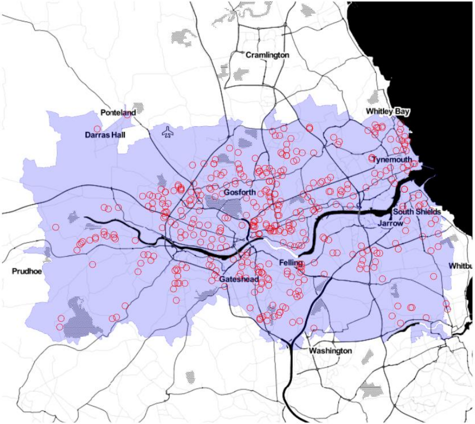
Methods and an R package for Bayesian inference with log-Gaussian Cox processes

http://cran.r-project.org/web/packages/lgcp

Benjamin M. Taylor
http://www.lancaster.ac.uk/staff/taylorb1
(Joint work with Tilman Davies, Barry Rowlingson and Peter Diggle)

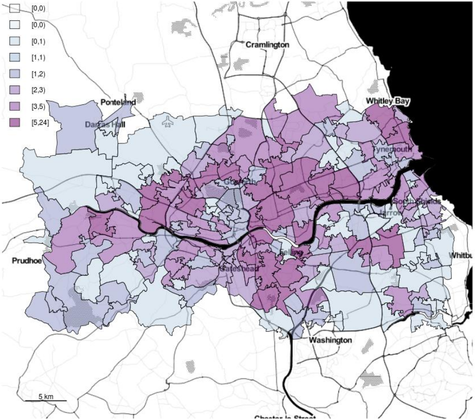
Spatial Point Process Data

- Primary biliary cirrhosis in Newcastle-Upon-Tyne
- 415 geo-referenced cases of definite or probable primary biliary cirrhosis (PBC) alive between 1987 and 1994.



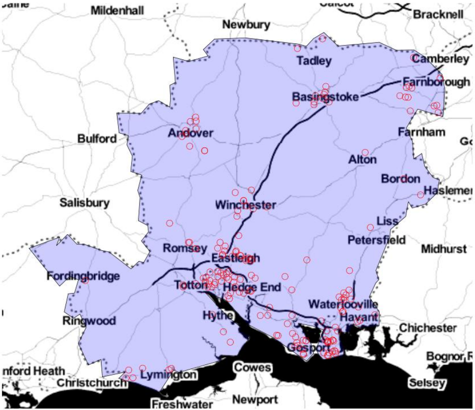
Aggregated Spatial Point Process Data

- Again, Primary biliary cirrhosis in Newcastle-Upon-Tyne
- This time suppose we only observe case numbers aggregated to regions



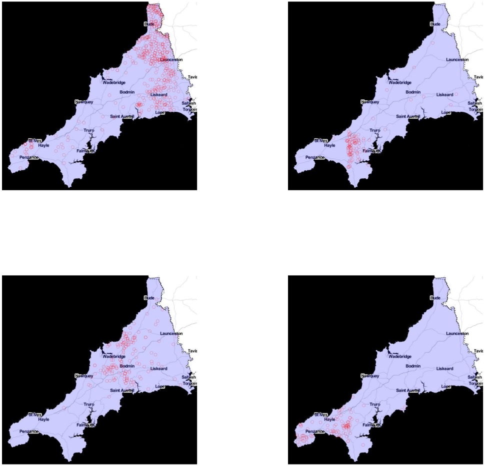
Spatiotemporal Point Process Data

- Ascertainment and Enhancement of Gastroenteric Infection Surveillance Statistics
- Calls to NHS direct



Multivariate Spatial Point Process Data

- 873 genotyped cases of bovine tuberclosis breakdowns in Cornwall herds (1989–2002)
- Is there spatial segregation? (related to potential transmission mechanisms)



The log-Gaussian Cox Process

Gaussian process: A stochastic process $Y(s)$ where $s \in \mathbb{R}^d$ is a Gaussian process if any finite collection $[Y(s_1), \ldots, Y(s_n)]$ has a multivariate Gaussian distribution.

(Spatial/ spatiotemporal) Poisson process: A stochastic process that counts the number of events in a given spatial/spatiotemporal region.

Intensity process: a non-negative valued stochastic process $\Lambda : S \mapsto [0, \infty)$ where $S \subset \mathbb{R}^d$; typically $d = 2$.

