

Eingereicht von Dipl.-Ing. Alexander Riha
Betreuung: Univ.-Prof. Dr. Andreas Koch
UNIGIS Salzburg, Zentrum für Geoinformatik
Universität Salzburg



Ein Vergleich räumlicher Interpolationsmethoden für Grundstückspreise im Bauland

Masterthesis

Eingereicht von Dipl.-Ing. Alexander Riha
Betreuung: Univ.-Prof. Dr. Andreas Koch
UNIGIS Salzburg, Zentrum für Geoinformatik
Universität Salzburg

DANKE

Herrn Univ.-Prof. Dr. Andreas KOCH für die Anregung, Betreuung und Möglichkeit
in seinem Fachgebiet eine Masterthesis anfertigen zu können,

Herrn Mag. Karl ATZMANSTORFER für die unterstützende Begleitung
während des gesamten UNIGIS Lehrganges,

Herrn Dipl.-Ing. Dieter LEITNER, GF der ZT - Datenforum reg. Gen. mbH
für die Bereitstellung der Daten,

meiner Frau Mag^a. Gabriele FORSTER-RIHA, MSc für die mentale Unterstützung
und die Zurückstellung der gemeinsamen Zeit zugunsten der Masterthesis

KURZFASSUNG

Die vorliegende Arbeit untersucht verschiedene, in kommerziellen GIS-Softwareprodukten üblicherweise implementierte, Interpolationsmethoden für Grundstückspreise im Bauland. Aktuelle Grundstückspreise als auch deren Entwicklung sind in der Wirtschaft und Verwaltung von großer Bedeutung, jedoch wenig transparent. Mithilfe flächendeckender Bodenpreiskarten könnte diesem Problem entgegengewirkt werden. Zu diesem Zweck müssen aus den punktuell vorhandenen Daten Schätzungen für die nicht beprobten Grundstücke berechnet werden. Hierzu stehen prinzipiell eine Reihe von Interpolationsmethoden zur Verfügung.

Die Arbeit untersucht unter Verwendung der GIS-Software von ArcGIS® die dort implementierten deterministischen Interpolationsverfahren Globale Polynominterpolation, lokale Polynominterpolation, Inverse Distance Weighted, Spline Interpolation sowie die geostatistischen Verfahren Ordinary und Universal Kriging. Die geostatistischen Methoden zeichnen sich durch die Nutzung der räumlichen Autokorrelation aus und haben zu einer stark zunehmenden Anwendung in den verschiedenen Disziplinen geführt.

Datengrundlage für die Untersuchungen bilden Grundstückspreise aus der Immobiliendatenbank „*immo netZT*“ der Firma ZT-Datenforum in Graz für die im Nordwesten der Steiermark liegenden Bezirke Schladming und Gröbming. Nach Aufbereitung, statistischen Analysen, Fehlerbereinigung und Übernahme der Daten in das GIS-System wurden für einzelne Gemeinden, die zwei Bezirke und den gesamten Datenbestand Schätzoberflächen gerechnet und das Ergebnis anhand der Kreuzvalidierung verglichen.

Dabei zeigte sich, dass allein auf dieser wenigen Datengrundlage die geostatistischen Methoden den deterministischen Methoden mit Ausnahme der Spline Interpolation überlegen sind. In Hinblick auf einen möglichst geringen mittleren quadratischen Fehler zwischen geschätzten und wahren Wert haben sich die Spline Interpolation mittels „completely regularized spline“ und die geostatistischen Methoden als gleichwertig erwiesen, wobei das Ordinary Kriging ausreichend ist und keine signifikanten Unterschiede zum Universal Kriging aufweist.

Im Ausblick wird hingewiesen, dass durch Hinzunahme geeigneter Zusatzinformationen die räumliche Struktur der Grundstückspreise besser beschrieben werden könnte und mittels des erweiterten geostatistischen Verfahrens Co-Kriging genauere Schätzungen zu erwarten sind. Zuletzt wird noch auf neue Forschungen zu Interpolationsmethoden für raumzeitliche Daten hingewiesen, welche die Zeit als weitere Dimension gleichzeitig behandeln.

INHALTSVERZEICHNIS

1. EINLEITUNG	1
1.1 NOTWENDIGKEIT DER RÄUMLICHEN INTERPOLATION	1
1.2 NUTZEN FÜR DEN IMMOBILIENMARKT	1
1.3 ZIELSETZUNG	2
2. PUBLIKATIONEN	3
3. GRUNDLAGEN	6
3.1 GRUNDLAGEN DER STATISTIK	6
3.1.1 Explorative Statistik	6
3.1.1.1 Grafische Darstellungen	7
3.1.1.2 Maßzahlen empirischer Verteilungen	8
3.1.2 Theoretische Verteilungen	11
3.1.2.1 Normalverteilung	11
3.1.2.2 Testen auf Normalverteilung	12
3.1.2.3 Box-Cox-Transformation	15
3.1.2.4 Ausreißerproblem	15
3.1.3 Schätzstatistik	17
3.2 GRUNDLAGEN DER GEOSTATISTIK	18
3.2.1 Theorie der regionalisierten Variablen	19
3.2.2 Stationarität	19
3.2.2.1 Strenge Stationarität	20
3.2.2.2 Stationarität zweiter Ordnung	20
3.2.2.3 Intrinsische Hypothese	20
3.2.3 Variographie	21
3.2.3.1 Autokorrelation	22
3.2.3.2 Empirisches Variogramm	24
3.2.3.3 Theoretisches Variogramm	26
3.2.3.4 Eigenschaften von Variogrammen	30
3.2.3.4.1 Verhalten im Ursprung	30
3.2.3.4.2 Anisotropie	31

3.2.3.4.3	<i>Drift</i>	32
3.2.3.4.4	<i>Proportionale Effekte</i>	33
3.2.3.4.5	<i>Loch-Effekt und Periodizität</i>	33
3.2.3.4.6	<i>Verschachtelte Strukturen</i>	33
4.	RÄUMLICHE INTERPOLATIONSVERFAHREN	35
4.1	THEORIE DER RÄUMLICHEN INTERPOLATION	35
4.2	RÄUMLICHE ANORDNUNG DER STICHPROBEN	36
4.3	DETERMINISTISCHE INTERPOLATIONSVERFAHREN	37
4.3.1	Polynom Interpolation	39
4.3.2	Inverse Distance Weighted Interpolation	40
4.3.3	Spline Interpolation	43
4.4	STOCHASTISCHE INTERPOLATIONSVERFAHREN	46
4.4.1	Simple Kriging	48
4.4.2	Ordinary Kriging	50
4.4.3	Universal Kriging	52
4.4.4	Lognormal Kriging	53
4.4.5	Trans-Gaussian Kriging	54
4.4.6	Co-Kriging	55
4.5	VALIDIERUNG VON SCHÄTZERGEBNISSEN	55
4.5.1	Kreuzvalidierung	55
4.5.2	Testen einer Kriginganordnung	58
5.	AUSWERTUNG UND ERGEBNISSE	59
5.1	DATENGRUNDLAGE	59
5.1.1	Immobilienbank immo net ZT	59
5.1.2	Digitale Katastralmappe	60
5.2	DATENAUFBEREITUNG	61
5.2.1	Explorative Datenanalyse	62
5.2.1.1	Fehlerbereinigung und Valorisierung	63
5.2.1.2	Ausreißeranalyse	63
5.2.1.3	Test auf Normalverteilung	65
5.2.1.4	Test auf Autokorrelation	67
5.2.1.5	Übernahme der Daten in ArcGIS	68

5.2.2 Variographie	69
5.2.2.1 Räumliche Statistik	70
5.2.2.2 Test auf Drift	70
5.2.2.3 Variogrammanalyse	73
5.2.2.3.1 Toleranzkriterien	73
5.2.2.3.2 Variogrammodell	74
5.2.2.3.3 Variogramparameter	75
5.2.2.3.4 Auswahl des Schätzverfahrens	75
5.2.2.3.5 Suchbereich der Schätzung	76
5.3 RÄUMLICHE INTERPOLATIONSERGEBNISSE	77
5.3.1 Deterministische Interpolation	78
5.3.2 Kriging Interpolation	80
6. SCHLUSSBEMERKUNGEN	83
6.1 ZUSAMMENFASSUNG DER VERWENDETEN METHODEN	83
6.2 ZUSAMMENFASSUNG DER ERGEBNISSE	84
6.3 AUSBLICK	84
7. LITERATURVERZEICHNIS	87
8. ANHANG	91

