

Mathematische Geologie II

Statistik regionalisierter Variablen

Geostatistik

H. Schaeben und S. Lindner
Mathematische Geologie und Geoinformatik
Institut für Geologie
TU Bergakademie Freiberg

SS 2000

Stand 10. April 2000

Contents

1	Vorwort	1
2	Einleitung	5
3	Interpolation, Approximation, Schätzen	7
3.1	1D Methoden	8
3.1.1	Stückweise konstanter Interpolant	8
3.1.2	Stückweise linearer Interpolant, lineare Interpolation	8
3.1.3	Polynomiale Interpolation, Lagrange's Interpolationspolynom	9
3.1.4	Polynomiale Approximation	10
3.1.5	Stückweise Polynome, polynomiale Splines	11
3.1.6	Radiale Basisfunktionen	11
3.1.7	Allgemeines zu 1D Methoden	12
3.2	2D Methoden	12
3.2.1	Stückweise konstanter Interpolant, Dirichlet - Thiessen - Voronoi Zer- legung	13
3.2.2	Lineare Interpolation über Dreiecken, Delaunay Triangulation	13
3.2.3	α -shapes	18
3.2.4	2D Polynome, Trendflächen-Analyse	18
3.2.5	Tensor-Produkt von polynomialem Splines	19
3.2.6	Radiale Basisfunktionen	19
3.3	3D, nD Methoden	19
3.4	Allgemeines zu deterministischen Methoden	19
4	Erkundende Daten Analyse	20
4.1	Der JURA-Datensatz	20
4.2	Häufigkeitstabelle kategorialer Zufallsvariablen	20
4.3	Kontingenztafel (cross table)	21
4.4	Histogramm und empirische Verteilungsfunktion	21
4.5	Elementare Statistiken (summary statistics)	21
4.6	Umgang mit extremen Werten, Datentransformationen	21

4.7	Bedingte Verteilungen (conditional distributions)	21
4.8	Streudiagramme (scatterplots)	22
4.9	Lineare und Rang-Korrelation	22
4.10	Karten georeferenzierter Daten	22
4.11	\mathbf{h} -Streudiagramm (\mathbf{h} -scattergram)	22
4.12	Maße für räumliche Variabilität und Stetigkeit	23
4.13	Indikator Transformation	24
4.14	Kreuz \mathbf{h} -Streudiagramm (cross \mathbf{h} -scattergram)	26
4.15	Maße für räumliche Kreuz-Variabilität und Kreuz-Stetigkeit	26
5	Theorie der Regionalisierten Variablen	28
5.1	Einführung	28
5.1.1	Bergbauliche Motivation 1 (“change of support”)	29
5.1.2	Bergbauliche Motivation 2 (“spatial continuity”)	29
5.1.3	Unzulänglichkeiten der klassischen Statistik - Modellannahme der unabhängig identisch verteilten Zufallsvariablen . .	30
5.1.4	Begriffe der klassischen Statistik, Elementare Regeln für das Rechnen mit Erwartungswerten und Varianzen	31
5.2	Zufallsfunktion, regionalisierte Zufallsvariablen, stochastischer Prozeß	32
5.2.1	Begriffe aus der Theorie der zufälligen Funktionen	33
5.2.2	Statistik zufälliger Funktionen	34
5.2.3	Modellannahme der strengen Stationarität	35
5.2.4	Modellannahme der schwachen Stationarität, Stationarität zweiter Ordnung	36
5.2.5	Modellannahme der intrinsischen Stationarität, Stationarität zweiter Ordnung der Inkremente	37
5.2.6	Diskussion der Stationaritätsbegriffe	39
5.2.7	Zusammenhang zwischen Kovarianzfunktion und Semivariogramm	39
5.2.8	Elementare Regeln für das Rechnen mit Erwartungswert- und Varianz- funktion	40
6	Der JURA–Datensatz in der Zusammenschau	42
7	Schließende Daten Analyse und Modellierung	45
7.1	Probenahme	45
7.2	Schätzer für Semivariogramm und Kovarianzfunktion	46
7.3	Modelle für Semivariogramm und Kovarianzfunktionen	46
7.4	Variographie	52

8 Kriging	56
8.1 Einfaches (simple) Punkt–Kriging	58
8.2 Duale Form des einfachen Punkt–Kriging	62
8.3 Gewöhnliches (ordinary) Punkt–Kriging	63
8.4 Einfaches vs gewöhnliches Punktkriging	67
8.5 Punktkriging mit Trend – universelles (universal) Punktkriging	68
8.6 k -IRF (Fontainebleau)	69
8.7 Weitere Kriging–Varianten	74
8.8 Zusammenfassung: Punktkriging	74
8.8.1 Kriging–Gewichte	74
8.8.2 Schätz–Nachbarschaft	77
8.8.3 Kriging–Varianz	77
8.8.4 Vergleich Meßwerte – geschätzte Werte	78
8.9 Tips für S–PLUS SPATIAL (5.Übung)	78
8.10 Blockkriging – Lagerstättenvorratsberechnung	82
8.10.1 Geostatistische Varianzen	82
8.10.2 Schätzprobleme und Schätzvarianzen	85
9 Modellierung lokaler Ungewißheit durch Indikator Kriging	87
9.1 Ziele der Modellierung der lokalen Ungewißheit	87
9.2 Lokale Verteilungen	89
9.3 Indikator Zugang – Nichtparametrische Geostatistik	90
9.3.1 Richtlinien zur Wahl der thresholds z_k , $k = 1, \dots, K$	90
9.3.2 Indikatorkriging	92
9.4 Optimale Schätzer	93
10 Odds and Ends	
Statt einer Zusammenfassung: Unvergeßliche Hinweise	96
11 Last words, final truths	98
12 Übungsaufgaben und S–PLUS script files	99
12.1 The JURA data set provided by J.-P. Dubois, IATE-Pédologie, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland	99
12.2 Aufgaben der ersten Übung	102
12.2.1 S–PLUS script file für die Aufgaben der ersten Übung	102
12.3 Aufgaben der zweiten Übung	107
12.3.1 S–PLUS script file für die Aufgaben der zweiten Übung	108
12.4 Aufgaben der dritten Übung	114
12.4.1 S–PLUS script file für die Aufgaben der dritten Übung	115
12.5 Aufgaben der vierten Übung	117
12.5.1 S–PLUS script file für die Aufgaben der vierten Übung	118

12.6 Aufgaben der fünften Übung	123
12.6.1 S-PLUS script file für die Aufgaben der fünften Übung	124
12.7 Aufgaben der sechsten Übung	127
12.7.1 S-PLUS script file für die Aufgaben der sechsten Übung	128
12.8 Aufgaben der siebten Übung	133
13 Synopsis	136
14 Checkliste – kriging 1,2,3,...	141

Chapter 1

Vorwort

Diese Vorlesung kann nur dann Wissen vermitteln, wenn sie nachgearbeitet wird. Dazu dient dieses Skript und die an dieser Stelle angegebene Literatur. Spezielle und weiterführende Literatur wird in den einzelnen Kapiteln angegeben.

Bibliography

- [1] Agterberg, F.P., 1974, Geomathematics: Elsevier
- [2] Akin, H., Siemes, H., 1988, Praktische Geostatistik - Eine Einführung für den Bergbau und die Geowissenschaften: Springer
- [3] Armstrong, M., 1998, Linear Geostatistics: Springer
- [4] Armstrong, M. und Dowd, P.A., 1994, Geostatistical Simulations: Kluwer Academic Publishers
- [5] Baafi, E.Y., Schofield, N.A., (eds.), 1997, Geostatistics Wollongong '96, vol. 1, vol 2: Kluwer Academic Publishers
- [6] Bucher, B., 1993, Optimierung von Grundwasser-Meßnetzen mit Kriging-Verfahren: Freiburger Schriften zur Hydrologie
- [7] Carr, J.R., 1995, Numerical Analysis for the Geological Sciences: Prentice Hall
- [8] Chilès, J.-P., Delfiner, P., 1999, Geostatistics: Modeling Spatial Uncertainty: Wiley
- [9] Christakos, G., 1992, Random Field Models in Earth Sciences: Academic Press
- [10] Clark, I., 1979, Practical Geostatistics: Applied Science Publishers
- [11] **Cressie, N.A.C., 1993, Statistics for Spatial Data - Revised Edition: Wiley**
- [12] David, M., 1977, Geostatistical Ore Reserve Estimation: Elsevier
- [13] Davis, J.C., 1986, Statistics and Data Analysis in Geology , 2nd ed.: Wiley
- [14] Deutsch, C.V, Journel, A.G., 1998, Geostatistical Software Library and User's Guide, Second Edition: Oxford University Press
- [15] Dimitrakopoulos, R., (ed.), 1994, Geostatistics for the Next Century: Kluwer Academic Publishers
- [16] **Goovaerts, P., 1997, Geostatistics for Natural Resources Evaluation: Oxford University Press**

- [17] Hohn, M.E., 1988, Geostatistics and petroleum geology: Van Nostrand Reinhold
- [18] Isaaks, E.H., Srivastava, R.M., 1989, An Introduction to Applied Geostatistics: Oxford University Press
- [19] Journel, A.G., 1989, **Fundamentals of Geostatistics in Five Lessons: American Geophysical Union**
- [20] Journel, A.G., Huijbregts, C., 1978, **Mining Geostatistics: Academic Press**
- [21] Kitanidis, P.K., 1997, Introduction to Geostatistics: Applications in Hydrogeology: Cambridge University Press
- [22] Krige, D., 1978, Lognormal - de Wijsian Geostatistics for Ore Evaluation: South African Institute of Mining and Mineralogy
- [23] Marsily, G. de, 1986, Quantitative Hydrogeology: Academic Press
- [24] Matheron, G., 1971, **The theory of regionalized variables and its applications: Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau, no 5**
- [25] Matheron, G., 1989, Estimating and Choosing: Springer
- [26] Müller, W.G., 1998, **Collecting Spatial Data: Physica Verlag**
- [27] Myers, J., 1997, Geostatistical Error Management Quantifying Uncertainty For Environmental Sampling and Mapping: Van Nostrand Reinhold
- [28] Olea, R., 1991, Geostatistical Glossary and Multilingual Dictionary: Oxford University Press
- [29] Olea, R., 1999, Geostatistics for Engineers and Earth Scientists: Kluwer Academic Publishers
- [30] Rendu, J.M., 1978, An Introduction to Geostatistical Methods and Mineral Evaluation: South African Institute of Mining and Mineralogy
- [31] Royle, A., 1980, Geostatistics: McGraw-Hill
- [32] Schafmeister, M.-T., 1999, Geostatistik für die hydrogeologische Praxis: Springer
- [33] Soares, A., (eds), 1993, Geostatistics Troia '92, vol. 1, vol 2: Kluwer Academic Publishers
- [34] Stein, M.L., 1999, Interpolation of Spatial Data - Some Theory for Kriging: Springer
- [35] Wackernagel, H., 1998, Multivariate Geostatistics (2nd completely revised version): Springer

- [36] Yarus, J.M., Chambers, R.L., (eds.), 1994, Stochastic Modeling and Geostatistics: Principles, Methods, and Case Studies: AAPG Computer Applications in Geology 3

Angewandte Statistik mit S-PLUS

- [37] MathSoft, 1995, S-PLUS Guide to Statistical and Mathematical Analysis, version 3.3: StatSci Division, MathSoft, Inc., Seattle
- [38] MathSoft, 1996, S+SpatialStats User's Manual, version 1.0: StatSci Division, MathSoft, Inc., Seattle
- [39] MathSoft, 1997, S-PLUS 4 Guide to Statistics: Data Analysis Products Division, MathSoft, Inc., Seattle
- [40] Venables, W.N. und Ripley, B.D., 1996, Modern Applied Statistics with S-PLUS: Springer
- [41] Krause, A., 1997, Einführung in S und S-PLUS: Springer
- [42] Böker, F., 1997, S-PLUS Learning by Doing: Lucius & Lucius
- [43] Everitt, B.S., 1994, A Handbook of Statistical Analyses using S-PLUS: Chapman & Hall
- [44] Schaffner, A. und Krahne, T., 1998, S-PLUS 4.0 unter Windows - Einführung und Leitfaden: MathSoft International Thomson Publishing

Chapter 2

Einleitung

Geostatistik wird hier als Kurzformel für die Anwendung stochastischer Prozesse und bedingter Wahrscheinlichkeiten in den Geowissenschaften, insbesondere zur Charakterisierung und Schätzung (Vorhersage) von georeferenzierten Daten, zum Beispiel Schwermetallkonzentrationen von Bodenproben oder Erzgehalte von Probekörpern aus einer vermuteten Lagerstätte, verwendet. Die Grundzüge dieser Methode werden entwickelt; dabei wird besondere Aufmerksamkeit auf das Wechselspiel zwischen naturwissenschaftlich-ingenieurmäßiger Praxis und der Aneignung mathematischer Konzepte zur Entwicklung von tragfähigen Modellen gelenkt.

Die Geologie entwickelte sich zunächst als beschreibende, empirische Wissenschaft. Durch die Anwendung neuer Methoden, die anderen Naturwissenschaften wie Physik, Chemie, Biologie etc. entliehen wurden, wurde die analytisch orientierte Entwicklung in Gang gesetzt. Diese Methoden sind heute im Gebäude der Geologie/Geowissenschaften fest etabliert.

Die Komplexität geowissenschaftlicher Zusammenhänge und der Grad der Idealisierung und Abstraktion der klassischen Mathematik verzögerte lange Zeit dauerhafte Kontakte zwischen beiden Wissenschaften. Erst mit der Verfügbarkeit von leistungsfähigen Rechnern, deren Entwicklung ihrerseits bestimmte Gebiete der Mathematik rasant vorantrieb, wurden mathematische Modelle in den Geowissenschaften sinnvoll anwendbar. Heute erlaubt die Verbreitung und Auswirkung der Anwendung mathematischer Modelle in den Geowissenschaften von einer Mathematisierung zu sprechen. Dieser Prozeß ist in anderen Naturwissenschaften, zum Beispiel Biologie, soweit fortgeschritten, daß "Biomathematik" als eigenständiges Fach schon aufgehört hat zu existieren und integraler Bestandteil verschiedener Richtungen innerhalb der Biologie geworden ist.

Geostatistik wendet man dann an, wenn die zu untersuchenden Phänomene so komplex sind, daß sie mit den Mitteln der klassischen Statistik nicht so erfaßt werden können, dass ihre Anwendung brauchbare Ergebnisse liefert. Geostatistik ist eine anwendungsorientierte Methode für den Naturwissenschaftler oder Ingenieur; sie ist eher eine Methode der mathematisch eingeweihten Modellierer als der Mathematiker. Ihre praktischen Anfänge gehen auf D. Krige im südafrikanischen Goldbergbau Anfang der 50er Jahre, ihre Theorie auf G. Matheron, Fontainebleau, France, in den 60er Jahren zurück. Heute hat sich neben Fontainebleau

eine weiteres Zentrum um den Matheron Schüler A. Journel in Stanford, CA, USA, gebildet.

Die Namensgebung legt nahe, die Ideen der Geostatistik mit Begriffen der Stochastik – Wahrscheinlichkeit und Statistik – zu entwickeln. Diesem Zugang folgt dieses Skript. An dieser Stelle sei aber betont, daß ein deterministisch orientierten Zugang existiert, in dem zum Beispiel das *experimentelle Semivariogramm* nicht als Schätzer eines Erwartungswert irgendeiner Zufallsfunktion sondern als diskretisierte Approximation eines wohldefinierten räumlichen Integral und *kriging* als mehrdimensionale Interpolation mit Hilfe radiauer Basisfunktion dargestellt werden kann. Um den Zusammenhang zwischen deterministisch und stochastisch motivierten Methoden, ihre Ähnlichkeiten und Unterschiede erfassen zu können, ist der Entwicklung der Geostatistik ein Kapitel vorangestellt, daß sich Aspekten der Interpolation und Approximation widmet.

Chapter 3

Interpolation, Approximation, Schätzen

Die Grundaufgabe besteht im folgenden Problem: Gegeben seien n verschiedene Punkte $(\mathbf{x}_i, f_i) \in D \subset \mathbb{R}^{p+1}$, $i = 1, \dots, n$. $f_i \in \mathbb{R}$ wird als $f(\mathbf{x}_i)$, $\mathbf{x}_i \in \mathbb{R}^p$, $i = 1, \dots, n$, interpretiert; also sind die \mathbf{x}_i die unabhängigen Variablen, die weiterhin als paarweise verschieden angenommen werden. $f_i = f(\mathbf{x}_i) \in \mathbb{R}$, $i = 1, \dots, n$, ist die abhängige Variable. Gesucht ist eine Vorschrift \hat{f} so, daß jedem $\mathbf{x} \in D \subset \mathbb{R}^p$ ein eindeutiger Wert $\hat{f}(\mathbf{x})$ zugeordnet wird und daß entweder

$$\text{Interpolation: } \hat{f}(\mathbf{x}_i) = f(\mathbf{x}_i), i = 1, \dots, n$$

oder

$$\text{Approximation: } \hat{f}(\mathbf{x}_i) \sim f(\mathbf{x}_i), i = 1, \dots, n$$

ist, wobei \sim noch spezifiziert werden muß. Die Wahl bzw. Konstruktion einer Methode zur Definition von \hat{f} wird wesentlich von den Eigenschaften abhängen, mit denen man sich die Funktion f ausgestattet vorstellt.

Konzeptionell kann man zwei Bestandteile der Methoden zur uni- und multivariaten Interpolation, Approximation, Schätzung unterscheiden

- die Nachbarschaftsbeziehung zwischen den Orten der Datenpunkte,
- die Zuordnung von Gewichten zu ausgewählten Nachbelpunkten,

sodaß die Methoden die Form

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^n w_i(\mathbf{x}) f(\mathbf{x}_i)$$

haben.

Im folgenden werden einige deterministische Methoden entsprechend dieser beiden Bestandteile dargestellt. Insbesondere sind multivariate Methoden von geowissenschaftlichem Interesse, bzw. solche univariaten Methoden, die eine Verallgemeinerung für den multivariaten Fall besitzen.

3.1 1D Methoden

In diesem Falle ist $p = 1$. Von Interesse in dieser Rückschau sind Methoden, die auf mehrdimensionale Verallgemeinerungen angelegt sind.

3.1.1 Stückweise konstanter Interpolant

Die an den Orten $x_i \in D$ beobachteten Werte $f_i = f(x_i)$ werden so über die Definitionsmenge D „verschmiert“, daß der Teilmenge von D , die näher an x_i ist als zu irgendeinem anderen Ort, der Wert f_i zugeordnet wird.

Die paarweise verschiedenen Ortskoordinaten $x_i \in D \subset \mathbb{R}$ seien der Größe nach geordnet $x_1 < x_2 < \dots < x_n$. An dieser Stelle sei angemerkt, daß eine Relation $<$ nur für den 1D Fall existiert! Im 1D Fall macht es Sinn, die Definitionsmenge als $D = [x_1, x_n] \subset \mathbb{R}^1$ zu definieren. Mit

$$v_i = x_i + \frac{x_{i-1} - x_i}{2} = \frac{x_i + x_{i+1}}{2}, \quad i = 1, \dots, n-1$$

definiert man eine Zerlegung von D in Teilmengen (Intervalle)

$$\begin{aligned} D_1 &= [x_1, v_1] \\ D_i &= (v_i, v_{i+1}] \\ D_n &= (v_{n-1}, x_n] \end{aligned}$$

und setzt

$$\hat{f}(x) = \sum_{i=1}^n f_i 1_{D_i}(x)$$

Die Teilmengen $D_i \subset \mathbb{R}^1$ lassen sich als die Mengen charakterisieren, für deren Elemente $x \in D_i$ der Abstand von x_i nicht größer als von irgendeinem anderen x_j ist.

Offensichtlich kann man diese Methode konzeptionell leicht für den mehrdimensionalen Fall verallgemeinern; unter dem Begriff Dirichlet - Thiessen - Voronoi Zerlegung (Parketierung) wird die zweidimensionale Verallgemeinerung im Zusammenhang mit der Delaunay Triangulation behandelt werden.

3.1.2 Stückweise linearer Interpolant, lineare Interpolation

Die Punkte $(x_i, f_i) \in \mathbb{R}^2$ werden durch ein Polygon, das heißt durch eine stückweise lineare Funktion mit „Knicken“ in den x_i , verbunden.

Variante (a): Jedem x_i , $i = 2, \dots, n-1$, wird eine Funktion w_i definiert auf $[x_{i-1}, x_{i+1}]$ durch

$$w_i(x) = \begin{cases} \frac{x - x_{i-1}}{x_i - x_{i-1}}, & x \in [x_{i-1}, x_i] \\ \frac{x_{i+1} - x}{x_{i+1} - x_i}, & x \in [x_i, x_{i+1}] \end{cases}$$

und x_1, x_n wird

$$\begin{aligned} w_1(x) &= \frac{x_2 - x}{x_2 - x_1}, \quad x \in [x_1, x_2] \\ w_n(x) &= \frac{x - x_{n-1}}{x_n - x_{n-1}}, \quad x \in [x_{n-1}, x_n] \end{aligned}$$

zugeordnet, und

$$\hat{f}(x) = \sum_{i=1}^n w_i(x) f_i$$

gesetzt.

Variante (b): Jedem Intervall $I_i = [x_i, x_{i+1}]$, $i = 1, \dots, n-1$, werden zwei Funktionen

$$\begin{aligned} F_i^{(l)}(x) &= \frac{x_{i+1} - x}{x_{i+1} - x_i} \\ F_i^{(r)}(x) &= \frac{x - x_i}{x_{i+1} - x_i} \end{aligned}$$

und dann

$$w_i(x) = F_i^{(l)}(x) + F_i^{(r)}(x)$$

zugeordnet, und

$$\hat{f}(x) = \sum_{i=1}^{n-1} w_i(x) f_i$$

gesetzt.

Man beachte, daß die Terme $\frac{x_{i+1}-x}{x_{i+1}-x_i}$, $\frac{x-x_i}{x_{i+1}-x_i}$ die baryzentrische Koordinaten des Punktes $x \in [x_i, x_{i+1}]$ darstellen.

3.1.3 Polynomiale Interpolation, Lagrange's Interpolationspolynom

Häufig werden an einen Interpolanten höhere Anforderungen bezüglich seiner Glattheit gestellt, als die stückweise konstanten (unstetigen) oder stückweise linearen (mit unstetiger erster Ableitung) erfüllen. Eine Lösung dieses Problems bietet die polynomiale Interpolation: zu n Punkten $(x_i, f_i) \in \mathbb{R}^2$ gibt es genau ein Polynom p vom Grade $n-1$ (Ordnung n) so, daß $p(x_i) = f_i$ gilt. Zu zwei Punkten ($n=2$) gibt es genau eine Gerade, zu drei Punkten ($n=3$) gibt es genau eine Parabel, zu vier Punkten ($n=4$) gibt es genau eine kubische Kurve, allgemein

$$\begin{aligned} p(x) &= \sum_{i=1}^n f_i L_i^{(n-1)}(x) \\ L_i^{(n-1)}(x) &= \frac{\prod_{j \neq i} (x - x_j)}{\prod_{j \neq i} (x_i - x_j)} = \prod_{j \neq i} \frac{(x - x_j)}{(x_i - x_j)} \end{aligned}$$

mit

$$\sum_{i=1}^n L_i^{(n)}(x) = 1$$

$$L_i^{(n)}(x_j) = \delta_{ij} = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases}$$

Dies ist das Lagrange Interpolationspolynom; es vermittelt die einzige Darstellung des Interpolationspolynoms in kardinaler Form, in der die Daten explizit erscheinen.

Es gilt zu beachten, daß der Grad (die Ordnung) des Polynoms direkt von der Anzahl der zu interpolierenden Punkte abhängt. Für große n neigt das Interpolationspolynom i.a. zu sehr starken Oszillationen, die seinen Wert für praktische Anwendungen sehr stark reduzieren; für $n > 6$ ist die Lagrange Interpolation i.a. praktisch unbrauchbar.

3.1.4 Polynomiale Approximation

Eine Möglichkeit, der unangenehmen Eigenschaft von Polynomen, für grosse Polynomgrade stark zu oszillieren, zu begegnen, besteht darin, die Idee der linearen Regression aus der bivariaten Statistik aufzugreifen, und ein Polynom mit festem Grad N

$$p(x) = \sum_{l=0}^N \alpha_l x^l$$

so an die Punkte (x_i, f_i) , $i = 1, \dots, n$, anzupassen, daß

$$\sum_{i=1}^n (f(x_i) - p(x_i))^2 \stackrel{!}{=} \min$$

Für $n \geq N$ ist dieses Problem eindeutig lösbar, die Lösung ist durch das entsprechende Normalen Gleichungssystem gegeben.

Für $N = 1$ sind die Rechenschritte formal identisch mit der Bestimmung einer Regression Geraden ohne daß hier ein expliziter Bezug zu irgendwelchen Verteilungsannahmen hergestellt wird. Erst durch Berücksichtigung der Modellannahme

$$f(x) = \hat{f}(x) + \varepsilon(x) = \alpha_0 + \alpha_1 x^1 + \varepsilon(x)$$

mit einem normalverteilten Term $\varepsilon(x)$ wird aus der Anpassung einer Geraden an eine Punktwolke das statistische Modell der linearen Regression.

3.1.5 Stückweise Polynome, polynomiale Splines

Wegen der unangenehmen Eigenschaft von Polynomen, für große Polynomgrade stark zu oszillieren, scheint es für einen Interpolanten mit besseren Eigenschaften nötig, die Anzahl der zu interpolierenden Punkte und den polynomialen Grad des Interpolanten zu entkoppeln. Anknüpfend an stückweise konstante oder stückweise lineare Interpolanten in der Variante (b) kann man versuchen, Interpolanten zu konstruieren, die stückweise, das heißt zwischen zwei "Knoten" (die nicht notwendig mit den Orten der Daten identisch sein müssen), mit einem Polynom eines gegebenen kleinen Grads übereinstimmen, als Ganzes aber kein Polynom sind. Dies läßt sich dadurch erreichen, daß man die Glattheitsanforderungen in den Knoten verglichen mit einem Polynom zurücknimmt. Ein Spline zu den Knoten v_0, v_1, \dots, v_N vom Grad n ist eine Funktion, die zwischen zwei Knoten mit einem Polynom vom Grad n übereinstimmt, und deren Ableitungen bis einschließlich der Ordnung $(n-1)$ stetig sind. Dies bedeutet, daß die n -te Ableitung des Splines in den Knoten unstetig sein darf. Für den Rand der Definitionsmenge muß man noch weitere spezielle Bedingungen angeben.

Der stückweise konstante Interpolant ist also ein Spline vom Grad 0 zu den Knoten v_i ; der stückweise lineare Interpolant ist ein Spline vom Grad 1, dessen Knoten mit den Orten der Daten identisch sind. Ein kubischer Spline ist eine Funktion, die zwischen zwei Knoten mit einem kubischen Polynom übereinstimmt und in den Knoten zweimal stetig differenzierbar ist.

Ein Spline zu den Knoten v_0, \dots, v_N vom Grad n ist charakterisiert durch $N+2$ Polynome vom Grad n , das heißt mit $n+1$ zu spezifizierenden Koeffizienten und n Glattheitsbedingungen an $N+1$ Knoten. Er besitzt also $(n+1)(N+2) - n(N+1) = n + N + 2$ freie Parameter, die man zur Lösung einer Interpolationsaufgabe zur Verfügung hat.

3.1.6 Radiale Basisfunktionen

Anknüpfend an stückweise lineare Interpolanten in der Variante (a) kann man Interpolanten konstruieren, indem man die hüttchenförmigen Funktionen w_i durch glattere glockenförmige Funktionen ersetzt. Dazu steht eine Fülle von bekannten Funktionen zur Verfügung; eine kleine Auswahl ist

$$\text{surface} \quad \chi(r) = r^{2n-1}, \quad n \in \mathbb{N} \quad (3.1)$$

$$\text{thin plate} \quad \chi(r) = r^{2n} \log r, \quad n \in \mathbb{N} \quad (3.2)$$

$$\text{multiquadric} \quad \chi(r) = (r^2 + c^2)^{1/2}, \quad c \in \mathbb{R} \quad (3.3)$$

$$\text{inverse multiquadric} \quad \chi(r) = (r^2 + c^2)^{-1/2}, \quad c \in \mathbb{R} \quad (3.4)$$

$$\text{shifted surface} \quad \chi(r) = (r^2 + c^2)^{n-1/2}, \quad n \in \mathbb{N}, \quad c \in \mathbb{R} \quad (3.5)$$

$$\text{shifted thin plate} \quad \chi(r) = (r^2 + c^2)^n \log(r^2 + c^2), \quad n \in \mathbb{N}, \quad c \in \mathbb{R} \quad (3.6)$$

$$\text{shifted logarithm} \quad \chi(r) = \log(r^2 + c^2), \quad c \in \mathbb{R} \quad (3.7)$$

$$\text{Gaussian} \quad \chi(r) = \exp(-cr^2), \quad c \in \mathbb{R}_+ \quad (3.8)$$

wobei $\chi_p(\mathbf{x}) = \chi(r_p)$ mit $r_p = \|\mathbf{x} - \mathbf{x}_p\|$, $\mathbf{x}, \mathbf{x}_p \in \mathbb{R}^d$, $d \in \mathbb{N}$.

Der Interpolant hat wieder die Form

$$\hat{f}(x) = \sum_{i=1}^n w_i(x) f_i = \sum_{i=1}^n f(x_i) w(x - x_i)$$

Dieser Zugang lässt sich sehr leicht für den mehrdimensionalen und auch nicht-euklidischen Fall (z.B. für Daten auf der Kugel) verallgemeinern.

3.1.7 Allgemeines zu 1D Methoden

Die oben vorgestellten verschiedenen Herangehensweisen an das klassische 1D Interpolationssproblem haben den linearen Ansatz

$$\hat{f}(x) = \sum_{i=1}^n w_i(x) f_i$$

und die beiden konzeptionellen Bestandteile

- Nachbarschaftsrelation: Welche f_i werden zur Bestimmung von $f(x)$ berücksichtigt, das heißt, welche $w_i > 0$?
- Gewichtung: Wie werden die w_i bestimmt?

gemeinsam.

Der lineare Ansatz erlaubt zwei zueinander duale Lesarten, bzw. Interpretationen, einmal (naiv) als Linearkombination der Daten, das heißt als gewichtetes Mittel der Daten,

$$\hat{f}(x) = \sum_{i=1}^n \underbrace{w_i(x)}_{\text{Gewicht}} \underbrace{f_i}_{\text{Datum}}$$

und auch als gewichtete Linearkombination von geeigneten Basisfunktionen

$$\hat{f}(x) = \sum_{i=1}^n \underbrace{w_i(x)}_{\text{Basisfunktionen}} \underbrace{f_i}_{\text{Gewicht}}$$

3.2 2D Methoden

In diesem Falle ist $p = 2$. Gegenstand sind mögliche Verallgemeinerungen der 1D Methoden. Die Orte $\mathbf{x}_i \in D \subset \mathbb{R}^2$ sind jetzt Punkte in der 2D Ebene und besitzen Koordinaten (x_i, y_i) .

3.2.1 Stückweise konstanter Interpolant, Dirichlet - Thiessen - Voronoi Zerlegung

In vollkommener Analogie zum 1D Fall ordnet man jedem Ort (x_i, y_i) seine Dirichlet Zelle zu, die definiert ist als die Teilmenge von $D_i \subset D \subset \mathbb{R}^2$, deren Elemente näher zum Ort (x_i, y_i) sind als zu irgendeinem anderen Ort. Die n Dirichlet Zellen bilden die Thiessen Zerlegung, Voronoi Parkettierung, etc. der Definitionsmenge D .

Seien D_i , $i = 1, \dots, n$, die Dirichlet Zellen, $D_i \cap D_j = \emptyset$, $i \neq j$, $\bigcup_{i=1}^n D_i = D \subset \mathbb{R}^2$ einer Datenmenge $((x_i, y_i), f_i)$. Ordnet man jedem Punkt $(x, y) \in D_i$ den Wert f_i zu, so erhält man den stückweise konstanten, an den Kanten der Dirichlet Zellen unstetigen Interpolanten

$$\hat{f}(x, y) = \sum_{i=1}^n f_i 1_{D_i}(x, y)$$

Die Orte (x_{i_j}, y_{i_j}) , $j = 1, \dots, J$, deren Dirichlet Zellen eine gemeinsame Kante mit der zu (x_i, y_i) gehörenden Dirichlet Zelle besitzen, werden auch die natürlichen oder starken Thiessen Nachbarn des Datenorts (x_i, y_i) genannt.

3.2.2 Lineare Interpolation über Dreiecken, Delaunay Triangulation

Dirichlet Zerlegung und Delaunay Triangulation sind duale mathematische Konstrukte; die Kenntnis des einen vermittelt direkt die Kenntnis des anderen.

Gegeben sei eine Punktmenge \mathcal{P} mit n paarweise verschiedenen nicht kollinearen Punkten $P_i = P(x_i, y_i)$, $i = 1, \dots, n$. Eine Triangulation \mathcal{T} der Punktmenge \mathcal{P} ist eine Menge von N Punktetripeln (P_i, P_j, P_k) so, daß

- jedes Tripel ein nicht-degeneriertes Dreieck definiert;
- jedes Dreieck keine anderen Punkte P_l enthält;
- der Durchschnitt des Inneren zweier Dreiecke leer ist;
- die Vereinigung aller Dreiecke die konvexe Hülle der Punkte P_i ist.

Eine Triangulation ist also eine Zerlegung der konvexen Hülle der Punkte P_i in Dreiecke.

Zu einer gegebenen Punktmenge existieren verschiedene Triangulierungen, jedoch gilt

$$\begin{aligned} N &= 2n - n_R - 2 \\ K &= 3n - n_R - 3 \end{aligned}$$

$$\begin{aligned} n - 2 &\leq N \leq 2n - 5 \\ 2n - 3 &\leq K \leq 3n - 6 \end{aligned}$$

wobei N die Anzahl aller Dreiecke, K die Anzahl aller Kanten, n_R die Anzahl der Punkte auf dem Rand der konvexen Hülle bezeichnet.

Eine Triangulation wird erst durch Berücksichtigung eines geeigneten Auswahlkriteriums eindeutig. Dabei zeichnet sich ein geeignetes Kriterium auch dadurch aus, daß es konstruktiv umsetzbar ist, das heißt, daß die Triangulation der Punkte entsprechend diesem Kriterium algorithmisch effektiv konstruiert werden kann. Ein Kriterium wird zunächst für die Triangulation eines strikt konvexen Vierecks formuliert. Mögliche Kriterien sind

- Triangulation gemäß der kürzeren Diagonale
- Triangulation gemäß der längeren Diagonale
- Triangulation gemäß der Maximierung der kleinsten Winkels (max-min Kriterium), Delaunay Triangulation
- Triangulation gemäß der Minimierung des größten Winkels (min-max Kriterium)
- Inkreis
- Umkreis
- Triangulation gemäß der Minimierung der Summe der Längen aller Kanten

Eine Triangulation heißt lokal optimal, wenn jedes strikt konvexe Viereck bezüglich des ausgewählten Kriteriums optimal trianguliert ist. Lokale Optimalität reicht i.a. nicht aus, um die Konstruktion einer Triangulation gemäß eines für Vierecke formulierten Kriteriums eindeutig zu machen. Zur Definition und operationellen Entscheidung über globale Optimalität müssen verschiedene lokal optimale Triangulationen nach dem gleichen Kriterium verglichen werden können.

Bezeichne $a(T)$, $T \in \mathcal{T}$, das Maß für die Form eines Dreiecks T der Triangulation \mathcal{T} , z.B. sei $a(T)$ der kleinste Innenwinkel des Dreiecks. Der Triangulation \mathcal{T} läßt sich dann der Vektor

$$a(\mathcal{T}) = (a_1, \dots, a_N), \quad a_i = a(T_i), \quad T_i \in \mathcal{T}, \quad i = 1, \dots, N$$

zuordnen, der als Komponenten die Werte $a(T_i)$, $T_i \in \mathcal{T}$ nach zunehmender Größe geordnet enthält, z.B. die Werte für die kleinsten Innenwinkel aller N Dreiecke nach zunehmender Größe geordnet. Dann heißt die Triangulation \mathcal{T}_1 bezüglich des Maßes a besser als die Triangulation \mathcal{T}_2 , falls $a(\mathcal{T}_1) > a(\mathcal{T}_2)$. Dabei bezeichnet $>$ hier die lexikographische Ordnung: für zwei Vektoren a, b bedeutet $a < b$ in diesem Fall, daß eine Zahl $m \leq N$ so existiert, daß $a_i = b_i$, $i = 1, \dots, m-1$ und $a_m < b_m$.

Die Triangulation \mathcal{T}^* zu einer Punktmenge \mathcal{P} ist bezüglich des Maßes a global optimal, falls $a(\mathcal{T}^*) \geq a(\mathcal{T})$ für alle Triangulationen \mathcal{T} von \mathcal{P} , das heißt, wenn ihr zugeordneter Vektor lexikographisch mindestens so groß ist wie alle übrigen. Da die Anzahl aller möglichen Triangulationen von \mathcal{P} endlich ist, ist diese Begriffsbildung sinnvoll.

Jede optimale Triangulation ist auch lokal optimal. Für die Triangulation nach dem max-min Kriterium gilt auch die Umkehrung, andere Kriterien mit dieser Eigenschaft sind nicht bekannt. Im allgemeinen gilt diese Umkehrung nicht.

Beispiel min-max Kriterium: Sei $a(T)$ der größte Innenwinkel des Dreiecks T , und sei $a(\mathcal{T}) = \max\{1/a(T), T \in \mathcal{T}\}$; falls für die beiden Triangulationen $\mathcal{T}_1, \mathcal{T}_2$ eines strikt konvexen Vierecks $\mathcal{T}_1 > \mathcal{T}_2$ gilt, dann ist \mathcal{T}_1 die bessere (richtige) Triangulation.

Dieses Beispiel zeigt, daß lokale Optimalität bezüglich des min-max Kriteriums nicht globale Optimalität (bezüglich des min-max Kriteriums) impliziert.

Die Triangulation \mathcal{T} einer Punktmenge \mathcal{P} , die dadurch definiert ist, daß alle starken Thiessen-Nachbarn durch Kanten von \mathcal{T} verbunden werden, ist durch besondere Eigenschaften ausgezeichnet. Geometrisch kann man sie sich entstanden denken, indem man in jedem Punkt der Punktwolke einen Kegel mit seiner Spitze aufstellt und die Kegelschnitte in die Ebene projiziert. Diese Triangulation ist lokal optimal bezüglich des max-min Kriteriums. Darüberhinaus ist diese Triangulation auch global optimal bezüglich des max-min Kriteriums. Diese Triangulation wird allgemein als Delaunay Triangulation bezeichnet. Sie ist dadurch ausgezeichnet, daß lokale Optimalität globale impliziert. Diese Eigenschaft kann man zur Konstruktion effektiver Algorithmen verwenden, die zunächst sukzessive strikt konvexe Vierecke richtig triangulieren und schließlich ohne weiteres zu einer global optimalen Delaunay Triangulation führen.

Man kann drei Grundtypen von Algorithmen zur Konstruktion der Delaunay Triangulation unterscheiden,

- Typ A: In einer vorgegebenen beliebigen Triangulation werden alle strikt konvexen Vierecke auf die Erfüllung des max-min Kriteriums (oder eines äquivalenten) überprüft; gegebenenfalls wird die Diagonale „umgeklappt“.
- Typ B: Startet mit drei Punkten, die ein Dreieck der Triangulation definieren. Dann werden sukzessiv Punkte hinzugefügt und die dabei entstehenden strikt konvexen Vierecke nach dem max-min Kriterium trianguliert. Die aktuelle Triangulation ist stets lokal optimal.
- Typ C: Startet mit einer groben Zerlegung von D in Teilmengen. Dann werden die Teilmengen lokal optimal bezüglich des max-min Kriteriums trianguliert. Schließlich werden die Triangulationen der Teilmengen so zusammengefügt, daß die lokale Optimalität erhalten bleibt.

Die Nachbarschaftsrelation einer 2D Methode zur Interpolation/Approximation auf der Grundlage einer Triangulation ist gerade durch die Triangulation selbst gegeben. Für lineare Interpolation werden die Gewichte der Nachbarn folgendermaßen bestimmt.

Zur Erhöhung der Effektivität werden lokale baryzentrische Koordinaten bezüglich jeden Dreiecks $T \in \mathcal{T}$ eingeführt. Die baryzentrischen Koordinaten (u, v, w) eines beliebigen Punk-

tes $P \in T = [P_{n_1}, P_{n_2}, P_{n_3}] \subset \mathbb{R}^2$ mit globalen Koordinaten (x, y) sind gegeben durch

$$\begin{aligned} u &= A_1/A \\ v &= A_2/A \\ w &= A_3/A \end{aligned}$$

mit

$$A = \left| \det \begin{pmatrix} x_{n_1} & x_{n_2} & x_{n_3} \\ y_{n_1} & y_{n_2} & y_{n_3} \\ 1 & 1 & 1 \end{pmatrix} \right|$$

$$A_1 = \left| \det \begin{pmatrix} x & x_{n_2} & x_{n_3} \\ y & y_{n_2} & y_{n_3} \\ 1 & 1 & 1 \end{pmatrix} \right|$$

$$A_2 = \left| \det \begin{pmatrix} x_{n_1} & x & x_{n_3} \\ y_{n_1} & y & y_{n_3} \\ 1 & 1 & 1 \end{pmatrix} \right|$$

$$A_3 = \left| \det \begin{pmatrix} x_{n_1} & x_{n_2} & x \\ y_{n_1} & y_{n_2} & y \\ 1 & 1 & 1 \end{pmatrix} \right|$$

Die baryzentrischen Koordinaten summieren sich zu 1.

Die baryzentrischen Koordinaten des Punktes $P = P_{n_1}(x_{n_1}, y_{n_1})$ sind $(1, 0, 0)$; der Punkt mit den baryzentrischen Koordinaten $(0, v, 1 - v)$ liegt auf der Kante P_{n_2}, P_{n_3} und teilt diese im Verhältnis $P_{n_2}, P : P, P_{n_3} = 1 - v : v$.

Die Berechnung dieser Determinanten erfolgt am günstigsten durch eine einfache Variante der Cramer'schen Regel, die z.B. für die Determinante A in

$$A = |(x_{n_1} - x_{n_2})(y_{n_2} - y_{n_3}) + (x_{n_2} - x_{n_3})(y_{n_2} - y_{n_1})|$$

besteht und demnach nur zwei Multiplikationen und vier Summationen erfordert. Durch geschickte (Um)Nummerierung der Dreieckseckpunkte kann man erreichen, daß alle Determinanten positiv sind und die Absolutzeichen nicht explizit berücksichtigt werden müssen.

Nun bezeichne $(x_{n_k}, y_{n_k}, f_{n_k})$, $k = 1, \dots, 3$, die Daten bezüglich der Orte $P_{n_k} \in \mathbb{R}^2$, $k = 1, \dots, 3$. Lineare Interpolation über dem Dreieck $T = [P_{n_1}, P_{n_2}, P_{n_3}]$ hat dann die einfache Form

$$\hat{f}(u, v, w) = u f_{n_1} + v f_{n_2} + w f_{n_3}$$

oder mit explizitem Bezug zu den globalen Koordinaten (x, y) des Punktes $P \in T \in \tilde{\mathcal{T}}$ mit lokalen baryzentrischen Koordinaten (u, v, w) bezüglich T

$$\hat{f}(x, y) = \hat{f}(u(x, y), v(x, y), w(x, y)) = u(x, y)f_{n_1} + v(x, y)f_{n_2} + w(x, y)f_{n_3}$$

Der so definierte Interpolant ist stückweise linear mit Unstetigkeiten der ersten Ableitung über den Dreieckskanten in allen Richtungen, die nicht mit der Dreieckskante übereinstimmen.

