Conference on Applied Statistics in Agriculture   2001 - 13th Annual Conference Proceedings

ANOTHER LOOK AT CLARK'S ADIT-SILVER SERIES

David W. Meek

Recommended Citation

ANOTHER LOOK AT CLARK'S ADIT-SILVER SERIES

David W. Meek
USDA-ARS-MWA-National Soil Tilth Laboratory
Ames, IA 50011-4420 USA

Abstract

The empirical semivariogram for Clark's adit-silver series has been the subject of several publications. Its use in the literature along with some other discussed considerations makes this series a suitable selection for a classroom or workshop semivariogram modeling exercise. My emphasis in this exercise is on estimating the scale of fluctuation (θ). Alternative unbounded, bounded asymptotic, and bounded transitional models are developed via weighted least-squares estimation for both regular and integral semivariogram parameterizations (ISV). Results are compared with Clark's recommendation along with some other traditional models, non-parametric models, and ad hoc numerical methods. When a given model fits well using the regular method, generally the ISV does also. When a given model fits poorly using the regular method, however, generally the ISV form fits much worse, or gives unrealistic parameter estimates, or diverges. While an unbounded rational polynomial exponential performs well and presents some interesting existence considerations, for practical purposes the series can be considered bounded and a hyperbolic tangent is selected as the best performing simple parameterization. The ISV for the hyperbolic tangent gives parameter and θ estimates closest to ad hoc independent values for them. In the spirit of nonparametric models, however, splined-line segments can perform extremely well if parameter parsimony and parameter interpretation are not deemed important considerations by a given modeler.

1. Introduction.

I work with soil scientists, hydrologists, and other environmental scientists. While some of the research projects that I am involved with are from designed experiments, many are not. Instead the analysis goals are related to problems with monitoring, sampling, or surveying the environment. Often an understanding of intrinsic variability of set environmental point or small area measurements is needed. Determination of aggregation domains, distances between independent observations, or the equivalent number of independent observations are then some of the questions of interest. This last year I was asked to conduct an in-house geostatistics workshop; included in it is this semivariogram exercise. I examine a well-known empirical semivariogram assuming the intrinsic hypotheses are satisfied, anisotropy is not an issue, etc.; the choice of series and reasons for its selection are given in the next section. Issues of stationarity and correlation scales for each variable are paramount. Positive-definiteness issues are considered via model choice and comparisons. The goal of the exercise is to select the model that best fits the series as well as best estimates the scale of fluctuation (a.k.a., correlation scale). A basic knowledge of empirical and model semivraiograms is assumed at least on the level of Chapter 2 in Clark (1979). A more formal treatment can be obtained in Cressie's (1993) Chapter 2.

Applied Statistics in Agriculture

2. Background and Methodology

2.1 Empirical Semivariogram Selection

The selected semivariogram example is Clark's (1979) adit-silver spatial series (in accumulation units of meters %); Clark lists the series values in Table 2.5 of her book where she denotes the lag increment as h and the spatial series as $\gamma(h)$. The complete series, shown Fig. 1, exhibits a two-level nested structure with the second level starting at about h = 81 with h in meters; thus it could be argued one could use the values of $\gamma(h)$ for h ≤ 80. Several published works (e.g., Clark, 1979; Shapiro and Botha, 1991; and Meek, 2001), however, have considered only the inner series for h ≤ 75, hence, this exercise does also. Values ranged from $\gamma(1) = 0.42$ to $\gamma(66) = 14.17$ with $\bar{\gamma}=8.16$ and the following quartiles: $\gamma_{25} = 5.98$, $\gamma_{50} = 9.34$, and $\gamma_{75} = 10.69$.

Also, as shown in Fig.1, ad hoc regularity tests are done to argue for a suitable intrinsic random function (see e.g., p. 59, Chilès and Delfiner, 1999). Using the left-hand axis, continuity and differentiability are suggested because $\gamma(h)/|h|^2 \to 0$ as $|h| \to \infty$ and $\gamma(h)/|h|^2 < 0.98/(h+1) \le 0.49$. A test for positive-definiteness (see e.g., p. 322 in Isaaks and Srivastava, 1989), such as all the eigenvalues being positive for the covariance matrix, cannot be done because the original data used to construct the series are no longer available and, were they, they could not be released to the public for confidentiality reasons (personal communication from I. Clark, 2001). Likewise, a robust version of the $\gamma(h)$ series cannot be derived. Hence, the beauty of this exercise. The modeler needs to consider both bounded and unbounded models and argue for the final model that is selected.

