Analysing spatial patterns of points

Adrian Baddeley
CSIRO & University of Western Australia

Includes joint work with
Ya-Mei Chang, Martin Hazelton, Jesper Møller,
Gopal Nair, Ege Rubak, Yong Song, Rolf Turner
Shapley galaxy concentration
Tree deaths in Perth’s groundwater catchment

+ tree death
● water bore
Murchison (Western Australia) gold deposits
Common elements
Common elements

- spatial locations of ‘things’/ ‘events’
Common elements

- spatial locations of ‘things’/ ‘events’

  spatial point pattern
Common elements

- spatial locations of ‘things’/ ‘events’

  spatial point pattern

- additional data
Common elements

- spatial locations of ‘things’/ ‘events’
  - spatial point pattern

- additional data
  - spatial covariates
Common elements

- spatial locations of ‘things’/ ‘events’
  
  spatial point pattern

- additional data
  
  spatial covariates

- we want to model the dependence of points on covariates,
Common elements

- spatial locations of ‘things’/ ‘events’
  
  **spatial point pattern**

- additional data
  
  **spatial covariates**

- we want to model the dependence of points on covariates, and the dependence *between* points (e.g. clustering)
A *point process* is a random mechanism that generates a random pattern of points.
A *point process* is any random mechanism that generates a random pattern of points.
A *point process* is any random mechanism that generates a random pattern of points.

The points could be
A *point process* is *any* random mechanism that generates a random pattern of points.

The points could be *things* (e.g. trees);
A point process is any random mechanism that generates a random pattern of points.

The points could be things (e.g. trees); events (e.g. accidents);
A point process is any random mechanism that generates a random pattern of points.

The points could be things (e.g. trees); events (e.g. accidents); detections (e.g. wild animal sightings);
A point process is any random mechanism that generates a random pattern of points.

The points could be things (e.g. trees); events (e.g. accidents); detections (e.g. wild animal sightings); sample points (e.g. geological drill holes);
A *point process* is any random mechanism that generates a random pattern of points.

The points could be *things* (e.g. trees); *events* (e.g. accidents); *detections* (e.g. wild animal sightings); *sample points* (e.g. geological drill holes); *absences* (e.g. no gold at this location);
A *point process* is any random mechanism that generates a random pattern of points.

The points could be *things* (e.g. trees); *events* (e.g. accidents); *detections* (e.g. wild animal sightings); *sample points* (e.g. geological drill holes); *absences* (e.g. no gold at this location); or *hypothetical / unobservable* things (e.g. all whales in ocean).

Long history:
A *point process* is *any* random mechanism that generates a random pattern of points.

The points could be *things* (e.g. trees); *events* (e.g. accidents); *detections* (e.g. wild animal sightings); *sample points* (e.g. geological drill holes); *absences* (e.g. no gold at this location); or *hypothetical / unobservable* things (e.g. all whales in ocean).

**Long history:**

1898: deaths by horse kick in the Prussian Cavalry.
A *point process* is any random mechanism that generates a random pattern of points.

The points could be *things* (e.g. trees); *events* (e.g. accidents); *detections* (e.g. wild animal sightings); *sample points* (e.g. geological drill holes); *absences* (e.g. no gold at this location); or *hypothetical / unobservable* things (e.g. all whales in ocean).

**Long history:**

1898: deaths by horse kick in the Prussian Cavalry.
1940: general theory of point processes.
The statement that "galaxies are a point process" is vacuous.
The statement that “galaxies are a point process” is vacuous. It just says that their locations were determined by some (unspecified) random mechanism.
The statement that “galaxies are a point process” is **vacuous**

It just says that their locations were determined by some (unspecified) random mechanism.

To make progress we need to specify a particular stochastic model — a *point process model*. 
Poisson point process

The simplest probability model for a point pattern is a *Poisson point process*
Poisson point process

The simplest probability model for a point pattern is a *Poisson point process*. 
The simplest probability model for a point pattern is a *Poisson point process*.

A.K.A. “Completely Random”— point locations are independent of each other; different areas of the pattern are independent of each other.
Statistical methodology for point processes
Until \( \sim 1990 \) it was widely believed that standard statistical methodology was "intractable" for spatial point processes.
Until ~ 1990 it was widely believed that standard statistical methodology was \textbf{“intractable”} for spatial point processes.
Until \( \sim 1990 \) it was widely believed that standard statistical methodology was “intractable” for spatial point processes (except Poisson).
Until \( \sim 1990 \) it was widely believed that standard statistical methodology was “intractable” for spatial point processes (except Poisson).