Im Zusammenhang mit linearer Interpolation über Dreiecken einer Triangulation ist die Delaunay Triangulation durch eine minimale Rauhigkeit des Interpolanten ausgezeichnet. Dieses Charakterisierung der Delaunay Triangulation stellt ihre Geometrie in einen Bezug zu Eigenschaften des über ihr definierten stückweise linearen Interpolanten und ist so nicht unmittelbar zu erwarten gewesen, da ja in ihren Konstruktionsprozeß die $f_i = f(\mathbf{x}_i)$ Werte gar nicht eingehen.

Das Rauhigkeitsmaß ist dabei die quadrierte L^2 Norm des Gradienten integriert über der triangulierten Menge.

Bei bestimmten Anwendungen kann eine übergeordnete "natürliche" Nachbarschaft der Daten, das heißt der Orte der beobachteten Werte, durch die Aufgabenstellung selbst bzw. durch zusätzliche Information, die nicht explizit in dem Datensatz (x_i, y_i, f_i) , $i = 1, \dots, n$, enthalten ist, existieren. Beispiele dafür sind digitalisierte Isolinien, wobei die Punkte auf einer Isolinie die natürlichen Nachbarn sind, oder Unstetigkeitslinien, wobei die Punkte auf jeweils einer Seite dieser Linie die jeweiligen natürlichen Nachbarn sind. In diesen Fällen ist es sinnvoll, diese natürlichen Nachbarschaftsverhältnisse zu bewahren und nicht durch irgendeine Triangulation zu ersetzen, das heißt zu zerstören. Oft kann man diese natürlichen Nachbarschaftsbeziehungen schon allein dadurch bewahren, daß man bestimmte Dreieckskanten von vornherein festsetzt, unabhängig vom Optimalitätskriterium und vom Algorithmus, das heißt, der Algorithmus darf diese Kanten nicht verändern, sondern muß sie bewahren. Im Falle der digitalisierten Isolinien werden unmittelbar benachbarte Punkte auf ein und derselben Isolinie durch solche "Zwangskanten" verbunden, im Falle der Unstetigkeitslinien werden unmittelbar benachbarte Punkte auf dieser Linie, bzw. Anfangs- und Endpunkt der Linie, durch Zwangskanten verbunden. In beiden Fällen erreicht man damit, daß nicht über die Zwangskanten hinweg sondern nur entlang dieser Zwangskanten interpoliert wird.

Dieses Konzept ist für die Delaunay Triangulation wegen ihrer ausgezeichneten Eigenschaft, daß lokale globale Optimalität impliziert, besonders einfach zu realisieren. Die sich unter Berücksichtigung von vorgegebenen Zwangskanten ergebende Triangulation ist überall bis auf durch die Zwangskanten bestimmte Teilmengen lokal optimal. Deshalb heißt diese Triangulation die bedingte Delaunay Triangulation.

Literaturhinweise

Okabe, A., Boots, B. & Sugihara, K., 1992, Spatial Tessellations: J. Wiley & Sons

Auerbach, S. & Schaeben, H., 1990, Surface representations reproducing given digitized contour lines: Math. Geol. 22, 723-742

Auerbach, S. & Schaeben, H., 1990, Computer aided geometric design of geologic surfaces and bodies: *Math. Geol.* 22, 957-987

Auerbach, S., Gmelig Meyling, R.H.J., Neamtu, M. & Schaeben, H., 1991, Approximation and geometric modeling with simplex b-splines associated with irregular triangles: *CAGD* 8, 67-87

Schaeben, H., Auerbach, S. & Schütz, E., 1996, Volumetrics and rendering of geologic bodies by three-dimensional geometric reconstruction from cross sections or contourlines: in Förster, A. & Merriam, D.F., (eds.), *Geologic Modeling and Mapping*, Plenum Publishing Corporation, 135-152

Schaeben, H. & Asmus, K., 1997, Recovering surfaces from their contourlines: in Pawłowski-Glahn, V., (ed.), *Proceedings International Association for Mathematical Geology*, CIMNE Barcelona, 355-360

3.2.3 α -shapes

<http://alpha.ncsa.uiuc.edu/alpha/>

3.2.4 2D Polynome, Trendflächen-Analyse

Die Idee der linearen Regression läßt sich für den 2D Fall dadurch verallgemeinern, daß man die univariaten durch bivariate Polynome von festem Polynomgrad N ersetzt.

$$p(x, y) = \sum_{\substack{k, l=0 \\ k+l \leq N}}^N \alpha_{k,l} x^k y^l$$

so an die Punkte (x_i, y_i, f_i) , $i = 1, \dots, n$, anzupassen, daß

$$\sum_{i=1}^n (f(x_i, y_i) - p(x_i, y_i))^2 \stackrel{!}{=} \min$$

Für $n \geq N$ ist dieses Problem eindeutig lösbar, die Lösung ist durch das entsprechende Normalen Gleichungssystem gegeben.

Unter Bezug auf die klassische Statistik und den expliziten Ansatz

$$f(x, y) = \hat{f}(x, y) + \varepsilon(x, y) = p(x, y) + \varepsilon(x, y)$$

mit einem bivariat normalverteilten Term $\varepsilon(x, y)$ wird aus dem Anpassen einer Fläche an eine Punktfolge die Trendflächenanalyse.

3.2.5 Tensor-Produkt von polynomialen Splines

$$\hat{f}(x, y) = \sum_{k,l} c_{k,l} \varphi_k(x) \varphi_l(y)$$

3.2.6 Radiale Basisfunktionen

1D Methoden lassen sich direkt, das heißt durch Verwendung des verallgemeinerten Abstands, übertragen.

3.3 3D, nD Methoden

3.4 Allgemeines zu deterministischen Methoden

In den bis jetzt angestellten Betrachtungen erscheinen die Kriterien zur Bestimmung der Nachbarschaftsrelation und ihrer Wichtung eher willkürlich. Bei Methoden auf der Grundlage einer Triangulation sind sie weitgehend entkoppelt, bei Methoden mit radialen Basisfunktionen ist die Wahl einer speziellen radialen Funktion von den Daten unabhängig.

Es gibt kein Maß für die Güte des Interpolanten, außer allgemeinen Kenntnissen über sein Verhalten, z.B. bei den Interpolationspolynomen. Um zu einer Beurteilung der Güte des Interpolanten zu gelangen, bedarf es der Kenntnisse bzw. Annahmen von Eigenschaften der deterministisch als Funktion f gegebenen “Gesamtheit”, von der man sich die an den Orten \mathbf{x}_i beobachteten Werte f_i vorstellt zu stammen, oder der Verteilung der stochastischen Gesamtheit, von der man sich die beobachteten Werte als Stichprobe vorstellt.

Abhilfe kann in dieser Situation also die Einführung von stochastisch begründeten Methoden leisten, die in irgendeiner Form Kenntnisse oder Annahmen über Eigenschaften der Gesamtheit berücksichtigen können und damit auch zu weitergehenden Aussagen führen. In diesem Fall spricht man von Schätzung. Einen gedanklichen Anknüpfungspunkt dazu liefert die Idee der Regression.

Chapter 4

Erkundende Daten Analyse

4.1 Der JURA–Datensatz

An 359 flächenhaft unregelmäßig verteilten Orten im Untersuchungsgebiet wurden an oberflächennahen Bodenproben 7 Schwermetall-Konzentrationen gemessen und Gesteinstyp und Flächennutzung als nominelle (kategoriale) Merkmale erfaßt.

Die geometrische Anordnung der 359 Orte ist das Resultat der Kombination einer regelmäßigen (regular) und einer geschachtelten (nested) Anordnung (configuration). Die regelmäßige Anordnung besteht aus einem quadratischen Raster mit 107 Rasterpunkten im Rasterabstand von 250m. Von diesen 107 Rasterpunkten wurden 38 zufällig ausgewählt. Die zufällige Auswahl hat zur Folge, daß der Anteil der 38 zufällig ausgewählten Rasterpunkte, der zum Gesteinstyp k , $k = 1, \dots, 5$, bzw zur Flächennutzung k' , $k' = 1, \dots, 4$, gehört, dem Anteil des Gesteinstyp k , bzw. der Flächennutzung k' an der Gesamtfläche des Untersuchungsgebiete entspricht.

Ausgehend von jedem dieser 38 zufällig ausgewählten Rasterpunkte wurde ein erster zusätzlicher Punkt im Abstand von 100m, ein zweiter zusätzlicher im Abstand von 40m vom ersten, ein dritter im Abstand von 16m vom zweiten und ein vierter im Abstand von 6m vom dritten gewählt; dabei wurden die Richtungen von einem zum nächsten Punkt zufällig gewählt.

Ziel dieser Anordnung ist ein möglichst gutes flächenhaftes Auflösungsvermögen über Distanzen zwischen 0 und 250m.

4.2 Häufigkeitstabelle kategorialer Zufallsvariablen

Die Verteilungen der beiden kategorialen Zufallsvariablen Gesteinstyp und Flächennutzung liest man an ihren Häufigkeitstabellen ab.

4.3 Kontingenztafel (cross table)

Die gemeinsame Verteilung von zwei kategorialen Zufallsvariablen, z.B. Gesteinstyp und Flächennutzung liest man an ihrer Kreuztabelle bzw. Kontingenztabelle ab. Sie liefert einen ersten Eindruck, welche Gesteinstyp - Flächennutzung Kombinationen besonders selten oder häufig sind.

4.4 Histogramm und empirische Verteilungsfunktion

Die Schwermetallkonzentrationen werden als stetige Zufallsvariable behandelt. Ihre Verteilungen stellt man als Histogramme oder kumulative empirische Verteilungsfunktionen dar. Mit Hilfe der Histogramme kommt man z.B. zu Aussagen über Symmetrie oder Schiefe und Uni- oder Multimodalität der Verteilung. An der kumulativen empirischen Verteilungsfunktion kann man instruktiv den Anteil der Meßwerte, die größer als ein gegebener Grenzwert sind, ablesen.

4.5 Elementare Statistiken (summary statistics)

Die Tabelle der elementaren Statistiken Mittelwert, Varianz, Standardabweichung, Quartile, Median, Schiefe, etc. vermittelt einen summarischen Einblick in die Daten.

4.6 Umgang mit extremen Werten, Datentransformationen

- Meßfehler; fehlerhafte Daten werden aus dem Datensatz entfernt. Zur Entscheidung, einen Meßwert als fehlerhaft zu charakterisieren, zieht man seinen Ort und die benachbarten Werte zu Rate.
- Proben von verschiedenen Gesamtheiten; extreme Daten werden als von einer anderen Gesamtheit stammend charakterisiert.
- Anwendung robuster Statistiken in Verbindung mit klassischen Statistiken.
- Anwendung von Datentransformationen zur Reduzierung des Einflusses extremer Meßwerte. Dabei ist die Rücktransformation das Problem.

4.7 Bedingte Verteilungen (conditional distributions)

Die Verteilungen bzw. Kennwerte der Verteilungen der Schwermetallkonzentrationen werden je nach Zugehörigkeit zu einem Gesteinstyp oder einer Flächennutzung untersucht. Man stellt

geringere Konzentrationen für Wald und den Gesteinstyp Argovian fest. Dabei sind Wald und Argovian schlecht korreliert (0.21). Bestimmt man den Anteil der Daten der Schwermetallkonzentrationen von Cd, Cu, Pb grösser als der Grenzwert für jeden Gesteinstyp bzw. jede Flächennutzung, stellt man z.B. keinen offensichtlichen Zusammenhang zwischen Cd, Pb Konzentrationen und Flächennutzung fest.

4.8 Streudiagramme (scatterplots)

Die Schwermetallkonzentrationen von Cd, Cu, und Pb sind wegen ihrer großen Anteile über dem jeweiligen Grenzwert von besonderem Interesse, die Konzentrationen von Ni, Zn sind am besten beprobt. Daher ist ein möglicher Zusammenhang zwischen diesen Paaren (Ni, Cd), (Zn, Cd), (Ni, Cu), (Zn, Cu), (Ni, Pb), (Zn; Pb) besonders interessant.

4.9 Lineare und Rang-Korrelation

Die Matrix der linearen Korrelationskoeffizienten mit Signifikanzniveaus und die Matrix der Rang Korrelationskoeffizienten mit Signifikanzniveaus gibt weitere Auskunft über Zusammenhänge. Große Abweichungen der linearen von den Rang Korrelationskoeffizienten zeigen entweder die Existenz eines nicht-linearen Zusammenhangs oder von Paaren extremer Werte an.

4.10 Karten georeferenzierter Daten

Jeder Meßwert besitzt (x, y) -Ortskoordinaten, sodaß der Datensatz kartenmäßig dargestellt werden kann, indem man die Schwermetallkonzentrationen klassifiziert, den Klassen Farben oder Grautöne zuweist und diese druckt. Indikatorkarten der Schwermetallkonzentrationen über dem Grenzwert zeigen das Ausmaß der Belastungen durch Cd und Pb an. Man erkennt, daß die Orte mit großen Konzentrationen nicht gleichmäßig verteilt sind, sondern räumliche Cluster bilden.

4.11 h -Streudiagramm (h -scattergram)

Das h -scattergram ist der scatter plot von $z(\mathbf{x}_\alpha)$, $z(\mathbf{x}_\alpha + \mathbf{h})$ und vermittelt eine Darstellung der räumlichen Variabilität. Es ist sehr nützlich bei der Identifikation extremer Werte.

Bei der Konstruktion benutzt man zu einem gegebenen Meßwert am Ort \mathbf{x}_α nicht nur den Meßwert am Ort $\mathbf{x}_\alpha + \mathbf{h}$, sondern auch die Werte an den Orten im Sektor, der durch $h \pm \Delta h$, $\theta \pm \Delta \theta$ gegeben ist.

Gewöhnlich nimmt die räumliche Variabilität mit zunehmendem Abstand h zu, die Punktewolke im h -scattergram erscheint zunehmend schlechter geschart um die Diagonale des ersten Quadranten.

4.12 Maße für räumliche Variabilität und Stetigkeit

Empirische Kovarianzfunktion

$$\begin{aligned} C(\mathbf{h}) &= \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} z(\mathbf{x}_\alpha) z(\mathbf{x}_\alpha + \mathbf{h}) - m_{-\mathbf{h}} m_{+\mathbf{h}} \\ m_{-\mathbf{h}} &= \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} z(\mathbf{x}_\alpha) \\ m_{+\mathbf{h}} &= \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} z(\mathbf{x}_\alpha + \mathbf{h}) \end{aligned}$$

Empirisches Korrelogramm

$$\begin{aligned} \rho(\mathbf{h}) &= \frac{C(\mathbf{h})}{\sqrt{\sigma_{-\mathbf{h}}^2 \sigma_{+\mathbf{h}}^2}} \\ \sigma_{-\mathbf{h}}^2 &= \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{x}_\alpha) - m_{-\mathbf{h}}]^2 \\ \sigma_{+\mathbf{h}}^2 &= \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{x}_\alpha + \mathbf{h}) - m_{+\mathbf{h}}]^2 \end{aligned}$$

Empirisches Semivariogramm

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha + \mathbf{h})]^2$$

dabei wird $z(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha + \mathbf{h})$ der \mathbf{h} -Zuwachs von z genannt

Geometrisch ist $\frac{1}{2}[Z(\mathbf{x}_\alpha) - Z(\mathbf{x}_\alpha + \mathbf{h})]^2$ der (orthogonale) Abstand des Punktes mit den Koordinaten $[Z(\mathbf{x}_\alpha), Z(\mathbf{x}_\alpha + \mathbf{h})]$ von der Diagonalen im ersten Quadranten. Damit kann man das Semivariogramm $\gamma(\mathbf{h})$ interpretieren als Trägheitsmoment bezüglich der Diagonalen des \mathbf{h} -Streudiagramms.

Die Entfernung $\|\mathbf{h}\|$ in Richtung θ , von der ab die Kovarianzfunktion bzw. das Korrelogramm annähernd 0 sind, heißt *Reichweite*. Das Semivariogramm hört bei der Reichweite auf, systematisch zuzunehmen, und schwankt für größere \mathbf{h} um den sogenannten *Schwellenwert*.

Obwohl per Definition $\gamma(0) = 0$, beobachtet man oft $\gamma(\mathbf{h}) \rightarrow \gamma(0) \neq 0, \mathbf{h} \rightarrow 0$. Diese scheinbare Unstetigkeit des Semivariogramms heißt *Nuggeteffekt* und ist charakteristisch für den Einfluß von Meßfehlern oder charakterisiert räumliche Variabilität über Entfernungen, die kleiner als die kleinste mit Meßwerten belegte Schrittweite \mathbf{h} sind.

4.13 Indikator Transformation

Die räumliche Variabilität kann für die größeren bzw. kleineren Werte verschieden sein. Um diese Unterschiede erfassen zu können, führt man Indikator Variable ein, indem man die Indikator Transformation auf die Daten anwendet und diese Werte als Realisationen der entsprechenden Indikator Variablen auffaßt.

Zur Charakterisierung der räumlichen Verteilung von z -Werten, die größer (oder kleiner) als ein vorgegebener Wert z_k sind, führt man die *Indikator-Transformation*

$$i(\mathbf{x}_\alpha; z_k) = \begin{cases} 1, & \text{falls } z(\mathbf{x}_\alpha) \leq z_k \\ 0, & \text{sonst} \end{cases}$$

ein.

Diese Indikator-Transformation führt man nacheinander für verschiedene Schranken z_k durch, z.B. für ausgesuchte Quantile der empirischen Verteilung von z .

Die Maße für räumliche Variabilität und Stetigkeit werden für Indikator-Transformierte analog definiert.

Empirische Indikator Kovarianzfunktion

$$\begin{aligned} C_I(\mathbf{h}; z_k) &= \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} i(\mathbf{x}_\alpha; z_k) i(\mathbf{x}_\alpha + \mathbf{h}; z_k) - F_{-\mathbf{h}}(z_k) F_{+\mathbf{h}}(z_k) \\ &= F(\mathbf{h}; z_k) - F_{-\mathbf{h}}(z_k) F_{+\mathbf{h}}(z_k) \\ F_{-\mathbf{h}}(z_k) &= \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} i(\mathbf{x}_\alpha; z_k) \\ F_{+\mathbf{h}}(z_k) &= \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} i(\mathbf{x}_\alpha + \mathbf{h}; z_k) \end{aligned}$$

Dabei bedeutet $F(\mathbf{h}; z_k)$ die gemeinsame Häufigkeit, daß $z(\mathbf{x}_\alpha) < z_k$ und $z(\mathbf{x}_\alpha + \mathbf{h}) < z_k$.

Für kleine Werte von z_k ist $F(\mathbf{h}; z_k)$ (und auch $C_I(\mathbf{h}; z_k)$) ein Maß für den räumlichen Zusammenhang kleiner Meßwerte mit Entfernung \mathbf{h} , je größer $F(\mathbf{h}; z_k)$, umso besser der räumliche Zusammenhang.

Für die Charakterisierung der großen z -Werte kann man die Indikator-Transformation

$$j(\mathbf{x}_\alpha; z_k) = 1 - i(\mathbf{x}_\alpha; z_k)$$

heranziehen. Es gilt

$$C_I(\mathbf{h}; z_k) = C_J(\mathbf{h}; z_k)$$

Empirisches Indikator Korrelogramm

$$\begin{aligned}\rho_I(\mathbf{h}; z_k) &= \frac{C_I(\mathbf{h}; z_k)}{\sqrt{\sigma_{-\mathbf{h}}^2(z_k)\sigma_{+\mathbf{h}}^2(z_k)}} \\ \sigma_{-\mathbf{h}}^2(z_k) &= F_{-\mathbf{h}}(z_k)[1 - F_{-\mathbf{h}}(z_k)] \\ \sigma_{+\mathbf{h}}^2(z_k) &= F_{+\mathbf{h}}(z_k)[1 - F_{+\mathbf{h}}(z_k)],\end{aligned}$$

wobei $\sigma_{-\mathbf{h}}^2(z_k)$ die Varianz der Anfangs-Indikatorwerte und $\sigma_{+\mathbf{h}}^2(z_k)$ die Varianz der End-Indikatorwerte bezeichnet

Empirisches Indikator–Semivariogramm

$$\gamma_I(\mathbf{h}; z_k) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [i(\mathbf{x}_\alpha; z_k) - i(\mathbf{x}_\alpha + \mathbf{h}; z_k)]^2$$

Das Indikator Semivariogramm ist ein Maß dafür, wie oft zwei z –Werte für Orte mit der Entfernung \mathbf{h} auf verschiedenen Seiten von z_k liegen. In diesem Sinne kann man das Indikator Semivariogramm als Übergangshäufigkeit interpretieren.

Das Indikator–Semivariogramm ist ziemlich robust, insbesondere kann das Indikator–Semivariogramm zum Median

$$\gamma_I(\mathbf{h}; z_{0.5})$$

|
Median

zusätzlich zum gewöhnlichen Semivariogramm benutzt werden, wenn dieses wegen extremer Werte zu erratisch erscheint.

Für kategoriale Variable

$$i(\mathbf{x}_\alpha; s_k) = \begin{cases} 1, & \text{falls } s(\mathbf{x}_\alpha) = s_k \\ 0, & \text{sonst.} \end{cases}$$

ist nur die Definition eines Indikator–Semivariogramms sinnvoll:

$$\gamma_I(\mathbf{h}; s_k) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [i(\mathbf{x}_\alpha; s_k) - i(\mathbf{x}_\alpha + \mathbf{h}; s_k)]^2$$

Die Interpretation als Übergangshäufigkeit ist für kategoriale Variable offensichtlich; je kleiner $\gamma_I(\mathbf{h}; s_k)$, desto besser der räumliche Zusammenhang des Zustands s_k .

4.14 Kreuz \mathbf{h} –Streudiagramm (cross \mathbf{h} –scattergram)

Das Kreuz \mathbf{h} –Streudiagramm ist der scatterplot von $z_i(\mathbf{x}_\alpha), z_j(\mathbf{x}_\alpha + \mathbf{h})$, wobei z_i, z_j verschiedene Attribute, z.B. Konzentrationen verschiedener Schwermetalle bezeichnen.

Für $\mathbf{h} = 0$ ist das Kreuz \mathbf{h} –Streudiagramm gerade der gewöhnliche scatter plot, das gewöhnliche Streudiagramm.

Für jedes $\mathbf{h} \neq 0$ gibt es zwei Kreuz \mathbf{h} –Streudiagramme: $z_i(\mathbf{x}_\alpha), z_j(\mathbf{x}_\alpha + \mathbf{h})$ und $z_j(\mathbf{x}_\alpha), z_i(\mathbf{x}_\alpha + \mathbf{h})$.

4.15 Maße für räumliche Kreuz–Variabilität und Kreuz–Stetigkeit

Empirische Kreuz–Kovarianzfunktion

$$\begin{aligned} C_{ij}(\mathbf{h}) &= \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} z_i(\mathbf{x}_\alpha) z_j(\mathbf{x}_\alpha + \mathbf{h}) - m_{i-\mathbf{h}} m_{i+\mathbf{h}} \\ m_{i-\mathbf{h}} &= \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} z_i(\mathbf{x}_\alpha) \\ m_{i+\mathbf{h}} &= \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} z_j(\mathbf{x}_\alpha + \mathbf{h}) \end{aligned}$$

$$\begin{aligned} \text{i.a. } C_{ij}(\mathbf{h}) &\neq C_{ij}(-\mathbf{h}) \\ \text{aber } C_{ij}(\mathbf{h}) &= C_{ji}(-\mathbf{h}) \end{aligned}$$

Empirisches Kreuz–Korrelogramm

$$\begin{aligned} \rho_{ij}(\mathbf{h}) &= \frac{C_{ij}(\mathbf{h})}{\sqrt{\sigma_{i-\mathbf{h}}^2 \sigma_{j+\mathbf{h}}^2}} \\ \sigma_{i-\mathbf{h}}^2 &= \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z_i(\mathbf{x}_\alpha) - m_{i-\mathbf{h}}]^2 \\ \sigma_{j+\mathbf{h}}^2 &= \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z_j(\mathbf{x}_\alpha + \mathbf{h}) - m_{j+\mathbf{h}}]^2 \end{aligned}$$

Empirisches pseudo Kreuz–Semivariogramm

$$\gamma_{ij}^p(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z_i(\mathbf{x}_\alpha) - z_j(\mathbf{x}_\alpha + \mathbf{h})]^2$$

Verändert man die Sichtweise von $[z_i(\mathbf{x}_\alpha), z_j(\mathbf{x}_\alpha + \mathbf{h})]$ zugunsten der Zuwächse

$$\left([z_i(\mathbf{x}_\alpha) - z_i(\mathbf{x}_\alpha + \mathbf{h})], [z_j(\mathbf{x}_\alpha) - z_j(\mathbf{x}_\alpha + \mathbf{h})] \right),$$

so erhält man das **empirische Kreuz–Semivariogramm**:

$$\gamma_{ij}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z_i(\mathbf{x}_\alpha) - z_i(\mathbf{x}_\alpha + \mathbf{h})][z_j(\mathbf{x}_\alpha) - z_j(\mathbf{x}_\alpha + \mathbf{h})].$$

Empirische Indikator Kreuz–Kovarianzfunktion

$$\begin{aligned} C_{ij}^I(\mathbf{h}; z_{ik}, z_{jk'}) &= \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} i(\mathbf{x}_\alpha; z_{ik}) i(\mathbf{x}_\alpha + \mathbf{h}; z_{jk'}) - F_{i_{-\mathbf{h}}}(z_{ik}) F_{j_{+\mathbf{h}}}(z_{jk'}) \\ &= F_{ij}(\mathbf{h}; z_{ik} z_{jk'}) - F_{i_{-\mathbf{h}}}(z_{ik}) F_{j_{+\mathbf{h}}}(z_{jk'}) \end{aligned}$$

Empirisches Indikator Kreuz–Korrelogramm

$$\begin{aligned} \rho_{ij}^I(\mathbf{h}; z_{ik}, z_{jk'}) &= \frac{C_{ij}^I(\mathbf{h}; z_{ik}, z_{jk'})}{\sqrt{\sigma_{i_{-\mathbf{h}}}^2(z_{ik}) \sigma_{j_{+\mathbf{h}}}^2(z_{jk'})}} \\ \sigma_{i_{-\mathbf{h}}}^2(z_{ik}) &= F_{i_{-\mathbf{h}}}[1 - F_{i_{-\mathbf{h}}}(z_{ik})] \\ \sigma_{j_{+\mathbf{h}}}^2(z_{jk'}) &= F_{j_{+\mathbf{h}}}[1 - F_{j_{+\mathbf{h}}}(z_{jk'})] \end{aligned}$$

Empirisches Indikator Kreuz–Semivariogramm

$$\gamma_{ij}^I(\mathbf{h}; z_{ik}, z_{jk'}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [i(\mathbf{x}_\alpha; z_{ik}) - i(\mathbf{x}_\alpha + \mathbf{h}; z_{ik})] - [i(\mathbf{x}_\alpha; z_{jk'}) - i(\mathbf{x}_\alpha + \mathbf{h}; z_{jk'})]$$

Chapter 5

Theorie der Regionalisierten Variablen

5.1 Einführung

Geostatistik wird hier als Kurzformel für die Anwendung stochastischer Prozesse und bedingter Wahrscheinlichkeiten in den Geowissenschaften verwendet. Ursprünglich bezog sich Geostatistik fast ausschließlich auf Lagerstättenvorratsberechnung, insbesondere im südafrikanischen Goldbergbau (Daniel Krige, kriging), heute bezieht sich der Begriff auf die räumliche Schätzung geo-referenzierter Daten, oder noch weiter gefaßt auf stochastisch begründete Methoden der räumlichen Interpolation/Approximation.

Mit Hilfe der Theorie und Statistik stochastischer Prozesse (zweiter Ordnung) wurde aus der empirisch - heuristisch begründeten Methode Krige's zur Lagerstättenvorratsberechnung die so definierte und von Georges Matheron und seiner Fontainebleau-Schule seit den 60er Jahren wesentlich geprägte Geostatistik. Wesentliche Beiträge zur Weiterentwicklung der Theorie, insbesondere auch ihrer deterministischen Seite, und der Anwendbarkeit auf neue Problemstellungen gehen seit Anfang der 80er Jahre von der Stanford-Schule und André Journel aus. “Until the late 1980s, geostatistics was essentially viewed as a means to describe spatial patterns and interpolate the value of the attribute of interest at unsampled locations. Geostatistics is now increasingly used to model the uncertainty about unknown values ...”, Goovaerts, p.3.

Die Zeitschrift “Mathematical Geology” der “International Association for Mathematical Geology” ist zu einem großen Teil der Geostatistik gewidmet.

Geostatistik wendet man dann an, wenn weniger aufwendige Methoden der klassischen Statistik oder der räumlichen Interpolation / Approximation / Schätzung keine brauchbaren Ergebnisse liefern, weil die Verhältnisse einer Lagerstätte bzw. Eigenarten eines räumlichen Phänomens zu komplex sind.

“Priority was given to practicality, a current trademark of geostatistics that explains its success and application in such diverse fields as mining, petroleum, soil science, oceanography, hydrogeology, remote sensing, and environmental sciences”, Goovaerts, p. 3. Geostatistik ist also eher eine Methode für den Anwender, also den Bergbauingenieur bzw. Geowissenschaftler, wobei sich “geo” hier wesentlich auf geo-referenzierte Daten bezieht, als für den Mathematiker.

5.1.1 Bergbauliche Motivation 1 (“change of support”)

Die Erfahrung lehrt, daß die Variabilität der beobachtbaren Größe “Erzgehalt” von ihrem materiellen Träger (support), das heißt vom geometrischen Inhalt des Körpers, dessen Erzgehalt bestimmt wurde, abhängt; und zwar in dem Sinne, daß die Variabilität der beobachteten Größe mit zunehmendem Volumen des Trägers abnimmt. Eine Plausibilitätsklärung für diese Erfahrung liefern die folgenden Überlegungen:

- im Mikrobereich kann der Erzgehalt zwischen 0% und 100% variieren, je nach dem, ob ein Erzmineral oder Nichterzmineral getroffen wurde;
- Erzgehalte von Proben (aus einer homogenen Lagerstätte) variieren beträchtlich;
- Erzgehalte von Abbaublöcken (aus einer homogenen Lagerstätte) variieren geringfügig;
- der Erzgehalt des ganzen Erzkörpers ist eine eindeutig bestimmte deterministische Größe.

Die Unterschiede in der Variabilität bezüglich Probekörpern oder Abbaublöcken nehmen mit zunehmender Homogenität der Lagerstätte ab.

In diesem Sinne gibt es eine beobachtbare absolute Größe “Erzgehalt” mit bestimmten (statistischen) Eigenschaften nicht; es gilt zwischen den Variablen “Erzgehalt von Probekörpern (aus Bohrkernen)”, “Erzgehalt von Abbaublöcken”, “Erzgehalt eines Erzkörpers” etc. zu unterscheiden. Diese Größen und ihre statistischen Eigenschaften sind i.a. nicht direkt vergleichbar, das heißt, man kann i.a. nicht von einem Verteilungsmerkmal (Varianz) der Größe “Erzgehalt von Probekörpern” auf das entsprechende Verteilungsmerkmal (Varianz) der Größe “Erzgehalt der Abbaublocke” schließen. Dazu wird auch auf das Beispiel in David, 1977, p. 45-48 verwiesen.

5.1.2 Bergbauliche Motivation 2 (“spatial continuity”)

Die Erfahrung lehrt, daß die tatsächlichen “Erzgehalte von Probekörpern” aus einer homogenen Lagerstätte

- umso ähnlicher sind, je näher die Probekörper im Erzkörper benachbart sind;
- unabhängig von einander sind, wenn die Distanz zwischen ihnen eine lagerstättenspezifische Größe überschritten hat

Dementsprechend können die Zufallsvariablen zur Beschreibung der Lagerstätte i.a. nicht als stochastisch unabhängig angenommen werden. Es steht also das mathematische Problem an, Phänomene mit der Eigenschaft der ortsbezogenen stochastischen Abhängigkeit der sie beschreibenden Zufallsvariablen mit einem anwendbaren mathematischen Modell zu behandeln.

5.1.3 Unzulänglichkeiten der klassischen Statistik - Modellannahme der unabhängig identisch verteilten Zufallsvariablen

Die klassische Statistik stellt Methoden bereit, wie man von Eigenschaften (z.B. Mittelwert, empirische Varianz, empirische Quantile) der n unabhängigen Realisationen (Daten) $z_i, i = 1, \dots, n$, auf Eigenschaften der zugrundeliegenden Gesamtheit, das heißt Zufallsvariablen Z (z.B. Erwartungswert, Varianz, Quantile) schließt, bis man die Gesamtheit bzw. die Zufallsvariable Z soweit wie möglich bzw. nötig bestimmt hat; im (unrealistischen) Idealfall bis zur Bestimmung ihrer Verteilung.

Behandelt man stochastisch abhängige Zufallsvariablen mit Methoden für stochastisch unabhängige Variablen, so führt dies i.a. zu schweren Fehlern bei Schätzungen.

Unter den Annahmen der klassischen Statistik werden die n Erzgehalte $z_i, i = 1, \dots, n$, der Probekörper als n unabhängige Realisationen einer einzigen Zufallsvariablen Z aufgefaßt. Dies ist gleichbedeutend damit, jeden einzelnen Erzgehalt z_i eines Probekörpers als Realisation einer Zufallsvariablen Z_i aufzufassen, wenn man annimmt, daß die Z_i die gleiche Verteilung (wie Z) besitzen, das heißt identisch verteilt sind, und stochastisch unabhängig sind, das heißt unabhängige Wiederholungen von Z sind. Insbesondere ist die gemeinsame Verteilung der Zufallsvariablen $Z_i, i = 1, \dots, n$, wegen ihrer stochastischen Unabhängigkeit gleich dem Produkt der Einzelverteilungen, also gleich dem n -fachen Produkt der Verteilung der Zufallsvariablen Z .

Dabei bedeutet stochastische Unabhängigkeit "praktisch", daß aus der Kenntnis von $n-1$ Realisationen $z_i, i = 1, \dots, n-1$, (Erzgehalte von $n-1$ Probekörpern) keinerlei zusätzliche Kenntnisse über die n -te Realisation z_n (Erzgehalt des n -ten Probekörpers) zu gewinnen ist; keiner der möglichen Werte für die n -te Realisation z_n ist aufgrund der Kenntnis der $n-1$ Realisationen $z_i, i = 1, \dots, n-1$, bevorzugt zu erwarten. Dies widerspricht den unter den bergbaulichen Motivationen angeführten Erfahrungen.

Da die einzelnen Verteilungen und damit auch die gemeinsame Verteilung der Zufallsvariablen Z_i nicht bekannt sind, sondern gerade Hauptgegenstand unseres Interesses sind, kann man die Annahme der stochastischen Unabhängigkeit und identischen Verteilung der Z_i in diesem Zusammenhang nicht mathematisch beweisen. Sie geht a priori in das zugrunde gelegte Modell ein. Die Annahme der stochastischen Unabhängigkeit und identischen Verteilung sollte durch außermathematische (geologische, geophysikalische, geochemische, etc.) Kenntnisse und Überlegungen gerechtfertigt sein. Beschreibt das mit diesen Annahmen behaftete Modell das interessierende Phänomen bezüglich gewisser Kriterien (Prognose der Gesamtmenge förderbaren Erzes) hinreichend gut, das heißt, stimmen aus dem Modell gewonnene Aussagen mit der beobachteten Realität überein, ist das Phänomen, soweit es von durch Beobachtungen belegtem Interesse ist, mit diesen Annahmen verträglich. Das heißt tatsächlich nicht, daß die beobachteten Zufallsvariablen unabhängig und identisch verteilt sind! Stimmen Modell und Realität nicht hinreichend gut überein, so kann man unter der Bedingung, daß alle Rechenoperationen des Modells richtig ausgeführt wurden, schließen, daß die beobachteten Zufallsvariablen nicht unabhängig oder nicht identisch verteilt sind, sodaß das Modell nicht

zum Phänomen paßt.

Die Methoden der klassischen Statistik (der Mathematischen Geologie I) sind hier also nur sehr beschränkt anwendbar, denn die klassische Statistik behandelt Daten als unabhängige Realisationen einer Zufallsvariablen. Nur für einen extrem homogenen Erzkörper liefern die Methoden der klassischen Statistik brauchbare Ergebnisse. In diesem eher unrealistischen Fall variieren die Erzgehalte der Probekörper nur sehr geringfügig, jeder einzelne gemessene Erzgehalt weicht nur sehr geringfügig vom Mittelwert der Gehalte der Probekörper ab; jeder einzelne gemessene Gehalt ist ein guter Schätzwert für den Erwartungswert der Gehalte der Probekörper und der Erwartungswert der Gehalte der Probekörper ist ein guter Schätzwert für den Erwartungswert der Gehalte der Abbaublöcke.

5.1.4 Begriffe der klassischen Statistik, Elementare Regeln für das Rechnen mit Erwartungswerten und Varianzen

Seien Z, Z_1, Z_2 reellwertige Zufallsvariablen, deren Erwartungswerte und Varianzen existieren.

Sei F die Verteilungsfunktion und f die Wahrscheinlichkeitsdichte der Zufallsvariablen Z

$$F(z) := \text{Prob}(Z \leq z) = E(I(z))$$

mit der Indikatorvariablen $I(z)$ definiert als

$$I_Z(z) := \begin{cases} 1 & \text{falls } Z \leq z \\ 0 & \text{sonst} \end{cases}$$

Der Erwartungswert der Zufallsvariablen Z ist definiert als

$$E(Z) := \int_{\mathbb{R}} z dF(z) = \int_{\mathbb{R}} z f(z) dz$$

und ihre Varianz als

$$\text{Var}(Z) := E([Z - E(Z)]^2) = E(Z^2) - E^2(Z) \geq 0$$

Es gilt

$$\begin{aligned} E(aZ + b) &= aE(Z) + b \quad (a, b \in \mathbb{R}) \\ \text{Var}(aZ + b) &= a^2 \text{Var}(Z) \\ E(Z_1 \pm Z_2) &= E(Z_1) \pm E(Z_2) \\ \text{Var}(Z_1 \pm Z_2) &= \text{Var}(Z_1) + \text{Var}(Z_2) \pm 2 \text{Cov}(Z_1, Z_2) \\ \text{mit } \text{Cov}(Z_1, Z_2) &= E([Z_1 - E(Z_1)][Z_2 - E(Z_2)]) \\ &= E(Z_1 Z_2) - E(Z_1)E(Z_2) \end{aligned}$$

Zwei Zufallsvariablen Z_1, Z_2 heißen unkorreliert, falls

$$\text{Cov}(Z_1, Z_2) = 0$$

Für unkorrelierte Zufallsvariablen Z_1, Z_2 gilt

$$\begin{aligned}\text{E}(Z_1, Z_2) &= \text{E}(Z_1)\text{E}(Z_2) \\ \text{Var}(Z_1 \pm Z_2) &= \text{Var}(Z_1) + \text{Var}(Z_2)\end{aligned}$$

Sind zwei Zufallsvariablen stochastisch unabhängig, dann sind sie auch unkorreliert; die Umkehrung gilt nicht.

Es gilt

$$\text{Cov}(Z, Z) = \text{Var}(Z)$$

und man bezeichnet $\text{Cov}(Z, Z)$ auch als Autokovarianz der Zufallsvariablen Z .

Die klassische Statistik behandelt die Folge der Meßwerte x_1, \dots, x_n als n Realisationen der unabhängig identisch verteilten Zufallsvariablen $Z_i, i = 1, \dots, n$.

$$\begin{aligned}P(Z_i \leq z) &= P(Z_j \leq z) = F(z) \\ \text{E}(Z_i) &= \text{E}(Z_j) = m \\ \text{Var}(Z_i) &= \text{Var}(Z_j) = v\end{aligned}$$

Jede gemeinsame Verteilung ist das Produkt der entsprechenden einzelnen Verteilungen

$$\begin{aligned}P(Z_{n_1} \leq z_{n_1}, \dots, Z_{n_k} \leq z_{n_k}) &\stackrel{\text{unabh.}}{=} \\ \prod_{l=1}^k P(Z_{n_l} \leq z_{n_l}) &\stackrel{\text{id. Vert.}}{=} \\ \prod_{l=1}^k P(Z_{n_l+m_l} \leq z_{n_l}) &\stackrel{\text{unabh.}}{=} \\ P(Z_{n_1+m_1} \leq z_{n_1}, \dots, Z_{n_k+m_k} \leq z_{n_k})\end{aligned}$$

Unabhängigkeit und identische Verteilung gehen als a-priori Annahmen in dieses Modell der klassischen Statistik ein und können dann nicht mathematisch bewiesen werden. Muß man diese Annahmen aufgeben, da das Modell nicht tauglich ist, rückt die Familie der endlichdimensionalen Verteilungen in den Mittelpunkt des Modells. Die Kenntnis der Einzelverteilungen ist unzulänglich, um die endlichdimensionalen Verteilungen zu bestimmen.

5.2 Zufallsfunktion, regionalisierte Zufallsvariablen, stochastischer Prozeß

Zur Überwindung der Einschränkungen der bis jetzt erinnerten Methoden der klassischen Statistik wird die Theorie der stochastischen Prozesse angewandt. Aus der neuen Sicht

der stochastischen Prozesse handelt es sich bei den Erzgehalten $z_\alpha, \alpha = 1, \dots, n$, der n Probekörper an den Orten $\mathbf{x}_\alpha \in D, \alpha = 1, \dots, n$, um **eine** diskrete Realisation $z(\mathbf{x}_\alpha), \mathbf{x}_\alpha \in D, \alpha = 1, \dots, n$, **einer** ortsabhängigen Zufallsfunktion $Z(\mathbf{x}), \mathbf{x} \in D$. Diese Zufallsfunktion $Z(\mathbf{x})$ heißt auch regionalisierte Variable (georeferenzierte Variable wäre eine - allerdings ungebräuchliche - Alternative). Diese Namensgebung soll die zwei wesentlichen Eigenschaften verdeutlichen

- für $\mathbf{x} = \mathbf{x}_\alpha \in D$ ist eine Zufallsfunktion $Z(\mathbf{x})$ lokal eine Zufallsvariable $Z(\mathbf{x}_\alpha)$;
- insgesamt sind die Zufallsvariablen, die eine Zufallsfunktion bilden, i.a. nicht unabhängig und nicht identisch verteilt

Der Übergang in der Schreibweise $z_\alpha = z(\mathbf{x}_\alpha)$ bzw. $Z_\alpha = Z(\mathbf{x}_\alpha)$ soll anzeigen, daß die (gegenseitige) räumliche Lage der analysierten Probekörper von nun an mit in Betracht gezogen wird.

Im Mittelpunkt unseres Interesse steht nun, Kenntnis über die gemeinsamen Verteilungen $P(Z(\mathbf{x}_\alpha) \leq z_1, Z(\mathbf{x}_\beta) \leq z_2, \dots)$ zu gewinnen. So wie die Verteilung einer einzelnen Zufallsvariablen diese vollständig beschreibt, beschreiben die gemeinsamen Verteilungen eine Zufallsfunktion vollständig. Da die Zufallsvariablen $Z_\alpha = Z(\mathbf{x}_\alpha)$ der Zufallsfunktion gerade nicht als unabhängig anzunehmen sind, sind die eindimensionalen Randverteilungen hier von geringerem Interesse, da sie alleine nicht die gemeinsamen Verteilungen bestimmen.

