Bayesian model selection for point process cluster models

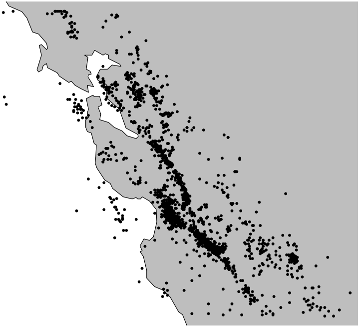
Thordis L. Thorarinsdottir

Norwegian Computing Center
www.nr.no/~thordis

Selbu, October 20 2012

Joint with Peter Guttorp

Model selection is performed for various reasons



- Which model is most appropriate for our modelling purposes/for our data?
- Which model requires least computational effort?
- Which model ist most flexible?

A brief introdution to point processes

A space or time point process is a **random collection of events** $X = \{x_i : i = 1, \ldots, n\}$, where x denotes either location within a spatial region D or time within a time-interval $[0, T]$.

The **intensity measure** of the process is given by

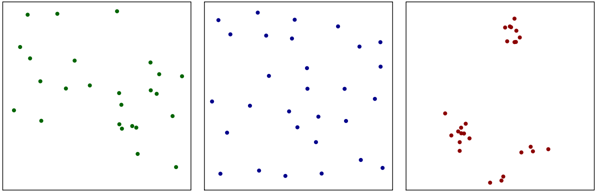
$$\Lambda(A) = \mathbb{E}\{\# \text{ of points in } A\}.$$

When modelling point processes, we are often interested in the **intensity function**, $\lambda(s)$,

$$\Lambda(A) = \int_A \lambda(s) ds$$

and we will here assume that such $\lambda(\cdot)$ exists.

Current model selection methods have two goals



1. To describe the features of the point pattern to distinguish between
   - **complete randomness** (Poisson process)
   - **repulsion**
   - **clustering**
2. To check the compatibility of a suggested/fitted model with the observed data

Goodness-of-fit tests are based on summary statistics

The most classical summary statistic is Ripley's K-function. It is a function of the **pair correlation function**

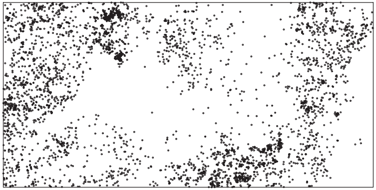
$$g(u,v) = \frac{\lambda^2(u,v)}{\lambda(u)\lambda(v)},$$

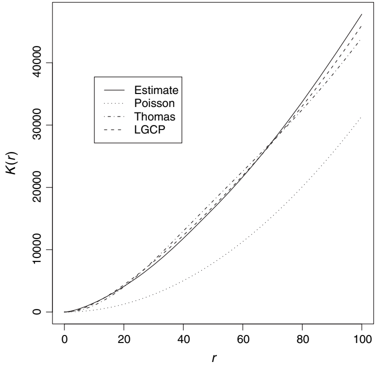
where λ is the intensity function and λ^2 is the second order product density for X. If $g(u,v) = g(u-v)$, the **K-function** is defined by

$$K(r) = \int_{b(0,r)} g(u)du, \quad r > 0.$$

Empirical estimates of the K-function are then compared to theoretical values for suggested models.

Goodness-of-fit for rain forest data based on the K-function





Møller and Waagepetersen (2007)

Type of pattern can be learned by analyzing interpoint distances

- The **empty space function** F is the cdf of the distance from an arbitrary location to the nearest point in X,

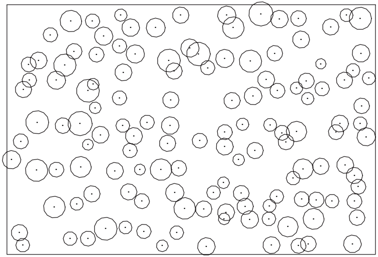
$$F(r) = \mathbb{P}(X \cap b(0,r) \neq \emptyset), \quad r > 0.$$

