Theory of Kriging

D G Rossiter

Nanjing Normal University, Geographic Sciences Department
南京师范大学地理学学院

30-November-2016

Overview

- Kriging is a **Best Linear Unbiased Predictor** (BLUP) of the value of an attribute at an unsampled location.
- "Best" is defined as the **lowest prediction variance** among all possible combination of weights for the weighted sum prediction.
- Derivation of the weights from this **optimization criterion**[1] depends on a **theory of random fields**.
  - This is a model of how the reality that we observe, and which we want to predict, is structured.
  - The model represents a random process with **spatial autocorrelation**.

[1] 2^{nd} section of this lecture

Theory of random fields

Presentation is based on R. Webster and M. Oliver, 2001 *Geostatistics for environmental scientists*, Chichester etc.: John Wiley & Sons, Ltd.; ISBN 0-471-96553-7

Notation: A point in space of any dimension is symbolized by a **bold-face** letter, e.g. $\mathbf{x}$. In 2D this is (x_1, x_2).

Theory of random fields – key idea

1. **Key idea**: The observed attribute values are only **one of many possible realisations** of a **random process** (also called a "stochastic" process)
2. This random processes is **spatially auto-correlated** , so that attribute values are somewhat **dependent**.
3. At each point $\mathbf{x}$, an **observed value** z is one possiblility of a **random variable** $Z(\mathbf{x})$
4. There is **only one reality** (which is sampled), but it is one **realisation of a process** that **could have** produced many realities. μ and variance σ^2 etc.
5. Cumulative distribution function (CDF): $F\{Z(\mathbf{x}; z)\} = \Pr[Z(\mathbf{x} \leq z_c)]$
6. the probability Pr governs the random process; this is where we can model **spatial dependence**

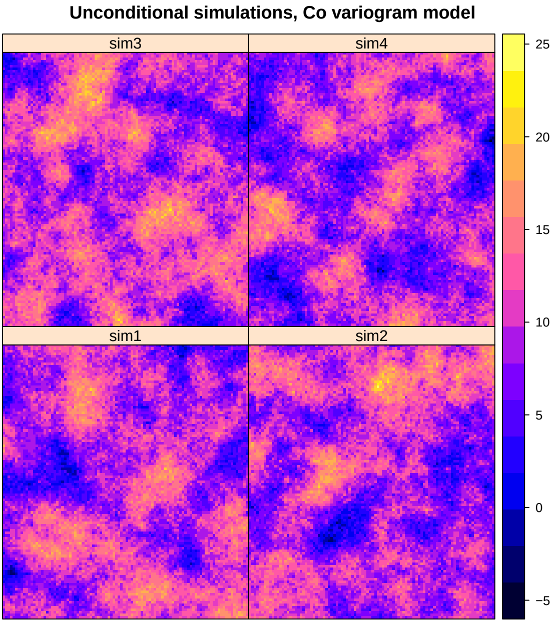
Random functions

- Each point has its **own** random process, but these all have the **same form** (same kind of randomness)
- However, there may be **spatial dependence** among points, which are therefore *not* independent
- As a whole, they make up a **stochastic process** over the whole field R
  - i.e., the observed values are assumed to result from some random process but one that **respects certain restrictions**, in particular spatial dependence
- The set of values of the random variable in the spatial field: $Z = \{Z(\mathbf{x}), \forall \mathbf{x} \in R\}$ is called a **regionalized variable**
- This variable is **doubly infinite**: (1) number of points; (2) possible values at each point

