Bayesian inference for spatiotemporal point processes

(driven by multivariate Gaussian processes)

Dani Gamerman

Departamento de Métodos Estatísticos - IM

Universidade Federal do Rio de Janeiro

Joint work with F. Gonçalves, G. Moreira, M. Paez, J. A. Pinto Jr and E. Reis

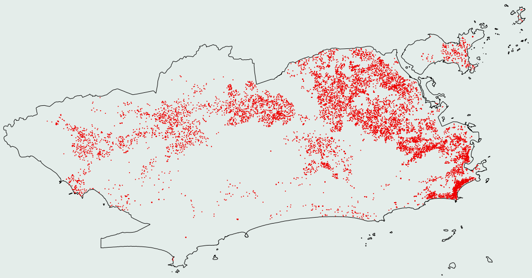
Point Process Workshop - TAMU, 29 September 2018

Content

1. Introduction

main types of spatial(-temporal) data:

- Geostatistics or continuous data
- areal data
- point pattern



This talk is concentrated on the latter

1.1. Models for spatial point patterns

The most usual model is the Poisson process

Governed by the intensity function $\lambda(s)$, for $s \in S$

[Typically, $S \subset R^d$.]

Notation: $Y \sim PP(\lambda)$

Also, Y can be identified with its observed locations $\{s_i\}_{i=1}^N$

Problems:

- object of interest (IF λ) is infinite-dimensional
- likelihood $l(\lambda; y) = \prod_{i=1}^N \lambda(s_i) \ \exp\left\{- \int_S \ \lambda(s) ds\right\}$

  depends on the entire IF and is unavailable analytically

Spatio-temporal Poisson processes

Assume discrete time over $\mathcal{T} = \{t_1, t_2, ...\} \equiv \{1, 2, ...\}$

[Continuous time is basically an added dimension over S]

Thus, $Y = (Y_1, Y_2, ...)$ with $Y_t \sim PP(\lambda_t)$, for $t \in \mathcal{T}$

Likelihood:

$$l(\lambda; y) = \prod_t \prod_{i=1}^{N_t} \lambda_t(s_{i,t}) \quad \exp\left\{ -\sum_t^{\infty} \int_S \lambda_t(s) ds \right\}$$

Cox processes (Cox, 1955):

- space only: λ is random
- space-time: $\{\lambda_t, \ t \in \mathcal{T}\}$ is random

Important special case: log Gaussian Cox process (LGCP)

$\log \lambda \sim GP$ (Moller et al, 1998)

Observation: Gaussian processes (GP)

$\eta \sim GP$ over $S \iff$

$\forall r, \forall \{s_1, ..., s_r\} \subset S, \ (\eta(s_1), ...\eta(s_r)) \sim MN$

1.2. Regression for point patterns

Standard approach for spatial models:

separate predictor into space(-time) component + regressor effect

PP: $\lambda_t(s) = g[\ \beta_{0,t}(s) + W'\beta\ ]$

[**LGCP**: $g =$ exponential]

Some applications show heterogeneity: separation is not reasonable

$\rightarrow$ space/space-time varying regression coefficients

continuous data: Gelfand, Banerjee and Gamerman (2005)

areal data: Gamerman, Moreira and Rue (2003)

Extension for point patterns

Regression coefficents vary over space-time: $\beta \;\to\; \beta_t(s)$

Extra complexity is only computational but....

much smaller than introduced by ∞ dimensionality of IF

From now on, β_0 incorporated into β (and 1 into W) $\to \lambda_t(s) = g[\,W'\beta_t(s)\,]$

More details about regressors in W:

- may vary over space/time (Benes et al, 2005) $\to$ easy

  [although it may bring in identifiability issues]

- may vary over configurations v (Liang et al, 2008; Diggle et al, 2010)

  $W \to W_v$ and $\lambda_t(s) \to \lambda_{t,v}(s)$

1.3. Dynamic Gaussian processes

Typically $\beta_t(s)$ vary smoothly over space-time

Smoothness over space provided by GP

Smoothness over time is also needed

One possible solution: DGP (Gamerman, 2010)

$$\beta_{t'} \;=\; G_{t,t'}\,\beta_t + w_t, \quad \text{where } w_t \sim GP$$

$$\text{with } \beta_1 \;\sim\; GP$$

Notation: $\beta \sim DGP$

Accomodates stochastic SV trend, seasonality, ...

Markov $\rightarrow$ Sparcity $\rightarrow$ computations can be improved

Continuous-time version in Brix and Diggle (2001)

More convolved forms in Wikle and Cressie (1999).

Both require approximations to work

We would like solutions to be as exact as possible

2. Model and inference problem

2.1. Model

Likelihood: $Y = (Y_1, Y_2, ...)$ with $Y_t \sim PP(\lambda_t)$, for $t \in \mathcal{T}$

Link function: $\lambda_t(s) = g[\ W'\beta_t(s)\ ]$

[If data varies over individual configurations $v \rightarrow$

$Y_t = \{Y_{t,v}\}$ with $Y_{t,v} \sim PP(\lambda_{t,v})$ and $\lambda_{t,v}(s) = g[W_v'\ \beta_t(s)]$, for $v \in \mathcal{V}$]

Prior: $\beta \sim DGP$

Hyperprior: $\theta \sim p(\theta)$

θ - all other unknowns (DGP hyperparameters)

This model includes many other models previously considered

Special cases:

- $\lambda(s) = g[\ \beta_0(s) + W'\beta\ ]$ (Benes et al., 2005)
- $\lambda_v(s) = g[\ \beta_0(s) + W_v'\ \beta\ ]$ (Liang et al, 2008; Diggle et al, 2010)
- $\lambda_t(s) = g[\ \beta_{0,t}(s)\ ]$ (Reis et al., 2013)
- $\cdots$

2.2. Inference (problem)

Prior density does not exist due to ∞ dimension of β

NB: finite dimension version do exist

Likelihood cannot be computed $\rightarrow$ Posterior cannot be computed!

Some solutions involve:

- parametric forms for the IF and/or β
- approximation of the integral in the likelihood

Solutions should be as exact as possible and:

- be model-based
- retain the IF as is, without any representation

3. Discretization

(Pinto Jr et al, 2015)

Some situations require regional rather than point-wise approach

In these cases, little is lost by assuming piece-wise constancy

$\lambda_t(s) = \lambda_{i,t}$, for $s \in R_i \subset S$

$\{R_i\}$ form a partition of S, with $a_i = vol(R_i)$

$card(\{R_i\})$ depends on the analyst

Most common link: exponential

Discretized model

Likelihood $\rightarrow l(\lambda; y) = \prod_t \prod_i \lambda_{i,t}^{N_{i,t}} \quad \exp\{-\sum_t \sum_i a_i \lambda_{i,t}\}$

[Basically equivalent to $N_{i,t} \sim Poisson(a_i \lambda_{i,t})$, $\forall (i,t)$]

Link function $\rightarrow \lambda_{i,t} = \exp\{W' \beta_{i,t}\}$

[$l(\lambda; y) = \prod_v l(\lambda_v; y)$ and $\lambda_{i,t,v} = \exp\{\, W_v' \, \beta_{i,t} \,\},$ with individual covariates]

Set-up $\rightarrow$ finite-dimensional

with parameters $\beta = (\beta_1, \beta_2, \ldots)$ where $\beta_t = \{\beta_{i,t}, \forall i\}$

