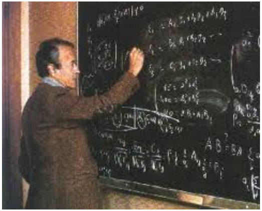
Spatial and spatio-temporal log-Gaussian Cox processes: re-defining geostatistics

Peter J Diggle

Lancaster University and University of Liverpool

- traditionally, a self-contained methodology for spatial
  prediction, developed at École des Mines, Fontainebleau, France

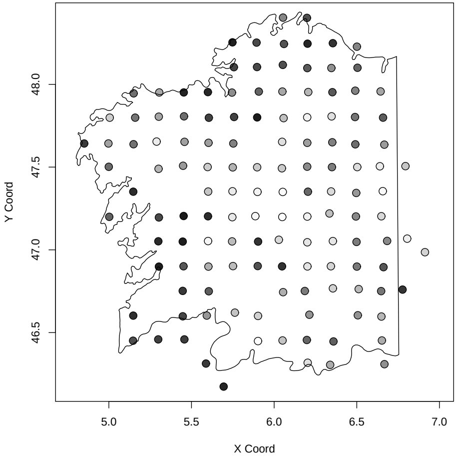


- nowadays, that part of spatial statistics which is
  concerned with data obtained by spatially discrete
  sampling of a spatially continuous process

Model-based Geostatistics (Diggle, Moyeed and Tawn, 1998)

- **the application of general principles of statistical modelling and inference to geostatistical problems**
- **which means:**
  - **formulate a model for the data**
  - **use likelihood-based methods of inference**
  - **answer the scientific question**

Geostatistical data and model : lead pollution in Galicia



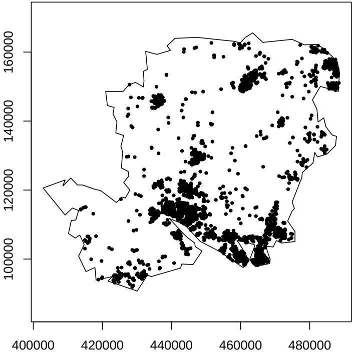
- **$S(x)$ pollution surface**
- **x_i sampling location**
- **Y_i measured pollution**
- **$Y_i | S(\cdot) \sim \mathrm{N}(S(x_i), \tau^2)$**

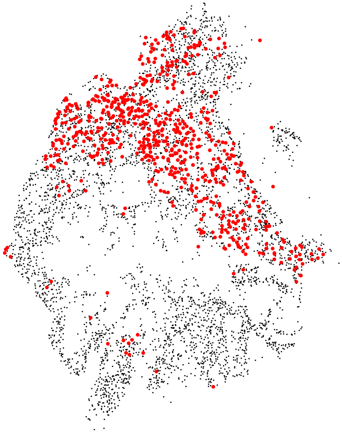
Point processes

- **Stochastic models for arrangements of points in space and/or time**

- **Scientific focus on understanding why the points are where they are:**
  - **absolutely**
  - **and/or relatively**

Point process data: two examples





Poisson process

- $\lambda(x)$: intensity function
- event-locations mutually independent, probability density proportional to $\lambda(x)$

Requires: $\lambda(x) \geq 0$ and $\int_A \lambda(x)dx < \infty$, for any finite region A

Cox process

- $\lambda(x)$ is unobserved realisation of stochastic process $\Lambda(x)$

Trans-Gaussian Cox process

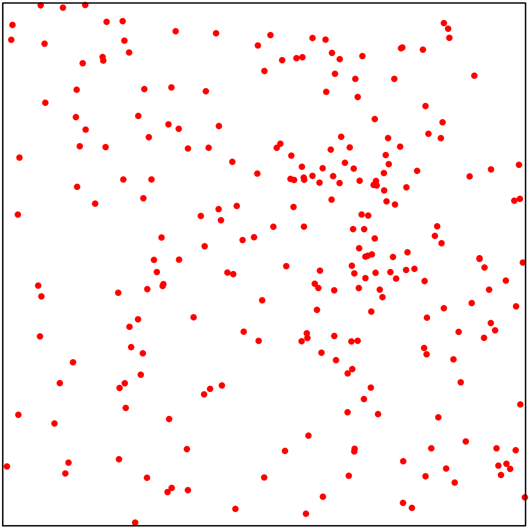
Cox process with:

