Decomposition of variance for spatial Cox processes

Rasmus Waagepetersen
Department of Mathematical Sciences
Aalborg University

Joint work with Abdollah Jalilian and Yongtao Guan

November 8, 2010

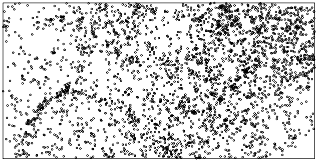
Background: Tropical rain forest ecology

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

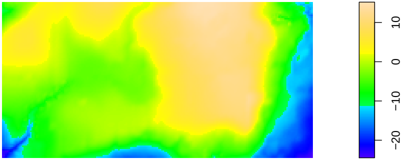
Key factors:

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

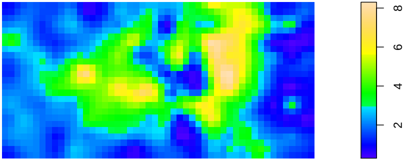
Example: *Capparis Frondosa*



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



Elevation



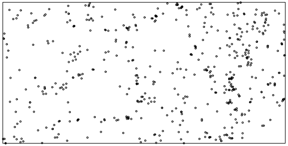
Potassium content in soil.

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

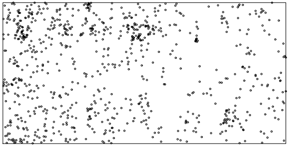
Example: modes of seed dispersal and clustering

Three species with different modes of seed dispersal:

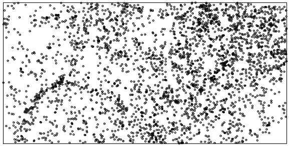
Acalypha Diversifolia explosive capsules



Loncocharpus Heptaphyllus wind



Capparis Frondosa bird/mammal



Is degree of clustering related to mode of seed dispersal ?

Quantify how much of the spatial variation is due to respectively environment and seed dispersal ?

Approach: Cox process model for joint effects of environment and seed dispersal.

Cox processes

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

Intensity function

$$\rho(u) = \mathbb{E}\Lambda(u)$$

Second-order product density

$$\rho^{(2)}(u, v) = \mathbb{E}\Lambda(u)\Lambda(v) = \mathbb{C}\mathrm{ov}[\Lambda(u), \Lambda(v)] + \rho(u)\rho(v)$$

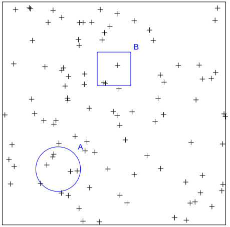
Pair correlation function

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

Mean and covariances of counts

A and B subsets of the plane

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



$$\mathbb{C}\mathrm{ov}[N(A), N(B)] = \int_{A\cap B} \mathbb{E}\Lambda(u)\mathrm{d}u + \int_A \int_B \mathbb{C}\mathrm{ov}[\Lambda(u), \Lambda(v)]\mathrm{d}u\mathrm{d}v$$
$$= \int_{A\cap B} \rho(u)\mathrm{d}u + \int_A \int_B \rho(u)\rho(v)[g(u,v) - 1]\mathrm{d}u\mathrm{d}v$$
$$= \text{Poisson variance} + \text{extra variance due to } \Lambda$$

Decomposition of variance for a count

Prediction of count $N(B)$ given Λ :

$$\hat{N}(B) = \mathbb{E}[N(B)|\Lambda] = \int_B \Lambda(u)\mathrm{d}u$$

$$\mathbb{V}\mathrm{ar} N(B) =$$

$$\text{variation} =$$

Decomposition of variance for a count

Prediction of count $N(B)$ given Λ :

$$\hat{N}(B) = \mathbb{E}[N(B)|\Lambda] = \int_B \Lambda(u)\mathrm{d}u$$

$$\mathbb{V}\mathrm{ar} N(B) = \mathbb{V}\mathrm{ar} \hat{N}(B)$$

variation = structured variation

Decomposition of variance for a count

Prediction of count $N(B)$ given Λ :

$$\hat{N}(B) = \mathbb{E}[N(B)|\Lambda] = \int_B \Lambda(u)\mathrm{d}u$$

$$\mathbb{V}\mathrm{ar} N(B) = \mathbb{V}\mathrm{ar} \hat{N}(B) + \mathbb{V}\mathrm{ar}\left[N(B) - \hat{N}(B)\right]$$

variation = structured variation + 'Poisson noise'

