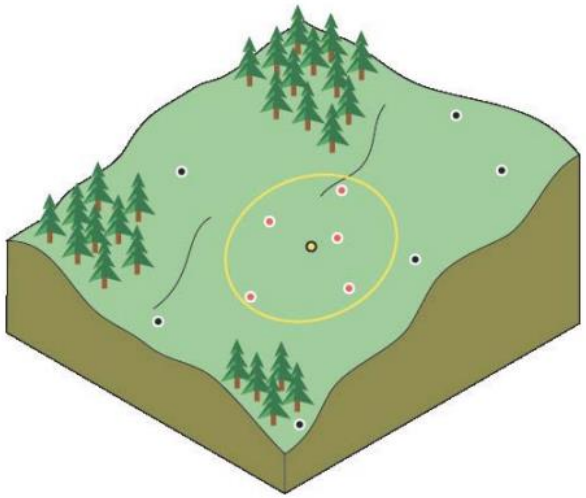
地理信息系统与遥感应用

空间地统计学

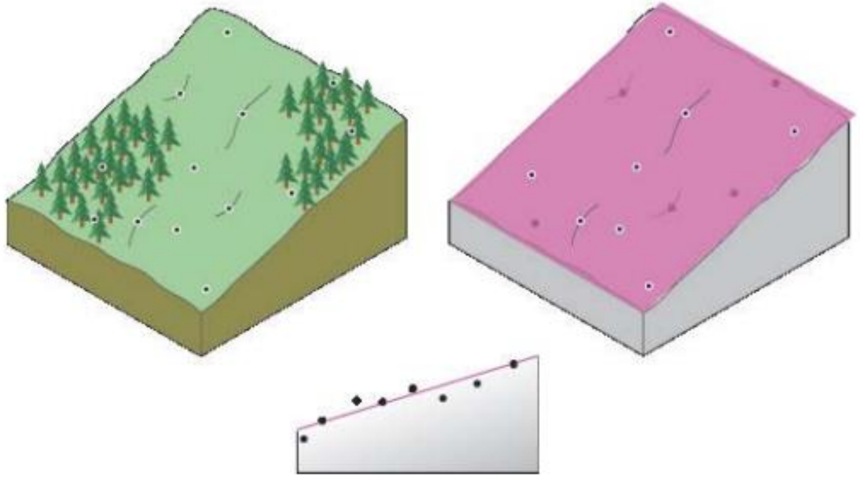
南方科技大学 · 环境科学与工程学院

田 勇

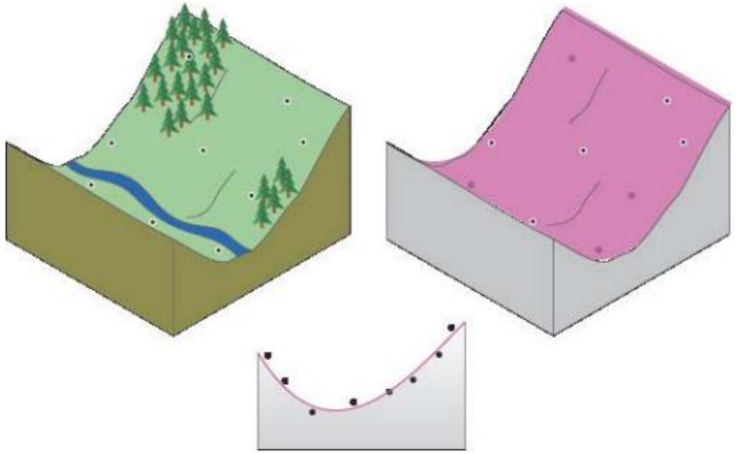
2018年11月19日

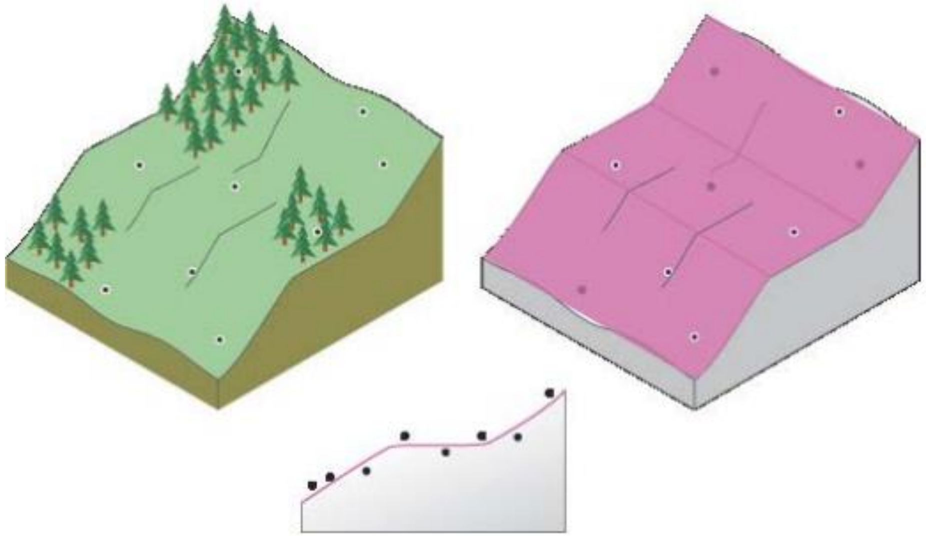


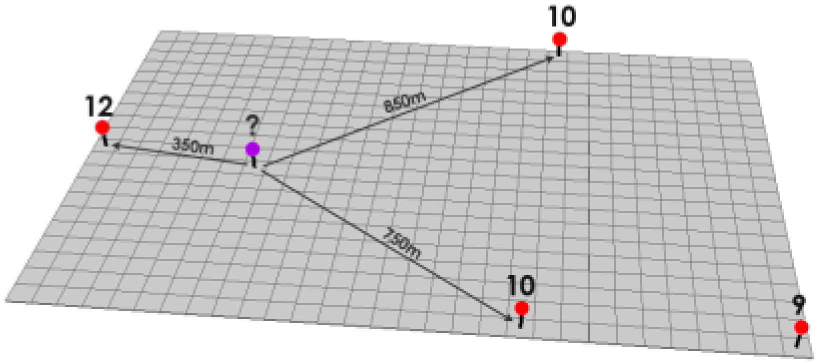
This is the basis for the Inverse Distance Weighting (IDW)

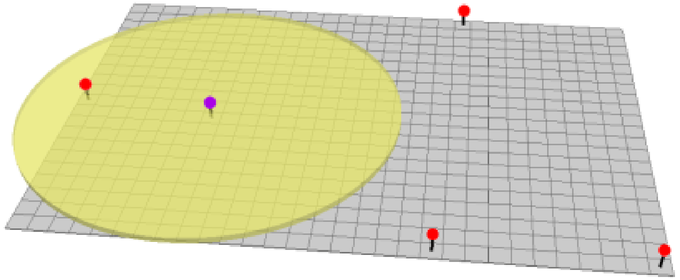


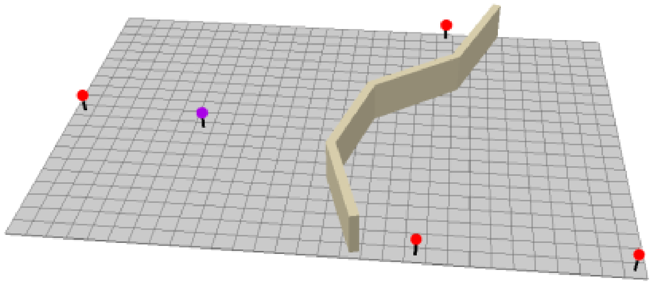
surface.

