Kriging Interpolation

GRG 460G - Advanced GIS

Eugenio Y. Arima

Department of Geography and the Environment
University of Texas at Austin

Motivation

- Spatial data often observed at irregular locations
- Need to predict values at unobserved locations
- Interpolation should:
    - Use spatial dependence
    - Provide uncertainty estimates
    - Be statistically optimal
- Kriging satisfies all three

What is Kriging?

Definition

Kriging is a spatial interpolation method that provides the *Best Linear Unbiased Estimator (BLUE)* of a spatial random field (provided assumptions hold).

- Originates from geostatistics (D. G. Krige)
- Treats spatial data as realizations of a stochastic process
- Explicitly models spatial covariance

Kriging Predictor

Suppose observations with known values Z at locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$.
We want to predict the Z value at location s_0

Linear Predictor

$$\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^{n} \lambda_i Z(\mathbf{s}_i)$$

- Predicted value is a linear combination of nearby known values
- Weights λ_i chosen optimally (minimum variance)
- How do we find the λ's? Depend on spatial covariance (or variogram)

$$\boldsymbol{\lambda} = \boldsymbol{C}^{-1}\boldsymbol{c}_0$$

Kriging Variance

- Method also returns the uncertainty of our prediction
- Depends only on sampling geometry and variogram
- Not on observed values

Formula (in matrix format) is:

$$\sigma_{s_0}^2 = \sigma^2 - \mathbf{c}^\top(s_0)\mathbf{C}^{-1}\mathbf{c}(s_0),$$

where $\sigma^2 = C(0)$ is the process variance.

Types of Kriging

- **Simple Kriging**: known mean
- **Ordinary Kriging**: unknown constant mean
- **Universal Kriging**: mean modeled as regression
- **Co-kriging**: multiple correlated variables
- **Empirical Bayesian Kriging**: subsetting and simulation that automates certain tasks

Simple Kriging

Assumptions

- Mean is known and constant: $\mathbb{E}[Z(\mathbf{s})] = m$
- Covariance function $C(\mathbf{h})$ known

Predictor

$$\hat{Z}(\mathbf{s}_0) = m + \sum_{i=1}^{n} \lambda_i \big(Z(\mathbf{s}_i) - m\big)$$

Objective Function

Minimize the prediction variance:

$$\min_{\boldsymbol{\lambda}} \ \mathrm{Var}\Big(Z(\mathbf{s}_0) - \hat{Z}(\mathbf{s}_0)\Big)$$

How do we get to that formula? Sketch...

I will not show the proof here, only a general outline.
If we expand the variance formula and take expectations, we obtain:

$$E\left[z(s_0) - m\right]^2 - 2\sum_{i=1}^{n} \lambda_i \, E\left[z(s_i) - m\right]\left[z(s_0) - m\right] +$$

$$+ \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \, E\left[z(s_i) - m\right]\left[z(s_j) - m\right]$$

which are the covariances

$$C(0) - 2\sum_{i=1}^{n} \lambda_i \, C(s_0, s_i) + \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i \lambda_j \, C(s_i, s_j)$$

To minimize, take the partial derivative w.r.t. λ and set equal zero, solve for λ

Much simplified in matrix notation

Minimization Problem in Matrix notation becomes

$$C\boldsymbol{\lambda} = \mathbf{c}_0$$

$$\boldsymbol{\lambda} = C^{-1}\mathbf{c}_0$$

- C: covariance matrix between observations
- C^{-1}: inverse of matrix C
- $\mathbf{c}_0$: covariance vector between observations and $\mathbf{s}_0$
- λ: vector of weights

No unbiasedness constraint is needed because m is known in simple kriging.

Ordinary Kriging

Assumptions

- Mean is unknown but constant
- Stationary covariance or variogram

Predictor

$$\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^{n} \lambda_i Z(\mathbf{s}_i)$$

Objective Function

Minimize the prediction variance:

$$\min_{\boldsymbol{\lambda}} \operatorname{Var}\Big(\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)\Big)$$

Cont.

Unbiasedness Constraint

$$\sum_{i=1}^{n} \lambda_i = 1$$

Minimization System

$$\begin{bmatrix} \boldsymbol{C} & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \mu \end{bmatrix} = \begin{bmatrix} \boldsymbol{c}_0 \\ 1 \end{bmatrix}$$

Universal Kriging

Assumptions

- Mean is spatially varying:

$$\mathbb{E}[Z(\mathbf{s})] = \mathbf{x}(\mathbf{s})^\top \boldsymbol{\beta}$$

- $\boldsymbol{\beta}$ unknown

Predictor

$$\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^{n} \lambda_i Z(\mathbf{s}_i)$$

Objective Function

Minimize:

$$\min_{\boldsymbol{\lambda}} \ \mathrm{Var}\Big(\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)\Big)$$

Cont.

