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

\[ Y_t = \alpha Y_{t-1} + W_t, \quad t = 2, 3, ... \]

\[ Z_t = \beta Y_t + \eta_t, \quad t = 1, 2, 3, ... \]

\[ \eta_t \sim iid (0, \sigma^2) \]

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

\[ Z(s_i) = Y(s_i) + \varepsilon(s_i) \]

\[ \varepsilon(s_i) \sim iid 0 \]
Hierarchical (statistical) models

- Data model
- Process model

In time series setting - state space models

Example

\[ Y_t = \alpha Y_{t-1} + W_t, \ t = 2, 3, \ldots \quad W_t \overset{\text{ind}}{\sim} (0, \sigma^2_W) \]

Process model

\[ Z_t = \beta Y_t + \eta_t, \ t = 1, 2, 3, \ldots \quad \eta_t \overset{\text{ind}}{\sim} (0, \sigma^2_W) \]

Data model

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

\[ Z(s_i) = Y(s_i) + \varepsilon(s_i), \quad \varepsilon(s_i) \sim \text{iid} \]
An unknown function is of interest, i.e. $Y(s), \ s \in D$

Standard problem:
- Observed $Z = (Z_1, ..., Z_m)$
- $Z_i = Y(s_i) + \varepsilon_i, \ i = 1, ..., n$
- Want to predict function in an unobserved position $s_0$, i.e. $Y(s_0)$

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.

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

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

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

\[ \min_{k,L} E\{(Y(s_0) - k - L^T Z)^2\} \]

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

\[ \min_{a(Z)} E\{(Y(s_0) - a(Z))^2\} \]

Know: \( a(Z) = E\{Y(s_0)|Z\} \)
Geostatistical models

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

- \((Y(s_1), \ldots, Y(s_m))\) is multivariate Gaussian for all \(m\) and \(s_1, \ldots, s_m\)

Need to specify

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

Assuming 2. order stationarity:

- \(\mu(s) = \mu, \forall s\)
- \(C_Y(s, s') = C_Y(s - s'), \forall s, s'\)

Common extension:

- \(\mu(s) = x(s)^T \beta\)

Often:

\[ Z(s_i) | Y(s_i), \sigma^2_\varepsilon \sim \text{independent } \text{N}(Y(s_i), \sigma^2_\varepsilon) \]
Covariance function and Variogram

Dependence can be specified through covariance functions or the Variogram

\[ 2\gamma_Y(h) \equiv \text{var}[Y(s + h) - Y(s)] \]
\[ = \text{var}[Y(s + h)] + \text{var}[Y(s)] - 2\text{cov}[Y(s + h), Y(s)] \]
\[ = 2C_Y(0) - 2C_Y(h) \]

- Variograms are more general than covariance functions
- Variogram can exist even if \( \text{var}[Y(s)] = \infty \)
- 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”.
- \( \gamma_Y(h) \) sometimes called semi-variogram
Stationary, isotropic, anisotropic

Strong stationarity    For any \((s_1, ..., s_m)\) and any \(h\)

\[
[Y(s_1), ..., Y(s_m)] = [Y(s_1 + h), ..., Y(s_m + h)]
\]

Stationarity in mean

\[
E(Y(s)) = \mu, \text{ for all } s \in D_A
\]

Stationarity covariance (Depend only on lag)

\[
\text{cov}(Y(s), Y(s + h)) = C_Y(h), \text{ for all } s, s + h \in D_A
\]

Isotropic covariance (depend only on length of lag)

\[
\text{cov}(Y(s), Y(s + h)) = C_Y(\|h\|) \text{ for all } s, s + h \in D_A
\]

Geometric anisotropy in covariance ("Rotated coordinate system")

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

Weak stationarity: Stationarity in mean and a stationarity covariance.
(recall time series)
Isotropic covariance functions/variograms

- Matern covariance function
  \[ C_Y(h; \theta) = \sigma_1^2 \left\{ 2^{\theta_2 - 1} \Gamma(\theta_2) \right\}^{-1} \left\{ ||h||/\theta_1 \right\}^{\theta_2} K_{\theta_2} \left( ||h||/\theta_1 \right) \]

- Powered-exponential
  \[ C_Y(h; \theta) = \sigma_1^2 \exp \left\{ - \left( ||h||/\theta_1 \right)^{\theta_2} \right\} \]

- Exponential
  \[ C_Y(h; \theta) = \sigma_1^2 \exp \left\{ - \left( ||h||/\theta_1 \right) \right\} \]

- Gaussian
  \[ C_Y(h; \theta) = \sigma_1^2 \exp \left\{ - \left( ||h||/\theta_1 \right)^2 \right\} \]

Note:
Many different ways to define the "Range parameter" \( \theta_1 \).

E.g. \( \sigma_1^2 \exp \left\{ -3 \left( ||h||/R \right)^\nu \right\} \)
Stationary spatial covariance functions
Simulation of stationary spatial random field

- Exponential, isotropic
- Gaussian, isotropic
- Exponential, axial anisotropy
- Exponential, geometric anisotropy
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 \( \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |C_Y(h)| \, dh < \infty \), then a valid real-valued covariance function can be written as

\[
C_Y(h) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \cos(\omega^T h) f_Y(\omega) \, d\omega
\]

where \( f_Y(\omega) \geq 0 \) is symmetric about \( \omega = 0 \).

\( f_Y(\omega) \): Spectral density of \( C_Y(h) \).
Nugget effect and Sill

We have

\[ C_Z(h) = \text{cov}[Z(s), Z(s + h)] \]
\[ = \text{cov}[Y(s) + \varepsilon(s), Y(s + h) + \varepsilon(s + h)] \]
\[ = \text{cov}[Y(s), Y(s + h)] + \text{cov}[\varepsilon(s), \varepsilon(s + h)] \]
\[ = C_Y(h) + \sigma_{\varepsilon}^2 I(h = 0) \]

Assume

\[ C_Y(0) = \sigma_Y^2 \]
\[ \lim_{h \to 0} [C_Y(0) - C_Y(h)] = 0 \]

Then

\[ C_Z(0) = \sigma_Y^2 + \sigma_{\varepsilon}^2 \]
\[ \lim_{h \to 0} [C_Z(0) - C_Z(h)] = \sigma_{\varepsilon}^2 = c_0 \]

Possible to include nugget effect also in Y-process.
Nugget/sill

Covariance function

Variogram

Nugget effect

Sill
Sill is a fixed finite level which the variogram converges towards at large lag (i.e. \( \|h\| \to \infty \)).

Variograms without a sill (\( \gamma(h) \to \infty \) as \( \|h\| \to \infty \)) has no equivalent correlation function.

The nugget effect is variability below the resolution in our model.

In the setting with point observations, i.e. \( Z_i = Y(s_i) + \varepsilon_i \) it is difficult (often impossible) to distinguish nugget effect in random field \( Y(s_i) \) and observation error \( \varepsilon_i \). This becomes a modelling choice.

Some softwares mixes nugget effect and observation error.
Estimation of variogram/covariance function

\[ 2\gamma_Z(h) \equiv \text{var}[Z(s + h) - Z(s)] \]

\[ \text{Const expectation} \quad = \quad E[Z(s + h) - Z(s)]^2 \]

Can estimate from “all” pairs having distance \( h \) between.

Problem: Few/no pairs for all \( h \)

Simplifications

- Isotropic: \( \gamma_Z(h) = \gamma_Z^0(||h||) \)
- Lag bin: \( 2\gamma_Z^0(h) = \text{ave}\{(Z(s_i) - Z(s_j))^2; ||s_i - s_j|| \in T(h)\} \)
- If covariates, use residuals
Boreality data

Empirical variogram: Boreality data
Testing for independence

If independence: $\gamma^0_Z(h) = \sigma_Z^2$

Test-statistic $F = \frac{\hat{\gamma}_Z(h_1)}{\hat{\sigma}_Z^2}$, $h_1$ 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 $Z$ (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

\((Z(s_1), ..., Z(s_m), Y(s_0))\) is multivariate Gaussian
Can use rules about conditional distributions:

\[
\begin{pmatrix}
X_1 \\
X_2
\end{pmatrix}
\sim MVN \left( \begin{pmatrix}
\mu_1 \\
\mu_2
\end{pmatrix}, 
\begin{pmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{pmatrix}
\right)
\]

\[
E(X_1|X_2) = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (X_2 - \mu_2)
\]

\[
\text{var}(X_1|X_2) = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}
\]

Need
- Expectations: As for ordinary linear regression
- Covariances: New!
Prediction in the spatial model

\[
\begin{pmatrix}
Y(s_0) \\
Z
\end{pmatrix}
\sim \text{MVN}\left(\begin{pmatrix}
\mu(s_0) \\
\mu_Z
\end{pmatrix}, \begin{pmatrix}
C_Y(s_0, s_0) & c(s_0) \\
c(s_0)^T & C_Z
\end{pmatrix}\right)
\]

\[c(s_0) = \text{cov}[Z, Y(s_0)]\]
\[= \text{cov}[Y, Y(s_0)]\]
\[= (C_Y(s_0, s_1), \ldots, C_Y(s_0, s_m))\]
\[= c_Y(s_0)\]
\[C_Z = \{C_Z(s_i, s_j)\}\]

\[C_Z(s_i, s_j) = \begin{cases} 
C_Y(s_i, s_i) + \sigma^2, & s_i = s_j \\
C_Y(s_i, s_j), & s_i \neq s_j 
\end{cases} \]

\[E(Y(s_0)|Z) = \mu(s_0) + c(s_0)^T C_Z^{-1}(Z - \mu_Z)\]

\[\text{Var}(Y(s_0)|Z) = C_Y(s_0, s_0) - c(s_0)^T C_Z^{-1} c(s_0)\]
Kriging = Prediction

Model

\[ Y(s) = x(s)^T \beta + \delta(s) \]
\[ Z_i = Y(s_i) + \varepsilon_i \]

Prediction of \( Y(s_0) \), i.e. in an unobserved location.

Linear predictors \( \{L^T Z + k\} \)

Optimal predictor minimize

\[
\text{MSPE}(L, k) \equiv E[Y(s_0) - L^T Z - k]^2
\]
\[ = \text{var}[Y(s_0) - L^T Z - k] + \{E[Y(s_0) - L^T Z - k]\}^2 \]

Note: Do not make any distributional assumptions
MSPE(L, k) = \text{var}[Y(s_0) - L^T Z - k] + \{E[Y(s_0) - L^T Z - k]\}^2
= \text{var}[Y(s_0) - L^T Z - k] + \{\mu_Y(s_0) - L^T \mu_Z - k\}^2

Second term is zero if \( k = \mu_Y(s_0) - L^T \mu_Z \).
First term (\( c(s_0) = \text{cov}[Z, Y(s_0)] \)):

\[ \text{var}[Y(s_0) - L^T Z - k] = C_Y(s_0, s_0) - 2L^T c(s_0) + L^T C_Z L \]

Derivative wrt \( L^T \):

\[ -2c(s_0) + 2C_Z L = 0 \]
\[ L^* = C_Z^{-1} c(s_0) \]
giving

\[ Y^*(s_0) = \mu_Y(s_0) + c(s_0)^T C_Z^{-1} [Z - \mu_Z] \]
\[ \text{MSPE}(L^*, k^*) = C_Y(s_0, s_0) - c(s_0)^T C_Z^{-1} c(s_0) \]
The Kriging predictions with a known mean is:

\[ c(s_0) = \text{cov}[Z, Y(s_0)] \]
\[ = \text{cov}[Y, Y(s_0)] \]
\[ = (C_Y(s_0, s_1), ..., C_Y(s_0, s_m)) \]
\[ = c_Y(s_0) \]
\[ C_Z = \{ C_Z(s_i, s_j) \} \]
\[ C_Z(s_i, s_j) = \begin{cases} 
C_Y(s_i, s_i) + \sigma^2, & s_i = s_j \\
C_Y(s_i, s_j), & s_i \neq s_j 
\end{cases} \]
\[ Y^*(s_0) = \mu_Y(s_0) + c_Y(s_0)^T C_Z^{-1}(Z - \mu_Z) \]
Recall now \( Y, Z \) are MVN

\[
\begin{pmatrix}
Z \\
Y(s_0)
\end{pmatrix} = \text{MVN} \left( \begin{pmatrix}
\mu_Z \\
\mu_Y(s_0)
\end{pmatrix}, \begin{pmatrix}
C_Z & c_Y(s_0)^T \\
c_Y(s_0) & C_Y(s_0, s_0)
\end{pmatrix} \right)
\]

Give

\[
Y(s_0) | Z \sim N(\mu_Y(s_0) + c(s_0)^T C_Z^{-1} [Z - \mu_Z], C_Y(s_0, s_0) - c(s_0)^T C_Z^{-1} c(s_0))
\]

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 $\mu(s)$, i.e. $f(s) = Y(s) - \mu(s)$.
- Set $f_i(s) = C_Y(s - s_i)$

Then:
- $F = C_Z$ ($= C_Y$ no observation error)
- $(F^TF)^{-1}F^T = C_Z^{-1}$ (symmetry)
- $\beta = C_Z^{-1}Z$
- $f(s_0) = \sum_{i=1}^{n} \beta_i C_Y(s_0 - s_i)$

Which gives the same result again:

$$f(s_0) = \beta^T c(s_0) = (Z^T)C_Z^{-1}c(s_0)$$

$$Y^*(s_0) = \mu(s_0) + c(s_0)^T C_Z^{-1}(Z - \mu_Z)$$

If we include $\mu(s)$ 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 parameters.
- 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

- Ordinary kriging
  - Plug-in estimates
  - Bayesian approach

- Non-Gaussian models
So far 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
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)

\[ Y^*(s_0) = \mu_Y(s_0) + c_Y(s_0)^T C_Z^{-1}(Z - \mu_Z) \]

Assuming

\[ E[Y(s)] = x(s)^T \beta \]
\[ Z(s_i) \mid Y(s_i), \sigma^2_{\varepsilon} \sim \text{ind.} \text{Gau}(Y(s_i), \sigma^2_{\varepsilon}) \]

Then

\[ \mu_Z = \mu_Y = X \beta \]
\[ C_Z = \Sigma_Y + \sigma^2_{\varepsilon} I \]
\[ c_Y(s_0) = (C_Y(s_0, s_1), ..., C_Y(s_0, s_m))^T \]

and

\[ Y^*(s_0) = x(s_0) \beta + c_Y(s_0)^T [\Sigma_Y + \sigma^2_{\varepsilon} I]^{-1}(Z - X \beta) \]
Kriging

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

$Y^*(s_0) = x(s_0)\beta + c_Y(s_0)^T C_Z^{-1}(Z - X\beta)$
Kriging

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

$$Y^*(s_0) = x(s_0)\beta + c_Y(s_0)^T C_z^{-1}(Z - X\beta)$$

Universal kriging, $EY(s) = x(s)^T \beta$, unknown

$$\hat{Y}(s_0) = x(s_0)^T \hat{\beta}_{gls} + c_Y(s_0)^T C_z^{-1}(Z - X\hat{\beta}_{gls})$$

$$\hat{\beta}_{gls} = [X^T C_z^{-1} X]^{-1} X^T C_z^{-1} Z$$

Possible to show the SK and UK are the limiting cases of BK

$\Sigma_0 \to 0 \Rightarrow BK \to SK$

$\Sigma_0 \to \infty \Rightarrow BK \to UK$
Kriging

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

$$Y^*(s_0) = x(s_0)\beta + c_Y(s_0)^T C_Z^{-1}(Z - X\beta)$$

Universal kriging, $EY(s) = x(s)^T \beta$, unknown

$$\hat{Y}(s_0) = x(s_0)^T \hat{\beta}_{gls} + c_Y(s_0)^T C_Z^{-1}(Z - X\hat{\beta}_{gls})$$

$$\hat{\beta}_{gls} = [X^T C_Z^{-1} X]^{-1} X^T C_Z^{-1} Z$$

Bayesian kriging, $EY(s) = x(s)^T \beta$, with $\beta \sim N(\beta_0, \Sigma_0)$

$$\hat{Y}(s_0) = x(s_0)^T \hat{\beta}_B + c_Y(s_0)^T C_Z^{-1}(Z - X\hat{\beta}_B)$$

$$\hat{\beta}_B = \beta_0 + \Sigma_0 X^T (X \Sigma_0 X^T + C_Z)^{-1} (Z - X \beta_0)$$

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

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

$$\hat{Y}(s_0) = \{c_Y(s_0) + \frac{1(1 - 1^T C_Z^{-1} c_Y(s_0))}{1^T C_Z^{-1} 1}\}^T C_Z^{-1} Z$$

$$= \hat{\mu}_{gls} + c_Y(s_0)^T C_Z^{-1} (Z - 1\hat{\mu}_{gls})$$

$$\hat{\mu}_{gls} = [1^T C_Z^{-1} 1]^{-1} 1^T C_Z^{-1} Z$$

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

$$\text{MSPE}(\lambda) = E(Y(s_0) - \lambda^T Z)^2$$

Unbiased constraint:

$$E[Y(s_0)] - E[\lambda^T Z] = 0$$

Prediction variance:

$$PV(\lambda) = C_Y(s_0, s_0) - 2\lambda^T c_Y(s_0) + \lambda^T C_Z \lambda$$
Ordinary kriging equations

We have: \( E[Y(s_0)] = \mu \) and \( E[\lambda^T Z] = \lambda^T E[Z] = \lambda^T 1 \mu \) thus

\[
\mu = \lambda^T 1 \mu \Rightarrow 1 = \lambda^T 1
\]

The problem then becomes:

\[
\min_{\lambda} PV(\lambda)
\]

subject to:

\[
1 = \lambda^T 1
\]

Solved by Lagrange multiplier, i.e. minimize

\[
C_Y(s_0, s_0) - 2\lambda^T c_Y(s_0) + \lambda^T C_Z \lambda - 2\kappa(\lambda^T 1 - 1)
\]

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

\[
-2c_Y(s_0) + 2C_Z \lambda = 2\kappa 1
\]

\[
\lambda^T 1 = 1
\]