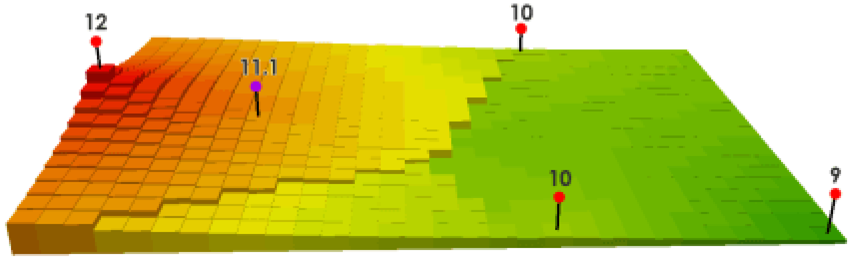


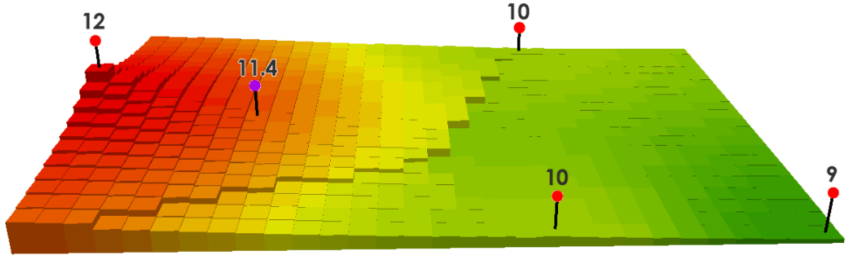












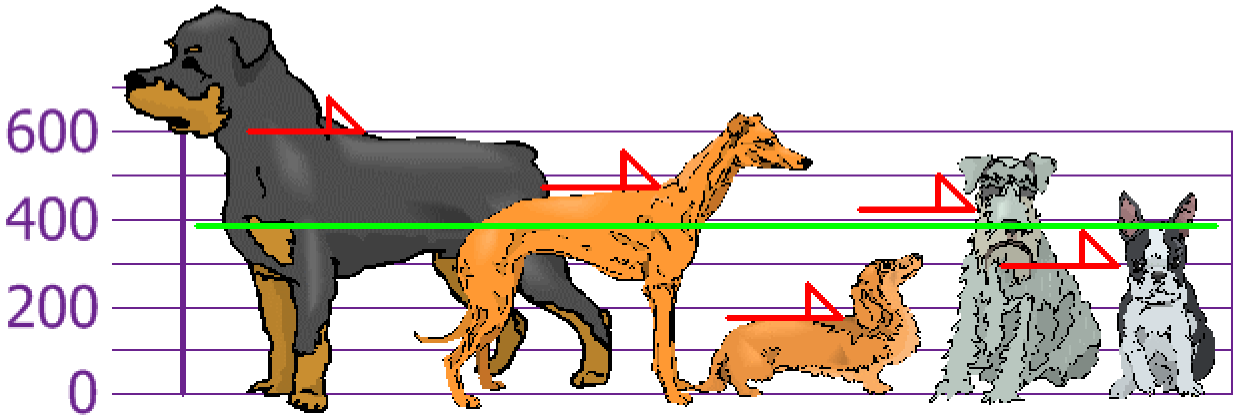
$$z_p = \frac{\sum_{i=1}^{n} \left(\frac{z_i}{d_i}\right)}{\sum_{i=1}^{n} \left(\frac{1}{d_i}\right)}$$

For a **power of 1**, that cell value is equal to:
((12/350) + (10/750) + (10/850)) / ((1/350) + (1/750) + (1/850)) = 11.1

For a **power of 2**, that cell value is equal to:
= ((12/350^2) + (10/750^2) + (10/850^2)) / ((1/350^2) + (1/750^2) + (1/850^2)) = 11.4

$$\sigma^2 = \frac{\sum (X - \mu)^2}{N}$$

Variance is a measurement of the spread between numbers in a data set. The variance measures how far each number in the set is from the mean.



$$\text{Mean} = \frac{600 + 470 + 170 + 430 + 300}{5} = \frac{1970}{5} = 394$$

$$\text{Variance: } \sigma^2 = \frac{206^2 + 76^2 + (-224)^2 + 36^2 + (-94)^2}{5}$$
$$= \frac{42,436 + 5,776 + 50,176 + 1,296 + 8,836}{5}$$
$$= \frac{108,520}{5} = 21,704$$

covariance is a measure of how much two random variables change together. If the greater values of one variable mainly correspond with the greater values of the other variable, and the same holds for the lesser values, i.e., the variables tend to show similar behavior, the covariance is positive.

$$\mathrm{cov}(X, Y) = \sum_{i=1}^{N} \frac{(x_i - \bar{x})(y_i - \bar{y})}{N}.$$

$$\mathrm{cov}(X, X) = \mathrm{Var}(X) \equiv \sigma^2(X)$$

$$\rho_{X,Y} = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y}$$

correlation coefficient is a measure of the linear dependence between two variables *X* and *Y*, giving a value between +1 and −1 inclusive, where 1 is total positive linear correlation, 0 is no linear correlation, and −1 is total negative linear correlation.

$$
\begin{aligned}
Var(x) &= E((x-E(x))^2) \\
&= E(x^2 - 2xE(x) + (E(x))^2) \\
&= E(x^2) - 2E(x)E(x) + (E(x))^2 \\
&= E(x^2) - 2(E(x))^2 + (E(x))^2 \\
&= E(x^2) - (E(x))^2
\end{aligned}
$$