Unbiasedness Constraints

$$\sum_{i=1}^{n} \lambda_i \mathbf{x}(\mathbf{s}_i) = \mathbf{x}(\mathbf{s}_0)$$

Minimization System

$$\begin{bmatrix} \boldsymbol{C} & X \\ X^\top & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\mu} \end{bmatrix} = \begin{bmatrix} \boldsymbol{c}_0 \\ \mathbf{x}(\mathbf{s}_0) \end{bmatrix}$$

Co-Kriging

Assumptions

- Multiple correlated variables:

$$Z^{(1)}(\mathbf{s}), \ldots, Z^{(p)}(\mathbf{s})$$

- Cross-covariances are modeled

Predictor

$$\hat{Z}^{(1)}(\mathbf{s}_0) = \sum_{j=1}^{p} \sum_{i=1}^{n_j} \lambda_{ij} Z^{(j)}(\mathbf{s}_i)$$

Objective Function

Minimize:

$$\min_{\{\lambda_{ij}\}} \operatorname{Var}\left(\hat{Z}^{(1)}(\mathbf{s}_0) - Z^{(1)}(\mathbf{s}_0) \right)$$

Cont.

Unbiasedness Constraints

Depend on kriging type (simple, ordinary, or universal), applied across all variables.

Minimization System

Block system involving:

$$\begin{bmatrix} C^{(11)} & \cdots & C^{(1p)} \\ \vdots & \ddots & \vdots \\ C^{(p1)} & \cdots & C^{(pp)} \end{bmatrix}$$

where each block is a (cross-)covariance matrix.

Why Use Kriging?

- Statistically optimal (BLUE)
- Explicit uncertainty quantification
- Flexible covariance modeling

Limitations

- Requires variogram estimation
- Sensitive to stationarity assumptions
- Computationally expensive for large n

Summary

- Kriging is model-based spatial interpolation
- Relies on spatial covariance structure
- Provides both predictions and uncertainty
- Foundation of modern geostatistics

Worked Example: Setup

Observed data: five points $s_1, \ldots, s_5$
Predict value s_0
Assume:

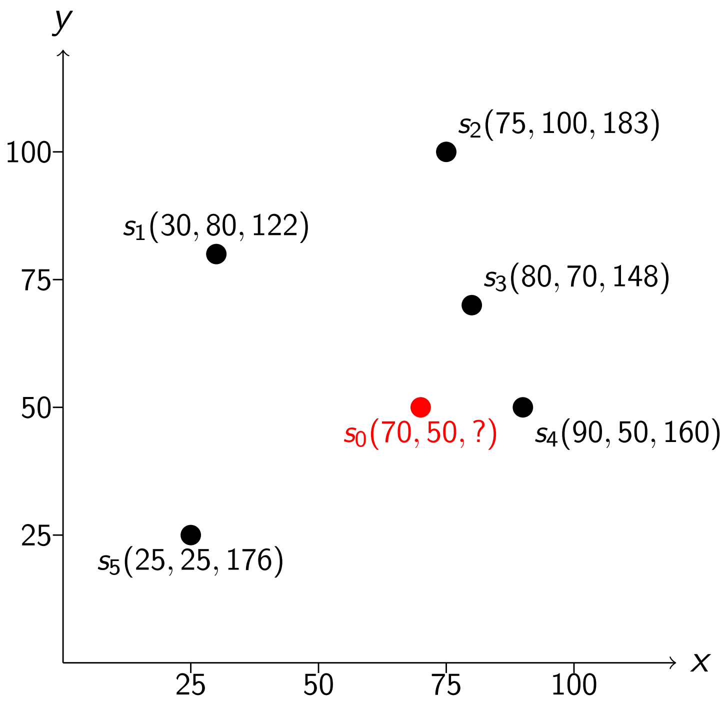
- Ordinary kriging (unknown mean but constant)
- Suppose we know the Covariance structure to be an exponential variogram model having a sill of 20.0 and a range of 100:

$$C(h) = 20e^{-3h/100}$$

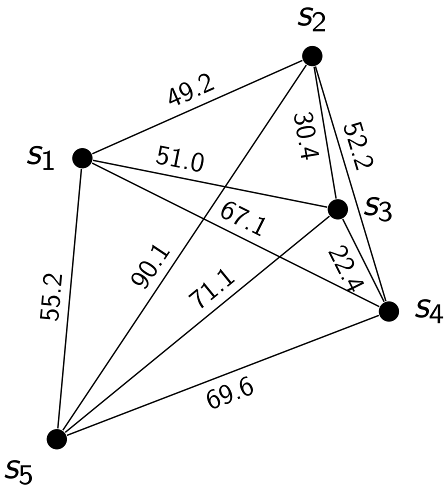
- h is the distance between a pair of points
- We want to find $\hat{\lambda}s$ by solving the matrix system:

$$\boldsymbol{C\lambda} = \boldsymbol{c_0} \text{ s.t. } \sum \lambda_i = 1$$

Point to be interpolated s_0 and known value points $s_1, \ldots, s_5$



Distances between known-value points



Covariance matrix from the known points

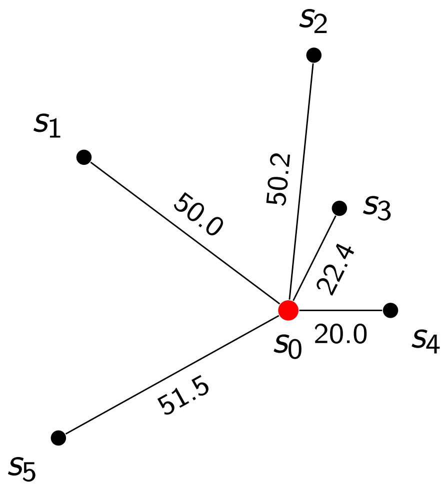
Recall that we know the Covariance structure to be an exponential variogram model having a sill of 20.0 and a range of 100:

$$C(h) = 20e^{-3h/100}$$

Using the given pairwise distances between $s_1, \dots, s_5$, we form the matrix $\mathbf{C} = [C(h(s_i, s_j))]$ with diagonal entries set to 20 because the distance h from a point to itself is zero.

$$\mathbf{C} = \begin{array}{c|ccccc} & s_1 & s_2 & s_3 & s_4 & s_5 \\ \hline s_1 & 20.0 & 4.57 & 4.33 & 2.67 & 3.82 \\ s_2 & 4.57 & 20.0 & 8.03 & 4.18 & 1.34 \\ s_3 & 4.33 & 8.03 & 20.0 & 10.21 & 2.37 \\ s_4 & 2.67 & 4.18 & 10.21 & 20.0 & 2.48 \\ s_5 & 3.82 & 1.34 & 2.37 & 2.48 & 20.0 \end{array}$$

Distances from s_0 to known value points



Covariance vector with the interpolation point

Similarly, we construct the covariance vector $\mathbf{c}(s_0)$ by plugging in the distances from the known points $s_1, \ldots, s_5$ to the point to be interpolated s_0 into the covariance function

$$C(h) = 20e^{-3h/100}.$$

Given the distances h

$$\begin{aligned} d(s_0, s_1) &= 50.0, \\ d(s_0, s_2) &= 50.2, \\ d(s_0, s_3) &= 22.4, \\ d(s_0, s_4) &= 20.0, \\ d(s_0, s_5) &= 51.5, \end{aligned}$$

we obtain

$$\mathbf{c}(s_0) = \begin{bmatrix} C(50) \\ C(50.2) \\ C(22.4) \\ C(20.0) \\ C(51.5) \end{bmatrix} = \begin{bmatrix} 4.46 \\ 4.44 \\ 10.21 \\ 10.98 \\ 4.27 \end{bmatrix}.$$

Ordinary kriging linear system

For ordinary kriging, the weights satisfy the augmented system

$$\begin{bmatrix} \mathbf{C} & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \mu \end{bmatrix} = \begin{bmatrix} \mathbf{c}(s_0) \\ 1 \end{bmatrix},$$

where μ is a Lagrange multiplier enforcing $\sum \lambda = 1$

Thus, the full system is

$$\begin{bmatrix} 20.00 & 4.57 & 4.33 & 2.67 & 3.82 & 1 \\ 4.57 & 20.00 & 8.03 & 4.18 & 1.34 & 1 \\ 4.33 & 8.03 & 20.00 & 10.21 & 2.37 & 1 \\ 2.67 & 4.18 & 10.21 & 20.00 & 2.48 & 1 \\ 3.82 & 1.34 & 2.37 & 2.48 & 20.00 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \mu \end{bmatrix} = \begin{bmatrix} 4.46 \\ 4.44 \\ 10.21 \\ 10.98 \\ 4.27 \\ 1 \end{bmatrix}.$$

Cont.