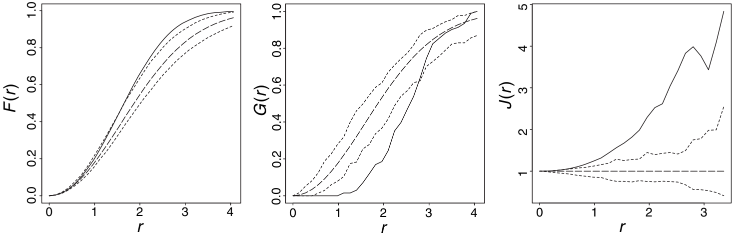
- The **nearest-neighbour function** $G(r)$ is the cdf of the distance between a 'typical' point in X and its nearest neighbour in X.
- The **J-function** incorporates both and is given by

$$J(r) = \frac{1 - G(r)}{1 - F(r)}, \text{ for } F(r) < 1$$

Estimates of these functions are then compared to the theoretial values for a Poisson process to detect repulsion or clustering.

Interpoint distances for Norwegian spruce data

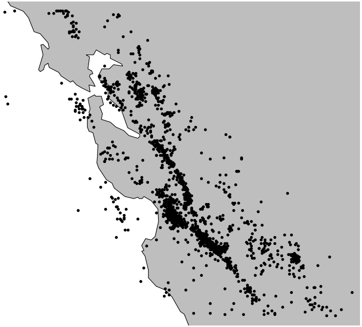




Møller and Waagepetersen (2007)

Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

We want to take this a step further



Can we detect the type of clustering mechanism underlying a point pattern based on formal tests?

Cluster point processes

We consider and **independent cluster process** $x = \cup\{\Phi_i + \tau_i\}$, where $\Psi = \{\tau_i\}$ is the **parent process** and Φ_i are **secondary offspring processes** that are inhomogeneous Poisson.

Conditional density of x w.r.t $x_1 \sim \mathrm{Po}(\zeta)$ on $B \subset \mathbb{R}^2$

$$p(x \mid \Psi, \alpha, \theta) = \exp\left(\zeta|B| - \int_B Z(\xi \mid \Psi, \alpha, \theta) d\xi\right) \prod_{\xi \in x} Z(\xi \mid \Psi, \alpha, \theta),$$

where $|\cdot|$ denotes area and

$$Z(\xi \mid \Psi, \alpha, \theta) = \sum_{\tau_i \in \Psi} \alpha_i k(\xi - \tau_i \mid \theta_i)$$

is the random intensity function for some kernel function k.

Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Cluster process models consist of three parts

- model for Ψ
  - random (Poisson)
  - repulsive (Strauss, Matérn)
  - clustered (multi-level)
  - homogeneous/inhomogeneous
- model for the cluster sizes
  - Poisson (distinct α_i, $\alpha_i = \alpha$)
- dispersion density k
  - normal (isotropic or not)
  - Cauchy
  - uniform on a disc
  - a mixture

We propose to perform the model selection within a Bayesian inference framework.

Our tool: Bayes factors

The **marginal likelihood** of the observed data x under model M is given by

$$m(x|M) = \int p(x|\theta, M)p(\theta|M)d\theta,$$

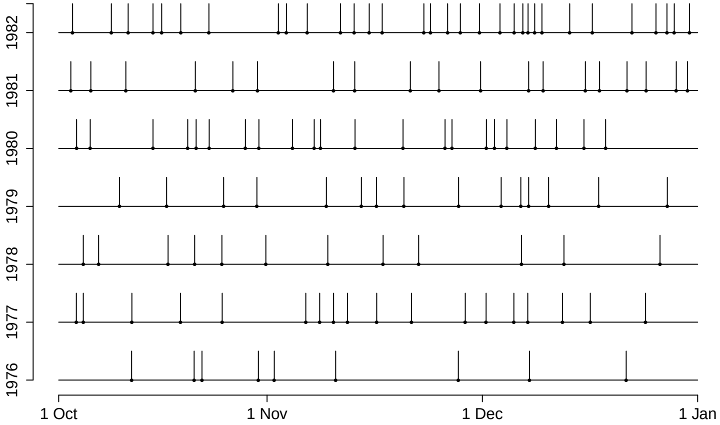
where $p(\cdot|\theta, M)$ is the likelihood and $p(\cdot|M)$ is the prior density for the parameter vector θ.