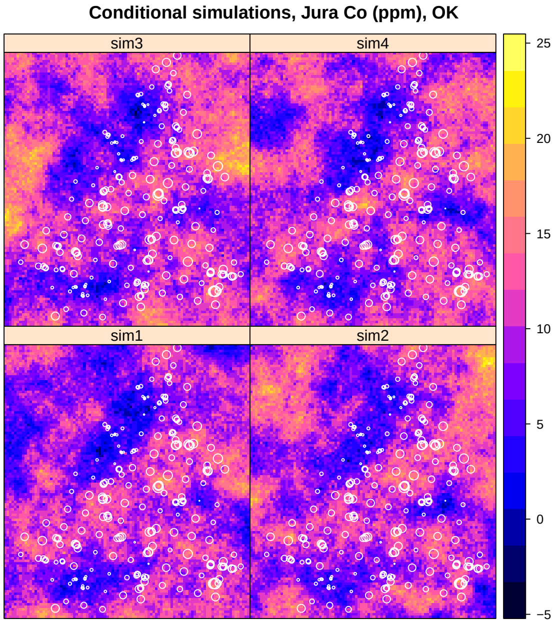
Simulated fields

- Equally probable **realizations** of the same **spatially-correlated random process**
- The fields match an assumed **variogram model**
- These are equally-probable **alternate realities**, assuming the given process
- We can determine which is most likely to be the one reality we have by **sampling.**
- Next pages: simulated fields on a 256x256 grid

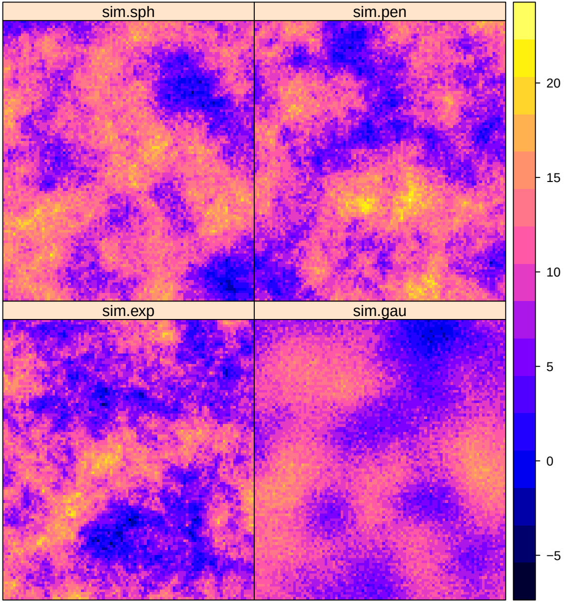
Four realizations of the same field – before sampling



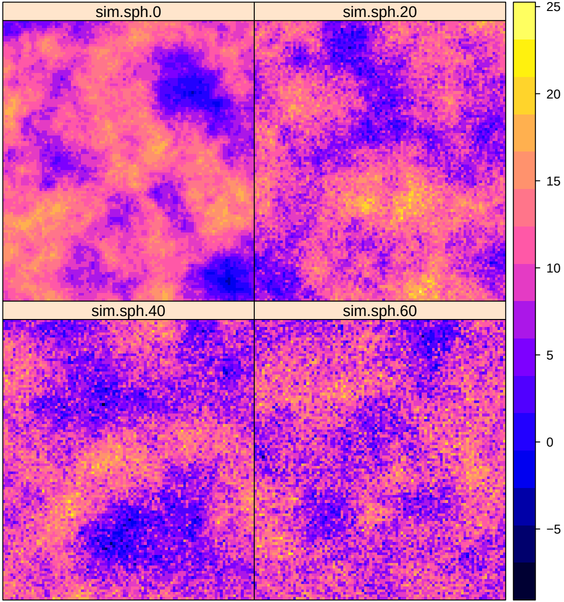
Four realizations of the same field – after sampling



Fields with same variogram parameters, different models

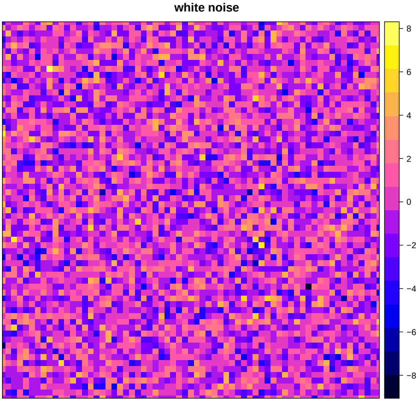


Fields with same model, different variogram parameters



Uncorrelated field with first-order stationarity (same expected value everywhere)

Corresponds to pure “nugget” variogram model.



First-order stationarity

- **Problem**: We have no way to estimate the **expected values** of the random process at each location $\mu(\mathbf{x})$ ...
- ... since we only have **one realisation** (what we actually measure), rather than the whole set of realisations that *could have been* produced by the random process.
- **Solution**: **assume** that the **expected values** at **all locations** in the field are the **same**:

$$E[Z(\mathbf{x}_i)] = \mu, \forall \mathbf{x}_i \in R$$

- This is called **first-order stationarity** of the random process
  - so μ is *not* a function of position $\mathbf{x}$.
- Then we can estimate the (common) expected value from the sample values and their spatial structure.

Problems with first-order stationarity

- It is often not plausible:
  1. We observe the mean value to be different in several regions (strata)
  2. We observe a regional trend
- In both cases there is a **process** that is not stationary which we can **model**:
  1. model the strata or trend, then the **residuals** may be first-order stationary → **Kriging with External Drift** or **Regression Kriging**)
  2. model a varying mean along with the local structure → **Universal Kriging**
- Another solution: study the **differences** between values, not the **values** themselves, and in a **“small”** region.

Spatial covariance

- Key idea: **nearby observations may be correlated**.
- Since it's the same variable, this is **autocorrelation**
- There is only **one realisation of the random field** per point, but **each point is a different realisation**, so in some sense they are *different* variables, which then have a covariance.
- **Key Insight:** Under certain assumptions (see below), this covariance can be considered to **depend only on the separation** (and possibly the direction) between the points.

Covariance

- Recall from non-spatial statistics: the **sample covariance** between **two variables** z_1 and z_2 observed at n points is:

$$\widehat{C}(z_1, z_2) = \frac{1}{n} \sum_{i=1}^{n} (z_{1_i} - \bar{z}_1) \cdot (z_{2_i} - \bar{z}_2)$$

- **Spatial version**: there is only **one variable** x:

$$\widehat{C}(\mathbf{x}_1, \mathbf{x}_2) \quad = \quad E[\{Z(\mathbf{x}_1) - \mu(\mathbf{x}_1)\} \cdot \{Z(\mathbf{x}_2) - \mu(\mathbf{x}_2)\}]$$

- Because of first-order stationarity, the expected values are the same, so:

$$\widehat{C}(\mathbf{x}_1, \mathbf{x}_2) \quad = \quad E[\{Z(\mathbf{x}_1) - \mu\} \cdot \{Z(\mathbf{x}_2) - \mu\}]$$

Second-order stationarity (1) – At one point

- **Problem**: The covariance **at one point** is its **variance**:

$$\sigma^2 = E[\{Z(\mathbf{x}_i) - \mu\}^2]$$

- This can not be estimated from one sample of the many possible realisations.
- **Solution**: **assume that the variance is the same finite value at all points**.
- Then we can estimate the variance of the process from the sample by considering all the random variables (at different points) together.
- This assumption is part of **second-order stationarity**

Second-order stationarity (2) – Over the spatial field

- **Problem**: The covariance equation as written is between all the points in the field. It is huge! And again, there is no way to estimate these from just one point pair per variable pair.
- **Solution** (the key insight): **Assume that the covariance between points depends only on their separation**
- Then we can **estimate** their covariance from a **large number of sample pairs**, all **separated** by (approximately) the same separation vector **h** (distance, possibly with direction).

Derivation of the covariance function

- **Autocovariance** ('auto' = same regionalized variable), at a separation $\mathbf{h}$:

$$\begin{aligned} C[Z(\mathbf{x}), Z(\mathbf{x}+\mathbf{h})] &= E[\{Z(\mathbf{x}) - \mu\} \cdot \{Z(\mathbf{x}+\mathbf{h}) - \mu\}] \\ &= E[\{Z(\mathbf{x})\} \cdot \{Z(\mathbf{x}+\mathbf{h})\} - \mu^2] \\ &\equiv C(\mathbf{h}) \end{aligned}$$

- **Autocorrelation**: Autocovariance normalized by total variance σ^2, which is the covariance at a point:

$$\rho(\mathbf{h}) = C(\mathbf{h}) / C(\mathbf{0})$$

- **Semivariance**: deviation of covariance at some separation from total variance:

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

Characteristics of Spatial Correlation functions

- symmetric: $C(\mathbf{h}) = C(-\mathbf{h})$ etc.
- range of $\rho(\mathbf{h}) \in [-1 \cdots 1]$
- Positive (covariance) or negative (variogram) **semi-definite matrices**; this restricts the choice of models
- Continuity, especially at **0**. But this is often not observed, the “nugget” effect.
  - Solved by adding a nugget structure to the spatial correlation model.

Problems with second-order stationarity

- It assumes the existence of a covariance and, so, a finite variance $\mathbf{Var}(Z(\mathbf{x})) = C(\mathbf{0})$
- This is often not plausible; in particular the covariance often increases without bound as the area increases.
- **Solutions**
  1. Study the **differences** between values, not the **values** themselves, and in a **"small"** region; then he covariances may be bounded → the **intrinsic hypothesis** (see next);
  2. So, model the **semi-variance**, not **co-variance**.
  3. This is a weaker assumption.

The Intrinsic Hypothesis

- Replace mean **values** $Z(\mathbf{x})$ with mean **differences**, which are the same over the whole random field, at least within some 'small' separation $\mathbf{h}$. Then the expected value is 0:

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

- Replace **covariance of values** with **variances of differences**:

$$\mathrm{Var}[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})] = E[\{Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})\}^2] = 2\gamma(\mathbf{h})$$

- The equations only involve the **difference in values** at a separation, not the **values**, so the necessary assumption of finite variance need only be assumed for the differences, a less stringent condition.
- This is the **intrinsic hypothesis**.

Using the empirical variogram to model the random process

- The semivariance of the separation vector $\gamma(\mathbf{h})$ is now given as the estimate of covariance in the spatial field.
- It models the **spatially-correlated component** of the **regionalized variable**
- We must go from the **empirical variogram** to a **variogram model** in order to be able to model the random process at any separation.

Ordinary Kriging

In this section we:

1. Present OK and its optimization criterion;
2. Derive a computable form of the prediction variance;
3. Derive the OK system of equations;
4. Show the solution to the OK syste,

This all depends on the theory of random fields,

Ordinary Kriging (OK) – Overview

- A **linear** predictor of the value at an **unknown** location, given the **locations** of a set of points and their known **values.**
- The linear predictor is a **weighted sum** of the known values.
- The weights are based on a model of **spatial autocorrelation** between the known values.
- This model is of the assumed **spatial autocorrelated random process** that produced a **random field.**
- We have **observed** the values of some attribute at some locations in this random field, we want to **predict** others.

OK as a weighted sum

- The estimated value $\hat{z}$ at a point $\mathbf{x}_0$ is predicted as the **weighted average** of the values at the observed points $\mathbf{x}_i$:

$$\hat{z}(\mathbf{x}_0) = \sum_{i=1}^{N} \lambda_i z(\mathbf{x}_i)$$

- The weights λ_i assigned to the observed points must **sum to 1**:

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

- Therefore, the prediction is **unbiased** with respect to the underlying random function Z:

$$E[\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)] = 0$$

What makes it "Ordinary" Kriging?

- The expected value (mean) is **unknown**, and must be estimated from the sample
  - If the mean is **known** we use **Simple Kriging** (SK)
- There is **no regional trend**
  - If so we use **Universal Kriging** (UK).
- There is **no feature-space predictor**, i.e. one of more other attributes, known at both the observation and prediction points, that helps explain the attribute of interest
  - If so we use **Kriging with External Drift** (KED) or **Regression Kriging** (RK).

Weighted-sum linear predictors

- There are many of these! The only restriction is $\sum_{i=1}^{N} \lambda_i = 1$
  - All weight to closest observation (nearest-neighbour) $\lambda_i = 1, \lambda_{j \neq i} = 0$.
  - Average of points within some specified distance (average in radius)
  - Average of some number of nearest points
  - Inverse distance-weighted averages within radius or some specified number
  - declustered versions of these
  - **kriging**: weights derived from the kriging equations
- **Which is "optimal"**?
  - To decide, we need an **optimization criterion**.

Optimization criterion

- "Optimal" depends on some **objective function** which can be **minimized** with the best weights;
- We choose the **variance of the prediction** as the **objective function**; i.e. we want to **minimize** the **uncertainty** of the prediction.

Prediction Variance

For **any** predictor (not just the kriging predictor):

- The **prediction** $\hat{z}(\mathbf{x}_0)$ at a given location $\mathbf{x}_0$ may be compared to the **true** value $z(\mathbf{x}_0)$; note the "hat" symbol to indicate an **estimated** value rather than a **measured** one.
- Even though we don't know the true value, we can write the expression for the **prediction variance**
- This is defined as the **expected value** of the **squared difference** between the **estimate** and the (unknown) **true value**:

$$\sigma^2(Z(\mathbf{x}_0)) \equiv E\left[\left\{\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)\right\}^2\right]$$

- If we can express this in some **computable form** (i.e. without the unknown true value) we can use it as an **optimality criterion**

Derivation of the prediction variance

(Based on P K Kitanidis, *Introduction to geostatistics: applications to hydrogeology*, Cambridge University Press, 1997, ISBN 0521583128; §3.9)

1. In OK, the estimated value is a **linear combination** of data values $\mathbf{x}_i$, with weights λ_i derived from the **kriging system**:

$$\hat{z}(\mathbf{x}_0) = \sum_{i=1}^{N} \lambda_i z(\mathbf{x}_i)$$

2. Kriging variance:

$$\sigma^2(Z(\mathbf{x}_0)) \equiv E\left[\left\{ \hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0) \right\}^2 \right]$$

3. Re-write the kriging variance with the weighted linear sum:

$$\sigma^2(Z(\mathbf{x}_0)) = E\left[\left\{ \sum_{i=1}^{N} \lambda_i z(\mathbf{x}_i) - Z(\mathbf{x}_0) \right\}^2 \right]$$

Expanding into parts

4. Add and subtract the unknown mean μ:

$$\sigma^2(z(\mathbf{x}_0)) = E\Big[\{\sum_{i=1}^{N}\lambda_i(z(\mathbf{x}_i)-\mu)-(Z(\mathbf{x}_0)-\mu)\}^2\Big]$$

5. Expand the square:

$$\begin{aligned}\sigma^2(Z(\mathbf{x}_0)) \quad = \quad & E\Big[(\sum_{i=1}^{N}\lambda_i z(\mathbf{x}_i)-\mu)^2 \\ & -2\sum_{i=1}^{N}\lambda_i(z(\mathbf{x}_i)-\mu)(Z(\mathbf{x}_0)-\mu) \\ & +(Z(\mathbf{x}_0)-\mu)^2\Big]\end{aligned}$$

Bring expectations into each term

6. Replace the squared single summation (first term) by a **double summation**, i.e. with separate indices for the two parts of the square (two observation points):

$$\left(\sum_{i=1}^{N} \lambda_i z(\mathbf{x}_i) - \mu\right)^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j (z(\mathbf{x}_i) - \mu)(z(\mathbf{x}_j) - \mu)$$

7. Bring the expectation into each term[2]:

$$\begin{aligned}
\sigma^2(Z(\mathbf{x}_0)) \quad = \quad & \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j E\left[(z(\mathbf{x}_i) - \mu)(z(\mathbf{x}_j) - \mu)\right] \\
& -2 \sum_{i=1}^{N} \lambda_i E\left[(z(\mathbf{x}_i) - \mu)(Z(\mathbf{x}_0) - \mu)\right] \\
& +E\left[(Z(\mathbf{x}_0) - \mu)^2\right]
\end{aligned}$$

[2] expectation of a sum is the sum of expectations

From expectations to covariances

8. The three **expectations** in the previous expression are the definitions of **covariance** or **variance**:
   1. $E\left[(z(\mathbf{x}_i)-\mu)(z(\mathbf{x}_j)-\mu)\right]$: **covariance** between **two observation points**
   2. $E\left[(z(\mathbf{x}_i)-\mu)(Z(\mathbf{x}_0)-\mu)\right]$: **covariance** between **one observation point** and the **prediction point**
   3. $E\left[(Z(\mathbf{x}_0)-\mu)^2\right]$: **variance** at the **prediction point**
9. So, replace the **expectations** with **covariances** and **variances**:

$$
\begin{aligned}
\sigma^2(Z(\mathbf{x}_0)) \quad = \quad & \sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j\mathrm{Cov}(z(\mathbf{x}_i),z(\mathbf{x}_j)) \\
& -2\sum_{i=1}^{N}\lambda_i\mathrm{Cov}(z(\mathbf{x}_i),Z(\mathbf{x}_0)) \\
& +\mathrm{Var}(Z(\mathbf{x}_0))
\end{aligned}
$$

How can we evaluate this expression?

- **Problem 1**: how do we know the **covariances** between any two points?
  - **Answer**: by applying a **covariance function** which **only depends on spatial separation** between them.
- **Problem 2**: how do we find the correct covariance function?
  - **Answer**: by fitting a **variogram model** to the **empirical variogram**
- **Problem 3**: how do we know the **variance** at any point?
  - **Answer**: The actual value doesn't matter
  - it will be eliminated in the following algebra . . .
  - . . . but it **must be the same at all points**
  - this is the assumption of **second-order stationarity**.

Stationarity (1)

- This is a term for **restrictions on the nature of spatial variation** that are required for OK to be correct
- **First-order** stationarity: the **expected values** (mean of the random function) at all locations in the random field are the **same**:

$$E[Z(\mathbf{x}_i)] = \mu, \forall \mathbf{x}_i \in R$$

- **Second-order** stationarity:
  1. The **variance** at any point is **finite** and the **same at all locations** in the field
  2. The **covariance structure** depends only on **separation** between point pairs

Stationarity (2)

- The concept of stationarity is often confusing, because stationarity refers to **expected** values, variances, or co-variances, rather than **observed** values.
  - Of course the actual values change over the field! That is exactly what we want to use to predict at unobserved points.
- First-order stationarity just says that **before we sampled**, the **expected** value at all locations was the same.
  - That is, we **assume** the values result from a **spatially-correlated process** with a constant **mean** – *not* constant **values**.
- Once we have some observation values, these influence the probability of finding values at other points, because of **spatial covariance**.

Unbiasedness

- An **unbiased** estimate is one where the **expectation** of the **estimate** equals the expectation of the **true** (unknown) value:

$$E\left[\hat{z}(\mathbf{x}_0\right] \equiv E\left[Z(\mathbf{x}_0)\right] = \mu$$

- We will estimate $E\left[\hat{z}(\mathbf{x}_0\right]$ as a weighted sum:

$$E\left[\hat{z}(\mathbf{x}_0\right] = \sum_{i=1}^{N} \lambda_i E\left[z(\mathbf{x}_i)\right] = \sum_{i=1}^{N} \lambda_i \mu = \mu \sum_{i=1}^{N} \lambda_i$$

- Since $E\left[\hat{z}(\mathbf{x}_0)\right] = \mu$ (unbiasedness), we must have

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

This is a **constraint** in the kriging system.

