Point-pattern analysis

D G Rossiter

Cornell University

April 18, 2020

Outline

Outline

What is a point pattern?

- A set of (georeferenced) point **locations** within a defined **region**
- Presumed to result from a **data-generating process** (DGP) operating over the region
- A DGP that results in the location of points is called a **point process**

  > *"a stochastic mechanism which generates a* **countable** *set of events" [7]*

- Details
  - Lines and polygons may be reduced to "points", e.g., by centroid, and treated as points
  - The **region must be carefully defined**, otherwise most statistics are distorted

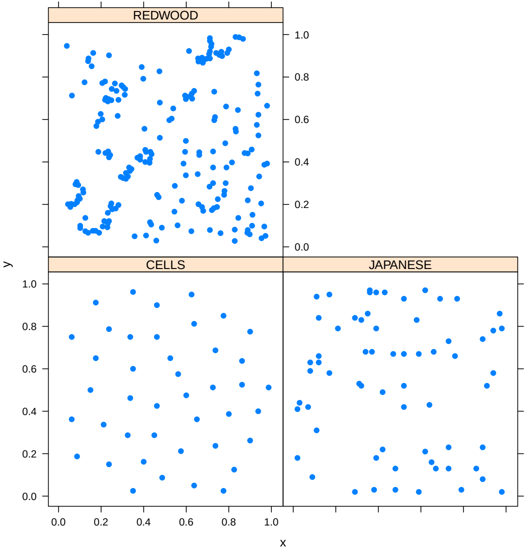
Examples

- Position of (one species of?) trees on a landscape
- Traces of meteor strikes
- Geomorphic features (reduced to points)
  - e.g., drumlin fields; Carolina bays
- Location of crime incidents
  - single type of crime, interaction between types of crime, interaction with point features, e.g., banks
- Distribution of grazing animals in a field
  - could study the evolution of the point-pattern over time
  - interaction between two species
- Bomb strikes around a target

Example used in ASDAR text

Redwood trees in a forest plot

Cell centres on a microscope slide



Japanese pines in a forest plot

coördinates normalized to a 1 x 1 square; reference: [3, Ch. 7]

Redwood trees

source: Strauss [14]
Hypotheses to test:

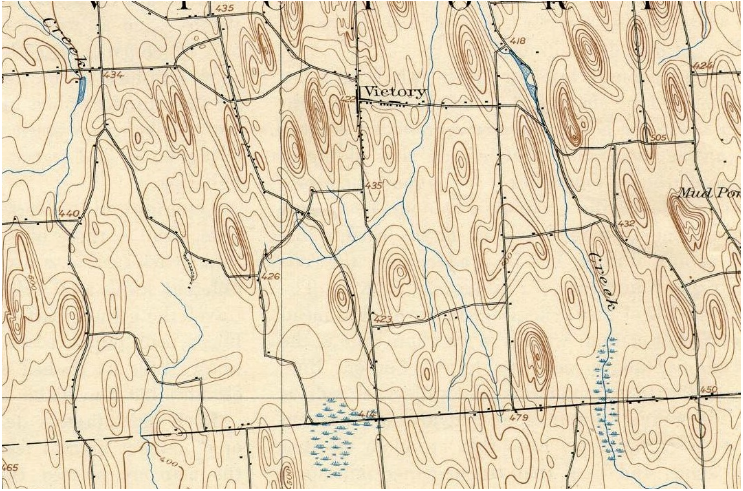
> *" It was felt that the seedlings would be* **scattered fairly randomly**, *except that* **a number of tight clusters** *would form around some of the . . . stumps present in the plot. A* **discontinuity in the soil**, *very roughly demarked by the diagonal line in the figure, was expected to cause a* **difference in clustering** *behaviour between regions I [upper left] and II [lower right]. Moreover, almost all the . . . stumps were situated in region II"*

area about 50 x 50 m = 0.25 ha

> *[1.8 m] "which was thought to be very roughly the range at which a pair of seedlings could 'interact' "*

So maximum density ≈ 1 500 trees

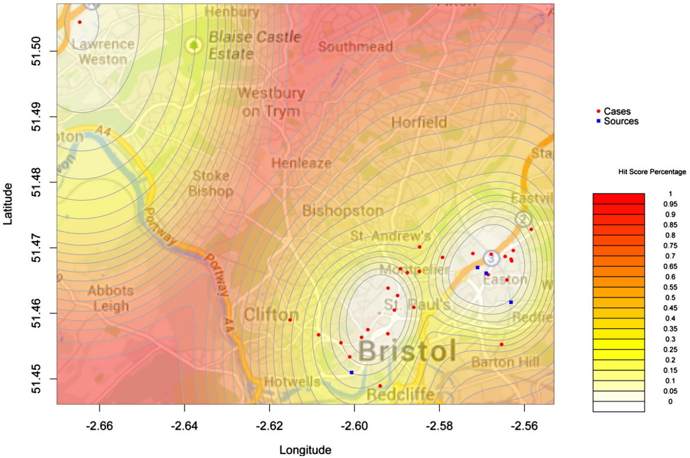
Example: drumlin field



Cayuga County, NY

tops of drumlins can be considered as "points"

Example: "crime" vs. suspect



source: Hauge, M. V. *et al.* (2016). *Tagging Banksy: using geographic profiling to investigate a modern art mystery*. **Journal of Spatial Science**, 1–6.
`http://doi.org/10.1080/14498596.2016.1138246`

Why analyze point patterns?

- We are interested in the **process** which produced the **pattern**
- We can only **observe** the pattern
- We want to **infer** the spatial data-generating process (sDGP)
  - make a **model** that fits the observed pattern
  - to confirm / reject / modify a geomorphological, ecological or social **theory**
- We may want to assign a **density** to every location in a region
  - probability of occurence, normalized by area
- We may want to **aggregate** counts / densities over some area

Process orders

1. **First**-order
   - just consider one set of points, each as **separate occurrences**
   - point distribution $\rightarrow$ suggests spatial (in-)homogeneity of process **intensity**
     - completely spatially-**random** (CSR)?
     - **clustered**?
     - **dispersed**?
     - **regularly-spaced**?
   - observed spatial **density**
2. **Second**-order process
   - **interaction** between (positions of) points
   - model **random** distribution, vs. **attraction**, vs. **repulsion**
3. both can (partially) depend on spatial **covariables**
   - e.g., regional trend or environmental factors

Poisson point process

- Example of a Data Generating Process (DGP)
- Named for the Poisson statistical distribution

$$f(k;\lambda) = \mathrm{Pr}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

- The number of points in a region is a **random variable** with a Poisson distribution, intensity λ

  homogeneous λ the same everywhere

  inhomogeneous λ can vary over space (cluster process)

- **Locations** of points in the region, given the intensity, are **completely random**.

