



Master Thesis

im Rahmen des
Universitätslehrganges „Geographical Information Science & Systems“
(UNIGIS MSc) am Zentrum für GeoInformatik (Z_GIS)
der Paris Lodron-Universität Salzburg

zum Thema

„Digitale Kartierung von Bodenschadstoffdaten“ Möglichkeiten und Grenzen im Schweizer Mittelland

vorgelegt von

Dipl. Ing. (FH) Kirsten Rehbein
u1123, UNIGIS MSc Jahrgang 2004

Zur Erlangung des Grades
„Master of Science (Geographical Information Science & Systems) – MSc(GIS)“

Gutachter:
Ao. Univ. Prof. Dr. Josef Strobl

Zürich, April 2006

Danksagung

Ich möchte mich beim UNIGIS-Lehrgangsteam, insbesondere bei Herrn Mag. Jürgen Praher und Herrn Mag. Michael Fally, bedanken, die jederzeit im Laufe des Studiums für Fragen zur Verfügung standen. Herrn Prof. Josef Strobl danke ich für seine hilfreichen Anregungen in Zusammenhang mit dieser Master Thesis.

Darüber hinaus möchte ich mich bei meinem Arbeitgeber, der Agroscope FAL Reckenholz, und insbesondere meinem Vorgesetzten Herrn Dr. André Desaulles bedanken, der mich sowohl finanziell als auch fachlich von Anfang an in meinem Vorhaben bezüglich des UNIGIS-Studiums unterstützt hat.

Ein ganz besonderer Dank geht an meinen Kollegen Herrn Dr. Armin Keller, der mir immer wieder ein wertvoller Diskussionspartner war und diese Arbeit Korrektur gelesen hat.

Ebenfalls danke ich Herrn Dr. Andreas Papritz von der ETH Zürich, dass er sich die Zeit genommen hat, mir bei geostatistischen Fragestellungen weiterzuhelfen.

Erklärung der eigenständigen Abfassung der Arbeit

Ich versichere, diese Master Thesis eigenständig angefertigt zu haben. Alle Ausführungen der Arbeit, die wörtlich oder sinngemäss übernommen wurden, sind entsprechend gekennzeichnet.

Kirsten Rehbein

Zürich, April 2006

Zusammenfassung

Mit der vorliegenden Studie wurde anhand der Schwermetalle Zink und Blei eine Vorgehensweise für die Erstellung von Bodenschadstoffkarten im Schweizer Mittelland erarbeitet und diskutiert. Ziel war die Darstellung der grossräumigen diffusen Hintergrundbelastung der Böden im Kanton Thurgau, ohne eine parzellenscharfe Interpretation zu ermöglichen. Der Kanton Thurgau verfügt über einen qualitativ und quantitativ sehr guten Bodenschadstoffdatensatz, dessen wichtigster Baustein den ganzen Kanton abdeckende land- und forstwirtschaftliche Untersuchungsraaster sind.

Für die räumliche Auswertung der Bodenschadstoffgehalte wurde ein Kriging in Verbindung mit einer robusten Regressionsmethode eingesetzt. Als erklärende Variablen im Regressionsmodell für die Streuung der Zink- und Bleigehalte im Boden stellten sich Nutzung, pH-Wert, Niederschlag, Bodentyp und Bodenart heraus. Für die Interpolation wurden Datensätze mit unterschiedlichen Nutzungs- und Belastungscharakteristiken erstellt. Auf Acker-, Grünland- und Waldstandorten wurden bis auf die Umgebung einer Verzinkerei flächendeckend Zinkgehalte geschätzt, die niedriger als der Richtwert der Verordnung über Belastungen des Bodens (VBBo, 1998) von 150 mg/kg waren. Eine Ausweitung des Datensatzes auf die Nutzungen Siedlung, Obst-, Gemüse- und Rebbau führte in einigen Siedlungsbereichen zu Schätzwerten von bis zu 1300 mg/kg. Mit den Messwerten der land- und forstwirtschaftlichen Standorte wurde in Bezug zu Prognoseintervall und Kreuzvalidierung die beste Schadstoffkarte erreicht. Die grosse Streuung der Daten, vor allem innerhalb der Siedlungsgebiete, konnte in der Regression und Interpolation nicht erfasst werden und führte zu einer Verletzung der Annahme der Stationarität. In der Folge wurden die Zinkgehalte für Acker-, Grünland- und Waldstandorte in der Nähe von Siedlungsgebieten überschätzt. Böden von Obst-, Gemüse- und Rebbau wurden durch den Siedlungseinfluss überlagert und waren daher bislang nicht beurteilbar. Die Karten sind maximal bis zu einem Massstab von 1:50.000 darstellbar, als sinnvoller wird ein Massstab von 1:100.000 erachtet.

Im Zuge der Auswertung wurde der Einfluss der robusten Schätzung im Regressionsmodell auf die Interpolation der Messwerte untersucht. Verwendet wurden letztlich die rohen Residuen. Der Einsatz von gewichteten Residuen führte zu einer starken Glättung hoher Zinkgehalte, während das grossräumige Muster der Zinkgehalte nahezu unverändert blieb. Der Einbezug einzelner erklärender Variablen im Regressionsmodell sowie der Ausschluss einiger weniger auffällig hoher Messwerte aus dem Datensatz kann die Aussagegenauigkeit der Zinkschätzkarte entscheidend verbessern.

Die erarbeitete Vorgehensweise zur räumlichen Analyse von Zinkgehalten im Boden erwies sich als genügend flexibel für die Analyse der Bleigehalte. Um die Übertragbarkeit dieses Vorgehens auch auf andere Kantone abzuschätzen, wurde zur Simulation geringerer Datendichten ein Datensatz iterativ ausgedünnt. Sofern Datensätze gleichmässig über ein Gebiet vorliegen, deuten die Ergebnisse für den Kanton Thurgau darauf hin, dass eine Datendichte von 0.45 Pkt/km² zu guten Ergebnissen führt, mit

0.22 Pkt/km² allerdings mit Einbussungen der Genauigkeit gerechnet werden muss. Diese Ergebnisse können allerdings nur ein erstes Indiz geben. Die Aussagekraft von Karten zur Darstellung von grossräumigen Bodenkonzentrationen in Bezug zur Messpunktdichte muss für weitere Regionen/Kantone geprüft werden.

Im Anschluss an diese Studie wird die entwickelte Methode auf andere Schwermetalle im Kanton Thurgau übertragen. Die Ergebnisse dienen als Exempel für mögliche weitere Kartierungen im Schweizer Mittelland. Voraussetzung sind dafür gleichmässig verteilte Datensätze mit einer ausreichenden Datendichte (die es noch genauer zu spezifizieren gilt) und sorgfältig dokumentierte Metainformationen.

Summary

In this study an approach is discussed and developed to produce soil pollution maps for the heavy metals zinc and lead in the Swiss midland. The aim was to illustrate the spacious diffuse background pollution in the soils of the canton of Thurgau, without creating the precision required to interpret individual plots of land. A very good soil pollution dataset exists for the canton of Thurgau, both in terms of quality and quantity, which is based on three agricultural and forestry surveys of the entire canton.

For the spatial statistics of heavy metal concentration a Kriging was used in combination with a robust regression. The best possible explanation for the variations in zinc and lead concentrations in the soil turned out to be due to the explanatory variables land use, pH, rainfall, soil type and soil texture. For the interpolation, datasets with different land use and contamination characteristics were produced. Except for the area around a galvanising plant, zinc concentrations on agricultural and forestry sites were estimated at levels below the recommended guide value of 150 mg/kg specified in the Ordinance relating to Impacts on the Soil (VBBo, 1998). Expanding the dataset to the land uses urban areas, orchards, vegetable gardening and vineyards, the estimated values for a number of urban areas increased up to levels of 1300 mg/kg. With regard to the confidence belt for individual observations and the cross validation, the best pollution map was produced from the dataset of agricultural and forestry sites. The high variation in the data, especially of the urban areas, could not be detected in the regression model or the interpolation and lead to violation of the assumptions of stationarity. As a consequence, zinc concentrations in agricultural and forestry sites close to urban areas were overestimated. Soils from orchards, vegetable gardening and vineyards were superimposed by the influence of the urban areas and thus could not be evaluated. The maps can be presented up to a scale of 1:50.000, however, a scale of 1:100.000 is considered wiser.

In the course of the spatial statistics, the influence of the robust estimation in the regression model on the interpolation of the measured values was examined and it was decided to use the raw residuals. Use of the weighted residuals lead to great smoothing of

higher zinc concentrations, while the spatial pattern of the zinc concentrations remained almost unchanged. Incorporating individual explanatory variables into the regression model or otherwise excluding exceptionally high values from the dataset can significantly improve the accuracy of the map.

The elaborated approach for the spatial statistics of zinc concentrations in soils proved to be flexible enough for the analyses of lead concentrations. To evaluate the transferability of this approach in other cantons, the datasets were iteratively thinned out to simulate lower densities of data. The results from the canton of Thurgau indicate that, as long as the datasets are evenly distributed throughout a region, a density of 0.45 points/km² provides good results, although at a density of 0.22 points/km², a loss of accuracy must be expected. However, these findings give only a first indication. The significance of maps illustrating large-scale spacious soil concentrations in relation to data densities needs to be examined in further cantons.

Following this study the developed approach will be used for other heavy metals in the canton of Thurgau. The results serve as an example for potential future surveys in the Swiss midland. Evenly distributed data with adequate densities (which needs to be specified) and carefully documented metadata are required.

Inhaltsverzeichnis

Zusammenfassung	I
Summary.....	II
1 Einleitung	1
2 Material und Methoden	4
2.1 Untersuchungsgebiet.....	4
2.2 Ablaufschema	8
2.3 Datenaufbereitung.....	9
2.3.1 Datenvalidierung	9
2.3.2 Grundlagenkarten	11
2.4 Datenselektion	11
2.4.1 Bodenschadstoffkarte Zink.....	11
2.4.2 Übertragbarkeit der Methode.....	12
2.5 Datenanalyse.....	13
2.5.1 Explorative Datenanalyse	13
2.5.2 Räumliche Datenanalyse	14
3 Ergebnisse und Diskussion.....	24
3.1 Datenaufbereitung.....	24
3.1.1 Datenvalidierung	24
3.1.2 Grundlagenkarten	28
3.2 Bodenschadstoffkarte Zink.....	30
3.2.1 Datensatz 673	31
3.2.2 Datensatz 793	43
3.2.3 Datensatz 1290	53
3.3 Übertragbarkeit der Methode.....	61
3.3.1 Datensatz 386/193	61
3.3.2 Datensatz Blei.....	71
4 Schlussfolgerungen, Empfehlungen und Ausblick.....	78
5 Literaturverzeichnis.....	81
Anhang A: Eliminierte Datensätze: Doppelte Werte pro Standort.....	85
Anhang B: Grundlagenkarten Kanton Thurgau.....	89
Anhang C: Prognoseintervalle der Schätzkarten für Zink.....	95
Anhang D: Nicht berücksichtigte Standorte Datensatz 1290	99

Verzeichnis der Figuren

Figur 2-1:	Biogeographische Regionen der Schweiz.....	4
Figur 2-2:	Gemeinden und Siedlungsstruktur im Kanton Thurgau	5
Figur 2-3:	Ablaufschema zur Erstellung der Bodenschadstoffkarten für den Kanton Thurgau	9
Figur 2-4:	Einteilung der Ausdünnungs-Sektoren im Untersuchungsgebiet	13
Figur 2-5:	Box Plot am Beispiel von Zinkgehalten auf Acker- und Grünlandnutzung.....	14
Figur 2-6:	Beispiel eines Tukey-Anscombe-Diagrammes mit Residuen und angepassten Werten (Fitted) für logarithmierte Zinkgehalte	17
Figur 2-7:	Beispielhafte Darstellung eines Semivariogramms (Nielsen & Wendroth, 2003)	18
Figur 2-8:	Beispielhafte Darstellung von theoretischen Variogrammfunktionen (Nielsen & Wendroth, 2003).....	19
Figur 3-1:	Standorte der Untersuchungskampagnen im Kanton Thurgau	25
Figur 3-2:	Verteilung der Schadstoffdaten (Totalgehalte) im Kanton Thurgau	28
Figur 3-3:	Zinkgehalte an den Untersuchungsstandorten Ausgangsdatensatz.....	31
Figur 3-4:	Verteilung der Zinkmesswerte Datensatz 673	33
Figur 3-5:	Univariate Statistik der erklärenden Variablen auf die Zink- Bodenbelastung Datensatz 673	33
Figur 3-6:	Zinkgehalte an den Untersuchungsstandorten Datensatz 673	34
Figur 3-7:	Streudiagramm der rohen Residuen gegen die angepassten Werte (Fitted) mit Nutzung Acker/Grünland (rot) und Wald (blau) (links), Histogramm der rohen Residuen (rechts), Zn673 Modell 1	37
Figur 3-8:	Streudiagramm der gewichteten Residuen gegen die angepassten Werte (Fitted) mit Nutzung Acker/Grünland (rot) und Wald (blau) (links), Histogramm der rohen Residuen (rechts), Zn673 Modell 1.....	37
Figur 3-9:	Streudiagramm der gewichteten Residuen gegen die angepassten Werte (Fitted) mit Nutzung Acker/Grünland (rot) und Wald (blau) (links), Histogramm der rohen Residuen (rechts), Zn673 Modell 2.....	37
Figur 3-10:	Semivariogramm der rohen Residuen Zn673 Modell 1 (links) und der gewichteten Residuen Zn673 Modell 1 (rechts)	38
Figur 3-11:	Semivariogramm der gewichteten Residuen Zn673 Modell 2	38
Figur 3-12:	Interpolierte Zinkgehalte auf Basis der rohen Residuen Zn673 Modell 1, interpoliert mit Simple Kriging	40
Figur 3-13:	Interpolierte Zinkgehalte auf Basis der gewichteten Residuen Zn673 Modell 1, interpoliert mit Simple Kriging	40

Figur 3-14:	Interpolierte Zinkgehalte auf Basis der gewichteten Residuen Zn673 Modell 2, interpoliert mit Ordinary Kriging	41
Figur 3-15:	Prognoseintervall des zweiseitigen 95 %-Quantils der interpolierten Zinkgehalte der Simple-Kriging-Vorhersagefehler, rohe Residuen Zn673 Modell 1.....	42
Figur 3-16:	Verteilung der Zinkmesswerte Datensatz 793	45
Figur 3-17:	Univariate Statistik der erklärenden Variablen auf die Zink- Bodenbelastung Datensatz 673	46
Figur 3-18:	Zinkgehalte an den Untersuchungsstandorten Datensatz 793	46
Figur 3-19:	Streudiagramm der rohen Residuen gegen die angepassten Werte (Fitted) mit Nutzung Acker/Grünland (rot) und Wald (blau) (links), Histogramm der rohen Residuen (rechts), Zn793 Modell 1	48
Figur 3-20:	Semivariogramm aller Residuen (links) und der Residuen ohne 4 auffällige Werte (rechts), Zn793 Modell 1	49
Figur 3-21:	Interpolierte Zinkgehalte auf Basis aller Residuen Zn793 Modell 1, interpoliert mit Ordinary Kriging.....	50
Figur 3-22:	Interpolierte Zinkgehalte auf Basis der Residuen ohne 4 auffällige Werte Zn793 Modell 1, interpoliert mit Simple Kriging.....	50
Figur 3-23:	Vergleich der interpolierten Zinkgehalte Zn793 Modell 1 (ohne 4 auffällige Werte) mit Zn673 Modell 1 (rohe Residuen)	51
Figur 3-24:	Verteilung der Zinkmesswerte Datensatz 1290	54
Figur 3-25:	Univariate Statistik der erklärenden Variablen auf die Zink- Bodenbelastung Datensatz 1290	55
Figur 3-26:	Zinkgehalte an den Untersuchungsstandorten Datensatz 1290	56
Figur 3-27:	Streudiagramm der rohen Residuen gegen die angepassten Werte (Fitted) mit Nutzung Acker/Grünland (rot), Wald (blau). Obst/Gemüse (gelb), Rebberge (grün) und Siedlungen (grau) (links), Histogramm der rohen Residuen (rechts), Zn1290, Modell 1	57
Figur 3-28:	Semivariogramm Zn1290, Modell 1.....	58
Figur 3-29:	Interpolierte Zinkgehalte Zn1290 Modell 1, interpoliert mit Ordinary Kriging.....	59
Figur 3-30:	Vergleich der interpolierten Zinkgehalte Zn1290 mit Zn793 (zur besseren Vergleichbarkeit mit Einbezug der 4 auffälligen Werte)	60
Figur 3-31:	Verteilung der Zinkmesswerte Datensatz 386 (oben), Datensatz 193 (unten)	62
Figur 3-32:	Univariate Statistik der erklärenden Variablen auf die Zink- Bodenbelastung Datensatz 386	63
Figur 3-33:	Univariate Statistik der erklärenden Variablen auf die Zink- Bodenbelastung Datensatz 193	64
Figur 3-34:	Standorte der Zinkgehalte von Zn386 und Zn193	65

Figur 3-35:	Streudiagramm der rohen Residuen gegen die angepassten Werte (Fitted) mit Nutzung Acker/Grünland (rot) und Wald (blau) (links), Histogramm der rohen Residuen (rechts), Zn386 Modell 1	66
Figur 3-36:	Streudiagramm der rohen Residuen gegen die angepassten Werte (Fitted) mit Nutzung Acker/Grünland (rot) und Wald (blau) (links), Histogramm der rohen Residuen (rechts), Zn193 Modell 1	66
Figur 3-37:	Semivariogramm Zn386 Modell 1, ohne 1 auffälligen Wert (links) und Zn193 Modell 1, ohne 4 auffällige Werte (rechts)	67
Figur 3-38:	Interpolierte Zinkgehalte Zn386 Modell 1 (ohne einen auffälligen Wert), interpoliert mit Simple Kriging	68
Figur 3-39:	Interpolierte Zinkgehalte Zn193 Modell 1 (ohne 4 auffällige Werte), interpoliert mit Ordinary Kriging.....	68
Figur 3-40:	Vergleich der interpolierten Zinkgehalte Zn386 mit Zn793.....	69
Figur 3-41:	Vergleich der interpolierten Zinkgehalte Zn193 mit Zn793	69
Figur 3-42:	Verteilung der Bleimesswerte Datensatz 793	72
Figur 3-43:	Univariate Statistik der erklärenden Variablen auf die Blei-Bodenbelastung Datensatz 793	72
Figur 3-44:	Bleigehalte an den Untersuchungsstandorten Datensatz 793	73
Figur 3-45:	Streudiagramm der rohen Residuen gegen die angepassten Werte (Fitted) mit Nutzung Acker/Grünland (rot) und Wald (blau) (links), Histogramm der rohen Residuen (rechts), Pb793 Modell 3	74
Figur 3-46:	Semivariogramm Pb793 Modell 3	75
Figur 3-47:	Interpolierte Bleigehalte Pb793 Modell 3, interpoliert mit Simple Kriging.....	76

Verzeichnis der Tabellen

Tabelle 2-1:	Relevante digitale Kartengrundlagen Kanton Thurgau	4
Tabelle 2-2:	Hauptnutzungstypen im Kanton Thurgau (Arealstatistik 92/97).....	6
Tabelle 2-3:	Bodenregionen im Kanton Thurgau (BÜK 50)	6
Tabelle 2-4:	Bodentypen im Kanton Thurgau (BÜK 50).....	7
Tabelle 2-5:	Bodenarten im Kanton Thurgau (BÜK 50)	8
Tabelle 3-1:	Mindestangaben für einen validen Datensatz	24
Tabelle 3-2:	Untersuchungen zu Bodenschadstoffen im Kanton Thurgau	26
Tabelle 3-3:	Erklärende Variablen der Bodenbelastung	29
Tabelle 3-4:	Allgemeine Statistische Kenngrössen Ausgangsdatsatz (in mg/kg)	30
Tabelle 3-5:	Beurteilungswerte (Totalgehalte) der Verordnung über Belastungen des Bodens (VBBo, 1998)	31
Tabelle 3-6:	Untersuchungskampagnen Datensatz 673	32
Tabelle 3-7:	Allgemeine statistische Kenngrössen der Zinkgehalte Datensatz 673	32
Tabelle 3-8:	Regressionsmodelle für logarithmierte Zinkgehalte Datensatz 673	36
Tabelle 3-9:	Kenngrössen der Kreuzvalidierung für den Datensatz Zn673	42
Tabelle 3-10:	Untersuchungskampagnen Datensatz 793	44
Tabelle 3-11:	Allgemeine statistische Kenngrössen der Zinkdaten Datensatz 793	45
Tabelle 3-12:	Regressionsmodell für logarithmierte Zinkgehalte Datensatz 793.....	47
Tabelle 3-13:	Kenngrössen der Kreuzvalidierung für den Datensatz Zn793.....	52
Tabelle 3-14:	Untersuchungskampagnen Datensatz 1290	53
Tabelle 3-15:	Allgemeine statistische Kenngrössen der Zinkdaten Datensatz 1290	54
Tabelle 3-16:	Regressionsmodell für logarithmierte Zinkgehalte Datensatz 1290.....	57
Tabelle 3-17:	Kenngrössen der Kreuzvalidierung für den Datensatz Zn1290	60
Tabelle 3-18:	Allgemeine statistische Kenngrössen der Zinkdaten Datensatz 386/193.....	62
Tabelle 3-19:	Regressionsmodelle für logarithmierte Zinkgehalte Datensatz 386/193.....	65
Tabelle 3-20:	Kenngrössen der Kreuzvalidierung für die Datensätze Zn386/193.....	70
Tabelle 3-21:	Allgemeine statistische Kenngrössen der Bleidaten Datensatz 793	71
Tabelle 3-22:	Regressionsmodell für die logarithmierten Bleigehalte Datensatz 793.....	74
Tabelle 3-23:	Kenngrössen der Kreuzvalidierung für den Datensatz Pb793	76

1 Einleitung

Um einen wirksamen vorsorgenden Bodenschutz betreiben zu können, werden verlässliche und bundesweite Informationen über Zustand, Belastung und Belastbarkeit der Böden benötigt (Schwarz et al., 1994). Der gesetzliche Auftrag der Schweizer Nationalen Bodenbeobachtung (NABO) wird in der Verordnung über Belastungen des Bodens (VBBo, 1998) geregelt und umfasst neben der zeitlichen auch die landesweite räumliche Erfassung von Bodenbelastungen.

Für eine landesweite räumliche Erfassung bedarf es zunächst einer Zusammenführung der verfügbaren Bodenschadstoffdaten. Zu diesem Zweck wurde im Jahr 2002 eine Machbarkeitsstudie für eine nationale Bodenschadstoff-Datenbank (NABODAT) mit Anbindung eines Geographischen Informationssystems (GIS) durchgeführt (Rehbein, 2004). Neben dem eigentlichen Betrieb einer solchen Datenbank stehen die verbesserte Nutzung und die Erarbeitung von fachlichen Grundlagen für den Bodenschutz im Zentrum. Zu den Kernaufgaben zählen statistische Auswertungen von Schadstoffbelastungen in Böden für spezifische Fragestellungen über Kantonsgrenzen hinweg, sowie deren räumliche Analyse und Interpolation mittels geostatistischer Methoden und anschließender Kartierung. Der im Anschluss an die Studie eingereichte Projektantrag zum Aufbau der NABODAT wurde im August 2005 vom Bundesamt für Umwelt (BAFU, ehemals BUWAL) genehmigt. Das Projekt befindet sich momentan in der Startphase.

Langfristiges Ziel der Zusammenführung von Bodendaten mit der NABODAT ist die Erstellung einer schweizweiten Bodenschadstoffkarte, die Anhaltspunkte über die diffuse grossräumige Bodenbelastung in der Schweiz liefert. Lokale anthropogen verursachte Belastungsflächen im Sinne von Altlasten(verdachts)flächen können und sollen mit diesem Instrument nicht beurteilt werden. Eine erste Zusammenführung der verfügbaren Bodenschadstoffdaten in 1996 (Keller & Desaulles, 2001) sowie die Resultate der Machbarkeitsstudie haben gezeigt, dass der Schweiz ein umfangreicher Datensatz zur Verfügung steht, der eine hohe Vergleichbarkeit hinsichtlich Probenahme und Analytik der Bodenproben aufweist. Damit ist grundsätzlich eine gute Ausgangsbasis für eine räumliche Analyse der chemischen Bodenbelastung gegeben.

Für die räumliche Interpolation von Bodendaten gibt es kein Standardverfahren, wohl aber Werkzeuge und Kriterien zu deren Anwendung. Aufgrund der Heterogenität des Mediums Boden und der Komplexität der beteiligten Prozesse und Einflussfaktoren sind räumlich-statistische Analysen bei bodenkundlichen Fragestellungen zudem eine besondere Herausforderung (Dobler et al., 2003). Isaaks & Srivastava (1989) verstehen die angewandte Geostatistik als eine Art Kunst, die somit weder komplett automatisierbar noch rein objektiv sein kann. In der Praxis steht eine Fülle von (Geo)Statistik-Software zur Verfügung, mit der auf schnellem Wege Ergebnisse erzielt werden können. Es obliegt aber der Fachperson zu überprüfen für welche Fälle die Annahmen zur mathematischen Ableitung der Interpolationsmethoden mehr oder minder erfüllt sind und dadurch die Aussagekraft der Ergebnisse geschmälert wird. Es gilt die Werkzeuge

fachgerecht einzusetzen und für die jeweilige Fragestellung die Kriterien zur Anwendung der Interpolationsmethode kritisch zu analysieren. In diesem Sinne wurde von Schulin et al. (1994) ein Rahmenkonzept bezüglich Strategien und detailliertem Vorgehen für regionale Erhebungen von Schadstoffbelastungen in Böden erarbeitet. Da eben die Variabilität des Bodens je nach Region sehr unterschiedlich sein kann, muss jede Erhebung diesen Besonderheiten Rechnung tragen. Die Möglichkeiten und der Nutzen von geostatistischen Methoden sind entsprechend unterschiedlich.

In der Schweiz wurden in den letzten Jahren einige Studien durchgeführt, die sich mit der Thematik der Interpolation von Schadstoffgehalten beschäftigt haben ((Meuli, 1997), (Steiger von et al., 1996), (Meuli et al., 1998), (Goovaerts et al., 1997)). Andere wiederum thematisieren eher die methodischen Grundlagen und deren Umsetzung (Papritz & Webster, 1995a; 1995b). In allen Studien wurde sich auf kleinere – teils belastete – Gebiete beschränkt, grossräumige Aussagen im kantonalen Massstab konnten bisher nicht durchgeführt werden.

In Deutschland dagegen wurde u. a. Ende der 90er Jahre im Bundesland Nordrhein-Westfalen (NRW) eine Vorgehensweise für die Erstellung von „Bodenbelastungskarten im Aussenbereich“ (Acker, Grünland, Wald) entwickelt (LUA NRW, 2000). Ein Leitfa-den für die Erstellung von Bodenbelastungskarten im Siedlungsgebiet von NRW ist momentan in Arbeit. Um eine breite Anwendung der Vorgehensweise zu erreichen, wurden die geostatistischen Möglichkeiten zur Erstellung der Karten stark minimiert, was zu unsachgemäsem Einsatz und äusserst fragwürdigen Ergebnissen führen kann. Im Jahr 2003 wurden im Auftrag der Bund-Länder-Arbeitsgemeinschaft Bodenschutz (LABO) die in den deutschen Bundesländern angewendeten Methoden zur flächigen Darstellung von Bodenschadstoffgehalten zusammengetragen (Hinterding et al., 2003a; 2003b) und „Empfehlungen für die Anwendung statistischer und geostatistischer Methoden zur flächenbezogenen Auswertung von Daten über Stoffgehalten in Böden“ (Dobler et al., 2003) abgeleitet. In Übereinstimmung mit Schulin et al. (1994) geben Dobler et al. (2003) allgemeine Kriterien und Überlegungen zur geostatistischen Analyse. Da jedoch jede Analyse einen Einzelfall darstellt, kann keine spezifische Vorgehensweise aufgezeigt werden.

Im Rahmen dieser Master Thesis sollen die Möglichkeiten und Grenzen einer flächen-deckenden Bodenschadstoffkartierung im Schweizer Mittelland erarbeitet werden. Ziel ist die Entwicklung von Karten, deren Massstabsbereich sowohl für nationale als auch für kantonale Belange von Interesse ist. Als Untersuchungsgebiet bietet sich hierfür der Kanton Thurgau mit einer umfangreichen Bodendatenbank an. Für die durchgeführte Studie wurden als Beispielparameter Zink und Blei ausgewählt. Aus Sicht der nationalen Bodenbeobachtung ist es ein Ziel die Möglichkeiten von digitalen Kartierungen im Massstabsbereich von etwa 1:200.000 bis 1:1.000.000 abzuklären. Aus Sicht der kantonalen Fachstelle, sind Kartierungen in möglichst grossem Massstab anzustreben (1:100.000 bis 1:200.000).

Daraus ergeben sich folgende zentrale Fragestellungen dieser Studie:

1. Welche Vorgehensweise ist für die Erstellung von Bodenschadstoffkarten im Schweizer Mittelland sinnvoll?
2. Mit welchen Schätzgenauigkeiten sind Bodenschadstoffkarten für den Kanton Thurgau möglich?
3. Bis zu welchem Massstab ist die Darstellung der Karten fachlich sinnvoll?
4. Welche Voraussetzungen müssen hinsichtlich Datendichte und Metainformationen für die Übertragung der Methodik auf andere Kantone gegeben sein?

In einem ersten Schritt erfolgt die Aufbereitung und Validierung der vorhandenen Punkt- und Flächeninformationen. Kernaspekt der (geo)statistischen Auswertung stellt die Entwicklung eines Regressionsmodells dar, deren Residuen in die Interpolation fliessen. Durch Variation der Messdaten wird mit Hilfe der Schätzfehler die bestmögliche Bodenschadstoffkarte für den Kanton Thurgau erzeugt. Um hinsichtlich der Datendichte, d.h. Bodenmessungen pro Flächeneinheit, die Möglichkeiten und den Nutzen der geostatistischen Methoden vom Kanton Thurgau auf andere Kantone grob abschätzen zu können, wurden die Messwerte ausgedünnt und erneut Karten erstellt.

Letztlich wird die gewählte Vorgehensweise diskutiert, die Ausgangslage im Kanton Thurgau beurteilt, die unter den diversen Bedingungen erzeugten Schadstoffkarten miteinander verglichen und interpretiert. Aus der Summe der Ergebnisse werden Schlussfolgerungen für die Übertragbarkeit der Methode auf andere Kantone abgeleitet.

Im Ergebnis liegen Bodenschadstoffkarten für Blei und Zink vor. Das entwickelte Vorgehen soll kurzfristig auf weitere Schwermetalle im Kanton Thurgau übertragen werden.

2 Material und Methoden

2.1 Untersuchungsgebiet

Als Untersuchungsgebiet wurde aufgrund seiner guten Bodenschadstoffdatenlage der Kanton Thurgau ausgewählt, der im nordöstlichen Schweizer Mittelland liegt (Figur 2-1).



Figur 2-1: Biogeographische Regionen der Schweiz

Zur Charakterisierung des Gebietes lassen sich vielfältige Informationen aus diversen Kartengrundlagen ableiten (Tabelle 2-1). So stehen digital verfügbare kantonale (wie die Bodenübersichtskarte des Kanton Thurgau (Presler et al., 2005)) sowie schweizweite Karten (wie die Arealstatistik,) zur Verfügung.

Tabelle 2-1: Relevante digitale Kartengrundlagen Kanton Thurgau

Geodatenatz	Erläuterung	Quelle	Daten-format	Massstab bzw. Gridgrösse
Landes-, Kantons-, Bezirks-, Gemeinde-grenzen (2004)	Administrative Grenzen der Schweiz	BFS ¹⁾	Polygone	-
PK 25 (1998)	Pixelkarten Kt. TG	Swisstopo ²⁾	Bild (Tiff)	1:25.000
BÜK TG (2005)	Bodenübersichtskarte Kt. TG	FaBo TG ³⁾	Polygone	1:50.000
Kantonaler Richtplan (2004)		ThurGIS ⁴⁾	Polygone	1:50.000
Arealstatistik (1992/97)	Landnutzung der Schweiz	BFS	Raster	100x100m
BEK 200 (2000)	Bodeneignungskarte der Schweiz (digitalisierte Version)	BFS	Polygone	1:200.000
GTK 200 (1963/67)	Vereinfachte Geotechnische Karte der Schweiz	BFS	Polygone	1:200.000
DHM 25 (2004)	Dig. Höhenmodell der Schweiz	BFS	Raster	1:25:000
HADES (1951/80)	Hydrol. Atlas der Schweiz	BAFU ⁵⁾	Raster	1000x1000m

¹⁾Bundesamt für Statistik ²⁾Bundesamt für Landestopographie ³⁾Fachstelle Bodenschutz Kanton Thurgau

⁴⁾Zentrum für Geoinformation Kanton Thurgau ⁵⁾Bundesamt für Umwelt

Insgesamt beträgt die Fläche des Kantons Thurgau 863 km². Davon werden laut kantonalem Richtplan 108 km² dem Siedlungsgebiet sowie weitere 25 km² dem Streusiedlungsgebiet zugeordnet, in Summe etwa 15 % des Kantonsgebietes. Rund ein Drittel der insgesamt 233.000 Einwohner lebt in den Gemeinden Frauenfeld, Kreuzlingen, Arbon, Amriswil und Weinfelden (Figur 2-2).



Figur 2-2: Gemeinden und Siedlungsstruktur im Kanton Thurgau

Das digitale Höhenmodell der Schweiz (DHM 25) stuft den Kanton zwischen 369 und 1010 m Höhe ein. Die tieferen Gebiete finden sich entlang der Flusstäler, insbesondere entlang der Thur und im Nordosten in Richtung Bodensee. Neben punktuellen Erhebungen bis 800 m Höhe vor allem im Norden im Bezirk Steckborn, steigt das Gelände im äussersten Süden des Kantons bis knapp über 1000 m Höhe an (Gemeinde Fischingen) (Anhang B- 1). In engem Zusammenhang zur Höhe steht erwartungsgemäss die Hangneigung. Überall dort, wo das Gelände merklich ansteigt, finden sich stark geneigte bis steile Hänge. Der Südzipfel des Kantons weist demnach vermehrt starke Hangneigungen auf (Anhang B- 2). Ähnliches zeigt sich für die Exposition. Im zum Bodensee abfallenden Gebiet im Norden des Kantons dominieren die nach Norden exponierten Hänge, während im Süden anschliessend in Richtung des Tals der Thur entsprechend die Südhänge überwiegen. Durch die zumeist von West nach Ost verlaufenden Höhenzüge scheinen Nord- und Südhänge im Kanton vorzuherrschen (Anhang B- 3).

Der mittlere Jahresniederschlag liegt laut des Hydrologischen Atlases der Schweiz (HADES) zwischen 880 und 1760 mm. Der nördliche Bereich des Kantons in Richtung Bodensee weist zumeist Jahresniederschläge von weniger als 1000 mm auf, während in Richtung Süden mit zunehmender Höhe auch die mittleren Niederschläge ansteigen (Anhang B- 4).

Die Landnutzung besteht überwiegend aus Acker und Grünland, gefolgt von Wald und Siedlungsgebieten (Anhang B- 5). Laut Arealstatistik 92/97 sind die in 6 Klassen aggregierten Hauptnutzungstypen folgendermassen über den Kanton verteilt (Tabelle 2-2):

Tabelle 2-2: Hauptnutzungstypen im Kanton Thurgau (Arealstatistik 92/97)

Nutzungstyp	Kürzel	Fläche [km ²]	Flächenanteil [%]
Acker/Grünland	AckGrü	468.35	54.3
Wald	Wal	214.52	24.9
Obst- und Gemüse	O+G	57.90	6.7
Rebberge	Reb	2.68	0.003
Siedlungsflächen	Sie	108.87	12,6
Sonstiges	Son	10.57	1,2
Gesamt		862.89	100

Informationen zum geogenen Ausgangssubstrat liefert die Bodenübersichtskarte des Kanton Thurgau (BÜK 50). Sie deckt insgesamt knapp 90 % des Kantonsgebietes ab; bei den nicht berücksichtigten Arealen handelt es sich um die Siedlungsgebiete. Die definierten Ausgangsgesteine wurden – teilweise aktualisiert – der Geologischen Übersichtskarte des Kanton Thurgau im Massstab 1:50.000 entnommen (Presler et al., 2005). Fast die Hälfte des Kantonsgebietes besteht aus Grundmoränenmaterial. Entlang der Thur finden sich in erster Linie Rückzugsschotter und Postglaziale Alluvien (Sand & Kies) und in den höheren Lagen die Obere Süsswassermolasse (OSM) bzw. nagelfluhreiche OSM. Ebenfalls auf Grundlage der Geologischen Übersichtskarte wurden für den Kanton Thurgau 10 Bodenregionen definiert und abgegrenzt. Die Bodeneignungskarte der Schweiz (EJPD et al., 1980) findet – entgegen ursprünglicher Planung – keine Berücksichtigung, da zu viele Unstimmigkeiten zwischen den beiden Kartenwerken existieren (Presler et al., 2005). Die Hälfte des Kantons ist entsprechend der Ausgangssubstrate mit Böden aus Grundmoräne bedeckt, gefolgt von Böden aus Molasse sowie Ebenen und Talsohlen mit Böden aus Schotter und sandig-kiesigem Alluvium (Tabelle 2-3, Anhang B- 6).

Tabelle 2-3: Bodenregionen im Kanton Thurgau (BÜK 50)

Bodenregion	Kürzel	Fläche [km ²]	Flächenanteil [%]
Ebenen und Talsohlen mit Böden aus Schotter und sandig-kiesigem Alluvium	R1	79.3	10.5
Ebenen und Talsohlen mit Böden aus feinkörnigem Alluvium	R2	17.9	2.4
Talmulden und Akkumulationslagen mit Kolluvialen Böden	R3	12.3	1.6
Gebiete mit Böden aus Grundmoräne	R4	355.0	46.8
Gebiete mit Böden aus lockerer Moräne	R5	65.0	8.6
Drumlinglebiete mit Böden aus Moräne	R6	35.7	4.7
Gebiete mit Böden aus Moräne und Molasse	R7	64.6	8.5
Gebiete mit Böden aus Molasse	R8	113.7	15.0
Bachschuttfächer	R9	15.0	2.0
Anthropogen stark veränderte Böden	R10	-	-
Gesamt		758.5	100

Im Rahmen der Erarbeitung der BÜK wurden sog. Bodenmuster definiert. Es handelt sich dabei um strukturierte Bodenkomplexe, die aus einem Hauptboden, Nebenboden und Begleitboden zusammengesetzt sind. Für die Fallstudie wird in der Folge ausschliesslich der Hauptboden, der die flächenmässig wichtigste Bodenform einer Bodeneinheit darstellt, berücksichtigt. Des Weiteren werden die Bodentypen nach Nutzung unterschieden. Es gibt ein Bodenmuster Landwirtschaftsböden und eines für Waldböden. Da diese nicht räumlich klar abgegrenzt werden konnten, ist für gleiche Gebiete oftmals ein Bodentyp Wald als auch ein Bodentyp Landwirtschaft definiert (Presler et al., 2005). Der dominierende Bodentyp im landwirtschaftlichen Gebiet ist die Braunerde, einzig im Tal der Thur finden sich fast ausschliesslich Fluvisole (Tabelle 2-4, Anhang B- 7). Die Waldböden setzten sich ebenfalls zu einem grossen Anteil aus Braunerden zusammen, vielfach allerdings auch in Form von Parabraunerden und Braunerde-Pseudogleyen. Als weniger stark entwickelte Böden treten im Wald vermehrt Regosole auf (Tabelle 2-4, Anhang B- 7).

Tabelle 2-4: Bodentypen im Kanton Thurgau (BÜK 50)

Bodentyp	Kürzel (FAL, 1997)	Landwirtschaft		Wald	
		Fläche [km ²]	Flächenanteil [%]	Fläche [km ²]	Flächenanteil [%]
Braunerde	B	449.1	68.0	101.6	27.1
Braunerde-Gley	V	10.9	1.6	2.8	0.8
Braunerde-Pseudogley	Y	-	-	38.3	10.2
Buntgley	W	1.0	0.2	1.4	0.4
Fahlgley	G	5.9	0.9	2.0	0.5
Fluvisol	F	66.6	10.1	25.1	6.7
Halbmoor	N	3.5	0.5	-	-
Kalkbraunerde	K	82.3	12.5	17.8	4.8
Parabraunerde	T	10.5	1.6	87.0	23.2
Regosol	O	28.2	4.3	60.3	16.1
Rendzina	R	-	-	0.5	0.1
Saure Braunerde	E	2.8	0.4	38.0	10.1
Gesamt		660.8	100	374.8	100

Die Bodenarten sind analog zur Definition der Bodentypen in Bodenarten in landwirtschaftlich geprägten Böden und Bodenarten in Waldböden unterteilt (Presler et al., 2005). Wie aus Tabelle 2-5 und Anhang B- 8 ersichtlich ist, dominieren in Landwirtschaftsböden flächendeckend Lehm Böden, gelegentlich sandig oder tonig. Lehmiger Schluff oder lehmreicher Sand treten als einzige andere Bodenarten nur vereinzelt im Kantonsgebiet auf. Ähnliches gilt für die Waldböden, die ebenfalls grösstenteils aus Lehm Böden bestehen. Sand oder Schluff als Hauptbodenarten treten lediglich vereinzelt auf (Tabelle 2-5, Anhang B- 8).

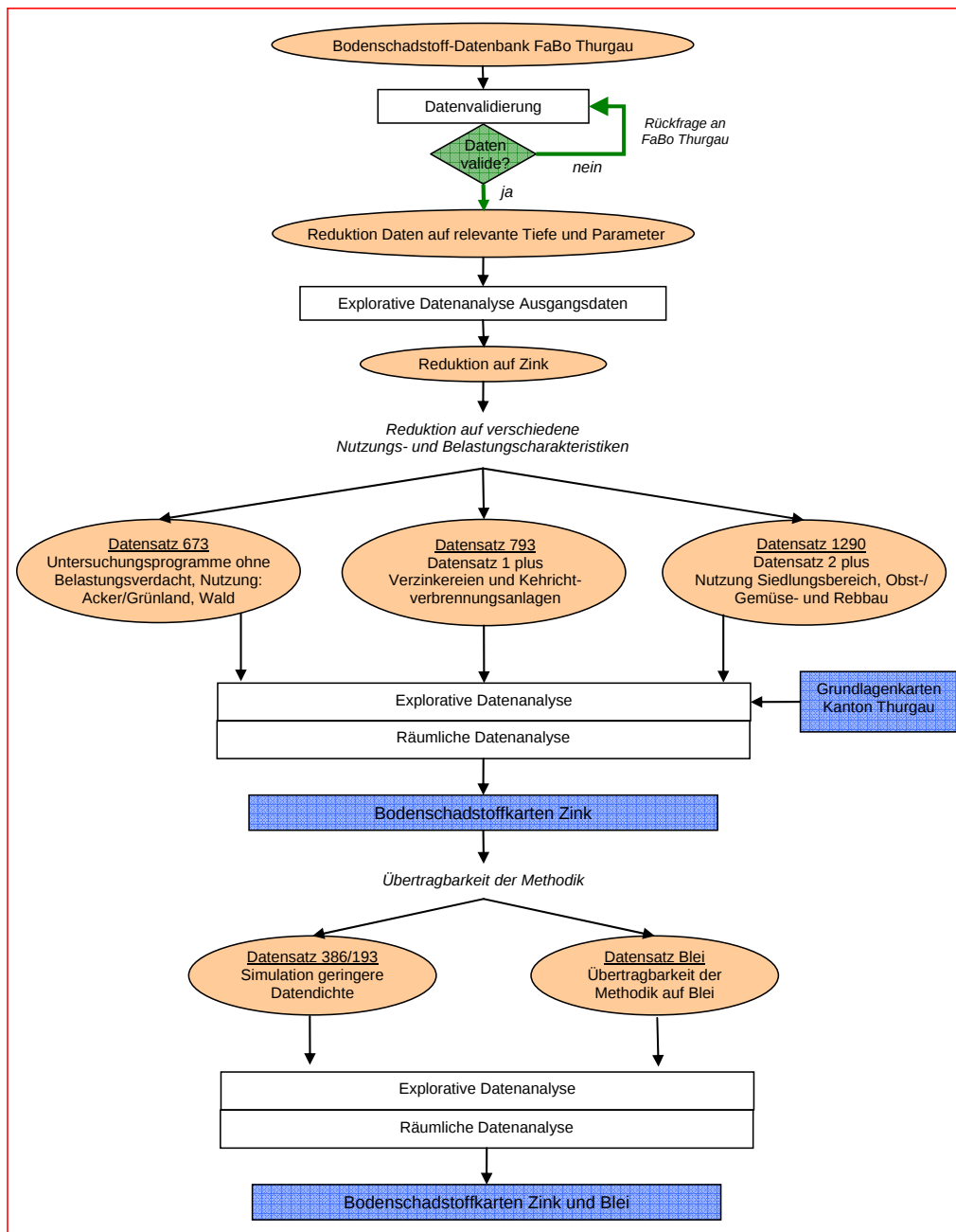
Tabelle 2-5: Bodenarten im Kanton Thurgau (BÜK 50)

Bodenart	Kürzel (FAL, 1997)	Landwirtschaft		Wald	
		Fläche [km ²]	Flächenanteil [%]	Fläche [km ²]	Flächenanteil [%]
Lehm	L	522.1	79.0	87.9	23.4
sandiger Lehm	sL	115.9	17.5	172.4	45.9
toniger Lehm	tL	15.0	2.3	73.7	19.6
lehmiger Sand	lS	-	-	19.7	5.2
lehmreicher Sand	lrS	7.1	1.1	9.2	2.5
schluffiger Sand	uS	-	-	2.1	0.6
lehmiger Schluff	lU	1.0	0.2	7.3	1.9
sandiger Schluff	sU	-	-	3.0	0.9
lehmiger Ton	lT	-	-	0.0	0.0
Gesamt		661.1	100	375.3	100

2.2 Ablaufschema

Die Vorgehensweise dieser Studie ist schematisch in Figur 2-3 dargestellt. Kernaspekt ist die Entwicklung eines Regressionsmodells im Rahmen der räumlichen Datenanalyse, die allerdings zunächst eine sorgfältige Datenvalidierung zwingend voraussetzt. Neben den Punktdaten sind für räumliche Auswertungen auch vorhandene digitale Flächendaten von grosser Bedeutung. Erst wenn alle notwendigen Angaben zu den Daten vorhanden sind, können die Daten auf die relevanten Informationen reduziert werden und die Statistik bzw. Geostatistik als nächster Block erfolgen. Dabei werden für die Erstellung der Bodenschadstoffkarte Zink verschiedene Datensätze mit unterschiedlichen Nutzungseinheiten und Belastungsniveaus gebildet. Die Studie unterteilt sich in 2 Abschnitte:

1. *Bodenschadstoffkarte Zink:*
Erstellung der bestmöglichen Bodenschadstoffkarte für Zink unter Verwendung aller geeigneten Bodenschadstoffdaten inklusive vorhandener Metainformationen.
2. *Übertragbarkeit der Methode:*
Simulation einer geringeren Datendichte und Übertragbarkeit der Methode auf den Parameter Blei.



