<u>Week 6.a Lecture Material</u> Experimental Variography

1. Key Concepts:
- calculate <u>frequency distribution of data pair separations</u>
- <u>identify potential problems with data</u>: non-stationarity [moving window analysis of mean and variance], heteroscedasticity, proportional effect, utilize h-scatterplots
- identify and <u>model the underlying correlation structure</u> in spatial data, including anisotropic structure, nugget variance and variogram range

2. Process for defining variogram correlation structure
- see Generalized Geostatistical Analysis and Modeling Sequence
- see also Olea (1994) assigned reading, VarioWin Help file

3. Techniques for working on and cleaning up experimental variograms:
- see Pannatier, Ch.2 for introduction to the use of VarioWin software
- <u>variography requires experience and practice</u>; some useful things to try on a "noisy" (ie. visually erratic) variogram (results will differ from dataset to dataset):

a. (see Hohn, p.59 ff): use <u>transforms</u> of the data for skewed distributions (e.g. normal-score, logarithmic transforms) to mitigate the effect of extreme values on the mean of the calculated statistic
b. use <u>other variogram measures</u> to account for skewed data or data with large outliers (e.g., inverted covariance, correlogram, standardized semivariogram)
c. <u>remove outliers</u> from the variogram using an h-scattergram

d. check for <u>sufficient number of pairs</u> at each lag distance (ideally greater than 30 to 50)
e. use a larger lag tolerance to get more pairs and a smoother variogram, but strive for the <u>minimum lag spacing that will give acceptable pair counts</u> near the origin
f. try <u>variable lags</u> (GeoEas or GSLIB), or <u>overlapping lags</u> (VarioWin or GSLIB) for noisy data sets
g. <u>maximum lag</u> should be no larger than about half the maximum pair separation distance
h. <u>start with an omnidirectional variogram</u> before trying to define directional variograms; there is no reason to expect to see structure in the directional variograms if the omnidirectional variogram is very noisy

i. look for <u>consistency of range and directionality</u> between different variogram measures to avoid erroneous conclusions; if the data are not skewed (or normal-score transformed data are used), and/or if the data are declustered, all these variogram measures should produce similar patterns of correlation structure
j. <u>to estimate variogram range</u>, use the standardized variogram, correlogram or mean absolute difference; <u>to constrain the nugget and sill</u> of the variogram use the semivariogram or inverted covariance

4. Problem Set **IV / V. Spatial Correlation 1 - Experimental Variography**
- manual calculation, comparison with VarioWin; initiate analysis of Walker Lake data

Words of Wisdom on Experimental Variogram Analysis

- use the smallest lag spacings that produce acceptable numbers of pairs (ideally, a minimum of 30-50 pairs per lag) in order to identify small-scale correlation; try not to infer short scale correlation structure from pairs at too large separations, as this will usually produce artificially large estimates of the nugget variance

- it is easy to mask spatial correlation by a poor choice of lag spacings; rarely can artificial correlation be generated where it does not exist

- always analyze histograms of sample spacing (calculated h values in a .cld file) before choosing initial lag spacings; import the ascii .cld file created by VarioWin into StatMost and look for the smallest histogram class intervals that give at least 30-50 samples per class interval - this will be a good starting estimate of the lag spacing, which can be iteratively refined calculating variograms with slightly different lags in order to find the best lag spacings (note that the best lag spacings may differ for an isotropic vs. anisotropic variograms of the same data!)

- if the data are clustered, use variable lag spacings for the first few lags (Note: VarioWin does not permit variable lag spacings; use GeoEas to do this)

- use the smallest directional tolerance angle that will retain acceptable numbers of data pairts per lag; use of too large a tolerance angle will mask anisotropy that may be present and will reduce the apparent degree of anistropy where it is identified

- indicator variograms are based on a categorized variable; the indicator can represent the presence/absence of a particular category (rock type, ore mineral, etc.), or whether a continuous variable is above or below a specified threshold; they are as valid as variograms of continuous variables, and are calculated in exactly the same way, except that the value of the indicator variable is only either 0 or 1 and its mean and variance represent the proportion of 0's and 1's

- if the spatial scales of the coordinate axes are very different (eg: thousands of meters in x,y, 10's of meters in z), rescale the axes so they have similar ranges; also check that the squared coordinate values do not exceed the computer's precision or the software's capability to handle large values (eg: VarioWin is limited to 5-digit coordinate values for x and y)

- be aware of the spatial process(es) which created the regionalized variable being analyzed and their potential impacts on spatial correlation: for example, if a sedimentary formation pinches out in one direction (variable thickness), the relative positions and physical properties of sediments are related in a non-orthogonal coordinate system; the spatial coordinates of the regionalized variable would have to be transformed to orthogonal stratigraphic coordinates before analysis (eg: see equation 4.5 above). In another example, if the goal of variogram analysis is to analyze sedimentary textural variations in depositional facies in which the original stratigraphic orientations of the sediments have been deformed by folding, then the non-orthogonal coordinate system of the folded strata must be transformed into an orthogonal system prior to spatial correlation analysis; however, if the spatial correlation analysis is aimed at analyzing the effects of thermal metamorphic redistribution of silica in those sediments, the spatial correlation analysis should probably start on the untransformed coordinates because the silica distribution may be controlled more by the metamosphism than the original stratigraphy

- always compare several estimators of spatial correlation (for example, semivariogram, inverted covariance and inverted correlogram). If the correlation structure represented in these various estimators is similar, confidence in the interpretation is bolstered; if not, then the underlying cause must be determined. For example, with clustered data, plotting the lag means and variances

