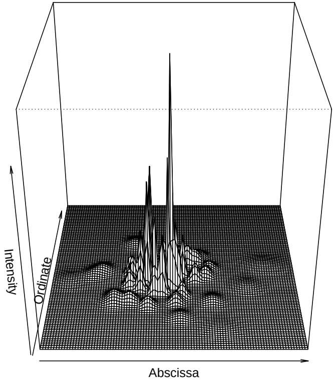
Doubly Inhomogeneous Cluster Point Processes

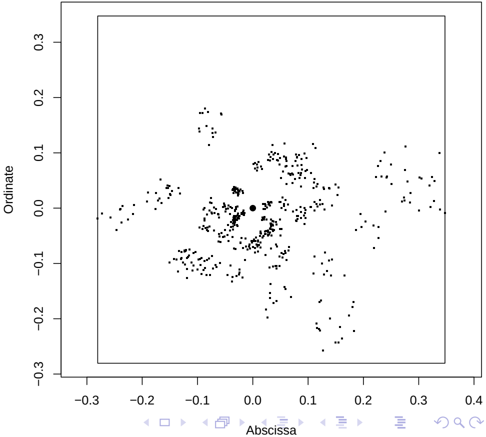
Tomáš Mrkvička, Samuel Soubeyrand

May 2016

Motivation - Group dispersal model

- It is dispersal model, where particles are released in groups by a single point source located at the origin of the planar space.





Motivation - Group dispersal model

- The number of released groups is Poisson distributed with mean value λ.
- The group sizes are Poisson distributed with mean value μ.
- The particles of a group are randomly deposited around a cluster center $c \in \mathbb{R}^2$ under the bivariate normal density:

$$k(u, \omega(\nu, c)) = \frac{1}{2\pi\nu||c||} \exp\left(-\frac{(u-c)'(u-c)}{2\nu||c||}\right), \; u \in \mathbb{R}^2,$$

whose dispersion linearly increases with the Euclidean distance $||c||$.

- Cluster centers are independently and identically drawn from the exponential dispersal kernel:

$$f(\beta, u) = \frac{1}{2\pi\beta^2} \exp\left(-||u||/\beta\right), \; u \in \mathbb{R}^2. \tag{1}$$

Motivation - Fish clustering model

- How fish cluster with respect to covariates.

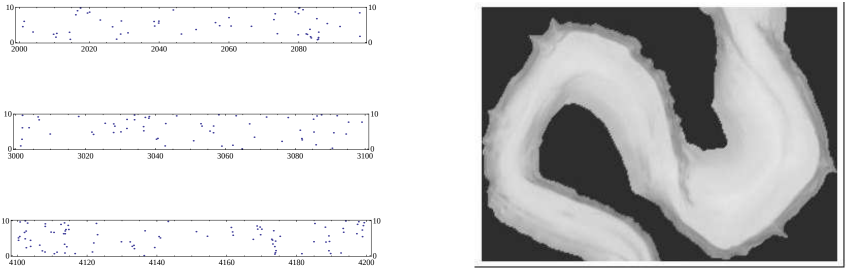


Figure: Left - Example of three birds eye of individual fish positions, coordinates are in meters. Right - Estimated inhomogeneity intensity function.

Covariates

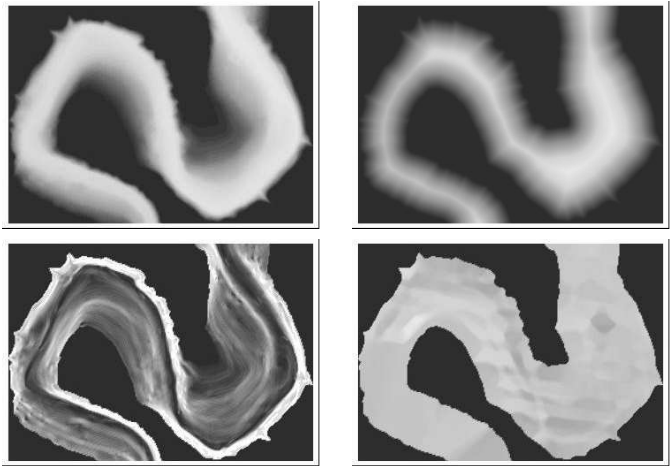


Figure: Four covariates, depth of the reservoir, distance from the bank, steepness of the bottom and light radiation. (Lighter colors correspond to the higher values.)

