# Spatial Data Analysis in R

## Point Patterns and Analysis

Eco 697DR – University of Massachusetts, Amherst – Spring 2022
Michael France Nelson

# Point Patterns

| What are point patterns useful for? | Characteristics of point patterns |
| --- | --- |
| • Describing configuration<br>  • Spatial arrangement of entities<br>• Inferring process<br>  • Testing/comparing different process hypotheses<br>  • Resource competition<br>  • Parent/offspring relationships | • Coordinates of points<br>• Marks (i.e. attributes)<br>  • Mark data type (integer, category, numeric, etc.)<br>• Generating process or model<br>• Homogeneous or inhomogeneous (stationarity/intensity)<br>• Isotropy (direction) |

# Configuration

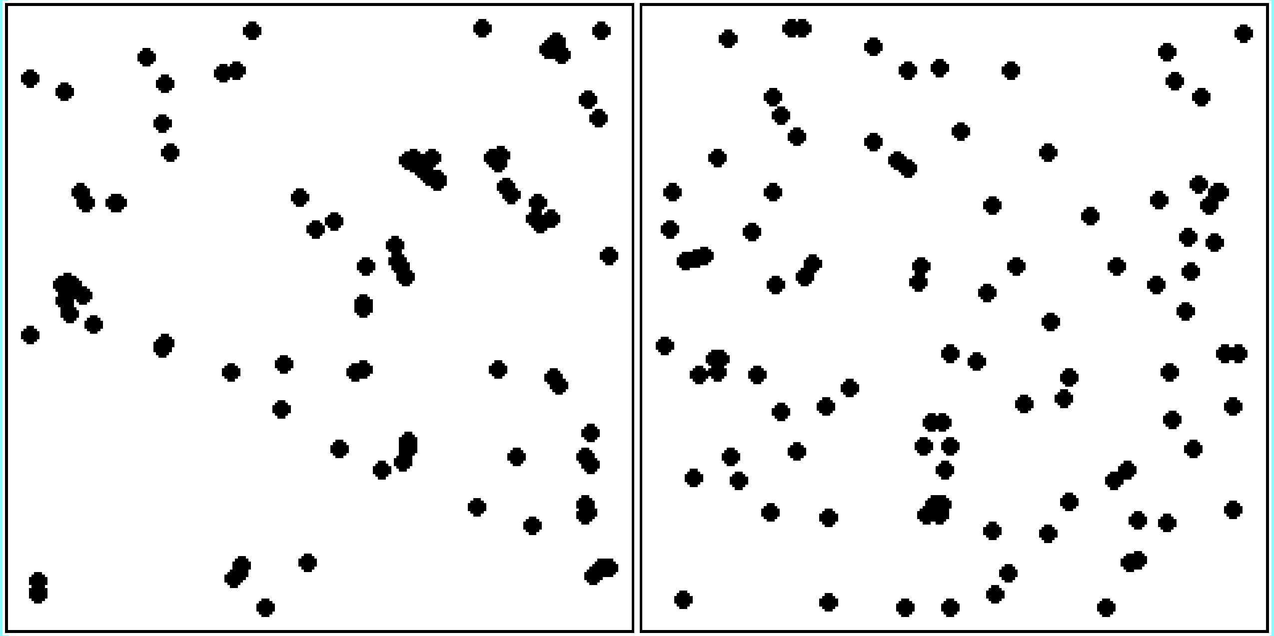
**Point patterns are all about configuration!**

Configuration and composition are different, but related, concepts.

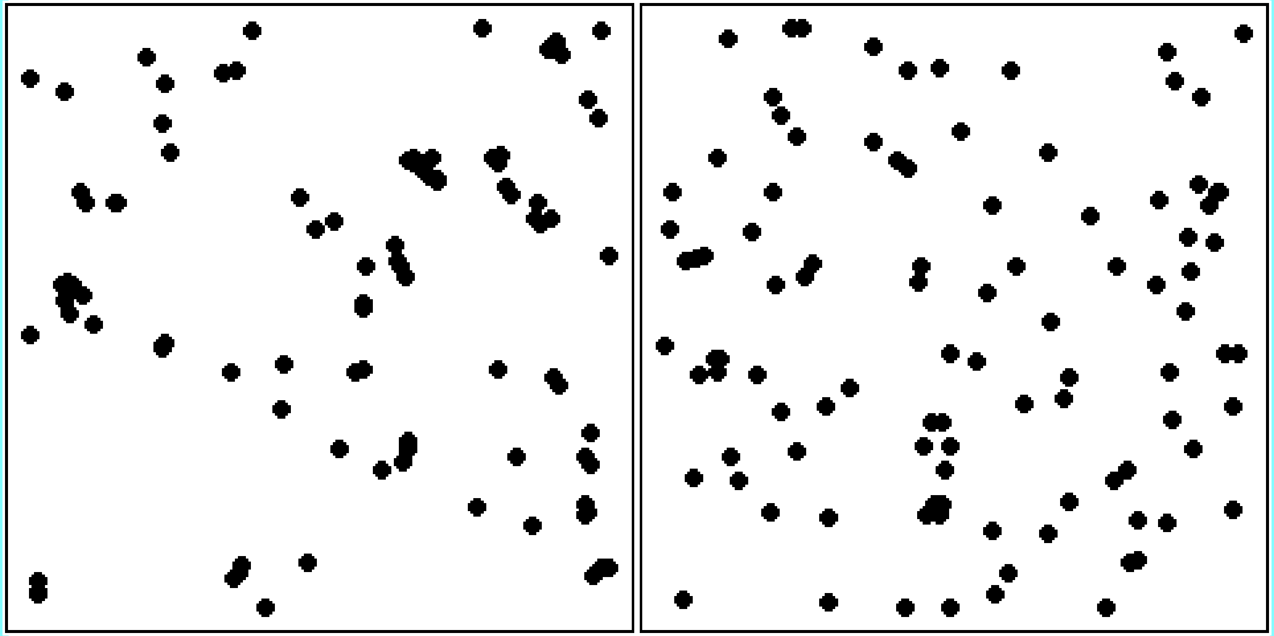
Why might one or both be relevant to your research context?

Stepping back, what are some broad *qualitative*, i.e. verbal, ways you could describe differences in configuration?

# Configuration of point patterns

| Configuration | How could you quantify these patterns? |
| --- | --- |
| Zooming in, what are some characteristics of configuration that you could measure to make your qualitative descriptions more concrete? |  |

# Configuration of point patterns

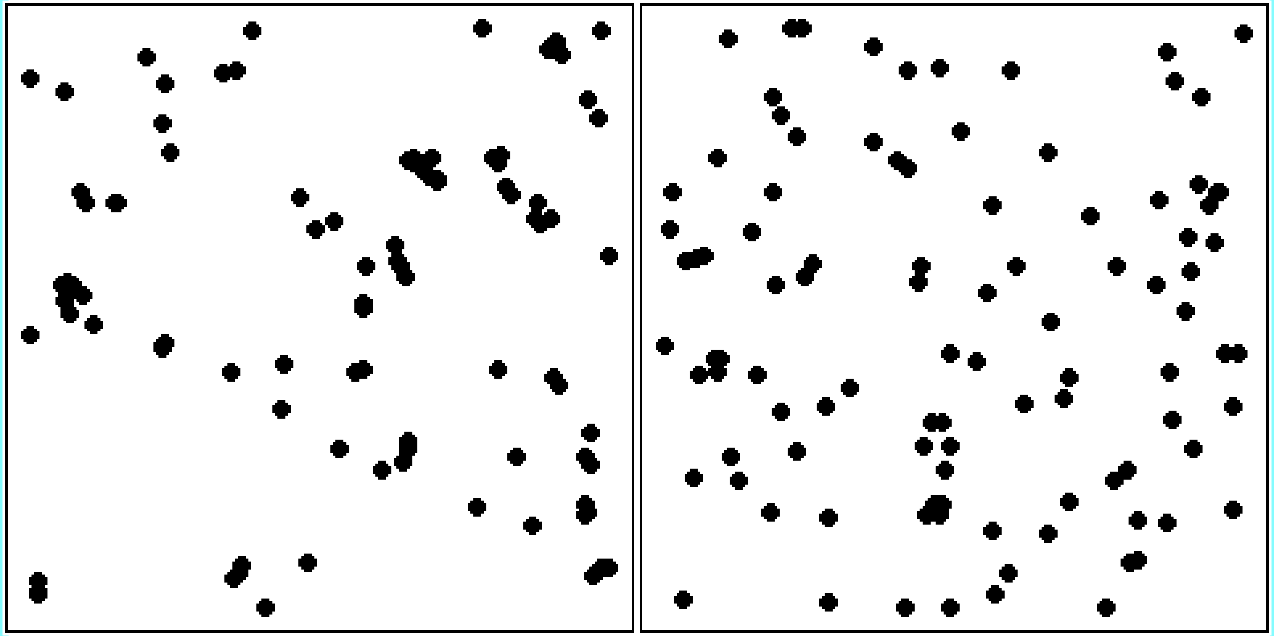
| Configuration | How could you quantify these patterns? |
|---|---|
| Zooming in, what are some characteristics of configuration that you could measure to make your qualitative descriptions more concrete?<br>• Density-based measures<br>• Distance-based measures |  |

# Configuration of point patterns

| Configuration | How could you quantify these patterns? |
|---|---|

Zooming in, what are some characteristics of configuration that you could measure to make your qualitative descriptions more concrete?

- Density-based measures
  - How many points are within distance r?
- Distance-based measures
  - What is the distribution of nearest-neighbor distances?



# Distance: Distance Matrices

We can represent pairwise distances between spatial objects (like points) using *distance matrices*.

Distance matrices are square with dimension n, where n is the number of spatial objects.

Distance matrices have $n^2$ elements – they can get very large for large datasets

We choose what type of distance to include.

- Euclidean, Manhattan, neighborhood, network

# Distance: Weight Matrices

**Weight matrix**

- Weight matrices are a kind of distance matrix.
- They help us define neighborhoods
- Weight matrices usually have higher values in cells representing nearby pairs of points.
- Weight matrix elements are kind of like a reciprocal of the distance matrix
- We can use an indicator function to populate the weight matrix

# Complete Spatial Randomness (CSR)

The x, and y, (and z for 3D space) coordinates are distributed *independently* and *uniformly* within the boundary.

What use is CSR?

# Poisson Distribution

What do we already know about the Poisson distribution?

# Poisson Distribution

**Poisson distribution: 1-parameter, lambda, distribution**

- The lambda is the *center* and the *spread*.
- Standard deviation equals the mean.

What is one key difference from the normal distribution?