Prior $\rightarrow \beta_{t'} = G_{t,t'} \, \beta_t + w_t,$ where $w_t \sim MN$

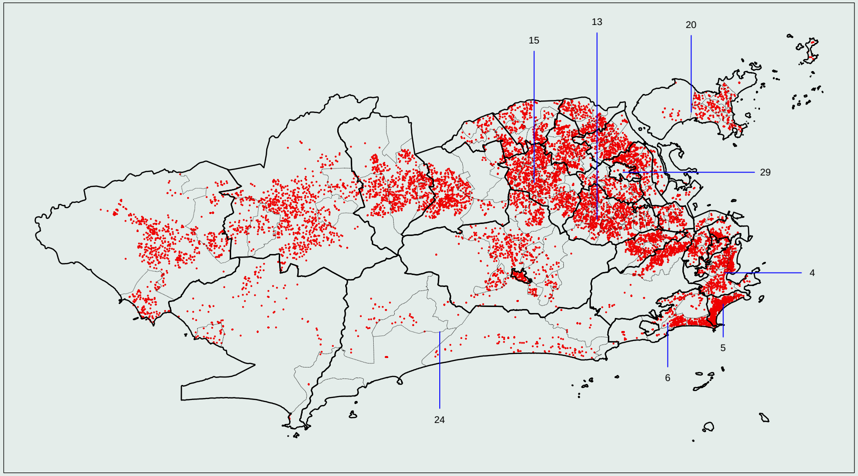
MN specification derived from GP, based on centroids of $\{R_i\}$

Comments:

- Similar to dynamic areal data (Vivar and Ferreira, 2009)
- MCMC for non-Gaussian TS data may be used (Gamerman, 1998; Fruhwirth-Schnatter and Wagner, 2006)
- Standard software (`Bugs`, `Jags`, etc) ok if T is not large
- Computational cost depend on $card(\{R_i\})$

Illustration (Pinto Jr et al, 2015)

Cardiovascular disease deaths in Rio de Janeiro, 2002-2007



Municipal health policy is organized by AR's or boroughs

No reason to consider within-region variation → discretization

Time span too small to capture temporal changes → purely spatial analysis

Models used

IF is function of explanatory variables:

- individual: age, gender, education and marital status
- spatial: socio-economic level (HDI) of the region

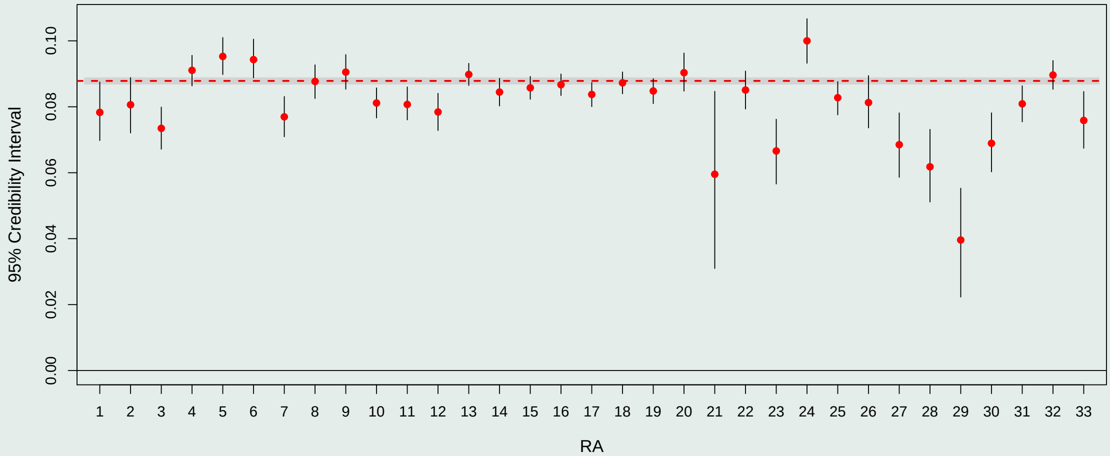
spatial effects were fixed, individual effects were SV

HDI were to handle effect of socio-economic level

$\rightarrow$ no residual effect of geography over intercept β_0

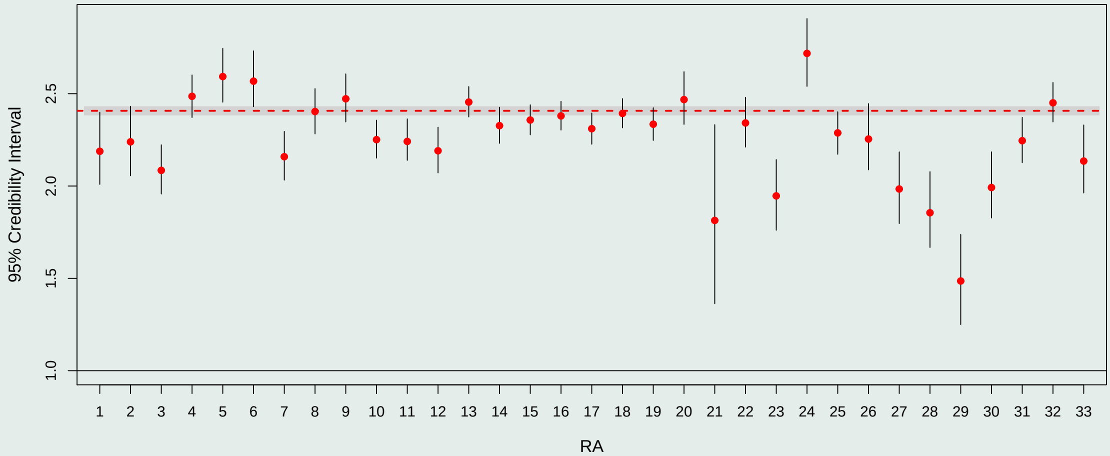
Offsets used to standardize over demography

Age effect: posterior median and 95% CI of coefficient



- space varying effects (SVE) over AR's and fixed effects (FE)
- largest coefficients: wealthier regions (4, 5, 6 and 24)
- smallest coefficients: Complexo do Alemão, a slum-town (region 29)

Relative risk of + 10 years: median and IC 95%



- variation of RR's: ∼ 50% (Alemão) to ∼ 270% (wealthier AR)
- FE model : ∼ 240%

Other results

Effects of individual covariates: smaller spatial variation

Some significant regional effects but not all

More evidence about relevance of SVE model

HDI coefficient highly relevant, as expected

SVE model: -10 and CI(95%)$=[-10.2, -9.7]$

There is still significant spatial variation of β_0

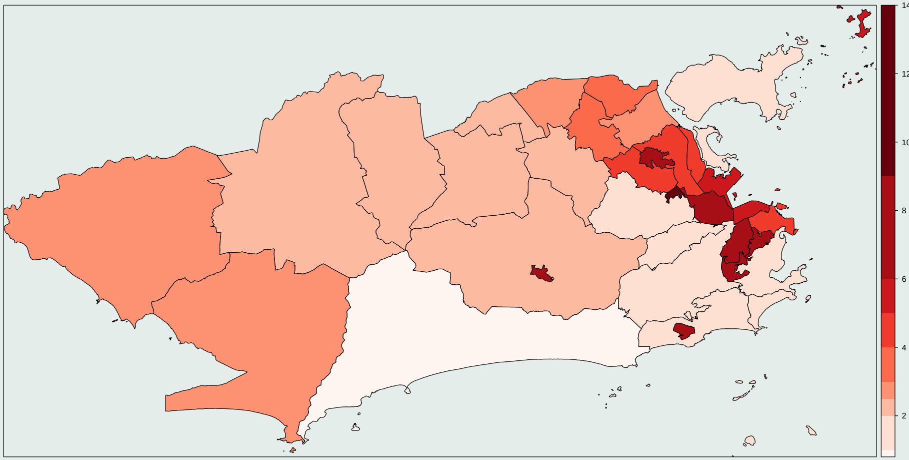
SVE model: better *fit* (DIC) than FE

SVE over AR's: much better *fit* than SVE over boroughs

Relative risk: *protective* configuration

(young female, living with someone and with education)

- comparisons against region 6, wealthiest
- slum-towns stand out: largest RR's (~ 14) but wider C.I.