Instead, statisticians developed specialised methods.
Statistical methodology for point processes

- Until ~ 1990 it was widely believed that standard statistical methodology was “intractable” for spatial point processes (except Poisson)

- Instead, statisticians developed specialised methods
  - summary statistics (Ripley’s $K$-function)
Until ~ 1990 it was widely believed that standard statistical methodology was "intractable" for spatial point processes (except Poisson).

Instead, statisticians developed specialised methods:

- summary statistics (Ripley’s $K$-function)
- composite likelihood
Until ~ 1990 it was widely believed that standard statistical methodology was “intractable” for spatial point processes (except Poisson).

Instead, statisticians developed specialised methods:

- summary statistics (Ripley’s $K$-function)
- composite likelihood
- Markov chain Monte Carlo
Until $\sim$ 1990 it was widely believed that standard statistical methodology was “intractable” for spatial point processes (except Poisson).

Instead, statisticians developed specialised methods:

- summary statistics (Ripley’s $K$-function)
- composite likelihood
- Markov chain Monte Carlo

Other sciences developed their own methods.
Statistical methodology for point patterns

Critique
Statistical methodology for point patterns

Critique

- Clunky

Inflexible, slow, temperamental
Critique

- **Clunky**
  Inflexible, slow, temperamental

- **Doesn’t answer real world questions**

  e.g. “How confident are you that there is gold in this region?”
Critique

■ **Clunky**

Inflexible, slow, temperamental

■ **Doesn’t answer real world questions**

  e.g. “How confident are you that there is gold in this region?”

■ **Immature**

  Doesn’t provide standard statistical tools e.g. confidence interval, goodness-of-fit, model diagnostics
Cyclones in NE Australia
“Cyclone frequency has been steadily decreasing from the 19th century to the present”
Linear regression
Linear regression
Linear regression: confidence interval
Linear regression: prediction interval

![Graph showing linear regression with prediction interval shaded area.](image-url)
Linear regression diagnostics: residuals

![Graph showing residuals vs x values]
Linear regression diagnostics: leverage
Linear regression diagnostics: influence
Nonparametric regression
Statistical tools
Statistical methodology
Point patterns in geoscience
Pixel Logistic Regression

Proposed by statistician John Tukey 1972, developed by geologist Frits Agterberg
Pixel Logistic Regression

Proposed by statistician John Tukey 1972, developed by geologist Frits Agterberg

- Divide survey region into pixels
Pixel Logistic Regression

Proposed by statistician John Tukey 1972, developed by geologist Frits Agterberg

- Divide survey region into pixels
- Set pixel value to 1 if it contains data points, otherwise 0
Pixel Logistic Regression

Proposed by statistician John Tukey 1972, developed by geologist Frits Agterberg

- Divide survey region into pixels

- Set pixel value to 1 if it contains data points, otherwise 0
Pixel Logistic Regression

Proposed by statistician John Tukey 1972, developed by geologist Frits Agterberg

- Divide survey region into pixels

- Set pixel value to 1 if it contains data points, otherwise 0

- Analyse 0/1 pixel values using logistic regression
Pixel Logistic Regression
Pixel Logistic Regression

For pixel $j$, set

$$y_j = \begin{cases} 
1 & \text{if pixel } j \text{ contains any data points} \\
0 & \text{otherwise}
\end{cases}$$
Pixel Logistic Regression

■ For pixel $j$, set

$$y_j = \begin{cases} 
1 & \text{if pixel } j \text{ contains any data points} \\
0 & \text{otherwise}
\end{cases}$$

■ For pixel $j$, let $z_j$ be the value of covariate $Z$ at the centre of pixel
Pixel Logistic Regression

For pixel $j$, set

$$y_j = \begin{cases} 
1 & \text{if pixel } j \text{ contains any data points} \\
0 & \text{otherwise}
\end{cases}$$

For pixel $j$, let $z_j$ be the value of covariate $Z$ at the centre of pixel.

Fit logistic regression of $(y_j)$ on $(z_j)$

$$\log \frac{p_j}{1 - p_j} = \beta_0 + \beta_1 z_j$$

where $p_j = P(Y_j = 1)$. 
Murchison gold deposits

Take $z_j = \text{distance from pixel } j \text{ to nearest fault line}$
Fitted presence probabilities $p_j$ in each pixel
Known “facts” about pixel logistic regression
Known “facts” about pixel logistic regression

According to GIS literature
Known “facts” about pixel logistic regression

According to GIS literature

- “logistic regression is a nonparametric technique”
Known “facts” about pixel logistic regression

According to GIS literature

- “logistic regression is a nonparametric technique”
- results depend on choice of pixel size
Known “facts” about pixel logistic regression

According to GIS literature

- “logistic regression is a nonparametric technique”
- results depend on choice of pixel size
- “difficult to interpret the fitted parameters”
Known “facts” about pixel logistic regression

According to GIS literature

- “logistic regression is a nonparametric technique”
- results depend on choice of pixel size
- “difficult to interpret the fitted parameters”
- small pixels $\Rightarrow$ numerical problems
Effect of pixel size
Effect of pixel size

- results using different pixel sizes are **incompatible**
Effect of pixel size

- Results using different pixel sizes are incompatible.
- There is no point process in continuous space that is consistent with logistic regression on every pixel grid.