**Poisson and uniform distributions in CSR**

- Points are uniformly and randomly distributed in space.
- Homogeneous/stationary point process
  - No change in intensity over the landscape

# Create a Poisson pattern

**One way to create a Poisson point pattern:**

- Define your intensity, a region of interest (ROI), and calculate the area of your ROI.
- The Poisson parameter (for the entire ROI) $\lambda$ is $intensity \times area$
- Generate a Poisson-distributed random number, n
- Generate n independently and uniformly-distributed x- and y-coordinates for the points.

What properties would such a pattern have?

Why is it called a Poisson pattern if we use the uniform distribution?

- Deep connection between uniform rates and the Poisson distribution.

# Poisson process and pattern

Number of points in independent samples (of the same area) are Poisson-distributed.

Link between the uniform and Poisson distributions!

Point counts are Poisson-distributed regardless of the area of your samples!

- But… what changes if you use different-sized samples?

When I first heard about this, I had to ‘prove’ to myself that this is true with some simulations in R.

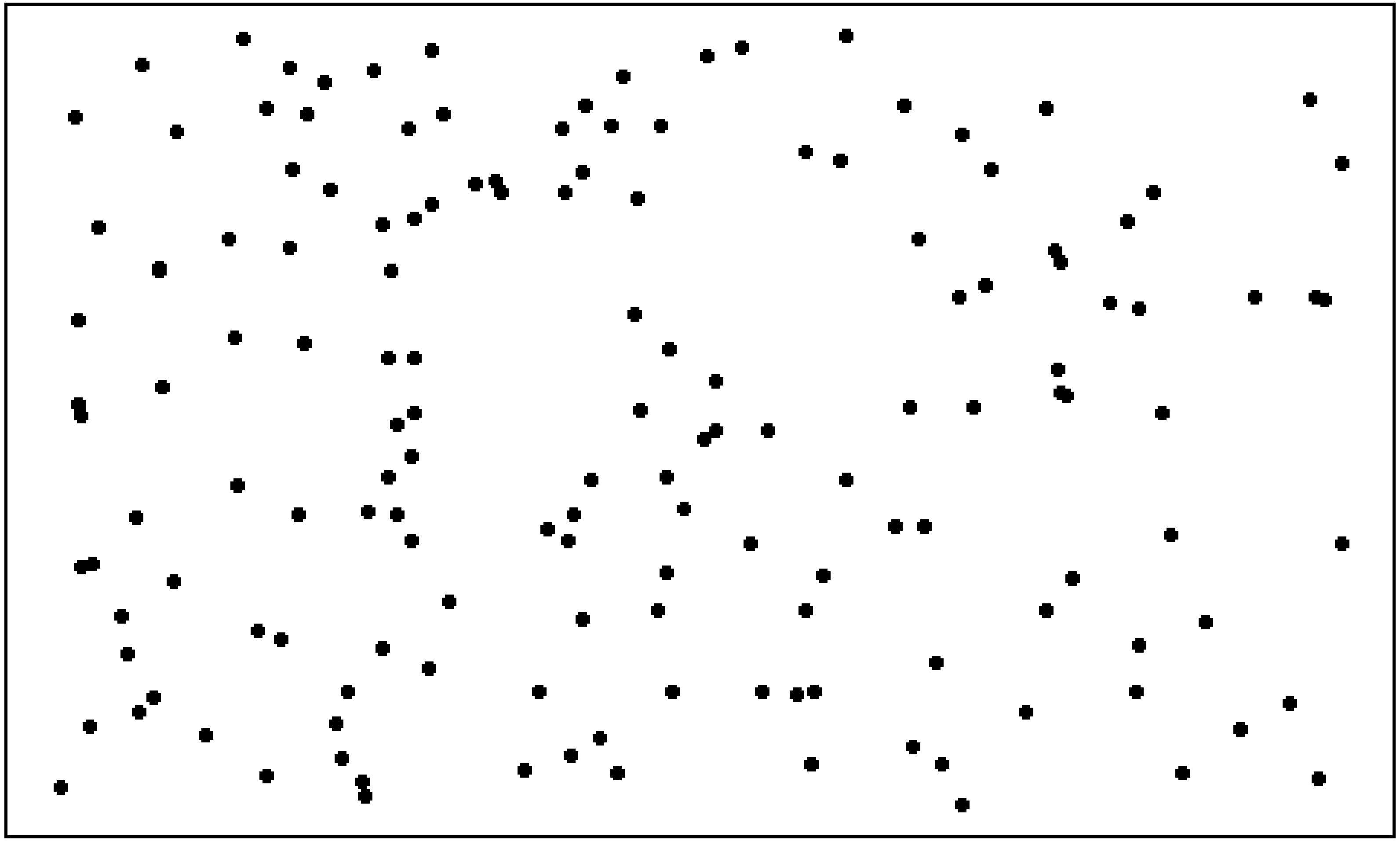
- Think about how you could perform such a simulation.

# Random Is Clumpier Than You Think

Shouldn't *random* feel evenly spaced?

We shouldn't see any patterns, right?

I was surprised by how clumpy randomness can look.



# Over- and underdispersion

I had imagined an *overdispersed* pattern.

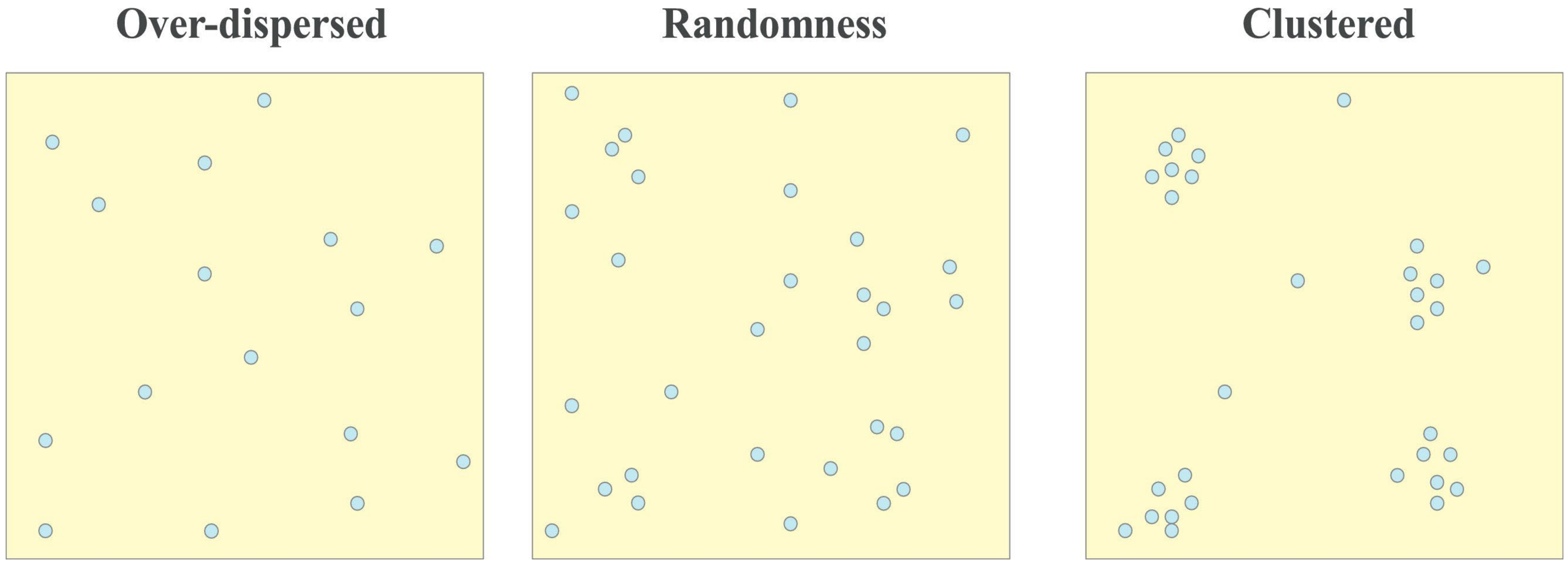



Figure 4 in Abraham et al. 2013 Modern statistical models for forensic fingerprint examinations: A critical review

# Quantifying aggregation

These functions assume a *homogeneous* process. What does that mean?