5.2.1 Begriffe aus der Theorie der zufälligen Funktionen

Sei $Z(\mathbf{x}), \mathbf{x} \in D \subset \mathbb{R}^p, p \geq 2$, eine Zufallsfunktion. Dann definiert man (unter der Voraussetzung, daß für jede ihrer Zufallsvariablen $Z(\mathbf{x}_0)$ mit beliebigem, aber festen $\mathbf{x}_0 \in D$ Erwartungswert und Varianz existieren)

Ein-Punkt Verteilungsfunktion	$F(\mathbf{x}; z) := \text{Prob}(Z(\mathbf{x}) \leq z)$
Zwei-Punkte Verteilungsfunktion	$F(\mathbf{x}_\alpha, \mathbf{x}_\beta; z_1, z_2) := \text{Prob}(Z(\mathbf{x}_\alpha) \leq z_1, Z(\mathbf{x}_\beta) \leq z_2)$
Z -Erwartungswertfunktion	$E(Z(\mathbf{x})) := m(\mathbf{x})$
Varianzfunktion	$\text{Var}(Z(\mathbf{x})) := E(Z^2(\mathbf{x})) - m^2(\mathbf{x})$
Zwei-Punkte Z -Kovarianzfunktion	$\text{Cov}(Z(\mathbf{x}_\alpha), Z(\mathbf{x}_\beta)) := E(Z(\mathbf{x}_\alpha)Z(\mathbf{x}_\beta)) - m(\mathbf{x}_\alpha)m(\mathbf{x}_\beta)$
Z -Variogramm ¹	$2\gamma(\mathbf{x}_\alpha, \mathbf{x}_\beta) := \text{Var}(Z(\mathbf{x}_\alpha) - Z(\mathbf{x}_\beta))$

¹ $2\gamma(\mathbf{x}_\alpha, \mathbf{x}_\beta) = E\left\{[(Z(\mathbf{x}_\alpha) - Z(\mathbf{x}_\beta))]^2\right\} - E(Z(\mathbf{x}_\alpha) - Z(\mathbf{x}_\beta))^2$
 $= E\left\{(Z(\mathbf{x}_\alpha) - Z(\mathbf{x}_\beta))^2\right\} - E^2\left\{Z(\mathbf{x}_\alpha) - Z(\mathbf{x}_\beta)\right\}$
 $= \quad \quad \quad \text{I} \quad \quad \quad + \quad \quad \quad \text{II}$

manche Autoren betrachten nur den Term I der Definition des Semivariogramms, das heißt, betrachten nur den Fall II = 0

Sei $\mathbf{Z}(\mathbf{x})$, $\mathbf{x} \in D \subset \mathbb{R}^p$, $p \geq 2$, eine multivariate Zufallsfunktion. Dann definiert man (unter der Voraussetzung, daß alle Erwartungswerte und Varianzen existieren)

$$\text{Rand-Verteilungsfunktionen} \quad F_i(\mathbf{x}; z) := \text{Prob}(Z_i(\mathbf{x}) \leq z)$$

$$\text{Zwei-Punkte Verteilungsfunktionen} \quad F_{ij}(\mathbf{x}_\alpha, \mathbf{x}_\beta; z_1, z_2) := \text{Prob}(Z_i(\mathbf{x}_\alpha) \leq z_1, Z_j(\mathbf{x}_\beta) \leq z_2)$$

$$Z\text{-Erwartungswertfunktionen} \quad E(Z_i(\mathbf{x})) := m_i(\mathbf{x})$$

Kreuz-Kovarianzfunktionen

$$\text{Cov}(Z_i(\mathbf{x}_\alpha), Z_j(\mathbf{x}_\beta)) := E([(Z_i(\mathbf{x}_\alpha) - m_i(\mathbf{x}_\alpha))(Z_j(\mathbf{x}_\beta) - m_j(\mathbf{x}_\beta))])$$

$$\begin{aligned} \text{Kreuz-Variogramme} \quad 2\gamma_{ij}(\mathbf{x}_\alpha, \mathbf{x}_\beta) &:= \text{Cov}([(Z_i(\mathbf{x}_\alpha) - Z_i(\mathbf{x}_\beta)), (Z_j(\mathbf{x}_\alpha) - Z_j(\mathbf{x}_\beta))]) \\ &= E([(Z_i(\mathbf{x}_\alpha) - Z_i(\mathbf{x}_\beta))(Z_j(\mathbf{x}_\alpha) - Z_j(\mathbf{x}_\beta))]) \end{aligned}$$

Es gilt

$$2\gamma(\mathbf{x}_\alpha, \mathbf{x}_\beta) = \text{Var}(Z(\mathbf{x}_\alpha)) + \text{Var}(Z(\mathbf{x}_\beta)) - 2\text{Cov}(Z(\mathbf{x}_\alpha), Z(\mathbf{x}_\beta))$$

Falls gilt

$$\begin{aligned} \text{Var}(Z(\mathbf{x}_\alpha)) &= \text{Var}(Z(\mathbf{x}_\beta)) = C(0) \\ \text{und} \quad \text{Cov}(Z(\mathbf{x}_\alpha), Z(\mathbf{x}_\beta)) &= C(\mathbf{x}_\alpha, \mathbf{x}_\beta) \end{aligned}$$

dann folgt

$$\gamma(\mathbf{x}_\alpha, \mathbf{x}_\beta) = C(0) - C(\mathbf{x}_\alpha, \mathbf{x}_\beta)$$

wobei $C(0)$ dem Schwellenwert eines Semivariogramms entspricht.

5.2.2 Statistik zufälliger Funktionen

Bei der Entwicklung der Theorie der zufälligen Funktionen zu einem anwendbaren Modell sieht man sich in dieser Situation einer weiteren Schwierigkeit gegenüber: Es ist nur eine einzige Realisation $z(\mathbf{x}_\alpha)$, $\alpha = 1, \dots, n$, nämlich die beobachteten Erzgehalte der am Ort $\mathbf{x}_\alpha \in D$ genommenen Probekörper, der zufälligen Funktion $Z(\mathbf{x})$ gegeben, das heißt, von der zur zufälligen Funktion $Z(\mathbf{x})$ gehörenden Zufallsvariablen $Z(\mathbf{x}_\alpha)$ ist nur eine einzige Realisation $z(\mathbf{x}_\alpha)$ gegeben. Ebenso wie man aus einer einzigen Realisation einer einzelnen Zufallsvariablen (in der univariaten Statistik) keine statistischen Rückschlüsse auf ihre Eigenschaften ziehen kann, kann man i.a. aus einer einzigen (diskreten!) Realisation einer Zufallsfunktion nicht auf ihre Eigenschaften schließen.

Im folgenden geht es also darum, Annahmen für das Modell der zufälligen Funktion zu entwickeln, damit aus einer schlüssigen Theorie ein anwendungsorientierter Kalkül wird. Es geht also darum, eine Zufallsfunktion, mit der ein Phänomen beschrieben werden soll, mit solchen “angenehmen” Eigenschaften zu versehen, daß man mit diesem Modell rechnen kann, das heißt, daß man auf der Grundlage einer einzigen Realisation Statistik betreiben kann. Es wird also für Zufallsfunktionen das mathematische Analogon zur Annahme der unabhängigen und identischen Verteilung der klassischen Statistik gesucht.

Der Zufallsfunktion $Z(\mathbf{x})$, von der die beobachteten Erzgehalte $z(\mathbf{x}_\alpha)$ der Probenkörper am Ort \mathbf{x}_α als eine Realisation aufgefaßt werden, werden a priori mathematisch “angenehme” Eigenschaften zugestanden. Man sucht zu den empirischen Daten und Kenntnissen ein Modell mit a priori angenehmen mathematischen Eigenschaften (und nicht irgendein Modell), das das ganze Phänomen hinreichend gut beschreibt. Deshalb muß wie in der klassischen Statistik außermathematisch festgestellt werden, daß das zu beschreibende Phänomen und das mit den künstlich eingeführten angenehmen Eigenschaften versehene Modell verträglich sind.

Diese angenehme Eigenschaft, die man dem Modell zuschreibt, um das Rechnen zu ermöglichen, besteht im wesentlichen darin, die Differenzen der Erzgehalte, deren Probekörper die gleiche (gerichtete) Entfernung von einander haben, $z(\mathbf{x}_\alpha + \mathbf{h}) - z(\mathbf{x}_\alpha)$ und $z(\mathbf{x}_j + \mathbf{h}) - z(\mathbf{x}_j)$, als unabhängig vom Ort jedes einzelnen beteiligten Probekörpers, als unabhängige Realisationen einer neuen Zufallsvariablen $\Delta(\mathbf{h})$, der Variablen der Zuwächse, aufzufassen.

Jetzt kann man mit dem arithmetischen Mittel der Differenzen der Erzgehalte von Probekörpern, die durch die Schrittweite \mathbf{h} von einander entfernt sind, nach klassischen statistischen Methoden den Erwartungswert von $\Delta(\mathbf{h}) = Z(\mathbf{x}_\alpha + \mathbf{h}) - Z(\mathbf{x}_\alpha)$, $\mathbf{x}_\alpha + \mathbf{h}, \mathbf{x}_\alpha \in D$ schätzen.

Da die Güte der Schätzung von der Anzahl der Realisationen abhängt, kann man schon jetzt sagen, daß die Schätzung des Erwartungswerts der Zuwächse für kurze Entfernung besser als für weite Entfernung sein wird.

Diese Annahme wird als Stationaritätsannahme bezeichnet; der Begriff ist der Zeitreihenanalyse entlehnt. Zutreffender wäre es, sie als Homogenitätsannahme zu bezeichnen. Diese Annahme bedarf der mathematischen Definition, wobei es verschiedene Varianten zu unterscheiden gilt. Es wird sich zeigen, daß diese Annahme (in der einen oder anderen Variante) geeignet ist, das Phänomen “Lagerstätte” so hinreichend gut zu beschreiben, wie es i.a. von der Praxis verlangt wird.

Im folgenden werden drei Varianten von Stationarität vorgestellt, die in unterschiedlicher mathematischer Form die heuristische Vorgabe bezüglich der Zuwächse beschreiben. Die mathematischen Formulierungen unterscheiden sich durch die Strenge ihrer Bedingungen, die sie einer Zufallsfunktion auferlegen.

5.2.3 Modellannahme der strengen Stationarität

Die erste Annahme der strengen Stationarität wird bezüglich einer Eigenschaft der endlich dimensionalen gemeinsamen Verteilungen definiert, aus der die gewünschte Eigenschaft der Zuwächse gefolgert werden kann. Die Aussagen, die die Momente (Erwartungswert, Varianz,

etc.) betreffen, gelten, falls diese Momente existieren. Gegenüber der klassischen Statistik wird die Annahme der Unabhängigkeit aufgegeben.

Eine Zufallsfunktion $Z(\mathbf{x})$, $\mathbf{x} \in D$, heißt streng stationär, falls alle endlichdimensionalen gemeinsamen Verteilungen translationsinvariant sind, das heißt

$$\begin{aligned} P(Z(\mathbf{x}_\alpha) \leq z_\alpha, \dots, Z(\mathbf{x}_{i+k}) \leq z_{i+k}) &= P(Z(\mathbf{x}_\alpha + \mathbf{h}) \leq z_\alpha, \dots, Z(\mathbf{x}_{i+k} + \mathbf{h}) \leq z_{i+k}) \\ &\quad \text{für alle } k \in \mathbb{N} \\ &\quad \text{für alle } \mathbf{x}_\alpha, \mathbf{x}_\alpha + \mathbf{h} \in D \end{aligned}$$

Aus der strengen Stationarität einer Zufallsfunktion folgt

- (i) $P(Z(\mathbf{x}_\alpha) \leq z) = P(Z(\mathbf{x}_\beta) \leq z)$, das heißt, alle eindimensionalen Marginalverteilungen sind identisch;
- (ii) falls $E(Z(\mathbf{x}))$ existiert, gilt $E(Z(\mathbf{x})) = m$ für alle $\mathbf{x} \in D$
- (iii) falls $\text{Cov}(Z(\mathbf{x}_\alpha), Z(\mathbf{x}_\beta))$ existiert, gilt

$$\text{Cov}(Z(\mathbf{x}_\alpha), Z(\mathbf{x}_\beta)) = C(\mathbf{x}_\alpha - \mathbf{x}_\beta)$$
- (iv) $\text{Var}(Z(\mathbf{x}_\alpha) - Z(\mathbf{x}_\beta)) = E\left\{(Z(\mathbf{x}_\alpha) - Z(\mathbf{x}_\beta))^2\right\}$
- (v) $\gamma(\mathbf{x}_\alpha, \mathbf{x}_\beta) = 2[C(0) - C(\mathbf{x}_\alpha - \mathbf{x}_\beta)]$
mit $C(0) = \text{Cov}(Z(\mathbf{x}), Z(\mathbf{x})) = \text{Var}(Z(\mathbf{x}))$

das heißt, die Erwartungswertfunktion ist auf ganz D konstant, die Kovarianz zweier Zufallsvariablen der Zufallsfunktion ist nur eine Funktion ihrer Distanz (und deren Richtung) und die Varianz der Inkremente $\Delta(\mathbf{x}_\alpha, \mathbf{x}_\beta) = Z(\mathbf{x}_\alpha) - Z(\mathbf{x}_\beta)$ ist translationsinvariant.

5.2.4 Modellannahme der schwachen Stationarität, Stationarität zweiter Ordnung

Im nächsten Schritt löst man sich von der strengen Bedingung an die endlichdimensionalen Verteilungen und definiert Stationarität bezüglich der Momente, deren Existenz dann also vorausgesetzt sein muß. Gegenüber der klassischen Statistik sind jetzt beide Annahmen (der unabhängigen und identischen Verteilungen) aufgegeben.

Eine Zufallsfunktion $Z(\mathbf{x})$, $\mathbf{x} \in D$, heißt schwach stationär, falls

$$\begin{aligned} E(Z^2(\mathbf{x})) &\leq \infty \\ E(Z(\mathbf{x})) &= m \quad \text{für alle } \mathbf{x} \in D \\ \text{Cov}(Z(\mathbf{x}_\alpha), Z(\mathbf{x}_\beta)) &= C(\mathbf{x}_\alpha - \mathbf{x}_\beta) \quad \mathbf{x}_\alpha, \mathbf{x}_\beta \in D \end{aligned}$$

das heißt, falls alle Momente zweiter Ordnung existieren, die Erwartungswertfunktion auf ganz D konstant ist und die Kovarianzfunktion zweier Zufallsvariablen nur eine Funktion ihrer Distanz (und deren Richtung) ist.

Aus der schwachen Stationarität einer Zufallsfunktion folgt

$$\begin{aligned} \text{(i)} \quad E(Z(\mathbf{x}_\alpha) - Z(\mathbf{x}_\beta)) &= 0 \\ \text{Var}(Z(\mathbf{x})) &= C(0) \end{aligned}$$

$$\text{(ii)} \quad \text{Var}(Z(\mathbf{x}_\alpha) - Z(\mathbf{x}_\beta)) = 2[C(0) - C(\mathbf{x}_\alpha - \mathbf{x}_\beta)]$$

das heißt, die Erwartungswertfunktion der Inkremente verschwindet und ihre Varianzfunktion ist translationsinvariant auf ganz D .

5.2.5 Modellannahme der intrinsischen Stationarität, Stationarität zweiter Ordnung der Inkremente

Im letzten Schritt löst man sich von der Bedingung der Existenz der Momente und formuliert direkt für die Objekte des eigentlichen Interesses, nämlich die Zuwächse, bzw. die ersten zwei Momente der Zuwächse.

Eine Zufallsfunktion $Z(\mathbf{x})$, $\mathbf{x} \in D$, genügt der intrinsischen Hypothese (schwache Stationarität ihrer Zuwächse), falls

$$\begin{aligned} E(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})) &= m(\mathbf{h}) \quad \text{für alle } \mathbf{x}, \mathbf{x} + \mathbf{h} \in D \\ \text{Var}(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})) &= 2\gamma(\mathbf{h}) \quad \text{für alle } \mathbf{x}, \mathbf{x} + \mathbf{h} \in D \end{aligned}$$

das heißt, falls die Erwartungswertfunktion und die Varianzfunktion der Inkremente auf ganz D nur eine Funktion der entsprechenden Distanzen (und ihrer Richtungen) ist. Ohne Beschränkung der Allgemeinheit kann man $m(\mathbf{h}) = 0$ annehmen.

Journel, A.G., Fundamentals of Geostatistics in Five Lessons: AGU (1989), p. 8

Random function

In any of the previous discussions, the two RV (random variables) X, Y can represent

- (i) either two different attributes measured at the same location, say porosity and permeability measured from the same core plug;
- (ii) the same attribute measured at two different locations in space, say, porosity at locations \mathbf{x} and $\mathbf{x} + \mathbf{h}$ distant of a vector \mathbf{h} : $X = Z(\mathbf{x})$, $Y = Z(\mathbf{x} + \mathbf{h})$ or
- (iii) two different attributes measured at two different locations, say, porosity at location \mathbf{x} and permeability at vector \mathbf{h} apart: $X = \phi(\mathbf{x})$, $Y = K(\mathbf{x} + \mathbf{h})$.

In all cases, the semivariogram γ_{XY} or the correlation ρ_{XY} would measure the degree of variability/similarity between the two variables X, Y .

The second case (ii) is of particular interest for spatial interpolation problems, where a whole field of a given attribute, $Z(\mathbf{x}), \mathbf{x} \in$ site A , has to be inferred (mapped) from a limited number of locations sampled for the same attribute.

Pooling all $n(\mathbf{h})$ pairs of data found on attribute Z over the same site/zone/layer A , these pairs being approximately distant of the same vector \mathbf{h} (in length and direction), one could estimate the variogram characteristic of spatial variability over A :

$$\gamma_{Z(\mathbf{x}), Z(\mathbf{x}+\mathbf{h})} = \hat{\gamma}_A(\mathbf{h}) = \frac{1}{n(\mathbf{h})} \sum_{\alpha=1}^{n(\mathbf{h})} [z(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha + \mathbf{h})]^2 \quad (5.29)$$

By varying the vector \mathbf{h} (in length and direction), that characteristic is made into a vectorial function $\gamma(\mathbf{h})$.

The experimental semi-variogram $\hat{\gamma}_A(\mathbf{h})$ is not the estimate of some elusive expected value of squared increments of a no less elusive set of random variables $Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h})$. It is the discrete estimate of a well-defined spatial integral defining an average over A :

$$\gamma_A(\mathbf{h}) = \frac{1}{A(\mathbf{h})} \int_{A \cap A_{-\mathbf{h}}} [z(\mathbf{x}) - z(\mathbf{x} + \mathbf{h})]^2 d\mathbf{x} \quad (5.30)$$

with: $A \cap A_{-\mathbf{h}}$ being the intersection of A and its translate $A_{-\mathbf{h}}$ by vector $-\mathbf{h}$. If $\mathbf{x} \in A \cap A_{-\mathbf{h}}$, the both \mathbf{x} and $\mathbf{x} + \mathbf{h}$ are in A . $A(\mathbf{h})$ is the measure (area/volume) of that intersection.

Just as the random variable $Z(\mathbf{x})$ and its distribution characterize the uncertainty about the attribute value at location \mathbf{x} , the random function $Z(\mathbf{x}), \mathbf{x} \in A$, defined as a set of dependent RV's, will characterize the joint spatial uncertainty over A . The variogram $2\gamma(\mathbf{h})$ of that random function (RF) characterizes the degree of spatial variability between any two RV's $Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h})$ distant of a vector \mathbf{h} . The variogram $2\gamma(\mathbf{h})$ is modeled after the spatial average $2\gamma_A(\mathbf{h})$ which can be estimated by the discrete sum $2\hat{\gamma}_A(\mathbf{h})$ defined in (5.29).

The modeling of $\gamma(\mathbf{h})$ after the spatial average $\gamma_A(\mathbf{h})$ amount to decide to average statistics over a given area/site/zone/layer A . That decision is necessarily subjective, usually based on geological criteria, and cannot be proven (or refuted) by statistical tests. Once the decision of modeling $\gamma(\mathbf{h})$ after $\gamma_A(\mathbf{h})$ is taken, $\gamma(\mathbf{h})$ just like $\gamma_A(\mathbf{h})$ is not any more location (\mathbf{x}) dependent within A :

$$2\gamma(\mathbf{h}) = E\{[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})]^2\} \quad \text{is independent of } \mathbf{x} \in A \quad (5.31)$$

The moment $\gamma(\mathbf{h})$ is said to be stationary. Stationarity is a model decision, not some intrinsic property of the actual distribution of z -values over A .

Once a model is obtained, the corresponding covariance model can be deduced by the classical relation:

$$C(\mathbf{h}) = \text{Constant} - \gamma(\mathbf{h}) \quad (5.32)$$

Indeed:

$$\begin{aligned}
 2\gamma(\mathbf{h}) &= E\{[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})]^2\} = \\
 [E\{Z^2(\mathbf{x})\} - m^2] &+ [E\{Z^2(\mathbf{x} + \mathbf{h})\} - m^2] - 2[E\{Z(\mathbf{x})Z(\mathbf{x} + \mathbf{h})\} - m^2] = \\
 \text{Var}\{Z(\mathbf{x})\} &+ \text{Var}\{Z(\mathbf{x} + \mathbf{h})\} - 2\text{Cov}\{Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h})\} = \\
 2[C(0) - C(\mathbf{h})], \text{ thus: } C(\mathbf{h}) &= C(0) - \gamma(\mathbf{h})
 \end{aligned}$$

The constant of formula (5.32) can either be set equal to the sill $\gamma(\infty)$ of the semivariogram *model* if it exists, or can be set to any arbitrary large value. That arbitrary constant should then be filtered out from all further algorithms, as will be the case in ordinary kriging, see Lesson III, System (4).

5.2.6 Diskussion der Stationaritätsbegriffe

Alle drei bisher eingeführten Stationaritätsannahmen sind verschiedene mathematische Gegenstücke ein und derselben außermathematischen Erfahrung, der räumlichen Wiederholung eines Phänomens; siehe dazu Benoit Mandelbrot, The Fractal Geometry of Nature: Freeman, San Francisco.

5.2.7 Zusammenhang zwischen Kovarianzfunktion und Semivariogramm

Für eine schwach stationäre Zufallsfunktion $Z(\mathbf{x}), \mathbf{x} \in D$, gilt²

$$\begin{aligned}
 \text{Var}(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})) &= E\left(\left[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})\right]^2\right) \\
 &= 2[C(0) - C(\mathbf{h})] = 2\gamma(\mathbf{h})
 \end{aligned}$$

das heißt, für schwachstationäre Zufallsfunktionen sind die Kovarianzfunktion und das Variogramm äquivalente Hilfsmittel.

$$\begin{aligned}
 ^2 \text{Var}(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})) &= E\left(\left[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})\right]^2\right) - E^2(Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})) \\
 &= 0 \quad \text{falls schwach stationär}
 \end{aligned}$$

$$\begin{aligned}
 \text{Var}(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})) &= \text{Var}(Z(\mathbf{x} + \mathbf{h})) + \text{Var}(Z(\mathbf{x})) - 2\text{Cov}(Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})) \\
 &= \text{Cov}(Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x} + \mathbf{h})) + \text{Cov}(Z(\mathbf{x}), Z(\mathbf{x})) - 2\text{Cov}(Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})) \\
 &= 2[C(0) - C(\mathbf{h})] \\
 &= 2\gamma(\mathbf{h})
 \end{aligned}$$

Für die Kovarianzfunktion C gilt

$$|C(\mathbf{h})| \leq C(0)$$

und für das Semivariogramm

$$\lim_{\mathbf{h} \rightarrow \infty} \frac{\gamma(\mathbf{h})}{\|\mathbf{h}\|^2} = 0 ,$$

das heißt, das Semivariogramm wächst schließlich langsamer als h^2 . Dies bedeutet für die Praxis, daß ein experimentelles Semivariogramm, das für große h mindestens wie h^2 , also schneller als parabolisch, wächst, anzeigt, daß das untersuchte Phänomen im gewählten Maßstab nicht durch eine intrinsische Zufallsfunktion modelliert werden kann, das heißt, daß eine Drift- (Trend-) Komponente vorhanden ist.

Weiterhin, falls

$$\lim_{\mathbf{h} \rightarrow \infty} C(\mathbf{h}) = 0$$

dann gilt

$$\lim_{\mathbf{h} \rightarrow \infty} \gamma(\mathbf{h}) = C(0)$$

Haben alle experimentellen Richtungs-Semivariogramme einen Schwellenwert, dann spricht dieses Verhalten für Stationarität.

5.2.8 Elementare Regeln für das Rechnen mit Erwartungswert- und Varianzfunktion

1. Sei Z^* eine endliche Linearkombination von Zufallsvariablen $Z(\mathbf{x}_\alpha)$ einer Zufallsfunktion $Z(\mathbf{x})$, $\mathbf{x} \in D$, das heißt

$$Z^* = \sum_{\alpha} \lambda_{\alpha} Z(\mathbf{x}_{\alpha}) , \quad \lambda_{\alpha} \in IR$$

dann gilt

$$\begin{aligned} 0 \leq \text{Var}(Z^*) &= \text{Var}\left(\sum_{\alpha} \lambda_{\alpha} Z(\mathbf{x}_{\alpha})\right) \\ &= \text{Cov}\left(\sum_{\alpha} \lambda_{\alpha} Z(\mathbf{x}_{\alpha}), \sum_{\beta} \lambda_{\beta} Z(\mathbf{x}_{\beta})\right) \\ &= \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} \text{Cov}(Z(\mathbf{x}_{\alpha}), Z(\mathbf{x}_{\beta})) \\ &= \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) = \bar{C}(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) \end{aligned}$$

Unter der Annahme der intrinsischen Stationarität existieren im allgemeinen die ersten beiden Momente der Inkremente, das heißt insbesondere das Semivariogramm, aber nicht die Kovarianzfunktion. In diesem Fall besitzen nur solche Linearkombinationen von Zufallsvariablen $\sum \lambda_\alpha Z_\alpha$ eine endliche Varianz, für die $\sum \lambda_\alpha = 0$ gilt

$$\text{Var}(Z^*) = - \sum_{\alpha} \sum_{\beta} \lambda_\alpha \lambda_\beta \gamma(\mathbf{x}_\alpha - \mathbf{x}_\beta) = -\bar{\gamma}(\mathbf{x}_\alpha - \mathbf{x}_\beta) \text{ falls } \sum \lambda_\alpha = 0$$

Da die Varianz nicht negativ sein kann, gilt dann wie vorher auch in diesem Fall

$$\text{Var}(Z^*) = -\bar{\gamma}(\mathbf{x}_\alpha - \mathbf{x}_\beta) \geq 0$$

das heißt $-\bar{\gamma}(\mathbf{x}_\alpha - \mathbf{x}_\beta)$ ist eine nicht negative Zahl.

Man braucht also zur Berechnung der Varianz im Falle $\sum \lambda_\alpha = 0$ ganz formal nur die Kovarianzfunktion C durch das Semivariogramm $-\gamma$ zu ersetzen.

2. Sei Z^* der Mittelwert einer Zufallsfunktion $Z(\mathbf{x})$, $\mathbf{x} \in D$, über einer Teilmenge $V \subset D$, das heißt

$$\begin{aligned} Z^* &= \frac{1}{|V|} \int_V Z(\mathbf{x}) d\mathbf{x} \\ |V| &= \int_V d\mathbf{x} = \text{vol}_p(V) \end{aligned}$$

Dann gilt

$$\begin{aligned} \text{Var}(Z^*) &= \text{Var}\left(\frac{1}{|V|} \int_V Z(\mathbf{x}) d\mathbf{x}\right) \\ &= \frac{1}{|V|^2} \text{Cov}\left(\int_V Z(\mathbf{x}_\alpha) d\mathbf{x}_\alpha, \int_V Z(\mathbf{x}_\beta) d\mathbf{x}_\beta\right) \\ &= \frac{1}{|V|^2} \int_V \int_V \text{Cov}(Z(\mathbf{x}_\alpha), Z(\mathbf{x}_\beta)) d\mathbf{x}_\alpha d\mathbf{x}_\beta \\ &= \frac{1}{|V|^2} \int_V \int_V K(\mathbf{x}_\alpha - \mathbf{x}_\beta) d\mathbf{x}_\alpha d\mathbf{x}_\beta = \bar{C}(V, V) \\ \text{Var}(Z^*) &= -\frac{1}{|V|^2} \int_V \int_V \gamma(\mathbf{x}_\alpha - \mathbf{x}_\beta) d\mathbf{x}_\alpha d\mathbf{x}_\beta = -\bar{\gamma}(V, V) \end{aligned}$$

Chapter 6

Der JURA–Datensatz in der Zusammenschau

An 359 flächenhaft unregelmäßig verteilten Orten im Untersuchungsgebiet wurden an oberflächennahen Bodenproben 7 Schwermetall-Konzentrationen gemessen und Gesteinstyp und Flächennutzung als nominelle (kategoriale) Merkmale erfaßt.

Die geometrische Anordnung der 359 Orte ist das Resultat der Kombination einer regelmäßigen (regular) und einer geschachtelten (nested) Anordnung (configuration). Die regelmäßige Anordnung besteht aus einem quadratischen Raster mit 107 Rasterpunkten im Rasterabstand von 250m. Von diesen 107 Rasterpunkten wurden 38 zufällig ausgewählt. Die zufällige Auswahl hat zur Folge, daß der Anteil der 38 zufällig ausgewählten Rasterpunkte, der zum Gesteinstyp k , $k = 1, \dots, 5$, bzw zur Flächennutzung k' , $k' = 1, \dots, 4$, gehört, dem Anteil des Gesteinstyp k , bzw. der Flächennutzung k' an der Gesamtfläche des Untersuchungsgebiets entspricht.

Ausgehend von jedem dieser 38 zufällig ausgewählten Rasterpunkte wurde ein erster zusätzlicher Punkt im Abstand von 100m, ein zweiter zusätzlicher im Abstand von 40m vom ersten, ein dritter im Abstand von 16m vom zweiten und ein vierter im Abstand von 6m vom dritten gewählt; dabei wurden die Richtungen von einem zum nächsten Punkt zufällig gewählt.

Ziel dieser Anordnung ist ein möglichst gutes flächenhaftes Auflösungsvermögen über Distanzen zwischen 0 und 250m.

Zusammenfassung der Ergebnisse der erkundenden Daten Analyse

1. Die **empirischen Verteilungsfunktionen** der Schwermetall-Konzentrationen zeigen an, daß viele Proben mit Cd und Pb kontaminiert sind, also Konzentrationen über dem gesetzlich festgesetzten größten zulässigen Wert besitzen, und daß der Anteil der Proben mit einer Cu Konzentration über dem größten zulässigen Wert kleiner ist.

2. Die **Histogramme** der Schwermetall-Konzentrationen zeigen an, daß die Verteilungen von Cd, Cu und Pb rechtsschief (positiv schiefl) sind.
3. Die **bedingten Verteilungen** der Schwermetall-Konzentrationen bezogen auf einen bestimmten Gesteinstyp bzw. eine bestimmte Flächennutzung und die **Kenntnisse dieser bedingten Verteilungen** zeigen an, daß die (im Mittel) kleinsten Schwermetall-Konzentrationen im Waldboden oder auf Argovian Gestein und daß für Boden unter Weideland (pasture) die (im Mittel) größten Konzentrationen für alle Schwermetalle gemessen wurden.
4. Die Anlaysse der **linearen und Rang-Korrelationskoeffizienten** zeigt an, daß die kontaminierenden Cd, Cu und Pb Konzentrationen signifikant positiv mit der Zn Konzentration, die besser beprobt ist, korrelieren.
5. Die **richtungsabhängigen Semivariogramme der Schwermetall-Konzentrationen** zeigen die Existenz von zwei räumlichen Strukturen der Variabilität der Schwermetall-Konzentrationen, eine mit einer kurzen Reichweite (ca. 200m) und eine mit einer langen Reichweite (ca. 1000m), an.

Die kurz-reichweitige Struktur ist dominierend für die kontaminierenden Cd, Cu und Pb Konzentrationen und für die Cr Konzentration, die lang-reichweitige Struktur ist dominierend für die Ni und Co Konzentrationen. Die richtungsabhängigen Semivariogramme der Zn Konzentration zeigen eine Kombination beider Strukturen zu etwa gleichen Anteilen an.

6. Die **Indikator-Semivariogramme des Gesteinstyps und der Flächennutzung** zeigen an, daß die kurz-reichweitige Struktur der Schwermetall-Konzentrationen der räumlichen Verteilung von Gesteinstyp und Flächennutzung und daß die lang-reichweitige Struktur der besonderen Abhängigkeit der Schwermetall-Konzentrationen von Argovian und Kimmeridgian Gestein entspricht.
7. Die **richtungsabhängigen Semivariogramme der Co und Ni Konzentrationen** zeigen für die SW-NE Richtung einen größeren räumlichen Zusammenhang, eine größere räumliche Ähnlichkeit oder Stetigkeit, an, die der Vorzugsorientierung des flächenhaften Verteilungsmusters der Gesteintypen Argovian und Kimmeridgian entspricht.

Die Richtungssemivariogramme der anderen Schwermetall-Konzentrationen zeigen keine flächenhafte Anisotropie an.

8. Die **Indikator-Semivariogramme der Cd, Ni und Zn Konzentrationen für verschiedene Schranken (thresholds)** zeigen für kleine Konzentrationen einen größeren räumlichen Zusammenhang als für große Konzentrationen an. Diese Beobachtung läßt die Existenz homogener Gebiete mit kleinen Konzentrationen und (größerer) Zonen mit mittleren und großen Konzentrationen vermuten.

9. Die **Kreuz–Semivariogramme der Schwermetall-Konzentrationen** zeigen eine kurz-reichweitige Abhangigkeit der Cd, Cu und Pb mit den Zn Konzentrationen und eine lang-reichweitige Abhangigkeit der Cd mit den Ni Konzentrationen an.

Chapter 7

Schließende Daten Analyse und Modellierung

7.1 Probenahme

Damit das Modell der Zufallsfunktion für das Schätzproblem anwendbar wird, müssen ihre Parameter, insbesondere Erwartungswertfunktion und Kovarianzfunktion bzw. Semivariogramm, aus den Daten bestimmt werden, da diese in den Kriging–Schätzalgorithmus eingehen.

Erste Voraussetzung für gute Schätzungen der Parameter ist, daß die Daten flächenhaft flächenbezogen repräsentativ sind. Für viele Datensätze aus den Geowissenschaften ist deren flächenhafte Repräsentativität durchaus fragwürdig, da die Proben nicht gleichmäßig oder regelmäßig flächenhaft verteilt sondern bevorzugt genommen wurden.

Auch wenn bevorzugte Probenahme objektiv begründet sein kann, widerspricht sie der idealen Voraussetzung für statistisch begründete Schlußfolgerung und verfälscht diese.

Falls die Vorzugsprobenahme, z.B. clustering, dokumentiert ist, besteht die Möglichkeit, die der Statistik abträglichen Auswirkungen zu demonstrieren und zu korrigieren.

Eine Methode zur Korrektur von clustering besteht darin, empirische arithmetische Mittelwerte

$$\hat{m} = \frac{1}{n} \sum_{\alpha} z(\mathbf{x}_{\alpha})$$

durch gewichtete Mittelwerte

$$\hat{m} = \frac{1}{|A|} \sum_{\alpha} w_{\alpha} z(\mathbf{x}_{\alpha})$$

zu ersetzen wobei w_{α} der Flächeninhalt der Thiessen–Fläche zum Ort \mathbf{x}_{α} ist, sodaß der Einfluß von clustered data reduziert wird.

7.2 Schätzer für Semivariogramm und Kovarianzfunktion

Der klassische Semivariogrammschätzer

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha + \mathbf{h})]^2$$

kann durch den Proportionalitätseffekt beeinträchtigt sein. Der Proportionalitätseffekt ist eine besondere Form der Heteroskedastizität, bei der die lokale Varianz für rechtsschiefe Verteilung proportional, für linksschiefe umgekehrt proportional zum lokalen Mittelwert ist.

Den Proportionalitätseffekt kann man durch scatter plots der lokalen Varianzen gegen lokale Mittelwerte auf der Grundlage von moving window statistics entdecken.

Falls die Mittelwerte und Varianzen zur Schrittweite \mathbf{h} als Funktion von \mathbf{h} das gleiche Verhalten zeigen, wird zusätzlich das relative Semivariogramm

$$\hat{\gamma}_{GR}(\mathbf{h}) = \frac{\hat{\gamma}(\mathbf{h})}{f(\hat{m}(\mathbf{h}))}$$

berechnet mit

$$\hat{m}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} z(\mathbf{x}_\alpha) + z(\mathbf{x}_\alpha + \mathbf{h}) = \frac{m_{-\mathbf{h}} + m_{+\mathbf{h}}}{2}$$

Für rechtsschiefe Verteilungen nimmt man gewöhnlich

$$f(\hat{m}(\mathbf{h})) = \hat{m}^2(\mathbf{h})$$

Ein guter Schätzer für die Kovarianzfunktion ist

$$\hat{C}(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} z(\mathbf{x}_\alpha)z(\mathbf{x}_\alpha + \mathbf{h}) - \hat{m}_{-\mathbf{h}}\hat{m}_{+\mathbf{h}}$$

7.3 Modelle für Semivariogramm und Kovarianzfunktionen

Wegen der mathematischen Eigenschaften einer Kovarianzfunktion

$$\sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) \geq 0$$

bzw. eines Semivariogramms

$$C(0) \sum_{\alpha} \lambda_{\alpha} \sum_{\beta} \lambda_{\beta} - \sum_{\alpha} \sum_{\beta} \lambda_{\alpha} \lambda_{\beta} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) \geq 0$$

stehen zur Modellierung der geschätzten Kovarianzfunktion nur positiv definite Funktionen (cf. Gneiting, T., Sasvári, Z., 1999, The characterization problem for isotropic covariance functions: Math.Geol. 31(1), 105-112) und zur Modellierung des geschätzten Variogramms nur bedingt negativ definite Funktionen zur Verfügung, wobei die Bedingung $\sum_{\alpha} \lambda_{\alpha} = 0$ ist.

Wackernagel, H., Multivariate Geostatistics: Springer (1995), pp. 37-40:

Positive definite function

A covariance function is a positive definite function¹. This means that the use of a covariance function $C(\mathbf{h})$ for the computation of the variance of a linear combination of $n + 1$ random variables $Z(\mathbf{x}_{\alpha})$ (any subset sampled from a second order random function) must be positive. It is necessarily linked to a positive semi-definite matrix C of covariances

$$\text{Var} \left(\sum_{\alpha=0}^n \lambda_{\alpha} Z(\mathbf{x}_{\alpha}) \right) = \sum_{\alpha=0}^n \sum_{\beta=0}^n \lambda_{\alpha} \lambda_{\beta} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) = \lambda^t C \lambda \geq 0$$

for any set of points \mathbf{x}_{α} and any set of weights λ_{α} (assembled into a vector λ).

The continuous covariance functions are characterized by Bochner's theorem as the Fourier transforms of positive measures. This topic is treated in more detail in the chapter on cross-covariance functions.

Conditionally negative definite function

The variogram is a *conditionally* negative definite function. The condition to guarantee the positivity of the variance of any linear combination of $n + 1$ random variables, subset of an intrinsic random function, is that the $n + 1$ weights λ_{α} sum up to zero. The variance of a linear combination of intrinsically stationary random variables is defined as

$$\text{Var} \left(\sum_{\alpha=0}^n \lambda_{\alpha} Z(\mathbf{x}_{\alpha}) \right) = - \sum_{\alpha=0}^n \sum_{\beta=0}^n \lambda_{\alpha} \lambda_{\beta} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) \geq 0 \quad \text{if } \sum_{\alpha=0}^n \lambda_{\alpha} = 0$$

i.e. any matrix Γ of variogram values is conditionally negative definite

$$[\lambda_{\alpha}]^t [\gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta})] [\lambda_{\alpha}] = \lambda^t \Gamma \lambda \geq 0 \quad \text{for } \sum_{\alpha=0}^n \lambda_{\alpha} = 0$$

To understand the meaning (sufficiency) of the condition on the sum of the weights, it is necessary to compute explicitly the variance of the linear combination.

¹For functions, in opposition to matrices, no distinction is usually made between “definite” and “semi-definite”.

As the random function $Z(\mathbf{x})$ is intrinsically stationary, only the expectation of its increments has a meaning. A trick enabling to construct increments is to insert an additional random variable $Z(\mathbf{0})$, placed arbitrarily at the origin $\mathbf{0}$ (assuming $\mathbf{0} \in \mathbf{D}$), multiplied by the zero sum weights

$$\begin{aligned}
 \text{Var} \left(\sum_{\alpha=0}^n \lambda_{\alpha} Z(\mathbf{x}_{\alpha}) \right) &= \text{Var} \left(-Z(\mathbf{0}) \cdot \underbrace{\sum_{\alpha=0}^n \lambda_{\alpha}}_0 + \sum_{\alpha=0}^n \lambda_{\alpha} Z(\mathbf{x}_{\alpha}) \right) \\
 &= \text{Var} \left[\sum_{\alpha=0}^n \lambda_{\alpha} (Z(\mathbf{x}_{\alpha}) - Z(\mathbf{0})) \right] \\
 &= \sum_{\alpha=0}^n \sum_{\beta=0}^n \lambda_{\alpha} \lambda_{\beta} \mathbb{E} \left[(Z(\mathbf{x}_{\alpha}) - Z(\mathbf{0})) \cdot (Z(\mathbf{x}_{\beta}) - Z(\mathbf{0})) \right] \\
 &= \sum_{\alpha=0}^n \sum_{\beta=0}^n \lambda_{\alpha} \lambda_{\beta} C_I(\mathbf{x}_{\alpha}, \mathbf{x}_{\beta})
 \end{aligned}$$

where $C_I(\mathbf{x}_{\alpha}, \mathbf{x}_{\beta})$ is the covariance of increments formed using the additional variable $Z(\mathbf{0})$.

We also introduce the additional variable $Z(\mathbf{0})$ into the expression of the variogram

$$\begin{aligned}
 \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) &= \frac{1}{2} \mathbb{E} \left[(Z(\mathbf{x}_{\alpha}) + Z(\mathbf{0}) - Z(\mathbf{x}_{\beta}) - Z(\mathbf{0}))^2 \right] \\
 &= \frac{1}{2} (2\gamma(\mathbf{x}_{\alpha}) + 2\gamma(\mathbf{x}_{\beta}) - 2C_I(\mathbf{x}_{\alpha}, \mathbf{x}_{\beta}))
 \end{aligned}$$

so that

$$C_I(\mathbf{x}_{\alpha}, \mathbf{x}_{\beta}) = \gamma(\mathbf{x}_{\alpha}) + \gamma(\mathbf{x}_{\beta}) - \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta})$$

which incorporates two non stationary terms \mathbf{x}_{α} and \mathbf{x}_{β} .

Coming back to the computation of the variance of the linear combination, we see, that the two non stationary terms are cancelled by the condition on the weights

$$\begin{aligned}
 \text{Var} \left(\sum_{\alpha=0}^n \lambda_{\alpha} Z(\mathbf{x}_{\alpha}) \right) &= \sum_{\alpha=0}^n \sum_{\beta=0}^n \lambda_{\alpha} \lambda_{\beta} C_I(\mathbf{x}_{\alpha}, \mathbf{x}_{\beta}) \\
 &= \underbrace{\sum_{\beta=0}^n \lambda_{\beta} \sum_{\alpha=0}^n \lambda_{\alpha} \gamma(\mathbf{x}_{\alpha})}_{0} + \underbrace{\sum_{\alpha=0}^n \lambda_{\alpha} \sum_{\beta=0}^n \lambda_{\beta} \gamma(\mathbf{x}_{\beta})}_{0} - \sum_{\alpha=0}^n \sum_{\beta=0}^n \lambda_{\alpha} \lambda_{\beta} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) \\
 &= - \sum_{\alpha=0}^n \sum_{\beta=0}^n \lambda_{\alpha} \lambda_{\beta} \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) \quad \text{for } \sum_{\alpha=0}^n \lambda_{\alpha} = \sum_{\beta=0}^n \lambda_{\beta} = 0
 \end{aligned}$$

When handling linear combinations of random variables, the variogram can only be used together with the conditions on the weights guaranteeing its existence. In particular, the variogram cannot in general be used in a simple kriging. Other forms of kriging, with constraints on the

weights, are required to use this tool that covers a wider range of phenomena than the covariance function.