Motivation - Fish clustering model

- The number of clusters is Poisson distributed with mean value λ.
- The cluster sizes are Poisson distributed with mean value $\mu(c),\ c \in \mathbb{R}^2$.
- The fish are randomly spread around a cluster center $c \in \mathbb{R}^2$ under the bivariate normal density:

$$k(u, \omega(c)) = \frac{1}{2\pi\omega(c)} \exp\left(-\frac{(u-c)'(u-c)}{\omega(c)}\right),\ u \in \mathbb{R}^2.$$

- Cluster centers are independently and identically drawn from inhomogeneous density $f(u)$.
- Assuming linearity the cluster parameters can be modeled by

$$\mu(c) = z(c)\beta_\mu^T,\ \ \omega(c) = z(c)\beta_\omega^T. \tag{2}$$

where $z = (z_1, \ldots, z_k)$ is a vector of covariates and $\beta = (\beta_1, \ldots, \beta_k)$ is a regression parameter.

Doubly inhomogeneous cluster process

- First step, cluster centers C is an inhomogeneous point process with intensity function $u \mapsto \lambda f(\beta, u)$, $u \in \mathbb{R}^2$.
- Second step, clusters X_c, $c \in C$, are independent processes with intensity functions depending on c and satisfying $u \mapsto N(\kappa(\mu, c))k(u - c, \omega(\nu, c))$, $u \in \mathbb{R}^2$.
- $N(\kappa(\mu, c))$ is the expected number of daughter points per cluster parameterized by $\kappa(\mu, c)$.
- $u \mapsto k(u, \omega(\nu, c))$ determines the spread of daughter points around its mother.
- The parametric functions κ and ω are assumed to be known.
- Doubly inhomogeneous Neyman-Scott point process = the spatial distribution of the cluster centers is Poisson, and the spatial distribution of the points of each cluster is also Poisson.

Doubly inhomogeneous Neyman-Scott process

The intensity:

$$\begin{aligned}\alpha(u) &= \mathbb{E}\left(\sum_{c \in C} N(\kappa(\mu, c))k(u - c, \omega(\nu, c))\right) \\ &= \lambda \int_{\mathbb{R}^2} N(\kappa(\mu, c))k(u - c, \omega(\nu, c))f(\beta, c)dc.\end{aligned} \tag{3}$$

The second order product density:

$$\begin{aligned}\rho^{(2)}(u, v) &= \alpha(u)\alpha(v) \\ &+ \lambda^2 \int_{\mathbb{R}^2} N(\kappa(\mu, c))k(u - c, \omega(\nu, c))k(v - c, \omega(\nu, c))f(\beta, c)dc.\end{aligned} \tag{4}$$

Doubly inhomogeneous Neyman-Scott process

The probability density of the doubly inhomogeneous Neyman-Scott point process X observed in a window W with respect to the homogeneous Poisson point process under the knowledge of C and all the parameters is:

$$p(X|C,\beta,\lambda,\mu,\nu) = \exp\left(|W| - \int_W \alpha(u)du\right)\prod_{x\in X}\alpha(x), \qquad (5)$$

where $|W|$ is the area of W.

Bayesian parameter estimation

• $p(C|\lambda, \beta)$ is the probability density of C under the knowledge of λ and β with respect to the homogeneous Poisson point process.
• The joint posterior distribution of C and the parameters is then:

$$p(C, \beta, \lambda, \mu, \nu|X) \propto p(X|C, \beta, \lambda, \mu, \nu)p(C|\lambda, \beta)p(\beta)p(\mu)p(\nu), \tag{6}$$

where $p(\beta)$, $p(\mu)$ and $p(\nu)$ denote prior probability densities.
• No prior for λ is required because it is, in our estimation procedure, a deterministic function of μ and β (Kopecky & Mrkvička, 2016). Indeed:

$$M = \lambda * \int_W N(\kappa(\mu, u))f(\beta, u)du.$$

Bayesian parameter estimation

Our MCMC simulation algorithm, which maximizes the posterior density, consists of
1) updating the process of centers C by Birth-Death-Move algorithm and
2) updating the parameters β, μ and ν by Metropolis-Hastings algorithm.

The priors were set to wide uniform distributions to reduce the effect of priors on estimation.

Does it work? - Simulation study

Four series of 100 simulations of the group dispersal model.

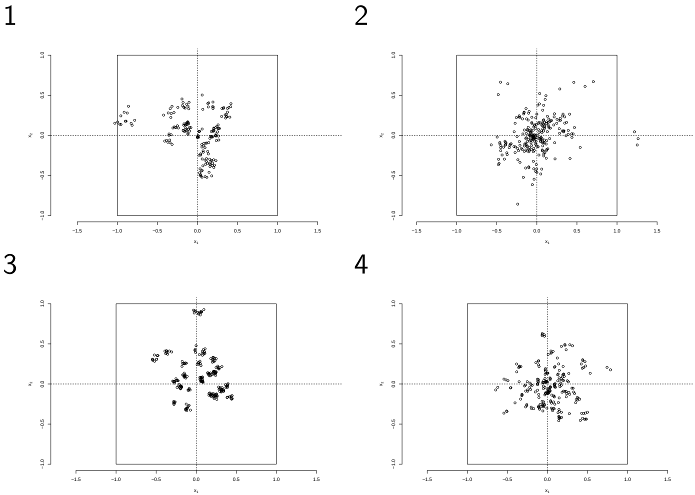


Figure: $\beta = 0.15$ and $(\lambda, \mu, \nu) = (25, 10, 0.005)$ (panel *1*), $(\lambda, \mu, \nu) = (100, 2.5, 0.005)$ (panel *2*), $(\lambda, \mu, \nu) = (25, 10, 0.001)$ (panel *3*), and $(\lambda, \mu, \nu) = (100, 2.5, 0.001)$ (panel *4*).