Small pixels

For small pixel size, pixel logistic regression is equivalent to assuming a Poisson point process with loglinear intensity

\[ \lambda(u) = \exp(\beta_0 + \beta_1 Z(u)) \]

where \( Z(u) \) is the covariate value at location \( u \).
Small pixels

For small pixel size, pixel logistic regression is equivalent to assuming a **Poisson point process** with loglinear intensity

$$
\lambda(u) = \exp(\beta_0 + \beta_1 Z(u))
$$

where $Z(u)$ is the covariate value at location $u$.

Warton & Shepherd, **Poisson point process models solve the “pseudo-absence problem” for presence-only data in ecology.** *Annals of Applied Statistics* 4 (2010)

Murchison gold deposits

Loglinear Poisson point process model

Predicted intensity of gold deposits
Murchison gold deposits

Loglinear Poisson point process model

Predicted intensity of gold deposits
(number of deposits per km$^2$)
Murchison gold deposits

Fitted coefficients:

\[
\beta_0 = -4.3 \\
\beta_1 = -0.3
\]
Murchison gold deposits

Fitted coefficients:

\[ \beta_0 = -4.3 \]
\[ \beta_1 = -0.3 \]

Interpretation:
Murchison gold deposits

Fitted coefficients:
\[ \beta_0 = -4.3 \]
\[ \beta_1 = -0.3 \]

Interpretation:
Near a fault, the intensity is \( e^{-4.3} = 0.01 \) deposits per \( \text{km}^2 \).
Murchison gold deposits

Fitted coefficients:

\[ \beta_0 = -4.3 \]
\[ \beta_1 = -0.3 \]

Interpretation:

Near a fault, the intensity is \( e^{-4.3} = 0.01 \) deposits per km\(^2\).

For every 1 km distance away from the nearest fault, the intensity drops by a factor \( e^{-0.3} = 0.77 \).
Murchison gold deposits

Loglinear Poisson point process model

Predicted intensity of gold deposits
(number of deposits per km$^2$)
Murchison gold deposits

Loglinear Poisson point process model

Predicted intensity of gold deposits
(number of deposits per km$^2$)

Assumes intensity is a loglinear function of distance to nearest geological fault.
Validating the model

Logistic regression assumes

\[ \lambda(u) = \exp(\beta_0 + \beta_1 X(u)) \]
Validating the model

Logistic regression assumes

$$\lambda(u) = \exp(\beta_0 + \beta_1 X(u))$$

What if the relationship is not log-linear?

$$\lambda(u) = \rho(X(u))$$
Validating the model

Logistic regression assumes

\[ \lambda(u) = \exp(\beta_0 + \beta_1 X(u)) \]

What if the relationship is not log-linear?

\[ \lambda(u) = \rho(X(u)) \]

How do we assess the evidence for/against a loglinear relationship?
An epiphany

"Hey, wait a minute! This is grass! We’ve been eating grass!"
Diagnostics for point process models
Diagnostics for point process models

Extend Tukey’s idea to diagnostics
Diagnostics for point process models

Extend Tukey’s idea to diagnostics

1. write down a diagnostic for logistic regression
Diagnostics for point process models

Extend Tukey’s idea to diagnostics

1. write down a diagnostic for logistic regression
   (rescale appropriately for pixel size)
Extend Tukey’s idea to diagnostics

1. write down a diagnostic for logistic regression
   (rescale appropriately for pixel size)

2. take very small pixels
Extend Tukey’s idea to diagnostics

1. write down a diagnostic for logistic regression (rescale appropriately for pixel size)

2. take very small pixels

3. interpret as a diagnostic for point processes
Diagnostics for point process models

Extend Tukey's idea to diagnostics

1. write down a diagnostic for logistic regression
   (rescale appropriately for pixel size)

2. take very small pixels

3. interpret as a diagnostic for point processes

Using this bridge, *tools from mainstream statistical science can be carried over to spatial point processes*
Diagnostics for point process models

