NRES_798_17_201501

Spatial Statistics

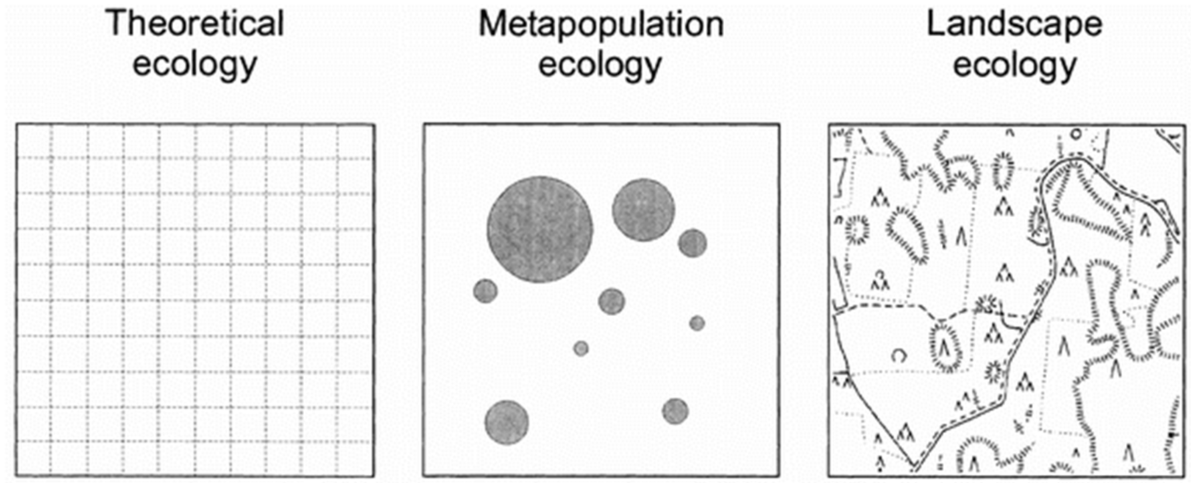
Fortin, M.-J., P. M. A. James, A. MacKenzie, S. J. Melles, and B. Rayfield. 2012. Spatial statistics, spatial regression, and graph theory in ecology. Spatial Statistics 1:100–109.

Liebhold, A. M., and J. Gurevitch. 2002. Integrating the statistical analysis of spatial data in ecology. Ecography 25:553–557.

Spatial ecology

- Island biogeography
- Metapopulation dynamics
- Niche vs neutral community structure
- Conservation ecology

Spatial ecology



Hanski, I. 1998. Metapopulation dynamics. Nature 396:41–49.

- Abiotic spatial heterogeneity
  - Habitat quality, resource availability, dispersal probability, mortality rate, colonization probability
- Biotic spatial heterogeneity
  - Territoriality, competition, predation, disease, local dispersal

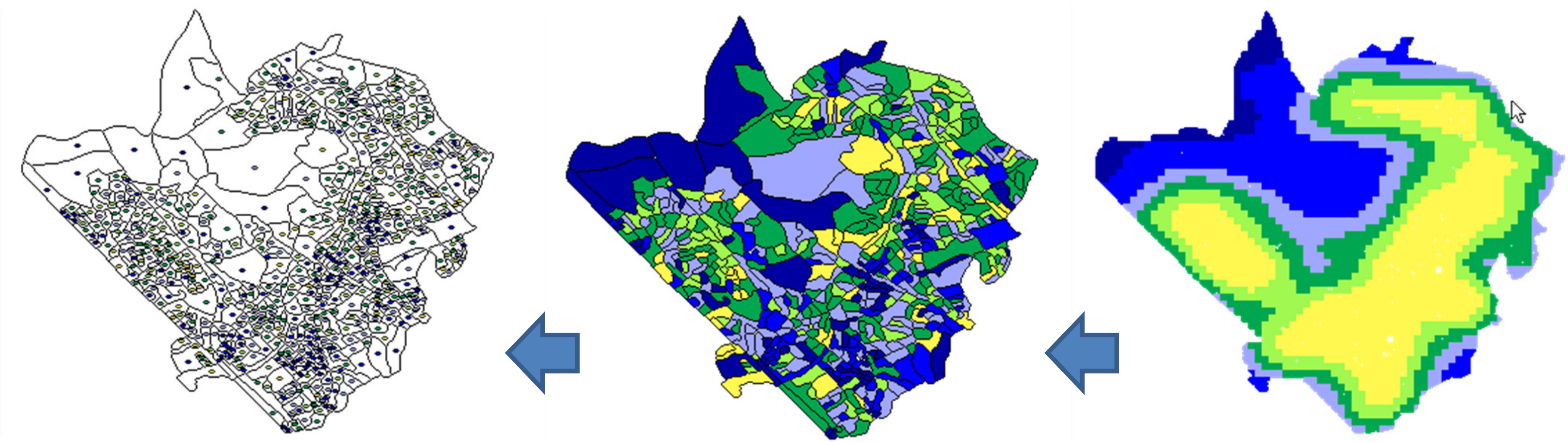
Spatial ecology

- Is there a spatial pattern?
- What process produces this spatial pattern?
- Does the “neighborhood” influence the process I am interested in?

- At what scale does spatial heterogeneity or pattern become apparent? (scaling)
- What is a “patch” and how can we describe groups of patches?
- Are there differences between or among study areas or landscapes with respect to spatial patterns?