Two models, M_0 and M_1, may then be compared by calculating the **Bayes factor**

$$\mathrm{B}_{01} = \frac{m(x|M_0)}{m(x|M_1)}.$$

Problem: $m(x|M)$ is usually intractable.

A very simple example: precipitation events at Whiteface Mountain, NY



We consider two competing models

Poisson process with intensity λ on $[0, T]$:

$$p(\mathbf{x}|\lambda, M_{\text{Po}}) = \lambda^n \exp\left(T(1-\lambda)\right).$$

Matérn Type III point process with thinning parameter R:

$$p(\mathbf{x}|\lambda, R, M_{\text{Ma}}) = \mathbb{1}\{\rho(\mathbf{x}) > R\}\lambda^n \exp\left(T + \lambda(nR - T)\right),$$

where $\rho(\mathbf{x})$ is the minimum interpoint distance in $\mathbf{x}$.
(Huber and Wolpert, 2009)

There is a strong evidence for repulsion

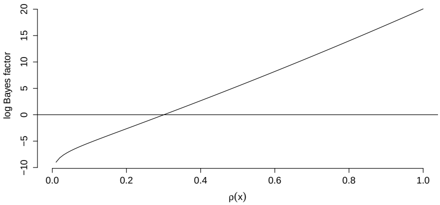
We assign prior distributions

$$p(\lambda) = 2e^{-2\lambda}, \quad p(R) = 1/T.$$

The resulting Bayes factor equals

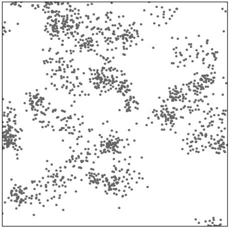
$$B_{\mathrm{Ma,Po}} = \frac{(7T+2)}{Tn^2}\left(\left(\frac{7T+2}{7T+2-n\rho(x)}\right)^n - 1\right) = 273618,$$

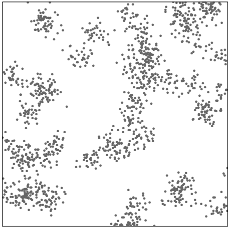
where $\rho(x) = 0.75$, $T = 92$, and $n = 127$.



Example 2: Detecting a mixture of two processes

Can we detect whether we have one cluster process, or a mixture of two such processes?

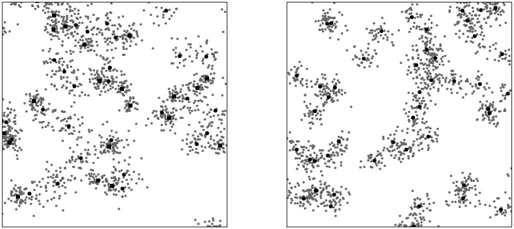




We test this under a modified Thomas model in a simulation study

Modified Thomas process (Diggle, Besag and Gleaves, 1976)

- an (unobserved) homogeneous Poisson cluster centre process
- Poisson cluster sizes
- a normal isotropic dispersion process



We are interested in comparing two models on $B = [0, 1]^2$

M_0: modified Thomas process (with true dispersion sd $\omega = 0.03$)

M_1: mixture of two such processes ($\omega_1 = 0.02, \omega_2 = 0.04$)

Model and inference

The random intensity function under M_1 is given by

$$Z(\xi|\Psi,\alpha,\omega) = \alpha\Big[\frac{1}{2\pi\omega_1^2}\sum_{c\in\Psi_1}\exp\Big(-\frac{\|c-\xi\|^2}{2\omega_1^2}\Big)+\frac{1}{2\pi\omega_2^2}\sum_{c\in\Psi_2}\exp\Big(-\frac{\|c-\xi\|^2}{2\omega_2^2}\Big)\Big],$$

with $\omega_1 < \omega_2$ for identifiability. The joint posterior distribution is

$$p(\psi,\kappa,\alpha,\omega|x) \propto p(x|Z(\cdot|\psi,\alpha,\omega))p(\psi_1|\kappa_1)p(\psi_2|\kappa_2)p(\kappa)p(\alpha)p(\omega),$$

and the MCMC simulation algorithm consists of

(a) updating the latent process ψ (via birth-death-move alg.);

(b) updating the parameters κ, α, ω (via MH or Gibbs sampling);

(c) proposing to jump between M_0 and M_1 (similar to Richardson and Green (1997)).

Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Reversible jump algorithm

We merge by

$$\psi' = \psi_1 \cup \psi_2, \quad \kappa' = \kappa_1 + \kappa_2, \quad \omega' = \sqrt{\frac{\kappa_1 \omega_1^2 + \kappa_2 \omega_2^2}{\kappa_1 + \kappa_2}}.$$

The split move has two degrees of freedom, $u_1, u_2 \sim \text{Beta}(2,2)$,

$$\kappa' = (\kappa_1', \kappa_2') = (u_1 \kappa, (1 - u_1)\kappa),$$

$$\omega' = (\omega_1', \omega_2') = \left(\sqrt{\frac{u_2}{u_1}} \omega, \sqrt{\frac{1 - u_2}{1 - u_1}} \omega \right).$$

Each point in ψ belongs to ψ_1' with probability κ_1'/κ or to ψ_2' with probability κ_2'/κ.

We need to balance the proposals

The density of the latent parent process ψ w.r.t. $y \sim \text{Po}(\zeta)$ is

$$p(\psi|\kappa) = \exp(|B|(\zeta - \kappa))\kappa^{n(\psi)}$$

and usually, the choice of ζ is irrelevant. Here, we set

$$\zeta = n(\psi)/|B|, \quad \zeta_1 = n(\psi_1)/|B|, \quad \zeta_2 = n(\psi_2)/|B|.$$

This gives

$$\log\Big(\frac{p(\psi_1'|\kappa_1')p(\psi_2'|\kappa_2')}{p(\psi|\kappa)}\Big)$$
$$= n(\psi_1')\Big[\log\frac{\kappa_1'}{\kappa} - \log\frac{n(\psi_1')}{n(\psi)}\Big] + n(\psi_2')\Big[\log\frac{\kappa_2'}{\kappa} - \log\frac{n(\psi_2')}{n(\psi)}\Big],$$

which penalizes for a lack of balance between intensites and point patterns.

Bayesian model selection outperforms AIC

We compare our method to model selection with AIC,

$$\text{AIC} = -2 \log L + 2k,$$

where L is the ML value and k is the number of parameters, based on maximum Palm likelihood estimation (Tanaka *et al.*, 2008).

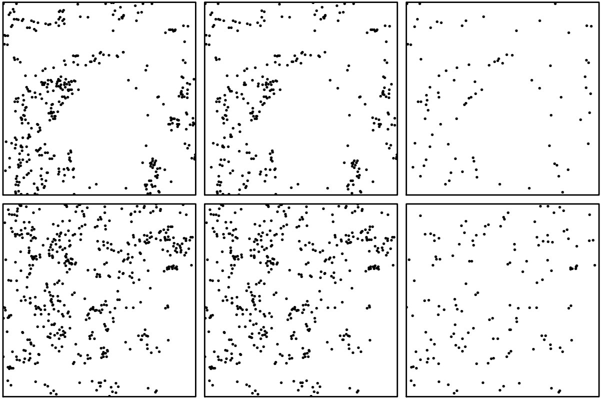
	AIC		BF	
	M_0	M_1	M_0	M_1
True model M_0	8	2	10	0
True model M_1	3	7	0	10

Classification results for 10 repetitions under each model

Example 2: Detecting second order structure in the parent process

- Can we detect whether the parent process is completely random (Poisson), or whether the structure is repulsive?
- We want to answer this question for a young Pacific silver fir (*Abies amabilis*) forest
- For this, we apply a Strauss model for the centre process with normal dispersion density and Poisson cluster sizes as before