($\overline{z}_{-h},\ \overline{z}_{+h}\ and\ \sigma^2_{-h}, \sigma^2_{+h}$) vs. h will show that closely-spaced pairs of data tend to have different lag means and possibly different lag variances than do widely-spaced data (to do this with VarioWin, click on Options|Numerical Results with a variogram window open, then File|Save As to save an ascii file of the numerical values of the experimental variograms and associated statistics like the lag means and variances; import this file into StatMost and plot $\overline{z}\ and\ \sigma^2$ vs. h. In another example, a few outlier values may greatly alter the apparent correlation structure; check the h-scatterplots and look for pairs involving one or two “troublesome” data points whose removal may substantially improve the look of the variograms. Note that the inverted covariance, correlogram and madogram, as well as other estimators of spatial correlation like the pairwise relative and general relative semivariograms (see Deutsch and Journel, 1997, p.45) are less sensitive to clustering and data outliers and therefore may provide cleaner estimates of spatial correlation

- above all, always use geologic judgement and other information to identify the properties of the sample data; the interpretation of variogram structure is subjective and should be based on prior experience and expert opinion. For example, if some data points badly affect the variogram, document their effect and remove them; if the process that generated the regionalized variable is well characterized, use that information to guide the selection of the “best” variogram estimators (eg: if subtle trends are known to exist in the data, an intrinsic or weakly stationary variogram structure might be expected; if the underlying process is known to be highly continuous, nugget effects might be expected to be small or negligible)

Variogram Analysis Software

3PLOT32 for Post Plot, Stationarity Analysis -

1. plot data to check locations for correct file!
2. check for obvious trends, data clustering, clustering of extreme values
3. use moving window to compare means, variances across the sample domain
4. plot variances vs means for all windows to identify non-stationarity
5. if extreme variations (criterion: ...), subdivide sample domain into sub-domains and analyze individually

PREVAR for Pair-Comparison File Creation -

1. set X, Y coordinate info
2. set max pair separation distance (NB for large data sets!)
3. debug .DAT file if necessary (for common coordinates, etc.)

VARIO2D for Experimental Variograms -

1. use variogram map to identify **strong** anisotropy (ignore if weak) and its direction
2. start with non-directional variograms, determine best lag spacing, max lag
3. for directional variograms, start with 30o tolerance, find max correlation range
4. when max range found, try smaller tolerances and check principal direction again
5. fine-tune lag spacing, tolerance angle, max bandwidth for max principal direction
6. repeat 3-5 for minimum principal direction
7. keep real time record of lag spacings used, tolerances, directions, etc.
7. look for internal consistency between different correlation structure estimators
8. if not consistent in directional variograms, or if noisy, use non-directional option as the only defensible interpretation
9. utilize lag cross-plot to eliminate outliers before saving (explain significance of plot)
10. save multiple experimental variograms (non-directional, max/min principal directions, to .VAR file

MODEL for Variogram Model Fitting -

1. select experimental variogram from .VAR file
2. starting with best correlation structure estimator(s), fit model, keep real time record of fitting parameters
3. use best estimator(s) to fix range; use inv.cov. to fix sill and nugget
4. check for consistency between different estimators’ fitted nuggets, ranges and sills
5. if necessary, iterate between different estimators to achieve consistency
6. save final fitting parameters to .MOD file

Week 6.b Lecture Material Modeling of Correlation Structure

1. Variogram Models
- the purpose of modeling the experimental variogram's correlation structure with a correlation function model is to produce a mathematically-concise and clean description which guarantees numerically-stable solutions to the matrix operations used in estimation and simulation algorithms
- see Isaaks and Srivastava (1989), Chapter 16, and Deutsch and Journel (1997), pp.24-26, for the types of models and model equations for correlation functions:

Second-order stationary models:
- nugget: $\gamma(h) = \gamma(0) = \text{constant}$
- spherical: $\gamma(h) = c \cdot [B]$ where $B = 1 \; if \; h \geq a$ and $B = \frac{3}{2}(\frac{h}{a}) - \frac{1}{2}(\frac{h}{a})^3 \; if \; h \leq a$
- exponential: $\gamma(h) = c \cdot [1 - \exp(\frac{-3h}{a})]$
- gaussian: $\gamma(h) = c \cdot \left[1 - \exp(\frac{-(3h)^2}{a^2})\right]$

Intrinsic models:
- linear (exponent = 1) and non-parabolic (exponent <2) power models: $\gamma(h) = |h|^{exponent}$
- periodic or hole-effect model: $\gamma(h) = c \cdot [1 - \cos(\frac{h}{a} \cdot \pi)]$
- logarithmic: see Hohn (1988); this type of structure is best modeled as a combination of exponential and power models, rather than a separate functional form (log model is not useful for spatial modeling)

- in all of the above, c represents the variogram sill, a represents the variogram range, and h is lag
- the hole-effect model can only be applied in one direction; practically, this means that the effective range(s) perpendicular to the direction of the maximum hole-effect must be so large as to negligibly contribute to the variogram's sill in those directions (a particular case of zonal anisotropy - see section 4 below)
- interactive modeling of experimental variograms is available in both GeoEas and VarioWin, using the four second-order stationary models as well as the linear model shown above
- these models can be be equally applied to describe spatial correlation as estimated by the semivariogram, standardized semivariogram, inverted covariance, inverted correlogram, madogram or other statistical estimators

2. Choosing a Model
- ensure that the model is consistent with the physical process responsible for the r.v.
eg: continuity at the origin (Gaussian model) for very continuous processes (eg: water table elevation); intrinsic models where a trend is present (eg: cyclic stratigraphy)
- decide whether small-scale correlation structure will be defined by the sample data alone or by external information (eg: presence or absence of a nugget variance); large nugget variance for processes known or inferred to have a large amount of small-scale variability

3. Nested Models
- multiple correlation function models can be combined linearly to model almost any degree of complexity in an experimental variogram
- for example, an experimental variogram that has an exponential form and a non-zero intercept can be represented as a combination of an exponential model and a nugget model:

$$\gamma(h) = \gamma(0) + c \cdot [1 - \exp(\tfrac{-3h}{a})]$$

- such “nested” models can be combined in any order and any number to satisfactorily reproduce the structure observed in the experimental variogram; however, it is generally advised not to “over-model” the variogram by using too many nested structures, because the physical significance of a large number of nested correlation structures usually cannot be defended
- the incorporation of multiple nested models is defensible where several scales of processes are known or inferred to be present (see example in Hohn, p.31-32); in such cases, each nested model structure represents a particular scale (range) of influence of the underlying spatial correlation structure and the resulting overall model of correlation structure is both an accurate representation of the experimental variogram and physically meaningful

4. Modeling Anisotropy in Spatial Correlation
- almost always, when anisotropy is present, directional variograms are modeled with orthogonal principal components of anisotropy (ie. directions of maximum and minimum ranges); this is mandated by the manner in which the directional variogram structures are combined to describe spatial correlation in all directions and the manner in which this information is implemented in kriging and simulation. However, the principal components of anisotropy in natural systems need not be perpendicular. Indeed, to identify and understand the underlying physical processes (eg: sedimentary deposition, metamorphic fabric, etc.), one does <u>not</u> assume orthogonality *a priori*; in such cases, variogram analysis is applied purely to identify the principal directions of anisotropy (see Davis et al, 1994) of the underlying physical process, and the empirically determined correlation structure represents potentially important information about the anisotropy of physical process
- <u>geometric anisotropy</u> represents a situation where the variogram sill is constant in all directions and variogram range varies with direction; this is the most common type of anisotropy and is the only type that can be modeled in VarioWin, GeoEas and GSLIB software; it is manifested in two directional variograms with the same sill values
- <u>zonal anisotropy</u> represents the opposite situation: range remains constant in all directions, whereas the variogram sill varies (see Isaaks and Srivastava, 1989, Fig. 16.7, for a graphical representation of these situations); this type of anisotropy cannot be represented directly in most software, but it can be viewed as a special case of geometric anisotropy and represented by <u>two</u> model structures, each of which has such a large range in one of the principal directions that the contribution of the model structure in that direction is negligible
- the nugget variance should always be viewed as an isotropic parameter; it is not physically meaningful to say that uncorrelated (random) variability at a point is different depending on the direction of looking away from that point. Instead, if the apparent nugget in two directional variograms is substantially different, it should be modeled as an additional anisotropic nested (non-nugget) structure with a maximum range so large as to add negligibly to the variogram at small lags

5. Introduction to Cross-variograms
- where two or more regionalized variables are correlated, the nature of spatial cross-correlation between the primary variable and a more extensively-sampled secondary variable can provide valuable information for estimation and simulation of the primary variable
- the calculation of the cross-variogram estimator is very similar to the direct variogram discussed in Week 5; for example, the expression for cross-covariance is exactly the same as that for covariance (equation 4.2):

$$C(h) = \frac{1}{n_h} \sum_{i=1}^{n_h} (z_i^a \cdot z_{i+h}^b) - \overline{z}_i^a \cdot \overline{z}_{i+h}^b$$

where z^a_i and z^b_{i+h} represent a pair of two different variables separated by a distance h; the same substitutions can be employed for all the estimators discussed in Week 5 except the cross-semivariogram, which is defined slightly differently than in equation (4.3):

$$\gamma(h) = \frac{1}{2n_h} \sum_{i=1}^{n_h} (z_i^a - z_{i+h}^a)(z_i^b - z_{i+h}^b)$$

- as in direct variography, all of these cross-variogram estimators are equivalent representations of spatial continuity (under the assumption of second-order stationarity) and can be used to define a model of the cross-correlation function from the experimental cross-variograms

- unlike direct variogram modeling, cross-variogram modeling is always done for the purpose of developing a model to be used in estimation or simulation; also unlike direct variogram modeling, the models that are to be used in estimation and simulation must obey a number of stringent constraints to ensure that the matrix solutions to the kriging equations exist and are numerically stable; the existence of these constraining conditions can be difficult and tedious to demonstrate, but fortunately, several general rules exist to simplify the problem; see Isaaks and Srivastava (1989), pp.390-398, and Goovaerts (1997), p. 175, for a discussion of these conditions

6. Problem Set **VI. Spatial Correlation 2 - Variogram Modeling**
- perform parameter fitting and define an isotropic as well as two directional correlation structure models for the Walker Lake data set

Words of Wisdom on Variogram Modeling

- when modeling directional variograms in VarioWin, open the .var file, hold <shift> down, and select both variograms (ie. max and min range directions) to import simultaneously; this places both variograms in the graph window and allows the sills of both to be automatically matched in a geometric anisotropy model of both variograms

- always use the smallest number of nested structures to produce an acceptable model fit to the experimental variogram

- when modeling geometric anisotropy in VarioWin's Model program, import the two principal directional variograms simultaneously and specify the degree of anistropy, so that the two variograms' sills are standardized to one another

