Residuals and diagnostics for spatial point processes

Adrian Baddeley

Joint work with Martin Hazelton (UWA), Rolf Turner (UNB), Jesper Møller (Aalborg)
1. Modelling point patterns
Swedish pines data: locations of 71 pine saplings in a $10 \times 10$ metre square.

Strand (1975)
Ripley (1981)
Seedlings and saplings of Japanese black pine in a $10 \times 10$ m square.

Numata (1964)
Ants’ nests of two species:
- Messor wasmanni
- Cataglyphis bicolor

Copper deposits (○) and lineaments (—) in a geological survey.

Berman & Turner (1986)
Realistic models may include:

- spatial inhomogeneity
Realistic models may include:

- spatial inhomogeneity
- interpoint interaction
Realistic models may include:

- spatial inhomogeneity
- interpoint interaction
- covariate effects
Practical model-fitting for spatial point patterns
Likelihoods are usually intractable.

- **Method of moments**
  - minimum contrast
    Bartlett (60’s), Ripley (77–88), Diggle (78+)
  - Takacs-Fiksel method
    Takacs/Fiksel (84)

- **Exact MLE for special models**
  - space-time models
    Brillinger (69+), Vere-Jones (75+), Daley/Vere-Jones/Ogata (80’s+), Diggle (01+)
  - log-Gaussian Cox processes
    Møller/Syversveen/Waagepetersen (01)

- **Computational strategies for MLE**
  - analytic approximation
    Ogata/Tanemura (84+)
  - correspondence with GLM
    [Brillinger], Berman/Turner (92), Lindsey (91+)
- Markov Chain Monte Carlo  
  Ripley (77), Penttinen (83+), Geyer/Møller (94+)
- stochastic approximation  
  Moyeed (91)

• Maximum pseudolikelihood
  - explicit  
    Besag (75+), Ripley (76+), Diggle/Stoyan/Ogata (94)
  - correspondence with GLM  
    Clyde/Strauss (91), Lawson (93), Baddeley/Turner (00)
  - pseudolikelihood + one Fisher scoring step  
    Ogata/Huang (01)

• Bayesian inference via MCMC
  Besag/Green (93), Green (95)
Main modelling tool:

conditional intensity $\lambda(u, X)$

of point process $X$ at location $u$
Papangelou conditional intensity

\[ \lambda(u, X) \, du = \mathbb{E}[\, dN(u) \mid X \setminus \{u\}] \]

conditional probability of getting a point of the process at location \( u \in \mathbb{R}^d \), given the rest of the process \( X \setminus \{u\} \)
Examples:

**homogeneous Poisson**, intensity $\lambda > 0$:

$$\lambda(u, x) \equiv \lambda$$

**inhomogeneous Poisson**, intensity function $\lambda(u)$:

$$\lambda(u, x) = \lambda(u)$$

**Strauss process**, parameters $\beta, \gamma, r$:

$$\lambda(u, x) = \beta \gamma^{t(u, x)}$$

where $t(u, x) = \# \{ i : \| x_i - u \| \leq r \}$
Baddeley & Turner (2000):
generic fitting algorithm for all models with loglinear conditional intensity

\[ \lambda(u, x) = \exp\{\theta H(u) + \eta G(u, x)\} \]

by maximum pseudolikelihood (transforming the pseudolikelihood into the likelihood of a GLM).

<table>
<thead>
<tr>
<th>Term</th>
<th>Point process interpretation</th>
<th>GLM analogue</th>
</tr>
</thead>
<tbody>
<tr>
<td>(H(u))</td>
<td>trend/covariate effects</td>
<td>(\leftrightarrow) linear predictor</td>
</tr>
<tr>
<td>(G(u, x))</td>
<td>interpoint interaction</td>
<td>(\leftrightarrow) error distribution</td>
</tr>
</tbody>
</table>
R interface analogous to glm:

```r
fit <- mpl(x, ~H, G)
```

