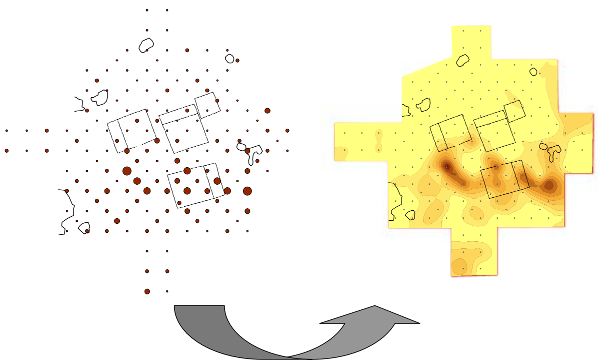
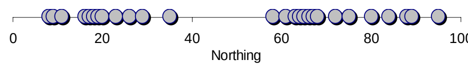


Interpolation

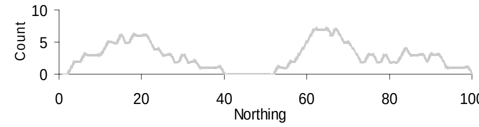
Recovery

The Point Process:



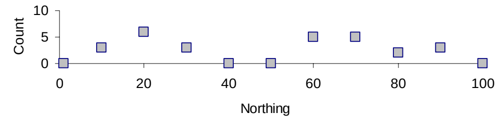
The Moving-Average Process:

(quadrat diameter=10)



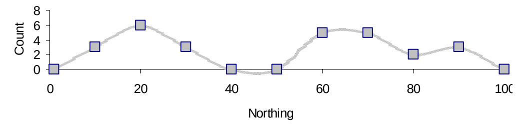
Sample the M-A Process

(quadrat spacing =10)



Analysis

Estimate the M-A Process from the sample:



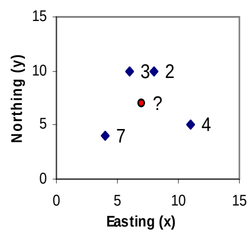
Interpolation

Many methods…

1. Inverse distance weighting (IDW)

2. Kriging

3. Others
   - -polynomial regression (local, global)
   - -TINs (triangulated irregular networks),
   - -splines (radial basis functions)



1. and 2. both make estimates of value of the z variable at an unsampled point in (x,y) space, as a *weighted average* of the values at nearby points, where z values are known.

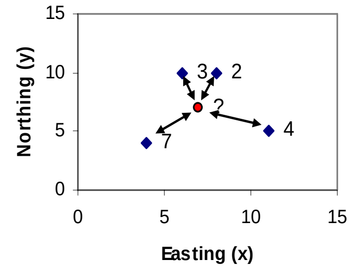
So….

$$\hat{z}_j = \sum_{i=1}^{n} w_i z_i$$

IDW

$$\hat{z}_j = \sum_{i=1}^{n} w_i z_i$$

$$w_i = \frac{1}{d_{ij}^{\,p}}$$

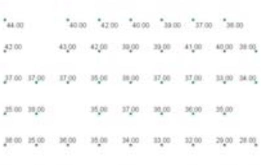


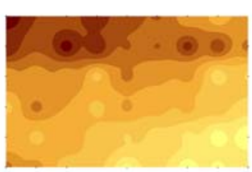
Point	x	y	z	distance	1/d	1/d * z	
1	11	5	4	4.47	0.22	0.89	
2	6	10	3	3.16	0.32	0.95	
3	8	10	2	3.16	0.32	0.63	
4	4	4	7	4.24	0.24	1.65	
Sum					1.09	4.13	3.78
5	7	7	?				

Pesky Questions about IDW:

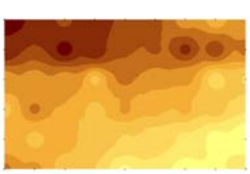
-what value for *p?*

-maximum value for *d*, beyond which d=0?

