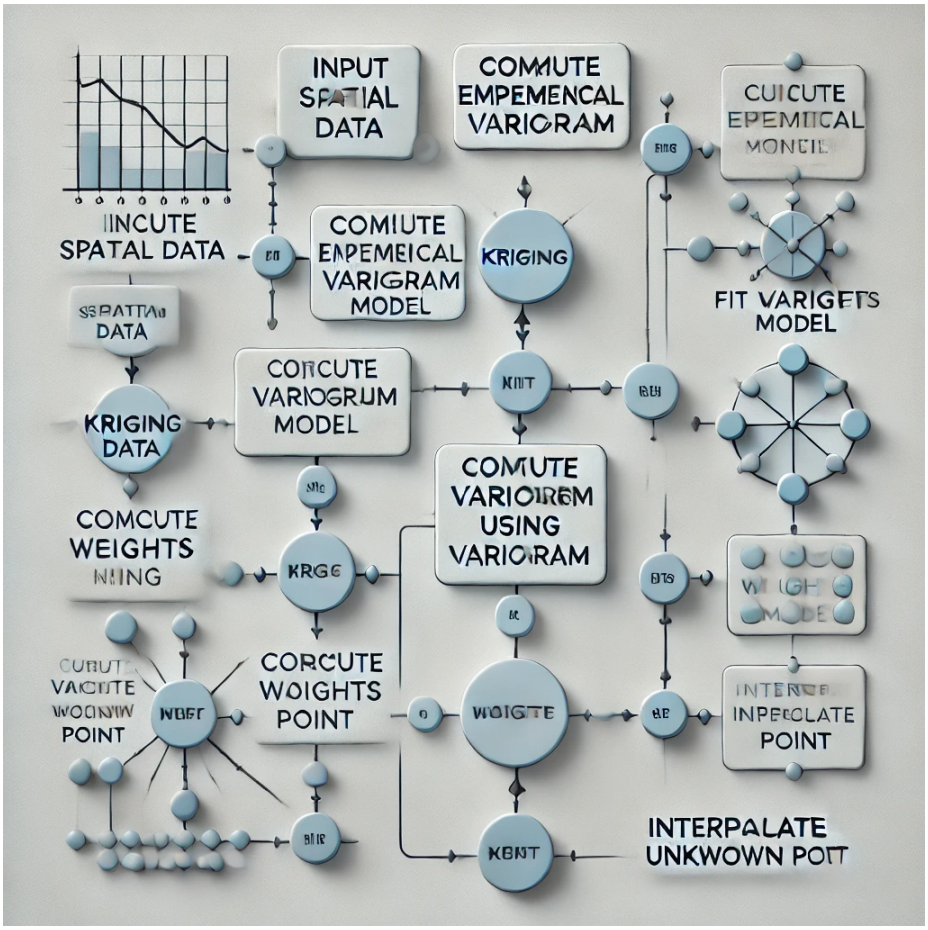
Introduction to Geostatistics

Mark Asch - ML-PREP

2025



Program

1. Introduction and Context
2. Geostatistics
3. Probability and Stochastic Processes
4. Variograms
5. Kriging
6. Bayesian Analysis (advanced)
7. Geostatistics and Machine Learning (advanced)

CONTEXT

1st law of Geography

The First Law of Geography, according to Waldo Tobler, is:

"everything is related to everything else, but near things are more related than distant things."

This first law is the foundation of the fundamental concepts of spatial dependence and spatial autocorrelation and is utilized specifically for the inverse distance weighting method for spatial interpolation and to support the regionalized variable theory for kriging. The first law of geography is the fundamental assumption used in all spatial analysis.

What is Geostatistics?

- Most properties of the environment, such as rainfall, plant nutrients in the soil, geological properties, pollutant distribution, epidemiology, are measured effectively at points between which there are large gaps.

- The environment is continuous, however, and environmental scientists typically want to know the values of those properties between the points, in the gaps; they want to predict in a spatial sense from their data, taking into account the locations of their observations.

- Geostatistics is based on the study of random fields, built upon stochastic processes that are themselves a generalization of random variables.

Geostatistics

Geostatistics is a branch of statistics for spatial or spatiotemporal datasets, in which the data consist of a finite sample of measured values, relating to an underlying spatially continuous phenomenon.