Cox point process

- Example of a Data Generating Process (DGP)
- A type of Poisson point process, but ...
- ... the **intensity**, i.e., Poisson parameter $\lambda(s)$, **varies** across the region ...
- ... according to some **stochastic process**

Outline

Intensity

- The simplest measure: how many points on average per unit area
  - observed point **density** $\leftarrow$ **intensity** of the process that produced them
- **homogeneous** process: $\lambda(s) = \lambda = n/|A|$
  - s = spatial location; n = number of points; $|A|$ = area
  - does not vary over the area; **expected** value the same everywhere
- **inhomogenous** process: $\lambda(s)$ varies over the area
- depends on the **scale** at which we examine it (the **bandwidth**)
  - **broad** (wide): all points are taken together and the process is by definition homogeneous
  - **fine** (narrow): random fluctuations lead to different intensity estimates, even of same process
- **narrow** bandwidth $\rightarrow$ "spiky" map; **wide** bandwidth $\rightarrow$ (over-?)smooth map
- try to **match** bandwidth with the scale of the sDGP

Kernel density estimate : objective

- We suppose the process is **in**homogeneous
- We want to estimate the **intensity** at all locations
  - i.e., what is the area-normalized **probability** of a point at each location?
- Only uses the point-pattern itself (no covariables, no trend)
- Non-parametric (no model of the underlying process)
- This can lead to **hypotheses** about the spatial **process** (sDGP)
  - overlay on presumed covariable to see if they match
  - e.g., vegetation density vs. soil type polygons

Kernel density estimation: concept

Simplest is the **spherical** kernel; in 2D this is **circular**:

$$\hat{\lambda}(x) = \frac{N(b(x,h))}{\pi h^2}$$

where:

- $b(x,h)$ is a disc of radius h centred at x
  - $\hat{\lambda}(x)$ is estimated density per unit square
  - h is the **bandwidth**
  - larger $\rightarrow$ smoother estimates as kernel moves across the map
- N is the number of points in the disc
- denominator πh^2 is the area of that circle, normalizes the density

Kernel density estimation : with kernel function

This simple spherical count can be generalized with a **kernel function** that gives more weight to "nearby" portions of the disc:

$$\hat{\lambda}(x) = \frac{1}{h^2} \sum_{i=1}^{n} \kappa \left(\frac{||x - x_i||}{h} \right) / q(||x||)$$

where:

- $|| \cdot ||$ is the signed **norm**, usually the Euclidean distance between the target position (centre of kernel) and an observed point
- $\kappa(u)$ is a bivariate, symmetric **kernel function** of $u = ||x - x_i||/h$ (see next slide)
- $q(||x||)$ is an **edge-correction** factor (unobservable points outside the boundary)
- h is the **bandwidth**

Smoothing kernel

Example: the **quartic** (a.k.a. **biweight**) kernel:

$$\kappa(u) = \begin{cases} \frac{3}{\pi}\left(1 - ||u^2||\right)^2 & \text{if} \quad u \in (-1, 1) \\ 0 & \text{otherwise} \end{cases}$$

- as points are further away from the centre of the kernel, they get less weight in the density estimate
- *u* is signed according to the coördinate system, but then squared
- outside the normalized bandwidth $|u| \geq 1$ any points are *not* included in the density for a given location
- this is controlled by the *h* parameter

Choice of bandwidth

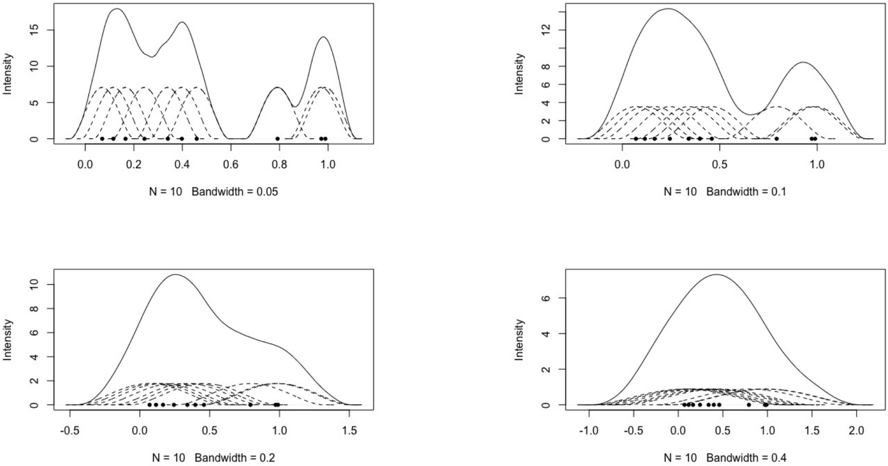
Which bandwidth "best" represents the (in)homogeneity of the point process?

> *"For any kernel function a* **small** *value of h may result in an estimated surface* $\hat{\lambda}(x)$ *that is* **too spiky**, *whereas a* **large** *h leads to smoother surfaces but may* **ignore local features** *of* $\hat{\lambda}(x)$.
>
> "**No simple recipe** *for the choice of bandwidth exists . . . .*
>
> "**Background information** *on the objects that form the pattern, such as dispersal distances for plants, might inform bandwidth choice . . .*
>
> *"[In] the absence of this the user should simply consider a number of values of h and choose the one that gives the* **most plausible result** *in the* **specific context**." – *Illian et al. [10, p. 115]*

Choice of bandwidth – relation to application

- How is "density" **perceived** in the application context?
  - Example: "dense" vegetation as cover for small animal (e.g., fox) vs. large animal (e.g., deer) – what is the radius that has to be "dense" before the animal feels secure using it as cover?
- What is a **realistic maximum density**?
  - If the process were homogeneous across the whole study area, what would be the maximum number of points per unit area? The kernel density at any point should not exceed this.
  - Example: mature trees with non-overlapping main canopies, radius $\approx 3\ m$, so no more than about $10000/(\pi \cdot 3)^2 \approx 350$ trees per hectare.
  - Example: redwood seedlings ≈ 6000 per hectare, ≈ 1500 in the example plot (Strauss)

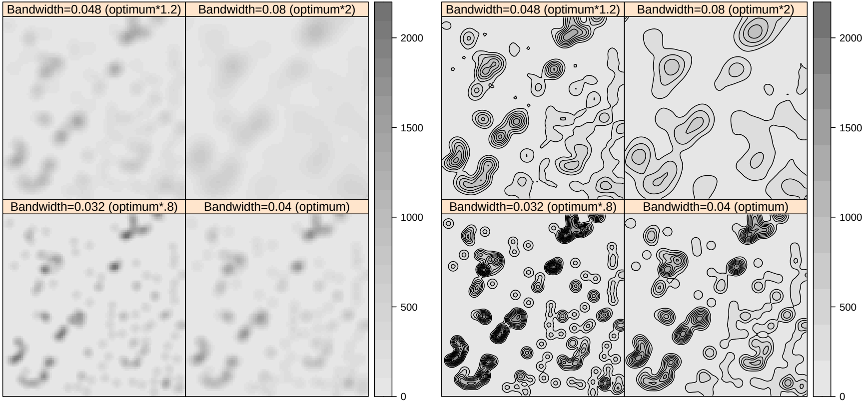
Effect of bandwidth: 1D example



$\lambda = 10$; solid line: biweight kernel density

dashed lines: contribution of each point to density

Effect of bandwidth: 2D example: redwood trees



– different bandwidth → different estimate;
– overall density 195 trees in the unit square
– density 2200 $(\text{unit})^{-1}$ in "hottest" spot / narrowest bandwidth → unrealistic ($>$ 1500)

