

## 3 Geostatistik

### 3.1 Gauß-Prozesse

#### Definition 3.1. (*Gauß-Zufallsfelder*)

$Y = \{Y(s), s \in D \subseteq \mathbb{R}^d\}$  heißt Gauß-Zufallsfeld der Dimension  $d$  (Gauß-ZF oder Gaussian random field (GRF) oder Gauß-Prozess):  $\Leftrightarrow$  Für beliebige Lokationen  $s_1, \dots, s_n$  ist  $(Y(s_1), \dots, Y(s_n))'$  multivariat normalverteilt.

Bezeichnungen:

- $\mu(s) = \mathbb{E}[Y(s)]$  : Erwartungswertfunktion, räumlicher Trend
- $\sigma^2(s) = \text{Var}(Y(s))$  : Varianzfunktion
- $c(s, s') = \text{Cov}(Y(s), Y(s'))$  : Kovarianzfunktion

#### Theorem 1.

Ein Gauß-Prozess wird durch  $\mu(s)$  und  $c(s, s')$  vollständig spezifiziert.

#### Beweis:

Für alle  $s_1, \dots, s_n \in D$  gilt:

$$(Y(s_1), \dots, Y(s_n))' \sim N_n \left( \begin{pmatrix} \mu(s_1) \\ \vdots \\ \mu(s_n) \end{pmatrix}, \begin{pmatrix} c(s_i, s_j)_{i=1, \dots, n; j=1, \dots, n} \end{pmatrix} \right)$$

□

#### 3.1.1 Stationarität

##### Definition 3.2. (*Schwach stationär*)

Ein räumlicher stochastischer Prozess  $Y = \{Y(s), s \in D\}$  heißt schwach stationär:  $\Leftrightarrow \mathbb{E}[Y(s)] = \mu = \text{const}$ ,  $\text{Cov}(Y(s), Y(s')) = c(s - s')$  für alle  $s, s' \in D$

##### Definition 3.3. (*Strenge Stationarität*)

Ein räumlicher stochastischer Prozess  $Y$  heißt streng stationär:  $\Leftrightarrow$  alle endlichdimensionale Verteilungen sind verschiebungsinvariant.

#### Theorem 2. (Strenge und schwache Stationarität)

1.  $Y$  streng stationär  $\Rightarrow Y$  schwach stationär
2.  $Y$  sei Gauß-Prozess. Dann gilt:  $Y$  streng stationär  $\Leftrightarrow Y$  schwach stationär.

### 3.1.2 Theoretisches (Ko-)Variogramm und Korrelationsfunktion

**Definition 3.4.** Sei  $Y = \{Y(s), s \in D\}$  ein räumlicher Prozess. Dann heißt

$$c(s, s') := \text{Cov}(Y(s), Y(s'))$$

Kovariogramm oder Kovarianzfunktion und

$$\gamma(s, s') := \frac{1}{2} \text{Var}(Y(s) - Y(s'))$$

(Semi-)Variogramm.

Es gilt:

$$\gamma(s, s') = \frac{1}{2} \text{Var}(Y(s)) + \frac{1}{2} \text{Var}(Y(s')) - c(s, s')$$

**Definition 3.5.** Sei  $Y = \{Y(s), s \in D\}$  ein stationärer Prozess und  $h := s - s'$ . Dann heißt  $c(h) = c(s - s') = c(s, s')$  stationäre Kovarianzfunktion und  $\rho(h) := \frac{c(h)}{c(0)} = \frac{c(h)}{\sigma^2}$  Korrelogramm oder Korrelationsfunktion.

Es gilt:

$$\begin{aligned} c(0) &= \sigma^2 = \text{Var}(Y(s)) = \text{const.} \\ c(h) &= c(-h) \quad \rho(h) = \rho(-h) \end{aligned}$$

#### **Definition 3.6. (Isotropie und Anisotropie)**

Ist  $c(s - s') = c(\|s - s'\|)$  nur eine Funktion des euklidischen Abstands von  $s$  und  $s'$ , so heißt  $Y$  bzw.  $c$  isotrop; ansonsten anisotrop.

#### **Definition 3.7. (Intrinsisch stationär)**

$Y$  intrinsisch stationär:  $\Leftrightarrow \mathbb{E}[Y(s)] = \mu, \quad \gamma(s, s') = \gamma(s - s')$ .

$\gamma(h)$  heißt stationäres Variogramm.

Es gilt:  $\gamma(h) = \gamma(-h), \quad \gamma(0) = 0$ .

Hängt  $\gamma(h)$  nur von  $\|h\|$  ab, so heißt  $\gamma$  isotrop.

#### **Theorem 3.**

$Y$  schwach stationär  $\Rightarrow Y$  intrinsisch stationär mit  $\gamma(h) = c(0) - c(h) = \sigma^2(1 - \rho(h))$ .

#### **Definition 3.8. ( $L^2$ -Stetigkeit, $L^2$ -Differenzierbarkeit)**

$Y$  heißt  $L^2$ -stetig:  $\Leftrightarrow \mathbb{E}[Y(s+h) - Y(s)]^2 \rightarrow 0$  für  $h \rightarrow 0$

$Y$  heißt  $L^2$ -differenzierbar:  $\Leftrightarrow$  Es existiert ein SP  $Y'$  mit

$$\mathbb{E} \left[ \frac{Y(s+h) - Y(s)}{h} - Y'(s) \right]^2 \rightarrow 0 \quad \text{für } h \rightarrow 0.$$

Später: wenn  $Y$  nicht  $L^2$ -stetig, liegt ein Nugget Effekt vor.  
Wir interessieren uns (erstmal) nur für stationäre Prozesse.

