The Variogram

Introduction

- Spatial Interpolation is the domain of geo-statistics, which deals with a random variable that is observed at fixed sample point locations. It aims at estimating a ***spatial field*** around the sample locations.
  In contrast, spatial statistical analysis deals with lattice/areal data (aggregation of points information within an area).

relationships between individuals

- The variogram focuses on estimating the ***second order variation*** of a stationary surface based on the observed values at the sampling points.

- A spatial field $Y(\mathbf{s}_i)$ can be decomposed into three components:

general trend   error term structured model it through individuals

$$Y(\mathbf{s}_i) = \underbrace{\mu(\mathbf{s}_i)}_{\substack{\text{first order =}\\ \text{exogenous trend}}} + \underbrace{\varepsilon'(\mathbf{s}_i)}_{\substack{\text{second order =}\\ \text{stochastic inter-dependence}}} + \underbrace{\varepsilon''(\mathbf{s}_i)}_{\substack{\text{white noise =}\\ \text{nugget effect}}}$$

totally random

- A ***regionalized random variable*** $\varepsilon'(\mathbf{s}_i)$ is continuous from point to point, but the spatial variability with respect to the locations $\mathbf{s}_i$ is too complex that it cannot be described with any tractable deterministic first order function (such as a trend surface).
  Thus, its spatial variation of the stochastic variable $\varepsilon'(\mathbf{s}_i)$ is best described by a covariance structure (function).

- The ***first order*** component can be generalized to any expected value of the random variable $E\left[Y(\mathbf{s}_i)\right] = \mu(\mathbf{s}_i)$ at an arbitrary location $\mathbf{s}_i$. For trend surfaces this expectation simply becomes $\mu(\mathbf{s}_i) = \mathbf{x}^T(\mathbf{s}_i) \cdot \boldsymbol{\beta}$

- The concept of stationarity applies to the ***covariance*** among the pair of random components $\varepsilon'(\mathbf{s}_i)+\varepsilon''(\mathbf{s}_i)$:
  - $E\left[\varepsilon'(\mathbf{s}_i)+\varepsilon''(\mathbf{s}_i)\right]=0$ and $Var\left[\varepsilon'(\mathbf{s}_i)+\varepsilon''(\mathbf{s}_i)\right]=const$

    relative location
  - $Cov\left[\left(\varepsilon'(\mathbf{s}_i)+\varepsilon''(\mathbf{s}_i)\right),\left(\varepsilon'(\mathbf{s}_j)+\varepsilon''(\mathbf{s}_j)\right)\right]$ does not depend on the absolute locations of $\mathbf{s}_i$ and $\mathbf{s}_j$, but only on their distance (and perhaps under *anisotropy* on the direction).
- The spatial arrangement of the sample points constitutes the ***support*** of the regionalized variable. Predictions outside the support become extrapolations.

  means applied area, once beyond this area, the discovered rule are not

Historical Background.

- South African mining engineer D.G. Krige. Early 1950 developed optimal core grade estimation from sample points.
- ***Meaning of "optimal"***: Interpolation weights $\lambda_i(\mathbf{s})$ are chosen to optimize the interpolation function $\hat{Y}(\mathbf{s})=\sum_{i=1}^{n}\lambda_i(\mathbf{s})\cdot Y(\mathbf{s}_i)$ for any prediction location $\mathbf{s}$ given the sampling locations $\mathbf{s}_i$,

  variance of prediction is smallest means best

  i.e., to provide a Best Linear Unbiased Estimate (BLUE) of the random field $\hat{Y}(\mathbf{s})$ at any location $\mathbf{s}$.
- Geo-statistics was established by Georges Matheron of the Centre of Morphologie Mathematique in Fontainebleu, France. (early 1960).
- However, similar aspatial estimation approaches are known for a long time in other branches of sciences.
- Due to this origin in the mining industry much of the lingo of geo-statistics is in terms of mining engineering (e.g., ***drift***, ***sill*** and ***nugget***)