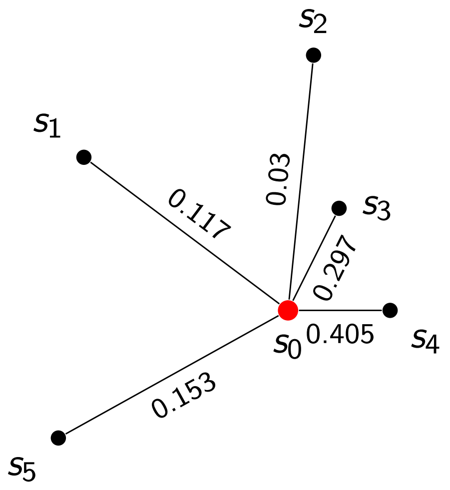
Solving the augmented system

$$\begin{bmatrix} \boldsymbol{\lambda} \\ \mu \end{bmatrix} = \begin{bmatrix} \mathbf{C} & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}(s_0) \\ 1 \end{bmatrix}$$

yields

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \mu \end{bmatrix} = \begin{bmatrix} 0.1171 \\ 0.0294 \\ 0.2967 \\ 0.4050 \\ 0.1529 \\ -0.961 \end{bmatrix}.$$

Weights



Estimated value at s_0

Using the previously computed ordinary kriging weights and the observed values

$$\begin{bmatrix} Z(s_1) \\ Z(s_2) \\ Z(s_3) \\ Z(s_4) \\ Z(s_5) \end{bmatrix} = \begin{bmatrix} 122 \\ 183 \\ 148 \\ 160 \\ 176 \end{bmatrix},$$

recall that the ordinary kriging estimator is

$$\hat{Z}(s_0) = \sum_{i=1}^{5} \lambda_i Z(s_i).$$

$$\hat{Z}(s_0) = (0.1171)(122) + (0.0294)(183) + (0.2967)(148) + \\ +(0.4050)(160) + (0.1519)(176).$$

$$\boxed{\hat{Z}(s_0) = 155.1}$$

Kriging variance of the prediction at s_0

The kriging variance quantifies the uncertainty of the prediction at location s_0. It is given by

$$\sigma^2_{s_0} = \sigma^2 - \mathbf{c}^\top(s_0)\mathbf{C}^{-1}\mathbf{c}(s_0),$$

where $\sigma^2 = C(0)$ is the process variance.

In this example,

$$\sigma^2 = 20, \qquad \mathbf{c}(s_0) = \begin{bmatrix} 4.46 \\ 4.44 \\ 10.21 \\ 10.98 \\ 4.27 \end{bmatrix}, \qquad \boldsymbol{\lambda} = \mathbf{C}^{-1}\mathbf{c}(s_0).$$

Thus,

$$\mathbf{c}^\top(s_0)\mathbf{C}^{-1}\mathbf{c}(s_0) = \mathbf{c}^\top(s_0)\boldsymbol{\lambda}.$$

Cont.

Substituting the values (the Lagrange multiplier cancels out...not included here),

$$\mathbf{c}^\top(s_0)\boldsymbol{\lambda} = (4.46)(0.1171) + (4.44)(0.0294) + (10.21)(0.2967) +$$

$$+ (10.98)(0.4050) + (4.27)(0.1519) = 8.78.$$

Therefore, the kriging variance at the estimated point is

$$\boxed{\sigma^2_{s_0} = 20 - 8.78 = 11.22}$$

leading to a 95% confidence interval of:

$$155.1 \pm \sqrt{11.22} * 1.96 = 155.1 \pm 6.56$$

What did we learn?

- Prediction is a weighted average satisfying unbiasedness
- The only information we need to calculate the weights is the (semi) variance (or covariance) of our data
- This is achieved through variography
- We need to use our data to create an empirical variogram

Variography: Definition

Definition

Variography is the geostatistical methodology used to quantify and model the spatial dependence of a regionalized variable through the *variogram*, which describes how data similarity decreases with distance.

Experimental Variogram

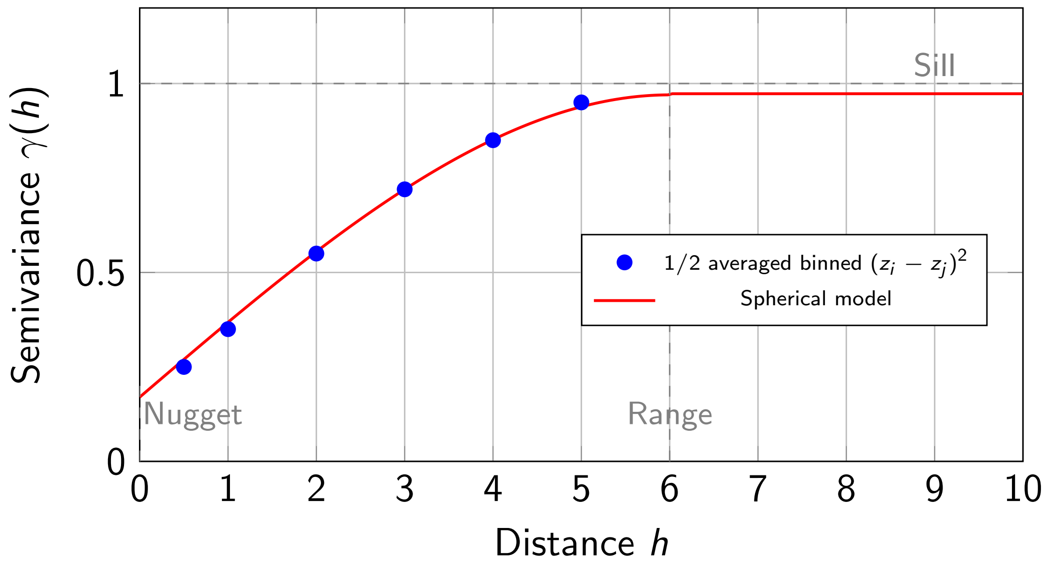
$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(\mathbf{s}_i) - Z(\mathbf{s}_i + \mathbf{h})]^2$$