2.2 Regression Modeling Procedures

The underlying modeling task then is to consider some nonnegative-definite functions to model a given series with suitable properties. In practice Isaaks and Srivastava (1989) recommend choosing from just a few of the best-known models. Shapiro and Botha (1991) point out that these forms are not always satisfactory and offer a broad class of nonparametric constrained optimization procedures as an alternative. Unfortunately these alternative model forms and procedures are unavailable in many geostatistics packages. Moreover they generally require mathematical programming software and experience. While ad hoc semiparametric methods are another possible alternative (see e.g., Meek, 2001), simple parametric models have their appeal and place. Using widely available nonlinear regression programs, other simple parametric models or simplified forms of nonparametric models with all the parameters explicit can be considered and evaluated. Moreover an analytic expression for the scale of fluctuation can generally be readily determined from each suitable model. Hence, for given empirical semivariograms, this latter approach may offer a practical compromise both conceptually and computationally.

The first column in Table 1 lists several popular models (see e.g., McBratney and Webster, 1986) and introduces some possible alternatives. Of course there are other possible models and families of models (see e.g., Chilés and Delfiner, 1999; McBratney and Webster, 1986). The considered semivariogram models, hereafter denoted *g(h)*, are listed by name in column 1 of

Table 1; by regular formula in column 2; and in integral form in column 3. The integral form is discussed in the next paragraph. Gotway's weighted routine and notation (1991) are adapted to estimate the parameters. Cressie (1993) argued for the given weighted minimization as a pragmatic compromise between the efficiency of generalized least squares (g.l.s.) and the simplicity of ordinary least squares. Moreover the method can be used as an initial step in obtaining g.l.s. estimators. The weight function, denoted *w(h)*, is simply $w(h) = N_h/g^2(h)$ with N_h representing the count or number of observation pairs associated with lag h. No count values could be included in the actual weight function for this exercise, however, because none are listed in Clark's Table 2.5.

Delay and de Marsily (1994) discuss the integral semivariogram (hereafter, *ISV*). The ISV method fits numerically integrated semivariogram values, known as a cumulative semivariogram (CSV), to the integrated form of each semivariogram model. The authors suggest using an ISV method when the original data are not from a regular grid and/or the γ(h) series values are undesirably dispersed. Both problems can make the regular semivariogram difficult to model. Clark's series (Fig. 1) may arguably be considered too dispersed because of the small scale pattern; hence the ISV parameters are also estimated for each model listed in column 3 of Table 1. The empirical CSVs are calculated using a trapezoidal rule integration. If need be, the ISV can be weighted. Most of the functions included in the ISV forms are available in standard regression packages. The regression for the hole-effect ISV, however, requires defining the sine integral special function, denoted Si(h). A series expansion can be used (see e.g., Eq. [5.2.14] on p. 232 in Abramowitz and Stegun, 1972). In this regression, I defined the general series term within a loop and let the number of terms be set by a convergence criterion which was 10^{-32}. The routine was tested over a range of h inputs; the output Si values agreed exactly with tabled values to the given number of significant figures but required a larger number of terms per value for larger h input values.

2.3 Background on the scale of fluctuation

Vanmarcke's book (1983) on random fields defines the concept of the scale of fluctuation and discusses numerical integration estimation methods for both spatial and temporal fields. The example in this paper is for one spatial dimension. Using the common covariance notation [c.f., Cressie, 1993], an empirical semivariogram, γ(h), is related to the covariagram, C(·), by the variance relationship $2\gamma(h) = 2(C(0) - C(h))$ where h is the separation distance and C(0) is the sill with C(0) > 0. In turn, the correlogram [a.k.a. *autocorrelation function*] is ρ(h) = C(h)/C(0). While formally γ(0) = 0, most often for γ(h) from real data sets the limit of a theoretical variogram model, g(h), as h → 0 is a finite positive value called the nugget; it is denoted c_0 with $c_0 \geq 0$. When there is a significant nugget effect, the sill in the correlation definition needs to be replaced with the partial sill, $C_{ps}(0) = C(0) - c_0$, hereafter, denoted c_s for brevity and to keep with the given notation. Under regularity conditions (see e.g., Cressie, 1993 or Vanmarcke, 1983), a linear scale of fluctuation, θ, is defined as

Applied Statistics in Agriculture

$$\theta = 2 \int_0^\infty \rho(h) dh \tag{1a}$$

where

$$\rho(h) = 1 - (\gamma(h) - c_0)/c_s \tag{1b},$$

for an empirical semivariogram or

$$\hat{\rho}(h) = 1 - (g(h) - c_0)/c_s \tag{1c},$$

for a semivariogram model.

For common semivariogram models that have sills or asymptotes, θ can be determined using Eq. (1a) with each $g(h)$ model under consideration substituted in Eq. (1c). For many popular models, the integrals of the $g(h)$ expressions are similar to those for the ISV which are listed in column 3 of Table 1; for $g(h)$ functions it is defined for, the related correlation functions are listed in column 4; and the resulting $\theta(a_s)$ expressions are listed in column 5. Asymptotic models like the exponential have no finite range but do have a simple, finite θ value expressed as a function of their a_s parameter. Also the concept of θ with the asymptotic models avoids having to stipulate a practical range definition such as the location where 95% of the sill value is achieved [see e.g., pp. 374-375 in Isaaks and Srivastava, 1989].

