

Development and Comparison of Residual Analysis Methods for Space-time Point Processes

Robert Alan Clements

October 5, 2012

Acknowledgements

Rick Paik Schoenberg
UCLA Department of Statistics

Danijel Schorlemmer
GFZ German Research Centre for Geosciences

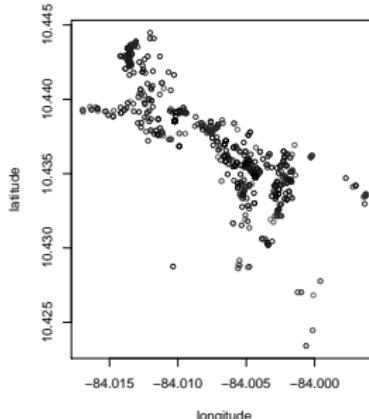
Alejandro Veen
IBM TJ Watson Research Center

Outline

Outline

Point processes

- N is:
 - a spatial-temporal point process
 - (locally) σ -finite defined on a compact set $\mathcal{S} \subset \mathbb{R}^d$
 - adapted to a filtration \mathcal{F} with probability measure \mathcal{P}
 - simple: points at distinct locations and times
- examples: earthquake occurrences, wildfires, crime locations



The conditional intensity

- N has \mathcal{F} -conditional intensity $\lambda(\mathbf{x}|\mathcal{H}_t) < \infty$
- $\lambda(\mathbf{x}|\mathcal{H}_t)$ is defined as the frequency with which events are expected to occur in \mathcal{S} around a specific point, conditional on the prior history \mathcal{H}_t (Daley and Vere-Jones, 2003)

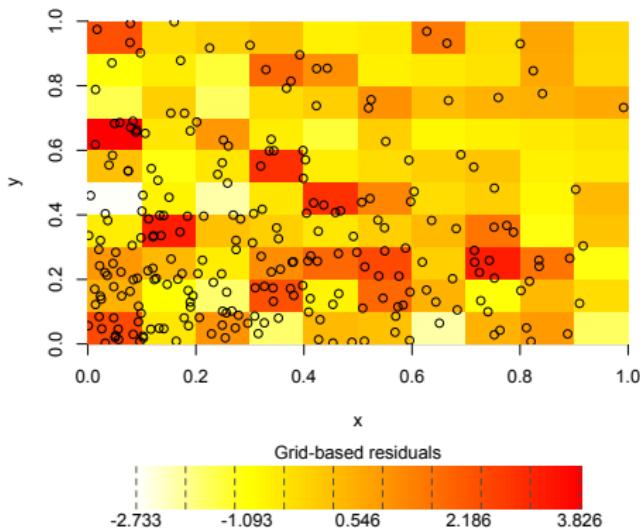
$$\lim_{\Delta x, \Delta y, \Delta t \downarrow 0} \frac{E[N\{(x, x+\Delta x) \times (y, y+\Delta y) \times (t, t+\Delta t)\} | \mathcal{H}_t]}{\Delta x \Delta y \Delta t}$$

- λ uniquely characterizes a simple point process

Outline

Pixel-based residuals

- \mathcal{S} is divided into bins, denoted B_i (Baddeley et al., 2005)



Pixel-based residuals

- Raw residuals:

$$R_R(B_i) = N(B_i) - \int_{B_i} \hat{\lambda}(\mathbf{x}|\mathcal{H}_t) d\mathbf{x}$$

- Pearson residuals:

$$R_P(B_i) = \sum_{(\mathbf{x}_i) \in B_i} \frac{1}{\sqrt{\hat{\lambda}(\mathbf{x}_i|\mathcal{H}_t)}} - \int_{B_i} \sqrt{\hat{\lambda}(\mathbf{x}|\mathcal{H}_t)} d\mathbf{x}$$

- Inverse residuals

$$R_I(B_i) = \sum_{(\mathbf{x}_i) \in B_i} \frac{1}{\hat{\lambda}(\mathbf{x}_i|\mathcal{H}_t)} - \int_{B_i} 1_{\{\hat{\lambda}(\mathbf{x}|\mathcal{H}_t) > 0\}} d\mathbf{x}$$

Pixel-based residuals

■ Deviance Residuals:

- model comparison tool
- compare log-likelihood within each bin
- deviance residual of model 1 vs. model 2:

$$R_D(B_i) = \sum_{(\mathbf{x}_i) \in B_i} \log \hat{\lambda}_1 - \int_{B_i} \hat{\lambda}_1 d\mathbf{x} - \left(\sum_{(\mathbf{x}_i) \in B_i} \log \hat{\lambda}_2 - \int_{B_i} \hat{\lambda}_2 d\mathbf{x} \right)$$

Pixel-based residuals

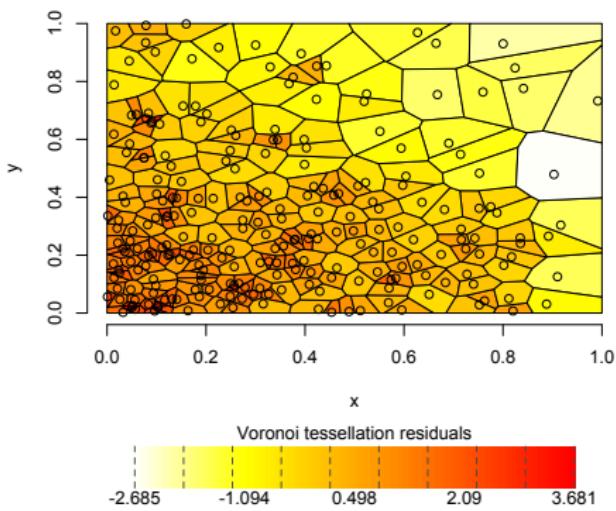
■ Problems:

- bin sizes are arbitrary
- residuals are highly skewed
- bins containing points dominate when depicted graphically
- large bins smooth out the residual

Tessellation residuals

- \mathcal{S} divided into cells using a Voronoi tessellation
- each cell, C_i contains 1 point

$$R_V(C_i) = \frac{1 - \int_{C_i} \hat{\lambda}(\mathbf{x}|\mathcal{H}_t) d\mathbf{x}}{\sqrt{\int_{C_i} \hat{\lambda}(\mathbf{x}|\mathcal{H}_t) d\mathbf{x}}}$$