- $\Lambda(x) = \mathcal{F}\{S(x)\}$
- $S(x) \sim$ Gaussian process

Log-Gaussian Cox process: $\mathcal{F}(\cdot) = \log(\cdot)$

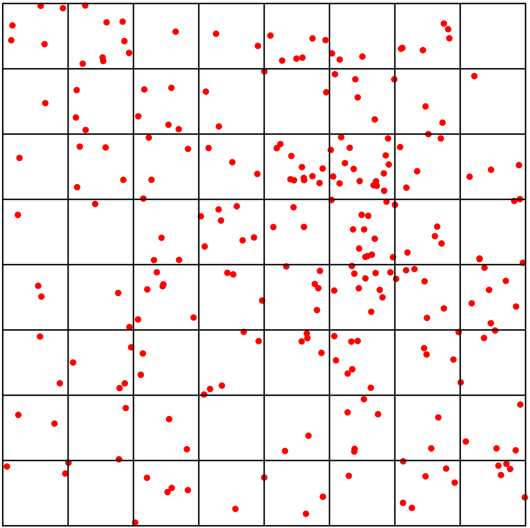
- introduced by Møller, Syversveen and Waagepetersen (1998)
- popular because of analytic tractiability (moments, etc)
- but any $\mathcal{F}(\cdot)$ OK if using Monte-Carlo methods of inference

LGCP as a geostatistical model

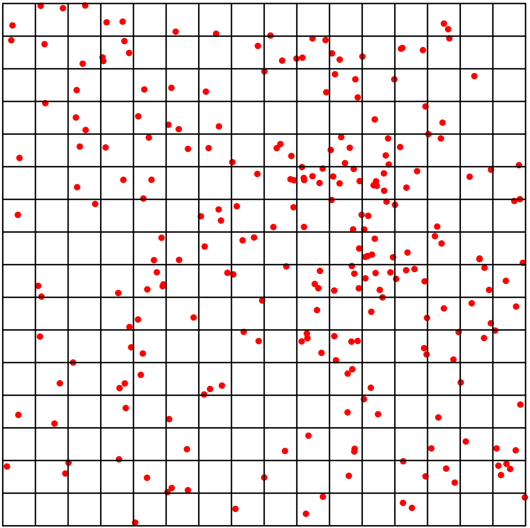


5	8	1	2	6	3	6	0
1	3	5	2	2	5	2	1
1	4	4	3	11	16	5	4
1	1	3	7	3	12	5	4
2	2	6	3	5	10	6	6
2	4	2	5	4	7	4	3
2	2	2	0	2	5	2	4
2	1	5	1	3	1	6	5

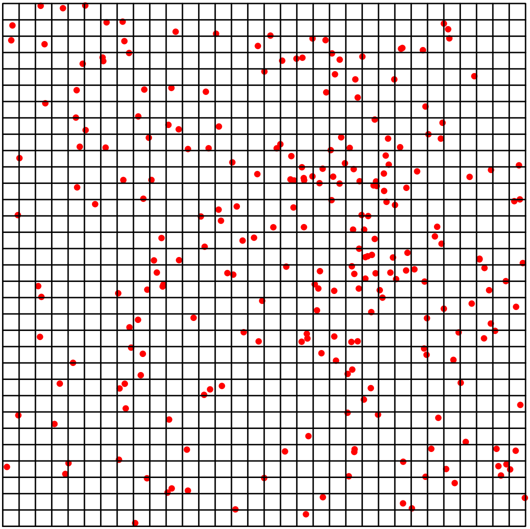
LGCP as a geostatistical model



LGCP as a geostatistical model



LGCP as a geostatistical model



Re-defining geostatistics

Old:

- unobserved, real-valued stochastic process $S = \{S(x) : x \in A\}$
- pre-specified locations $\{x_i \in A : i = 1, \dots, n\}$ measurements

  $Y = \{Y_i : i = 1, \dots, n\}$ at locations x_i
- use model for $[S, Y] = [S][Y|S]$ to predict S

New:

- unobserved, real-valued stochastic process $S = \{S(x) : x \in A\}$
- data D
- use model for $[S, D] = [S][D|S]$ to predict S

New definition shifts focus from data to problems

LGCP model-fitting

Ingredients:

- latent Gaussian process S
- data D
- parameters θ

Predictive inference via MCMC or INLA

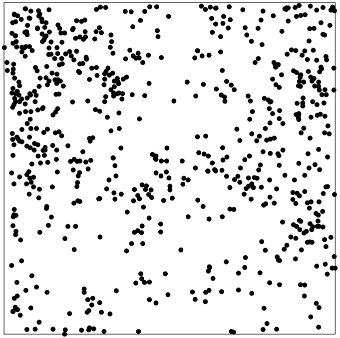
$$[\theta, \mathsf{S}, \mathsf{D}] = [\theta][\mathsf{S}|\theta][\mathsf{D}|\mathsf{S}, \theta] \Rightarrow [\mathsf{S}|\mathsf{D}] = \int [\mathsf{S}|\mathsf{D}; \theta][\theta|\mathsf{D}]d\theta$$

$$\int [\mathsf{S}|\mathsf{D}; \theta][\theta|\mathsf{D}]d\theta \approx [\mathsf{S}|\mathsf{D}; \hat{\theta}] \text{ ?}$$

