Ecological applications of spatial point process theory – examples of spatial complexity

Janine Illian

CREEM, University of St Andrews, Scotland

Lisbon, September 2010

overview

- biodiversity and spatial point processes
- measuring biodiversity
- understanding biodiversity
- conserving biodiversity

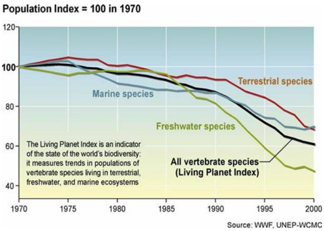
biodiversity

a definition...

"(...) the variability among living organisms from all sources including, inter alia, terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part ; this includes diversity within species, between species and of ecosystems"
(United Nations Environment Programme 1992)

biodiversity and ecosystem functioning

- increasing anthropogenic impact on ecosystems
- recent decades have seen an increasing threat to (species) biodiversity worldwide
- one of the key agreements at the 1992 Earth Summit in Rio de Janeiro is the convention on biological diversity



aim

"achieve by 2010 a significant reduction of the current rate of biodiversity loss"

biodiversity and ecosystem dynamics

- increased effort to maintain diversity or stop decline
- ecologists seek to understand community dynamics
  - how can large numbers of species co-exist ?
  - do depauperate systems function differently from communities with higher diversity ?
  - loss of biodiversity = loss of functionality ?

individual based models

- coexistence is concerned with **inter-individual interactions**
- interactions take place in a spatial context
- interact mainly with immediate neighbours due to limited mobility

$\Rightarrow$ modelling from the individuals' perspective has become popular in (plant) ecology :

- spatial birth and death processes
- dynamic models
- mechanistic models...

spatial point process methods

- describe and model the patterns formed by the locations of objects in space
- models capture the characteristics of spatial point patterns in a finite number of parameters ;
- simplest model is **Poisson process** ("complete spatial randomness")
  - random number of points
  - points are independent

- spatial point pattern is a **realisation** from a spatial point process (= random variable)
- the locations of any objects can be modelled – plants, animals, stars, cells, cities...

spatial point process methods

- many motivating examples for spatial point processes derived from ecology
- **but :** rarely used to directly answer topical ecological questions, e.g. related to biodiversity

aim

- *use existing spatial point process methods in the context of biodiversity research*
- *develop methodology that is suitable for this purpose*

It is essential to

⇒ be able to measure biodiversity

- develop measures of spatial (or local) biodiversity

⇒ understand biodiversity

- describe processes that organise ecosystems sustain biodiversity

⇒ preserve biodiversity

- understand suitable habitat for individual species

spatial biodiversity

measures of biodiversity : consider diversity of the ecosystem as a whole and typically ignore local spatial structure

aim

assess biodiversity by taking on the **individuals' perspective**

- *characterise the neighbourhood of each individual in terms of species diversity*
- *construct measures of (spatial) biodiversity*

