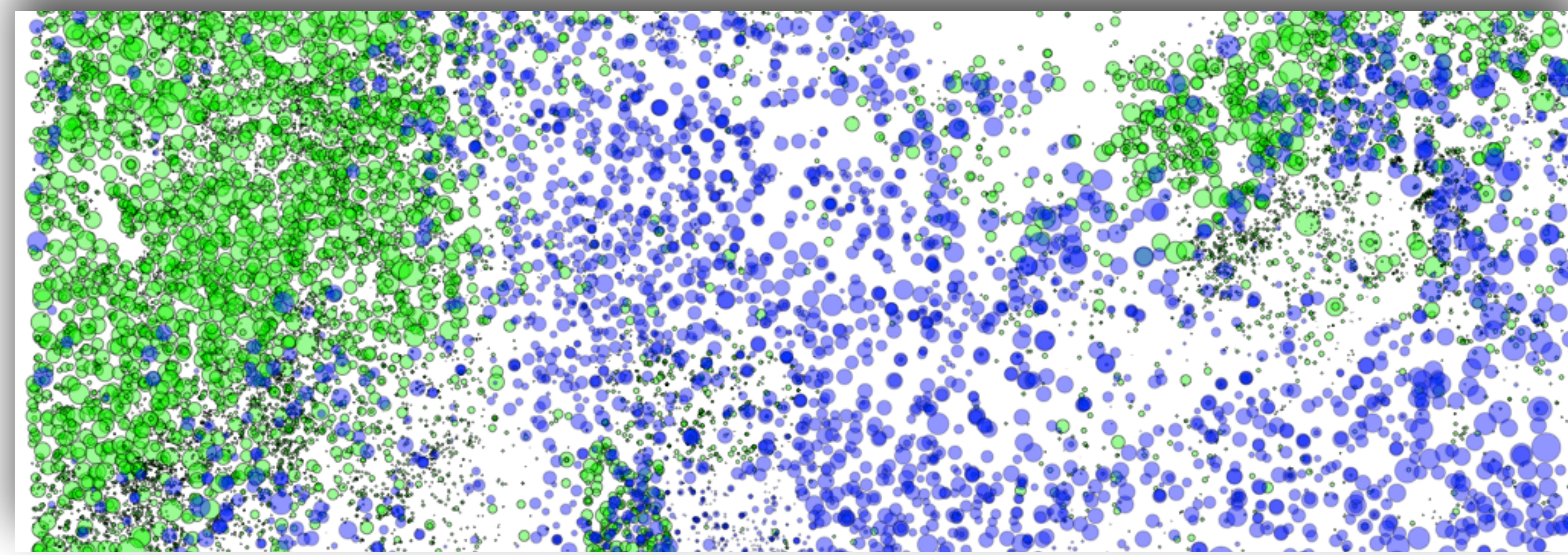
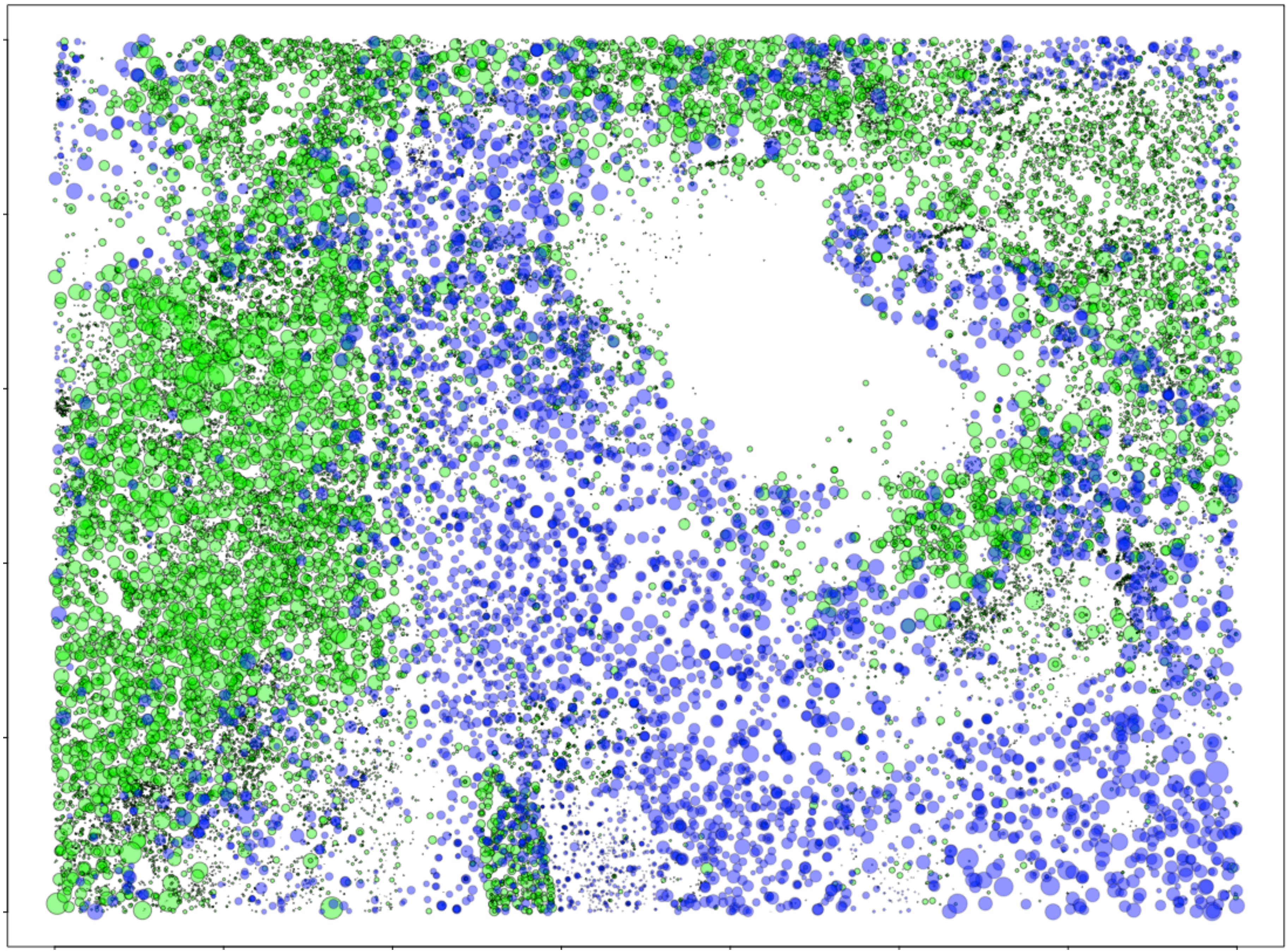
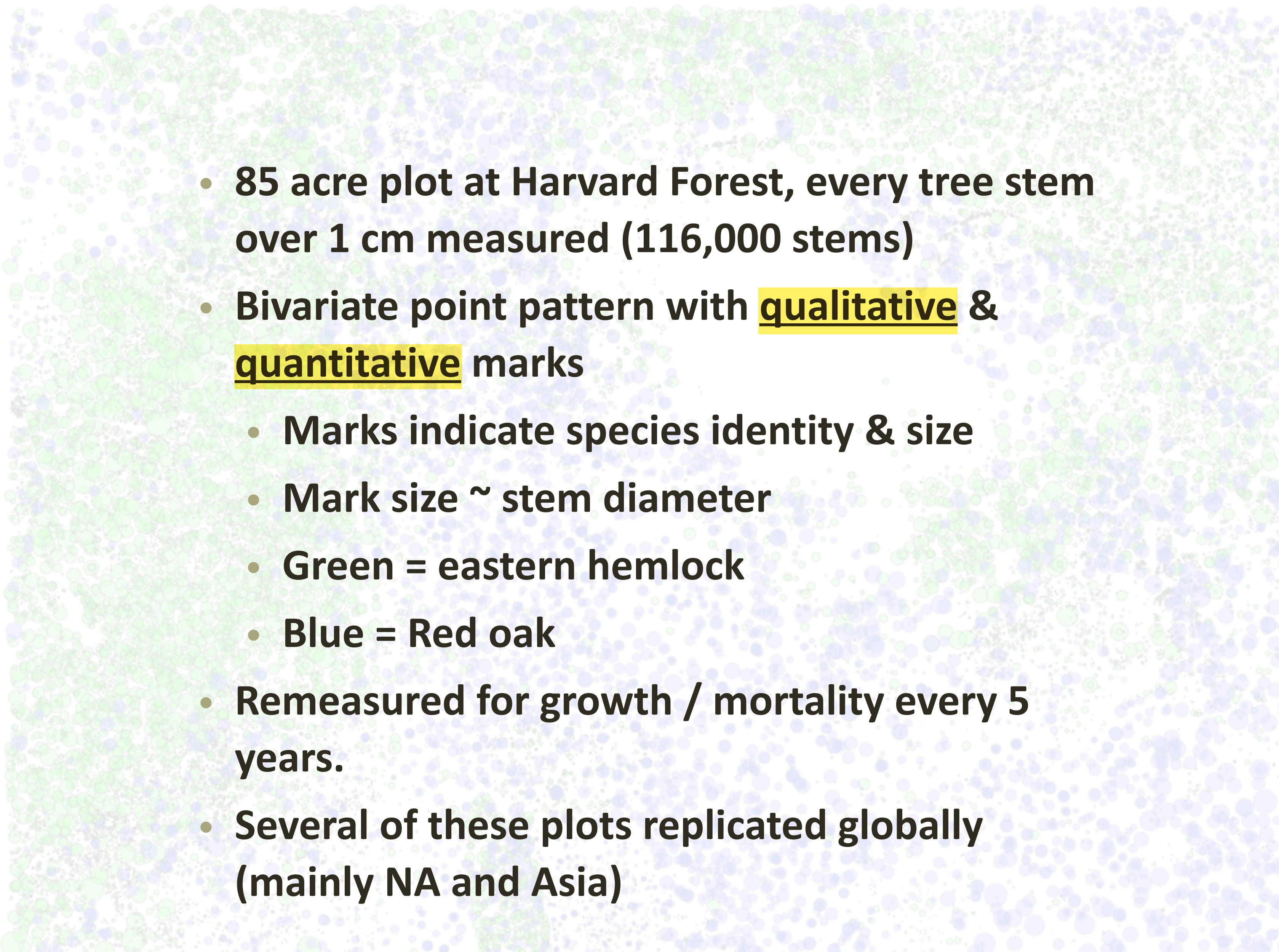
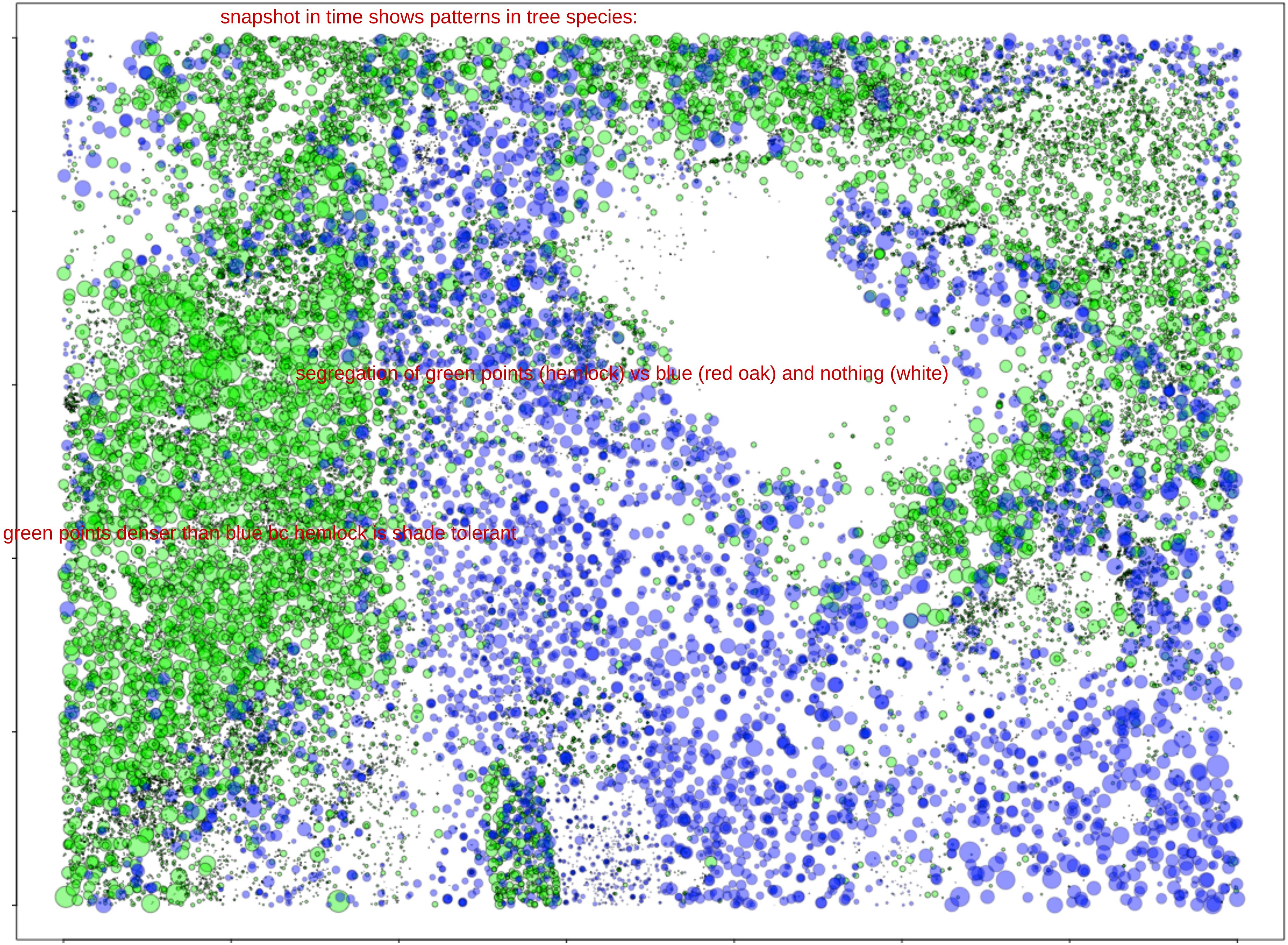
Inferring process from pattern: Point-pattern analysis