2.4 Evaluation Criteria

Evaluation is guided by common modeling statistics, common sense, and some ad hoc alternative estimates. Let n be the number of values in the series; p be the number of parameters in the model; CSS be weighted corrected sum of squares; ESS be weighted error total sum of squares; PRESS be the weighted predictive error sum of squares; and MSE be weighted error mean square. Some common regression modeling evaluation statistics are the weighted coefficient of determination, $R^2 = 1 - ESS/CSS$, its simple jackknife counterpart, $R_p^2 = 1 - PRESS/CSS$ (see e.g., p. 432 in Montgomery and Peck, 1982), McBratney and Webster's (1986) shortened version of the Akaike Information Criterion (AIC,) where $AIC \approx n \times \ln(ESS) + 2 \times p$, and Fuller's (1996) adjusted mean square error (MPE) with $MPE = MSE \times (n+p)/n$. For the first two, higher values are better and for the last two, lower values are better. Absolute values for any of these statistics must be taken with caution because the weighted CSS varies somewhat between the different models. The values of R_p^2 are not reported for purposes of brevity. The statistic is simply used as an additional comparative tool. Similarly, the parameter T values and some of the other listed statistics are also considered but only comparatively because the values in the series are not independent; hence, they too are also omitted from the tabled results for brevity. A more general cross-validation to compare predictions, would present inference problems (see e.g. p. 104 in Cressie, 1993). Moreover the original adit-silver observations are unavailable for this purpose.

Ad hoc independent semiparametric estimates for c_0, c_s, and θ can be employed to obtain $47 \le \theta \le 49$, with $c_0 \approx 0.1$ and $11.7 \le c_0 + c_s \le 12.0$ (Meek, 2001). If the series values for $76 \le h \le 80$ are

included in the later methodology then the estimates are $c_0 \approx 0.1$, $c_0 + c_s \approx 12.0$, $\theta \approx 49.2$. If Shapiro and Botha's (1991) fig. 2 fit is considered valid then by constructing the related correlogram, interpolating, then numerically integrating $\hat{\rho}(h)$ for the plot, one finds that $c_0 \approx 0$, $c_0 + c_s \approx 11.6$, and $\theta \approx 48$. Hence, based on all these estimates and model performance statistics, comparisons and selections can be made.

3. Results and Discussion

3.1 Clark's Analysis

Results are summarized in Table 2. Table 2 is rich in information and hence is best carefully worked through section by section. The sections are arranged by model/analysis type. When a model name appears twice consecutively, the first one is without the nugget parameter, c_0. The values not enclosed by the brackets are for the regular regression analyses, while the values inside the brackets are for the ISV analyses.

Clark's 'eyeball' examination of the series is summarized in the three models listed in the top section of Table 2; she selected the spherical model. Her argument is that the spherical model interpolates the rising portion better than the exponential model does. The bounded line is a preliminary way to get initial estimates for the spherical model. Of consequence for this modeling exercise, she sets the standard for the performance of alternative models to meet or exceed: $R^2 = 0.976$, AIC $= 23.3$, and 100 MPE $= 1.82$.

3.2 Modeling Evaluation

In each of the remaining lower sections of Table 2, full inverse variance weighted regression estimates for both the regular and ISV model forms are presented. Regression diagnostics (see e.g., Ch.3 in Montgomery and Peck, 1982) indicate that the Gotway's (1991) weight function is generally suitable for the regular semivariogram models. The ISV models, however, may be better estimated with a weaker weight, $\epsilon^2(h) = g(h)$. The R_p^2 results closely match those for R^2 and are part of the basis for stating the next finding. Generally when a given model fits very well using the regular method, the ISV does also. When a given model fits poorly using the regular method, however, generally the ISV form fits much worse, or gives unrealistic parameter estimates, or diverges. In each of the lower sections in Table 2 there is at least one traditional $g(h)$ model and one or more of the nontraditional alternatives that were introduced in Table 1 or suggested in the methods section. Recall Isaaks and Srivastava (1989) argue for just using the few well-known forms. For this series the development of alternative suitable functions, however, proves to be worthwhile because they have better overall fits. Importantly, better sill and θ estimates are obtained, generally each with values higher than those recommended by Clark. Relevantly and of particular importance to kriging, the nugget is generally not zero. Specific results by model class support these findings.

Notice that the Gaussian, hole, and rational quadratic models are obviously poor fits, much worse than any of the others. In fact, based on the performance statistics only the exponential and hyperbolic tangent models are viable asymptotic forms; likewise, the spherical and bounded

Applied Statistics in Agriculture

quadratic are the only traditional transitional forms; 3 and 4 splined-line segments are the only forms in the splined-line segments model class; and the rational polynomial exponential is the only possible model for the unbounded model class. Five models have $R^2 \approx 0.980$ but only the 4 splined-line segments model with c_0 is rounded down; the rest are rounded up. The absolute interpolation order of the remaining four is 4-splined-lines model without c_0, the hyperbolic tangent with c_0, the hyperbolic tangent without c_0, and then the rational polynomial exponential. The absolute difference in R^2 between the first and last is about 0.0009. Overall, the exponential fit interpolates better than the spherical model fit. No c_0 is presented for the exponential model because a fit including it gives a $c_0 < 0$ estimate. Similarly in the results reported for other models, the absence of parameter and statistics means the model either diverged or converged to unrealistic parameter estimates, a condition which is generally just $c_0 < 0$. Based on interpolation and parsimony, the bounded quadratic model without c_0 is the best transitional model choice while the hyperbolic tangent is the best overall choice of the models listed in the middle two sections.

In the next section forms with splined-line segments are considered. As discussed by many researchers, including Cressie (1993) and Shapiro and Botha (1991), there are many implicit assumptions involved when modeling $\gamma(h)$. The latter authors addressed some of the underlying premises and presented some examples of nonnegative definite functions via nonparametric curves fit to the very same selected $\gamma(h)$ example from Clark (1979). The methodology is not actually nonparametic but the curve form properties are what are important rather than values of the local regression parameters. In the spirit of Shapiro and Botha's (1991) fig. 2 nonparametric approach, I parametrically add linear segments to the bounded line model. This approach can still be considered a transitional model if the last segment has no slope. I stopped at 4 segments because adding one more became difficult and did little to improve the fit which is already better in interpolation based on R^2 than that for any other model considered in Table 2. The last result, however, is at the cost of added parameters which are mostly spent in trying to adequately model the rising section of the $\gamma(h)$ series. Although not having the highest R^2 value, overall the performance statistics for this class of models favor the choice of the 3 splined-line segments version, especially given the ISV results. Interestingly the splined-line segments approach may offer some interesting insight into a relevant modeling practice. To get starting values for an intrinsically nonlinear g(h) model often eyeball or ad hoc methods are used. A popular ad hoc estimation procedure for c_0 takes the intercept from a linear regression, preferably weighted, for the first three or more values in the series (notice the parameters should be similar to the first segment of a splined-line segments model). Similarly for the far end, the mean of say the last 5% or more of the series is used (notice the parameter is essentially the sill for the last segment for a splined-line segments model). My semiparametic alternative utilizes these ideas to avoid trying to model the rise section altogether (Meek, 2001).

Finally, consider the results presented in the last section of Table 2. While none of the usual conventional unbounded models like the power model fit very well, based on R^2 the rational polynomial exponential alternative interpolates almost as well as the hyperbolic tangent and the 4 splined-line segments model. Moreover its AIC and 100 MPE values are the lowest overall.

Conference on Applied Statistics in Agriculture
Kansas State University
362   **Kansas State University**

3.3 Model Selection and Discussion

Selections by model type are shown in Fig. 2. The performance statistics in Table 2 represent the average or overall behavior over the entire h series. The local performance of the fit, however, is important in some applications. So let's consider the fit on the sill and near the origin. The open squares in each frame of Fig. 2 are the series values for $76 \leq h \leq 80$. While the models were developed for the series values in $1 \leq h \leq 75$, the graphs depict the predictions for $1 \leq h \leq 80$. Clockwise from the upper left graph in Fig. 2, the weighted prediction root-mean-square (RMS) errors between each model's prediction and open square values are 0.13, 0.20, 0.15, and 0.09, respectively. In comparison, for $1 \leq h \leq 8$ the respective RMS errors are 0.22, 0.28, 0.20, and 0.20. In all of the bounded models the inclusion of c_0 in the model reduced the local RMS error.

Of these four selected models the bounded quadratic is obviously the poorest choice and so is eliminated first. Not surprisingly the 4 splined-line segments model gives the c_0 and c_s estimates that are closest to those inferred from Shapiro and Botha's (1991) fig. 2. Moreover it has the highest R^2. All versions of the hyperbolic tangent give the c_s or the c_0 and c_s estimates that are closest to my ad hoc values for them; moreover, the θ estimates are the closest of all to values given by the alternative estimators. Based on the comparative performance statistics, however, the unbounded rational polynomial exponential and the bounded hyperbolic tangent both look like viable choices. Unfortunately, the inclusion of values $76 \leq h \leq 80$ in the latter two models gives the same overall pattern of comparative results. The rational polynomial exponential interpolates the right-most values of the series somewhat better than the hyperbolic tangent but remember that the latter form is an asymptotic model with the sill as its asymptote.

