Semi-parametric estimates of the long-term background trend, periodicity, and clustering effect in crime data

Jiancang Zhuang (ISM, Japan)

Jorge Mateau (University Jaume I of Castellon, Spain)

On stochastic reconstruction

- **Zhuang et al 2002, JASA --- ETAS model, application in earthquakes**
- **Zhuang et al 2004, JGR --- Stochastic reconstruction**
- **Zhuang, 2006, JRSSB --- Residual analysis**
- **Marsan et al 2006, Science**
- **Mohler et al, 2011, JASA, application to crime data**

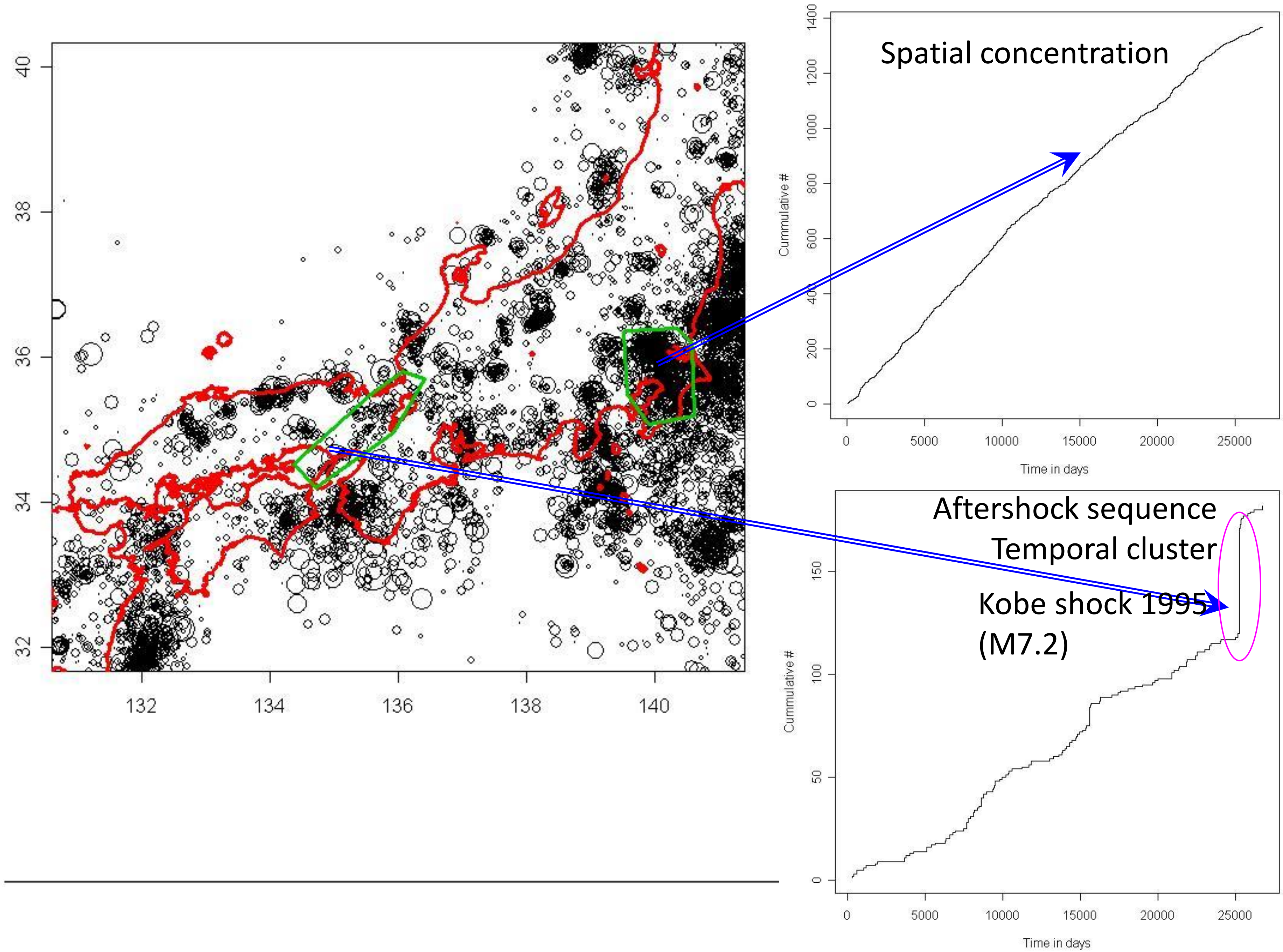
Point process & Conditional intensity

A **point process** N is a random measure that $\Pr\{N(B) < \infty\} = 1$ for any regualr bounded B.

Conditional intensity for a space-time point process N:

$$\lambda(t, x)dtdx = N(\, dt \times dx \mid H_t \,)$$

where $H_t = \sigma(N[(-\infty, t) \times B]), B$ is a Borel set.