Choice of bandwidth – Diggle's method

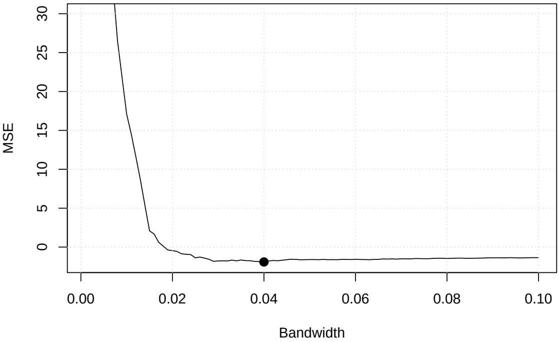
Diggle [8] provides a bandwidth estimation method for the Cox process: Minimize the mean square error (MSE) of the kernel smoothing estimator vs. actual counts.

$$\mathrm{MSE}(r) = E\left\{\left[\hat{\lambda}_r(x) - \lambda_r(x)\right]^2\right\}$$

- Try a series of bandwidths
- For each, compute the kernel density at each grid point, compare to counts
- Summarize across area
- Graph MSE vs. bandwidth, look for first minimum

Simpler: $h = 1/\sqrt{n}$.

Diggle's method – redwood trees



0.04 on unit square $\approx 0.04 * 50 = 2$ m at field scale; Strauss used 1.8 m
Homogeneous density: $1/\sqrt{195} = 0.072$.

Complete spatial randomness (CSR)

- The simplest **null hypothesis** of spatial distribution
- Points are distributed randomly; **no interaction**
  - no **clustering** (attraction, preferential conditions ...)
  - no **dispersion** (repulsion, limited resource in area ...)
- Produced by a (spatially) **homogeneous Poisson process**
- Constant **intensity** $\lambda(s) = \lambda > 0, \forall s \in A$
- The **probability** of any number of points in the same-sized region is the same, across the entire field.

A naïve approach – spatially-discretized Poisson distribution

- Spatially homogeneous **Poisson process**: count of "rare" events (points) in a discrete area follows the Poisson distribution
  - $f(k; \lambda) = \Pr(X = k) = (\lambda^k e^{-\lambda})/k!$
- a single parameter: process **intensity** $\lambda = \mu(X) = \sigma^2(X)$
  - i.e., the mean count, and its variance, are the intensity
- **Test** for Poisson process: **discretize** the area, compute λ, count the points, compare to Poisson distribution
- Q: how fine a discretization? A: about half of the cells should have Poisson expectation 0

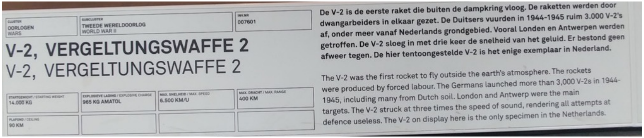
Example of Poisson counts: V-2 rocket strikes

- **point process**: WW2: 1,358 German V-2[1] rockets aimed at London (UK)
  - direct precursor of US Redstone rocket used in early manned space programme
- could not be shot down, so all that did not malfunction struck
- **point pattern**: clustered? dispersed? random?
  - inference about V-2 guidance system: ballistic (set at launch only) or adjusted en-route to target as in Cold War ICBMs?[2]
  - "once the rockets are up, who cares where they come down? / That's not my department, says Wernher von Braun."
    – Tom Lehrer, *Wernher von Braun* (1964)
  - **practical implications** for locating industry, population, civil defence, fire brigades . . .

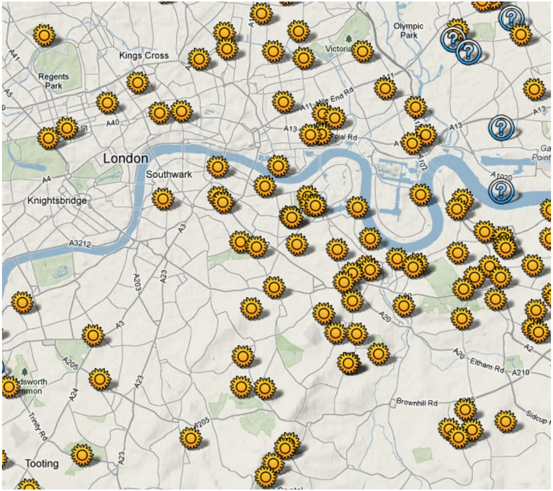
[1] "*Vergeltungswaffe*" = "Vengeance Weapon"

[2] `https://titanmissilemuseum.org/`

Dutch Military Museum, Soest (NL)



V-2 strikes point pattern



Source: `http://londonist.com/2009/01/london_v2_rocket_sitesmapped.php`

V-2 example: computations

- source: Clarke [6]
- Discretize 144 km² area into $n = 576$ squares of 0.25 km²
  - this corresponds to the bandwidth
- Poisson distribution:
  - Total hits 537, so intensity $\lambda = 537/576 = 0.932 \approx 1$
  - $f(0, 0.932) = e^{-0.932} = 0.39377$, i.e., about 40% of cells with no expected hits; $f(0, 0.932) \cdot 576 \approx 227$ grid cells
    - this is probably why he chose 0.25 km² cells
  - $f(2, 0.932) = (0.932^2 e^{-0.932})/2 = 0.171$; $f(2, 0.932) \cdot 576 = 98.54 \approx 99$ grid cells with exactly 2 expected hits

V-2 example: results

n	expected	actual
0	226.74	229
1	211.39	211
2	98.54	93
3	30.62	35
4	7.14	7
≥ 5	1.57	1

$\chi^2 = \sum \frac{(O-E)^2}{E} = 1.17$; $\Pr(\chi^2(4) > 1.17) = 0.88$

Conclusion: hits are not provably different from the null hypothesis of a homogeneous Poisson process; within the target area the distribution is CSR

The G function

- A more sophisticated approach for evaluating CSR and deviations from it
- Measures the distribution of distances from an arbitrary point to its **nearest neighbour**. The **empirical** function is:

$$
\begin{aligned}
d_i &= \min_j \{d_{ij}, \forall j \neq i \in S\}, i = 1, \ldots, n \\
\widehat{G}(r) &= \frac{\{\# d_i : d_i \leq r, \forall i\}}{n}
\end{aligned}
$$