ABKÜRZUNGSVERZEICHNIS

ArcGIS®	Geoinformationssystem des Softwareherstellers ESRI Inc.
BEV	Bundesamt für Eich- und Vermessungswesen
CRS	Completely Regularized Spline
DKM	Digitale Katastralmappe
DXF	Drawing Interchange File Format (für den CAD-Datenaustausch)
EDA	Explorative Datenanalyse
GB	Gerichtsbezirk
GDB	Grundstücksdatenbank
GIS	Geografisches Informationssystem
GPI	Globale Polynom Interpolation
IDW	Inverse Distance Weighted
KDB	Koordinatendatenbank
KE	Kernel Estimation
KG	Katastralgemeinde
LPI	Lokale Polynom Interpolation
MS®	Microsoft
NNA	Nearest Neighbour Analysis
NNI	Nearest Neighbour Index
NNI	Natural Neighbour Interpolation
OK	Ordinary Kriging
PI	Polynom Interpolation
PPA	Point Pattern Analysis
RBF	Radiale Basisfunktion
RMSE	Root Mean Square Error
RMSPE	Root Mean Square Prediction Error
SEM	Standard Error of the Mean
SK	Simple Kriging
SI	Spline Interpolation
SPSS®	Software der Fa. SPSS Inc., 2009 durch IBM übernommen zur statistischen Analyse von Daten
TIN	Triangulated Irregular Network
TPS	Thin Plate Spline
UK	Universal Kriging
VMR	Variance/Mean Ratio

ABKÜRZUNGSVERZEICHNIS

Abb.	Abbildung
bzw.	beziehungsweise
d. h.	das heißt
engl.	englisch
Gl.	Gleichung
inkl.	inklusive
Kap.	Kapitel
sog.	sogenannte
S.	Seite
Tab.	Tabelle
u. a.	unter anderem
usw.	und so weiter
vgl.	vergleiche
z. B.	zum Beispiel
z. T.	zum Teil
%	Prozent

ABBILDUNGSVERZEICHNIS

Abb.3.1: Histogramm der Baulandgrundstückspreise pro m ² (ohne Valorisierung)	7
Abb.3.2: Box-Plot der Baulandgrundstückspreise pro m ² (ohne Valorisierung)	8
Abb.3.3: Asymmetrische Verteilung mit Darstellung des Modus, Median und arithmetischen Mittel	9
Abb.3.4: Normalverteilung	11
Abb.3.5: Normalisierter QQ-Plot	13
Abb.3.6: (Semi-)Variogramm	25
Abb.3.7: Kovarianz/Variogramm	26
Abb.3.8: Variogrammmodelle – (a) Sphärisches Modell, (b) Exponentielles Modell, (c) Gaußsches Modell, (d) Robustes Modell, (e) Nugget Modell, (f) Lineares Modell, (g) Potenz Modell, (h) 2D-Loch Effekt, (i) Kardinal-Sinus Modell (alle aus FUCHS, 2011)	30
Abb.3.9: Verschachtelte Struktur	34
Abb.4.1: Grundaufgabe der Interpolation (PETERS, 2008)	35
Abb.4.2: Verteilungsmuster (a) zufällig-random, (b) gleichmäßig-regular (c) gruppiert-clustered	35
Abb.4.3: Illustration des Konzeptes der bzgl. Anisotropie korrigierten effektiven Distanz (TOMCZAK, 1998)	42
Abb.4.4: IDW „bull-eyes“-Effekt	42
Abb.4.5: (a) Suchtechnik die n nächste Nachbarn um den Stützpunkt identifizieren (b) Quadrantensuche die eine gleichmäßige Verteilung der Nachbarn um den Schätzpunkt gewährleistet (c) Oktantensuche für die Nachbarn um den Stützpunkt (DAVIS, 2002)	43
Abb.4.6: Grafischer Vergleich zwischen IDW (links) und Spline (rechts) Interpolation (DAVIS, 2002)	44
Abb.5.1: Internetportal des ZT-Datenforum (http://www.zt.co.at) [30.04.2011]	59
Abb.5.2: Webplattform des ZT-Datenforum (http://web4you.at) [30.04.2011]	60
Abb.5.3: Histogramm der valorisierten Grundstückspreise pro m ² im Gerichtsbezirk	63
Abb.5.4: Box-Plot der valorisierten Grundstückspreise pro m ² im GB Schladming (links) und GB Gröbming (rechts)	64
Abb.5.5: Walsh-Test für valorisierte Grundstückspreise pro m ² auf Gerichtsbezirksebene	65
Abb.5.6: Untersuchungsgebiet und Datenwerte der Grundstückspreise Links GB Schladming, rechts GB Gröbming	70

Abb.5.7: Trendanalyse des gesamten Datensatzes	72
Abb.5.8: Voronoi-Diagramm der Gemeinde Mitterberg	72
Abb.5.9: Interpolationsoberfläche der Grundstückspreise	79
Abb.5.10: Assistent des Geostatistical Analyst in ArcGIS®	80

TABELLENVERZEICHNIS

Tab.5.1: Extremwerte der Grundstückspreise pro m ² aus den Originaldaten	61
Tab.5.2: Statistische Maßzahlen der valorisierten Grundstückspreise pro m ² im Gerichtsbezirk	62
Tab.5.3: Statistik der Prüfgröße B/s für die Gemeinden im GB Schladming	66
Tab.5.4: Statistik der Prüfgröße B/s für die Gemeinden im GB Gröbming	67
Tab.5.5: Moran's Index, Geary's C-Index und NN-Index für die Gerichtsbezirke	68
Tab.5.6: Ergebnis der Trendanalyse	71
Tab.5.7: Bewertung der Gemeinden für die Kriging Interpolation	77
Tab.5.8: Vergleich der deterministischen Interpolationsverfahren	79
Tab.5.9: Vergleich der Kriging Interpolationsverfahren	81
Tab.5.10: Vergleich beste deterministische Interpolation - Kriging	82

1 EINLEITUNG

1.1 Notwendigkeit der räumlichen Interpolation

Die meisten Karten sind Darstellungen eines kontinuierlichen Phänomens basierend auf Beobachtungen an diskreten Punkten (DAVIS, 2002). Das bedeutet, raumbezogene Daten werden im Allgemeinen punktuell erfasst und an verschiedenen Orten die Werte der Beobachtungsvariable gemessen. Um quantifizierte und genaue Informationen über räumliche Variationen eines zu untersuchenden Phänomens zu erhalten, wird mit räumlich kontinuierlich variierenden Oberflächen gearbeitet. Aus den punktuellen Stichproben-Messungen sollen Einzelwerte an anderen nicht beprobten Orten über das Untersuchungsgebiet geschätzt werden. Während die Geländeoberfläche ein anschauliches Beispiel für den Typ einer zumeist kontinuierlich variierenden Oberfläche ist, liegen bei vielen thematischen Werteoberflächen (Klima- und Wetterdaten, Grundwasseroberflächen, Schadstoffkonzentrationen) komplizierte Verhältnisse vor. Aufgabe der Interpolation ist es, Näherungswerte für einen unbekanntes Ort zu ermitteln bzw. aus den wenigen punktuellen Messungen ein möglichst gutes Bild und Verständnis der räumlichen Variation thematischer Oberflächen in einem Untersuchungsgebiet zu gewinnen.

Interpolationen bzw. zuverlässige Schätzungen erweisen sich dann als schwierig, wenn zusätzliche Faktoren berücksichtigt werden müssen. Dies führt zu den geostatistischen Verfahren, die Korrelationen zwischen den Beobachtungswerten erlauben. Eine wesentliche Voraussetzung für zuverlässige Ergebnisse mittels diesen Verfahren ist eine eingehende Datenanalyse.

1.2 Nutzen für den Immobilienmarkt

Der Immobilienmarkt hat eine große volkswirtschaftliche Bedeutung und hat sich zu einem globalen Markt entwickelt. Nicht zuletzt durch die Ende 2008 weite Teile der Welt betreffende Wirtschaftskrise, mit ausgelöst durch die Immobilienkrise in den USA, ist klar, wie wichtig der Immobilienmarkt geworden ist. Die Preise am Grundstücksmarkt sind in den Bereichen Recht, Wirtschaft und Verwaltung von großer Bedeutung, insbesondere in der Liegenschaftsbewertung. Liegenschaftswerte bilden die Grundlage von Investitionen, steuerlichen Angelegenheiten, Aufteilung von Gemeinschaftskosten, Enteignungsentschädigungen, Verlassenschaftsabhandlungen, dienen als Beweis für Sicherstellungen (Belehnungen) usw. Für die Bewertung von Liegenschaften sind für Immobiliensachverständige zuverlässige Informationen über das Preisniveau des örtlichen Liegenschaftsmarktes notwendig.

Dem Vergleichswertverfahren wird dabei eine aus seiner Plausibilität resultierende Vorrangigkeit beigemessen. Wenn genügend Vergleichswerte (tatsächlich erzielte Kaufpreise) existieren, ist dieses Verfahren eine einfache und zuverlässige Methode. In der Vergangenheit erzielte Kaufpreise können prinzipiell valorisiert, das heißt wertmäßig angepasst werden.

Oft sind die einzigen Informationsquellen Anzeigen in den Medien, die häufig auf nur wenige Grundstücke beschränkt sind, weshalb auf Schätzungen von Experten zurückgegriffen wird. Bodenpreiskarten, erstellt durch Interpolation von tatsächlich erzielten Kaufpreisen, würden deutlich das Preisniveau der Grundstücke visualisieren und wären geeignet Entscheidungshilfen zu bieten. Diese Konstruktion bietet weiters die Möglichkeit der Analyse von räumlicher Veränderung in der Bewertung.