Types of spatial data

- Point patterns
  - location of an event such as trees in a forest or bird nests
- Area data
  - Data that can be separated into zones that differ in intensity
    - E.g. Soil type, time since disturbance, density or number of species within an area
- Continuous data
  - Events that change across space
    - E.g. gradients of precipitation or salinity

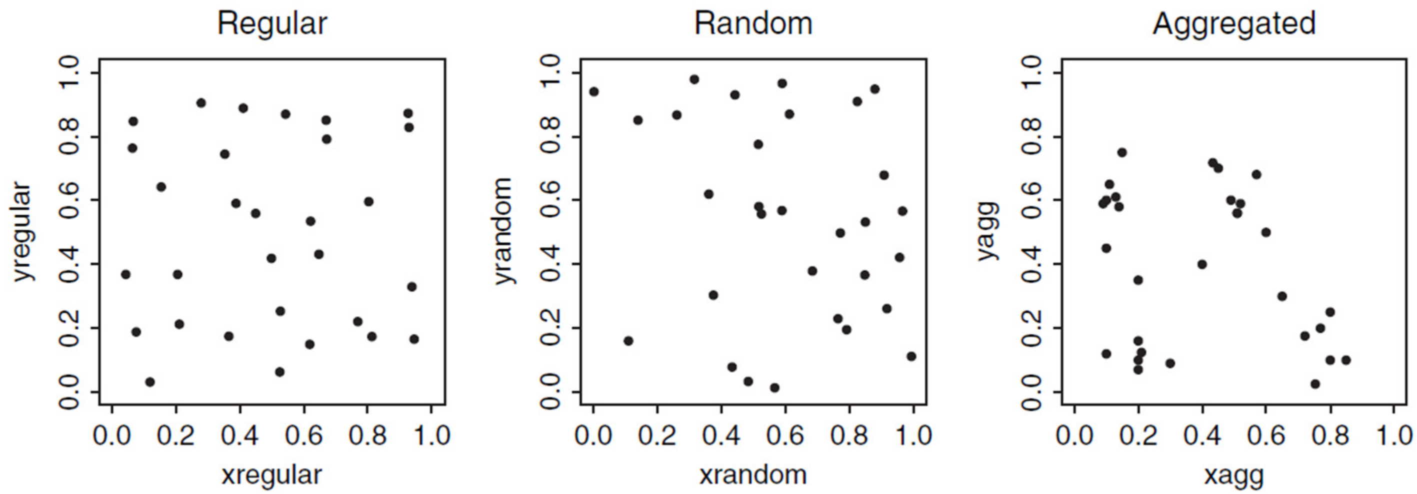


point pattern   area data   continuous process

Spatial statistics

- Test for aggregation
- Tests for spatial autocorrelation
- Maps of continuous response variables (kriging)
- Modeling spatially explicit responses (spatial impacts of neighbors)

Test for spatial randomness



- Ripley's K
- Calculated based on number of spatial points in a circular search window

$$\widehat{K}(t) = \frac{\lambda^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} I_t(e_i, e_j)}{n}$$

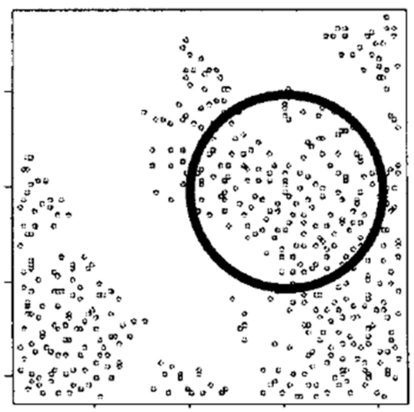
$$\text{for } i \neq j \text{ and } t > 0$$

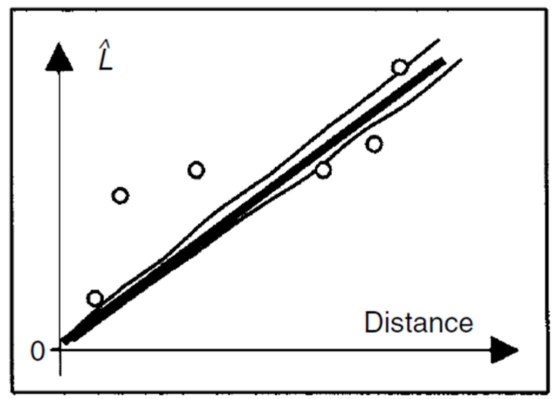
Ripley's K and L

$$\widehat{K}(t) = \frac{\lambda^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n} I_t(e_i, e_j)}{n}$$

$$\text{for } i \neq j \text{ and } t > 0$$

$$\widehat{L}(t) = t - \left(\frac{\widehat{K}(t)}{\pi}\right)^{1/2}$$

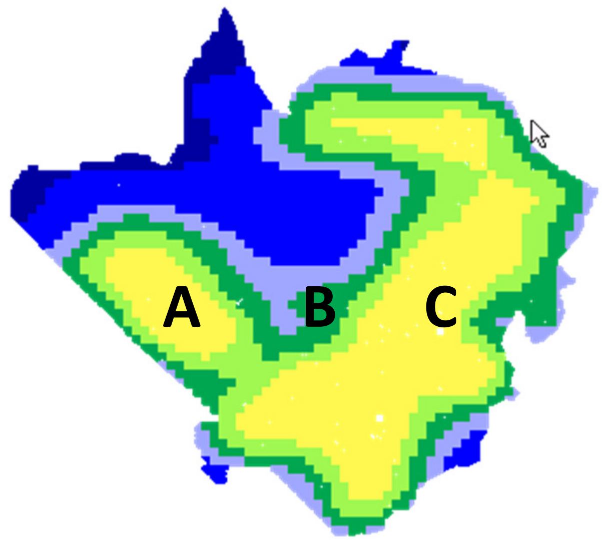




- Expected value of L(t) under a Poisson process is zero
- Positive values = spatial clustering
- Negative values = spatial segregation