Applications

- **intensity estimation: hickories in Lansing Woods**
- **spatial segregation: BTB in Cornwall**
- **disease atlases: lung cancer mortality in Spain**
- **real-time spatial health surveillance: AEGISS**

Intensity estimation: hickories in Lansing Woods

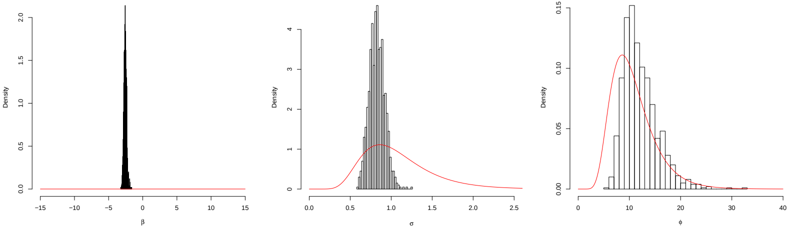


- use data to construct non-parametric estimate of $\lambda(x)$
- lots of existing methods, but inferential standing unclear
- LGCP approach enables predictive inference

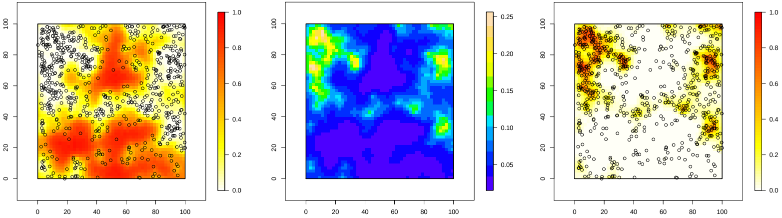
$$\begin{aligned} \Lambda(x) &= \exp\{S(x)\} \\ S(\cdot) &\sim \mathrm{SGP}(\beta, \sigma^2, \rho(u)) \\ \rho(u) &= \exp(-u/\phi) \end{aligned}$$

Intensity estimation: Lansing Woods results

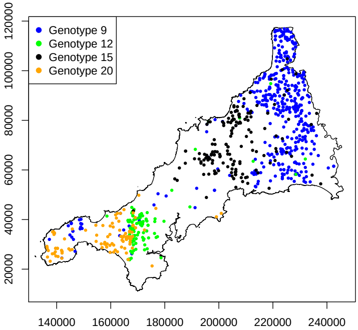
Parameter estimation



Prediction



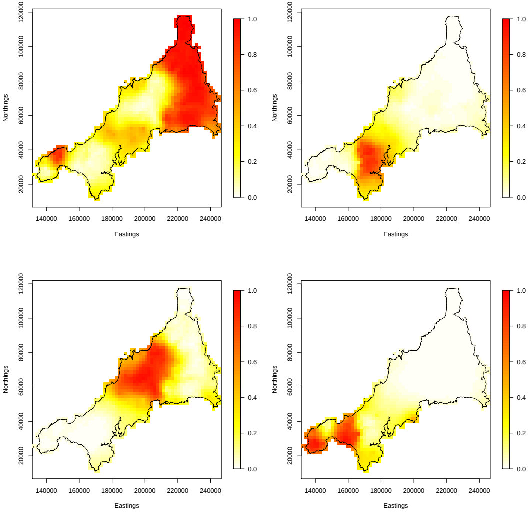
Spatial segregation: BTB in Cornwall



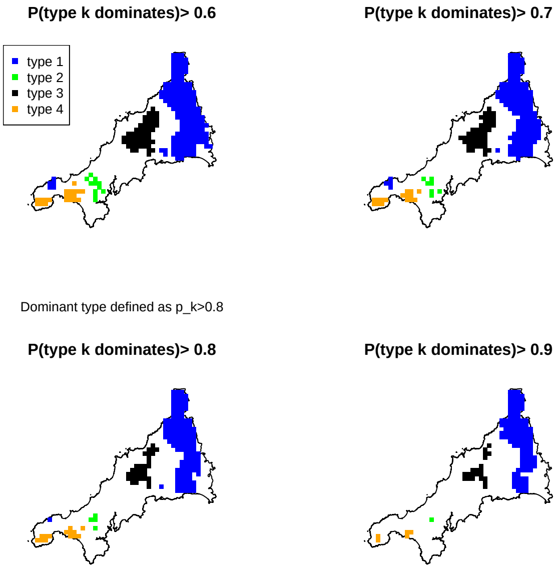
- $\Lambda_k(x) = \exp\{\beta_k + S_0(x) + S_k(x)\} : k = 1, \ldots, m$
- $S_0(x)$ **not identifiable:**