Basic Kriging models

first order component equal to 0

- For ***simply kriging*** it is assumed that $E\left[Y(\mathbf{s}_i)\right]=0\ \forall s_i \in \Re$, that is $\mu(\mathbf{s}_i)=0$ is given externally
- For ***ordinary kriging*** we assume that $E\left[Y(\mathbf{s}_i)\right]=c\ \forall s_i \in \Re$ has a constant expectation $\mu(\mathbf{s}_i)=c$ unequal to zero. Thus the expectation does not dependent on the location
- For ***universal kriging*** we assume that $E\left[Y(\mathbf{s}_i)\right]=\mu(\mathbf{s}_i)\ \forall \mathbf{s}_i \in \Re$ has spatially varying expectation, e.g., $\mu(\mathbf{s}_i)=\mathbf{x}^T(\mathbf{s}_i)\cdot\boldsymbol{\beta}$.

  for instance, trend surface value
- The ordinary and universal kriging models can always be reduced to the simple kriging model by first estimating the expectation component $\hat{\mu}(\mathbf{s}_i)$ and then transforming the observed spatial field $Y(\mathbf{s}_i)$ by $\left[Y(\mathbf{s}_i)-\hat{\mu}(\mathbf{s}_i)\right]$ because $E\left[Y(\mathbf{s}_i)-\hat{\mu}(\mathbf{s}_i)\right]=0$.

  The expectation component $\hat{\mu}(\mathbf{s}_i)$ must be estimated by ***feasible generalized least squares*** with an estimated covariance structure because the observed spatial field $Y(\mathbf{s}_i)$ is autocorrelated, which violates the **OLS** independence assumption.

  (Recall *ordinary least squares* requires that the individual observations are independent of each other, i.e., $Cov\left[Y(\mathbf{s}_i),Y(\mathbf{s}_j)\right]=0$ for all $\mathbf{s}_i \neq \mathbf{s}_j$)

Deterministic Inverse Distance Weighted Interpolator

- Inverse distance weighted interpolation is used frequently to obtain a first idea about the spatial variability of a spatial field $Y(\mathbf{s}_i)$.
- Its spatial dependence (i.e., covariation) among the sample points is determined ***determistically***.

- It is defined as

$$\hat{Y}(\mathbf{s}) = \frac{\sum_{i=1}^{n}[h(\mathbf{s}_i,\mathbf{s})]^{-\alpha} \cdot Y(\mathbf{s}_i)}{\sum_{i=1}^{n}[h(\mathbf{s}_i,\mathbf{s})]^{-\alpha}}$$

  with $h(\mathbf{s}_i,\mathbf{s})$ being the distance between the sample location $\mathbf{s}_i$ and the prediction location $\mathbf{s}$ and $\alpha \geq 0$ a smoothing parameter.

- By convention the smoothing value $\alpha = 2$ is used:
  - For $\alpha = 0$ extreme smoothing is achieved with $\hat{Y}(\mathbf{s}) = \frac{\sum_{i=1}^{n} Y(\mathbf{s}_i)}{n}$ at all prediction locations $\mathbf{s}$.
  - For $\alpha = \infty$ the smoothed pattern resembles that of a Voronoi polygon with each tile having a value equal to $\hat{Y}(\mathbf{s})$= $Y(\mathbf{s}_i)$.
- For prediction locations $\mathbf{s}_i = \mathbf{s}$ the predicted value is $\hat{Y}(\mathbf{s})$= $Y(\mathbf{s}_i)$.
- In deterministic smoothing the standard error of the prediction $\hat{Y}(\mathbf{s})$ cannot be evaluated, because no distribution assumptions are made.

