Lecture 8: Variogram Estimation

Math 586

Standard case is very similar to covariance estimation, but with unequal spacing:

$$\text{pick} \quad \delta > \min_{\texttt{all } i,j} |\mathbf{x}_i - \mathbf{x}_j|$$

$$A_k = \{(i,j) \ : \ (k-1)\delta < |\mathbf{x}_i - \mathbf{x}_j| \leq k\delta\}$$

$$L_k := \#\{A_k\} \quad \text{number of points}$$

Isotropic case

$$\hat{\gamma}(\overline{d}_k) = \frac{1}{2L_k} \sum_{(i,j)\in A_k} [V(\mathbf{x}_i) - V(\mathbf{x}_j)]^2 \tag{1}$$

$\overline{d}_k$ = average distance separation of points in A_k.
Note: $\overline{V}$ is not needed.
δ = smoothing parameter (bin size). L_k should be reasonably large (say, ≥ 30).
Max distance used should be $\leq 0.25 \max_{i,j} |\mathbf{x}_i - \mathbf{x}_j|$

Non-isotropic case

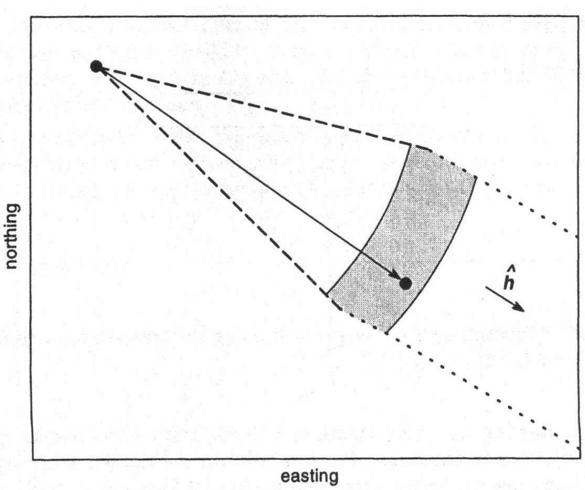


Figure 5.3. A distance class for semivariogram estimation along direction $\hat{h}$. All points in the shaded area belong to the same distance class relative to the point in the upper left corner. The dotted line denotes the lateral bounds and the segmented line the angular tolerance. The lag interval is the distance between the arc segments bounding the shaded area.

from Olea (1999)

Distance/angular class (see sector above):

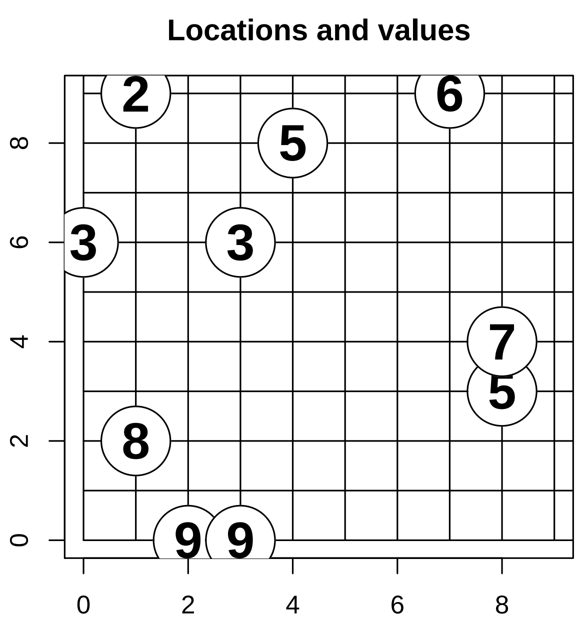
$A_{kr} = \{(i,j)\ :\ (k-1)\delta < |\mathbf{x}_i - \mathbf{x}_j| \leq k\delta\ \ \textbf{and}\ \text{angle of}\ \mathbf{h} = \mathbf{x}_j - \mathbf{x}_i \in (\phi_r - \alpha, \phi_r + \alpha)\}$