Superposed residuals

- simulate points in \mathcal{S} with rate

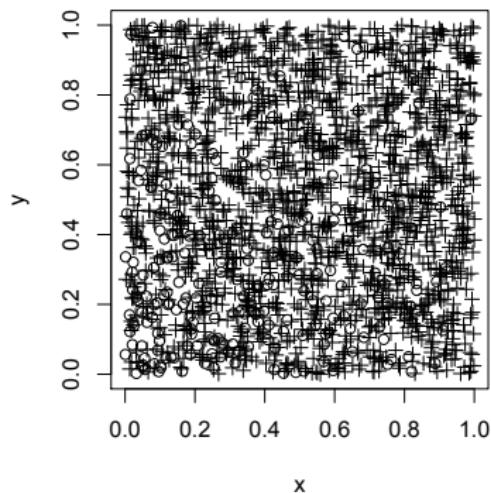
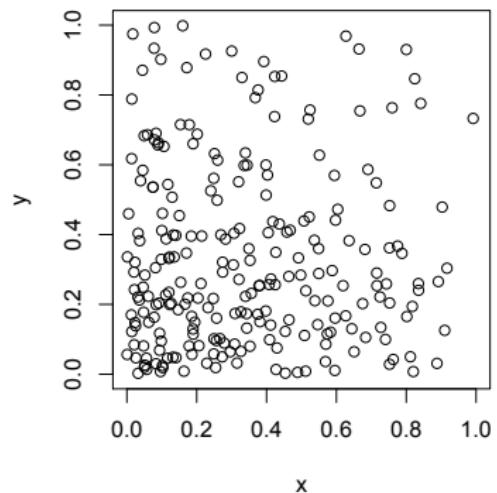
$$c - \hat{\lambda}(\mathbf{x}|\mathcal{H}_t) \text{ where } c = \sup\{\hat{\lambda}(\mathbf{x}|\mathcal{H}_t)\} \text{ over } \mathcal{S}$$

- essentially, superimpose a point process onto the existing point process, with conditional intensity

$$\hat{\lambda}_{\mathcal{S}}(\mathbf{x}|\mathcal{H}_t) = c - \hat{\lambda}(\mathbf{x}|\mathcal{H}_t)$$

- the result, called *superposed* residuals, is homogeneous Poisson, with rate c , if fitted model is correct (Brémaud, 1981)

Superposed residuals



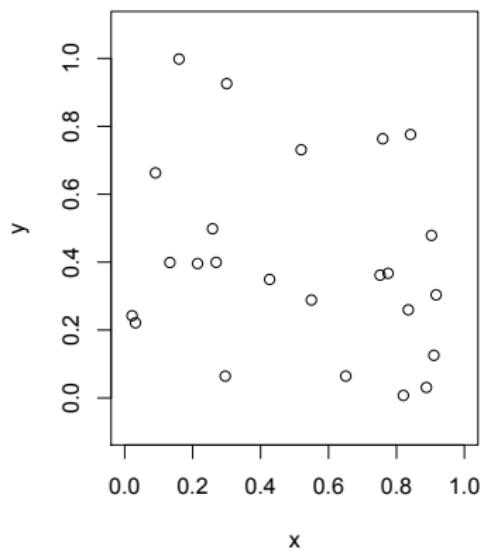
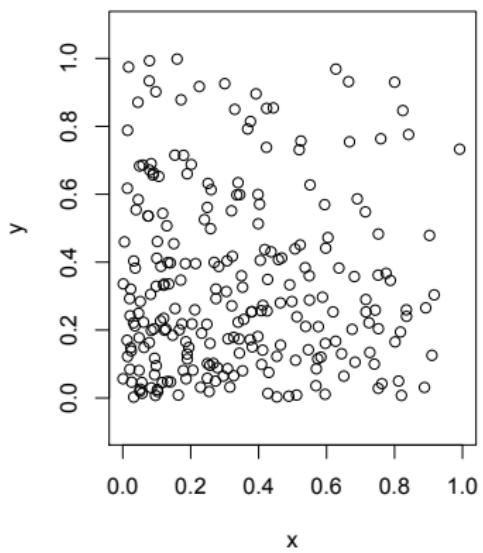
Thinned residuals

- keep each observation in \mathcal{S} with probability

$$\frac{b}{\hat{\lambda}(x_i, y_i, t_i | \mathcal{H}_t)} \text{ where } b = \inf\{\hat{\lambda}(\mathbf{x} | \mathcal{H}_t)\} \text{ over } \mathcal{S}$$

- remaining points, called *thinned* residuals, is homogeneous Poisson, with rate b , if fitted model is correct (Schoenberg, 2003)

Thinned residuals



Weaknesses of superposed and thinned residuals

■ Superposed:

- if c is large, too many residuals
- hard to interpret
- drowning out of data

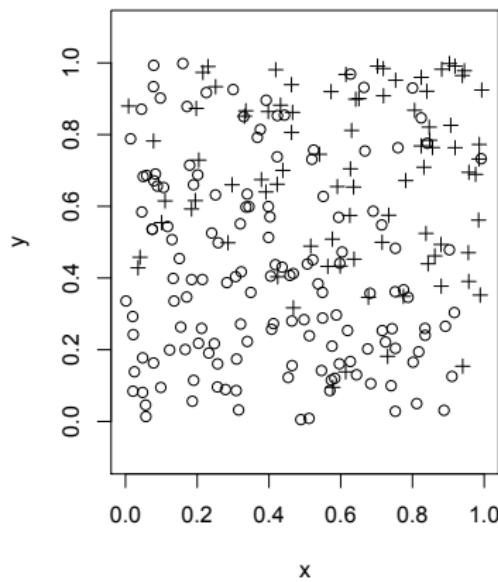
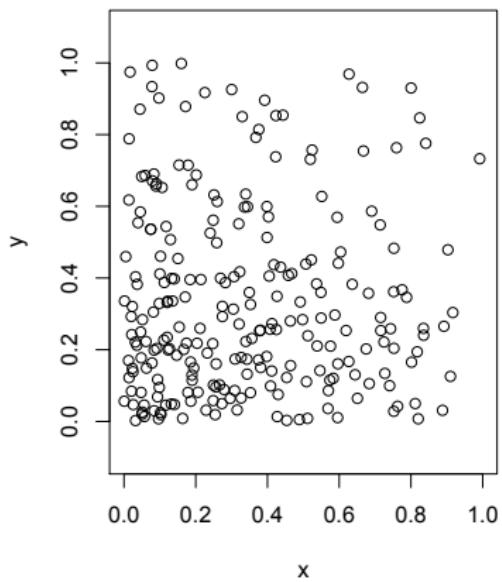
■ Thinned:

- if b is small, very few or no points
 - we can use *approximate thinned* residuals to fix this
 - let $j > b$, with $j \ll \frac{1}{|\mathcal{S}|} \int_{\mathcal{S}} \hat{\lambda} dS$
 - keep each point with probability $\frac{j}{\hat{\lambda}(x_i, y_i, t_i | \mathcal{H}_t)}$
 - *approximate thinned residuals* is approximately homogeneous
- we can do better!

Super-thinned residuals

- thin points if $\hat{\lambda}(x_i, y_i, t_i | \mathcal{H}_t) \geq k$, $\inf\{\hat{\lambda}\} \leq k \leq \sup\{\hat{\lambda}\}$
 - keep each point with probability $\frac{k}{\hat{\lambda}(x_i, y_i, t_i | \mathcal{H}_t)}$
- simulate points if $\hat{\lambda}(\mathbf{x} | \mathcal{H}_t) < k$
 - simulate points with rate $k - \hat{\lambda}(\mathbf{x} | \mathcal{H}_t)$
- the result, called *super-thinned* residuals, is homogeneous Poisson, with rate k , if fitted model is correct (Clements et al., 2011)

Super-thinned residuals



Super-thinned residuals

- some choices of k are more powerful than others
- how should we choose k ?
- common sense choices:
 - choose k such that the same number of points are thinned and superposed

$$k = \frac{1}{|\mathcal{S}|} \int_{\mathcal{S}} \hat{\lambda} d\mathcal{S}$$

- choose k such that the fewest number of points are thinned and superposed (the median of $\hat{\lambda}$)

Tests of homogeneity

- Ripley's K-function (Ripley, 1977):

$$\hat{K}(r) = \frac{|\mathcal{S}|}{N^2} \sum_i \sum_{j \neq i} \frac{1}{w(s_i, s_j)} \mathbf{1}_{\{|s_i - s_j| \leq r\}}$$

- N-test: compare number of residuals to a Poisson distribution
- inhomogeneous, or weighted, K-function (Baddeley et al., 2002; Veen and Schoenberg, 2005):

$$\hat{K}_w(r) = \sum_i \frac{1}{\hat{\lambda}(\mathbf{x}_i | \mathcal{H}_t)} \sum_{j \neq i} \frac{1}{\hat{\lambda}(\mathbf{x}_j | \mathcal{H}_t)} \mathbf{1}_{\{|s_i - s_j| \leq r\}}$$

- $K(r) > \pi r^2 \Rightarrow$ clustering, $K(r) < \pi r^2 \Rightarrow$ inhibition
- use $L(r) - r = \sqrt{\frac{K(r)}{\pi}} - r$ instead

Outline

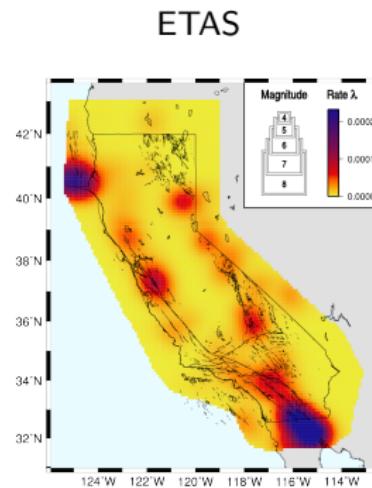
CSEP

Collaboratory for the Study of Earthquake Predictability

- international collaboration, with testing centers in:
 - Japan, Switzerland, New Zealand, and the US
- earthquake forecast models submitted by researchers in the field
- short-term (daily) and long-term (3-month, 5-year) forecasts
- testing regions
 - California
 - Japan
 - Northwest/Southwest Pacific
 - New Zealand
 - Global

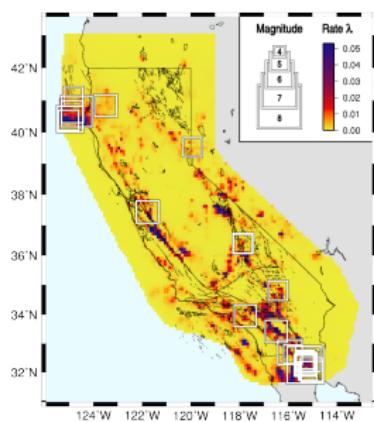
CSEP earthquake forecast models

- A (Helmstetter et al. (2007))
- B (Kagan et al. (2007))
- C (Shen et al. (2007))
- ETAS (Ogata and Zhuang, (2006))
- STEP (Gerstenberger et al. (2005))