  $p_k(x) = \{\Lambda_k(x)\}/\{\sum_{j=1}^{m} \Lambda_j(x)\} = \exp[-\sum_{j \neq k}\{\beta_j + S_j(x)\}]$

BTB in Cornwall: estimated type-specific probability surfaces

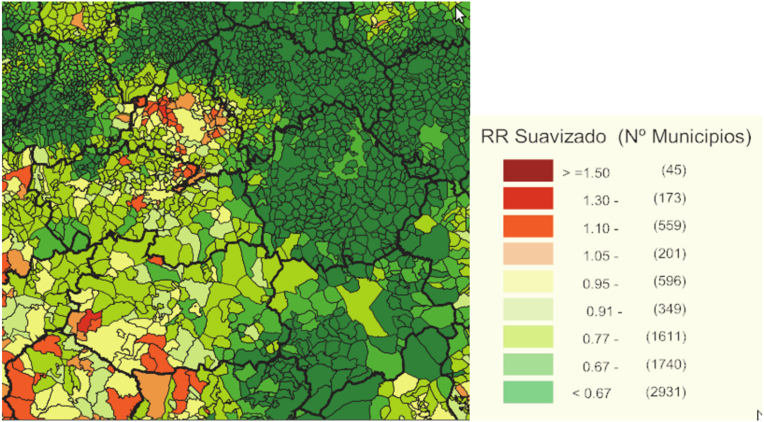


BTB in Cornwall: areas of type-specific (0.8+) dominance



Disease atlases: lung cancer mortality in Spain

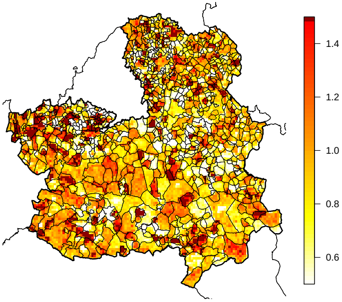
Spatially discrete approach: population denominators and risk-factor information aggregated to area-level averages (MRF model)



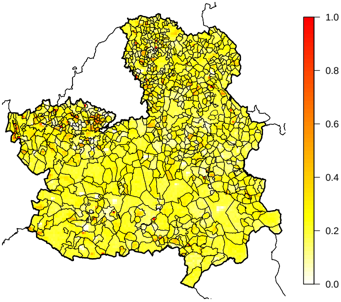
Disease atlases: lung cancer mortality in Spain

Spatially continuous approach: population denominators and risk-factor information in principle at point level (LGCP model)

$$\Lambda(x) = d(x)\exp\{z(x)'\beta + S(x)\}$$



Predicted risk surface



P(relative risk > 1.1)

Building spatio-temporal dependence

- **Separable**

$$\rho(u, v) = \rho_S(u; \alpha)\rho_T(v; \beta)$$

- **Non-separable empirical**

$$\rho(u, v) = \rho_S(u; \alpha; \beta)\rho_T(v)\{1 + f(u, v; \theta)\}$$

- **Non-separable mechanistic**

$$S(x, t) = \lim_{\delta \to 0} \left\{ \int h_\delta(u) S(x - u, t - \delta) du + Z_\delta(x, t) \right\}$$

In conclusion

1. **Statistical modelling should be driven by the underlying scientific problem, rather than by the format of available data**
2. **Most (but not all) natural phenomena are spatially continuous**
3. **Surprisingly many of the problems to which spatial methods can make a useful contribution reduce to an application of Bayes' Theorem**

$$[S, D] = [S][D|S] \Rightarrow [S|D]$$

4. **Modelling an unobserved stochastic process and assigning a prior distribution to an unknown parameter are mathematically equivalent but scientifically different activities**
5. **Modelling as a route to empirical prediction and modelling as a route to understanding a physical/biological mechanism demand different approaches**

Acknowledgements

CHICAS, Lancaster University : Paula Moraga, Barry Rowlingson, Ben Taylor

MRC: Methodology Research Grant G0902153

Diggle, P.J., Moraga, P., Rowlingson, B. and Taylor, B. (2013). Spatial and spatio-temporal log-Gaussian Cox processes: extending the geostatistical paradigm. *Statistical Science*, 28, 542–563.