  - this is the number ($\#$) of points which have *at least one* neighbour within some threshold distance r, *normalized* by the total number of points n in the pattern S.
- This is a **continuous** function of r – no need to discretize

G function under CSR

The result of the homogeneous Poisson process is the **theoretical** function:

$$G(r) = 1 - e^{\lambda \pi r^2}$$

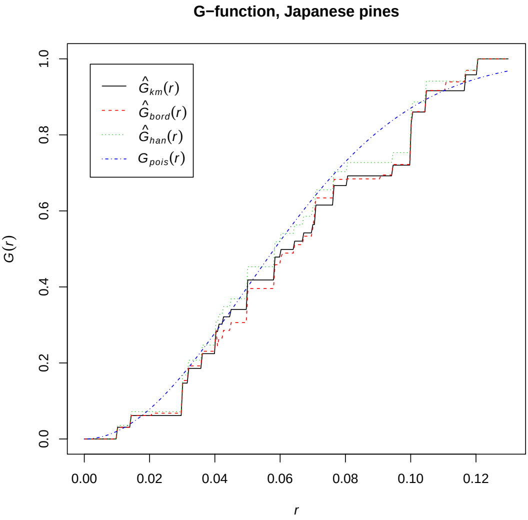
where λ is the process **intensity**, i.e., mean number of points per unit area.

Clustered patterns: $\widehat{G}(r) > G(r)$ (more nearby points than expected under CSR)

Dispersed (regular) patterns: $\widehat{G}(r) < G(r)$ (fewer ...)

These are all evaluated at *any* threshold (radius) r, can infer radius of clustering/dispersion

Example G function



- G_{pois}: theoretical CSR
- G_{km}: empirical distribution
- G_{bord}, G_{han}: border-corrected empirical distributions

Example G function: interpretation

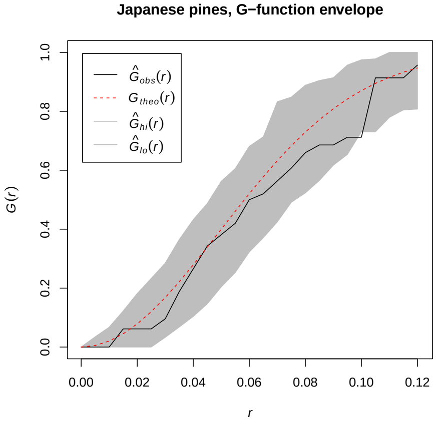
- All points have a nearest neighbour within 0.13 normalized units
  - compute inter-point distances with `nndist`
- Empirical closely follows theoretical CSR
- Some deviations (dispersal, below the theoretical line) near 0.02 and 0.10 normalized units
- The border has little effect

Envelope

A method to determine **confidence intervals** for the G function:

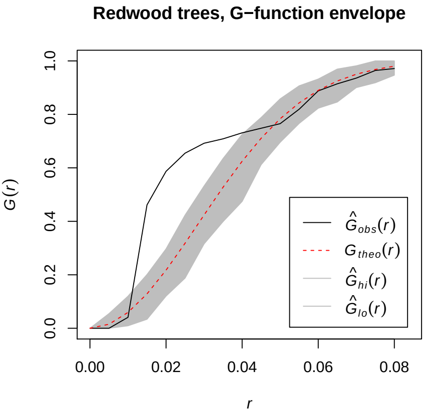
1. compute overall intensity λ
2. repeatedly **simulate** a CSR process with this intensity
3. compute G function for each simulated process
4. the observed pattern is assumed to be a **single realization** of the process
5. realized G function inside the *envelope* $\rightarrow$ evidence that the null hypothesis of CSR can not be rejected

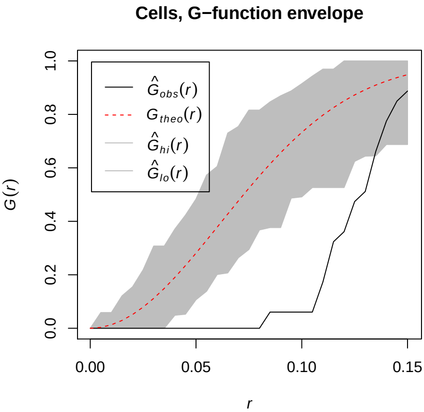
Envelope – example



Conclusion: can not reject the null hypothesis of CSR

Examples of clustered and dispersed patterns





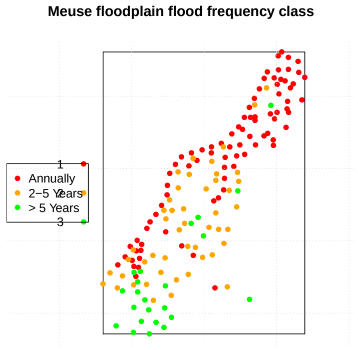
clustered: $\widehat{G}(r) > G(r)$

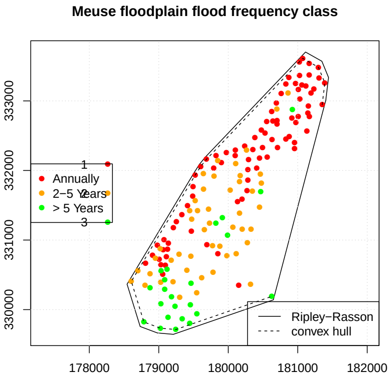
dispersed: $\widehat{G}(r) < G(r)$

The F function

- another way to examine first-order properties
- also called the **empty space** function
- distribution of the distances from an arbitrary **location** (not necessarily a point) to its nearest **observed point**
  - measures the average empty space between observed points.
- it has the same theoretical distribution as the G function
- **sensitive to window size** if there is “empty” space at edges

Effect of window size on F function: windows

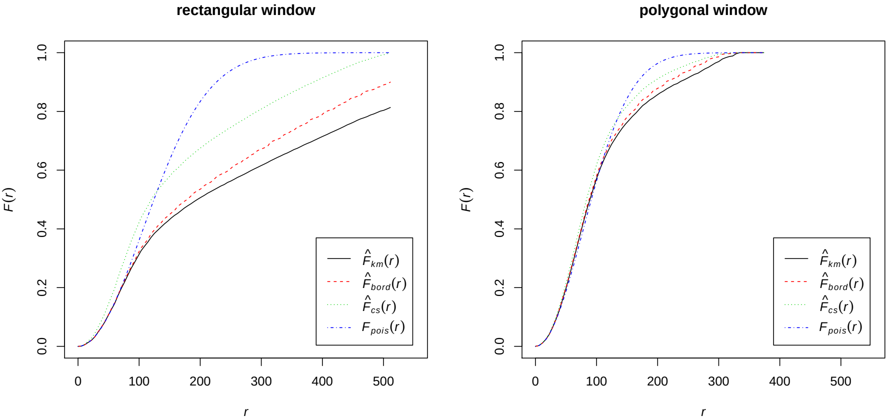




Rectangular

limiting polygon

Effect of window size on F function: results



polygon: much higher proportion of nearest points found within given r than for rectangular bounding box

The J function

- Combines G (point-to-point) and F (space-to-point)
- $J(r) = (1 - G(r))/(1 - F(r))$
- Expected value under CSR $= 1$
  - because G and F have the same expectation under CSR
- $J(r) < 1$ implies clustering, $J(r) > 1$ implies dispersion
- advantage: not sensitive to edge effects

Outline

Second-order properties

- these measure the **interactions** between “events” of one or more sDGP
  - the “events” result in observed points of one or two two types
- **competition** (dispersal) within or between processes
  - two trees can not occupy the same position (within some radius)
  - allelopathy: chemicals from one species spread out to some radius, prevent others from growing
- **synergy** (clustering) within or between processes
  - seedlings from one tree sprout nearby? seed dispersal, soil type . . .
  - earthquakes facilitated by fracking wells?