1.3 Zielsetzung

Aufgabe dieser Arbeit ist es, für Grundstückspreise verschiedene Interpolationsmethoden anzuwenden und zu vergleichen und als Ergebnis vertrauenswürdige Verteilungskarten herzustellen.

Die Punktdaten wurden von der Firma ZT-Datenforum in Graz aus der Immobiliendatenbank „*immo netZT*“ zur Verfügung gestellt und beinhalten die Grundstückspreise für Bauland, landwirtschaftliche Nutzung und Wald im Gerichtsbezirk Schladming - Steiermark. Die geometrischen Daten wurden für die 14 Katastralgemeinden des Gerichtsbezirks über das Bundesamt für Eich- und Vermessungswesen (BEV) bezogen. Die Daten der Grundstückspreise sind einer statistischen Analyse zu unterziehen, die verschiedenen Interpolationsmethoden (IDW, Spline und Kriging) auszuwerten und zu vergleichen und das am meisten geeignete Verfahren zu bestimmen.

Für diese Untersuchungen stehen prinzipiell eine Reihe von Methoden zur Verfügung, die angewandten Methoden sind aber vor allem auf die verwendete Statistik-Software IBM® SPSS® Statistics Student Version 18.0, R-Programm 2.12.2 und Crimestat 3.3, sowie der GIS-Software ArcGIS® 10.0 von ESRI® mit der Erweiterung ArcGIS Geostatistical Analyst beschränkt.

2 PUBLIKATIONEN

Im deutschsprachigen Raum finden sich wenig Veröffentlichungen zum Thema Interpolationsverfahren für Grundstückspreise, andere Anwendungen und Untersuchungen belegen die prinzipielle Anwendbarkeit dieser Verfahren in den verschiedenen Disziplinen.

Anhand von Daten über die Grundwasserbeschaffenheit verwendet GRAMS (2000) in Ihrer Doktorarbeit nur das geostatistische Interpolationsverfahren Ordinary Kriging für 121 unregelmäßig verteilte Messstellen. Ausführlich werden die Eignung des Datensatzes zur korrekten Regionalisierung überprüft und Variogrammodelle beschrieben.

GERLACH (2001) vergleicht in Ihrer Diplomarbeit die zwei Interpolationsverfahren Inverse-Distance-Weighted (IDW) und Universal Kriging anhand von Windgeschwindigkeitsdaten, die in zwei Untersuchungsgebieten, zum einem mit 18 und zum anderen mit 14 unregelmäßig verteilten Messstationen ermittelt wurden. In einem Gebiet erweist sich IDW, im anderen Gebiet Universal Kriging als bessere Interpolationsmethode. Auf den Vergleich der Interpolationsmethoden wird sehr ausführlich eingegangen, auf die Wahl der Interpolationsparameter allerdings wenig. Festzuhalten ist, dass für eine Interpolation eine sehr geringe Anzahl von Stützpunkten (18 bzw. 14 Messstellen) vorhanden ist.

In seiner Doktorarbeit untersucht SPICKERMANN (2005) verschiedene räumliche Interpolationsverfahren (IDW, Spline, Natural Neighbour, Ordinary Kriging) anhand von Daten über Pflanzenkrankheiten. Die Stichproben sind in einem Raster gleichmäßig angeordnet, was ausführlich begründet wird. Die besten Interpolationsergebnisse werden durch Kriging erreicht, nur für Daten mit zufälliger Verteilung war IDW zu bevorzugen. Auf die Variogrammanalyse wird ausführlich eingegangen.

In seiner Diplomarbeit vergleicht PETERS (2008) die räumlichen Interpolationsverfahren IDW, Spline und Ordinary Kriging für Ertragswerte im Weinbau. Verwendet wurde eine regelmäßige Stichprobenanordnung, sodass für die verschiedenen Weine 27 bis 47 Stichproben zu Verfügung standen. Für die Bodenwerte konnten 92 Stützpunkte zur Interpolation verwendet werden. Auch hier ergibt das Krigingverfahren das beste Ergebnis.

SCHAUPPENLEHNER (2008) untersucht in seiner Dissertation die grundsätzliche Eignung von Punktdaten der österreichischen Finanzbodenschätzung für die Erstellung hoch aufgelöster Bodenqualitätskarten mithilfe statistischer als auch geostatistischer Interpolationsverfahren. Dabei zeigte sich, dass sie geostatistischen Methoden mit Ausnahme des Ordinary-Kriging bessere Ergebnisse liefern als das multiple Regressionsmodell. Als bestes Modell hat sich für die Fragestellung das Regression Kriging-Modell erwiesen, wenngleich die Unterschiede zu Universal Kriging und Co-Kriging geringfügig sind.

In seiner umfassenden Dissertation behandelt GAU (2010) den Einfluss des Anwenders im geostatistischen Modellierungsprozess am Beispiel der Baugrundmodellierung. Im Mittelpunkt praxisnaher Untersuchungen steht im Wesentlichen der Aspekt der Anwendung unter Beschreibung der eingesetzten Methoden. Mit einer geostatistischen Anwendung ist jedoch die Notwendigkeit von Entscheidungen gegeben, die vom Benutzer oft unbewusst getroffen werden. Die Einflussnahme des Anwenders erstreckt sich auf die qualitative Auswahl von Parametern (z. B. Auswahl von Variogrammodellen) aber auch auf die quantitative Festlegung von Einflussgrößen (z. B. Toleranzkriterien, Variogrammparameter). Wesentliches Ergebnis der Arbeit ist, dass die Einflussnahme des Benutzers erheblich ist. Bedingt durch Prinzipien der Geostatistik hat der Anwender insbesondere methodenspezifische Parameter und Optionen auszuwählen, die zwar scheinbar eine objektive Modellierung gestatten, im Kern jedoch weiterhin subjektiv bleiben. Mit steigendem Detaillierungsgrad der Entscheidungen entziehen sich die jeweiligen Parameter, die sich bisweilen nur ansatzweise mit den im Phänomen verwirklichten Strukturen parallelisieren lassen, immer weiter der gedanklichen Nahvollziehbarkeit. Die in dieser Arbeit getroffenen Aussagen lassen sich, insbesondere zur Variographie und zum Kriging, prinzipiell auch auf andere Anwendungen als der Baugrundmodellierung übertragen.

Interpolation und Analyse für Grundstückspreise untersucht DUMFARTH (1996) in seiner Diplomarbeit für Bodenpreisoberflächen der Stadt Salzburg. Getestet und verwendet wurden die in der Software ArcInfo implementierten Verfahren IDW und das als dort bezeichnete Minimum Curvature Spline bezeichnete CRS (completely regularized spline with smoothing and tension). Da die Ergebnisse der Qualitätstests mit CRS meist besser waren, wurde die der Studie zugunsten des Splineverfahrens entschieden. Kriging setzt grundsätzlich eine homogene Variabilität der Werte über den Raum voraus (Normalverteilung der Daten) und daher wegen Ausreißer auf dieses stabile Interpolationsverfahren zurückgegriffen und Kriging nicht untersucht wurde.

PÖCHINGER (2009) analysiert in seiner Masterarbeit Grundstückspreise für die Steiermark anhand statistischer Methoden. Ziel der Arbeit war jedoch nicht Verteilungskarten zu erzeugen, sondern einen Preisindex zu ermitteln, der für die Valorisierung verwendet werden kann. Für den Gerichtsbezirk Schladming wurde dies exemplarisch durchgeführt.

Mit diesen Daten setzt die gegenständliche Masterthesis fort, um mittels Interpolation direkte Schätzwerte für aktuelle Grundstückspreise zu erhalten.

3 GRUNDLAGEN

3.1 Grundlagen der Statistik

Statistik wird einerseits als eigenständige mathematische Disziplin über das Zusammenfassen, die Analyse, die Charakterisierung und Interpretation von Daten betrachtet, andererseits als Teilgebiet der Mathematik, insbesondere der Stochastik, angesehen. Die Stochastik befasst sich mit dem Beschreiben und Untersuchen von Zufallsexperimenten. Statistik ist eine auf Mathematik basierende Methode zur Analyse von quantitativen Daten. Sie verbindet dabei empirische Daten mit theoretischen Modellen.

Die Statistik kann in drei Teilbereiche eingeteilt werden:

- Die *deskriptive (beschreibende) Statistik* zur Gliederung, Zusammenfassung, Darstellung und Analyse vorhandener Daten. Mit ihren Methoden verdichtet man quantitative Daten zu Tabellen, grafischen Darstellungen und Maßzahlen.
- Die *induktive (schließende) Statistik* oder *Interferenzstatistik* leitet aus den Daten einer Stichprobe Eigenschaften der zugehörigen Grundgesamtheit ab. Die Wahrscheinlichkeitstheorie liefert die Grundlagen für die erforderlichen Schätz- und Prüfverfahren.
- Die *explorative (analytische) Statistik* ist methodisch eine Zwischenform der beiden vorgenannten Teilbereiche, erlangt als Anwendungsform jedoch zunehmend eigenständige Bedeutung. Mittels deskriptiver Verfahren und induktiver Testmethoden werden mögliche Zusammenhänge zwischen Daten in vorhandenen Datenbeständen systematisch gesucht und zugleich in Ihrer Stärke und Ergebnissicherheit bewertet.