Spatial concentration
Aftershock sequence
Temporal cluster
Kobe shock 1995
(M7.2)
Cummulative #
Time in days
0
5000
10000
15000
20000
25000
200
400
600
800
1000
1200
1400
50
100
150
132
134
136
138
140
32
34
36
38
40

Space-time ETAS model

- Time varying seismicity rate (conditional intensity or stochastic intensity)

$$\lambda(t,x,y) = \mu(x,y) + \sum_{i:t_i<t} \kappa(m_i) g(t-t_i) f(x-x_i, y-y_i, m_i)$$

Contribution from background seismicity

Contribution from the *i*-th event

Pr{event *j* is from background}

$$\phi_j = \frac{\mu(x_j, y_j)}{\lambda(t_j, x_j, y_j)}$$

Pr{event *j* is from *i*}

$$\rho_{ij} = \frac{g(t_j - t_i) f(x_j - x_i, y_j - y_i; m_j)}{\lambda(t_j, x_j, y_j)}$$

Thinning method

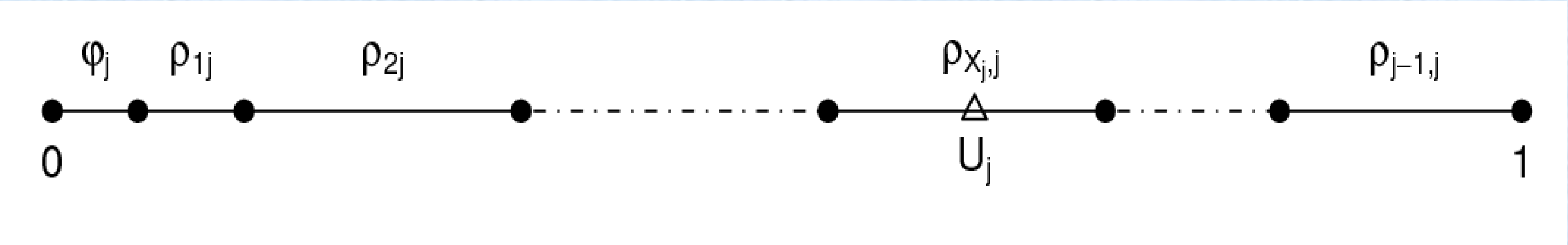
- For each event j

Pr{event j is from background} $\varphi_j = \dfrac{s(m_j)\mu(x_j, y_j)}{\lambda(t_j, x_j, y_j, m_j)}$

Pr{event j is from i} $\rho_{ij} = \dfrac{s(m_j)g(t_j - t_i)f(x_j - x_i, y_j - y_i)}{\lambda(t_j, x_j, y_j, m_j)}$

Stochastic declustering: Set event j to be a background event or a child of event 1, 2, …, according to probabilities φ_j or $\rho_{1j}, \rho_{2j}, \ldots, \rho_{j-1,j}$ respectively

Stochastic declustering method



Algorithm: Generate a uniform random number U_j on [0, 1], set K satisfy

$$\varphi_j + \sum_{i=1}^{K} \rho_{ij} \le U_j < \varphi_j + \sum_{i=1}^{K+1} \rho_{ij}$$

Estimation problems

- ETAS model – conditional intensity

(nonparametric part)

(parametric part)

$$\lambda(t,x,y) = \mu(x,y) + \sum_{i:t_i<t} \kappa(m_i)\, g(t-t_i)\, f(x-x_i, y-y_i, m_i)$$

How to estimate background seismicity?

How to estimate clustering parameters?

Estimation problems

- ETAS model – conditional intensity

$$\lambda(t,x,y) = \mu(x,y) + \sum_{i:t_i<t} \kappa(m_i) g(t-t_i) f(x-x_i, y-y_i, m_i)$$

How to estimate time-free total seismicity $\lambda(x,y)$?

↓

Kernel, spline, tessellation, histogram, …

How to estimate background seismicity?

↓

?

How to estimate clustering parameters?

↓

Maximum likelihood estimate if background seismicity **μ** is known

Estimation problems

- ETAS model – conditional intensity

$$\lambda(t,x,y) = \mu(x,y) + \sum_{i:t_i<t} \kappa(m_i)\, g(t-t_i)\, f(x-x_i, y-y_i, m_i)$$

How to estimate time-free total seismicity $\lambda(x,y)$?

How to estimate background seismicity?

How to estimate clustering parameters?

Kernel, spline, tessellation, histogram, …

Kernel, spline, tessellation, histogram, …, with each event weighted by φ_j

Estimation problems

- Time varying seismicity rate (conditional intensity or stochastic intensity)

$$\lambda(t,x,y)=\mu(x,y)+\sum_{i:t_i<t}\kappa(m_i)g(t-t_i)f(x-x_i,y-y_i,m_i)$$

Kernel

Kernel with each event weighted by φ_j

$$\hat{\mu}(x,y)=\frac{1}{T}\sum_j \varphi_j h(x-x_j,y-y_j;d)$$

$$\hat{\lambda}(x,y)=\frac{1}{T}\sum_i h(x-x_i,y-y_i;d)$$

Solution—estimating parameters and background rate simultaneously

Algorithm:

1. Assume an initial background rate.
2. Using MLE to estimate parameters in the clustering structures.
3. Using the assumed background and estimated clustering parameters to evaluate φ_j .
4. Using φ_j to get a better background rate.
5. Update the background rate by this better one.
6. Repeat Steps 2 to 5 until results converge.