Figur 2-3: Ablaufschema zur Erstellung der Bodenschadstoffkarten für den Kanton Thurgau

2.3 Datenaufbereitung

2.3.1 Datenvalidierung

Ziel der Datenvalidierung ist zu prüfen, welche relevanten und untereinander vergleichbaren Punktdaten für die Bearbeitung vorliegen. Dazu gehört zunächst eine Kontrolle der Daten auf ihre Lesbarkeit und Vollständigkeit.

Mindestanforderungen

Darüber hinaus gilt es zu prüfen, ob gewisse Mindestanforderungen an die Daten erfüllt sind. Zum Beispiel, ob die zur Interpretation und Vergleichbarkeit der Daten nötigen Informationen zum Untersuchungsprogramm, zum Standort, zur Probenahme und zur Analytik vorliegen. Traten im Rahmen dieser Validierung Unklarheiten auf, wurden Rückfragen an die Fachstelle Bodenschutz erforderlich.

Relevante Daten

Nicht alle validen Daten sind für die Erstellung der Bodenschadstoffkarten relevant. So liegen Daten aus unterschiedlichen Probenahmetiefen vor, die nicht miteinander vergleichbar sind. Dank Vorgabe der VBBo entstammt ein Grossteil der Bodenschadstoffdaten in der Schweiz aus einer Tiefe von 0-20 cm. Ebenso ist die Auswahl der Schadstoffe, für die Bodenschadstoffkarten erstellt werden können, begrenzt. Nur Parameter, die mit einer ausreichenden Dichte über das Kantonsgebiet beprobt wurden, erlauben eine flächenhafte Auswertung.

Für die Entwicklung einer Vorgehensweise für die Erstellung von Bodenschadstoffkarten im Schweizer Mittelland beschränkt sich die Studie auf Pb und Zn als Beispielparаметer. Für die übrigen relevanten Parameter werden lediglich für eine grobe Einordnung der Schadstoffgehalte allgemeine Kenngrössen aufgezeigt.

Plausibilitätsprüfung

Abschliessend werden die relevanten Daten einer Plausibilitätsprüfung unterzogen, in der die Wertespanssen der interessierenden Parameter, deren Einheiten, Bestimmungsgrenzen und die angegebenen Koordinaten auf ihre Glaubhaftigkeit untersucht werden. Eingabefehler oder fehlende Angaben in der Datenbasis lassen sich so bereits im Vorfeld eliminieren.

Prüfung auf Homogenität und Stationarität

Eine Grundvoraussetzung für die räumliche Statistik von Geodaten ist deren Stationarität und Homogenität. Eine Messgrösse wird als räumliche Zufallsvariable aufgefasst, d.h. die Messgrösse variiert räumlich als Ergebnis eines stochastischen Zufallsprozesses. Die Stationarität ist dann erfüllt, wenn die Wahrscheinlichkeit einen bestimmten Messwert bei beliebig vielen Wiederholungen unter genau den gleichen Bedingungen zu erhalten an jedem Ort des Untersuchungsgebietes gleich hoch ist (Webster & Oliver, 2001).

Ist die Stationarität gewährleistet, gilt es zu prüfen, ob die Daten ebenfalls homogen sind. Messwerte können nur dann miteinander verglichen werden, wenn sie hinsichtlich ihrer Probenahme und Analytik vergleichbar sind. Man spricht von Homogenität, wenn man annehmen kann, dass die einzelnen Messungen einer Grundgesamtheit entstammen (Papritz, 2003).

Die Beachtung der Homogenität erfolgt zum einen im Zuge der Datenvalidierung, indem die Daten auf solche reduziert werden, die ausreichend Metainformationen besitzen. Damit kann ein hinsichtlich Probenahme und Analytik vergleichbarer Datensatz erstellt werden. Darüber hinaus werden zur Erstellung der Bodenschadstoffkarte verschiedene Datensätze selektiert, was eine Homogenität wie auch Stationarität der für die Interpolation berücksichtigten Messwerte mit sich bringt (vgl. Kapitel 2.4).

2.3.2 Grundlagenkarten

Mit dem Hinzuziehen der digitalen Grundlagenkarten werden verschiedene Ziele verfolgt. Topographische Karten dienen dem Überblick des Untersuchungsgebietes. Für die spätere geostatistische Auswertung sind in erster Linie thematische Karten interessant, wie z.B. Landnutzung oder Geologie. Diese und andere Faktoren können den Schadstoffgehalt im Boden entscheidend beeinflussen und müssen daher zu den Messwerten – soweit möglich – in Beziehung gesetzt werden. Sie werden als erklärende Variablen bezeichnet.

Liegen im Untersuchungsgebiet für diese erklärenden Variablen Karten vor, können Metainformationen, die in der Datenbank vielleicht nicht oder nur unvollständig vorhanden sind, kompensiert werden. Die Zuordnung erfolgt in einem GIS durch eine Punkt-in-Polygon-/ bzw. Raster-Analyse. Die ermittelten Variablen fließen neben der Zielvariablen (dem gemessenen Schadstoffgehalt) in die geostatistischen Auswertungen ein.

Als erklärende Variablen können ausschliesslich flächendeckend vorhandene Standortinformationen verwendet werden, um das Regressionsmodell (vgl. Kap. 2.5.2) auf jede Zelle im Untersuchungsgebiet anwenden zu können. Dies schliesst Faktoren wie „Projekt“ oder „Labor“, die ebenfalls als erklärende Variablen im Modell fungieren könnten, von vornherein aus.

2.4 Datenselektion

2.4.1 Bodenschadstoffkarte Zink

Als Ausgangsdaten werden alle diejenigen Daten bezeichnet, die nach der Datenvalidierung als relevant und plausibel für die Zielvariablen selektiert wurden. Aus Gründen der Homogenität und Stationarität der Bodengehalte wurden folgende Datensätze mit unterschiedlichen Nutzungs- und Belastungscharakteristiken gebildet:

Datensatz 673

Der Ausgangsdatsatz wurde in einem ersten Schritt stark eingeschränkt. Ziel war es, einen möglichst „sauberen“ Datensatz zu erzeugen, dessen Messwerte die angestrebte diffuse Hintergrundbelastung durch Schadstoffe im Boden bestmöglich repräsentieren.

Lokale Spezialfälle, die räumlich nicht verallgemeinerbar sind oder nicht der Hintergrundbelastung entsprechen, sollten zunächst keine Berücksichtigung finden.

Als „sauber“ betrachtet wurden alle Standorte auf den Nutzungen Acker, Grünland oder Wald, die nicht aus Untersuchungsprogrammen mit explizitem Belastungsverdacht stammen und somit keine lokale Belastungsspitzen darstellen. Ebenfalls nicht berücksichtigt wurden zunächst die Messdaten aus dem Siedlungsgebiet, Obst-/Gemüsebau und den Rebbergen. Sie repräsentieren nicht die diffuse Hintergrundbelastung.

Datensatz 793

Mit dem Ausschluss von Daten von Belastungsuntersuchungen befindet man sich auf der sicheren Seite, läuft jedoch Gefahr auch Messwerte auszuschliessen, die zwar im Rahmen von Belastungsuntersuchungen erhoben wurden, aber dennoch räumlich korrelieren können. Dies betrifft die Untersuchungskampagnen im Umfeld der Verzinkereien und Kehrichtverbrennungsanlagen (KVAs), die im Zuge der ersten Datensatzerweiterung wieder einbezogen wurden.

Datensatz 1290

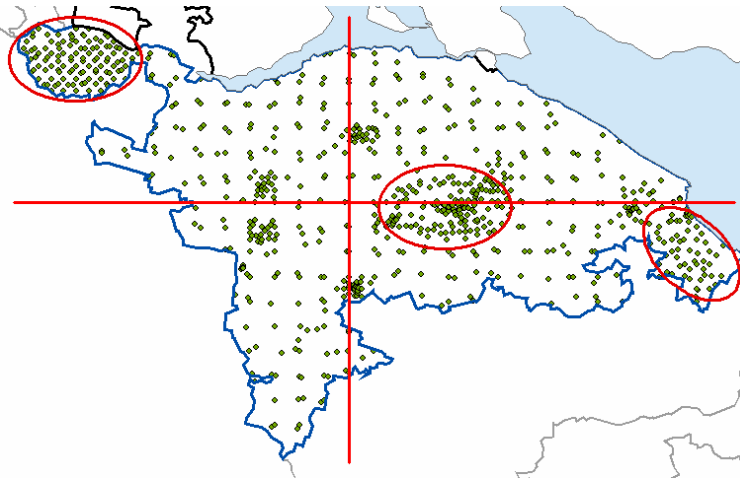
In der zweiten Erweiterung des Datensatzes geht es um die Einbeziehung der Messdaten, die auf Standorten im Siedlungsgebiet, im Obst- oder Gemüsebau oder in Rebbergen gewonnen wurden. Neben weiteren Daten der bereits in den vorhergehenden Selektionsschritten ausgewählten Projekte wurden nun auch neue Projekte berücksichtigt, wie z.B. die Rasteruntersuchungen in Siedlungsgebieten oder die Untersuchungskampagnen in aktuellen oder ehemaligen Rebbergen. Lediglich einzelne Projekte, wie Schiessplatzuntersuchungen, wurden auch in dieser zweiten und letzten Erweiterung nicht berücksichtigt, da es sich um einzeln auftretende lokal begrenzte Schadstoffanreicherungen handelt, die geostatistisch weder relevant noch auswertbar sind.

2.4.2 Übertragbarkeit der Methode

Datensatz 386/193

Langfristiges Ziel ist die Übertragung der Methode auf weitere Gebiete des Mittellandes. Der Kanton Thurgau verfügt über einen umfangreichen Datensatz, der nicht in anderen Gebieten als gegeben vorausgesetzt werden kann. Das betrifft in erster Linie die Anzahl und Verteilung der Messpunkte.

Zur Abschätzung der Übertragbarkeit der Ergebnisse auf andere Kantone wurden daher durch Ausdünnung der Messpunkte weitere Datensätze mit geringerer Dichte erzeugt. Dafür wurde der Datensatz 793 um 50 % und um 75 % der Messwerte stratifiziert zufällig ausgedünnt. Als Straten wurden 7 Gebiete behandelt, einerseits 4 (nach Anzahl Untersuchungsstandorte gleichgrosse) Quadranten des Kantonsgebiet sowie separat die 3 Gebiete in Diessenhofen, Weinfelden und Arbon, welche verstärkt beprobt worden sind (Figur 2-4).



Figur 2-4: Einteilung der Ausdünnungs-Sektoren im Untersuchungsgebiet

Datensatz Blei

Zur Überprüfung der entwickelten Methode anhand eines weiteren Parameters werden die Bleimesswerte analog zum Zink-Datensatz 793 verwendet.

2.5 Datenanalyse

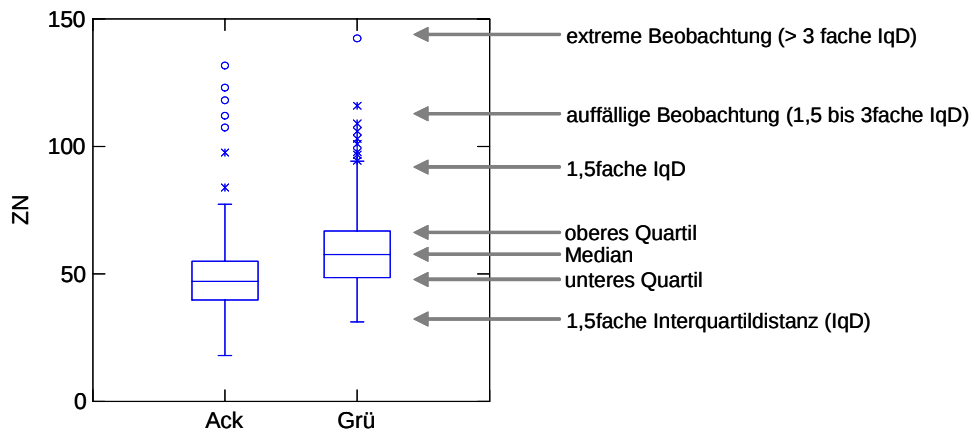
2.5.1 Explorative Datenanalyse

Statistische Kenngrößen und Normalverteilung

Die statistischen Kenngrößen dienen zunächst der Einordnung der Größenverhältnisse der betrachteten Daten. Dafür wurden Mittelwert, Median, Standardabweichung, Perzentile etc. ermittelt. Die Verteilung der Daten wurde mittels Histogrammen visualisiert und bei einer Schiefe (Skewness) von > 1 log-transformiert (entspricht in dieser Studie dem dekadischen Logarithmus). Das gleiche gilt für die erklärenden Variablen, die ebenfalls für die Einbeziehung in das Regressionsmodell annähernd normalverteilt sein sollten.

Einfluss erklärende Variablen

Um Unterschiede der Schadstoffgehalte hinsichtlich Standorteigenschaften zu analysieren, wurden Box-Plots eingesetzt, die einen raschen Vergleich zwischen Stichproben ermöglichen (Stahel, 2002). Figur 2-5 gibt die statistischen Kennzahlen der Box-Plot Abbildung wieder.



Figur 2-5: Box Plot am Beispiel von Zinkgehalten [mg/kg] auf Acker- und Grünlandnutzung

Für die kontinuierlichen Variablen werden Streudiagramme eingesetzt, die ebenso rasch Zusammenhänge zwischen erklärender Variable und Zielvariable aufdecken.

Räumliche Verteilung

Neben den statistischen Auswertungen ist es ebenfalls hilfreich die Daten räumlich zu visualisieren. So erhält man auf einen Blick nicht nur einen Eindruck in welchen Größenordnungen die Daten liegen, sondern darüber hinaus, wie sie räumlich verteilt sind und wo die statistisch auffälligen Werte liegen.

2.5.2 Räumliche Datenanalyse

Unter einer räumlichen Datenanalyse versteht man im Prinzip die Antwort auf die Frage „Wo gibt es wie viel?“. Die Geostatistik basiert dabei auf der Theorie der regionalisierten Zufallsvariablen. Eine regionalisierte Zufallsvariable ist eine Variable $z(s)$, die die Werte einer Grösse in Abhängigkeit vom Ort x angibt. Solche räumlichen Zufallsprozesse $Z(s)$ beinhalten sowohl einen deterministischen ($\mu(s)$) als auch einen stochastischen Anteil ($E(s)$):

$$Z(s) = \mu(s) + E(s) \quad (2.1)$$

Der deterministische Anteil besteht aus einer Trendkomponente, die für das instationäre Verhalten des Gesamtprozesses verantwortlich ist. Die stochastische Abhängigkeit des Prozesses wird durch eine autokorrelative Komponente und eine rein zufällige Komponente, häufig als weisses Rauschen bezeichnet, gebildet (Heinrich, 1992). Unter Autokorrelation versteht man, dass sich räumlich benachbarte Punkte in ihrer Wertedimension ähnlicher sind als weiter entfernte Punkte.

Zur bestmöglichen Zerlegung des deterministischen und des stochastischen Anteils wird in dieser Studie das räumliche Interpolationsverfahren Kriging in Verbindung mit einer Regression durchgeführt. Diskutiert wird diese Methode unter anderem von Hengl et al. (2005; 2003; 2004), Moyeed und Papritz (2002), Papritz und Moyeed (1999), Papritz und Dubois (1999) oder Papritz und Stein (1999). Beim Kriging handelt es sich um ein

Interpolationsverfahren, welches ermöglicht unbekannte Werte (z.B. Schadstoffgehalte) zwischen vorhandenen gemessenen Werten zu schätzen. Im Gegensatz zu anderen Interpolationsverfahren wie der Bildung von Thiessen Polygonen oder der Inverse-Distance-Weighted (IDW) Methode, werden beim Kriging in Form von sog. Semivariogrammen die räumliche Korrelation benachbarter Punkte mit einbezogen und – besonders wichtig – Schätzfehler abgeleitet. In den letzten Jahren hat die Bedeutung von so genannten „hybriden“ Interpolationsverfahren zugenommen. Sie kombinieren Kriging mit weiteren Informationen. Dazu gehört auch das Kriging in Verbindung mit einer Regression oder das co-kriging (Einbezug einer zweiten Variablen). Im Ergebnis vieler Studien hat sich gezeigt, dass ein Kriging kombiniert mit einer Regression detailliertere und präzisere Vorhersagen liefert als die herkömmlichen einfachen Krigingmethoden oder auch das co-kriging (Hengl et al., 2004). Letzteres ist sinnvoll, wenn die eine oder mehrere erklärenden Variablen nicht flächendeckend, sondern nur punktuell vorliegen (Hinterding et al., 2003a).

In dieser Studie wird die Zielvariable Zink im Kanton Thurgau mittels einer Regression angepasst und die resultierenden Residuen anschliessend interpoliert. Neben der Variation der Datensätze werden verschiedene Einstellungen hinsichtlich Regressionsmodell, Auswahl der Residuen und Interpolation diskutiert. Zur besseren Einschätzung der erzeugten Bodenschadstoffkarten werden zudem für einen Datensatz die Bleigehalte interpoliert und die Ergebnisse mit der entsprechenden Zinkkarte verglichen.

Die Grundlagen der Theorie des verwendeten Regressionsmodells, der Variographie und der Interpolation sowie deren Evaluation werden im Folgenden kurz dargestellt.

Regressionsmodell

Zur Ermittlung des deterministischen Anteils wird das Modell einer multiplen linearen Regression angewendet. Ziel ist die vollständige Eliminierung der möglicherweise in den Daten vorhandenen Drift mit dem Ergebnis unabhängiger Residuen, die den stochastischen Anteil repräsentieren. Die Abhängigkeit einer Zielgrösse (z.B. der Zinkmesswerte) von einer Ausgangsgrösse (z.B. dem pH-Wert) kann gut grafisch in einem einfachen Streudiagramm dargestellt werden (vgl. Kap. 2.5.1), womit das Wesentliche des Zusammenhangs sofort ersichtlich wird (Stahel, 2002). Wenn der Zusammenhang einer Zielgrösse allerdings zu mehreren erklärenden Variablen $X^{(1)}, X^{(2)}, \dots, X^{(m)}$ untersucht werden soll, reichen die grafischen Mittel nicht mehr aus und es wird das Modell der multiplen linearen Regression verwendet, welches lautet:

$$Y_i = \alpha + \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + \dots + \beta_m x_i^{(m)} + E_i. \quad (2.2)$$

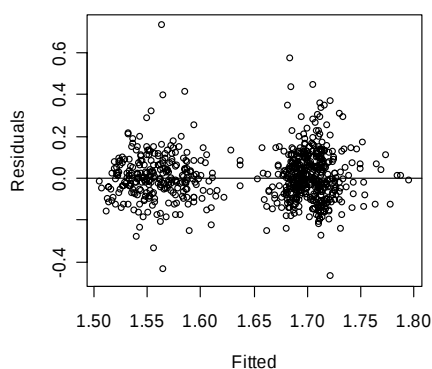
mit:	Y_i	Zielgrösse Y
	α	Achsenabschnitt
	β	Koeffizienten (1,2,...m)
	E_i	zufällige Abweichungen (Fehler) mit der Varianz σ^2

Die klassische Schätzung der Koeffizienten α und β_j erfolgt über die „Methode der Kleinsten Quadrate“ (least squares). Grundannahme für die Schätzung der Koeffizienten ist eine Normalverteilung der Variablen. Ist dies nicht exakt, sondern - trotz Transformation - höchstens annäherungsweise der Fall, können von der Normalverteilung abweichende Werte das Modell entscheidend beeinflussen. Gerade bei Umweltdaten ist eine rechtsschiefe Verteilung weit verbreitet. Eine Möglichkeit zur Beseitigung der Langschwänzigkeit, die bei der „Methode der Kleinsten Quadrate“ zu einer nicht optimalen Verteilung der Fehler führt, besteht in der Eliminierung von Ausreißern, deren Detektion nicht trivial ist. Eine bessere Lösung ist die Verwendung einer Schätzung, die einer (leicht) langschwänzigen Verteilung entspricht (Stahel, 2002). Man spricht von „Robusten Methoden“. Diese robusten Methoden führen zu einer effizienteren Verteilung der Fehler, indem Beobachtungen mit extremen Residuen heruntergewichtet werden. Je langschwänziger eine Verteilung, desto effizienter der Einsatz einer robusten Methode.

Als Kriterium für die Robustheit eines Schätzers dienen die Einflussfunktion und der Bruchpunkt. Die empirische Einflussfunktion zeigt den Einfluss einer zusätzlichen Beobachtung (x_0) zu einer gegebenen Stichprobe auf eine bestimmte Schätzung. Bei einer robusten Schätzung ist die Einflussfunktion beschränkt. Je niedriger der absolute Wert der Einflussfunktion, desto besser die Robustheit. So ist beispielsweise der Median ein sehr robuster Schätzer, er wird kaum von der Hinzunahme eines Wertes $\rightarrow \pm\infty$ beeinflusst. Der Mittelwert $\bar{x}$ dagegen wird $\bar{x} \rightarrow \pm\infty$, wenn $x_{(0)} \rightarrow \pm\infty$. Ein einfaches Mass zur Erfassung dieses Aspektes der Robustheit ist der Bruchpunkt. Dabei wird nicht mehr nur eine Beobachtung variiert, sondern iterativ eine beliebige Auswahl von m -Beobachtungen aus den n vorhandenen. Der Bruchpunkt des Medians z.B. liegt erst bei 50 % der Werte, bevor der Median beliebig gross wird oder anders ausgedrückt „zusammenbricht“. Der Bruchpunkt des Mittelwertes liegt dagegen bei 0 %. Unter dem Bruchpunkt versteht man somit den maximalen Anteil von Beobachtungen m/n , für den die Schätzung noch nicht zusammenbricht.

In der vorliegenden Fallstudie wird als robuster Schätzer der MM-Schätzer verwendet, der auf einem Huber-Typ M-Schätzer beruht. Als M-Schätzer bezeichnet man einen Maximum-Likelihood-Typ Schätzer, der sowohl nicht robust sein kann wie die gewöhnliche Kleinste-Quadrate Schätzung, oder robust wie z.B. die Funktion nach Huber oder Tukey. Die robuste Schätzung kann auch als eine gewichtete Kleinste-Quadrate Schätzung aufgefasst werden, die über eine so genannte „Tuning Konstante“ gesteuert wird. Beim MM-Schätzer wird der Huber-Typ M-Schätzer mit einer wieder absinkenden Funktion (ψ) kombiniert, die Beobachtungen mit grossen Residuen allmählich, je nach Betrag des Residuums, ignoriert. (Je nach Auswertung kann es sinnvoll sein, die rohen oder die über die ψ -Funktion weiter gewichteten Residuen zu verwenden.) Der MM-Schätzer kann als sehr robust bezeichnet werden, da er sowohl einen hohen Bruchpunkt als auch eine hohe Effizienz zeigt. Einen Überblick zum Thema der robusten Statistik bieten u. a. Publikationen von Huber (1981), Hampel (1986) und Venables (1994).

In der Fallstudie wird mit dem Open-Source (Geo)Statistikprogramm „R“ gearbeitet, welches in Form der „rlm-Funktion“ eine robuste multiple lineare Regression mit Verwendung des MM-Schätzers bietet. Zur Einschätzung der Güte der Regression und somit zur Auswahl des besten Regressionsmodells dient in erster Linie das Bestimmtheitsmass (R^2), das Quadrat der multiplen Korrelationen, also der Korrelationen zwischen den Beobachtungen Y_i und den angepassten Werten (fitted values). Es zeigt im Rahmen der Fallstudie z.B. an, wie viel Prozent der Streuung der Zinkgehalte im Boden mit den in der Regression verwendeten erklärenden Variablen beschrieben werden kann. Neben dem Bestimmtheitsmass können eine Reihe von grafischen Darstellungen zur Residuenanalyse eingesetzt werden. In dieser Studie wird das Tukey-Anscombe-Diagramm eingesetzt, welches Residualwerte und angepasste Werte gegenüberstellt. Annahme für den Fehlerterm, also die Residuen ist, dass sie a) im Mittel gleich Null sind, b) konstante Varianz aufweisen und c) gegenseitig unabhängig sind (Papritz, 2003). Letzteres wird mit dem Kapitel Variographie beschrieben, während die ersten beiden Annahmen mittels Figur 2-6 schnell ersichtlich werden. Sie sind erfüllt, wenn die Punktwolke symmetrisch um die Nulllinie schwankt.



Figur 2-6: Beispiel eines Tukey-Anscombe-Diagrammes mit Residuen und angepassten Werten (Fitted) für logarithmierte Zinkgehalte

Die Auswahl des besten Regressionsmodells für einen Datensatz geschieht jedoch nicht nur auf Grundlage der ausgegebenen Kenngrößen. Ebenso wichtig ist die Überlegung, ob die auf das Modell einflussnehmenden erklärenden Variablen aus fachlicher Sicht sinnvoll sind. Nur so können Scheinzusammenhänge verhindert werden. Führt darüber hinaus eine Variable lediglich zu einem geringfügig besseren Bestimmtheitsmass, sollte gegebenenfalls auf den Einbezug dieser Variable verzichtet werden, um die Anzahl Freiheitsgrade nicht unnötig zu reduzieren.

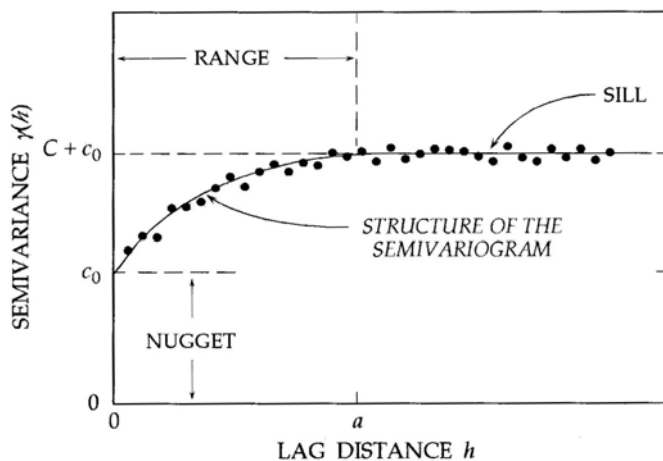
Variographie

Voraussetzung für ein Kriging ist die Kenntnis über die Autokorrelation der vorhandenen Messgrößen, in diesem Falle der Residualwerte der Regression. Grundannahme ist, dass sich benachbarte Punkte stärker ähneln als weiter entfernte Punkte, d. h. die Varianzen zweier Punkte werden grösser, je weiter sie auseinander liegen. Ausgedrückt wird diese Autokorrelation in der Semivarianz $\gamma(h)$, welche die halbe, mittlere, quadratische

euklidische Distanz h zwischen den Messwerten $z(x_i)$ und $z(x_i+h)$ an den Orten x_i und x_i+h darstellt. Sie wird über die folgende Formel definiert:

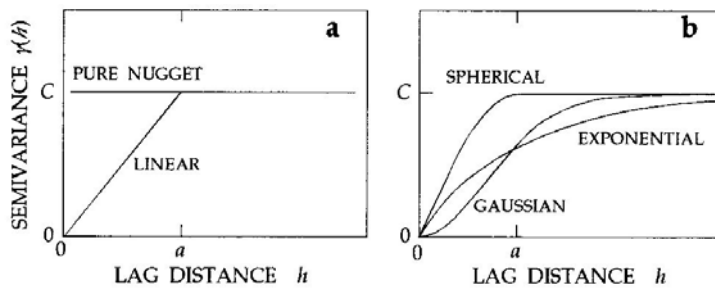
$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [z(x_i) - z(x_i+h)]^2. \quad (2.3)$$

Zur Verdeutlichung der räumlichen Abhängigkeiten werden sog. Semivariogramme erstellt, die die Semivarianzen in Abhängigkeit zur Entfernung zeigen. Wie in Figur 2-7 dargestellt, muss damit gerechnet werden, dass die Semivarianz zwischen zwei benachbarten Punkten, auch wenn deren Entfernung noch so gering ist, nicht ganz Null wird. Es bleibt gewöhnlich ein Rauschen übrig, die sog. Nugget-Varianz. Mit zunehmender Distanz nimmt die Semivarianz dann zu, bis ein Schwellenwert (Sill) erreicht wird, ab dem die Differenz zwischen Wertepaaren nicht mehr grösser wird. Der „Range“ markiert den Bereich, in dem die Semivarianzen noch zunehmen.



Figur 2-7: Beispielhafte Darstellung eines Semivariogramms (Nielsen & Wendroth, 2003)

Die Darstellung der Wertepaare wird als experimentelles Variogramm bezeichnet. Um die räumliche Information für die Interpolation nutzbar zu machen, wird versucht, eine bestmögliche Regressionsfunktion an die Punkte anzupassen. Man spricht vom theoretischen Variogramm. Hierfür kommen verschiedene Modelle in Frage, von denen einige an dieser Stelle kurz thematisiert werden. Trifft die Grundannahme zu, dass die Ähnlichkeit zweier Messwerte mit zunehmender Distanz abnimmt und in einen Schwellenwert übergeht, wird ein experimentelles Variogramm durch ein lineares (in diesem Fall begrenzt), sphärisches, exponentielles oder gaussian Modell abgebildet (Figur 2-8). Ist dagegen die Varianz eines Wertepaares unabhängig von der Entfernung immer annähernd gleich gross, spricht man von einem Pure-Nugget-Variogramm, welches keinerlei räumliche Abhängigkeiten erkennen lässt und entsprechende Auswirkungen auf die Schätzung hat. Im Rahmen der Studie werden die entsprechenden Effekte aufgezeigt und diskutiert.



Figur 2-8: Beispielhafte Darstellung von theoretischen Variogrammfunktionen (Nielsen & Wendroth, 2003)

Die Gleichungen dieser Modelle sind gegeben als

$$\text{Pure Nugget: } \gamma(h) = \begin{cases} 0 & h=0 \\ C & h>0 \end{cases} \quad (2.4)$$

$$\text{Linear: } \gamma(h) = \begin{cases} 0h/a & 0 \leq h \leq a \\ C & h > a \end{cases} \quad (2.5)$$

$$\text{Sphärisch: } \gamma(h) = \begin{cases} 0 \left[\frac{3h}{2a} - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right] & 0 \leq h \leq a \\ C & h > a \end{cases} \quad (2.6)$$

$$\text{Exponentiell: } \gamma(h) = C[1 - \exp(-h/a)] \quad h \geq 0 \quad (2.7)$$

$$\text{Gaussian: } \gamma(h) = C \left\{ 1 - \exp \left[- (h/a)^2 \right] \right\} \quad h \geq 0 \quad (2.8)$$

, wobei die Semivarianz γ für eine Distanz h mit a als Skalierungsfaktor für den Range sowie den Schwellenwert C ausgedrückt wird.

Alle diese Modelle sind isotrop, d.h. die Semivarianz der Messwerte hängt nicht von der Richtung des Distanzvektors ab. Ist dies der Fall, spricht man von Anisotropie. Grundsätzlich kann nicht erwartet werden, dass die Variation der Schadstoffdaten in alle Richtungen gleich ist. Auch Bodeneigenschaften, als erklärende Variablen für Bodenschadstoffdaten, sind in aller Regel richtungsabhängig. Anisotropie trifft dann zu, wenn der Schwellenwert C für verschiedene Richtungen in unterschiedlicher Entfernung erreicht wird. Im Modell wird in diesem Falle ein Winkel in die Richtung mit der grössten Autokorrelationsdistanz aufgenommen.

Zur Erstellung der Semivariogramme wird die Software Surfer 8 sowie der Geostatistical Analyst im ArcGIS 9.0 eingesetzt. Vor allem letzteres Tool bietet vielfältige Möglichkeiten zur Identifizierung von räumlich schlecht korrelierten Stützpunkten. Eine davon ist die Erstellung einer „Semivariogramm-Cloud“, in der die Semivarianz eines jeden Wertepaares ausgewiesen wird. Durch Markierung einer oder mehrerer aufgetra-

gener Semivarianzen werden die entsprechenden Stützpunkte in der Karte signalisiert. Oft genügt der Ausschluss weniger hoher Werte, um die angezeigten Semivarianzen entscheidend zu reduzieren.

Interpolation - Kriging

Zur Interpolation der im Zuge der Regression erhaltenen Residuen (stochastischer Anteil mit mehr oder weniger grosser noch enthaltener Drift) werden zwei verschiedene Kriging-Verfahren eingesetzt. Zum einen das Ordinary Kriging und zum anderen das Simple Kriging.

Ziel des Kriging ist es, die Werte $z(x_1), z(x_2), \dots, z(x_n)$ einer Zufallsvariable Z an einer nicht beprobten Stelle $x_1, x_2, \dots, x_n$ zu schätzen. Die in Form der Semivariogramme ermittelte parametrisierte Information über die räumliche Korrelationsstruktur der betrachteten regionalisierten Variable ist Grundlage für das Schätzverfahren.

Das gebräuchlichste Kriging-Verfahren ist das **Ordinary Kriging** (Webster & Oliver, 2001). Es wird von einem konstanten, aber unbekannten Erwartungswert $\hat{Z}(x_0)$ ausgegangen. Als Schätzer wird der „beste lineare erwartungstreue Schätzer“ BLUP (best linear unbiased estimator) verwendet, der neben der Erwartungstreue auch die Bedingung einer minimierten Schätzvarianz erfüllt. Er ist gegeben durch

$$\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i z(x_i) \quad (2.9)$$

, wobei $\hat{Z}(x_0)$ den Schätzwert an der Stelle x_0 bezeichnet und λ_i die aus der Semivariogrammfunktion resultierenden Gewichtungen, deren Summe gleich 1 ist:

$$\sum_{i=1}^n \lambda_i = 1 \quad (2.10)$$

Die Wahl der Gewichte erfolgt so, dass die Varianz zwischen $\hat{Z}(x_0)$ und $Z(x_0)$ gegen 0 geht. Diese Minimierung der Schätzvarianz wird erreicht, wenn

$$\sum_{i=1}^n \lambda_i \gamma(x_i, x_j) + \psi(x_0) = \gamma(x_j, x_0) \quad \text{für } j = 1, 2, \dots, n \quad (2.11)$$

, wobei $\gamma(x_i, x_j)$ die Semivarianz zwischen den bekannten Messpunkten x_i und x_j und $\gamma(x_j, x_0)$ die Semivarianz zwischen dem bekannten Messpunkt x_j und dem unbekannten Punkt x_0 bezeichnet. Mit ψ wird ein so genannter Lagrange Multiplikator einbezogen. Mit den Gleichungen 2.9 und 2.10 ist ein System von $n+1$ linearen Gleichungen gegeben, aus denen sich die Unbekannten $\lambda_1, \lambda_2, \dots, \lambda_n$ und ψ ermitteln lassen. Die Schätzvarianz $\delta^2(x_0)$ von $\hat{Z}(x_0)$ ist dann gegeben durch

$$\delta^2[(x_0)] = \sum_{i=1}^n \lambda_i \gamma(x_i, x_0) + \psi. \quad (2.12)$$

Die dargestellten Formeln beschreiben den Lösungsweg eines Punktkrigings, der analog für ein Blockkriging angewendet werden kann. Der Unterschied besteht lediglich darin, dass anschliessend über die gewählte Blockgrösse gemittelt wird. Im Rahmen dieser Studie wurde ein Blockkriging mit einer Gridgrösse von 1ha durchgeführt.

Ein weiteres in dieser Studie angewendetes Kriging-Verfahren ist das so genannte **Simple Kriging**. Bei dieser Methode wird ebenfalls von einem konstanten aber diesmal bekannten Erwartungswert μ ausgegangen. Der Schätzwert $\hat{Z}(x_0)$ ergibt sich aus folgender Gleichung:

$$\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i z(x_i) + \left\{ 1 - \sum_{i=1}^n \lambda_i \right\} \mu \quad (2.13)$$

Die Gewichte müssen bei diesem Verfahren nicht mehr in der Summe 1 sein. Sie lassen sich errechnen, in dem n Gleichungen mit n Unbekannten, diesmal ohne den Lagrange Multiplikator, gelöst werden. Die Gleichung lautet demnach

$$\sum_{i=1}^n \lambda_i \gamma(x_i, x_j) = \gamma(x_j, x_0). \quad \text{für } j = 1, 2, \dots, n \quad (2.14)$$

Die zugehörige Krigingvarianz ist analog dazu gegeben durch

$$\delta^2[(x_0)] = \sum_{i=1}^n \lambda_i \gamma(x_i, x_0). \quad (2.15)$$

Die Wahl, wann ein Ordinary und wann ein Simple Kriging angewendet wird, hängt von der räumlichen Korrelation der Residuen ab. Grundsätzlich geht die in dieser Studie gewählte Vorgehensweise (Kriging in Verbindung mit einer Regression) auf die Idee des Universal Kriging zurück, d.h. Kriging unter Berücksichtigung eines Trends. Da mit der eingesetzten Software, dem Geostatistical Analyst (Erweiterung im ArcGIS 9.0 der Firma Esri), ein Universal Kriging mit Integrierung der Regressionsfunktion nicht möglich ist, werden die Residuen separat interpoliert und die Regressionsfunktion im Anschluss manuell mittels des „Raster Calculators“ rückgerechnet. Das bietet allerdings die Möglichkeit, je nach räumlicher Korrelation der Residuen ein passendes Interpolationsverfahren auszuwählen. Idealerweise wird mit dem Regressionsmodell die Drift in den Daten vollständig eliminiert, aber die Fülle der erklärenden Variablen auf die Bodenbelastung macht dieses Ziel oft unmöglich.

Im Falle von noch räumlich korrelierenden Residuen wird in dieser Studie ein Simple Kriging angewendet. Es wird vorausgesetzt, dass der Erwartungswert $\hat{Z}(x_0)$ an jeder Stelle x_0 bekannt ist, da durch die Variogrammfunktion gut beschrieben. Schwanken die Werte unabhängig um einen Schwellenwert, kann der Erwartungswert nicht als bekannt

angesehen werden, die Semivarianzen werden durch ein Pure-Nugget-Modell nur annähernd beschrieben und in Folge dessen ein Ordinary-Kriging angewendet.

Wie viele Stützpunkte $x_1, x_2, \dots, x_n$ für die Schätzung von x_0 herangezogen werden, hängt vom Range des Semivariogramms ab. Allerdings werden für die Interpolation eine Mindest- und Maximalstützpunktanzahl gewählt von 12 bis 24, d.h. es müssen mindestens 12 umgebende Messwerte und maximal 24 in die Schätzung von x_0 einfließen. Weiter entfernte werden je nach Variogramm durch eine geringe Gewichtung allerdings kaum Einfluss nehmen können. Eine Anzahl von mehr als 24 Messwerten, auch wenn diese noch innerhalb des Ranges liegen, führt zu keinen verbesserten Ergebnissen, sondern lediglich zu einem aufwendigeren Gleichungssystem. Als Gridgrösse für die Schätzung wird (aufgrund der vorliegenden Grundlagenkarten) eine Auflösung von 1 ha gewählt. Für erste kantonsweiten Aussagen wäre ein Raster mit 500x500 m oder 1x1 km völlig ausreichend gewesen, hätte allerdings eine Aggregation von Nutzungsdaten mit entsprechenden Spielregeln bedurft.

Prognoseintervall und Kreuzvalidierung

In dieser Studie werden hinsichtlich Datenanzahl, Datendichte, Regressionsmodell, Variographie und Interpolation verschiedene Annahmen getroffen und die Ergebnisse miteinander verglichen. Um Aussagen über die Genauigkeit der Interpolation machen zu können, werden Prognoseintervalle gebildet. Die eigentliche Güte der gesamten Vorgehensweise, d.h. von der Anpassung der Regressionsfunktion, über die Variogrammfunktion hin zur Schätzfunktion, kann anhand dieser Intervalle jedoch nicht aufgezeigt werden, dazu wird die so genannte Kreuzvalidierung hinzugezogen.

Das Prognoseintervall (PI) auf Basis der Schätzvarianzen der Interpolation dient dazu aufzuzeigen, in welchen Bereichen die geschätzten Gehalte im Boden aufgrund der geostatistischen Annahmen schwanken können. Im Gegensatz zum Vertrauensintervall wird nicht über eine gegebene, aber unbekannte Beobachtung etwas ausgesagt, sondern über eine Zufallsvariable (Stahel, 2002). In diesem Fall wird die Zufallsabweichung einer Beobachtung mit einbezogen. Dementsprechend ist das Prognoseintervall (auch Vorhersageintervall genannt) breiter als das Vertrauensband (auch Konfidenzintervall genannt). Zur Berechnung wird der geschätzte deterministische Anteil ($\hat{\alpha} + \hat{\beta}x_0$), ermittelt mit dem Regressionsmodell, mit einem geeigneten Quantil t der t-Verteilung (95 %-Quantil zweiseitig für > 30 Freiheitsgrade) und der Standardabweichung des Schätzwertes $\hat{\delta}(x_0)$ in Beziehung gesetzt:

$$PI = (\hat{\alpha} + \hat{\beta}x_0) \pm t * \hat{\delta}(x_0) \quad (2.16)$$

Mittels der Kreuzvalidierung werden die Abweichungen zwischen Mess- und Schätzwerten aufgezeigt. Dazu wird die Interpolation n -mal wiederholt und zwar jeweils mit $n-1$ Daten. Das heisst, es wird sukzessive ein Stützpunkt nach dem anderen ausgeschlossen, die Interpolation durchgeführt und anschliessend die Differenz (der Fehler)

zwischen Messwert $z(x_i)$ und geschätztem Wert $\hat{z}(x_i)$ ermittelt. Dann wird der Stützpunkt wieder einbezogen und der nächste eliminiert.

Als Kenngrößen werden der mittlere Fehler ME (mean error) und die Variabilität der Vorhersage RMSSE (root-mean-square standardized error) gebildet. Der mittlere Fehler ist gegeben durch

$$ME = \frac{1}{N} \sum_{i=1}^N [z(x_i) - \hat{z}(x_i)]. \quad (2.17)$$

Er sollte unter der Annahme einer erwartungstreuen Schätzung beim Kriging idealerweise bei 0 liegen. Um die Variabilität der Vorhersage vom Messwert korrekt einzuschätzen, wird jeder Fehler mit seiner geschätzten Vorhersagestandardabweichung dividiert und somit der RMSSE bestimmt:

$$RMSSE = \frac{[z(x_i) - \hat{z}(x_i)]}{\hat{\sigma}(x_i)} \quad (2.18)$$

Im Idealfall ist dieser gleich 1. Bei einem RMSSE von > 1 wird die Variabilität in der Vorhersage unterschätzt, bei < 1 überschätzt (ESRI, 2001).

Um auch hier eine robuste Auswertung zu erstellen, werden der Median des Fehlers MedE und der Median des Absolutbetrages der Fehler MAD (median absolute deviation) als robuste Schätzer gegenübergestellt. Sie sind gegeben durch:

$$MedE = med[z(x_i) - \hat{z}(x_i)] \quad (2.19)$$

$$MAD = med\{|[z(x_i) - \hat{z}(x_i)] - med[z(x_i) - \hat{z}(x_i)]|\} \quad (2.20)$$

Ebenfalls unter der Annahme eines erwartungstreuen Schätzers sollten die beiden Kenngrößen gegen 0 gehen. Zur Evaluation des geostatistischen Modells wird in dieser Studie in erster Linie der MAD betrachtet, wobei der RMSSE nicht ausser Acht gelassen werden darf, als Mass für eine systematische Unter- oder Überschätzung des Schätzfehlers.

