

Bedingte und unbedingte Fehler bei geostatistischen Vorhersagen  
– forstwissenschaftliche Fallstudien –



Dissertation  
zur Erlangung des Doktorgrades der Fakultät für  
Forstwissenschaften und Waldökologie der  
Georg-August-Universität Göttingen

**Bedingte und unbedingte Fehler bei  
geostatistischen Vorhersagen  
– forstwissenschaftliche Fallstudien –**

vorgelegt von  
– Andreas Dominik Cullmann –  
geboren 1975

Göttingen, 2006

**Cullmann, Andreas Dominik:**

Bedingte und unbedingte Fehler bei geostatistischen Vorhersagen  
– forstwissenschaftliche Fallstudien –

Erster Berichterstatter: Prof. Dr. Joachim Saborowski

Zweiter Berichterstatter: Prof. Dr. Christoph Kleinn

Dritter Berichterstatter: Prof. Dr. Martin Schlather

Tag der mündlichen Prüfung: 16. März 2007

Dies ist ein Ausdruck des Inhaltes unter

<http://resolver.sub.uni-goettingen.de/purl/?webdoc-1441>.

Dieser Inhalt ist unter einem Creative Commons Namensnennung 2.0 Deutschland Lizenzvertrag lizenziert. Um die Lizenz anzusehen, gehen Sie bitte zu <http://creativecommons.org/licenses/by/2.0/de/> oder schicken Sie einen Brief an Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

There is no probability in itself. There are only probabilistic models.  
The only question that really matters, in each particular case, is  
whether this or that probabilistic model, in relation to this or that real  
phenomenon, has or has not an objective meaning.

---

*(Georges Matheron in Matheron (1989))*



# Inhaltsverzeichnis

<b>Tabellenverzeichnis</b>	<b>ix</b>
<b>Abbildungsverzeichnis</b>	<b>xi</b>
<b>1. Einleitung</b>	<b>1</b>
<b>2. Zielführende Darstellung geostatistischer Methoden</b>	<b>9</b>
2.1. Räumliche Zufallsprozesse . . . . .	9
2.1.1. Schätzung der räumlichen Abhängigkeit . . . . .	11
2.1.2. Erwartungswertschätzung . . . . .	13
2.2. Räumliche Vorhersagen . . . . .	14
2.2.1. Lineare Prädiktoren . . . . .	15
2.2.2. Genauigkeit linearer Prädiktoren . . . . .	17
2.2.3. Blockvorhersagen . . . . .	18
2.2.4. Nicht-lineare Prädiktoren . . . . .	20
2.2.5. Prognosegüte . . . . .	23
2.3. Bedingte Verteilungen . . . . .	24
2.3.1. Anwendungen bedingter lokaler Verteilungsfunktionen	24
2.3.2. Approximation bedingter lokaler Verteilungsfunktionen	25
2.4. Bedingte Simulationen . . . . .	27
<b>3. Bedingte Vorhersagefehler in gauß'schen Prozessen</b>	<b>33</b>
3.1. Herleitung . . . . .	33
3.2. Fallstudie: Prognoseintervalle . . . . .	38
<b>4. Bedingte Verteilungen in der Anwendung</b>	<b>47</b>
4.1. Fallstudie: Magnesium in Baden-Württemberg . . . . .	48
4.2. Fallstudie: Simulierte Daten . . . . .	54
<b>5. Diskussion</b>	<b>69</b>
5.1. Bedingte Fehler und Konfidenzen . . . . .	69
5.2. Approximation bedingter lokaler Verteilungsfunktionen . . . . .	71
<b>6. Zusammenfassung</b>	<b>73</b>
<b>Danksagung</b>	<b>77</b>
<b>A. Elektronischer Briefverkehr</b>	<b>79</b>

<b>B. Umformungen</b>	<b>83</b>
B.1. Transponieren inverser Kovarianzmatrizen . . . . .	83
B.2. Zu Kapitel 2.1.2 . . . . .	83
B.3. Zu Kapitel 2.2.1 . . . . .	83
B.4. Zu Kapitel 2.2.2 . . . . .	83
B.5. Zu Kapitel 3.1 . . . . .	84
<b>Literaturverzeichnis</b>	<b>85</b>
<b>Lebenslauf</b>	<b>89</b>

## Tabellenverzeichnis

1.1. Bedingte und unbedingte Wahrscheinlichkeiten. . . . .	6
3.1. Deskriptive Statistiken: Magnesium. . . . .	40
3.2. Extrema und Lage, Fehler. . . . .	43
4.1. Deskriptive Statistiken: Magnesium. . . . .	49
4.2. Hermitesche Koeffizienten. . . . .	51
4.3. Semivariogramme, Magnesium. . . . .	51
4.4. Deskriptive Statistiken für $\psi_{(578)}^1$ , $\psi_{(578)}^2$ , $z_{(578)}^1$ und $z_{(578)}^2$ . . . . .	57
4.5. Semivariogramme, simulierte Daten. . . . .	60
4.6. Bedingte Momente einiger Variabler und $\hat{\mu}$ . . . . .	62
4.7. Maße der Prognosegüte, $\psi_{(578)}^1$ und $\psi_{(578)}^2$ . . . . .	67



# Abbildungsverzeichnis

1.1. Stichprobenwerte. . . . .	1
1.2. Interpolation einer räumlichen Variablen. . . . .	2
1.3. Meßstellenanordnung mit unterschiedlichen Meßreihen. . . . .	4
1.4. Bedingte und unbedingte Realisation. . . . .	5
2.1. Gauß'sche Autokorrelation. . . . .	12
2.2. Approximation einer bedingten lokalen Verteilungsfunktion. . . . .	26
2.3. Überdeckungswahrscheinlichkeit. . . . .	28
2.4. Meßwerte, Prognosen und partielle Realisation. . . . .	29
2.5. Höhenprofile. . . . .	29
3.1. Gewicht des Erwartungswertes. . . . .	37
3.2. Magnesiumgehalte in Fichtennadeln. . . . .	39
3.3. Verteilungsfunktion Magnesium. . . . .	40
3.4. Semivariogramm, Magnesiumgehalt. . . . .	41
3.5. Unbedingte Fehler, Magnesium. . . . .	43
3.6. $R_{duo}$ , $GE$ , Magnesium. . . . .	44
3.7. Relative Differenzen $R_{duc}$ , Magnesium. . . . .	45
4.1. Aufteilung der Meßstellen. . . . .	48
4.2. Gauß'sche Anamorphosen, Magnesium. . . . .	50
4.3. Semivariogramme, Magnesium. . . . .	50
4.4. Maße der Prognosegüte für $z^1$ und $z^2$ . . . . .	53
4.5. Überdeckungswahrscheinlichkeiten für $z^1$ und $z^2$ . . . . .	55
4.6. Gauß'sche Anamorphosen, simulierte Daten. . . . .	58
4.7. Semivariogramme, simulierte Daten. . . . .	59
4.8. Auswahl bedingter lokaler Verteilungsfunktionen. . . . .	61
4.9. Simulierte Werte $z_{(578)}^1$ und $z_{(578)}^2$ . . . . .	62
4.10. Prognosen zum Schwellenwert $q_{0.5,z_{(578)}^1}$ . . . . .	63
4.11. Absolute Fehler zum Schwellenwert $q_{0.5,z_{(578)}^2}$ . . . . .	63
4.12. Maße der Prognosegüte I. . . . .	65
4.13. Maße der Prognosegüte II. . . . .	66



# 1. Einleitung

Die meisten forstlichen Daten haben eines gemeinsam: sie stammen aus Waldflächen. Sie werden entweder, wie bei einer Messung der Grundfläche pro Hektar mit einer Winkelzählprobe, an einen Punkt oder, wie bei einer Zählung der Stammzahl in einem Probekreis, auf einer Fläche definierter Form und Größe erhoben. In beiden Fällen ist der Ort der Datenaufnahme bekannt, daher können die Meßwerte in Karten eingetragen werden. Solch eine Karte kann lückenlos sein, falls die Probeflächen das gesamte Untersuchungsgebiet abdecken.

Meist erfolgt die Datenaufnahme aber nicht flächendeckend, sondern stichprobenhaft: Abbildung 1.1 zeigt eine Karte mit Stichprobendaten der Immissionsökologischen Waldzustandserfassung (IWE) in Baden-Württemberg. Die mittleren Magnesiumgehalte in cg pro kg Trockensubstanz in Fichtennadeln wurden im Jahr 1994 an 578 Stichprobenpunkten gemessen; der größte Teil der Karte besteht aber aus weißen Flecken<sup>1</sup>. Wünschenswerter sind Karten, die flächendeckende Informationen darstellen. Um die weißen Flecken – also Orte, an denen keine Daten erhoben wurden – der Karte in Abbildung 1.1 ohne kostenverursachende zusätzliche Datenaufnahme zu füllen, gibt es verschiedene Interpolationsverfahren<sup>2</sup> (Webster u. Oliver (2001), Kapitel 3 gibt eine erste Übersicht). Die Interpolationsergebnisse können dann in Karten wie in Abbildung 1.2(a) dargestellt werden.

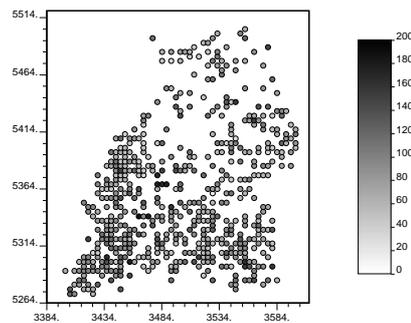


Abbildung 1.1.: Stichprobenwerte.

Nehmen wir an, daß den Daten ein räumlicher Zufallsprozess, eine unendlichdimensionale Zufallsfunktion zugrundeliegt. Die Meßwerte wären dann der beobachtete Teil einer *Realisation* dieses Prozesses – oder „zusammengehörige“ Realisationen mehrerer beobachteter Zufallsvariabler mit einer gemeinsamen Verteilung. Dann können wir einige der Interpolationsalgorithmen, wie Regressionsmodelle, als *Prädiktoren* und deren interpolierte Werte als *Prädiktionen* oder *Prognosen* des unbeobachteten Teils der Realisation des Prozesses – oder der zugehörigen Realisationen unbeobachteter Zufallsvariabler der gleichen gemeinsamen Verteilung – interpretieren. Aufgrund des unterstellten stochastischen Modells ist es möglich, Aussagen über die Genauigkeit der Prädiktoren, üblicherweise in Form des mittleren quadratischen Prognosefehlers (*mspe*), zu treffen und diese ebenfalls in Karten wie der in Abbildung 1.2(b)

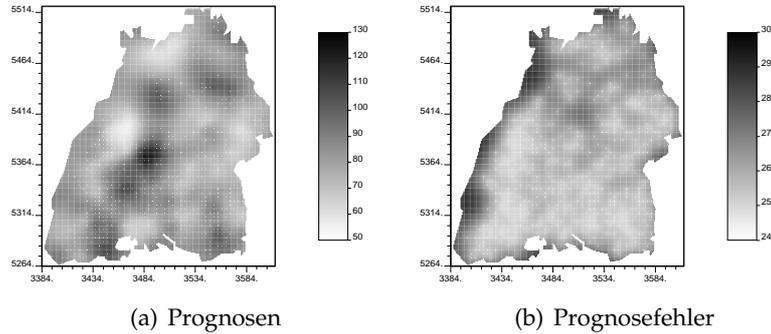


Abbildung 1.2.: Interpolation einer räumlichen Variablen.

darzustellen.

Die einfachste Art der linearen Regression ist der arithmetische Mittelwert der Meßwerte. Das zugehörige Regressionsmodell für  $N$  an den Orten  $\mathbf{s}_1, \dots, \mathbf{s}_N \in \mathbb{R}^2$  gemessene Daten  $z(\mathbf{s}_1), \dots, z(\mathbf{s}_N)$  lautet

$$\begin{aligned} Z(\mathbf{s}_n) &= \beta_0 + \varepsilon(\mathbf{s}_n) \\ &= \mu + \varepsilon(\mathbf{s}_n) \quad \forall n = 1, \dots, N \end{aligned}$$

und

$$\hat{\mu} = \sum_{n=1}^N \frac{1}{N} Z(\mathbf{s}_n)$$

ist der beste lineare unverzernte Schätzer (BLUE) für den Parameter  $\mu$ , falls alle Residuen den Erwartungswert Null, gleiche Varianzen und gleiche Kovarianzen untereinander haben (vergleiche Searle (1971), S. 89):

$$E[\varepsilon(\mathbf{s}_n)] = 0 \quad \forall n \quad (1.0.1a)$$

$$\text{var}[\varepsilon(\mathbf{s}_n)] = \sigma^2 \quad \forall n \quad (1.0.1b)$$

$$\text{cov}(\varepsilon(\mathbf{s}_n), \varepsilon(\mathbf{s}_m)) = \rho \quad \forall n \neq m. \quad (1.0.1c)$$

Sind die Kovarianzen nicht alle gleich Null, zeigt Cressie (1991), S. 13f, am Beispiel eines gauß'schen Prozesses, daß die Varianz des Erwartungswertschätzers  $\hat{\mu}$  größer ist als unter Annahme unabhängiger Residuen. Bei Missachtung einer vorhandenen Abhängigkeit überschätzen wir also die Genauigkeit von  $\hat{\mu}$  und berechnen zu schmale Prognoseintervalle der Form

$$\left[ \hat{\mu} - \sigma_g / \sqrt{n} \cdot u_{1-\frac{\alpha}{2}}; \hat{\mu} + \sigma_g / \sqrt{n} \cdot u_{1-\frac{\alpha}{2}} \right]$$

<sup>1</sup>Das liegt auch daran, daß nicht ganz Baden-Württemberg mit Fichtenwald bedeckt ist, vor allem aber daran, daß in den Fichtenbeständen nur Magnesiumgehalte einzelner Bäume an den Knotenpunkten eines  $4 \times 4$ -km Gitters gemessen wurden.

<sup>2</sup>In dieser Arbeit verstehe ich unter Interpolation das Errechnen von Werten einer Funktion an Stellen, die zwischen Stellen mit bekannten Werten liegen; ohne dabei weitere Anforderungen (wie Reproduktion der bekannten Werte) zu stellen. Regressionen und Simulationen sind in diesem Sinne Interpolationen.

auf den wahren Erwartungswert  $\mu$ , wobei  $\sigma_g$  die Standardabweichung des gauß'schen Prozesses und  $u_{1-\frac{\alpha}{2}}$  das  $1 - \frac{\alpha}{2}$ -Quantil der Standardnormalverteilung ist.

Der arithmetische Mittelwert aus den Daten kann als Prognose

$$p(\mathbf{s}_0, \mathbf{z})_{\hat{\mu}} = \hat{z}(\mathbf{s}_0) = \sum_{n=1}^N \frac{1}{N} z(\mathbf{s}_n)$$

der zugehörigen Realisation  $z(\mathbf{s}_0)$  der nicht beobachteten Zufallsvariablen  $Z(\mathbf{s}_0)$  dienen. Der entsprechende Prädiktor

$$p(\mathbf{s}_0, \mathbf{Z})_{\hat{\mu}} = \hat{Z}(\mathbf{s}_0) = \sum_{n=1}^N \frac{1}{N} Z(\mathbf{s}_n), \quad (1.0.2)$$

ist eine Funktion der  $N$  beobachteten Zufallsvariablen  $Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_N)$  (und damit selbst eine Zufallsvariable) und der beste lineare unverzerrte Prädiktor ( $p_1(\mathbf{s}_0, \mathbf{Z}, \theta)$ ) für  $Z(\mathbf{s}_0)$  unter allen Prädiktoren der Form  $\hat{Z}(\mathbf{s}_0) = \hat{\beta}_0$ , falls 1.0.1 erfüllt sind. Sind  $Z(\mathbf{s}_0)$  und  $\hat{\mu}$  stochastisch unabhängig und die Varianzen aller Zufallsvariablen identisch (1.0.1b), so ist der *mspe* des Prädiktors 1.0.2 nach Searle (1971), S. 91

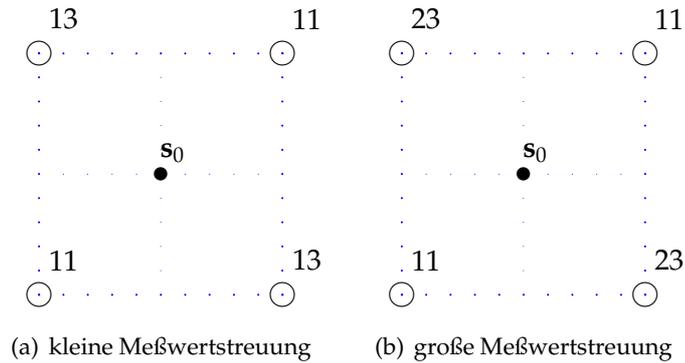
$$E \left[ \left\{ p(\mathbf{s}_0, \mathbf{Z})_{\hat{\mu}} - Z(\mathbf{s}_0) \right\}^2 \right] = \sigma_0^2 \{1 + 1/N\}, \quad (1.0.3)$$

wobei  $\sigma_0^2$  die Varianz des betrachteten Prozesses an der Stelle  $\mathbf{s}_0$  ist.

Die Annahme, daß alle Paare aus je zwei Zufallsvariablen  $\varepsilon(\mathbf{s}_n)$  und  $\varepsilon(\mathbf{s}_m)$  (und damit die Variablen, deren Realisationen die Daten sind) die selbe Kovarianz haben (1.0.1c), ist sehr streng. Daß zudem durch diese Annahme alle Daten ungeachtet ihrer Entfernung von  $\mathbf{s}_0$  bei der Berechnung der Prognose das Gewicht  $1/N$  erhalten, leuchtet bei vielen ökologischen Daten nicht ein. Vielmehr scheint es so, als seien sich beispielsweise Jahresniederschlagssummen tendenziell ähnlicher, je kleiner der Abstand zwischen ihren Meßstellen ist.

Wird die Annahme identischer Kovarianzen aufgehoben, sprechen wir von allgemeinen linearen Regressionen, die ganz allgemein dazu dienen können, den Wert einer unbeobachteten Zufallsvariablen aus dem Wert eines beobachteten Zufallsvektors vorherzusagen (vergleiche Harville (1985), S. 132); der Zufallsvektor ist dabei  $[Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_N)]^T = \mathbf{Z}$ , sein Wert  $[z(\mathbf{s}_1), \dots, z(\mathbf{s}_N)]^T = \mathbf{z}$ . Um allgemeine lineare Regressionen als Prädiktoren nutzen zu können, müssen aber die Kovarianzen bekannt oder zumindest schätzbar sein<sup>3</sup>. Da bei forstlichen Daten meist die zur Kovarianzschätzung nötigen Wiederholungen fehlen, sie aber als ökologische Daten oft die angesprochene Eigenheit haben, sich mit abnehmendem Abstand zwischen zwei Beobachtungsorten stärker zu ähneln, wird angenommen, daß die Kovarianz zwischen den Variablen an zwei

<sup>3</sup>Die Forderung nach bekannten Kovarianzen scheint schwer zu erfüllen, aber für nicht einfache lineare Regressionen muß sogar bekannt oder zumindest anzunehmen sein, daß alle Kovarianzen identisch sind (1.0.1c).



**Abbildung 1.3.:** Messstellenanordnung mit unterschiedlichen Meßreihen.

verschiedenen Orten eine Funktion nur des Abstandsvektors zwischen den beiden Orten ist. Diese Eigenschaft wird in der Literatur als eine Eigenschaft sogenannter stationärer Prozesse angesehen und ermöglicht die Schätzung der Kovarianzen aus Wiederholungen von Abstandsvektorklassen.

Allgemeine lineare Regressionen, die auf der Annahme einer räumlichen Abhängigkeit beruhen und die  $Z(\mathbf{s}_n)$  als Prädiktoren verwenden, werden *Kriging*verfahren genannt. Sie ersetzen den Prädiktor 1.0.2 durch Prädiktoren der Form

$$\hat{Z}_k(\mathbf{s}_0) = p(\mathbf{s}_0, \mathbf{Z}) = \sum_{n=1}^N \lambda_n(\mathbf{s}_0) Z(\mathbf{s}_n) + \lambda_0(\mathbf{s}_0),$$

die jedem Ort ein eigenes Gewicht  $\lambda_n$  zuweisen und den Prädiktor 1.0.2 als Sonderfall einschließen.

Eine Eigenschaft der Krigingprädiktoren ist, daß die Meßwerte in die Berechnung ihrer *mspe* ( $\sigma_k^2(\mathbf{s}_0)$ ) nicht eingehen, wie am Beispiel des *mspe* 1.0.3 leicht zu erkennen ist. Die Prognosefehler eines Krigingprädiktors an den Stellen  $\mathbf{s}_0$  in Abbildung 1.3(a) und Abbildung 1.3(b) sind daher identisch, obwohl die Prognose in Abbildung 1.3(a) „sicherer“ wirkt als die in Abbildung 1.3(b).

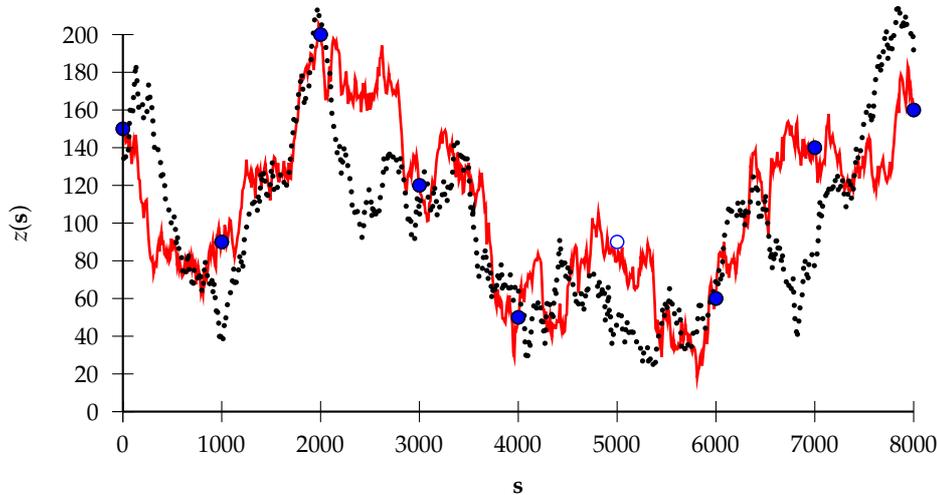
Während durch die Erhebung der Daten ein Zufallsvektor zu einem Vektor bekannter Werte ( $\mathbf{Z} = \mathbf{z}$ ) und somit ein Prädiktor  $p(\mathbf{s}_0, \mathbf{Z})$  durch die Bedingung  $\mathbf{Z} = \mathbf{z}$  zur Prognose

$$\hat{z}_k(\mathbf{s}_0) = p(\mathbf{s}_0, \mathbf{z}) = \sum_{n=1}^N \lambda_n(\mathbf{s}_0) z(\mathbf{s}_n) + \lambda_0(\mathbf{s}_0),$$

wird, bleibt sein Fehler von den Daten unabhängig, der Fehler der Prognose unbekannt.

Das bedeutet, daß die *mspe* über alle möglichen Realisationen hinweg berechnet werden. Abbildung 1.4 zeigt zwei Realisationen eines Zufallsprozesses über dem Intervall  $[0, 8000]$ : Eine (—), aus der die an den Stellen  $\mathbf{s}_n = -4.5, -3.5, \dots, 4.5$  beobachteten Daten stammen können und eine andere mögliche ( $\cdots$ ).

Wenn wir aus den Meßwerten die Prognose für die Stelle  $\mathbf{s}_0=0$  berechnen, interessiert uns die mittlere quadratische Abweichung des Prädiktors



**Abbildung 1.4.:** Datenreproduzierende (—) und beliebige Realisation (---) eines Zufallsprozesses über dem Intervall  $[0,8000]$ . Die dunklen Punkte stellen die Daten, der Kreis über (5000) die Prognose dar.

von nur den Realisationen, aus denen die Daten stammen können. Diese durch die beobachteten Daten bedingten mittleren quadratischen Prognosefehler ( $\sigma_{k|Z=z}^2(\mathbf{s}_0)$ ) sind nicht prinzipiell berechenbar; in der Praxis werden daher meist die *mspe*, also die mittleren quadratischen Abweichungen über alle möglichen Realisationen, angegeben. Diese sollten zur besseren Unterscheidung unbedingte mittlere quadratische Prognosefehler genannt werden.

Aus dem bedingten mittleren quadratischen Prognosefehler können wir das *exakte Prognoseintervall* für den wahren Wert  $z(\mathbf{s}_0)$  zur Konfidenz  $1 - \alpha$  ableiten:

$$I_{\alpha, Z=z}(\mathbf{s}_0) = [z_a; z_b] : P[Z(\mathbf{s}_0) \in [z_a; z_b] | \mathbf{Z} = \mathbf{z}] = 1 - \alpha.$$

Unter Annahme einer gauß'schen Verteilung des räumlichen Prozesses verwenden wir hierfür

$$I_{\alpha, Z=z}^K(\mathbf{s}_0) = [\hat{z}_k(\mathbf{s}_0) - \sigma_{k|Z=z}(\mathbf{s}_0) \cdot u_{1-\frac{\alpha}{2}}, \hat{z}_k(\mathbf{s}_0) + \sigma_{k|Z=z}(\mathbf{s}_0) \cdot u_{1-\frac{\alpha}{2}}]$$

wobei  $u_{1-\frac{\alpha}{2}}$  das  $1 - \frac{\alpha}{2}$ -Quantil der Standardnormalverteilung ist. Benutzen wir hier, wie beispielsweise von Cressie (1991), S. 122, vorgeschlagen, den unbedingten  $\sigma_k^2(\mathbf{s}_0)$  anstelle des bedingten mittleren quadratischen Prognosefehlers  $\sigma_{k|Z=z}^2(\mathbf{s}_0)$ , so wäre die Breite des resultierenden *cressieschen Prognoseintervalles* von den Meßwerten unabhängig und für die Beispiele in Abbildung 1.3 gleich.

Betrachten wir nur die obere Grenze eines exakten Prognoseintervalles, so liefert sie die durch die Messungen bedingte Wahrscheinlichkeit, mit welcher der Prozeß an einer Stelle  $\mathbf{s}_0$  einen Schwellen- oder Grenzwert  $z_c$  unterschreitet:

$$P[Z(\mathbf{s}_0) \leq z_c | \mathbf{Z} = \mathbf{z}].$$

Solche bedingten Wahrscheinlichkeiten werden immer dann benötigt, wenn uns weniger der interpolierte Wert  $\hat{z}(\mathbf{s}_0)$  an einer Stelle  $\mathbf{s}_0$  interessiert sondern

**Tabelle 1.1.:** Aus den Meßreihen in Abbildung 1.3 resultierende bedingte und unbedingte Wahrscheinlichkeiten.

Meßreihe	$P[Z(\mathbf{s}_0) \leq 14 \mid \mathbf{Z} = \mathbf{z}]$	$P[Z(\mathbf{s}_0) \leq 14]$
Abbildung 1.3(a)	0.80154	0.89705
Abbildung 1.3(b)	0.20287	0.89705

vielmehr die Frage, wie sicher der wahre Wert  $z(\mathbf{s}_0)$  den Schwellenwert  $z_c$  unterschreitet: Im Hinblick auf eine Kalkungsmaßnahme interessiert uns nicht allein, ob der interpolierte pH-Wert 5.5 oder 5.8 beträgt oder ob er größer als eine Schwelle von beispielsweise  $z_c = 3.8$  ist; wichtig ist, mit welcher Wahrscheinlichkeit der wahre Wert,  $z(\mathbf{s}_0)$ , 3.8 über- oder unterschreitet. Auch hier interessiert uns ein bedingtes Maß anstatt eines unbedingten Maßes: Wir erwarten zum Beispiel in Abbildung 1.3, daß die Wahrscheinlichkeit, daß der Prozeß an der Stelle  $\mathbf{s}_0$  den Wert  $z_c=14$  nicht überschreitet in Abbildung 1.3(a) deutlich größer ist als in Abbildung 1.3(b), wenn wir beiden Meßreihen den selben zugrundeliegenden Prozeß unterstellen. Nehmen wir an, daß dieser gauß' sch ist mit konstantem Erwartungswert  $\mu = 10$  und räumlicher Autokovarianz

$$C(\mathbf{h}) = 10 \exp(-1.5|\mathbf{h}|),$$

wobei  $\mathbf{h}$  den Abstandsvektor zweier Orte bezeichne.

Wenn der Prognosepunkt im Mittelpunkt eines Quadrates mit Seitenlänge 1 liegt, dessen Eckpunkte die Datenpunkte bilden, so zeigt Tabelle 1.1 die nach 2.3.3 berechneten bedingten und die unbedingten Wahrscheinlichkeiten: die aus der Meßreihe in Abbildung 1.3(a) resultierende bedingte Wahrscheinlichkeit ist wie erwartet größer als die sich aus der Meßreihe in Abbildung 1.3(b) ergebende – die unbedingten Wahrscheinlichkeiten sind gleich und laut Johnson u. Kotz (1970), S. 40 gegeben durch

$$\Phi\left(\frac{z_c - \mu}{\sqrt{\text{var}[Z]}}\right) = 0.89705; \quad z_c = 14, \mu = 10, \text{var}[Z] = C(0) = 10.$$

Betrachten wir die bedingten Wahrscheinlichkeiten zu nicht nur einem sondern zu allen möglichen Schwellenwerten, so erhalten wir die bedingte lokale Verteilungsfunktion

$$F_{Z(\mathbf{s}_0)|\mathbf{Z}=\mathbf{z}}(z) = P[Z(\mathbf{s}_0) \leq z \mid \mathbf{Z} = \mathbf{z}] \quad \forall z \in \mathbb{R}.$$

Diese Funktion stellt den höchsten aufgrund der erhobenen Daten möglichen Wissensstand über den Prozeß an der Stelle  $\mathbf{s}_0$  dar: Sie liefert bedingte Unterschreitungswahrscheinlichkeiten, exakte Prognoseintervalle, mit dem zugehörigen bedingten Erwartungswert den optimalen Prädiktor und mit der bedingten Varianz dessen bedingten erwarteten Prognosefehler.

Betrachten wir wieder den pH-Wert des Bodens: Die Kenntnis seiner durch  $N$  Messungen bedingten Verteilungsfunktion nicht nur an einer sondern an potentiell allen Stellen eines Waldgebietes ermöglicht es uns, neben der optimalen Vorhersage einzelner pH-Werte, Teilflächen auszuweisen, in denen

---

die Wahrscheinlichkeit, daß der pH-Wert einen Grenzwert unterschreitet, eine Toleranz  $\alpha$  überschreitet. Die Genauigkeit der Ausweisung solcher Kalkungsflächen kann dann nur durch verbesserte Anordnung der Messungen oder zusätzliche Messungen erhöht werden.

Aus diesem Grund ist Näherung der bedingten lokalen Verteilungsfunktion aus den Meßwerten Thema dieser Arbeit.

Wie wir in Kapitel 2.2 sehen werden, ist im Spezialfall eines gauß'schen Prozesses die bedingte lokale Verteilungsfunktion die Normalverteilungsfunktion jener Variablen, deren bedingter Erwartungswert dem Prädiktor des Einfachen Krigings und deren bedingte Varianz dem mittleren quadratischen Prognosefehler dieses Prädiktors entspricht. Da Einfaches Kriging für gewöhnlich nicht möglich ist (es erfordert die Kenntnis einer normalerweise unbekanntem Erwartungswertfunktion) und für schwach stationäre Prozesse durch Gewöhnliches Kriging ersetzt wird, liegt es nahe, die bedingte lokale Verteilungsfunktion zu approximieren durch die Normalverteilung, deren bedingter Erwartungswert gleich dem Prädiktor des Gewöhnlichen Krigings und deren bedingte Varianz gleich dem bedingten mittleren quadratischen Prognosefehler dieses Prädiktors ist. Wie bereits erwähnt, sind die bedingten mittleren quadratischen Prognosefehler meist nicht verfügbar und werden durch die unbedingten ersetzt.

In Kapitel 3 werde ich daher die Differenz zwischen unbedingtem und bedingtem mittleren quadratischen Prognosefehler des Gewöhnlichen Krigings in gauß'schen Prozessen betrachten. Die Herleitung und Analyse dieser Differenz geschieht in Kapitel 3.1, in Kapitel 3.2 folgt daraus – am Beispiel von Daten der Immissionsökologischen Waldzustandserfassung in Baden-Württemberg – die Abschätzung des Fehlers, den wir begehen, wenn wir statt der exakten oder der durch Gewöhnliches Kriging genäherten exakten Prognoseintervalle für den Wert  $z(s_0)$  die cressieschen Prognoseintervalle berechnen.

In Kapitel 2 gebe ich eine Einführung in die im Rahmen dieser Arbeit relevanten Methoden der Geostatistik: Kapitel 2.1 stellt die Grundlage der räumlichen Statistik, die Annahme eines Zufallsprozesses, vor. In Kapitel 2.2 stelle ich gängige Prädiktionsverfahren und ihre Genauigkeiten dar. Um die Herleitung in Kapitel 3.1 vorzubereiten, weiche ich dabei in den Kapiteln 2.2.1 und 2.2.2 von der üblichen Herangehensweise ab. Die bedingten lokalen Verteilungsfunktionen, ihre Anwendungsmöglichkeiten sowie ihre Approximation mit den aus Kapitel 2.2 bekannten Verfahren und die Beurteilung dieser Approximation sind Thema des Kapitel 2.3. Kapitel 2 schließt mit einem Überblick über die Möglichkeiten und – hinsichtlich der Approximation bedingter lokaler Verteilungsfunktionen – Grenzen der geostatistischen Simulationsverfahren.

In Kapitel 4 werden drei der vorgestellten Approximationsverfahren für bedingte lokale Verteilungsfunktionen auf die aus Kapitel 4.1 bekannten Daten angewendet. Da Funktionswerte bedingter lokaler Verteilungsfunktionen – also Wahrscheinlichkeiten – Modelle sind und nicht meßbar, lassen sie sich für gewöhnlich nicht kreuzvalidieren – es sei denn, wir verwenden mit einem bekannten Zufallsprozeß simulierte Daten wie in Kapitel 4.2.

Die Arbeit schließt mit der Diskussion der Ergebnisse der Kapitel 3 und 4 in Kapitel 5.



## 2. Zielführende Darstellung geostatistischer Methoden

### 2.1. Räumliche Zufallsprozesse

Grundlage jedes geostatistischen Verfahrens ist die Annahme, daß die betrachteten Meßwerte, also die pH-Werte, die Brusthöhendurchmesser oder die Jahresniederschlagssummen, eine Stichprobe einer Realisation eines räumlichen Zufallsprozesses sind: Bezeichnet  $S_{(N)} = \{\mathbf{s}_n \in D, n = 1, \dots, N\}$  die Menge, deren Elemente die Koordinaten der  $N$  Orte sind, an denen die Meßwerte

$$z_{(N)} = \{z(\mathbf{s}) : \mathbf{s} \in S_{(N)}\} \quad (2.1.1)$$

erhoben wurden, dann werden die Meßwerte als Stichprobe einer Realisation

$$z = \{z(\mathbf{s}) : \mathbf{s} \in D\} \quad (2.1.2)$$

des räumlichen Zufallsprozesses

$$Z = \{Z(\mathbf{s}) : \mathbf{s} \in D\} \quad (2.1.3)$$

über dem Untersuchungsgebiet  $D$  oder als Realisation des über  $S_{(N)}$  definierten Teils

$$Z_{(N)} = \{Z(\mathbf{s}) : \mathbf{s} \in S_{(N)}\} \quad (2.1.4)$$

des Prozesses 2.1.3 aufgefaßt. Olea (1999), S. 11 bezeichnet 2.1.2 allgemein als partielle Realisation.

Die Matrix der Koordinaten ist dann  $\mathbf{s}_{(N)} = [\mathbf{s}_1, \dots, \mathbf{s}_N]^T$ , die geordnete Stichprobe

$$\mathbf{z} = [z(\mathbf{s}_1), \dots, z(\mathbf{s}_N)]^T$$

eine Realisation des Zufallsvektors

$$\mathbf{Z} = [Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_N)]^T. \quad (2.1.5)$$

Weiter nehmen wir an, daß der Prozeß  $\{Z(\mathbf{s}) : \mathbf{s} \in D\}$  sich in

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \varepsilon(\mathbf{s}) = \sum_{j=0}^p f_j(\mathbf{s}) \beta_j + \varepsilon(\mathbf{s}) = \mathbf{x}^T(\mathbf{s}) \boldsymbol{\beta} + \varepsilon(\mathbf{s}) \quad \forall \mathbf{s} \in D,$$

zerlegen läßt (vergleiche Cressie (1991), S. 151), wobei  $\mathbf{x}(\mathbf{s}) = [f_0(\mathbf{s}), \dots, f_p(\mathbf{s})]^T$  einen Vektor beliebiger bekannter Funktionen,  $\boldsymbol{\beta} = [\beta_0, \dots, \beta_p]^T$  einen Vektor (gewöhnlich) unbekannter Parameter bezeichnet.