1. residuals

2. leverage

3. influence

4. partial residual

5. nonparametric smooth
1. Residuals

In linear regression, if $\hat{y}_i$ is the fitted mean for observation $y_i$, the residuals are

$$r_i = y_i - \hat{y}_i$$
1. Residuals

In linear regression, if $\hat{y}_i$ is the fitted mean for observation $y_i$, the residuals are

$$r_i = y_i - \hat{y}_i$$

The residuals should not show a systematic pattern. If they do, this suggests that the relationship between $x$ and $y$ has not been correctly modelled.
Murchison gold deposits: smoothed Pearson residuals
Murchison gold deposits: smoothed Pearson residuals

Diagnostics for Poisson point process models

1. residuals
2. leverage
3. influence
4. partial residual
5. nonparametric smooth
2. Leverage

In linear regression of $y$ against $x$, the fitted values $\hat{y}_i$ can be written

$$\hat{y}_i = \sum_j h_{ij} y_j.$$ 

The diagonal coefficient $h_{ii}$ is the leverage: it measures how strongly the fitted value $\hat{y}_i$ depends on the observed value $y_i$. 
2. Leverage

In linear regression of \( y \) against \( x \), the fitted values \( \hat{y}_i \) can be written

\[
\hat{y}_i = \sum_j h_{ij} y_j.
\]

The diagonal coefficient \( h_{ii} \) is the leverage: it measures how strongly the fitted value \( \hat{y}_i \) depends on the observed value \( y_i \).

Large values of leverage are associated with the observations which, because of their covariate value, have a potentially strong influence on the fitted model.
For a Poisson point process with loglinear intensity

$$\lambda_{\beta}(u) = \exp(\beta^T Z(u))$$

the leverage function is

$$h(u) = \lambda(u)Z(u)\mathcal{I}^{-1}Z(u)^T$$
Leverage for Poisson point process

For a Poisson point process with loglinear intensity

\[ \lambda_\beta(u) = \exp(\beta^T Z(u)) \]

the leverage function is

\[ h(u) = \lambda(u) Z(u) \mathcal{I}^{-1} Z(u)^T \]

Murchison data: leverage

Leverage for fit
Diagnostics for Poisson point process models

1. residuals
2. leverage
3. influence
4. partial residual
5. nonparametric smooth
3. Influence

In a linear model (etc), the influence of the $i$th observation is

$$s_i = \frac{2}{p} \log \frac{L(\hat{\theta})}{L(\hat{\theta}_{(-i)})}$$

where $L$ is the likelihood, $\hat{\theta}$ is the estimate of the parameter $\theta$ using all the data, $\hat{\theta}_{(-i)}$ is the estimate using all the data except the $i$th observation, and $p$ is the number of parameters.
3. Influence

In a linear model (etc), the influence of the $i$th observation is

$$s_i = \frac{2}{p} \log \frac{L(\hat{\theta})}{L(\hat{\theta}_{(-i)})}$$

where $L$ is the likelihood, $\hat{\theta}$ is the estimate of the parameter $\theta$ using all the data, $\hat{\theta}_{(-i)}$ is the estimate using all the data except the $i$th observation, and $p$ is the number of parameters.

Large values of influence are associated with the observations which, because of their atypical response and high leverage, actually had a strong effect on the fitted model.
Influence for loglinear Poisson model

For the loglinear Poisson point process, the influence of data point $x_i$ is

$$m_i = \frac{1}{p} Z(x_i) \mathcal{I}_\hat{\beta}^{-1} Z(x_i)^\top.$$
Influence for loglinear Poisson model

For the loglinear Poisson point process, the influence of data point $x_i$ is

$$m_i = \frac{1}{p} Z(x_i) \mathcal{I}_\hat{\beta}^{-1} Z(x_i)^\top.$$ 

Murchison data: influence

Influence for fit
Murchison data: influence

Influence for fit
Murchison data: influence

Influence for fit

Large circle at left identifies an outlier or anomaly
Diagnostics for Poisson point process models

1. residuals
2. leverage
3. influence
4. partial residual
5. nonparametric smooth
4. Partial residuals

In linear regression

\[ y = ax + b \]

the partial residual (aka component-plus-residual) is

\[ r_i = \hat{b} x_i + \frac{y_i - \hat{y}_i}{\hat{\sigma}^2}. \]

A smoothed plot of \( r_i \) against \( x_i \) gives an estimate of the true relationship between \( x \) and \( y \).
Partial residuals for spatial point process model

For loglinear Poisson point process, the smoothed partial residuals are

\[ \hat{h}(z) = \hat{\beta} z + \sum_i \frac{\hat{\lambda}(x_i)^{-1} k(Z(x_i) - z)}{\int_W k(Z(u) - z) \, du} - 1 \]

where \( k \) is a smoothing kernel (on the space of values of \( Z \)).
Partial residuals for spatial point process model

For loglinear Poisson point process, the smoothed partial residuals are

\[ \hat{h}(z) = \hat{\beta} z + \sum_i \hat{\lambda}(x_i)^{-1} k(Z(x_i) - z) \int_W k(Z(u) - z) \, du - 1 \]

where \( k \) is a smoothing kernel (on the space of values of \( Z \)).