Questions on boundedness and θ existence are thus at hand. For this series, the performance of the rational polynomial exponential certainly allows room for one to play the devil's advocate. For practical purposes, however, if a splined-line segments or other suitable model fits the series very well then there is a sill at the right-most subinterval of h. Shapiro and Botha's (1991) constrained approach is actually meant to guarantee the curve's behavior. Also remember the results of the regularity tests. Alternatively, one can perform some more data analysis. While running means for the outer series values tend to increase as the averaging interval is shortened and moved to the right (from ≈ 11.3 for the last 25 to ≈ 12 for the last 5), the mean seems to stabilize in the interval $71 \leq h \leq 79$. Analysis of change with smoothed numerical derivatives reveals the same behavior as the running mean analysis. Still other ad hoc and more formal tests are possible and discussed in many references like Chilés and Delfiner (1999), Cressie (1993), or Shapiro and Botha (1991). In reality, however, one cannot say for certain that there is a true model or class of models. Rather, empiricism, custom, goals, judgement, experience, and sometimes software capabilities dictate the choice. When and if possible, prior knowledge of the process, its covariance structure model, etc. should be considered. Here consistency of the selected model is based on comparison with several different ad hoc analyses.

Applied Statistics in Agriculture

4. Conclusions

Given its history and properties, Clark's adit-silver series is an ideal choice for a semivariogram modeling exercise. For this series, nontraditional unbounded, asymptotic, and transitional alternatives to the popular semivariogram models can be formulated and shown to characterize the series behavior better than any traditional model does. In future workshops the series should include the values for $75 \leq h \leq 80$. Practical considerations support the choice of a bounded model. All viable options indicate $c_0 + c_s$ and θ values higher than those argued for by Clark. Additionally most alternatives improve with $c_0 > 0$. Considering the comparative performance statistics and parameter estimates, my choice for a bounded model is the hyperbolic tangent. The θ estimates from the nugget form of the ISV for the hyperbolic tangent is closest to the argued value from the ad hoc estimators.

Acknowledgments

References

Abramowitz, M. and Stegun, I. 1972. Handbook of mathematical functions with formulas, graphs, and mathematical tables, 9th Printing. Dover Pub. Inc., New York, 1046 pp.

Chilés, J.P. and Delfiner, P. 1999. Geostatistics: Modeling spatial uncertainty. J. Wiley & Sons, New York, 695 pp.

Clark, I. 1979. Practical geostatistics. Applied Science Publishers, London, 129 pp.

Cressie, N.A.C. 1993. Statistics for spatial data. J. Wiley & Sons, New York, 900 pp.

Delay, F. and de Marsily, G. 1994. The integral semivariogram: A powerful method for adjusting the semivariogram in geostatistics. Mathematical Geology 26 (3), 301-321.

Fuller, W.A. 1996. Introduction to statistical time series. J. Wiley & Sons, New York, 698 pp.

Gotway, C.A. 1991. Fitting semivariogram models by weighted least squares. Computers & Geosciences 17, 171-172.

Isaaks, E.H. and Srivastava, R.M. 1989. An introduction to applied geostatistics. Oxford Univ. Press, New York, 561 pp.

McBratney, A.B. and Webster, R. 1986. Choosing functions for semivariograms of soil properties and fitting them to sampling estimates. J. Soil Science 37, 617-639.

Meek, D. 2001. A semiparametric method for estimating the scale of fluctuation. Computers & Geosciences, 1245-1251.

Montgomery, D.C., and Peck, E.A. 1982. Introduction to linear regression analysis. John Wiley & Sons, New York, 504 pp.

Ratkowsky, D.A. 1989. Handbook of nonlinear regression models. Marcel Dekker, Inc., New York, 241 pp.

Shapiro, A. and Botha, J.D. 1991. Variogram fitting with a general class of nonnegative definite functions. Computational Statistics & Data Analysis. 11, 87-96.

Vanmarcke, E. 1983. Random Fields: Analysis and Synthesis. MIT Press, Cambridge, 382 pp.

Table 1.

Properties of some selected semivariogram models.

