Spatial analysis of tropical rain forest plot data

Rasmus Waagepetersen
Department of Mathematical Sciences
Aalborg University

December 11, 2010

Tropical rain forest ecology

Fundamental questions: which factors influence the spatial distribution of rain forest trees and what is the reason for the high biodiversity of rain forests ?

Key factors:

- environment: topography, soil composition,...
- seed dispersal limitation: by wind, birds or mammals...

Tropical rain forest ecology

Fundamental questions: which factors influence the spatial distribution of rain forest trees and what is the reason for the high biodiversity of rain forests ?

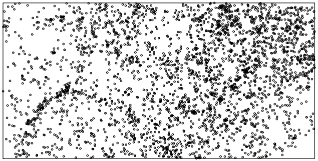
Key factors:

- environment: topography, soil composition,...
- seed dispersal limitation: by wind, birds or mammals...

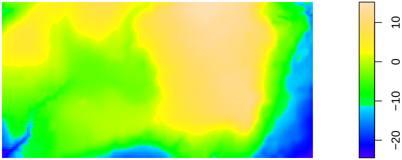
Outline:

- data examples
- introduction to spatial point processes
- applications to tropical rain forest data

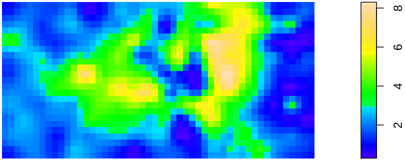
Example: *Capparis Frondosa* and environment



- *observation window W* $= 1000$ m $\times$ 500 m
- seed dispersal$\Rightarrow$ *clustering*
- environment $\Rightarrow$ *inhomogeneity*



Elevation



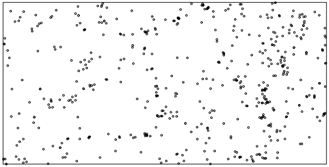
Potassium content in soil.

Quantify dependence on environmental variables and seed dispersal using statistics for spatial point processes.

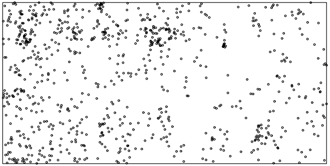
Example: modes of seed dispersal and clustering

Three species with different modes of seed dispersal:

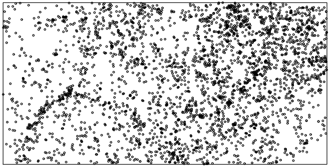
Acalypha Diversifolia explosive capsules



Loncocharpus Heptaphyllus wind



Capparis Frondosa bird/mammal

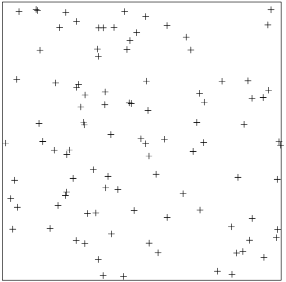


Is degree of clustering related to mode of seed dispersal ?

Spatial point process

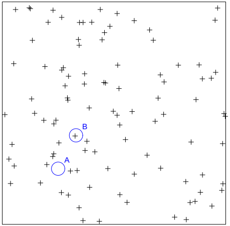
Spatial point process: random collection of points

(finite number of points in bounded sets)



Intensity function and product density

X: spatial point process. A and B small subregions.

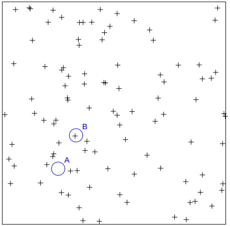


Intensity function and product density

$\mathbf{X}$: spatial point process. A and B small subregions.

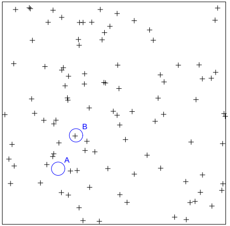
Intensity function of point process $\mathbf{X}$

$\rho(u)|A| \approx P(\mathbf{X} \text{ has a point in A}), \quad u \in A$



Intensity function and product density

X: spatial point process. A and B small subregions.