3.1.1 Explorative Statistik

Empirische Daten, deren Struktur und Auffälligkeiten interessieren, werden mit den Methoden der deskriptiven und der (erkundenden) explorativen Datenanalyse (EDA) untersucht, ohne dass ein Modell vorausgesetzt wird. Dies gilt insbesondere für Daten irgendwelcher Teilgesamtheiten, die nicht auf sorgfältig geplante Art gewonnen worden sind. Aufgabe der EDA ist das Aufspüren und die zusammenfassende Darstellung unbekannter Strukturen von (umfangreichen) Stichproben bei unbekannter Grundgesamtheit, kaum geplanter Datenerhebung, Unklarheit über die Auswahl geeigneter Modelle und quantitativ nicht fassbarer Aussagegenauigkeit (SACHS & HEDDERICH, 2009). In diesem Kapitel werden grafische Darstellungen und

Maßzahlen zusammengefasst, die bei der Charakterisierung der Daten Anwendung finden.

3.1.1.1 Grafische Darstellungen

Ein erster Überblick über die Datenverteilung kann durch ein Histogramm gewonnen werden. In diesen werden die in Klassen zusammengefassten Messwerte und ihre Häufigkeit aufgetragen (Abb. 3.1). Damit kann die zugrunde liegende Verteilungsform (gleichmäßig, symmetrisch, asymmetrisch) abgeschätzt werden. Des Weiteren sind aus dem Histogramm anschaulich Extremwerte ersichtlich, die mögliche Ausreißer darstellen.

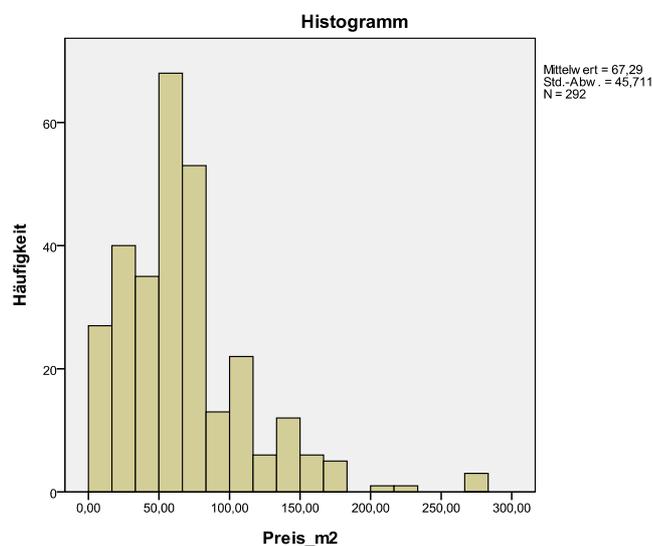


Abb. 3.1: Histogramm der Baulandgrundstückspreise pro m² (ohne Valorisierung)

Eine weitere grafische Darstellungsform ist der sogenannte Box-Plot (Abb. 3.2). Dafür werden alle Daten der Größe nach geordnet und die Quartildistanz (I_{50}) aus der Differenz des dritten und ersten Quantils bestimmt. Diese umfasst damit 50% der Messwerte, d. h. 25% der Messwerte liegen unterhalb und 25% liegen oberhalb der Quartildistanz (siehe Gl. 3.19).

Im Box-Plot wird die Quartildistanz als Box entlang der Ordinate in einem Koordinatensystem dargestellt, mit einer inneren Linie, die den Median repräsentiert. Ferner werden der kleinste und größte Wert markiert, sofern sie keine Ausreißer sind. Werte die um mehr als das dreifache der Boxhöhe außerhalb liegen (Extremwerte) werden als Stern markiert, Werte die mehr als das eineinhalbfache außerhalb liegen (Ausreißer) als Kreis (BÜHL, 2010). Dieser nicht parametrische Test ist unabhängig von der Verteilung, da er sich auf den Median (Kap. 3.1.2.4) bezieht.

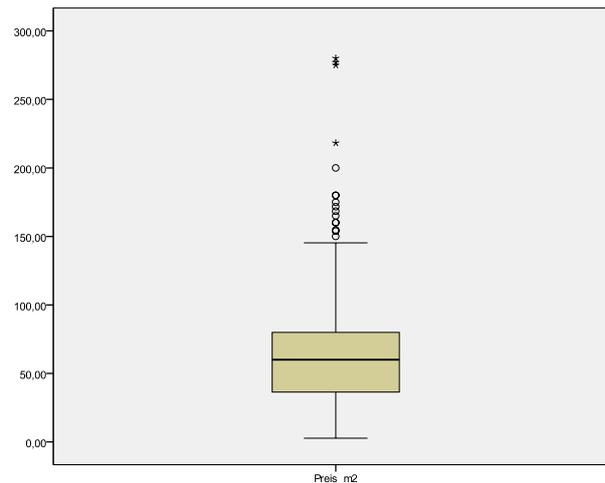


Abb. 3.2: Box-Plot der Baulandgrundstückspreise pro m² (ohne Valorisierung)

3.1.1.2 Maßzahlen empirischer Verteilungen

Häufigkeitsdiagramme vermitteln einen visuellen Eindruck von der Form der empirischen Verteilung und in ihnen kommen die Charakteristika dieser wie Symmetrie, Schiefe usw. quasi gleichzeitig zum Ausdruck, allerdings nur ungenau (BAHRENBERG et al., 1999). Für genaue Aussagen zur Beschreibung der Charakteristika verwendet man Maßzahlen, die sich auf die Lage, Streuung oder Gestalt der Verteilung beziehen.

Als Lagemaße werden jene Maßzahlen bezeichnet, die die Häufigkeitsverteilung durch einen zentralen (mittleren) Wert repräsentieren. Die meistverwendeten sind Modus, Median und arithmetisches Mittel.

- Der Modus (Modalwert) ist derjenige Variablenwert, der am häufigsten vorkommt, d. h. wo die empirische Verteilung $F(x)$ ihr Maximum erreicht.

$$\bar{m}_{Mod} = F(x)_{max} \quad (3.1)$$

- Der Median (Zentralwert) teilt die der Größe nach geordnete Datenreihe (x_1, x_2, \dots, x_n) in zwei gleich große Hälften.

$$\begin{aligned} \bar{m}_{Med} &= \frac{x_{n+1}}{2} && n \text{ ungerade} \\ \bar{m}_{Med} &= \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} && n \text{ gerade} \end{aligned} \quad (3.2)$$

- Unter dem arithmetischen Mittel (Mittelwert) versteht man

-

(3.3)

In symmetrischen Verteilungen fallen das arithmetische Mittel, der Median und der Modus zusammen, während bei asymmetrischen Verteilungen der Median zwischen arithmetischem Mittel und Modus liegt (Abb. 3.3).

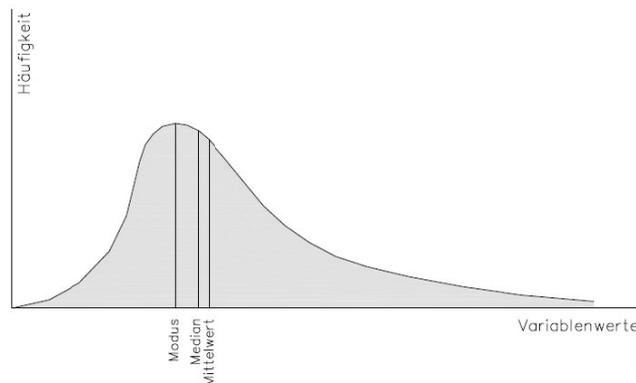


Abb. 3.3: Asymmetrische Verteilung mit Darstellung des Modus, Median und arithmetischen Mittel

Der arithmetische Mittelwert spielt in der Statistik eine dominierende Rolle als Lageparameter, da er für theoretische Verteilungen leicht zu berechnen ist.

Die Streuungsmaße Spannweite, mittlere Abweichung, Varianz, Standardabweichung und Variationskoeffizient sind weitere Maßzahlen zur Kennzeichnung der Datenvariation einer Stichprobe (x_1, x_2, \dots, x_n) .

- Die Spannweite B ist als Breite des gesamten Wertebereichs einer Variablen definiert, also als Differenz zwischen dem größten (Maximum) und kleinsten Wert (Minimum).

(3.4)

- Die mittlere Abweichung ist als mittlerer Abstand der Datenwerte vom arithmetischen Mittel definiert:

(3.5)

- Die Varianz s^2 ist die mittlere quadratische Abweichung der Datenwerte vom arithmetischen Mittel:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (3.6)$$

- Als Standardabweichung bezeichnet man die Wurzel aus der Varianz:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (3.7)$$

Varianz und Standardabweichung haben in der Statistik eine größere Bedeutung als die mittlere Abweichung, da sie sich leicht für theoretische Verteilungen berechnen lassen.