Variograms can be characterized on the basis of continuous covariance functions: a function $\gamma(\mathbf{h})$ is a variogram, if for any positive constant c the exponential of $-c \gamma(\mathbf{h})$ is a covariance function

$$C(\mathbf{h}) = e^{-c \gamma(\mathbf{h})} \quad c > 0$$

This remarkable relation, based on a theorem by SCHOENBERG, links (through the defining kernel of the Laplace transform) the conditionally negative functions with the positive definite functions (see also CHOQUET).

The covariance function

$$C(\mathbf{h}) = \exp\left\{-\frac{|\mathbf{h}|^p}{a}\right\} \text{ with } 0 < p \leq 2 \text{ and } a > 0$$

which is related by Schoenberg's theorem to the power variogram model, defines the family of stable covariance functions. The case $p = 2$ (Gaussian covariance function) is pathological: it corresponds to a deterministic random function (MATHERON), which is contradictory with the randomness. The case $p = 1$ defines the exponential covariance function.

Modelle für Semivariogramme

Reiner Nugget-Effekt

$$g(h) = \begin{cases} 0, & \text{falls } h = 0 \\ 1, & \text{sonst} \end{cases}$$

Sphärisches Semivariogramm

$$g(h) = \begin{cases} 1.5 \frac{h}{a} - 0.5 \frac{h^3}{a^3}, & \text{falls } h < a \\ 1, & \text{sonst} \end{cases}$$

Wackernagel, H., Multivariate Geostatistics: Springer (1995), pp. 43-45:

Derivation of the spherical covariance

Imagine a universe with Poisson points, i.e. a 3D-space with points \mathbf{x}_P scattered randomly following a uniform distribution along each coordinate and summing up to θ points per volume unit on

average. A counting function $N(\mathcal{V})$ is defined which counts the number of Poisson points contained in a volume \mathcal{V} .

Consider the random function $Z(\mathbf{x}) = N(\mathcal{B}_x)$ which is the count of the number of Poisson points contained in a ball \mathcal{B} centered on a point \mathbf{x} . Clearly \mathcal{B}_x represents the volume of influence of diameter d around a point \mathbf{x} which determines the value of $Z(\mathbf{x})$. The problem at hand is to calculate the covariance function of the random function $Z(\mathbf{x})$.

An indicator function $\mathbf{1}_{\mathcal{B}}(\mathbf{x}')$ is constructed indicating whether a location \mathbf{x}' is inside a ball centered at \mathbf{x}

$$\mathbf{1}_{\mathcal{B}}(\mathbf{x}') = \begin{cases} 1, & \text{if } \mathbf{x}' \in \mathcal{B}_x \\ 0, & \text{if } \mathbf{x}' \notin \mathcal{B}_x \end{cases}$$

A function $\mathcal{K}(\mathbf{h})$, the *geometric covariogram*, measures the volume of the intersection of a ball \mathcal{B} with a copy \mathcal{B}_h of it translated by a vector \mathbf{h}

$$\begin{aligned} \mathcal{K}(\mathbf{h}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{1}_{\mathcal{B}}(\mathbf{x}') \mathbf{1}_{\mathcal{B}}(\mathbf{x}' + \mathbf{h}) d\mathbf{x}' = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{1}_{\mathcal{B}}(\mathbf{x}') \mathbf{1}_{\mathcal{B}_h}(\mathbf{x}') d\mathbf{x}' \\ &= |\mathcal{B} \cap \mathcal{B}_h| \end{aligned}$$

Conversely, it is worth noting that the intersection $\mathcal{B} \cap \mathcal{B}_{-h}$ of the ball with a copy of itself translated by $-\mathbf{h}$ represents the set of points $\mathbf{x}' \in \mathcal{B}$ which have a neighbor $\mathbf{x}' + \mathbf{h}$ within the ball[, as shown in figure 6.4.]

$$\mathcal{K}(\mathbf{h}) = \int_{\mathbf{x}' \in \mathcal{B} \cap \mathcal{B}_{-h}} d\mathbf{x}' = |\mathcal{B} \cap \mathcal{B}_{-h}|$$

The covariance of $Z(\mathbf{x})$ can now be expressed as

$$C(\mathbf{h}) = \mathbb{E}[N(\mathcal{B})N(\mathcal{B}_h)] - \mathbb{E}[N(\mathcal{B})]\mathbb{E}[N(\mathcal{B}_h)]$$

and as the counts $N(\mathbf{V})$ are independent in any subvolume

$$\begin{aligned} C(\mathbf{h}) &= \mathbb{E}[N(\mathcal{B} \cap \mathcal{B}_h)^2] - \mathbb{E}^2[N(\mathcal{B} \cap \mathcal{B}_h)] \\ &= \theta |\mathcal{B} \cap \mathcal{B}_h| \\ &= \theta \mathcal{K}(\mathbf{h}) \end{aligned}$$

Calculating explicitly the volume of the intersection of two spheres of equal size whose centers are separated by a vector \mathbf{h} yields the formula for the spherical covariance

$$C(\mathbf{h}) = \begin{cases} \theta |\mathcal{B}| \left(1 - \frac{3}{2} \frac{|\mathbf{h}|}{d} + \frac{1}{2} \frac{|\mathbf{h}|^3}{d^3}\right) & \text{for } 0 \leq |\mathbf{h}| \leq d \\ 0 & \text{for } |\mathbf{h}| > d \end{cases}$$

where $\theta|\mathcal{B}| = \theta\pi d^3/6 = C(\mathbf{0})$ represents the variance of $Z(\mathbf{x})$ and $|\mathcal{B}|$ is the volume of the spheres.

The diameter d of the spheres is equal to the range of the covariance function as it indicates the distance at which the covariance vanishes. The range of the spherical covariance function is the maximal distance at which the volumes of influence of two random variables $Z(\mathbf{x})$ and $Z(\mathbf{x} + \mathbf{h})$ can overlap and share information.

In applications large objects (as compared to the scale of the investigation) can condition the spatial structure of the data. The maximal size of these morphological objects in a given direction can often be read from the experimental variogram and interpreted as the range of a spherical model.

The shape of objects conditioning the morphology of a regionalized variable may not be spherical in many applications. This will result in anisotropic behavior of the variogram.

Exponentielles Semivariogramm

$$g(h) = 1 - \exp\left(\frac{-3h}{a}\right)$$

Gauß–Variogramm

$$g(h) = 1 - \exp\left(\frac{-3h^2}{a}\right)$$

Potenz–Variogramm

$$g(h) = h^\omega \quad 0 < \omega < 2$$

Für Semivariogramme mit Schwellenwert ist die entsprechende Kovarianz–Modellfunktion gegeben durch

$$c(\mathbf{h}) = 1 - \gamma(\mathbf{h})$$

Für unbeschränkte Semivariogramm–Modelle existiert keine entsprechende Kovarianz–Modellfunktion.

Im allgemeinen anisotropen Fall ist das Semivariogramm nicht nur vom Abstand $h = \|\mathbf{h}\|$, sondern auch von der Richtung \mathbf{h} abhängig.

Man unterscheidet in der Geostatistik

- geometrische Anisotropie, bei der die Richtungsvariogramme die gleiche Form und den gleichen Schwellenwert, aber verschiedene Reichweiten besitzen
- und zonale Anisotropie, bei der auch der Schwellenwert richtungsabhängig ist.

7.4 Variographie

Unter Variographie versteht man alle Aspekte der Semivariogramm- bzw. Kovarianzfunktion-Modellierung. Sie beruht auf mathematischen Eigenschaften des stochastischen Modells der Zufallsfunktion, bezieht aber ausdrücklich außermathematische Kenntnisse über das Phänomen ein. Diese Modellierung ist subjektiv Benutzer gesteuert; sie ist keine mathematisch-statistische Operation.

Grundlagen der Modellierung

- empirische Semivariogramm- bzw. Kovarianzfunktionswerte
- zulässige Modellfunktion - positive Linearkombinationen von Semivariogramm-Modellen (Schachtelung von Semivariogramm-Modellen)
- zusätzliche Kenntnisse

Benutzer-Entscheidungen

- isotropes oder anisotropes Modell
 - Undeutliche Anisotropie in den Richtungsvariogrammen ohne qualitative Begründung ist eher vernachlässigbar.
 - Schwerwiegende qualitative Kenntnisse können zur Anwendung eines anisotropen Modells führen, auch wenn die Anisotropie in den Daten nicht zu erkennen ist.
- Wieviele und welche Semivariogramm-Modelle?
 - so wählen, daß das Modell so einfach wie möglich wird
 - lineares Verhalten für kleine \mathbf{h} meistens geeignet
 - glatter als linear (für kleine \mathbf{h}), etwa parabolisch = Gauß, nur bei speziellen Problemen, z.B. Potentialflächen
- Welche Parameter-Werte?
 - Nugget-Effekt
 - * Ist die kleinste Schrittweite des empirischen Semivariogramms groß, wird der Nuggeteffekt leicht überschätzt.
 - * Im Falle von clustered data und/oder Proportionalitätseffekt zieht man relative Semivariogramme hinzu.
 - * Gauß-Modelle müssen mit einem Nugget-Effekt versehen werden.
 - * gewöhnlich als isotrop modelliert

- Schwellenwert

Der Schwellenwert sollte nicht gleich der empirischen Varianz gesetzt werden, cf. Barnes, R.J., 1991, The variogram sill and the sample variance: Math. Geol. 23(4), 673-678.

Das Variogramm ist allgemein definiert als

$$2\gamma(\mathbf{h}) = E\{[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})]^2\} \quad (7.1)$$

Falls die Schrittweite \mathbf{h} größer als die Reichweite ist, d.h. falls $Z(\mathbf{x})$ und $Z(\mathbf{x} + \mathbf{h})$ unkorreliert (nicht unabhängig !) sind, dann ist der Wert des Semi-Variogramms gleich der Varianz der Gesamtheit. Es gilt nämlich

$$\begin{aligned} 2\gamma(\mathbf{h}) &= E\{[Z(\mathbf{x}) - \mu - Z(\mathbf{x} + \mathbf{h}) + \mu]^2\} \\ &= E\{[Z(\mathbf{x}) - \mu]^2 - 2[Z(\mathbf{x}) - \mu][Z(\mathbf{x} + \mathbf{h}) - \mu] + [Z(\mathbf{x} + \mathbf{h}) - \mu]^2\} \\ &= E\{[Z(\mathbf{x}) - \mu]^2\} - 2E\{[Z(\mathbf{x}) - \mu][Z(\mathbf{x} + \mathbf{h}) - \mu]\} + E\{[Z(\mathbf{x} + \mathbf{h}) - \mu]^2\} \\ &= \sigma^2 - 2\text{Cov}[Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h})] + \sigma^2 \end{aligned}$$

also

$$\gamma(\mathbf{h} | \mathbf{h} > \text{Reichweite}) = \sigma^2 \quad (7.2)$$

Im Falle von realen Daten ist weder das Variogramm noch die Varianz (der Gesamtheit) bekannt. Ist in dieser Situation die Gleichung 7.2 im allgemeinen nützlich und hilfreich?

Die Stichprobe bestehe aus n Werten $Z_i = Z(\mathbf{x}_i), i = 1, \dots, n$ die an den Meßstellen $\mathbf{x}_i, i = 1, \dots, n$ gemessen worden seien. Deren Mittelwert und Stichproben-Varianz seien

$$\begin{aligned} M &= \frac{1}{n} \sum_{i=1}^n Z_i \\ S^2 &= \frac{1}{n} \sum_{i=1}^n (Z_i - M)^2 \end{aligned} \quad (7.3)$$

Der mittlere Semi-Variogrammwert aller n^2 Wertepaare sei

$$\Gamma_n = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \gamma(\mathbf{x}_i, \mathbf{x}_j) \quad (7.4)$$

Dann gilt

$$\begin{aligned}
 \Gamma_n &= \frac{1}{2n^2} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}\{(Z_i - Z_j)^2\} \\
 &= \frac{1}{2n^2} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}\{[(Z_i - M) - (Z_j - M)]^2\} \\
 &= \frac{1}{2n^2} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}\{(Z_i - M)^2 - 2(Z_i - M)(Z_j - M) + (Z_j - M)^2\} \\
 &= \mathbb{E}\left[\frac{1}{2n^2} \sum_{i=1}^n \sum_{j=1}^n \{(Z_i - M)^2 - 2(Z_i - M)(Z_j - M) + (Z_j - M)^2\}\right] \\
 &= \mathbb{E}\left[\frac{1}{2n^2} \sum_{i=1}^n \sum_{j=1}^n (Z_i - M)^2 - \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n (Z_i - M)(Z_j - M) + \frac{1}{2n^2} \sum_{i=1}^n \sum_{j=1}^n (Z_j - M)^2\right] \\
 &= \mathbb{E}\left[\frac{1}{2n} \sum_{i=1}^n (Z_i - M)^2 - \frac{1}{n^2} \sum_{i=1}^n (Z_i - M) \sum_{j=1}^n (Z_j - M) + \frac{1}{2n} \sum_{j=1}^n (Z_j - M)^2\right] \\
 &= \mathbb{E}\left[\frac{1}{2n} \sum_{i=1}^n (Z_i - M)^2 + \frac{1}{2n} \sum_{j=1}^n (Z_j - M)^2\right] \\
 &= \mathbb{E}\left[\frac{1}{n} \sum_{i=1}^n (Z_i - M)^2\right]
 \end{aligned} \tag{7.5}$$

Der mittlere Semi–Variogrammwert aller n^2 Wertepaare ist also gleich dem Erwartungswert der Stichproben–Varianz

$$\Gamma_n = \mathbb{E}[S^2] \tag{7.6}$$

d.h. für ortsabhängige nicht als unabhängig angenommene Zufallsvariablen ist der Erwartungswert der Stichproben–Varianz eine Funktion der gesamten geographischen Datenkonfiguration und des gesamten Variogramms, nicht nur der Schwellenwert des Semi–Variogramms.

Falls die Meßorte gleichmäßig über ein Gebiet verteilt sind, dessen Ausmaße ein Vielfaches der Reichweite sind, dann gehen in die Berechnung des mittleren Variogrammwertes aller Paare sehr viele Paare mit einem Abstand größer als die Reichweite und wenige Paare mit einem Abstand kleiner als die Reichweite ein. Der mittlere Variogrammwert ist also der Mittelwert vieler Werte, die gleich dem Schwellenwert sind, und weniger Werte, die kleiner als der Schwellenwert sind. In diesem Fall ist die Stichproben–Varianz eine sinnvolle erste Schätzung des Schwellenwerts. Als heuristisches Kriterium kann man benutzen, daß die Daten gleichmäßig in einem Gebiet von der Größe der dreifachen Reichweite verteilt sein sollen.

Falls die Daten aus einem Gebiet stammen, dessen Ausmaße von der gleichen Größenordnung wie die Reichweite sind, ist die Stichproben–Varianz keine geeignete Anfangsschätzung des Schwellenwerts, da sie den Schwellenwert unterschätzen würde.

Allgemein kann man feststellen, daß im Falle eines im plot der experimentellen Semi–Variogramme deutlich erkennbaren Schwellenwerts dieser Schwellenwert als Schätzung der Varianz der Gesamtheit benutzt werden kann, aber die Stichproben–Varianz nicht als Schätzung des Schwellenwerts benutzt werden soll.

- Reichweite
Bei unzureichendem Datendargebot wird die Reichweite leicht unterschätzt.
- Anisotropie–Parameter
Hauptachsen des Anisotropie–Ellipsoids bestimmen

Wie gut ist das Modell?

Darauf gibt es keine Antwort, weil kein objektives Maß für die Beurteilung der Güte der Anpassung existiert — auch Kreuzprüfung (cross–validation) ist kein solches, da z.B. auch Implementierungsparameter (z.B. der Suchstrategie) in die Berechnung der Kreuzprüfung eingehen.

Chapter 8

Kriging

Auf der Basis der n Meßwerte für ein ortsabhängiges Merkmal $z(\mathbf{x}_\alpha)$, $\alpha = 1, \dots, n$, soll der Wert dieses Merkmals am Ort \mathbf{x} geschätzt werden, und zwar durch eine Linearkombination der Meßwerte

$$z^*(\mathbf{x}) = \sum_{\alpha} \lambda_\alpha(\mathbf{x}) z(\mathbf{x}_\alpha)$$

Im Rahmen des stochastischen Modells einer Zufallsfunktion werden $z(\mathbf{x}_\alpha)$ als Realisationen der regionalisierten Variablen $Z(\mathbf{x}_\alpha)$ und der Schätzwert $z^*(\mathbf{x})$ (estimate) als Realisation des linearen Regressionsschätzers $Z^*(\mathbf{x})$ (estimator), der selber wieder eine regionale Zufallsvariable ist, aufgefaßt:

$$Z^*(\mathbf{x}) - m(\mathbf{x}) = \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_\alpha(\mathbf{x}) [Z(\mathbf{x}_\alpha) - m(\mathbf{x}_\alpha)]$$

mit

$$\begin{aligned} m(\mathbf{x}) &= E[Z(\mathbf{x})] \\ m(\mathbf{x}_\alpha) &= E[Z(\mathbf{x}_\alpha)] \end{aligned}$$

Sowohl die Anzahl der zu berücksichtigenden Punkte als auch ihre Gewichtung hängen vom Ort \mathbf{x} ab. Gewöhnlich benutzt man nur die Punkte in einer bestimmten Nachbarschaft von \mathbf{x} .

Ein guter Schätzer zeichnet sich durch Unverzerrtheit (unbiasedness)

$$E[Z^*(\mathbf{x})] = E[Z(\mathbf{x})]$$

bzw.

$$E[Z^*(\mathbf{x}) - Z(\mathbf{x})] = 0$$

und eine kleine Schätzvarianz

$$\sigma_E^2(\mathbf{x}) = \text{Var}[Z^*(\mathbf{x}) - Z(\mathbf{x})] \xrightarrow{!} \min$$

aus.

Diese Eigenschaften führen zum **blue** — best linear unbiased estimator — und zu dem mathematischen Optimierungsproblem

$$\sigma_E^2(\mathbf{x}) \xrightarrow{!} \min$$

unter Berücksichtigung der Nebenbedingung

$$\mathbb{E}[Z^*(\mathbf{x}) - Z(\mathbf{x})] = 0$$

Gemäß den Stationaritätsannahmen, die nötig sind, um das stochastische Modell der regionalisierten Variable operabel zu machen, zerlegt man die regionalisierte Zufallsvariable

$$Z(\mathbf{x}) = R(\mathbf{x}) + m(\mathbf{x})$$

in eine stationäre Komponente $R(\mathbf{x})$ und eine Trend- (Drift-) Komponente $m(\mathbf{x})$, wobei

$$m(\mathbf{x}) = \mathbb{E}[Z(\mathbf{x})]$$

und

$$\begin{aligned} \mathbb{E}[R(\mathbf{x})] &= 0 \\ \text{Cov}[R(\mathbf{x}), R(\mathbf{x} + \mathbf{h})] &= \mathbb{E}[R(\mathbf{x})R(\mathbf{x} + \mathbf{h})] \\ &= C_R(\mathbf{h}) \end{aligned}$$

Je nach Wahl des Modells für die Trend- (Drift-) Komponente unterscheidet man drei verschiedene Kriging-Varianten:

- **einfaches (simple) Kriging (SK)** setzt konstanten und bekannten Trend voraus

$$m(\mathbf{x}) = m \quad \text{für alle } \mathbf{x} \in D$$

- **gewöhnliches (ordinary) Kriging (OK)** geht von einem unbekannten, in der Nachbarschaft W von \mathbf{x} konstantem Trend aus

$$m(\mathbf{x}') = m \quad \text{für alle } \mathbf{x}' \in W(\mathbf{x})$$

- **Kriging mit einem Trendmodell (KT)**, früher auch als *universelles (universal) Kriging (UK)* bezeichnet, geht von dem Modell

$$m(\mathbf{x}') = \sum_{k=0}^K a_k f_k(\mathbf{x}') \quad \text{für alle } \mathbf{x}' \in W(\mathbf{x})$$

für den unbekannten Trend aus, also von unbekannten a_k und bestimmten Funktionen f_k .

Man setzt $f_0(\mathbf{x}') = 1$, so daß für $K = 0$ Kriging mit einem Trendmodell zum gewöhnlichen Kriging wird.

8.1 Einfaches (simple) Punkt–Kriging

$$\begin{aligned}
Z_{SK}^*(\mathbf{x}) &= \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_\alpha^{SK}(\mathbf{x}) [Z(\mathbf{x}_\alpha) - m] + m \\
&= \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_\alpha^{SK}(\mathbf{x}) Z(\mathbf{x}_\alpha) + [1 - \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_\alpha^{SK}(\mathbf{x})] m \\
&= \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_\alpha^{SK}(\mathbf{x}) Z(\mathbf{x}_\alpha) + \lambda_m^{SK} m
\end{aligned}$$

Der einfache Kriging–Schätzer $Z_{SK}^*(\mathbf{x})$ ist unverzerrt

$$\mathbb{E}(Z_{SK}^*(\mathbf{x})) = m ,$$

also

$$\mathbb{E}(Z_{SK}^*(\mathbf{x}) - Z(\mathbf{x})) = 0 .$$

Weiter, wegen

$$\begin{aligned}
Z_{SK}^*(\mathbf{x}) - Z(\mathbf{x}) &= [Z_{SK}^*(\mathbf{x}) - m] - [Z(\mathbf{x}) - m] \\
&= \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_\alpha^{SK}(\mathbf{x}) [Z(\mathbf{x}_\alpha) - m] - [Z(\mathbf{x}) - m] \\
&= \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_\alpha^{SK}(\mathbf{x}) R(\mathbf{x}_\alpha) - R(\mathbf{x}) \\
&= R_{SK}^*(\mathbf{x}) - R(\mathbf{x})
\end{aligned}$$

gilt

$$\begin{aligned}
\sigma_E^2(\mathbf{x}) &= \text{Var}(Z_{SK}^*(\mathbf{x}) - Z(\mathbf{x})) \\
&= \text{Var}(R_{SK}^*(\mathbf{x}) - R(\mathbf{x})) \\
&= \text{Var}(R_{SK}^*(\mathbf{x})) + \text{Var}(R(\mathbf{x})) - 2\text{Cov}(R_{SK}^*(\mathbf{x}), R(\mathbf{x})) \\
&= \sum_{\alpha=1}^{n(\mathbf{x})} \sum_{\beta=1}^{n(\mathbf{x})} \lambda_\alpha^{SK}(\mathbf{x}) \lambda_\beta^{SK}(\mathbf{x}) C_R(\mathbf{x}_\alpha - \mathbf{x}_\beta) + C_R(0) - 2 \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_\alpha^{SK}(\mathbf{x}) C_R(\mathbf{x}_\alpha - \mathbf{x}) \\
&= Q(\lambda_1^{SK}(\mathbf{x}), \lambda_2^{SK}(\mathbf{x}), \dots, \lambda_{n(\mathbf{x})}^{SK}(\mathbf{x})) \xrightarrow{!} \min
\end{aligned}$$

Die optimalen Gewichte $\lambda_\alpha^{SK}(\mathbf{x})$, für die Q sein Minimum annimmt, findet man, indem man die entsprechenden partiellen Ableitungen Null setzt.

$$\frac{1}{2} \frac{\partial Q}{\partial \lambda_\alpha^{SK}(\mathbf{x})} = \sum_{\beta=1}^{n(\mathbf{x})} \lambda_\beta^{SK}(\mathbf{x}) C_R(\mathbf{x}_\alpha - \mathbf{x}_\beta) - C_R(\mathbf{x}_\alpha - \mathbf{x}) = 0 , \quad \alpha = 1, \dots, n(\mathbf{x})$$

Wegen

$$\begin{aligned} C_R(\mathbf{h}) &= \mathbb{E}(R(\mathbf{x})R(\mathbf{x} + h)) \\ &= \mathbb{E}((Z(\mathbf{x}) - m)(Z(\mathbf{x} + h) - m)) \\ &= C(\mathbf{h}) \end{aligned}$$

erhält man schließlich das lineare Gleichungssystem

$$\sum_{\beta=1}^{n(\mathbf{x})} \lambda_{\beta}^{SK}(\mathbf{x}) C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) = C(\mathbf{x}_{\alpha} - \mathbf{x}), \quad \alpha = 1, \dots, n(\mathbf{x})$$

Für die nach diesem Gleichungssystem bestimmten optimalen Gewichte $\lambda_{\alpha}^{SK}(\mathbf{x})$ nimmt die Schätzvarianz $\sigma_E^2(\mathbf{x})$ ihren minimalen Wert

$$\sigma_{SK}^2(\mathbf{x}) = C(0) - \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}^{SK}(\mathbf{x}) C(\mathbf{x}_{\alpha} - \mathbf{x})$$

an. Dieser Ausdruck ergibt sich aus der allgemeinen Gleichung für die Schätzvarianz $\sigma_E^2(\mathbf{x})$, wenn man anstelle von $\lambda_{\beta}^{SK}(\mathbf{x}) C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta})$ den laut Krige-Gleichungssystem äquivalenten Ausdruck $C(\mathbf{x}_{\alpha} - \mathbf{x})$ einsetzt:

$$\begin{aligned} \sigma_E^2(\mathbf{x}) &= \sum_{\alpha=1}^{n(\mathbf{x})} \sum_{\beta=1}^{n(\mathbf{x})} \lambda_{\alpha}^{SK}(\mathbf{x}) \lambda_{\beta}^{SK}(\mathbf{x}) C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) + C(0) - 2 \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}^{SK}(\mathbf{x}) C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) \\ &= \sum_{\alpha=1}^{n(\mathbf{x})} C(\mathbf{x}_{\alpha} - \mathbf{x}) + C(0) - 2 \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}^{SK}(\mathbf{x}) C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) \\ &= C(0) - \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}^{SK}(\mathbf{x}) C(\mathbf{x}_{\alpha} - \mathbf{x}) \end{aligned}$$

Anmerkung: In dem Ausdruck für diese Kriging-Varianz kommen die Meßwerte $z(\mathbf{x}_{\alpha})$ bzw. die ihnen entsprechenden regionalisierten Variablen $Z(\mathbf{x}_{\alpha})$ nicht vor!

Schreibt man das Krige-Gleichungssystem ausführlich hin

$$\begin{pmatrix} C(\mathbf{x}_1 - \mathbf{x}_1) & \cdots & C(\mathbf{x}_1 - \mathbf{x}_{n(\mathbf{x})}) \\ \vdots & \vdots & \vdots \\ C(\mathbf{x}_{n(\mathbf{x})} - \mathbf{x}_1) & \cdots & C(\mathbf{x}_{n(\mathbf{x})} - \mathbf{x}_{n(\mathbf{x})}) \end{pmatrix} \begin{pmatrix} \lambda_1^{SK}(\mathbf{x}) \\ \vdots \\ \lambda_{n(\mathbf{x})}^{SK}(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} C(\mathbf{x}_1 - \mathbf{x}) \\ \vdots \\ C(\mathbf{x}_{n(\mathbf{x})} - \mathbf{x}) \end{pmatrix},$$

und komprimiert dann in Matrix-Schreibweise zu

$$K_{SK} \lambda_{SK}(\mathbf{x}) = k_{SK},$$

kann man die Lösung angeben als

$$\lambda_{SK}(\mathbf{x}) = K_{SK}^{-1} k_{SK}$$

und die Kriging–Varianz als

$$\begin{aligned}\sigma_{SK}^2 &= C(0) - \lambda_{SK}^t(\mathbf{x}) k_{SK} \\ &= C(0) - k_{SK}^t K_{SK}^{-1} k_{SK}\end{aligned}$$

Das Gleichungssystem des einfachen Kriging besitzt eine eindeutige Lösung und die Kriging–Varianz ist positiv, falls $K_{SK} = C(\mathbf{x}_\alpha - \mathbf{x}_\beta)$ positiv definit ist, d.h. in der Praxis, falls

- (i) $\mathbf{x}_\alpha \neq \mathbf{x}_\beta$ für $\alpha \neq \beta$
- (ii) das gewählte Kovarianzmodell zulässig ist.

Kriging besitzt die Interpolationseigenschaft, d.h.

$$z_{SK}^*(\mathbf{x}_\alpha) = z(\mathbf{x}_\alpha) , \quad \alpha = 1, \dots, n$$

Zusammenfassend vermitteln

$$Z_{SK}^*(\mathbf{x}) = \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_\alpha^{SK}(\mathbf{x}) Z(\mathbf{x}_\alpha) + \left[1 - \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_\alpha^{SK}(\mathbf{x}) \right] m$$

und

$$\lambda_{SK}(\mathbf{x}) = K_{SK}^{-1} k_{SK}$$

folgende Lesart: Wenn sich der Ort \mathbf{x} des zu schätzenden Wertes $z_{SK}^*(\mathbf{x})$ weiter von den beteiligten Meßorten entfernt, dann wird $C(\mathbf{x}_\alpha - \mathbf{x})$ kleiner, während $C(\mathbf{x}_\alpha - \mathbf{x}_\beta)$ unverändert bleibt. In diesem Fall nehmen die SK–Gewichte ab und das Gewicht des stationären Mittelwertes nimmt zu. Je weiter der Ort des zu schätzenden Wertes von den Datenorten entfernt ist, desto mehr ist der Schätzwert durch das stationäre Mittel bestimmt.

Journel, A.G., Fundamentals of Geostatistics in Five Lessons: AGU (1989), pp. 11-12

Independent case

Assume that the n RV data Z_α , $\alpha = 1, \dots, n$, are independent one from another; they are then called “independent variables”, the unknown RV Z_0 being the “dependent variable”.

The data covariance matrix \underline{K} thus reduces to a diagonal matrix with for diagonal elements the data variances

$$\underline{K} = [C_{\alpha\beta}] = \begin{bmatrix} C_{11} & & \underline{0} \\ & \ddots & \\ \underline{0} & & C_{nn} \end{bmatrix},$$

i.e. $C_{\alpha\beta} = \begin{cases} 0, & \text{for } \alpha \neq \beta \\ C_{\alpha\alpha}, & \text{for } \alpha = \beta \end{cases}$

The SK system (12)¹ is then written:

$$\lambda_\alpha C_{\alpha\alpha} = C_{\alpha 0} \implies \lambda_\alpha = C_{\alpha 0} / C_{\alpha\alpha}, \alpha = 1, \dots, n$$

Introducing the coefficient of correlation:

$$\rho_{\alpha 0} = C_{\alpha 0} / \sqrt{C_{\alpha\alpha} C_{00}} \implies \lambda_\alpha = \rho_{\alpha 0} \frac{\sqrt{C_{00}}}{\sqrt{C_{\alpha\alpha}}}, \alpha = 1, \dots, n$$

and the SK estimator (9)² becomes:

$$[Z_0 - m_0]^* = \sum_{\alpha=1}^n \rho_{\alpha 0} \frac{\sqrt{C_{00}}}{\sqrt{C_{\alpha\alpha}}} \cdot [Z_\alpha - m_\alpha]$$

i.e., with the standard deviation notation: $\sigma_\alpha = \sqrt{C_{\alpha\alpha}}$

$$\left[\frac{Z_0 - m_0}{\sigma_0} \right]^* = \sum_{\alpha=1}^n \rho_{\alpha 0} \left[\frac{Z_\alpha - m_\alpha}{\sigma_\alpha} \right] \quad (8.16)$$

In presence of independent data, the SK (regression) weights $\lambda_\alpha, \alpha = 1, \dots, n$ are none other than the correlation coefficients $\rho_{0\alpha}$ between each datum and the unknown (dependent) variable Z_0 .

The greater the correlation $\rho_{0\alpha}$ of a particular datum with Z_0 , the greater its weight.

¹Das simple kriging(SK) System in Gleichung (12) lautet

$$\sum_{\beta=1}^n \lambda_\beta C_{\alpha\beta} = C_{\alpha 0}, \alpha = 1, \dots, n$$

²Der SK Sch"atzer in Gleichung (9) lautet

$$[Z_0^* - m_0] = [Z_0 - m_0]^* = \sum_{\alpha=1}^n \lambda_\alpha [Z_\alpha - m_\alpha]$$

Note that:

$$\sum_{\alpha=1}^n \rho_{\alpha 0} \neq 1, \text{ usually.}$$

$$\sigma_{SK}^2 = C_{00} \left[1 - \sum_{\alpha} \lambda_{\alpha} \frac{C_{\alpha 0}}{C_{00}} \right] = C_{00} [1 - \sum_{\alpha} \rho_{\alpha 0}^2]$$

Moreover, if the n data are independent of the unknown Z_0 , i.e., if $\rho_{\alpha 0} = 0$, the SK estimator reduces to the mean: $Z_0^* = m_0$.

8.2 Duale Form des einfachen Punkt-Kriging

Es war

$$z_{SK}^*(\mathbf{x}) = \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}^{SK}(\mathbf{x}) [z(\mathbf{x}_{\alpha}) - m] + m$$

also $z_{SK}^*(\mathbf{x})$ eine Linearkombination der Meßwerte $z(\mathbf{x}_{\alpha})$, $\alpha = 1, \dots, n(\mathbf{x})$.

Die optimalen Gewichte $\lambda_{\alpha}^{SK}(\mathbf{x})$, die die Kriging-Varianz minimieren, sind durch das lineare Gleichungssystem

$$\sum_{\beta=1}^{n(\mathbf{x})} \lambda_{\beta}^{SK}(\mathbf{x}) C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) = C(\mathbf{x}_{\alpha} - \mathbf{x}), \quad \alpha = 1, \dots, n(\mathbf{x})$$

bestimmt, in Matrix-Form

$$K_{SK} \lambda_{SK}(\mathbf{x}) = k_{SK}$$

$$\lambda_{SK}(\mathbf{x}) = K_{SK}^{-1} k_{SK}$$

Damit

$$\begin{aligned} Z_{SK}^*(\mathbf{x}) - m &= \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}^{SK}(\mathbf{x}) [Z(\mathbf{x}_{\alpha}) - m] \\ &= (Z(\mathbf{x}_1) - m, \dots, Z(\mathbf{x}_{n(\mathbf{x})}) - m) \begin{pmatrix} \lambda_1^{SK}(\mathbf{x}) \\ \vdots \\ \lambda_{n(\mathbf{x})}^{SK}(\mathbf{x}) \end{pmatrix} \\ &= R_{\alpha}^t \lambda_{SK}(\mathbf{x}) \\ &= R_{\alpha}^t K_{SK}^{-1} k_{SK} \\ &= d_{SK}^t k_{SK} \\ &= \sum_{\alpha=1}^{n(\mathbf{x})} d_{\alpha}^{SK}(\mathbf{x}) C(\mathbf{x}_{\alpha} - \mathbf{x}) \end{aligned}$$

wobei

$$\begin{aligned} d^{SK} &= K_{SK}^{-1} R_\alpha \\ K_{SK} d^{SK} &= R_\alpha \\ \sum_{\beta=1}^{n(\mathbf{x})} d_\alpha^{SK} C(\mathbf{x}_\alpha - \mathbf{x}_\beta) &= z(\mathbf{x}_\alpha) - m \end{aligned}$$

Die Gewichte $\lambda_{SK}(\mathbf{x})$ sind also als Linearkombinationen der Kovarianzen $C(\mathbf{x}_\alpha - \mathbf{x})$ bestimmt. Deshalb kann der Schätzwert $z_{SK}^*(\mathbf{x})$ selbst auch als Linearkombination dieser Kovarianzen plus stationäres Mittel m aufgefaßt werden

$$z_{SK}^*(\mathbf{x}) = \sum_{\alpha=1}^{n(\mathbf{x})} d_\alpha^{SK}(\mathbf{x}) C(\mathbf{x}_\alpha - \mathbf{x}) + m$$

wobei die Koeffizienten d_α^{SK} als duale Kriging–Gewichte bezüglich der Kovarianzen $C(\mathbf{x}_\alpha - \mathbf{x})$ bezeichnet werden.

Sie werden durch die Forderungen

$$z_{SK}^*(\mathbf{x}_\alpha) = \sum_{\beta=1}^{n(\mathbf{x}_\alpha)} d_\beta^{SK}(\mathbf{x}_\alpha) C(\mathbf{x}_\beta - \mathbf{x}_\alpha) + m \stackrel{!}{=} z(\mathbf{x}_\alpha)$$

bestimmt.

Damit vermittelt duales Kriging den formalen Anschluß an deterministische Interpolation mit Hilfe radialer Basisfunktionen.

8.3 Gewöhnliches (ordinary) Punkt–Kriging

Verallgemeinerung des einfachen Krigings, indem nicht mehr vorausgesetzt wird, daß der Erwartungswert auf ganz D , also global, konstant ist, sondern daß er nur noch lokal in einer Umgebung $W(\mathbf{x}) \subset D$, der Schätzumgebung, von \mathbf{x} , konstant ist. Diese lokal konstanten Erwartungswerte (Mittelwerte) sind zunächst nicht bekannt und müssen selbst geschätzt werden – wieder durch Kriging.

$$Z^*(\mathbf{x}) = \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_\alpha(\mathbf{x}) Z(\mathbf{x}_\alpha) + \underbrace{\left[1 - \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_\alpha(\mathbf{x}) Z(\mathbf{x}_\alpha) \right]}_{\stackrel{!}{=} 0 \text{ für } \lambda_\alpha^{OK}(\mathbf{x})} m(\mathbf{x})$$

also

$$Z_{OK}^*(\mathbf{x}) = \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_\alpha^{OK}(\mathbf{x}) Z(\mathbf{x}_\alpha) \quad \text{mit} \quad \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_\alpha^{OK}(\mathbf{x}) = 1$$

Es gilt

$$E(Z_{OK}^*(\mathbf{x}) - Z(\mathbf{x})) = \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}^{OK} m(\mathbf{x}) - m(\mathbf{x}) = 0$$

also Unverzerrtheit, und Minimierung der Schätzvarianz $\sigma_E^2(\mathbf{x})$ unter Berücksichtigung der Nebenbedingung $\sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}^{OK}(\mathbf{x}) = 1$ führt zur Definition der *Lagrange-Funktion*

$$\sigma_E^2(\mathbf{x}) + 2\mu_{OK}(\mathbf{x}) \left[\sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}^{OK}(\mathbf{x}) - 1 \right] =: L(\lambda_1^{OK}(\mathbf{x}), \dots, \lambda_{n(\mathbf{x})}^{OK}(\mathbf{x}); 2\mu_{OK}(\mathbf{x}))$$

mit dem Lagrange-Parameter $2\mu_{OK}(\mathbf{x})$.

Jetzt liefern die $n(\mathbf{x}) + 1$ partiellen Ableitungen

$$\begin{aligned} \frac{1}{2} \frac{\partial L}{\partial \lambda_{\alpha}^{OK}(\mathbf{x})} &= \sum_{\beta=1}^{n(\mathbf{x})} \lambda_{\beta}^{OK}(\mathbf{x}) C_R(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) - C_R(\mathbf{x}_{\alpha} - \mathbf{x}) + \mu_{OK}(\mathbf{x}) = 0, \quad \alpha = 1, \dots, n(\mathbf{x}) \\ \frac{1}{2} \frac{\partial L}{\partial \lambda_{\alpha}^{OK}(\mathbf{x})} &= \sum_{\beta=1}^{n(\mathbf{x})} \lambda_{\beta}^{OK}(\mathbf{x}) - 1 = 0 \end{aligned}$$

die Bedingungen für minimale Schätzvarianz

$$\begin{aligned} \sum_{\beta=1}^{n(\mathbf{x})} \lambda_{\beta}^{OK}(\mathbf{x}) C_R(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) + \mu_{OK}(\mathbf{x}) &= C_R(\mathbf{x}_{\alpha} - \mathbf{x}), \quad \alpha = 1, \dots, n(\mathbf{x}) \\ \sum_{\beta=1}^{n(\mathbf{x})} \lambda_{\beta}^{OK}(\mathbf{x}) &= 1 \end{aligned}$$

In der Praxis identifiziert man $C_R(\mathbf{h})$, die Kovarianzfunktion der Residuen, mit $C(\mathbf{h})$, der globalen Z -Kovarianzfunktion, obwohl der Erwartungswert nicht mehr als global konstant vorausgesetzt wurde.

Auf diese Weise erhält man

$$\begin{aligned} \sum_{\beta=1}^{n(\mathbf{x})} \lambda_{\beta}^{OK}(\mathbf{x}) C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) + \mu_{OK}(\mathbf{x}) &= C(\mathbf{x}_{\alpha} - \mathbf{x}), \quad \alpha = 1, \dots, n(\mathbf{x}) \\ \sum_{\beta=1}^{n(\mathbf{x})} \lambda_{\beta}^{OK}(\mathbf{x}) &= 1 \end{aligned}$$

Für die so bestimmten optimalen Gewichte $\lambda_{\beta}^{OK}(\mathbf{x})$ nimmt die Schätzvarianz σ_E^2 ihren minimalen Wert

$$\sigma_{OK}^2(\mathbf{x}) = C(0) - \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}^{OK}(\mathbf{x}) C(\mathbf{x}_{\alpha} - \mathbf{x}) - \mu_{OK}(\mathbf{x})$$

an.

Benutzt man den bekannten Zusammenhang

$$C(\mathbf{h}) = C(0) - \gamma(\mathbf{h})$$

zwischen Z -Kovarianzfunktion und Semivariogramm, kann man das OK-Gleichungssystem mit Hilfe des Semivariogramms schreiben

$$\begin{aligned} \sum_{\beta=1}^{n(\mathbf{x})} \lambda_{\beta}^{OK}(\mathbf{x}) \gamma(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) - \mu_{OK}(\mathbf{x}) &= \gamma(\mathbf{x}_{\alpha} - \mathbf{x}), \quad \alpha = 1, \dots, n(\mathbf{x}) \\ \sum_{\beta=1}^{n(\mathbf{x})} \lambda_{\beta}^{OK}(\mathbf{x}) &= 1 \end{aligned}$$

Das OK-Gleichungssystem kann man also sowohl in der Schreibweise der Kovarianzfunktion als auch des Semivariogramms darstellen; das SK-Gleichungssystem kann man nur in der Schreibweise der Kovarianzfunktion darstellen.

Die Existenz der Schätz-Varianz $\text{Var}(Z_{SK}^*(\mathbf{x})Z(\mathbf{x}))$ ist unter der Annahme der schwachen Stationarität unbedingt gesichert, da die Varianz aller beteiligten Zufallsvariablen und also auch ihrer Linearkombinationen gesichert ist. Unter der Annahme der schwachen Stationarität der Zuwächse (intrinsische Stationarität) ist die Existenz der Schätz-Varianz zunächst nicht gesichert. Nur für Linearkombinationen, deren Koeffizienten sich zu 0 summieren, ist die Existenz ihrer Varianz gesichert. Für die Koeffizienten der Linearkombination des OK-Schätzers $Z_{SK}^*(bmx)$ impliziert die Forderung der Unverzerrtheit gerade, daß sie sich zu 1 summieren; die Koeffizienten der Linearkombination $Z_{SK}^*(\mathbf{x})Z(\mathbf{x})$ summieren sich also gerade zu 0, das heißt die Existenz der Schätz-Varianz ist durch die aus der Forderung nach Unverzerrtheit des Schätzers folgende Bedingung $\sum \lambda_{\alpha} = 1$ gesichert. Eine solche Bedingung gibt es aber im Formalismus des SK nicht; deshalb kann man SK nur unter der Annahme der schwachen Stationarität und mit Hilfe der Kovarianzfunktion darstellen.