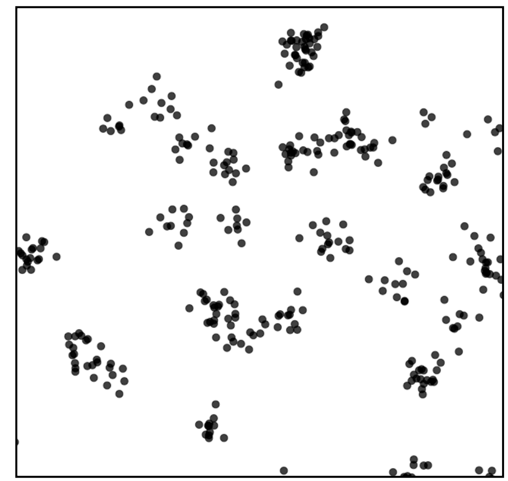


- 85 acre plot at Harvard Forest, every tree stem over 1 cm measured (116,000 stems)
- Bivariate point pattern with qualitative & quantitative marks
  - Marks indicate species identity & size
  - Mark size ~ stem diameter
  - Green = eastern hemlock
  - Blue = Red oak
- Remeasured for growth / mortality every 5 years.
- Several of these plots replicated globally (mainly NA and Asia)


snapshot in time shows patterns in tree species:
segregation of green points (hemlock) vs blue (red oak) and nothing (white)
green points denser than blue bc hemlock is shade tolerant

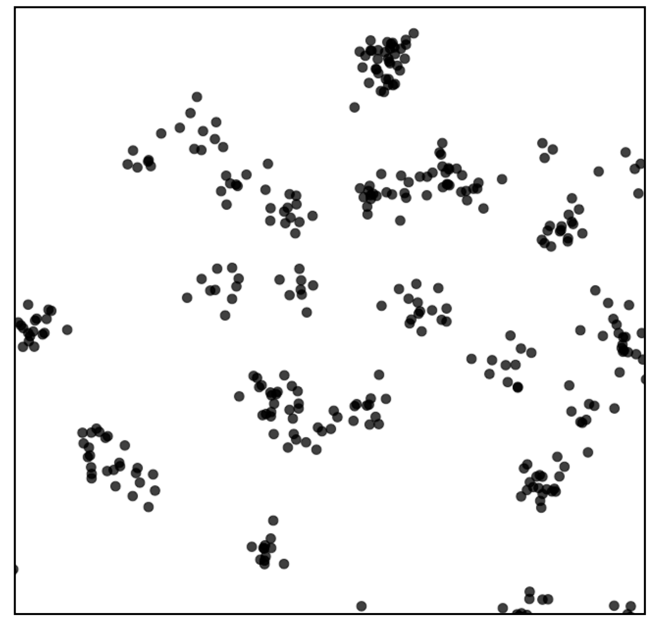
What is point-pattern analysis and why is it used?

- Point-pattern analysis is the statistical analysis of spatial point pattern data
- Locations of ecological objects in space
  - Trees in a forest
  - Occurrence of coral species
  - Gopher mounds in a prairie
  - Bird nest locations
  - lightening strikes, etc, etc
  - typically 2-D
- Spatial patterns conserve an imprint of past processes and therefore we may recover information on these processes through statistical analysis of the pattern



What is point-pattern analysis and why is it used?

- Primary question of interest: What are the potential ecological causes (i.e., processes) behind the observation distribution of the objects in space (i.e., patterns)?
- Two primary lines of research around this question:
  1. Point-pattern analysis (spatially explicit)
  2. Species distribution modeling (spatial implicit for the most part)



What is point-pattern analysis and why is it used?

- **Point-pattern analysis**
  - Environment is homogeneous & heterogeneity is seem as a complication (can deal with this using covariates, much like species distribution modeling) control for the role of the covariate controlling the spatial pattern differences
  - Focus is on small-scale interactions between points
  - Typically involves the use of “null” models
- **Species distribution modeling**
  - What determines the presence-absence / abundance of ecological objects in space?
  - The environment is heterogeneous and covariates can explain spatial distribution.
  - Interactions between objects (esp. autocorrelation) is typically ignored
  - Some of the most popular presence-only SDM methods (e.g., MaxEnt) are in essence ‘point process models’…

What is point-pattern analysis and why is it used?

- Spatial segregation hypothesis (Pacala 1997):
  - Seedlings of the same species tend to be aggregated (seed dispersal limitation)…
  - Therefore, there is interspecific segregation in plant communities, such that…
  - Most plants tend to compete with con-specific plants at a local scale…
  - This segregation of species contributes to coexistence by hindering competitive exclusion

Ecology, 91(7), 2010, pp. 2110–2120

Evidence for the spatial segregation hypothesis: a test with nine-year survivorship data in a Mediterranean shrubland

José Raventós,[1,4] Thorsten Wiegand,[2] and Martín De Luis[3]

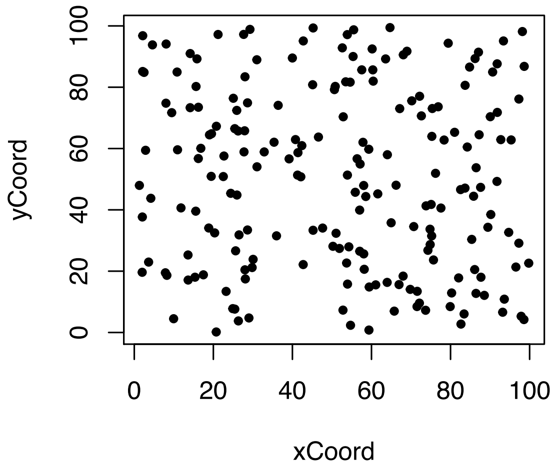
- Raventos et al (2010):
  - Long-term dataset of fully mapped seedling emergence and subsequent survival in a Mediterranean shrub land
  - Used Point-pattern analysis (PPA) to test the spatial segregation hypothesis
  - Seedling emergence: intraspecific aggregation and interspecific segregation
  - Survival: Used PPA to ask: are you more likely to survive if you are near a conspecific versus heterospecific neighbor?
  - Mortality controlled almost entirely by intraspecific interactions

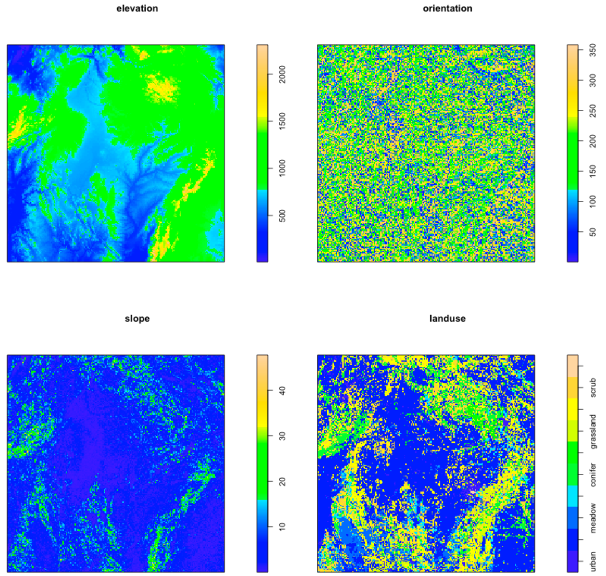
seedlings near the parent plants lead same species to be near each other

survival was almost entirely based on proximity to nearest competing individuals

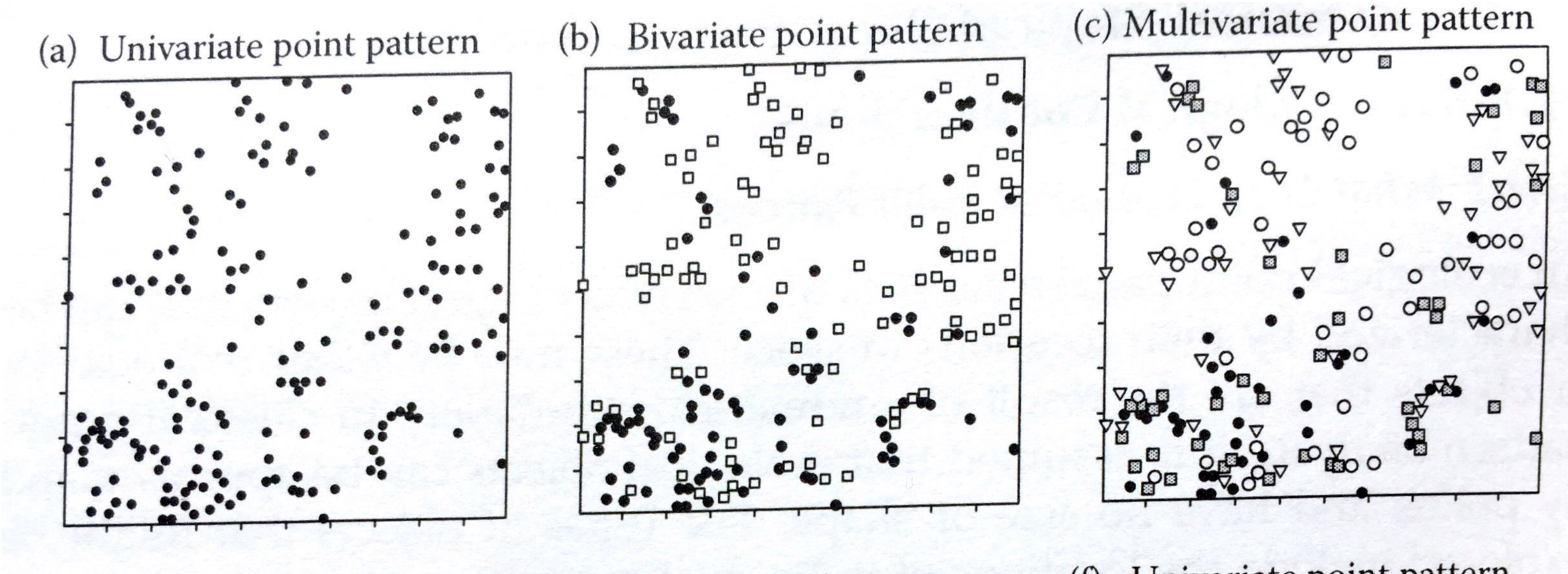
Spatial point patterns

- A point pattern gives the *locations* of data (termed "events") occurring in a region
- **Points represent a sample from some generating *point process***
- Often the events have *attributes* (termed *marks*) associated with them
  - Categorical / qualitative marks (species, status)
  - Continuous / quantitative marks (tree diameter, age, etc.)
- May also include *covariates*