- Für Vergleichszwecke müssen die absoluten Streuungsmaße relativiert werden. Dazu dient der Variationskoeffizient, der die Standardabweichung als prozentuellen Anteil am arithmetischen Mittel ausdrückt:

$$v = \frac{s}{\bar{m}} \quad \text{alle } x_i > 0 \quad (3.8)$$

Ein Koeffizient größer als eins zeigt das Vorhandensein einiger unregelmäßig hoher Werte an, die einen signifikanten Einfluss auf die statistische Auswertung haben können (ISAAKS & SRIVASTAVA, 1989).

Die Gestalt (Form) der Verteilung kann durch die Maßzahlen Schiefe und Exzess charakterisiert werden. Für eine Stichprobe (x_1, x_2, \dots, x_n) mit Mittelwert \bar{m} und Standardabweichung s versteht man

- die Schiefe als

$$a_3 = \frac{\sum_{i=1}^n (x_i - \bar{m})^3}{ns^3} \quad (3.9)$$

Sie charakterisiert die Symmetrie ($a_3=0$) bzw. Asymmetrie einer Verteilung ($a_3>0$ linksteilige, $a_3<0$ rechtsteilige Verteilung).

- Der Exzess (Kurtosis) charakterisiert die Wölbung und berechnet sich nach

$$a_4 = \frac{\sum_{i=1}^n (x_i - \bar{m})^4}{ns^4} - 3 \quad (3.10)$$

Die Wölbung wird als normal bezeichnet für $a_4=0$, als flach für $a_4<0$ und als steil für $a_4>0$.

3.1.2 Theoretische Verteilungen

Verteilungen von Grundgesamtheiten (Populationen) bzw. von Zufallsvariablen nennt man theoretische Verteilungen. Unter ihnen hat die Normalverteilung eine herausragende Bedeutung, insbesondere durch den zentralen Grenzwertsatz (central limit theorem). Er beschreibt die sehr markante Eigenschaft, Zufallserscheinungen, die sich als Summe einer genügend großen Anzahl voneinander unabhängiger Einzelwirkungen ergeben, von denen jeder zur Summe nur einen unbedeutenden Beitrag liefert, durch das Modell der Normalverteilung darstellen und bewerten zu können (SACHS & HEDDERICH, 2009, BRONSTEIN & SEMENDJAJEW, 1987).

Dieses Kapitel befasst sich mit statistischen Methoden zur Schätzung einer Grundgesamtheit aus einer Stichprobe und zeigt, wie diese Schätzungen für Tests verwendet werden können.

3.1.2.1 Normalverteilung

Der zentrale Grenzwertsatz liefert häufig die Rechtfertigung empirische (Stichproben-) Verteilungen mit genügend großen Stichprobenumfang als Normalverteilung zu approximieren. Außerdem basieren zahlreiche statistische Schätz- und Prüfverfahren auf der Normalverteilung. Sehr nützlich daran ist, dass die Maßzahlen Mittelwert und Standardabweichung der deskriptiven Statistik diese Verteilung beschreiben können.

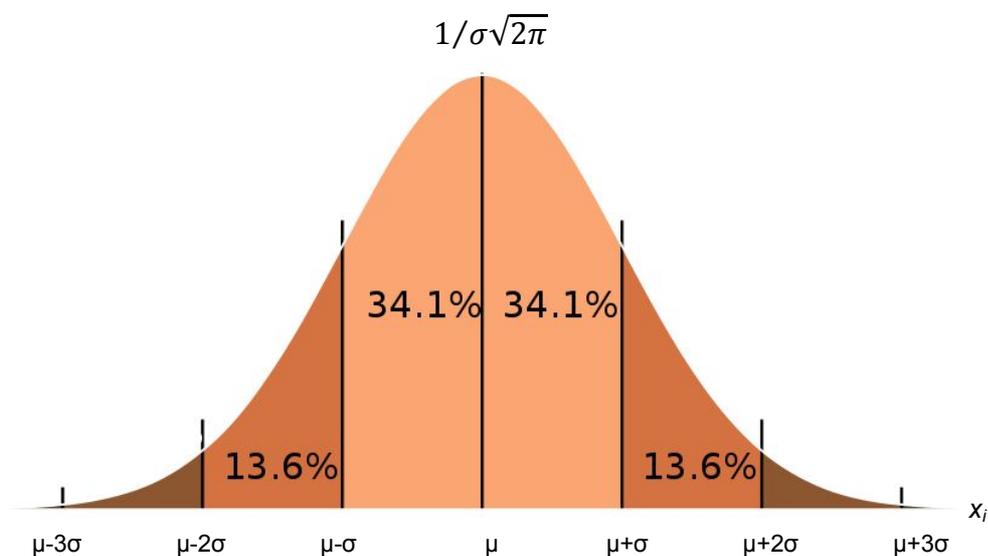


Abb. 3.4: Normalverteilung

Abbildung 3.4 zeigt die Normalverteilung, mit dem wahren Mittel- bzw. Erwartungswert μ und der Standardabweichung σ , die wegen ihrer Gestalt oft auch als Glockenkurve bezeichnet wird. Sie ist unimodal, symmetrisch und erreicht das Maximum für $x=\mu$.

Der wahre Mittelwert μ berechnet sich analog zur empirischen Verteilung (Gl. 3.3) mit

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (3.11)$$

und die Standardabweichung zu

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{N}} \quad (3.12)$$

(Der Großbuchstabe N soll anzeigen, dass es sich um die Größe der Population handelt, im Vergleich zum Kleinbuchstaben n für den Stichprobenumfang).

Bei der Berechnung der Standardabweichung für Stichproben nach Gl. (3.7) wird in der Regel durch $(n-1)$ dividiert, und zwar dann, wenn die Maßzahlen der Stichprobe als Schätzung der entsprechenden Parameter der Grundgesamtheit dienen sollen.

Setzt man für eine normalverteilte Variable X die Transformation

$$Z = \frac{x - \mu}{\sigma} \quad (3.13)$$

ein, erhält man die standardisierte Normalverteilung mit Erwartungswert null und Standardabweichung eins. Aufschlussreich für die Beurteilung einer normalverteilten Variablen sind die zentralen Schwankungsintervalle:

$$\begin{aligned} \mu \pm \sigma & \text{ oder } z = \pm 1 & \text{ mit } 68,27\% \text{ Anteil der Grundgesamtheit} \\ \mu \pm 2\sigma & \text{ oder } z = \pm 2 & \text{ mit } 95,45\% \text{ Anteil der Grundgesamtheit} \\ \mu \pm 3\sigma & \text{ oder } z = \pm 3 & \text{ mit } 99,73\% \text{ Anteil der Grundgesamtheit} \end{aligned} \quad (3.14)$$

3.1.2.2 Testen auf Normalverteilung

Zahlreiche statistische Untersuchungsmethoden und geostatistische Verfahren setzen zumindest annähernd normalverteilte Daten voraus. Bevor diese Methoden angewendet werden können, muss deshalb diese Annahme überprüft werden. Auch hier stehen wieder grafische und numerische Methoden zur Verfügung.

- Das normalisierte **QQ-Diagramm** (Quantil-Quantil Diagramm, oder Normalverteilungsdiagramm) ist eine grafische Darstellung, ob die Daten als hinreichend normalverteilt angesehen werden können. Dabei werden die der Größe nach kumulativ geordneten Daten mit der Verteilungsfunktion der Standardnormalverteilung verglichen (Abb. 3.5). Der Vergleich erfolgt in Bezug auf eine Referenzgerade. Abweichungen zwischen der theoretischen Normalverteilung und der empirischen Verteilung werden durch Abweichungen der Werte von der Geraden deutlich.

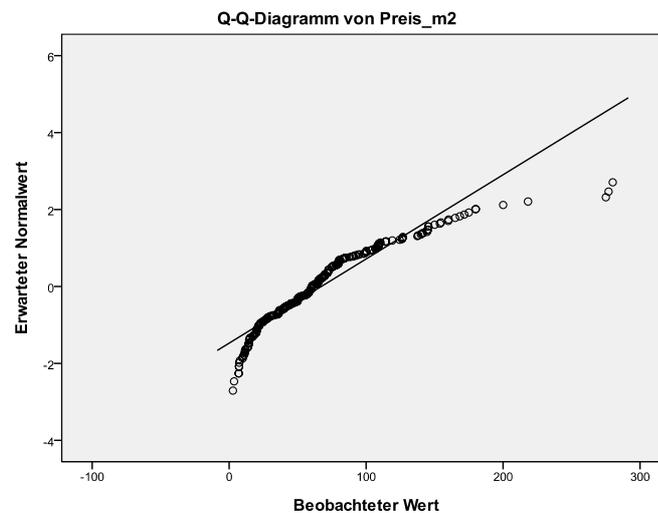


Abb. 3.5: Normalisierter QQ-Plot

- Ist der Quotient der Spannweite zur Standardabweichung zu klein oder zu groß, dann zeigt dies eine Abweichung von der Normalverteilung an.

$$\frac{\text{Spannweite}}{\text{Standardabweichung}} = \frac{B}{s} \quad (3.15)$$

Liegt das Verhältnis bei einer Stichprobe nicht innerhalb der kritischen Grenzen laut Tabelle von PEARSON & STEPHENS (1964), dann kann das Vorliegen einer Normalverteilung auf den entsprechenden Signifikanzniveaus abgelehnt werden. Tabelle 1 des Anhangs zeigt den kritischen Bereich des Quotienten B/s auszugsweise mit einem Signifikanzniveau von 5% und 10%.

