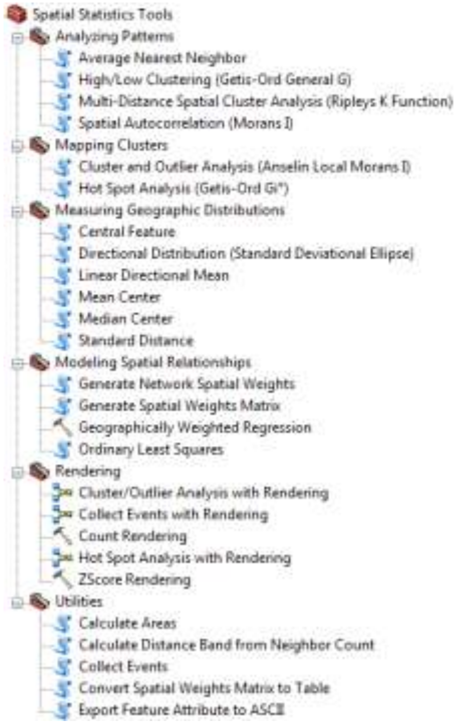
ArcToolBox: Spatial Statistics



Autocorrelation & Statistic Inference

- Positive autocorrelation -> redundant sampling
- The redundancy causes the selection of samples that are similar and thus, underestimates the variation of the population
- A smaller variation (standard error) results in an overestimation of z score

$$Z=(X_i-X_{mean})/SE$$

- Which means the sample mean is further away from the population mean and indicates the inclination to reject null hypothesis when it actually is true (i.e., commit a type I error)

Example

- Sample values: 0, 100 StDev = 70.71
- Sample values: 0, 100, 100, 100   StDev = 50

Moran's I

$$I = \left(\frac{n}{\sum_i \sum_j w_{ij}} \right) \left(\frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \right)$$

Examples of w_{ij}

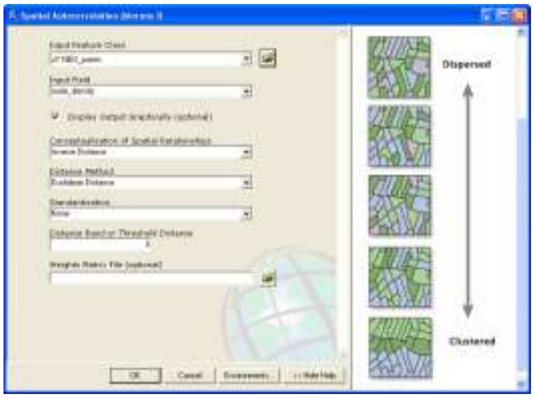
w_{ij} = 1 / d_{ij}
w_{ij} = 1 if i touches j, else 0

+1: clustering (positive spatial autocorrelation)
0: random
-1: dispersion (negative spatial autocorrelation)

Analyzing Patterns Toolset

Spatial Autocorrelation (Moran’s I)

- Moran's I -> +1.0: clustering
- Moran’s I -> -1.0: dispersion



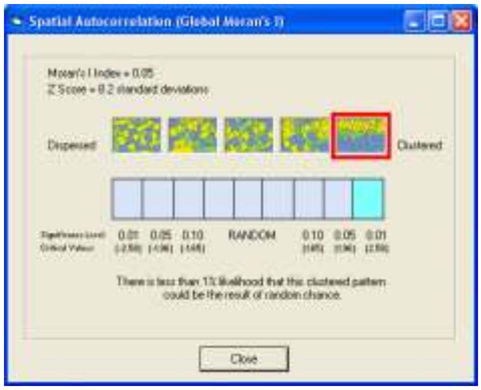
Conceptualization of Spatial Relationships

Specifies how spatial relationships between features are conceptualized.

- Inverse Distance—The impact of one feature on another feature decreases with distance.
- Inverse Distance Squared—Same as Inverse Distance, but the impact decreases more sharply over distance.
- Fixed Distance Band—Everything within a specified critical distance is included in the analysis; everything outside the critical distance is excluded.
- Zone of Indifference—A combination of Inverse Distance and Fixed Distance Band. Anything up to a critical distance has an impact on your analysis. Once that critical distance is exceeded, the level of impact quickly drops off.
- Polygon Contiguity (First Order)—The neighbors of each feature are only those with which the feature shares a boundary. All other features have no influence.
- Get Spatial Weights From File—Spatial relationships are defined in a spatial weights file. The pathname to the spatial weights file is specified in the Weights Matrix File parameter.

