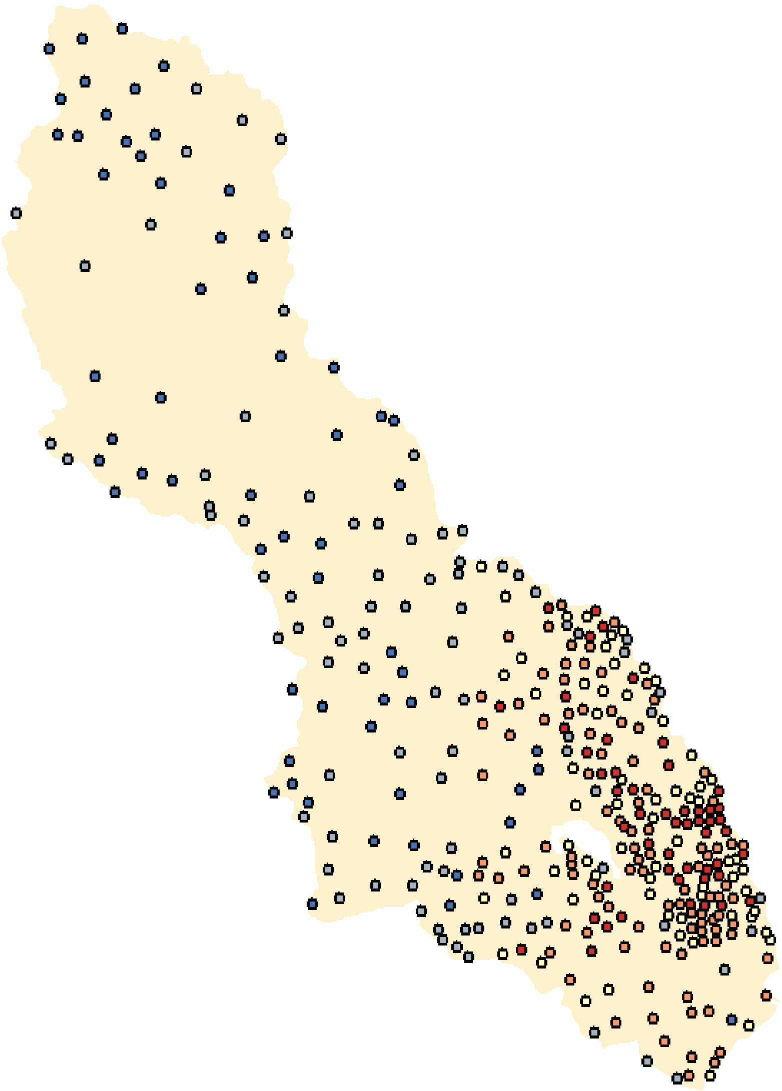
Spatial Point Pattern Analysis

Jiquan Chen
Prof of Ecology, University of Toledo

EEES6980/MATH5798, UT

LANDSCAPE ECOLOGY &
ECOSYSTEM SCIENCE

Point variables in nature



A point process is a discrete stochastic process of which the underlying distribution in not continuous, e.g.

- Distribution of feeding bark beetles;
- Tree distribution
- Inland lakes in terrestrial ecosystems;
- Cities at regional and/or continental scales;
- Students in a classroom;
- etc.

Introduction

- Point pattern analysis looks for patterns in the spatial location of events
  - ✓ "Events" are assigned to points in space
  - ✓ e.g. infection by bird-flu, site where firm operates, place where crime occurs, redwood seedlings

- Point pattern analysis has the advantage that it is not directly dependent on zone definitions (MAUP)

From Steve Gibbons

Types of Point data

- Univariate, Bivariate, Multivariate
- 1, 2, and 3 Dimensions (x,y,z)

From unknown

Point Pattern Analysis

- Pattern may change with scale!

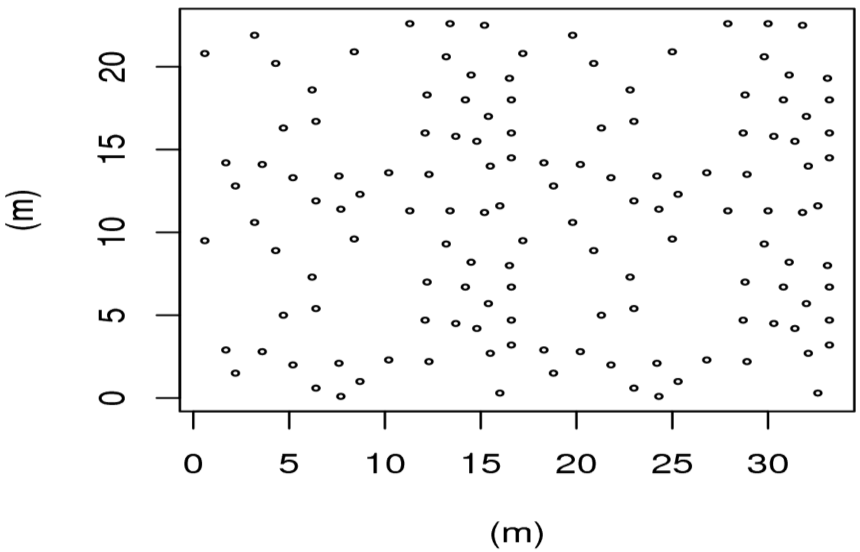
- Test statistic calculated from data vs. expected value of statistic under CSR (complete spatial randomness)

From unknown

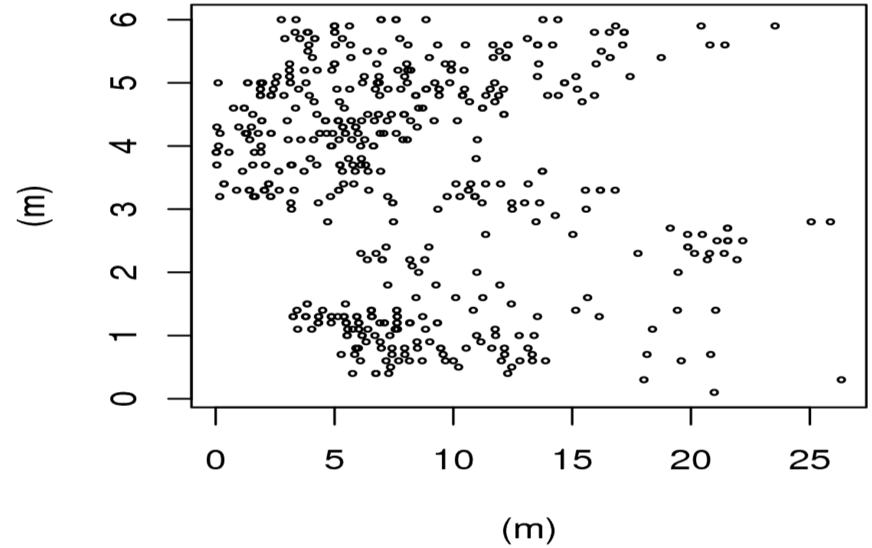
Types of Point Patterns

- Random (CSR)
- Overdispersed (spaced or regular)
- Underdispersed (clumped or aggregated)

