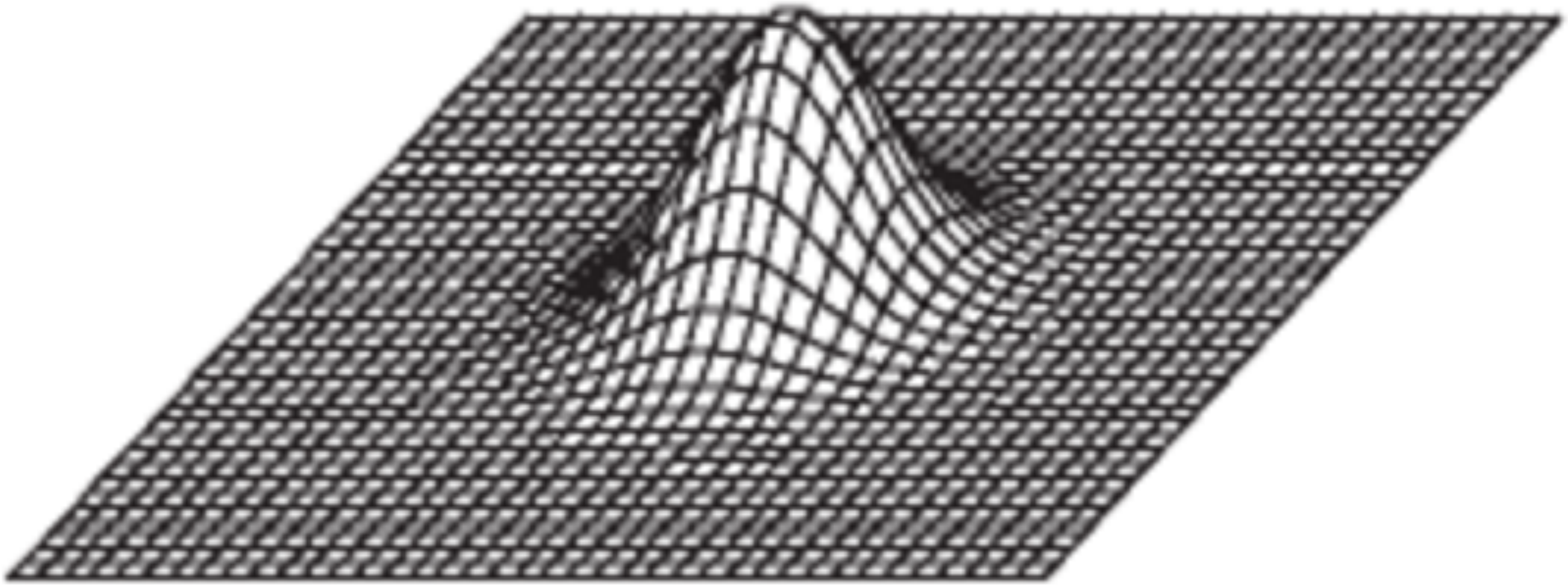
Spatial Analysis

Rachel & Kat (with a little help from Gill Green, Geography)

Looking for Spatial Auto-correlation

(c) Single bump



Formula for Moran's I

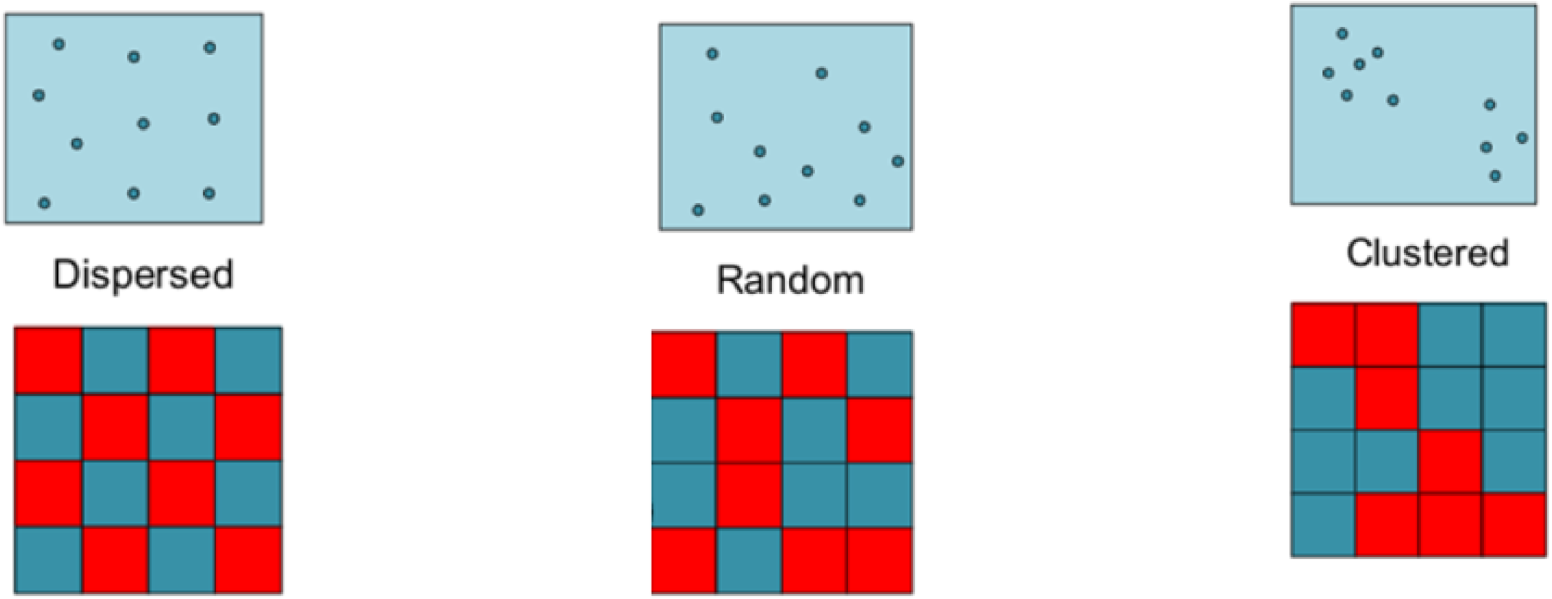
$$I = \frac{N \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\left(\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \right) \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Where:

N is the number of observations (points or polygons)
$\bar{X}$ is the mean of the variable
X_i is the variable value at a particular location
X_j is the variable value at another location
W_{ij} is a weight indexing location of *i* relative to *j*