## 2. Zielführende Darstellung geostatischer Methoden

---

Die Bestandteile des Prozesses sind also eine Erwartungswertfunktion

$$E[Z(\mathbf{s})] = \mu(\mathbf{s}) = \sum_{j=0}^p f_j(\mathbf{s}) \beta_j = \mathbf{x}^T(\mathbf{s}) \boldsymbol{\beta} \quad \forall \mathbf{s} \in D; \quad (2.1.6)$$

und ein Rest- oder Fehlerprozeß  $\varepsilon = \{\varepsilon(\mathbf{s}) : \mathbf{s} \in D\}$ , dessen Erwartungswert 0 über ganz  $D$  ist und von dem wir annehmen, daß er eine räumliche Abhängigkeit, also eine Kovarianzfunktion besitzt:

$$\begin{aligned} \text{cov}(\varepsilon(\mathbf{s}_n), \varepsilon(\mathbf{s}_m)) &= E[\{\varepsilon(\mathbf{s}_n) - E[\varepsilon(\mathbf{s}_n)]\} \{\varepsilon(\mathbf{s}_m) - E[\varepsilon(\mathbf{s}_m)]\}] \\ &= C(\mathbf{s}_n, \mathbf{s}_m) \quad \forall \mathbf{s}_n, \mathbf{s}_m \in D. \end{aligned} \quad (2.1.7)$$

Dies ist, da für alle  $\mathbf{s} \in D$  gilt  $E[\varepsilon(\mathbf{s})] = 0$  und  $\varepsilon(\mathbf{s}) = Z(\mathbf{s}) - E[Z(\mathbf{s})]$ , nach Ripley (1981), S. 45 zugleich auch die Kovarianzfunktion des Prozesses  $Z$ :

$$\begin{aligned} \text{cov}(Z(\mathbf{s}_n), Z(\mathbf{s}_m)) &= E[\{Z(\mathbf{s}_n) - E[Z(\mathbf{s}_n)]\} \{Z(\mathbf{s}_m) - E[Z(\mathbf{s}_m)]\}] \\ &= C(\mathbf{s}_n, \mathbf{s}_m) \quad \forall \mathbf{s}_n, \mathbf{s}_m \in D. \end{aligned}$$

Dementsprechend kann der Zufallsvektor 2.1.5 zerlegt werden:

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

wobei  $\mathbf{X}(\mathbf{s})$  eine  $N \times (p+1)$  Matrix mit Elementen  $x_{nj} = f_{j-1}(\mathbf{s}_n)$  und  $\varepsilon_n = \varepsilon(\mathbf{s}_n)$  ist. Der Erwartungswertvektor von  $\mathbf{Z}$  ist

$$\begin{aligned} \boldsymbol{\mu} &= [E[Z(\mathbf{s}_1)], \dots, E[Z(\mathbf{s}_N)]]^T, \\ (C(\mathbf{s}_n, \mathbf{s}_m))_{n,m=1,\dots,N} &= \boldsymbol{\Sigma} \end{aligned} \quad (2.1.8)$$

ist Kovarianzmatrix sowohl des Teilprozesses  $Z_{(N)}$  als auch des über  $S_{(N)}$  definierten Teils des Fehlerprozesses  $\boldsymbol{\varepsilon}$ , und

$$[\text{cov}(Z(\mathbf{s}_1), Z(\mathbf{s}_0)), \dots, \text{cov}(Z(\mathbf{s}_N), Z(\mathbf{s}_0))]^T = \mathbf{c} \quad (2.1.9)$$

ist der Kovarianzvektor zwischen dem Teilprozeß an den Meßstellen und dem unbeobachteten Teilprozeß an einem Prognoseort  $\mathbf{s}_0$ .

Die verschiedenen linearen Krigingverfahren treffen unterschiedliche Annahmen über die Erwartungswertfunktion 2.1.6, sie setzen aber alle voraus, daß die Kovarianzmatrix 2.1.8 und der Kovarianzvektor 2.1.9 - oder entsprechende Matrizen des Semivariogrammes - bekannt sind.

Um die benötigten Kovarianzfunktionen ohne Meßwiederholungen aus den Daten schätzen zu können, nehmen wir an, daß der Prozeß ergodisch (vergleiche Cressie (1991), S. 53ff) und damit stationär im Sinne einer von mehreren möglichen Definitionen ist. Die Definition, daß der Erwartungswert 2.1.6 konstant über  $D$  und die Kovarianzfunktion 2.1.7 (und mit ihr 2.1.8 und 2.1.9) eine Funktion des Abstandsvektors zwischen zwei Orten, nicht aber deren absoluter Lage ist,

$$\begin{aligned} E[Z(\mathbf{s})] &= \mu \quad \forall \mathbf{s} \in D, \\ C(\mathbf{s}_n - \mathbf{s}_m) &= C(\mathbf{h}) \geq 0 \quad \forall \mathbf{s}_n, \mathbf{s}_m \in D, \end{aligned}$$

wird in der Literatur (zum Beispiel bei Cressie (1991), S. 53) meist Schwache Stationarität oder Stationarität zweiter Ordnung des Prozesses  $Z$  genannt. Die Kovarianzfunktion muß positiv definit sein (vergleiche Cressie (1991), S. 68), woraus folgt, daß jede aus ihr bildbare Kovarianzmatrix  $\Sigma$  nichtnegativ definit<sup>1</sup> ist. Die abgeschwächte Forderung, daß

$$E[Z(\mathbf{s})] = \mu \quad \forall \mathbf{s} \in D, \quad (2.1.10)$$

und

$$\gamma(\mathbf{h}) = \frac{1}{2} \text{var}[Z(\mathbf{s}) - Z(\mathbf{s} + \mathbf{h})] \geq 0 \quad \forall \mathbf{s}, \mathbf{s} + \mathbf{h} \in D, \quad (2.1.11)$$

gilt, wird intrinsische Stationarität, die Funktion  $\gamma(\mathbf{h})$  Semivariogrammfunktion genannt. Ein schwach stationärer Prozeß ist immer auch intrinsisch stationär, für Maße räumlicher Abhängigkeit gilt dann  $\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$ . Prozesse, deren Semivariogramm- oder Kovarianzfunktionen nur von der Länge des Vektors  $\mathbf{h}$ , nicht aber seiner Richtung abhängen, werden isotrope Prozesse genannt.

### 2.1.1. Schätzung der räumlichen Abhängigkeit

Die räumliche Abhängigkeit eines Prozesses wird meist als Semivariogramm dargestellt. Chilès u. Delfiner (1999), S. 38 unterscheiden zwischen 2.1.11, dem Semivariogramm des Prozesses  $Z$  (2.1.3), und dem Semivariogramm der Stichprobe  $z_{(N)}$  (2.1.1),

$$\gamma_S(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{N(\mathbf{h})} \{z(\mathbf{s}_n) - z(\mathbf{s}_m)\}^2.$$

Die Schätzung des Semivariogrammes des Prozesses erfolgt mit dem sich aus dem Semivariogramm der Stichprobe ergebenden Schätzer

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{N(\mathbf{h})} \{Z(\mathbf{s}_n) - Z(\mathbf{s}_m)\}^2,$$

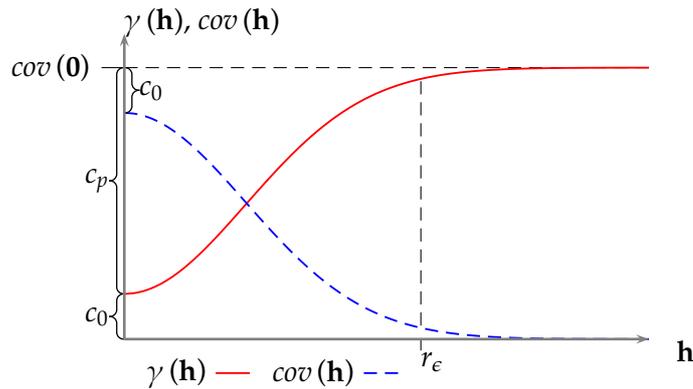
wobei  $N(\mathbf{h}) = \{(\mathbf{s}_n, \mathbf{s}_m) : \mathbf{s}_n - \mathbf{s}_m = \mathbf{h}; n, m = 1, \dots, N\}$  alle Ortspaare sind, die den Abstand  $\mathbf{h}$  aufweisen und  $|N(\mathbf{h})|$  ihre Anzahl bezeichnet (vergleiche Cressie (1991), S. 69). Journel u. Huijbregts (1978), S. 194 empfehlen, nur solche Paare zu betrachten, deren Abstand kleiner als die Hälfte des betrachteten Untersuchungsgebietes ist.

Ist das Histogramm der Meßwerte sehr schief („long-tailed“), empfehlen Chilès u. Delfiner (1999), S. 142 die Verwendung robuster Semivariogrammschätzer; Cressie (1991) widmet diesen das Kapitel 2.4.3, Chilès u. Delfiner

<sup>1</sup>Nach Harville (1997), S. 210 ist eine Matrix  $\Sigma \in \mathbb{R}^{n \times n}$  nichtnegativ definit, wenn sie

$$\mathbf{x}^T \Sigma \mathbf{x} \geq 0 \quad \forall \mathbf{x} \in \mathbb{R}^n$$

erfüllt.



**Abbildung 2.1.:** Gauß'sche Semivariogramm- und Kovarianzfunktion,  $c_p$  bezeichnet nach Cressie (1991), S. 61, 67 den partiellen Schwellenwert,  $c_0$  den Nugget-Effekt und der Schwellenwert ist  $cov(0) = c_p + c_0$ .

(1999) das Kapitel 2.2.5. Um eine genügend genaue Schätzung von  $\gamma(\mathbf{h})$  zu gewährleisten, empfehlen Journel u. Huijbregts (1978), S. 194  $|N(\mathbf{h})| \geq 30, 50$ . Chilès u. Delfiner (1999), S. 38 empfiehlt  $|N(\mathbf{h})| \geq 50$ , Olea (1999), S. 71 und Cressie (1991), S. 70 zitieren Journel u. Huijbregts (1978). Da kaum ein Datensatz diese Anforderungen erfüllt, werden meist Abstandsvektorklassen gebildet und Schätzer für  $\gamma$  berechnet. Werden beispielsweise die Abstandsvektorklassen als  $\{(l \times \mathbf{h} \pm \mathbf{h}/2), l = 0, \dots, L\}$  definiert, ergeben sich  $L + 1$  Schätzer

$$\hat{\gamma}_l(\mathbf{h}) = \frac{1}{2|N_l(\mathbf{h})|} \sum_{N_l(\mathbf{h})} \{Z(\mathbf{s}_n) - Z(\mathbf{s}_m)\}^2,$$

$$N_l(\mathbf{h}) = \{(\mathbf{s}_n, \mathbf{s}_m) : \mathbf{s}_n - \mathbf{s}_m \in (l \times \mathbf{h} \pm \mathbf{h}/2); n, m = 1, \dots, N\}.$$

An die mit diesem Schätzer berechneten Schätzwerte wird eine bedingt negativ definite Funktion angepaßt (Cressie (1991), Kapitel 2.5.2), um die räumliche Abhängigkeit für beliebige Distanzen angeben zu können - Cressie (1991), Kapitel 2.6 diskutiert verschiedene Anpassungsmethoden, neben der rein visuellen Anpassung auch Kleinste Quadrate. Geeignete bedingt negativ definite Funktionen finden sich bei Chilès u. Delfiner (1999), Kapitel 2.5. Abbildung 2.1 zeigt die um ein Nuggetmodell

$$\gamma(\mathbf{h}) = \begin{cases} c_0; & c_0 > 0 & \text{wenn } |\mathbf{h}| > 0, \\ 0 & & \text{wenn } |\mathbf{h}| = 0 \end{cases}$$

auch Nugget-Effekt genannt, erweiterte Gauß'sche Semivariogramm- und Kovarianzfunktion eines schwach stationären Prozesses, sie ist nach Hoeting u. a. (2006), S. 88 ein Grenzfall des flexiblen Matern-Modelles.

Die Gauß'sche Semivariogrammfunktion wird von Chilès u. Delfiner (1999), S. 85 in ihrer Grundform mit

$$\gamma(\mathbf{h}) = c_p \left[ 1 - \exp\left(-\frac{\mathbf{h}^2}{a_0^2}\right) \right]$$

angegeben. In dieser Formulierung entspricht  $r_e$ , die effektive Reichweite, bei der eine asymptotische Semivariogrammfunktion etwa 95 Prozent ihres Schwellenwertes erreicht,  $\sqrt{3} \cdot a_0$ . Bei dieser und anderen asymptotischen Semivariogrammfunktionen haben daher verschiedene Autoren verschiedene Darstellungen der Grundform eingeführt, diese wurden in Softwarelösungen übernommen und führen in der Praxis zu unterschiedlichen Parametrisierungen eigentlich identischer Semivariogrammfunktionen.

### 2.1.2. Erwartungswertschätzung

Die in Kapitel 2.2.1 vorgestellten linearen Krigingprädiktoren benötigen eine bekannte (Einfaches Kriging) oder geschätzte (Gewöhnliches und Universelles Kriging) Erwartungswertfunktion.

Bei der Schätzung unterscheiden wir danach, ob die Erwartungswertfunktion konstant über  $D$  ist (2.1.10) oder nicht. Ist die Erwartungswertfunktion konstant, können wir Gewöhnliches Kriging anwenden; ist die Erwartungswertfunktion nicht konstant, können wir durch Universelles Kriging den Einfluß externer erklärender Variabler – bei der Interpolation der Jahresmitteltemperatur zum Beispiel könnten dies die Geländehöhe über NN und die geographischen Koordinaten sein – berücksichtigen.

#### Schätzung eines räumlich variablen Erwartungswertes

Ist  $\beta$  und damit die Erwartungswertfunktion 2.1.6 des Prozesses unbekannt, ist nach Cressie (1991), S. 20f

$$\hat{\beta}_{gls} = (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-1} \mathbf{Z} \quad (2.1.12)$$

der beste lineare unverzerrte Schätzer für  $\beta$ . Somit ist

$$\hat{\mu}_{gls}(\mathbf{s}) = \mathbf{x}^T(\mathbf{s}) \hat{\beta}_{gls} \quad \forall \mathbf{s} \in D \quad (2.1.13)$$

der beste lineare unverzerrte Schätzer für 2.1.6 und

$$\text{var}[\hat{\beta}_{gls}] = E\left[(\hat{\beta}_{gls} - \beta)(\hat{\beta}_{gls} - \beta)^T\right] = (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \quad (2.1.14)$$

die Varianz von 2.1.12 (vergleiche Cressie (1991), S. 21).

#### Schätzung eines konstanten Erwartungswertes

Ist die Erwartungswertfunktion konstant über  $D$ , so gilt in 2.1.6  $p = 0 \wedge f_0(\mathbf{s}) = 1 \wedge \beta_0 = \mu$ . Daraus folgt  $\mathbf{x}^T(\mathbf{s}) = \mathbf{x}(\mathbf{s}) = 1, \mathbf{X} = \mathbf{1}, \beta = \beta_0 = \mu$  und

$$\hat{\mu}(\mathbf{s}) = \hat{\mu} = (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1} \mathbf{1}^T \Sigma^{-1} \mathbf{Z} = \mathbf{a}^T \mathbf{Z} \quad \forall \mathbf{s} \in D \quad (2.1.15)$$

ist der beste lineare unverzerrte Schätzer für  $\mu$ . Die Summe der Gewichte  $\mathbf{a}^T$  ist  $(\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1} \mathbf{1}^T \Sigma^{-1} \mathbf{1} = 1, E[\hat{\mu}] = E[\mathbf{a}^T \mathbf{Z}] = \mathbf{a}^T E[\mathbf{Z}] = \mathbf{a}^T \boldsymbol{\mu} = \mu$  der Erwartungs-

wert<sup>2</sup> von  $\hat{\mu}$  und nach Olea (1999), S. 10, Lemma 2.4 ist

$$E[(\hat{\mu} - \mu)^2] = \text{var}[\hat{\mu}] = \frac{1}{\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}} \quad (2.1.16)$$

die Varianz<sup>3</sup> von  $\hat{\mu}$  (eine genauere Darstellung der Umformung findet sich in Kapitel B.2).

## 2.2. Räumliche Vorhersagen

Für einen Prädiktor  $p(\mathbf{s}_0, \mathbf{Z})$  kann der unbedingte mittlere quadratische Prognosefehler (vergleiche Chilès u. Delfiner (1999), S. 154) folgendermaßen zerlegt werden

$$E\left[\{p(\mathbf{s}_0, \mathbf{Z}) - Z(\mathbf{s}_0)\}^2\right] = \text{var}[p(\mathbf{s}_0, \mathbf{Z}) - Z(\mathbf{s}_0)] + \{E[p(\mathbf{s}_0, \mathbf{Z}) - Z(\mathbf{s}_0)]\}^2. \quad (2.2.1)$$

Dies vereinfacht sich zu

$$E\left[\{p(\mathbf{s}_0, \mathbf{Z}) - Z(\mathbf{s}_0)\}^2\right] = \text{var}[p(\mathbf{s}_0, \mathbf{Z}) - Z(\mathbf{s}_0)],$$

falls der Prädiktor unverzerrt und somit  $E[p(\mathbf{s}_0, \mathbf{Z}) - Z(\mathbf{s}_0)] = 0$  ist. Der unbedingte mittlere quadratische Prognosefehler ist der über alle möglichen Realisationen  $\mathbf{z}$  von  $\mathbf{Z}$  hinweg erwartete quadratische Prognosefehler. Anstatt nur jene Realisationen zu betrachten, für die  $\mathbf{z}=\mathbf{Z}$  gilt, wird die durch die Datenerhebung gewonnene Information nicht berücksichtigt: der bedingte mittlere quadratische Prognosefehler ist eine Funktion der Anordnung der Prognose- und Meßstellen sowie der räumlichen Abhängigkeit. Ist diese bekannt, kann er a priori zur Bewertung eines Stichprobendesigns herangezogen werden.

Der im Hinblick auf die Beurteilung der Güte der Prognose interessantere bedingte mittlere quadratische Prognosefehler ist nach Cressie (1991), S. 111

$$E\left[\{p(\mathbf{s}_0, \mathbf{Z}) - Z(\mathbf{s}_0)\}^2 \mid \mathbf{Z}\right] = \text{var}[Z(\mathbf{s}_0) \mid \mathbf{Z}] + \{E[Z(\mathbf{s}_0) \mid \mathbf{Z}] - p(\mathbf{s}_0, \mathbf{Z})\}^2. \quad (2.2.2)$$

und für gewöhnlich nicht berechenbar, da der bedingte Erwartungswert  $E[Z(\mathbf{s}_0) \mid \mathbf{Z}]$  unbekannt ist. Dieser bedingte Erwartungswert, meist eine nichtlineare Funktion, hat den kleinsten mittleren quadratischen Prognosefehler aller Prädiktoren, ist unverzerrt (vergleiche Harville (1985), S. 133) und daher nach Cressie (1991), S. 108, der optimale Prädiktor

$$p_{opt}(\mathbf{s}_0, \mathbf{Z}) = E[Z(\mathbf{s}_0) \mid \mathbf{Z}].$$

---

<sup>2</sup> Falls der Prozess  $Z$  schwach stationär und unkorreliert ist, wird  $\boldsymbol{\Sigma}$  zu  $C(0) \cdot \mathbf{I}$  und

$$\hat{\mu} = \frac{\mathbf{1}^T \mathbf{Z}}{\mathbf{1}^T \mathbf{1}} = \frac{1}{N} \sum Z(\mathbf{s}_i).$$

<sup>3</sup>Falls der Prozess  $Z$  schwach stationär und unkorreliert ist, wird  $\boldsymbol{\Sigma}$  zu  $C(0) \cdot \mathbf{I}$  und

$$E[(\hat{\mu} - \mu)^2] = \frac{1}{\mathbf{1}^T (\mathbf{I}C(0))^{-1} \mathbf{1}} = \frac{C(0)}{\mathbf{1}^T \mathbf{1}} = \frac{C(0)}{N}.$$

Sein bedingter mittlerer quadratischer Prognosefehler ist

$$E \left[ \left\{ p_{opt}(\mathbf{s}_0, \mathbf{Z}) - Z(\mathbf{s}_0) \right\}^2 \mid \mathbf{Z} \right] = E \left[ \{E[Z(\mathbf{s}_0) \mid \mathbf{Z}] - Z(\mathbf{s}_0)\}^2 \mid \mathbf{Z} \right] = var [Z(\mathbf{s}_0) \mid \mathbf{Z}], \quad (2.2.3)$$

die bedingte Varianz; sein unbedingter mittlerer quadratischer Prognosefehler,

$$\begin{aligned} E \left[ \left\{ p_{opt}(\mathbf{s}_0, \mathbf{Z}) - Z(\mathbf{s}_0) \right\}^2 \right] &= E \left[ \{E[Z(\mathbf{s}_0) \mid \mathbf{Z}] - Z(\mathbf{s}_0)\}^2 \right] \\ &= E \left[ E \left[ \{E[Z(\mathbf{s}_0) \mid \mathbf{Z}] - Z(\mathbf{s}_0)\}^2 \mid \mathbf{Z} \right] \right] \\ &= E [var [Z(\mathbf{s}_0) \mid \mathbf{Z}]] \end{aligned} \quad (2.2.4)$$

wird wegen  $var [Y] = E [var [Y|X]] + var [E [Y|X]]$  (vergleiche Mood u. a. (1974), S. 159) zu  $var [Z(\mathbf{s}_0)] - var [E [Z(\mathbf{s}_0) \mid \mathbf{Z}]]$  und daher von Cressie (1991), S. 108 mit  $var [Z(\mathbf{s}_0)] - var [p_{opt}(\mathbf{s}_0, \mathbf{Z})]$  angegeben.

Da der bedingte Erwartungswert für gewöhnlich nicht bekannt ist, wurde eine Vielzahl von Prädiktoren für  $Z(\mathbf{s}_0)$  entwickelt, die unter verschiedenen Randbedingungen optimal sind. Cressie (1991), Kapitel 5.9.1 gibt eine Übersicht über die meisten dieser Verfahren, wir werden in Kapitel 2.2.1 die verbreitetsten linearen – Einfaches, Gewöhnliches und Universelles Kriging – und in Kapitel 2.2.4 einige nicht-lineare Prädiktoren betrachten.

**Wenn  $Z$  ein gauß'scher Prozeß ist,** ist  $var [Z(\mathbf{s}_0) \mid \mathbf{Z}]$  unabhängig von  $\mathbf{Z}$  (vergleiche Johnson u. Kotz (1972), S. 41) und daher gilt

$$var [Z(\mathbf{s}_0) \mid \mathbf{Z}] = E [var [Z(\mathbf{s}_0) \mid \mathbf{Z}]] \quad (2.2.5)$$

was nach 2.2.3 und 2.2.4 bedeutet, daß dann der bedingte und der unbedingte mittlere quadratische Prognosefehler des optimalen Prädiktors identisch sind:

$$E \left[ \left\{ p_{opt}(\mathbf{s}_0, \mathbf{Z}) - Z(\mathbf{s}_0) \right\}^2 \mid \mathbf{Z} \right] = E \left[ \left\{ p_{opt}(\mathbf{s}_0, \mathbf{Z}) - Z(\mathbf{s}_0) \right\}^2 \right]. \quad (2.2.6)$$

### 2.2.1. Lineare Prädiktoren

In der geostatistischen Literatur werden die linearen Prädiktoren der Verständlichkeit halber meist in der Form

$$p(\mathbf{s}_0, \mathbf{Z}) = \sum_{n=1}^N \lambda_n Z(\mathbf{s}_n)$$

angegeben. Ich wähle in diesem Kapitel einen etwas komplizierter wirkenden Ansatz, der auf den drei verschiedenen Annahmen über die Erwartungswertfunktion – bekannt, unbekannt aber linear in  $\mathbf{x}(\mathbf{s})$ , unbekannt aber konstant – beruht, um in Kapitel 3.1 die Differenz zwischen unbedingten und bedingten mittleren quadratischen Prognosefehlern leichter beschreiben zu können.

Ist die Erwartungswertfunktion bekannt, ist nach Cressie (1991), S. 173, (3.4.59)

$$\begin{aligned} p_{sk}(\mathbf{s}_0, \mathbf{Z}, \boldsymbol{\beta}) &= \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z} + (\mathbf{x}^T(\mathbf{s}_0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}) \boldsymbol{\beta} \\ &= \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z} + \mu(\mathbf{s}_0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} = \hat{Z}_{sk}(\mathbf{s}_0) \end{aligned} \quad (2.2.7)$$

der beste lineare unverzerrte Prädiktor für  $Z(\mathbf{s}_0)$  und das Interpolationsverfahren wird *Einfaches Kriging* genannt. Es gilt offenbar auch

$$\hat{Z}_{sk}(\mathbf{s}_0) = \mu(\mathbf{s}_0) + \mathbf{c}^T \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \boldsymbol{\mu}),$$

das heißt, der Einfache Krigingprädiktor ist nichts anderes als der bekannte Erwartungswert an der Stelle  $\mathbf{s}_0$  plus eine Linearkombination der Abweichungen der  $Z(\mathbf{s}_n)$  von ihrem jeweiligen Erwartungswert  $\mu(\mathbf{s}_n)$ .

Ist die Erwartungswertfunktion unbekannt aber linear in  $\mathbf{x}(\mathbf{s})$ , muß  $\boldsymbol{\beta}$  geschätzt werden (siehe Kapitel 2.1.2). Der beste lineare unverzerrte Prädiktor wird zu

$$\begin{aligned} p_{uk}(\mathbf{s}_0, \mathbf{Z}, \hat{\boldsymbol{\beta}}_{gls}) &= \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z} + (\mathbf{x}^T(\mathbf{s}_0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}) \hat{\boldsymbol{\beta}}_{gls} \\ &= \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z} + \hat{\mu}_{gls}(\mathbf{s}_0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu}_{gls} = \hat{Z}_{uk}(\mathbf{s}_0), \end{aligned} \quad (2.2.8)$$

und wir nennen die Interpolation *Universelles Kriging* (vergleiche Cressie (1991), S. 173, (3.4.61)).

Die Differenz zwischen den Prädiktoren des Universellen und des Einfachen Krigings, also zwischen 2.2.8 und 2.2.7, ist

$$\begin{aligned} \hat{Z}_{uk}(\mathbf{s}_0) - \hat{Z}_{sk}(\mathbf{s}_0) &= (\mathbf{x}^T(\mathbf{s}_0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}) (\hat{\boldsymbol{\beta}}_{gls} - \boldsymbol{\beta}) = (\hat{\mu}_{gls}(\mathbf{s}_0) - \mu(\mathbf{s}_0)) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_{gls} - \boldsymbol{\mu}). \end{aligned} \quad (2.2.9)$$

Nehmen wir an, daß die unbekannte Erwartungswertfunktion eine Konstante ist, daß also 2.1.10 gilt; dann wird, wie schon zu 2.1.15 erläutert wurde,  $\mathbf{x}(\mathbf{s})$  zu  $\mathbf{1}$ ,  $\mathbf{X}$  zu  $\mathbf{1}$ ,  $\hat{\boldsymbol{\beta}}_{gls}$  zu  $\hat{\mu}$  und  $\boldsymbol{\mu}$  zu  $\mathbf{1}\mu$ . Der beste lineare unverzerrte Prädiktor vereinfacht sich zu

$$\begin{aligned} p_{ok}(\mathbf{s}_0, \mathbf{Z}) &= \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z} + (\mathbf{1} - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}) \hat{\mu} \\ &= \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z} + \hat{\mu} - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \hat{\mu} = \hat{Z}_{ok}(\mathbf{s}_0) \end{aligned} \quad (2.2.10)$$

und das Verfahren heißt *Gewöhnliches Kriging*. Da ich für diese Darstellung des Gewöhnlichen Krigings keinen Textbeleg kenne, zeige ich die Übereinstimmung mit der Darstellung von  $\hat{Z}_{ok}(\mathbf{s}_0)$  bei Cressie (1991), S. 123 in Anhang B.3.

Wenn, wie beim Gewöhnlichen Kriging, die Erwartungswertfunktion konstant ist, wird der Prädiktor des Einfachen Krigings (2.2.7) zu

$$\hat{Z}_{sk}(\mathbf{s}_0) = \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z} + \mu - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \mu$$

und der Unterschied zwischen den Prädiktoren des Gewöhnlichen und des Einfachen Krigings ist

$$\hat{Z}_{ok}(\mathbf{s}_0) - \hat{Z}_{sk}(\mathbf{s}_0) = (\mathbf{1} - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}) (\hat{\mu} - \mu). \quad (2.2.11)$$

Den vorgestellten Prädiktoren ist gemeinsam, daß sie bedingt verzerrt sind, daß also

$$E \left[ \hat{Z}_k(\mathbf{s}_0) - Z(\mathbf{s}) \mid \mathbf{Z} \right] \neq 0; \quad k = sk, ok, uk \quad (2.2.12)$$

gilt. Dabei werden große Werte von  $Z(\mathbf{s})$  tendenziell unter- und kleine Werte tendenziell überschätzt, diese Eigenschaft ist als Glättung durch Kriging (vergleiche Olea (1999), S. 26f) bekannt.

**Wenn  $Z$  ein gauß'scher Prozeß ist,** ist  $E[Z(\mathbf{s}_0) \mid \mathbf{Z}]$  linear in  $\mathbf{Z}$  (vergleiche Cressie (1991), S. 109) und identisch mit dem Prädiktor des Einfachen Krigings (vergleiche z.B. Cressie (1991), S. 110):

$$p_{opt}(\mathbf{s}_0, \mathbf{Z}) = E[Z(\mathbf{s}_0) \mid \mathbf{Z}] = \hat{Z}_{sk}(\mathbf{s}_0). \quad (2.2.13)$$

### 2.2.2. Genauigkeit linearer Prädiktoren

Der mittlere quadratische Prognosefehler des Prädiktors des Einfachen Krigings, 2.2.7, ist nach Cressie (1991), S. 359, (5.9.14),

$$E \left[ \left\{ \hat{Z}_{sk}(\mathbf{s}_0) - Z(\mathbf{s}_0) \right\}^2 \right] = C(0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{c} = \sigma_{sk}^2(\mathbf{s}_0), \quad (2.2.14)$$

seine Herleitung aus 2.2.1 findet sich bei Olea (1999), S. 15f. Der mittlere quadratische Prognosefehler des Prädiktors des Universellen Krigings kann folgendermaßen zerlegt werden:

$$\begin{aligned} E \left[ \left\{ \hat{Z}_{uk}(\mathbf{s}_0) - Z(\mathbf{s}_0) \right\}^2 \right] &= E \left[ \left\{ (\hat{Z}_{sk}(\mathbf{s}_0) - Z(\mathbf{s}_0)) + (\hat{Z}_{uk}(\mathbf{s}_0) - \hat{Z}_{sk}(\mathbf{s}_0)) \right\}^2 \right] \\ &= E \left[ \left\{ \hat{Z}_{sk}(\mathbf{s}_0) - Z(\mathbf{s}_0) \right\}^2 \right] \\ &\quad + 2E \left[ (\hat{Z}_{sk}(\mathbf{s}_0) - Z(\mathbf{s}_0)) (\hat{Z}_{uk}(\mathbf{s}_0) - \hat{Z}_{sk}(\mathbf{s}_0)) \right] \\ &\quad + E \left[ \left\{ \hat{Z}_{uk}(\mathbf{s}_0) - \hat{Z}_{sk}(\mathbf{s}_0) \right\}^2 \right]. \end{aligned}$$

Nach Harville (1985), Kapitel 3.3, sind  $(\hat{Z}_{sk}(\mathbf{s}_0) - Z(\mathbf{s}_0))$  und  $(\hat{Z}_{uk}(\mathbf{s}_0) - \hat{Z}_{sk}(\mathbf{s}_0))$  unkorreliert<sup>4</sup> und da  $\hat{Z}_{sk}(\mathbf{s}_0)$  unverzerrt,  $E[\hat{Z}_{sk}(\mathbf{s}_0) - Z(\mathbf{s}_0)]$  also gleich 0 ist, gilt

$$\begin{aligned} \sigma_{uk}^2(\mathbf{s}_0) &= E \left[ \left\{ \hat{Z}_{uk}(\mathbf{s}_0) - Z(\mathbf{s}_0) \right\}^2 \right] \\ &= E \left[ \left\{ \hat{Z}_{sk}(\mathbf{s}_0) - Z(\mathbf{s}_0) \right\}^2 \right] + E \left[ \left\{ \hat{Z}_{uk}(\mathbf{s}_0) - \hat{Z}_{sk}(\mathbf{s}_0) \right\}^2 \right], \end{aligned}$$

was durch Einsetzen von 2.2.14 und 2.2.9

$$\begin{aligned} &C(0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{c} \\ &+ (\mathbf{x}^T(\mathbf{s}_0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}) E \left[ (\hat{\boldsymbol{\beta}}_{gls} - \boldsymbol{\beta}) (\hat{\boldsymbol{\beta}}_{gls} - \boldsymbol{\beta})^T \right] (\mathbf{x}^T(\mathbf{s}_0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{X})^T \end{aligned}$$

<sup>4</sup>Rivoirard (1994), S. 13 und andere Autoren nennen Unkorreliertheit zweier Variabler  $X, Y$  ( $cov(X, Y)=0$ ) Orthogonalität der Variablen  $X, Y$ .

(vergleiche Cressie (1991), S. 174, (3.4.64)) und durch Einsetzen von 2.1.14

$$\begin{aligned} \sigma_{uk}^2(\mathbf{s}_0) = & C(0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{c} \\ & + \left( \mathbf{x}^T(\mathbf{s}_0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} \right) \left( \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} \right)^{-1} \left( \mathbf{x}^T(\mathbf{s}_0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} \right)^T, \end{aligned}$$

ergibt, die Varianz des Universellen Krigings (vergleiche Cressie (1991), S. 155).

Falls die Erwartungswertfunktion eine Konstante über  $D$  ist, ändert sich in der obigen Zerlegung der Index  $uk$  zu  $ok$ , es gilt  $\mathbf{x}(\mathbf{s}_0) = 1$  und  $\mathbf{X} = \mathbf{1}$  (wie zu 2.1.15 erläutert) und der mittlere quadratische Prognosefehler des Prädiktors des Gewöhnlichen Krigings ergibt sich als

$$\begin{aligned} E \left[ \left\{ \hat{Z}_{ok}(\mathbf{s}_0) - Z(\mathbf{s}_0) \right\}^2 \right] = & C(0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{c} \\ & + \left\{ \left( 1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \right) \right\}^2 E \left[ (\hat{\mu} - \mu)^2 \right], \end{aligned} \quad (2.2.15a)$$

woraus durch Einsetzen von 2.1.16

$$\begin{aligned} \sigma_{ok}^2(\mathbf{s}_0) = & C(0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{c} \\ & + \left( 1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \right) \left( \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \right)^{-1} \left( 1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \right), \end{aligned} \quad (2.2.15b)$$

die Varianz des Gewöhnlichen Krigings wird (vergleiche Cressie (1991), S. 123, zur Umformung siehe Anhang B.4).

Die Differenz zwischen den mittleren quadratischen Prognosefehlern des Einfachen Krigings und des Gewöhnlichen Krigings beträgt (vergleiche Olea (1999), S. 61)

$$\sigma_{ok}^2(\mathbf{s}_0) - \sigma_{sk}^2(\mathbf{s}_0) = \left( 1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \right) \left( \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \right)^{-1} \left( 1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \right). \quad (2.2.16)$$

**Wenn  $Z$  ein gauß'scher Prozeß ist,** ist der Prädiktor des Einfachen Krigings gleich dem bedingten Erwartungswert (siehe 2.2.13), der bedingte mittlere quadratische Prognosefehler des optimalen Prädiktors ist gleich dem unbedingten mittleren quadratischen Prognosefehler des optimalen Prädiktors (siehe 2.2.6) und daher gilt

$$E \left[ \left\{ E[Z(\mathbf{s}_0) | \mathbf{Z}] - Z(\mathbf{s}_0) \right\}^2 \mid \mathbf{Z} \right] = E \left[ \left\{ E[Z(\mathbf{s}_0) | \mathbf{Z}] - Z(\mathbf{s}_0) \right\}^2 \right] = \sigma_{sk}^2(\mathbf{s}_0). \quad (2.2.17)$$

### 2.2.3. Blockvorhersagen

Interessieren uns nicht Punktprognosen sondern Block- oder Flächenmittelwerte

$$Z(B) = \begin{cases} \int_B Z(\mathbf{s}) d\mathbf{s} / |B|, & |B| > 0 \\ \text{ave}\{Z(\mathbf{s}) : \mathbf{s} \in B\}, & |B| = 0 \end{cases}, \quad |B| = \int_B d\mathbf{s},$$

ändern sich in der Definition 2.1.3 des Zufallsprozesses, nun

$$Z(B) = \mathbf{x}^T(B) \boldsymbol{\beta} + \varepsilon(B),$$

die Ausdrücke, die vom Koordinatenvektor ( $\mathbf{s}$ ) abhängen:

$$\mathbf{x}(B) = (f_0(B), \dots, f_p(B))^T, f_j(B) = \int_B f_j(\mathbf{s}) d\mathbf{s} / |B|, j = 0, \dots, p, |B| = \int_B d\mathbf{s} \text{ und}$$

$$\varepsilon(B) = \int_B \varepsilon(\mathbf{s}) d\mathbf{s} / |B| \text{ (vergleiche Cressie (1991), Kapitel 5.2).}$$

Die Zerlegungen des Prozesses sei wie in Kapitel 2.1, die Kovarianzfunktion wird zu einer Funktion

$$C(U, V) = \frac{1}{|U||V|} \int_U \int_V \text{cov}(\mathbf{v}, \mathbf{u}) d\mathbf{v} d\mathbf{u}$$

zwischen zwei Blöcken  $U$  und  $V$  (vergleiche Cressie (1991), S. 285), beziehungsweise zu einer Funktion

$$C(\mathbf{s}, V) = \frac{1}{|V|} \int_V \text{cov}(\mathbf{s}, \mathbf{v}) d\mathbf{v}$$

zwischen einem (Daten-) Punkt  $\mathbf{s}$  und einem (Prognose-) Block  $V$  (vergleiche Chilès u. Delfiner (1999), S. 195). Geschätzt wird sie nach Olea (1999), S. 197ff durch

$$\hat{C}(U, V) = \frac{1}{N(U)N(V)} \sum_{n=1}^{N(U)} \sum_{m=1}^{N(V)} \text{cov}(\mathbf{u}_n, \mathbf{v}_m),$$

beziehungsweise

$$\hat{C}(\mathbf{s}, V) = \frac{1}{N(V)} \sum_{m=1}^{N(V)} \text{cov}(\mathbf{s}, \mathbf{v}_m);$$

Journal u. Huijbregts (1978), S. 97 empfehlen dabei eine Diskretisierung  $N(U) = N(V) = 6$  für Flächen. Die Krigingprädiktoren und ihre unbedingten mittleren quadratischen Prognosefehler (vergleiche beispielsweise Cressie (1991), S. 155) benötigen die selbe Blockdiskretisierung wie die Schätzung der räumlichen Abhängigkeit.