Covariance and Semi-variogram

- Most geo-spatial surfaces exhibit some degree of ***spatial persistence***. I.e. they exhibit ***positive spatial autocorrelation***.
- The auto-covariance $C(\mathbf{s}_i,\mathbf{s}_j)$ between two points is

$$C(\mathbf{s}_i,\mathbf{s}_j) = E\left[\left(Y(\mathbf{s}_i) - \underbrace{\boxed{\mu(\mathbf{s}_i)}}_{\text{residual}}\right)\cdot\left(Y(\mathbf{s}_j) - \mu(\mathbf{s}_j)\right)\right]$$ or in terms of the autocorrelation $$\rho(\mathbf{s}_i,\mathbf{s}_j) = \frac{C(\mathbf{s}_i,\mathbf{s}_j)}{\sigma(\mathbf{s}_i)\cdot\sigma(\mathbf{s}_j)}$$

- For an ***isotropic stationary*** spatial process the covariance $C(\mathbf{s}_i,\mathbf{s}_j)$ reduces to $C(\mathbf{s}_i,\mathbf{s}_j) \Rightarrow C(\mathbf{s}_i - \mathbf{s}_j) \Rightarrow C(h_{ij})$, which depends only on the distance h_{ij} between points $\mathbf{s}_i = (x_i, y_i)^T$ and $\mathbf{s}_j = (x_j, y_j)^T$.

  Note: ***spherical*** distances are better suited for small-scale maps to accommodate the curvature of the earth.

- For an ***isotropic stationary*** spatial process we get for the dissimilarity $Y(\mathbf{s}+\mathbf{h}) - Y(\mathbf{s})$:

  - $E[Y(\mathbf{s}+\mathbf{h}) - Y(\mathbf{s})] = E[Y(\mathbf{s}+\mathbf{h})] - E[Y(\mathbf{s})] = 0$

    for the individual expectations being ***zero*** or ***constant*** over space

    => Works for ***simple*** and ***ordinary*** kriging because any constant mean surface in the difference cancels out.

    => Consequently, we ***do not need to know the mean*** as long as it is constant at all locations

  - Otherwise we need to work with regression residuals, because their expectations are supposed to be identical within the study area: $E\left[\left(\varepsilon'(\mathbf{s}_i) + \varepsilon''(\mathbf{s}_i)\right) - \left(\varepsilon'(\mathbf{s}_j) + \varepsilon''(\mathbf{s}_j)\right)\right] = 0$

    => Works for ***universal*** kriging.

  - The variance of the distance at a distance $h = |\mathbf{h}|$ between two locations $Y(\mathbf{s}+\mathbf{h}) - Y(\mathbf{s})$ can be expressed as semi-variogram $\gamma(h)$ for simple and ordinary kriging:

$$\underbrace{Var[Y(\mathbf{s}+\mathbf{h}) - Y(\mathbf{s})]}_{=2\cdot\gamma(h)} = \underbrace{Var[Y(\mathbf{s}+\mathbf{h})]}_{=\sigma^2} + \underbrace{Var[Y(\mathbf{s})]}_{=\sigma^2} - 2\cdot\underbrace{C[Y(\mathbf{s}+\mathbf{h}), Y(\mathbf{s})]}_{=C(h)}$$

$$2\cdot\gamma(h) = 2\cdot\sigma^2 - 2\cdot C(h)$$

$$\gamma(h) = \sigma^2 - C(h)$$

because $Var(X \pm Z) = Var(X) + Var(Z) \pm 2\cdot C(X,Z)$.

- The term ***intrinsic stationarity*** refers to the stationarity of spatial difference $[Y(\mathbf{s}+\mathbf{h})-Y(\mathbf{s})]$ of regionalized random variable $Y(\mathbf{s})$ at two locations and not to the random variable $Y(\mathbf{s})$ itself. If the random field $Y(\mathbf{s})$ is weakly stationary then $[Y(\mathbf{s}+\mathbf{h})-Y(\mathbf{s})]$ must be also intrinsic stationary.