Name	$g(h)$[a]	ISV	$\rho(h)$	θ[b]
		Unbounded Models		
Linear	$c_0 + bh$	$c_0h + 0.5bh^2$	-	-
Power	$c_0 + bh^p$	$c_0h + (bh^{p+1})/(p+1)$	-	-
Logarithmic	$a\log(h)$	$a(h\log(h) - h)$	-	-
Rational Quadratic	$c_0 + bh^2/(1 + h^2/a)$	$(c_0 + ba)h - ba^{3/2}\arctan(h/a^{1/2})$	-	-
		Asymptotic Models		
Gaussian	$c_0 + c_s(1 - \exp(-(h/a_s)^2))$	$(c_0 + c_s)h - c_s\,a_s\,\pi^{1/2}(\text{erf}\,((\sqrt{2}h/a_s) - 0.5)$	$\exp(-(h/a_s)^2)$	$\pi^{1/2}a_s$
Exponential	$c_0 + c_s(1 - \exp(-(h/a_s)))$	$c_0h + c_s(h + a_s(\exp(-(h/a_s)-1)))$	$\exp(-(h/a_s))$	$2a_s$
Hyperbolic Tangent[c]	$c_0 + c_s\tanh(h/a_s)$	$c_0h + c_s\,a_s\log(\cosh(h/a_s))$	$1 - \tanh(h/a_s)$	$\ln(4)a_s$
Hole Effect[d]	$c_0 + c_s(1 - a_s\sin(h/a_s)/h)$	$c_0h + c_s(h - a_s\text{si}(h/a_s))$	$a_s\sin(h/a_s)/h$	$\frac{1}{2}\pi\,a_s$
Cauchy(p=1)	$c_0 + c_s(1 - (1 + (h/a_s)^2)^{-1})$	$c_0h + c_s(h - a_s\arctan(h/a_s))$	$(1 + (h/a_s)^2)^{-1}$	πa_s
		Transitional Models		
Bounded Line	$\begin{cases} c_0 + c_s h/a_s, & 0 < h \le a_s \\ c_0 + c_s, & h > a_s \end{cases}$	$\begin{cases} c_0h + 0.5c_sh^2/a_s, & 0 < h \le a_s \\ (c_0 + c_s)h - 0.5c_sa_s, & h > a_s \end{cases}$	$\begin{cases} 1 - h/a_s, & 0 < h \le a_s \\ 0, & h > a_s \end{cases}$	a_s
Bounded Quadratic[e]	$\begin{cases} c_0 + c_s(2(h/a_s) - (h/a_s)^2), & 0 < h \le a_s \\ c_0 + c_s, & h > a_s \end{cases}$	$\begin{cases} c_0h + c_s((h^2/a_s) - 0.33(h^3/a_s^2)), & 0 < h \le a_s \\ (c_0 + c_s)h - 0.33c_sa_s, & h > a_s \end{cases}$	$\begin{cases} 1 + (h/a_s)^2 - 2(h/a_s), & 0 < h \le a_s \\ 0, & h > a_s \end{cases}$	$2a_s/3$
Spherical	$\begin{cases} c_0 + c_s(1.5(h/a_s) - 0.5(h/a_s)^3), & 0 < h \le a_s \\ c_0 + c_s, & h > a_s \end{cases}$	$\begin{cases} c_0h + c_s(0.75(h^2/a_s) - 0.125(h^4/a_s^3)), & 0 < h \le a_s \\ (c_0 + c_s)h - 0.375c_sa_s & h > a_s \end{cases}$	$\begin{cases} 1 + 0.5(h/a_s)^3 - 1.5(h/a_s), & 0 < h \le a_s \\ 0, & h > a_s \end{cases}$	$3a_s/4$
Circular	$\begin{cases} c_0 + c_s(1 - (2/\pi)(\cos^{-1}(h/a_s) - (h/a_s)(1-(h/a_s)^2)^{1/2})), & 0 < h \le a_s \\ c_0 + c_s, & h > a_s \end{cases}$	$\begin{cases} (c_0+c_s)h + (2c_s/\pi)((a_s^2-h^2)^{1/2}(2a_s^2+h^2)/3a_s^2 - h\cos^{-1}(h/a_s)), & 0 < h \le a_s \\ (c_0+c_s)h, & h > a_s \end{cases}$	$\begin{cases} (2/\pi)(\cos^{-1}(h/a_s) - (h/a_s)(1-(h/a_s)^2)^{1/2}), & 0 < h \le a_s \\ 0, & h > a_s \end{cases}$	$8a_s/3\pi$

[a] Model Parameters: c_0 = Nugget; c_S = Partial Sill ($C(0) - c_0$), where $C(0)$ = Sill; a_s = Range or asymptotic scaling parameter; θ = Scale of Fluctuation. Note that by definition, $\gamma(0) = 0$, thus $g(0) = 0$ for all models.

[b] Note that for θ from the Gaussian model, $\pi^{1/2} \approx 1.772$; for the tanh model, $\ln(4) \approx 1.386$; for the hole-effect model, $\frac{1}{2}\pi \approx 1.571$; for the circular model, $8/3\pi \approx 0.8488$. For θ to exist for the Cauchy model, $p > \frac{1}{2}$.

[c] Notice that formally $\tanh(x) = 2\text{logistic}(2x) - 1$ where $\text{logistic}(x) = 1/(1 + \exp(-x))$. This model is a very simple form of an artificial neural net as well as a common growth curve. Hence the use of this model invites the consideration of other basis functions from Fourier and wavelet series as well as related growth functions.

[d] The hole-effect model and related forms may also be known as the *wave model* or the *periodic model*.

[e] The spherical model is really a partial term cubic polynomial form splined to a plateau (a linear segment without a slope); hence it could be called a *Bounded Cubic* polynomial. A second degree polynomial form of the spherical model, the *Bounded Quadratic* model listed above may be an interesting alternative to consider for a $\gamma(h)$ model.

Table 1.

Properties of some selected semivariogram models.

