Chapter 4 - Fundamentals of spatial processes
Lecture notes

Odd Kolbjørnsen and Geir Storvik

January 30, 2017

Spatial processes

- Typically correlation between nearby sites
- Mostly positive correlation
- Negative correlation when competition

Part of a space-time process

- Temporal snapshot
- Temporal aggregation

Statistical analysis

- Incorporate spatial dependence into spatial statistical models
- Active research field
- Computer intensive tasks gives specialized software

Hierarchical (statistical) models

- Data model
- Process model

In time series setting - state space models

Hierarchical (statistical) models

- Data model
- Process model

In time series setting - state space models
Example

$$
\begin{aligned}
Y_t &= \alpha Y_{t-1} + W_t, t = 2, 3, \ldots \qquad W_t \overset{ind}{\sim} (0, \sigma_W^2) \qquad \text{Process model} \\
Z_t &= \beta Y_t + \eta_t, t = 1, 2, 3, \ldots \qquad \eta_t \overset{ind}{\sim} (0, \sigma_W^2) \qquad \text{Data model}
\end{aligned}
$$

We will do similar type of modelling now, separating the process model and the data model:

$$
Z(\mathbf{s}_i) = Y(\mathbf{s}_i) + \varepsilon(\mathbf{s}_i), \quad \varepsilon(\mathbf{s}_i) \sim \text{iid}
$$

Spatial prediction

An unknown function is of interest, i.e. $Y(\mathbf{s}),\ \mathbf{s} \in D$

Standard problem:

- Observed $\mathbf{Z} = (Z_1, ..., Z_m)$
- $Z_i = Y(\mathbf{s}_i) + \varepsilon_i, i = 1, ..., n$
- Want to predict function in an unobserved position $\mathbf{s}_0$, i.e. $Y(\mathbf{s}_0)$

Multiple methods, assumptions and framework are different.

- Interpolation, Regression, Splines
- Kriging
- Hierarchical model (with a Gaussian random field)

Methods for spatial prediction

- Regression: Find function of minimum least squares.

$$f(\mathbf{s}) = \sum_{i=1}^{K} \beta_i f_i(\mathbf{s}) = \beta^T \mathbf{f}(\mathbf{s})$$

Know: $\beta = (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T\mathbf{Z}$, with $F_{ij} = f_j(s_i)$.

- Kriging: Find optimal unbiased linear predictor under Squared error loss.

$$\min_{k,\mathbf{L}} E\{(Y(\mathbf{s}_0) - k - \mathbf{L}^T\mathbf{Z})^2\}$$

- Hierarchical model: Find optimal predictor under Squared error loss.

$$\min_{a(\mathbf{Z})} E\{(Y(\mathbf{s}_0) - a(\mathbf{Z}))^2\}$$

Know: $a(\mathbf{Z}) = E\{Y(\mathbf{s}_0)|\mathbf{Z}\}$

Geostatistical models

Assume $\{Y(\mathbf{s})\}$ is a *Gaussian process*

- $(Y(\mathbf{s}_1), ..., Y(\mathbf{s}_m))$ is multivariate Gaussian for all m and $\mathbf{s}_1, ..., \mathbf{s}_m$

Need to specify

- $\mu(\mathbf{s}) = E(Y(\mathbf{s}))$
- $C_Y(\mathbf{s}, \mathbf{s}') = \text{cov}(Y(\mathbf{s}), Y(\mathbf{s}'))$

Assuming 2. order stationarity:

- $\mu(\mathbf{s}) = \mu, \quad \forall \mathbf{s}$
- $C_Y(\mathbf{s}, \mathbf{s}') = C_Y(\mathbf{s} - \mathbf{s}'), \quad \forall \mathbf{s}, \mathbf{s}'$

Common extension:

- $\mu(\mathbf{s}) = \mathbf{x}(s)^T \boldsymbol{\beta}$

Often:

$$Z(\mathbf{s}_i)|Y(\mathbf{s}_i), \sigma_\varepsilon^2 \sim \text{independent N}(Y(\mathbf{s}_i), \sigma_\varepsilon^2)$$

Covariance function and Variogram

Dependence can be specified through covariance functions or the *Variogram*

$$
\begin{aligned}
2\gamma_Y(\mathbf{h}) \equiv& \text{var}[Y(\mathbf{s}+\mathbf{h}) - Y(\mathbf{s})] \\
=& \text{var}[Y(\mathbf{s}+\mathbf{h})] + \text{var}[Y(\mathbf{s})] - 2\text{cov}[Y(\mathbf{s}+\mathbf{h}), Y(\mathbf{s})] \\
=& 2C_Y(\mathbf{0}) - 2C_Y(\mathbf{h})
\end{aligned}
$$

- Variograms are more general than covariance functions
- Variogram can exist even if $\text{var}[Y(\mathbf{s})] = \infty$!
- In variograms relative changes are modelled model rather than the process it self.
- In Geostatistics it is common to use variograms.
- All formulas "Covariance formulas" have corresponding "Variogram formulas".
- $\gamma_Y(\mathbf{h})$ sometimes called semi-variogram