φ_j: Estimate of probability that event j is of background

Uses of stochastic declustering

- To inverse clustering features (*Zhuang et al, 2004*)

- Empirical functions (histograms) of weighted samples
  - ρ_{ij} : *i* triggers ρ_{ij} children, not 1 child
  - φ_i: we get φ_i background events, not 1 background event

Stochastic reconstruction: inverting clustering features

- Empirical functions (histograms) of weighted samples
  - ρ_{ij} : i triggers ρ_{ij} children, not 1 child
  - φ_i: we get φ_i background events, not 1 background event

Example

Empirical p.d.f. for offspring locations

$$\hat{f}_R(r) = \frac{\sum_{i,j} \rho_{ij} I(|r_{ij} - r| < \Delta r / 2)}{\Delta r \sum_{i,j} \rho_{ij}}$$

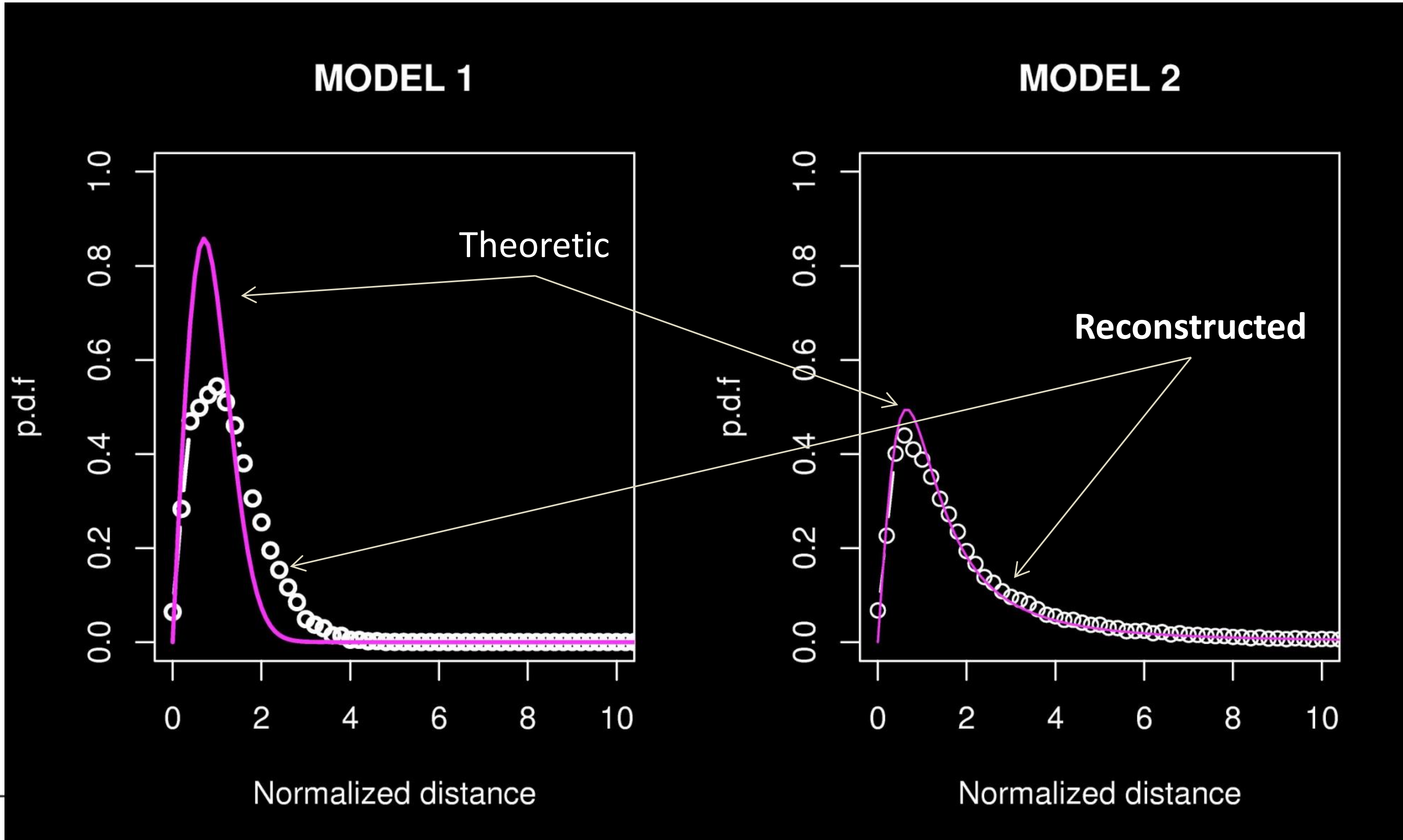
r, r_{ij} : Standardized distance between a parent and its direct offspring.

Sample weight

Results of location distributions for JMA earthquake catalog

Model 1: a short range decay (2-D gaussian => Rayleigh)

Model 2: a long range decay (inverse power => $Cr(1+r^2)^{-q}$)



Non-parametric estimation of both background rate and clustering structures

- Algorithm
- 1. Assuming some initial guess of model formation, obtain φ_j and ρ_{ij}
- 2. Estimate background rate and each component in the clustering part by using φ_j and ρ_{ij}.
- 3. Update φ_j and ρ_{ij}, and back to step 2 until convergence is reached.