3 Ergebnisse und Diskussion

3.1 Datenaufbereitung

3.1.1 Datenvalidierung

Mindestanforderungen

Die von der Fachstelle Bodenschutz Thurgau zur Verfügung gestellte Datenbank ist lesbar und auf den ersten Blick vollständig. Da nur zu einigen Projekten Berichte vorliegen, waren einige mündliche Nachfragen bei der Fachstelle erforderlich, um die für eine räumliche Auswertung der Bodenschadstoffdaten zwingend notwendigen Metainformationen zu erhalten (Tabelle 3-1). Als besonders wichtig ist hier die Angabe zum Untersuchungsprogramm zu nennen. Wenn das Ziel einer Untersuchungskampagne bekannt ist, lassen sich u. U. bereits Rückschlüsse auf die gemessenen Schadstoffdaten ziehen. Bis auf das geogene Ausgangsmaterial liegen alle benötigten Informationen vor. Letzteres ist nicht zwingend, jedoch wünschenswert für spätere Interpretationen der Schadstoffgehalte und wurde daher aus den vorliegenden Grundlagenkarten (vgl. Kap. 3.1.2) zugeordnet.

Tabelle 3-1: Mindestangaben für einen validen Datensatz

Thematik	• Information	vorhanden
Angaben zum Untersuchungsprogramm	• Titel / Veranlassung der Untersuchung • Probenehmer	✓ ✓
Angaben zum Standort	• Koordinaten (möglichst auf 10 m genau) • Landnutzung • geogenes Ausgangsmaterial • Angaben zur Umgebungssituation (Belastungsverdacht)	✓ ✓ teilweise ✓
Angaben zur Probenahme	• Probenahmedatum • Probenahmetiefe • Probenart	✓ ✓ ✓
Angaben zur Analytik	• Labor • Aufschluss • Analyse • Messeinheit	✓ ✓ ✓ ✓

Alle Messwerte mit den dazugehörigen notwendigen Metainformationen wurden zusammengetragen und eindeutige Identifikationsnummern für jeden Datensatz, jede Probe, Standort und jedes Projekt vergeben.

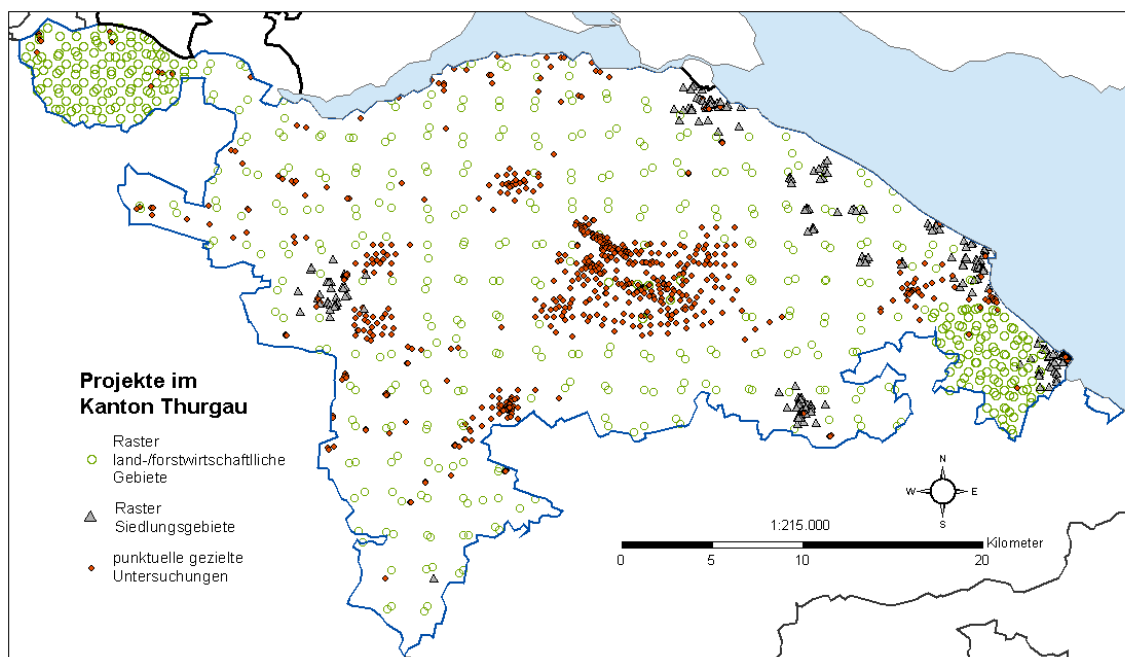
Vorhandene Schadstoffdaten

In der Datenbank des Kanton Thurgau liegen Bodenschadstoffdaten zu 34 Untersuchungskampagnen vor, die zwischen 1991 und 2001 durchgeführt wurden (Tabelle 3-2). In erster Linie wurden Schwermetalle der Verordnung über Belastungen des Bodens (VBBo, 1998) und in vereinzelt Kampagnen auch polyzyklische aromatische Koh-

lenwasserstoffe (PAK) und polychlorierte Biphenyle (PCB) analysiert. Insgesamt liegen Daten zu 1660 Standorten vor. Einige Standorte wurden im Rahmen nachfolgender Untersuchungsprogramme wiederholt beprobt.

Je nach Intention des Projektes wurden die Standorte zufällig (teilweise rasterförmig) oder gezielt ausgewählt. Bei den umfangreichen Projekten handelt es sich um Rasterbeprobungen, die entweder in land- und forstwirtschaftlichen Gebieten (Projekt Nr. 1, 3, 5, 30) oder in Siedlungsbereichen (Nr. 2, 12) durchgeführt wurden. Die Beprobungen der land- und forstwirtschaftlichen Gebiete erstrecken sich flächendeckend über das gesamte Kantonsgebiet, wie in Figur 3-1 dargestellt. Auffällig war, dass im Rasternetz immer 2 Messstellen direkt nebeneinander liegen. Rückfragen bei der Fachstelle Bodenschutz haben ergeben, dass grundsätzlich ein Waldstandort und ein Acker-, Grünland- oder Intensivstandort in unmittelbarer Nähe zueinander beprobt worden ist.

Gezielte Probenahmen erfolgten bei Nachbeprobungen der Rasteruntersuchungen, entlang von Strassen, in Rebbaugeländen, auf ehemaligen Schiessplätzen, im Umfeld geplanter oder existierender Kehrlichtverbrennungsanlagen (KVA), Abwasserreinigungsanlagen (ARA), Verzinkereien oder Bitumenfabriken, auf geplanten Deponiestandorten oder im Rahmen von Baugesuchen.



Figur 3-1: Standorte der Untersuchungskampagnen im Kanton Thurgau

Tabelle 3-2: Untersuchungen zu Bodenschadstoffen im Kanton Thurgau

Nr	Projekt	Thema	Stao(n)	Nutzung ¹⁾	relevante Parameter
1	P95RAS	SM in Land- und Forstwirtschaft	352	Wal, Ack, Grü, O+G, Son	Pb,Cd,Cr,Co,Cu,Ni, Hg,Zn,Fl,Tl,Mo
2	P94SIE	SM in Siedlungsböden (Lehmann, 1995c)	167	Grü, Sie, O+G	Pb,Cd,Cr,Co,Cu,Ni, Hg,Zn,As,Fl,Pak,Pcb
3	P93LW DI	SM in intens. Landwirtschaftsbezirken	114	Ack, Wal, Grü, O+G, Sie	Pb,Cd,Cr,Co,Cu,Ni, Hg,Zn,As,Fl
4	P92KVAW	Untersuchung vor KVA	103	Grü, Ack, Wal, O+G, Sie, Reb	Pb,Cd,Cr,Co,Cu,Ni, Hg,Zn,As,Fl
5	P93LW OT	SM in intens. Landwirtschaftsbezirken	93	Grü, O+G, Sie, Ack, Wal	Pb,Cd,Co,Cr,Cu,Ni, Hg,Zn,As,Fl,Tl,Mo
6	P95STR	SM entlang von Strassen (Lehmann, 1996)	90	Grü, Son, Wal, Ack	Pb,Cd,Cr,Co,Cu,Ni, Hg,Zn,Fl
7	P00SA	Schiessanlagen	70	Grü, Ack, Sie	Pb
8	P95RB OB	Kupferbelastung in Rebbergen (Lehmann, 1998)	68	Reb, Grü, O+G, Ack, Sie	Pb,Cd,Cu,Zn
9	P00REB	Kupferbelastung in Rebbergen	60	Reb, Grü, O+G, Ack	Pb,Cd,Cu,Zn
10	P92KVAW G	Untersuchung vor KVA	60	O+G, Ack, Wal, Sie	Pb,Cd,Cu,Hg,Zn,As,Fl
11	P98ARBON	Verdichtung P95RAS	53	Sie, Grü, Wal	Pb,Cd,Cu,Zn
12	P97SIE	SM in Siedlungsböden (Rutz, 1997b)	46	Sie	Pb,Cd,Cu,Zn,Pak
13	P99BST	Überprüfung hoher Werte im Kanton	41	Sie, Ack, Grü, Wal, Son	Pb,Cd,Cr,Cu,Ni,Hg, Zn,Mo
14	P94ARA MT	SM in der Umgebung einer ARA (Rutz, 1997a)	40	Ack, Grü, O+G	Pb,Cd,Cu,Zn
15	P93VZB	SM in der Umgebung einer Verzinkerei (1995a; Lehmann, 1995b)	38	Grü, O+G, Wal, Ack, Sie	Pb,Cd,Co,Cr,Cu,Ni, Hg,Zn,Fl
16	P92KVAH	SM in der Umgebung einer KVA (Lehmann, 1993)	31	Grü, O+G, Wal, Ack, Sie	Pb,Cd,Co,Cr,Cu,Ni, Hg,Zn,As,Fl
17	P92KVAM	SM in der Umgebung einer KVA (Lehmann, 1993)	30	Ack, Grü, Wal, O+G	Pb,Cd,Cr,Co,Cu,Ni, Hg,Zn,As,Fl
18	P99ARA	SM auf Klärschlammflächen	30	Ack, Grü	Pb,Cd,Cr,Cu,Ni,Hg, Zn,Mo,Pak,Pcb
19	P93DEP B	UVP für geplante Deponie	25	Grü, Wal, Ack, O+G	Pb,Cd,Cr,Co,Cu,Ni, Hg,Zn,As,Fl
20	P93DEP KB	UVP für geplante Deponie	25	Grü, Ack, O+G	Pb,Cd,Cr,Co,Cu,Ni, Hg,Zn,As,Fl
21	P93DEP RO	UVP für geplante Deponie	25	Grü, O+G, Ack, Sie	Pb,Cd,Cr,Co,Cu,Ni, Hg,Zn,As,Fl
22	P93VZW	SM in der Umgebung einer Verzinkerei (1995a; Lehmann, 1995b)	20	Ack, Grü, Son, Sie, Wal	Pb,Cd,Cr,Co,Cu,Ni, Hg,Zn,Fl
23	P93VZE	SM in der Umgebung einer Verzinkerei (1995a; Lehmann, 1995b)	16	Grü, O+G, Ack	Pb,Cd,Cu,Hg,Zn
24	P96PAK	PAK in Umgebung von 2 Bitumenfabriken	15	Ack, Son, Wal	Pak

¹⁾ Ack=Ackerland/Kunstwiese, Grü=Dauergrünland/Naturwiese, Wal=Wald, Sie=Siedlung, O+G=Obst-/Gemüse, Reb=Rebberge, Son=Sonstige

²⁾ einige Standorte doppelt beprobt (→ 1660)

Tabelle 3-2: Untersuchungen zu Bodenschadstoffen im Kanton Thurgau (Fortsetzung)

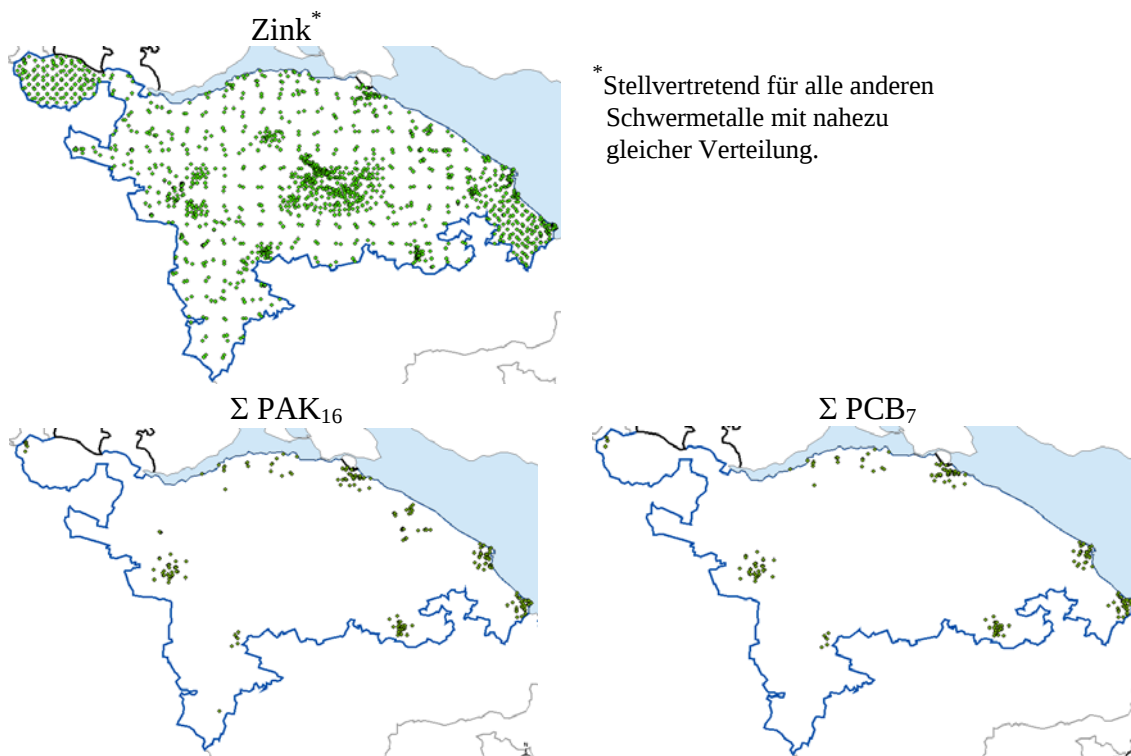
Nr	Projekt	Thema	Stao(n)	Nutzung ¹⁾	relevante Parameter
25	P93VZB-G	SM in der Umgebung einer Verzinkerei	12	Sie, Grü, O+G	Pb,Cd,Cr,Co,Cu,Ni, Zn
26	P01ALTLAST EN	Schiessanlage	10	Grü	Pb
27	P93N1 AW	Sanierung des Lützelburg-Viaduktes (Rutz, 1995)	8	Wal, Grü, Son	Pb,Cd,Cu,Zn
28	P99 BG	Baugesuchprüfung	7	Grü, Sie, Ack	Pb,Cd,Cu,Zn
29	P00 BG	Baugesuchprüfung	6	Grü	Pb,Cd,Cu,Zn
30	P95RAS AUW	Teilprojekt P95RAS	6	Ack, Wal, O+G	Pb,Cd,Cr,Co,Cu,Ni, Hg,Zn,Fl
31	P00WAN	Spielplatzuntersuchung	5	Sie	Pb,Cd,Cu,Hg
32	P01 BG	Baugesuchprüfung	4	Grü, Ack	Pb,Cd,Cu,Zn
33	P97RAS 2	Verdichtung von P94SIE	4	Wal	Pb,Cd,Cr,Co,Cu,Ni, Hg,Zn,Fl
34	P97VZB-P	SM in der Umgebung einer Verzinkerei	2	Wal	Pb,Cd,Cu,Zn
Gesamtanzahl Standorte:			1676²⁾		

¹⁾ Ack=Ackerland/Kunstwiese, Grü=Dauergrünland/Naturwiese, Wal=Wald, Sie=Siedlung, O+G=Obst-/Gemüse, Reb=Rebberge, Son=Sonstige

²⁾ einige Standorte doppelt beprobt (→ 1660)

Relevante Daten

Als relevante Daten für die Fallstudie wurden die in der VBBo berücksichtigten anorganischen und organischen Parameter (Totalgehalte) der Tiefe 0-20 cm ausgewählt. Lediglich Molybdän und Fluor wurden als nicht prioritär bewertet und von vornherein ausgeschlossen. Für die anderen Parameter war die Verteilung der Messwerte über den Kanton ausschlaggebend. Aus Figur 3-2 wird ersichtlich, dass weder die PAK- noch die PCB-Daten für eine räumliche Auswertung genutzt werden konnten. Als Parameter mit einer gut verteilten Datenbasis über das Kantonsgebiet kamen demnach nur die Schwermetalle Pb, Cd, Cr, Cu, Ni, Hg und Zn für weitere Auswertungen in Frage.



Figur 3-2: Verteilung der Schadstoffdaten (Totalgehalte) im Kanton Thurgau

Durch diese Eingrenzung auf die relevanten Daten fielen alle 15 Standorte des Projektes P96PAK (Nr. 24) weg. Ebenfalls nicht berücksichtigt werden konnten 2 weitere Standorte, die ausserhalb der Tiefe 0-20 cm beprobt wurden; ein Standort im Rahmen der Sanierung des Lützelburg Viaduktes (P93N1 AW) und einer bei einer Profiluntersuchung in der Umgebung der Verzinkerei Bettwiesen (P97VZB_P).

Bei Standorten mit Mehrfachbeprobungen wurden die Messwerte analysiert und 51 Messwerte aufgrund von Nachmessungen, abweichenden Labors oder Mittelwertbildungen eliminiert (Anhang A).

Plausibilitätsprüfung

Bis auf wenige Ausnahmen sind die Datensätze des Kanton Thurgau plausibel. Die Wertespans pro Parameter sind in Verbindung mit ihren Messeinheiten sinnvoll und die Werte unterhalb der Bestimmungsgrenze eindeutig identifizierbar (Cadmium 1 Wert und Quecksilber ~5 % der Werte). Sie wurden für alle weiteren Auswertungen auf die halbe Bestimmungsgrenze gesetzt. Die Angabe der Koordinaten lässt ebenfalls keinen Fehler erkennen; weder kamen Duplikate der Standorte vor, noch liegen Standorte ausserhalb der Kantons Grenzen (Figur 3-1).

3.1.2 Grundlagenkarten

Zur Darstellung und Interpretation der punktuellen Messungen bedarf es entsprechender digitaler Flächendaten. Auf Grundlage der in Kap. 2.1 genannten kantonalen und schweizweiten Geodaten wurden die Grundlagenkarten erzeugt, die für die weiteren

geostatistischen Auswertungen von Bedeutung sind (Anhang B). Um die Bodenschadstoffgehalte in späteren Auswertungen auf einen potenziellen Einfluss der Siedlungsgebiete testen zu können, wurden um die grösseren zusammenhängenden Siedlungsflächen ($> 1 \text{ km}^2$) Pufferzonen im Abstand bis 500 m, 500 - 1000 m und mehr als 1000 m Entfernung gebildet (Anhang B- 9). Ebenfalls stellt der pH-Wert eine entscheidende Grösse für das Verhalten von Schwermetallen im Boden dar. Um diesen Faktor für die spätere Interpolation nutzbar zu machen, wurden die Einzelwerte im Untersuchungsgebiet zunächst selber interpoliert. Mit 1589 Messungen liegt nahezu zu allen der 1643 Standorte ein pH-Wert vor. Als Interpolationsverfahren wurde ein Ordinary-Kriging durchgeführt.

Als erklärende Variablen wurden 11 Standorteigenschaften definiert (Tabelle 3-3), die im besten Falle allen 1643 relevanten Standorten im Untersuchungsgebiet zugeordnet sind. Einige Faktoren liessen sich direkt aus der Datenbank entnehmen (wie die Nutzung), andere Faktoren mussten aus den erstellten Grundlagenkarten abgeleitet werden (wie der Bodentyp oder die Höhe).

Tabelle 3-3: Erklärende Variablen der Bodenbelastung

	Variable	Quelle	kont. Variable	kateg. Variable (n-Kategorien)	n-Standorten zugeordnet
Nutzung	Nutzung*	DB	-	6	1643
Siedlung	Siedlungsgebiet*	Kant. Richtplan	-	2	1643
	Distanz zur Siedlung*	Kant. Richtplan	-	4	1643
Geologie	Ausgangsgestein	DB	-	10	387
	Ausgangssubstrat	BÜK	-	29	1269
	Bodenregion*	BÜK	-	9	1269
Bodentyp	Bodengruppe	DB	-	4	14
	Hauptbodentyp*	BÜK	-	12	1219
Bodenart	Bodenart	DB	-	7	500
	Hauptbodenart*	BÜK	-	9	1220
Höhe	Höhe*	DHM	x	-	1643
Hangneigung	Hangneigung*	DHM	-	7	1643
Exposition	Exposition*	DHM	-	8	1643
mittlerer Jahresniederschlag	Niederschlag*	HADES	x	-	1643
pH-Wert	pH*	DB	x	-	1589
org. Kohlenstoff	Corg	DB	x	-	490

*Faktor wird für Regressionsmodell verwendet, Kategorien vgl. Kap. 2.1, (Presler et al., 2005), (FAL, 1997)

Für das Ausgangsgestein, den Bodentyp und die Bodenart liegen nur vereinzelt Informationen in der Datenbank vor, so dass für diese Variablen auf die aus der BÜK abgeleiteten Angaben zurückgegriffen wurde. Limitierender Faktor war der Gehalt an organischem Kohlenstoff, für den keine Karten behelfsmässig hinzugezogen werden können. Da damit für diesen Faktor nur Angaben für weniger als ein Drittel der Daten vorliegen, konnte der Faktor nicht weiter berücksichtigt werden.

Auch wenn sich die Landnutzungen Ackerland und Grünland durchaus deutlich in ihren Schadstoffgehalten unterscheiden können, mussten – aufgrund der fehlenden Differenzierung der Arealstatistik – diese beiden Nutzungsformen für die Erstellung des Regressionsmodells aggregiert werden. Es wurden die in Tabelle 2-2 dargestellten Nutzungstypen verwendet. Ansonsten wäre eine spätere Rückrechnung des Regressionsmodells pro Zelle nicht möglich gewesen.

3.2 Bodenschadstoffkarte Zink

Als relevante Daten für die Fallstudie haben sich die Messwerte für Pb, Cd, Cr, Cu, Ni, Hg und Zn in der Tiefe 0-20 cm an insgesamt 1643 Standorten herausgestellt. Diese Daten stellten den Ausgangsdatensatz für alle weiteren Auswertungen dar. In diesem Kapitel werden die statistischen Eckdaten dieser Ausgangsdaten mittels der allgemeinen statistischen Kenngrößen sowie der statistischen und räumlichen Verteilung aufgezeigt. Alle weiterführenden Auswertungen wurden in erster Linie für den Parameter Zink durchgeführt und anschliessend auf Blei übertragen.

Pro Parameter lagen zwischen 980 und 1634 Datensätze verteilt über den Kanton Thurgau vor (Tabelle 3-4). Nicht an jedem Standort wurden alle Parameter bestimmt. Im Mittel wurden bei den meisten Parametern die Richtwerte der Verordnung über Belastungen des Bodens (VBBo, 1998) deutlich unterschritten (Tabelle 3-5). Ausnahme waren einige extrem hohe Werte für Blei und Kupfer. Deren Median lag allerdings ebenfalls merklich tiefer, was auf rechtsschiefe Verteilungen hindeutet. Die Maximalwerte lagen für alle Parameter klar oberhalb der Richtwerte, im Falle von Cadmium und Kupfer wurde darüber hinaus der Prüfwert überschritten und bei Blei lagen Werte oberhalb des Sanierungswertes vor.

Tabelle 3-4: Allgemeine Statistische Kenngrößen Ausgangsdatensatz (in mg/kg)

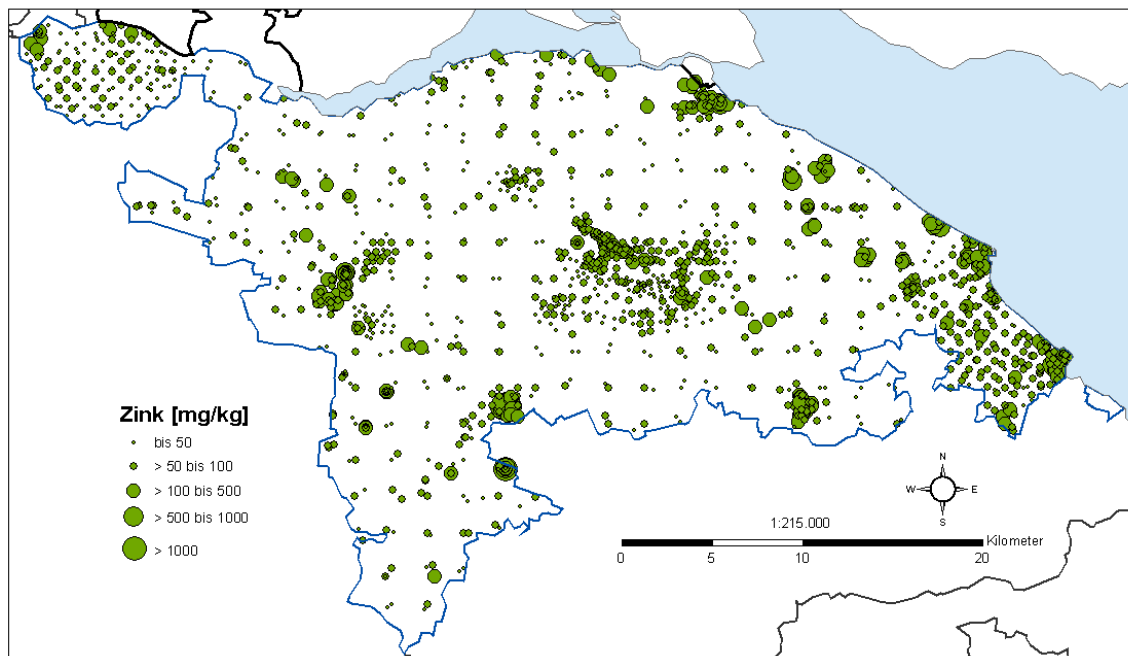
	Pb	Cd	Cr	Cu	Ni	Hg	Zn
Anzahl Datensätze	1634	1553	980	1544	980	1262	1549
Minimum	6.70	0.03	1.97	3.98	4.70	0.01	13.60
Maximum	6207.00	4.54	262.00	538.00	130.00	4.18	1985.00
Median	24.90	0.26	24.60	21.60	21.70	0.07	58.10
Mittelwert	92.92	0.32	26.28	42.36	23.32	0.11	101.44
Standardabweichung	342.54	0.25	13.86	71.40	9.00	0.22	172.91
Variationskoeffizient	3.69	0.80	0.53	1.69	0.39	2.05	1.70
Schiefe	10.47	6.01	10.80	4.33	3.30	13.04	5.88
Exzess	139.75	66.83	171.81	20.48	25.26	212.44	40.52
25. Perzentil	19.80	0.20	20.50	15.30	18.20	0.05	44.00
75. Perzentil	41.76	0.35	29.00	33.65	26.35	0.10	82.50
95. Perzentil	287.88	0.70	41.25	144.60	38.38	0.21	283.35
99. Perzentil	1590.05	1.33	56.45	440.06	55.38	0.76	1160.00

Tabelle 3-5: Beurteilungswerte (Totalgehalte) der Verordnung über Belastungen des Bodens (VBBö, 1998) (in mg/kg)

	Pb	Cd	Cr	Cu	Ni	Hg	Zn
Richtwert	50	0.8	50	40	50	0.5	150
Prüfwert	200	2	-	150	-	-	-
Sanierungswert	1000-2000*	20-30	-	1000	-	-	2000

*je nach Nutzung

Für die betrachteten anorganischen Parameter lagen die Daten lückenlos über das Kantonsgebiet vor, wie bereits in Figur 3-2 im Zuge der Auswahl der relevanten Daten dargestellt wurde. Aus Figur 3-3 lässt sich zudem am Beispiel von Zink erkennen, dass sich höhere Gehalte meist geballt auf einem Haufen befanden, eher im dichter besiedelten Raum als im ländlichen Raum (vgl. Anhang B- 9). Gleiches gilt für die anderen Schwermetalle.

**Figur 3-3: Zinkgehalte an den Untersuchungsstandorten Ausgangsdatensatz**

3.2.1 Datensatz 673

Wie aus Tabelle 3-6 hervorgeht, blieben nach Ausschluss der Projekte mit einem gezielten Belastungsverdacht noch 12 der 34 Projekte übrig – etwa die Hälfte der ursprünglichen 1643 Standorte. Wurden zudem nur noch Standorte mit der Nutzung Acker, Grünland oder Wald berücksichtigt, erhielt man einen Datensatz mit Messwerten zu 673 Standorten (vgl. Tabelle 3-2), der die Forderungen nach Homogenität und Stationarität der Daten erfüllte.

Tabelle 3-6: Untersuchungskampagnen Datensatz 673

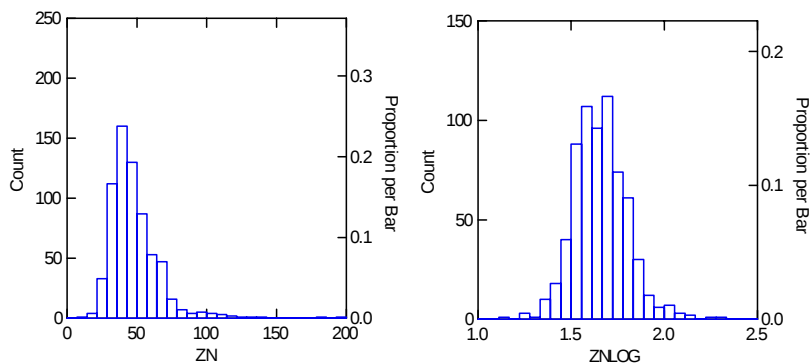
Nr	Projekt	Thema	Stao(n) (Ack, Grü, Wal)
1	P95RAS	Schwermetalle in Land- und Forstwirtschaft	318
3	P93LW DI	Schwermetalle in intensiven Landwirtschaftsbezirken	111
4	P92KVAW	Untersuchung vor dem Bau einer KVA	81
5	P93LW OT	Schwermetalle in intensiven Landwirtschaftsbezirken	45
10	P92KVAW G	Untersuchung vor dem Bau einer KVA	37
19	P93DEP B	UVP für geplante Deponie	22
20	P93DEP KB	UVP für geplante Deponie	21
21	P93DEP RO	UVP für geplante Deponie	18
28	P99 BG	Baugesuchprüfung	5
29	P00 BG	Baugesuchprüfung	6
30	P95RAS AUW	Teilprojekt P95RAS	5
32	P01 BG	Baugesuchprüfung	4
Summe			673

3.2.1.1 Explorative Datenanalyse

Für jeden der verbleibenden 673 Standorte lagen Zinkwerte vor (Tabelle 3-7). Die Reduktion der Standorte und somit auch der Daten bedeutete eine deutlich veränderte Werteverteilung, wobei erwartungsgemäss in erster Linie der Mittelwert stark herabgesetzt wurde (vgl. Tabelle 3-4). Dieser wich nun nicht mehr so stark vom Median ab und selbst das 99. Perzentil lag mittlerweile unterhalb des Richtwerts. Auch wenn Median und Mittelwert dichter beieinander lagen und die Schiefe stark abgenommen hatte, lässt sich in Figur 3-4 erkennen, dass die Daten noch nicht annähernd normalverteilt waren. Eine log-Transformation führte zu einer deutlichen Abnahme der Schiefe, die log-Zinkgehalte waren annähernd normalverteilt (Figur 3-4).

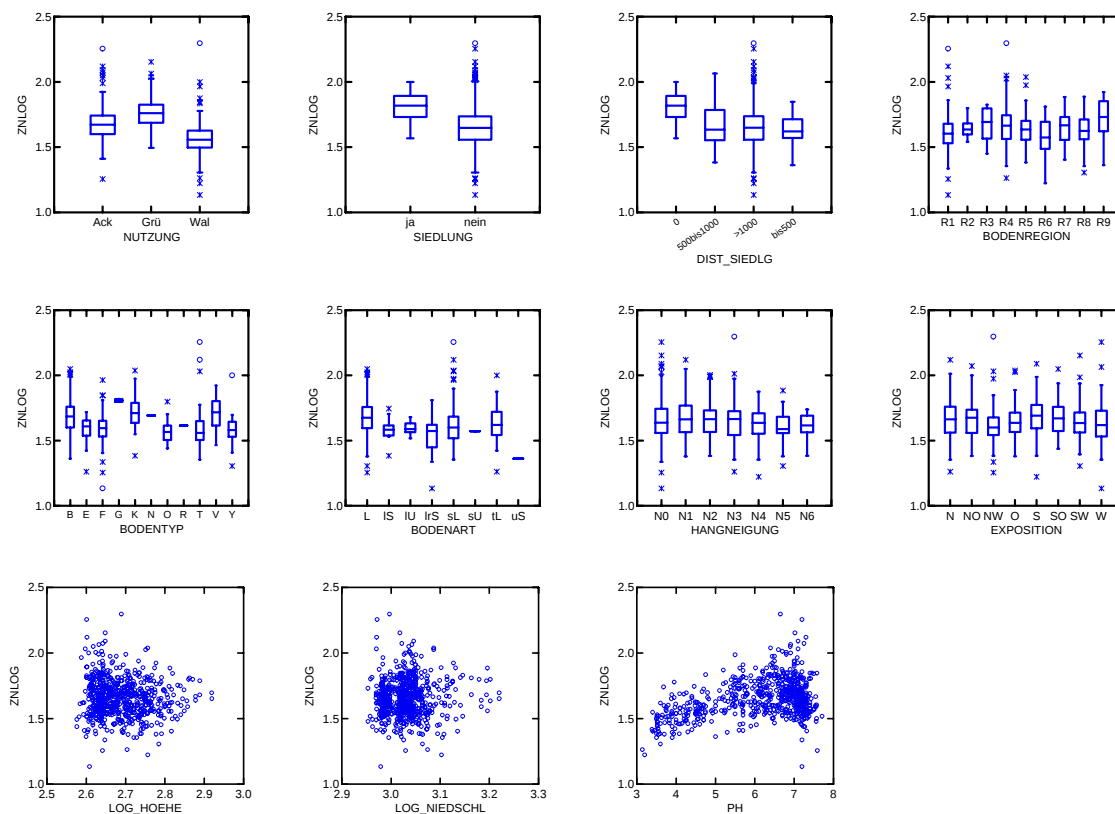
Tabelle 3-7: Allgemeine statistische Kenngrössen der Zinkgehalte Datensatz 673 (in mg/kg)

	Zn*
Anzahl Datensätze	673
Minimum	13.60
Maximum	198.00
Median	44.50
Mittelwert	47.96
Standardabweichung	18.23
Variationskoeffizient	0.38
Schiefe	2.56
Exzess	13.04
25. Perzentil	36.10
75. Perzentil	55.00
95. Perzentil	77.17
99. Perzentil	115.1
	Zn(log)*
Schiefe	0.45
Exzess	1.41



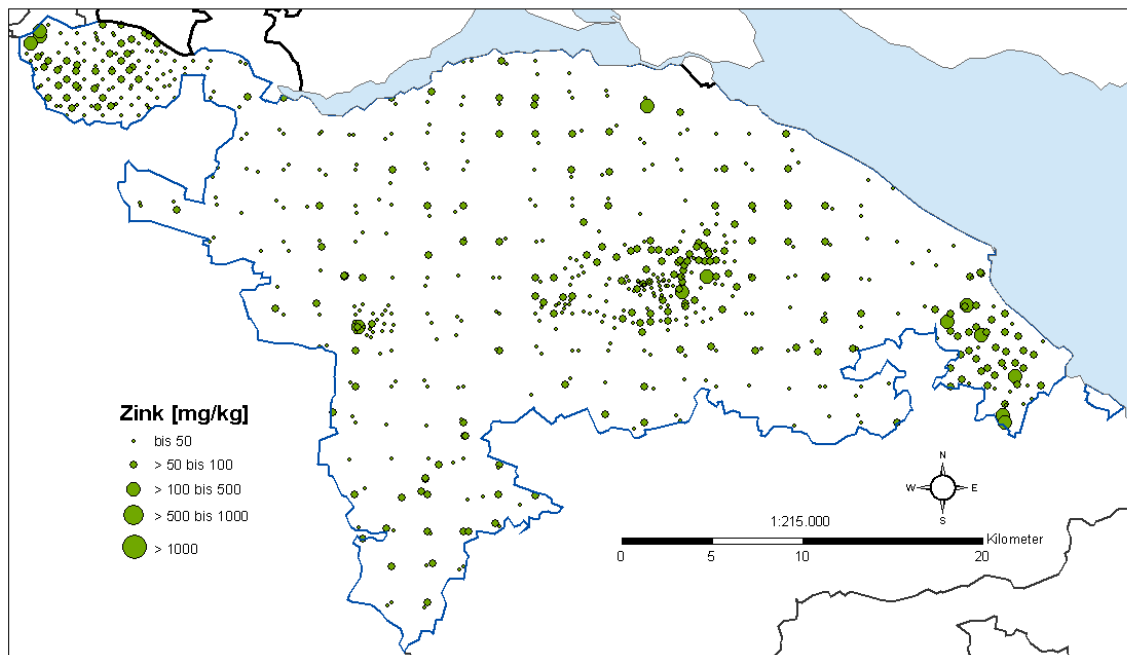
Figur 3-4: Verteilung der Zinkmesswerte Datensatz 673 (links Zn-, rechts logZn-Werte in mg/kg)

Die univariate Statistik diente der Erkenntnis derjenigen erklärenden Variablen, die für die Regression bedeutsam werden können. So zeigt Figur 3-5, dass die logarithmierten Zinkgehalte auf Grünland signifikant höher waren als diejenigen im Wald. Gleiches galt für Messwerte im Siedlungsgebiet, die klar oberhalb der Werte des Aussenbereichs lagen. Die Entfernung zur Siedlung zeigte dabei keinen merklichen Effekt. Ebenfalls noch Unterschiede in den Zinkgehalten zeigten sich je Bodentyp und pH-Wert. Je weiter der pH-Wert absank, desto geringer waren die Zinkgehalte. Zurückzuführen ist dieser Effekt vermutlich auf die bei einem Grenz-pH-Wert von 5,5 bis 6 bereits zunehmende Mobilisierung von Zink (Scheffer & Schachtschabel, 2002). Diese wiederum könnte denn auch die Ursache für die geringeren Gehalte im Wald sein, dessen Schadstoffimmissionen aufgrund des Auskämmeffektes aus der Luft ungleich höher sein sollten.



Figur 3-5: Univariate Statistik der erklärenden Variablen auf die Zink-Bodenbelastung Datensatz 673, logZn in mg/kg

Figur 3-6 zeigt die Verteilung der Messpunkte mit entsprechenden Zinkgehalten. Der Vergleich mit Figur 3-3 bestätigt, dass ein Grossteil der hohen Zinkgehalte weder in Acker-/Grünland- noch in Waldböden gemessen wurden. Relativ hohe Gehalte fanden sich für diesen Datensatz in der Region Weinfelden, sowie im Bezirk Diessenhofen und Arbon.



Figur 3-6: Zinkgehalte an den Untersuchungsstandorten Datensatz 673

3.2.1.2 Räumliche Datenanalyse

Regressionsanalyse

Multivariat ausgewertet mit der in Kap. 2.5.2 beschriebenen robusten linearen Regressionsfunktion ergaben sich 2 Modelle, die für die räumliche Schätzung eingesetzt wurden. Zur Ermittlung dieser wurden zuvor zahlreiche Modelle berechnet, in denen schrittweise je nach Bestimmtheitsmass erklärende Variablen einbezogen und gegebenenfalls wieder eliminiert wurden. Eines der beiden aus den vielen Berechnungen als sinnvoll eingestuftes Modell (Modell 1) bezog die erklärenden Variablen Nutzung, pH-Wert und Niederschlag mit ein. Für die Nutzung zeigte sich – wie aus der univariaten Statistik erwartet werden konnte – ein negativer Koeffizient für Wald im Vergleich zu Acker/Grünland (Tabelle 3-8). Die beiden kontinuierlichen Variablen pH-Wert und Niederschlag besaßen einen positiven Koeffizienten, d.h. je höher der pH-Wert und der Niederschlag, desto höher die Zinkgehalte. Auffällig war die Höhe des Koeffizienten für Niederschlag. Dieser war sowohl aus fachlicher Sicht wie auch aufgrund der univariaten Statistik nicht zu erwarten und musste daher auf Wechselwirkungen mit anderen Variablen zurückgeführt werden.

Ein weiteres Modell (Modell 2) bezog zusätzlich den Bodentyp und die Bodenart mit ein. Die Koeffizienten der bisherigen Variablen veränderten sich kaum. Bis auf den Bodentyp „Fluvisol“ wurde für alle Bodentypen im Vergleich zum Regosol ein leicht positiver Koeffizient vorgegeben; deutlich höher lediglich für die „Saure Braunerde“, der ähnlich wie der Koeffizient für Niederschlag fachlich nicht logisch erschien. Den verschiedenen Bodenarten wurden ausschliesslich geringe Koeffizienten zugeordnet.

Durch die Einbeziehung dieser Variablen wurden die Freiheitsgrade deutlich reduziert, da die Bodenübersichtskarte nicht das komplette Kantonsgebiet abdeckt und somit die Information nicht allen Punkten zugeordnet werden konnte. Auf der anderen Seite stieg allerdings das Bestimmtheitsmass, d.h. ein grösserer Anteil der Streuungen der Zinkwerte konnte erklärt werden. Im Modell 1 betrug dies rund 37 % der Streuungen, im Modell 2 etwa 42 %. Der Faktor Siedlung oder Siedlungsnähe konnte für diesen Datensatz nicht berücksichtigt werden, da keine repräsentative Anzahl an Standorten im Siedlungsgebiet vorhanden war.

$$\begin{aligned} \text{Modell 1: } \log_{10}(Zn) = & \alpha + \beta_1 * \text{Nutzung} + \beta_2 * pH + \\ & \beta_3 * \log_{10}(\text{Niederschlag}) + E_i \end{aligned} \quad (3.1)$$

$$\begin{aligned} \text{Modell 2: } \log_{10}(Zn) = & \alpha + \beta_1 * \text{Nutzung} + \beta_2 * pH + \\ & \beta_3 * \log_{10}(\text{Niederschlag}) + \\ & \beta_4 * \text{Bodentyp} + \beta_5 * \text{Bodenart} + E_i \end{aligned} \quad (3.2)$$

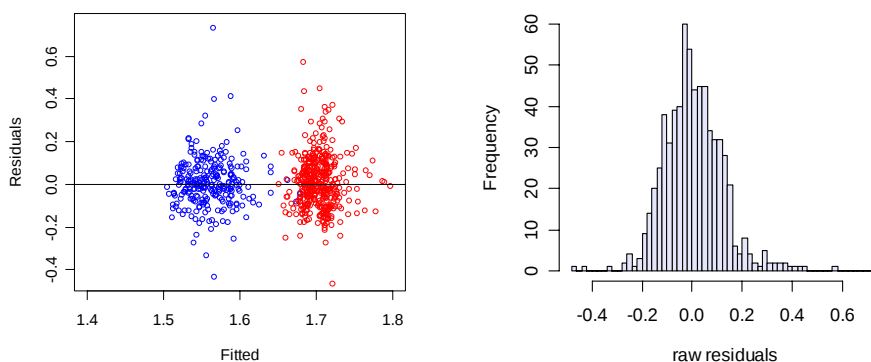
Tabelle 3-8: Regressionsmodelle für logarithmierte Zinkgehalte Datensatz 673

Koeffizient	Kategorie	Zn673, Modell 1	Zn673, Modell 2*
α		-0.067	0.106
β_1 Nutzung	AckGrü	0	0
	Wal	-0.120	-0.098
β_2 pH		0.019	0.029
$\beta_3 \log_{10}(\text{Niederschlag})$		0.542	0.457
β_4 Bodentyp	Regosol		0
	Fluvisol		-0.067
	Rendzina		-
	Kalkbraunerde		0.048
	Braunerde		0.027
	Parabraunerde		0.006
	Saure Braunerde		0.244
	Braunerde-Pseudogley		0.029
	Braunerde-Gley		-
	Buntgley		-
	Fahlgley		-
	Halbmoor		-
β_5 Bodenart	L		0
	IU		-
	IrS		-0.051
	sL		-0.009
	tL		0.031
	IS		-
	IT		-
	sU		-
	uS		-
Anzahl Freiheitsgrade		662	534
adjusted R ²		0.366	0.422

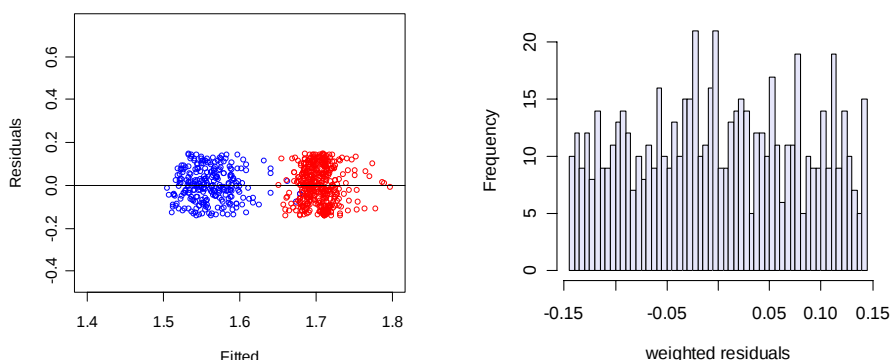
* einige Kategorien waren entweder nicht vorhanden oder Anzahl Zinkwerte pro Kategorie <15

Die Gewichtung der Residuen zur robusten Schätzung der Regressionskoeffizienten kann mit der Tuning-Konstante mehr oder weniger strikt (konservativ) vorgenommen werden. Im Folgenden soll der Unterschied zwischen den rohen und den gewichteten Residuen und deren Effekt auf die Interpolation dargestellt werden. Für das Modell 1 wurden daher beide Formen der Residuen betrachtet, für Modell 2 anschliessend nur noch die gewichteten.