Stationary, isotropic, anisotropic

Strong stationarity For any $(\mathbf{s}_1, ..., \mathbf{s}_m)$ and any $\mathbf{h}$

$$[Y(\mathbf{s}_1), ..., Y(\mathbf{s}_m)] = [Y(\mathbf{s}_1 + \mathbf{h}), ..., Y(\mathbf{s}_m + \mathbf{h})]$$

Stationarity in mean

$$\mathrm{E}(Y(\mathbf{s})) = \mu, \text{ for all } \mathbf{s} \in D_A$$

Stationarity covariance (Depend only on lag)

$$\mathrm{cov}(Y(\mathbf{s}), Y(\mathbf{s} + \mathbf{h})) = C_Y(\mathbf{h}), \text{ for all } \mathbf{s}, \mathbf{s} + \mathbf{h} \in D_A$$

Isotropic covariance (depend only on length of lag)

$$\mathrm{cov}(Y(\mathbf{s}), Y(\mathbf{s} + \mathbf{h})) = C_Y(\|\mathbf{h}\|) \text{ for all } \mathbf{s}, \mathbf{s} + \mathbf{h} \in D_A$$

Geometric anisotropy in covariance ("Rotated coordinate system")

$$\mathrm{cov}(Y(\mathbf{s}), Y(\mathbf{s} + \mathbf{h})) = C_Y(\|\mathbf{Ah}\|), \text{ for all } \mathbf{s}, \mathbf{s} + \mathbf{h} \in D_A$$

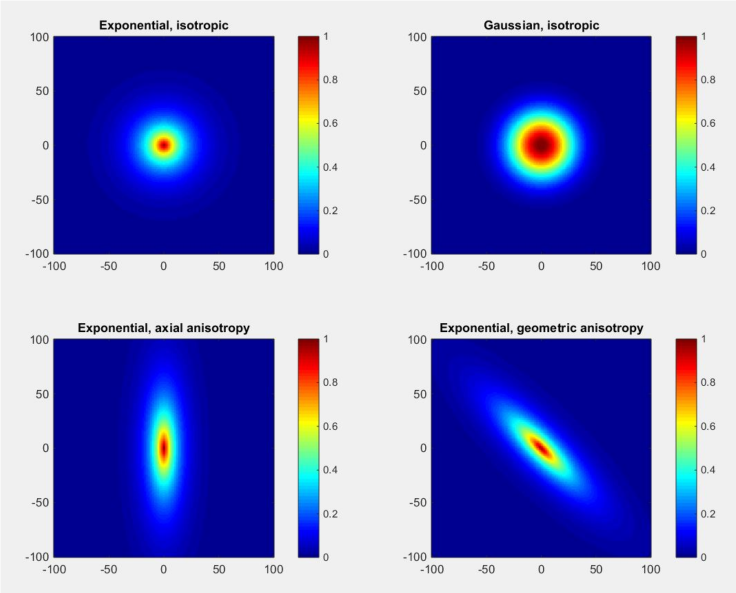
Weak stationarity : Stationarity in mean and a stationarity covariance.
(recall time series)

Isotropic covariance functions/variograms

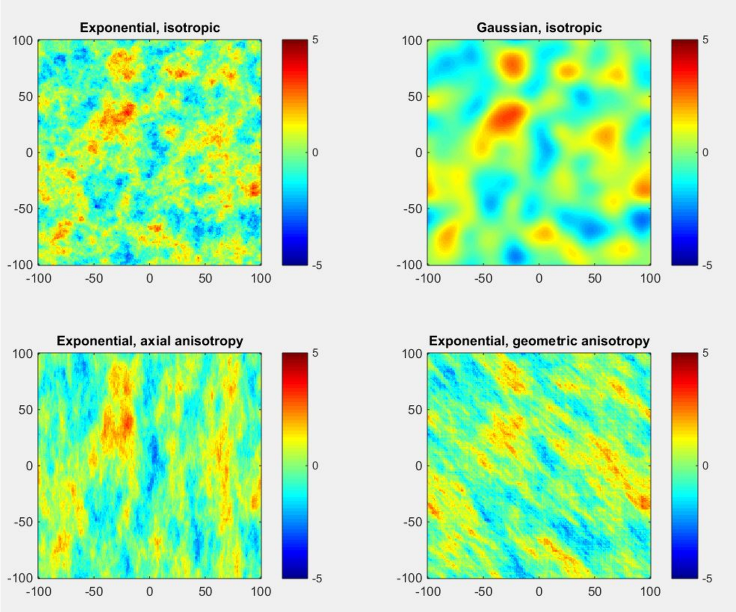
- Matern covariance function
$$C_Y(\mathbf{h};\boldsymbol{\theta}) = \sigma_1^2\{2^{\theta_2-1}\Gamma(\theta_2)\}^{-1}\{||\mathbf{h}||/\theta_1\}^{\theta_2}K_{\theta_2}(||\mathbf{h}||/\theta_1)$$
- Powered-exponential
$$C_Y(\mathbf{h};\boldsymbol{\theta}) = \sigma_1^2\exp\{-(||\mathbf{h}||/\theta_1)^{\theta_2}\}$$
- Exponential
$$C_Y(\mathbf{h};\boldsymbol{\theta}) = \sigma_1^2\exp\{-(||\mathbf{h}||/\theta_1)\}$$
- Gaussian
$$C_Y(\mathbf{h};\boldsymbol{\theta}) = \sigma_1^2\exp\{-(||\mathbf{h}||/\theta_1)^2\}$$

