



Master Thesis

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

zum Thema

„Bestimmung der räumlichen Verteilung der Feinstaubbelastung im Bereich groß- flächiger Tagebauareale“

GIS-gestützte Modellierung und Darstellung am Beispiel
des Rheinischen Braunkohlereviers

vorgelegt von

Heinz-Jürgen Jansen

10349, UNIGIS MSc Jahrgang 2016

Zur Erlangung des Grades

„Master of Science (Geographical Information Science & Systems) – MSc(GIS)“

Korschenbroich, 07.09.2021

Danksagung

Mein besonderer Dank gilt dem UNIGIS Lehrgangsteam, welches mir unter Berücksichtigung meiner privaten und beruflichen Situation durch die Gewährung einer weiteren Verlängerung überhaupt erst ermöglicht hat, das Studium abschließen zu können.

Auch möchte ich mich herzlich bei meinem Betreuer Herrn Dr. Robert Marschallinger bedanken, der in der entscheidenden Phase meiner Master Thesis die Betreuung übernommen hat und durch seine fachliche Unterstützung eine sehr wertvolle Hilfe und zugleich ein wesentlicher Motivator für mich war.

Ohne die Zurverfügungstellung geeigneter, validierter Feinstaub-Messwerte durch das Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen wäre das Thema nicht in der gewünschten Form realisierbar gewesen. Hier gilt mein besonderer Dank Herrn Jürgen Friesel und Herrn Michael Niedenzu aus dem Fachbereich 42 „Kontinuierliches Luftqualitätsmessnetz“ des Landesamtes, die mich mit Rat und Tat unterstützt und die gewünschten Daten stets umgehend bereitgestellt haben.

Außerdem möchte ich mich herzlich bei Frau Mag. Maria Embacher bedanken, die mich in der letzten Phase meiner Master Thesis motivierend und unterstützend begleitet hat.

Erklärung der eigenständigen Abfassung der Arbeit

Ich versichere, diese Master Thesis ohne fremde Hilfe und ohne Verwendung anderer als der angeführten Quellen angefertigt zu haben, und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat. Alle Ausführungen der Arbeit, die wörtlich oder sinngemäß übernommen wurden, sind entsprechend gekennzeichnet.

Korschenbroich, den 07.09.2021

Heinz-Jürgen Jansen

Zusammenfassung

Hintergrund: Das Thema Feinstaub ist in den letzten Jahren verstärkt in den Fokus der Öffentlichkeit gerückt. Die von Feinstaub ausgehende Gesundheitsgefahr ist seit langem bekannt und konnte in zahlreichen Studien umfangreich belegt werden. Lange Zeit nicht berücksichtigt wurde hingegen, dass neben den bekannten Emittenten auch der offene Abbau von Bodenschätzen in Tagebauen zu einer erheblichen Feinstaubemission führen kann. Die Kenntnis um die raum-zeitliche Verteilung der Belastung ist für die Tagebauanrainer von großer Bedeutung. Ziel ist es, eine Interpolationsmethode unter Einbeziehung der Landnutzung zu selektieren und zu modifizieren, die mit frei zugänglichen Daten gute Schätzergebnisse liefert. Als Untersuchungsgebiet wurde das Rheinische Braunkohlerevier mit einem dicht besiedelten Umfeld, in dem bereits eine hohe Feinstaubbelastung vorhanden ist, gewählt.

Methodik: Vergleichen wurde in einem 2D+t-Modell die Validität eines Simple Krigings und eines Simple Cokrigings unter Einbeziehung der Landnutzung auf Basis der Corine Land Cover Daten. Als sekundäre Variable floss der vom belgischen VITO ursprünglich für die Trendbereinigung entwickelte numerische Landnutzungsindikator, der für jeden beliebigen Punkt berechnet werden kann, ein. Die Daten wurden durch Eliminierung von vorhandenen Trends und mehrfacher Glättung für die Interpolationen aufbereitet. Die Validierung der Ergebnisse erfolgte mittels QQ-Plots und RMSE. Eine Kreuzvalidierung stand nicht zur Verfügung.

Ergebnisse: Simple Kriging lieferte auf Basis der wenigen Messdaten zufriedenstellende Ergebnisse (RMSE-Wert: $4,55 \mu\text{g}/\text{m}^3$). Das Kokriging unter Einbeziehung des Landnutzungsindikators mit und ohne Extrema brachte nicht die erwartete Verbesserung des Ergebnisses (RMSE-Werte: $10,40/10,49 \mu\text{g}/\text{m}^3$). Als Ursache für die nicht zufriedenstellenden Ergebnisse konnten die Kreuzvariogramm-Modelle identifiziert werden. Durch große Semivarianzen im Nahbereich, die dann wieder abfallen, und einer in bestimmten Distanzbereichen negativen Korrelation zwischen den Kokriging-Variablen, ist die Konstruktion eines auch für kleine Distanzen passenden Variogramm-Modells nicht möglich. Dies führt beim Kokriging zu Extrema im Bereich der Messpunkte. Abseits der Messpunkte sind die Ergebnisse indessen gut.

Schlussfolgerung: Es konnte gezeigt werden, dass Simple Kriging nach einer entsprechenden Aufbereitung der Daten ein geeigneter Ansatz für die Fragestellung ist. Die Verwendung des Landnutzungsindikators beim Kokriging ist ohne Modifikation jedoch nicht sinnvoll. Eine Weiterentwicklung des Kokriging-Ansatzes durch den Einbezug anderer sekundärer Kokriging-Variablen zur Verbesserung der Ergebnisse ist als nächster Schritt vorstellbar.

Abstract

Background: In recent years, the topic “particulate matter” (PM) has come into the focus of public attention. Health risks of PM have been known for a long time and are confirmed by several studies. Most PM emitter are well known but it is almost unknown, that surface mining may release high quantities of PM too. Therefore, the knowledge of the distribution of fine dust is very important for residents close to open pit mines. The aim of the study was to select and modify an interpolation method, that takes land use into account to obtain satisfactory prediction results based on data for free use. As study area, the Renish brown coal area in Western Germany was chosen. The neighbouring areas of the open pit mines are partially densely populated; they already have a very high level of PM pollution.

Methods: Using a 2D+t model, the validity of Simple kriging and Simple Cokriging was compared, taking into account the land use based on the Corine Land Cover data. As secondary cokriging variable a numeric land use indicator developed by the Belgian VITO institute was used. It is computable for any point in the area. Before kriging, existing trends were eliminated and smoothing of data (transitions) was performed. QQ-Plots and RMSE values were used for data validation. Since not implemented in the geostatic tool, a cross-validation was not done.

Results: Considering the small number of sample points Simple kriging delivered satisfactory results (RMSE-value: $4,55 \mu\text{g}/\text{m}^3$). In contrast cokriging taking the land use indicator (with and without extrema) into account did not produce the expected better results (RMSE-values: $10,40/10,49 \mu\text{g}/\text{m}^3$). As a cause of the unsatisfactory results, the constructed cross-variogram models were identified. Due to the huge semivariances at short distances, which fall off and then increase by increasing distances, and a negative correlation between the cokriging variables in certain distance ranges, a good fitting cross-variogram model – especially for short distances – could not be constructed. This produced very high prediction values at the sample points. Away from the sample points the prediction values are good as the QQ-Plots have shown.

Conclusion: It could be demonstrated that Simple kriging after measurement data preparation is a suitable approach for the research question with regard to further studies. The land use indicator as a secondary cokriging variable cannot be recommended – its use would require prior modifications. A further development of the cokriging model which would include other variables like meteorological parameters as secondary cokriging variable is conceivable.

Inhaltsverzeichnis

Danksagung	II
Erklärung der eigenständigen Abfassung der Arbeit	III
Zusammenfassung	IV
Abstract	V
Inhaltsverzeichnis	VI
Abbildungsverzeichnis	VIII
Tabellenverzeichnis	XI
Abkürzungsverzeichnis	XII
1 Einleitung	1
1.1 Motivation	1
1.2 Problemstellung und Relevanz	4
1.3 Untersuchungsgebiet	7
1.4 Forschungsziele und Einordnung der Arbeit	12
1.5 Nicht-Ziele.....	14
1.6 Zielgruppe.....	14
2 Methodik	15
2.1 Luftqualitätsmessungen in Nordrhein-Westfalen.....	15
2.2 Deskriptive Statistik	18
2.3 Explorative a-räumliche und räumliche Datenanalyse	19
2.4 Regressions- und Korrelationsanalyse.....	24
2.5 Interpolationsverfahren.....	27
2.6 RIO-Modell	38
2.7 Anwendungsfall.....	46

3	Ergebnisse	57
3.1	Deskriptive Statistik	57
3.2	Explorative Räumliche Datenanalyse (ESDA).....	58
3.3	Regressions- und Korrelationsanalyse.....	69
3.4	Variographie und Interpolationen.....	72
4	Diskussion	82
4.1	Eignung der gewählten Verfahren und Tools.....	82
4.2	Aussagekraft des Modells zur Feinstaubbelastung.....	85
4.3	Bestätigung der Hypothese	86
5	Schlussfolgerungen und Ausblick	87
	Literaturverzeichnis	89
Anhang A	Weitere Messdaten	92
Anhang B	Landnutzungsindikatoren	94
Anhang C	CORINE Land Cover Daten	95
Anhang D	Erweitertes RIO-Modell	96
Anhang E	Visual Basic- und Python-Skripte	100
Anhang F	Kriging- und KoKriging-Parameter	112

Abbildungsverzeichnis

Abb. 1	Tagebau Garzweiler II, Kohlekraftwerke BOA und Neurath im Januar 2018	1
Abb. 2	Braunkohleförderung und Verwendung in Deutschland 2018	8
Abb. 3	Revierkarte Rheinisches Braunkohlerevier	8
Abb. 4	Tagebau Garzweiler II im Januar 2018	9
Abb. 5	Untersuchungsgebiet mit Messstationen LANUV und Landnutzung.....	10
Abb. 6	Untersuchungsgebiet (Physische Karte), Westermann, 2019	11
Abb. 7	Messstationen LANUV am Tagebaurand	12
Abb. 8	Histogramm PM ₁₀ -Messwerte 01.09.2018.....	21
Abb. 9	Histogramm gesamte PM ₁₀ -Messwerte	22
Abb. 10	Trendanalyse PM ₁₀ -Messwerte 08.09.2018	24
Abb. 11	Regressionsgerade PM ₁₀ -Mittelwerte - Landnutzungsindikator	25
Abb. 12	Beispiel-Semivariogramm erläutert (β , gauß-transformiert).....	31
Abb. 13	Trendfunktionen PM ₁₀ (Wochentage/Wochenende), (Janssen et al., 2008)...	42
Abb. 14	Attributtabelle Puffer_RCL_RODE_DIS (β -Berechnung)	49
Abb. 15	Pufferzonen Landnutzungsindikator	51
Abb. 16	De-/Retrending-Interpolationsraster.....	53
Abb. 17	Liniendiagramm Verlauf PM ₁₀ -Tagesmittelwerte	58
Abb. 18	Histogramme PM ₁₀ -Messwerte 01.09.2018 - 08.09.2018.....	59
Abb. 19	Histogramm PM ₁₀ -Messwerte.....	60
Abb. 20	Hist. PM ₁₀ -Messwerte nach erster GT	60
Abb. 21	Hist. PM ₁₀ -Messwerte nach Detrending	61
Abb. 22	Hist. PM ₁₀ -Messwerte nach zweiter GT	61
Abb. 23	Hist. PM ₁₀ -Messwerte nach SK/RT/GRT	61
Abb. 24	Hist. PM ₁₀ -Messwerte nach SCK/RT/GRT	61
Abb. 25	Histogramm PM ₁₀ -Messwerte nach SCK(β -Range)/RT/GRT	62
Abb. 26	Histogramm β (alle Punkt)	62
Abb. 27	Histogramm β nach GT (alle Punkte)	62
Abb. 28	Histogramm β nach GT ohne Ausreißer (β -Range)	63
Abb. 29	Normal QQ-Plots PM ₁₀ -Messwerte 01.09.2018	63
Abb. 30	Normal QQ-Plots PM ₁₀ -Messwerte 02.09.2018	63

Abb. 31	Normal QQ-Plots PM ₁₀ -Messwerte 03.09.2018	64
Abb. 32	Normal QQ-Plots PM ₁₀ -Messwerte 04.09.2018	64
Abb. 33	Normal QQ-Plots PM ₁₀ -Messwerte 05.09.2018	64
Abb. 34	Normal QQ-Plot PM ₁₀ -Messwerte 06.09.2018	64
Abb. 35	Normal QQ-Plot PM ₁₀ -Messwerte 07.09.2018	65
Abb. 36	Normal QQ-Plot PM ₁₀ -Messwerte 08.09.2018	65
Abb. 37	Raum-Zeit-Diagramm PM ₁₀ -Tagesmittelwerte (01.09. - 08.09.2018).....	66
Abb. 38	Trendanalyse PM ₁₀ -Messwerte 01.09.2018	66
Abb. 39	Trendanalyse PM ₁₀ -Messwerte 02.09.2018	66
Abb. 40	Trendanalyse PM ₁₀ -Messwerte 03.09.2018	67
Abb. 41	Trendanalyse PM ₁₀ -Messwerte 04.09.2018	67
Abb. 42	Trendanalyse PM ₁₀ -Messwerte 05.09.2018	67
Abb. 43	Trendanalyse PM ₁₀ -Messwerte 06.09.2018	67
Abb. 44	Trendanalyse PM ₁₀ -Messwerte 07.09.2018	67
Abb. 45	Trendanalyse PM ₁₀ -Messwerte 08.09.2018	67
Abb. 46	Trendanalyse PM ₁₀ - IDW Power 0	68
Abb. 47	Trendanalyse β Messstationen	68
Abb. 48	Trendanalyse β zusätzliche Punkte	68
Abb. 49	Trendanalyse β Punkte gesamt	68
Abb. 50	Regressionsgerade PM ₁₀ -Mittelwerte - Landnutzungsindikator β	69
Abb. 51	Histogramm β Messstationen	70
Abb. 52	Histogramm PM ₁₀ -Mittelwerte	70
Abb. 53	Regressionsanalyse PM ₁₀ -Mittelwerte - β ohne Extrema	70
Abb. 54	Regressionsanalyse PM ₁₀ (GT, DT, GT) - β (GT)	71
Abb. 55	Variogramm PM ₁₀ (GT, DT, GT)	72
Abb. 56	Variogramm β (GT)	72
Abb. 57	Kreuzvariogramm PM ₁₀ (GT, DT, GT) / β (GT)	73
Abb. 58	Experimentelle Kreuzvariogramme PM ₁₀ (GT, DT, GT) / β (GT).....	73
Abb. 59	Variogramm β -Range (GT).....	74
Abb. 60	Kreuzvariogramm PM ₁₀ (GT, DT, GT) / β -Range (GT).....	74
Abb. 61	Anisotropie der Gauß-transformierten β -Werte	75
Abb. 62	Interpolationsraster Simple Kriging PM ₁₀ (GRT/RT/GRT).....	75
Abb. 63	Interpolationsraster Simple CoKriging PM ₁₀ / β (GRT/RT/GRT).....	76

Abb. 64 Interpolationsraster Simple CoKriging PM_{10}/β -Range (GRT/RT/GRT).....	76
Abb. 65 QQ-Plot SK PM_{10} - PM_{10} -Messwerte.....	77
Abb. 66 QQ-Plot SCK PM_{10}/β - PM_{10} -Messwerte.....	77
Abb. 67 QQ-Plot SCK PM_{10}/β -Range - PM_{10} -Messwerte	77
Abb. 68 β -Landnutzungsindikorkarte	97
Abb. 69 Trendbereinigung Mittelwert am Beispiel von NO_2 , (Janssen et al., 2008)....	98
Abb. 70 Karte der PM_{10} -Jahresmittelwerte 2006 (RIO/OK), (Janssen et al., 2008)	99
Abb. 71 Simple Kriging PM_{10} Parameter	112
Abb. 72 Simple CoKriging PM_{10}/β Parameter.....	112
Abb. 73 Simple CoKriging PM_{10}/β -Range Parameter	113

Tabellenverzeichnis

Tab. 1	Flächen, Kohleinhalt und Fördermengen der Tagebaue	9
Tab. 2	Messstationen des LANUV NRW im Untersuchungsgebiet	17
Tab. 3	Messwerte Feinstaub PM10 (24H Mittelwert)	18
Tab. 4	Klasseneinteilung Histogramme	21
Tab. 5	Univariate und multivariate Kriging-Verfahren (Auswahl)	33
Tab. 6	Definition RIO-Klassen (RCL), Janssen et al. (2008)	41
Tab. 7	Verwendung RIO-Modell (Pro/Contra)	44
Tab. 8	Eingesetzte GIS-Software und Verwendung	47
Tab. 9	Zuordnung z-Koordinate - Untersuchungstag	48
Tab. 10	Verarbeitungsschritte Kriging/Kokriging	55
Tab. 11	Deskriptive Statistik: Maßzahlen der Stichprobe	57
Tab. 12	RMSE Interpolationen	77
Tab. 13	Messwerte Windgeschwindigkeit (24H Mittelwert)	92
Tab. 14	Messwerte Windrichtung vektoriell (24H Mittelwert)	92
Tab. 15	Messwerte Niederschlag (24H-Summe)	93
Tab. 16	Landnutzungsindikatoren Messstationen	94
Tab. 17	Landnutzungsindikatoren Zusätzliche Punkte	94
Tab. 18	Entwicklung der CORINE Land Cover Daten, Copernicus, Stand 02/2020	95
Tab. 19	Corine Land Cover (CLC) Klassen (Level 1- 3)	95

Abkürzungsverzeichnis

Abb.	Abbildung
ASE	Average Standard Error
BLUP	Best Linear Unbiased Predictor
BUND	Bund für Umwelt und Naturschutz Deutschland e.V.
CLC	CORINE Land Cover
CORINE	Coordination of Information on the Environment
DEBRIV	Deutscher Braunkohlen-Industrie-Verein e.V.
DT	Detrending
EDSA	Exploratory Spatial Data Analysis
EMEP	European Monitoring and Evaluation Programme
EU	Europäische Union
etc.	et cetera
GIS	Geographisches Informationssystem
GPI	Globale Polynomiale Interpolation
GSLIB	Geostatistical Software Library (Stanford University)
GRT	Gauß-Rücktransformation
GT	Gauß-Transformation
ha	hectare/Hektar
Hist.	Histogramm
IDW	Inverse Distance Weighted
IK	Indicator Kriging
Kap.	Kapitel
KT	Kriging with a Trend
LANUV	Landesamt für Natur, Umwelt und Verbraucherschutz NRW
LNI	Landnutzungsindikator β (RIO)
LUQS	Luftqualitätsüberwachungssystem

MAE	Mean Absolute Error
math.	mathematisch
MMU	Minimal Mapping Unit
OK	Ordinary Kriging
PM	Particulate Matter
QQ-Plot	Quantil-Quantil-Plot
RCL	RIO Class
RIO	Residual Interpolation Optimised
RMSE	Root Mean Square Error
RT	Retrending
RWE	Rheinisch-Westfälisches Elektrizitätswerk
SGeMS	Stanford Geostatistical Modeling Software
SCK	Simple Cokriging
SK	Simple Kriging
SR	Suchradius
sog.	sogenannt
Tab.	Tabelle
UK	Universal Kriging
VB	Visual Basic
VITO	Vlaamse Instelling voor Technologisch Onderzoek
vgl.	vergleiche
vs.	versus
z.B.	zum Beispiel

1 Einleitung

1.1 Motivation

Die Tagebaue des Rheinischen Braunkohlereviers mit ihren gewaltigen Dimensionen und die weithin sichtbaren Braunkohlekraftwerke prägen das Bild der Landschaft im südlichen Niederrheinischen Tiefland und der südlich angrenzenden Niederrheinischen Bucht im äußersten Westen Deutschlands.



Abb. 1 Tagebau Garzweiler II, Kohlekraftwerke BOA und Neurath im Januar 2018 ¹

Das Rheinische Revier ist mit einem Kohlevorrat von ca. 55 Mrd. t die größte Braunkohlelagerstätte Europas. Der Abbau von ca. 30 Mrd. t gilt als technisch realisierbar und wirtschaftlich sinnvoll. Die aktuell genehmigten Tagebaue des Reviers erschließen ca. 3 Mrd. t abbaubare Braunkohle.

Die wirtschaftliche Nutzung der Braunkohle im Rheinischen Revier begann schon im 18. Jahrhundert in kleineren Gruben. Mit der in Europa einsetzenden Industrialisierung im 19. Jahrhundert folgte Ende des Jahrhunderts auch der industrielle Abbau der Braunkohle. In den im Vergleich zu heutigen Maßstäben noch kleinen Tagebauen kamen bereits

¹ Quelle: Eigene Fotoaufnahme, Datum der Aufnahme: 27.01.2018

Großgeräte zum Einsatz. Großtagebaue mit einer Betriebsfläche von mehreren tausend Hektar, wie sie heute vorzufinden sind, entstanden erst ab den 1950er-Jahren als Folge einer Wanderung der Abbaugebiete in nordwestlicher und westlicher Richtung in ländlichere Gebiete mit einer geringeren Bevölkerungsdichte. Es kam dabei zu einer Verlagerung der Abbaufelder in zunehmende Tiefen.

Die wachsenden Dimensionen der Tagebaue vergrößerten in der zweiten Hälfte des letzten Jahrhunderts massiv den Einfluss auf die Landschaft, den Wasserhalt und die Luftqualität im Revier. Die Anrainer des Braunkohlereviere waren dabei verschiedenen Belastungen durch den Tagebaubetrieb ausgesetzt.

In den 1970er- und 1980er-Jahre lag der Fokus der Öffentlichkeit primär auf dem Thema Umsiedlung der vom Tagebau betroffenen Orte und dem damit verbundenen Verlust der „Heimat“ für deren Bewohner. In dem ländlich geprägten Gebiet mit größtenteils gewachsenen Dorfgemeinschaften waren und sind die Folgen der Umsiedlung für hier lebenden Menschen oftmals sehr weitreichend. Entsprechend formierte sich insbesondere in dieser Phase großer Widerstand, der in den Medien seinerzeit omnipräsent war.

Daneben begann Mitte der 1970er-Jahre die öffentliche Diskussion um die Folgen der Sumpfungsmaßnahmen zur Trockenlegung der Tagebaue, die den Abbau der Braunkohle erst ermöglichen. Die Grundwasserabsenkung hat Auswirkung auf den Wasserhaushalt der gesamten Region weit über den Tagebaubereich hinaus. So kommt es zu Schädigungen grundwasserabhängiger Feuchtgebiete wie etwa im Naturpark Schwalm-Nette. Auch die Trinkwassergewinnung wird durch die durch das Abpumpen bedingte Verlagerung unterirdischer Wasserscheiden mitunter massiv beeinträchtigt. Dies führt zu einer zunehmenden Mineralisierung und Aufhärtung des Wassers. Dies ist insofern als besonders problematisch anzusehen, da die Niederrheinische Bucht das größte Grundwasserreservegebiet Nordrhein-Westfalens beherbergt.

Ein weiteres seinerzeit bereits bekanntes Problem des Braunkohleabbaus ist die Emission von mit Schwermetallen und Radionukliden belastetem Staub. Man betrachtete zu diesem Zeitpunkt jedoch primär nur den sichtbaren und nicht inhalierbaren Grobstaub mit einem Partikeldurchmesser von mehr als 10 µm.

Allgemein nahezu unbekannt war lange Zeit hingegen, dass durch den Abbauvorgang der Braunkohle im Tagebau und dort die offene Zwischenlagerung der Kohle in sogenannten Kohlebunkern am Tagebaurand erhebliche Mengen Feinstaub freigesetzt werden. Erstmals aufmerksam auf dieses Problem machten der Bund für Umwelt und Naturschutz

Deutschland e.V. (BUND) und die lokale Bürgerinitiative Bürgergemeinschaft e.V. für die Gemeinde Niederzier (BG Niederzier) im September 2003². Das nordrhein-westfälische Umweltministerium veranlasste daraufhin erste mobile Messungen in Tagebaunähe. Die gemessenen Feinstaub-Tagesmittelwerte (PM₁₀) überschritten dabei an teils deutlich mehr als 35 Tagen im Jahr den in der EU-Richtlinie 1999/30/EG vorgegebenen Grenzwert von 50 µg/m³.

Als Konsequenz stellten die zuständigen Bezirksregierungen Düsseldorf und Köln Aktionspläne – im weiteren Verlauf Luftreinhaltepläne genannt – auf. Vom Umweltministerium NRW wurde auf Basis der Pläne die Einrichtung zusätzlicher stationärer Messstationen für die Luftqualitätsüberwachung im Rheinischen Braunkohlerevier beauftragt, die vom Landesamt für Natur, Umwelt und Verbraucherschutz (LANUV) NRW betrieben werden.

Die Aufstellung der Luftreinhaltepläne und die damit verbundenen Maßnahmen haben in den vergangenen Jahren zu einer Reduzierung der Feinstaubbelastung im Tagebauumfeld geführt. Das zeigt die tendenziell sinkende Zahl an Grenzwertüberschreitungen in den letzten Jahren im Vergleich zum Zeitraum vor Inkrafttreten der Luftreinhaltepläne. Das Thema hat jedoch nicht an Relevanz verloren, da immer noch eine hohe Belastung vorhanden ist, die durch den Klimawandel möglicherweise wieder steigen könnte. Lange niederschlagsarme Phasen wie in den Sommerhalbjahren 2018, 2019 und 2020 führen durch die geringere Auswaschung zu einer höheren Staubbelastung - inklusive Feinstaub - in der Luft.

Die Einrichtung und der Betrieb von Messstationen sind jedoch sehr kostenintensiv. Es stehen daher nur wenige punktuelle Messdaten zur Verfügung, die für das gesamte betrachtete Gebiet eine bedingte Aussagekraft haben. Um eine valide Aussage für die Fläche treffen zu können, ist die möglichst genaue Kenntnis der räumlichen Verteilung der Feinstaubbelastung von wesentlicher Bedeutung.

² BUND Landesverband NRW, „Braunkohle und Gesundheit - Das Feinstaub-Problem“, Pulheim 2006

Diese Masterarbeit soll dazu beitragen, ein Schätzverfahren für die räumliche Verteilung der Feinstaubbelastung (PM₁₀) im Rheinischen Braunkohlerevier und des teils sehr dicht besiedelten Umlandes mit den Methoden der Geoinformatik zu entwickeln und zu validieren. Das Verfahren soll mit einem geringen Arbeitsaufwand und frei verfügbaren Daten Ergebnisse hoher Genauigkeit liefern. Im Fokus steht dabei eine möglichst realistische Einschätzung der Feinstaubbelastung der Bevölkerung im Untersuchungsgebiet.

Die Ergebnisse der Masterarbeit könnten sowohl für die vom Tagebau betroffenen Kommunen interessant sein, die so zeitnah einen Überblick über die Luftbelastung erhalten und auf dieser Basis Maßnahmen ergreifen können, als auch für Umweltverbände, die die Entwicklung der Feinstaubbelastung im gesamten Rheinischen Revier weiter verfolgen und bewerten wollen.

1.2 Problemstellung und Relevanz

Das Thema Feinstaub und dessen schädliche Wirkung auf die menschliche Gesundheit sind in den letzten Jahren verstärkt in den Fokus der Öffentlichkeit gerückt. So wird in Deutschland aktuell eine kontroverse Diskussion über Fahrverbote für Diesel-PKW in städtischen Ballungsräumen geführt, bei der auch die regelmäßige Überschreitung der Feinstaub-Grenzwerte eine wesentliche Rolle spielt.

Als Feinstaub (PM₁₀) bezeichnet man Partikel unterschiedlicher chemischer Zusammensetzung und Herkunft, deren Durchmesser kleiner als 10 Mikrometer ist, und die sich daher in der Atmosphäre physikalisch weitgehend wie Gase verhalten. Die gesundheitsschädliche Wirkung von Feinstaub auf den menschlichen Organismus und die damit verbundene erhöhte Mortalität konnten durch verschiedene Zeitreihen- und Kohortenstudien belegt werden (Englert, 2007, Pope und Dockery, 2006).

Feinstaub wird primär über die Lunge aufgenommen. Ein Bezug zu einem gehäuften Auftreten von Atemwegserkrankungen scheint daher naheliegend und konnte durch zahlreiche Studien belegt werden (Farina et al., 2013, Tabaku et al., 2011). Sehr kleine - ultrafeine - Partikel ($< 2,5 \mu\text{m}$) können tief in die Lunge eindringen, hierüber in die Blutbahn gelangen und je nach chemischer Zusammensetzung zu verschiedenen weiteren Erkrankungen des Herz-Kreislaufsystems, des Stoffwechsels wie z.B. Diabetes oder auch

zu Veränderungen der DNA mit Krebs als Folge führen (Vidale und Campana, 2018, De Prins et al., 2013, Chen et al., 2013).

Remy et al. (2010) konnten bei einer Untersuchung der Auswirkung einer dauerhaften Feinstaubbelastung (PM_{10}) von mehr als $20 \mu\text{g}/\text{m}^3$ - Grenzwert EU-Direktive - auf die menschliche Gesundheit in den drei Belgischen Städten Brüssel, Antwerpen und Lüttich belegen, dass 5,5 Prozent der Sterberate auf die erhöhte PM_{10} -Belastung zurückzuführen sind.

Verkehr, Industrie und Haushalte sind durch chemische (Verbrennung) und physikalische Prozesse (Verarbeitung, Abrieb, etc.) als anthropogene Feinstaubquellen hinlänglich bekannt. Weniger bekannt ist dagegen, dass auch andere anthropogene Quellen wie z.B. der offene Abbau von oberflächennahen Bodenschätzen in Tagebauen eine mitunter erhebliche Belastung für die Bewohner des direkten Umfeldes darstellen (Ghose und Majee, 2007, Rojano et al., 2018, Pless-Mulloli et al., 2000).

Mit zunehmender Sensibilisierung von Politik und Behörden für dieses Thema werden zusätzliche Messstationen in Tagebaunähe eingerichtet. So betreibt beispielsweise das Landesamt für Natur, Umwelt und Verbraucherschutz (LANUV) Nordrhein-Westfalen seit einigen Jahren mehrere Messstationen im Rheinischen Braunkohlerevier, welches inmitten des dicht besiedelten Städtevierecks Düsseldorf-Mönchengladbach-Aachen-Köln im Westen von Nordrhein-Westfalen liegt.

Die Feinstaubbelastung in besiedelten, vor allem in dicht besiedelten urbanen Räumen zu kennen, um die Gefährdung für die dort lebende Bevölkerung einschätzen zu können, ist daher von großer Bedeutung. Jedoch beschränken sich Feinstaubmessungen insbesondere aus Kostengründen meist auf wenige Messorte, bei denen entweder eine hohe Belastung durch Industrie oder Verkehr zu erwarten ist (Emittentennähe), oder die Hintergrundbelastung gemessen werden soll. Die (Echtzeit-)Feinstaubbelastung an den einzelnen Messorten zu kennen, ist für die Beurteilung der Belastung der zwischen den Messstationen liegenden Siedlungsflächen nicht ausreichend. Es können anhand der punktuellen Messwerte nur begrenzt Rückschlüsse gezogen werden. Für eine Beurteilung der Belastung dieser Siedlungsflächen wäre es viel interessanter zu wissen, wie sich die räumliche Verteilung der Feinstaubkonzentration darstellt.

Räumliche Interpolationsmethoden – in Kombination mit geostatistischen Verfahren -, die mittlerweile zum Umfang vieler GIS-Produkte proprietärer und OpenSource-Anbieter

gehören, bieten hierfür geeignete Möglichkeiten (Li und Heap, 2014, Pinichka et al., 2017, Oliver und Webster, 2014).

Die Eignung univariater Interpolationsverfahren wurde in den vergangenen Jahrzehnten hinlänglich diskutiert und darf mit den nicht unerheblichen Einschränkungen als bekannt vorausgesetzt werden. Zur Erzielung genauerer Ergebnisse wurden in neueren Studien weitere Parameter in die Berechnung der Schätzwerte einfließen lassen. So entwickelten Brauer et al. (2003) als Teil einer multizentrischen Studie zum Auftreten von Asthma in der Kindheit durch verkehrsbedingte Luftverschmutzung eine Methode zur Bestimmung der individuellen Belastung der Bevölkerung, die neben den reinen Schadstoffmesswerten noch die Parameter Verkehrsnetz und Verkehrsdichte berücksichtigt. Dabei wurde in einem bestimmten Puffer um die Messstation beispielsweise mittels GIS die Verkehrsweeinfrastuktur ermittelt und in die Regressionsanalyse zur Bestimmung der Schätzwerte einfließen lassen (Brauer et al., 2003). Das Forschungsteam konnte beweisen, dass dieses Vorgehen bessere Ergebnisse bei der Regressionsanalyse liefert als bei der ausschließlichen Berücksichtigung der Messwerte.

Einen ähnlichen Ansatz verfolgen Janssen et al. (2008), die zur Bestimmung der Schätzwerte für NO_2 und PM_{10} neben den Messwerten auch die Landnutzung auf Basis der CORINE Land Cover data (CLC) in einem Radius von 2 km um jede Messstation berücksichtigen. Die Anteile der vorher gewichteten Landnutzungsklassen werden berechnet und fließen als spezifischer Faktor für jede Messstation in das Ordinary Kriging (OK) der Messwerte ein (Janssen et al., 2008).

Janssen et al. (2008) konnten beweisen, dass das von ihnen entwickelte sog. RIO-Modell unter Einbeziehung der Landnutzung bei gering topographisch strukturierten Ländern wie Belgien im Vergleich zu Standard-Interpolationsverfahren wie OK oder IDW bessere Ergebnisse liefert.

Neben der Landnutzung können verschiedene weitere Größen wie z.B. Straßenverkehr, Bevölkerungsdichte, Meteorologie oder Topographie als zweiter Parameter für die Berechnung der Schätzwerte verwendet werden (Ross et al., 2007, Beelen et al., 2009).

1.3 Untersuchungsgebiet

1.3.1 Rheinisches Braunkohlerevier

Das Rheinische Braunkohlerevier als Kernbereich des Untersuchungsgebietes erstreckt sich über den mittleren und nördlichen Teil der Niederrheinischen Bucht und den äußersten Süden des Niederrheinischen Tieflandes. Bezogen auf die Verwaltungsgrenzen im Bundesland Nordrhein-Westfalen liegt der überwiegende Teil im Regierungsbezirk Köln und der verbleibende Teil (nördlichen Bereich des Tagebaus Garzweiler I/II) im Regierungsbezirk Düsseldorf. Es ist mit einer Fördermenge von 86,3 Millionen Tonnen Braunkohle im Jahr 2018 das größte Abbaugbiet in Deutschland vor dem Lausitzer Revier (60,7 Mio. t) und dem Mitteldeutschen Revier (19,3 Mio. t).

Das Rheinische Revier wird in vier Teilgebiete (Nordrevier, Westrevier, Mittleres Revier und Südrevier) unterteilt. Das südwestlich von Köln gelegene Südrevier ist bereits seit den 1960er-Jahren ausgekohlt und wurde vollständig rekultiviert. Das unmittelbar westlich von Köln gelegene mittlere Revier ist ebenfalls ausgekohlt und wurde weitgehend rekultiviert. Aktive Tagebaue befinden sich noch im Nordrevier, das im Dreieck Mönchengladbach/ Düren/Köln liegt, und im nordwestlich von Aachen gelegen Westrevier. Mit Fortschreiten des Tagebaus laufen Rekultivierungsmaßnahmen auch im Nord- und Westrevier bereits seit Jahrzehnten. Braunkohle wird derzeit noch in den Tagebauen Garweiler I/II, Hambach (beide Nordrevier) und Inden I/II (Westrevier) abgebaut.

Braunkohlenförderung und -verwendung in Deutschland 2018

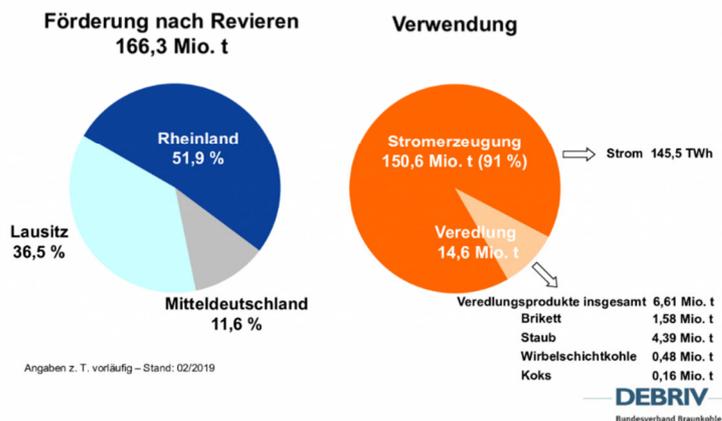


Abb. 2 Braunkohleförderung und Verwendung in Deutschland 2018³

Die Zukunft der Braunkohleförderung im Rheinischen Revier scheint derzeit ungewiss. In dem im Januar 2019 veröffentlichten Abschlussbericht der Kohlekommission der Deutschen Bundesregierung⁴ ist der vollständiger Kohleausstieg für Ende 2038 (2035) vorgesehen. Durch die aktuell geführte Debatte über den voranschreitenden Klimawandel könnte die Planung für einen vorzeitig Ausstieg deutlich an Dynamik gewinnen.

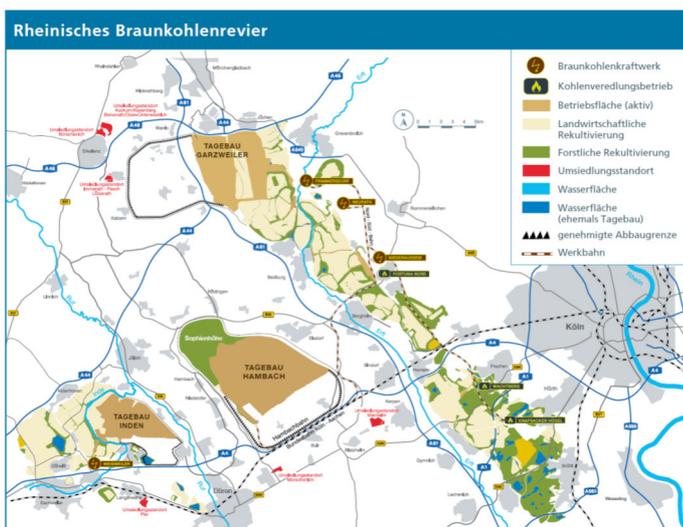


Abb. 3 Revierkarte Rheinisches Braunkohlerevier⁵

³ Quelle: Deutscher Braunkohlen-Industrie-Verein e.V. (DEBRIV), Stand: 02/2019

⁴ BMWi, Abschlussbericht Kommission „Wachstum, Strukturwandel und Beschäftigung“, Berlin 2019

⁵ Quelle: Deutscher Braunkohlen-Industrie-Verein e.V. (DEBRIV), Stand:08/2017

Der Tagebau Hambach ist mit einer Betriebsfläche von ca. 4.400 ha und einer jährlichen Kohlefördermenge von ca. 40 Millionen Tonnen Braunkohle der größte Tagebau im Rheinischen Revier, gefolgt vom Tagebau Garzweiler und dem deutlich kleineren Tagebau Inden.

Tab. 1 Flächen, Kohleinhalte und Fördermengen der Tagebaue 6

Tagebau	Garzweiler	Hambach	Inden
Größe genehmigtes Abbaufeld (ha)	11.400	8.500	4.500
Betriebsfläche (ha)	3.200	4.380	1.700
Kohleinhalt (Mio. t)	1.100	1.350	260
Kohleförderung pro Jahr (Mio. t)	35	40	20
Abraumleistung pro Jahr (Mio. m ³)	130 -150	220 - 250	70 -90



Abb. 4 Tagebau Garzweiler II im Januar 2018⁷

1.3.2 Tagebauumfeld

Die Bedeutung der Lage der Tagebauareale zeigt sich erst, wenn man das größtenteils dicht besiedelte Umland des Rheinischen Reviers betrachtet. Dies stellt einen entscheidenden Unterschied zu den anderen deutschen Braunkohlerevieren dar, bei denen auch das Umland ländlich geprägt und daher dünn besiedelt ist.

⁶ Quelle: RWE AG, Stand:12/2017

⁷ Quelle: Eigene Fotoaufnahme, Datum der Aufnahme: 27.01.2018

musste, kaum größere Waldflächen vorhanden. Es ist beabsichtigt, im Rahmen der Rekultivierungsmaßnahmen den Waldanteil im Revier im Vergleich zum Ausgangsniveau vor dem Abbau zu erhöhen.⁸

Dicht besiedelte urbane Bereiche befinden sich vor allem nördlich (Mönchengladbach, Kreis Neuss) und östlich (Rheinschiene), aber auch südwestlich (Städteregion Aachen) des Reviers. Die Topographie des Untersuchungsgebietes ist insgesamt wenig strukturiert. Lediglich am südwestlichen (Nordeifel) und nordöstlichen (Bergisches Land) Rand sind kleine Teile des Rheinischen Schiefergebirges einbezogen.