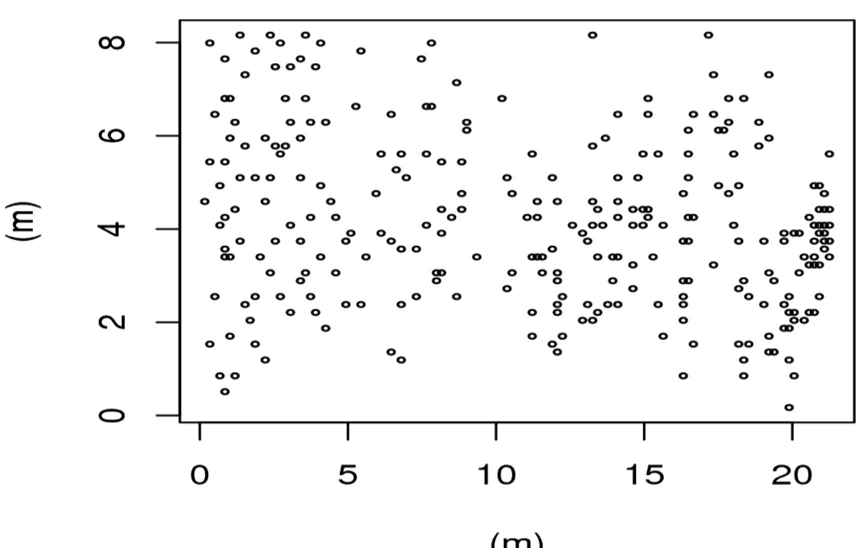
From unknown



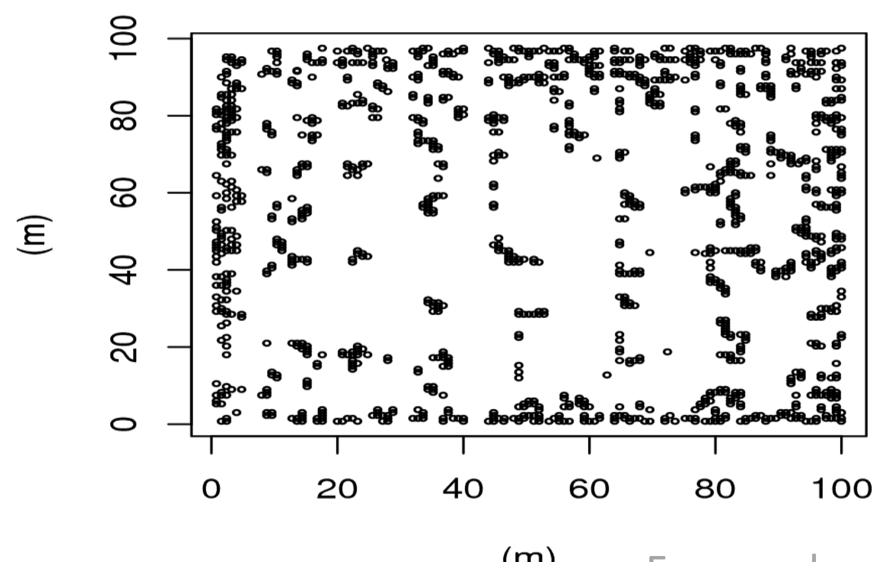
PG 93



Waterfowl



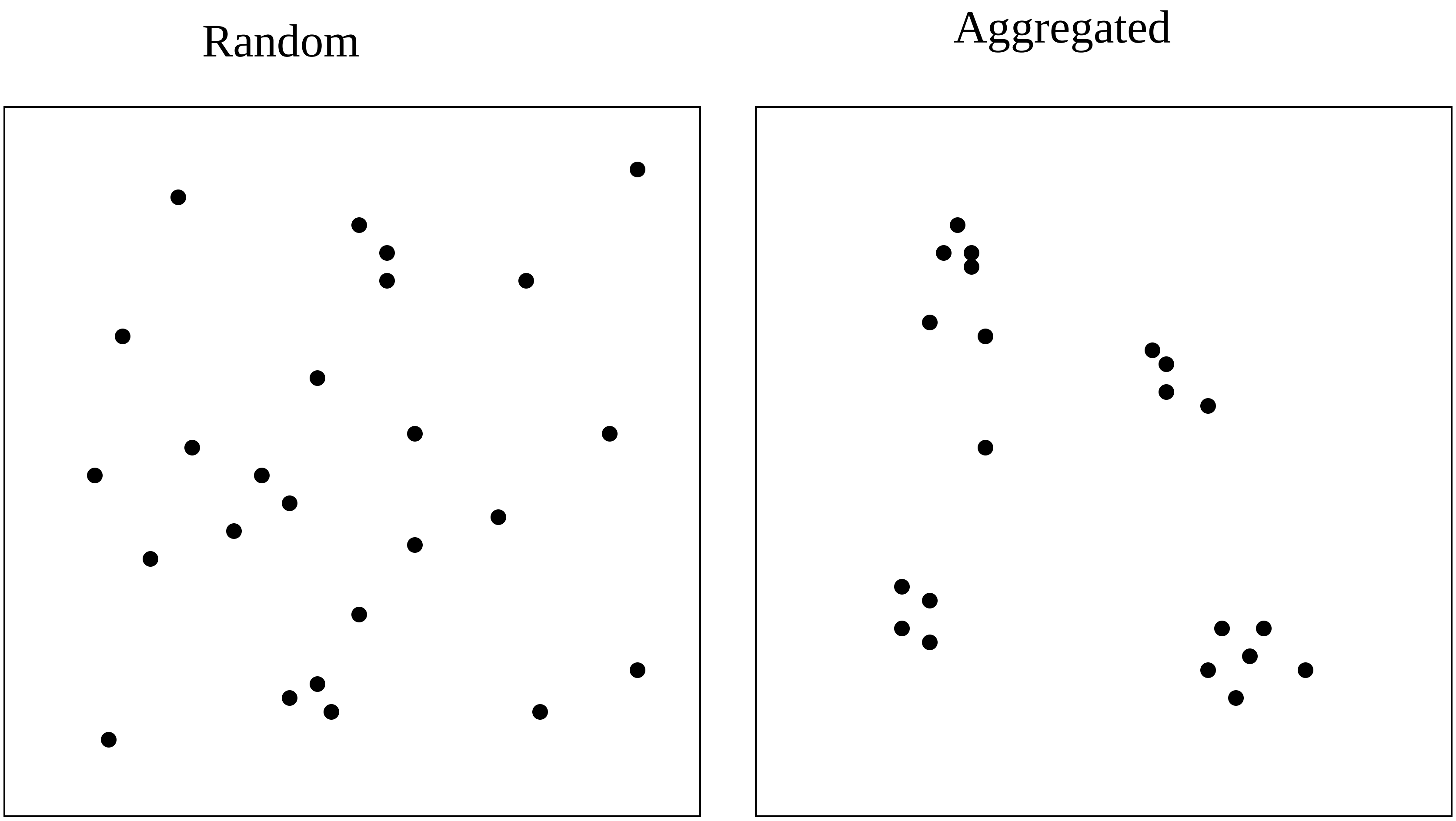
Rice 5



Corn 3

From unknown

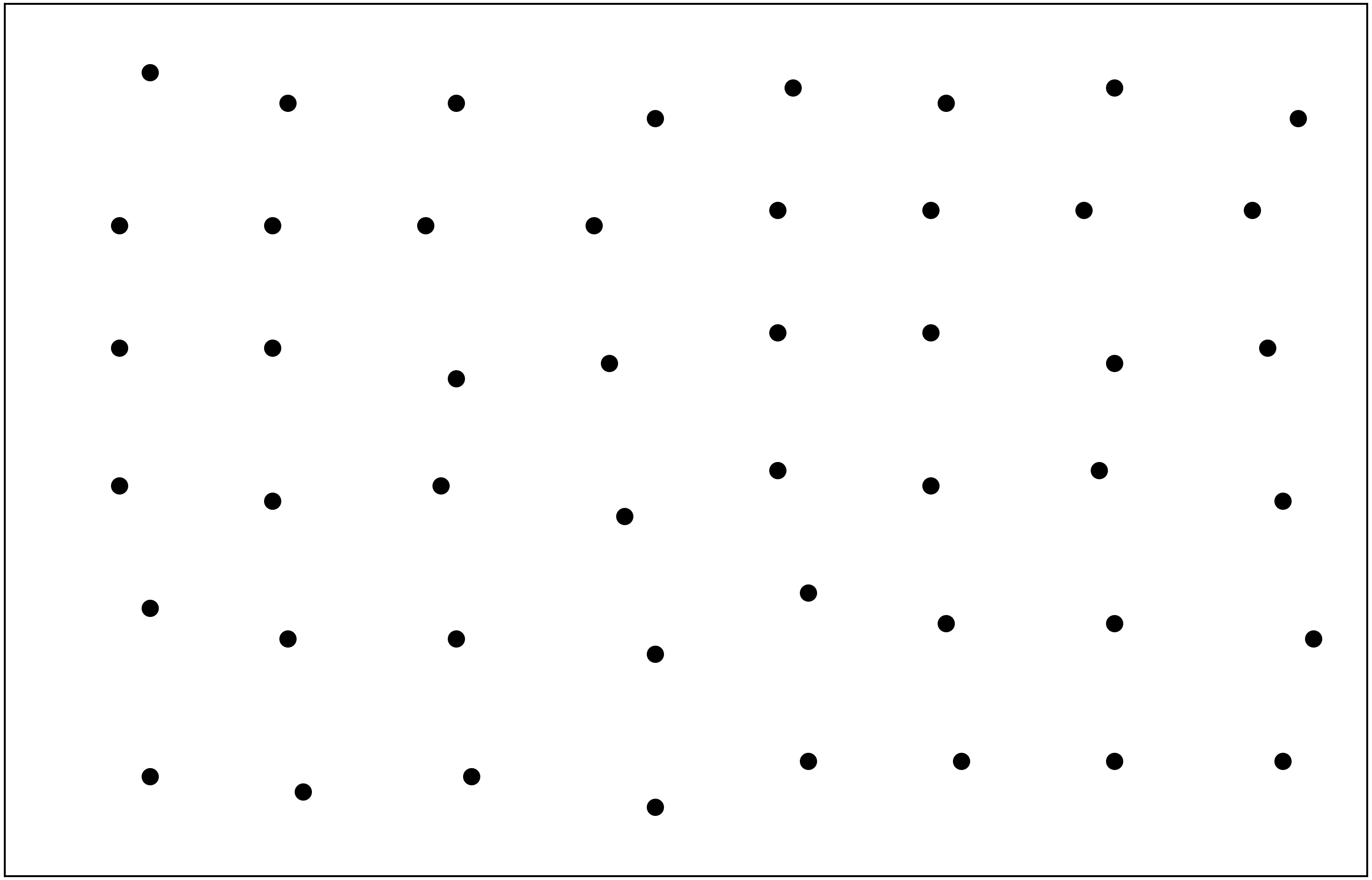
Spatial point patterns



From Steve Gibbons

Spatial point patterns

Regular



From Steve Gibbons

Spatial data exploration

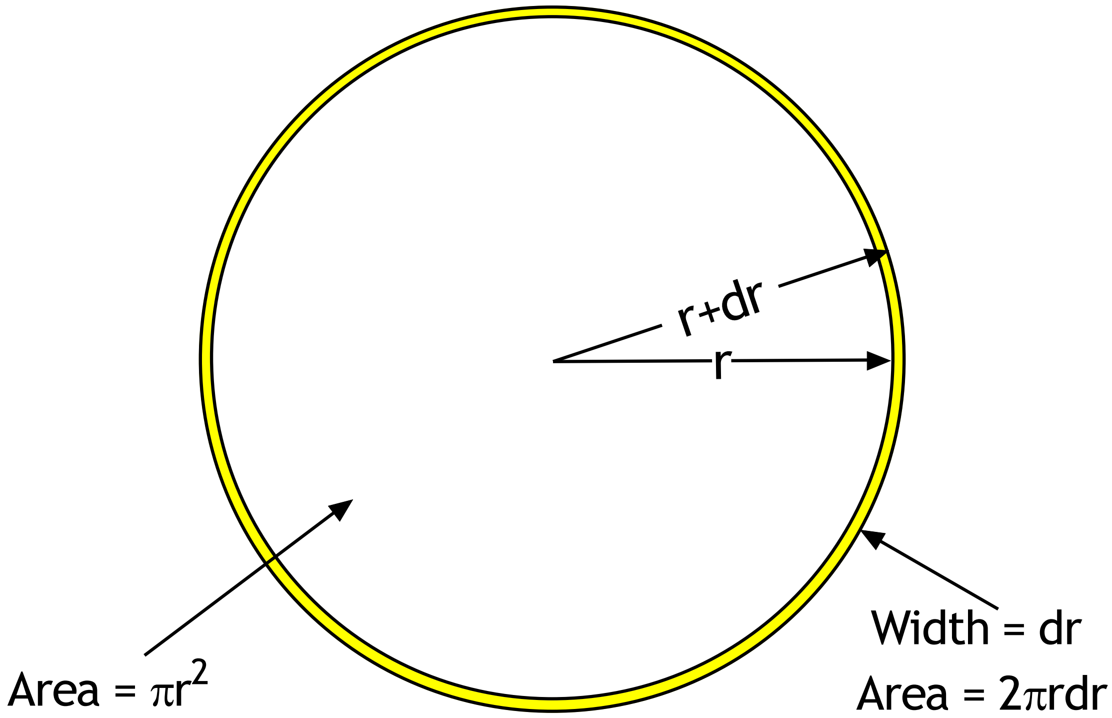
- Point (event) based statistics
  - Typically analysis of point-pair distances
  - Points vs events
  - Distance metrics: Euclidean, spherical, L_p or network
  - Weighted or unweighted events
  - Events, NOT computed points (e.g. centroids)
  - Classical statistical models vs Monte Carlo and other computational methods

Spatial data exploration

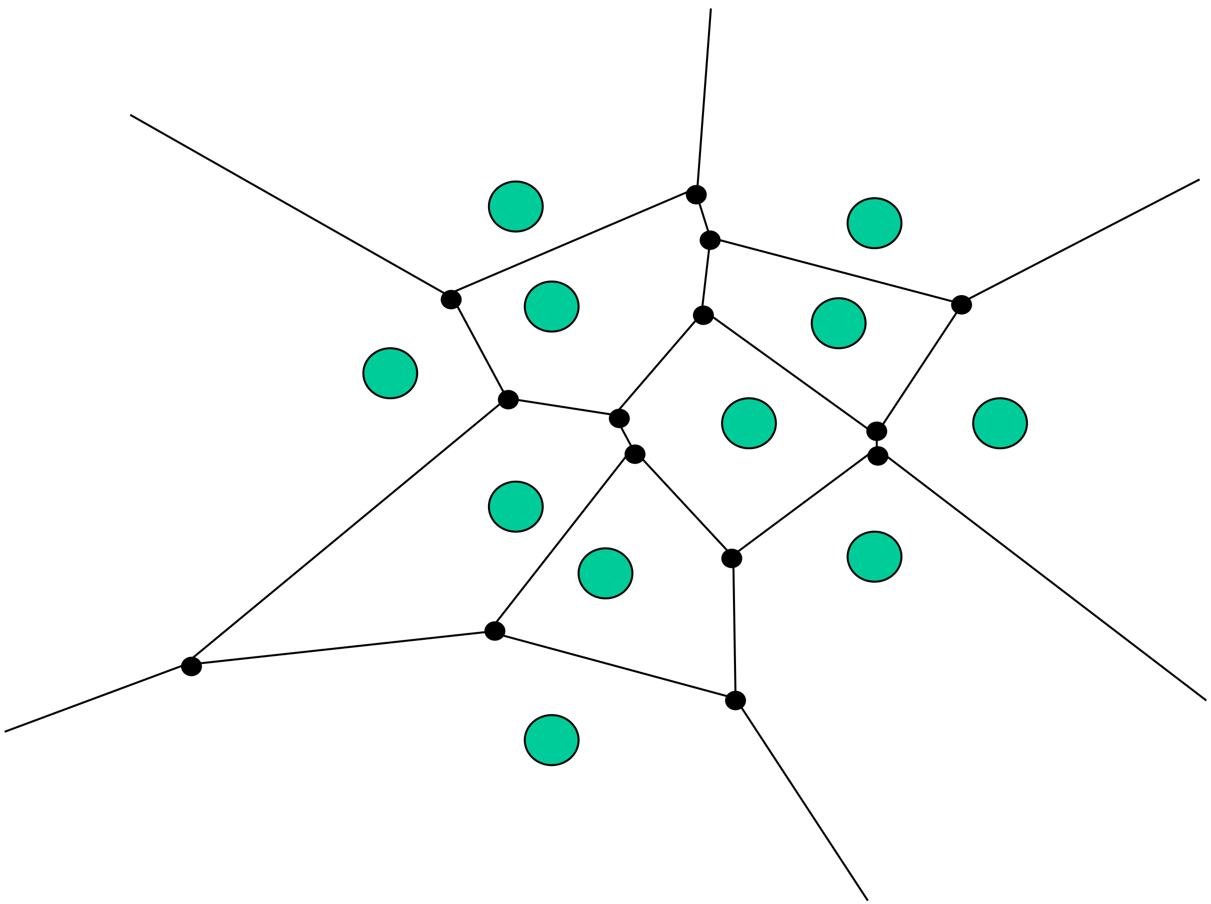
- Point (event) based statistics
  - Basic Nearest neighbour (NN) model
    - Input coordinates of all points
    - Compute (symmetric) distances matrix **D**
    - Sort the distances to identify the 1st, 2nd,...*k*th nearest values
    - Compute the mean of the observed 1st, 2nd, ...*k*th nearest values
    - Compare this mean with the expected mean under Complete Spatial Randomness (CSR or Poisson) model

Spatial data exploration

- Point (event) based statistics – NN model



Nearest Neighbor Search Structure



Spatial data exploration

- Point (event) based statistic s – NN model
  - Mean NN distance: $\mu = \dfrac{1}{2\sqrt{m}}$
  - Variance: $\mu_2 = \dfrac{(4-\pi)}{4\pi m}$
  - NN Index (Ratio): $R = \bar{r}_o / \bar{r}_e$
  - Z-transform: $z = (\bar{r}_o - \bar{r}_e)/\sigma_e \sim \mathrm{N}(0,1)$, where $\sigma_e = \mu_2/\sqrt{\mathrm{n}} = 0.261358/\sqrt{mn}$