Ripley's K function

- measures the **number** of events (points) up to a given distance from an event (point)
  - so, counts **all neighbours** up to that distance
- if $E[.]$ is expectation, $N_0(s)$ is the number of events up to distance s from any event, λ is process intensity:

$$\widehat{K}(s) = \lambda^{-1} E[N_0(s)]$$

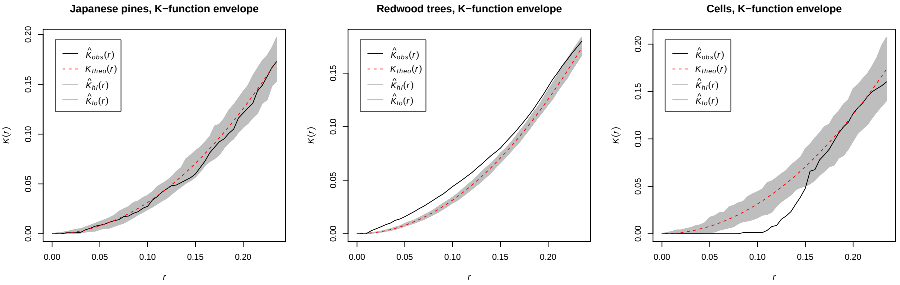
Computing the univariate K function

- unbiased estimator:

$$\widehat{K}(s) = (n(n-1))^{-1}|A| \sum_{i=1}^{2} \sum_{j \neq i} w_{ij}^{-1} |\{x_j : d(x_i, x_j) \leq s\}|$$

- weights w_{ij}: proportion of $|A|$ occupied by circle centred at x_i with radius $d(x_i, x_j)$
  - corrects for edge effects
- for a homogeneous process, $K(s) = \pi s^2$
  - i.e., number is proportional to circle **area**
- application: graph $\widehat{K}$ and K vs. radius s, compare actual to theoretical at each s
- can compute envelopes as for G and F functions

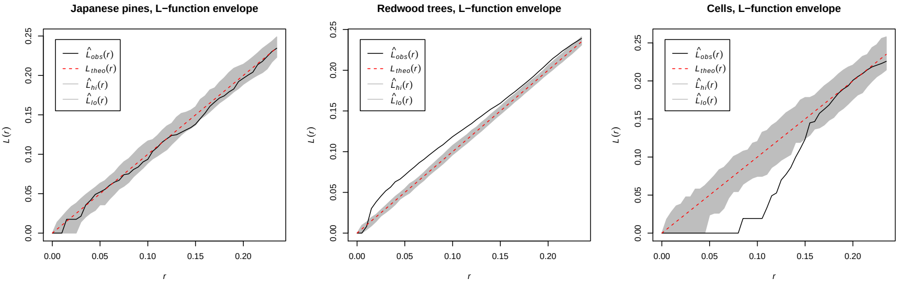
K function envelopes



- *redwoods*: more than expected events near to an arbitrary event; "attraction" $\rightarrow$ clustering
- *cells*: fewer nearby events, after 0.15 units there is no more "repulsion" $\rightarrow$ dispersal; gives scale of sDGP
- **interpret** in terms of sDGP (note the radius)

Besag's L function

This is a linearization of the K-function which makes it easier to compare theoretical and actual values at narrow separations (low values of the radius):
$L(r) = \sqrt{\frac{K(r)}{\pi}}$

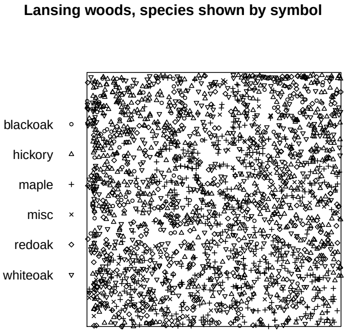


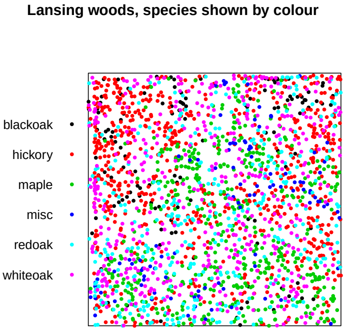
Outline

Marked point patterns

- The previous measures were just considering point **location**, with no information about what the point represents.
- But a point pattern can be **marked**: each point is **labelled** with some attribute
  - Example: tree species in a forest plot (**categorical** mark)
  - Example: size of trees in a forest plot (**continuous** mark)
  - Example: **time** of observation of a point (e.g., wildfires)
- Can analyze each sub-pattern separately (as with unmarked patterns)
- Can analyze **interactions** between patterns

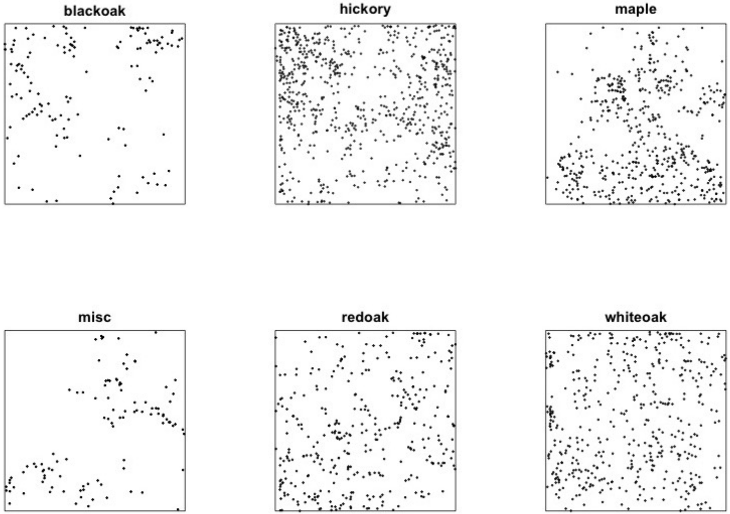
Example marked pattern





Location of trees in Lansing woods, **marked** by their **species**

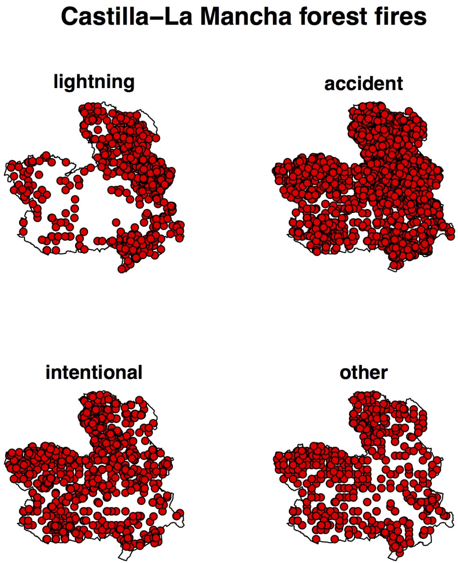
Marked pattern divided into unmarked patterns



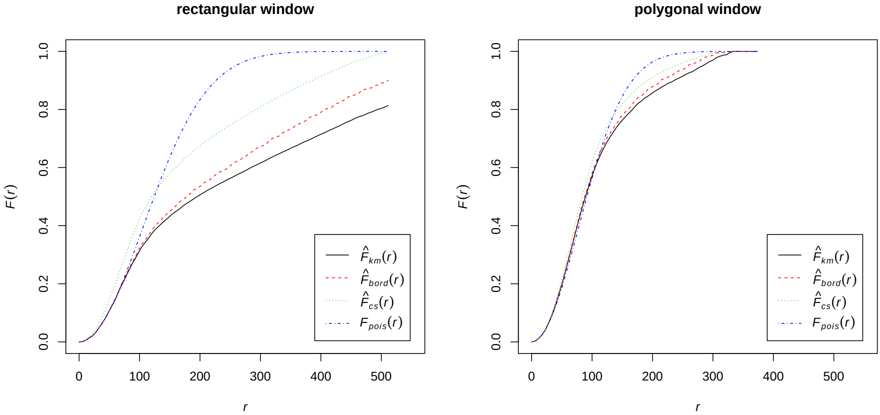
Interactions in marked patterns

- Q: Do patterns with different marks "influence" each other?
  - Simpler: what is their relation?
- A: Bi- and multi-variate versions of G, K, L, J functions (not F)
  - $G_{ij}(r)$: the distribution of the distance from a typical point of type (mark) i to its nearest point of type j.
  - $K_{ij}(r)$: given intensity λ_j of type j, $\lambda_j K_{ij}(r)$ is the expected number of *additional* points of type j within a distance r of a typical point of type i.
  - empirical $>$ theoretical $\rightarrow$ clustering & vice-versa.

Example marked pattern: forest fire type

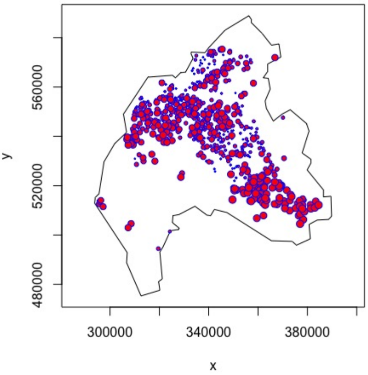


Example crossed-K function



Intentional vs. caused by lightning
After 15 km radius there is “repulsion”

Marked pattern: time of occurrence



Foot-and-mouth disease, northern Cumbia (England), 2001; from R package `stpp`, dataset `fmd`

more recent → larger symbol

Simple spatio-temporal analysis

- Marked point-pattern, marks as time of occurrence
- Use the crossed K etc. functions to assess interaction

Outline

Models of spatial data-generating processes

- We often want to infer the sDGP that produced the observed pattern
- We have competing **models**
  - CSR, attraction, repulsion
  - dependence on spatially-distributed **covariables**
  - interaction with other sDGP

Model development and evaluation

- formulate hypothetical model
  - based on *a priori* knowledge or theory
  - example from V-2 rockets: no in-flight guidance system
  - example from ecology: allelopathy
  - example from criminology: distance from source, “no action” buffer
- parameterize with observed pattern(s) and possibly covariables
- evaluate **goodness-of-fit**
  - good fit $\rightarrow$ **evidence** (*not* “proof”!) for the hypothesized sDGP

A general model of sDGP

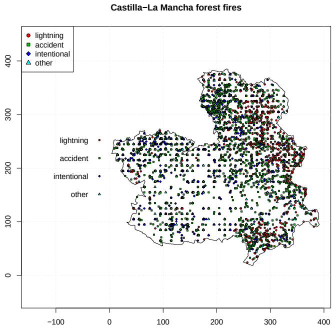
- spatial **trend** – a function of the coördinates
- **interaction** between events
- influence of **covariates** (other than trend) on events
- General model:

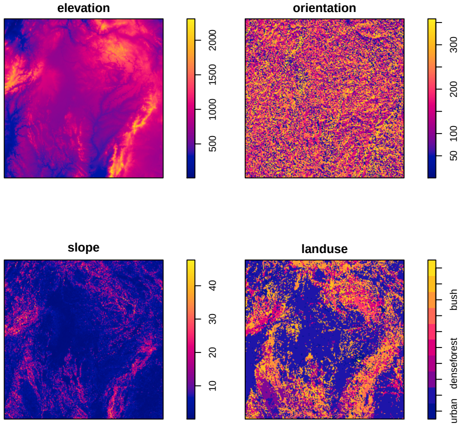
$$\lambda(s, \mathbf{x}) = \exp(\psi^T B(s) + \phi^T C(s, \mathbf{x}))$$

  - $B(s)$: depends only on location (trend and/or spatial covariates)
  - $C(s, \mathbf{x})$ also depends on other points
  - either may be absent (simpler process)
  - details in Baddeley and Turner [1]
- compute with `spatstat` package function `ppm`

Example sDGP covariates

200 m grid covariates

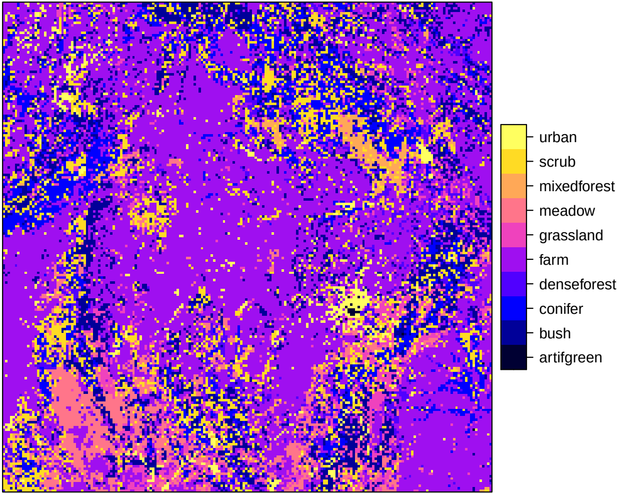




Pattern of fires

spatial covariates

sDGP covariate – land cover classes



sDGP models

1. **Null**: complete spatial randomness (CSR)
   - homogeneous Poisson process
2. **Trend**: regional trend in the coördinates in conditional density
   - **in**homogeneous Poisson process
   - may be due to a trend in an environmental factor, e.g., rainfall
3. **Covariates**: dependence on mapped covariates
   - e.g., some land uses more prone to fire
4. **Interaction**: clustering or repulsion between points: **Strauss** processes
5. **Combination** of trend and/or covariates and/or interaction

Strauss process

- sDGP causes clustering or repulsion between points
- $\lambda(u, \mathbf{x}) = \beta\gamma^{t(u,\mathbf{x})}$
  - $\lambda(u, \mathbf{x})$: intensity of pattern $\mathbf{x}$ at location u;
  - β: overall homogeneous intensity;
  - γ: **interaction** parameter $0 \leq \gamma \leq 1$
  - $t(u, \mathbf{x})$: the number of points closer than the **interaction radius** r
- Interpretation:
  - $\gamma = 0 \implies \lambda = 0$: no chance of finding another point;
  - $\gamma < 1$: chance of a second point is reduced;
  - $\gamma = 1$ equivalent to a Poisson process (CSR)
- Specify r, fit γ from observed pattern

Model selection

- Based on reasonable **hypotheses** of the sDGP
- Evaluate by statistical **likelihood** that the observed pattern was generated by the hypothesized sDGP
- Compare **alternate models** by **relative** likelihood
- Interpret **parameters**: strength of inter-point interaction; coefficients of trend surfaces or covariate models; relative strength of trend and interaction components
  - Do we now understand the sDGP better? Are our hypotheses confirmed, rejected, modified?

Comparing models by likelihood

1. **Poisson** model: CSR
2. **Landuse** model: observed density depends only on landuse
3. **Strauss process** model: observed density depends only on interaction between events
4. Landuse + Strauss model: **combined**

```
            model likelihood
1         Poisson    -8562.0
2         Landuse    -8532.2
3         Strauss    -6749.9
4 Landuse + Strauss    -6730.8
```

Landuse explains little; interaction explains a lot; combined is best

Interpreting model parameters

Interaction: $\gamma > 1$ suggests clustering; β is overall log-density

```
> exp(coef(m.strauss.4))
(Intercept) Interaction
   0.017543    1.080151
```

Land use:

```
Fitted coefficients for trend formula:
       (Intercept)        landusefarm      landuseconifer
          -4.41477            0.28745             0.59341
```

Higher density in coniferous forest than on farms

Trend surface model

```
> (m.ts1 <- ppm(clmfires.i, trend = ~polynom(x, y, 1),
     interaction = NULL))

Nonstationary Poisson process

Log intensity:  ~x + y

Fitted trend coefficients:
(Intercept)           x           y
 -3.2878809  -0.0057327   0.0028847

              Estimate        S.E.     CI95.lo     CI95.hi Ztest    Zval
(Intercept) -3.2878809 0.07317041 -3.4312923 -3.1444695   *** -44.935
x           -0.0057327 0.00027073 -0.0062633 -0.0052021   *** -21.175
y            0.0028847 0.00030260  0.0022916  0.0034778   ***   9.533
```

Comparing trend surface models by ANOVA

1. **Poisson** model: CSR
2. **First-order trend** in the coördinates
3. **Second-order trend** in the coördinates

```
> anova(m.ts2, m.ts1, m.pois)

Analysis of Deviance Table

Model 1: ~x + y + I(x^2) + I(x * y) + I(y^2)   Poisson
Model 2: ~x + y   Poisson
Model 3: ~1   Poisson
  Npar Df Deviance
1    6
2    3 -3     -126
3    1 -2     -505
```

Second-order model is a bit better than the others.

Prediction from sDGP models

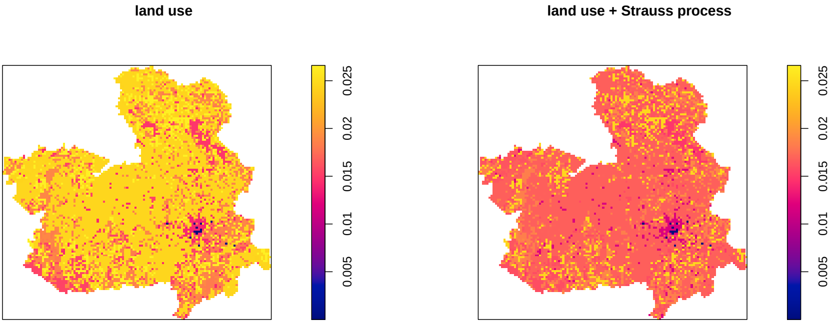
- Once an sDGP model is **fit**, it has **parameters**
  - e.g., coefficients of a trend surface or covariate model, strength of inter-point interaction.
- This model can be applied to new situations: across an area (from trend, points), with renewed covariates
- `spatstat` function `predict.ppm` predicts from models fit with `ppm`.

Predictions

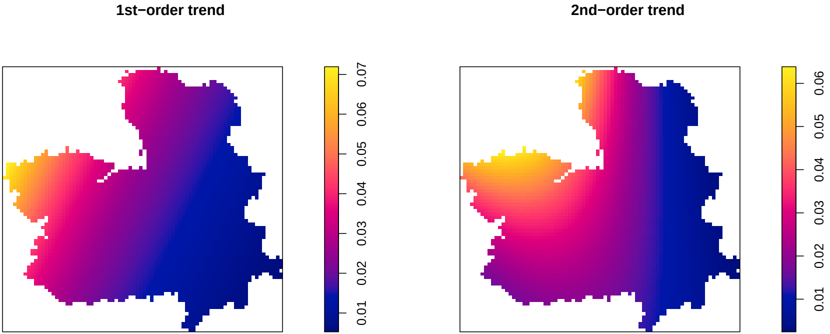
```
> pred.lu.strauss <- predict(m.lu.strauss.4,
                             covariates=clmfires.extra$clmcov200)
> summary(pred.lu.strauss)

real-valued pixel image
128 x 128 pixel array (ny, nx)
enclosing rectangle: [4.1311, 391.38] x [18.565, 385.19] kilometres
dimensions of each pixel: 3.03 x 2.8642 kilometres
Image is defined on a subset of the rectangular grid
Subset area = 79462.0730449286 square kilometres
Subset area fraction = 0.56
Pixel values (inside window):
range = [2.2603e-07, 0.024494]
integral = 1402.8
mean = 0.017653
```

