Theoretical Semivariogram Models (Chapter 5.5.3 - Bailey & Gatrell)

Thus far, we have studied various empirical estimates of the semivariogram, both classical & robust, isotropic & anisotropic. These are not ideal for the purpose of *modeling* spatial data for the following reasons.

1. The estimated semivariogram only gives estimates at particular lag distances. Later, when we look at predicting a value of the response variable Y at a new location (say Y_0), we will need $\text{Cov}(Y_i, Y_0)$ for each site $i = 1, \ldots, n$. We will likely not have the required covariance estimates since the distances between Y_0 and other sites are probably not the same as the lag distances used.

2. The matrix of sample covariances, $\widehat{\boldsymbol{C}}$, may not be a valid covariance matrix. In order to be valid, a matrix must contain covariances such that the variance of any linear combination of the Y_i's is nonnegative (i.e.: For any $n \times 1$ vector $\boldsymbol{l}$, Var $(\boldsymbol{l'Y}) = \boldsymbol{l'}\widehat{\boldsymbol{C}}\boldsymbol{l} \geq 0$). Matrices with this property are said to be nonnegative definite. If the sample covariances are used to estimate $\boldsymbol{C}$, there is no guarantee that this nonnegative definite condition is satisfied.

3. The sample covariances do not necessarily decrease smoothly with distance, as some are estimated based on more pairs of points than others. If you believe that the shape of the covariance function should be smooth theoretically, a better estimate can be obtained by pooling the information from different lags together and using a smooth curve to describe the covariance function.

Some of the more commonly used semivariogram models for describing the covariance function are outlined and discussed below. For each model, the functional form is given as well as a plot of the function. These models are of two basic types: those with a sill, and those without. Models without a sill are often referred to as transition models. There are many additional models other than those given here, but these are some of the more useful models and are all supported by **R**.

For most of these models, the parameters have the following interpretation:

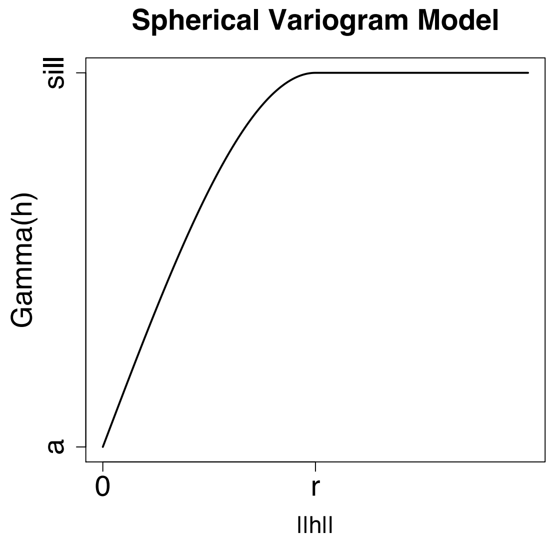
$$a = \text{nugget effect}, \qquad \sigma^2 = \text{sill}, \qquad r = \text{range}$$

$\boxed{\text{Spherical Model}}$: This model is valid for dimensions 1, 2, & 3 only.

$$\gamma(\boldsymbol{h}) = \left\{ \begin{array}{ll} 0 & \boldsymbol{h} = \boldsymbol{0} \\ a + (\sigma^2 - a)\left\{\frac{3}{2}\frac{||\boldsymbol{h}||}{r} - \frac{1}{2}\left(\frac{||\boldsymbol{h}||}{r}\right)^3\right\} & 0 < ||\boldsymbol{h}|| \leq r \\ a + (\sigma^2 - a) & ||\boldsymbol{h}|| \geq r \end{array} \right\},$$

where $a \geq 0$, $\sigma^2 \geq a$, $r \geq 0$.

- The spherical model reaches the sill at $||\boldsymbol{h}|| = r$. The model looks nearly linear at small lags. The tangent at $||\boldsymbol{h}|| = 0$ would cross the sill at a distance of $\frac{2}{3}r$.
- The spherical model is particularly good for modeling spatial correlation which decreases approximately linearly with the separation distance, and is assumed to be zero beyond a certain distance. This is probably the most commonly used variogram structure in practice.