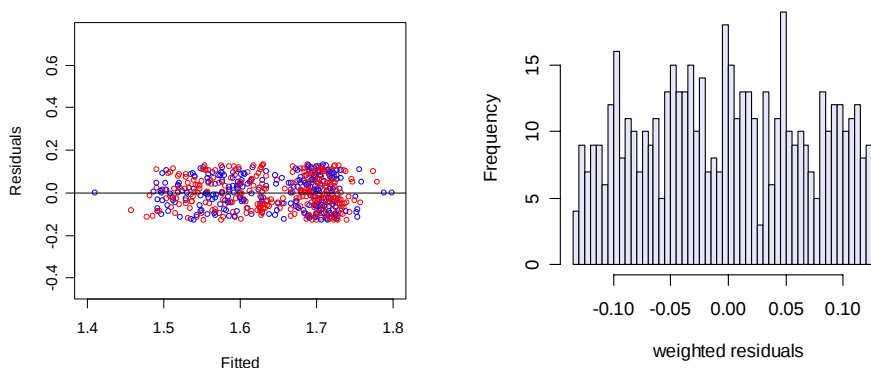
Sowohl die rohen als auch die gewichteten Residuen der beiden Modelle, aufgetragen gegen die angepassten Werte, schwankten symmetrisch um die Nulllinie. Es gab nur wenige auffällige Werte bei den rohen Residuen (Figur 3-7 bis Figur 3-9). Im Modell 1 fiel eine deutliche Segregation der Residuen auf, was auf die Nutzung zurückzuführen ist. Mit der Hinzunahme der erklärenden Variablen der BÜK (Modell 2) verwischte sich diese Zweiteilung. Während die rohen Residuen des Modells 1 annähernd normalverteilt waren, sorgte die Gewichtung der Residuen für eine geringere Streubreite der Residuen (-0.15 bis 0.15) und eine relativ gleichmässige Verteilung pro Werteklasse, was ebenso für das Modell 2 galt.



Figur 3-7: Streudiagramm der rohen Residuen gegen die angepassten Werte (Fitted) mit Nutzung Acker/Grünland (rot) und Wald (blau) (links), Histogramm der rohen Residuen (rechts), Zn673 Modell 1, Angaben in mg/kg



Figur 3-8: Streudiagramm der gewichteten Residuen gegen die angepassten Werte (Fitted) mit Nutzung Acker/Grünland (rot) und Wald (blau) (links), Histogramm der rohen Residuen (rechts), Zn673 Modell 1, Angaben in mg/kg



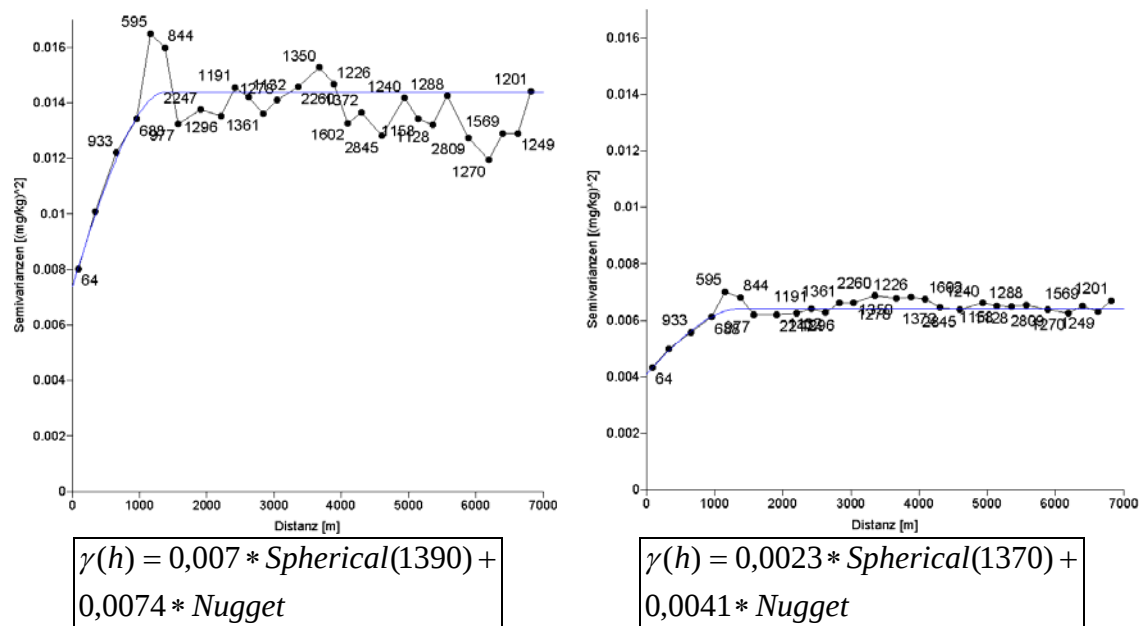
Figur 3-9: Streudiagramm der gewichteten Residuen gegen die angepassten Werte (Fitted) mit Nutzung Acker/Grünland (rot) und Wald (blau) (links), Histogramm der rohen Residuen (rechts), Zn673 Modell 2, Angaben in mg/kg

Variographie

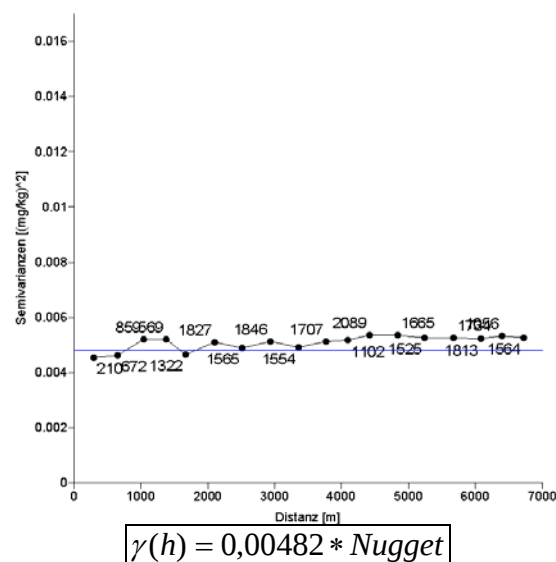
In der anschliessenden Variogrammanalyse zeigten die aus Modell 1 resultierenden Residualwerte auf log-Basis noch deutliche räumliche Abhängigkeiten (d.h. der Trend in den Daten konnte mit Hilfe des Regressionsmodells nicht gänzlich eliminiert werden). Die Residuen liessen sich am besten an ein sphärisches Modell anpassen. Die Reichweite der rohen und gewichteten Residuen war mit rund 1400 m nahezu gleich, während

sich Schwellenwert und Nugget deutlich unterschieden. Die rohen Residuen zeigten erwartungsgemäss höhere mittlere Semivarianzen (Figur 3-10).

Das Modell 2 unter Einbeziehung des Bodentyps und der Bodenart liess keine räumlichen Abhängigkeiten für die gewichteten Residuen mehr erkennen (Figur 3-11). Entsprechend resultierte ein Pure-Nugget-Modell mit einem Nuggeteffekt von $0,00482 \text{ mg}^2/\text{kg}^2$, der im Bereich der mittleren Semivarianzen der gewichteten Residuen von Modell 1 lag.



Figur 3-10: Semivariogramm der rohen Residuen Zn673 Modell 1 (links) und der gewichteten Residuen Zn673 Modell 1 (rechts)



Figur 3-11: Semivariogramm der gewichteten Residuen Zn673 Modell 2

Interpolation

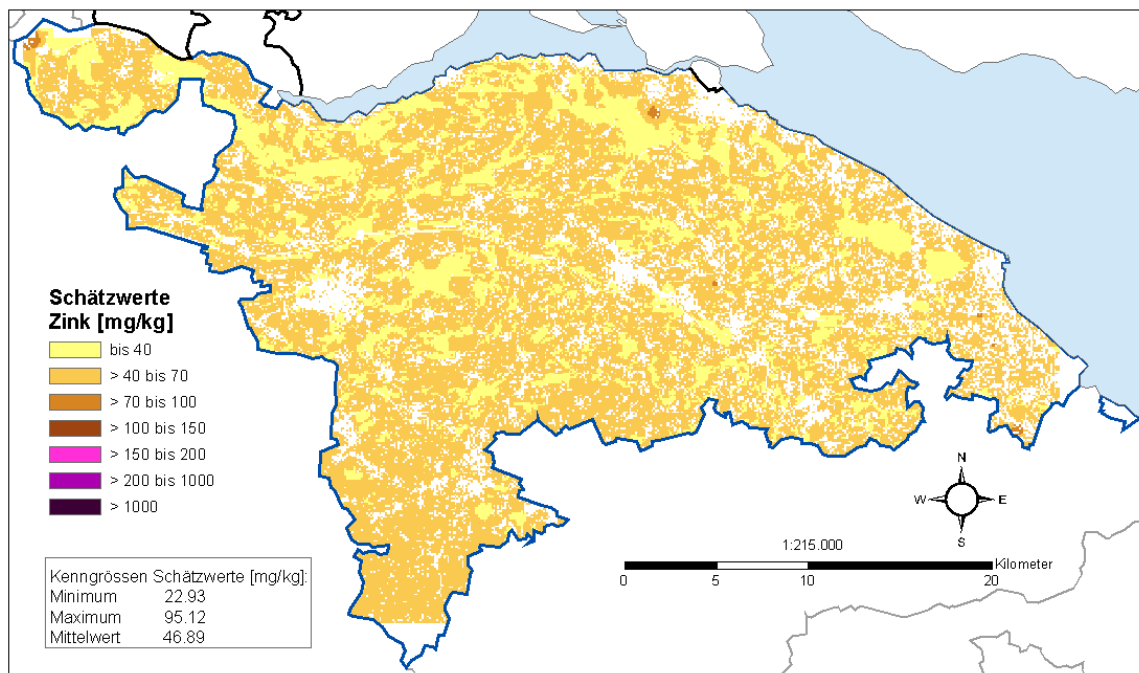
Für das Modell 1 in Verbindung mit der Anpassung der Semivarianzen an ein sphärisches Variogrammodell wurden die Residuen mit der Simple Kriging Methode interpoliert. Damit wurde ein Kriging mit einem konstanten, bekannten Erwartungswert durchgeführt (vgl. Kap. 2.5.2). Im Falle der Residuen von Modell 2 wurde ein Ordinary Kriging verwendet, da diese durch ein Pure-Nugget-Variogramm angepasst werden und von einem konstanten, aber nicht bekannten Erwartungswert ausgegangen wird. In beiden Fällen wurden mindestens 12 und maximal 24 Stützpunkte berücksichtigt.

Im Mittel wiesen die Schätzkarten für die 3 Varianten fast identische Zinkgehalte auf, die Streuung der Schätzwerte war allerdings deutlich verschieden (Figur 3-12 bis Figur 3-14). Die geringste Streuung zeigte die Karte der geschätzten Zinkgehalte auf Basis der gewichteten Residuen ohne die Parameter der BÜK (Modell 1). Nahm man für dieses Modell die rohen Residuen wurden Maximalwerte bis 95 mg/kg geschätzt, anstelle von zuvor 63 mg/kg. Modell 2 führte mit dem zahlenmässig reduzierten Datensatz und der Verwendung eines Ordinary Kriging zu ebenfalls breiter streuenden Schätzwerten als die gewichteten Residuen von Modell 1. Ein möglicher Grund liegt in der Wahl der Kriging-Methode. Durch die Verwendung eines Simple Kriging für Modell 1 wurde ein bekannter Erwartungswert angenommen, der die Streubreite der Daten minimiert. Grundsätzlich wurden die Messwerte (vgl. Tabelle 3-7) in allen 3 Karten durch die robuste Erstellung des Regressionsmodells deutlich geglättet.

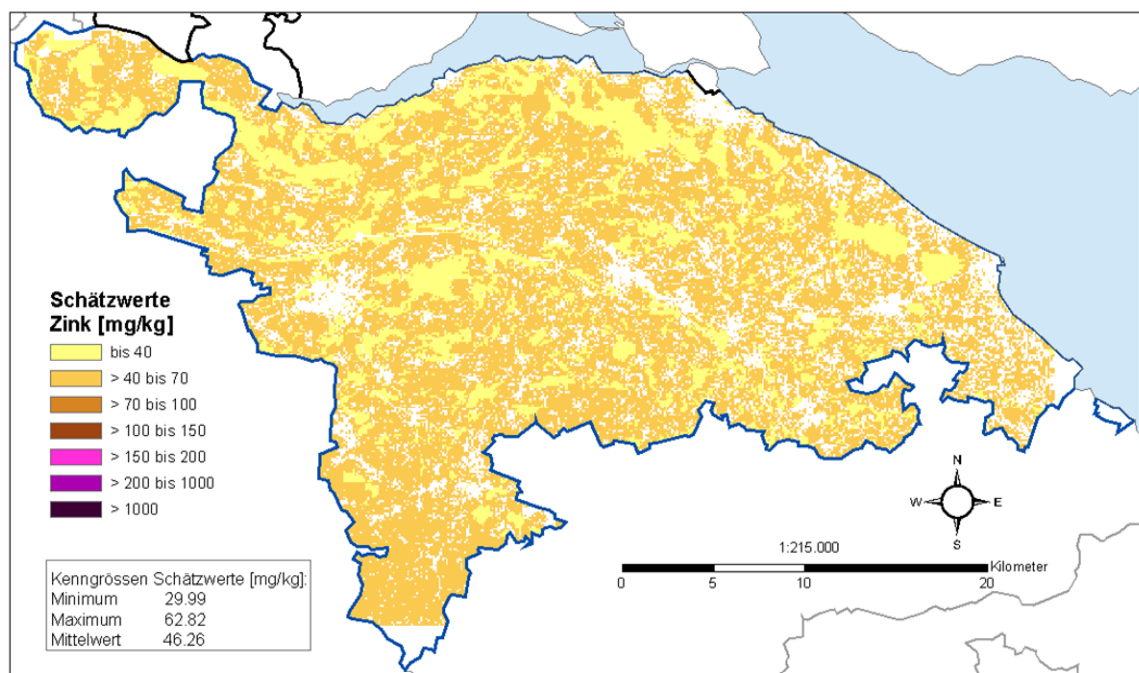
Das räumliche Muster der interpolierten Zinkgehalte war geprägt durch die in der Regression verwendeten erklärenden Variablen bzw. deren Regressionskoeffizienten. Daher war das Muster der gewichteten und der rohen Residuen von Modell 1 nahezu identisch. In erster Linie prägte die Landnutzung das Muster dieser Karten. Die Bereiche mit geschätzten Zinkgehalten von < 40 mg/kg waren zum Grossteil der Nutzung Wald zuzuordnen. Höhere Gehalte im Wald zeigten sich lediglich in Bereichen, in denen die mittleren Jahresniederschläge hoch sind (v. a. im höher gelegenen Süden des Kantons). Dort wurde der Nutzungseinfluss durch den im Modell hohen Einfluss des Niederschlags überdeckt. Der pH-Wert spielte durch den geringen Koeffizienten erwartungsgemäss keine Rolle. Gehalte von > 70 mg/kg wurden lediglich für wenige Stellen, u. a. in den Gemeinden Tägerwil und Schlatt (Figur 2-2) vorhergesagt.

Die mit den Residuen von Modell 2 interpolierten Zinkgehalte zeigten teilweise ein ähnliches Muster. Allerdings zeigte die Karte entsprechend der Regressionskoeffizienten für den Bodentyp dort, wo als Bodentyp eine „Saure Braunerde“ vorherrscht, höhere Schätzwerte und im Tal der Thur, wo der „Fluvisol“ überwiegt, geringere Gehalte.

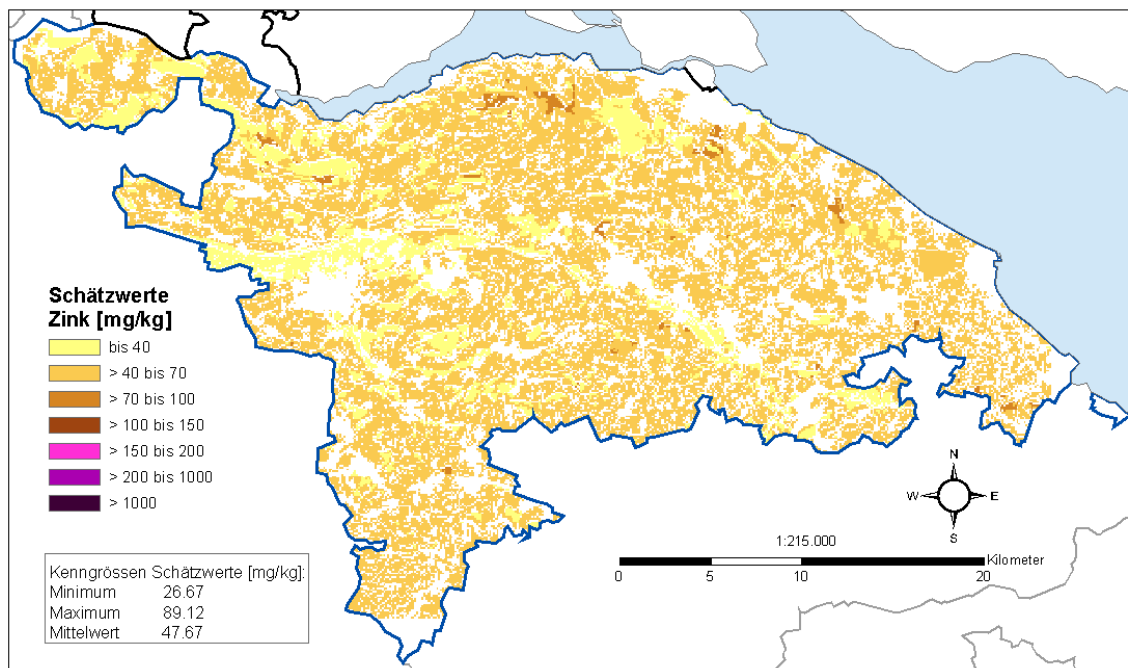
Auf die Darstellung der Stützpunkte wurde in diesen und den folgenden Karten verzichtet, da die Struktur der Schätzergebnisse schlecht ersichtlich wäre.



Figur 3-12: Interpolierte Zinkgehalte auf Basis der rohen Residuen Zn673 Modell 1, interpoliert mit Simple Kriging



Figur 3-13: Interpolierte Zinkgehalte auf Basis der gewichteten Residuen Zn673 Modell 1, interpoliert mit Simple Kriging



Figur 3-14: Interpolierte Zinkgehalte auf Basis der gewichteten Residuen Zn673 Modell 2, interpoliert mit Ordinary Kriging

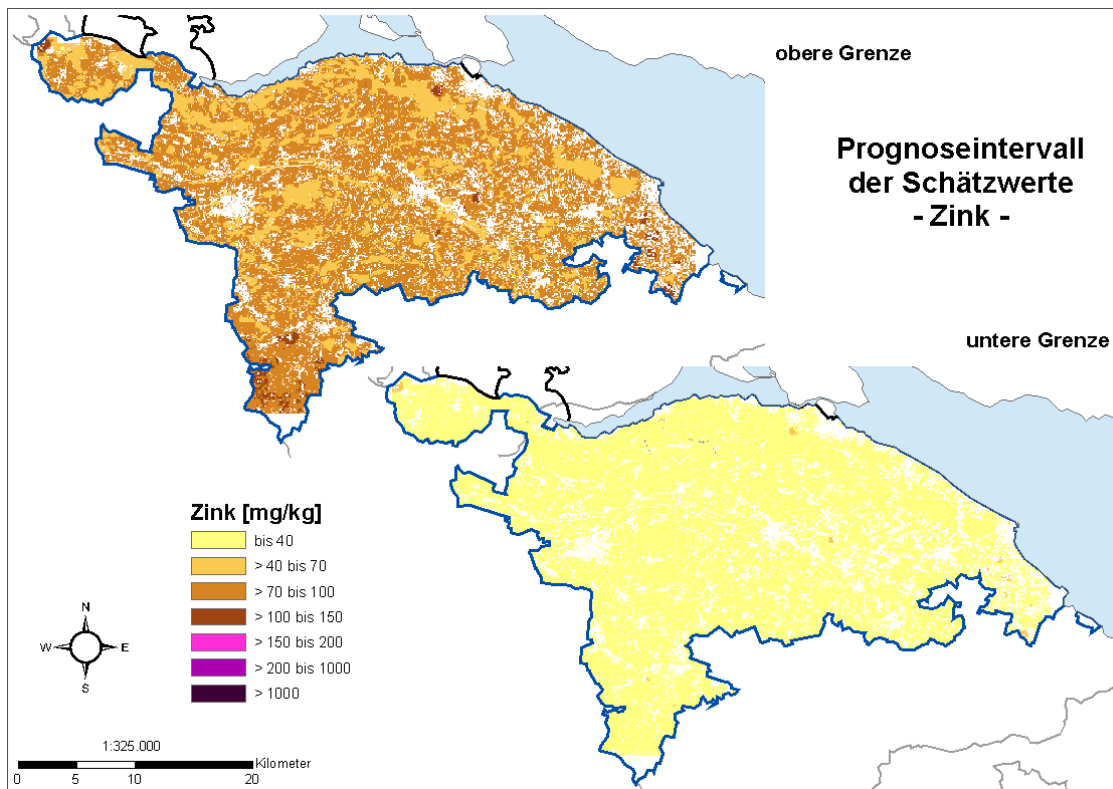
Prognoseintervall und Kreuzvalidierung

Jede Schätzung ist mit einem Fehler behaftet. Da die Krigingvarianzen auf log-Residualbasis wenig aussagekräftig sind, wurden die Prognoseintervalle, die in Form der rücktransformierten Schätzwerte dargestellt werden können, ausgewiesen. Sie zeigen an, in welchem Bereich der Erwartungswert mit einer Wahrscheinlichkeit von 95 % liegt. In den Karten aufgetragen wurden die obere und untere Grenze des Intervalls.

Für das Modell 1 mit Verwendung der rohen Residuen bedeutete dies, dass die von 23 mg/kg bis 95 mg/kg geschätzten Zinkgehalte mit einer 95 %igen Wahrscheinlichkeit im Bereich von 14 mg/kg bis 154 mg/kg lagen (Figur 3-15). Die Intervalle waren aufgrund der Verwendung der log-Daten für die Interpolation nicht symmetrisch. Der Richtwert der VBBo (1998) von 150 mg/kg Zink wurde mit dem 95 % Intervall nur in 3 einzelnen Schätzzellen überschritten (Gemeinde Tägerwil). Ein Grossteil des Untersuchungsgebietes wies aber mit dieser Interpolationsmethode ein Prognoseintervall von kleiner 100 mg/kg Zink im Boden auf.

Für die Interpolation mit den gewichteten Residuen, sowohl für Modell 1 als auch für Modell 2, waren die Vorhersageintervalle enger. Die Schätzung der gewichteten Residuen wurde damit als genauer prognostiziert, was auf den höheren Homogenisierungsgrad der Residuen durch Wahl der robusteren Kriterien zurückzuführen ist (Anhang C-1, Anhang C-2). Das Prognoseintervall mit 95 %iger Wahrscheinlichkeit erreichte den

Richtwert nicht mehr annähernd. Die maximale obere Grenze lag im Modell 1 bei 89 mg/kg und im Modell 2 bei 123 mg/kg.



Figur 3-15: Prognoseintervall des zweiseitigen 95 %-Quantils der interpolierten Zinkgehalte der Simple-Kriging-Vorhersagefehler, rohe Residuen Zn673 Modell 1

Die Ergebnisse der Kreuzvalidierung sind in Tabelle 3-9 dargestellt. Die sehr kleinen absoluten Fehler sind auf die log-Residualwerte als Ausgangsgrößen zurückzuführen.

Bestes Ergebnis zeigte das Modell 2 mit den gewichteten Residuen. Die Variabilität der Vorhersage wurde am genauesten geschätzt und der MAD zeigte die geringsten absoluten Fehler im Vergleich zu den beiden anderen Validierungen. Modell 1 mit den rohen Residuen zeigte den schlechtesten MAD, dafür wurde bei diesem Modell mit den gewichteten Residuen die Variabilität in der Vorhersage etwas überschätzt.

Tabelle 3-9: Kenngrößen der Kreuzvalidierung für den Datensatz Zn673 (in log mg/kg)

Kenngrößen Kreuzvalidierung	Zn673, Modell 1* (rohe Residuen)	Zn673, Modell 1 (gewichtete Residuen)	Zn673, Modell 2 (gewichtete Residuen)
ME	-0.003044	-0.002253	-0.0008494
RMSSE	1.005	0.974	1.002
MedE	0.0027	0.0007	0.0028
MAD	0.0654	0.0586	0.0538

Zusammenfassende Diskussion

Die genaueste Karte der Zinkgehalte für die vermutet unbelasteten 673 Standorte auf Acker, Grünland und Wald wurde mit Einbezug der Bodenparameter der BÜK erzielt, auch wenn dadurch die Anzahl der Stützpunkte von 668 auf 549 herabgesenkt wurde. Dafür sprach zum einen die räumliche Unabhängigkeit der Residuen des Modell 2, die darauf hinwies, dass mit dem Regressionsmodell ein erheblicher Anteil der räumlichen Streuung der Zinkgehalte erklärt wurde. Der verbleibende stochastische Anteil lieferte dazu durch die Verwendung eines Ordinary Kriging in Verbindung mit dem Pure-Nugget-Modell die Schätzung mit dem engsten Prognoseintervall und führte letztlich zu der in der Kreuzvalidierung aufgezeigten besten Vorhersage. Darüber hinaus zeigte das räumliche Muster der Ergebniskarten, dass durch Einbezug der Variablen Bodentyp und Bodenart ins Regressionsmodell die geschätzten Zinkgehalte räumlich noch stärker differenziert werden konnten. Dies befürwortet eine sorgfältige und umfassende Ausarbeitung von Grundlagenkarten, die in die räumlichen Analysen der Zielvariablen mit einbezogen werden.

Die Verwendung von gewichteten Residuen eines Regressionsmodells führte durch die Homogenisierung der Residuen zu geringeren Semivarianzen ($1/3$ der rohen Residuen) und damit zu einer deutlichen Glättung der Schätzwerte mit entsprechend engerem Prognoseintervall. In der Kreuzvalidierung führte dies zu merklich geringeren Abweichungen der Schätzung vom tatsächlichen Messwert und damit zu einer besseren Voraussage. Allerdings wurde dabei die Variabilität der Vorhersage etwas unterschätzt.

3.2.2 Datensatz 793

Der bislang strikt reduzierte Datensatz wurde in einer ersten Erweiterung um 120 Datensätze ausgebaut. Dazu zählten zwei Untersuchungen im Umfeld von Kehrichtverbrennungsanlagen (KVAs) und drei im Umfeld von Verzinkereien, auch diesmal allerdings wieder beschränkt auf die Nutzungen Acker, Grünland und Wald. Ebenfalls mit einbezogen wurden 9 Standorte aus 2 Belastungsuntersuchungen (Projekt P94ARA MT, P93N1 AW), die als Referenzstandorte gedient haben und somit als unbelastete zum Zeitpunkt der Probenahme angesehen werden konnten. Für die weiteren Auswertungen wurden damit Daten von 793 Standorten verwendet (Tabelle 3-10).

Tabelle 3-10: Untersuchungskampagnen Datensatz 793

Nr	Projekt	Thema	Stao(n) (Ack, Grü, Wal)
1	P95RAS	SM in Land- und Forstwirtschaft	318
3	P93LW DI	SM in intens. Landwirtschaftsbezirken	111
4	P92KVAW	Untersuchung vor KVA	81
5	P93LW OT	SM in intens. Landwirtschaftsbezirken	45
10	P92KVAW G	Untersuchung vor KVA	37
14	P94ARA MT	SM in der Umgebung einer ARA	2
15	P93VZB	SM in der Umgebung einer Verzinkerei	32
16	P92KVAH	SM in der Umgebung einer KVA	23
17	P92KVAM	SM in der Umgebung einer KVA	28
19	P93DEP B	UVP für geplante Deponie	22
20	P93DEP KB	UVP für geplante Deponie	21
21	P93DEP RO	UVP für geplante Deponie	18
22	P93VZW	SM in der Umgebung einer Verzinkerei	17
23	P92VZE	SM in der Umgebung einer Verzinkerei	11
27	P93N1 AW	Sanierung des Lützelburg-Viaduktes	7
28	P99 BG	Baugesuchprüfung	5
29	P00 BG	Baugesuchprüfung	6
30	P95RAS AUW	Teilprojekt P95RAS	5
32	P01 BG	Baugesuchprüfung	4
Summe			793

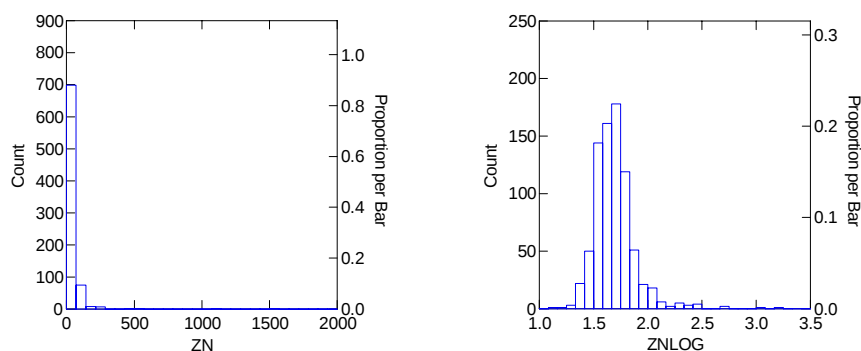
orange = zusätzliche Projekte

3.2.2.1 Explorative Datenanalyse

Für alle neu ausgewählten Standorte lagen Zinkmesswerte vor (Tabelle 3-11). Die Erweiterung des Datensatzes führte erwartungsgemäss zu einer merklichen Anhebung der mittleren Zinkgehalte. Ebenfalls lag das 99. Perzentil wieder klar oberhalb des Richtwertes. Gleichzeitig nahm allerdings wieder durch die Hinzunahme der hohen Zinkgehalte die Rechtsschiefe der Daten zu, was sich durch Logarithmierung deutlich verbessern liess (Figur 3-16) und als annähernd normalverteilt betrachtet wurde, wenn auch die Schiefe noch immer bei > 2 lag. Im Zuge der räumlichen Datenanalyse wurden die für die Schiefe verantwortlichen hohen Zinkgehalte genauer betrachtet und gegebenenfalls eliminiert.

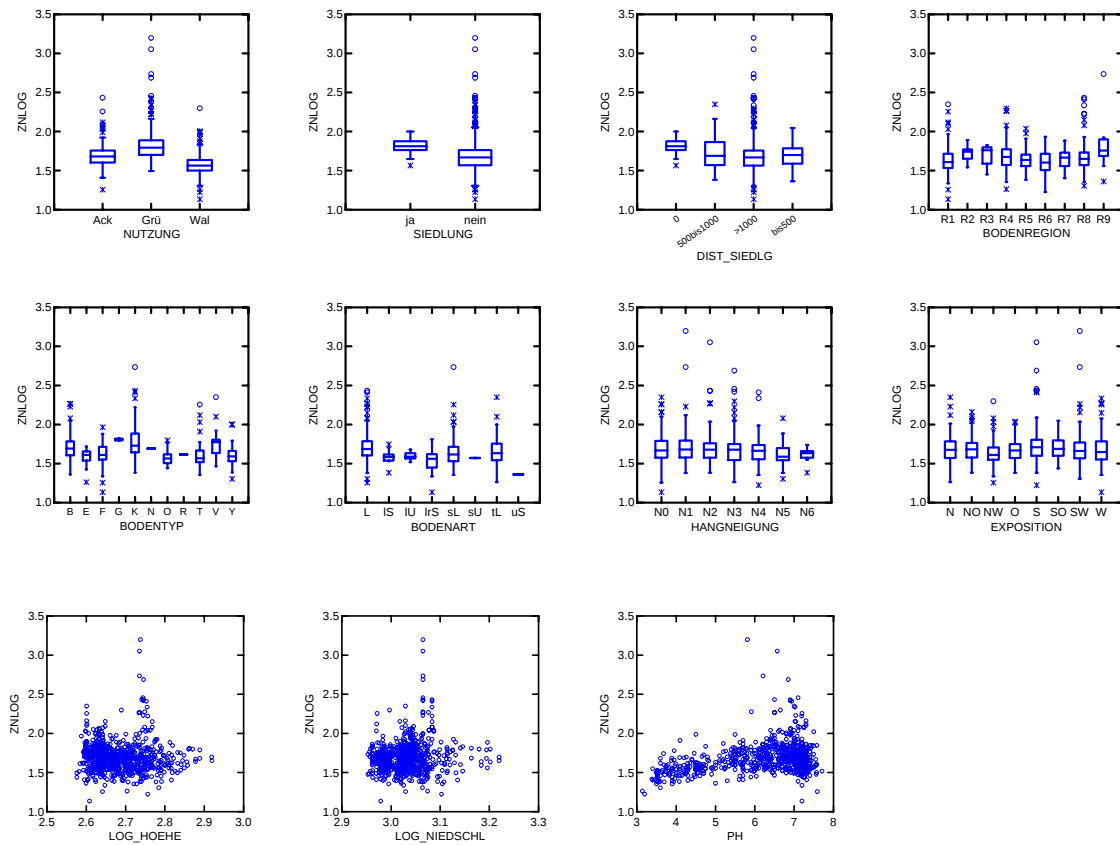
Tabelle 3-11: Allgemeine statistische Kenngrößen der Zinkdaten Datensatz 793 (in mg/kg)

	Zn
Anzahl Datensätze	793
Minimum	13.60
Maximum	1570.00
Median	47.00
Mittelwert	57.37
Standardabweichung	75.80
Variationskoeffizient	1.32
Schiefte	14.37
Exzess	252.87
25. Perzentil	37.10
75. Perzentil	59.00
95. Perzentil	100.00
99. Perzentil	260.42
	Zn(log)*
Anzahl Datensätze	793
Schiefte	2.11
Exzess	10.63

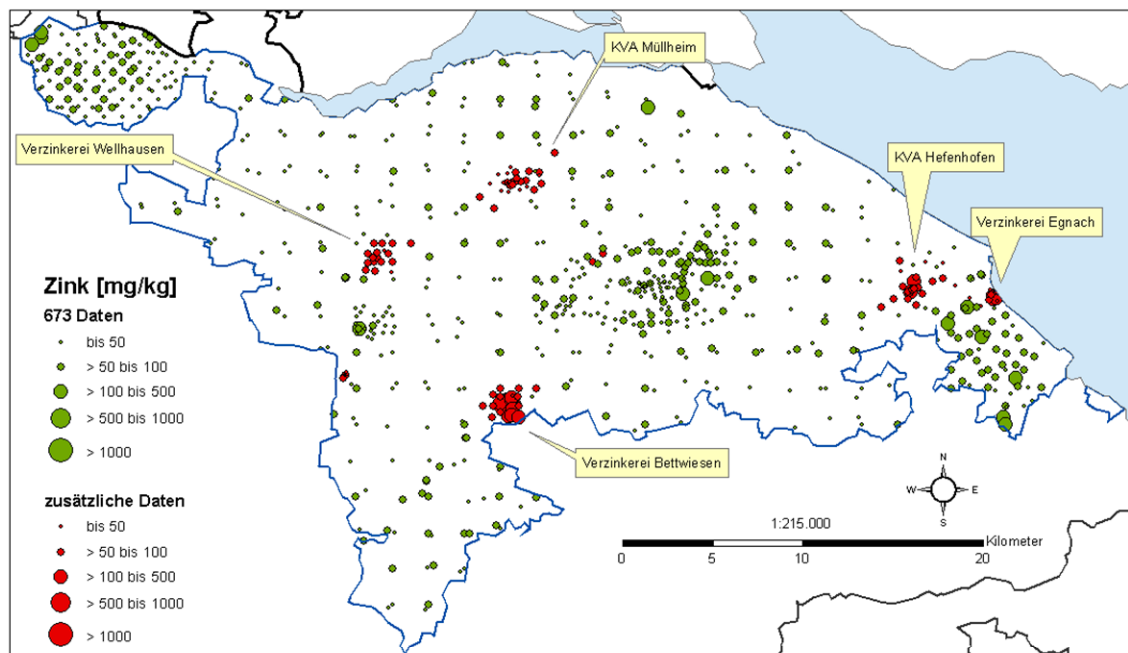
**Figur 3-16: Verteilung der Zinkmesswerte Datensatz 793 (links Zn-, rechts logZn-Werte in mg/kg))**

Auch wenn eine Reihe von hohen Zinkgehalten hinzukamen, änderte sich kaum das Verhältnis der Zinkdifferenzen der einzelnen Kategorien der erklärenden Variablen (Figur 3-17). Die Waldstandorte zeigten noch immer deutlich geringere Zinkgehalte als auf Acker-/Grünlandstandorten. Ebenfalls gab es erneut Unterschiede zwischen Siedlungsgebieten und Aussenbereich, allerdings nicht mehr so deutlich wie beim Datensatz 673. Unverändert blieben die Unterschiede beim Bodentyp und beim pH-Wert.

Die 120 Datensätze der hinzugenommenen Spezialuntersuchungen verteilten sich vorwiegend auf einige wenig Gebiete. Deshalb änderte sich an der Verteilung der Stützpunkte über den Kanton hinweg nur wenig. Figur 3-18 verdeutlicht, dass vereinzelt erhöhte Zinkgehalte in Böden in der Umgebung von Verzinkereien und KVA's von bis über 1000 mg/kg gemessen wurden.



Figur 3-17: Univariate Statistik der erklärenden Variablen auf die Zink-Bodenbelastung Datensatz 673, logZn in mg/kg



Figur 3-18: Zinkgehalte an den Untersuchungsstandorten Datensatz 793

3.2.2.2 Räumliche Datenanalyse

Für die Erstellung der Zinkkarte mit den Daten der Verzinkereien und KVAs wurden trotz der Ergebnisse der räumlichen Datenanalyse der reduzierten Daten nicht die Parameter der BÜK ins Regressionsmodell einbezogen. Zu viele der neu hinzugezogenen Daten hätten in diesem Falle nicht berücksichtigt werden können, da es sich teilweise um Standorte im Siedlungsgebiet handelt, die von der BÜK nicht abgedeckt werden. Darüber hinaus wurden in dieser und den folgenden Analysen ausschliesslich die rohen Residuen für die Variographie und Interpolation verwendet. Die Heruntergewichtung von auffälligen Werten im Zug der Erstellung der gewichteten Residuen nivellierte die Werte der Verzinkereien und KVAs nahezu vollständig, was bekanntermassen fern der Realität wäre. Die damit einhergehenden Unsicherheiten in der Aussagegenauigkeit galt es entsprechend zu diskutieren.

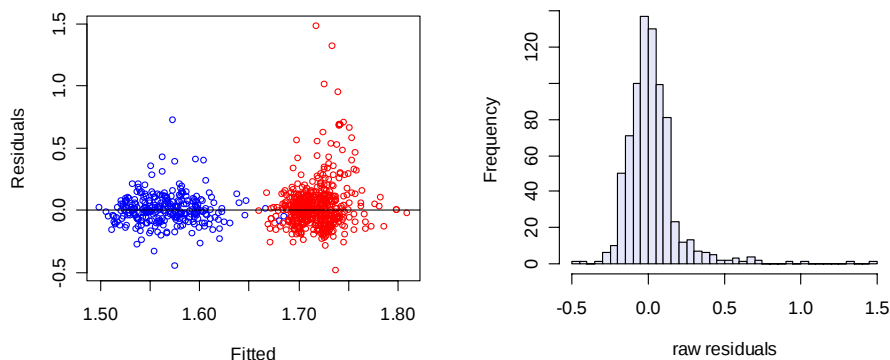
Regressionsanalyse

Analog zum Datensatz 673 liessen sich die Zinkgehalte des Datensatz 793 am besten mit den erklärenden Variablen Nutzung, pH-Wert und Niederschlag ausdrücken. Somit wurde erneut das Modell 1 verwendet. Trotz der neuen Zinkgehalte und der erhöhten Zahl an Freiheitsgraden ergaben sich kaum Unterschiede, weder für die Koeffizienten noch für das Bestimmtheitsmass. Letzteres sank von 0,366 auf 0,357 ab (Tabelle 3-12). Der deterministische Anteil in den Streuungen blieb damit auch bei einer Hinzunahme von 120 Messwerten mit teilweise stark erhöhten Zinkgehalten gleich.

Tabelle 3-12: Regressionsmodell für logarithmierte Zinkgehalte Datensatz 793

Koeffizient	Kategorie	Zn793, Modell 1
α		-0.043
β_1 Nutzung	AckGrü	0
	Wal	-0.126
β_2 pH		0.022
$\beta_3 \log_{10}(\text{Niederschlag})$		0.533
Anzahl Freiheitsgrade		753
adjusted R ²		0.357

Die bereits im Histogramm der logarithmierten Zinkgehalte gezeigte Rechtsschiefe der Daten (Figur 3-16) liess sich auch in den Residuen des Regressionsmodells wieder finden (Figur 3-19). Die Residuen schwankten grundsätzlich – aufgetragen gegen die angepassten Werte – gut um die Nulllinie. Dabei hatten aber die log-Residuen > 0,5 deutlich zugenommen, was sich auch im Histogramm zeigte. Im Zuge der räumlichen Analyse musste daher der Ausschluss einzelner auffälliger Werte geprüft werden. Die äusserst scharfe Trennung zwischen Residualwerten der Nutzung Acker/Grünland und Wald blieb unverändert.



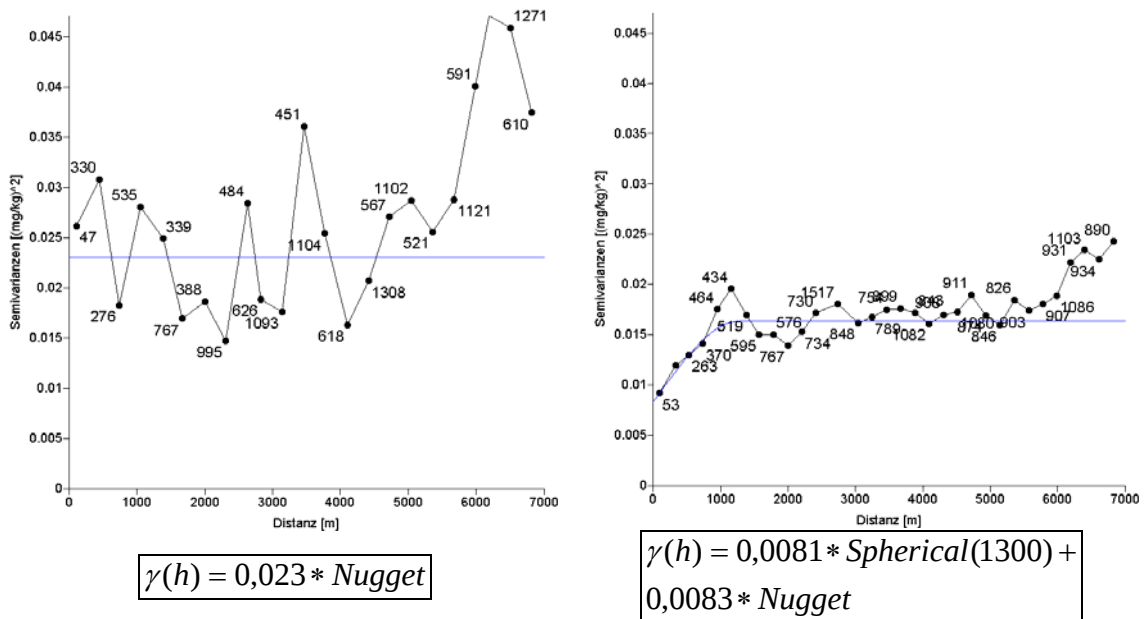
Figur 3-19: Streudiagramm der rohen Residuen gegen die angepassten Werte (Fitted) mit Nutzung Acker/Grünland (rot) und Wald (blau) (links), Histogramm der rohen Residuen (rechts), Zn793 Modell 1, Angaben in log mg/kg

Variographie

Die Rechtsschiefe der Residuen hatte Auswirkungen auf die Variographie und Interpolation. Wie stark dabei einzelne Werte bei weit über 700 Stützpunkten das Ergebnis beeinflussen konnten, wurde mit den folgenden Auswertungen deutlich. Im Zuge der Variographie wurden auf Grundlage der Semivariogramm-Cloud 4 räumlich auffällige Werte im Bereich der Verzinkerei Bettwiesen, die gleichzeitig die höchsten Residualwerte darstellen, eliminiert. In der Folge wurden die weiteren Auswertungen sowohl für den Datensatz aller Residuen (771 Stützpunkte) durchgeführt, als auch für den um die 4 Residuen bereinigten Datensatz (767 Stützpunkte).

Die Variogramme der beiden Datensätze zeigten deutliche Unterschiede (Figur 3-20). So zeigte der Datensatz mit allen Residuen keinerlei räumliche Zusammenhänge, gerade im Nahbereich waren die Semivarianzen eher höher als geringer. Einzige Möglichkeit war die Anpassung an ein Pure-Nugget-Modell. Nach Eliminierung der 4 stark erhöhten Zinkgehalte wiesen die Konzentrationen mit Anpassung eines sphärischen Modells an das experimentelle Variogramm eine Autokorrelation bis etwa 1300 m auf und einen klar tieferen Schwellenwert. In beiden Variogrammen wurde eine Anisotropie eingestellt, im Falle der vollständigen Residuen ein 30°Fächer in Richtung NO-SW und im Falle der reduzierten Residuen ein 45°Fächer in Richtung WNW-OSO.

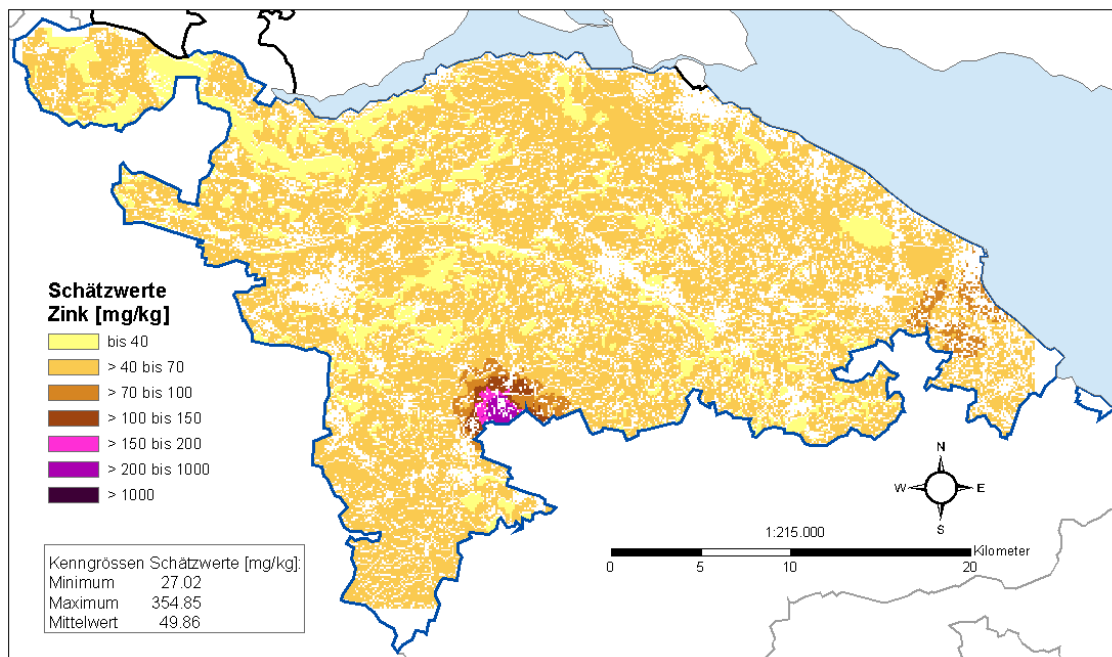
Im Gegensatz zum entsprechenden Variogramm der 673 im vorigen Kapitel betrachteten Zinkwerte (Modell 1 mit rohen Residuen) erhöhten sich die Semivarianzen nach Einbezug der Verzinkereien und KVAs nur minimal (Residuen ohne die auffälligen Werte). Das sprach für die räumlich gute Korrelation der neu hinzugekommenen Daten.



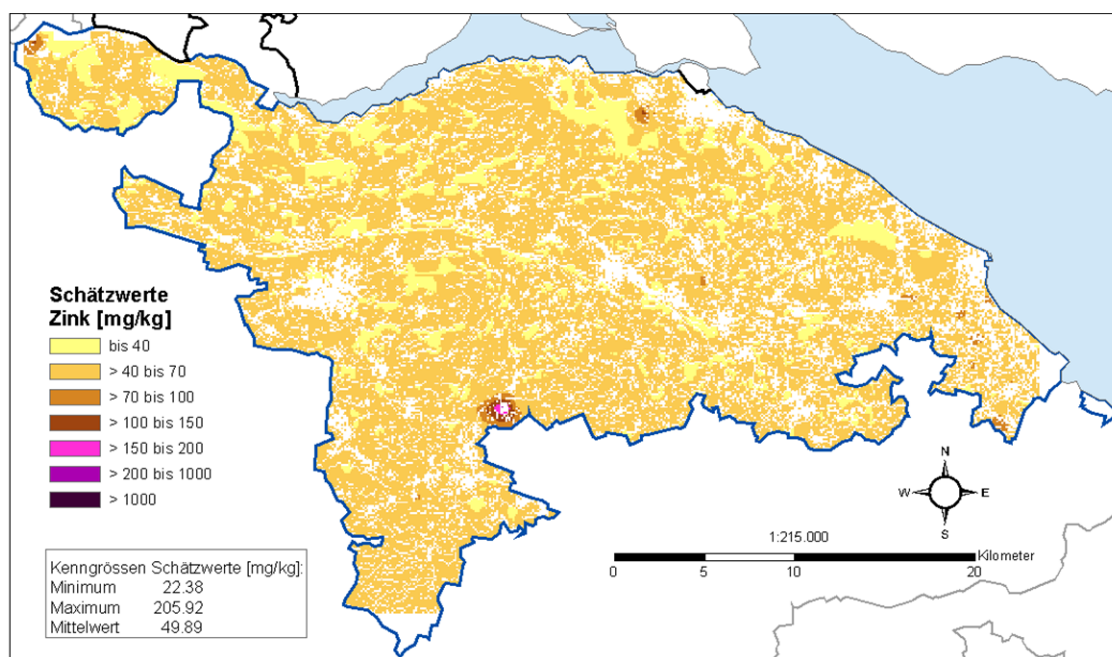
Figur 3-20: Semivariogramm aller Residuen (links) und der Residuen ohne 4 auffällige Werte (rechts), Zn793 Modell 1

Interpolation

Analog zu den vorhergehenden Interpolationen wurde im Falle aller Residuen aufgrund des Pure-Nugget-Modells ein Ordinary Kriging durchgeführt und im Falle der Residuen ohne die 4 Werte ein Simple Kriging. Der Ausschluss der höchsten 4 Werte führte zu einer drastischen Minimierung des Maximums, der Mittelwert und das Minimum blieben in etwa gleich (Figur 3-21, Figur 3-22). In beiden Fällen wurden die Zinkwerte verglichen mit dem Mittelwert der gemessenen Gehalte von 57 mg/kg merklich geglättet auf < 50 mg/kg im Mittel für die Schätzwerte. Dennoch fiel vor allem das Gebiet der Verzinkerei Bettwiesen (vgl. Figur 3-18) als Gebiet mit erhöhten Zinkgehalten, nämlich klar oberhalb des Richtwertes, in beiden Karten auf. Die Eliminierung der 4 Werte hatte nicht nur im entsprechenden Bereich Auswirkungen auf die Schätzwerte, sondern über die gesamte Karte. Teilweise resultieren höhere, teilweise niedrigere Gehalte. Zurückzuführen war dies auf die voneinander abweichende Variogrammfunktion und die Interpolationsmethode.



Figur 3-21: Interpolierte Zinkgehalte auf Basis aller Residuen Zn793 Modell 1, interpoliert mit Ordinary Kriging

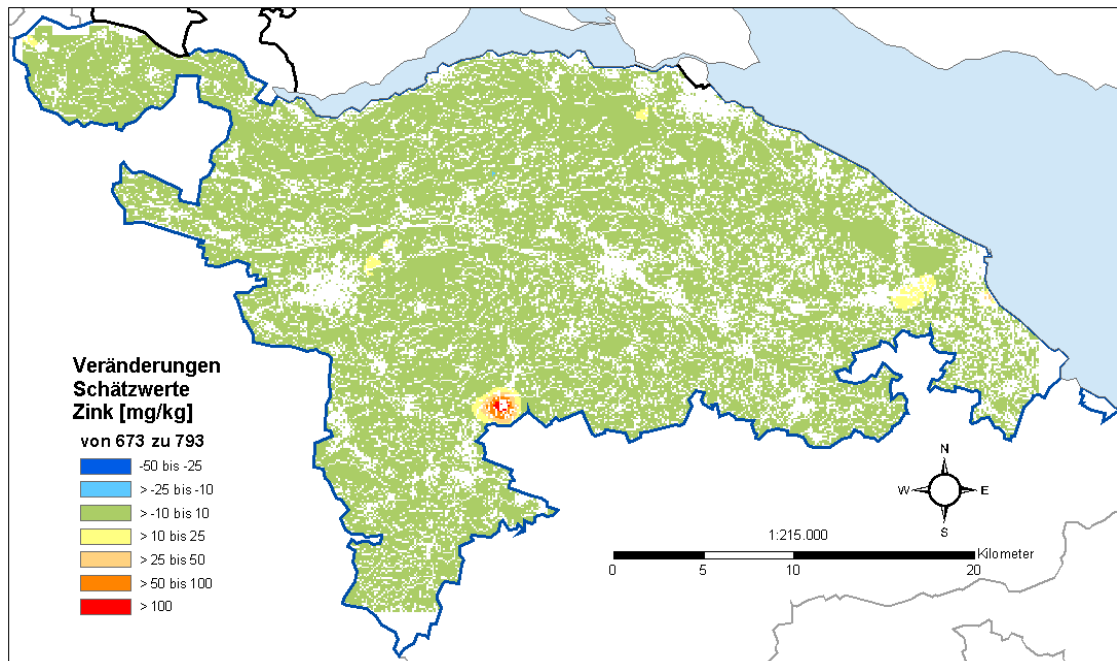


Figur 3-22: Interpolierte Zinkgehalte auf Basis der Residuen ohne 4 auffällige Werte Zn793 Modell 1, interpoliert mit Simple Kriging

Das Muster der Zinkgehalte ähnelte bis auf das Gebiet der Verzinkerei Bettwiesen dem der interpolierten 673 Datensätze. Auch hier trat durch den Einbezug der erklärenden Variablen des Modell 1 die Nutzung hervor, allerdings lange nicht mehr so eindeutig. Dennoch zeigten die Waldgebiete auch hier tendenziell geringere Zinkgehalte im Boden als die Acker-/Grünlandflächen. Lediglich die Waldgebiete im Süden zeigten erneut durch den hohen Jahresniederschlag höhere Gehalte an. Werte von über 70 mg/kg wur-

den für die gleichen kleinen Gebiete (v. a. Gemeinden Tägerwil und Schlatt) prognostiziert.

Insgesamt wurde der Mittelwert nach Hinzunahme der Werte ein wenig angehoben, auch wenn für den Grossteil der Kantonsfläche ± 10 mg/kg identische Schätzgehalte prognostiziert wurden (bemerkenswerterweise aber eben auch verringerte Gehalte) (Figur 3-23). Die höheren Gehalte zeigten sich erwartungsgemäss deutlich in der Umgebung der drei Verzinkereien, besonders Bettwiesen und Egnach, sowie in der Umgebung der KVA Hefenhofen (Figur 3-18).



Figur 3-23: Vergleich der interpolierten Zinkgehalte Zn793 Modell 1 (ohne 4 auffällige Werte) mit Zn673 Modell 1 (rohe Residuen) (dargestellt sind die Differenzen Zn793 – Zn673)

Prognoseintervall und Kreuzvalidierung

Die Prognoseintervalle der beiden Schätzungen zeigten unterschiedliche Bereiche als Schwankungsbereiche an. Im Falle der vollständigen Residuen lagen die Werte mit 95 %iger Wahrscheinlichkeit zwischen 13 mg/kg und 723 mg/kg. Ohne die auffälligen Werte waren die Schätzwerte wesentlich zuverlässiger, d.h. das Prognoseintervall enger. Die Schätzwerte schwankten mit gleicher Wahrscheinlichkeit nur noch zwischen 13 mg/kg und 328 mg/kg (Anhang C- 3, Anhang C- 4).

Im Gegensatz zum Prognoseintervall der entsprechenden interpolierten Karte der 673 Datensätze lagen damit vermehrt Zinkgehalte > 150 mg/kg als obere Grenze vor. Neben der Gemeinde Tägerwil nun auch in der Gemeinde Schlatt (äusserster Nordwesten) und im Umfeld der Verzinkei Bettwiesen.

Auch die Kreuzvalidierung schätzte das Modell ohne die 4 auffälligen Werte als das

bessere Gesamtmodell ein (Tabelle 3-13). Allerdings unterschied sich der MAD kaum voneinander. Dafür war der mittlere Fehler und die systematische Überschätzung der Werte deutlich weniger stark ausgeprägt.

Der Vergleich zu den 673 Daten fiel, wie auch der Vergleich der Prognoseintervalle, ein wenig, aber nicht entscheidend schlechter aus. Der MAD stieg ein wenig an und die Werte wurden systematisch etwas überschätzt.

Tabelle 3-13: Kenngrößen der Kreuzvalidierung für den Datensatz Zn793 (in log mg/kg)

Kenngrößen Kreuzvalidierung	Zn793, Modell 1 (alle Residuen)	Zn793, Modell 1 (Residuen ohne 4 auffällige Werte)
ME	0.01016	0.003862
RMSSE	0.8607	0.9731
MedE	0.0104	0.0098
MAD	0.0682	0.0680

Zusammenfassende Diskussion

Durch die Hinzunahme der Daten der Verzinkereien und KVAs ergab sich ein leicht rechtsschiefer log-Datensatz, dessen Zinkgehalte unverändert am besten mit den erklärenden Variablen Nutzung, pH-Wert und Niederschlag erklärt werden konnten. Die Regressionskoeffizienten blieben aufgrund der robusten Schätzung nahezu unverändert. Allerdings wurde für diesen Datensatz keine Autokorrelationsstruktur gefunden, da einzelne lokal erhöhte Zn-Gehalte (die aufgrund der Verwendung der rohen Residuen als Extremwerte erhalten blieben) in der Nähe von Verzinkereien die Semivarianzen stark erhöhten und so die Autokorrelation überlagerten. Entsprechend wurde ein Pure-Nugget-Variogramm angepasst und ein Ordinary Kriging durchgeführt. Im Ergebnis zeigte das 95 % Prognoseintervall Zinkgehalte von über 100 mg/kg im Boden, im Umfeld der Verzinkereien und im Bezirk Arbon von über 150 mg/kg und damit wesentlich breitere Vorhersageintervalle als zuvor beim Datensatz 673. Dass allerdings die Variabilität in den Daten deutlich überschätzt wurde zeichnete sich bereits durch die Kreuzvalidierung ab, auch wenn der absolute Fehler nicht entscheidend höher war als beim Datensatz 673 mit rohen Residuen.

Welche Auswirkungen allerdings einige wenige Werte bei fast 800 betrachteten Datensätzen in der räumlichen Datenanalyse haben konnten, zeigten die Ergebnisse des Datensatz 793 nach Eliminierung von 4 räumlich auffällig hohen Werten. (Dies ist jedoch nur bei einer Verwendung der rohen Residuen der Fall, im Fall von gewichteten Residuen wäre kein Unterschied auszumachen, da bereits alle auffällig hohen Residualwerte heruntergewichtet worden wären.) Es konnte eine räumliche Abhängigkeit der Residuen festgestellt werden, deren Semivarianzen kaum mehr höher waren als beim analogen Ansatz des Datensatz 673 (rohe Residuen) und zu einem vergleichbaren sphärischen Variogrammmodell führten. Die mittels eines Simple Kriging interpolierten Schätzwerte waren sowohl den maximalen Schätzwert als auch die maximale Grenze des Prognoseintervalls betreffend klar geringer als ohne Eliminierung der 4 Werte. Sie führten nur

noch im Umfeld der Verzinkerei Bettwiesen zu geschätzten Zinkgehalten oberhalb des Richtwertes. Das Muster der Schätzwerte entsprach, bis auf die Umgebung der Verzinkerei, dem Muster der Schätzkarte auf Basis der 673 Datensätze (rohe Residuen). Die höhere Streuung in den Daten führte dennoch zu etwas breiteren Prognoseintervallen und damit einer geringeren Genauigkeit. Betrachtete man die Ergebnisse der Kreuzvalidierung liess sich kein grosser Unterschied zum Datensatz 673 mit rohen Residuen erkennen. Der MAD wurde etwas höher und die Variabilität in den Daten etwas überschätzt. Dafür aber wurden für einige Gebiete Zusatzinformationen berücksichtigt, die die Karte als realitätsnäher einstufen lassen konnten. Die Karte mit Einbezug der Daten der Verzinkereien und KVAs wurde daher eindeutig präferiert. Eine sorgfältige Evaluation auffälliger Werte ist allerdings unbedingt erforderlich.

3.2.3 Datensatz 1290

In einer zweiten Erweiterung wurden nun auch die Standorte im Nutzungen Siedlungsbereich, Obst- und Gemüsebau und in den Rebbergen eingeschlossen. Damit wurden 1290 der 1549 für den Kanton Thurgau vorhanden Zinkuntersuchungen ausgewertet (Tabelle 3-14).

Tabelle 3-14: Untersuchungskampagnen Datensatz 1290

Nr	Projekt	Thema	Stao(n) (alle Nutzungen *)
1	P95RAS	SM in Land- und Forstwirtschaft	345
2	P94SIE	SM in Siedlungsböden	162
3	P93LW DI	SM in intens. Landwirtschaftsbezirken	114
4	P92KVAW	Untersuchung vor KVA	103
5	P93LW OT	SM in intens. Landwirtschaftsbezirken	93
8	P95RB OB	Kupferbelastung in ehemaligen/aktuellen Rebbergen	68
9	P00REB	Kupferbelastung in ehemaligen/aktuellen Rebbergen	60
10	P92KVAW G	Untersuchung vor KVA	59
12	P97SIE	SM in Siedlungsböden	46
14	P94ARA MT	SM in der Umgebung einer ARA	3
15	P93VZB	SM in der Umgebung einer Verzinkerei	38
16	P92KVAH	SM in der Umgebung einer KVA	31
17	P92KVAM	SM in der Umgebung einer KVA	30
19	P93DEP B	UVP für geplante Deponie	25
20	P93DEP KB	UVP für geplante Deponie	25
21	P93DEP RO	UVP für geplante Deponie	25
22	P93VZW	SM in der Umgebung einer Verzinkerei	18
23	P92VZE	SM in der Umgebung einer Verzinkerei	15
27	P93N1 AW	Sanierung des Lützelburg-Viaduktes	7
28	P99 BG	Baugesuchprüfung	7
29	P00 BG	Baugesuchprüfung	6
30	P95RAS AUW	Teilprojekt P95RAS	6
32	P01 BG	Baugesuchprüfung	4
Summe			1290

orange = zusätzliche Projekte

*bis auf die Nutzung, die „sonstigen“ Nutzungen zugeordnet wurden, wie z.B. „Nassstandorte“, Standorte mit „Unproduktiver Vegetation“ oder „Gebüsch, Strauchvegetation“

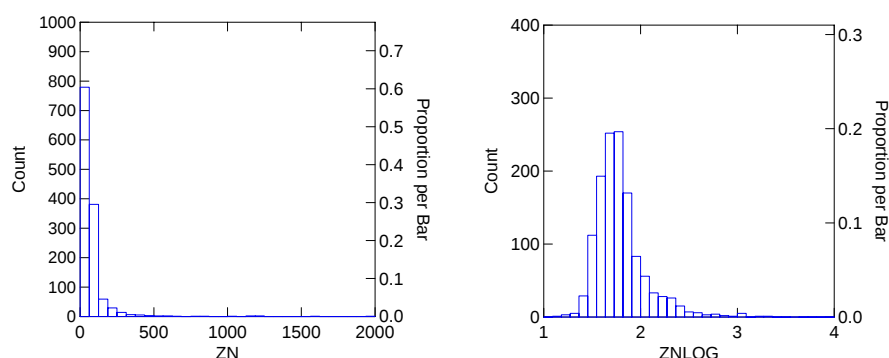
Die nicht berücksichtigten 259 Messwerte wurden für die hier durchgeführte räumliche Analyse als nicht relevant beurteilt, da es sich hierbei um lokale Belastungspunkte oder Belastungsuntersuchungen entlang von Strassen handelt, die nicht Gegenstand der Zielsetzungen dieser Studie waren (Anhang D- 1).

3.2.3.1 Explorative Datenanalyse

Durch die Erweiterung um mehr als 500 Datensätze veränderten sich die Kenngrößen entscheidend (Tabelle 3-15). Nicht nur der Mittelwert stieg deutlich an, auch der Median wurde merklich nach oben verschoben. Das 95. Perzentil lag klar oberhalb des Richtwertes und das Maximum erreichte fast den Sanierungswert für Zink. Allerdings waren die Daten nun nicht mehr so schief verteilt wie beim Datensatz 793 und konnten nach Logarithmierung eher wieder als annähernd normalverteilt angesehen werden (Figur 3-24).

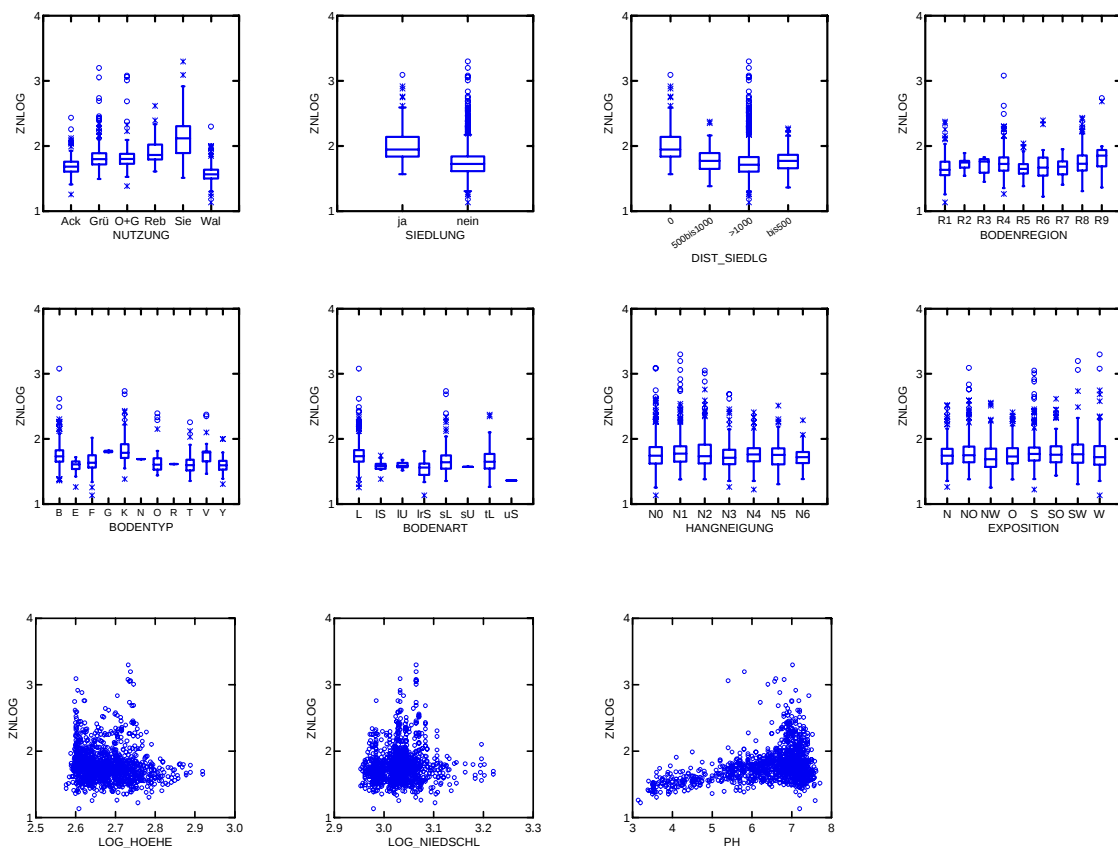
Tabelle 3-15: Allgemeine statistische Kenngrößen der Zinkdaten Datensatz 1290 (in mg/kg)

	Zn
Anzahl Datensätze	1290
Minimum	13.60
Maximum	1985.00
Median	55.50
Mittelwert	80.42
Standardabweichung	117.22
Variationskoeffizient	1.46
Schiefte	8.82
Exzess	103.05
25. Perzentil	42.38
75. Perzentil	74.10
95. Perzentil	197.00
99. Perzentil	550.80
	Zn(log)*
Anzahl Datensätze	1290
Schiefte	1.71
Exzess	4.97



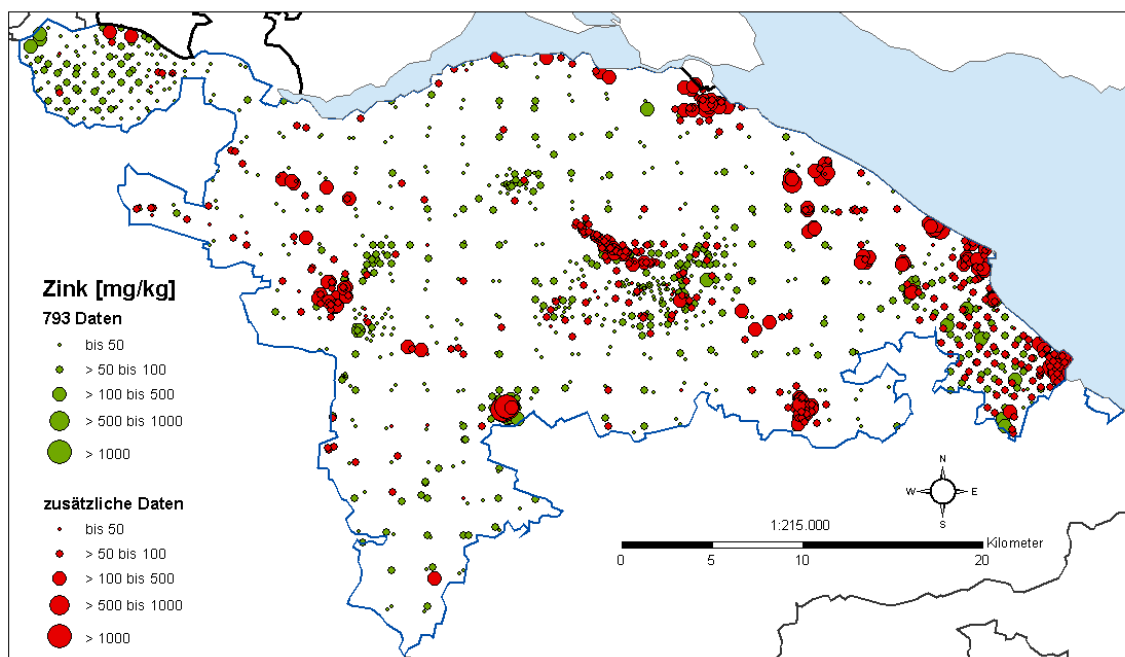
Figur 3-24: Verteilung der Zinkmesswerte Datensatz 1290 (links Zn-, rechts logZn-Werte in mg/kg)

Durch die Hinzunahme der neuen Nutzungen Obst-/Gemüse, Rebberge und Siedlungen verstärkten sich die Unterschiede zwischen den im Modell berücksichtigten Klassen der erklärenden Variablen weiter (Figur 3-25). Besonders hohe Differenzen lagen zwischen den Wald- und den Siedlungsgebieten vor. Der Unterschied zwischen Aussenbereich und Siedlungsgebieten war wieder ein wenig angewachsen, während die Differenzen zwischen den einzelnen Bodentypen unverändert blieben und der Einfluss des pH-Werte durch die Einbeziehung der hohen Werte im Siedlungsbereich keinen eindeutigen Effekt mehr zeigte.



Figur 3-25: Univariate Statistik der erklärenden Variablen auf die Zink-Bodenbelastung Datensatz 1290, logZn in mg/kg

Die neuen Daten traten zumeist lokal verdichtet auf, v. a. in den Siedlungsbereichen, die im Zuge der Rasteruntersuchung 1994 beprobt wurden (Figur 3-26). Auf den ersten Blick liessen sich diesen Untersuchungen im Mittel deutlich höhere Zinkgehalte zuweisen als im Landwirtschaftsraster.



Figur 3-26: Zinkgehalte an den Untersuchungsstandorten Datensatz 1290

3.2.3.2 Räumliche Datenanalyse

Analog zum Vorgehen der 793 Standorte wurden auch bei diesen 1290 Datensätzen nicht die Parameter der BÜK ins Regressionsmodell aufgenommen. Ebenso wurden ausschliesslich die rohen Residuen für die räumliche Analyse verwendet (vgl. Kap. 3.2.2.2).

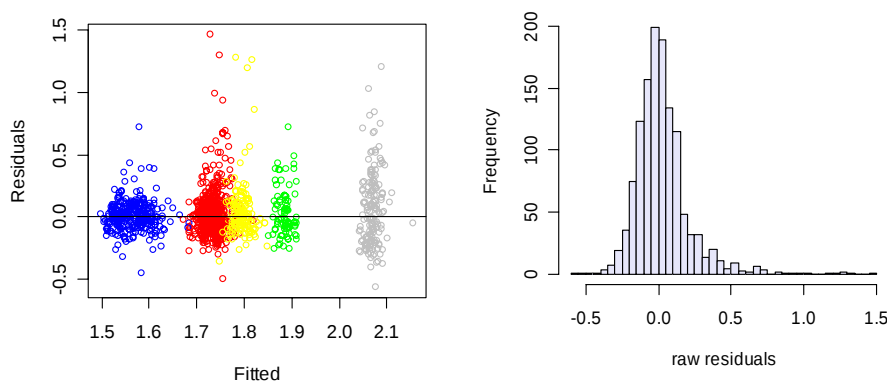
Regressionsanalyse

Für das Regressionsmodell wurden die gleichen erklärenden Variablen verwendet (Variable Nutzung aber nun mit 5 Klassen), wie bereits in den Datensätzen 673 und 793 mit dem Modell 1. Der erneute Einsatz des Modells 1 erlaubte einen direkten Vergleich der folgenden Analysen. Die deutlichen Unterschiede pro Nutzungsklasse in der univariaten Statistik bestätigten sich auch in der multivariaten Auswertung, die Siedlungsgebiete besaßen einen deutlich positiven Koeffizienten, während der Wald einen im Vergleich zu Acker/Grünland unveränderten negativen Koeffizienten aufwies (Tabelle 3-16). Die Koeffizienten der anderen erklärenden Variablen blieben ebenfalls in etwa unverändert. Als Resultat der grossen Nutzungsunterschiede stieg das Bestimmtheitsmass entscheidend an; mit $R^2 = 0,57$ konnte das Modell einen grossen Teil der Streuungen in den Zinkgehalten erklären.

Tabelle 3-16: Regressionsmodell für logarithmierte Zinkgehalte Datensatz 1290

Koeffizient	Kategorie	Zn1290, Modell 1
α		0.070
β_1 Nutzung	AckGrü	0
	Obst+Gemüse	0.063
	Rebberge	0.145
	Siedlung	0.329
	Wal	-0.137
β_2 pH		0.025
$\beta_3 \log_{10}(\text{Niederschlag})$		0.494
Anzahl Freiheitsgrade		1230
adjusted R^2		0.570

Wie schon beim Datensatz 793 zeigte sich die leicht rechtsschiefe Verteilung der log-Daten auch wieder in den Residualwerten, insgesamt aber schwankten die Residuen gegen die angepassten Werte symmetrisch um die Nulllinie (Figur 3-27). Die Residuen der neuen Nutzungstypen bewegten sich im gleichen Bereich wie die beiden zuvor betrachteten aggregierten Nutzungstypen (Acker/Grünland, Wald). Erneut sehr bemerkenswert war die scharfe Trennung der Nutzungen.



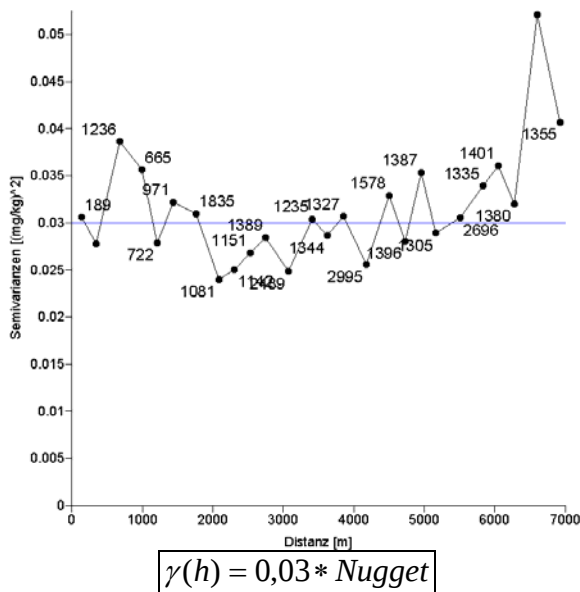
Figur 3-27: Streudiagramm der rohen Residuen gegen die angepassten Werte (Fitted) mit Nutzung Acker/Grünland (rot), Wald (blau). Obst/Gemüse (gelb), Rebberge (grün) und Siedlungen (grau) (links), Histogramm der rohen Residuen (rechts), Zn1290, Modell 1, Angaben in mg/kg

Variographie

In der Variogrammanalyse liessen die Semivarianzen der Messwertpaare keinen räumlichen Zusammenhang erkennen (Figur 3-28). Im Gegensatz zu den Daten der Verzinke-reien und KVAs handelte es sich bei den neuen Daten um lokale, scharf abgrenzbare und damit nicht stetige Belastungen über die Fläche. Auch wenn das Bestimmtheits-mass der Regression einen hohen Anteil an erklärbarer Streuung (57 %) angab ist zu vermuten, dass die Gesamtstreuung innerhalb der einzelnen Klassen nicht ausreichend erklärt werden konnte. Es musste daher davon ausgegangen werden, dass die Daten

noch über eine Drift verfügten, diese aber von den immer noch hohen Semivarianzen zwischen benachbarten Punkten überlagert wurde. Durch den Ausschluss von einigen wenigen Messwerten (wie beim Datensatz 793 geschehen) war es diesmal nicht möglich, eine räumliche Abhängigkeit im Datensatz zu erreichen.

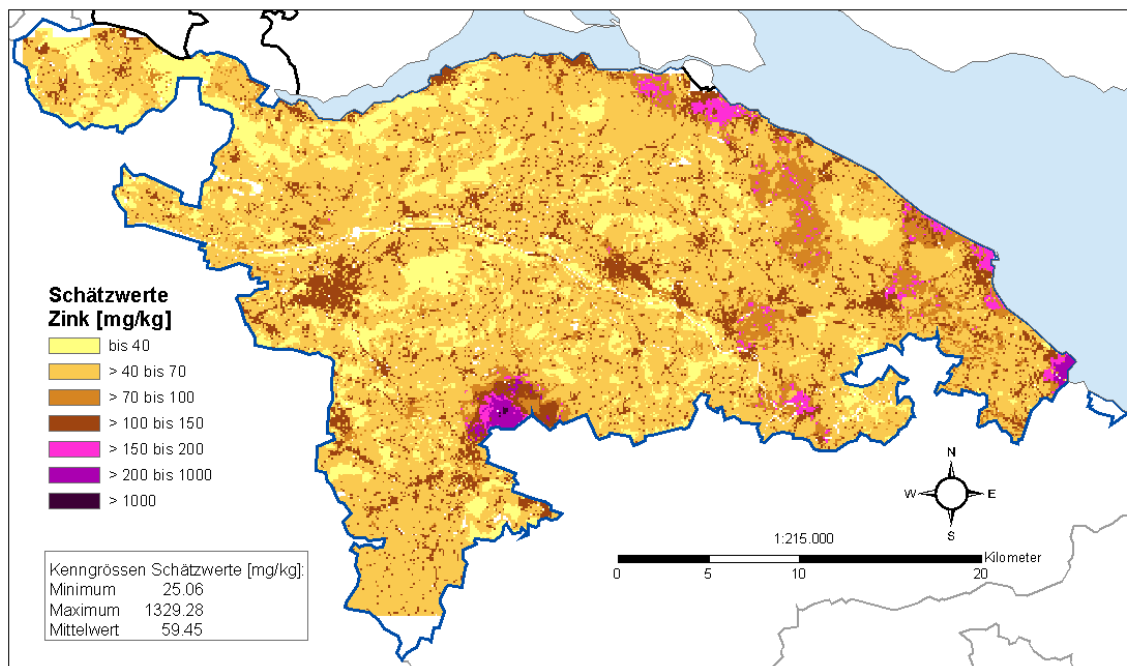
Der Nugget des Variogramms der log-Residuen lag bei $0,03 \text{ mg}^2/\text{kg}^2$ und damit deutlich höher als bei allen Variogrammen zuvor. Angepasst werden konnte nur ein Pure-Nugget-Variogramm. Es wurde eine Anisotropie eingestellt mit einem 30° Suchfächer in ONO-WSW-Richtung.



Figur 3-28: Semivariogramm Zn1290, Modell 1

Interpolation

Durch Einbezug der Nutzungen Siedlungen, Obst/Gemüse und Rebbau wurde die Aussage auf nahezu das gesamte Kantonsgebiet erweitert (Figur 3-29). Die Schätzwerte gingen bis auf über 1300 mg/kg hoch und es zeigte nicht mehr nur das Gebiet im Umfeld der Verzinkerei Bettwiesen geschätzte Zinkgehalte oberhalb des Richtwertes. Flächige Belastungen oberhalb des Richtwertes wurden auch für die Siedlungsgebiete Bischofszell, Arbon, Romanshorn und Kreuzlingen ausgegeben sowie in der Umgebung der Verzinkerei Egnach. Für die vier Siedlungsgebiete lagen rasterförmig Datensätze vor, die im Rahmen einer speziellen Untersuchung zu Schwermetallen in Böden von Siedlungen (Lehmann, 1995c) beprobt wurden. Insgesamt stieg der Mittelwert von knapp 50 mg/kg auf knapp 60 mg/kg an.

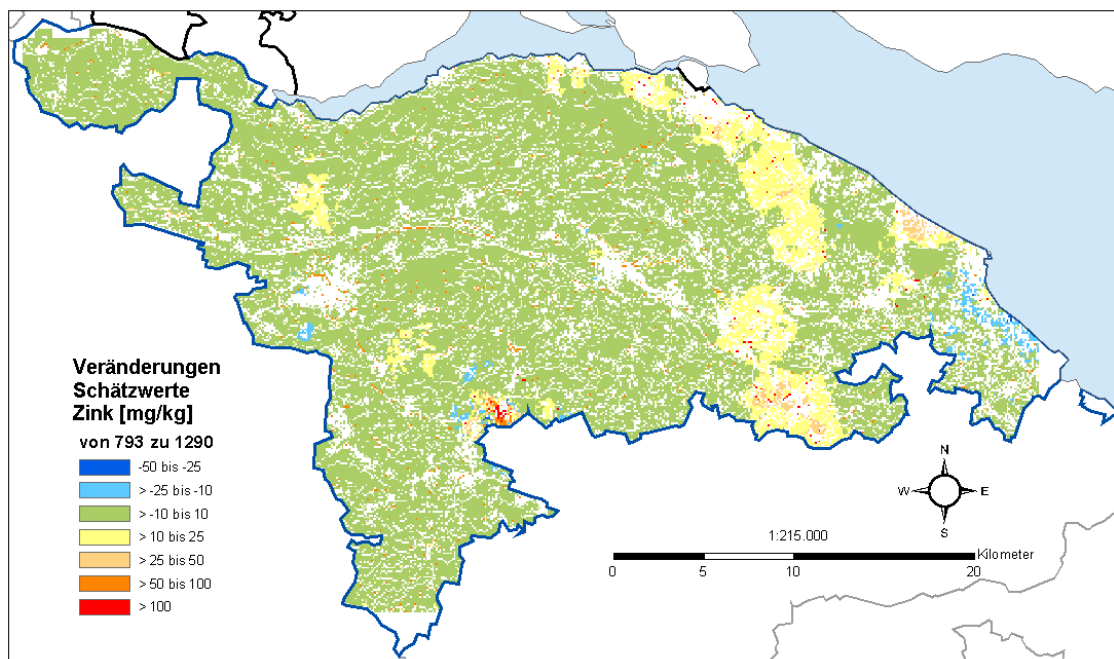


Figur 3-29: Interpolierte Zinkgehalte Zn1290 Modell 1, interpoliert mit Ordinary Kriging

Das Muster der Schätzwerte war wiederum stark durch die Nutzung geprägt. Entsprechend der Koeffizienten des Regressionsmodells war der Siedlungseinfluss sehr dominant, aber auch die anderen Nutzungen Obst-/Gemüsebau und die Rebberge liessen sich klar erkennen. Teilweise wurde dieses durch die Nutzung bedingte Muster allerdings vom Einfluss der unterschiedlichen Niederschläge überlagert, wie z.B. im westlichen Teil mit geringeren Niederschlägen.

Der Nachteil dieser Karte zeigt sich jedoch bereits im Nordosten des Kantons in Richtung Bodensee. Im Bereich der Siedlungsgebiete und der intensiv durch Obst- und Gemüsebau geprägten Gebiete wurden die Zinkgehalte flächig überschätzt. Es ist anzunehmen, dass die scharf abgrenzbaren höheren Gehalte dieser Nutzungen über die jeweilige Nutzungsgrenze „hinweg“ interpoliert wurden und somit z.B. in Waldgebieten, die eigentlich über geringere Gehalte verfügen sollten, die Zinkgehalte auf über 40 mg/kg steigen liessen.

Für den Grossteil der Acker-, Grünland- und Waldgebiete blieb die Schätzung der Zinkwerte dennoch ± 10 mg/kg gleich (Figur 3-30). Die angesprochenen Gebiete im Nordosten wurden hinsichtlich der Zinkgehalte um 10-50 mg/kg höher geschätzt, ohne dass neue Punkte für diese Nutzungen hinzugekommen waren. Spätestens mit dieser Differenzkarte wurde deutlich, dass die Karte der 1290 Datensätze mit äusserster Vorsicht zu behandeln und zu interpretieren ist.



Figur 3-30: Vergleich der interpolierten Zinkgehalte Zn1290 mit Zn793 (zur besseren Vergleichbarkeit mit Einbezug der 4 auffälligen Werte) (dargestellt sind die Differenzen Zn1290 – Zn793)

Prognoseintervall und Kreuzvalidierung

Wie gross diese Ungenauigkeit in der Schätzung der Zinkgehalte auf Grundlage der 1290 Daten war, belegten die Karten der Prognoseintervalle. Statt eines Prognoseintervalls von 13 mg/kg bis 328 mg/kg des Datensatzes 793, lagen die Werte des Datensatz 1290 mit einer Wahrscheinlichkeit von 95 % zwischen 11 mg/kg und 2998 mg/kg, d.h. bei einem maximalen Schätzwert von rund 1300 mg/kg konnte der Wert um bis zu 125 % nach oben schwanken (Anhang C- 5).

Unter Berücksichtigung der MADs der beiden vorhergehenden Datensätze mit demselben Regressionsmodell (Modell 1) und der Verwendung von rohen Residuen zwischen 0,0654 und 0,0680 wurde ersichtlich, wie viel grösser der Fehler unter Verwendung der Daten aller Nutzungsformen wurde (0,0808). Der RMSSE zeigte an, dass die Zinkgehalte zudem merklich systematisch überschätzt wurden (Tabelle 3-17).

Tabelle 3-17: Kenngrössen der Kreuzvalidierung für den Datensatz Zn1290 (in log mg/kg)

Kenngrössen Kreuzvalidierung	Zn1290, Modell 1
ME	0.009261
RMSSE	0.9169
MedE	0.0083
MAD	0.0808

Zusammenfassende Diskussion

Der Einbezug der Daten des Siedlungsgebietes, Obst- und Gemüsebaus und der Rebberge führte sowohl punktuell als auch lokal begrenzt zu erheblich höheren Zinkgehalten. Auch wenn der nutzungsabhängige Trend in den Daten mit dem gewählten Regressionsmodell generell erfasst wurde, war die Streuung innerhalb der Nutzungsklassen bei weitem grösser als zwischen den Nutzungsgruppen. Im Gegensatz zum Datensatz 793 zeigten die Residuen auch nach Ausschluss einiger stark erhöhter Werte keinen räumlichen Zusammenhang. Für die punktuellen Einträge der Obst-/Gemüsebau- und Rebbaustandorte wurde die Annahme der Stationarität nicht erfüllt. Gleiches galt für die Standorte im Siedlungsgebiet, deren Zinkgehalte auf benachbarten Standorten teilweise stark unterschiedlich waren, es bestand keine räumliche Abhängigkeit.

Resultierend führten die hohen Gehalte der neuen Nutzungstypen überwiegend zu einer Überschätzung der Zinkgehalte auf umliegenden Acker-, Grünland- oder Waldstandorten. Das Prognoseintervall wurde ebenfalls merklich breiter bis hin zu einer oberen Grenze von nahezu 3000 mg/kg. Die Kreuzvalidierung resultierte in einem grösseren absoluten Fehler und die Variabilität wurde systematisch überschätzt. Die Aussagegenauigkeit der Interpolation der 1290 Daten musste damit als deutlich verschlechtert bewertet werden.

3.3 Übertragbarkeit der Methode

3.3.1 Datensatz 386/193

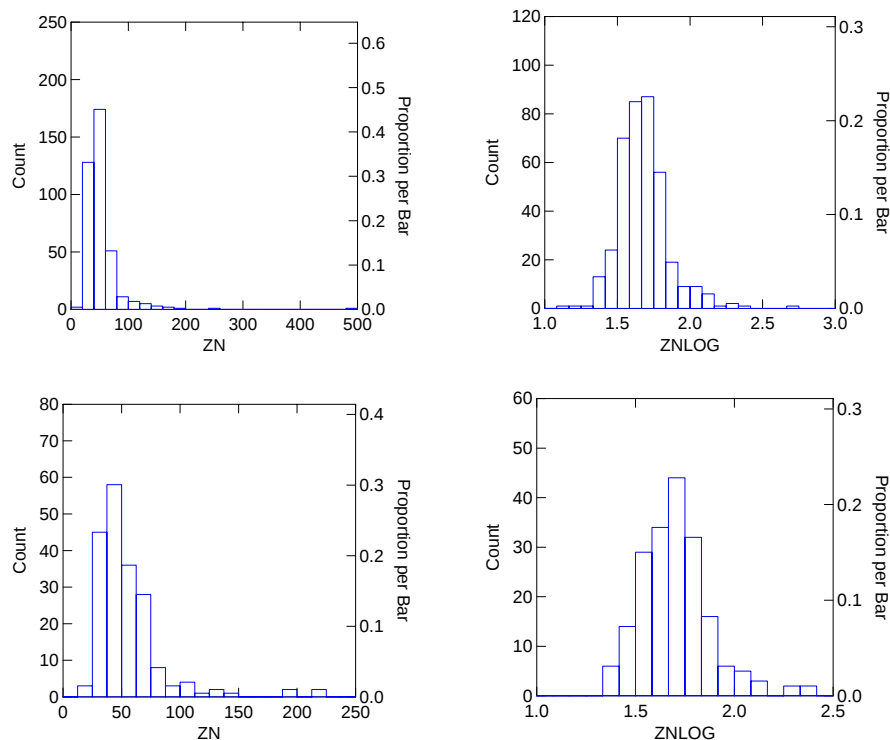
Der Datensatz für den Kanton Thurgau ist im Gegensatz zu den meisten anderen Kantonen äusserst umfangreich und bietet v. a. aufgrund der Rasterbeprobungen im land- und forstwirtschaftlichen Gebiet eine sehr gute Grundlage für eine kantonsweite Bewertung der Schwermetallgehalte im Boden. Zur Simulation der möglichen Datendichten in anderen Kantonen wurden in diesem Kapitel die Daten des Kantons Thurgau um 50 % und um 75 % ausgedünnt ausgewertet.

3.3.1.1 Explorative Datenanalyse

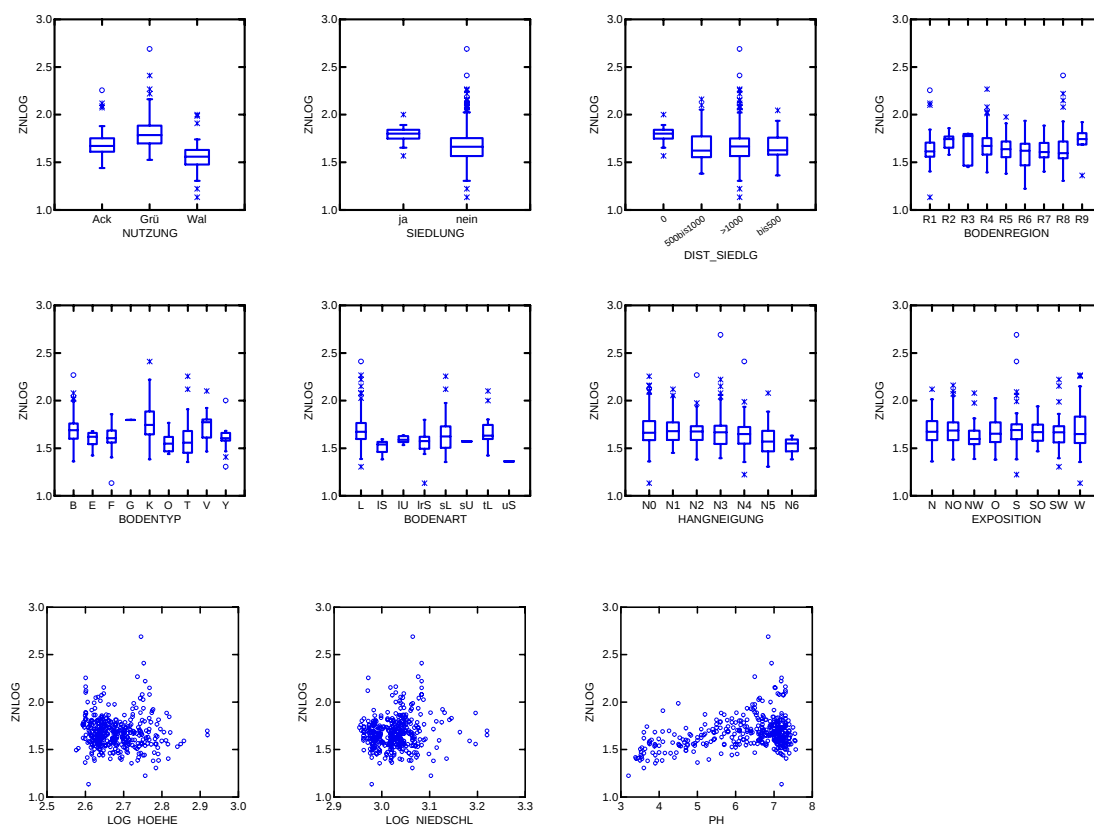
Für die Auswertung lagen damit Datensätze mit 386 (Zn386) und 193 (Zn193) Zinkdaten vor (Tabelle 3-18). Durch die Ausdünnung der Daten wurden die hohen Daten der ursprünglich 793 Standorte deutlich reduziert, für Zn386 allerdings zufällig stärker – bis auf das Maximum – als für Zn193. Beide Datensätze waren nach der log-Transformation annähernd normalverteilt (Figur 3-31).

Tabelle 3-18: Allgemeine statistische Kenngrößen der Zinkdaten Datensatz 386/193 (in mg/kg)

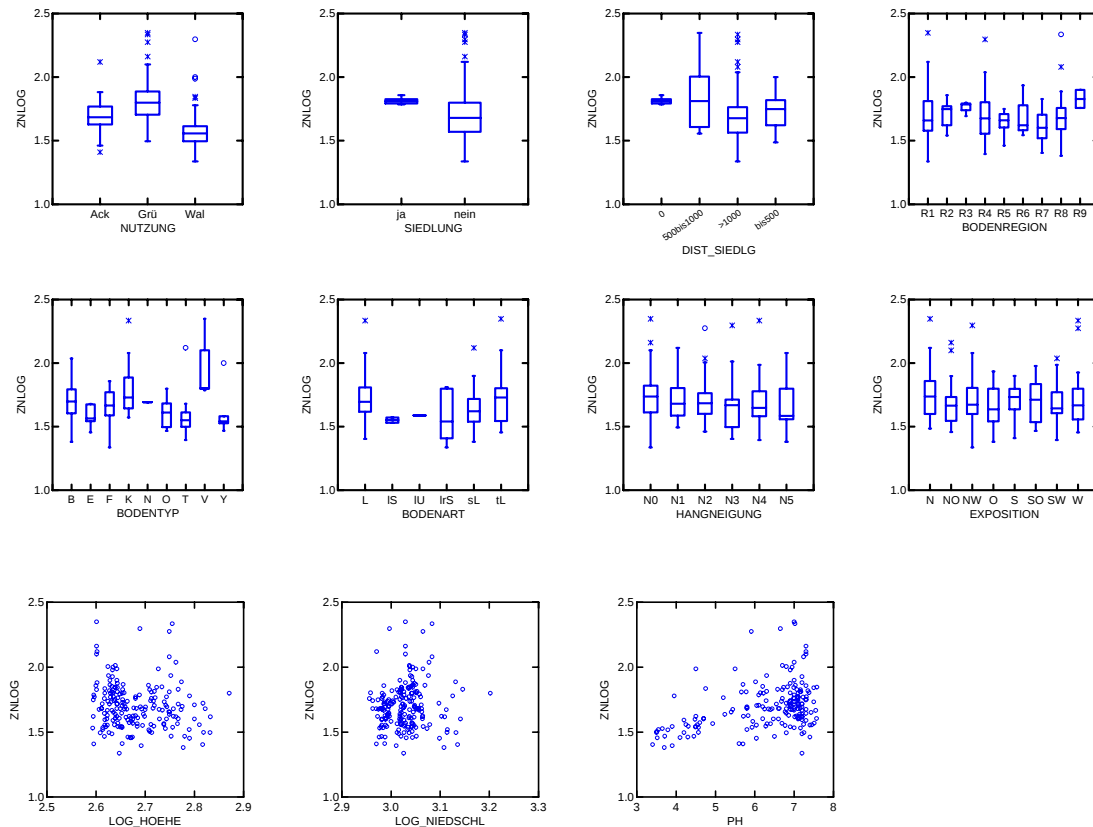
	Zn386	Zn193
Anzahl Datensätze	386	193
Minimum	13.60	21.70
Maximum	488.00	223.00
Median	46.40	47.90
Mittelwert	52.54	55.33
Standardabweichung	33.67	30.04
Variationskoeffizient	0.64	0.54
Schiefe	6.97	3.10
Exzess	76.86	12.62
25. Perzentil	36.90	37.98
75. Perzentil	57.50	63.00
95. Perzentil	100.60	102.70
99. Perzentil	174.83	208.26
	Zn(log)	Zn(log)
Anzahl Datensätze	386	193
Schiefe	1.19	1.00
Exzess	4.27	1.90

**Figur 3-31: Verteilung der Zinkmesswerte Datensatz 386 (oben), Datensatz 193 (unten) (links Zn-, rechts logZn-Werte in mg/kg)**

Die univariate Statistik der erklärenden Variablen zeigte für den Datensatz 386 keine grossen Unterschiede zum Datensatz 793. Die grössten Zinkdifferenzen zwischen den einzelnen Klassen wiesen erneut Nutzung, Siedlungsnähe, Bodentyp und pH-Wert auf (Figur 3-32 und Figur 3-33). Mit der Eliminierung weiterer Standorte allerdings (Datensatz 193), wiesen die im Siedlungsbereich gemessenen Zinkgehalte auf keine Unterschiede mehr zu Daten im Aussenbereich hin. Allerdings wurde die Anzahl der Standorte im Siedlungsbereich zufällig stark ausgedünnt.

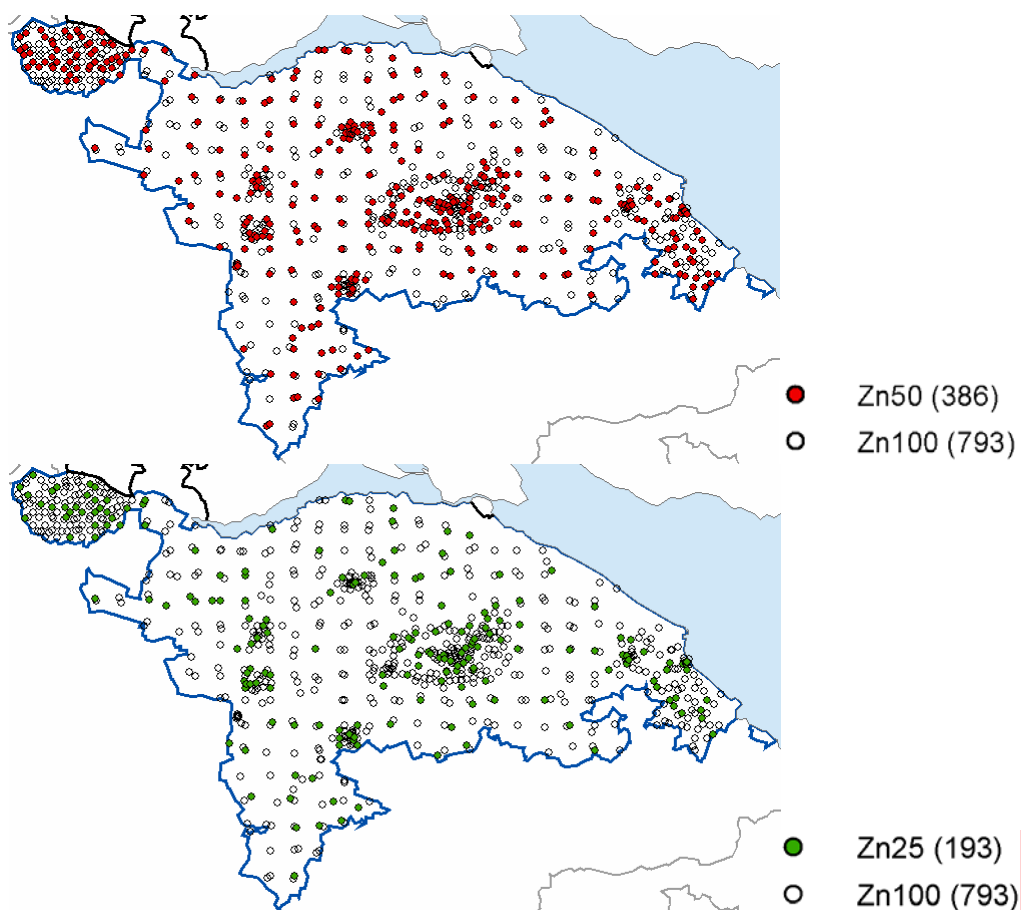


Figur 3-32: Univariate Statistik der erklärenden Variablen auf die Zink-Bodenbelastung Datensatz 386, logZn in mg/kg



Figur 3-33: Univariate Statistik der erklärenden Variablen auf die Zink-Bodenbelastung Datensatz 193, logZn in mg/kg

Die resultierende Verteilung der ausgedünnten Daten ist in Figur 3-34 dargestellt. Auch wenn die 3 grössten Ballungsgebiete separat ausgedünnt wurden, liegen in diesen Bereichen immer noch gehäuft Daten vor. Vor Ausdünnung betrug die Stützpunktdichte im Kanton Thurgau (basierend auf Datensatz 793) absolut betrachtet $0,92 \text{ Punkte/km}^2$, nach 50 %iger Ausdünnung noch $0,45 \text{ Punkte/km}^2$ und mit 25 % der Daten nur noch $0,22 \text{ Punkte/km}^2$. Aufgrund der Besonderheit der zwei unmittelbar benachbarten Stützpunkte des land- und forstwirtschaftlichen Rasters ändert sich die effektive Datendichte nach Ausdünnung von 50 % der Daten jedoch noch nicht massgeblich. Vielfach blieb mindestens einer der beiden Stützpunkte erhalten. Erst mit 25 % der Daten fallen des öfteren beide benachbarte Stützpunkte weg und verringern so entscheidend die Datendichte. Auf Grundlage der von Keller und Desaulles (2001) zusammengeführten Bodenschadstoffdaten wurde angenommen, dass man sich mit letzterer Punktdichte von $\sim 0,25 \text{ Punkten/km}^2$ im Bereich anderer Kantone bewege.



Figur 3-34: Standorte der Zinkgehalte von Zn386 und Zn193

3.3.1.2 Räumliche Datenanalyse

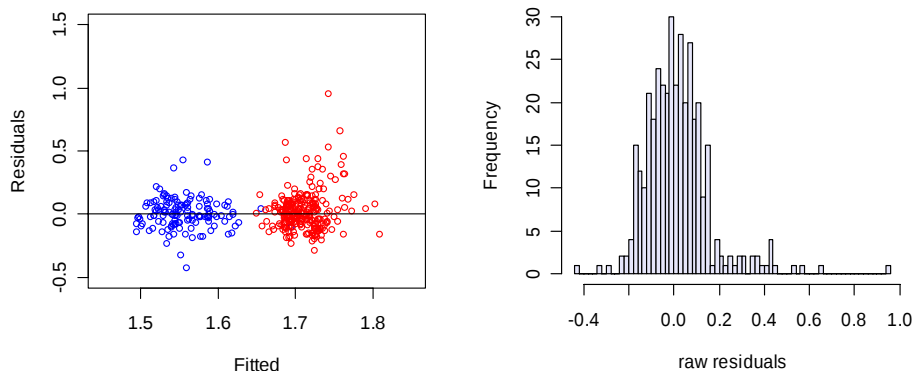
Regressionsanalyse

Die Regressionsanalyse erfolgte analog zur Analyse des Datensatzes 793, d.h. es wurde erneut Modell 1 verwendet. Im Unterschied zum Datensatz 793 veränderten sich in erster Linie die Konstante und der Koeffizient für den Niederschlag. Zudem stieg mit zunehmend ausgedünnten Daten das Bestimmtheitsmass ein wenig an (Tabelle 3-19).

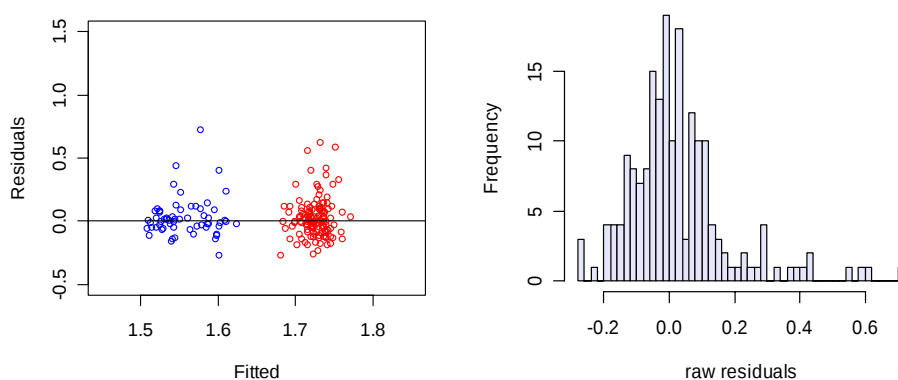
Tabelle 3-19: Regressionsmodelle für logarithmierte Zinkgehalte Datensatz 386/193

Koeffizient	Kategorie	Zn386, Modell 1	Zn193, Modell 1
α		-0.483	0.513
β_1 Nutzung	AckGrü	0	0
	Wal	-0.132	-0.133
β_2 pH		0.023	0.026
$\beta_3 \log_{10}(\text{Niederschlag})$		0.675	0.343
Anzahl Freiheitsgrade		366	179
adjusted R^2		0.368	0.389

Die Residuen wurden so ausgedünnt, dass sie weiterhin ausreichend symmetrisch um die Nulllinie schwankten (Figur 3-35 und Figur 3-36). Eine leichte Rechtsschiefe der Residuen blieb allerdings bestehen. Auffällig waren für beide Datensätze einige hohe Residualwerte, die im Zuge der Variographie kritisch geprüft werden mussten.



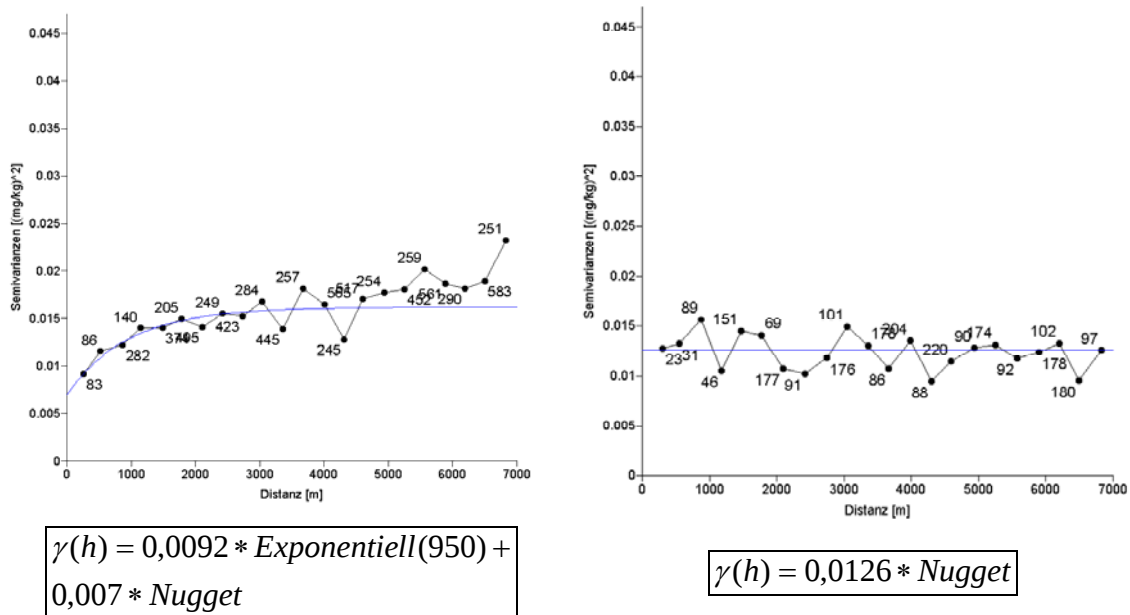
Figur 3-35: Streudiagramm der rohen Residuen gegen die angepassten Werte (Fitted) mit Nutzung Acker/Grünland (rot) und Wald (blau) (links), Histogramm der rohen Residuen (rechts), Zn386 Modell 1, Angaben in mg/kg



Figur 3-36: Streudiagramm der rohen Residuen gegen die angepassten Werte (Fitted) mit Nutzung Acker/Grünland (rot) und Wald (blau) (links), Histogramm der rohen Residuen (rechts), Zn193 Modell 1, Angaben in mg/kg

Variographie

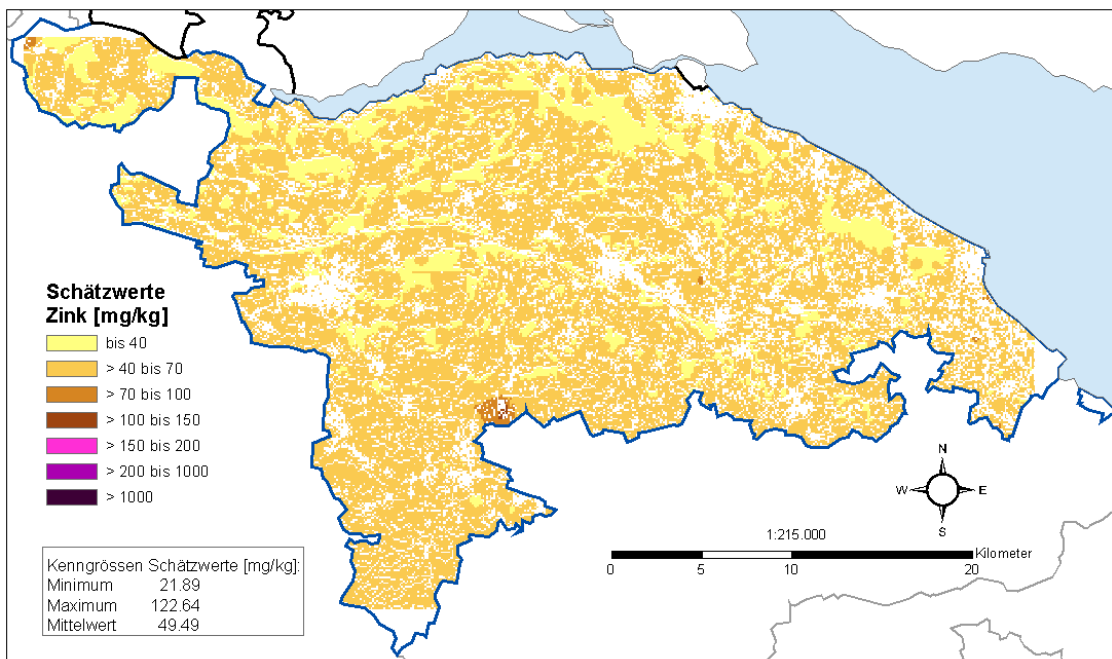
Die im Histogramm der Residuen auffälligen hohen Werte hatten teilweise Einfluss auf die räumliche Korrelation. So wurde für den Datensatz Zn386 der höchste Wert ausgeschlossen und für Zn193 die höchsten 4 Residualwerte. Im Ergebnis korrelierten die log-Residuen von Zn386 räumlich miteinander (Anisotropie mit 60° Suchfächer in NO-SW-Richtung) und es liess sich als bestes Modell eine exponentielle Funktion anpassen (Figur 3-37). Für Zn193 war auch nach Ausschluss der 4 höchsten Werte keine räumliche Korrelation mehr zu finden, die Daten schwankten aber geringfügiger um den Schwellenwert von 0.0126. In beiden Fällen lagen die Semivarianzen im Bereich der Semivarianzen aller 793 Daten (ohne 4 auffällige Werte), was für eine repräsentative Stichprobe sprach.



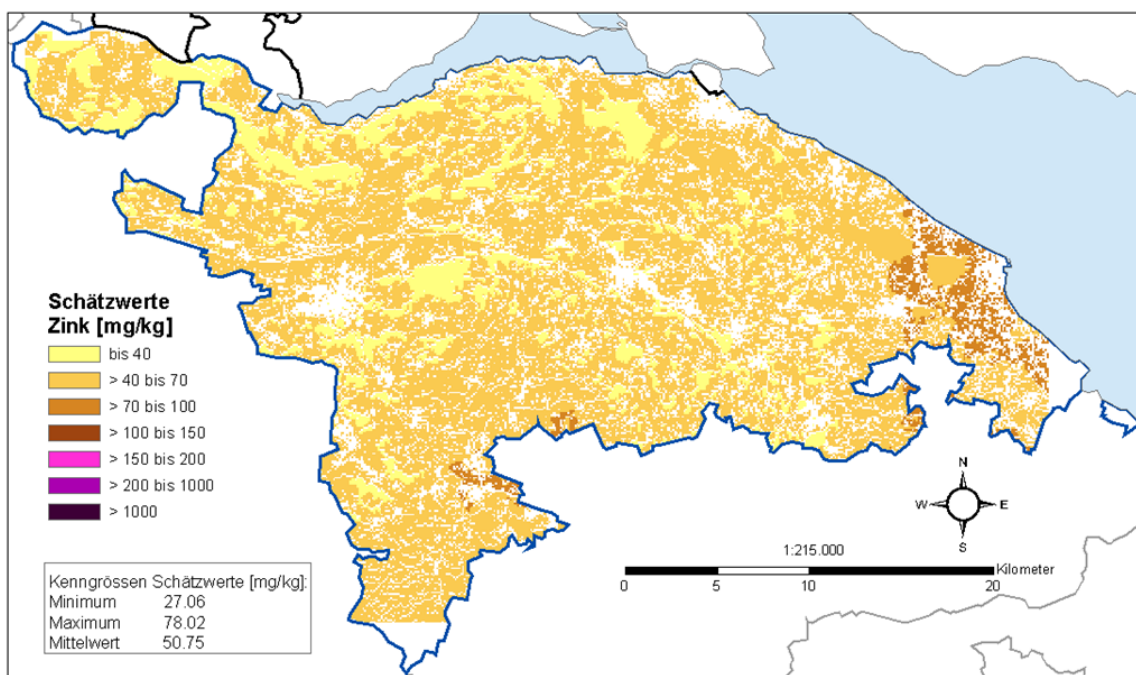
Figur 3-37: Semivariogramm Zn386 Modell 1, ohne 1 auffälligen Wert (links) und Zn193 Modell 1, ohne 4 auffällige Werte (rechts)

Interpolation

Im Mittel blieben die Schätzwerte durch die Reduzierung nahezu unberührt, die Maximalwerte wurden allerdings mit zunehmender Ausdünnung geglättet (Figur 3-38 und Figur 3-39). Nach Ausdünnung von 50 % der Daten ergab sich noch ein fast unverändertes Muster der Schätzwerte, was durch die kaum verschlechterte Stützpunkumgebung zurückzuführen war. Die Kombination der erklärenden Variablen Nutzung und Niederschlag prägten das Bild und die Bereiche oberhalb von 70 mg/kg Zink blieben identisch. Erst nach der weiteren zufälligen Ausdünnung waren Unterschiede erkennbar. Nutzung und Niederschlag blieben zwar unverändert Hauptkriterium, allerdings fielen nun andere Flächen im Kanton mit Werten von über 70 mg/kg auf. Die Umgebung der Verzinkerei trat nicht mehr mit höheren Zinkgehalten hervor, dafür aber grossflächig Bereiche im Bezirk Arbon ganz im Osten des Kantons, sowie kleinere Gebiete westlich und östlich der Verzinkerei Bettwiesen.



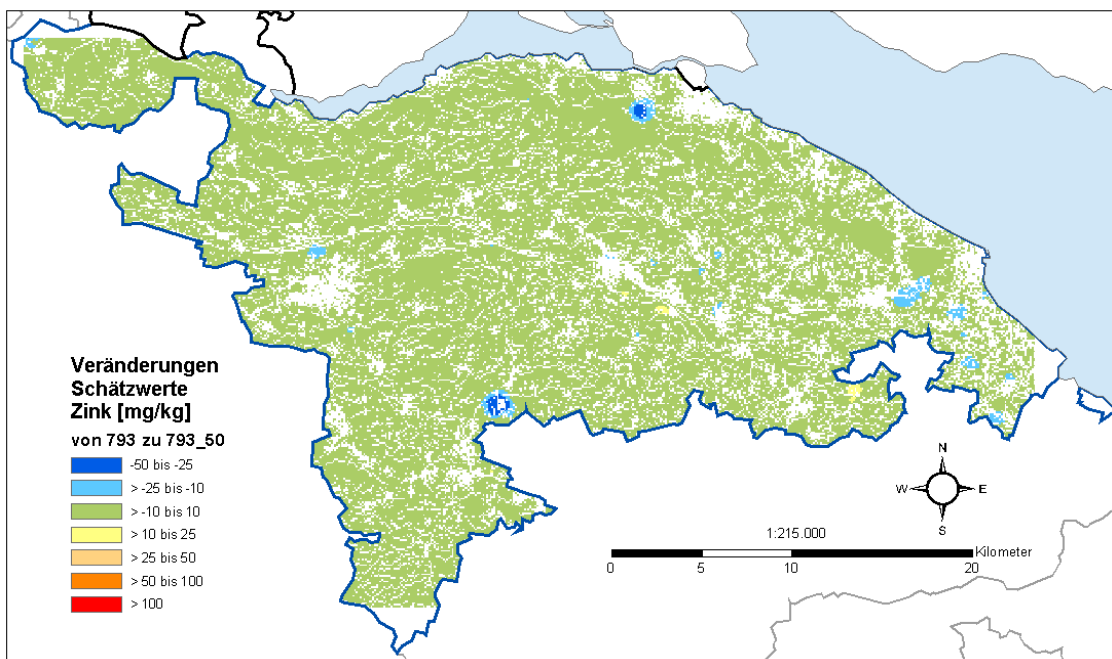
Figur 3-38: Interpolierte Zinkgehalte Zn386 Modell 1 (ohne einen auffälligen Wert), interpoliert mit Simple Kriging



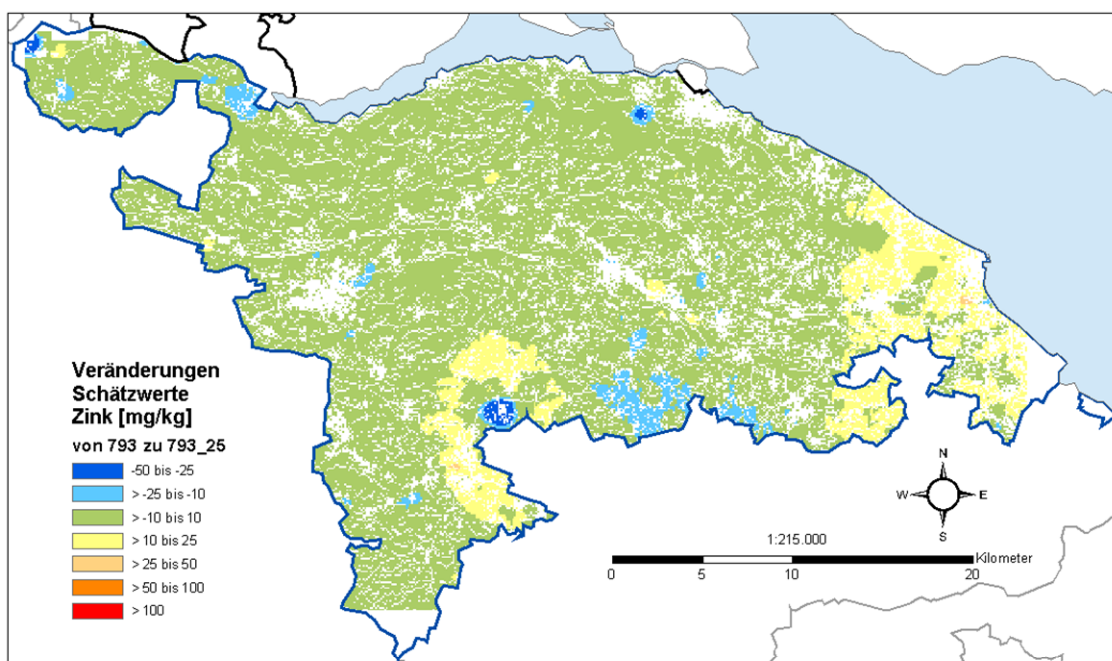
Figur 3-39: Interpolierte Zinkgehalte Zn193 Modell 1 (ohne 4 auffällige Werte), interpoliert mit Ordinary Kriging

Ein Vergleich der Karten mit den Schätzwerten der Interpolation mit 793 Standorten zeigte, wie schon aus dem Muster der Schätzwerte ersichtlich, insbesondere nach Ausdünnung von 50 % der Daten kaum Veränderung, die Schätzwerte lagen um ± 10 mg/kg in derselben Größenordnung (Figur 3-40 und Figur 3-41). Nur wenige Ge-

biete werden nun um > 10 mg/kg niedriger geschätzt, wie in der Umgebung der Verzinkei Bettwiesen. Der weitere Ausschluss von Daten führte zu stärkeren Veränderungen. Es dominierten zwar immer noch die Flächen mit einer Abweichung von ± 10 mg/kg, aber Gebiete mit deutlicher Veränderung der Schätzwerte um 10 - 25 % hin zu höheren Zinkgehalten (rund um die Verzinkei Bettwiese und im östlichen Kantonsteil) bzw. zu niedrigeren Gehalten (im südlichen Gebiet) nahmen zu.



Figur 3-40: Vergleich der interpolierten Zinkgehalte Zn386 mit Zn793 (dargestellt sind die Differenzen Zn 386 - Zn793)



Figur 3-41: Vergleich der interpolierten Zinkgehalte Zn193 mit Zn793 (dargestellt sind die Differenzen Zn193 – Zn793)

Prognoseintervall und Kreuzvalidierung

Das Prognoseintervall der um 50 % ausgedünnten Daten veränderte sich ebenfalls nicht merklich, der Wahrscheinlichkeitsbereich blieb im Mittel gleich (Anhang C- 3 und Anhang C- 6), auch wenn das Gesamtintervall durch die geringere Streuung in den Daten enger wurde. Gleiches galt für den 25 % Datensatz, allerdings wurde für diese Interpolation ein Pure-Nugget-Modell verwendet, welches die Werte per se stärker glättet (Anhang C- 7). Das Muster der Vorhersagebereiche allerdings änderte sich analog zum Muster der Schätzwerte deutlich. Es wurden für andere Gebiete und zudem weitaus grossflächiger Schätzwerte mit einer 95 %igen Wahrscheinlichkeit zwischen 100 mg/kg und 150 mg/kg vorhergesagt.

Der Vergleich der Schätzwerte mit den tatsächlichen Messwerten im Rahmen der Kreuzvalidierung zeigte letztlich ein ähnliches Bild. Die Ergebnisse von Datensatz 386 unterschieden sich kaum von denen von Datensatz 793 (Tabelle 3-20 und Tabelle 3-13). Die absoluten Abweichungen waren sogar ein wenig geringer, dafür wurde die Variabilität der Daten etwas stärker systematisch überschätzt. Die Fehler für die Interpolation des Datensatz 193 waren dagegen erwartungsgemäss grösser, allerdings bei weitem nicht so gross wie im Falle der 1290 Datensätze (vgl. Tabelle 3-17). Hierbei muss allerdings beachtet werden, dass sich die Kreuzvalidierung jeweils auf eine unterschiedliche Auswahl an Messungen bezieht.

Tabelle 3-20: Kenngrössen der Kreuzvalidierung für die Datensätze Zn386/193 (in log mg/kg)

Kenngrössen Kreuzvalidierung	Zn386, Modell 1	Zn193, Modell 1
ME	0.003138	-0.0000868
RMSSE	0.9424	1.079
MedE	0.0077	0.0071
MAD	0.0653	0.0738

Zusammenfassende Diskussion

Die Ausdünnung des Datensatz 793 um 50 % auf 386 Datensätze zeigte aufgrund der kaum verschlechterten Stützpunktumgebung wenig Unterschiede bezüglich der geschätzten Zinkgehalte. Das deutete darauf hin, dass eine Datendichte von etwa 0,5 Punkten/km² noch gute Ergebnisse liefert, sofern die Stützpunkte die Region gut abdecken. Allerdings muss bedacht werden, dass es sich um eine zufällige Ausdünnung der Daten handelte. Es ist denkbar, dass entscheidende Werte nicht oder gerade eben mittels der stratifizierten zufälligen Auswahl der Daten ausgeschlossen wurden. Dennoch können die Interpolationsergebnisse für den reduzierten Datensatz 386 als viel versprechend interpretiert werden, da generell das räumliche Muster (Schätzwerte) als auch die Genauigkeit der Schätzungen sich gegenüber 793 Stützpunkten kaum veränderte.

Der Datensatz 193 lieferte deutlich abweichende Ergebnisse mit einer geringeren Aussagegenauigkeit. Die im östlichen und südlichen Kantonsgebiet sowie rund um Bettwiesen grösseren Abweichungen zu den Schätzwerten des Datensatz 793 deuteten darauf hin, dass die Qualität der Schätzung gelitten hat. Hinzu kamen die höheren absoluten Abweichungen in der Kreuzvalidierung.

3.3.2 Datensatz Blei

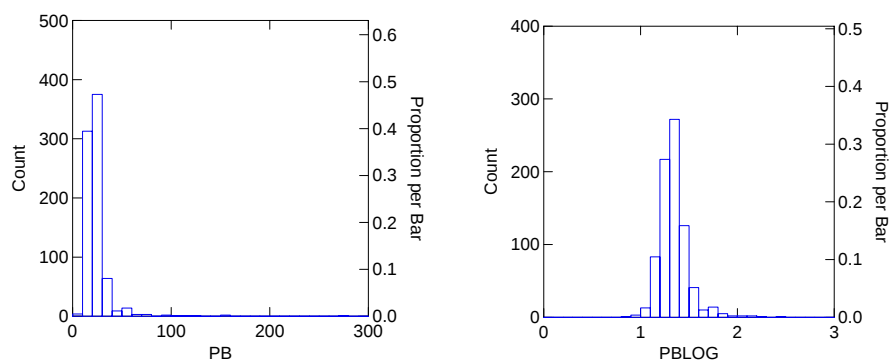
Zur Überprüfung der Vorgehensweise sollte die bislang ausschliesslich für Zink angewendete Methodik am Beispiel des Datensatz 793 (Einbeziehung der Verzinkereien und KVAs) auf einen zweiten Parameter übertragen werden. Ausgewählt wurde Blei als ein weiterer insbesondere atmosphärisch eingetragener Bodenschadstoff. Soweit sinnvoll wurden die Einstellungen hinsichtlich Regressionsmodell und Interpolation verwendet, die bereits für die 793 Zinkdaten angewendet wurden. Damit wurde eine bessere Vergleichbarkeit der Karten der Schätzwerte und der Evaluation gewährleistet.

3.3.2.1 Explorative Datenanalyse

Die Bleiwerte lagen im Mittel deutlich unterhalb des Richtwertes der VBBo von 50 mg/kg, der erst mit dem 99. Perzentil überschritten wurde. Dennoch wurde ebenfalls der Prüfwert von 200 mg/kg überschritten. Wie auch die Zinkgehalte waren die Bleimesswerte stark rechtsschief, nach Logarithmierung aber annähernd normalverteilt, wenn auch die Schiefe noch immer > 1 war (Tabelle 3-21, Figur 3-42).

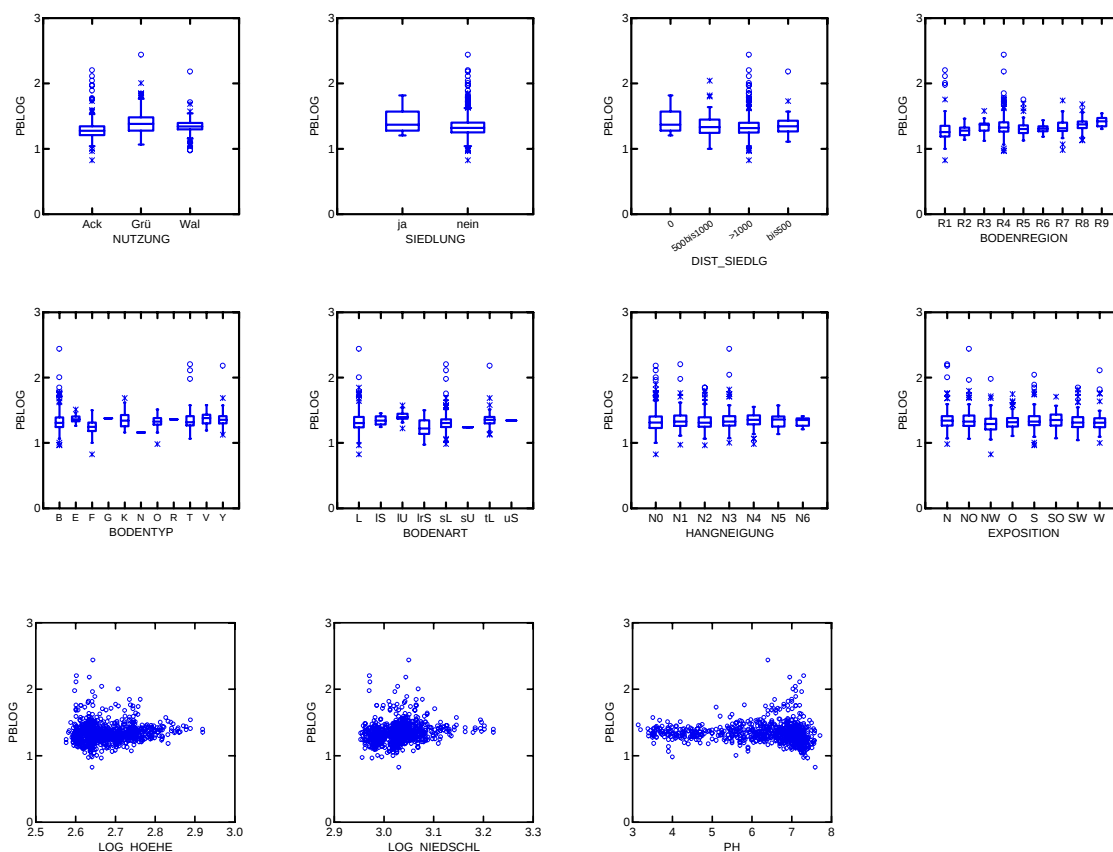
Tabelle 3-21: Allgemeine statistische Kenngrössen der Bleidaten Datensatz 793 (in mg/kg)

	Pb
Anzahl Datensätze	793
Minimum	6.7
Maximum	275.00
Median	21.00
Mittelwert	23.95
Standardabweichung	15.45
Variationskoeffizient	0.65
Schiefe	8.36
Exzess	106.01
25. Perzentil	17.90
75. Perzentil	25.40
95. Perzentil	39.00
99. Perzentil	85.84
	Pb(log)*
Anzahl Datensätze	793
Schiefe	1.74
Exzess	7.45



Figur 3-42: Verteilung der Bleimesswerte Datensatz 793 (links Pb-, rechts logPb-Werte in mg/kg)

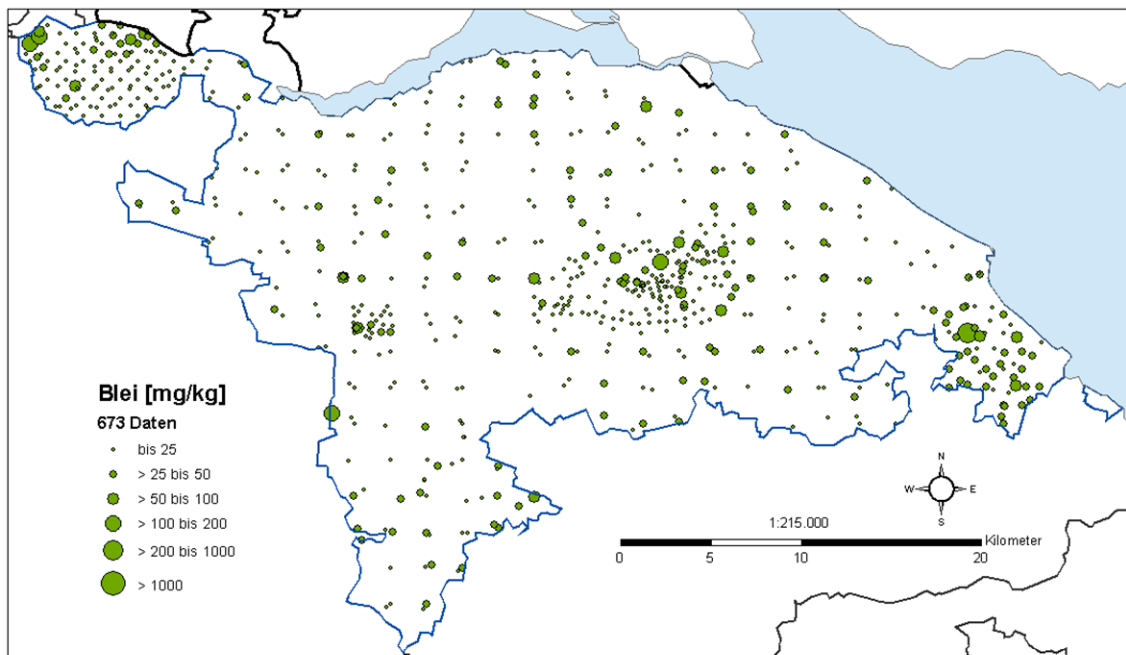
Im Gegensatz zu den logarithmierten Zinkdaten gab es zwischen den einzelnen Klassen der verschiedenen erklärenden Variablen weniger signifikante Differenzen (Figur 3-43). Vor allem die Nutzung und der pH-Wert näherten sich einem Niveau an. Dagegen schienen Parameter der BÜK, Bodentyp und Bodenart Einfluss zu haben und auch der Niederschlag, was es in der Regressionsanalyse zu überprüfen galt.



Figur 3-43: Univariate Statistik der erklärenden Variablen auf die Blei-Bodenbelastung Datensatz 793, logPb in mg/kg

Wie für Zink wurden an allen ausgewählten 793 Standorten die Bleigehalte im Boden bestimmt, die räumliche Verteilung war somit identisch. Die höchsten Bleimesswerte

fanden sich analog in den Bezirken Diessenhofen, Weinfelden und Arbon, sowie ebenfalls im Bereich der Verzinkerei Bettwiesen (Figur 3-44).



Figur 3-44: Bleigehalte an den Untersuchungsstandorten Datensatz 793

3.3.2.2 Räumliche Datenanalyse

Regressionsanalyse

Angestrebt war, möglichst die Einstellungen der Datenanalyse der Daten 793 zu verwenden, d.h. Regressionsmodell 1. Das hätte bedeutet, Nutzung, pH und Niederschlag ins Modell aufzunehmen. Wie sich aber bereits in Figur 3-43 angedeutet hatte, bedeutete ein solches Modell für Blei ein äusserst geringes Bestimmtheitsmass ($R^2 = 0,113$), d.h. es konnten lediglich 11 % der Varianzen in den Bleidaten erklärt werden. Nahm man daher doch wieder die Parameter der BÜK (Bodentyp, Bodenart) hinzu und liess pH aussen vor, ergab sich kein gutes, aber immerhin leicht besseres Ergebnis für das Modell ($R^2 = 0,17$). Da allerdings die Bodenart kaum etwas zur Verbesserung des R^2 beitrug, aber stattdessen die Anzahl Freiheitsgrade deutlich absenkte, wurde sie letztlich doch nicht berücksichtigt. Die resultierenden Koeffizienten für die Nutzung waren entsprechend kleiner und in diesem Falle sogar leicht positiv für die Variable Wald (0.026), für den Niederschlag blieb die Grössenordnung in etwa bestehen (0.573). Die Koeffizienten für die Bodentypen zeigten mit Werten von < 0.1 einen ähnlich geringen Einfluss wie die Nutzung im Verhältnis zum Niederschlage (Tabelle 3-22).

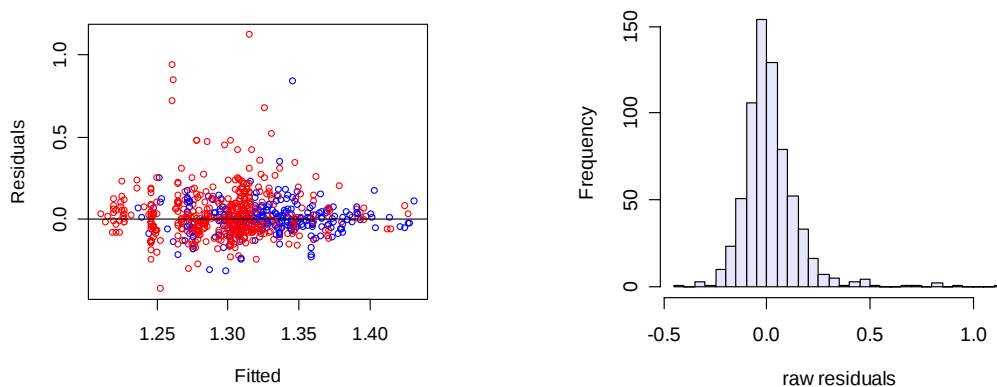
$$\text{Modell 3: } \log_{10}(\text{Zn}) = \alpha + \beta_1 * \text{Nutzung} + \beta_2 * \text{Bodentyp} + \beta_3 * \log_{10}(\text{Niederschlag}) + E_i \quad (3.3)$$

Tabelle 3-22: Regressionsmodell für die logarithmierten Bleigehalte Datensatz 793

Koeffizient	Kategorie	Pb793, Modell 3*
α		-0.444
β_1 Nutzung	AckGrü	0
	Wal	0.026
β_2 Bodentyp	Regosol	0
	Fluvisol	-0.040
	Rendzina	-
	Kalkbraunerde	0.037
	Braunerde	0.011
	Parabraunerde	0.002
	Saure Braunerde	0.047
	Braunerde-Pseudogley	0.019
	Braunerde-Gley	0.079
	Buntgley	-
	Fahlgley	-
	Halbmoor	-
$\beta_3 \log_{10}(\text{Niederschlag})$		0.573
Anzahl Freiheitsgrade		663
adjusted R ²		0.170

* einige Kategorien waren entweder nicht vorhanden oder Anzahl Zinkwerte pro Kategorie <15

Für Blei gab es erwartungsgemäss aufgrund der vorherigen Ergebnisse nicht mehr die strikte Trennung zwischen Acker/Grünland und Wald (Figur 3-45). Insgesamt streuten die Residuen gleichmässig um die Nulllinie, lediglich einige wenige höhere Residualwerte liessen die leichte Rechtsschiefe in den log-Residuen erkennen:

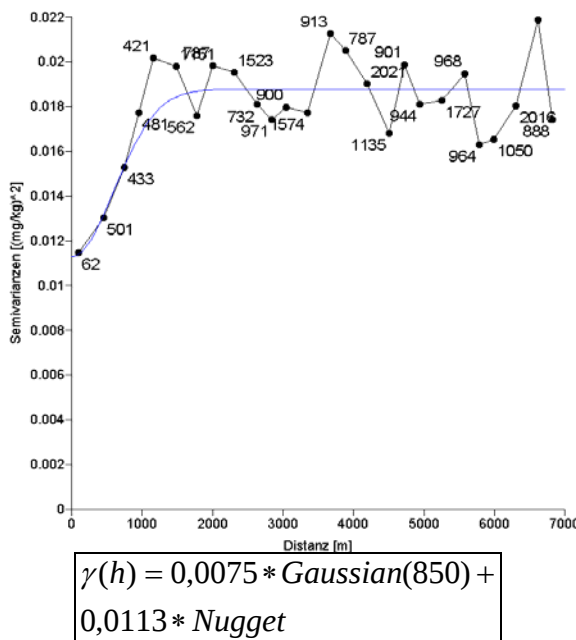


Figur 3-45: Streudiagramm der rohen Residuen gegen die angepassten Werte (Fitted) mit Nutzung Acker/Grünland (rot) und Wald (blau) (links), Histogramm der rohen Residuen (rechts), Pb793 Modell 3, Angaben in mg/kg

Variographie

Ebenfalls erwartungsgemässe liess sich der deterministische Anteil der Streuungen in den Daten mit dem verwendeten Regressionsmodell nicht abdecken, die Daten zeigten noch eine räumliche Abhängigkeit (Drift) (Figur 3-46). Diese konnte mit einer Gaussian-Variogrammfunktion am exaktesten beschrieben werden. Der Range der Daten lag bei 850 m bei einem 60°-Suchfächer in O-W-Richtung. Im Gegensatz zu den Vari-

ogrammen von Zink ergab sich – entsprechend des geringen R^2 – ein partieller Sill, der deutlich kleiner als der Nugget-Effekt war, d.h. es war ein kräftiges nicht erklärbares Grundrauschen vorhanden. Im Falle der Zinkvariogramme war der Nugget-Effekt höchstens gleich gross, was allerdings immer noch einem grossen Grundrauschen entspricht.

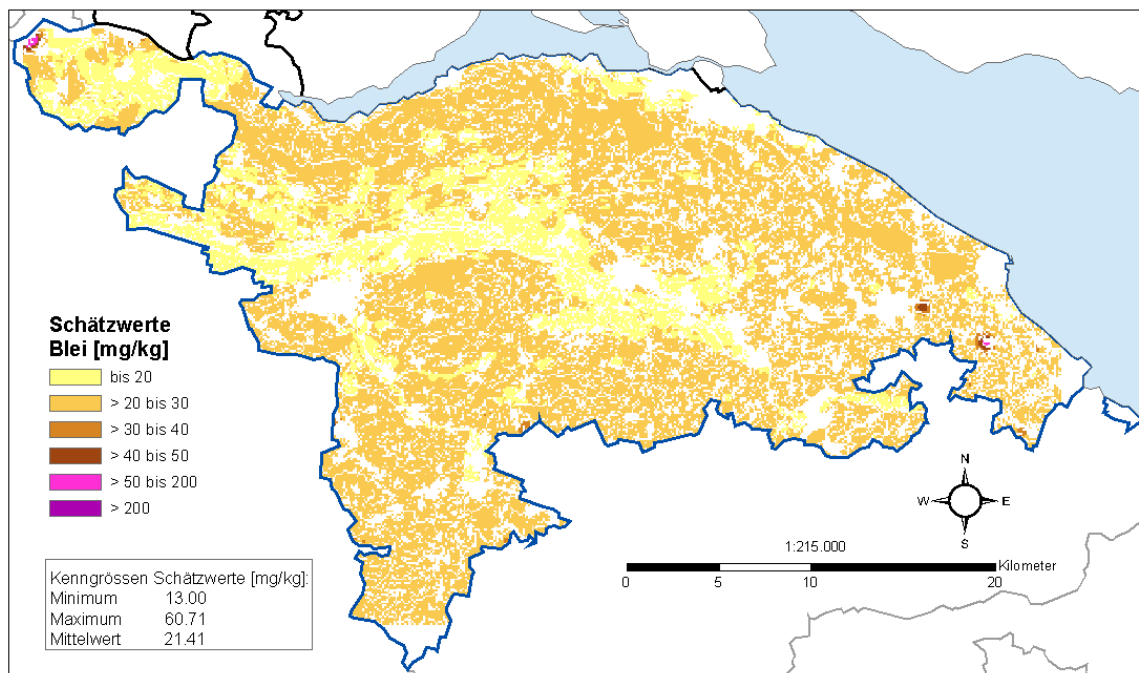


Figur 3-46: Semivariogramm Pb793 Modell 3

Interpolation

Die Interpolation wurde mit einem Simple Kriging durchgeführt. Die Messwerte wurden stark geglättet, das Maximum lag nur noch bei 61 mg/kg anstelle von 275 mg/kg tatsächlich gemessenem Bleigehalt. Die hohen gemessenen Werte wurden demnach als nicht repräsentativ für das jeweils betrachtete Gebiet ausgewiesen. Zudem wurde mittels des Simple Kriging von einem bekannten Erwartungswert ausgegangen, was ebenfalls eine Glättung der Werte bedingt. Dennoch wurden auch für diesen Parameter in der Gemeinde Schlatt, im äussersten Nordwesten des Kantons, sowie im Umfeld der KVA Hefenhofen kleine Gebiete mit Überschreitungen des Richtwertes ausgewiesen (Figur 3-47).

Das Muster der Schätzwerte wurde in erster Linie durch den hohen Regressionskoeffizienten des Niederschlags geprägt. Dort, wo die mittleren Jahresniederschläge geringer sind, v. a. im Tal der Thur und im Bezirk Diessenhofen, lagen die interpolierten Bleigehalte bei unter 20 mg/kg. Zudem haben sich im Flusstal erwartungsgemäss überwiegend Fluvisole ausgebildet, für die im Regressionsmodell ein negativer Koeffizient vergeben wurde, sprich mit geringeren Gehalten auf diesem Bodentyp gerechnet wurde. Im übrigen Kantonsgebiet wurden grossflächig Gehalte von 20-40 mg/kg vorhergesagt.



Figur 3-47: Interpolierte Bleigehalte Pb793 Modell 3, interpoliert mit Simple Kriging

Prognoseintervall und Kreuzvalidierung

Auch wenn die Schätzwerte fast ausschliesslich auf unter 30 mg/kg geschätzt wurden, reichte in vielen Bereichen das 95 % Prognoseintervall bis an den Richtwert für Blei heran (Anhang C- 8). Zu den Rändern des Kantons hin, insbesondere im Süden wurden Prognoseintervalle bis etwa 100 mg/kg ausgegeben.

Die Kreuzvalidierung zeigte, dass die Variabilität in den Daten kaum systematisch unter- oder überschätzt wurde, der RMSSE war nahezu 1 (Tabelle 3-23). Der MAD war deutlich niedriger als der für die Zink-Interpolationen, ist aber nicht mit diesen zu vergleichen, da die Bleigehalte grundsätzlich geringer sind und der MAD die absoluten Fehler darstellt und keine relativen.

Tabelle 3-23: Kenngrössen der Kreuzvalidierung für den Datensatz Pb793 (in log mg/kg)

Kenngrössen Kreuzvalidierung	Pb793, Modell 3
ME	-0.003769
RMSSE	0.9997
MedE	0.0137
MAD	0.0559

Zusammenfassende Diskussion

Die Karte der Bleigehalte zeigte Schadstoffgehalte in ähnlich geringen Konzentrationen bezüglich der VBBo-Beurteilungswerte, wie es für Zink der Fall war; die Richtwerte wurden selten überschritten. Allerdings handelte es sich ebenfalls in erster Linie um die Hintergrundbelastung in den Böden, ohne Berücksichtigung der Gehalte in Siedlungsgebieten, Obst- und Gemüsebau und der Rebberge. Im Gegensatz zu Zink liess sich mittels der vorhandenen erklärenden Variablen wesentlich weniger Varianz der Messwerte erklären. Dies führte im Vergleich zu Zink zum etwas breiteren Vorhersageintervall. Das Muster der Schätzwerte war dennoch eindeutig auf die in Regression verwendeten erklärenden Variablen zurückzuführen. Die Kreuzvalidierung zeigte ein sehr positives Bild. Die Werte wurden weder deutlich unter- noch überschätzt. Der Mittelwert der Fehler und der Median deuteten auf eine gute Schätzung hin.

4 Schlussfolgerungen, Empfehlungen und Ausblick

Mit den interpolierten Bodenschadstoffkarten für Zink und Blei wurden Aussagen entwickelt, die die grossräumige diffuse Hintergrundbelastung der Böden widerspiegeln. Es konnten und sollten keine parzellenscharfen Informationen dargestellt werden.

Die Voraussetzungen im Kanton Thurgau für eine Bodenschadstoffkarte sind hinsichtlich Quantität und Qualität der vorhandenen Daten als aussergewöhnlich gut zu bewerten. Wichtigster Baustein sind Rasteruntersuchungen im land- und forstwirtschaftlichen Bereich. Mittels der in der Datenbank dokumentierten Metainformationen und des für den Kanton vorhandenen Kartenmaterials stehen weit mehr als die notwendigen Mindestinformationen für die Bewertung der gemessenen Schadstoffgehalte zur Verfügung.

Die gewählte Interpolationsmethodik eines Krigings in Verbindung mit einer Regression erfordert eine vertiefte statistische Analyse. Die in Frage kommenden erklärenden Variablen liegen nicht immer als Metainformationen direkt vor, sondern müssen aus Kartengrundlagen aufbereitet und den Messpunkten zugeordnet werden. Im Falle der Zinkgehalte konnte aber gezeigt werden, dass die Daten eindeutig einen Trend besaßen (R^2 bis 0,57) und damit der grosse Aufwand gerechtfertigt war, um den deterministischen Anteil bestmöglich vom stochastischen zu trennen.

Die Verwendung von rohen oder (je nach Tuning-Konstante mehr oder weniger stark) gewichteten Residuen als Ergebnis der Regression kann die Interpolationsergebnisse stark verändern und ist daher je nach Fragestellung zu entscheiden. Wenn es allein um die Einschätzung der zu erwartenden Zink-Hintergrundbelastung im land- und forstwirtschaftlichen Bereich geht, sollte die Karte der 673 Datensätze mit Verwendung der Parameter der BÜK und den gewichteten Residuen verwendet werden. Sie liefert die exakteste Schätzung. Mit Einbeziehung höherer Werte, die räumlich aber immer noch korrelieren (Datensatz 793), ist die starke Glättung der Daten nicht mehr wünschenswert. In diesem Fall sollten die weniger robusten rohen Residuen vorgezogen werden, die noch mehr Information beinhalten. Es bedarf dann aber einer sorgfältigen Betrachtung auffälliger Werte. Welche Auswirkungen bereits 4 einzelne Messwerte auf die räumliche Autokorrelation benachbarter Werte und damit auf die Variogrammfunktion und Interpolation eines umfangreichen Datensatzes haben können, zeigte deutlich die Auswertung der 793 Datensätze.

Als erklärende Variablen für die Zinkgehalte im Boden im Kanton Thurgau haben sich v. a. Nutzung, pH-Wert, Niederschlag, Bodentyp und Bodenart herausgestellt. Die Siedlungsnähe erklärt ebenfalls räumliche Unterschiede der Zinkkonzentrationen im Boden. Dabei kann die Einbeziehung oder der Ausschluss einzelner erklärender Variablen das Ergebnis, vom Modell bis zur Aussagegenauigkeit der Karte, entscheidend beeinflussen.

Die durchgeführten Interpolationen zeigten, dass das 95 % Prognoseintervall der geschätzten Zinkgehalte im land- und forstwirtschaftlichen Bereich im Kanton Thurgau

die Richtwerte der VBBo nahezu nicht überschreitet. Im Umfeld der Verzinkereien oder KVAs werden hingegen die Richtwerte lokal deutlich überschritten. Für die weiteren Nutzungen im Siedlungsbereich, Obst- und Gemüsebau oder in den Rebbergen konnte keine gesicherte Aussage getroffen werden. Aufgrund der hohen Streuung der Messwerte innerhalb dieser Nutzungsklassen resultierte neben dem mit dem Regressionsmodell geschätzten deterministischen Anteil ein grosser stochastischer Anteil der Streuung der Zinkgehalte. Die Residuen erfüllten die Annahme der Stationarität nur bedingt. Dadurch war eine Grundvoraussetzung der Interpolationsmethode verletzt, was zu einer Überschätzung von Zinkgehalten in der Umgebung dieser Nutzungen führte und zu einem sehr breiten Prognoseintervall bis hin zu einer oberen Grenze von 3000 mg/kg. Die Schätzgenauigkeit hatte damit im Vergleich zu den Karten der Datensätze 673 und 793 erheblich abgenommen. Wenn im Zuge der Regression der deterministische Anteil nicht vollständig erfasst werden kann, ist der Einbezug von Standorten mit erhöhten Schadstoffgehalten nur dann fachlich sinnvoll, wenn es sich um räumlich korrelierende Messwerte (wie bei den Verzinkereien oder KVAs) handelt oder aber genügend Messpunkte die Anomalien in lokal eng begrenzten Gebieten abdecken. Zur Einschätzung der Nutzungen Obst-/Gemüsebau und Rebberge sollte die Methode ohne die Einbeziehung der Siedlungsgebiete separat angewendet werden. Zusammen mit den Siedlungsbereichen werden die Effekte dieser Nutzungen überlagert.

Ziel der Schätzkarten ist eine kantonale Übersichtskarte zu Schadstoffen im Boden, d.h. die Ausweisung von Gebieten, in denen tendenziell mit höheren Schadstoffgehalten gerechnet werden muss. Für kleinräumige Massstäbe bis hin zu parzellenscharfen Fragestellungen der Bodenbelastung lassen die Karten **keine gesicherten Aussagen** zu. Mit der gewählten Gridgrösse von 1 ha (zurückzuführen auf die Auflösung der Arealstatistik) liessen sich die räumlich differenzierten Muster mit unterschiedlicher Grösse der Prognoseintervalle gut darstellen. Für eine kantonsweite Aussage sollte aber zukünftig versucht werden, die Einheiten auf z.B. 500 x 500 m oder 1 x 1 km² zu aggregieren. In der vorliegenden Studie musste noch darauf verzichtet werden, da es für eine Aggregation der Nutzungsdaten entsprechender „Spielregeln“ bedurft hätte. Abgebildet werden sollten die Karten in einem Massstab von maximal 1:50.000, sinnvoller nur bis 1:100.000. Mit diesen Massstäben sind alle notwendigen Details für die Ausweisung der diffusen Hintergrundbelastung gut darstellbar, erlauben aber gleichzeitig keine zu kleinräumigen Interpretationen.

Um eine ähnlich gute Schätzung der Schadstoffgehalte auch in anderen Kantonen zu erzielen bedarf es eines geeigneten Beprobungsmusters und einer Homogenität der Daten. Für das Beprobungsmuster ist neben der Anzahl der Messpunkte deren Verteilung über das Untersuchungsgebiet ausschlaggebend. Eine gleichmässige Verteilung der Beprobungsstandorte über das Untersuchungsgebiet mit einzelnen verdichteten Teilgebieten ist für die räumliche Analyse günstiger als einzelne intensiv beprobte Teilgebiete. Allerdings sind Rasterbeprobungen in den Kantonen – wie im Kanton Thurgau – die Ausnahme, was die Ausgangslage erschwert. Welche Datendichte es dann in einem Kanton bedarf, ist aus den Ergebnissen noch nicht zweifelsfrei abzuleiten. Die Ergeb-

nisse im Kanton Thurgau deuten aber darauf hin, dass auch noch mit der Hälfte der Daten ($\sim 0,5$ Punkte/km²) ähnlich gute Ergebnisse erzielt werden können wie mit zuvor 0,9 Punkte/km², sofern die Daten räumlich gut verteilt sind. Mit nur noch 0,22 Punkten/km² scheint die Qualität der Schätzung klar abzunehmen. Die Ergebnisse können aber nur als ein erstes Indiz angesehen werden, welches nicht ohne weiteres auf andere Kantone übertragen werden kann. Für eine zuverlässigere Aussage sollten weitere Testdatensätze untersucht werden. Als zweites Kriterium sind die vorhandenen Metainformationen von immenser Bedeutung. Ohne die nötigen Mindestangaben ist ein Messwert für eine Interpolation nutzlos, da die Homogenität der Daten als Grundvoraussetzung einer Interpolation nicht gewährleistet werden kann. Daher sollte unbedingt die nötige Zeit zur Dokumentation der benötigten Metadaten aufgewendet werden. Nur so ist darüber hinaus eine langfristige Nutzung der mit viel Geld erhobenen Messdaten gesichert.

Die entwickelte Methodik hat sich als ausreichend flexibel für die Übertragung auf Blei herausgestellt. Die räumliche Verteilung und der (wenn auch geringere) Zusammenhang mit den erklärenden Variablen der Bleigehalte im Boden konnte erfasst und dargestellt werden. Die Kreuzvalidierung zeigte eine gute Schätzung der Bleidaten ohne systematische Unter- oder Überschätzung.

Im Anschluss an diese Studie soll kurzfristig die entwickelte Methodik auf die anderen relevanten Schwermetalle im Kanton Thurgau (Cd, Cr, Cu, Hg und Ni) angewendet und wo nötig verbessert werden. Insbesondere die Besonderheiten der Nutzungen im Siedlungsbereich, Obst-, Gemüse- und Rebbau und deren Auswirkungen auf umgebende land- und forstwirtschaftlich genutzte Flächen sollen vertieft analysiert werden. Für eine vereinfachte Auswertung sollen Auswertungstools im Statistikprogramm „R“ mit der geostatistischen Erweiterung GStat erstellt werden. Mit diesem Tool kann ein Kriging in Verbindung mit einer Regression integriert gelöst werden. Der in dieser Studie verwendete Geostatistical Analyst von Esri verfügt nicht über die notwendigen Funktionen.

Mit weiteren räumlichen Analysen von Bodengehalten im Schweizer Mittelland können zusätzliche wertvolle Erfahrungen zur Interpolation und Kartierung gesammelt werden. Ist die Erstellung einer Schadstoffkarte in einem Kanton aufgrund der Qualität der Datenlage nur schlecht oder gar nicht möglich, erlaubt das Vorgehen Aussagen zur Optimierung des Messnetzes, d.h. wo weitere Messungen wertvoll sind.

Eine wichtige Rolle wird in diesem Zusammenhang zukünftig die im Aufbau begriffene nationale Bodenschadstoffdatenbank (NABODAT) spielen. Mit dieser sollen valide interpretierbare, d.h. mit den nötigen Metainformationen versehene Daten, zusammengestellt und vor allem für die Kantone verfügbar gemacht werden.

5 Literaturverzeichnis

- Dobler, L., Hinterding, A., Müller, A., Gerlach, N. & Gabel, F., 2003. Geostatistische und statistische Methoden und Auswerteverfahren für Geodaten mit Punkt- bzw. Flächenbezug, Abschlussbericht Teil 3: Empfehlungen für die Anwendung statistischer und geostatistischer Methoden zur flächenbezogenen Auswertung von Daten über Stoffgehalte in Böden. Bund/Länder-Arbeitsgemeinschaft Bodenschutz (LABO), 55 pp.
- EJPD, EVD & EDI, 1980. Bodeneignungskarte der Schweiz 1:200.000. Grundlagen für die Raumplanung. EDMZ, Bern.
- ESRI, 2001. Using ArcGIS Geostatistical Analyst. ESRI, 300 pp.
- FAL, 1997. Kartieren und Beurteilen von Landwirtschaftsböden. Schriftenreihe der FAL Nr. 24. Eidg. Forschungsanstalt für Agrarökologie und Landbau (FAL), Zürich.
- Goovaerts, P., Webster, R. & Dubois, J.-P., 1997. Assessing the risk of soil contamination in the Swiss Jura using indicator geostatistics. *Environmental and Ecological Statistics*, 1997(4): pp. 31-48.
- Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. & Stahel, W.A., 1986. *Robust Statistics: The Approach Based on Influence Functions*. Wiley, N.Y.
- Heinrich, U., 1992. Untersuchungen zur Validität flächenhafter Schätzungen diskreter Messungen kontinuierlicher raumzeitlicher Prozesse. Dissertation. Christian-Albrechts-Universität, Kiel. 124 pp.
- Hengl, T., Heuvelink, G.B.M. & Rossiter, D.G., 2005. About regression-kriging: from theory to interpretation of results. (Preprint submitted to *Computers and geosciences*)
- Hengl, T., Heuvelink, G.B.M. & Stein, A., 2003. Comparison of kriging with external drift and regression-kriging. Technical note. ITC, 17 pp. (Available on-line at http://www.itc.nl/library/Academic_output/)
- Hengl, T., Heuvelink, G.B.M. & Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, 120 (2004): pp. 75-93.
- Hinterding, A., Müller, A., Gerlach, N. & Gabel, F., 2003a. Geostatistische und statistische Methoden und Auswerteverfahren für Geodaten mit Punkt- bzw. Flächenbezug, Abschlussbericht Teil 1: Grundlagen. Bund/Länder-Arbeitsgemeinschaft Bodenschutz (LABO), 145 pp.
- Hinterding, A., Müller, A., Gerlach, N. & Gabel, F., 2003b. Geostatistische und statistische Methoden und Auswerteverfahren für Geodaten mit Punkt- bzw. Flächenbezug, Abschlussbericht Teil 2: Auswertbeispiele. Bund/Länder-Arbeitsgemeinschaft Bodenschutz (LABO), 114 pp.
- Huber, P.J., 1981. *Robust Statistics*. Wiley, N.Y.
- Isaaks, E.H. & Srivastava, R.M., 1989. *An Introduction to Applied Geostatistics*. Oxford University Press, 561 pp.
- Keller, T. & Desales, A., 2001. Böden der Schweiz – Schadstoffgehalte und Orientierungswerte (1990-1996). Umwelt-Materialien Nr. 139. Bundesamt für Umwelt, Wald und Landschaft (BUWAL), CH-3003 Bern. 115 pp.
- Lehmann, R., 1993. Schwermetallgehalte der Böden in der Umgebung der Kehrlichtverbrennungsanlagen Hefenhofen und Müllheim. Amt für Umweltschutz und Wasserwirtschaft, Bodenschutzfachstelle Kanton Thurgau, Frauenfeld. 41 pp.

- Lehmann, R., 1995a. Schwermetallbelastung in der Umgebung der 3 Verzinkereien Bettwiesen, Egnach und Wellhausen - Teilbericht "Ergebnisse der Boden-, Pflanzen- und Trinkwasseruntersuchungen". Amt für Umweltschutz und Wasserwirtschaft, Bodenschutzfachstelle Kanton Thurgau, Frauenfeld. 38 pp.
- Lehmann, R., 1995b. Schwermetallbelastung in der Umgebung der 3 Verzinkereien Bettwiesen, Egnach und Wellhausen - Zusammenfassung der Untersuchungsergebnisse. Amt für Umweltschutz und Wasserwirtschaft, Bodenschutzfachstelle Kanton Thurgau, Frauenfeld. 19 pp.
- Lehmann, R., 1995c. Schwermetallbelastung von Böden in Siedlungen. Amt für Umweltschutz und Wasserwirtschaft, Bodenschutzfachstelle Kanton Thurgau, Frauenfeld. 8 pp. (Interner Bericht)
- Lehmann, R., 1996. Schwermetallbelastung der Böden entlang von Strassen. Amt für Umweltschutz und Wasserwirtschaft, Bodenschutzfachstelle Kanton Thurgau, Frauenfeld. 6 pp. (Interner Bericht)
- Lehmann, R., 1998. Kupferbelastung in Böden bestehender und ehemaliger Rebberge am Ottenberg. Amt für Umweltschutz und Wasserwirtschaft, Bodenschutzfachstelle Kanton Thurgau, Frauenfeld. 11 pp. (Interner Bericht)
- LUA NRW, 2000. Leitfaden zur Erstellung digitaler Bodenbelastungskarten - Teil I: Aussenbereiche. LUA-Merkblätter 24. Landesumweltamt Nordrhein-Westfalen (LUA NRW), Essen. 115 pp.
- Meuli, R., 1997. Geostatistical Analysis of Regional Soil Contamination by Heavy Metals. Diss. ETH No. 12121. Swiss Federal Institute of Technology Zürich, 191 pp.
- Meuli, R., Schulin, R. & Webster, R., 1998. Experience with the replication of regional survey of soil pollution. *Environmental Pollution*, 101: pp. 311-320.
- Moyeed, R.A. & Papritz, A., 2002. An empirical comparison of kriging methods for nonlinear spatial point prediction. *Mathematical Geology*, 34: pp. 365-386.
- Nielsen, D.R. & Wendroth, O., 2003. Spatial And Temporal Statistics - Sampling Field Soils and their Vegetation. *GeoEcology textbook*. Catena Verlag GmbH, 398 pp.
- Papritz, A., 2003. Einführung in Geostatistik und Stichprobentheorie. ETH Zürich, Institut für Terrestrische Ökologie, Schlieren. 117 pp. (Vorlesungsskript)
- Papritz, A. & Dubois, J.-P., 1999. Mapping Heavy Metals in Soil by (Non-)Linear Kriging: An empirical validation. Kluwer Academic Publishers.
- Papritz, A. & Moyeed, R.A., 1999. Linear and Non-Linear Kriging Methods: Tools for Monitoring Soil Pollution. In: Barnett, V., Stein, A. and Turkman, K.F. (Editors), *Statistics for the Environment 4: Pollution Assessment and Control*. John Wiley & Sons, Ltd, pp. 303-336.
- Papritz, A. & Stein, A., 1999. Spatial prediction by linear kriging. In: Stein, A., van der Meer, F. and Gorte, B. (Editors), *Spatial Statistics for Remote Sensing*. Kluwer Academic Publishers, pp. 83-113.
- Papritz, A. & Webster, R., 1995a. Estimating Temporal Change in Soil Monitoring: I. Statistical Theory. *European Journal of Soil Science*, 46(1): pp. 1-12.
- Papritz, A. & Webster, R., 1995b. Estimating Temporal Change in Soil Monitoring: II. Sampling from Simulated Fields. *European Journal of Soil Science*, 46(1): pp. 13-27.
- Presler, J., Zürcher, M. & Kaufmann, G., 2005. Bodenübersichtskarte Kanton Thurgau: Schlussbericht. Amt für Umwelt des Kantons Thurgau, Amt für Raumplanung des Kantons Thurgau, Frauenfeld. 104 pp. (vorläufige Version)

- Rehbein, K., 2004. Machbarkeitsstudie für eine nationale Bodenschadstoff-Datenbank (NABODAT) mit GIS-Anbindung. Agroscope FAL Reckenholz, Eidg. Forschungsanstalt für Agrarökologie und Landbau, Zürich. 97 pp. (interner Bericht)
- Rutz, W., 1995. Bodenuntersuchungen im Zusammenhang mit der Sanierung des Lützelzelmurg-Viaduktes der N1 bei Aawangen 1993/1994. Amt für Umweltschutz und Wasserwirtschaft, Bodenschutzfachstelle Kanton Thurgau, Frauenfeld. 11 pp. (Interner Bericht)
- Rutz, W., 1997a. Schwermetallgehalte von Böden in der Umgebung der Abwasserreinigungsanlage Weinfelden. Amt für Umweltschutz und Wasserwirtschaft, Bodenschutzfachstelle Kanton Thurgau, Frauenfeld. 9 pp. (Interner Bericht)
- Rutz, W., 1997b. Zustand von Hausgartenböden. Amt für Umweltschutz und Wasserwirtschaft, Bodenschutzfachstelle Kanton Thurgau, Frauenfeld. 8 pp.
- Scheffer, F. & Schachtschabel, 2002. Lehrbuch der Bodenkunde. 15. Auflage. Spektrum Akademischer Verlag, Heidelberg, 593 pp.
- Schulin, R., Webster, R., Meuli, R., Atteia, O., Dubois, J.-P., Steiger von, B. & Védý, J.-C., 1994. Regional Soil Contamination Surveying. Environmental Documentation No. 25 - Soil. Federal Office of Environment, Forests and Landscape (FOEFL), Berne. 60 pp.
- Schwarz, S., Dvorak, A., Falkner, T. & Riss, A., 1994. Einrichtung eines Bodeninformationssystems in Österreich. UBA-BE-017. Bundesministerium für Umwelt, Jugend und Familie, Wien.
- Stahel, W.A., 2002. Statistische Datenanalyse. 4. Auflage. Vieweg Verlag, Braunschweig.
- Steiger von, B., Webster, R., Schulin, R. & Lehmann, R., 1996. Mapping heavy metals in polluted soil by disjunctive kriging. *Environmental Pollution*, 94(2): pp. 205-215.
- VBBo, 1998. Verordnung über Belastungen des Bodens (VBBo) vom 1. Juli 1998. SR 814.12.
- Venables, W.N. & Ripley, B.D., 1994. Modern Applied Statistics with S-Plus. Statistics and Computing. Springer, 462 pp.
- Webster, R. & Oliver, M., 2001. Geostatistics for Environmental Scientists. Statistics in Practice. John Wiley & Sons, Ltd, 271 pp.

Anhang

Anhang A: Eliminierte Datensätze: Doppelte Werte pro Standort

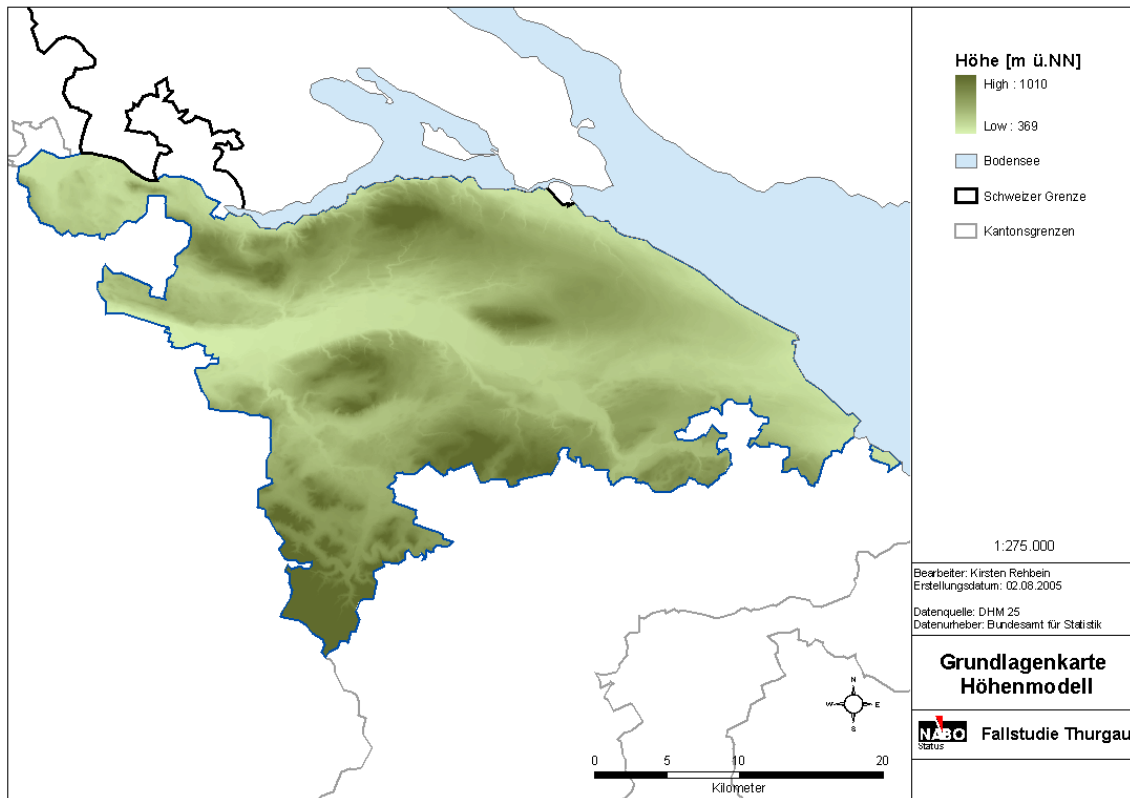
CASE	ProjID	Projekt	StaoID	ProbID	Grund für Eliminierung
Standorte im gleichen Projekt nachgemessen:					
570	2	P94SIE	281	863	Stao wurde nachgemessen
572	2	P94SIE	285	683	Stao wurde nachgemessen
713	2	P94SIE	1638	595	Nachmessung eliminiert (anderes Labor, fast gleiche Werte) nur Textur und Corg übernommen
1513	13	P99BST	280	1965	Stao wurde nachgemessen
1517	13	P99BST	285	1963	Stao wurde nachgemessen
1520	13	P99BST	286	1964	Stao wurde nachgemessen
1548	14	P94ARA MT	733	2003	Nachmessung eliminiert (weniger Parameter, fast gleiche Werte)
1557	14	P94ARA MT	813	1975	Nachmessung eliminiert (weniger Parameter, fast gleiche Werte)
1561	14	P94ARA MT	820	1971	Nachmessung eliminiert (weniger Parameter, fast gleiche Werte)
1565	14	P94ARA MT	826	1981	Nachmessung eliminiert (weniger Parameter, fast gleiche Werte)
1572	14	P94ARA MT	843	1977	Nachmessung eliminiert (weniger Parameter, fast gleiche Werte)
1574	14	P94ARA MT	852	1985	Nachmessung eliminiert (weniger Parameter, fast gleiche Werte)
1580	14	P94ARA MT	864	1999	Nachmessung eliminiert (weniger Parameter, fast gleiche Werte)
1582	14	P94ARA MT	867	1990	Nachmessung eliminiert (weniger Parameter, fast gleiche Werte)
1592	14	P94ARA MT	934	1996	Nachmessung eliminiert (weniger Parameter, fast gleiche Werte)
1595	14	P94ARA MT	939	2005	Nachmessung eliminiert (weniger Parameter, fast gleiche Werte)
1862	25	P93VZB-G	616	2370	Stao wurde nachgemessen
Standorte in späteren Projekten nachgemessen:					
37	1	P95RAS	166	4	in P99BST überprüft
118	1	P95RAS	379	261	in P99BST überprüft
546	2	P94SIE	235	691	in P99BST überprüft, nur Textur und Corg übernommen
568	2	P94SIE	272	901	Probenart Bohrg(1)
571	2	P94SIE	285	683	in P99BST überprüft
593	2	P94SIE	1131	643	in P99BST überprüft, nur pH, Textur und Corg übernommen
607	2	P94SIE	1179	649	in P99BST überprüft, nur pH, Textur und Corg übernommen
611	2	P94SIE	1194	547	Probenart Bohrg(1)
624	2	P94SIE	1252	546	Probenart Bohrg(1)
651	2	P94SIE	1427	548	Probenart Bohrg(1)
678	2	P94SIE	1489	717	in P99BST überprüft, nur pH, Textur und Corg übernommen
696	2	P94SIE	1588	545	Probenart Bohrg(1)
1414	11	P98ARBON	1619	1802	Probenart Bohrg(3)
1420	11	P98ARBON	1626	1834	Probenart Bohrg(3)

Anhang A: Eliminierte Datensätze: Doppelte Werte pro Standort (Fortsetzung)

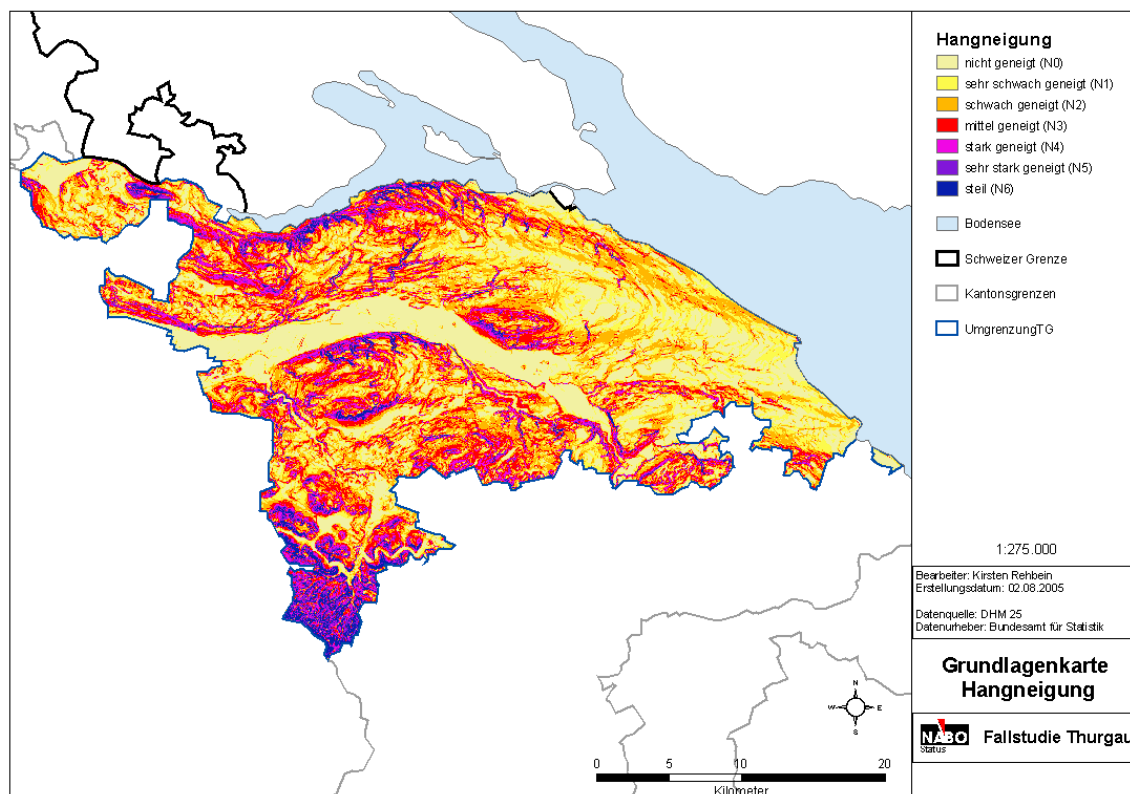
CASE	ProjID	Projekt	StaoID	ProbID	Grund für Eliminierung
1432	11	P98ARBON	1638	1820	als Mittelwert in Case 714 eingeflossen
1433	11	P98ARBON	1638	1832	Probenart Bohrg(3)
1438	11	P98ARBON	1642	1833	Probenart Bohrg(3)
1449	11	P98ARBON	1652	1801	Probenart Bohrg(3)
1516	13	P99BST	285	1955	Stao 2x nachbeprob, Werte in Case 1518 als Mittelwerte eingeflossen
1848	23	P93VZE	1502	2349	als Mittelwert in Case 995 eingeflossen
1863	25	P93VZB-G	616	2370	Probenart Bohrg(4), aber Pb, Cr, Cu, Ni übernommen
1864	25	P93VZB-G	637	2382	als Mittelwert in Case 1617 eingeflossen
1866	25	P93VZB-G	647	2381	als Mittelwert in Case 1623 eingeflossen
1886	27	P93N1 AW	267	2402	Werte nach Brückensanierung eliminiert
1889	27	P93N1 AW	271	2400	Werte nach Brückensanierung eliminiert
1891	27	P93N1 AW	277	2406	Werte nach Brückensanierung eliminiert
1893	27	P93N1 AW	287	2416	Werte nach Brückensanierung eliminiert
1895	27	P93N1 AW	289	2418	Werte nach Brückensanierung eliminiert
1897	27	P93N1 AW	292	2414	Werte nach Brückensanierung eliminiert
1922	31	P00WAN	226	2439	Probenart Bohrg(3)
1928	33	P97RAS 2	28	2469	Probenart Bohrg(1)
1929	33	P97RAS 2	95	2468	Probenart Bohrg(1)
1930	33	P97RAS 2	446	2483	Probenart Bohrg(1)
1931	33	P97RAS 2	470	2470	Probenart Bohrg(1)
51	Cases bzw. Datensätze eliminiert				

Anhang B: Grundlagenkarten Kanton Thurgau

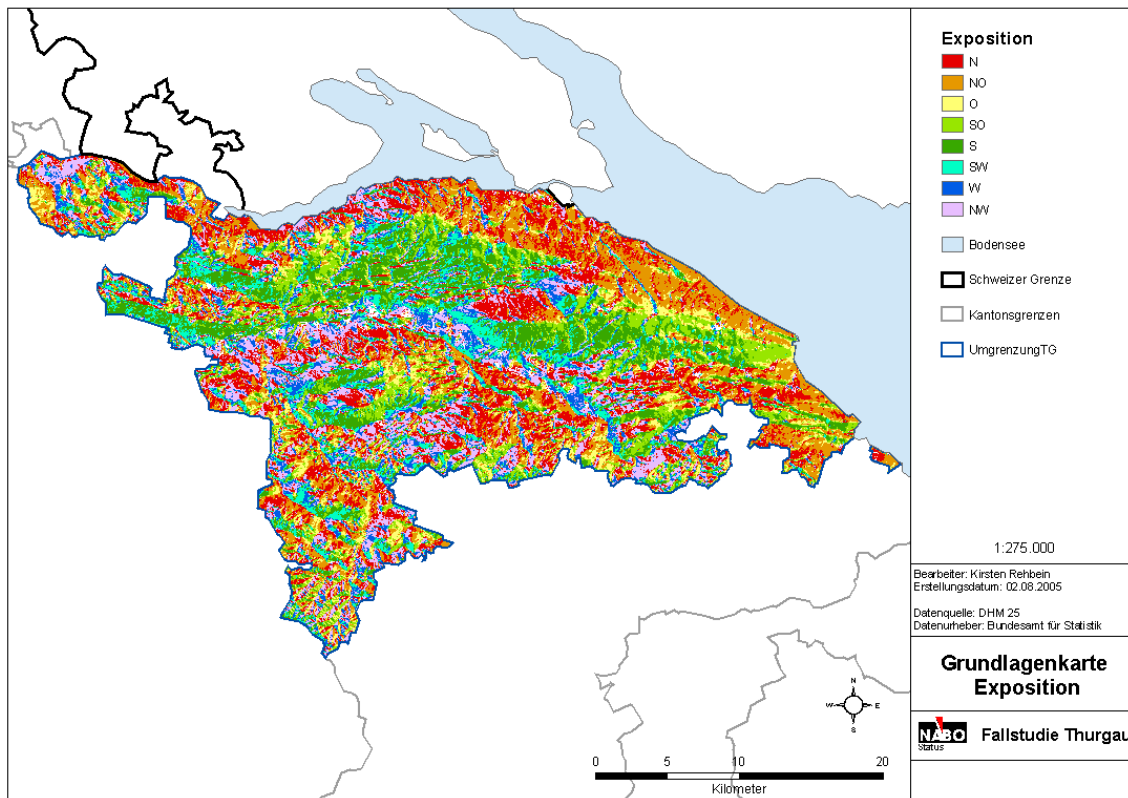
Anhang B- 1: Karte des Höhenmodells



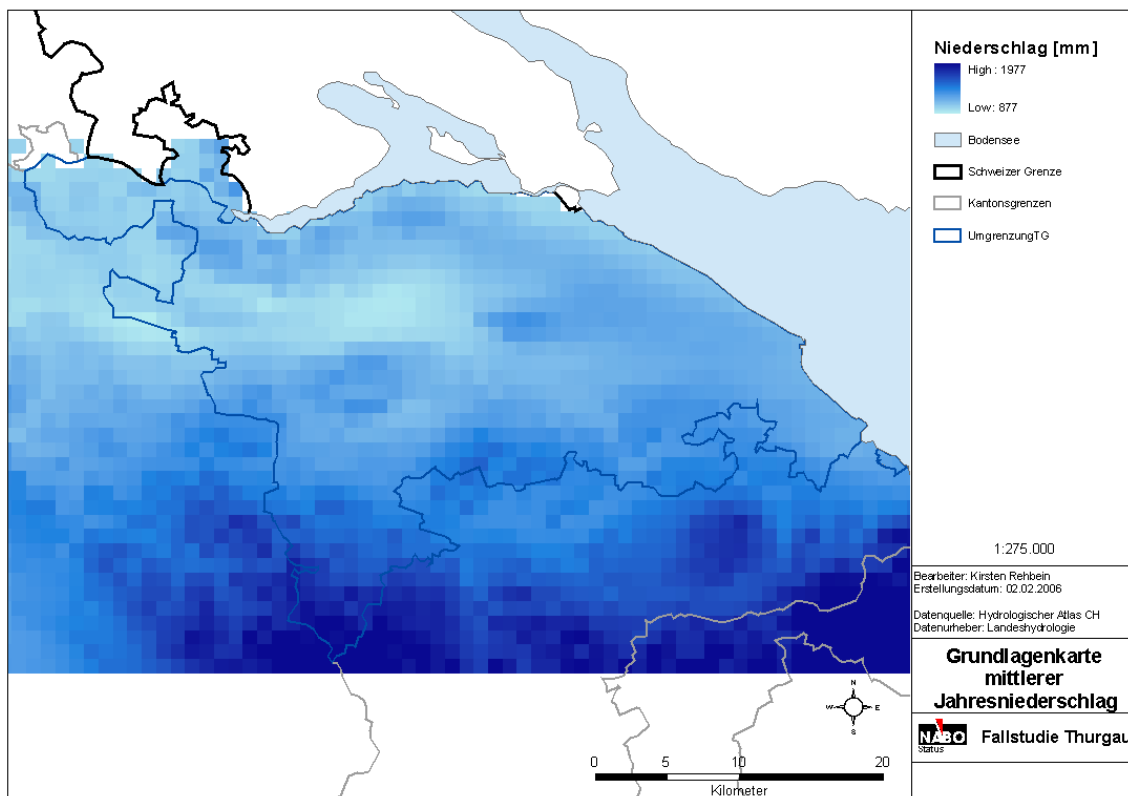
Anhang B- 2: Karte der Hangneigung



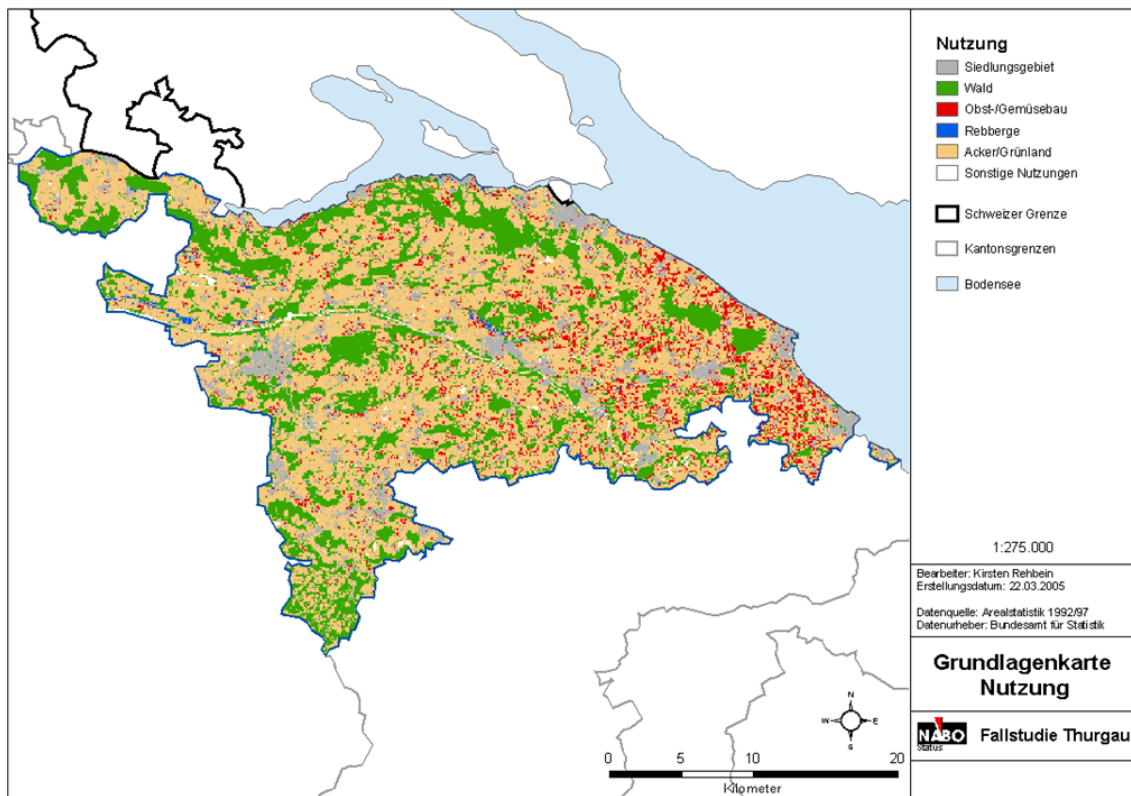
Anhang B- 3: Karte der Exposition



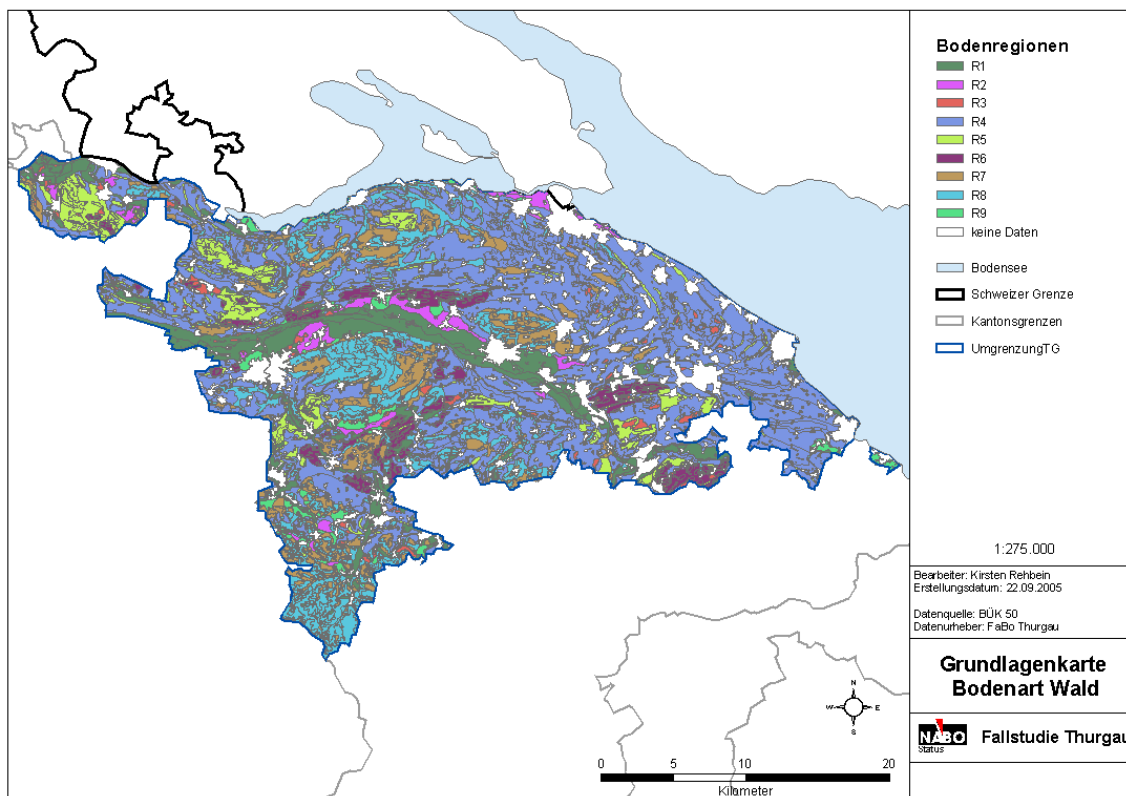
Anhang B- 4: Karte der Jahresniederschläge



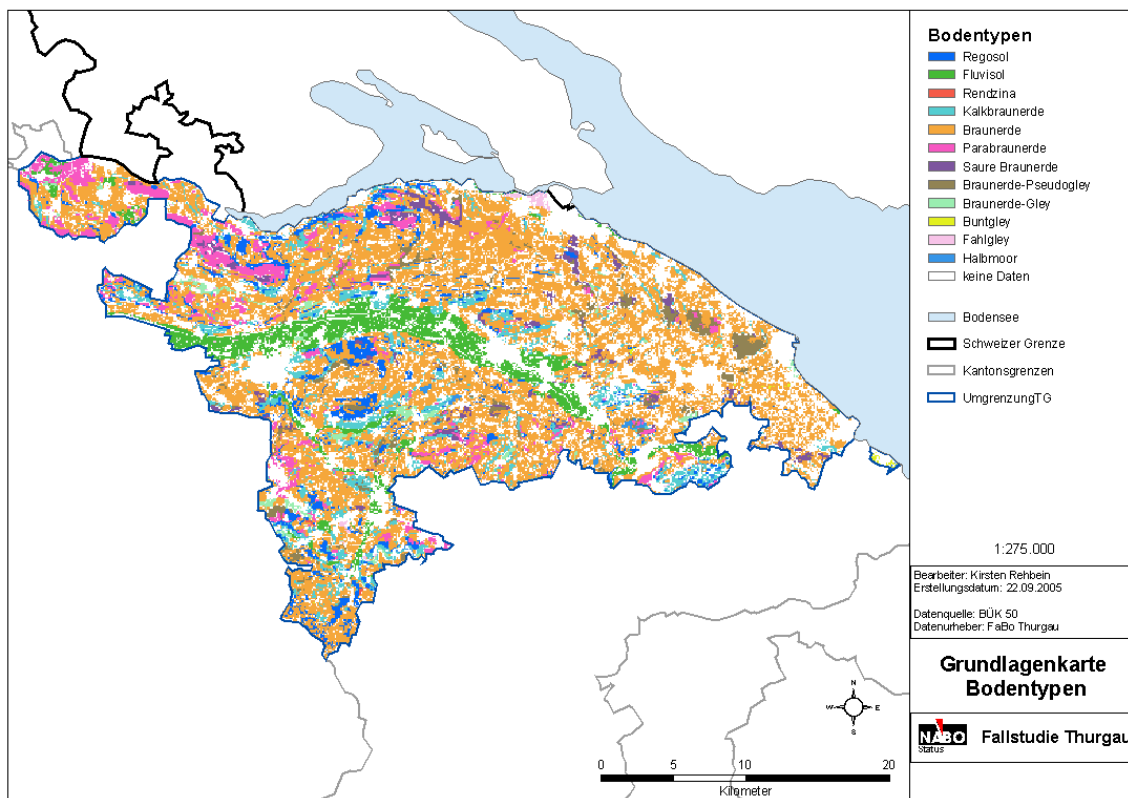
Anhang B- 5: Karte der Landnutzung



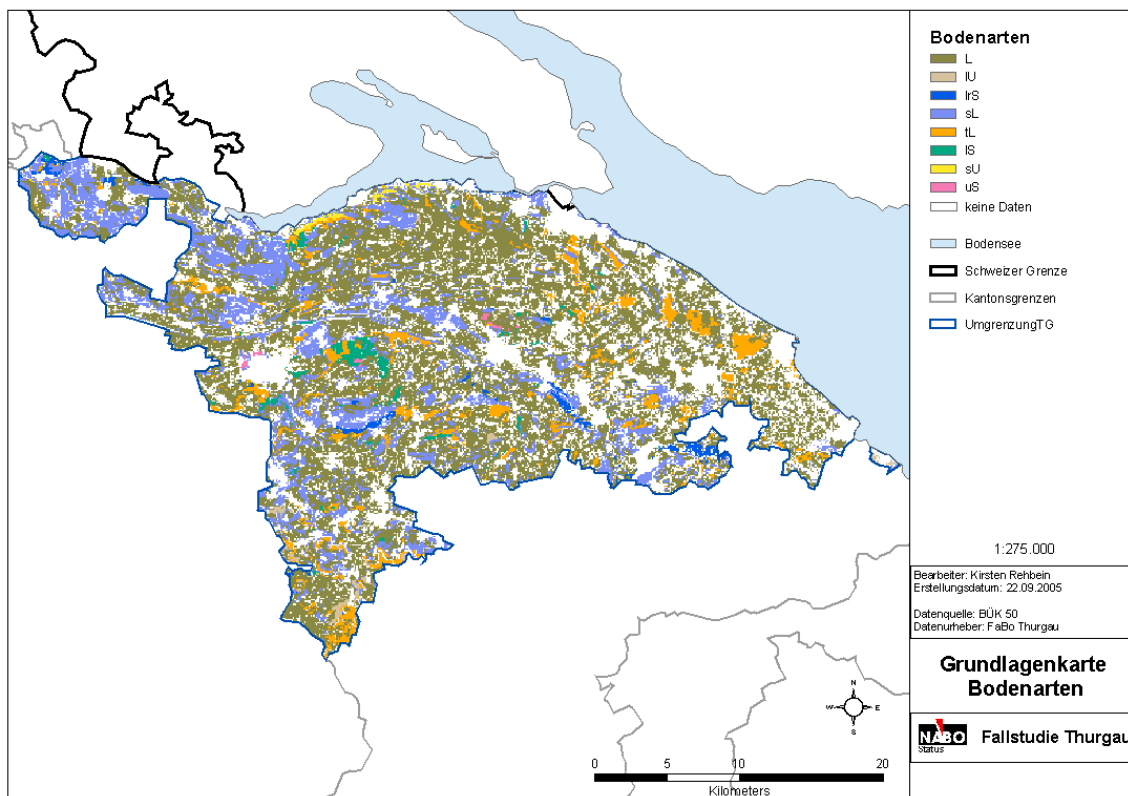
Anhang B- 6: Karte der Bodenregionen



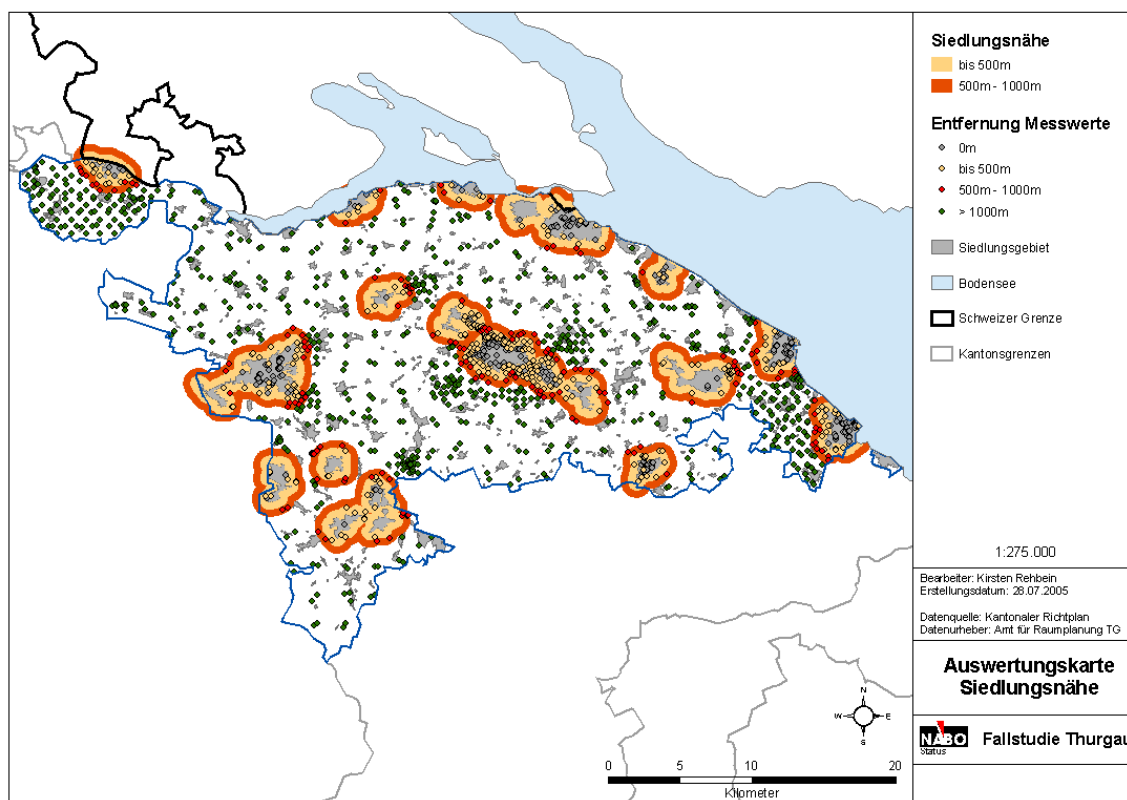
Anhang B- 7: Karte der Bodentypen



Anhang B- 8: Karte der Bodenarten

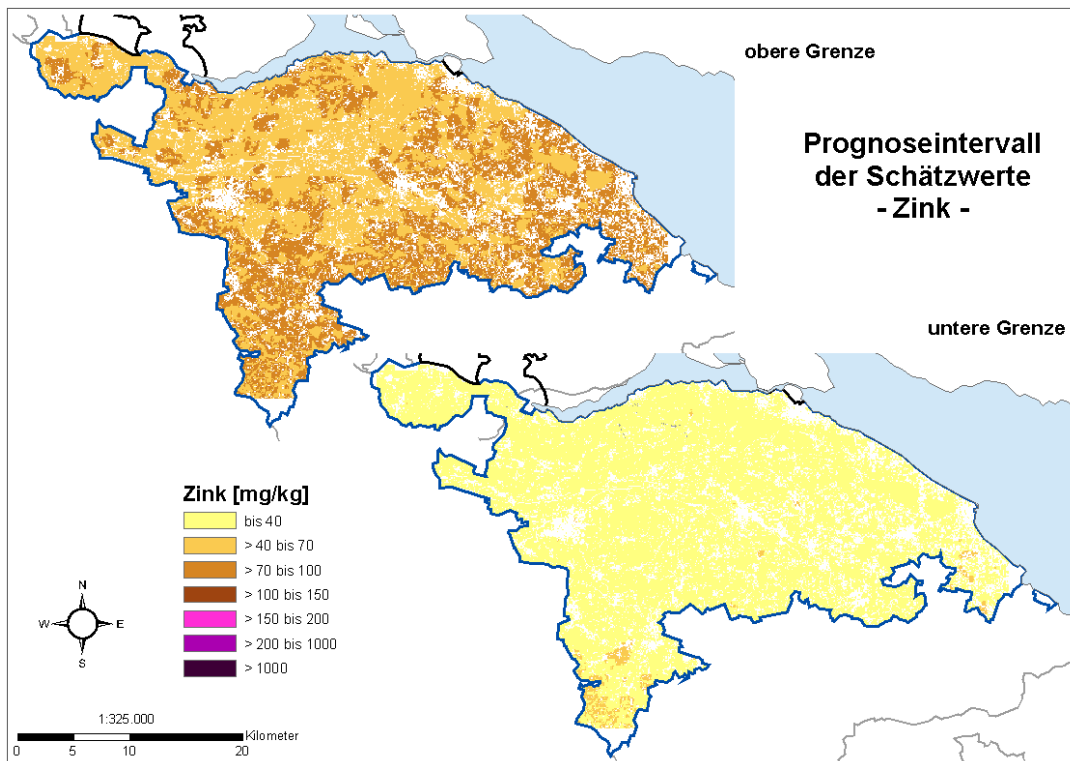


Anhang B- 9: Karte der Siedlungsnähe

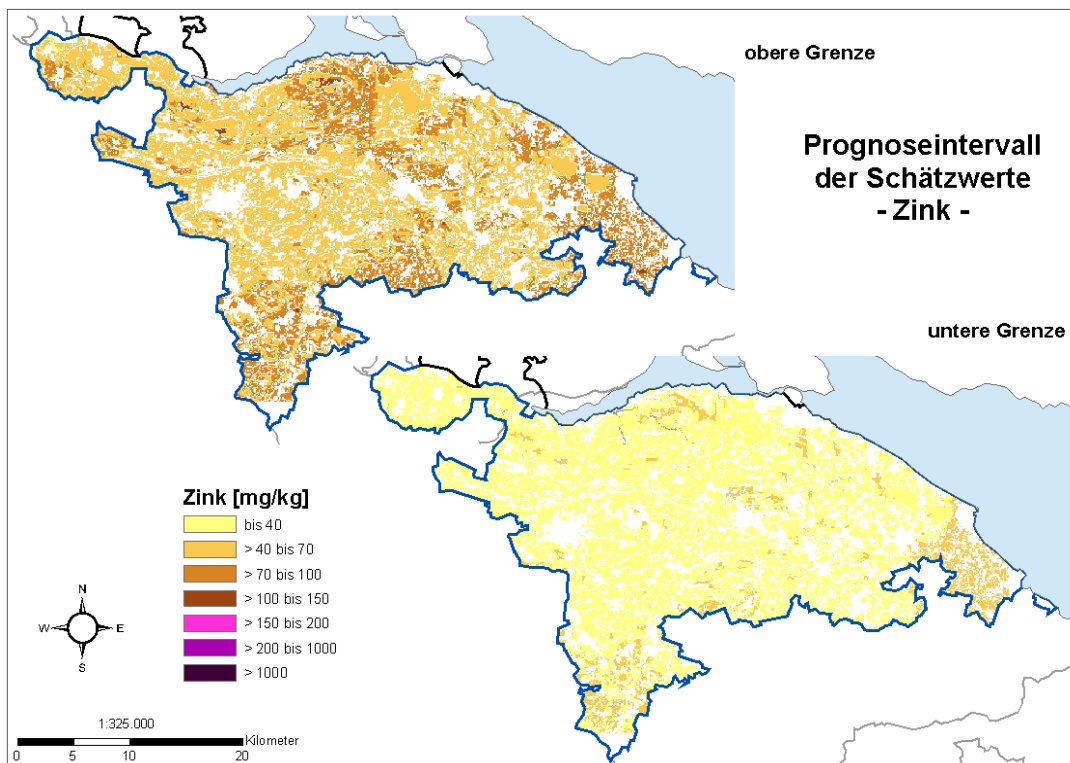


Anhang C: Prognoseintervalle der Schätzkarten für Zink

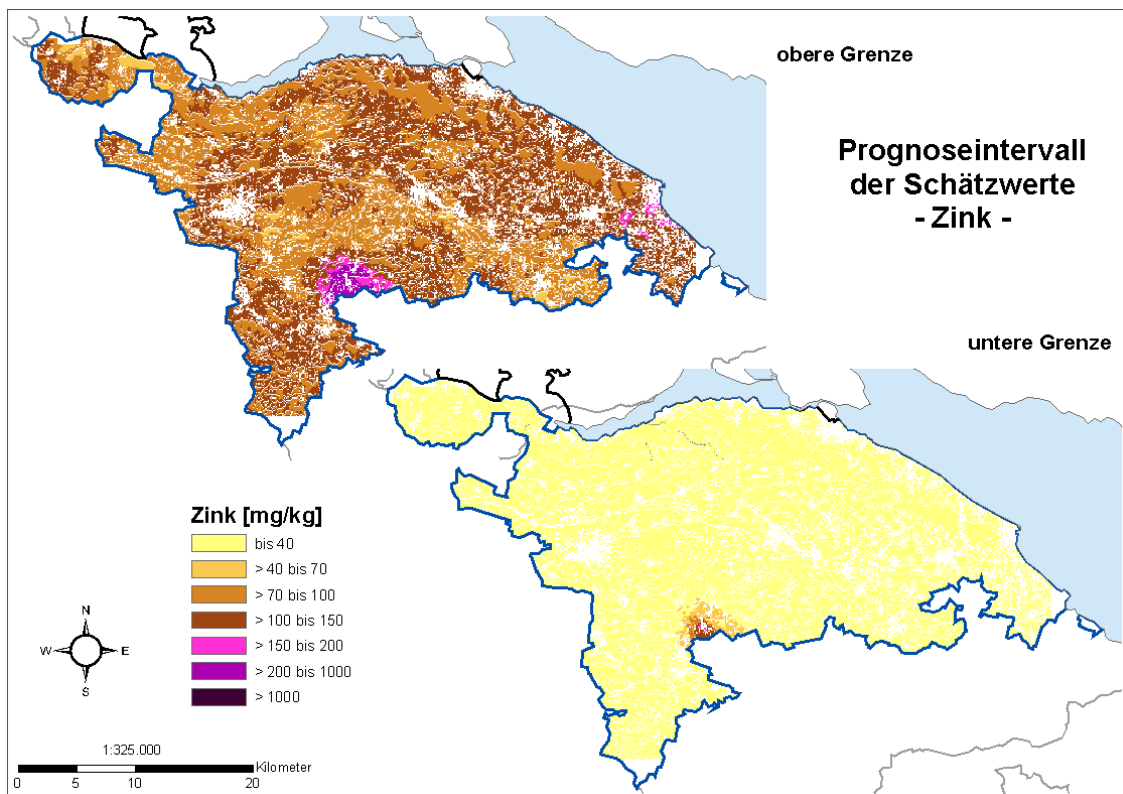
Anhang C- 1: Prognoseintervall des zweiseitigen 95 %-Quantils der interpolierten Zinkgehalte der Simple-Kriging-Vorhersagefehler, gewichtete Residuen, Zn673 Modell 1



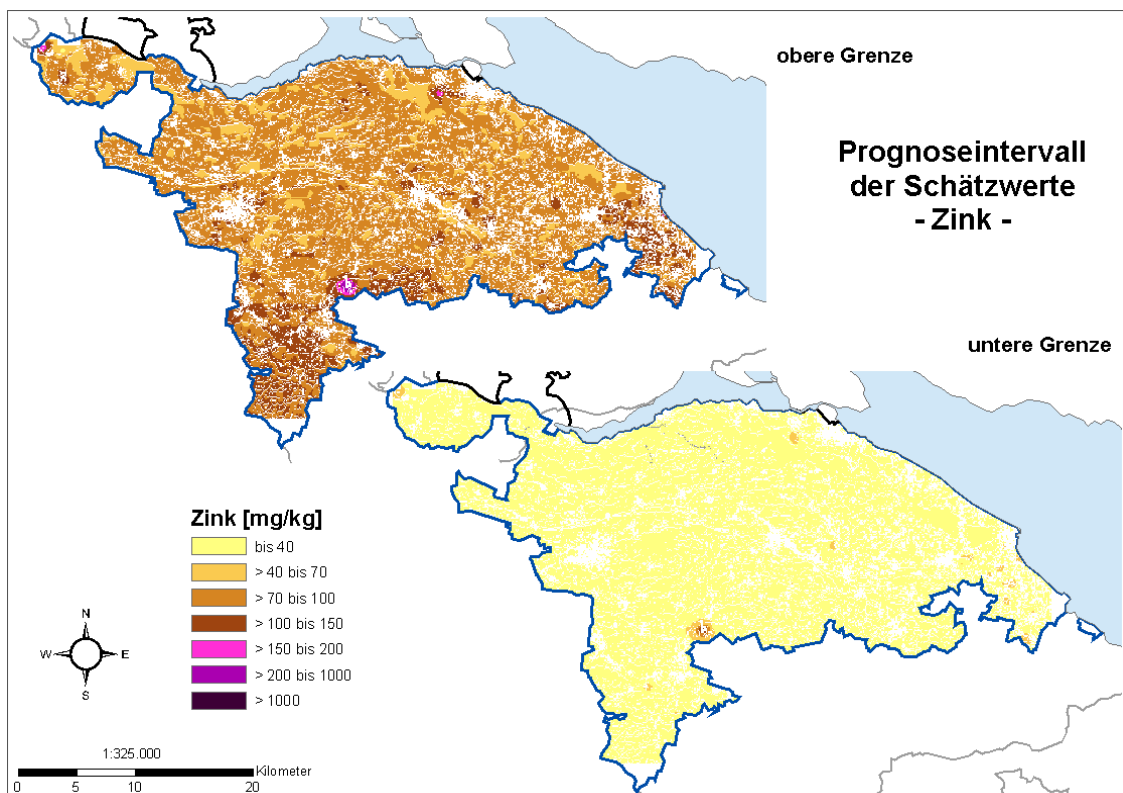
Anhang C- 2: Prognoseintervall des zweiseitigen 95 %-Quantils der interpolierten Zinkgehalte der Ordinary-Kriging-Vorhersagefehler, gewichtete Residuen, Zn673 Modell 2



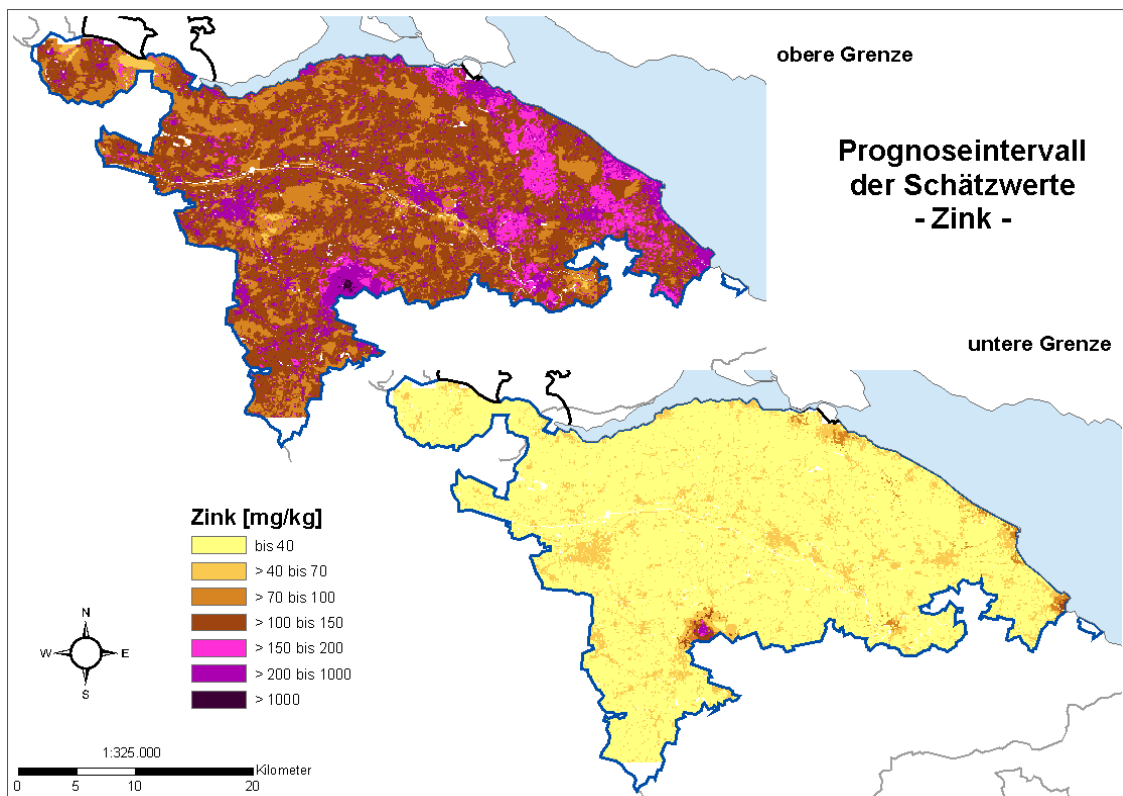
Anhang C- 3: Prognoseintervall des zweiseitigen 95 %-Quantils der interpolierten Zinkgehalte der Ordinary-Kriging-Vorhersagefehler, alle Residuen, Zn793 Modell 1



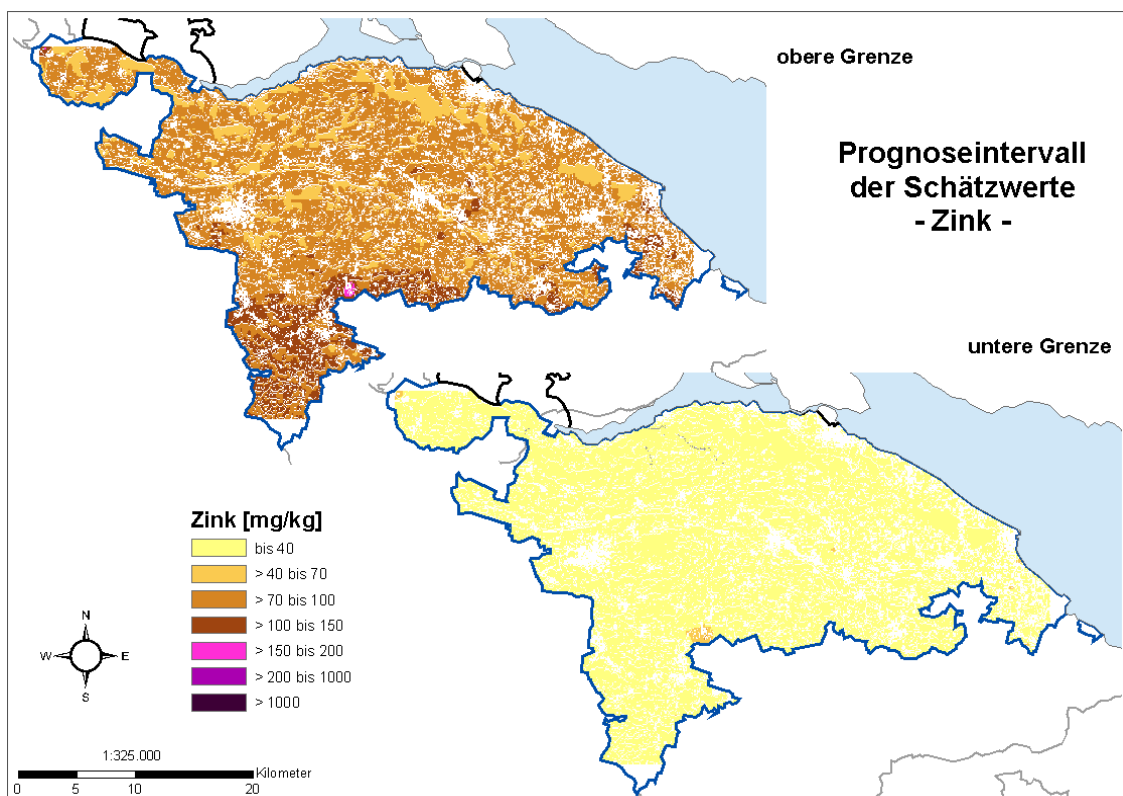
Anhang C- 4: Prognoseintervall des zweiseitigen 95 %-Quantils der interpolierten Zinkgehalte der Simple-Kriging-Vorhersagefehler, Residuen ohne 4 auffällige Werte, Zn793 Modell 1



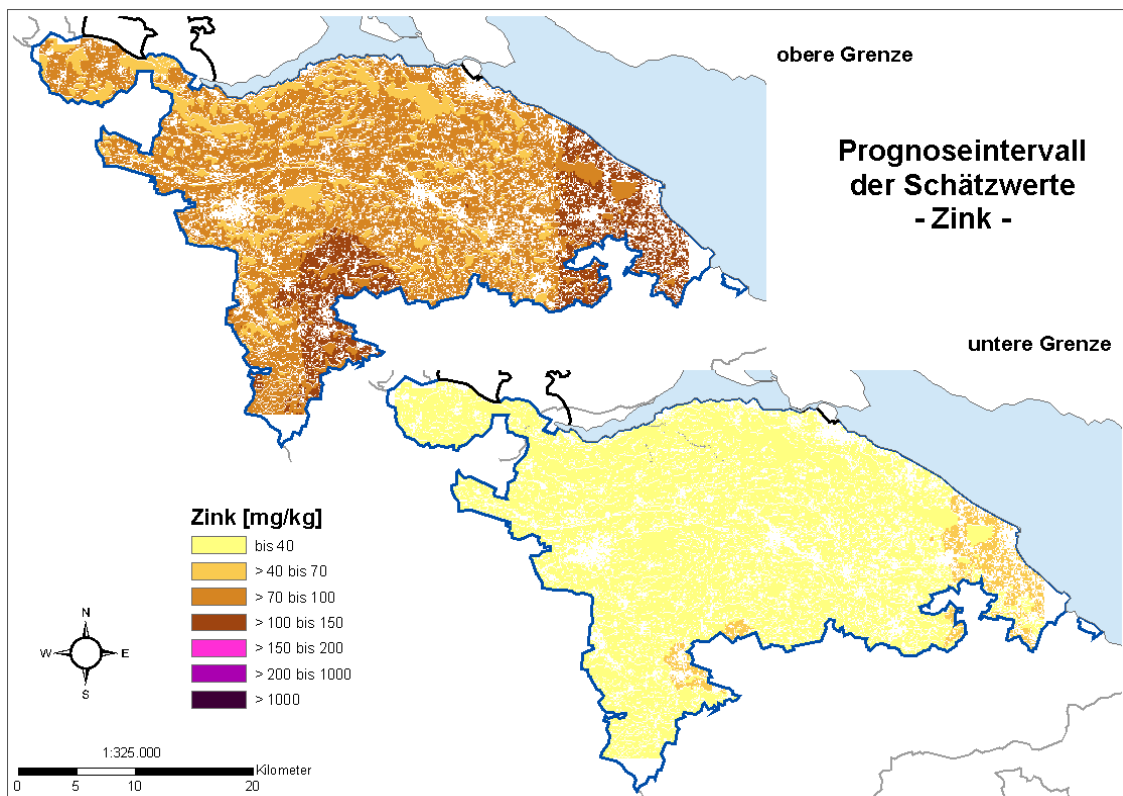
Anhang C- 5: Prognoseintervall des zweiseitigen 95 %-Quantils der interpolierten Zinkgehalte der Ordinary-Kriging-Vorhersagefehler, Zn1290 Modell 1



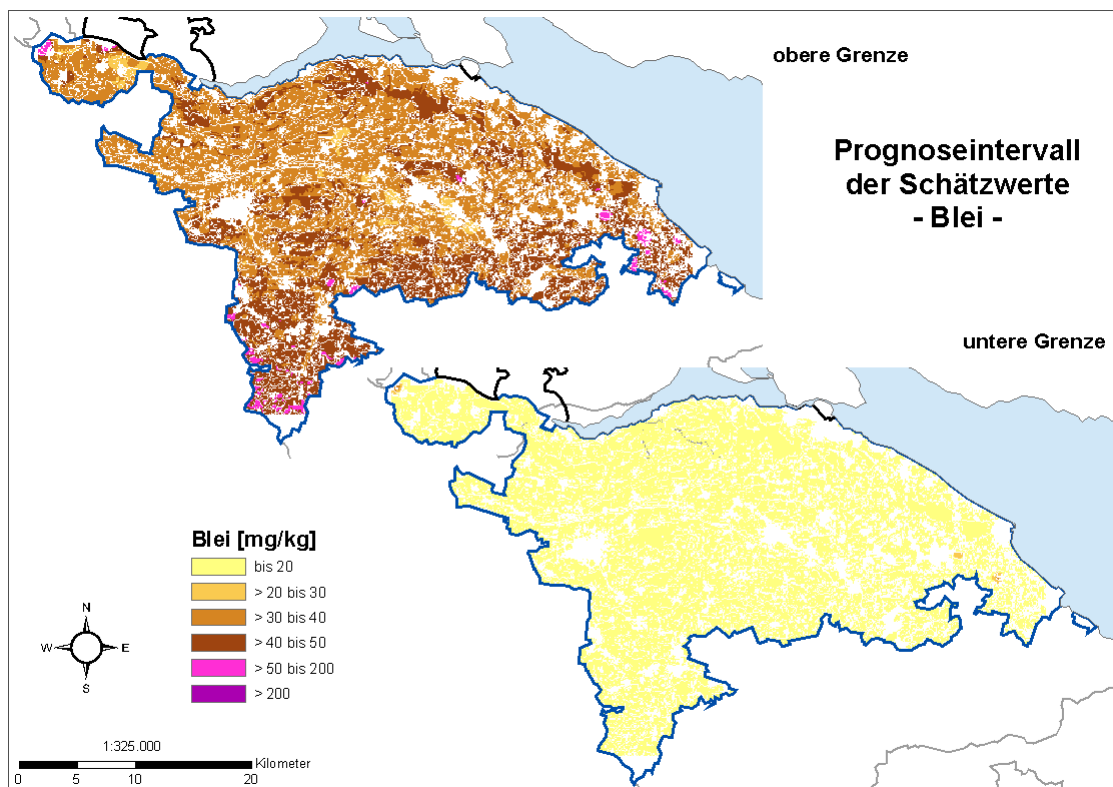
Anhang C- 6: Prognoseintervall des zweiseitigen 95 %-Quantils der interpolierten Zinkgehalte der Simple-Kriging-Vorhersagefehler, Zn386 Modell 1



Anhang C- 7: Prognoseintervall des zweiseitigen 95 %-Quantils der interpolierten Zinkgehalte der Ordinary-Kriging-Vorhersagefehler, Zn193 Modell 1



Anhang C- 8: Prognoseintervall des zweiseitigen 95 %-Quantils der interpolierten Bleigehalte der Simple-Kriging-Vorhersagefehler, Pb793 Modell 3



Anhang D: Nicht berücksichtigte Standorte Datensatz 1290

Nr	Projekt	Thema	Begründung
6	P95STR	SM entlang von Strassen	War für die Fragestellung nicht sinnvoll
7	P00SA	Schiessanlagen	Kein Zink gemessen
11	P98ARBON	Verdichtung P95RAS	Hohe Werte vermutlich auf Giessereisande zurückzuführen, lokale Belastungen, nicht flächenrepräsentativ
13	P99BST	Überprüfung hoher Werte im Kanton	Lokale Belastungen, nicht flächenrepräsentativ
18	P99ARA	SM auf Klärschlammflächen	Lokale Belastungen, nicht flächenrepräsentativ
25	P93VZB-G	SM in der Umgebung einer Verzinkerei	Spezialuntersuchung in Gemüse
26	P01ALTLASTEN	Schiessanlage	Kein Zink gemessen
31	P00WAN	Spielplatzuntersuchung	Kein Zink gemessen
34	P97VZB-P	SM in der Umgebung einer Verzinkerei	Einzelne Profilmessung