Issues of Scale in Spatial Ecology - 2

How do we detect changes in process with scale?

A survey of quantitative methods …

“first law of geography: everything is related to everything else, but near things are more related than distant things.”

–Waldo Tobler

“first law of geography: everything is related to everything else, but near things are more related than distant things.”

In essence, Tobler’s law describes what we call **autocorrelation or spatial dependence:** the similarity of observations based on the distance (in space or time) between them

Statistics for identifying scale

- Many methods are available, we will focus on a few of the common approaches
- Different techniques are closely related with similar assumptions
  - all techniques are sensitive to sampling intervals (grain) and record length (extent)
  - in other words, the methods themselves are sensitive to scale
  - reliability can depend on the “stationarity” of the data
    - Stationarity = lack of trend or same mean, variance, and *isotropy* throughout the study area
      - Isotropy = the pattern is the same in all directions

most patterns are anisotropic

Useful methods for quantifying pattern of spatial (or temporal) structure include:

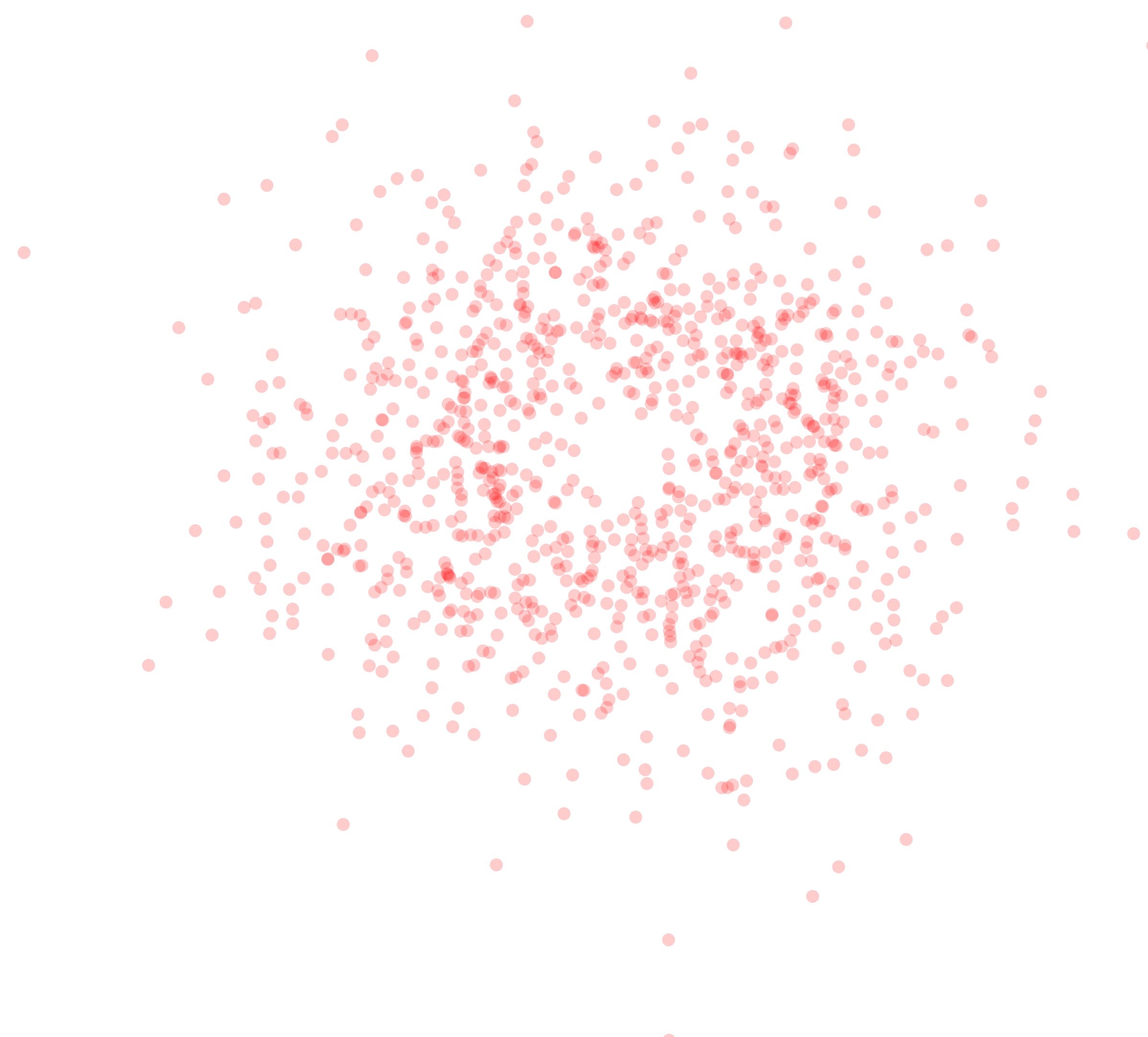
1. Variogram / Semivariance / Kriging
2. Correlogram / Autocorrelation
3. Spectral analysis
4. Point pattern analyses

kriging not strictly a method for quantification instead a method for interpolation

Semivariogram

Two common uses:

1. Structure recognition
2. Optimal interpolation
   (e.g., kriging)



Semivariogram

assume:
that further H distances away will be more dissimilar
-assume that at some point dissimilarity will be asymptotic (ie. can't be more dissimilar)

Examine the dissimilarity in values measured at two points some distance apart.

1. Rearrange data into **pairs** separated by distance *h*

2. Calculate squared distances by summation over all possible pairs at that distance *h* (*ideally at least 50 pairs for each value of h*)

3. Analyze how patterns of dissimilarities changes as a function of *h*

Note that the common implementation assumes *stationarity* (ie. that the trend is the same)

Semivariance

$$\hat{\gamma} = 1/2N(h)\sum_{i=1}^{N(h)}(X_i - X_{i+h})^2$$

divide by 2 because we are counting the data twice

- *N(h)* = number of lagged data pairs
- *h* = lagged distance
- *X*'s = data observations

Analyze dissimilarities

Lag(1) = difference between adjacent points

Lag(2) = difference between points 2 units apart

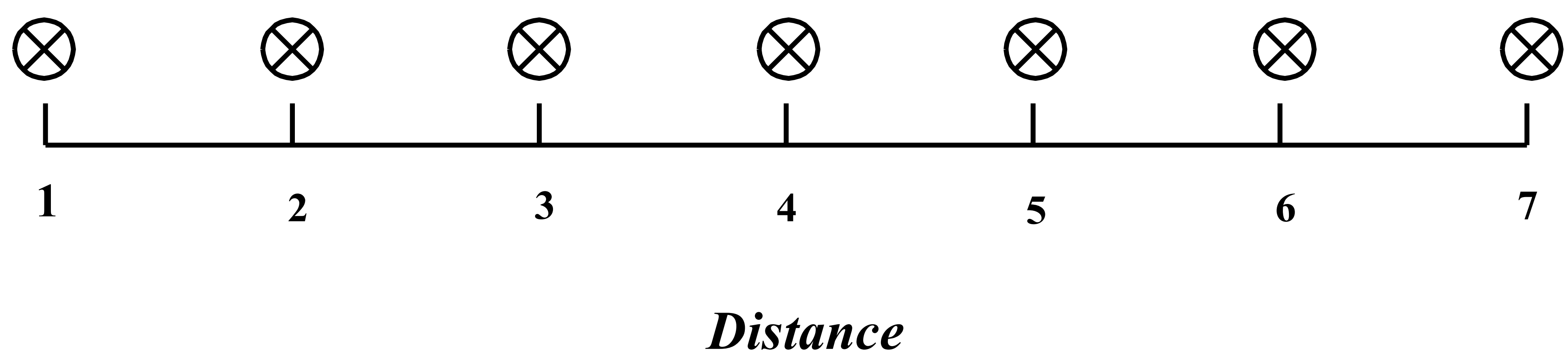
.

.

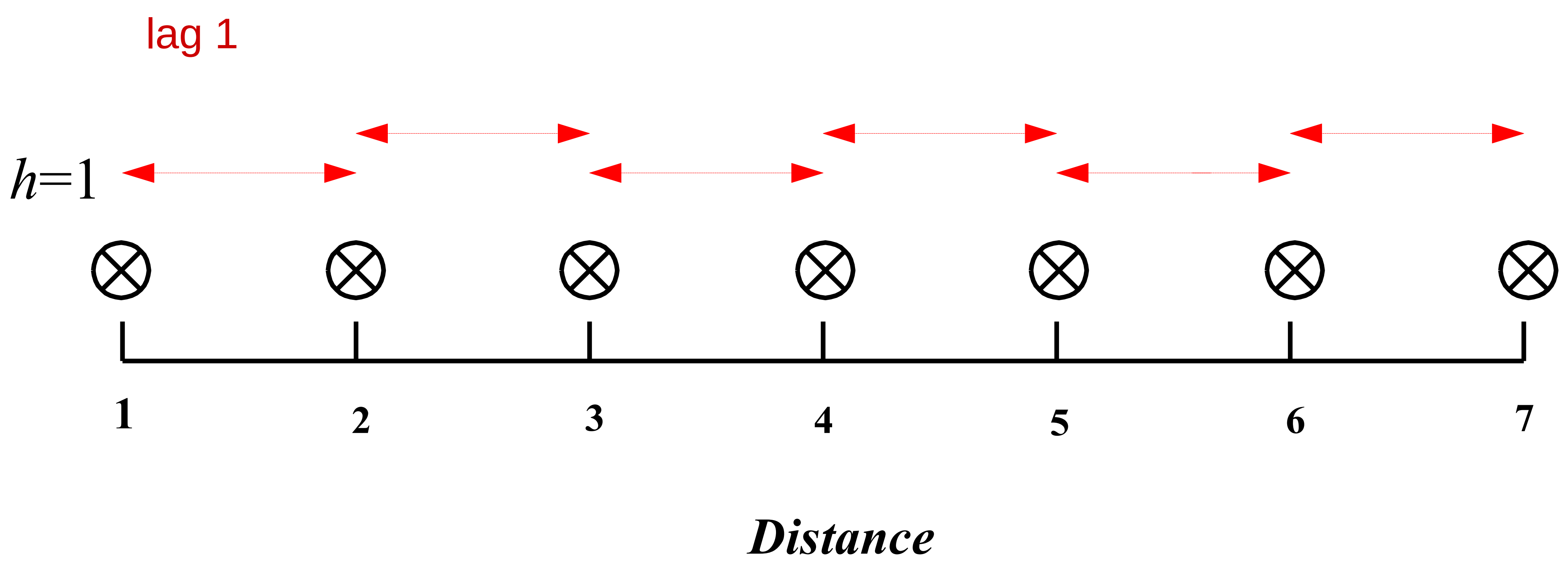
.