Cox process: An inhomogenegous Poisson process with stochastic intensity Λ.

log-Gaussian Cox process: a Cox process where $\log \Lambda$ is a Gaussian process

The log-Gaussian Cox Process

If X is a Cox process, then conditional on Λ, we have for any finite observation window $W \subseteq S$

$$\text{number of } X \text{ in } W \sim \text{Poisson}\left(\int_W \Lambda(s)ds\right)$$

- We think of X as being continuous in space
- In practice, for computational reasons, we treat X as piecewise constant on a fine grid.

Example: Spatial Point Process

Let $X(s)$ denote the number of events in a grid cell containing s.

$$\begin{aligned} X(s) &\sim \text{Poisson}[\Lambda(s)] \\ \Lambda(s) &= C_A \lambda(s) \exp\{Z(s)\beta + Y(s)\} \end{aligned}$$

- C_A is the area of the grid cell containing s (regular grid).
- $Z(s)$ is a vector of covariates particular to the grid cell containing s
- $Y(s)$ is the value of the Gaussian process Y in the grid cell containing s

Bayesian Inference

Given some observed data X, we use the posterior density:

$$\pi(\beta, Y, \eta|X) \propto \pi(X|\beta, Y, \eta)\pi(\beta, Y, \eta)$$

for Bayesian inference.

- $\eta = \{\sigma, \phi\}$ are parameters that control the properties of Y (respectively variance and spatial dependence)
- We assume $\pi(X|\beta, Y, \eta) = \pi(X|\beta, Y)$

Markov Chain Monte Carlo

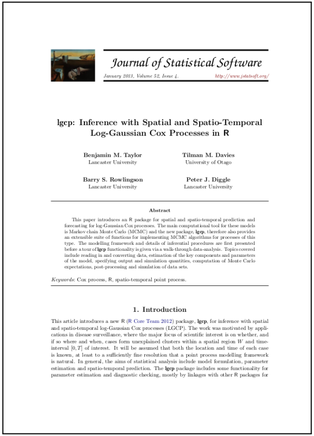
Independent block proposal scheme. A mix of:

- Langevin kernels for Y and β
- A random walk kernel for η

Algorithm tuning:

- Design the proposal covariance using a quadratic approximation to the posterior.
- Adaptively tune the chain to maintain acceptance rate of 0.574 to suit Langevin kernels

lgcp Log-Gaussian Cox Processes in R



Journal of Statistical Software
January 2013, Volume 52, Issue 4. http://www.jstatsoft.org/

lgcp: Inference with Spatial and Spatio-Temporal Log-Gaussian Cox Processes in R

Benjamin M. Taylor
Lancaster University

Tilman M. Davies
University of Otago

Barry S. Rowlingson
Lancaster University

Peter J. Diggle
Lancaster University

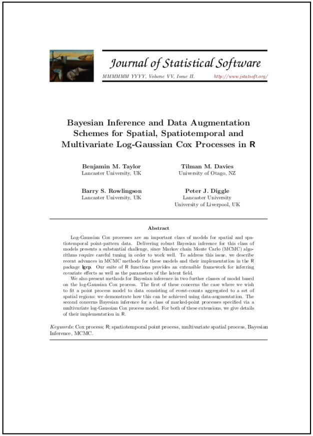
Abstract

This paper introduces an R package for spatial and spatio-temporal prediction and forecasting for log-Gaussian Cox processes. The main computational tool for these models is Markov chain Monte Carlo (MCMC) and the new package, **lgcp**, therefore also provides an extensible suite of functions for implementing MCMC algorithms for processes of this type. The modelling framework and details of inferential procedures are first presented before a tour of **lgcp** functionality is given via a walk-through data-analysis. Topics covered include reading in and converting data, estimation of the key components and parameters of the model, specifying output and simulation quantities, computation of Monte Carlo expectations, post-processing and simulation of data sets.

Keywords: Cox process, R, spatio-temporal point process.