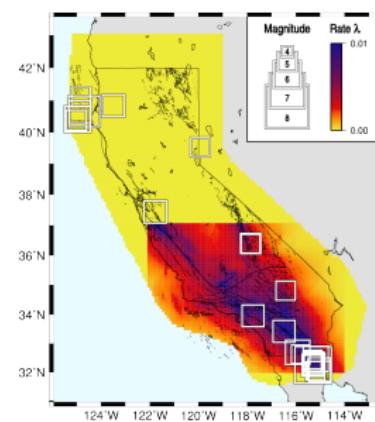


CSEP earthquake forecast models

A

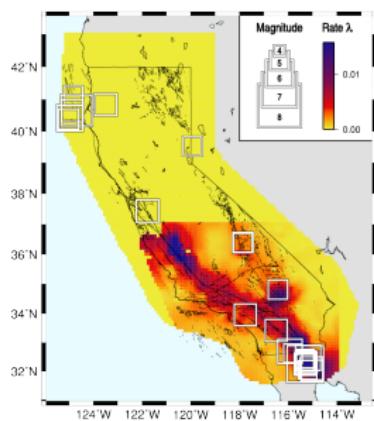


B

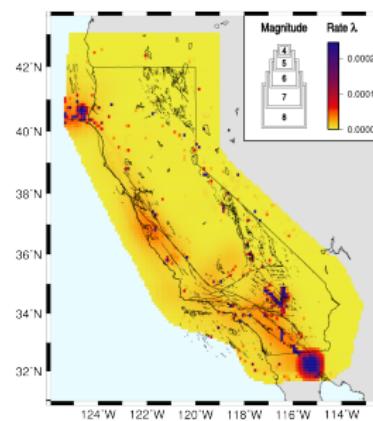


CSEP earthquake forecast models

C



STEP



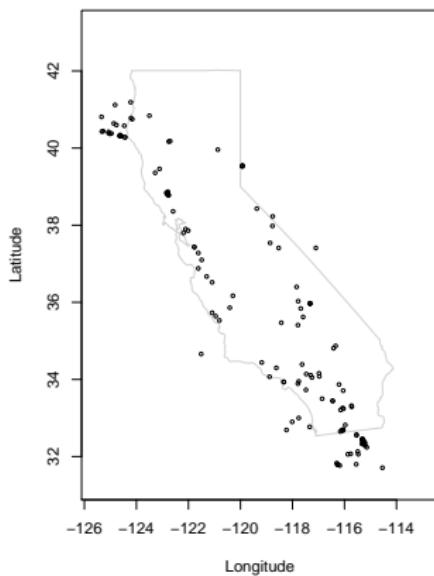
CSEP earthquake forecast models

- 5-year models only forecast earthquakes of magnitude ≥ 4.95
- problem: very few earthquake occurrences this large (only 13 in the testing region)
- use the tapered Gutenberg-Richter law to extrapolate forecasts to earthquakes of magnitude ≥ 3.95

$$10^{-b(m-m_{min})} \exp \left\{ 10^{1.5(m_{min}-m_c)} - 10^{1.5(m-m_c)} \right\}$$

Earthquake catalog

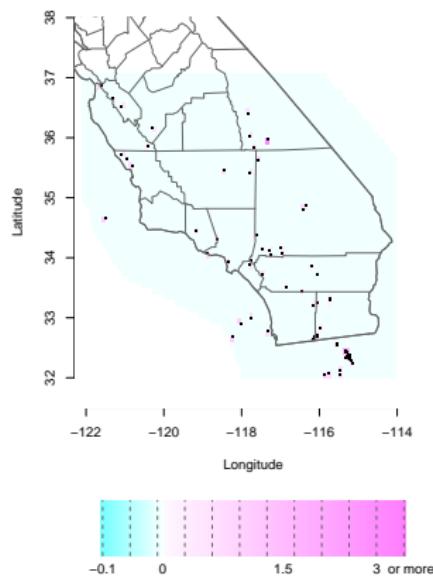
locations of 142 earthquakes of magnitude ≥ 3.95 , 1/1/2006 - 9/1/2009



CSEP tests

- currently numerical summaries and error diagrams are used to test goodness-of-fit
 - the N-test (number of events)
 - the L-test (likelihood)
 - the R-test (likelihood-ratio for comparing models)
 - the Molchan-test (Molchan 1990; Molchan 1997; Zaliapin & Molchan 2004; Kagan 2009)
 - the ROC-test (Swets 1973)
- Problems: no information about where or when the model is fitting poorly

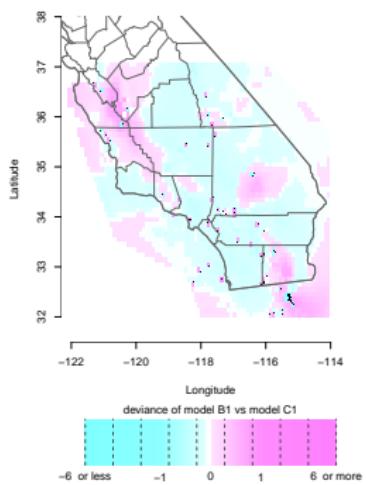
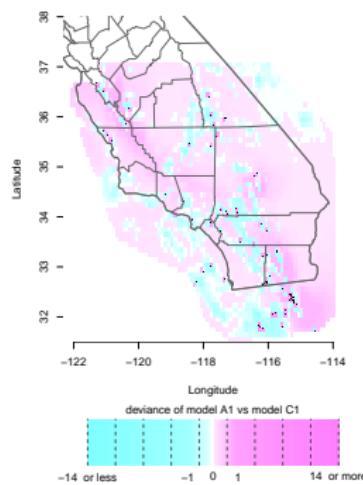
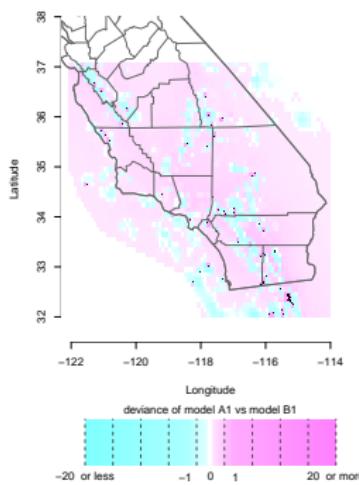
Pearson residuals results



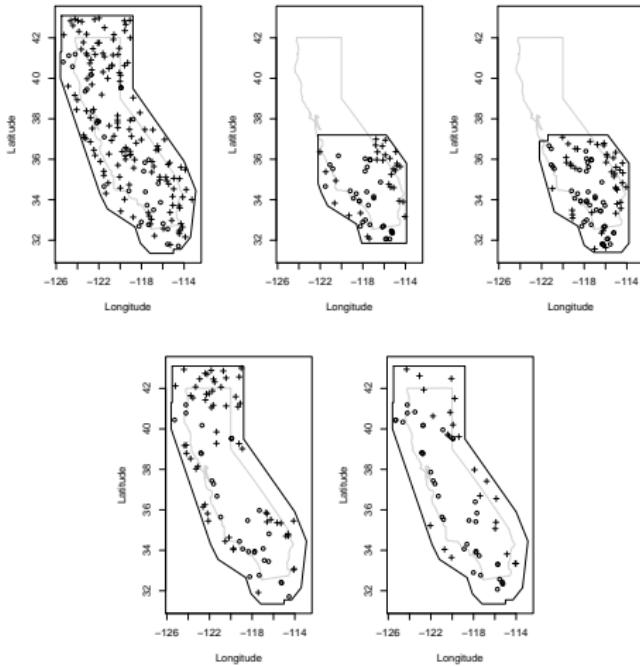
Large Pearson residuals for model B

- Imperial fault zone
 $R_P = 2.817$
- Panamint Valley fault zone
(similar for ETAS)