Interpreting Moran's I

-1 --- --- 0 --- --- 1

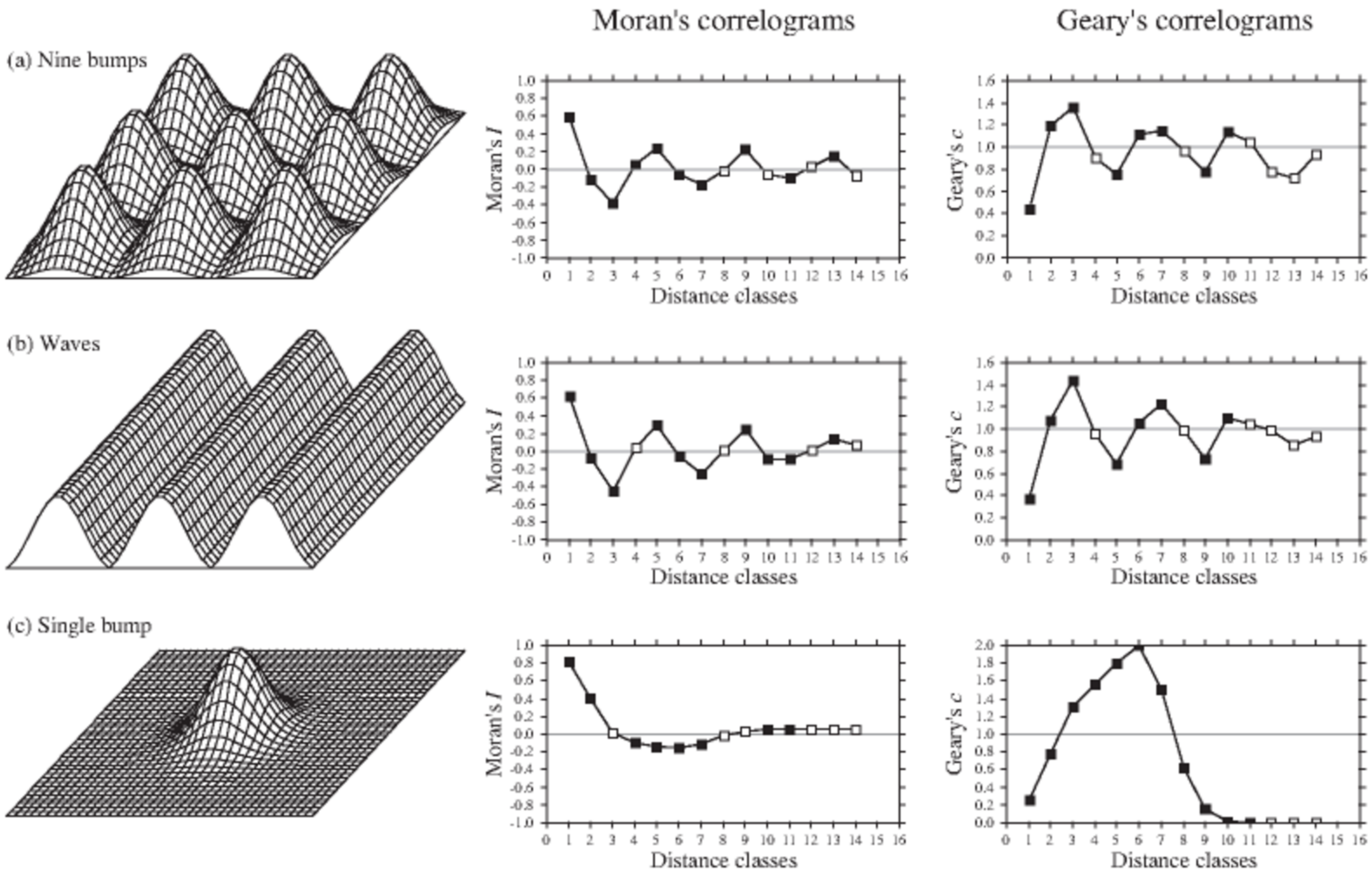


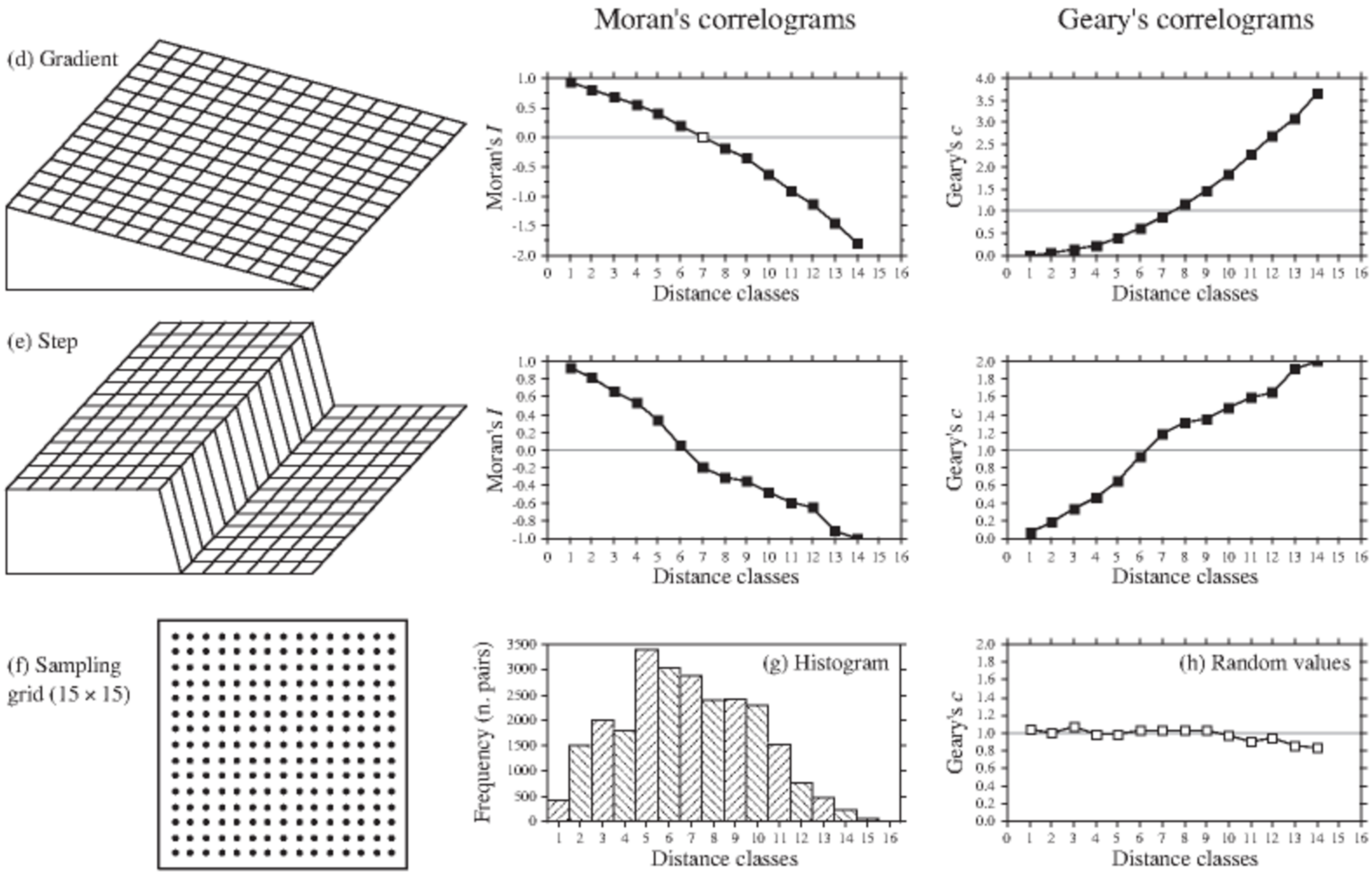
6. Geary's C

- Harder to interpret than Moran's I, but effectively the inverse of Moran's I.
- Moran's I null hypothesis expects that neighbors do not co-vary in a consistent way (covariance structure). Geary's C null hypothesis is that there is no consistency to the differences between neighbors – i.e., sometimes the differences are large and sometimes small.