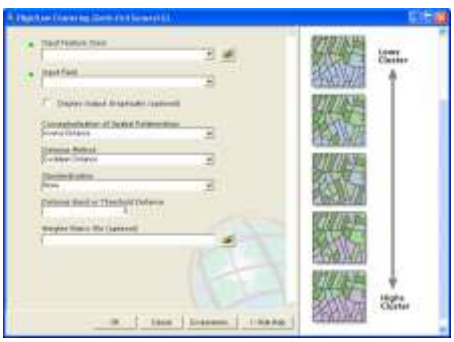
Output

- Moran's Index = 0.05
- Expected Index = -0.005
- Variance = 4.84E-5
- Z Score = 8.22



Analyzing Patterns Toolset

High/Low Clustering (Getis-Ord General G) – Inferential Statistics



Null Hypothesis: there is no spatial clustering of values

- Positive Z value: high values are clustered
- Negative Z value: low values are clustered

Getis-Ord General G

The General G statistic of overall spatial association is given as:

$$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j} x_i x_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j}, \ \forall j \neq i \quad (1)$$

where x_i and x_j are attribute values for features i and j, and $w_{i,j}$ is the spatial weight between feature i and j.

The z_G-score for the statistic is computed as:

$$z_G = \frac{G - \mathrm{E}[G]}{\sqrt{\mathrm{V}[G]}} \quad (2)$$

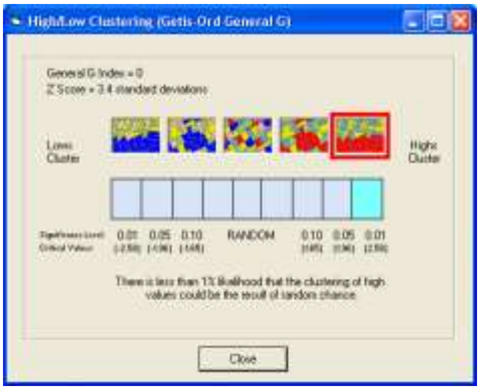
where:

$$\mathrm{E}[G] = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j}}{n(n-1)}, \ \forall j \neq i \quad (3)$$

$$\mathrm{V}[G] = \mathrm{E}[G^2] - \mathrm{E}[G]^2 \quad (4)$$

Output

- Observed General G = 3.49E-5
- Expected General G = 3.18E-5
- General G Variance = 8.66E-13
- Z Score = 3.44 Standard Deviations



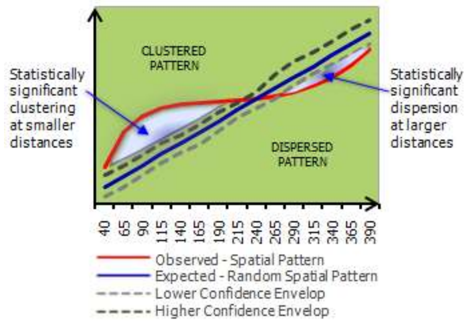
Ripley's K – L(d)

- Ripley's K function illustrates how the spatial clustering or dispersion of events changes when the distance band changes.

- Observed K (average number of events inside a circle of radius d)
  - Count # of events within distance band distance for each location
  - Calculate the mean of counts
  - Divide the mean by the average density (N/Area) within study area
  - Repeat for all distance bands (*d*)