1. Introduction

This article introduces a new R (R Core Team 2012) package, **lgcp**, for inference with spatial and spatio-temporal log-Gaussian Cox processes (LGCP). The work was motivated by applications in disease surveillance, where the major focus of scientific interest is on whether, and if so where and when, cases form unexplained clusters within a spatial region W and time-interval $[0, T]$ of interest. It will be assumed that both the location and time of each case is known, at least to a sufficiently fine resolution that a point process modelling framework is natural. In general, the aims of statistical analysis include model formulation, parameter estimation and spatio-temporal prediction. The **lgcp** package includes some functionality for parameter estimation and diagnostic checking, mostly by linkages with other R packages for



Journal of Statistical Software
MMMMMM YYYY, Volume VV, Issue II. http://www.jstatsoft.org/

Bayesian Inference and Data Augmentation Schemes for Spatial, Spatiotemporal and Multivariate Log-Gaussian Cox Processes in R

Benjamin M. Taylor
Lancaster University, UK

Tilman M. Davies
University of Otago, NZ

Barry S. Rowlingson
Lancaster University, UK

Peter J. Diggle
Lancaster University
University of Liverpool, UK

Abstract

Log-Gaussian Cox processes are an important class of models for spatial and spatiotemporal point-pattern data. Delivering robust Bayesian inference for this class of models presents a substantial challenge, since Markov chain Monte Carlo (MCMC) algorithms require careful tuning in order to work well. To address this issue, we describe recent advances in MCMC methods for these models and their implementation in the R package **lgcp**. Our suite of R functions provides an extensible framework for inferring covariate effects as well as the parameters of the latent field.

We also present methods for Bayesian inference in two further classes of model based on the log-Gaussian Cox process. The first of these concerns the case where we wish to fit a point process model to data consisting of event-counts aggregated to a set of spatial regions: we demonstrate how this can be achieved using data-augmentation. The second concerns Bayesian inference for a class of marked-point processes specified via a multivariate log-Gaussian Cox process model. For both of these extensions, we give details of their implementation in R.

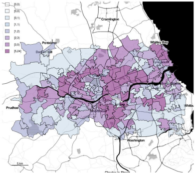
Keywords: Cox process; R; spatiotemporal point process, multivariate spatial process, Bayesian Inference, MCMC.

Aggregated Spatial Point Process Data

- Observe number of events T_i in each region A_i
- Augment the list of parameters, $\{\beta, \eta, Y\}$, with an additional variable N, the unobserved cell count (akin to X)
- Seek to sample from:

$$\pi(\beta, \eta, Y, N | T_1, \ldots, T_m).$$

- Use a Gibbs scheme, alternately sampling from $\pi(\beta, \eta, Y | N, T_{1:m})$ and $\pi(N | \beta, \eta, Y, T_{1:m})$

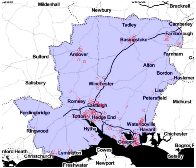


Spatiotemporal Point Process Data

Model:

$$
\begin{aligned}
X(s,t) &= \text{Poisson}[R(s,t)] \\
R(s,t) &= C_A \lambda(s,t) \exp\{Z(s,t)\beta + Y(s,t)\}
\end{aligned}
$$

- Similar in notation to spatial case: (s,t) instead of (s)
- Assume a separable spatiotemporal covariance function for Y:



$$\mathrm{cov}[Y(s_1,t_1),Y(s_2,t_2)] = \sigma^2 \exp\{-||s_2 - s_1||/\phi - |t_2 - t_1|/\theta\}.$$

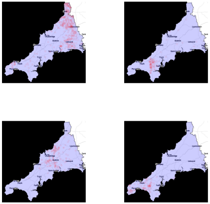
- θ determines the temporal correlation in the process Y.

Multivariate Spatial Point Process Data

Model for multitype process with K types, for $k \in 1, \ldots, K$,

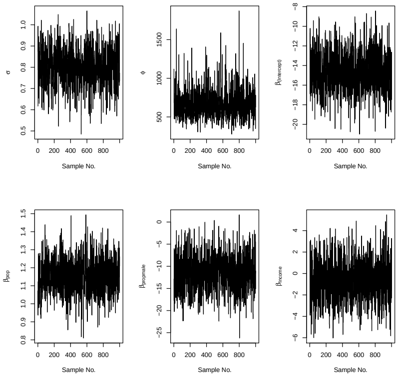
$$
\begin{aligned}
X_k(s) &= \text{Poisson}[R_k(s)] \\
R_k(s) &= C_A \lambda_k(s) \exp\{Z(s)_k \beta_k + Y_k(s) + Y_{K+1}(s)\}
\end{aligned}
$$