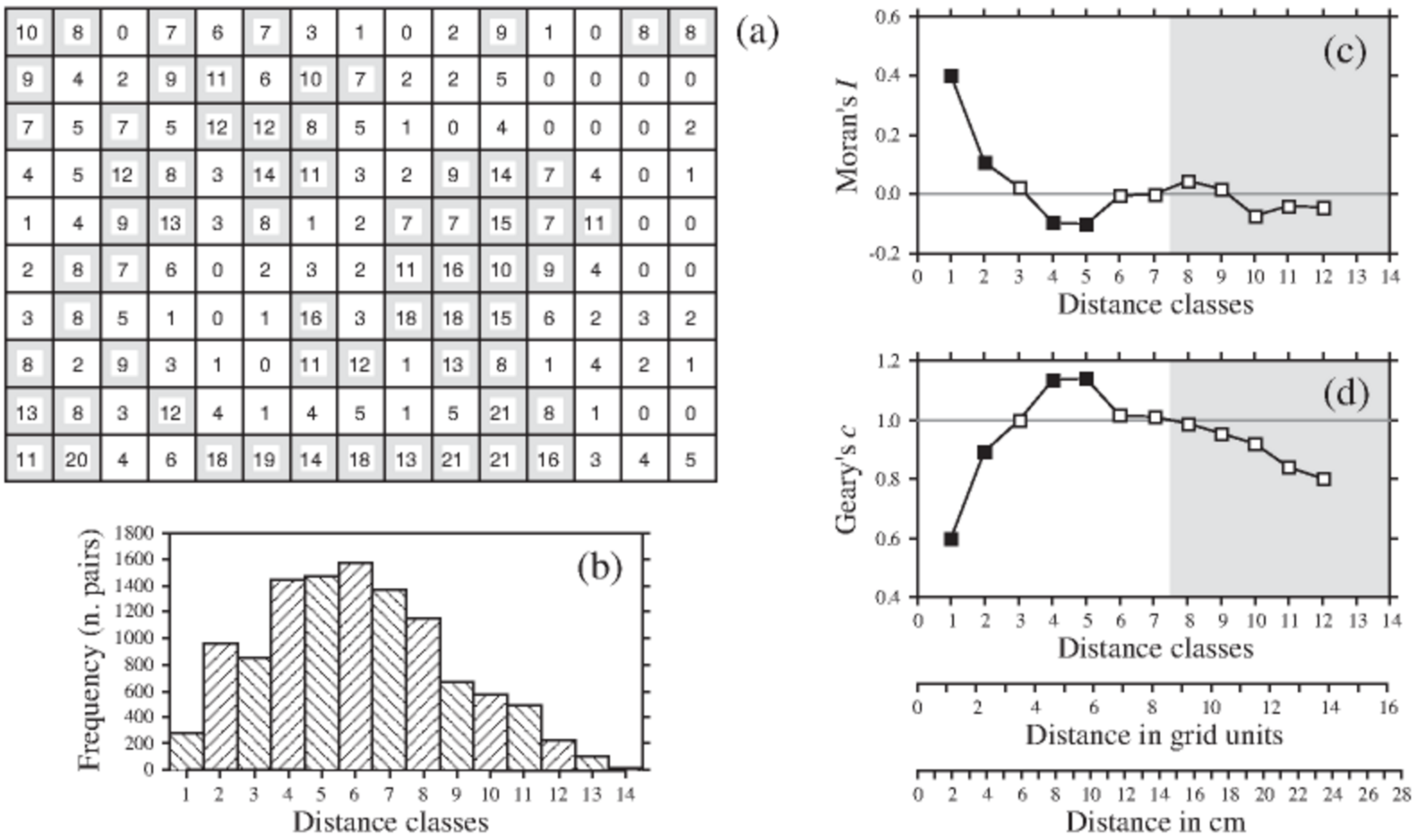
$$C = \frac{(N-1)\sum_i \sum_j w_{ij}(X_i - X_j)^2}{2W\sum_i (X_i - \bar{X})^2}$$

where N is the number of spatial units indexed by i and j; X is the variable of interest; $\bar{X}$ is the mean of X; w_{ij} is a matrix of spatial weights; and W is the sum of all w_{ij}.


Moran's correlograms
Geary's correlograms
(a) Nine bumps
(b) Waves
(c) Single bump
Moran's I
Geary's c
Distance classes

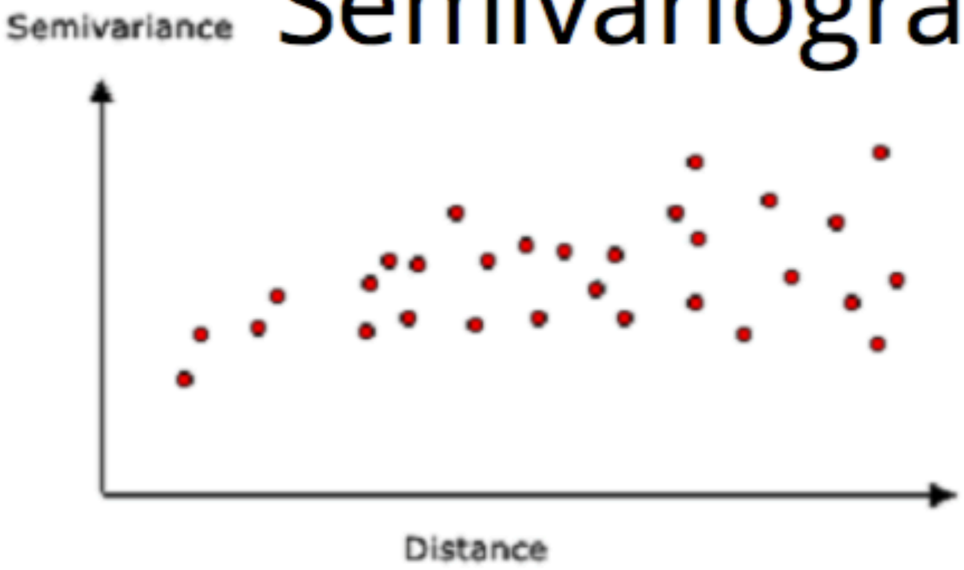


Moran's correlograms
Geary's correlograms
(d) Gradient
(e) Step
(f) Sampling grid (15 × 15)
(g) Histogram
(h) Random values
Moran's I
Geary's c
Frequency (n. pairs)
Distance classes