Lag(n/2) = difference between points **1/2 the extent** of the study area (n = # of points) the maximum parallized distance apart

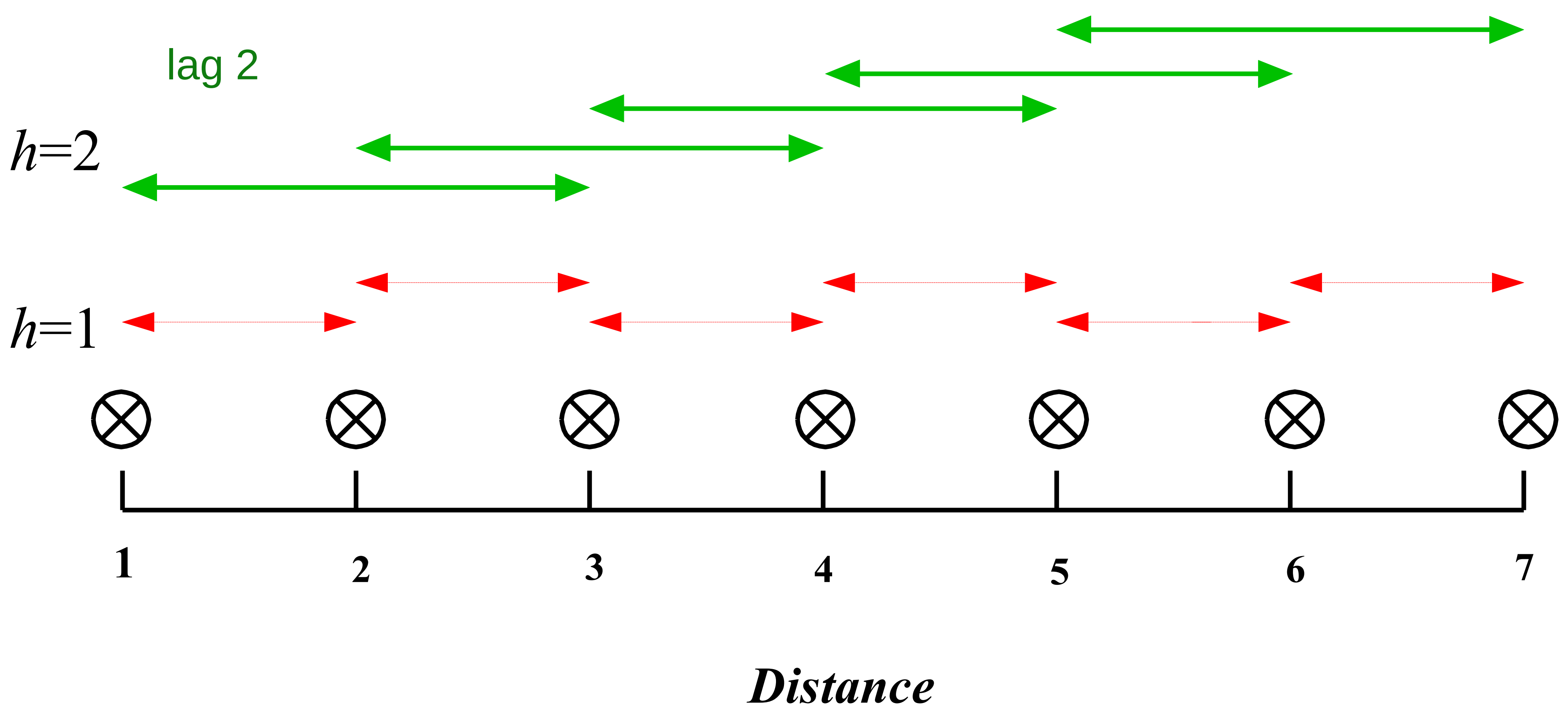
Counting lags



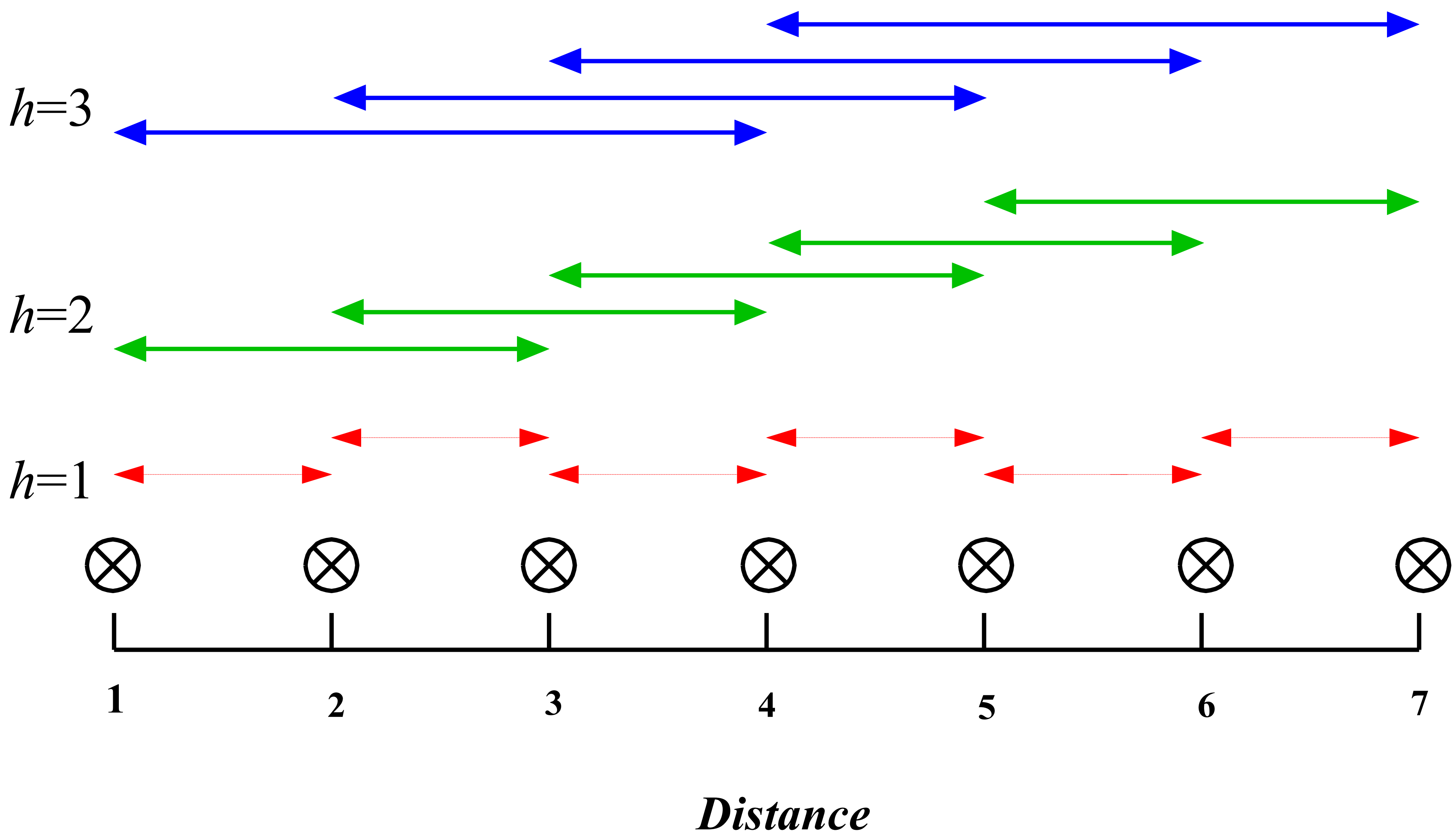
Counting lags



Counting lags



Counting lags



Semivariogram

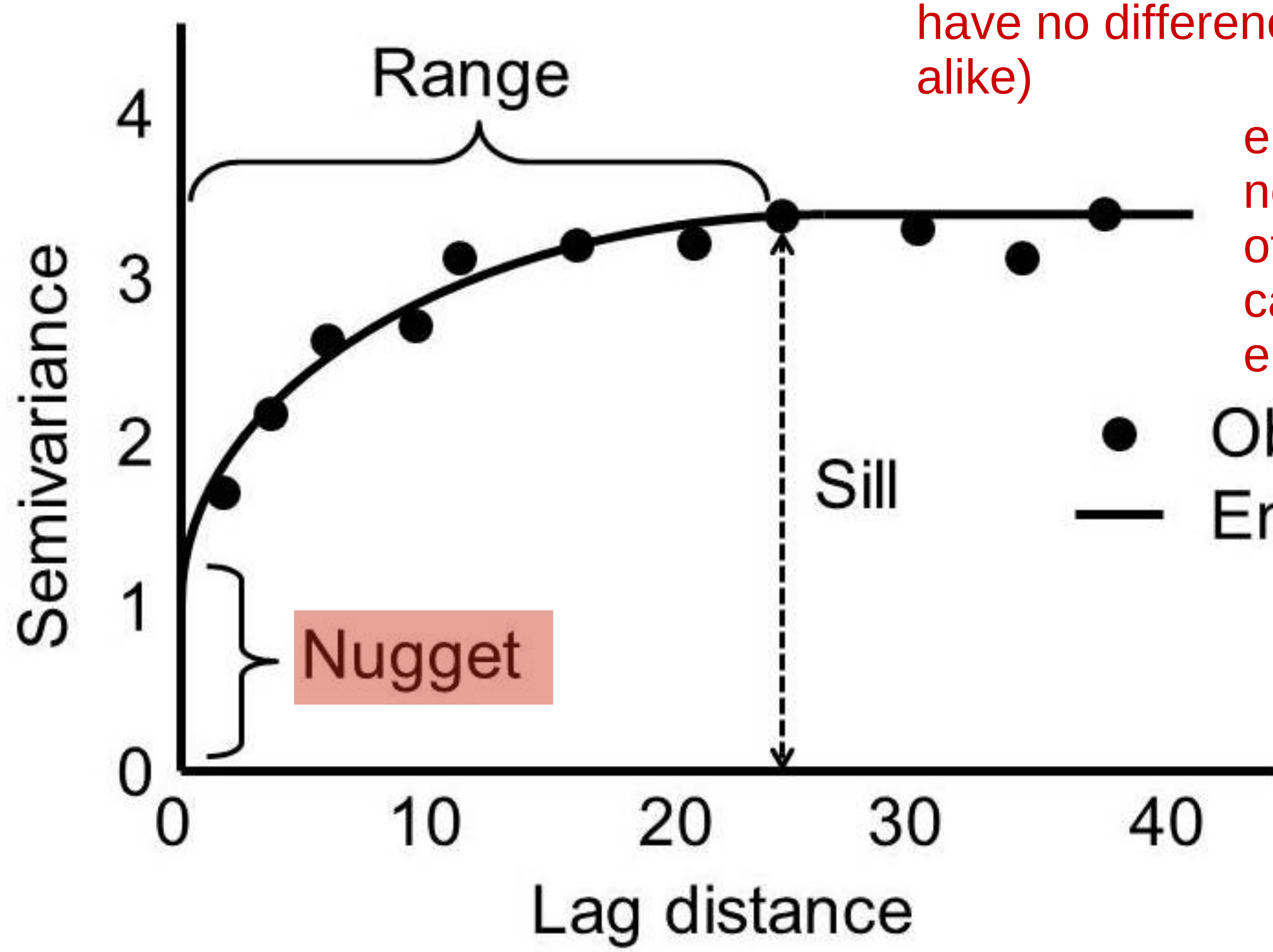
nugget = value at *h*=0 where the line intersects the y-axis

- indicates measurement errors or spatial sources of variation at distances smaller than the sampling interval or both

- In theory, nugget should always = 0

nugget should be zero because two points in the exact same location (ie. lag of 0) should have no difference (they should be exactly alike)

environmentally there may be no dissimilarity but there are other things that could be causing dissimilarity besides environment

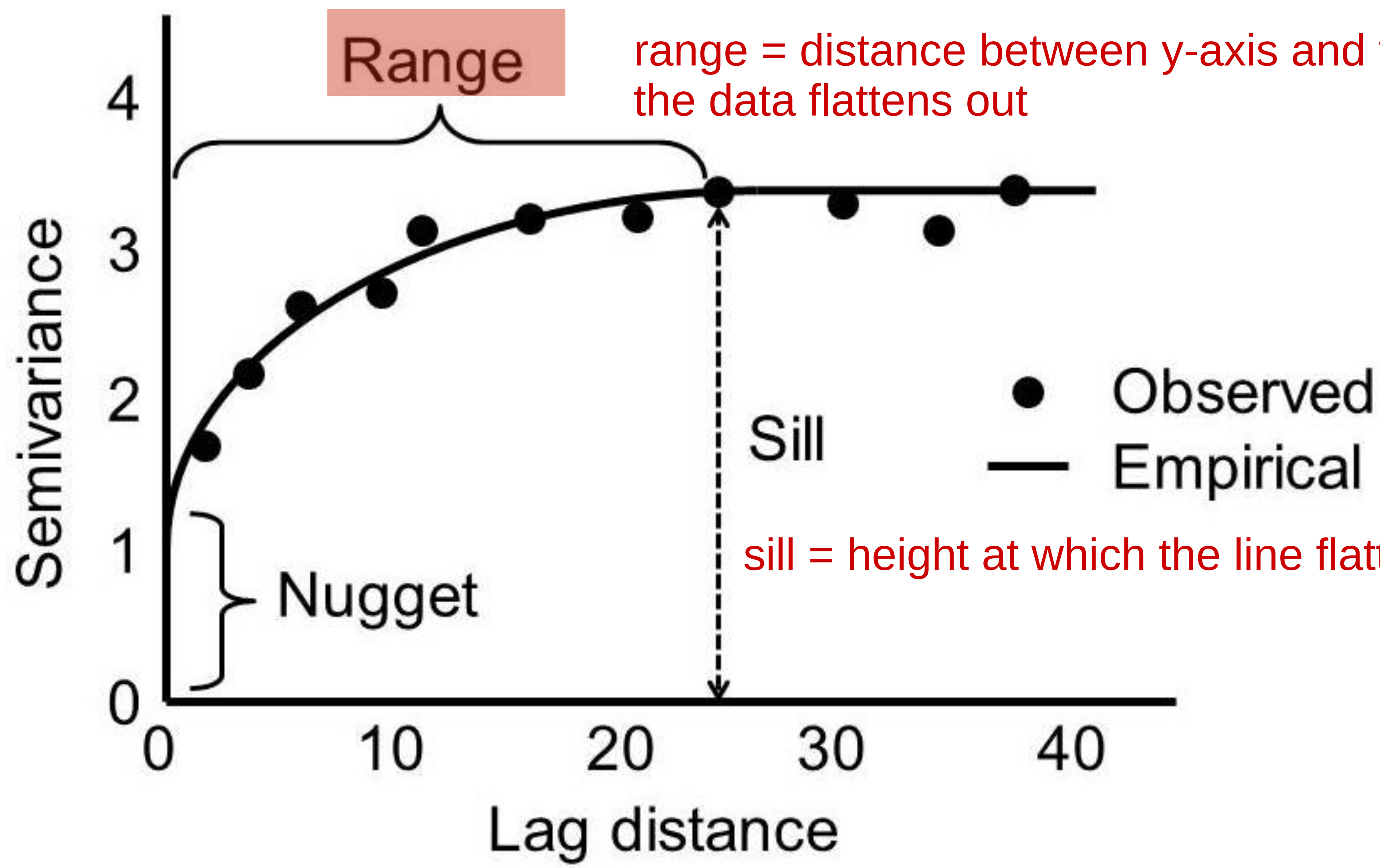


Semivariogram

Range

if the sampling interval is too big, you may miss the spatial structure due to points being far distances from one another and missing small dissimilarity variation

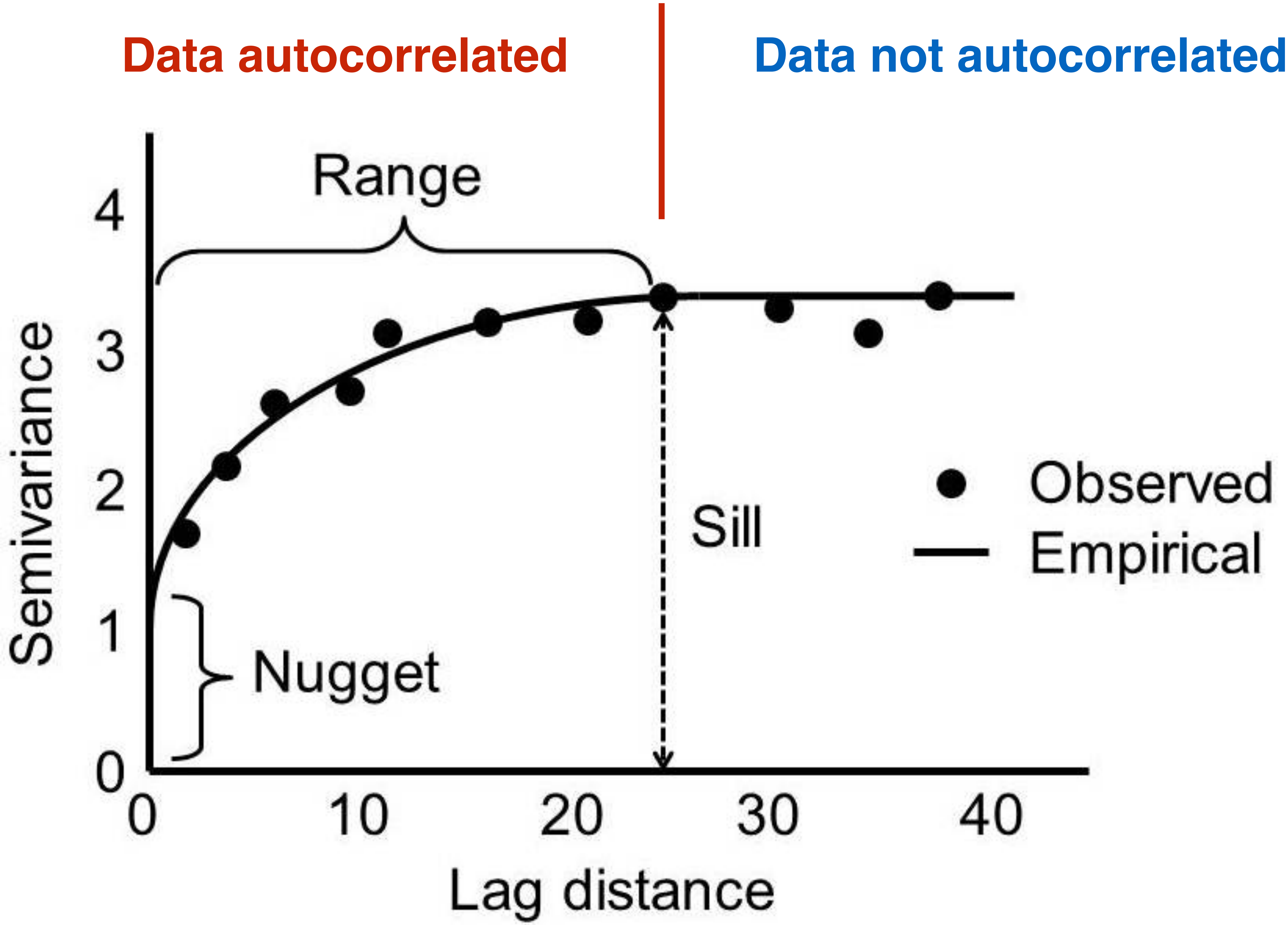
- Distance within which sampled points are spatially dependent
- if semivariogram is flat THERE IS NO SPATIAL STRUCTURE