- $X_k(s)$ is the number of events of type k in the computational grid cell containing the point s
- Y_{K+1} captures areas of high or low intensity that are common to all types

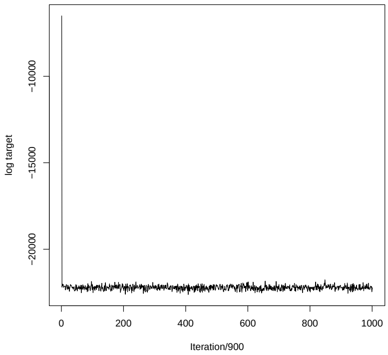


Convergence/Mixing Diagnostics

```
R> traceplots(lg)
```

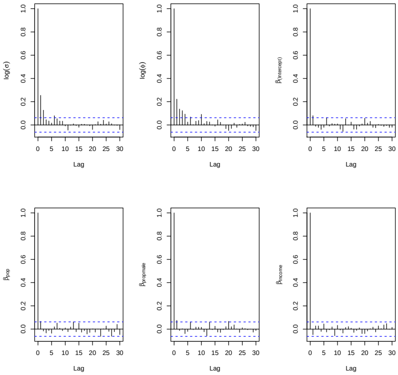



```
R> plot(ltar(lg),lty="s")
```

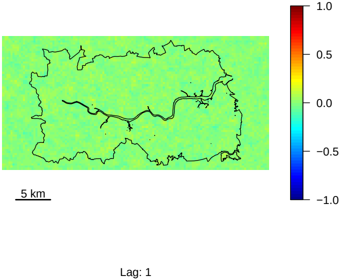


Convergence/Mixing Diagnostics

```
R> parautocorr(lg)
```

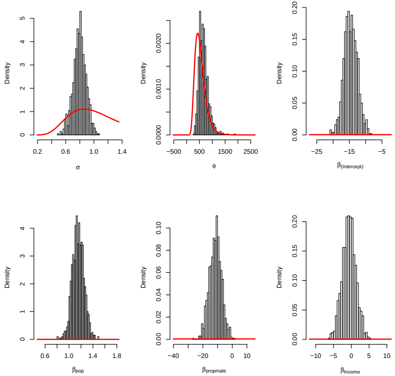



```
R> lagch <- c(1, 5, 15)
R> Sacf <- autocorr(lg, lagch,
+                   inWindow = NULL)
```



Plots of Prior and Posterior

```
R> priorpost(lg)
```



Tabulate Results

```
R> parsum <- parsummary(lg, LaTeX = TRUE)
R> require("miscFuncs")
R> latextable(parsum, rownames = rownames(parsum),
+       colnames = c("Parameter", colnames(parsum)), digits = 4)
```

Parameter	median	lower 95% CRI	upper 95% CRI
σ	0.7999	0.6033	0.9695
ϕ	637.1	389.3	1098
$\exp(\beta_{(Intercept)})$	4.111×10^{-7}	8.735×10^{-9}	3.019×10^{-5}
$\exp(\beta_{pop})$	3.162	2.633	3.84
$\exp(\beta_{propmale})$	1.328×10^{-5}	3.937×10^{-9}	4.62×10^{-2}
$\exp(\beta_{Income})$	0.5449	1.425×10^{-2}	23.05
$\exp(\beta_{Employment})$	52.73	0.2343	9981
$\exp(\beta_{Education})$	0.9961	0.9797	1.012
$\exp(\beta_{Barriers})$	0.982	0.9594	1.007
$\exp(\beta_{Crime})$	0.9034	0.6956	1.223
$\exp(\beta_{Environment})$	1.015	0.9967	1.035

```
R> textsummary(lg)
```

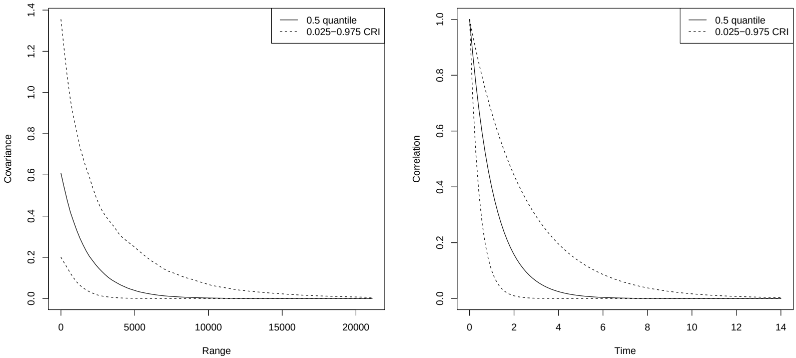
A summary of the parameters of the latent field is as follows. The parameter σ had median 8×10^{-1} (95% CRI 0.603 to 0.97) and the parameter ϕ had median 637 (95% CRI 389 to 1098).