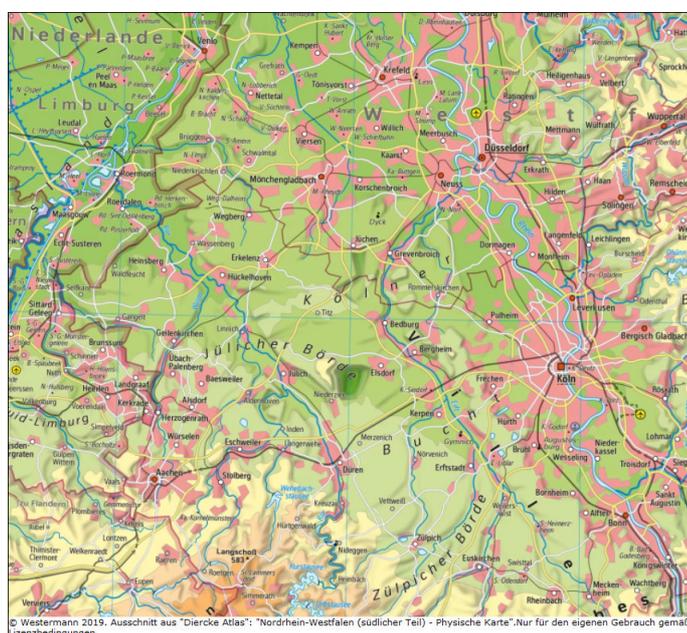


Abb. 6 Untersuchungsgebiet (Physische Karte), Westermann, 2019

Im definierten Untersuchungsgebiet betreibt das Landesamt für Natur, Umwelt und Verbraucherschutz (LANUV) Nordrhein-Westfalen insgesamt 25 Messstationen, an denen die Feinstaubbelastung (Hintergrund, Verkehr und Industrie) gemessen wird. Um die Interpolationsergebnisse nicht zu verfälschen, wurden alle Messstationen im Untersuchungsgebiet ausgeklammert, an denen die verkehrsbedingte Feinstaubbelastung gemessen wird. Es verbleiben 16 Stationen, die in der Karte verzeichnet sind. Vier Stationen liegen unmittelbar am Tagebaurand (Elsdorf-Berrendorf, Jackerath, Jüchen-Hochneukirch und Niederzier). An allen anderen Stationen, an denen überwiegend die Hintergrundbelastung gemessen wird, befinden sich in einer maximalen Entfernung von

⁸ RWE Power AG, Rekultivierung im Rheinland, Essen/Köln 2008, S. 10

38 km – euklidische Distanz – zum Tagebaurand. Die Messstationen sind nicht gleichmäßig über das Untersuchungsgebiet verteilt. So liegen westlich des Reviers keine Messstationen.

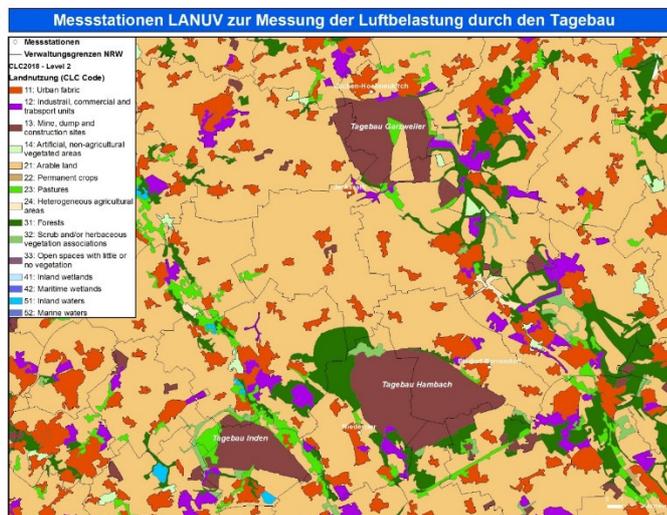


Abb. 7 Messstationen LANUV am Tagebaurand

1.4 Forschungsziele und Einordnung der Arbeit

1.4.1 Einordnung der Arbeit

Ziel dieser Arbeit ist die vergleichende Prüfung der Eignung geostatistischer Interpolationsverfahren (Kriging) unter Einbeziehung der CORINE Land Cover Landnutzungsdaten als zweiten Parameter (Kokriging) für die Darstellung der räumlichen Verteilung der Feinstaubbelastung (PM10) in Tagebauarealen und deren Umgebung mit wenigen Messstationen am Beispiel des Rheinischen Braunkohlereviers und der angrenzenden, teils dicht besiedelten urbanen Gebiete. Durch den Vergleich mit dem univariaten Interpolationsverfahren Simple Kriging soll die Leistungsfähigkeit des entwickelten Modells ermittelt werden. Die Wahl der Kriging-Methoden basiert auf den Ergebnissen mehrerer wissenschaftlicher Artikel, die die Eignung einzelner Verfahren für unterschiedliche Problemstellungen aufzeigen (Li und Heap, 2011, Li und Heap, 2014, Oliver und Webster, 2014, Akkala et al., 2010).

Im Fokus steht demzufolge die Bestimmung der tatsächlichen Feinstaubbelastung für die Bevölkerung an jedem beliebigen Ort im Untersuchungsgebiet anhand der vorliegenden, validierten Messwerte (Tagesmittelwerte). Ein wichtiges zusätzliches Ziel ist die

Entwicklung einer Methode, die mit wenigen Messdaten und zusätzlich frei verfügbaren Daten eine hohe Ergebnisgenauigkeit bei geringem Aufwand und niedrigen Kosten aufweist. Einrichtung und Betrieb von Messstationen sind mit hohen Kosten verbunden.

1.4.2 Forschungsfrage

Hieraus leitet sich die Hypothese ab, die es zu beweisen oder widerlegen gilt:

„Kriging liefert unter Einbeziehung der Landnutzung für die Darstellung der raumzeitlichen Verteilung der Feinstaubbelastung (PM_{10}) im Bereich großflächiger Tagebauareale genauere Ergebnisse“

1.4.3 Operative Teilziele - Lösungsansatz

Aus den validierten Tagesmittelwerten der Feinstaubbelastung (PM_{10}) an ausgewählten Messstationen im Untersuchungsgebiet soll im ersten Schritt nach einer Trendbereinigung und Transformation mittels Kriging die flächenhafte Verteilung der Feinstaubkonzentration bestimmt werden. Die Validierung der Ergebnisse der univariaten Interpolationen erfolgt durch QQ-Plots und RMSE-Betrachtung.

Im zweiten Schritt werden die CORINE Landnutzungsdaten (CLC2018) für die Berechnung der Interpolationsflächen hinzugezogen. Diese fließen als zweiter Parameter für das Kokriging durch die Nutzung des im belgischen RIO-Modell (Kapitel 2.6) verwendeten Landnutzungsindikators β in modifizierter Form ein. Vorhandene Trends im Untersuchungsgebiet sind hierbei zu berücksichtigen. Die Validierung der Ergebnisse dieser multivariaten Interpolationen erfolgt ebenfalls durch QQ-Plots und RMSE-Betrachtung. Im dritten Schritt werden die Interpolationsergebnisse miteinander verglichen und die Eignung der einzelnen Methoden im Hinblick auf die aufgestellte Hypothese diskutiert.

Wie verschiedene Studien bereits zeigen konnten, ist bei einer Einbeziehung der Landnutzungsdaten grundsätzlich mit einer weiteren Erhöhung der Genauigkeit der Schätzwerte zu rechnen (Janssen et al., 2008, Beelen et al., 2009).

1.5 Nicht-Ziele

- Betrachtung des Erklärungswertes des Modells im Hinblick auf die Ermittlung potentieller Emittenten (Prozessmodellierung)
- Betrachtung des Vorhersagewertes des Modells - Wie könnte sich die Feinstaubbelastung entwickeln?
- Einbeziehung der Topographie des Untersuchungsgebietes unter Verwendung eines Höhenmodells

Bei einer stark strukturierten Topographie wie beispielsweise in den Alpenländern wäre die zusätzliche Einbeziehung eines Höhenmodells unumgänglich. Im wenig strukturierten Untersuchungsgebiet kann dieser Einfluss jedoch vernachlässigt werden (Nagl und Spangl, 2007).

1.6 Zielgruppe

Ein wesentlicher Aspekt dieser Arbeit ist die Verwendung frei verfügbarer oder leicht zu beschaffender Daten als Basis für die interpolationsgestützte Ermittlung der flächenhaften Feinstaubbelastung. Deterministische und probabilistische Interpolationsverfahren gehören heute als Werkzeuge für diese Aufgabe zum Standard vieler proprietärer und freier GI-Softwareprodukte, die in Behörden und der freien Wirtschaft eingesetzt werden. Unter diesen Voraussetzungen ist mit dem angewendeten Verfahren eine flächenhafte Ermittlung der Feinstaubbelastung in topographisch gering strukturierten Gebieten schnell und mit geringen Kosten umsetzbar.

Dies könnte für Umweltbehörden interessant sein, die aufgrund nationaler und EU-weiter Richtlinien eine Bewertung der Luftqualität - in Echtzeit - für ihren Zuständigkeitsbereich vornehmen müssen oder die Luftgüte kartographisch präsentieren wollen. Mithilfe der Methode erstellten Luftgütekarten könnten auch bei Umweltorganisationen Verwendung finden, um beispielsweise eine Problemstellung zu veranschaulichen und die eigenen Ziele zu untermauern. Auf Basis der Ergebnisse dieser Arbeit sind weitere Untersuchungen denkbar, die in stärker topographisch strukturierten Gebieten ein Höhenmodell einbeziehen oder unter Berücksichtigung meteorologischer Daten eine weitere Erhöhung der Genauigkeit der Interpolationsergebnisse anstreben.

2 Methodik

2.1 Luftqualitätsmessungen in Nordrhein-Westfalen

2.1.1 Messverfahren

Das im Jahr 2007 gegründete Landesamt für Natur, Umwelt und Verbraucherschutz NRW (LANUV)⁹ ist als wissenschaftlich-technische Fachbehörde des Landes Nordrhein-Westfalen neben anderen Themenbereichen auch für die Überwachung der Luftqualität zuständig. Das Landesamt betreibt zur Erfüllung dieser Aufgabe das Luftqualitätsüberwachungssystem (LUQS), welches kontinuierliche und diskontinuierliche Messungen umfasst und verschiedene Möglichkeiten zur Darstellung der gemessenen Luftqualitätsdaten bietet. Kontinuierliche Messungen führt das LANUV zur Bestimmung der Außenluftkonzentration von Feinstaub (PM₁₀/PM_{2.5}), Ozon (O₃), Schwefeldioxid (SO₂), Stickstoffdioxid (NO₂), Stickstoffmonoxid (NO) sowie zur Ermittlung meteorologischer Parameter wie Temperatur, Windrichtung, Windstärke und Niederschlag durch. Das Messnetz umfasst aktuell insgesamt 164 über das Bundesland NRW verteilte Messstationen. An 78 Stationen werden im Rahmen kontinuierlicher Messungen halbstündlich die Feinstaubwerte (PM₁₀) ermittelt.

Aus Sicht der Messtechnik ist die Messung von PM₁₀ schwieriger und problembehafteter als die Messung von anderen Luftschadstoffen wie O₃, SO₂, NO₂ oder NO, da Feinstaub im Gegensatz zu den anderen genannten Verbindungen chemisch nicht einheitlich zusammengesetzt ist. Feinstaub ist eine Kombination aus vielen Stoffen wie beispielsweise Nitraten, Sulfaten, Kohlenstoffverbindungen, Ammoniumverbindungen, Siliziumverbindungen oder auch Wasser, die unterschiedliche physikalische und chemische Eigenschaften haben (Harrison et al., 1997).

Die genaueste Methode der Feinstaubmessung - die sogenannte Referenzmethode - ist das Sammeln des Feinstaubes auf einem Filter, der nach der Messung im Labor gewogen

⁹ Landesamt für Natur, Umwelt und Verbraucherschutz NRW (LANUV); Hauptsitz: Leibnizstr. 10, 45659 Recklinghausen; Webseite LANUV zum Thema Luft: www.lanuv.nrw.de/umwelt/luft

wird. Diese Methode liefert zwar die präzisesten Werte, ist aber sehr aufwändig. Zudem können nur Tagemittelwerte bestimmt werden, die nicht zeitnah zur Verfügung stehen.

Für die kontinuierliche Messung von Feinstaub gibt es verschiedene Methoden, die hier nicht näher betrachtet werden sollen. Diese Methoden bieten den Vorteil zeitnah regelmäßig aktualisierte Messwerte zu erhalten, die erhöhte Konzentrationen schnell erkennen lassen und somit eine direkte Analyse bzw. Maßnahmen ermöglichen. Die Messwerte sind jedoch weniger genau und müssen noch korrigiert, d.h. in einem Referenzverfahren kalibriert werden.

Hierzu wird für jedes kontinuierliche Feinstaub-Messverfahren ein eigener Kalibrierfaktor ermittelt, der zur Korrektur der gemessenen Werte verwendet wird. Zur Bestimmung der Faktoren greift man auf Vergleichsmessungen zurück. An einigen LANUV-Messstationen wird sowohl mit dem Referenzverfahren als auch mit den kontinuierlichen Messverfahren gemessen. Aus den Ergebnissen der Vergleichsmessungen wird dann der Kalibrierfaktor berechnet. Die Kalibrierung erfolgt jeweils nach Ablauf eines Kalenderjahres. Anschließend können die Ergebnisse der kontinuierlichen Messverfahren bereinigt werden.

2.1.2 Messstationen im Untersuchungsgebiet

Von den 25 Messstationen im Untersuchungsgebiet, an denen das Landesamt für Natur, Umwelt und Verbraucherschutz NRW kontinuierliche Messungen der Feinstaubbelastung (PM10) vornimmt, wurden 16 Stationen ausgewählt. Keine Berücksichtigung fanden Messstationen, die in unmittelbarer Nähe von Straßen die Feinstaubbelastung durch den Verkehr messen. Die Konzentration der meisten Luftschadstoffe nimmt mit zunehmender Entfernung von der Straße ab und geht in einer maximalen Entfernung von 570 m in die Hintergrundbelastung über; auch wenn dieser Trend für PM10 nicht so ausgeprägt ist (Karner et al., 2010). Die an diesen Stationen gemessenen Werte hätten somit bei den Schätzverfahren je nach Parametrisierung einen zu großen Einfluss auf weiter entfernt liegende Bereiche im Untersuchungsgebiet.

Tab. 2 Messstationen des LANUV NRW im Untersuchungsgebiet

Internes Kürzel	Stationsname	Attribut ¹⁰	UTM-Koordinaten		Meteorologische Parameter		
			E32 U	N32 U	Wind-richtung	Wind-stärke	Nieder-schlag
AABU	Aachen-Burtscheid	B	295020	5626578	x	x	x
BONN	Bonn-Auerberg	B	364754	5624181	x	x	
LOER	Düsseldorf-Lörick	B	341743	5679982			
ELSB	Elsdorf-Berrendorf	I	330989	5644503	x	x	
HUE2	Hürth	I	350417	5638209			
JACK	Jackerath	I	321314	5656678	x	x	
JHNC	Jüchen-Hochneukirch	I	322818	5663728	x	x	
CHOR	Köln-Chorweiler	B	351630	5654105	x	x	x
RODE	Köln-Rodenkirchen	B	358285	5639512	x	x	x
KREF	Krefeld-Linn	B	335631	5690013			
LEV2	Leverkusen-Manfort	B	360103	5654936			
MGRH	M.-gladb.-Rheydt	B	319970	5670169	x	x	x
NETT	Nettetal-Kaldenkirchen	B	304636	5689916	x	x	
NERH	Neuss (Rheinhafen)	B	339667	5676748	x	x	
NIZI	Niederzier	I	321990	5639919	x	x	
EIFE	Simmerath Eifel	B	307805	5614792	x	x	x

Das Landesamt stellt neben den aktuellen Daten auch validierte Tages-, Monats- und Jahresmittelwerte für PM₁₀ zur Verfügung.

2.1.3 Messdaten

Das LANUV hat am 18.02.2019 die folgenden validierten Messdaten für den Zeitraum vom 01.09.2018 bis 08.09.2018 im Excel-Format zur Verfügung gestellt:

- Feinstaub (PM10): 24 Stunden-Mittelwert [$\mu\text{g}/\text{m}^3$]
(Tabelle 3)
- Windgeschwindigkeit: 24 Stunden-Mittelwert [m/s]
(Anhang A, Tab. 13)
- Windrichtung vektoriell: 24 Stunden-Mittelwert [$^\circ$]
(Anhang A, Tab. 14)
- Niederschlag: 24 Stunden-Summe [mm]
(Anhang A, Tab. 15)

¹⁰ Abkürzungen im Feld Attribut: B = Background (Hintergrundbelastung), I = Industrie (Belastung durch Industrie inklusive Tagebau)

Tab. 3 Messwerte Feinstaub PM10 (24H Mittelwert)

Station (Kürzel)	Feinstaub PM10 24H [$\mu\text{g}/\text{m}^3$]							
	01.09.2018	02.09.2018	03.09.2018	04.09.2018	05.09.2018	06.09.2018	07.09.2018	08.09.2018
AABU	13,39	15,28	19,85	36,31	24,53	23,57	14,09	8,72
BONN	14,26	12,71	27,19	30,1	31,22	32,23	15,18	16,18
CHOR	13,73	11,86	16,13	27,12	28,45	26,46	16,26	13,19
EIFE	10,86	11,3	21,74	36,81	19,98	27,4	9,79	7,34
ELSB	15,46	16,54	30	40,08	40,91	34,55	14,19	13,62
HUE2	14,74	13,97	19,21	29,98	33,86	29,61	16,67	12,42
JACK	21,92	21,42	31,6	40,25	38,99	37,48	15,01	10,5
JHNK	13,06	11,26	26,88	37,44	48,81	28,84		
KREF	10,69	6,96	18,75	25,91	22,42	15,79	7,97	6,99
LEV2	6,7	4,78	10,59	18,55	18,78	20,25	9,38	6,04
LOER	11,92	11,11	21,24	30,79	27,99	29,82	18,27	16,2
MGRH	13,45	13,79	26,02	36,05	29,49	30	11,8	10,64
NERH	12,25	10,56	26,73	41,76	39,95	31,49	25,29	18,44
NETT	10,49	12,13	22,41	32,25	33,68		10,55	10,55
NIZI	21,8	20,3	29,22	40,28	37,6	29,54	13,41	10,78
RODE	13,19	14,2	19,16	25,71	26,63	29,56	14,78	11,82

Der Grenzwert¹¹ für den PM10-Tagemittelwert wurde an keiner der 16 Messstationen im Untersuchungsgebiet an keinem Tag im Untersuchungszeitraum vom 01.09.2018 bis 08.09.2018 überschritten.

Die Messdaten zur Windgeschwindigkeit, zur vektoriellen Windrichtung und zum Niederschlag sind für diese Arbeit nicht relevant, da keine meteorologischen Parameter berücksichtigt werden. Für potentielle weitere Forschungsarbeiten, die auf dieser Arbeit aufbauen, sind die meteorologischen Messdaten im Anhang A aufgeführt.

2.2 Deskriptive Statistik

Jeder Datensatz weist Charakteristika auf, die es mit den Parametern der deskriptiven Statistik zu erkunden gilt. Ein grundlegendes Verständnis der Daten ist die Voraussetzung für eine korrekte Durchführung aller weiteren Schritte. Betrachtet werden hierbei aräumliche Maßzahlen, die die oftmals zahlreichen empirischen Werte zu wenigen aussagekräftigen Maßzahlen zusammenfassen. Diese beschreiben die zentrale Tendenz, Streuung und Verteilung einer Stichprobe. Wesentliche Informationen können beispielsweise aus den Maßzahlen für Mittelwert, Spannweite (Range), Streuung und auch dem Variationskoeffizienten, der die Streuung in Relation zum Mittelwert ausdrückt, gezogen werden.

¹¹ Grenzwert Tagesmittelwert (PM10): 50 $\mu\text{g}/\text{m}^3$, maximale erlaubte Überschreitungen pro Jahr: 35

Der Modus wird nach der Schätzformel für metrische Daten von Bahrenberg et al. (1992) auf Basis der Histogramm-Klasseneinteilung in Kap. 2.3.2 berechnet:

$$Md = U + \left(\frac{f - fl}{2f - fl - fr} \right) * b$$

mit

- Md = Modus
- U = untere Klassengrenze der am stärksten besetzten Klasse
- f = Anzahl der Elemente in der am stärksten besetzten Klasse
- fl = Anzahl der Elemente in der linken Nachbarklasse
- fr = Anzahl der Elemente in der rechten Nachbarklasse
- b = konstante Klassenbreite

2.3 Explorative a-räumliche und räumliche Datenanalyse

Bevor es zur Anwendung der geostatistischen Schätzverfahren kommt, gilt es, einen detaillierten Überblick über und ein Verständnis für die a-räumlichen und räumlichen Variablen im Untersuchungsgebiet zu erlangen. Hierfür stellt die explorative räumliche Datenanalyse (ESDA) verschiedene Möglichkeiten wie beispielsweise zur Untersuchung der Lage-Verteileigenschaften zur Verfügung.

Der Zugang zu einem tieferen Einblick in die Eigenschaften der Daten erfolgt visuell über grafische Darstellungen der Häufigkeitsverteilung der Werte (Histogramm, Normalisierter Quantil-Quantil-Plot) und der räumlichen Relationen (Voronoi-Diagramm, Trendanalyse). Die ESDA-Ergebnisse werden für jeden Untersuchungstag separat betrachtet und ausgewertet. Die Interpolationen erfolgen auf Basis des in Kapitel 2.7.2 beschriebenen 2D + t Modells, welches die Daten aus dem gesamten Untersuchungszeitraum berücksichtigt.

2.3.1 Kurvendiagramm

Kurvendiagramme gehören nicht üblicherweise zu den Methoden der explorativen Datenanalyse. Da hier jedoch die PM₁₀-Tagesmittelwerte von acht Untersuchungstagen vorliegen, bietet das Liniendiagramm eine gute visuelle Möglichkeit, die Verläufe der Messwerte aller Messstationen miteinander zu vergleichen. Die a-räumlichen Darstellung hilft bereits gut bei der Identifizierung von Ausreißern.

Das Kurvendiagramm wurde mit Grapher 16¹² erstellt. Die vier speziell für die Luftqualitätsmessung in Tagebaunähe eingerichteten Messstationen sind sowohl im Diagramm (dickere Linien) als auch in der Tabelle (Fettschrift) hervorgehoben.

2.3.2 Histogramm

In einem Histogramm werden nach Einteilung der Variablenwerte in Klassen die absoluten oder relativen Klassenhäufigkeiten meist als Blockdiagramm graphisch dargestellt. Mit Hilfe des Histogramms ist es so möglich, einen schnellen Überblick über die a-räumliche Verteilung der Werte einer Variablen zu bekommen.

Zur besseren visuellen Erfassung des Histogramms sollten bei der Erstellung einige Grundsätze, die nicht allgemein gültig aber hilfreich sind, beachtet werden. Die Klassenbreiten sind konstant und die Klassengrenzen als einfache Zahlenwerte zu wählen. Zur zentralen Frage der Festlegung der Klassenzahl gibt es keine feste Regel, da dabei die Fragestellung zur beabsichtigten Aussage des Histogramms eine wichtige Rolle spielt (Bahrenberg et al., 1992). Grundsätzlich gilt, dass die Anzahl der Elemente und die Zahl der Klassen positiv korrelieren. Dabei kann das Verhältnis zwischen beiden nach der Faustregel von Sturges bestimmt werden (Bahrenberg et al., 1992).

$$k = 1 + \log_2 n \quad (n=\text{Anzahl der Messwerte})$$

Für die vorliegenden Messwerte bedeutet dies:

Untersuchungstag 1-5 (n=16):

$$k = 1 + \log_2 16 = 5,00$$

Untersuchungstag 6-8 (n=15):

$$k = 1 + \log_2 15 = 4,91 (\approx 5)$$

¹² GOLDEN SOFTWARE, LLC, 809 14th Street, Golden, Colorado 80401 USA

Tab. 4 Klasseneinteilung Histogramme

Klasse:	Klassengrenzen ($\mu\text{g}/\text{m}^3$):
1	[0,00;10,00]
2	(10,00;20,00]
3	(20,00;30,00]
4	(30,00;40,00]
5	(40,00;50,00]

Die Wahl der Klassenbreite und der Klassengrenzen orientiert sich am Minimum und Maximum der Messwerte aller Untersuchungstage (Minimum: $4,68 \mu\text{g}/\text{m}^3$, Maximum: $48,81 \mu\text{g}/\text{m}^3$), um die Vergleichbarkeit der Histogramme zu gewährleisten.

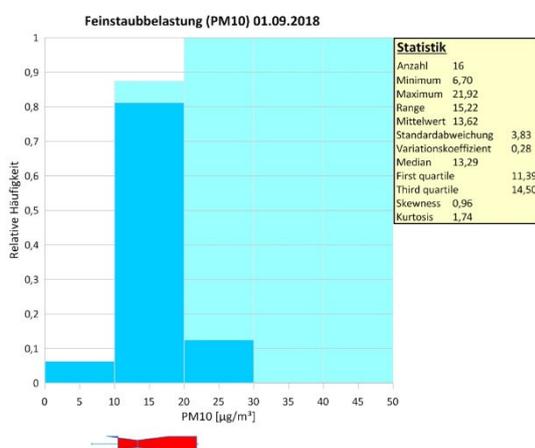
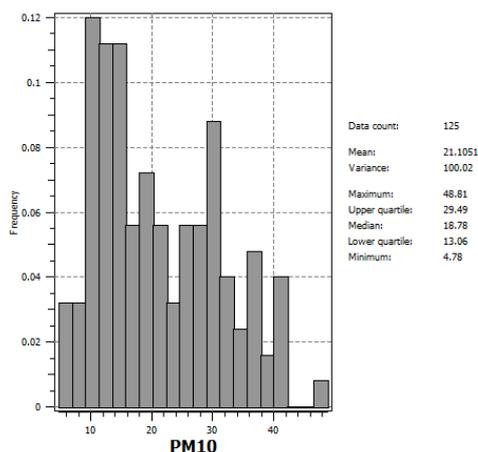


Abb. 8 Histogramm PM₁₀-Messwerte 01.09.2018

Im Histogramm werden die relativen Häufigkeiten (dunkelblaue Balken) und die kumulierten relativen Häufigkeiten (hellblaue Balken) dargestellt. Unter dem Histogramm ist ein Box-Whisker-Plot platziert. Der mittlere Querstrich kennzeichnet den Median. Die rot gefüllten Boxen sind horizontal vom zehnten und vom neunzigsten Perzentil begrenzt. Der horizontale Strich zeigt den Range der Messwerte.

Die Histogramme für die einzelnen Untersuchungstage sollen helfen, die Verteilungscharakteristika der tageweisen PM₁₀-Messungen im Untersuchungszeitraum aufzudecken. Wichtiger für die Variographie ist ein gutes Verständnis der Verteilung der gesamten PM₁₀-Messdaten und Landnutzungsindikatoren, da die Interpolationen auf Grundlage der raumzeitlichen Daten (2D +t) in SGeMS¹³ erfolgt (Kap. 2.7).

¹³ The Stanford Geostatistical Modeling Software (SGeMS), Download: <http://sgems.sourceforge.net>



Die Histogramme hierfür wurden mit SGeMS erstellt.

Abb. 9 Histogramm gesamte PM10-Messwerte

2.3.3 Normalisierter Quantil-Quantil-Plot

Eine weitere und genauere Möglichkeit, eine Aussage über die a-räumliche Verteilung der Daten zu treffen, bietet der normalisierte Quantil-Quantil-Plot. Dabei werden die empirischen Daten mit der Verteilfunktion der Standardnormalverteilung verglichen. Beim Vergleich betrachtet man die Abweichung von Punkten, welche die der Größe nach kumulativ geordneten Daten repräsentieren, in Bezug auf eine Referenzgerade, welche die Standardnormalverteilung darstellt. Je mehr Punkte und je weiter diese Punkte von der Bezugsgeraden entfernt liegen, desto weniger entspricht die empirische Verteilung einer Standardnormalverteilung.

Wenn eine deutliche Abweichung von der Normalverteilung vorzuliegen scheint, kann zusätzlich ein Vergleich der logarithmierten Daten mit der Standardnormalverteilung vorgenommen werden. Findet man hier geringe Abweichungen von der Normalverteilung, sollten die logarithmierten Daten für die Interpolationen verwendet werden. Die normalisierten QQ-Plots für die PM₁₀-Messdaten der acht Untersuchungstage wurden mit der ArcGIS-Erweiterung Geostatistical Analyst auf Basis der nicht-logarithmierten und der logarithmierten Daten erstellt und miteinander verglichen.

2.3.4 Raum-Zeit-Diagramm

Raum-Zeit-Diagramme gehören wie Kurvendiagramme üblicherweise nicht zu den Standard-ESDA-Methoden. Im Kontext der vorliegenden Messwertzeitreihe bieten diese

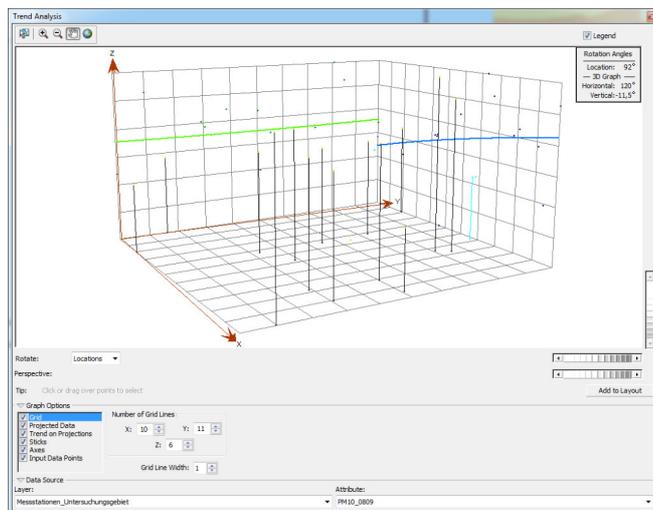
jedoch die Möglichkeit, die Messwertverläufe im Untersuchungszeitraum räumlich an den Messstationen verortet darzustellen. Visuell können so die räumliche Verteilung der Messstationen und die Messwertverläufe auf einen Blick erfasst werden. Räumliche und zeitliche Charakteristika werden somit schnell identifiziert und bieten Orientierungspunkte für die weitere explorative Betrachtung der Daten.

Das Raum-Zeit-Diagramm wurde mit Voxler¹⁴ in Form eines Scatterplots erstellt. Die 16 Messstationen des LANUV im Untersuchungsgebiet sind mit ihren 2-dimensionalen Koordinaten (UTM 32U) in der xy-Ebene dargestellt. Um die einzelnen Tage für jede Messstation gut sichtbar darstellen zu können, wurde durch eine Massstabsanpassung die z-Koordinate um den Faktor 3500 überhöht. Der unterste Würfel repräsentiert damit den ersten Messtag und der oberste Würfel den letzten Messtag.

2.3.5 Trendanalyse

Als räumlicher Trend werden richtungsabhängige Unterschiede in der Merkmalsausprägung der zu untersuchenden Variablen bezeichnet. Im Falle eines Trends sind die Variablenwerte exogen beeinflusst. Die Kenntnis um die Existenz von Trends ist für die weitere geostatistische Analyse und die Interpretation der Ergebnisse wesentlich. So benötigen beispielweise einige Interpolationsverfahren trendbereinigte Daten zur Berechnung der Interpolationsflächen. Unterschieden werden von vornherein erklärbare und nicht von vornherein erklärbare Trends. Letztere werden als Anisotropie bezeichnet und häufig nicht durch eine Trendbereinigung eliminiert. Im Rahmen der Trendanalyse werden im dreidimensionalen Raum Trendoberflächen generiert. Die räumliche Verortung erfolgt über die x- und y-Koordinaten, während die z-Koordinate den Attributwert darstellt.

¹⁴ GOLDEN SOFTWARE, LLC, 809 14th Street, Golden, Colorado 80401 USA



Nicht waagrecht verlaufende Linien zeigen Trends an. Trendverläufe können linear oder nichtlinear sein. Die Diagramme werden mit ArcGIS - Geostatistical Analyst - generiert.

Abb. 10 Trendanalyse PM₁₀-Messwerte 08.09.2018

Neben der obigen Trendanalyse, die räumliche Trends für die einzelnen Untersuchungstage detektiert, erfolgt zusätzlich über eine IDW Power 0 Interpolation (Kap. 2.5.2) eine raum-zeitliche Analyse von potentiell vorhandenen Trends. Diese Betrachtung ist wichtig, da Kriging und Kokriging unter Verwendung eines 2D + t Modells (Kap. 2.7.2) mit dreidimensionalen Daten durchgeführt wird. Die IDW-Interpolation erfolgt mit Voxler. Die raumzeitliche Änderung der PM₁₀-Konzentration kann so sehr gut visualisiert und Trends leichter identifiziert werden.

2.4 Regressions- und Korrelationsanalyse

Die Regressionsanalyse ist ein Strukturen prüfendes Verfahren und beschäftigt sich mit bi- und multivariaten Verteilungen. Als allgemein statistisches Verfahren ist sie auch für die räumliche Analyse geeignet. Der Blick richtet sich auf den Zusammenhang zwischen zwei oder mehreren Variablen. Zusammenhang bedeutet dabei zum einen die Stärke der Beziehung zwischen zwei oder mehreren Variablen, die durch Korrelationsanalysen bestimmt wird, und zum anderen die Form der Beziehung, die durch Regressionsanalysen ermittelt wird (Bahrenberg et al., 1992). Mit der Stärke der Beziehung wird ausgedrückt, wie gut die Streuung einer oder mehrerer unabhängiger Variablen die Streuung einer abhängigen Variablen zu erklären vermag. Die Zusammenhänge zwischen den Variablen können linear, exponentiell, U-förmig oder auch zufällig und damit nicht vorhanden sein.

Bei den meisten Regressionsanalysen müssen die Variablen metrisch skaliert sein, wie auch für die nachfolgend beschriebene Technik der linearen Einfachregression, die nur zwei Variablen, für die ein linearer Zusammenhang vermutet wird, betrachtet.

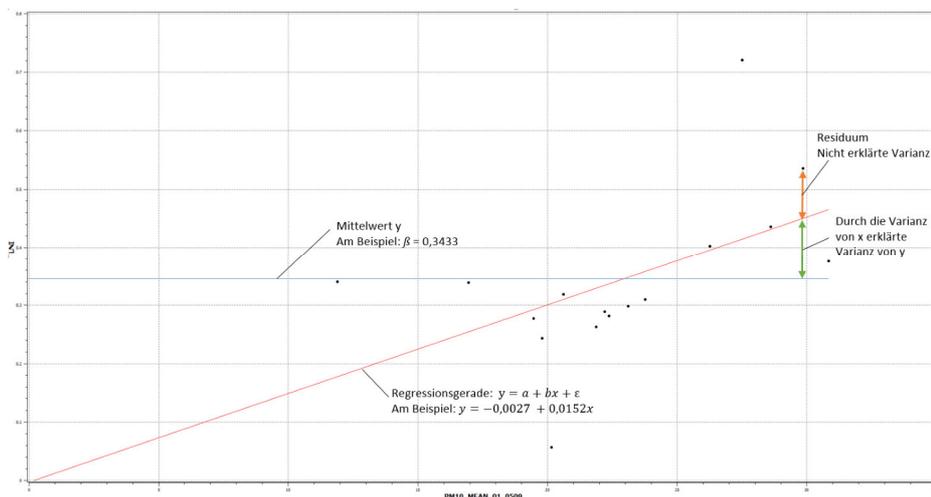


Abb. 11 Regressionsgerade PM₁₀-Mittelwerte - Landnutzungsindikator

Lineare Einfachregression (Bahrenberg et al., 1992):

$$y = a + bx + \varepsilon$$

mit a = Regressionskonstante (Schnittpunkt der Geraden mit der y-Achse)
 b = Regressionskoeffizient (positive oder negative Steigerung der Geraden)
 ε = Zufallsfehler (z.B. durch Messungenauigkeiten)

Zur Bestimmung des Gütemaßes einer Regressionsgeraden werden die senkrechten Abstände zwischen Messpunkt und Modellpunkt auf der Geraden (Residuen) quadriert, um zu vermeiden, dass sich bei der Addition gleichgroße positive und negative Abstände aufheben, aber auch, um diese zu gewichten. Die kleinste Summe der Abweichungsquadrate (Varianzen) für eine Regressionsfunktion zeigt das beste Ergebnis an. Die Analyse der Residuen ist ein wesentlicher Beitrag zur Beurteilung der Eignung einer Regressionsanalyse. Die folgenden beiden Punkte sollten dabei beachtet werden.

- Ausreißer wirken sich negativ auf die Bildung einer Regressionsgeraden aus
- Systematische Abweichungen von Residuen in bestimmten Bereichen - z.B. ein zunehmend wachsender Streubereich - zeigen an, dass für den unterstellten Zusammenhang eine oder mehrere weitere Variablen zu berücksichtigen sind (Bahrenberg et al., 1992)

Als Erweiterung der bivariaten Regressionsanalyse sei noch die multivariate Regressionsanalyse erwähnt, bei der der Zusammenhang von mehreren unabhängigen Variablen zur Erklärung der abhängigen Variablen untersucht wird. Hierbei kann der als Multikollinearität bezeichnete Fall auftreten, dass bei den unabhängigen Variablen untereinander Abhängigkeiten auftreten. Das meist nicht einfache zu erreichende Ziel ist dann eine Beschränkung auf die wesentlichen - also möglichst wenige - Variablen, die die abhängige Variable bestmöglich zu erklären vermögen.

Der Pearsonsche Produktmoment-Korrelationskoeffizient r_{xy} ist ein Maß für die Stärke des Zusammenhangs zwischen den Variablen x und y und kann einen Wert zwischen -1 und $+1$ annehmen (Bahrenberg et al., 1992).

- $r_{xy} = 1$: absolut gleichsinnige Korrelation
- $r_{xy} = 0$: keine Korrelation
- $r_{xy} = -1$: absolut gegensinnige Korrelation

Die Stärke des linearen Zusammenhangs ist umso größer, je näher r_{xy} bei 1 oder -1 liegt.

Im Kontext der Fragestellung dieser Arbeit dienen Regressions- und Korrelationsanalyse der Aufdeckung eines möglichen Zusammenhangs im Hinblick auf dessen Ausprägung zwischen den empirischen Feinstaubmesswerten und des RIO-Landnutzungsindikators β . Dabei soll die grundsätzliche Eignung von β als sekundäre Variable bei einem Kokring-Verfahren geprüft werden. Die Feinstaubdaten der Messstationen für die Analyse basieren auf dem Mittelwert der PM_{10} -Messwerte der ersten fünf Untersuchungstage. Die letzten drei Untersuchungstage werden bei der Mittelwertbildung nicht berücksichtigt, da von jeweils einer Station ein Messwert fehlt. Durch die starken Schwankungen der Messwerte im Untersuchungszeitraum (Kap. 3.2.1) könnte es zu Verzerrungen des Acht-Tage-Mittelwertes bei den betroffenen beiden Stationen kommen. Die statistische Signifikanz der Ergebnisse der Regressions- und Korrelationsanalyse kann auf dieser Grundlage angezweifelt werden, da die Mittelwertbildung nur stichprobenartig für einen sehr kurzen Zeitraum - ob fünf oder acht Tage - erfolgt. Die Ergebnisse, die aus einer Stichprobe gewonnen werden, sind daher als Schätzwerte für die Parameter der bivariaten Grundgesamtheit anzusehen (Bahrenberg et al., 1992). Eine Überprüfung der Korrelation mit weiteren Messdaten, die derzeit nicht zur Verfügung stehen, ist daher zur Absicherung des Ergebnisses anzuraten.

2.5 Interpolationsverfahren

Das Ziel von Interpolationsverfahren ist die Generierung einer kontinuierlichen Fläche von Schätzwerten auf Basis der an den Messpunkten gemessenen bzw. bestimmten Werten. Das Grundprinzip der Verfahren basiert auf der Annahme, dass räumliche Phänomene voneinander abhängig sind. Je geringer die räumliche Distanz zwischen den Phänomenen ist, desto ähnlicher sind ihre Merkmalsausprägungen. Tobler (1970) beschreibt diesen Zusammenhang im „ersten Gesetz der Geographie“: „everything is related to everything else, but near things are more related than distant things.“

2.5.1 Räumliche Autokorrelation

Die Kenntnis von Prozessen, die die Werteausprägung einer Variablen beeinflussen, ist als Vorüberlegung zur Variographie (Kapitel 2.5.3) essentiell. Diese Prozesse können exogen oder endogen verursacht sein. Bei exogen verursachten Prozessen liegt ein räumlicher Trend vor, bei dem die Werteausprägung einer Variablen durch eine oder mehrere weitere Variablen beeinflusst wird. Bei endogen verursachten Prozessen wird der Wert einer Variablen von den Werten derselben Variablen benachbarter Punkte beeinflusst. Diesen Zusammenhang bezeichnet man als räumliche Autokorrelation.