- K- and L- functions
- g function
- G function

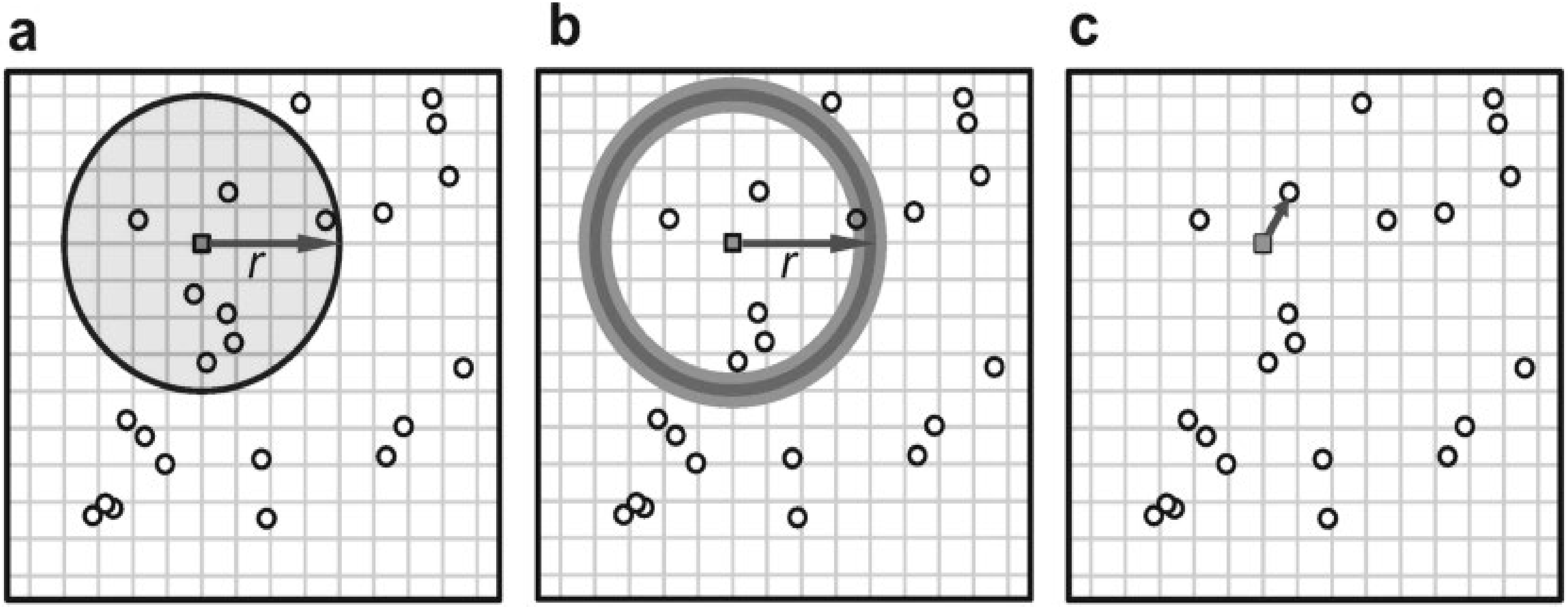



Figure 4.3 in Fletcher and Fortin, 2018

# Area based functions

**Ripley's K Function**

E is the number of points within radius r

$$K(r) = \frac{E}{\lambda}$$

Equation 4.2 in Fletcher and Fortin 2018

If E is large at radius r: evidence for clustering

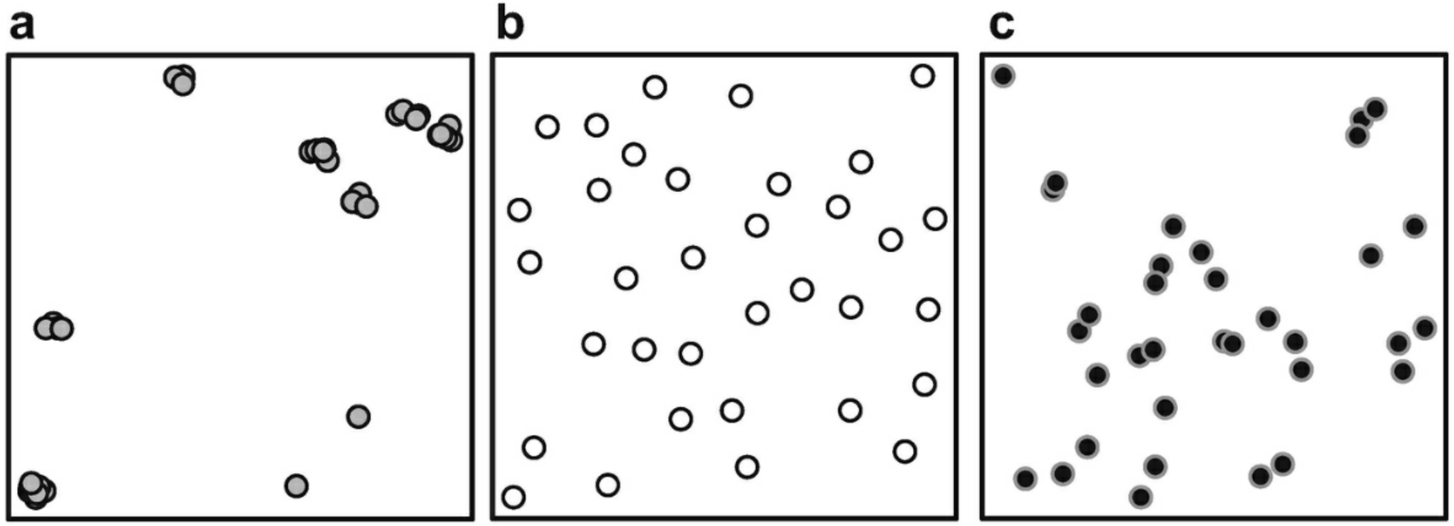



Figure 4.1 in Fletcher and Fortin 2018

If E is small: evidence for overdispersion

- Lambda is just the number of points per unit area (the intensity).
- Expected number of points with constant intensity (lambda) is proportional to the area of a circle with radius r:

$$E_{\mathrm{CSR}} = \lambda \pi r^2$$

- Lambda terms cancel, and we're left with the circle's area:

$$K(r) = \frac{E}{\lambda}$$

$$K(r)_{\mathrm{CSR}} = \pi r^2.$$

- The function is quadratic
  - F+F states the function is exponential, it should say quadratic.

# L Function

- Non-linearity is an undesirable property of the K function
- Since it's quadratic, we can just take the square root to get the L function:
- The L function is an increasing function under CSR:

$$\widehat{L}(r) = \sqrt{\frac{K(r)}{\pi}} = r.$$

- It's convenient to subtract the radius so we can compare L to zero.

# Interpreting L

- Under CSR, L(r) is always zero
- If there is an excess of nearby points within a circle of radius r, L is large.
  - This happens when points are clumped together in clusters.
- If there are too few points, L is less than r
  - This happens when patterns are overdispersed.

- Points can be clustered at one spatial scale and random at another:
- Significance of L is estimated with simulation envelopes

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

# Estimating K and L

(note the hat)

Let's dissect this equation! …Actually, let's use the 4.8 form….

$$\widehat{K}(r) = A \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{w_{ij} I_r\left(d_{ij} < r\right)}{n^2}$$

# Estimating K and L

Weight matrix

Indicator function

n/A is just lambda

$$\widehat{K}(r) = A \sum_{\substack{i=1 \\ i \neq j}}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{w_{ij} I_r (d_{ij} < r)}{n^2}$$

Sample size normalization.

Double summation indicates pairs of points

# K components

**Indicator function**

- simpler than it looks
- How many points are within a circle of radius r?

**Distance and weight matrices**

- Distance is calculated relative to the point of interest radius
- Weight Matrix

**Number of points: How many points are in our ROI?**

- Just add them up!
- Divide by the area of the ROI to get the intensity

**Summations**

- rows of distance matrix

# Estimating K and L: In Words

1. Calculate the intensity, lambda: Total number of points divided by the area
2. For each point, count the number of points within a circle of radius r
3. Sum all of the counts
    1. Apply the edge correction (if applicable)
4. Divide the sum by number of pairs of points
5. Divide the entire thing by lambda

# K and ring functions

K and L are cumulative: they count number of points within a circle

- K(5) includes the value of K(3)
- Lots of clustering or overdispersion at small radii can obscure patterns at larger r

What if we want to know density in a specific *range* of distances?

- The pair correlation function, g(r) (lowercase g), is analogous to K, or L, but quantifies the points within a ring.
- g(r) is not dominated by patterns at small r.

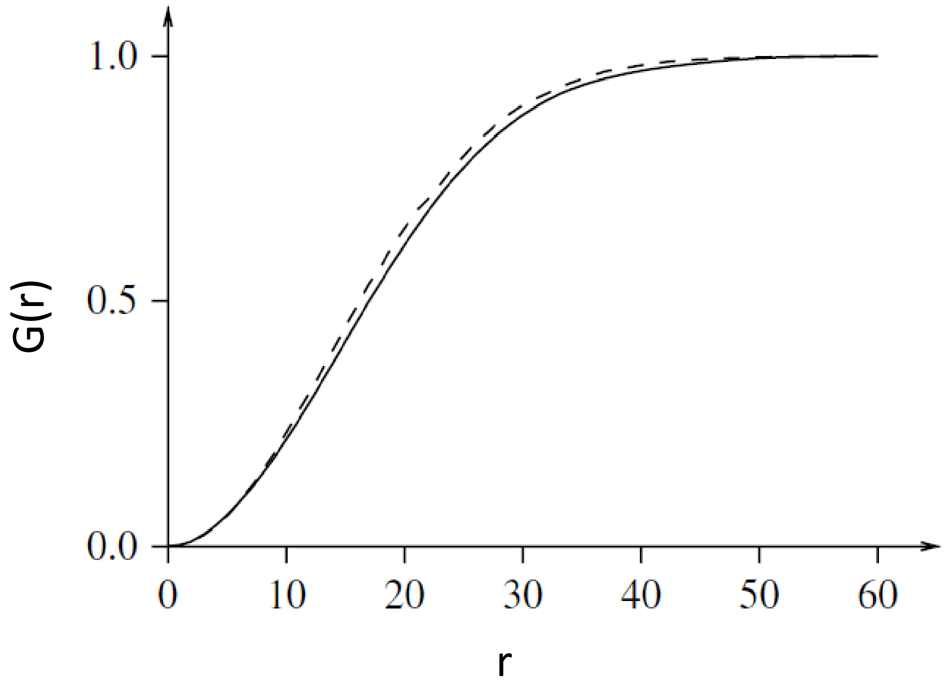
# Nearest-neighbor based functions

G function: G(r) = probability that nearest neighbor is less than r units away

- Also known as D(r)
- For a randomly chosen point, how likely is it that the nearest point is within a circle of radius r?
- G function is like a cumulative distribution function: it is monotonically increasing.
- If all points are within r, then G(r) = 1

$$\widehat{G}(r) = \frac{1}{N} \sum_{i=1}^{n} 1\,(d_i \le r)$$

$$G(r)_{\mathrm{CSR}} = 1 - \exp\left(-\lambda \pi r^2\right)$$




From figure 4.11 in Illian et al. 2008
Statistical Analysis and Modelling of Spatial Point Patterns

# Point generating processes

**CSR: Poisson process**

**Inhomogeneous Poisson patterns: non-constant intensity**

- Cox process: intensity $\lambda$ varies in space

**Offspring point patterns: clustering**

- Parent points are CSR.
- Matérn: offspring are CSR within a radius of parent points.
- Thomas: offspring are Normally-distributed about parent

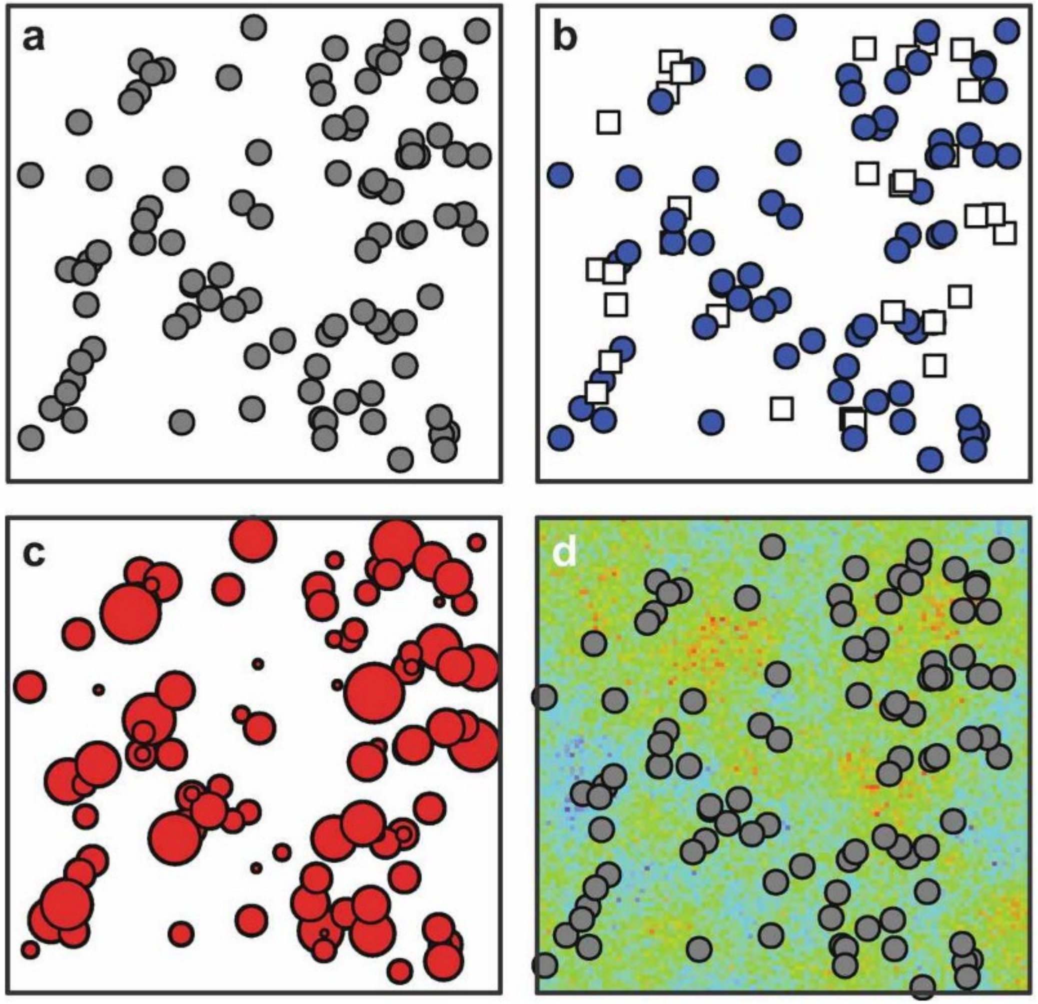
**Processes with repulsion: overdispersion**