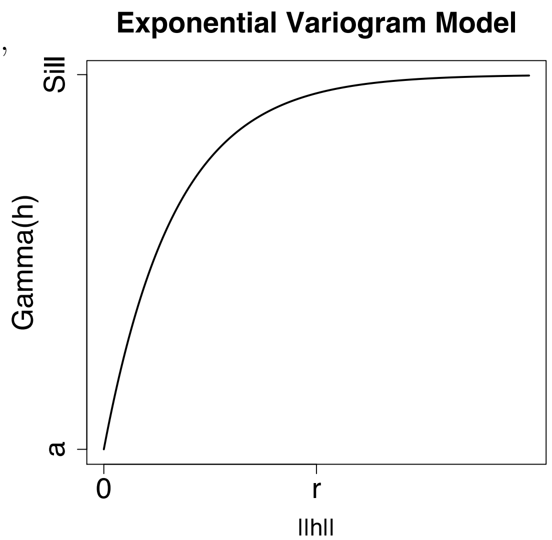
$\boxed{\text{Exponential Model}}$: This model is valid for all dimensions.

$$\gamma(\boldsymbol{h}) = \left\{ \begin{array}{ll} 0 & \boldsymbol{h} = \boldsymbol{0} \\ a + (\sigma^2 - a)\left(1 - e^{-3||\boldsymbol{h}||/r}\right) & \boldsymbol{h} \neq \boldsymbol{0} \end{array} \right\},$$

where $a \geq 0$, $\sigma^2 \geq a$, $r \geq 0$.

- The exponential model reaches the sill only asymptotically, as $||\boldsymbol{h}|| \to \infty$. Consequently, the model is parameterized so that:

  r = lag distance at which $\gamma(\boldsymbol{h}) = .95(\sigma^2)$.



- The exponential model has a similar shape to the spherical model but reaches the sill more quickly. The tangent at $||\boldsymbol{h}|| = 0$ would cross the sill at a distance of $\frac{1}{3}r$.

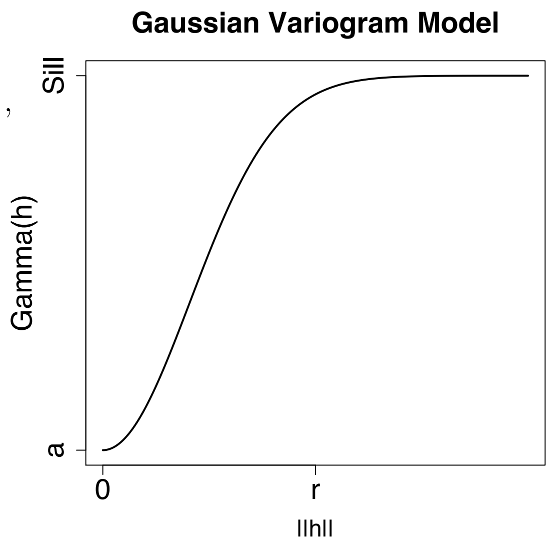
$\boxed{\text{Gaussian Model}}$: This model is valid for all dimensions.

$$\gamma(\boldsymbol{h}) = \left\{ \begin{array}{ll} 0 & \boldsymbol{h} = \boldsymbol{0} \\ a + (\sigma^2 - a)\left(1 - e^{-3\left(\frac{||\boldsymbol{h}||}{r}\right)^2}\right) & \boldsymbol{h} \neq \boldsymbol{0} \end{array} \right\},$$

where $a \geq 0$, $\sigma^2 \geq a$, $r \geq 0$.

- As with the exponential model, the Gaussian model reaches the sill only asymptotically, and is also parameterized so that:

  r = lag distance at which $\gamma(\boldsymbol{h}) = .95(\sigma^2)$.



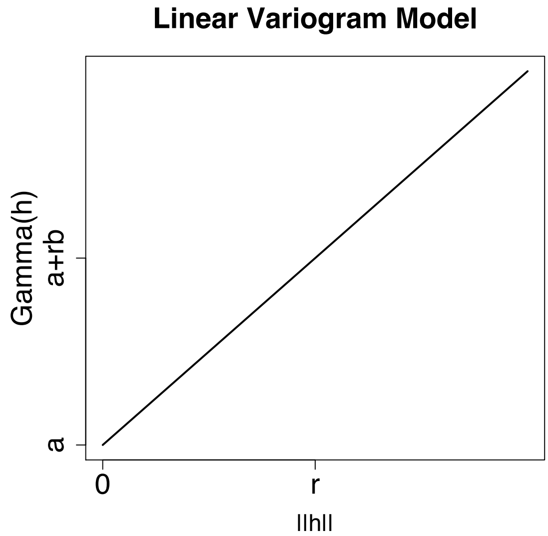
- The Gaussian model is used when the data exhibit strong continuity at short lag distances (i.e.: when the spatial correlation between two nearby points is very high).