PROBABILITY

and

STOCHASTIC PROCESSES

Probability Theory

- Probability space: $(\Omega, \mathcal{F}, \mathrm{P})$
- Random variable: $X : \Omega \to \mathbb{R}$, $X(\omega) = x$, $\omega \in \Omega$
- Probability distribution: $F_X(x) = \mathrm{P}(X \leq x)$
- Mean and variance: μ_X, σ_X^2
- Covariance and correlation: σ_{XY}^2, ρ_{XY}

Statistics

- Sample mean and variance: m, s^2
- Estimation and inference
- Bayesian statistics

Stochastic Processes

Stochastic Process

A *stochastic process* is a family of random variables $\{Z(t;\omega),\ t \in T\}$ where T is an index set, $\omega \in \Omega$ is a realization in the probability space $(\Omega, \mathcal{F}, \mathrm{P})$.

- The index set can be spatial, temporal or spatio-temporal. We will use as notation:

  $\Rightarrow$ $\mathbf{s} \in D$, $\mathbf{s} = (x,t)$, $D \subset \mathbb{R}^2 \times \mathbb{R}$, $x \in \mathbb{R}^2$.

  $\Rightarrow$ for each realization, $\omega \in \Omega$, we obtain a (deterministic) sample trajectory of the process/field

- Probability Distribution: the law of $Z(s)$ is described by the joint probability function $F_Z(z_1, \ldots, z_n) = \mathrm{P}\left[Z(s_1) < z_1, \ldots, Z(s_n) < z_n\right]$, which is rarely available.

- Stationarity: a (weakly) stationary time series has statistical properties or moments (e.g., mean and variance) that do not vary in time.

Strong all distributions are invariant under translation of indices

Weak/2nd order first and second moments are invariant under translation of indices

- Gaussian processes: often, in nature, complex multivariable phenomena can be described by Gaussian probability distributions. A random function/process is Gaussian if all its finite-dimensional distributions are multivariate Gaussian. In this case, 2nd order stationarity is equivalent to strong stationarity.

Correlation

The problem of characterizing the spatial variability of measurements $z(x)$ reduces to that of characterizing the **correlations** between pairs of random variables $Z(x_i)$ and $Z(x_j)$ (or $Z(x)$ and $Z(x+h)$) that constitute the random field $\{Z(x), x \in D\}$.

Stationarity

> [Matheron et al. 1965]: "For inference to be possible, we need to introduce additional assumptions about the random function $Z(s)$ so as to reduce the number of parameters on which its law depends. This is the purpose of the stationary assumption we are going to define: a stationary function is repeated itself in some form in space, and this repetition once again makes statistical inference possible from a single realisation."

Definition 1. A random function is strictly (strongly) stationary if the joint distribution of $Z(s_i)$ is the same as that of $Z(s_i + h)$ for any translation h.

Definition 2. A random function is weakly (2nd order) stationary if its mean and variance are invariant by translation: $\mu(s+h) = \mu(s)$, $\mathrm{E}\left[(Z(s) - \mu)^2\right] = \sigma^2$ and $\mathrm{Cov}\left[Z(s+h), Z(s)\right] = C(h)$.

Definition 3. A random function is intrinsically stationary if $\mathrm{E}\left[Z(s+h) - Z(s)\right] = 0$ and $\mathrm{Var}\left[Z(s+h) - Z(s)\right] = 2\gamma(h)$, the variogram, a function of h only.

Stationarity

Strong $\Rightarrow$ Weak $\Rightarrow$ Intrinsic

Remark. Intrinsic stationarity allows a maximum of heterogeneity, but still enables the defintion of a variogram that can be used as the basis of inference (kriging). Each geospatial context requires a careful choice of the type of stationarity that should be used.

- Decision table:

homogeneity	HIGH	$\longrightarrow$		LOW
stationarity	strong	2nd order weak	intrinsic	local
restriction	all moments	mean & covariance	mean & variogram	nbhd.

GEOSTATISTICS

Basic Geostatistical Tools

- The *variogram* is the central tool of geostatistics. It enables scientists to assess whether their data are spatially correlated and to what extent.

- With a suitable model the variogram can be combined with the data to predict by *kriging*, which in its simplest form is just weighted averaging (IDW—see below).

- Kriging is an *optimal* method of prediction in that it provides unbiased estimates with minimum variance (MVUE, BLUE, BLUP).