Wenn beispielsweise mittels Einfachen Krigings aus Punktdaten  $z_{(N)}$  der Blockmittelwert  $z(V_0)$  prognostiziert werden soll, ändern sich in der Darstellung des Prädiktors,

$$\hat{Z}_{sk}(\mathbf{s}_0) = \mathbf{c}^T \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \boldsymbol{\mu}) + \mu(\mathbf{s}_0),$$

und in der seines mittleren quadratischen Prognosefehlers,

$$\sigma_{sk}^2(\mathbf{s}_0) = C(0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{c},$$

lediglich einige Kovarianzausdrücke: Der Kovarianzvektor  $\mathbf{c}$  zwischen Daten- und Zieleinheit wird sich in diesem Fall  $\mathbf{c} = [C(\mathbf{s}_1, V_0), \dots, C(\mathbf{s}_N, V_0)]^T$ , wobei  $C(\mathbf{s}_n, V_0)$  durch

$$\hat{C}(\mathbf{s}_n, V_0) = \frac{1}{N(V_0)} \sum_{m=1}^{N(V_0)} \text{cov}(\mathbf{s}_n, \mathbf{v}_m);$$

genähert wird. Die Varianz der Zieleinheit  $C(0)$  ändert sich von  $C(\mathbf{s}_0, \mathbf{s}_0)$  zu  $C(V_0, V_0)$ , sie wird durch

$$\hat{C}(V_0, V_0) = \frac{1}{N(V_0)N(V_0)} \sum_{n=1}^{N(V_0)} \sum_{m=1}^{N(V_0)} \text{cov}(\mathbf{v}_n, \mathbf{v}_m)$$

genähert. Je nach Symmetrie der Blockdiskretisierung läßt sich dieser Ausdruck noch vereinfachen. Da die Punktdaten weiterhin Punktdaten bleiben, bleibt die Kovarianzmatrix  $\Sigma$  zwischen den Datenpunkten unverändert. Die Prognose für einen Block entspricht dabei dem Mittelwert der Prognosen innerhalb des Blockes:

$$\hat{z}(V_0) = \frac{1}{N(V_0)} \sum_{n=1}^{N(V_0)} \hat{z}(\mathbf{v}_n).$$

Analoges gilt nicht für den mittleren quadratischen Prognosefehler des Prädiktors für den Block (vergleiche Olea (1999), S. 205f).

Ein großer Nachteil Krigings von Blöcken besteht darin, daß es nur mit linearen Formen des Krigings in der beschriebenen Art funktioniert. Für die im folgenden Kapitel vorgestellten nicht-linearen Verfahren gilt, daß Indikator-Kriging – von Ausnahmen abgesehen – nicht angewandt werden kann (siehe Chilès u. Delfiner (1999), S. 437) und Disjunktives Kriging nur für eine enge Auswahl an unterstellten Prozeßmodellen möglich ist (siehe Chilès u. Delfiner (1999), Kapitel 6.5). Ich werde daher im Rahmen dieser Arbeit nicht weiter auf das Kriging von Blöcken eingehen.

#### 2.2.4. Nicht-lineare Prädiktoren

Jedes nicht-lineare Krigingverfahren ist ein Einfaches oder Gewöhnliches Kriging nicht-linear transformierter Daten (Deutsch u. Journel (1998), S. 75). Im Gegensatz zu den linearen Krigingverfahren dienen sie meist nicht zur Prognose der Realisation  $z(\mathbf{s}_0)$  des Prozesses an einer Stelle sondern zur Schätzung der bedingten Wahrscheinlichkeit, daß der Prozeß an einer Stelle einen Schwellenwert  $z_c$  nicht überschreitet:

$$P[Z(\mathbf{s}_0) \leq z_c \mid \mathbf{Z} = \mathbf{z}]. \quad (2.2.18)$$

Die einfachste nicht-lineare Transformation ist die Binärtransformation

$$I(\text{Bedingung}) = \begin{cases} 1 & \text{wenn die Bedingung erfüllt ist,} \\ 0 & \text{sonst.} \end{cases}$$

Journel (1983) schlägt das auf ihr beruhende *Indikator-Kriging* vor, um bedingte Wahrscheinlichkeiten zu schätzen. Dazu werden die Indikator-Prozesse  $I(z_c) = \{I(Z(\mathbf{s}) \leq z_c) : \mathbf{s} \in D\}$  definiert. Diese Prozesse besitzen in Abhängigkeit vom jeweiligen Schwellenwert  $z_c$  eine eigene räumliche Abhängigkeit.

Der Vektor der transformierten Daten

$$\mathbf{i} = [I(z(\mathbf{s}_1) \leq z_c), \dots, I(z(\mathbf{s}_N) \leq z_c)]^T$$

ist eine Realisation des über  $\mathbf{s}_{(N)}$  definierten Vektors

$$\mathbf{I} = [I(Z(\mathbf{s}_1) \leq z_c), \dots, I(Z(\mathbf{s}_N) \leq z_c)]^T;$$

der aus dem Datenvektor berechnete Indikator-Prädiktor  $\hat{I}(Z(\mathbf{s}_0) \leq z_c)$  ist nach Journel (1983), S. 459, (41) eine Schätzung der bedingten Wahrscheinlichkeit 2.2.18, nach Cressie (1991), S. 282, eine Schätzung der bedingten Wahrscheinlichkeit

$$P[Z(\mathbf{s}_0) \leq z_c \mid \mathbf{I} = \mathbf{i}]$$

und damit eine möglicherweise schlechte Näherung für 2.2.18.

Werden verschiedene Schwellenwerte  $z_c; c = 1, \dots, C$  betrachtet, muß für jeden Schwellenwert die räumliche Abhängigkeit neu geschätzt werden und es ist möglich, daß  $\hat{I}(Z(\mathbf{s}_0) \leq z_i) > \hat{I}(Z(\mathbf{s}_0) \leq z_j)$ , obwohl  $z_i < z_j$ , was die Interpretation der Prognosen als Wahrscheinlichkeiten verbietet. Vermeiden läßt sich dies durch die Annahme, daß die räumlichen Abhängigkeiten zu allen Schwellenwerten Vielfache einer gemeinsamen Funktion sind; das *Mosaik Modell* garantiert beispielsweise solche proportionalen räumlichen Abhängigkeiten (vergleiche Chilès u. Delfiner (1999), S. 384f). Wird die gemeinsame räumliche Abhängigkeit als die des Indikator-Prozesses zum Median, für den die Hälfte der transformierten Daten Null ist, geschätzt, sprechen Deutsch u. Journel (1998), S. 78f von *Median-Indikator-Kriging*.

Die Forderung proportionaler räumlicher Autokorrelationen ist eine ausgesprochen strenge, die Alternative ein Indikator-Kriging mit getrennten Schätzungen der räumlichen Abhängigkeiten zu jedem betrachteten Schwellenwert und anschließenden Korrekturen, um die Ordnungsrelationen

$$\hat{I}(Z(\mathbf{s}_0) \leq z_i) < \hat{I}(Z(\mathbf{s}_0) \leq z_j) \quad \forall z_i < z_j$$

zu garantieren.

Werden beim Indikator-Kriging bei der Schätzung zu einem Schwellenwert  $z_c$  nicht nur die zu diesem sondern die zu allen  $C$  Schwellenwerten transformierten Daten und neben der Kovarianzfunktion des Indikatorprozesses  $I(z_c)$  auch die  $C - 1$  Kreuzkovarianzfunktionen zwischen  $I(z_c)$  und den Indikatorprozessen zu allen anderen Schwellenwerten berücksichtigt, sprechen wir von *Indikator-Cokriging* (vergleiche Chilès u. Delfiner (1999), S. 386f). Für Mosaik-Prozesse sind (Indikator-)Kriging und (Indikator-)Cokriging identisch (vergleiche Chilès u. Delfiner (1999), S. 384).

*Disjunktives Kriging* ist ebenfalls ein Kriging nicht-linear transformierter Daten, allerdings sind beliebige Transformationen  $f(\cdot)$  möglich. Sein Prädiktor ist

$$p_{dk}(\mathbf{s}_0, \mathbf{Z}) = \sum_{n=1}^N f_n(Z(\mathbf{s}_n)) \quad \forall n = 1, \dots, N,$$

wobei

$$f_n(Z(\mathbf{s}_n)) = \sum_k f_{n,k} \cdot I(f_n(Z(\mathbf{s}_n)) = f_{n,k}).$$

Disjunktives Kriging eines Indikators ist daher gleichbedeutend mit vollständigem Indikator-Cokriging (vergleiche Rivoirard (1994), S. 19).

Disjunktives Kriging ist benötigt nach Cressie (1991), S. 279, die Kenntnis der bivariaten Verteilungen

$$F_{Z(\mathbf{s}_n), Z(\mathbf{s}_m)}(z(\mathbf{s}_n), z(\mathbf{s}_m)) = P[Z(\mathbf{s}_n) \leq z(\mathbf{s}_n), Z(\mathbf{s}_m) \leq z(\mathbf{s}_m)].$$

Da diese für gewöhnlich nicht bekannt sind, unterstellen wir in der praktischen Umsetzung daher ein Modell für die bivariaten Verteilungen (vergleiche Rivoirard (1994), S. 22) – beziehungsweise für ihre Transformationen.

Gauß'sches Disjunktives Kriging unterstellt gauß'sche bivariate Verteilungen und nutzt die Tatsache, daß sich mittels *Hermitescher Polynome*  $H_k(\cdot)$  durch

$$f(Z(\mathbf{s}_n)) = \sum_{k=0}^{\infty} f_k H_k(Z(\mathbf{s}_n))$$

beliebige Funktionen einer Variablen  $Z(\mathbf{s}_n)$  ausdrücken lassen. Speziell läßt sich eine Variable  $Z(\mathbf{s}_n)$  ausdrücken als eine Summe von Hermiteschen Polynomen einer standardnormalverteilten Variablen  $Y(\mathbf{s}_n)$ :

$$Z(\mathbf{s}_n) = \Phi(Y(\mathbf{s}_n)) = \sum_{k=0}^{\infty} \phi_k H_k(Y(\mathbf{s}_n)),$$

wobei in der Praxis  $k$  nach oben begrenzt ist. Webster u. Oliver (2001), Kapitel 10.5 nutzen  $H_0$  bis  $H_7$ , Rivoirard (1994), S. 43, empfiehlt für die meisten Fälle weniger als ein Dutzend Hermitesche Polynome. Obige Transformation ist mit  $Y(\mathbf{s}_n) = \Phi^{-1}(Z(\mathbf{s}_n))$  invertierbar, falls  $\Phi$  injektiv ist; ich werde später die Transformation einer Variablen in eine standardnormalverteilte Variable mittels Hermitescher Polynome wie in der Literatur üblich als *Gauß'sche Anamorphose* bezeichnen.

Da die Hermiteschen Polynome untereinander unkorreliert sind (vergleiche Rivoirard (1994), S. 40), vereinfacht sich Gauß'sches Disjunktives Kriging dann von Cokriging zu Kriging (vergleiche Rivoirard (1994), S. 43). Der Prädiktor des Gauß'schen Disjunktiven Krigings ergibt sich aus Rivoirard (1994), S. 43, (5.6), unter Berücksichtigung von  $H_0(Y(\mathbf{s}_0)) = 1$  (vergleiche Rivoirard (1994), S. 40) als

$$\hat{Z}_{dk}(\mathbf{s}_0) = \sum_k \phi_k \hat{H}_k(Y(\mathbf{s}_0)) \quad (2.2.19)$$

Die Hermiteschen Polynome  $H_k(Y(\mathbf{s}_0))$  werden dabei durch

$$\hat{H}_k(Y(\mathbf{s}_0)) = \sum_{n=1}^N \lambda_{n,k} H_k(Y(\mathbf{s}_n))$$

geschätzt, wobei die Gewichte  $\lambda_{n,k}$  die Lösungen des Krigingssystemes

$$\sum_{n=1}^N \lambda_{n,k} \text{cov}(H_k(Y(\mathbf{s}_n)), H_k(Y(\mathbf{s}_m))) = \text{cov}(H_k(Y(\mathbf{s}_n)), H_k(Y(\mathbf{s}_0))) \quad \forall m = 1, \dots, N$$

sind.

### 2.2.5. Prognosegüte

Die Güte von Prognosen kann durch Kreuzvalidierungsverfahren beurteilt werden. Das einfachste Verfahren teilt dabei die  $N$  Orte, an denen die Meßwerte erhoben wurden, in zwei disjunkte Mengen  $\mathbb{K}$  und  $\mathbb{M}$ ;  $\mathbb{K}$  ist dabei vom Umfang  $|\mathbb{K}| = K$  und  $\mathbb{M}$  vom Umfang  $|\mathbb{M}| = N - K$ . Aus der Modellierungsstichprobe  $\{z(\mathbf{s}) : \mathbf{s} \in \mathbb{M}\}$  werden nun für ein Prognoseverfahren  $v$  die Prognosen  $\{\hat{z}_v(\mathbf{s}) : \mathbf{s} \in \mathbb{K}\}$  berechnet.

Naheliegende Maße zur Beurteilung der Prognosegüte sind nun der mittlere absolute Vorhersagefehler,

$$MEP_v = \frac{1}{K} \sum_{k=1}^K (z(\mathbf{s}_k) - \hat{z}_v(\mathbf{s}_k)) \quad (2.2.20)$$

und der mittlere quadratische Vorhersagefehler

$$MSEP_v = \frac{1}{K} \sum_{k=1}^K \{z(\mathbf{s}_k) - \hat{z}_v(\mathbf{s}_k)\}^2. \quad (2.2.21)$$

Das Verhältnis

$$BR_v = \frac{\{MEP_v\}^2}{MSEP_v}$$

aus diesen beiden Fehlern gibt den Anteil des Bias am mittleren quadratischen Vorhersagefehler an.

Da 2.2.21 quadratisch auf einzelne große Fehler reagiert, ist, wenn mehrere Verfahren angewendet werden, der mittlere Rang des absoluten Vorhersagefehlers ein stabileres Maß für die Prognosegüte eines Verfahrens:

$$MR_v = \frac{1}{K} \sum_{k=1}^K r_{v,k}, \quad (2.2.22)$$

wobei  $r_{v,k}$  den Rang des absoluten Vorhersagefehlers  $|z(\mathbf{s}_k) - \hat{z}_v(\mathbf{s}_k)|$  des  $v$ -ten der betrachteten Verfahren bei der Prognose von  $z(\mathbf{s}_k)$  bezeichnet. Die Standardabweichung des Ranges des absoluten Vorhersagefehlers,

$$SDR_v = \sqrt{\frac{1}{K-1} \sum_{k=1}^K \{r_{v,k} - MR_v\}^2},$$

gibt dabei an, wie konsistent die Beurteilung durch 2.2.22 über alle  $K$  Punkte ist.

### 2.3. Bedingte Verteilungen

Der Prozeß 2.1.3 ist nach Cressie (1991), S. 52 durch seine endlich-dimensionalen Randverteilungsfunktionen

$$F_{Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_m)}(z(\mathbf{s}_1), \dots, z(\mathbf{s}_m)) = P[Z(\mathbf{s}_1) \leq z(\mathbf{s}_1), \dots, Z(\mathbf{s}_m) \leq z(\mathbf{s}_m)],$$

$$m \geq 1,$$

$$z(\mathbf{s}_i) \in \mathbb{R}, n = 1, \dots, m$$

definiert. Die eindimensionale Randverteilungsfunktion des Prozesses an einer Stelle  $\mathbf{s}_0$  ist dann gegeben durch

$$F_{Z(\mathbf{s}_0)}(z(\mathbf{s}_0)) = P[Z(\mathbf{s}_0) \leq z(\mathbf{s}_0)], \quad \forall z(\mathbf{s}_0) \in \mathbb{R}$$

und die durch die Kenntnis der Stichprobe  $z_{(N)}$ , also der Realisation des Teilprozess  $Z_{(N)}$ , bedingte eindimensionale Randverteilungsfunktion ist dann

$$F_{Z(\mathbf{s}_0)|\mathbf{Z}=\mathbf{z}}(z(\mathbf{s}_0)) = P[Z(\mathbf{s}_0) \leq z(\mathbf{s}_0) | \mathbf{Z} = \mathbf{z}] \quad \forall z(\mathbf{s}_0) \in \mathbb{R}, \quad (2.3.1)$$

mit  $\mathbf{z}=[z(\mathbf{s}_1), \dots, z(\mathbf{s}_N)]^T$  und  $\mathbf{Z}=[Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_N)]^T$  (siehe Kapitel 2.1). Im folgenden werde ich die eindimensionale Randverteilungsfunktion des Prozesses an einer Stelle  $\mathbf{s}_0$  als *unbedingte lokale Verteilungsfunktion* bezeichnen; wenn sie durch die Stichprobe bedingt ist, werde ich sie als *bedingte lokale Verteilungsfunktion* bezeichnen.

Für eine beliebige Zahl  $m = K$  ist

$$F_{Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_K)}(z(\mathbf{s}_1), \dots, z(\mathbf{s}_K)) = P[Z(\mathbf{s}_1) \leq z(\mathbf{s}_1), \dots, Z(\mathbf{s}_K) \leq z(\mathbf{s}_K)];$$

$$\mathbf{s}_k : k = 1, \dots, K \in D$$

die unbedingte multivariate Randverteilungsfunktion und die bedingte multivariate Randverteilungsfunktion wird durch

$$F_{Z(\mathbf{s}_{N+1}), \dots, Z(\mathbf{s}_{N+K})|\mathbf{Z}=\mathbf{z}}(z(\mathbf{s}_{N+1}), \dots, z(\mathbf{s}_{N+K})) =$$

$$P[Z(\mathbf{s}_{N+1}) \leq z(\mathbf{s}_{N+1}), \dots, Z(\mathbf{s}_{N+K}) \leq z(\mathbf{s}_{N+K}) | \mathbf{Z} = \mathbf{z}];$$

$$\mathbf{s}_n : n = 1, \dots, N \in D, \mathbf{s}_k : k = N + 1, \dots, N + K \in D \quad (2.3.2)$$

definiert.

#### 2.3.1. Anwendungen bedingter lokaler Verteilungsfunktionen

– Die naheliegendste Anwendung der bedingten lokalen Verteilungsfunktion 2.3.1 ist die Berechnung der bedingten Wahrscheinlichkeit, daß der Prozess an einer Stelle einen Schwellenwert  $z_c$  nicht überschreitet: sie ergibt sich direkt aus 2.3.1 durch  $z(\mathbf{s}_0)=z_c$ . Die zugehörige Überschreitungswahrscheinlichkeit ist

$$P[Z(\mathbf{s}_0) > z_c | \mathbf{Z} = \mathbf{z}] = 1 - F_{Z(\mathbf{s}_0)|\mathbf{Z}=\mathbf{z}}(z_c).$$

Wir sind daher in der Lage, Gebiete auszuweisen, für die ein Grenzwert mit einer gegebenen Wahrscheinlichkeit  $1 - \alpha$  überschritten wird:

$$\{\mathbf{s} \in D : P[Z(\mathbf{s}) > z_c | \mathbf{Z} = \mathbf{z}] \leq \alpha\}.$$

- Exakte Prognoseintervalle zum Niveau  $\alpha$  für den wahren Wert  $z(\mathbf{s}_0)$ ,

$$I_{\alpha, Z=z}(\mathbf{s}_0) = [z_a; z_b] : P[z_a \leq Z(\mathbf{s}_0) \leq z_b | \mathbf{Z} = \mathbf{z}] = 1 - \alpha$$

lassen sich durch Lösung von

$$\alpha/2 = \int_{-\infty}^{z_a} f_{Z(\mathbf{s}_0)|\mathbf{Z}=\mathbf{z}}(z) dz \Leftrightarrow z_a = F_{Z(\mathbf{s}_0)|\mathbf{Z}=\mathbf{z}}^{-1}(\alpha/2) \text{ und}$$

$$\alpha/2 = \int_{z_b}^{\infty} f_{Z(\mathbf{s}_0)|\mathbf{Z}=\mathbf{z}}(z) dz \Leftrightarrow z_b = F_{Z(\mathbf{s}_0)|\mathbf{Z}=\mathbf{z}}^{-1}(1 - \alpha/2)$$

bestimmen (vergleiche Mood u. a. (1974), Kapitel 4.2).

- Die Kenntnis der bedingten lokalen Verteilungsfunktion beinhaltet die Kenntnis des optimalen Prädiktors

$$E[Z(\mathbf{s}_0) | \mathbf{Z}] = \int_{-\infty}^{\infty} z(\mathbf{s}_0) f_{Z(\mathbf{s}_0)|\mathbf{Z}}(z(\mathbf{s}_0)) dz(\mathbf{s}_0)$$

und seines bedingten mittleren quadratischen Prognosefehlers, der bedingten Varianz (siehe Kapitel 2.2),

$$var[Z(\mathbf{s}_0) | \mathbf{Z}] = \int_{-\infty}^{\infty} \{E[Z(\mathbf{s}_0) | \mathbf{Z}] - z(\mathbf{s}_0)\}^2 f_{Z(\mathbf{s}_0)|\mathbf{Z}}(z(\mathbf{s}_0)) dz(\mathbf{s}_0).$$

- Darüber hinaus werden durch Kriging approximierte bedingte lokale Verteilungsfunktionen bei sequentiellen Simulationsverfahren benötigt (siehe Kapitel 2.4).

### 2.3.2. Approximation bedingter lokaler Verteilungsfunktionen

Jedes Krigingverfahren läßt sich zur Approximation bedingter lokaler Verteilungsfunktionen nutzen. Bei den linearen Verfahren wird dabei der bedingte Erwartungswert durch die Prognose und die bedingte Varianz durch einen mittleren quadratischen Prognosefehler ersetzt; sie benötigen daher die Annahme einer Verteilung, die durch die bedingten ersten beiden Momente ihrer Variablen vollständig definiert ist.

Im einfachsten Fall nehmen wir an, daß der Prozeß  $Z$  gauß'sch ist. Sind dann Erwartungswert- und Kovarianzfunktion bekannt, liefert Einfaches Kriging die bedingten ersten beiden Momente der Variablen  $Z(\mathbf{s}_0)$  und die bedingte lokale Verteilungsfunktion folgt einer Normalverteilung (vergleiche Stein (1999), S. 3):

$$F_{Z(\mathbf{s}_0)|\mathbf{Z}=\mathbf{z}}(z(\mathbf{s}_0)) \sim N(\hat{z}_{sk}(\mathbf{s}_0), \sigma_{sk}^2(\mathbf{s}_0)). \quad (2.3.3)$$

Da die Erwartungswertfunktion für gewöhnlich unbekannt ist, böte sich trotz 2.2.12

$$\hat{F}_{Z(\mathbf{s}_0)|\mathbf{Z}=\mathbf{z}}(z(\mathbf{s}_0)) \sim N(\hat{z}_k(\mathbf{s}_0), \sigma_{k|\mathbf{Z}=\mathbf{z}}^2(\mathbf{s}_0)), k = uk, ok$$

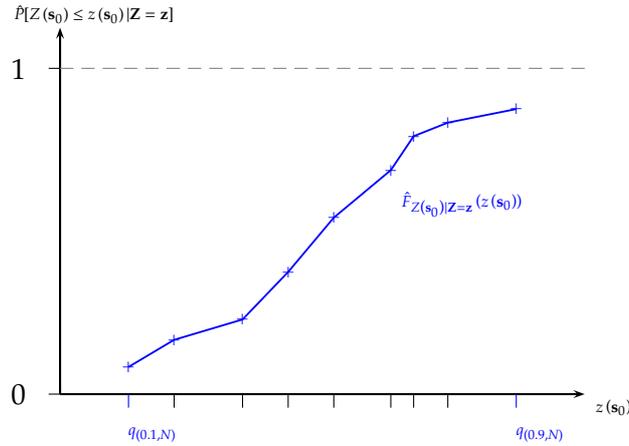


Abbildung 2.2.: Approximation einer bedingten lokalen Verteilungsfunktion.

als Approximation für  $F_{Z(\mathbf{s}_0)|Z=\mathbf{z}}(z(\mathbf{s}_0))$  an. Es ist aber  $\sigma_{k|Z=\mathbf{z}}(\mathbf{s}_0)$  im Regelfall ebenfalls unbekannt ist, weshalb der bedingte durch den unbedingten mittleren quadratischen Prognosefehler ersetzt wird, das heißt in der Regel

$$\hat{F}_{Z(\mathbf{s}_0)|Z=\mathbf{z}}(z(\mathbf{s}_0)) \sim N(\hat{z}_{uk}(\mathbf{s}_0), \sigma_{uk}^2(\mathbf{s}_0)),$$

beziehungsweise bei konstanter Erwartungswertfunktion

$$\hat{F}_{Z(\mathbf{s}_0)|Z=\mathbf{z}}(z(\mathbf{s}_0)) \sim N(\hat{z}_{ok}(\mathbf{s}_0), \sigma_{ok}^2(\mathbf{s}_0)). \quad (2.3.4)$$

Die Approximation 2.3.4 werde ich von hieran den *Polfeldt-Ansatz* nennen, da ich sie in Anwendung erstmals bei Polfeldt (1999) sah, der in Anlehnung an die – von mir cressieschen genannten – Prognoseintervalle aus Cressie (1991), S. 155 minimale empfehlenswerte Abstände  $|z_b - z_a|$  zwischen zwei aufeinanderfolgenden Schwellenwerten herleitet.

Wird keine Annahme über die Art der Verteilung des zugrundeliegenden Prozesses getroffen, so können wir bedingte lokale Verteilungsfunktionen approximieren, indem wir bedingte lokale Unterschreitungswahrscheinlichkeiten zu möglichst vielen Schwellenwerten mit einem nicht-linearen Kriging prognostizieren und zwischen diesen prognostizierten Funktionswerten der bedingte lokale Verteilungsfunktion linear interpolieren. In Abbildung 2.2 wurde für eine Stelle  $\mathbf{s}_0$  zu neun Schwellenwerten  $q_{(0.1,N)}, \dots, q_{(0.9,N)}$ , die den Quantilen der empirischen Verteilung der  $N$  Meßwerte entsprachen, Unterschreitungswahrscheinlichkeiten durch Indikator-Kriging (siehe Kapitel 2.2.4) approximiert.

Die neun Kreuze stellen also  $\hat{F}_{Z(\mathbf{s}_0)|Z=\mathbf{z}}(q_{(0.i,N)})$ ;  $i = 1, \dots, 9$  dar, während die sie verbindende Linie eine weitere lineare Interpolation ist.

Da ich im Rahmen dieser Arbeit nur approximierte Werte von Verteilungsfunktionen an Stellen  $q_{(0.1,N)} \leq z(\mathbf{s}_0) \leq q_{(0.9,N)}$  betrachte, benötige ich keine

Extrapolation, um die Schwänze der Verteilung zu approximieren. Vorschläge zur Extrapolation finden sich bei Deutsch u. Journal (1998).

### Beurteilung approximierter bedingter lokaler Verteilungsfunktionen

Die vorgestellten Verfahren zur Approximation bedingter lokaler Verteilungsfunktionen liefern keine Angaben über die Güte der Approximation der bedingten lokalen Verteilungsfunktion mit der Methode  $v$ , oder über die Genauigkeit einzelner approximierter Wahrscheinlichkeiten

$$\hat{P}_v [Z(\mathbf{s}_0) \leq z(\mathbf{s}_0) \mid \mathbf{Z} = \mathbf{z}] = \hat{F}_{v, Z(\mathbf{s}_0) \mid \mathbf{Z} = \mathbf{z}}(z(\mathbf{s}_0)).$$

Falls für eine Menge  $\mathbb{K} = \{\mathbf{s}_k; k = 1, \dots, K\}$  die bedingten lokalen Verteilungsfunktionen

$$F_{Z(\mathbf{s}_k) \mid \mathbf{Z} = \mathbf{z}}(z(\mathbf{s}_k))$$

bekannt sind, bieten sich die in Kapitel 2.2.5 vorgestellten Maße zur Beurteilung der Vorhersagegenauigkeit für einzelne Wahrscheinlichkeiten an, indem  $z(\mathbf{s}_k) - \hat{z}_v(\mathbf{s}_k)$  durch  $F_{Z(\mathbf{s}_k) \mid \mathbf{Z} = \mathbf{z}}(z(\mathbf{s}_k)) - \hat{F}_{v, Z(\mathbf{s}_k) \mid \mathbf{Z} = \mathbf{z}}(z(\mathbf{s}_k))$  ersetzt wird. Dies ist beispielsweise möglich, wenn der zugrundeliegende Prozeß ein gauß'scher Zufallsprozess mit bekannter Erwartungswertfunktion ist und die bedingte lokale Verteilungsfunktion 2.3.3 alle Wahrscheinlichkeiten für beliebige  $\mathbf{s}_0$  liefert.

In praktischen Anwendungen sind die bedingten lokalen Verteilungsfunktionen nicht bekannt; um die Genauigkeit der approximierten bedingten lokalen Verteilungsfunktionen dennoch einschätzen zu können, berechnen wir *Überdeckungswahrscheinlichkeiten*  $f_v(p)$  (Papritz u. Moyeed (1999), S. 324). Wenn  $\hat{q}_v(\mathbf{s}_k; p)$  das  $p$ -Quantil der approximierten bedingten lokalen Verteilungsfunktion an der Stelle  $\mathbf{s}_k$  ist, wenn also angenommen wird, daß

$$P [Z(\mathbf{s}_k) \leq \hat{q}_v(\mathbf{s}_k; p) \mid \mathbf{Z} = \mathbf{z}] = p,$$

so sollte

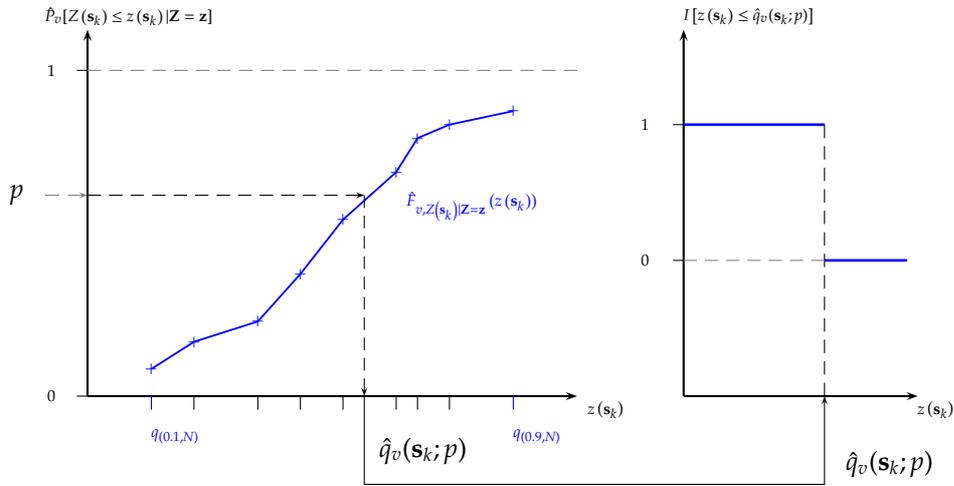
$$f_v(p) = \frac{1}{K} \sum_{k=1}^K I [z(\mathbf{s}_k) \leq \hat{q}_v(\mathbf{s}_k; p)],$$

der Anteil der Beobachtungen, die tatsächlich kleiner als die zugehörigen geschätzten Quantile sind, mit  $p$  möglichst gut übereinstimmen. Abbildung 2.3 zeigt die Ermittlung des Indikators in obiger Gleichung für eine Stelle  $\mathbf{s}_k$  und ein nicht-lineares Verfahren.

Vergleiche eines Streudiagrammes von  $f_v(p)$  über  $p$  mit der ersten Winkelhalbierenden geben Auskunft über die Güte eines Verfahrens: Wenn die bedingten lokalen Verteilungsfunktionen im Mittel gut approximiert sind, ist  $f_v(p)$  ungefähr gleich  $p$  für alle  $p$ .

## 2.4. Bedingte Simulationen

Berechnen wir aus den Meßwerten  $z_{(N)}$  (Kreise in Abbildung 2.4) die Krigingprognosen nicht nur für eine sondern für viele Prognosestellen  $s_{(K)} =$



**Abbildung 2.3.:** Bestimmung eines Inkinators  $I[z(\mathbf{s}_k) \leq \hat{q}_v(\mathbf{s}_k; p)]$  bei der Berechnung einer Überdeckungswahrscheinlichkeit.

$\{\mathbf{s}_k \in D, k = 1, \dots, K\}$ , erhalten wir eine Menge  $\hat{z}_{(K)} = \{\hat{z}(\mathbf{s}) : \mathbf{s} \in s_{(K)}\}$  (punktirierte Linie in Abbildung 2.4). Diese enthalt eine Prognose fur jedes Element der partiellen Realisation  $z_{(K)} = \{z(\mathbf{s}) : \mathbf{s} \in s_{(K)}\}$  (Kreuze in Abbildung 2.4).

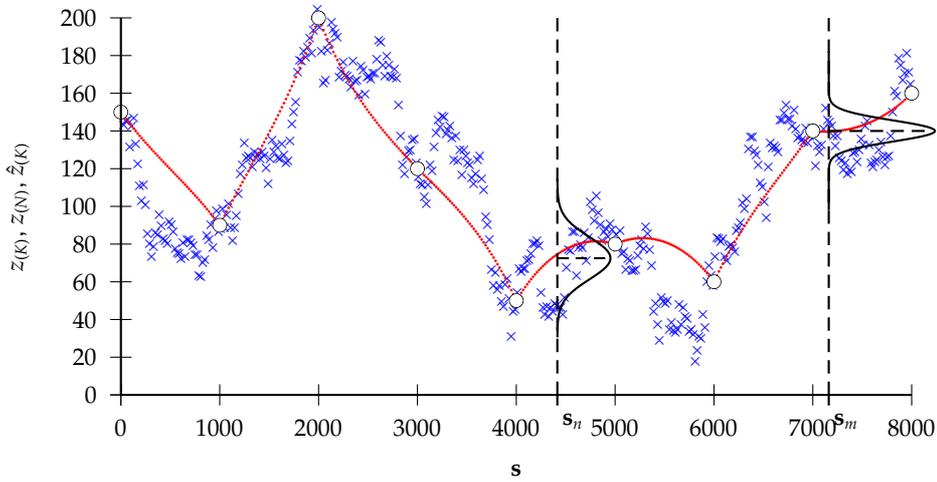
Stellen wir uns vor, in Abbildung 2.5 sei das Hohenprofil eines Sollingtransektes als dicke Linie dargestellt. Ein Radfahrer, der die neun bekannten Hohen zum Kriging nutzte und die Prognosen, also die Punkte, fur die zuruckzulegende Strecke hielt, erlebte eine bose uberraschung: Die wahre Strecke ware weitaus langer und der Radfahrer legte mehr Hohenmeter als vorhergesagt zuruck.

Die Lange der Linie, welche die Prognosen verbindet, scheint ein schlechter Naherungswert fur die wahre Strecke zu sein, besser scheinen die beiden dunnen Linien in Abbildung 2.5 die Struktur des Hohenprofils wiederzugeben. Wahrend die gebrochene Linie einfach eine weitere Realisation  $z_{(K)}^1$  des selben Prozesses, dessen Realisation auch das Hohenprofil ist, darstellt, gilt fur  $z_{(K)}^2$ , die ungebrochene dunne Linie:  $z^2(\mathbf{s}_1) = z(\mathbf{s}_1), \dots, z^2(\mathbf{s}_N) = z(\mathbf{s}_N)$ . Die Realisation  $z_{(K)}^1$  ist Ergebnis einer unbedingten Simulation, also einer zufalligen Ziehung aus

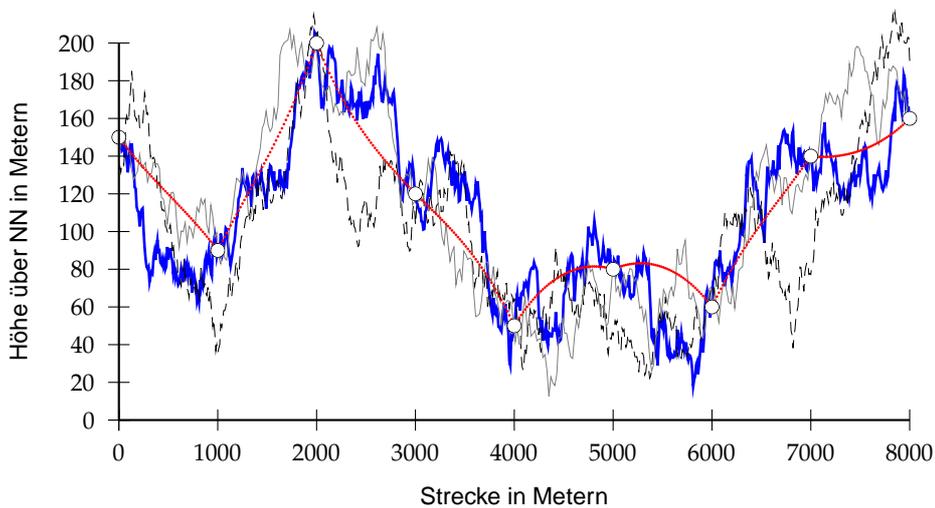
$$F_{Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_K)}(z(\mathbf{s}_1), \dots, z(\mathbf{s}_K)) = P[Z(\mathbf{s}_1) \leq z(\mathbf{s}_1), \dots, Z(\mathbf{s}_K) \leq z(\mathbf{s}_K)],$$

$z_{(K)}^2$  stammt aus einer durch die Stichprobe bedingten Simulation, ist also eine zufallige Ziehung aus 2.3.2 und daher  $z_{(K)}$  ahnlicher als  $z_{(K)}^1$  dies ist.

Bedingte Simulationen reproduzieren die Streuung des betrachteten Phanomens und sind daher geeignet, einen realistischen Eindruck der raumlichen Variabilitat zu vermitteln. Wameling (2002), Kapitel 2.7, bietet einen uberblick uber die verschiedenen Simulationsalgorithmen, diese lassen sich nach der Art,



**Abbildung 2.4.:** Aus den  $N = 9$  Meßwerten  $z_{(N)}$  (Kreise) berechnete Prognosen (punktirierte Linie)  $\hat{z}_{(K)}$  für eine partielle Realisation  $z_{(K)}$  (Kreuze) über einem Transekt  $D := \{s \in \mathbb{R} \mid 0 \leq |s| \leq 8000\}$  sowie bedingte lokale Dichtefunktionen für  $s_n$  und  $s_m$ .



**Abbildung 2.5.:** Höhenprofil einer Radstrecke (dicke Linie), bekannte Höhen (Kreise), an 800 Stellen vorhergesagtes (punktirierte Linie), bedingt (dünne Linie) und unbedingt (dünne gebrochene Linie) simuliertes Höhenprofil.

in der die Bedingung umgesetzt wird, in drei Klassen einteilen: Direkt bedingte, durch Kriging bedingte und sequentielle Verfahren (die Chilès u. Delfiner (1999), nach Tabelle 7.1 zu den direkten Verfahren zählen).

### Die direkt bedingten Verfahren

benötigen die Kenntnis der bedingten gemeinsamen Verteilungsfunktion 2.3.2 (vergleiche Chilès u. Delfiner (1999), S. 452), sie wird meist als eine multivariat Gauß'sche angenommen.

### Die durch Kriging bedingten Algorithmen

sind bei Cressie (1991), S. 207ff näher beschreiben. Sie benötigen Einfaches Kriging sowie unbedingte Simulationen und daher die Kenntnis der unbedingten gemeinsamen Verteilung (vergleiche Chilès u. Delfiner (1999), S. 452) welche wiederum zumeist als gauß'sch angenommen wird und daher durch Erwartungswertfunktion und räumliche Abhängigkeit definiert ist (vergleiche Johnson u. Kotz (1972), S. 40). Eine durch Kriging bedingte Simulation des Prozesses an der Stelle  $s_0$  entspricht dem Ziehen aus der Verteilung, die der unbedingten Simulation zugrundeliegt und deren Erwartungswert<sup>5</sup> gleich dem Prädiktor des Einfachen Krigings und deren Varianz gleich dem mittleren quadratischen Prognosefehler des Einfachen Krigings ist (Wameling (2002), S. 31).

### Die sequentiellen Verfahren

faktorisieren (vergleiche Goovaerts (1997), Kapitel 8.2) die bedingte gemeinsame Verteilungsfunktion 2.3.2, so daß immer nur die durch die Daten und alle bisher simulierten Werte bedingte lokale Verteilungsfunktion (vergleiche Chilès u. Delfiner (1999), S. 452), an einer Stelle bekannt sein muß:

$$\begin{aligned} & F_{Z(\mathbf{s}_{N+1}), \dots, Z(\mathbf{s}_{N+K}) | Z(\mathbf{s}_1)=z(\mathbf{s}_1), \dots, Z(\mathbf{s}_N)=z(\mathbf{s}_N)} (z(\mathbf{s}_{N+1}), \dots, z(\mathbf{s}_{N+K})) \\ &= F_{Z(\mathbf{s}_{N+1}) | Z(\mathbf{s}_1)=z(\mathbf{s}_1), \dots, Z(\mathbf{s}_N)=z(\mathbf{s}_N)} (z(\mathbf{s}_{N+1})) \\ & \times F_{Z(\mathbf{s}_{N+2}) | Z(\mathbf{s}_1)=z(\mathbf{s}_1), \dots, Z(\mathbf{s}_{N+1})=z(\mathbf{s}_{N+1})} (z(\mathbf{s}_{N+2})) \\ & \times \dots \\ & \times F_{Z(\mathbf{s}_{N+K}) | Z(\mathbf{s}_1)=z(\mathbf{s}_1), \dots, Z(\mathbf{s}_{N+K-1})=z(\mathbf{s}_{N+K-1})} (z(\mathbf{s}_{N+K})) \end{aligned}$$

Für  $K = 1$  ist also die bedingte gemeinsame Verteilungsfunktion 2.3.2 gleich der bedingten lokalen Verteilungsfunktion 2.3.1:

$$F_{Z(\mathbf{s}_{N+1}) | Z(\mathbf{s}_1)=z(\mathbf{s}_1), \dots, Z(\mathbf{s}_N)=z(\mathbf{s}_N)} (z(\mathbf{s}_{N+1})).$$

Bei der *Sequentiellen Gauß'schen-Simulation* wird ein gauß'scher Prozess unterstellt, die bedingte lokale Verteilungsfunktion 2.3.1 ist dann eine Normalverteilungsfunktion mit Erwartungswert  $\hat{Z}_{sk}(\mathbf{s}_0)$  und Varianz  $\sigma_{sk}^2(\mathbf{s}_0)$  (siehe 2.2.13 und 2.2.17) und eine bedingte Simulation des Prozesses an der Stelle  $\mathbf{s}_0$  entspricht dem Ziehen aus der  $N(\hat{Z}_{sk}(\mathbf{s}_0), \sigma_{sk}^2(\mathbf{s}_0))$ -Verteilung.

---

<sup>5</sup>Variablen, nicht ihre Verteilungen besitzen Momente – allerdings begründen Mood u. a. (1974), S. 65 den Ausdruck des Momentes einer Verteilungsfunktion. Um sprachliche Unge-tümte wie „Ziehen aus der Verteilung, für die der Erwartungswert einer Variablen, die diese Verteilung besitzt, gleich vierzehn ist“ zu vermeiden, spreche ich manchmal von Momenten einer Verteilung. Gemeint sind immer die Momente der Variablen.

Bei der *Sequentiellen Indikator-Simulation* werden die bedingten lokalen Verteilungsfunktionen am jeweils betrachteten Prognoseort durch eine der Varianten des Indikator-Krigings approximiert, für einfaches Indikator-Kriging ist die bedingte Simulation des Prozesses an der betrachteten Stelle dann gleich dem Ziehen aus der durch Indikator-Kriging approximierten bedingten lokalen Verteilungsfunktion.

Die *Direkte Sequentielle Simulation mit Histogrammreproduktion* (vergleiche Oz u. a. (2003)) berechnet bedingte lokale Verteilungen auf Basis einer Quantiltransformation: Bezeichnet  $y$  eine standardnormalverteilte Variable und  $G(y)$  die Standardnormalverteilung  $N(0, 1)$ . Dann läßt sich eine beliebig gemäß einer Verteilungsfunktion  $F$  verteilte Variable  $z$  durch die Quantiltransformation  $y = G^{-1}(F(z))$  in eine standardnormalverteilte transformieren und diese Transformation durch  $z = F^{-1}(G(y))$  umkehren. Die Direkte Sequentielle Simulation mit Histogrammreproduktion berechnet nun für viele  $\mu_y \in [-3.5, 3.5]$ ,  $\sigma_y \in ]0, 2]$  und  $p^l \in [0, 1]$ ,  $l = 1, \dots, L$

$$z^l = F^{-1}\left(G\left(\mu_y + G^{-1}(p^l) * \sigma_y\right)\right).$$

Aus den  $z^l$  läßt sich für feste  $\mu_y, \sigma_y$  eine approximierte bedingte lokale Verteilungsfunktion  $\hat{F}_{\mu_y, \sigma_y}(z)$  aufstellen. Für eine zu simulierende Stelle  $\mathbf{s}_0$  werden  $\hat{Z}_{sk}(\mathbf{s}_0)$  und  $\sigma_{sk}^2(\mathbf{s}_0)$  berechnet und ein Wert aus jener approximierten bedingten lokalen Verteilungsfunktion gezogen, deren Mittelwert und Varianz  $\hat{Z}_{sk}(\mathbf{s}_0)$  und  $\sigma_{sk}^2(\mathbf{s}_0)$  am nächsten sind. Der simulierte Wert wird dann noch durch die Differenz zwischen den Momenten der approximierten bedingten lokalen Verteilungsfunktion und den Krigingwerten (vergleiche Oz u. a. (2003), Gleichung (4)) korrigiert. Dieser Ansatz geht also davon aus, daß einfaches Kriging angewandt werden kann und  $\hat{Z}_{sk}(\mathbf{s}_0)$  eine gute Näherung für  $E[Z(\mathbf{s}_0) | \mathbf{Z}]$  ist; von der Sequentiellen Gauß'schen-Simulation unterscheidet ihn also die Annahme, daß die bedingte lokale Verteilungsfunktion der Näherung  $\hat{F}_{\mu_y, \sigma_y}(z(\mathbf{s}_0))$  statt  $N(\hat{Z}_{sk}(\mathbf{s}_0), \sigma_{sk}^2(\mathbf{s}_0))$  entspricht. Nach Oz u. a. (2003), S. 47 entspricht  $\hat{F}_{\mu_y, \sigma_y}(z(\mathbf{s}_0))$  exakt der durch hermitesche Polynome und Disjunktives Kriging approximierten bedingten lokalen Verteilungsfunktion.

Gemeinsam ist allen Simulationsverfahren, daß sie Annahmen über die bedingte lokale Verteilungsfunktion treffen. Wiederholte Anwendung eines Simulationsalgorithmus für eine Stelle  $\mathbf{s}_0$  reproduziert diese angenommene bedingte lokale Verteilungsfunktion. Die Bedingtheit der lokalen Verteilungsfunktion wird entweder – wie bei den direkt bedingten und den sequentiellen Verfahren – direkt, oder – wie bei den durch Kriging bedingten Verfahren – indirekt durch ein Krigingverfahren hergestellt. Geostatistische Simulationsverfahren, die kein Kriging benutzen (LU-Decomposition und Direkte Sequentielle Simulation mit Histogrammreproduktion), treffen Annahmen, die identisch zur Approximation durch ein entsprechendes Krigingverfahren sind: LU-Decomposition unterstellt einen Gauß'schen Zufallsprozeß und die Kenntnis der Erwartungswertfunktion, die bedingte lokale Verteilungsfunktion folgt

## 2. Zielführende Darstellung geostatistischer Methoden

---

dann der Normalverteilung  $N(\hat{z}_{sk}(\mathbf{s}_0), \sigma_{sk}^2(\mathbf{s}_0))$  und ist durch Einfaches Kriging bekannt. Direkte Sequentielle Simulation mit Histogrammreproduktion führt eine gauß'sche Anamorphose und eine Rücktransformation durch und approximiert die bedingte lokale Verteilungsfunktion, welche sich durch ein entsprechendes Gauß'sches Disjunktives Kriging ergibt.

Bedingte geostatistische Simulationsverfahren sind daher Ziehungen aus bedingte lokale Verteilungsfunktion, die durch Krigingverfahren – oder exakt äquivalente Verfahren – approximiert wurden. Sie bieten daher bei der Approximation bedingter lokaler Verteilungsfunktionen gegenüber den Krigingverfahren keinen Vorteil – falls wir die Ziehung nicht oft genug wiederholen erhalten wir sogar nur eine grobe Näherung der durch das entsprechende Krigingverfahren approximierten bedingten lokalen Verteilungsfunktion.

### 3. Bedingte Vorhersagefehler in gauß'schen Prozessen

Wie wir aus Kapitel 2.3.2 wissen, ist im Spezialfall eines gauß'schen Prozesses die bedingte lokale Verteilungsfunktion die Normalverteilungsfunktion (2.3.3) jener Variablen, deren bedingter Erwartungswert dem Prädiktor des Einfachen Krigings (siehe 2.2.13) und deren bedingte Varianz dem mittleren quadratischen Prognosefehler dieses Prädiktors (siehe 2.2.17) entspricht. Da in der Anwendung Einfaches Kriging meist der unbekanntem Erwartungswertfunktion wegen nicht verfügbar ist und durch Universelles Kriging beziehungsweise – für konstante Erwartungswertfunktionen – durch Gewöhnliches Kriging ersetzt wird, liegt es nahe, die bedingte lokale Verteilungsfunktion durch die Normalverteilungsfunktion einer Variablen mit bedingtem Erwartungswert  $\hat{Z}_{ok}(\mathbf{s}_0)$  und bedingter Varianz  $\sigma_{ok|Z=z}^2(\mathbf{s}_0)$  zu approximieren. Da der bedingte mittlere quadratische Prognosefehler des Gewöhnlichen Krigings bei unbekanntem bedingtem Erwartungswert (siehe 2.2.2) ebenfalls unbekannt ist, wird er durch  $\sigma_{ok}^2(\mathbf{s}_0)$  substituiert und die bedingte lokale Verteilungsfunktion mit dem Polfeldt-Ansatz (siehe 2.3.4) genähert:

$$\hat{F}_{Z(\mathbf{s}_0)|Z=z}(z(\mathbf{s}_0)) \sim N(\hat{z}_{ok}(\mathbf{s}_0), \sigma_{ok}^2(\mathbf{s}_0)).$$

Um die durch das Ersetzen des bedingten durch den unbedingten mittleren quadratischen Prognosefehler entstehende Ungenauigkeit abzuschätzen, wird in Kapitel 3.1 der bedingte mittlere quadratische Prognosefehler des Gewöhnlichen Krigings und die Differenz  $\sigma_{ok}^2(\mathbf{s}_0) - \sigma_{ok|Z=z}^2(\mathbf{s}_0)$  unter der Normalverteilungsannahme hergeleitet und analysiert, um in Kapitel 3.2 zur Betrachtung der Unterschiede zwischen den bedingten und den unbedingten Prognoseintervallen,  $I_{\alpha, Z=z}^{OK}(\mathbf{s}_0)$  und  $I_{\alpha}^C(\mathbf{s}_0)$ , für Werte  $z(\mathbf{s}_0)$  an Punkten aus einer Menge  $S_{(0)}$  von Prognosepunkten zu dienen.

#### 3.1. Herleitung

Da der bedingte mittlere quadratische Prognosefehler, 2.2.2, eines Prädiktors  $p(\mathbf{s}_0, \mathbf{Z})$  vom bedingten Erwartungswert abhängt, läßt er sich im allgemeinen nicht bestimmen. Nehmen wir aber an, daß  $Z$  ein gauß'scher Zufallsprozeß ist, so ist der Prädiktor des Einfachen Krigings gleich dem bedingten Erwartungswert (siehe 2.2.13) und die Varianz des Einfachen Krigings gleich der bedingten Varianz (siehe 2.2.17). Dadurch läßt sich der bedingte mittlere quadratische Prognosefehler eines Prädiktors als

$$E \left[ \{p(\mathbf{s}_0, \mathbf{Z}) - Z(\mathbf{s}_0)\}^2 \mid \mathbf{Z} \right] = \sigma_{sk}^2(\mathbf{s}_0) + \{\hat{z}_{sk}(\mathbf{s}_0) - p(\mathbf{s}_0, \mathbf{Z})\}^2$$

### 3. Bedingte Vorhersagefehler in gauß'schen Prozessen

---

darstellen. Für Universelles Kriging ergibt sich durch Einsetzen von 2.2.14 und 2.2.9 der bedingte mittlere quadratische Prognosefehler

$$E \left[ \left\{ \hat{Z}_{uk}(\mathbf{s}_0) - Z(\mathbf{s}_0) \right\}^2 \mid \mathbf{Z} \right] = C(0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{c} \\ + \left\{ \left( \hat{\mu}_{gls}(\mathbf{s}_0) - \mu(\mathbf{s}_0) \right) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_{gls} - \boldsymbol{\mu}) \right\}^2 = \sigma_{uk|Z=\mathbf{z}}^2(\mathbf{s}_0).$$

Falls die Erwartungswertfunktion konstant über  $D$  (2.1.10 erfüllt) ist, vereinfacht sich dieser Ausdruck zum mittleren quadratischen Prognosefehler des Gewöhnlichen Krigings:

$$E \left[ \left\{ \hat{Z}_{ok}(\mathbf{s}_0) - Z(\mathbf{s}_0) \right\}^2 \mid \mathbf{Z} \right] = C(0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{c} \\ + \left\{ \left( 1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \right) (\hat{\mu} - \mu) \right\}^2 = \sigma_{ok|Z=\mathbf{z}}^2(\mathbf{s}_0). \quad (3.1.1)$$

Die Differenz zwischen unbedingtem und bedingtem mittleren quadratischen Prognosefehler des Gewöhnlichen Krigings, 2.2.15a - 3.1.1, ist

$$\left\{ \left( 1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \right) \right\}^2 E \left[ (\hat{\mu} - \mu)^2 \right] - \left\{ \left( 1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \right) (\hat{\mu} - \mu) \right\}^2 \\ = \left( 1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \right)^2 \left\{ E \left[ (\hat{\mu} - \mu)^2 \right] - (\hat{\mu} - \mu)^2 \right\}. \quad (3.1.2)$$

Der relative Fehler, den wir begehen, wenn wir den bedingten durch den unbedingten mittleren quadratischen Prognosefehler beim Gewöhnlichen Kriging ersetzen, ist also

$$\frac{\left( 1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \right)^2 \left\{ E \left[ (\hat{\mu} - \mu)^2 \right] - (\hat{\mu} - \mu)^2 \right\}}{C(0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{c} + \left( 1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} \right)^2 \cdot (\hat{\mu} - \mu)^2}. \quad (3.1.3)$$

Es soll nun versucht werden, diesen Fehler abzuschätzen.

**Der Faktor**  $\left\{ E \left[ (\hat{\mu} - \mu)^2 \right] - (\hat{\mu} - \mu)^2 \right\}$  bestimmt das Vorzeichen des Zählers von 3.1.3. Aus Kapitel 2.1.2 wissen wir, daß  $\hat{\mu}_{gls}$  eine gewichtete Summe  $\mathbf{a}^T \mathbf{Z}$  ist (siehe 2.1.15). Wenn  $\mathbf{Z}$  multivariat normalverteilt ist, folgt daher für die Verteilung von  $\hat{\mu}$  nach Anderson (1958), S. 19

$$\hat{\mu}_{gls} \sim N(\mathbf{a}^T \boldsymbol{\mu}, \mathbf{a}^T \boldsymbol{\Sigma} \mathbf{a})$$

und somit bei konstanter Erwartungswertfunktion nach 2.1.16

$$\hat{\mu} \sim N(\mu, (\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1})^{-1}) \text{ und} \\ \left( \frac{\hat{\mu} - \mu}{\sqrt{E \left[ (\hat{\mu} - \mu)^2 \right]}} \right)^2 \sim \chi_1^2.$$

Die Wahrscheinlichkeit, daß die quadratische Differenz zwischen dem geschätzten und dem wahren Erwartungswert eines schwach stationären gauß'schen Zufallsprozesses größer ist als die Erwartung dieser Differenz,  $P\left[(\hat{\mu} - \mu)^2 > E\left[(\hat{\mu} - \mu)^2\right]\right] = P\left[\frac{(\hat{\mu} - \mu)^2}{E\left[(\hat{\mu} - \mu)^2\right]} > 1\right] = P\left[\chi_1^2 > 1\right]$ , beträgt damit etwa 32%. Die Differenz 3.1.2 und mit ihr der relative Fehler 3.1.3 ist also mit einer Wahrscheinlichkeit von etwa 0.68 positiv, der bedingte ist dann kleiner als der unbedingte mittlere quadratische Prognosefehler.

**Der Faktor  $(1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1})^2$**

entspricht dem quadrierten Gewicht des Erwartungswertes beim Kriging: Wir können die linearen Prädiktoren 2.2.7, 2.2.8 und 2.2.10 als

$$\hat{p}_m(\mathbf{s}_0) = \mathbf{c}^T \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \mathbf{m}) + m(\mathbf{s}_0); \quad m(\mathbf{s}) = \mu(\mathbf{s}), \hat{\mu}_{gls}(\mathbf{s}), \hat{\mu}(\mathbf{s}) \quad \forall \mathbf{s} \in D$$

und bei konstanter Erwartungswertfunktion als

$$\mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z} + (1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}) m; \quad m = \mu, \hat{\mu} \quad (3.1.4)$$

darstellen. Hier wird deutlich, daß  $\lambda = (\mathbf{c}^T \boldsymbol{\Sigma}^{-1})^T$  den  $N$ -dimensionalen Vektor der Gewichte der Daten  $\mathbf{Z}$  und die Zahl  $(1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}) = 1 - \lambda^T \mathbf{1}$  (also „Eins minus Summe der Gewichte der Daten“) das Gewicht des geschätzten oder bekannten Erwartungswertes  $m$  angibt. Beide Ausdrücke sind Funktionen nur der Anordnung der Prognose- und Meßstellen sowie der räumlichen Abhängigkeit und lassen sich für gegebene Anordnungen und Autokorrelationen berechnen.

Der Wert dieses Gewichtes  $(1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1})$

- ist zwar grundsätzlich nicht auf das Intervall  $[0; 1]$  beschränkt, liegt aber meist in ihm (siehe Anhang A),
- ist klein, wenn die zur Prognose benutzten Daten (relativ zur Struktur der räumlichen Abhängigkeit) nahe beim Prognoseort  $\mathbf{s}_0$  liegen (vergleiche Rivoirard (1994), S. 67),
- und ist sogar oft nahe Null (vergleiche Rivoirard (1984), S. 67f).

Das Verhalten des Gewichtes des Erwartungswertes beim Kriging läßt sich durch die folgenden Überlegungen charakterisieren:

Je kleiner die Distanz  $\mathbf{s}_0 - \mathbf{s}_l$  zwischen dem Prognoseort und einem Datum  $Z(\mathbf{s}_l)$  wird, desto ähnlicher werden sich  $C(0)$  und  $C(\mathbf{s}_0 - \mathbf{s}_l)$ ; falls die Kovarianzfunktion keinen Nugget-Effekt aufweist, gilt sogar  $\lim_{(\mathbf{s}_0 - \mathbf{s}_l) \rightarrow 0} C(\mathbf{s}_0 - \mathbf{s}_l) = C(0)$ .

Geschieht dies, indem der Prognoseort an die  $l$ -te der fixen Meßstellen heranbewegt wird, verändern sich gleichzeitig die übrigen Distanzen  $\mathbf{s}_0 - \mathbf{s}_n \quad \forall \mathbf{s}_n \neq \mathbf{s}(l)$  kaum, wenn die übrigen Datenpunkte ausreichend weit von  $\mathbf{s}(l)$  entfernt liegen – wie es beispielsweise bei einem regelmäßigen Stichprobengitter der Fall ist. Dadurch geht das  $l$ -te Gewicht  $\lambda_l$  im Vektor  $\lambda = (\mathbf{c}^T \boldsymbol{\Sigma}^{-1})^T$  gegen 1, während

sich die übrigen Gewichte  $\lambda_n, n \neq l$  Null nähern. Dies führt dazu, daß einerseits  $\lambda^T \mathbf{1}$  immer ähnlicher 1 (und damit  $(1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1})$  immer ähnlicher Null) wird und daß sich andererseits  $\lambda^T \mathbf{c}$  in 2.2.14:

$$\sigma_{sk}^2(\mathbf{s}_0) = C(0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{c} = C(0) - \lambda^T \mathbf{c}$$

immer mehr an  $C(0)$  annähert. Ist schließlich  $\mathbf{s}_n = \mathbf{s}_l$ , wird das Gewicht  $\lambda_l = 1$ , alle anderen Gewichte und damit auch  $(1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1})$  werden gleich Null. Dann wird 3.1.4 zu

$$\hat{p}_{m,l}(\mathbf{s}_0, \mathbf{Z}) = 1 \cdot Z_l + (1 - 1) m$$

und gibt exakt  $Z(\mathbf{s}_0)$  wieder während die Krigingvarianzen aus Kapitel 2.2.2 gleich Null werden, was sich für Einfaches Kriging und Gewöhnliches Kriging durch

$$\sigma_{sk}^2(\mathbf{s}_0) = C(0) - \lambda^T \mathbf{c} = C(0) - C(0) = 0$$

und (siehe 2.2.16)

$$\sigma_{ok}^2(\mathbf{s}_0) = \sigma_{sk}^2(\mathbf{s}_0) + 0 \cdot (\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1})^{-1} \cdot 0$$

leicht zeigen läßt.

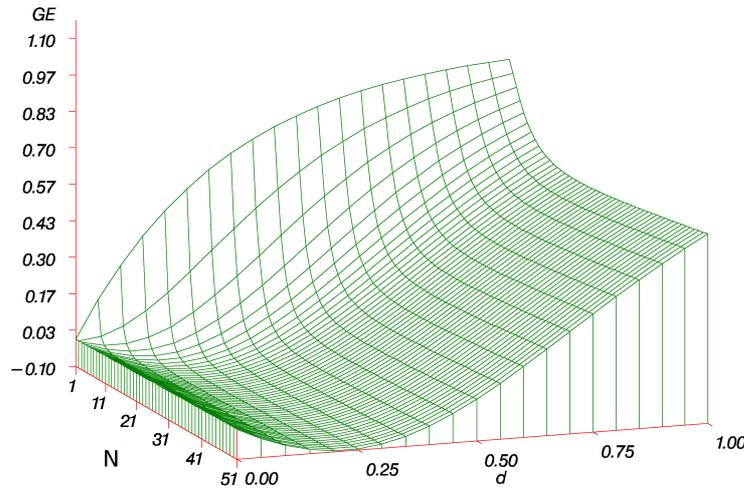
Abbildung 3.1 zeigt das Gewicht des Erwartungswertes (GE) für  $N$  Daten, die in gleichmäßigen Abständen auf einem Kreis mit Radius  $r$  um die Prognosestelle  $\mathbf{s}_0$  angeordnet sind, basierend auf einer isotropen exponentiellen Kovarianzfunktion

$$\text{cov}(|\mathbf{h}|) = c_p \cdot \exp\left(-\frac{3|\mathbf{h}|}{r_c}\right)$$

als Maß der räumlichen Abhängigkeit. Mit Abnahme der relativen Distanz  $d = \frac{r}{r_c}$  wird  $(1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1})$  rasch klein, wird zwischenzeitlich negativ und nähert sich dann Null; die Anzahl der zur Verfügung stehenden Daten nimmt ab einem Wert größer als 12 kaum noch Einfluß auf das Gewicht des Erwartungswertes. Rivoirard (1984), Kapitel 2-7, beobachtet für verschiedene Semivariogrammfunktionen und Meßstellenanordnungen, daß  $(1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1})$  abnimmt, solange Daten hinzugefügt werden, ohne daß negative Gewichte  $\lambda_n$  – beispielweise durch den „Screen Effekt“ (vergleiche Journal u. Huijbregts (1978), S. 346) bei dem Daten durch räumlich näherliegende Daten verdeckt werden – für einzelne Daten entstehen.

Es ist also anzunehmen, daß aufgrund des oft kleinen Gewichtes des Erwartungswertes beim Kriging – zumal es in 3.1.2 quadratisch eingeht – die Differenz  $\sigma_{ok}^2(\mathbf{s}_0) - \sigma_{ok|Z=z}^2(\mathbf{s}_0)$  für die meisten Anwendungen nahe Null liegt. Selbst wenn sie, bei größeren Werten für  $(1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1})$ , steigt, wächst mit dem Gewicht des Erwartungswertes beim Kriging auch der Nenner in 3.1.3, so daß der relative Fehler 3.1.3

$$\frac{\sigma_{ok}^2(\mathbf{s}_0) - \sigma_{ok|Z=z}^2(\mathbf{s}_0)}{\sigma_{ok|Z=z}^2(\mathbf{s}_0)}$$



**Abbildung 3.1.:** Gewicht des Erwartungswertes ( $GE$ ) über  $N$ , der Anzahl der Daten, und  $d$ , der Entfernung der Daten von der Prognosestelle relativ zur effektive Reichweite der räumlichen Abhängigkeit.

klein sein wird.

### Lineares Kriging in schwach stationären gauß'schen Prozessen.

Aus den vorangegangenen Betrachtungen des linearen Krigings in schwach stationären Prozessen und des Gewichtes des Erwartungswertes bei demselben ergibt sich eine Betrachtung des linearen Krigings in schwach stationären gauß'schen Prozessen, die bei dem Vergleich der unbedingten mit den exakten Prognoseintervallen im Kapitel 3.2 hilfreich sein kann. Analog zur Darstellung 3.1.4 des linearen Krigingprädiktors in schwach stationären Prozessen lassen sich die Varianzen des Einfachen und des Gewöhnlichen Krigings (2.2.14 und 2.2.15a) durch

$$\begin{aligned}\sigma_{sk}^2(\mathbf{s}_0) &= C(0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{c} + \left(1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}\right)^2 \cdot 0 \\ \sigma_{ok}^2(\mathbf{s}_0) &= C(0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{c} + \left(1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}\right)^2 \cdot E\left[(\hat{\mu} - \mu)^2\right],\end{aligned}\quad (3.1.5)$$

ausdrücken. Betrachten wir zusätzlich 3.1.1:

$$\sigma_{ok|Z=z}^2(\mathbf{s}_0) = C(0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{c} + \left(1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}\right)^2 \cdot (\hat{\mu} - \mu)^2,$$

den bedingten mittleren quadratischen Prognosefehler des Gewöhnlichen Krigings, ergibt sich folgende Interpretation: Die Gewöhnliche Krigingvarianz ist gleich der Prozeßvarianz abzüglich der Varianz des Einfachen Krigingprädiktors (siehe Anhang B.5) und zuzüglich eines Strafterms für Unkenntnis des (konstanten) Erwartungswertes.

– Ist der Erwartungswert bekannt, gibt es keine Strafe,  $\hat{Z}_{sk}(\mathbf{s}_0)$  ist der optimale Prädiktor und sein bedingter mittlerer quadratischer Prognosefehler ist die

Einfache Krigingvarianz – in die das Gewicht des Erwartungswertes beim Einfaches Kriging nicht eingeht.

– Ist der Erwartungswert unbekannt, wird der lineare Prädiktor 3.1.4 sub-optimal und zu dem des Gewöhnlichen Krigings. Sein (bedingter) mittlerer quadratischer Prognosefehler steigt um den quadratischen Fehler der Erwartungswertschätzung  $((\hat{\mu} - \mu)^2)$  multipliziert mit dem quadratischen Gewicht des Erwartungswertes beim Kriging: Je stärker der nun geschätzte Erwartungswert in die Prädiktion eingeht, desto gravierender ist es, wenn wir uns verschätzen.

– Ist aber, wie in der Praxis üblich, mit dem Erwartungswert zugleich auch der quadratische Fehler seiner Schätzung unbekannt, ersetzen wir letzteren durch seine eigene Erwartung  $E[(\hat{\mu} - \mu)^2] = (\mathbf{1}^T \mathbf{\Sigma}^{-1} \mathbf{1})^{-1}$  (siehe 2.1.16) und der mittlere quadratische Prognosefehler von  $\hat{Z}_{ok}(\mathbf{s}_0)$  ist unbedingt.

– Erst bei  $(\hat{\mu} - \mu)^2 > 2 \cdot E[(\hat{\mu} - \mu)^2]$  ist die Differenz  $\sigma_{ok|Z=Z}^2(\mathbf{s}_0) - \sigma_{ok}^2(\mathbf{s}_0)$  (siehe 3.1.2) im Betrag größer als der Strafterm für den unbekanntes Erwartungswert in 3.1.5. Der Fehler der Näherung des bedingten mittleren quadratischen Prognosefehler des Gewöhnlichen Krigings durch die Krigingvarianz des Gewöhnlichen Krigings ist also nur in einem kleinen Teil der zu erwartenden Fälle größer als der in der Praxis selten zu vermeidende Fehler, der entsteht, wenn wir die Erwartungswertfunktion nicht kennen und Einfaches Kriging durch Gewöhnliches Kriging substituieren.

Aus der Betrachtung von 3.1.4 und 3.1.5 läßt sich darüberhinaus ersehen, daß wenn der Kovarianzvektor  $\mathbf{c}$  ein Nullvektor ist (was bei räumlicher Unabhängigkeit der Fall ist, aber auch eintreten kann, wenn die Meßstellen zu weit von  $\mathbf{s}_0$  entfernt sind) und damit das Gewicht des Erwartungswertes beim Kriging gleich Eins ist, Gewöhnliches Kriging der Erwartungswertschätzung mit dem mittleren quadratischen Prognosefehler

$$\sigma_{ok}^2(\mathbf{s}_0) = C(0) - 0 + 1 \cdot E[(\hat{\mu} - \mu)^2] = var[Z(\mathbf{s}_0)] + var[\hat{\mu}]$$

(vergleiche 1.0.3) entspricht. Im Falle vorhandener räumlicher Abhängigkeit ist die Differenz

$$E[\{\hat{\mu} - Z(\mathbf{s}_0)\}^2] - \sigma_{ok}^2(\mathbf{s}_0) = \mathbf{c}^T \mathbf{\Sigma}^{-1} \mathbf{c} + \left(1 - (1 - \mathbf{c}^T \mathbf{\Sigma}^{-1} \mathbf{1})^2\right) \cdot E[(\hat{\mu} - \mu)^2]$$

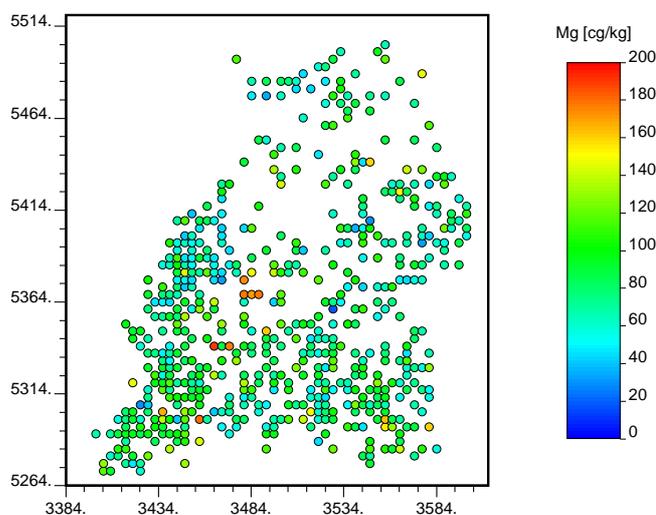
fast<sup>1</sup> immer positiv und somit  $\hat{Z}_{ok}(\mathbf{s}_0)$  ein genauere Prädiktor als  $\hat{\mu}$ .

