

SPATIAL STATISTICS

21/05

- (1.) • Geostatistical data
- (2.) • lattice data (areal data)
- (3.) • Point processes

The models we'll see tackle the dependence when there is sort of basic assumption that is:

"when 2 data have been observed in places that are close are more dependent of 2 data

observed in places that are far away", with

"closedness" introduced by a distance (we decide)

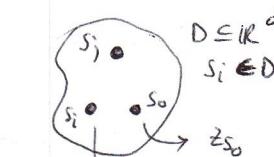
we assume that we have some data (observations) and these observations refer to some geogr. locations

(1.) $s_1, \dots, s_n \in D \subseteq \mathbb{R}^d$ fixed : geographical locations
 z_{s_1}, \dots, z_{s_n} r.v. observations

Goals :

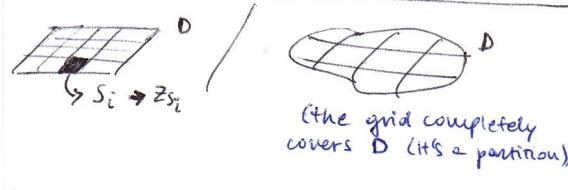
- Models (of z_s , $j = \dots$)
- Study the spatial dependence

PREDICTION → **KRIGING**



$D \subseteq \mathbb{R}^d$
 $s_i \in D$
 z_{s_i} → observation in the zone S_i
 $s_0 \notin D$
 z_{s_0} → we have z_{s_j} , $j = \dots$ and we want to predict z_{s_0} in S_0 which is not a known zone

(2.) s_1, \dots, s_n grid locations (fixed but not points)
 z_{s_i} observation in S_i



Goals :

- Model
- Clustering (we don't need prediction)

(3.) Point patterns:

$\{s_1, \dots, s_n\}$ random locations

- Standard point process
- Market point process (locations + observations attached)

(s_i, z_{s_i})
↑ random site ↑ obs. features

Ex. (earthquake zone, magnitude)

s_1, s_2, \dots, s_n

We have random locations in the sense that:
when they're realized they're the observations themselves

Goals :

- is there a clustering?
- is there more presence in some areas than in others?

Examples: earthquakes happen in random locations? There are areas more subjected than others?

GEOSTATISTICAL DATA

Example: $s_1, \dots, s_n \in D$

z_{s_1}, \dots, z_{s_n} random variables

$$z_{s_i} = \sum_{e=0}^E a_e f_e(s_i) + \delta_{s_i}$$

↑ regression coefficients (β) ↑ regressors

"residuals" (ε)

$s_i \in D$

Goal: estimate the parameters a

First: How these δ 's are distributed?

$$\mathbb{E}[\delta_{s_i}] = 0$$

$$\text{Cov}(\delta_{s_i}, \delta_{s_j}) \neq 0$$

since the spatial closedness induces a special dependence

$$\underline{\delta} = [\delta_{s_1}, \dots, \delta_{s_n}]$$

$$\text{Cov}(\underline{\delta}) = \sum (\neq \sigma^2 I)$$

If we forget about spatial dependence and we use OLS:

$$\text{OLS: } \hat{\alpha}^{\text{OLS}} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \underline{z}$$

↓
design matrix ↗ vector of observations

$$\text{Cov}(\hat{\alpha}^{\text{OLS}}) = (\mathbf{F}^T \mathbf{F})^{-1} (\mathbf{F}^T \Sigma \mathbf{F}) (\mathbf{F}^T \mathbf{F})^{-1}$$

$$\text{GLS: } \hat{\alpha}^{\text{GLS}} = (\mathbf{F}^T \Sigma^{-1} \mathbf{F})^{-1} \mathbf{F}^T \Sigma^{-1} \underline{z}$$

$$\text{Cov}(\hat{\alpha}^{\text{GLS}}) = (\mathbf{F}^T \Sigma^{-1} \mathbf{F})^{-1}$$

$$\text{Cov}(\hat{\alpha}^{\text{OLS}}) - \text{Cov}(\hat{\alpha}^{\text{GLS}}) \geq 0$$

How to measure:

SPATIAL DEPENDENCE

s_1, \dots, s_n site

z_{s_1}, \dots, z_{s_n} observations : we assume that these observations come from a random process (z_s is a random process, which is a collection of random variables)

- Assumptions:
1. $E[z_s] < \infty \quad \forall s \in D$
 2. $\text{Var}(z_s) < \infty \quad \forall s \in D$

It's not enough tho. We have to make more sever constraints to estimate the spatial dependence to perform prediction:

Def. Spatial mean of $\{z_s, s \in D\}$:

$$m_s = E[z_s] \quad s \in D$$

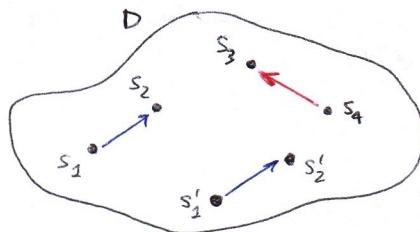
Def. Covariance function:

$$C(s_1, s_2) = \text{Cov}(z_{s_1}, z_{s_2})$$

Def. $\{z_s, s \in D\}$ second order stationary if:

1. $E[z_s] = m \quad \forall s \in D$
 2. $\text{Cov}(z_{s_1}, z_{s_2}) = \frac{C(s_1 - s_2)}{\uparrow} \quad s_1, s_2 \in D$
- (Note of notations:
 $s_i \in \mathbb{R}^k \quad k \neq 1$)