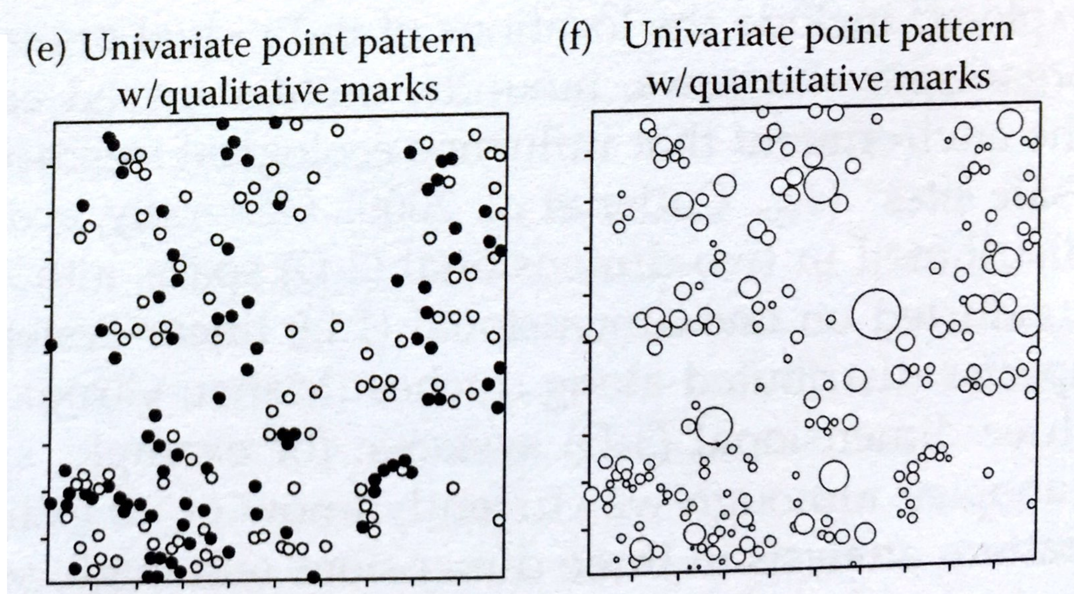




Spatial point patterns



- Univariate - all points are of the same type, only location matters
- Bi - and multi-variate point patterns - relationships between two or more types of objects



Wiegand & Moloney

Basic steps in point-pattern analysis

1. Determine the data type

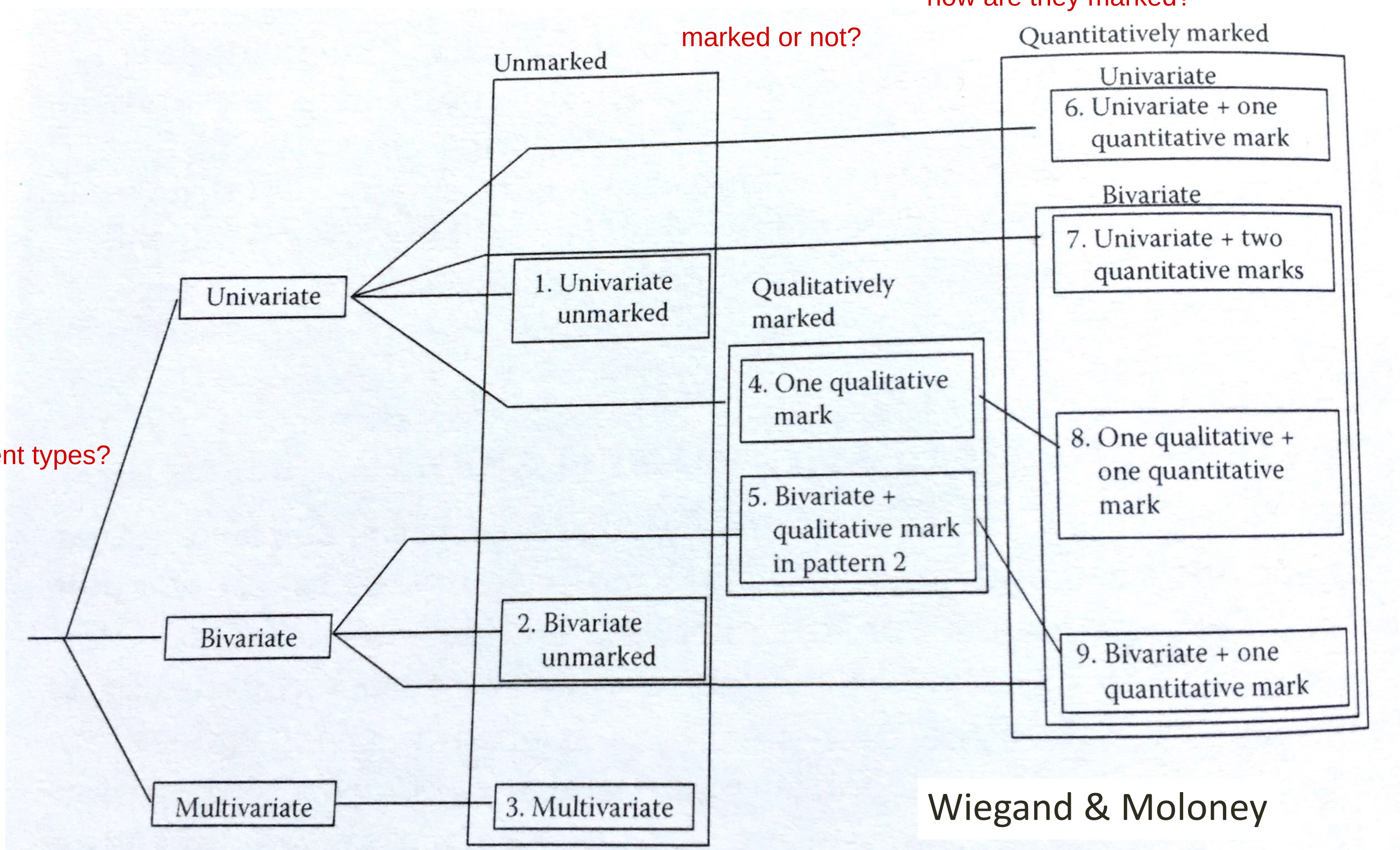
Basic steps in point-pattern analysis

1. Determine the data type

how are they marked?

marked or not?

how many event types?



Wiegand & Moloney

Basic steps in point-pattern analysis

1. Determine the data type & decide whether it is *homogeneous* or *heterogeneous*
   - *homogenous* - environmental conditions and processes influencing the pattern are the same everywhere within the observation window
   - *heterogenous* - the pattern results from environmental heterogeneity or a process that is not constant across the study window.

Basic steps in point-pattern analysis

1. Determine the data type & decide whether it is *homogeneous* or *heterogeneous*
2. Select appropriate summary statistic(s)
   - some are ad hoc methods with little statistical theory to support them (e.g., average distance from a point to its nearest neighbor)

Basic steps in point-pattern analysis

1. Determine the data type & decide whether it is *homogeneous* or *heterogeneous*
2. Select appropriate summary statistic(s)
3. Select appropriate *null models* and *point-process models*
   - a *null model* tests whether there is pattern in the data beyond that generated by purely random processes (e.g., Poisson Point Process)
   - *point-process models:* Formulate and fit a statistical model to a point pattern to represent complex spatial structures or a hypothesized process

bring in covariates to try to control for underlying structure

Basic steps in point-pattern analysis

1. Determine the data type & decide whether it is *homogeneous* or *heterogeneous*
2. Select appropriate summary statistic(s)
3. Select appropriate null models and point-process models
4. Compare observed data to null model
    - the summary statistic serves as the test statistic and is calculated for both the observed (data) the simulated patterns (null or process model)

Basic steps in point-pattern analysis

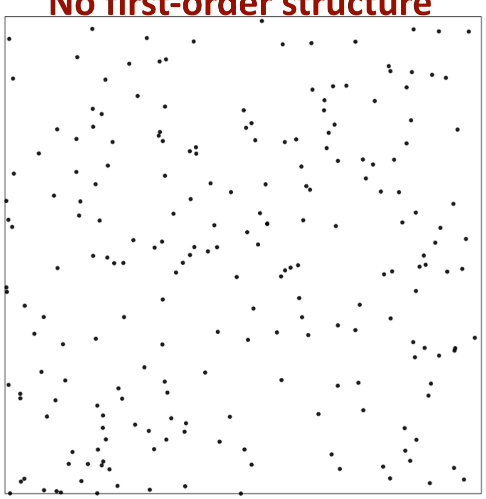
1. Determine the data type & decide whether it is *homogeneous* or *heterogeneous*
2. Select appropriate summary statistic(s)
3. Select appropriate null models and point-process models
4. Compare observed data to null model
    - the summary statistic serves as the test statistic and is calculated for both the observed (data) the simulated patterns (null or process model)

THIS CAN BE CONSIDERED A GENERAL STRATEGY FOR ANY PROCESS-PATTERN ANALYSIS

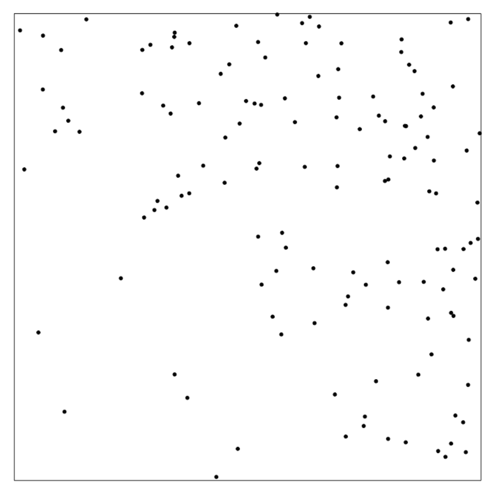
Characterizing spatial pattern

- **First-order structure**
  - General, global trends - e.g., variation in ***intensity*** (density of points) across space
  - Location of any individual event is independent of the location of other events
  - A point pattern **WITH** first-order structure is called *non-stationary* or *inhomogeneous*