### 3.2. Fallstudie: Prognoseintervalle

Die verwendeten Daten entstammen der Immissionsökologischen Waldzustandserfassung in Baden-Württemberg aus dem Jahr 1994. Aus den Magnesiumgehalten in Fichtennadeln der Nadeljahrgänge 1994, 1992 und 1990 wurden

---

<sup>1</sup>Die Differenz könnte negativ werden, wenn das Gewicht des Erwartungswertes beim Kriging größer als Eins würde.

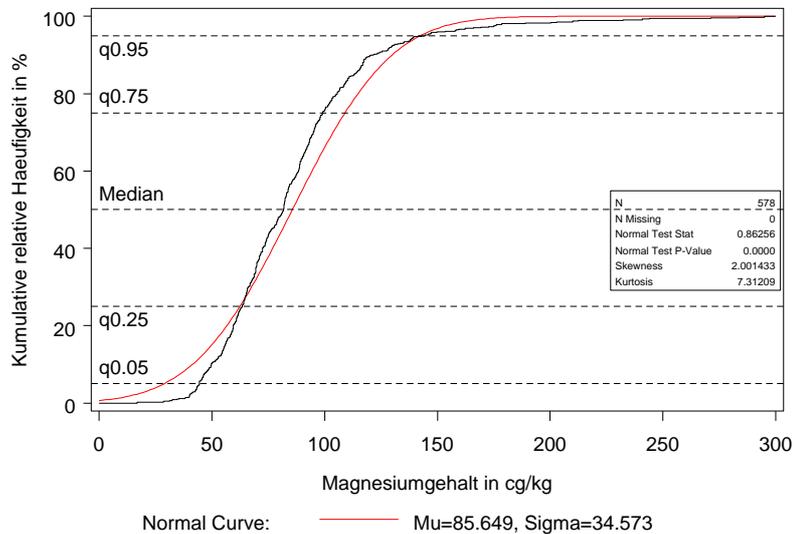


**Abbildung 3.2.:** Magnesiumgehalte in Fichtennadeln, Baden-Württemberg 1994, Gauß-Krüger-Koordinaten in Kilometern.

mittlere Magnesiumgehalte – im folgenden als Magnesiumgehalte bezeichnet – je Stichprobenpunkt bestimmt. Niedrige Magnesiumgehalte führen zu schlechten Stickstoff-Magnesium-Verhältnissen und Nadelvergilbung bei Fichten. Sie können nach Evers u. Hüttl (1992) zu Zuwachsverlusten bei Nadelbäumen führen, gelten als eine der Ursachen der neuartigen Waldschäden (vergleiche beispielsweise Liu u. Huettl (1991)) und sind häufig Anlaß für Düngungen mit Magnesiakalken; Danneberg (2001) gibt die Magnesiummangelgrenze für Fichte mit 70 bis 80  $cg/kg$  an. Der Schlußfolgerung des Waldzustandsberichtes 2006 der Forstlichen Versuchs- und Forschungsanstalt Baden-Württemberg (Meining u. von Wilpert (2006), S. 52) zufolge tritt Magnesiummangel zwar in den letzten Jahren als Schadursache in den Hintergrund, dennoch wird auch hier weiterhin die Düngung mit magnesiumhaltigen Kalken empfohlen.

Die Magnesiumgehalte sind in Abbildung 3.2 in einer Aufsicht dargestellt: Die Immissionsökologische Waldzustandserfassung in Baden-Württemberg folgt einer Rasterstichprobe auf dem gemeinsamen Stichprobennetz der Immissionsökologischen Waldzustandserfassung, der terrestrischen Waldschadensinventur und der Bodenzustandserhebung. Der minimale Abstand zweier Punkte dieses Stichprobennetzes beträgt 4 Kilometer. Punkte, die einen grünblauen bis blauen Ton haben, kennzeichnen Magnesiummangel; auffällig sind ein Magnesiummangelgebiet im Mittleren Schwarzwald südöstlich von Offenburg sowie einzelne sehr hohe Werte am oberen Neckar. Ein Trend, also eine über dem Untersuchungsgebiet nicht konstante Erwartungswertfunktion, ist nicht zu erkennen. Die Annahme  $E[Z(s_0)] = \mu$  und die mit ihr einhergehende Substitution des Universellen Krigings durch Gewöhnliches Kriging erscheint daher gerechtfertigt.

Abbildung 3.3 zeigt die empirische Verteilung der Meßwerte, einige Kennzahlen dieser Verteilung finden sich in Tabelle 3.1. Die empirische Verteilung ist relativ schief: die lognormale Verteilung, die Stichprobenmittelwert und Stich-



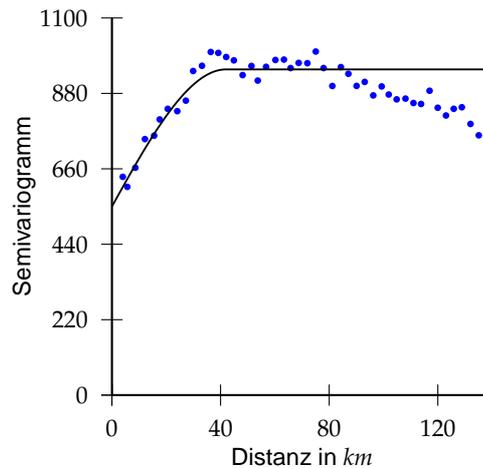
**Abbildung 3.3.:** Empirische kumulative Verteilungsfunktion der Magnesiumgehalte.

**Tabelle 3.1.:** Deskriptive Statistiken; N bezeichnet die Anzahl der Meßwerte, IQA den Interquartilsabstand.

Statistik	Wert	Statistik	Wert
N	578	Kurtosis	7.31
Median	81.67	Minimum	16.83
Varianz	1195.28	Mittelwert	85.65
Schiefe	2.00	Maximum	295.17
	Interquartilsabstand		35.50

probenvarianz aus Tabelle 3.1 reproduziert, besitzt eine Schiefe von etwa 1.276. Die empirische Verteilung der Meßwerte ist keine eindimensionale Randverteilung des Prozesses, von dem wir im Rahmen der folgenden geostatistischen Auswertung annehmen werden, er liege den Daten zugrunde.

Die Annahme, daß der empirischen Verteilung der Meßwerte eine Normalverteilung zugrundeliegt, ist für einen schwach stationär Gauß'schen Prozeß notwendige, aber nicht hinreichende Bedingung. Selbst die Annahme, nicht der gesamte Prozeß sondern nur die gemeinsame Verteilung der vorherzusagenden Zufallsvariablen  $Z(s_0)$  und des Vektors  $\mathbf{Z}$  sei multivariat normal, ist „praktisch nicht nachweisbar“ (Chilès u. Delfiner (1999), S. 382). Es mag Datensätze geben, bei denen die Annahme, sie stammten aus einem Gauß'schen Prozeß, – aufgrund ihrer empirischen Verteilung unter Berücksichtigung der angenommenen Erwartungswertfunktion – naheliegender erscheint. Aber „da Daten niemals gauß'sch sind“ (Cressie (1991), S. 134), treffe ich die Normalvertei-



**Abbildung 3.4.:** Empirisches Semivariogramm der Meßwerte (Punkte) und angepaßte sphärische Semivariogrammfunktion mit  $c_0 = 550$ ,  $c_p = 400$  und  $r_\epsilon = 42$ .

lungsannahme dennoch für die vorliegenden Daten, um beispielhaft an realen Meßwerten den bedingten mit dem unbedingten mittleren quadratischen Prognosefehler des Gewöhnlichen Krigings vergleichen zu können.

An das empirische Semivariogramm der Daten in Abbildung 3.4 ein isotropes sphärisches Semivariogramm mit Nugget-Effekt

$$\gamma(|\mathbf{h}|) = \begin{cases} c_0 + c_p \cdot \left[ \frac{3}{2} \cdot \frac{|\mathbf{h}|}{r_\epsilon} - \frac{1}{2} \cdot \frac{|\mathbf{h}|^3}{r_\epsilon^3} \right] & \text{wenn } 0 < |\mathbf{h}| \leq r_\epsilon, \\ c_0 + c_p & \text{wenn } |\mathbf{h}| \geq r_\epsilon, \\ 0 & \text{wenn } |\mathbf{h}| = 0 \end{cases}$$

angepaßt, das als Semivariogramm des Prozesses angenommen wird. Die Wahl der Parameter fällt dabei auf  $c_0 = 550$  für den Nugget-Effekt,  $c_p = 400$  für den Schwellenwert und  $r_\epsilon = 42$  für die effektive Reichweite. Die effektive Reichweite ist damit relativ zu den Abständen zwischen den Meßstellen mit einem Faktor von maximal  $\frac{42\text{km}}{4\text{km}} = 10.5$  sehr groß. Mit diesem Semivariogramm und der Meßstellenverteilung aus Abbildung 3.2 ergibt sich auf zwei Nachkommastellen gerundet der geschätzte konstante Erwartungswert nach 2.1.15 als  $\hat{\mu} = 87.44$  und seine Varianz nach 2.1.16 als  $E[(\hat{\mu} - \mu)^2] = 16.12$ .

Unabhängig von den Meßwerten lassen sich nun für ein gegebenes Prognosegitter die unbedingten mittleren quadratischen Prognosefehler des Einfachen und des Gewöhnlichen Krigings sowie das Gewicht des Erwartungswertes beim Kriging berechnen. Das verwendete Prognosegitter deckt nicht nur die Fichtenbestände Baden-Württembergs sondern das gesamte Bundesland ab. Wie sinnvoll die Prognose eines mittleren Magnesiumsgehaltes in Fichtennadeln für einen Weinberg, einen Kartoffelacker oder ein Industriegebiet ist, kann bezweifelt werden. Aber für die Prognose nur der Fichtenbestände ist die Datenlage „zu gut“: aufgrund der relativ zur Reichweite des Semivariogrammes geringen Entfernung der nächstgelegenen Datenpunkte wird das Gewicht des

### 3. Bedingte Vorhersagefehler in gauß'schen Prozessen

---

Erwartungswertes beim Kriging in den Fichtenbeständen sehr klein sein. Um die Unterschiede zwischen den nachfolgend definierten Gewöhnlichen und cressieschen Prognoseintervallen auch für größere Werte des Gewichtes des Erwartungswertes beim Kriging zu betrachten, benutze ich deswegen ganz Baden-Württemberg als Prognosegebiet.

Unter der getroffenen Normalverteilungsannahme ist wegen 2.3.3 das wegen 2.3.3 auf Einfachem Kriging basierende exakte Prognoseintervall für  $z(\mathbf{s}_0)$

$$I_{\alpha, Z=z}^{SK}(\mathbf{s}_0) = \left[ \hat{z}_{sk}(\mathbf{s}_0) - \sigma_{sk}(\mathbf{s}_0) \cdot u_{1-\frac{\alpha}{2}}; \hat{z}_{sk}(\mathbf{s}_0) + \sigma_{sk}(\mathbf{s}_0) \cdot u_{1-\frac{\alpha}{2}} \right] : \\ P \left[ \hat{z}_{sk}(\mathbf{s}_0) - \sigma_{sk}(\mathbf{s}_0) \cdot u_{1-\frac{\alpha}{2}} \leq Z(\mathbf{s}_0) \leq \hat{z}_{sk}(\mathbf{s}_0) + \sigma_{sk}(\mathbf{s}_0) \cdot u_{1-\frac{\alpha}{2}} \right] = 1 - \alpha,$$

wobei  $u_{1-\frac{\alpha}{2}}$  das  $1 - \frac{\alpha}{2}$ -Quantil der Standardnormalverteilung bezeichnet. Es wird durch Gewöhnliches Kriging mit

$$I_{\alpha, Z=z}^{OK}(\mathbf{s}_0) = \left[ \hat{z}_{ok}(\mathbf{s}_0) - \sigma_{ok|Z=z}(\mathbf{s}_0) \cdot u_{1-\frac{\alpha}{2}}; \hat{z}_{ok}(\mathbf{s}_0) + \sigma_{ok|Z=z}(\mathbf{s}_0) \cdot u_{1-\frac{\alpha}{2}} \right],$$

dem *Gewöhnlichen Prognoseintervall*, approximiert; der Polfeldt-Ansatz entspricht dann er Ersetzung von  $I_{\alpha, Z=z}^{OK}(\mathbf{s}_0)$  durch das cressiesche Prognoseintervall

$$I_{\alpha}^{Cr}(\mathbf{s}_0) = \left[ \hat{z}_{ok}(\mathbf{s}_0) - \sigma_{ok}(\mathbf{s}_0) \cdot u_{1-\frac{\alpha}{2}}; \hat{z}_{ok}(\mathbf{s}_0) + \sigma_{ok}(\mathbf{s}_0) \cdot u_{1-\frac{\alpha}{2}} \right].$$

Das Verhältnis

$$\frac{B(I_{\alpha}^{Cr}(\mathbf{s}_0)) - B(I_{\alpha, Z=z}^{OK}(\mathbf{s}_0))}{B(I_{\alpha, Z=z}^{OK}(\mathbf{s}_0))}$$

der Breiten  $B(\cdot)$  der Prognoseintervalle läßt sich zu

$$\frac{\sigma_{ok}(\mathbf{s}_0) - \sigma_{ok|Z=z}(\mathbf{s}_0)}{\sigma_{ok|Z=z}(\mathbf{s}_0)} = R_{duc}(\mathbf{s}_0)$$

kürzen, die Breite des Gewöhnliche Prognoseintervalles  $I_{\alpha, Z=z}^{OK}(\mathbf{s}_0)$  ergibt sich als  $B(I_{\alpha, Z=z}^{OK}(\mathbf{s}_0)) = B(I_{\alpha}^{Cr}(\mathbf{s}_0)) \cdot (1 + R_{duc})^{-1}$ . Da die Breiten der Prognoseintervalle und ihre Verhältnisse proportional zu den Quadratwurzeln der mittleren quadratischen Prognosefehler sind, werde ich in diesem Kapitel statt der mittleren quadratischen Prognosefehler ihre Quadratwurzeln betrachten und sie als Fehler bezeichnen.

In Abbildung 3.5 ist kein Unterschied zwischen den von den Meßwerten unabhängigen Fehlern  $\sigma_{ok}(\mathbf{s}_0)$  und  $\sigma_{sk}(\mathbf{s}_0)$  zu erkennen,  $\sigma_{sk}(\mathbf{s}_0)$  ist hier aufgrund der Annahme eines gauß'schen Prozesses der Fehler des optimalen Prädiktors und wegen 2.2.3 gleich der bedingten Standardabweichung der Variablen  $Z(\mathbf{s}) \quad \forall \mathbf{s} \in D$ .

Die relative Differenz

$$\frac{\sigma_{ok}(\mathbf{s}_0) - \sigma_{sk}(\mathbf{s}_0)}{\sigma_{sk}(\mathbf{s}_0)} = R_{duo}(\mathbf{s}_0)$$

in Tabelle 3.2 zeigt, daß es – wenn auch relativ zu  $\sigma_{sk}(\mathbf{s}_0)$  kleine – Unterschiede zwischen  $\sigma_{ok}(\mathbf{s}_0)$  und  $\sigma_{sk}(\mathbf{s}_0)$  gibt: Sehr viele sehr kleine Werte und wenige

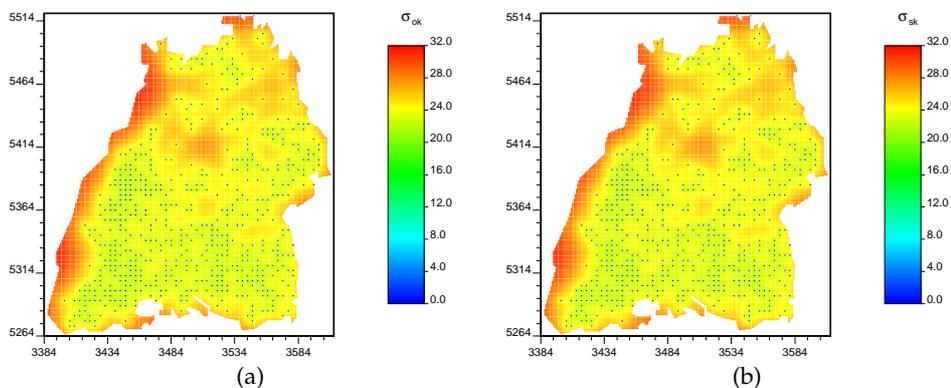
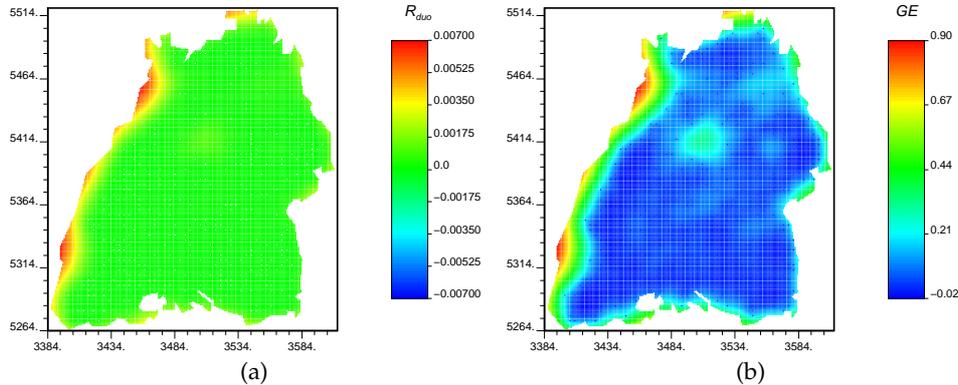


Abbildung 3.5.: Unbedingte Fehler des Gewöhnlichen Krigings (a) und des Einfachen Krigings (b).

Tabelle 3.2.: Extrema, Median und arithmetisches Mittel bei unterschiedlichen relativen Abständen  $\frac{(\hat{\mu}-\mu)^2}{E[(\hat{\mu}-\mu)^2]}$ , bel. steht für konstante Werte bei beliebigen Abständen.

$\frac{(\hat{\mu}-\mu)^2}{E[(\hat{\mu}-\mu)^2]}$	Statistik	Minimum	Maximum	Median	Mittel
bel.	$\sigma_{sk}^2(\mathbf{s}_0)$	0	937.1596	562.4708	582.6520
bel.	$\sigma_{ok}^2(\mathbf{s}_0)$	0	950.0342	562.5342	583.3223
bel.	$\hat{\mu}$	87.4404	87.4404	87.4404	87.4404
bel.	$E[(\hat{\mu}-\mu)^2]$	16.1207	16.1207	16.1207	16.1207
bel.	$(1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1})$	-0.0105	0.8936	0.0638	0.1277
bel.	$R_{duo}$	0	0.0068	0.000059	0.000444
0	$\sigma_{ok Z=z}(\mathbf{s}_0)$	0	30.6130	23.7164	23.8896
0	$R_{duc}$	0	0.0067	0.000059	0.000441
1	$\sigma_{ok Z=z}(\mathbf{s}_0)$	0	30.8226	23.7178	23.9017
1	$R_{duc}$	0	0	0	0
2	$\sigma_{ok Z=z}(\mathbf{s}_0)$	0	31.0307	23.7243	23.9710
2	$R_{duc}$	-0.0068	0	-0.000059	-0.000444



**Abbildung 3.6.:** Relative Differenz  $R_{duo}$  zwischen  $\sigma_{ok}(\mathbf{s}_0)$  und  $\sigma_{sk}(\mathbf{s}_0)$ , identisch zu  $R_{duc}$  für  $\hat{\mu} = \mu$  (a) und Gewicht des Erwartungswertes beim Kriging (b).

große führen zu einem arithmetischen Mittel, das eine Zehnerpotenz über dem Median liegt, das Maximum liegt wiederum eine Zehnerpotenz höher. Die wenigen hohen Werte liegen, wie Abbildung 3.6(a) zeigt, fast ausschließlich in den Randgebieten Baden-Württembergs – weit weg von jeglicher Meßstelle.

Daß die relativen Differenzen  $R_{duo}$  so klein sind, läßt sich durch den Strafterm in 3.1.5 erklären: Er ist für den unbedingten mittleren quadratischen Prognosefehler des Gewöhnlichen Krigings  $(1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1})^2 E[(\hat{\mu} - \mu)^2]$ .

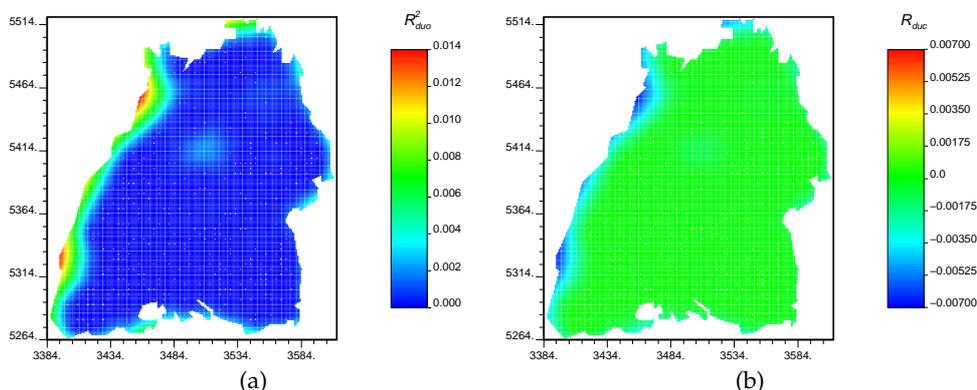
$E[(\hat{\mu} - \mu)^2]$  ist aufgrund des hohen Stichprobenumfangs sehr klein im Vergleich zum mittleren quadratischen Prognosefehler des Einfachen Krigings (siehe Tabelle 3.2). Das Quadrat des in Abbildung 3.6(b) gezeigten Gewichtes des Erwartungswertes beim Kriging weist einen Median von etwa 0.06 auf, die mittleren quadratischen Prognosefehler sind sich daher sehr ähnlich, ihre relative Differenz

$$\frac{\sigma_{ok}^2(\mathbf{s}_0) - \sigma_{sk}^2(\mathbf{s}_0)}{\sigma_{sk}^2(\mathbf{s}_0)} = R_{duo}^2$$

meist nahe Null (Abbildung 3.7(a)). Wenn sich  $\sigma_{ok}^2(\mathbf{s}_0)$  und  $\sigma_{sk}^2(\mathbf{s}_0)$  ähnlich sind, gilt dies auch für ihre Wurzeln, deren relative Differenz  $R_{duo}$  dann sehr klein sein muß.

Interessanter als die Größenordnung der unbedingten Fehler ist ihre Relation zu den bedingten Fehlern  $R_{duc}$  und damit die relative Breite der cressieschen Prognoseintervalle. Um diese angeben zu können, muß  $(\hat{\mu} - \mu)^2$  bekannt sein, ich werde daher im folgenden  $(\hat{\mu} - \mu)^2$  in Anlehnung an die Darstellung des linearen Krigings in schwach stationären gauß'schen Prozessen in Kapitel 3.1 unterschiedliche Werte annehmen lassen:

– Im Extremfall einer exakten Mittelwertschätzung  $\hat{\mu} = \mu$  ist  $(\hat{\mu} - \mu)^2 = 0$  und aus 3.1.1 ergibt sich  $\sigma_{ok|Z=Z}^2(\mathbf{s}_0) = \sigma_{sk}^2(\mathbf{s}_0)$  und daher gilt  $R_{duc} = R_{duo}$ . Da  $R_{duc}$  überall nach Tabelle 3.2 nicht-negativ ist, sind die Gewöhnlichen Prognoseintervalle also über dem gesamten Untersuchungsgebiet höchstens so breit



**Abbildung 3.7.:** Relative Differenz  $R_{duo}^2$  zwischen unbedingten quadratischen Fehlern (a) und relative Differenz  $R_{duc}$  zwischen bedingten Fehlern bei  $(\hat{\mu} - \mu)^2 = 2 \cdot E[(\hat{\mu} - \mu)^2]$  (b).

wie die cressieschen, in den Gebieten nahe den Meßstellen und damit in den Fichtenbeständen allerdings auch nicht nennenswert schmaler. Mit zunehmender Differenz zwischen dem Erwartungswert und seinem Schätzwert bleiben  $\sigma_{sk}^2(\mathbf{s}_0)$  und  $\sigma_{ok}^2(\mathbf{s}_0)$  konstant, während der bedingte mittlere quadratische Prognosefehler des Gewöhnlichen Krigings steigt,

– bis er bei  $(\hat{\mu} - \mu)^2 = E[(\hat{\mu} - \mu)^2]$  gleich dem unbedingten mittleren quadratischen Prognosefehler  $\sigma_{ok}^2(\mathbf{s}_0)$  ist und  $R_{duc}$  über dem gesamten Untersuchungsgebiet Null ist (Tabelle 3.2), die cressieschen und die Gewöhnlichen Prognoseintervalle sind identisch.

– Wird die Differenz  $(\hat{\mu} - \mu)^2$  größer als ihre Erwartung  $E[(\hat{\mu} - \mu)^2]$ , wird  $R_{duc}$  negativ und die Gewöhnlichen Prognoseintervalle werden breiter als die cressieschen.

– Im nur mit Wahrscheinlichkeit  $P[(\hat{\mu} - \mu)^2 \geq 2E[(\hat{\mu} - \mu)^2]] = 1 - P[\chi_1^2 < 2] \approx 0.157$  auftretenden Fall, daß die quadratische Abweichung der Erwartungswertschätzung das Doppelte ihrer Erwartung ist, bewegt sich  $R_{duc}$  (mit umgekehrtem Vorzeichen) in der Größenordnung von  $R_{duo}$  (siehe Abbildung 3.7(b)). Die Gewöhnlichen Prognoseintervalle sind nun in großen Entfernungen von den Meßstellen breiter als die cressieschen – um etwa den gleichen Faktor, mit dem die cressieschen Prognoseintervalle kleiner sind als die exakten Prognoseintervalle  $I_{\alpha, Z=z}^{SK}(\mathbf{s}_0)$ .

Dieses Verhalten der Breite der Gewöhnlichen Prognoseintervalle ist dabei unabhängig von der (im Fall der betrachteten Daten sehr geringen) Varianz  $E[(\hat{\mu} - \mu)^2]$ , es hängt ab vom – in der Praxis unbekanntem – Verhältnis  $(\hat{\mu} - \mu)^2 \cdot (E[(\hat{\mu} - \mu)^2])^{-1}$ . Selbst bei gründlich mißglückter Erwartungswertschätzung, wenn also der aus den Daten geschätzte Erwartungswert weit stärker als zu erwarten vom Erwartungswert des angenommenen Zufallspro-

### 3. Bedingte Vorhersagefehler in gauß'schen Prozessen

---

zesses abweicht (die Wahrscheinlichkeit eines solchen Falles läßt sich durch einen vergrößerten Stichprobenumfang nicht verringern), unterscheiden sich die Breiten der cressieschen Prognoseintervalle von denen der Gewöhnlichen in der gleichen Größenordnung wie von den wahren exakten Prognoseintervallen.

## 4. Bedingte Verteilungen in der Anwendung

In diesem Kapitel werden bedingte lokale Verteilungsfunktionen, also die bedingten eindimensionalen Randverteilungen

$$F_{Z(\mathbf{s}_0)|Z=\mathbf{z}}(z(\mathbf{s}_0)) = P[Z(\mathbf{s}_0) \leq z(\mathbf{s}_0) | \mathbf{Z} = \mathbf{z}], \quad \forall z(\mathbf{s}_0) \in S_{(0)}$$

für eine Menge  $S_{(0)}$  von Prognosepunkten approximiert. Ziel ist es, zu prüfen, wie gut sich verschiedene Methoden für nicht gauß'sche Daten eignen. Eingesetzt werden dabei der Polfeldt-Ansatz als lineares Verfahren sowie die nicht-linearen Verfahren Gauß'sches Disjunktives Kriging und Median-Indikator-Kriging. Ich wähle die beiden nicht-linearen Verfahren dabei trotz ihrer strengen Anforderungen – die Annahme der gauß'schen bivariaten Verteilungen der gauß'schen Anamorphosen ist nicht überprüfbar; die Voraussetzung der für Median-Indikator-Kriging notwendigen proportionalen räumlichen Abhängigkeit ist wahrscheinlich nicht erfüllt. Es bleiben daher drei Möglichkeiten: Das Verfahren aufgrund der nicht beizubehaltenden Annahme nicht anwenden, ein anderes Verfahren anwenden – beispielsweise Indikator-Cokriging, das so schwierig umzusetzen ist, daß es außer bei Methodenstudien so gut wie nicht zur Anwendung kommt (vergleiche Chilès u. Delfiner (1999), S. 387) –, oder das Verfahren trotz der nicht beizubehaltenden Annahme anwenden und Fehler in Kauf nehmen. Man mag einwenden, dies sei eine Methodenstudie, ich werde mich aber auf Methoden konzentrieren, die wenigstens die Aussicht auf eine zukünftige Bedeutung in der Praxis besitzen, Indikator-Cokriging gehört meines Erachtens nicht dazu. Ich entscheide mich daher für die dritte Möglichkeit und wende Median-Indikator-Kriging an. Der Polfeldt-Ansatz unterstellt ursprünglich einen gauß'schen Zufallsprozeß und erbringt dann, wie wir im vorangegangenen Kapitel gesehen haben, sehr gute Ergebnisse, da sich die Verteilungen  $N(\hat{Z}_{sk}(\mathbf{s}_0), \sigma_{sk}^2(\mathbf{s}_0))$  und  $N(\hat{Z}_{ok}(\mathbf{s}_0), \sigma_{ok}^2(\mathbf{s}_0))$  zumeist kaum unterscheiden. Es ist aber gängige Praxis (vergleiche Cressie (1991), S. 122, Zimmerman u. Cressie (1992), S. 41), 95%-Prognoseintervalle durch  $[\hat{z}_k(\mathbf{s}_0) - 1.96 \cdot \sigma_k(\mathbf{s}_0); \hat{z}_k(\mathbf{s}_0) + 1.96 \cdot \sigma_k(\mathbf{s}_0)]$  zu approximieren. Dies entspricht der Anwendung des Polfeldt-Ansatz für Prognoseintervalle (siehe Kapitel 3.2, Seite 42) in nicht gauß'schen Prozessen; ich werde daher den Polfeldt-Ansatz auch bei nicht gauß'schen Prozessen verwenden, um die Einschränkung seiner Leistungsfähigkeit bei Verletzung der Normalverteilungsannahme beurteilen zu können.

In Kapitel 4.1 werden die drei Verfahren hinsichtlich der Güte ihrer Prognosen sowie der Güte ihrer Approximationen der bedingten lokalen Verteilungsfunktionen miteinander verglichen.

In Kapitel Kapitel 4.2 werden die gleichen Verfahren auf simulierte Daten angewandt, um die Genauigkeit prognostizierter Wahrscheinlichkeiten der drei Verfahren vergleichen zu können. Der zur Simulation benutzte Prozess

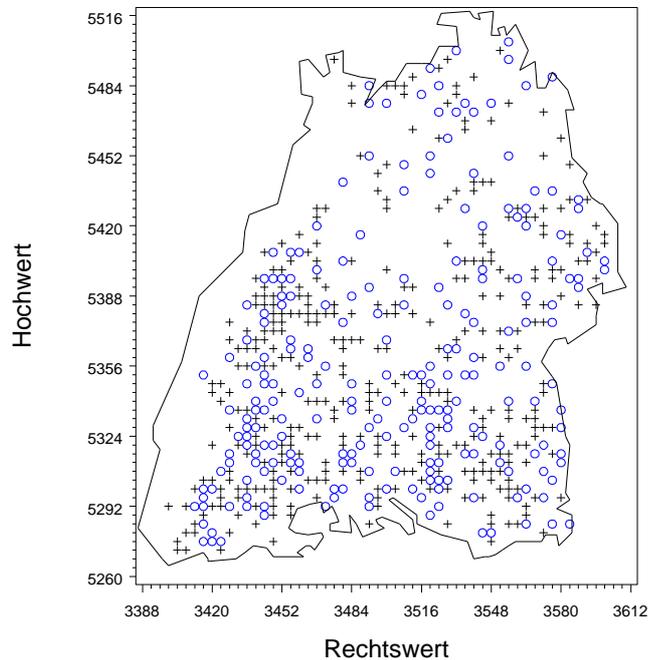


Abbildung 4.1.: Aufteilung der Meßstellen in  $S_{(1)}$  (Kreuze) und  $S_{(2)}$  (Kreise).

orientiert sich dabei an den in Kapitel 3.2 beschriebenen Daten; die Punkte, an denen die Stichproben aus Realisationen des Prozess gezogen werden, stimmen mit den Stichprobenpunkten der Daten aus Kapitel 3.2 überein.

#### 4.1. Fallstudie: Magnesium in Baden-Württemberg

Um die Prognosen der drei betrachteten Verfahren mit den in Kapitel 2.2.5 vorgestellten Maßen der Prognosegüte kreuzvalidieren und die Überdeckungswahrscheinlichkeiten aus Kapitel 2.3.2 berechnen zu können, werden die Meßstellen der Daten aus Kapitel 3.2 nach der in Abbildung 4.1 abgebildeten zufälligen Auswahl in zwei Mengen  $S_{(1)}$  vom Umfang 351 und  $S_{(2)}$  vom Umfang 227 aufgeteilt. Sie definieren die Stichproben  $z^1 = \{z(\mathbf{s}_n) : \mathbf{s}_n \in S_{(1)}\}$  und  $z^2 = \{z(\mathbf{s}_n) : \mathbf{s}_n \in S_{(2)}\}$ . Für jedes der betrachteten Verfahren – Polfeldt-Ansatz, Indikator-Kriging und Disjunktives Kriging – wird einmal  $z^1$  als Modellierungs- und  $z^2$  als Validierungsstichprobe genutzt und einmal  $z^1$  als Validierungs- und  $z^2$  als Modellierungsstichprobe. Die Menge  $S_{(0)}$  der Prognosepunkte ist dann jeweils die Menge der Meßstellen der Validierungsstichprobe. Die Bezeichnung der Prognosegütemaße und der Überdeckungswahrscheinlichkeiten richtet sich dabei immer nach der Modellierungsstichprobe, der  $MEP_v$  für  $z^2$  ist also der mittlere absolute Prognosefehler des  $v$ -ten Verfahrens mit  $z^1$  als Validierungs- und  $z^2$  als Modellierungsstichprobe.

**Tabelle 4.1.:** Deskriptive Statistiken für alle Magnesium-Messungen (erste Zeile), der Stichprobe  $z^1$  (zweite Zeile) und der Stichprobe  $z^2$ ; N bezeichnet die Anzahl der Meßwerte, Mittel das arithmetische Mittel der Stichprobe, Min. das Minimum, Max. das Maximum und IQA den Interquartilsabstand.

N	Mittel	Varianz	Schiefe	Min.	Median	Max.	IQA
578	85.6488	1195.28	2.00143	7.31209	81.6667	295.167	35.5000
351	85.9644	1335.54	2.01207	7.43734	81.1667	295.167	38.1667
227	85.1608	982.95	1.92139	6.30327	81.8333	244.167	31.6667

Beider Stichproben Lagemaße unterscheiden sich nur geringfügig von denen des gesamten Datensatzes, der in Tabelle 4.1 mit  $z^0$  bezeichnet ist. Varianz und Interquartilsabstand der Stichprobenwerte sind für  $z^1$  größer als für  $z^0$ , bei  $z^2$  hingegen kleiner als die des gesamten Datensatzes.

Die nicht-linearen Verfahren benötigen, wie in Kapitel 2.3.2 erläutert, zur Approximation bedingter lokaler Verteilungsfunktionen eine endliche Anzahl von Schwellenwerten. Für diese werden Unterschreitungswahrscheinlichkeiten prognostiziert und der Verlauf der Verteilungsfunktion zwischen den Prognosen durch lineare Interpolation approximiert. Ich wähle als Schwellenwerte die neun Dezile der empirischen Verteilung der Meßwerte der jeweiligen Modellierungsstichprobe, da die geeigneteren Dezile der jeweiligen bedingten lokalen Verteilungsfunktion von  $Z(s_0)$  unbekannt sind und, wären sie bekannt, zu einer eigenen Parametrisierung der nicht-linearen Verfahren für jede Prognosestelle  $s_0$  führten. Die Approximationen des Polfeldt-Ansatz können hingegen kontinuierlich für beliebige Wahrscheinlichkeiten bestimmt werden.

Die Gauß'schen Anamorphosen des Gauß'schen Disjunktiven Krigings nutzen die hermiteschen Polynome  $H_0$  bis  $H_{19}$ . Nach Abbildung 4.2 gelingen die Quantiltransformationen der beiden Stichproben etwa gleich gut. Tabelle 4.2 zeigt die dabei verwendeten Koeffizienten  $f_1$  bis  $f_{19}$ : Da die Koeffizienten ab  $k = 7$  sehr klein sind, reichten wohl auch die Polynome  $H_0$  bis  $H_7$  aus. Da aber die Rechenzeit des Algorithmuses des Gauß'schen Disjunktiven Krigings inklusive der Quantiltransformation auf derzeit verfügbaren Rechenanlagen für diese beiden Anwendungen jeweils unter zehn Minuten lag, sehe ich keinen Anlaß,  $k$  besonders niedrig zu halten.