where h is the separation distance (lag) and $N(h)$ is the number of data pairs.

Variography Process

- Compute the experimental variogram from sampled data
  - Pass through the center of the cloud of binned values
  - Pass as closely as possible to the averaged values
- Knowledge of phenomenon may dictate the shape of the model as well as its nugget, range, sill, etc.
- Analyze anisotropy and directional effects (more on this later)
- Fit a theoretical variogram model (e.g., spherical, exponential, Gaussian)
- Validate the model using cross-validation
- Use the variogram in kriging or other spatial predictions

Empirical Variogram and Fitted Model



Note: x-axis and y-axis values are for display purposes only

The Variogram

Definition

$$\gamma(\mathbf{h}) = \frac{1}{2}\mathbb{E}[(Z(\mathbf{s}) - Z(\mathbf{s}+\mathbf{h}))^2]$$

- Measures spatial dissimilarity
- Primary tool for modeling spatial dependence
- Related to covariance:

  $$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h})$$

  where $C(0)$ is the global variance σ^2, a constant.
- By tradition, we typically model the (semi)variance $\gamma(\mathbf{h})$ but $C(\mathbf{h})$ would be just as easy.

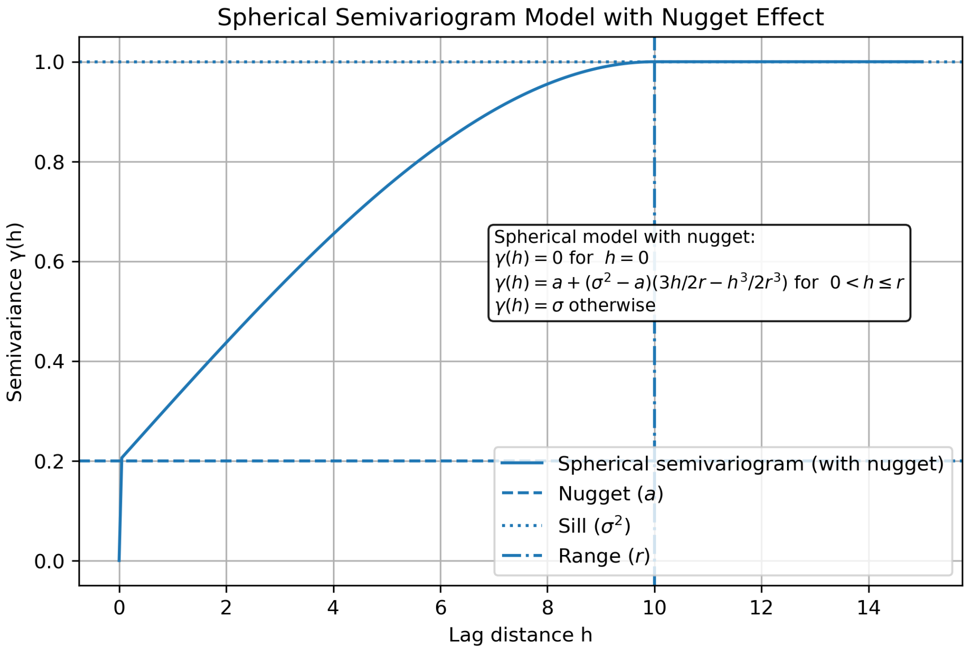
Variogram Components

- **Nugget**: microscale variation or measurement error
- **Sill**: total variance
- **Range**: distance beyond which dependence is negligible