Examples:

```r
mpl(X, ~1, Strauss(r=1))  # stationary Strauss process

mpl(X, ~ x + y, Poisson()) # inhomogeneous Poisson process with intensity
\[ \lambda(x, y) = \exp\{\theta_0 + \theta_1 x + \theta_2 y\} \]
Japanese black pines.
Inhomogeneous Poisson model with log-quadratic intensity.
Fitted intensity function.
Japanese black pines.
Inhomogeneous Strauss process model with log-quadratic trend.
Fitted conditional intensity.
Problem:
How to assess the fit of a point process model fitted to spatial data?
Classical exploratory methods in spatial statistics compare the data with the uniform Poisson process — which is usually the \textbf{null} model in a modelling context.
2. Residuals
In linear regression, generalised linear modelling, etc., plots of the residuals are used to

- identify outliers
- detect systematic departures from model
- detect lurking variables
- validate distributional assumptions
What are the “residuals” for a fitted spatial point process model, and how do we use them?
Aside: for point processes on the line, with

\[ N_t = \text{number of arrivals up to time } t \]

and directed conditional intensity

\[ \lambda_t = \mathbb{E}[dN_t \mid N_s, s < t] \]

the residual process

\[ R_t = N_t - \int_0^t \lambda_s \, ds \]

is a zero mean martingale, so that for any predictable \( h_t \)

\[ \mathbb{E} \left[ \int_0^t h_s \, dR_s \right] = \mathbb{E} \left[ \sum_i h_{ti} - \int_0^t h_s \lambda_s \, ds \right] = 0 \]
For spatial point processes, define the residual process
\[dR(u) = dN(u) - \lambda(u, X) \, du\]

Integral over a set \(B \subset \mathbb{R}^d\)
\[R(B) = N(B) - \int_B \lambda(u, X) \, du\]
satisfies
\[\mathbb{E}[R(B)] = 0.\]
For a spatial point process $X$ the “residual random measure”

$$R(B) = N(B) - \int_B \lambda(u, X) \, du$$

is a zero-mean set-indexed martingale, and

$$\mathbb{E} \int_B h(u, X \setminus u) \, dR(u) = 0$$

for all integrable $h$. 
**Nguyen-Zessin identity**

\[
\mathbb{E} \left[ \sum_{x_i \in X} h(x_i, X \setminus x_i) \right] = \int \mathbb{E}[h(u, X) \lambda(u, X)] \, du
\]

for all nonnegative integrable functions \( h \).

X. Nguyen & H. Zessin, Punktprozesse mit Wechselwirkung.
Taking $h(u, X) = 1\{u \in B\}$ yields

$$\mathbb{E}[N(B)] = \mathbb{E} \int_B \lambda(u, X) \, du$$
3. Stoyan-Grabarnik weights
D. Stoyan & P. Grabarnik (1991)
Diagnostic for fitted point process models.
Taking
\[ h(u, X) = \frac{1\{u \in B\}}{\lambda(u, X)} \]
in Nguyen-Zessin yields
\[
\mathbb{E} \left[ \sum_{x_i \in B} \frac{1}{\lambda(x_i, X)} \right] = |B|
\]
Stoyan & Grabarnik (1991) proposed:

- fit model to data
- compute conditional intensity of fitted model
  \[ \hat{\lambda}(u, x) = \lambda_{\hat{\theta}}(u, x) \]
- to each data point \( x_i \), attach “weight”
  \[ m_i = \frac{1}{\hat{\lambda}(x_i, x)} \]

“exponential energy weights”
Stoyan & Grabarnik (1991) suggested:

(a) extreme values of individual marks may indicate ‘outliers’;

(b) groups of extreme values may indicate regions of irregularity;

(c) global departures between the left and right sides of equation may be used to test goodness-of-fit or to test convergence of MCMC samplers.

(Did not present any practical examples...)

4. Diagnostic plots for spatial trend
Example 1:

An inhomogeneous Poisson process with intensity

$$\lambda(x, y) = 300 \exp\{-3x\}$$

in the unit square.
Mark plot:
circles centred at the points of the data pattern with radii proportional to the Stoyan-Grabarnik exponential energy weights.
Mark plot of exponential energy weights

Correct model  Incorrect model (uniform Poisson)
Individual weights $m_i$ are rather difficult to interpret. The Nguyen–Zessin identity relates to sums of weights, i.e. to the residual measure.