- The semi-variogram is defined as $\gamma(h)=\sigma^2-C(h) \Leftrightarrow C(h)=\sigma^2-\gamma(h)$, thus the covariance can be directly calculated from the semi-variogram. It only depends on the distance h between the locations $\mathbf{s}+\mathbf{h}$ and $\mathbf{s}$.

- Relationship between the covariance, correlation and the semi-variogram

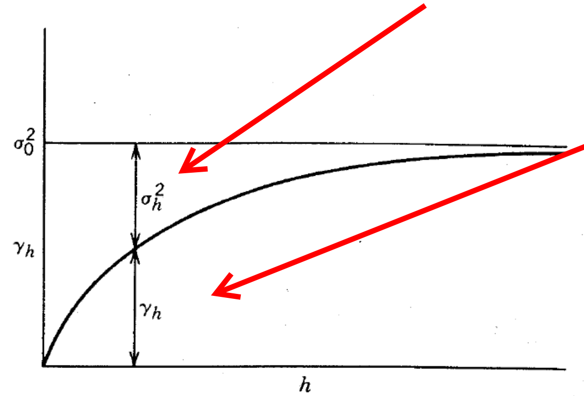


FIGURE 4.43 Relationship between semivariance γ and autocovariance σ^2 for a stationary regionalized variable. σ_0^2 is the variance of the observations, or the autocovariance at lag 0. For values of h beyond the range, $\gamma_h = \sigma_0^2$.

- Theoretical ***properties*** of the semi-variogram for intrinsic stationarity fields:

  - At $h=0 \Rightarrow \gamma(0)=\sigma^2-\underbrace{C(0)}_{=\sigma^2}=0$

- At $h \to \infty \Rightarrow \gamma(\infty) = \underbrace{\sigma^2}_{Var(Y(\mathbf{s}))} - \underbrace{C(\infty)}_{=0} = \sigma^2$. This is the ***sill***.
- The ***threshold distance*** h where the sill is reached is called ***range***.
- If at some distance range the sample data exhibit ***negative*** spatial autocorrelation then the semi-variogram overshoots the sill.

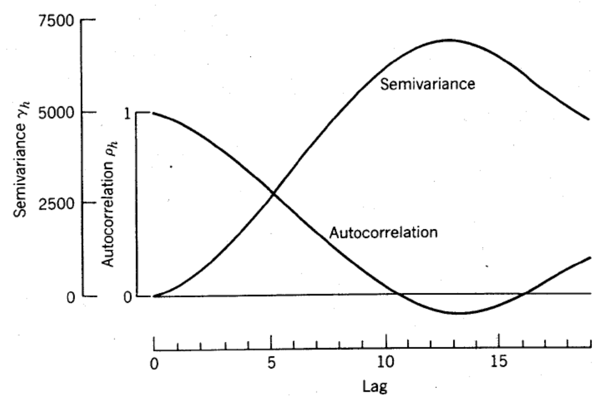


FIGURE 4.44 Relationship between semivariance γ_h and autocorrelation ρ_h for a stationary regionalized variable.

if the semi variance keep increasing, then would violate the stationary assumption (sill is variance.)