Model predictions – covariate, Strauss process



Model predictions – trend surface

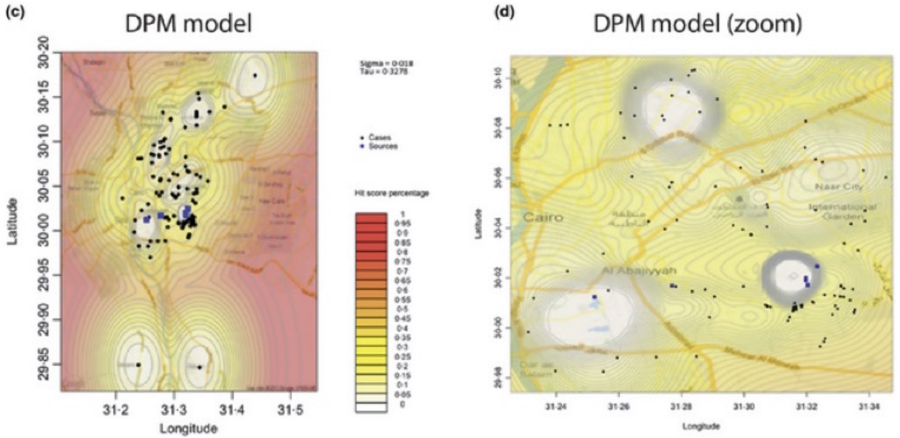


Outline

Empirical source finding

- **Objective**: find **possible sources** from a set of **occurrences**.
- **Approach**: empirical criminal geographic targeting (CGT)
- **Approach**: Dirichlet Process Mixture (DPM)
  - no predefined number of clusters; algorithm finds most probable and their location

Example of source finding



139 *Plasmodium vivax* cases in Cairo, Egypt
observed data points; black circles; empirically-identified **sources**:blue squares.

source: Verity et al. [15]

Bayesian models

- Explicit probability models
- Can incorporate knowledge via **prior probabilities**
  - distribution and parameters, e.g., normal with a prior mean and variance
  - e.g., weights of passengers and luggage on a flight
- **Update** probabilities based on **evidence**
  - for realistic models computationally-intensive (e.g., Markov chain Monte Carlo)
- Result is a **posterior probability** of parameters of the chosen distribution

Bayesian Hierarchical Models

- The model of the effect (e.g., spatial pattern of a disease) is specified as a **hierarchical** set of layers
- Each layer accounts for different sources of spatial variation
- E.g., Besag et al. [2]: sum of:
  1. a **spatially-correlated** variable;
  2. an **area-independent** effect (local heterogeneity)

INLA

- **"Integrated Nested Laplace Approximation"** to the posterior marginals of model parameters
- INLA computes only **relative** posterior distributions for *latent Gaussian models*
  - these are enough for many applications, e.g., relative risks in disease mapping
- References: Rue et al. [13], Illian et al. [11]; Bivand et al. [4]; `http://www.r-inla.org`

Outline

Further reading

Theory: Boots and Getis [5], Diggle [7], Ripley [12], Illian et al. [10] (also as e-book)

Applications: Gatrell et al. [9]

In R: Bivand et al. [3, Ch. 7] (also as e-book); Baddeley and Turner [1]

Web pages

- Yuri Zhukov (Harvard):
  `http://www.people.fas.harvard.edu/~zhukov/spatial.html`, unit 4
- Roger Bivand (Bergen):
  `http://www.bias-project.org.uk/ASDARcourse/`, unit 5

R packages

- **spatial**: Functions for Kriging and Point Pattern Analysis (Ripley, Bivand, Venables)
- **spatstat**: Spatial Point Pattern Analysis, Model-Fitting, Simulation, Tests (Baddely, Turner *et al.*)
- **splancs**: Spatial and Space-Time Point Pattern Analysis (Bivand, Rowlingson, Diggle *et al.*)
- **Rgeoprofile**: Geographic profiling in R (Stevenson, Verity, Nichols, LeComber)
- **R-INLA**: Integrated Nested Laplace Approximation (INLA)

See also `https://cran.r-project.org/web/views/Spatial.html`

References I

[1] A. Baddeley and R. Turner. Modelling spatial point patterns in R. In A. Baddeley, P. Gregori, J. Mateu, R. Stoica, D. Stoyan, J. Berger, S. Fienberg, J. Gani, K. Krickeberg, I. Olkin, and B. Singer, editors, *Case Studies in Spatial Point Process Modeling*, volume 185 of *Lecture Notes in Statistics*, pages 23–74. Springer New York, 2006. ISBN 978-0-387-31144-9.

[2] J. Besag, J. York, and A. Mollié. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–59, 1991.

[3] R. S. Bivand, E. J. Pebesma, and V. Gómez-Rubio. *Applied Spatial Data Analysis with R*. UseR! Springer, 2008. `http://www.asdar-book.org/`.

[4] Roger Bivand, Virgilio Gómez-Rubio, and Høavard Rue. Spatial data analysis with R-INLA with some extensions. *Journal of Statistical Software*, 63(20), 2015.

[5] B. N. Boots and A. Getis. *Point pattern analysis*. Number v. 8 in Scientific geography series. Sage Publications, Newbury Park, Calif, 1988. ISBN 0803922450.

[6] R. D. Clarke. An application of the Poisson distribution. *Journal of the Institute of Actuaries*, 72:481, 1946.

[7] P. Diggle. *Statistical analysis of spatial point patterns*. Arnold, London, 2nd edition, 2003. ISBN 0122158504.

[8] Peter Diggle. *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns*. CRC Press, Boca Raton, third edition. edition, 2014. ISBN 978-1-4665-6024-6. doi: 10.1201/b15326.

References II

[9] A. C. Gatrell, T. C. Bailey, P. J. Diggle, and B. S. Rowlingson. Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers*, 21(1):256, 1996. doi: 10.2307/622936.

[10] Janine Illian, Antti Penttinen, Helga Stoyan, and Dietrich Stoyan. *Statistical analysis and modelling of spatial point patterns*. John Wiley, 2008. ISBN 9780470014912.

[11] Janine B. Illian, Sara Martino, Sigrunn H. Sørbye, Juan B. Gallego-Fernández, María Zunzunegui, M. Paz Esquivias, and Justin M. J. Travis. Fitting complex ecological point process models with integrated nested Laplace approximation. *Methods in Ecology and Evolution*, 4(4):305–315, Apr 2013. doi: 10.1111/2041-210x.12017.

[12] B D Ripley. *Spatial statistics*. John Wiley and Sons, New York, 1981.

[13] Havard Rue, Sara Martino, and Nicolas Chopin. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 71:319–392, 2009. doi: 10.1111/j.1467-9868.2008.00700.x.

[14] David J. Strauss. A model for clustering. *Biometrika*, 62(2):467–475, 1975. doi: 10.2307/2335389.

[15] Robert Verity, Mark D. Stevenson, D. Kim Rossmo, Richard A. Nichols, and Steven C. Le Comber. Spatial targeting of infectious disease control: identifying multiple, unknown sources. *Methods in Ecology and Evolution*, 5(7):6470–655, Jul 2014. doi: 10.1111/2041-210X.12190.