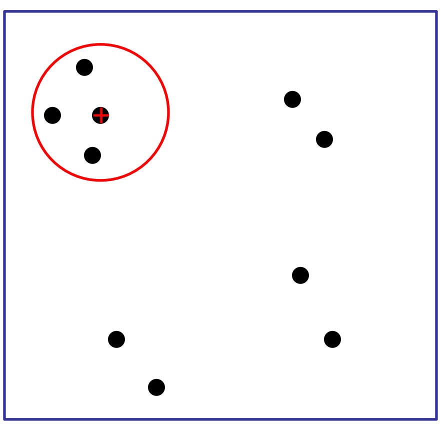
- Expected K (random)

$$E(K_r) = \lambda \pi d^2 / \lambda = \pi d^2$$
$$\lambda = N / Area$$



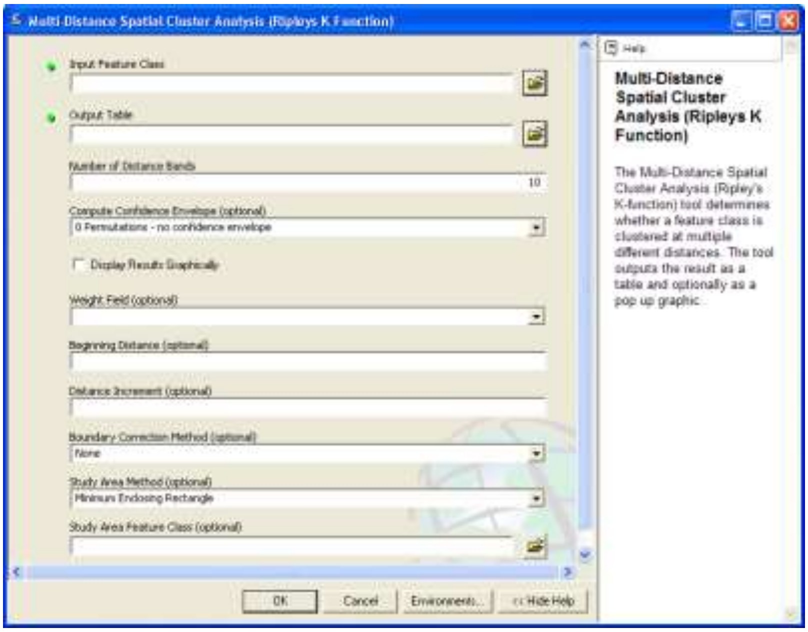
K-Function

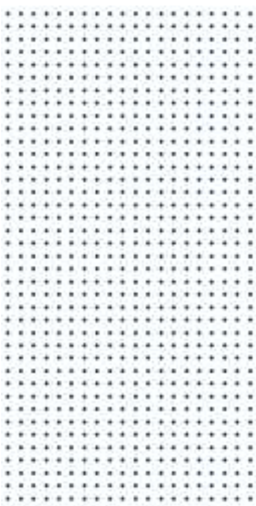
- Study Area = 100 sq unit, N = 10
- Average density (λ) = 10/100 = 0.1 (point/unit area)
- d = 1.67

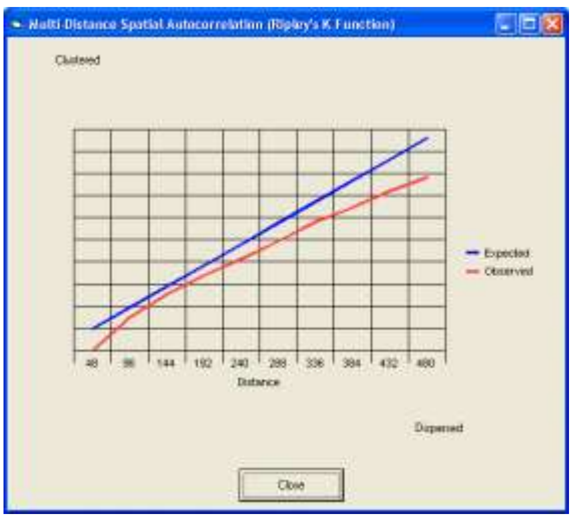


(4 + 3 + 4 + 3 + 2 + 2 + 2 + 2 + 1 + 1)/10 = 2.4 (average points per d)
K = 2.4/0.1 = 24
E(K) = 3.1416 x 1.67^2 = 8.76158775