(***Zhuang, 2006, JRSSB; Marsan and Lenglin, 2008***)

Burglary data in Los Angeles

(*MOHLER et al, 2008, JASA*)

- Conditional intensity or stochastic intensity

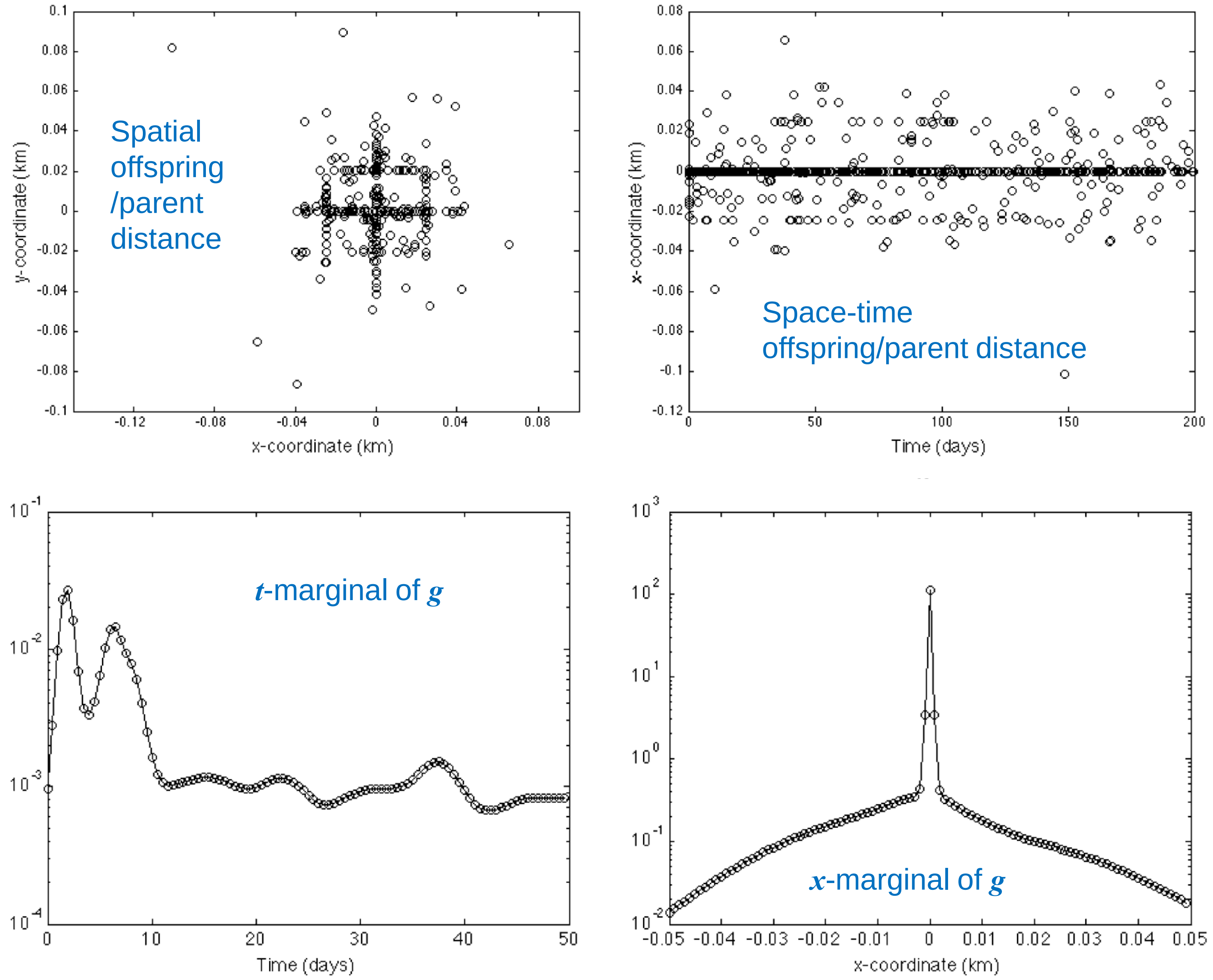
$$\lambda(t,x,y) = \nu(t)\mu(x,y) + \sum_{\{k:t_k<t\}} g(t-t_k, x-x_k, y-y_k).$$

- Data by the Los Angeles Police Department

5376 reported residential burglaries in an 18 km by 18 km region of the San Fernando Valley in Los Angeles occurring during the years 2004 and 2005.

Each burglary is associated with a reported time window over which it could have occurred, often a few hour span (for instance, the time span over which a victim was at work), and we define the time of burglary to be the midpoint of each burglary window.

