

Universität Ulm, Fakultät für Mathematik und Wirtschaftswissenschaften
Abteilung Stochastik

Bayessche Ansätze in der Geostatistik



im Rahmen des Seminars

Bayessche Ansätze in der Bildanalyse

im Sommersemester 2006

bei Prof.Dr.V. Schmidt
Jun.-Prof.Dr.E. Spodarev

Vorgelegt von
Bernhard Clasen

interne Betreuung durch Stefanie Eckel

Inhaltsverzeichnis

- 1 Einleitung
 - 1.1 Übersicht
 - 1.2 Vorgehensweise

- 2 Einführung und Motivation
 - 2.1 Einführung in des geostatistische Modell
 - 2.2 Zwei einführende Beispiele

- 3 Anwendungen des Bayesschen Paradigmas bei räumlichen Daten
 - 3.1 Modelle stationärer, räumlicher Prozesse
 - 3.1.1 Grundlegendes Modell
 - 3.1.2 Isotropes Modell

- 4 Modellbasierte Geostatistik
 - 4.1 Allgemeines geostatistisches Modell
 - 4.2 Das Gaußsche Modell
 - 4.2.1 Prognosen mit Hilfe des Gaußschen Modells
 - 4.3 Parameter-Schätzung der Kovarianzstruktur
 - 4.3.1 Variogramm-Analyse
 - 4.3.2 Maximum-Likelihood Schätzung
 - 4.4 Prognose durch Einsetzen
 - 4.4.1 Gaußsches Modell
 - 4.4.2 Das transformierte Gaußsche Modell
 - 4.5 Bayessche Folgerungen für das lineare Gauß-Modell
 - 4.5.1 Feste Korrelationsparameter
 - 4.5.2 Unsicherheiten in allen Korrelationsparametern

- 5 Fallstudie: Schweizer Regenfalldaten

Bayessche Ansätze in der Geostatistik

1 Einleitung

1.1 Übersicht

Die Geostatistik hat gerade in den vergangenen Jahren stark an Bedeutung gewonnen und in unterschiedlichsten Bereichen Anwendung gefunden.

Ihr zentraler Aspekt ist die Analyse zeitlich-räumlicher Daten. Ziel dabei ist es, aus Punktmessungen Aussagen über Flächen oder gar Volumina machen zu können. Verschiedene Schätz- und Prognosemethoden sind hierbei hilfreiche Werkzeuge zur Problemlösung. Im Bayesschen Ansatz wird formal nicht zwischen Prognose- und Schätzproblemen unterschieden.

1.2 Vorgehensweise

In Kapitel 2 wird zunächst eine kurze Einführung in den Begriff „Geostatistik“ geben. Dazu ist es zweckmäßig, das geostatistische Modell grob zu umreißen und anschließend mit zwei einführenden Beispielen Probleme in diesem Bereich näher zu bringen.

Um eine theoretische Grundlage für weitere Analysen zu schaffen werden in Kapitel 3 Anwendungen der Bayesschen Paradigmas bei räumlichen Daten vorgestellt.

Kapitel 4 befasst sich daraufhin mit konkreten Methoden der modell-basierten Geostatistik. Hierbei wird das Gauß-Modell als Basis für Prognose- und Schätzmethoden fungieren. Den letzten Abschnitt (Kapitel 5) bildet die Fallstudie „Schweizer Regenfalldaten“. Hiermit soll ein umfassendes Beispiel für ein geostatistisches Problem gegeben und die zuvor betrachteten Methoden bezüglich ihrer praktischen Anwendung diskutiert werden.

2 Einführung und Motivation

2.1 Einführung in das geostatistische Modell

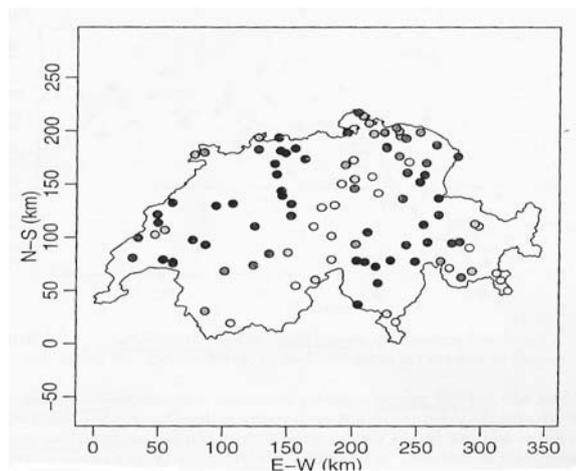
Die Geostatistik umfasst den Bereich der räumlichen Statistik, der sich mit stetiger, räumlicher Variation beschäftigt. Im Mittelpunkt steht dabei ein räumliches Phänomen $s(x)$, welches in einem stetigen räumlichen Bereich $A \subset \mathbb{R}^2$ existiert. Dieses kann als eine Realisierung eines stochastischen Prozesses $S(\bullet) = \{S(x) : x \in A\}$ gesehen werden. Dieser ist generell nicht direkt beobachtbar. Die Daten setzen sich aus Messungen Y_1, \dots, Y_n an Orten x_1, \dots, x_n zusammen, wobei Y_i als „verrauschte“ Version von $S(x_i)$ gesehen werden kann. Weiterhin wird vorausgesetzt, dass das sog. „sampling design“ für x_1, \dots, x_n deterministisch oder stochastisch ist, jedoch unabhängig vom Prozess $S(\bullet)$.¹

2.2 Zwei einführende Beispiele

Schweizer Regenfalldaten

¹ Vgl. Möller, J., (2003), S.43

Dieser Datensatz wurde häufig für empirische Vergleiche von verschiedenen Methoden der räumliche Interpolation verwendet. Als Ziel dieser Untersuchung stand die Konstruktion einer räumlichen Karte von Niederschlagswerten in der Schweiz. Die Daten bestehen aus Messungen des Regens am 8. Mai 1986 an 467 Orten in der Schweiz. 100 Datenpunkte wurden dazu benützt, ein Modell für Prognosen zu erstellen. Die restlichen 367 Punkte dienten dazu, die prognostizierten Werte empirisch zu überprüfen.²