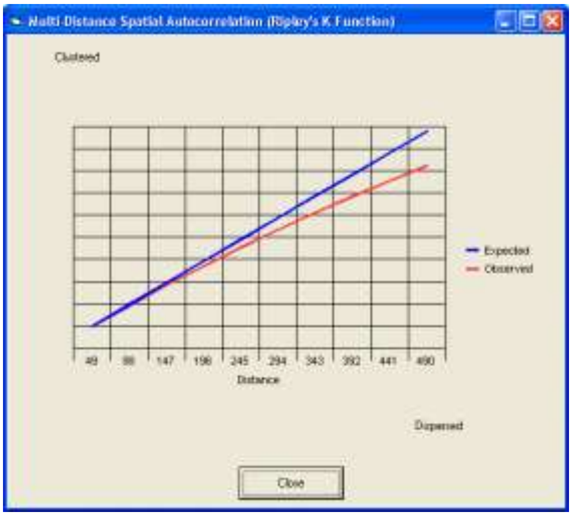
Multi-Distance Spatial Cluster Analysis: Ripley's k-function



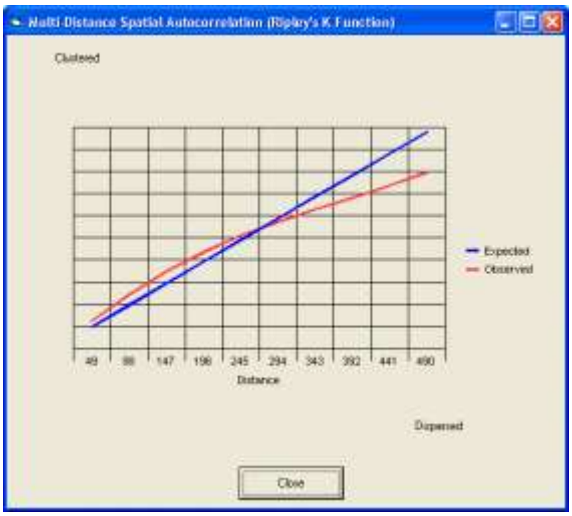




Grid points are 50 meters apart



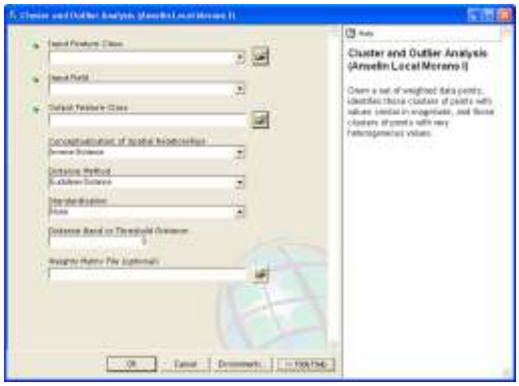
Multi-Distance Spatial Autocorrelation (Ripley's K Function)
Clustered
49
98
147
196
245
294
343
392
441
490
Distance
Expected
Observed
Dispersed
Close



Multi-Distance Spatial Autocorrelation (Ripley's K Function)
Clustered
49
98
147
196
245
294
343
392
441
490
Distance
Expected
Observed
Dispersed
Close

Mapping Clusters Toolset

Cluster & Outlier Analysis (Anselin Local Moran's I - LISA)

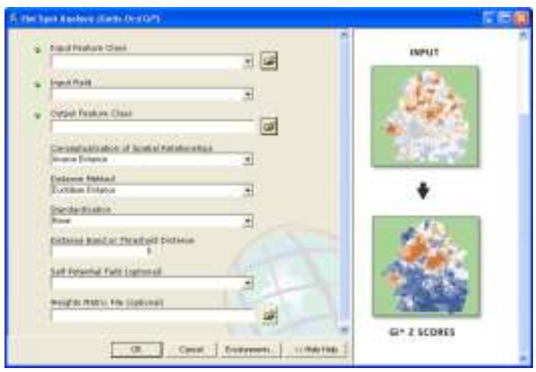


Outputs

- LMi
- LMz
  - High positive Z: surrounding values are similar (either high or low)
  - Very negative Z: surrounding values are dissimilar

Mapping Clusters Toolset

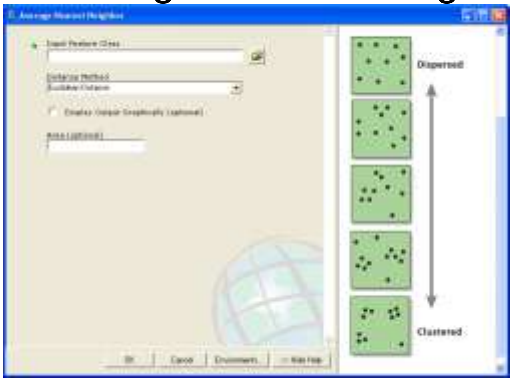
Hot Spot Analysis (Getis-Ord Gi*)