Common Models

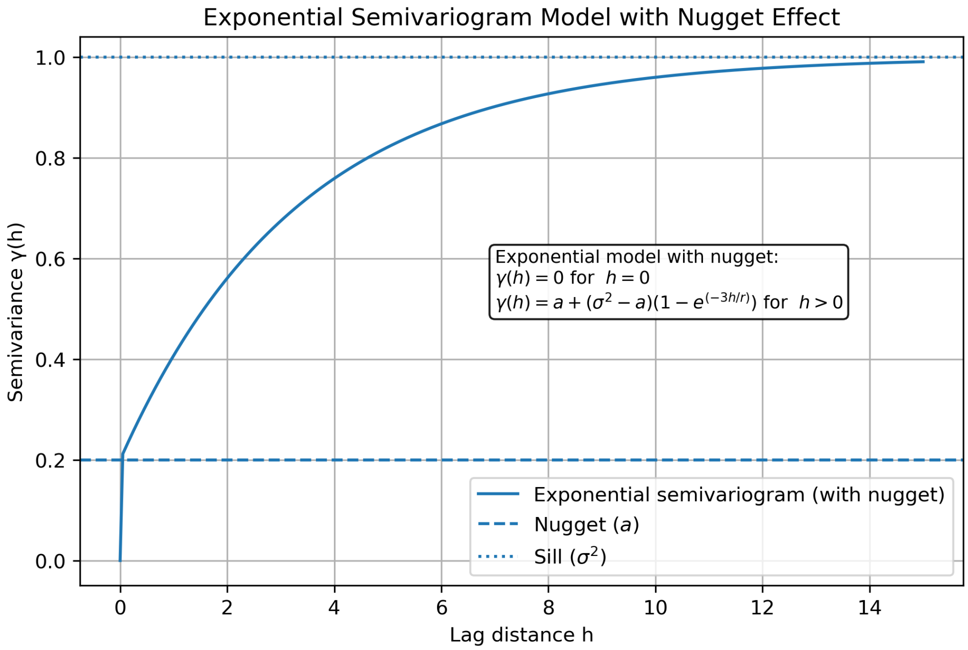
- Spherical
- Tetra-, Penta-spherical
- Exponential
- Gaussian
- Circular
- K-, J-Bessel
- Many others...

Spherical Variogram



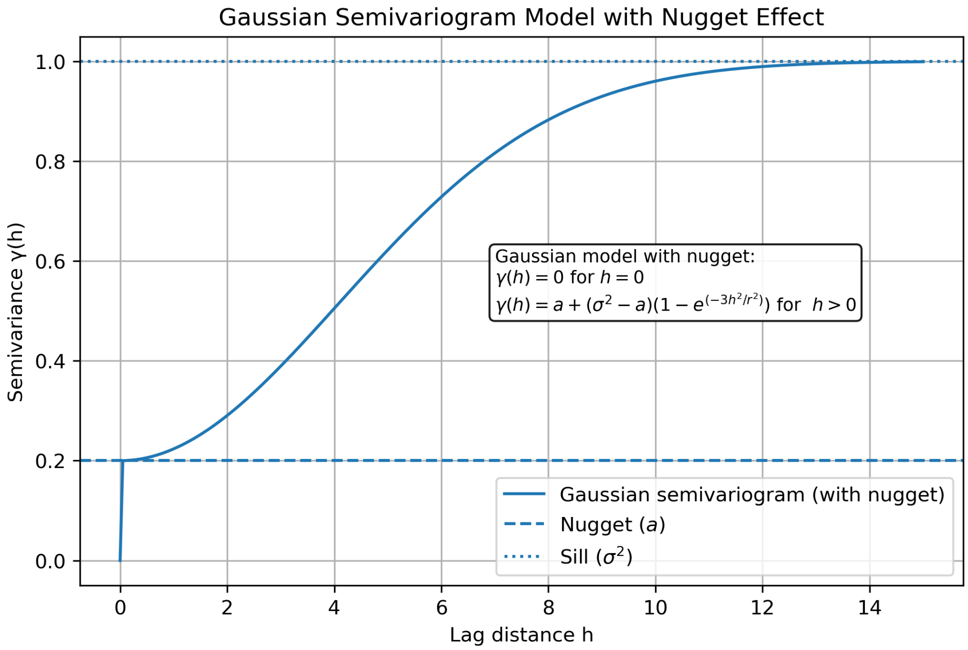
Note: x-axis and y-axis values are for display purposes only

Exponential Variogram



Note: x-axis and y-axis values are for display purposes only

Gaussian Variogram



Note: x-axis and y-axis values are for display purposes only

Spatial Random Field

Let $Z(\mathbf{s})$ denote a spatial process at location $\mathbf{s} \in \mathbb{R}^d$.