Bei der Aufdeckung von räumlicher Autokorrelation wird nach dem Vorhandensein von Stationarität gesucht. Stationarität liegt vor, wenn die Varianz der Werteausprägung von Punktpaaren nicht von der Lage im Untersuchungsgebiet, sondern nur von der Distanz und Richtung der Punktpaare abhängt. Ein Zustand der in der Realität selten vorkommt. Dennoch ist das Wissen um den Einfluss der Stationarität auf die (Mess-)Daten wichtig, um die Qualität der Ergebnisse geostatistischer Verfahren möglichst gut einschätzen zu können. Das Fehlen von Stationarität - Instationarität - weist zudem auf einen räumlichen Trend hin.

2.5.2 Deterministische Interpolationsverfahren

Diese Verfahren gewichten die räumliche Nähe der Messpunkte zu den zu schätzenden Punkten (Distanzkriterium). Nur diese paarweisen Distanzen gehen in die Berechnung

ein. Sie berechnen die Ähnlichkeit benachbarter Punkte rein mit Hilfe mathematischer Verfahren.

Eine Differenzierung Deterministischer Verfahren kann anhand zweier Kriterien vorgenommen werden. Zum einen erfolgt eine Unterscheidung zwischen globalen Methoden, die für jede Schätzung die gesamten Daten berücksichtigen, und lokalen Methoden, die nur eine räumliche Teilmenge der Daten heranziehen. Zum anderen werden exakte Methoden, bei denen die Messpunkte stets in der Interpolationsfläche liegen, und approximative Methoden, bei denen dies nicht zutreffen muss, unterschieden. Ein in vielen GIS-Programmen implementiertes und häufig verwendetes, einfaches deterministisches Verfahren ist Invers Distance Weighting (IDW) als lokale, exakte Methode. IDW gewichtet die Distanz zwischen Punktepaaren invers, sodass im Suchradius näher gelegene Punkte ein höheres Gewicht erhalten als weiter entfernt liegende Punkte. Die Distanz kann potenziert und somit noch höher gewichtet werden.

IDW eignet sich grundsätzlich für die Interpolation der PM_{10} -Werte als Vergleichsmethode mit Kriging und Co-Kriging. Ein wesentlicher Nachteil dieses vergleichsweise einfachen Verfahrens ist eine Schätzoberfläche, die je nach Gewichtung der Distanz mehr oder weniger stark zerklüftet ist und daher für die von den Messpunkten weiter entfernt liegenden Punkte keine zufriedenstellenden Ergebnisse liefert. Bei räumlich ungleichmäßig verteilten, teils geclusterten Messpunkten wie im Fall der berücksichtigten LANUV-Messstationen im Untersuchungsgebiet ist die Verwendung von IDW zudem nicht sinnvoll (PL und Srivastava, 1989). Die meisten räumlichen Merkmale variieren ohnehin auf eine so komplexe Weise, dass die Variation der Werte nicht rein deterministisch beschrieben werden kann (Haining et al., 2010). IDW wird aus den genannten Gründen daher zwar nicht als Vergleichsmethode in dieser Arbeit berücksichtigt, findet jedoch bei der Trendbereinigung Verwendung (Kapitel 2.7).

2.5.3 Variographie

Um eine Interpolation mit geostatistischen Verfahren wie Kriging, die neben dem Distanzkriterium auch autokorrelative Zusammenhänge zwischen den Messpunkten berücksichtigen, durchführen zu können, bedarf es der Aufdeckung dieser Beziehungsregeln durch die Variographie. Matheron (1963) definierte und nutzte erstmals

Variogramme auf Basis eines probabilistischen/stochastischen Ansatzes für eine Stichprobe, um die räumliche Variation der Werteausprägung zu analysieren.

Als Grundüberlegung dient die bereits diskutierte Annahme, dass in räumlicher Nähe liegende Punkte ähnlichere Werte aufweisen als weiter voneinander entfernt liegende Punkte. Die Varianz der Werte nimmt somit mit größer werdender Distanz der Punkte zu. Die Variographie basiert auf der Hypothese, dass Zufallsprozesse die Werteausprägung einer Variablen im Raum bestimmen. Diese Zufallsprozesse haben jedoch eine räumliche Abhängigkeit. Natürlich unterliegen die Prozesse physikalischen Gesetzen, die die Variation in der Werteausprägung beeinflussen. Eine Vielzahl interagierender Prozesse führt jedoch zu einem so komplexen Ergebnis, dass die Variation in der Werteausprägung als zufällig generiert erscheint (Webster, 2000).

Die Variographie gruppiert unter Annahme des Stationariätskriteriums Punktpaare, die in Bezug auf Richtung und Distanz ähnlich sind. Auf dieser Grundlage wird ein Variogramm-Modell entwickelt. Das sog. Semivariogramm stellt das Verhältnis zwischen der (Semi-)Varianz und der Distanz der Messwerte dar. Es zeigt keine richtungsspezifischen Unterschiede der Varianz auf. Für die Aufdeckung richtungsspezifischer Variationen, die sich nicht aus einem globalen Trend ableiten lassen, sondern von diesem unabhängig existieren (Anisotropie), werden Semivariogramm-Oberflächen verwendet. Anisotrope Phänomene sind nicht auf eine Ursache zurückführbar und damit aus geostatistischer Sicht Zufallsprozesse.

Das Gruppieren von Punktpaaren erfolgt in vier Schritten:

- 1) Messung der Distanzen von allen Messpunkten zu allen anderen Messpunkten über die quadratische euklidische Distanz.
- 2) Alle Punktpaare werden hinsichtlich ihrer Distanz und Richtung gemäß der Stationaritätsannahme in Bins gruppiert.
- 3) Für jede Gruppe wird der Mittelwert der Punktpaare (Zentroid) ermittelt.
- 4) Die Gruppenmittelwerte werden in das Semivariogramm eingetragen.

Die Distanz wird im Semivariogramm entlang der x-Achse in Form von Distanzintervallen (Bin-Breite), die als Lags bezeichnet werden, abgetragen. Bei Semivariogramm-Modellen, die alle Bins enthalten, können für jeden Lag auch mehrere Punkte eingetragen sein, da die Bins den Entfernungsklassen zugeordnet werden. Andere

Modelle mitteln die Varianzen in einem Lag, so dass für jeden Lag, falls dieser einen Wert enthält, genau ein Punkt dargestellt wird.

Als Orientierung zur Festlegung einer geeigneten Lag-Größe (lag_{size}) und Lag-Anzahl (lag_{num}) dient folgende Formel¹⁵:

$$lag_{size} * lag_{num} = \frac{1}{2} d_{max}$$

d_{max} ist die maximale Distanz eines Punktpaares im Untersuchungsgebiet.

Die Formel sollte jedoch nicht als Faustformel betrachtet werden, da häufig eine explorative Herangehensweise erforderlich ist, um optimale Parametereinstellungen zu erlangen. Dabei werden die Semivariogramme mit unterschiedlichen Parametereinstellungen vergleichend bewertet.

Die Festlegung der Lags und Bins beeinflusst sensitiv das Ergebnis. Geringe Lag-Distanzen, die einer hohen Anzahl Bins entsprechen, ermöglichen die Analyse lokaler Variationen. Die Gefahr liegt hierbei in einem steilen Anstieg der Semivariogrammkurve und somit eines Overfittings des Modells.

Berechnung der Semivarianz¹⁶:

$$\gamma(h) = \frac{1}{2} * \frac{1}{n(h)} \sum_{i,j=1}^{n(h)} (\chi_i - \chi_j)^2$$

mit χ_i = Variablenwert am Ort i
 χ_j = Variablenwert am Ort j
 h = Distanz zwischen i und j (Lag-Distanz)
 n = Anzahl der Punkte im Distanzintervall h

Es genügt, die Varianz von $(\chi_i - \chi_j)^2$ zu bestimmen, da die Varianz von $(\chi_j - \chi_i)^2$ gleich ist und daher nicht mehr berücksichtigt werden muss. Hieraus resultiert der Präfix Semi in den Wörtern Semivarianz und Semivariogramm.

¹⁵ UNIGIS MSc, Modul 8: Geostatistik, Lektion 10: Variographie, Seite 21

¹⁶ UNIGIS MSc, Modul 8: Geostatistik, Lektion 10: Variographie, Seite 10

Semivariogramm:

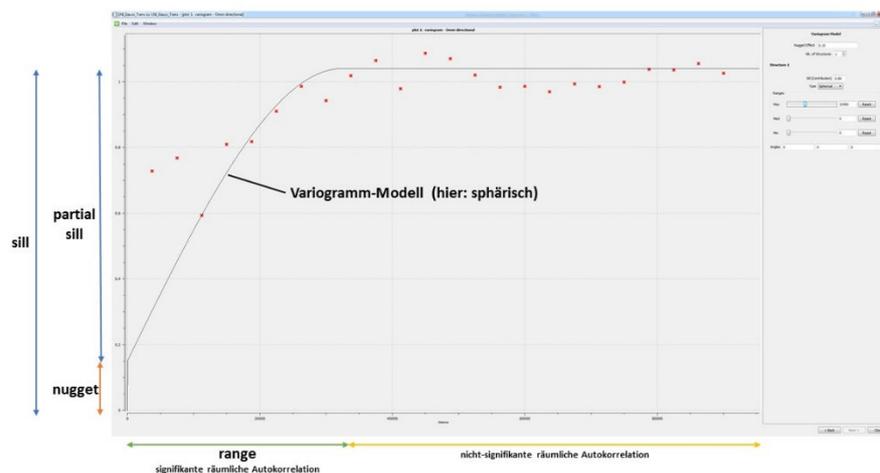


Abb. 12 Beispiel-Semivariogramm erläutert (β , gauß-transformiert)

Semivariogramm Begriffe:

Die roten Kreuze im Semivariogramm bilden die empirischen Semivarianzen ab. Die graue Linie repräsentiert das theoretische Semivariogramm, das optimal an die Punktwolke angepasst sein sollte. Die Anpassung erfolgt je nach verwendetem Tool manuell oder mittels einer Regressionsanalyse über alle Lags (Minimierung der Abweichquadrate) unter Verwendung eines Semivariogramm-Modells. Häufig verwendete Modelle sind das sphärische und das exponentielle Modell.

Der *range* bezeichnet das Distanzintervall, in dem die Semivarianzen zunehmen und die räumliche Autokorrelation signifikant ist. Dahinter bleiben die Semivarianzen mehr oder weniger konstant. Die räumliche Autokorrelation ist hier nicht mehr signifikant.

Mit *nugget* bzw. *nugget*-Effekt bezeichnet man geringe Semivarianzen im Nahbereich - in der Nähe des Koordinatenursprungs -, die durch Zufallsfehler (z.B. Messfehler) bedingt sind. Die Unschärfe des Modells steigt mit größer werdendem *nugget*-Wert.

sill bezeichnet den dem *range* entsprechenden Abschnitt entlang der y-Achse bis zum Erreichen des maximalen Semivarianzwertes.

partial sill ist der Bereich oberhalb des *nuggets* bis zum Erreichen des maximalen Semivarianzwertes. Dieser ist der Wert der Differenz von *sill* und *nugget*.

Die räumliche Variabilität von vielen untersuchten Merkmalen ist oft schlecht durch ein Variogramm charakterisierbar. Bei der Verwendung von Semivariogrammen ist dies zu beachten. So haben verschiedene Merkmale mit einer sehr unterschiedlichen räumlichen Variabilität oftmals sehr ähnliche Semivariogramme (Caers, 2005).

Als Voraussetzung für die Anwendung eines Kokriging-Verfahrens (Kap. 2.5.5), bei denen neben der primären Variablen (hier: PM_{10}) eine sekundäre Variable (hier: β , siehe Kap. 2.6.3) zur Berechnung der Schätzoberfläche herangezogen wird, ist neben den Semivariogrammen für die beiden Variablen auch ein Kreuzvariogramm erforderlich. Als multivariates Äquivalent zum Semivariogramm wird im Kreuzvariogramm anstatt der Varianz die Summe der Kovarianz der beiden Variablen in Abhängigkeit der Distanz berücksichtigt.

Sowohl die Semi- als auch die Kreuzvariogramme für diese Arbeit wurden mit SGeMS erstellt. In SGeMS sind mit dem sphärischen, exponentiellen und dem Gaußschen Modell nur drei Semivariogramm-Modelle implementiert, die für die Erstellung der benötigten theoretischen Semivariogrammen jedoch vollkommen ausreichend sind. Bei der Modellierung der Semivariogramme in SGeMS ist zu beachten, dass der *partial sill* verwendet wird (Sill Contribution).

2.5.4 Kriging-Verfahren

Kriging-Verfahren gehen über deterministische Verfahren (Kap. 2.5.2) hinaus, da sie nicht nur die Distanz zwischen den Stützpunkten sondern auch eine potentielle räumliche Autokorrelation der Messpunkte untereinander berücksichtigen (Kap. 2.5.1). Kriging ist eine geostatistische Methode und kombiniert mathematische und statistische Verfahren. Letztere gehen von der Annahme zufälliger räumlicher Variationen der Werteausprägungen aus. Die Schätzung der unbekanntenen Werte beruht auf Wahrscheinlichkeiten.

Merkmale von Kriging-Verfahren:

- Kombination mathematischer und statistischer Verfahren
- Räumliche Autokorrelation fließt in die Wertevorhersage ein
- Liefern neben den eigentlichen Interpolationsflächen (*prediction surfaces*) auch die dazu gehörigen Fehlerflächen (*error surfaces*)
- Exakte Interpolationen sofern *nugget*-Effekt = 0

Geostatistische Methoden beruhen auf der Annahme, dass die Werteausprägung eines Merkmals zufällig im Raum variiert. Die nachfolgende Formel¹⁷ beschreibt diese Annahme.

$$Z(s) = \mu(s) + \varepsilon(s)$$

mit $Z(s)$ = Variable Z am Ort s
 $\mu(s)$ = Mittelwert μ am Ort s
 $\varepsilon(s)$ = Autokorrelierter Zufallsfehler ε am Ort s

Der Wert einer Variablen Z am Ort s wird durch einen Mittelwert, der als konstant über das Untersuchungsgebiet angenommen wird, und einen räumlich autokorrelierten Zufallsfehler bestimmt. Die angenommene rein zufällige Abweichung vom Mittelwert bedeutet, dass kein räumlicher Trend vorliegt. Da der Mittelwert tatsächlich nicht bekannt ist, kann durchaus ein Trend vorliegen. Je nach spezifischer Ausgangssituation können verschiedene Kriging-Verfahren, die sich vor allem im Umgang mit räumlichen Trends unterscheiden, für eine Problemlösung angewendet werden.

Nachfolgende Tabelle zeigt einen Auszug gängiger Kriging-Verfahren in den Umweltwissenschaften (Li und Heap, 2014).

Tab. 5 Univariate und multivariate Kriging-Verfahren (Auswahl)

Univariat	Multivariat
▪ Simple kriging (SK)	▪ Universal kriging (UK)
▪ Ordinary kriging (OK)	▪ SK with varying local mean (SKlm)
▪ Kriging with a trend (KT)	▪ Kriging with an external drift (KED)
▪ Factorial kriging (FK)	
▪ Indicator kriging (IK)	
▪ Disjunctive Kriging (DK)	

Zu den wichtigsten Verfahren gehören Ordinary kriging, Universal kriging, Simple kriging und Indicator kriging, die bei der Auswahl der für die Fragestellung dieser Arbeit geeigneten Verfahren eine wesentliche Rolle spielen (Kap. 2.7), und daher näher betrachtet werden. Die vier Verfahren sind in den meisten Geostatistik Tools - wie auch im hier verwendeten SGeMS - implementiert.

¹⁷ UNIGIS MSc, Modul 10: Geostatistik, Lektion 10: Probabilistische Interpolation, Seite 6

Ordinary kriging (OK):

Bezogen auf obige Formel geht dieses Verfahren von der Annahme aus, dass der Mittelwert unbekannt, aber über das gesamte Untersuchungsgebiet konstant ist ($\mu(s) = m$). OK erfordert daher kein Wissen über Stationarität und den Mittelwert im gesamten Untersuchungsgebiet. OK ist sensitiv gegenüber Trends, nicht aber gegenüber Anisotropie. OK kann dennoch bei Daten, bei denen ein Trend vermutet wird, für die Bestimmung der Schätzwerte genutzt werden, da ein Muster der Werteausprägung nie eindeutig durch die räumliche Autokorrelation bedingt ist, sondern auch aus einem Trend resultieren kann (Johnston et al., 2001). Klar detektierbare Trends müssen vor der Durchführung eines OK eliminiert werden.

Simple kriging (SK):

Simple kriging geht ebenfalls von der Annahme aus, dass der Mittelwert m über das Untersuchungsgebiet konstant ($\mu(s) = m$) und zudem bekannt ist. Damit ist auch der autokorrelierte Fehler ε am Ort s exakt bekannt, was in der Realität selten der Fall sein dürfte (Johnston et al., 2001). SK ist wie OK sensitiv gegenüber Trends und erfordert eine Normalverteilung der Daten. Falls die Ausgangsdaten nicht normalverteilt sind, was oftmals der Fall ist, sollte eine Transformation vorgenommen werden (Johnston et al., 2001).

Universal kriging (UK):

Beim Universal kriging ist $\mu(s)$ eine deterministische Funktion und wird somit als nicht konstant betrachtet. Der autokorrelierte Fehler $\varepsilon(s)$ wird ans zufällig angenommen. Der Mittelwert aller $\varepsilon(s)$ ist 0. UK ist letztlich eine Regressionsanalyse mit den räumlichen Koordinaten als erklärende Variablen (Johnston et al., 2001). UK ist nicht sensitiv gegenüber Trends. Wenn ein Trend bei der explorativen Analyse der Daten detektiert wird, sollte UK statt OK für die Schätzung verwendet werden. UK ist auch für die Interpolation transformierter Daten geeignet. In diesem Fall ist ein vorhandener Trend vorher jedoch zu eliminieren (Johnston et al., 2001). UK sollte das primär gewählte Verfahren sein, wenn ein Trendverlauf in den Daten extrapoliert werden muss (Li und Heap, 2014).

Indicator kriging (IK):

IK nimmt wie OK $\mu(s)$ als unbekannt und konstant über das gesamte Untersuchungsgebiet an. $Z(s)$ ist beim IK eine binäre Variable. Die Zuweisung des binären Wertes erfolgt entweder durch die Verwendung eines Grenzwertes (0 bei Unterschreitung und 1 bei Überschreitung) oder direkt im Fall, dass die beobachteten Daten binär vorliegen (Johnston et al., 2001). IK ist als in SGeMS implementiertes Verfahren der Vollständigkeit halber aufgeführt, findet jedoch bei der Auswahl des Interpolationsverfahrens keine Berücksichtigung, da die PM_{10} -Messdaten und β -Daten weder binär vorliegen, noch die Überschreitung eines Grenzwertes betrachtet wird.

Die Frage, ob ein räumlicher Trend in den Daten vorliegt, ist für die Wahl eines geeigneten Kriging-Verfahrens ein wichtiges Kriterium. Wenn Trends nicht eliminiert werden, ist dies bei der Interpretation der Ergebnisse trendsensitiver Verfahren wie beispielweise OK zu berücksichtigen. Kriging-Verfahren erfordern meist keine Normalverteilung der Daten. Datentransformation und Trendeliminierung helfen jedoch dabei, die Annahme von Normalverteilung und Stationarität als Grundidee des Krigings zu rechtfertigen (Johnston et al., 2001). Normalverteilte Daten sind bei einigen Verfahren auch für die Erstellung von Quantil-Plots und Wahrscheinlichkeitskarten notwendig.

Eine genaue Vorhersage der Schätzwerte ist selbst bei vielen, nicht geclusterten und über das Untersuchungsgebiet verteilten empirischen Werten nicht möglich. Die neben den Interpolationsflächen generierten Fehlerflächen (*error surfaces*) auf Grundlage der Kriging-Varianz quantifizieren die Unsicherheit des Schätzwertes für jeden Ort auf der Schätzoberfläche. In der Fehleroberfläche haben Lokationen in der Nähe von Messpunkten im Allgemeinen kleinere Fehler. Die Kriging-Varianz wird häufig fälschlicherweise als Maß für die Genauigkeit der Schätzwerte betrachtet. Sie ist jedoch unabhängig von den Datenwerten und hängt nur von der räumlichen Konfiguration der zu schätzenden Punkte ab (Remy et al., 2009). Für die Validierung der Interpolationsergebnisse müssen andere Methoden herangezogen werden (Kap. 2.5.6).

2.5.5 Kokriging

Beim Kokriging wird für die Berechnung der Schätzwerte eine zweite Variable hinzugezogen. Die Schätzung der Werte für die primäre Variable ist damit deutlich

aufwendiger, da neben der Autokorrelation für diese auch die Autokorrelation für die sekundäre Variable und Kreuzkorrelation zwischen beiden Variablen mit in die Berechnung einfließt (Johnston et al., 2001). Bei guter Korrelation der Variablen ist eine Verbesserung der Interpolationsergebnisse zu erwarten.

Für das Kokriging können grundsätzlich alle bereits näher betrachteten Kriging-Verfahren verwendet werden, die in einem sehr unterschiedlichen Umfang in Geostatistik-Tools implementiert sind. SGeMS bietet nur Simple Cokriging (SCK) und Ordinary Cokriging (OCK) an.

Die Werte der sekundären Variablen müssen nicht zwingend an den Messpunkten der primären Variablen vorliegen. Eine möglichst gleichmäßige Verteilung der sekundären Variablen im Untersuchungsgebiet kann bei guter Korrelation zwischen beiden Variablen die Vorhersagegenauigkeit für die erste Variable erhöhen. Eine um den Faktor drei bis fünf erhöhte Anzahl an Werten der sekundären Variablen gegenüber der primären Variablen stellt eine gute Relation für das Kokriging dar. Überwiegt die sekundäre Variable deutlich, wird beim Kokriging der Einfluss der primären Variablen durch die Sekundäre nahezu gänzlich überdeckt. Die Gewichtung der Variablen beim Kokriging kann auch über den Nugget der beiden Variablen gesteuert werden. Ein Herabsetzen des Nuggets erhöht das Gewicht der jeweiligen Variablen und Hochsetzen senkt den Einfluss.

2.5.6 Validierung

Die Prüfung der Güte des gewählten Modells durch die Analyse und Beurteilung der Schätzergebnisse ist der zentrale Punkt am Ende eines jeden Schätzverfahrens (PL und Srivastava, 1989). In der Geostatistik stehen hierfür verschiedene Möglichkeiten zur Verfügung.

Kreuzvalidierung (cross validation):

Bei der Kreuzvalidierung wird nacheinander jeder Messpunkt aus dem Interpolationsmodell entfernt und für diesen Punkt ein Schätzwert errechnet. Die Werte aus der empirischen Messung und der Schätzung werden für jeden Messpunkt auf der Grundlage des Semivariogramms miteinander verglichen. Der Differenzbetrag (Residuum) sollte dabei möglichst klein sein. Die Summe der quadrierten Residuen über alle Messpunkte ist ein Gütemaß für das Interpolationsmodell. Die graphische

Darstellung der Validierungsergebnisse erfolgt mit Hilfe eines standardisierten Histogramms. Die um den Nullpunkt liegenden geringsten Differenzen sollten die höchste Häufigkeit aufweisen. Die Über- und Unterschätzungen sollten normalverteilt sein. Eine andere Möglichkeit der graphischen Interpretation bietet der auf die Normalverteilung bezogene QQ-Plot - empirische versus geschätzte Werte. Im Idealfall liegen alle Punkte auf einer Diagonalen. Als dritte Möglichkeit gilt der Standardfehler. Dieser gibt an, welchen durchschnittlichen Fehler man machen würde, wenn man das Experiment wiederholen würde.

Validation (validation):

Bei der Validation wird vor der Erstellung des Semivariogramm-Modells die Ausgangsdatenmenge in eine Trainings- und eine Testdatenmenge unterteilt. Die Trainingsdatenmenge umfasst etwa 60-80% der Ausgangswerte. Mit dieser Menge wird das Modell generiert und mit der verbleibenden Testmenge geprüft. Diese Verfahren sind bei der Regressionsanalyse oder bei der Neuronale-Netze-Modellierung gebräuchlich. Analog zur Kreuzvalidierung geben die Differenzen zwischen den beobachteten Werten (Testwerten) und den geschätzten Werten (Modell- und Trainingswerten) eine Auskunft über die Güte des Modells.

Fehlerflächen (error surfaces):

Fehleroberflächen, die bei einer Interpolation neben den Schätzoberflächen generiert und bereits in Kapitel 2.5.4 diskutiert werden, zeigen die Bereiche geringer und hoher Schätzfehler auf. Diese können im Kontext mit den Interpolationsoberflächen Hinweise auf die Ursache von regional aufgetretenen Schätzfehlern geben. Eine zu geringe Punktdichte in einem Teil des Untersuchungsgebietes wäre ein klassisches Beispiel hierfür. Die Interpretation einer Fehleroberfläche hängt von dem Fehlerparameter (RMSE, Standard Error, Kriging-Varianz, etc.), auf den sich diese bezieht, ab.

Die Begriffe RMSE (Root Mean Square Error) und ASE (Average Standard Error), die bei der Beurteilung der Interpolationsergebnisse von Bedeutung sind, seien hier noch erläutert.

Der RMSE ist die Wurzel der Abweichquadrate (Messwert vs. Schätzung) und bezieht sich auf das konkrete Interpolationsmodell, welches aus dem Semivariogramm generiert wird.

Formel (Li und Heap, 2011):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

mit y_i = gemessener Wert
 \hat{y}_i = geschätzter Wert (Interpolation)
 n = Anzahl der Messwerte

Der ASE als gemittelter Standardfehler bezieht sich auf die Untersuchung bzw. die Untersuchungsreihe und damit auf eine Mehrzahl von Modellen. Dieser gibt an, welcher Fehler im Durchschnitt auftritt, wenn eine Untersuchung, also die Messung der Werte häufig wiederholt wird.

2.6 RIO-Modell

2.6.1 Entwicklung des ersten Modells für Ozon

Das RIO-Modell (RIO: residual interpolation optimised) ist ein trendbereinigtes, auf Kriging basiertes Interpolationsmodell, das für die Anwendung in Belgien vom flämischen VITO-Institut¹⁸ entwickelt wurde. Primäre Verwendung findet das Modell bei der landesweiten Beurteilung der Luftqualität in Echtzeit.

Die Grundidee des Modells ist die Bereinigung von lokalen räumlichen Trends im Untersuchungsgebiet vor der Durchführung der Interpolation (Hooyberghs et al., 2006). Dabei wird in einem ersten Schritt der lokale Charakter der Messwerte durch eine Trendbereinigung entfernt. Die so trendbereinigten Daten werden dann mittels Ordinary kriging interpoliert. Abschließend wird ein Retrending der Interpolationsergebnisse durchgeführt, um den lokalen Bias wieder hinzuzufügen (Janssen et al., 2008). Die Transformation der an den Messstationen gemessenen Werte in lageunabhängige Quantitäten führt zu einer räumlichen Homogenität. Die Homogenität der Messwerte ist eine essenzielle Voraussetzung für die korrekte Anwendung jedes Interpolations-

¹⁸ Vlaamse Instelling voor Technologisch Onderzoek (VITO), Boeretang 200, 2400 Mol, Belgien

verfahrens und wird bei der Interpolation von Messdaten von Luftschadstoffen oft missachtet (Janssen et al., 2008).

Ausgangspunkt für die Entwicklung des ersten RIO-Modells in den Jahren 2004 bis 2006 war die Notwendigkeit, auf Basis weniger Messstationen (2006: 38) die räumliche Verteilung der Ozonbelastung für das gesamte belgische Territorium zu bestimmen, um nationale und EU-Richtlinien zur Luftreinhaltung einhalten und die Gefährdung der Bevölkerung besser einschätzen zu können.

Troposphärisches Ozon ist ein sekundärer Luftschadstoff, der durch sogenannten Titrationseffekt in Bodennähe teilweise wieder eliminiert wird. Die Bevölkerungsdichte korreliert positiv mit dem Titrationseffekt, da anthropogen durch Industrie und Verkehr emittierte Stickoxide hierbei eine wesentliche Rolle spielen (Hooyberghs et al., 2006). Die Titration verursacht einen lokalen Trend, der auf der Basis der an jedem Ort bekannten Bevölkerungsdichte bereinigt werden kann. (Hooyberghs et al., 2006).

Die trendbereinigten und logarithmierten Werte, die am ehesten einer Normalverteilung ähneln, werden mittels Ordinary Kriging interpoliert. Die geringe Werteanzahl erschwert dabei die Bestimmung einer guten Korrelation zum Teil deutlich (Hooyberghs et al., 2006). Da neben den aktuellen auch historische Daten für jede Messstation zur Verfügung stehen, fließen die Mittelwerte der über einen bestimmten Zeitraum an einer Station gemessenen Werte in die Kalkulation der räumlichen Korrelation zwischen den Messstationen ein. Dieser Ansatz ist kennzeichnend für das RIO-Modell.

Für die Evaluierung der beschriebenen Methode werden die Daten zusätzlich nach dem IDW-Verfahren ohne und mit vorheriger Trendbereinigung (R_IDW) interpoliert. Der Vergleich basiert auf der „leaving-out-one“-Methode. Als Evaluationskriterium wird unter anderem der RMSE (Kap. 2.5.6), der auf der Differenz zwischen dem gemessenen und dem interpolierten Wert basiert, verglichen. Dabei zeigt sich, dass der RMSE-Wert von IDW über R_IDW nach RIO immer weiter abnimmt (Hooyberghs et al., 2006).

2.6.2 Corine Land Cover (CLC) Daten

Das in Kapitel 2.6.3 ausführlich diskutierte erweiterte RIO-Modell basiert auf der Einbeziehung von CORINE Bodenbedeckungs-/Landnutzungsdaten (CLC). Das Land-Cover Projekt ist Teil des CORINE-Programms der europäischen Kommission zur einheitlichen Sammlung, Koordination und Sicherstellung der Konsistenz von

Umweltinformationen innerhalb der EU. Es wurde mit dem Ziel spezifiziert, die Datenerfassung für Landbedeckung und Landnutzung in Europa zwecks der Entwicklung von EU-weiten Umweltschutzrichtlinien zu standardisieren. Die Ersterfassung der Daten begann 1985, um für das Referenzjahr 1990 (CLC1990) eine Datenbasis aufzubauen. Die vektorbasierten Daten und die hieraus resultierenden Karten werden für die meisten Länder durch visuelle Interpretation hochauflösender Satellitenbilder generiert. In einigen, wenigen Ländern werden halbautomatische Lösungen praktiziert, die zusätzlich z.B. nationale In-situ-Daten einbeziehen. Nach der Erfassung der ersten Datenbasis für das Bezugsjahr 1990 erfolgten bis dato vier Updates für die Bezugsjahre 2000, 2006, 2012 und 2018 in stetig kürzeren Produktionszeiträumen. Die Entwicklung der CLC-Daten von 1986 bis heute und deren Spezifikationen sind in Tab. 18 (Anhang C) aufgeführt.

Für das CLC-Projekt wird auf Basis der verwendeten Satellitendaten ein Kartenmaßstab von 1:100.000 gewählt, der eine für den Verwendungszweck ausreichend genaue Darstellung der Landbedeckung/-nutzung liefert und zu den gebräuchlichsten Maßstäben in den am Projekt beteiligten Ländern gehört. Die Mindestkartiereinheit (MMU) für Flächenobjekte beträgt 25 ha und die Mindestkartierbreite für lineare Objekte 100 m. Kleinere Flächen werden der nächstpassenden Klasse zugeordnet.

Seit der Ersterfassung für das Bezugsjahr 1990 wird die Landbedeckung/-nutzung auf unterster Ebene (Level 3) in 44 Klassen eingeteilt, von denen 37 Klassen für Deutschland bedeutsam sind. Die 44 Level 3-Klassen werden auf mittlerer Ebene (Level 2) zu 15 Klassen und letztere auf oberster Ebene (Level 1) zu 5 Hauptklassen zusammengefasst (Tab. 19). Jeder Klasse ist ein eindeutiger CLC-Code zugeordnet.

Die CLC-Daten sind für alle Bezugsjahre frei verfügbar und können auf der Copernicus-Webseite als Vektor- oder Rasterdatei heruntergeladen werden. In dieser Arbeit werden die aktuellen CLC-Daten (CLC2018) als Vektordaten im ESRI Geodatabase Format verwendet¹⁹.

¹⁹ Download: <https://land.copernicus.eu/pan-european/corine-land-cover/clc2018?tab=download>

2.6.3 Erweitertes RIO-Modell und Einbeziehung der Landnutzung

Eine Weiterentwicklung des von Hooyberghs et al. (2006) entwickelten RIO-Modells für Ozon stellt das von Janssen et al. (2008) überarbeitete RIO-Modell dar, welches um die Luftschadstoffe Stickstoffdioxid (NO₂) und Feinstaub (PM₁₀) erweitert wurde. Die Grundidee des RIO-Modells der Trendbereinigung und damit der Entfernung des lokalen Charakters der Messwerte vor der Interpolation und des Retrendings nach der Interpolation bleibt bestehen. Der wesentliche Unterschied zum ersten RIO-Modell liegt in der Einbeziehung der in Kapitel 2.6.2 beschriebenen CLC-Landnutzungsdaten zur Trendbereinigung. Das erweiterte Modell kann hierdurch für alle drei Luftschadstoffe angewendet werden. Die Bevölkerungsdichte als Parameter für die Trendbereinigung findet in diesem Modell keine Berücksichtigung mehr, da der in Kapitel 2.6.1 beschriebene Titrationseffekt nur für Ozon relevant ist.

Die Einbindung der Landnutzung erfolgt durch die Definition des sogenannten Landnutzungsindikators β für jeden Luftschadstoff an jeder Messstation. Der Landnutzungsindikator β drückt als Einzelparameter die Korrelation zwischen den lokalen Landnutzungscharakteristika und dem lokalen Luftverschmutzungsniveau aus (Formel: Anhang D). Dafür wird zuerst der Flächenanteil der Landnutzungsclassen in einem Radius von 2 km um die Messstation ermittelt. Zur Vereinfachung des Modells werden die 44 CLC-Klassen zu 11 RIO-Landnutzungsclassen (RCL) zusammengefasst (Tab. 6). Die CLC-Klassen 111, 121, 122, 123 und 124, bei denen eine hohe Schadstoffemission zu erwarten ist, werden unverändert in das RIO-Modell übernommen.

Tab. 6 Definition RIO-Klassen (RCL), Janssen et al. (2008)

RIO class	Description	CLC classes
RCL1	Continuous urban fabric	111
RCL2	Discontinuous urban fabric, green and sport	112, 141, 142
RCL3	Industrial or commercial units	121
RCL4	Road and rail networks and associated land	122
RCL5	Port areas	123
RCL6	Airports	124
RCL7	Mine, dump and construction sites	131, 132, 133
RCL8	Arable land	211, 212, 213
RCL9	Agricultural areas	221, 222, 223, 231, 241, 242, 243, 244
RCL10	Forest and semi natural areas	311, 312, 313, 321, 322, 323, 324, 331, 332, 333, 334, 335
RCL11	Wetlands and water bodies	411, 412, 421, 422, 423, 511, 512, 521, 522, 523

Wenn für jede Messstation ein charakterisierender β -Wert vorliegt, kann eine Relation zwischen der Landnutzung und Luftverschmutzungseigenschaften formuliert werden. Dafür werden die Langzeit-Mittelwerte der Luftschadstoffe an jeder Messstation gegen die korrespondierenden β -Parameter der Messstationen aufgezeichnet. Dieser Plot repräsentiert die Trendfunktion und wird für jeden Luftschadstoff einzeln erstellt.

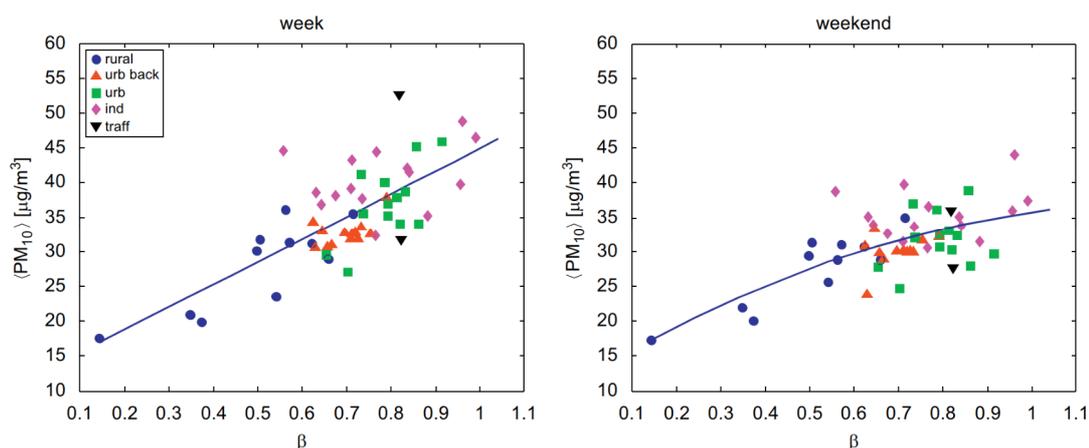


Abb. 13 Trendfunktionen PM10 (Wochentage/Wochenende), (Janssen et al., 2008)

PM₁₀ zeigt die schwächste Relation zwischen Landnutzung und Luftqualität von allen drei untersuchten Luftschadstoffen (Janssen et al., 2008). Almeida et al. (2005) konnten zudem zeigen, dass natürliche Quellen signifikant das PM_{10-2,5}-Niveau beeinflussen, während anthropogene Quellen primär einen signifikanten Einfluss auf Ultrafeinstaub (PM_{2,5}) haben.

Der Landnutzungsindikator β kann nicht nur für jede Messstation sondern für jeden Ort im Untersuchungsgebiet bestimmt und in Form eines regelmäßigen Rasters dargestellt werden. Um eine solche Rasterkarte auf Basis der CLC-Karte zu generieren, wird für jede einzelne Rasterzelle die Verteilung der RIO-Klassen im definierten Puffer von 2 km um die Rasterzelle betrachtet (Abb. 68). Zusätzlich zu den aktuellen Schadstoffkonzentrationen liegen auch historisierte Messdaten für die Stationen vor. Bei der Analyse dieser Zeitreihendaten sind ähnliche Trends für die Standardabweichung der NO₂ und PM₁₀-Messwerte beobachtbar. Die Grundidee des RIO-Modells ist, beide Relationen - Mittelwert der aktuellen Daten und die Standardabweichung der Zeitreihendaten - zur Trendbereinigung der an den Messstationen gemessenen Werte zu verwenden. Die RIO-Interpolationsroutine basiert auf Ordinary kriging (OK), jedoch weicht die verwendete Kriging-Methode in einem wesentlichen Punkt vom Standard Ordinary kriging ab. Beim Standard-OK bezieht sich das Variogramm bzw. die räumliche

Korrelationsfunktion ausschließlich auf die Eingabewerte. Bei der Interpolation von Luftschadstoffen besteht jedoch häufig das Problem, dass nur wenige Messstationen vorhanden sind. Die niedrige Anzahl kann die statistische Signifikanz des Variogramms bzw. der Korrelationsfunktion beeinträchtigen (Janssen et al., 2008). Um dieses Problem zu lösen und die statistische Signifikanz zu verbessern, wird die Korrelationsfunktion im RIO-Modell auf Basis des gesamten historischen Datensatzes, der die Messdaten mehrerer Jahre beinhaltet, konstruiert.

Nach der Interpolation mit den trendbereinigten Messdaten wird ein Retrending vorgenommen, um für jeden räumlich-interpolierten Punkt die lokalen Landnutzungscharakteristika wieder einfließen zu lassen. Das Retrending erfolgt durch die Verwendung des spezifischen β -Wertes, der für jede Zelle im Interpolationsraster bestimmt wird.

Die Validierung der RIO-Methode erfolgt mit einer Kreuzvalidierungsanalyse nach dem „leaving-one-out“-Prinzip für jede Messstation. Die Qualität des Modells wird anhand des RMSE-Wertes, des Biases und des MAE überprüft. Zwecks weiterer Evaluierung der Gesamtleistung des Modells werden die RIO-Ergebnisse mit Fourth Power IDW und Standard-OK - letzteres ohne Trendbereinigung und Retrending - verglichen. Janssen et al. (2008) konnten mit Ausnahme einiger, weniger Messstationen belegen, dass das RIO-Modell eine bessere Performanz als die Standard-Interpolationsmethoden aufweist. Neben der statistischen Auswertung evaluiert man die Performanz des RIO-Modells auch mit Hilfe von Luftqualitätskarten (Abb. 70), die mit unterschiedlichen Interpolationstechniken (RIO/OK) generiert wurden.