10 8 0 7 6 7 3 1 0 2 9 1 0 8 8
9 4 2 9 11 6 10 7 2 2 5 0 0 0 0
7 5 7 5 12 12 8 5 1 0 4 0 0 0 2
4 5 12 8 3 14 11 3 2 9 14 7 4 0 1
1 4 9 13 3 8 1 2 7 7 15 7 11 0 0
2 8 7 6 0 2 3 2 11 16 10 9 4 0 0
3 8 5 1 0 1 16 3 18 18 15 6 2 3 2
8 2 9 3 1 0 11 12 1 13 8 1 4 2 1
13 8 3 12 4 1 4 5 1 5 21 8 1 0 0
11 20 4 6 18 19 14 18 13 21 21 16 3 4 5
(a)
(b)
Frequency (n. pairs)
Distance classes
(c)
Moran's I
Distance classes
(d)
Geary's c
Distance classes
Distance in grid units
Distance in cm

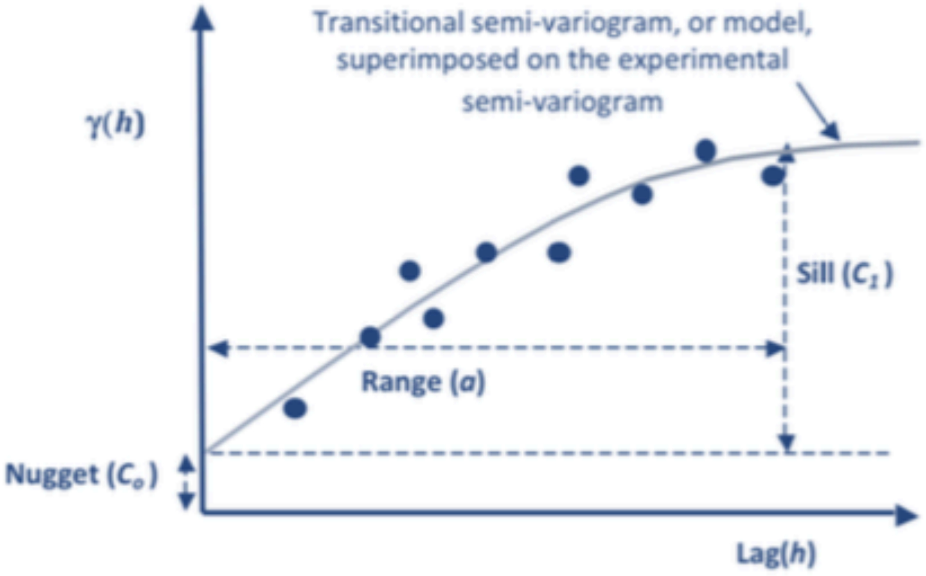
Semivariogram



- Variance between points is represented over space. The above shows how variance between point values tends to increase as distance between points increases.
- The values of the points could be rainfall, elevation, demographic continuous data, number of people killed in a battle, percentage of gold in a soil sample…

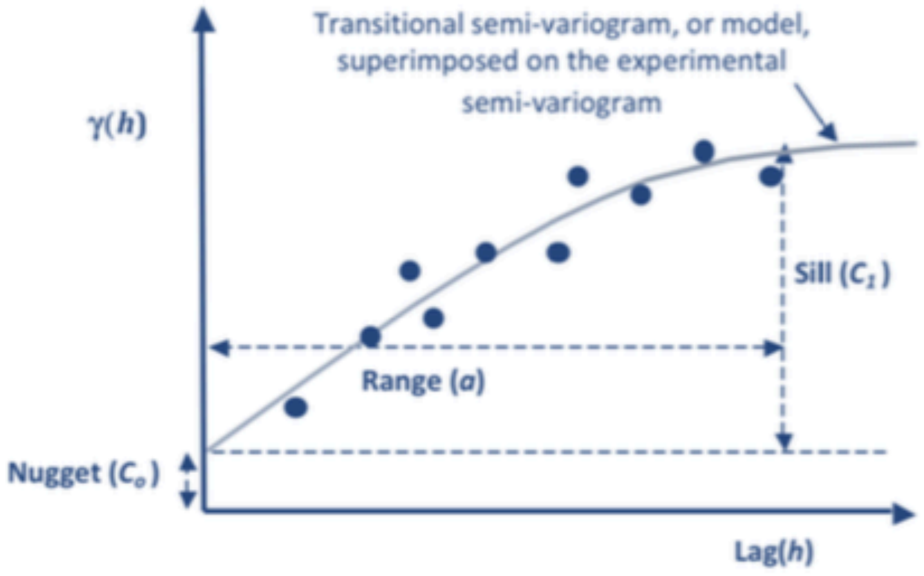
Semivariogram

- **Nugget** represents subgrid scale variation that cannot be estimated by reasons of the sampling grid spacing or measurement error.
- **Range** is a measure of the degree of association or correlation between data points represented in terms of distance. How far we can go from a data point before going beyond the extend of its influence.



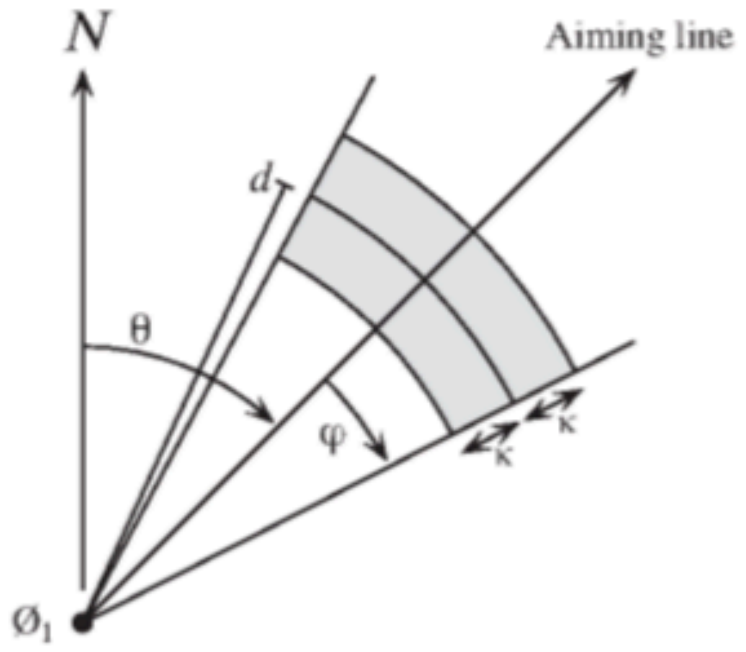
Semivariogram

- **Sill** is the value of the semi-variance as the lag tends towards infinity – in non-standardized data it is equal to the total variance of the dataset.
- **Lag** refers to tolerance around distance bands. Bands can be based on Euclidean distance or other theoretically important measures of distance (e.g. log).