range = distance between y-axis and the point at which the data flattens out

sill = height at which the line flattens out

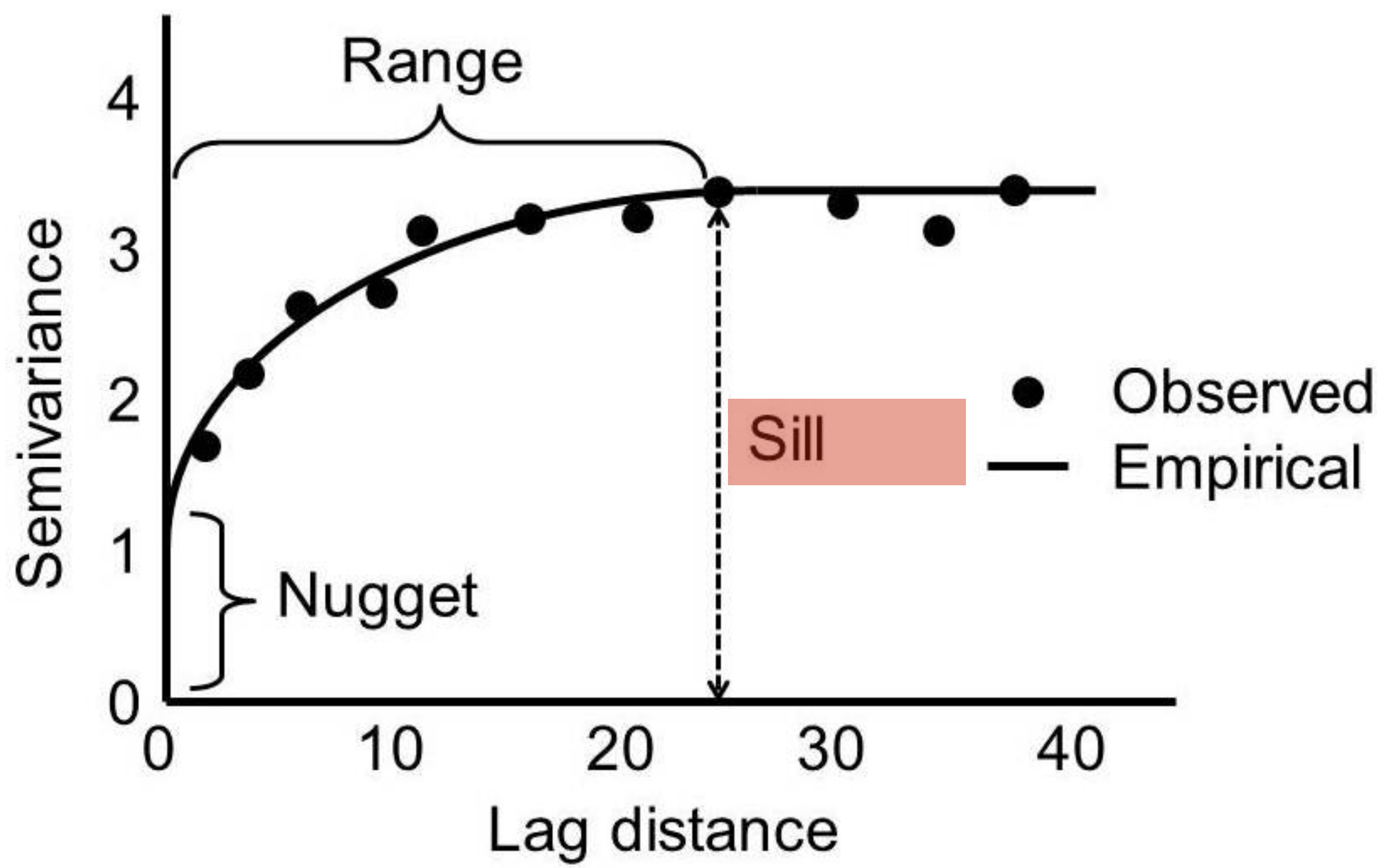
Semivariogram



Semivariogram

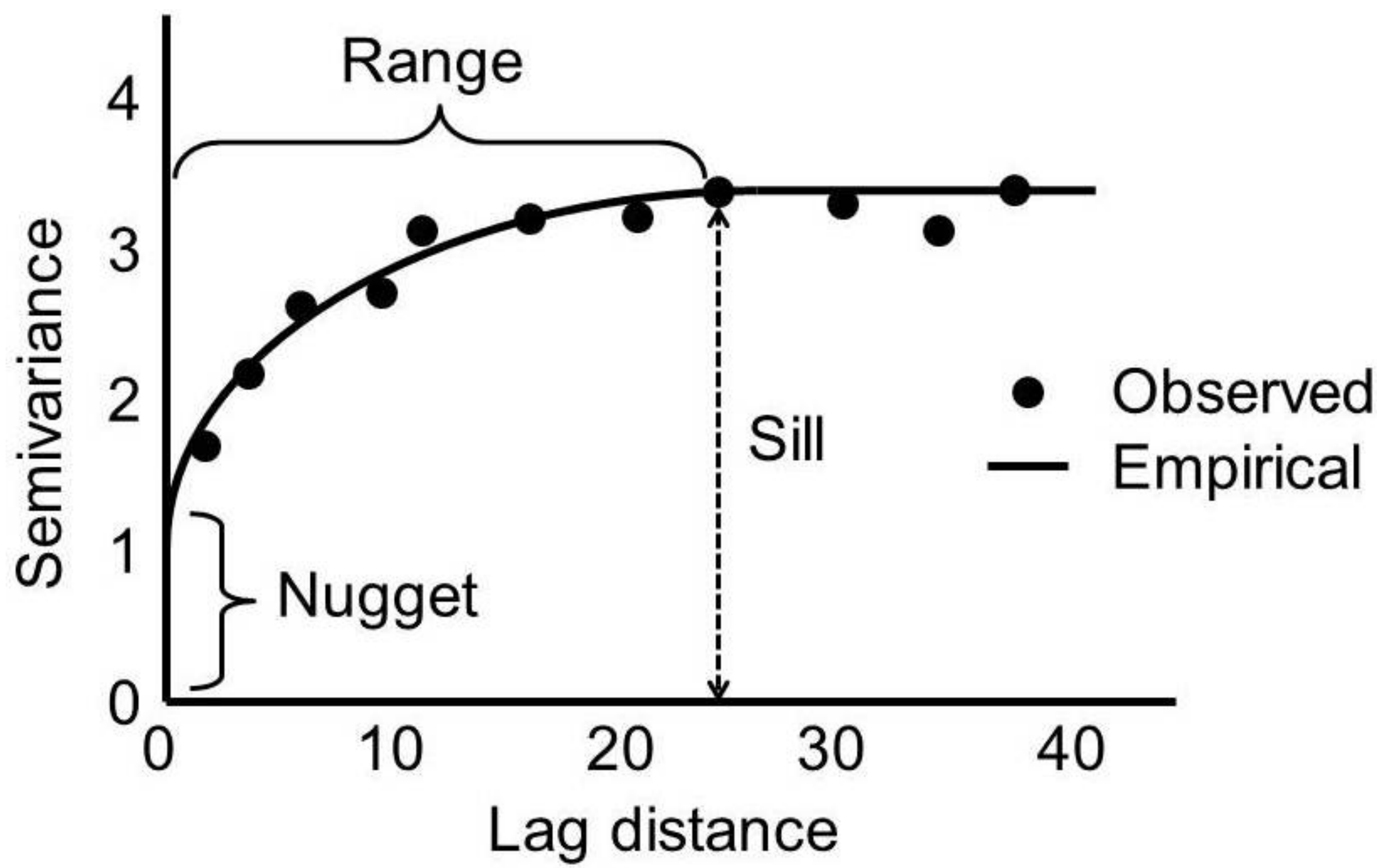
sill

- Height of the range, where the semivariogram flattens out



Semivariogram

Key point: identifies the scale at which points are spatially dependent

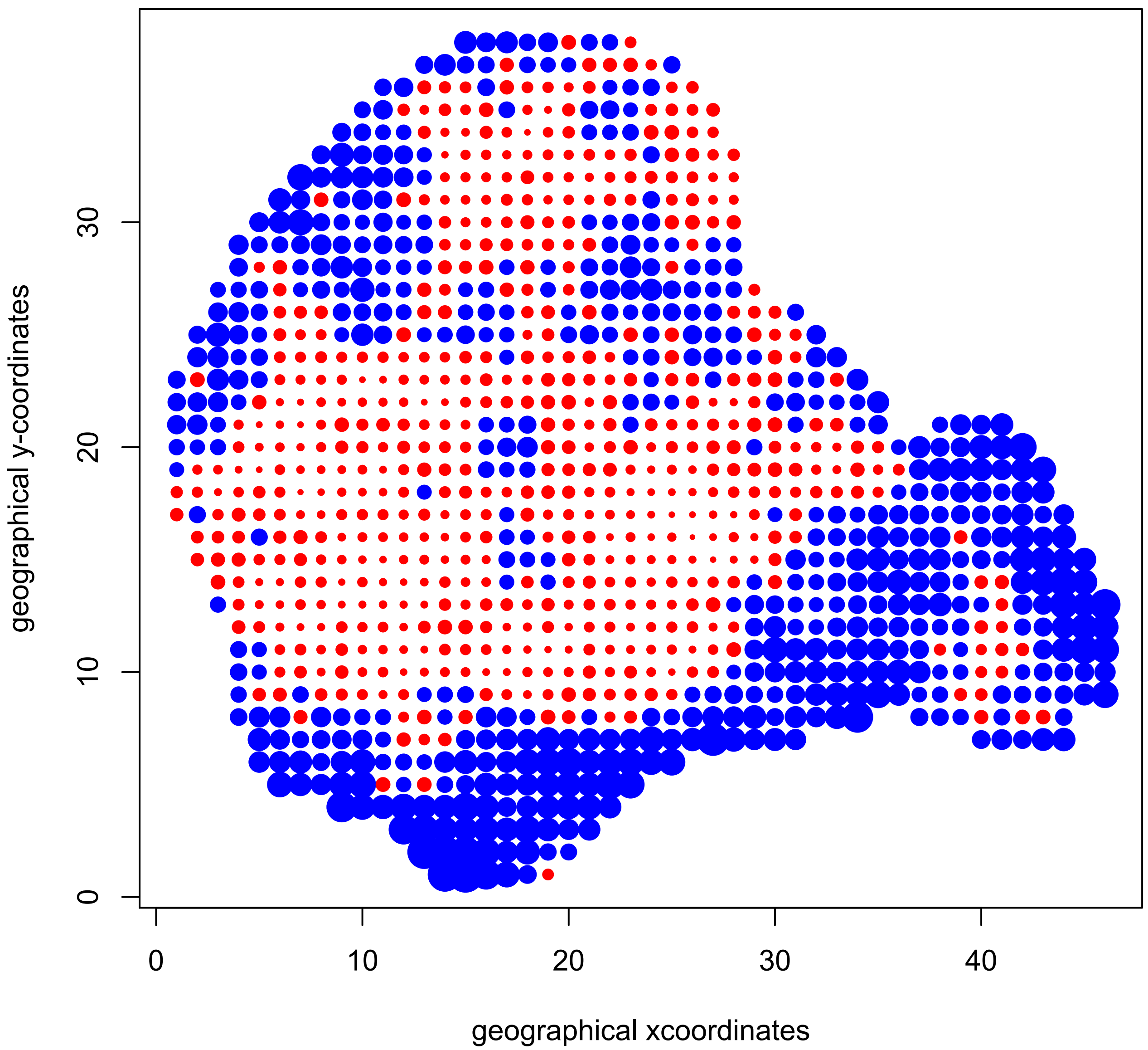


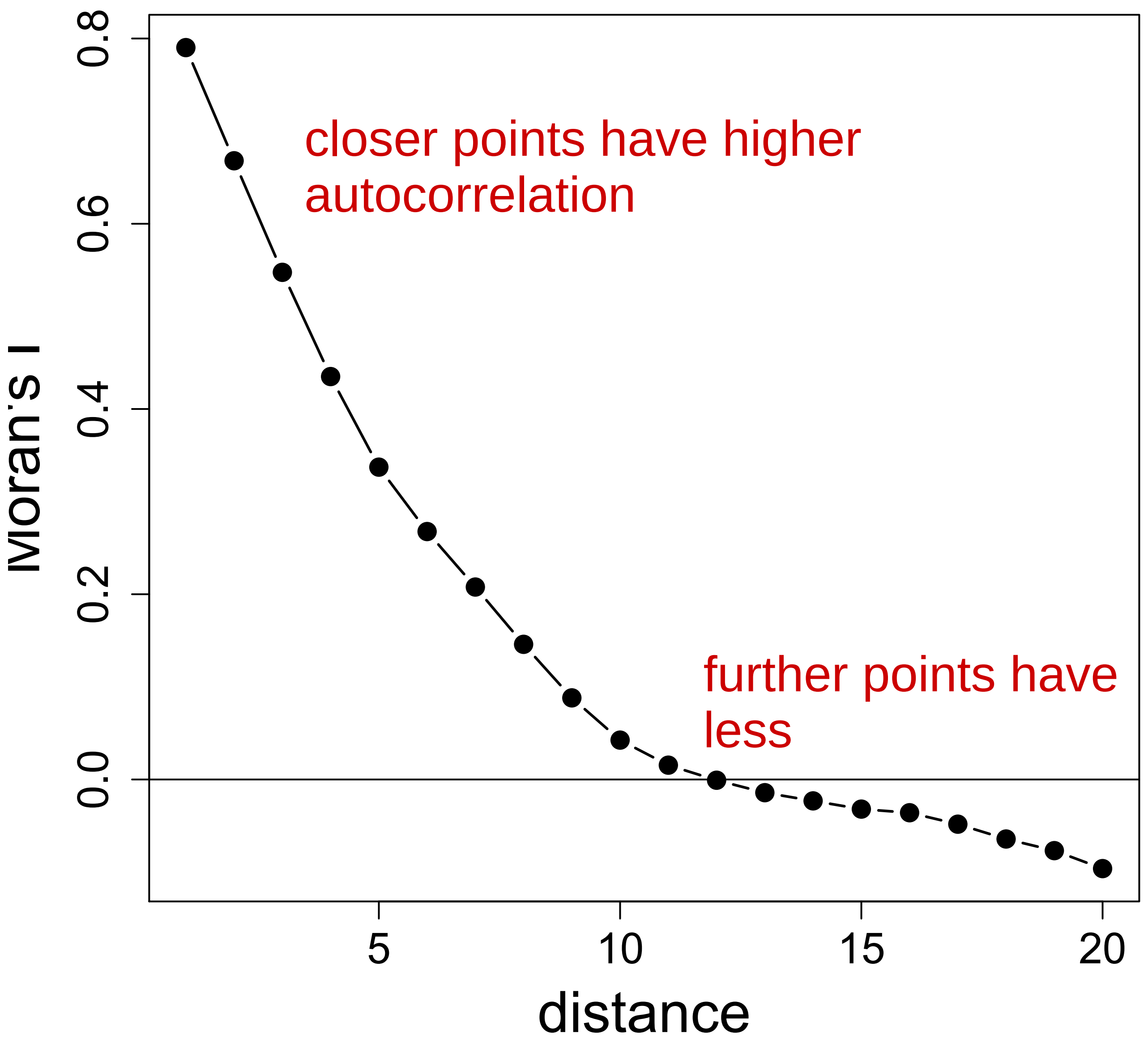
2. Correlogram / Autocorrelation

- general principles similar to semivariance
- *correlations* (instead of dissimilarities) estimated between lagged pairs of data
- a **correlogram** visualizes results (aka ‘autocorrelation plot’)
  - Plot of Moran’s I against distance classes is a Correlogram
  - Moran’s I = degree of correlation between values of a variable as a function of spatial location (similar to Pearson’s coefficient)
    - Varies from -1 (negative autocorrelation) to 1 (positive autocorrelation)
    - 0 = no spatial structure