4 Steps of Geostatistics

1. Define an area/site/population D, assumed to be homogeneous enough (see Stationarity) to warrant statistical averaging within it.

2. Collect all the data/measurements within D and calculate the experimental h-characteristics of spatial variability, in particular the empirical (estimated/approximate) variogram $\hat{\gamma}$

3. Smooth and complete the partial experimental image from step 2 into a model variogram $\gamma(h)$ for all inter-distance vectors h.

4. Use the model of $\gamma(h)$ from step 3, or the corresponding covariance model $C(h)$, to perform one of the well-proven regression techniques of kriging.

Variograms

- bivariate distribution
- variogram
- empirical/approximate/sample variogram
- relation with covariance

Variograms - Bivariate Distribution

- Recall, for a single random variable, X,

  $\Rightarrow$ distribution: $F_X(x)$
  $\Rightarrow$ expectation: $\mu = \mathrm{E}[X]$
  $\Rightarrow$ variance:

$$\sigma^2 = \mathrm{Var}[X] = \mathrm{E}[(X-\mu)^2] = \mathrm{E}[X^2] - \mu^2$$

- In earth and environmental sciences, what is often most important is the pattern of dependencies between 2 random variables, X and Y, that can represent

  $\Rightarrow$ 2 differrent attributes measured at the same location
  $\Rightarrow$ the same attribute measured at 2 different locations, x and $x+h$; then $X = Z(x)$ and $Y = Z(x+h)$
  $\Rightarrow$ 2 attributes measured at 2 locations

- For a pair of random variables, X and Y, we have

  $\Rightarrow$ joint distribution: $F_{XY}(x,y) = \mathrm{P}\{X \leq x \text{ and } Y \leq y\}$

⇒ expectations: $\mu_X = \mathrm{E}[X]$, $\mu_Y = \mathrm{E}[Y]$

⇒ covariance:

$$\begin{aligned}\mathrm{Cov}\,[X,Y] = \sigma_{XY} &= \mathrm{E}\,[(X-\mu_X)(Y-\mu_Y)] \\ &= \mathrm{E}\,[XY] - \mu_X\mu_Y\end{aligned}$$

and autocovariance

$$\sigma^2(X) = \mathrm{Cov}\,[X,X] = \mathrm{Var}\,[X] \geq 0$$

⇒ empirical (sample, estimated) covariance

$$\begin{aligned}s^2_{XY} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - m_X)(y_i - m_Y) \\ &= \frac{1}{n}\sum_{i=1}^{n} x_i y_i - m_X m_Y,\end{aligned}$$

where the empirical means are defined by

$$m_X = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad m_Y = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$\Rightarrow$ correlation:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$
$$= \frac{\mathrm{Cov}\left[X, Y\right]}{\sqrt{\mathrm{Var}\left[X\right] \mathrm{Var}\left[Y\right]}} \in [-1, +1]$$

$\Rightarrow$ variogram:

$$2\gamma_{XY} = \mathrm{E}\left[(X - Y)^2\right]$$

Covariance vs. Variogram

- Variogram γ_{XY} is a measure of **variability**.
- Covariance σ_{XY} and correlation ρ_{XY} are measures of **similarity**.

- For standardized variables, $X' = (X - \mu_X)/\sigma_X$ and $Y' = (Y - \mu_Y)/\sigma_Y$, with mean zero and variance one,

$$\gamma_{X'Y'} = 1 - \rho_{XY},$$

so when the variogram γ_{XY} increases, the covariance σ_{XY} decreases—more variability implies less similarity. See 1st Law of Geography...

Empirical Variogram

- Let $N(h) = \{(x_i, x_j) : x_i - x_j = h,\ i, j = 1, \ldots, n\}$ be the set of all pairs of points separated by a distance (vector) h and with cardinality (total number of pairs) $|N(h)|$

- Method-of-moments estimator of the variogram is then

$$\hat{\gamma}(h) = \frac{1}{2\,|N(h)|} \sum_{i,j \in N(h)} [z(x_i) - z(x_j)]^2$$

- Note: the choice of $N(h)$ will depend on the isotropy or anisotropy of the region D and, in practice, the experimental variogram is computed using lag vectors h of length up to a distance of half the diameter of the region

Variogram Model (I)

- The experimental variogram $\hat{\gamma}(h)$ provides a first estimate of the assumed underlying theoretical variogram $\gamma(h)$, but due to its numerous limitations cannot be used directly.