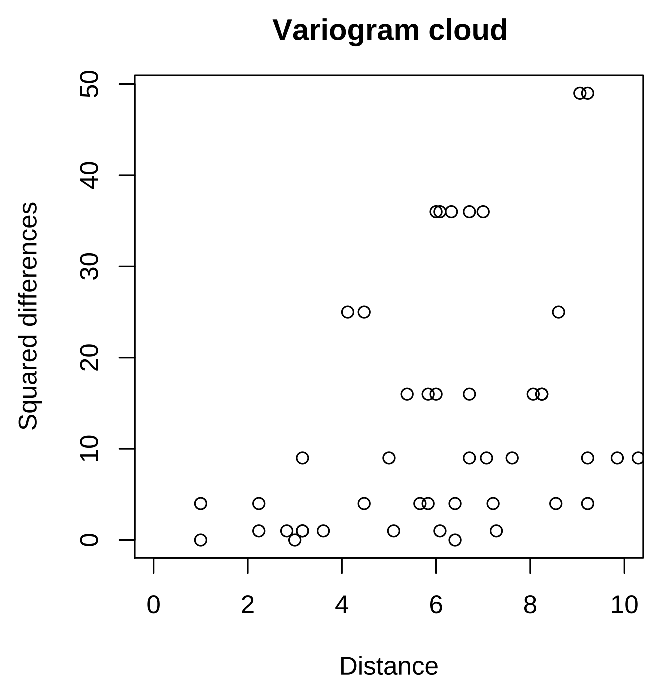
Play with δ, α, ϕ etc. to get smooth enough estimates. Other estimators are possible (Cressie & Hawkins, 1980).

"Empirical variograms" are thus obtained. Usually fit standard forms to ensure admissibility.

Examples:

1. A simple computational example.





Locations and values:

point #	1	2	3	4	5	6	7	8	9	10
X	4	2	0	8	7	1	3	1	3	8
Y	8	0	6	3	9	9	0	2	6	4
V	5	9	3	5	6	2	9	8	3	7

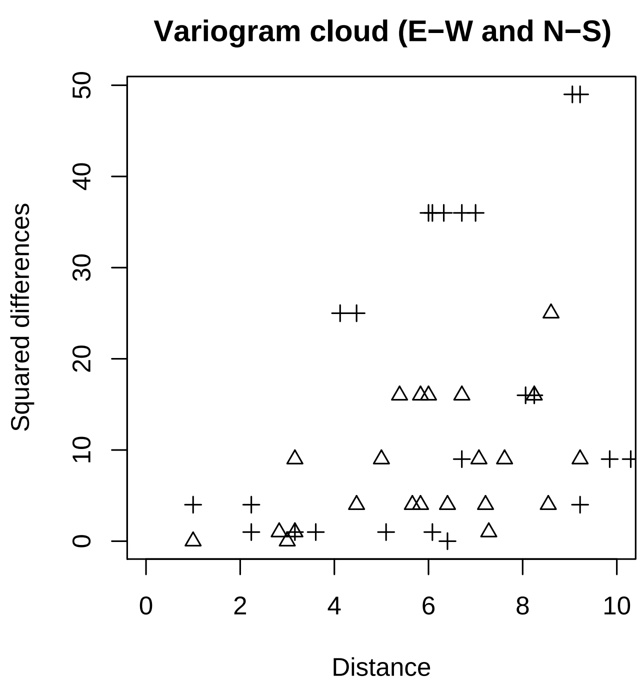
For the points $i = 1, 2, ..., 10$, compute distances between pairs and plot a "Variogram cloud", with distance on the x-axis and $(V(\mathbf{x}_i) - V(\mathbf{x}_j))^2$ on the y-axis.

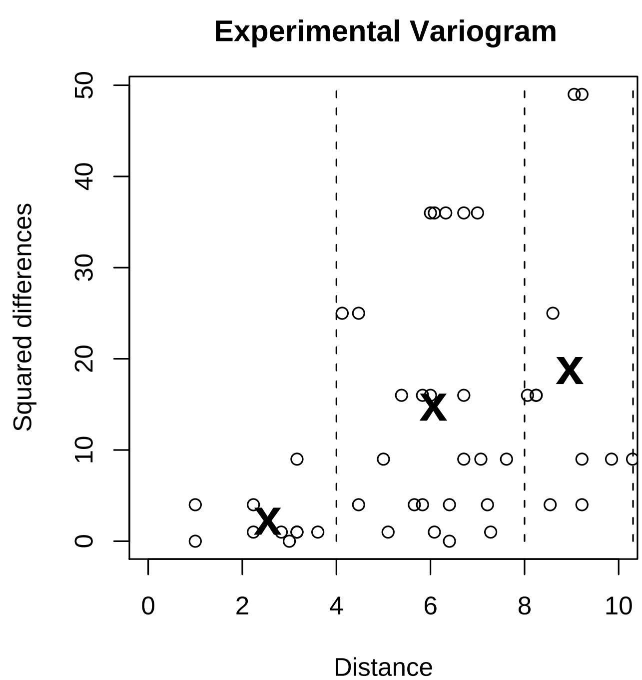
Distances (upper right) and squared differences (lower left)