Wackernagel, H., Multivariate Geostatistics: Springer (1995), p. 75:

Authorized use of the variogram in the computation of the variance of the estimation error

The estimation variance $\sigma_E^2 = \text{var}(Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0))$ is the variance of the linear combination

$$Z^*(\mathbf{x}_0) - Z(\mathbf{x}_0) = \sum_{\alpha=1}^n \omega_{\alpha} Z(\mathbf{x}_{\alpha}) - 1 \cdot Z(\mathbf{x}_0) = \sum_{\alpha=0} \omega_{\alpha} Z(\mathbf{x}_{\alpha})$$

with a weight ω_0 equal to -1 and

$$\sum_{\alpha=0} \omega_{\alpha} = 0$$

Thus the condition that the weights numbered from 1 to n sum up to one also implies that the use of the variogram is *authorized* in the computation of the variance of the estimation error.

The variogram is authorized for ordinary kriging, but not for simple kriging, because the latter does not include a constraint on the weights.

Für das Semivariogramm der Residuen gilt

$$\begin{aligned} 2\gamma(\mathbf{h}) &= E\left\{[Z(\mathbf{x}') - Z(\mathbf{x}' + \mathbf{h})]^2\right\} \\ &= E\left\{[R(\mathbf{x}') + m(\mathbf{x}') - R(\mathbf{x}' + \mathbf{h}) - m(\mathbf{x}' + \mathbf{h})]^2\right\} \\ &= E\left\{[R(\mathbf{x}') - R(\mathbf{x}' + \mathbf{h})]^2\right\} \end{aligned}$$

da

$$m(\mathbf{x}') = m(\mathbf{x}' + \mathbf{h}) \quad \text{für alle } \mathbf{x}', \mathbf{x}' + \mathbf{h} \in W(\mathbf{x})$$

In der Praxis modelliert man deshalb gewöhnlich ein Semivariogramm, für numerische Berechnung benutzt man dann eine sogenannte Pseudo-Kovarianz

$$A - \gamma(\mathbf{h})$$

mit einer so hinreichend großen Zahl A , daß

$$A - \gamma(\mathbf{h}) \geq 0$$

für alle h , die numerisch relevant sind, da sie numerisch effizienter anzuwenden ist.

Der lokal konstante unbekannte Erwartungswert wird geschätzt durch

$$m_{OK}^*(\mathbf{x}) = \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha m}^{OK}(\mathbf{x}) Z(\mathbf{x}_\alpha)$$

Die Forderung der Unverzerrtheit

$$E(m_{OK}^*(\mathbf{x}) - m(\mathbf{x})) = 0$$

führt zu

$$\sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha m}^{OK} = 1$$

Die Forderung der minimalen Schätzvarianz

$$\sigma_E^2(\mathbf{x}) = \text{Var}(m_{OK}^*(\mathbf{x}) - m(\mathbf{x})) \xrightarrow{!} 0$$

führt über

$$\sigma_E^2(\mathbf{x}) = \sum_{\alpha=1}^{n(\mathbf{x})} \sum_{\beta=1}^{n(\mathbf{x})} \lambda_{\alpha m}^{OK}(\mathbf{x}) \lambda_{\beta m}^{OK}(\mathbf{x}) C(\mathbf{x}_\alpha - \mathbf{x}_\beta) \xrightarrow{!} 0$$

zu

$$\begin{aligned} \sum_{\beta=1}^{n(\mathbf{x})} \lambda_{\beta m}^{OK}(\mathbf{x}) C(\mathbf{x}_\alpha - \mathbf{x}_\beta) + \mu_m^{OK}(\mathbf{x}) &= 0, \quad \alpha = 1, \dots, n(\mathbf{x}) \\ \sum_{\beta=1}^{n(\mathbf{x})} \lambda_{\beta m}^{OK}(\mathbf{x}) &= 1 \end{aligned}$$

Dieses lineare Gleichungssystem ist in den bekannten Größen unabhängig von x : Für zwei Orte \mathbf{x} und \mathbf{x}' , für die die gleiche Schätzumgebung benutzt wird, erhält man identische Lösungen und also identische Schätzer des lokal konstant angenommenen Erwartungswertes.

8.4 Einfaches vs gewöhnliches Punktkriging

Im allgemeinen wird gewöhnliches Kriging in der Praxis bevorzugt, da die Voraussetzungen schwächer sind. Es lässt sich zeigen, daß gewöhnliches Kriging äquivalent zu folgendem ist:

- zuerst die lokal konstanten Mittelwerte mittels OK zu schätzen,
- dann im Formalismus von SK den global konstant angenommenen Erwartungswert m lokal durch $m_{OK}^*(\mathbf{x})$ zu ersetzen und dann SK durchzuführen; formelhaft

$$\begin{aligned} Z_{OK}^*(\mathbf{x}) &= \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_\alpha^{SK}(\mathbf{x}) [Z(\mathbf{x}_\alpha) - m_{OK}^*(\mathbf{x})] + m_{OK}^*(\mathbf{x}) \\ &= \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_\alpha^{SK}(\mathbf{x}) Z(\mathbf{x}_\alpha) + \lambda_m^{SK}(\mathbf{x}) m_{OK}^*(\mathbf{x}) \end{aligned}$$

Zwischen SK und OK Schätzer besteht daher der Zusammenhang

$$\begin{aligned} Z_{OK}^*(\mathbf{x}) &= Z_{SK}^*(\mathbf{x}) - \lambda_m^{SK}(\mathbf{x}) m + \lambda_m^{SK}(\mathbf{x}) m_{OK}^*(\mathbf{x}) \\ &= Z_{SK}^*(\mathbf{x}) + \lambda_m^{SK}(\mathbf{x}) [m_{OK}^*(\mathbf{x}) - m] \end{aligned}$$

Der Unterschied zwischen SK und OK besteht dann im allgemeinen darin, daß der OK–Schätzwert in Gebieten kleiner Meßwerte kleiner als der SK–Schätzwert ist, da m_{OK}^* kleiner

als m ist, und umgekehrt, daß der OK–Schätzwert in Gebieten großer Werte größer als der SK–Schätzwert ist, da m_{OK}^* größer als m ist.

Die Einbeziehung der Schätzwerte für die nur lokal als konstant angenommenen Erwartungswerte bedeutet tatsächlich, daß die OK–Schätzung die Existenz eines globalen Trends – also Nichtstationarität – durch ein einfaches Modell – stückweise konstante Erwartungswerte – berücksichtigt.

8.5 Punktkriging mit Trend – universelles (universal) Punktkriging

Die nächste denkbare Verallgemeinerung der Krige–Schätzung besteht darin, komplexere Modelle für den Trend zuzulassen. Im Sinne deterministischer Approximationsmethoden wird der Erwartungswert in der Schätznachbarschaft durch

$$m(\mathbf{x}') = \sum_{k=0}^K a_k(\mathbf{x}') f_k(\mathbf{x}')$$

mit $a_k(\mathbf{x}') = a_k$ für alle $\mathbf{x}' \in W(\mathbf{x})$ modelliert. Dabei sind die Funktionen f_k bekannt, die unbekannten Koeffizienten müssen nach irgendeinem Kriterium bestimmt werden.

Durch die Wahl geeigneter Nebenbedingungen kann man den KT–Schätzer wieder als Linearkombination der Variablen $Z(\mathbf{x}_\alpha)$ allein darstellen. Folgt man den formalen Schritten zur Entwicklung von OK, kommt man zu einem nur unwesentlich veränderten Gleichungssystem.

Zur Anwendung des KT–Schätzers muß jedoch der Trend und die Kovarianzfunktion der Residuen bestimmt werden.

Ersteres kann man durch Kriging der unbekannten Koeffizienten im Modellansatz des Trends erledigen. Der zweite Schritt ist wesentlich schwieriger und führt an die Grenzen des Krigings auf der Grundlage des Semivariogramms.

Ein **unzulässiger** Weg besteht darin,

- ein Trendmodell nach der Methode der kleinsten Quadrate anzupassen und dann die Residuen nach

$$\hat{r}(\mathbf{x}_\alpha) = z(\mathbf{x}_\alpha) - \hat{m}(\mathbf{x}_\alpha)$$

zu bestimmen,

- ein Semivariogramm $\gamma_{\hat{R}}(\mathbf{h})$ aufgrund der $\hat{r}(\mathbf{x}_\alpha)$ zu bestimmen und damit in den KT–Algorithmus einzusteigen.

Das so bestimmte $\gamma_{\hat{R}}$ ist stark verzerrt und hängt wesentlich von der Methode der Trendmodellierung ab.

besser: einfaches Kriging der Residuen $\hat{r}(\mathbf{x}_\alpha)$ unter Benutzung von $\gamma_{\hat{R}}$, und dann wieder $\hat{m}(\mathbf{x})$ addieren:

$$\hat{z}(\mathbf{x}) = \hat{r}(\mathbf{x}) + \hat{m}(\mathbf{x})$$

Eine mathematische Lösung stellen die k -IRF, die intrinsischen zufälligen Funktionen k -ter Ordnung, und die Anwendung verallgemeinerter Kovarianzfunktionen zur Verfügung.

In der Praxis bleibt man im allgemeinen bei gewöhnlichem Kriging OK.

8.6 k -IRF (Fontainebleau)

Den Krige-Formalismus kann man für die Fälle SK, OK und KT (UK) ohne Rückgriffe auf stochastische Modellannahmen entwickeln. Dadurch und insbesondere auch durch den mit der dualen Form des krigings vermittelten Zusammenhang zu deterministischen Methoden der Interpolation durch radiale Basisfunktionen zeigt sich, daß diese Verfahren sehr allgemein sind, d.h. im Prinzip keinerlei einschränkender Hypothesen oder Modellannahmen bedürfen.

Vor der praktischen Anwendung steht allerdings das Problem der Schätzung der Kovarianzfunktion oder des Semivariogramms, und genau an dieser Stelle halten die stochastischen Annahmen der Stationarität mit ganzer Macht (wieder) Einzug in das Modell. Wie steht es aber unter der Annahme einer Drift- (Trend-) Komponente $E(Z(\mathbf{x})) = m(\mathbf{x}) \neq \text{const}$, $\mathbf{x} \in D$, um die Schätzung der Kovarianzfunktion oder des Semivariogramms aus?

Zunächst könnte man wiederum daran denken, so lokal schwache Stationarität zu fordern, daß

$$\text{Cov}(Z(\mathbf{x}_1), Z(\mathbf{x}_2)) = C(\mathbf{x}_1, \mathbf{x}_2) = C(\mathbf{x}_1 - \mathbf{x}_2)$$

gilt. Nach der Zerlegung der Zufallsfunktion Z in einen Drift- (Trend-) und einen Fluktuationsterm

$$Z(\mathbf{x}) = m(\mathbf{x}) + R(\mathbf{x})$$

gilt dann

$$\text{Cov}(Z(\mathbf{x}_1), Z(\mathbf{x}_2)) = \text{Cov}(R(\mathbf{x}_1), R(\mathbf{x}_2))$$

so daß es nahe liegt, daran zu denken, mit der Kovarianzfunktion oder dem Semivariogramm der geschätzten Fluktuationen zu rechnen. Dadurch würde man jedoch einen systematischen Fehler (bias, Verzerrung) in das Modell einführen, siehe MATHERON, 1971, p. 188.

Aus dieser Situation könnten zwei sehr vertrackte Wege führen:

- Methoden der statistischen Schlußfolgerung zu entwickeln, die auf dem Semivariogramm der Fluktuationen basieren. Solche Methoden haben sich jedoch als sehr mühsam und in gewisser Weise als unnötig herausgestellt;

- Mit der Annahme eines lokal einfachen (zum Beispiel linearen) Semivariogramms zu arbeiten und nachträglich anhand gewisser experimenteller Kriterien zu überprüfen, inwieweit diese Annahme gerechtfertigt war. Auch diese Vorgehensweise führt zu komplizierten mathematischen Problemen.

Auf diesem Kenntnisstand kommt man nun zu folgender

(i) Schlußfolgerung

- Drift- (Trend-) Komponente nur am Rand – stationäres Modell benutzen und am Rand Bedingungen der Unverzerrtheit berücksichtigen;
- Drift- (Trend-) Komponente nicht deutlich ausgeprägt – für kurze bis mittlere Abstände stationäres Modell benutzen;
- Eine Richtung ohne Drift- (trend-) Komponente – Annahme, daß das wahre Semivariogramm isotrop ist und Modellanpassung für die Richtung ohne Drift- (Trend-) Komponente;
- Überall deutlich ausgeprägte polynomiale Drift- (Trend-) Komponente – Schätzung des Drift- (Trend-) Terms und statistische Analyse der Residuen – good luck!!

(ii) Schlußfolgerung

- Methoden der Geostatistik so zu verallgemeinern, daß sie die mathematische Eleganz bei der Behandlung stationärer Phänomene behalten – Anwendbarkeit??

Konzept der Zuwächse

Mit SK und OK wurde schon vorgeführt, wie es möglich ist, Schätzprobleme nur unter Benutzung des Semivariogramms, d.h. der Varianz der Zuwächse oder – wie man sagt – der Varianz der Differenzen erster Ordnung, zu lösen.

Die Bedeutung der Differenzen erster Ordnung kann man beim Aufstellen des Kriging-Systems erkennen; sie liegt in der Tatsache, daß unter der Annahme, daß $E(Z(\mathbf{x}))$ eine unbekannte Konstante m für alle $\mathbf{x} \in D$ ist, der Erwartungswert der Differenzen erster Ordnung unabhängig vom Wert m verschwindet. Aus dieser Sichtweise kann man sagen, daß die Differenzen erster Ordnung einen konstanten Drift- (Trend-) Term verschwinden lassen, er wird herausgefiltert. Dann gilt

$$\text{Var}(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})) = E([Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]^2)$$

und das Semivariogramm kann unverzerrt durch die quadrierten experimentellen Differenzen geschätzt werden.

Es läßt sich allerdings zeigen, daß dieser Filtereffekt bezüglich Konstanten nicht genügt, um in Gegenwart einer Drift- (Trend-) Komponente das Semivariogramm unverzerrt schätzen

zu können. Deshalb ist es natürlich, jetzt daran zu denken, nicht nur Konstanten, sondern polynomiale Terme bis zur Ordnung k , $k \in \mathbb{N}$, herauszufiltern, d.h. zu Zuwächsen der Ordnung k überzugehen bzw. die Ordnung der Zuwächse solange zu erhöhen, bis diese Zuwächse dann gerade schwach stationär sind. Diese Eigenschaft haben gerade die Differenzen höherer Ordnung.

Als Beispiel sei der Operator der endlichen Differenzen k -ter Ordnung betrachtet

$$\begin{aligned}\Delta_h^1 Z(x) &= Z(x + h) - f(x) \\ \Delta_h^2 Z(x) &= \Delta_h^1 Z(x + h) - \Delta_h^1 Z(x) = \\ &= Z(x + 2h) - 2Z(x + h) + Z(x) \\ \Delta_h^3 Z(x) &= \Delta_h^1 Z(x + 2h) - \Delta_h^1 (2Z(x + h)) + \Delta_h^1 Z(x) = \\ &= Z(x + 3h) - 3Z(x + 2h) + 3Z(x + h) - f(x) \\ \Delta_h^k Z(x) &= \Delta_h^1 (\Delta_h^{k-1} Z(x)) = \sum_{l=1}^k (-1)^{k-l} \binom{n}{k} Z(x + lh)\end{aligned}$$

1D Beispiel

$$\begin{aligned}Z(x) &= R(x) + m(x) \\ m(x) &= a_3 x^3 + a_2 x^2 + a_1 x + a_0\end{aligned}$$

Dann ist

$$\begin{aligned}\Delta_h^1 m(x) &= a_3 h^3 + (3a_3 x + a_2) h^2 + (3a_3 x^2 + 2a_2 x + a_1) h \\ &\quad \text{der konstante Anteil ist verschwunden} \\ \Delta_h^2 m(x) &= 6a_3 h^3 + 2a_2 h^2 \\ &\quad \text{der lineare Anteil ist verschwunden} \\ \Delta_h^3 m(x) &= 6a_3 h^3 \\ &\quad \text{der quadratische Anteil ist verschwunden} \\ \Delta_h^k m(x) &= 0 \text{ für alle } k \geq 4 \\ &\quad \text{die ganze Funktion ist verschwunden}\end{aligned}$$

Der Differenzenoperator wirkt ähnlich wie der Differentialoperator.

Konzept der Zuwächse k -ter Ordnung

Um dieses Konzept fruchtbar zu machen, sind noch zwei Verallgemeinerungen nötig. Die Punkte, zwischen denen Differenzen gebildet werden sollen, sind als Punkte des \mathbb{R}^p , $p = 1, 2, 3$ zu behandeln und nicht länger notwendig als äquidistant anzunehmen.

Es werden im folgenden Zufallsfunktionen, die durch ihre Zuwächse k -ter Ordnung definiert sind, betrachtet, das heißt Ausdrücke der Form

$$\sum_{\alpha} \lambda_{\alpha} Z(\mathbf{x}_{\alpha})$$

wobei λ_α Bedingungen der Form

$$\sum_{\alpha} \lambda_\alpha f^l(\mathbf{x}_\alpha) = 0 \text{ für alle } l \leq k$$

für Monome f^l , $l \leq k$, zu erfüllen haben.

2D Beispiel: Die Funktionen f^l sind

$$\begin{aligned} k = 0 : \quad & x^0 = 1 \\ k = 1 : \quad & 1, x, y \\ k = 2 : \quad & 1, x, y, xy, x^2, y^2 \end{aligned}$$

Für diesen Fall nehmen die Bedingungen die Form

$$\begin{aligned} k = 0 : \quad & \sum \lambda_\alpha = 0 \\ k = 1 : \quad & \sum \lambda_\alpha = 0, \quad \sum \lambda_\alpha x_\alpha = 0, \quad \sum \lambda_\alpha y_\alpha = 0 \\ k = 2 : \quad & \sum \lambda_\alpha = 0, \quad \sum \lambda_\alpha x_\alpha = 0, \quad \sum \lambda_\alpha y_\alpha = 0, \quad \sum \lambda_\alpha x_\alpha y_\alpha = 0, \quad \sum \lambda_\alpha x_\alpha^2 = 0, \quad \sum \lambda_\alpha y_\alpha^2 = 0 \end{aligned}$$

an.

Man kann hier deutlich eine Verallgemeinerung der 1D Differenzen k -ter Ordnung erkennen; so ist in 2D die Linearkombination

$$Z(-1, 0) + Z(1, 0) + Z(0, -1) + Z(0, 1) - 4Z(0, 0)$$

eine verallgemeinerte Differenz erster Ordnung; sie ist keine verallgemeinerte Differenz zweiter Ordnung, da

$$\begin{aligned} \sum \lambda_\alpha x_\alpha^2 &= 1 * (-1)^2 + 1 * (+1)^2 + 1 * 0^2 + 1 * 0^2 \neq 0 \\ \sum \lambda_\alpha y_\alpha^2 &= 1 * 0^2 + 1 * 0^2 + 1 * (-1)^2 + 1 * (+1)^2 \neq 0 \end{aligned}$$

Verallgemeinerte Kovarianzen k -ter Ordnung

Die Theorie zeigt, daß die Varianz der verallgemeinerten Differenzen k -ter Ordnung mit einer Funktion $K(h)$, der sogenannten verallgemeinerten Kovarianzfunktion k -ter Ordnung, ausgedrückt werden kann. Für $k = 0$, den Fall einfacher Zuwächse, ist die verallgemeinerte Kovarianz 0-ter Ordnung bis auf das Vorzeichen identisch mit dem Semivariogramm.

Sei $\sum \lambda_\alpha Z(\mathbf{x}_\alpha)$ ein Zuwachs k -ter Ordnung; dann schreibt man seine Varianz als

$$\text{Var}\left(\sum \lambda_\alpha Z(\mathbf{x}_\alpha)\right) = \sum \sum \lambda_\alpha \lambda_\beta K(\mathbf{x}_\alpha - \mathbf{x}_\beta)$$

Diese Schreibweise ist insbesondere für den Krige-Fehler $Z - Z^*$ unter Berücksichtigung der Unverzerrtheitsbedingung anwendbar, denn $Z - Z^*$ ist selbst wieder ein Zuwachs k -ter Ordnung.

Von diesem Punkt aus gehen die Berechnungen in vollständig analoger Art voran wie unter Verwendung der gewöhnlichen Kovarianz und liefern ein klassisches Krige – System. Zur praktischen Lösung fehlt jetzt nur noch eine Schätzung und ein Modell für die verallgemeinerte Kovarianzfunktion k –ter Ordnung, so wie man vorher aus den Daten über das experimentelle Semivariogramm das benötigte Modellsemivariogramm durch Anpassung entwickelt hat.

Verallgemeinerte polynomiale Kovarianzen k –ter Ordnung

Eine verallgemeinerte Kovarianzfunktion k –ter Ordnung kann (ebenso wenig wie die gewöhnliche Varianz) nicht durch jede beliebige Funktion dargestellt werden. Die zur Darstellung geeigneten Funktionen müssen bestimmte Bedingungen erfüllen, die im wesentlichen sicher stellen, daß die Varianz eines Zuwachses k –ter Ordnung nicht–negativ ist. Diese Bedingungen sind jedoch viel schwächer als die Bedingungen zur Darstellung einer gewöhnlichen Kovarianzfunktion. Als Folge davon ist die Klasse der theoretischen Modelle sehr viel größer. Insbesondere enthält die Klasse der theoretischen verallgemeinerten Kovarianzen der k –ten Ordnung alle Polynome vom Grad kleiner gleich $2k + 2$. Die 2D Modelle kann man zusammenfassen

Drift	Ordnung k	# Nebenbedingungen	Modelle polynomialer Kovarianzen
konstant	0	0	$K(h) = C\delta + A_1 h $
linear	1	3	$K(h) = C\delta + A_1 h + A_3 h ^3$
quadratisch	2	6	$K(h) = C\delta + A_1 h + A_3 h ^3 + A_5 h ^5$

mit den Bedingungen $C \geq 0$, $A_1 \leq 0$, $A_5 \leq 0$, $A_3 \geq -10/3\sqrt{A_1A_5}$ und dem Nugget–Effekt δ .

Die Klasse dieser Modelle für $k = 1$ ist hinreichend groß, um die interessierenden praktischen Probleme beschreiben zu können; für den gewöhnlichen Fall $k = 0$ stehen sehr viel mehr Modelle (spärisches, exponentielles, etc.) als nur das lineare, welches alleine nicht ausreichen würde, zur Verfügung.

Anpassung des Modells

Es bleibt übrig, die Ordnung k und die entsprechenden Koeffizienten A_i des Kovarianzmodells zu bestimmen.

Die Idee k zu bestimmen besteht im wesentlichen darin, der Reihe nach $z^*(\mathbf{x}_0)$ für einige Orte \mathbf{x}_0 , für die $z(\mathbf{x}_0)$ bekannt ist, aus seinen Nachbarwerten so zu schätzen, daß nur die Unverzerrtheitsbedingungen geändert werden, d.h. die Ordnungen $k = 0, 1, 2, \dots$ werden der Reihe nach durchgespielt. Die Fehler $Z(\mathbf{x}_0) - Z^*(\mathbf{x}_0)$ werden bewertet und der Wert für k als wahr angenommen, für den der bewertete Fehler am günstigsten ist.

Die Bestimmung der zugehörigen A_i geschieht im wesentlichen durch Probieren. Zunächst versucht man solche K –Modelle, deren Koeffizienten bis auf einen verschwinden. Die Güte der Anpassung durch dieses Modell wird durch den Vergleich einer experimentellen mit einer

nach dem angenommenen Polynom berechneten Varianz gemessen. Eines dieser Modelle ist oft ausreichend.

Zur Überprüfung der Gültigkeit des Modells kann man möglicherweise zum Schluß einige Meßwerte kriegen, als ob sie unbekannt wären, indem man das gewählte k und die besten Kovarianzmodelle der Reihe nach anwendet. Danach kann man sich endgültig auf eine Modell festlegen.

8.7 Weitere Kriging–Varianten

- Faktoren–Kriging

Zerlegung der praktisch meßbaren Zufallsfunktion $Z(\mathbf{x})$ gemäß der sie beschreibenden geschachtelten Struktur ihrer räumlichen Variabilität, d.h. gemäß des geschachtelten Semivariogramms, als Linearkombination ursächlicher Faktoren, die möglicherweise nicht praktisch meßbar sind.

8.8 Zusammenfassung: Punktkriging

8.8.1 Kriging–Gewichte

Der Vektor der OK Gewichte wird nach

$$\begin{pmatrix} \lambda_{\beta}^{OK}(\mathbf{x}) \\ \mu_{OK}(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}) & 1 \\ 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} C(\mathbf{x}_{\alpha} - \mathbf{x}) \\ 1 \end{pmatrix}$$

berechnet. Dieses Gleichungssystem berücksichtigt die Nachbarschaft zwischen den Orten \mathbf{x}_{α} und \mathbf{x} durch den Term $C(\mathbf{x}_{\alpha} - \mathbf{x})$ und die Redundanz in den Daten durch $C(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta})$. Dies geschieht nicht durch das gewöhnliche Astandsmaß vermittelt durch die Euklidische Metrik, sondern durch das verallgemeinerte Abstandsmaß vermittelt durch die Kovarianzfunktion.

Journel, A.G., Fundamentals of Geostatistics in Five Lessons: AGU (1989), pp. 13-14

Data orthogonalization

In most practical situations, the n data Z_{α} , $\alpha = 1, \dots, n$ are dependent on each other. The idea is to define from the initial set of dependent data Z_{α} , $\alpha = 1, \dots, n$, a new set of independent data Y_{α} , $\alpha = 1, \dots, n$, on which the simpler regression (8.16) could be applied.

Given any initial data vector: $\underline{Z}_{\alpha}^T = [Z_1, \dots, Z_n]$ with non-diagonal covariance matrix \underline{K} , there exists a linear transform $\underline{Y}_{\alpha} = \underline{A} \cdot \underline{Z}_{\alpha}$ which will transform \underline{Z}_{α} into a vector \underline{Y}_{α} with an identity diagonal covariance matrix, i.e., the n transformed RV's \underline{Y}_{α} are uncorrelated.

The initial matrix \underline{K} being a covariance matrix always admits the following LU decomposition:

$$\underline{K} = \underline{L} \cdot \underline{U} \quad (8.24)$$

with: \underline{L} being a lower triangular matrix $[\backslash^0]$

and: $\underline{U} = \underline{L}^T$ being an upper triangular matrix $[0\backslash]$, [Golub and van Loan 1983, p.54].

The inverse of \underline{K} always exists and is written:

$$\underline{K}^{-1} = \underline{U}^{-1} \cdot \underline{L}^{-1} = \underline{L}^{-1} \cdot \underline{U}^{-1}, \text{ since } \underline{K} \text{ and } \underline{K}^{-1} \text{ are symmetric}$$

Consider the linear transform $\underline{A} = \underline{L}^{-1}$:

$$\underline{Y}_\alpha = \underline{L}^{-1} \cdot \underline{Z}_\alpha \Rightarrow \underline{Y}_\alpha^T = \underline{Z}_\alpha^T \cdot \underline{U}^{-1} \quad (8.25)$$

The covariance matrix of the \underline{Y}_α 's is written:

$$\begin{aligned} \underline{K}_Y &= E \{ \underline{Y}_\alpha \cdot \underline{Y}_\alpha^T \} = E \{ \underline{L}^{-1} \underline{Z}_\alpha \underline{Z}_\alpha^T \underline{U}^{-1} \} \\ &= \underline{L}^{-1} \cdot E \{ \underline{Z}_\alpha \underline{Z}_\alpha^T \} \cdot \underline{U}^{-1} \\ &= \underline{L}^{-1} \underline{K} \underline{U}^{-1} = \underline{L}^{-1} (\underline{L} \underline{U}) \underline{U}^{-1} = \underline{I} \end{aligned}$$

Thus the transform $\underline{Y}_\alpha = \underline{L}^{-1} \cdot \underline{Z}_\alpha$ “uncorrelates” the data. The new data, however, remains correlated to the unknown RV Z_0 , with covariance matrix:

$$\underline{k}_Y = E \{ Z_0 \underline{Y}_\alpha \} = \underline{L}^{-1} \cdot E \{ Z_0 \underline{Z}_\alpha \} = \underline{L}^{-1} \cdot \underline{k}$$

Consider the SK estimate based on the \underline{Y}_α data:

$$[Z_0 - m_0]^{**} = \sum_{\alpha=1}^n \nu_\alpha [Y_\alpha - m_{Y_\alpha}] = \underline{\nu}^T \cdot \underline{L}^{-1} \cdot \underline{R}_\alpha$$

The corresponding SK system (14)³ is written, with $\underline{K}_Y = \underline{I}$

$$\underline{I} \cdot \underline{\nu} = \underline{k}_Y \Rightarrow \underline{\nu} = \underline{k}_Y$$

Thus:

$$[Z_0 - m_0]^{**} = \underline{k}_Y^T \cdot \underline{L}^{-1} \cdot \underline{R}_\alpha = \underline{k}^T \cdot (\underline{U}^{-1} \underline{L}^{-1}) \cdot \underline{R}_\alpha = \underline{k}^T \cdot \underline{K}^{-1} \cdot \underline{R}_\alpha$$

i.e. $[Z_0 - m_0]^{**} \equiv [Z_0 - m_0]^*$, as defined in (15)⁴.

Thus, the SK algorithm as described by the system (14) is the sequence of two steps:

³Die Gleichung (14) lautet

$$\underline{K} \underline{\lambda} = \underline{k} \iff \underline{\lambda} = \underline{K}^{-1} \cdot \underline{k}$$

⁴Die Gleichung (15) lautet

$$[Z_0 - m_0]^* = \underline{\lambda}^T \cdot \underline{R}_\alpha = \underline{k}^T \cdot \underline{K}^{-1} \cdot \underline{R}_\alpha$$

mit der Matrix der Residuen

$$\underline{R}_\alpha = [Z_1 - m_1, \dots, Z_n - m_n]$$

- (i) an orthogonalization of the data

$$\underline{Y}_\alpha = \underline{L}^{-1} \cdot \underline{Z}_\alpha, \text{ with: } \underline{K} = \underline{L} \cdot \underline{L}^T$$

- (ii) a straight weighting of the orthogonal data \underline{Y}_α by their correlation with the unknown Z_0 :

$$[Z_0 - m_0]^* = \underline{k}_Y^T \cdot [\underline{Y}_\alpha - m_{\underline{Y}_\alpha}]$$

The first step accounts for the redundancy between the data Z_α . Once that redundancy is corrected, a straightforward correlation weighting, of type (8.16), is performed.

Die aus den Daten abgeleitete bzw. an die Daten angepaßte Z -Kovarianzfunktion bestimmt also die Nachbarschaftsverhältnisse zwischen den Daten, indem sie die Struktur der räumlichen Variabilität beschreibt, und die Gewichte, mit denen Nachbarn im Schätzalgorithmus versehen werden. Dabei hängen die Krige-Gewichte nur von der Gestalt der Kovarianzfunktion bzw. des Semivariogramms, also vom (relativen) Nugget-Effekt, von der Anisotropie und Reichweite, ab, aber nicht vom globalen Schwellenwert oder von einem Faktor, mit dem die Kovarianzfunktion oder das Semivariogramm multipliziert werden können.

Wackernagel, H., Multivariate Geostatistics: Springer (1995), pp. 109-110

Sensitivity to choice of variogram model

The most important characteristic for the choice of the variogram model is the interpretation of the behavior at the origin. The type of continuity assumed for the regionalized variable under study has immediate implications for ordinary kriging.

The behavior at the origin can be of three types

- (i) *discontinuous*, i.e. with a nugget-effect component. The presence of a nugget-effect has the consequence that the kriging estimates will be also discontinuous: each data location will be a point of discontinuity of the estimated surface.
- (ii) *continuous, but not differentiable*. For example a linear behavior near the origin implies that the variogram is not differentiable at the origin. It has the effect that the kriging estimator is not differentiable at the data locations and that the kriged surface is linear in the immediate neighborhood of the data points.
- (iii) *continuous and differentiable*. A quadratic behavior at the origin will generate a quadratic behavior of the kriged surface at the data locations. Models that are differentiable at the origin have the power to extrapolate outside the range of the data in kriging. This implies negative ordinary kriging weights.

Most variogram models are fairly robust with respect to kriging. There however is one pathological model: The so-called “Gaussian” variogram model. This model belongs to the family of *stable variogram models* (that bear this name because of their analogy with the characteristic function of the stable distributions)

$$\gamma(\mathbf{h}) = b \left(1 - e^{-\frac{|\mathbf{h}|^p}{a}} \right) \text{ with } 0 < p \leq 2 \text{ and } a, b > 0$$

For a power p equal to 2 we have the Gaussian variogram model

$$\gamma(\mathbf{h}) = b \left(1 - e^{-\frac{|\mathbf{h}|^2}{a}} \right)$$

which is infinitely differentiable at the origin. The implication is that the associated random function also has this property. This means, that the function is analytic and, if known over a small area, it could be predicted just anywhere as all derivatives are known. Clearly this model is unrealistic in most applications (besides the fact that it is purely deterministic). Furthermore the use of such an unrealistic model has consequences on kriging which are illustrated in the following example.

8.8.2 Schätz–Nachbarschaft

Zur Bestimmung einer geeigneten Schätz–Nachbarschaft sollte man berücksichtigen,

- daß die Zuverlässigkeit der Z –Kovarianzfunktion für wachsende Schrittweiten h abnimmt,
- daß die Benutzung einer lokalen Schätz–Nachbarschaft es erlaubt, die Voraussetzung der globalen Stationarität in die lokaler Stationarität abzumildern,
- daß der screen–Effekt den Einfluß weiter entfernter Meßwerte zusätzlich vermindert,
- daß die numerische Effizienz sich wie $(n(\mathbf{x}))^3$ verhält.

8.8.3 Kriging–Varianz

Die Kriging–Varianz

- ist abhängig von der Z –Kovarianzfunktion,

- ist abhängig von der geometrischen Konfiguration der Meßwerte,
- unabhängig von den Meßwerten.

Kriging führt zu sehr ähnlichen Schätzvarianzen für Schätzwerte mit gleicher lokaler Konfiguration der Meßwerte.

Intuitiv sollte die Schätzvarianz bei gleicher lokaler Konfiguration der Meßwerte für sehr unterschiedliche Meßwerte größer sein als für sehr ähnliche Meßwerte!

8.8.4 Vergleich Meßwerte – geschätzte Werte

Glättungseffekt von Kriging:

- Die geschätzten Werte zeigen weniger Variabilität als die Meßwerte.
- Große Konzentrationen werden tendenziell unterschätzt, kleine Konzentrationen werden tendenziell überschätzt. Diese bedingte Verzerrtheit führt zu häufigen Fehlklassifizierungen.

Fast verschwindende Rangkorrelation von tatsächlichem Schätzfehler und Kriging–Varianz zeigt an, daß die Kriging–Varianz alleine ein unzureichendes Maß für die Güte der Schätzung vermittelt.

8.9 Tips für S–PLUS SPATIAL (5.Übung)

Experimentelles logarithmisches Semivariogramm (Aufgabe 3 b)

Bildung des natürlichen Logarithmus von Cd und Abspeichern dieser neuen Variable lnCd in der Datei jura259

S>jura259\$lnCd<-log(Cd)

Berechnung des experimentellen Variogramms von lnCd, Verfahrensweise analog zur dritten Übung

S>module(spatial)
S>lnCd.var<-variogram(lnCd ~ loc(X,Y), data=jura259, azimuth=90,
+ tol.azimuth=90, nlag=10, lag=0.2, tol.lag=0.1, maxdist=2.0,
+ na.action='na.omit')

*Anpassung eines Variogramm-Modells, 3 mögliche Varianten: Anpassung eines ungeschachtelten Modells in S–PLUS mit der Funktion **model.variogram** (siehe dritte*

Übung), oder Anpassung eines geschachtelten Modells in S-PLUS durch Verknüpfung der Funktionen `plot` und `lines` (siehe unten), oder Auslagerung der Daten (Koordinaten und Variablenwerte) aus S-PLUS in eine ASCII-Datei und Durchführung der Modellanpassung mit VARIOWIN (siehe dritte und zweite Übung)

1. Öffnen eines Grafik-Fensters

`S>trellis.device(win.graph)`

2. Darstellung des experimentellen Variogramms von `lnCd`

`S>plot(lnCd.var)`

3. Erzeugung eines Vektors (einer Datei `vdist`) der Abstände, für die das Variogramm-Modell berechnet werden soll

`S>vdist<-c(seq(0.01,0.09,by=0.01),seq(0.1,2.0,by=0.1))`

4. Erzeugung des (geschachtelten) Variogramm-Modells, die mit ??? markierten Stellen müssen durch Zahlenwerte ersetzt werden

`S>lnCd.varmodel<-spher.vgram(vdist,range=???,sill=???,
+ nugget=???) +spher.vgram(vdist,range=???,sill=???)`

5. Darstellung des Variogramm-Modells auf der Grafik experimentellen Variogramms

`S>lines(vdist,lnCd.varmodel)`

6. ggf. Verbesserung des Modells durch mehrfache Wiederholung der letzten beiden Schritte; um die Linien verworfener Modelle vom Bild zu löschen, muß außerdem vor der Eingabe von `lines(...)` der `plot(lnCd.var)` Befehl wiederholt werden

vergleichende Darstellung der experimentellen Variogramme von `Cd` und `lnCd` sowie ihrer Modelle `Cd.varmodel` und `lnCd.varmodel`

1. nochmalige Erzeugung des experimentellen Variogramms (siehe dritte Übung, Aufgabe 4)

`S>Cd.var<-variogram(Cd ~ loc(X,Y), data=jura259,
+ azimuth=90,tol.azimuth=90,nlag=10,lag=0.2,tol.lag=0.1,
+ maxdist=2.0,na.action='na.omit')`

und des geschachtelten Variogramm-Modells für `Cd` (vierte Übung, Aufgabe3);

wenn nur eine Struktur (ungeschachtelt) angepaßt werden soll, dann nur ersten Teil der folgenden Formel mit entsprechend geänderten `range` und `sill` Parametern eingeben

`S>Cd.varmodel<-spher.vgram(vdist,range=0.2,sill=0.3,
+ nugget=0.3) +spher.vgram(vdist,range=1.3,sill=0.26)`

2. Vorbereitung des Bildschirms zur gleichzeitigen Darstellung von 2 Bildern

S>par(mfrow=c(2,1))

3. Darstellung des exp. Variogramms von Cd

S>plot(Cd.var, type='b')

4. Überlagerung des Bildes mit dem Variogramm-Modell für Cd

S>lines(vdist,Cd.varmodel)

5. Darstellung des exp. Variogramms von lnCd

S>plot(lnCd.var, type='b')

4. Überlagerung des Bildes mit dem Variogramm-Modell für lnCd

S>lines(vdist,lnCd.varmodel)

lognormal kriging, Aufgabe 4

Lognormales kriging ist kriging der logarithmierten Variable

$$Y(\mathbf{x}) = \ln Z(\mathbf{x})$$

mit anschließender Rücktransformation der Krigingergebnisse (Schätzwert und Schätzvarianz) auf die Skala der nichtlogarithmierten Variable.

Es gilt

$$\begin{aligned} m_Z &= \exp\left(m_Y + \frac{\sigma_Y^2}{2}\right) \\ \sigma_Z^2 &= m_Z^2 (\exp(\sigma_Y^2) - 1) \\ C_Z(\mathbf{h}) &= m_Z^2 (\exp(C_Y(\mathbf{h})) - 1) \end{aligned}$$

Für die Rücktransformation darf nicht einfach die dem Logarithmieren entgegengesetzte Operation des Exponentierens auf den Schätzwert angewendet werden, da der Schätzwert kein einzelner Wert, sondern eine gewichtete Summe ist. Der Schätzwert $Z_K^*(\mathbf{x})$ und die Schätzvarianz $\sigma_{Z_K^*}^2(\mathbf{x})$ auf der Skala der nichtlogarithmierten Variable Z , ausgedrückt mit Hilfe des einfachen Krigeschätzwertes $Y_K^*(\mathbf{x})$ und der Schätzvarianz $\sigma_{Y_K^*}^2(\mathbf{x})$ des einfachen krigings der logarithmierten Variable Y , werden nach folgenden Formeln berechnet:

- Krigeschätzwert $Z_{SK}^*(\mathbf{x})$

$$Z_{SK}^*(\mathbf{x}) = \exp\left(Y_{SK}^*(\mathbf{x}) + \frac{\sigma_{Y_{SK}^*}^2(\mathbf{x})}{2}\right)$$

- Krigeschätzvarianz $\sigma_{Z_{SK}^*}^2(\mathbf{x})$

$$\sigma_{Z_{SK}^*}^2(\mathbf{x}) = m_Z^2 \exp(\sigma_Y^2) \left(1 - \exp(-\sigma_{Y_{SK}^*}^2(\mathbf{x})) \right)$$

Alternativ mit Mittelwert und Stichproben–Varianz der logarithmierten Zufallsvariablen

$$m_Y = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \sum_{i=1}^n \ln z_i$$

$$\sigma_Y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - m_Y)^2$$

kann man

$$\exp(m_Y) = \exp\left(\frac{1}{n} \sum_{i=1}^n \ln z_i\right) = \left(\prod_{i=1}^n x_i\right)^{\frac{1}{n}} = m_X^{\text{geom}}$$

und

$$P(\exp(m_Y - m\sigma_Y) < \exp(Y) \leq \exp(m_Y + m\sigma_Y)), \quad m = 1, 2, 3$$

betrachten.

Vorbereitung des Kriging der logarithmierten Cd–Werte, für ??? sind die in Aufgabe 3 bestimmten Parameter einzusetzen

- für Kovarianzmodell mit einer Struktur und evtl. Nuggeteffekt
 $S > \text{lnCd.krige} \leftarrow \text{krige}(\text{lnCd} \sim \text{loc}(\mathbf{X}, \mathbf{Y}), \text{data}=\text{jura259}, \text{covfun}=\text{???},$
 $\quad + \text{range}=\text{???}, \text{sill}=\text{???}, \text{nugget}=\text{??})$
- für Kovarianzmodell mit zwei oder drei Strukturen und evtl. Nuggeteffekt; Verwendung der Funktion `krignest`, diese Funktion ist nicht standardmäßig in S-PLUS enthalten, sie muß ggf. mit dem Befehl `dget(''A:\krignest'')` von Diskette geladen werden
 $S > \text{lnCd.krignest} \leftarrow \text{krignest}(\text{lnCd} \sim \text{loc}(\mathbf{X}, \mathbf{Y}), \text{data}=\text{jura259}, \text{covfun}=\text{???},$
 $\quad + \text{range}=\text{???}, \text{sill}=\text{???}, \text{nugget}=\text{??}, \text{covfun2}=\text{??}, \text{range2}=\text{??}, \text{sill2}=\text{??},$
 $\quad + \text{covfun3}=\text{??}, \text{range3}=\text{??}, \text{sill3}=\text{??})$

Erzeugung einer Datei newXY mit den Koordinaten der Schätzpunkte
 S>newXY<-jura[260:359,1:2]

Kriging der logarithmierten Cd-Werte

S>lnCd.Pred<-predict.krige(lnCd.krignest (bzw. lnCd.krige), newXY, se.fit=T)

Berechnung des Krige-Schätzwertes auf der Skala der nichtlogarithmierten Variable Cd

S>lnCd.back<-exp(lnCd.Pred\$fit + lnCd.Pred\$se/2)

Berechnung des Krige-Schätzvarianz auf der Skala der nichtlogarithmierten Variable Cd

S>lnCd.backse<-lnCd.back * (exp(lnCd.Pred\$se) - 1)

8.10 Blockkriging – Lagerstättenvorratsberechnung

8.10.1 Geostatistische Varianzen

Die Ausdehnungsvarianz wahrer Gehalte

Der Mittelwert Z_v über einer Teilmenge $v \subset D$ soll auf den Mittelwert Z_V über einer Teilmenge $v \subset V \subset D$ “ausgedehnt” werden.