Does it work? - Simulation study

- For the estimation procedure, we used the extended window $W_{\text{ext}} = [-1.1, 1.1] \times [-1.1, 1.1]$.
- The MCMC chains were run with 1 000 000 steps.
- We discarded 150 000 steps as the burn-in.
- The posterior characteristics were calculated from every tenth step of the chain to avoid substantial dependencies.

Does it work? - Simulation study

Parameter	β	λ	μ	ν
Series 1				
True value	0.150	25.0	10.0	0.0050
Average posterior mean	0.174	18.7	14.0	0.0055
Average 95%-posterior int.	[0.122,0.253]	[15.5,22.8]	[11.4,16.8]	[0.0045,0.0065]
Cov. of true value by 95% int.	93	28	23	79
Series 2				
True value	0.150	100.0	2.50	0.0050
Average posterior mean	0.158	143	2.75	0.0034
Average 95%-posterior int.	[0.133,0.190]	[80.5,224.5]	[1.52,4.49]	[0.0019,0.0055]
Cov. of true value by 95% int.	92	67	68	48
Series 3				
True value	0.150	25.0	10.0	0.0010
Average posterior mean	0.165	22.6	11.5	0.0010
Average 95%-posterior int.	[0.120,0.230]	[19.7,23.0]	[10.0,13.5]	[0.0009,0.0013]
Cov. of true value by 95% int.	89	53	56	90
Series 4				
True value	0.150	100.0	2.50	0.0010
Average posterior mean	0.153	93.7	2.72	0.0010
Average 95%-posterior int.	[0.130,0.181]	[74.5,117.4]	[2.15,3.37]	[0.0008,0.0013]
Cov. of true value by 95% int.	93	81	86	95

Real data

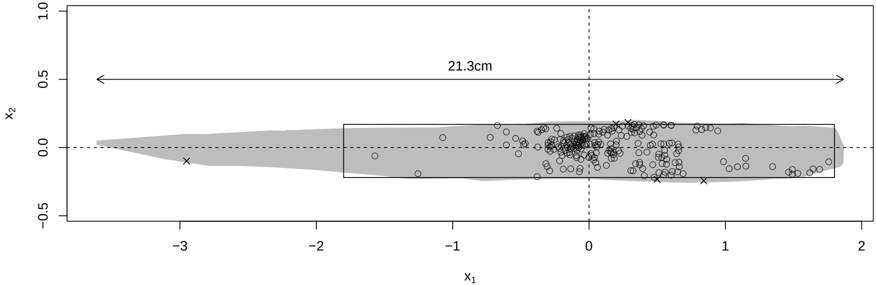


Figure: Locations (open circles and crosses) of brown rust lesions on a wheat leaf (grey shape). The pathogen lesions were formed after the dispersal of spores emitted from one mother lesion located at the intersection of the two dashed lines. The estimation algorithm was applied to the point pattern within the rectangle (open circles; 224 points), whereas the lesions outside the rectangle (crosses; 5

Real data - test of independent dispersal

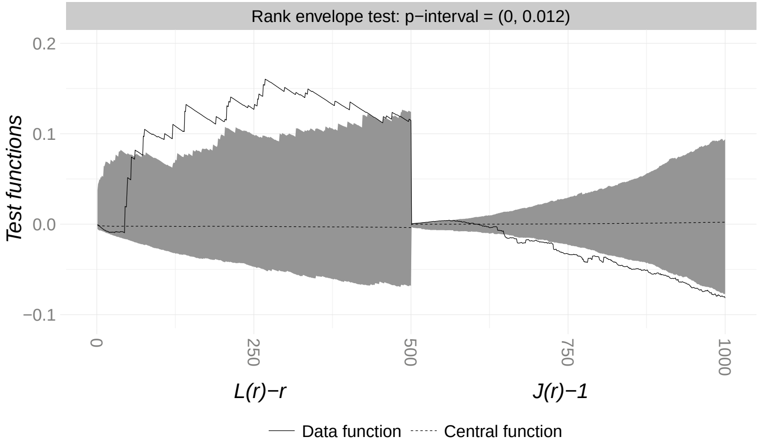


Figure: Rank envelope test (Myllymaki et al. 2015, Mrkvicka et al. 2015) of independent dispersal applied with the inhomogeneous *L*-function over the interval of distances [0.005, 0.3] and the inhomogeneous *J*-function over the interval of distances [0.005, 0.1].