	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	0	8.25	4.47	6.40	3.16	3.16	8.06	6.71	2.24	5.66
[2,]	16	0.00	6.32	6.71	10.30	9.06	1.00	2.24	6.08	7.21
[3,]	4	36.00	0.00	8.54	7.62	3.16	6.71	4.12	3.00	8.25
[4,]	0	16.00	4.00	0.00	6.08	9.22	5.83	7.07	5.83	1.00
[5,]	1	9.00	9.00	1.00	0.00	6.00	9.85	9.22	5.00	5.10
[6,]	9	49.00	1.00	9.00	16.00	0.00	9.22	7.00	3.61	8.60
[7,]	16	0.00	36.00	16.00	9.00	49.00	0.00	2.83	6.00	6.40
[8,]	9	1.00	25.00	9.00	4.00	36.00	1.00	0.00	4.47	7.28
[9,]	4	36.00	0.00	4.00	9.00	1.00	36.00	25.00	0.00	5.39
[10,]	4	4.00	16.00	4.00	1.00	25.00	4.00	1.00	16.00	0.00

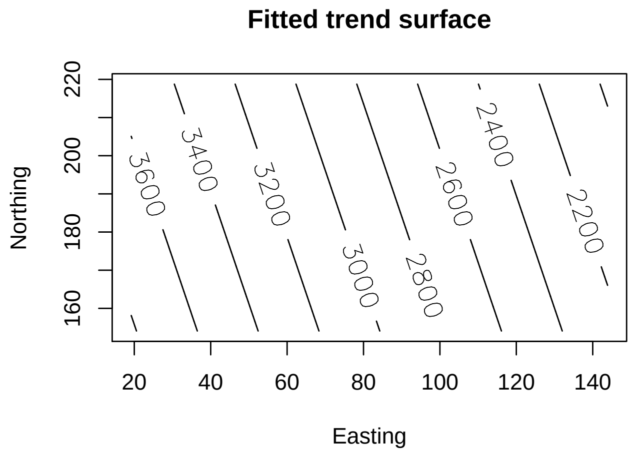
Also, we can mark the points on the variogram cloud according to the sector where the vector $\mathbf{h}_{ij} = \mathbf{x}_i - \mathbf{x}_j$ belongs: in the plot below left, NS-sector pairs are marked by + and EW-sector pairs are marked by triangles (sector width corresponds to the value of $\alpha = 45°$).

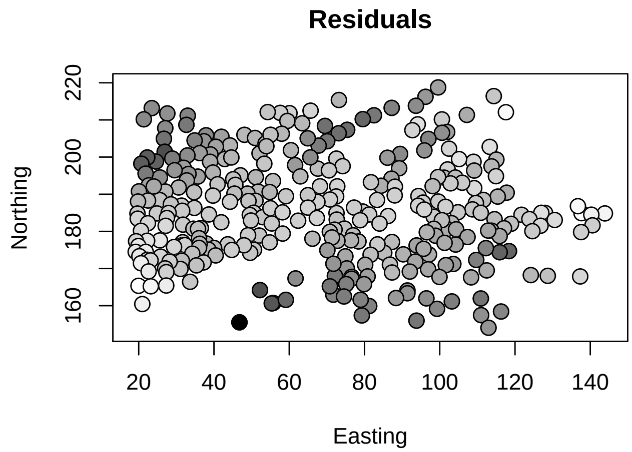
The points of the empirical variogram computed using (1) can be thought of as averages of x- and y-values of the variogram cloud (below right, need to divide by 2). Non-isotropic variograms are computed similarly.