- We have to fit a variogram function to the empirical variogram, i.e. replace it by a parametric, theoretical/model variogram. We can choose among:

  $\Rightarrow$ Nugget-effect model.
  $\Rightarrow$ Bounded linear model.
  $\Rightarrow$ Spherical model.
  $\Rightarrow$ Exponential model.
  $\Rightarrow$ Gaussian model.
  $\Rightarrow$ Matérn models.

Variogram Models (II)

model	Variogram, $\gamma(t)$
Linear	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 t & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
Spherical	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 & \text{if } t \geq 1/\phi \\ \tau^2 + \sigma^2 \left[\frac{3}{2}\phi t - \frac{1}{2}(\phi t)^3\right] & \text{if } 0 < t \leq 1/\phi \\ 0 & \text{otherwise} \end{cases}$
Exponential	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi t)) & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
Powered exponential	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\lvert\phi t\rvert^p)) & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
Gaussian	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi^2 t^2)) & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
Rational quadratic	$\gamma(t) = \begin{cases} \tau^2 + \frac{\sigma^2 t^2}{(1+\phi t^2)} & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
Wave	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2(1 - \frac{\sin(\phi t)}{\phi t}) & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
Power law	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 t^\lambda & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
Matérn	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left[1 - \frac{(2\sqrt{\nu}t\phi)^\nu}{2^{\nu-1}\Gamma(\nu)} K_\nu(2\sqrt{\nu}t\phi)\right] & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$
Matérn at $\nu = 3/2$	$\gamma(t) = \begin{cases} \tau^2 + \sigma^2 \left[1 - (1 + \phi t)\exp(-\phi t)\right] & \text{if } t > 0 \\ 0 & \text{otherwise} \end{cases}$

Variogram Model Choice (III)

- **Linear**: A model in which influence is determined proportional to distance. At first glance, it seems plausible, but it's actually not used, due to difficult interpretation with the covariogram.

- **Spherical**: A model in which influence completely disappears beyond a certain distance. This is a reasonable choice for many data sets.

- **Exponential**: The simplest and most understandable model where influence decreases exponentially as distance increases. It is sufficient for basic, academic projects.

- **Matérn**: Among the given models, it can be used in a wide variety of situations due to its inclusion of τ^2 which relates to the y-intercept, σ^2 that determines scale, and ϕ, ν that are smoothing parameters that affect the shape itself. Matérn is considered the most powerful

and most widely-used model. The K_ν appearing in the equations refers to the Modified Bessel function of the first kind.

Kriging

Definition. Kriging is a group of geostatistical techniques to interpolate the value of a random field (e.g., the elevation, z, of the landscape as a function of the geographic location) at an unobserved location from observations of its value at nearby locations.

- Kriging is an estimation at an unsampled location calculated as the weighted sum of the adjacent sampled points.
- The weighting factors depend on a model of spatial correlation, with the calculation of the weighting factors done by minimizing the error variance of a given or assumed model of the data with regard to the spatial distribution of the observed data points.
- This model of spatial distribution is where the semivariogram is essential. Knowing what model to use and why depends on many factors that we will need to consider.
- Types of kriging:

$\Rightarrow$ simple - 2nd-order stationary; μ known, constant

$\Rightarrow$ ordinary - intrinsically stationary; μ unknown, constant

$\Rightarrow$ universal - intrinsically stationary; μ unknown

Kriging	Stationarity	Assumptions
simple	2nd-order, weak	μ known, constant
ordinary	intrinsic	μ unknown, constant
universal	intrinsic	μ unknown

Inverse Distance Weighting (IDW)

- IDW is a deterministic method for multivariate, spatial interpolation
- The interpolated value, $u(x)$, at an unknown point x where we have no measurement available, is obtained by weighting with the inverse distance to the known points, $d(x_i, x)$ as follows

$$u(x) = \begin{cases} \frac{\sum_{i=1}^{n} w_i(x) u_i}{\sum_{i=1}^{n} w_i(x)}, & \text{if } d(x, x_i) \neq 0, \\ u_i, & \text{if } d(x, x_i) = 0 \text{ for some } i, \end{cases}$$

where the inverse weight

$$w_i(x) = \frac{1}{d(x, x_i)^p}$$

Simple Kriging (I)

- Recall: kriging produces an interpolation function based on a covariance or variogram model, derived from the data, rather than an *a priori* model of the interpolating function, as would be the case in deterministic, polynomial interpolation, for example.