4. Augmentation (Gonçalves and Gamerman, 2018)

Situation seems unsolvable directly

Solution: Poisson thinning (Lewis and Shedler, 1979)

Algorithm to simulate from $Y \mid \lambda \sim PP(\lambda)$ - space only

0) set $\lambda^* = \sup_s \lambda(s)$

1) sample $X \sim PP(\lambda^*)$ $\leftarrow$ augmented data

[sample $K \sim Poisson(\lambda^* vol(S))$ and

distribute K points $\{s_k\}_{k=1}^K$ uniformly over S.]

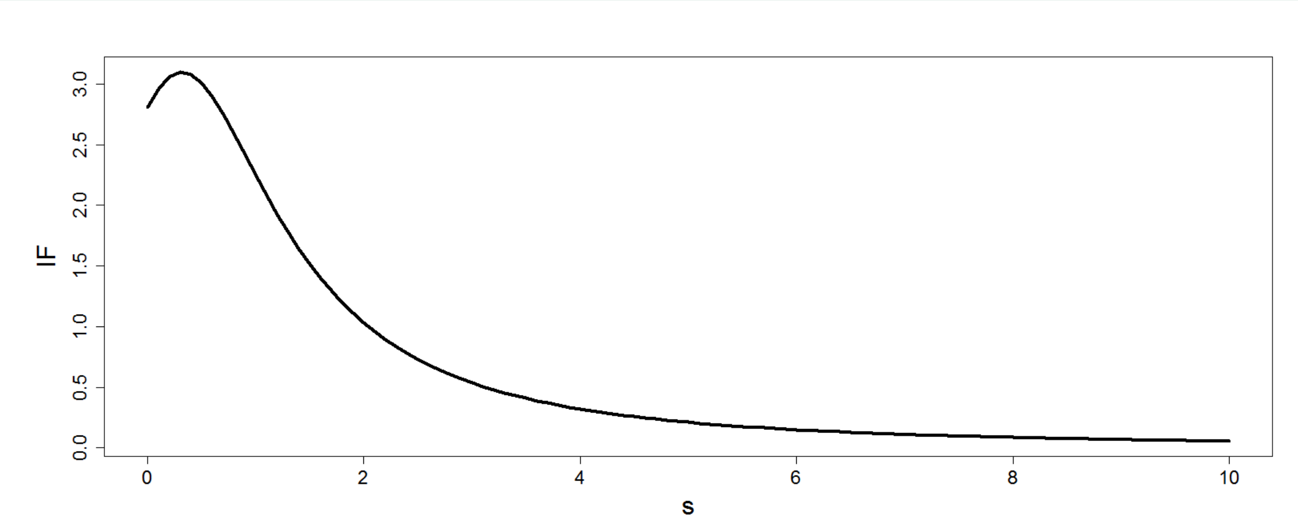
2) retain each point s_k with probability $\lambda(s_k)/\lambda^*, \forall k$

3) $\{s_i\}_{i=1}^N$ retained points are a sample from Y

Obs: $X \setminus Y = \tilde{Y} \sim PP(\lambda^* - \lambda)$, $\tilde{Y}$ independent of Y and $\tilde{y} \equiv \{s_i\}_{i=N+1}^K$