**Eigenschaften:**  $Y$  stationär  $\Rightarrow$

- $Y$  ist  $L^2$ -stetig  $\Leftrightarrow \rho(h)$  stetig in  $h = 0$
- $Y$   $k$ -mal  $L^2$ -differenzierbar  $\Leftrightarrow \rho(h)$   $2k$ -mal in  $h = 0$  differenzierbar.
- $Y$  ist  $L^2$ -stetig  $\Leftrightarrow \gamma(h)$  stetig in  $h = 0$
- $\gamma(h) = \text{const}$  für  $h > 0 \Rightarrow \text{Cov}(Y(s), Y(s')) = 0$  für beliebige  $s \neq s'$ .

Die räumliche Korrelation wird bei stationären GRF also durch  $\rho(h)$  bzw.  $\gamma(h)$  bestimmt.

## 3.2 Empirisches Variogramm

In diesem Abschnitt nichtparametrische Schätzung des (Ko-)Variogramms als exploratives Werkzeug. Später definieren wir uns parametrische Modelle, die auch parametrische Schätzung erlauben.

### 3.2.1 Schätzer

#### (a) Empirisches Variogramm

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

mit  $N(h) = \{(s_i, s_j) : s_i - s_j = h; i, j = 1, \dots, n\}$

**Herleitung:**

$$\begin{aligned} 2\gamma(h) &= \text{Var}(Y(s_i) - Y(s_j)) \Leftrightarrow \\ \gamma(h) &= \frac{1}{2} \text{Var}(Y(s_i) - Y(s_j)) \\ &= \frac{1}{2} (\mathbb{E}[(Y(s_i) - Y(s_j))^2] - (\mathbb{E}[Y(s_i) - Y(s_j)])^2) \\ &= \frac{1}{2} (\mathbb{E}[(Y(s_i) - Y(s_j))^2] - (\mathbb{E}[Y(s_i)] - \mathbb{E}[Y(s_j)])^2) \\ &= \frac{1}{2} (\mathbb{E}[(Y(s_i) - Y(s_j))^2] - (\mu - \mu)^2) \\ &= \frac{1}{2} (\mathbb{E}[(Y(s_i) - Y(s_j))^2]) \Rightarrow \\ \hat{\gamma}(h) &= \frac{1}{2|N(h)|} \sum_{i,j \in N(h)} (y(s_i) - y(s_j))^2 \end{aligned}$$

- Analoger Schätzer für Kovariogramm verzerrt.
- Nachteil:  $|N(h)|$  meist klein, oft  $|N(h)| = 1 \Rightarrow$  Schätzer instabil.

(b) **Empirisches Variogramm nach Matheron**

Für isotrope, intrinsisch stationäre Prozesse. Einteilung der euklidischen Distanz  $d_{ij} = d(s_i, s_j)$  zwischen zwei Lokationen in  $K$  Klassen (Bins):

$$H_k = \{d_{ij} : h_{k-1} \leq d_{ij} < h_k\}, \quad k = 1, \dots, K; \quad h_0 = 0$$

$$\hat{\gamma}_k = \frac{1}{2|N_k|} \sum_{i,j \in N_k} (Y(s_i) - Y(s_j))^2$$

$$N_k = \{(s_i, s_j) : s_i - s_j \in H_k\}$$

Fragen / Probleme:  $K = ?$ , Wahl der Klassengrenzen.

(c) **Constant nearest-neighbour estimator**

$$\hat{\gamma}_{CNN}(h) = \frac{\sum_{i < j} \mathbb{1}\{-\delta < h - \|s_i - s_j\| < \delta\} (Y(s_i) - Y(s_j))^2}{\sum_{i < j} \mathbb{1}\{-\delta < h - \|s_i - s_j\| < \delta\}}$$

(d) **Variable nearest-neighbour estimator**

$$\hat{\gamma}_{VNN}(h) = \frac{\sum_{i < j} \frac{1}{\delta_0(\|s_i - s_j\|)} K\left(\frac{h - \|s_i - s_j\|}{\delta \delta_0(\|s_i - s_j\|)}\right) (Y(s_i) - Y(s_j))^2}{\sum_{i < j} \frac{1}{\delta_0(\|s_i - s_j\|)} K\left(\frac{h - \|s_i - s_j\|}{\delta \delta_0(\|s_i - s_j\|)}\right)},$$

wobei  $K$  den Kern und  $\delta$  die Bandbreite bezeichnen.  $\delta_0$  erlaubt bessere Anpassung, oft wird  $\delta_0 \equiv 1$  angenommen.  $\delta$  kann AMISE-optimal gewählt werden.