Ein Maß für die strukturelle Variabilität zwischen Z_v und Z_V ist die “Ausdehnungsvarianz”:

$$Z_v = \frac{1}{|v|} \int_v Z(\mathbf{x}) d\mathbf{x}$$

$$Z_V = \frac{1}{|V|} \int_V Z(\mathbf{x}) d\mathbf{x}$$

$$\begin{aligned} \text{Var}(Z_V - Z_v) &= \text{Cov}(Z_V, Z_V) + \text{Cov}(Z_v, Z_v) - 2\text{Cov}(Z_V, Z_v) \\ &= \frac{1}{|V|^2} \int_V \int_V C(\mathbf{x}_1 - \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 + \frac{1}{|v|^2} \int_v \int_v C(\mathbf{x}_1 - \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \\ &\quad - \frac{2}{|v| |V|} \int_V \int_v C(\mathbf{x}_1 - \mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \\ &= \bar{C}(V, V) + \bar{C}(v, v) - 2\bar{C}(v, V) \end{aligned}$$

oder

$$\text{Var}(Z_V - Z_v) = 2\bar{\gamma}(v, V) - \bar{\gamma}(v, v) - \bar{\gamma}(V, V)$$

mit

$$\bar{\gamma}(v, V) = \frac{1}{|V|} \int_V \left[\frac{1}{|v|} \int_v \gamma(\mathbf{x}_1 - \mathbf{x}_2) d\mathbf{x}_1 \right] d\mathbf{x}_2$$

als Maß für die mittlere strukturelle Variabilität zwischen v und V , relativen Abstand zwischen v und V ; $\bar{\gamma}(v, v)$, $\bar{\gamma}(V, V)$ berücksichtigt die Geometrie von v bzw. V .

Die Verteilungsvarianz oder Dispersionsvarianz wahrer Gehalte

Der Träger V sei zusammengesetzt aus n (kleinen) volumengleichen Trägern v_i , $i = 1, \dots, n$, d.h.

$$V = \bigcup_i v_i, \quad v_i \cap v_j = \emptyset \quad \text{falls } v_i \neq v_j \\ |v_i| = |v| \quad \text{für alle } i = 1, \dots, n$$

Setzt man

$$Z_V = \frac{1}{n} \sum_{i=1}^n Z_{v_i} \\ \text{und } S^2 = \frac{1}{n} \sum_{i=1}^n (Z_{v_i} - Z_V)^2,$$

dann gilt für den Erwartungswert der Zufallsvariablen S^2 , deren Realisierung als experimentelle Varianz s^2 interpretiert wird:

$$\begin{aligned} E(S^2) &= E\left(\frac{1}{n} \sum_i (Z_{v_i} - Z_V)^2\right) \\ &= \frac{1}{n} \sum_i [\bar{C}(V, V) + \bar{C}(v_i, v_i) - 2\bar{C}(v_i, V)] \\ &= \bar{C}(V, V) + \bar{C}(v, v) - 2\bar{C}(V, V) \\ &= \bar{C}(v, v) - \bar{C}(V, V). \end{aligned}$$

Dieser wird als Varianz der Verteilung und mit $\sigma^2(v/V)$ bezeichnet (Abb.8.1).

$$\sigma^2(v/V) = \bar{\gamma}(V, V) - \bar{\gamma}(v, v) = D^2(v, V)$$

Die Verteilungsvarianz $\sigma^2(v/V)$ ist gleich dem Mittelwert bezüglich i der Ausdehnungsvarianzen $\text{Var}(Z_V - Z_{v_i})$ von v_i nach V .

$$\begin{aligned} D^2(o/V) = \sigma^2(o/V) &= \bar{\gamma}(V, V) - \bar{\gamma}(o, o) \\ D^2(V/D) = \sigma^2(V/D) &= \bar{\gamma}(D, D) - \bar{\gamma}(V, V) \end{aligned}$$

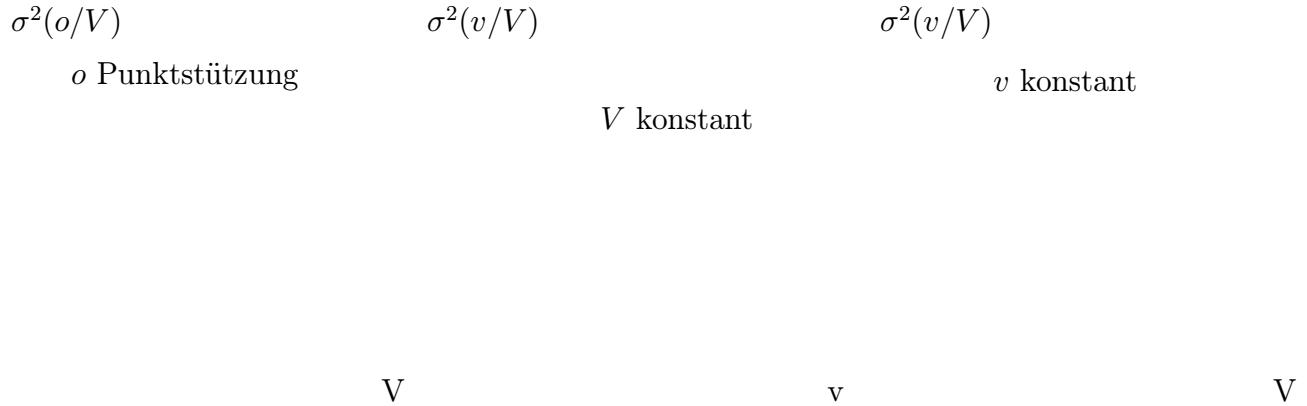


Figure 8.1: Schematische Darstellung der Volumen-Varianz-Beziehung

$$\begin{aligned} D^2(o/V) + D^2(V/D) &= D^2(o/D) \\ \implies D^2(o/D) &> D^2(V/D) \end{aligned}$$

Daraus folgt unmittelbar die “Volumen-Varianz-Beziehung”:

$$\sigma^2(o/D) = \sigma^2(o/V) + \sigma^2(V/D) \quad \text{für } o \subset V \subset D$$

und weiter

$$\sigma^2(o/D) > \sigma^2(V/D)$$

Weiterhin erkennt man hier einen Beziehungspunkt zur klassischen Statistik. Für einen Punkt o gilt nämlich

$$\begin{aligned} \sigma^2(0/D) &= \bar{\gamma}(D, D) \\ \lim_{D \rightarrow \infty} \sigma^2(o/D) &= \lim_{h \rightarrow \infty} \gamma(h) = C(0) = \text{Var}(Z) \end{aligned}$$

falls das Semivariogramm einen Schwellenwert besitzt.

Insgesamt stellt das Konzept der Verteilungsvarianz einen quantitativen Ausdruck für die Erfahrung dar, daß und wie die Variabilität von der Stützung abhängt.

8.10.2 Schätzprobleme und Schätzvarianzen

Ein einfaches Schätzproblem

Der Mittelwert

$$Z_V = \frac{1}{|V|} \int_V Z(\mathbf{x}) d\mathbf{x} \quad V \subset D$$

soll durch eine geeignete Linearkombination von $Z(\mathbf{x}_i)$, $\mathbf{x}_i \in V$, geschätzt werden

$$Z^* = \sum_i \lambda_i Z(\mathbf{x}_i), \quad \lambda_i \in I\mathbb{R}$$

Diese Schätzung ist mit einem Fehler behaftet, für dessen Varianz gilt

$$\begin{aligned} \text{Var}(Z_V - Z^*) &= \text{Cov}(Z_V, Z_V) + \text{Cov}(Z^*, Z^*) - 2 \text{Cov}(Z_V, Z^*) \\ &= \bar{C}(V, V) + \sum_i \sum_j \lambda_i \lambda_j C(\mathbf{x}_i - \mathbf{x}_j) - \frac{2}{|V|} \sum_i \lambda_i \int_V C(\mathbf{x}_i - \mathbf{x}) d\mathbf{x} \end{aligned}$$

Diese Varianz wird als Schätzvarianz und mit $\sigma_E^2(Z_V, Z^*)$ bezeichnet:

$$\sigma_E^2(Z_V, Z^*) = \frac{2}{|V|} \sum_i \lambda_i \int_V \gamma(\mathbf{x}_i - \mathbf{x}) d\mathbf{x} - \sum_i \sum_j \lambda_i \lambda_j \gamma(\mathbf{x}_i - \mathbf{x}_j) - \bar{\gamma}(V, V).$$

Ein elementares Schätzproblem

Im Zusammenhang mit Schätzproblemen wird eine neue Interpretation des Variogramms möglich. Soll nämlich $Z = Z(\mathbf{x}_i + \mathbf{h})$ geschätzt werden durch $Z^* = Z(\mathbf{x}_i)$, dann gilt im intrinsischen Fall für die Schätzvarianz

$$\text{Var}(Z - Z^*) = \text{Var}(Z(\mathbf{x}_i + \mathbf{h}) - Z(\mathbf{x}_i)) = 2\gamma(h)$$

d.h. das Semivariogramm kann als fundamentale Schätzvarianzfunktion interpretiert werden.

Ein typisches Schätzproblem

Der Mittelwert $Z_V(\mathbf{x}_0)$ über einen Träger V an der Stelle \mathbf{x}_0

$$Z_V(\mathbf{x}_0) = \frac{1}{|V|} \int_V Z(\mathbf{x}) d\mathbf{x}$$

soll durch eine endliche Linearkombination Z_V^* von Mittelwerten $Z_{v_i}(\mathbf{x}_i)$ über Trägern v_i an den Stellen \mathbf{x}_i

$$\begin{aligned} Z_V^*(\mathbf{x}_0) &= \sum_i \lambda_i Z_{v_i}(\mathbf{x}_i) \\ &= \sum_i \lambda_i \frac{1}{|v_i|} \int_{v_i} Z(\mathbf{x}) d\mathbf{x} \end{aligned}$$

so geschätzt werden, daß die Schätzung (i) unverzerrt und (ii) ihre Schätzvarianz minimal ist, d.h.

$$\begin{aligned} \mathbb{E}(Z_V - Z_V^*) &\stackrel{!}{=} 0 \\ \sigma_E^2 = \text{Var}(Z_V - Z_V^*) &\stackrel{!}{=} \min = \sigma_K^2 \end{aligned}$$

Chapter 9

Modellierung lokaler Ungewißheit durch Indikator Kriging

Während zunächst die Herleitung und Anwendung eines optimalen Schätzers und seiner Schätzfehler–Varianz im Mittelpunkt stand, wendet man sich nun der Modellierung der Ungewißheit bezüglich der unbekannten Größe zu. Solche Modelle versetzen den Anwender dann in die Lage, das Risiko im Zusammenhang mit einem Entscheidungsprozeß abzuschätzen, z.B. das Risiko der Fehlklassifizierung kontaminiert oder nichtkontaminiert Böden. Die Modellierung der Ungewißheit führt auch zu neuen Schätzern, die optimal bezüglich anderer Kriterien als dem willkürlichen “kleinste – Quadrate” Kriterium sind.

9.1 Ziele der Modellierung der lokalen Ungewißheit

Bestimmung und kartenmäßige Darstellung

- der Wahrscheinlichkeiten, daß lokal ein vorgegebener Wert übertroffen wird;
- des Risikos der Fehlklassifizierung;
- des Bedarfs zusätzlicher Information.

Fehlklassifizierung der kleinen Werte, falls $Z^* > t$

$$\text{risk } \alpha(Z^*, t) = P(Z \leq t | Z^* > t; (n))$$

und der großen Werte, falls $Z^* < t$,

$$\text{risk } \beta(Z^*, t) = P(Z > t | Z^* \leq t; (n))$$

Wenn Z^* zunimmt, dann wird das Risiko β kleiner und α größer, wenn Z^* abnimmt, dann wird das Risiko β größer und α kleiner.

Theoretische Möglichkeiten der Modellierung der Ungewißheit bestehen durch

- (i) Konfidenz Intervalle
- (ii) bedingte Verteilungen
 - multiGauß Zugang
 - Indikator Zugang

1. ist der durch die klassische univariate Statistik eröffnete Zugang. Er setzt voraus, daß,

- (i) der Schätzfehler $Z^*(\mathbf{x}) - Z(\mathbf{x})$ die Realisation einer normalverteilten Zufallsvariable ist,
- (ii) die Schätzfehler Varianz $\sigma_E^2(\mathbf{x})$ unabhängig von den Daten ist.

(i) impliziert insbesondere die Symmetrie der Verteilung, diese widerspricht den Beobachtungen, daß große Werte eher unterschätzt und kleine Werte eher überschätzt werden, (ii) ist zwar erfüllt, widerspricht aber der Intuition, daß die Schätzfehler Varianz von den Daten abhängig sein sollte.

Dieser Zugang ist mit den Einschränkungen (i) und (ii) Bestandteil des Kriging Paradigmas, zunächst einen optimalen Schätzwert und dann das zugehörende Konfidenzintervall anzugeben.

2. vermittelt einen gänzlich neuen Zugang: Aus der Modellierung der Ungewißheit wird ein angepaßt optimaler Schätzer entwickelt.

Der zentrale Begriff dieses Zugangs ist die lokale Verteilung bzw. die bedingte Verteilung.

$\{Z(\mathbf{x}), \mathbf{x} \in D\}$ bezeichnet wieder die Zufallsfunktion, $z(\mathbf{x}_\alpha)$, $\alpha = 1, \dots, n$ die Daten. Die bedingte Verteilungsfunktion

$$F(\mathbf{x}; z|(n)) = P(Z(\mathbf{x}) \leq z|(n)),$$

die durch die den Meßwerten \mathbf{x}_α , $\alpha = 1, \dots, n$ innewohnende Information (n) bedingt ist, modelliert die lokale Ungewißheit

$$\begin{aligned} P(Z(\mathbf{x}) \in [a, b]|(n)) &= F(\mathbf{x}; b|(n)) - F(\mathbf{x}; a|(n)) \\ P(Z(\mathbf{x}) > b|(n)) &= 1 - F(\mathbf{x}; b|(n)) \end{aligned}$$

Diese Aussagen über die lokale Ungewißheit sind unabhängig von irgendeinem Schätzwert!

Eine Folge von bedingten Ein-Punkt Verteilungen $(F(\mathbf{x}_i, z|(n)), i = 1, \dots, I)$ modelliert nicht die gemeinsame räumliche Ungewißheit, d.h. vermittelt keine Kenntnis der gemeinsamen Mehr-Punkt Verteilungen $P(Z(\mathbf{x}_1) \leq 1, \dots, Z(\mathbf{x}_I) \leq z_I|(n))$. (Dies ist in perfekter Analogie zur klassischen multivariaten Statistik; die Randverteilungen erlauben keine Aussagen über die gemeinsamen Verteilungen.)

Aussagen über die räumlich (globale) Ungewißheit kann man durch bedingte Simulation erreichen.

9.2 Lokale Verteilungen

In der Praxis hat man oft vor Messungen Informationen über mögliche Werte einer Messung, die man als Realisationen einer Zufallsvariablen oder Zufallsfunktion interpretiert, d.h. Informationen der Form

$$z(\mathbf{x}) \in [z_{\min}, z_{\max}]$$

Aufgrund allgemeiner Kenntnisse und Informationen ergibt sich beispielsweise für die Cd-Konzentrationen im Untersuchungsgebiet

$$z(\mathbf{x}) \in [0, 6] \text{ in ppm}$$

Ohne zusätzliche Information müßte man von einer lokalen Gleichverteilung ausgehen, also

$$F(\mathbf{x}; z) = \begin{cases} 0 & z \leq 0 \\ z/6 & z \in (0, 6] \\ 1 & z > 6 \end{cases}$$

unabhängig von $\mathbf{x} \in D$. Daraus folgte

$$\begin{aligned} P(Z(\mathbf{x}) > 0.8) &= 1 - F(\mathbf{x}; 0.8) \\ &= 0.83 \end{aligned}$$

Die Annahme der Stationarität erlaubt, die Stichprobenverteilung $F_n(z)$ als globale Information einzuführen und die lokalen Verteilungen gegenüber der Gleichverteilungsannahme so zu korrigieren

$$F(\mathbf{x}; z) = P(Z \leq z) = F_n(z)$$

daß folgt

$$\begin{aligned} P(Z(\mathbf{x}) > 0.8) &= \frac{\#\{z(\mathbf{x}_i) > 0.8\}}{n} \\ &= \frac{\#\{z_i > 0.8\}}{n} \end{aligned}$$

Bis jetzt ist das Modell der lokalen Ungewißheit noch unabhängig vom Ort, also noch nicht lokal. Die Ortsabhängigkeit wird im nächsten Schritt sinngemäß so eingeführt, daß die lokale Wahrscheinlichkeit, daß die Cd-Konzentration einen gegebenen Wert überschreitet, in der Umgebung großer Meßwerte [größer als der gegebene Wert] größer ist als in der Umgebung kleiner Meßwerte [kleiner als der gegebene Wert].

Die räumliche Abhängigkeit soll also ausgenutzt werden, um das Modell der Ungewißheit lokal, d.h. ortsabhängig von den Meßwerten der Umgebung zu machen. Dazu schreibt man ausführlich:

$$F(\mathbf{x}; z|n(\mathbf{x}))$$

oder kürzer

$$F(\mathbf{x}; z|(n)) .$$

9.3 Indikator Zugang – Nichtparametrische Geostatistik

Den Zugang zu $F(\mathbf{x}; z|(n))$ können Indikator transformierte Zufallsvariablen bzw. Meßwerte bereitstellen.

Es sei erinnert, daß

$$i(\mathbf{x}_\alpha; z_k) = \begin{cases} 1 & \text{falls } z(\mathbf{x}_\alpha) \leq z_k \\ 0 & \text{sonst} \end{cases}$$

$$I(\mathbf{x}; z) = \begin{cases} 1 & \text{falls } Z(\mathbf{x}) \leq z \\ 0 & \text{sonst} \end{cases}$$

und daß aus

$$P(Z(\mathbf{x}) \leq z) = F(\mathbf{x}; z)$$

folgt

$$E(I(\mathbf{x}; z)) = 1 \times P(Z(\mathbf{x}) \leq z) + 0 \times P(Z(\mathbf{x}) > z) \\ = P(Z(\mathbf{x}) \leq z)$$

$$E(I(\mathbf{x}; z_1); I(\mathbf{x} + h; z_2)) = F(h; z_1, z_2) \\ = P(Z(\mathbf{x}) \leq z_1, Z(\mathbf{x} + h) \leq z_2)$$

Man stellt sich vor, die Funktion $F(\mathbf{x}; z|(n))$ durch eine Folge von diskreten Werten

$$F(\mathbf{x}; z_k|(n)) = P(Z(\mathbf{x}) \leq z_k|(n)), \quad k = 1, \dots, K$$

zu approximieren.

Der Indikatorzugang basiert wesentlich auf

$$F(\mathbf{x}; z_k|(n)) = E(I(\mathbf{x}; z_k)|(n))$$

Man kann mathematisch beweisen, daß der Kriging–Schätzwert für den Indikatorwert $i(\mathbf{x}; z_k)$ zugleich der kleinste–Quadrate Schätzwert seines bedingten Erwartungswertes ist. Daraus folgt, daß man die Werte $F(\mathbf{x}; z_k|(n))$ als Kriging Schätzwerte des unbekannten Indikators $i(\mathbf{x}; z_k)$ unter Berücksichtigung benachbarter Indikator–transformierter Meßwerte verstehen kann.

9.3.1 Richtlinien zur Wahl der thresholds $z_k, \quad k = 1, \dots, K$

- Dezentile der empirischen (Stichproben) Verteilungsfunktion

- kritische Werte, z.B. gesetzliche Grenzwerte, als threshold berücksichtigen. Dann braucht man für diesen Wert nicht zu approximieren/interpolieren.
- Anzahl der thresholds in kritischen Bereichen, z.B. größer als gesetzlicher Grenzwert, erhöhen
- threshold-Werte kleiner als das erste Dezentil und größer als das neunte Dezentil sollte man vermeiden.

A.G.JURNEL: Do not go beyond 80% percentile in indicator kriging; never beyond the largest threshold!

Nach der Wahl der K threshold Werte z_1, \dots, z_K wird jede zur Verfügung stehende Information bezüglich dieser thresholds Indikator-transformiert, d.h. aus jeder Information wird ein K -Vektor $(i(\mathbf{x}, z_1), \dots, i(\mathbf{x}, z_K))$ erstellt.

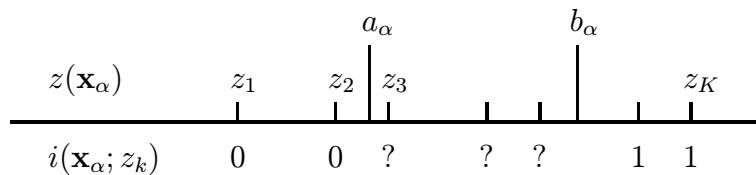
Indikator-transformierte der Meßwerte $z_\alpha, \alpha = 1, \dots, n$ haben die Form

$$i(\mathbf{x}_\alpha; z_k) = \begin{cases} 1 & \text{falls } z(\mathbf{x}_\alpha) \leq z_k \quad k = 1, \dots, K \\ 0 & \text{sonst} \end{cases}$$

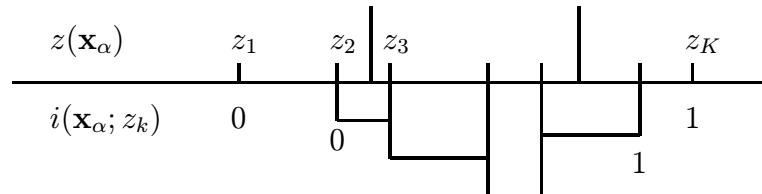
Aus einem Meßwert $z(\mathbf{x}_\alpha)$ wird also eine Folge $(0, 0, \dots, 1, 1, 1)$.

Indikator-transformierte Intervall-Information der Form $z(\mathbf{x}_\alpha) \in (a_\alpha, b_\alpha]$ hat die Form

$$i(\mathbf{x}_\alpha; z_k) = \begin{cases} 1 & b_\alpha \leq z_k \\ \text{NA} & \\ 0 & a_\alpha > z_k \end{cases}$$



Indikator-transformierte mittelbare Information (soft data) hat die Form von fuzzy infor-



mation:

Werte zwischen 0 und 1, die ein Maß für die Möglichkeit vermitteln, daß $z(\mathbf{x}_\alpha) \in (z_{k-1}, z_k]$

9.3.2 Indikatorkriging

Ziel ist

$$F(\mathbf{x}; Z_k|(n)) = P(Z(\mathbf{x}) \leq z_k|(n))$$

für einen beliebigen Ort $\mathbf{x} \in D$ zu bestimmen, wobei die bedingende Information (n) aus Informationen in der Schätzumgebung $W(\mathbf{x})$ um \mathbf{x} besteht. Der Kriging Schätzwert i^* für den Indikator i kann als Modell der bedingten lokalen Verteilung herangezogen werden.

$$F^*(\mathbf{x}; Z_k|(n)) = E^*(I(\mathbf{x}; z_k)|(n)) = i_{kriging}^*(\mathbf{x}; z_k)$$

Für den Fall, daß nur Indikator-transformierte Meßwerte in der Schätzumgebung berücksichtigt werden sollen, erhält man für den Kriging Indikator Schätzer

$$[I^*(\mathbf{x}; z_k) - E(I(\mathbf{x}; z_k))] = \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}(\mathbf{x}; z_k) [I(\mathbf{x}_{\alpha}; z_k) - E(I(\mathbf{x}_{\alpha}; z_k))]$$

Einfaches Indikator Kriging (sIK)

Es gilt

$$E(I(\mathbf{x}; z_k)) = F_n(z_k)$$

und damit

$$\begin{aligned} F_{sIK}^*(\mathbf{x}; z_k|(n)) &= I_{SK}^*(\mathbf{x}; z_k) = F_n(z_k) + \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}^{SK}(\mathbf{x}; z_k) [I(\mathbf{x}_{\alpha}; z_k) - F_n(z_k)] \\ &= \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}^{SK}(\mathbf{x}; z_k) I(\mathbf{x}_{\alpha}; z_k) + \left(1 - \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}^{SK}(\mathbf{x}; z_k)\right) F(z_k) \\ &= \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}^{SK}(\mathbf{x}; z_k) I(\mathbf{x}_{\alpha}; z_k) + \lambda_m^{SK}(\mathbf{x}; z_k) F(z_k) \end{aligned}$$

Die sIK Gewichte sind bestimmt durch ein einfaches Kriging System linearer Gleichungen

$$\sum_{\beta=1}^{n(\mathbf{x})} \lambda_{\beta}^{SK}(\mathbf{x}; z_k) C_I(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}; z_k) = C_I(\mathbf{x}_{\alpha} - \mathbf{x}; z_k), \quad \alpha = 1, \dots, n(\mathbf{x})$$

wobei $C_I(h; z_k)$ die Kovarianzfunktion der Indikator Zufallsvariablen $I(\mathbf{x}; z_k)$ zum threshold z_k ist.

Gewöhnliches Indikator Kriging (oIK)

Es gilt

$$E(I(\mathbf{x}'; z_k)) = \text{const für alle } \mathbf{x}' \in W(\mathbf{x})$$

aber unbekannt und damit

$$F_{oIK}^*(\mathbf{x}; z_k | (n)) = I_{OK}^*(\mathbf{x}; z_k) = \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}^{OK}(\mathbf{x}; z_k) I(\mathbf{x}; z_k)$$

und die oIK Gewichte sind bestimmt durch

$$\begin{aligned} \sum_{\beta=1}^{n(\mathbf{x})} \lambda_{\beta}^{OK}(\mathbf{x}; z_k) C_I(\mathbf{x}_{\alpha} - \mathbf{x}_{\beta}; z_k) + \mu_{OK}(\mathbf{x}; z_k) &= C_I(\mathbf{x}_{\alpha} - \mathbf{x}; z_k) \\ \sum_{\beta=1}^{n(\mathbf{x})} \lambda_{\beta}^{OK}(\mathbf{x}; z_k) &= 1 \end{aligned}$$

Wahrscheinlichkeitskriging (probability kriging PK)

Man setzt

$$\begin{aligned} F_{PK}^*(\mathbf{x}; z_k | (n)) &= \sum_{\alpha=1}^{n(\mathbf{x})} \lambda_{\alpha}(\mathbf{x}; z_k) I(\mathbf{x}_{\alpha}; z_k) + \sum_{\alpha=1}^{n(\mathbf{x})} \nu_{\alpha}(\mathbf{x}; z_k) Z_R(\mathbf{x}_{\alpha}) \\ Z_R(\mathbf{x}_{\alpha}) &= \frac{1}{n} \text{Rang}(Z(\mathbf{x}_{\alpha})) \\ z_R(\mathbf{x}_{\alpha}) &= \frac{1}{n} \text{Rang}(z(\mathbf{x}_{\alpha})) \end{aligned}$$

Zusammenfassung

Das Ergebnis von IK ist in jedem Fall ein Vektor, dessen Komponenten Wahrscheinlichkeiten sind, die eine diskrete Approximation der lokalen Verteilung darstellen.

9.4 Optimale Schätzer

Die klassische univariate Statistik kennt folgende Schätzer

$$\begin{aligned} P(Z \leq z_0 | (n)) &= \frac{\#\{z_j \leq z_0\}}{\#\{z_k\}} = F_n(z_0) \\ &= \frac{1}{n} \sum_{j=1}^n i(z_j; z_0) \end{aligned}$$

$$\begin{aligned}
z^{*(1)} &= \frac{1}{n} \sum_{j=1}^n z_j \\
z^{*(2)} &= F_n^{-1}(0.5) \\
z^{*(3)} &= \max_{j=1, \dots, n} z_j \\
z^{*(4)} &= \frac{\max z_j - \min z_j}{2} \\
z^{*(5)} &= \frac{1}{n-2} \sum_{j=1}^{n-2} z_j, \quad z_j \neq z_{\max}, \quad z_j \neq z_{\min}
\end{aligned}$$

Die nichtparametrische Geostatistik führt folgende Schätzer $z^*(\mathbf{x})$ von $z(\mathbf{x})$ ein. Im folgenden bezeichnet

$$e(\mathbf{x}) = z^*(\mathbf{x}) - z(\mathbf{x})$$

den Schätzfehler und $L(e)$ die Verlust Funktion (loss function), die ein Maß für die Auswirkung (impact) des Fehlers e ist, z.B.

$$L(e) = e^2$$

bewirkt, daß Fehler $e < 1$ geringer und Fehler $e > 1$ stärker bewertet werden.

Da der Schätzfehler im allgemeinen unbekannt ist, kann $L(e)$ zunächst nicht "abgelesen" werden, auch wenn L gewählt ist.

Das Modell der lokalen Ungewißheit $F(\mathbf{x}; z|(n))$ kann man jetzt anwenden, um den bedingten Erwartungswert von L

$$E(L(Z^* - Z)|(n)) = \int L(Z^* - Z) dF(\mathbf{x}; z|(n)) = \varphi(Z^*; \mathbf{x})$$

zur Bestimmung eines bezüglich der Verlust Funktion L optimalen Schätzers zu benutzen.

Der optimale Schätzer bezüglich der Verlust Funktion L ist so bestimmt, daß $\varphi(Z^*; \mathbf{x})$ minimal wird.

Beispiele:

•

$$L(e) = e^2$$

führt zu

$$z_L^*(\mathbf{x}) = z_E^*(\mathbf{x}) = \int z dF(\mathbf{x}; z|(n)) ,$$

also zum Erwartungswert der Verteilung $F(\mathbf{x}; z|(n))$.

•

$$L(e) = |e|$$

führt zu

$$z_L^*(\mathbf{x}) = F_n^{-1}(\mathbf{x}; 0.5|(n)) ,$$

also zum Median der Verteilung $F(\mathbf{x}; z|(n))$.

•

$$L(e) = \begin{cases} w_1 e & \text{falls } e > 0 \\ w_2 |e| & \text{falls } e \leq 0 \end{cases}$$

führt zu

$$\begin{aligned} z_L^*(\mathbf{x}) &= F^{-1}(\mathbf{x}; p|(n)) \\ p &= \frac{w_2}{w_1 + w_2} \end{aligned}$$

•

$$L(e) = \begin{cases} 0 & \text{falls } e = 0 \\ \text{const} & \text{falls } e \neq 0 \end{cases}$$

$z_L^*(\mathbf{x})$ ist dann der Modalwert der Verteilung.

Chapter 10

Odds and Ends Statt einer Zusammenfassung: Unvergeßliche Hinweise

- Daten–Modell
Datum = Drift–Term + Fluktuations–Term
welche Bedeutung (Interpretation) schließt die Lesart Datum = Trend–Term + Residual–Term in sich ein?
- Stützung (support)
punktformig, Probenkörper, ...
- geometrische Konfiguration der Meßorte (“Stützpunkte”)
- Stationarität
strenge Stationarität: Invarianz aller endlich–dimensionalen Randverteilungen (“alle Zufallsvariablen $Z(\mathbf{x})$, $\mathbf{x} \in D$ sind identisch verteilt”).
schwache Stationarität: Die Erwartungswertfunktion ist konstant, die Kovarianzfunktion ist ortsunabhängig und hängt nur von der Distanz ab (“alle Zufallsvariablen $Z(\mathbf{x})$, $\mathbf{x} \in D$ haben die gleiche Varianz”).
intrinsische Stationarität: schache Stationarität der Zuwächse, die Varianzfunktion der Zuwächse ist ortsunabhängig.
- Bedeutung der Größenordnung (“Maßstab”)
eine Drift–Komponente bezüglich einer Größenordnung kann zu einer Fluktuations–Komponente bezüglich einer größeren Größenordnung gehören.
klein(st)e Fluktuationen (“noise”) können durch eine Mikrostruktur (“nugget–effect”) bezüglich kleinerer Größenordnung verursacht sein.
- Isotropie
Die Kovarianzfunktion ist unabhängig von der Richtung

- Anisotropie

geometrische: gleiche Form und gleicher Schwellenwert, richtungsabhängige Reichweiten
 zonal: gleiche Form, richtungsabhängiger Schwellenwert und Reichweite

- kriging

Verkleinerung des kriging–Gleichungssystems – Schätzumgebung

Verminderung der Anzahl der kriging–Gleichungssysteme

Schnelle Berechnung von $\bar{\gamma}, \bar{C}$

Geeignete Daten–File Organisation

Stabile und schnelle Algorithmen zur Lösung linearer Gleichungssysteme

Wahl der Anzahl der Schritte (total number of lags): $N(|\mathbf{h}|) > 30$,

Wahl der Schrittweite (lag increment) $|\mathbf{h}| < d_{\max}/2$,

Wahl des Toleranzmaßes (lag tolerance)

Falls das Verhältnis von Nugget–Effekt zu Schwellenwert $C_0/C(0) > 1/2$, dann ist Mißtrauen gegen den Modellierungsversuch angebracht; bei $C_0/C(0) \leq 1/3$ kann man sich auf der sicheren Seite wähnen.

Optimale Schätzung mit KT = optimale Schätzung der Driftkomponente + optimale OK–Schätzung der optimalen Fluktuationskomponente

a–priori Varianz $\text{Var}(Z) < \infty$

Verteilungs– (Dispersions–) Varianz

$\lim_{D \rightarrow \infty}$ Verteilungs–Varianz = a–priori Varianz

Chapter 11

Last words, final truths

“... beware that uncertainty is not intrinsic to the phenomenon under study: rather it arises from our imperfect knowledge of that phenomenon, it is data-dependent and most importantly model-dependent, that model specifying our prior concept (decision) about the phenomenon. No model, hence no uncertainty measure, can ever be objective: the point is to accept that limitation and document clearly all aspects of the model”, Goovaerts, p. 442.

Chapter 12

Übungsaufgaben und S-PLUS script files

12.1 The JURA data set provided by J.-P. Dubois, IATE-Pédologie, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

Der Jura Data Set wurde in folgenden Original-Veröffentlichungen behandelt:

- Atteia, O., Dubois, J.-P., Webster, R., 1994, Geostatistical analysis of soil contamination in the Swiss Jura: Environmental Pollution 86, 315-327
- Webster, R., Atteia, O., Dubois, J.-P., 1994, Coregionalization of trace metals in the soil in the Swiss Jura: European Journal of Soil Science 45, 205-218

Beschreibung der Variablen

Spatial coordinates and values of categorial and continuous attributes at the 359 sampled sites:

- Koordinaten X [km], Y [km]
- Gesteinstyp: 1: Argovian, 2: Kimmeridgian, 3: Sequanian, 4: Portlandian, 5: Quaternary
- Landnutzung: 1: Forest, 2: Pasture (Weide(land), Wiese, Grasland), 3: Meadow (Wiese, Flur, Matte, Anger), 4: Tillage (Ackerland, bestelltes Land)
- Cadmium Cd[ppm] (0.8)
- Kupfer Cu[ppm] (50.0)
- Blei Pb[ppm] (50.0)

- Kobalt Co[ppm] (25.0)
- Chrom Cr[ppm] (75.0)
- Nickel Ni[ppm] (50.0)
- Zink Zn[ppm] (200.0)

Die Schwermetallgehalte sind analytische Meßwerte und werden als stetige Zufallsvariable behandelt; der Wert in Klammern ist der Grenzwert. Die Angaben zur Geologie (Gesteinstyp) und Flächennutzung sind vergleichsweise eher interpretierte Informationen ("soft data") und werden als kategoriale Zufallsvariablen behandelt.

Für 100 der insgesamt 359 Meßorte werden für die statistische Analyse und Schätzung (Vorhersage) nur die Angaben über Gesteinstyp, Landnutzung, und Gehalte von Nickel und Zink benutzt. Die gemessenen Werte für Cadmium, Kupfer, Blei, Kobalt, Chrom werden nur zum Vergleich der geschätzten und gemessenen Werte benutzt, um so zu empirischen Aussagen über die Güte der Schätzung zu kommen.

Explorative Datenanalyse

- (univariate) Häufigkeitstabelle der kategorialen Zufallsvariablen Gesteinstyp und Flächennutzung
- Kontingenztafel (bivariate Häufigkeitstabelle) für die beiden kategorialen Zufallsvariablen Gesteinstyp und Flächennutzung
- Histogramme für alle Schwermetallkonzentrationen
- kumulative Verteilungsfunktion aller Schwermetallkonzentrationen
- Einfache Statistiken (summary statistics)
- mittlere Schwermetallkonzentration für jeden Gesteinstyp, jede Flächennutzung
- Anteil der Daten der Schwermetallkonzentrationen von Cd, Cu, Pb über dem Grenzwert für jeden Gesteinstyp, jede Flächennutzung
- Streudiagramme (scatterplots) (X,Y): (Ni, Cd), (Zn, Cd), (Ni, Cu), (Zn, Cu), (Ni, Pb), (Zn; Pb)
- Matrix der linearen Korrelationskoeffizienten mit Signifikanzniveaus und Matrix der Rang- Korrelationskoeffizienten mit Signifikanzniveaus

Aufgaben der Statistik ortsabhängiger Zufallsvariablen

- Beschreibung der Form (pattern) der räumlichen stochastischen Abhängigkeit
- stochastische Modellierung der räumlichen Verteilung Schätzung von Werten für Orte ohne Meßwerte
- Abschätzung der Wahrscheinlichkeiten, daß Grenzwerte überschritten werden Angabe von Orten für zusätzliche Messungen, die einen möglichst großen Informationsgewinn vermitteln
- Modellierung der Aussage-Unsicherheit

12.2 Aufgaben der ersten Übung

(i) **Häufigkeitstafel**

Wie groß ist der Anteil der Fläche, der dauernd mit Gras bedeckt ist?

(ii) **empirische Verteilungsfunktion**

Wie groß ist der Anteil der Schwermetallkonzentrationen, die größer als der jeweilige Grenzwert (Cd 0.8, Cu 50.0, Pb 50.0, Co 25.0, Cr 75.0, Ni 50.0, Zn 200.0) sind? Welche Schwermetallgehalte sind also von besonderem Interesse?

(iii) **summary statistics**

Summary statistics der Schwermetallkonzentrationen - welche Verteilungen sind ausgesprochen schief? Sind die besonders interessanten Schwermetallgehalte schief verteilt, und falls ja, was bedeutet dies für die weitere Analyse?

(iv) **Kontingenztafel**

Wieviele der 20 Kategorien der (4x5) Flächennutzung - Gesteinstyp Kontingenztafel enthalten weniger als 10 Proben entsprechend 4%?

(v) **Lineare und Rangkorrelation**

Welche Paare von Schwermetallkonzentrationen sind durch die größten linearen Korrelationskoeffizienten bzw. durch die größten Rang-Korrelationskoeffizienten ausgezeichnet? Interpretation?

(vi) **Bedingte Verteilung**

Für welche Flächennutzungen sind die Schwermetallkonzentrationen (im Mittel) am größten? Interpretation?