Sowohl beim Kriging als auch bei der Anwendung der Trendfunktion kommt es zu Fehlern, die es zu spezifizieren gilt. Beim Ordinary kriging erhält man neben dem Interpolationsergebnis einen Wert für die Error-Varianz σ^2_R , die ein Maß für die Unsicherheit ($1 \sigma_R$) des Ergebnisses darstellt. Ein weiterer Fehler tritt bei der Anwendung der Trendfunktion während des Retrending-Prozesses auf.

2.6.4 Verwendung des RIO-Modells im Kontext der Fragestellung

Die Auswahl des RIO-Modells als potentielle Vergleichsmethode zum Kriging erfolgte anhand der Ergebnisse einer Machbarkeitsstudie zum Thema „Flächenhafte Darstellung von Luftschadstoffen“ des Österreichischen Umweltbundesamtes, die vergleichend

verschiedene Verfahren zur Ermittlung der flächenhaften Immissionsbelastung analysiert (Nagl und Spangl, 2007).

Ob und in welchen Umfang das RIO-Modell als Vergleichsmethode geeignet ist, hängt von verschiedenen Faktoren, wie beispielsweise der Eignung der Datenbasis, der verwendeten Software oder auch der vorhandenen Programmierkenntnisse ab. Tab. 7 führt Argumente Pro und Contra einer Einbeziehung des RIO-Modells unter Berücksichtigung der gegebenen Rahmenbedingungen auf.

Tab. 7 Verwendung RIO-Modell (Pro/Contra)

Pro	Contra
<ul style="list-style-type: none"> ▪ Basiert auf frei verfügbaren Landnutzungsdaten (CLC), die im Vektor- und Rasterformat zur Verfügung gestellt werden 	<ul style="list-style-type: none"> ▪ Keine Langzeitmesswerte (Zeitreihendaten) vorhanden. Korrelationsfunktion kann nicht auf historischen Messdaten konstruiert werden
<ul style="list-style-type: none"> ▪ CLC-Daten sind in den meisten GI-Tools als Vektor- oder Rasterdaten ohne Aufbereitung und damit leicht zu implementieren 	<ul style="list-style-type: none"> ▪ Beeinträchtigung der statistischen Signifikanz der Korrelationsfunktion bei wenigen Messpunkten möglich, da nicht auf historische Daten zurückgegriffen werden kann
<ul style="list-style-type: none"> ▪ Topographische Struktur des Untersuchungsgebietes ähnelt sehr der Belgiens (benachbarte, teils angrenzende Regionen) 	<ul style="list-style-type: none"> ▪ PM₁₀ zeigt im RIO-Modell von allen Luftschadstoffen die schwächste Relation zwischen Luftschadstoff und Landnutzung
<ul style="list-style-type: none"> ▪ Landnutzungsindikator β ist eine metrische Variable, die grundsätzlich als sekundäre Variable für ein Kokriging in Frage kommt 	<ul style="list-style-type: none"> ▪ De- und Re-Trending der Messwerte ist aufwendiger. Steht als Funktion in GI-Tools nicht zur Verfügung und muss programmiert werden (Im RIO-Modell mit Python gelöst)
<ul style="list-style-type: none"> ▪ Berechnung von β ist einfach 	
<ul style="list-style-type: none"> ▪ Quantitativer Emittentenbezug durch Schadstoffemissionskoeffizient a_i als Bestandteil von β 	

Für die Untersuchung liegen die validierten PM₁₀-Tagesmittelwerte von 16 Messstationen an 8 aufeinander folgenden Tagen vor. Insgesamt stehen 125 von 128 möglichen Messwerten zur Verfügung (Eine Messstation lieferte an zwei Tagen und eine weitere an einem Tag keine Messwerte). Langzeit-Mittelwerte für PM₁₀ sind nicht vorhanden.

Die Relation von PM_{10} -Messwerten und Landnutzungsindikator kann hierdurch nicht analog zum RIO-Modell formuliert werden. Es ist nur ein Bezug auf die Messwerte im achttägigen Untersuchungszeitraum möglich. Diese sind aufgrund der jahreszeitlichen Schwankungen von PM_{10} -Messwerten bezogen auf die im RIO-Modell verwendeten Jahresmittelwerte möglicherweise nicht charakteristisch, da sie nur einen sehr kurzen Zeitraum abbilden. Zudem basiert aus Gründen der statistischen Signifikanz die Trendfunktion im RIO-Modell auf dem Mittelwert der aktuellen Daten und der Standardabweichung der Zeitreihendaten. Ohne die Zeitreihendaten und bei einer so geringen Gesamtzahl Messstationen ist die Bildung der Trendfunktion und damit die Anwendung der Trendeliminierung analog zum RIO-Modell nicht sinnvoll und wird hier nicht weiter berücksichtigt.

So wird zwar die Grundidee des RIO-Modells der Bereinigung eines räumlichen Trends durch den Bezug auf die lokale Landnutzung in Form des Indikators β in Kombination mit Langzeit-Messdaten nicht übernommen, dennoch bietet β als metrischer Wert die Möglichkeit einer Verwendung als sekundäre Variable in einem Kokriging-Verfahren.

Ein wesentlicher Vorteil des Landnutzungsindikators ist, dass dieser neben den Messpunkten für jeden beliebigen Punkt im Untersuchungsgebiet berechnet werden kann. Anzahl und Lage zusätzlicher Punkte sind somit frei definierbar und können für die Verwendung von β als sekundäre Variable in einem Kokriging-Verfahren optimal angepasst werden. Optimal bedeutet eine gleichmäßige Verteilung der zusätzlichen Punkte im Untersuchungsgebiet (Vermeidung von Clustern) und eine sinnvolle Bestimmung der Relation zwischen primärer und sekundärer Variablen, um den Einfluss der sekundären Variablen beim Kokriging nicht unter- oder überzubewerten.

Vor der Verwendung von β als sekundäre Variable ist zu prüfen, in welchem Maß β an den 16 Messstationen mit den PM_{10} -Messwerten korreliert. Da die Korrelationsanalyse mit Gauß-transformierten Daten durchgeführt wird, sind die Schritte zur Einbindung von β im Rahmen der Darstellung der konkreten Methodik in Kapitel 2.7 beschrieben.

2.7 Anwendungsfall

2.7.1 Software und Tools

Kriging und Kokriging sollten anfangs mit dem Geostatistical Analyst Tool in ArcGIS Desktop 10.7 (ArcMap) als ausschließliche Software für die Schätzverfahren durchgeführt werden. Für die Interpolation von 3D bzw. 2D+t Daten kann der Geostatistical Analyst jedoch nicht verwendet werden, da dieser nur 2D Daten verarbeiten kann. Die Interpolation dieser Daten ist nur mit 3D Analyst möglich, der bei den Kriging-Verfahren auf OK und UK beschränkt ist und kein Kokriging anbietet. Durch die eingeschränkten Möglichkeiten in ArcGIS wurde die Suche nach leistungsfähigeren Programmen mit einem größeren Funktionsumfang im Bereich Variographie und Interpolation erforderlich. Die Wahl fiel letztlich auf SGeMS, da die meisten SGeMS-Tools auf den klassischen Geostatistik-Verfahren für Kriging und Simulation basieren und viele Möglichkeiten der Parametrisierung bieten. SGeMS ist darüber hinaus frei verfügbar, wird seit 2011 jedoch nicht mehr weiter entwickelt. Der Fokus richtet sich bei SGeMS auf Fragestellungen der Geologie. Dennoch eignen sich die implementierten Verfahren durchaus für die Verarbeitung von Luftschadstoffdaten. Für den Datenimport in SGeMS müssen diese im GSLIB-Format vorliegen.

Zur explorativen Analyse der Messdaten und des Landnutzungsindikatoren werden neben ArcGIS Desktop 10.7 auch Grapher 16 und Voxler 4 verwendet. Letztere Tools sind nicht frei verfügbar, können aber als Trial-Version oder mit einer Studentenlizenz genutzt werden. Der Einsatz von Grapher und Voxler ist für die Analyse der Daten nicht notwendig, da die Daten auch mit Geostatistical Analyst zufriedenstellend analysiert werden können. Gerade Voxler bietet aber sehr gute Möglichkeiten der Visualisierung, die auch im Hinblick auf Umfang und Funktionalität deutlich über den Geostatistical Analyst hinausgehen und dabei helfen können, die Ausgangsdaten besser zu verstehen.

Tab. 8 Eingesetzte GIS-Software und Verwendung

Software/Tool	Verwendung
ArcGIS ArcMap 10.6/10.7 inklusive Extension Geostatistical Analyst	<ul style="list-style-type: none"> ▪ ESDA (Normal QQ-Plots, Trendanalyse) ▪ Berechnung Landnutzungsindikator für Messstationen und zusätzliche Punkte ▪ Erstellung Hintergrundkarten zur Verwendung in Voxler ▪ Erstellung Karten für Master Thesis
Grapher 16	<ul style="list-style-type: none"> ▪ ESDA (Histogramme) ▪ Erstellung Kurvendiagramm
Voxler 4	<ul style="list-style-type: none"> ▪ Datenvisualisierung (Scatterplot, Volumenrendering) ▪ IDW-Interpolation als Basis für das De- und Retrending der PM₁₀-Messwerte
SGeMS-beta-x64	<ul style="list-style-type: none"> ▪ ESDA (Histogramme 2D + t, QQ-Plots) ▪ Regressionsanalyse ▪ Datentransformation ▪ Variographie ▪ Interpolation (SK und SCK) ▪ Validierung (QQ-Plots)

2.7.2 2D+t Modell

Es war ursprünglich vorgesehen, die flächenhafte Feinstaubbelastung nur für zwei anhand der Ergebnisse der deskriptiven Statistik und der EDSA ausgewählten Untersuchungstage zu schätzen. Die Dichte des PM₁₀-Messnetzes ist jedoch mit 16 Stationen im 5920 km² umfassenden Untersuchungsgebiet sehr gering (0,27 Messstationen pro 100 km²). Für drei von acht Untersuchungstagen liegen aufgrund fehlender Daten nur 15 Messwerte vor. Die geringe Zahl kann, wie in Kapitel 2.6.3 bereits diskutiert wurde, die statistische Signifikanz des Variogramms beeinflussen.

Um die Datenbasis für die Interpolation zu erhöhen, fließen daher die PM₁₀-Messwerte aus dem gesamten Untersuchungszeitraum in Schätzung ein (125 statt 15 bzw. 16 Messwerte). Neben der räumlichen Verteilung der Feinstaubbelastung wird damit auch die zeitliche Veränderung über den Untersuchungszeitraum in einer dreidimensionalen Darstellung (2D+t) betrachtet.

Die Zeit wird als z-Koordinate metrisch skaliert. Das für diesen Zweck in den Naturwissenschaften häufig verwendete Julianische Datum findet hier jedoch keine Berücksichtigung, da die Zahlenwerte hoch sind. In Kombination mit den 6- bzw. 7-stelligen x-/y-Koordinaten (UTM) erfordert dies bei rasterbezogenen Berechnungen einen höheren Rechenaufwand in SGeMS, der möglichst reduziert werden sollte.

Des Weiteren führt man eine Skalentransformation mit dem Faktor 10.000 durch, um eine passende Relation zu den UTM-Koordinaten (Einheit: Meter) herzustellen und den zeitlichen Verlauf so sichtbar zu machen. Ohne Transformation wäre die Dreidimensionalität des Rasters nicht erkennbar. Analog zu den x-/y-Koordinaten umfasst der Extrapolationsbereich 2000 Koordinateneinheiten (Abstand: Äußere Punkt - Rastergrenze).

Tab. 9 Zuordnung z-Koordinate - Untersuchungstag

z-Koordinate	Untersuchungstag	Datum
2000	1	01.09.2018
12000	2	02.09.2018
22000	3	03.09.2018
32000	4	04.09.2018
42000	5	05.09.2018
52000	6	06.09.2018
62000	7	07.09.2018
72000	8	08.09.2018

2.7.3 Implementierung RIO-Landnutzungsindikator β

Der Landnutzungsindikator β ist ortsgebunden und bleibt über die Zeit (Untersuchungszeitraum) konstant. β ist somit eine zweidimensionale Variable. Die Berechnung des Landnutzungsindikators erfolgt in ArcMap, da die CLC-Landnutzungsdaten als zweidimensionaler Vektordatensatz vorliegen und ArcMap über die benötigten Geoverarbeitungstools verfügt, die unter Verwendung des arcpy-Moduls leicht in Python implementiert werden können.

Die Bestimmung von β ist für jeden beliebigen Punkt im Untersuchungsgebiet möglich, sodass neben den Messstationen weitere Punkte für die Wahl einer passenden Relation zwischen primärer und sekundärer Variablen beim Kokriging hinzugefügt werden

können. Eine zudem gleichmäßige Verteilung zusätzlicher β -Punkte führt zu einer Abdeckung des gesamten Untersuchungsgebietes bei der sekundären Kokriging-Variablen ohne eine zu starke Gewichtung einzelner Bereich etwa durch Punktecluster. Wenn Messstationen geclustert sind, werden oftmals sehr ähnliche Werte an diesen gemessen. Bei einer guten Interpolation sollte der Einfluss geclusteter (Mess-)Punkte auf den zu bestimmenden Schätzwert verringert werden (Hooyberghs et al., 2006). Eine gleichmäßige, rasterförmige Verteilung der β -Punkte ist in Bezug auf die geringe Zahl PM_{10} -Messstationen und deren ungleichmäßiger Verteilung im Untersuchungsgebiet beim Kokriging daher sinnvoll. Das Verhältnis zwischen PM_{10} -Messpunkten und β -Punkten beträgt ca. 1:8,5 (16:137). Die Berechnung des Landnutzungsindikators für die Messstationen wird in ArcMap manuell durchgeführt, während die Berechnung für die zusätzlichen Punkte mit einem Python-Skript (Anhang E) erfolgt.

Berechnung des Landnutzungsindikators β - Messstationen:

- Import der vom LANUV bereitgestellten Messdaten (validierte Tagesmittelwerte) in ArcMap (Layer: *Messstationen_Untersuchungsgebiet*)
- Import der CLC2018-Vektordaten in ArcMap (Layer: *Untersuchungsgebiet_CLC2018_RCL*)
- In der Attributtabelle des Layers *Untersuchungsgebiet_CLC2018_RCL* (ArcMap) Feld für RCL-Code hinzufügen und jeder CLC-Klasse den entsprechenden RCL-Code über die Feldberechnung zuweisen (VB-Code: siehe Anhang E)
- In der Attributtabelle des Layers *Untersuchungsgebiet_CLC2018_RCL* Feld für Verschmutzungskoeffizienten a hinzufügen und jedem RCL-Code den entsprechenden a -Wert über die Feldberechnung zuweisen (VB-Code: siehe Anhang E)
- Pufferzone (2000m) um jeden Punkt einzeln generieren (Buffer)
- Jede einzelne Pufferzone aus dem Layer *Untersuchungsgebiet_CLC2018_RCL* ausschneiden (Clip)
- Polygone mit dem gleichen RCL-Code für jeden Puffer zusammenfügen (Dissolve)

Puffer_RCL_RODE_DIS						
FID*	Shape *	RCL	Shape Length	Shape Area	RCL	Coeff a
1	Polygon ZM	1	5103,25528	435998,271585	1	1,16
2	Polygon ZM	2	34819,986807	7500029,778975	2	1
3	Polygon ZM	3	4552,243368	539193,69541	3	2,07
4	Polygon ZM	8	1593,573926	97656,916952	8	0,64
5	Polygon ZM	9	4548,779608	441109,071021	9	0,64
6	Polygon ZM	10	23228,703039	2455187,196258	10	0
7	Polygon ZM	11	6629,792517	1097186,450513	11	0

Abb. 14 Attributtabelle Puffer_RCL_RODE_DIS (β -Berechnung)

- Landnutzungsindikator β für jeden Puffer berechnen (Math. Formel: Anhang D)
- Ermittelte β -Werte als neues Feld zum Layers „Messstationen_Untersuchungsgebiet“ hinzufügen
- Attributtabelle des Layers „Messstationen_Untersuchungsgebiet“ zur weiteren Verarbeitung in SGeMS und Voxler als csv-Datei exportieren

Berechnung des Landnutzungsindikators β - zusätzlich generierte Punkte:

Python-Skript *AdditionalPointsLNI.py* (Anhang E) - Verarbeitungsschritte:

- Dichte des Punkterasters festlegen (Anzahl Punkte auf x- und y-Achse); hartkodiert 11*11 Punkte. Modifizierung der Punktezahl ist durch Anpassen im Skript möglich.
- UTM-Koordinaten der Punkte berechnen und in csv-Datei speichern
- Shapefile aus csv-Datei generieren
- Features aus Shapefile in GDB-Feature-Class importieren (ArcMap Desktop)
- Pufferzone (2000m) um jeden Punkt generieren
- RCL-Features im gesamten Untersuchungsgebiet zusammenfügen
- Untersuchungsgebiet mit zusammengeführten RCL-Features auf Basis der gebildeten Puffer splitten
- Landnutzungsindikator β für jeden Punkt berechnen (Math. Formel: Anhang D)
Inhalt der bereits generierten csv-Datei einlesen und ergänzt um die Z-Werte (Faktor 10000 und 12000) und den berechneten β in neue csv-Datei schreiben

Die csv-Dateien aus beiden Berechnungen werden abschließend zusammengefügt und direkt oder angepasst als GSLIB-Datei in SGeMS importiert.

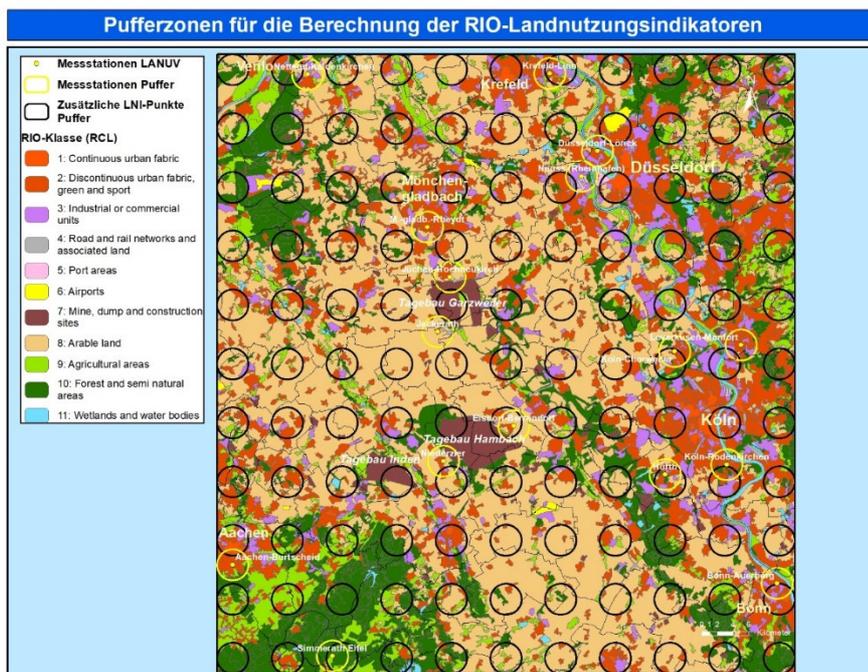


Abb. 15 Pufferzonen Landnutzungsindikator

Abb. 15 zeigt die 2000m-Pufferzonen um die Messpunkte (gelb) und die rasterförmig angeordneten zusätzlichen Punkte (schwarz), deren enthaltene RIO-Klassen über den zugehörigen Verschmutzungskoeffizienten a_i in die Berechnung von β des jeweiligen Punktes einfließen (Math. Formel: Anhang D).

2.7.4 Datenaufbereitung

Datentransformation:

Das Ziel einer Transformation ist, die Varianzen im Untersuchungsgebiet konstanter zu machen und die Daten einer Normalverteilung anzunähern. Daten, die eine deutlich schiefe Verteilung aufweisen, führen bei der Variogramm-Kalkulation zu Problemen, da die Extremwerte einen signifikanten Einfluss auf die Bildung des Variogramms haben. Darüber hinaus erfordern viele geostatistische Verfahren, dass die Daten in einer Normalverteilung vorliegen. Die Transformation der Daten von einer in eine andere Verteilung wird durch eine Anpassung der Quantile einer Verteilung auf einfachem Weg erreicht.

Werden die Ausgangsdaten vor der Durchführung eines geostatistischen Verfahrens transformiert, muss nach Abschluss des Verfahrens eine Rücktransformation erfolgen. Liegt nach einer Trendbereinigung bereits transformierter Daten abermals eine deutlich schiefe

Verteilung vor, sollten die trendbereinigten Daten vor der Durchführung der Interpolation erneut transformiert werden.

Sowohl bei den PM₁₀-Messwerten als auch bei den β -Werten erfolgt eine Gaußsche Histogramm-Transformation, um die bestmögliche Annäherung an eine Gaußsche Normalverteilung zu erzielen. Nach der Trendbereinigung der bereits transformierten PM₁₀-Messwerte wird eine zweite Gaußsche Histogramm-Transformation durchgeführt. Die Gaußsche-Normalverteilung ist die beste Verteilung für Interpolationen mit einer linearen Schätzfunktion (Hooyberghs et al., 2006). Die Datentransformation eliminiert keine Trends. Diese bleiben auch nach der Transformation bestehen und müssen gesondert betrachtet werden.

Trendbereinigung:

Trendsensitive Interpolationsverfahren wie das gewählte Simple Kriging (Kapitel 2.7.5) erfordern eine Trendbereinigung im Vorfeld der Interpolation, um die Annahme des Vorliegens von Stationarität zu rechtfertigen (Johnston et al., 2001). SGeMS beinhaltet kein Tool zum Detrending von Daten. Die Trendbereinigung muss daher gesondert vorgenommen werden.

Als einfaches Verfahren zur Trendbereinigung kann ein konstanter Wert von den Daten abgezogen werden. Dieser muss nach der Interpolation für jeden Schätzwert (Rasterzelle) wieder addiert werden. Die Werte für das De- und Retrending liegen bei einer Interpolation von 2D+t Daten in Form eines Trendrasters vor. Bei der Ermittlung der Werte sollten die Trendverläufe in den Daten berücksichtigt werden, um eine möglichst gute Glättung für die trendbereinigten Daten zu erzielen.

Die Werte für das De- und Retrending der Gauß-transformierten PM₁₀-Daten werden durch eine IDW Power1 Interpolation in Voxler unter Berücksichtigung aller Daten ermittelt. Ein Power-Wert von 1 bedeutet eine nur geringe Senkung der Gewichtung benachbarter Messpunkte mit zunehmender Distanz. Durch die Berücksichtigung aller Messwerte und einer in Relation hohen Gewichtung entfernt liegender Punkte wird ein geglättetes Interpolationsraster generiert, das abgeschwächte Trends der Ausgangsdaten noch erhält und somit gut für die Trendbereinigung geeignet ist.

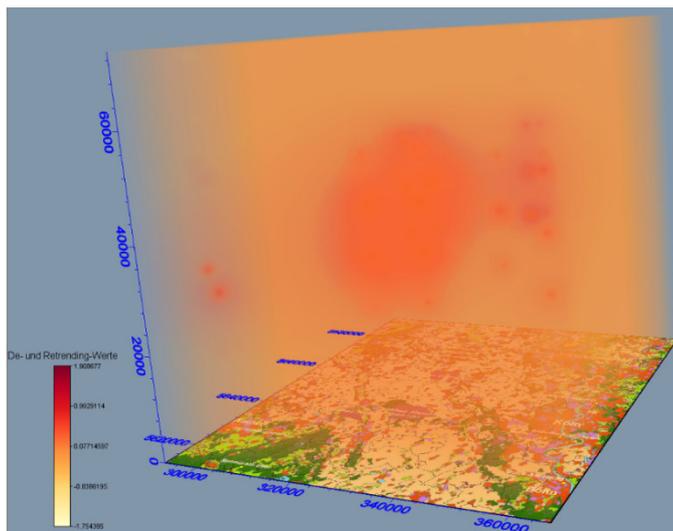


Abb. 16 zeigt das Ergebnis der IDW Power 1 Interpolation. Die Schätzwerte werden für das De- und Retrending genutzt. Die Trends sind noch gut zu erkennen.

Abb. 16 De-/Retrending-Interpolationsraster

Voraussetzung für ein korrektes Retrending sind identische Raster in SGeMS, wo das Kriging erfolgt, und Voxler, sodass jeder Zelle im SGeMS-Raster die passende Zelle im Voxler-Raster zugeordnet werden kann. Dabei sind die unterschiedlichen Bezugspunkte der Rasterzell-Koordinaten in SGeMS (Zellmitte) und Voxler (linke untere Ecke) zu berücksichtigen.

Auch beim Detrending der Gauß-transformierten PM_{10} -Werte (SGeMS), die als Punktemenge und nicht als Raster vorliegen, erfolgt eine Zuordnung der SGeMS Punkt-Koordinaten zu passenden Voxler Rasterzell-Koordinaten. Hierfür nutzt man das Ergebnis einer IDW Power 0 Interpolation der PM_{10} -Daten in Voxler mit einem kleinen Suchradius. Im Interpolationsraster weisen nur sehr wenige - optimalerweise eine - Zellen in der Nähe der Messpunkte Werte ungleich 0 auf. Diese selektiert das Python-Skript *PM10DetrendDataPreparation.py* (Anhang E). Die Zuordnung erfolgt anschließend in Excel.

2.7.5 Auswahl des Kriging-Verfahrens

Die Wahl der Geostatistik Modellierungssoftware SGeMS für Kriging und Kokriging der dreidimensionalen Daten ($2D + t$) schränkt die Auswahl eines geeigneten Interpolationsverfahrens bereits auf die implementierten Verfahren ein (Kap. 2.7.1). Für das Kriging stehen in SGeMS mit Simple kriging, Ordinary kriging, Kriging with a trend (UK) und

Indicator kriging und für das Kokriging mit Simple Cokriging und Ordinary Cokriging allerdings die mit am häufigsten verwendeten Verfahren zur Verfügung.

Da beim Indicator Kriging $Z(s)$ eine binäre Variable ist (Kap. 2.5.4), scheidet IK von vornherein aus. Für die weitere Eingrenzung ist die Betrachtung der Ergebnisse der explorativen a-räumlichen und räumlichen Datenanalyse (Kap. 3.2) und der Variographie (Kap. 3.4.1) wichtig. Diese zeigen, dass sowohl eine Histogramm-Transformationen aufgrund von Ausreißern in den Verteilungen der PM_{10} - und β -Werte als auch eine Trendbereinigung für die PM_{10} -Werte erforderlich sind (Kap. 3.2.5).

Kriging with a Trend ist nicht trendsensitiv und könnte für die transformierten, noch nicht trendbereinigten Daten verwendet werden. In SGeMS sind beim KT jedoch nur vereinfachte, richtungsbezogene Angaben zu vorliegenden Trends (linear, quadratisch, etc.) für die Trendeliminierung möglich. Die in Kapitel 2.7.4 beschriebene Trendbereinigungsmethode unter Verwendung des Ergebnisraters einer IDW-Interpolation bietet dagegen eine für den PM_{10} -Datensatz zugeschnittene Möglichkeit der Trendeliminierung, die im Vergleich mit KT zu besseren Ergebnissen führen sollte. Durch die Verwendung der gesonderten Trendbereinigung findet KT keine weitere Berücksichtigung.

Das trendsensitive Ordinary Kriging war ursprünglich die nach einer Trendbereinigung für das Kriging und Kokriging favorisierte Methode. Durch die notwendig gewordene Gauß-Transformation der Ausgangsdaten scheidet OK jedoch aus, da beim OK eine weitere hier nicht gewünschte Glättung der Daten vorgenommen wird.

Simple Kriging setzt voraus, dass der Mittelwert m über das Untersuchungsgebiet konstant und bekannt ist (Kap. 2.5.4). Letzteres ist zwar nicht der Fall, jedoch kann man nach Trendbereinigung und Transformation m als konstant über das Untersuchungsgebiet annehmen. Folglich können die Mittelwerte der transformierten und trendbereinigten PM_{10} - und β -Daten für Simple Kriging und Simple CoKriging verwendet werden. Bei der abschließenden Betrachtung des Kokrigings scheidet OK aus dem oben bereits genannten Grund aus. Nach Abwägung aller Kriterien wurden Simple Kriging und Simple Cokriging hinsichtlich der vorliegenden Daten als die am besten geeigneten Interpolationsverfahren gewählt.

2.7.6 Validierung der Interpolationsergebnisse

Um die Qualität der Interpolationsergebnisse von Kriging und Kokriging vergleichen zu können, muss abschließend eine Validierung durchgeführt werden. SGeMS ist ein leistungsfähiges Programm für EDSA, Variographie, Kriging und Simulation, umfasst jedoch keine gesonderten Tools bzw. Funktionen für die Validierung der Ergebnisse. Validierungsmethoden wie etwa die Kreuzvalidierung können über Python-Skripte eingebunden werden.

Eine Möglichkeit der Validierung bieten aber die in SGeMS für die EDSA vorgesehenen allgemeinen QQ-Plots, die die Quantile einer numerischen Variablen den Quantilen einer zweiten numerischen Variablen gegenüberstellen und so die Verteilungen der beiden Variablen miteinander vergleichen. Wenn die Verteilungen gleich sind, liegen die Punkte im QQ-Plot auf einer 45°-Geraden. Bei guten Interpolationsergebnissen sollte dies im Vergleich mit der Ausgangsvariablen der Fall sein. Je weiter die Punkte von der Geraden entfernt liegen, desto unähnlicher sind die Verteilungen und desto schlechter die Schätzwerte.

Neben den in SGeMS implementierten QQ-Plots wird mittels des Python-Skriptes *CalculateRMSE.py* (Anhang E) der RMSE-Wert der Schätzwerte ermittelt. Hierfür muss in SGeMS das Ergebnistraster als csv-Datei exportiert werden. Auf eine Kreuzvalidierung wird wegen des Programmieraufwandes verzichtet. Die Beurteilung der Güte der gewählten Kriging- und Kokriging-Modelle erfolgt damit ausschließlich auf Basis der allgemeinen QQ-Plots und des RMSE-Wertes.

2.7.7 Verarbeitungsschritte Kriging/Kokriging

Tab. 10 Verarbeitungsschritte Kriging/Kokriging

	Verarbeitungsschritt	Tool/Python-Skript
1	Erstellen einer GSLIB-Datei mit PM ₁₀ -/LNI-Daten	Editor
2	Datenimport in SGeMS	
3	Gaußsche Transformation der PM ₁₀ -/LNI-Werte	SGeMS
4	Export der Gauß-transformierten Daten als csv-Dateien	SGeMS
5	Import der Gauß-transformierten PM ₁₀ -Daten in Voxler	

6	Interpolation der Gauß-transformierten PM ₁₀ -Daten; IDW Power 0 (Basis für die Zuordnung zwischen Voxler- und SGeMS-Rasterzellen) + Power 1 (Werte für das De- und Retrending)	Voxler
7	Export der Rasterdaten als csv-Dateien	
8	Zuordnung der Messpunkt-Koordinaten zu Voxler-Rasterzellen als Basis für die Trendbereinigung der PM ₁₀ -Werte (Voxler IDW Power1 Raster)	<i>PM10DetrendDataPreparation.py</i> ²⁰ , Excel
9	Trendbereinigung der Gauß-transformierten PM ₁₀ -Werte (Ausgabedateien: csv- und GISLIB-Format)	<i>PM10DetrendData.py</i> ²⁰
10	Wertebereich der Gauß-transformierten LNI-Daten zur Eliminierung von Extrema einschränken und als weitere Spalte an csv-Datei hängen	<i>DataSplitLNI.py</i> ²⁰
11	Gauß-transformierte LNI-Werte mit und ohne Extrema mit den trendbereinigten PM ₁₀ -Werten zusammenfügen (Ausgabedatei: GSLIB-Format)	<i>PM10DetrendedDataAddLNI.py</i> ²⁰
12	Datenimport nach SGeMS	
13	Gaußsche Transformation der trendbereinigten PM ₁₀ -Werte	SGeMS
14	Variogramphie Kriging / KoKriging	SGeMS
15	Simple Kriging PM ₁₀ / Simple Kokriging PM ₁₀ -LNI	SGeMS
16	Gaußsche Rücktransformation der PM ₁₀ -Schätzwerte	SGeMS
17	Export der Interpolationsraster als csv-Datei	SGeMS
18	x-/y-/z-Koordinaten zur csv-Datei hinzufügen	<i>AddCoordinatesToSGeMSExport.py</i> ²⁰
19	Retrending der PM ₁₀ -Schätzwerte (Ausgabedatei: GSLIB-Format)	<i>PM10RetrendData.py</i> ²⁰
20	Datenimport nach SGeMS	
21	Gaußsche Rücktransformation der retrendeten PM ₁₀ -Schätzwerte	SGeMS
22	Erstellen der QQ-Plots für Validierung	SGeMS
23	Export der Interpolationsraster	SGeMS
24	Berechnung RMSE der Interpolationsergebnisse	<i>CalculateRMSE.py</i> ²⁰

²⁰ Python-Skripte: Anhang E

3 Ergebnisse

3.1 Deskriptive Statistik

Tab. 11 Deskriptive Statistik: Maßzahlen der Stichprobe

	01.09.18	02.09.18	03.09.18	04.09.18	05.09.18	06.09.18	07.09.18	08.09.18
Anzahl Messwerte	16	16	16	16	16	15	15	15
Anzahl fehlende Messwerte	0	0	0	0	0	1	1	1
Minimum	6,70	4,78	10,59	18,55	18,78	15,79	7,97	6,04
Maximum	21,92	21,42	31,6	41,76	48,81	37,48	25,29	18,44
Range	15,22	16,64	21,01	23,21	30,03	21,69	17,32	12,4
Mittelwert	13,62	13,01	22,92	33,09	31,46	28,44	14,18	11,56
Median	13,29	12,42	22,08	34,15	30,36	29,56	14,19	10,78
Modus²¹	15,22	15,00	22,73	35,00	(30,00)	26,00	14,45	13,89
Standardabweichung	3,83	4,22	5,66	6,62	8,32	5,39	4,25	3,57
Variationskoeffizient	0,281	0,324	0,247	0,200	0,264	0,190	0,300	0,309
Skewness	0,964	0,231	-0,385	-0,578	0,339	-0,861	1,043	0,31
Kurtosis	1,738	0,711	-0,148	-0,285	-0,349	1,313	2,347	-0,443

²¹ Berechnung vgl. Kap. 2.2; Modus für den 05.09.2018 nicht eindeutig bestimmbar, da zwei Klassen mit der höchsten Werteanzahl vorliegen.

3.2 Explorative Räumliche Datenanalyse (ESDA)

3.2.1 Kurvendiagramm Messwertreihen

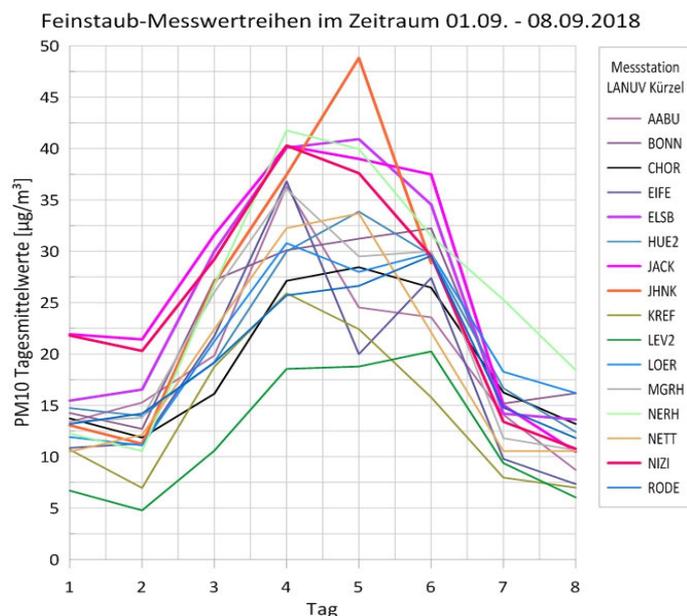


Abb. 17 Liniendiagramm Verlauf PM₁₀-Tagesmittelwerte

Anhand des Kurvendiagramms kann man bereits zwei wichtige Aussagen treffen. Zum einen ist gut zu erkennen, dass die höchsten PM₁₀-Tagesmittelwerte an allen Messstationen in der Mitte des Untersuchungszeitraums (Tag 4 + 5) gemessen wurden, zum anderen, dass an fast allen Tagen die höchsten PM₁₀-Werte an den Stationen in Tagebaunähe, die im Diagramm hervorgehoben sind, gemessen wurden. Hier sticht insbesondere der Peak der Feinstaubbelastung an der Messstation Jüchen-Hochneukirch (JHNK) am 05.09.18 (Tag 5) hervor. Im gesamten Untersuchungszeitraum wird der Grenzwert für PM₁₀-Tagesmittelwerte von 50 µg/m³ nicht überschritten.

3.2.2 Histogramme

PM₁₀-Tagesmittelwerte nach Untersuchungstag:

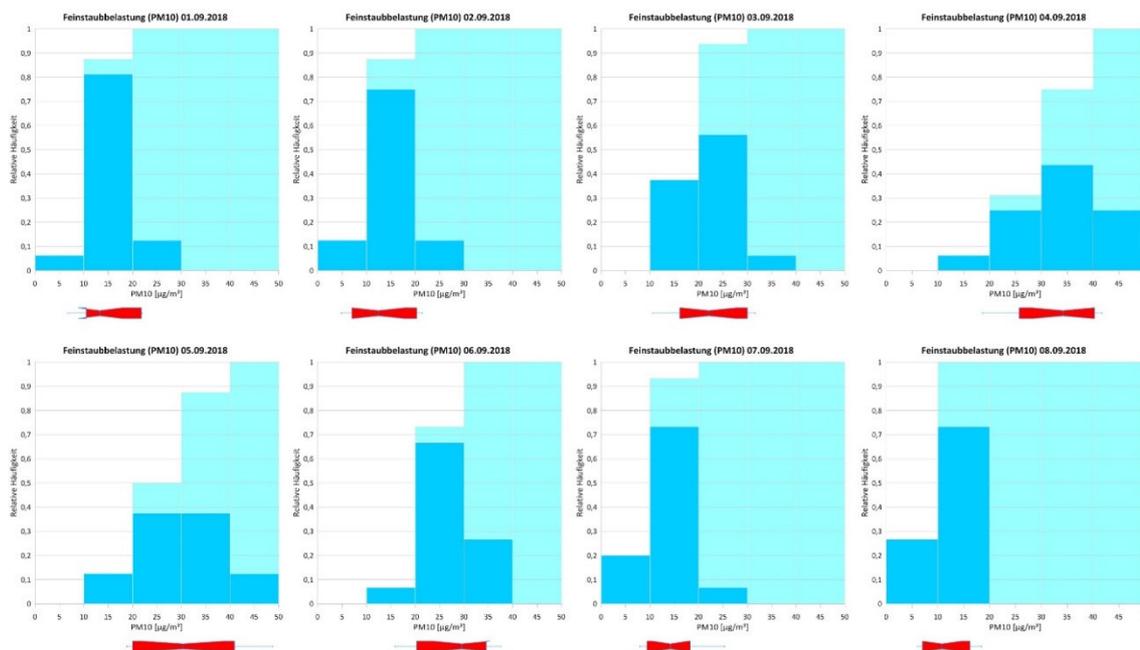


Abb. 18 Histogramme PM₁₀-Messwerte 01.09.2018 - 08.09.2018

An den ersten beiden Tagen wurde an allen Stationen eine geringe bis mittlere Feinstaubbelastung gemessen. Die Verteilungen sind rechtsschief (Skewness: 0,96/0,23) und weisen eine höhere Dichte als bei einer Normalverteilung auf (Kurtosis: 1,74/0,71) auf. Die Streuung der Werte ist sehr gering.

Am dritten Untersuchungstag sind die Messwerte deutlich höher, und der Range ist breiter geworden. Die Verteilung zeigt jetzt eine Linksschiefe (Skewness: -0,39) und eine geringere Dichte als bei einer Normalverteilung (Kurtosis: -0,15). Vor allem die Symmetrie weicht gut erkennbar von einer Normalverteilung ab.

Am vierten Untersuchungstag steigen die Messwerte weiter an und erreichen bezogen auf das arithmetische Mittel und den Median den Höhepunkt im Untersuchungszeitraum. Die Verteilung ist linksschief (Skewness: -0,58) und stärker streuend als bei einer Normalverteilung (Kurtosis: -0,28).