Intensity function of point process **X**

$\rho(u)|A| \approx P(\mathbf{X}$ has a point in A), $\quad u \in A$

Second order product density

$\rho^{(2)}(u, v)|A||B| \approx P(\mathbf{X}$ has a point in each of A and $B) \quad u \in A,\ v \in B$

Pair correlation and K-function

Pair correlation function

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

NB: independent points $\Rightarrow \rho^{(2)}(u,v) = \rho(u)\rho(v) \Rightarrow g(u,v) = 1$

Pair correlation and K-function

Pair correlation function

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

NB: independent points $\Rightarrow \rho^{(2)}(u,v) = \rho(u)\rho(v) \Rightarrow g(u,v) = 1$

K-function

$$K(t) = \int_{\|h\| \le t} g(h)\mathrm{d}h$$

(provided $g(u,v) = g(u-v)$ i.e. $\mathbf{X}$ second-order reweighted stationary, Baddeley, Møller, Waagepetersen, 2000)

Pair correlation and K-function

Pair correlation function

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

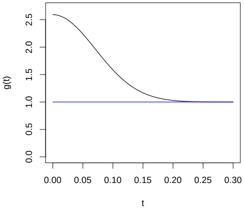
NB: independent points $\Rightarrow \rho^{(2)}(u,v) = \rho(u)\rho(v) \Rightarrow g(u,v) = 1$

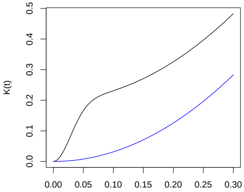
K-function

$$K(t) = \int_{\|h\| \le t} g(h)\mathrm{d}h$$

(provided $g(u,v) = g(u-v)$ i.e. $\mathbf{X}$ second-order reweighted stationary, Baddeley, Møller, Waagepetersen, 2000)

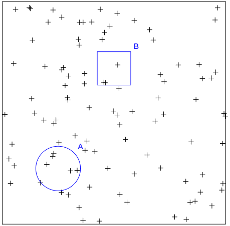
Examples of pair correlation and K-functions:





Mean and covariances of counts

A and B subsets of the plane. $N(A)$ and $N(B)$ random numbers/counts of points in A and B.



$$\mathbb{E}[N(A)] = \mu(A) = \int_A \rho(u)\mathrm{d}u$$

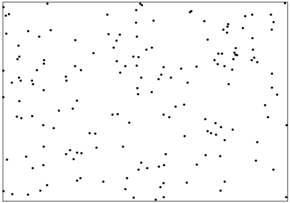
$$\mathbb{C}\mathrm{ov}[N(A), N(B)] = \int_{A\cap B} \rho(u)\mathrm{d}u + \int_A \int_B \rho(u)\rho(v)[g(u,v)-1]\mathrm{d}u\mathrm{d}v$$

NB: can compute means and covariances for any sets A and B !
(in contrast to quadrat count methods)

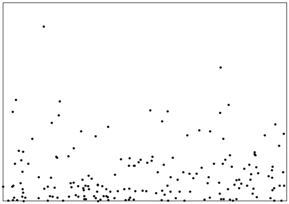
The Poisson process

X is a Poisson process with intensity function $\rho(\cdot)$ if for any bounded region B:

1. $N(B)$ is Poisson distributed with mean $\mu(B) = \int_B \rho(u)\mathrm{d}u$
2. Given $N(B) = n$, the n points are independent and identically distributed with density proportional to intensity function $\rho(\cdot)$.