joint work with T. Rajala

aspects of biodiversity

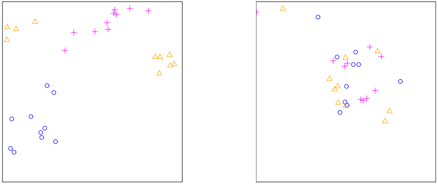
spatial biodiversity : two aspects

scattering : spatial distribution of individual species

exposure : spatial mixing of all species

- scattering : clustered versus regular
- exposure : mingled versus segregated

example : segregated + clustered versus mingled + clustered

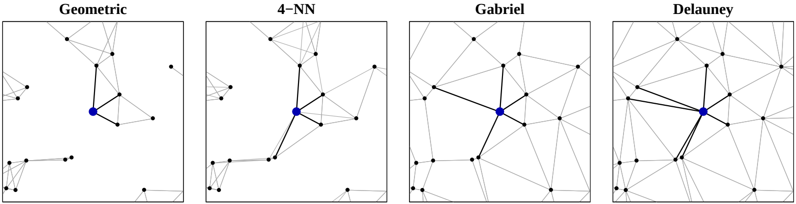


neighbourhood structure

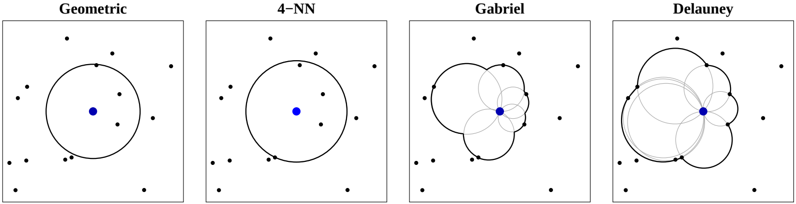
associate spatial pattern with a graph $G = (N, E)$

- locations are **vertexes** ; vertex set N
- points are connected by **edges** to indicate neighbourhood relationship ; edge set E
- different "rules" for connecting pairs of points generate different types of graphs

graphs and neighbourhoods



graphs and neighbourhoods



idea

Use graph theoretical framework to define a general and unified class of **spatial biodiversity measures***.*

generalisation

Both aspects of spatial biodiversity can be expressed in terms of graphs :

⇒ existing measures are defined using geometric graph (fixed distance) ; scale dependent

⇒ generalise current approaches using different neighbourhoods

scattering

- Ripley's K-function
- pair correlation function

exposure

- mingling index
- spatial Simpson index...

generalisation

Both aspects of spatial biodiversity can be expressed in terms of graphs :
$\Rightarrow$ existing measures are defined using geometric graph (fixed distance) ; scale dependent
$\Rightarrow$ generalise current approaches using different neighbourhoods

exposure

- mingling index
- spatial Simpson index...

notation

$N = \bigcup_{\tau=1}^{S} N_\tau$ is a multitype point pattern with S different species. For a graph $G = (N, E)$, a point $x \in N$ and $\tau \in \mathcal{S} = \{1, \ldots, S\}$ let

$$\deg(x) := \sum_{y \in N \setminus \{x\}} \mathbf{1}(x \to y)$$

be the number of emanating edges (**degree** of x) and denote the mean degree across all points as $\overline{\deg}$. Let

$$\pi_\tau := \frac{\overline{\deg_\tau}}{\overline{\deg}}$$

denote the mean proportion of type τ neighbours amongst all neighbours of the typical point.

mingling index

For a type τ, consider $\tilde{M}_\tau$, the fraction of points different from type τ among all the individuals' neighbours

$$\tilde{M}_\tau := \frac{\overline{\deg}_{\tau,\neq}}{\overline{\deg}_{\tau,\neq} + \overline{\deg}_{\tau,=}}$$

Then define a mingling index as

$$M := \frac{1}{S} \sum_{\tau=1}^{S} M_\tau .$$

Other similar measures, including a spatial Simpson index and a spatial Shannon index may be considered.

results

- simulation study shows that the choice of the neighbourhood (= graph) has an effect on the performance of the measures
- k-nearest neighbour graph best separates exposure from scattering

- when the geometric graph is used with real data (rainforest), local biodiversity appears to decreased over the years
- however, there is a bias since the graph also captures the scattering effect
- with the k-nearest neighbour graph, local biodiversity **increases** over the years !

conclusions

- graph theory provides a general framework for spatial biodiversity indices to account for different types of neighbourhoods
- naively using an unsuitable neighbourhood structure can produce biased and misleading results

complexity = large choice of neighbourhoods = choice of graph

understanding biodiversity

- reducing biodiversity loss requires an understanding of the mechanisms that **maintain** biodiversity
- studies have focused on highly diverse communities such as tropical rainforests and other biodiversity hot spots
- understand ecological mechanisms that allow large numbers of species to coexist in these communities

joint work with G. Evans and R. King

understanding biodiversity

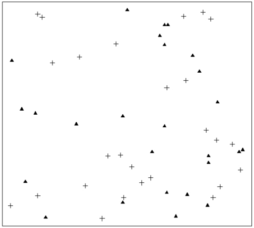
- coexistence concerns **inter-individual interactions**
- large numbers of species
  $\Rightarrow$ large numbers of potential interactions
  $\Rightarrow$ large numbers parameters
- use methods of **model comparison** to decrease complexity