\therefore COVARIogram : "function that quantifies the covariance between two element of the process"



what we are assuming is: In every \bullet the mean is the same and if two points are separated by a separated vector (\rightarrow) and other two points are separated by the same separating vector then the two locations (two couples of pairs) will be characterized by the same type of covariance (the DIRECTION MATTERS: even if (s_3, s_4) have a separated vector (\rightarrow) of the same length as (\rightarrow), the direction is $\neq \Rightarrow$ the covariance is \neq)

Properties: (algebraic)

1. Symmetry: $C(-h) = C(h) \quad \forall h \in \mathbb{R}^d$

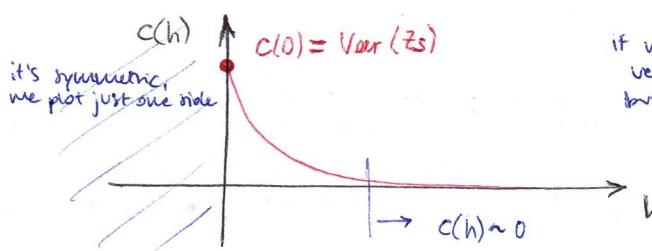
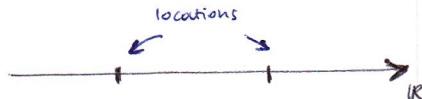
2. Bounded: $|C(h)| \leq C(0) \quad \forall h \in \mathbb{R}^d$

CAUCHY-SCHWARTZ
(covariance is dominated by variance)

3. Positive definite: $\sum_{ij} \lambda_i \lambda_j C(s_i - s_j) \geq 0$

$\forall \lambda_i, \lambda_j \in \mathbb{R}$
 $s_i, s_j \in D$

Example: $D \subset \mathbb{R}$
Covariogram in \mathbb{R}^d



If we have small values of length of separating vector we have a high covariance among data, but at a certain point the locations will be so far from each other that the correlation between data will disappear
so, typically $c(h) \xrightarrow{h \rightarrow \infty} 0$

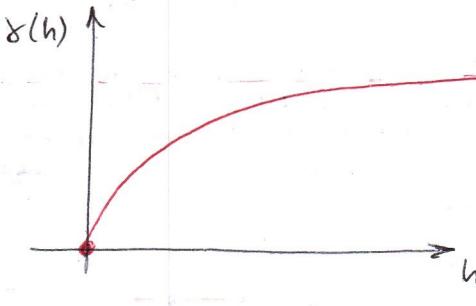
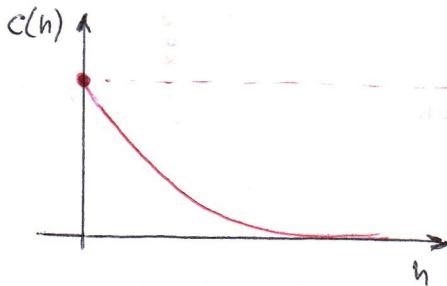
Def. Variogram (under 2nd order stationarity):

$$\begin{aligned} 2\gamma(s_1 - s_2) &= \text{Var}(z_{s_1} - z_{s_2}) \\ (\text{alternative measure of dependence}) \quad &\downarrow \\ &= E[(z_{s_1} - z_{s_2})^2] - (m_{s_1} - m_{s_2})^2 \quad (\text{under stationarity}) \end{aligned}$$

Note: $2\gamma(s_1 - s_2) = \text{Var}(z_{s_1}) + \text{Var}(z_{s_2}) - 2\text{Cov}(z_{s_1}, z_{s_2})$
 \downarrow
 $= 2c(0) - 2c(s_1 - s_2)$

$$\begin{aligned} \gamma(s_1 - s_2) &= c(0) - c(s_1 - s_2) \\ &:= \text{SEMIVARIOGRAM} \end{aligned}$$

(once we know the covariogram we know the variogram (and often viceversa))



Properties: (algebraic)

1. Symmetry: $\gamma(-h) = \gamma(h)$

2. Null at 0: $\gamma(0) = 0$

K. Conditional negative definite: $\sum_{ij} \lambda_i \lambda_j \gamma(s_i - s_j) \leq 0$ $\lambda_i, \lambda_j \in \mathbb{R}$

Remember: when we're estimating the variogram/covariogram we have to check that they fulfill these properties

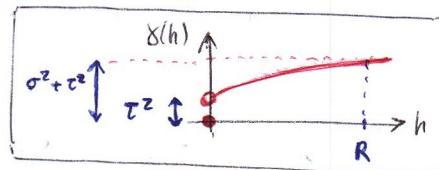
$$\sum \lambda_i = 0$$

$$\forall s_i, s_j \in D$$

Structural properties:

τ^2 nugget effect

(related to the kind of smoothness expect from the realizations of our variogram) and regularity that we have of the process as functions



the semi-covariogram is $\gamma(0) = 0$ but it may be DISCONTINUED

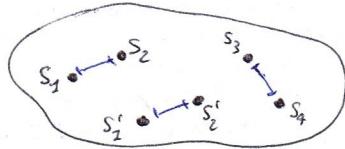
partial sill (σ^2); the higher σ^2 the higher the overall variability of the process

R range : value of h st. the variogram reaches the sill

$\rightarrow R$ quantify the amount of dependence that we have in our data ($R = \text{range of dependence in the field}$)

Def. Isotropy : A 2nd order stationary field is isotropic if

$$\text{Cov}(z_{s_i}, z_{s_j}) = C(\|s_i - s_j\|) \quad s_i, s_j \in D$$



the (length) of the distance establishes if two locations (couple of pairs) have the same covariance with this definition is more appropriate makes $b(h) \uparrow h$ and considering b as a distance