Semivariogram: Directionality

- Directionality is isotropic (non-directional compared to anistropic) over the entire surface in semivariogram calculation.
- In order to make it anistropic, selection bands can be placed on a geographic axis to analyze only specific points.
- In regional studies, a semivariogram per region would be best.



Search parameters for pairs of points in directional variograms and correlograms. From an observed study site $Ø_1$, an aiming line is drawn in the direction determined by angle θ (usually by reference to the geographic north, indicated by N in the figure). The angular tolerance parameter φ determines the search zone (grey) laterally whereas parameter κ sets the tolerance along the aiming line for each distance class d. Points within the search window (in gray) are included in the calculation of $I(d)$, $c(d)$ or $\gamma(d)$.

Multivariate variograms???

- Cross-variogram: partitions covariance between variables among distance class *d*
  - Creates variogram matrix *C*
- Can compute spatial cross-correlograms as in Chapter 12 (detailed example in book)
- Plots of spatial covariance, semi-variance, Geary's *C* and r(d) are equivalent under 2nd order stationarity

Multivariate variograms???

- Multivariate Mantel correlogram
  - Compare Matrix Y of multivariate distances to model matrix X, different for each distance class *d*
  - Plot Mantel stats against distance classes to give multivariate correlogram
  - Must correct for multiple tests
  - See example 836-837 (digital page numbers)
  - When model matrix *X* is is a similarity matrix, positive Mantel stat = positive spatial autocorrelation

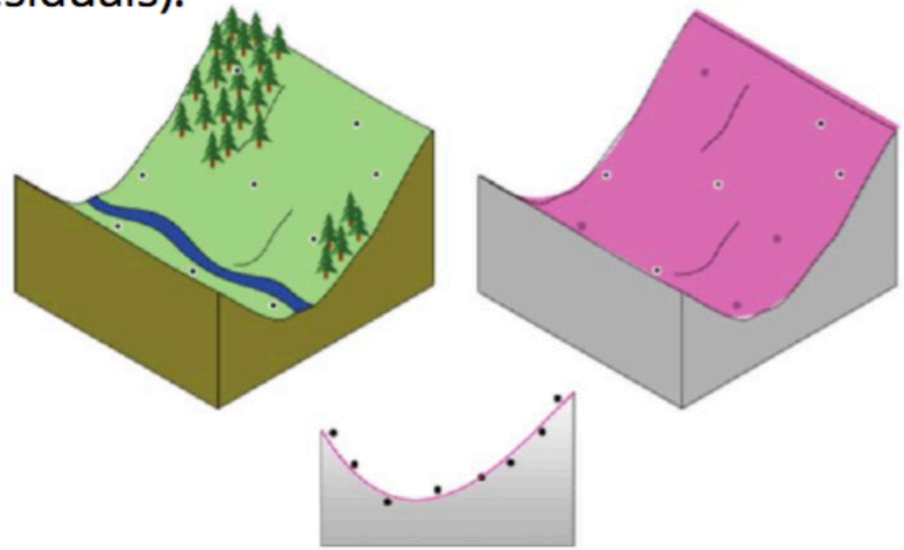
R packages

- Compute empirical variograms
  - *variog()* of geoR
  - *variogram()* of nlme
  - *vario()* of pastecs
  - *est.variogram()* of sgeostat
- Multivariate variograms
  - *mso()* of vegan
  - *Mantel.correlog()* of vegan

Changing gears to INTERPOLATION TECHNIQUES (13.2)

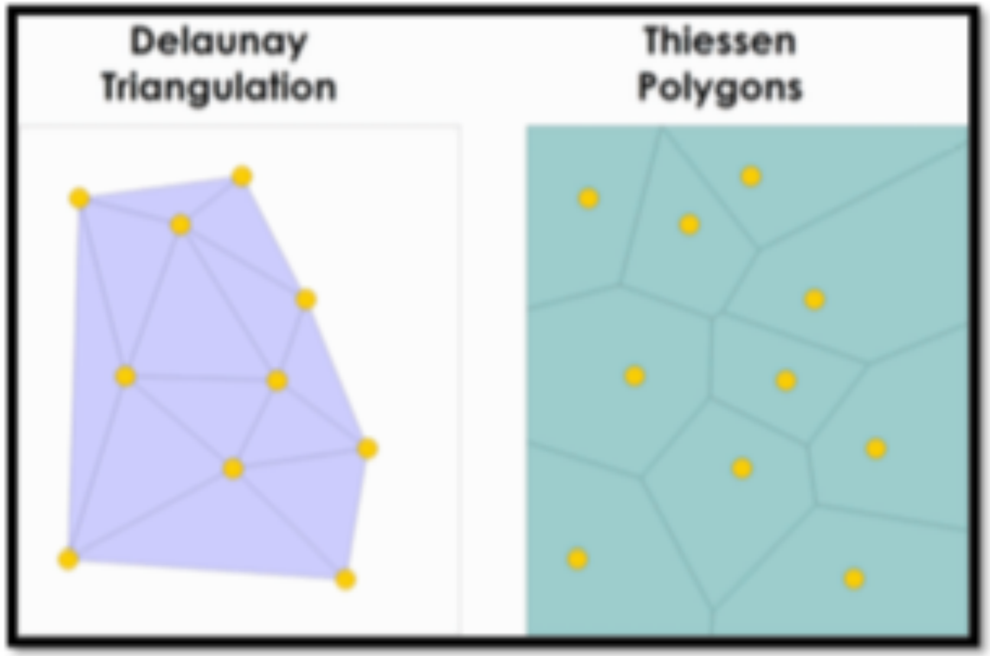
- Trend-surface analysis: uses polynomials to fit surface
  - If extent of trend = size of study area, can model using degree 3, 2, or 1 polynomial

- Adds polynomials to make more bends... Second order polynomial below.
- Least-squares approach to interpolation (lowest residuals).



More interpolations

- Voronoi Polygons: give each node of the grid a value based on the point closest to it
- Delaunay triangulation: assign central value based on value of three vertices of triangle
- Search circle: draw circle of some distance around each grid point and use values of points within circle



Inverse Distance Weighted (IDW)

- Inverse Distance Weighted (IDW) is a method of interpolation that estimates cell values by averaging the values of sample data points in the neighborhood of each processing cell.
- The closer a point is to the center of the cell being estimated, the more influence, or weight, it has in the averaging process.

