Anwendung von Rang- und log-normaler Transformation zur Verbesserung der räumlichen Strukturanalyse (Variographie) und der Schätzung mittels Kriging am Beispiel von geochemischen Daten aus dem Oderbruch.

Bachelor-Arbeit im Rahmen des Studiengangs Statistik

vorgelegt durch:

Claudio Checchia

Sommersemester 2005

Fachbereich Geowissenschaften
Fachrichtung Geoinformatik
Freie Universität Berlin
Malteser Str. 74-100
12249 Berlin

Selbständigkeitserklärung

Ich habe die vorliegende Arbeit selbständig verfasst. Andere als die angegebenen Hilfsmittel habe ich nicht benutzt. Die Arbeit wird gleichzeitig dem Fachbereich Geowissenschaften, Fachrichtung Angewandte Geologie zur Korrektur vorgelegen.

Berlin, den

(Claudio Checchia)

Inhaltsverzeichnis

1. Datengrundlage und Zielstellung 1
1.1 Einleitung 1
1.2 Das Oderbruch 1
1.3 Deskriptive Analyse der Daten 2
1.4 Räumliche Verteilung der Messpunkte 5

2 Das Verfahren 6
2.1 Variographie und Kriging, Kreuzvalidation 6
2.2 Transformation 10
2.2.1 log-normale Transformation 10
2.2.2 Rangtransformation 12
2.3 Variogramme 13

3. Ergebnisse der Schätzung und Interpretation 16
3.1 Isolinienkarten der Natrium-Schätzwerte 16
3.2 Bestimmung der oberen Grenze nach log-normaler Transformation 21
3.3 Bestimmung der oberen Grenze nach Rangtransformation 22
3.4 Vergleich der Schätzungen 23

4. Zusammenfassung 29

5. Literatur

6 Anhang

1 Datengrundlage und Zielstellung
1.1 Einleitung

Eine wichtige Zielstellung von geochemischen Untersuchungen ist die Schätzung von Werten an Stellen, für die keine Beobachtung vorliegt. Solche Schätzungen sind aussagekräftiger, wenn sie mit Konfidenzintervallen, also mit Wahrscheinlichkeiten, belegt werden.
Ist die Verteilung eines chemischen Parameters bekannt, lassen sich Konfidenzintervalle angeben, anhand der beschreibenden Verteilungs-Parameter Mittelwert und Streuung kann man leicht obere Grenzen ausrechnen, also Werte, die mit einer vorgegebenen Wahrscheinlichkeit nicht überschritten werden.

Leider entsprechen zahlreiche Verteilungen von im Gelände beobachteten Messgrößen nicht einer Normalverteilung. Um dennoch Konfidenzintervalle angeben zu können, transformiert man die Daten derart, dass sie einer bekannten Verteilung folgen.
Als geeignet werden hier die log-normale Transformation und die Rangtransformation angesehen, da sie den Einfluss von Ausreißern reduzieren.

Eine Gegenüberstellung der Ergebnisse schließt diese Arbeit ab. Zweck der Untersuchung ist es zu zeigen, inwieweit Daten-Transformationen Auswirkungen auf die Schätzergebnisse haben.

Die Arbeit gliedert sich in folgende Teile:

a) kurze Beschreibung des Untersuchungsgebiets

b) Deskriptive Statistik

c) Krigingverfahren

d) Transformation der Daten und räumliche Analyse

e)Darstellung und Vergleich der Ergebnisse

1.2 Das Oderbruch

Die rund 800 qkm weite Ebene wurde in ihrer heutigen Ausprägung durch jahrhunderte lange Begradigung der Oder und Trockenlegung gewonnen. Sie erstreckt sich in NW-SE-Richtung parallel zum Flusslauf auf einer Länge von 56 km mit einer durchschnittlichen Breite von 15 km und ist von Hochflächen begrenzt. Da sich die hier behandelten Messpunkte auf deutscher Seite befinden, stellt die Oder im Osten und im Nordosten die Begrenzung des untersuchten Gebiets dar.
Das Oderbruch ist ein Gebiet, das tiefer als die Oberfläche der eingedeichten Oder liegt. Durch diese Besonderheit herrschen effluente Verältnisse: Das Oderwasser drückt das Grundwasser des Oderbruchs hoch. Zahlreiche Entwässerungsgräben, zum Teil Altarme der Oder, im nördlichen Teil der Flusslauf der Alten Oder, entwässern das Gelände mit einer generellen Fließrichtung nach NW (LIEDHOLZ, 2001).

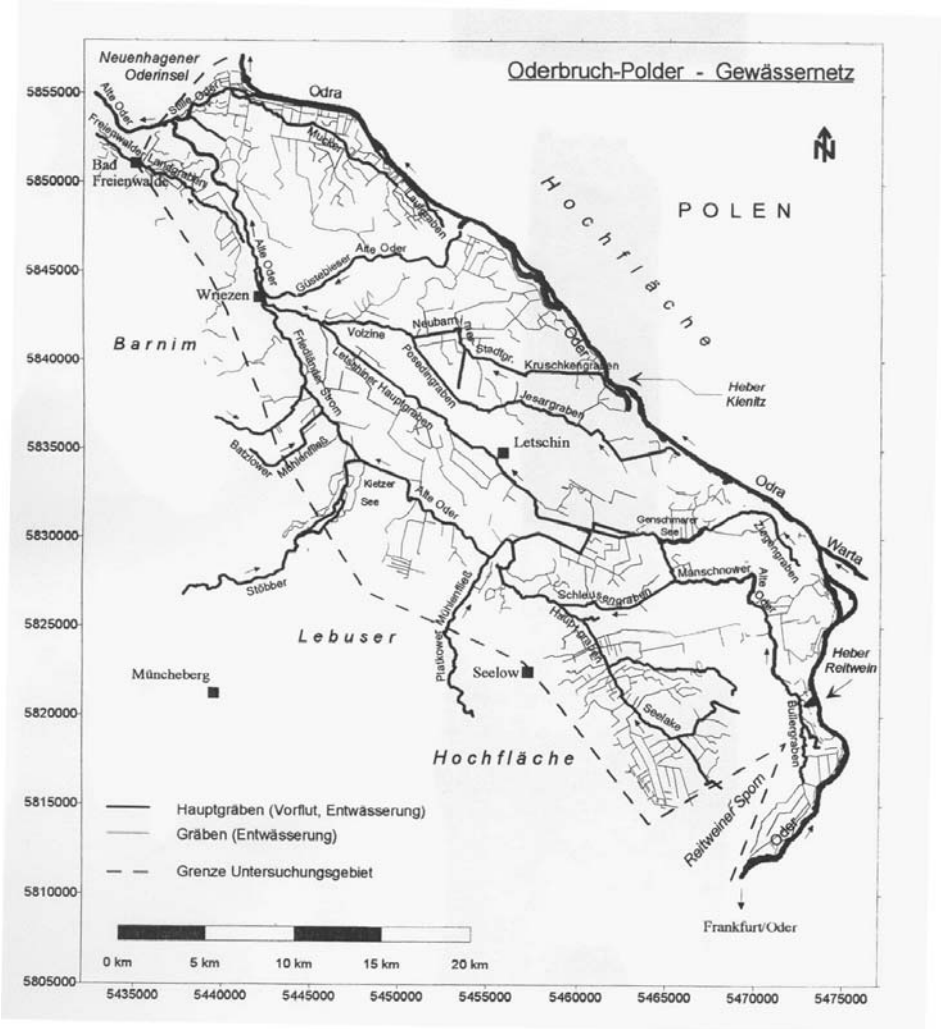

Abb. 1: Oderbruchkarte mit Entwässerungsnetz (aus LIEDHOLZ, 2001).

1.3 Deskriptive Analyse der Daten

Hydrochemische Daten, gewonnen aus dem Oderbruch-Grundwasser durch Beprobung an insgesamt 116 Messstellen, bilden die Datengrundlage dieser Untersuchung.
Aus diesen Daten werden die Natrium-Gehalte der 116 Proben herausgesucht und sind Gegenstand aller folgenden Berechnungen.
Zunächst wird auf die statistische Verteilung der Natrium-Werte eingegangen.

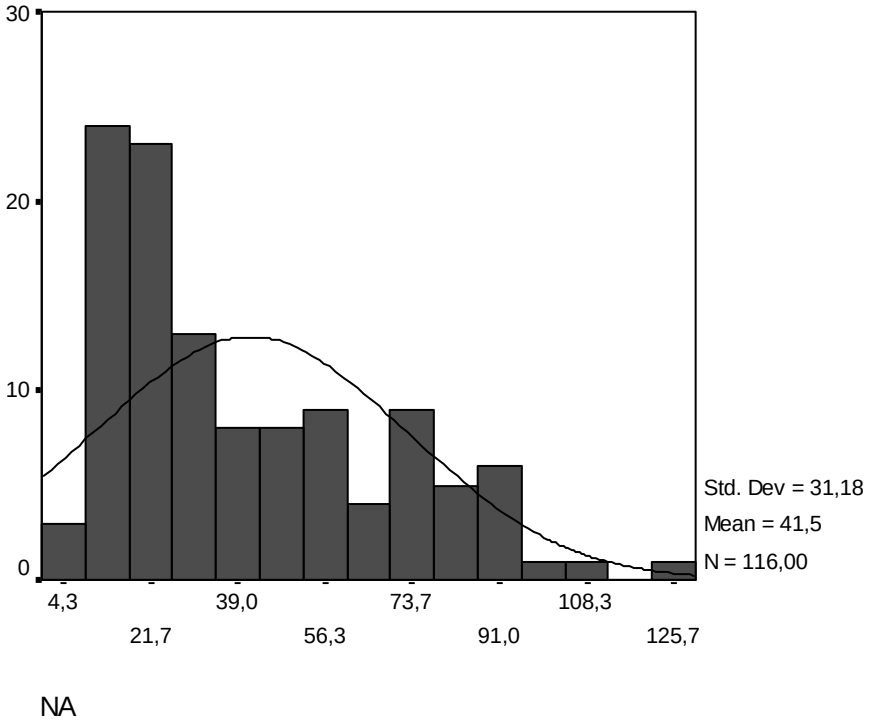

Abb. 2: Statistische Verteilung der Natrium-Werte

Wie in Abb. 2 deutlich zu sehen, ergeben die Daten eine linksschiefe Verteilung. Mittels des Tests nach Kolmogorov-Smirnov wird überprüft, inwieweit die dargestellte Verteilung als normalverteilt bezeichnet werden kann.

One-Sample Kolmogorov-Smirnov Test

		NA
N		116
Normal Parameters	Mean	41,4922
	Std. Deviation	31,1760
Most Extreme Differences	Absolute	,167
	Positive	,167
	Negative	-,124
Kolmogorov-Smirnov Z		1,801
Asymp. Sig. (2-tailed)		,003

a Test distribution is Normal.
b Calculated from data.

Tab. 1: K-S-Test

Tabelle 1 zeigt, dass die Annahme einer Normalverteilung mit einer Irrtumswahrscheinlichkeit von 0,03 abgelehnt werden kann.

Es folgt eine Überprüfung auf Ausreißer.

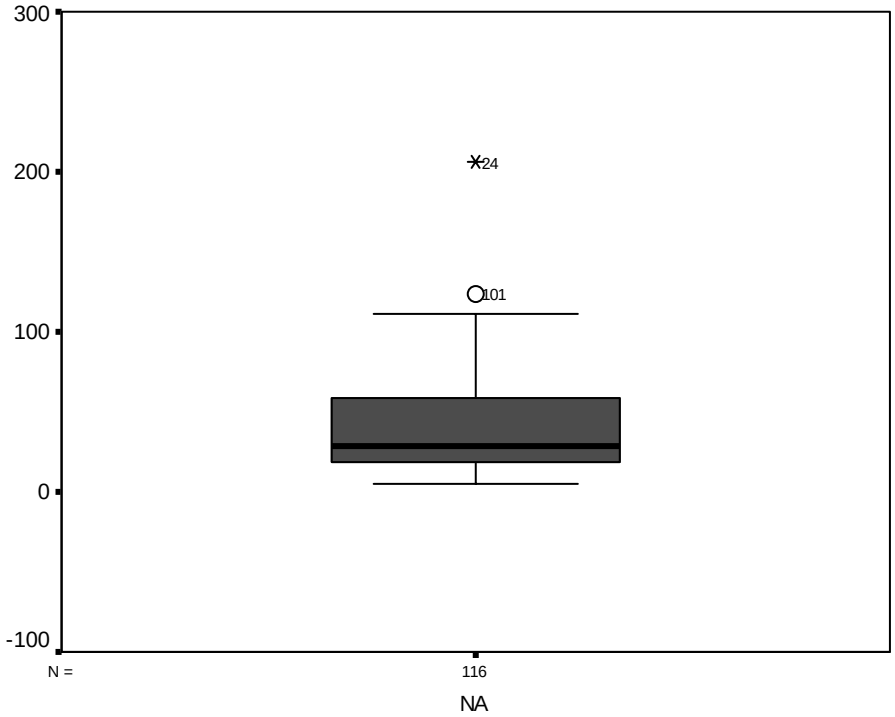

Abb. 3: Box-Plot der Natrium-Werte.

Die Boxplot-Darstellung weist auf die Existenz von zwei möglichen Ausreißern. Es sind die Nr. 24 und 101, die höchsten Natrium-Werte. Ausreißer nach unten hin sind nicht auszumachen. Um einschätzen zu können, wie stark die Ausreißer von den übrigen Werten abweichen, werden die Summen der quadratischen Abweichungen vom Mittelwert – einmal einschließlich der Ausreißer (Nenner), ein weiteres Mal ohne Ausreißer (Zähler) – in Beziehung gebracht. Dieses Verhältnis ist auch Basis des Ausreißer-Tests nach GRUBBS und BECK (siehe GRUBBS et al, 1972).
Die Werte werden der Größe nach aufsteigend geordnet, dann wird folgende Testgröße berechnet:

$$\frac{SQA_{(n-2)}}{SQA_{(n)}} = \frac{\sum_{(i)=1}^{n-2} \left(x_{(i)} - \overline{x}_{(n-2)}\right)^2}{\sum_{i=1}^{n} \left(x_i - \overline{x}_n\right)^2} = 0{,}69249$$

Bei einer deutlichen Abweichung von 1 können die beiden größten Messwerte als Ausreißer betrachtet werden. Im vorliegendem Datensatz ergibt sich ein Testwert von 0,69249<<1, sodass die Messpunkte 24 und 101 als Ausreißer bezeichnet werden können. Sie werden bei der Variogramm-Berechnung nicht berücksichtigt (siehe Kap. 2.1); für die Kriging-Schätzungen werden sie jedoch einbezogen, da ein gänzlicher Ausschluss dieser Werte eine Verzerrung der Datengrundlage bedeuten würde.