Murchison data: partial residual plot
Diagnostics for Poisson point process models

1. residuals
2. leverage
3. influence
4. partial residual
5. nonparametric smooth
5. Nonparametric estimate of covariate effect

Suppose that, instead of the loglinear model, the point process intensity depends on covariate \( Z \) through

\[
\lambda(u) = \rho(Z(u))
\]

where the function \( \rho \) is to be estimated.
5. Nonparametric estimate of covariate effect

Suppose that, instead of the loglinear model, the point process intensity depends on covariate $Z$ through

$$\lambda(u) = \rho(Z(u))$$

where the function $\rho$ is to be estimated.

A nonparametric ‘relative density’ estimate of $\rho$ is

$$\hat{\rho}(z) = \frac{\sum_i k(Z(x_i) - z)}{\int_W k(Z(u) - z) \, du}$$
Suppose that, instead of the loglinear model, the point process intensity depends on covariate $Z$ through

$$\lambda(u) = \rho(Z(u))$$

where the function $\rho$ is to be estimated.

A nonparametric ‘relative density’ estimate of $\rho$ is

$$\hat{\rho}(z) = \frac{\sum_i k(Z(x_i) - z)}{\int_W k(Z(u) - z) \, du}$$

Murchison data: smoothed effect estimate
Tree deaths in Perth’s groundwater catchment

+ tree death
• water bore
Spatially varying death rate

Tree deaths per km$^2$
Density of live trees

Compiled from 300,000 tree locations (detected from aerial imagery)

Y.M. Chang
Spatially varying death risk

Deaths per thousand trees
Cumulative residual deaths vs distance from bore

Wallace et al
Monitoring native vegetation on an urban groundwater supply mound using airborne digital imagery

Chang et al
Spatial statistical analysis of tree deaths using airborne digital imagery
Covariate data

Terrain elevation

Depth to water table
Effect of depth to water table

Nonparametric estimate
Effect of terrain elevation, groundwater recharge

Partial residuals

Shapley galaxy concentration
Locally fitted cluster model
Technical outcomes
Technical outcomes

- New tools for spatial point patterns
Technical outcomes

- new tools for spatial point patterns
  - leverage & influence
  - partial residuals
  - nonparametric estimates of covariate effect
  - local fitting
Technical outcomes

- New tools for spatial point patterns
  - leverage & influence
  - partial residuals
  - nonparametric estimates of covariate effect
  - local fitting

- New fast approximations for point process properties
Technical outcomes

- new tools for spatial point patterns
  - leverage & influence
  - partial residuals
  - nonparametric estimates of covariate effect
  - local fitting

- new fast approximations for point process properties

- proof-of-concept software & demonstrations on real data
The new diagnostics apply to Poisson and non-Poisson point processes fitted by $M$-estimators, where the surrogate likelihood is of the same functional form

$$\log L(\theta) = \sum_i \log s_\theta(x_i) - \int_W s_\theta(u) \, du$$

where $s_\theta(\cdot)$ is known analytically.
The new diagnostics apply to Poisson and non-Poisson point processes fitted by $M$-estimators, where the surrogate likelihood is of the same functional form

$$
\log L(\theta) = \sum_i \log s_\theta(x_i) - \int_W s_\theta(u) \, du
$$

where $s_\theta(\cdot)$ is known analytically.

- Poisson point process, maximum likelihood
The new diagnostics apply to Poisson and non-Poisson point processes fitted by $M$-estimators, where the surrogate likelihood is of the same functional form

$$\log L(\theta) = \sum_i \log s_\theta(x_i) - \int_W s_\theta(u) \, du$$

where $s_\theta(\cdot)$ is known analytically.

- Poisson point process, maximum likelihood
- Poisson point process, robust $M$-estimator
The new diagnostics apply to Poisson and non-Poisson point processes fitted by $M$-estimators, where the surrogate likelihood is of the same functional form

$$\log L(\theta) = \sum_i \log s_\theta(x_i) - \int_W s_\theta(u) \, du$$

where $s_\theta(\cdot)$ is known analytically.