data set

- biodiverse plant community consisting of a total of 67 plant species from Cataby, Western Australia
- spatial locations of more than 3000 plants in a 22×22 metre plot
- individual plants have existed in the same location for a very long time
- nutrient and water levels in the soil may be considered homogeneous

approach

- model interactions within the plant community using a spatial point process model ; a spatial **Gibbs process**
  - random number of points
  - points **not** independent (interaction)
- no need to include environmental conditions in the model
- initially consider two species (*Banksia menziesii* and *Banksia attenuata*)



Gibbs process approach

Let ξ and ϕ represent the two point patterns. Set $\xi = \{\xi_1, \dots, \xi_{n_1}\}$ and $\phi = \{\phi_1, \dots, \phi_{n_2}\}$.
Consider an interaction function $s(\xi|\phi)$

$$s(\xi|\phi) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} h(\|\xi_i - \phi_j\|),$$

where $\|\xi_i - \phi_j\|$ is the Euclidean distance between ξ_i and ϕ_j.
Choose h as :

$$h(r) = \begin{cases} (1 - (r/R)^2)^2 & \text{if} \quad 0 < r \leq R \\ 0 & \text{otherwise,} \end{cases}$$

where R is a fixed interaction radius.

Gibbs process approach

The (pseudo)likelihood can be expressed as a function of the intensity parameters $\beta = \{\beta_1, \beta_2\}$ and the interaction parameters $\gamma = \{\gamma_{11}, \gamma_{22}, \gamma_{12}\}$

$$f(\upsilon, \theta) = \alpha \beta_1^{n_1} \beta_2^{n_2} \gamma_{11}^{s(\upsilon_1|\upsilon_1)} \gamma_{22}^{s(\upsilon_2|\upsilon_2)} \gamma_{12}^{s(\upsilon_1|\upsilon_2)},$$

α is an intractable normalising constant.

- saturated model contains all possible interactions γ_{11}, γ_{22} and γ_{12}
- submodels can be defined as combinations of the presence or absence of the three interaction parameters
- in total there are eight possible models

Bayesian approach

- adopt a Bayesian approach to obtain inference on the model parameters
- joint posterior distribution of the parameters is formed by combining the likelihood of the data with the corresponding prior distribution of the parameters
- the posterior distribution of the saturated model can be written as :

$$\pi(\theta|\upsilon) \;\; \propto \;\; f(\upsilon|\theta)p(\theta),$$

where $p(\theta)$ denotes the prior on the model parameters

Bayesian approach

- obtain posterior estimates of interest using a Markov chain Monte Carlo (MCMC) algorithm
- additional model uncertainty ; extend the posterior distribution to incorporate this uncertainty

$\Rightarrow$ treat the model itself as a discrete parameter

- form the joint posterior distribution over both parameter and model space

Bayesian approach

To explore the posterior distribution and to obtain posterior summary statistics, use a Reversible Jump MCMC approach :

Step 1. Update the parameters, conditional on the model, using the Metropolis Hastings algorithm.

Step 2. Update the model itself using a reversible jump step.

Metropolis Hastings algorithm used in the first step is a standard random walk Metropolis update

results

- model with the highest posterior probability corresponds to the saturated model
- all models with non-negligible support contain the between species interaction parameter, γ_{12}

$\Rightarrow$ very important factor in the underlying dynamics

- posterior probabilities for γ_{11} and γ_{22} are 0.849 and 0.834
- all interactions are negative (= repulsive)

conclusion

- model comparison used to avoid over-parametrisation
- model comparison difficult in frequentist approach ; very approximate
- Bayesian approach combined with reversible jump MCMC provides elegant methods for model comparison
- extend to more than two species ; running time increases...