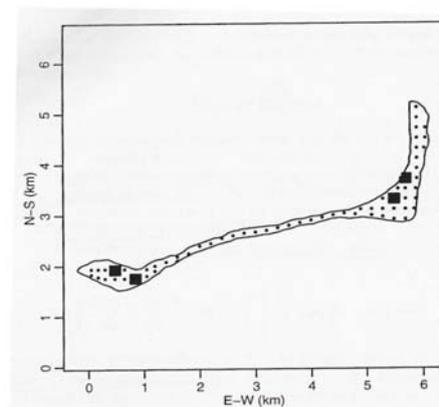


(Quelle: Möller, J., (2003), S.70)

Abbildung 2.1. Schweizer Regenfalldaten an den Orten x_1, \dots, x_{100} . Die Grauskalierung von weiß (kleine Werte) bis schwarz (hohe Werte) gibt die Menge des gefallenen Regens, y_1, \dots, y_{100} , wieder.

Residuale Kontamination auf der Rongelap Insel

Diese Daten beruhen auf Messungen der residualen Kontamination auf der Rongelap Insel im Pazifik, in dessen Gebiet die USA in 50er Jahren Atomtests durchführten. Ziel war es hierbei, eine Entscheidung bezüglich einer Wiederbesiedelung der Insel treffen zu können. Dazu wurden verrauschte Messungen Y_i der Cäsiumkonzentration über einem Netz von Messorten x_i in 200 m Abständen gemacht. Später wurde in Quadraten in 40 m Abständen gemessen. Abbildung 2.2 zeigt eine Karte des „sampling design“.³



(Quelle: Möller, J., (2003), S.46)

² Vgl. Möller, J., (2003), S.44

³ Vgl. Möller, J., (2003), S.44/45

3 Anwendungen des Bayesschen Paradigmas bei räumlichen Daten

3.1 Modelle stationärer, räumlicher Prozesse

3.1.1 Grundlegendes Modell

Das Modell, mit welchem in diesem Abschnitt gearbeitet wird, ist

$$Y(x) = \mu(x) + \omega(x) + \varepsilon(x),$$

wobei die $\mu(x) = x^T(x)\beta$ Struktur des Mittelwertes ist. Die Residuen setzen sich aus zwei Teilen zusammen, einem räumlichen und einem nicht-räumlichen. Die räumliche Komponente wird durch $\omega(x)$ dargestellt, der Realisierung eines null-zentrierten stationären Gaußprozesses. Daneben wird $\varepsilon(x)$ als unkorrelierter, reiner Fehlerterm gesehen, der den nicht-räumlichen Teil ausmacht. Während $\omega(x)$ die Parameter σ^2 (Varianz des Gaußprozesses, sog. „Schwelle“) und ϕ (Reichweite der Messungen) einführt, kommt durch $\varepsilon(x)$ dessen Varianz τ^2 als neuer Modellparameter hinzu.⁴

3.1.2 Isotropes Modell

Zunächst sei bemerkt, dass die Spezifizierung der Korrelationsfunktion als Funktion einer Separation zwischen zwei Seiten ein stationäres Modell ergibt. Geht man einen Schritt weiter und spezifiziert die Korrelationsfunktion nur durch den Abstand $\|x_i - x_j\|$, so erhält man *Isotropie*. Verschiedene Arten der Korrelationsfunktion werden in späteren Abschnitten besprochen.⁵

Betrachtet werden nun die Daten $Y_i(x), i = 1, \dots, n$, und sei $Y = (Y(x_1), \dots, Y(x_n))$.

Y sei multivariat-normalverteilt, d.h. $Y \sim N(X\beta, \Sigma)$, wobei $\Sigma = \sigma^2 H(\phi) + \tau^2 I$ die Kovarianzmatrix sei. Das zentrale Problem bei einer Bayesschen Analyse hierbei ist, eine geeignete Definition für Σ zu finden.

Sammelt man alle Modellparameter in einem Vektor $\theta = (\beta, \sigma^2, \tau^2, \phi)$, so benötigt eine Bayessche Lösung eine passende a priori Verteilung $p(\theta)$.

Parameterschätzungen können dann aus der a posteriori Verteilung

$$p(\theta) \propto f(y | \theta) p(\theta),$$

wobei

⁴ Vgl. Carlin, B.P., Banerjee, S., (2004), S.129

⁵ Vgl. Carlin, B.P., Banerjee, S., (2004), S.130

$$Y | \theta \sim N(X\beta, \sigma^2 H(\phi) + \tau^2 I),$$

bestimmt werden. Für die verschiedenen Parameter werden im Allgemeinen unabhängige a priori Verteilungen gewählt:

$$p(\theta) = p(\beta) p(\sigma^2) p(\tau^2) p(\phi)$$

Als nützlich haben sich die multivariate Normalverteilung für β sowie die inverse Gammaverteilung für σ^2 und τ^2 erwiesen. Für ϕ wird oft eine Gammaverteilung verwendet.⁶

Häufig sind Aussagen über den einzelnen Parameter von Interesse. Ein Punktschätzer oder ein denkbare Intervall für β z.B. ergibt sich aus

$$\begin{aligned} p(\beta | y) &= \iiint p(\beta, \sigma^2, \tau^2, \phi | y) d\sigma^2 d\tau^2 d\phi \\ &\propto p(\beta) \iiint f(y | \theta) p(\sigma^2) p(\tau^2) p(\phi) d\sigma^2 d\tau^2 d\phi \end{aligned} \quad 7$$