Real data - estimation

- The MCMC chain was performed with 5 000 000 steps,
- first 150 000 steps were discarded as burn-in,
- every 10th step were used for calculation of posterior characteristics and
- the extended window was $W_{\text{ext}} = [-1.90, 1.90] \times [-0.32, 0.27]$.

Parameter	β	λ	μ	ν
Posterior mean	0.45	162	5.87	0.0081
Posterior median	0.44	155	5.65	0.0078
95%-posterior interval	[0.31,0.71]	[86,273]	[3.2,10.0]	[0.005,0.014]

Real data - estimation

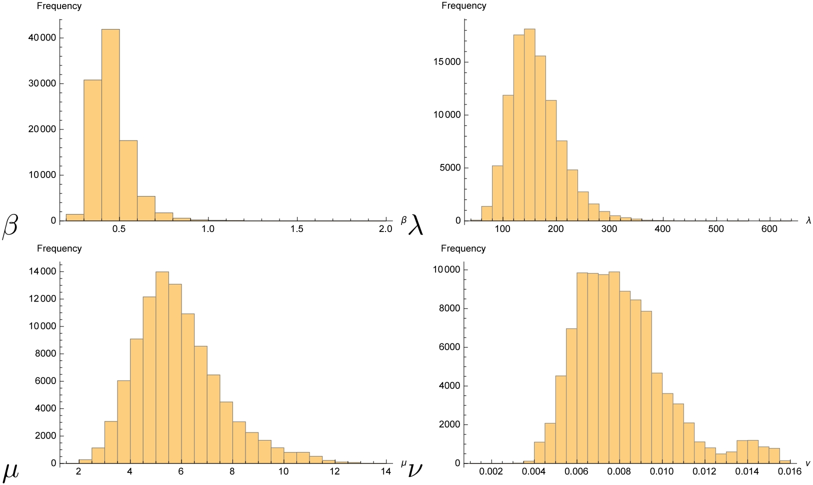


Figure: Histograms of estimated posterior distributions of model parameters.

Real data - estimation

- β= 0.45, implies that the mean dispersal distance of group centers is about 0.22, i.e. about the half of the maximum width of the leaf. Thus, a non-negligible amount of spores dispersed at the scale of the leaf is lost and/or is deposited on other leaves.
- μ=5.87, indicates that, at the leaf scale, groups of lesions with posterior mean size equal to 5.87 are formed.
- ν=0.0081 indicates that unidimensional standard deviation $\sqrt{\nu||c||}$ of the distances between the points of a group and their group center is about 0.06 at a distance $||c|| = 0.5$ from the mother lesion, 0.09 at a distance $||c|| = 1.0$, and 0.11 at a distance $||c|| = 1.5$.

Real data - test of group dispersal

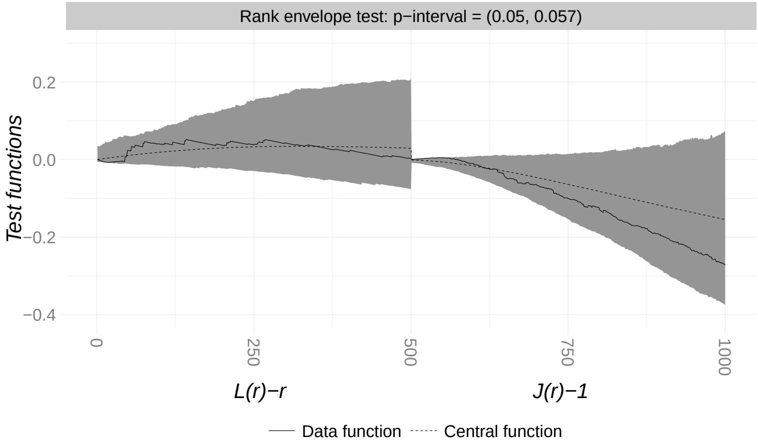


Figure: Rank envelope test for the estimated group dispersal model applied with the inhomogeneous *L*-function over the interval of distances [0.005, 0.3] and the inhomogeneous *J*-function over the interval of distances [0.005, 0.1].

Thank you for your attention.

References.

- Myllymaki M., Mrkvicka T., Seijo H., Grabarnik P., Hahn U.: Global envelope tests for spatial processes, JRSSB, 2016.
- Mrkvicka T., Myllymäki M., Hahn U.: Multiple Monte Carlo Testing with Applications in Spatial Point Processes, submitted to Statistics and computing http://arxiv.org/abs/1506.01646, 2016.