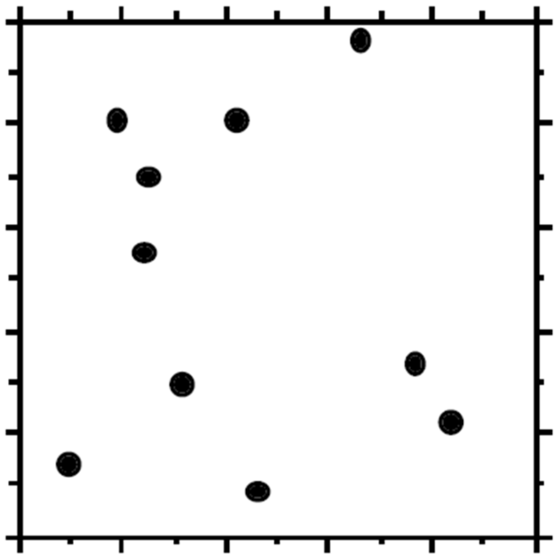
Correlogram

- Autocorrelation value plotted against distance classes



Correlogram

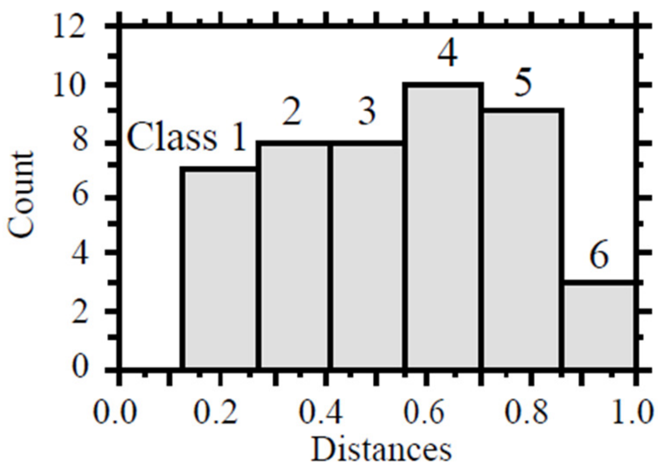
(a)



(b)

$$\mathbf{D} = \begin{bmatrix} 0.00 & 0.52 & 0.74 & 0.20 & 0.31 & 0.29 & 0.72 & 0.72 & 0.59 & 0. \\ 0.52 & 0.00 & 0.27 & 0.41 & 0.27 & 0.75 & 0.52 & 0.25 & 0.45 & 0. \\ 0.74 & 0.27 & 0.00 & 0.58 & 0.44 & 1.00 & 0.74 & 0.37 & 0.70 & 0. \\ 0.20 & 0.41 & 0.58 & 0.00 & 0.15 & 0.49 & 0.76 & 0.65 & 0.63 & 0. \\ 0.31 & 0.27 & 0.44 & 0.15 & 0.00 & 0.59 & 0.68 & 0.51 & 0.57 & 0. \\ 0.29 & 0.75 & 1.00 & 0.49 & 0.59 & 0.00 & 0.76 & 0.90 & 0.64 & 0. \\ 0.72 & 0.52 & 0.74 & 0.76 & 0.68 & 0.76 & 0.00 & 0.40 & 0.13 & 0. \\ 0.72 & 0.25 & 0.37 & 0.65 & 0.51 & 0.90 & 0.40 & 0.00 & 0.39 & 0. \\ 0.59 & 0.45 & 0.70 & 0.63 & 0.57 & 0.64 & 0.13 & 0.39 & 0.00 & 0. \\ 0.23 & 0.53 & 0.67 & 0.12 & 0.26 & 0.50 & 0.87 & 0.77 & 0.74 & 0. \end{bmatrix}$$

(c)



(d)

$$\mathbf{D} = \begin{bmatrix} - & & & & & & & & & \\ 3 & - & & & & & & & & \\ 5 & 2 & - & & & & & & & \\ 1 & 2 & 4 & - & & & & & & \\ 2 & 2 & 3 & 1 & - & & & & & \\ 2 & 5 & 6 & 3 & 4 & - & & & & \\ 5 & 3 & 5 & 5 & 4 & 5 & - & & & \\ 5 & 1 & 2 & 4 & 3 & 6 & 2 & - & & \\ 4 & 3 & 4 & 4 & 4 & 4 & 1 & 2 & - & \\ 1 & 3 & 4 & 1 & 1 & 3 & 6 & 5 & 5 & - \end{bmatrix}$$

Correlogram correlations

- Autocorrelation calculation
  - Pearson's correlation coefficient

+ +
- -
+ -

$$\rho = \frac{\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\left[\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2\right]^{1/2}},$$

Variance

  - Moran's I
  - Geary's C

Moran's I

Pearson's correlation

$$\rho = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2\right]^{1/2}},$$

weight for observation pair (neighbouring function)

Moran's I

$$I = \frac{n}{S_0}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}w_{ij}(x_i-\bar{x})(x_j-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2},$$

Sum of all weights

Moran’s I

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2},$$

- Positive correlation   I > 0 (I = 1 perfect correlation)
- Negative correlation I < 0 (I = -1 perfect dispersion
- No correlation I = 0

- Null hypothesis expected value of I :   E(I) = -1/(N -1)