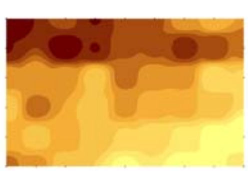




p=2, max(d)=6



p=2, max(d)=2.3



p=4, max(d)=6

More Pesky Question about IDW:

-Why base weights on Inverse Distance at all?

-Why not make try to make the weights optimal in a statistical sense,

- ***they minimize sum of squared prediction errors***
- ***-they are unbiased***

Kriging

-An interpolation method in which the weights depend on the spatial autocorrelation structure of the data, as reflected in the variogram, AND that produces estimates of Z that are designed to minimize the sum of squared prediction errors.

The Kriging Approach :

Predict z at point $\mathbf{s_0}$ as a weighted average of known values at points $\mathbf{s_i}$:

$$\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^{N} \lambda_i Z(\mathbf{s}_i)$$

Find weights (λ_i) such that

$$\left(Z(\mathbf{s}_0) - \sum_{i=1}^{N} \lambda_i Z(\mathbf{s}_i) \right)^2$$

To do this we need a model: Regionalized Variable Theory

$$Z(\mathbf{s}) = m(\mathbf{s}) + e'(\mathbf{s}) + e''$$

Z(**s**): the value of the RV at place **s**
m(**s**): the deterministic component or trend (i.e. the mean or expected value at **s**)
e'(**s**): the locally spatially-dependent component
e'': spatially *independent* noise

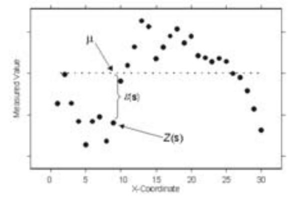
Ordinary Kriging: Assumptions

1. m(**s**) is constant across the study area – no trend or "drift" :
E[Z(**s**) – Z(**s + h**)] = 0.

2. variance of differences at different places is a function of distance (and direction) only, not absolute location:
E[{Z(**s**) – Z(**s + h**)}^{2}] = E [{e'(**s**) - e'(**s** +**h**)}^{2}]
=2γ (**h**) = semivariance

Question:

How limiting is the no-trend assumption?



If 1 and 2 are true, then the semivariance can be estimated from the sample data:

$$\hat{\gamma}_{\mathbf{h}} = \frac{1}{2n} \sum_{i=1}^{n} \{ z(\mathbf{s}_i) - z(\mathbf{s}_i + \mathbf{h}) \}^2$$

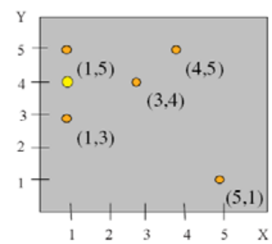
← This is the geostatistical version.

The i'th point in space

The point lag distance (and direction) h away from the i'th point in space.

$$\hat{\gamma}_{\mathbf{h}} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (z_i - z_j)^2}{2 \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}}$$

← We have seen this version of the formula before; *c.f.* Geary's C



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

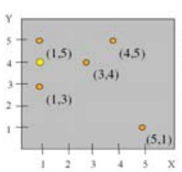
The Problem:

Predict the value at 1,4.

The Kriging solution:

1. Model spatial dependence using the (semi)variogram.
2. Use linear algebra to generate an ***Best, Linear, Unbiased Prediction (BLUP).***

1. The data



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

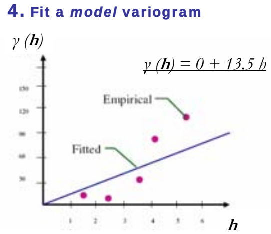
2. Compute all distance and squared difference pairs.

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

3. Combine over chosen lags.

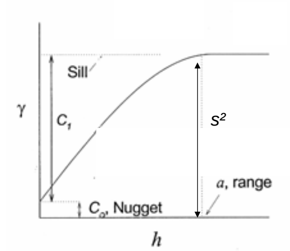
Lag Distance	Pairs Distance	Av. Distance	Semivariance	Average
1+-2	1.414, 2	1.707	12.5, 12.5	12.5
2+-3	2.236, 2.236, 3	2.491	12.5, 0, 0	4.167
3+-4	3.606, 3.606	3.606	50, 12.5	31.25
4+-5	4.472, 4.123	4.298	50, 112.5	81.25
5+	5.657	5.657	112.5	112.5

4. Fit a *model* variogram



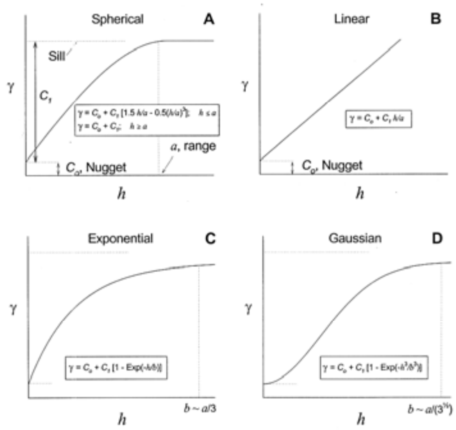
Variogram Lingo

- **Sill**: for larger values of h the variogram levels out, indicating that there no longer is any auto correlation between data points. *The sill should be equal to the* ***variance (s^2)*** *of the z values.*
- **Range**: is the value of h where the sill occurs (or 95% of the value of the sill). This is the distance beyond which pairs of values are no longer autocorrelated.
- **Nugget variance**: a non-zero value for γ when h = 0. Produced by various sources of unexplained error (e.g. measurement error).



Common Model Variograms

Variogram Models:
Differently shaped curves, defined by different equations.



5. Use the Model variogram to get a matrix of semivariances for all points pairs, based on their (Euclidean) distances, using: $\gamma(h) = 0 + 13.5\,h$

		(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

→ Γ

6. Use the Model variogram to get a vector of semivariances for the unknown point the knowns, based on their (Euclidean) distances, using $\gamma(h) = 0 + 13.5\,h$

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

→ g

7. Solve the kriging equations for ***lambda***. In matrix algebra form:

$$\Gamma * \lambda = \mathbf{g}$$

5. Use the Model variogram to get a matrix of semivariances for all points pairs, based on their (Euclidean) distances.

		(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

$$\Gamma * \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \end{bmatrix} =$$

6. Use the Model variogram to get a vector of semivariances for the unknown point the knowns, based on their (Euclidean) distances.

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

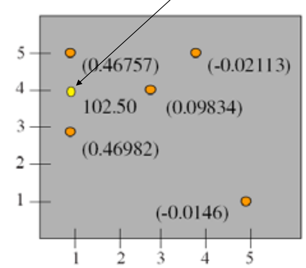
→ g

7. Solve the kriging equations for ***lambda***. In matrix algebra form:

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

7. Use the weights ***lambda*** *to make the prediction*:

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



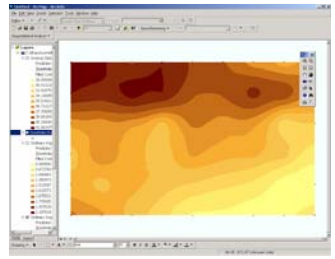
Values:
(1,5) = 100
(3,4) = 105
(1,3) = 105
(4,5) = 100
(5,1) = 115

7. Use the weights ***lambda*** *to estimate the prediction error*:

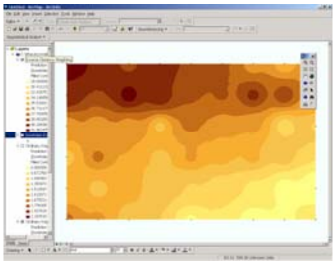
G Vector	Weights (λ)	g Vector Times Weights
13.5	0.46757	6.312195
27.0	0.09834	2.65518
13.5	0.46982	6.34257
42.69	-0.02113	-0.90204
67.5	-0.0146	-0.9855
1	-0.18281	-0.18281
	Kriging Variance	13.2396
	Kriging Std Error	3.6386