Decomposition of variance for a count

Prediction of count $N(B)$ given Λ :

$$\hat{N}(B) = \mathbb{E}[N(B)|\Lambda] = \int_B \Lambda(u)\mathrm{d}u$$

$$\mathbb{V}\mathrm{ar}N(B) = \mathbb{V}\mathrm{ar}\hat{N}(B) + \mathbb{V}\mathrm{ar}\left[N(B) - \hat{N}(B)\right]$$

variation = structured variation + 'Poisson noise'

Further, (Z=environmental variables)

$$\mathbb{V}\mathrm{ar}\Lambda(u) =$$

variation of Λ =

Decomposition of variance for a count

Prediction of count $N(B)$ given Λ :

$$\hat{N}(B) = \mathbb{E}[N(B)|\Lambda] = \int_B \Lambda(u)\mathrm{d}u$$

$$\mathbb{V}\mathrm{ar}N(B) = \mathbb{V}\mathrm{ar}\hat{N}(B) + \mathbb{V}\mathrm{ar}\left[N(B) - \hat{N}(B)\right]$$

variation = structured variation + 'Poisson noise'

Further, (Z=environmental variables)

$$\mathbb{V}\mathrm{ar}\Lambda(u) = \mathbb{V}\mathrm{ar}\hat{\Lambda}(u) \qquad \hat{\Lambda}(u) = \mathbb{E}[\Lambda(u)|Z]$$

variation of Λ = variation due to environment

Decomposition of variance for a count

Prediction of count $N(B)$ given Λ :

$$\hat{N}(B) = \mathbb{E}[N(B)|\Lambda] = \int_B \Lambda(u)\mathrm{d}u$$

$$\mathbb{V}\mathrm{ar} N(B) = \mathbb{V}\mathrm{ar} \hat{N}(B) + \mathbb{V}\mathrm{ar}\left[N(B) - \hat{N}(B)\right]$$

variation = structured variation + ‘Poisson noise’

Further, (Z=environmental variables)

$$\mathbb{V}\mathrm{ar}\Lambda(u) = \mathbb{V}\mathrm{ar}\hat{\Lambda}(u) + \mathbb{V}\mathrm{ar}\left[\Lambda(u) - \hat{\Lambda}(u)\right] \quad \hat{\Lambda}(u) = \mathbb{E}[\Lambda(u)|Z]$$

variation of Λ = variation due to environment + other sources

Measure of influence of environmental covariates Z:

$$R^2 = \frac{\mathbb{V}\mathrm{ar}\hat{\Lambda}(u)}{\mathbb{V}\mathrm{ar}\Lambda(u)} = \frac{\mathbb{V}\mathrm{ar}\mathbb{E}[\Lambda(u)|Z]}{\mathbb{V}\mathrm{ar}\Lambda(u)}$$

(right hand side does not depend on u in case of stationary environment)

Analogy to linear regression:

$$SSR = \mathbb{E}\int_W \left[\hat{\Lambda}(u) - \rho(u)\right]^2 \mathrm{d}u = |W|\mathbb{V}\mathrm{ar}\hat{\Lambda}(u)$$

$$SST = \mathbb{E}\int_W [\Lambda(u) - \rho(u)]^2 \mathrm{d}u = |W|\mathbb{V}\mathrm{ar}\Lambda(u)$$

Then

$$R^2 = \frac{SSR}{SST}$$

Additive model for Λ

Additive model:

$$\Lambda(u) = \widetilde{Z}(u) + \Lambda_0(u)$$

$\widetilde{Z}$ contribution from environment:

$$\widetilde{Z}(u) = \beta Z(u)^{\mathsf{T}} \quad Z(u) = (Z_1(u), \ldots, Z_p(u))$$

Λ_0: structured variation (e.g. seed dispersal) independent of environment.