Output Gi (Z score)

- Higher positive (or lower negative) G_i -> stronger association of high or low values
- G_i near 0: no apparent concentration of hot or cool spots

Average Nearest Neighbor



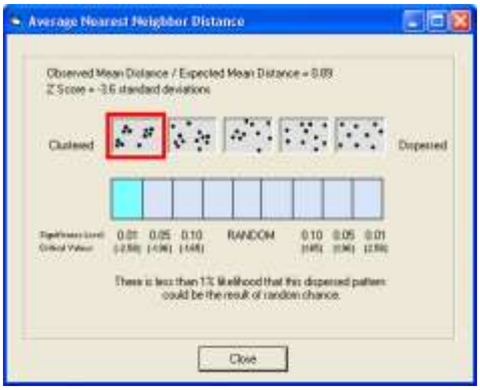
- Nearest Neighbor Ratio = Observed Mean Dist / Expected Mean Dist
  >> 1 (Dispersed) = 1 (Random) << 1 (Clustered)
- Expected Mean Dist is based on a hypothetical random distribution

$$E(d) = \frac{1}{2\sqrt{\lambda}}$$

$$\lambda = N / Area$$

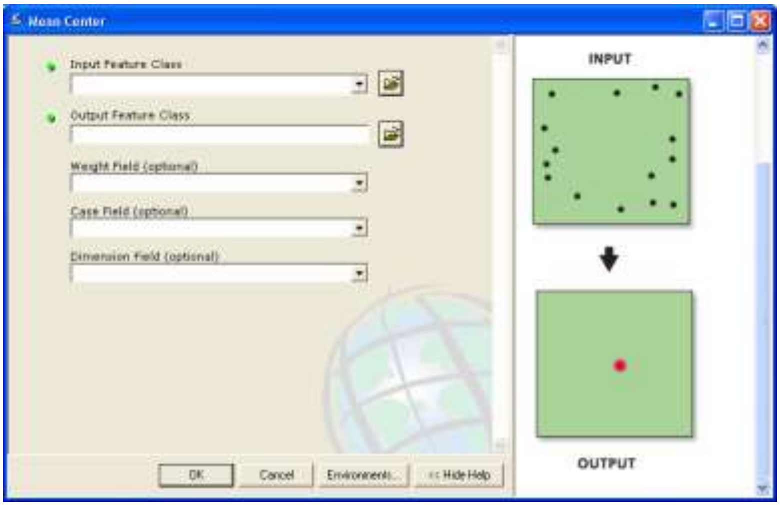
Output

- Nearest Neighbor Observed Mean Distance = 3756.9
- Expected Mean Distance = 4215.6
- Nearest Neighbor Ratio = 0.89
- Z Score = -3.55 Standard Deviations



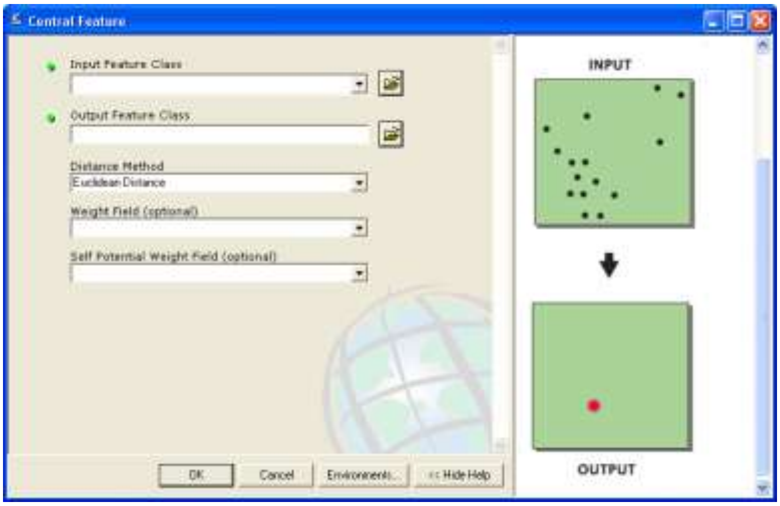
Measuring Geographic Distribution Toolset