Spatial data exploration

- Point (event) based statistics
  - Issues
    - Are observations n discrete points?
    - Sample size (esp. for k^{th} order NN, k>1)
    - Model requires density estimation, m
    - Boundary definition problems (density and edge effects) – affects all methods
    - NN reflexivity of point sets
    - Limited use of frequency distribution
    - Validity of Poisson model vs alternative models

Spatial data exploration

- Frequency distribution of nearest neighbour distances, i.e.
  - The frequency of NN distances in distance bands, say 0-1 km, 1-2 km, etc.
  - The cumulative frequency distribution is usually denoted
    - $G(d) = \#(d_i < r)/n$ where d_i are the NN distances and n is the number of measurements, or
    - $F(d) = \#(d_i < r)/m$ where m is the number of random points used in sampling

Spatial data exploration

- Computing G(d) [computing F(d) is similar]
- Find all the NN distances
- Rank them and form the cumulative frequency distribution
- Compare to expected cumulative frequency distribution:

$$G(r) = 1 - e^{-m\pi r^2}$$

- Similar in concept to K-S test with quadrat model, but compute the critical values by simulation rather than table lookup

Complete Spatial Randomness

- The simplest "null hypothesis" regarding spatial point patterns
- The number of events N(A) in any planar region A with area |A| follows a Poisson distribution with mean:

$$p(N(A)=n)=\frac{\left(\lambda|A|\right)^{n}}{n!}e^{-\lambda|A|} \qquad \lambda|A|$$

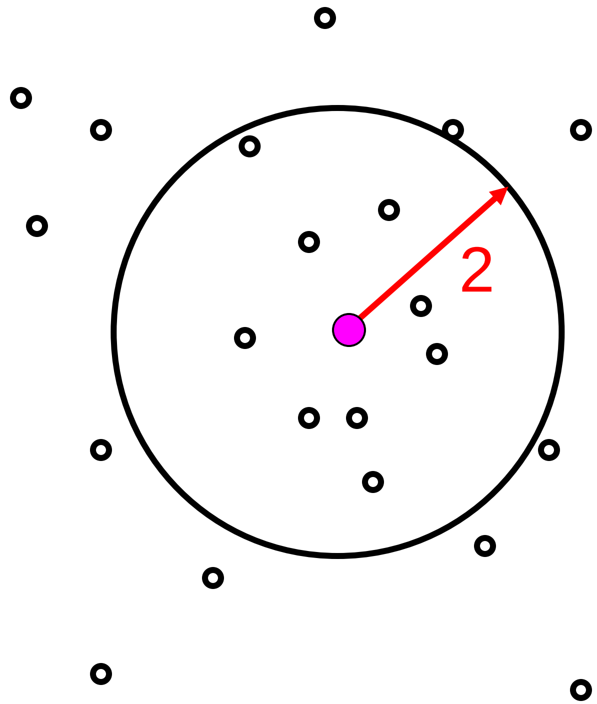
- Given N(A)= n, the events in A are an independent random sample from the uniform distribution on A
- Poisson process has constant "intensity"
- *Intensity* is the expected number of events per unit area λ
- Also mean = variance See Diggle p.47

From Steve Gibbons

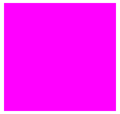
Kernel intensity/density estimates

- A simple kernel intensity estimate using a "uniform" kernel

$$\lambda(s) = \frac{N(C(s,r))}{\pi r^2}$$

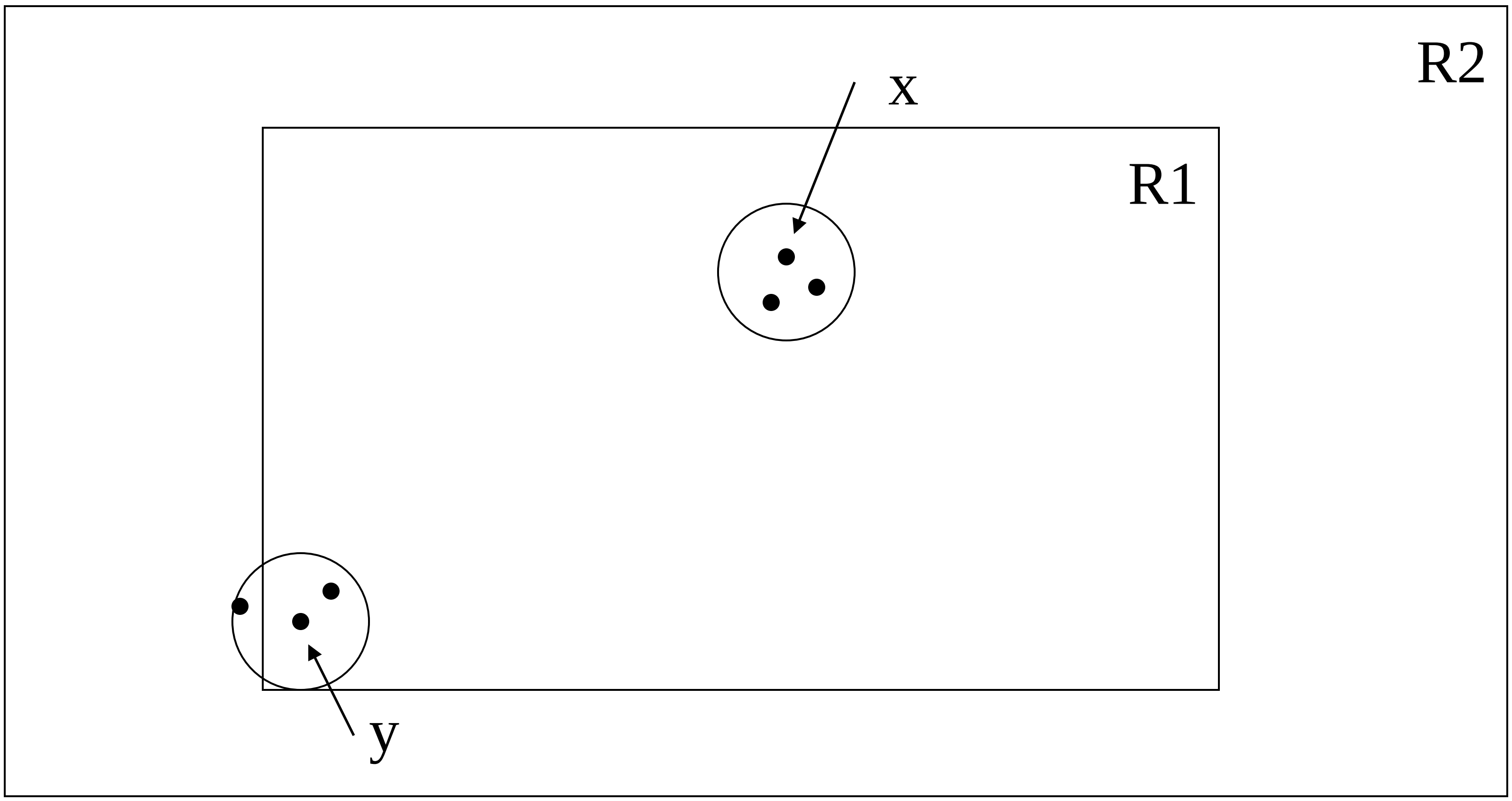


$$\lambda(s) = \frac{9}{\pi 2^2} = 0.716$$



From Steve Gibbons

Edge effects



From Steve Gibbons

Correcting Edge Effects

- Intensity estimated lower at point y than at point x

- Corrections can be based on
  - % area of circle within R1
  - % circumference of circle within R1
  - [circumference easier to calculate]
  - drawing buffer zones

From Steve Gibbons

K function

- The "K function" is the expected number of events within distance d of an event, divided by mean intensity in the study area (i.e. number of events/ area)

$$K(d) = \frac{E\left[N(d)\right]}{\lambda} \text{ or } \frac{E\left[C\left(N(x,d)\right)\right]}{\lambda}$$

From Steve Gibbons

Ripley's K

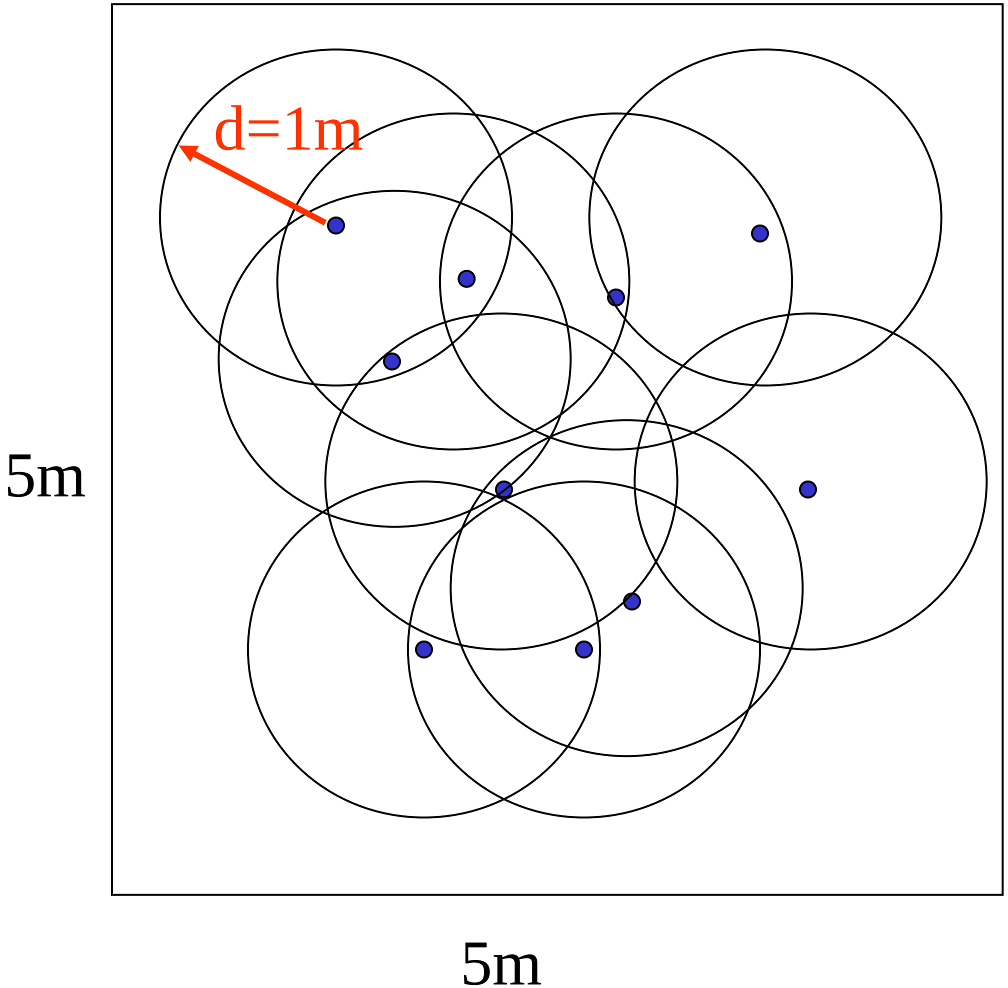
- Ripley's (1976) estimator of K

$$\hat{K}(d) = \frac{|A|}{n^2} \sum_{i=1}^{i=n} \sum_{j \neq i} I\left\{ \left\| s_i - s_j \right\| < d \right\}$$

- Where |A| means area of study area A, and $\left\| s_i - s_j \right\|$ means distance between s_i and s_j
- Also need to take care of edge effects
- If events uniformly distributed with intensity λ then expected number of events within distance d is $\lambda \pi d^2$
- So expected K(d) under uniform distribution (CSR) is πd^2