Burglary data in Los Angeles (*MOHLER et al, 2008, JASA*)

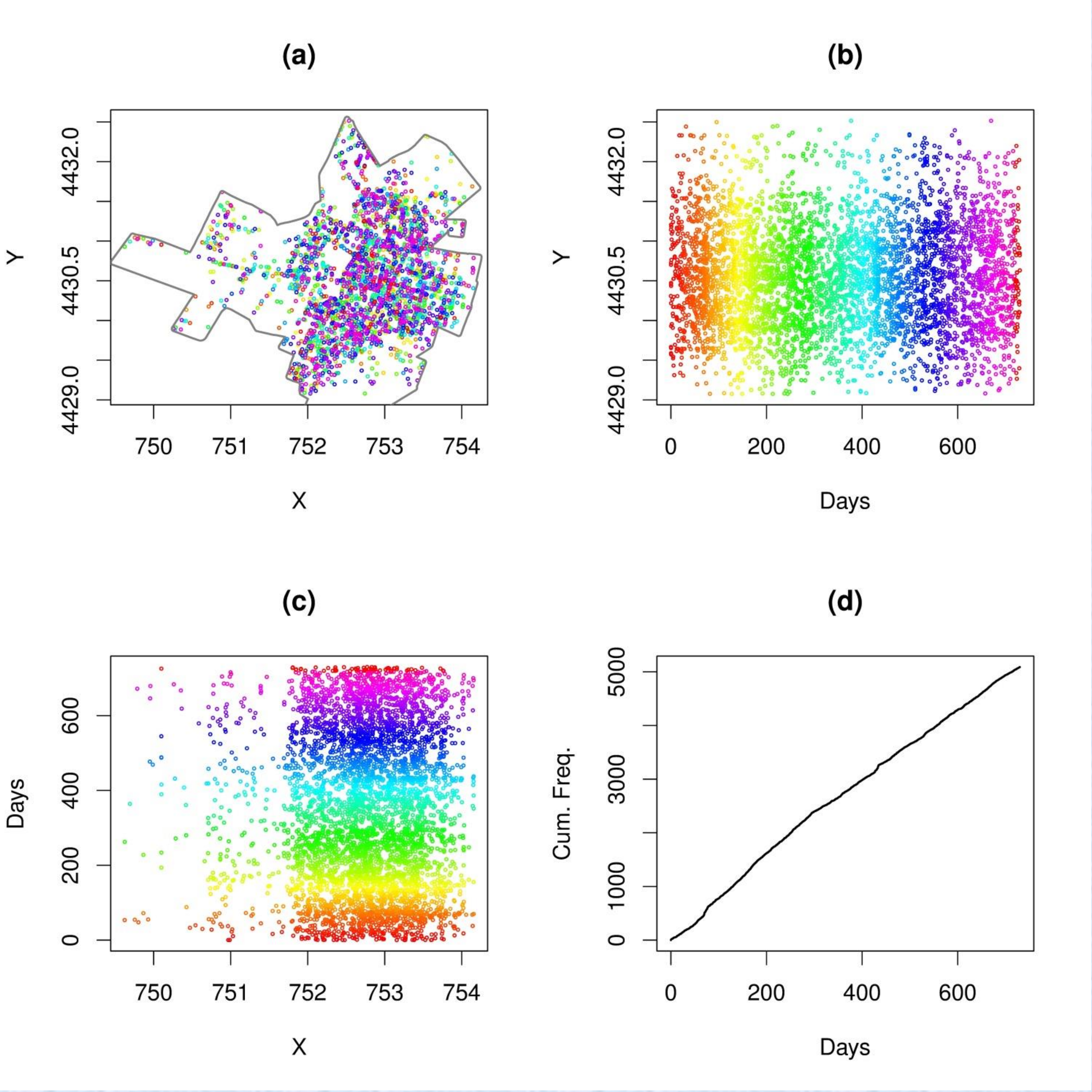


Estimating long-term background trend, periodicity, and clustering effect

Application to Robbery related violence in Castellon, Spain, 2013-2015

Dataset

Robbery related violence in Castellon, Spain, 2012-2013



Model

$$\boldsymbol{\lambda(t,x,y) = \mu_0\, \mu_t(t)\, \mu_d(t)\, \mu_w(t)\, \mu_b(x,y) \\ + A \sum_{i:t_i<t} g(t-t_i)\, f(x-x_i, y-y_i)}$$

t (day): time (x,y) (km): location

Background terms: all normalized to have average 1.

$\mu_t(t)$: trend $\mu_d(t)$: daily periodicity $\mu_w(t)$: weekly periodicity

$\mu_b(x,y)$: spatial inhomogeneity of background

Triggering terms: both normalized to be pdf

$g(t)$: temporal triggering response

$f(x,y)$: spatial triggering response.

μ_0 and A: constants, relaxing coeficients

Stochastic reconstruction

$$\hat{g}(t) \propto \sum_{ij} \rho_{ij}\, I\left(t_j - t_i \in [t - \Delta, t + \Delta]\right),$$

$$\hat{f}(x, y) \propto \sum_{ij} \rho_{ij}\, I\left(x_j - x_i \in [x - \Delta_x, x + \Delta_x]\right) I\left(y_j - y_i \in \left[y - \Delta_y, x + \Delta_y\right]\right),$$

$$\rho_{ij} = \frac{A g\left(t_j - t_i\right) h\left(x_j - x_i, y_j - y_i\right)}{\lambda\left(t_j, x_j, y_j\right)}, \qquad \text{for } j < i$$

We use kernel functions instead of simple histogram and correct the edge effects.