Deviance residuals results



Super-thinned residuals results



Results

- CSEP will be implementing these methods beginning January, 2011
- extrapolate forecasts for more meaningful results
- Pearson residuals ⇒ model improvements only in pixels containing earthquakes
- deviance residuals ⇒ preferred model pixel-by-pixel and overall
 - model A over models B and C
 - model C over model B
 - ETAS over STEP
- super-thinning ⇒ model improvements in all areas, but requires careful examination

Outline

The simulated models

Hawkes process

- useful for modeling earthquakes (Ogata 1988, 1998)
- self-exciting cluster process
 - **background points** (immigrants) trigger offspring points (1st generation descendants)
 - offspring points, in turn, trigger their own offspring points (2nd, 3rd, ... generation descendants)
- modeled by

$$\lambda(\mathbf{x}|\mathcal{H}_t) = \mu(\mathbf{x}, \mathbf{y}) + \sum_{\{i: t_i < t\}} g(x - x_i, y - y_i, t - t_i)$$

Cluster models

$$\lambda_{cl}(\mathbf{x}|\mathcal{H}_t) = \min \left\{ \gamma, \left[\mu(\mathbf{x}) + K_0 \sum_{\{i: t_i < t\}} g(t - t_i, x - x_i, y - y_i) \right] \right\}$$

($\mu = 0.02, \gamma = 3, K_0 = 8/9$)

- four g functions with different spatial clustering
 - exponential (squared distance) ($\beta = 7$)
 - halfnormal ($\sigma = 1$)
 - uniform ($\phi = 0, \omega = 2$)
 - pareto ($\eta = 5, \theta = 0.15$)
- temporal clustering component - exponential distribution ($\alpha = 1.5$)

Cluster models

$$\lambda_{cl}(\mathbf{x}|\mathcal{H}_t) = \min \left\{ \gamma, \left[\mu(\mathbf{x}) + K_0 \sum_{\{i: t_i < t\}} g(t - t_i, x - x_i, y - y_i) \right] \right\}$$

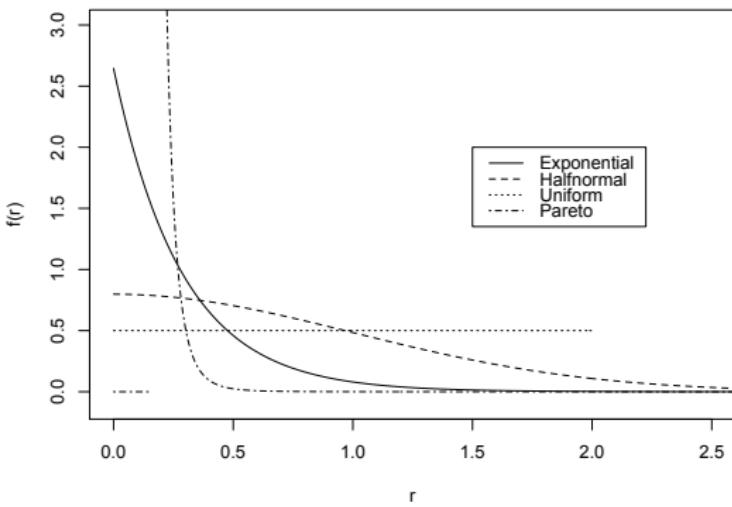
$$g_1(., ., .) = \frac{\alpha\beta}{\pi} \exp \left\{ -\alpha(t - t_i) - \beta((x - x_i)^2 + (y - y_i)^2) \right\}$$

$$g_2(., ., .) = \frac{\alpha}{2\sigma^2\pi} \exp \left\{ -\alpha(t - t_i) - \frac{1}{2} \left(\frac{\sqrt{(x - x_i)^2 + (y - y_i)^2}}{\sigma} \right)^2 \right\}$$

$$g_3(., ., .) = \begin{cases} \frac{\alpha}{\pi(\omega-\phi)(\omega+\phi)} e^{-\alpha(t-t_i)} & \text{if } \phi \leq \sqrt{(x - x_i)^2 + (y - y_i)^2} \leq \omega; \\ 0 & \text{if } \sqrt{(x - x_i)^2 + (y - y_i)^2} < \phi \text{ or } \sqrt{(x - x_i)^2 + (y - y_i)^2} > \omega. \end{cases}$$

$$g_4(., ., .) = \frac{\alpha(\eta-1)\theta^{\eta-1}}{2\pi((x - x_i)^2 + (y - y_i)^2)^{\frac{\eta+1}{2}}} e^{-\alpha(t-t_i)}$$

Spatial clustering



Inhibition models

$$\lambda_{inh}(\mathbf{x}|\mathcal{H}_t) = \delta + \left[\mu(\mathbf{x}) - K_0 \sum_{\{i: t_i < t\}} g(t - t_i, x - x_i, y - y_i) \right]^+$$

($\delta = 0.01, \mu = 0.25$)

- four g functions with different spatial inhibition
 - exponential (squared distance)
 - halfnormal
 - uniform
 - pareto
- temporal component - exponential distribution

Inhibition models

$$\lambda_{inh}(\mathbf{x}|\mathcal{H}_t) = \delta + \left[\mu(\mathbf{x}) - K_0 \sum_{\{i:t_i < t\}} g(t - t_i, x - x_i, y - y_i) \right]^+$$