- Inhibition zone
- Hard-core and soft-core processes: minimum core distance

# Marked Point Patterns

# Marks

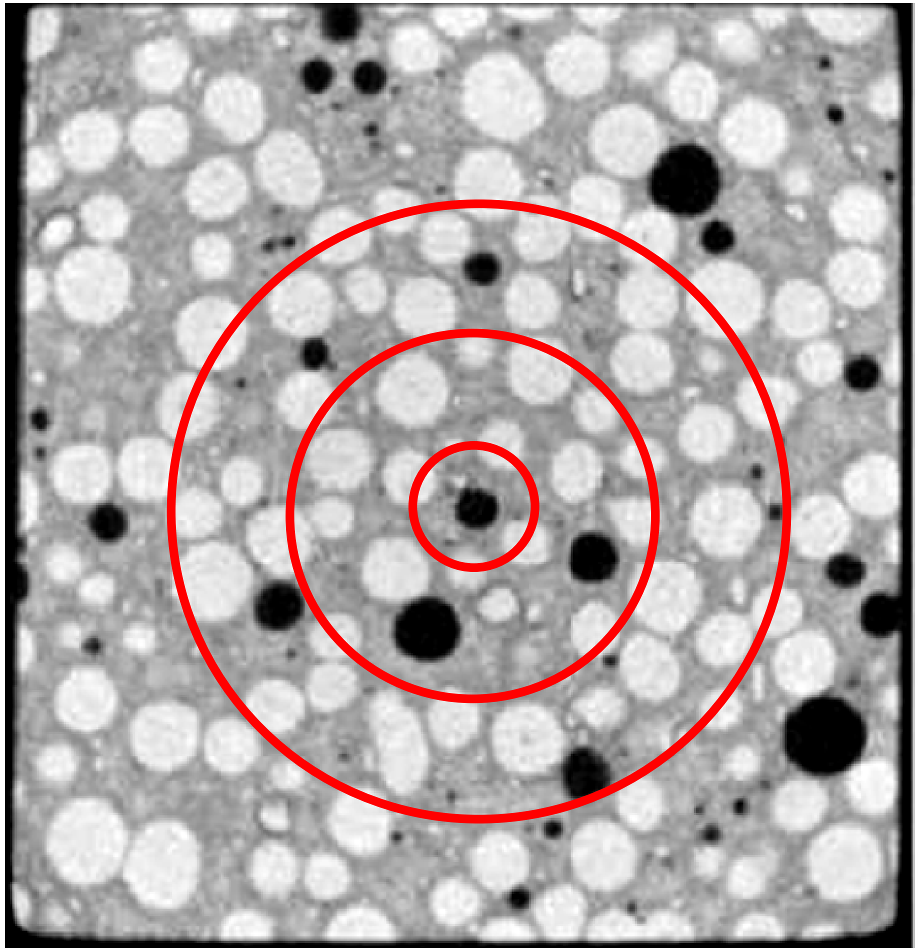
- So far, we've only talked about unmarked, or univariate, point patterns.
- Points can have characteristics (marks)
- For example:
  - Species (categorical)
  - Tree diameter (numeric)
  - Presence/absence
  - Elevation
- Points can have multiple marks




F + F *Fig. 4.2* Some common characteristics of point patterns.

# Marks: The Bivariate K Function

- Is one class of point clustered *with respect to* the other type?
  - The univariate asks: Are points clustered with respect to their own type?
- Per Illian et al 2008: "$K_{ij}(r)$ is the mean number of points of type j in a disc (sphere) of radius r centered at the typical point of type i."



Grains and air pores in concrete.
Figure 5.7 in Illian et al. 2008

# Interpreting Point Patterns

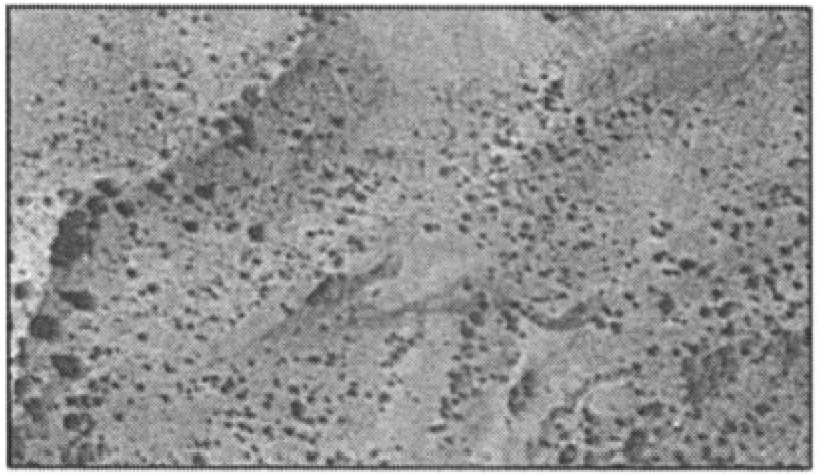
# Recall the definitions

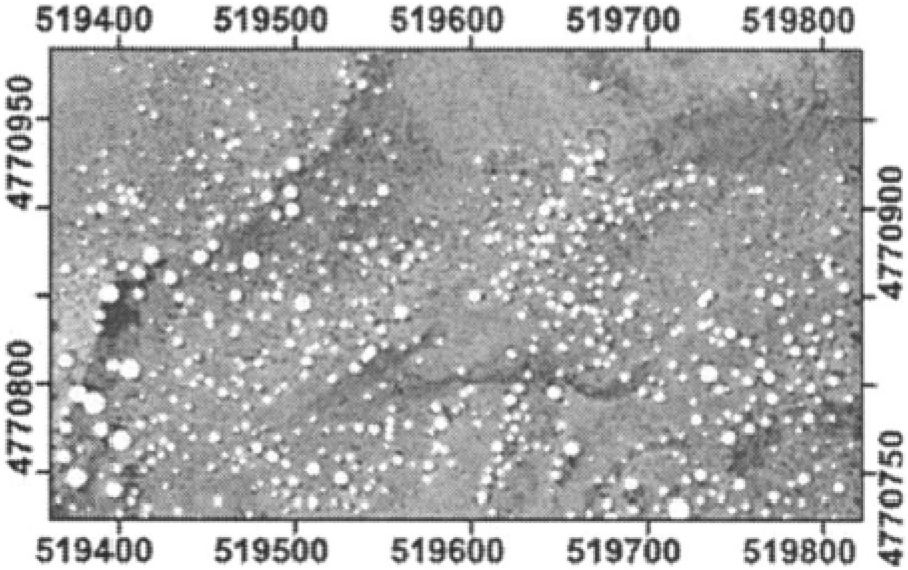
| Density-based measures | Neighbor based measures |
|---|---|
| • How many points are within a circle of radius r? | • What is the probability that my nearest neighbor is within a circle of radius r? |

# An Example: Sagebrush/Juniper Steppe

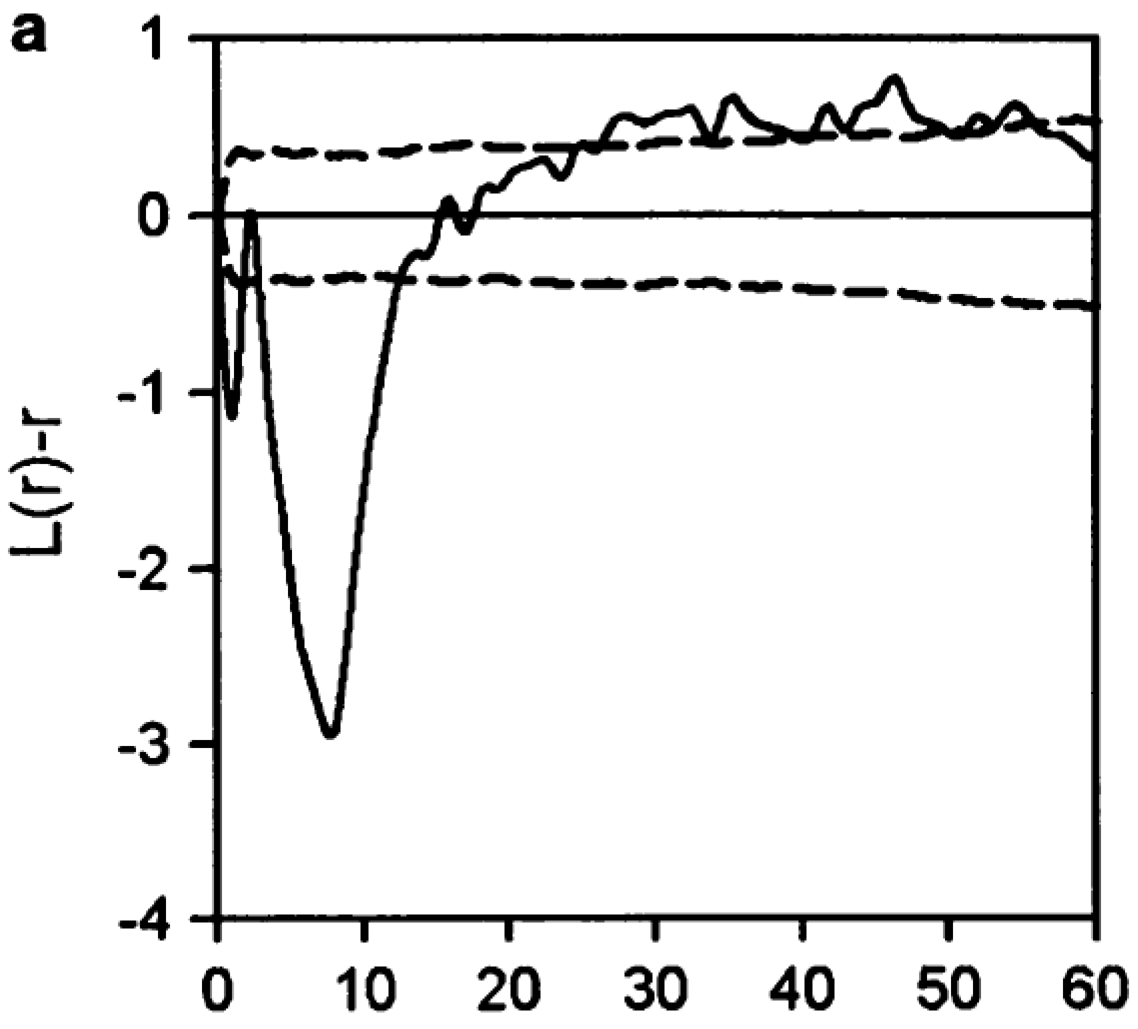
**Spatial patterns on the sagebrush steppe/Western juniper ecotone**