Um nun eine Prognoseaussage für Y an einem neuen Punkt x_0 zu erhalten, betrachten wir als nächsten Schritt das sog. Bayessche „Kriging“. Mit den Notationen $Y_0 \equiv Y(x_0)$ und $\hat{x}_0 \equiv x(x_0)$ ergibt sich eine Lösung für den unbekanntem Wert $Y(x_0)$ dadurch, die Prognoseverteilung

$$\begin{aligned} p(y_0 | y, X, \hat{x}_0) &= \int p(y_0, \theta | y, X, \hat{x}_0) d\theta \\ &= \int p(y_0 | y, \theta, \hat{x}_0) p(\theta | y, X) d\theta \end{aligned}$$

zu finden, wobei $p(y_0 | y, X, \hat{x}_0)$ eine Normalverteilung hat, welche sich aus der gemeinsamen Verteilung von Y_0 und den Daten Y ergibt.⁸

⁶ Vgl. Carlin, B.P., Banerjee, S., (2004), S.131

⁷ Vgl. Carlin, B.P., Banerjee, S., (2004), S.131

⁸ Vgl. Carlin, B.P., Banerjee, S., (2004), S.132

4 Modellbasierte Geostatistik

4.1 Allgemeines geostatistisches Modell

Zunächst sollen die folgenden Notationen für das Modell festgelegt werden:

- (x_i, y_i) : Daten, mit x_1, \dots, x_n als Messungsorte und y_1, \dots, y_n als Messungen, y_i Realisierungen von $Y_i = Y(x_i)$
- $\{x_i : i = 1, \dots, n\}$: „*sampling design*“ ($x \in A$, wobei A die Studienregion sei)
- $Y(\bullet) = \{Y(x) : x \in A\}$: *Messungsprozess*
- $S(\bullet) = \{S(x) : x \in A\}$: *Signalprozess* (unbeobachtbar)
- $T = T(S(\bullet))$: *Ziel der Prognose*

In der Praxis sind Y_1, \dots, Y_n „verrauschte“ Versionen von $S(x_1), \dots, S(x_n)$.⁹

Ein geostatistisches Modell ist eine Spezifizierung einer gemeinsamen Verteilung des Messungsprozesses und des Signalprozesses in der Form

$$[S(\bullet), Y(\bullet)] = [Y(\bullet) | S(\bullet)][S(\bullet)],$$

wobei $[\bullet]$ „die Verteilung von“ meint. Ein Prädiktor (Prognosewert) für T ist jede Funktion $\hat{T} = \hat{T}(Y)$, wobei $Y = (Y_1, \dots, Y_n)^T$. Der Prädiktor bzgl. des minimalen quadratischen Fehlers im Mittelwert minimiert $\text{MSE}(\hat{T}) = E[(T - \hat{T})^2]$. Folgende Aussage kann gemacht werden.¹⁰

Proposition 1

Vorausgesetzt sei, dass $\text{Var}[T] < \infty$; dann ist der MSE-Prädiktor von T $\hat{T} = E_T[T | Y]$, mit dem zugehörigen quadratischen Prognosefehler im Mittelwert $E[(T - \hat{T})^2] = E_Y \text{Var}_T[T | Y]$.¹¹

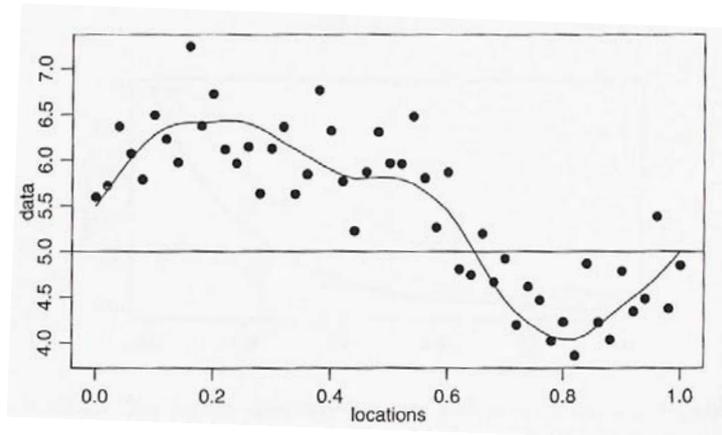
Für Punktprognosen verwendet man oft $E[T | y]$, den MSE-Prädiktor ausgewertet an den beobachteten y . Als Schätzer für den erhaltenen Fehler wird die sog. *Prognosevarianz* $\text{Var}[T | y]$ verwendet. Die komplette Lösung für ein Prognoseproblem sollte jedoch in einer Wahrscheinlichkeitsverteilung $[T | y]$ ausgedrückt werden, der sog. *Prognoseverteilung*. Diese fällt unter Annahme des Bayesschen Paradigmas mit der a posteriori Verteilung von T zusammen.¹²

⁹ Vgl. Möller, J., (2003), S.45/46

¹⁰ Vgl. Möller, J., (2003), S.46

¹¹ Vgl. Möller, J., (2003), S.46

¹² Vgl. Möller, J., (2003), S.47



(Quelle: Möller, J., (2003), S.70)

Abbildung 4.1 Simulation eines Gaußschen Modells; Daten Y_1, \dots, Y_n (Punkte), das Signal $S(\cdot)$ (durchgehende Kurve) und der Mittelwert μ (horizontale Linie)

4.2 Das Gaußsche Modell

In seiner grundlegenden Form ist das gaußsche Modell dadurch charakterisiert, dass $S(\cdot)$ ein stationärer Gaußprozess ist mit $E[S(x)] = \mu$, $\text{Var}[S(x)] = \sigma^2$ und Korrelationsfunktion $\rho(u) = \text{Corr}[S(x), S(x')]$, wobei $u = \|x - x'\|$ der euklidischen Abstand zwischen x und x' sei. Die Verteilung von Y_i ist eine Normalverteilung mit Mittelwert $S(x_i)$ und Varianz τ^2 .¹³

Eine äquivalente Formulierung hierfür wäre, dass

$$Y_i = S(x_i) + Z_i : i = 1, \dots, n,$$

wobei die Z_1, \dots, Z_n unabhängig und identische verteilt seien mit $Z_i \sim N(0, \tau^2)$. Die Verteilung von Y sei eine multivariate Gaußverteilung,

$$Y \sim N(\mu \mathbf{1}, \sigma^2 \mathbf{R} + \tau^2 \mathbf{I}),$$

wobei $\mathbf{1}$ einen Vektor mit n Einsen darstellt, \mathbf{I} die $n \times n$ Einheitsmatrix und \mathbf{R} die Einträge $R_{ij} = \rho(u_{ij})$ hat, wobei $u_{ij} = \|x_i - x_j\|$ ist.