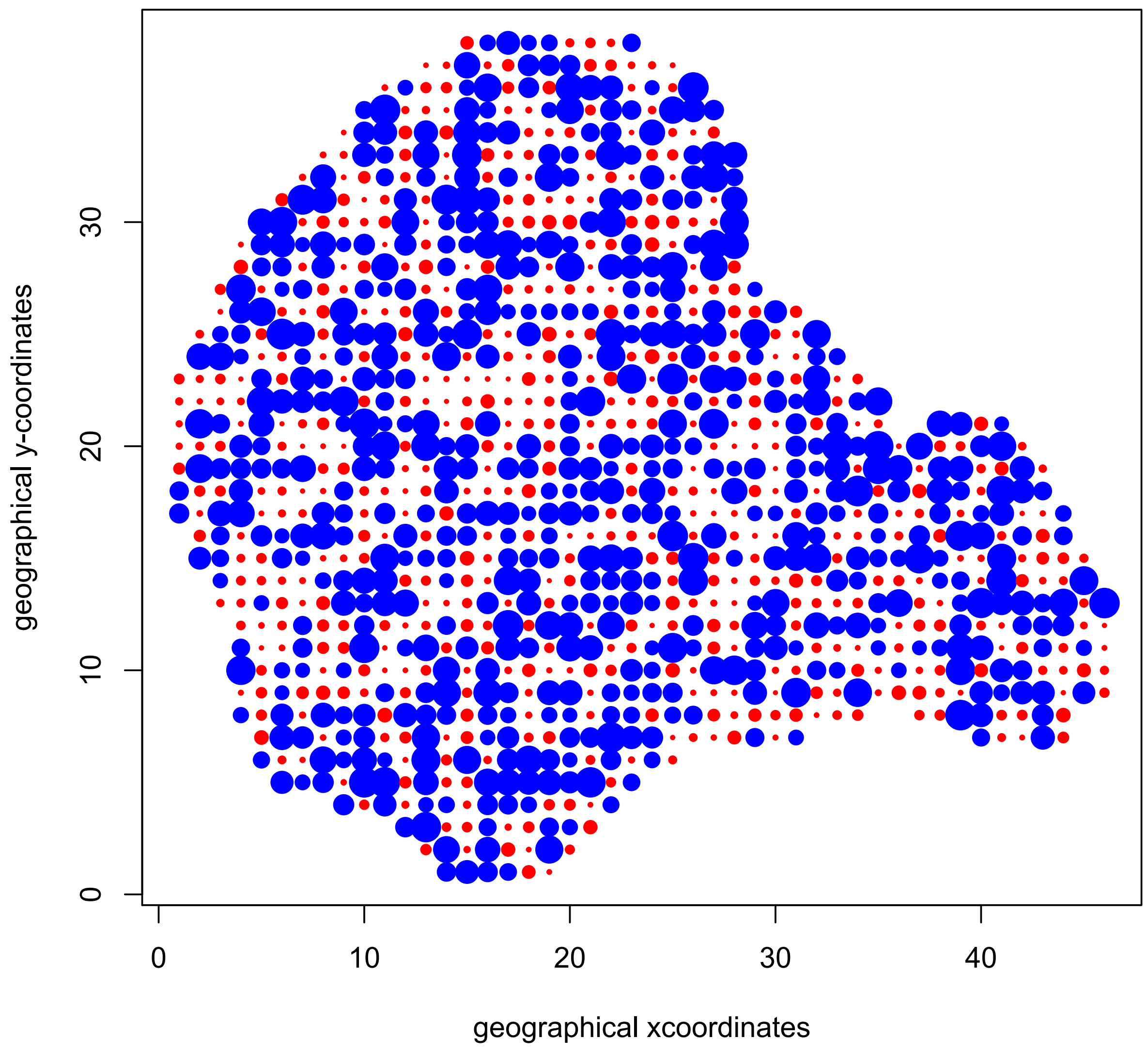
Spatial pattern vs. correlogram

blue points tend to be grouped and red points tend to be grouped



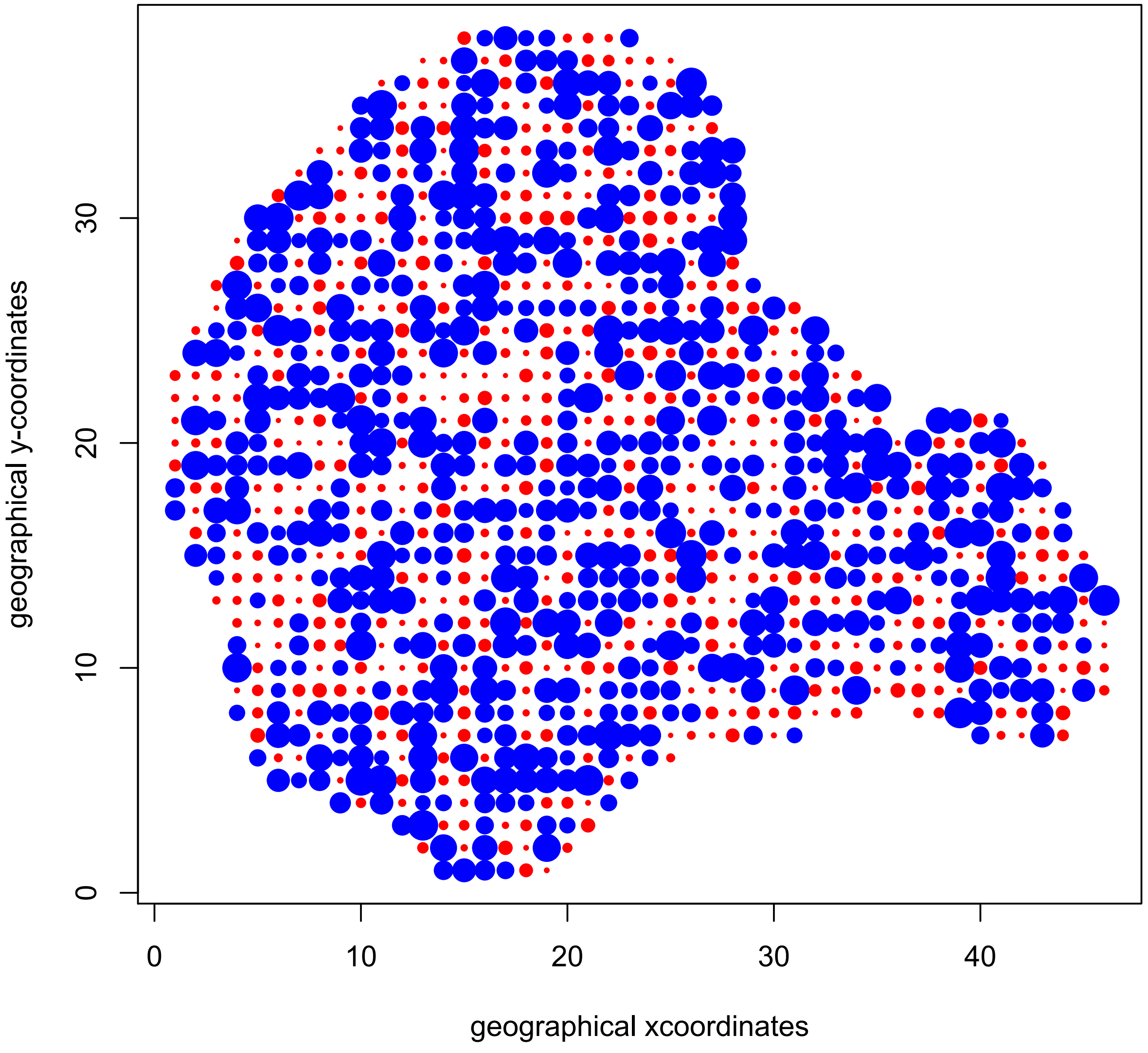


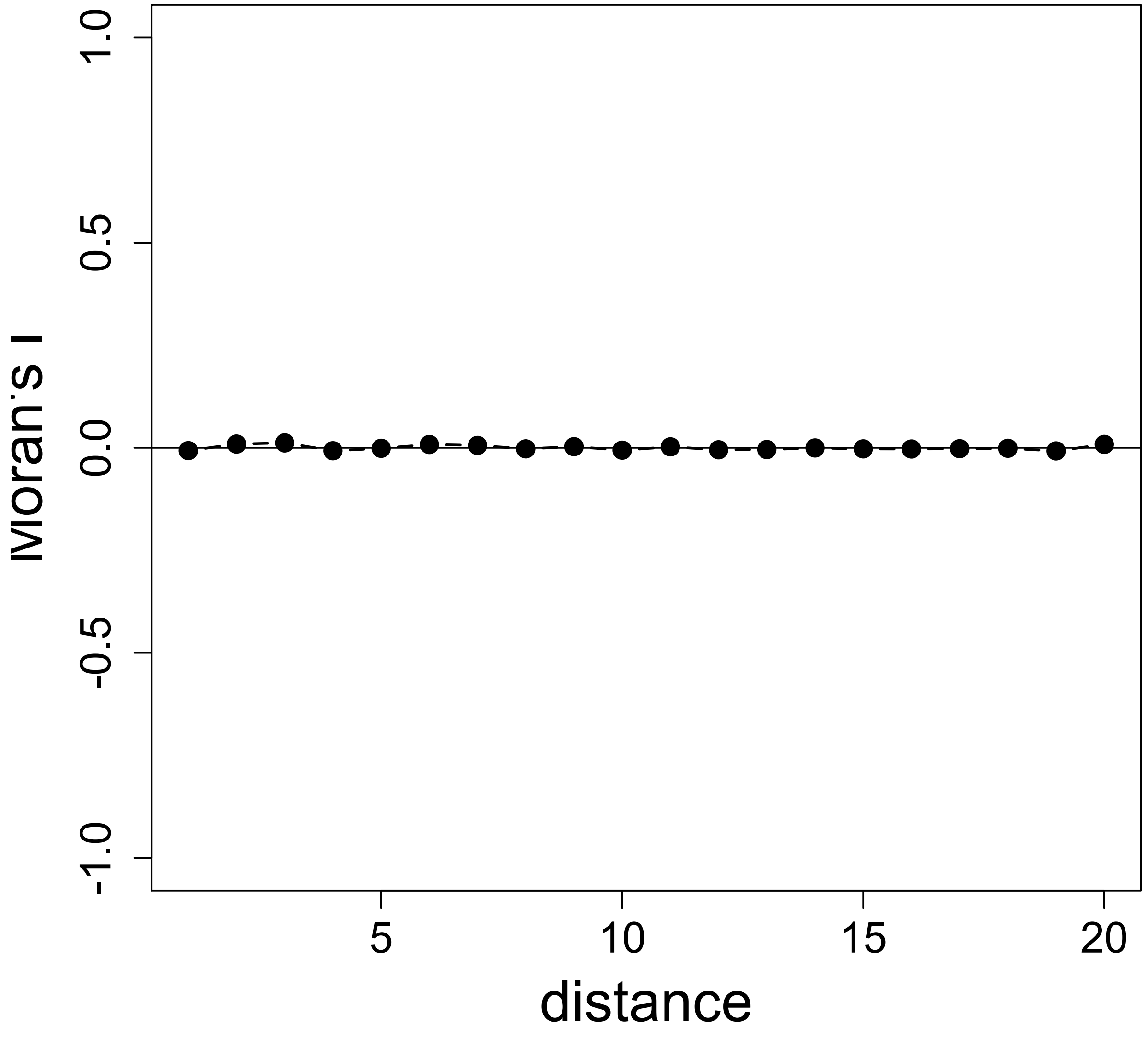
Spatial pattern vs. correlogram



Spatial pattern vs. correlogram

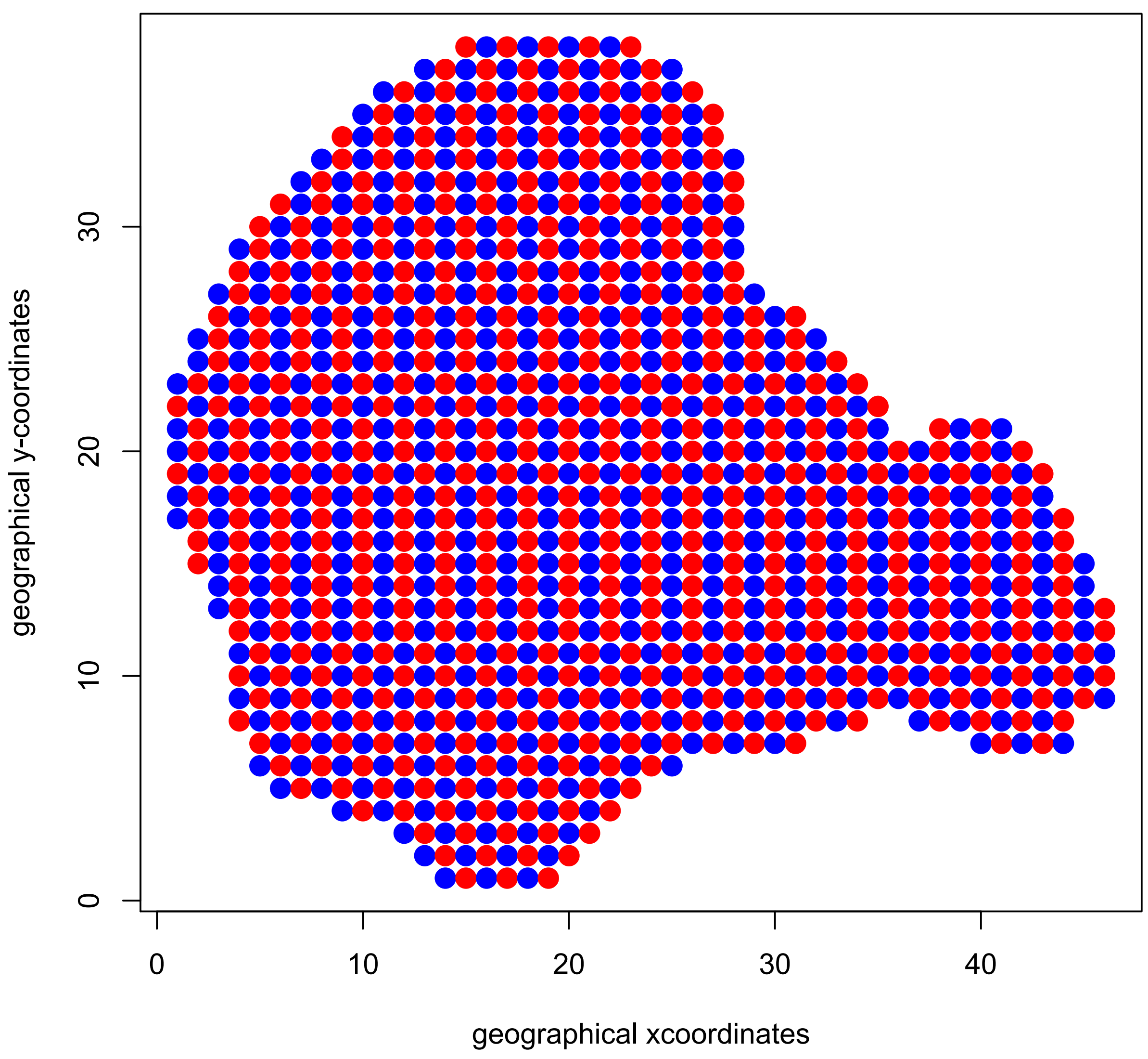
random pattern so no autocorrelation

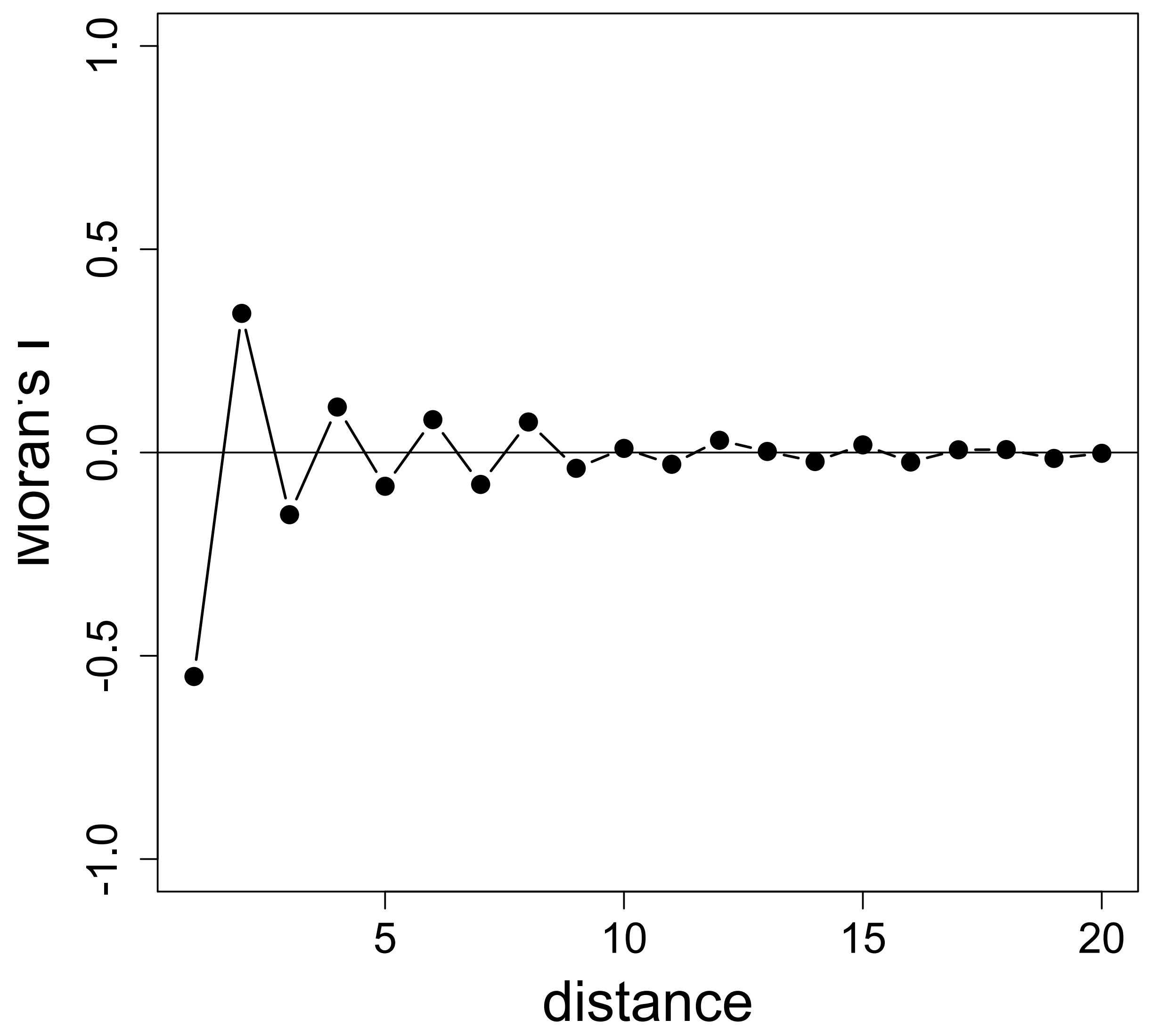




Spatial pattern vs. correlogram

oscillating to no spatial autocorrelation





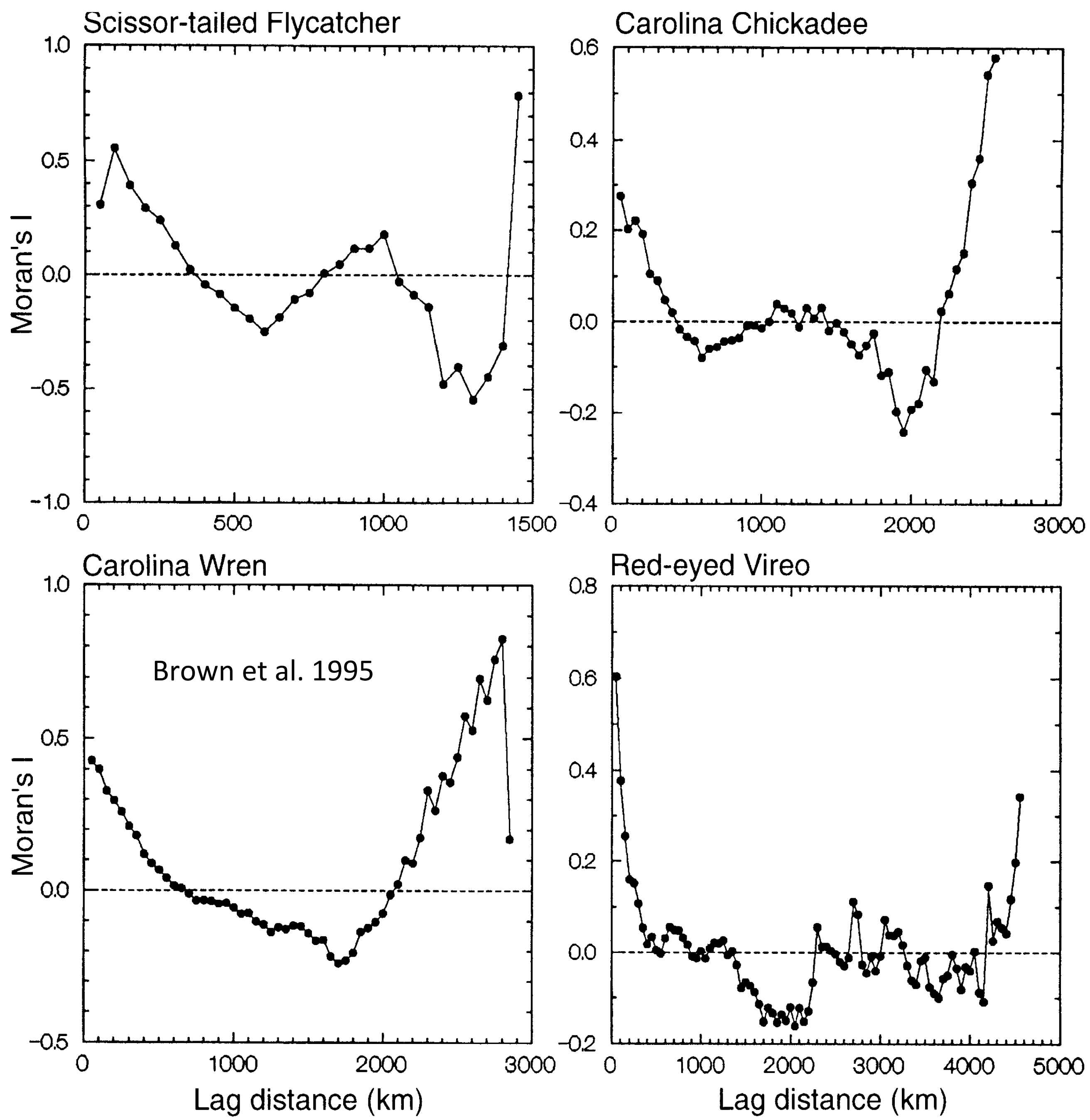
checkerboard

doesn't occur in nature

the birds have 2 scales of dependence