Am fünften Untersuchungstag sank die Feinstaubbelastung im Mittel wieder leicht. Markant im Box-Whisker-Plot ist der deutliche Ausreißer nach oben. Die Verteilung ist leicht rechtsschief (Skewness: 0,34) und weist eine stärkere Streuung als bei einer

Normalverteilung auf (Kurtosis: -0,35). Die Verteilungshäufigkeit erreicht in keiner Klasse ein absolutes Maximum und ist daher möglicherweise nicht unimodal.

Am sechsten Tag sank die Feinstaubbelastung weiter, sodass es zu einer Verschiebung in den mittleren Bereich des Histogramms kommt. Mit einer Kurtosis von 1,31 ist die Verteilung sehr kompakt und mit einer Skewness von -0,86 linksschief.

Am siebten Untersuchungstag erreichte die durchschnittliche Feinstaubbelastung in etwa wieder das Niveau der beiden ersten Tage. Die Verteilung ist rechtsschief (Skewness: 1,04). Es zeigt sich eine sehr geringe Streuung der Werte (Kurtosis: 2,35). Die im Mittel geringsten Feinstaubwerte wurden am letzten Untersuchungstag gemessen. Im Box-Whisker-Plot können zwei Ausreißer identifiziert werden, die jedoch nicht deutlich unter dem 10. bzw. über dem 90. Perzentil liegen. Die Verteilung ist leicht rechtsschief und mit einer Kurtosis von -0,44 etwas weniger kompakt als eine Normalverteilung.

Bei der Darstellung von nur 5 Klassen in einem Histogramm lässt sich anhand der grafischen Verteilung nur bedingt eine Aussage treffen, da die Gestalt schwer zu bestimmen ist. Erschwerend kommt bei den acht Histogrammen hinzu, dass zur Gewährleistung der Vergleichbarkeit der Stichproben ein Wertebereich und die Klassengrenzen für den gesamten Untersuchungszeitraum bestimmt wurden. Die Wertebereiche für die einzelnen Tage sind hingegen kleiner, sodass sich die Werte nur auf zwei bis vier, jedoch nie auf alle fünf Klassen verteilen. Ob und inwieweit die Verteilung je Stichprobe von einer Normalverteilung abweicht, ist aus den Histogrammen allein nicht ableitbar.

PM₁₀-Tagesmittelwerte gesamt (2D+t):

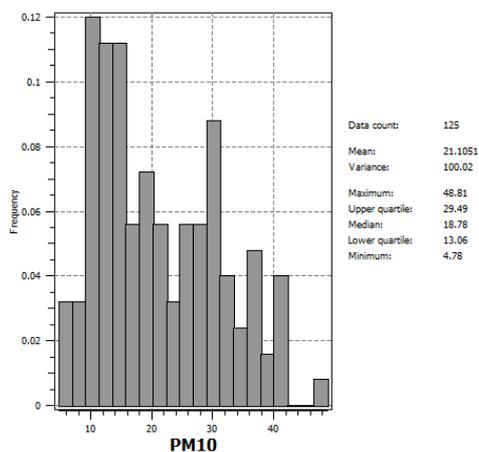


Abb. 19 Histogramm PM₁₀-Messwerte

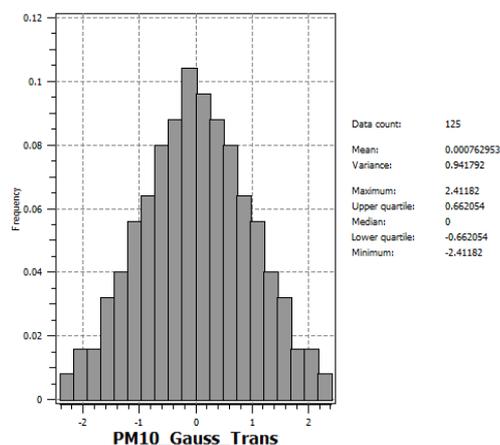


Abb. 20 Hist. PM₁₀-Messwerte nach erster GT

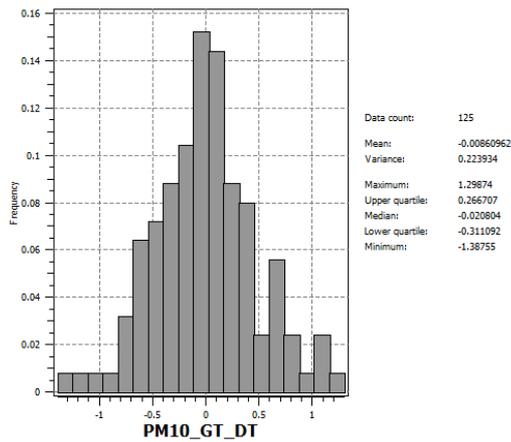


Abb. 21 Hist. PM₁₀-Messwerte nach Detrending

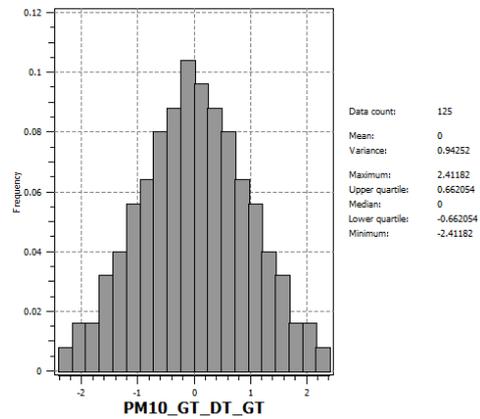


Abb. 22 Hist. PM₁₀-Messwerte nach zweiter GT

Die PM₁₀-Mittelwerte aller Untersuchungstage sind bimodal verteilt (Abb. 19). Die Verteilung ist mit einem deutlich erkennbaren Ausreißer als Maximum rechtsschief (vgl. Kap. 3.2.1). Die bimodale Verteilung erfordert im ersten Schritt eine Gaußsche Transformation der Daten zum Erreichen einer unimodalen Normalverteilung (Ergebnis vgl. Abb. 20), bevor der Trend eliminiert werden kann.

Nach der Trendbereinigung ist die Verteilung zwar immer noch unimodal, jedoch nicht mehr symmetrisch und kompakter als eine Normalverteilung (Abb. 21). Die Bildung von Ausreißern erfordert eine weitere Gaußsche Transformation, die zu einem zufriedenstellenden Ergebnis führt (vgl. Abb. 22).

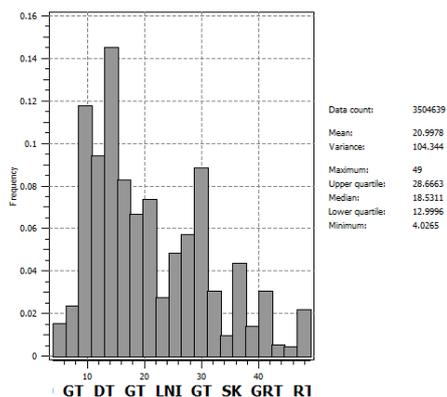


Abb. 23 Hist. PM₁₀-Messwerte nach SK/RT/GRT

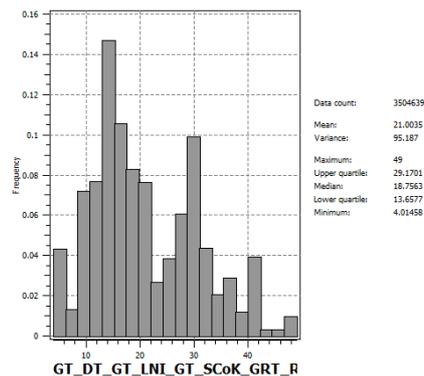


Abb. 24 Hist. PM₁₀-Messwerte nach SCK/RT/GRT

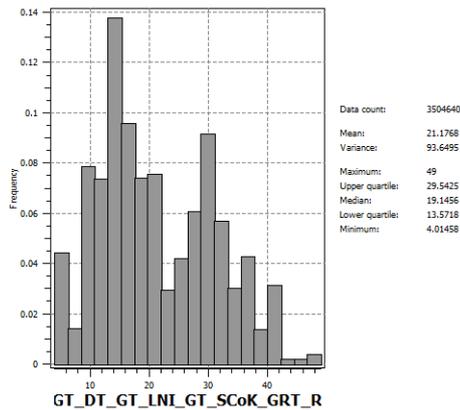


Abb. 25 Histogramm PM10-Messwerte nach SCK(β-Range)/RT/GRT

Die PM₁₀-Werte der Ergebnistraster sind nach Simple Kriging bzw. Kokriging, Retrending und zweifacher Rücktransformation wie die PM₁₀-Messwerte bimodal und rechtsschief verteilt, jedoch unterscheiden sich die Verteilungen erkennbar von der der PM₁₀-Messwerte. Die Histogramme nach beiden Kokriging-Interpolationen sind nahezu identisch (vgl. Abb. 24 und Abb. 25).

Nach dem SK sind die Werte im Vergleich zu den Messwerten (Varianz: 100.0) etwas stärker (Varianz: 104,3) und nach den SCK (Varianzen: 93,6/95,2) etwas schwächer gestreut. SK betont die Maxima und SCK die Minima. Das Histogramm nach SK (Abb. 23) hat die größte Ähnlichkeit mit dem Histogramm der PM₁₀-Messwerte.

Landnutzungsindikator β: Messstationen und zusätzliche Punkte (2D+t):

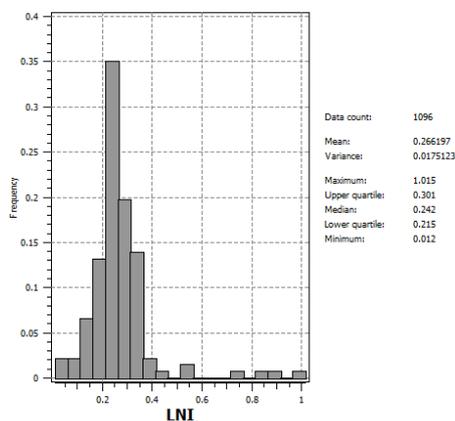


Abb. 26 Histogramm β (alle Punkt)

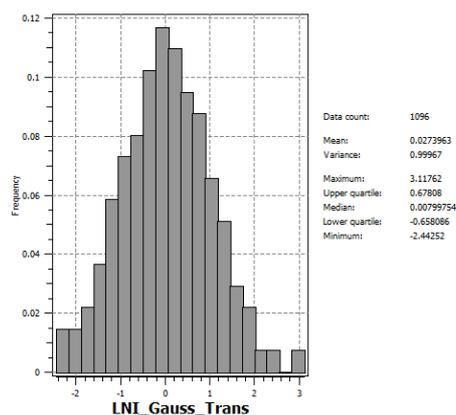


Abb. 27 Histogramm β nach GT (alle Punkte)

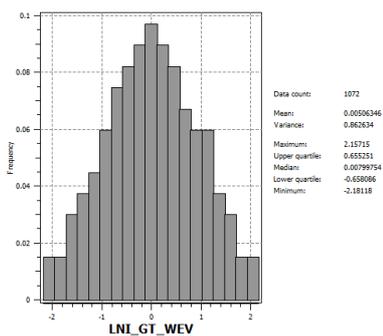


Abb. 28 Histogramm β nach GT ohne Ausreißer (β -Range)

Die Verteilung der Landnutzungsindikatoren (Abb. 26) ist unimodal und weist eine ausgeprägte Rechtsschiefe mit deutlichen Ausreißern (Mittelwert: 0,266 / Maximum: 1,015) auf. Für die Verwendung als sekundäre Variable beim Kokringing ist eine Gauß-Transformation erforderlich (Ergebnis vgl. Abb. 27). Auch nach der Transformation ist immer noch ein Ausreißer vorhanden. Erst durch die Eliminierung der Extremwerte wird die Verteilung nahezu symmetrisch (Abb. 28).

3.2.3 Normal QQ-Plots

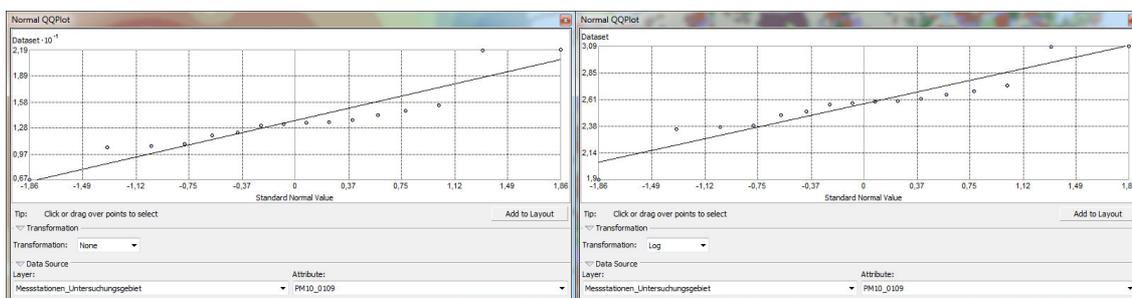


Abb. 29 Normal QQ-Plots PM₁₀-Messwerte 01.09.2018

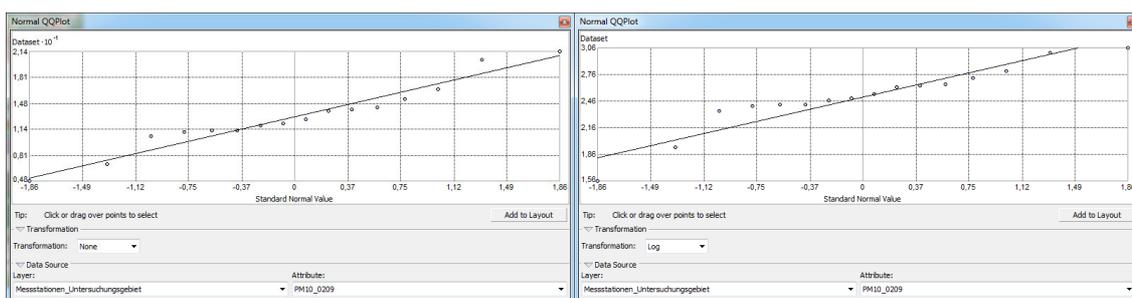


Abb. 30 Normal QQ-Plots PM₁₀-Messwerte 02.09.2018

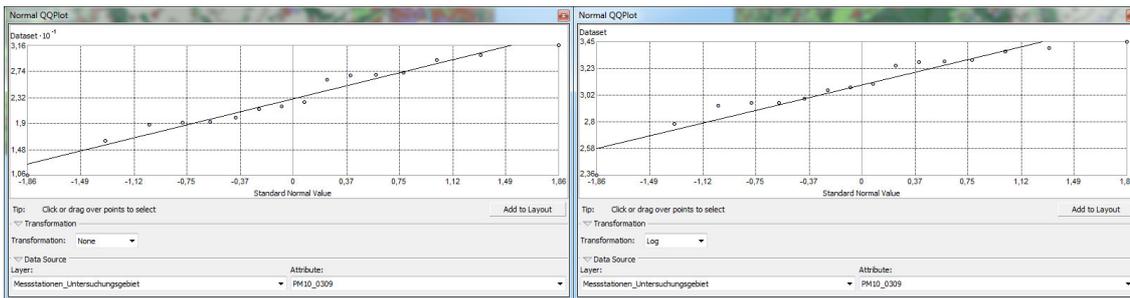


Abb. 31 Normal QQ-Plots PM₁₀-Messwerte 03.09.2018

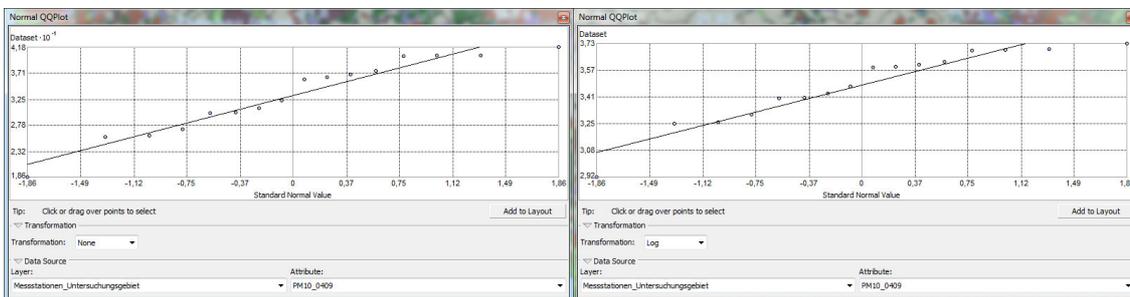


Abb. 32 Normal QQ-Plots PM₁₀-Messwerte 04.09.2018

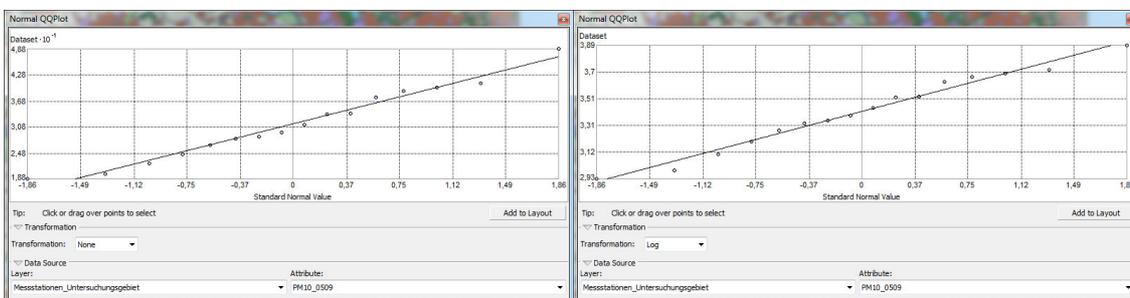


Abb. 33 Normal QQ-Plots PM₁₀-Messwerte 05.09.2018

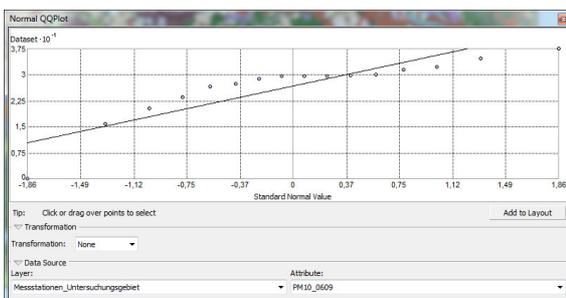


Abb. 34 Normal QQ-Plot PM₁₀-Messwerte 06.09.2018

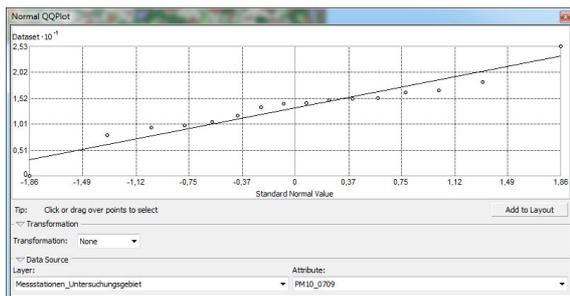


Abb. 35 Normal QQ-Plot PM₁₀-Messwerte 07.09.2018

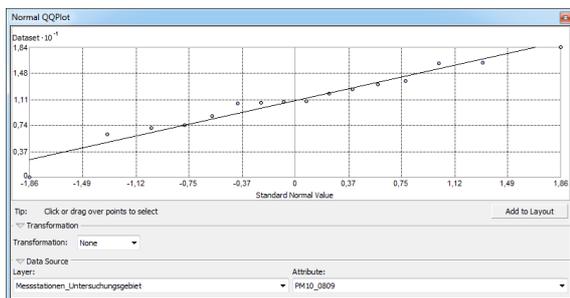


Abb. 36 Normal QQ-Plot PM₁₀-Messwerte 08.09.2018

Es zeigen sich für alle Untersuchungstage unterschiedlich stark ausgeprägte Abweichungen von der Normalverteilung. Die größten Abweichungen findet man in den Randbereichen (niedrigste und höchste Werte). Die QQ-Plots für die logarithmierten Daten liegen nur für die ersten fünf der acht Untersuchungstage vor und werden daher nicht weiter betrachtet. Dies ist wahrscheinlich in den fehlenden Messwerten an den letzten drei Untersuchungstagen begründet.

3.2.4 Raum-Zeit-Diagramm

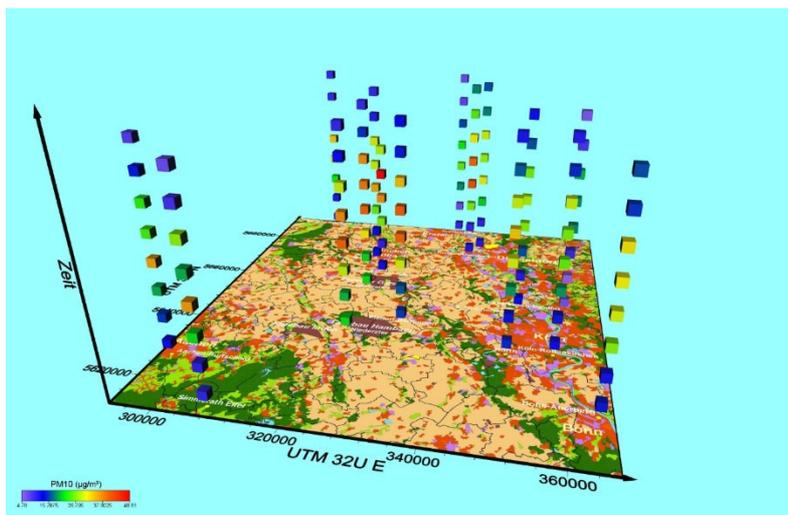


Abb. 37 Raum-Zeit-Diagramm PM₁₀-Tagesmittelwerte (01.09. - 08.09.2018)

Mit Hilfe dieser einfachen 3-dimensionalen Darstellung lassen sich bereits zwei wichtige Aussagen treffen. Wie schon im Kurvendiagramm (Kap. 3.2.1) zu sehen war, wurden die höchsten PM₁₀-Tagesmittelwerte in der Mitte des Untersuchungszeitraumes gemessen. Aus räumlicher Sicht sind die höchsten Messwerte an den Messstationen in der Mitte und im Südwesten des Untersuchungsgebietes zu finden. Dies sind interessanterweise aber nicht unbedingt überraschenderweise vor allem die Messstationen, die in Tagebaunähe eingerichtet worden sind.

3.2.5 Trendanalyse

PM₁₀-Messwerte:

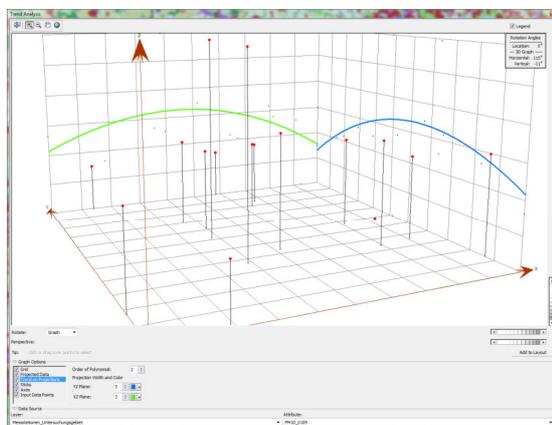


Abb. 38 Trendanalyse PM₁₀-Messwerte 01.09.2018

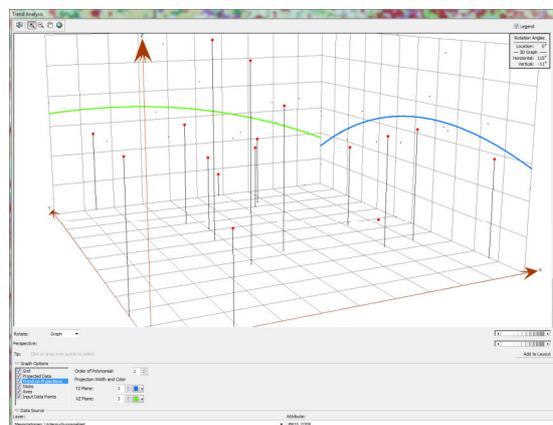


Abb. 39 Trendanalyse PM₁₀-Messwerte 02.09.2018

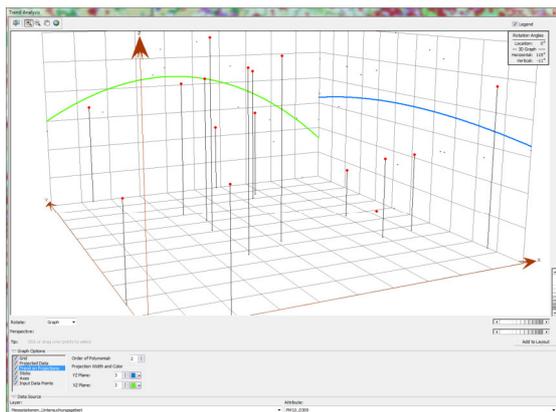


Abb. 40 Trendanalyse PM₁₀-Messwerte 03.09.2018

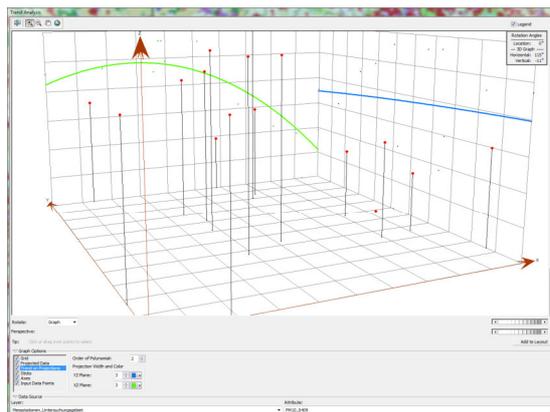


Abb. 41 Trendanalyse PM₁₀-Messwerte 04.09.2018

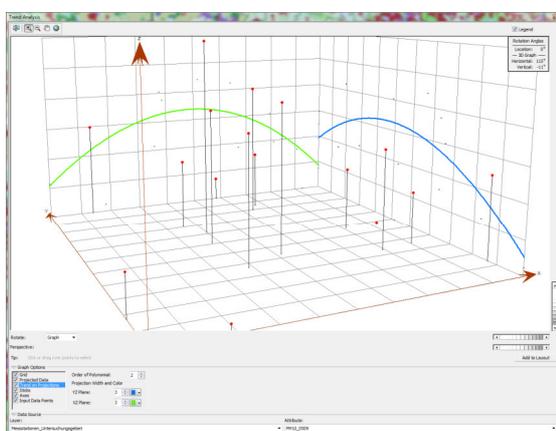


Abb. 42 Trendanalyse PM₁₀-Messwerte 05.09.2018

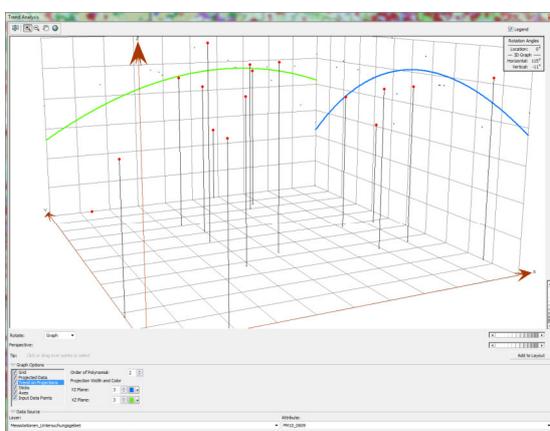


Abb. 43 Trendanalyse PM₁₀-Messwerte 06.09.2018

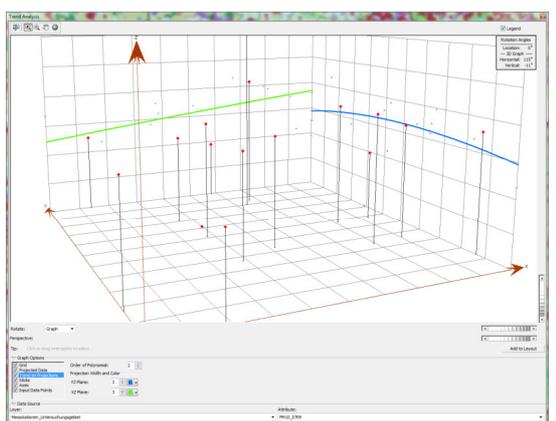


Abb. 44 Trendanalyse PM₁₀-Messwerte 07.09.2018

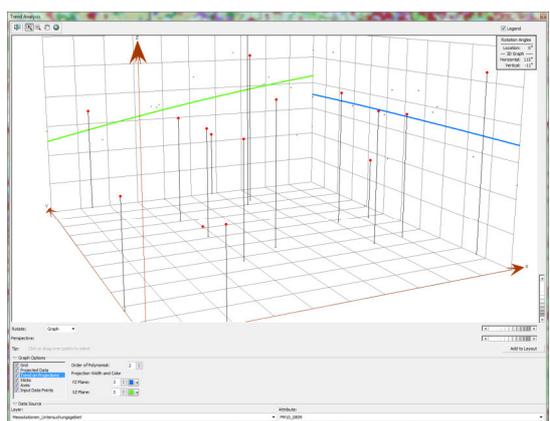


Abb. 45 Trendanalyse PM₁₀-Messwerte 08.09.2018

Die Trendanalyse der einzelnen Untersuchungstage zeigt überwiegend und teilweise ausgeprägte polynomische Trends zweiter Ordnung in Nord-Süd und West-Ost-Richtung. Daneben sind auch lineare Trends vorhanden. Hierbei fallen stärkere lineare Trends in West-Ost-Richtung an den letzten beiden Untersuchungstagen auf (Abb. 44, Abb. 45).

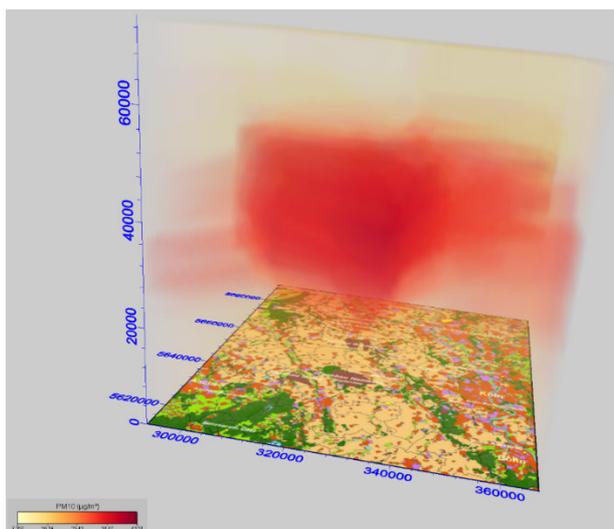


Abb. 46 Trendanalyse PM₁₀ - IDW Power 0

Die Analyse raum-zeitlicher Trends für den gesamten Untersuchungszeit-raum (Abb. 46) bestätigt, was sich bereits im Kurvendiagramm (Kap. 3.2.1) und Raum-Zeit-Diagramm (Kap. 3.2.4) angedeutet hat. Es liegen quadratische Trends in allen Richtungen (2D+t) vor, die für das trendsensitive SK eliminiert werden müssen.

Landnutzungsindikator β :

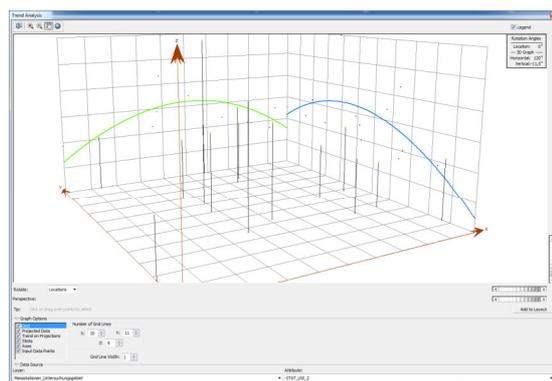


Abb. 47 Trendanalyse β Messstationen

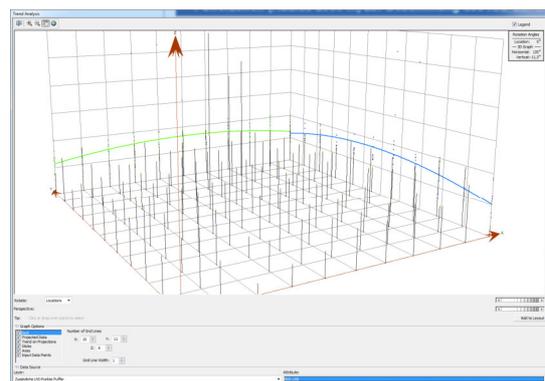


Abb. 48 Trendanalyse β zusätzliche Punkte

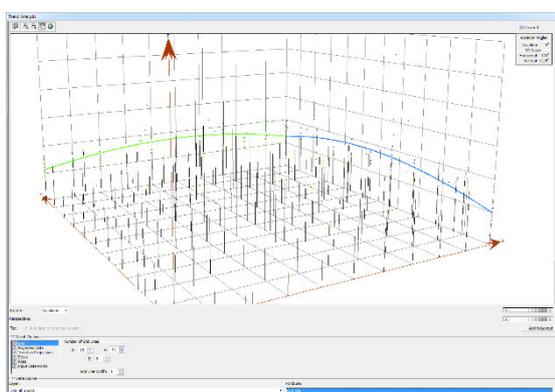


Abb. 49 Trendanalyse β Punkte gesamt

Bei der Betrachtung von β der 16 Messstationen (Abb. 47) zeigt sich ein deutlich ausgeprägter polynominaler Trend zweiter Ordnung sowohl in Nord-Süd- als auch in West-Ost-Richtung.

Die in einem gleichförmigen Raster über das gesamte Untersuchungsgebiet verteilten zusätzlichen Punkte, die den Einfluss von β als sekundäre Variable im Kokriging erhöhen sollen, weisen in beiden Richtungen nur einen schwachen polynominalen Trend zweiter

Ordnung für β auf (Abb. 48). Der ausgeprägte Trend bei den Messstationen dürfte damit zusammenhängen, dass die Stationen häufig an Orten mit einer zu erwartenden hohen Feinstaubbelastung und entsprechender Landnutzung (hoher β -Wert) installiert werden. Abb. 49 zeigt die Trendanalyse aller Punkte (16 Messstationen + 121 zusätzliche Punkte). Die Trendverläufe in Abb. 48 und Abb. 49 sind nahezu identisch. Der ausgeprägte Trend bei den Messstationen wird durch die hohe Zahl zusätzlicher Punkte (Relation ca. 1:8,5) deutlich abgeschwächt.

Die Hinzunahme der zusätzlichen Punkte, die primär zur Verbesserung des Kokriging-Ergebnisses dienen, bereinigt den Trend im ursprünglichen β -Datensatz als Nebeneffekt damit bereits weitgehend. Eine gesonderte Trendbereinigung wird nicht mehr durchgeführt.

3.3 Regressions- und Korrelationsanalyse

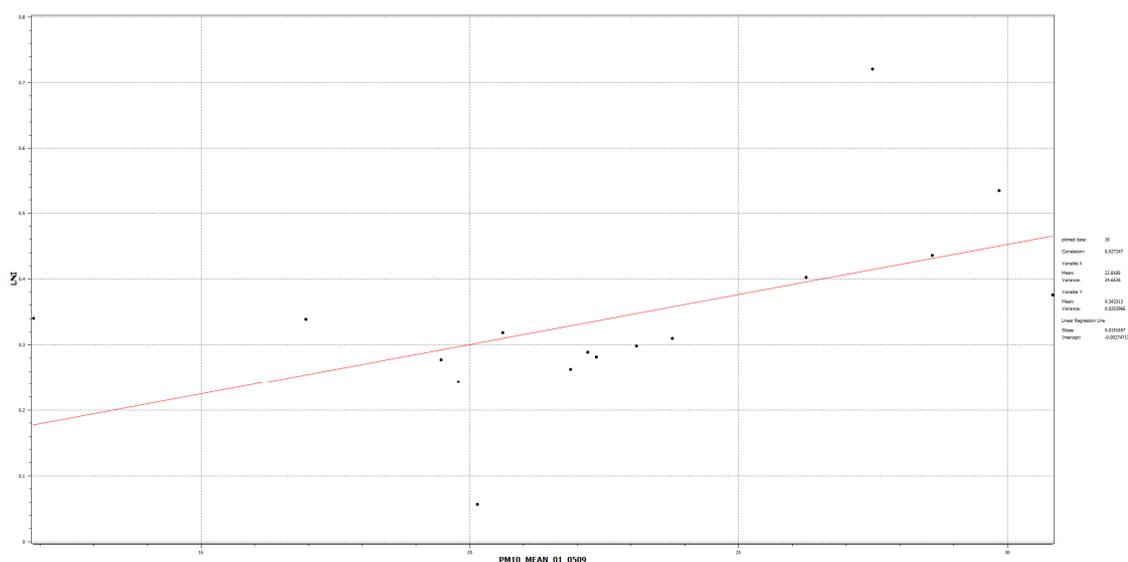


Abb. 50 Regressionsgerade PM₁₀-Mittelwerte - Landnutzungsindikator β

Korrelationskoeffizient r : 0,5272

Die PM₁₀-Mittelwerte und β korrelieren gleichsinnig. Mit einem Koeffizienten von 0,5272 ist die Korrelation jedoch nicht sehr ausgeprägt.

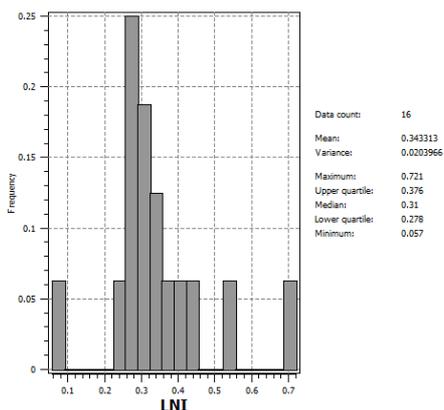


Abb. 51 Histogramm β Messstationen

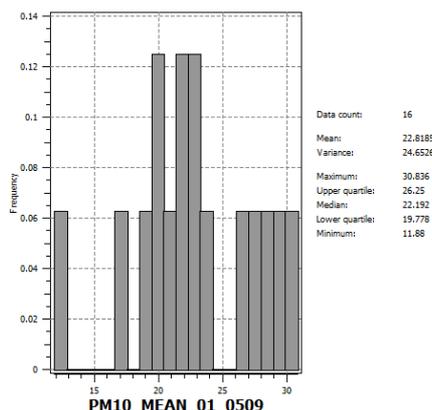


Abb. 52 Histogramm PM₁₀-Mittelwerte

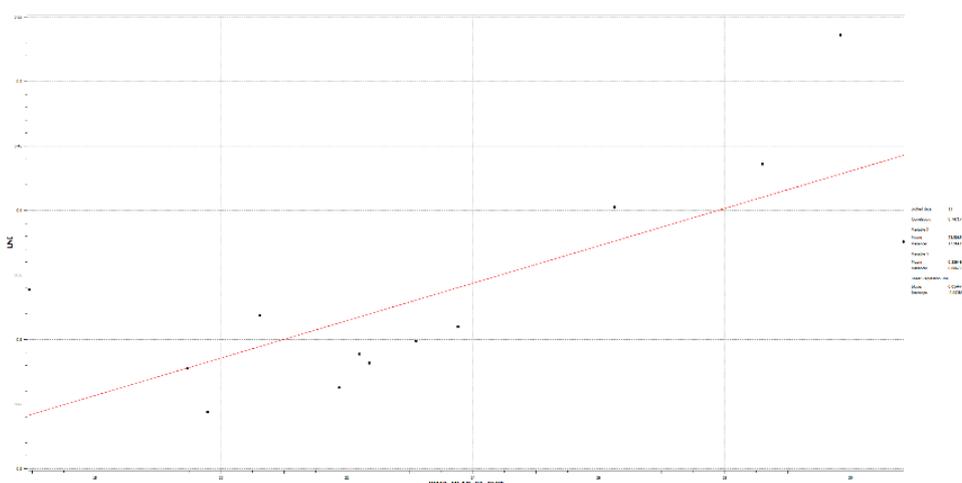


Abb. 53 Regressionsanalyse PM₁₀-Mittelwerte - β ohne Extrema

Korrelationskoeffizient r (ohne Min/Max β): 0,5843

Korrelationskoeffizient r (ohne Min/Max β + Min PM₁₀): 0,7472

Einen negativen Einfluss auf die Bildung der Regressionsgeraden und den Korrelationskoeffizienten haben Ausreißer beider Variablen (Min/Max β + Min PM₁₀), die in den Histogrammen (Abb. 51 + Abb. 52) gut zu identifizieren sind. Ohne Berücksichtigung der Ausreißer ist ein linearer Zusammenhang deutlicher zu erkennen (Abb. 53) und die Stärke der Bindung zwischen β und PM₁₀ nimmt zu ($r = 0,5272 \rightarrow r = 0,7472$). Systematische Abweichungen der Residuen fallen nicht auf.

Der Landnutzungsindikator β ist grundsätzlich als zweiter Parameter für das Kokriging geeignet. Gegebenenfalls sind die Extremwerte für das Kokriging auszuklammern.

Da das Kokriging entgegen des ursprünglichen Ansatzes, die Interpolationen zweidimensional für die einzelnen Untersuchungstage durchzuführen, für 2D+t-Daten erfolgt, muss die Stärke der Korrelation zwischen primärer und sekundärer Variablen hierfür ebenfalls ermittelt werden.

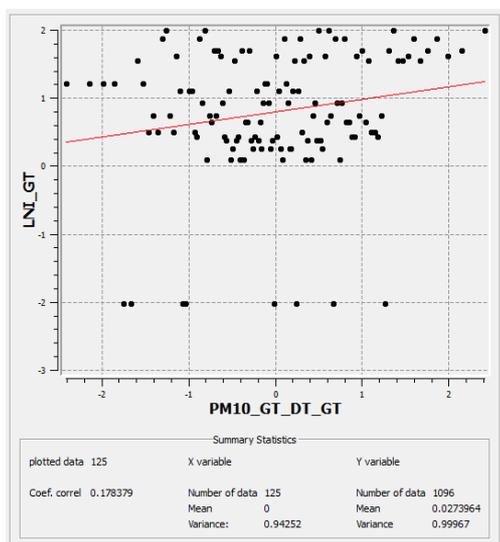


Abb. 54 Regressionsanalyse $PM_{10}(GT, DT, GT) - \beta(GT)$

Korrelationskoeffizient r : 0,1783

Die Korrelation zwischen den zweifach Gauß-transformierten, trendbereinigten PM_{10} -Werten und den Gauß-transformierten β -Werten ist mit $r = 0,1783$ sehr schwach. Das horizontale Punktemuster resultiert aus einem über die Zeit konstanten β -Wert für jede Messstation. Bei sich im Untersuchungszeitraum ändernden PM_{10} -Werten für jeden Messpunkt bleibt Landnutzungsindikator konstant. Dies wirkt sich negativ auf die Korrelation zwischen beiden Variablen aus. Die Korrelation zwischen den zweifach Gauß-transformierten, trendbereinigten PM_{10} -Werten und den Gauß-transformierten β -Werten ohne Extrema (β -Range) ist mit $r = 0,1920$ nur geringfügig stärker.

3.4 Variographie und Interpolationen

3.4.1 Variographie Kriging und Kokriging

Variogramme:

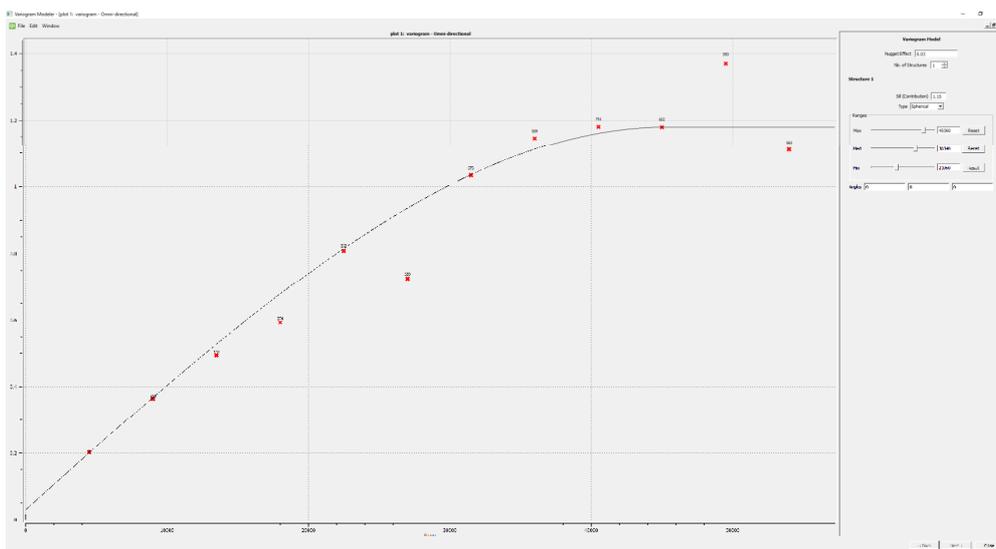


Abb. 55 Variogramm PM₁₀ (GT, DT, GT)

Type: Spherical
 Nugget: 0,03
 Sill(Contribution): 1,15
 Ranges: Max: 45360, Med: 38340, Min: 21060
 Angles: Azimuth: 0, Dip: 0, Rake: 0

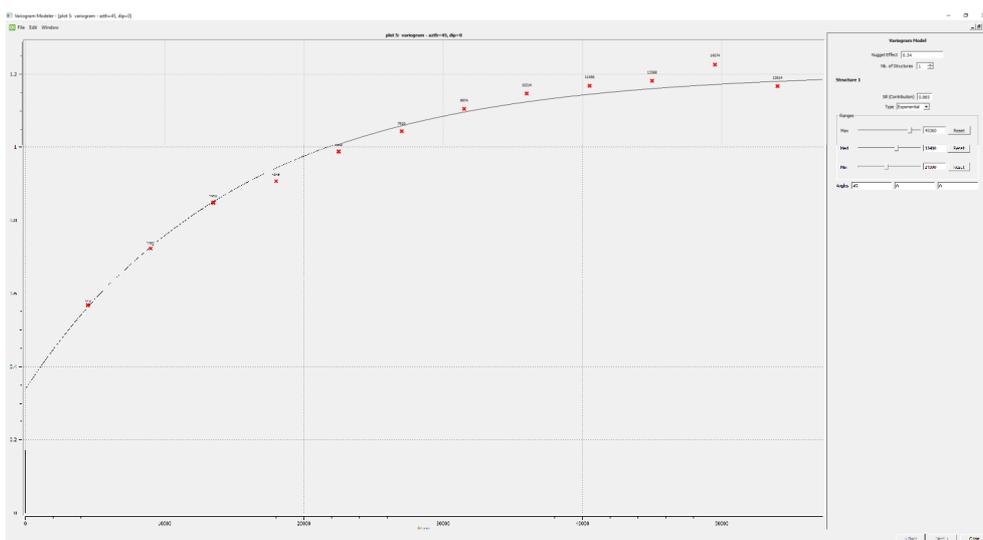


Abb. 56 Variogramm β (GT)

Type: Exponential
 Nugget: 0,34
 Sill(Contribution): 0,865

Ranges: Max: 45360, Med: 33480, Min: 24300
 Angles: Azimuth: 45, Dip: 0, Rake: 0

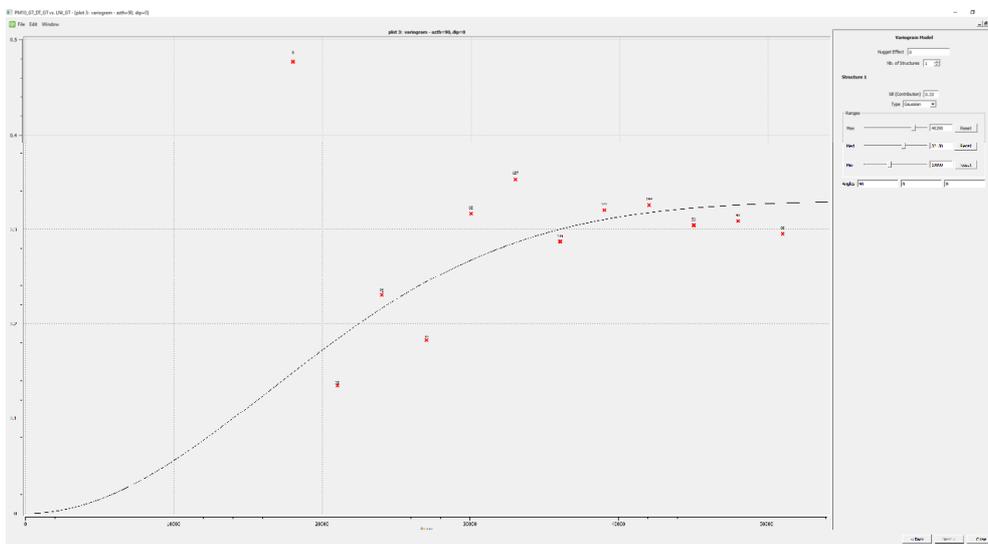


Abb. 57 Kreuzvariogramm PM₁₀ (GT, DT, GT) / β (GT)

Type: Gaussian
 Nugget: 0
 Sill(Contribution): 0,33
 Ranges: Max: 40290, Med: 32130, Min: 19890
 Angles: Azimuth: 90, Dip: 0, Rake: 0

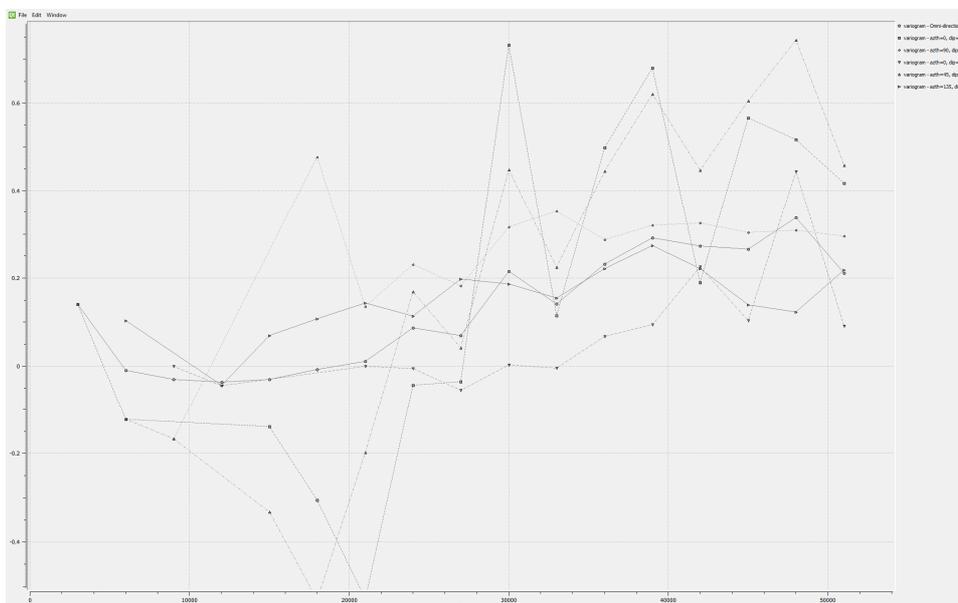


Abb. 58 Experimentelle Kreuzvariogramme PM₁₀ (GT, DT, GT) / β (GT)

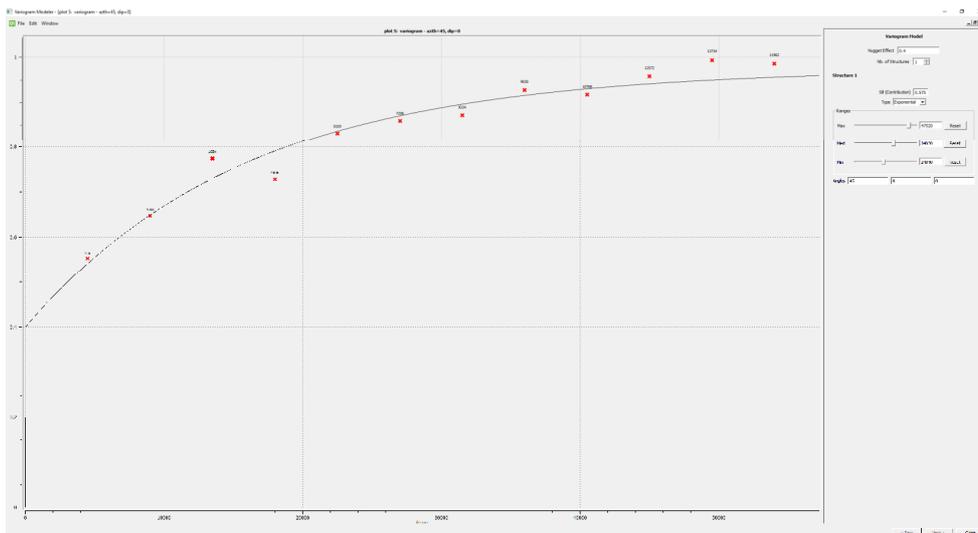


Abb. 59 Variogramm β -Range (GT)

Type: Exponential
 Nugget: 0,4
 Sill(Contribution): 0,575
 Ranges: Max: 47520, Med: 34020, Min: 24840
 Angles: Azimuth: 45, Dip: 0, Rake: 0

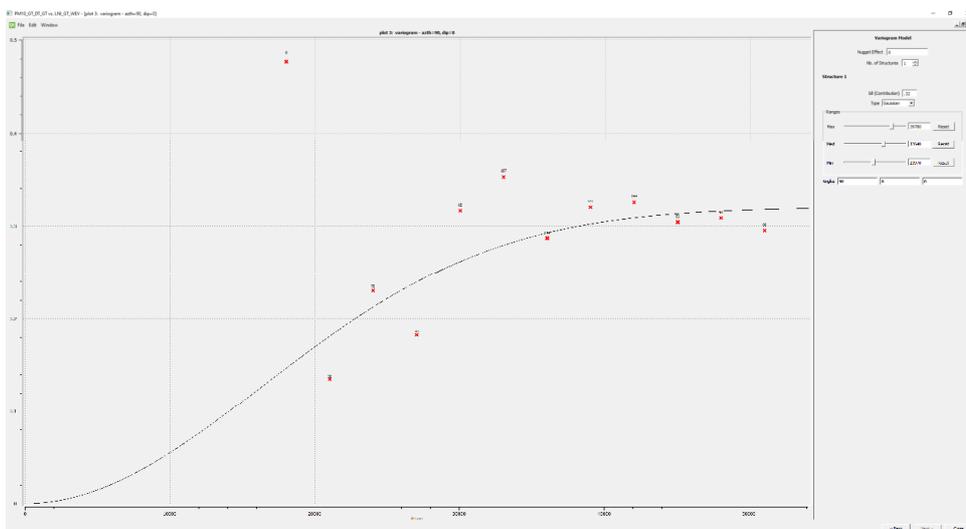


Abb. 60 Kreuzvariogramm PM_{10} (GT, DT, GT) / β -Range (GT)

Type: Gaussian
 Nugget: 0
 Sill(Contribution): 0,32
 Ranges: Max: 39780, Med: 32640, Min: 23970
 Angles: Azimuth: 90, Dip: 0, Rake: 0

Die Variogramme für PM_{10} (Abb. 55) und β/β -Range (Abb. 56, Abb. 59) weisen je nach Richtung einen unterschiedlich hohen Sill auf. Dies ist ein Hinweis auf zonale Anisotropie. Da die Zonen größtenteils nicht parallel verlaufen, sondern eher

konzentrisch angeordnet sind, ist die Verwendung einer in eine Richtung ausgeprägter Suchellipse trotz vorliegender Anisotropie nicht sinnvoll.

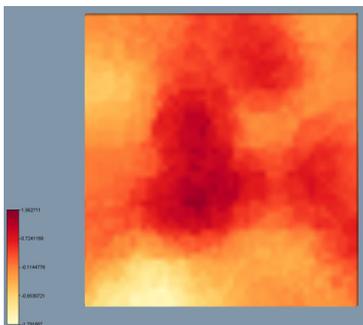


Abb. 61 Anisotropie der Gauß-transformierten β -Werte²²

Abb. 61 verdeutlicht dies am Beispiel der Gauß-transformierten β -Werte, auch wenn hier noch ein schwacher Trend zu berücksichtigen ist (Kap. 3.2.5).

3.4.2 Kriging und Kokriging

Interpolationsraster:

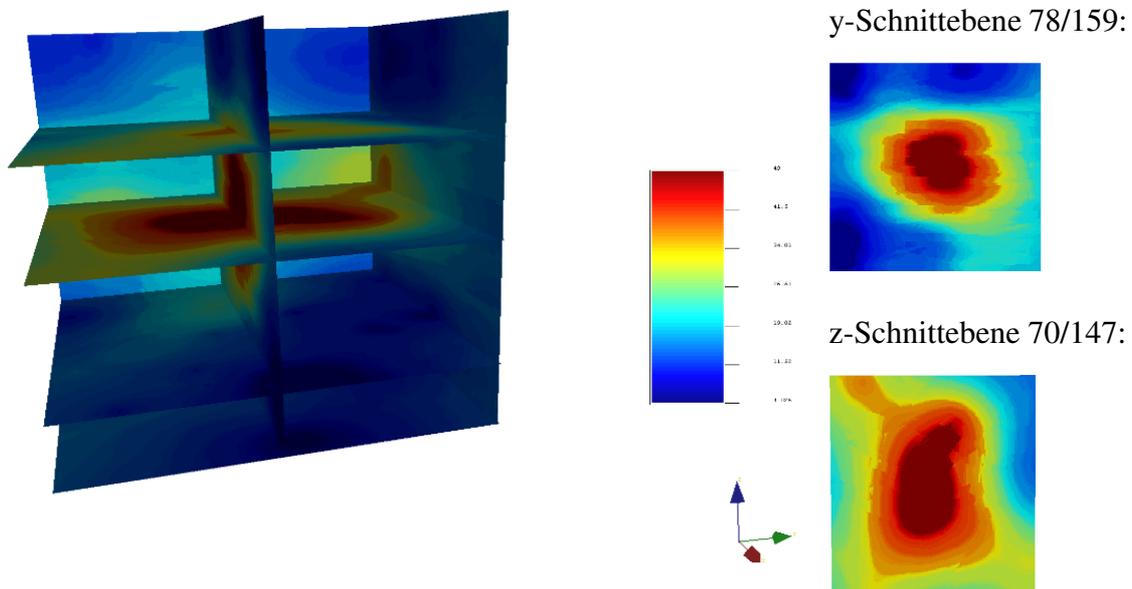


Abb. 62 Interpolationsraster Simple Kriging PM_{10} (GRT/RT/GRT)

²² IDW-Interpolation (Power 0, Smooth 0, SR: 20.000 m) mit Voxler

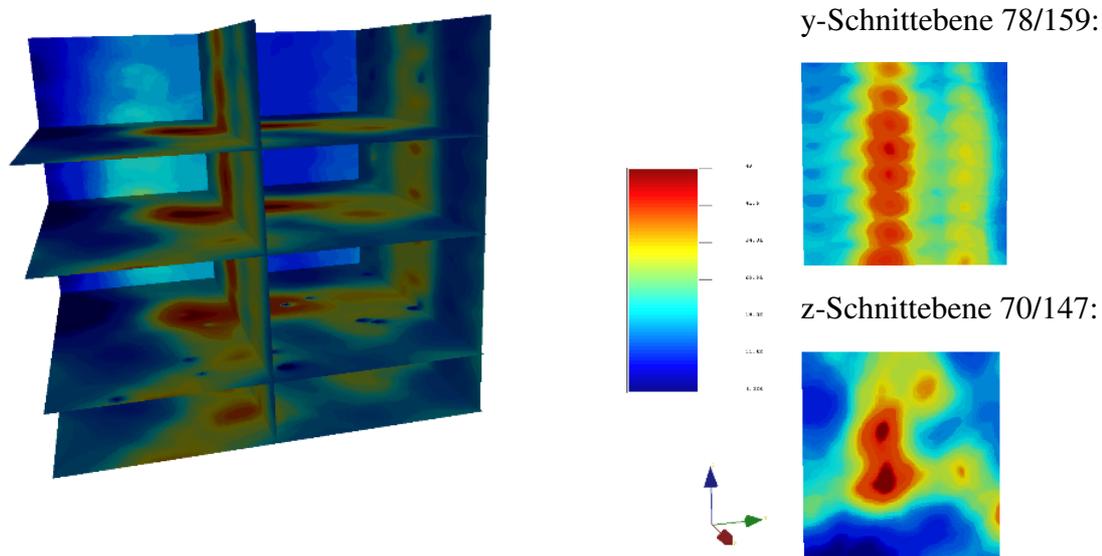


Abb. 63 Interpolationsraster Simple CoKriging PM₁₀/β (GRT/RT/GRT)

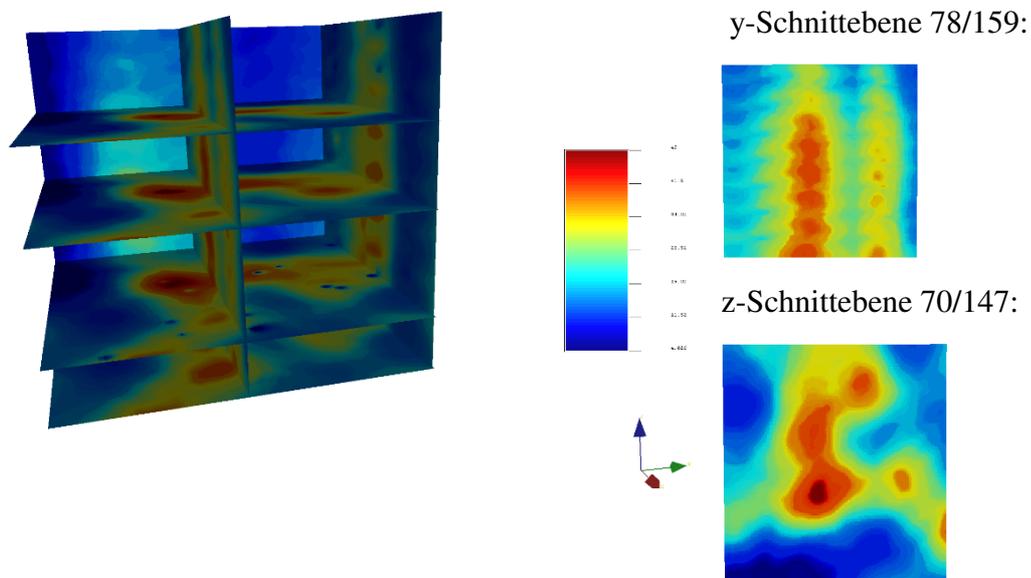


Abb. 64 Interpolationsraster Simple CoKriging PM₁₀/β-Range (GRT/RT/GRT)

QQ-Plots:

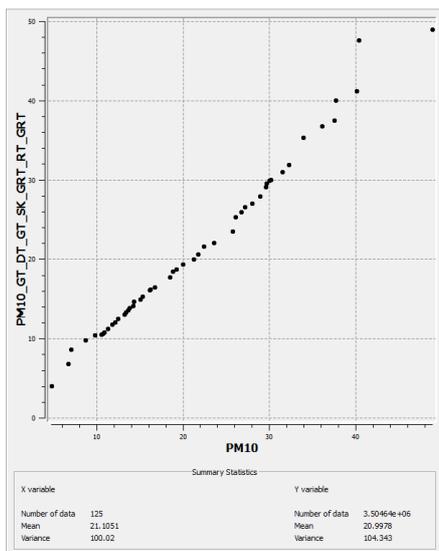


Abb. 65 QQ-Plot SK PM₁₀ - PM₁₀-Messwerte

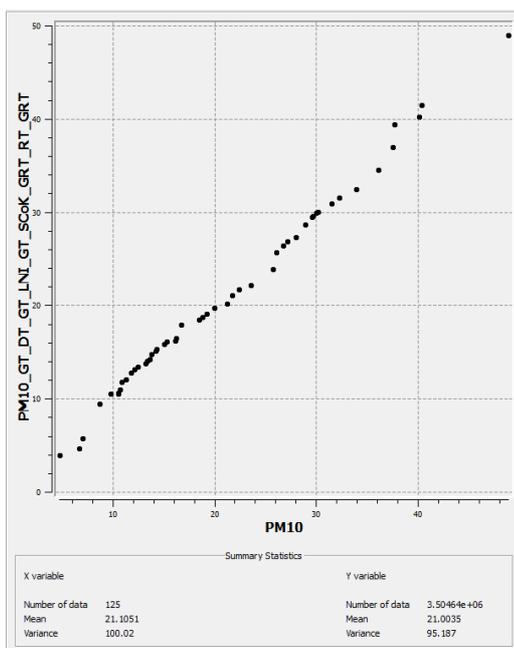


Abb. 66 QQ-Plot SCK PM₁₀/β - PM₁₀-Messwerte

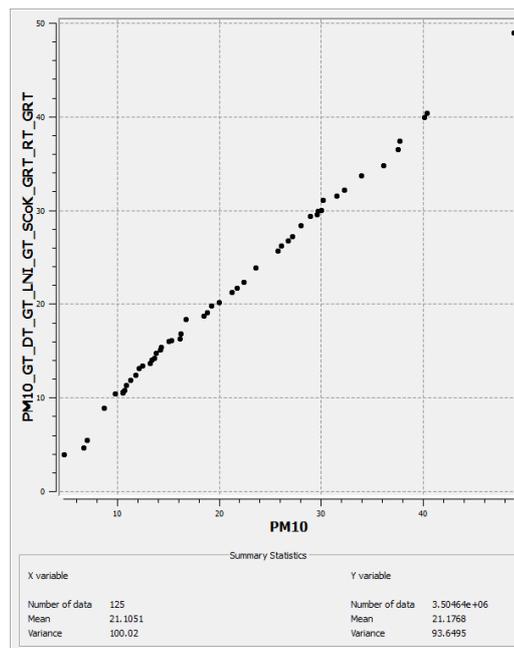


Abb. 67 QQ-Plot SCK PM₁₀/β-Range - PM₁₀-Messwerte

RMSE:

Tab. 12 RMSE Interpolationen

Verfahren	RMSE (µg/m ³)
Simple Kriging PM ₁₀	4,55
Simple CoKriging PM ₁₀ /β	10,40
Simple CoKriging PM ₁₀ /β-Range	10,49

3.4.3 Vergleich und Bewertung der Ergebnisse

Variographie:

Die Bildung eines Semivariogramm-Modells anhand eines experimentellen Variogramms, welches nur auf wenigen (Mess-)Werten basiert, gestaltet sich schwierig. Dies wäre für die PM₁₀-Daten im Fall der Betrachtung einzelner Untersuchungstage (2D) mit 15 bzw. 16 Messwerten zutreffend gewesen. Durch die Entscheidung, ein 2D+t-Modell zu verwenden, erhöht sich die Zahl der PM₁₀-Messwerte auf 125 und damit auch die Aussagekraft des experimentellen Variogramms. Durch eine explorative Herangehensweise mit unterschiedlichen Parameterwerten wurde bei allen Variogrammen versucht, ein geeignetes Variogramm-Modell zu konstruieren.

Beim experimentellen PM₁₀-Variogramm zeigt sich bei der Wahl von maximalen Punktdistanzen größer ca. 50.000 m ein Hole-Effekt - Absacken der Semivarianz nach stetig steigenden Werten. Zur Eliminierung des Hole-Effekts müsste an dieser Stelle im Variogramm ein Cosinus-Glied eingefügt werden. Solche Möglichkeiten bieten jedoch nur bestimmte Tools wie WinGSLibrary, nicht aber SGeMS. Hiervon wird daher abgesehen und eine maximale Distanz von 54.000 m gewählt. Durch den niedrigen Nugget-Wert von 0,03 (vgl. Abb. 55) ist die Unschärfe des Modells gering. Als primäre Variable beim Kokriging erhalten die PM₁₀-Werte hierdurch ein höheres Gewicht.

In beiden experimentellen Variogrammen des Gauß-transformierten Landnutzungsindikators β (mit und ohne Ausreißer) sind die Semivarianz-Punkte so angeordnet, dass ein exponentielles Variogramm-Modell leicht zu konstruieren ist (Abb. 56, Abb. 59). Um eine bestmögliche Anpassung des Modells an die Semivarianz-Punkte zu erreichen, wird ein vergleichsweise hoher Nugget-Wert gewählt. Dieser erhöht die Unschärfe des Modells und senkt den Einfluss von β als sekundäre Variable im Kokriging.

Die Generierung der beiden Kreuzvariogramm-Modelle (Abb. 57, Abb. 60) ist sehr schwierig. Zur besseren Analyse kleinräumiger Variationen wird die Anzahl der Lags im Vergleich zu den anderen Variogrammen auf 17 erhöht. Die Semivarianzen sind bereits bei geringen Distanzen sehr hoch und fallen dann ab, um anschließend wieder anzusteigen. Erschwerend hinzu kommt eine abschnittsweise negative Korrelation zwischen den Variablen hinzu, die sich je nach Richtung über einen unterschiedlichen großen Distanzbereich erstreckt und zu negativen y-Werten der betroffenen Semivarianz-

Punkte führt (vgl. Abb. 58). Die negativen Werte können bei der Bildung des Variogramm-Modells nicht berücksichtigt werden.

Die beste Anpassung an die positiven Semivarianzpunkte bietet in beiden Fällen ein Gaußsches Modell mit einem Nugget-Wert von 0. Keine Berücksichtigung findet dabei der erste Semivarianz-Punkt, der die größte Semivarianz aufweist. Das Residuum des Punktes ist durch diese Konstellation maximal. Bei geringen Punktdistanzen bildet das gewählte Kreuzvariogramm-Modell damit nicht die experimentell ermittelten Werte ab, was sich wiederum negativ auf das Kokriging auswirkt.

Interpolationsraster:

Abb. 62, Abb. 63 und Abb. 64 zeigen die Interpolationsraster inklusive der nach dem Kriging folgenden Bearbeitungsschritte (zweifache Rücktransformation und Retrending der Daten). Im Simple Kriging Raster (Abb. 62) sind die Übergänge leicht fragmentiert mit der Bildung von einzelnen kleinen Artefakten in den Grenzbereichen. Durch die alleinige Berücksichtigung der PM₁₀-Messwerte sind als negativer Effekt die Bereiche ähnlicher Schätzwerte weit ausgedehnt. So werden in Bereiche ohne Messpunkte die Werte der nächstgelegenen Messpunkte weit hinein interpoliert, was zu Verzerrungen führt. Hier ist gerade der westliche Bereich des Untersuchungsgebietes ohne eine einzige Messstation als problematisch anzusehen. Die in der ESDA identifizierten räumlichen und zeitlichen Trends spiegeln sich im SK-Raster gut wider. Insbesondere die y- und die z-Schnittebene aus dem zentralen Rasterbereich zeigen die quadratischen Trendverläufe mit den höchsten Werten in der Mitte der Untersuchungsgebietes und in der Mitte des Untersuchungszeitraums anschaulich.

In den beiden Simple Cokriging Rastern (Abb. 63, Abb. 64), die sich sehr ähneln, sind die Übergänge weniger fragmentiert als im SK-Raster und damit fließender. Der Einfluss der sekundären Variablen ist in den z-Schnittebenen gut zu erkennen. Die Areale mit höheren Schätzwerten sind kleiner als im SK-Raster. Ein markantes Beispiel ist die hervorgehobene Schnittebene 70, in der im SK-Raster kaum blaue Bereiche niedriger Schätzwerte zu finden sind, während die gleiche Ebene in den SCK-Rastern im Nordwesten, Nordosten und Süden größere Gebiete mit niedrigeren Schätzwerten aufweist (Abb. 63, Abb. 64). Genau dies sind Bereiche mit sehr niedrigen β -Werten und korrespondierender Landnutzung (z.B. Waldgebiete). Im Vergleich beider SCK-Raster

ist die Interpolationsoberfläche bei der Verwendung von β ohne Extrema (β -Range) im Kokriging etwas glatter.

Das Problem des verwendeten Kreuzvariogramm-Modells, das in bestimmten Distanzbereichen nicht das experimentelle Variogramm widerspiegelt, wurde bereits ausführlich diskutiert und zeigt sich in den SCK-Rastern in Form von lochartigen Artefakten im Bereich der Messpunkte. Dort werden sehr hohe Werte geschätzt, die sich deutlich von Schätzwerten der direkten Umgebung abheben. In z-Richtung werden die SCK-Raster verzerrt (siehe y-Schnittebene 78), was sich durch vertikale, längliche Artefakte zeigt. Auf die Berechnung der PM_{10} -Schätzwerte in z-Richtung hat der über die Zeit konstante β -Wert als sekundäre Variable einen stark angleichenden Einfluss. Dies muss ebenfalls als gravierender Nachteil des Kokrigings gewertet werden.

QQ-Plots:

Der Vergleich der Verteilungen der Schätzwerte mit der Verteilung der PM_{10} -Messwerte zeigt für alle drei Interpolationen eine gute und fast gleiche Annäherung an die Messwertverteilung (Abb. 65, Abb. 66, Abb. 67). Im QQ-Plot der durch SK ermittelten Schätzwerte weicht ein Punkt deutlich nach oben ab und zeigt eine Überschätzung an. Alle anderen Punkte bilden bis auf die äußeren Quantile nahezu eine Gerade und zeigen damit in diesem Bereich die im Vergleich beste Annäherung an die Verteilung der Messwerte. Der Vergleich der beiden KoKriging QQ-Plots ergibt eine geringere Abweichung von Messwertverteilung bei der Verwendung des Landnutzungsindikators ohne Berücksichtigung von Ausreißern (vgl. Abb. 67). Die Gesamtanalyse der QQ-Plots ergibt eine in etwa gleiche Eignung der Interpolationen ohne erkennbare wesentliche Vorteile für ein Verfahren.

RMSE:

Die Abweichungen der Schätz- von den Messwerten sind im Vergleich zum Kokriging beim Kriging deutlich geringer. Mit einem RMSE-Wert von $4,55 \mu\text{g}/\text{m}^3$ (ca. 10% des Wertebereiches) weist das Kriging zwar keinen guten, aber noch einen akzeptablen Wert auf (vgl. Tab. 12). Die RMSEs beider Kokriging-Interpolationen ($10,40 \mu\text{g}/\text{m}^3/10,49 \mu\text{g}/\text{m}^3$) sind in Bezug auf den genannten Wertebereich als schlecht einzustufen. Eine Analyse der Schätz-/Messwertabweichungen für jeden Messpunkt (Python-Skript

CalculateRMSE.py, Anhang E) zeigt für das Kokriging große Abweichungen bei Extremwerten und eher kleine Abweichungen in den mittleren Quartilen.

4 Diskussion

4.1 Eignung der gewählten Verfahren und Tools

Kriging-Verfahren ermöglichen eine optimale Schätzung von Schadstoffwerten im Hinblick auf deren flächenhafte oder räumliche Verteilung ohne Bias (BLUP) und mit einer geringen Varianz (Oliver und Webster, 2014). Diese sind daher auch für die Schätzung der Luftschadstoffkonzentration an Punkten zwischen den Messpunkten in einem Untersuchungsgebiet sehr gut geeignet. Durch die Implementierung in vielen GIS-Systemen für die geostatistische Analyse gehört Kriging heute zu den am häufigsten verwendeten Interpolationsverfahren (Oliver und Webster, 2014). Eignung und Verfügbarkeit von Interpolationsverfahren waren in Kombination mit den Erkenntnissen der explorativen Datenanalyse wesentliche Kriterien für die Wahl von Kriging und Kokriging als Interpolationsverfahren für die Fragestellung dieser Arbeit. Durch ihre Komplexität sind Kriging-Verfahren in ihrer Anwendung nicht trivial. Ein grundlegendes Verständnis der Daten, der Variographie und der Kriging-Parameter ist für die Erzielung guter Ergebnisse unabdingbar, wie auch die Ergebnisse dieser Arbeit zeigen.

Die weite Verbreitung und die meist einfache Handhabung der Tools ermöglicht es dem Benutzer, Kriging ohne die notwendige Kenntnis des Zusammenhangs zwischen Daten und Interpolationsergebnis anzuwenden (Oliver und Webster, 2014). Das Ergebnis einer Interpolation sollte daher immer ausreichend validiert und hinterfragt werden. Wie sich gezeigt hat, spielt neben der Kenntnis auch Erfahrung eine große Rolle, da die Variographie und die Parametrisierung meist eine explorative Herangehensweise erfordert. Im Rahmen dieser Arbeit wurden etwa 80 Interpolationen (OK, KT, SK, SCK) mit unterschiedlichen Parametern in SGeMS durchgeführt und miteinander verglichen. Obwohl Ordinary Kriging das für Fragestellung der Interpolation von Luftschadstoffen am häufigsten verwendete Verfahren ist, fiel die Wahl auf Simple Kriging. Vorhandene Trends konnten gut eliminiert und die Daten durch Transformation sehr nah an eine Normalverteilung gebracht werden. Somit war die Voraussetzung für die Annahme eines konstanten Mittelwertes im Untersuchungsgebiet als Bedingung für die Verwendung von SK und SCK geschaffen.

Kriging:

Die Kriging-Ergebnisse sind durch die geringe Anzahl an Messpunkten und deren ungleichmäßiger Verteilung im Untersuchungsgebiet bedingt aussagekräftig. Wie auch die Betrachtung der Fehleroberfläche zeigt, sind sie die besten Ergebnisse in Bereichen mit vielen Messstationen und die schlechtesten in Bereichen ohne oder mit sehr wenigen Messstationen zu erwarten. Auch wenn die nächstgelegenen Messpunkte in Bereichen ohne Messstation weiter entfernt liegen, haben diese immer noch einen wesentlichen Einfluss auf die Bildung der dortigen Schätzwerte. Dies führt zu Verzerrungen, wie gut an den Interpolationsrastern des SK zu erkennen ist (vgl. Abb. 62). Aussagen zu den SK-Ergebnissen sind daher immer unter Berücksichtigung der Anzahl und Lage der Messpunkte zu treffen.

Bei dem für die Beurteilung der Qualität der Ergebnisse genutzten RMSE-Wert ist zu beachten, dass dieser sensitiv auf Ausreißer reagiert, weil große Fehler stärker gewichtet werden (Hernandez-Stefanoni und Ponce-Hernandez, 2006). Eine vergleichende Bewertung mit anderen Fehlerparameter wäre hier sicher sinnvoll gewesen, konnte jedoch aufgrund des zusätzlichen Aufwandes nicht realisiert werden.

Ein Vergleich mit Studien, denen eine ähnliche Fragestellung zugrunde liegt, zeigt, dass der RMSE-Wert für die Beurteilung der Qualität der Simple Kriging - Ergebnisse als zufriedenstellend eingestuft werden kann. Beelen et al. (2009) erzielten beim Ordinary und Universal Kriging von PM_{10} -Werten RMSE-Werte auf einem in Relation zum Wertebereich vergleichbaren Niveau.

Kokriging:

Die Kokriging-Ergebnisse sind, wie in Kapitel 3.4.3 bereits dargestellt und diskutiert wurde, aufgrund der hohen Schätzwertabweichungen an den Messpunkten und den Verzerrungen in z-Richtung insgesamt nicht zufriedenstellend. Es macht dabei keinen wesentlichen Unterschied, ob der Wertebereich der Landnutzungsindikatoren zum Ausschluss von Extrema eingeschränkt wird oder nicht. Die Ergebnisse beider Kokriging-Interpolationen weichen nur geringfügig voneinander ab.

Ein differenzierter Blick auf die Kokriging-Ergebnisse ist dennoch sinnvoll, da, wie die Auswertung der QQ-Plots zeigt, die Verteilungen beider Kokriging-Interpolationen einschließlich Nachbearbeitung der Verteilung der Messwerte sehr nahe kommen und damit das SCK abgesehen von den beschriebenen Probleme auf z-Ebene durchaus

akzeptable Ergebnisse liefert. Auch die Fehleroberfläche deutet abseits der Messpunkte auf bessere Ergebnisse als beim SK hin.

Für die Verwendung des Landnutzungsindikators β als sekundäre Variable beim Kokriging spricht, dass dieser für jeden beliebigen Punkt im Untersuchungsgebiet berechnet werden kann. So sind Lokation und Anzahl der Punkte frei wählbar und nicht von den Bedingungen einer zweiten Messvariablen abhängig. Die Verteilung der Punkte im Untersuchungsgebiet und das Verhältnis zur primären Kokriging-Variablen kann damit optimal gestaltet werden, um möglichst gute Interpolationsergebnisse zu erzielen. Wie die Ergebnisse zeigen, stellen die sich teils sprunghaft ändernden β -Werte bei benachbarten Punkten ein Problem dar. Die Autokorrelation kann hierdurch bereits in einem geringen Distanzbereich zwischen den Punkten sehr schwach und damit die Semivarianz sehr groß sein. Dies wirkt sich insbesondere negativ auf die Bildung des Kreuzvariogramms aus (vgl. Kap. 3.4.3). Die Landnutzung und die damit verbundene Schadstoffemission (Verschmutzungskoeffizient α) ändert sich mitunter abrupt, wenn beispielsweise ein Waldgebiet an ein Industrieareal angrenzt. In Bereichen, in denen sich die Landnutzung und damit der Landnutzungsindikator β nicht sprunghaft ändern, scheint das Kokriging bessere Ergebnisse zu liefern.

Ein weiterer wesentlicher negativer Punkt ist die geringe Korrelation zwischen den zweifach Gauß-transformierten und trendbereinigten PM_{10} -Werten und den Gauß-transformierten β -Werten. Durch die Änderung der Messwerte an einem Messpunkt im Untersuchungszeitraum ändert sich damit auch stetig die Relation zum β -Wert des Messpunktes. Eine über Zeit die konstante sekundäre Kokriging-Variable, die nur auf der räumlichen Ebene variiert, wirkt sich negativ auf die Darstellung des zeitlichen Verlaufs in einem 2D+t-Modell aus.