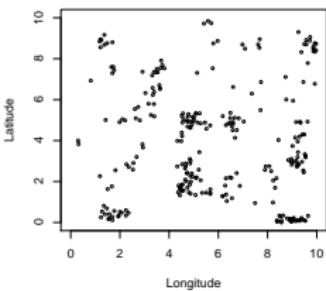
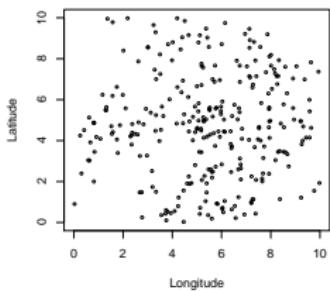
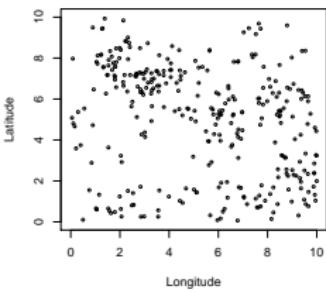
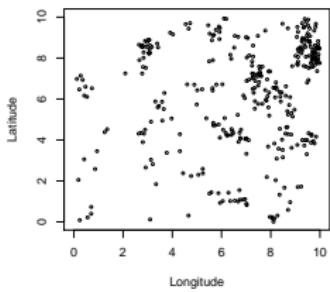
$$g_1(.,.,.) =$$

$$g_2(.,.,.) = \text{ same as above.}$$

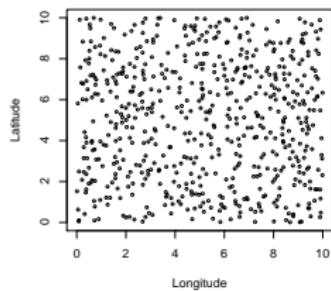
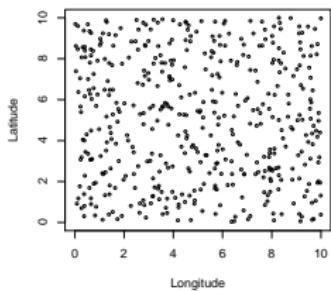
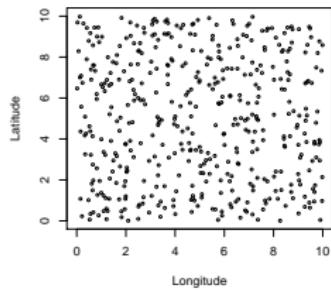
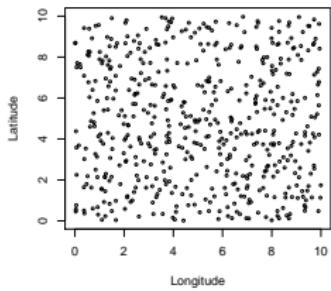
$$g_3(.,.,.) =$$

$$g_4(.,.,.) =$$

Cluster models examples



Inhibition models examples



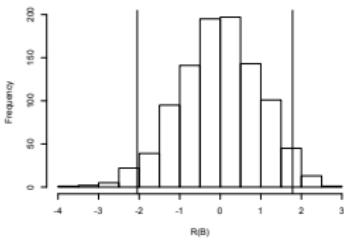
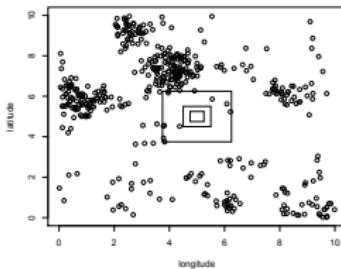
Methodology

- Simulations:
 - simulate 1000 realizations from each model
 - Exponential, Halfnormal, Uniform, Pareto cluster
 - Exponential, Halfnormal, Uniform, Pareto inhibition
 - “fit” each model to each simulation
 - cluster → cluster
 - inhibition → inhibition
 - use same parameter values
 - 8000 simulations
 - estimate distribution of residuals using true model
 - compare residuals from 3 alternative models to distribution

Methodology - pixel-based

- constant pixel, B_c
 - three pixel sizes
 - estimate a distribution of residuals
 - $q_{.025} = 2.5\text{th percentile}$, $q_{.975} = 97.5\text{th percentile}$

$$\rho_p = 1 - \frac{\sum_{i=1}^{1000} 1_{\{q_{.025} < R^i(B_c) < q_{.975}\}}}{1000}$$



Preliminary results

- raw, Pearson, and inverse residuals - nothing stands out
- $\rho_p \nearrow$ as pixel size \nearrow
 - need to test more sizes
- performance is better with inhibition models
- tessellation residuals are disappointing
- deviance residuals perform best overall

Preliminary results

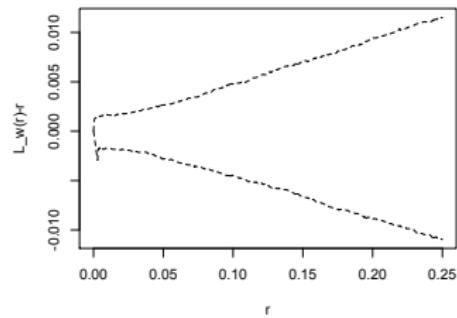
True model	Null model	Deviance residuals		
		bin	side	length
Cluster		0.5	1.0	1.5
Exponential	<i>Halfnormal</i>	0.07	0.07	0.14
	<i>Uniform</i>	0.34	0.39	0.50
	<i>Pareto</i>	0.02	0.01	0.01
Halfnormal	<i>Exponential</i>	0.10	0.18	0.22
	<i>Uniform</i>	0.09	0.13	0.19
	<i>Pareto</i>	0.06	0.02	0.03
Uniform	<i>Exponential</i>	0.02	0.16	0.29
	<i>Halfnormal</i>	0.04	0.05	0.11
	<i>Pareto</i>	0.22	0.12	0.23
Pareto	<i>Exponential</i>	0.08	0.13	0.14
	<i>Halfnormal</i>	0.05	0.08	0.11
	<i>Uniform</i>	0.01	0.10	0.23

True model	Null model	Deviance residuals		
		bin	side	length
Inhibition		0.5	1.0	1.5
Exponential	<i>Halfnormal</i>	0.13	0.39	0.79
	<i>Uniform</i>	0.38	0.65	0.90
	<i>Pareto</i>	0.43	0.96	1.00
Halfnormal	<i>Exponential</i>	0.33	0.65	0.84
	<i>Uniform</i>	0.12	0.14	0.14
	<i>Pareto</i>	0.06	0.14	0.47
Uniform	<i>Exponential</i>	0.05	0.23	0.83
	<i>Halfnormal</i>	0.03	0.03	0.04
	<i>Pareto</i>	0.09	0.29	0.69
Pareto	<i>Exponential</i>	0.77	0.94	1.00
	<i>Halfnormal</i>	0.74	0.78	0.91
	<i>Uniform</i>	0.07	0.35	0.72