Kriging vs. IDW

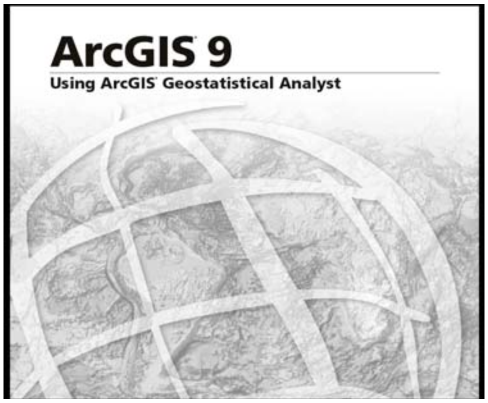
Cool – thanks to the elegant statistical model that tailors weights to the the data.



Uncool – thanks to the arbitrary choices for p and max(d).



Now on Toolkit:



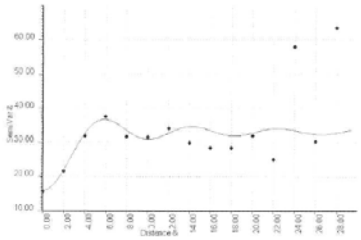


FIGURE 5. Neanderthal variogram (average of the two minimum dates by square, selected dataset). Vertical axis: semi-variance of the ¹⁴C dates; horizontal axis: geographical distance (x100 km).

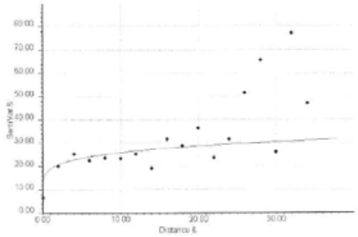
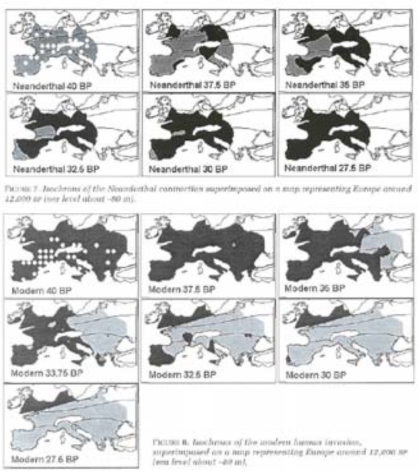


FIGURE 6. Modern human variogram (average of the two maximum dates by square, selected dataset). Vertical axis: semi-variance of the ¹⁴C dates; horizontal axis: geographical distance (x100 km).



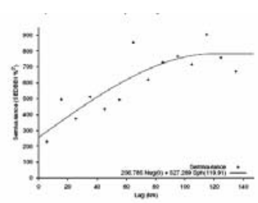


Fig. 3. Omnidirectional variogram for SEDBB I. Nug. is nugget, Sph. is spherical.

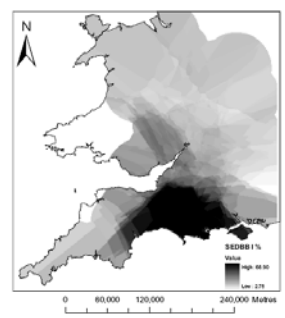


Fig. 4. Map of SEDBB I%, derived using OK, 1000 m cells.

C.D. Lloyd, P.M. Atkinson / Journal of Archaeological Science 31 (2004) 151–165

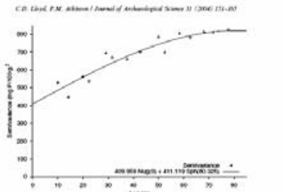


Fig. 7. Omnidirectional variogram of soil phosphate. Nug. is nugget, Sph. is spherical, DD is decimal degrees.

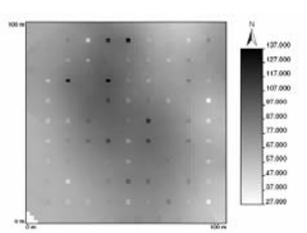


Fig. 9. Map of soil phosphate produced using punctual OK, 2 m cells. Scale is in mg P/100 g of soil.