- Knowledge of the mean gives a lot of information, and may even be sufficient. However, in the real world we require many repetitions or large data volumes to obtain satisfactory estimates of the mean.

- With a known mean, we can develop a very simple theory that gives the kriging estimator some nice properties.

Simple Kriging - Formulation (II)

- **Assumptions**:
  1. Mean μ is known and constant for all $x \in D$.
  2. $Z(x)$ is 2nd-order stationary, with known covariance function

$$C(h) = \mathrm{Cov}\left[Z(x+h), Z(x)\right] = \mathrm{E}\left[Z(x+h)Z(x)\right] - \mu^2$$

- **Predictor**: the predictor at any point $x_0 \in D$ is the sum of the known mean μ and the weighted average of the difference of $Z(x_i)$ and μ

$$\hat{Z}(x_0) = \mu + \sum_{i=1}^{n} \lambda_i \left(Z(x_i) - \mu\right)$$

- **Optimization**: the SK predictor is the optimal, minimum variance, unbiased estimator, obtained by solving the necessary optimality condition for the λ_i

$$C\boldsymbol{\lambda}^{\mathrm{SK}} = \mathbf{c},$$

where

$$C = \begin{bmatrix} \mathrm{Cov}(x_1, x_1) & \cdots & \mathrm{Cov}(x_1, x_n) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(x_n, x_1) & \cdots & \mathrm{Cov}(x_n, x_n) \end{bmatrix},$$

$$\boldsymbol{\lambda}^{\mathrm{SK}} = \left[\lambda_1^{\mathrm{SK}}, \ldots, \lambda_n^{\mathrm{SK}}\right]^{\mathrm{T}},$$

$$\mathbf{c} = \left[\mathrm{Cov}(x_0, x_1), \ldots, \mathrm{Cov}(x_0, x_n)\right]^{\mathrm{T}}.$$

- **Error analysis**: calculate the SK estimation variance,

$$\sigma^2_{\mathrm{SK}}(x_0) = \mathrm{Cov}(0) - \mathbf{c}^{\mathrm{T}}\boldsymbol{\lambda}^{\mathrm{SK}},$$

which can then be used for constructing confidence intervals.

Simple Kriging - Algorithm (III)

1. Calculate each term of the matrix C.

2. Calculate each term of the right-hand side vector $\mathbf{c}$.

3. Solve the linear system of equations

$$C\boldsymbol{\lambda}^{\mathrm{SK}} = \mathbf{c}.$$

4. Compute the estimate

$$\hat{Z}^{\mathrm{SK}}(x_0) = \mu + \sum_{i=1}^{n} \lambda_i \left(Z(x_i) - \mu\right).$$

5. Calculate the simple kriging estimation variance

$$\sigma^2_{\mathrm{SK}}(x_0) = \mathrm{Cov}(0) - \mathbf{c}^{\mathrm{T}}\boldsymbol{\lambda}^{\mathrm{SK}}.$$

Ordinary Kriging

- In most practical situations the mean $\mu(x)$ is not known.
- An obvious approach would be to estimate it and subtract it from the data and thus recover the zero-mean case. This approach is commonly used for processing time series, where it is known as "detrending."
- The problem is that estimated residuals are not the same as true residuals, they depend on how the mean is estimated, and the statistical properties of the whole procedure are difficult to analyze.

  $\Rightarrow$ The ordinary and universal kriging approaches provide an optimal solution that involves only one estimation step.

Ordinary Kriging - Formulation (II)

- **Assumptions**:

  1. Mean μ is constant but unknown for all $x \in D$.
  2. $Z(x)$ is intrinsically stationary, with known variogram

$$2\gamma(h) = \mathrm{Var}\left[Z(x) - Z(x+h)\right].$$

- **Predictor**: the predictor at any point $x_0 \in D$ is the linear combination of the random functions $Z(x_i)$

$$\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i),$$

  subject to the constraint

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

- **Optimization**: the OK predictor is the optimal, minimum variance, unbiased estimator, obtained by solving the necessary optimality condition for the λ_i augmented by the Lagrange multiplier ν that enforces the constraint

$$G\boldsymbol{\lambda}^{\mathrm{OK}} = \mathbf{g},$$

where

$$G = \begin{bmatrix} \gamma x_1, x_1) & \cdots & \gamma(x_1, x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n, x_1) & \cdots & \gamma x_n, x_n) & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix},$$

$$\boldsymbol{\lambda}^{\mathrm{OK}} = \left[\lambda_1^{\mathrm{OK}}, \ldots, \lambda_n^{\mathrm{OK}}, -\nu\right]^{\mathrm{T}},$$

$$\mathbf{g} = \left[\gamma(x_0, x_1) \cdots \gamma(x_0, x_n)\ 1\right]^{\mathrm{T}}.$$

- **Error analysis**: calculate the OK estimation variance,

$$\sigma^2_{\mathrm{OK}}(x_0) = \mathbf{g}^{\mathrm{T}}\boldsymbol{\lambda}^{\mathrm{OK}},$$

which can then be used for constructing confidence intervals.

Ordinary Kriging - Algorithm (III)

1. Calculate each term of the matrix G.

2. Calculate each term of the right-hand side vector $\mathbf{g}$.

3. Solve the linear system of equations

$$G\boldsymbol{\lambda}^{\mathrm{OK}} = \mathbf{g}.$$

4. Compute the estimate

$$\hat{Z}^{\mathrm{OK}}(x_0) = \mathbf{z}^{\mathrm{T}}\boldsymbol{\lambda}^{\mathrm{OK}},$$

   where

$$\mathbf{z} = \left[Z(x_1) \cdots Z(x_n)\ 0\right]^{\mathrm{T}}.$$

5. Calculate the ordinary kriging estimation variance

$$\sigma^2_{\mathrm{OK}}(x_0) = \mathbf{g}^{\mathrm{T}}\boldsymbol{\lambda}^{\mathrm{OK}}.$$

Ordinary Kriging - 2nd order (IV)

- Under the stronger assumption of 2nd order stationarity, but still with an unknown mean, we can derive an OK algorithm.

- We then simply replace the semivariogram values $\gamma(\cdot, \cdot)$ by the covariances $\mathrm{Cov}(\cdot, \cdot)$.

- Despite being more restrictive, this algorithm is widely used for two reasons:

  $\Rightarrow$ processes for which there is a semivariogram and no covariance are of rare and limited interest,

  $\Rightarrow$ the system of equations based on the covariance is better behaved than that for intrinsic random functions.

Ordinary Kriging - Properties (V)

- Simple kriging shares most of its properties with ordinary kriging.
- Ordinary kriging has the following properties:

1. minimum mean square error,
2. estimation interval not restricted to the data interval,
3. screen effect,
4. declustering ability,
5. exact interpolation with zero kriging variance,
6. inability to handle duplicate sampling sites,
7. independence from translation of the reference axes,

8. dependence upon the sampling pattern, and

9. independence of the kriging variance from individual observations.

Universal Kriging

- Models that presume constancy of the mean are inadequate for the characterization of attributes that have a clear, systematic variation.

- Universal kriging is a generalization that takes simple kriging one step beyond the ordinary kriging expansion.

- In addition to removing the requirement of simple kriging that the mean be known, universal kriging fully removes from ordinary kriging the need to assume a constant mean.

Universal Kriging - Formulation (II)

- **Assumptions**:

  1. Mean μ is unknown for all $x \in D$.
  2. The residuals

$$Y(x) = Z(x) - \mathrm{E}\left[Z(x)\right]$$

  are intrinsically stationary, with constant expectation and known variogram

$$2\gamma_Y(h) = \mathrm{Var}\left[Y(x) - Y(x+h)\right].$$

  3. The unknown expectation (drift), can be modeled as a linear combination of polynomial basis functions

$$\mu(x) = \sum_{i=0}^{k} a_k f_k(x)$$

  with $f_0(x) = 1$.