Homogeneous: $\rho = 150/0.7$



Inhomogeneous: $\rho(x, y) \propto \mathrm{e}^{-10.6y}$

Back to rain forest: parametric models for intensity and pair correlation

Study influence of covariates

$$Z(u) = (Z_1(u), \ldots, Z_p(u))$$

using log-linear model for intensity function:

$$\log \rho(u; \beta) = \beta Z(u)^{\mathsf{T}} \Leftrightarrow \rho(u; \beta) = \exp(\beta Z(u)^{\mathsf{T}})$$

where

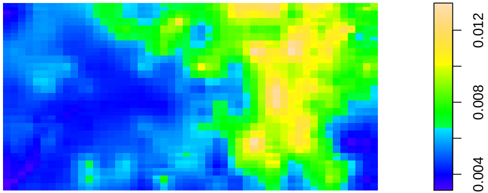
$$\beta Z(u)^{\mathsf{T}} = \beta_1 Z_1(u) + \beta_2 Z_2(u) + \ldots + \beta_p Z_p(u)$$

Capparis Frondosa and Poisson process ?

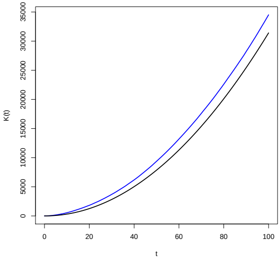
Fit model with covariates elevation and Potassium.

Fitted intensity function

$\rho(u;\hat{\beta}) = \exp(\hat{\beta}_0 + \hat{\beta}_1 \text{Elev}(u) + \hat{\beta}_2 \text{K}(u))$:



Estimated K-function and $K(t) = \pi t^2$-function for Poisson process:



Not Poisson process - aggregation due to unobserved factors (e.g. seed dispersal)

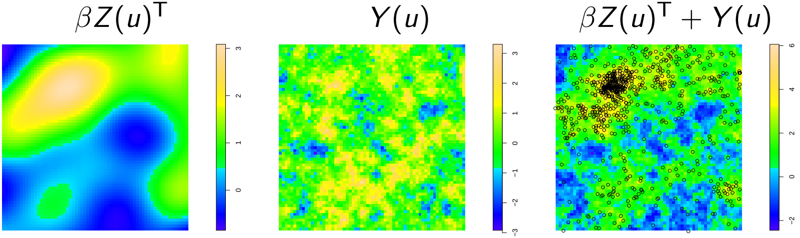
Cox processes

X is a *Cox process* driven by the random intensity function Λ if, conditional on $\Lambda = \lambda$, **X** is a Poisson process with intensity function λ.

Example: log Gaussian Cox process (Møller, Syversveen, W, 1998)

$$\log \Lambda(u) = \beta Z(u)^{\mathsf{T}} + Y(u)$$

where $\{Y(u)\}$ Gaussian random field.



Intensity and pair correlation function for log Gaussian Cox processes

Log linear intensity

$$\log \rho(u;\beta) = \mu + Z(u)\beta^\mathsf{T}$$

Intensity and pair correlation function for log Gaussian Cox processes

Log linear intensity

$$\log \rho(u;\beta) = \mu + Z(u)\beta^{\mathsf{T}}$$

Pair correlation function:

$$g(u-v;\psi) = \exp[c(u-v;\sigma^2,\alpha)], \quad \psi = (\sigma^2,\alpha)$$

where σ^2 variance of Gaussian field and $c(\cdot;\alpha)$ covariance function.

Intensity and pair correlation function for log Gaussian Cox processes

Log linear intensity

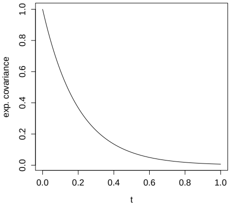
$$\log \rho(u;\beta) = \mu + Z(u)\beta^{\mathsf{T}}$$

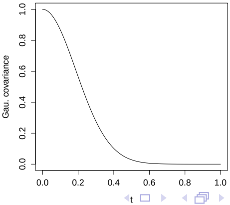
Pair correlation function:

$$g(u-v;\psi) = \exp[c(u-v;\sigma^2,\alpha)], \quad \psi = (\sigma^2,\alpha)$$

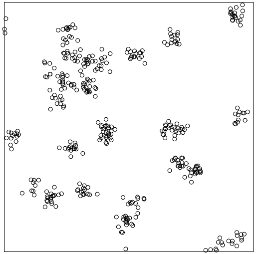
where σ^2 variance of Gaussian field and $c(\cdot;\alpha)$ covariance function.