Mean Center / Median Center



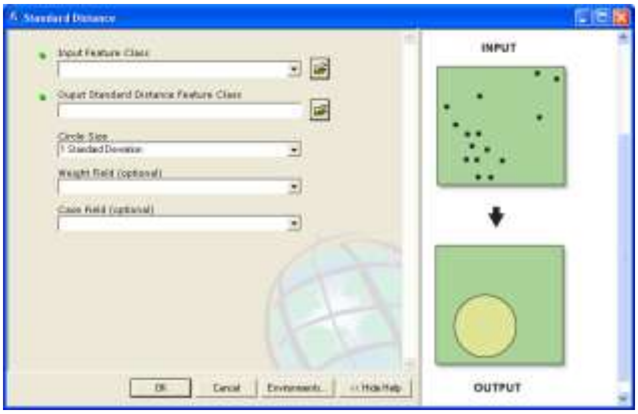
Measuring Geographic Distribution Toolset

Central Feature



Measuring Geographic Distribution Toolset

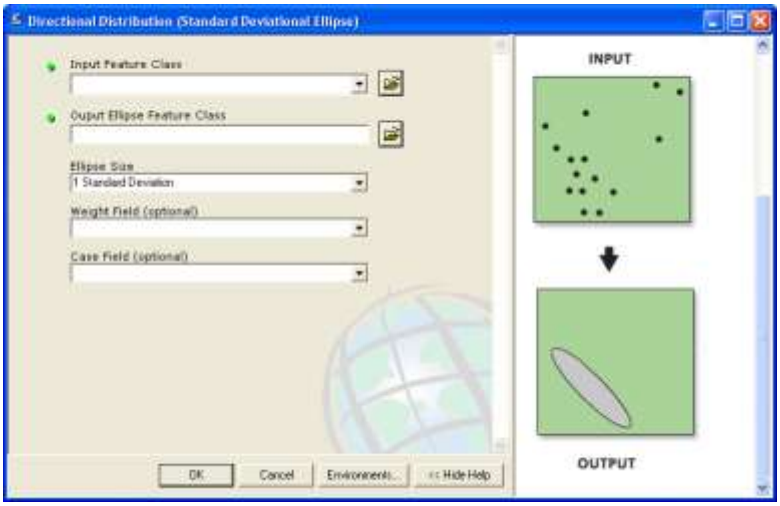
Standard Distance



- 1 stdv (68.26%)
- 2 stdv (95.46%)
- 3 stdv (99.73%)

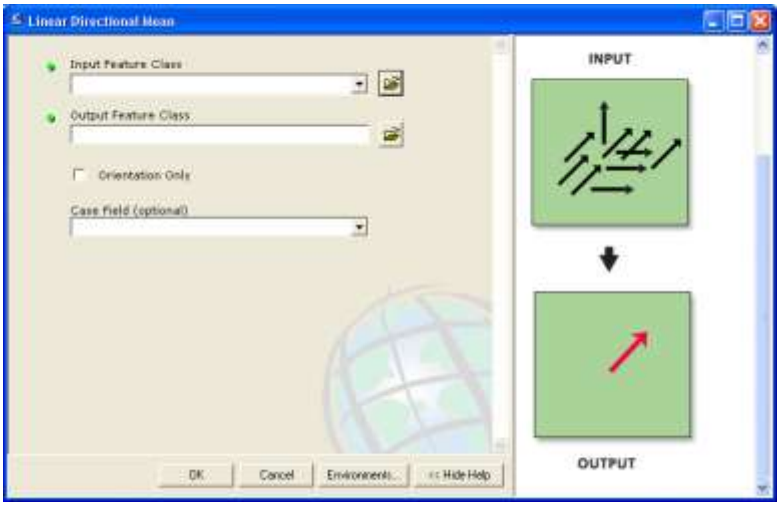
Measuring Geographic Distribution Toolset

Directional Distribution



Measuring Geographic Distribution Toolset

Linear Directional Mean



Modeling Spatial Relationships