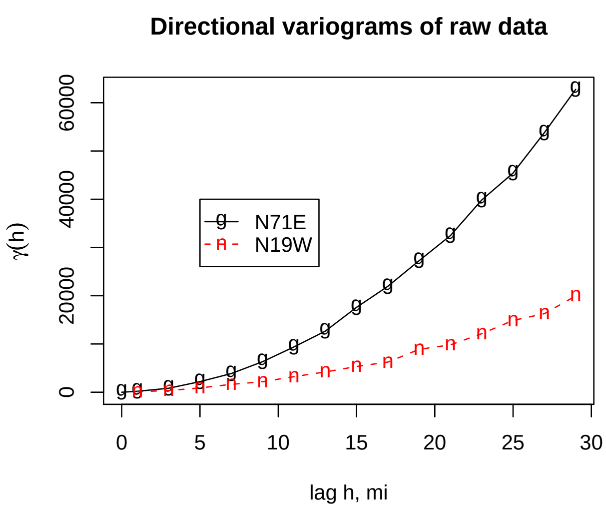


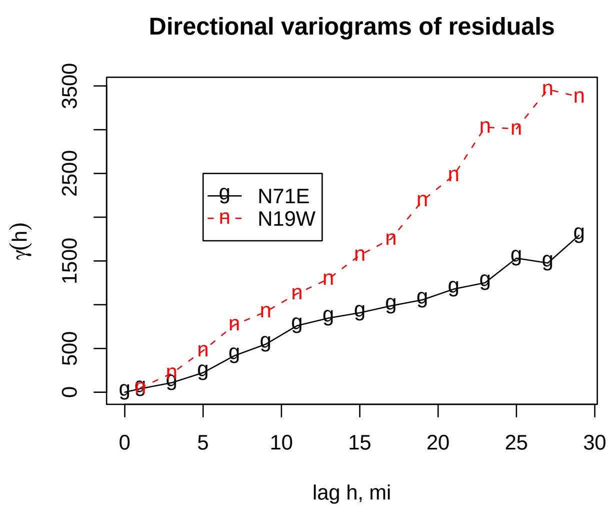
2. North Plains (Kansas) aquifer data (Olea), Water Table Elevation (ft)





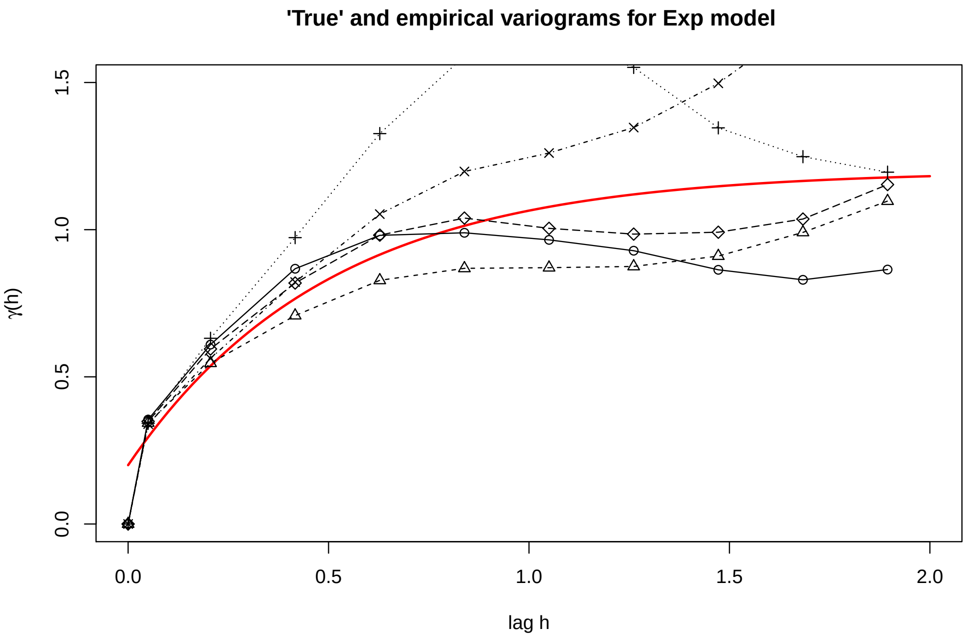
Linear trend $4714 - 12.6x - 4.26y$, spatially correlated residuals.
Compute variograms in the direction of the trend, that is, the angle $\phi = \arctan(4.259/12.562) \approx 19°$, and the direction normal to it (2 variogram zones)