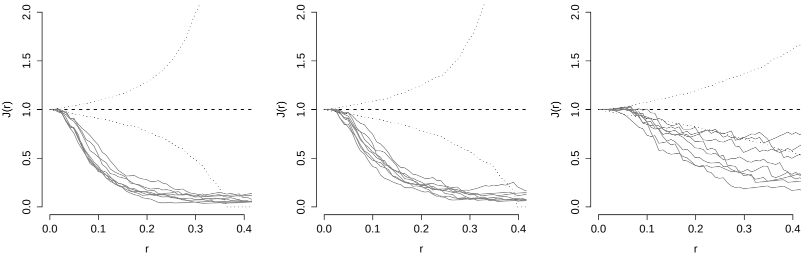
Pacific silver fir data set



The data consists of locations of Pacific silver fir trees at eight 6×6m plots at Findley Lake Reserve, WA. The area was clear-cut in 1957; we have observations from 1978, 1990, and 2009.

Norsk Regnesentral
NORWEGIAN COMPUTING CENTER

Distributions of interpoint distances indicate clustering



(a) 1978 data (b) 1990 data (c) 2009 data

Summary statistics based on the *J*-function

Model for the latent parent process

The **Strauss model** (Strauss, 1975) has density

$$p(\psi \,|\, \nu, \gamma, R) = \mathcal{Z}(\nu, \gamma, R)\nu^{n(\psi)}\gamma^{s(\psi,R)} \propto p^*(\psi \,|\, \nu, \gamma, R)$$

where $\mathcal{Z}$ is an intractable normalizing constant and

$$s(\psi, R) = \sum_{\tau_i, \tau_{i'} \in \psi} \mathbb{1}\{\|\tau_i - \tau_{i'}\| \leq R\}.$$

- for $\gamma = 1$, this is the Poisson model
- for $0 < \gamma < 1$, cluster centres closer than R are discouraged
- for $\gamma = 0$, cluster centres closer than R are not allowed

Auxilliary variable MH algorithms make full inference possible

Exchange algorithm by Murray *et al.* (2006)

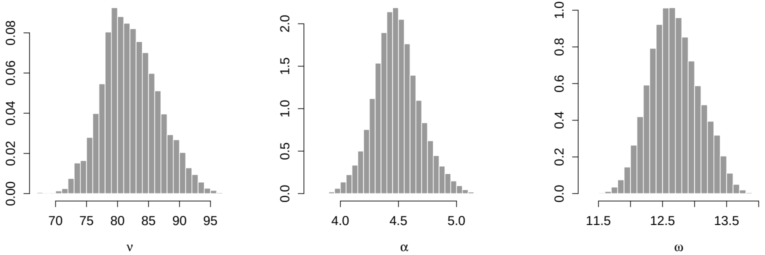
for $\theta \in \{\nu, \gamma, R\}$ do

1. sample θ' from $q(\theta, \cdot)$
2. generate a new latent process η from $p(\cdot \mid \theta')$ (Berthelsen and Møller, 2003; `spatstat`)
3. compute
$$r = \frac{q(\theta', \theta) p(\theta') p^*(\psi \mid \theta')}{q(\theta, \theta') p(\theta) p^*(\psi \mid \theta)} \frac{p^*(\eta \mid \theta)}{p^*(\eta \mid \theta')}$$
4. accept θ' with probability $\min\{1, r\}$

end

(Møller *et al.*, 2006; Murray *et al.*, 2006; Liang and Jin, 2011)