Aus den Meßwerten der Stichproben und ihren Transformationen mit den jeweiligen gauß'schen Anamorphosen beziehungsweise zum Median-Indikator ergeben sich die in Abbildung 4.3 gezeigten empirischen isotropen Semivariogramme. Der großen Stichprobenumfänge wegen sind die Schätzungen der jeweiligen mittleren Semivariogrammwerte  $\hat{\gamma}_l(|\mathbf{h}|)$  ziemlich stabil: Das Minimum der Anzahl  $|N(|\mathbf{h}|)|$  der in eine Abstandsvektorklasse fallenden Punktpaare beträgt 161 bei Stichprobe  $z^1$  und 55 bei Stichprobe  $z^2$ , für die meisten Abstandsvektorklassen liegen aber mehr als eintausend Punktpaare vor. Wie aufgrund der Stichprobenvarianzen zu erwarten war, weist das empirische Semivariogramm der Magnesiumwerte von  $z^1$  einen deutlich höheren Schwellen-



**Tabelle 4.2.:** Koeffizienten  $f_k$  der hermiteschen Polynome  $H_k$  bei der Gauß'schen Anamorphose von  $\psi_{(578)}^1$  und  $\psi_{(578)}^2$ .

	k	0	1	2	3	4
$z^1$	0+	85.964388	-25.168818	6.385013	-2.111460	0.339671
	5+	0.391146	-0.441027	0.137675	0.120995	-0.111726
	10+	0.004838	0.037803	-0.023854	-0.003591	0.015084
	15+	-0.004569	-0.006082	0.003813	0.001458	-0.001779
$z^2$	0+	85.160793	-21.607654	5.306800	-1.746345	0.211964
	5+	0.530437	-0.456063	0.033434	0.196130	-0.090736
	10+	-0.050460	0.048611	0.001924	-0.016428	0.007149
	15+	0.002710	-0.005472	0.001151	0.002725	-0.001448

**Tabelle 4.3.:** Semivariogramme:  $r_\epsilon$  ist die effektive Reichweite,  $c_p$  der partielle Schwellenwert und  $c_0$  der Nugget-Effekt (siehe Kapitel 2.1.1); MG bezeichnet die untransformierte, MI die zum Median-Indikator transformierte und GA die mit der gauß'schen Anamorphosen transformierte Variable.

Stichprobe	Variable	Type	$r_\epsilon$	$c_p$	$c_0$
$z^1$	MG	Sphärisch	42.0	600.0	1000.0
$z^2$	MG	Exponentiell	20.0	720.0	280.0
$z^1$	GA	Sphärisch	40.0	1.0	0.0
$z^2$	GA	Sphärisch	25.0	1.0	0.0
$z^1$	MI	Sphärisch	42.0	0.1100	0.1500
$z^2$	MI	Exponentiell	20.0	0.1000	0.1500

wert auf als das der zweiten Stichprobe. Die Reichweite des Semivariogrammes ist für  $z^1$  ebenfalls größer, die räumliche Abhängigkeit also ausgeprägter, eine dreidimensionale Darstellung von  $z^1$  wirkte also glatter, nicht so „zappelig“ wie eine Darstellung von  $z^2$ . Der Schwellenwert des Semivariogrammes von  $z^1$  weicht deutlich von der Stichprobenvarianz ab, obwohl beide als Schätzungen der Prozeßvarianz  $C(0)$  interpretiert werden können. Da die Stichprobenvarianz die räumliche Abhängigkeit außer acht läßt, ist der Unterschied nicht überraschend – die Tatsache, daß die räumliche Abhängigkeit bei  $z^2$  nicht so stark ausgeprägt ist, erklärt die stärkere Übereinstimmung der Stichprobenvarianz von  $z^2$  mit dem Schwellenwert des Semivariogrammes. Diese Struktur ändert sich durch die beiden Transformationen nicht, auffällig ist die extreme Glättung der empirischen Semivariogramme durch die gauß'sche Anamorphose. Form und Parameter der angepassten Modelle finden sich in Tabelle 4.3.

### Prognosegüte

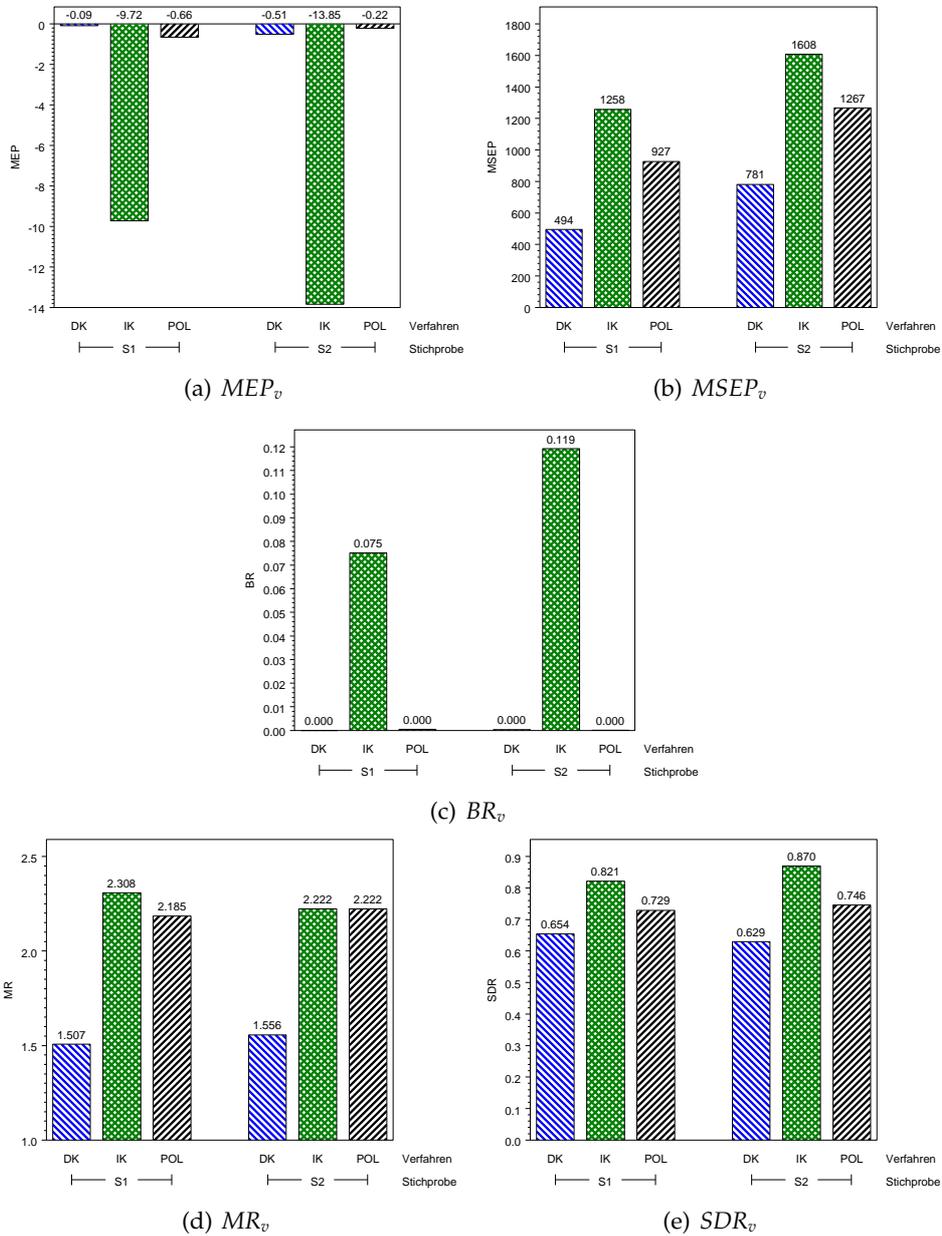
Um die Prognosegüte eines Verfahrens zu beurteilen, betrachten wir die in Kapitel 2.2.5 vorgestellten Maße. Dabei werden die Prognosen der drei Verfahren anhand der Messungen in der jeweiligen Validierungsstichprobe bewertet. Die Prädiktoren sind 2.2.10 für den Polfeldt-Ansatz und 2.2.19 für Gauß'sches Disjunktives Kriging. Indikator-Kriging wird für gewöhnlich nicht zur Prognose des Wertes  $z(\mathbf{s}_0)$  genutzt; da wir dennoch eine solche Prognose zum Vergleich der Prognosegüten der Verfahren betrachten wollen, benutze ich

$$\hat{Z}_{(IK)}(\mathbf{s}_0) = \sum_{i=1}^9 q_{0,i,N} \cdot \left( P[Z(\mathbf{s}_0) \leq q_{0,i,N}] - I(i > 1) \cdot \left( P[Z(\mathbf{s}_0) \leq q_{0,(i-1),N}] \right) \right),$$

also die Summe der gewichteten Dezile der jeweiligen Modellierungsstichprobe, als Prädiktor des Median-Indikator-Krigings. Die Gewichte entsprechen den Funktionswerten der durch Indikator-Kriging approximierten bedingten lokalen Dichtefunktion an den Stellen der Quantile der Modellierungsstichprobe. Dieser Prädiktor ist also vom Typ  $\hat{x} = \sum_i x_i \cdot P[X = x_i]$  und daher eine Approximation eines Erwartungswertes  $EX = \int_{-\infty}^{\infty} x \cdot f(x) dx$ . Er ist damit – wie auch die übrigen Krigingprädiktoren – ein kleinste-Quadrate-Prädiktor und somit beispielsweise dem Median der durch Median-Indikator-Kriging approximierten bedingten lokalen Verteilungsfunktion,  $\hat{z}_{q_{0.5}}(\mathbf{s}_0)$ , vorzuziehen.

Die Maße der Prognosegüte zeigt Abbildung 4.4. Median-Indikator-Kriging liegt in der Bewertung durch den mittleren absoluten Vorhersagefehler (siehe Abbildung 4.4(a)) und in der Folge auch in der Bewertung durch das Verhältnis aus quadriertem mittleren absoluten Vorhersagefehler und mittlerem quadratischem Vorhersagefehler,  $BR_v$ , weit hinter den beiden anderen Verfahren. Der mittlere quadratische Vorhersagefehler des Median-Indikator-Kriging ist ebenfalls deutlich größer als die des Polfeldt-Ansatzes und des Gauß'schen Disjunktiven Krigings. Die Betrachtung des mittleren Ranges,  $BR_v$ , zeigt, daß das Indikator-Kriging nicht deutlich schlechter als der Polfeldt-Ansatz bewertet wird. Die relativ bessere Bewertung des Median-Indikator-Krigings durch  $BR_v$  (verglichen mit jener durch  $MSEP_v$ ) legt nahe, daß das Indikator-Kriging einige sehr schlechte Prognosen ergibt, es aber auch häufig genauere Prognosen ergibt als der Polfeldt-Ansatz. Daß die Verfahren bei der Prognose der einzelnen Validierungsdaten häufig die Reihenfolge in der Genauigkeit wechseln, zeigt auch Abbildung 4.4(e): die Standardabweichungen der Ränge der absoluten Vorhersagefehler sind relativ hoch – eine diskrete uniforme Verteilung (vergleiche Mood u. a. (1974), S. 86) mit drei möglichen Ausprägungen besitzt eine Standardabweichung von etwa 0.816. Die drei Verfahren wechseln also in etwa so häufig den Rang des absoluten Vorhersagefehler wie eine gleichverteilte Variable. In einer Gesamtbewertung über alle Gütemaße hinweg erscheint mir Gauß'sches Disjunktives Kriging das bei der Prognose beste Verfahren zu sein.

#### 4.1. Fallstudie: Magnesium in Baden-Württemberg



**Abbildung 4.4.:** Maße der Prognosegüte für die Stichproben  $z^1$  und  $z^2$  (hier mit S1 und S2 bezeichnet); DK steht für Gauß'sches Disjunktives Kriging, IK für Median-Indikator-Kriging und POL für den Polfeldt-Ansatz.

### Überdeckungswahrscheinlichkeiten

Die Güte der Approximationen der bedingten lokalen Verteilungsfunktionen wird mit dem Konzept der aus Kapitel 2.3.2 bekannten Überdeckungswahrscheinlichkeiten beurteilt. In Abbildung 4.1 finden sich Diagramme der Überdeckungswahrscheinlichkeiten über den erwarteten Wahrscheinlichkeiten für alle drei Approximationsverfahren.

In der linken Spalte sind die Streudiagramme, deren Überdeckungswahrscheinlichkeiten mit  $z^1$  als Modellierungs- und  $z^2$  als Validierungsstichprobe ermittelt wurden, dargestellt. Beim Polfeldt-Ansatz sind hier die Überdeckungswahrscheinlichkeiten für kleine erwartete Wahrscheinlichkeiten zu groß und für große erwartete Wahrscheinlichkeiten zu klein. Es sind also für niedrige Werte  $p$  mehr als  $N \cdot p$  approximierte Quantile  $\hat{q}_v(\mathbf{s}_k; p)$  grösser als der jeweilige Meßwert  $z(\mathbf{s}_k)$ . Dieses Ergebnis ähnelt der bekannten Tendenz des gewöhnlichen Krigings, aufgrund der bedingten Verzerrtheit des Prädiktors  $p_{ok}(\mathbf{s}_0, \mathbf{Z})$  (siehe 2.2.12), große Werte zu unter- und kleine Werte zu überschätzen (vergleiche Olea (1999), S. 141). Für das Gauß'sches Disjunktives Kriging gilt dies für alle erwarteten Werte  $p$  und je größer  $p$  wird, desto stärker weicht die Überdeckungswahrscheinlichkeit von  $p$  ab, allerdings bleibt die maximale Differenz  $f(p) - p$  unter der des Polfeldt-Ansatzes. Den besten Eindruck erweckt Indikator-Kriging, dessen Überdeckungswahrscheinlichkeiten anscheinend zufällig und gering um die Winkelhalbierende streuen.

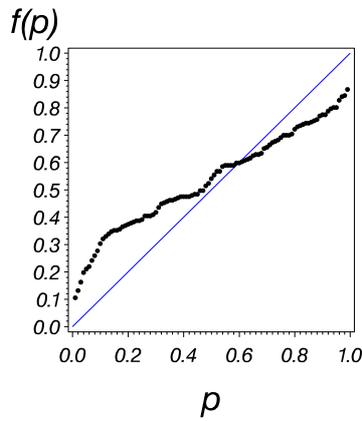
Die Tendenz der Beurteilung der drei Approximationsverfahren bleibt für die umgekehrte Konstellation mit  $z^2$  als Modellierungs- und  $z^1$  als Validierungsdatensatz (in Abbildung 4.1 in der rechten Spalte) erhalten: für den Polfeldt-Ansatz sind wiederum die Überdeckungswahrscheinlichkeiten für kleine  $p$  zu groß und für große  $p$  zu klein, für das Gauß'sches Disjunktives Kriging sind die Überdeckungswahrscheinlichkeiten noch zu groß, wenn auch nicht mehr systematisch mit steigender erwarteter Wahrscheinlichkeit zunehmend zu groß. Die Überdeckungswahrscheinlichkeiten für Indikator-Kriging weichen bei sehr großen wie bei sehr kleinen erwarteten Wahrscheinlichkeiten stark von ebendiesen ab.

## 4.2. Fallstudie: Simulierte Daten

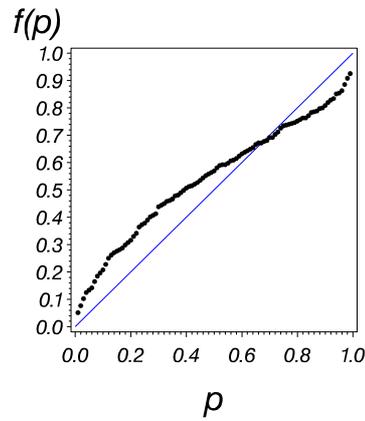
Da Wahrscheinlichkeiten nicht gemessen werden können, ist eine Kreuzvalidierung der Median-Indikator-Kriging-Prognose

$$\hat{F}_{MIK,Z(3472,5292)|Z^1=z^1}(59.5) = 0.26$$

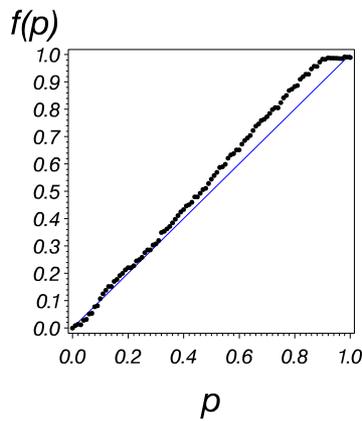
aus dem vorherigen Kapitel – also der mit 26% prognostizierten, durch die Stichprobe  $z^1$  bedingten Wahrscheinlichkeit, daß der Prozeß an der Stelle  $\mathbf{s}_0 = (3472, 5292)$  das 0.2-Quantil der empirischen Verteilung der Stichprobenwerte  $z^1$ , 59.5 cg/kg Mg, nicht überschreitet und somit eindeutig als Magnesiummangelstandort gelten muß – nicht möglich. Möglich wird eine solche Validierung nur, wenn die betrachteten Daten  $z_{(N)}$  Ziehungen aus einem bekannten räumlichen Zufallsprozess  $Z = \{Z(\mathbf{s}) : \mathbf{s} \in D\}$  sind, denn mit dem Pro-



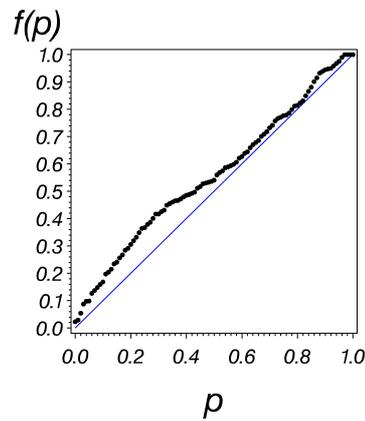
(a) Polfeldt-Ansatz



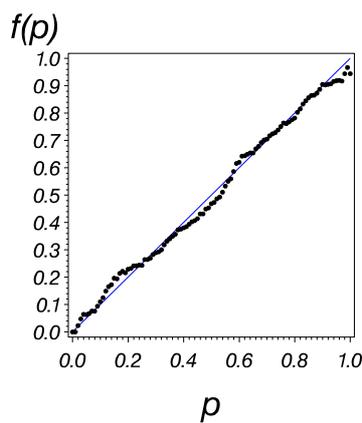
(b) Polfeldt-Ansatz



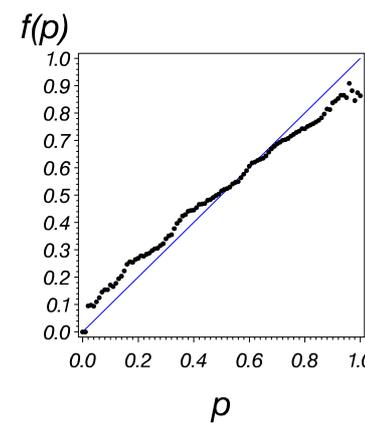
(c) Disjunktives Kriging



(d) Disjunktives Kriging



(e) Indikator-Kriging



(f) Indikator-Kriging

**Abbildung 4.5.:** Überdeckungswahrscheinlichkeiten  $f(p)$  über  $p$  nach Verfahren  $v = \{\text{Polfeldt-Ansatz, Indikator-Kriging, Disjunktives Kriging}\}$ ; Stichprobe  $z^1$  (links) und Stichprobe  $z^2$  (rechts).

#### 4. Bedingte Verteilungen in der Anwendung

---

zeß sind auch die bedingten lokalen Verteilungsfunktionen

$$F_{Z(\mathbf{s})|Z=\mathbf{z}}(z(\mathbf{s})) = P[Z(\mathbf{s}) \leq z(\mathbf{s}) | \mathbf{Z} = \mathbf{z}] \quad \forall \mathbf{s} \in D, z(\mathbf{s}) \in \mathbb{R}$$

bekannt. Ich simuliere daher mittels der LU-Decomposition (vergleiche beispielsweise Wameling (2002), S. 28f) zwei partielle Realisationen  $\psi_{(578)}^1$  und  $\psi_{(578)}^2$  des über den 578 aus Kapitel 3.2 bekannten Meßstellen definierten Teiles  $\Psi_{(578)}$  eines schwach stationären gauß'schen Prozesses. Der betrachtete Prozeß  $\Psi = \{\Psi(\mathbf{s}) : \mathbf{s} \in D\}$  ist dabei durch seine Erwartungswertfunktion

$$E[\Psi(\mathbf{s})] = 4.365 \quad \forall \mathbf{s} \in D$$

und die isotrope exponentielle Kovarianzfunktion

$$C_{\Psi}(|\mathbf{h}|) = c_p \cdot \exp\left(-\frac{3|\mathbf{h}|}{r_{\epsilon}}\right) + c_0 \quad (4.2.1)$$

mit  $c_p = 0.0656$ ,  $r_{\epsilon} = 42$  und  $c_0 = 0.088$  definiert. Da der Prozeß  $\Psi$  gauß'sch ist, wird der Polfeldt-Ansatz den beiden nicht-linearen Verfahren bei der Approximation bedingter lokaler Verteilungsfunktionen deutlich überlegen sein. Sein Erwartungswert und die Parameter der räumlichen Autokovarianzfunktion sind derart gewählt, daß sich für die folgende lognormale Transformation Stichproben ergeben, die den Magnesiumgehalten der Immissionsökologischen Waldzustandserfassung aus Kapitel 3.2 möglichst ähnlich sind. Um die Güte der prognostizierten Wahrscheinlichkeiten der Verfahren unter nicht gauß'schen Rahmenbedingungen miteinander zu vergleichen, betrachte ich daneben die beiden Transformationen

$$z_{(578)}^1 = \left\{ e^{(\psi(\mathbf{s}_n))} : \psi(\mathbf{s}_n) \in \psi_{(578)}^1 \right\} \text{ und } z_{(578)}^2 = \left\{ e^{(\psi(\mathbf{s}_n))} : \psi(\mathbf{s}_n) \in \psi_{(578)}^2 \right\}.$$

Sie sind partielle Realisationen des lognormalen Prozesses  $Z = e^{(\Psi)}$ , welcher die ersten beiden Momente

$$E[Z(\mathbf{s})] = e^{(E[\Psi(\mathbf{s})] + 0.5 \cdot \text{var}[\Psi(\mathbf{s})])} = 84.92767 \quad \forall \mathbf{s} \in D,$$

$$\text{cov}(Z(\mathbf{s}_n), Z(\mathbf{s}_m)) = E[Z(\mathbf{s}_n)] E[Z(\mathbf{s}_m)] (e^{\text{cov}(\Psi(\mathbf{s}_n), \Psi(\mathbf{s}_m))} - 1) \quad \forall \mathbf{s}_n, \mathbf{s}_m \in D,$$

besitzt (vergleiche Johnson u. Kotz (1970), S. 115). Um die partiellen Realisationen sprachlich von ihren im Rahmen der nicht-linearen Verfahren – Median-Indikator-Kriging und Gauß'sches Disjunktives Kriging – erstellten Transformationen zu unterscheiden, werde ich im Folgenden  $\psi_{(578)}^1$  und  $\psi_{(578)}^2$  sowie ihre Transformationen  $z_{(578)}^1$  und  $z_{(578)}^2$  nur noch als Stichproben bezeichnen.

Das Kovarianzmodell  $C_Z(|\mathbf{h}|)$  läßt sich nicht direkt aus dem bekannten Kovarianzmodell 4.2.1 des zugehörigen gauß'schen Zufallsprozesses  $\Psi$  herleiten, da die funktionale Form durch die Transformation zerstört wird (siehe Anhang A). Für jede Stelle  $\mathbf{s}$  im Untersuchungsgebiet gilt aber für die bedingte lokale Wahrscheinlichkeit

$$P[Z(\mathbf{s}) \leq z(\mathbf{s}) | \mathbf{Z} = \mathbf{z}],$$

**Tabelle 4.4.:** Deskriptive Statistiken der Stichproben; N bezeichnet den Stichprobenumfang, IQA den Interquartilsabstand, Shapiro den Wert der Teststatistik des Shapiro-Wilk-Tests auf Normalität und p-Wert den zu diesem Wert gehörigen p-Wert.

Statistik	$\psi_{(578)}^1$	$\psi_{(578)}^2$	$z_{(578)}^1$	$z_{(578)}^2$
N	578	578	578	578
Mittel	4.365	4.546	84.08	100.90
Schiefe	-0.013	-0.169	1.0827	1.0817
Varianz	0.134	0.139	1005.63	1442.67
Maximum	5.386	5.676	218.35	291.75
Median	4.368	4.559	78.90	95.50
Minimum	3.148	3.282	23.29	26.63
IQA	0.491	0.489	39.34	46.94
Shapiro	0.999	0.997	0.94	0.95
p-Wert	0.983	0.409	0.00	0.00

daß sie gleich der bekannten bedingten Wahrscheinlichkeit

$$P[\Psi(\mathbf{s}) \leq \psi(\mathbf{s}) | \Psi = \psi]$$

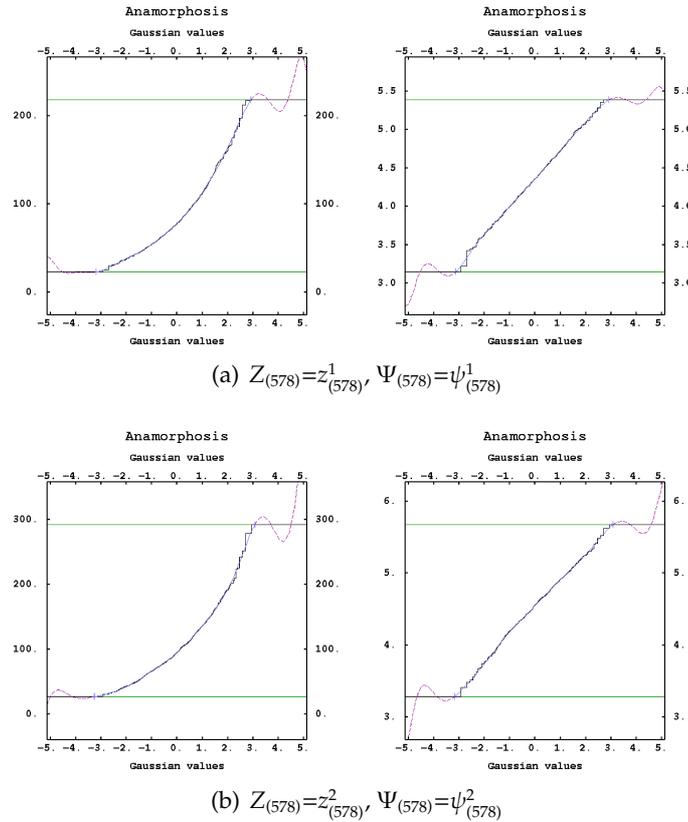
ist.

Tabelle 4.4 gibt die Charakteristika der vier Stichproben wieder: Das arithmetische Stichprobenmittel von  $\psi_{(578)}^1$  ist nahezu identisch mit dem Erwartungswert des Prozesses;  $\psi_{(578)}^2$  hingegen ist eine Stichprobe, deren Werte im Mittel deutlich über  $E[\Psi(\mathbf{s})]$  liegen. Beider Stichproben Varianzen sind etwas niedriger als  $C(0) = 0.1536$ , was sich wie im letzten Kapitel durch die nicht beachtete räumliche Abhängigkeit erklärt, und für beide Stichproben wird die Annahme einer Normalverteilung mit dem Shapiro-Wilk-Test zum Konfidenzniveau  $\alpha = 0.05$  beibehalten. Allerdings handelt es sich bei der betrachteten Verteilung um die von Beobachtungen (räumlich) abhängiger Variabler, Tests wie der Shapiro-Wilk-Test, die unabhängige, identisch verteilte Variablen voraussetzen, liefern daher verzerrte Ergebnisse (siehe Pardo-Igúzquiza u. Dowd (2004)). Dennoch wird nach Shapiro-Wilk die Nullhypothese der multivariaten Normalverteilung von  $F_{Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_K)}(z(\mathbf{s}_1), \dots, z(\mathbf{s}_K))$   $\mathbf{s}_k \in D$  für  $k = 1, \dots, K$  für beide Stichproben richtigerweise beibehalten. Die lognormal transformierten Stichproben „erben“ die Eigenschaften der gauß'schen hinsichtlich Lage und Streuung,  $z_{(578)}^1$  ähnelt der in Kapitel 3.2 beschriebenen Stichprobe bis auf die geringere Schiefe stark.

Jede der vier Stichproben wird nun zum eigenen Median binär und mittels der hermiteschen Polynome  $H_0$  bis  $H_{19}$  transformiert, es entstehen also insgesamt zwölf Datensätze, wobei die Indikator-Transformationen der normalen und der zugehörigen lognormalen Stichprobe identisch sind:

$$I(\Psi(\mathbf{s}) \leq \psi(\mathbf{s})) \equiv I(e^{(\Psi(\mathbf{s}))} \leq e^{(\psi(\mathbf{s}))}) \quad \forall \mathbf{s} \in D.$$

#### 4. Bedingte Verteilungen in der Anwendung



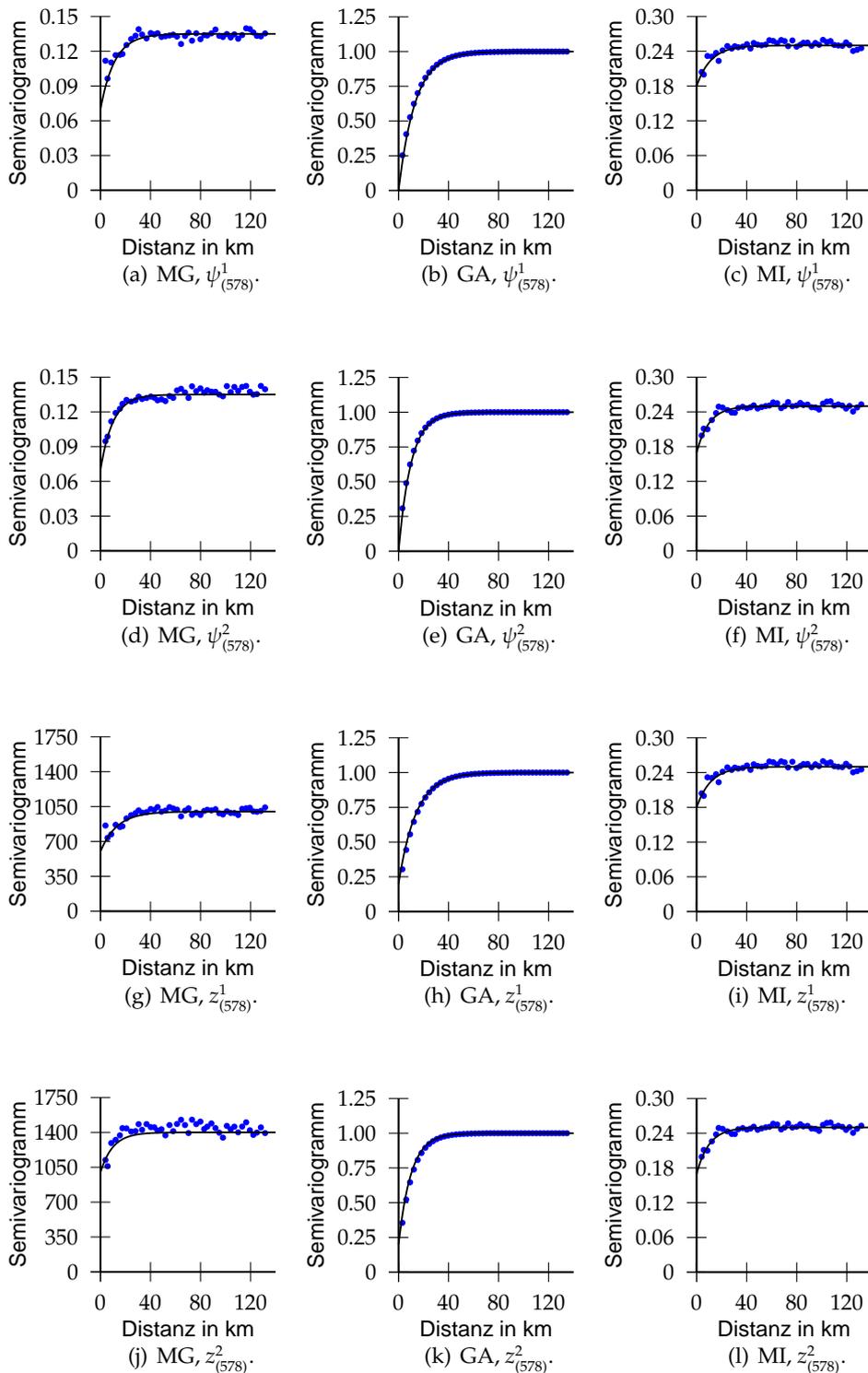
**Abbildung 4.6.:** Gauß'sche Anamorphosen der simulierten Daten, lognormal (links) und gauß'sch (rechts).

Die Koeffizienten der gauß'schen Anamorphosen für die lognormalen Stichproben  $z_{(578)}^1$  und  $z_{(578)}^2$  werden auch wie im vorigen Kapitel sehr schnell klein, die Transformationen in Abbildung 4.6 scheinen gut zu gelingen. Für die Stichproben  $\psi_{(578)}^1$  und  $\psi_{(578)}^2$  entsprechen sie in etwa den linearen Transformationen  $\frac{\psi(s_n) - 4.365}{\sqrt{0.134}}$  und  $\frac{\psi(s_n) - 4.546}{\sqrt{0.139}}$ , welche sich aus Tabelle 4.4 ergeben.

Die empirischen und angepassten Semivariogramme aller Stichproben und ihrer Transformationen zeigt Abbildung 4.7, Tabelle 4.5 enthält die entsprechenden Parameter.

Alle angepassten Semivariogrammfunktionen sind wie das des Prozesses vom exponentiellen Typus und ihre effektiven Reichweiten sind durchweg kleiner oder gleich der Reichweite des Prozeß-Semivariogrammes. Die Reichweiten der Semivariogramme aller Transformationen von  $\psi_{(578)}^1$  und  $z_{(578)}^1$  liegen dabei über den entsprechenden effektiven Reichweiten für  $\psi_{(578)}^2$  und  $z_{(578)}^2 - \psi_{(578)}^1$  war also eine glattere Stichprobe als  $\psi_{(578)}^2$ . Die Median-Indikator-Semivariogramme einer gauß'schen Stichprobe und ihrer lognormalen Transformation sind identisch, da, wie erwähnt, die Indikatoren gleich sind.

Es werden wie in Kapitel 4.1 die Wahrscheinlichkeiten zu neun Schwel-



**Abbildung 4.7.:** Semivariogramme; MG bezeichnet die untransformierten, MI die zum Median-Indikator transformierten und GA die mit den gauß'schen Anamorphosen transformierten simulierten Werte; Punkte stellen die Werte der empirischen Semivariogramme dar, die Linien das jeweilige angepasste Modell.

#### 4. Bedingte Verteilungen in der Anwendung

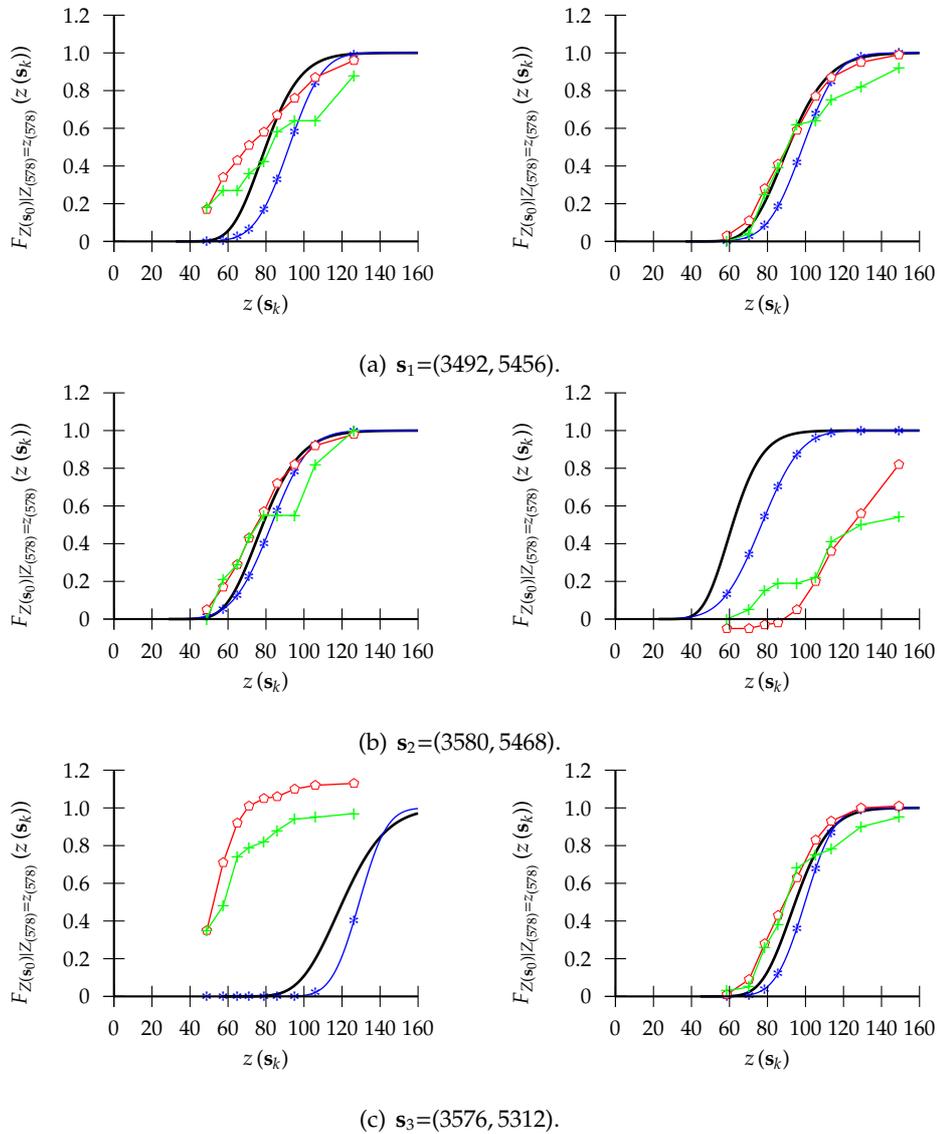
**Tabelle 4.5.:** Exponentielle Semivariogrammfunktionen:  $r_\epsilon$  ist die effektive Reichweite,  $c_p$  der partielle Schwellenwert und  $c_0$  der Nugget-Effekt (siehe Kapitel 2.1.1); MG bezeichnet die untransformierten, MI die zum Median-Indikator transformierten und GA die mit den gauß'schen Anamorphosen transformierten simulierten Werte.

