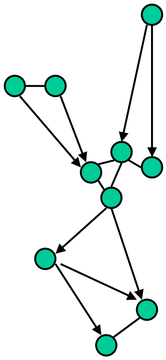
is Poisson $(\exp(\alpha_i + \sum_{j\in\partial i} \theta_{ij} x_j))$

For integrability, θ_{ij} must be ≤ 0, so this only models negative dependence: very limited use.

Hierarchical models
and
hidden Markov processes

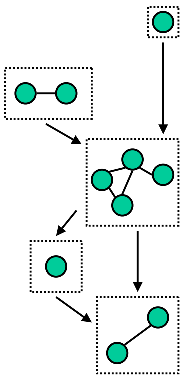
Chain graphs

- If both directed and undirected edges, but no directed loops:
- can rearrange to form global DAG with undirected edges within blocks



Chain graphs

- If both directed and undirected edges, but no directed loops:
- can rearrange to form global DAG with undirected edges within blocks
- Hammersley-Clifford within blocks

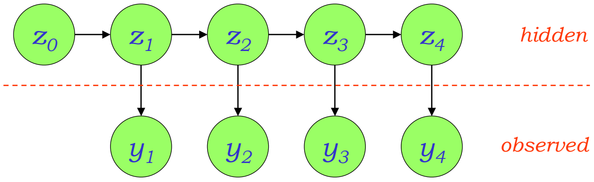


Hidden Markov random fields

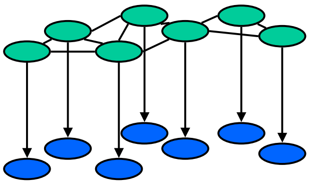
- We have a lot of freedom modelling spatially-dependent continuously-distributed random fields on regular or irregular graphs
- But very little freedom with discretely distributed variables
- ⇒ use hidden random fields, continuous or discrete
- compatible with introducing covariates, etc.

Hidden Markov models

e.g. Hidden Markov chain



Hidden Markov random fields



Unobserved dependent field

Observed conditionally-independent discrete field

(a chain graph)

Spatial epidemiology applications

relative risk

cases $Y_i \sim Poisson(\lambda_i e_i)$ expected cases

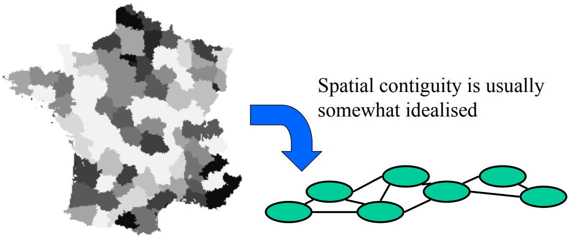
independently, for each region *i*. Options:

- $\log(\lambda_i)$ CAR, CAR+white noise (BYM, 1989)
- Direct modelling of $\text{cov}(\log(\lambda_i))$,e.g. SAR
- Mixture/allocation/partition models:

$$E(Y_i) = \lambda_{z_i} e_i$$

- Covariates, e.g.: $E(Y_i) = \lambda_{z_i} e_i \exp(x_i^T \beta)$

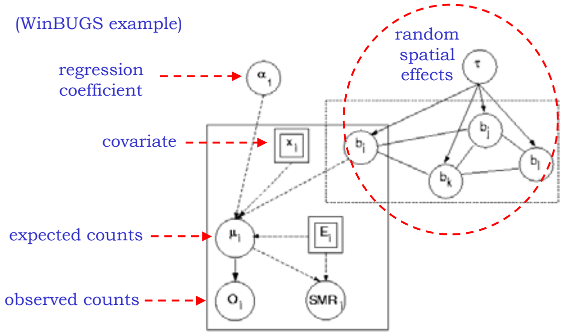
Spatial epidemiology applications



Spatial contiguity is usually somewhat idealised

CAR model for lip cancer data

(WinBUGS example)



Example of an allocation model

Richardson & Green (*JASA*, 2002) used a hidden Markov random field model for disease mapping