(vii) **Bedingte Häufigkeiten**

Für welche Gesteinstypen sind die Cd- und Pb- Belastungen, d.h. die Anteile der Cd- und Pb- Konzentrationen über dem jeweiligen Grenzwert, am geringsten? (Die bedingten Häufigkeiten berechnet man z.B. für Cd nach der Formel

$$H(\text{ Cd } > 0.8 | \text{ Rock } = n) = \#\{ \text{ Cd } > 0.8 \text{ und } \text{ Rock } = n \} / \#\{ \text{ Rock } = n \}, n = 1, \dots, 5$$

12.2.1 S-PLUS script file für die Aufgaben der ersten Übung

```
jura<-read.table("C:/jura259.dat", header=T, skip=17) # Einlesen
attach(jura) # im Suchpfad nach oben stellen

table(Land) # absolute Häufigkeitstafel der Flächennutzung
100*table(Land[1:259])/length(Land[1:259]) # relative Häufigkeitstafel

sCd<-sort(Cd) # aufsteigend sortierte Cd Konzentrationen
```

```
length(Cd) # Anzahl der Cd Konzentrationen
length(sCd) # Anzahl der aufsteigend sortierten Cd Konzentrationen
edf<-(1:length(sCd))/length(sCd) # empirische Verteilungsfunktion der Cd Konzentrationen
plot(sCd,edf,type="S",xlab="Cd Concentration (ppm)",ylab="cumulative frequency")
abline(v=0.8) # vertikale Linie beim Grenzwert 0.8
Cd259<-Cd[1:259] # die ersten 259 Cd Konzentrationen; die nicht NA Cd Konzentrationen
length(Cd259[Cd259>0.8])/length(Cd259) # relative Häufigkeit h(Cd > 0.8)

sCu<-sort(Cu)
length(Cu)
length(sCu)
edf<-(1:length(sCu))/length(sCu)
plot(sCu,edf,type="S",xlab="Cu Concentration (ppm)",ylab="cumulative frequency")
abline(v=50.0)
Cu259<-Cu[1:259]
length(Cu259[Cu259>50.0])/length(Cu259)

sPb<-sort(Pb)
length(Pb)
length(sPb)
edf<-(1:length(sPb))/length(sPb)
plot(sPb,edf,type="S",xlab="Pb Concentration (ppm)",ylab="cumulative frequency")
abline(v=50.0)
Pb259<-Pb[1:259]
length(Pb259[Pb259>50.0])/length(Pb259)

sCo<-sort(Co)
length(Co)
length(sCo)
edf<-(1:length(sCo))/length(sCo)
plot(sCo,edf,type="S",xlab="Co Concentration (ppm)",ylab="cumulative frequency")
abline(v=25.0)
Co259<-Co[1:259]
length(Co259[Co259>25.0])/length(Co259)

sCr<-sort(Cr)
length(Cr)
length(sCr)
edf<-(1:length(sCr))/length(sCr)
plot(sCr,edf,type="S",xlab="Cr Concentration (ppm)",ylab="cumulative frequency")
abline(v=75.0)
Cr259<-Cr[1:259]
```

```
length(Cr259[Cr259>75.0])/length(Cr259)

sNi<-sort(Ni)
length(Ni)
length(sNi)
edf<-(1:length(sNi))/length(sNi)
plot(sNi,edf,type="S",xlab="Ni Concentration (ppm)",ylab="cumulative frequency")
abline(v=50.0)
length(Ni[Ni>50.0])/length(Ni)

sZn<-sort(Zn)
length(Zn)
length(sZn)
edf<-(1:length(sZn))/length(sZn)
plot(sZn,edf,type="S",xlab="Zn Concentration (ppm)",ylab="cumulative frequency")
abline(v=200.0)
length(Zn[Zn>200.0])/length(Zn)

summary(jura) # summary
par(mfrow=c(3,3)) # Aufteilung des Bildschirms
hist(Cd) # Histogramm f\"ur Cd
hist(Cu)
hist(Pb)
hist(Co)
hist(Cr)
hist(Ni)
hist(Zn)
par(mfrow=c(3,3))
hist(Cd,nclass=20) # Histogramm f\"ur Cd mit 20 Klassen
hist(Cu,nclass=20)
hist(Pb,nclass=20)
hist(Co,nclass=20)
hist(Cr,nclass=20)
hist(Ni,nclass=20)
hist(Zn,nclass=20)
par(mfrow=c(1,1))

100*table(Rock[1:259],Land[1:259])/length(Land[1:259]) # Kontingenztafel

jurappm<-jura[,5:11] # Datensatz der Konzentrationen in ppm
pairs(jurappm) # plot aller Paare
cor(jura[1:259,5:11],jura[1:259,5:11]) # Korrelationsmatrix
```

```

Ni259<-Ni[1:259]
Zn259<-Zn[1:259]
cor.test(Cu259,Pb259,method="pearson") # Linearer Korrelationstest
cor.test(Cu259,Pb259,method="spearman") # Rangkorrelationstest
cor.test(Co259,Ni259,method="pearson")
cor.test(Co259,Ni259,method="spearman")

cor.test(Cd259,Ni259,method="pearson")
cor.test(Cd259,Ni259,method="spearman")
cor.test(Cu259,Ni259,method="pearson")
cor.test(Cu259,Ni259,method="spearman")
cor.test(Pb259,Ni259,method="pearson")
cor.test(Pb259,Ni259,method="spearman")
cor.test(Cd259,Zn259,method="pearson")
cor.test(Cd259,Zn259,method="spearman")
cor.test(Cu259,Zn259,method="pearson")
cor.test(Cu259,Zn259,method="spearman")
cor.test(Pb259,Zn259,method="pearson")
cor.test(Pb259,Zn259,method="spearman")
cor.test(Ni259,Zn259,method="pearson")
cor.test(Ni259,Zn259,method="spearman")

#trellis.device(win.graph)
par(mfrow=c(3,2))
plot(Ni259,Cd259,xlab="Ni",ylab="Cd",main="jurappm") # scatter plot
plot(Zn259,Cd259,xlab="Zn",ylab="Cd",main="jurappm")
plot(Ni259,Cu259,xlab="Ni",ylab="Cu",main="jurappm")
plot(Zn259,Cu259,xlab="Zn",ylab="Cu",main="jurappm")
plot(Ni259,Pb259,xlab="Ni",ylab="Pb",main="jurappm")
plot(Zn259,Pb259,xlab="Zn",ylab="Pb",main="jurappm")

par(mfrow=c(1,1))
xyplot(Cd259~Ni259, data=jura, main="scatter plot Cd = f(Ni)")
xyplot(Cd259~Zn259, data=jura, main ="scatter plot Cd = f(Zn)")
xyplot(Cu259~Ni259, data=jura, main="scatter plot Cu = f(Ni)")
xyplot(Cu259~Zn259, data=jura, main ="scatter plot Cu = f(Zn)")
xyplot(Pb259~Ni259, data=jura, main="scatter plot Pb = f(Ni)")
xyplot(Pb259~Zn259, data=jura, main ="scatter plot Pb = f(Zn)")

tapply(Cd[1:259], Land[1:259], mean) # Mittelwerte von Cd bez\uglich jeder Fl\achennutz
tapply(Cu[1:259], Land[1:259], mean)
tapply(Pb[1:259], Land[1:259], mean)

```

```
tapply(Co[1:259], Land[1:259], mean)
tapply(Cr[1:259], Land[1:259], mean)
tapply(Ni[1:359], Land[1:359], mean)
tapply(Zn[1:359], Land[1:359], mean)

length(Cd259[Cd259>0.8 & Rock[1:259]==1])/length(Cd259[Rock[1:259]==1]) # bedingte relative Häufigkeit h(Cd > 0.8 | Rock = 1)
length(Cd259[Cd259>0.8 & Rock[1:259]==2])/length(Cd259[Rock[1:259]==2])
length(Cd259[Cd259>0.8 & Rock[1:259]==3])/length(Cd259[Rock[1:259]==3])
length(Cd259[Cd259>0.8 & Rock[1:259]==4])/length(Cd259[Rock[1:259]==4])
length(Cd259[Cd259>0.8 & Rock[1:259]==5])/length(Cd259[Rock[1:259]==5])
length(Cu259[Cu259>50.0 & Rock[1:259]==1])/length(Cu259[Rock[1:259]==1])
length(Cu259[Cu259>50.0 & Rock[1:259]==2])/length(Cu259[Rock[1:259]==2])
length(Cu259[Cu259>50.0 & Rock[1:259]==3])/length(Cu259[Rock[1:259]==3])
length(Cu259[Cu259>50.0 & Rock[1:259]==4])/length(Cu259[Rock[1:259]==4])
length(Cu259[Cu259>50.0 & Rock[1:259]==5])/length(Cu259[Rock[1:259]==5])
length(Pb259[Pb259>50.0 & Rock[1:259]==1])/length(Pb259[Rock[1:259]==1])
length(Pb259[Pb259>50.0 & Rock[1:259]==2])/length(Pb259[Rock[1:259]==2])
length(Pb259[Pb259>50.0 & Rock[1:259]==3])/length(Pb259[Rock[1:259]==3])
length(Pb259[Pb259>50.0 & Rock[1:259]==4])/length(Pb259[Rock[1:259]==4])
length(Pb259[Pb259>50.0 & Rock[1:259]==5])/length(Pb259[Rock[1:259]==5])

detach("jura") # in den Suchpfad einordnen
rm(jura, jurappm) # Aufruf "aumen, Objekte 1" "oschen"
rm(Cd259, Cu259, Pb259, Co259, Cr259, Ni259, Zn259)
rm(sCd, sCu, sPb, sCo, sCr, sNi, sZn, edf)
```

12.3 Aufgaben der zweiten Übung

(i) **Delaunay Triangulierung, Dirichlet–Thiessen–Voronoi Zerlegung**

Die Delaunay Triangulierung erhält man aus der Dirichlet - Thiessen - Voronoi Zerlegung, indem man jeden Punkt mit seinen starken Thiessen Nachbarn durch Dreiecksanten verbindet (siehe dazu auch das Vorlesungsskript).

- (a) Konstruieren Sie geometrisch am gegebenen Beispiel der Delaunay Triangulation das entsprechende Thiessen Polygon zum Punkt $P4$.
- (b) Geben Sie die allgemeine Konstruktionsvorschrift in einem Satz an.

(ii) **Lineare Interpolation über Dreiecken**

Konstruieren Sie geometrisch die Isolinie zum Wert $z = 15.0$ nach der Methode der linearen Interpolation über den gegebenen Dreiecken der bedingten Delaunay Triangulation der Punkte der digitalisierten Isolinien zu den Werten $z_1 = 10.0, z_2 = 20.0$.

(iii) **Karten und Indikator–Karten**

Konstruieren Sie am PC

- (a) Karten der flächenhaften Cd, Cu, Pb und Zn Verteilung. Welche Erkenntnis bezüglich der flächenhaften Verteilungen vermitteln diese Karten?
- (b) Indikator-Karten für Cd, Cu und Pb Konzentrationen über dem jeweiligen Grenzwert. Welche Erkenntnis bezüglich der flächenhaften Verteilungen vermitteln diese Karten?
- (c) Indikatkarten für Cd Konzentrationen über den Quantilen (Dezentilen) $z_{0.2}, z_{0.6}$ und $z_{0.8}$. Welche Erkenntnis bezüglich der flächenhaften Verteilungen vermitteln diese Karten?
- (d) eine Karte der Flächennutzung - Können Sie ein flächenhaftes "Muster" erkennen?
- (e) eine Karte der Gesteinstypen - Können Sie ein Muster erkennen?

Welche Erkenntnisse vermitteln die Karten bezüglich der flächenhaften Verteilungen?

(iv) **Experimentelle Semivariogramme**

Erstellen Sie experimentelle Semivariogramme für die Cd Schwermetallkonzentration

- (a) in Richtung (direction, azimuth) $\theta = 22.5, 67.5, 112.5, 157.5$ mit (angular tolerance, azimuth tolerance) $\Delta\theta = 22.5$, (lag spacing, lag tolerance) $\Delta h = 100\text{m}$
- (b) omnidirektional mit $\Delta h = 100\text{m}$
- (c) omnidirektional mit $\Delta h = 100\text{m}$ des entsprechenden logarithmisch transformierten Schwermetallgehalt.

Machen Sie für jedes Semivariogram Aussagen zu seinen Parametern Nuggeteffekt, Reichweite und Schwellenwert.

(v) **Experimentelles Indikator Semivariogramm zu einem threshold**

Erstellen Sie ein omnidirektionales experimentelles Indikator Semivariogramme für den Indikator bezüglich der Quantile (Dezentile) $z_{0.2}$, $z_{0.6}$ und $z_{0.8}$ der Cd-Schwermetallkonzentrationen. Tip: lag spacing $\Delta h = 100\text{m}$. Interpretieren Sie die Semivariogramme mit Hilfe von Aussagen zu ihren Parametern Nuggeteffekt, Reichweite und Schwellenwert.

12.3.1 S-PLUS script file für die Aufgaben der zweiten Übung

```
jura<-read.table("C:/jura259.dat",header=T,skip=17)
jura
summary(jura[,5:11])
attach(jura)
module(spatial) # Modul für räumliche Statistik laden

scaled.plot(X,Y,main="alle 359 Messorte") # maßstabstreuer plot der Messorte
jura.hull<-chull(X,Y) # konvexe Hülle aller Messorte
polygon(X[jura.hull],Y[jura.hull],dens=0) # Polygon der konvexen Hülle
scaled.plot(X,Y,type="n",main="259 berücksichtigte Messorte") # Messorte
points(X[Cd!="NA"],Y[Cd!="NA"]) # Messorte
polygon(X[jura.hull],Y[jura.hull],dens=0)
scaled.plot(X,Y,type="n",main="100 nicht berücksichtigte Messorte") # Messorte, für die Cd zu NA gesetzt wurde
points(X[Cd=="NA"],Y[Cd=="NA"]) # Messorte, für die Cd zu NA gesetzt wurde
polygon(X[jura.hull],Y[jura.hull],dens=0)

scaled.plot(X,Y,type="n",main="klassifizierte Cd Konzentrationen")
Cd259<-Cd[1:259]
Cd259.groups<-cut(Cd259,breaks=c(0,0.8,1.5,3.0,4.5,6.0)) # Klassifizierung in 5 Klassen
Cd259.groups
Cd259.groups<-as.numeric(Cd259.groups)
text(X,Y,Cd259.groups)
polygon(X[jura.hull],Y[jura.hull],dens=0)

scaled.plot(X,Y,type="n",main="klassifizierte Cu Konzentrationen")
Cu259<-Cu[1:259]
Cu259.groups<-cut(Cu259,breaks=c(0,50,90,130,170))
Cu259.groups
Cu259.groups<-as.numeric(Cu259.groups)
text(X,Y,Cu259.groups)
polygon(X[jura.hull],Y[jura.hull],dens=0)
```

```

scaled.plot(X,Y,type="n",main="klassifizierte Pb Konzentrationen")
Pb259<-Pb[1:259]
Pb259.groups<-cut(Pb259,breaks=c(0,50,100,150,200,250))
Pb259.groups
Pb259.groups<-as.numeric(Pb259.groups)
text(X,Y,Pb259.groups)
polygon(X[jura.hull],Y[jura.hull],dens=0)

scaled.plot(X,Y,type="n",main="klassifizierte Zn Konzentrationen")
Zn259<-Zn[1:259]
Zn259.groups<-cut(Zn259,breaks=c(0,200,215,230,245,260))
Zn259.groups
Zn259.groups<-as.numeric(Zn259.groups)
text(X,Y,Zn259.groups)
polygon(X[jura.hull],Y[jura.hull],dens=0)

X259<-X[1:259]
Y259<-Y[1:259]

scaled.plot(X,Y,type="n",main="Cd Konzentration kleiner/groesser als Grenzwert")
points(X259[Cd<=0.8],Y259[Cd<=0.8],pch="s") # alphanumerischer plot
points(X259[Cd>0.8],Y259[Cd>0.8],pch="L")
polygon(X[jura.hull],Y[jura.hull],dens=0)

ICd259<-as.numeric(Cd259<=0.8) # I Indikator zum <Grenzwert, ICd259 = 1 falls Cd <=0.8
JCd259<-1-ICd259 # J Indiaktor = 1 - I
scaled.plot(X,Y,type="n",main="J Indikator zum Cd Grenzwert") # plot
text(X,Y,JCd259) # bin\"arer plot
polygon(X[jura.hull],Y[jura.hull],dens=0)

ICu259<-as.numeric(Cu259<=50.0)
JCu259<-1-ICu259
scaled.plot(X,Y,type="n",main="J Indiaktor zum Cu Grenzwert")
text(X,Y,JCu259)
polygon(X[jura.hull],Y[jura.hull],dens=0)

IPb259<-as.numeric(Pb259<=50.0)
JPb259<-1-IPb259
scaled.plot(X,Y,type="n",main="J Indikator zum Pb Grenzwert")
text(X,Y,JPb259)
polygon(X[jura.hull],Y[jura.hull],dens=0)

```

```

Cd259q<-quantile(Cd,c(0.2,0.35,0.6,0.8),na.rm=TRUE) # Quantile
Cd259q
Cd259q<-quantile(Cd259,c(0.2,0.35,0.6,0.8)) # Quantile
Cd259q
ICd259q80<-as.numeric(Cd<=quantile(Cd,0.8,na.rm=TRUE)) # I Indikator zum 0.8 Quantil
ICd259q80
scaled.plot(X,Y,type="n",main="I Indikator zum 0.8 Quantil Cd")
# I Indikator Karte
text(X,Y,ICd259q80)
polygon(X[jura.hull],Y[jura.hull],dens=0)

scaled.plot(X,Y,type="n",main="I Indikator zum 0.8 Quantil Cd")
points(X[ICd259q80==0],Y[ICd259q80==0],pch=3,cex=0.7)
points(X[ICd259q80==1],Y[ICd259q80==1],pch=1,cex=0.7)
polygon(X[jura.hull],Y[jura.hull],dens=0)

JCd259q80<-1-ICd259q80 # J indicator of threshold z(0.8)
JCd259q80
scaled.plot(X,Y,type="n",main="J Indikator zum 0.8 Quantil Cd")
# J indicator map
text(X,Y,JCd259q80)
polygon(X[jura.hull],Y[jura.hull],dens=0)

ICd259q60<-as.numeric(Cd<=quantile(Cd,0.6,na.rm=TRUE)) # I Indikator zum 0.6 Quantil
JCd259q60<-1-ICd259q60 # J indicator of threshold z0.6
JCd259q60
scaled.plot(X,Y,type="n",main="J Indikator zum 0.6 Quantil Cd")
# J indicator map
text(X,Y,JCd259q60)
polygon(X[jura.hull],Y[jura.hull],dens=0)

ICd259q35<-as.numeric(Cd<=quantile(Cd,0.35,na.rm=TRUE)) # I Indikator zum 0.35 Quantil
JCd259q35<-1-ICd259q35 # J indicator of threshold z0.35
JCd259q35
scaled.plot(X,Y,type="n",main="J Indikator zum 0.35 Quantil Cd") # J indicator map
text(X,Y,JCd259q35)
polygon(X[jura.hull],Y[jura.hull],dens=0)

ICd259q20<-as.numeric(Cd<=quantile(Cd,0.2,na.rm=TRUE)) # I Indikator zum 0.2 Quantil
JCd259q20<-1-ICd259q20 # J indicator of threshold z0.2
JCd259q20
scaled.plot(X,Y,type="n",main="J Indikator zum 0.2 Quantil Cd") # J indicator map

```

```

text(X,Y,JCd259q20)
polygon(X[jura.hull],Y[jura.hull],dens=0)

scaled.plot(X,Y,type="n",main="Flaechennutzung")
text(X,Y,Land) # alphanumerischer plot der Flächennutzung
polygon(X[jura.hull],Y[jura.hull],dens=0)

Land
Land==1
as.numeric(Land==1)
scaled.plot(X,Y,type="n",main="Karte der Flaechennutzung 1")
text(X,Y,as.numeric(Land==1)) # indicator map der Flächennutzung 1
polygon(X[jura.hull],Y[jura.hull],dens=0)
scaled.plot(X,Y,type="n",main="Karte der Flaechennutzung 2")
text(X,Y,as.numeric(Land==2))
polygon(X[jura.hull],Y[jura.hull],dens=0)
scaled.plot(X,Y,type="n",main="Karte der Flaechennutzung 3")
text(X,Y,as.numeric(Land==3))
polygon(X[jura.hull],Y[jura.hull],dens=0)
scaled.plot(X,Y,type="n",main="Karte der Flaechennutzung 4")
text(X,Y,as.numeric(Land==4))
polygon(X[jura.hull],Y[jura.hull],dens=0)

scaled.plot(X,Y,type="n",main="Karte der Flaechennutzung")
points(X[Land==1],Y[Land==1],pch=1,cex=0.7)
points(X[Land==2],Y[Land==2],pch=2,cex=0.7)
points(X[Land==3],Y[Land==3],pch=3,cex=0.7)
points(X[Land==4],Y[Land==4],pch=4,cex=0.7)
polygon(X[jura.hull],Y[jura.hull],dens=0)
legend(-1,5,legend=c("Wald","Weide","Wiese","Acker"),marks=c(1,2,3,4))

scaled.plot(X,Y,type="n",main="Gesteinstyp") # alphanumerischer plot des Gesteinstyp
text(X,Y,Rock)
polygon(X[jura.hull],Y[jura.hull],dens=0)

az<-c(22.5,67.5,112.5,157.5) # 4 Richtungen
Cd.var2<-variogram(Cd~loc(X,Y),data=jura,
azimuth=az,tol.azimuth=22.5,nlag=10,lag=0.2,
tol.lag=0.1,maxdist=2.0,na.action="na.omit") # 4 richtungsabhängige Semivariogramme
plot(Cd.var2,type="l",xlab="directional semivariograms")
Cd.var2

```

```

Cd.var1<-variogram(Cd~loc(X,Y),data=jura,azimuth=90.0,
tol.azimuth=22.5,nlag=10,lag=0.2,
tol.lag=0.1,maxdist=2.0,na.action="na.omit") # Semivariogramm zur Richtung 90
plot(Cd.var1,type="l",xlab="unidirectional semivariogram of Cd")
Cd.var1

Cd.var<-variogram(Cd~loc(X,Y),data=jura,
azimuth=90,tol.azimuth=90, nlag=10, lag=0.2,
tol.lag=0.1, maxdist=2.0, na.action="na.omit") # richtungsunabh\"angiges Semivariogramm
plot(Cd.var,type="b",xlab="omnidirectional semivariogram of Cd")
Cd.var

Cd
logCd<-log(Cd)
logCd # logarithmierte Cd Konzentrationen
logCd.var<-variogram(logCd~loc(X,Y),data=jura,
azimuth=90,tol.azimuth=90, nlag=10, lag=0.2,
tol.lag=0.1, maxdist=2.0, na.action="na.omit") # und ihr richtungsunabh\"angiges Semivariogramm
plot(logCd.var,type="b",xlab="omnidirectional semivariogram of logCd")
logCd.var

ICd259q20.var2<-variogram(ICd259q20~loc(X,Y),data=jura,
azimuth=90,tol.azimuth=90,nlag=10,lag=0.2,
tol.lag=0.1,maxdist=2.0,na.action="na.omit") # I Indikator-Semivariogramm zum 20\% Cd-Quotienten
plot(ICd259q20.var2,type="l",xlab="omnidirectional I indicator semivariogram of q20 of Cd")
ICd259q20.var2

JCd259q20.var2<-variogram(JCd259q20~loc(X,Y),data=jura,
azimuth=90,tol.azimuth=90,nlag=10,lag=0.2,
tol.lag=0.1,maxdist=2.0,na.action="na.omit") # J Indikator-Semivariogramm zum 20\% Cd-Quotienten
plot(JCd259q20.var2,type="l",xlab="omnidirectional J indicator semivariogram of q20 of Cd")
JCd259q20.var2

ICd259q35.var2<-variogram(ICd259q35~loc(X,Y),data=jura,
azimuth=90,tol.azimuth=90,nlag=10,lag=0.2,
tol.lag=0.1,maxdist=2.0,na.action="na.omit")
plot(ICd259q35.var2,type="l",xlab="omnidirectional I indicator semivariogram of q35 of Cd")
ICd259q35.var2

ICd259q60.var2<-variogram(ICd259q60~loc(X,Y),data=jura,
azimuth=90,tol.azimuth=90,nlag=10,lag=0.2,
tol.lag=0.1,maxdist=2.0,na.action="na.omit")

```

```
plot(ICd259q60.var2,type="l",xlab="omnidirectional I indicator semivariogram of q60 of 0
ICd259q60.var2

ICd259q80.var2<-variogram(ICd259q80~loc(X,Y),data=jura,
azimuth=90,tol.azimuth=90,nlag=10,lag=0.2,
tol.lag=0.1,maxdist=2.0,na.action="na.omit")
plot(ICd259q80.var2,type="l",xlab="omnidirectional I indicator semivariogram of q80 of 0
ICd259q80.var2

detach("jura")
rm(jura,jura.hull)
rm(X259,Y259,Cd259,Cu259,Pb259,Zn259)
rm(Cd259.groups,Cu259.groups,Pb259.groups,Zn259.groups)
rm(ICd259,JCd259,ICu259,JCu259,IPb259,JPb259)
rm(Cd259q,ICd259q20,ICd259q35,ICd259q60,ICd259q80,JCd259q20,JCd259q35,JCd259q60,JCd259q80
rm(az,Cd.var,Cd.var1,Cd.var2)
rm(logCd,logCd.var)
rm(ICd259q20.var2,JCd259q20.var2,ICd259q35.var2,ICd259q60.var2,ICd259q80.var2)
```

12.4 Aufgaben der dritten Übung

(i) **Experimentelle Indikator–Semivariogramme einer kategorialen Variablen**

Erstellen Sie omnidirektionale experimentelle Indikator Semivariogramme für die häufigsten Flächennutzungen Land=1 (forest, Wald), Land=2 (pasture, Weide), Land=3 (meadow, Wiese). Tip: lag spacing $\Delta h = 100\text{m}$. Interpretieren Sie die Semivariogramme mit Hilfe von Aussagen zu ihren Parametern Nuggeteffekt, Reichweite und Schwellenwert.

(ii) **Experimentelle Indikator–Semivariogramme einer kategorialen Variablen**

Erstellen Sie experimentelle Indikator Semivariogramme für den Gesteinstyp Rock=1 (Argovian) für die Richtungen (direction, azimuth) theta =22.5, 67.5, 112.5, 157.5 mit (angular tolerance, azimuth tolerance) delta theta = 22.5, und (lag spacing, lag tolerance) delta h =100m Interpretieren Sie die Semivariogramme mit Hilfe von Aussagen zu ihren Parametern Nuggeteffekt, Reichweite und Schwellenwert.

(iii) **Variogramm Wolke**

Die Variogrammwolke stellt die Verteilung der Varianz der Meßwerte bezüglich aller Punktpaare für jede Schrittweite dar.

- (a) Was lässt sich über die Gestalt dieser Verteilung für die Cd Konzentrationen aussagen? Welche Konsequenzen hat dies?
- (b) Sind die großen Semivariogrammwerte auf die tatsächliche Schiefe der Verteilung zurückzuführen oder als untypische auffällige Beobachtungen zu qualifizieren?

(iv) **Schwellenwert und Stichprobenvarianz (ohne PC)**

- (a) Warum ist die empirische Varianz (Stichprobenvarianz) aller Meßwerte unter der Annahme entsprechender Stationaritätsbedingungen im allgemeinen kein geeigneter Schätzer für den Schwellenwert des zugehörigen empirischen Semivariogramms?
- (b) Stellen Sie den Wert des theoretischen Semivariogramms für Schrittweiten größer als die Reichweite, also für unkorrierte regionalisierte Zufallsvariablen mit Hilfe von Erwartungswert, Varianz und Kovarianz aller Zufallsvariablen dar.
- (c) Es lässt sich zeigen, da der Erwartungswert der theoretischen Varianz der Zufallsvariablen unter der Annahme, daß der theoretische Mittelwert der Zufallsvariablen bekannt ist, gleich dem arithmetischen Mittel aller n^2 Werte des theoretischen Semivariogramms der n Zufallsvariablen ist. Diskutieren Sie unter Berücksichtigung dieses Zusammenhangs, wann die empirische Varianz ein geeigneter bzw. kein geeigneter Schätzer für den Schwellenwert des empirischen Semivariogramms ist.

(v) **Experimentelle Indikator–Semivariogramme**

Erstellen Sie ein omnidirektionales experimentelles Indikator Semivariogramme für den Indikator bezüglich des Medians der Cd-Schwermetallkonzentrationen. Tip: lag spacing $\Delta h = 100\text{m}$. Interpretieren Sie die Semivariogramme mit Hilfe von Aussagen zu ihren Parametern Nuggeteffekt, Reichweite und Schwellenwert.

(vi) **Semivariogramm–Modellierung, Variographie**

Modellieren Sie das experimentelle omnidirektionale Semivariogramm des Cr-Schwermetallgehalts mit Hilfe isotroper Modell-Semivariogramme. Wieviele Strukturen benötigen und verwenden Sie? Welches sind die Parameter dieser Strukturen?

(vii) **Semivariogramm–Modellierung, Variographie**

Modellieren Sie das experimentelle omnidirektionale Semivariogramm eines der Schwermetallgehalte Cd, Cu, Pb mit Hilfe isotroper Modell-Semivariogramme. Wieviele Strukturen benötigen und verwenden Sie? Welches sind die Parameter dieser Strukturen?

12.4.1 S-PLUS script file für die Aufgaben der dritten Übung

```
jura<-read.table("C:/programme/splus4/jura/jura259.dat",header=T,skip=17)
jura
attach(jura)
module(spatial)

ILand1<-as.numeric(Land==1) # I Indikator f\"ur Fl\"achennutzung 1
ILand1
ILand1.var<-variogram(ILand1~loc(X,Y),data=jura,
azimuth=90,tol.azimuth=90,nlag=10,lag=0.2,
tol.lag=0.1,maxdist=2.0,na.action="na.omit")
plot(ILand1.var,type="l",xlab="omnidirectional variograms of forest")
ILand1.var

ILand2<-as.numeric(Land==2)
ILand2
ILand2.var<-variogram(ILand2~loc(X,Y),data=jura,
azimuth=90,tol.azimuth=90,nlag=10,lag=0.2,
tol.lag=0.1,maxdist=2.0,na.action="na.omit")
plot(ILand2.var,type="l",xlab="omnidirectional variograms of pasture")
ILand2.var

ILand3<-as.numeric(Land==3)
ILand3
ILand3.var<-variogram(ILand3~loc(X,Y),data=jura,
azimuth=90,tol.azimuth=90,nlag=10,lag=0.2,
```

```

tol.lag=0.1,maxdist=2.0,na.action="na.omit")
plot(ILand3.var,type="l",xlab="omnidirectional variograms of meadow")
ILand3.var

IRock1<-as.numeric(Rock==1)
IRock1
az<-c(22.5,67.5,112.5,157.5)
IRock1.var2<-variogram(IRock1~loc(X,Y),data=jura,
azimuth=az,tol.azimuth=22.5,nlag=10,lag=0.2,
tol.lag=0.1,maxdist=2.0,na.action="na.omit")
plot(IRock1.var2,type="l",xlab="directional variograms of Argovian")
IRock1.var2

Cd.vcloud<-variogram.cloud(Cd~loc(X,Y), data=jura, na.action=na.omit)
plot(Cd.vcloud,main="variogram cloud")
Cd.vcloud1<-variogram.cloud(Cd~loc(X,Y), data=jura, na.action=na.omit, maxdist=2.0)
plot(Cd.vcloud1,main="variogram cloud, maxdist =2.0")
Cd.vcloud2<-update(Cd.vcloud1, maxdist=1.3)
plot(Cd.vcloud2,main="variogram cloud, maxdist=1.3")
Cd.vcloud3<-update(Cd.vcloud1, maxdist=0.7)
plot(Cd.vcloud3,main="variogram cloud, maxdist=0.7")
Cd.vcloud4<-update(Cd.vcloud1, maxdist=0.3)
plot(Cd.vcloud4,main="variogram cloud, maxdist=0.3")

Cd.vcloud2<-variogram.cloud(Cd~loc(X,Y), data=jura, na.action=na.omit, maxdist=1.3)
boxplot(Cd.vcloud2,main="variogram cloud, boxplot of squared increments, maxdist=1.3")
Cd.vcloud2<-variogram.cloud(Cd~loc(X,Y), data=jura, na.action=na.omit, maxdist=1.3)
boxplot(Cd.vcloud2, mean=T, pch.mean="o",main="variogram cloud, boxplot of squared increments, m

Cd.vcloudx<-variogram.cloud(Cd~loc(X,Y), data=jura, na.action=na.omit, maxdist=1.3,
fun=function(zi,zj) sqrt(abs(zi-zj))/2)
boxplot(Cd.vcloudx, mean=T, pch.mean="o",main="variogram cloud, boxplot of increments, m

detach("jura")
rm(jura)
rm(ILand1,ILand2,ILand3)
rm(ILand1.var,ILand2.var,ILand3.var)
rm(IRock1,IRock1.var2)
rm(Cd.vcloud,Cd.vcloud1,Cd.vcloud2,Cd.vcloud3,Cd.vcloud4)
rm(Cd.vcloudx)

```

12.5 Aufgaben der vierten Übung

(i) Variographie, Modellieren eines Semivariogramms

Modellieren Sie das experimentelle omnidirektionale Semivariogramm des Cd Schwermetallgehalts mit Hilfe eines isotropen Modell-Semivariogramms.

- (a) Wieviele Strukturen einer geschachtelten Struktur benötigen und verwenden Sie?
- (b) Welches sind die Parameter dieser Strukturen?

(ii) kriging

Die dem Jura Datensatz zugrundeliegende geometrische Meßpunkt-Konfiguration resultiert aus der Überlagerung eines regelmäßigen quadratischen Rasters mit 107 Meßpunkten im Abstand von 250m und 38 Clustern von je 4 um einen Rasterpunkt geschachtelten Meßpunkten. Berechnen Sie für die 100 Testorte, an denen Cd als NA vereinbart wurde, den Krige-Schätzwert $z^*(x_\alpha)$ für Cd und die entsprechende Krige-Fehlervarianz $s_K^2(x)$, d.h. den minimalen Wert der Schätzvarianz unter Verwendung des angepaßten Semivariogramms

$$\gamma(h) = 0.30g_0(h) + 0.30\text{sph}(|h|/200) + 0.26\text{sph}(|h|/1300)$$

wobei $g_0(h)$ das Nugget-Effekt Model und sph das sphärische Semivariogramm Model bezeichnet.

- (a) Verschaffen Sie sich einen Überblick über die geometrische Anordnung der 259 Meßpunkte und der 100 Testpunkte. Erkennen Sie ein Muster? Beschreiben Sie es!
- (b) Vergleichen Sie in einem scatter plot die geschätzten mit den wahren Cd Konzentrationen und berechnen Sie den entsprechenden Rangkorrelationskoeffizient. Interpretieren Sie den Vergleich insbesondere hinsichtlich der Schätzung kleiner und großer wahrer Cd Konzentrationen und hinsichtlich der Forderung der Unverzerrtheit des kriging Schätzers.
- (c) Vergleichen Sie in einem scatter plot den wahren absoluten Fehler $|z^*(x_\alpha) - z(x_\alpha)|$ mit der Krige-Fehler-Standardabweichung $\sigma_k(x)$ und berechnen Sie den entsprechenden Rangkorrelationskoeffizienten. Interpretieren Sie beide Vergleiche!
- (d) Vergleichen Sie das omnidirektionale experimentelle Semivariogramm der Cd Konzentrationen der 259 Meßpunkte und das omnidirektionale experimentelle Semivariogramm der Cd Konzentrationen aller 359 Meß- und Testpunkte bezüglich der Reichweite, des relativen Nugget-Effekts und des Schwellenwerts. Interpretation!
- (e) Vergleichen Sie die Anzahl der Paare $(Z(x_\alpha), Z(x_\alpha + h))$, aus denen der Wert des empirischen Semivariogramms für h zwischen 0m und 500m berechnet wird!

- (f) Vergleichen Sie beide omnidirektionale empirischen Semivariogramme mit Ihrem angepaßten Modell-Semivariogramm und mit dem angepassten geschachtelten Modell-Semivariogramm

$$\gamma(h) = 0.30g_0(h) + 0.30\text{sph}(|h|/200) + 0.26\text{sph}(|h|/1300)$$

(iii) **Einfaches kriging und lineare Regression (ohne PC)**

Entwickeln Sie "einfaches kriging" für den Fall, daß die den Meßwerten z_α entsprechenden Zufallsvariablen $Z_\alpha = Z(\mathbf{x}_\alpha)$, $\alpha = 1, \dots, n$, als die stochastisch unabhängigen Zufallsvariablen und der Schätzer $Z_0^* = Z^*(\mathbf{x}_0)$ als die abhängige Zufallsvariable angenommen wird, d.h. unter der Annahme der klassischen Statistik.

- (a) Welche spezielle Form nimmt die Kovarianzmatrix des SK Gleichungssystems an, welche Werte nehmen die von Null verschiedene Matrixelemente (Kovarianzen) an?
- (b) Bestimmen Sie die Lösung dieses SK Gleichungssystems, d.h. bestimmen Sie die SK Gewichte.
- (c) Welche Form hat also der Schätzer für diesen speziellen Fall?
- (d) Welche Form hat der Schätzer Z_0^* unter der zusätzlichen Annahme, daß er von den Z_α , $\alpha = 1, \dots, n$, stochastisch unabhängig ist?
- (e) Welche Form nimmt das Semivariogramm an (Schwellenwert!), wie sieht die Kovarianzfunktion aus? (Unter welcher zusätzlichen Annahme über die der Aufgabenstellung hinaus sind die Begriffe Semivariogramm und Kovarianzfunktion praktisch sinnvoll? Wie ändern sich die Antworten auf a)-d) unter Berücksichtigung dieser zusätzlichen Modellannahme und der speziellen Form des Semivariogramms und der Kovarianzfunktion?)

(iv) **SK – schwache Stationarität, OK – intrinsische Stationarität (ohne PC)**

Warum kann man das SK Gleichungssystem ausschließlich mit Hilfe der Kovarianzfunktion ableiten und darstellen, während man das OK Gleichungssystem sowohl mit Hilfe der Kovarianzfunktion als auch des Semivariogramms darstellen kann? (Hinweis: Die Verwendung der Kovarianzfunktion bezieht sich auf die Modellannahme der schwachen Stationarität [Stationarität der Momente], die Verwendung des Semivariogramms bezieht sich auf die Modellannahme der intrinsischen Stationarität [schwache Stationarität der Inkremente]. In welchem Fall ist die Existenz der (zu minimierenden) Schätzvarianz unbedingt gesichert, in welchem Fall ist sie nur bedingt gesichert? Wie lautet diese Bedingung?)

12.5.1 S-PLUS script file für die Aufgaben der vierten Übung

```
jura<-read.table("C:/jura259.dat", header=T, skip=17)
attach(jura)
```

```

module(spatial)

jura.hull<-chull(X,Y) # konvexe H\\"ulle aller Me{\ss}orte
scaled.plot(X,Y,main="alle 359 Messorte") # ma{\ss}stabstreuer plot der Me{\ss}orte
polygon(X[jura.hull],Y[jura.hull],dens=0) # Polygon der konvexen H\\"ulle
scaled.plot(X,Y,type="n",main="259 beruecksichtigte Messorte")
points(X[Cd!="NA"],Y[Cd!="NA"],col=4) # Me{\ss}orte
polygon(X[jura.hull],Y[jura.hull],dens=0)
scaled.plot(X,Y,type="n",main="100 nicht beruecksichtigte Messorte")
points(X[Cd=="NA"],Y[Cd=="NA"],col=8) # Me{\ss}orte, f\"ur die Cd zu NA gesetzt wurde
polygon(X[jura.hull],Y[jura.hull],dens=0)
scaled.plot(X,Y,type="n",main="alle 359 Messorte")
points(X[Cd!="NA"],Y[Cd!="NA"],pch=18,col=4) # Me{\ss}orte
points(X[Cd=="NA"],Y[Cd=="NA"],pch=5,col=8) # Me{\ss}orte, f\"ur die Cd zu NA gesetzt wurde
polygon(X[jura.hull],Y[jura.hull],dens=0)

vdist<-1:100
vrangle<-30
plot(vdist,exp.vgram(distance=vdist,range=vrangle),type="l",xlab=" practical range = 3 %",
vrangle,main="exponential semivariogramme")
plot(vdist,spher.vgram(distance=vdist,range=vrangle),type="l",xlab="spherical",main="spherical
semivariogramme")
plot(vdist,gauss.vgram(distance=vdist,range=vrangle),type="l",xlab="practical range = 1.7 %",
vrangle,main="Gaussian semivariogramme")
plot(vdist,linear.vgram(distance=vdist,slope=0.3),type="l",xlab="linear",main="linear semivariogramme")
plot(vdist,power.vgram(distance=vdist,slope=0.3,range=0.5),type="l",xlab="power",main="power semivariogramme")

vdist<-1:20/10
vdist
plot(vdist,
spher.vgram
(distance=vdist,range=0.5,sill=0.56,nugget=0.3),type="l",xlab="spherical")
plot(vdist,
0.30*spher.vgram
(distance=vdist,range=0.2,sill=1.0,nugget=1.0) +
0.26*spher.vgram
(distance=vdist,range=1.3,sill=1.0,nugget=0.0),
type="l",xlab="nested spherical structure")

Cd.var<-variogram(Cd~loc(X,Y),data=jura,
azimuth=90,tol.azimuth=90,nlag=20,lag=0.1,
tol.lag=0.05,maxdist=2.0,na.action="na.omit")

```

```
plot(Cd.var,type="b",xlab="omnidirectional variogram of Cd")
Cd.var

Cd.var<-variogram(Cd~loc(X,Y),data=jura,
azimuth=90,tol.azimuth=90,nlag=10,lag=0.2,
tol.lag=0.1,maxdist=2.0,na.action="na.omit")
plot(Cd.var,type="b",xlab="omnidirectional variogram of Cd")
Cd.var
model.variogram(Cd.var,fun=spher.vgram,range=0.2,sill=0.6,nugget=0.2)

sumspherPG.cov<-function(vdist, ...)
{
  spher.cov(distance=vdist,range=0.2,sill=0.3,nugget=0.3, ...) +
  spher.cov(distance=vdist,range=1.3,sill=0.26,nugget=0.0, ...)
}

sumspherCHS.cov<-function(vdist, ...)
{
  spher.cov(distance=vdist,range=0.2,sill=0.29,nugget=0.3, ...) +
  spher.cov(distance=vdist,range=1.3,sill=0.16,nugget=0.0, ...)
}

X259<-X[1:259]
Y259<-Y[1:259]
Cd259<-Cd[1:259]
Cd.krige<-krige(Cd259~loc(X259,Y259),covfun=sumspherPG.cov)

newXY<-jura[260:359,1:2]
newXY
names(newXY)
names(newXY)<-c("X259","Y259")
names(newXY)
Cd.Prd<-predict.krige(Cd.krige, newXY, se.fit=T)
Cd.Prd

JURAC<-read.table("C:/juraset.dat",header=T,skip=17)
CdTrue<-JURAC$Cd[260:359]
CdTrue
var(CdTrue)
var(Cd.Prd$fit)
trellis.device(win.graph)
plot(CdTrue,Cd.Prd$fit,main="estimated vs true Cd values")
```

```

abline(0,1,lty=3,col=8)
abline(0.8,0,lty=3)
good.points<-identify(CdTrue,Cd.Prd$fit)
bad.points<-identify(CdTrue,Cd.Prd$fit)
good.points
bad.points
trellis.device(win.graph)
scaled.plot(X,Y,main="alle 359 Messorte")
points(X[good.points],Y[good.points],col=4)
points(X[bad.points],Y[bad.points],col=8)
polygon(X[jura.hull],Y[jura.hull],dens=0)
length(Cd.Prd$fit[Cd.Prd$fit<CdTrue])
cor.test(CdTrue,Cd.Prd$fit,method="spearman")

trellis.device(win.graph)
plot(Cd.Prd$se.fit,abs(CdTrue-Cd.Prd$fit),main="true error vs. krige standard error of Cd")
good.points<-identify(Cd.Prd$se.fit,abs(CdTrue-Cd.Prd$fit))
bad.points<-identify(Cd.Prd$se.fit,abs(CdTrue-Cd.Prd$fit))
good.points
bad.points
trellis.device(win.graph)
scaled.plot(X,Y,main="alle 359 Messorte")
points(X[good.points],Y[good.points],col=4)
points(X[bad.points],Y[bad.points],col=8)
polygon(X[jura.hull],Y[jura.hull],dens=0)
cor.test(Cd.Prd$se.fit,abs(CdTrue-Cd.Prd$fit),method="spearman")

trellis.device(win.graph)
vdist<-c(seq(0.01,0.09,by=0.01),seq(0.1,2.0,by=0.1))

CdPG.var<-variogram(Cd[1:259]^loc(X[1:259],Y[1:259]),
azimuth=90,tol.azimuth=90,nlag=10,lag=0.2,
tol.lag=0.1,maxdist=2.0,na.action="na.omit")
plot(CdPG.var,type="b",xlab="omnidirectional variogram of 259 Cd")
CdPG.var

Cd.var<-variogram(Cd^loc(X,Y),data=JURAC,
azimuth=90,tol.azimuth=90,nlag=10,lag=0.2,
tol.lag=0.1,maxdist=2.0,na.action="na.omit")
plot(Cd.var,type="b",xlab="omnidirectional variogram of all 359 Cd")
Cd.var

```

```

plot(CdPG.var,type="b",xlab="omnidirectional variogram of 259 Cd")
Cd.varmodelPG<-%@%
spher.vgram(vdist,range=0.2,sill=0.3,nugget=0.3)+spher.vgram(vdist,range=1.3,sill=0.26,nugget=0.3)
lines(vdist,Cd.varmodelPG,col=8)

plot(Cd.var,type="b",xlab="omnidirectional variogram of all 359 Cd")
Cd.varmodel<-spher.vgram(vdist,range=0.2,sill=0.29,nugget=0.3)+spher.vgram(vdist,range=1.3,sill=0.26,nugget=0.3)
lines(vdist,Cd.varmodel)

plot(CdPG.var,type="b",xlab="omnidirectional variogram of 259 Cd")
Cd.varmodel<-spher.vgram(vdist,range=0.2,sill=0.29,nugget=0.3)+spher.vgram(vdist,range=1.3,sill=0.26,nugget=0.3)
Cd.varmodelPG<-%@%
spher.vgram(vdist,range=0.2,sill=0.3,nugget=0.3)+spher.vgram(vdist,range=1.3,sill=0.26,nugget=0.3)
lines(vdist,Cd.varmodel)
lines(vdist,Cd.varmodelPG,col=8)

par(mfrow=c(1,3))
plot(CdPG.var,type="b",xlab="omnidirectional variogram of 259 Cd values")
Cd.varmodelPG<-%@%
spher.vgram(vdist,range=0.2,sill=0.3,nugget=0.3)+spher.vgram(vdist,range=1.3,sill=0.26,nugget=0.3)
lines(vdist,Cd.varmodelPG,col=8)
plot(CdTrue,Cd.Prd$fit,main="estimated vs true Cd values")
abline(0,1,lty=3,col=8)
abline(0.8,0,lty=3)
plot(Cd.Prd$se.fit,abs(CdTrue-Cd.Prd$fit),main="true error vs. krige standard error of Cd")
par(mfrow=c(1,1))

detach("jura")
rm(jura)
rm(X259,Y259,Cd259)
rm(vdist,vrange)
rm(Cd.var,CdPG.var)
rm(Cd.varmodel,Cd.varmodelPG)
rm(sumspherPG.cov, sumspherCHS.cov)
rm(newXY,Cd.krige,Cd.Prd,CdTrue,JURAC)
rm(good.points,bad.points)

```

12.6 Aufgaben der fünften Übung

(i) Stochastisches Schätzen vs deterministische Interpolation

Erläutern Sie den durch die duale Form des SK vermittelten Zusammenhang zwischen "kriging" und Interpolation durch die Anwendung radialer Basisfunktionen. Charakterisieren Sie dazu das Verhältnis zwischen irgendeiner speziellen radialen Basisfunktion und einem angepaßten Modell für die Kovarianzfunktion. An welcher Stelle kommen also die stochastischen Modellannahmen der Stationarität ins Spiel?