- As seen above, the Gaussian semivariogram function is S-shaped, much like one-half of the Gaussian (normal) distribution. This function gives correlations of nearly one for early lags, is concave upward until the inflection point of $r/\sqrt{6}$, at which point the curve begins to flatten.

$\boxed{\text{Linear Model}}$: This model is valid for all dimensions.

$$\gamma(\boldsymbol{h}) = \left\{ \begin{array}{ll} 0 & \boldsymbol{h} = \boldsymbol{0} \\ a + b||\boldsymbol{h}|| & \boldsymbol{h} \neq \boldsymbol{0} \end{array} \right\},$$

where $a \geq 0$, $b \geq 0$.

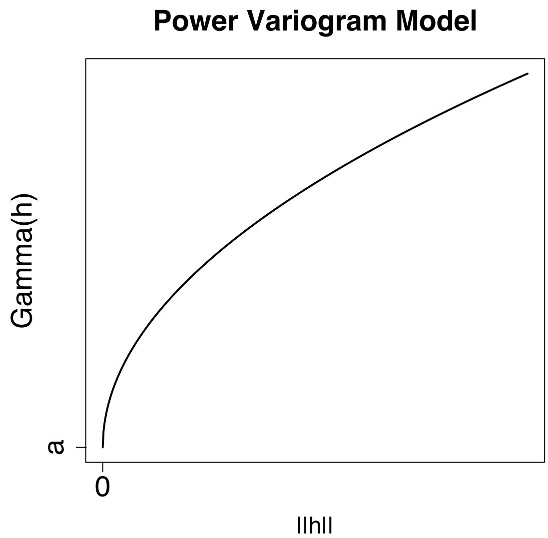


- The linear variogram model does not reach a sill, so the second parameter in this model is labeled "b" instead of "σ^2". When is this model appropriate?

$\boxed{\text{Power Model}}$: This model is valid for all dimensions.

$$\gamma(\boldsymbol{h}) = \left\{ \begin{array}{ll} 0 & \boldsymbol{h} = \boldsymbol{0} \\ a + b||\boldsymbol{h}||^{\lambda} & \boldsymbol{h} \neq \boldsymbol{0} \end{array} \right\},$$

where $0 \leq \lambda < 2$, $a \geq 0$, $b \geq 0$.



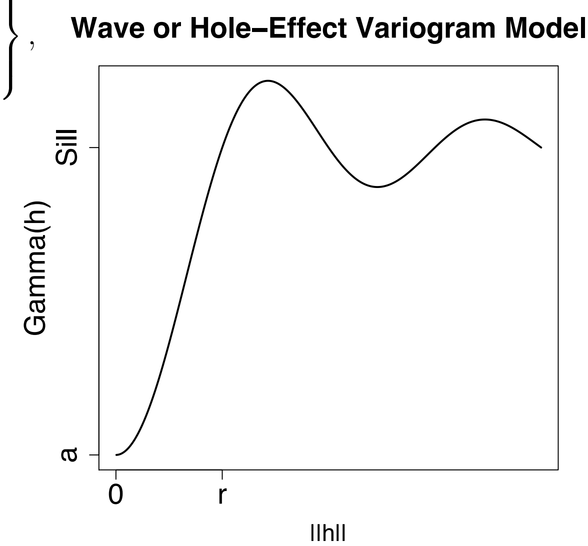
- As with the linear variogram model, the power model does not reach a sill, and so labels the second parameter "b" instead of "σ^2". Any power between 0 and 2 may be used to construct a valid power variogram model. In the figure above, a power of 0.5 was used.

- Both the power and linear variogram models are appropriate if there is long-range correlation or if samples were not collected at a sufficiently large distance to reach the point where pairs of points are uncorrelated.

Wave or Hole-Effect Model: This model is valid for dimensions 1, 2, & 3 only.

$$\gamma(\boldsymbol{h}) = \begin{Bmatrix} 0 & \boldsymbol{h} = \boldsymbol{0} \\ a + (\sigma^2 - a)\left\{1 - \dfrac{\sin(\pi||\boldsymbol{h}||/r)}{\pi||\boldsymbol{h}||/r}\right\} & \boldsymbol{h} \neq \boldsymbol{0} \end{Bmatrix},$$

where $a \geq 0$, $\sigma^2 \geq a$, $r \geq 0$.