No first-order structure



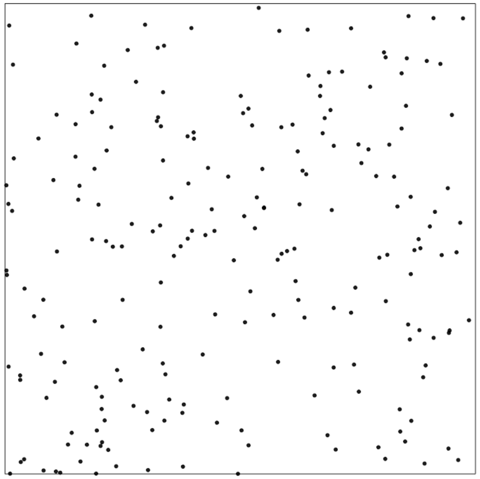
First-order structure



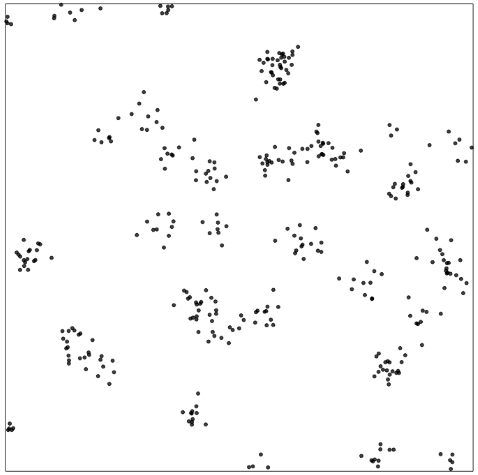
Characterizing spatial pattern

- **Second-order structure**
  - ***Interactions*** between events
  - Location of an event is NOT independent of the location of other events
    - competition between individuals (segregation)
    - Dispersal (aggregation)

No first- or second-order structure



Second-order structure



Spatial point patterns

- Many questions can be asked regarding point patterns
  1. **Intensity**: Expected number of points per unit area (constant versus inhomogeneous)
     - the analog of the mean or expected value of random variable
     - Summary statistic: *intensity function*: $\lambda(x)$ = mean number of points per unit area (n/A)
     - May be constant (uniform/homogeneous) or vary from location to location (inhomogeneous) - in other words: Is there first-order structure?
     - Investigating intensity should be one of the first steps in analyzing point patterns

Spatial point patterns

- Many questions can be asked regarding point patterns
  1. Intensity: Expected number of points per unit area (constant versus inhomogeneous)
  2. **Interactions**: dependence between points in a point pattern
     - Is there second-order structure?
       - *Is the spacing between tree saplings greater than would be expected for a random pattern?*
     - Segregation of points with different attributes
       - *Is there spatial variation in the density and age of trees?*
     - Dependence between points of different types
       - *Does species A intentionally live close to species B?*
     - Many possible second-order summary statistics
       - depends on data type, hypotheses to be tested, and nature of the comparisons to be made
     - Summary stats examining interactions between points are based on the spatial relationships between ***pairs of points***

Spatial point patterns

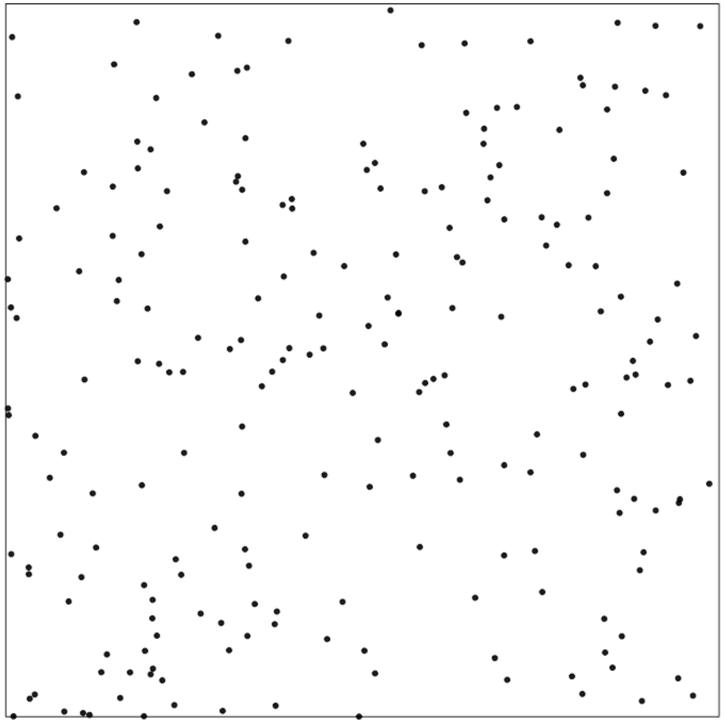
- Many questions can be asked regarding point patterns
  1. Intensity: Expected number of points per unit area (constant versus inhomogeneous)
  2. Interactions: dependence between points in a point pattern
  3. **Covariate effects**: can influence intensity (first order effect in most cases - but need to allow for covariate effects BEFORE studying interactions between points).
     - Does tree density depend on elevation / soil type / etc?
     - After accounting for variation in tree density due to elevation, is there evidence for clustering of trees?

Spatial point patterns

- Summary statistics can be classified two ways:
  1. **Event-related** - describe spatial structure of the pattern from the viewpoint of the events themselves and summarize properties of the neighborhood around each event. (think “plants-eye view” of a forest).
  2. **Location-related** - describe spatial structure of the pattern from the viewpoint of *test locations (not the events themselves)*, which are placed within the observation window independently of the events in the pattern.

Point-pattern analysis

- The most basic point pattern we want to understand is one in which locations of points are generated without any underlying process - i.e., "at random" (no first- OR second-order structure). We call this (1) complete spatial randomness ~~(CRS)~~ (CSR) or a (2) homogenous / uniform Poisson point process (HPP).

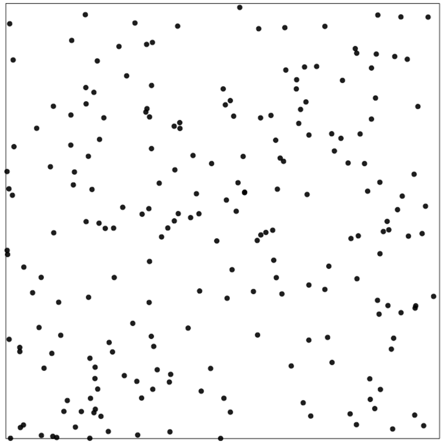


Point-pattern analysis

- The most basic point pattern we want to understand is one in which locations of points are generated without any underlying process - i.e., "at random" (no first- OR second-order structure). We call this (1) complete spatial randomness (CRS) or a (2) homogenous / uniform Poisson point process (HPP).
  - Point process: a stochastic mechanism (process) that generates a countable set of events
  - the NUMBER of events is random, as is their LOCATION
  - Events are distributed such that there are no regions where the events are more/less likely to occur (no first-order structure) and the presence of a given event does not modify the probability of other events appearing nearby (no second-order structure).
  - Stationary (constant intensity) & isotropic (invariant to rotation)

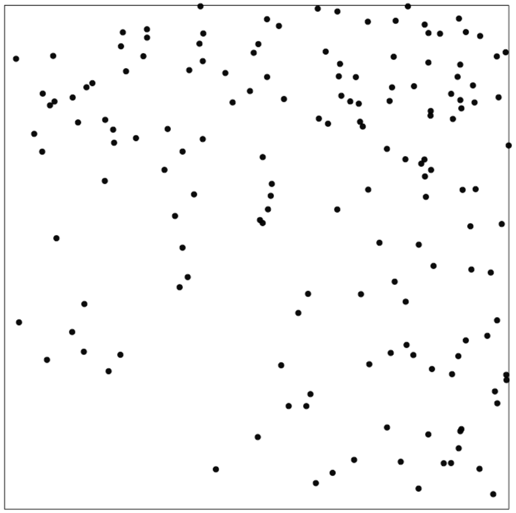
The **Homogeneous** Poisson Point Process

- Emits a Poisson-distributed number of points with mean λ over area A in an N-dimensional space (N is almost aways = 2).
- The intensity λ of the process is constant and follows a Poisson distribution with an expected mean of λA.
- Coordinates in each of these N dimensions are uniformly distributed and independent of other dimensions (in short, no interactions between points).



The **Inhomogeneous** Poisson Point Process

- ‘Heterogeneous’ extension of CSR that allows for non-constant intensity (due to covariates)
- Independence between events still holds (i.e., no interactions between points), but events are more likely in some areas than others.
- For example, assume seeds are randomly dispersed according to a Poisson process, and seeds randomly germinate independently of each other, with germination probability depending on a gradient in local soil conditions. The resulting pattern of plants is an inhomogeneous Poisson process.

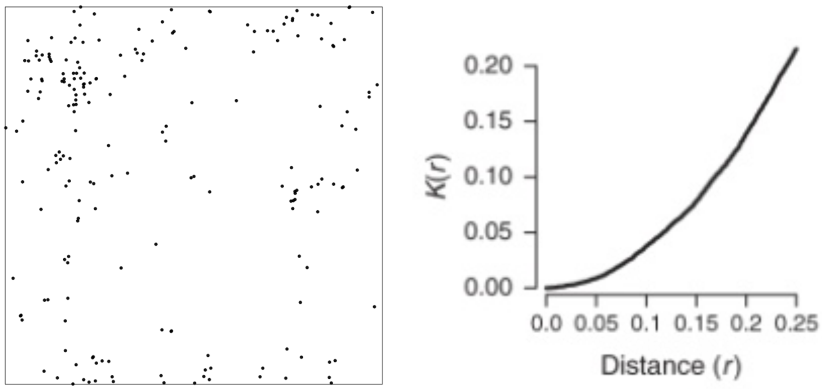


Point-pattern analysis

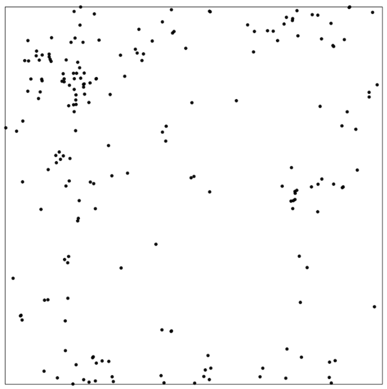
- CSR/HPP/IPP are “null models” for spatial point patterns
- We can compare our data to these expected distributions under a ‘random process’ to identify dispersion or clustering, etc.
- Goal is to estimate parameters of the most likely statistical distribution that generated the pattern.
- All point pattern analyses require a WINDOW
- The analysis window sets the extent (scale) of the analysis and therefore will influence the perceived pattern.