From Steve Gibbons

Ripley's K



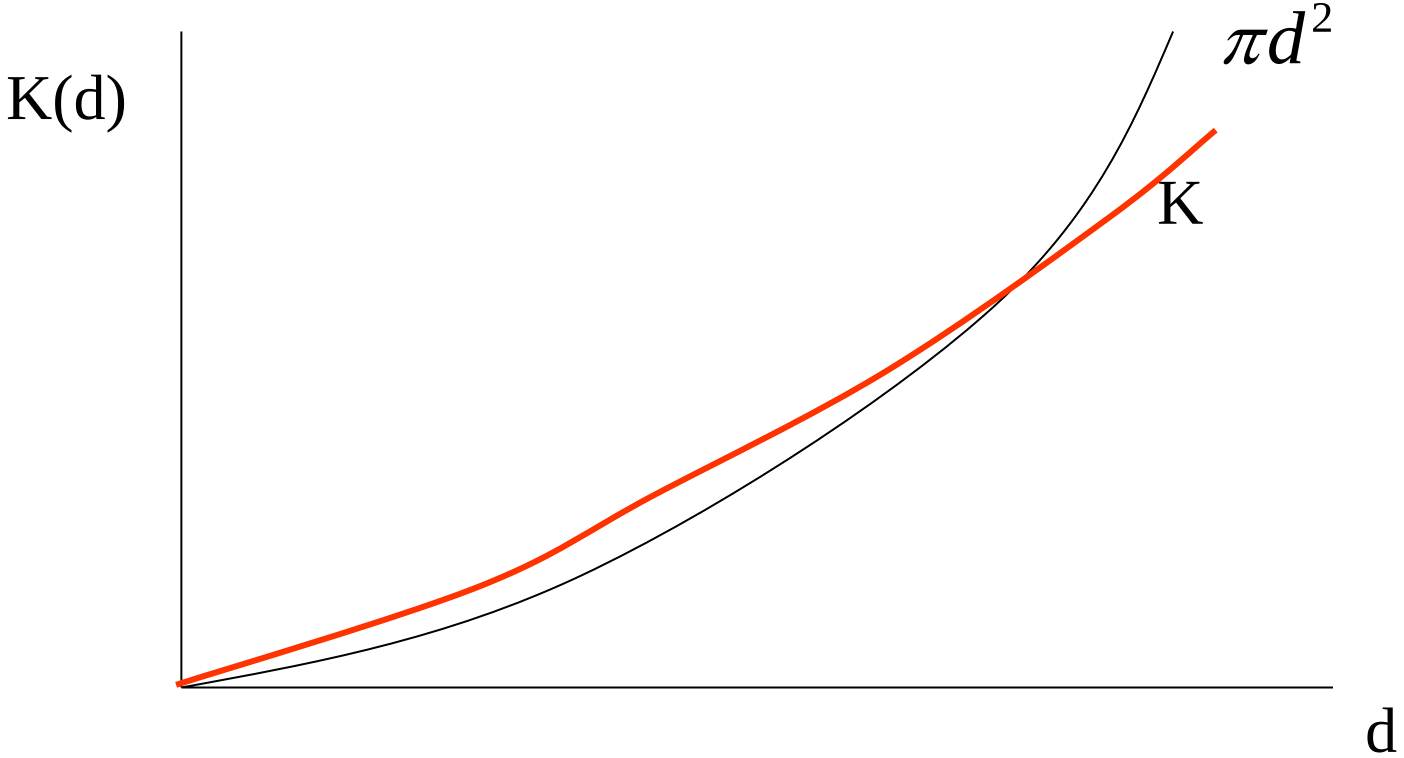
If uniform $K(1) = \pi = 3.14$

N(1)
2
3
2
1
2
1
0
2
1
2
$E[N(1)] = 16/10$
$\lambda = 10/25 = 0.4$
$K(1) = 1.6/0.4 = 4$

From Steve Gibbons

Checking for clustering

- Under CSR with uniform intensity expect $K(d) = \pi d^2$



From Steve Gibbons

Hypothesis tests

- Sampling distribution of these spatial point process statistics is often unknown
- Possible to derive analytical point-wise confidence intervals for kernel estimates
- But more generally use “monte-carlo”, “bootstrap” and random assignment methods

From Steve Gibbons

Methods

- Distance to neighbor
  - sample
- Refined Nearest Neighbor
  - randomization
- Second-order point pattern analysis

Second-order Point Pattern Analysis: Ripley's K

"Used to analyse the mapped positions of events in the plane… and assume a complete census…"

From unknown

ELSEVIER

Forest Ecology and Management 120 (1999) 219–233

Forest Ecology and Management

Forest structure in space: a case study of an old growth spruce-fir forest in Changbaishan Natural Reserve, PR China

Jiquan Chen[a,*], Gay A. Bradshaw[b]

[a]*School of Forestry and Wood Products, Michigan Technological University, Houghton, MI 49931, USA*
[b]*Forest Science Laboratory, USDA Forest Service, 3200 Jefferson Way, Corvallis, OR 97331, USA*

Received 8 September 1998; accepted 1 December 1998

sity index). We were concerned with three main questions: (1) How are various tree species distributed across the stand (e.g., random, clustered) and at what scales; and (2) Are these patterns consistent with observed species' functions within the stand, life histories, and inter-species interactions? Specifically, we sought to quantify the spatial distributions of trees of different species and height classes (i.e., sub-populations) and examine inter-species and intra-species size-class interactions in terms of vertical and horizontal canopy structure.

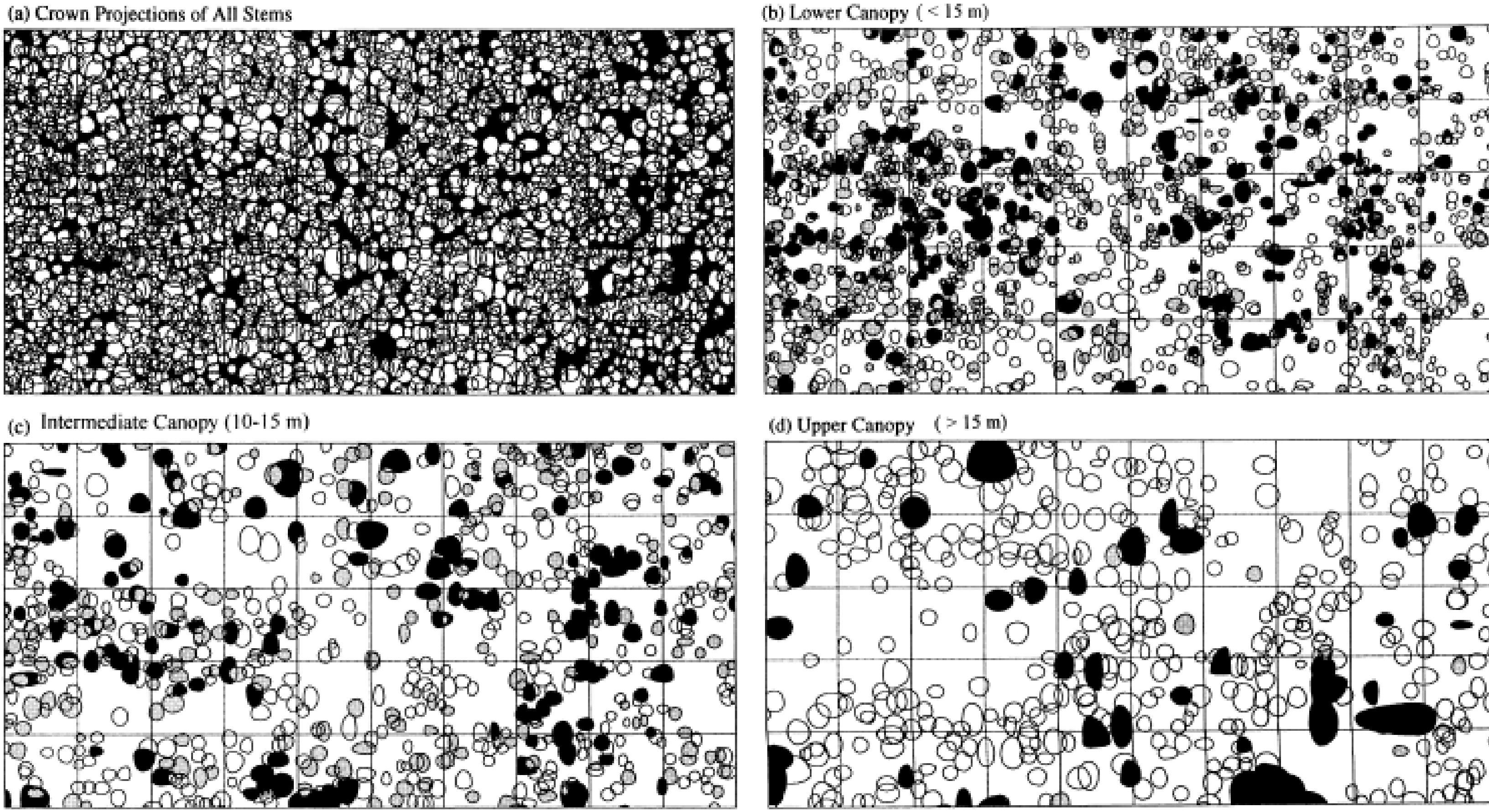


Fig. 6. Crown projections of trees in: (a) all three layers combined, (b) lower canopies, (c) intermediate canopies, and (d) upper canopies. Crowns of spruce, fir, birch, and *other* are shaded in (b), (c), and (d) as open, light-shaded, black, and heavy-shaded, respectively. Grid size is 20 m.

J. Chen, G.A. Bradshaw / Forest Ecology and Management 120 (1999) 219–233

Chen & Bradshaw 1999

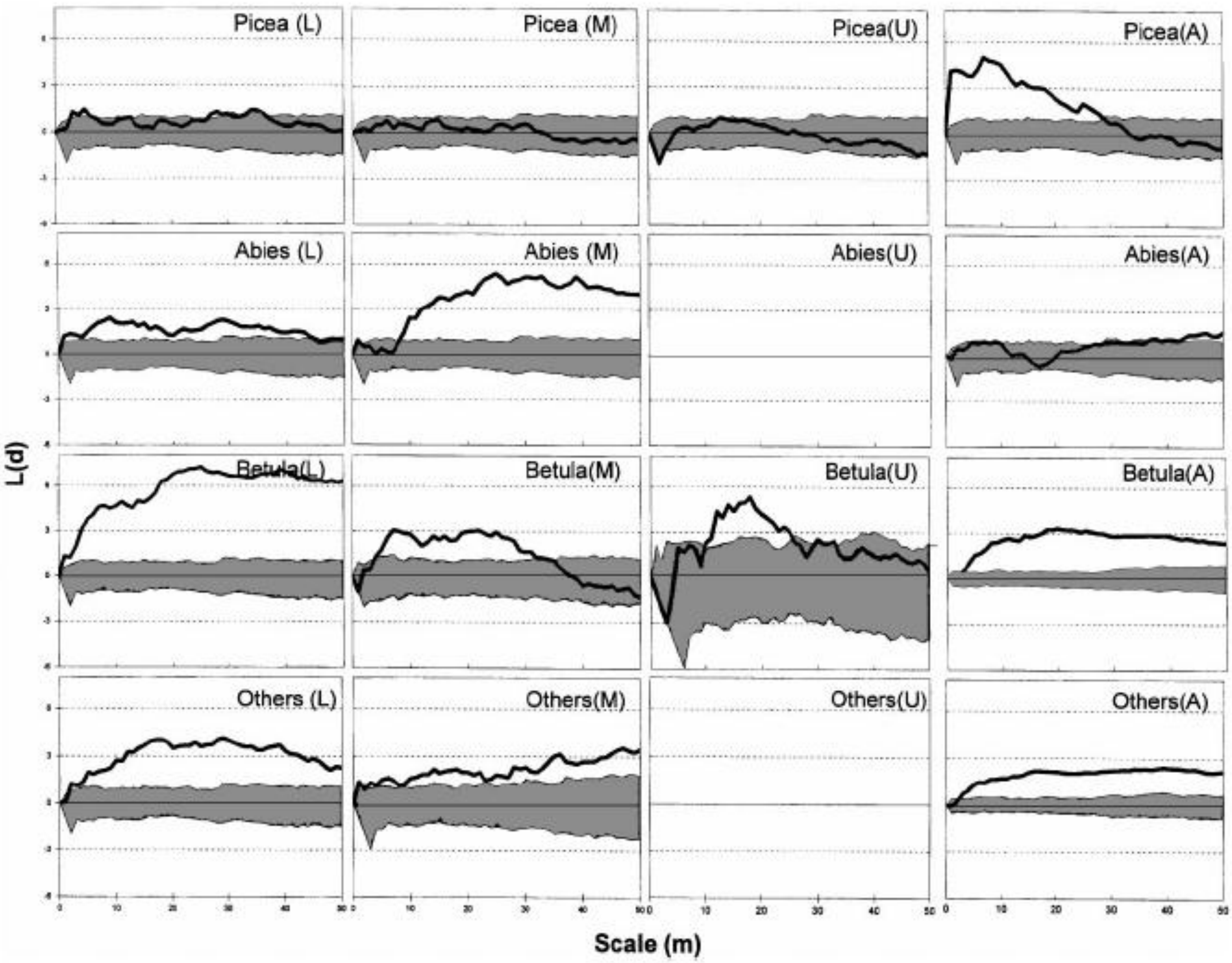
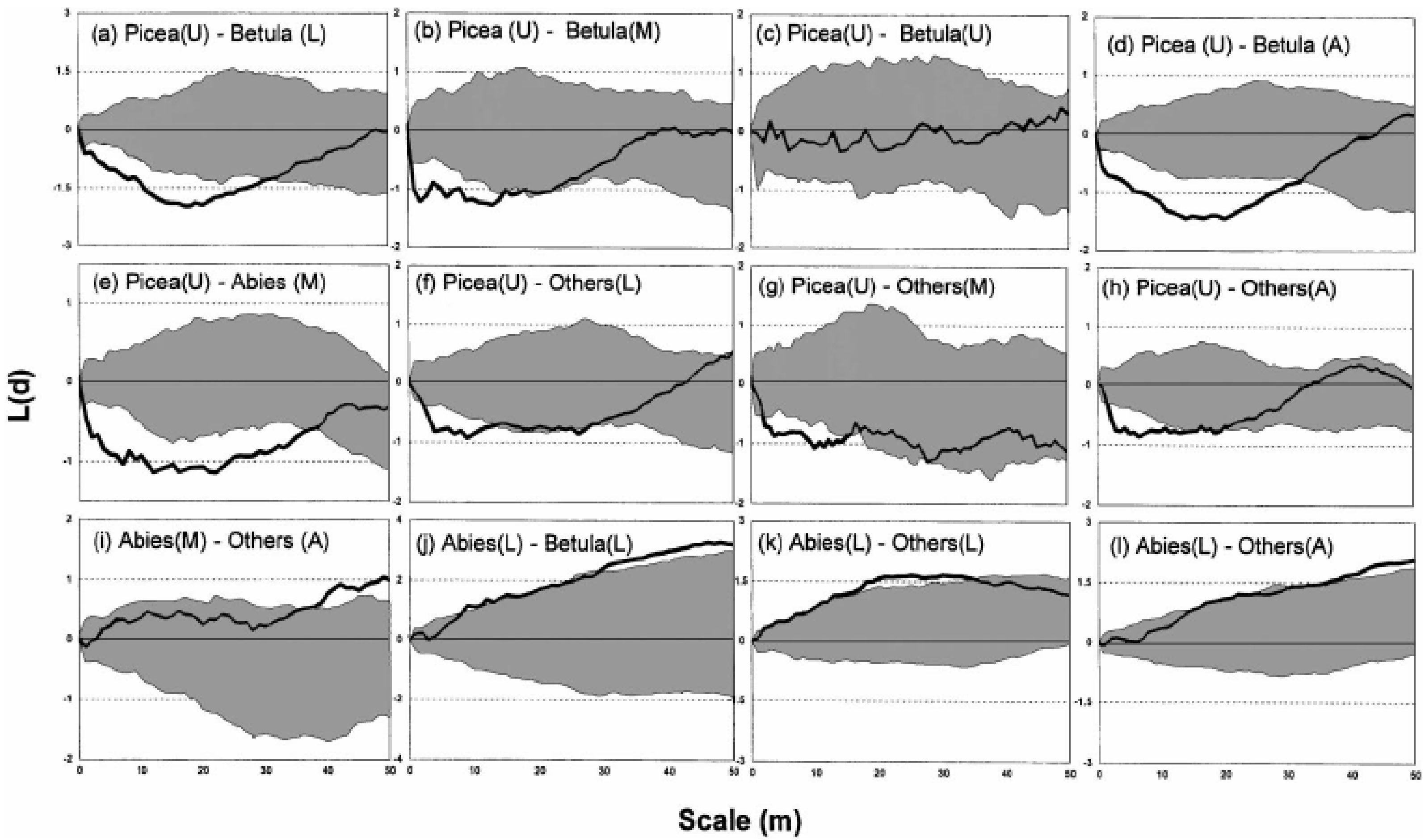