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\,\mathrm{cov}(X,\ Y)$$

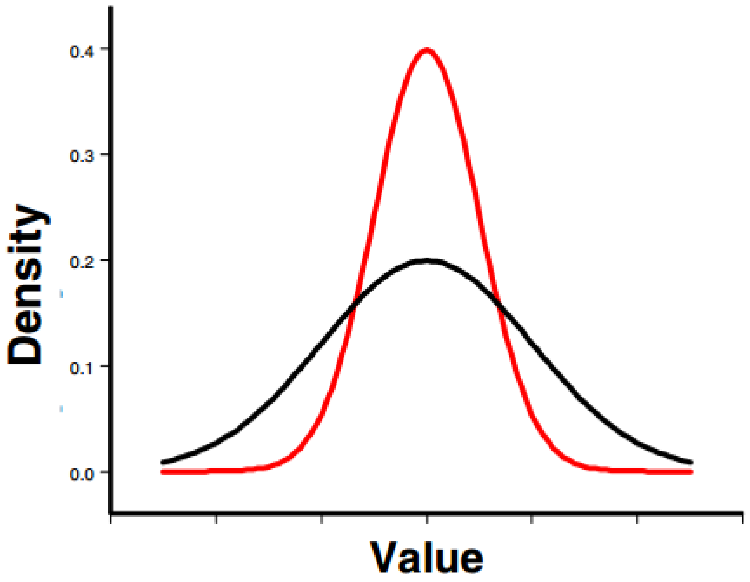
$$Var(ax+by) = a^2 Var(x) + b^2 Var(y) + 2Cov(x,y)$$

$$Cov(x,y) = E((x-E(x))(y-E(y)))$$

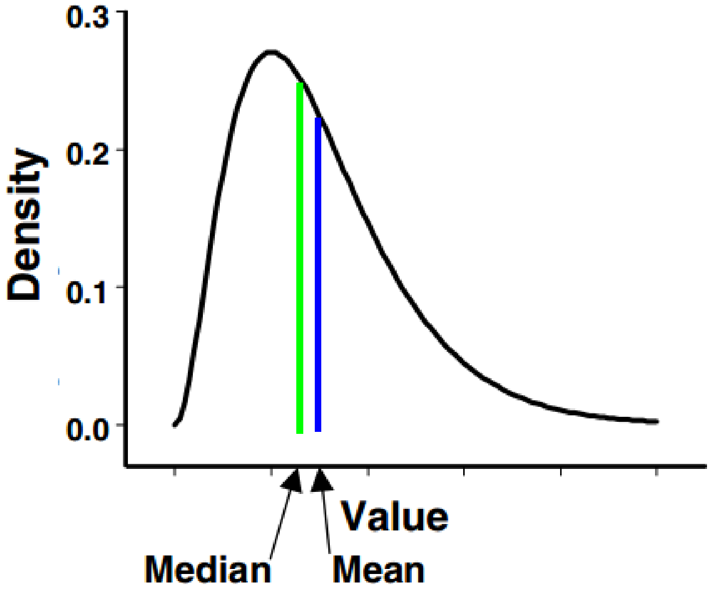
Cov(X, Y)=E(XY)-E(X)E(Y)

$$\mathrm{cov}(a\,X + b\,Y,\ Z) = a\,\mathrm{cov}(X,\ Z) + b\,\mathrm{cov}(Y,\ Z)$$

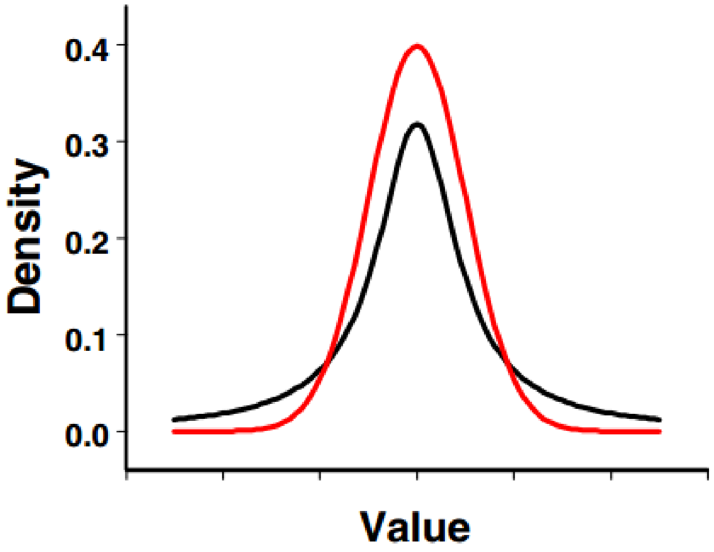
$$
\begin{aligned}
\mathrm{var}\left(\sum_{i=1}^{n} a_i X_i\right) &= E\left[\left(\sum_{i=1}^{n} a_i X_i\right)^2\right] - \left(E\left[\sum_{i=1}^{n} a_i X_i\right]\right)^2 \\
&= E\left[\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j X_i X_j\right] - \left(E\left[\sum_{i=1}^{n} a_i X_i\right]\right)^2 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j E[X_i X_j] - \left(\sum_{i=1}^{n} a_i E[X_i]\right)^2 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j E[X_i X_j] - \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j E[X_i]E[X_j] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \left(E[X_i X_j] - E[X_i]E[X_j]\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \,\mathrm{cov}(X_i, X_j)
\end{aligned}
$$