Note:
Many different ways to define the "Range parameter" θ_1.
e.g. $\sigma_1^2\exp\{-3(||\mathbf{h}||/R)^{\nu}\}$

Stationary spatial covariance functions

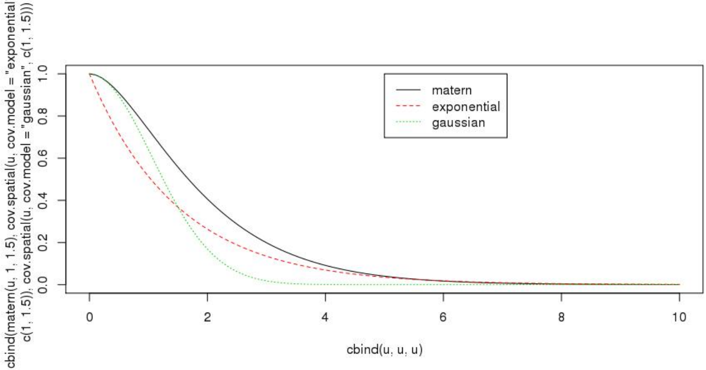


Simulation of stationary spatial random field



Isotropic Covariance

Benefit of isotropic assumption is that we only need 1D functions



Note: not all covariance functions / variograms valid in 1D are valid as isotropic covariance functions in higher dimensions.

Bochner's theorem

A covariance function needs to be positive definite

Theorem (Bochner, 1955)

If $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |C_Y(\mathbf{h})| d\mathbf{h} < \infty$, *then a valid real-valued covariance function can be written as*

$$C_Y(\mathbf{h}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \cos(\boldsymbol{\omega}^T \mathbf{h}) f_Y(\boldsymbol{\omega}) d\boldsymbol{\omega}$$

where $f_Y(\boldsymbol{\omega}) \geq 0$ *is symmetric about* $\boldsymbol{\omega} = 0$.

$f_Y(\boldsymbol{\omega})$: Spectral density of $C_Y(\mathbf{h})$.

Nugget effect and Sill

We have

$$\begin{aligned} C_Z(\boldsymbol{h}) =& \mathrm{cov}[Z(\mathbf{s}), Z(\mathbf{s}+\mathbf{h})] \\ =& \mathrm{cov}[Y(\mathbf{s}) + \varepsilon(\mathbf{s}), Y(\mathbf{s}+\mathbf{h}) + \varepsilon(\mathbf{s}+\mathbf{h})] \\ =& \mathrm{cov}[Y(\mathbf{s}), Y(\mathbf{s}+\mathbf{h})] + \mathrm{cov}[\varepsilon(\mathbf{s}), \varepsilon(\mathbf{s}+\mathbf{h})] \\ =& C_Y(\mathbf{h}) + \sigma_\varepsilon^2 I(\mathbf{h} = \mathbf{0}) \end{aligned}$$

Assume

$$\begin{aligned} C_Y(\mathbf{0}) =& \sigma_Y^2 \\ \lim_{\mathbf{h}\to 0}[C_Y(\mathbf{0}) - C_Y(\mathbf{h})] =& 0 \end{aligned}$$

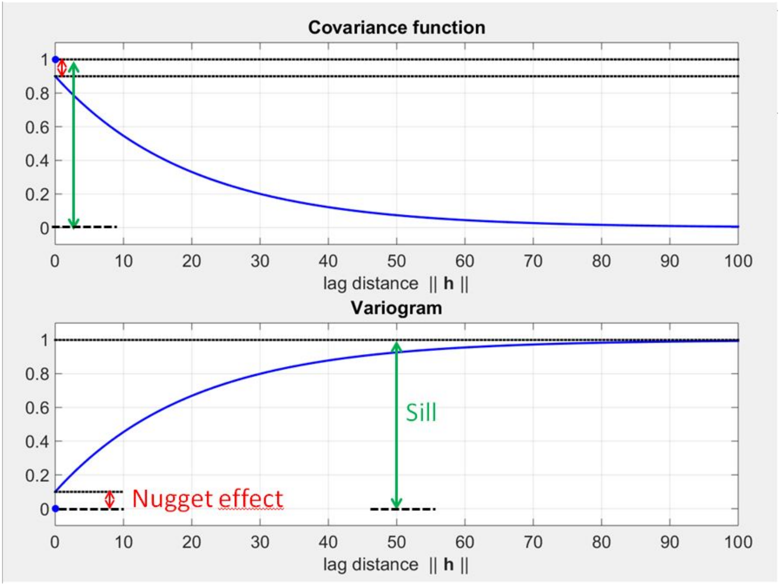
Then

$$C_Z(\mathbf{0}) = \sigma_Y^2 + \sigma_\varepsilon^2 \qquad \text{Sill}$$

$$\lim_{\mathbf{h}\to 0}[C_Z(\mathbf{0}) - C_Z(\mathbf{h})] = \sigma_\varepsilon^2 = c_0 \qquad \text{Nugget effect}$$