Stichprobe	Variable	$r_\epsilon$	$c_p$	$c_0$
Prozeß	MG	42.0	0.0656	0.0880
$\psi_{(578)}^1$	MG	34.0	0.0650	0.0700
$\psi_{(578)}^2$	MG	30.0	0.0650	0.0700
$z_{(578)}^1$	MG	40.0	400.0	600.0
$z_{(578)}^2$	MG	30.0	400.0	1000.0
$\psi_{(578)}^1$	GA	40.0	1.0	0.0
$\psi_{(578)}^2$	GA	30.0	1.0	0.0
$z_{(578)}^1$	GA	42.0	0.8	0.2
$z_{(578)}^2$	GA	30.0	0.8	0.2
$\psi_{(578)}^1, z_{(578)}^1$	MI	34.0	0.0700	0.1800
$\psi_{(578)}^2, z_{(578)}^2$	MI	30.0	0.0800	0.1700

lenwerten betrachtet, die Schwellenwerte entsprechen den Dezilen der empirischen Verteilung der jeweiligen Stichprobe. Zur Prognose wird ein Gitter  $S_{(0)} = \{s_k \in D, k = 1, \dots, K = 1714\}$  genutzt, welches die Vervollständigung des Stichprobennetzes über ganz Baden-Württemberg darstellt.

Abbildung 4.8 zeigt für drei beispielhafte Punkte bedingte lokale Verteilungsfunktionen des lognormalen Prozesses und ihre jeweiligen Approximationen durch den Polfeldt-Ansatz, Gauß'sches Disjunktives Kriging und Median-Indikator-Kriging. Auf der linken Seite sind die durch die Stichprobe  $z_{(578)}^1$  bedingten lokalen Verteilungsfunktionen abgebildet, die rechte Seite zeigt die Verteilungen für  $z_{(578)}^2$ . Für den Punkt  $s_1 = (3492, 5456)$  liegt der durch beide Stichproben  $z_{(578)}^1$  und  $z_{(578)}^2$  bedingte Erwartungswert nahe dem aus der jeweiligen Stichprobe geschätzten konstanten Erwartungswert  $\hat{\mu}$  des Prozesses (siehe Tabelle 4.6). Für  $Z_{(578)} = z_{(578)}^1$  läßt sich nicht eindeutig entscheiden, welches Verfahren die bedingte lokale Verteilungsfunktion am besten approximiert: der Polfeldt-Ansatz unterschätzt die Dezile systematisch, die nicht-linearen überschätzen kleine und unterschätzen große Dezile. Für die Bedingung durch  $Z_{(578)} = z_{(578)}^2$  scheint Gauß'sches Disjunktives Kriging den anderen beiden Verfahren überlegen. Klare Entscheidungen gegen die nicht-linearen Verfahren verlangen die Approximationen in Abbildung 4.8(b), rechts und in Abbildung 4.8(c) links; Tabelle 4.6 weist für diese beiden Fälle extrem von  $\hat{\mu}$  abweichende bedingte Erwartungswerte aus.

Diese resultieren jeweils aus nahebei gelegenen extremen Meßwerten, wie Abbildung 4.9 zeigt. Daß die nicht-linearen Verfahren solche Verteilungen mit

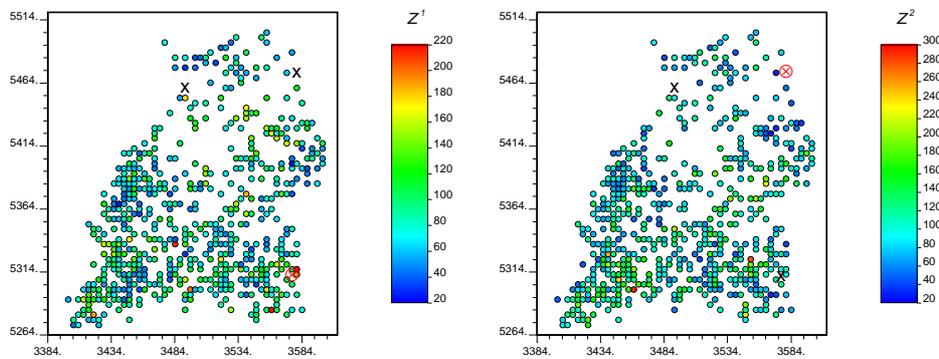


**Abbildung 4.8.:** Bedingte lokale Verteilungsfunktionen (Linie) und ihre Approximationen durch den Polfeldt-Ansatz (Asteriske), Indikator-Kriging (Kreuze) und Disjunktives Kriging (Fünfecke) an verschiedenen Orten für  $Z = \mathbf{z}_{(578)}^1$  (links) und  $Z = \mathbf{z}_{(578)}^2$  (rechts).

#### 4. Bedingte Verteilungen in der Anwendung

**Tabelle 4.6.:** Bedingte Momente einiger Variabler und aus den bedingenden Stichproben geschätzter Erwartungswert,  $\alpha_3$  bezeichnet die Schiefe.

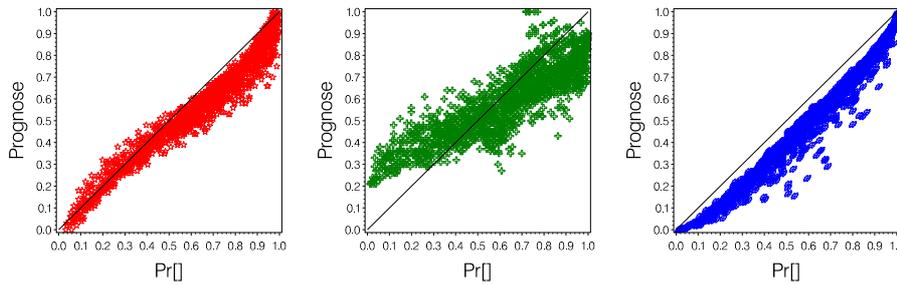
$\mathbf{B}$	$k$	$E[Z(\mathbf{s}_k) \mathbf{B}]$	$var[Z(\mathbf{s}_k) \mathbf{B}]$	$\alpha_3[Z(\mathbf{s}_k) \mathbf{B}]$	$\hat{\mu}_{(578)}$
$\mathbf{Z} = \mathbf{z}_{(578)}^1$	1	81.14194	216.8155	0.55038	83.7355
$\mathbf{Z} = \mathbf{z}_{(578)}^1$	2	79.83803	260.1324	0.614295	
$\mathbf{Z} = \mathbf{z}_{(578)}^1$	3	121.6624	336.8141	0.455976	
$\mathbf{Z} = \mathbf{z}_{(578)}^2$	1	92.40675	281.1946	0.55038	97.6943
$\mathbf{Z} = \mathbf{z}_{(578)}^2$	2	62.80282	160.9656	0.614295	
$\mathbf{Z} = \mathbf{z}_{(578)}^2$	3	95.70303	208.415	0.455976	



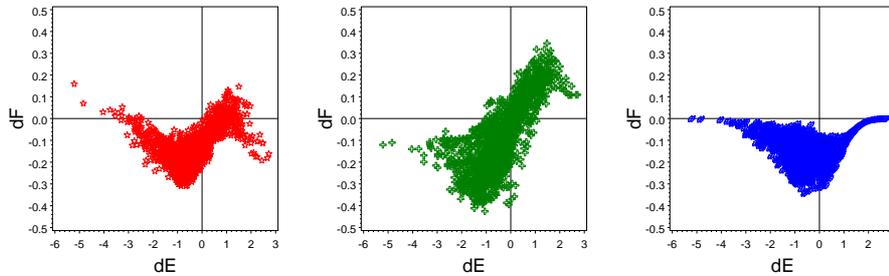
**Abbildung 4.9.:** Simulierte Werte  $z_{(578)}^1$  (links) und  $z_{(578)}^2$  (rechts); die Kreuze markieren die drei Prognosestellen aus Abbildung 4.8, eingekreiste Kreuze bezeichnen Orte, an denen die nicht-linearen Verfahren schlecht approximieren.

extremen Erwartungswerten schlecht approximieren, läßt sich am Beispiel des Indikator-Krigings anschaulich erklären: Liegt eine betrachtete Prognosestelle  $\mathbf{s}_0$  nahe einer Meßstelle mit sehr hohem Meßwert (wie es an der Stelle  $\mathbf{s}_3 = (3576, 5312)$  für  $z_{(578)}^1$  der Fall ist), ist der bedingte Erwartungswert an  $\mathbf{s}_0$  sehr hoch. Durch die Indikator-Transformation zu einem Schwellenwert, der kleiner als der Meßwert ist, wird aus dem sehr hohen Wert (ungeachtet seiner Differenz vom Schwellenwert) „nur“ eine Null, der Erwartungswert wird dadurch unterschätzt.

Es fällt auf, daß die Schiefen der bedingten lokalen Verteilungsfunktionen, welche nur von den bedingten lokalen Varianzen des gauß'schen Prozesses abhängen (siehe Johnson u. Kotz (1970), S. 115) und daher für  $Z_{(578)} = z_{(578)}^1$  und  $Z_{(578)} = z_{(578)}^2$  identisch sind, allesamt deutlich unter den Schiefen der Stichproben liegen. Die Schiefen der bedingten lokalen Verteilungsfunktionen aller 1714 Prognoseorte lagen im Intervall  $[0.332; 0.745]$ : Die bedingten lokalen Varianzen des gauß'schen Prozesses sind kleiner als seine Varianz  $C(0)$ , die bedingten



**Abbildung 4.10.:** Prognostizierte Wahrscheinlichkeiten zum Schwellenwert  $q_{0.5,z^1_{(578)}}$  über  $P\left[Z(\mathbf{s}_k) \leq q_{0.5,z^1_{(578)}} \mid \mathbf{Z} = \mathbf{z}^1_{(578)}\right]$ . Von links nach rechts: Gauß'sches Disjunktives Kriging, Median-Indikator-Kriging und Polfeldt-Ansatz.



**Abbildung 4.11.:** Absoluter Fehler  $dF$  der prognostizierte Wahrscheinlichkeiten zum Schwellenwert  $q_{0.5,z^2_{(578)}}$  über der normierten Distanz  $dE$  zwischen  $E\left[Z(\mathbf{s}_k) \mid \mathbf{Z} = \mathbf{z}^2_{(578)}\right]$  und  $q_{0.5,z^2_{(578)}}$ . Von links nach rechts: Gauß'sches Disjunktives Kriging, Median-Indikator-Kriging und Polfeldt-Ansatz.

lokalen Schiefen des lognormalen Prozesses sind daher relativ gering. Dies führt dazu, daß die bedingten lokalen Verteilungsfunktion (wie beispielsweise die in Abbildung 4.8) weniger schief sind, als ich es aufgrund der Momente der lognormalen Stichproben erwartet hatte. Daher fallen die (symmetrischen) Approximationen des Polfeldt-Ansatz besser als erwartet aus.

Um die Betrachtung etwas zu verdichten, lassen sich für einzelne Schwellenwerte  $q_{0,i,z}$  alle 1714 Prognosen einer Stichprobe über den zu prognostizierenden Wahrscheinlichkeiten wie in Abbildung 4.10 darstellen. Für den abgebildeten Schwellenwert  $q_{0.5,z^1_{(578)}} = 71.062$  streuen die aus  $z^1_{(578)}$  berechneten Prognosen des Median-Indikator-Kriging am stärksten, die Prognosen des Polfeldt-Ansatzes sind deutlich verzerrt, die Prognosen des Gauß'schen Disjunktiven Krigings wirken auf mich am besten.

Dieser Eindruck verstärkt sich noch, wenn wir Abbildung 4.11 betrachten: Die Abbildung zeigt die absoluten Fehler

#### 4. Bedingte Verteilungen in der Anwendung

---

$$dF = F_{Z(\mathbf{s}_k)|Z=z_{(578)}^2} \left( q_{0.5, z_{(578)}^2} \right) - \hat{F}_{v, Z(\mathbf{s}_k)|Z=z_{(578)}^2} \left( q_{0.5, z_{(578)}^2} \right)$$

des jeweiligen Verfahrens  $v$  über der – um der Vergleichbarkeit willen mit der bedingten lokalen Standardabweichung – normierten Differenz zwischen dem bedingten lokalen Mittelwert und dem betrachteten Schwellenwert

$$dE = \frac{E \left[ Z(\mathbf{s}_k) | Z = z_{(578)}^2 \right] - q_{0.5, z_{(578)}^2}}{\sqrt{\text{var} \left[ Z(\mathbf{s}_k) | Z = z_{(578)}^2 \right]}}$$

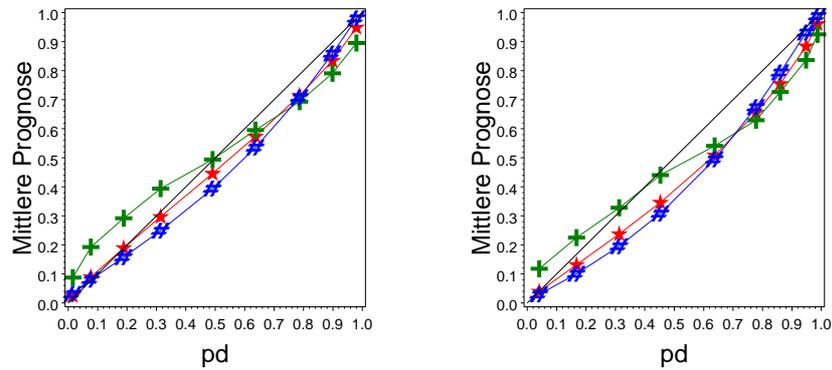
Hier wird deutlich, daß die Unterschreitungswahrscheinlichkeiten zum Schwellenwert  $q_{0.5, z_{(578)}^2} = 85.567$  durch Gauß'sches Disjunktives Kriging am besten vorhergesagt werden und daß der Polfeldt-Ansatz Unterschreitungswahrscheinlichkeiten für bedingte lokale Verteilungsfunktionen, deren Erwartungswert relativ dicht am betrachteten Schwellenwert liegt, stärker verzerrt und nicht präziser vorhersagt als die nicht-linearen Verfahren.

Eine weitere Verdichtung bieten die Prognosegütemaße: Sie lassen sich im Mittel über alle 1714 Prognosepunkte je betrachtetem Schwellenwert oder über alle Prognoseorte und Schwellenwerte hinweg gemittelt berechnen. Die Abbildungen 4.12 und 4.13 zeigen die Prognosegütemaße für die beiden lognormalen Stichproben; um eine einheitlich Skalierung der Achsen zu erreichen, werden die Prognosegütemaße dabei nicht über dem betrachteten Schwellenwert  $q_{0.i}(\mathbf{s}_k)$  sondern über

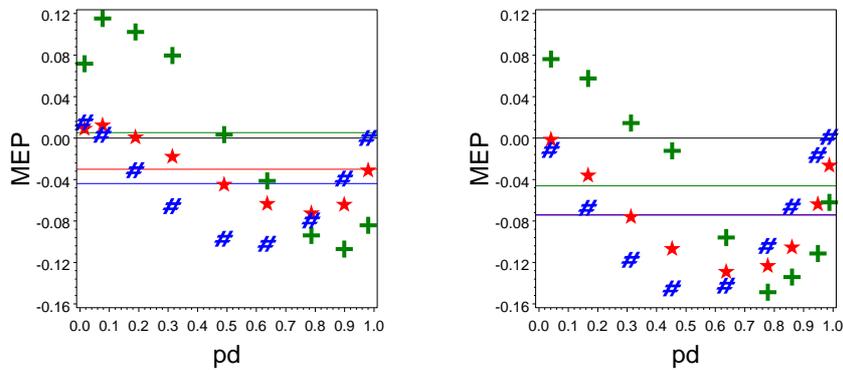
$$pd = \frac{1}{K} \sum_{k=1}^K P \left[ Z(\mathbf{s}_k) \leq q_{0.i}(\mathbf{s}_k) | Z_{(578)} = z_{(578)}^1, z_{(578)}^2 \right],$$

der über alle Prognoseorte gemittelten Wahrscheinlichkeit, daß der Prozeß an der betrachteten Stelle das  $i$ -te Dezil nicht überschreitet, dargestellt. Darstellungen der mittleren vorhergesagten Unterschreitungswahrscheinlichkeit je Schwellenwert über der mittleren Unterschreitungswahrscheinlichkeit zu diesem Schwellenwert (siehe Abbildung 4.12(a)) zeigen im Gegensatz zu den in der Darstellung ähnlichen Überdeckungswahrscheinlichkeiten in Kapitel 4.1 keine Überlegenheit der nicht-linearen Verfahren. Während Median-Indikator-Kriging einen deutlich kleineren mittleren Fehler  $MEP_v$  als die anderen beiden Verfahren aufweist, ist der mittlere quadratische Fehler  $MSEP_v$  des Median-Indikator-Krigings größer als die des Polfeldt-Ansatzes und des Gauß'schen Disjunktiven Krigings.

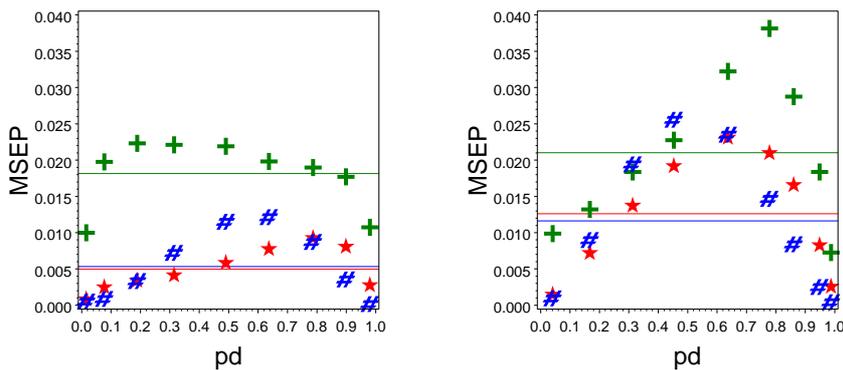
Der mittlere Rang des absoluten Fehlers (siehe Abbildung 4.13(a)) zeigt für beide Stichproben eine leichte Überlegenheit des Polfeldt-Ansatzes, Median-Indikator-Kriging wird am schlechtesten bewertet. Allerdings sind die Standardabweichungen der Ränge durchgängig recht hoch, die Verfahren wechseln also über den 1714 Prognosepunkten häufig den Rang des absoluten Prognosefehlers. Alle Verfahren zeigen in Abbildung 4.13(b) starke Verzerrungen für einzelne Dezile  $q_{0.i,z}$  (siehe auch Abbildung 4.10, dort sind die Prognosen des Polfeldt-Ansatzes durchweg zu niedrig), Median-Indikator-Kriging zeigt aufgrund seines hohen  $MSEP_v$  die geringste relative Verzerrung. Dies ist nicht



(a) Mittlere Approximation

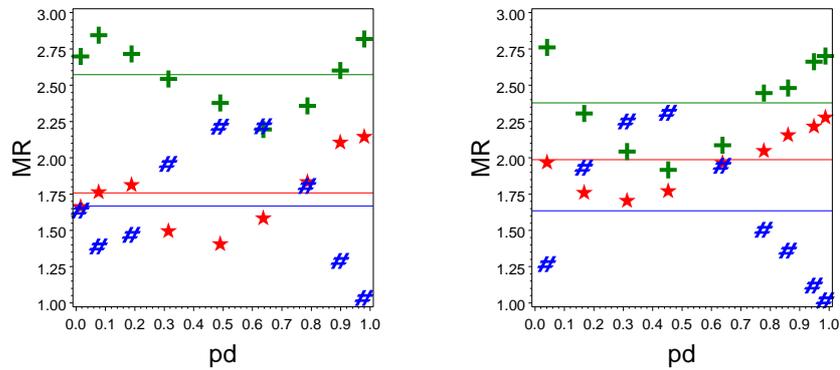


(b)  $MEP_v$

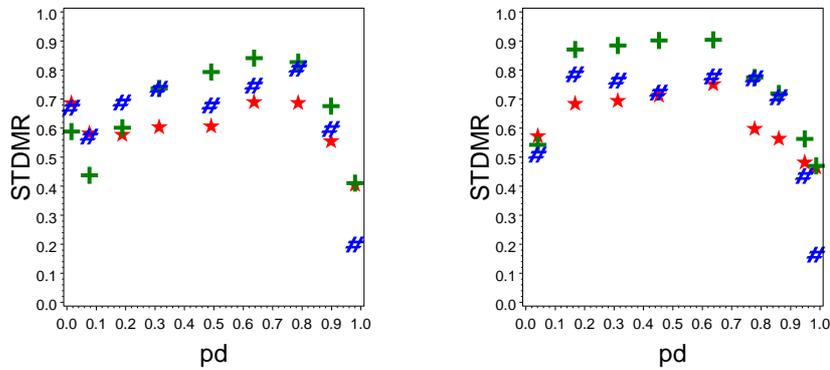


(c)  $MSEP_v$

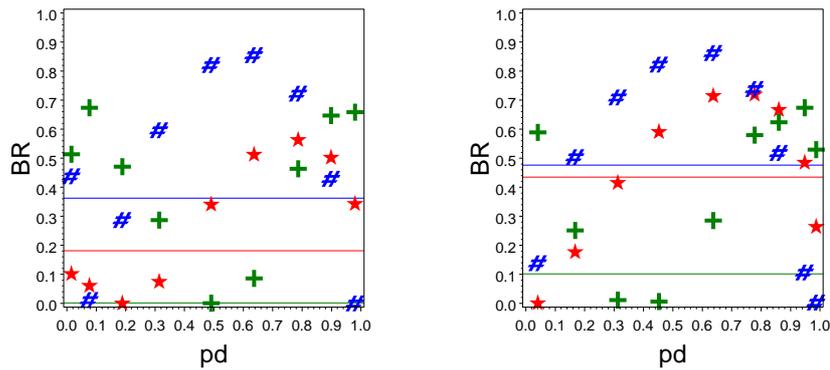
**Abbildung 4.12.:** Maße Prognosegüte I für den Polfeldt-Ansatz (Doppelkreuze), Gauß'sches Disjunktives Kriging (Sterne) und Median-Indikator-Kriging (Kreuze);  $Z_{(578)}=z_{(578)}^1$  (links) und  $Z_{(578)}=z_{(578)}^2$  (rechts), die horizontalen Linien entsprechen den arithmetischen Mitteln je Verfahren über alle Schwellenwerte und Prognoseorte.



(a)  $MR_v$



(b)  $SDR_v$



(c)  $BR_v$

**Abbildung 4.13.:** Maße Prognosegüte II für den Polfeldt-Ansatz (Doppelkreuze), Gauß'sches Disjunktives Kriging (Sterne) und Median-Indikator-Kriging (Kreuze);  $Z_{(578)}=z_{(578)}^1$  (links) und  $Z_{(578)}=z_{(578)}^2$  (rechts), die horizontalen Linien entsprechen den arithmetischen Mitteln je Verfahren über alle Schwellenwerte und Prognoseorte.

zwingend eine positive Bewertung: Eine geringe relative Verzerrung bei gleichzeitig hohem  $MSEP_v$  kennzeichnet ein genaues, aber unpräzises Verfahren, eine hohe relative Verzerrung bei niedrigem  $MSEP_v$  hingegen ein präzises aber ungenaues Verfahren. Unverzerrtheit ist nur ein, wenn auch meist für wichtig erachtetes, Bewertungskriterium für ein Verfahren – für gewöhnlich wird die Genauigkeit in Form des mittleren quadratischen Prognosefehlers bewertet, ist die *relative* Verzerrung neutral für die Bewertung.

Daß die Repräsentativität der Stichprobe bezüglich des Erwartungswertes, ob also  $\hat{\mu}$  ein guter Schätzwert für  $E[Z(s)]$  ist, offensichtlich keinen Einfluß auf die relative Bewertung des Polfeldt-Ansatzes besitzt, ist dadurch zu erklären, daß für die meisten Prognoseorte genug relativ zur räumlichen Abhängigkeit naheliegende Daten verfügbar sind und das Gewicht des Erwartungswertes beim Kriging daher überwiegend sehr klein ist.

Für die Realisationen  $\psi_{(578)}^1$  und  $\psi_{(578)}^2$  des gauß'schen Prozesses ist, wie sich Tabelle 4.7 entnehmen läßt, der lineare Polfeldt-Ansatz erwartungsgemäß den nicht-linearen Verfahren deutlich überlegen.

**Tabelle 4.7.:** Maße der Prognosegüte für  $\psi_{(578)}^1$  und  $\psi_{(578)}^2$ ; Verf. gibt die Verfahren Gauß'sches Disjunktives Kriging (gDK), Median-Indikator-Kriging (MIK) und den Polfeldt-Ansatz (Pol) an.

$\Psi$	Verf.	$q_i$	$MEP_v$	$MSEP_v$	$MR_v$	$SDR_v$	$BR_v$
$\psi_{(578)}^1$	gDK	1	0.0155	0.0013	2.1739	0.3822	0.1837
$\psi_{(578)}^1$	gDK	2	0.0299	0.0036	2.1184	0.3391	0.2447
$\psi_{(578)}^1$	gDK	3	0.0315	0.0044	2.1581	0.3912	0.2271
$\psi_{(578)}^1$	gDK	4	0.0223	0.0040	2.1377	0.3996	0.1233
$\psi_{(578)}^1$	gDK	5	0.0024	0.0037	2.1097	0.3579	0.0016
$\psi_{(578)}^1$	gDK	6	-0.0153	0.0041	2.1674	0.4301	0.0572
$\psi_{(578)}^1$	gDK	7	-0.0298	0.0047	2.1301	0.3866	0.1882
$\psi_{(578)}^1$	gDK	8	-0.0323	0.0042	2.1336	0.3571	0.2458
$\psi_{(578)}^1$	gDK	9	-0.0176	0.0014	2.1021	0.3124	0.2262
$\psi_{(578)}^1$	MIK	1	0.0717	0.0100	2.8250	0.3801	0.5139
$\psi_{(578)}^1$	MIK	2	0.1194	0.0210	2.8746	0.3366	0.6781
$\psi_{(578)}^1$	MIK	3	0.1025	0.0223	2.8250	0.3952	0.4710
$\psi_{(578)}^1$	MIK	4	0.0797	0.0221	2.8390	0.3740	0.2869
$\psi_{(578)}^1$	MIK	5	0.0033	0.0219	2.8676	0.3591	0.0005
$\psi_{(578)}^1$	MIK	6	-0.0414	0.0198	2.8022	0.4142	0.0862
$\psi_{(578)}^1$	MIK	7	-0.0939	0.0190	2.8471	0.3696	0.4637
$\psi_{(578)}^1$	MIK	8	-0.1070	0.0177	2.8541	0.3661	0.6473
$\psi_{(578)}^1$	MIK	9	-0.0841	0.0107	2.8932	0.3127	0.6592
$\psi_{(578)}^1$	Pol	1	0.0003	0.0000	1.0012	0.0341	0.0395
$\psi_{(578)}^1$	Pol	2	0.0007	0.0000	1.0070	0.0834	0.0666
$\psi_{(578)}^1$	Pol	3	0.0011	0.0000	1.0169	0.1378	0.0756

4. Bedingte Verteilungen in der Anwendung

Tabelle 4.7.: (forstgesetzt)

$\Psi$	Verf.	$q_i$	$MEP_v$	$MSEP_v$	$MR_v$	$SDR_v$	$BR_v$
$\psi_{(578)}^1$	Pol	4	0.0012	0.0000	1.0233	0.1548	0.0709
$\psi_{(578)}^1$	Pol	5	0.0011	0.0000	1.0228	0.1530	0.0537
$\psi_{(578)}^1$	Pol	6	0.0008	0.0000	1.0303	0.1782	0.0335
$\psi_{(578)}^1$	Pol	7	0.0005	0.0000	1.0228	0.1568	0.0135
$\psi_{(578)}^1$	Pol	8	0.0002	0.0000	1.0123	0.1249	0.0027
$\psi_{(578)}^1$	Pol	9	-0.0000	0.0000	1.0047	0.0763	0.0006
$\psi_{(578)}^2$	gDK	1	0.0093	0.0017	1.9597	0.5729	0.0520
$\psi_{(578)}^2$	gDK	2	-0.0096	0.0055	1.9854	0.6845	0.0170
$\psi_{(578)}^2$	gDK	3	-0.0383	0.0085	2.0379	0.7243	0.1721
$\psi_{(578)}^2$	gDK	4	-0.0621	0.0109	2.1383	0.6565	0.3545
$\psi_{(578)}^2$	gDK	5	-0.0805	0.0126	2.1814	0.5690	0.5131
$\psi_{(578)}^2$	gDK	6	-0.0782	0.0114	2.0735	0.4381	0.5370
$\psi_{(578)}^2$	gDK	7	-0.0666	0.0089	2.1120	0.4713	0.4981
$\psi_{(578)}^2$	gDK	8	-0.0395	0.0044	2.1418	0.4285	0.3519
$\psi_{(578)}^2$	gDK	9	-0.0157	0.0013	2.1908	0.5060	0.1861
$\psi_{(578)}^2$	MIK	1	0.0763	0.0099	2.7631	0.5147	0.5890
$\psi_{(578)}^2$	MIK	2	0.0577	0.0132	2.5251	0.7405	0.2513
$\psi_{(578)}^2$	MIK	3	0.0146	0.0184	2.4428	0.8053	0.0116
$\psi_{(578)}^2$	MIK	4	-0.0123	0.0228	2.4300	0.8172	0.0066
$\psi_{(578)}^2$	MIK	5	-0.0959	0.0322	2.5770	0.7095	0.2854
$\psi_{(578)}^2$	MIK	6	-0.1487	0.0382	2.8046	0.4913	0.5796
$\psi_{(578)}^2$	MIK	7	-0.1340	0.0288	2.7625	0.5207	0.6238
$\psi_{(578)}^2$	MIK	8	-0.1113	0.0184	2.7987	0.4581	0.6739
$\psi_{(578)}^2$	MIK	9	-0.0621	0.0073	2.6657	0.6119	0.5300
$\psi_{(578)}^2$	Pol	1	-0.0114	0.0004	1.2771	0.5485	0.2905
$\psi_{(578)}^2$	Pol	2	-0.0307	0.0021	1.4895	0.6688	0.4513
$\psi_{(578)}^2$	Pol	3	-0.0420	0.0033	1.5193	0.6318	0.5317
$\psi_{(578)}^2$	Pol	4	-0.0456	0.0036	1.4317	0.6124	0.5725
$\psi_{(578)}^2$	Pol	5	-0.0413	0.0029	1.2415	0.4820	0.5885
$\psi_{(578)}^2$	Pol	6	-0.0312	0.0017	1.1219	0.3786	0.5650
$\psi_{(578)}^2$	Pol	7	-0.0224	0.0010	1.1254	0.3851	0.5218
$\psi_{(578)}^2$	Pol	8	-0.0100	0.0002	1.0595	0.2533	0.4021
$\psi_{(578)}^2$	Pol	9	-0.0028	0.0000	1.1435	0.3960	0.2441

## 5. Diskussion

### 5.1. Bedingte Fehler und Konfidenzen

Einer der größten Nachteile der Krigingverfahren ist die Unabhängigkeit ihrer Varianz, also ihres mittleren quadratischen Prognosefehlers, von den zur Vorhersage genutzten Daten. Für die Berechnung des bedingten Fehlers ist die Kenntnis des bedingten Erwartungswertes und der bedingten Varianz vonnöten, die Abschätzung des Unterschiedes zwischen unbedingtem und bedingtem mittleren quadratischen Prognosefehler kann daher nur für einen Spezialfall gelingen: Wenn wir annehmen, daß die beobachteten Daten eine partielle Realisation eines schwach stationären gauß'schen Zufallsprozesses sind, dann ist die Differenz  $\sigma_{ok}^2(\mathbf{s}_0) - \sigma_{ok|Z=z}^2(\mathbf{s}_0)$  zwischen unbedingtem und bedingtem mittleren quadratischen Prognosefehler des Gewöhnlichen Krigings eine Funktion nur des quadratisch eingehenden Gewichtes des Erwartungswertes beim Kriging und der Differenz aus quadratischem Fehler der Erwartungswertschätzung und seiner Erwartung (siehe 3.1.3, 3.1.5). Der Einfluß der Meßwerte wird in diesem Fall also auf die aus ihnen erfolgende Erwartungswertschätzung und auf das Gewicht dieser Schätzung beim Kriging reduziert.

Während der Fehler der Erwartungswertschätzung in der Praxis unbekannt ist, kennen wir für den erwähnten Spezialfall die Verteilung des Quotienten aus quadratischem Fehler der Erwartungswertschätzung und seiner Erwartung: er folgt einer  $\chi_1^2$ -Verteilung und ist mit einer Wahrscheinlichkeit von  $Prob\chi_1^2 < 1 \approx 68\%$  kleiner als Eins. Ist dieser Quotient kleiner Eins, der quadratische Fehler also kleiner als seine Erwartung, führt das Ersetzen des bedingten mittleren quadratischen Prognosefehlers durch den unbedingten zu einer konservativen Schätzung der Genauigkeit der Prognose, das heißt wir könnten die tatsächliche höhere bedingte Präzision besser einschätzen, wenn der bedingte mittlere quadratische Prognosefehler bekannt oder schätzbar wäre.

Der bestimmende Faktor der Differenz  $\sigma_{ok}^2(\mathbf{s}_0) - \sigma_{ok|Z=z}^2(\mathbf{s}_0)$  ist das quadratische Gewicht des Erwartungswertes beim Kriging. Es kann für eine Prognosestelle – bei bekannter räumlicher Abhängigkeitsfunktion noch vor Erhebung der Daten – aus der räumlichen Verteilung der Meßstellen und der Prognosestelle bestimmt werden und liegt bei geeigneter Anordnung der Meßstellen meist nahe Null, womit auch  $\sigma_{ok}^2(\mathbf{s}_0) - \sigma_{ok|Z=z}^2(\mathbf{s}_0)$  in gauß'schen Prozessen meist nahe Null ist. In Anbetracht der Tatsache, daß für gauß'sche Prozesse der Einfache Krigingprädiktor der bedingte Erwartungswert und sein bedingter mittlere quadratische Prognosefehler gleich seinem unbedingten mittleren quadratischen Prognosefehler ist, wirkt dieses Ergebnis plausibel.

Selbst bei mäßig großem Gewicht des Erwartungswertes beim Kriging ist die relative Differenz  $(\sigma_{ok}^2(\mathbf{s}_0) - \sigma_{ok|Z=z}^2(\mathbf{s}_0)) \cdot (\sigma_{ok|Z=z}^2(\mathbf{s}_0))^{-1} = R_{duc}^2$  vernachlässigbar

klein, da auch  $\sigma_{ok|Z=z}^2(\mathbf{s}_0)$  mit steigendem Gewicht des Erwartungswertes beim Kriging größer wird.

Die Breite des exakten Prognoseintervalles für den Wert  $z(\mathbf{s}_0)$  ergibt sich als  $B_{Z=z} = B \cdot (1 + R_{duc})^{-1}$  aus der Breite  $B$  des cressieschen Prognoseintervalles. Wie in Kapitel 3.2 gezeigt wurde, ist  $R_{duc}$  für kleine  $R_{duc}^2$  ebenfalls klein und bewegt sich in der Mehrzahl der Anwendungen in der Größenordnung des Fehlers, der durch Unkenntnis des Erwartungswertes entsteht. Die relative Differenz zwischen dem von Cressie (1991), S. 122, vorgeschlagenen (cressieschen) Prognoseintervall und dem exakten Prognoseintervall ist daher für die meisten Prognosen, bei denen die Daten als Stichprobe eines schwach stationären gauß'schen Prozesses angesehen werden, vernachlässigbar klein.

Bei schwach ausgeprägter räumlicher Abhängigkeit zwischen Prognose- und Meßstellen entsteht neben dem hohen Gewicht des Erwartungswertes beim Kriging das folgende weitere Problem: Das Semivariogramm des Prozesses wird beim Kriging für gewöhnlich durch ein an das empirische Semivariogramm der Stichprobe angepaßtes Modell geschätzt. Dadurch sind sämtliche verwendeten Kovarianzmatrizen und -vektoren nur Schätzungen und wir kennen statt des besten linearen unverzerrten Prädiktors  $p_1(\mathbf{s}_0, \mathbf{Z}, \boldsymbol{\theta})$  bei bekannten Kovarianzparametern  $\boldsymbol{\theta}$  nur den EBLUP (Stein (1999), S. 200), den empirischen besten linearen unverzerrten Prädiktor  $p_2(\mathbf{s}_0, \mathbf{Z}, \hat{\boldsymbol{\theta}})$ . Dementsprechend berechnen wir anstatt des theoretischen Krigingfehlers  $m_1(\boldsymbol{\theta}) = E\left[\{Z(\mathbf{s}_0) - p_1(\mathbf{s}_0, \mathbf{Z}, \boldsymbol{\theta})\}^2\right]$  den Fehler  $m_1(\hat{\boldsymbol{\theta}})$ , der leider nicht dem relevanten Fehler  $m_2(\boldsymbol{\theta}) = E\left[\{Z(\mathbf{s}_0) - p_2(\mathbf{s}_0, \mathbf{Z}, \hat{\boldsymbol{\theta}})\}^2\right]$  entspricht. Abgesehen von einfachen Spezialfällen (Harville u. Jeske (1992), S. 725) ist für  $m_2(\boldsymbol{\theta})$  noch immer keine geschlossene Darstellung bekannt (vergleiche Zimmerman (2006), S. 664f). Nach Zimmerman (2006), S. 665 wird in der Praxis  $m_2(\boldsymbol{\theta})$  mit  $m_1(\boldsymbol{\theta})$  gleichgesetzt und durch  $m_1(\hat{\boldsymbol{\theta}})$  geschätzt. Im gleichen Artikel zeigt er auch, daß Stichprobennetze, die eine möglichst gute Schätzung von  $\boldsymbol{\theta}$  ermöglichen, starke Klumpungen der Stichprobenpunkte aufweisen – im Gegensatz zu gleichmäßigen Netzen, die den maximalen Krigingfehler über einem Prognosegebiet minimieren. Eine gute Schätzung  $\hat{\boldsymbol{\theta}}$  bei gleichzeitig kleinem Fehler  $m_1(\hat{\boldsymbol{\theta}})$  erscheint daher nicht möglich. Zimmerman u. Cressie (1992), Beispiel 3, zeigen für einen schwach stationären isotropen gauß'schen Zufallsprozeß mit sphärischer Kovarianzfunktion, daß bei sehr schwacher räumlicher Abhängigkeit die relative Verzerrung  $\left|E\left[m_1(\hat{\boldsymbol{\theta}})\right] - m_2(\boldsymbol{\theta})\right| \cdot (m_2(\boldsymbol{\theta}))^{-1}$  über dreißig Prozent steigen kann.

Es ist anzunehmen, daß eine Analogie für  $E\left[\{Z(\mathbf{s}_0) - p(\mathbf{s}_0, \mathbf{Z}, \hat{\boldsymbol{\theta}})\}^2 \mid \mathbf{Z}\right]$  gilt, der Nachweis ist aber meines Wissensstandes erst noch zu erbringen.

In der Praxis würde ich daher die von Cressie (1991), S. 122 und 155, vorgeschlagene Näherung exakter Prognoseintervalle in einem als schwach stationär und gauß'sch angenommenen Prozeß als ausreichend betrachten: In den Fällen, in denen die relative Differenz aus unbedingtem und bedingtem mittleren quadratischen Prognosefehler aufgrund mangelnder räumlicher Abhängigkeit zwischen Prognose- und verfügbaren Meßstellen und damit hohem Gewicht

des Erwartungswertes beim Kriging merklich groß wird, besteht ohnehin die Gefahr, daß jeder mit geschätzter Kovarianzfunktion berechnete Krigingfehler den jeweiligen wahren nur unzureichend nähert und die aus ihm abgeleiteten Prognoseintervalle – ob bedingt oder nicht – sehr ungenau sind.

## 5.2. Approximation bedingter lokaler Verteilungsfunktionen

Die bedingte lokale Verteilungsfunktion stellt den größten aufgrund der erhobenen Stichprobe möglichen Wissenstand über den der Stichprobe zugrundeliegenden räumlichen Zufallsprozeß an einer Prognosestelle dar. Sie beinhaltet mit dem bedingten Erwartungswert den optimalen Prädiktor und das bedingte Prognoseintervall zum Konfidenzniveau  $1 - \alpha$ .

Alle Krigingverfahren lassen sich zur Approximation der bedingten lokalen Verteilungsfunktion nutzen, sie liefern somit Näherungswerte für den bedingten Erwartungswert und für die Grenzen des bedingten Prognoseintervalles. Eine andere Klasse geostatistischer Verfahren, die bedingten Simulationen, unterstellen die gleichen Annahmen wie entsprechende Krigingverfahren oder nutzen diese sogar direkt, und liefern bei ausreichender Iteration gute Näherungen der Ergebnisse des jeweiligen Krigingverfahrens. Die in der praktischen Umsetzung einfachsten auf Kriging basierenden Approximationsverfahren sind der Polfeldt-Ansatz, Gauß'sches Disjunktives Kriging und Median-Indikator-Kriging, die jeweils die Modellierung lediglich einer Semivariogrammfunktion und keine Modellierung der Erwartungswertfunktion benötigen. Der Polfeldt-Ansatz unterstellt, daß der zugrundeliegende Zufallsprozeß schwach stationär und gauß'sch ist, approximiert den bedingten Erwartungswert durch die Prognose des Gewöhnlichen Krigings und die bedingte Varianz durch den unbedingten mittleren quadratischen Prognosefehler des Gewöhnlichen Krigingprädiktors. Gauß'sches Disjunktives Kriging und Median-Indikator-Kriging prognostizieren Funktionswerte der bedingten lokalen Verteilungsfunktion zu einem Satz diskreter Schwellenwerte, zwischen diesen Prognosen wird linear interpoliert. Gauß'sches Disjunktives Kriging nimmt an, daß die Variablen, deren Beobachtungen die transformierten Meßwerte sind, bivariat normalverteilt sind, während Median-Indikator-Kriging die Annahme proportionaler räumlicher Abhängigkeiten aller zum Satz der Schwellenwerte definierter Indikatorprozesse unterstellt.

Bei der Kreuzvalidierung der Prognosen einzelner Meßwerte aus zwei Stichproben mit schiefen Verteilungen zeigt sich Gauß'sches Disjunktives Kriging den anderen beiden Verfahren überlegen. Deren Annahmen – schwach stationärer gauß'scher Prozess für den Polfeldt-Ansatz und proportionale räumliche Abhängigkeiten für Median-Indikator-Kriging – im Gegensatz zu der des Gauß'schen Disjunktiven Krigings wahrscheinlich nicht erfüllt sind. Daß alle drei Verfahren mit der variableren, schiefen Stichprobe  $z^1$  als Modellierungstichprobe ungenauere Prognosen treffen als mit der Stichprobe  $z^2$ , läßt sich anhand Polfeldt-Ansatz anschaulich erklären: er geht von gauß'schen – also symmetrischen – bedingten lokalen Verteilungsfunktionen aus. Diese Annah-

me wird für jede Prognose zu der Annahme, daß der Vektor  $[Z(s_0), \dots, Z(s_N)]^T$  multivariat normalverteilt ist und hängt hauptsächlich von der Modellierungstichprobe ab. Je schief und variabler sie ist, desto ungenauer werden die Prognosen des Polfeldt-Ansatz.

Die Beurteilung der approximierten bedingten lokalen Verteilungsfunktionen durch Überdeckungswahrscheinlichkeiten fällt für den Polfeldt-Ansatz schlechter aus als für die beiden nicht-linearen Verfahren (siehe Abbildung 4.1). Dabei sind die Überdeckungswahrscheinlichkeiten für den Polfeldt-Ansatz für kleine erwartete Wahrscheinlichkeiten zu groß und für große erwartete Wahrscheinlichkeiten zu klein; diese Struktur findet sich auch bei Papritz u. Moyeed (1999). Erklärt werden kann sie durch die Betrachtung der Prognosen für bekannte Wahrscheinlichkeiten eines simulierten lognormalen Prozesses: Der Polfeldt-Ansatz trifft die am stärksten verzerrten Prognosen der Unterscheidungswahrscheinlichkeiten – sowohl im Mittel aller Prognoseorte zu den meisten Schwellenwerten als auch im Mittel über alle Schwellenwerte und speziell für bedingte lokale Verteilungsfunktionen, deren bedingter Erwartungswert relativ nahe am betrachteten Schwellenwert liegt. Die Prognosen des Median-Indikator-Kriging sind die unpräziesesten, während Gauß'sches Disjunktives Kriging einen Mittelweg bezüglich Präzision und Genauigkeit darstellt. Die Betrachtung einzelner bedingter lokaler Verteilungsfunktionen zeigt hingegen eine deutlich bessere Approximation der Verteilungen von Variablen mit extremem bedingtem Erwartungswert durch den Polfeldt-Ansatz: die nicht-linearen Verfahren sind nicht in der Lage, stark genug auf extreme, nahe der Prognosestelle gelegene Meßwerte zu reagieren.

Während also die nicht-linearen Verfahren im Mittel über viele Prognosestellen, also über Flächen, genauere Approximationen der bedingten lokalen Verteilungsfunktionen sehr schiefer Prozesse liefern, ist dieser Vorteil bei mäßig schiefen, aber immer noch klar nicht gauß'schen Prozessen nicht mehr festzustellen. Zur Näherung der bedingten lokalen Verteilungsfunktionen an einer Stelle ist der Polfeldt-Ansatz sogar oft besser geeignet – zumal für die betrachteten lognormalen Prozesse die bedingten lokalen Verteilungsfunktionen deutlich weniger schief sind als die Verteilung der Stichprobenwerte. Die gute Bewertung des Polfeldt-Ansatzes gilt insbesondere für Stellen mit relativ zur räumlichen Abhängigkeit nahegelegenen extremen Meßwerten, da die transformierenden Verfahren auf diese extreme bedingte Erwartungswerte verursachenden Messungen nicht so flexibel wie der Polfeldt-Ansatz reagieren können (siehe Seite 62).

## 6. Zusammenfassung

Die meisten Merkmale, die wie Entnadelungsprozente, mittlere Grundflächen oder  $h/d$ -Verhältnisse Ziel forstlicher Datenerhebungen sind, haben einen räumlichen Bezug. Für viele dieser Messungen muß angenommen werden, daß die an unterschiedlichen Stellen beprobten Merkmale nicht voneinander unabhängig, sondern miteinander korreliert sind. Nimmt die Korrelation zwischen zwei Stellen mit deren abnehmender räumlicher Distanz zu, sprechen wir von räumlicher Autokorrelation.

Diese Art der räumliche Abhängigkeit machen sich die Krigingverfahren zunutze, um genauere Prognosen der Merkmalswerte an nicht erhobenen Stellen in der Fläche zu treffen als dies beispielsweise mit dem durch den Mittelwert der Messungen geschätzten Erwartungswert möglich ist. Die Struktur der räumliche Abhängigkeit, wie stark sich also das Merkmal an unterschiedlichen Stellen selbst beeinflußt, wird dabei aus den Messungen geschätzt.

Neben der Prognose eines lokalen Merkmalswertes liefern die Krigingverfahren mit dem mittleren quadratischen Prognosefehler auch ein Maß für die Genauigkeit des zur Prognose genutzten Prädiktors. Da der mittlere quadratische Prognosefehler eine Funktion der räumlichen Anordnung von Prognose- und Meßstellen unter Annahme der aus den Messungen geschätzten räumlichen Abhängigkeit ist, reagiert er für eine Prognosestelle nicht auf die ihn direkt umgebenden Meßwerte – im Gegensatz zur Prognose, die von naheliegenden Meßwerten stark geprägt wird. Diese Unbedingtheit des mittleren quadratischen Prognosefehlers ist ein zentraler Kritikpunkt an den Krigingverfahren: Die Variabilität des Merkmales an einer von untereinander sehr ähnlichen Messungen umgebenen Stelle wird durch die räumliche Abhängigkeit stärker eingeschränkt werden als die Variabilität des Merkmales an einer von untereinander sehr unterschiedlichen Messungen umgebenen Stelle. Um die Genauigkeit der Prognose abzubilden, sollte daher ein durch die Daten bedingter mittlerer quadratischer Prognosefehler an der „homogeneren“ Stelle kleiner sein als der bedingte mittlere quadratische Prognosefehler an der „inhomogeneren“ Stelle.

Die Krigingverfahren betrachten die Meßwerte und alle vorherzusagenden Werte als Ausprägung einer räumlich autokorrelierten multivariaten Zufallsfunktion; ein vorherzusagender Wert ist damit die unbeobachtete Ausprägung einer eindimensionalen Randverteilung. Die Kenntnis der durch die Meßwerte bedingten eindimensionalen Randverteilung, der bedingten lokalen Verteilungsfunktion, ermöglichte durch den bedingten Erwartungswert die optimale Prognose der unbeobachteten Ausprägung dieser Verteilung sowie die Bestimmung des bedingten mittleren quadratischen Prognosefehlers und die Ableitung exakter Prognoseintervalle für den unbekanntem Merkmalswert.

Diese für gewöhnlich unbekannte bedingte lokale Verteilungsfunktion kann auf verschiedene Arten approximiert werden.

Im einfachsten Fall können wir annehmen, daß der der Zufallsfunktion zugrundeliegende Prozeß gauß'sch ist, die bedingte lokale Verteilungsfunktion entspricht dann der Normalverteilungsfunktion einer Variablen, deren bedingter Erwartungswert die Einfache Krigingprognose und deren bedingte Varianz der bedingte mittlere quadratische Prognosefehler des Einfachen Krigingprädiktors ist. Dank der Annahme eines gauß'schen Prozesses ist dieser mit dem unbedingten mittleren quadratischen Prognosefehler des Einfachen Krigingprädiktors identisch. Da Einfaches Kriging die Kenntnis der normalerweise unbekanntem Erwartungswertfunktion des Prozesses voraussetzt, wird an seiner Statt oft Gewöhnliches Kriging durchgeführt. Die bedingte lokale Verteilungsfunktion wird dann als die Normalverteilungsfunktion einer Variablen approximiert, deren bedingter Erwartungswert gleich der Gewöhnlichen Krigingprognose und deren bedingte Varianz gleich dem unbedingten mittleren quadratischen Prognosefehler des Gewöhnlichen Krigingprädiktors ist. Es wird also neben der Näherung eines Krigingverfahrens durch ein anderes auch der bedingte durch den unbedingten mittleren quadratischen Prognosefehler ersetzt; das Verfahren wird Polfeldt-Ansatz genannt. Um die Güte der Approximation einschätzen zu können, wird im Rahmen dieser Arbeit für den beschriebenen Fall die Differenz zwischen unbedingtem und bedingtem Fehler des Gewöhnlichen Krigingprädiktors hergeleitet und analysiert.

In Kapitel 3.2 werden unter Annahme eines gauß'schen Zufallsprozesses die von Cressie (1991), S. 122 vorgeschlagenen Prognoseintervalle für den unbekanntem Merkmalswert mit jenen verglichen, die wir erhalten, wenn wir anstatt der unbedingten die bedingten mittleren quadratischen Prognosefehler des Gewöhnlichen Krigings nutzen. Es zeigt sich, daß die Differenzen zwischen diesen beiden Arten von Prognoseintervallen meist sehr gering sind und im Betrag die Differenzen aus den von Cressie (1991), S. 122 vorgeschlagenen und den exakten Prognoseintervallen selten überschreiten.

Nicht-lineare Krigingverfahren bieten alternative Approximationen der bedingten lokalen Verteilungsfunktion, dabei werden die Daten zu einem Satz von Schwellenwerten nicht-linear transformiert und Funktionswerte der bedingten lokalen Verteilungsfunktion für die Schwellenwerte vorhergesagt. Die nicht-linearen Verfahren arbeiten mit unterschiedlichen Annahmen über den transformierten Prozeß, im Allgemeinen nimmt die Vielschichtigkeit der Umsetzung eines Verfahrens mit der Stärke der Annahmen ab.

Die nicht-linearen Verfahren Median-Indikator-Kriging und Gauß'sches Disjunktives Kriging, die beide starke Annahmen an den zugrundeliegenden Prozeß stellen und daher in der Umsetzung praxistauglich sind, werden in zwei weiteren Fallstudien mit dem Polfeldt-Ansatz verglichen. In Kapitel 4.1 werden anhand von wahrscheinlich nicht aus einem gauß'schen Prozeß stammenden Daten der Immissionsökologischen Waldzustandserfassung in Baden-Württemberg die Prognosegüten der drei Verfahren mittels Kreuzvalidierung und die Approximationsgüten der bedingten lokalen Verteilungsfunktionen durch Überdeckungswahrscheinlichkeiten bewertet. Während der Polfeldt-Ansatz bei der Bewertung der Prognosegüten dem Median-Indikator-Kriging ebenbürtig ist und beide Verfahren hinter das gauß'sche Disjunktives Kriging zurückfallen, fällt die Bewertung der nicht-linearen Verfahren bei der Approxi-

---

mationsgüte deutlich besser aus.

Um dieses Ergebnis zu analysieren, werden die drei Verfahren in Kapitel 4.2 auf simulierte Daten angewandt. Mit dem zur Simulation genutzten Prozeß sind auch die bedingten lokalen Verteilungsfunktionen bekannt, es können daher sowohl einzelne Approximationen mit den bedingten lokalen Verteilungsfunktionen verglichen werden als auch zu verwendeten Schwellenwerten prognostizierte Funktionswerte der bedingten lokalen Verteilungsfunktionen kreuzvalidiert werden. Für einen gauß'schen Zufallsprozeß zeigt sich der Polfeldt-Ansatz erwartungsgemäß den nicht-linearen Verfahren klar überlegen, bei der lognormalen Transformation der Daten sollte die Wahl des Verfahrens von der Fragestellung abhängen: Median-Indikator-Kriging liefert die im Mittel am wenigsten verzerrten aber ungenauesten Prognosen, der Polfeldt-Ansatz approximiert bedingte lokale Verteilungsfunktionen mit extremem Erwartungswert am besten, während Gauß'sches Disjunktives Kriging den geringsten mittleren quadratischen Vorhersagefehler und den kleinsten Rang des absoluten Vorhersagefehlers aufweist.



## Danksagung

Ich danke

Prof. Dr. Joachim Saborowski vom Institut für Forstliche Biometrie und Informatik für die Betreuung und Anleitung,

Prof. Dr. Dr. h. c. Bransilav Sloboda und dem Institut für Forstliche Biometrie und Informatik für die dingliche und menschliche Infrastruktur,

Prof. Dr. Joachim Saborowski vom Institut für Forstliche Biometrie und Informatik, Prof. Dr. Christoph Kleinn vom Institut für Waldinventur und Waldwachstum und Prof. Dr. Martin Schlather vom Institut für Mathematische Stochastik für die Begutachtung,

Prof. Ing. Lubomir Scheer von der Technischen Universität Zvolen für das Ermöglichen eines Forschungsaufenthaltes,

Dr. Jaques Rivoirard vom Centre de Geostatistique de l'Ecole des Mines de Paris für die Zusendung seiner Dissertation (Rivoirard (1984)),

Dr. Klaus-Hermann von Wilpert von der Forstlichen Versuchsanstalt Baden-Württemberg, dafür, daß er die Daten der Immissionsökologischen Waldzustandserfassung zur Verfügung stellte,

Dr. Almuth Wameling für die Infektion mit dem Geostatistikvirus,

der Deutschen Forschungsgemeinschaft für die Finanzierung des Projektes „Bedingte Vorhersagefehler für Punkt- und Flächenvorhersagen bei Stichprobeninventuren im Wald“.



## A. Elektronischer Briefverkehr

### Zum Gewicht des Erwartungswertes beim Kriging

Sent: Tue 1/4/2005 3:02 PM  
To: ai-geostats@unil.ch  
Subject: [ai-geostats] the sum of the simple kriging weights

Hi all,  
the simple kriging weights do not sum to a constant.

From Olea (1999): 'Geostatistics for engineers and earth scientists', Kluwer, Boston, p.60: '(c) The possibility of having negative weights implies that the estimate is not confined to the data interval;...'

To me, this sounds like as if the sum of the simple kriging weights is restricted to some interval, especially since I've never seen any sum of weights not in  $[0,1]$ . Does anybody know any hint, evidence or proof of such a restriction?

Thanks a lot.  
Andreas Dominik Cullmann

---

Subject: Re: the sum of the simple kriging weights  
From: Isobel Clark <drisobelclark@yahoo.co.uk>  
Date: Tue, 4 Jan 2005 16:15:44 +0000 (GMT)  
To: Andreas Dominik Cullmann <acullma@gwdg.de>  
CC: AI Geostats mailing list <ai-geostats@unil.ch>  
Andreas

The simple kriging weights are not constrained in any way, either in sum or in sign.

Solving the simple kriging system gives a regression equation which 'best' predicts the value at the specified location. This is usually adjusted to be unbiased by applying the 'rest' ( $1 - \text{sum of weights}$ ) to the global average.

I have had cases where the simple kriging weights added up to 2.

Isobel  
<http://geocosse.bizland.com>

## A. Elektronischer Briefverkehr

---

*Subject: RE: [ai-geostats] the sum of the simple kriging weights*  
*From: "Colin Daly" <Colin.Daly@roxar.com>*  
*Date: Tue, 4 Jan 2005 16:51:03 -0000*  
*To: "Andreas Dominik Cullmann" <acullma@gwdg.de>, <ai-geostats@unil.ch>*  
*Andreas*

There is no restriction - although in most cases the values will lie between 0 and 1. In principle the weights can sum to anything - the value (1-sum\_of\_weights) is the weight associated to the global mean value. If all your data are very far from the point to be estimated - and the variogram range is short, then the weight on the mean will be close to 1 (so that the sum of weights on the data will be near zero)

With a 'quirky' variogram such as the gaussian, you can construct scenarios (perhaps using the screen effect) where the sum of weights is outside the [0,1] interval.

If you have a math background - then simple kriging is the unrestricted projection of the random variable at the point you are trying to estimate onto the linear space generated by the known data using the covariance as scalar product. (This would convince you that there are no restrictions on the weights)

Regards  
Colin Daly

---

### Zur Kovarianzfunktion eines lognormal Prozesses

*Subject: [ai-geostats] lognormal covariogram function*  
*From: Andreas Dominik Cullmann <acullma@gwdg.de>*  
*Date: Wed, 20 Jul 2005 18:41:02 +0200*  
*To: geostat <ai-geostats@unil.ch>*

Dear all,  
it's known how to derive the covariance values cov(h) of a lognormal process Z(s) from those of its corresponding gaussian process Y(s).  
Is it possible to derive the covariance function (i.e. form and parameters) of Z(s) from the one of Y(s)?  
[...]  
Thanks and regards,  
Dominik

---

*Subject: Re: [ai-geostats] lognormal covariogram function*  
*From: Denis Marcotte <denis.marcotte@polymtl.ca>*  
*Date: Wed, 20 Jul 2005 16:53:49 -0400*  
*To: Andreas Dominik Cullmann <acullma@gwdg.de>*

Dear Dominik,  
The covariance function of Z will be  
 $C_z(h) = (m_z)^2 (\exp(C_y(h)) - 1)$

---

I guess this is what you used to compute covariance values.  
If  $C_y$  was a spherical model for example than  $C_z$  is no longer a spherical model (and vice-versa). The logarithmic transformation does not preserve the model shape. In fact, there is no simpler expressions than the one above.

Denis Marcotte



## B. Umformungen

### B.1. Transponieren inverser Kovarianzmatrizen

Für die folgenden Berechnungen ist oft die Kenntnis der Transposition inverser Kovarianzmatrizen vonnöten. Da  $\Sigma$  symmetrisch ist ( $\Sigma = \Sigma^T$ ), ist nach Harville (1997), S. 82, (2.3), (2.4) auch  $\Sigma^{-1}$  symmetrisch. Nach Harville (1997), S. 5, (2.13) ist  $(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$ , für  $\mathbf{A} = \mathbf{1}^T$  und  $\mathbf{B} = \Sigma^{-1}$  gilt also  $(\mathbf{1}^T \Sigma^{-1})^T = \Sigma^{-1T} \mathbf{1}^{TT} = \Sigma^{-1} \mathbf{1}$ .

### B.2. Zu Kapitel 2.1.2

Die Varianz des Erwartungswertschätzers  $\hat{\mu}$  läßt folgendermaßen umformen:

$$\begin{aligned} \text{var} [\hat{\mu}] &= E [(\hat{\mu} - \mu)^2] = \text{var} [\mathbf{a}^T \mathbf{Z}] = \mathbf{a}^T \Sigma \mathbf{a} \\ &= \left( (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1} \mathbf{1} \Sigma^{-1} \right) \Sigma \left( (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1} \mathbf{1} \Sigma^{-1} \right)^T \\ &= (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1} \mathbf{1} \Sigma^{-1} \Sigma (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1} (\mathbf{1}^T \Sigma^{-1})^T \\ &= (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1} \mathbf{1} (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1} \Sigma^{-1} \mathbf{1} \\ &= (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1} (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1} \mathbf{1}^T \Sigma^{-1} \mathbf{1} \\ &= \frac{1}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}}. \end{aligned}$$

### B.3. Zu Kapitel 2.2.1

Nach Cressie (1991), S. 123 ist der Prädiktor des Gewöhnlichen Krigings

$$p_{ok}(\mathbf{s}_0, \mathbf{Z}) = \left( \mathbf{c} + \mathbf{1} \frac{1 - \mathbf{c}^T \Sigma^{-1} \mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \right)^T \Sigma^{-1} \mathbf{Z},$$

dies wird ausmultipliziert zu

$$\mathbf{c}^T \Sigma^{-1} \mathbf{Z} + \mathbf{1} (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1} \mathbf{1}^T \Sigma^{-1} \mathbf{Z} - \mathbf{c}^T \Sigma^{-1} \mathbf{1} (\mathbf{1}^T \Sigma^{-1} \mathbf{1})^{-1} \mathbf{1}^T \Sigma^{-1} \mathbf{Z}$$

und durch Einsetzen von 2.1.15

$$\mathbf{c}^T \Sigma^{-1} \mathbf{Z} + \mathbf{1} \hat{\mu} - \mathbf{c}^T \Sigma^{-1} \mathbf{1} \hat{\mu}.$$

### B.4. Zu Kapitel 2.2.2

Nach Cressie (1991), S. 123 ist der unbedingte mittlere quadratische Prognosefehler des Gewöhnlichen Krigings

$$\begin{aligned} \sigma_{ok}^2(\mathbf{s}_0) &= C(0) - \left( \mathbf{c} + \mathbf{1} \frac{1 - \mathbf{c}^T \Sigma^{-1} \mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \right)^T \Sigma^{-1} \mathbf{c} + \frac{1 - \mathbf{c}^T \Sigma^{-1} \mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \\ &= C(0) - \mathbf{c}^T \Sigma^{-1} \mathbf{c} + (-\mathbf{1}^T \Sigma^{-1} \mathbf{c} + \mathbf{1}) \frac{1 - \mathbf{c}^T \Sigma^{-1} \mathbf{1}}{\mathbf{1}^T \Sigma^{-1} \mathbf{1}} \end{aligned}$$

## B. Umformungen

---

Und da  $1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}$  ein Skalar ist, gilt  $1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} = (1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1})^T$  und somit

$$\sigma_{ok}^2(\mathbf{s}_0) = C(0) - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{c} + (1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1})^T (\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1})^{-1} (1 - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}).$$

### B.5. Zu Kapitel 3.1

Die Varianz des Einfachen Krigingprädiktors ist  $\text{var}[\mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z} - \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \mu(\mathbf{s}_0)]$ , wobei  $\mathbf{c}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\mu} + \mu(\mathbf{s}_0)$  bei gegebenem Prozeß und festgelegter Anordnung der  $N$  Meßstellen und des Prognoseortes eine konstante Zahl ist, als variabler Teil bleibt also  $\mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z}$ . Da nach Olea (1999), S. 10, Lemma 2.4 jede Linearkombination des Typs  $\boldsymbol{\lambda}^T \mathbf{Z}$  die Varianz  $\boldsymbol{\lambda}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\lambda}$  besitzt, gilt

$$\text{var}[\mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z}] = \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma} (\mathbf{c}^T \boldsymbol{\Sigma}^{-1})^T = \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma} \boldsymbol{\Sigma}^{-1} \mathbf{c} = \mathbf{c}^T \boldsymbol{\Sigma}^{-1} \mathbf{c}.$$

# Literaturverzeichnis

Once a map is drawn people tend to accept it as reality.

---

(Bert Friesen,  
nach Chilès u. Delfiner (1999),  
S. 150)

## Anderson 1958

Anderson, T. W.: *An Introduction to Multivariate Statistical Analysis*. New York : Wiley, 1958

## Chilès u. Delfiner 1999

Chilès, J.-P. ; Delfiner, P.: *Geostatistics – modelling spatial uncertainty*. New York : Wiley, 1999

## Cressie 1991

Cressie, N. A. C.: *Statistics for spatial data*. New York : Wiley, 1991

## Danneberg 2001

Danneberg, : *Wald(boden)sanierung*. Wien: Forstliche Bundesversuchsanstalt, 2001

## Deutsch u. Journel 1998

Deutsch, C. V. ; Journel, André G.: *GSLIB: Geostatistical software library and user's guide*. New York : Oxford University Press, 1998

## Evers u. Hüttl 1992

Evers, F.H. ; Hüttl, R.F.: *Magnesium-, Calcium- und Kaliummangel bei Waldbäumen – Ursachen, Symptome, Behebung*. Freiburg: Forstliche Versuchs- und Forschungsanstalt Baden-Württemberg, 1992

## Goovaerts 1997

Goovaerts, P.: *Geostatistics for natural resources evaluation*. New York : Oxford University Press, 1997

## Harville 1985

Harville, D. A.: Decomposition of Prediction Error. In: *Journal of the American Statistical Association* 80 (1985), März, Nr. 389, S. 132–138

## Harville 1997

Harville, D. A.: *Matrix Algebra from a Statistician's Perspective*. New York : Springer, 1997

## Harville u. Jeske 1992

Harville, D. A. ; Jeske, D. R.: Mean Squared Error of Estimation or Prediction Under a General Linear Model. In: *Journal of the American Statistical Association* 87 (1992), September, Nr. 419, S. 724–731

## Hoeting u. a. 2006

Hoeting, Jennifer A. ; Dobra, Richard A. ; Minka, Andrew A. ; Tost, Sandra E.: Model Selection for Geostatistical Models. In: *Ecological Applications* 16 (2006), Nr. 1, S. 87–98

## Johnson u. Kotz 1970

Johnson, N. L. ; Kotz, S.: *Distributions in Statistics: Continuous Univariate Distributions – 1*. New York : Wiley, 1970

**Johnson u. Kotz 1972**

Johnson, N. L.; Kotz, S.: *Distributions in Statistics: Continuous Multivariate Distributions*. New York : Wiley, 1972

**Journel 1983**

Journel, André G.: Nonparametric estimation of spatial distributions. In: *Mathematical Geology* 18 (1983), S. 119–140

**Journel u. Huijbregts 1978**

Journel, André G.; Huijbregts, C. J.: *Mining geostatistics*. London : Academic Press, 1978

**Liu u. Huettl 1991**

Liu, J. C.; Huettl, R. F.: Relations between damage symptoms and nutritional status of Norway spruce stands (*Picea abies* Karst.) in southwestern Germany. In: *Nutrient Cycling in Agroecosystems* 27 (1991), jan, Nr. 1, S. 9–22

**Matheron 1989**

Matheron, G.: *Estimating and choosing : an essay on probability in practice*. Berlin : Springer, 1989

**Meining u. von Wilpert 2006**

Meining, S.; von Wilpert, K. von: Waldzustandsbericht 2006 der Forstlichen Versuchs- und Forschungsanstalt Baden-Württemberg. (2006)

**Mood u. a. 1974**

Mood, A. M.; Graybill, F. A.; Boes, D. C.: *Introduction to the theory of statistics*. Auckland : McGraw-Hill, 1974

**Olea 1999**

Olea, R. A.: *Geostatistics for engineers and earth scientists*. Boston : Kluwer, 1999

**Oz u. a. 2003**

Oz, Bora; Durrleman, Clayton V.; Thomas, T.; Xu, YuLong: DSSIM-HR: A FORTRAN 90 program for direct sequential simulation with histogram reproduction. In: *Computers & Geosciences* 29 (2003), feb, Nr. 1, S. 39–51

**Papritz u. Moyeed 1999**

Papritz, A.; Moyeed, R. A.: Linear and Non-linear Kriging Methods: Tools for Monitoring Soil Pollution. In: Borchers, V. (Hrsg.); Thomas, K. (Hrsg.); Sill, A. (Hrsg.): *Statistics for the Environment 4: Statistical Aspects of Health and the Environment*. Wiley, 1999, S. 303–336

**Pardo-Igúzquiza u. Dowd 2004**

Pardo-Igúzquiza, E.; Dowd, P. A.: Normality Tests for Spatially Correlated Data. In: *Mathematical Geology* 36 (2004), August, Nr. 6, S. 659–681

**Polfeldt 1999**

Polfeldt, T.: On the quality of contour maps. In: *Environmetrics* 10 (1999), S. 785–790

**Ripley 1981**

Ripley, B. D.: *Spatial Statistics*. New York : Wiley, 1981

**Rivoirard 1984**

Rivoirard, J.: *Le Comportement des Poids de Krigeage*. Fontainebleu, Ecole des Mines de Paris, Thèse de Docteur-Ingénieur en Sciences et Techniques Minières, Option Géostatistique, 1984

**Rivoirard 1994**

Rivoirard, J.: *Introduction to disjunctive kriging and non-linear geostatistics*. Oxford : Oxford University Press, 1994

**Searle 1971**

Searle, S.R.: *Linear Models*. New York : John Wiley & Sons, 1971

**Stein 1999**

Stein, M. L.: *Interpolation of spatial data: some theory for kriging*. New York : Springer, 1999

**Wameling 2002**

Wameling, A.: *Räumliche Genauigkeit geostatistischer Interpolationsverfahren in Forst- und Umweltwissenschaften*. Cuvillier, 2002

**Webster u. Oliver 2001**

Webster, R. ; Oliver, M. A.: *Geostatistics for environmental scientists*. Chichester : Wiley, 2001

**Zimmerman 2006**

Zimmerman, D. L.: Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. In: *Environmetrics* 17 (2006)

**Zimmerman u. Cressie 1992**

Zimmerman, D. L. ; Cressie, N. A. C.: Mean squared prediction error in the spatial linear model with estimated covariance parameters. In: *Annals of the Institute of Statistical Mathematics* 44 (1992), mar, Nr. 1, S. 27–43



# Lebenslauf

Andreas Dominik Cullmann, geboren am 17.06.1975 in Freiburg im Breisgau

## Studium

Seit 2003/02	Doktorand am Institut für forstliche Biometrie und Informatik der Universität Göttingen.
2002/08	Master of Science mit Studienschwerpunkt Waldökosystemanalyse und Informationsverarbeitung
2000/02	Bachelor of Science
1997/10 – 2002/08	Studium der Forstwissenschaften an der Universität Göttingen.
1996/10 – 1997/08	Studium der Geoökologie an der Universität Potsdam.

## Schulabschluß

1995/06	Abitur am Faust-Gymnasium Staufen.
---------	------------------------------------

## Studienbegleitende und berufliche Tätigkeiten

Seit 2002/11	Wissenschaftlicher Angestellter am Institut für forstliche Biometrie und Informatik der Universität Göttingen.
2002/10	Wissenschaftliche Hilfskraft am Institut für forstliche Biometrie und Informatik der Universität Göttingen.
1999/05 – 2002/10	Studentische Hilfskraft am Institut für forstliche Biometrie und Informatik der Universität Göttingen.
1999/05 – 2002/04	Studentischer Vertreter im Prüfungsausschuß der Fakultät für Forstwissenschaften und Waldökologie der Universität Göttingen.
1998/04 – 2001/02	Studentischer Vertreter in der Studienkommission der Fakultät für Forstwissenschaften und Waldökologie der Universität Göttingen.
1998/04 – 2000/03	Finanzreferent der Fachschaft der Fakultät für Forstwissenschaften und Waldökologie der Universität Göttingen.

## Praktika

2002/05 – 2002/08	Furlong Forest Surveys, Wales.
1998/08 – 1998/09	Staatliches Forstamt Bovenden.
1998/02 – 1998/03	Staatliches Forstamt Bovenden.