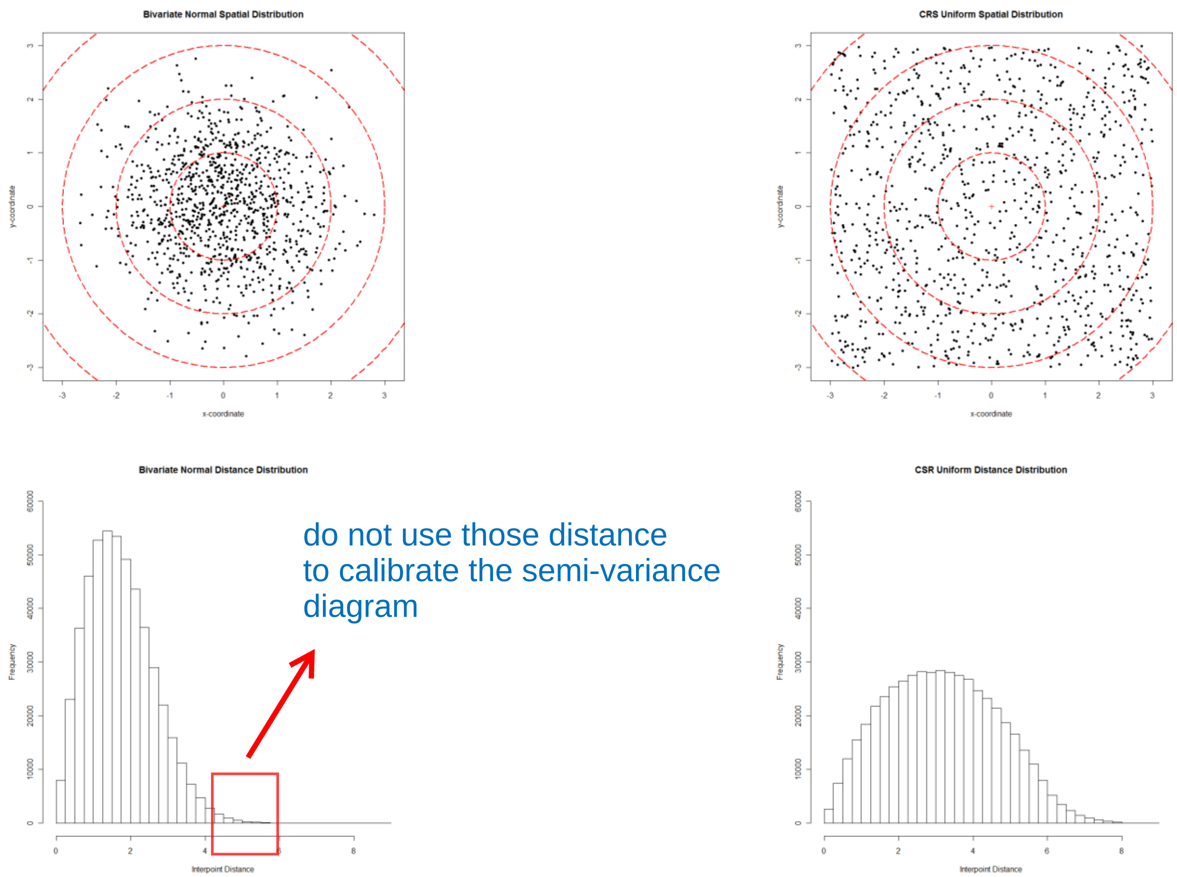
- If a semi-variogram is equal to the sill at all distances $h \geq 0$ then there is
  - no spatial autocorrelation in the underlying stochastic component $\varepsilon'(\mathbf{s}_i)$ and
  - $\gamma(h) = Var(\varepsilon''(\mathbf{s}_i)\ )$ for all distances h.
    Consequently, spatial interpolation becomes meaningless because $\varepsilon''(\mathbf{s}_i) \sim i.i.d.$, thus it does not exhibit a spatial relationship.
- How fast the semi-variogram approaches the sill, if at all, for large hs is an indication of the presence of non-stationarities.

    - One expects that spatial dependencies are fading out at some threshold distance h.
    - If this is not the case then non-stationarities are present.

- For non-stationary spatial fields the ***ergodasticity*** assumption is violated because effects of a regionalized random variable will not be forgotten even at very high distances.
- In theory, $\gamma(0+\delta)$ must almost equal zero for any incremental small distance value δ.
  If this is not the case we observe a nugget effect.
- The ***nugget effect*** arises because the regionalized variable is
  [a] so erratic over a short distance that the semi-variogram goes from zero to the level of the nugget effect in a distance less than the smallest sampling interval h_{min} between two points $\mathbf{s}+\mathbf{h}$ and $\mathbf{s}$.
  [b] or the spatial process exhibits a discontinuity over a short distances (e.g. moving from inside a nugget to the surrounding environment).
  measurement error
  [c] or our measurements $Y(\mathbf{s})$ at locations $\mathbf{s}$ are impacted by some error $Var[\varepsilon''(\mathbf{s}_i)] \neq 0$.
  => Implication: A ***predicted surface*** $\hat{Y}(\mathbf{s})$ does not need to go exactly through the observed value $Y(\mathbf{s}_i)$ at the sampling locations $\mathbf{s}_i$.