Possible to include nugget effect also in Y-process.

Nugget/sill



Covariance function
1
0.8
0.6
0.4
0.2
0
0
10
20
30
40
50
60
70
80
90
100
lag distance || h ||
Variogram
Sill
Nugget effect
lag distance || h ||

Nugget/sill

- Sill is a fixed finite level which the variogram converges towards at large lag (i.e. $\|\mathbf{h}\| \to \infty$).
- Variograms without a sill ($\gamma(\mathbf{h}) \to \infty$ as $\|\mathbf{h}\| \to \infty$) has no equivalent correlation function.
- The nugget effect is variability below the resolution in our model.
- In the setting with point observations, i.e. $Z_i = Y(\mathbf{s}_i) + \varepsilon_i$ it is difficult (often impossible) to distinguish nugget effect in random field $Y(\mathbf{s}_i)$ and observation error ε_i. This becomes a modelling choice.
- Some softwares mixes nugget effect and observation error.

Estimation of variogram/covariance function

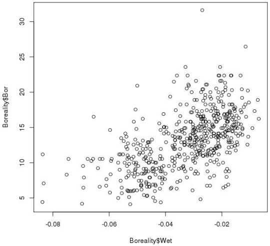
$$2\gamma_Z(\mathbf{h}) \equiv \text{var}[Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})]$$
$$\overset{\text{Const expectation}}{=} E[Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})]^2$$

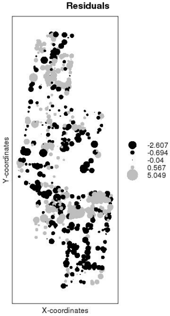
Can estimate from "all" pairs having distance **h** between.
Problem: Few/no pairs for all **h**
Simplifications

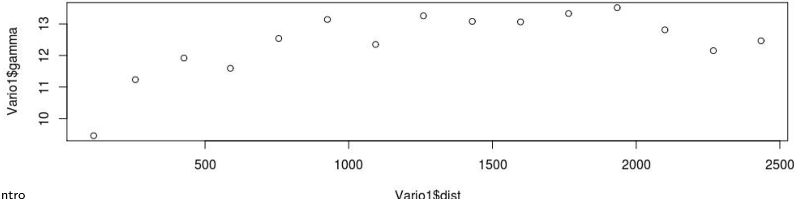
- Isotropic: $\gamma_Z(\mathbf{h}) = \gamma_Z^0(||\mathbf{h}||)$
- Lag bin: $2\widehat{\gamma}_Z^0(h) = \text{ave}\{(Z(\mathbf{s}_i) - Z(\mathbf{s}_j))^2; ||\mathbf{s}_i - \mathbf{s}_j|| \in T(h)\}$
- If covariates, use residuals

Boreality data

Empirical variogram: Boreality data

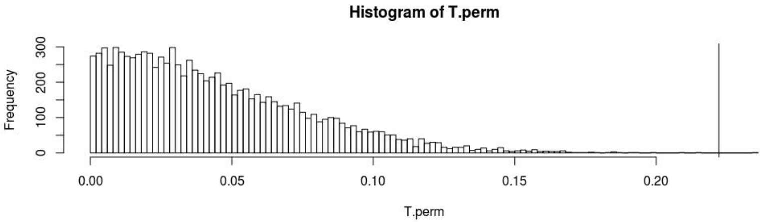






Testing for independence

If independence: $\gamma_Z^0(h) = \sigma_Z^2$
Test-statistic $F = \widehat{\gamma}_Z(h_1)/\widehat{\sigma}_Z^2$, h_1 smallest observed distance
Reject H_0 for $|F - 1|$ large
Permutation test: Keep spatial positions, and keep values of residuals, but scramble the pairing, i.e. reassign residuals to new spatial positions.
Recalculate F for all permutations of $\mathbf{Z}$ (or for a random sample of permutations)
If observed F is above 97.5% percentile, reject H_0
Boreality example:



P-value = 0.0001

Prediction in multivariate Gaussian models

$(Z(\mathbf{s}_1), ..., Z(\mathbf{s}_m), Y(\mathbf{s}_0))$ is multivariate Gaussian
Can use rules about conditional distributions:

$$\begin{pmatrix}\mathbf{X}_1\\ \mathbf{X}_2\end{pmatrix} \sim MVN\left(\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22}\end{pmatrix}\right)$$

$$E(\mathbf{X}_1|\mathbf{X}_2) = \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\mu}_2)$$

$$\text{var}(\mathbf{X}_1|\mathbf{X}_2) = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}$$

Need

- Expectations: As for ordinary linear regression
- Covariances: New!