Cox process **X** union of independent Cox processes with random intensity functions $\widetilde{Z}(u)$ and $\Lambda_0(u)$.

Assume $\widetilde{Z}$ and Λ_0 non-negative and stationary.

R^2 for additive model

$$R^2 = \frac{\sigma^2_{\widetilde{Z}}}{\sigma^2_{\widetilde{Z}} + \sigma^2_0}$$

$$\sigma^2_{\widetilde{Z}} = \mathbb{V}\mathrm{ar}\widetilde{Z}(u) \quad \text{and } \sigma^2_0 = \mathbb{V}\mathrm{ar}\Lambda_0(u)$$

Log-linear model

$$\Lambda(u) = \Lambda_0(u) \exp[\widetilde{Z}(u)]$$

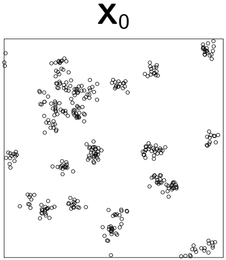
Interpretation in terms of survival of seedlings:

$\mathbf{X}_0$ seedlings: stationary Cox process with random intensity function Λ_0.

$\mathbf{X}$ thinning of $\mathbf{X}_0$ with survival depending on environment $\widetilde{Z}$.

$\widetilde{Z}$ does not need to be non-negative.

Example: Cox/cluster process: Inhomogeneous Thomas process

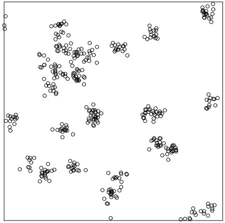


Λ_0 shot-noise process $\Rightarrow$ $\mathbf{X}_0$ cluster process:

Offspring distributed according to Gaussian density around Poisson parents

Example: Cox/cluster process: Inhomogeneous Thomas process

$\mathbf{X}_0$



Λ_0 shot-noise process $\Rightarrow$ $\mathbf{X}_0$ cluster process:

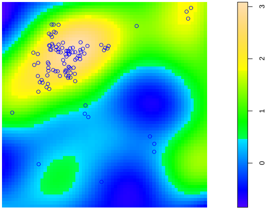
Offspring distributed according to Gaussian density around Poisson parents

Inhomogeneity: offspring in $\mathbf{X}_0$ survive according to probability

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

depending on covariates (independent thinning).

$\mathbf{X}$



R^2 for log-linear model

$$R^2 = \frac{\sigma^2_{\exp \widetilde{Z}}}{\sigma^2_{\exp \widetilde{Z}} + \sigma_0^2 \left[\sigma^2_{\exp \widetilde{Z}} + \rho^2_{\exp \widetilde{Z}}\right]}$$

$$\sigma^2_{\exp \widetilde{Z}} = \mathbb{V}\mathrm{ar} \exp[\widetilde{Z}(u)]$$
$$\rho_{\exp \widetilde{Z}} = \mathbb{E} \exp[\widetilde{Z}(u)]$$
$$\sigma_0^2 = \mathbb{V}\mathrm{ar} \Lambda_0(u)$$

Models for $c_0(u-v) = \mathbb{C}\mathrm{ov}[\Lambda_0(u), \Lambda_0(v)]$

NB: any positive definite function is a covariance function but not necessarily for a non-negative random process Λ_0. Use covariance functions from explicit constructions of Λ_0.

Log-Gaussian:

$$\Lambda_0(u) = \exp[Y(u)] \quad c_0(u-v) = \rho_0^2 \exp[\mathbb{C}\mathrm{ov}(Y(u), Y(v))] - \rho_0^2$$

where Y Gaussian field and $\rho_0 = \mathbb{E}\Lambda_0(u)$.