Methodology - transformation

- use k values ranging from b (min) to c (max)
- evaluate homogeneity of residuals with $L_w(r)$
- estimate a distribution for $L_w(r)$ based on 1000 homogeneous Poisson processes of rate k
-

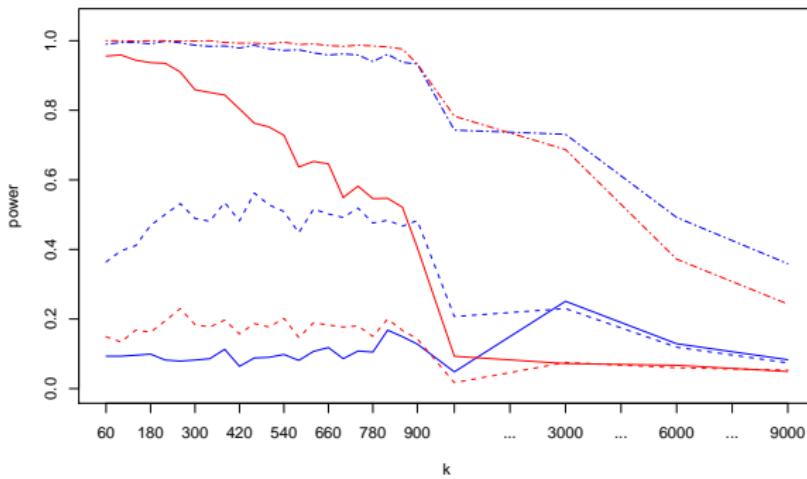
$$\rho_t = 1 - \frac{\sum_{i=1}^{1000} \mathbf{1}_{\{q_{.025} < L_w^i(r) < q_{.975}\}}}{1000}$$



Preliminary results

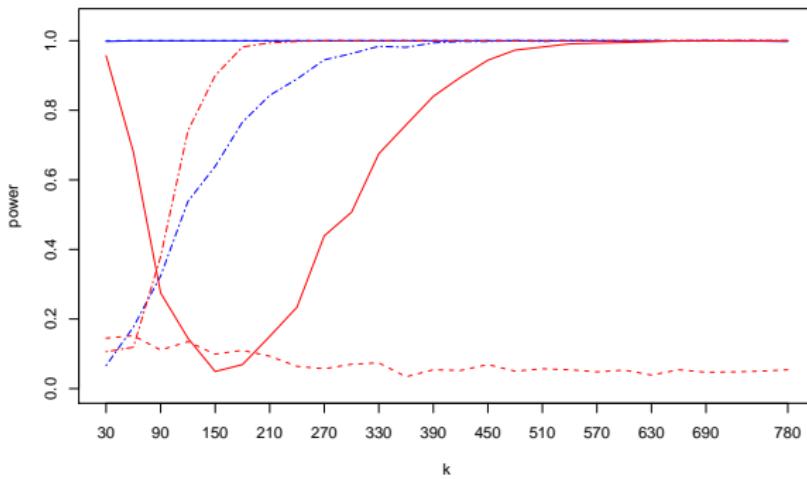
- cluster models:
 - $\rho_t \searrow$ as $k \nearrow$
- inhibition models:
 - performance is better with inhibition models
 - $\rho_t \nearrow$ as $k \nearrow$

Super-thinning power - cluster



Blue lines are power curves for the Exponential simulations. Solid line: Halfnormal. Dashed line: Uniform. Dashed-dotted line: Pareto. Red lines are power curves for the Halfnormal simulations. Solid line: Exponential. Dashed line: Uniform. Dashed-dotted line: Pareto.

Super-thinning power - inhibition



Blue lines are power curves for the Exponential simulations. Solid line: Halfnormal. Dashed line: Uniform. Dashed-dotted line: Pareto. Red lines are power curves for the Halfnormal simulations. Solid line: Exponential. Dashed line: Uniform. Dashed-dotted line: Pareto.

Outline

stppResid

- an R package for residual analysis of space-time point processes
- **stpp** - a space-time point process object
- contains functions for
 - **thinresid**: thinned residuals
 - **supresid**: superposed residuals
 - **superthin**: super-thinned residuals
 - **gridresid**: raw, Pearson, inverse residuals
 - **devresid**: deviance residuals
 - **tessresid**: tessellation residuals
 - summaries
 - plots

stppResid: stpp()

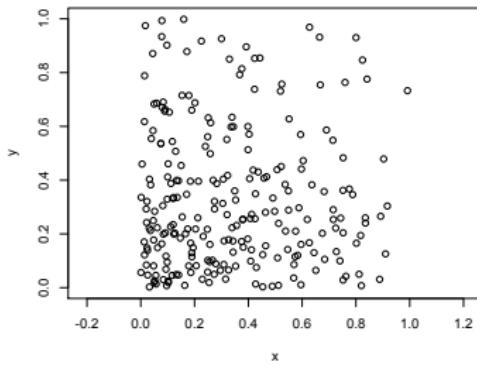
stpp object

```
stpp(x, y, t, stw = stwin(xcoord, ycoord, tcoord))
```

- **x:** x coordinates of observed points
- **y:** y coordinates of observed points
- **t:** t coordinates of observed points
- **stw** defines the space-time window
 - **xcoord:** x limits
 - **ycoord:** y limits
 - **tcoord:** t limits

stppResid: stpp()

```
> load(sim)
> X <- stpp(sim$x, sim$y, sim$t) #use default space-time window
> plot(X)
```