Examples: $\sigma^2 \exp(-\|u-v\|/\alpha)$ and $\sigma^2 \exp(-\|u-v\|^2/\alpha)$





Cluster process: Inhomogeneous Thomas process (W, 2007)



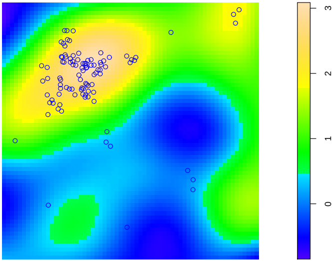
Parents stationary Poisson point process intensity κ

Offspring distributed around mothers according to Gaussian density with standard deviation ω

Inhomogeneity: offspring survive according to probability

$$p(u) \propto \exp(Z(u)\beta^{\mathsf{T}})$$

depending on covariates (independent thinning).



Cluster process as Cox process

Inhomogeneous Thomas is a Cox process with intensity function

$$\Lambda(u) = \Lambda_0(u)\exp[Z(u)\beta^{\mathsf{T}}]$$

where

$$\Lambda_0(u) = \sum_{v \in C} k(u - v)$$

and k Gaussian density with standard deviation ω.

More generally, k can be any bivariate probability density function
- but not all choices lead to explicit pair correlation function.

Intensity and pair correlation function for cluster-Cox processes

Log linear intensity

$$\log \rho(u; \beta) = \mu + Z(u)\beta^{\mathsf{T}}$$

Intensity and pair correlation function for cluster-Cox processes

Log linear intensity

$$\log \rho(u;\beta) = \mu + Z(u)\beta^{\mathsf{T}}$$

Pair correlation function for inhomogeneous Thomas:

$$\begin{aligned} g(u-v;\psi) &= 1 + \exp(-\|u-v\|^2/(4\omega)^2)/(4\omega^2\kappa\pi) \\ &= 1 + \sigma^2 \exp(-\|u-v\|^2/\alpha), \quad \psi = (\kappa,\omega) \text{ or } \psi = (\sigma^2,\alpha) \end{aligned}$$

Intensity and pair correlation function for cluster-Cox processes

Log linear intensity

$$\log \rho(u;\beta) = \mu + Z(u)\beta^{\mathsf{T}}$$

Pair correlation function for inhomogeneous Thomas:

$$\begin{aligned} g(u-v;\psi) &= 1 + \exp(-\|u-v\|^2/(4\omega)^2)/(4\omega^2\kappa\pi) \\ &= 1 + \sigma^2 \exp(-\|u-v\|^2/\alpha), \quad \psi = (\kappa,\omega) \text{ or } \psi = (\sigma^2,\alpha) \end{aligned}$$

Pair correlation function for more general Bessel cluster processes:

$$g(u-v;\psi) = 1 + \sigma^2 \frac{(\|u-v\|/\alpha)^{\nu} K_{\nu}(\|u-v\|/\alpha)}{2^{\nu-1}\Gamma(\nu)}$$

Special case $\nu = 1/2$:

$$g(u-v;\psi) = 1 + \sigma^2 \exp(-\|u-v\|/\alpha)$$

Parameter estimation

Possibilities:

1. Maximum likelihood estimation (Monte Carlo computation of likelihood function)
2. Simple estimating functions based on intensity function and pair correlation function - inspired by methods for count variables: least squares, composite likelihood, quasi-likelihood,...

Example: composite likelihood I (Schoenberg, 2005; W, 2007)

Consider indicators $X_i = 1[N_i > 0]$ for presence of points in cells C_i. $P(X_i = 1) = \rho_\beta(u_i)|C_i|$.