- Measure of “global” spatial autocorrelation

Geary’s C

- Inversely related to Moran’s I
- More sensitive to local spatial autocorrelation

Moran’s I: $$I(d) = \frac{\frac{1}{W}\sum_{h=1}^{n}\sum_{i=1}^{n} w_{hi}(y_h - \bar{y})(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad \text{for } h \neq i$$

Geary’s c: $$c(d) = \frac{\frac{1}{2W}\sum_{h=1}^{n}\sum_{i=1}^{n} w_{hi}(y_h - y_i)^2}{\frac{1}{(n-1)}\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad \text{for } h \neq i$$

Geary's C

Geary's c:

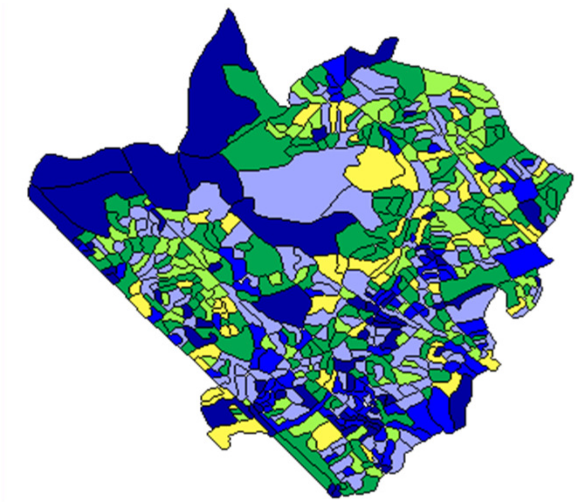
$$c(d) = \frac{\frac{1}{2W}\sum_{h=1}^{n}\sum_{i=1}^{n} w_{hi}(y_h - y_i)^2}{\frac{1}{(n-1)}\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad \text{for } h \neq i$$

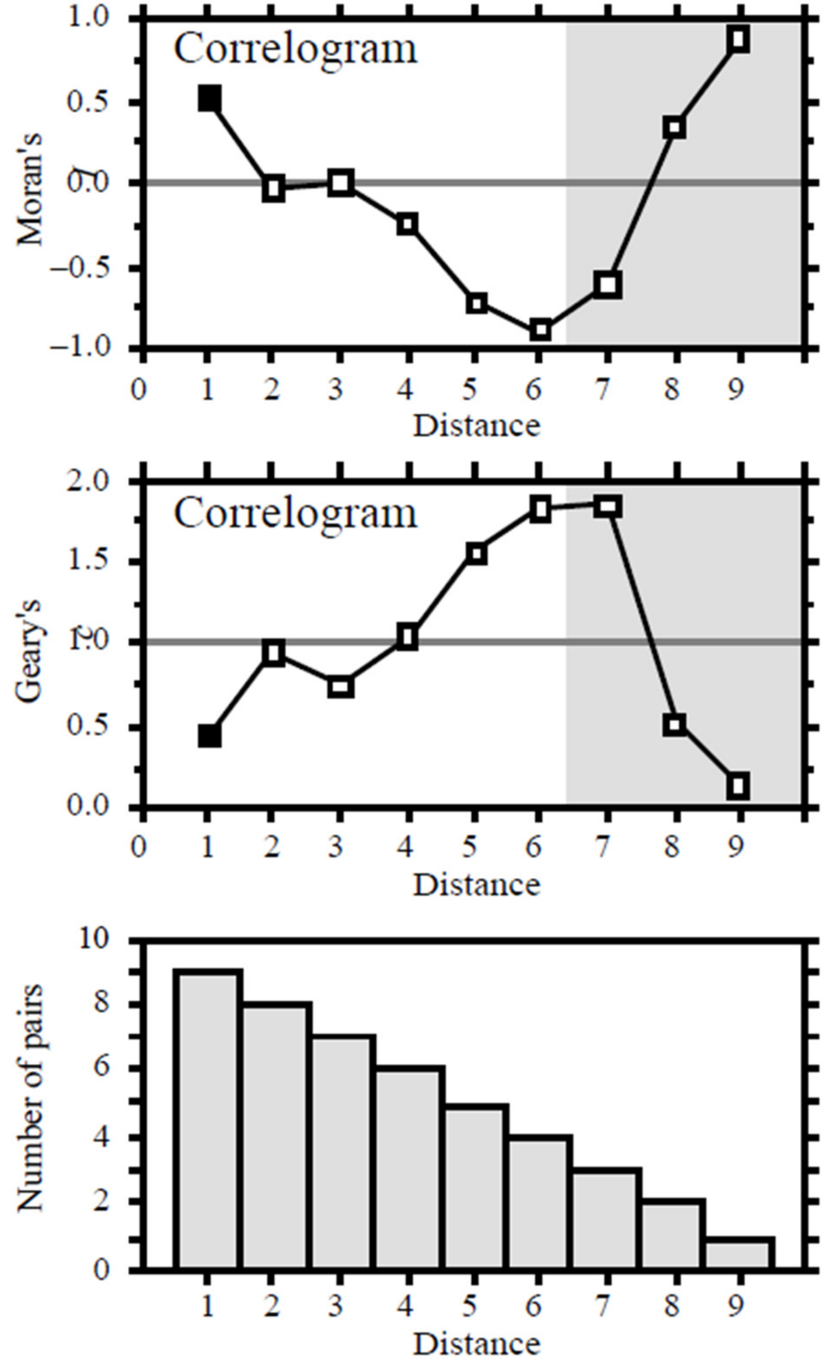
- Range : 0 to 2
- No correlation C = 1
- Positive correlation 0 < C < 1
- Negative correlation C > 1