Estimation of the semi-variogram

- Distribution of inter-sample point distances within a bound study area



Notes:

- for higher distances the number of point pairs is decreasing again.
- For the normal distributed locations from the map center the squared distances h^2 are $h^2 \sim \chi^2_{df=2}$.

- Steps of variogram estimation:

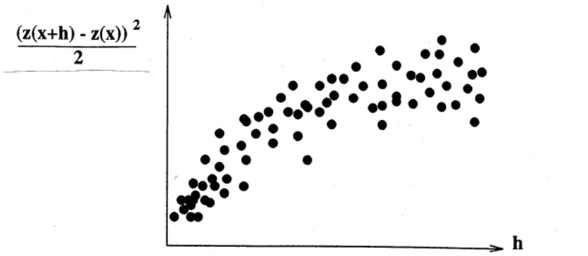


Figure 4.2: Plot of the dissimilarities γ^* against the spatial separation $\mathbf{h}$ of sample pairs: a variogram cloud.

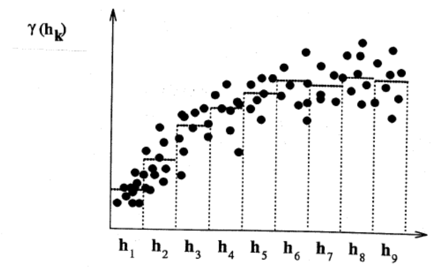


Figure 4.3: The experimental variogram is obtained by averaging the dissimilarities γ^* for given classes $\mathfrak{H}_k$.

none linear weighted regression

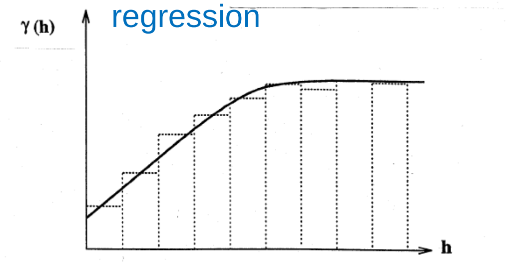


Figure 4.4: The sequence of average dissimilarities is fitted with a theoretical variogram function.

1. Plot pairwise squared semi-dissimilarities $\left(Y(\mathbf{s}+\mathbf{h})-Y(\mathbf{s})\right)^2/2$ for distances from zero to approx. half the ***diameter*** (spatial extend in either direction) of the study region.

2. Take the averages of $\left(Y(\mathbf{s}+\mathbf{h})-Y(\mathbf{s})\right)^2/2$ in ***distance bands***, which are also called ***bins***.

3. Fit a ***theoretical semi-variogram functional form*** to the mean levels in each bin:
   => Elementary textbooks suggest fitting it by visual inspection of the sill, range and nugget effect
   => Advanced methods use non-linear weighted statistical curve fitting techniques

   - The number of squared dissimilarities in each bin determines the weights.
   - A large number of pair-wise dissimilarities in a bin indicate higher precision of the estimate mean of the particular bin.

   => There are also approaches which do not require binning and use the pairwise squared semi-dissimilarities directly (see Bivand et al. p 228)

- What do we gain by estimating the semi-variogram instead of the covariance structure directly?