Stochastic reconstruction

$$\hat{\mu}_t(t) \propto \sum_i w_i^{(t)}\, I(t_i \in [t-\Delta, t+\Delta])\,, \qquad w_i^{(t)} = \frac{\mu_t(t_i)\,\mu_b(x_i, y_i)}{\lambda(t_i, x_i, y_i)}$$

$$\hat{\mu}_d(t) \propto \sum_i w_i^{(d)}\, I(t_i - \lfloor t_i \rfloor \in [t-\Delta, t+\Delta])\,, \qquad w_i^{(d)} = \frac{\mu_d(t_i)\,\mu_b(x_i, y_i)}{\lambda(t_i, x_i, y_i)}$$

$$\hat{\mu}_w(t) \propto \sum_i w_i^{(w)}\, I\left(t_i - 14 \times \left\lfloor \frac{t_i}{14} \right\rfloor \in [t-\Delta, t+\Delta]\right) \qquad w_i^{(w)} = \frac{\mu_w(t_i)\,\mu_b(x_i, y_i)}{\lambda(t_i, x_i, y_i)}$$

$$\hat{\mu}_b(x, y) \propto \sum_i \varphi_i I(x_i \in [x-\Delta_x, x+\Delta_x]) I\left(y_i \in \left[y-\Delta_y, x+\Delta_y\right]\right)$$

$$\varphi_i = \frac{\mu_0\,\mu_t(t_i)\mu_d(t_i)\mu_w(t_i)\,\mu_b(x_i, y_i)}{\lambda(t_i, x_i, y_i)}$$

Stochastic reconstruction

$$\hat{\boldsymbol{\mu}}_t(\boldsymbol{t}) \propto \sum_i \boldsymbol{w}_i^{(t)}\, \boldsymbol{I}(\boldsymbol{t}_i \in [\boldsymbol{t} - \boldsymbol{\Delta}, \boldsymbol{t} + \boldsymbol{\Delta}]),$$

$$\boldsymbol{w}_i^{(t)} = \frac{\boldsymbol{\mu}_t(\boldsymbol{t}_i)\, \boldsymbol{\mu}_b(\boldsymbol{x}_i, \boldsymbol{y}_i)}{\boldsymbol{\lambda}(\boldsymbol{t}_i, \boldsymbol{x}_i, \boldsymbol{y}_i)}$$

Given a spatiotemporal point process *N* equipped with a conditional intensity $\lambda(t,x)$, if $h(t,x)$ is a predictable marked process, then for any fixed interval *T* and region *S*,

$$E\left[\sum_{(t_i,x_i)\in N\cap T\times S} h(t_i,x_i)\right] = E\left[\iint_S \int_T h(t,x)\lambda(t,x)dtdx\right]$$

providing that h is nonnegetive or either sides of the above exists.

Stochastic reconstruction

$$\hat{\boldsymbol{\mu}}_{\boldsymbol{t}}(\boldsymbol{t}) \propto \sum_{\boldsymbol{i}} \boldsymbol{w}_{\boldsymbol{i}}^{(\boldsymbol{t})} \, \boldsymbol{I}(\boldsymbol{t}_{\boldsymbol{i}} \in [\boldsymbol{t} - \boldsymbol{\Delta}, \boldsymbol{t} + \boldsymbol{\Delta}]) \, , \qquad \boldsymbol{w}_{\boldsymbol{i}}^{(\boldsymbol{t})} = \frac{\boldsymbol{\mu}_{\boldsymbol{t}}(\boldsymbol{t}_{\boldsymbol{i}}) \, \boldsymbol{\mu}_{\boldsymbol{b}}(\boldsymbol{x}_{\boldsymbol{i}}, \boldsymbol{y}_{\boldsymbol{i}})}{\boldsymbol{\lambda}(\boldsymbol{t}_{\boldsymbol{i}}, \boldsymbol{x}_{\boldsymbol{i}}, \boldsymbol{y}_{\boldsymbol{i}})}$$

$$h(t,x,y) = \frac{\mu_t(t)\mu_b(x,y)}{\lambda(t,x,y)}$$

$T = [t - \Delta, t + \Delta], S = \text{whohle area}$

$$\begin{aligned} E\left[\sum_{(t_i,x_i,y_i)\in N\cap T\times S} h(t_i,x_i,y_i)\right] &= E\left[\int_S \int_T h(t,x)\lambda(t,x)dtdxdy\right] \\ &= E\left[\int_T \int_S \mu_t(t)\mu(x,y)dxdydt\right] \\ &= \int_T \mu_t(t)\,dt \, \iint_S \mu_b(x,y)dxdy \\ &\propto \mu_t(t)\,\Delta \end{aligned}$$

Estimation of relaxing coefficients

Update μ_0 and A through maximizing the likelihood function

$$\log L = \sum_{i=1}^{n} \log \lambda(t_i, x_i, y_i) - \int_0^T \iint_S \lambda(s, u, v)\, du\, dv\, ds$$