- The wave or hole-effect model is generally used when there is some periodicity in the data resulting in a hole effect. We have seen the occurrence of such an effect in the Walker Lake U-data, as a result of preferential sampling.
- One other common example of where the hole-effect might occur is in agricultural field studies where there are rows and furrows. The correlation between observations in rows may be high, whereas that between observations in a row and in the furrow may be lower, resulting in an alternating pattern of high and low correlations whose amplitude dies down as the distance between observations gets larger.
- The range in the hole-effect model is the shortest distance at which the semivariogram equals σ^2. This will occur on the initial rise in the variogram function. Because of the periodicity, this model contains both positive *and* negative correlations.

There are other types of variogram models, not mentioned here, which could also be considered. Two of these include:

1. Matern class of models: These models are highly flexible models around the nugget effect, and so are best for modeling complicated behavior near the nugget effect. A good reference for this class of models is: Handcock & Stein, *Technometrics*, 1993.
2. Rational Quadratric Model: This model has near-one correlations for small lags (much like the Gaussian model), but rises much more steeply than the Gaussian model.

Nested Variogram Structures: If the empirical semivariogram does not seem to follow any of the standard structures, it is possible to combine structures to obtain a variogram with the characteristics of more than one of the standard structures. A linear combination of valid variogram structures is also a valid variogram model and is called a **nested variogram structure**.

Letting $\gamma_0(\boldsymbol{h}), \gamma_1(\boldsymbol{h}), \ldots, \gamma_k(\boldsymbol{h})$ be valid variogram structures, any function of the form:

$$\gamma(\boldsymbol{h}) = \sum_{i=0}^{k} w_i \gamma_i(\boldsymbol{h}) \qquad \text{is a nested variogram structure.}$$

Example: Suppose we have a nested structure with a nugget effect $w_0 = 1$, and two additional structures, one of exponential form and one of spherical form. The components of this semivariogram model are:

1. Nugget Effect: $w_0 = 1$.
2. First Structure: exponential with contribution $w_1 = 1$ and range $r_1 = 2$.
$$w_1 = 1,\ r_1 = 2,\ \gamma_1(\boldsymbol{h}) = \left\{ \begin{array}{ll} 0 & \boldsymbol{h} = \boldsymbol{0} \\ 1 - e^{-3||\boldsymbol{h}||/r_1} & \boldsymbol{h} \neq \boldsymbol{0} \end{array} \right\}.$$
3. Second Structure: spherical with contribution $w_2 = 2$ and range $r_2 = 10$.
$$w_1 = 2,\ r_2 = 10,\ \gamma_2(\boldsymbol{h}) = \left\{ \begin{array}{ll} 0 & \boldsymbol{h} = \boldsymbol{0} \\ 1.5\left(\frac{||\boldsymbol{h}||}{r_2}\right) - 0.5\left(\frac{||\boldsymbol{h}||}{r_2}\right)^3 & 0 < ||\boldsymbol{h}|| \leq r_2 \\ 1 & ||\boldsymbol{h}|| > r_2 \end{array} \right\}.$$

Combining these structures then, this variogram model is:

$$\begin{aligned} \gamma(\boldsymbol{h}) &= w_0 + w_1\gamma_1(\boldsymbol{h}) + w_2\gamma_2(\boldsymbol{h}) \\ &= \left\{ \begin{array}{ll} 0 & \text{for } \boldsymbol{h} = \boldsymbol{0} \\ 1 + \left(1 - e^{-3||\boldsymbol{h}||/2}\right) + 2\left(1.5\left(\frac{||\boldsymbol{h}||}{10}\right) - 0.5\left(\frac{||\boldsymbol{h}||}{10}\right)^3\right) & \text{for } 0 < ||\boldsymbol{h}|| \leq 10 \\ 1 + \left(1 - e^{-3||\boldsymbol{h}||/2}\right) + 2 & \text{for } ||\boldsymbol{h}|| > 10 \end{array} \right\}. \end{aligned}$$

- In this model, the sill is $w_0 + w_1 + w_2 = 4$. The range is the distance, $||\boldsymbol{h}||$, at which $\gamma(\boldsymbol{h}) = (.95)(4) = 3.8$ (i.e.: the range is 7.3).
- The variogram above lies between the exponential variogram $1 + 3(1 - e^{-3||\boldsymbol{h}||/2})$, and the spherical variogram $1 + 3\left(1.5\left(\frac{||\boldsymbol{h}||}{10}\right) - 0.5\left(\frac{||\boldsymbol{h}||}{10}\right)^3\right)$. Why would we use such a nested variogram?

- Plots of the nested variogram, as well as the exponential and spherical variograms from which the nested variogram is structured, are given below: