Concepts

Geostatistical model

- The experimental variogram is used to analyze the spatial structure of the data from a regionalized variable $z(\mathbf{x})$.
- It is fitted with a nested variogram model, thus providing the structure function of a random function.
- The regionalized variable (reality) is viewed as one realization of the random function $Z(\mathbf{x})$, which is a collection of random variables.

Kriging: a linear regression method for estimating point values (or spatial averages) at any location of a region.

Conditional simulation: simulation of an ensemble of realizations of a random function, conditional upon data — for non-linear estimation.
**Concepts**

Geostatistical model

- The experimental variogram is used to analyze the spatial structure of the data from a regionalized variable \( z(x) \).
- It is fitted with a nested variogram model, thus providing the structure function of a random function.
- The regionalized variable (reality) is viewed as one realization of the random function \( Z(x) \), which is a collection of random variables.

**Kriging**: a *linear* regression method for estimating point values (or spatial averages) at any location of a region.

**Conditional simulation**: simulation of an ensemble of realizations of a random function, conditional upon data — for *non-linear* estimation.
Geostatistical model

- The experimental variogram is used to analyze the spatial structure of the data from a regionalized variable $z(x)$.
- It is fitted with a nested variogram model, thus providing the structure function of a random function.
- The regionalized variable (reality) is viewed as one realization of the random function $Z(x)$, which is a collection of random variables.

**Kriging:** a *linear* regression method for estimating point values (or spatial averages) at any location of a region.

**Conditional simulation:** simulation of an ensemble of realizations of a random function, conditional upon data — for *non-linear* estimation.
Stationarity

For the top series:
- stationary mean $m$ and covariance function $C(h)$

For the bottom series:
- mean and variance are not stationary,
- actually the realization of a non-stationary process without drift.

Both types of series can be characterized with a variogram.
Kriging of the mean of a random function
Spatially Correlated Data

Sample locations $x_\alpha$ in a spatial domain:

With spatial correlation we need to consider that:
- each sample point plays a different role in estimating the mean of the spatial domain,
- distances to neighboring points play a role.

How should samples thus be weighted in an optimal way?
Using the arithmetic mean:

\[ M^* = \frac{1}{n} \sum_{\alpha=1}^{n} Z(x_\alpha) \]

all samples get the same weight: \( \frac{1}{n} \)

We rather need an estimator:

\[ M^* = \sum_{\alpha=1}^{n} w_\alpha Z(x_\alpha) \]

with weights \( w_\alpha \) reflecting the spatial correlation.
We assume *translation-invariance* of the mean:

\[ \forall x \in D : \quad \mathbb{E}[Z(x)] = m \]

and of the covariance:

\[ \forall \mathbf{x}_\alpha, \mathbf{x}_\beta \in D \text{ with } \mathbf{x}_\alpha - \mathbf{x}_\beta = \mathbf{h} : \quad \text{cov}(Z(\mathbf{x}_\alpha), Z(\mathbf{x}_\beta)) = C(\mathbf{x}_\alpha - \mathbf{x}_\beta) = C(\mathbf{h}) \]
Unbiased estimator

The estimation error in our statistical model:

\[
\underbrace{M^*}_{\text{estimated value}} - \underbrace{m}_{\text{true value}}
\]

should be zero on average:

\[
E\left[ M^* - m \right] = 0
\]

No bias: the estimator \( M^* \) does not on average yield a value that is different from \( m \).
No bias

Bias is avoided using weights of unit sum:

\[
\sum_{\alpha=1}^{n} w_\alpha = 1
\]

Consider:

\[
\mathbb{E}\left[M^* - m\right] = \mathbb{E}\left[\sum_{\alpha=1}^{n} w_\alpha Z(x_\alpha) - m\right]
\]

\[
= \sum_{\alpha=1}^{n} w_\alpha \mathbb{E}\left[Z(x_\alpha)\right] - m - m
\]

\[
= m \sum_{\alpha=1}^{n} w_\alpha - m = 0
\]
The variance of the estimation error is:

\[
\sigma^2_E = \text{var}(M^* - m) = \mathbb{E}\left[(M^* - m)^2\right] - \mathbb{E}\left[M^* - m\right]^2
\]

\[
= \mathbb{E}\left[M^{*2} - 2mM^* + m^2\right]
\]

\[
= \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} w_{\alpha} w_{\beta} \mathbb{E}\left[Z(x_{\alpha}) Z(x_{\beta})\right] - 2m \sum_{\alpha=1}^{n} w_{\alpha} \mathbb{E}\left[Z(x_{\alpha})\right] + m^2
\]

\[
\Rightarrow \sigma^2_E = \sum_{\alpha=1}^{n} \sum_{\beta=1}^{n} w_{\alpha} w_{\beta} C(x_{\alpha} - x_{\beta})
\]
Minimal estimation variance

Aim: weights $w_\alpha$ that produce a minimal estimation variance:

\[
\text{minimize } \sigma_E^2 \quad \text{subject to } \sum_{\alpha=1}^{n} w_\alpha = 1
\]

The objective function $\varphi$ has $n+1$ parameters:

\[
\varphi(w_1, \ldots, w_n, \mu) = \text{var}(M^* - m) - 2\mu \left( \sum_{\alpha=1}^{n} w_\alpha - 1 \right)
\]

where $\mu$ is a Lagrange multiplier.

Setting partial derivatives to zero, we obtain $n+1$ equations:

\[
\forall \alpha : \frac{\partial \varphi(w_1, \ldots, w_n, \mu)}{\partial w_\alpha} = 0, \quad \frac{\partial \varphi(w_1, \ldots, w_n, \mu)}{\partial \mu} = 0
\]
Kriging equations

The method of Lagrange yields the equation system for the optimal weights $w_{KM}^{\alpha}$ of the estimation of the mean:

$$\begin{align*}
\sum_{\beta=1}^{n} w_{\beta}^{KM} C(x_{\alpha} - x_{\beta}) - \mu_{KM} &= 0 \quad \text{for } \alpha = 1, \ldots, n \\
\sum_{\beta=1}^{n} w_{\beta}^{KM} &= 1
\end{align*}$$

The variance at the minimum:

$$\sigma_{KM}^2 = \mu_{KM}$$

is equal to the Lagrange multiplier.
Special case: no spatial correlation

When the covariance model is only *nugget-effect*:

\[ C(x_\alpha - x_\beta) = \begin{cases} \sigma^2 & \text{if } x_\alpha = x_\beta \\ 0 & \text{if } x_\alpha \neq x_\beta \end{cases} \]

the kriging of the mean system simplifies to:

\[
\begin{align*}
\sum_{\beta=1}^{n} w_{\alpha}^{KM} \sigma^2 &= \mu_{KM} \quad \text{for } \alpha = 1, \ldots, n \\
\sum_{\beta=1}^{n} w_{\beta}^{KM} &= 1
\end{align*}
\]

The solution weights are all equal:

\[ w_{\alpha}^{KM} = \frac{1}{n} \]

\[ M^* = \frac{1}{n} \sum_{\alpha=1}^{n} Z(x_\alpha) \quad \text{with variance} \quad \mu_{KM} = \sigma_{KM}^2 = \frac{1}{n} \sigma^2 \]
Special case: no spatial correlation

When the covariance model is only *nugget-effect*:

\[ C(\mathbf{x}_\alpha - \mathbf{x}_\beta) = \begin{cases} \sigma^2 & \text{if } \mathbf{x}_\alpha = \mathbf{x}_\beta \\ 0 & \text{if } \mathbf{x}_\alpha \neq \mathbf{x}_\beta \end{cases} \]

the kriging of the mean system simplifies to:

\[
\begin{cases}
  w_{\alpha}^{\text{KM}} \sigma^2 = \mu_{\text{KM}} \\
  \sum_{\beta=1}^{n} w_{\beta}^{\text{KM}} = 1
\end{cases}
\]

for \( \alpha = 1, \ldots, n \)

The solution weights are all equal:

\[ w_{\alpha}^{\text{KM}} = \frac{1}{n} \]

\[
M^* = \frac{1}{n} \sum_{\alpha=1}^{n} Z(\mathbf{x}_\alpha) \quad \text{with variance} \quad \mu_{\text{KM}} = \frac{\sigma_{\text{KM}}^2}{\frac{1}{n} \sigma^2}
\]
Estimation at an unsampled location

Sample locations $\mathbf{x}_\alpha$ (blue dots) in a spatial domain $\mathcal{D}$:

Aim: estimate $Z^\star$ at an unsampled location $\mathbf{x}_0$. 
Ordinary kriging

The estimate $Z^*$ is a weighted average of data values $Z(x_\alpha)$:

$$Z^*(x_0) = \sum_{\alpha=1}^{n} w_\alpha Z(x_\alpha) \quad \text{with} \quad \sum_{\alpha=1}^{n} w_\alpha = 1$$

The weights $w^{OK}_\alpha$ of the Best Linear Unbiased Estimator (BLUE) are solution of the system:

$$\begin{cases} 
\sum_{\beta=1}^{n} w^{OK}_\beta \gamma(x_\alpha-x_\beta) + \mu_{OK} = \gamma(x_\alpha-x_0) & \text{for} \quad \alpha = 1, \ldots, n \\
\sum_{\beta=1}^{n} w^{OK}_\beta = 1
\end{cases}$$

Minimal variance: $\sigma_{OK}^2 = \mu_{OK} + \sum_{\alpha=1}^{n} w^{OK}_\alpha \gamma(x_\alpha-x_0)$
Ordinary kriging

The estimate \( Z^* \) is a weighted average of data values \( Z(x_\alpha) \):

\[
Z^*(x_0) = \sum_{\alpha=1}^{n} w_\alpha Z(x_\alpha) \quad \text{with} \quad \sum_{\alpha=1}^{n} w_\alpha = 1
\]

The weights \( w^{\text{OK}}_\alpha \) of the Best Linear Unbiased Estimator (BLUE) are solution of the system:

\[
\begin{align*}
\sum_{\beta=1}^{n} w^{\text{OK}}_\beta \gamma(x_\alpha - x_\beta) + \mu_{\text{OK}} &= \gamma(x_\alpha - x_0) \quad \text{for} \quad \alpha = 1, \ldots, n \\
\sum_{\beta=1}^{n} w^{\text{OK}}_\beta &= 1
\end{align*}
\]

Minimal variance:

\[
\sigma^2_{\text{OK}} = \mu_{\text{OK}} + \sum_{\alpha=1}^{n} w^{\text{OK}}_\alpha \gamma(x_\alpha - x_0)
\]
The behavior of kriging weights
Geometric anisotropy
Spherical variogram, weights sum up to 100%

Isotropic model:

- 25% at each of the four corners
- L represents the horizontal distance

Anisotropic (ranges — horizontal 1.5L, vertical 0.75L):
- 25% at each of the four corners
- L represents the horizontal distance

Geometric anisotropy
Spherical variogram, weights sum up to 100%

Isotropic model:

Anisotropic (ranges — horizontal 1.5L, vertical .75L):
The screen effect
Spherical variogram with range 2L

Left sample is at 1L and right at 2L from target

\[ \sigma_{OK}^2 = 1.14 \]

65.6%  
A

34.4%  
B

Introducing an extra sample

\[ \sigma_{OK}^2 = 0.87 \]

49.1%  
C

Adding the sample C screens off the sample B.
The screen effect
Spherical variogram with range 2L

Left sample is at 1L and right at 2L from target

\[ \sigma_{OK}^2 = 1.14 \]

Introducing an extra sample

\[ \sigma_{OK}^2 = 0.87 \]

Adding the sample C screens off the sample B.
Filtering noisy images by cokriging
Trace elements are usually masked by instrumental noise.

Data: $512 \times 512$ image of phosphorus (P) trace elements.

Images for chrome (Cr) and (Ni) are less noisy.

Geostatistical filtering is used to remove the noise.
Structural analysis

Image of phosphorus

Nested variogram

\[ \gamma(h) = 384 \text{nug}(h) + 75 \exp(h) + 13 |h| \]
Filtering the nugget-effect

Raw image of phosphorus

Filtered image
Multivariate data

P

Cr

Ni
Multivariate structural analysis
Direct and cross variograms

Matrix variogram model: \[ G(h) = B_0 \text{nug}(h) + B_1 \exp(h) + B_1 |h| \]
Filtering the nugget-effect
Phosphorus

Filtered by kriging

Filtered by cokriging
Space-time filtering
Séguret & Huchon, JGR 1990
Earth magnetism

- Magnetic anomalies are essential to study earth history.
- Magnetism is influenced by several external factors like:
  - solar wind explaining daily fluctuations (period: 24 hours)
  - rotation of the moon around the earth (period: 28 days)
  - solar perturbations (half-year cycle)

Available data:

**SEAPERC campaign (Ifremer, 1986)** Data from a research vessel about magnetism over a fractured area of 111 km² off Peru.

**Fluctuations of earth magnetism** Measurements at a Peruvian observatory for the time period of the campaign.
Daily fluctuation of earth magnetism
Huancayo observatory (Peru): 22 to 28/08/1986

Time series (6 days)

Variogram
SEAPERC campaign
Ship moves along a profile in 12 hours

Map

Study area
Measurements at observatory and along ship route

Perturbations diurnes $D(t)$ (observatoire fixe de HUANCAYO)

$\text{Magnetisme } Z(t) \text{ perturbe par } D(t)$

 absence de donnees du fait d'une panne du magnetometre
Filtering the daily fluctuations of magnetism

Space-time model: $Z(x, t) = Y(x) + m(t)$

With time perturbations

After geostatistical filtering
We have seen:

- how to set up a variogram model and a corresponding random function model with several components,
- that these components can be extracted by kriging,
- that this applies to multivariate or space-time filtering problems.

These methods are based on estimators that are **linear** combinations (weighted averages) of data.

However we often are asked to estimate statistics that are **not linearly** related to data. We will see next how to provide answers by geostatistical simulation.
JP Chilès and P Delfiner.  
*Geostatistics: Modeling Spatial Uncertainty.*  

G Matheron.  
*The Theory of Regionalized Variables and its Applications.*  

S Séguret and P Huchon.  
Trigonometric kriging: a new method for removing the diurnal variation from geomagnetic data.  

H Wackernagel.  
*Multivariate Geostatistics: an Introduction with Applications.*  