which reduces to

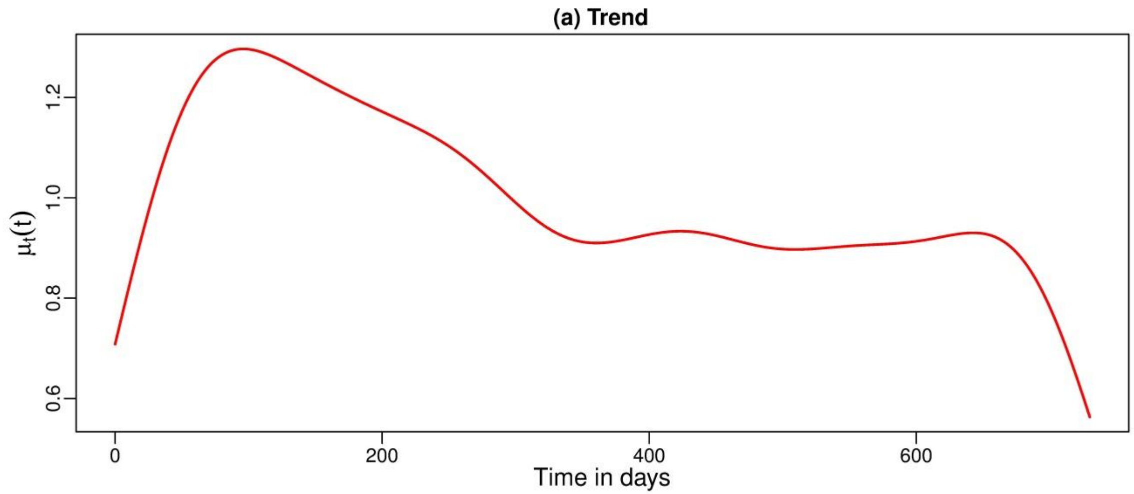
$$A^{(k+1)} = \frac{n - \sum_{i=1}^{n} \varphi_i^{(k)}}{G} \qquad U = \int_0^T \iint_S \mu_t(t)\mu_d(t)\mu_w(t)\mu_b(x,y) dxdydt$$

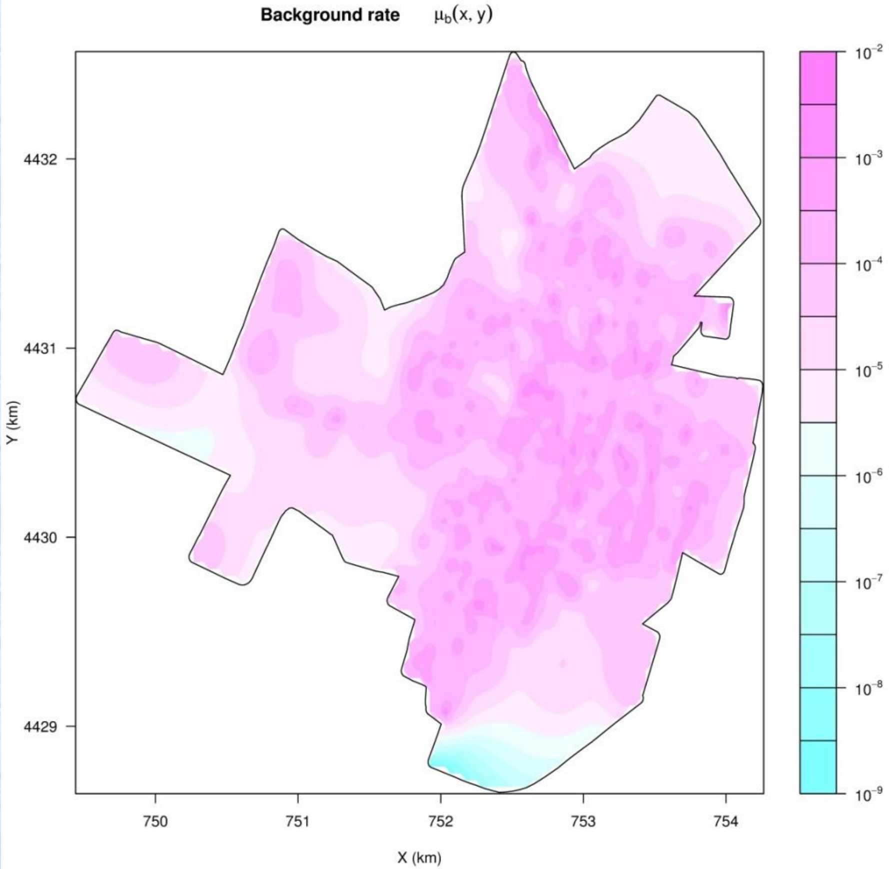
$$\mu_0^{(k+1)} = \frac{n - A^{(k+1)} G}{U} \qquad G = \sum_{i}^{n} \int_0^T \iint_S g(t - t_i) f(x - x_i, y - y_i) dxdydt$$

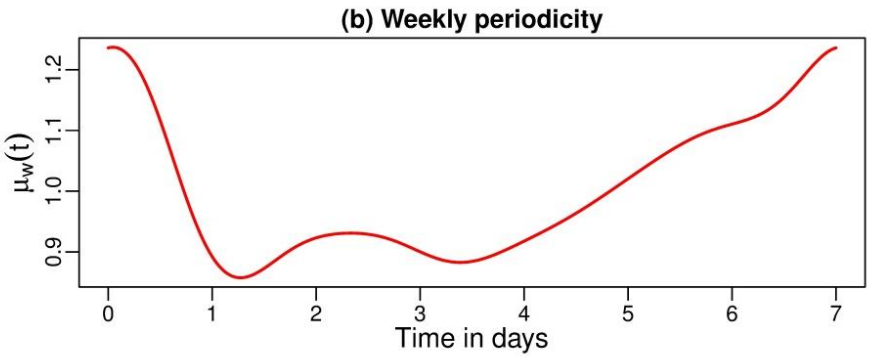
$$\varphi_i^{(k)} = \frac{\mu_0^{(k)}\, \mu_t(t_i)\, \mu_d(t_i)\mu_w(t_i)\mu_b(x_i, y_i)}{\mu_0^{(k)}\, \mu_t(t_i)\, \mu_d(t_i)\mu_w(t_i)\mu_b(x_i, y_i) + A^{\{(k)\}} \sum_{j:\, t_j < t_i} g(t_j - t_i)\, h(x_j - x_i, y_j - y_i)}$$