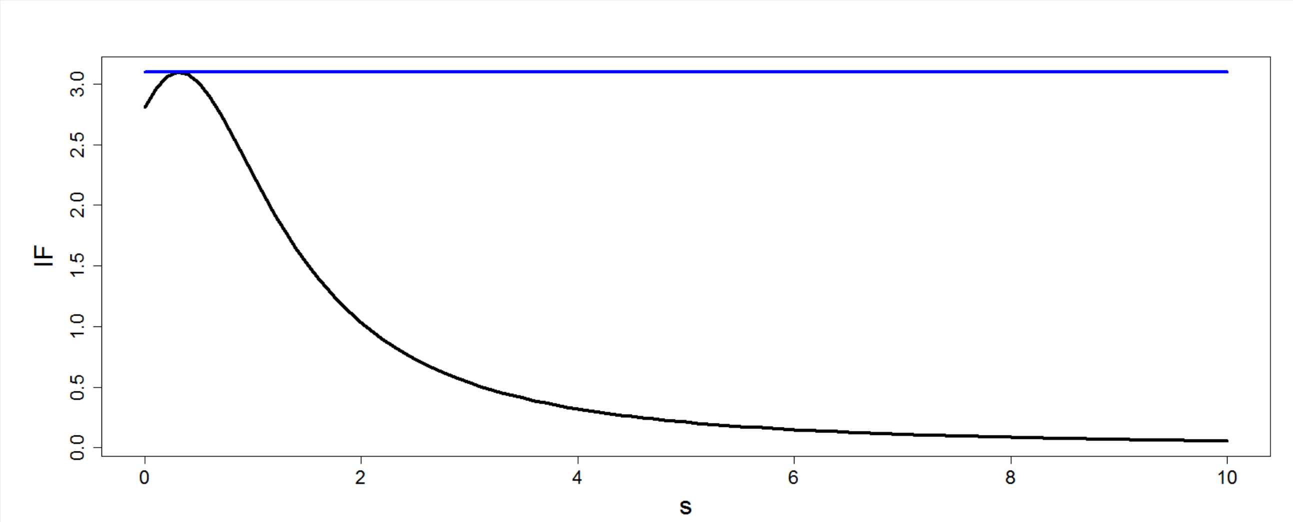
Graphical version of the algorithm

start: IF over the real line



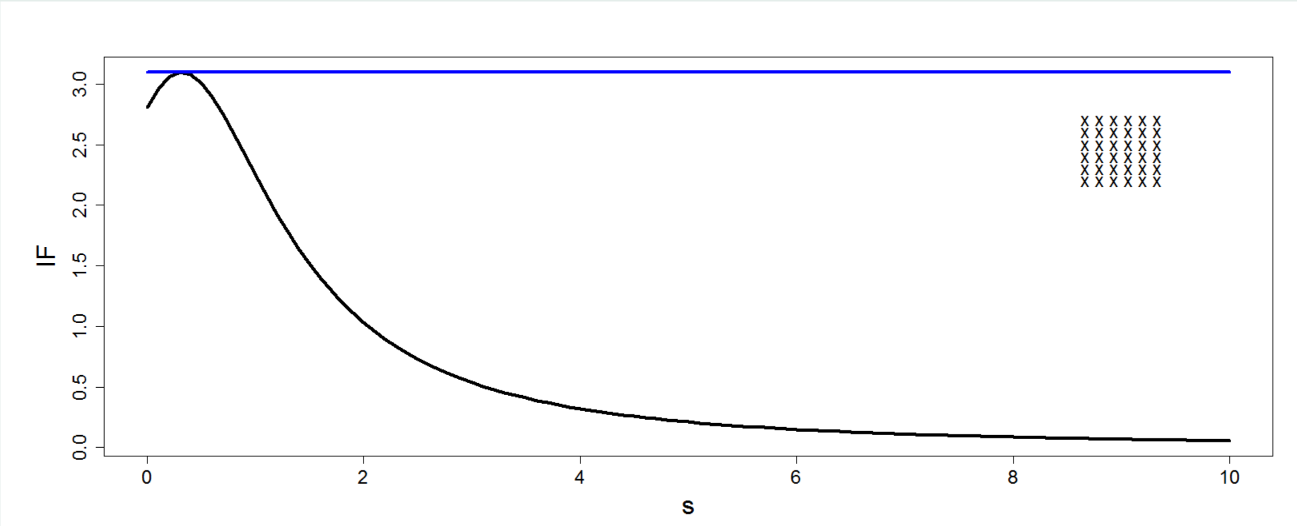
Graphical version of the algorithm

step 0: setting upper limit λ^*



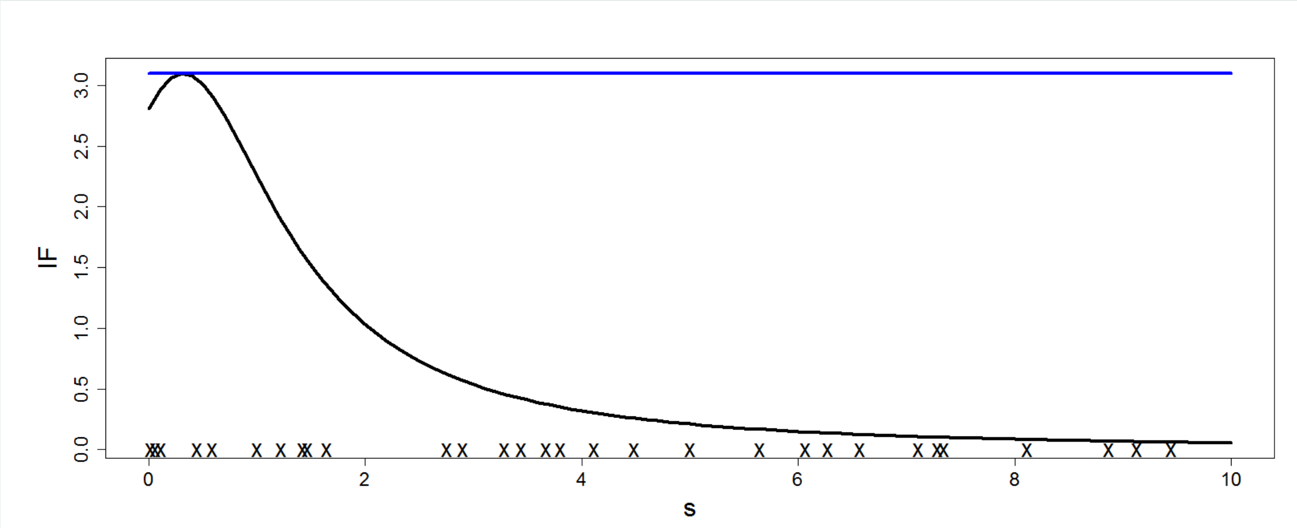
Graphical version of the algorithm

step 1a: drawing K, the number of locations



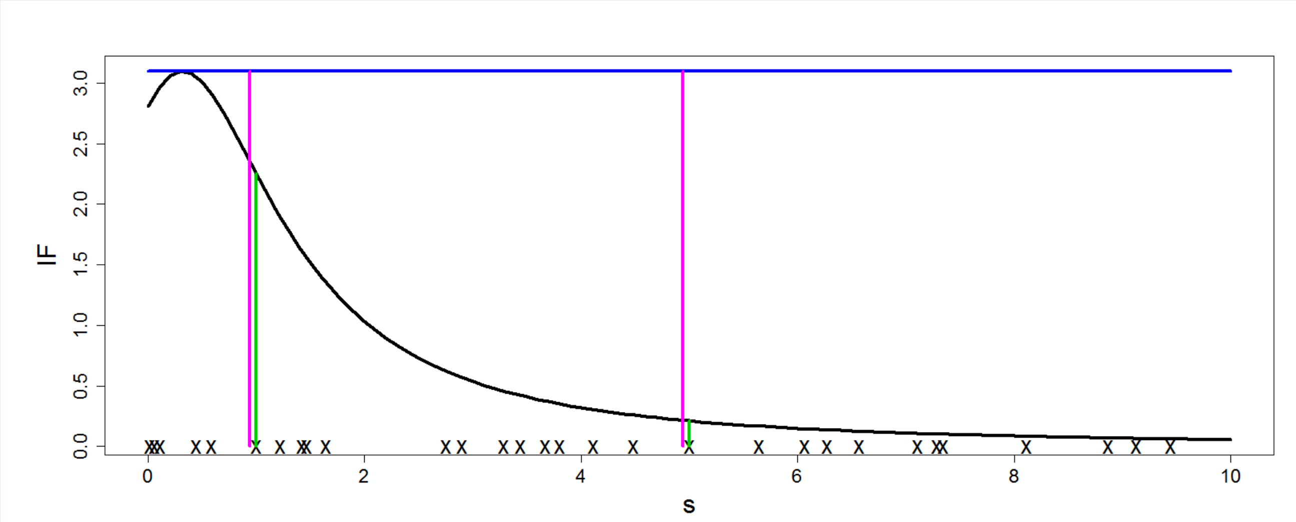
Graphical version of the algorithm

step 1b: distributing the locations $\{s_k\}$



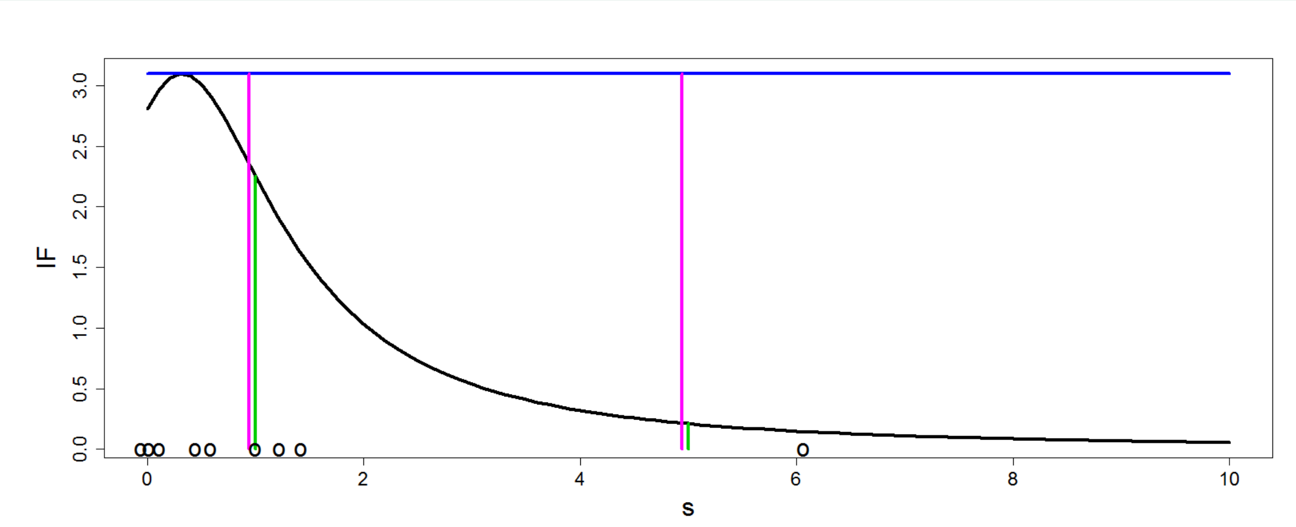
Graphical version of the algorithm

step 2: evaluating thinning probabilities



Graphical version of the algorithm

step 3: drawing the thinned locations $\{s_i\}$



Analytic version of the algorithm

$$\pi(K, x, y \mid \lambda) = \pi(K \mid \lambda^*) \times \pi(x \mid K) \times \pi(y \mid x, \lambda)$$

$$= e^{-\lambda^* vol(S)} \ \frac{[vol(S)\lambda^*]^K}{K!} \times \left[\frac{1}{vol(S)}\right]^K$$

$$\times \prod_{i=1}^{N} \frac{\lambda(s_i)}{\lambda^*} \times \prod_{j=1}^{K-N} \left[1 - \frac{\lambda(s_j)}{\lambda^*}\right]$$

Notation: x and y are the observed values of X and Y

Extension to general space-time case

Algorithm to simulate from $Y \mid \lambda \sim PP(\lambda)$

Recall that $Y \mid \lambda \sim PP(\lambda) \Leftrightarrow Y_t \mid \lambda_t \sim PP(\lambda_t), \forall t$