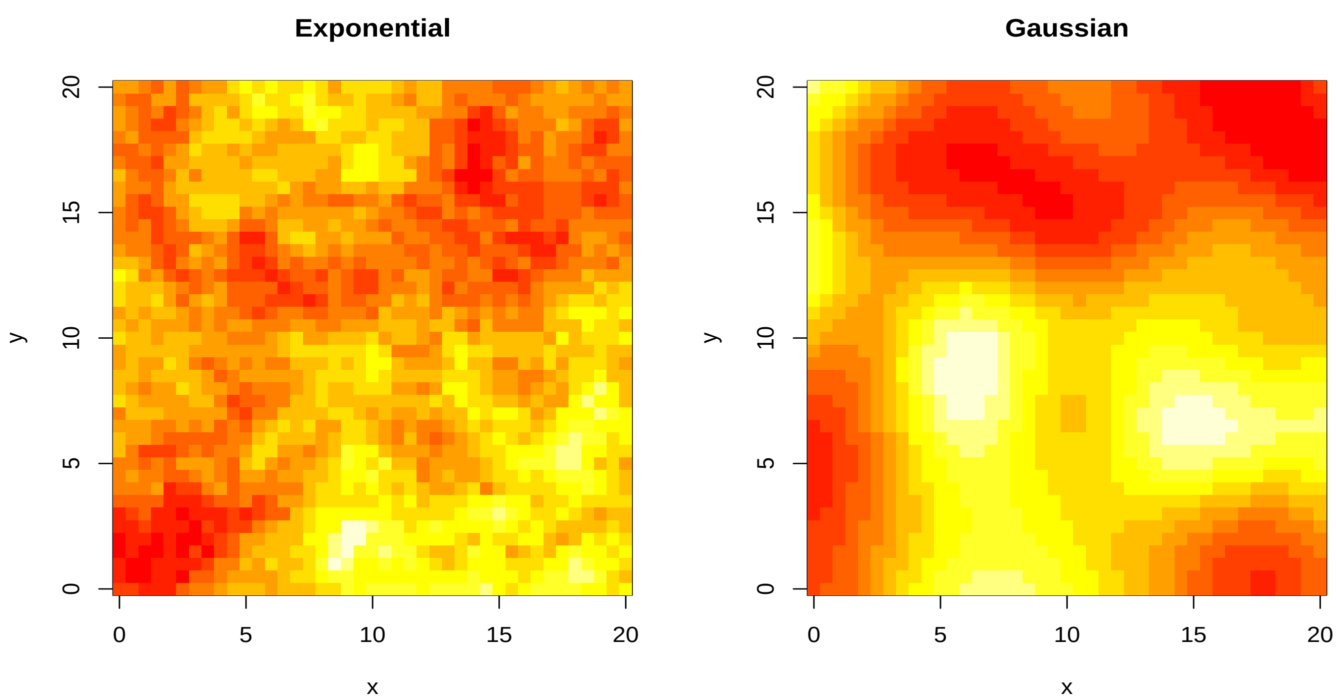


Raw data variogram shows quadratic shape $\Rightarrow$ indicative of a trend.
Both variograms look non-isotropic.

3. Simulation studies.
Example: exponential model with mean=0, variance=1, nugget=0.2, scale=0.5
Five simulated data sets (square grid of size 10x10 and spacing of 0.2):



4. Random fields simulated from different variogram models:



Sill = 1, scale = 5 in both. Notice the difference in smoothness.

Problems/things to watch for:

i. Effects of δ ("smoothing parameter"): can lead to oscillating # of points in each class. Variogram cloud plots, $[V(\mathbf{x}+\mathbf{h})-V(\mathbf{x})]^2$ vs. k can be useful. Vary δ and look at # in classes.

ii. outliers (very large/small observations) can greatly affect the variogram estimates. Exclude large extremes, re-estimate.

iii. Bi-modal, multiple population case - look at data and histograms

iv. trend removal may affect estimates

v. errors: missing values replaced by -1, small numbers replacing 0 in log transforms etc.

Variogram model fitting

Olea 1999 *Geostatistics...*:

- maximum likelihood (Kitanidis, p. 98) good but expensive, relies on Gaussian assumptions

- weighted least squares fit to empirical variograms.

- nested models:

$$\gamma(\mathbf{h}) = \sum_{i=1}^{m} w_i \gamma_i(\mathbf{h})$$

  – valid variograms; sill $\sigma^2 = \sum w_i \sigma_i^2$. Temptation to overfit.