- Eine weitere grobe Aussage über Normalverteilung geben die sogenannten „Sigma-Regeln“. Die „**Drei-Sigma-Regel**“ besagt, dass beim Vorliegen einer Normalverteilung außerhalb der Grenzen $\mu \pm 3\sigma$ kaum noch gültige Werte vor-

liegen. Für kleine Stichproben ist diese Regel jedoch zu konservativ (SACHS & HEDDERICH, 2009).

- Die Abweichung einer empirischen Verteilung vom Modell einer Normalverteilung kann durch die **Schiefe** oder durch den **Exzess** beurteilt werden, deren Maßzahlen bereits in den Gl. (3.9) und (3.10) angegeben sind, wobei für die Berechnung der Standardabweichung s der Nenner $(n-1)$ durch n zu ersetzen ist. Für Normalverteilung sollte die Schiefe nahe null und der Exzess nahe drei liegen. Ein stark negativer Exzess weist auf eine bimodale Verteilung hin.
- Ein weitverbreiteter nicht parametrischer Test zur Überprüfung der beobachteten Daten auf (Normal-)Verteilung ist der **Kolmogoroff-Smirnov-Anpassungstest**. Der Test basiert auf dem maximalen Abstand der empirischen Verteilungsfunktion mit der ausgewählten speziellen (theoretischen) Verteilungsfunktion, z. B. der Normalverteilung. Wenn der Wert dieser Prüfgröße größer ist als der kritische Wert, der entsprechenden Tabellen entnommen bzw. für $n > 40$ nach einer einfachen Formel bestimmt werden kann, wird die Annahme der Nullhypothese (Normalverteilung) verworfen. Der Vorteil des K-S-Test ist, dass sie nicht von den beobachteten Verteilung abhängt und exakt bestimmt werden kann. Die theoretische Verteilungsfunktion muss stetig und hinsichtlich der Verteilungsparameter (Lage, Varianz, Schiefe, Exzess) vollständig definiert sein. Der Test ist sehr stabil und unanfällig. Zu beachten ist jedoch, dass der K-S-Test sensitiver in der Mitte der Verteilung als an den Rändern entscheidet und für den Fall, dass die Parameter aus der Stichprobe geschätzt werden müssen der Test nicht genau und konservativ ist. Besser ist dann für den speziellen Fall der Normalitätsprüfung die **Lilliefors-Modifikation** des K-S-Tests.
- Der **Shapiro-Wilk Test** prüft die Hypothese, dass die Daten einer Stichprobe einer normalverteilten Variablen zugeordnet werden können. Dazu wird die Verteilung des Quotienten aus zwei Schätzungen für σ^2 betrachtet (Gl. 3.16). Die Prüfgröße berechnet sich mittels eines ersten Schätzers im Zähler, wie die Varianz einer Stichprobe aussehen müsste, wenn sie aus einer normalverteilten Grundgesamtheit stammte, und vergleicht diese erwartete Varianz mit einem zweiten Schätzer im Nenner für die tatsächliche Varianz der Stichprobe.

$$\widehat{W} = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{b^2}{(n-1)s^2} \quad (3.16)$$

Die erwartete Varianz im Zähler wird mittels der Methode der kleinsten Quadrate geschätzt für die Steigung einer Regressionsgeraden im QQ-Dia-

gramm, der Schätzer im Nenner ist die Stichprobenvarianz. Für den Fall, dass eine Normalverteilung vorliegt, sollten diese beiden Schätzungen nahe zusammen liegen und somit das Verhältnis eins ergeben. Der Shapiro-Wilk Test zeichnet sich gegenüber den anderen Testverfahren durch seine hohe Teststärke aus. Der 1965 entwickelte Test war ursprünglich nur imstande Stichproben vom Umfang $3 < n < 50$ zu untersuchen. Verbesserungen durch ROYSTON (1982) und RAHMAN & GOVINDARAJULU (1997) erweiterten den Einsatzbereich des Test auf $3 < n < 5000$. In der Statistiksoftware SPSS und 'R' wird der Shapiro-Wilk Test durch die Funktion nach Royston berechnet, die bis zu einer Stichprobengröße $n < 2000$ möglich ist.

3.1.2.3 Box-Cox-Transformation

Die Annahme annähernd normalverteilter Daten ermöglicht den Einsatz von parametrischen Tests, die in der Regel einfacher anzuwenden sind und eine höhere Teststärke aufweisen. Viele reale Daten erfüllen diese Voraussetzung jedoch nicht, allerdings kann eine geeignete Transformation dies erzwingen.

Mit der Box-Cox-Transformation (Gl. 3.17) nach G.E.P. Box und D.R. Cox steht eine Transformation zur Verfügung mit der Daten in eine Normalverteilung überführt werden können.

$$\begin{aligned} T_\lambda(x) &= \frac{x^\lambda - 1}{\lambda}, \quad \text{für } \lambda \neq 0 \\ T_0(x) &= \log x, \quad \text{für } \lambda = 0 \end{aligned} \quad (3.17)$$

Für die Bestimmung eines geeigneten Wertes für λ können QQ-Diagramme dienen. Für alle Werte eines bestimmten Definitionsbereich wird ein λ gewählt, dass zwischen den transformierten Daten und den (theoretischen) Quantilen der Standardnormalverteilung zu einem möglichst hohen Korrelationskoeffizienten führt.

3.1.2.4 Ausreißerproblem

Extrem niedrige oder hohe Werte innerhalb der Daten werden als Ausreißer bezeichnet und können unter gewissen Umständen gestrichen werden, falls ihre Korrektheit fraglich ist. Messfehler oder Beurteilungsfehler müssen eliminiert werden, wenn dies sachlich begründet werden kann. Saubere statistische Entscheidungen einen Messwert als Ausreißer zu identifizieren sind nur selten möglich. Zur Erkennung starker Abweichungen der Normalverteilung sind folgende Varianten möglich:

- Nach der **Tschebyscheff Ungleichung** (3.18) muss ein Intervall von 4 Standardabweichungen verwendet werden (**4-Sigma-Bereich**), um sicherzustellen, dass zumindest 93,75% der Daten einer beliebigen Verteilung in dieses Intervall fallen. Bei mindestens 10 Einzelwerten (besser $n \gg 25$) darf ein Wert als Ausreißer verworfen werden, wenn er außerhalb des Bereiches $(\bar{m} \pm 4s)$ liegt, wobei Mittelwert und Standardabweichung ohne den ausreißerverdächtigen Wert berechnet werden.

$$P(|x - \mu| < k\sigma) \geq 1 - 1/k^2 \quad (3.18)$$

- Ein **robustes Verfahren** für die Erkennung von Ausreißern kann auf der Grundlage der Quartile und der Quartildistanz abgeleitet werden:

$$Q_1 - k(Q_3 - Q_1) \leq \text{kein Ausreißer} \leq Q_3 + k(Q_3 - Q_1), \quad (3.19)$$

wobei für k Werte zwischen 1.5 und 3 gewählt werden. Diese Methode wird für Box-Plots verwendet, in denen Werte die außerhalb der Quartildistanz liegen besonders markiert und einzeln dargestellt werden.

- Der **Walsh-Test** ist ein nicht-parametrisches Verfahren, um Ausreißer in einem Datensatz zu detektieren. Der Test erfordert zwar eine große Anzahl von Datenwerten ($n > 220$ für ein Signifikanzniveau von $\alpha=0.05$, $n > 60$ für $\alpha=0.10$, für $n < 60$ ist der Test nicht anwendbar), kann dafür aber auch für verteilungsfreie Daten angewendet werden (WALSH, 1951, 1953). Für eine aufsteigend sortierte Datenreihe (x_1, x_2, \dots, x_n) werden folgende Schritte durchgeführt:

1. A-priori Festlegung der möglichen Ausreißer $r \geq 1$.

2. Berechnung von:

$$c = \text{ceil}(\sqrt{2n}), k = r + c, b^2 = 1/\alpha \quad \text{und} \\ a = \frac{1+b\sqrt{\frac{c-b^2}{c-1}}}{c-b^2-1} \quad (3.20)$$

ceil() Aufrundungsfunktion zur nächst größeren ganzen Zahl

3. Die r kleinsten Werte sind Ausreißer, falls

$$x_r - (1+a)x_{r+1} + ax_k < 0. \quad (3.21)$$

4. Die r größten Werte sind Ausreißer, falls

$$x_{n+1-r} - (1+a)x_{n-r} + ax_{n+1-k} > 0. \quad (3.22)$$

- Der **Grubbs-Test** kann zur Erkennung von Ausreißern in normalverteilten Daten verwendet werden. Der Test entdeckt jeweils einen einzelnen Ausreißer, der aus den Daten eliminiert wird und verläuft iterativ solange, bis kein Ausreißer mehr erkannt wird. Die Prüfgröße ist definiert als die größte absolute Abweichung vom Mittelwert \bar{m} der Stichprobe relativiert zur Standardabweichung s .

$$\hat{G} = \frac{\max(|x_i - \bar{x}|)}{s} \quad (3.23)$$

Ist der dieser Wert größer als der entsprechende kritische Wert, der nach (3.24) berechnet werden kann, dann gibt es mindestens einen Ausreißer.

$$G_{n,\alpha} = \frac{(n-1)}{\sqrt{n}} \sqrt{\frac{t_{n-2,\alpha/2n}^2}{n-2+t_{n-2,\alpha/2n}^2}} \quad (3.24)$$

Dabei ist $t_{n-2,\alpha/2n}$ der kritische Wert der t-Verteilung mit $(n-2)$ Freiheitsgraden und dem Signifikanzniveau von $\alpha/2n$.

