4. Variograms

The covariogram and its normalized form, the correlogram, are by far the most intuitive methods for summarizing the structure of spatial dependencies in a covariance stationary process. However, from an estimation viewpoint such functions present certain difficulties (as will be discussed further in Section 4.10 below). Hence it is convenient to introduce a closely related function known as the *variogram*, which is widely used for estimation purposes.

4.1 Expected Squared Differences

To motivate the notion of a variogram for a covariance stationary process, $\{Y(s): s \in R\}$, we begin by considering any pair of component variables, $Y_s = Y(s)$ and $Y_v = Y(v)$, and computing their *expected squared difference*:

(4.1.1) $$E[(Y_s - Y_v)^2] = E[Y_s^2 - 2Y_sY_v + Y_v^2] = E(Y_s^2) - 2E(Y_sY_v) + E(Y_v^2)$$

To relate this to covariograms, note that if $\|s - v\| = h$, then by (3.2.3) and (3.2.4),

(4.1.2) $$C(h) = \text{cov}(Y_s, Y_v) = E[(Y_s - \mu)(Y_v - \mu)] = E[Y_sY_v - Y_s\mu - \mu Y_v + \mu^2]$$

$$= E(Y_sY_v) - E(Y_s)\mu - \mu E(Y_v) + \mu^2$$

$$= E(Y_sY_v) - \mu^2 - \mu^2 + \mu^2 = E(Y_sY_v) - \mu^2$$

$$\Rightarrow \; E(Y_sY_v) = C(h) + \mu^2$$

Exactly the same argument with $s = v$ shows that

(4.1.3) $$E(Y_s^2) = C(0) + \mu^2 = E(Y_v^2)$$

Hence by substituting (4.1.2) and (4.1.3) into (4.1.1) we see that expected squared differences for all $s, v \in R$ with $\|s - v\| = h$ can be expressed entirely in terms of the covariogram, C, as

(4.1.4) $$E[(Y_s - Y_v)^2] = 2 \cdot [C(0) - C(h)]$$

To obtain a slightly simpler relation, it is convenient to suppress the factor "2" by defining the associated quantity,

(4.1.5) $$\boxed{\gamma(h) = \tfrac{1}{2} E[(Y_s - Y_v)^2] \quad , \quad \|s - v\| = h}$$

and observing from (4.1.4) that with this definition we obtain the following simple identity for all distances , h :

(4.1.6) $$\boxed{\gamma(h) = C(0) - C(h) = \sigma^2 - C(h)}$$

From (4.1.6) it is thus evident that the "scaled" expected squared differences in (4.1.5) define a unique function of distance which is intimately related to the covariogram. For any given covariance stationary process, this function is designated as the *variogram*, γ, of the process. Moreover, it is also evident that this variogram is uniquely constructible from the covariogram. But the converse is not true. In particular since (4.1.6) also implies that

(4.1.7) $$\boxed{C(h) = \sigma^2 - \gamma(h)}$$

it is clear that in addition to the variogram, γ, one must also know the variance, σ^2, in order to construct the covariogram.[1] Hence this variance will become an important parameter to be estimated in all models of variograms developed below.

Before proceeded further with our analysis of variograms it is important to stress that the above terminology is not completely standard. In particular, the expected squared difference function in (4.1.4) is often designated as the "variogram" of the process, and its scaled version in (4.1.5) is called the "semivariogram" [as for example in Cressie (1993, p.58-59) and Gotway and Waller (2004, p.274)]. (This same convention is used in the *Geostatistical Analyst* extension in ARCMAP.) But since the scaled version in (4.1.5) is the *only* form used in practice [because of the simple identity in (4.1.7)] it seems most natural to use the simple term "variogram" for this function, as for example in [BG, p.162].[2]

4.2 The Standard Model of Spatial Dependence

To illustrate the relation in (4.1.7) it is most convenient to begin with the simplest and most commonly employed model of spatial dependence. Recall from the Ocean Depth Example in Section 3.3.1 above, that the basic hypothesis there was that nearby locations tend to experience similar concentration levels of plankton, while those in more widely separated locations have little to do with each other. This can be formalized most easily in terms of correlograms by simply postulating that correlations are high (close to unity) for small distances, and fall monotonely to zero as distance increases. This same general hypothesis applies to a wide range of spatial phenomena, and shall be referred to here as the *standard model* of spatial dependence. Given the relation between correlograms and covariograms in (3.3.13), it follows at once that covariograms for the standard model, i.e., *standard covariograms*, must fall monotonely from $C(0) = \sigma^2$ toward zero, as illustrated

[1] However, assuming that $\lim_{h\to\infty} C(h) = 0$, it follows from (4.1.6) that $\lim_{h\to\infty} \gamma(h) = \sigma^2$. So σ^2 is in principle obtainable from γ as the asymptote (sill) in Figure 4.2 below.

[2] See also the "lament" regarding this terminology in Schabenberger and Gotway (2005, p.135).

in Figure 4.1 below. The right end of this curve has intentionally been left rather vague. It may reach zero at some point, in which case covariances will be exactly zero at all greater distances. On the other hand, this curve may approach zero only asymptotically, so that covariance is positive at all distances but becomes arbitrarily small. Both cases are considered to be possible under the standard model (as will be illustrated in Section 4.6 below by the "spherical" and "exponential" variogram models).

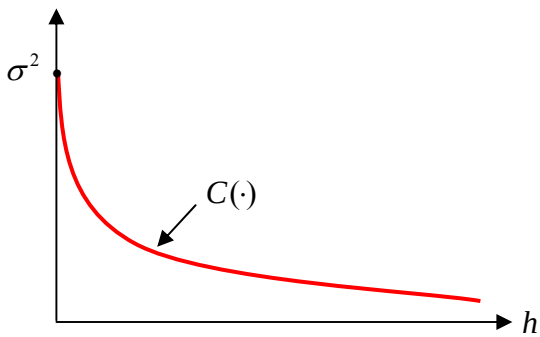


Figure 4.1. Standard Covariogram

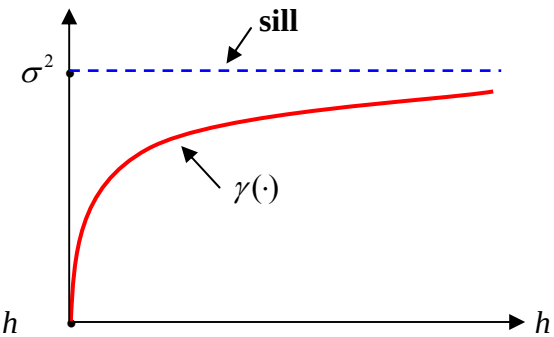


Figure 4.2. Standard Variogram

On the right in Figure 4.2 is the associated *standard variogram*, which by (4.1.6) above must necessarily start at zero and rise monotonely toward the value σ^2. Graphically this implies that the standard variogram must either reach the dashed line in Figure 4.2, designated as the *sill*, or must approach this sill asymptotically.[3]

But while this mathematical correspondence between the standard variogram and covariogram is quite simple, there are subtle differences in their interpretation. The interpretation of standard *covariograms* is straightforward, since decreases in (positive) covariance at large distances are naturally associated with decreases in spatial dependence. But the associated increase in the standard *variogram* is somewhat more difficult to interpret in a simple way. If we recall from (4.1.5) that these variogram values are proportional to *expected squared differences*, then is reasonable to conclude that such differences should increase as variables become less similar (i.e., less positively dependent). But as a general rule, it would still appear that the simplest approach to interpreting variogram behavior is to describe this behavior in terms of the corresponding covariogram.

4.3 Non-Standard Spatial Dependence

Since the analysis to follow will focus almost entirely on the standard model, it is of interest to consider one example of a naturally occurring stationary process that exhibits non-standard behavior. As a more micro version of the Ocean Depth Example in Section 3.3.1 above, suppose that one is interested in measuring variations in ocean depth due to wave action on the surface. Figure 4.3 below depicts an idealized measurement scheme

[3] As noted by [BG, p.162] the scaling by ½ in (4.1.5) is precisely to yield a "sill" which is associated with σ^2 rather than $2\sigma^2$.

involving a set of (yellow) corks at locations $\{s_i : i = 1,..,n\}$ that are attached to vertical measuring rods, allowing them to bob up and down in the waves. The set of cork *heights*, $H_i = H(s_i)$, on these n rods at any point of time can be treated as a sample of size n from a spatial stochastic process, $\{H(s) : s \in R\}$, of wave heights defined with respect to some given ocean region, R.

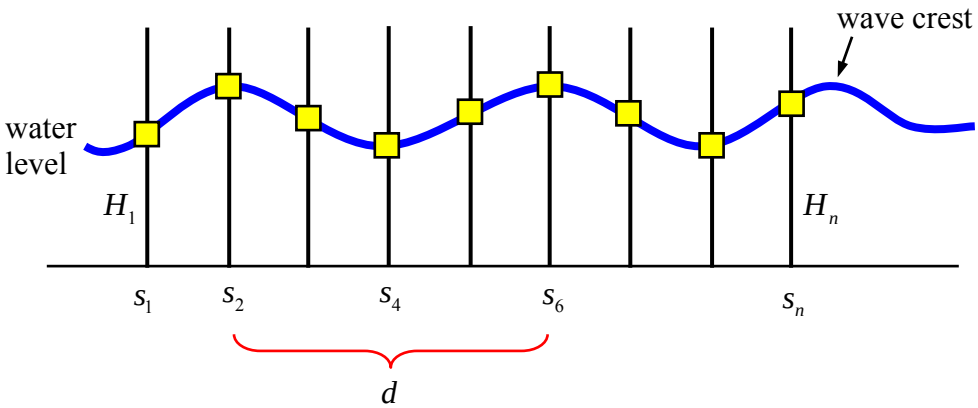


Figure 4.3. Measurement of Wave Heights

Here the fluctuation behavior of corks should be essentially the same over time at each location. Moreover, any dependencies among cork heights due to the smoothness of wave actions should depend only on the spacing between their positions in Figure 4.3. Hence the homogeneity and isotropy assumptions of spatial stationarity in Section 3.3.1 should apply here as well, so that in particular, $\{H(s) : s \in R\}$ can be treated as a covariance stationary process.

But this process has additional structure implied by the natural spacing of waves. If this spacing is denoted by d, then it is clear that for corks separated by distance d, such as those at locations s_2 and s_6 in Figure 4.3, whenever a wave crest (or trough) occurs at one location it will tend to occur at the other as well. Hence pairs of location separated by a distance d should exhibit a *positive correlation* in wave heights, as shown in the covariogram of Figure 4.4 below. However, for locations spaced at around half this distance, such as s_2 and s_4 in Figure 4.3, the opposite should be true: whenever a crest (or trough) occurs at one location, a wave trough (or crest) will tend to occur at the other. Hence the wave heights at such locations can be expected to exhibit *negative correlation*, as is also illustrated by the covariogram in Figure 4.4.

Finally, it should be clear that distances between wave crests are themselves subject to some random variation (so that distance d in Figure 4.3 should be regarded as the *expected distance* between wave crests). Thus, in a manner similar to the standard model, one can expect that wave heights a distant locations will be statistically unrelated. This in turn implies that the positive and negative correlation effects above will gradually

dampen as distance increases. Hence this process should be well represented by the "damped sine wave" covariogram shown in Figure 4.4.[4]

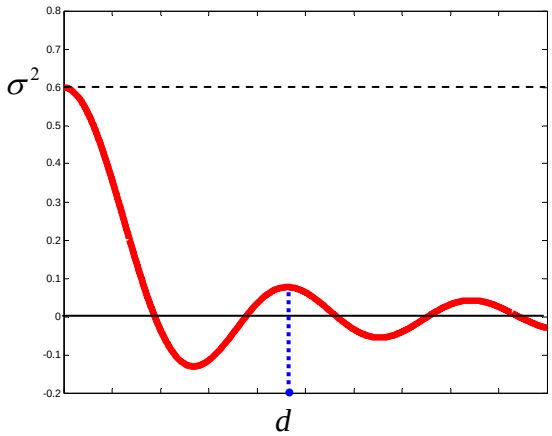


Figure 4.4. Wave Covariogram

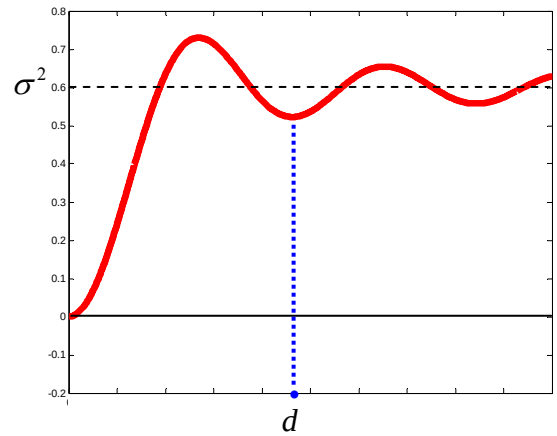


Figure 4.5. Wave Variogram

Finally, the associated *variogram* for this process [as defined by (4.1.6)] is illustrated in Figure 4.5 for sake of comparison. If the variance, σ^2, in Figure 4.4 is again take to define the appropriate *sill* for this variogram (as shown by the horizontal dashed line in Figure 4.5) then it is clear that the values of this variogram now oscillate around the sill rather that approach it monotonely. Hence this sill is only meaningful at larger distances, where wave heights no longer exhibit any significant correlation.

4.4 Pure Spatial Independence

A second example of a covariance stationary process, $\{Y(s): s \in R\}$, which is far more extreme, is the case of *pure spatial independence*, in which distinct random components, $Y(s)$ and $Y(v)$, have no relation to each other – no matter how close they are in space. Mathematically this implies that $\text{cov}[Y(s), Y(v)] = 0$ for all distinct s and v. But since $\text{cov}[Y(s), Y(s)] = \sigma^2 > 0$ for all s, this in turn implies that the covariogram, C, for such a process must exhibit a *discontinuity at the origin*, as shown on the left in Figures 4.6.

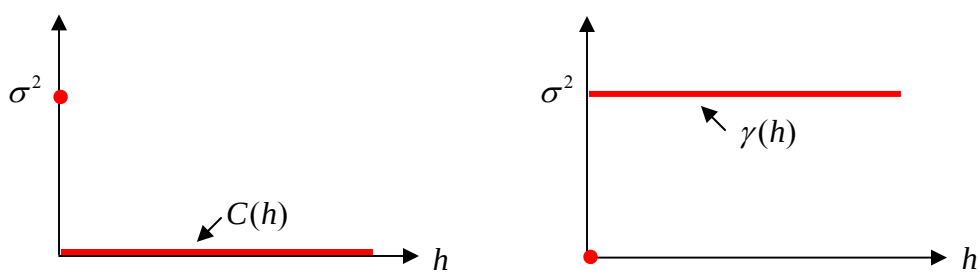


Figure 4.6. Pure Spatial Independence

[4] A mathematical model of this type of covariogram is given in expression 4.6.9 below.

Hence by definition, the corresponding variogram, γ, for pure spatial spatial independence (shown on the right in Figure 4.6) must also exhibit a *discontinuity at the origin*, since $\gamma(0)=0$ and $\gamma(h)=\sigma^2>0$ for all $h>0$.

Such processes are of course only mathematical idealizations, since literally all physical processes must exhibit some degree of smoothness (even at small scales). But if independence holds at least approximately at sufficiently small scales then this idealization may be reasonable. For example, if one considers a sandy desert region, R, and lets $D(s)$ denote the depth of sand at any location, $s \in R$, then this might well constitute a smooth covariance stationary process, $\{D(s): s \in R\}$, which is quite consistent with the standard model of Section 3.5 (or perhaps even the "wave model" of Section 3.6 if wind effects tend to ripple the sand). But in contrast to this, suppose that one considers an alternative process $\{W(s): s \in R\}$ in which $W(s)$ now denotes the *weight* of the topmost grain of sand at location s (or perhaps the *diameter* or *quartz content* of this grain). Then while is it reasonable to suppose that the distribution of these weights is the same at each location s (and is thus a homogeneous process as in Section 3.3.1 above), there need be little relation whatsoever between the specific weights of adjacent grains of sand. So at this scale, the process $\{W(s): s \in R\}$ is well modeled by *pure spatial independence*.

4.5 The Combined Model

The *standard model* in Section 4.2 and the model of *pure spatial independence* in Section 4.4 can be viewed as two extremes: one with continuous positive dependence gradually falling to zero, and the other with zero dependence at all positive distances. However, many actual processes are well represented by a mixture of the two. This can be illustrated by a further refinement of the Ocean Depth Example in Section 3.3.1. Observe that while mobile organisms like zooplankton have some ability to cluster in response to various stimuli, the ocean also contains a host of inert debree (dust particles from the atmosphere, and skeletal remains of organisms, etc.) which bear little relation to each other. Hence in addition to the spatially correlated errors in sonar depth measurements created by zooplankton, there is a general level of "background noise" created by debree particles that is best described in terms of spatially independent errors.

If these two types of measurement errors at location s are denoted repectively by $\varepsilon_1(s)$ and $\varepsilon_2(s)$, then a natural refinement of the depth measurement model in (3.3.1) would be to postulate that total measurement error, $\varepsilon(s)$, is the sum of these two components.

(4.5.1) $$\boxed{\varepsilon(s) = \varepsilon_1(s) + \varepsilon_2(s) \quad , \quad s \in R}$$

Moreover, it is also reasonable to assume that these error components are *independent* (i.e., that the distribution of zooplankton is not influenced by the presence or absence of debree particles). More formally, it may be assumed that $\varepsilon_1(s)$ and $\varepsilon_2(v)$ are independent random variables for every pair of locations, $s, v \in R$. With this assumption it then

follows (see section A2.1 in Appendix A2) that the covariogram, C, of error process ε must be the *sum* of the separate covariograms, C_1 and C_2, for component processes ε_1 and ε_1, i.e., that for any $h \geq 0$,

(4.5.2) $$\boxed{C(h) = C_1(h) + C_2(h)}$$

More generally, any covariance stationary process, $\{Y(s) : s \in R\}$, with covariogram of the form (4.5.2) will be said to satisfy the *combined model* of covariance stationary processes. Covariogram C_1 then represents the *spatially dependent component* of this process, and covariogram C_2 represents its *spatially independent component*.[5]

To see the graphical form of this combined model, observe first that by setting $h = 0$ in (4.5.2) it also follows that

(4.5.3) $$\sigma^2 = C(0) = C_1(0) + C_2(0) = \sigma_1^2 + \sigma_2^2$$

where σ_1^2 and σ_2^2 are the corresponding variances for the spatially dependent and independent components, respectively. Hence the covariogram for the combined process in (4.5.2) is given by Figure 4.7 below:

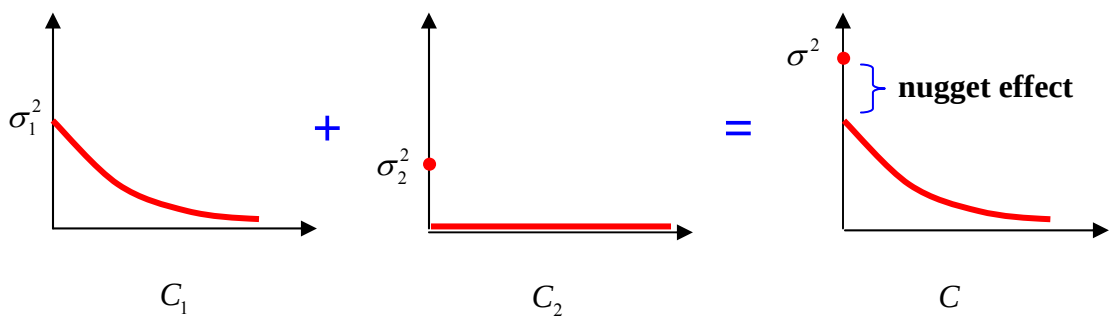


Figure 4.7. Covariogram for Combined Model

In this graphical form it is clear that the covariogram for the combined model is essentially the same as that of the standard model, except that there is now a discontinuity at the origin. This local discontinuity is called the *nugget effect* in the combined model,[6] and the magnitude of this effect (which is simply the variance, σ_2^2, of the pure independent component) is called the *nugget*. Note that by definition the ratio, σ_2^2 / σ^2,

[5] This combined model is an instance of the more general decomposition in Cressie (1993, pp.112-113) with C_1 reflecting the "smooth" component, W, and C_2 reflecting the "noise" components, $\eta + \in$.

[6] This term originally arose in mining applications where there are often microscale variations in ore deposits due to the presence of occasional nuggets of ore [as discussed in more detail by Cressie (1993,p.59)]. In the present context, such a "nugget effect" would be modeled as an independent micro component of a larger (covariance stationary) process describing ore deposits.

gives the relative magnitude of this effect, and is designated as the *relative nugget effect*. For example, if the relative nugget effect for a given covariogram is say .75, then this would indicate that the underlying process exhibits relatively little spatial dependence.

Next we consider the associated variogram for the combined model. If γ denotes the variogram of the combined process in (4.5.1) then we see from (4.1.6) together with (4.5.2) and (4.5.3) that

$$\text{(4.5.4)} \qquad \gamma(h) = \sigma^2 - C(h) = (\sigma_1^2 + \sigma_2^2) - [C_1(h) + C_2(h)]$$

$$= [\sigma_1^2 - C_1(h)] + [\sigma_2^2 - C_2(h)]$$

$$\Rightarrow \boxed{\gamma(h) = \gamma_1(h) + \gamma_2(h)}$$

where γ_1 and γ_2 are the variograms for the spatially dependent and independent components, respectively. Hence it follows that variograms add as well, and yield a corresponding combined variogram as shown in Figure 4.8 below:

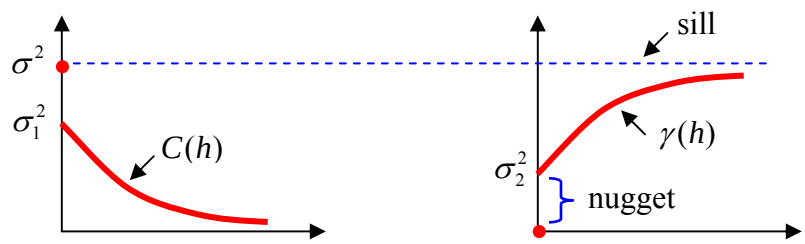


Figure 4.8. Summary of the Combined Model

4.6 Explicit Models of Variograms

While the combined model above provides a useful conceptual framework for variograms and covariograms, it is not sufficiently explicit to be estimated statistically. We require explicit mathematical models that are (i) qualitatively consistent with the combined model, and (ii) are specified in terms of a small number of parameters that can be estimated.[7]

[7] There is an additional technical requirement covariograms yield well-defined covariance matrices, as detailed further in the Appendix to Part III (Corollary 2.p.A3-70).

4.6.1. The Spherical Model

The simplest and most widely used variogram model is the *spherical variogram*, defined for all $h \geq 0$ by:

$$\gamma(h;r,s,a) = \begin{cases} 0 & , \quad h = 0 \\ a + (s-a)\left(\dfrac{3h}{2r} - \dfrac{h^3}{2r^3}\right) & , \quad 0 < h \leq r \\ s & , \quad h > r \end{cases} \tag{4.6.1}$$

Here parameters (r,s,a) of γ are assumed to satisfy $r,s > 0, a \geq 0$ with $s \geq a$. [Note that the argument, h, of function γ is separated from it parameters, (r,s,a), by a semicolon[8]] To interpret these parameters, it is useful to consider the spherical variogram shown in Figure 4.9 below with $(r = 6$, $s = 4$, $a = 1)$:

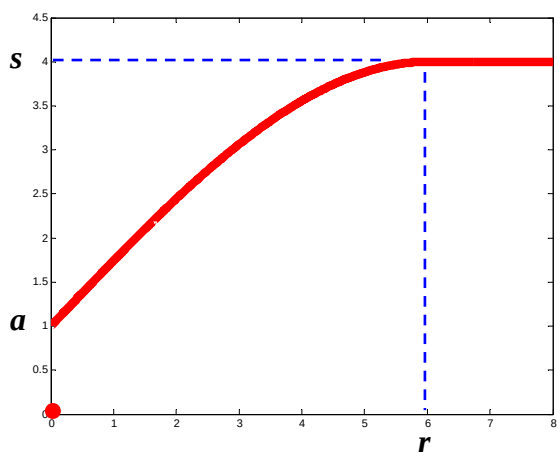


Figure 4.9. Spherical Variogram

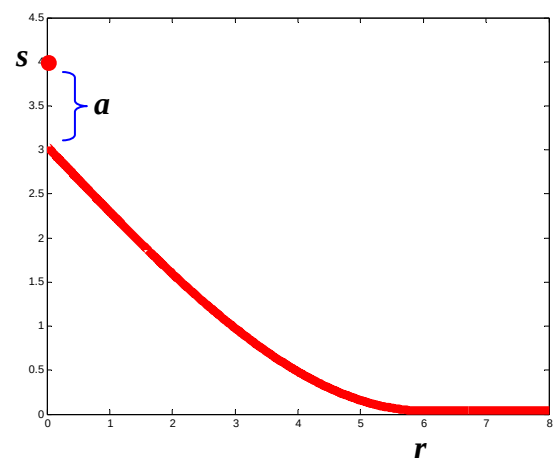


Figure 4.10. Spherical Covariogram

A comparison of Figure 4.9 with the right hand side of Figure 4.8 shows that parameter, s, corresponds to the *sill* of the variogram and parameter, a, corresponds to the *nugget* [as can also be seen by letting h approach zero in the (4.6.1)]. So for this particular example the *relative nugget effect* is $a/s = 1/4$. Note finally that since the spherical variogram reaches the sill at value, r [as can also be seen by setting $h = r$ in (4.6.1)], this implies that the corresponding covariogram in Figure 4.10 *falls to zero* at r. Hence the parameter, r, denotes the maximum range of positive spatial dependencies, and is

[8] More generally the expression, $f(x_1,..,x_n;\theta_1,..,\theta_k)$, is taken to denote a function, f, with arguments $(x_1,..,x_n)$ and parameters $(\theta_1,..,\theta_k)$.

designated simply as the *range* of the variogram (and corresponding covariogram). These same notational conventions for *range*, *sill* and *nugget* will be used throughout.[9]

The formal *spherical covariogram* corresponding to expression (4.6.1) is immediately obtainable from (4.1.7) [with $s = \sigma^2$], and is given by:

(4.6.2)
$$\boxed{C(h\,;r,s,a) = \begin{cases} s & , \quad h = 0 \\ (s-a)\left(1 - \dfrac{3h}{2r} + \dfrac{h^3}{2r^3}\right) & , \quad 0 < h \le r \\ 0 & , \quad h > r \end{cases}}$$

Together, (4.6.1) and (4.6.2) will be called the *spherical model*. One can gain further insight into the nature of this model by differentiating (4.6.2) in the interval, $0 < h \le r$, to obtain:

(4.6.3)
$$\frac{dC}{dh} = (s-a)\left(\frac{3h^2}{2r^3} - \frac{3}{2r}\right) = (s-a)\left(\frac{3}{2r}\right)\left(\frac{h^2}{r^2} - 1\right)$$

Hence we see that

(4.6.3)
$$\frac{dC}{dh} = 0 \;\Leftrightarrow\; h = r$$

Moreover, by differentiating once more we see that

(4.6.4)
$$\frac{d^2C}{dh^2} = (s-a)\left(\frac{3}{2r}\right)\left(\frac{2h}{r^2}\right) > 0$$

whenever the sill is greater than the nugget (i.e., $s - a > 0$). Thus, except for the extreme case of pure independence, this function is always "bowl shaped" on the interval $0 < h \le r$, and has a unique differentiable minimum at $h = r$. Hence this spherical covariogram yields a combined-model form with finite range that falls smoothly to zero. These properties (together with its mathematical simplicity) account for the popularity of the spherical model.

All explicit variogram applications in these notes will employ this *spherical model*. However, it is of interest at this point to consider one alternative model which is also in wide use.

[9] Note that the use of "*s*" to denote *sill* should not be confused with the use of "$s = (s_1, s_2)$" to denote spatial locations. Also, since the symbol, *n*, is used to denote *sample size,* we choose to denote the nugget by "*a*" rather than "*n*".

4.6.2 The Exponential Model

While the spherical model is smooth in the sense of continuous differentiability, it makes the implicit assumption that correlations are exactly zero at all sufficiently large distances. But in some cases it may be more appropriate to assume that while correlations may become arbitrarily small at large distances, they never vanish. The simplest model with this property is the *exponential variogram*, defined for all $h \geq 0$ by,

(4.6.5)
$$\boxed{\gamma(h\,;r,s,a) = \begin{cases} 0 & , \quad h = 0 \\ a + (s-a)\left(1 - e^{-3h/r}\right) & , \quad h > 0 \end{cases}}$$

with corresponding *exponential covariogram*, defined for all $h \geq 0$ by,

(4.6.6)
$$\boxed{C(h\,;r,s,a) = \begin{cases} s & , \quad h = 0 \\ (s-a)e^{-3h/r} & , \quad h > 0 \end{cases}}$$

Together, this variogram-covariogram pair is designated as the *exponential model*, and is illustrated in Figures 4.11 and 4.12 below, using the same set of parameter values $(r = 6\,,\ s = 4\,,\ a = 1)$ as for the spherical model above.

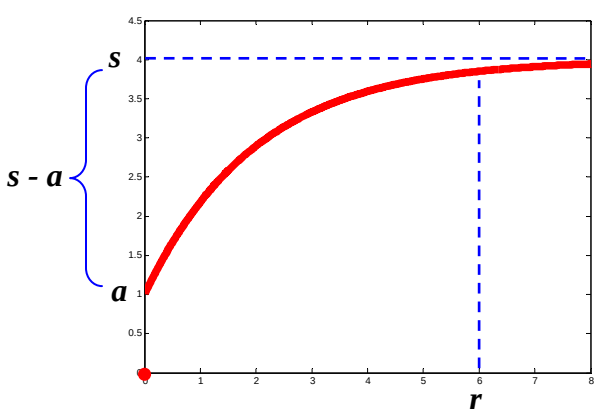


Figure 4.11. Exponential Variogram

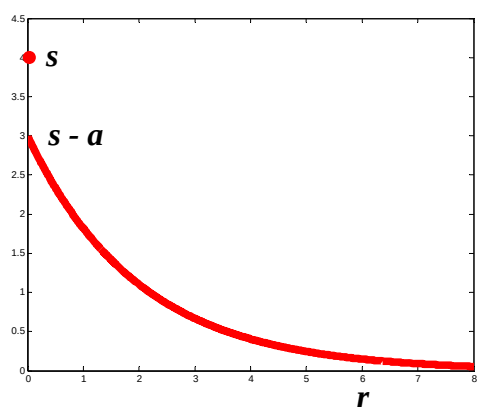


Figure 4.12. Exponential Covariogram

Here it is clear that the sill, s, and nugget, a, play the same role as in the spherical model. However, the "range" parameter, r, is more difficult to interpret in this case since spatial dependencies *never fall to zero*. To motivate the interpretation of this parameter,

observe first that since spatial dependencies are only meaningful at positive distances, it is natural to regard the quantity $s-a$ in Figure 4.12 as the *maximal covariance* for the underlying process.[10] In these terms, the *practical range* of spatial dependency is typically defined to be the smallest distance, r, beyond which covariances are no more than 5% of the maximal covariance. To see that r in (4.6.6) in indeed the practical range for this covariogram, observe simply that since $e^{-x} = .05 \Leftrightarrow x = -\ln(.05) = 2.9957 \approx 3$, it follows that

$$h = r \Rightarrow e^{-3h/r} \approx .05 \Rightarrow \gamma(r) \approx (s-a)(.05) \tag{4.6.7}$$

Note finally that in terms of the corresponding variogram (which plays the primary role in statistical estimation of the exponential model), the quantity $s-a$ in Figure 4.11 is usually called the *partial sill*.[11]

4.6.3 The Wave Model

Finally, it is of interest to consider a mathematical model of the nonstandard "wave" dependence example in Section 4.3 above. Here it is not surprising that the appropriate variogram for this *wave model* is given by a damped sin wave as follows,[12]

$$\boxed{\gamma(h;r,s,a) = \begin{cases} 0 & , \ h=0 \\ a+(s-a)\left(1-w\dfrac{\sin(h/w)}{h}\right) & , \ h>0 \end{cases}} \tag{4.6.8}$$

where the parameter, w, denotes the *wave intensity*. Here the corresponding covariogram is given by:

$$\boxed{C(h;r,s,a) = \begin{cases} s & , \ h=0 \\ (s-a)\left(w\dfrac{\sin(h/w)}{h}\right) & , \ h>0 \end{cases}} \tag{4.6.9}$$

The wave covariogram and variograms shown in Figures 4.4 and 4.5 above are in fact the instances of this wave model with $(w = 0.6, a = 0, s = 0.6)$.

---

[10] More generally, this maximal covariance for any combined model in Figure 4.7 is seen to be given by the variance, σ_1^2, of the (continuous) spatially dependent component.

[11] Indeed this quantity plays such a central role that variograms are often defined with the partial sill as an explicit parameter rather than the sill itself. See for example the spherical and exponential (semi) variogram models in Cressie (1993, p.61). See also the Geostatistical Analyst example in Section 4.9.2 below.

[12] This is also referred to as the *hole-effect model* [as in Cressie (1993, p.623)], and in particular, is given this designation in the *Geostatistical Analyst* kriging option of ARCMAP.

4.7 Fitting Variogram Models to Data

There are many approaches to fitting possible variogram models to spatial data sets, as discussed at length in Cressie (1993, section 2.4) and Schabenberger and Gotway (2004, sections 4.4-4.6). Here we consider only the standard two-stage approach most commonly used in practice (as for example in *Geostatistical Analyst*). The basic idea of this approach is to begin by constructing a direct model-independent estimate of the variogram called the "empirical variogram". This empirical variogram is then used as intermediate data to fit specific variogram models. We consider each of these steps in turn.

4.7.1 Empirical Variograms

An examination of (4.1.5) suggests that for any given set of spatial data $\left(y(s_i): i = 1,..,n\right)$ and distance, h, there is an obvious estimator of the variogram value, $\gamma(h)$, namely "half the average value of $\left(y(s_i) - y(s_j)\right)^2$ for all pairs of locations s_i and s_j separated by distance h". However, one problem with this estimator is that (unlike K-functions) the value $\gamma(h)$ refers to point pairs with distance $\|s_i - s_j\|$ *exactly* equal to h. Since in any finite sample there will generally be at most one pair that are separated by a given distance h (except for data points on regular grids, as discussed below), one must necessarily *aggregate* point pairs (s_i, s_j) with similar distances and hence estimate $\gamma(h)$ at only a small number of representative distances for each aggregate. The simplest way to do so is to partition distances into intervals, called *bins*, and take the average distance, h_k, in each bin k to be the appropriate representative distances, called *lag distances*, as shown in Figure 4.13 below:

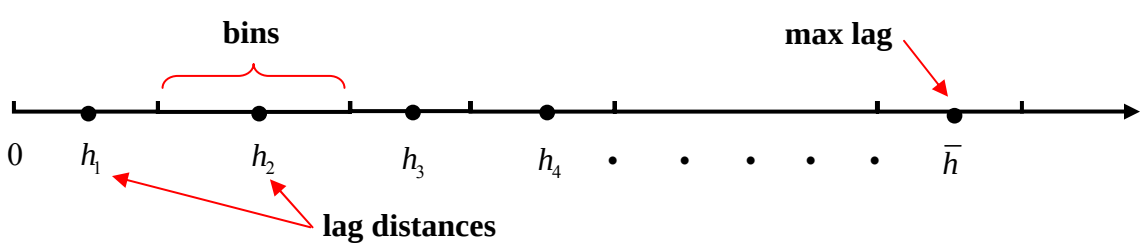


Figure 4.13. Lag Distances and Bins

More formally, if N_k denotes the *set of distance pairs*, (s_i, s_j), in bin k, [with the *size* (number of pairs) in N_k denoted by $|N_k|$], and if the distance between each such pair is denoted by $h_{ij} = \|s_i - s_j\|$, then the *lag distance*, h_k, for bin k is defined to be

(4.7.1) $$h_k = \frac{1}{|N_k|}\sum_{(s_i,s_j)\in N_k} h_{ij}$$

To determine the size of each bin, the most common approach is to make all bins the same size, in order to insure a uniform approximation of lag distances within each bin. However there is an implicit tradeoff here between approximation of lag distances and the number of point pairs used to estimate the variogram at each lag distance. Here the standard *rule of thumb* is that each bin should contain at least 30 point pairs,[13] i.e., that

(4.7.2) $$|N_k| \geq 30$$

Next observe that the choice of the *maximum lag distance* (*max-lag*), $\bar{h}$, (in Figure 4.13) also involves some implicit restrictions. First, for any given set of sample points, $\{s_i : i = 1,..,n\} \subset R$, one cannot consider lag distances greater than the *maximum pairwise distance*,

(4.7.3) $$h_{\max} = \max\left\{\left\|s_i - s_j\right\| : i < j \leq n\right\}$$

in this sample since no observations are available. Moreover, practical experience has shown that even for lag distances close to $h_{\max}$ the resulting variogram estimates tend to be unstable [Cressie (1985, p.575)]. Hence, in a manner completely analogous to the rule of thumb for K-functions [expression (4.5.1) of Part I], it is common practice to restrict $\bar{h}$ to be no greater than half of $h_{\max}$, i.e.,

(4.7.4) $$\bar{h} \leq \frac{h_{\max}}{2}$$

Hence our basic rule for constructing bins is choose a system of bins $\{N_k : k = 1,..,\bar{k}\}$ of *uniform size*, such that the max-lag, $\bar{h} = h_{\bar{k}}$, is as large as possible subject to (4.7.3) and (4.7.4). More formally, if the biggest distance in each bin k is denoted by $d_k = \max_{(s_i,s_j)\in N_k} d_{ij}$, then our procedure (in the MATLAB program **variogram.m** discussed below) is to choose a *maximum bin number*, $\bar{k}$, and *maximum distance* (*max-dist*), $\bar{d}$, such that [14]

(4.7.5) $$|N_1| = \cdots = |N_{\bar{k}}| \geq 30$$

(4.7.6) $$\bar{h} = h_{\bar{k}} < d_{\bar{k}} = \bar{d}$$

[13] Notice that this rule of thumb is reminiscent of that for the Central Limit Theorem used in the Clark-Evans test of Section 3.2.2 in Part I (and in Section 3.1.3 above). Note also that some authors recommend there be at least 50 pairs in each bin [as for example in Schabenberger and Gotway (2005, p.153)].
[14] This is essentially a variation on the "practical rule" suggested by Cressie (1985, p.575).

(Here the default value of $\bar{d}$ is $h_{\max}/2$ and the default value of $\bar{k}$ is 100 bins.) With these rules for constructing bins and associated lag distances, it then follows from (4.1.5) that for any given set of sample points, $\{s_i : i = 1,..,n\} \subset R$, with associated data, $\{y(s_i) : i = 1,..,n\}$, an appropriate estimate of the variogram value, $\gamma(h_k)$, at each lag distance, $h_k \le \bar{h}$, is given by half the average squared differences $\left(y(s_i) - y(s_j)\right)^2$ over all point pairs (s_i, s_j) in N_k, i.e.,

(4.7.7) $$\boxed{\hat{\gamma}(h_k) = \frac{1}{2|N_k|}\sum_{(s_i,s_j)\in N_k}\left(y(s_i) - y(s_j)\right)^2}$$

This set of estimates at each lag distance is designated as the *empirical variogram*.[15] More formally, if for any given set of (ordered) lag distances, $\{h_k : k = 1,..,\bar{k}\}$, the associated variogram estimates in (4.7.7) are denoted simply by $\hat{\gamma}_k = \hat{\gamma}(h_k)$, then the *empirical variogram* is given by the set of pairs $\{(h_k, \hat{\gamma}_k) : k = 1,..,\bar{k}\}$. An schematic example of this empirical variogram construction is given in Figure 4.14 below:

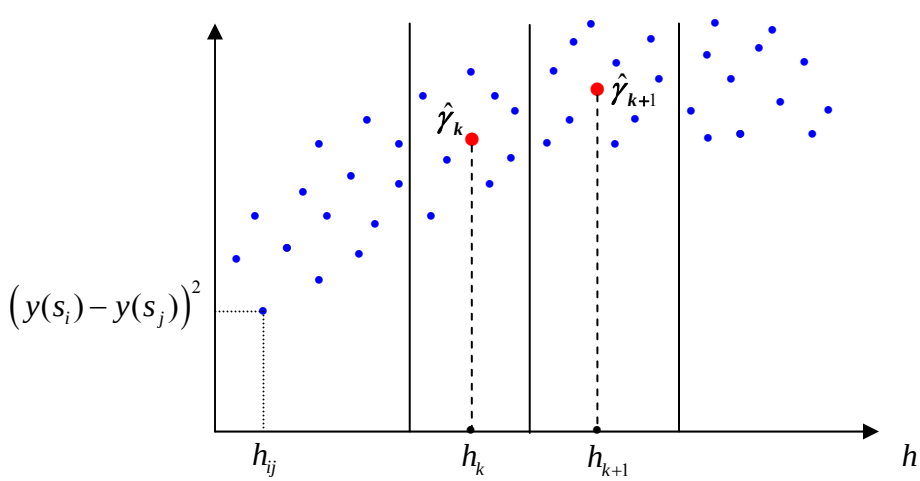


Figure 4.14 Empirical Variogram Construction

Here the *blue dots* correspond to squared-difference pairs, $\left(y(s_i) - y(s_j)\right)^2$, plotted against distances, $h_{ij} = \|s_i - s_j\|$, for each point pair, (s_i, s_j), [as illustrated for one point in the lower left corner of the figure]. The vertical lines separate the bins, as shown for bins

[15] The empirical variogram is also known as *Matheron's estimator*, in honor of its originator [Schabenberger and Gotway (2005, Section 4.4.1)].

k and k+1. So in bin k, for example, there is one blue dot for every point pair, $(s_i, s_j) \in N_k$. The *red dot* in the middle of these points denotes the pair of average values, $(h_k, \hat{\gamma}_k)$, representing all points in that bin. Hence the empirical variogram consists of all these average points, one for each bin of points. [Schematics of such empirical variograms are shown (as *blue dots*) in Figure 4.15 below. An actual example of an empirical variogram is shown in Figure 4.19 below.]

While this empirical variogram will be used to fit all variograms in these notes, it should be mentioned that a number of modifications are possible. First of all, while the use of average distances, h_k, in each bin k has certain statistical advantages (to be discussed below), one can also use the median distance, or simply the midpoint of the distance range. Similarly, while uniformity of bin sizes in (4.7.5) will also turn out to have certain statistical advantages for fitting variograms in our framework (as discussed below), one can alternatively require uniform widths of bins.

In addition, it has been observed by Cressie and Hawkins (1980) [also Cressie (1993, Section 2.4.3)] that estimates involving squared values such as (4.8.7) are often dominated by a few large values, and are thus sensitive to *outliers*. Hence these authors propose several "robust" alternatives to (4.7.7) based on square roots and median values of absolute differences.

Finally it should be noted that a number of fitting procedures in use actually drop this initial stage altogether, and fit variogram models directly in terms of the original data, $\{y(s_i) : i = 1,..,n\}$.[16] In such approaches, the empirical variogram is essentially replaced by a completely disaggregated version called the *variogram cloud,* where each point pair (s_i, s_j) is treated as a separate "bin", and where $\gamma_{ij} = \gamma\left(\left\|s_i - s_j\right\|\right)$ is estimated by the single sample, $\hat{\gamma}_{ij} = \left(y(s_i) - y(s_j)\right)^2$.[17] While this approach can in many cases be more powerful statistically, it generally requires stronger modeling assumptions. Moreover, it turns out that such methods are not only very sensitive to these modeling assumptions, but can also be less stable for smaller data sets. Finally, and most important from practical viewpoint, plots of the empirical variogram tend to be visually much more informative that plots of the entire variogram cloud, and in particular, can often help to suggest appropriate model forms for the variogram itself. [An example is given in Figure 4.20 below.] Hence we choose to focus on the classical empirical-variogram approach.[18]

[16] Most prominent among these is the method of *maximum likelihood*, as detailed for example in Schabenberger and Gotway (2005, Section 4.5.2). [This general method of estimation will also be developed in more detail in Part III of these notes for fitting spatial regression models.]
[17] An example is given in Figure 4.19 below.
[18] For additional discussion see the section on "Binning versus Not Binning" in Schabenberger and Gotway (2005, Section 4.5.4.3). See also the excellent discussion in Reilly and Gelman (2007).

4.7.2 Least-Squares Fitting Procedure

Given an empirical variogram, $\{(h_k, \hat{\gamma}_k) : k = 1,..,\bar{k}\}$, together with a candidate variogram model, $\gamma(h\,; r, s, a)$ [such as the spherical model in (4.7.1)], the task remaining is to find parameter values, $(\hat{r}, \hat{s}, \hat{a})$, for this model that yield a "best fit" to the empirical variogram data. The simplest and most natural approach is to adopt a "least squares" strategy, i.e., to seek parameter values, $(\hat{r}, \hat{s}, \hat{a})$, that solve the following (nonlinear) *least-squares* problem:

(4.7.8) $$\boxed{\min_{(r,s,a)} \sum_{k=1}^{\bar{k}} \left(\hat{\gamma}_k - \gamma(h_k; r, s, a)\right)^2}$$

While this procedure will be used to fit all variograms in these notes, it is important to note some shortcomings of this approach. First of all, since squared deviations are being used in (4.7.8), it again follows that this least-squares procedure is sensitive to outliers. As with all least-squares procedures, one can attempt to mitigate this problem by using an appropriate weighting scheme, i.e., by considering the more general *weighted least-squares* problem:

(4.7.9) $$\boxed{\min_{(r,s,a)} \sum_{k=1}^{\bar{k}} w_k \left(\hat{\gamma}_k - \gamma(h_k; r, s, a)\right)^2}$$

for some set of appropriate nonnegative weights $\left(w_k : k = 1,..,\bar{k}\right)$. A very popular choice for these weights [first proposed by Cressie (1985)] is to set:[19]

(4.7.10) $$w_k = \frac{|N_k|}{\gamma(h_k; r, s, a)^2} \quad , \quad k = 1,..,\bar{k}$$

Here the numerator simply places more weight on those terms with more samples. The denominator is approximately proportional to the variance of the estimates, $\hat{\gamma}_k$,[20] so that the effect of both the numerator and denominator is to place more weight on those terms for which the estimates, $\hat{\gamma}_k$, are most reliable. However, it has been pointed out by others that the inclusion of the unknown parameters (r, s, a) in these weights can create certain instabilities in the estimation procedure [see for example Zhang et al. (1995) and Müller (1999, Section 4)]. Moreover, since our constant bin sizes in (4.7.5) eliminate variation in the sample weights, we choose to use the simpler unweighted least-squares procedure in (4.7.8).