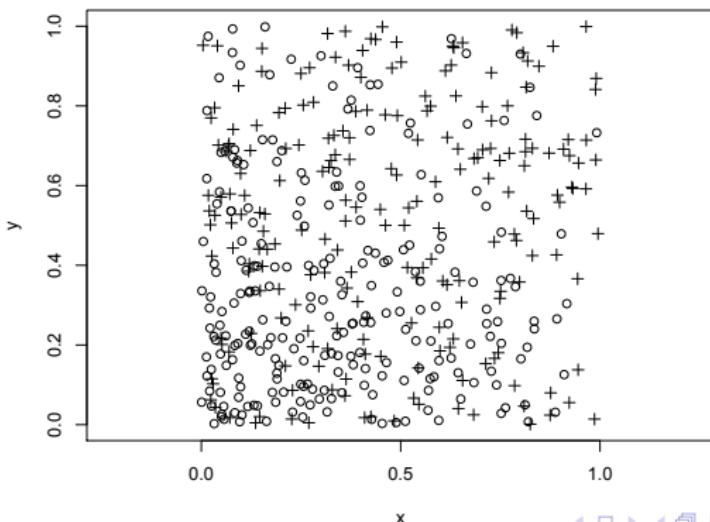
stppResid: superthin()

Super-thinned residuals for model ci1

```
> ci1 <- function(X, theta) {theta[1]*exp(-theta[2]*X$x -  
theta[3]*X$y - theta[4]*X$t)}  
> stresiduals <- superthin(X, cifunction = ci1, theta = c(2500,  
5, 5, 10), k = 250)  
> summary(stresiduals)  
Superthin rate: 250  
Number of residuals: 456  
Expected number of residuals: 250  
One-tailed p-value: 6.140031e-32
```

stppResid: superthin()

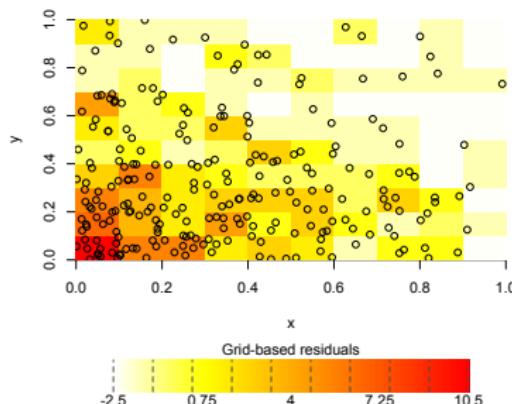
```
> plot(stresiduals)
```



stppResid: gridresid()

Raw residuals for model ci2

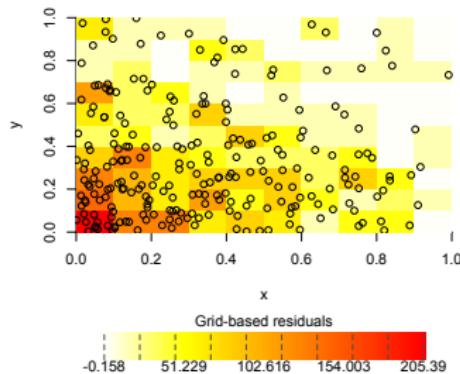
```
> ci2 <- function(X, theta = NULL) {rep(250, length(X$x))}  
#homogeneous Poisson model  
> gresiduals <- gridresid(X, ci2)  
> plot(gresiduals)
```



stppResid: gridresid()

Pearson residuals for model ci2

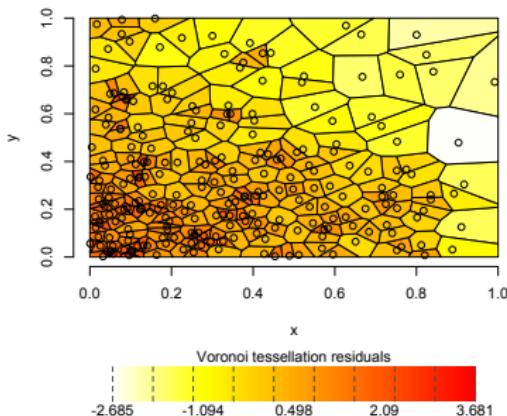
```
> gresiduals <- gridresid(X, ci2, resid = "pearson")
> plot(gresiduals)
```



stppResid: tessresid()

Tessellation residuals for model ci2

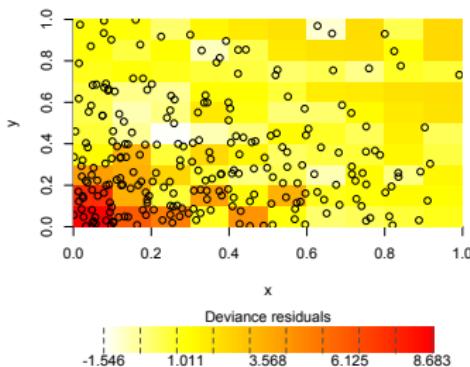
```
> tsresiduals <- tessresid(X, ci2)
> plot(tsresiduals)
```



stppResid: devresid()

Deviance residuals for model ci1 vs. ci2 on a 10x10 grid

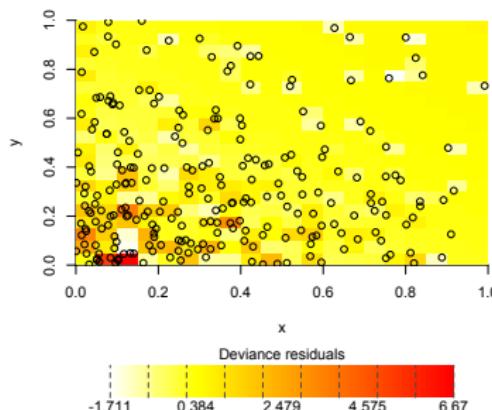
```
> deviance <- devresid(X, ci1, ci2, theta1 = c(3000, 2, 2, 2),  
theta2 = NULL)  
> plot(deviance)
```



stppResid: devresid()

Deviance residuals for model ci1 vs. ci2 on a 20x20 grid

```
> deviance <- devresid(X, ci1, ci2, theta1 = c(3000, 2, 2, 2),  
theta2 = NULL, grid = c(20, 20))  
> plot(deviance)
```



Future updates

- implement support for marked space-time point processes -
 $\lambda(x, y, t, m)$
 - support for non-rectangular observation regions
-
- improve Monte Carlo integration algorithms

```
> system.time(devresid(sim, ci1, ci2, theta1 = 3000, theta2 = NULL, grid = c(20, 20)))  
user      system elapsed  
2042.176  22.622 2162.165
```

Contact

Thank you for attending!

Contact info:
Robert Clements
Department of Statistics, UCLA
clements@stat.ucla.edu