From point-pairs to separation vectors

- As written above, the covariances between **all point-pairs** must be determined separately, and since there is only one realization of the random field, it's imposible to know these from the observations.
- However, because of **second-order stationarity**, we can assume that the covariances between any two points depend **only on their separation** and a single **covariance function**.
- So rather than try to compute all the covariances, we just need to know this function, then we can apply it to any point-pair, just by knowing their **separation** and this function.
- So, **OK is only as good as the covariance model!**

Replace point-pairs with separation vectors

Continuing the derivation of the OK equations:

10. Substitute the **covariance function** of **separation** $\mathbf{h}$ into the expression:

$$\begin{aligned}\sigma^2(Z(\mathbf{x}_0)) &= \sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j \mathrm{Cov}(\mathbf{h}(i,j)) \\ &\quad -2\sum_{i=1}^{N}\lambda_i \mathrm{Cov}(\mathbf{h}(i,0)) \\ &\quad +\mathrm{Cov}(\mathbf{0})\end{aligned}$$

- $\mathbf{h}(i,0)$ is the separation between the observation point $\mathbf{x}_I$ and the point to be predicted $\mathbf{x}_0$.
- $\mathbf{h}(i,j)$ is the separation between two observation points $\mathbf{x}_i$ and $\mathbf{x}_j$.
- $\mathrm{Cov}(\mathbf{0})$ is the variance of the random field at a point

From covariances to semivariances

11. Replace covariances by **semivariances**, using the relation $\mathrm{Cov}(\mathbf{h}) = \mathrm{Cov}(\mathbf{0}) - \gamma(\mathbf{h})$:

$$\sigma^2(Z(\mathbf{x}_0)) = -\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j\gamma(\mathbf{h}(i,j)) + 2\sum_{i=1}^{N}\lambda_i\gamma(\mathbf{h}(i,0))$$

Replacing covariances by semivariances changes the sign.

- First term: **depends on the covariance structure of the known points**; the greater the product of the two weights for a given semivariance, the lower the prediction variance (note $-$ sign)
- Second term: **depends on the covariance between the point to be predicted and the known points**

- This is now a **computable** expression for the **kriging variance** at any point $\mathbf{x}_0$, given the **locations** of the observation points $\mathbf{x}_i$, once the **weights** λ_i are known.

Computing the weights

- Q: **How do we compute the weights** λ to predict at a given point?
- A: We compute these weights for each point to be predicted, by an **optimization criterion**, which in OK is **minimizing the kriging variance**.

Objective function (1): Unconstrained

- In a minimization problem, we must define an **objective function** f to be minimized. In this case, it is the kriging variance in terms of the N weights λ_i:

$$f(\lambda) = 2\sum_{i=1}^{N} \lambda_i \gamma(\mathbf{x}_i, \mathbf{x}_0) - \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_i \lambda_j \gamma(\mathbf{x}_i, \mathbf{x}_j)$$

- This expression is **unbounded** and can be trivially solved by setting all weights to 0. We must add **another constraint** to **bound** it.

Objective function (2): Constrained

- The added constraint is **unbiasedness**: the **weights must sum to 1**.
- This is added as term in the function to be optimized, along with a new argument to the function, the **LaGrange multipler** ψ

$$
\begin{aligned}
f(\lambda, \psi) \quad = \quad & 2 \sum_{i=1}^{N} \lambda_i \gamma(\mathbf{x}_i, \mathbf{x}_0) \\
& - \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j \gamma(\mathbf{x}_i, \mathbf{x}_j) \\
& - 2\psi \left\{ \sum_{i=1}^{N} \lambda_i - 1 \right\}
\end{aligned}
$$

- The last term $\equiv 0$, i.e. the prediction is unbiased; ψ must be added as a variable so there is one variable per equation

Minimization

- **Minimize** by setting all $N+1$ **partial derivatives** to zero (N prediction points; 1 constraint):

$$\begin{aligned} \frac{\partial f(\lambda_i, \psi)}{\partial \lambda_i} &= 0, \forall i \\ \frac{\partial f(\lambda_i, \psi)}{\partial \psi} &= 0 \end{aligned}$$