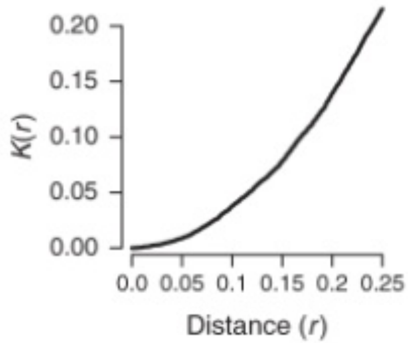
Null model example

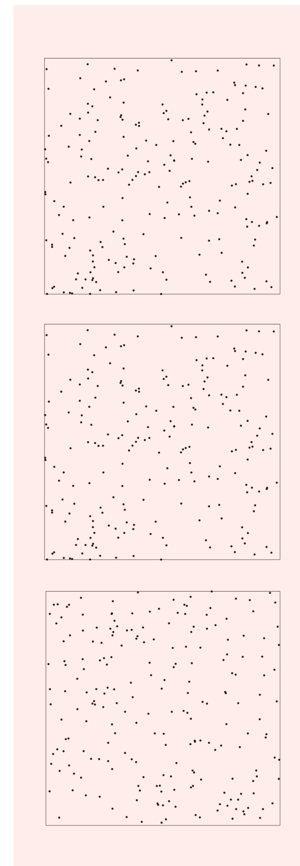
(1) Fit summary statistic to data



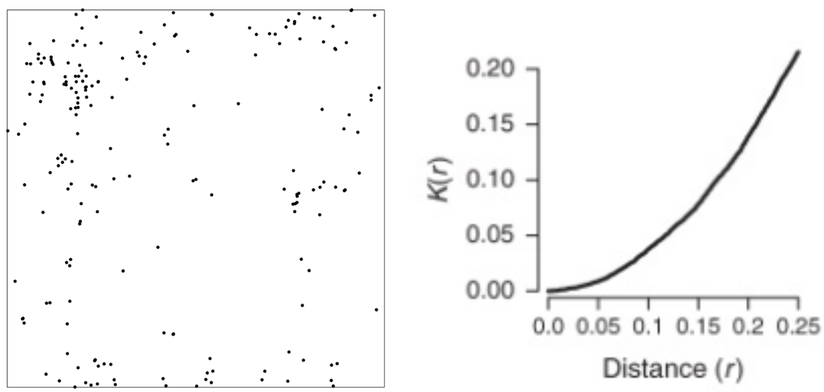
(2) Simulate the **null pattern** many times (~100+)

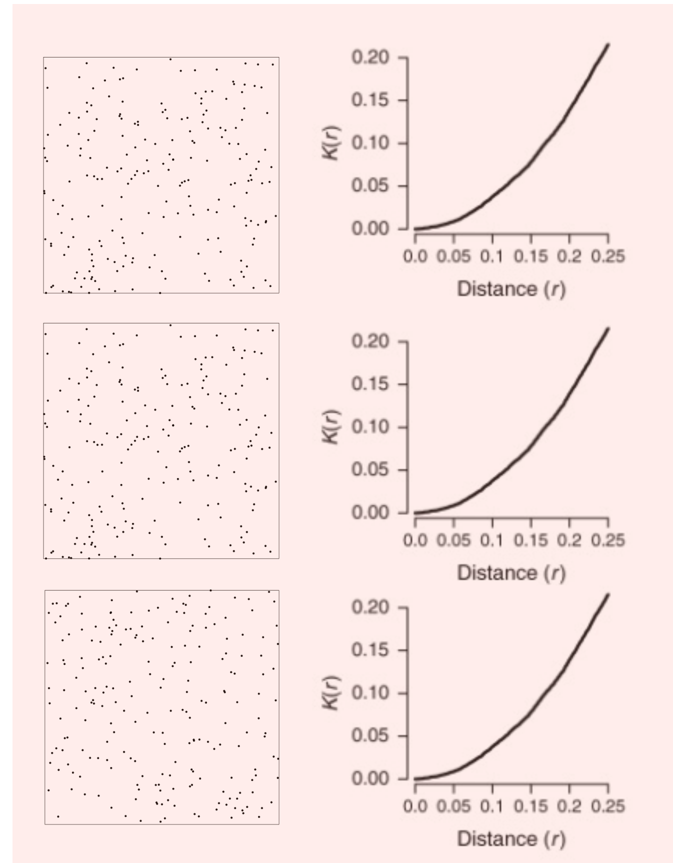




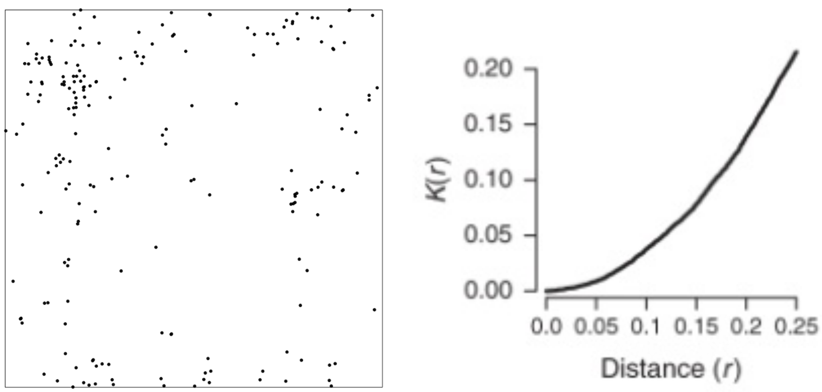


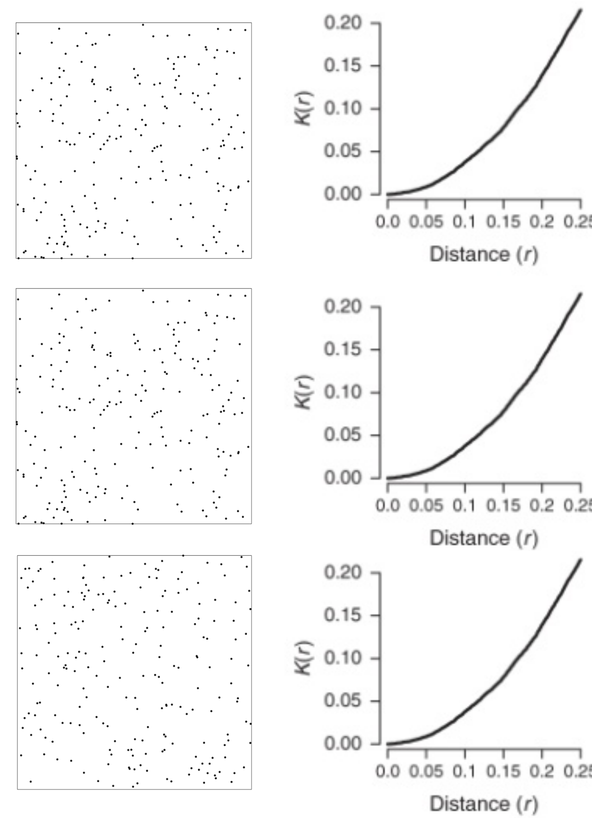
(3) Fit summary stat to *simulated* null data

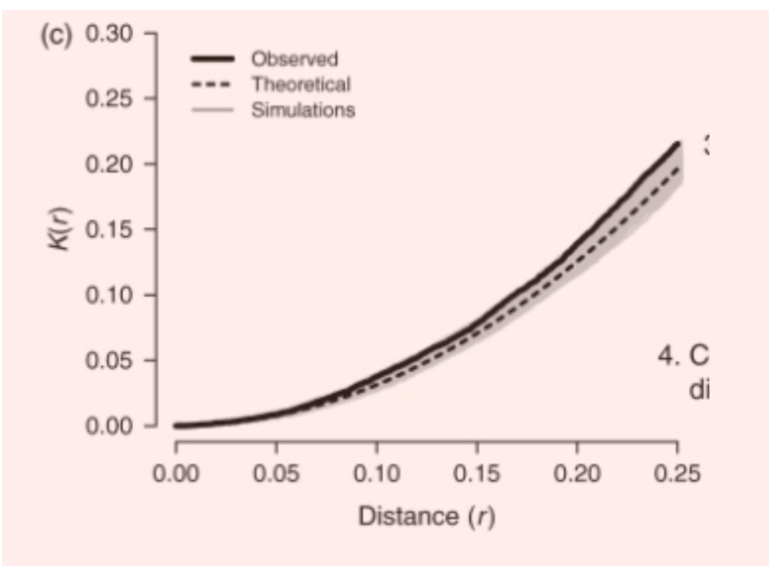




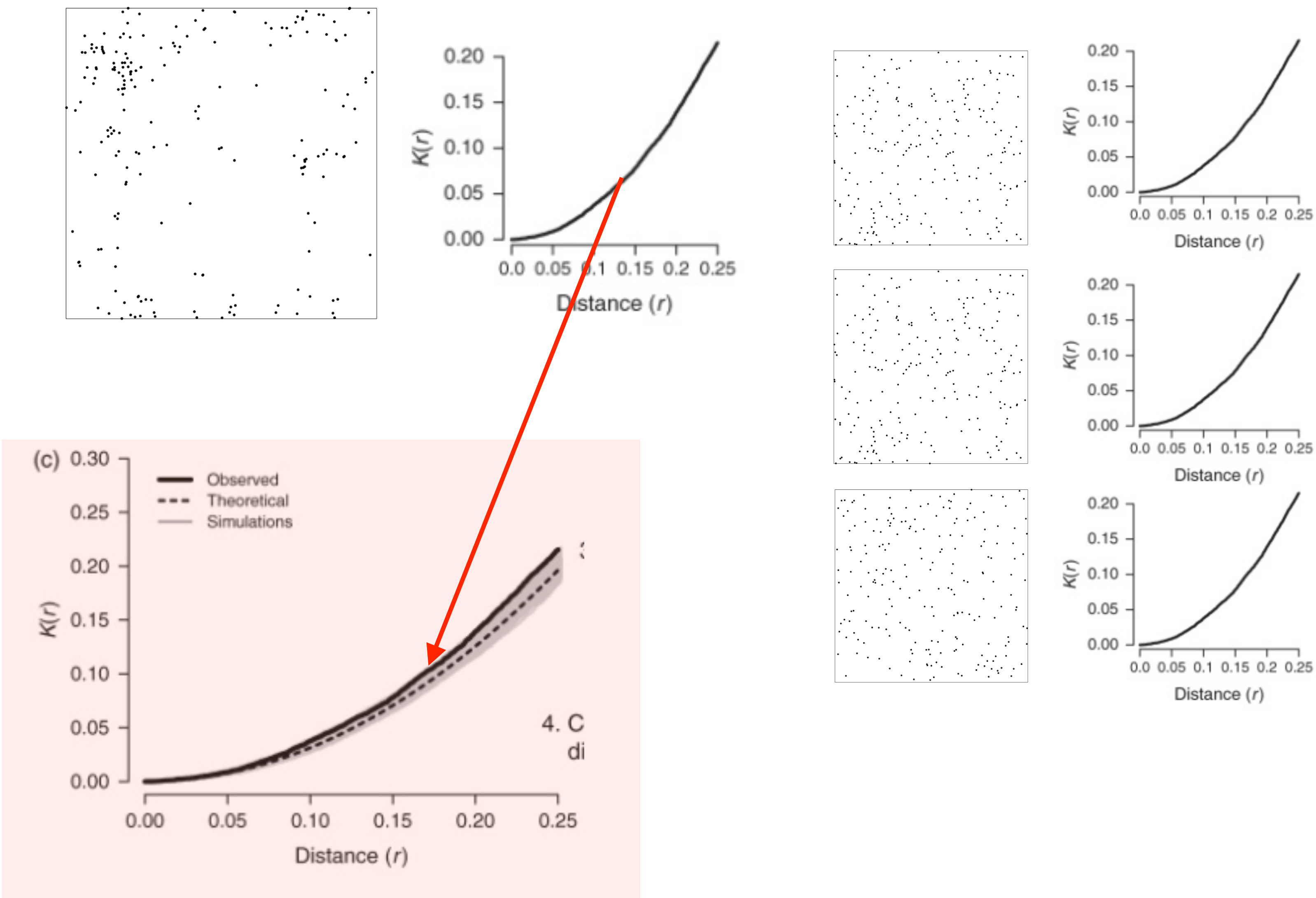
(4) Compare observed value with distribution from null model simulations



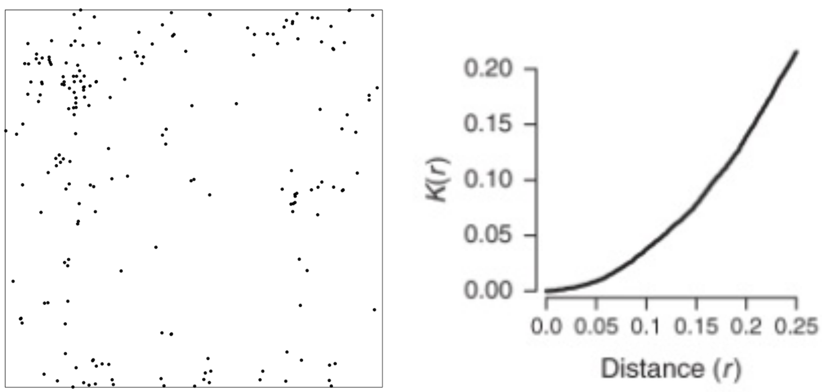


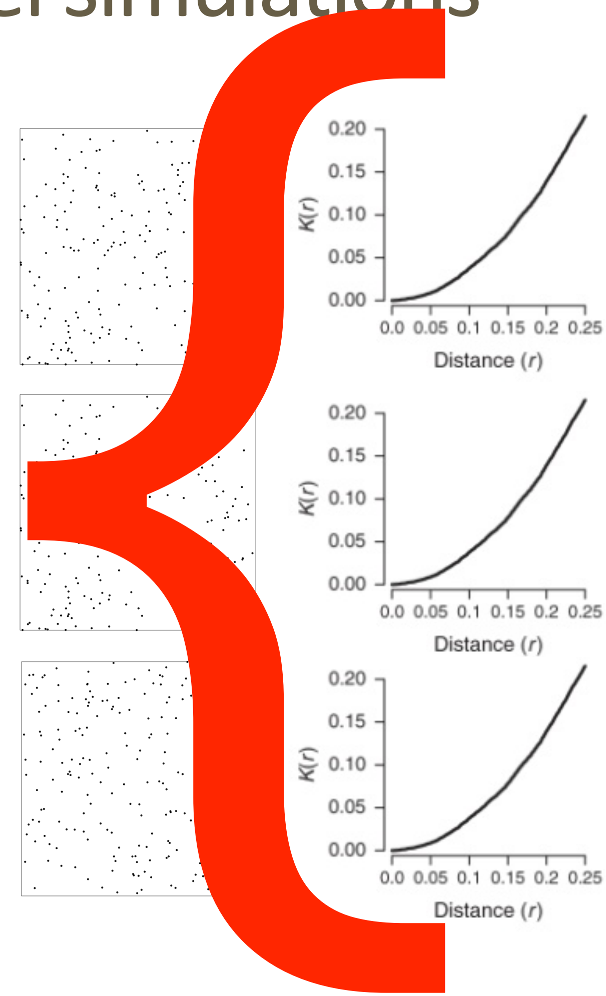


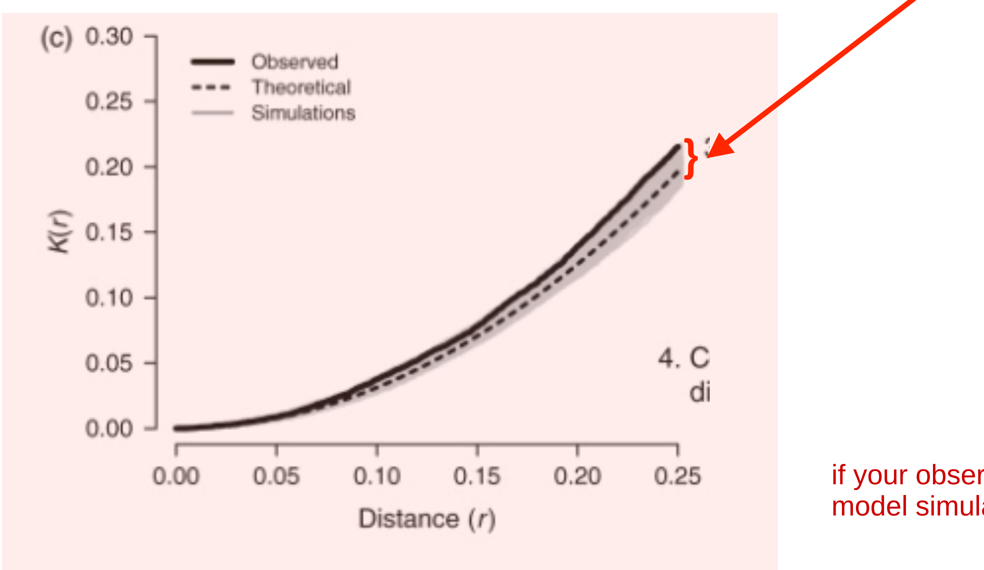
(4) Compare observed value with distribution from null model simulations



(4) Compare observed value with distribution from null model simulations



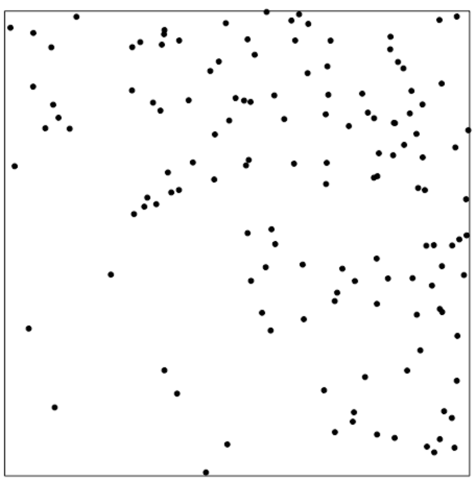




if your observed line is significantly below the lines predicted by the base model simulations, it is likely spatially structured

Point-pattern analysis - Tests for CSR

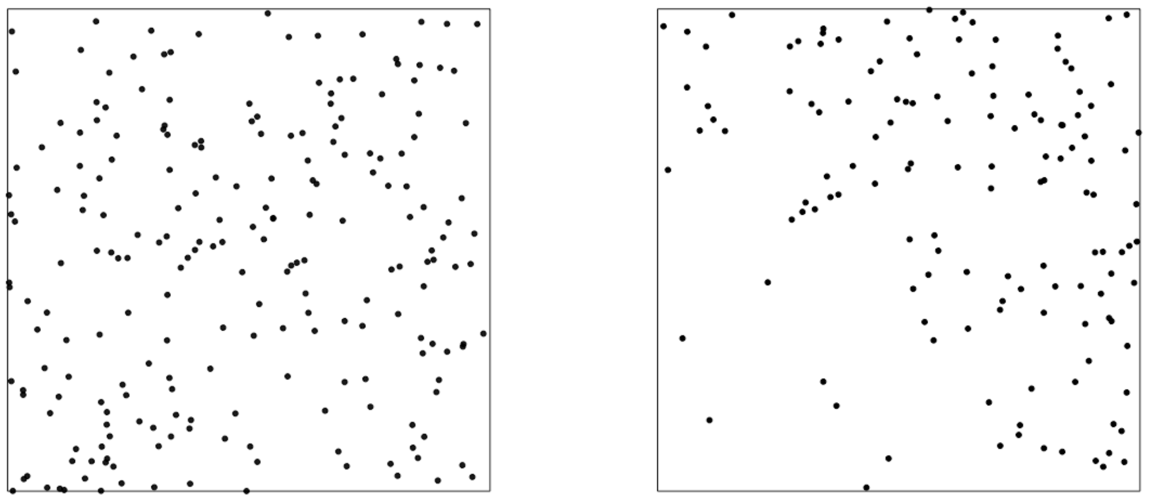
- Quadrat counting test (χ^2) then use a Chi Squared test to tell you whether the points depart from randomness
  - Lacks information



5	9	10	9	8
5	4	6	9	14
0	5	5	3	12
1	1	2	4	8
1	1	2	4	8

Point-pattern analysis - Tests for CSR

- Quadrat counting test (χ^2)
  - Lacks information



Point-pattern analysis - Tests for CSR

- Quadrat counting test (χ^2)
  - Lacks information

5 8.7 -1.2	8 8.7 -0.23	3 8.7 -1.9	8 8.7 -0.23	9 8.7 0.11
10 8.7 0.45	9 8.7 0.11	9 8.7 0.11	12 8.7 1.1	6 8.7 -0.91
8 8.7 -0.23	12 8.7 1.1	12 8.7 1.1	5 8.7 -1.2	12 8.7 1.1
9 8.7 0.11	6 8.7 -0.91	6 8.7 -0.91	7 8.7 -0.57	10 8.7 0.45
16 8.7 2.5	17 8.7 2.8	5 8.7 -1.2	5 8.7 -1.2	8 8.7 -0.23

5 5.4 -0.19	9 5.4 1.5	10 5.4 2	9 5.4 1.5	8 5.4 1.1
5 5.4 -0.19	4 5.4 -0.62	6 5.4 0.24	9 5.4 1.5	14 5.4 3.7
0 5.4 -2.3	5 5.4 -0.19	5 5.4 -0.19	3 5.4 -1	12 5.4 2.8
1 5.4 -1.9	1 5.4 -1.9	2 5.4 -1.5	4 5.4 -0.62	8 5.4 1.1
1 5.4 -1.9	1 5.4 -1.9	2 5.4 -1.5	4 5.4 -0.62	8 5.4 1.1

Point-pattern analysis - Tests for CSR

- Quadrat counting test (χ^2)
  - Lacks information

# of points

5 8.7 -1.2	8 8.7 -0.23	3 8.7 -1.9	8 8.7 -0.23	9 8.7 0.11
10 8.7 0.45	9 8.7 0.11	9 8.7 0.11	12 8.7 1.1	6 8.7 -0.91
8 8.7 -0.23	12 8.7 1.1	12 8.7 1.1	5 8.7 -1.2	12 8.7 1.1
9 8.7 0.11	6 8.7 -0.91	6 8.7 -0.91	7 8.7 -0.57	10 8.7 0.45
16 8.7 2.5	17 8.7 2.8	5 8.7 -1.2	5 8.7 -1.2	8 8.7 -0.23

5 5.4 -0.19	9 5.4 1.5	10 5.4 2	9 5.4 1.5	8 5.4 1.1
5 5.4 -0.19	4 5.4 -0.62	6 5.4 0.24	9 5.4 1.5	14 5.4 3.7
0 5.4 -2.3	5 5.4 -0.19	5 5.4 -0.19	3 5.4 -1	12 5.4 2.8
1 5.4 -1.9	1 5.4 -1.9	2 5.4 -1.5	4 5.4 -0.62	8 5.4 1.1
1 5.4 -1.9	1 5.4 -1.9	2 5.4 -1.5	4 5.4 -0.62	8 5.4 1.1

Point-pattern analysis - Tests for CSR

- Quadrat counting test (χ^2)
  - Lacks information

Expected # of points

under a uniform process

5 8.7 -1.2	8 8.7 -0.23	3 8.7 -1.9	8 8.7 -0.23	9 8.7 0.11
10 8.7 0.45	9 8.7 0.11	9 8.7 0.11	12 8.7 1.1	6 8.7 -0.91
8 8.7 -0.23	12 8.7 1.1	12 8.7 1.1	5 8.7 -1.2	12 8.7 1.1
9 8.7 0.11	6 8.7 -0.91	6 8.7 -0.91	7 8.7 -0.57	10 8.7 0.45
16 8.7 2.5	17 8.7 2.8	5 8.7 -1.2	5 8.7 -1.2	8 8.7 -0.23

5 5.4 -0.19	9 5.4 1.5	10 5.4 2	9 5.4 1.5	8 5.4 1.1
5 5.4 -0.19	4 5.4 -0.62	6 5.4 0.24	9 5.4 1.5	14 5.4 3.7
0 5.4 -2.3	5 5.4 -0.19	5 5.4 -0.19	3 5.4 -1	12 5.4 2.8
1 5.4 -1.9	1 5.4 -1.9	2 5.4 -1.5	4 5.4 -0.62	8 5.4 1.1
1 5.4 -1.9	1 5.4 -1.9	2 5.4 -1.5	4 5.4 -0.62	8 5.4 1.1

Point-pattern analysis - Tests for CSR

- Quadrat counting test (χ^2)
  - Lacks information

5 8.7 -1.2	8 8.7 -0.23	3 8.7 -1.9	8 8.7 -0.23	9 8.7 0.11
10 8.7 0.45	9 8.7 0.11	9 8.7 0.11	12 8.7 1.1	6 8.7 -0.91
8 8.7 -0.23	12 8.7 1.1	12 8.7 1.1	5 8.7 -1.2	12 8.7 1.1
9 8.7 0.11	6 8.7 -0.91	6 8.7 -0.91	7 8.7 -0.57	10 8.7 0.45
16 8.7 2.5	17 8.7 2.8	5 8.7 -1.2	5 8.7 -1.2	8 8.7 -0.23

chi squared residual

χ^2 residual

5 5.4 -0.19	9 5.4 1.5	10 5.4 2	9 5.4 1.5	8 5.4 1.1
5 5.4 -0.19	4 5.4 -0.62	6 5.4 0.24	9 5.4 1.5	14 5.4 3.7
0 5.4 -2.3	5 5.4 -0.19	5 5.4 -0.19	3 5.4 -1	12 5.4 2.8
1 5.4 -1.9	1 5.4 -1.9	2 5.4 -1.5	4 5.4 -0.62	8 5.4 1.1
1 5.4 -1.9	1 5.4 -1.9	2 5.4 -1.5	4 5.4 -0.62	8 5.4 1.1

so there is spatial structure to this data

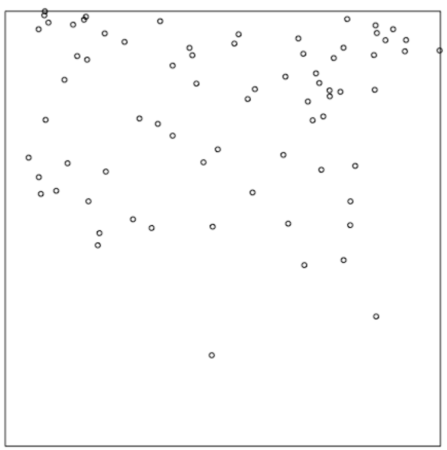
Point-pattern analysis - Tests for CSR

- Kolmogorov-Smirnov test of CSR
  - Compares observed & expected distributions of the values of a spatial *covariate*
  - Can also be used to test for goodness-of-fit for spatial models

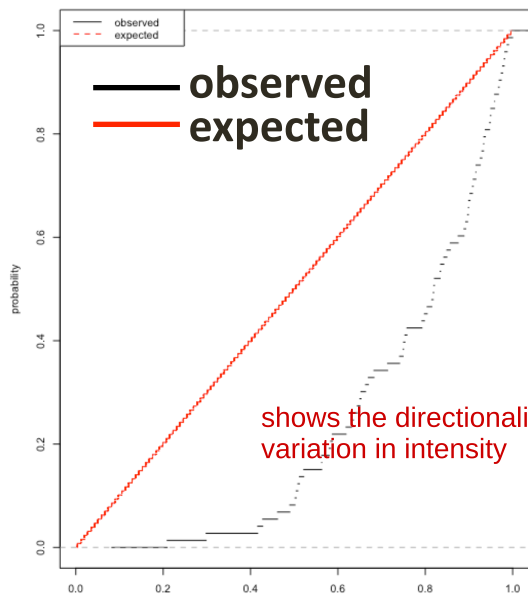
Point-pattern analysis - Tests for CSR

- Kolmogorov-Smirnov test of CSR
  - Compares observed & expected (theoretical) distributions of the values of a spatial *covariate*
  - Can also be used to test for goodness-of-fit for spatial models

Inhomogeneous Poisson Process (IPP)

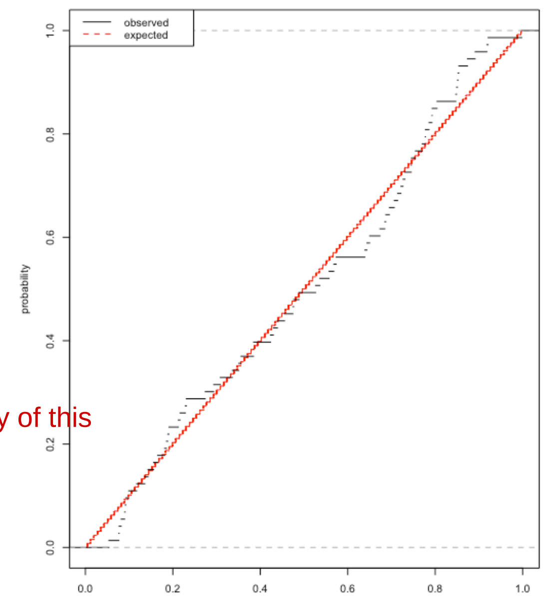


K-S test, y-direction



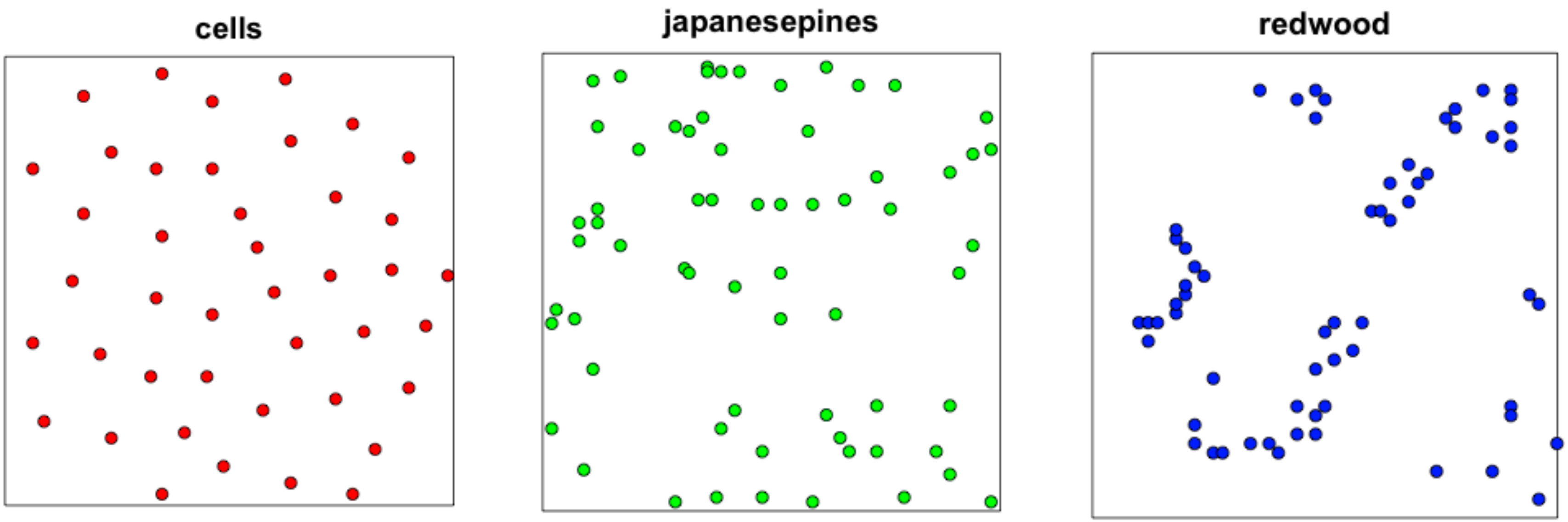
shows the directionality of this variation in intensity

K-S test, x-direction



Point-pattern analysis – Distance methods

- Distance methods try to answer: Are there interactions between points?
  - empty space distances
  - nearest neighbor distances
  - pairwise distances



more segregated than normal
ie. evenly spaced

maybe first order structure unclear

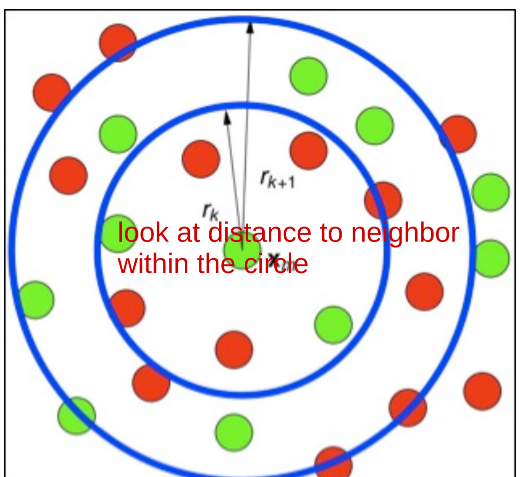
possible clustering 2nd order structure
maybe 1st order too

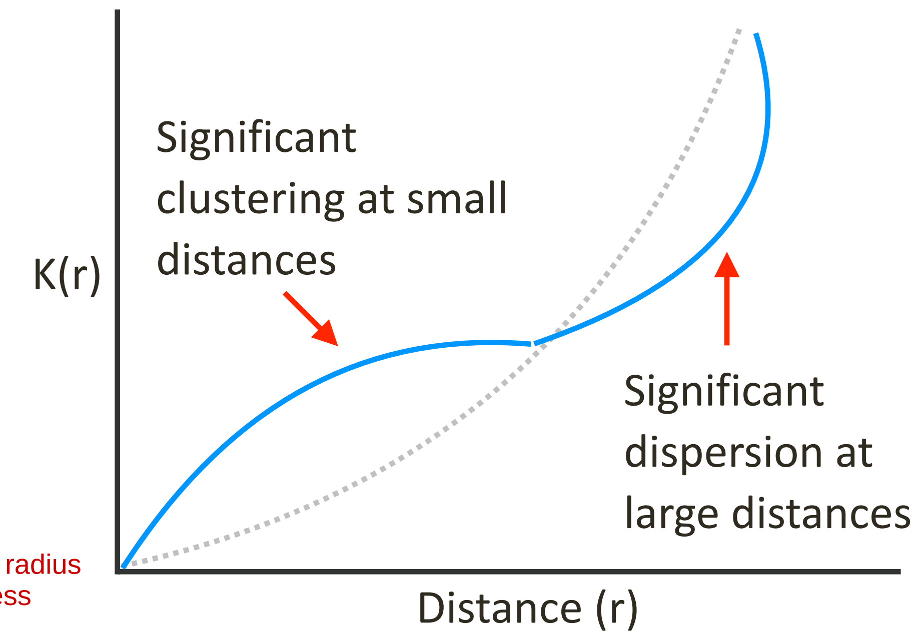
similar to the pair correlation function but slightly different

Point-pattern analysis – Distance methods

- Ripley's K:
  - Counts the number of events within defined distances from each point
  - Compare counts to the number of events expected under CSR
    - larger than expected = ? clustering or aggregated
    - smaller than expected = ? dispersed or segregated
  - Pair correlation function - version of Ripley's K that uses rings rather than circles

Ripley's K is performed on every point





go to each point and count the number of events within a defined radius
compare that number with expectation under complete randomness

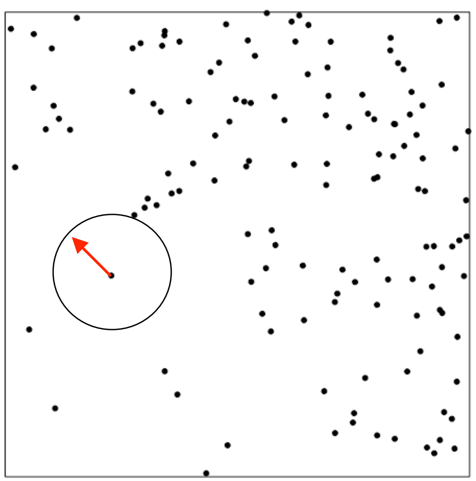
uses every point as a focal point

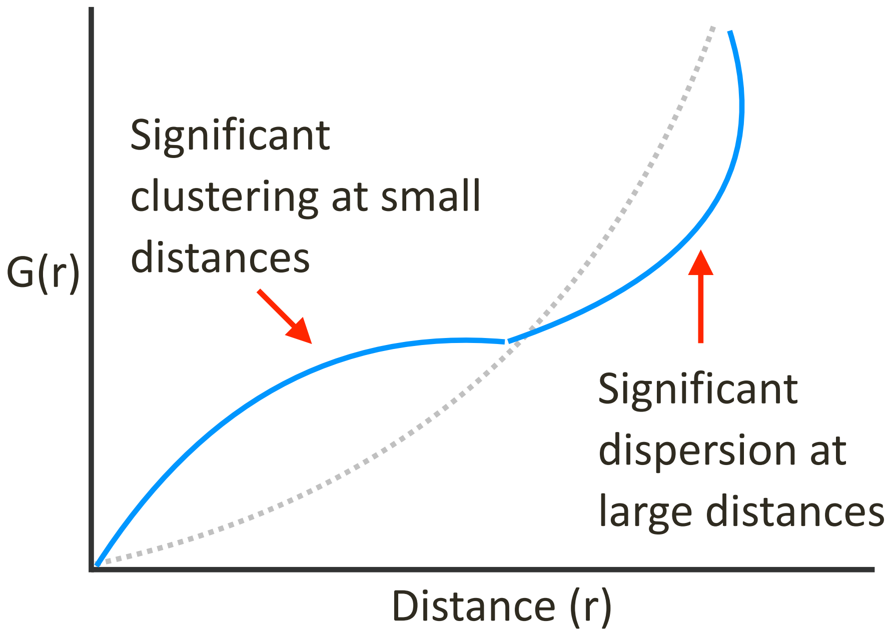
Point-pattern analysis – Distance methods

for point pattern analysis, your sampling cannot take place on say a grid because that imparts its own spatial structure

- Nearest neighbor distance distributions: G(r)
  - G(r) - event-to-event distribution
  - Distribution of distances *r* to the nearest neighbor, measured ***from an event*** in the pattern
  - Interpretation of departure from the null expectation the same as for Ripley's K.

looks at number of nearest neighbors around a focal point



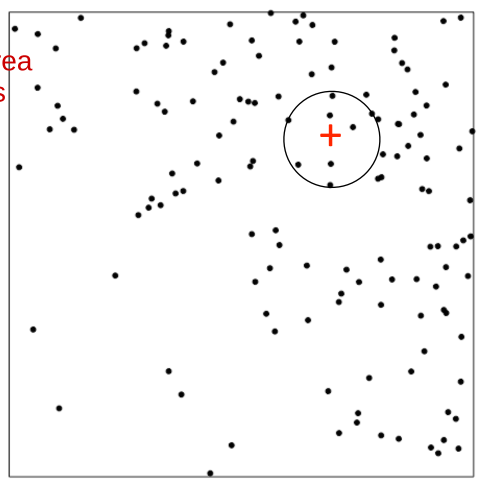


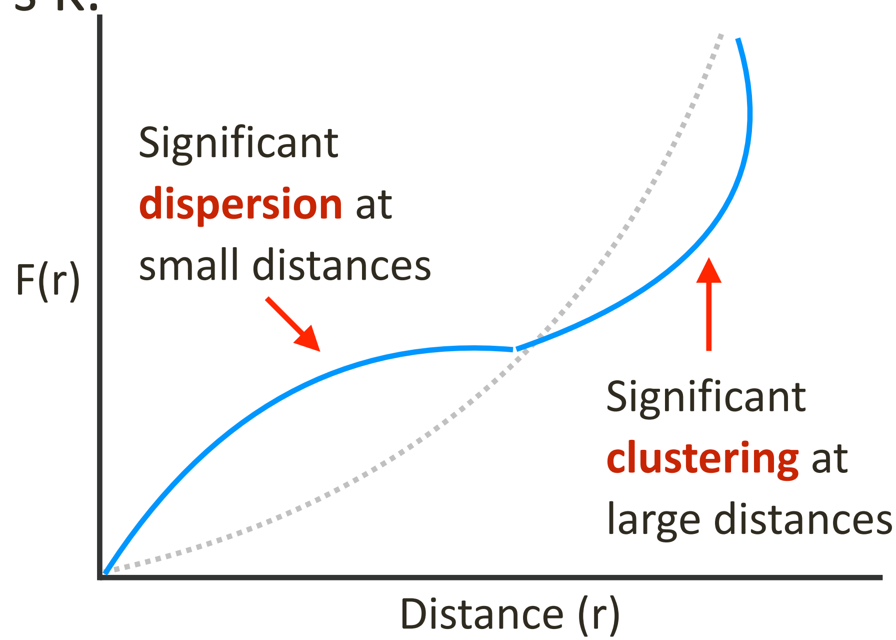
Point-pattern analysis – Distance methods

- Empty space function F(r) & Spherical contact distribution H(r):
  - The distribution of distances *r* from ***an "arbitrary" test location*** to the nearest event of the pattern
  - Test locations can be random or on a grid
  - Characterizes the empty space in a pattern
  - Interpretation of departure from the null expectation ***opposite that*** of Ripley's K.

focus measurements on non-events within the region

ie. the red x is on an empty area and then draws a radius to examine the points (events)





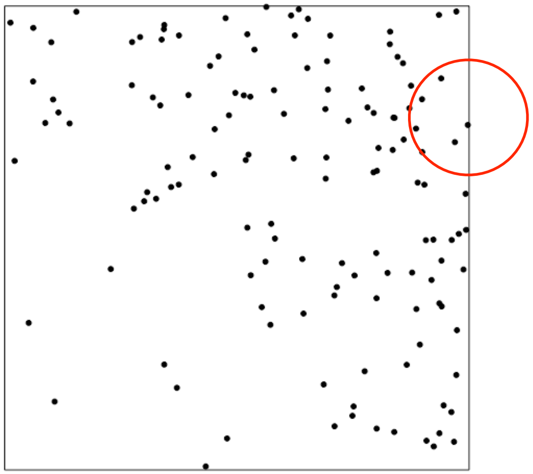
best practice in point pattern analysis to use multiple statistics and compare them

Point-pattern analysis – Adjusting for inhomogeneity

- If a point pattern is spatially ***inhomogeneous***, then analysis of the pattern must account for this inhomogeneity.
- Modified version of the K-function (Ripley's K), weighted by *spatial variation in intensity*.
- in R: Kinhom(X, λ), where λ is a ***function*** that describes variation in intensity

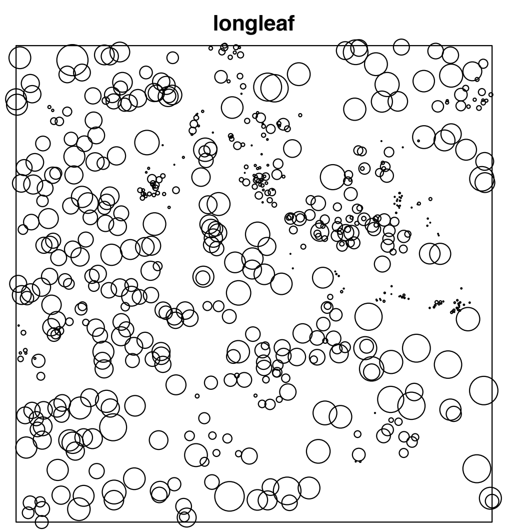
Point-pattern analysis – Distance methods

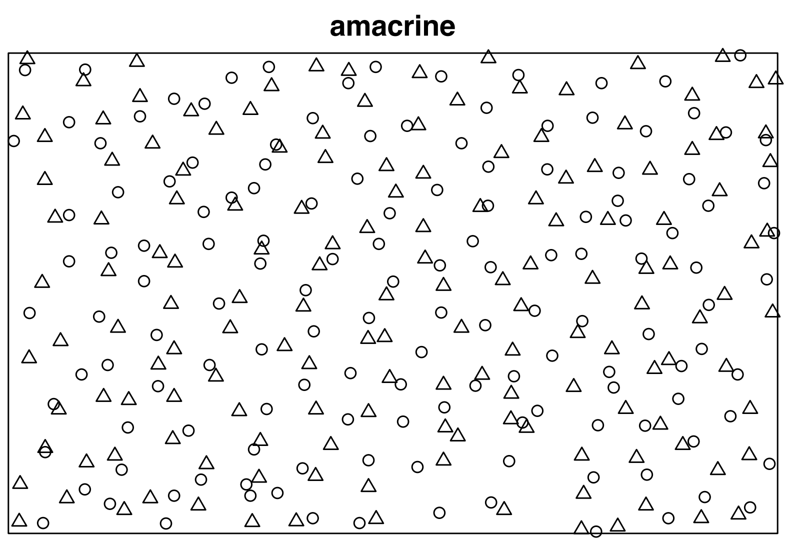
- Important caveats:
  - Most distance methods assume that the point process is stationary / homogeneous
  - If the process is not stationary, could misinterpret test results as indicating interactions between events (2nd order), when in fact that could result from variations in intensity (1st order).
  - Distance-based summaries do not completely characterize the process
  - All can be sensitive to **edge effects**, but there are often corrections for this



Marked point patterns

- Marks are interpreted as an ***additional coordinate (z)*** for the event.





Marked point patterns

- Many summary functions have been extended to marked point patterns (categorical / qualitative marks).
- Neighbor-distance functions for marked point patterns are called “cross-type” or “*i*-to-*j*” summary functions because they **compare distances from an event of type *i* to an event of type *j***.
- Also defined are “*i*-to-any” summaries (dot functions) = the distance from any event of type *i* to an event of any type.
- “Marked correlation function” = a measure of the dependence between the marks of two events of the process a distance *r* apart

Marked point patterns - example

Ant / bird / fish nest data: Two species of ants / birds / fish build nests in a desert / forest / estuary. We want to investigate ecological interaction between the species, and between different nests of the same species. The locations of all nests are mapped, and marked by species.

- These data can be analyzed as a marked point process consisting of two different types of events.
- The 'mark' attached to each event is its species (a categorical variable or qualitative mark).
- Possible analyses to consider:
  - Poisson marked point process to check for interactions between marked events
    - Do the species tend to be aggregated or segregated? In other words, does species A tend to occur near species B?
  - Treat one of the species as a covariate and analyze the other species conditional on it.
    - Does the intensity of one species influence the intensity of the other?

Challenges of linking spatial pattern & process…

1. **The problem of equifinality** - The same process may generate many different spatial patterns. Different processes may create the same spatial pattern.
   - Levin (1992): “There are many roads to Rome; and in general, there will be many conceivable mechanisms that could give rise to any set of patterns.”

Challenges of linking spatial pattern & process…

1. **The problem of equifinality** - The same process may generate many different spatial patterns. Different processes may create the same spatial pattern.
2. Assigning causality with confidence can be difficult because several dynamic processes may interact and modify the spatial patterns in a complex manner

Challenges of linking spatial pattern & process…

1. **The problem of equifinality** - The same process may generate many different spatial patterns. Different processes may create the same spatial pattern.
2. Assigning causality with confidence can be difficult because several dynamic processes may interact and modify the spatial patterns in a complex manner
3. Well-defined, but ***non-random*** processes can produce patterns indistinguishable from apparent random assembly (in other words, the pattern may appear random, but was generated by a non-random process).

Challenges of linking spatial pattern & process…

1. **The problem of equifinality** - The same process may generate many different spatial patterns. Different processes may create the same spatial pattern.
2. Assigning causality with confidence can be difficult because several dynamic processes may interact and modify the spatial patterns in a complex manner
3. Well-defined, but ***non-random*** processes can produce patterns indistinguishable from apparent random assembly.
4. Process and patterns are interrelated - the generating process also may be the result of the spatial pattern rather than being the source of the pattern

…but, some good news

1. A single process *can create* a single precise pattern
2. Processes *can and do create* highly structured patterns
3. The impact of pattern on process may not act at the same scales as the impact of process on pattern - and so it can be possible to tease part this interplay.

Strong spatial signals of biological processes can and do exist in many datasets, you just need to be careful in your interpretation and keep you mind open to other possibilities.

Avoiding pitfalls

1. Most studies consider only one or two summary statistics, four or five may be required to capture the properties of patterns with complex spatial structures.
2. Careful selection of an appropriate null model is key - but this is challenging and requires experience
3. Experimentation (including through simulation) is one way to increase confidence in linking pattern and process.
4. If this topic interests you, I highly recommend “Handbook of Spatial Point-Pattern Analysis in Ecology” by Wiegand and Moloney