Simply repeat algorithm for space only, $\forall t$

$\{\{s_{i,t}\}_{i=1}^{N_t}, \forall t\}$ retained points are a sample from Y

$\{\{s_{i,t}\}_{i=N_t+1}^{K_t}, \forall t\}$ discarded points are sample from $\tilde{Y}$

$\tilde{Y} = (\tilde{Y}_1, \tilde{Y}_2, ...)$ where $X_t \setminus Y_t = \tilde{Y}_t \sim PP(\lambda_t^* - \lambda_t), \forall t$

Analytic version of the algorithm: general case

$$
\begin{aligned}
\pi(\{K_t\}, x, y \mid \lambda) &= \prod_t \{\pi(K_t \mid \lambda_t^*) \times \pi(x_t \mid K_t) \times \pi(y_t \mid x_t, \lambda_t)\} \\
&= \prod_t \Bigg\{ e^{-\lambda_t^* vol(S)} \frac{[vol(S)\lambda_t^*]^{K_t}}{K_t!} \times \left[\frac{1}{vol(S)}\right]^{K_t} \\
&\qquad \times \prod_{i=1}^{N_t} \frac{\lambda_t(s_{i,t})}{\lambda_t^*} \times \prod_{j=1}^{K_t - N_t} \left[1 - \frac{\lambda_t(s_{j,t})}{\lambda_t^*}\right] \Bigg\}
\end{aligned}
$$

Crucial point: integral disappeared and likelihood depends on a finite subset of λ

4.1. Model redefined - space only

Redefinition needed due to presence of upper bound λ^* on IF

Likelihood: $Y \sim PP(\lambda)$

Link function: $\lambda(s) = \lambda^* \, g[W'\beta(s)]$

Prior for β: $\beta \sim GP$

Prior for λ^*: $\lambda^* \sim \pi_\lambda$

Hyperprior: $\theta \sim \pi(\theta)$

θ - all other unknowns (GP hyperparameters)

Comments:

.

- link g must be bounded to $[0,1] \to$ any d.f works
  - common choice: logistic (or sigmoidal)
  - we chose probit: $g = \Phi$, the d.f. of $N(0,1)$
  - indistinguishable but much better computationally
- obvious option for π_λ: $G(a_\lambda, b_\lambda)$ - conjugate

Notation: β_N - latent GP β at N locations of Y

Similar notation for W_N, β_K, W_K, β_{K-N}, W_{K-N}, ...

Inference

Unknown quantities: $\tilde{Y}$ (and K), λ^*, β, θ

(augmented) likelihood: $l\left(K, \tilde{y}, \beta, \lambda^*; y\right) = \pi\left(K, x, y \mid \lambda^*, \beta\right)$

$$
\begin{aligned}
l\left(K, \tilde{y}, \beta, \lambda^*; y\right) &= e^{-\lambda^* vol(S)} \frac{[\lambda^*]^K}{K!} \\
&\times \prod_{i=1}^{N} \Phi[W_i' \beta(s_i)] \prod_{j=1}^{K-N} \Phi[-W_j' \beta(s_j)]
\end{aligned}
$$

NB: likelihood depends on $\beta = (\beta_K, \beta_{-K})$ only through β_K!!!

$\rightarrow l\left(K, \tilde{y}, \beta, \lambda^*; y\right) = l\left(K, \tilde{y}, \beta_K, \lambda^*; y\right)$

Set-up $\rightarrow$ finite-dimensional (as with discretization)

Inference (cont.)

(joint) prior: $\pi\left(\beta_K, \lambda^*, \theta\right) = \pi_{GP}\left(\beta_K \mid \theta\right) \pi_\lambda\left(\lambda^*\right) \pi(\theta)$

where $\beta_K \mid \theta \sim MN$, with moments given by GP

(joint) posterior: $\pi\left(K, \tilde{y}, \beta_K, \lambda^*, \theta \mid y\right)$

$$\pi\left(K, \tilde{y}, \beta_K, \lambda^*, \theta \mid y\right) \propto l\left(K, \tilde{y}, \beta_K, \lambda^*; y\right) \times \pi\left(\beta_K, \lambda^*, \theta\right)$$

Computation

Unknown quantities: $\tilde{Y}, \beta_K, \lambda^*, \theta$ (and β_{-K})

Performed via block MCMC: samples from full conditional

- $\tilde{Y}$

$\tilde{Y}$, Y independent $\rightarrow [\, \tilde{Y} \mid \cdot \,] = [\, \tilde{Y} \mid \lambda^*, \beta \,] = PP(\, \lambda^* \Phi[-W(s)'\beta(s)] \,)$

sample $\tilde{y}$ obtained by thinning from $X \sim PP(\lambda^*)$ but requires β_{-K}

- β_K [Full conditional is multivariate skew normal $\rightarrow$ Gibbs step]

$$\pi(\beta_K \mid \cdot) \propto \prod_{i=1}^{N} \Phi[W_i'\beta(s_i)] \prod_{j=1}^{K-N} \Phi[-W_j'\beta(s_j)] \; \pi(\beta_K \mid \theta)$$

- λ^* [Gamma prior $\rightarrow$ conditional conjugacy]

- θ [Metropolis-Hastings steps]

Inference for β_{-K}

Required for inference about IF and other functionals

Example: $\Lambda(C) = \int_C \lambda(s) ds$

$\{\tilde{s}_1, ..., \tilde{s}_G\} \subset S$ - new set of locations

Denote $\beta_G = (\beta(\tilde{s}_1), ..., \beta(\tilde{s}_G))$

Extended set of unknowns: $K, \tilde{Y}, \beta_K, \beta_G, \lambda^*, \theta$

Full conditional for β_G is $\pi(\beta_G \mid \beta_K, \theta)$

Can be trivially added as an extra step to MCMC

Just like your standard kriging in (Bayesian) Geostatistics

Comments

- other MCMC blocking schemes are possible

Example: $(K, \tilde{Y}, \beta_{K-N})$, β_N, λ^* and θ

- estimation of IFs via augmentation was introduced by Adams et al (2009)

they did not consider time, covariates, ST variations, ...

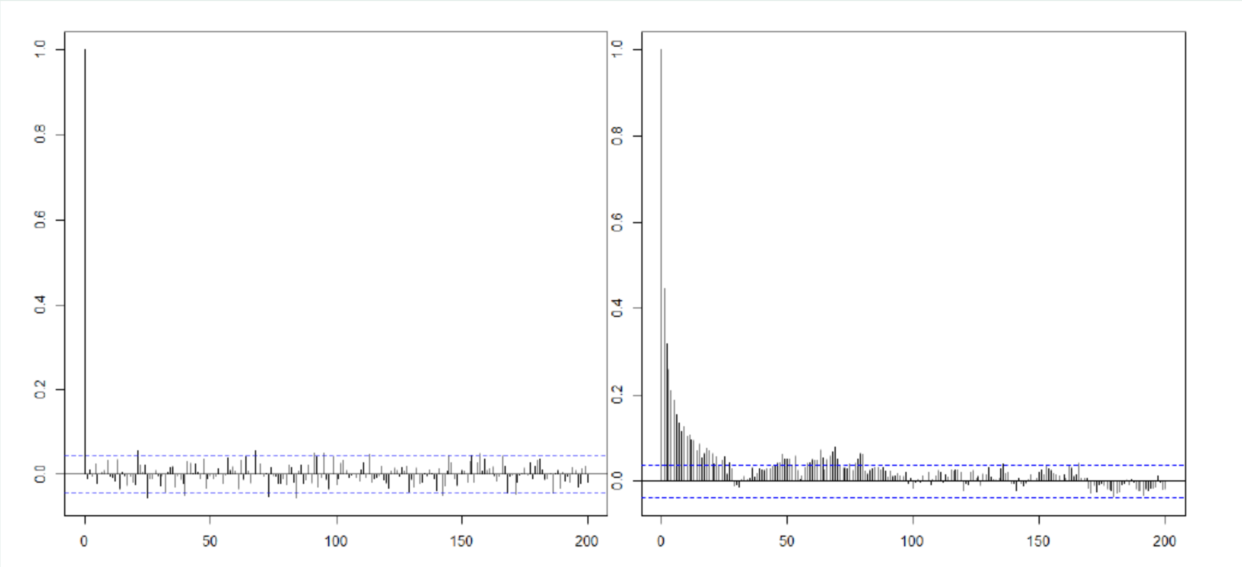
they use Hamiltonian MC

Comments (cont.)