Prediction in the spatial model

$$\begin{pmatrix} Y(\mathbf{s}_0) \\ \mathbf{Z} \end{pmatrix} \sim MVN\left(\begin{pmatrix} \mu(\mathbf{s}_0) \\ \boldsymbol{\mu}_\mathbf{Z} \end{pmatrix}, \begin{pmatrix} C_Y(\mathbf{s}_0,\mathbf{s}_0) & \mathbf{c}(\mathbf{s}_0) \\ \mathbf{c}(\mathbf{s}_0)^T & \boldsymbol{C}_\mathbf{Z} \end{pmatrix} \right)$$

$$\begin{aligned}
\mathbf{c}(\mathbf{s}_0) =& \text{cov}[\mathbf{Z}, Y(\mathbf{s}_0)] \\
=& \text{cov}[\mathbf{Y}, Y(\mathbf{s}_0)] \\
=& (C_Y(\mathbf{s}_0,\mathbf{s}_1), ..., C_Y(\mathbf{s}_0,\mathbf{s}_m)) \\
=& \mathbf{c}_Y(\mathbf{s}_0) \\
\mathbf{C}_Z =& \{C_Z(\mathbf{s}_i,\mathbf{s}_j)\} \\
C_Z(\mathbf{s}_i,\mathbf{s}_j) =& \begin{cases} C_Y(\mathbf{s}_i,\mathbf{s}_i) + \sigma_\varepsilon^2, & \mathbf{s}_i = \mathbf{s}_j \\ C_Y(\mathbf{s}_i,\mathbf{s}_j), & \mathbf{s}_i \neq \mathbf{s}_j \end{cases}
\end{aligned}$$

$$\text{E}(Y(\boldsymbol{s}_0)|\mathbf{Z}) = \mu(\boldsymbol{s}_0) + \mathbf{c}(\mathbf{s}_0)^T \boldsymbol{C}_\mathbf{Z}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu}_{\boldsymbol{Z}})$$

$$\text{Var}(Y(\boldsymbol{s}_0)|\mathbf{Z}) = C_Y(\boldsymbol{s}_0,\boldsymbol{s}_0) - \mathbf{c}(\mathbf{s}_0)^T \boldsymbol{C}_\mathbf{Z}^{-1}\mathbf{c}(\mathbf{s}_0)$$

Kriging = Prediction

Model

$$
\begin{aligned}
Y(\mathbf{s}) =& \mathbf{x}(\mathbf{s})^T \boldsymbol{\beta} + \boldsymbol{\delta}(\mathbf{s}) \\
Z_i =& Y(\mathbf{s}_i) + \varepsilon_i
\end{aligned}
$$

Prediction of $Y(\mathbf{s}_0)$, i.e. in an unobserved location.
Linear predictors $\{\mathbf{L}^T\mathbf{Z} + k\}$
Optimal predictor minimize

$$
\begin{aligned}
\text{MSPE}(\mathbf{L}, k) \equiv& E[Y(\mathbf{s}_0) - \mathbf{L}^T\mathbf{Z} - k]^2 \\
=& \text{var}[Y(\mathbf{s}_0) - \mathbf{L}^T\mathbf{Z} - k] + \{E[Y(\mathbf{s}_0) - \mathbf{L}^T\mathbf{Z} - k]\}^2
\end{aligned}
$$

Note: Do not make any distributional assumptions

Kriging

$$\begin{aligned}\text{MSPE}(\mathbf{L}, k) =& \text{var}[Y(\mathbf{s}_0) - \mathbf{L}^T\mathbf{Z} - k] + \{E[Y(\mathbf{s}_0) - \mathbf{L}^T\mathbf{Z} - k]\}^2 \\ =& \text{var}[Y(\mathbf{s}_0) - \mathbf{L}^T\mathbf{Z} - k] + \{\mu_Y(\mathbf{s}_0) - \mathbf{L}^T\boldsymbol{\mu}_z - k]\}^2\end{aligned}$$

Second term is zero if $k = \mu_Y(\mathbf{s}_0) - \mathbf{L}^T\boldsymbol{\mu}_Z$.
First term ($\mathbf{c}(\mathbf{s}_0) = \text{cov}[\mathbf{Z}, Y(\mathbf{s}_0)]$):

$$\text{var}[Y(\mathbf{s}_0) - \mathbf{L}^T\mathbf{Z} - k] = C_Y(\mathbf{s}_0, \mathbf{s}_0) - 2\mathbf{L}^T\mathbf{c}(\mathbf{s}_0) + \mathbf{L}^T\mathbf{C}_Z\mathbf{L}$$

Derivative wrt $\mathbf{L}^T$:

$$\begin{aligned}&-2\mathbf{c}(\mathbf{s}_0) + 2\mathbf{C}_Z\mathbf{L} = \mathbf{0} \\ &\mathbf{L}^* = \mathbf{C}_Z^{-1}\mathbf{c}(\mathbf{s}_0)\end{aligned}$$

giving

$$\begin{aligned}Y^*(\mathbf{s}_0) =& \mu_Y(\mathbf{s}_0) + c(\mathbf{s}_0)^T\mathbf{C}_Z^{-1}[\mathbf{Z} - \boldsymbol{\mu}_Z] \\ \text{MSPE}(\mathbf{L}^*, k^*) =& C_Y(\mathbf{s}_0, \mathbf{s}_0) - \mathbf{c}(\mathbf{s}_0)^T\mathbf{C}_Z^{-1}\mathbf{c}(\mathbf{s}_0)\end{aligned}$$

Kriging (Simple Kriging)

The Kriging predictions with a known mean is:

$$\begin{aligned}
\mathbf{c}(\mathbf{s}_0) &= \text{cov}[\mathbf{Z}, Y(\mathbf{s}_0)] \\
&= \text{cov}[\mathbf{Y}, Y(\mathbf{s}_0)] \\
&= (C_Y(\mathbf{s}_0, \mathbf{s}_1), ..., C_Y(\mathbf{s}_0, \mathbf{s}_m)) \\
&= \mathbf{c}_Y(\mathbf{s}_0) \\
\mathbf{C}_Z &= \{C_Z(\mathbf{s}_i, \mathbf{s}_j)\} \\
C_Z(\mathbf{s}_i, \mathbf{s}_j) &= \begin{cases} C_Y(\mathbf{s}_i, \mathbf{s}_i) + \sigma_\varepsilon^2, & \mathbf{s}_i = \mathbf{s}_j \\ C_Y(\mathbf{s}_i, \mathbf{s}_j), & \mathbf{s}_i \neq \mathbf{s}_j \end{cases}
\end{aligned}$$

$$Y^*(\mathbf{s}_0) = \mu_Y(\mathbf{s}_0) + \mathbf{c}_Y(\mathbf{s}_0)^T \mathbf{C}_Z^{-1}(\mathbf{Z} - \boldsymbol{\mu}_Z)$$

Gaussian assumptions

Recall now $\mathbf{Y}, \mathbf{Z}$ are MVN

$$\begin{pmatrix}\mathbf{Z}\\ \mathbf{Y}(\mathbf{s}_0)\end{pmatrix} = \text{MVN}\left(\begin{pmatrix}\boldsymbol{\mu}_Z\\ \boldsymbol{\mu}_Y(\mathbf{s}_0)\end{pmatrix}, \begin{pmatrix}\mathbf{C}_Z & \mathbf{c}_Y(\mathbf{s}_0)^T\\ \mathbf{c}_Y(\mathbf{s}_0) & C_Y(\mathbf{s}_0,\mathbf{s}_0)\end{pmatrix}\right)$$

Give

$$Y(\mathbf{s}_0)|\mathbf{Z} \sim N(\mu_Y(\mathbf{s}_0) + c(\mathbf{s}_0)^T\mathbf{C}_Z^{-1}[\mathbf{Z}-\boldsymbol{\mu}_Z], C_Y(\mathbf{s}_0,\mathbf{s}_0) - \mathbf{c}(\mathbf{s}_0)^T\mathbf{C}_Z^{-1}\mathbf{c}(\mathbf{s}_0))$$

Same as kriging!
The book derive this directly without using the formula for conditional distribution

Compare to interpolation

- Assume no observation error
- Consider deviations from $\mu(\mathbf{s})$,i.e. $f(\mathbf{s}) = Y(\mathbf{s}) - \mu(\mathbf{s})$.
- Set $f_i(\mathbf{s}) = C_Y(\mathbf{s} - \mathbf{s}_i)$

Then:

- $\mathbf{F} = \mathbf{C_Z}$ ($= \mathbf{C_Y}$ no observation error)
- $(\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T = \mathbf{C_Z}^{-1}$ (symmetry)
- $\beta = \mathbf{C_Z}^{-1}\mathbf{Z}$
- $f(\mathbf{s}_0) = \sum_{i=1}^n \beta_i C_Y(\mathbf{s}_0 - \mathbf{s}_i)$

Which gives the same result again:

$$f(\mathbf{s}_0) = \beta^T\mathbf{c}(\mathbf{s}_0) = (\mathbf{Z}^T)\mathbf{C_Z}^{-1}\mathbf{c}(\mathbf{s}_0)$$

$Y^*(\mathbf{s}_0) = \mu(\mathbf{s}_0) + c(\mathbf{s}_0)^T\mathbf{C_Z}^{-1}(\mathbf{Z} - \boldsymbol{\mu}_\mathbf{Z})$ If we include $\mu(\mathbf{s})$ we get

Spatial prediction comments

- Given the mean and covariance function Kriging and the a Gaussian random field model give identical spatial correlations.
- There are stronger assumptions underlying the Gaussian model than needed in Kriging
- Kriging is the optimal linear predictor for any distribution (not only gaussian)
- For the Gaussian model the linear prediction is optimal (among all), for other distributions there might be other predictors which are better.
- The Gaussian model is easier to extend using a Hierarchical approach
- In Kriging and Gaussian random field, "interpolating functions" are determined by data
- In Kriging it is common to use plugin estimate for the covariance function
- The hierarchical approach is suited to include uncertainty on model parametrers.
- the Mean Squared Prediction Error (MSPE) is used for both Kriging and hierarchical model (HM)

So fa and Next:

- Simple kriging
  - Linear predictor
  - Assume parameters known
  - Equal to conditional expectation

So fa and Next:

- Simple kriging
    - Linear predictor
    - Assume parameters known
    - Equal to conditional expectation
- Unknown parameters
    - Ordinary kriging
    - Plug-in estimates
    - Bayesian approach
- Non-Gaussian models

Unknown parameters

So far assumed parameters known, what if unknown?

- Direct approach -Universal and Ordinary kriging (for mean parameters)
- Plug-in estimate/Empirical Bayes
- Bayesian approach

Kriging (cont)

$$Y^*(\mathbf{s}_0) = \mu_Y(\mathbf{s}_0) + \mathbf{c}_Y(\mathbf{s}_0)^T \mathbf{C}_Z^{-1}(\mathbf{Z} - \boldsymbol{\mu}_Z)$$

Assuming

$$\begin{aligned} E[Y(\mathbf{s})] =& \mathbf{x}(\mathbf{s})^T \boldsymbol{\beta} \\ Z(\mathbf{s}_i)|Y(\mathbf{s}_i), \sigma_\varepsilon^2 \sim& \text{ind.Gau}(Y(\mathbf{s}_i), \sigma_\varepsilon^2) \end{aligned}$$

Then

$$\begin{aligned} \boldsymbol{\mu}_Z =& \boldsymbol{\mu}_Y = \mathbf{X}\boldsymbol{\beta} \\ \mathbf{C}_Z =& \boldsymbol{\Sigma}_Y + \sigma_\varepsilon^2 \mathbf{I} \\ \mathbf{c}_Y(\mathbf{s}_0) =& (C_Y(\mathbf{s}_0, \mathbf{s}_1), ..., C_Y(\mathbf{s}_0, \mathbf{s}_m))^T \end{aligned}$$

and

$$Y^*(\mathbf{s}_0) = \mathbf{x}(\mathbf{s}_0)\boldsymbol{\beta} + \mathbf{c}_Y(\mathbf{s}_0)^T [\boldsymbol{\Sigma}_Y + \sigma_\varepsilon^2 \mathbf{I}]^{-1}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})$$

Kriging

Simple kriging, $EY(\mathbf{s}) = \mathbf{x}(\mathbf{s})^T\beta$, known

$$Y^*(\mathbf{s}_0) = \mathbf{x}(\mathbf{s}_0)\beta + \mathbf{c}_Y(\mathbf{s}_0)^T\mathbf{C}_Z^{-1}(\mathbf{Z} - \mathbf{X}\beta)$$

Kriging

Simple kriging, $EY(\mathbf{s}) = \mathbf{x}(\mathbf{s})^T\boldsymbol{\beta}$, known

$$Y^*(\mathbf{s}_0) = \mathbf{x}(\mathbf{s}_0)\boldsymbol{\beta} + \mathbf{c}_Y(\mathbf{s}_0)^T\mathbf{C}_Z^{-1}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})$$

Universal kriging, $EY(\mathbf{s}) = \mathbf{x}(\mathbf{s})^T\boldsymbol{\beta}$, unknown

$$\begin{aligned}\widehat{Y}(\mathbf{s}_0) =& \mathbf{x}(\mathbf{s}_0)^T\widehat{\boldsymbol{\beta}}_{gls} + \mathbf{c}_Y(\mathbf{s}_0)^T\mathbf{C}_Z^{-1}(\mathbf{Z} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{gls})\\ \widehat{\boldsymbol{\beta}}_{gls} =& [\mathbf{X}^T\mathbf{C}_Z^{-1}\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{C}_z^{-1}\mathbf{Z}\end{aligned}$$

Kriging

Simple kriging, $EY(\mathbf{s}) = \mathbf{x}(\mathbf{s})^T\boldsymbol{\beta}$, known

$$Y^*(\mathbf{s}_0) = \mathbf{x}(\mathbf{s}_0)\boldsymbol{\beta} + \mathbf{c}_Y(\mathbf{s}_0)^T\mathbf{C}_Z^{-1}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})$$

Universal kriging, $EY(\mathbf{s}) = \mathbf{x}(\mathbf{s})^T\boldsymbol{\beta}$, unknown

$$\begin{aligned}\widehat{Y}(\mathbf{s}_0) =& \mathbf{x}(\mathbf{s}_0)^T\widehat{\boldsymbol{\beta}}_{gls} + \mathbf{c}_Y(\mathbf{s}_0)^T\mathbf{C}_Z^{-1}(\mathbf{Z} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{gls})\\ \widehat{\boldsymbol{\beta}}_{gls} =& [\mathbf{X}^T\mathbf{C}_Z^{-1}\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{C}_z^{-1}\mathbf{Z}\end{aligned}$$

Bayesian kriging, $EY(\mathbf{s}) = \mathbf{x}(\mathbf{s})^T\boldsymbol{\beta}$, with $\beta \sim N(\beta_0, \boldsymbol{\Sigma}_0)$

$$\begin{aligned}\widehat{Y}(\mathbf{s}_0) =& \mathbf{x}(\mathbf{s}_0)^T\widehat{\boldsymbol{\beta}}_B + \mathbf{c}_Y(\mathbf{s}_0)^T\mathbf{C}_Z^{-1}(\mathbf{Z} - \mathbf{X}\widehat{\boldsymbol{\beta}}_B)\\ \widehat{\boldsymbol{\beta}}_B =& \beta_0 + \boldsymbol{\Sigma}_0\mathbf{X}^T(\mathbf{X}\boldsymbol{\Sigma}_0\mathbf{X}^T + \mathbf{C}_Z)^{-1}(\mathbf{Z} - \mathbf{X}\beta_0)\end{aligned}$$

Possible to show the SK and UK are the limiting cases of BK
$\Sigma_0 \to 0 \Rightarrow BK \to SK$
$\Sigma_0 \to \infty \Rightarrow BK \to UK$

Ordinary kriging

Ordinary kriging, $EY(\mathbf{s}) = \mu$, unknown (special case of UK)

$$\begin{aligned}\widehat{Y}(\mathbf{s}_0) =& \{\mathbf{c}_Y(\mathbf{s}_0) + \frac{\mathbf{1}(1 - \mathbf{1}^T\mathbf{C}_Z{-}1\mathbf{c}_y(\mathbf{s}_0))}{\mathbf{1}^T\mathbf{C}_Z^{-1}\mathbf{1}}\}^T\mathbf{C}_Z^{-1}\mathbf{Z} \\ =& \widehat{\mu}_{gls} + \mathbf{c}_Y(\mathbf{s}_0)^T\mathbf{C}_Z^{-1}(\mathbf{Z} - \mathbf{1}\widehat{\mu}_{gls}) \\ \widehat{\mu}_{gls} =& [\mathbf{1}^T\mathbf{C}_Z^{-1}\mathbf{1}]^{-1}\mathbf{1}^T\mathbf{C}_z^{-1}\mathbf{Z}\end{aligned}$$

To minimize the MSPE, we make an unbiased estimate with the minimum variance.

$$\text{MSPE}(\boldsymbol{\lambda}) = E(Y(\mathbf{s}_0) - \boldsymbol{\lambda}^T\mathbf{Z})^2$$

Unbiased constraint:

$$E[Y(\mathbf{s}_0)] - E[\boldsymbol{\lambda}^T\mathbf{Z}] = 0$$

Prediction variance:

$$PV(\boldsymbol{\lambda}) = C_Y(\mathbf{s}_0, \mathbf{s}_0) - 2\boldsymbol{\lambda}^T c_Y(\mathbf{s}_0) + \boldsymbol{\lambda}^T\mathbf{C}_Z\boldsymbol{\lambda}$$

Ordinary kriging equations

We have: $E[Y(\mathbf{s}_0)] = \mu$ and $E[\boldsymbol{\lambda}^T\mathbf{Z}] = \boldsymbol{\lambda}^T E[\mathbf{Z}] = \boldsymbol{\lambda}^T\mathbf{1}\mu$ thus

$$\mu = \boldsymbol{\lambda}^T\mathbf{1}\mu \Rightarrow 1 = \lambda^T\mathbf{1}$$

The problem then becomes:

$$\min_{\lambda} PV(\boldsymbol{\lambda})$$

subject to:

$$1 = \lambda^T\mathbf{1}$$

Solved by Lagrange multiplier, i.e. minimize

$$C_Y(\mathbf{s}_0, \mathbf{s}_0) - 2\boldsymbol{\lambda}^T c_Y(\mathbf{s}_0) + \boldsymbol{\lambda}^T\mathbf{C}_Z\boldsymbol{\lambda} - 2\kappa(\boldsymbol{\lambda}^T\mathbf{1} - 1)$$

Differentiation wrt $\boldsymbol{\lambda}^T$ gives (which is combined to the final result presented on previous page):

$$\begin{aligned} -2c_Y(\mathbf{s}_0) + 2\mathbf{C}_Z\boldsymbol{\lambda} &= 2\kappa\mathbf{1} \\ \boldsymbol{\lambda}^T\mathbf{1} &= 1 \end{aligned}$$