- AIC (Akaike information criterion) and BIC (Bayesian information criterion) to choose between models.

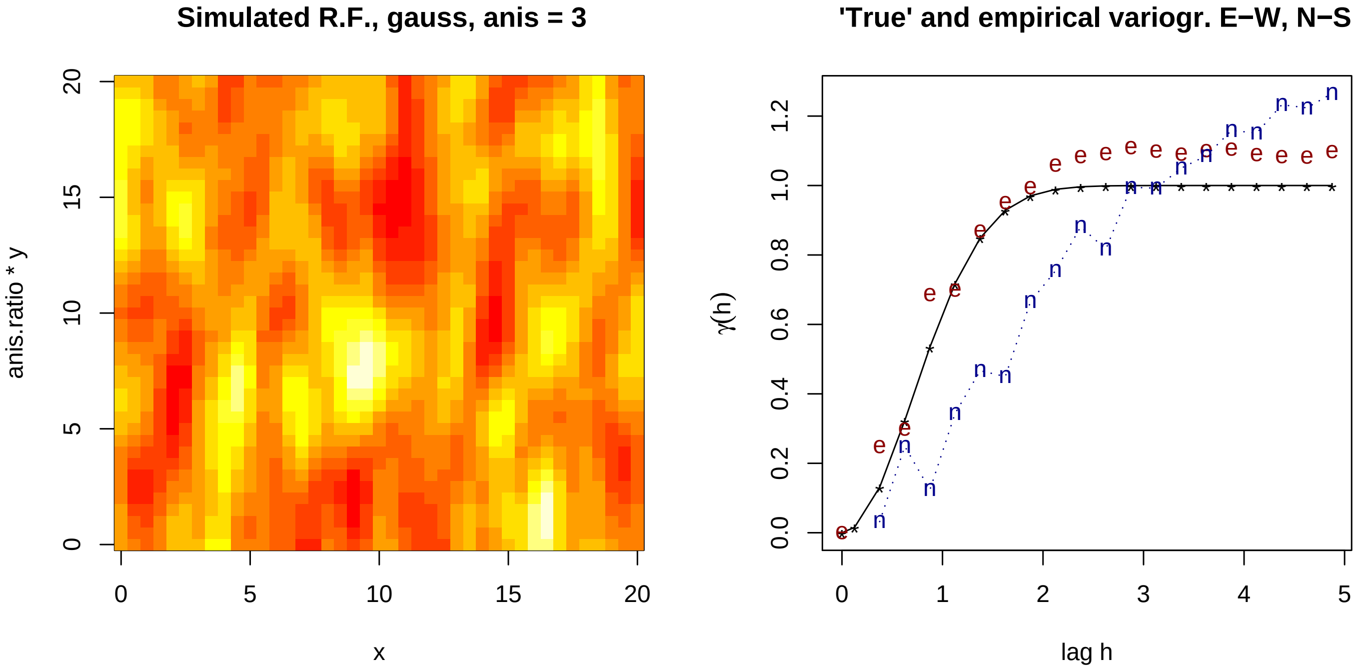
$$AIC = n\log(RSS/n) + 2p \qquad BIC = n\log(RSS/n) + p\log n,$$

  where RSS = residual sum of squares. Both add a penalty term for p= # of parameters in the variogram model. Occam razor: simple models are usually the best!

Geometric anisotropy

Similar to isotropic case, but space distorted. Ellipses instead of circles.
Add two extra parameters - *anisotropy angle* and *ratio*.

The example below is for a simulated anisotropic r.f. and empirical variogram with two angular zones.



(scale = 1, N-S is stretched by anisotropy factor of 3, aniso. angle = 0°)

To fit an anisotropic model:
When aniso.angle ϕ is 0 or $\pi/2$, then compute "reduced distance"

$$h^* = \sqrt{(h_1/a_1)^2 + (h_2/a_2)^2}$$

The ratio a_1/a_2 defines how stretched the ellipses are. For example, let $\gamma_{EW}(h_1) = 1 - e^{-h_1}$, $\gamma_{NS}(h_2) = 1 - e^{-h_2/3}$.
If take $a_1 = 1$, $a_2 = 3$, then $\gamma(h_1, h_2) = 1 - e^{-h^*}$ will coincide with γ_{EW}, γ_{NS} above.
Naturally, this requires us to use the same models for both N-S and E-W directions.