- Discrete intervals $[h_k - h_{k+1})$ are replaced by a function in the continuous distances h between two points.
- The semi-variogram is a simple function in just a few parameters.
- We do not need to know the mean of the spatial field as long as it is constant.
- We are not bound to the set of observed distances among observed ***sample point*** pairs.
  This gives us the flexibility to estimate the co-variation between any two points as long as we know their distance.
  Thus we can add ***prediction points*** and immediately calculate the covariation of these points with the given sample points.
- The semi-variogram is a smooth function and eliminates sampling variations from an empirically observed covariance matrix.

- The covariance matrix between the spatial objects can be derived from the theoretical forms of semi-variogram function $\gamma(h)$: $C(h) = \sigma^2 - \gamma(h)$
- However, only ***specific functional specifications*** for a semi-variogram function are permitted because the covariance matrix $C(h)$ has to satisfy the specific conditions:
  [a] ***symmetry*** and all its eigenvalues are positive
  [b] ***positive definiteness*** for any value of h.
- This leads to a selected number of ***feasible*** functional semi-variogram model specifications. For instance:

- <u>Spherical model:</u>

$$\gamma(h)=\begin{cases}0 & h=0\\ a+(\sigma^2-a)\cdot\left(\dfrac{3\cdot h}{2\cdot r}-\dfrac{h^3}{2\cdot r^3}\right) & 0<h\le r\\ \sigma^2 & otherwise\end{cases}$$

  a is the nugget effect, σ^2 the sill and r is the range.
  Discuss function's behavior at $h=0$, $h=r$ and beyond.

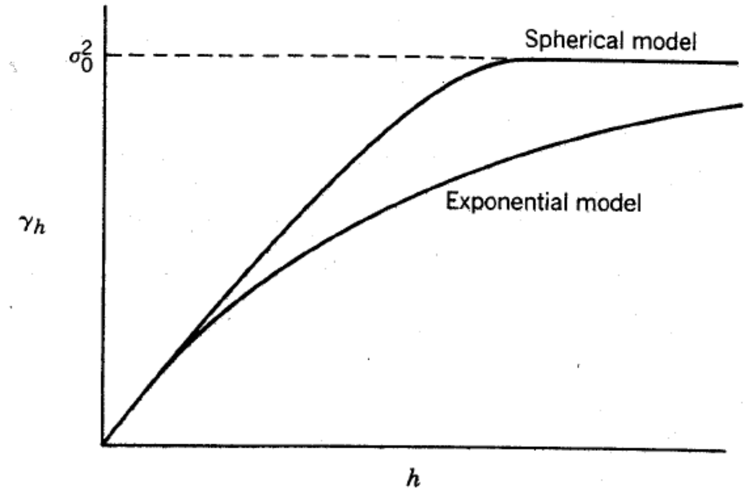


FIGURE 4.47 Exponential and spherical models of semivariogram. Both models have same initial slope and sill. After Clark (1979).

- <u>Exponential model:</u>

$$\gamma(h)=\begin{cases}0 & h=0\\ a+(\sigma^2-a)\cdot\left[1-\exp(-3h/r)\right] & h>0\end{cases}$$

  The exponential model will never exactly reach the sill, however, for large *h* it will come extremely close to it. It can model mild non-stationarities.

- For more models and a discussion of their properties see Waller & Gotway pp277-280.

Extensions:

- The semi-variogram model can be relaxed to ***accommodate simple forms anisotropy*** (beyond the scope of this lecture). Here the sill or the range may differ depending on the direction of the distance measurement (see the Geo-Statistical Analyst in ArcGIS)

- The semi-variogram can be ***generalized to a linear combination*** of several semi-variogram functions such as one for shorter ranges and one for longer ranges, e.g., $\gamma(h) = \lambda \cdot \gamma_1(h) + (1-\lambda) \cdot \gamma_2(h)$ (see Bivand et al. p 224 and **`gstat::vgm(…, add.to=vgmComponent)`**) divided data into two sets(short distance and long distance)
- One needs to question, however, if the more complex semi-variogram models add sufficiently new information and what their underlying theoretical justification is?