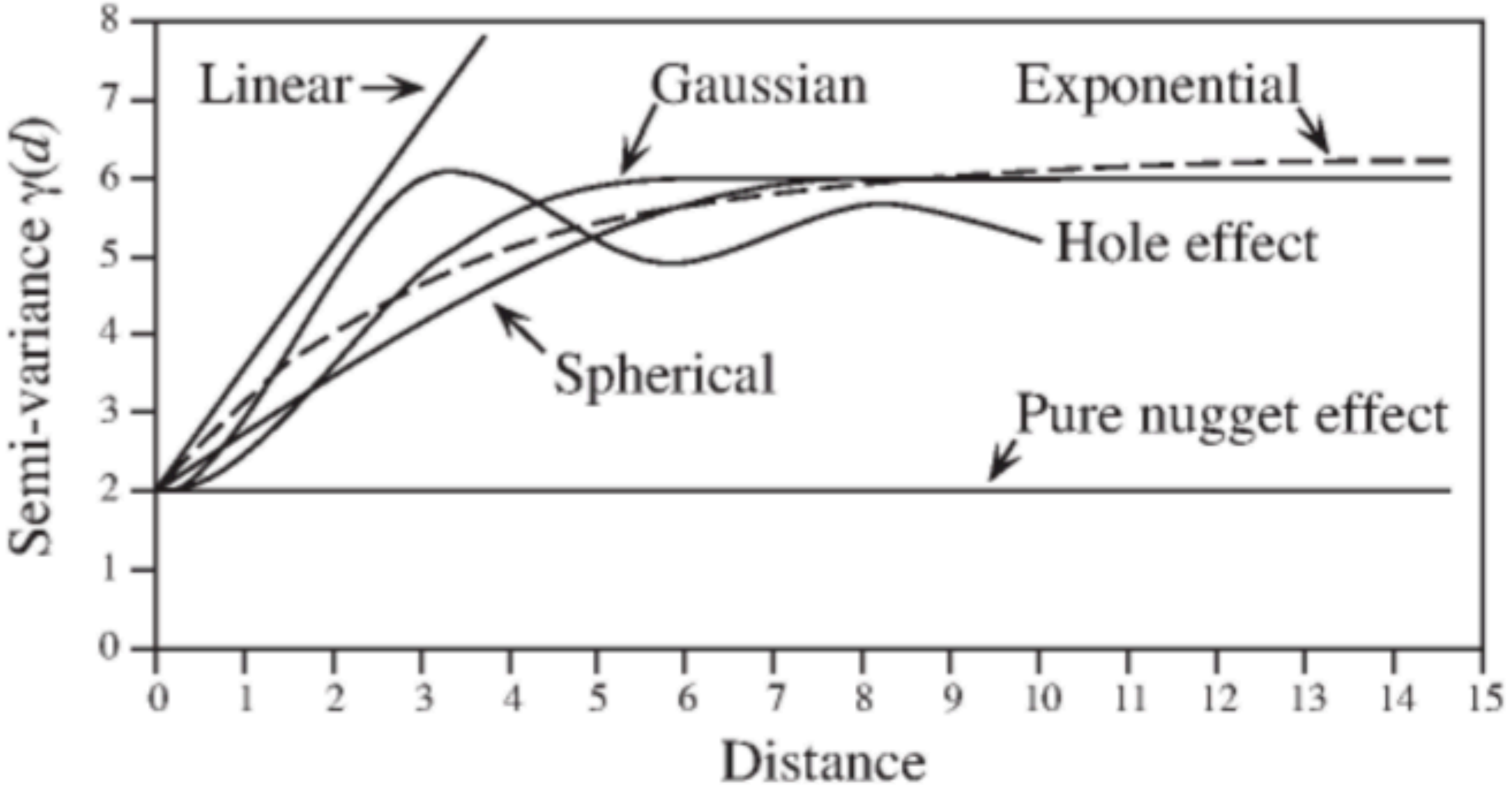
Kriging

- Similar to IDW in that it weighs surrounding values.
- But weights are not just based on distance, they are based on overall spatial arrangement of points and the variance of values over space.
- Uses a semivariogram model.
- Semivariogram modeling (maps variance over surface depending on layout of points and point values).
- The semivariogram is based on modeling the (squared) differences in the z-values as a function of the distances between the known points.

Kriging

- The first step in ordinary kriging is to construct a semivariogram from the scatter point set to be interpolated.
- A semivariogram consists of two parts:
  - an experimental semivariogram
  - a model semivariogram.
- The experimental semivariogram is found by calculating the variance (g) of each point in the set with respect to each of the other points and plotting the variances versus distance (h) between the points.

Models to fit to semivariogram



Some confusing math

$$\mathbf{C} \cdot \mathbf{w} = \mathbf{d}$$

$$\begin{bmatrix} c_{11} & \dots & c_{1n} & 1 \\ \cdot & \dots & \cdot & 1 \\ \cdot & \dots & \cdot & 1 \\ c_{n1} & \dots & c_{nn} & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ \cdot \\ \cdot \\ w_n \\ \lambda \end{bmatrix} = \begin{bmatrix} d_1 \\ \cdot \\ \cdot \\ d_n \\ 1 \end{bmatrix} \qquad \textbf{(13.22)}$$

where **C** is the covariance matrix among the n points $Ø_i$ used in the estimation, i.e. the semi-variances corresponding to the distances separating the various pair of points, provided by the variogram model; **w** is the vector of weights to be estimated (with the constraint that the sum of weights must be 1); and **d** is a vector containing the covariances between the various points $Ø_i$ and the grid node to be estimated. This is where a variogram model becomes essential; it provides the weighting function for the entire map and is used to construct matrix **C** and vector **d** for each grid node to be estimated. Element λ in vector **w** is a Lagrange parameter (as in Section 4.4) introduced to minimize the variance of the estimates under the constraint $\Sigma w_i = 1$ (unbiasedness condition). The solution to this linear system is obtained by matrix inversion (Section 2.8):

$$\mathbf{w} = \mathbf{C}^{-1}\,\mathbf{d} \qquad \textbf{(13.23)}$$

More Kriging

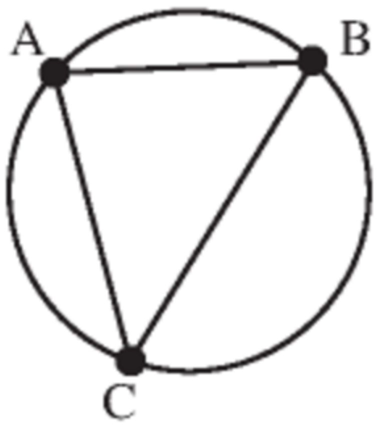
- Can estimate error
- If anisotropy is present, can use Kriging in variograms of 2+ directions
- Cross-validation provides measure of fit
  - Remove observation, then estimate it

- Measure of fit of other interpolated data: mean absolute error, mean squared error, Euclidean distance, correlation coefficient