Fig. 3. Ripley's *K* for four tree species groups located in three canopy layers: L – lower layer (<10 m), M – middle or intermediate layer (10–15 m), and U – upper layer (>15 m). A indicates all stems of the species. The Monte Carlo envelope (shaded area) is constructed at the 95% confidence level. The square root transformation, L(*d*), of Ripley's *K* were applied (see Section 2).

J. Chen, G.A. Bradshaw / Forest Ecology and Management 120 (1999) 219–233

Chen & Bradshaw 1999

Bi-Ripley K



J. Chen, G.A. Bradshaw / Forest Ecology and Management 120 (1999) 219–233

Fig. 4. Bivariate Ripley's K between two point patterns showing significant (95%) attractive or repulsive responses. L – lower layer (<10 m), M – middle or intermediate layer (10–15 m), and U – upper layer (>15 m). A indicates all stems of the species. The square root transformation, L(*d*), of Ripley's *K* were applied (see Moeur 1993).

Chen & Bradshaw 1999

Coherent Spatial Relationships of Species Distribution and Production in An Old-growth *Pseudotsuga-tsuga* Forest

Jiquan Chen*[1], Bo Song [1], Melinda Moeur[2], Mark Rudnicki[1], Ken Bible [3], Dave C. Shaw[3], Malcolm North[4], Dave M. Braun[3], and Jerry F. Franklin[3]

1. Dept. of Earth, Ecological and Environmental Science, University of Toledo, Toledo, OH 436062, USA.
2. USDA Forest Service, Rocky Mountain Research Station, Moscow, ID 83843, USA
3. College of Forest Resources, University of Washington, Seattle, WA 98195, USA
4. USDA Forest Service, Pacific Southwest Research Station, Fresno, CA 93710, USA

Hypotheses

1) all trees have a clumped distribution in small size classes but a regular distribution in large size classes;

2) there is no significant attraction or repulsion between different tree species at a fine scale (<50 m) but;

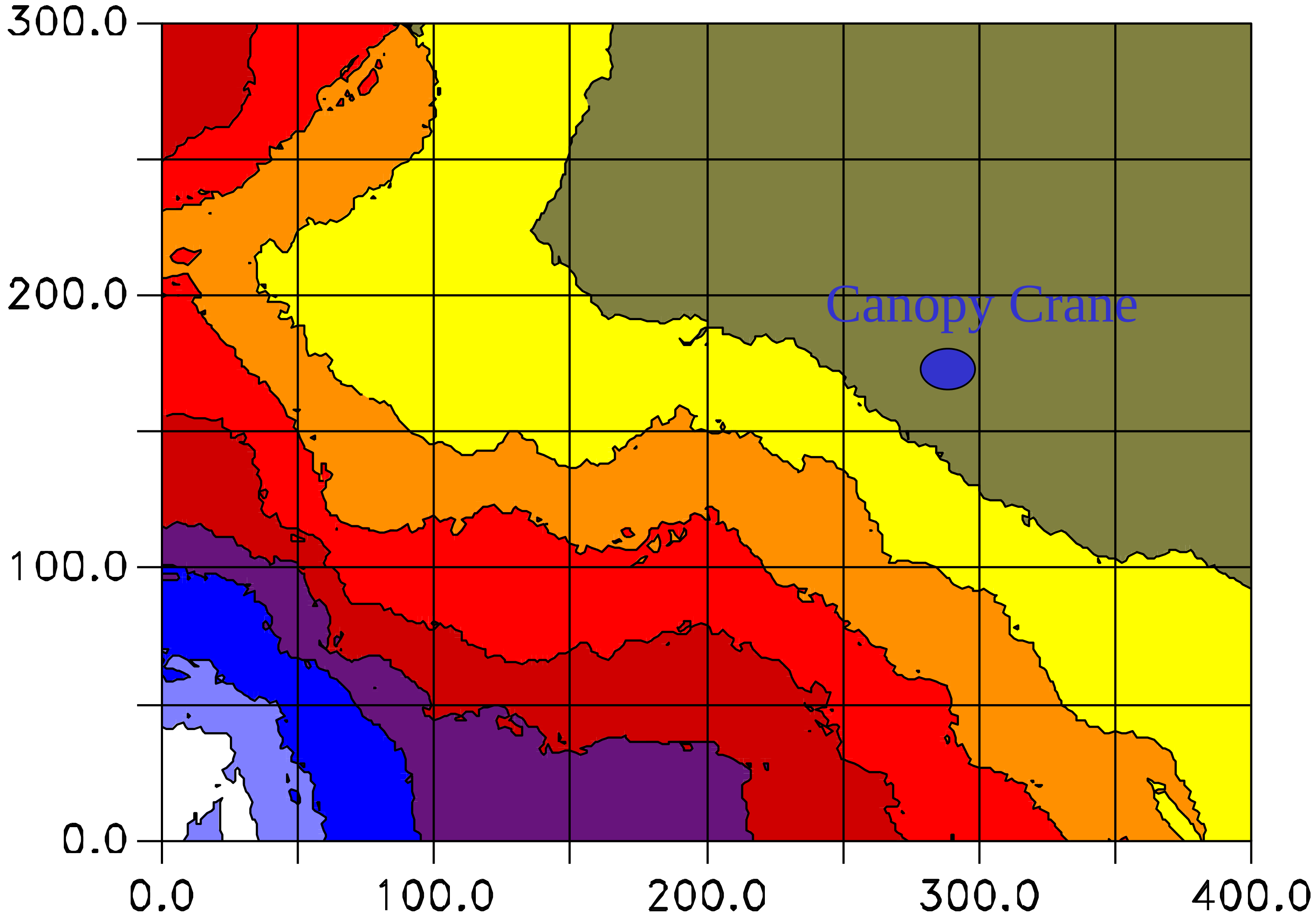
3) species will be significantly clumped at large scales (i.e., larger than canopy gaps); and

4) species distribution will indicate site production (i.e., above-ground biomass, AGB) because community composition and tree size directly determined AGB and other biomass measures across the stand.

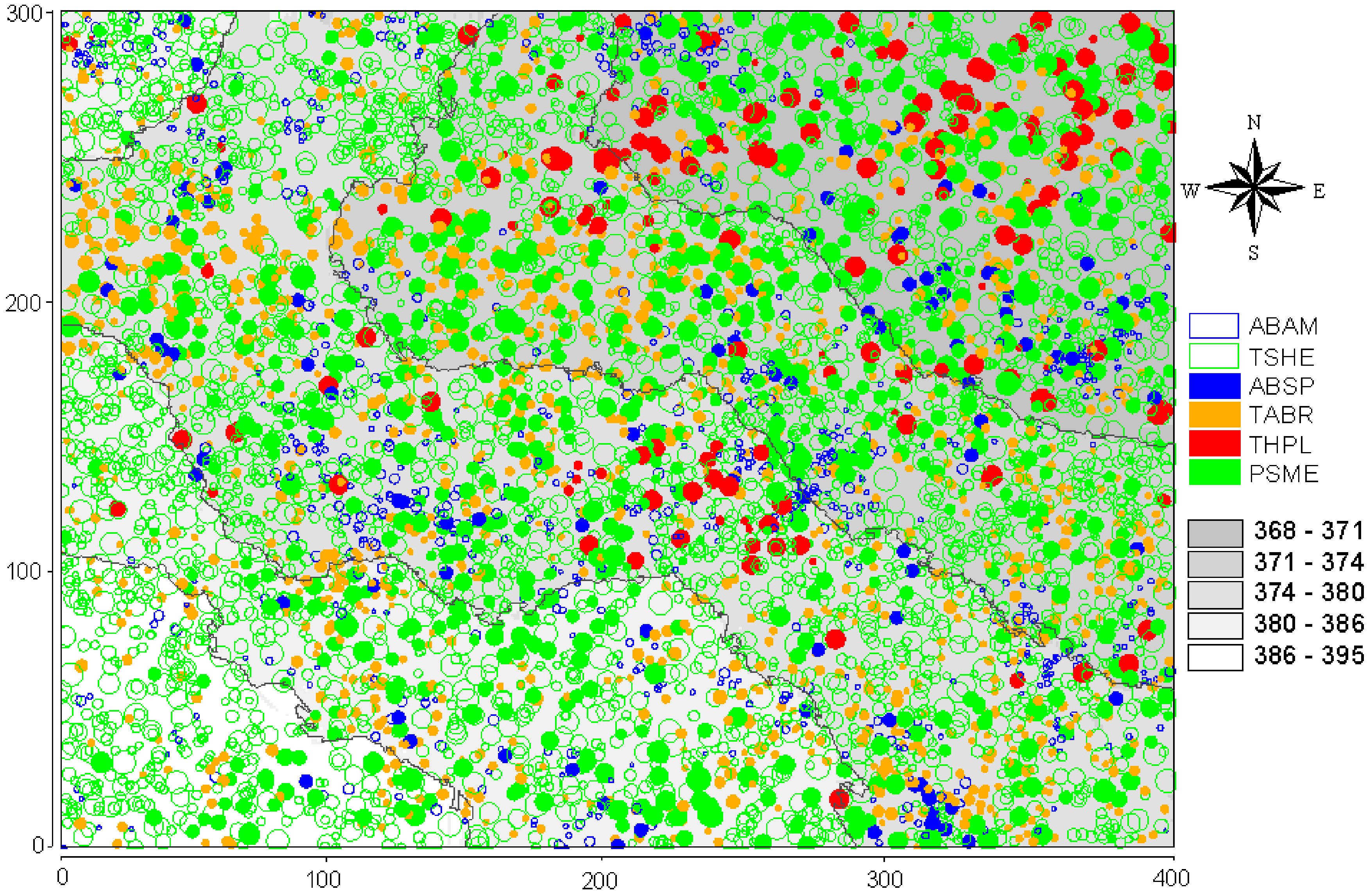
Methods

- Stem Map
- Point pattern Analysis (Ripley’s K function)
- Semivariance analysis and Kriging
- Spatial correlation between composition and production