**3.2.2 Bezeichnungen**(a) **Nugget-Effekt**

Empirisch ist oft zu beobachten, dass das Variogramm bei  $\|h\| = 0$  springt. Sei  $Y(s)$  der wahre, stationäre Prozess und  $Z(s)$  ein beobachteter Prozess mit

$$Z(s) = Y(s) + \varepsilon(s), \quad \varepsilon_i \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2)$$

Dann gilt:

$$c_Z(h) = \begin{cases} c_Y(h) + \sigma_\varepsilon^2, & \text{falls } \|h\| > 0 \\ c_Y(0), & \text{falls } \|h\| = 0 \end{cases}$$

(b) **Schwellenwert/Sill**

Mit *Schwellenwert* oder *Sill* bezeichnet man den Wert, den das Vario-

ogramm erreicht, wenn der Abstand  $h$  gegen unendlich geht. Zieht man von diesem Wert den Nugget-Effekt ab, so erhält man den *Partial Sill*.

(c) **Range**

Mit Range wird der Abstand bezeichnet, ab dem das Variogramm den Schwellenwert fast erreicht. Punkte, deren Abstand größer ist als der Range werden als unabhängig voneinander betrachtet. Das hat zur Folge, dass dann, wenn das Variogramm positiv konstant ist (außer am Ursprung, wo es null ist), kein Zusammenhang zwischen  $Z(s_i)$  und  $Z(s_j)$  für alle  $s_i$  und  $s_j$  besteht.

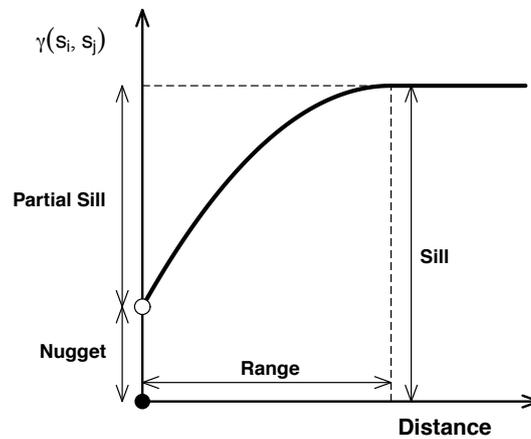


Abbildung 3: Variogramm

### 3.3 Parametrische Modelle für Korrelationsfunktionen und (Ko-)Variogramme

#### 3.3.1 Gültigkeit von Korrelationsfunktionen

Gültige Modelle für Kovariogramme müssen positiv semidefinit sein, d.h.

$$\sum_{i=1}^n \sum_{j=1}^n a_i a_j c(s_i - s_j) \geq 0 \quad \text{für beliebige } a_i, a_j, s_i, s_j$$

erfüllen. Völlig analog für Korrelationsfunktionen. Dies sichert:

$$\text{Var}(a_1 Y(s_1) + \dots + a_n Y(s_n)) \geq 0.$$

**Es gilt:**

$c_1(h)$  und  $c_2(h)$  gültige Kovarianzfunktionen  $\Rightarrow$

$$b_1 c_1(h) + b_2 c_2(h) \quad \text{für } b_1 > 0 \text{ und } b_2 > 0$$

$$c_1(h) \cdot c_2(h)$$

sind gültige Korrelationsfunktionen.

**Wünschenswerte Eigenschaften:**

- $\rho(h)$  fallend in  $h$ ,  $\rho(h) \rightarrow 0$  für  $h \rightarrow \infty$ .
- Wenigstens 1 Parameter steuert, wie schnell  $\rho(h) \rightarrow 0$  für  $h \rightarrow \infty$  geht.

**Theorem 4. (Bochner-Kriterium für isotrope Korrelationsfunktionen)**

$\rho(u)$  ist gültige isotrope Korrelationsfunktion  $\Leftrightarrow$

$\rho(u)$  ist charakteristische Funktion einer symmetrischen ZV  $X \Leftrightarrow$

$$\rho(u) = \int_{-\infty}^{+\infty} \exp(iux) f(x) dx$$

Weiterführende Literatur: *Gneiting and und P. Guttorp [2007]*

### 3.3.2 Übliche Korrelationsfunktionen

**Erinnerung:** Es gilt unter Isotropie:  $\gamma(u) = c(0) - c(u) = \sigma^2(1 - \rho(u))$ .

(a) **Potenz-Exponentialfamilie** (Abb. 4)

$$\rho(u) = \exp\{-(u/\phi)^\kappa\}, \quad \phi > 0, \quad 0 < \kappa \leq 2$$

$\kappa = 1$  : Exponential-Korrelationsfunktion,

$\kappa = 2$  : Gauß-Korrelationsfunktion.

(b) **Sphärische Familie** (Abb. 5)

$$\rho(u) = \begin{cases} 1 - \frac{3}{2}(u/\phi) + \frac{1}{2}(u/\phi)^3, & 0 \leq u \leq \phi \\ 0, & u > \phi \end{cases}$$

Problem: Nicht differentierbar in  $u = \phi \Rightarrow$  Schwierigkeiten bei ML-Schätzung.

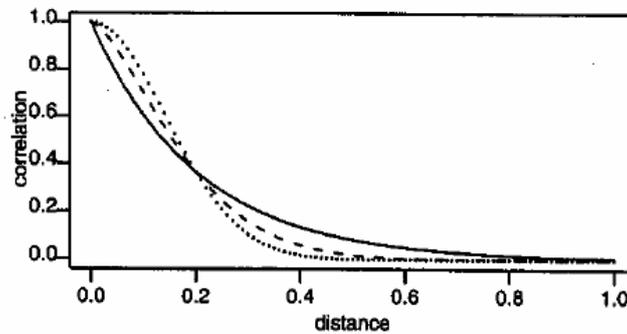


Abbildung 4: Powered-Exponential Korrelationsfunktion mit  $\phi = 0.2$  und  $\kappa = 1$  (durchgezogene Linie),  $\kappa = 1.5$  (gestrichelte Linie) und  $\kappa = 2$  (gepunktete Linie).

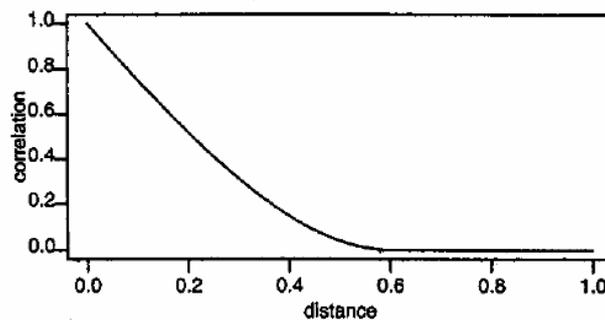


Abbildung 5: Sphärische Korrelationsfunktion mit  $\phi = 0.64$ .

(c) **Matérn-Familie** (Abb. 6)

$$\rho(u) = \frac{1}{2^{\kappa-1} \Gamma(\kappa)} (u/\phi)^\kappa K_\kappa(u/\phi), \quad \phi > 0, \quad \kappa > 0$$

$K_\kappa$ : Besselfunktion, i.A.nicht in geschlossener Form anzugeben. Numerische Approximationen in [Abramowitz and Stegun \[1972\]](#). Explizite Formeln für  $\kappa = 0.5, 1.5, \dots$  möglich:

$$\begin{aligned} \rho(u; \phi = 1, \kappa = 0.5) &= \exp(-u) \\ \rho(u; \phi = 1, \kappa = 1.5) &= \exp(-u)(1 + u) \\ \rho(u; \phi = 1, \kappa = 2.5) &= \exp(-u) \left( 1 + u + \frac{1}{3}u^2 \right) \end{aligned}$$

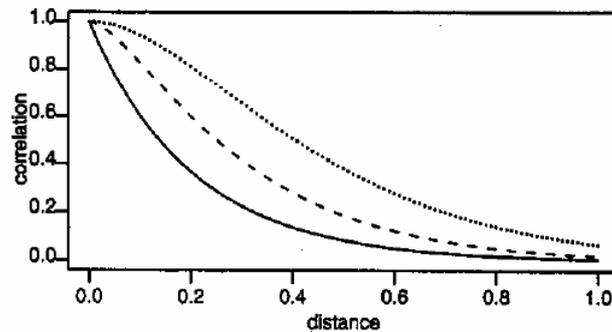


Abbildung 6: Matérn-Korrelationsfunktion mit  $\phi = 0.2$  und  $\kappa = 1$  (durchgezogene Linie),  $\kappa = 1.5$  (gestrichelte Linie) und  $\kappa = 2$  (gepunktete Linie).

$$\rho(u; \phi = 1, \kappa = 3.5) = \exp(-u) \left( 1 + u + \frac{2}{5}u^2 + \frac{1}{15}u^3 \right)$$

$\kappa \rightarrow \infty$  : Gauß-Korrelationsfunktion

Obwohl nur numerisch greifbar gebräuchlichste Korrelationsfunktion, vgl. Abschnitt 4.

### 3.4 Anisotrope und nicht-stationäre Prozesse

Arten der Anisotropie:

- Geometrische Anisotropie: Sill identisch, Range unterschiedlich
- Zonale Anisotropie: Sill unterschiedlich

**Erkennung:** Berechne Variogramme in verschiedenen Raumrichtungen.

#### Geometrische (Range-)Anisotropie

Wähle in isotroper Korrelationsfunktion  $\rho(u)$  für  $u$  statt  $u^2 = \|h\|^2$  eine nicht-euklidische Distanz, z.B.:

$$u^2 = h' R(\psi) S(\delta) R(\psi) h$$

mit

$$R(\psi) = \begin{pmatrix} \cos \psi & -\sin \psi \\ \sin \psi & \cos \psi \end{pmatrix} \quad \text{Rotationsmatrix, } \psi \in (0, 2\pi)$$

$$S(\delta) = \begin{pmatrix} \delta^{-1} & 0 \\ 0 & 1 \end{pmatrix} \quad \text{Dehnungsmatrix, } \delta \geq 1$$

**Zonale Anisotropie** (siehe Ecker and Gelfand [2003])

$\rho_l(u)$ ,  $l = 1, \dots, p$  seien gültige 2D isotrope Korrelationsfunktionen  $\Rightarrow \rho(h) = \rho_1(\sqrt{h' B_1 h}), \dots, \rho_p(\sqrt{h' B_p h})$  mit  $B_l$  pos. definit ist gültige Korrelationsfunktion in  $\mathbb{R}^2$ .

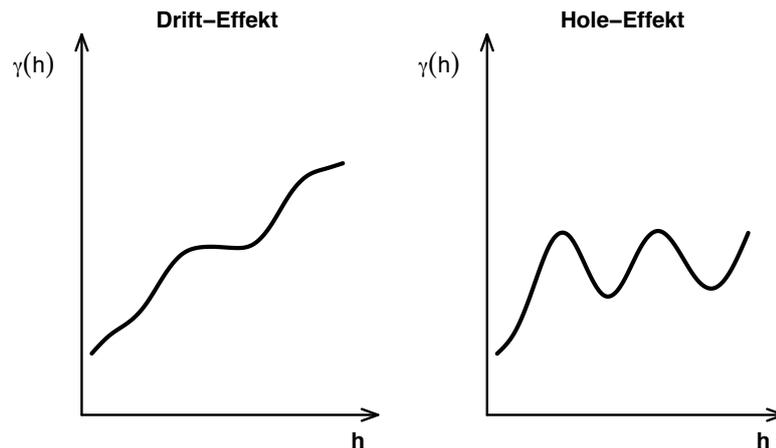


Abbildung 7: Der Drift Effekt weist auf Nicht-Stationarität hin. Der Hole Effekt weist auf regelmäßige Strukturen hin.

### 3.5 Kriging

Ziel: Schätzung der (wahren) Oberfläche. Vorerst seien (deterministische) Erwartungswertfunktion und Korrelationsfunktion bekannt.

#### 3.5.1 Einfaches Kriging

**Gegeben:** Daten  $y(s_1), \dots, y(s_n)$

**Daten-Modell:**  $Y(s_i) = Z(s_i) + \varepsilon(s_i)$

$Z$  sei ein (stationäres) Zufallsfeld mit  $\mu(s)$  und  $c(s, s') = \sigma^2 \rho(s, s')$

$\varepsilon(s_i)$  sei i.i.d. mit  $\mathbb{E}[\varepsilon(s_i)] = 0$  und Varianz  $\sigma_\varepsilon^2$

**Gesucht:** linearer Prädiktor für  $Z(s_0)$  für alle  $s_0 \in D$  mit minimalem mittleren quadratischen Prädiktionsfehler.

$$\hat{Z}(s_0) = l'Y + K$$

Wegen Verschiebungssatz gilt:

$$\text{MSPE}(l, K) = \mathbb{E}[Z(s_0) - l'Y - K]^2 = \text{Var}(Z(s_0) - l'Y - K) + \{\mathbb{E}[Z(s_0) - l'Y - K]\}^2$$

$$\mathbb{E}[Z(s_0) - l'Y - K] = \mu(s_0) - l'\mu_Y - K, \quad \text{wobei } \mu_Y = (\mu(s_1), \dots, \mu(s_n))$$

$$\text{Var}(Z(s_0) - l'Y - K) = \text{Var}(Z(s_0)) - 2l'\text{Cov}(Z(s_0), Y) + l'C_Y l$$

wobei  $K$  fest ist.

$$\text{Cov}(Z(s_0), Y) = (c(s_0, s_1), c(s_0, s_2), \dots, c(s_0, s_n))' =: c(s_0)$$

$$C_Y = (C_Y(s_i, s_j))_{\substack{i=1, \dots, n \\ j=1, \dots, n}} \quad \text{mit} \quad C_Y(s_i, s_j) = \begin{cases} c(s_i, s_i) + \sigma_\varepsilon^2, & \text{für } i = j \\ c(s_i, s_j), & \text{sonst} \end{cases}$$

$$\text{Ableiten nach } l \Rightarrow l^* = C_Y^{-1}c(s_0)$$

$$\text{Ableiten nach } K \Rightarrow K^* = \mu(s_0) - c(s_0)'C_Y^{-1}\mu_Y$$

Daraus folgt, der lineare Prädiktor ist

$$\widehat{Z}(s_0) = \mu(s_0) + c(s_0)'C_Y^{-1}(Y - \mu_Y)$$

Die Kriging-Varianz ist dann

$$\sigma_Y^2(s_0) = \text{MSPE}(l^*, K^*) = c(s_0, s_0) - c(s_0)'C_Y^{-1}c(s_0)$$

### Bemerkungen:

(1) Fall  $\mu(s) \equiv \mu$  heißt gewöhnliches (ordinary) Kriging.

(2) Fall  $\mu(s) = x(s)'\beta$  mit bekanntem  $\beta$  heißt universelles Kriging.

**Beispiel 22.** Sei  $\mu(s) = \mu = 0$ ,  $\text{Var}(Z(s)) = 1$  und  $\sigma_\varepsilon^2 = 0.5$ .

Gegeben: Zwei Punkte  $s_1 = (1, 0)$  und  $s_2 = (2, 0)$ ,  $s_1$  näher an  $s_0 = (0.5, 0)$  als  $s_2$ .

Annahme: es gelte die Exponentialkorrelationsfunktion mit  $\sigma^2 = 1$ .

$$c(s_1, s_0) = 0.61, \quad c(s_2, s_0) = 0.22 \quad \text{und} \quad c(s_1, s_2) = 0.37$$

Also

$$c(s_0) = (0.61, 0.22)'$$

und

$$C_Y = \begin{pmatrix} 1 & 0.37 \\ 0.37 & 1 \end{pmatrix} + 0.5\mathbb{I} = \begin{pmatrix} 1.5 & 0.37 \\ 0.37 & 1.5 \end{pmatrix}$$

Kriging ergibt

$$\widehat{Z} = 0.395 Y(s_1) + 0.049 Y(s_2)$$

wobei 0.395 und 0.049 die Kriging-„Gewichte“ sind. Die zugehörige Varianz ist

$$\sigma_Y^2 = 1 - (0.61, 0.22) \begin{pmatrix} 1.5 & 0.37 \\ 0.37 & 1.5 \end{pmatrix}^{-1} \begin{pmatrix} 0.61 \\ 0.22 \end{pmatrix} = 0.7508$$

also unabhängig von  $Y$ !

**Bemerkungen:**

- Größerer Nugget-Effekt  $\rightarrow$  Größere Kriging-Varianz.
- Größerer Sill  $\rightarrow$  Größere Kriging-Varianz.
- Größerer Range  $\rightarrow$  Kleinere Kriging-Varianz.
- Kriging-Varianz klein, wo mehr Daten (Trick: erhebe weitere Daten dort, wo Kriging-Varianz groß).

### 3.5.2 Kriging im Gauß-Modell

Nun Normalverteilungsannahme.

**Erster Ansatz: additives Modell**

$$Y(s) = \mu(s) + Z(s) + \varepsilon(s)$$

$\mu(s)$ : deterministisch

$Z(s)$ : stationäres Gaußzufallsfeld mit  $\mathbb{E}[Z(s)] = 0$  und  $\text{Var}(Z(s)) = \tau^2$

$\varepsilon(s)$ : i.i.d. mit  $\mathbb{E}[\varepsilon(s)] = 0$  und  $\text{Var}(\varepsilon(s)) = \sigma^2$

Gelegentlich auch

$$Y(s) = \mu(s) + W(s) + V(s) + \varepsilon(s)$$

$W(s)$  und  $V(s)$  unabhängige und stationäre GZF mit  $\rho_W$  und  $\rho_V$  gegeben und  $\rho_V$  geht schnell gegen 0.

$\rightarrow$  „Mikroskalenvariation“, vgl. Bsp. 2 (Rongelap Daten).

Oft aber  $V(s)$  und  $W(s)$  nicht unterscheidbar.

**Alternativer Ansatz: hierarchisches Modell**

$$Y(s)|Z(s), \sigma_\varepsilon^2 \stackrel{\text{i.n.d.}}{\sim} N(Z(s), \sigma_\varepsilon^2)$$

$Z(s)$  Gauss-Zufallsfeld mit  $\mathbb{E}[Z(s)] = \mu(s)$ ,  $\text{Var}(Z(s)) = \tau^2$ ,  $\rho_Z(h)$  gegeben.

**Die beiden Ansätze sind identisch.** Es gilt jeweils:

$$\mathbb{E}[Y(s)] = \mu(s) \text{ und } \text{Cov}(Y(s_i), Y(s_j)) \text{ wie oben}$$

$$\begin{aligned}
Y &= (Y(s_1), \dots, Y(s_n))' \\
R &= (v_{ij}) = (\rho(s_i - s_j)) \\
\mu &= (\mu(s_1), \dots, \mu(s_n))' \\
Y &\sim N(\mu, \tau R + \sigma^2 \mathbb{I}) \\
r &:= (r_1, \dots, r_n)'; \quad r_i = \rho(s_0 - s_i)
\end{aligned}$$

Dann gilt auch für  $Z_0 = Z(s_0)$

$$\begin{pmatrix} Z_0 \\ Y \end{pmatrix} \sim N \left( \begin{pmatrix} \mu(s_0) \\ \mu \end{pmatrix}, \begin{pmatrix} \tau^2 & \tau^2 r' \\ \tau^2 r' & \tau^2 R + \sigma^2 \mathbb{I} \end{pmatrix} \right)$$

$Z_0|Y$  ist normalverteilt mit:

$$\mathbb{E}[Z_0|Y] = \mu(s_0) + \tau^2 r' (\tau^2 R + \sigma^2 \mathbb{I})^{-1} (Y - \mu) =: \hat{Z}_0$$

$$\text{Var}(Z_0|Y) = \tau^2 - \tau^2 r' (\tau^2 R + \sigma^2 \mathbb{I})^{-1} \tau^2 r$$

Entspricht also dem einfachen Kriging-Schätzer und dessen Varianz.

### 3.5.3 Inferenz

Seien nun  $\mu(s) = X'(s)\beta$  und  $\rho(h)$  unbekannt.

**Bemerkung:**  $X$  kann parametrisches Modell für Oberfläche, etwa  $\mu(s) = \beta_0 + \beta_1 s_X + \beta_2 s_Y + \beta_3 s_X^2 + \beta_4 s_Y^2 + \beta_5 s_X s_Y$ , sein und/oder räumlich vorliegende Kovariable (für die Prädiktion müssen diese aber auf ganz  $D$  vorliegen).

(a) **Maximum-Likelihood-Schätzer:**

$$Y(s_i) = X'(s_i)\beta + Z(s_i) + \varepsilon(s_i)$$

$\theta$  Korrelationsparameter,  $\alpha := (\sigma^2, \tau^2, \theta)$

$$\text{Cov}(Y) = \sigma^2 \mathbb{I} + \tau^2 R(\theta) =: \Sigma(\alpha)$$

Das allgemeine lineare Modell ergibt sich zu

$$Y \sim N(X\beta, \Sigma(\alpha))$$

Log-Likelihood (bei Normalverteilungsannahme)

$$l(\beta, \alpha) = -\frac{1}{2} (\log |\Sigma(\alpha)| + (y - X\beta)' \Sigma^{-1}(\alpha) (y - X\beta))$$

Für gegebenes  $\alpha$  ist

$$\hat{\beta}(\alpha) = (X' \Sigma^{-1}(\alpha) X)^{-1} X' \Sigma^{-1}(\alpha) y$$

ML-Schätzer für  $\beta$ . Einsetzen in  $l(\beta, \alpha)$  ergibt die *Profil-log-Likelihood* für  $\alpha$

$$l(\alpha) := l(\hat{\beta}(\alpha), \alpha).$$

Maximieren liefert ML-Schätzer  $\hat{\alpha}$ . Einsetzen in  $\hat{\beta}(\alpha)$  liefert  $\hat{\beta}(\hat{\alpha})$

**Problem:** Verzerrung in kleineren Stichproben  $\Rightarrow$  Übergang zu REML.

### (b) Restringierte ML (REML) - Schätzung

Ziel: Bias-Reduktion

#### Herleitung 1

- Transformiere  $Y = X\beta + Z + \varepsilon$  linear zu  $Y^* = AY$  so, dass Verteilung von  $Y^*$  nicht von  $\beta$  abhängt  
 $\Rightarrow A$  so, dass  $AX = 0$ . Geht z.B. mit  $A = \mathbb{I} - X(X'X)^{-1}X'$ .
- Schätze  $\alpha$  nach ML-Prinzip für  $Y^*$ .
- Ergebnis:  $\hat{\alpha}_{REML}$  maximiert die restringierte log-Likelihood  $l^*(\alpha) =$

$$-\frac{1}{2} \left( \log |\Sigma(\alpha)| + \log |X' \Sigma^{-1}(\alpha) X| + (y - X\hat{\beta}(\alpha))' \Sigma^{-1}(\alpha) (y - X\hat{\beta}(\alpha)) \right)$$

(Alternative: ersetze  $\hat{\beta}(\alpha)$  durch  $\beta$  und schätze  $\alpha$  und  $\beta$  simultan.)

#### Herleitung 2

- Für das hierarchische Modell gilt

$$\begin{aligned} Y|\beta, Z, \alpha &\sim N(X\beta + Z, \sigma^2\mathbb{I}) \\ Z &\sim N(0, \tau^2 R(\theta)) \end{aligned}$$

- Sei  $p(y|\beta, Z, \alpha)$  zugehörige Dichte. Bilde marginale Dichte (mit  $p(\beta) \propto \text{const}$ )

$$p(\alpha|y) = \int p(y|\beta, Z, \alpha) d\beta dp(Z)$$

und maximiere bezüglich  $\alpha$ .

- Es gilt (Harville [1974]) bis auf multiplikative Konstanten

$$\log p(\alpha | y) = l^*(\alpha)$$

Nach Diggle et al. [2003] ist REML in Ordnung für gewöhnliches Kriging, ansonsten ist die MSE-Güte aber unklar; REML sensitiv bei Missspezifikation von  $\mu(s)$  (allgemeines Problem).

Manchmal wird

$$l_{re}(\alpha, \beta) = -\frac{1}{2} (\log |\Sigma(\alpha)| + \log |X' \Sigma^{-1}(\alpha) X| + (y - X\beta)' \Sigma^{-1}(\alpha) (y - X\beta))$$

als restringierte log-Likelihood bezeichnet und zum simultanen Schätzen von  $\alpha, \beta$  maximiert.

### (c) Bayes-Schätzer

Als Hierarchisches Modell:

$$Y(s) | Z(s), \sigma_\varepsilon^2 \sim N(Z(s), \sigma_\varepsilon^2)$$

Prioris:

$$\left. \begin{array}{l} \sigma_\varepsilon^2 \sim \text{IG}(\cdot, \cdot) \\ Z(s) \sim \text{GZF}(\tau^2, \theta) \\ \tau^2 \sim \text{IG}(\cdot, \cdot) \\ \theta \sim ? \end{array} \right\} \text{semi-konjugierte}$$

Posteriori  $\propto$  Priori  $\cdot$  Likelihood

Vollständig bedingte Posteriori:

$$\begin{array}{l} Z(s) | \cdot \sim N(\cdot, \cdot) \\ \tau^2 | \cdot \sim \text{IG}(\cdot, \cdot) \\ \sigma^2 | \cdot \sim \text{IG}(\cdot, \cdot) \\ \theta | \cdot \sim ? \quad (\text{oft } \kappa \text{ fix}) \end{array}$$

Prädiktive Posteriori

$$Y(s_0) | Z, \sigma^2 \sim N(Z(s_0), \sigma_\varepsilon^2)$$

### 3.5.4 Generalisiertes Kriging

Geostatistische Modelle für nicht-normalverteilte Zielvariablen, z.B.  $Y$  binär, binomial, kategorial oder Zählvariable, bzw. Erweiterung des generalisierten

linearen Modells um geostatistische Komponente. Vgl. Beispiele 2 (Rongelap) und 3 (Waldschäden).

### Hierarchisches Modell

- (1)  $Y(s_i) | W(s_i)$ ,  $i = 1, \dots, n$  bedingt unabhängig, Dichten aus Exponentialfamilien
- (2)  $W$  stationärer Gauß-Prozess,  $\mathbb{E}[W(s)] = 0$
- (3)  $\mu_i = \mathbb{E}[Y(s_i) | W(s_i)] = h(\eta_i)$ ,  $\eta_i = x_i' \beta + W(s_i)$ ,  $x_i = x(s_i)$

**Speziell:**  $x_i' \beta = \mu$  **Gewöhnliches Kriging**

$$\widehat{W}(s) = \mathbb{E}[W(s) | Y], \quad Y = (Y(s_1), \dots, Y(s_n))'$$

bzw.:  $\widehat{\eta}(s) = x_i' \widehat{\beta} + \widehat{W}(s)$  ist generalisierter linearer Kriging-Prädiktor für  $W$  bzw.  $\eta$  an der Stelle  $s$ .

### Statistische Inferenz

Zu schätzen sind:  $\beta$  Regressionsparameter,  $\theta$  Kovarianzparameter,  $W(s)$  an Beobachtungsstellen  $s_1, \dots, s_n$  und an Vorhersagestelle  $s$ .

Sei die (bedingte) Likelihood, mit  $W = (W(s_1), \dots, W(s_n))'$

$$p(y | W, \beta) = \prod_{i=1}^n f_i(y(s_i) | W(s_i), \beta)$$

z.B.

$$f_i(y(s_i) | W(s_i), \beta) = \pi_i^{y(s_i)} (1 - \pi_i)^{1-y(s_i)} \quad \text{für binäre } Y(s_i)$$

und

$$\pi_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \quad \text{bei Logit-Modell}$$

(a) **Bayes-Ansatz**

$$\begin{aligned} W | \theta &\sim N_n(\cdot, \cdot) \\ p(\beta) &\propto \text{const} \quad \text{oder} \quad \beta \sim N(\cdot, \cdot) \\ \sigma^2, \tau^2 &\sim \text{IG}(\cdot, \cdot) \\ \theta &\sim ? \end{aligned}$$

MCMC mit Sampling aus vollständig bedingten Dichten:

$(\beta   \cdot)$	Metropolis-Hastings
$(W   \cdot)$	Metropolis-Hastings
$(\tau^2   \cdot)$	IG mit aufdatierten Parametern
$(\sigma^2   \cdot)$	IG mit aufdatierten Parametern
$(\theta   \cdot)$	?

(b) **Penalisierter Likelihood-Ansatz**

$$(1) \quad l_{pen}(\beta, W | \theta) = l(\beta, W) + \underbrace{\frac{1}{\tau^2} W' R^{-1}(\theta) W}_{\text{vgl. Kapitel 6}} \rightarrow \max_{\beta, W}$$

$$\begin{aligned} l(\beta, W) &= \log p(y | W, \beta) \\ r_{ij} &= \rho(\|s_i - s_j\| | \theta) \end{aligned}$$

(2)  $\tau^2, \theta, \sigma^2$  mit REML

(1) und (2) iterieren

(c) **MCML: Monte Carlo Maximum Likelihood (Christensen [2004])**Mit  $\varphi$  wird im Folgenden der Vektor der Parameter bezeichnet.

$$\begin{aligned} L(\varphi) &= f(y|\varphi) = \int f(y|W)f(W|\varphi)dW \\ &\propto \int \frac{f(y|W)f(W|\varphi)}{\tilde{f}(y)}dW = \int \frac{f(y|W)f(W|\varphi)}{f(y|W)\tilde{f}(W)}\tilde{f}(W|y)dW \\ &= \tilde{\mathbb{E}} \left[ \frac{f(W|\varphi)}{\tilde{f}(W)} \middle| y \right] \end{aligned}$$

 $\tilde{f}(W)$  ist eine Vorschlagsdichte für  $W$  auf  $\mathbb{R}^n$ .

$$\tilde{f}(y) = \int f(y|W)\tilde{f}(W)dW$$

(i) Monte Carlo Approximation

$$L_M(\varphi) = \frac{1}{m} \sum_{j=1}^m \frac{f(W(j)|\varphi)}{\tilde{f}(W(j))}$$

wobei  $W(j)$  Ziehungen aus  $\tilde{f}(\cdot|y)$  sind. Wahl von  $\tilde{f}$  kritisch, Vorschlag  $\tilde{f} = f(\cdot|\varphi_0)$ , einige Iterationen für neues  $\varphi$ .

(ii) Maximiere  $L_M$  über Newton-Raphson iterativ

$$\hat{\beta}(W(j)) = (X' C^{-1} X)^{-1} C^{-1} W(j)$$

Varianz der Kovarianzstruktur

$$\hat{\sigma}^2(W(j)) = \frac{1}{n} (W(j) - X \hat{\beta}(W(j)))' C^{-1} (W(j) - X \hat{\beta}(W(j)))$$

mit

$$C = R(\theta) + \tau^2 \mathbb{I}_n$$

(iii)  $\hat{\beta}$  und  $\hat{\sigma}^2$  Plugin einsetzen in  $L_M$ . Maximiere  $\tilde{L}_M(\theta) = L_M(\hat{\beta}, \hat{\sigma}^2, \theta)$ . Maximiere numerisch ( $C^{-1}$  kritisch).