Statistische Tests zur Ausreißeranalyse stehen in Softwareprogrammen selten zur Verfügung, weshalb in den meisten Fällen auf das Box-Plot Verfahren zurückgegriffen wird. Im 'R'-Programm ist der Grubbs-Test im Paket outliers() neben anderen Verfahren implementiert. In SPSS werden für die Berechnung der Ausreißer einfach die fünf höchsten und niedrigsten Werte angezeigt. Die statistischen Tests sollten zumindest mit einem Signifikanzniveau von 10% durchgeführt werden.

3.1.3 Schätzstatistik

Der Mittelwert \bar{m} einer Stichprobe kann oder kann keine genaue Schätzung des wahren Mittelwertes μ der Grundgesamtheit sein. Wenn man einige Stichproben zufällig aus der Grundgesamtheit nimmt und für jede den Mittelwert berechnet, kann statistisch gezeigt werden, dass diese Mittelwerte ebenfalls normalverteilt mit eigenem Mittelwert und Standardabweichung sind. Die Standardabweichung der Verteilung der Mittelwerte aus Stichproben ist eine wichtige statistische Größe und wird als Standardfehler (SEM – standard error of the mean) bezeichnet. Dieser berechnet sich für ein Stichprobenumfang n nach

$$SEM = \sigma_{\bar{m}} = \frac{\sigma}{\sqrt{n}} \quad (3.25)$$

Der Standardfehler ist wichtig für das Testen von Hypothesen und kann auch verwendet werden, um Konfidenzintervalle um den wahren Mittelwert μ zu schätzen. Üblicherweise sind die wahren Parameter μ und σ der Grundgesamtheit unbekannt. Falls man Daten von nur einer Stichprobe hat, stehen nur der empirische Mittelwert \bar{x} und die empirische Standardabweichung s zur Verfügung. Um den Standardfehler zu berechnen, können diese Werte in Gl. (3.25) eingesetzt werden und man erhält damit:

$$SEM = s_{\bar{m}} = \frac{s}{\sqrt{n}} \quad (3.26)$$

Zur Entscheidungsfindung, ob ein Mittelwert einer Stichprobe signifikant von einem erwarteten Mittelwert μ differiert, kann die Beziehung

$$t = \frac{\bar{m} - \mu}{SEM} \quad (3.27)$$

herangezogen werden. Diese Maßzahl folgt jedoch der sog. t -Verteilung, weshalb speziell für kleine Stichprobenumfänge kleiner als 100, die anfällig für Stichprobenfehler sind, die entsprechende Korrektur notwendig sind (MCKILLUP & DARBY, 2010). Das Konfidenzintervall von 95% (Signifikanzniveau 5%) von $\pm 1.96\sigma$ vergrößert sich für eine Stichprobengröße $n=4$ auf $\pm 3.182\sigma$, $n=60$ auf $\pm 2.001\sigma$, $n=200$ auf $\pm 1.972\sigma$ und $n=1000$ auf $\pm 1.962\sigma$. Unter Berücksichtigung dieser Tatsache können Hypothesen zum Stichprobenmittel \bar{m} getestet werden, ohne die Parameter der Grundgesamtheit zu kennen.

3.2 Grundlagen der Geostatistik

Der Begriff räumliche Statistik wird für einen großen Bereich von statistischen Modellen und Methoden verwendet, mit welchen Analysen räumlich referenzierter Daten durchgeführt werden. Innerhalb der räumlichen Statistik befasst sich die Geostatistik mit der Beschreibung und Analyse räumlicher, ortsabhängiger Daten natürlicher Phänomene, die als Realisationen von Zufallsfunktionen gesehen werden.

Über viele Jahre wurde die Geostatistik vor allem für geologische Datensätze im Bereich der Lagerstättenkunde eingesetzt und weiterentwickelt. Inzwischen hat sie auch in vielen Zweigen der Naturwissenschaften Einzug gefunden, vor allem in den Geo- und Umweltwissenschaften wie z. B. Botanik, Geologie, Geotechnik, Hydrologie, Klimatologie, Meteorologie und Ökologie (HUNSAKER et al., 2001)

3.2.1 Theorie der regionalisierten Variable

Das Grundkonzept der Geostatistik ist eine regionalisierte (ortsabhängige) Variable $Z(x)$, welche Eigenschaften aufweist zwischen zufälliger und deterministischer Variable. Im Gegensatz zur zufälligen Variable, besitzen die ortsabhängigen Variablen Kontinuität (Stetigkeit) von Punkt zu Punkt, aber die Änderungen der Variablen sind so komplex, dass sie nicht mit einer deterministischen Funktion allein beschrieben werden können. Die ortsabhängige Variable $Z(x)$, die die Werte einer Messung in Abhängigkeit vom Ort x darstellt, wird als Zufallsvariable betrachtet und damit als Ergebnis eines Zufallsprozesses aufgefasst. Die Messungen an einem Ort stellen demnach Realisationen der Zufallsvariable dar. Eine Behandlung solcher Daten mittels klassischer Methoden der induktiven Statistik würde zu einer systematisch fehlerhaften Analyse führen, da die Daten nicht unabhängig sind, sondern aufgrund ihrer Lagebeziehung eine gewisse Redundanz im Informationsgehalt aufweisen (GAU, 2010).

Quantifiziert werden kann diese graduelle Abhängigkeit mit der *Theorie der regionalisierten Variablen*, die vom französischen Mathematiker MATHERON (1963, 1965) unter Verwendung der empirischen Arbeiten von KRIGE (1951) entwickelt wurde. Die regionalisierte Variable lässt sich in eine deterministische, eine autokorrelative und eine rein zufällige Komponente unterteilen. Die Zufallsfunktion beschreibt die gegenseitige Abhängigkeit der einzelnen Werte, d. h. deren *Autokorrelation*. Die Modellierung der regionalisierten Variable als Zufallsfunktion ist analog dem Konzept einer Zeitreihenanalyse mit dem Unterschied, dass die Zufallsfunktion über den zwei- bzw. dreidimensionalen Raum anstatt über die Zeit definiert ist. Zu jedem Ort innerhalb des Untersuchungsgebietes existiert eine Zufallsvariable. Die Beobachtungen im Untersuchungsgebiet bilden eine einzige Zufallsrealisation, analog wie eine beobachtete Zeitreihe eine einzelne Realisation der gesamten Zeitreihen ist. Dieses theoretische Konstrukt erlaubt es die regionalisierte Variable mit wahrscheinlichkeitstheoretischen Methoden zu behandeln, auch wenn nur eine einzige Beobachtung an einem Ort existieren kann (DAVIS, 2002). Das zeitliche Mittel vieler Realisationen kann durch das räumliche Mittel einer Realisation ersetzt werden.

3.2.2 Stationarität

Das theoretische Modell geht davon aus, dass an jedem Datenpunkt eine Verteilungsfunktion $F(x)$ existiert und der Messwert eine zufällige Realisation aus dieser Verteilung darstellt. Dabei ist die Stationarität eine wichtige Annahme, deren Existenz eine unabdingbare Voraussetzung ist. Es sind unterschiedliche Stationaritätsannahmen verwendbar, die in Bezug auf ihre Strenge qualitativ abgestuft sind.

3.2.2.1 Strenge Stationarität

Die strenge Stationarität besagt, dass die Verteilungsfunktion der Zufallsvariable $Z(x)$ im Untersuchungsgebiet unveränderlich ist, das heißt, dass in allen Punkten des Raumes die gleiche, aber unbekannte Verteilungsfunktion während des Prozesses vorliegt, aus der an jedem Punkt ein Wert realisiert wird. Für alle Beobachtungen des Zufallsprozesses $Z(x)$ gilt, dass sie aus einer Wahrscheinlichkeitsdichte (Dichtefunktion) stammen und nicht von verschiedenen Verteilungen an einzelnen Orten. Bei Nachweis der Momente der Verteilung gilt Stationarität, welche aber bei begrenzten empirischen Daten nicht verifiziert werden kann.