ESTIMATE SPATIAL DEPENDENCE

: typically through the variogram

→ Estimate the variogram

Assumptions: 2nd order stationary & isotropy

$$Z \propto (S_1 - S_2) = \mathbb{E}[(Z_{S_1} - Z_{S_2})^2] + (m_{S_1} - m_{S_2})^2$$

$\downarrow \text{Var}(Z_{S_1} - Z_{S_2})$

since we have stationarity

Empirical Estimate:

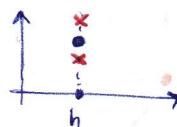
$$2\hat{\delta}(h) = \frac{1}{|N(h)|} \sum_{(i,j) \in N(h)} (z_{s_i} - z_{s_j})^2$$

$$N(h) = \{(i,j) : \|s_i - s_j\| = h\}$$

If we have our data:

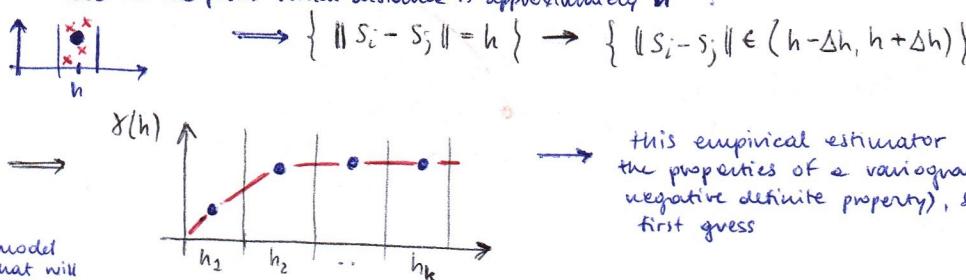


We want to estimate the variogram at a given distance b :



we take all the pairs of the data which distance is b , we take the square difference and take an average of these differences

Problem: we don't have so many points (for all the b_i).
 we divide and we have a tolerance of distances: $b \rightarrow \Delta b$
 and we take all the pairs which distance is approximately b



We need a model
that we know that will
fit all the properties

Model: parametric families

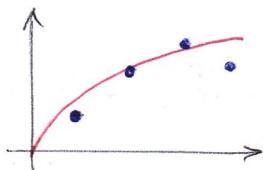
$$z \gamma(h, \theta)$$

$\theta \in \Theta \rightarrow$ we have just to estimate θ

- $\gamma_1 + \gamma_2$ valid variogram model if γ_1 and γ_2 are valid
 - $\lambda \gamma_1$ valid variogram if γ_1 valid, $\lambda > 0$

We have to find:

$\theta \rightarrow$ best parameter vector in Θ to fit the empirical estimate



- $\hat{\theta}$ = empirical estimate
- $\hat{\theta}_S$ = spherical model based on $\hat{\theta}$
($\hat{\theta}$ estimated in order to best fit
the empirical estimate)

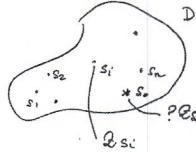
OLS: (for example)

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{k=1}^K (\delta(h_k) - \delta(h_k, \theta))^2$$

real random variables

s_1, \dots, s_n sites in $D \subseteq \mathbb{R}^d$
 z_{s_1}, \dots, z_{s_n} obs.

Goal: Build a predictor for z_{s_0}



How to make prediction at a location not observed?

Random field $\{z_s, s \in D\} \rightarrow$ the problem of prediction is to predict one element of this random field

In general: $\hat{z}_{s_0} = f(z_{s_1}, \dots, z_{s_n})$ where f is "best":

$$f: \min \mathbb{E}[(z_{s_0} - f(z_{s_1}, \dots, z_{s_n}))^2] \quad (*)$$

Solution: $\mathbb{E}[z_{s_0} | z_{s_1}, \dots, z_{s_n}]$

Properties:

(1) It minimizes:

$$\mathbb{E}[(z_{s_0} - f(z_{s_1}, \dots, z_{s_n}))^2 | z_{s_1}, \dots, z_{s_n}]$$

(2) Unbiased:

$$\mathbb{E}[\mathbb{E}[z_{s_0} | z_{s_1}, \dots, z_{s_n}]] = \mathbb{E}[z_{s_0}]$$

(3) Interpolator: $s_i \rightarrow s_i$

$$\mathbb{E}[z_{s_i} | z_{s_1}, \dots, z_{s_n}] = z_{s_i}$$

If the target is one of the sample sites then the predictor (conditional expectation) coincides with the data itself

(4) If $\{z_s, s \in D\}$ is Gaussian (If the field is gaussian = all the collections of observations that we get from this random field will be a gaussian vector)

$$\mathbb{E}[z_{s_0} | z_{s_1}, \dots, z_{s_n}] = \lambda_0 + \sum_i \lambda_i z_{s_i}$$

The problem is: if our field is not gaussian then the function f of our data can be highly non-linear. How can we find this function without putting gaussian assumptions on data?

That's why Kriging doesn't look for a conditional expectation but it looks for a linear predictor.

→ We look for the best linear combination of data that minimizes (*).