- Mean function:

$$\mathbb{E}[Z(\mathbf{s})] = m(\mathbf{s})$$

- Covariance function:

$$\text{Cov}(Z(\mathbf{s}), Z(\mathbf{s}+\mathbf{h})) = C(\mathbf{h})$$

- Often assume intrinsic or second-order stationarity

Stationarity Assumptions

Assumes that (some) statistical properties like mean, variance, and covariance of a spatial process are constant throughout a space domain and do not depend on absolute location but rather on the distance (separation) between them.

Stationarity Assumptions

Strong Stationarity

- Not very relevant here.
- Essentially it says that all joint distributions of a spatial process are invariant over space.

Second-Order Stationarity

- Constant mean: $\mathbb{E}[\mathbf{Z}(\mathbf{s})] = \mu \ \forall \mathbf{s}$
- Covariance depends only on lag $\mathbf{h}$, not on absolute location

$$Cov(\mathbf{Z}(\mathbf{s}), \mathbf{Z}(\mathbf{s}+\mathbf{h})) = \mathbf{C}(\mathbf{h})$$

Stationarity Assumptions

Intrinsic Stationarity

- Mean of increments is zero

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

- Variogram exists and is finite

$$Var(\mathbf{Z}(\mathbf{s}+\mathbf{h}) - \mathbf{Z}(\mathbf{s})) = 2\gamma(h)$$

What Is Anisotropy?

- **Anisotropy** occurs when spatial dependence varies by **direction**.
- Similarity between locations depends on:
  - Distance *and*
  - Direction
- Contrast with isotropy, where dependence depends only on distance.

Intuition

- Points 1 km apart east–west may be more correlated than points 1 km apart north–south.
- Spatial continuity is stronger in some directions.
- Reflects underlying physical or social processes.

Directional Dependence

- In an anisotropic process:

$$\gamma(\mathbf{h}_1) \neq \gamma(\mathbf{h}_2)$$

  even if:

$$|\mathbf{h}_1| = |\mathbf{h}_2|$$

- Variograms differ by direction.

Examples

- Soil properties vary smoothly along river channels.
- Geological strata create directional continuity.
- Air pollution influenced by prevailing winds.
- Urban phenomena shaped by street networks.

Why Anisotropy Matters

- Kriging relies on an accurate spatial dependence model.
- Ignoring anisotropy can:
  - Bias predictions
  - Over-smooth in some directions
  - Underestimate uncertainty

Impact on Variogram Modeling

- Directional variograms reveal anisotropy.
- Isotropic models average over directions.
- Leads to poorly fitted range and sill parameters.

Uncertainty and Interpretation

- Prediction variance depends on the covariance structure.
- Misspecified anisotropy yields incorrect confidence intervals.
- Anisotropy often has clear physical meaning.

How Geostatistics Handles Anisotropy

- Coordinate rotation and rescaling.
- Anisotropic variogram and covariance models.
  - Anisotropic model reaches the sill more rapidly in some directions than others
  - Length of longer axis: major range
  - Length of shorter axis: minor range

Isotropic vs. Anisotropic Variogram

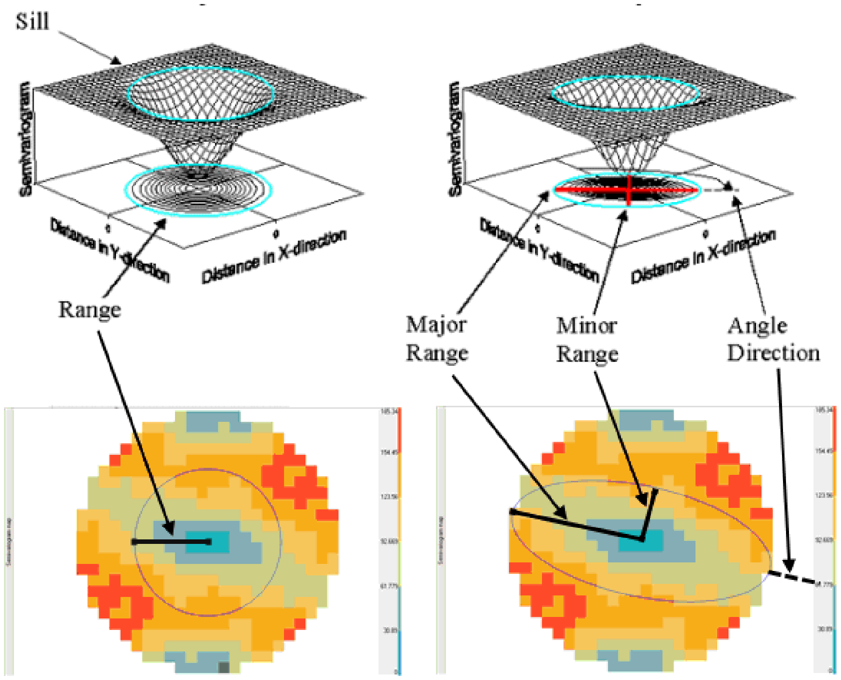


Figure: Directional Variography. Source: ESRI

Isotropy vs. Anisotropy

Isotropy	**Anisotropy**
Depends only on distance	Depends on distance and direction
Circular variograms	Elliptical variograms
Simpler but restrictive	More realistic in practice