[19] In particular, this is the weighted least-squares procedure used in *Geostatistical Analyst*.
[20] This approximation is based on the important case of *normally distributed* spatial data.

Finally it should also be noted that this least-square procedure is implicitly a *constrained* minimization problem since it is required that (i) $r \geq 0$ and (ii) $s \geq a \geq 0$. In the present setting, however, nonnegativity of both r and s is essentially guaranteed by the nonnegativity of the empirical variogram itself. But nonnegativity of the nugget, a, is much more problematic, and can in some cases *fail to hold*. This is illustrated by the schematic example shown on the left in Figure 4.15 below, where a spherical variogram model (red curve) has been fitted to a set of hypothetical empirical variogram data (blue dots). Here it is clear that the best fitting spherical variogram does indeed involve a negative value for the estimated nugget, $\hat{a}$.

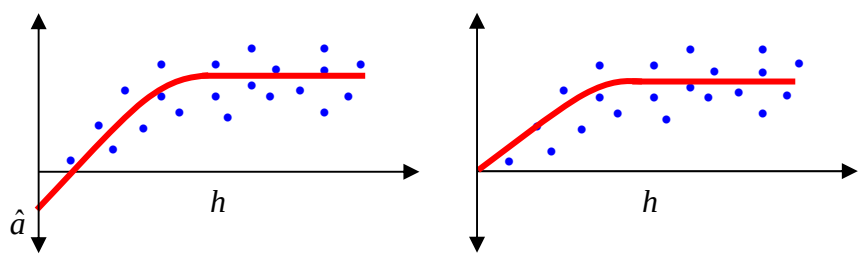


Figure 4.15. Negative Nugget Problem

Hence in such cases, it is natural to impose the additional constraint that $a = 0$, and then solve the reduced minimization problem in the remaining unknown parameters, (r, s):

(4.7.11) $$\boxed{\min_{(r,s)} \sum_{k=1}^{\bar{k}} \left(\hat{\gamma}_k - \gamma(h_k; r, s, 0)\right)^2}$$

The solution to this reduced problem, shown schematically above will yield the "closest approximation" to the solution of (4.8.8) with a feasible value for the nugget, a. It is this two-stage fitting procedure that will be used (implicitly) whenever nuggets are negative.

4.8 The Constant-Mean Model

Our next objective is to develop a practical illustration of variogram estimation. But to do so, it is important to begin by recalling that covariance stationarity was originally motivated in the context of our *general modeling framework* in Section 1.2 above, where it was assumed that spatial random variables are of the form:

(4.8.1) $$Y(s) = \mu(s) + \varepsilon(s) \quad , \quad s \in R$$

and where covariance stationarity is actually a property of the *unobserved* residual process, $\left(\varepsilon(s): s \in R\right)$. Hence variogram estimation for any given set of spatial data, $\left(y(s_i): i = 1,..,n\right)$, must generally been done as part of a larger modeling effort in which *both* the variogram and the spatial trend function $\left(\mu(s): s \in R\right)$ are modeled explicitly. One can then consider iterative fitting procedures in which the spatial trend function is first fitted from the data, say by $\left(\hat{\mu}(s_i): i = 1,..,n\right)$, to yield *residual estimates*,

(4.8.2) $$\hat{\varepsilon}(s_i) = y(s_i) - \hat{\mu}(s_i) \ , \ i = 1,..,n$$

that are in turn used to fit the variogram model. Much of the present section on Continuous Spatial Data Analysis will be devoted to this larger modeling-and-estimation problem. Hence to develop a meaningful example of variogram estimation at this point, it is necessary to make stronger assumptions about the general framework in (4.9.1) above.

In particular, we now assume that the entire process $\left(Y(s): s \in R\right)$ is itself *covariance stationary*. By (3.2.6) through (3.2.8) this equivalent to assuming that in addition to covariance stationarity of the residual process in the second term of (4.8.1), the spatial trend function in the first term is *constant*, so that

(4.8.3) $$Y(s) = \mu + \varepsilon(s) \ , \ s \in R$$

for some (possibly unknown) scalar, μ. Under these conditions it follows at once that

(4.8.4) $$E\left[\left(Y(s) - Y(v)\right)^2\right] = E\left[\left(\mu + \varepsilon(s) - \mu - \varepsilon(v)\right)^2\right] = E\left[\left(\varepsilon(s) - \varepsilon(v)\right)^2\right]$$

for all $s, v \in R$, so that by definition the variograms for the Y-process and the ε-process are *identical*:

(4.8.5) $$\gamma_Y(h) = \gamma_\varepsilon(h) \ , \ h \geq 0$$

Hence, under these assumptions we see that for any given spatial data, $\left(y(s_i): i = 1,..,n\right)$, the residual variogram, γ_ε, can be estimated directly in terms of the empirical variogram,

(4.8.6) $$\hat{\gamma}_Y(h_k) = \frac{1}{2\left|N(h_k)\right|} \sum\nolimits_{(s_i, s_j) \in N(h_k)} \left(y(s_i) - y(s_j)\right)^2 \ , \ k = 1,..,\bar{k}$$

for the *observable* Y-process. This approach will be illustrated in the following example.

4.9 Example: Nickel Deposits on Vancouver Island

The following example is taken from [BG, pp.150-151] and is based on sample data from Vancouver Island in British Columbia collected by the Geological Survey of Canada. This data set [contained in the ARCMAP file (**…\projects\nickel\nickel.mxd**)], extends over the area at the northern tip of the island shown in Figure 4.16 below. The area outlined in red denotes the full extent of the data site. For purposes of this illustration, a smaller set of 436 sample sites was selected, as shown by the dots in Figure 4.17.

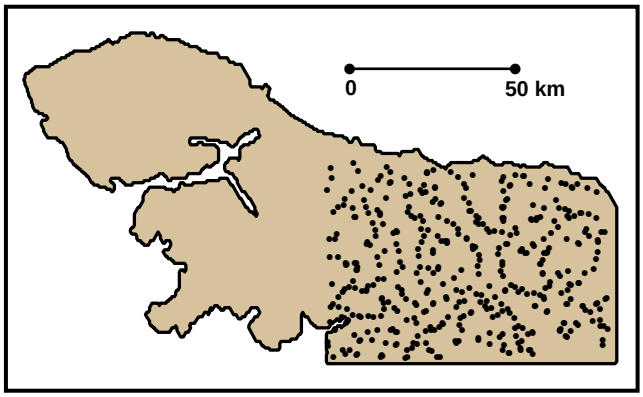


Figure 4.16. Vancouver Sample Area **Figure 4.17. Vancouver Sample Area**

Note the curvilinear patterns of these sample points. As with many geochemical surveys, samples are here taken mainly along stream beds and lake shores, where minerals deposits are more likely to be found. In particular, samples of five different ore types were collected. The present application will focus on deposits of *Nickel* ore. [In class Assignments 3 and 4 you will study deposits of Cobalt and Manganese at slightly different site selections.] This Nickel data is shown in the enlarged map below, where Nickel concentration in water samples is measured in parts per million (ppm).

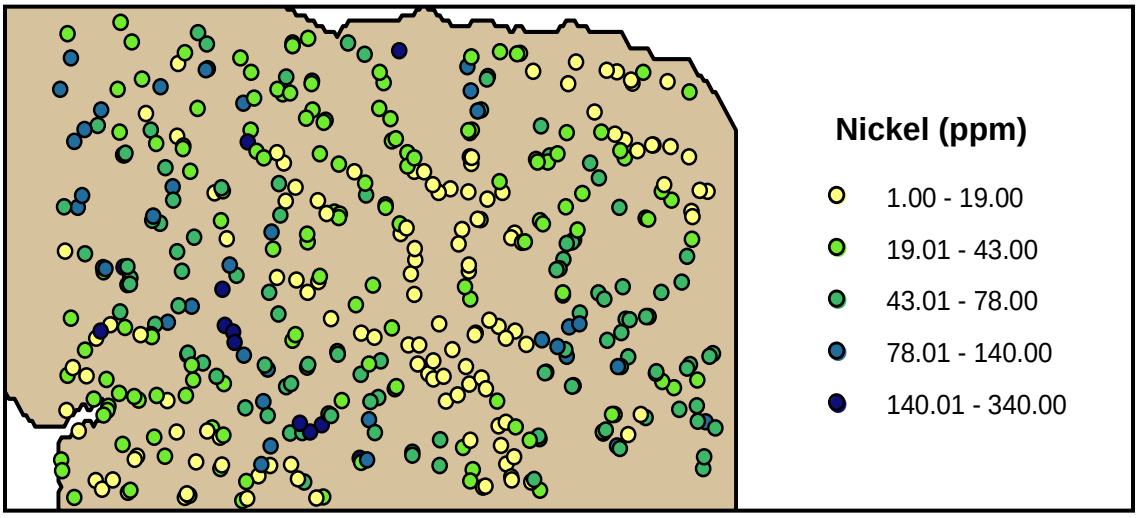


Figure 4.18. Nickel Data

Since the mapped data exhibits strong similarities between neighboring values (at this physical scale), we can expect to find a substantial range of *spatial dependence* in this data. Notice however that the covariance-stationarity assumption of *Isotropy* in (3.3.5) [and (3.3.3)] is much more questionable for this data. Indeed there appear to be diagonal "waves" of high and low values rippling through the site. An examination of Figure 4.16 above shows that these waves are roughly parallel to the Pacific coastline, and would seem to reflect the history of continental drift in this region.[21] Hence our present assumption of covariance stationarity is clearly an over-simplification of this spatial data pattern. We shall see this more clearly in the variogram estimation procedure to follow.

4.9.1 Empirical Variogram Estimation

Given these $n = 436$ sites $(s_i : i = 1,..,n)$ together with their corresponding *nickel measurements*, $y_i = y(s_i)$, our first objective is to construct an empirical variogram for this data as in (4.8.5) above. This procedure is operationalized in the MATLAB program, **variogram_plot.m**. To use this program, the data from **Nickel.mxd** has been imported to the MATLAB workspace file, **nickel.mat**. The 436 x 3 matrix, **nickel**, contains the coordinate + nickel data (s_{i1}, s_{i2}, y_i) for each location $i = 1,..,n$. By opening the program, **variogram_plot.m**, it can be seen that a matrix of this form is the first required input. Next, recall from Section 4.7.1 that along with this data, there are two inputs for defining an appropriate set of distance bins, namely the *maximum bin number*, $\bar{k}$, and the *maximum distance* (*max-dist*), $\bar{d}$. These parameter options are specified in an **opts** structure (similar to that in the program **clust_sim.m** of Section 3.5 in Part I). Here we shall start with the *default* values, $\bar{k} = 100$, and $\bar{d} = h_{\max}/2 = 48{,}203$ meters, so that there is no need to specify this structure. Hence by typing the simple command:

>> **variogram_plot(nickel);**

one obtains a plot of the *empirical variogram*, as shown in Figure 4.19 below.

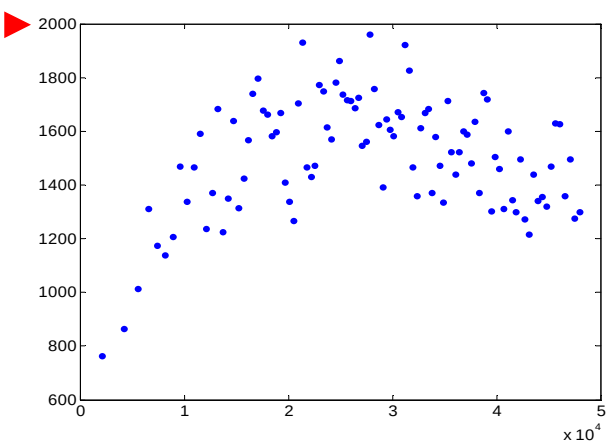


Figure 4.19. Empirical Variogram

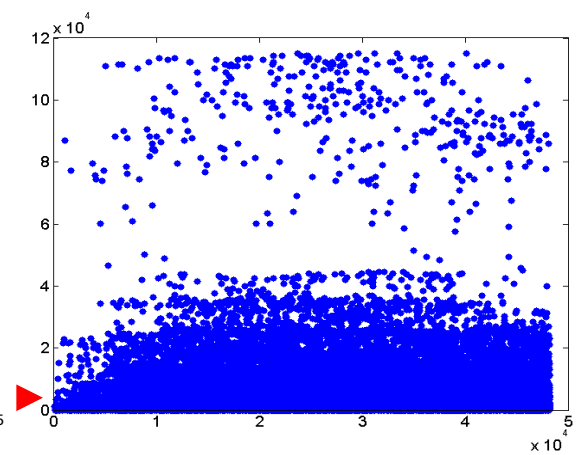


Figure 4.20. Variogram Cloud

[21] In fact these waves are almost mirror images of the Cascadia subduction zone that follows the coastline immediately to the west of Vancouver Island.

Here the point scatter does rise toward a "sill", as in the classical case illustrated in Figure 4.8 above. So it appears that one should obtain a reasonable fit using the spherical model in Figure 4.9 [from expression (4.6.1)]. But before fitting this model, there are a number of additional observations to be made.

First, for purposes of comparison, the corresponding *variogram cloud* is plotted in Figure 4.20.[22] Notice first that while the horizontal (distance) scales of these two figures are the same, the vertical (squared difference) scales are very different. In order to include the full point scatter in the variogram cloud, the maximum squared-difference value has been increased from 2000 in Figure 4.19 to around $120,000\ (=12\times10^4)$ in Figure 4.20. For visual comparison, the value 2000 is shown by a red arrow in both figures. So while the empirical variogram does indeed look "classical" in nature, it is difficult to draw many inferences about the shape of the true variogram from the wider scatter of points exhibited by the variogram cloud. The reason for this is that while the empirical variogram shows *mean* estimates of the variogram at $\bar{k}=100$ selected lag distances, the variogram cloud contains the squared *y*-differences for *each* of the 70,687 individual pairs, (s_i, s_j), with $d_{ij}\le\bar{d}$. Hence about all that can be seen from this "cloud" of points is that there are a considerable number of *outliers* that are very much larger than the mean values at each distance. But fortunately this pattern of outliers is fairly uniform across the distance spectrum, and hence should not seriously bias the final result in this particular case. On the other hand, if outliers were more concentrated in certain distance ranges (as is often typical for the larger distance values), then this might indicate the need to "trim" some of these outliers before proceeding. In short, while the variogram cloud may provide certain useful diagnostic information, the empirical variogram is usually far more informative in terms of the possible shapes of the true variogram.

Next, it should be noted that in addition to the variogram plot, one obtains the following screen output

MAXDIST = 48203.698

which is precisely $\bar{d}$ above. To compare this with the max-lag distance, $\bar{h}$, note first that there are a number of optional outputs for this program as well. First, the actual values of the empirical variogram, $\{(h_k,\hat{\gamma}_k): k=1,..,\bar{k}\}$, are contained in the matrix, **DAT**, where each row contains one $(h_k,\hat{\gamma}_k)$ pair. This can be seen by running the full command,

>> [DAT,maxdist,bin_size,bin_last] = variogram_plot(nickel);

and then clicking on the matrix, **DAT**, in the workspace to display the empirical variogram. In particular, the value $\bar{h}$ corresponds to the last element of the first column and can be obtained with the command [>> **DAT(end,1)**] yielding $\bar{h}=47984$. This is smaller than $\bar{d}$ since $\bar{h}$ is somewhere in the middle of the last bin (as in Figure 4.13 above), and $\bar{d}$ is by definition the *outer edge*, $d_{\bar{k}}$, of this last bin.

[22] This was constructed using the MATLAB program, **variogram_cloud_plot.m**.

As for the additional outputs, **maxdist** is precisely the screen output above, and the value, **bin_size = 707**, tells you how many point pairs there are in each bin [as in condition (4.7.5) above]. In this application there are many more than 30 point pairs in each bin, so that the maximum number of bins, $\bar{k} = 100$, is precisely the number realized. However, if the number of sample points had been sufficiently small, then **bin_size = 30** , would be a binding constraint in (4.7.5) , and there could well be fewer than 100 bins.[23] Finally, the value, **bin_last**, is simply a count of points in the last bin, to check whether it is significantly smaller than the rest. This will only occur if $\bar{d}$ is chosen to be very close to the maximum pairwise distance, $h_{\max}$, and hence will rarely occur in practice.[24]

As one last observation, recall from the "wave" pattern in Figure 4.17 above that one may ask whether this effect is picked up by the empirical variogram at larger distances. By using the measurement tool in ARCMAP and tracing a diagonal line in the direction of these waves (from lower left to upper right), it appears that a reasonable value of maxdist to try is $\bar{d} = 80{,}000$ meters. To do so, we can run the program with this option as follows:

```
>> opts.maxdist = 80000;

>> variogram_plot(nickel,opts);
```

We then obtain the empirical variogram in Figure 4.21b, where the previous variogram has been repeated in Figure 4.21a for ease of comparison:

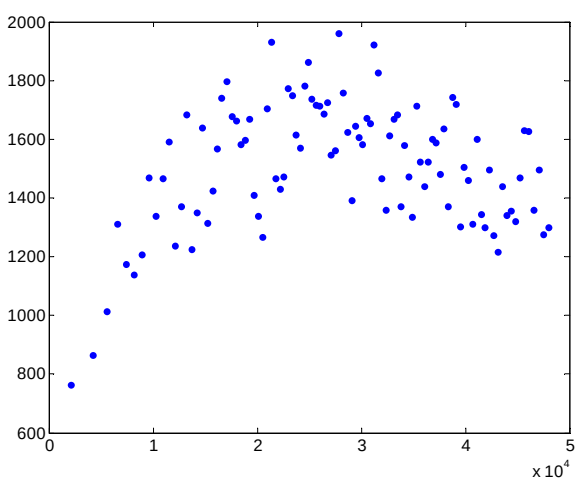


Figure 4.21a. Max Distance = 48,203

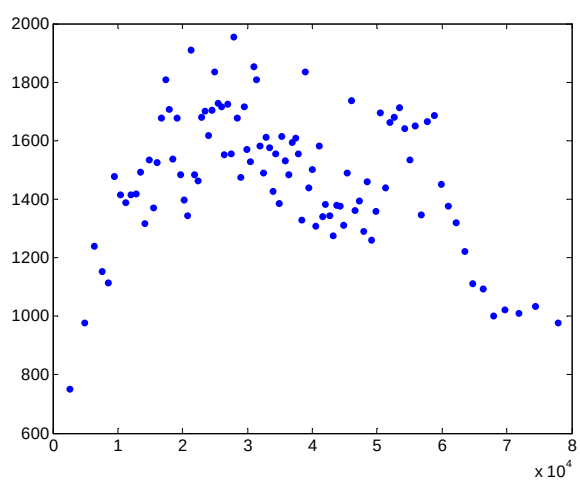


Figure 4.21b. Max Distance = 80,000

[23] For example if $n = 50$ so that the number of distinct point pairs is 50(49)/2 = 12256 < 30(100), then there would surely be fewer than 100 bins.

[24] For example, if one were to set **opts.maxdist = 95000**, which is very close to $h_{\max}$ in the present example, then the last bin will indeed have fewer points than the rest.

Notice that while the vertical (squared difference) scales for these two figures are the same, the horizontal distance scales are now different (reflecting the different maximum distances specified). Moreover, while the segment of Figure 4.21b up to 50,000 $(=5\times10^4)$ meters is qualitatively similar to Figure 4.210a, the bins and corresponding lag distances are not the same as in Figure 4.21a. Hence it is more convenient to show *separate plots* of these two empirical variograms rather than try to superimpose them on the same scale. Given this scale difference, it is nonetheless clear that the slight dip in the empirical variogram on the left, starting at about 40,000 meters, becomes much more pronounced at the larger lag distances shown on the right. Recall (from the corresponding covariograms) that this can be interpreted to mean that pairs of *y*-values (nickel measurements) separated by more than 40,000 meters tend to be more similar (positively correlated) than those separated by slightly smaller distances. Finally, by again using the measurement tool in ARCMAP, it can be seen that the spacing of successive waves is about 40,000 meters. So it does appear that this effect is being reflected in the empirical variogram.

As a final caveat however, it should be emphasized that the most extreme dip in Figure 4.21b occurs at lag distances close to $h_{\max}$, where variogram estimates tend to be very unreliable. In addition, there are "edge effects" created by this rectangular sample region that may add to the unreliability of comparisons at larger distances.

4.9.2 Fitting a Spherical Variogram

Recall from Section 4.6.1 above that all variogram applications in these notes (as well as the class assignments) will involve fitting spherical variogram models to empirical-variogram data. [Other models can easily be fitted using the *Geostatistical Analyst* (GA) extension in ARCMAP, as illustrated below.] For purposes of the present application, we shall adhere to the restriction in (4.7.4) that $\bar{d}$ not exceed $h_{\max}/2$, and hence shall use only the empirical variogram in Figure 4.19 (and 4.21a) constructed under this condition. To fit a spherical variogram model to this empirical-variogram data, we shall use the simple nonlinear least-squares procedure in (4.7.8) above.

Fitting Procedure using MATLAB

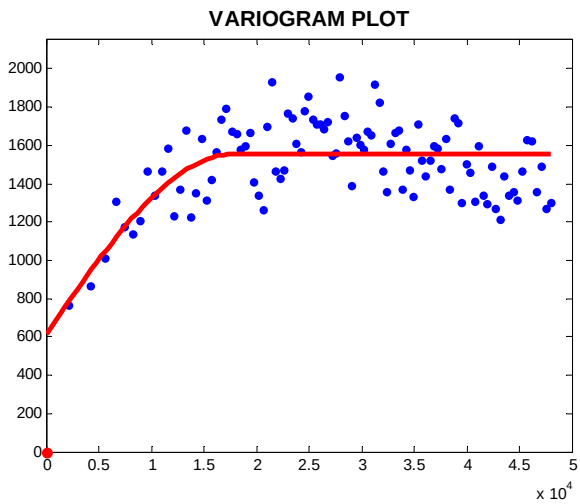
This is operationalized in the MATLAB program, **var_spher_plot.m**.[25] Since this program uses exactly the same inputs as those detailed for **variogram_plot.m** in Section 4.9.1 above, there is no need for further discussion of inputs. Hence a spherical variogram model can be fitted in the present application with the command:

>> **var_spher_plot(nickel);**

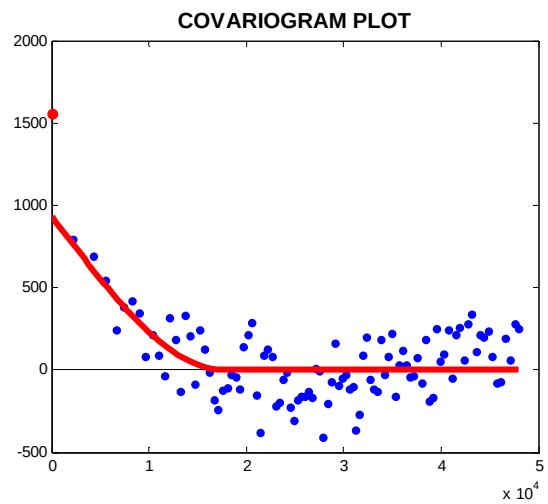
The first output of this fitting procedure is the spherical variogram plot shown in Figure 4.22 below, where the blue dots are the empirical variogram points, and the estimated

[25] One can also use the *weighted nonlinear least-squares* procedure in (4.8.9) and (4.8.10) above, which is programmed in **var_spher_wtd_plot.m**.

spherical variograms is shown in red. If you click **Enter** again you will see the associated covariogram plot, as shown in Figure 4.23 below.



4.22 Fitted Spherical Variogram



4.23 Derived Spherical Covariogram

Here it must be emphasized that this covariogram is *not* being directly estimated. Rather, the estimates $(\hat{r},\hat{s},\hat{a})$ obtained for the spherical variogram are substituted into (4.6.2) in order to obtain the corresponding covariogram. Hence it is more properly designated as the *derived spherical covariogram.* Similarly, the blue dots shown in this figure are simply an inverted reflection of the empirical variogram shown in Figure 4.22. However, they can indeed be similarly interpreted as the *derived empirical covariogram* corresponding to the empirical variogram in Figure 4.22. To do so, recall first from (4.1.7) that for all distances, h, it must be true that $C(h)=\sigma^2-\gamma(h)$. But since each empirical variogram point $(\hat{\gamma}_k,h_k)$ by definition yields an estimate of $\gamma(h_k)$, namely $\hat{\gamma}_k=\hat{\gamma}(h_k)$, and since the sill value, $\hat{s}$, is by definition an estimate of σ^2, i.e., $\hat{s}=\hat{\sigma}^2$, it is natural to use (4.1.7) to estimate the covariogram at distance h_k by

$$\text{(4.9.1)} \qquad \hat{C}(h_k)\equiv\hat{\sigma}^2-\hat{\gamma}(h_k)=\hat{s}-\hat{\gamma}(h_k)$$

Hence by letting $\hat{C}_k=\hat{C}(h_k)$, it follows that the set of points, $\{(h_k,\hat{C}_k):k=1,..,\bar{k}\}$, obtained is precisely the *derived empirical covariogram* in Figure 4.23 corresponding to the empirical variogram, $\{(h_k,\hat{\gamma}_k):k=1,..,\bar{k}\}$, in Figure 4.22.[26]

As mentioned earlier, the advantage of displaying this derived covariogram is that it is much easier to interpret that the estimated variogram. To do so, we begin by noting that in addition to these two diagrams, the program **var_spher_plot.m** also yields a screen

[26] In particular, the vertical component, $\hat{\gamma}_k$, of each variogram point $(h_k,\hat{\gamma}_k)$ has simply been shifted to the new value, $\hat{C}_k=\hat{s}-\hat{\gamma}_k$.

display of the parameter estimates $(\hat{r}, \hat{s}, \hat{a})$ [along with maxdist, $\bar{d}$, and the number of iterations in the optimization procedure[27]], as shown in Figure 4.24 below.

SPHERICAL VARIOGRAM:

RANGE	**17769.160**
SILL	**1554.658**
NUGGET	**618.044**
MAXDIST	**48203.698**

ITERATIONS = 126

Figure 4.24. Parameter Estimates

In particular, the RANGE ($\hat{r} = 17769.160$ *meters*) denotes the distance beyond which there is estimated to be no statistical correlation between nickel values.[28] In Figure 4.22, this corresponds to the distance at which the variogram first "reaches the sill". But this offers little in the way of statistical intuition. In Figure 4.23 on the other hand, it is clear that this is the distance at which *covariance* (*and hence correlation*) *first falls to zero*. This is the key difference between these two representations. Notice also that the vertical axis in Figure 4.23 has been shifted relative to Figure 4.22, in order to depict the *negative* covariance values in the cluster of values around the zero line.

Turning to the other estimated parameters, note first from Figure 4.23 that the SILL ($\hat{s} = 1554.658$) is seen to be precisely the estimated *variance* of individual nickel values (i.e., the estimated covariance at "zero distance"). Similarly, the NUGGET ($\hat{a} = 618.044$) is seen to be that part of the individual variance that not related to spatial dependence among neighbors. Since in this case the *relative nugget effect*, 0.398 (= 618.044/1554.658), is well below 0.5, it is evident that there is a substantial degree of local spatial dependence among nickel values. So in summary, it should be clear that while the variogram model is useful for obtaining these parameter estimates, $(\hat{r}, \hat{s}, \hat{a})$, the derived covariogram model is far more useful for interpreting them.

Fitting Procedure using ARCMAP

Before proceeding, it is of interest to compare this estimated spherical variogram with the fitting procedure used in the *Geostatistical Analyst* (GA) extension in ARCMAP (Version 10). The results of this procedure applied to the nickel data in the ARCMAP file, **nickel.mxd**, are shown in Figure 4.25 below.

[27] Note that if ITERATIONS exceeds 600, you will get an error message telling you that the algorithm failed to converge in 600 iterations (which is the default maximum number of iterations allowed).
[28] Notice also that this RANGE value is considerably below the MAXDIST (48203.698 *meters*), indicating that the range of spatial dependence among nickel values is well captured by this empirical variogram

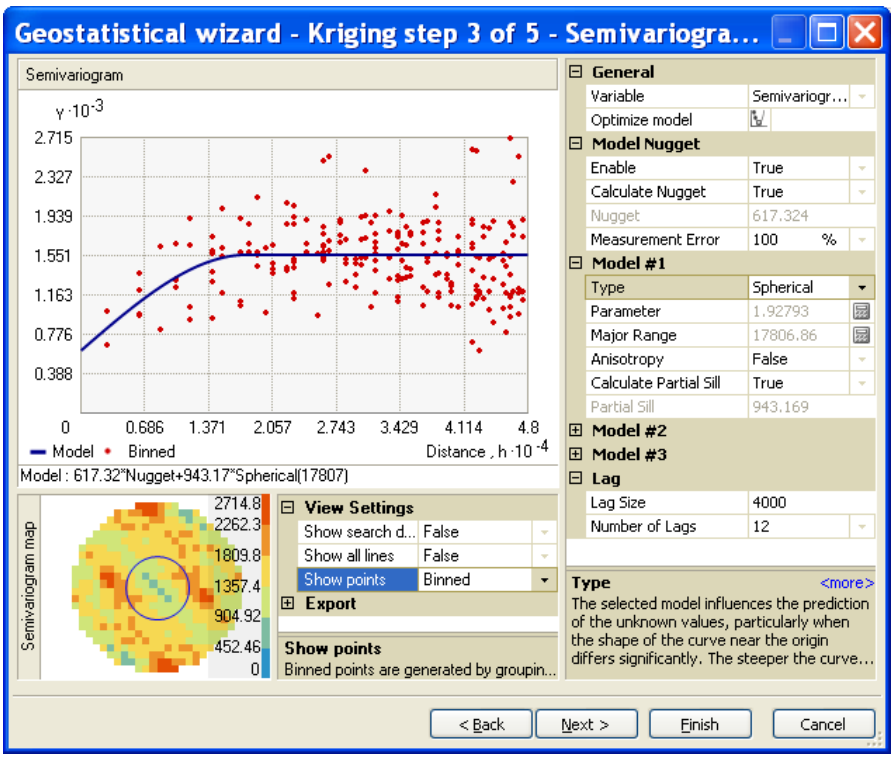


Figure 4.25. Spherical Variogram Fit in Geostatistical Analyst

Note first that the title of this window is "Semivariogram" rather than "Variogram" (as discussed at the end of Section 4.1 above). Since the full details of this variogram fitting procedure are given in Assignment 3, it suffices here to concentrate on the estimated parameter values. However, it is important to point out one aspect of this procedure that is crucial for parameter estimation. Recall from the discussion of Figure 4.13 that one must define appropriate *bins* for the empirical variogram. Since the "default" option for bin definitions in GA is rather complex compared to **ver_spher_plot.m**, it is most convenient to define the bin sizes in GA manually in order to make them (roughly) comparable to those in **ver_spher_plot**. To do so, recall from Figure 4.24 the MAXDIST value is close to 48000 meters. So by setting the number of lags to 12 and choosing a constant bin size of 4000 meters (as seen in the **Lag** window in the lower right of Figure 4.25), we will obtain a maximum distance of exactly 48000 meters (as seen on the distance axis of the variogram plot). Note also that in the **Model #1** window we have chosen Type = "Spherical", indicating that a spherical variogram is to be fitted.

The fitted *spherical variogram* is shown by the blue curve in the figure, and the *empirical variogram* is shown by red dots. Note that while the number of lags (12) is considerably smaller than the number of bins (100) used in **ver_spher_plot**, there actually appear to be

more red dots here than there are blue dots in Figure 4.22 above. The reason for this can be seen by considering the circular pattern of squares in the lower left corner of the figure. Starting from the center and moving to the right, one can count 12 squares, which denote the 12 lag distances. Hence as the figure shows, point pairs are here distinguished not only by the length of the line between them (distance) but also the direction of this line (angle). Each square thus defines a "bin" of point pairs with similar separation distances and angles. So the number of bins here is much larger than 12.[29] While these directional distinctions are important for fitting *anisotropic variogram models* in which the isotropy assumption of covariance stationarity is relaxed, we shall not explore such models in these notes.[30] Hence, under our present isotropy assumption, the appropriate empirical variogram in GA is constructed by using each of these squares as a separate bin with "lag distance" equal to the average distance between point pairs with distance-angle combinations in that square.

Next observe that in addition to the different binning conventions, the actual estimation procedure used in GA is more complex than the simple least-squares procedure used in **ver_spher_plot** [and is essentially an elaboration of the weighted least-squares approach of Cressie in shown in expressions (4.7.9) and (4.7.10) above]. So it should be clear that the resulting spherical model estimate will not be the same as in Figure 4.22 above. In particular, the estimated *range* and *nugget* in this case are given, respectively, by "Major Range" (= 17806.86) and "Nugget" (= 617.32). However, the "Sill" is here replaced by "Partial Sill" (= 943.17). Hence, recalling from the discussion at the end of Section 4.6.2 that "Sill = Partial Sill + Nugget", it follows the corresponding *sill* is here given by 1560.5 (= 943.17 + 617.32). A comparison of the parameter estimates using both MATLAB and GA in this example (Figure 4.26 below) show that in spite of the differences above, they are qualitatively very similar.

	MATLAB	**GA**
Range	**17769.2**	**17806.9**
Sill	**1554.7**	**1552.6**
Nugget	**618.0**	**617.3**

Figure 4.26. Parameter Estimates

[29] In this example, the number of bins is given approximately by $\pi 12^2 \approx 452$. However the number of red dots is actually *half* the number of bins, since each bin has a "twin" in the opposite direction. Hence the number of red dots in this case is given approximately by 226, which is still much larger than 100.
[30] For a detailed discussion of such anisotropic models see Gotway and Waller (2004, Section 2.8.5).