Name	$g(h)^a$	ISV	$\rho(h)$	θ^b
Unbounded Models				
Linear	$c_0 + bh$	$c_0h + 0.5bh^2$	-	-
Power	$c_0 + bh^p$	$c_0h + (bh^{p+1})/(p+1)$	-	-
Logarithmic	$a\log(h)$	$a(h\log(h) - h)$	-	-
Rational Quadratic	$c_0 + bh^2/(1 + h^2/a)$	$(c_0 + ba)h - ba^{3/2}\arctan(h/a^{1/2})$	-	-
Asymptotic Models				
Gaussian	$c_0 + c_s(1 - \exp(-(h/a_s)^2))$	$(c_0 + c_s)h - c_s\,a_s\,\pi^{1/2}(\text{erf}\,((\sqrt{2}h/a_s) - 0.5)$	$\exp(-(h/a_s)^2)$	$\pi^{1/4}a_s$
Exponential	$c_0 + c_s(1 - \exp(-(h/a_s)))$	$c_0h + c_s(h + a_s(\exp(-(h/a_s))-1))$	$\exp(-(h/a_s))$	$2a_s$
Hyperbolic Tangent[c]	$c_0 + c_s\tanh(h/a_s)$	$c_0h + c_s\,a_s\log(\cosh(h/a_s))$	$1 - \tanh(h/a_s)$	$\ln(4)a_s$
Hole Effect[d]	$c_0 + c_s(1 - a_s\sin(h/a_s)/h)$	$c_0h + c_s(h - a_s\text{si}(h/a_s))$	$a_s\sin(h/a_s)/h$	$\frac{1}{2}\pi\,a_s$
Cauchy(p=1)	$c_0 + c_s(1 - (1 + (h/a_s)^2)^{-1})$	$c_0h + c_s(h - a_s\arctan(h/a_s))$	$(1 + (h/a_s)^2)^{-1}$	πa_s
Transitional Models				
Bounded Line	$\begin{cases} c_0 + c_sh/a_s, & 0 < h \le a_s \\ c_0 + c_s, & h > a_s \end{cases}$	$\begin{cases} c_0h + 0.5c_sh^2/a_s, & 0 < h \le a_s \\ (c_0 + c_s)h - 0.5c_sa_s, & h > a_s \end{cases}$	$\begin{cases} 1 - h/a_s, & 0 < h \le a_s \\ 0, & h > a_s \end{cases}$	a_s
Bounded Quadratic[e]	$\begin{cases} c_0 + c_s(2(h/a_s) - (h/a_s)^2), & 0 < h \le a_s \\ c_0 + c_s, & h > a_s \end{cases}$	$\begin{cases} c_0h + c_s((h^2/a_s) - 0.33(h^3/a_s^2)), & 0 < h \le a_s \\ (c_0 + c_s)h - 0.33c_sa_s, & h > a_s \end{cases}$	$\begin{cases} 1 + (h/a_s)^2 - 2(h/a_s), & 0 < h \le a_s \\ 0, & h > a_s \end{cases}$	$2a_s/3$
Spherical	$\begin{cases} c_0 + c_s(1.5(h/a_s)-0.5(h/a_s)^3), & 0 < h \le a_s \\ c_0 + c_s, & h > a_s \end{cases}$	$\begin{cases} c_0h + c_s(0.75(h^2/a_s)-0.125(h^4/a_s^3)), & 0 < h \le a_s \\ (c_0 + c_s)h - 0.375c_sa_s, & h > a_s \end{cases}$	$\begin{cases} 1 + 0.5(h/a_s)^3 - 1.5(h/a_s), & 0 < h \le a_s \\ 0, & h > a_s \end{cases}$	$3a_s/4$
Circular	$\begin{cases} c_0 + c_s(1-(2/\pi)(\cos^{-1}(h/a_s)-(h/a_s)(1-(h/a_s)^2)^{1/2})), & 0 < h \le a_s \\ c_0 + c_s, & h > a_s \end{cases}$	$\begin{cases} (c_0+c_s)h+(2c_s/\pi)((a_s^2-h^2)^{1/2}(2a_s^2+h^2)/3a_s^2-h\cos^{-1}(h/a_s)), & 0 < h \le a_s \\ (c_0+c_s)h, & h > a_s \end{cases}$	$\begin{cases} (2/\pi)(\cos^{-1}(h/a_s)-(h/a_s)(1-(h/a_s)^2)^{1/2}), & 0 < h \le a_s \\ 0, & h > a_s \end{cases}$	$8a_s/3\pi$

[a] Model Parameters: c_0 = Nugget; c_s = Partial Sill ($C(0) - c_0$), where $C(0)$ = Sill; a_s = Range or asymptotic scaling parameter; θ = Scale of Fluctuation. Note that by definition, $\gamma(0) = 0$, thus $g(0) = 0$ for all models.

[b] Note that for θ from the Gaussian model, $\pi^{1/2} \approx 1.772$; for the tanh model, $\ln(4) \approx 1.386$; for the hole-effect model, $\frac{1}{2}\pi \approx 1.571$; for the circular model, $8/3\pi \approx 0.8488$. For θ to exist for the Cauchy model, $p > \frac{1}{2}$.