KRIGING: Look for the Best Linear Unbiased Predictor

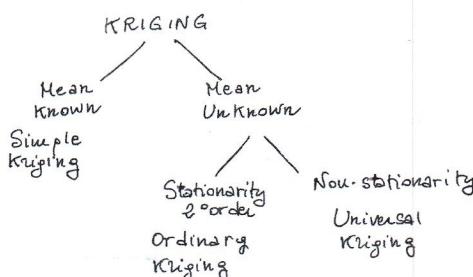
$$\hat{z}_{s_0} = \lambda_0 + \sum_i \lambda_i^* z_{s_i}$$

where $\lambda_0^*, \lambda_1^*, \dots, \lambda_n^*$ solve

$$\begin{aligned} \min_{\lambda} & \mathbb{E}[(z_{s_0} - (\lambda_0 + \sum_i \lambda_i z_{s_i}))^2] \\ \text{s.t.} & \mathbb{E}[\lambda_0 + \sum_i \lambda_i z_{s_i}] = \mathbb{E}[z_{s_0}] \end{aligned}$$

- linear
- unbiased
- must interpolate the data

(K)



We can distinguish different types of Kriging based on the assumptions on the field.

ORDINARY KRIGING

Assumptions: (1) $\{z_s, s \in D\}$ 2-order stationary random field

$$\bullet \mathbb{E}[z_s] = m \quad \forall s \in D$$

$$\bullet \text{Cov}(z_{s_1}, z_{s_2}) = C(s_1 - s_2) \quad \forall s_1, s_2 \in D$$

(2) C is known (while m is unknown)

covariogram (\hat{C})

Kriging Problem:

$$\hat{z}_{s_0} = \lambda_0 + \sum_i \lambda_i z_{s_i}$$

Find $\lambda_0, \lambda_1, \dots, \lambda_n$

$$\min \mathbb{E}[(z_{s_0} - \hat{z}_{s_0})^2]$$

$$\text{s.t. } \mathbb{E}[\hat{z}_{s_0}] = \mathbb{E}[z_{s_0}]$$

$\hat{z}_{s_0}^{\lambda}$ = general predictor (linear) from our data

(1)

(I) Unbiasedness → we have to transform it in a constraint for our weights

$$\mathbb{E}[\lambda_0 + \sum_i \lambda_i z_{s_i}] = \lambda_0 + \sum_i \lambda_i \mathbb{E}[z_{s_i}]$$

$$= \lambda_0 + \sum_i \lambda_i m = \mathbb{E}[z_{s_0}] = m$$

" = " we're imposing it

$$\rightarrow \begin{cases} \lambda_0 = 0 \\ \sum_i \lambda_i = 1 \end{cases}$$

Uniform unbiasedness

: Whatever the value of the m we are guaranteed that the predictor is unbiased

$$(II) \hat{z}_{s_0} = \sum_i \lambda_i z_{s_i} \quad (\lambda_0 = 0)$$

$$\Phi(\hat{z}, \mu) = \frac{\mathbb{E}[(z_{s_0} - \hat{z})^2]}{(A)} + \frac{\mu}{(B)} \frac{\sum_i \lambda_i - 1}{(B)}$$

Unconditional error

unbiasedness constraint

Lagrangian
(μ Lagrange multiplier)

for convenience
(since it's a Lagrange multiplier (μ) we can do it)

$$\begin{aligned}
 & \text{under unbiasedness} \\
 (A) &= \text{Var}(z_{s_0} - \sum_i \lambda_i z_{s_i}) \\
 &= \text{Var}(z_{s_0}) + \text{Var}(\sum_i \lambda_i z_{s_i}) - 2 \text{Cov}(z_{s_0}, \sum_i \lambda_i z_{s_i}) \\
 &\approx C(s_0) + \sum_i \lambda_i \lambda_j \text{C}(z_{s_i}, z_{s_j}) - 2 \sum_i \lambda_i C(z_{s_0}, z_{s_i}) \\
 &= [C(s_0) + \sum_i \sum_j \lambda_i \lambda_j C(s_i, s_j) - 2 \sum_i \lambda_i C(s_0, s_i)]
 \end{aligned}$$

$$\begin{aligned}
 \text{Var}(z^* s) &= \lambda^T \Sigma \lambda \\
 \text{Cor}(\sum_i \lambda_i z_{s_i}, \sum_j \lambda_j z_{s_j}) &
 \end{aligned}$$

Expression of the variance of the prediction error under unbiasedness

$$\Phi(\lambda, \mu) = C(s_0) + \sum_i \lambda_i \lambda_j C(s_i, s_j) - 2 \sum_i \lambda_i C(s_0, s_i) + \mu (\sum_i \lambda_i - 1)$$

$$\begin{aligned}
 \min_{\lambda, \mu} \Phi(\lambda, \mu) &\rightarrow \left\{ \begin{array}{l} \frac{\partial \Phi}{\partial \lambda_i} = 0 \quad i=1 \dots n \\ \frac{\partial \Phi}{\partial \mu} = 0 \end{array} \right. \\
 &\text{because of the double sum} \\
 &\left\{ \begin{array}{l} \frac{\partial \Phi}{\partial \lambda_i} = 2 \sum_j \lambda_j C(s_i, s_j) - 2 C(s_0, s_i) + 2\mu = 0 \quad i=1 \dots n \\ \frac{\partial \Phi}{\partial \mu} = \sum_i \lambda_i - 1 = 0 \end{array} \right.
 \end{aligned}$$