- generally, it is OK to adopt the variogram estimator that produces the most interpretable and cleanest correlation structure; the relative estimators (eg: correlogram, standardized semivariogram) and the madogram cannot be used to estimate the nugget and sill, but provide equivalent measures of the range; in contrast, either the semivariogram, madogram, or inverted covariance can be used to estimate nugget and sill, as well as the range

- model a variogram's structure on the basis of preexisting information or expert knowledge of the system under study, rather than relying on automated fitting or synthetic measures of best fit (eg: VarioWin's goodness-of-fit statistic)

- when modeling a variogram for kriging purposes, the choice of nugget variance is not critical, but should be considered in light of the type of kriging to be performed (eg: point vs. block); however, when modeling for purposes of stochsastic simulation, all or part of the nugget variance should be modeled with a nested structure of short range, so that simulation can better represent small-scale variability at small grid spacings

- above all, always use geologic judgement and other information to best model the sample data; the modeling of variogram structure is subjective and should be based on prior experience and expert opinion. For example, if a few number of pairs define one variogram point give it less weight when evaluating a model fit than points defined by a large number of pairs. If the process that generated the regionalized variable is well understood, use that information to guide the selection of the "best" variogram models (eg: if subtle trends are known to exist in the data, an intrinsic or weakly stationary variogram structure might be expected; if the underlying process is known to be highly continuous, nugget effects might be expected to be small or negligible)

<u>Week 7-8 Lecture Material</u> Kriging Concept, Introduction to Kriging System of Equations

1. Spatial Estimation
- estimation of values of a variable at unsampled locations through interpolation
- see Hohn, p.101, for a summary of the steps in contour mapping; the key step is interpolation
- commonly used techniques of interpolation: (see Cressie, 1993, section 5.9 for extensive review and references)
    - a. trend-surface fitting
    - b. nearest-neighbor data weighting
        - moving average (problems with data clustering)
        - triangulation, polygonal declustering / averaging
        - inverse distance weighting
        - statistical weighting (kriging)

- <u>kriging</u> is a collection of generalized linear regression techniques for minimizing estimation variance defined by a prior covariance model (Deutsch and Journel, p.14)
- kriging is a B.L.U.E. - best, linear, unbiased estimator:
    - best = minimized estimation error
    - linear = determined by a linear (statistically) weighted average of nearest neighbors (like inverse distance weighting)
    - unbiased = sum of all the estimation errors is zero
- unlike kriging, the other methods of estimation are either not the "best" estimates (eg: trend surface) or are not necessarily unbiased estimates (eg: moving average)

2. Why Kriging?

Deterministic interpolation techniques such as inverse distance and triangulation do not take into account a model of the spatial correlation (ie. the variogram). One might be able to obtain a contour map and subjectively think that the spatial process is represented in the contours, but objectively, the interpolated values cannot be defended. Kriging relies on the spatial correlation reflected in the available data and so represents a global view of all the data as well as nearest neighbor influence.

Furthermore, kriging allows you to quantify the quality of your predictions via the kriging variance. You can do the same for deterministic techniques, but it's quite tedious; you'd still need to model the variogram and derive the estimation weights.

Another advantage of kriging is that it takes data clustering (redundant data) and possible spatial anisotropies into account much more easily than other interpolation techniques. Furthermore, more advanced geostatistical techniques such as indicator kriging and simulations permit the incorporation of “soft" or qualitative information in a quantitative manner.

Kriging requires more work in data analysis and variogram modeling, which takes time and effort. Nevertheless, the results from kriging are generally superior, more realistic and generally more defensible compared to techniques such as inverse distance or triangulation.

3. Types of Kriging
- point vs. block: estimation at a scale equivalent to the scale of "point" measurements vs. estimation of the average value over a larger scale; see Goovaerts (1997), Fig.5.9, p.158
- simple (SK) vs. ordinary (OK): kriging with prior knowledge of a stationary global mean vs. kriging in a moving local window where the local mean is stationary and defined by the data within the moving window
- universal kriging (UK): spatial estimation of a r.v. which contains a trend component
- cokriging (CoK): spatial estimation of a primary r.v. constrained or informed by spatial variations of, and cross-correlation with, one or more secondary variables

4. Applications of Kriging
- spatial estimation (interpolation)
- data analysis (identifying and separating different scales of variation)
    - see Goovaerts (1997), Table 5.1, p.164
    - several types of kriging that can be used for analyzing scales of variability:
        - factorial kriging (Goovaerts, 1997, p.160)
        - kriging the trend (similar to UK, but the goal is to estimate the trend within the data and not to incorporate the trend in generating spatial estimates from data)
        - kriging with an external drift (similar to cokriging but with a secondary variable defined at all grid locations)

5. Global vs. Local Trends, Non-Ergodicity, Non-Stationarity
- for general comments, see Cressie, pp. 52-58; for comments related to stochastic simulation problems, see Deutsch and Journel, 1992, p.126-130
- ergodicity and stationarity are central to the theoretical justification for inferring R.F. properties from variograms, and crucial to the unbiased nature of kriging interpolation; how, then, is it possible to treat data such as contaminant plumes or ore bodies (ie. local anomalies) in the geostatistical framework?
- the existence of a spatial trend is central to questions of both ergodicity and stationarity
- first and foremost is the issue of scale: for second-order stationarity, the mean and variance over the study area are assumed to be independent of location; this would only be possible if the spatial dimensions on which the criterion of location-independence is based are very large relative to the local variations of the regionalized variable, so that the effect of a local high (eg: contaminant plume) on the regionalized variable's overall distribution would not be sufficient to markedly change the confidence interval about the mean
- fortunately, practitioners have found that global stationarity is not as important as <u>local stationarity</u>; for example, local stationarity could be demonstrated if, within the transition region of the variogram, a windowed average of the regionalized variable did not drift outside the 95% confidence interval of the surrounding windowed averages
- the issue of ergodicity could be treated in a similar manner, although in practice proving ergodic behavior is rarely possible; for example, if a trend exists, then the regionalized variable is non-stationary and hence non-ergodic (Cressie, 1993, p.56); but by what criteria should the existence of a trend be defined so that the data can be de-trended? even after de-trending, ergodicity can only be demonstrated if the variogram achieves a well-defined sill at large lags and the spatial data represent a normal population distribution (ibid., p. 56)