complexity = large number of connections = re-weighting edges

preserving biodiversity

- in ecological applications, the patterns are often large and data-collection time-consuming and expensive
- example : Malaysian rainforest data ; 320903 stems from 817 species...
- often only possible to collect data on a single spatial pattern
- majority of practical applications in ecology analyse individual patterns

BUT :

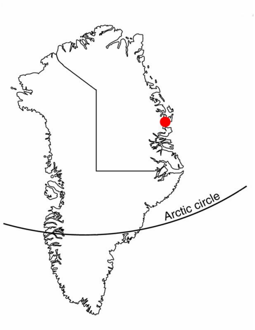
- sometimes the accessible observation area is too small and a single spatial pattern may contain insufficient information
- common in the context of animal ecology or for rare species

joint work with R. King and S. King

motivation

- often it is possible to collect point pattern data at repeated points in time, where the observed patterns may be considered independent given the environment
- aim is to fit a **single model** to replicated patterns to gain a better understanding of the spatial behaviour of the observed individuals
- **here :** emphasis on analysis of replicated patterns taking differences between replicates into account

muskoxen data



Zackenberg research station, east Greenland

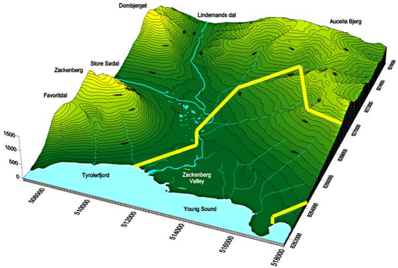
- spatial locations of muskoxen herds recorded at different time points throughout different years
- average number of points in year 2005 20.8, sd = 10.34 (min = 8)
- habitat choice : does the spatial distribution depend on habitat quality ?

muskoxen data

Zackenberg research station, east Greenland

- spatial locations of muskoxen herds recorded at different time points throughout different years
- average number of points in year 2005 20.8, sd = 10.34 (min = 8)
- habitat choice : does the spatial distribution depend on habitat quality ?

an example – replicated spatial point patterns



- animals move in and out of the observation area
- habitat changes over time (e.g. snow melts)
- group interactions change over time (e.g. in mating season)
- fit one "big" model to all time points together

⇒ treat "time" as random factor and habitat (and group type) as fixed factors in a spatial point process model

- animals move in and out of the observation area
- habitat changes over time (e.g. snow melts)
- group interactions change over time (e.g. in mating season)
- fit one "big" model to all time points together

⇒ treat "time" as random factor and habitat (and group type) as fixed factors in a spatial point process model

parameter estimation

- previous work took frequentist approach, applying the Berman-Turner device ; approximate the integral in the pseudolikelihood by a finite sum
- (approximate) estimates based on software for generalised linear mixed models using penalised quasi-likelihood
- here : **Bayesian approach**

Strauss process

Observe the spatial point pattern $\mathbf{x}_i$ at time points $i = 1, \ldots, T$ within region $W \subset \mathbb{R}^2$. The overall data set is denoted by $\mathbf{x} = \{\mathbf{x}_1, \ldots, \mathbf{x}_T\}$, a series of spatial point patterns, corresponding to the different time points.

Consider a Strauss process for point pattern $i = 1, \ldots, T$ with parameters :

β_i - the intensity ; and

γ_i - the interaction.

We have that $\beta_i > 0$ and $\gamma_i \in [0, 1]$; small values of the interaction term corresponds to strong inhibition.

The interaction radius, r, is assumed to be known.