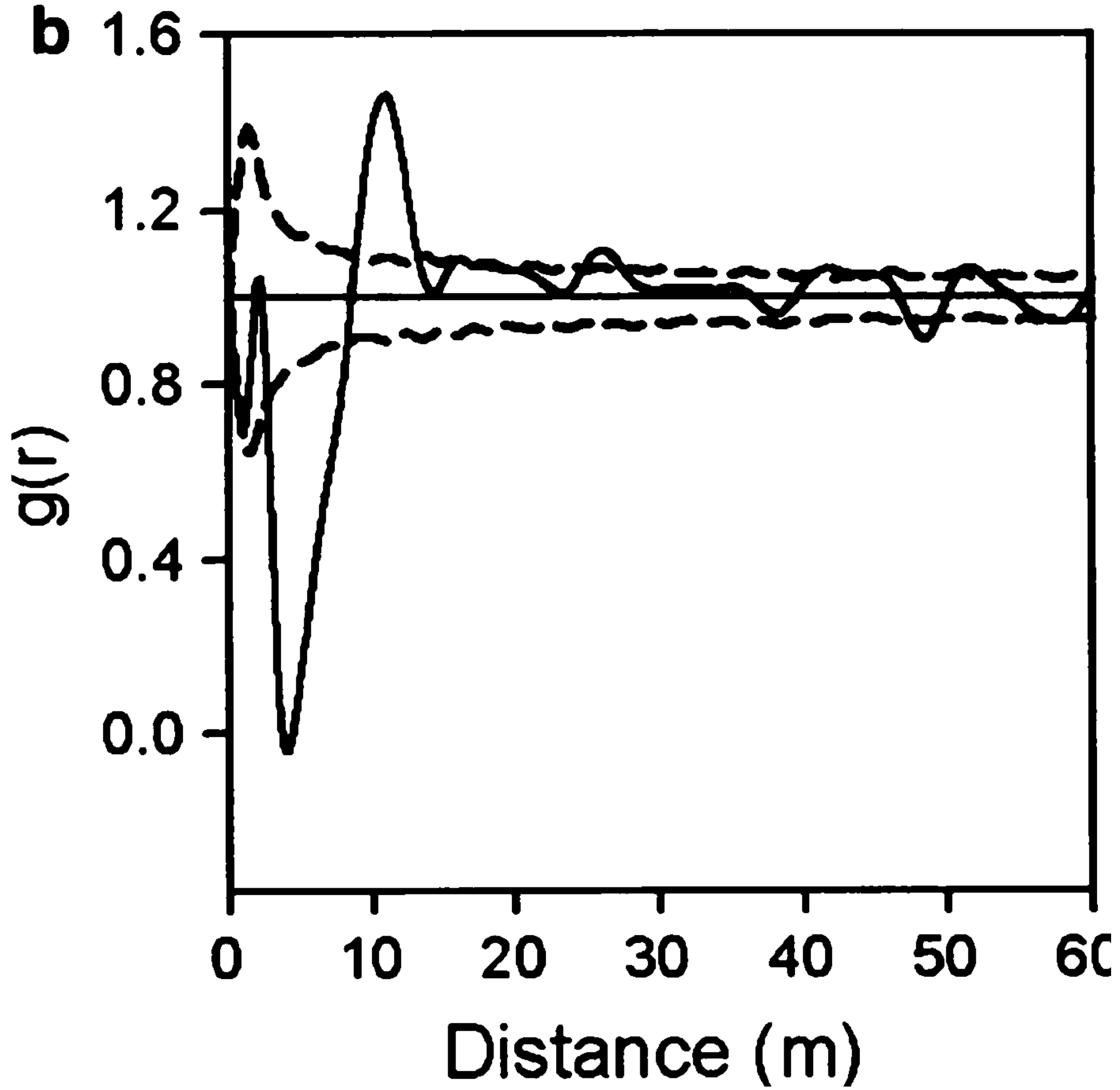
**Eva K. Strand · Andrew P. Robinson · Stephen C. Bunting**