$$y_i \sim \text{Poisson}(\lambda_{z_i} e_i)$$

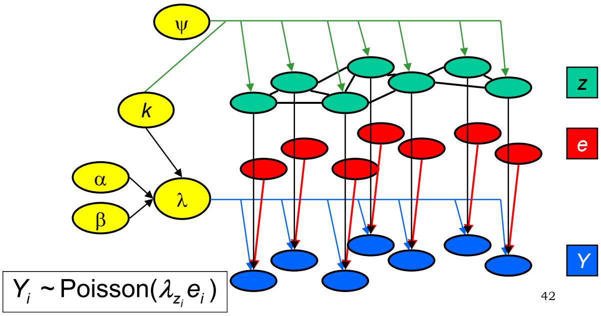
observed incidence

relative risk parameters

hidden MRF

expected incidence

Chain graph for disease mapping based on Potts model

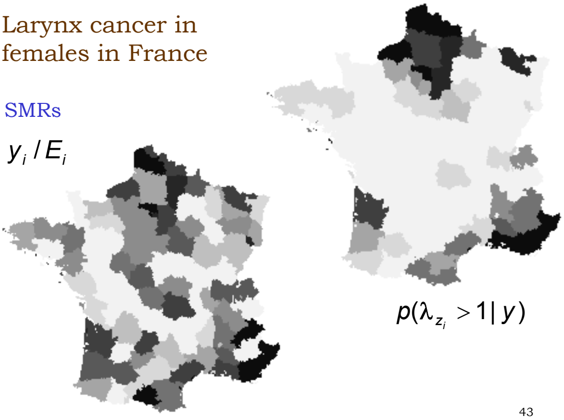


$Y_i \sim \text{Poisson}(\lambda_{z_i} e_i)$

Larynx cancer in females in France

SMRs

y_i / E_i



$p(\lambda_{z_i} > 1 | y)$

Continuously indexed data

Continuously indexed fields

The basic model is the Gaussian random field $\{Z(s), s \in E \subseteq \Re^2\}$ with $E(Z(s)) = \mu(s)$ and $\operatorname{cov}(Z(s), Z(t)) = c(s,t)$

Translation-invariant or fully stationary (isotropic) cases have $E(Z(s)) = \mu$ and $\operatorname{cov}(Z(s), Z(t)) = c(s-t)$ or $\operatorname{cov}(Z(s), Z(t)) = c(\| s-t \|)$,resp.

Geostatistics and kriging

- There is a huge literature on a group of methodologies originally developed for geographical and geological data
- The main theme is prediction of (functionals of) a random field based on observations at a finite set of locations

---

Ordinary kriging

- $\{Z(s), s \in E \subseteq \Re^2\}$ is a random process, we have observations $\{Z(s_1), Z(s_2), \ldots, Z(s_n)\}$ and we wish to predict $g(Z(.))$, e.g. a "block average" $\int_B Z(s) ds / |B|$
- The usual basis is least-squares prediction, using a model for the mean and covariance of $\{Z(s)\}$ estimated from the data

---

Ordinary kriging

The usual assumption is that $\{Z(s)\}$ is intrinsically stationary, i.e. has 2nd order structure

$$E(Z(s+h) - Z(s)) = 0$$

$$\operatorname{var}(Z(s+h) - Z(s)) = 2\gamma(h)$$

for all s

$\gamma(h)$ is called the semi-variogram

This is somewhat weaker than full 2nd-order stationarity

Ordinary kriging

The optimal solution to the prediction problem in terms of the semivariogram follows from standard linear algebra arguments; an empirical estimate of the semivariogram is then plugged in.

Variants of kriging

Kriging without intrinsic stationarity (& a model instead of empirical estimates)

Co-kriging (multivariate)

Robust kriging

Universal kriging (kriging with regression)

Disjunctive (nonlinear) kriging

Indicator kriging

Connections with splines

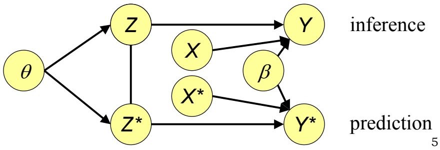
Bayesian geostatistics

(Diggle, Moyeed and Tawn, *Appl Stat*, 1998)

Given data (s_i, x_i, Y_i), build model starting with a Gaussian random field $\{Z(s), s \in E \subseteq \Re^2\}$ with $E(Z(s)) = 0$ and $\text{cov}(Z(s), Z(t)) = c_\theta(s-t)$

Set $Y_i \mid Z(.) \sim f(y_i \mid M_i)$ where $M_i = E(Y_i \mid Z(.))$

and $h(M_i) = Z(s_i) + x_i^T \beta$



Point data

Point processes

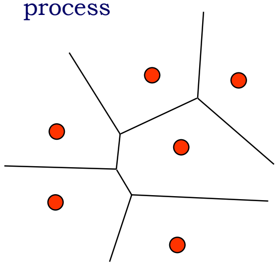
- (inhomogeneous) Poisson process
- Neyman-Scott process
- (log Gaussian) Cox process
- Gibbs point process
- Markov point process
- Area-interaction process

Analysis of spatial point pattern

- Very strong early emphasis on modelling clustering and repelling alternatives to homogeneous Poisson process (*complete spatial randomness*)
- May be different effects at different scales
- Interpretations in terms of mechanisms, e.g. in ecology, forestry

Point process as parametrisation of space

Voronoi tessellation of random point process



Flexible modelling of surfaces: step functions, polynomials, …

Rare disease point data

- Regard locations of cases as Poisson process with highly structured intensity process
  - Covariates
  - Spatial dependence

$Y(ds) =$ number of cases in *ds*

$$E(Y(ds)) = \lambda(s)e(ds)$$

$$E(Y(A)) = \int_{s \in A} \lambda(s)e(ds)$$

Models without covariates 1

Cox process $E(Y(ds)) = \lambda(s)e(ds)$

where $\lambda(s)$ is a random field, e.g.