Die Spezifizierung der Korrelationsfunktion legt die Glätte des entstehenden Prozesses $S(\cdot)$ fest. Mathematisch ist die Glätte eines Prozesses sein Grad an Differenzierbarkeit. Ein Prozess wird als stetig bzgl. des quadratischen Mittelwerts bezeichnet, falls für alle x $E[\{S(x) - S(x')\}^2] \rightarrow 0$, wenn $\|x - x'\| \rightarrow 0$. Analog wird ein Prozess als differenzierbar bzgl.

des quadratische Mittelwerts bezeichnet, falls $E[\{\frac{S(x) - S(x')}{\|x - x'\|} - S'(x)\}^2] \rightarrow 0$ wenn

$\|x - x'\| \rightarrow 0$. Ein Zusammenhang zwischen Prozess und Korrelationsfunktion offenbart sich in folgender Aussage.

Proposition 2

¹³ Vgl. Möller, J., (2003), S.47

Sei $S(\bullet)$ ein stationärer Gaußscher Prozess mit Korrelationsfunktion $\rho(u) : u \in \mathbb{R}$. Dann gilt:

- $S(\bullet)$ ist stetig bzgl. des quadratischen Mittelwertes $\Leftrightarrow \rho(u)$ ist stetig in $u=0$
- $S(\bullet)$ ist k -mal differenzierbar bzgl. des quadratischen Mittelwertes
 $\Leftrightarrow \rho(u)$ ist mindestens $2k$ -mal differenzierbar in $u=0$ ¹⁴

In diesem Zusammenhang sollen nun einige Familien von Korrelationsfunktionen vorgestellt werden. Als eine attraktive Variante gilt die *Matérn* Familie. Ihre algebraische Form ist gegeben durch

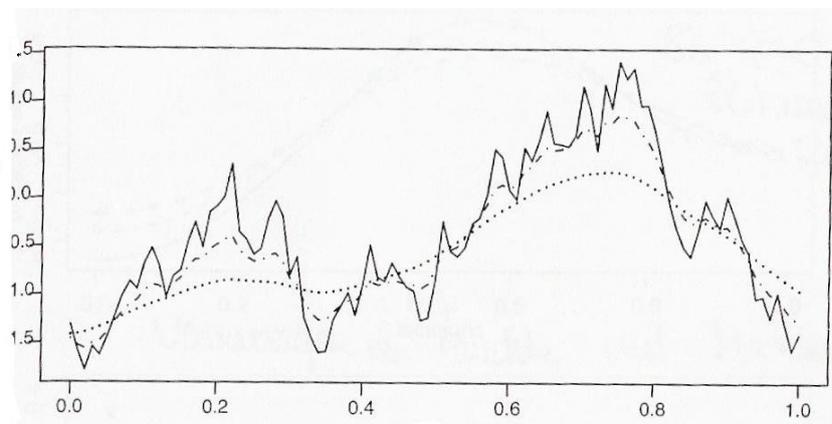
$$\rho(u) = \{2^{k-1} \Gamma(k)\}^{-1} (\frac{u}{\phi})^k K_k(\frac{u}{\phi}),$$

wobei $k > 0$ und $\phi > 0$ Parameter sind und $K_k(\bullet)$ eine Besselfunktion der Ordnung k ist. Die Attraktivität dieser Familie ergibt sich aus der Eigenschaft, dass der Prozess $S(\bullet)$ $[k-1]$ -mal differenzierbar ist, wobei $[k]$ die größte natürliche Zahl kleiner oder gleich k seien soll. ¹⁵

Eine andere Familie beinhaltet die Exponentialfamilie,

$$\rho(u) = \exp\{-\frac{|u|}{\phi}\},$$

definiert für $\phi > 0$ und $0 < k \leq 1$. Der zugehörige Prozess $S(\bullet)$ ist zwar stetig für $k < 2$, allerdings nicht differenzierbar. ¹⁶



(Quelle: Möller, J., (2003), S.51)

Abbildung 4.2 Simulationen eines Gaußschen Prozesses mit Matérn Korrelationsfunktion; $\phi = 0.2$ und $k=0.5$ (durchgehende Linie), $k=1$ (gestrichelte Linie) und $k=2$ (gepunktete Linie).

In der Geostatistik häufig verwendet ist die Sphärische Familie. Sie zeichnet sich durch die Form

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

aus, wobei hier $\phi > 0$ einziger Parameter ist. Mit nur einem Parameter fehlt ihr die Flexibilität der Matérn Klasse. ¹⁷

¹⁴ Vgl. Möller, J., (2003), S.48

¹⁵ Vgl. Möller, J., (2003), S.48

¹⁶ Vgl. Möller, J., (2003), S.49

4.2.1 Prognosen mit Hilfe des Gaußschen Modells

Das Ziel der Prognose ist wie bisher $T = S(x_0)$, der Wert des Signalprozesses am Punkt x_0 , wobei x_0 nicht unbedingt im „sampling design“ enthalten sein muss. Im Gaußschen Modell ist $[T, Y]$ multivariat-gaußverteilt, die Prognosevarianz $\text{Var}[T | Y]$ und die Prognoseverteilung $[T | Y]$ können mit Hilfe folgender Aussage bestimmt werden, wobei $\hat{T} = E[T | Y]$.