1.4 Räumliche Verteilung der Messpunkte

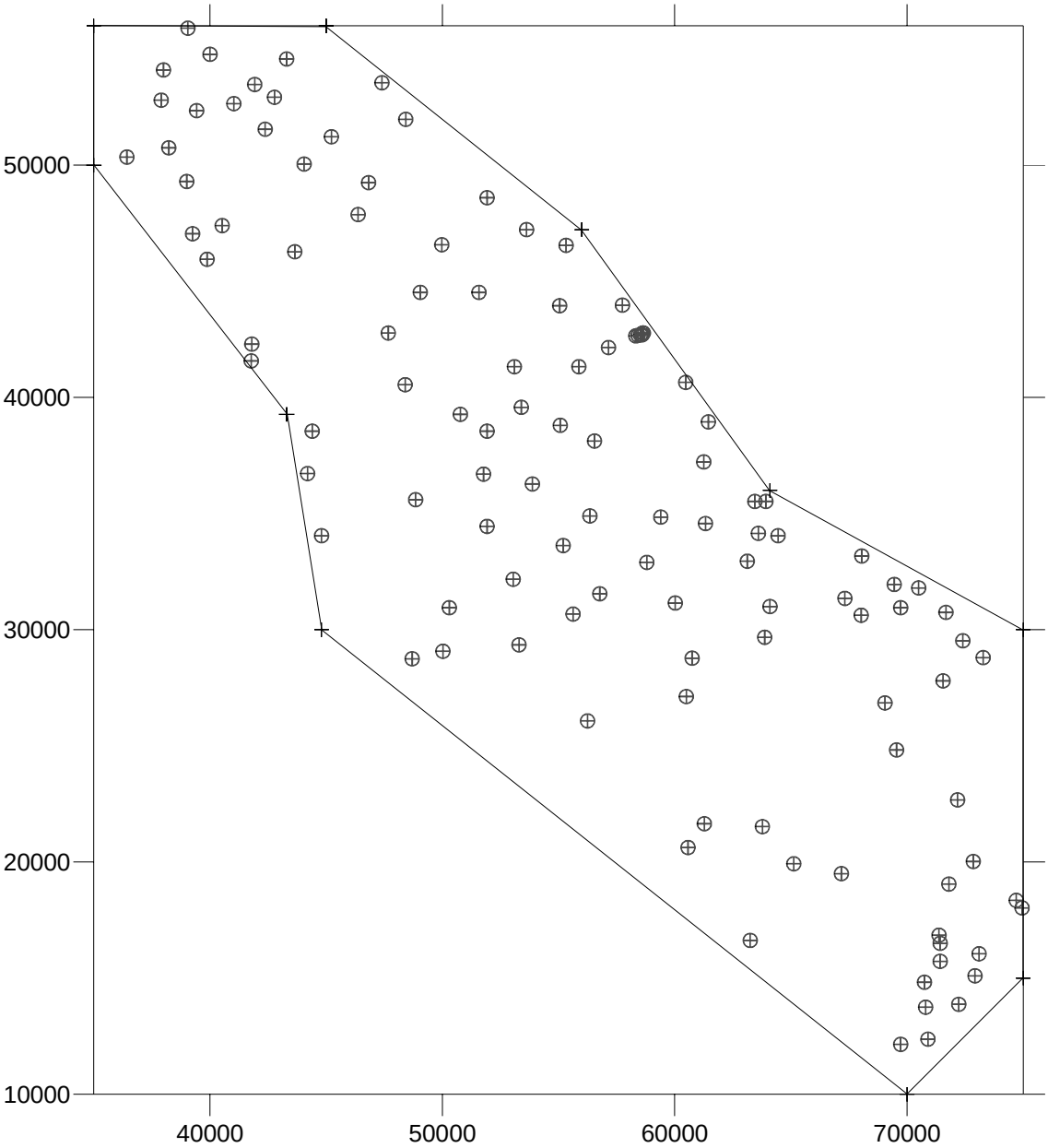

Abb. 4: Räumliche Verteilung der Messstellen

Abb. 4 zeigt die Lage der untersuchten Messpunkte. Berücksichtigt man die Größe des untersuchten Geländes, kann man die Verteilung der Messstellen als sehr gut bezeichnen. Lücken sind lediglich im südlichen Bereich des Geländes erkennbar; Clusterartige Punktwolken, die eine höhere Messstellendichte widerspiegeln, sind für die folgenden Schätzungen nicht von Nachteil.

2. Das Verfahren
2.1Variographie und Kriging, Kreuzvalidation

Die räumliche Punktschätzung (in dieser Arbeit: im zweidimensionalen Raum) beinhaltet die Einbeziehung der gemessenen Daten mit einem Verfahren, welches näher gelegene Messwerte stärker berücksichtigt, als weiter entfernte. Es wird kaum von Interesse sein, flächendeckend den Mittelwert anzugeben, sondern man wird bemüht sein, lokale Minima und Maxima herauszuarbeiten und durch Isolinien Bereiche gleicher Konzentration zu kennzeichnen.

Ausgangspunkt für die Geostatistik ist die Annahme, dass man an jedem Punkt x_i des untersuchten Gebietes G einen unbekannten Wert z hat, der als Zufallsvariable Z charakterisiert wird. Um die Ortsabhängigkeit der Zufallsvariablen zu verdeutlichen, schreibt man $Z(x_i)$. Die Stelle x_i stammt aus einem n-dimensionalen Raum und ist folglich ein Vektor mit n Komponenten (in der vorliegenden Arbeit betrachtet man einen zweidimensionalen Raum, da die dritte Dimension durch die oberflächennahe Beprobung vernachlässigt werden kann). Eine Menge von ortsabhängigen, d.h. regionalisierten Zufallsvariablen in einem Gebiet G heißt Zufallsfunktion. Eine Messkampagne liefert k Beobachtungen, d.h. $z(x_i)$, $i=1,2,\ldots,k$, ist eine Realisation der Zufallsfunktion.

Um die räumliche Variabilität zu beschreiben, betrachtet man die Kovarianzen zwischen den Zufallsvariablen $Z(x_i)$ und $Z(x_j)$, $i \neq j$. Alternativ kann man schreiben: Z(x) und Z(x+h), mit h als den Abstand zwischen den zwei Punkten. Man erhält dann die Autokovarianzfunktion:

$$Cov[Z(x),Z(x+h)] = E\,[(Z(x) - E[Z(x)])\,(Z(x+h) - E[Z(x+h)])]$$

man bildet das Variogramm, definiert als Varianz der Inkremente (Differenzen):

$$\begin{aligned} 2\gamma[Z(x),Z(x+h)] &= Var[Z(x) - Z(x+h)] \\ &= E\,[((Z(x) - Z(x+h)) - E[(Z(x) - Z(x+h))])^2] \end{aligned}$$

Um an diese theoretische Funktion ein Modell anpassen zu können, werden einige vereinfachende Annahmen getroffen.
Strenge Stationarität, oder Stationarität I. Ordnung, bedeutet, dass sich die Verteilungsfunktion der Zufallsvariablen Z(x) bei Verschiebung nicht ändert (translation invariance, s. WACKERNAGEL, 1995). Diese Annahme wird abgeschwächt: Es wird Stationarität der ersten beiden Momente Erwartungswert und Varianz der Zufallsfunktion angenommen (Stationarität II. Ordnung).
Es gilt dann: Die regionalisierte Zufallsvariable Z(x) ist dann stationär (II. Ordnung), wenn im Betrachtungsraum G der Erwartungswert konstant, d.h. unabhängig vom Ort x ist und die Autokovarianzfunktion nur von h abhängt:

Erwartungswert: $E[Z(x)] = \mu$
Autokovarianz: $E[(Z(x) - \mu)\,(Z(x+h) - \mu)] = C(h)$

Eine darüber hinaus gehende Abschwächung erfolgt durch die Annahme, dass Stationarität der ersten beiden Momente der Differenzen (Inkremente) von Punktepaaren gilt (interinsische Hypothese).

Da das Variogramm als Varianz der Inkremente Z(x) – Z(x+h) definiert ist, ergibt sich:

$$2\gamma(h) = Var[Z(x) - Z(x+h)] = E\,[(Z(x) - Z(x+h))^2],$$

und es hängt nur von h ab.

Das Variogramm ist eine graphische Darstellung der lokalen Variabilität der Daten. Es bildet die Basis aller geostatistischen Untersuchungen.

Zunächst wird das experimentelle Variogramm berechnet:

$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (z(x_i) - z(x_i + h))^2,$$

mit N(h)=Anzahl der Probenpaare im Abstand h.

Bei unregelmäßiger Punkteverteilung im Raum wird dieser in Abstands- und Richtungsklassen geteilt und der Abstand h definiert ein Intervall, wobei alle Messpunkte, die sich innerhalb der Abstandsintervallen befinden, zur Berechnung herangezogen werden.

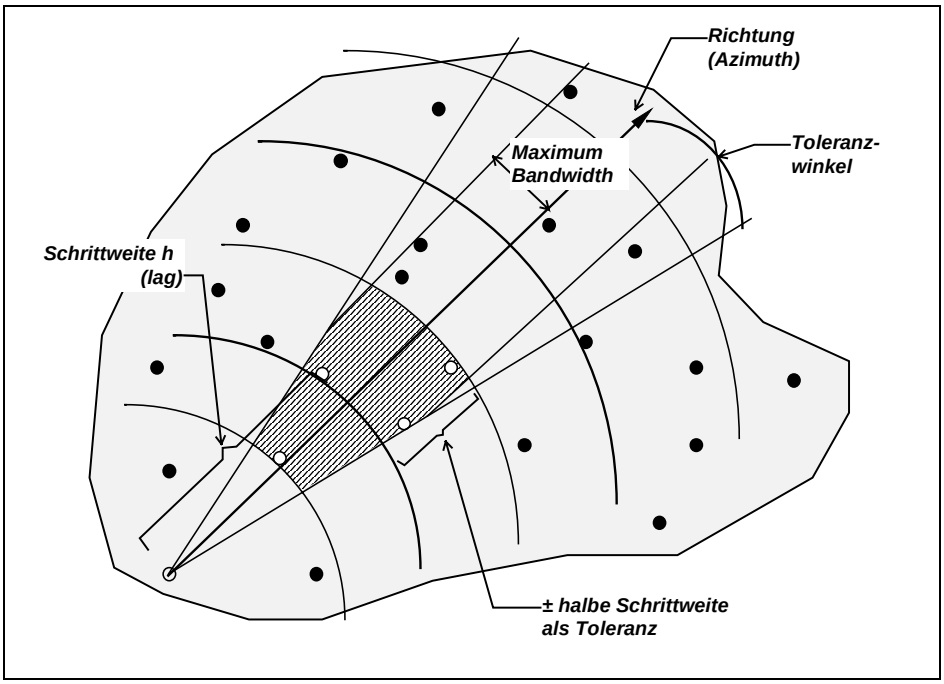

Abb.5 Empirische Variogrammberechnung (aus BURGER et al.)

Abbildung 5 zeigt ein Beispiel für eine Geländeeinteilung zur Berechnung eines Variogramms. Die in weiß hervorgerufene Messpunkte innerhalb des schraffierten Bereiches sind diejenigen Messpunkte, die zur Berechnung herangezogen werden. Es folgen die Berechnungen für 2h, 3h, usw. bis zu einer maximalen Distanz, die die halbe Gebietslänge nicht übersteigen sollte. Der Vorgang wird für jeden Messpunkt wiederholt, d.h. das Gebiet wird flächendeckend untersucht.

Die errechneten Variogrammwerte werden gegen den Abstand h aufgetragen und ergeben so eine Kurve, an die eine Modellfunktion angepasst werden kann.
Zulässig sind positiv definite Modelle. Das geschätzte Modell folgt im untersuchten Fall einer Exponentialfunktion, bzw. einem sphärischen Modell:

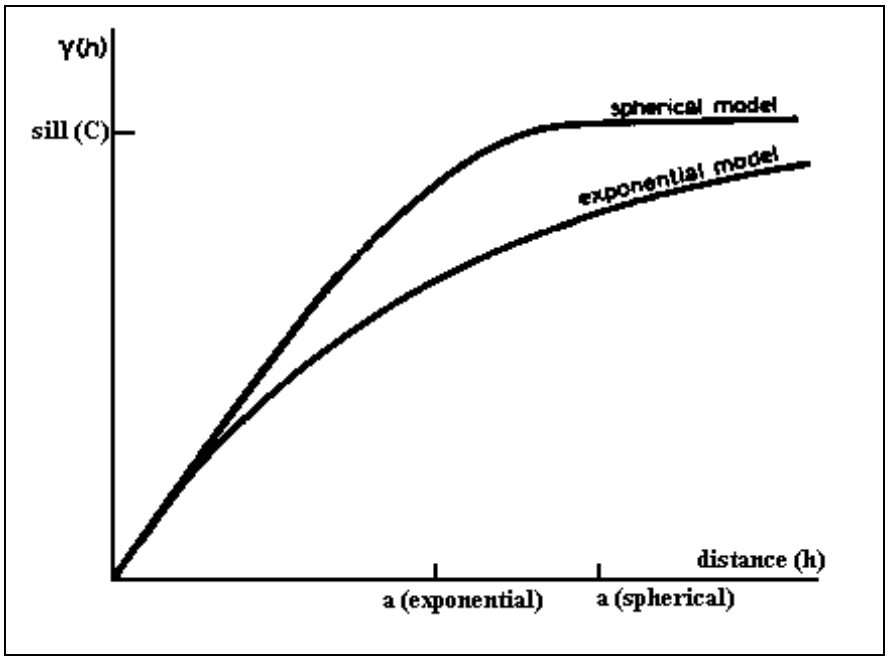

Abb. 6: Darstellung des sphärischen und des exponentialen Modells (aus BURGER et al.)

Das exponentiale Modell wird beschrieben durch:

$$\gamma_{(h)} = C[1 - \exp(-\frac{|h|}{a})]$$

Es wird im wesentlichen durch zwei Kenngrößen bestimmt: Die Reichweite (engl. range, mit a bezeichnet) und der Grenz- oder Schwellenwert (engl. sill, mit C bezeichnet). Die Reichweite gibt an, in welchem Radius die umgebenden Messpunkte Einfluss auf einen Punkt haben, d.h. in welchem Radius die umliegenden Punkte mit dem betrachteten Punkt korreliert sind.

Das sphärische Modell ist im Kurvenverlauf ähnlich und wird beschrieben durch:

$$\gamma_{(h)} = C[1{,}5\frac{|h|}{a} - 0{,}5(\frac{|h|}{a})^3] \qquad \text{für } |h| \le a\,;$$

und $\gamma_{(h)} = C$ für $|h| > a$.

Das Variogramm berücksichtigt anisotropes Verhalten von Variablen durch Verzerrung der Suchradien entlang einer bevorzugten Richtungsachse.
Die berechneten Variogramme werden in Kap. 3 angegeben.

Die Kriging-Methode (nach dem südafrikanischen Geologen D.G. KRIGE, s. AKIN et al. 1988, ISAAKS et al. 1989) besteht in der Aufgabe, den besten linearen unverzerrten Schätzer (best linear unbiased estimator, kurz BLUE) zu finden. Sie basiert also auf der Kleinst-Quadrate-Methode.
Es werden an der Stelle x_0 den umgebenden $Z(x_i)$-Werten Gewichte zugeteilt.
Die Gewichte λ_i erhält man durch Lösung eines linearen Gleichungssystems unter folgenden Voraussetzungen:

a) Unverzerrtheit:

$$E[z^*(x_0)-\mu_0] = 0$$

b) eines minimalen Schätzfehlers:

$$E[(z^*(x_0)-\mu_0)^2] = MIN$$

mit $\mu_0 = E[Z(x_0)]$

Bedingung a) führt zur Gleichung:

$$\Sigma \lambda_i = 1$$

Bedingung b führt auf das lineare Gleichungssystem:

$$\begin{pmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1n} & 1 \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & 1 \\ \gamma_{n1} & \gamma_{n2} & \dots & \gamma_{nn} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \eta \end{pmatrix} = \begin{pmatrix} \gamma_{10} \\ \gamma_{20} \\ \vdots \\ \gamma_{n0} \\ 1 \end{pmatrix}$$

mit i,j = 1,…,n; $\gamma_{ij} = \gamma(x_i-x_j)$, d.h. der Wert für den betrachteten Abstand wird am modellierten Variogramm abgelesen und mit η als Lagrange-Multiplikator, über dem die Nebenbedingung $\Sigma \lambda_i = 1$ eingebracht wird.
Nach Umrechnung liefert das obige Gleichungssystem die Gewichte λ_i.

Den gewichteten Schätzwert z* erhält man mit:

$$z^*(x_0) = \Sigma \lambda_i Z(x_i)$$

Für die Kriging-Schätzvarianz ergibt sich mit diesen Gewichten:

$$s^{*2} = E[(z^*(x_0)-\mu_0)^2] = \Sigma \lambda_i \gamma(x_0-x_i)-\mu_0$$

Das Kriging-Verfahren liefert somit für jeden beliebigen Punkt innerhalb des untersuchten Geländes einen Schätzwert z* und einen Schätzfehler s* (siehe Kap. 3 für die Auswertung).

Die Methode der Kreuzvalidation nimmt eine wichtige Rolle bei der Suche nach einem optimalen Variogrammmodell ein. Sie besteht darin, einzelne gegebene Messpunkte zunächst auszuschließen, eine Schätzung vorzunehmen und schließlich aus den entstandenen Abweichungen eine Residualanalyse durchzuführen. Eine Beurteilung der Schätzung kann anhand der Abweichung des Schätzwertes $z^*(x_i)$ vom Datenwert $z(x_i)$ vorgenommen werden. Die Schätzung sollte ohne systematischen Fehler sein, d.h.:

$$E[z^*(x_i) - z(x_i)] = 0$$

Durch Berücksichtigung der Kriging-Standardabweichung σ_i erhält man die standardisierte Differenz:

$$E[z^*(x_i) - z(x_i)/ \sigma_i] = 0$$

Außer als bei der Auswahl des besten Modells für das experimentelle Variogramm behilflich zu sein, kann die Kreuzvalidation eingesetzt werden, um lokale Ausreißer zu entdecken, sowie um das optimale Schätzverfahren zu bestimmen.

In der vorliegenden Arbeit wird die mittlere Differenz zum Quadrat

$$1/N \sum_{i=1}^{N} (z^*(x_i) - z(x_i))^2$$

bzw. die mittlere standardisierte Differenz zum Quadrat

$$1/N \sum_{i=1}^{N} (z^*(x_i) - z(x_i))^2/ \sigma_i$$

errechnet, um das jeweils beste Modell für das experimentelle Variogramm zu bestimmen (s. Kap. 2.3).

2.2 Transformationen

Die Datenanalyse in Kap. 1 ergeben, dass die Natrium-Werte nicht einer Normalverteilung folgen. Um die Vorteile einer bekannten Verteilung nutzen zu können, müssen die Daten sinnvoll transformiert werden.

2.2.1 Log-normale Transformation

Nach näherem Betrachten der Daten liegt nahe, eine lognormale Verteilung anzunehmen (s. Kap. 1.3). Um dies zu überprüfen, werden die Logarithmen ausgerechnet und in einem Histogramm aufgetragen (s. auch LIMPERT et al. 2001):

Es gilt:

$y = \ln(Z)$

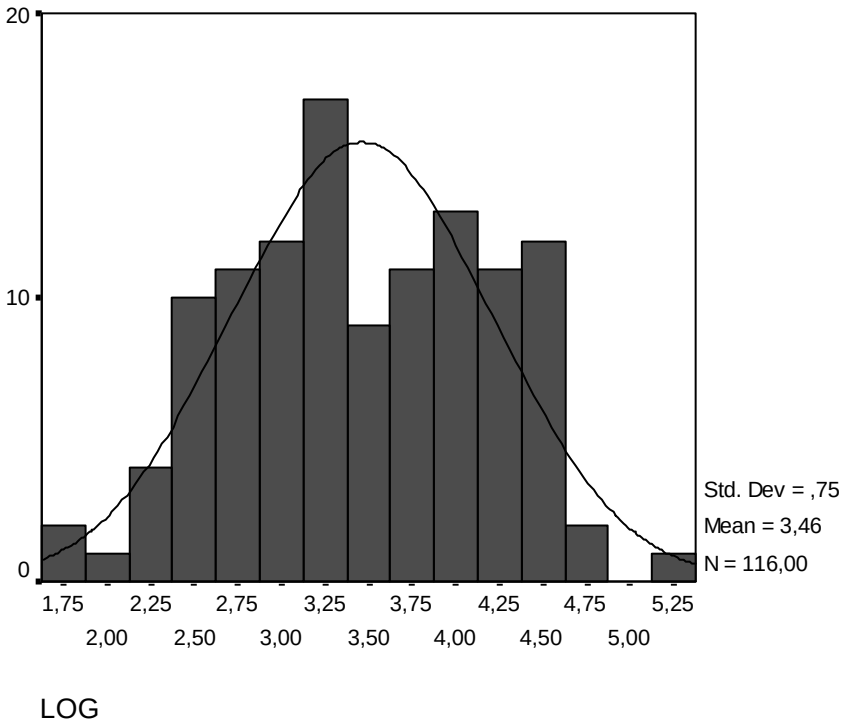

Abb. 7: Log-normale Verteilung der Daten

Im Vergleich zu Abb. 2, ist das Histogramm der logarithmierten Werte symmetrischer und näher einer Normalverteilung.

One-Sample Kolmogorov-Smirnov Test

		LOG
N		116
Normal Parameters	Mean	3,4619
	Std. Deviation	,7462
Most Extreme Differences	Absolute	,081
	Positive	,057
	Negative	-,081
Kolmogorov-Smirnov Z		,871
Asymp. Sig. (2-tailed)		,434

a Test distribution is Normal.
b Calculated from data.

Tab. 2: K-S-Test der logarithmierten Werte

Die Hypothese einer Normalverteilung lässt sich nicht ablehnen.

Die Berechnungen der Schätzungen mittels Kriging werden mit den transformierten Werten vorgenommen. Die Schätzwerte werden mit y*, bzw. s_y* für die geschätzte Standardabweichung bezeichnet. Anschließend müssen diese rücktransformiert werden. Dies erfolgt für jeden geschätzten Gitterpunkt durch die Funktion z* = exp(y*), bzw. s* = exp(s_y*). Diese Werte werden als Median und Streufaktor bezeichnet (SACHS).

2.2.2 Rangtransformation

DAVID (1981) und BALAKRISHNAN (1998) unterscheiden strikt zwischen *order statistics* und *rank order statistics* mit dem Hinweis, dass *order statistics* nicht nur den Rang der Daten berücksichtigt, sondern auch den Betrag. BÜNING (1994) weist daraufhin, dass der offensichtliche Informationsverlust durch Berücksichtigung der Ränge allein erfahrungsgemäß zu keiner Ungenauigkeit bei den Ergebnissen führt.
Im Rahmen dieser Untersuchung dienen die Rangdaten (*rank order*) als Datengrundlage für die spätere Schätzung.
Die Natrium-Werte werden der Größe nach geordnet und jedem Datenwert wird ein Ordnungsindex von 1 bis n vergeben. Eine Normierung durch $r = i/(n+1)$ liefert die Rangvariable mit Werten zwischen 0 und 1. Ein Problem stellen in der Regel Bindungen dar, also gleiche Ausgangswerte für verschiedene Messpunkte (s. BÜNING et al., 1994). Da sie im untersuchten Datensatz nur dreimal vorkommen, kann ihr Einfluss als sehr gering, bzw. vernachlässigbar klein eingestuft werden. Die Rangwerte ergeben selbstverständlich eine Gleichverteilung.
Nach Berechnung der r*-Schätzwerte durch Kriging erfolgt die Rücktransformation auf die Natrium-Schätzwerte z*. Die Rückrechnung erfolgt durch eine Rückrechentabelle, die durch die Rangtransformation entstanden ist. Für die Werte, die außerhalb des durch die Rangwerte definierten Intervalls liegen (siehe Kap. 3, Bestimmung der oberen Grenze) geschieht dies durch eine Polynomfunktion, die durch das Auftragen der Natrium-Ausgangswerten gegen die dazu ermittelten Rangwerte geschätzt werden kann. Die Schätzwerte für die Standardabweichung s_r* werden ebenfalls zu s* rücktransformiert.

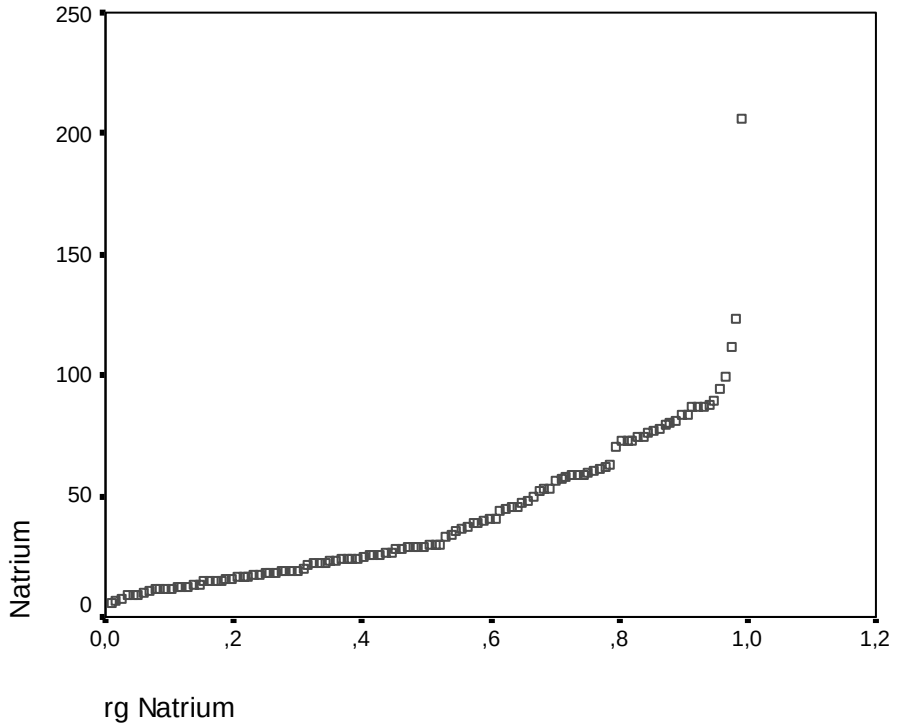

Abb. 8: Scatter-plot der Natrium-Werte gegen die dazugehörigen Rangwerte

Obige Kurve wurde durch die nichtlineare Regressionsfunktion geschätzt:

$z^* = 8{,}511 + 99{,}369 * r^{*2}$

Die Überprüfung der Regressionsfunktion fällt mit einem Bestimmtheitsmaß von 0,894 zufrieden stellend aus.

2.3 Variogramme

Durch die gegebene Grundwasserfließrichtung kann für das untersuchte Gebiet kein isotropes Verhalten der Variablen angenommen werden. Es werden daher Variogramme geschätzt, die ihre maximale Reichweite in Grundwasserfließrichtung (entlang einer SE-NW-Achse) haben. Senkrecht dazu ergibt sich die Richtung minimaler Reichweite. Der sich ergebenden Ellipse wird der Suchradius angepasst.
Es werden Variogramme für die untransformierten, die logarithmierten, sowie für die rangtransformierten Natriumwerte berechnet. Die Berechnungen erfolgten durch das Programm SURFER, dabei ist es für den späteren Vergleich wichtig, bei allen Schätzungen dieselben Klassenintervalle (Abstand und Richtung) zu verwenden.
Es werden zunächst die Natrium-(Ausgangs-)Werte untersucht:

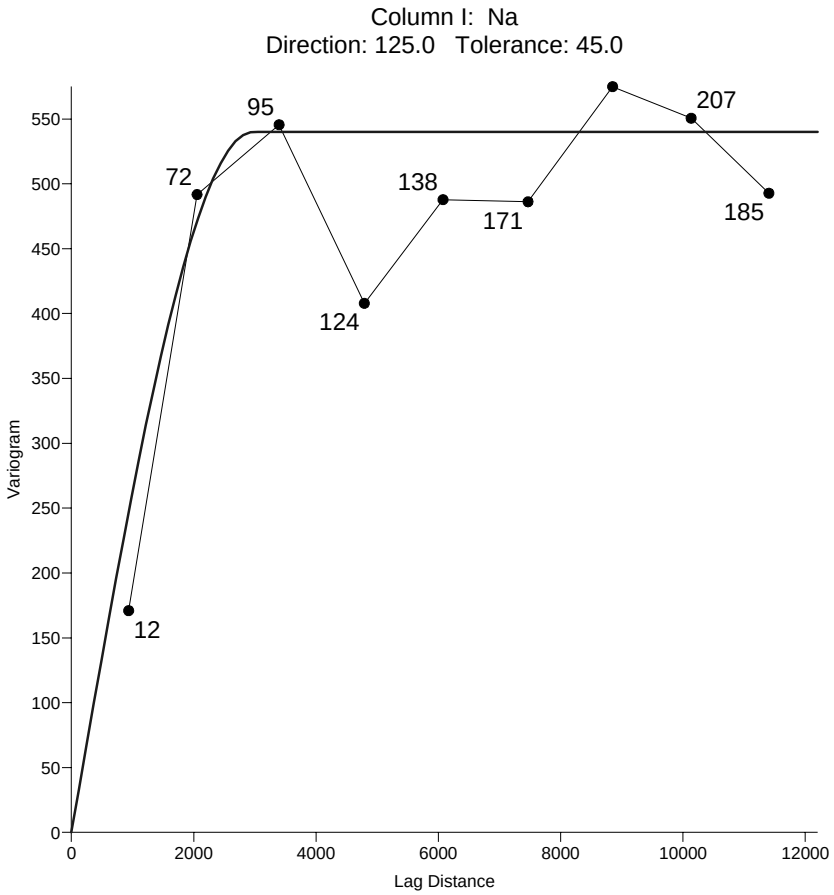

Abb. 9: Variogramm der Na-Werte

Aufgetragen sind die Entfernung der Klassen-Intervalle (auf der Abszisse) und die zugehörige Daten-Variabilität (Ordinate). Es sind neun Intervalle dargestellt, wobei die Anzahl der Intervalle frei wählbar ist. Die untersuchte Richtung wird mit 125°[1] angegeben.

[1] Auf einer Skala, bei der 0° Osten (und damit einer Ost-West-Achse) entspricht, 90° Norden (Nord-Süd-Achse) und 125° Nord-Nordwesten (NNW-SSE-Achse), beschreibt sie eine Richtung parallel zum Oderlauf.

In der Abbildung ist der rasche Anstieg der Daten-Variabilität mit zunehmender Entfernung erkennbar; ab einer Entfernung von etwa 3000 m ist kein gegenseitiger Einfluss mehr vorhanden.
Für jede Abstandsklasse ist ferner die Anzahl der Punktepaare angegeben, die in die Berechnung eingegangen sind. Die dicke Kurve kennzeichnet das aus den empirischen Daten geschätzte Modell, welches für die weiteren Berechnungen (s. Kap.3) verwendet wird.
Der Anisotropiefaktor ergibt sich durch das Verhältnis des größten zum kleinsten Radius der betrachteten Fläche. Um den Anisotropiefaktor mit der geringsten Varianz zu finden, wurde das Kreuzvalidation-Verfahren angewandt: Die Ergebnisse dieses 'Herantastens' sind in Tab. 3 enthalten.

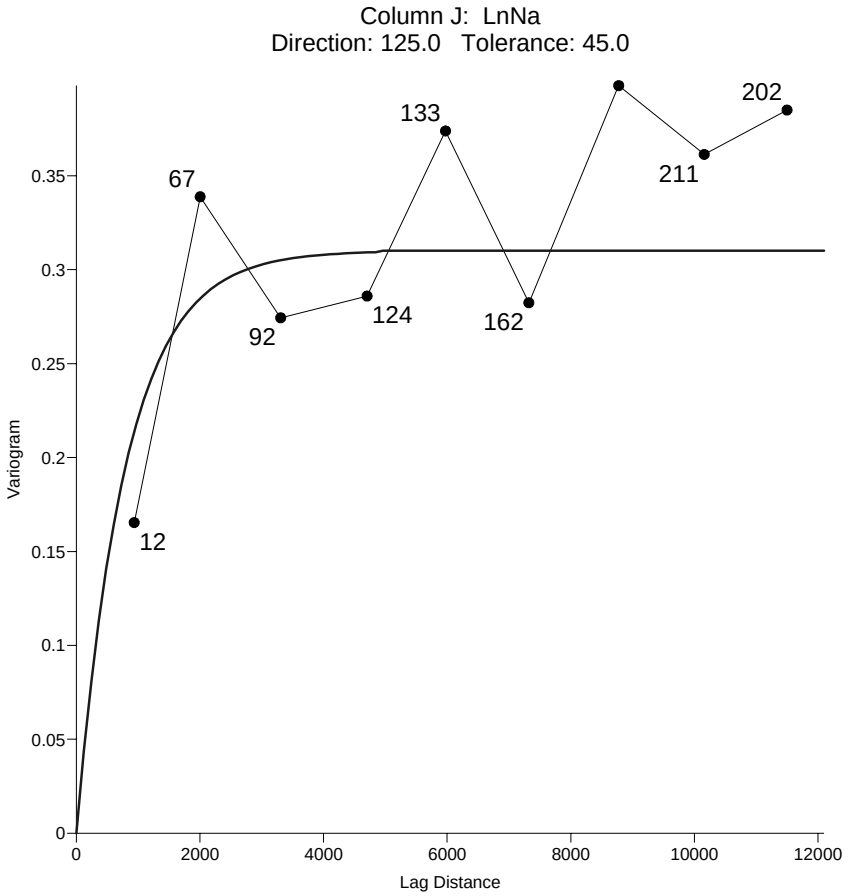

Abb. 10: Variogramm der logarithmierten Werte

Das Variogramm der log-normal transformierten Werte lässt sich im Gegensatz zu dem der nicht transformierten nur schwer durch ein Modell anpassen. Die unterschiedlich hohe Variabilität der Daten mit zunehmender Entfernung wird durch die logarithmierten Werte deutlicher.
Es folgt die Betrachtung des Variogramms der rangtransformierten Werte (Abb. 11). Dieses lässt sich ebenfalls nur in grober Annäherung modellieren. Das Modell ist dem der logarithmierten Werte ähnlich.

Erwartungsgemäß zeigen alle Variogramme vergleichbare Reichweiten. Die Grenz- oder Schwellenwerte lassen sich aufgrund der Transformationen nicht miteinander vergleichen.

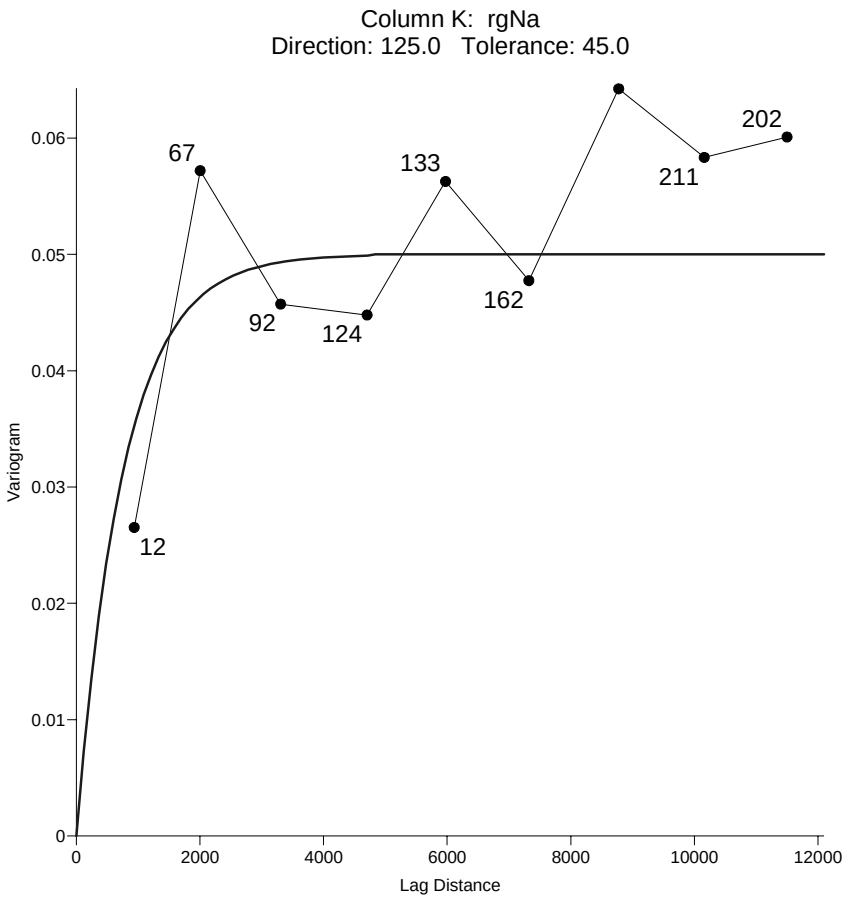

Abb. 11: Variogramm der Rangwerte

Wie in Kap. 2.1 erwähnt, kann die Auswahl nach dem besten Variogramm-Modell nach Auswertung durch Kreuzvalidation getroffen werden (s. Kap. 2.1).

Die Ergebnisse der Kreuzvalidation, sowie ergänzende Informationen zu den Variogrammen sind in der folgenden Tabelle enthalten (Tab. 3).
Für die nicht transformierten Ausgangsdaten wird ein optimaler Anisotropiefaktor von 2,0 ermittelt, für die transformierten Werte jeweils 2,2. Diese Werte sind dahin zu interpretieren, dass die generelle Grundwasserfließrichtung entscheidend auf die Ausbreitung (Diffusion) der Natrium-Ionen wirkt.
Dabei fallen die Varianzen hoch aus: Für die nicht transformierten Natriumwerte ergibt sich eine mittlere Varianz von 1215,43 $(mg/l)^2$, die einer mittleren Standardabweichung von 34,86 mg/l entspricht. Diese lässt sich durch Ausreißer erklären: Insbesondere, wenn innerhalb der Suchellipse nur wenige Messstellen liegen, erhöhen Ausreißer die Varianz deutlich.

Natrium (untransformiert)			
Modell:	sphärisch		
Grenzwert (sill):	540 (mg/l)²		
Reichweite:	3100 m		
Suchradius in m:	max: 4000	min: 2000	
Anisotropiefaktor:	1,8	**2,0**	2,2
Mittl. Varianz in (mg/l)²:	1239,08	1215,43	1234,82

Log-normal transformierte Natriumwerte			
Modell:	exponential		
Grenzwert (sill):	0,31		
Reichweite:	800 m		
Suchradius in m:	max: 4000	min: 2000	
Anisotropiefaktor:	2,0	**2,2**	2,4
Mittl. Varianz:	0,50227	0,50208	0,50213

Rangtransformierte Natriumwerte			
Modell:	exponential		
Grenzwert (sill):	0,05		
Reichweite:	800 m		
Suchradius in m:	max: 4000	min: 2000	
Anisotropiefaktor:	2,0	**2,2**	2,4
Mittl. Varianz:	0,06745	0,06744	0,06745

Tab.3 : Zusammenfassung der Variogramm-Modelle mit Angabe der durch Kreuzvalidation ermittelten Varianz.

3 Ergebnisse der Schätzung und Interpretation

3.1 Isolinienkarten der Natrium-Schätzwerte

Sind die Variogramme bestimmt, können die Werte für das gesamte untersuchte Gebiet mittels Kriging ausgerechnet und als Isolinienkarte dargestellt werden.
Zunächst wurden die (nicht transformierten) Messdaten direkt in die Kriging-Schätzung gegeben.

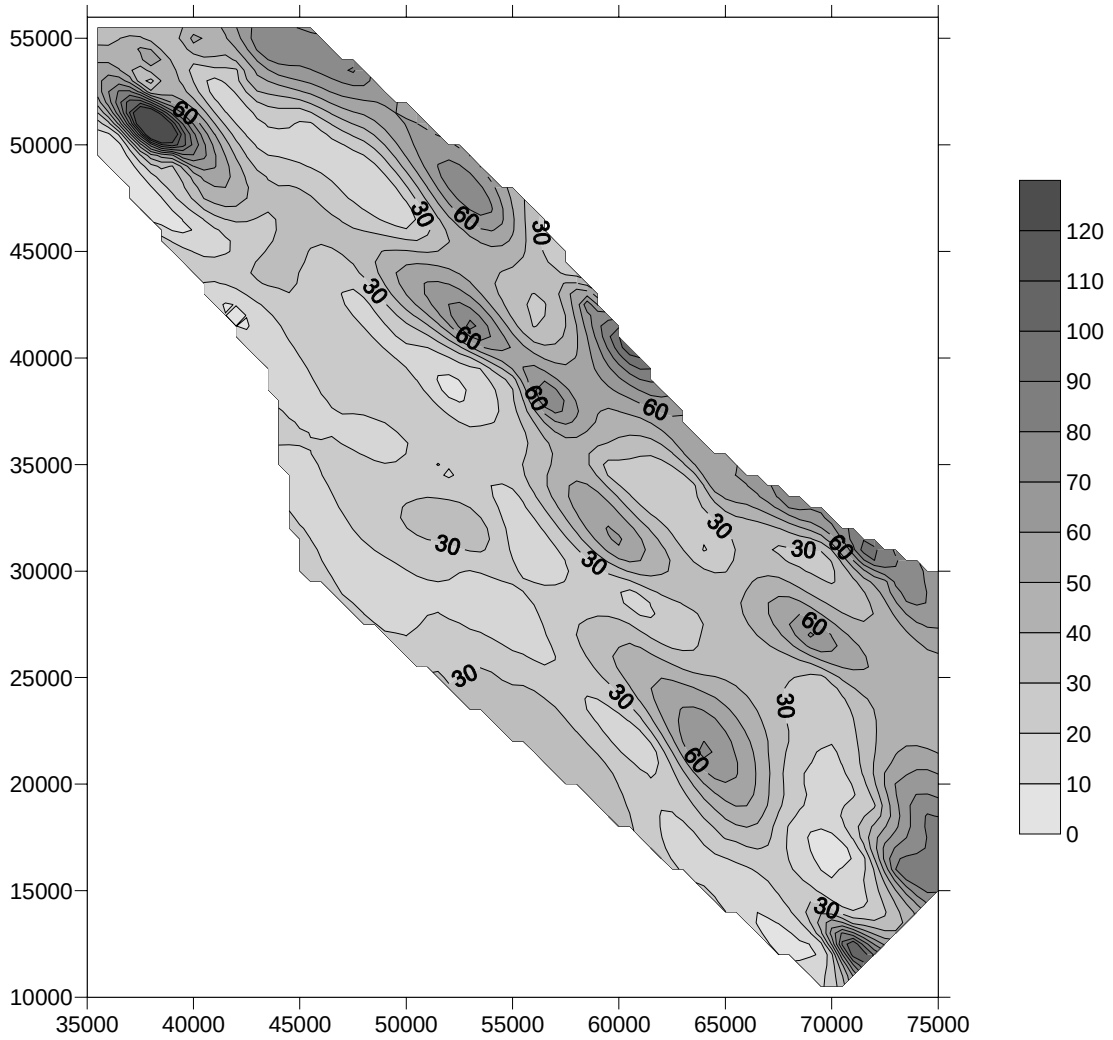

Abb. 12: Natrium-Gehalt in mg/l geschätzt nach den Ausgangsdaten (Anisotropie-Faktor 2.0)

Im Isolinienplan (Abb. 12) macht sich die modellierte Anisotropie entlang einer NW-SE-Achse in den länglichen Linsen bemerkbar. Erkennbar ist ein Trend zu höheren Werten mit Annäherung zur nordöstlichen Grenze, die dem Flusslauf entspricht. Im südöstlichen Teil der Karte ist zu berücksichtigen, dass der Oderlauf hier entlang einer Nord-Süd-Achse verläuft (außerhalb des Isolinienplans); auch hier sind die erhöhten Werte mit der Nähe zur Oder zu erklären. Die sehr hohen Schätzungen im äußersten Nordwesten[2] und im äußersten Süden der Karte sind jedoch durch einzelne, sehr hohe Ausgangswerte (Ausreißer, s. Kap.1) zu erklären und entsprechen nicht dem allgemeinen Trend. Der Wertebereich zwischen 10 und 60 mg/l beherrscht weite Teile des Geländes, hohe Werte (über 60 mg/l) sind auf kleinräumige Bereiche begrenzt.

Eine robustere Schätzung kann durch die Verwendung der logarithmierten Werte erreicht werden. Die rücktransformierten Werte sind in der folgenden Abbildung dargestellt (Abb. 12). Auffällig ist die Dominanz des Wertebereichs zwischen 10 und 40 mg/l und die geringe Ausprägung von hohen Werten.

[2] Der höchste Wert in Abb. 11 beträgt 182,2 mg/l. Er wird zu dem Bereich 'größer als 120 mg/l gezählt. Werte über 120 mg/l tauchen nur in der nordöstlichen lokalen Anomalie auf.

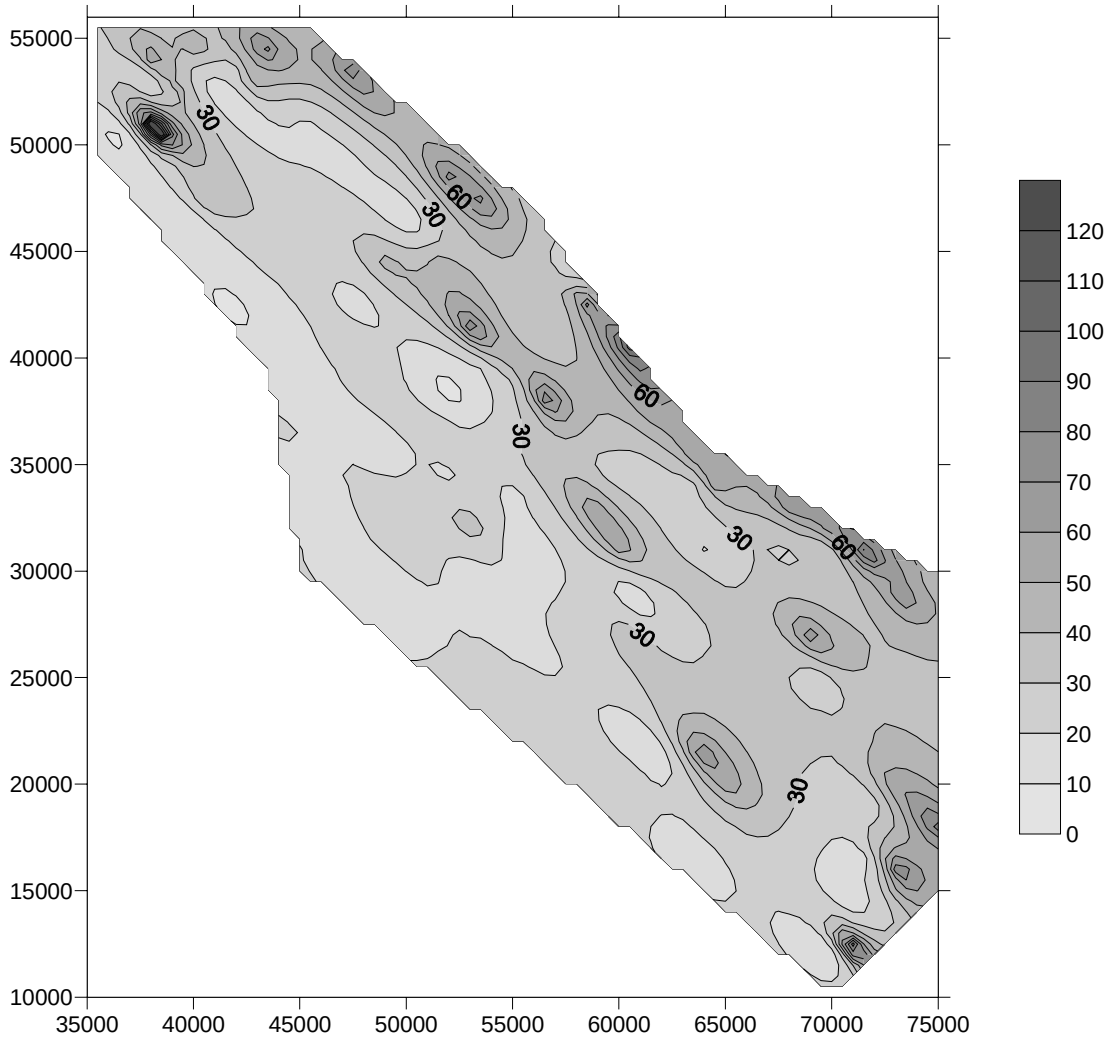

Abb.13: Natrium-Gehalt in mg/l geschätzt nach log-normaler Transformation und anschließender Rücktransformation (Anisotropiefaktor 2,2).

Konnte man in Abb. 12 eine grobe Unterteilung in einer flussfernen Region mit niedrigen Werten und einer flussnahen mit höheren Werten machen, zeigt Abb. 13 eine flächenmäßig beinahe konstante Natrium-Konzentration mit Werten unter 40 mg/l. Ein schmaler Streifen entlang der Oder, sowie einzelne 'Inseln' höherer Konzentration komplettieren das Bild. Auffällig ist die Tatsache, dass trotz kleinräumiger Ausbreitung, die Bereiche höherer Konzentration gleich hohe Werte erreichen, wie in Abb. 11 dargestellt[3]. Wäre die Darstellung in Abb. 13 dreidimensional, würde man kleine Hügel mit sehr steilen Flanken beobachten. Ebenfalls auffällig ist die Ähnlichkeit in beiden Abbildungen in der Ausbreitung der niedrigen Werte (unter 30 mg/l).

Ebenfalls robust ist die Schätzung nach Rangtransformation (siehe unten Abb. 14). Hier ist vor allem der Wertebereich zwischen 30 und 60 mg/l stark vertreten. Hohe Werte erfahren im Gegensatz zu den vorigen Abbildungen eine flächenmäßig stärkere Berücksichtigung, insgesamt lassen sich die abgebildeten Werte gleichmäßiger auf die einzelnen Wertebereiche verteilen**.** Im Unterschied zu den vorherigen Abbildungen sind extrem hohe Werte (über 90 mg/l) kaum vertreten, bzw. kommen nicht vor[4].

[3] In Abb. 13 beträgt der höchste geschätzte Wert 136,3 mg/l. Auch hier taucht er lediglich im nordöstlichen 'Fleck' auf.

[4] Der höchste Wert in Abb. 14 beträgt 101,3mg/l, die Einteilung der Farbskala wurde dennoch beibehalten, um eine Vergleichbarkeit nach Farbton zu gewährleisten.

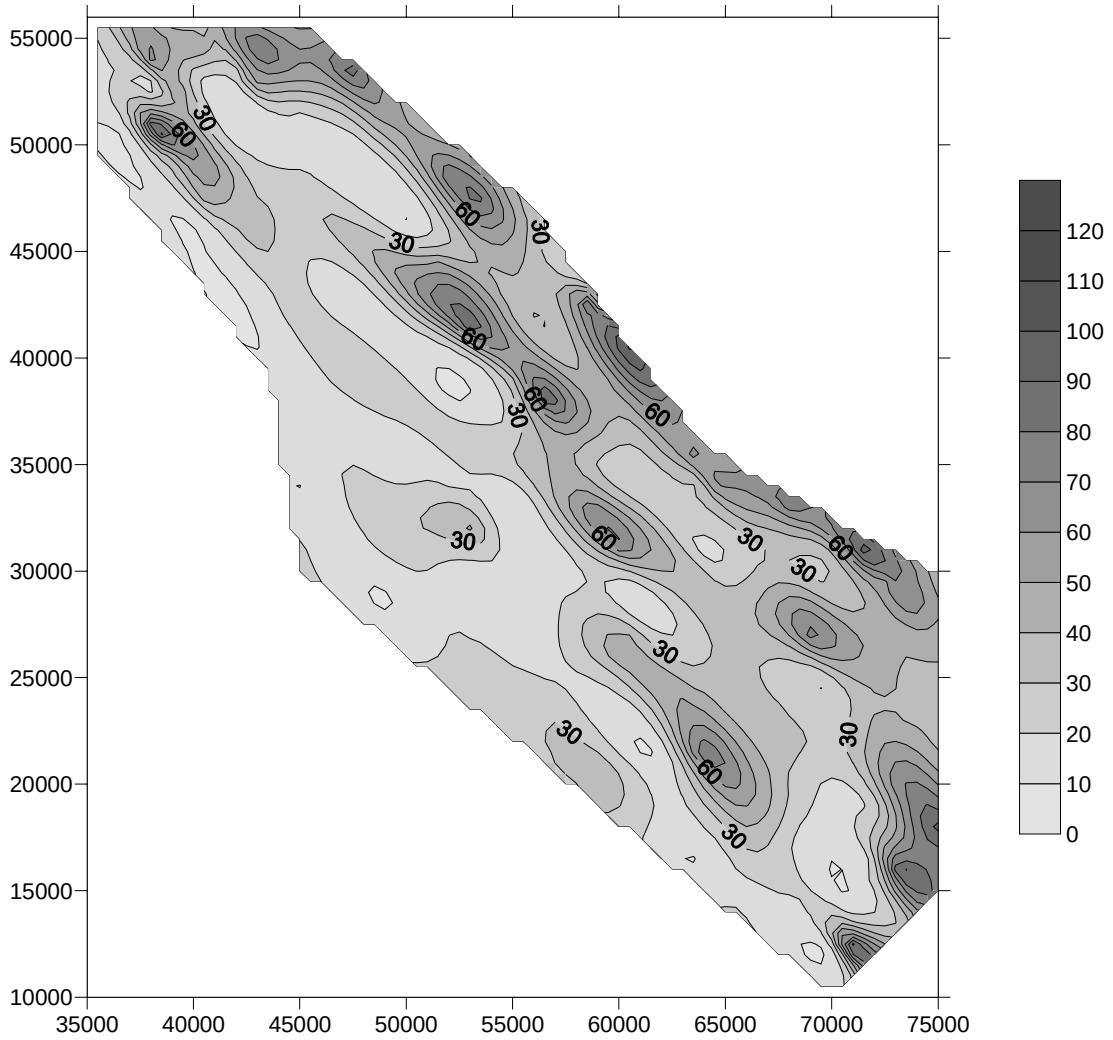

Abb. 14: Natrium-Gehalt in mg/l geschätzt nach Rangtransformation und anschließender Rücktransformation (Anisotropiefaktor 2,2).

Der Glättungseffekt der Rangtransformation wird in Kap. 3.4 bei den direkten Vergleichen deutlich.
Ergänzend stellt Tab. 4 eine Übersicht der einfachen deskriptiven Statistik aus den drei vorgestellten Schätzungen zusammen.

Geschätzte Größe	Transformation, Rücktransformation	Min	Max	Mittelwert	Stand.abw.
Natriumgehalt	keine	0,159	182,202	36,307	19,280
Natriumgehalt	Log-normal	5,916	136,320	31,059	13,939
Natriumgehalt	rang	8,550	101,273	32,611	18,044

Tab. 4: Deskriptive Statistik aus den Schätzungen des Natriumgehaltes ohne, mit log-normaler und mit Rang-Transformation.

Aus Tabelle 4 erkennt man, dass die Transformationen eine kleinere Streuung der Werte bewirkt haben. Deutlich zeigt sich der Einfluss der Transformationen auf den Mittelwert, der nach unten korrigiert wird. Ebenfalls große Unterschiede sind in der Schätzung von sehr hohen Werten (Ausreißer): Erwartungsgemäß fallen die Schätzungen nach Transformation robuster aus.

Die oben dargestellten Isolinienkarten sind als Übersicht sehr hilfreich, sie können jedoch aus zweierlei Hinsicht nicht den Anforderungen einer hydrochemischen Analyse genügen:

- Von Interesse ist meist die Konzentration ab einem Grenzwert, eine äquidistante Aufteilung ist vor allem in den niedrigen Bereichen nicht von Bedeutung;
- Die Schätzungen bedürfen eines Vertrauensintervalls, die sie in ihrer Aussage festigt.

Meist ist von Interesse, einen oberen Grenzwert anzugeben, der mit einer möglichst hohen Wahrscheinlichkeit (z.B. 0.95) nicht überschritten wird.

Im folgenden Kapitel werden die oberen 95%-Vertrauensgrenzen berechnet und als Isolinienkarten dargestellt. Die Berechnungen können nur für die Datensätze erfolgen, deren jeweilige Verteilung bekannt ist. In der vorliegenden Arbeit wurde diese Berechnung unter der Annahme der log-normalen Verteilung und für eine Gleichverteilung der rang-transformierten Daten vorgenommen.

3.2 Bestimmung der oberen Grenze nach log-normaler Transformation

Da y = ln(Z) normalverteilt ist, liegt der Erwartungswert von y mit einer Wahrscheinlichkeit von 0.90 im Intervall $y^* +/- 1{,}65 s_y^*$. Bei einseitiger Betrachtung liegt er mit einer Wahrscheinlichkeit von 0.95 unterhalb von $z^*+1{,}65 s_y^*$.
Nach der Rücktransformation ergibt sich die obere Schranke durch:

$$z_{95}^* = \exp(z^*+1{,}65 s_y^*)$$

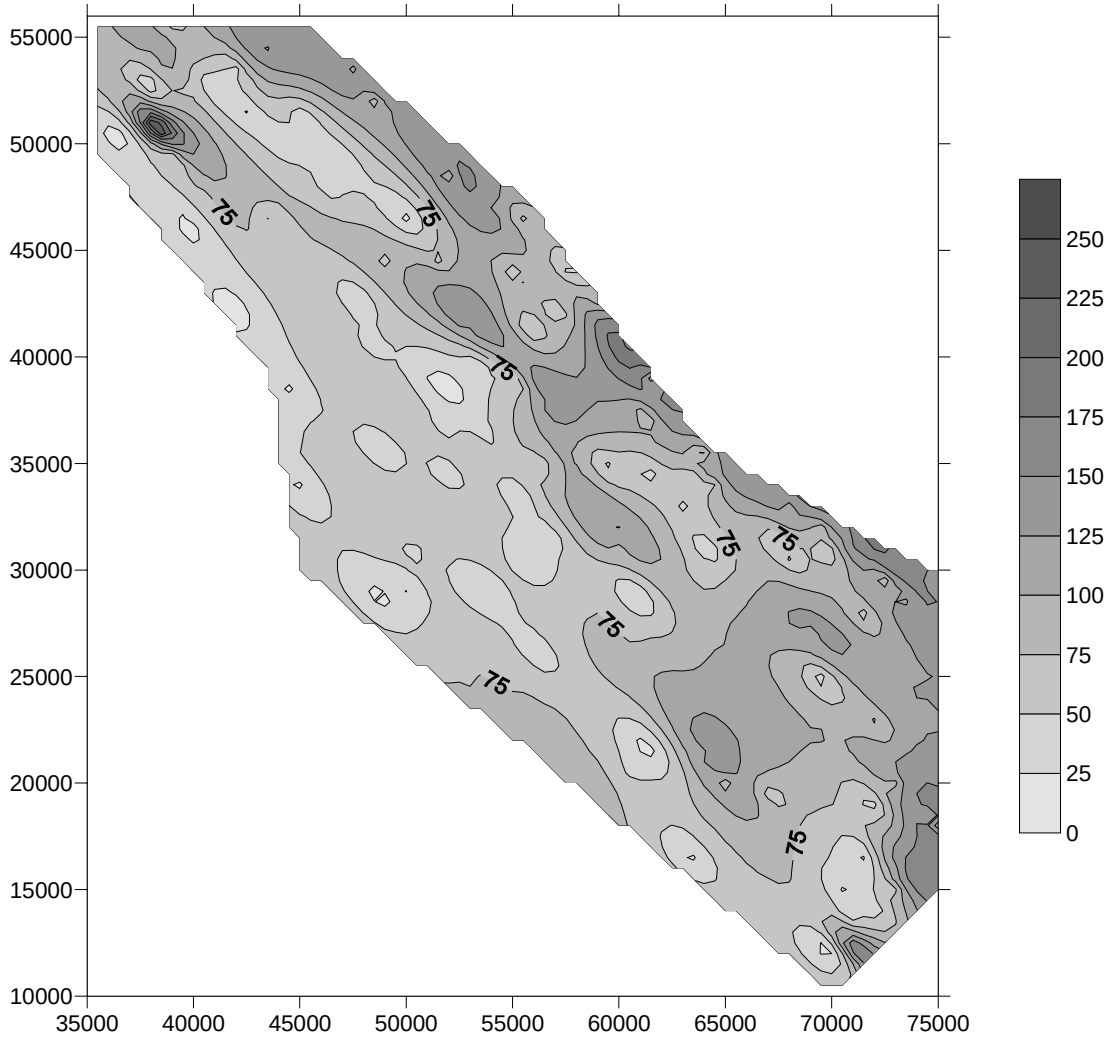

Abb. 15: Räumliche Verteilung der oberen 95%-Grenze nach log-normaler Transformation. Angaben in mg/l.

Ähnlich wie in Abb. 13 dominieren wenige Wertebereiche (hier: Werte zwischen 50 und 125 mg/l) große Teile der Karte. Höhere Werte beschränken sich meist auf die Randgebiete. Extreme Werte (über 200 mg/l) kommen nur im äußersten Nordwesten vor, auffällig ist hierbei die geringe Ausbreitung in der Fläche trotz der sehr hohen Werte.

3.3 Bestimmung der oberen Grenze nach Rangtransformation

Nimmt man an, dass im Intervall um die durch Kriging errechneten r*-Werte eine Gleichverteilung vorliegt, kann man folgende Annahmen treffen:

Sei $\mu_r = 0{,}5$ der Median einer zwischen 0 und 1 gleich verteilten Funktion der Dichte r, dann ist die Varianz von r definiert durch:

$$Var(r) = 1/12 = 1/N\ \Sigma(z_i - 0{,}5)^2$$

und $a = s_r^* \sqrt{12}$

Dann kann man die Dichtefunktion angeben:

$$f(x) = 1/a \text{ für } r^* - a/2 \leq x \leq r^* + a/2$$
$$f(x) = 0 \text{ sonst}$$

Daraus folgt für die obere 95%-Grenze: $x \geq r^* + 0{,}9 \cdot a/2$.
Die Ergebnisse zeigt Abb. 16:

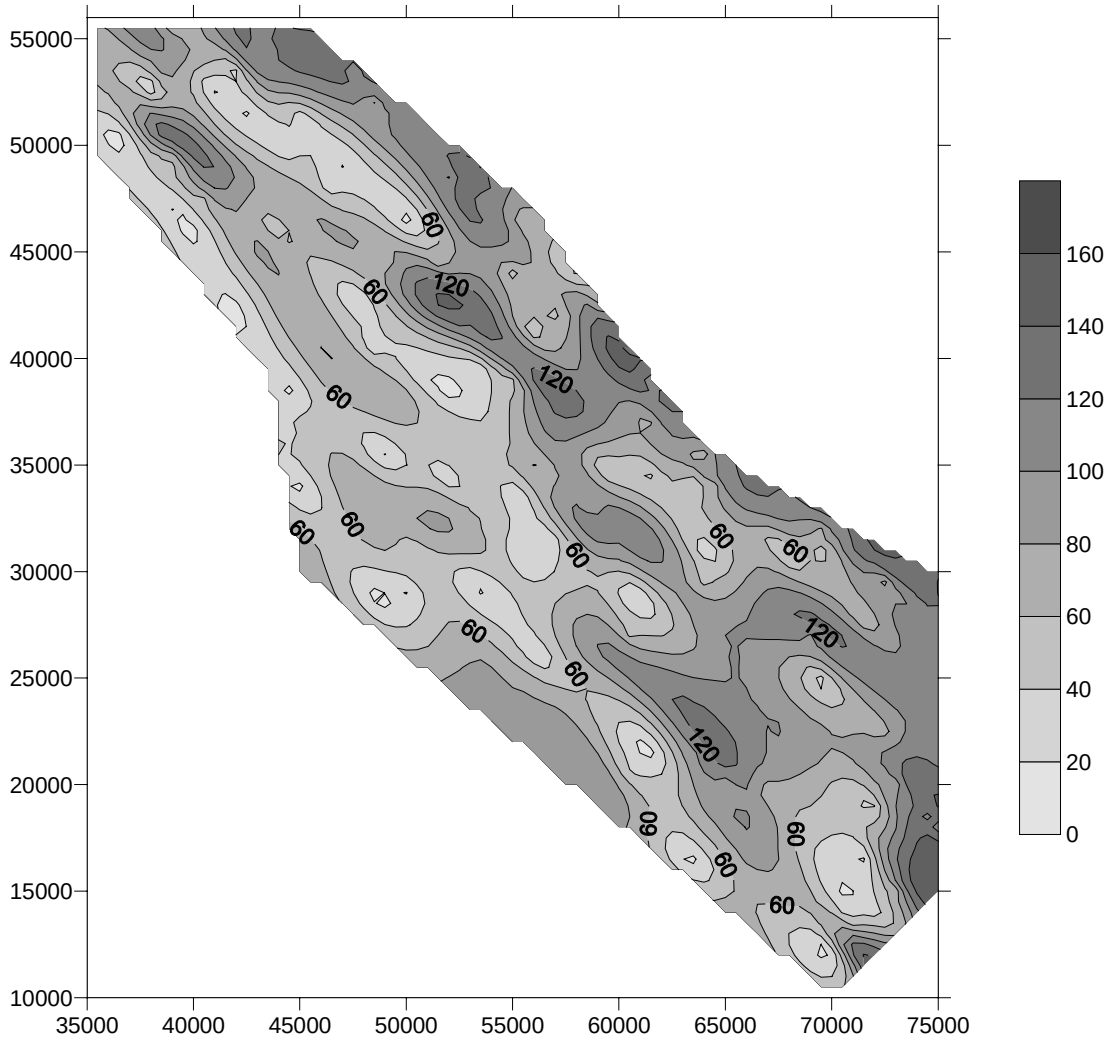

Abb. 16: Räumliche Verteilung der oberen 95%-Grenze nach Rangtransformation. Angaben in mg/l.

Obwohl die räumliche Verteilung der Minima und Maxima in beiden Abbildungen ähnlich ist, unterscheiden sich die Schätzwerte ganz erheblich.
Auch hier sind Ähnlichkeiten zur Isolinienkarte der Schätzwerte z* (Abb. 13). Auffällig ist neben der besseren Verteilung auf die Wertebereiche das Fehlen von extrem hohen Werten (über 200 mg/l).

3.4 Vergleich der Schätzungen

Da die bisher dargestellten Isolinienpläne für jeden beliebigen Punkt innerhalb des untersuchten Gebietes einen Schätzwert angeben, ist es möglich, Isolinienpläne übereinander zu legen und Differenzen zu bilden und somit einen direkten Vergleich zu erzeugen.
Bevor die Schätzergebnisse miteinander verglichen werden können, wird ein Korrekturfaktor eingefügt (siehe auch: JOURNEL et al., 1978). Dieser ergibt sich aus der Differenz der jeweiligen Mittelwerte:

$$\varepsilon_{ij} = \overline{x}_i - \overline{x}_j$$

Die Daten werden derart verändert, dass die untersuchten Datensätze den gleichen Mittelwert haben.
Beispiel: Bei der Berechnung des Korrekturfaktors für den Vergleich der Schätzwerte, die auf Basis der Natrium-Ausgangswerte berechnet wurden, mit den Schätzwerten, die eine log-normale Transformation erfahren haben, ergibt sich eine Differenz der Mittelwerte von 5,2384 mg/l:

$$\varepsilon_{Na,\ln Na} = \overline{x}_{Na} - \overline{x}_{\ln Na} = 36{,}2970 - 31{,}0586 = 5{,}2384 mg/l$$

Anschließend werden die Werte mit dem kleineren Mittelwert um diesen Differenzbetrag durch einfache Addition erhöht.
Die Korrektur wird für alle folgenden Isolinien-Vergleiche vorgenommen.

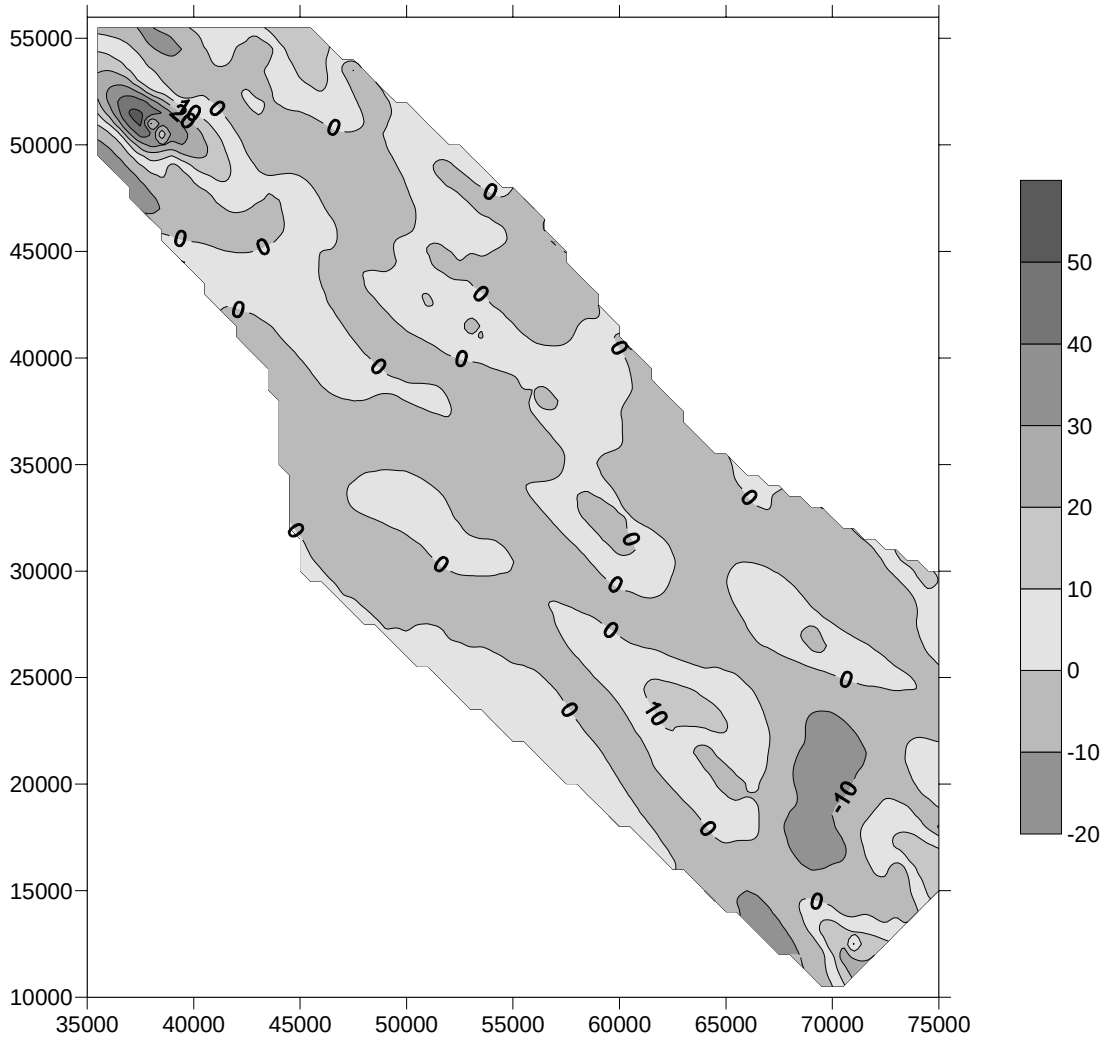

Abb. 17: Differenz der Schätzwerte: Berechnung auf Basis der Originaldaten abzgl. Werte aus der Berechnung auf Basis log-transformierter Daten. Angaben in mg/l.

Abb. 17 zeigt den direkten Vergleich der Schätzwerte aus Abb. 12 (Schätzung durch Verwendung der nicht transformierten Ausgangsdaten) und der Werte aus Abb. 13 (Schätzung nach log-normaler Transformation und anschließende Rücktransformation). Positive Werte geben für die betrachtete Stelle eine höhere Schätzung durch Verwendung der Originaldaten wider, während negative Werte eine höhere Schätzung durch Verwendung der Daten nach log-normaler Transformation anzeigen.

Zunächst fällt auf, dass die positiven Abweichungen höhere Werte erreichen. Bei genauerer Betrachtung erkennt man jedoch, dass mit Ausnahme der nordwestlichen Anomalie die positiven Differenzbeträge kaum 20 mg/l übersteigen.
Innerhalb der auffälligen nordwestlichen Anomalie sind bei näherer Betrachtung zwei helle Bereiche erkennbar: Diese sind als Maxima der Schätzung mit den logarithmierten Daten zu interpretieren und da diese Maxima sehr steile Flanken aufweisen (vgl. Abb. 13) fallen die Schätzungen mit den Originaldaten in unmittelbarer Umgebung höher aus.
Ferner erkennt man, dass negative Differenzen stets in der Nähe von Bereichen positiver Abweichungen vorkommen. Die Schätzungen mit den Originaldaten zeigen eine höhere ’Trägheit’, indem Maxima eine größere Umgebung beeinflussen: Sie bewirken hohe Schätzungen in der unmittelbaren Umgebung und niedrige Schätzung in der weiteren Umgebung.

Für die übrigen Bereiche des untersuchten Gebietes (vor allem im zentralen Bereich) sind keine nennenswerten Unterschiede festzustellen. Deutliche Unterschiede treten dort auf, wo eine starke Datenvariabilität zu beobachten ist.

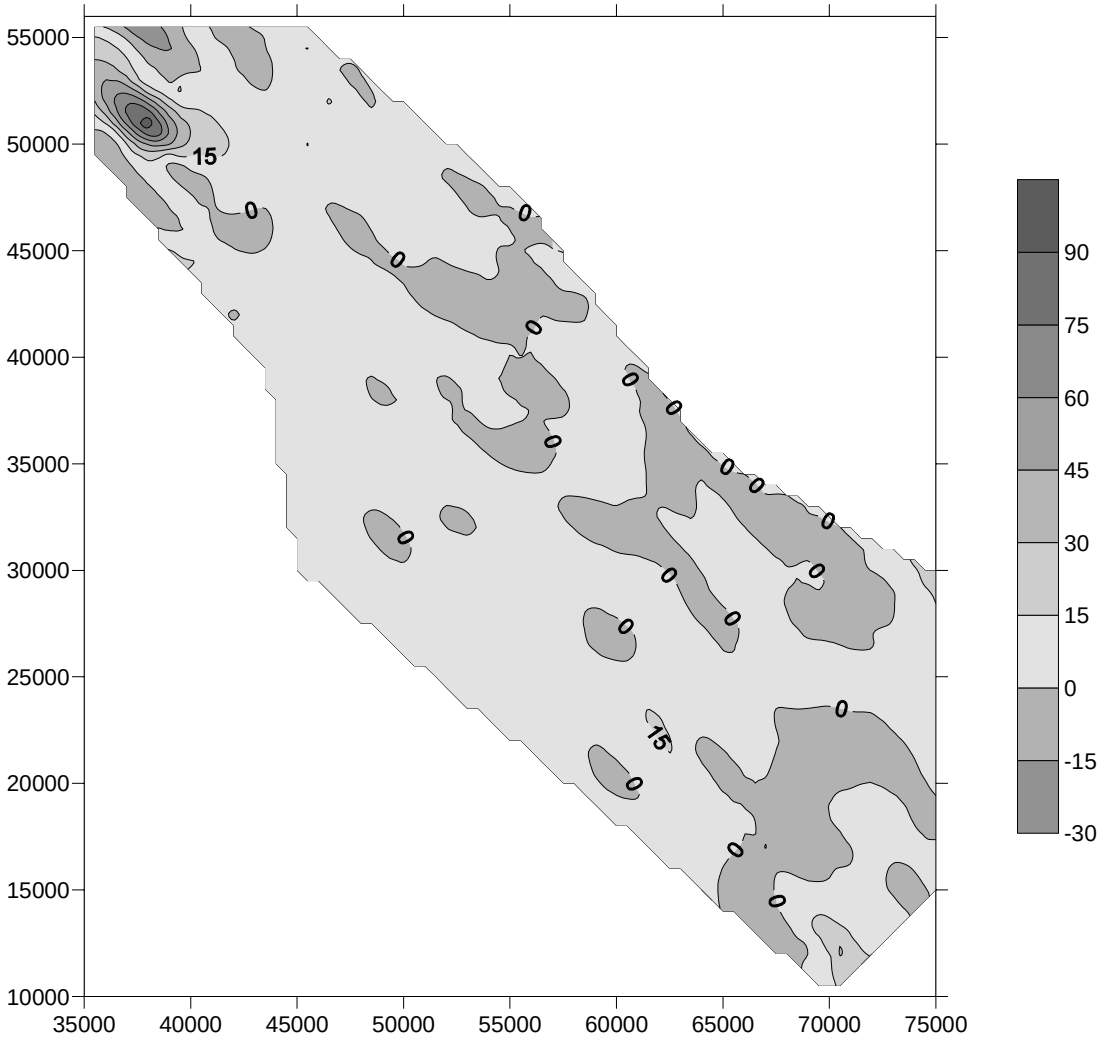

Abb. 18: Differenz der Schätzung auf Basis der Originaldaten abzgl. der Schätzwerte nach Rangtransformation. Angaben in mg/l.

Der Differenzenberechnung aus Abb. 18 ist ebenfalls eine Korrektur nach Vergleich der Mittelwerte vorausgegangen. Zu erkennen ist ein Übergewicht positiver Differenzen. Für weite Bereiche des untersuchten Gebiets fallen diese nicht größer als 15 mg/l aus (gelb gekennzeichnete Fläche); diese Bereiche entsprechen den Flächen niedriger Natriumwerte. Eine mögliche Erklärung liegt in dem Einfluss einzelner, hoher Messwerte auf die Schätzung mit den nicht transformierten Originaldaten. Das könnte bedeuten, dass eine Überschätzung der Natriumwerte durch Berechnung mit nicht transformierten Daten für die dem Flusslauf fernen Bereiche vorliegt.
Anders verhält es sich mit den Bereichen mittlerer Schätzung. Hier liegt eine allgemein höhere Schätzung nach Berechnung mit den rangtransformierten Daten vor (siehe grün gekennzeichnete Flächen in Abb. 18). Da hohe Ausgangswerte in der Häufigkeitsverteilung schwach vertreten sind (s. Abb. 2), erfahren mittlere Werte in der Rangdarstellung eine Aufwertung.

Im Bereich der nordwestlichen Anomalie ist die glättende Wirkung der Rangtransformation erkennbar: Extrema werden deutlich relativiert, indem sie stark den Werten ihrer Umgebung angepasst werden.

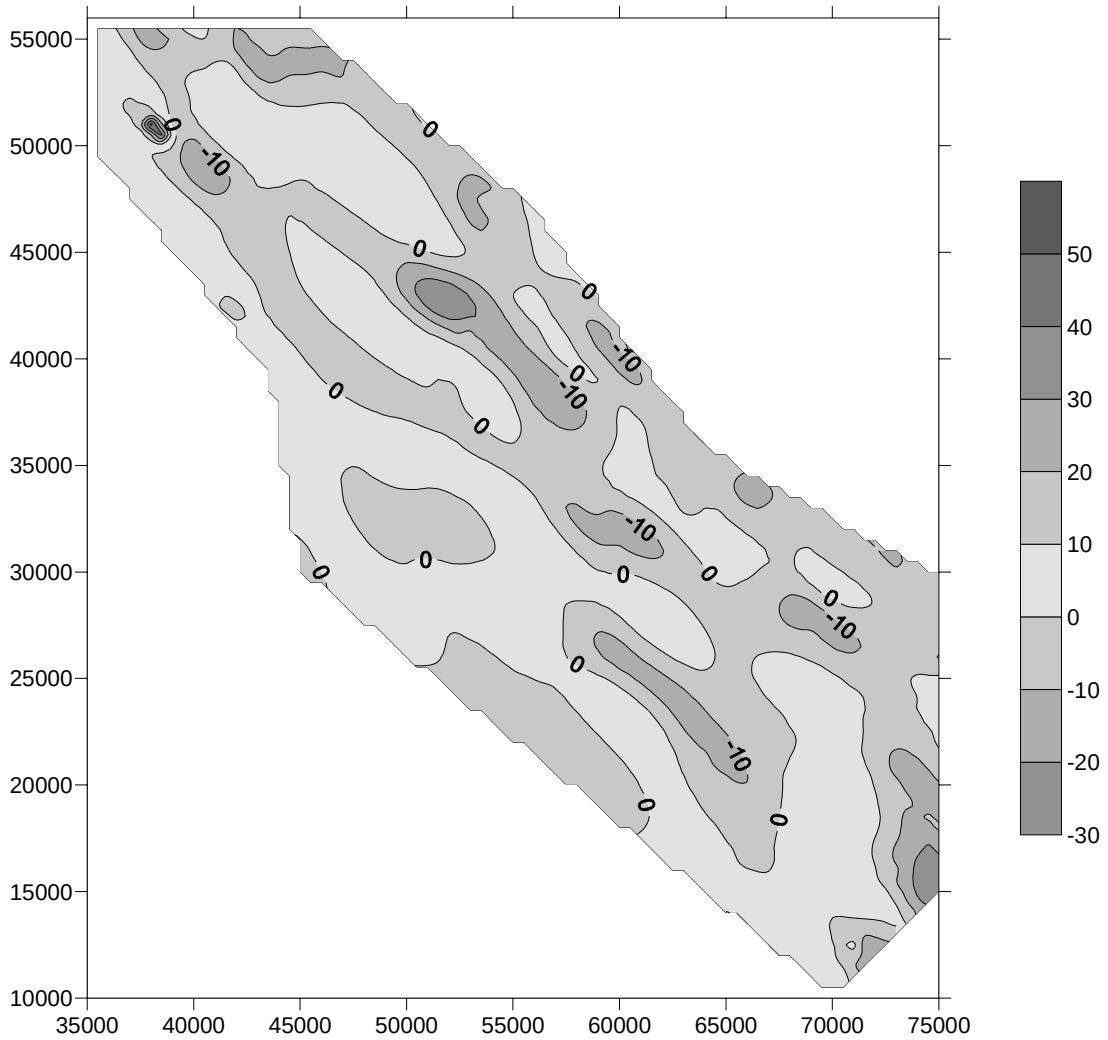

Abb. 19: Differenz der Schätzwerte nach log-normaler Transformation abzgl. der Schätzwerte nach Rangtransformation. Angaben in mg/l.

Der direkte Vergleich der Schätzungen, die mit transformierten Daten vorgenommen worden sind, ist in Abb. 19 dargestellt. Die Differenzen ergeben sich durch den Betrag aus der Schätzung mit log-normal transformierten Daten abzüglich des Betrages aus der Schätzung mit den rangtransformierten Daten.
Von der nordwestlichen Anomalie mit extrem hoher Schätzung durch die log-normal transformierten Daten abgesehen, ist ein flächenmäßig leichtes Übergewicht der Schätzungen nach Rangtransformation zu erkennen. Es betrifft die Bereiche, die Ausgangswerte mittlerer Größe enthalten. Wie bereits in der Beschreibung von Abb. 18 erwähnt, werden mittlere Werte durch die Rangtransformation aufgewertet.
Anhand einer Anomalie an der südlichen 'Spitze' des untersuchten Geländes lässt sich die unterschiedliche Schätzung ebenfalls gut erkennen: In diesem Bereich befindet sich eine Messstelle mit einem extrem hohen Ausgangswert (123,6 mg/l). Bei näherem Hinsehen in Abb. 18 ist ein sehr kleiner Kringel (kaum größer als ein Punkt)zu sehen, der dem Standort

der Messstelle entspricht und mit gelber Farbe gekennzeichnet ist: An dieser Stelle ist die Schätzung nach log-normaler Transformation erwartungsgemäß sehr hoch ausgefallen. Für die unmittelbare Umgebung, östlich und südöstlich von dem Kringel, für die keine Messwerte vorliegen, hat die Schätzung nach Rangtransformation höhere Werte berechnet.

Mittels Kreuzvalidation werden die Ergebnisse der beiden Schätzungen nach Transformation miteinander verglichen und die Rangtransformation liefert das bessere Ergebnis: Die mittlere Abweichung von den Ausgangsdaten beträgt 14,84 mg/l (19,86 mg/l bei der log-normalen Transformation).

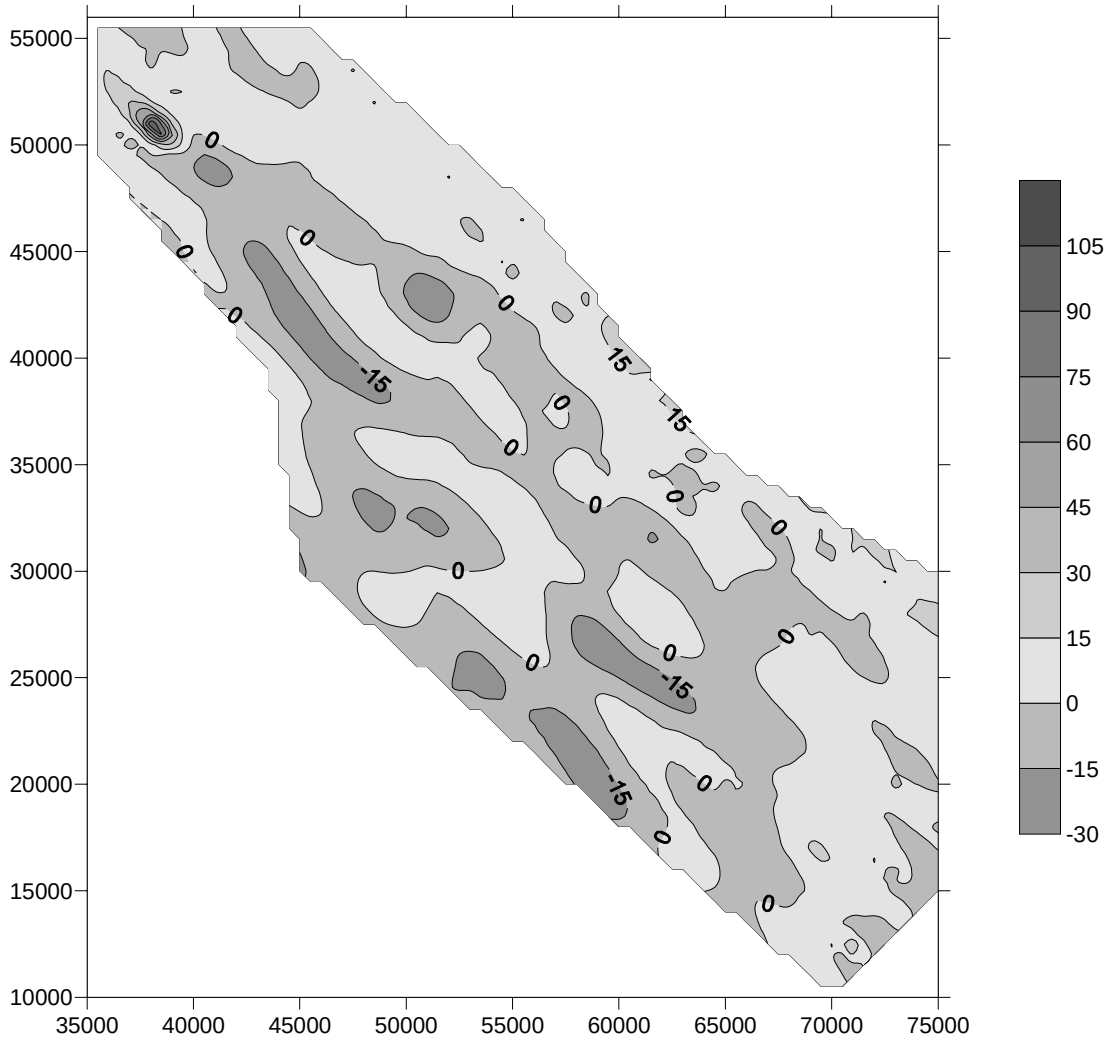

Abb. 20: Differenz in der Schätzung der oberen Grenze: Schätzung mit log-normal transformierten Daten abzgl. Schätzung mit rangtransformierten Daten. Angaben in mg/l.

Schließlich wird ein Vergleich der oberen Grenzen (auf dem Niveau einer 95%-iger Wahrscheinlichkeit) durchgeführt.
Abbildung 20 enthält die Differenzen aus der Schätzung der oberen Grenze nach Berechnung mit den log-normal transformierten Daten abzüglich der Schätzung der oberen Grenze, die auf Basis der rangtransformierten Daten vorgenommen worden sind.
Für die Bereiche, die eine Differenz nicht größer als 15 mg/l im Betrag haben ist ein flächenmäßig ausgeglichenes Verhältnis zu beobachten. Der Verlauf der 0-Isolinie belegt es. In der Anomalie im Nordwesten zeigt sich die beinahe punktuelle (im Sinne von kaum Fläche einnehmende) Schätzung nach log-normaler Transformation. Ebenfalls kleine Flächen mit

einer höheren Schätzung mit den logarithmierten Werten sind entlang des Flusslaufes zu sehen.
Auffälliger sind jene Bereiche, in denen die Schätzung nach Rangtransformation höhere Werte liefert. In Abb. 20 ist lediglich zu erkennen, dass es sich hierbei um dem Flusslauf ferne Teilgebiete handelt. Eine Erweiterung von Abb. 20 ist in Abb. 21 enthalten: Hier ist der Standort der untersuchten Messstellen dem Isolinienplan von Abb. 20 hinzugefügt worden. Wie man sieht, stimmen die oben beschriebenen Gebiete mit einer höheren Schätzung nach Rangtransformation mit den größeren 'Lücken' im Messstellennetz überein.

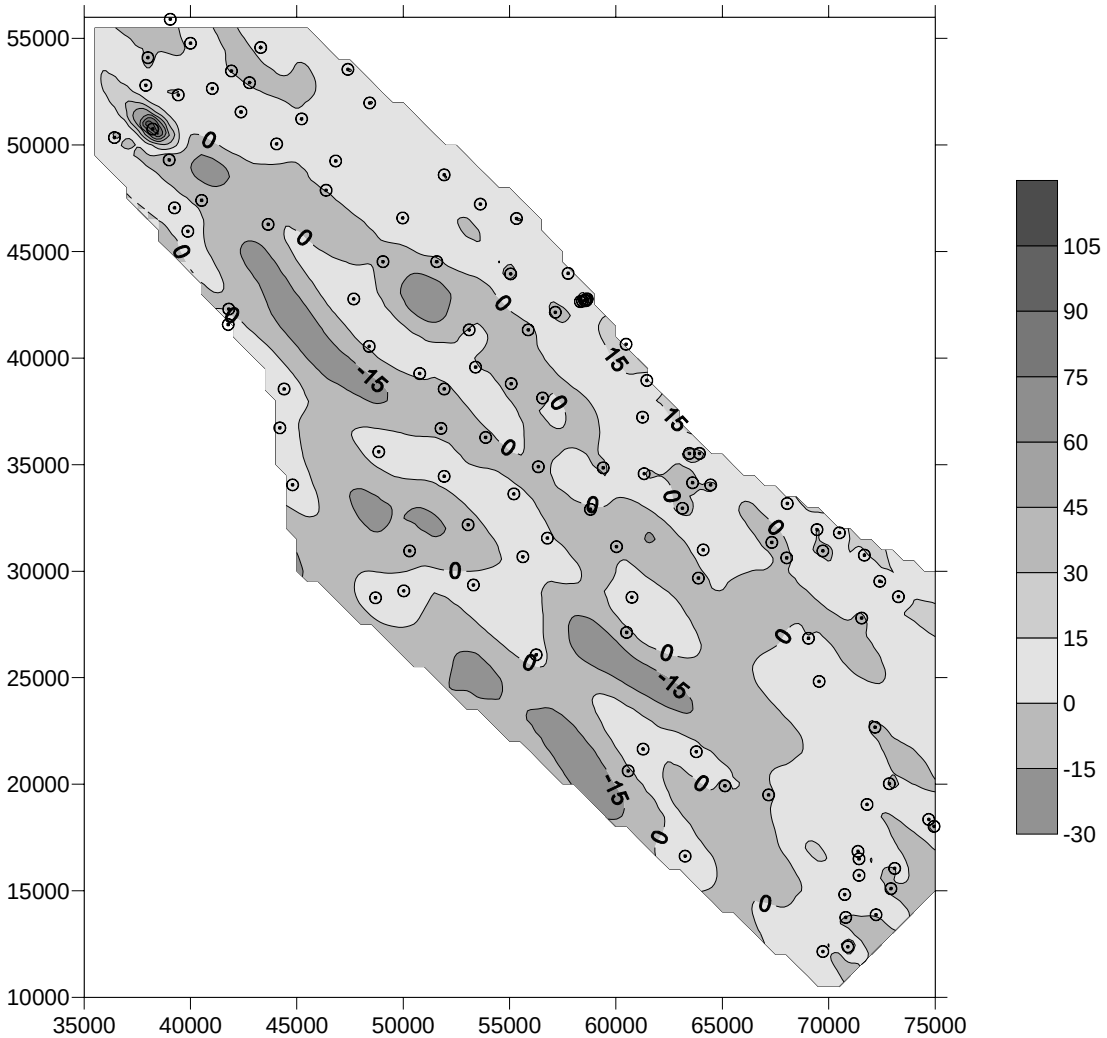

Abb. 21: Erweiterung von Abb. 20 mit Standort der untersuchten Messstellen.

Dies ist ein weiterer Beleg dafür, dass die Berechnung nach Rangtransformation eine flächenhaft gleichmäßige Schätzung bewirkt. Log-normal transformierte Werte lassen hingegen größere Unterschiede innerhalb einer kurzen Distanz zu.

Anschließend bleibt zu untersuchen, inwieweit sich die betragsmäßig hohe Differenzen quantifizieren lassen.
Abb. 22 zeigt die Häufigkeitsverteilung der Differenzen der Obere-Grenze-Berechnungen. Grundlage hierzu sind Schätzwerte an 3350 Punkten innerhalb des Isolinienplans (Punktegitter). Man erkennt, dass nur wenige Werte betragsmäßig größer als 40 mg/l ausfallen. Diese Erkenntnis relativiert den ersten Eindruck einer deutlich schiefen Verteilung aus dem Isolinienplan.

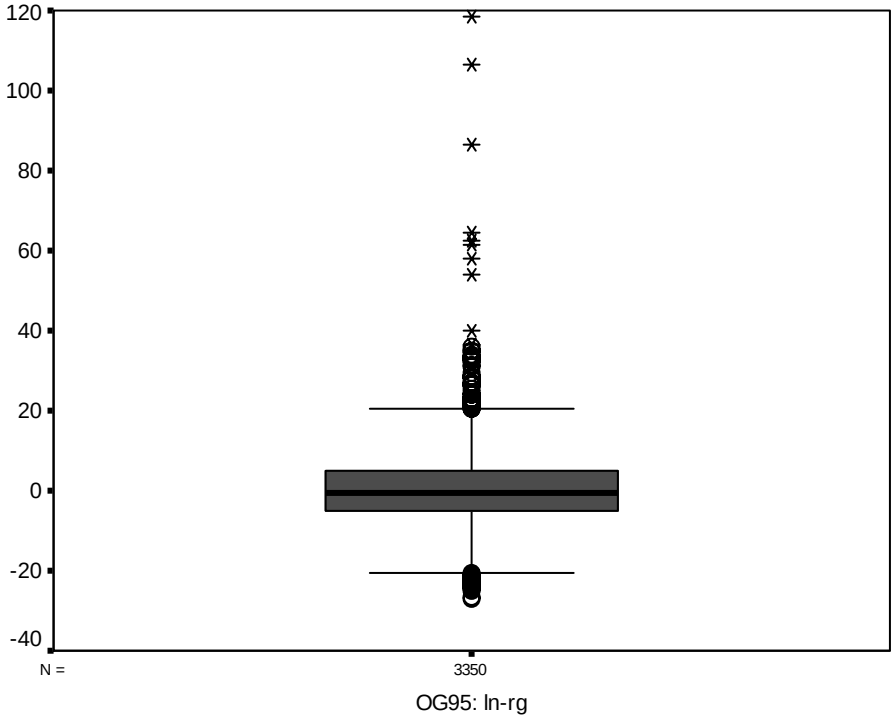

Abb. 22: Boxplot der Differenzen der oberen Grenzen. Angaben in mg/l.

4. Zusammenfassung

Die Häufigkeitsverteilung der untersuchten Natrium-Gehalte macht eine Umwandlung der Ausgangswerte hinsichtlich der Angabe von Erwartungswert und Konfidenzintervall erforderlich. Um robuste Schätzungen vornehmen zu können, eignen sich log-normale, sowie Rang-Transformation der Daten. Diese werden durchgeführt.
Für die Berechnung der Schätzwerte im zweidimensionalen Raum wird das Kriging-Verfahren angewandt, das die räumliche Ausdehnung der untersuchten Variablen und den räumlichen Zusammenhang der Messwerte untereinander berücksichtigt.
Die Auswahl des besten Variogramm-Modells erfolgt nach Überprüfung mehrerer Faktoren: Anisotropie des Mediums mit bevorzugter Ausbreitungsrichtung und Anisotropiestärke (Faktor), Größe des Suchradius, Einteilung der Abstandsklassen, Anpassung des errechneten Variogramms durch ein Modell. Somit sind gleichzeitig mehrere Fehlerquellen gegeben.
Die vorgestellten Variogramme zeigen, dass die gegenseitige räumliche Beeinflussung der Messwerte nur innerhalb eines Radius von maximal 4000 m in Grundwasserfließrichtung (und maximal 2000 m in der dazu senkrechten Richtung) stattfindet.
Die Kriging-Schätzungen werden sowohl mit den Ausgangswerten, als auch mit den transformierten Werten vorgenommen. Für die Isolinienpläne werden die umgewandelten Daten anschließend rücktransformiert.
Anhand der Transformationen werden auch obere Grenzen für das 95%ige Konfidenzintervall berechnet. Auch diese Werte werden anschließend rücktransformiert.

Im direkten Vergleich der Schätzungen sind für weite Bereiche des untersuchten Gebiets geringe Abweichungen festzustellen. Große Unterschiede treten dort auf, wo Ausreißer zu finden sind.
Ferner ist zu beobachten, dass eine Kriging-Schätzung mit log-normal transformierten Daten an Stellen, die Ausreißer enthalten, lokal sehr begrenzte Anomalien hervorrufen, ohne die weitere Umgebung zu beeinflussen. Eine Rang-Transformation führt hingegen bei der

Schätzung dazu, dass Anomalien durch eine Anpassung der Umgebung (für die keine Messwerte vorliegen) betragsmäßig kleiner ausfallen, jedoch in der Fläche eine weitere Ausdehnung erfahren.
Ebenfalls zu beobachten ist die Tatsache, dass Schätzungen mit log-normal transformierten Daten größere Lücken im Messstellennetz mit niedrigeren Werten versehen, als es im Fall von rangtransformierten Daten geschieht.

5 Literatur

AKIN, H. & SIEMES, H.(1988): Praktische Geostatistik – Eine Einführung für den Bergbau und die Geowissenschaften. Springer Verlag, Berlin, 304 S.;

BALAKRISHNAN, N. & RAO, C.R.(1998): Order Statistics: Theory and Methods. Handbook of Statistics 16, Elsevier, Amsterdam, 731 S.;

BÜNING, H. & TRENKLER, G.(1994): Nichtparametrische statistische Methoden. de Gruyter, Berlin, 485 S.;

BURGER, H., PAWLOWSKY-GLAHN, V., SCHAFMEISTER, M.-Th. (2001): Grundbegriffe der Geostatistik – Eine Einführung in die geostatistische Modellierung und räumliche Interpolation im Rahmen des Weiterbildenden Studiums Mathematische Methoden und Modelle in den Geowissenschaften, 65 S. [Unveröff.]

DAVID, H.A. (1981): Order Statistics. Wiley, New York, 358 S.

GRUBBS, F.E. & BECK, G. (1972): Extension of sample sizes and percentage points for significance tests of outlying observations, Technometrics, vol. 14, S. 847-854;

ISAAKS, E.H. & SRIVASTAVA, R.M.(1989): Applied Geostatistics. Oxford University Press, New York, 561 S.;

JOURNEL, A.G., HUIJBREGTS, C.J. (1978): Mining Geostatistics. Academic Press, London

LIEDHOLZ, T. (2001): Die hydrochemischen Verhältnisse im oberflächennahen Grundwasser des Oderbruchs. Berliner Geowissenschaftliche Abhandlungen, Reihe A, Bd. 209, 131 S.;

LIMPERT, E., STAHEL, W. A., ABBT, M.(2001): Log-normal distributions across the sciences: Keys and Clues. BioScience, vol. 51 No. 5, S. 341-352;

SACHS, L. (1999): Angewandte Statistik, Berlin, Springer, 881 S.;

WACKERNAGEL, H. (1995): Multivariate Geostatistics – An Introduction with Applications. Springer, Berlin, 387 S.;

6 Anhang

Datengrundlage mit Kennzeichnung der Messstelle, Hoch- und Rechtswert, Natrium-Gehalt in mg/l:

ID	x	y	Na	ID	x	y	Na
555	69550	24825	22,3	2334	61450	38950	80,9
952	40525	47400	39,9	2335	60475	40650	111,7
955	44050	50050	17,5	2336	56550	38125	87,8
959	49050	44525	45,6	2337	55075	38800	29,6
960	55050	43950	38,6	2339	61325	34575	23,9
961	44800	34050	11,8	2340	56775	31550	20,2
962	53400	39575	19,4	2341	64450	34050	57,9
966	61250	37225	51,9	2342	55200	33625	18
970	55625	30675	12,3	2344	63875	29675	34
2024	41025	52650	13,6	2347	63600	34150	27,8
2025	39425	52350	45,3	2380	67325	31350	29,3
2026	40000	54775	56,6	2382	68050	33175	76,9
2027	42775	52925	47,9	2383	69450	31950	57,2
2038	39050	55900	26,7	2384	71675	30750	99,4
2039	38000	54100	62,4	2389	68025	30625	28,5
2040	43300	54575	79,8	2390	69725	30950	29,4
2059	37900	52800	16,6	2391	70500	31800	74,4
2061	47400	53550	83,4	2420	48700	28750	9,6
2064	48425	51975	58,7	2423	50025	29075	16,6
2069	41925	53475	23,9	2427	53300	29350	12,2
2100	36425	50350	7,2	2460	56250	26075	17
2104	43650	46275	28,8	2461	65125	19925	58,8
2105	42375	51550	14,9	2464	60500	27125	36,8
2128	38225	50750	206,1	2471	61275	21650	8,8
2129	39000	49300	30,2	2472	60750	28775	12,5
2137	39250	47050	15	2486	60575	20625	25
2140	51575	44525	46,8	2490	63775	21525	74,8
2143	46375	47875	23,2	2504	69050	26850	73,2
2144	45225	51225	18,9	2506	72175	22675	40,9
2145	49975	46575	11,9	2509	72850	20025	58,9
2148	47675	42775	16	2513	73275	28800	72,9
2149	53100	41325	86,7	2514	71550	27800	38,8
2150	46825	49250	14,8	2515	72400	29525	53,2
2154	51925	48600	76,2	2520	67175	19500	37,4
2157	53625	47225	77,6	2521	71800	19050	22,5
2180	55875	41325	29	2587	63250	16625	11,2
2181	57150	42150	33,5	2622	74700	18350	83,4
2182	57750	43975	22,9	2623	73100	16050	87
2183	55325	46550	44,3	2624	71425	16500	14,5
2271	53050	32175	40,2	2625	71425	15725	19
2272	51925	38550	5,5	2626	70800	13750	24,4
2273	48850	35600	15,7	2628	69725	12150	9,4
2275	48400	40550	22,5	2631	70900	12375	123,6
2278	44200	36725	21,8	2632	72225	13875	23,7
2279	50775	39275	16,7	2633	72925	15100	49,4
2280	53875	36275	26,7	2634	74950	18025	87
2281	51925	34450	18,3	2636	71375	16850	18,8
2283	50300	30950	25,8	2638	70750	14825	11,6
2287	51775	36700	25,4	7503	39875	45950	9,3

2291	44400	38550	13,3	7509	41800	42300	6,5
2321	59400	34850	25,3	7552	41775	41575	10,8
2324	56350	34900	35,7	9241	58650	42775	43,9
2325	58800	32900	61,4	9244	58625	42775	79,9
2326	60025	31150	62,6	9245	58600	42700	72,9
2329	64100	31000	18,3	9247	58575	42700	70,7
2330	63125	32950	28,7	9248	58525	42700	89,2
2332	63925	35525	53,3	9249	58425	42675	94,7
2333	63450	35525	59,2	9250	58325	42650	60,8