- Null hypothesis expected value of C : E(C) = 1

- More sensitive to local spatial autocorrelation

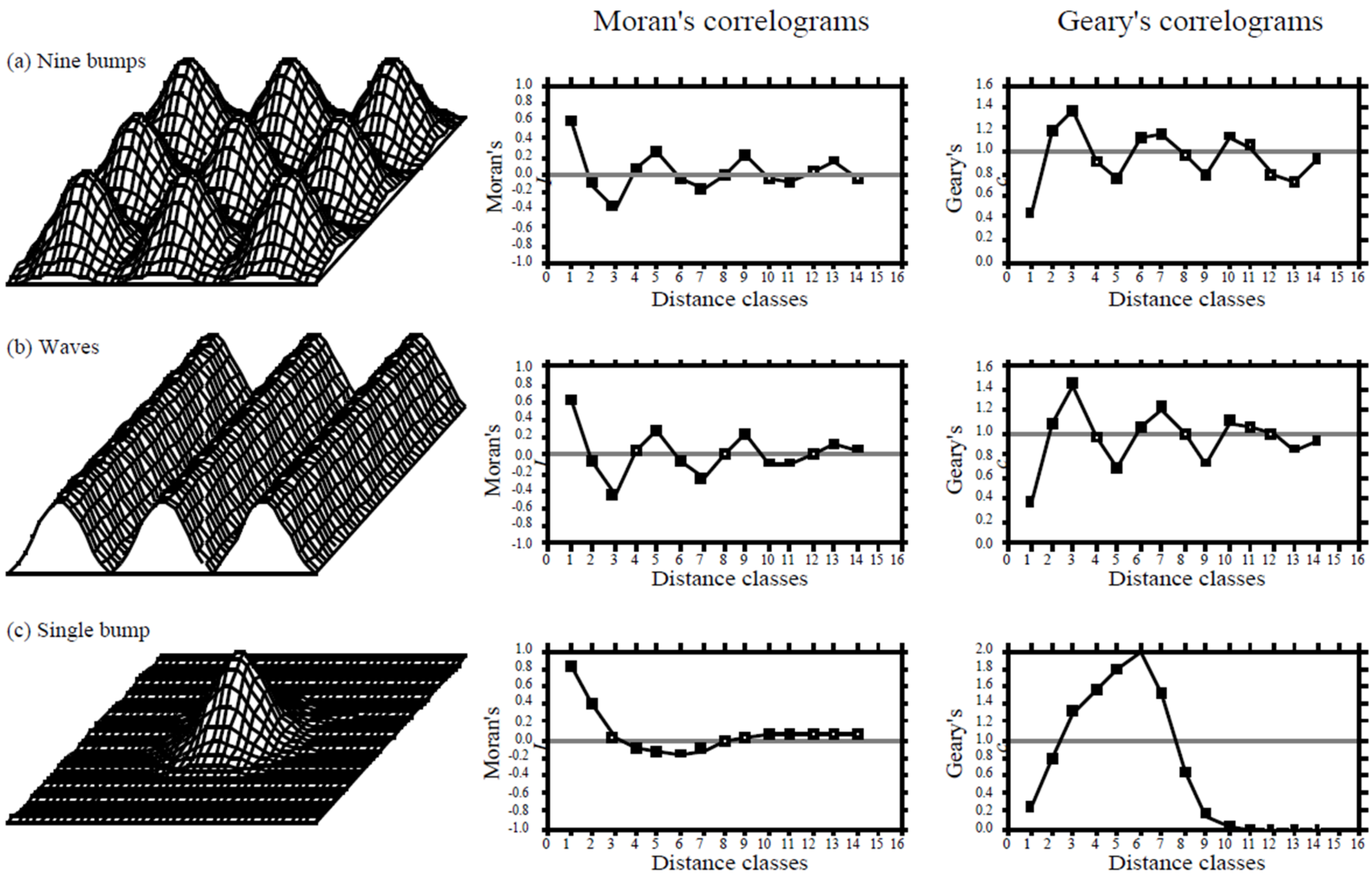
Correlogram





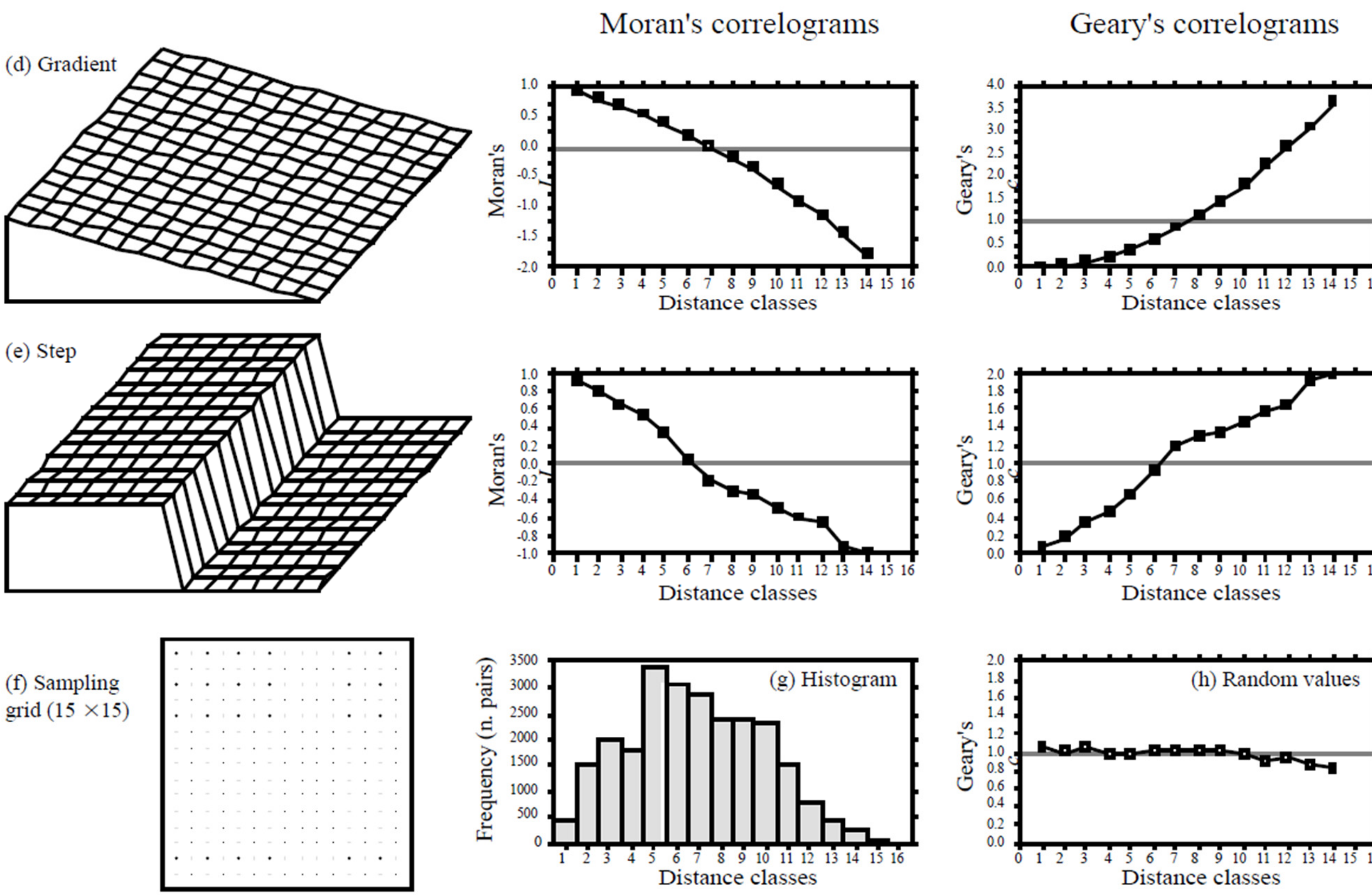