Models for $c_0(u-v) = \mathbb{C}\mathrm{ov}[\Lambda_0(u), \Lambda_0(v)]$

NB: any positive definite function is a covariance function but not necessarily for a non-negative random process Λ_0. Use covariance functions from explicit constructions of Λ_0.

Log-Gaussian:

$$\Lambda_0(u) = \exp[Y(u)] \quad c_0(u-v) = \rho_0^2 \exp[\mathbb{C}\mathrm{ov}(Y(u), Y(v))] - \rho_0^2$$

where Y Gaussian field and $\rho_0 = \mathbb{E}\Lambda_0(u)$.

Shot-noise:

$$\Lambda_0(u) = \sum_{u \in C} \alpha k(u-v) \quad c_0(u-v) = \kappa\alpha^2 \int_{\mathbb{R}^2} k(w)k(w+v-u)\mathrm{d}w$$

where C homogeneous Poisson with intensity κ and $k(\cdot)$ probability density.

Matérn class

Λ_0 shot-noise process: sum of $K_{(\nu-1)/2}$ Bessel densities $k(\cdot)$ centered around points of homogeneous Poisson process.

$$c_0(h) = \kappa\alpha^2 \int_{\mathbb{R}^2} k(w)k(w+v-u)\mathrm{d}w = \sigma_0^2 \frac{(\|h\|/\eta)^\nu K_\nu(\|h\|/\eta)}{2^{\nu-1}\Gamma(\nu)}$$

With $\nu = 1/2$ (exponential covariance function)

$$c_0(h) = \sigma_0^2 \exp(-\|h\|/\eta)$$

With '$\nu = \infty$' ('Gaussian' covariance function)

$$c_0(h) = \sigma_0^2 \exp[-(\|h\|/\eta)^2]$$

Parameter estimation (β)

$\mathbf{X}$ observed within $W \subset \mathbb{R}^2$

Estimate β and $\psi = (\sigma_0^2, \eta, \nu)$ using $\mathbf{X}|Z$.

First-order composite likelihood:

$$\mathrm{CL}_1(\beta) = \sum_{u \in \mathbf{X} \cap W} \log \rho(u|Z;\beta) - \int_W \rho(u|Z;\beta)\mathrm{d}u$$

$\rho(\cdot|Z,\beta)$ intensity function for $\mathbf{X}|Z$.

For additive model:

$$\rho(u|Z,\beta) = \rho_0 + \beta Z(u)^{\mathsf{T}} \quad \rho_0 = \mathbb{E}\Lambda_0(u)$$

Estimation of ψ

Second-order composite likelihood (given $\hat{\beta}$):

$$\mathrm{CL}_2(\psi|\hat{\beta}) = \sum_{\substack{u,v\in\mathbf{X}\cap W\\ \|u-v\|\le R}}^{\neq} \log \rho^{(2)}(u,v|Z;\hat{\beta},\psi) - \iint_{\|u-v\|\le R} \rho^{(2)}(u,v|Z;\hat{\beta},\psi)\mathrm{d}u\mathrm{d}v$$

$\rho^{(2)}(\cdot,\cdot|Z;\beta,\psi)$ second-order product density for $\mathbf{X}|Z$.

For additive model:

$$\rho^{(2)}(\cdot,\cdot|Z;\beta,\psi) = \rho(u|Z,\beta)\rho(v|Z,\beta) + c_0(u-v;\psi)$$

Environmental variances

Z observed on grid $G = \{u_i\}_{i=1,\dots,M}$ and $\widehat{\widetilde{Z}}(u) = \widehat{\beta} Z(u)^{\mathsf{T}}$.