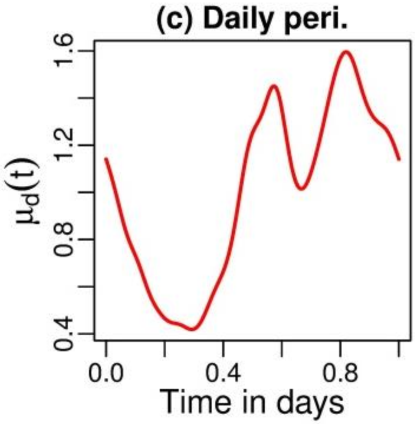
Results

$\mu_0 = 0.791$



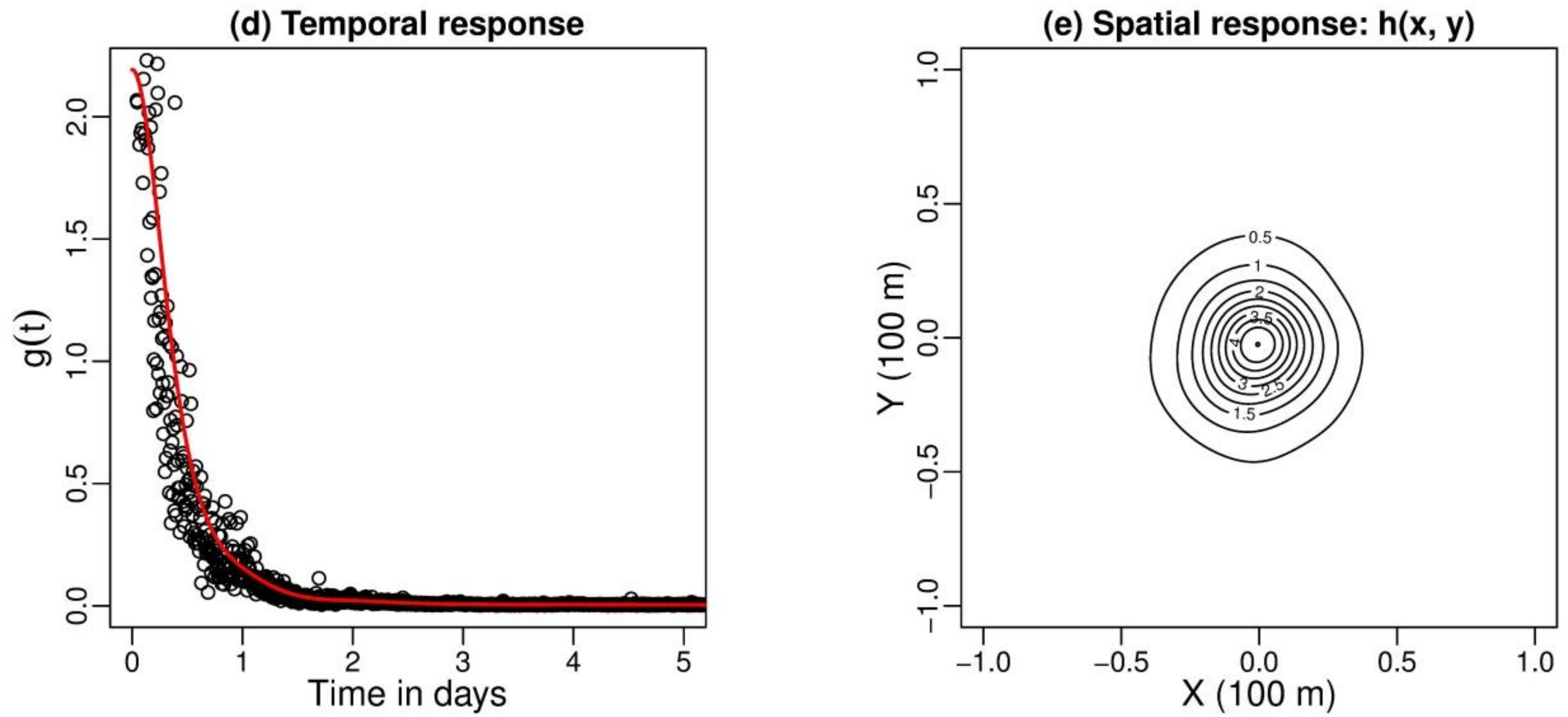






Results

$A = 0.029$



Conclusions

1. Stochastic reconstruction helps visualizing the structure of family trees in the observations of a branching process together with uncertainties.

2. Based on the theory of residual analysis, stochastic reconstruction provides us a non-parametric method for estimating each individual characteristic in a wide range of branching models.

3. New ingredients are added in the analysis: (a) periodicity in background and (b) relaxation coefficients.


Thank
You