- Distance at which correlogram lines cross the E(I), E(C) value indicates “patch size” or “spatial range of the pattern”

Correlograms



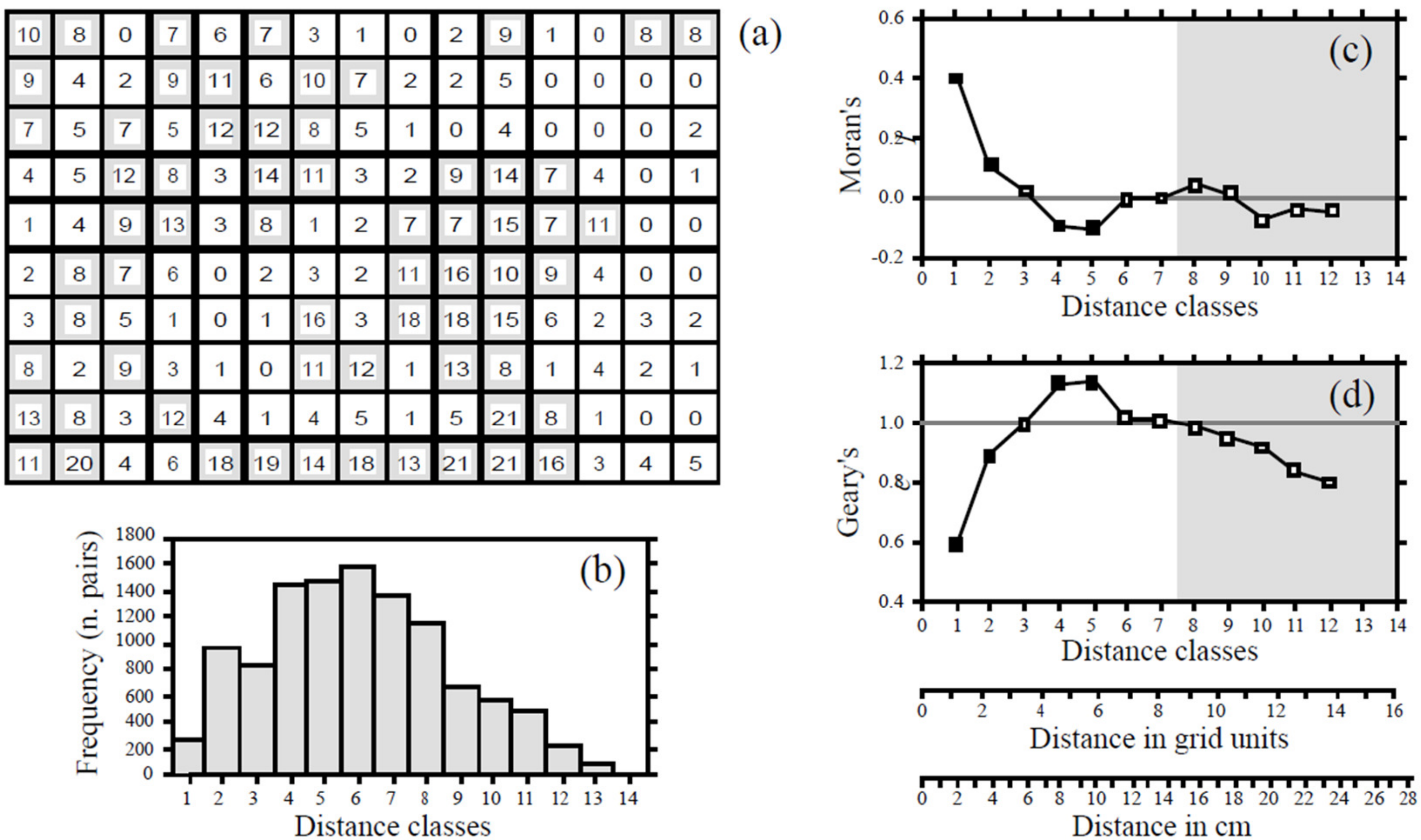
Legendre, P., and L. F. J. Legendre. 1998. Numerical Ecology. Elsevier.