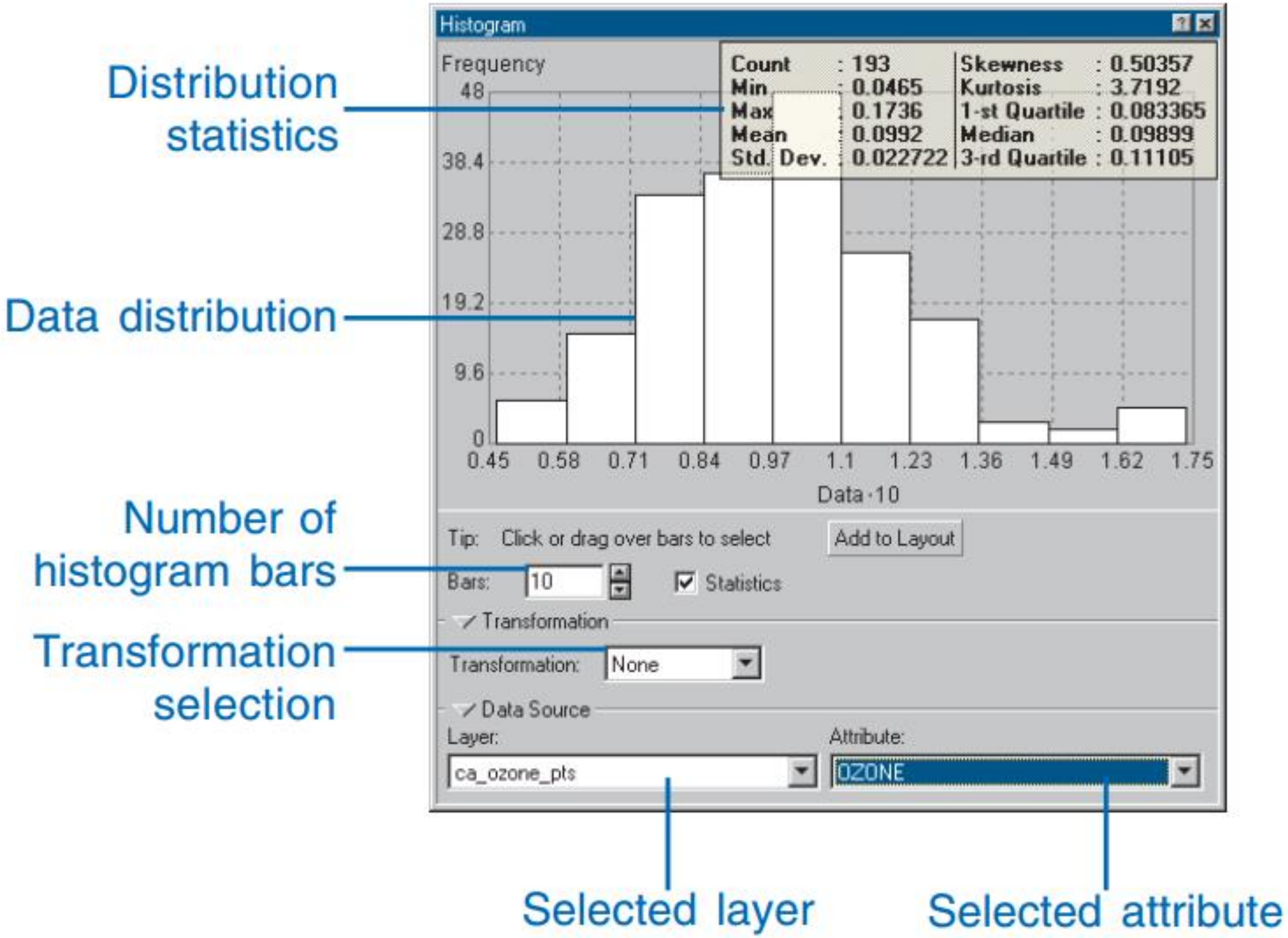
0.4
0.3
0.2
0.1
0.0
Density
Value

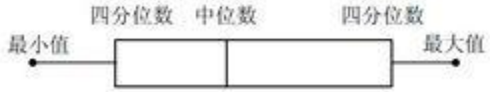


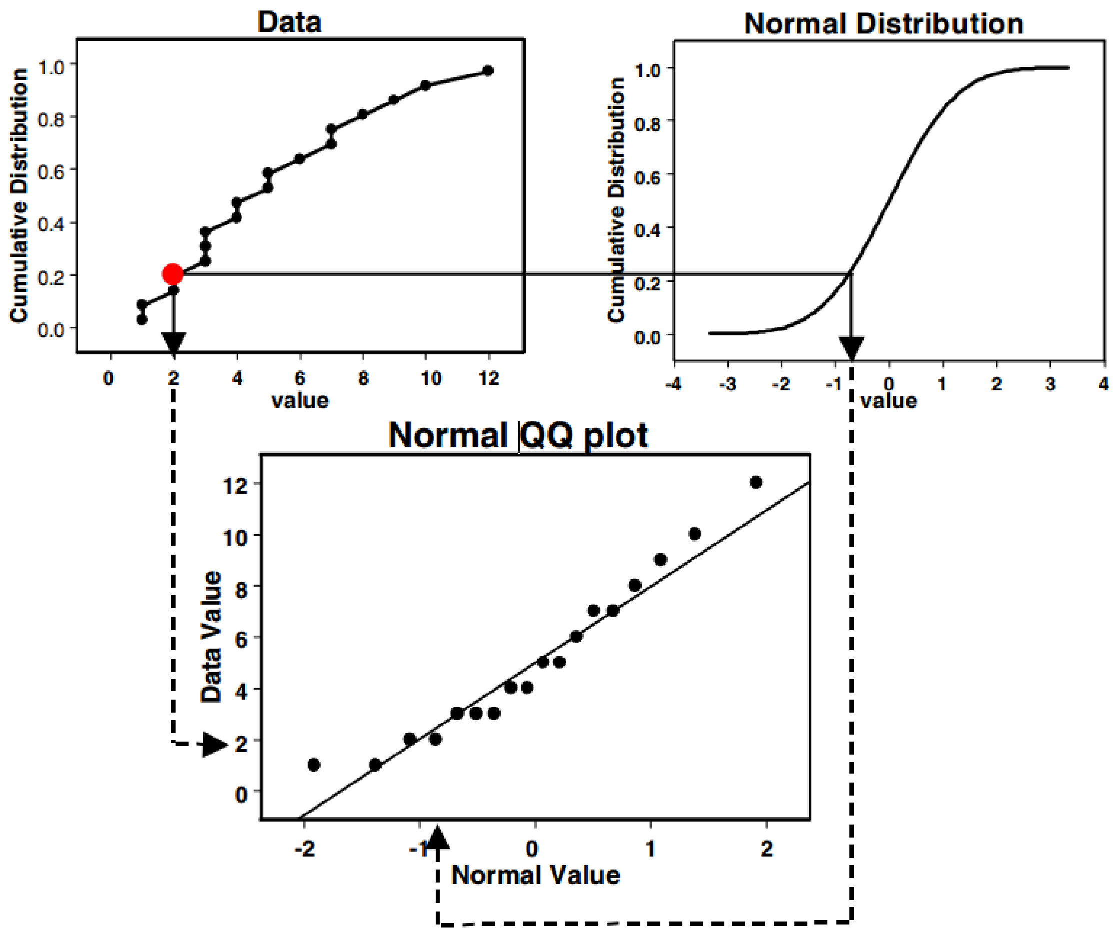
0.3
0.2
0.1
0.0
Density
Value
Median
Mean