- **Predictor**: the UK estimator at any point $x_0 \in D$ is the linear combination of the random functions $Z(x_i)$

$$\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i),$$

subject to the constraint

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

- **Optimization**: the UK predictor is the optimal, minimum variance, unbiased estimator, obtained by solving the necessary optimality condition for the λ_i augmented by Lagrange multipliers ν_i that enforces the constraint of the drift model,

$$A\boldsymbol{\lambda}^{\mathrm{UK}} = \mathbf{a},$$

where

$$A = \begin{bmatrix} \gamma x_1, x_1) & \cdots & \gamma(x_1, x_n) & 1 & f_1(x_1) & \cdots & f_k(x_1) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(x_n, x_1) & \cdots & \gamma x_n, x_n) & 1 & f_1(x_1) & \cdots & f_k(x_1) \\ 1 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ f_1(x_1) & \cdots & f_1(x_n) & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\ f_k(x_1) & \cdots & f_k(x_n) & 0 & 0 & \cdots & 0 \end{bmatrix},$$

$$\boldsymbol{\lambda}^{\mathrm{UK}} = \left[\lambda_1^{\mathrm{UK}} \ \ldots \ \lambda_n^{\mathrm{UK}} \ \nu_0 \ \ldots \ \nu_k\right]^{\mathrm{T}},$$

$$\mathbf{a} = \left[\gamma_Y(x_0, x_1) \ \cdots \ \gamma_Y(x_0, x_n) \ 1 \ f_1(x_0) \ \ldots \ f_k(x_0)\right]^{\mathrm{T}}.$$

- **Error analysis**: calculate the UK estimation variance,

$$\sigma^2_{\mathrm{UK}}(x_0) = \mathbf{a}^{\mathrm{T}} \boldsymbol{\lambda}^{\mathrm{UK}},$$

which can then be used for constructing confidence intervals.

Universal Kriging - Algorithm (III)

1. Calculate each term of the matrix A.

2. Calculate each term of the right-hand side vector $\mathbf{a}$.

3. Solve the linear system of equations

$$A\boldsymbol{\lambda}^{\mathrm{UK}} = \mathbf{a}.$$

4. Compute the estimate

$$\hat{Z}^{\mathrm{UK}}(x_0) = \mathbf{z}^{\mathrm{T}}\boldsymbol{\lambda}^{\mathrm{UK}},$$

   where

$$\mathbf{z} = \left[Z(x_1) \cdots Z(x_n)\ 0\right]^{\mathrm{T}}.$$

5. Calculate the universal kriging estimation variance

$$\sigma^2_{\mathrm{UK}}(x_0) = \mathbf{a}^{\mathrm{T}}\boldsymbol{\lambda}^{\mathrm{UK}}.$$

Universal Kriging

As long as the assumptions are met, the universal kriging algorithm can be employed to characterize any spatial attribute regardless of its physical nature.

Universal Kriging - 2nd order (IV)

- Under the stronger assumption of 2nd order stationarity, but still with an unknown mean, we can derive a UK algorithm.

- We then simply replace the semivariogram values $\gamma(\cdot, \cdot)$ by the covariances $\mathrm{Cov}(\cdot, \cdot)$.

- Despite being more restrictive, this algorithm is widely used for two reasons:

  $\Rightarrow$ Attributes with infinite variance are a curiosity rather than a common occurrence.

  $\Rightarrow$ The system of equations based on the covariance is better behaved than that for intrinsic random functions.

Universal Kriging - Properties (V)

Universal kriging shares with ordinary kriging all its properties:

- it is a minimum mean square error, unbiased, exact interpolator that
- automatically corrects for clustering in the sampling, and
- the observations take weights under a screen effect.
- The method is fairly robust to parameter misspecifications and inadequate assumptions.
- Notice that although the drift model must be specified, it is handled automatically in the systems of equations without the necessity of resorting to the estimation of the coefficients a_i for the model.

- Perhaps the most serious drawback of universal kriging is the need to specify the semivariogram.

  $\Rightarrow$ The method requires the semivariogram of the residuals, not the semivariogram of the original regionalized variable.

  $\Rightarrow$ The easiest and most common way to estimate the semivariogram of the residuals is to do the modeling along trend-free directions in the original data.

References

1. C. Wikle, A. Zammit-Mangion, N. Cressie. *Spatio-Temporal Statistics with R*. CRC Press. 2019. https://spacetimewithr.org/

2. P. Diggle, P. Ribeiro. *Model-based Geostatistics*. Springer. 2007.

3. M. Oliver, R. Webster. *Basic Steps in Geostatistics: The Variogram and Kriging*. Springer. 2007.

4. INSEE. *Handbook of Spatial Analysis: Theory and practical application with R*. 2018. https://www.insee.fr/en/information/3635545