- in practice, ergodicity is frequently assumed but rarely demonstrated and, mostly, not well understood (ibid., p.54); it is almost never verifiable except in a weak sense where the hypothesis cannot be rejected (ibid., p.57)

- a related scale issue which bears on trend analysis is: at what scale does the existence of a trend threaten the geostatistical analysis? for example, in this context "scale" could be an inverse function of the order of the polynomial used to de-trend a data set; thus, a first-order model of a trend would demonstrate a clear regional trend underlying the data, but a tenth-order polynomial describes the presence of many localized highs and lows, each of which could be construed as a local-scale trend
- the answer ultimately lies in the experimental variogram: as stated in Week 4, second-order stationarity is demonstrated in variograms which achieve a sill at large lags; if the semivariogram does not achieve a constant sill at large lags, a trend is present and the intrinsic hypothesis must be invoked; only the lowest-order polynomial trend surface is acceptable, in order that the trend surface residuals describe a semivariogram sill, at which point second-order stationarity is achieved; if in addition, the data are normally-distributed, ergodicity can sometimes be demonstrated (Cressie, 1993, p.56)

- key points:
  - the assumption of ergodicity is crucial to the validity of all geostatistical analysis and is always invoked, but rarely can be proved
  - the existence of a trend is suggested by lack of attainment of a sill in the variogram; the presence of a sill implies second-order stationarity
  - second-order stationarity ensures unbiased kriging; a non-stationary data set can be de-trended with the lowest-order polynomial surface required to achieve a variogram sill or handled with <u>universal kriging</u>, <u>kriging with a trend</u>, or <u>disjunctive kriging</u>
  - de-trending is necessary only when local stationarity cannot be demonstrated; thus, kriging of contaminant plume concentrations does not require removal of the "trend" of the plume concentrations as long as the scale of variation of plume concentrations is appropriate to both variogram structure and the nearest-neighbor search distances employed in kriging (see below)

6. Conecptualized Development of the Kriging System (for exact derivation, see Isaaks & Srivasatava, Ch.12, p. 279-289)
- see Fig. 4.1 and Table 4.1 in Clark for a simple example; also Isaaks and Srivastava, p. 290 ff.
- an estimate at either a point (or an areal or volumetric "block") can be derived from a weighted linear average of surrounding known values:

(7.1) $$z^* = w_1 z_1 + w_2 z_2 + ... + w_i z_i = \sum w_i \cdot z_i$$ (linear estimate)

where z* is the estimated value at an unsampled location and the z_i are known values, weighted by factors w_i
- the simplest case is where the w_i are equal, giving a simple arithmetic average of known values
- this could be applied globally, in which case z* would be the global average of all data points;

alternatively, the average could be computed for some local neighborhood of known values, as in a moving window averaging technique
- inverse distance weighting methods define w_i as a function of 1/separation distance between each neighborhood point z_i and the estimation point z* eg:

(7.2) $$w_i = (\frac{1}{h_i^n}) / \Sigma(\frac{1}{h_i^n})$$

where h is the separation distance between point i and the estimation point *, n is a power factor (usually 2) to give greatest weight to the closest neighbors, and the w_i sum to 1 (linear, unbiased estimate): $\Sigma w_i = 1$

- a measure of how good is the estimate z* can be defined as an estimation error (residual), r:

$$r_i = z_i^{true} - z_i^*$$

- in the absence of a trend, the mean of the r_i would have an average of zero
- similarly, the estimation variance, σ^2_K, can be defined as:

(7.3) $$\sigma_K^2 = \frac{1}{k} \Sigma (z_i^{true} - z_i^*)^2$$

where k is the number of estimated values and z_i^{true} are the actual values at each location (which are never known, so this definition of estimation variance is not very useful)

- note that the definition of the estimation variance and the definition of the semivariogram or cross-semivariogram statistic are conceptually equivalent if the true and estimated values are viewed as a regionalized variable compared at zero lag; ie. $\sigma_K^2 = 2\gamma(0)$
- by generalizing this conceptual relationship to any lag, $\gamma(h)$, and using the relationship that exists between semivariogram and covariance under second-order stationary conditions:

(7.4) $$\sigma_K^2 \sim 2\gamma(h) = 2(\sigma_z^2 - C_{io}) = 2(C_{oo} - C_{io})$$

we see that the estimation variance can be thought of as related to covariances: covariances between neighboring data, i, and the estimation point, o (C_{io}); and the covariance at zero lag (C_{oo}, which is equal to the sample variance, σ_z^2)

- equation (7.4) is a non-rigorous, conceptual demonstration of the existence of a relationship between the estimation variance, σ^2_K, and the covariances between data points and estimation location; in practice, the estimation variance has to be rigorously defined in terms of contributions to the covariance between local neighbor data points and the estimation point as well as between local neighbor data points, and generalized to permit estimates over a block (area or volume) as well as a point
- see Isaaks & Srivastava, p.281-283 for the derivation of the error variance in terms of the weights and covariances for point kriged estimates;
- for estimation at a point (point kriging) or at the center of a block (block kriging):

(7.5) $$\underbrace{\sigma_K^2 = \sigma_z^2 + \sum \sum w_i w_j C_{ij} - 2 \sum w_i C_{io}}_{\text{(Point)}} = \underbrace{\bar{C}_{oo} + \Sigma \Sigma w_i w_j C_{ij} - 2 \Sigma w_i \bar{C}_{io}}_{\text{(Block)}}$$

where the $C_{ij}, \bar{C}_{io}$ terms represent the covariance contributions arising from the spatial distance between neighborhood data points, and the covariance contribution between a neighborhood data point and the estimation point (or the average covariance contribution between a neighborhood data point and the estimation block), as calculated from the correlation function model deduced from the data

- note that the difference between point and block kriging is the definition of C_{io} *and* C_{oo}: in point kriging, C_{oo} is the same as the global sample variance, σ^2_z, whereas in block kriging $\bar{C}_{oo}$ is the average covariance between all points within a block; in point kriging, $C_{io}$is the covariance between neighboring data and the estimation point, whereas in block kriging, $\bar{C}_{io}$ is the average covariance between neighboring data points and the block-average of all points within a block

- if all we did was to calculate weights which sum to 1, we would not have realized the power of kriging, because the values so estimated would define an estimation variance, σ^2_K, which may or may not be the minimum variance had some other combination of linear weights been used (ie. only a L.U.E. rather than a B.L.U.E.)
- on the other hand, if all we did was to minimize the estimation variance with respect to the weights:

$$\frac{\partial \sigma_K^2}{\partial w_i} = 0 \qquad \text{for i = 1, 2, ... n (n equations in n unknowns)}$$

this would not satisfy the requirement that $\Sigma\, w_i = 1$ (the "linear, unbiased" in B.L.U.E.)

- thus, in order to achieve a “true B.L.U.” estimator, we need to satisfy n+1 equations (minimum error variance, plus n equations analogous to 7.2, defining the weights), with only n unknowns (the weights); this is an over-determined system, lacking a unique solution (ie. there are many combinations of the w_i that would satisfy the n+1 equations)

- to get around this problem an additional (artificial) unknown, the LaGrangian Multiplier (λ), must be introduced so as to increase the number of unknowns to match the number of equations; this is done by adding a fourth term to the RHS of equation (7.5) in such a way that it will not change the original equality of (7.5) (see Isaaks & Srivastava, p. 282-284; Clark, Ch.5)
- for point kriging:

$$(7.6) \qquad \sigma_K^2 = \bar{C}_{oo} + \Sigma\,\Sigma\, w_i w_j C_{ij} - 2\,\Sigma\, w_i\, \bar{C}_{io} + 2\lambda(\Sigma\, w_i - 1)$$

note that $\Sigma w_i = 1$, so the added term $2\lambda(\Sigma w_i - 1)$ is zero, and hence (7.6) and (7.5) are equivalent

- the minimization is then performed on the expression $\sigma_K^2 - \lambda(\Sigma\, w_i - 1)$ with respect to the n+1 variables, w_i and λ, yielding the following equations:

n equations of type $\frac{\partial \sigma_K^2}{\partial w_i} = 0$ and one equation of type $\frac{\partial \lambda(\sum w_i - 1)}{\partial \lambda} = 0$

- the minimization with respect to λ ($\partial/\partial\lambda$) yields the original unbiasedness constraint:

$$(7.7) \qquad \Sigma\, w_i = 1$$

and minimization with respect to each of the w_i yields equations for each data point (i = 1 to n) of the form (Isaaks & Srivastava, p.286-7). In the case of block kriging the n equations have the form:

$$(7.8) \qquad \sum_{j=1}^{n}(w_j \bar{C}_{ij}) + \lambda = \bar{C}_{io}$$

where the symbol $_o$ stands for the block at which an estimate is made.

- the system of equations given by (7.7) and (7.8) is called the "ordinary kriging system", and can be written in matrix notation as:

$$(7.8) \qquad \underset{\mathbf{C}}{\begin{bmatrix} C_{11} & C_{12} & \ldots & C_{1n} & 1 \\ . & . & & . & . \\ . & & . & . & . \\ . & & . & . & . \\ C_{n1} & C_{n2} & \ldots & C_{nn} & 1 \\ 1 & 1 & \ldots & 1 & 0 \end{bmatrix}} \bullet \underset{\mathbf{w}}{\begin{bmatrix} w_1 \\ . \\ . \\ . \\ w_n \\ \lambda \end{bmatrix}} = \underset{\mathbf{D}}{\begin{bmatrix} C_{01} \\ . \\ . \\ . \\ C_{0n} \\ 1 \end{bmatrix}}$$

- the solution for the kriging weights is then: $\mathbf{w} = \mathbf{C}^{-1} \cdot \mathbf{D}$

7. Conceptual Interpretation of the Kriging Matrices:
- the **D** matrix of covariances can be viewed as an analog to the weights used in inverse distance estimation methods; that is, like an inverse distance weight, the covariance between a data point and a point being estimated decreases as the distance between the two increases
- the $\mathbf{C}^{-1}$ matrix is a multiplier that scales the D weights to sum to unity, assuring an unbiasedness estimate; in effect, it accounts for data redundancy (clustering) by automatically giving lower weight to closely-spaced data points
- both the **D** and the $\mathbf{C}^{-1}$ matrices express separation as a statistical distance rather than a geometric distance; they carry information on the data's spatial correlation characteristics while at the same time weighting the data for the effects of data clustering (see Isaaks and Srivastava, 1989, p.299-301); in effect, the calculated kriging weights reflect the effect of spatial correlation and the effect of data positions