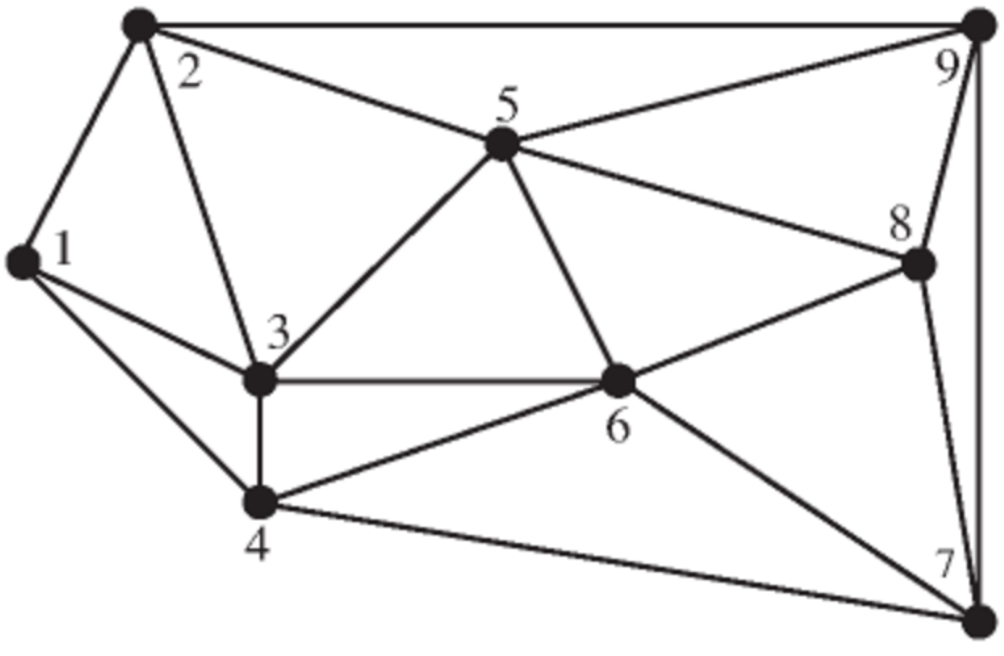
R packages (note: also available in most GIS programs)

- Trend-surface analysis
  - *lm()* of stats
- Adjusting variogram models to data
  - *eyefit()* of geoR
- Conventional kriging
  - *krige.conv()* in geoR

Spatial Clustering



Point identifiers	Coordinates X	Coordinates Y
1	0	3
2	1	5
3	2	2
4	2	1
5	4	4
6	5	2
7	8	0
8	7.5	3
9	8	5



19 edges form the Delaunay triangulation:

1–2	1–3	1–4	2–3	2–5	2–9	3–4
3–5	3–6	4–6	4–7	5–6	5–8	5–9
6–7	6–8	7–8	7–9	8–9		

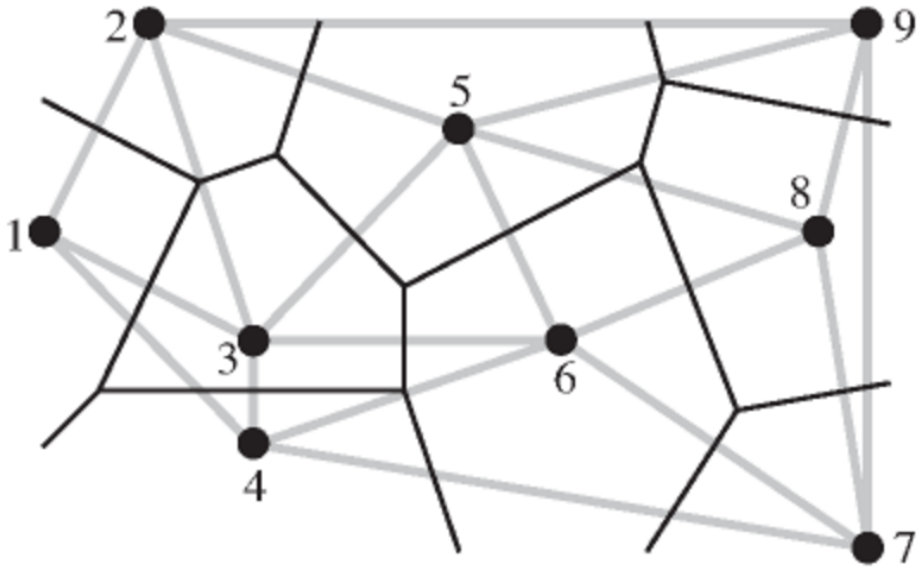
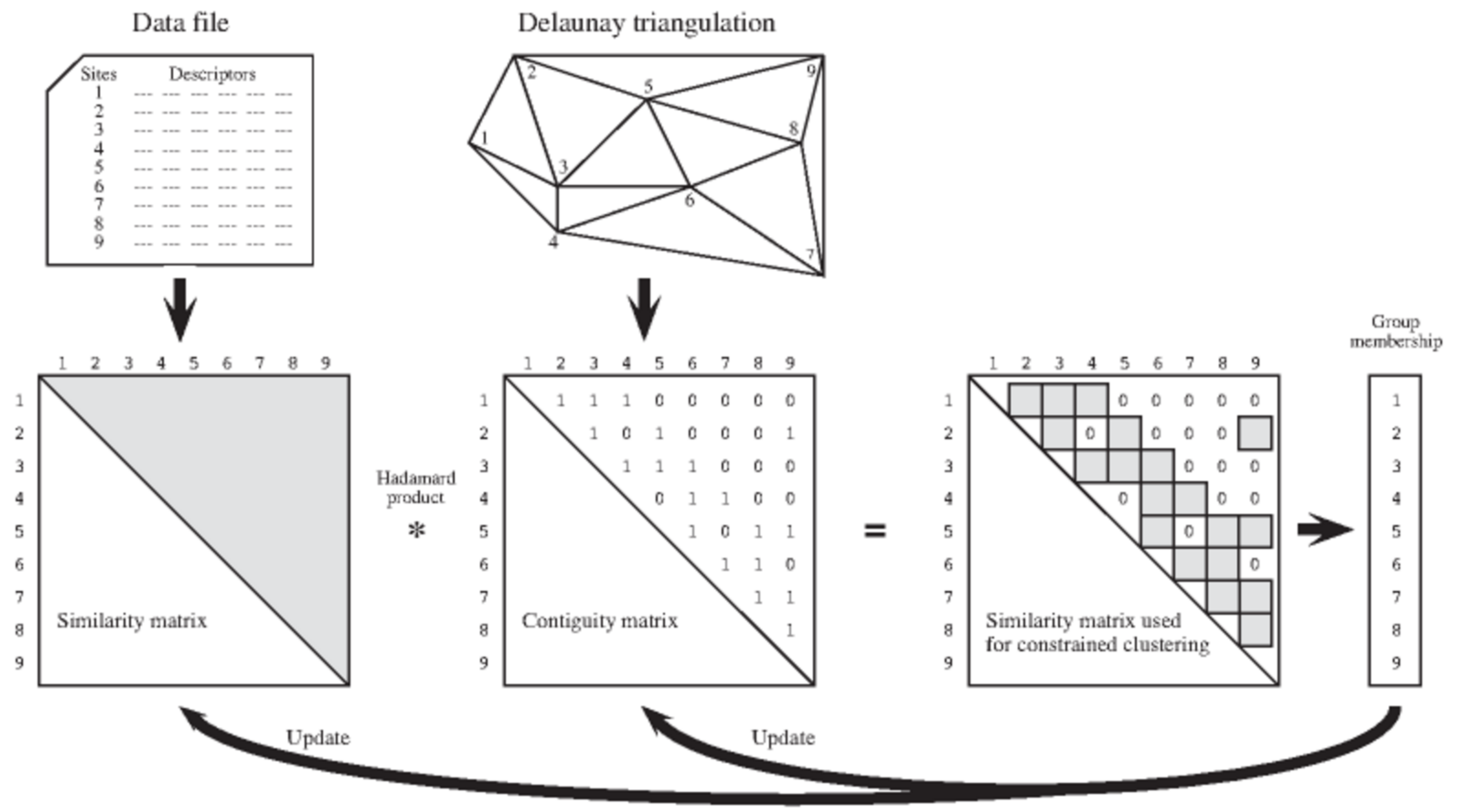
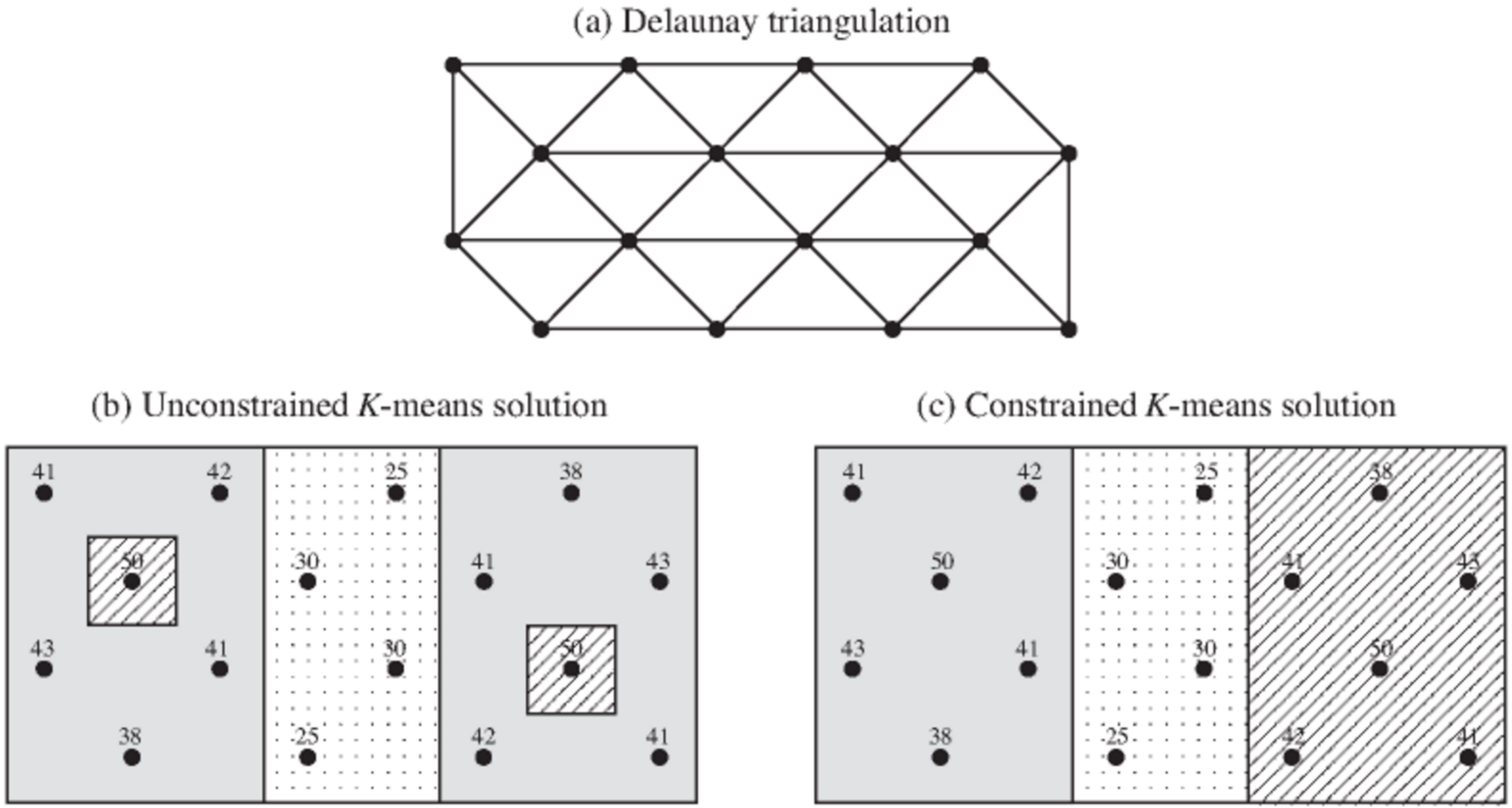