- G & G (2018): very good mixing despite Gibbs steps

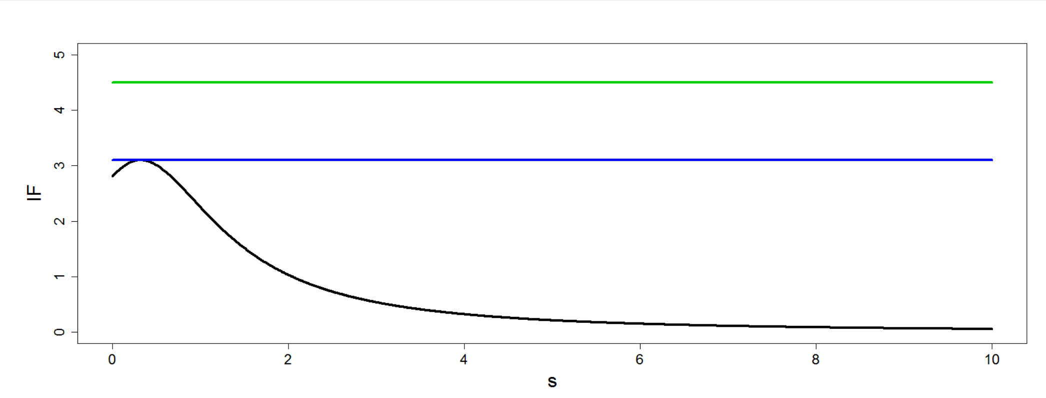
ACF of -2 log posterior density

G & G (2018) **Adams et al. (2009)**



Comments (cont.)

- any $\lambda^\dagger > \lambda^*$ is also valid but is less efficient



extra care must be exercised when setting priors

- λ with substantial variation in magnitude $\rightarrow K \gg N$

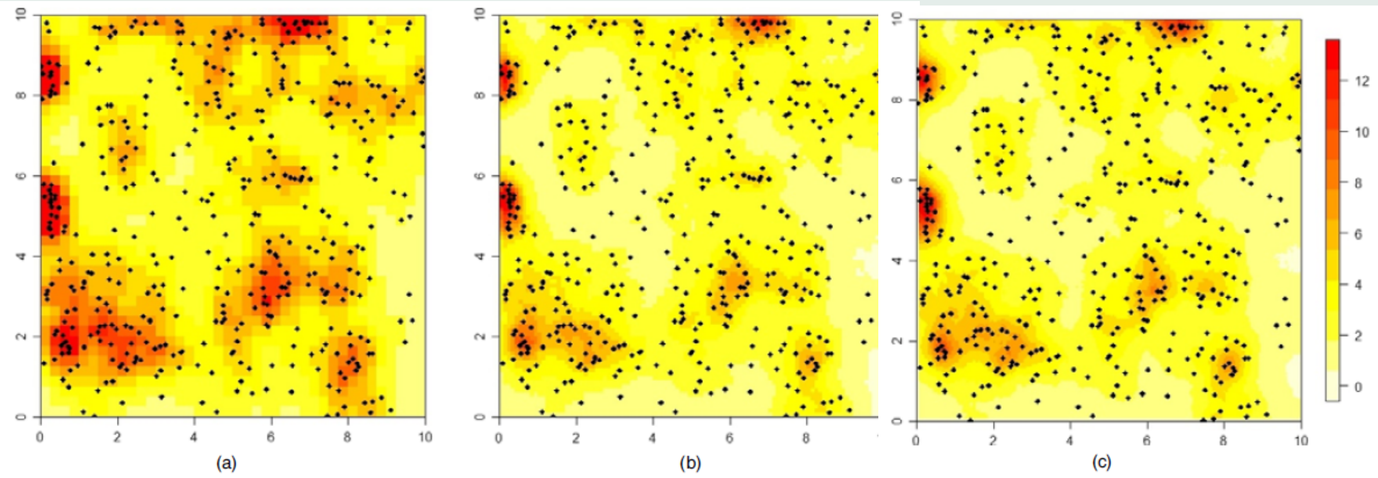
only 20% of the drawn points above are "retained"

augmentation becomes computationally inefficient

Illustration

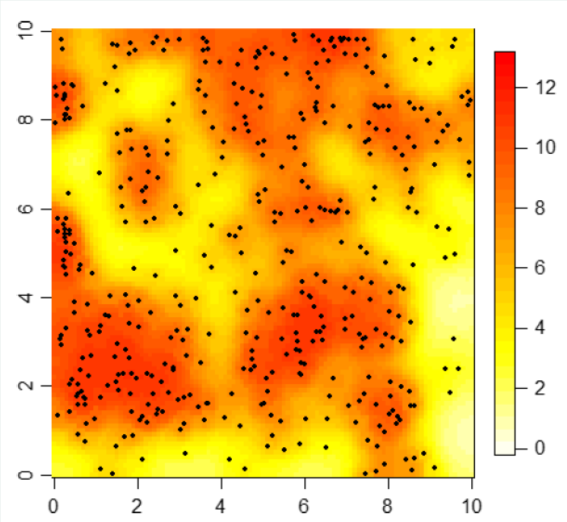
Oaks in Lansing Woods, USA (Baddeley et al, 2015)

Discretized IF - `lgcp` (Taylor et al, 2013)



No. of regions: (a) - 1.600; (b) - 10.000; (c) - 40.000

Estimated IF - continuous estimation



IF seems smoother in continuous space

$E[\Lambda([0,4]^2) \mid y] = 126.5$ $(se_{MC} = .47\%)$, real N= 126

4.2. Model redefined - general space-time case

Redefinition needed due to presence of upper bounds $\{\lambda_t^*, \forall t\}$ on IF

Likelihood: $Y_t \sim PP(\lambda_t)$, for $t \in \mathcal{T}$

Link function: $\lambda_t(s) = \lambda_t^* \; \Phi[W'\beta_t(s)]$

Prior for $\{\beta_t, \forall t\}$: $\beta \sim DGP$

Prior for $\{\lambda_t^*, \forall t\}$: $(\lambda_1^*, \lambda_2^*, ...) \sim \pi_\lambda$

Hyperprior: $\theta \sim \pi(\theta)$

Options for π_λ:

- equal: $\lambda_t^* = \lambda^*$, $\forall t$
- independent: $\prod_t \pi(\lambda_t^*)$
- beta evolution model (Gamerman et al, 2013)

Inference

Unknown quantities: $\{K_t\}, \tilde{Y}, \beta, \{\lambda_t^*\}, \theta$

(augmented) likelihood:

$$l\left(\{K_t\}, \tilde{y}, \beta, \{\lambda_t^*\}; y\right) = \; = \prod_t l\left(K_t, \tilde{y}_t, \beta_{K_t}, \lambda_t^*; y\right)$$

NB: likelihood depends on $\beta = (\beta_K, \beta_{-K})$ only through $\beta_K = \{\beta_{K_t}\}$!