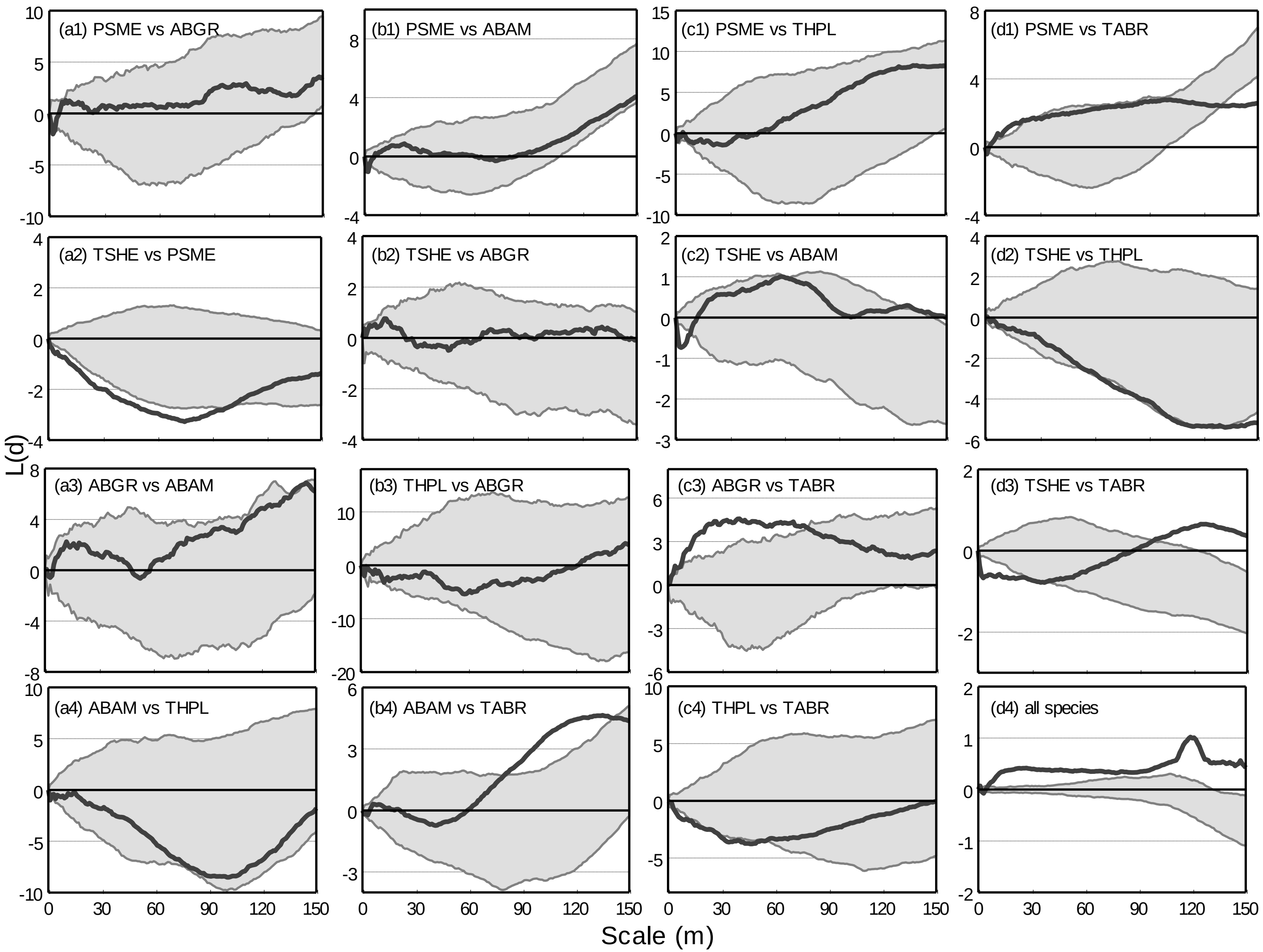
Elevation Gradient across the 12 ha plot at WRCCRF



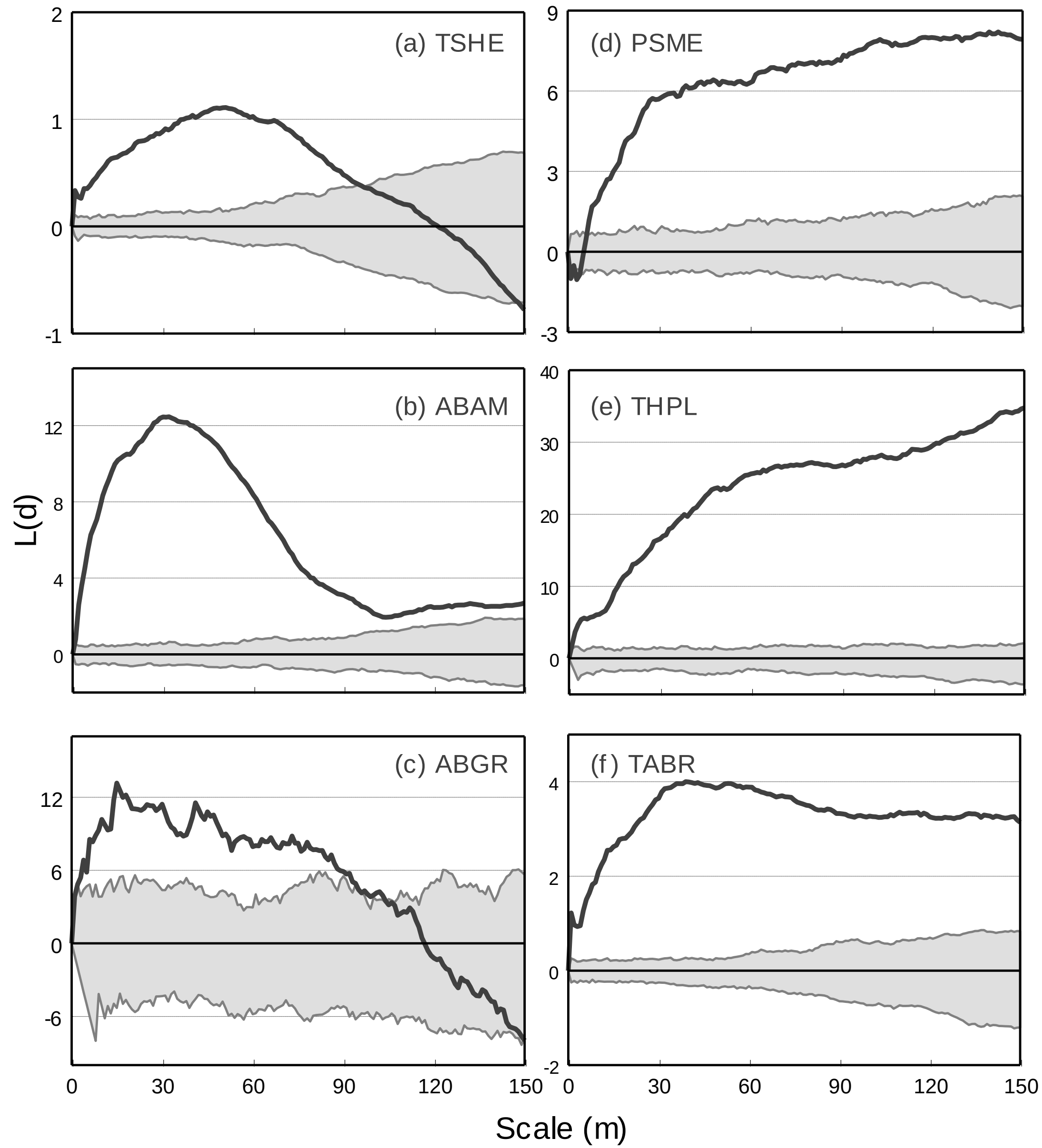
Spatial distribution of tree species on a topographic map in the 12 ha (400 x 300 m) plot facing north.

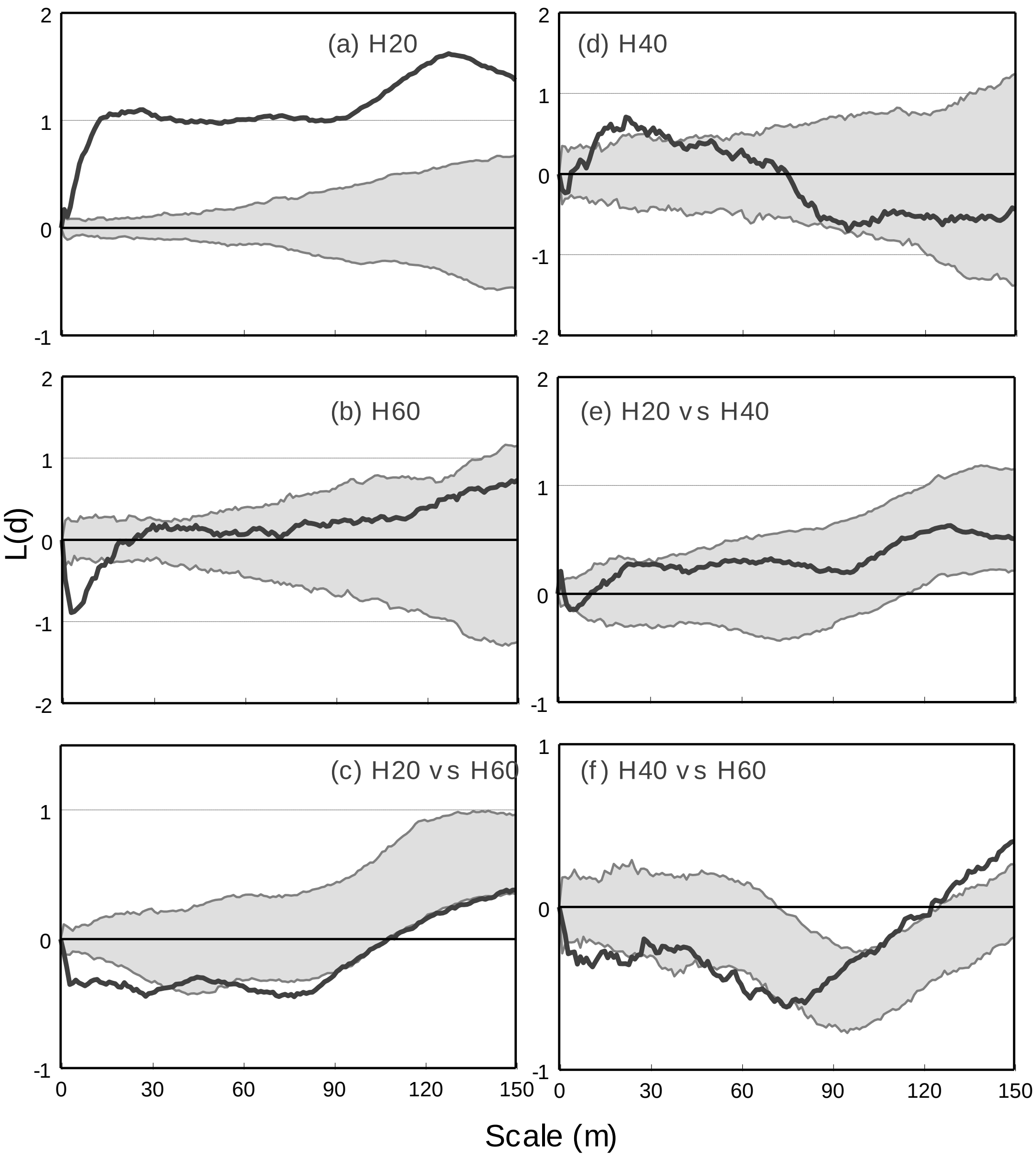


Ripley's K statistics for six major species in an old-growth Douglas-fir forest based on stem-mapped data. The Monte Carlo envelope (shaded area) was constructed at the 95% confidence level with 100 Monte Carlo simulations.



Bivariate Ripley's K statistics for all combinations of any two of the six major species (TSHE, PSME, ABGR, ABAM, THPL, and TABR) showing significant (95%) random (L(d) falls within the 95% CI envelope), attractive (L(d) falls above the 95% CI envelope), or repulsive (L(d) falls below the 95% CI envelope) based on 100 Monte Carlo simulations.