(ii) Ein elementares Schätzproblem (ohne PC)

Es soll $Z = Z(\mathbf{x}_i + \mathbf{h})$ durch $Z^* = Z(\mathbf{x}_i)$ geschätzt werden. Geben Sie die entsprechende Schätzvarianz für die Annahme der intrinsischen Stationarität an.

(iii) Experimentelles Semivariogramm und experimentelles logarithmisches Semivariogramm

(a) Erstellen Sie Histogramme der Cd - und der logarithmierten Cd Konzentrationen; erstellen Sie "summary statistics" für die Cd - und die logarithmierten Cd Konzentrationen. Vergleichen Sie die Verteilungen.

(b) Erstellen Sie omnidirektionale experimentelle Semivariogramme für die Cd - und die logarithmierten Cd Konzentrationen. Vergleichen Sie die beiden experimentellen Semivariogramme und interpretieren Sie die zu beobachtenden Unterschiede insbesondere bezüglich ihrer Ursache(n).

(iv) lognormal kriging

Berechnen Sie aus den logarithmierten Cd Konzentrationen für die 100 Testorte, an denen Cd als NA vereinbart wurde, den Krige-Schätzwert $z^*(\mathbf{x}_\alpha)$ für Cd und die entsprechende Krige-Fehlervarianz $s_K^2(bmx)$, d.h. den minimalen Wert der Schätzvarianz unter Verwendung der angepaßten Kovarianzfunktion mit Nugget-Effekt, Schwellenwert und Reichweite.

(a) Vergleichen Sie in einem scatter plot die geschätzten mit den wahren Cd Konzentrationen und berechnen Sie den entsprechenden Rangkorrelationskoeffizient.

(b) Vergleichen Sie in einem scatter plot den wahren absoluten Fehler $|z^*(\mathbf{x}_\alpha) - z(\mathbf{x}_\alpha)|$ mit der Krige-Fehler-Standardabweichung $\sigma_K(\mathbf{x})$ und berechnen Sie den entsprechenden Rangkorrelationskoeffizienten.

(c) Interpretieren Sie beide Vergleiche!

(v) lognormal kriging vs kriging

Vergleichen Sie die Ergebnisse für kriging auf der Grundlage von logarithmierten Cd Konzentrationen mit den Ergebnissen für kriging auf der Grundlage der Cd Konzentrationen.

12.6.1 S-PLUS script file für die Aufgaben der fünften Übung

```

JURAC<-read.table("C:/juraset.dat",header=T,skip=17)
attach(JURAC)
module(spatial)

scaled.plot(X,Y)
JURAC.hull<-chull(X,Y)
polygon(X[JURAC.hull],Y[JURAC.hull],dens=0)
scaled.plot(X[1:259],Y[1:259])
polygon(X[JURAC.hull],Y[JURAC.hull],dens=0)
scaled.plot(X[260:359],Y[260:359])
polygon(X[JURAC.hull],Y[JURAC.hull],dens=0)

CdTrue<-JURAC$Cd[260:359]
jura259<-JURAC[1:259,]
jura100<-JURAC[260:359,]
detach("JURAC")
attach(jura259)
X259<-jura259$X
Y259<-jura259$Y
Cd259<-jura259$Cd

summary(Cd259)
hist(Cd259,nclasss=20)

Cd259.var<-variogram(Cd259~loc(X259,Y),data=jura259,
azimuth=90,tol.azimuth=90,nlag=10,lag=0.2,
tol.lag=0.1,maxdist=2.0,na.action="na.omit")
plot(Cd259.var,type="b",xlab="omnidirectional variogram of 259 Cd")
plot(Cd259.var,type="b",xlab="omnidirectional variogram of 259 Cd")
lags<-c(seq(0.01,0.09,by=0.01),seq(0.1,2.0,by=0.1))
Cd259.varmodel<-spher.vgram(lags,range=0.2,sill=0.3,nugget=0.3)+spher.vgram(lags,range=1
lines(lags,Cd259.varmodel)
Cd259.var

logCd259<-log(Cd259)
logCd259
summary(logCd259)
hist(logCd259,nclass=20)
mz <- mean(Cd259)
my <- mean(logCd259)

```

```

sz2 <- var(Cd259)
sy2 <- var(logCd259)
mz
my
exp(my+0.5*sy2)
sz2
sy2
mz^2.*(exp(sy2)-1.)

RANGE <- 0.8
SILL <- 0.425
NUGGET <- 0.1
logCd259.var<-variogram(logCd259~loc(X259,Y259),data=jura259,
  azimuth=90,tol.azimuth=90, nlag=10, lag=0.2,
  tol.lag=0.1, maxdist=2.0, na.action="na.omit")
plot(logCd259.var,type="b",xlab="omnidirectional variogram of 259 logCd")
logCd259.varmodel<-spher.vgram(lags,range=RANGE,sill=SILL,nugget=NUGGET)
+spher.vgram(lags,range=0.0,sill=0.0)
plot(logCd259.var,type="b",xlab="omnidirectional variogram of logCd")
lines(lags,logCd259.varmodel,col=8)
logCd259.var

logCd259x.var<-variogram(logCd259~loc(X259,Y259),data=jura259,
  azimuth=90,tol.azimuth=90, nlag=20, lag=0.1,
  tol.lag=0.05, maxdist=2.0, na.action="na.omit")
plot(logCd259x.var,type="b",xlab="omnidirectional variogram of logCd")
lines(lags,logCd259.varmodel,col=8)
logCd259x.var

#model.variogram(logCd259.var,fun=spher.vgram,range=RANGE,sill=SILL,nugget=NUGGET)

sumspher.cov<-function(lags, ...)
{
  spher.cov(distance=lags,range=RANGE,sill=SILL,nugget=NUGGET, ...) +
  spher.cov(distance=lags,range=0.001,sill=0.001,nugget=0.001, ...)
}

logCd.krige<-krige(logCd259~loc(X259,Y259),data=jura259,
covfun=sumspher.cov)

newXY<-jura100[,1:2]
names(newXY)<-c("X259","Y259")

```

```
logCd.Prd<-predict.krige(logCd.krige,newXY,se.fit=T)
Cd.back<-exp(logCd.Prd$fit + logCd.Prd$se.fit/2)
Cd.backse<-sqrt(mz^2*exp(sy2)*(1.-exp(-logCd.Prd$se.fit^2.)))

var(CdTrue)
var(Cd.back)
plot(CdTrue,Cd.back,main="estimated vs true Cd values")
abline(0,1,lty=3)
abline(0.8,0,lty=3)
length(Cd.back[Cd.back<CdTrue])
cor.test(CdTrue,Cd.back,method="spearman")
plot(Cd.backse,abs(CdTrue-Cd.back),main="true error vs. krige standard error")
cor.test(Cd.backse,abs(CdTrue-Cd.back),method="spearman")

par(mfrow=c(1,3))
plot(logCd259.var,type="b",xlab="omnidirectional variogram of logCd")
lines(lags,logCd259.varmodel,col=8)
plot(CdTrue,Cd.back,main="estimated vs true Cd values")
abline(0,1,lty=3,col=8)
abline(0.8,0,lty=3)
plot(Cd.backse,abs(CdTrue-Cd.back),main="true error vs. krige standard error of Cd value")
par(mfrow=c(1,1))

detach("jura259")
rm(jura259)
rm(JURAC,JURAC.hull,jura100,CdTrue)
rm(X259,Y259,Cd259)
rm(mz,sz2,my,sy2)
rm(Cd259.var,lags,Cd259.varmodel)
rm(RANGE,SILL,NUGGET,logCd259.var,logCd259x.var,logCd259.varmodel)
rm(sumspher.cov,logCd.krige,newXY)
rm(logCd.Prd,Cd.back,Cd.backse)
```

12.7 Aufgaben der sechsten Übung

(i) kriging der Cd Konzentrationen

Die dem Jura Datensatz zugrundeliegende geometrische Meßpunkt-Konfiguration resultiert aus der Überlagerung eines regelmäßigen quadratischen Rasters mit 107 Meßpunkten im Abstand von 250m und 38 Clustern von je 4 um einen Rasterpunkt geschachtelten Meßpunkten.

- (a) Verschaffen Sie sich einen Überblick über die geometrische Anordnung der 259 Meßpunkte und der 100 Testpunkte. Erkennen Sie ein Muster? Beschreiben Sie es!
- (b) Vergleichen Sie das omnidirektionale experimentelle Semivariogramm der Cd Konzentrationen der 259 Meßpunkte und das omnidirektionale experimentelle Semivariogramm der Cd Konzentrationen aller 359 Meß- und Testpunkte bezüglich der Reichweite, des relativen Nugget-Effekts und des Schwellenwerts. Interpretation!
- (c) Vergleichen Sie die Anzahl der Paare $(Z(x_\alpha), Z(x_\alpha + h))$, aus denen der Wert des empirischen Semivariogramms für h zwischen 0m und 500m berechnet wird!
- (d) Vergleichen Sie beide omnidirektionale empirischen Semivariogramme mit Ihrem angepaßten Modell-Semivariogramm und mit dem angepaßten geschachtelten Modell-Semivariogramm

$$\gamma(h) = 0.30g_0(h) + 0.30\text{sph}(|h|/200) + 0.26\text{sph}(|h|/1300)$$

- (e) Vergleichen Sie vielfach omnidirektionale experimentelle Semivariogramme der Cd Konzentrationen an etwa 259 zufällig ausgewählten Meßpunkten mit dem angepaßten geschachtelten Modell-Semivariogramm aus (d)! Was beobachten Sie?
- (f) Beziehen Sie in die Interpretation der scatterplots der Übung 4 Aufgabe 4 und Übung 5 Aufgabe 4 Mittelwert, Spannweite und Standardabweichung der 259 zum Schätzen benutzten Cd Konzentrationen, der 359 tatsächlich vorhandenen Cd Konzentrationen und der 100 Cd Konzentrationen die zum Vergleich mit den geschätzten benutzt werden.

(ii) Nicht parametrische Geostatistik Indikator kriging (ohne PC)

Ziel des Indikator Zugangs zur Modellierung lokaler Ungewißheit ist eine Näherung der lokalen Verteilung $F(x; z|(n))$ unter Berücksichtigung der als räumlich abhängig angenommenen Meßwerte.

- (a) Welche Näherung liefern die Indikator-Transformierten unter der Annahme der klassischen Statistik, daß die Zufallsvariablen $Z(x_\alpha), \alpha = 1, \dots$, stochastisch unabhängig sind?
- (b) Was ist das Ergebnis von Indikator kriging in jedem Fall?

(iii) **Experimentelles Indikator Semivariogramm**

Erstellen Sie ein omnidirektionales experimentelle Indikator Semivariogramme für den Cd Indikator zum Wert 0.8 (zum Median).

(iv) **Indikator kriging**

Berechnen Sie aus den Cd Indikatorwerten zum Wert 0.8 (zum Median) die kriging Indikator-Schätzwerte für die 100 Testorte, an denen Cd als NA vereinbart wurde. Was bedeuten diese Schätzwerte? Stellen Sie sie kartenmäßig dar und interpretieren Sie diese Karte!

12.7.1 S-PLUS script file für die Aufgaben der sechsten Übung

```

JURAC<-read.table("C:/juraset.dat",header=T,skip=17)
attach(JURAC)
Cd<-JURAC$Cd
Cd259PG<-JURAC$Cd[1:259]
CdTrue<-JURAC$Cd[260:359]

module(spatial)

scaled.plot(X,Y)
JURAC.hull<-chull(X,Y)
polygon(X[JURAC.hull],Y[JURAC.hull],dens=0)
scaled.plot(X[260:359],Y[260:359])
polygon(X[JURAC.hull],Y[JURAC.hull],dens=0)
scaled.plot(X[1:259],Y[1:259])
polygon(X[JURAC.hull],Y[JURAC.hull],dens=0)

Cd.var<-variogram(Cd~loc(X,Y),data=JURAC,azimuth=90,tol.azimuth=90,
  nlag=10,lag=0.2,tol.lag=0.1,maxdist=2.0,na.action="na.omit")
plot(Cd.var,type="b",xlab="omnidirectional variogram of all 359 Cd")
plot(Cd.var,type="b",xlab="omnidirectional variogram of all 359 Cd")
vdist<-c(seq(0.01,0.09,by=0.01),seq(0.1,2.0,by=0.1))
Cd.varmodel<-spher.vgram(vdist,range=0.2,sill=0.29,nugget=0.3)+spher.vgram(vdist,range=1
lines(vdist,Cd.varmodel)
Cd.var

scaled.plot(X[260:359],Y[260:359])
polygon(X[JURAC.hull],Y[JURAC.hull],dens=0)
scaled.plot(X[1:259],Y[1:259])
polygon(X[JURAC.hull],Y[JURAC.hull],dens=0)
CdPG.var<-variogram(Cd[1:259]~loc(X[1:259],Y[1:259]),azimuth=90,tol.azimuth=90,

```

```
nlag=10,lag=0.2,tol.lag=0.1,maxdist=2.0,na.action="na.omit")
plot(CdPG.var,type="b",xlab="omnidirectional variogram of 259 Cd")
plot(CdPG.var,type="b",xlab="omnidirectional variogram of 259 Cd")
Cd.varmodel<-spher.vgram(vdist,range=0.2,sill=0.29,nugget=0.3)+spher.vgram(vdist,range=1
Cd.varmodelPG<-%%@
spher.vgram(vdist,range=0.2,sill=0.3,nugget=0.3)+spher.vgram(vdist,range=1.3,sill=0.26,n
lines(vdist,Cd.varmodel)
lines(vdist,Cd.varmodelPG,col=8)
CdPG.var

mean(Cd)
mean(Cd259PG)
mean(CdTrue)
var(Cd)
var(Cd259PG)
var(CdTrue)
sqrt(var(Cd))
sqrt(var(Cd259PG))
sqrt(var(CdTrue))

rlist<-rank(runif(length(Cd)), na.last = T)
rlist
jura259choice<-JURAC[rlist>100,]
length(jura259choice$Cd)
jura259choice
scaled.plot(jura259choice$X,jura259choice$Y)
polygon(X[JURAC.hull],Y[JURAC.hull],dens=0)

jura100<-JURAC[rlist<=100,]
length(jura100$Cd)
jura100
scaled.plot(jura100$X,jura100$Y)
polygon(X[JURAC.hull],Y[JURAC.hull],dens=0)

detach("JURAC")
rm(JURAC)
attach(jura259choice)
summary(jura259choice)

Cd259<-jura259choice$Cd
Cu259<-jura259choice$Cu
Pb259<-jura259choice$Pb
```

```

X259<-jura259choice$X
Y259<-jura259choice$Y

Cd259.var<-variogram(Cd259~loc(X259,Y259),data=jura259choice,azimuth=90,tol.azimuth=90,
  nlag=10,lag=0.2,tol.lag=0.1,maxdist=2.0,na.action="na.omit")
plot(Cd259.var,type="b",xlab="omnidirectional variogram of 259 randomly chosen Cd")
plot(Cd259.var,type="b",xlab="omnidirectional variogram of 259 randomly chosen Cd")
Cd.varmodel<-spher.vgram(vdist,range=0.2,sill=0.29,nugget=0.3)+spher.vgram(vdist,range=1
lines(vdist,Cd.varmodel)
lines(vdist,Cd.varmodelPG,col=8)

mean(Cd)
mean(jura259choice$Cd)
mean(jura100$Cd)
var(Cd)
var(jura259choice$Cd)
var(jura100$Cd)
sqrt(var(Cd))
sqrt(var(jura259choice$Cd))
sqrt(var(jura100$Cd))

#model.variogram(Cd.var,fun=spher.vgram,range=0.2,sill=0.6,nugget=0.2)

sumspherPG.cov<-function(vdist, ...)
{
  spher.cov(distance=vdist,range=0.2,sill=0.29,nugget=0.3, ...) +
  spher.cov(distance=vdist,range=1.3,sill=0.16,nugget=0.0, ...)
}

sumspherCHS.cov<-function(vdist, ...)
{
  spher.cov(distance=vdist,range=0.2,sill=0.29,nugget=0.3, ...) +
  spher.cov(distance=vdist,range=1.3,sill=0.16,nugget=0.0, ...)
}

sumspher.cov<-sumspherCHS.cov

Cd.krige<-krige(
  Cd259~loc(X259,Y259),covfun=sumspher.cov)

newXY<-jura100[,1:2]
newXY

```

```

names(newXY)
names(newXY)<-c("X259","Y259")
names(newXY)
Cd.Prd<-predict.krige(Cd.krige, newXY, se.fit=T)
Cd.Prd

CdTrue
plot(CdTrue,Cd.Prd$fit,main="estimated vs true Cd values")
abline(0,1,lty=3)
abline(0.8,0,lty=3)
cor.test(CdTrue,Cd.Prd$fit,method="spearman")
plot(Cd.Prd$se.fit,abs(CdTrue-Cd.Prd$fit),main="true error vs. krige standard error")
cor.test(Cd.Prd$se.fit,abs(CdTrue-Cd.Prd$fit),method="spearman")

ICd259<-as.numeric(Cd259<=0.8) # I Indikator zum <Grenzwert, ICd259 = 1 falls Cd <=0.8
JCd259<-1-ICd259 # J Indiaktor = 1 - I
scaled.plot(X,Y,type="n",main="J Indikator zum Cd Grenzwert") # plot
text(X,Y,JCd259) # bin\arer plot
polygon(X[jura.hull],Y[jura.hull],dens=0)

ICd259.var2<-variogram(ICd259~loc(X,Y),data=jura,
azimuth=90,tol.azimuth=90,nlag=10,lag=0.2,
tol.lag=0.1,maxdist=2.0,na.action="na.omit")
plot(ICd259.var2,type="l",xlab="omnidirectional I indicator variogram of treshold 0.8 of
ICd259.var2

ICd259med<-as.numeric(Cd<=quantile(Cd259,0.5,na.rm=TRUE)) # Quantile
ICd259med
scaled.plot(X,Y,type="n",main="I Indikator zum Median Cd") # I Indikator Karte
text(X,Y,ICd259med)
polygon(X[jura.hull],Y[jura.hull],dens=0)

ICd259med.var2<-variogram(ICd259med~loc(X,Y),data=jura,
azimuth=90,tol.azimuth=90,nlag=10,lag=0.2,
tol.lag=0.1,maxdist=2.0,na.action="na.omit")
plot(ICd259med.var2,type="l",xlab="omnidirectional I indicator variogram of median of Cd
ICd259med.var2

detach("jura259choice")
rm(JURAC.hull)
rm(rlist,jura259choice,jura100)
rm(vdist,Cd.var,Cd.varmodel,Cd.varmodelPG,CdPG.var,Cd259.var)

```

```
rm(X259,Y259,Cd259,Cd259PG,Cd)
rm(newXY,Cd.krige,Cd.Prd,CdTrue)
```

12.8 Aufgaben der siebten Übung

- (i) Checkliste kriging von a-z
 - (a) Worin besteht die fundamentale Annahme der klassischen univariaten mathematischen Statistik bezüglich der Zufallsvariablen, als deren Realisationen man die Meßwerte interpretiert?
 - (b) Worin besteht die entsprechende fundamentale Annahme der angewandten Statistik bezüglich der Meßwerte?
 - (c) Wie groß ist die Wahrscheinlichkeit, daß eine stetigverteilte reellwertige Zufallsvariable Z den Wert z_0 aus der Menge der reellen Zahlen annimmt?
 - (d) Was ist die Varianz einer Zufallsvariablen, was ist die Kovarianz von zwei Zufallsvariablen, falls (i) die beiden Zufallsvariablen identisch sind, (ii) die beiden Zufallsvariablen nicht identisch sind?
 - (e) Was mißt der Pearson Korrelationskoeffizient, was mißt der Spearman Korrelationskoeffizient?
 - (f) Was ist lineare Regressionsanalyse und was ist ihr Ziel?
 - (g) Was ist multiple lineare Regressionsanalyse und was ist ihr Ziel?
 - (h) Worin besteht die fundamentale Annahme der (mathematischen) Geostatistik bezüglich der Zufallsfunktion, als deren Realisation man die Meßwerte interpretiert?
 - (i) Worin besteht die entsprechende fundamentale Annahme der (angewandten) Geostatistik bezüglich der Meßwerte?
 - (j) Was ist ein Semivariogramm; was ist eine Kovarianzfunktion?
 - (k) Worin unterscheidet sich kriging von anderen Interpolationsmethoden, die auf der Anwendung radialer Basisfunktionen beruhen?
 - (l) Wozu braucht man in der praktischen Anwendung des kriging die fundamentale Annahme der Geostatistik?
 - (m) Welches sind die Parameter eines Semivariogramms, was bedeuten sie?
 - (n) Wodurch ist der Schwellenwert eines omnidirektionalen (theoretischen) Semivariogramms gegeben, falls er existiert?
 - (o) Benutzt man die a-priori Varianz aller Meßwerte zur Bestimmung des Schwellenwerts eines experimentellen Semivariogramms?
 - (p) Welche Forderungen an einen Schätzer führen zu kriging?
 - (q) Welche Parameter und Eigenschaften eines angepaßten Modell-Semivariogramms bestimmen die kriging Gewichte, von welchen sind die kriging Gewichte unabhängig?

- (r) Werden Meßwerten an Meßorten, deren Distanz größer ist als die Reichweite des (für diese Richtung zuständigen) angepaßten Semivariogramms, kriging Gewichte 0 zugeordnet?
 - (s) Was versteht man unter dem screen effect des kriging?
 - (t) Worin unterscheidet sich kriging insbesondere von der Interpolationsmethode der "quadrierten inversen Distanzen"?
 - (u) In welchem Verhältnis steht kriging zu multipler linearer Regression?
 - (v) Ist ein kriging Schätzwert immer in der Spannweite der Meßwerte? Begründen Sie Ihre Antwort!
 - (w) Unter welcher zusätzlichen Annahme kann man die kriging Varianz benutzen, um mit ihrer Hilfe Konfidenzintervalle für Schätzwerte zu bestimmen?
 - (x) Was ist die Indikator-Transformierte einer Zufallsvariablen?
 - (y) Welcher Zusammenhang besteht zwischen der Verteilung einer Zufallsvariablen und dem Erwartungswert ihrer Indikator-Transformierten?
 - (z) Was ist das Ergebnis von Indikator kriging?
- (ii) Nicht parametrische Geostatistik Indikator kriging
 Ziel des Indikator Zugangs zur Modellierung lokaler Ungewißheit ist eine Näherung der lokalen Verteilung $F(\mathbf{x}; z|n)$ unter Berücksichtigung der als räumlich abhängig angenommenen Meßwerte.
- (a) Welche Näherung liefern die Indikator-Transformierten unter der Annahme der klassischen Statistik, daß die Zufallsvariablen $Z(\mathbf{x}_\alpha)$, $\alpha = 1, \dots, n$, stochastisch unabhängig sind?
 - (b) Was ist das Ergebnis von Indikator kriging in jedem Fall?
- (iii) kriging der Cd Konzentrationen
 Die dem Jura Datensatz zugrundeliegende geometrische Meßpunkt-Konfiguration resultiert aus der Überlagerung eines regelmäßigen quadratischen Rasters mit 107 Meßpunkten im Abstand von 250m und 38 Clustern von je 4 um einen Rasterpunkt geschachtelten Meßpunkten.
- (a) Verschaffen Sie sich einen Überblick über die geometrische Anordnung der 259 Meßpunkte und der 100 Testpunkte.
 - (b) Erkennen Sie ein Muster? Beschreiben Sie es!
 - (c) Vergleichen Sie das omnidirektionale experimentelle Semivariogramm der Cd Konzentrationen der 259 Meßpunkte und das omnidirektionale experimentelle Semivariogramm der Cd Konzentrationen aller 359 Meß- und Testpunkte bezüglich der Reichweite und des relativen Nugget-Effekts und des Schwellenwerts. Interpretation!

(d) Vergleichen Sie die Anzahl der Paare $(Z(\mathbf{x}_\alpha), Z(\mathbf{x}_\alpha + \mathbf{h}))$, aus denen der Wert des empirischen Semivariogramms für \mathbf{h} zwischen 0m und 500m berechnet wird!

(e) Vergleichen Sie beide omnidirektionale empirischen Semivariogramme mit Ihrem angepaßten Modell-Semivariogramm und mit dem angepaßten geschachtelten Modell-Semivariogramm

$$\gamma(\mathbf{h}) = 0.30\gamma_0(\mathbf{h}) + 0.30\text{sph}(|\mathbf{h}|/200) + 0.26\text{sph}(|\mathbf{h}|/1300)$$

(f) Vergleichen Sie vielfach omnidirektionale experimentelle Semivariogramme der Cd Konzentrationen an etwa 259 zufällig ausgewählten Meßpunkte mit dem angepaßten geschachtelten Modell-Semivariogramm aus (e)! Was beobachten Sie?

(g) Beziehen Sie in die Interpretation der scatterplots der Übung 4 Aufgabe 4 und Übung 5 Aufgabe 4 Mittelwert, Spannweite und Standardabweichung der 259 zum Schätzen benutzten Cd Konzentrationen, der 359 tatsächlich vorhandenen Cd Konzentrationen und der 100 Cd Konzentrationen die zum Vergleich mit den geschätzten benutzt werden.

(iv) Experimentelles Indikator Semivariogramm

Erstellen Sie ein omnidirektionales experimentelle Indikator Semivariogramme für den Cd Indikator zum Wert 0.8 (zum Median).

(v) Indikator kriging

Berechnen Sie aus den Cd Indikatorwerten zum Wert 0.8 (zum Median) die kriging Indikator-Schätzwerte für die 100 Testorte, an denen Cd als NA vereinbart wurde. Was bedeuten diese Schätzwerte? Stellen Sie sie kartenmäßig dar und interpretieren Sie diese Karte!

Chapter 13

Synopsis

aus: Schaeben, H., 1988, Geology and Mathematics - Progressing Mathematization of Geology: Geol. Rdschau 77, 591-607

Characterization and management of natural resources - Recoverable ore reserve estimation: geostatistics

So far we have confined ourselves to rather conceptually oriented models, therefore leaving the question of accuracy and precision of numbers in geology unmentioned. Here, we face a completely different problem. Geostatistics stems from economic and engineering/mining geology. It was developed in response to problems of reserve estimation, assignment of confidence limits to estimates, and the spatial variability of the quantity to be estimated, eg ore grade. Thus, it is essentially devoted to the problem of accurately and precisely estimating recoverable ore reserves because of their economic significance. To this end, geostatistics designs measures of spatial continuity and applies the theory of stochastic processes and thus statistics of dependent events. To be more specific, geostatistics offers a solution to the problem of estimating a regionalized variable, the two major characteristics of which are that it is a random function (not a family of identically and independently distributed stochastic variables) defined on a given spatial domain and that it can be related to different (not fixed) supports at each point of the domain of definition. Thus, its mathematics/statistics is designed to account for mining experiences that ore grades in neighboring places are likely to be similar and tend to be independent beyond a range of influence, and that the variability of the ore grade depends largely on the volume of its reference support. Actually, the ore grade of a given ore body is a deterministic though unknown quantity, while the ore grade of crystal grains may fluctuate between 0 and 100%. To avoid any possible misreading, geostatistics does not provide a model of the geological processes that created the ore body.

In contrast to trend surface analysis which addresses a value at a datum point as the sum of a deterministic function and a random error, geostatistics assumes a value at a datum point as the sum of a regional component and a residual. This makes a large difference. As WHITTEN (1984) concludes 'trend surface analysis ... provides an inventory of the data

for the samples studied. While the method can be used for testing some hypotheses, the descriptive role has been dominant together with (commonly misleading) implicit or explicit statements about interpolation and/or extrapolation from the sampled to the more-extensive target populations'. However, '... geostatistics started elaborating their own methods and their own mathematical formalism which is nothing else than an abstract formulation of secular mining experience. This formalism has inherited from its statistical origin a language in which one still speaks of variance and covariance, including however in those notions a new content. This similarity in vocabulary must not deceive', MATHERON (1963). The main tools of geostatistics are the variogram and an interpolation scheme called kriging. Textbooks of geostatistics in this sense of the Fontainebleau School are AKIN & SIEMES (1988), CLARK (1979), DAVID (1977), JOURNEL & HUIJBREGTS (1978), RENDU (1978), ROYLE et al (1971/75)

For a rather pragmatical but to me also revitalizing view of geostatistics and particularly of kriging and the various variances associated with it the reader is urgently referred to JOURNEL (1986b). Only recently it has been elaborated that spline interpolation is equivalent to kriging with a given generalized covariance, DUBRULE (1983, 1984), MATHERON (1981), SALKAUSKAS (1982), WATSON (1984). As kriging when reduced to a mere interpolation method is often seen as one of the few interpolation methods of genuine geoscientific origin, it should be appealing to geologists to find kriging so closely related to the mathematical theory and practice of splines, the most advanced technique of numerical evaluation. Also, new applications to interpolation subject to constraints appear possible now, see DUBRULE & KOSTOV (1986), JOURNEL (1986a). In practical applications the assumption of some kind of stationarity (strong, weak, or intrinsic) and the assumption of Gaussian normality in some instances has been critically debated ever since geostatistics surfaced from the Fontainebleau school of mines, MATHERON (1963, 1971); an early discussion of this topic is given by MATHERON (1978) himself.

With respect to mathematically modeling assumptions concerning (multi-) normality only the latest contribution (JOURNEL 1987) in removing a remaining Gaussian flavor shall be mentioned. Summarizing, geology and especially geostatistics seem to be most instrumental and demanding in developing a non-parametric distribution-free statistical science as suggested and designed for instance by PARZEN (1977, 1979).

The problem of assumptions concerning stationarity shall be discussed in more detail, as it may illuminate the dialectics of deterministic and stochastic approaches in mathematical modeling in geology. In a real world problem, a geostatistician has to cope with an inhomogeneous, anisotropic phenomenon which was regionally ill-distributedly sampled; each of these features amounting to a highly erratic variogram. Thus, much effort has been undertaken to find robust estimators of the variogram within the concept of classical geostatistics which assumes the phenomenon to be studied is an incompletely, ie discretely sampled single realization of a stochastic process, the parameters of which have to be estimated to be used in some spatial prediction procedures. Now, the assumption of some kind of stationarity is required to do any statistically based calculations, and next the implicit hypothesis of ergodicity which guarantees the equality of ensemble and spatial averages is used to read available ensemble

means as spatial means. This probabilistic framework is most useful when the primary interest is a better understanding of the underlying genetic processes as required for instance for extrapolation purposes since extrapolation always requires an implicit faith in some underlying genetic process. Within this framework the use of the classical variogram and its estimators seems favorable.

However, if the primary interest is to use what the data reveal about the peculiarities of the sampled domain most efficiently in some interpolation procedure for the well defined finite domain, spatial averages should be used directly since ensemble and spatial average are no longer equal. Thus, if not interested in the stochastic process but exclusively in the unique realization the ergodicity hypothesis may be dropped and all calculations based on spatial averages. This deterministic side of geostatistics, see JOURNEL (1985), MATHERON (1971), has just now been fully developed into a deterministic non-ergodic framework by ISAAKS & SRIVASTAVA (1987) to meet the latter requirements.

This framework essentially relates to the finite domain of interest; in terms of stochastic processes, it provides a view through a finite window at eg a Wiener-Levy process. Then, all the spatially truncated moments exist but to the expense that they do no longer equal the ensemble moments. Therefore, now the spatial covariance can be directly estimated and used to interpolate.

Essentially the same question of deterministic vs stochastic methodology arises eg in the statistical analysis of geochemical data which are most likely not to be drawn from a Gaussian population. Since the sample size is always finite, all empirical moments exist. They may be used as parameters to describe this individual sample but should not be interpreted as estimates of the parameters of the true distribution of the population nor be used in statistical inference, unless some additional knowledge or reasonable assumptions concerning the distribution are available.

Today, geostatistical methods are applied to a vast variety of problems in analyzing spatially distributed data.

Literatur

AKIN, H., SIEMES, H.(1988): Praktische Geostatistik - Eine Einführung für den Bergbau und die Geowissenschaften. - Springer, Heidelberg

CLARK, I.(1979): Practical geostatistics. - London (Applied Science Publishers).

DAVID, M.(1977): Geostatistical ore reserve estimation. - Amsterdam (Elsevier).

DUBRULE, O.(1984): Comparing splines and kriging. - Computers and Geosciences 10, 327 - 338.

DUBRULE, O.(1983): Two methods with different objectives: splines and kriging. - Math. Geol. 15, 245 - 257.

DUBRULE, O., KOSTOV, C.(1986): An interpolation method taking into account inequality constraints: I. Methodology. - Math. Geol. 18, 33 - 51. II. Practical approach. - Math. Geol. 18, 53 - 73.

ISAAKS, E.H., SRIVASTAVA, R.M.(1987): Spatial continuity measures for probabilistic and deterministic geostatistics. - Preprint, paper presented at the MGUS '87 Conference held at Redwood City, CA, April 13 - 15, 1987.

JOURNEL, A.G.(1987): New distance measures: the route towards truly non-gaussian geostatistics. - Preprint, paper presented at the MGUS '87 Conference held at Redwood City, CA, April 13 - 15, 1987.

JOURNEL, A.G.(1986a): Constrained interpolation and qualitative information - the soft kriging approach. - Math. Geol. 18, 269 - 286.

JOURNEL, A.G.(1986b): Geostatistics - models and tools for the earth sciences. - Math. Geol. 18, 119 - 140.

JOURNEL, A.G.(1985): The deterministic side of geostatistics. - Math. Geol. 17, 1 - 15.

JOURNEL, A.G., HUIJBREGTS, C.(1978): Mining geostatistics. - London (Academic Press).

KRIGE, D.C.(1978): Lognormal-de Wijsian geostatistics for ore evaluation. - Johannesburg (South African Institute of Mining and Mineralogy).

MATHERON, G.(1981): Splines and kriging: their formal equivalence. - in Merriam, D.F., (ed.), Down to earth statistics: solutions looking for geological problems. - Syracuse University Geology Contribution 8, 17 - 27.

MATHERON, G.(1978): Estimer et choisir. - Fontainebleau (ENS des Mines de Paris).

MATHERON, G.(1971): The theory of regionalized variables and its applications. - Fontainebleau (Les cahiers du centre de morphologie mathematique de Fontainebleau 5).

MATHERON, G.(1963): Principles of geostatistics. - Econ. Geol. 58, 1246 - 1266.

PARZEN, E.(1979): A density-quantile function perspective on robust estimation. - in Lammer, R.L., Wilkinson, G.N. (eds.), Robustness in statistics, 237 - 258, New York (Academic Press).

PARZEN, E.(1977): Nonparametric statistical data science (a unified approach based on density estimation and testing for 'white noise'). Buffalo, NY (Technical report 47, Statistical

Science Division, SUNY).

RENDU, J.M.(1978): An introduction to geostatistical methods of mineral evaluation. - Johannesburg (South African Institute of Mining and Mineralogy).

ROYLE, A.G.(1971/75): A practical introduction to geostatistics (Course Manuscript, 3 Parts including geostatistical tables and exercises in geostatistics): Leeds (Department of Mining and Mineral Sciences, The University of Leeds).

SALKAUSKAS, K.(1982): Some relationships between surface splines and kriging. - in Schempp,W., Zeller,K. (eds.), Multivariate approximation theory II, 313 - 325, Proceedings of the conference held at the Mathematical Research Institute at Oberwolfach, W-Germany, Feb. 8-12(1982), Basel (Birkhäuser Verlag).

TURCOTTE, D.L.(1986a): A fractal approach to the relationship between ore grade and tonnage. - Economic Geology 81, 1528 - 1532.

WATSON, G.S.(1984): Smoothing and interpolation by kriging and with splines. - Math. Geol. 16, 601 - 615.

Chapter 14

Checkliste – kriging 1,2,3,...

- (i) Worin besteht die fundamentale Modellannahme der klassischen mathematischen Statistik bezüglich der Zufallsvariablen/Zufallsvektoren, als deren Realisationen man die Meßwerte interpretiert?
- (ii) Worin besteht die entsprechende fundamentale Modellannahme der angewandten Statistik bezüglich der Meßwerte?
- (iii) Wie groß ist die Wahrscheinlichkeit, daß eine stetigverteilte reellwertige Zufallsvariable Z den Wert $z_0 \in \mathbb{R}^1$ aus der Menge der reellen Zahlen annimmt, $P(Z = z_0) = ?$
- (iv) Wie groß ist die Wahrscheinlichkeit, daß die stetigverteilte reellwertige Zufallsvariable $Z(x_i)$, $x_i \in D$, Werte größer als $z \geq z_i$ aus der Menge der reellen Zahlen unter der Bedingung der Beobachtungen $Z(x_j) = z_j$, $j = 1, \dots, n$, annimmt, $P(Z(x_i) > z | Z(x_j) = z_j, j = 1, \dots, n) = ?$ für $z \geq z_i$.
- (v) Wie häufig nimmt der I -Indikator zum Median eines stetigverteilten ortsabhängigen Merkmals in einer entsprechenden Karte den Wert 1 an?
- (vi) Unter welcher Voraussetzung an die flächenhafte Verteilung der zugehörenden Meßorte ist die Häufigkeit, mit der der I -Indikator zum Median eines stetigverteilten ortsabhängigen Merkmals in einer entsprechenden Karte den Wert 1 annimmt, eine näherungsweises Maß für den Flächenanteil, der durch die Bedingung $Z(x) \leq \text{Median}$ charakterisiert ist?
- (vii) Was ist die Varianz einer Zufallsvariablen, was ist die Kovarianz von zwei Zufallsvariablen, falls (i) die beiden Zufallsvariablen identisch sind, (ii) die beiden Zufallsvariablen nicht identisch sind?
- (viii) Was mißt der Pearson Korrelationskoeffizient, was mißt der Spearman Korrelationskoeffizient?
- (ix) Was ist lineare Regressionsanalyse und was ist ihr Ziel?

- (x) Was ist multiple lineare Regressionsanalyse und was ist ihr Ziel?
- (xi) Worin besteht die fundamentale Modellannahme der mathematischen Geostatistik bezüglich der Zufallsfunktion, als deren einzige Realisation man die Meßwerte interpretiert, damit man Statistik betreiben kann?
- (xii) Worin besteht die entsprechende fundamentale Modellannahme der angewandten Geostatistik bezüglich der Meßwerte?
- (xiii) Was bedeutet die Stationaritätsannahme der mathematischen Geostatistik bezüglich einer Zufallsfunktion für die Ortsabhängigkeit oder Ortsunabhängigkeit der zugehörigen Zufallsvariablen und ihrer Zuwächse (ohne Formeln!)
- (xiv) Kann man die Stationaritätsannahme der mathematischen Geostatistik bezüglich der Zufallsfunktion mit einem statistischen Test prüfen?
- (xv) Wozu braucht man in der praktischen Anwendung des kriging die Stationaritätsannahme der Geostatistik?
- (xvi) Was beschreibt ein Semivariogramm; was beschreibt eine Kovarianzfunktion?
- (xvii) Worin unterscheidet sich kriging von anderen Interpolationsmethoden, die auf der Anwendung radialer Basisfunktionen beruhen?
- (xviii) Wie groß ist der "Nuggeteffekt" eines theoretischen Semivariogramms, $\gamma(0) = ?$
- (xix) Wie wird der Nuggeteffekt eines experimentellen Semivariogramms bestimmt?
- (xx) Was beschreibt der Nuggeteffekt eines Semivariogramms?
- (xxi) Wie wird die Reichweite eines experimentellen Semivariogramms bestimmt?
- (xxii) Was beschreibt die Reichweite eines Semivariogramms?
- (xxiii) Für zwei ortsabhängige isotrope Merkmale $Z_A(x), Z_B(x), x \in D$, gelte für die Reichweiten ihrer Semivariogramme $R_A < R_B$. Für welches Merkmal ist der räumliche Zusammenhang stärker ausgeprägt?
- (xxiv) Wie wird der Schwellenwert eines experimentellen Semivariogramms bestimmt?
- (xxv) Für welche Größe ist der Schwellenwert eines beschränkten experimentellen Semivariogramms ein Schätzer?
- (xxvi) Gibt es unbeschränkte Semivariogramme?
- (xxvii) Wodurch ist der Schwellenwert eines omnidirektionalen (theoretischen) Semivariogramms gegeben, falls er existiert?

- (xxviii) Wie groß ist die Reichweite eines reinen Nugget-Effekt Semivariogramms?
- (xxix) Warum ist die empirische Varianz aller Meßwerte zur Bestimmung des Schwellenwerts eines experimentellen Semivariogramms selbst im isotropen Fall im allgemeinen nicht geeignet?
- (xxx) Welche Eigenschaft eines experimentellen Semivariogramms kann man als Diagnostik interpretieren, daß die Modellannahme der Stationarität offensichtlich nicht erfüllt ist?
- (xxxi) Welche Forderungen an einen Schätzer führen zu kriging?
- (xxxii) Hängen die nach einfachem oder gewöhnlichem kriging bestimmten Gewichte von den beobachteten Meßwerten ab?
- (xxxiii) Welche Parameter und Eigenschaften eines angepaßten Modell-Semivariogramms bestimmen die kriging Gewichte, von welchen sind die kriging Gewichte unabhängig?
- (xxxiv) Werden Meßwerten an Meßorten, deren Distanz vom Schätzort größer als die Reichweite des (für diese Richtung zuständigen) angepaßten Semivariogramms ist, im allgemeinen kriging Gewichte 0 zugeordnet?
- (xxxv) Was versteht man unter dem screen effect des kriging?
- (xxxvi) Worin unterscheidet sich kriging insbesondere von der Interpolationsmethode der "quadrierten inversen Distanzen"?
- (xxxvii) In welchem Verhältnis steht kriging zu multipler linearer Regression?
- (xxxviii) Ist ein kriging Schätzwert immer in der Spannweite der Meßwerte? Begründen Sie Ihre Antwort!
- (xxxix) Aus einer Stichprobe $z(x_i), i = 1, \dots, n$, werden Schätzwerte $\hat{z}_j = \hat{z}(x_j), j = 1, \dots, m, x_i \neq x_j$ für alle i, j , der wahren Werte $z(x_j)$ nach gewöhnlichem kriging berechnet. Für welche Menge $\{\hat{z}_j | j = 1, \dots, m\}, \{z_j | j = 1, \dots, m\}$ ist die unter Vernachlässigung der Ortsabhängigkeit berechnete empirische Varianz kleiner?
- (xl) Zeigen Sie, daß für einen Meßort x_α der SK-geschätzte mit dem gemessenen Wert übereinstimmt, also $z *_{SK}(x_\alpha) = z(x_\alpha)$ für jeden beliebigen Meßort x_α .
- (xli) Hängt die kriging Varianz von den beobachteten Meßwerten ab?
- (xlii) Wie kann man die kriging Varianz interpretieren?
- (xliii) Unter welcher zusätzlichen Modellannahme kann man die kriging Varianz benutzen, um mit ihrer Hilfe Konfidenzintervalle für Schätzwerte zu bestimmen?
- (xliv) Welche theoretische Eigenschaft der kriging-Schätzmethode wird durch die beiden Abbildungen empirisch bestätigt?

- (xlv) Welche theoretische Eigenschaft der kriging-Varianz wird durch die beiden Abbildungen empirisch bestätigt?
- (xlvi) Was ist die Indikator-Transformierte einer Zufallsvariablen?
- (xlvii) Welcher Zusammenhang besteht zwischen der Verteilung einer Zufallsvariablen und dem Erwartungswert ihrer Indikator-Transformierten?
- (xlviii) Was ist das Ergebnis von Indikator kriging?