Spatial pattern vs. correlogram

abundances highly correlated at close distances then uncorrelated then highly correlated at high distances



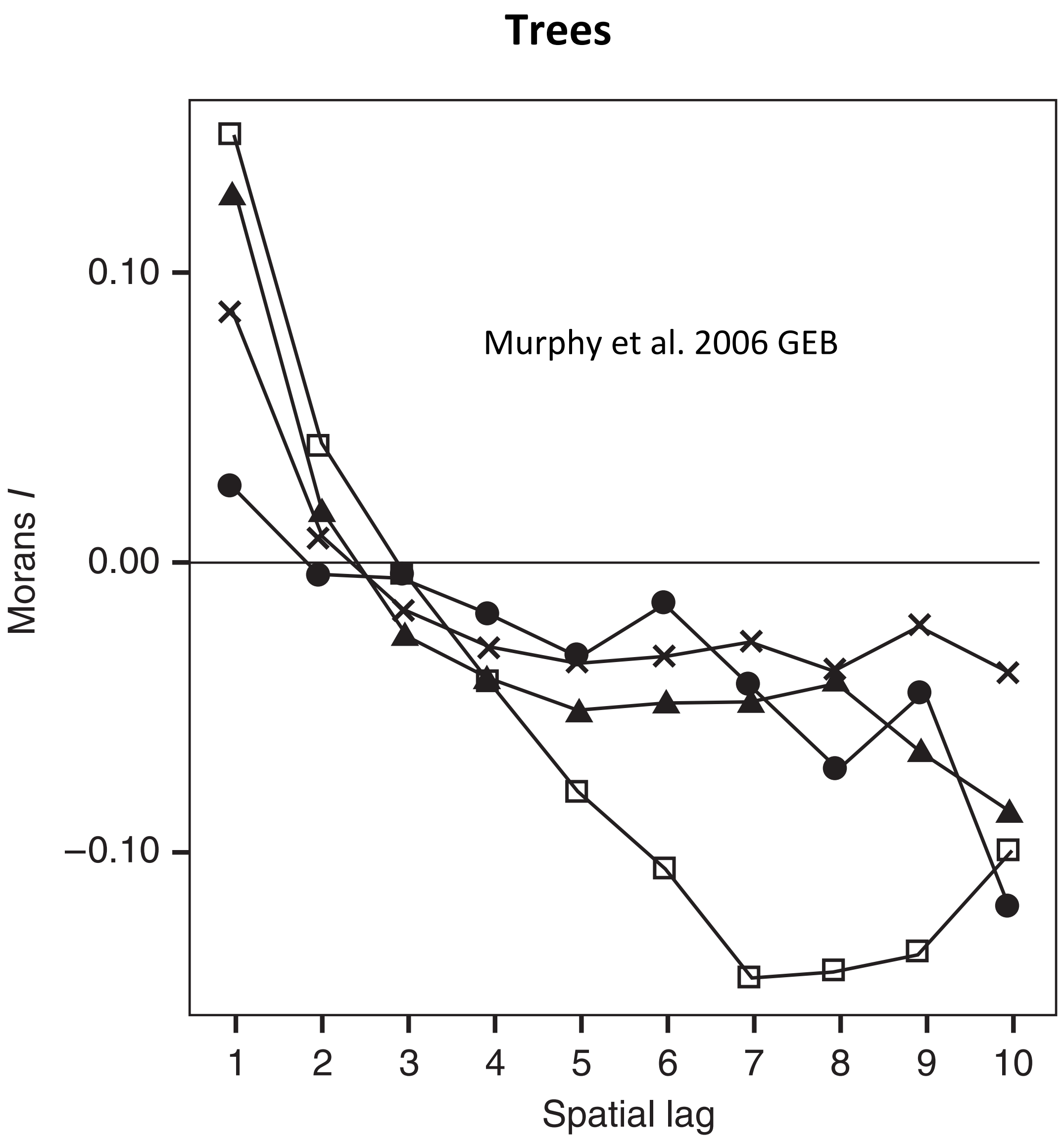
Spatial autocorrelation of bird abundance across their geographic ranges

What pattern of abundance might explain these correlograms?

abundance center describes high correlation closeby as move to compare center to edges you decline then compare edges to edges the correlation increases

Spatial pattern vs. correlogram

can use monte carlo to assign statistical significance to the points (ie. assign open or closed points to show it). As go towards 0, no autocorrelation and no significance



Spatial autocorrelation of tree species abundance across their geographic ranges

What pattern of abundance might explain these results?

latitudinal trend in abundance because the trees tend to be more abundant in certain latitudes than others

Interpreting correlograms

- provides visualization of change in space (or time)
- indicates positive and negative autocorrelation
- Restrictions:
  - data should be equally spaced
  - residuals should be normally distributed
  - requires at least ~50 data pairs