Software/Tools:

Das für das Kriging gewählte SGeMS ist eine sehr leistungsfähige geostatistische Modellierungssoftware für geostatistische Interpolationen und Simulationen mit einem breiten Spektrum an implementierten Geostatistik-Tools (Remy et al., 2009). SGeMS eignet sich vor allem für die Verarbeitung von 3D-Daten mit sehr guten Visualisierungsmöglichkeiten wie dem Volume Explorer für die Betrachtung und Analyse von Raster-schnittebenen (Remy et al., 2009). Die Entwicklung von SGeMS wurde an den Bedürfnissen und Anforderungen von Power-Usern ausgerichtet (Remy et al., 2009).

Dadurch ist die Handhabung der Geostatistik-Tools für einen in geostatistischen Interpolationsverfahren unerfahrenen Anwender nicht trivial und bedarf einer intensiven Einarbeitung. Dies kann in Bezug auf die Verwendung in einer Masterarbeit als Nachteil gesehen werden. Es überwiegen aufgrund der Leistungsfähigkeit bei 3D-Interpolationen und der Parametrisierungsmöglichkeiten im Vergleich zu Geostatistik-Tools anderer GIS-Systeme vor allem beim verwendeten 2D+t-Modell deutlich die Vorteile.

Die professionelle Ausrichtung auf in der Regel programmiererfahrene User hat jedoch den Nachteil, dass von der Implementierung von Validierungsmethoden weitestgehend abgesehen wurde. So ist die mittlerweile schon zum Standard gehörende Kreuzvalidierung nicht enthalten und müsste programmiert werden. Auch alle für die Einschätzung der Qualität der Interpolationsergebnisse wichtigen Fehlerparameter (RMSE, ASE, MAE, etc.) fehlen. Um eine Aussage zur Qualität der SK- und SCK-Ergebnisse treffen zu können, wurde ein Python-Skript zur Berechnung des RMSE programmiert (Anhang E). Auf die Einbindung weiterer Fehlerparameter wurde aus Zeitgründen verzichtet.

Ein weiteres Problem sind im Programm enthaltene Bugs, die teilweise zu Abstürzen und damit zu Datenverlust führen. Identifizierte Fehler sollten daher wenn möglich mit einem Workaround umgangen werden. Mit einer Behebung der Bugs ist nicht zu rechnen, da SGeMS seit 2011 nicht mehr weiterentwickelt und supportet wird.

4.2 Aussagekraft des Modells zur Feinstaubbelastung

Trotz der vor allem beim Kokriging aufgetretenen Probleme lassen sich aus den Interpolationsergebnissen Aussagen zur Feinstaubbelastung im Braunkohlerevier und den angrenzenden Gebieten treffen. Die Analyse mit dem SGeMS Volume Explorer zeigen bei allen drei Interpolationen in Relation zum restlichen Untersuchungsgebiet stets die höchsten PM₁₀-Konzentrationen im Bereich der Tagebaue und der direkt angrenzenden Gebiete. Diese sind durchweg höher als in den durch Verkehr und Industrie stark belasteten Ballungsräumen Düsseldorf und Köln. Auch die Übergangszonen zwischen dem Revier und den größeren Städten im Umfeld weisen höhere PM₁₀-Konzentration auf. Eine bessere Differenzierung ist beim Kokriging unter Einbeziehung der Landnutzung möglich. Die Bereiche sind hier zerklüfteter und nicht so großflächig wie im Kriging-Raster. Das entspricht eher den realen Verhältnissen.

Der Schwerpunkt der Feinstaubbelastung liegt über fast den gesamten Untersuchungszeitraum hinweg in allen Interpolationsrastern im Bereich der Tagebaue und zeigt damit eine klare Tendenz, die bereits bei der Trendanalyse zu erkennen war. Die Ergebnisse sind nicht statisch signifikant, da nur ein kleiner Zeitraum von 8 Tagen im Spätsommer betrachtet wurde. Bei der Beurteilung der Feinstaubbelastung sind jahrzeitliche Schwankungen und wetterbedingte Einflüsse ebenfalls zu berücksichtigen. Niederschläge beispielweise können durch Auswaschung der PM₁₀-Partikel zu einer raschen Senkung der Feinstaubkonzentration führen.

Das RIO-Modell für Belgien zeigt in Bezug auf PM₁₀, dass in urbanisierten Regionen erwartungsgemäß höhere β -Werte zu finden sind, aber auch in einigen ländlichen Gebieten höhere β -Werte beobachtet werden (Janssen et al., 2008). Letztere Regionen sind charakterisiert durch ausgeprägte landwirtschaftliche Aktivitäten, welche einen nicht unerheblichen Einfluss auf die PM₁₀-Bildung aufgrund von primären Emissionen und Emissionen von PM₁₀-Vorläufern haben (Erisman und Schaap, 2004, Pinder et al., 2007). Dieser Punkt ist auch für das Rheinische Braunkohlerevier relevant, welches in einem wenig bewaldetem, stark landwirtschaftlich geprägtem Gebiet liegt. Vor allem in längeren Trockenheitsphasen dürfte es hier zu keiner unerheblichen Feinstaubemission durch die Landwirtschaft kommen. Es wäre daher interessant zu wissen, welchen Anteil diese Emissionen an den gemessenen PM₁₀-Werten in an den Messstationen in Tagebaunähe haben.

4.3 Bestätigung der Hypothese

Die Hypothese wird durch die Ergebnisse der Arbeit nicht bestätigt. Die deutlichen schlechteren RMSE-Werte beim Kokring und die durch die Verwendung des nicht für ein Kokring geeigneten RIO-Landnutzungsindikators auftretenden Probleme führen zu keiner Verbesserung des Kriging-Ergebnisses.

Janssen et al. (2008) und Hooyberghs et al. (2006) konnten zeigen, dass die Landnutzung über den Landnutzungsindikator als Wert für das De- und Retrending der Daten, für das dieser auch primär entwickelt wurde, in ein Kriging-Verfahren einfließen kann und zu einer Verbesserung des Ergebnisses führt. Die Hypothese hätte für eine Abgrenzung zum RIO-Modell hinsichtlich der Verwendung des Landnutzungsindikators den Begriff Kokring enthalten müssen.

5 Schlussfolgerungen und Ausblick

Die abschließende Bewertung der Interpolationsergebnisse basiert auf nur zwei Validierungsmethoden und ist hierdurch in ihrer Aussagekraft limitiert. Alle zu den Interpolationsergebnissen getroffenen Aussagen sind in diesem Kontext zu betrachten. Die Implementierung weiterer Validierungsmethoden wie etwa die sehr häufig verwendete und aussagekräftige Kreuzvalidierung und weiterer wichtiger Fehlerparameter war aufgrund des hohen Programmieraufwandes im Rahmen dieser Arbeit nicht möglich. Für eine anschließende differenziertere Analyse der Ergebnisse wäre die Einbeziehung weiterer Validierungsmethoden sinnvoll. Erst danach ist eine vollumfassende Bewertung der Ergebnisse möglich.

Das Simple Kriging der PM_{10} -Daten liefert unter Berücksichtigung der geringen Zahl an Messstationen und deren ungleichmäßiger Verteilung im Untersuchungsgebiet mit der vorgenommen Datenaufbereitung (Kap. 2.7.4) zufriedenstellende Ergebnisse. Wichtig für die Anwendung von Simple Kriging sind trendbereinigte und bei Bedarf durch Transformation geglättete Daten, da SK von einem konstanten Mittelwert ausgeht und daher sehr sensitiv auf Abweichungen reagiert. Eine Verwendung von SK ist im untersuchten Kontext für eine erste Aussage zur raum-zeitlichen Verteilung der Feinstaubbelastung valide.

Die Kokriging-Ergebnisse sind, wie in Kapitel 3.4.3 bereits dargestellt und diskutiert wurde, aufgrund der sehr hohen Schätzwertabweichungen an den Messpunkten und den Verzerrungen in z -Richtung nicht zufriedenstellend. Der Landnutzungsindikator β ist für ein $2D+t$ -Modell als sekundäre Kokriging-Variable nicht ausreichend geeignet. Die Verwendung von β als sekundäre Kokriging-Variable in einem zweidimensionalen Modell (z.B. bei Betrachtung einzelner Untersuchungstage) kann durch die deutlich bessere Korrelation mit den PM_{10} -Werten und dem Wegfall eines über die Zeit konstanten β -Wertes jedoch sinnvoll und damit Thema weiterer Untersuchungen sein.

Ein möglicher Ansatz zur Verbesserung des experimentellen Kreuzvariogramms und damit zur Optimierung des Kreuzvariogrammmodells in einem $2D+t$ -Modell wäre die Einrichtung von Pufferzonen zwischen unterschiedlichen Landnutzungsklassen, deren Verschmutzungskoeffizient a aus den Verschmutzungskoeffizienten a_i der angrenzenden Landnutzungsklassen als Mittelwert berechnet wird. Damit werden die teils sprunghaften

Änderungen des Verschmutzungskoeffizienten beim Übergang von einer Klasse in eine andere abgeschwächt, dass die Übergänge fließender werden. Dies beeinflusst auch den Landnutzungsindikator β , da die eingerichteten Pufferzonen über ihre Fläche im β -Puffer und ihren Verschmutzungskoeffizienten a in die Berechnung von β einfließen. Die β -Werte benachbarter Punkte können sich hierdurch in begrenztem Umfang angleichen und möglicherweise besser korrelieren. Das wiederum hätte einen positiven Einfluss auf die Kreuzvariogrammbildung, wenn hierdurch negative Werte im experimentellen Kreuzvariogramm reduziert oder gar vermieden werden können. Die Festlegung der Pufferzonenbreite bedarf einer explorativen Herangehensweise.

Auf Basis der in dieser Arbeit entwickelten Datenaufbereitung für das Simple Kriging sind weitere Untersuchungen vorstellbar, die andere Parameter wie etwa meteorologische Daten als sekundäre Variable in ein Simple Cokriging einbeziehen. In der durch große Freiflächen (Landwirtschaft) geprägten und wenig strukturierten Landschaft im Untersuchungsgebiet dürften bei vorherrschenden Westwinden die Windverhältnisse (Richtung, Geschwindigkeit) für den Transport von Luftschadstoffen keine unwesentliche Rolle spielen.

Zur Erzielung besserer Interpolationsergebnisse wäre zudem auch eine Erhöhung der Anzahl der Messstationen wünschenswert. Gerade durch die Einrichtung in Gebieten, die bisher noch nicht abgedeckt sind (z.B. der Westen des Untersuchungsgebietes), ließen sich über das gesamte Gebiet genauere Ergebnisse erzielen. Dabei ist zu beachten, dass die Hintergrundbelastung und nicht die spezifische Belastung durch bestimmte Emittenten (Verkehr, Industrie, etc.), die das Ergebnis verzerren würden, gemessen wird. Aufgrund der hohen Kosten für Einrichtung und Betrieb von Messstationen ist eine Erweiterung des Luftschadstoff-Messnetzes jedoch nicht zu erwarten.

Literaturverzeichnis

- AKKALA, A., DEVABHAKTUNI, V. & KUMAR, A. 2010. Interpolation techniques and associated software for environmental data. Hoboken, USA.
- ALMEIDA, S. M., PIO, C. A., FREITAS, M. C., REIS, M. A. & TRANCOSO, M. A. 2005. Source apportionment of fine and coarse particulate matter in a sub-urban area at the Western European Coast. *Atmospheric Environment*, 39, 3127-3138.
- BAHRENBERG, G., GIESE, E. & NIPPER, J. 1992. *Statistische Methoden in der Geographie. Bd. 1, Univariate und bivariate Statistik.*
- BEELEN, R., HOEK, G., PEBESMA, E., VIENNEAU, D., DE HOOGH, K. & BRIGGS, D. J. 2009. Mapping of background air pollution at a fine spatial scale across the European Union. *Science of the Total Environment*, 407, 1852-1867.
- BRAUER, M., HOEK, G., VAN VLIET, P., MELIEFSTE, K., FISCHER, P., GEHRING, U., HEINRICH, J., CYRYS, J., BELLANDER, T., LEWNE, M. & BRUNEKREEF, B. 2003. Estimating long-term average particulate air pollution concentrations: Application of traffic indicators and geographic information systems. *Epidemiology*, 14, 228-239.
- CAERS, J. 2005. *Petroleum geostatistics*, Society of Petroleum Engineers Richardson, TX.
- CHEN, H., BURNETT, R. T., KWONG, J. C., VILLENEUVE, P. J., GOLDBERG, M. S., BROOK, R. D., VAN DONKELAAR, A., JERRETT, M., MARTIN, R. V., BROOK, J. R. & COPES, R. 2013. Risk of incident diabetes in relation to long-term exposure to fine particulate matter in Ontario, Canada. *Environmental health perspectives, July 2013, Vol.121(7), pp.804-10*, 804.
- DE PRINS, S., KOPPEN, G., JACOBS, G., DONS, E., VAN DE MIEROOP, E., NELEN, V., FIERENS, F., INT PANIS, L., DE BOEVER, P., COX, B., NAWROT, T. S. & SCHOETERS, G. 2013. Influence of ambient air pollution on global DNA methylation in healthy adults: A seasonal follow-up. *Environment International, September 2013, Vol.59, pp.418-424*, 418.
- ENGLERT, N. 2007. Effects of fine particular matter pollution on mortality and life expectancy. *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz, January 2007, Vol.50(1), pp.112-8*, 112.
- ERISMAN, J. W. & SCHAAP, M. 2004. The need for ammonia abatement with respect to secondary PM reductions in Europe. *Environmental Pollution*, 129, 159-163.
- FARINA, F., SANCINI, G., BATTAGLIA, C., TINAGLIA, V., MANTECCA, P., CAMATINI, M. & PALESTINI, P. 2013. Milano Summer Particulate Matter (PM10) Triggers Lung Inflammation and Extra Pulmonary Adverse Events in Mice. *PLoS ONE, 25 February 2013, Vol.8(2)*, <xocs:firstpage xmlns:xocs=""/>.
- GHOSE, M. & MAJEE, S. 2007. Characteristics of Hazardous Airborne Dust Around an Indian Surface Coal Mining Area. *An International Journal Devoted to Progress in the Use of Monitoring Data in Assessing Environmental Risks to Man and the Environment*, 17.
- HAINING, R. P., KERRY, R. & OLIVER, M. A. 2010. Geography, spatial data analysis, and geostatistics: An overview. *Geographical Analysis*, 42, 7-31.
- HARRISON, R. M., SMITH, D. J. T., PIO, C. A. & CASTRO, L. M. 1997. Comparative receptor modelling study of airborne particulate pollutants in Birmingham (United

- Kingdom), Coimbra (Portugal) and Lahore (Pakistan). *Atmospheric Environment*, 31, 3309-3321.
- HERNANDEZ-STEFANONI, J. L. & PONCE-HERNANDEZ, R. 2006. Mapping the spatial variability of plant diversity in a tropical forest: Comparison of spatial interpolation methods. *Environmental Monitoring and Assessment*, 117, 307-334.
- HOOYBERGHS, J., MENSINK, C., DUMONT, G. & FIERENS, F. 2006. Spatial interpolation of ambient ozone concentrations from sparse monitoring points in Belgium. *Journal of Environmental Monitoring*, 8, 1129-1135.
- JANSSEN, S., DUMONT, G., FIERENS, F. & MENSINK, C. 2008. Spatial interpolation of air pollution measurements using CORINE land cover data. *Atmospheric Environment*, 2008, Vol.42(20), pp.4884-4903, 4884.
- JOHNSTON, K., VER HOEF, J. M., KRIVORUCHKO, K. & LUCAS, N. 2001. *Using ArcGIS geostatistical analyst*, Esri Redlands.
- KARNER, A. A., EISINGER, D. S. & NIEMEIER, D. A. 2010. Near-Roadway Air Quality: Synthesizing the Findings from Real-World Data. *Environmental Science & Technology*, 44, 5334-5344.
- LI, J. & HEAP, A. D. 2011. A review of comparative studies of spatial interpolation methods in environmental sciences: Performance and impact factors. *Ecological Informatics*, 2011, Vol.6(3), pp.228-241, 228.
- LI, J. & HEAP, A. D. 2014. Spatial interpolation methods applied in the environmental sciences: A review. *Environmental Modelling and Software*, March 2014, Vol.53, pp.173-189, 173.
- MATHERON, G. 1963. Principles of geostatistics. *Economic Geology*, 58, 1246-1266.
- NAGL, C. & SPANGL, W. 2007. Flächenhafte Darstellung von Luftschadstoffen. *Machbarkeitsstudie zur Berechnung der PM10-, PM2.*
- OLIVER, M. A. & WEBSTER, R. 2014. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. *Catena*, 113, 56-69.
- PINDER, R. W., ADAMS, P. J. & PANDIS, S. N. 2007. Ammonia emission controls as a cost-effective strategy for reducing atmospheric particulate matter in the eastern United States. *Environmental Science & Technology*, 41, 380-386.
- PINICHKA, C., MAKKA, N., SUKKUMNOED, D., CHARİYALERTSAK, S., INCHAI, P. & BUNDHAMCHAROEN, K. 2017. Burden of disease attributed to ambient air pollution in Thailand: A GIS-based approach. *Plos One*, 12, 18.
- PL, I. E. & SRIVASTAVA, R. 1989. An introduction to applied geostatistics. Oxford, Oxford University Press.
- PLESS-MULLOLI, T., KING, A., HOWEL, D., STONE, I. & MEREFIELD, J. 2000. PM 10 levels in communities close to and away from opencast coal mining sites in Northeast England. *Atmospheric Environment*, 2000, Vol.34(19), pp.3091-3101, 3091.
- POPE, C. A. & DOCKERY, D. W. 2006. Health effects of fine particulate air pollution: Lines that connect. *Journal of the Air & Waste Management Association*, 56, 709-742.
- REMY, N., BOUCHER, A. & WU, J. 2009. *Applied Geostatistics with SGeMS: A user's guide*.
- ROJANO, R., MANZANO, C., TORO, R., MORALES, R., RESTREPO, G. & LEIVA, M. 2018. Potential local and regional impacts of particulate matter emitted from one of the world's largest open-pit coal mines. *An International Journal*, 601.
- ROSS, Z., JERRETT, M., ITO, K., TEMPALSKI, B. & THURSTON, G. D. 2007. A land use regression for predicting fine particulate matter concentrations in the New York City region. *Atmospheric Environment*, 41, 2255-2269.

- TABAKU, A., BEJTJA, G., BALA, S., TOCI, E. & RESULI, J. 2011. Effects of air pollution on children's pulmonary health. *Atmospheric Environment*, 2011, Vol.45(40), pp.7540-7545, 7540.
- TOBLER, W. R. 1970. COMPUTER MOVIE SIMULATING URBAN GROWTH IN DETROIT REGION. *Economic Geography*, 46, 234-240.
- VIDALE, S. & CAMPANA, C. 2018. Ambient air pollution and cardiovascular diseases: From bench to bedside. *European Journal of Preventive Cardiology*, 25, 818-825.
- WEBSTER, R. 2000. Is soil variation random? *Geoderma*, 97, 149-163.

Anhang A Weitere Messdaten

Tab. 13 Messwerte Windgeschwindigkeit (24H Mittelwert)

Station (Kürzel)	Windgeschwindigkeit 24H [m/s]							
	01.09.2018	02.09.2018	03.09.2018	04.09.2018	05.09.2018	06.09.2018	07.09.2018	08.09.2018
AABU	1,599	1,812	1,109	1,157	1,207	1,586	3,004	2,999
BONN	1,689	1,694	1,364	1,342	1,301	1,49	2,638	2,11
CHOR	1,668	1,955	1,371	1,209	1,335	1,71	2,263	1,912
EIFE	1,906	2,129	1,429	1,324	1,418	1,81	2,283	2,154
ELSB								
HUE2								
JACK	1,497	1,746	1,283	1,172	1,121	1,417	1,577	1,664
JHNK								
KREF								
LEV2								
LOER								
MGRH	2,039	2,582	1,35	1,203	1,187	1,542	1,941	1,755
NERH	2,415	3,736	1,861	1,359	1,477	2,625	3,48	2,773
NETT	2,361	3,247	1,814	1,488	1,694	2,777	4,531	3,768
NIZI	1,937	2,362	1,644	1,807	1,489	2,288	3,259	2,898
RODE	1,615	1,673	1,266	1,188	1,304	1,348	1,712	1,422

Tab. 14 Messwerte Windrichtung vektoriell (24H Mittelwert)

Station (Kürzel)	Windrichtung vektoriell 24H [°]							
	01.09.2018	02.09.2018	03.09.2018	04.09.2018	05.09.2018	06.09.2018	07.09.2018	08.09.2018
AABU	67,73	68,68	333,45	317,63	278,22	259,57	233,06	216,2
BONN	352,95	117,31	322,6	26,54	345,64	288,79	249,9	214,6
CHOR	4,63	40,25	5,42	12,63	318,78	292,86	245,7	217,11
EIFE	55,45	84,66	10,73	61,29	65,23	298,79	261,01	254,88
ELSB								
HUE2								
JACK	20,52	47,29	341,75	355,45	321,85	291,02	225,24	212,84
JHNK								
KREF								
LEV2								
LOER								
MGRH	26,19	45,51	357,9	5,87	342,4	279,52	242,13	231,74
NERH	35,35	53,82	24,76	320,41	14,9	272,12	244,01	218,38
NETT	42,28	55,37	346,9	343,04	341,71	279,54	248,32	225,38
NIZI	39,87	65,47	351,75	344,58	295,12	274,48	249,13	241,87
RODE	15,39	85,8	12,49	56,18	9,84	292,57	253,54	212,53

Tab. 15 Messwerte Niederschlag (24H-Summe)

Station (Kürzel)	Niederschlag 24H Summe [mm]							
	01.09.2018	02.09.2018	03.09.2018	04.09.2018	05.09.2018	06.09.2018	07.09.2018	08.09.2018
AABU	0	0	0	0	1,4	1,1	1,6	0
BONN								
CHOR	0	0	0	11,8	0,1	8,5	0,2	0
EIFE	0	0	1,6	0,3	2,1	5	2,1	0
ELSB								
HUE2								
JACK								
JHNK								
KREF								
LEV2								
LOER								
MGRH	0	0	0	0	0	8,6	0,6	0
NERH								
NETT								
NIZI								
RODE	0	0	0	0,1	6,6	2,1	0,1	0

Anhang B Landnutzungsindikatoren

Tab. 16 Landnutzungsindikatoren Messstationen

Station (Kürzel)	UTM-Koordinaten		θ
	32U E	32U N	
AABU	295020	5626578	0,2630
BONN	364754	5624181	0,2990
CHOR	351630	5654105	0,2780
EIFE	307805	5614792	0,0570
ELSB	330989	5644503	0,4360
HUE2	350417	5638209	0,2820
JACK	321314	5656678	0,3760
JHNC	322818	5663728	0,7210
KREF	335631	5690013	0,3390
LEV2	360103	5654936	0,3410
LOER	341743	5679982	0,3190
MGRH	319970	5670169	0,3100
NERH	339667	5676748	0,4030
NETT	304636	5689916	0,2890
NIZI	321990	5639919	0,5360
RODE	358285	5639512	0,2440

Tab. 17 Landnutzungsindikatoren Zusätzliche Punkte

UTM-Koordinaten			θ	UTM-Koordinaten			θ	UTM-Koordinaten			θ
32U E	32U N			32U E	32U N			32U E	32U N		
295000	5614500	0,0390	323000	5614500	0,1380	351000	5614500	0,2300			
295000	5622100	0,1930	323000	5622100	0,2270	351000	5622100	0,1880			
295000	5629700	0,3490	323000	5629700	0,3280	351000	5629700	0,1430			
295000	5637300	0,2420	323000	5637300	0,2440	351000	5637300	0,2940			
295000	5644900	0,2280	323000	5644900	0,8950	351000	5644900	0,3180			
295000	5652500	0,2420	323000	5652500	0,2220	351000	5652500	0,2480			
295000	5660100	0,2380	323000	5660100	1,0150	351000	5660100	0,2540			
295000	5667700	0,2180	323000	5667700	0,2550	351000	5667700	0,2210			
295000	5675300	0,1350	323000	5675300	0,3310	351000	5675300	0,2150			
295000	5682900	0,2810	323000	5682900	0,2430	351000	5682900	0,2290			
295000	5690500	0,1510	323000	5690500	0,2170	351000	5690500	0,1990			
302000	5614500	0,2000	330000	5614500	0,2110	358000	5614500	0,3130			
302000	5622100	0,2780	330000	5622100	0,2030	358000	5622100	0,1500			
302000	5629700	0,1720	330000	5629700	0,2190	358000	5629700	0,2900			
302000	5637300	0,2320	330000	5637300	0,3090	358000	5637300	0,3280			
302000	5644900	0,2690	330000	5644900	0,5250	358000	5644900	0,3490			
302000	5652500	0,2180	330000	5652500	0,2990	358000	5652500	0,3570			
302000	5660100	0,2500	330000	5660100	0,2380	358000	5660100	0,2440			
302000	5667700	0,1000	330000	5667700	0,2330	358000	5667700	0,2080			
302000	5675300	0,1860	330000	5675300	0,2550	358000	5675300	0,2230			
302000	5682900	0,2220	330000	5682900	0,2390	358000	5682900	0,2310			
302000	5690500	0,2930	330000	5690500	0,3310	358000	5690500	0,1940			
309000	5614500	0,0650	337000	5614500	0,2170	365000	5614500	0,0630			
309000	5622100	0,0120	337000	5622100	0,2030	365000	5622100	0,3410			
309000	5629700	0,2740	337000	5629700	0,2180	365000	5629700	0,3400			
309000	5637300	0,3120	337000	5637300	0,2540	365000	5637300	0,2780			
309000	5644900	0,2830	337000	5644900	0,2630	365000	5644900	0,1980			
309000	5652500	0,2430	337000	5652500	0,2960	365000	5652500	0,1710			
309000	5660100	0,2170	337000	5660100	0,2200	365000	5660100	0,2320			
309000	5667700	0,2300	337000	5667700	0,2200	365000	5667700	0,2210			
309000	5675300	0,2390	337000	5675300	0,3020	365000	5675300	0,2990			
309000	5682900	0,1950	337000	5682900	0,2070	365000	5682900	0,2200			
309000	5690500	0,1840	337000	5690500	0,3120	365000	5690500	0,2800			
316000	5614500	0,1130	344000	5614500	0,3240						
316000	5622100	0,1580	344000	5622100	0,2150						
316000	5629700	0,1150	344000	5629700	0,2610						
316000	5637300	0,8510	344000	5637300	0,1770						
316000	5644900	0,2890	344000	5644900	0,3690						
316000	5652500	0,2190	344000	5652500	0,2320						
316000	5660100	0,3430	344000	5660100	0,1590						
316000	5667700	0,2020	344000	5667700	0,3010						
316000	5675300	0,2350	344000	5675300	0,3420						
316000	5682900	0,3170	344000	5682900	0,3190						
316000	5690500	0,2670	344000	5690500	0,2200						

Anhang C CORINE Land Cover Daten

Tab. 18 Entwicklung der CORINE Land Cover Daten, Copernicus, Stand 02/2020²³

	CLC1990	CLC2000	CLC2006	CLC2012	CLC2018
Satellite data	Landsat-5 MSS/TM single date	Landsat-7 ETM single date	SPOT-4/5 and IRS P6 LISS III dual date	IRS P6 LISS III and RapidEye dual date	Sentinel-2 and Landsat-8 for gap filling
Time consistency	1986-1998	2000 +/- 1 year	2006 +/- 1 year	2011-2012	2017-2018
Geometric accuracy, satellite data	≤ 50 m	≤ 25 m	≤ 25 m	≤ 25 m	≤ 10 m (Sentinel-2)
Min. mapping unit/width	25 ha / 100m	25 ha / 100m	25 ha / 100m	25 ha / 100m	25 ha / 100 m
Geometric accuracy, CLC	100 m	better than 100 m	better than 100 m	better than 100 m	better than 100 m
Thematic accuracy, CLC	≥ 85% (probably not achieved)	≥ 85% (achieved) [13]	≥ 85%	≥ 85% (probably achieved)	≥ 85%
Change mapping (CHA)	not implemented	boundary displacement min. 100 m; change area for existing polygons ≥ 5 ha; for isolated changes ≥ 25 ha	boundary displacement min. 100 m; all changes ≥ 5 ha are to be mapped	boundary displacement min. 100 m; all changes ≥ 5 ha are to be mapped	boundary displacement min. 100 m; all changes ≥ 5 ha are to be mapped
Thematic accuracy, CHA	-	not checked	≥ 85% (achieved)	≥ 85%	≥ 85%
Production time	10 years	4 years	3 years	2 years	1.5 years
Documentation	incomplete metadata	standard metadata	standard metadata	standard metadata	standard metadata
Access to the data (CLC, CHA)	unclear dissemination policy	dissemination policy agreed from the start	free access for all users	free access for all users	free access for all users
Number of countries involved	26 (27 with late implementation)	30 (35 with late implementation)	38	39	39

Tab. 19 Corine Land Cover (CLC) Klassen (Level 1- 3)

Level 1	Level 2	Level 3
1 Artificial surfaces	11 Urban fabric	111 Continuous urban fabric
		112 Discontinuous urban fabric
	12 Industrial, commercial and transport units	121 Industrial or commercial units
		122 Road and rail networks and associated land
		123 Port areas
		124 Airports
	13 Mine, dump and construction sites	131 Mineral extraction sites
		132 Dump sites
		133 Construction sites
	14 Artificial, non-agricultural vegetated areas	141 Green urban areas
142 Sport and leisure facilities		
2 Agricultural areas	21 Arable land	211 Non-irrigated arable land
		212 Permanently irrigated land
		213 Rice fields
	22 Permanent crops	221 Vineyards
		222 Fruit trees and berry plantations

²³ Quelle: [Copernicus Land Monitoring Service: CORINE Land Cover](#)

		223 Olive groves
	23 Pastures	231 Pastures
	24 Heterogeneous agricultural areas	241 Annual crops associated with permanent crops
		242 Complex cultivation patterns
		243 Land principally occupied by agriculture, with significant areas of natural vegetation
		244 Agro-forestry areas
3 Forest and semi natural areas	31 Forests	311 Broad-leaved forest
		312 Coniferous forest
		313 Mixed forest
	32 Scrub and/or herbaceous vegetation associations	321 Natural grasslands
		322 Moors and heathland
		323 Sclerophyllous vegetation
		324 Transitional woodland-shrub
	33 Open spaces with little or no vegetation	331 Beaches, dunes, sands
		332 Bare rocks
		333 Sparsely vegetated areas
		334 Burnt areas
		335 Glaciers and perpetual snow
	4 Wetlands	41 Inland wetlands
412 Peat bogs		
42 Maritime wetlands		421 Salt marshes
		422 Salines
		423 Intertidal flats
5 Water bodies		51 Inland waters
	512 Water bodies	
	52 Marine waters	521 Coastal lagoons
		522 Estuaries
		523 Sea and ocean

Anhang D Erweitertes RIO-Modell

Landnutzungsindikator β :

$$\beta = \log \left[1 + \frac{\sum_i a_i \times n_{RCLi}}{\sum_i n_{RCLi}} \right] \quad (\text{Janssen et al., 2008})$$

Der Index i läuft über alle RIO-Klassen im gebildeten Puffer um die Messstation. n_{RCLi} ist die Anzahl der Pixel der Klasse i innerhalb des Puffers und a_i ist ein auf die

Schadstoffemission bezogener Koeffizient für jede RIO-Klasse und jeden der drei Luftschadstoffe. Der β -Parameter wird aus dem Logarithmus der gewichteten und normalisierten Summe der RIO-Klassenverteilung unter Berücksichtigung eines klassenspezifischen Verschmutzungsfaktors im Puffer gebildet. Er gibt Auskunft über die Landnutzungscharakteristika in unmittelbarer Nähe der Messstation und wird genutzt, um diesen Ort zu klassifizieren (Janssen et al., 2008).

Zur vereinfachten Ermittlung des Koeffizienten a_i für jede RIO-Klasse, der die Bedeutung einer bestimmten Klasse an der Konzentration des Luftschadstoffes gewichtet, werden im ersten Schritt folgende Annahmen zugrunde gelegt (Janssen et al., 2008):

- Landnutzungsklassen wie Wald oder Gewässer (RCL10+11) beinhalten keine Emittenten und haben daher keinen Einfluss auf die Konzentration eines Luftschadstoffes. Die Koeffizienten für beide Klassen können auf null gesetzt werden.
- Die RIO-Klasse 2 (nicht durchgängige städtische Prägung) kann als Klasse angesehen werden, die einen wesentlichen Einfluss auf die Konzentration eines Luftschadstoffes hat. Aus Gründen der Normalisierung wird der Koeffizient a_2 auf 1 gesetzt.

Damit müssen noch für 8 der 11 RIO-Klassen die Koeffizienten a_i bestimmt werden, die in einem numerischen Verfahren optimiert werden. Eine detaillierte Analyse des Verfahrens zeigt, dass die bestmögliche Optimierung wesentlich von der Wahl der Initialwerte für die Koeffizienten a_i abhängt (Janssen et al., 2008). Basierend auf der initialen Festlegung der Koeffizienten a_i kann eine erste Schätzung des β -Parameters für jede Luftmessstation vorgenommen werden.

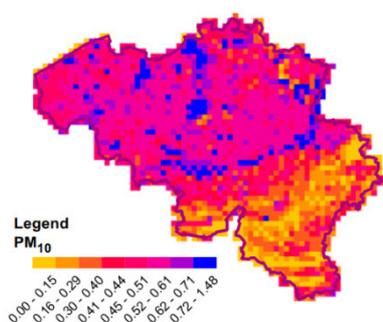
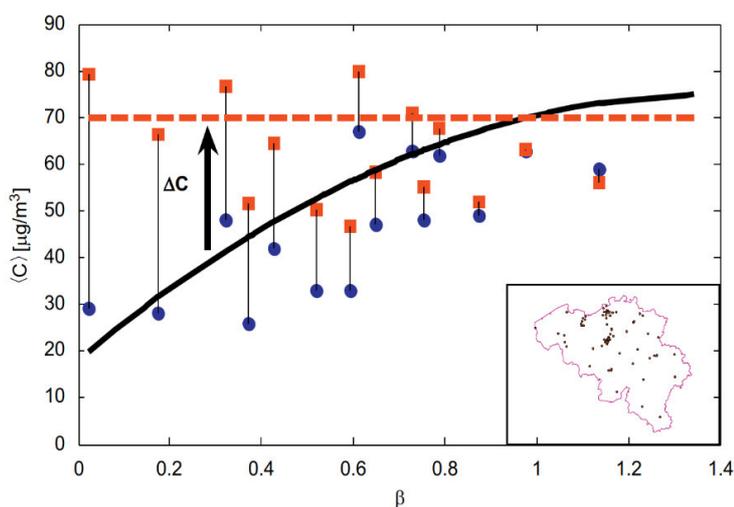


Abb. 68 β -Landnutzungsindikatorkarte

Trendbereinigung:

Die Trendbereinigung des Mittelwertes erfolgte mit einer simplen linearen Translation. Entsprechend des β -Wertes der Messstation wurde eine Korrekturverschiebung ΔC aus der Differenz eines willkürlichen Referenzniveaus und der Trendfunktion bestimmt. Zur Trendbereinigung wurde dann das ΔC zum Messwert hinzugefügt (Abb. 69).



In nebenstehender Abbildung repräsentieren die Punkte die Messwerte, die Quadrate die trendbereinigten Werte, die durchgezogene Linie die Trendfunktion und die gestrichelte Linie das Referenzniveau.

Abb. 69 Trendbereinigung Mittelwert am Beispiel von NO₂, (Jansen et al., 2008)

Liegt die Trendfunktion unterhalb des Referenzniveaus, wird ΔC zum Messwert addiert und im umgekehrten Fall subtrahiert.

Die Bereinigung des Trends für die Standardabweichung der Zeitreihe erfolgt nach folgender Formel (Jansen et al., 2008):

$$x_{\text{det}} = (x - \langle x \rangle) \frac{\sigma_{\text{ref}}}{\sigma(\beta)} + \langle x \rangle$$

Dabei ist σ_{ref} das Referenzniveau der Trendbereinigung, $\sigma(\beta)$ die theoretische Standardabweichung entsprechend des β -Wertes der Messstation und $\langle x \rangle$ der Konzentrationsmittelwert über die gesamte Zeitreihe.

Validierung:

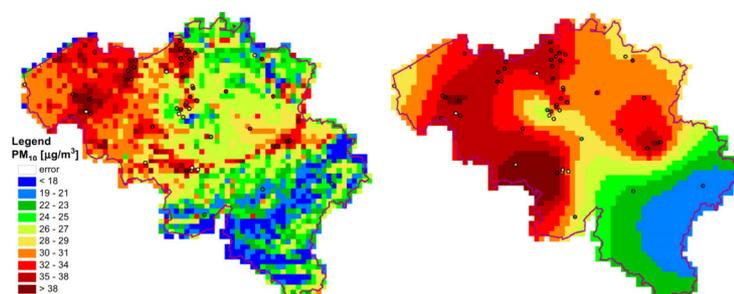


Abb. 70 Karte der PM₁₀-Jahresmittelwerte 2006 (RIO/OK), (Janssen et al., 2008)

Die Karten in Abb. 70 zeigen den Jahresmittelwert der PM₁₀-Konzentration in Belgien für das Jahr 2006. Der Wert für jeden Pixel der Karte wurde aus den Interpolationen aller verfügbaren stündlichen Messungen im Jahr 2006, deren Ergebnisse anschließend über das Jahr gemittelt werden, bestimmt.

Da die Messstationen für die PM₁₀-Messung im Jahr 2006 meist in Emittentennähe (Urbane Bereiche, Industrie und Verkehr) und nur vereinzelt im ländlichen Raum lagen, wird beim Standard-OK (Abb. 70, rechte Karte) eine hohe Belastung auch zwischen den Messpunkten ermittelt, was zu einer deutlichen Verfälschung des Ergebnisses führt (Janssen et al., 2008). Bei der mit der RIO-Methode erstellten Karte (Abb. 70, linke Karte) zeigt sich ein anderes Bild. Hier sind hohe Konzentrationen tatsächlich nur auf urbane Gebiete beschränkt. In ländlichen Gebieten finden sich die erwartbaren niedrigeren Werte.