Example: composite likelihood I (Schoenberg, 2005; W, 2007)

Consider indicators $X_i = 1[N_i > 0]$ for presence of points in cells C_i. $P(X_i = 1) = \rho_\beta(u_i)|C_i|$.

Composite Bernouilli likelihood

$$\prod_{i=1}^{n}(P(X_i = 1))^{X_i}(1-P(X_i = 1))^{1-X_i} \equiv \prod_{i=1}^{n}\rho_\beta(u_i)^{X_i}(1-\rho_\beta(u_i)|C_i|)^{1-X_i}$$

has limit ($|C_i| \to 0$)

$$L(\beta) = [\prod_{u \in \mathbf{X} \cap W} \rho(u;\beta)] \exp(-\int_W \rho(u;\beta)\,\mathrm{d}u)$$

Example: composite likelihood I (Schoenberg, 2005; W, 2007)

Consider indicators $X_i = 1[N_i > 0]$ for presence of points in cells C_i. $P(X_i = 1) = \rho_\beta(u_i)|C_i|$.

Composite Bernouilli likelihood

$$\prod_{i=1}^{n}(P(X_i = 1))^{X_i}(1-P(X_i = 1))^{1-X_i} \equiv \prod_{i=1}^{n}\rho_\beta(u_i)^{X_i}(1-\rho_\beta(u_i)|C_i|)^{1-X_i}$$

has limit ($|C_i| \to 0$)

$$L(\beta) = [\prod_{u \in \mathbf{X} \cap W} \rho(u;\beta)] \exp(-\int_W \rho(u;\beta)\,\mathrm{d}u)$$

Estimate $\hat{\beta}$ maximizes $L(\beta)$.

NB: $L(\beta)$ formally equivalent to likelihood function of a Poisson process with intensity function $\rho_\beta(\cdot)$.

Example: minimum contrast estimation for ψ

Computationally easy approach if $\mathbf{X}$ second-order reweighted stationary so that K-function well-defined.

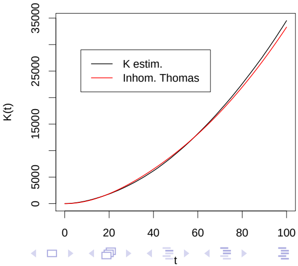
Estimate of K-function (Baddeley, Møller and W, 2000):

$$\hat{K}_\beta(t) = \sum_{u,v \in \mathbf{X} \cap W} \frac{1[0 < \|u - v\| \leq t]}{\rho(u;\beta)\rho(v;\beta)} e_{u,v}$$

Unbiased if β 'true' regression parameter.

Minimum contrast estimation: minimize squared distance between theoretical K and $\hat{K}$:

$$\hat{\psi} = \underset{\psi}{\operatorname{argmin}} \int_0^r \left(\hat{K}_{\hat{\beta}}(t) - K(t;\psi)\right)^2 \mathrm{d}t$$



Two-step estimation

Obtain estimates $(\hat{\beta}, \hat{\psi})$ in two steps

1. obtain $\hat{\beta}$ using composite likelihood
2. obtain $\hat{\psi}$ using minimum contrast

Clustering and mode of seed dispersal

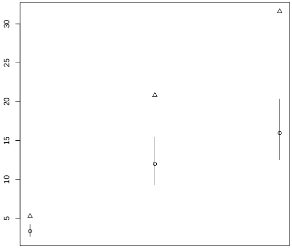
Fit Thomas cluster process with log linear model for intensity function.

Acalypha and *Capparis*: positive dependence on elevation and potassium (significantly positive coefficients $\hat{\beta} = (0.02, 0.005)$ and $\hat{\beta} = (0.03, 0.004)$).