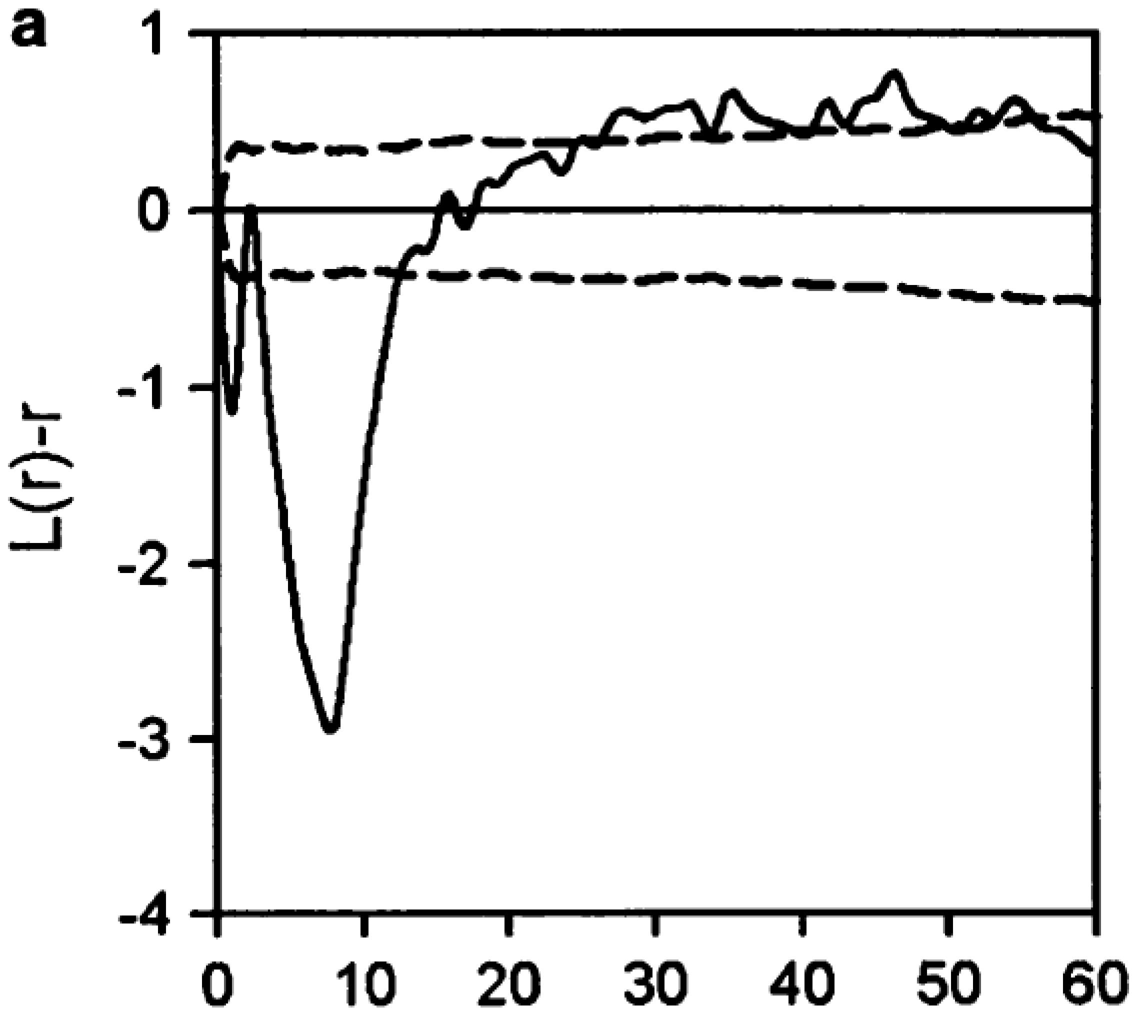


# L and g functions for Juniper



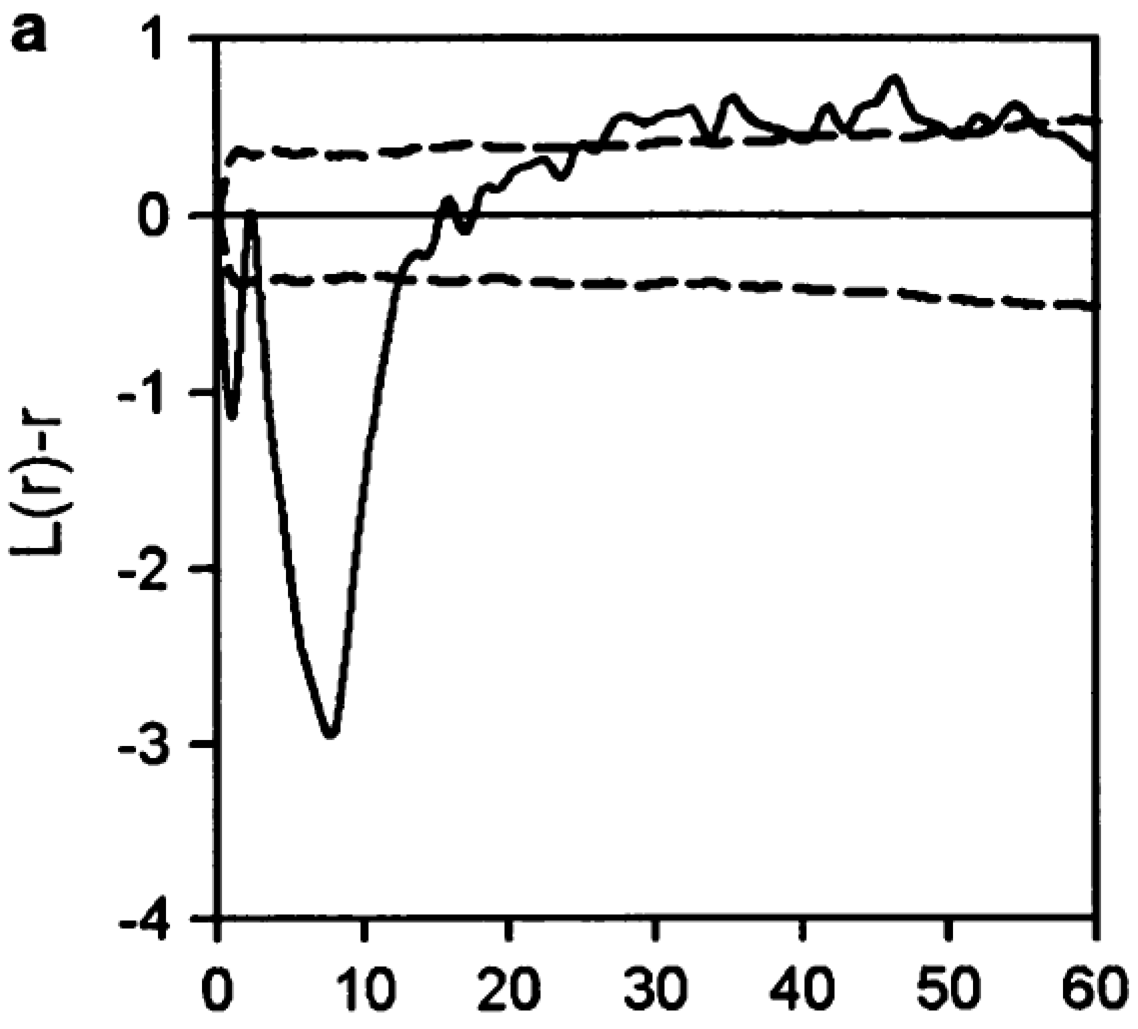



# L and function for Juniper



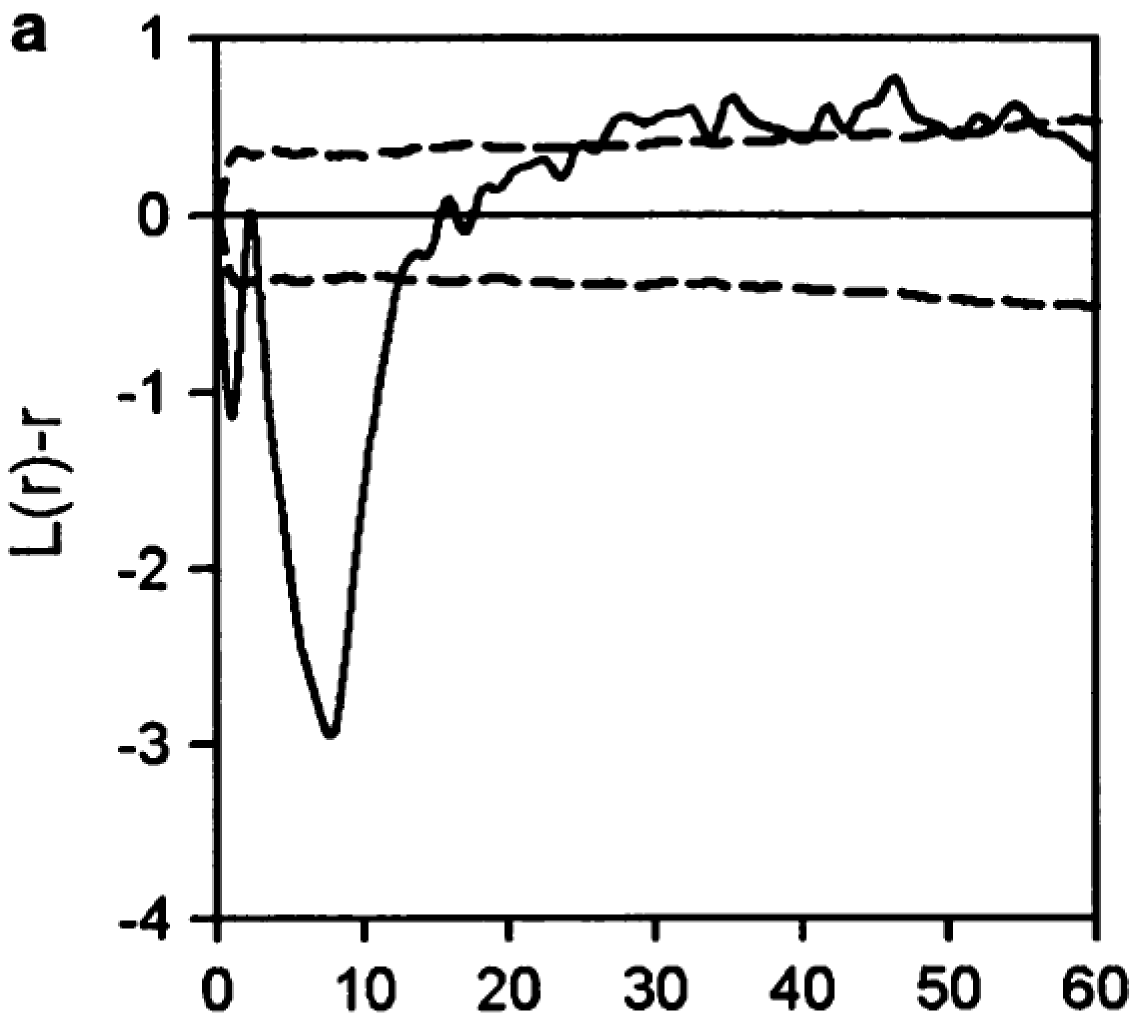

- How to interpret this L-function?

# L and function for Juniper



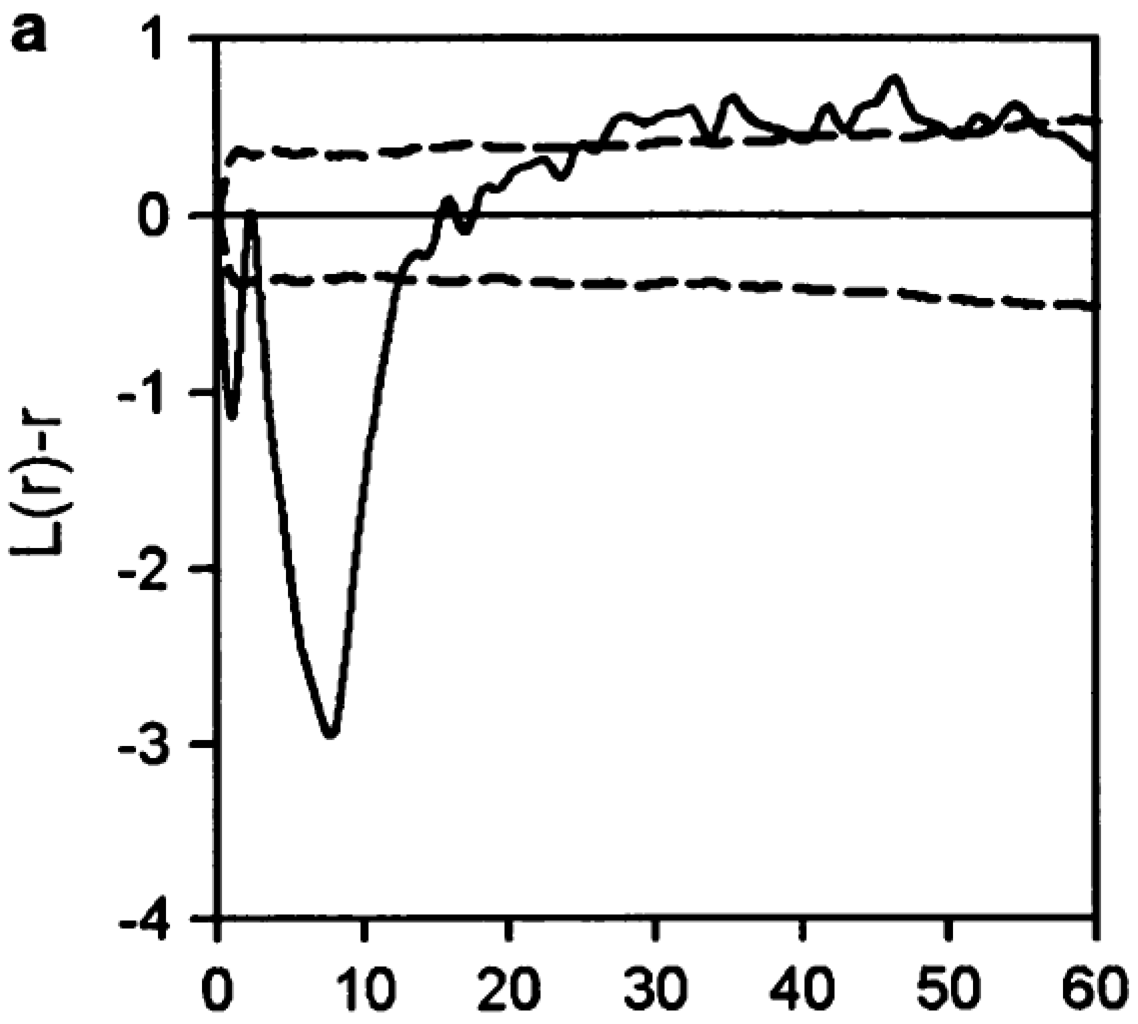

- How to interpret this L-function?
  - L<0 indicates overdispersion
  - L>0 indicates clustering
  - Dotted lines are a 95% simulation envelope

# L and function for Juniper




- How to interpret this L-function?
  - L<0 indicates overdispersion
  - L>0 indicates clustering
  - Dotted lines are a 95% simulation envelope
  - Fine-scale inhibition
  - What process(es) might lead to the observed pattern?

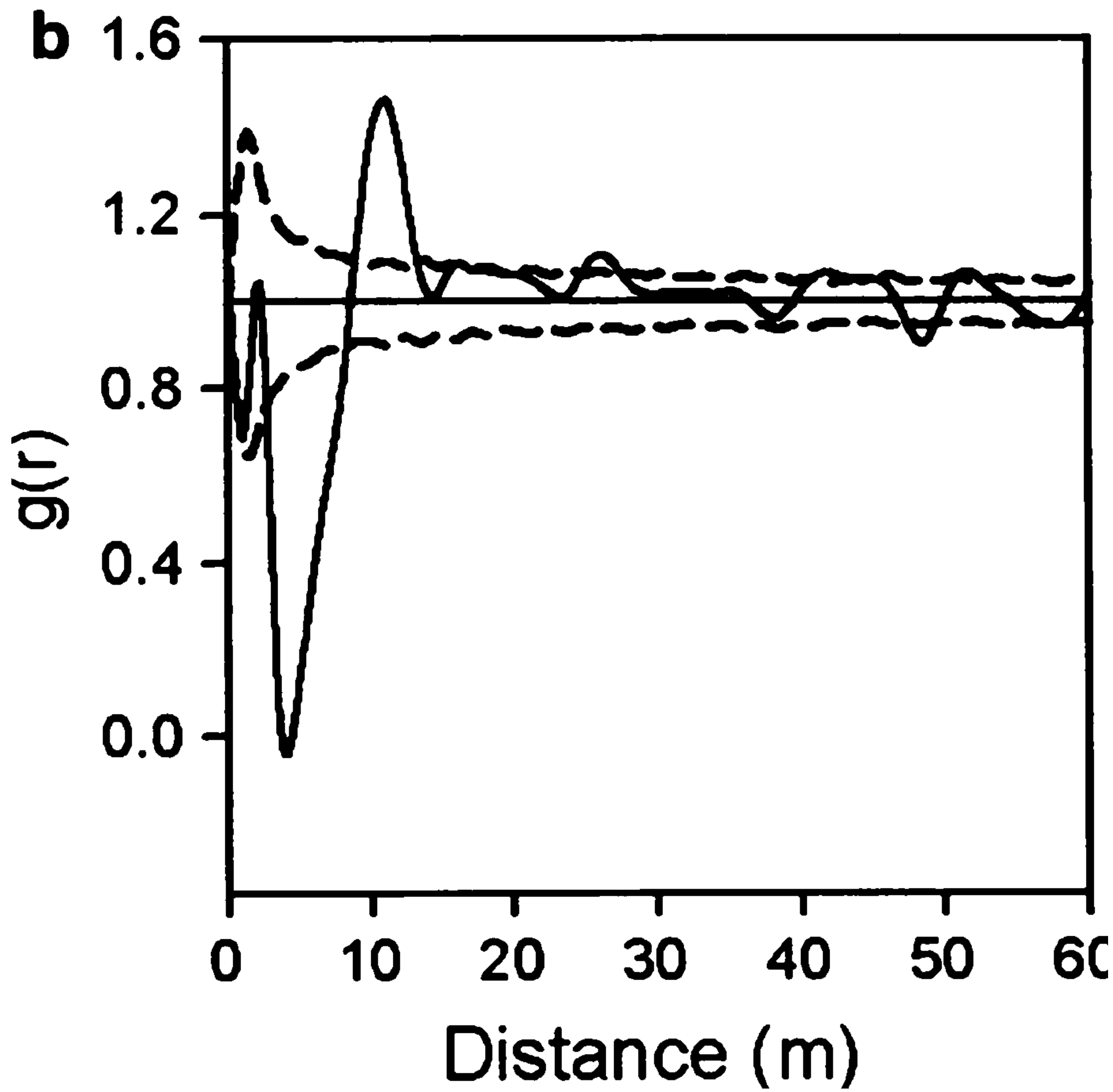
# L and function for Juniper




- How to interpret this L-function?
  - L<0 indicates overdispersion
  - L>0 indicates clustering
  - Dotted lines are a 95% simulation envelope
  - Fine-scale inhibition
  - What process(es) might lead to the observed pattern?
    - Competition for space?
    - Competition for water?

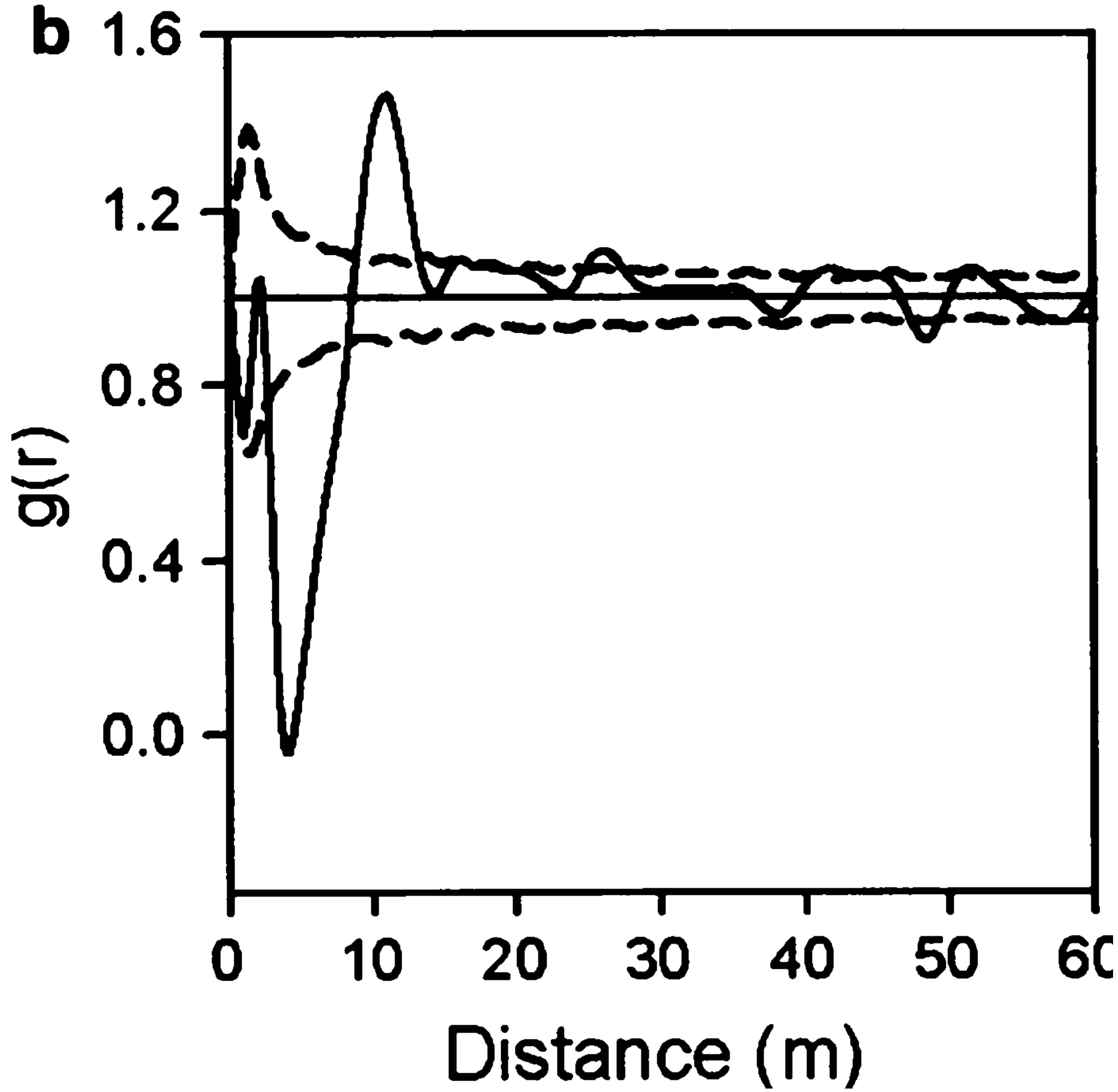
# g function for Juniper

- Recall that g(r) is a ‘ring’ function
  - g(r) = 1 under CSR
- How to interpret this pairwise correlation function: g(r)?
- How does it differ from the L function?
- What could lead to this pattern?



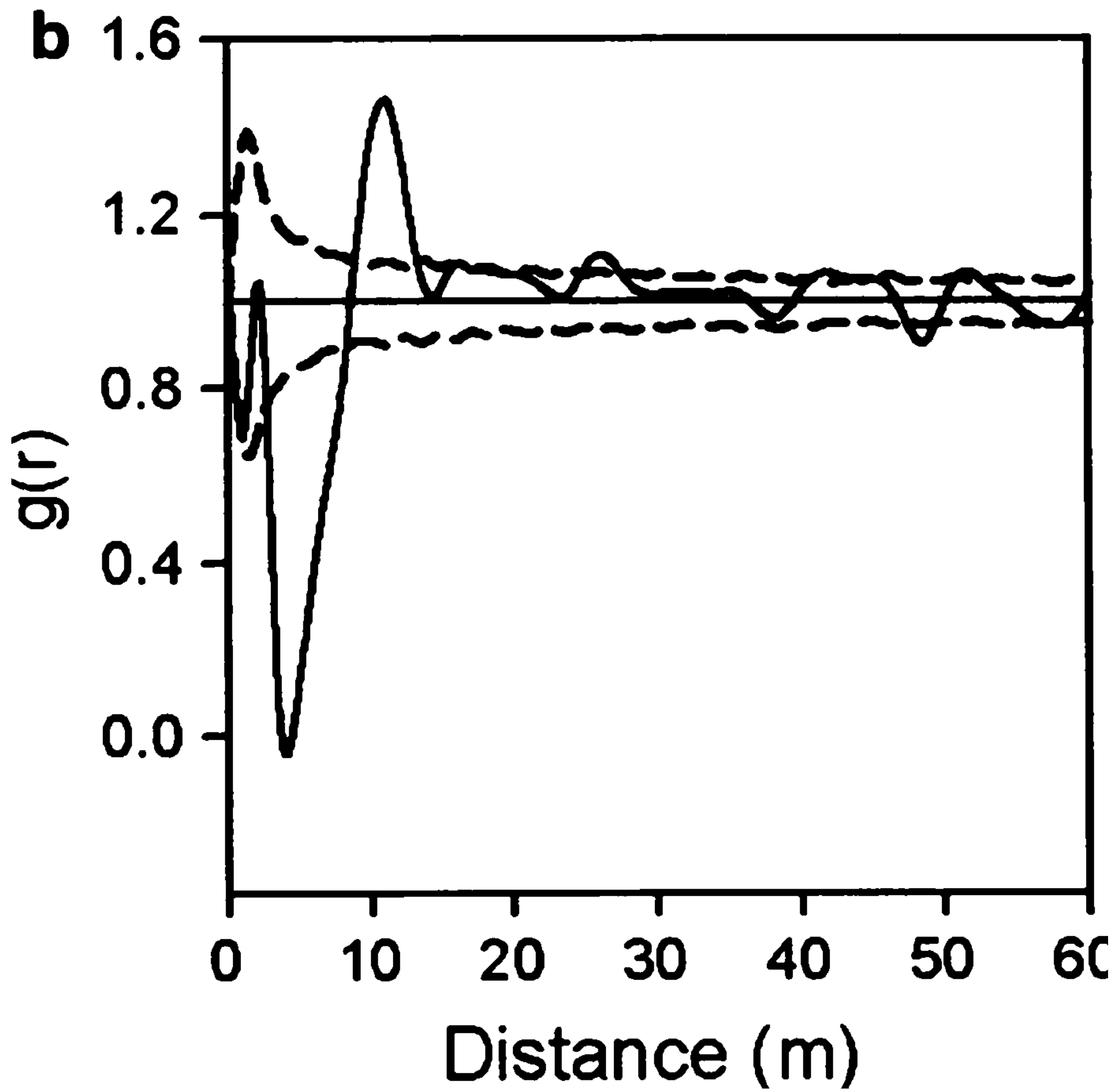
# g function for Juniper

- Recall that g(r) is a 'ring' function
  - g(r) = 1 under CSR
- How to interpret this pairwise correlation function: g(r)?
  - Amount of clustering at radius r
- How does it differ from the L function?
  - Not cumulative
- What could lead to this pattern?