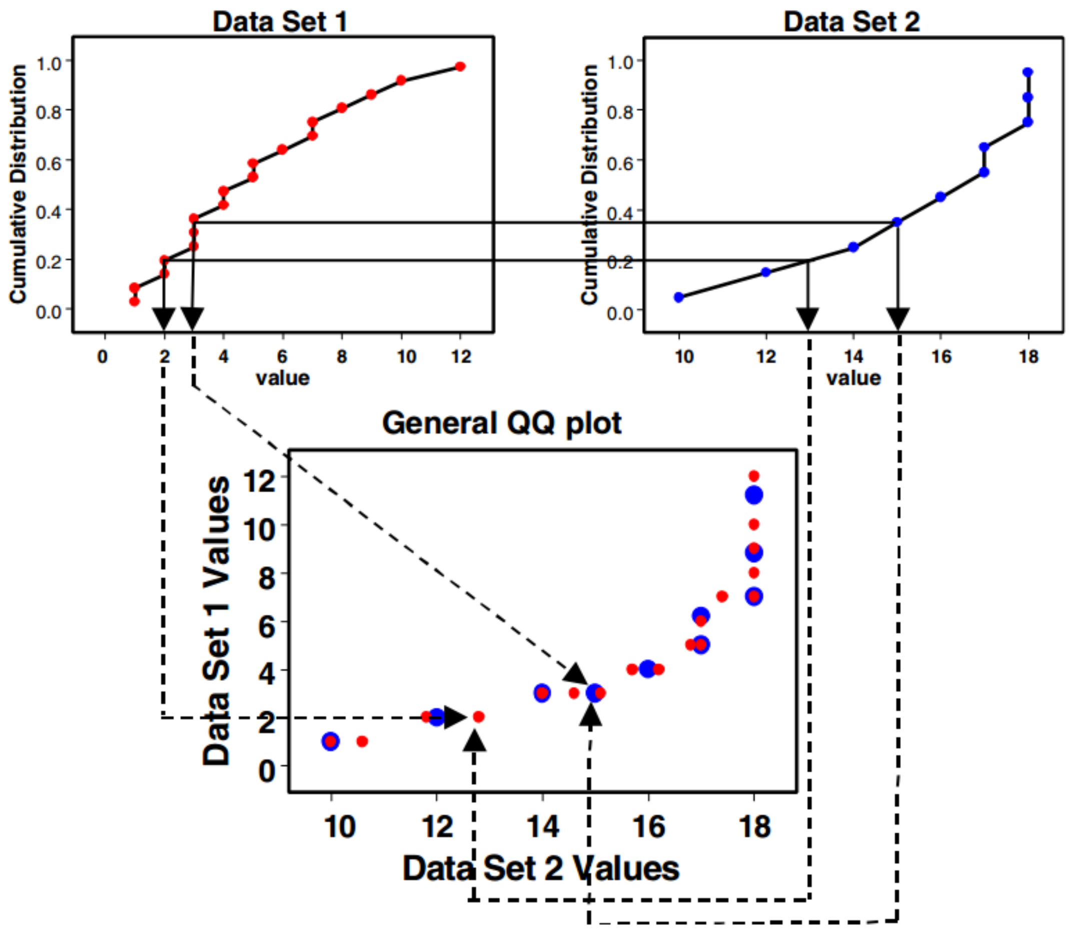
0.4
0.3
0.2
0.1
0.0
Density
Value


Distribution statistics
Data distribution
Number of histogram bars
Transformation selection
Selected layer
Selected attribute
Histogram
Frequency
48
38.4
28.8
19.2
9.6
0
Count : 193
Min : 0.0465
Max : 0.1736
Mean : 0.0992
Std. Dev. : 0.022722
Skewness : 0.50357
Kurtosis : 3.7192
1-st Quartile : 0.083365
Median : 0.09899
3-rd Quartile : 0.11105
0.45 0.58 0.71 0.84 0.97 1.1 1.23 1.36 1.49 1.62 1.75
Data · 10
Tip: Click or drag over bars to select
Add to Layout
Bars: 10
Statistics
Transformation
Transformation: None
Data Source
Layer:
ca_ozone_pts
Attribute:
OZONE


四分位数
中位数
四分位数
最小值
最大值



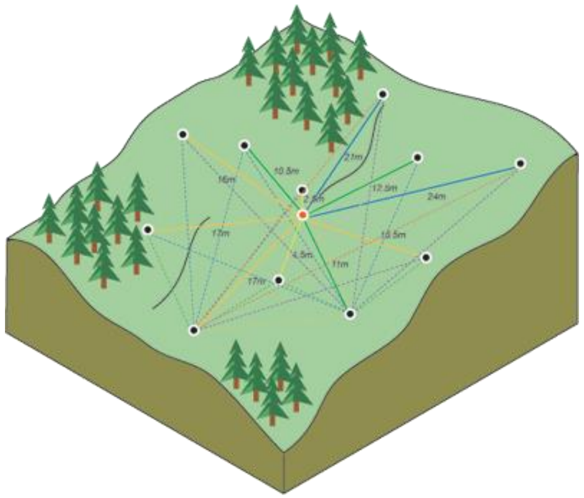
Data
Normal Distribution
Cumulative Distribution
value
Normal QQ plot
Data Value
Normal Value

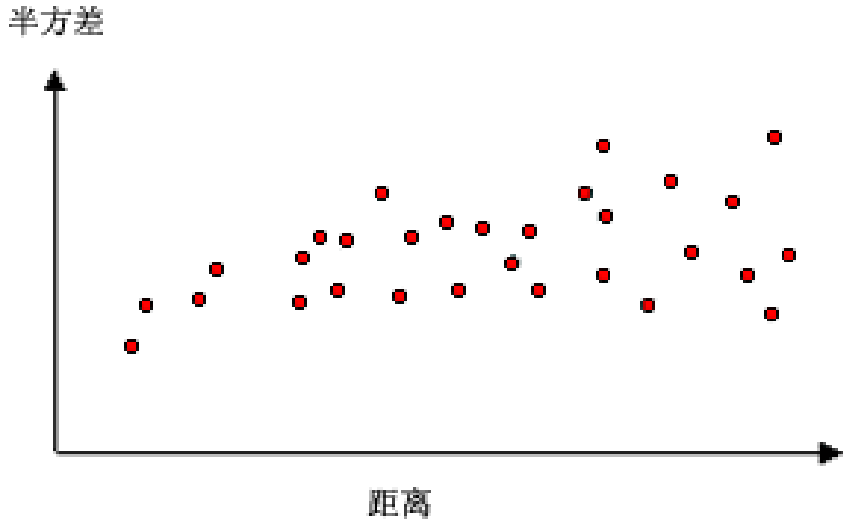


Data Set 1
Data Set 2
Cumulative Distribution
value
General QQ plot
Data Set 1 Values
Data Set 2 Values

For observations z_i, $i = 1, \dots, k$ at locations $x_1, \dots, x_k$ the empirical variogram $\hat{\gamma}(h)$ is defined as

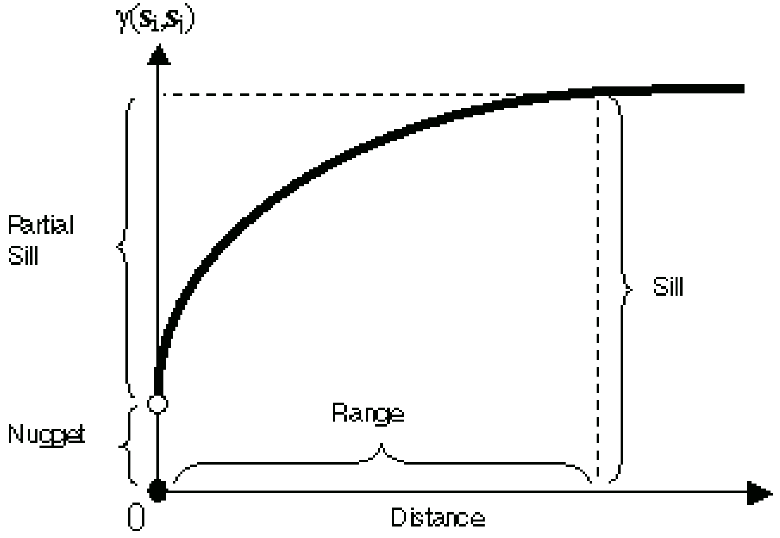
$$\hat{\gamma}(h) := \frac{1}{2|N(h)|} \sum_{(i,j) \in N(h)} |z_i - z_j|^2$$