Appropriate tools are:
(1) the mark sum measure $M$ defined by

$$M(B) = \sum_{x_i \in B} m_i$$

which satisfies $\mathbb{E}[M(B)] = |B|$, or

(2) a kernel-smoothed version

$$s(u) = e(u) \sum_{x_i} m_i \ k(x_i - u)$$

where $k$ is a smoothing kernel and $e(u)$ is an edge correction.

Satisfies $\mathbb{E}[s(u)] = 1$. 
Kernel-smoothed exponential energy weight measure
Kernel-smoothed exponential energy weights

Correct model

Incorrect model
Kernel-smoothed exponential energy weights

Correct model  Incorrect model
Kernel-smoothed exponential energy weights

Correct model

Incorrect model
Lurking variable plot:
Suppose $Z(u)$ is a spatial covariate observable at any location $u$. Write $V(t) = \{ u \in W : Z(u) \leq t \}$. Compute
\[
g(t) = \sum_{x_i \in V(t)} m_i
\]
compare with
\[
g_0(t) = |V(t)|
\]
If the model is correct, $\mathbb{E}[g(t)] = g_0(t)$. 

Lurking variable plot for $x$ coordinate
Lurking variable plots

Correct model

Incorrect model
Variances: For any Gibbs process $X$

$$\text{var}\left[\sum_{x_i \in B} \frac{1}{\lambda(x_i, X)}\right] = \int_B \int_B \mathbb{E}\left[\frac{\lambda(u, v; X)}{\lambda(u, X \cup \{v\}) \lambda(v, X \cup \{u\})}\right] \, du \, dv$$

$$+ \int_B \mathbb{E}\left[\frac{1}{\lambda(u, X)}\right] \, du - |B|^2$$

where $\lambda(u, v; X)$ is the two-point conditional intensity.
For the nonuniform Poisson process with intensity function \( \lambda(\cdot) \)

\[
\text{var} \left[ \sum_{x_i \in B} \frac{1}{\lambda(x_i, X)} \right] = \int_B \frac{1}{\lambda(u)} \, du.
\]
Standard presentation of diagnostic plots based on exponential energy weights.
Alternative (using derivative of lurking variable plots)
Diagnostics for (incorrect) uniform Poisson model
Stoyan & Grabarnik (1991) suggested:

(a) extreme values of individual marks may indicate ‘outliers’; — no

(b) groups of extreme values may indicate regions of irregularity; — yes

(c) global departures between the left and right sides of equation may be used to test goodness-of-fit or to test convergence of MCMC samplers. — no
Counterexample: if the fitted model is uniform Poisson, then $\hat{\lambda} = n/|W|$ so

$$m_i = \frac{|W|}{n}$$

for all $i$, and

$$\sum_i m_i = |W|$$

exactly, for any nonempty point pattern!
Example 2:
Strauss process, $\gamma = 0.1$
(strong interpoint interaction)
Diagnostic plots (exponential energy weights) for fitted model of correct form.
Diagnostic plots (exponential energy weights) for true model.
Diagnostic plots (exponential energy weights) for incorrect model (uniform Poisson).
Problems with these plots:

(1) high variability
(2) insensitivity to interpoint interaction
High variability: analogous to variance inflation in Horvitz-Thompson estimator.

For any pairwise interaction process,

$$\text{var} \left[ \sum_{x_i \in B} \frac{1}{\lambda(x_i, X)} \right] = \int_B \int_B \frac{1}{c(u, v)} \, du \, dv$$

$$+ \int_B \mathbb{E} \left[ \frac{1}{\lambda(u, X)} \right] \, du - |B|^2$$

where $c$ is pair potential. For Strauss process

$$c(u, v) = \gamma^{1\{||u-v|| \leq r\}}$$

Variance is large when $\gamma$ is small.
uniform Poisson
Fitted model:

Data: Streuss

smoothed

exponential energy weights

cumulative exponential energy weights
If the fitted model is uniform Poisson with intensity $\lambda$, then

$$\hat{\lambda} = \frac{n}{|W|}$$

so $m_i = |W|/n$, and $\mathbb{E}[s(u)] \equiv 1$ for any stationary point process.
5. General residuals
Stoyan & Grabarnik assign diagnostic weights
$m_i = 1/\lambda(x_i, \mathbf{x})$ to the data points $x_i$ only.

This is possible by a felicitous choice of integrand $h$ in the
N–Z identity

$$