(joint) prior: $\pi\left(\beta_K, \{\lambda_t^*\}, \theta\right)$

$$\pi\left(\beta_K, \{\lambda_t^*\}, \theta\right) = \pi_{DGP}\left(\beta_K \mid \theta\right) \pi_{BEM}\left(\{\lambda_t^*\}\right) \pi(\theta)$$

(joint) posterior: $\pi\left(\{K_t\}, \tilde{y}, \beta_K, \{\lambda_t^*\}, \theta \mid y\right)$

Other quantities of interest:

- inference for β_{-K}
- prediction beyond observed times

Computation

Unknown quantities: $\tilde{y}$, $\beta_K, \{\lambda_t^*\}, \theta$

Performed via MCMC: full conditionals basically as before

Only relevant differences are:

- β_K

As before (but sparcity may be used to improve computations)

- $\{\lambda_t^*\}$

BEM prior $\rightarrow$ easy sampling from full conditional

non-Gaussian FFBS (Gamerman et al, 2013)

- β_{-K} and prediction for future times

both straightforward sampling from full conditional

Illustrations

1. New Brunswick fires (Baddeley et al, 2015)

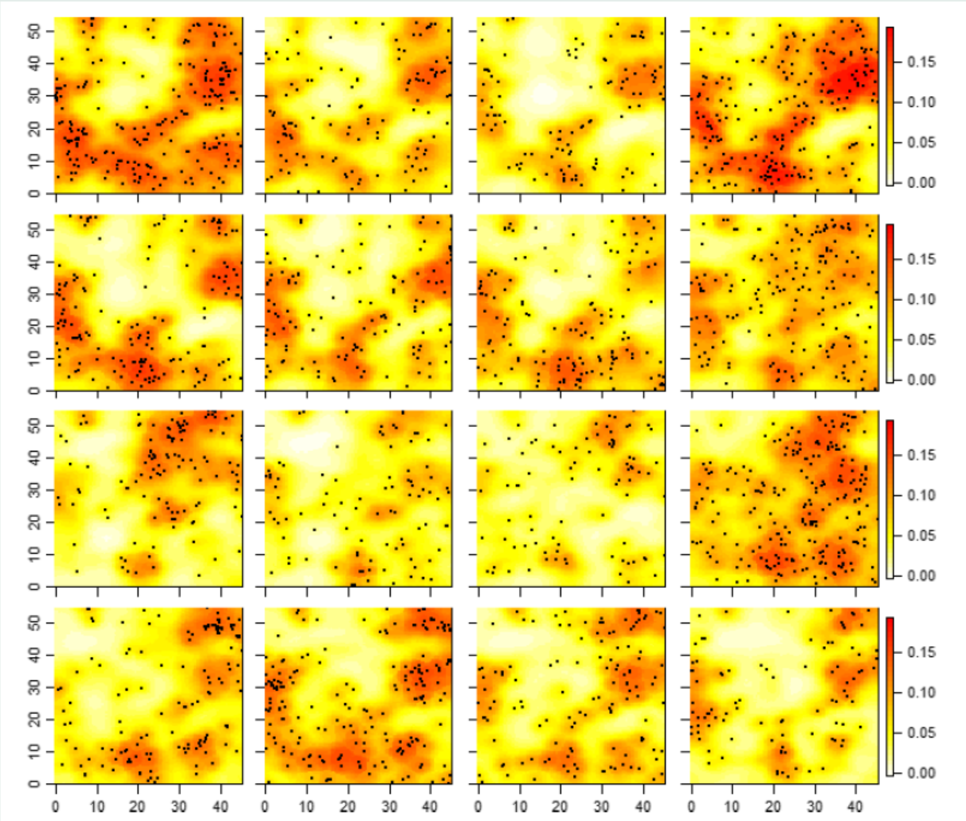
Map of the study region

Data aggregated over space

Year	*Number of fires*
1987	216
1989	120
1990	102
1991	211
1992	155
1993	123
1994	136
1995	169
1996	122
1997	94
1998	86
1999	224
2000	140
2001	194
2002	127
2003	94

- ST variation of IF
- No covariates included

New Brunswick fires - estimated IF



Illustrations (cont.)

2. Simulated ST data with seasonality

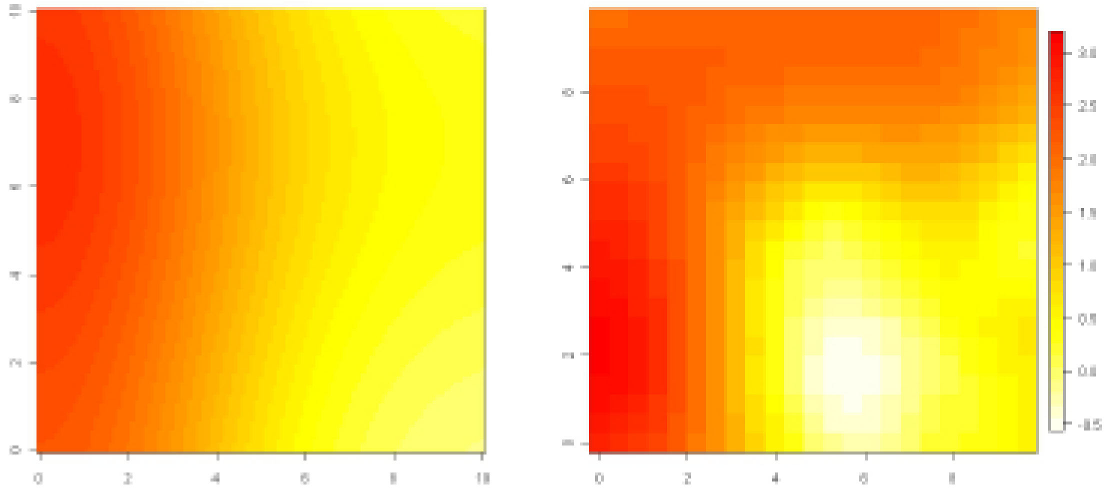
Model: $\lambda_t(s) = \lambda^* \, \Phi[\, \beta_{0,t}(s) + \beta_1(s) \cos(2\pi t/p + \phi) \,]$