$$\widehat{\sigma}^2_{\widetilde{Z}} = \frac{1}{M} \sum_{u \in G} \left\{ \widehat{\widetilde{Z}}(u) - \widehat{\rho}_{\widetilde{Z}} \right\}^2$$

and

$$\widehat{\sigma}^2_{\exp \widetilde{Z}} = \frac{1}{M} \sum_{u \in G} \left\{ \exp \left[\widehat{\widetilde{Z}}(u) \right] - \widehat{\rho}_{\exp \widetilde{Z}} \right\}^2$$

where

$$\widehat{\rho}_{\widetilde{Z}} = \frac{1}{M} \sum_{u \in G} \widehat{\widetilde{Z}}(u) \qquad \widehat{\rho}_{\exp \widetilde{Z}} = \frac{1}{M} \sum_{u \in G} \exp \left[\widehat{\widetilde{Z}}(u) \right]$$

Results for rain forest data

Covariates elevation and potassium for Acalypha and Capparis. Nitrogen and phosphorous for Lonchocarpus.

Qualitative similar results for additive and log-linear model regarding dependence on covariates (β).

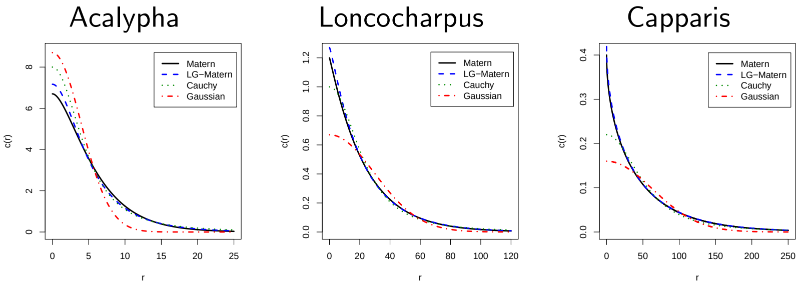
Species	model for Λ	c_0	$\mathrm{CL}_2(\widehat{\psi}\|\widehat{\beta})$	R^2
Acalypha	log-linear	'Gaussian'	-1239.8	0.01
		Matérn	-1221.0	0.02
		LG-Matérn	-1204.6	0.01
	additive	'Gaussian'	-1641.7	0.01
		Matérn	-1623.4	0.01
		LG-Matérn	-1623.4	0.01

LG-Matérn: $c_0(h) = \rho_0^2[\exp(c(h)) - 1]$ where $c(\cdot)$ Matérn.

Results continued

Species	model for Λ	c_0	$\mathrm{CL}_2(\widehat{\psi}\vert\widehat{\beta})$	R^2
Loncocharpus	log-linear	'Gaussian'	-3204	0.17
		Matérn	-3156	0.10
		LG-Matérn	-3153	0.10
	additive	'Gaussian'	-4081	0.11
		Matérn	-4055	0.07
		LG-Matérn	-4055	0.07
Capparis	log-linear	'Gaussian'	-76657	0.37
		Matérn	-76325	0.19
		LG-Matérn	-76311	0.19
	additive	'Gaussian'	-81139	0.23
		Matérn	-80685	0.16
		LG-Matérn	-80685	0.16

Fitted covariance functions (log-linear model)



Some conclusions

Estimated R^2 sensitive to choice of model: especially c_0.

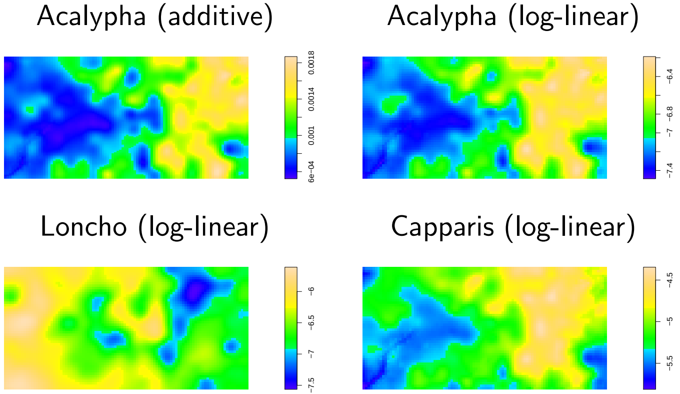
Best fit with log-linear model (interpretation in terms of survival).