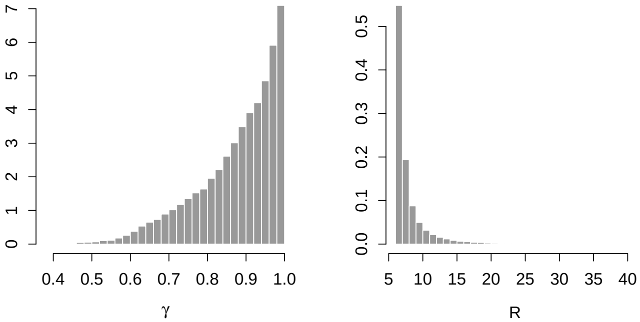
Posterior distributions



(a) Centre intensity (b) Cluster size (c) Dispersion sd (in cm)

Results based on the data from 1978 only

Posterior distributions



(d) Interaction (e) Range (in cm)

Results based on the data from 1978 only

The temporal aspect

No reprocduction was recorded in the tree stand over the observation period.

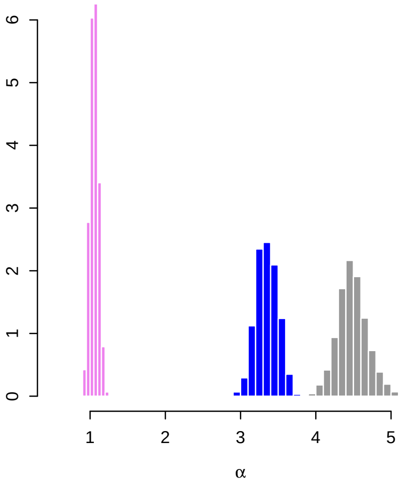
We can thus assume that the cluster centres and the dispersion process don't change over time.

We assign α a conjugate $\Gamma(a, b)$ prior which results in

$$\alpha \mid \{x_j\}, \{\psi_j\}, \omega \sim \Gamma\Big(a + \sum_j n(x_j), b + \sum_j \sum_{\tau_i \in \psi_j} \int_B \frac{1}{2\pi\omega^2} \exp\Big(-\frac{\|\xi - \tau_i\|^2}{2\omega^2}\Big) d\xi\Big).$$

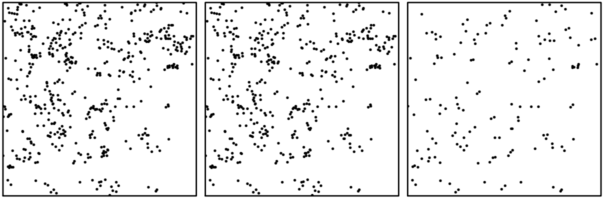
It is easy to obtain new posterior samples for α conditional on the other parameters in the model.

Cluster size development in time



Cluster sizes for observations from **1978**, **1990**, and **2009**

Extensions to the inference procedure



- Formal testing procedure to determine whether $\gamma = 1$
- More rigorous method to deal with the identifiability issues than to simply assume that $R \geq r$ for some small $r > 0$
- Fully spatio-temporal analysis of the data

Conclusions

- Bayesian model selection allows us to compare different models for the clustering mechanism underlying clustered point patterns
- When the marginal likelihood is intractable, the model comparison can be complicated for non-nested models
- Recent advances in MH algorithms have made precise inference possible for a large class of point process models with intractable likelihoods

Selected references

1. Guttorp, P. and Thorarinsdottir, T. L. (2012): Bayesian inference for non-Markovian point processes. E. Porcu, J.M. Montero, and M. Schlather (Eds.), *Advance and Challenges in Space-Time Processes Modelling of Natural Events*, 79-102. Springer.
2. Murray, I., Ghahramani, Z., and MacKay, D.J.C. (2006): MCMC for doubly-intractable distributions. *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence*, 359-366.
3. Møller, J. and Waagepetersen, R. P. (2007): Modern statistics for spatial point processes. *Scandinavian Journal of Statistics*, **34**, 643-684.
4. Richardson, S. and Green, P.J. (1997): On Bayesian analysis of mixtures with an unkown number of components. *Journal of the Royal Statistical Society, Series B*, **59**, 731-792.
5. Tanaka, U., Ogata, Y., and Stoyan, D. (2008): Parameter estimation and model selection for Neyman-Scott point processes. *Biometrical Journal*, **49**, 1-15.