Proposition 3

Sei $X = (X_1, X_2)$ multivariat-gaußverteilt mit Mittelwertvektor $\mu = (\mu_1, \mu_2)$ und Kovarianzmatrix

$$\Sigma = \begin{vmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{vmatrix}.$$

Dann ist die bedingte Verteilung von X_1 bei gegebenem $X_2 = x_2$ ebenfalls eine multivariate Gaußverteilung mit $X_1 | X_2 = x_2 \sim N(\mu_{1|2}, \Sigma_{1|2})$, wobei

$$\mu_{1|2} = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2)$$

und

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}.^{18}$$

Im geostatistischen Modell ist $[T, Y]$ multivariat-gaußverteilt mit Mittelwertvektor $\mu 1$ und Kovarianzmatrix

$$\begin{vmatrix} \sigma^2 & \sigma^2 r^T \\ \sigma^2 r & \tau^2 I + \sigma^2 R \end{vmatrix},$$

wobei r ein Vektor mit Einträgen $r_i = \rho(\|x_0 - x_i\|) : i = 1, \dots, n$ ist. Wendet man nun Proposition 3 für $X_1 = T, X_2 = Y$ hier an, so ergibt sich ein MSE-Prädiktor für T in Form von

$$\hat{T} = \mu + \sigma^2 r^T (\tau^2 I + \sigma^2 R)^{-1} (y - \mu 1) \quad (4.1)$$

mit einer Prognosevarianz

$$\text{Var}[T | y] = \sigma^2 - \sigma^2 r^T (\tau^2 I + \sigma^2 R)^{-1} \sigma^2 r.$$

Die Konstruktion der Fläche $\hat{S}(\bullet)$, wobei hier für jeden Punkt x_0 $\hat{T} = \hat{S}(x_0)$ durch (4.1) gegeben ist, nennt man „simple kriging“. Folgende Eigenschaften von $\hat{S}(\bullet)$ sollten dabei beachtet werden:

¹⁷ Vgl. Möller, J., (2003), S.50

¹⁸ Vgl. Möller, J., (2003), S.50/51

- (i) $\hat{S}(\bullet)$ interpoliert (d.h. $\hat{S}(x_i) = y_i$) die Daten genau dann, wenn $\tau^2 = 0$. Denn in diesem Fall ist $Y(x_i) = S(x_i)$ für $i=1, \dots, n$.
- (ii) $\hat{S}(\bullet)$ übernimmt die analytische Glätte der Korrelationsfunktion, z.B. ergibt $k \leq 0.5$ in der Matérn-Familie eine nichtdifferenzierbare Fläche $\hat{S}(\bullet)$, während für $k > 0.5$ $\hat{S}(\bullet)$ eine glatte, differenzierbare Oberfläche ist.
- (iii) Bei Modellen, in denen $\rho(u) \rightarrow 0$, falls $u \rightarrow \infty$, ist $\hat{S}(x_0) \approx \mu$ für Punkte x_0 , die weit genug entfernt von den restlichen Punkten x_i liegen. Befindet sich x_0 nahe bei einem oder mehreren x_i , so wird $\hat{S}(x_0)$ stärker durch die y_i 's der angrenzenden Punkte beeinflusst.¹⁹

4.3 Parameter-Schätzung der Kovarianzstruktur

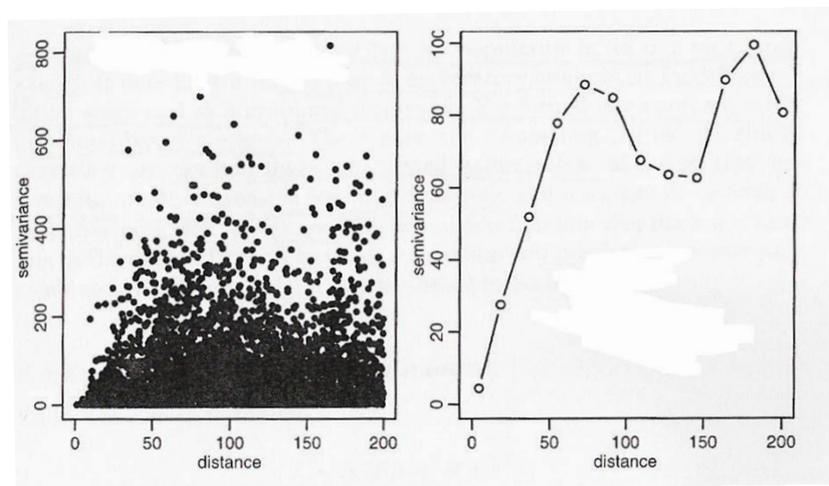
4.3.1 Variogramm-Analyse

Einen Überblick über die Struktur der zweiten Momente eines räumlichen stochastischen Prozesses $Y(\bullet)$ gibt das Variogramm. Dabei handelt es sich um eine Funktion der Form

$$V(x, x') = \frac{1}{2} \text{Var}\{Y(x) - Y(x')\}.$$

Im linearen Gaußschen Modell mit $u = \|x - x'\|$ ist

$$V(u) = \tau^2 + \sigma^2 \{1 - \rho(u)\}.$$
²⁰



(Quelle: Möller, J., (2003), S.57)

Abbildung 4.3 Die Variogrammwolke (links) und das empirische Variogramm (rechts) für die Schweizer Regenfalldaten

Einen optischen Überblick über die räumliche Korrelation von Daten bietet die *Variogrammwolke*. Sie ist ein Scatterplot der Punkte (u_{ij}, v_{ij}) , welche aus den Größen

¹⁹ Vgl. Möller, J., (2003), S.55/56

²⁰ Vgl. Möller, J., (2003), S.56

$$u_{ij} = \|x_i - x_j\|$$

$$v_{ij} = \frac{(y_i - y_j)^2}{2}$$

bestimmt werden. Hat der zugrunde liegende Prozess einen räumlich variierenden Mittelwert $\mu(x)$, so bietet die Variogrammwolke keinen sinnvollen Überblick.

Eine stabilere Variante der Variogrammwolke ist das *empirische Variogramm* $\bar{V}(\bullet)$. Man bestimmt $\bar{V}(\bullet)$ aus dem Durchschnitt über solche v_{ij} 's, für die $|u - u_{ij}| < h/2$, wobei u eine Separationsdistanz darstellt und h eine bezüglich der Darstellung gewählte Weite ist.

Eine Art, um Kovarianzparameter zu schätzen, ist das Einfügen einer parametrischen Kovarianzfunktion in das empirische Variogramm. In der Praxis wird dies oft ohne formale Kriterien, sozusagen per Auge, ausgeführt. Alternativ werden aber auch gewöhnliche Methoden oder Methoden der kleinsten Quadrate verwendet. Diese schätzen die Kovarianzparameter θ durch Minimieren von

$$S(\theta) = \sum_k \omega_k [\bar{V}(u_k) - V(u_k; \theta)]^2,^{21}$$

wobei $\omega_k = 1$ für die gewöhnlichen kleinsten Quadrate ist, während für gewichtete Quadrate ω_k die Anzahl der Paare von Messungen ist, welche in $\bar{V}(u_k)$ mit einwirken.

4.3.2 Maximum-Likelihood Schätzung

Im Gaußschen Modell ist

$$Y \sim N(F\beta, \sigma^2 R + \tau^2 I)$$

wobei F die $n \times p$ Matrix der Kovarianzen ist und β der Vektor der Parameter. Die log-likelihood Funktion ist dann

$$l(\beta, \tau, \sigma, \phi, k) \propto -0.5 \{ \log |(\sigma^2 R + \tau^2 I)| + (y - F\beta)^T (\sigma^2 R + \tau^2 I)^{-1} (y - F\beta) \}.$$

Durch Maximieren erhält man hieraus die ML-Schätzer für die Modellparameter.

Zur Berechnung reparametrisiert man zunächst $v^2 = \frac{\tau^2}{\sigma^2}$ und bezeichnet $V = (R + v^2 I)$. So ergibt sich eine Maximierung der log-LH Funktion für

$$\hat{\beta}(V) = (F^T V^{-1} F)^{-1} F^T V^{-1} y$$

und

$$\hat{\sigma}^2(V) = n^{-1} (y - F\hat{\beta})^T V^{-1} (y - F\hat{\beta}).^{22}$$

Substituiert man nun $(\hat{\beta}(V), \hat{\sigma}^2(V))$ in die log-Likelihood Funktion, so erhält man die reduzierte log-Likelihood

²¹ Vgl. Möller, J., (2003), S.57/58

²² Vgl. Möller, J., (2003), S.59

$$l(v^2, \phi, k) \propto -0.5\{n \log |\hat{\sigma}^2(\mathbf{V})| + \log |\mathbf{V}|\}.$$

Diese muss dann numerisch optimiert werden um daraufhin durch Rücksubstitution $\hat{\sigma}^2$ und $\hat{\beta}$ zu erhalten.²³

4.4 Prognose durch Einsetzen

In diesem Abschnitt wird der Ansatz vorgestellt, Prognosen zu erhalten, in dem Schätzer für unbekannt Parameter in das Modell einsetzt, so als wären sie die wahren Werte. Dabei muss beachtet werden, dass die Varianz zwischen den wahren und den geschätzten Werten vernachlässigt wird und dieser Ansatz somit recht optimistisch ist.

4.4.1 Gaußsches Modell

Für das Gaußsche Modell haben wir bereits gesehen, dass der MSE-Prädiktor für $T = S(x_0)$

$$\hat{T} = \mu + \sigma^2 \mathbf{r}^T (\tau^2 \mathbf{I} + \sigma^2 \mathbf{R})^{-1} (\mathbf{y} - \mu \mathbf{1})$$

ist mit einer Prognosevarianz

$$\text{Var}[T | \mathbf{y}] = \sigma^2 - \sigma^2 \mathbf{r}^T (\tau^2 \mathbf{I} + \sigma^2 \mathbf{R})^{-1} \sigma^2 \mathbf{r}.$$

Eine Prognose durch Einsetzen besteht darin, in den obigen Gleichungen die Parameter durch ihre Schätzer zu ersetzen.

4.4.2 Das transformierte Gaußsche Modell

Zusätzliche Flexibilität in ein Gaußsches Modell bringt die Einführung eines Parameters λ , welcher eine Box-Cox Transformation definiert. Das daraus resultierende Modell nimmt an, dass die Daten $\mathbf{y} = (y_1, \dots, y_n)$ durch

$$\tilde{y}_i = h_\lambda(y_i) = \begin{cases} (y_i^\lambda - \lambda) / \lambda & \text{falls } \lambda \neq 0 \\ \log y_i & \text{falls } \lambda = 0 \end{cases}$$

transformiert werden können.²⁴

Für dieses Modell nehmen wir an, dass $Y(x_0)$ das Ziel der Prognose ist und bezeichnen $T_\lambda = h_\lambda(Y(x_0))$. Der MSE-Prädiktor \hat{T}_λ und die zugehörige Prognosevarianz $\text{Var}[T_\lambda | \mathbf{y}]$ können wieder durch Einsetzen von Schätzern für die Mittelwert- und Kovarianzparameter bestimmt werden. Für $\lambda = 0$ erhalten wir

$$\hat{T} = \exp(\hat{T}_0 + 0.5 \text{Var}[T_0 | \mathbf{y}])$$

²³ Vgl. Möller, J., (2003), S.59

²⁴ Vgl. Möller, J., (2003), S.56/57

mit Prognosevarianz

$$\text{Var}[T | y] = \exp(2\hat{T}_0 + \text{Var}[T_0 | y])(\exp(\text{Var}[T_0 | y]) - 1).$$

Für $\lambda > 0$ kann man \hat{T} und $\text{Var}[T | y]$ durch eine Summe von Momenten der Normalverteilung approximieren. Für $\lambda = 0.5$ erhält man z.B.

$$\hat{T} \approx (0.5\hat{T}_{0.5} + 1)^2 + 0.25\text{Var}[T_{0.5} | y]$$

mit Varianz

$$\text{Var}[T | y] \approx (0.5\hat{T}_{0.5} + 1)^4 + 1.5(0.5\hat{T}_{0.5} + 1)^2\text{Var}[T_{0.5} | y] + 3(\text{Var}[T_{0.5} | y])^2 / 16.$$

4.5 Bayessche Folgerungen für das lineare Gauß-Modell

Der Bayessche Ansatz behandelt die Modellparameter als Zufallsvariablen, und somit wird kein formaler Unterschied mehr gemacht zwischen einer Parameterschätzung und einem Prognoseproblem.

4.5.1 Feste Korrelationsparameter

In diesem Ansatz nun nehmen wir zunächst an, dass alle Parameter der Korrelationsfunktion bekannt sind, $\tau^2 = 0$ und eine Unsicherheit nur in den Parametern β und σ^2 vorhanden sei. In diesem Fall kann die Prognoseverteilung analytisch bestimmt werden.

Für festes ϕ ist die a priori Familie für (β, σ^2) die Gauß-Skalierte Inverse χ^2 -Verteilung. So ergeben sich für β und σ^2 die a priori Verteilungen

$$[\beta | \sigma^2, \phi] \sim N(m_b, \sigma^2 V_b) \text{ und } [\sigma^2 | \phi] \sim \chi_{\text{ScI}}^2(n_\sigma, S_\sigma^2),$$

wobei $\chi_{\text{ScI}}^2(n_\sigma, S_\sigma^2)$ eine Verteilung mit einer Dicht der Form

$$\pi(z) \propto z^{-(n_\sigma/2+1)} \exp(-n_\sigma S_\sigma^2 / 2z), z > 0 \text{ ist.}$$

Als Kurzform schreiben wir dies als

$$[\beta, \sigma^2 | \phi] \sim N\chi_{\text{ScI}}^2(m_b, V_b, n_\sigma, S_\sigma^2) \quad (4.2)$$

Unter Verwendung des Bayesschen Theorems ist die a posteriori Verteilung der Parameter

$$[\beta, \sigma^2 | y, \phi] \sim N\chi_{\text{ScI}}^2(\tilde{\beta}, V_{\tilde{\beta}}, n_\sigma + n, S^2), \quad (4.3)$$

wobei $V_{\tilde{\beta}} = (V_b^{-1} + F^T R^{-1} F)^{-1}$, $\tilde{\beta} = V_{\tilde{\beta}}(V_b^{-1} m_b + F^T R^{-1} y)$ und

$$S^2 = \frac{n_\sigma S_\sigma^2 + \mathbf{m}_b^T \mathbf{V}_b^{-1} \mathbf{m}_b + \mathbf{y}^T \mathbf{R}^{-1} \mathbf{y} - \tilde{\boldsymbol{\beta}}^T \mathbf{V}_\beta^{-1} \tilde{\boldsymbol{\beta}}}{n_\sigma + n}. \quad (4.4)$$

4.5.2 Unsicherheiten in allen Korrelationsparametern

Nun soll in allen Modellparametern Unsicherheit sein. Wir betrachten ein Modell ohne Messungsfehler, d.h. $\tau^2 = 0$ und einem einzigen Korrelationsparameter ϕ . Wir nehmen als a priori Verteilung $\pi(\boldsymbol{\beta}, \sigma^2, \phi) = \pi(\boldsymbol{\beta}, \sigma^2 | \phi) \pi(\phi)$ an, ein Produkt von (4.2) und einer passenden Dichte von ϕ .²⁶

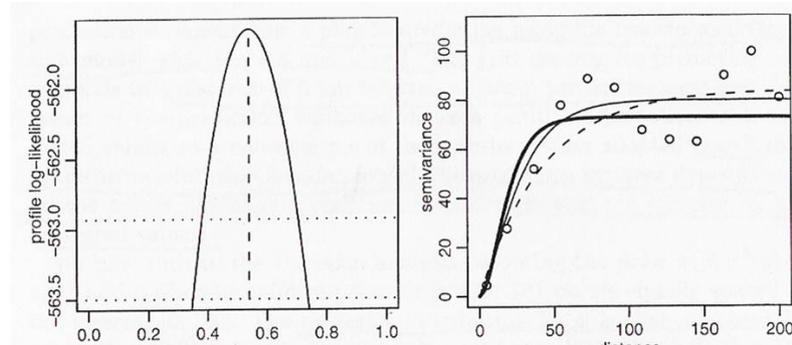
Die a posteriori Verteilung für die Parameter ist dann gegeben durch

$$p(\boldsymbol{\beta}, \sigma^2, \phi | \mathbf{y}) = p(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \phi) p(\phi | \mathbf{y}),$$

wobei $[p(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \phi)]$ durch (4.3) gegeben ist und

$$p(\phi | \mathbf{y}) \propto \pi(\phi) | \mathbf{V}_\beta |^{-\frac{1}{2}} | \mathbf{R} |^{-\frac{1}{2}} (S^2)^{-\frac{n_\sigma + n}{2}}, \quad (4.5)$$

wo \mathbf{V}_β und S^2 durch (4.3) und (4.4) gegeben sind.



(Quelle: Möller, J., (2003), S.66)

Abbildung 5.1 links: Verhalten der log-LH für den Transformationsparameter λ

Rechts: geschätzte Variogramme (Kreise) für die transformierten Daten ($\lambda = 0.5$) verglichen mit den theoretischen Matérn Modell mit gleichen Parametern wie die MLH Schätzer. Matérn mit $k=0.5$ (gestrichelte Linie), $k=1$ (dicke Linie) und $k=2$ (dünne Linie)

5 Fallstudie: Schweizer Regenfalldaten

In dieser Studie verwenden wir 100 Daten-Orte der Schweizer Regenfalldaten für die Modellsimulation. Es wird ein transformiertes Gaußsches Modell verwendet mit Matérn Korrelationsfunktion.

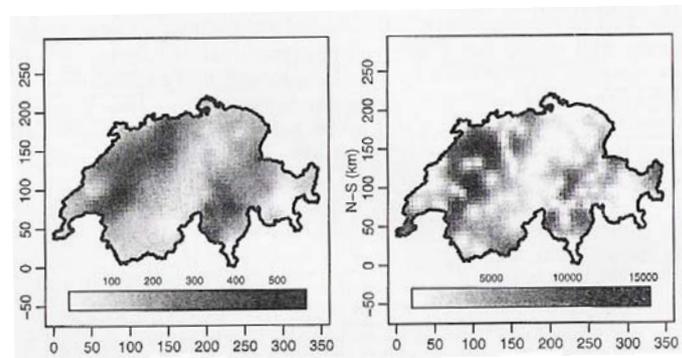
²⁵ Vgl. Möller, J., (2003), S.63

²⁶ Vgl. Möller, J., (2003), S.64

Abbildung 5.1 zeigt im linken Bild, dass weder die log-Transformation $\lambda = 0$ noch ein nicht-transformiertes Gauß-Modell ($\lambda = 1$) für die Daten haltbar ist. Das rechte Bild zeigt das empirische Variogramm mit eingefügter Korrelationsfunktion

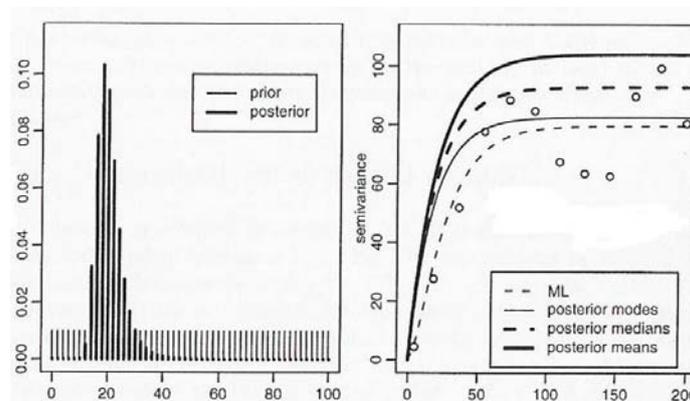
In Abbildung 5.2 sieht man die Punktprognosen der Regenfallwerte und zugehörigen Prognosevarianzen, welche mit Hilfe einer Prognose durch Einsetzen unter Verwendung des transformierten Gaußschen Modells mit $\lambda = 0.5$ und $k=1$ erzeugt wurden. Die Werte der Prognosevarianz sind positiv korreliert mit den Prognosewerten, was eine Konsequenz der Transformation ist.²⁷

Als nächstes führen wir eine kurze Bayessche Analyse durch. Dazu nehmen wir an, dass die a priori $\pi(\beta, \sigma^2 | \phi) \propto 1/\sigma^2$ ist und eine diskrete a priori Verteilung für ϕ vorliegt mit 101 Punkten, welche sich gleichmäßig im Intervall $[0;100]$ verteilen. Die a posteriori Verteilung für ϕ ergibt sich aus der Berechnung von (4.5) für jeden diskreten Wert. Das linke Bild in Abbildung 5.3 zeigt die angenommene a priori und die berechnete a posteriori, während das rechte Bild Variogramme zeigt, die auf verschiedenen Zusammenfassungen der a posteriori Verteilung $[\sigma^2, \phi | y]$ und den ML-Schätzern ($\hat{\sigma}^2, \hat{\phi}$) basieren.²⁸



(Quelle: Möller, J., (2003), S.68)

Abbildung 5.2 Die Prognosen (links) und die Prognosevarianzen (rechts) der Schweizer Regenfalldaten.



(Quelle: Möller, J., (2003), S.69)

Abbildung 5.3 links: Angenommene a priori Verteilung und zugehörige a posteriori Verteilung. Rechts: Variogramme basierend auf Zusammenfassungen der a posteriori und dem ML-Schätzer.

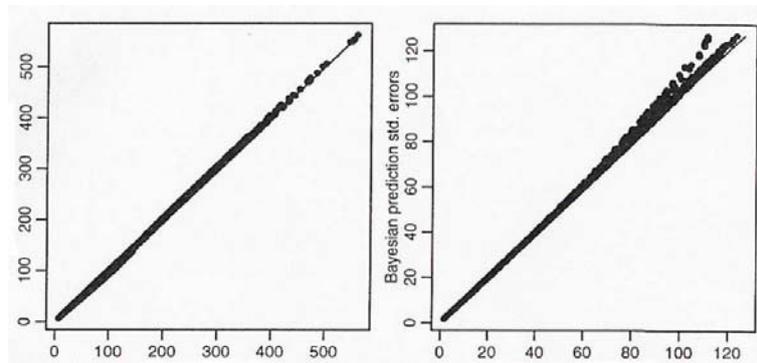
Die starke Konzentration von Punkten entlang der Diagonalen im linken Bild von Abbildung 5.4 zeigt, dass die Bayesschen Punktprognosen sich nicht von den Prognosen unterscheiden, die durch Einsetzen bestimmt wurden. Das rechte Bild zeigt, dass die Varianz beim Einsetzen

²⁷ Vgl. Möller, J., (2003), S.67

²⁸ Vgl. Möller, J., (2003), S.68

die Varianz der Prognoseverteilung leicht unterschätzt.²⁹

Die Daten in diesem Beispiel zeichnen sich durch ihr außergewöhnlich gutes Verhalten aus. In der Praxis findet man meist stärker verrauschte Daten mit größeren Unsicherheiten.



(Quelle: Möller, J., (2003), S.70)

Abbildung 5.4 Vergleich der Prognose durch Einsetzten mit den Bayesschen Prognosewerten (links) und den zugehörigen Standardfehlern (rechts)

²⁹ Vgl. Möller, J., (2003), S.70

Literaturverzeichnis

J. Moeller, Ed. *Spatial Statistics and Computational Methods*. Springer, LNS 173. 2003

S. Banerjee, B. P. Carlin. *Hierarchical Modelling and Analysis for Spatial Data*. Chapman & Hall, 2004.