$\log(\lambda(s)) = Z(s)$ is Gaussian – 'log Gaussian Cox process' (Moller, Syversveen and Waagepetersen, 1998)

Models without covariates 2

Smoothed Gamma random field

(Wolpert and Ickstadt, 1998)

$\lambda(s) = \int k(s,u)\Gamma(du)$

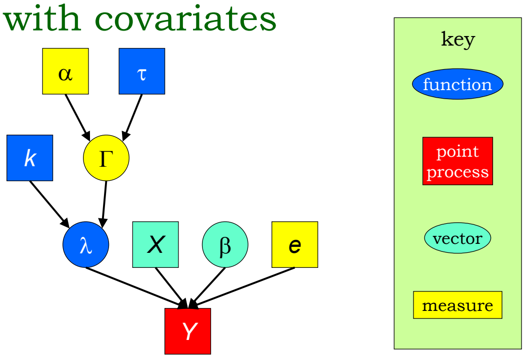
where $k(s,u)$ is a kernel function

and $\Gamma(du) \sim \Gamma(\alpha(du), \tau(u))$

$\lambda(s)$ is a sum of smoothed gamma-distributed impulses

-- example of shot-noise Cox process

DAG for Gamma RF model with covariates



Models without covariates 3

Voronoi tessellation models

(PJG,1995; Heikkinen and Arjas, 1998)

$$\lambda(s) = \sum_k \lambda_k I[s \in A_k]$$

where $\{A_k\}$ are cells of Voronoi tessellation of an unobserved point process and $\{\lambda_k\}$ might be independent or dependent (e.g. CAR model for logs)

Introducing covariates

With covariates $\{X_j(s)\}$ measured at case locations s, usual formulation is multiplicative

$$\lambda(s) \Rightarrow \lambda(s)\exp\left\{\sum_j X_j(s)\beta_j\right\}$$

but occasionally additive

$$\lambda(s) \Rightarrow \lambda(s) + \sum_j X_j(s)\beta_j$$

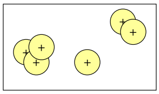
+ data-dependent constraints on parameters

Markov point processes

Rich families of non-Poisson point processes can be defined by specifying their densities (Radon-Nikodym derivatives) w.r.t. unit-rate Poisson process, e.g. pairwise interaction models

$$p(s) = \alpha\beta^{n(s)}\prod_{(i,j)} g(s_i, s_j)$$

(e.g. $g(s_i,s_j)=\gamma<1$ if $d(s_i,s_j)<\rho$, 1 otherwise), and area-interaction models



$$p(s) = \alpha\beta^{n(s)}\gamma^{-|s\oplus T|}$$

Note formal similarity to Gibbs lattice models

Marginal distribution of #points usually not explicit

Object processes

- Poisson processes of objects (lines, planes, flats, ...)
- Coloured triangulations....

Aggregation

Aggregation coherence and ecological bias

- Commonly, covariates and responses are spatially indexed differently, and for most models this poses coherence problems (linear Gaussian case the main exception)
- E.g. areally-aggregated response Y_i=$Y(A_i)$, and continuously indexed covariate $X(s)$

$$E(Y(A_i)) = \int_{s\in A_i} \lambda(s)\exp(X(s)\beta)e(ds)$$

Aggregation coherence and ecological bias

$$E(Y(A_i)) = \int_{s\in A_i} \lambda(s)\exp(X(s)\beta)e(ds)$$

Even with uniform $\lambda(s)e(ds)$, this is not of form

$$E(Y(A_i)) = \lambda_i^{\oplus}\exp(\overline{X}(A_i)\beta)$$

where

$$\overline{X}(A_i) = \int_{A_i} X(s)ds/|A_i|$$

$\Rightarrow$ (mis-specification) bias in estimation of β.

Need to know spatial variation in covariate

Aggregation coherence and ecological bias

Additive formulation

$$E(Y(A_i)) = \int_{s \in A_i} (\lambda(s) + X(s)\beta) e(ds)$$

avoids this problem, as does the Ickstadt and Wolpert approach, to some extent

---

Invited talks on spatial statistics

Fri 09:00

- Brad Carlin: space & space-time CDF models, air pollutant data
- Jon Wakefield: ecological fallacy

Fri 10:45

- Montserrat Fuentes: spatial design, air pollution
- Doug Nychka: filtering for weather forecasting

Fri 17:30

- Susie Bayarri: validating computer models
- Arnoldo Frigessi: localisation of GSM phones

Sat 10:45

- Rasmus Waagepetersen: Poisson-log Gaussian processes
- Adrian Baddeley: point process diagnostics

Spatial processes and statistical modelling

Peter Green
University of Bristol, UK
IMS/ISBA, San Juan, 24 July 2003

P.J.Green@bristol.ac.uk
http://www.stats.bris.ac.uk/~peter/PR