Kriging system:

$$\begin{cases} \sum_j \lambda_j C(s_i, s_j) + \mu = C(s_0, s_i) & i=1 \dots n \\ \sum_i \lambda_i = 1 \end{cases}$$

matrix form

$$\begin{bmatrix} C(s_1, s_1) & C(s_1, s_2) & \dots & C(s_1, s_n) \\ C(s_2, s_1) & C(s_2, s_2) & \dots & C(s_2, s_n) \\ \vdots & \vdots & \ddots & \vdots \\ C(s_n, s_1) & C(s_n, s_2) & \dots & C(s_n, s_n) \end{bmatrix} \begin{bmatrix} 1 \\ \lambda_1 \\ \lambda_2 \\ \vdots \\ 1 \end{bmatrix} = \begin{bmatrix} C(s_0, s_1) \\ C(s_0, s_2) \\ \vdots \\ C(s_0, s_n) \\ 1 \end{bmatrix}$$

if we solve this we get our weights which guarantee the unbiasedness constraint

$$\Sigma = \text{Cov}(s) \quad \Sigma_{ij} = \text{Cov}(z_{s_i}, z_{s_j}) = C(s_i, s_j)$$

$$\begin{bmatrix} \Sigma & 1 & 1 & \dots & 1 \\ 1 & 0 & \mu & & 1 \end{bmatrix} = \begin{bmatrix} g_0 \\ \vdots \\ 1 \end{bmatrix} \quad (K1) \quad \begin{array}{l} \text{---} \\ \text{---} \\ \text{---} \end{array} \quad \begin{array}{l} \text{---} \\ \text{---} \\ \text{---} \end{array} \quad \begin{array}{l} \text{---} \\ \text{---} \\ \text{---} \end{array} \quad \begin{array}{l} \text{---} \\ \text{---} \\ \text{---} \end{array}$$

$$\text{Solve system (K1)} \rightarrow \hat{\lambda}^* \rightarrow z_{s_0}^* = \lambda^T \hat{\Sigma}$$

$$\text{Kriging Variance: } \sigma_{\text{OK}}^2(s_0) = \Phi(\hat{\lambda}^*, \mu^*) = C(s_0) - \sum_i \hat{\lambda}_i^* C(s_0, s_i) - \mu^*$$

Note: Assumption: C is known! But in fact we don't know it!
(we have to estimate it from the data)

- (1) Estimate $\hat{\gamma} \rightarrow \hat{C} \rightarrow \hat{\Sigma}$
- (2) Solve Kriging System (K1) with $\hat{\Sigma}$ in place of Σ

Plug-in Kriging predictor

$$\begin{aligned} ? \hat{\gamma} \rightarrow \hat{C} : \quad \hat{y}(h) &= C(0) - C(h) \\ &C(h) = C(0) - \hat{y}(h) \\ &\uparrow \\ &\lim_{h \rightarrow 0} \hat{y}(h) \end{aligned}$$



$$\begin{bmatrix} \sum_{i=1}^n \frac{1}{\lambda_i} \\ 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} g_0 \\ 1 \end{bmatrix} \quad (K1)$$

$$\sigma_{\text{OK}}^2 = \sigma^2_{\text{ORDINARY KRIGING}}$$

if we don't have the asymptote then we can't go from $\hat{y}(h)$ to $y(h)$
(the process in this case is not even 2nd order stationary)

D UNIVERSAL KRIGING (we are no more assuming STATIONARITY)

Model: $z_{s_i} \dots z_{s_n}$ obs from $\{s_i, s_{\text{ED}}\}$ non-stationary

$$\begin{aligned} z_g &= m_s + s_s, \quad s_{\text{ED}} \\ m_s &= E[z_s] \quad \text{drift} \\ s_s &= z_s - m_s \quad \text{residual} \end{aligned}$$

Assumptions: $m_s = \sum_{l=0}^L a_l f_l(s)$, a_l unknown coeffs.
 $f_l(s)$ regressors, known over D

- s_s : $\{s_i, s_{\text{ED}}\}$ random field \rightarrow 2nd order stationary random field for the residuals
- $E[s_s] = 0$
- $\text{Cor}(s_{s_i}, s_{s_j}) = C(s_i, s_j) = \text{Cor}(z_{s_i}, z_{s_j})$

$$\begin{aligned} \text{Kriging problem: } z_{s_0}^* &= \lambda_0 + \sum_i \lambda_i z_{s_i} \\ \text{Find } \lambda_0, \lambda_1, \dots, \lambda_L \text{ to solve problem (K)} \end{aligned}$$

- (1) Unbiasedness

$$E[z_{s_0}^*] = E[z_{s_0}]$$

$$E[z_{s_0}^*] = \lambda_0 + \sum_i \lambda_i m_{s_i} = \lambda_0 + \sum_i \lambda_i \sum_l a_l f_l(s_i)$$

$$E[z_{s_0}^*] = m_{s_0} = \sum_{l=0}^L a_l f_l(s_0)$$

$$\lambda_0 + \sum_l a_l \sum_i \lambda_i f_l(s_i) = \sum_l a_l f_l(s_0)$$

$$\begin{cases} \lambda_0 = 0 \\ \sum_l a_l f_l(s_i) = f_l(s_0) \quad l=0 \dots L \end{cases}$$

- (2) Objective function

$$\Phi(\lambda, \mu) = E[(z_{s_0} - z_{s_0}^*)^2] + 2 \sum_{l=0}^L \mu_l (\sum_i \lambda_i f_l(s_i) - f_l(s_0))$$

we have $L+1$ constraints
so we have to put $L+1$ lagrangian multipliers

same as before
(A)