4.10 Variograms versus Covariograms

Before applying these methods to analyze spatially dependent data, it is appropriate to return to the question of why variograms are preferable to covariograms in terms of *estimation*. To do so, we start by showing that for any spatial stochastic process, $\left(Y(s): s \in R\right)$, satisfying the covariance stationarity condition (3.3.7) above, the "standard" sample estimator of covariance is biased.

4.10.1 Biasedness of the Standard Covariance Estimator

First recall from expression (3.3.7) that for any distance $h > 0$ the covariogram value, $C(h)$, is by definition

(4.10.1) $$C(h) = \text{cov}\left[Y(s_1), Y(s_2)\right]$$

for any $s_1, s_2 \in R$ with $\|s_1 - s_2\| = h$. Hence suppose for sake of simplicity that we are able to draw n sample pairs, $[y_1(s_{1i}), y_2(s_{2i})] = (y_{1i}, y_{2i})$, from this process with $\|s_{1i} - s_{2i}\| = h$ holding *exactly* for all $i = 1,..,n$. In this context, the standard sample estimator for the covariance value in (4.10.1) is given by

(4.10.2) $$\hat{C}(h) = \tfrac{1}{n-1}\sum_{i=1}^{n}\left(y_{1i} - \bar{y}_1\right)\left(y_{2i} - \bar{y}_2\right)$$

with sample means denoted by $\bar{y}_j = (1/n)\sum_{i=1}^{n} y_{ji}$, $j = 1,2$. Here division by $n-1$ (rather than the seemingly more natural choice of division by n) ensures that if these sample pairs $[(y_{1i}, y_{2i})$, $i = 1,..,n]$ were *independent* draws from jointly distributed random variables (Y_1, Y_2) with covariance given by (4.10.1), then $\hat{C}(h)$ in (4.10.2) would be an *unbiased* estimator of $C(h)$. However, if these pairs are *not* independent, then it is shown in Appendix A2.2 that the actual expectation of $\hat{C}(h)$ is given by

(4.10.3) $$E[\hat{C}(h)] = C(h) - \tfrac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j \neq i}\text{cov}(Y_{1i}, Y_{2j})$$

Notice first that if these sample pairs were independent then by definition each covariance, $\text{cov}(Y_{1i}, Y_{2j})$, with $i \neq j$ must be zero so that (4.10.3) would reduce to $E[\hat{C}(h)] = C(h)$, and $\hat{C}(h)$ would indeed be an unbiased estimator. But for the more classical case of *nonnegative spatial dependences*, all covariances in the second term of (4.10.3) must either be positive or zero. Hence for this classical case it is clear that there will in general be a considerable *downward bias* in this estimator. Moreover, without prior knowledge of the exact nature of such dependencies, it is difficult to correct this bias in any simple way. It is precisely this difficulty that motivates the need for alternative approaches to modeling spatial dependencies.

4.10.2 Unbiasedness of Empirical Variograms for Exact-Distance Samples

To motivate the use of variograms for modeling spatial dependencies, we begin by recalling from (4.1.7) that the covariogram, $C(h)$, is entirely determined by the variogram, $\gamma(h)$, together with the (nonspatial) variance parameter, σ^2. Hence if the empirical variogram, $\hat{\gamma}(h)$, can be shown to yields a *unbiased* estimate of $\gamma(h)$, then this will surely offer a better approach to capturing spatial dependencies.

There is one case in which this is possible, namely when there exist *multiple* pairs, $[Y(s_{1i}), Y(s_{2i}) : i = 1,..,n_h]$, each separated by the *same* distance h, i.e., satisfying the condition that $\|s_{1i} - s_{2i}\| = h$ for all $i = 1,..,n_h$. In particular, if spatial samples form a *regular lattice*, as illustrated by the small set of red dots in Figure 4.27 below, then there will generally be a set of representative distances for which this is true. In particular, the symmetry of such lattices implies that distance values such as h_1, h_2, and h_3 in the figure will occur for many different point pairs.

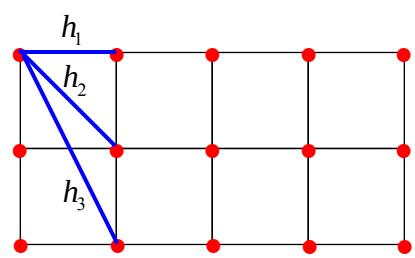


Figure 4.27. Regular Lattice of Sample Points

More generally, whenever there exists a representative range of distinct distance values, $\{h_k : k = 1,..,\bar{k}\}$, at which a substantial set of *exact-distance pairs*,

(4.10.4) $$N_k = \{(s_1, s_2) : \|s_1 - s_2\| = h_k\}$$

can be sampled at each h_k, then the associated *empirical variogram*, $\{(h_k, \hat{\gamma}_k) : k = 1,..,\bar{k}\}$, in (4.7.7) will indeed provide a meaningful *unbiased* estimate of the true variogram, $\gamma(h_k)$, at each of these distance values.[31] To see this, it is enough to recall from (4.1.5) that $E\left[\left(Y(s_{1i}) - Y(s_{2i})\right)^2\right] = 2 \cdot \gamma(h_k)$ for all $(s_1, s_2) \in N_k$, and hence that

[31] Here the qualifier "meaningful" is meant to distinguish this estimator from one in which there is no possibility of eventually accumulating a large set of sample pairs, N_k, for each h_k.

(4.10.5) $$E[\hat{\gamma}(h_k)] = E\left[\frac{1}{2|N_k|}\sum_{(s_1,s_2)\in N_k}\left(Y(s_1)-Y(s_2)\right)^2\right]$$

$$= \frac{1}{2|N_k|}\sum_{(s_1,s_2)\in N_k} E\left[\left(Y(s_1)-Y(s_2)\right)^2\right]$$

$$= \frac{1}{2|N_k|}\sum_{(s_1,s_2)\in N_k} 2\cdot\gamma(h_k)$$

$$= \frac{2|N_k|}{2|N_k|}\cdot\gamma(h_k) = \gamma(h_k)$$

So regardless of the size of each exact-distance set, N_k, this empirical variogram will *always* yield an unbiased estimate of the true variogram, $\gamma(h_k)$, at each distance $k = 1,..,\bar{k}$. Hence if in addition it is true that each of these sets is sufficiently large, say $|N_k| \geq 30$, then this empirical variogram should provide a reliable estimate of the true variogram.

Finally, it should be noted that if one is able to *choose* the pattern of samples to use in studying a given spatial stochastic process, $(Y(s): s \in R)$, then such regular lattices have the practical advantage of providing a *uniform coverage* of region R. This is particularly desirable for interpolating *unobserved* values in R (as discussed in detail in Part 6 below). It is for this reason that much attention is focused on *regular lattice samples* of such processes [as for example in Cressie (1993, p.69) and Waller and Gotway (2004, p.281)].[32]

4.10.3 Approximate Unbiasedness of General Empirical Variograms

For the general case of irregular samples, where exact-distance sets rarely contain more than one observation, it is necessary to rely on the binning procedure developed in Section 4.7.1 above. The "Nickel" example in Section 2.4 above provides a good illustration of such a case where regular sample patterns are impractical if not impossible. In this more typical setting, it is difficult to find much discussion in the literature about the bias of empirical variogram estimates created by binning.[33]

However, it is not difficult to show that if the true variogram is reasonably smooth, then one can at least bound the bias in a rather simple way. In particular, if by "smooth" we

[32] It should be mentioned again that these references define empirical variograms with respect to the more general notion of stationarity mentioned in footnote 6 of Section 3.2 above. So the exact-distance sets used here are replaced by "exact-difference sets".

[33] One noteworthy exception is the interesting analysis of "clustered" sampling schemes by Reilly and Gelman (2007).

mean that the variogram, $\gamma(h)$, is *locally linear* in the sense that its values are well approximated by linear functions on sufficiently small intervals, then one can bound the bias of the general empirical variogram in (4.7.7) in terms of these linear approximations. To be more specific, suppose that the true variogram is given by red curve in Figure 4.27 below, and that the set of bins chosen for estimating this (unknown) function are shown schematically as in Figure 4.28 below [where by definition each bin, $k = 1,..,\bar{k}$, is defined by the interval of separation distances, $d_{k-1} \le h < d_k$ (with $d_0 = 0$)].

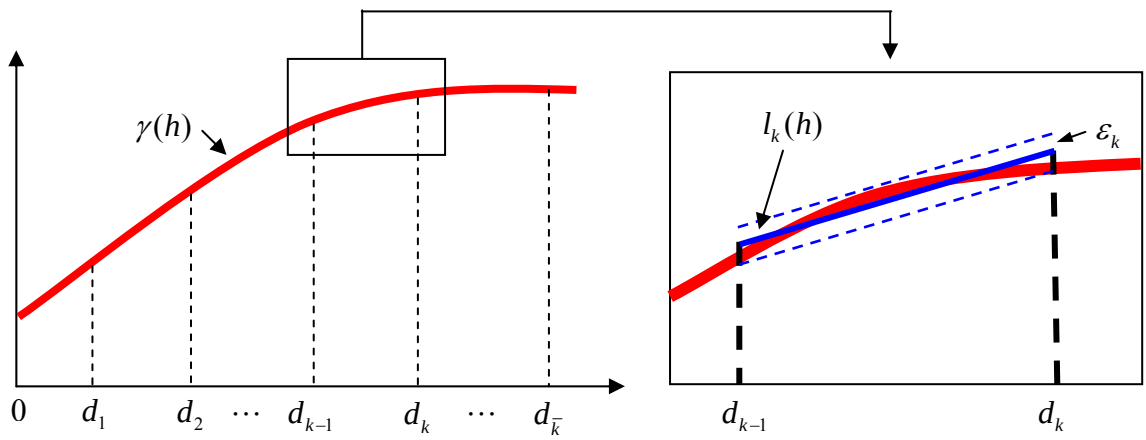


Figure 4.28 Bins for Variogram Estimation **Figure 4.29. Local Linear Approximation**

Here the variogram, $\gamma(h)$, illustrated is assumed to be an instance of the "combined model" in Figure 4.8 above. In addition, it is assumed that $\gamma(h)$ is sufficiently smooth to allow the section of the curve on each bin to be roughly approximated by a linear function. This is illustrated for a typical bin interval, $[d_{k-1}, d_k)$, by the solid blue line in Figure 4.29. This *linear approximation function*, denoted by

(4.10.6) $$l_k(h) = a_k \cdot h + b_k$$

(with slope, a_k, and intercept, b_k) has been implicitly chosen to minimize the maximum deviation, $|\gamma(h) - l_k(h)|$, over the interval $d_{k-1} \le h < d_k$. If this maximum deviation is denoted by ε_k, then the variogram, $\gamma(h)$, is said to have an ε_k*-linear approximation* on bin k. With these definitions, it is shown in Appendix A2.3 that in terms of this ε_k-linear approximation, *the maximum bias in the empirical variogram estimate of* $\gamma(h_k)$ *can never exceed* $2\varepsilon_k$*, i.e.,*

(4.10.7) $$|E[\hat{\gamma}(h_k)] - \gamma(h_k)| \le 2\varepsilon_k$$

Of course one cannot know the value of ε_k without knowing the true variogram itself. So the bound in (4.10.7) is simply a qualitative result showing that if $\gamma(h)$ is assumed to be sufficiently smooth to ensure that the *maximum deviation*, $\varepsilon = \max\{\varepsilon_k : k = 1,..,\bar{k}\}$, for the given bin partition is "small", then the bias in the empirical variogram, $\{(h_k, \hat{\gamma}_k) : k = 1,..,\bar{k}\}$, will also be "small". In other words, for variograms with good "piece-wise linear approximations" on the given set of bins, empirical variogram estimates can be expected to exhibit only minimal bias.