Bivariate Ripley's K statistics for all combinations of any two height classes for H20 (<20 m), H40 (20-40 m), and H60 (>40 m) showing significant (95% confidence intervals) showing repulsive, attractive, and random relationships based on 100 Monte Carlo simulations. Tree heights were calculated using models developed by Song (1998) and Ishii et al. (2000).

Spatial distribution of infected trees across the stand

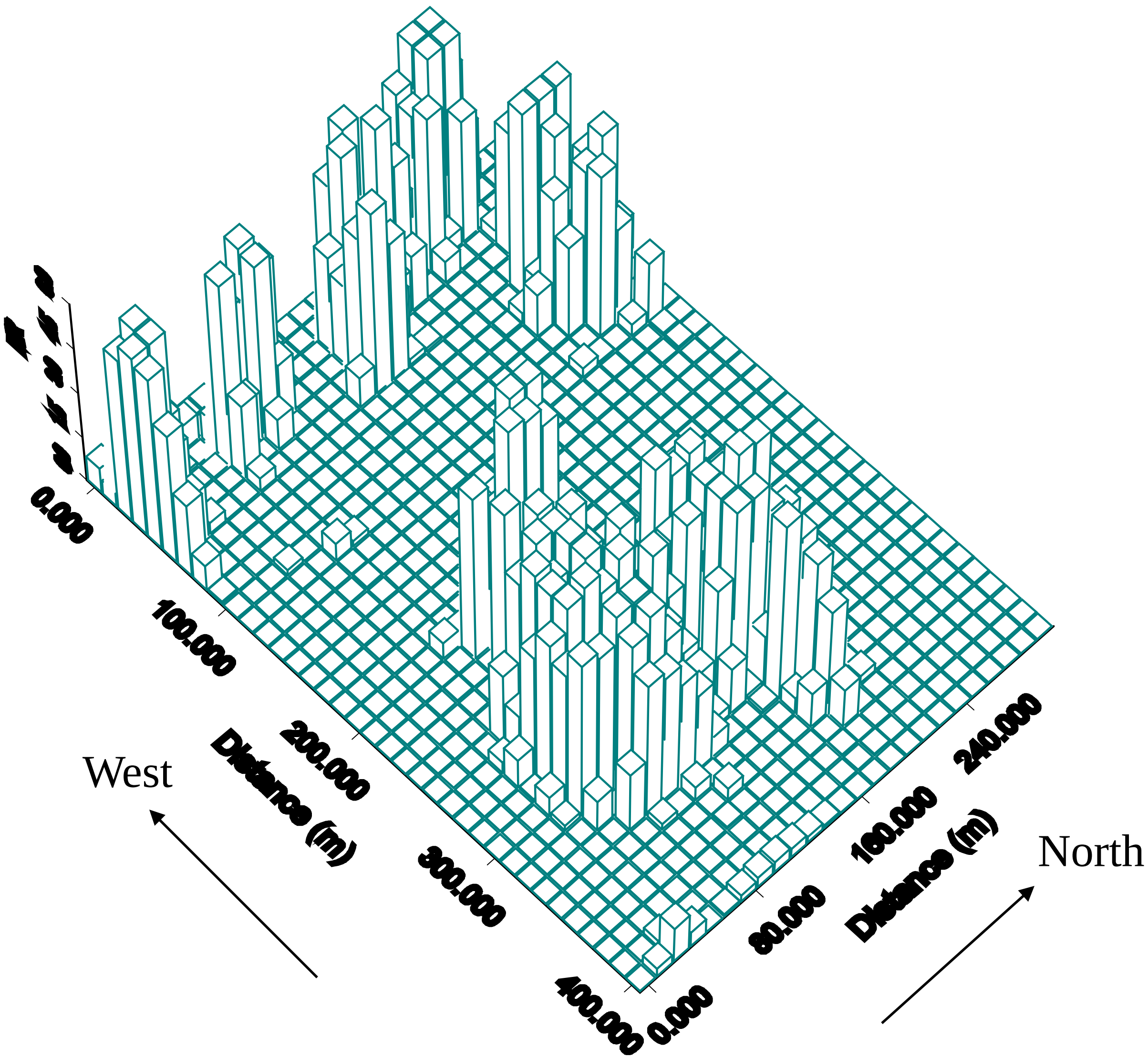


Fig. 6

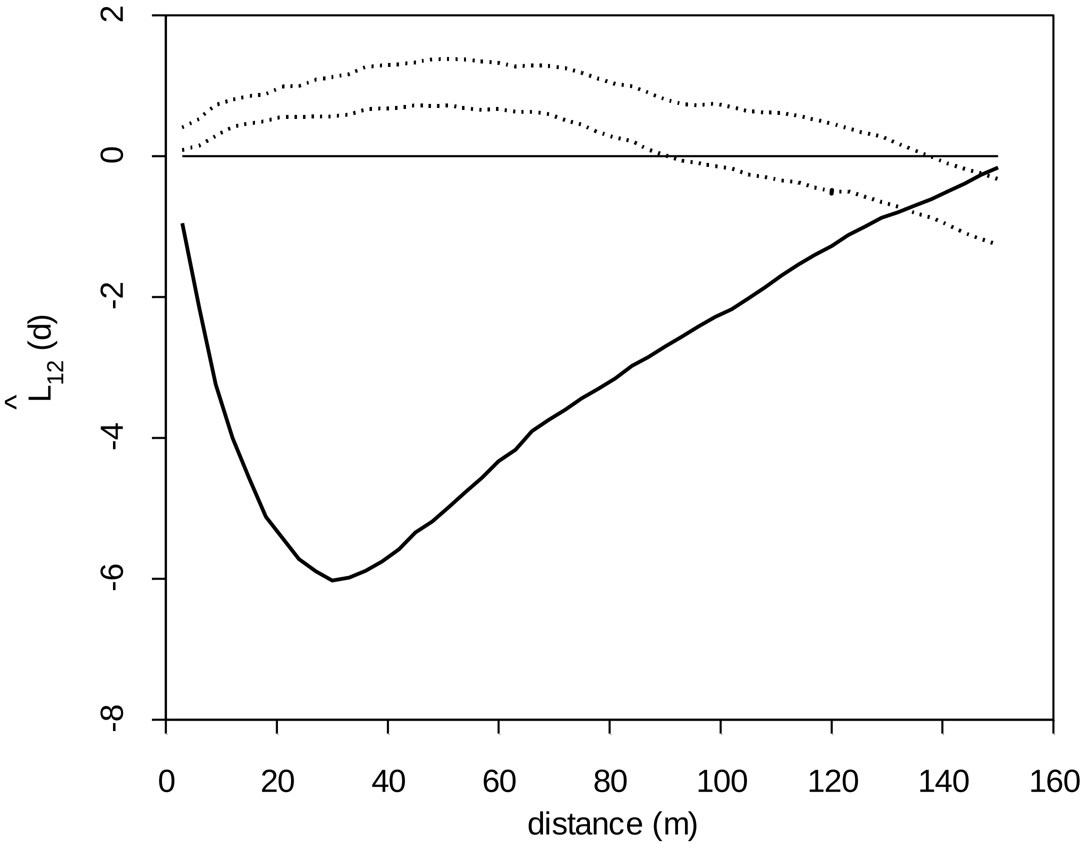


Fig. 7

Spatial distribution of infected trees across the stand

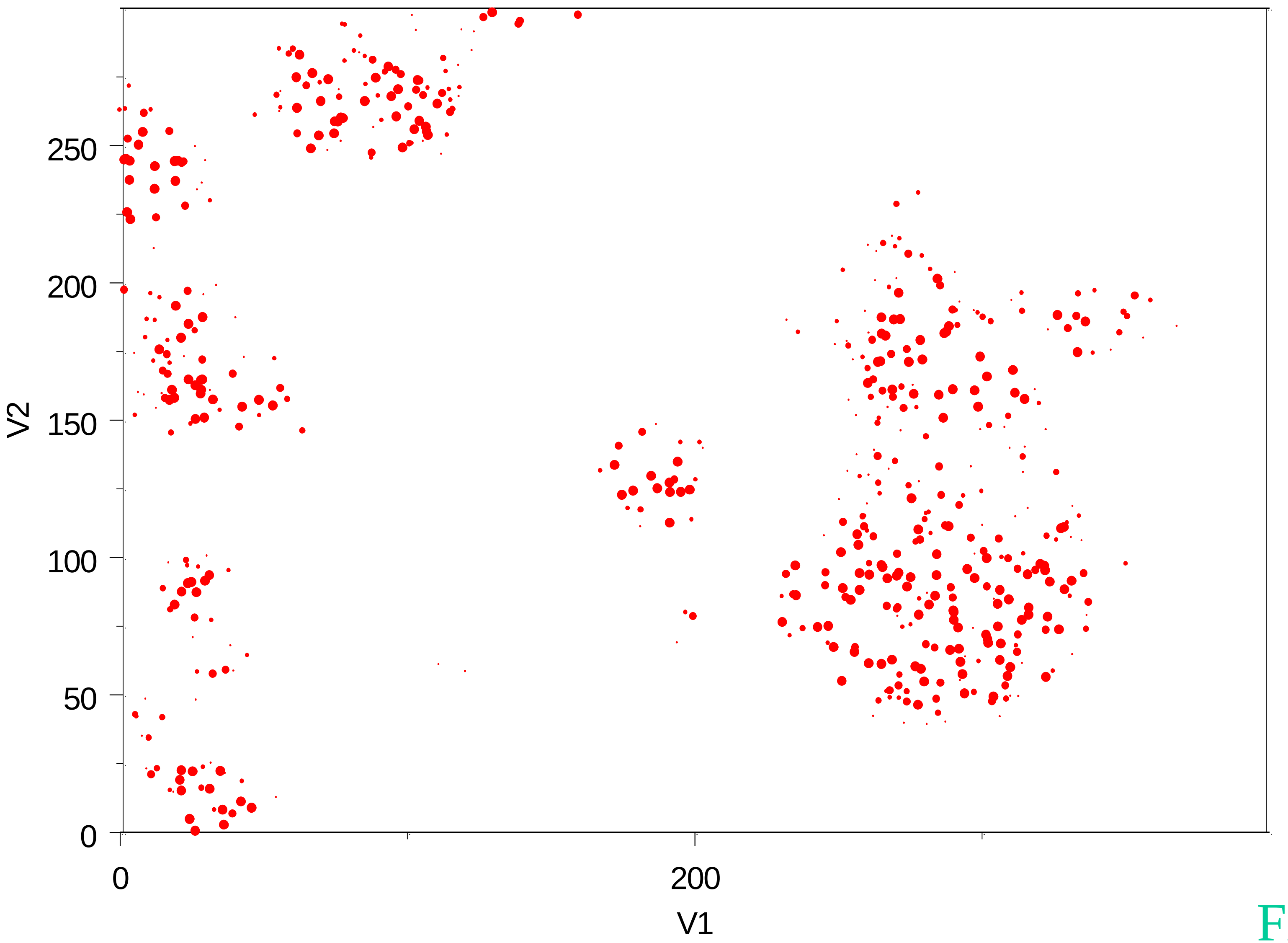


Fig. 8

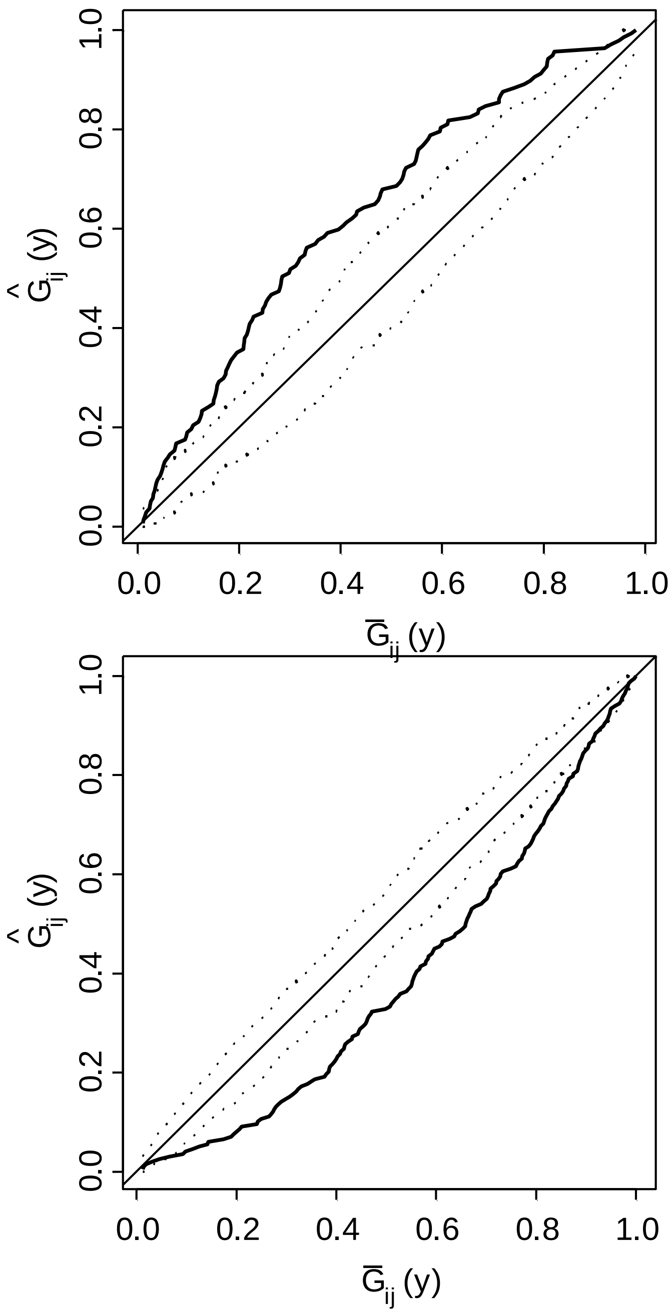
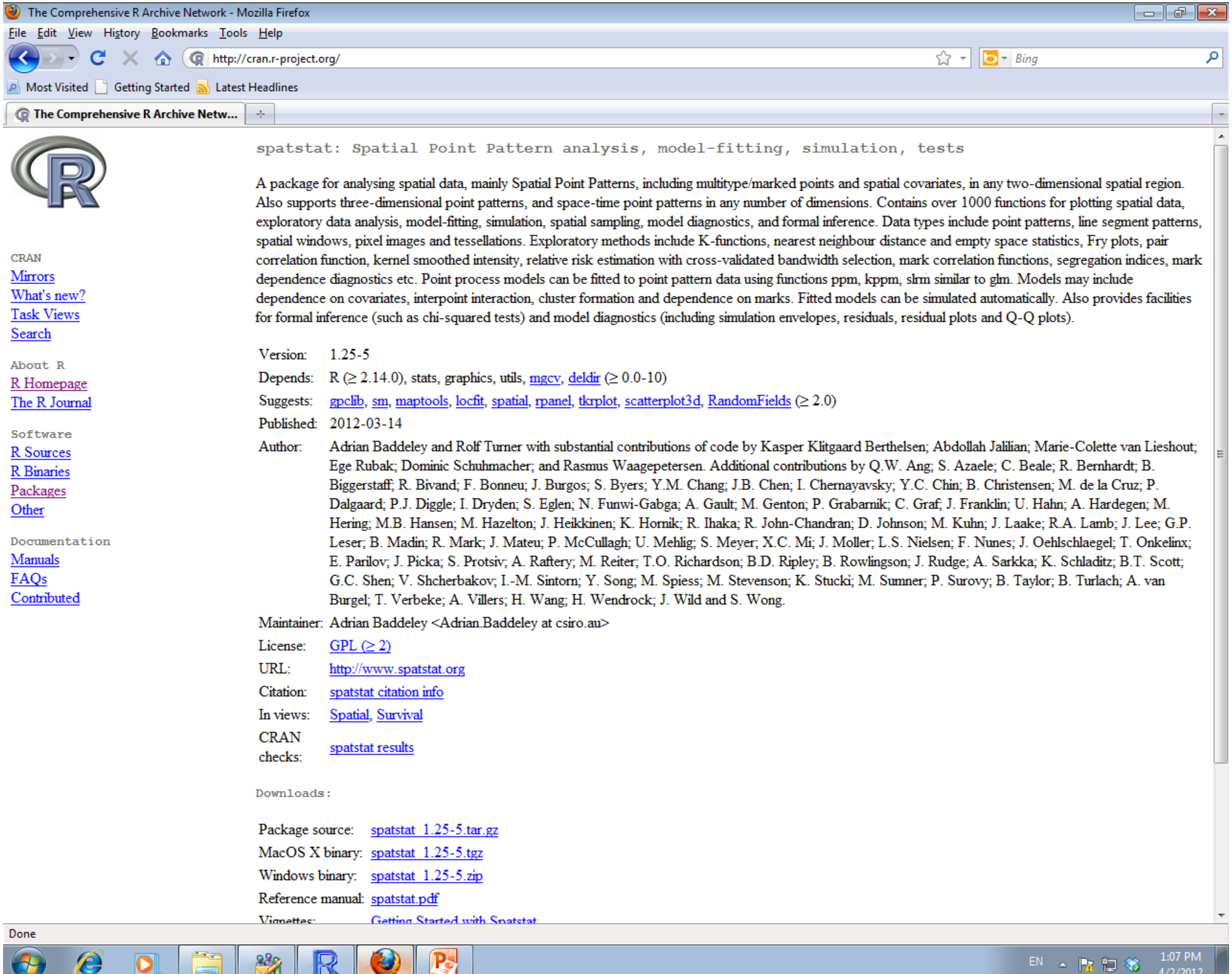

1.0
0.8
0.6
0.4
0.2
0.0
$\hat{G}_{ij}$ (y)
0.0 0.2 0.4 0.6 0.8 1.0
$\bar{G}_{ij}$ (y)
1.0
0.8
0.6
0.4
0.2
0.0
$\hat{G}_{ij}$ (y)
0.0 0.2 0.4 0.6 0.8 1.0
$\bar{G}_{ij}$ (y)

Fig. 9

R resource webpage: http://cran.r-project.org/


The Comprehensive R Archive Network - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://cran.r-project.org/

Most Visited Getting Started Latest Headlines

The Comprehensive R Archive Netw...

CRAN
Mirrors
What's new?
Task Views
Search

About R
R Homepage
The R Journal

Software
R Sources
R Binaries
Packages
Other

Documentation
Manuals
FAQs
Contributed

spatstat: Spatial Point Pattern analysis, model-fitting, simulation, tests

A package for analysing spatial data, mainly Spatial Point Patterns, including multitype/marked points and spatial covariates, in any two-dimensional spatial region. Also supports three-dimensional point patterns, and space-time point patterns in any number of dimensions. Contains over 1000 functions for plotting spatial data, exploratory data analysis, model-fitting, simulation, spatial sampling, model diagnostics, and formal inference. Data types include point patterns, line segment patterns, spatial windows, pixel images and tessellations. Exploratory methods include K-functions, nearest neighbour distance and empty space statistics, Fry plots, pair correlation function, kernel smoothed intensity, relative risk estimation with cross-validated bandwidth selection, mark correlation functions, segregation indices, mark dependence diagnostics etc. Point process models can be fitted to point pattern data using functions ppm, kppm, slrm similar to glm. Models may include dependence on covariates, interpoint interaction, cluster formation and dependence on marks. Fitted models can be simulated automatically. Also provides facilities for formal inference (such as chi-squared tests) and model diagnostics (including simulation envelopes, residuals, residual plots and Q-Q plots).

Version:	1.25-5
Depends:	R (≥ 2.14.0), stats, graphics, utils, mgcv, deldir (≥ 0.0-10)
Suggests:	gpclib, sm, maptools, locfit, spatial, rpanel, tkrplot, scatterplot3d, RandomFields (≥ 2.0)
Published:	2012-03-14