(B')

$$\phi = C(0) + \sum_i \sum_j \lambda_i \lambda_j C(s_i - s_j) - 2 \sum_i \lambda_i C(s_0 - s_i) + 2 \sum_{L=0}^L \mu_L (\sum_i \lambda_i f_L(s_i) - f_L(s_0))$$

$$\min_{\lambda, \mu} \phi(\lambda, \mu) \Rightarrow \begin{cases} \frac{\partial \phi}{\partial \lambda_i} = 0 \\ \frac{\partial \phi}{\partial \mu_L} = 0 \end{cases}$$

$$\begin{cases} \frac{\partial \phi}{\partial \lambda_i} = \sum_j \lambda_j C(s_i - s_j) - \sum_i \mu_L f_L(s_i) + \sum_{L=0}^L \mu_L f_L(s_0) = 0 \\ \frac{\partial \phi}{\partial \mu_L} = \sum_i \lambda_i f_L(s_i) - f_L(s_0) = 0 \end{cases}$$

Universal Kriging system:

$$\begin{cases} \sum_j \lambda_j C(s_i - s_j) + \sum_L \mu_L f_L(s_i) = C(s_0 - s_i) \\ \sum_i \lambda_i f_L(s_i) = f_L(s_0) \end{cases}$$

Block-Matrix form:

$$\begin{bmatrix} \Sigma & F \\ F^T & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} g_0 \\ f_0 \end{bmatrix} \quad (K2)$$

Looking at the results we can conclude that: universal kriging is nothing but an expanded ordinary kriging where we allow for a variability in the drift term.

$$\begin{array}{l} \Sigma \text{ covariance matrix} \\ F \text{ design matrix} \\ Z = FZ + g \\ D \text{ matrix of zeros} \\ f_0 \text{ design vector} \\ f_0 = f_L(s_0) \end{array}$$

$$\begin{bmatrix} \Sigma & F \\ F^T & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} g_0 \\ f_0 \end{bmatrix} \quad (K2)$$

Solve system $\rightarrow Z^* \rightarrow Z_{s_0}^* = Z^* \underline{Z}$

Value of the L^{th} regressor at the new location

Note: Assumption that Σ is known

But we don't know the residuals at the beginning of the analysis!

- If Σ is unknown: we should estimate $\hat{\Sigma} \rightarrow \hat{C} \rightarrow \hat{\Sigma}$. We know how to estimate $\hat{\Sigma}$ from stationary data, but here we're assuming that data are not stationary. We estimate $\hat{\Sigma}$ based on the residuals (since they're from a 2nd stationary random field)
- If $\hat{s}_1, \dots, \hat{s}_n$ given
 - 1) Estimate $\hat{s}_1, \dots, \hat{s}_n$
 - 2) Estimate variogr. from $\hat{s}_1, \dots, \hat{s}_n$ and get $\hat{\gamma} \rightarrow \hat{C} \rightarrow \hat{\Sigma}$
- If $\hat{s}_1, \dots, \hat{s}_n$ unknown:
 - 1) Estimate $\hat{s}_1, \dots, \hat{s}_n$
 - 2) Estimate variogr. from $\hat{s}_1, \dots, \hat{s}_n$ and get $\hat{\gamma} \rightarrow \hat{C} \rightarrow \hat{\Sigma}$

Residuals estimation:

$$\begin{aligned} \hat{s}_{s_i} &= s_{s_i} - \hat{m}_{s_i} \\ \hat{m}_{s_i} &= \sum_l \hat{\alpha}_l f_l(s_i) \quad \rightarrow \hat{\alpha}^{OLS} = (F^T \Sigma^{-1} F)^{-1} F^T \Sigma^{-1} Z \end{aligned}$$

we can't proceed since we don't know Σ (which is the one we're trying to estimate)

\rightarrow Iterative algorithm

$$0. \text{ Initialize } \hat{\alpha} = \hat{\alpha}^{OLS} = (F^T F)^{-1} F^T Z \quad (\perp \Sigma)$$

$$1. \text{ Compute } \hat{s}_1, \dots, \hat{s}_n$$

$$2. \text{ Estimate } \hat{\gamma} \rightarrow \hat{C} \rightarrow \hat{\Sigma}$$