Data generation: $\beta_0 \sim DGP$ and deterministic β_1

Estimation: $\beta_0 \sim DGP$ and $\beta_1 \sim GP$

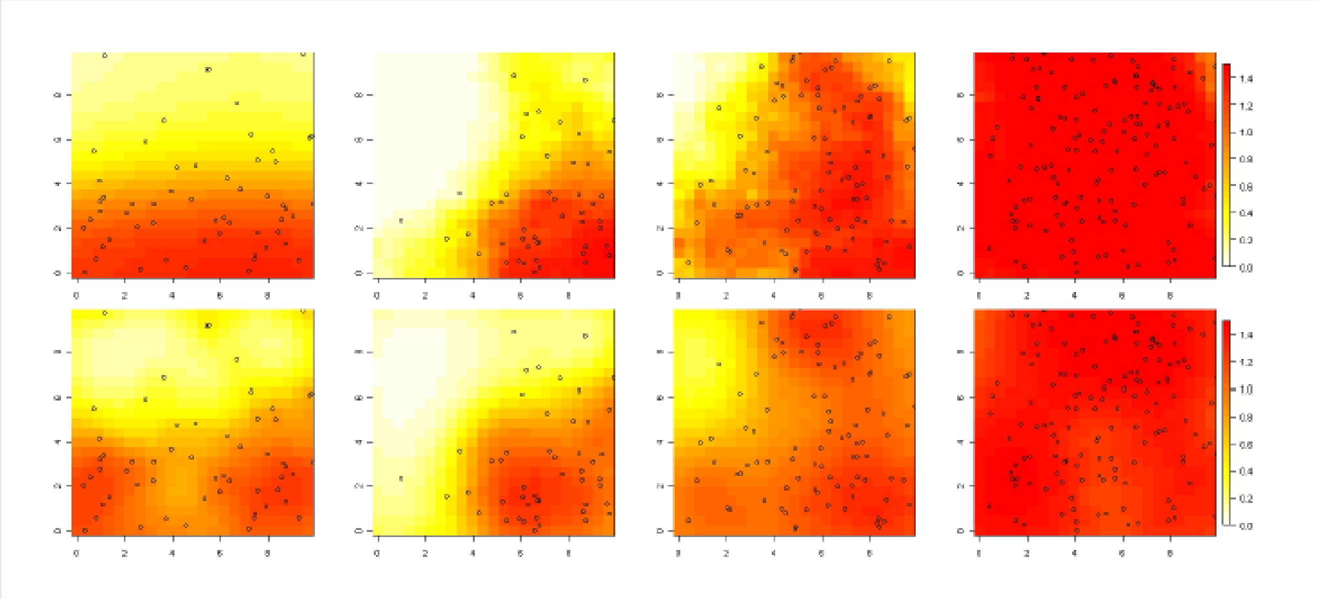
Estimation of β_1

deterministic **estimated**



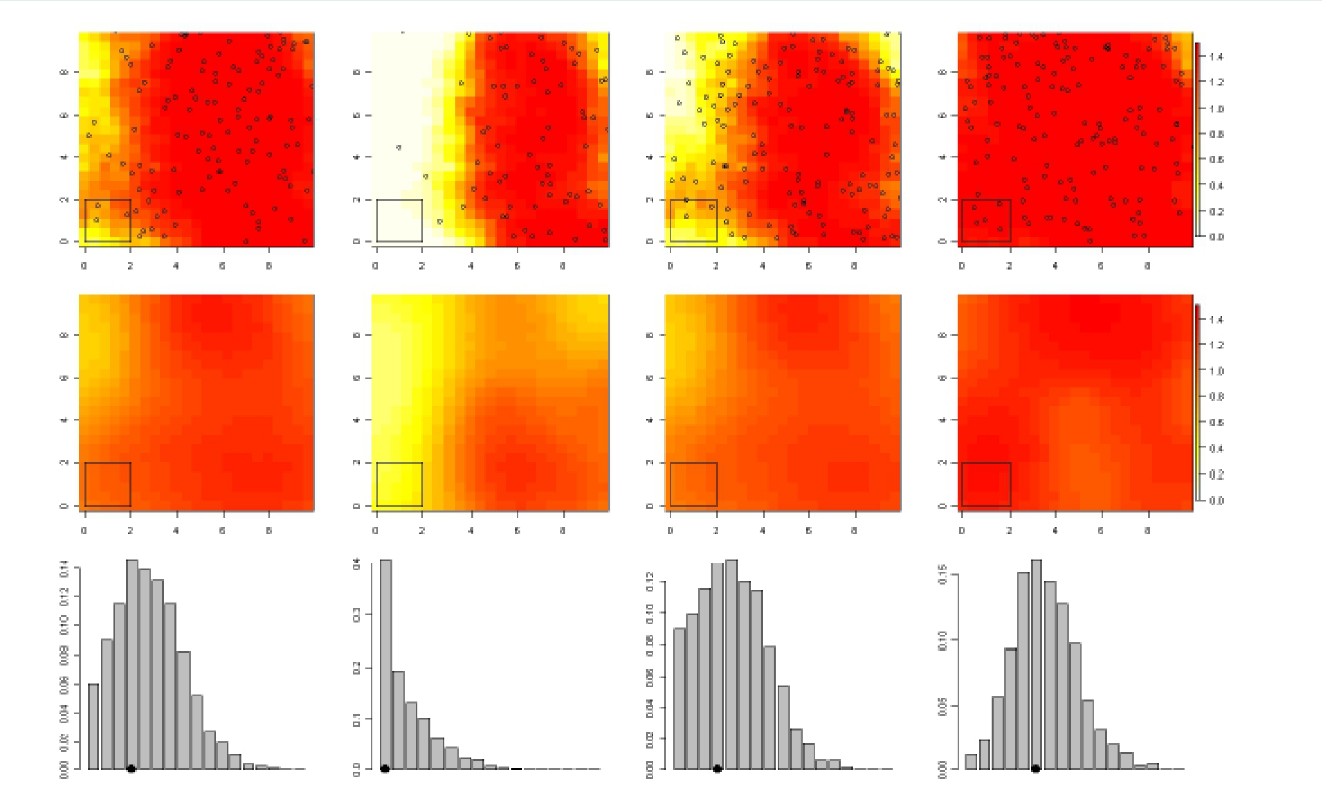
Estimated IF for $t = 1, 6, 11, 16$

true (top), estimated (bottom)



Predicted IF for $t = 17, 18, 19, 20$

true (top), predicted IF (middle) and $N([0,2]^2)$ (bottom)



5. Final comments

- discretization is useful when regional effects are more meaningful than

  point-wise

  computational cost depends on number of regions

- augmentation also *leads* to discretization

  computational cost depends on number of locations

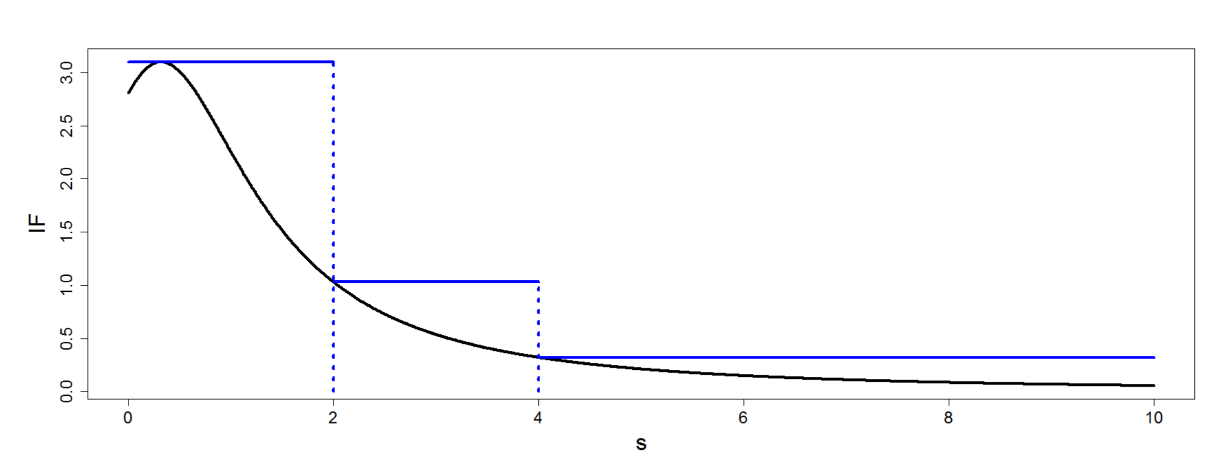
  (augmented + real)

- computation speeded by approximating GP's

  Shirota & Banerjee (2018)

Final comments (cont.)

- augmentation can be more efficient by varying λ^*



60% of the drawn points are "retained" [instead of 20%]

Final comments (cont.)

- GP is useful model for smooth IF

  not useful for jumps, discontinuities, non-stationarity, ...

- Alternatives available from Geostatistics literature

  partition into local GPs (Kim et al, 2005; Gramacy and Lee, 2007)

  particularly attractive from computational perspective

- our approaches allows for partition of β (and λ^*, with augmentation)

- individual configurations v: $\lambda^* \rightarrow \lambda^*(v)$. Example: $\lambda^*(v) = \lambda_0^* \, g(v)$.

- presence-only data: active area of application of point process

  a bit odd but common in Ecology; currently advising a Ph.D. project

Thank you!

dani@im.ufrj.br

www.statpop.com.br