Author:	Adrian Baddeley and Rolf Turner with substantial contributions of code by Kasper Klitgaard Berthelsen; Abdollah Jalilian; Marie-Colette van Lieshout; Ege Rubak; Dominic Schuhmacher; and Rasmus Waagepetersen. Additional contributions by Q.W. Ang; S. Azaele; C. Beale; R. Bernhardt; B. Biggerstaff; R. Bivand; F. Bonneu; J. Burgos; S. Byers; Y.M. Chang; J.B. Chen; I. Chernayavsky; Y.C. Chin; B. Christensen; M. de la Cruz; P. Dalgaard; P.J. Diggle; I. Dryden; S. Eglen; N. Funwi-Gabga; A. Gault; M. Genton; P. Grabarnik; C. Graf; J. Franklin; U. Hahn; A. Hardegen; M. Hering; M.B. Hansen; M. Hazelton; J. Heikkinen; K. Hornik; R. Ihaka; R. John-Chandran; D. Johnson; M. Kuhn; J. Laake; R.A. Lamb; J. Lee; G.P. Leser; B. Madin; R. Mark; J. Mateu; P. McCullagh; U. Mehlig; S. Meyer; X.C. Mi; J. Moller; L.S. Nielsen; F. Nunes; J. Oehlschlaegel; T. Onkelinx; E. Parilov; J. Picka; S. Protsiv; A. Raftery; M. Reiter; T.O. Richardson; B.D. Ripley; B. Rowlingson; J. Rudge; A. Sarkka; K. Schladitz; B.T. Scott; G.C. Shen; V. Shcherbakov; I.-M. Sintorn; Y. Song; M. Spiess; M. Stevenson; K. Stucki; M. Sumner; P. Surovy; B. Taylor; B. Turlach; A. van Burgel; T. Verbeke; A. Villers; H. Wang; H. Wendrock; J. Wild and S. Wong.
Maintainer:	Adrian Baddeley <Adrian.Baddeley at csiro.au>
License:	GPL (≥ 2)
URL:	http://www.spatstat.org
Citation:	spatstat citation info
In views:	Spatial, Survival
CRAN checks:	spatstat results

Downloads:

Package source:	spatstat_1.25-5.tar.gz
MacOS X binary:	spatstat_1.25-5.tgz
Windows binary:	spatstat_1.25-5.zip
Reference manual:	spatstat.pdf
Vignettes:	Getting Started with Spatstat

Done

EN 1:07 PM 4/2/2012

Questions?

LANDSCAPE ECOLOGY &
ECOSYSTEM SCIENCE

http://research.eeescience.utoledo.edu/lees/

Spatial data exploration

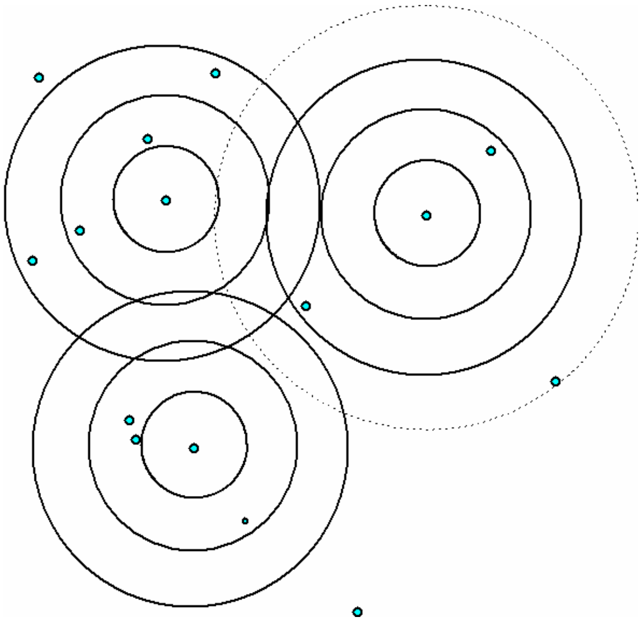
- Point (event) based statistics – clustering (ESDA)
  - Is the observed clustering due to natural background variation in the population from which the events arise?
  - Over what spatial scales does clustering occur?
  - Are clusters a reflection of regional variations in underlying variables?
  - Are clusters associated with some feature of interest, such as a refinery, waste disposal site or nuclear plant?
  - Are clusters simply spatial or are they spatio-temporal?

Spatial data exploration

- Point (event) based statistics – clustering
  - k^{th} order NN analysis
  - Cumulative distance frequency distribution, G(r)
  - Ripley K (or L) function – single or dual pattern
  - PCP
  - Hot spot and cluster analysis methods

Spatial data exploration

- Point (event) based statistics – Ripley K or L



- Construct a circle, radius *d*, around each point (event), *i*
- Count the number of other events, labelled *j*, that fall inside this circle
- Repeat these first two stages for all points *i*, and then sum the results
- Increment *d* by a small fixed amount
- Repeat the computation, giving values of K(*d*) for a set of distances, *d*
- Adjust to provide 'normalised measure' L:

$$L(d)=\sqrt{\frac{K(d)}{\pi}}-d$$

Spatial data exploration

- Point (event) based statistics – comments
  - CSR vs PCP vs other models
  - Data: location, time, attributes, error, duplicates
    - Duplicates: deliberate rounding, data resolution, genuine duplicate locations, agreed surrogate locations, deliberate data modification
  - Multi-approach analysis is beneficial
  - Methods: choice of methods and parameters
  - Other factors: borders, areas, metrics, background variation, temporal variation, non-spatial factors
  - Rare events and small samples
  - Process-pattern vs cause-effect
  - ESDA in most instances