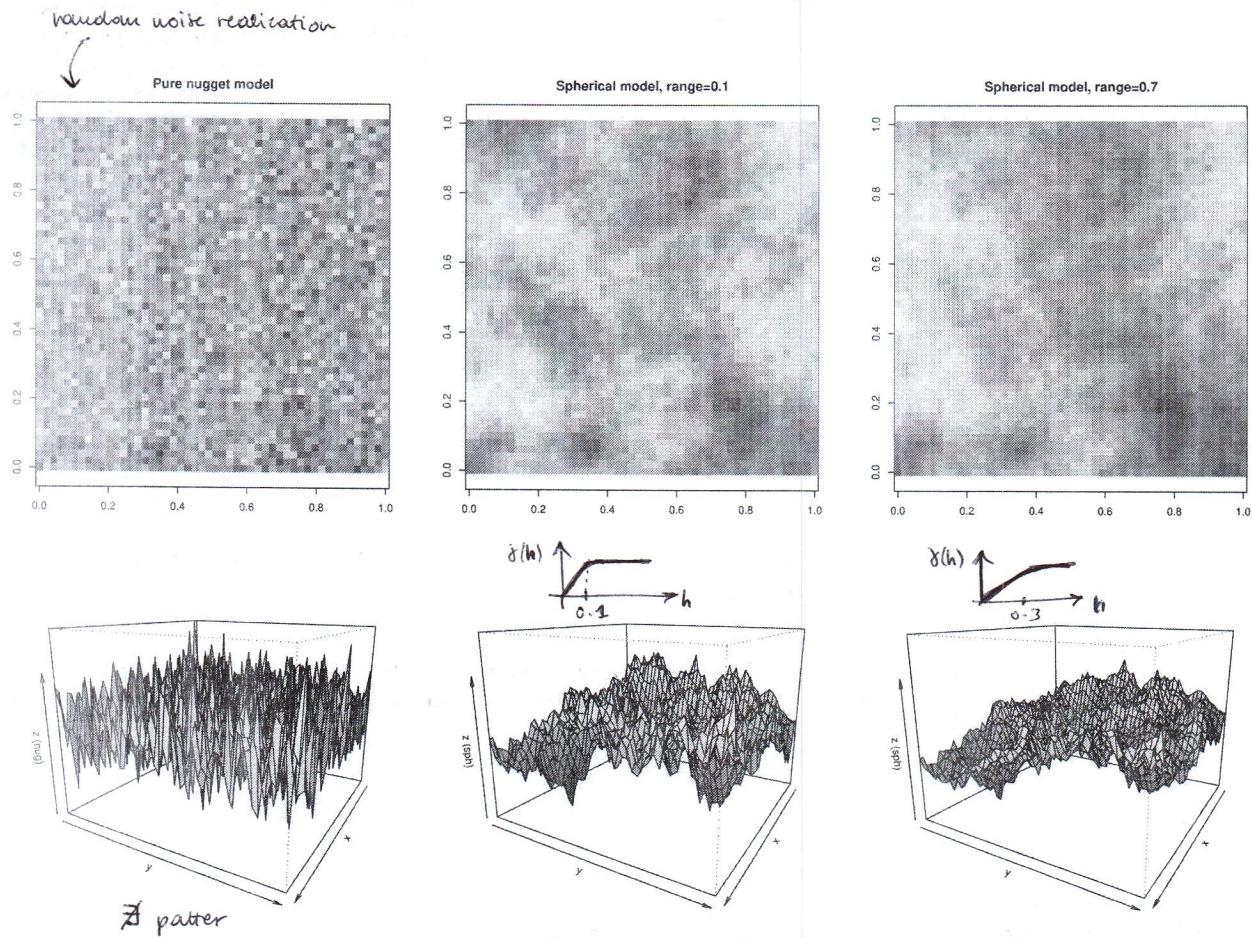
$$3. \text{ Update } \hat{\alpha} = \hat{\alpha}^{OLS} = (F^T \hat{\Sigma}^{-1} F)^{-1} F^T \hat{\Sigma}^{-1} Z$$

Until convergence

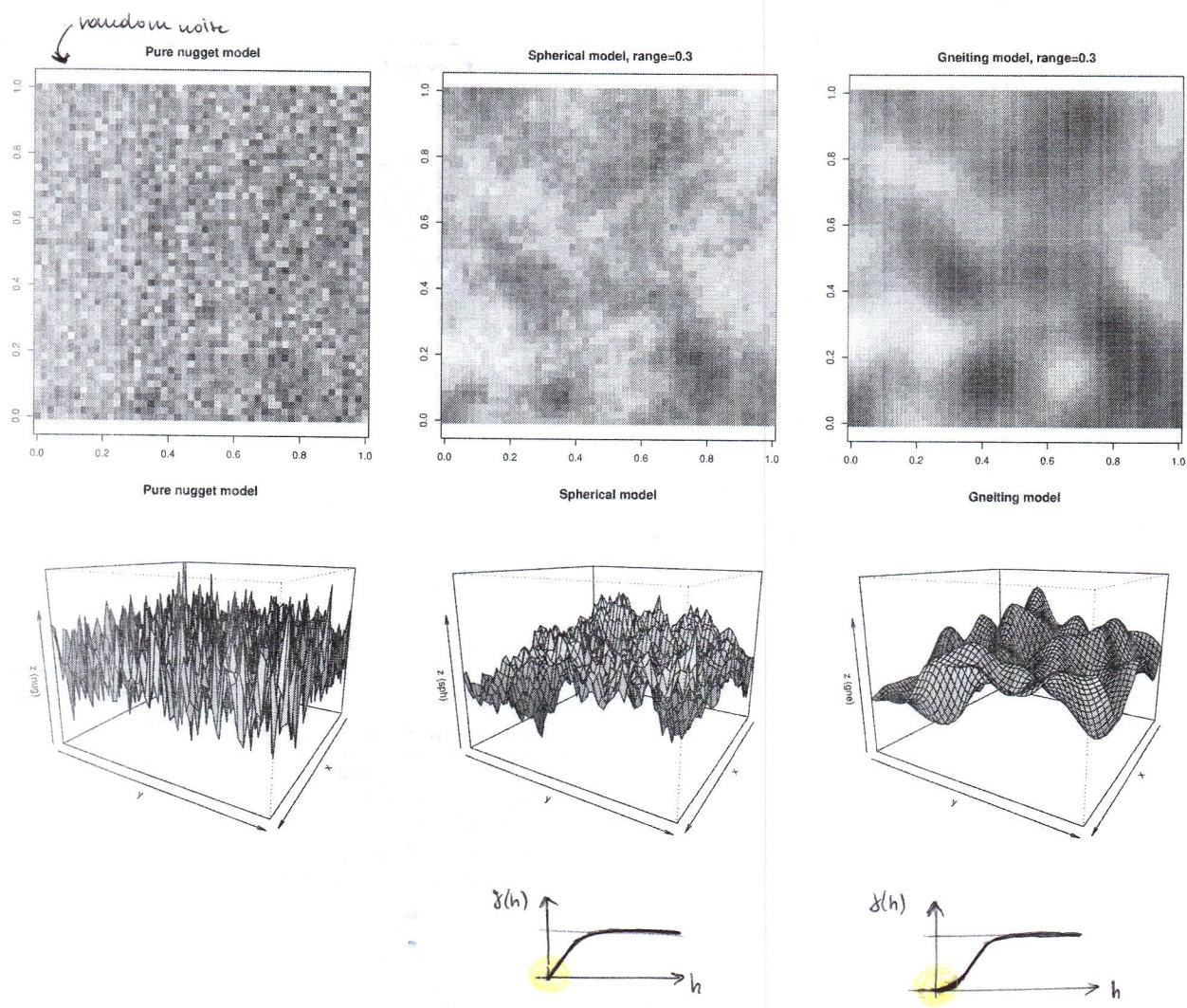
$\rightarrow \hat{\Sigma}$ plug-in in (K2) \rightarrow Plug-in Kriging predictor