Figure 13.24 Delaunay triangulation (grey lines) and influence polygons (black lines) for the nine points of Fig. 13.22.

Constrained cluster analysis.




(a) Delaunay triangulation
(b) Unconstrained K-means solution
41 42 25 38
50 30 41 43
43 41 30 50
38 25 42 41
(c) Constrained K-means solution
41 42 25 38
50 30 41 43
43 41 30 50
38 25 42 41

Unconstrained & constrained ordination maps (13.4)

- Decompose variation into ordination axes (e.g. PCA)
- Map different ordination axes onto geographic surface
- Compare different maps
- Ordinations may be constrained to geographic structures
- Numerical example on 867-869 (digital pages)

- R packages: refer to Sections 9.5 (unconstrained) and 11.7 (constrained)

Spatial modelling through canonical analysis (13.5) ????????

- Model variation of variables as linear combination of environmental vairables <u>and</u> geographic coordinates
- Want to isolate effect of spatial structures separate from environment
  - Use partial regression/canonical regression & spatially constrained ordination maps
- Numerical example on 871-872 (digital pages)

- R packages: refer to Section 11.7

Additional resources (thanks, Dr. Green!)

This is free and a great resource written by the best in the field:

http://www.spatialanalysisonline.com/HTML/index.html

Here are some other options...

https://books.google.ca/books?hl=en&lr=&id=MzN_BwAAQBAJ&oi=fnd&pg=PP1&dq=textbook+spatial+statistics&ots=NK-fvX246J&sig=tE6bG-YjVDTRTfuWJk06yE3LWzY#v=onepage&q=textbook%20spatial%20statistics&f=false

https://books.google.ca/books?id=c0EP_6eYsjAC&pg=PA24&lpg=PA24&dq=fischer+spatial+statistics&source=bl&ots=JBG8rk1C_Z&sig=yRy_clf9R167sZxZWnUdwpo4flQ&hl=en&sa=X&ved=0ahUKEwjn5-CS1v_LAhVQyGMKHSXYDAsQ6AEIJDAA#v=onepage&q&f=false

https://books.google.ca/books/about/Elementary_Statistics_for_Geographers.html?id=p7YMOPuu8ugC

Fuentes taught me spatial stats:

ftp://cdsarc.u-strasbg.fr/contrib/MF/Handbook_of_Spatial_Statistics.pdf