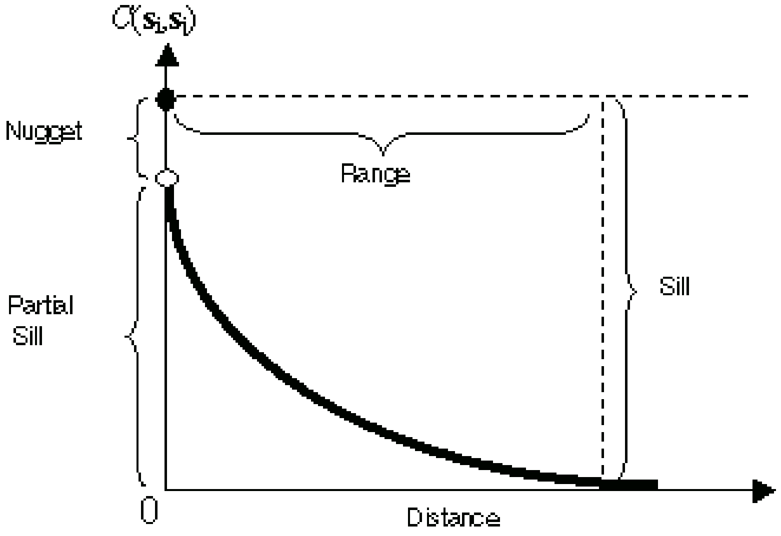




$$Y(\mathbf{s}_i, \mathbf{s}_j) = \text{sill} - C(\mathbf{s}_i, \mathbf{s}_j),$$

$$C(\mathbf{s}_i, \mathbf{s}_j) = \text{cov}(Z(\mathbf{s}_i), Z(\mathbf{s}_j)),$$



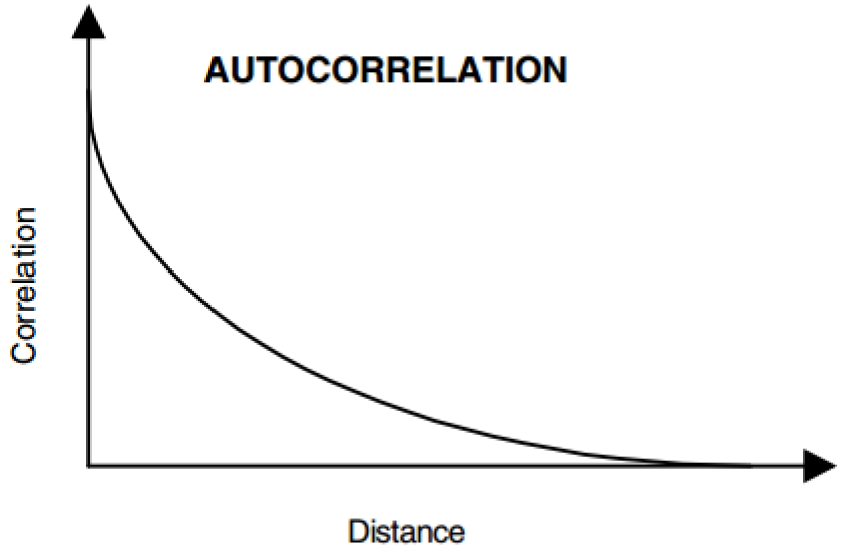


If two locations, s_i and s_j, are close to each other in terms of the distance measure of $d(s_i, s_j)$, then you expect them to be similar, so the difference in their values, $Z(\mathbf{s}_i) - Z(\mathbf{s}_j)$, will be small. As s_i and s_j get farther apart, they become less similar, so the difference in their values, $Z(\mathbf{s}_i) - Z(\mathbf{s}_j)$, will become larger.

Covariance is a scaled version of correlation. So, when two locations, $\mathbf{s}_i$ and $\mathbf{s}_j$, are close to each other, you expect them to be similar, and their covariance (a correlation) will be large. As $\mathbf{s}_i$ and $\mathbf{s}_j$ get farther apart, they become less similar, and their covariance becomes zero.

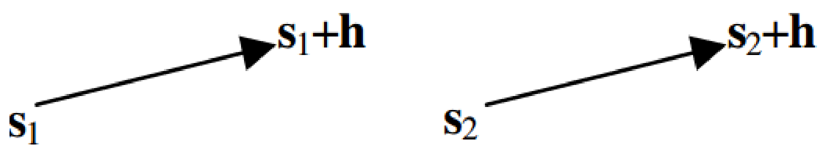
$$Y(\mathbf{s}_i, \mathbf{s}_j) = \text{sill} - C(\mathbf{s}_i, \mathbf{s}_j),$$

Understanding the kriging model



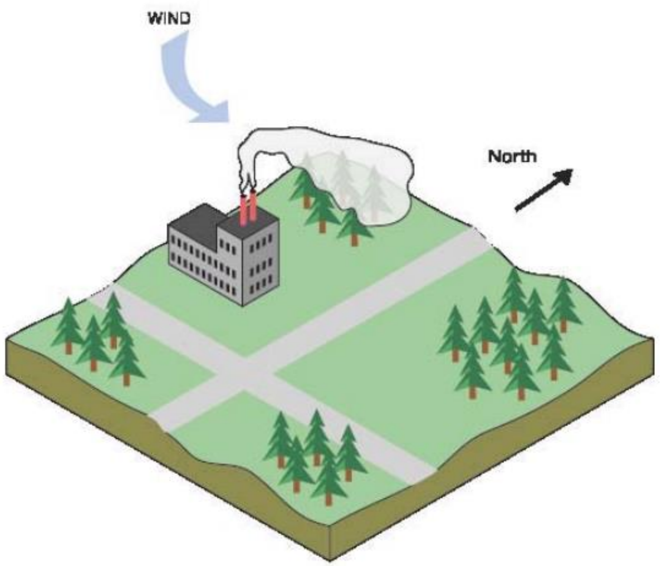
$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \varepsilon(\mathbf{s}),$$

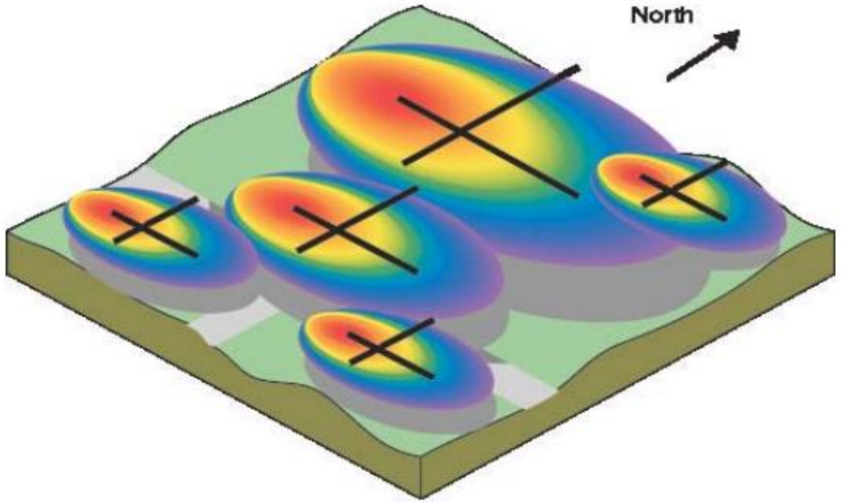
where $Z(\mathbf{s})$ is the variable of interest, decomposed into a deterministic trend $\mu(\mathbf{s})$, and random, autocorrelated errors form $\varepsilon(\mathbf{s})$. The symbol $\mathbf{s}$ simply indicates the location; think of it as containing the spatial x- (longitude) and y- (latitude) coordinates. Variations on this formula form the basis for all of the different types of kriging, and it is worth a little effort to become familiar with it.

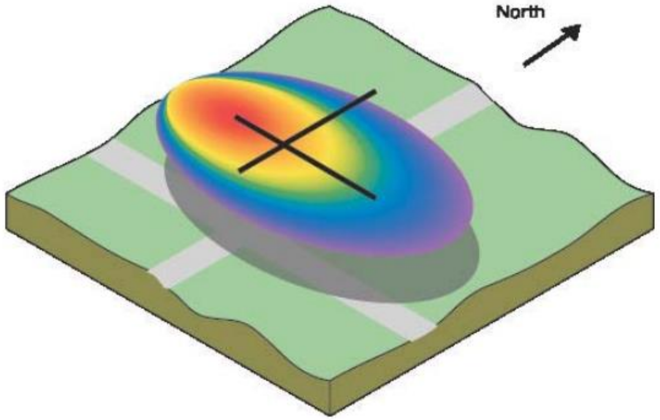


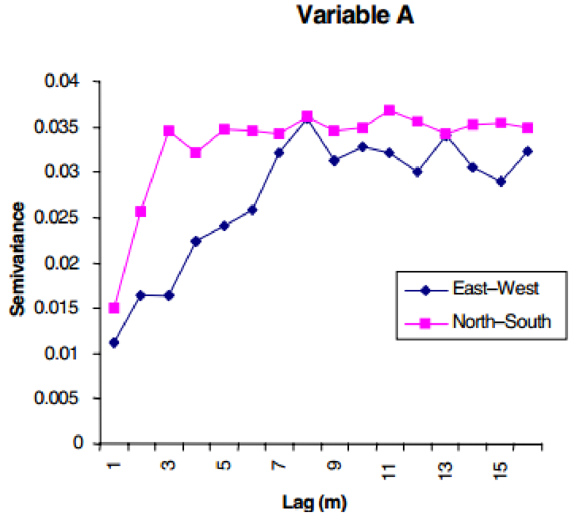
$$\mu(s) = \beta_0 + \beta_1 x + \beta_2 y + \beta_3 x^2 + \beta_4 y^2 + \beta_5 xy,$$

Directional influences—trend and anisotropy

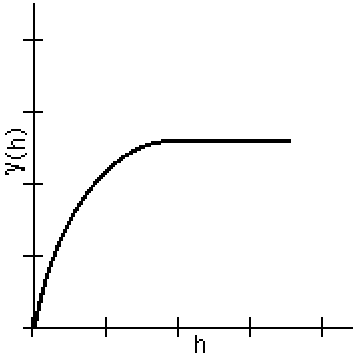








SPHERICAL（球形模型）

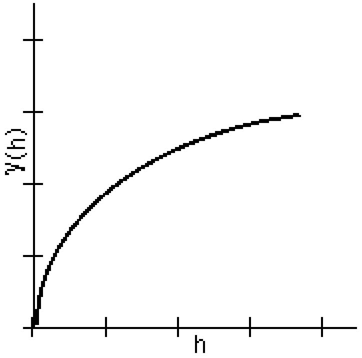


$$\gamma(\mathbf{h}) = c_0 + c\left(\frac{3h}{2a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right) \quad 0 < h \le a$$

$$\gamma(\mathbf{h}) = c_0 + c \quad h > a$$

$$\gamma(0) = 0$$

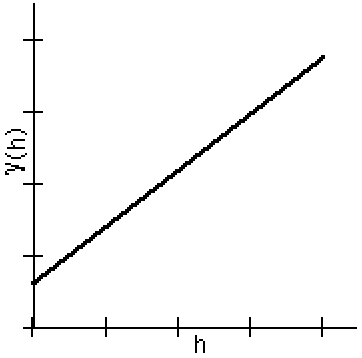
EXPONENTIAL（指数模型）



$$\gamma(\mathbf{h}) = c_0 + c\left(1 - exp\left(\frac{-h}{r}\right)\right) \quad h > 0$$

$$\gamma(0) = 0$$

LINEAR（线性模型）

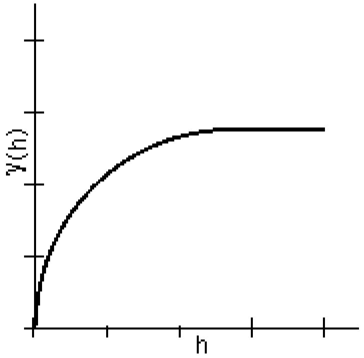


$$\gamma(\mathbf{h}) = c_0 + c\left(\frac{h}{a}\right) \quad 0 < h \le a$$

$$\gamma(\mathbf{h}) = c_0 + c \quad h > a$$

$$\gamma(0) = 0$$

CIRCULAR （圆形模型）



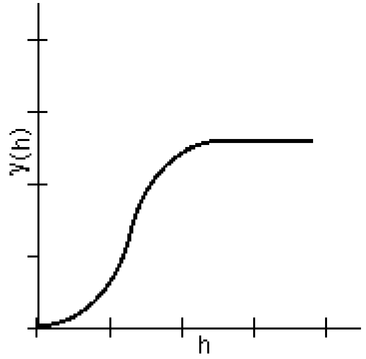
$$\gamma(\mathbf{h}) = c_0 + c\left(1 - \frac{2}{\pi}\cos^{-1}\left(\frac{h}{a}\right) + \sqrt{1 - \frac{h^2}{a^2}}\right)$$