Best fit with (LG)-Matérn (heavy tails for covariance/cluster density).

Largest R^2 for Capparis (bird/mammal seed dispersal), smallest for Acalypha (explosive capsules).

Stationary environment ?

$\widehat{\widetilde{Z}}$:



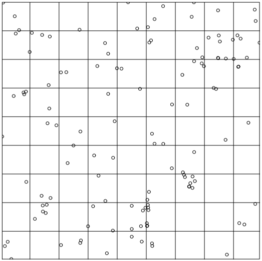
Handling possible non-stationarity topic for further research.

Composite likelihood obtained from binary random field

Random count variables: $N_i = \#\mathbf{X} \cap C_i$ number of points in C_i.

Disjoint subdivision $W = \cup_{i=1}^{n} C_i$ in 'cells' C_i.

$X_i = 1[N_i > 0]$ binary random variable.
$P(X_i = 1) = \rho(u_i; \beta)|C_i|$.



Bernouilli composite likelihood

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

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

$$\mathrm{CL}_1(\beta) = \sum_{u \in \mathbf{X} \cap W} \log \rho(u; \beta) - \int_W \rho(u; \beta) \mathrm{d}u$$

Second-order composite likelihood obtained from binary variables associated with joint presence of points in pairs of cells C_i and C_j:

$$X_{ij} = 1[N_i > 0, N_j > 0] \quad (= N_i N_j \text{ when } C_i \text{ and } C_j \text{ small })$$

where

$$P(X_{ij} = 1) = \mathbb{E} N_i N_j = \rho^{(2)}(u, v; \beta, \psi)|C_i||C_j|$$

Alternative: GEE

Observation vector

$$Y = [N_1, \ldots, N_m]$$

Mean vector

$$\mu = \left[\rho(u_1)|C_1|, \ldots, \rho(u_m)|C_m|\right]$$

Working covariance matrix

$$V = [V_{ij}]_{ij}$$

$$V_{ii} = \mathbb{V}\mathrm{ar} N_i = \rho(u_i)|C_i| + \rho(u_i)^2[g(u_i, u_i; \psi) - 1]$$

$$V_{ij} = \mathbb{C}\mathrm{ov}[N_i, N_j] = \rho(u_i)\rho(u_j)[g(u_i, u_j; \psi) - 1]$$

where $g(\cdot; \psi)$ working pair correlation function.

GEE

$$[Y-\mu]V^{-1}D \qquad D = \frac{\mathrm{d}\mu^{\mathsf{T}}}{\mathrm{d}\beta}$$

has limiting form (using Neuman series for V^{-1})

$$\sum_{u \in \mathbf{X} \cap W} \frac{\rho'(u;\beta)}{\rho(u)} - \int_W \rho'(u;\beta)\frac{\rho(u;\beta)}{\rho(u)}\mathrm{d}u$$

$$+ \sum_{v_0 \in \mathbf{X} \cap W} \frac{1}{\rho(v_0)} \sum_{k=1}^{\infty} \int_{W^k} \prod_{l=1}^{k} \big(\rho(v_{l-1})[g(v_{l-1}, v_l) - 1]\big) \rho'(v_k;\beta)\mathrm{d}v_1 \cdots \mathrm{d}v_k$$

$$- \int_W \frac{\rho(v_0;\beta)}{\rho(v_0)} \sum_{k=1}^{\infty} \int_{W^k} \prod_{l=1}^{k} \big(\rho(v_{l-l})[g(v_{l-1}, v_l) - 1]\big) \rho'(v_k;\beta)\mathrm{d}v_1 \cdots \mathrm{d}v_k$$

Topic of current research: truncate infinite sum or approximate integral ?