Take-Home Message

- Anisotropy means spatial dependence varies by direction.
- It is common in real-world spatial data.
- Properly modeling anisotropy is essential for:
    - Accurate prediction
    - Correct uncertainty quantification
    - Physical interpretability

Additional Slides

Relationship between spatial covariance and variance

Covariance

$$\boldsymbol{C}(\boldsymbol{h}) = \mathbb{E}\left[(\boldsymbol{Z}(\boldsymbol{s}) - \mu)(\boldsymbol{Z}(\boldsymbol{s}+\boldsymbol{h}) - \mu\right]$$

Variance

$$\boldsymbol{C}(\boldsymbol{h}) = \mathbb{E}\left[(\boldsymbol{Z}(\boldsymbol{s}) - \mathbb{E}[\boldsymbol{Z}])^2\right]$$

In the notation we've been using, the semivariogram is related to the covariogram as follows:

$$\gamma(s_i, s_j) = \sigma^2 - C(s_i, s_j)$$

where σ^2 is the sill.
Bottom line: Because of this equivalence (there are exceptions), you can perform prediction using either function.

Mean Increments vs. No Trend

Intrinsic Stationarity

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

- Assumption on *differences*
- Variogram exists
- Mean need not exist explicitly

No Trend (Constant Mean)

$$\mathbb{E}[Z(\mathbf{s})] = m$$

- Mean exists and is constant
- Assumption of ordinary kriging

Constant mean ⇒ zero mean increments but not conversely.
Takeaway Zero mean increments imply stationarity of differences, not necessarily absence of a trend in the mean.

Connection to Universal Kriging

If the mean has a trend:

$$\mathbb{E}[Z(\mathbf{s})] = \mathbf{x}(\mathbf{s})^\top \boldsymbol{\beta}$$

- Mean increments are generally nonzero
- Ordinary kriging is invalid
- Universal kriging models the trend explicitly

Types of Anisotropy

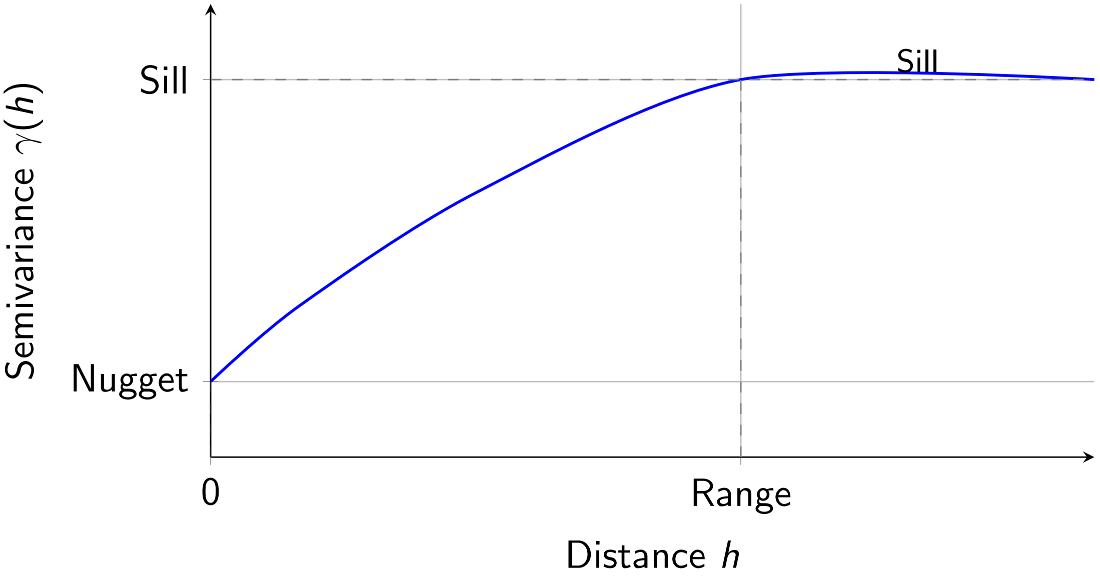
Geometric Anisotropy

- Range changes with direction.
- Sill and nugget remain constant.
- Variogram contours are elliptical.

Zonal Anisotropy

- Sill varies by direction.
- Some directions may have very long or infinite range.

Variogram Components: Nugget, Sill, and Range



The variogram increases with distance until the sill is reached, indicating the limit of spatial dependence.