[c] Notice that formally $\tanh(x) = 2\text{logistic}(2x) - 1$ where $\text{logistic}(x) = 1/(1 + \exp(-x))$. This model is a very simple form of an artificial neural net as well as a common growth curve. Hence the use of this model invites the consideration of other basis functions from Fourier and wavelet series as well as related growth functions.

[d] The hole-effect model and related forms may also be known as the *wave model* or the *periodic model*.

[e] The spherical model is really a partial term cubic polynomial form splined to a plateau (a linear segment without a slope); hence it could be called a *Bounded Cubic* polynomial. A second degree polynomial form of the spherical model, the *Bounded Quadratic* model listed above may be an interesting alternative to consider for a $\gamma(h)$ model.

Figure 1. Clark's complete adit-silver series from Table 2.5 in her book (Clark, 1979). The $\gamma(h)$ units are (m %)^2 where h is the distance increment in meters. The analyses in this paper concern only the values for $1 \leq h \leq 75$.
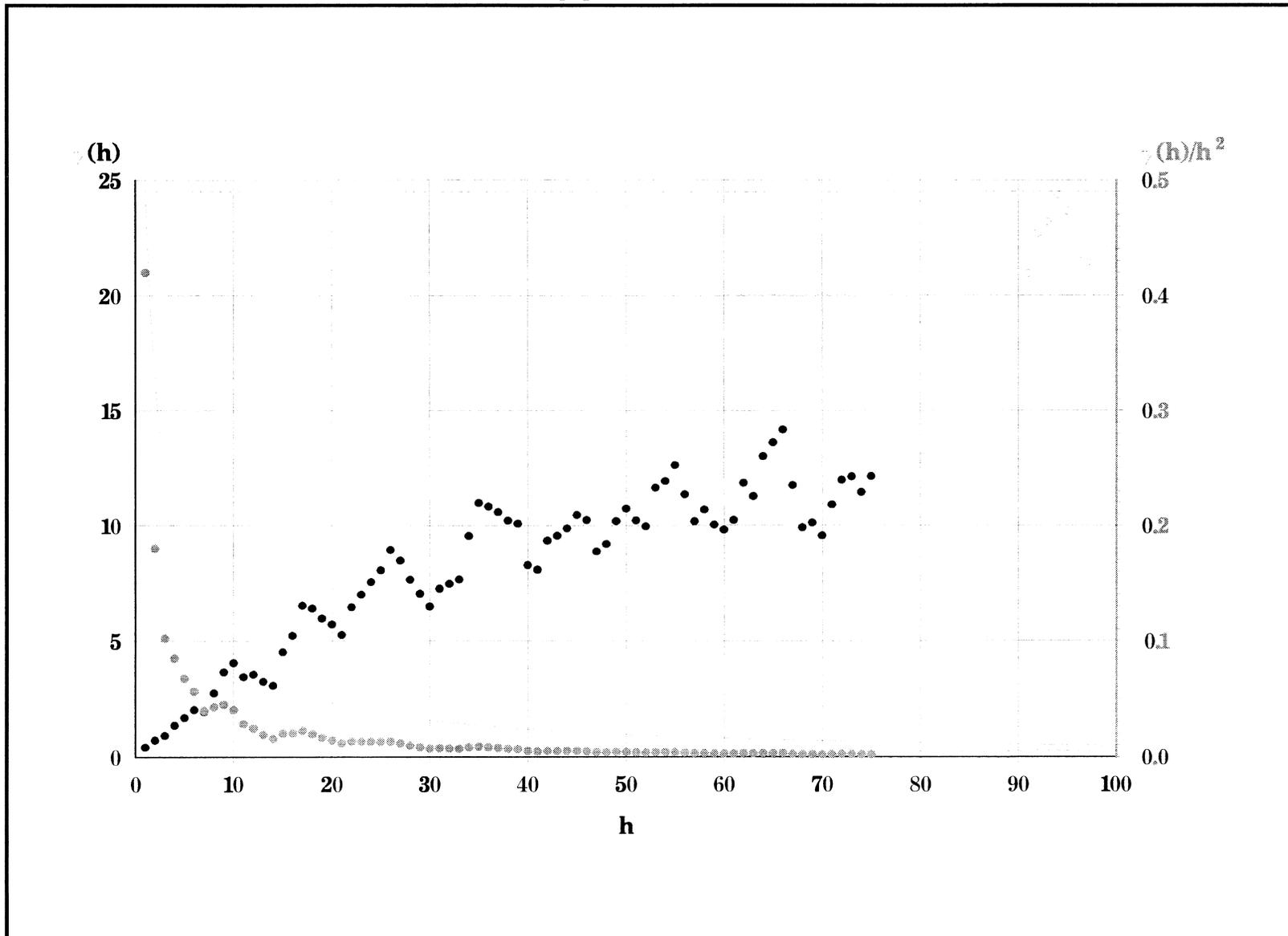
Figure 2. Selected models by class. Black circles and lines represent the semivariogram and its models while the gray circles and lines represent the corresponding correlogram and correlation function. Open squares are the series values for $76 \le h \le 80$.