8. Point vs. Block Kriging:
- use point kriging where an estimate (or its uncertainty) at a given point is required
- if estimating statistically averaged values of an area or volume (a "block"), use block kriging

- the dimensions of the block are determined by the measurement process that produced the data being kriged (eg: a volume equivalent to the radius of influence of well tests or slug tests, or the incremental block volume extracted in a mine stoping operation)
- for information on how block averages and data point-to-block averages are computed, see Isaaks and Srivastava (1989), Chapter13

9. Notes on Scale-Dependent Measurements:
- all geoscience measurements are made on a sample of the earth at a particular scale eg: petrofabric measurements on core samples, on outcrops, on aerial photos
- measurements of a regionalized variable made at one scale (eg: core-scale measurements of ore grade or permeability) should not be used to obtain kriging estimates at another scale (truckload ore grade scale, or well-test scale), unless an appropriate correction is applied to account for the scale effect; this problem is dealt with at the end of this course

10. Cokriging:
- in data sets containing more than one regionalized variable which are correlated with one another, the kriging estimate and estimation variance of one variable could be further reduced if spatial correlation information on another, more widely sampled variable, were incorporated in the kriging system
- beware that in attempting cokriging, the correlation function models used to describe the covariance of the R.F. include the direct variograms of each variable, plus cross-variograms of each pair of variables considered (all in their directional forms, if anisotropy is present); this can lead to a large number of variograms to infer and model, and requires that the sills and nuggets values of all direct and cross-variograms satisfy a number of additional consistency constraints to ensure that the solution to the kriging equations exists and is unique; see Isaaks and Srivastava, Ch. 17 for more information

4. Problem Set **VII. and VIII. Kriging with GeoEas**
- use your subjective "best" and "worst" variogram models from the variogram modeling of the Walker Lake data set to perform kriging estimation for both cases

Kriging Methods in Order of Power and Complexity:

Simple Kriging - stationary global mean

Ordinary Kriging, Indicator Kriging - stationary local mean

Cokriging - bivariate or multivariate data

OK with External Drift - secondary variable at all grid locations

Kriging the Trend (Universal Kriging) - defining or incorporating a trend

Factorial Kriging - decomposing local and regional scales of variability

Week 9-10 Lecture Material Practical Implementation of Kriging

Any estimation problem involves at least three steps:

a. gridding the sample domain;
b. performing estimation at grid nodes;
c. interpolation (visual or contour-based).

The latter is a straightforward application of well-tested algorithms, and does not concern us. The former two steps, however, are at the heart of successful estimation, and also lay the foundation for understanding simulation techniques.

1. Gridding

- too coarse a grid leads to excessive smoothing (loss of information), whereas if too many grid points are used (where "too many" is relative to the number of data available), computational efficiency is sacrificed with no gain in accuracy

- a starting point for choosing a grid node spacing is the optimum cell size determined by DECLUS; this node spacing will be optimal in the sense that it is based on the amount of data and their spatial distribution; however, this should be considered a starting point, only

- the density of grid nodes as well as their actual x,y locations is important to consider; the grid node locations represent the center of blocks when block kriging is implemented

- if point kriging is used, the grid node locations <u>must not</u> coincide with data locations (otherwise spatial estimation at that point does not occur, and a "bull's eye" effect results); if block kriging is used as a convenient way to get around this problem, the block size should be small enough to justify the estimates so derived as quasi-point estimates

2. Estimation

- is based on a weighted average of data in a search neighborhood
- this is the most crucial step in kriging, and the choice of estimation parameters here is key to the success or failure of the kriging project
- several sets of parameters must be specified; these can be grouped into those defining

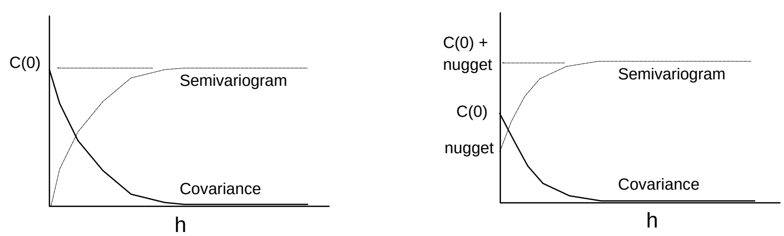
a. the geometry of the search neighborhood;
b. the number of data within the search area to be used in estimation; and
c. criteria for defining the closest neighbors

a. The Search Neighborhood

- the search is confined to a local neighborhood for several good reasons:
  - to reduce computation time by eliminating data points likely to have very low weights;
  - to be able to accommodate varying local means; and still
  - satisfy the assumption of second-order stationarity for the kriging system of equations.
- the needn't extend any farther than the range of the variogram, but could be set much smaller, if the data points are fairly evenly distributed (not clustered)

- to more efficiently handle various arrangements of data, the search area can be circular or elliptical, and can be subdivided into sectors, each of which is searched independently

b. The Number of Data
- this is one of the most important search parameters
- recall that: "variograms" (semivario., inv. covariance, etc.) are the graphical representations of a correlation function that is used to define the functional parameters of the covariance for kriging; that the covariance can be viewed as a statistical analog to inverse-distance weighting; and that under second-order stationarity, the variogram and covariance are related; ie. for a zero nugget, $C(0)$ = variogram sill height = σ^2, and $C(h)$ levels off at 0; with a nugget, $C(h)$ still levels off at zero, but starts with a discontinuity at $h = 0$ such that $C(h) \neq C(0)$ for $h > 0$:



- so if we view the covariance function as a statistical weighting function analogous to inverse-distance schemes, then in a situation with zero nugget, distant data are statistically dissimilar to the point being estimated and therefore receive a low weight (= low covariance); however, for a nugget that is large relative to the population variance (variogram sill), even distant samples have more of a statistical similarity to the point being estimated (because of the random, non-correlated nugget variance) and hence will have relatively larger kriging weights; in the extreme case, with a pure nugget variogram, all samples at all distances receive equal weight

- ideally, the number of nearest neighbors is chosen so that the least significant weight is $\lesssim 0.01$ (in GeoEas, use the debug option <NumLock> to view the kriging weights for each grid node)
- in practice, start with a large number and reduce to the minimum necessary for stability and to achieve an adequate estimation result (no or few null estimates)

c. The Closest Neighborhood Data Points
- when more data occur within the specified search neighborhood, only the "closest" n_{max} data points are used for the kriging estimation at that grid node
- the method of determining the "closest" points can be either a geometric distance (same as in inverse-distance methods) or a variogram "distance"
- the variogram method is not actually a distance, but the statistical similarity between two points as estimated by the variogram value corresponding to the pair's spatial separation; eg: for strong spatial anisotropy (geometric or zonal), or for a hole-effect correlation function (not permissable in GeoEas), the statistical similarity in different directions may be quite different, in which case,

the variogram "distance" is preferrable; in all other cases, the geometric or Euclidean distance is adequate (note: GSLIB uses only variogram distance, no choice is available)

3. Useful Rules of Thumb:
- point kriging exactly reproduces data values if they fall on grid nodes
- otherwise, the kriged values honor the data; also, in block kriging, the data values are not reproduced exactly but are honored
- greater nugget value causes more smoothing in the kriged surface
- coarser grid spacing causes more smoothing in the kriged surface
- start with the most conservative search strategy (circular search, single sector, large numbers of max/min neighbors, block discretization with DECLUS grid spacing, and search radius equal to maximum variogram range)
- fine-tune parameters in priority: point kriging (if necessary), reduced numbers of max/min neighbors, use search ellipse with axes set to variogram ranges/direction, try smaller ellipse axes

4. Impacts of Correlation Function Parameters on Kriging (Isaaks & Srivastava, p.301, ff.)
- note that all of the following have varying degrees of effect on kriged estimates depending on data distribution, spatial correlation, etc. and, other than the effect described below, the effects are case-specific :
    - variogram sill height (no effect on estimates and weights, only on kriging variance)
    - variogram shape (Gaussian model gives greater weight to nearest data, more pronounced "screening" leading to negative kriging weights)
    - nugget (a large nugget produces estimates more like a simple averaging of local data)
    - variogram range (neighboring data appear statistically closer when variogram range is large; very small range = large statistical distance, similar to effect of a large nugget)
    - anisotropy (range and/or sill differs with direction, hence statistical "distance" differs)

5. Cross-Validation (Isaaks and Srivastava, Ch.15)
- although every solution of the kriging system of equations produces a "best", linear, unbiased estimate, the criterion of "best" is only as good as how accurately the experimental variogram defines the R.F's spatial correlation structure, how well the variogram model fits the R.F., what type of kriging is performed and how good our kriging search strategy proves to be
- there are many sources of potential variation (and error) in calculating and modeling variograms and in obtaining kriging estimates with a variety of search strategies; therefore, a methodology exists to compare the results of different variogram models and kriging strategies to identify the approaches that best reproduce the existing data
- this process is known as cross-validation, in which data are dropped one at a time and compared with an estimate at the same location (single-sample comparison, with replacement); or
- jack-knifing, in which several data points are dropped at a time (multiple-sample comparison, without replacement)

- applications of these techniques are used:
    - to compare different estimation techniques to determine the best approach
    - to compare different nearest neighbor search strategies

- to decide on the "best" variogram function model
- as another *a posteriori* tool of exploratory data analysis (eg: Hohn, p.125-132), together with variogram outlier analysis, and reproduction of global statistics by the kriged estimates

- estimation errors (residuals) can be evaluated simply as a univariate statistical distribution (non-spatial information ignored), or the spatial arrangement of residuals can be mapped to provide additional information on potential estimation problems and improvements (ie. another EDA tool)

- in other applications, such as classification of a site for contaminant cleanup, minimization of the cross-validation residuals may be the most important objective (eg: to minimize the proportion of soil that is misclassified as contaminated or non-contaminated), rather than obtaining the "best" estimates of the level of contamination

- cross-validation is not a "goodness-of-fit" tool; it should only be used to choose the "best" kriging approach in clear recognition of the following caveats:
  - cross-validation cannot help choose between models of different sill (sill does not influence kriging weights) or different relative nugget (undefined due to lack of sample data at zero spacing)
  - sparse and clustered data may be unrepresentative; a model that best reproduces such data may not yield good predictions at unsampled locations
  - cross-validation cannot identify the weakness of an estimation model: whether it is due to the decision of stationarity, the variogram model, or the kriging strategy

6. Problem Set **IX. Kriging Cross-Validation, Introduction to GSLIB's Kriging**
- using cross-validation, compare the two kriged estimates derived in the kriging problem to decide on the best variogram model / kriged estimate of the two