- In the differential equation with respect to ψ, all the λ are constants, so the first two terms differentiate to 0; in the last term the ψ differentiates to 1 and we are left with the unbiasedness condition:

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

The Ordinary Kriging system

- In addition to unbiasedness, the partial derivatives with respect to the λ_i give N equations (one for each λ_i) in $N+1$ unknowns (the λ_i plus the LaGrange multiplier ψ):

$$\sum_{j=1}^{N} \lambda_j \gamma(\mathbf{x}_i, \mathbf{x}_j) + \psi = \gamma(\mathbf{x}_i, \mathbf{x}_0), \quad \forall i$$

- This is now a **system of $N+1$ equations in $N+1$ unknowns** and can be solved by linear algebra.

Solving the OK system

1. For the system as a whole: compute the semivariances **between all pairs of observed points** $\gamma(\mathbf{x}_i, \mathbf{x}_j)$ from their separation, according to the **variogram model**
2. At each point $\mathbf{x}_0$ to be predicted:
    1. **Compute the semivariances** $\gamma(\mathbf{x}_i, \mathbf{x}_0)$ from the separation between the point and the observed values, according to the **variogram model**
    2. **Solve simultaneously** for the weights and multiplier
    3. **Compute the predicted value** as the **weighted average** of the observed values, using the computed weights
    4. **Compute the kriging variance**.

Matrix form of the OK system

$$\mathbf{A}\lambda = \mathbf{b}$$

$$\mathbf{A} = \begin{bmatrix} \gamma(\mathbf{x}_1, \mathbf{x}_1) & \gamma(\mathbf{x}_1, \mathbf{x}_2) & \cdots & \gamma(\mathbf{x}_1, \mathbf{x}_N) & 1 \\ \gamma(\mathbf{x}_2, \mathbf{x}_1) & \gamma(\mathbf{x}_2, \mathbf{x}_2) & \cdots & \gamma(\mathbf{x}_2, \mathbf{x}_N) & 1 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \gamma(\mathbf{x}_N, \mathbf{x}_1) & \gamma(\mathbf{x}_N, \mathbf{x}_2) & \cdots & \gamma(\mathbf{x}_N, \mathbf{x}_N) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix}$$

$$\lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \\ \psi \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} \gamma(\mathbf{x}_1, \mathbf{x}_0) \\ \gamma(\mathbf{x}_2, \mathbf{x}_0) \\ \vdots \\ \gamma(\mathbf{x}_N, \mathbf{x}_0) \\ 1 \end{bmatrix}$$

Inside the OK Matrices

The **block matrix** notation shows the semivariances and LaGrange multiplier explicitly:

$$\mathbf{A} = \begin{pmatrix} \Gamma & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix}$$

$$\lambda = \begin{bmatrix} \Lambda \\ \psi \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} \Gamma_0 \\ 1 \end{bmatrix}$$

Solution of the OK system

- This is a system of $N+1$ equations in $N+1$ unknowns, so can be solved if $\mathbf{A}$ is positive definite; this is guaranteed by using authorized models:

$$\lambda = \mathbf{A}^{-1}\mathbf{b}$$

- **Predict** at a point $\mathbf{x}_0$, using the weights:

$$\hat{Z}(\mathbf{x}_0) = \sum_{i=1}^{N} \lambda_i z(\mathbf{x}_i)$$

- The **kriging variance** at a point is:

$$\hat{\sigma}^2(\mathbf{x}_0) = \mathbf{b}^T\lambda$$

- The last element of λ is ψ, which depends on covariance structure of the observed points.

End

Key points to remember about OK:

1. It depends on the theory of **random fields.**
2. It requires the assumption of 1st and 2nd order **stationarity**
3. Computations are based on a **model** of spatial autocorrelation
4. Its prediction is the Best Linear Unbiased Predictor (BLUP), if "best" means **"lowest prediction variance".**