\mathbb{E} \left[ \sum_{x_i \in X} h(x_i, X \setminus x_i) \right] = \int \mathbb{E}[h(u, X)\lambda(u, X)] \, du
$$

Taking $h(u, X) = 1/\lambda(u, X)$ yields

$$
\mathbb{E} \left[ \sum_{x_i \in B} \frac{1}{\lambda(x_i, X)} \right] = |B|
$$
For other choices of $h$, we may interpret N–Z identity as defining a \textit{residual measure}, with mass

$$r(x_i) = h(x_i, x \setminus x_i) \quad \text{for each data point}$$

and density

$$r(u) \, du = -h(u, x) \lambda(u, x) \, du$$

over all (non-data) points $u$. 
Example 1:
$h \equiv 1$ yields *raw residual measure*

\[
\begin{align*}
    r(x_i) &= 1 \\
    r(u) \, du &= -\lambda(u, x) \, du
\end{align*}
\]
Example 2:
\[ h(u, \mathbf{x}) = \frac{1}{\sqrt{\lambda(u, \mathbf{x})}} \] yields *Pearson residual measure*

\[
\begin{align*}
    r(x_i) & = \frac{1}{\sqrt{\lambda(x_i, \mathbf{x})}} \\
    r(u) \, du & = -\sqrt{\lambda(u, \mathbf{x})} \, du
\end{align*}
\]

Pearson residual measure has unit variance,

\[
\text{var} R(B) = |B|.
\]
Pearson residuals

all residuals

negative residuals
Pearson residuals.
Data: Strauss.
Fitted model: Strauss.
Pearson residuals.
Data: Strauss.
True Strauss model.
Pearson residuals.

Data: Strauss.

Fitted model: uniform Poisson.
6. Q–Q plots
**Problem:** if the fitted model is uniform Poisson (say), our residual plots are insensitive to the presence of interpoint interaction.
uniform Poisson

Fitted model

Data: Strauss

smoothed

exponential energy weights

cumulative exponential energy weights

cumulative exponential energy weights
This could have been expected from the analogy:

<table>
<thead>
<tr>
<th>Term</th>
<th>Point process interpretation</th>
<th>GLM analogue</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H(u)$</td>
<td>trend/covariate effects</td>
<td>$\leftrightarrow$ linear predictor</td>
</tr>
<tr>
<td>$G(u, x)$</td>
<td>interpoint interaction</td>
<td>$\leftrightarrow$ error distribution</td>
</tr>
</tbody>
</table>

Plots of residuals against covariates are good for assessing trend, but not for validating the assumptions about the error distribution.

For this we should look at the distribution of residuals.
The kernel-smoothed residual field $s(u)$ does carry information about interpoint interaction. Spatial fluctuations in $s$ are more extreme when the pattern is more clustered.
Special case:

Fitted model is uniform Poisson, kernel $k$ is uniform on the disc $b(0, r)$. Then

$$s(u) = c.N(b(u, r))$$

Note

$$S = \max_{u \in W} N(b(u, r))$$

is known as the scan statistic and is able to discriminate between different types of interpoint interaction.
Q–Q plot of residuals

Compute $s(u)$ on a grid of pixel locations $u_j$ in $W$.
Plot the order statistics of $s(u_j)$ against the expected order statistics of $s(u)$ for the fitted model (estimated by simulation).
Q–Q plot of kernel smoothed field $s$.
Data: uniform Poisson.
Fitted model: uniform Poisson
Q-Q plot of kernel smoothed field $s$.
Data: Strauss.
Fitted model: uniform Poisson
Conclusions

• Residuals for a spatial point pattern are defined at all locations, not just at the data points.

• Useful analogy with residual plots in classical settings.

• Departures from systematic (trend) part of model assessed by plotting kernel-smoothed residuals against covariate.

• Departures from random (interpoint interaction) part of model assessed by making Q–Q plot of smoothed residuals.

• Some existing *ad hoc* techniques are recovered.
In progress

• adjust for effect of estimating $\theta$

• use two-point conditional intensity to define second-order residuals

• residuals from associated GLM
References

Stoyan, D. and Grabarnik, P.

Baddeley, A. and Turner, R.

Software package *spatstat* and forthcoming papers