Note: Here we're not reporting the Universal Kriging variance since it's based on Σ (and it considers Σ known) so, since Σ it's estimated, this variance is an underestimation of variability

DIFFERENCE IN THE RANGE : (both linear, no nugget)

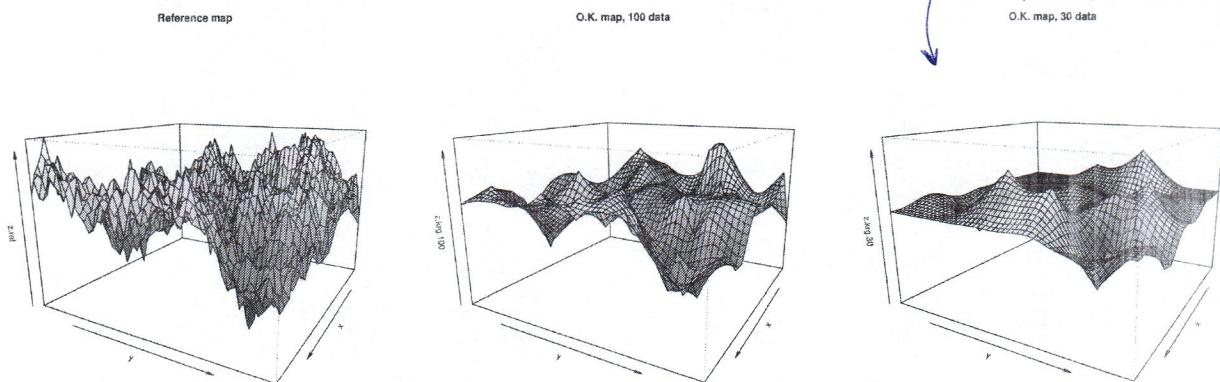


DIFFERENCE IN LINEAR/QUADRATIC : (both $R=0.3$, both no nugget)



the more zero-derivatives we have in 0 the more smooth the realizations

When we have less data it becomes something like "a tends"; it goes to the mean when we go far from the data (so if data were uncorrelated the best prediction that it could give is the mean)



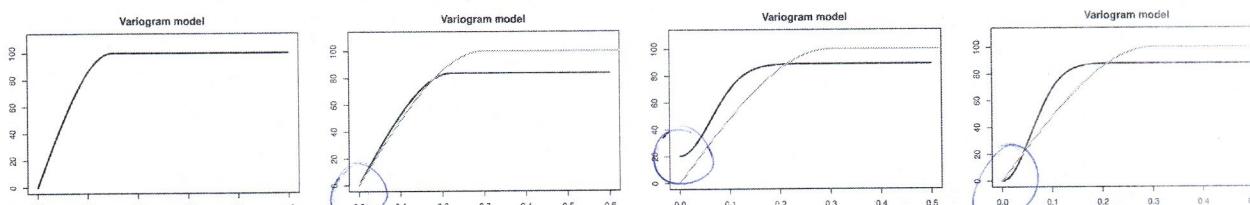
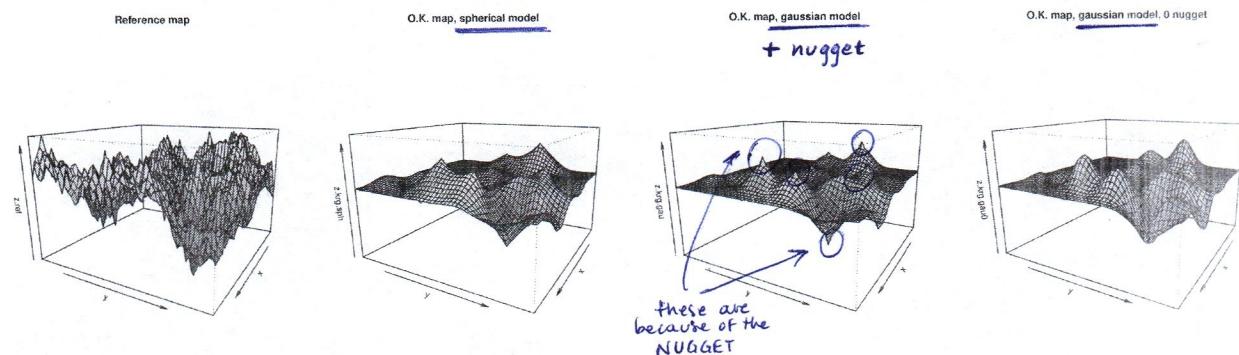
Real realization of the field

Kriging with 100 data
(kriging reconstruction)

Kriging with 30 data

these predictions are smoother than the real map
(that's because they're based on a spherical model, which is linear)

What also matters is the way in which we allow the prediction to go to the mean when we go far from the data:



We're interpolating the data but then, as soon as we go to zero ($h \rightarrow 0$: distance from the data) we have a jump \Rightarrow Kriging is interpolating but is discontinuous

We're assuming a quadratic / super-quadratic structure (for smoothness) near zero