The following effects were found to be significant: each unit increase in `propmale` led to a reduction in relative risk with median 1.33×10^{-5} (95% CRI 3.94×10^{-9} to 4.62×10^{-2}); each unit increase in `pop` led to a increase in relative risk with median 3.16 (95% CRI 2.63 to 3.84).

The remainder of the . . .

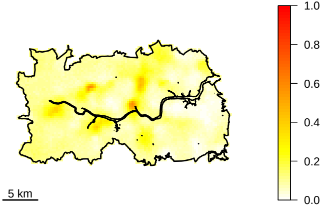
Posterior Correlation Function

```
R> postcov(lg)
```



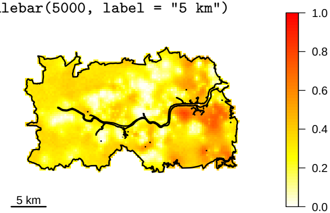
Exceedance Probabilities

```
R> ep <- exceedProbs(c(1.5, 2, 5, 10))
R> ex <- lgcp:::expectation.lgcpPredict(lg, ep)
R> plotExceed(ex[[1]], "ep", lg,
+       zlim = c(0,1), asp = 1,
+       axes = FALSE, xlab = "",
+       ylab = "", sub = "")
R> scalebar(5000, label = "5 km")
```



$\mathbb{P}\{\text{Relative Risk} > 2\}$

```
R> sp <- exceedProbs(c(2/3, 1/2, 1/5, 1/10),
+                                    direction = "lower")
R> su <- lgcp:::expectation.lgcpPredict(lg, sp)
R> plotExceed(su[[1]], "sp", lg,
+       zlim = c(0, 1), asp = 1,
+       axes = FALSE, xlab = "",
+       ylab = "", sub = "")
R> scalebar(5000, label = "5 km")
```



$$\mathbb{P}\left\{\text{Relative Risk} < \frac{1}{2}\right\}$$

Conditional Probabilities, Segregation Probabilities

Conditional Probabilities:

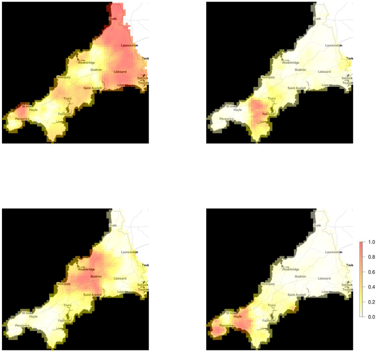
$$p_k = \mathbb{P}\{\text{a case at location } s \text{ is of type } k | \text{there is a case at } s\}$$

Segmentation Probabilities:

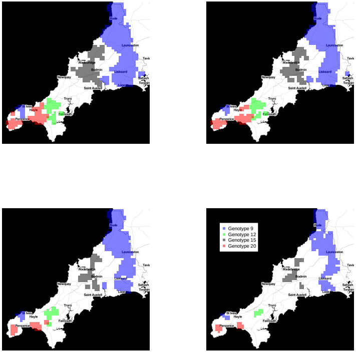
- Let $A_k(c, q)$ denote the set of locations x for which $\mathbb{P}\{p_k(x) > c|X\} > q$.
- c is the "dominance probability"
- $c, q \to 1 \Rightarrow A_k(c, p) \to \emptyset$
- happens more slowly in a highly segregated pattern compared with a weakly segregated one
- Compute $\mathbb{P}\{p_k(x) > c|X\}$

Conditional Probabilities, Segregation Probabilities

`R> condProbs(lg)`



`R> segProbs(lg,domprob=0.8)`



Conclusions

- Four classes of log-Gaussian Cox processes
- A software package, `lgcp`, for MCMC-based inference for these models
- . . . more functionality . . .