Anhang E Visual Basic- und Python-Skripte

VB-Skriptcode für die Zuweisung eines RCL-Codes für jede CLC-Klasse (Tab. 6) - ArcMap Feldberechnung:

```

select case [Code_18]
  case 111
    rioclass = 1
  case 112,141,142
    rioclass = 2
  case 121
    rioclass = 3
  case 122
    rioclass = 4
  case 123
    rioclass = 5
  case 124
    rioclass = 6
  case 131,132,133
    rioclass = 7
  case 211,212,213
    rioclass = 8
  case 221,222,223,231,241,242,243,244
    rioclass = 9
  case 311,312,313,321,322,323,324,331,332,333,334,335
    rioclass = 10
  case else
    rioclass = 11
end select

RCL = rioclass

```

VB-Skriptcode für die Zuweisung eines Verschmutzungskoeffizienten a für jeden RCL-Code - ArcMap Feldberechnung:

```

select case [RCL]
  case 1
    rcla = 1.16
  case 2
    rcla = 1.00
  case 3
    rcla = 2.07
  case 4
    rcla = 2.23
  case 5
    rcla = 2.65
  case 6
    rcla = 1.01
  case 7
    rcla = 10.99
  case 8
    rcla = 0.64
  case 9
    rcla = 0.64
  case else
    rcla = 0.00
end select

Coeff_a = rcla

```

Python-Skript *AdditionalPointsLNI.py*

```
#####
#####
#script_name      :AdditionalPointsLNI.py
#description      :Generierung eines Punkterasters und Berechnung des RIO-Landnutzungsindikators (LNI) für jeden zusaetzlichen
Punkt
#
#                Ergaenzung der LNI-Werte der PM10-Messstationen - Optimierung der sekundaeren Variablen fuer Co-Kriging
#author          :Heinz-Juergen Jansen
#date            :20210112
#python_version  :2.7.15
#####
#####

# Import Python-Module
#
import arcpy, shapefile, csv, math
from arcpy import env

# Variablendefinition
#
# UTM32-Ursprungskordinaten im Untersuchungsgebiet
e32UTM = 295000
n32UTM = 5614500

# Raeumliche Ausdehnung des Untersuchungsgebietes
xExpansion = 70000
yExpansion = 76000

# ArcMap-Umgebung
dataDir = "D:/UNIGIS/UNIGIS MSc/Master Thesis/Daten/"
MD = arcpy.mapping.MapDocument(dataDir + 'Master Thesis.mxd')
AD = MD.activeDataframe
env.workspace = dataDir + "Master_Thesis.gdb"

# Hauptprogramm
#
print("Skript zum Erstellen eines Punkterasters zur Berechnung zusätzliche LNI-Werte im Untersuchungsgebiet.\n")

# Festsetzen der Rastergröße
#
xNumber = 11 # Anzahl der Punkte in x-Richtung
yNumber = 11 # Anzahl der Punkte in y-Richtung
numberOfPoints = xNumber * yNumber
print("Rastergröße: " + str(xNumber) + "*" + str(yNumber) + " Punkte")

# Abstand Punkte in x-/y-Richtung berechnen
#
xPointDist = int(round(xExpansion / (xNumber - 1)))
yPointDist = int(round(yExpansion / (yNumber - 1)))
print("Abstand der Punkte in x-Richtung (m):" + str(xPointDist))
print("Abstand der Punkte in y-Richtung (m):" + str(yPointDist) + "\n")

# csv-Datei für Verarbeitung in ArcMap generieren
#
print("csv-File generieren:"),
csvFile = open(dataDir + "add_LNI_points.csv","w")
#csvFile.write('Point_ID,E32_UTM,N32_UTM\n') # Header, Verwendung nur beim Einlesen von xy-Daten
i = 1

while yNumber > 0:
    xNumberCount = xNumber
    e32UTM0 = e32UTM
    while xNumberCount > 0:
        csvFile.write("AP" + str(i) + "," + str(e32UTM0) + "," + str(n32UTM) + "\n")
        i += 1
        e32UTM0 += xPointDist
        xNumberCount -= 1
        n32UTM += yPointDist
        yNumber -= 1

csvFile.close()
print("OK\n")

# Shapefile aus csv-Datei generieren
#
print("Shapefile aus csv-File generieren:"),
shpFile = dataDir + "add_LNI_points.shp"
pointID,e32UTM,n32UTM = [],[],[]
# Inhalt csv-Datei in Listen auslesen
with open(dataDir + "add_LNI_points.csv", "rb") as csvfile:
    csvRead = csv.reader(csvfile, delimiter = ',')
```

```

for j,row in enumerate(csvRead):
    pointID.append(row[0])
    e32UTM.append(int(row[1]))
    n32UTM.append(int(row[2]))

# Shapefile erzeugen und füllen
wSHP = shapefile.Writer(shpFile)
wSHP.autoBalance = 1
wSHP.field("POINT_ID","C")
wSHP.field("E32_UTM","N")
wSHP.field("N32_UTM","N")
k = 0

while k < numberOfPoints:
    wSHP.point(e32UTM[k], n32UTM[k])
    wSHP.record(pointID[k], e32UTM[k], n32UTM[k])
    k += 1

wSHP.close()
print("OK\n")

# Features aus Shapefile in GDB-Feature-Class importieren
#
print("Features aus Shapefile in GDB-Feature-Class importieren:"),
sr = arcpy.SpatialReference(25832)
arcpy.DefineProjection_management(shpFile, sr) # Projektion festlegen
alnipFC = "add_LNI_points"
outDir = dataDir + "Master_Thesis.gdb"
arcpy.FeatureClassToGeodatabase_conversion(shpFile, outDir)

# FeatureClass als Layer hinzufügen
alnipShpLayer = arcpy.mapping.Layer(alnipFC)
arcpy.mapping.AddLayer(AD, alnipShpLayer, "TOP")
allLayersMT = arcpy.mapping.ListLayers(MD,None,AD)
arcpy.RefreshTOC()
arcpy.RefreshActiveView()
print("OK\n")

# Buffer (2000 m) fuer jeden Punkt erzeugen
#
print("Buffer generieren:"),
bufferFC = "add_LNI_points_Buffer"
arcpy.Buffer_analysis(in_features="add_LNI_points", out_feature_class=bufferFC, buffer_distance_or_field="2000 Meters",
line_side="FULL", line_end_type="ROUND", dissolve_option="NONE", dissolve_field="", method="PLANAR")
print("OK\n")

# RCL-Features im Untersuchungsgebiet zusammenfügen
#
print("Features aus Layer 'Untersuchungsgebiet_CLC2018_RCL' zusammenfügen:"),
arcpy.Dissolve_management(in_features="Untersuchungsgebiet_CLC2018_RCL",
out_feature_class="Untersuchungsgebiet_CLC2018_RCL_DIS", dissolve_field="RCL;Coeff_a", statistics_fields="",
multi_part="MULTI_PART", unsplit_lines="DISSOLVE_LINES")
print("OK\n")

# Splitten der Bufferzonen in RCL-FeatureClass
#
print("Buffer mit Layer 'Untersuchungsgebiet_CLC2018_RCL' verschneiden und splitten:"),
arcpy.Split_analysis(in_features="Untersuchungsgebiet_CLC2018_RCL_DIS", split_features=bufferFC, split_field="POINT_ID",
out_workspace=outDir, cluster_tolerance="")
print("OK\n")

# RIO-Landnutzungsindikator(LNI) fuer jeden Punkt berechnen
#
print("LNI für jeden Punkt berechnen:"),
sum_a_area, sum_area = [],[]
inFCField = "Prod_a_area"
l = 0

while l < numberOfPoints:
    inFC = "AP" + str(l + 1)
    arcpy.AddField_management(inFC, inFCField, "Double")
    arcpy.CalculateField_management(inFC, inFCField, "!Coeff_a! * !Shape_Area!", "PYTHON_9.3")
    sum_aa = 0
    sum_a = 0
    rows = arcpy.SearchCursor(inFC)

    for m in rows:
        sum_aa = m.getValue("Prod_a_area") + sum_aa
        sum_a = m.getValue("Shape_Area") + sum_a

```

```

sum_a_area.append(sum_aa)
sum_area.append(sum_a)
del m, rows, sum_aa, sum_a
l += 1

fnameLNI = "RIO_LNI"
arcpy.AddField_management(bufferFC, fnameLNI, "Double")
pointLNI = []
n = 0
rows = arcpy.UpdateCursor(bufferFC)

for o in rows:
    pointLNI.append(math.log10(1 + (sum_a_area[n]/sum_area[n])))
    o.setValue(fnameLNI, pointLNI[n])
    rows.updateRow(o)
    n += 1

del rows
print("OK\n")

# Neues csv-File inkl. LNI fuer Aufbereitung als GSLIB-Datei generieren
#
print("Neue csv-File inkl. LNI generieren:");
with open(dataDir + "add_LNI_points.csv", "r") as csvInFile, \
    open(dataDir + "add_LNI_points_GSLIB_ready.csv", "wb") as csvOutFile:
    csvRead = csv.reader(csvInFile)
    csvWrite = csv.writer(csvOutFile)
    p = 0
    newVal = "-9966699"

    for q in csvRead:
        t10Val = 2000
        t12Val = 2000
        newVal2 = str(round(pointLNI[p], 3))
        newValLNI = newVal2.replace(",", ".")
        q.append(str(t10Val))
        q.append(str(t12Val))
        q.append(newVal)
        q.append(newValLNI)
        while t10Val <= 72000:
            csvWrite.writerow(q[1:])
            t10Val += 10000
            t12Val += 12000
            q[3] = t10Val
            q[4] = t12Val
            p += 1

print("OK")

# Map-Dokument speichern
#
MD.save()

```

Python-Skript *PM10DetrendDataPreparation.py*

```

#####
#script_name      :PM10DetrendDataPreparation.py
#description      :Determinate grid cells from Voxler grid (IDW Power 0) which match the locations of the PM10 sample points
#author          :Heinz-Juergen Jansen
#date            :20210425
#python_version  :2.7.15
#####

# Import Python-Module
#
import csv

# Variablendefinition
#
dataDir = "D:/UNIGIS/Master_Thesis/Daten/Voxler/"
inFile = "PM10_GT_IDW_Power0_SR475.csv"
outFile = "PM10_GT_IDW_Power0_Values.csv" # Ausgabedatei im csv-Format zur weiteren Verarbeitung in Excel; manueller Abgleich
Messpunkt-Koordinaten - Voxler-Rasterzellen
noDataValue = -999

# Selektion von Rasterzellen aus einem Voxler-Raster (IDW Power 0, optimaler Suchradius 475m), denen ein Wert zugeordnet wurde
und doppelte Einträge herausfiltern
#

```

```

print("Aus csv-File " + inFile + " Zeilen mit Eintraegen für Gauss-transformierte PM10-Werte selektieren und Doppelte herausfiltern:"),
with open(dataDir + inFile, "r") as csvInFile, \
    open(dataDir + outFile, "wb") as csvOutFile:
    csvRead = csv.reader(csvInFile, delimiter=',')
    csvWrite = csv.writer(csvOutFile)
    valuePM10GTList = []
    for i in csvRead:
        valuePM10GT = float(i[3])
        if valuePM10GT != 0:
            if valuePM10GT not in valuePM10GTList:
                csvWrite.writerow(i)
                valuePM10GTList.append(valuePM10GT)
print("OK")

```

Python-Skript *PM10DetrendData.py*

```

#####
#script_name      :PM10DetrendData.py
#description      :Detrending of the Gauss-transformed PM10 data by subtraction a detrend value determined from related IDW Power
1 data (Voxler)
#author          :Heinz-Juergen Jansen
#date            :20210701
#python_version  :2.7.15
#####

# Import Python-Module
#
import csv

# Variablendefinition
#
dataDir = " D:/UNIGIS/Master_Thesis/Daten/"
dataDirV = dataDir + "Voxler/"
dataDirS = dataDir + "SGeMS/"
inFile1 = "PM10_IDW_Power1.csv" # IDW Power 1 Raster - Export aus Voxler
inFile2 = "Grid_Points_Voxler_Sample_Points.csv" # Zuordnung Messpunkte - Voxler IDW Raster
inFile3 = "PM10_Gauss_Trans.csv" # Gauss-transformierte PM10-Daten - Export aus SGeMS
dataFile = "Values_for_Detrending.csv"
outFile = "PM10_Detrended_Data_SGEMS.csv"
outFile2 = "PM10_Detrended_Data_SGEMS.gslib" # Ausgabedatei im GSLIB-Format mit den trendbereinigten PM10-Werten
noDataValue = -999

# Dictionaries zur Speicherung der Eingabewerte definieren
file1Dict = dict()
file2Dict = dict()

# Werte fuer die Trendbereinigung aus Exportdatei der IDW Power 1 Interpolation (Voxler-Raster) selektieren
#
print("Werte fuer die Trendbereinigung aus Voxler-Exportdatei " + inFile1 + " selektieren: "),

# Input-File inFile1 einlesen
with open(dataDirV + inFile1, "r") as csvInFile1:
    csvRead1 = csv.reader(csvInFile1, delimiter=',')
    #csvRead1.next() # Kopfzeile ueberspringen
    for i in csvRead1:
        valueUTME32GCF1 = str(round(float(i[0]), 3)) # GC = Grid Cell
        valueUTMN32GCF1 = str(round(float(i[1]), 3))
        valueDateGCF1 = str(round(float(i[2]), 3))
        # Dictionary-Schluesssel fuer den Vergleich mit inFile2 generieren
        file1key = valueUTME32GCF1 + '_' + \
            valueUTMN32GCF1 + '_' + \
            valueDateGCF1
        # Speichern der Werte in file1Dict
        file1Dict[file1key] = i

# Input-file inFile2 einlesen
with open(dataDirV + inFile2, "r") as csvInFile2:
    csvRead2 = csv.reader(csvInFile2, delimiter=',')
    csvRead2.next() # Kopfzeile ueberspringen
    for j in csvRead2:
        valueUTME32GCF2 = str(round(float(j[0]), 3)) # GC = Grid Cell
        valueUTMN32GCF2 = str(round(float(j[1]), 3))
        valueDateGCF2 = str(round(float(j[2]), 3))
        # Dictionary-Schluesssel fuer den Vergleich mit inFile1 generieren
        file2key = valueUTME32GCF2 + '_' + \
            valueUTMN32GCF2 + '_' + \
            valueDateGCF2
        # Speichern der Werte in file2Dict
        file2Dict[file2key] = [int(j[3]), int(j[4]), int(j[5]),
            int(j[6])]

```

```

# Vergleich der Dictionary-Schlüssel von file1Dict und file2Dict
with open(dataDirV + dataFile, "wb") as csvOutFile1:
    for key in sorted(file1Dict.keys()):
        csvWrite1 = csv.writer(csvOutFile1)
        # Pruefung, ob Schlüssel in file2Dict vorhanden ist
        if key in file2Dict:
            #print(inFile2 + " also contains value " + key) # Optionale Ausgabe der gefundenen Rasterzell-Koordinaten
            # Speichern der selektierten Werte fuer die Trendbereinigung
            file1Dict[key].extend(file2Dict[key])
            csvWrite1.writerow(file1Dict[key])

print("OK\n")

# Trendbereinigung der Gauss-transformierten PM10-Messwerte + Generierung einer csv-Datei )
#
print("Trendbereinigung der Gauss-transformierten PM10-Messwerte und Erzeugen einer csv-Datei: "),
with open(dataDirS + inFile3, "r") as csvInFile3, \
    open(dataDirS + outFile, "wb") as csvOutFile2:
    csvRead3 = csv.reader(csvInFile3, delimiter=',')
    csvWrite2 = csv.writer(csvOutFile2, delimiter=',')
    csvRead3.next()

    for k in csvRead3:
        valuePM10GT = float(k[3])
        if valuePM10GT == noDataValue:
            k.append(noDataValue)
            csvWrite2.writerow(k)
        else:
            with open(dataDirV + dataFile, "r") as csvInFile4:
                csvRead4 = csv.reader(csvInFile4, delimiter=',')
                for l in csvRead4:
                    if k[0] == l[4] and k[1] == l[5] and k[2] == l[6]:
                        #print(str(k[0]) + '==' + str(l[4]) + ' and ' + str(k[1]) + '==' + str(l[5]) + ' and ' + str(k[2]) + '==' + str(l[6]))
                        #valueDetData = round(float(l[3]), 5) # Inkl. Anpassung auf Nachkommastellen der PM10_GT-Werte
                        valueDetData = float(l[3])
                        valuePM10GTDet = round(valuePM10GT - valueDetData, 6)
                        k.append(valuePM10GTDet)
                        csvWrite2.writerow(k[0:])

print("OK\n")

# Aus zuvor generierter csv-Datei zusaetzlich eine GSLIB-Datei erstellen als Basis fuer den Import in SGeMS
#
print("Aus zuvor generierter csv-Datei eine GSLIB-Datei erstellen als Basis fuer den Import in SGeMS: "),

#Header fuer GSLIB-Datei erstellen
gslibHeader = [
    "PM10_24H 01.09. - 08.09.2018 Gauss-transformiert und trendbereinigt\n",
    "5\n",
    "E32_UTM\n",
    "N32_UTM\n",
    "t_10000\n",
    "PM10_Gauss_Trans\n",
    "PM10_Gauss_Trans_DeTrend\n"
]
headerFileName = "GSLIB_Header"
headerFile = file(dataDirS + headerFileName, "w")
headerFile.writelines(gslibHeader)
headerFile.close()

# Zusammenfuegen GSLIB-Header + csv-Datei
gslibFile = open(dataDirS + outFile2, "a")
for m in open(dataDirS + headerFileName):
    gslibFile.write(m)
for n in open(dataDirS + outFile):
    gslibFile.write(n)
gslibFile.close()

print("OK")

```

Python-Skript *DataSplitLNI.py*

```

#####
#script_name      :DataSplitLNI.py
#description      :Limit value range of Gauss-transformed LNI data (SGeMS export) to exclude extreme values from Cokriging and add
new data to
#                 csv-file with Gauss-transformed LNI data as a new column.
#author           :Heinz-Juergen Jansen
#date             :20210809
#python_version   :2.7.15
#####

```

```

# Import Python-Module
#
import csv, shutil

# Variablendefinition
#
dataDir = "D:/UNIGIS/Master_Thesis/Daten/SGEMS/"
inFile = "LNI_Gauss_Trans.csv"
copyFile = "LNI_Gauss_Trans_Range.csv"
noDataValue = -999

# Begrenzung des LNI-Wertebereichs
#
print("Definition des Landnutzungsindikator-Wertebereichs, der fuer das CoKriging genutzt werden soll (Kappung der Extremwerte):\n")
LNImin = input("LNI-Wert Minimum: ")
LNImax = input("LNI-Wert Maximum: ")
print("")
outFile = "LNI_Gauss_Trans_Range_" + str(LNImin).replace(".", "_") + "-" + str(LNImax).replace(".", "_") + ".csv"

# Wertebereich der LNI-Werte einschaerken und Ergebnis als neue Spalte an die Ausgangsdaten anhaengen
#
print("Wertebereich der LNI-Werte einschaerken und Ausgangsdaten aus Datei " + inFile + " um die neuen Dateien erweitern: "),
with open(dataDir + inFile, "r") as csvInFile, \
    open(dataDir + outFile, "wb") as csvOutFile:
    csvRead = csv.reader(csvInFile, delimiter=',')
    csvWrite = csv.writer(csvOutFile)

    headerRow = next(csvRead)
    newColumn = "LNI_Gauss_Trans_WEV" # WEV = without extrem values
    headerRow.append(newColumn)
    csvWrite.writerow(headerRow)
    count = 0

    for i in csvRead:
        valueLNI = float(i[3])
        if valueLNI > LNImin and valueLNI < LNImax:
            i.append(valueLNI)
            csvWrite.writerow(i)
        else:
            i.append(noDataValue) # Zuweisung des No-Data-Values fuer Punkte ohne LNI-Wert
            csvWrite.writerow(i)
        count += 1
print("OK\n")

# Genenrierte csv-Datei kopieren. Eine csv-Datei ohne Angabe des Wertebereichs im Dateinamen wird als Input-File fuer das Skript
PM10DetrendedDataAddLNI.py benötigt
#
print("csv-File " + outFile + " nach " + copyFile + " kopieren: "),
shutil.copyfile(dataDir + outFile, dataDir + copyFile)
print("OK\n")

# Anzahl der NoData Values und Punkte mit NoData Values nach Einschraenkung des Wertebereichs
#
print("Anzahl LNI No-Data-Values: " + str(count))
print("Anzahl LNI-Punkte mit No-Data-Values: " + str(count/8)) # Division durch die Anzahl der Untersuchungstage

```

Python-Skript *PM10DetrendedDataAddLNI.py*

```

#####
#script_name      :PM10DetrendedDataAddLNI.py
#description      :Add Gauss-transformed LNI data and Gauss-transformed LNI data with limited value range to detrended PM10 data
#author          :Heinz-Juergen Jansen
#date            :20210804
#python_version  :2.7.15
#####

# Import Python-Module
#
import csv

# Variablendefinition
#
dataDir = "D:/UNIGIS/Master_Thesis/Daten/SGEMS/"
inFile1 = "PM10_Detrended_Data_SGEMS.csv"

```

```

inFile2 = "LNI_Gauss_Trans_Range.csv"
outFile1 = "PM10_Detrended_Data_LNI_SGEMS.csv"
outFile2 = "PM10_Detrended_Data_LNI_SGEMS.gslib"

# Dictionaries zur Speicherung der Eingabewerte definieren
file1Dict = dict()
file2Dict = dict()

# Gauss-transformierte und zusätzlich trendbereinigte PM10-Daten mit Gauss-transformierten LNI-Daten (mit und ohne Extrema)
zusammenfuegen
#
print("Dateien " + inFile1 + " und " + inFile2 + " in Dictionaries einlesen und in neuer csv-Datei " + outFile1 + " zusammenfuegen: "),

# Input-File inFile1 einlesen
with open(dataDir + inFile1, "r") as csvInFile1:
    csvRead1 = csv.reader(csvInFile1, delimiter=',')
    #csvRead1.next() # Kopfzeile ueberspringen
    for i in csvRead1:
        valueUTME32SPFile1 = str(i[0]) # SP = Sample Point
        valueUTMN32SPFile1 = str(i[1])
        valueDateSPFile1 = str(i[2])
        # Dictionary-Schlüssel fuer den Vergleich mit inFile2 generieren
        file1key = valueUTME32SPFile1 + '_' + \
            valueUTMN32SPFile1 + '_' + \
            valueDateSPFile1
        # Speichern der Werte in file1Dict
        file1Dict[file1key] = i # Einlesen der vollstaendigen Zeile in file1Dict

# Input-File inFile2 einlesen
with open(dataDir + inFile2, "r") as csvInFile2:
    csvRead2 = csv.reader(csvInFile2, delimiter=',')
    csvRead2.next() # Kopfzeile ueberspringen
    for j in csvRead2:
        valueUTME32APFile2 = str(int(j[0])) # AP = Additional Point
        valueUTMN32APFile2 = str(int(j[1]))
        valueDateAPFile2 = str(int(j[2]))
        # Dictionary-Schlüssel fuer den Vergleich mit inFile1 generieren
        file2key = valueUTME32APFile2 + '_' + \
            valueUTMN32APFile2 + '_' + \
            valueDateAPFile2
        # Speichern der Werte in file2Dict
        file2Dict[file2key] = [float(j[3]),float(j[4])] # Nur Einlesen der Spalten LNI_Gauss_Trans und LNI_Gauss_Trans_WEV aus inFile2

# Vergleich der Dictionary-Schlüssel von file1Dict und file2Dict
with open(dataDir + outFile1, "wb") as csvOutFile1:
    for key in sorted(file1Dict.keys()):
        csvWrite1 = csv.writer(csvOutFile1, delimiter=',')
        # Pruefung, ob Schlüssel in file2Dict vorhanden ist
        if key in file2Dict:
            file1Dict[key].extend(file2Dict[key]) # Zusammenfuegen der PM10- und LNI-Daten
            csvWrite1.writerow(file1Dict[key]) # Speichern der zusammengefuegten Daten

print("OK\n")

# Aus zuvor generierter csv-Datei zusätzlich eine GSLIB-Datei erstellen als Basis fuer den Import in SGeMS
#
print("Aus zuvor generierter csv-Datei " + outFile1 + " die GSLIB-Datei " + outFile2 + " erstellen als Basis fuer den Import in SGeMS: "),

#Header fuer GSLIB-Datei erstellen
gslibHeader = [
    "PM10_24H 01.09. - 08.09.2018 Gauss-transformiert und trendbereinigt + LNI Gauss-transformiert und zusätzlich ohne
    Extrema\n",
    "7\n",
    "E32_UTM\n",
    "N32_UTM\n",
    "t_10000\n",
    "PM10_GT\n",
    "PM10_GT_DT\n",
    "LNI_GT\n",
    "LNI_GT_WEV\n" # WEV: without extreme values
]
headerFileName = "GSLIB_Header_PM10_DeTrend_LNI"
headerFile = file(dataDir + headerFileName, "w")
headerFile.writelines(gslibHeader)
headerFile.close()

# Zusammenfuegen GSLIB-Header + csv-Datei
gslibFile = open(dataDir + outFile2, "w")
for k in open(dataDir + headerFileName):
    gslibFile.write(k)
for l in open(dataDir + outFile1):
    l = l.replace("999.0", "999")
    gslibFile.write(l)
gslibFile.close()

```

```
print("OK")
```

Python-Skript *AddCoordinatesToSGeMSExport.py*

```
#####
#script_name      :AddCoordinatesToSGeMSExport.py
#description      :Add x/y/z-coordinates to SGeMS export file, since SGeMS grid export files do not include coordinates
#author          :Heinz-Juergen Jansen
#date            :20210812
#python_version  :2.7.15
#####

# Import Python-Module
#
import csv

# Variabledefinition
#
dataDir = "D:/UNIGIS/Master_Thesis/Daten/SGeMS/"
#inFile = "PM10_GT_DT_GT_SK_GRT.csv"
inFile = "PM10_GT_DT_GT_LNI_GT_SCoK_GRT.csv"
#outFile = "PM10_GT_DT_GT_SK_GRT_ADDCOORD.csv"
outFile = "PM10_GT_DT_GT_LNI_GT_SCoK_GRT_ADDCOORD.csv"
xMin = 293250
xMax = 366750
yMin = 5612750
yMax = 5692250
zMin = 250
zMax = 73750
cellSize = 500 # SGeMS Zellengroesse in x/y/z-Richtung
xCoord = [] # Liste zum Speichern der x-Koordinaten
yCoord = [] # Liste zum Speichern der y-Koordinaten
zCoord = [] # Liste zum Speichern der z-Koordinaten

# x/y/z-Koordinaten in getrennten Listen speichern
#
zCount = zMin
while zCount <= zMax:
    yCount = yMin
    while yCount <= yMax:
        xCount = xMin
        while xCount <= xMax:
            xCoord.append(xCount)
            yCoord.append(yCount)
            zCoord.append(zCount)
            xCount += cellSize
        yCount += cellSize
    zCount += cellSize

# x/y/z-Koordinaten zur SGeMS Exportdatei hinzufuegen
#
print("x/y/z-Koordinaten zum csv-File " + inFile + " hinzufuegen: "),
with open(dataDir + inFile, "r") as csvInFile, \
    open(dataDir + outFile, "wb") as csvOutFile:
    csvRead = csv.reader(csvInFile, delimiter=',')
    csvWrite = csv.writer(csvOutFile, delimiter=',')
    headerRow = next(csvRead)
    listCount = 0

    for i in csvRead:
        valuePM10RT = float(i[0]) # Wert fuer das Retrending
        i.remove(i[0]) # Löschen des ersten Wertes, damit die Koordinaten an den Zeilenanfang geschrieben werden koennen
        i.append(xCoord[listCount])
        i.append(yCoord[listCount])
        i.append(zCoord[listCount])
        i.append(valuePM10RT)
        csvWrite.writerow(i)
        listCount += 1

print("OK")
```

Python-Skript *PM10RetrendData.py*

```
#####
#script_name      :PM10RetrendData.py
#description      :Retrend the PM10 data by subtraction of a detrend value determined from related PM10 IDW Power 1 data
generated in Voxler
#author          :Heinz-Juergen Jansen
#date            :20210804
#python_version  :2.7.15
#####

# Import Python-Module
#
import csv

# Variablendefinition
#
dataDir = " D:/UNIGIS/Master_Thesis/Daten/"
dataDirV = dataDir + "Voxler/"
dataDirS = dataDir + "SGeMS/"
#inFile1 = "PM10_GT_DT_GT_SK_GRT_ADDCOORD.csv" # PM10 Daten nach Simple Kriging und erster Ruecktransformation (SGeMS
Export) und Hinzufuegen der Koordinaten
inFile1 = "PM10_GT_DT_GT_LNI_GT_SCoK_GRT_ADDCOORD.csv" # PM10 Daten nach CoKriging und erster Ruecktransformation
(SGeMS Export) und Hinzufuegen der Koordinaten
inFile2 = "PM10_IDW_Power1.csv" # PM10 IDW Power 1 Raster - Export aus Voxler
dataFile = "PM10_GT_DT_GT_SK_GRT_COORD_RT-DATA.csv"
#outFile1 = "PM10_GT_DT_GT_SK_GRT_RT.csv"
outFile1 = "M10_GT_DT_GT_LNI_GT_SCoK_GRT_RT.csv"
#outFile2 = "PM10_GT_DT_GT_SK_GRT_RT.gslib"
outFile2 = "PM10_GT_DT_GT_LNI_GT_SCoK_GRT_RT.gslib"

# Dictionaries zur Speicherung der Eingabewerte definieren
file1Dict = dict()
file2Dict = dict()

# Werte fuer das Retrending aus Exportdatei der IDW Power 1 Interpolation (Voxler) selektieren
#
print("Aus csv-File " + inFile1 + " die fuer das Retrending benoetigten Werte selektieren: "),

# Input-File inFile1 einlesen
with open(dataDirS + inFile1, "r") as csvInFile1:
    csvRead1 = csv.reader(csvInFile1, delimiter=',')
    #csvRead1.next() # Kopfzeile ueberspringen
    for i in csvRead1:
        valueUTME32GCF1 = str(i[0]) # GC = Grid Cell
        valueUTMN32GCF1 = str(i[1])
        valueDateGCF1 = str(i[2])
        # Dictionary-Schlüssel fuer den Vergleich mit inFile2 generieren
        file1key = valueUTME32GCF1 + '_' + \
            valueUTMN32GCF1 + '_' + \
            valueDateGCF1
        # Speichern der Werte in file1Dict
        file1Dict[file1key] = i

# Input-File inFile2 einlesen
with open(dataDirV + inFile2, "r") as csvInFile2:
    csvRead2 = csv.reader(csvInFile2, delimiter=',')
    #csvRead2.next() # Kopfzeile ueberspringen
    for j in csvRead2:
        valueUTME32GCF2 = str(int(j[0]) + 250) # GC = Grid Cell
        valueUTMN32GCF2 = str(int(j[1]) + 250)
        valueDateGCF2 = str(int(j[2]) + 250)
        # Dictionary-Schlüssel fuer den Vergleich mit inFile1 generieren
        file2key = valueUTME32GCF2 + '_' + \
            valueUTMN32GCF2 + '_' + \
            valueDateGCF2
        # Speichern der Werte in file2Dict
        file2Dict[file2key] = [float(j[3])] # Einlesen des Retrending-Wertes in file2Dict

# Vergleich der Dictionary-Schlüssel von file1Dict und file2Dict
with open(dataDirS + dataFile, "wb") as csvOutFile1:
    for key in sorted(file1Dict.keys()):
        csvWrite1 = csv.writer(csvOutFile1)
        # Pruefung, ob Schlüssel in file2Dict vorhanden ist
        if key in file2Dict:
            # Speichern der selektierten Retrending-Werte in csv-Datei fuer die weitere Verarbeitung
            file1Dict[key].extend(file2Dict[key])
            csvWrite1.writerow(file1Dict[key])

print("OK\n")
```

```

# Retrending der PM10-Werte nach Kriging und erster Ruecktransformation
#
print("Retrending der PM10-Messwerte und Erzeugen der csv-Datei " + str(outFile1) + ":",
with open(dataDirS + dataFile, "r") as csvInFile3, \
    open(dataDirS + outFile1, "wb") as csvOutFile2:
    csvRead3 = csv.reader(csvInFile3, delimiter=',')
    csvWrite2 = csv.writer(csvOutFile2, delimiter=' ')
    #csvRead3.next() # Kopfzeile ueberspringen
    for k in csvRead3:
        valuePM10SK = float(k[3]) # Simple Kriging Schaetzwert
        valuePM10Ret = float(k[4]) # Retrending-Wert (Voxler IDW Power 1)
        valuePM10SKRet = round(valuePM10SK + valuePM10Ret, 6) # Retrending der Schaetzwerte durch Addition des Retrending-
Wertes
        k.append(valuePM10SKRet)
        csvWrite2.writerow([k[0],k[1],k[2],k[5]])

print("OK\n")

# Aus zuvor generierter csv-Datei zusaetzlich eine GSLIB-Datei erstellen als Basis fuer den Import in SGeMS
#
print("Erstellen der GSLIB-Datei " + str(outFile2) + " aus zuvor generierter csv-Datei als Basis fuer den Import in SGeMS: "),

#Header fuer GSLIB-Datei erstellen
gslibHeader = [
    "PM10_24H 01.09. - 08.09.2018 Retrended\n",
    "4\n",
    "E32_UTM\n",
    "N32_UTM\n",
    "t_10000\n",
    #"PM10_GT_DT_GT_SK_GRT_RT\n" # Eintrag fuer Simple Kriging
    "PM10_GT_DT_GT_LNI_GT_SCoK_GRT_RT\n" # Eintrag fuer Simple CoKriging
]
headerFileName = "GSLIB_Header_RT"
headerFile = file(dataDirS + headerFileName, "w")
headerFile.writelines(gslibHeader)
headerFile.close()

# Zusammenfuegen GSLIB-Header + csv-Datei
gslibFile = open(dataDirS + outFile2, "w ")
for m in open(dataDirS + headerFileName):
    gslibFile.write(m)
for n in open(dataDirS + outFile1):
    gslibFile.write(n)
gslibFile.close()

print("OK")

```

Python-Skript *CalculateRMSE.py*

```

#####
#script_name      :CalculateRMSE.py
#description      :Calculate RMSE(Root Mean Square Error) for the interpolation result grid (after retransformations and retrending)
#author          :Heinz-Juergen Jansen
#date            :20210817
#python_version  :2.7.15
#####

# Import Python-Module
#
import csv

# Variablendefinition
#
dataDir = "D:/UNIGIS/Master_Thesis/Daten/SGeMS/"
inFile2 = "PM10_Measured_Value_To_Grid_Cell.csv" # Zuordnung PM10-Messwert zur Rasterzelle im Interpolationsrater
noDataValue = "-999.0"

# Dictionaries zur Speicherung der Eingabewerte definieren
file1Dict = dict()
file2Dict = dict()

# Werte aus csv-Dateien in Dictionaries einlesen
#
print("Root Mean Square Error(RMSE) fuer das Interpolationsergebnis berechnen: \n")

# Auswahl des Export-Rasters (SGeMS)
print("Auswahl des Interpolationsverfahrens: ")
inpoMe = input("1. Kriging | 2. CoKriging | 3. CoKriging mit LNI-Range: ")
if inpoMe == 1:

```

```

inFile1 = "PM10_GT_DT_GT_SK_GRT_RT_GRT.csv"
elif inpoMe == 2:
inFile1 = "PM10_GT_DT_GT_LNI_GT_SCoK_GRT_RT_GRT.csv"
elif inpoMe == 3:
inFile1 = "PM10_GT_DT_GT_LNI_GT_WEV_SCoK_GRT_RT_GRT.csv"
else:
print("Ungueltige Eingabe!")
quit()

# Input-File inFile1 einlesen
with open(dataDir + inFile1, "r") as csvInFile1:
csvRead1 = csv.reader(csvInFile1, delimiter=',')
csvRead1.next() # Kopfzeile ueberspringen
for i in csvRead1:
valueUTME32GCFile1 = str(i[0]) # GC = Grid Cell
valueUTMN32GCFile1 = str(i[1])
valueDateGCFile1 = str(i[2])
# Dictionary-Schluesssel fuer den Vergleich mit inFile2 generieren
file1key = valueUTME32GCFile1 + '_' + \
valueUTMN32GCFile1 + '_' + \
valueDateGCFile1
file1Dict[file1key] = [float(i[3])] # PM10-Schaetzwert am Messpunkt (Rasterzelle) in Dictionary file1Dict einlesen

# Input-File inFile2 einlesen
with open(dataDir + inFile2, "r") as csvInFile2:
csvRead2 = csv.reader(csvInFile2, delimiter=',')
csvRead2.next() # Kopfzeile ueberspringen
for j in csvRead2:
valueUTME32GCFile2 = str(j[0]) # GC = Grid Cell
valueUTMN32GCFile2 = str(j[1])
valueDateGCFile2 = str(j[2])
# Dictionary-Schluesssel fuer den Vergleich mit inFile1 generieren
file2key = valueUTME32GCFile2 + '_' + \
valueUTMN32GCFile2 + '_' + \
valueDateGCFile2
file2Dict[file2key] = [float(j[3])] # PM10-Messwert am Messpunkt (Rasterzelle) in Dictionary file2Dict einlesen

# RMSE berechnen
#
n = 0 # Zaehler fuer Messpunkte, an denen ein Messwert vorliegt
ndvc = 0 # Zaehler fuer Messpunkte mit No-Data-Value
PM10ResSQSum = 0 # Summe der Abweichquadrate
for key in sorted(file1Dict.keys()):
# Pruefung, ob Schluesssel in file2Dict vorhanden ist
if key in file2Dict:
# Speichern der selektierten Werte fuer das Retrending
PM10EstValue = str(file1Dict[key])[1:-1] # PM10EstValue: PM10 Estimated Value
PM10MValue = str(file2Dict[key])[1:-1] # PM10MValue: PM10 Measured Value
if PM10MValue != noDataValue:
PM10ResSQ = (round(float(PM10EstValue), 2) - float(PM10MValue))**2
PM10ResSQSum = PM10ResSQSum + PM10ResSQ
n += 1
print("Messpunkt: " + str(n))
print("Schaetzwert: " + str(round(float(PM10EstValue), 2)))
print("Messwert: " + PM10MValue + "\n")
else:
ndvc +=1
print("Anzahl No-Data-Values: " + str(ndvc) + "\n")
RMSE = round((PM10ResSQSum/n)**0.5, 2)
print("RMSE fuer Raster \'" + str(inFile1)[0:-4] + "\': " + str(RMSE) + "\n")

```

Anhang F Kriging- und KoKriging-Parameter

Simple Kriging PM₁₀:

The screenshot shows two panels of the software interface. The left panel is titled 'General and Data' and contains the following settings:

- Kriging Grid:** Grid: PM10_LNI_KG, Region: (empty), New Property Name: PM10_GT_DT_GT_SK, Simple Kriging (SK) selected, Mean: 0.
- Block Kriging:** Unchecked.
- Hard Data:** Object: Grid: PM10_24_LNI_GT_DT, Region: (empty), Property: PM10_GT_DT_GT.
- Search Ellipsoid:** Conditioning data: Min: 4, Max: 12. Ranges: Max: 54000, Med: 48000, Min: 12000. Azimuth: 0, Dip: 0, Rake: 0.

The right panel is titled 'General and Data Variogram' and contains the following settings:

- Load existing model... (button)
- Nugget Effect: 0.03
- Nb of Structures: 1
- Structure 1: Contribution: 1.15, Type: Spherical.
- Table:

	Max	Med	Min
Ranges	45360	38340	21060
Azimuth	0	0	0
Dip	0	0	0
Rake	0	0	0

Abb. 71 Simple Kriging PM₁₀ Parameter

Simple CoKriging PM₁₀/β:

The screenshot shows four panels of the software interface for Simple CoKriging PM₁₀/β parameter configuration.

Panel 1 (Left): 'General and Data' with 'Simple Kriging (SK)' selected and 'SK mean of Z1 and Z2' set to 0.0.0273964. 'Cokriging Type' is set to 'Full Cokriging'.

Panel 2 (Middle-Left): 'General and Data Variogram' with 'Assign Data to Kriging Grid' checked. It shows settings for 'Primary Hard Data Grid' (PM10_24_LNI_GT_DT) and 'Secondary Hard Data Grid' (LNI_GT). Search ellipsoid settings are provided for both primary and secondary variables.

Panel 3 (Middle-Right): 'General and Data Variogram' for 'Variogram for primary variable C11(h)'. It shows 'Structure 1' with Contribution: 1.15, Type: Spherical, and a table of ranges and angles.

Panel 4 (Right): 'Variogram of Secondary variable C22(h)'. It shows 'Structure 1' with Contribution: 0.865, Type: Exponential, and a table of ranges and angles.

Panel 5 (Bottom-Middle): 'Full Cokriging Parameters' for 'Cross-variogram C12(h)'. It shows 'Structure 1' with Contribution: 0.33, Type: Gaussian, and a table of ranges and angles.

Abb. 72 Simple CoKriging PM₁₀/β Parameter

Simple CoKriging PM10/ β -Range:

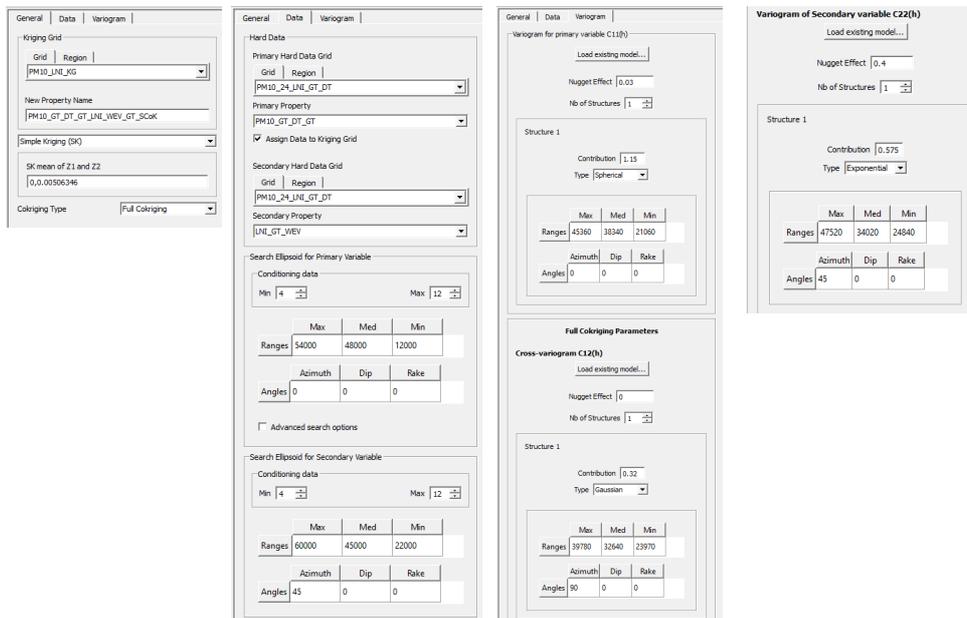


Abb. 73 Simple CoKriging PM10/ β -Range Parameter