Correlogram



Correlogram example

- Counts of adult barnacles per plate (1.7 by 1.7)



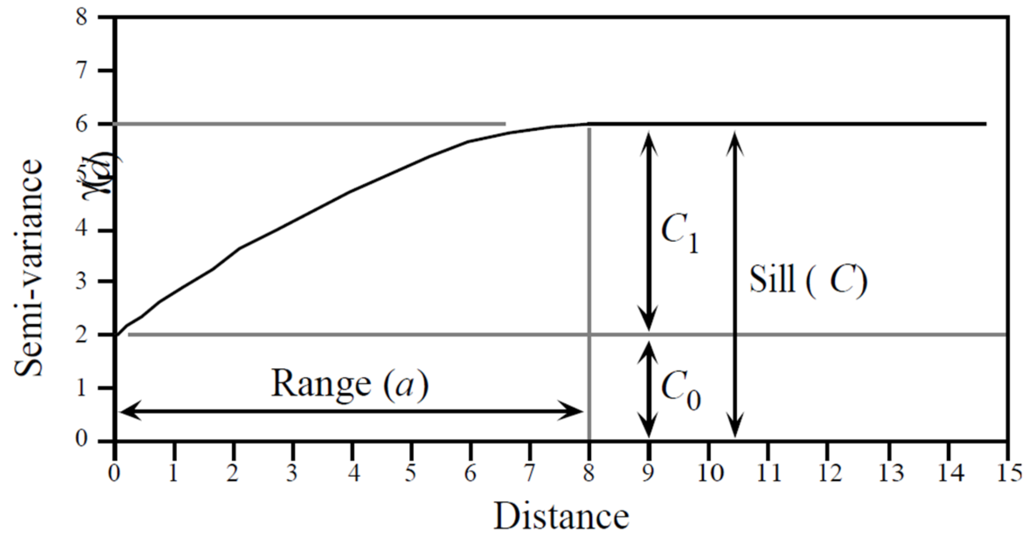
Variograms (semi-variograms)

- Based in geostatistical methods
- Used to identify and **model** spatial patterns
- Spatial structure of the data is estimated based on spatial variance (experimental variogram)
- Estimated spatial structure and parameters are then used in kriging models

$$\gamma(d) = \frac{1}{2W} \sum_{h=1}^{n} \sum_{i=1}^{n} w_{hi} (y_h - y_i)^2 \quad \text{for } h \neq i$$

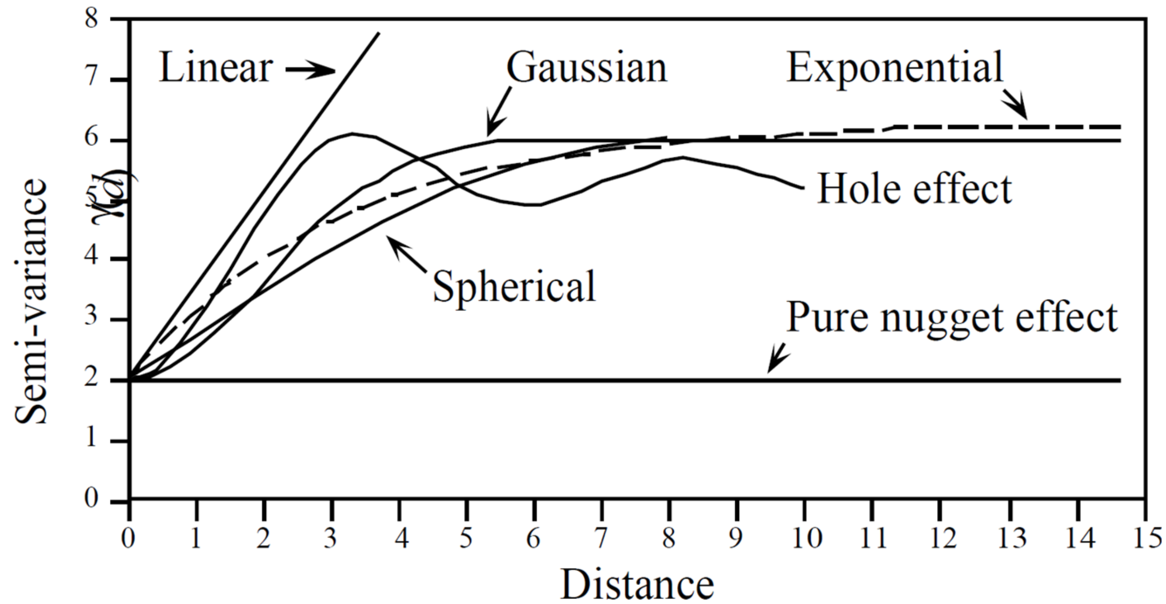
- Non-standardized form of Geary's C

Variograms



- Shape of the experimental variogram can be used to characterize the spatial structure of the data
  - Spatial range (a): distance over which the underlying spatial process occurs
  - Nugget effect (Co): estimate of error inherent in the measurement
  - Sill (C): quantification of the intensity of the spatial pattern