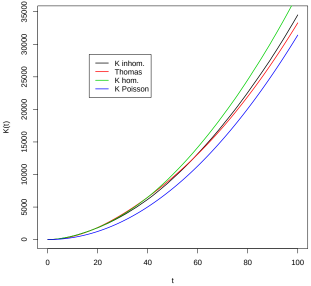
Loncocharpus: negative dependence on nitrogen and phosphorous ($\hat{\beta} = (-0.03, -0.16)$).

Recall ω = 'width' of clusters.

Estimates of ω for explosive, wind and bird/mammal:



Estimates of K-functions for bird/mammal dispersed species



Triangles: model without covariates.

Decomposition of variance for rain forest tree point patterns (Shen, Jalilian, W, in progress)

Question: how much of the spatial variation for rain forest trees is due to environment ?

Decomposition of variance for rain forest tree point patterns (Shen, Jalilian, W, in progress)

Question: how much of the spatial variation for rain forest trees is due to environment ?

Variance of a count $N(B)$ (number of points in region B) for a stationary Cox process (constant intensity ρ):

Decomposition of variance for rain forest tree point patterns (Shen, Jalilian, W, in progress)

Question: how much of the spatial variation for rain forest trees is due to environment ?

Variance of a count $N(B)$ (number of points in region B) for a stationary Cox process (constant intensity ρ):

$$\mathbb{V}\mathrm{ar} N(B) = \int_B \rho \mathrm{d}u$$

Variance=Poisson variance

Decomposition of variance for rain forest tree point patterns (Shen, Jalilian, W, in progress)

Question: how much of the spatial variation for rain forest trees is due to environment ?

Variance of a count $N(B)$ (number of points in region B) for a stationary Cox process (constant intensity ρ):

$$\mathbb{V}\mathrm{ar} N(B) = \int_B \rho \mathrm{d}u + \int_B \int_B \rho^2 [g(u,v) - 1] \mathrm{d}u \mathrm{d}v$$

Variance=Poisson variance+Extra variance due to random intensity

Decomposition of variance for log linear random intensity:

Decomposition of variance for log linear random intensity:

$$\mathbb{V}\mathrm{ar} \log \Lambda(u) = \mathbb{V}\mathrm{ar} \beta Z(u)^{\mathsf{T}}$$

Variance=Environment

Decomposition of variance for log linear random intensity:

$$\mathbb{V}\mathrm{ar}\log\Lambda(u) = \mathbb{V}\mathrm{ar}\beta Z(u)^{\mathsf{T}} + \mathbb{V}\mathrm{ar}\, Y(u) = \sigma_Z^2 + \sigma_Y^2$$

Variance=Environment+Seed dispersal

Decomposition of variance for log linear random intensity:

$$\mathbb{V}\mathrm{ar} \log \Lambda(u) = \mathbb{V}\mathrm{ar} \beta Z(u)^{\mathsf{T}} + \mathbb{V}\mathrm{ar}\, Y(u) = \sigma_Z^2 + \sigma_Y^2$$

Variance=Environment+Seed dispersal

Note $\tilde{Z}(u) = \beta Z(u)^{\mathsf{T}}$ regarded as stationary random process.

Decomposition of variance for log linear random intensity:

$$\mathbb{V}\mathrm{ar} \log \Lambda(u) = \mathbb{V}\mathrm{ar} \beta Z(u)^\mathsf{T} + \mathbb{V}\mathrm{ar}\, Y(u) = \sigma_Z^2 + \sigma_Y^2$$

Variance=Environment+Seed dispersal

Note $\tilde{Z}(u) = \beta Z(u)^\mathsf{T}$ regarded as stationary random process.

Estimate β and σ_Y^2 using two-step approach.

Simple empirical estimate of σ_Z^2:

$$\hat{\sigma}_Z^2 = \frac{1}{n_G} \sum_{u \in G} (\tilde{Z}(u) - \bar{\tilde{Z}})^2$$

Compute

$$\frac{\sigma_Z^2}{\sigma_Z^2 + \sigma_Y^2} \quad \text{and} \quad \frac{\sigma_Y^2}{\sigma_Z^2 + \sigma_Y^2}$$

Decomposition of variance for log linear random intensity:

$$\mathbb{V}\mathrm{ar} \log \Lambda(u) = \mathbb{V}\mathrm{ar} \beta Z(u)^{\mathsf{T}} + \mathbb{V}\mathrm{ar}\, Y(u) = \sigma_Z^2 + \sigma_Y^2$$

Variance=Environment+Seed dispersal

Note $\tilde{Z}(u) = \beta Z(u)^{\mathsf{T}}$ regarded as stationary random process.

Estimate β and σ_Y^2 using two-step approach.

Simple empirical estimate of σ_Z^2:

$$\hat{\sigma}_Z^2 = \frac{1}{n_G} \sum_{u \in G} (\tilde{Z}(u) - \bar{\tilde{Z}})^2$$

Compute

$$\frac{\sigma_Z^2}{\sigma_Z^2 + \sigma_Y^2} \quad \text{and} \quad \frac{\sigma_Y^2}{\sigma_Z^2 + \sigma_Y^2}$$

Can also define closely related "R^2" summarizing how much of variation in Λ is due to Z.

Additive model for random intensity function (Jalilian, Guan, W, in progress)

Alternative to log linear model:

$$\Lambda(u) = \beta Z(u)^{\mathsf{T}} + Y(u)$$

Cox process superposition of point processes with (random) intensity functions $\tilde{Z}(u) = \beta Z(u)^{\mathsf{T}}$ and $Y(u)$

Straightforward variance decomposition for Λ:

$$\mathbb{V}\mathrm{ar}\Lambda(u) = \mathbb{V}\mathrm{ar}\tilde{Z}(u) + \mathbb{V}\mathrm{ar}\,Y(u) = \sigma_Z^2 + \sigma_Y^2$$

$$R^2 = \frac{\sigma_Z^2}{\sigma_Z^2 + \sigma_Y^2}$$

Results

Consider pair correlation functions of the form

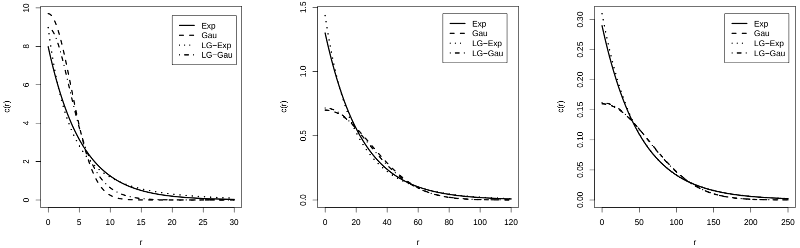
$$g(u-v;\sigma^2,\alpha) = 1 + \sigma^2 \exp(-\|u-v\|^\delta/\alpha) \quad \delta = 1 \text{ or } \delta = 2$$

Species	Λ	δ	R^2	Goodness of fit ("AIC")
Acalypha	log linear	1	0.01	1178
	log linear	2	0.01	1198
	additive	1	0.01	1565
	additive	2	0.01	1582
Lonchocarpus	log linear	1	0.10	3053
	log linear	2	0.17	3105
	additive	1	0.06	4001
	additive	2	0.10	4026
Capparis	log linear	1	0.25	4938
	log linear	2	0.38	5230
	additive	1	0.20	8736
	additive	2	0.33	9157

Best fit with log linear model and $\delta = 1$. Largest R^2 for bird/mammal dispersion. Smallest for explosive capsules.

Fitted pair correlation functions

Plots show $g(u - v) - 1$:



Final remarks

- Log-linear model gives better fit than additive
- Better fit with exponential covariance function than with Gaussian
- Gaussian covariance function (Thomas process) tails decay too fast
- value of R^2 related to mode of seed dispersal ?

Thank you for your attention !