$$0 < h \le a$$

$$\gamma(\mathbf{h}) = c_0 + c \quad h > a$$

$$\gamma(0) = 0$$

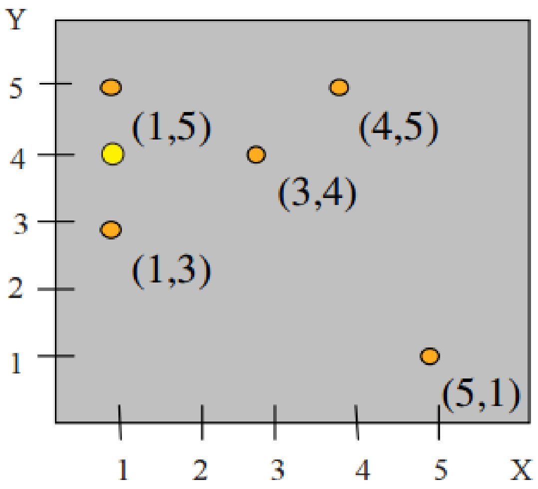
GAUSSIAN （高斯模型）



$$\gamma(\mathbf{h}) = c_0 + c\left(1 - exp\left(-\frac{h^2}{r^2}\right)\right) \quad h > 0$$

$$\gamma(0) = 0$$

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$



Values:

at (1,5) observe = 100

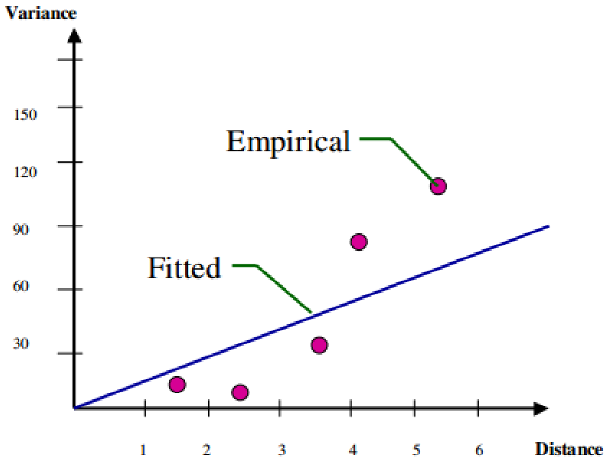
at (3,4) observe = 105

at (1,3) observe = 105

at (4,5) observe = 100

at (5,1) observe = 115

Locations	Distance Cal.	Distances	Difference^2	Semivariance
(1,5),(3,4)	sqrt[(1-3)^2 + (5-4)^2]	2.236	25	12.5
(1,5),(1,3)	sqrt[0^2 + 2^2]	2	25	12.5
(1,5),(4,5)	sqrt[3^2 + 0^2]	3	0	0
(1,5),(5,1)	sqrt[4^2 + 4^2]	5.657	225	112.5
(3,4),(1,3)	sqrt[2^2 + 1^2]	2.236	0	0
(3,4),(4,5)	sqrt[1^2 + 1^2]	1.414	25	12.5
(3,4),(5,1)	sqrt[2^2 + 3^2]	3.606	100	50
(1,3),(4,5)	sqrt[3^2 + 2^2]	3.606	25	12.5
(1,3),(5,1)	sqrt[4^2 + 2^2]	4.472	100	50
(4,5),(5,1)	sqrt[1^2 + 4^2]	4.123	225	112.5



	Binning the Empirical Semivariogram			
Lag Distance	Pairs Distance	Av. Distance	Semivariance	Average
1+-2	1.414, 2	1.707	12.5, 12.5	12.5
2+-3	2.236, 2.236, 3	2.491	12.5, 0, 0	4.167
3+-4	3.606, 3.606	3.606	50, 12.5	31.25
4+-5	4.472, 4.123	4.298	50, 112.5	81.25
5+	5.657	5.657	112.5	112.5

Semivariance = Slope * Distance

Semivariance = $13.5 * h$

Semivariance = $13.5 * 2.236 = 30.19$

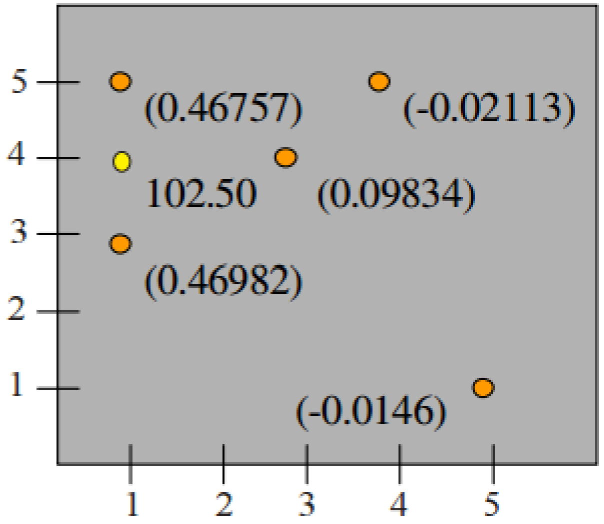
	(1, 5)	(3, 4)	(1, 3)	(4, 5)	(5, 1)	
	Γ Matrix (Gamma)					
(1, 5)	0	30.19	27.0	40.5	76.37	1
(3, 4)	30.19	0	30.19	19.09	48.67	1
(1, 3)	27.0	30.19	0	48.67	60.37	1
(4, 5)	40.5	19.09	48.67	0	55.66	1
(5, 1)	76.37	48.67	60.37	55.66	0	1
	1	1	1	1	1	0

$\lambda = \Gamma^{-1} * \mathbf{g}$

Inverse of Γ Matrix (Gamma)					
-0.02575	0.00704	0.0151	0.00664	-0.00303	0.3424
0.00704	-0.04584	0.01085	0.02275	0.0052	-0.22768
0.0151	0.01085	-0.02646	-0.00471	0.00522	0.17869
0.00664	0.02275	-0.00471	-0.02902	0.00433	0.28471
-0.00303	0.0052	0.00522	0.00433	-0.01173	0.42189
0.3424	-0.22768	0.17869	0.28471	0.42189	-41.701

Point	Distance	g Vector for (1,4)
(1,5)	1	13.5
(3,4)	2	27.0
(1,3)	1	13.5
(4,5)	3.162	42.69
(5,1)	5	67.5
		1

Weights	Values	Product	
0.46757	100	46.757	
0.09834	105	10.3257	
0.46982	105	49.3311	
-0.02113	100	-2.113	
-0.0146	115	-1.679	
-0.18281		102.6218	Kriging Predictor



Values:

(1,5) = 100

(3,4) = 105

(1,3) = 105

(4,5) = 100

(5,1) = 115