Strauss process

Include a random effects component to the Strauss process by

$$\begin{aligned}\log \beta_i &= \theta_1 + \phi_{1,i} \\ \text{logit } \gamma_i &= \theta_2 + \phi_{2,i},\end{aligned}$$

where,

$$\begin{aligned}\phi_{1,i} &\sim N(0, \sigma_1^2) \\ \phi_{2,i} &\sim N(0, \sigma_2^2).\end{aligned}$$

Set $\boldsymbol{\theta} = \{\theta_1, \theta_2\}$, $\boldsymbol{\phi}_1 = \{\phi_{1,1}, \ldots, \phi_{1,T}\}$, $\boldsymbol{\phi}_2 = \{\phi_{2,1}, \ldots, \phi_{2,T}\}$ and $\boldsymbol{\sigma}^2 = \{\sigma_1^2, \sigma_2^2\}$.
Take Bayesian approach and fit a random effects model using an **auxiliary variable approach**

Strauss process

$\boldsymbol{\theta}$ and $\boldsymbol{\sigma}^2$ are model parameters.
We also treat the random effect terms ϕ as parameters (or auxiliary variables) to be estimated (equivalent to a hierarchical prior within a Bayesian analysis). Form the joint posterior distribution of the model parameters and random effects terms.

$$\pi(\boldsymbol{\theta}, \boldsymbol{\sigma}^2, \phi | \mathbf{x}) \quad \propto \quad PL(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\sigma}^2, \phi) p(\boldsymbol{\theta}) p(\phi | \boldsymbol{\sigma}^2) p(\boldsymbol{\sigma}^2),$$

where $PL(\mathbf{x} | \boldsymbol{\theta}, \boldsymbol{\sigma}^2, \phi)$ denotes the pseudo-likelihood.
Conditional on all the parameter values, the pseudo-likelihood can be calculated.

Strauss process

The posterior distribution of the **model parameters** can be obtained by taking the marginal posterior distribution,

$$\pi(\boldsymbol{\theta}, \boldsymbol{\sigma}^2|\mathbf{x}) = \int \pi(\boldsymbol{\theta}, \boldsymbol{\sigma}^2|\mathbf{x}) d\phi.$$

Obtain a sample from the joint posterior distribution $\pi(\boldsymbol{\theta}, \boldsymbol{\sigma}^2, \phi|\mathbf{x})$ based on MCMC (single-update Metropolis-Hastings).

muskoxen data

the simple Strauss process is too simplistic for the **muskoxen data**

- there is some indication that the interaction is not repulsive but **attractive**
- use **area interaction process**
- pattern highly inhomogeneous ; use e.g. altitude and vegetation index as covariates, i.e. as a fixed effects

results

- results indicate that intensity and interaction vary strongly with time
- both altitude and vegetation index (and their interaction) highly influential

conclusion

- models with mixed effects can model (conditionally independent) replicated patterns if single patterns are too small
- Bayesian context particularly useful (e.g. with regard to **inference**)
- for muskoxen data a large number of further covariates is available ; reversible jump approaches for **model comparison**

complexity = replicated patterns = fixed and random components in graph

complexity occurs in many disguises

$\Rightarrow$ measuring spatial biodiversity

- individuals interact mainly with their neighbours
- How to choose the neighbourhood ?
- What graph ?

complexity occurs in many disguises

$\Rightarrow$ measuring spatial biodiversity

- individuals interact mainly with their neighbours
- How to choose the neighbourhood ?
- What graph ?

$\Rightarrow$ understanding biodiversity

- large numbers of (potential) intra- and inter-species interactions
- use methods for model comparison to choose appropriate model
- How to weight (or discard) different edges ?

complexity occurs in many disguises

⇒ measuring spatial biodiversity

- individuals interact mainly with their neighbours
- How to choose the neighbourhood ?
- What graph ?

⇒ understanding biodiversity

- large numbers of (potential) intra- and inter-species interactions
- use methods for model comparison to choose appropriate model
- How to weight (or discard) different edges ?

⇒ preserving biodiversity

- rare or protected species of particular conservation interest
- very small patterns (few points) but replicates exist
- need to account for variation across replicates
- random (and fixed) components in graphs

Complexity occurs in many disguises

- issues not restricted to ecology or biodiversity research
- similar issues with similar structural properties in many other contexts...

Acknowledgment :
I would like to thank the Zackenberg Basic programmes, in particular BioBasis and GeoBasis who collected the muskoxen data for making the data available to me and in particular Ditte Katrine Henrichsen for introducing me to the data and many useful discussions.