For other aniso.angles: first, rotate coordinate axes to line up with aniso.angle.

Zonal anisotropy: different sills - **not** stationary. (See Isaaks & Srivastava, p. 377). Models layering effects.

More sophisticated kind of anisotropy (Chilès and Delfiner):

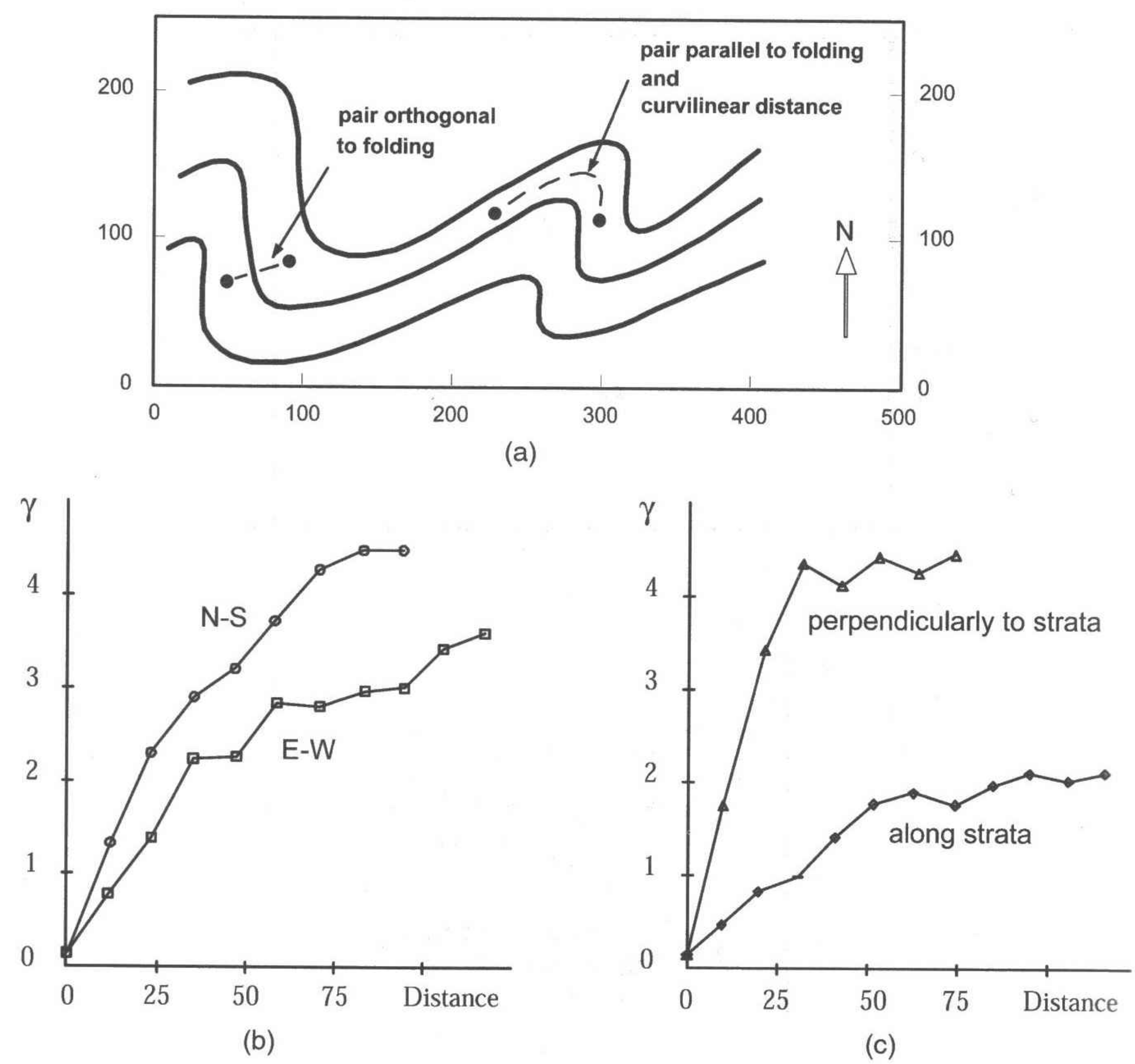


FIGURE 2.5. Calculation of a sample variogram in folded beds: (a) stratification pattern; (b) variogram calculated in the Euclidean system; (c) variogram calculated according to the stratification.