# g function for Juniper

- Recall that g(r) is a ‘ring’ function
  - g(r) = 1 under CSR
- How to interpret this pairwise correlation function: g(r)?
  - Amount of clustering at radius r
- How does it differ from the L function?
  - Not cumulative
  - Fine-scale clustering
  - Moderate scale inhibition
- What could lead to this pattern?
  - Fine-scale competition for water/space?
  - Moderate scale parent-offspring effects?



# Tampa Neighborhoods – Hurricane Irma

- In-process research!
- Three Neighborhoods
  - Two are affluent/upper middle-class
    - Lots of single-family homes
  - One is 'up and coming'
    - Lots of rentals and apartments
- Most significant defect, pre-hurricane
  - Analyzed spatial patterns of pre-hurricane dead trees
- Type of failure post-hurricane
  - Branch failure
  - Whole tree failure

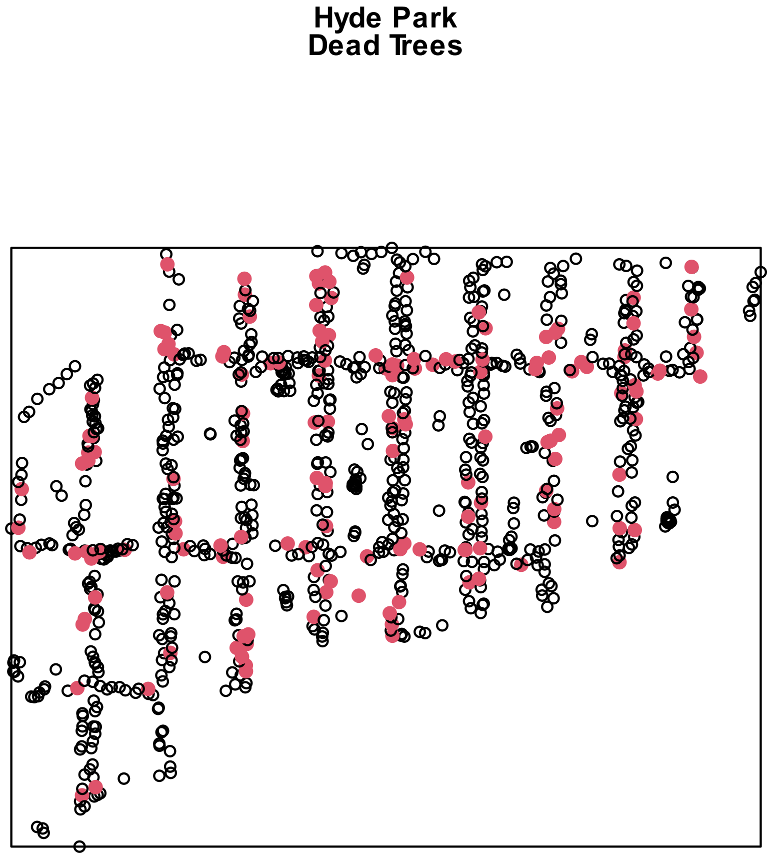
# Tampa Neighborhoods – Hurricane Irma

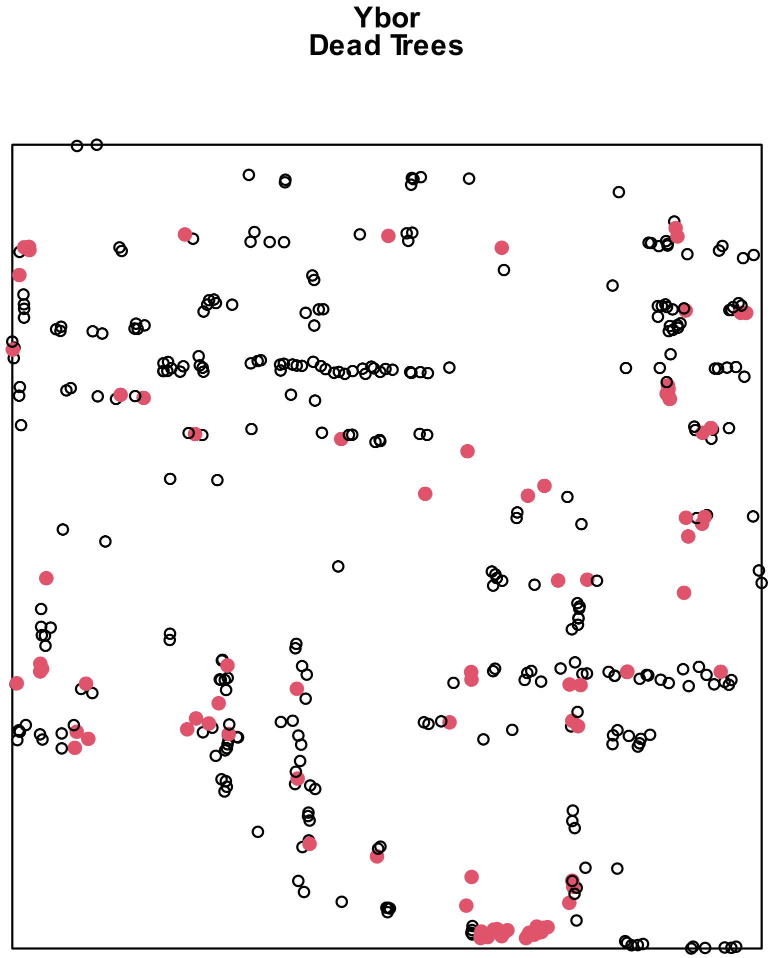
- Let's do some brainstorming:
- What factors might affect patterns of tree defects and failures?
  - Socioeconomic?
  - Hurricane?

# Tampa Neighborhoods – Hurricane Irma

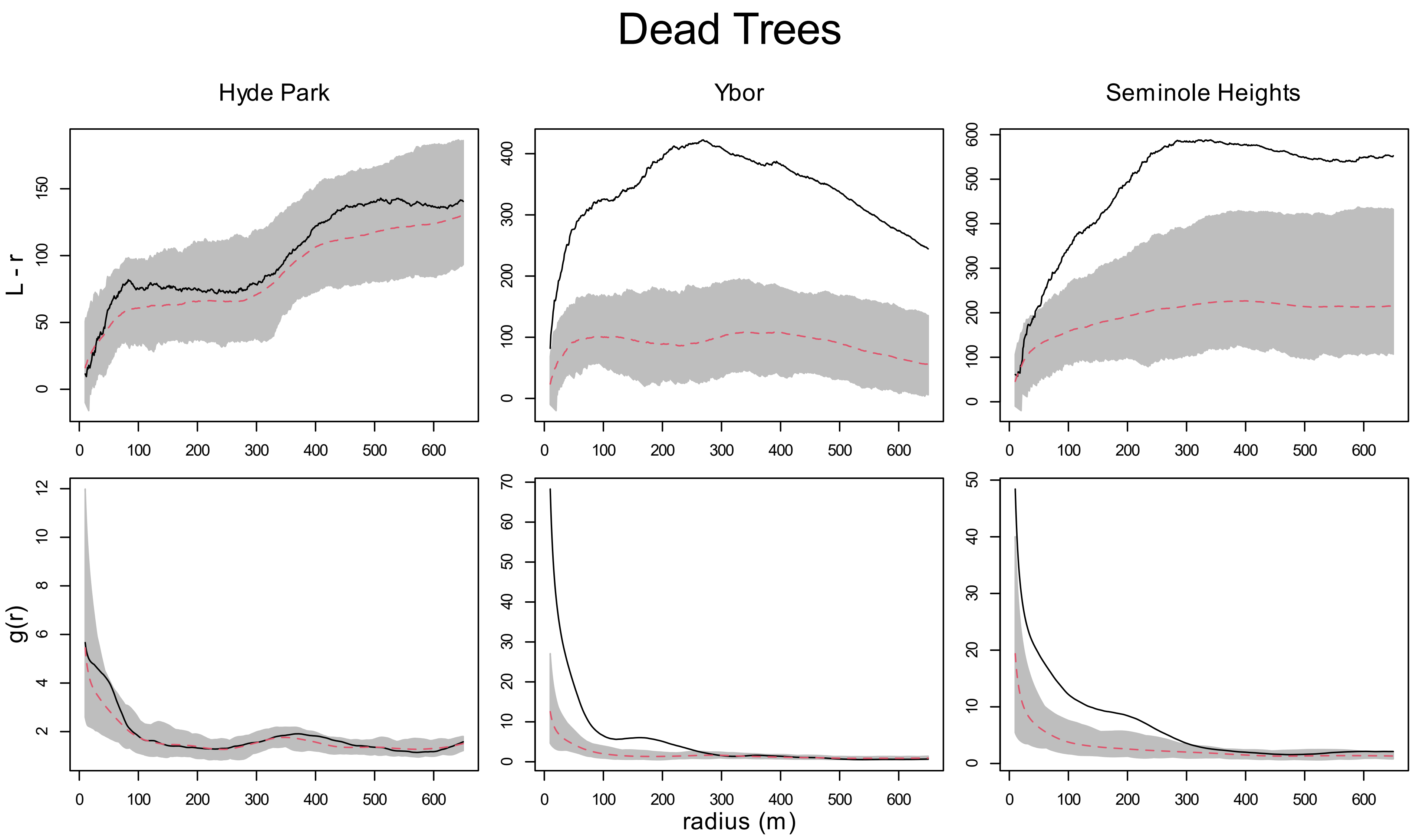
- Before you look at the figures, what do you predict?
- Pre-hurricane: point pattern of dead trees
  - Do you expect spatial clustering, why or why not?
  - Do you think the patterns will be similar in all the neighborhoods, why or why not?
- Post-hurricane: point pattern of tree failures
  - Do you think there'll be clustering?  What might cause it?

# Pre-Hurricane Dead Trees




# Univariate L and PCF for Dead Trees



# Post Hurricane Failures