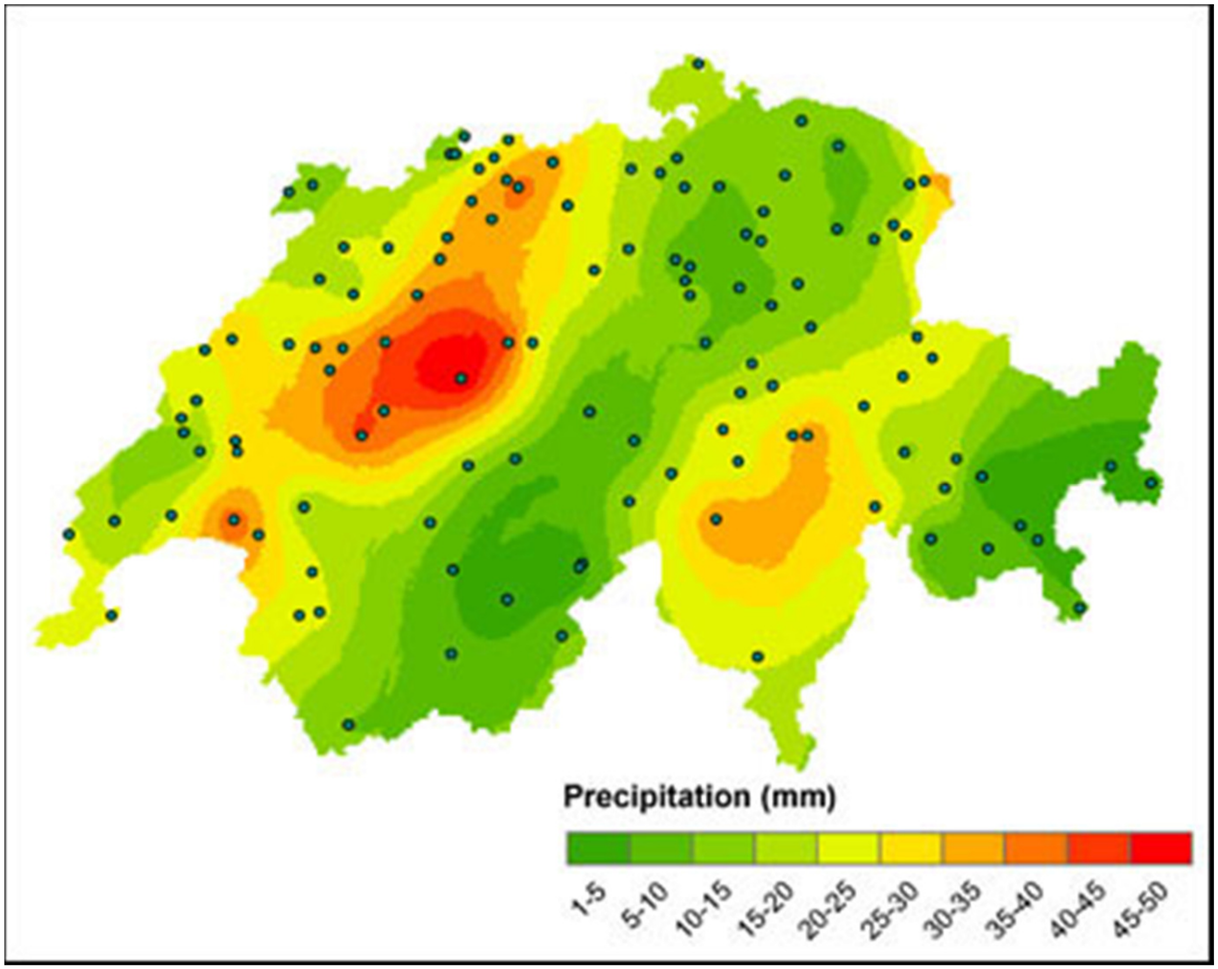
Spatial model selection and parameterization



- Spherical model: $\gamma(d) = C_0 + C_1\left[1.5\frac{d}{a} - 0.5\left(\frac{d}{a}\right)^3\right]$ if $d \le a$; $\gamma(d) = C$ if $d > a$.
- Exponential model: $\gamma(d) = C_0 + C_1\left[1 - \exp\left(-3\,\frac{d}{a}\right)\right]$.
- Gaussian model: $\gamma(d) = C_0 + C_1\left[1 - \exp\left(-3\,\frac{d^2}{a^2}\right)\right]$.

Kriging: used defined spatial model to predict values at unmeasured locations

Kriging



Precipitation (mm)
1-5
5-10
10-15
15-20
20-25
25-30
30-35
35-40
40-45
45-50

Spatial autoregressive models

- Simultaneous autoregressive (SAR) models
- Conditional autoregressive (CAR) models

$$E\left(y_i \mid y_{-i}\right) = x_i\beta + \sum_{j=1}^{n} c_{ij}\left[y_j - x_i\beta\right]$$