3.2.2.2 Stationarität zweiter Ordnung

Die Stationarität zweiter Ordnung (schwache Stationarität) stellt eine abgeschwächte Form der Stationarität dar, bei der lediglich die beiden ersten Momente Mittelwert und Standardabweichung der jeweiligen Verteilung über den gesamten Raum gleich, d. h. unabhängig von der Datenpunktlage sind. Die Beschreibung der meisten Verteilungen ist anhand von Mittelwert und Standardabweichung für praktische Fälle ausreichend. Mit anderen Worten muss der Erwartungswert von $Z(x)$ für alle Punkte x konstant sein und die Kovarianzfunktion zwischen zwei beliebigen Punkten x und $x+h$ nur vom Vektor h und nicht vom Punkt x abhängen. Somit soll gelten:

$$E[Z(x)] = \bar{m} \quad (3.28)$$

und

$$Cov[x, x+h] = Cov[h] = E[Z(x) \cdot Z(x+h)] - \bar{m}^2 \quad (3.29)$$

3.2.2.3 Intrinsische Hypothese

Die Stationaritätsanforderung lässt sich in Form der so genannten **intrinsischen Hypothese** weiter abschwächen, die besagt, dass die Differenz von Werten der Messgrößen, deren Datenpunkte um den Vektor h voneinander entfernt liegen, nur eine Funktion dieses Abstandsvektors und nicht abhängig vom Ort selbst sein darf. Mit anderen Worten wird nur noch gefordert, dass der Erwartungswert der Differenz der Werte $Z(x+h)$ und $Z(x)$ gleich null ist und dass diese beiden Punkte für alle h eine endliche Varianz aufweisen. Die Erfüllung der intrinsischen Hypothese bezieht sich damit nur auf die Inkremente der Zufallsfunktion. Sie wird daher auch als Zuwachstationarität zweiter Ordnung bezeichnet. Die Inkremente der Funktion

werden als schwach stationär angenommen, d. h. Erwartungswert (Mittelwert) und Varianz von $Z(x+h)-Z(x)$ existieren und sind unabhängig vom Ort x :

$$E[Z(x+h) - Z(x)] = 0, \quad (3.30)$$

für die Varianz

$$\text{Var} [Z(x+h) - Z(x)] = E[Z(x+h) - Z(x)]^2 = 2\gamma(h) \quad (3.31)$$

und schließlich

$$\gamma(h) = \frac{1}{2} \text{Var}[Z(x+h) - Z(x)]. \quad (3.32)$$

Die Funktion $\gamma(h)$ wird als Semivariogramm (kurz Variogramm) bezeichnet. Sie ist das Grundwerkzeug für Strukturanalysen von Phänomenen als auch für Schätzungen. Regionalisierte Variablen, die stationär sind, erfüllen immer die intrinsische Hypothese aber nicht umgekehrt. Später werden wir sehen, falls eine regionalisierte Variable stationär ist, gibt es eine Äquivalenz zwischen ihrem Variogramm $\gamma(h)$ und der Kovarianz $C(h)$.

3.2.3 Variographie

Der Begriff Variographie hat sich für Methoden etabliert, die der Festlegung der Grenzen der homogenen Zone oder der Nachbarschaft für die Schätzung und der rechnerischen Ermittlung der vorhandenen Autokorrelation dienen. Allgemein lässt sich für den Ablauf von Prozessen festhalten, dass Zustände eines Objektes bzw. seiner räumlichen Nachbarn die zukünftigen Zustände des Objektes beeinflussen können, indem ein Transfer Information zwischen räumlich benachbarten Objekten stattfindet. Allerdings kann der Zustand eines Objektes nicht vorausgesagt werden (deterministische Abhängigkeit), sondern nur mit einer gewissen Wahrscheinlichkeit (stochastische Abhängigkeit). Generell lassen sich zwei unterschiedliche Steuerungskomponenten ausmachen:

- exogene Steuerung
Gewisse Faktoren außerhalb des eigentlichen Prozesses wirken auf den Ablauf der Prozessvariablen ein.
- endogene Steuerung
Die Entwicklung des Prozesses steuert aus sich heraus. Es findet ein prozessimmanenter Transfer Information auf benachbarte Raumeinheiten

statt. Dieser Transfer bewirkt, dass räumlich benachbarte Objekte ähnliche bzw. stark entgegengesetzte (alternierende) Werte der Prozessvariablen besitzen. Die Tendenzen benachbarter Lokalitäten setzen sich also fort. Dieses Phänomen bei Prozessen nennt man Erhaltungsneigung. Statistisch wirkt sich die Erhaltungsneigung als stochastische Abhängigkeit aus.

3.2.3.1 Autokorrelation

Als erster Schritt vor einer quantitativen Bestimmung einer räumlichen Erhaltungsneigung muss eine Festlegung der Nachbarn erfolgen, wobei die Nachbarschaft richtungsabhängig definiert wird, da sich für verschiedene Richtungen unterschiedliche Erhaltungsneigungen ergeben können. Die Festlegung der Nachbarschaft im Raum geschieht durch Raumschrittweiten $h \geq 1$. Bei räumlichen Daten bzgl. einer Raumschrittweite k üben mehrere Nachbarn Einfluss auf die Ausprägung in der betreffenden Raumeinheit i aus. Der Gesamteinfluss aller Nachbarn bzgl. der Raumschrittweite k lässt sich dann durch die gewichtete Summe und über alle Punkte und deren Nachbarn über die Autokorrelationsfunktion mittels der $I(h)$ -Koeffizienten bestimmen (BAHRENBERG et al., 2008):

$$I(h) = \frac{\frac{\sum_{i=1}^n \sum_{j=1}^n \omega_{ij}^{(h)} x_i x_j}{\sum_{i=1}^n \sum_{j=1}^n \omega_{ij}^{(h)}}}{\frac{\sum_{i \in L(h)} x_i^2}{n_h}} \quad (3.33)$$

mit $L(h) = \{j \mid j \text{ hat Nachbarn in der Raumschrittweite } h\}$

n_h = Zahl der Raumeinheiten mit Nachbarn der Raumschrittweite k

$\omega_{ij}^{(h)}$ Gewichte ($1 \leq i, j \leq n$, n = Anzahl der Raumelemente)

$\omega_{ij}^{(h)} = \begin{cases} > 0 & \text{falls für } k \text{ die Raumeinheit } j \text{ zu } i \text{ benachbart ist} \\ = 0 & \text{sonst.} \end{cases}$

$x_i(h)$... Ausprägung der Variablen X in der Raumeinheit i

d .h. $I(k)$ ist das gewichtete arithmetische Mittel der kreuzweise multiplizierten Werte der Variablen X in Relation zu dem Mittel der quadrierten x -Werte. Die Signifikanz der räumlichen Autokorrelationsfunktion gegen die Hypothese der stochastischen Unabhängigkeit lässt sich prüfen.

In Zusammenhang mit der Untersuchung räumlicher Autokorrelationen ist die Stationarität bzw. Instationarität eine wichtige Eigenschaft. Man spricht von einem stationären raumvarianten Muster, wenn für jede Raumeinheit i das Mittel der

möglichen Prozessrealisationen identisch ist und die Kovarianz zwischen den Wertvariablen für die Raumeinheit i und derjenigen für die h Raumschritte entfernte Raumeinheit j nicht von den beiden Raumeinheiten, sondern ausschließlich von der Raumschrittweite h abhängt. Räumliche Strukturen, in denen die Werte in einer Richtung ansteigen, die also einen linearen Trend aufweisen sind instationär und die räumlichen Autokorrelationskoeffizienten führen in einem solchen Fall zu Fehlschätzungen. Räumliche Trends lassen sich durch sogenannte Trendoberflächenanalysen extrahieren. Es gibt zahlreiche Verfahren, um räumliche Autokorrelation zu bestimmen. Einfache Indizes, die herauszufinden versuchen, ob räumliche Autokorrelation für eine Variable im Untersuchungsgebiet existiert, sind der **MORAN's I-Index** und der **GEARY's C-Index**. Das Grundprinzip beider Indizes ist gleich, unterschiedlich ist die Interpretation des Begriffes der räumlichen Ähnlichkeit. Räumliche Autokorrelation ist in beiden Fällen das gewichtete Mittel zwischen den Punkten. Bei MORAN's I-Index wird die räumliche Ähnlichkeit über die Abweichung der Punkte vom Mittelwert zum Ausdruck gebracht:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n \lambda_{ij} (x_i - \bar{m})(x_j - \bar{m})}{\sum_{i=1}^n (x_i - \bar{m})^2 \left(\sum_{i=1}^n \sum_{j=1}^n \lambda_{ij} \right)} \quad (3.34)$$

mit n ... Anzahl der Raumelemente
 x_i ... Variablenwert an der Position i
 x_j ... Variablenwert an der Position j
 \bar{m} ... Mittelwert der Variable x
 λ_{ij} ... Gewichte = $1/d_{ij}$ (inverse Distanzgewichtung, je näher desto ähnlicher)

Bei Stationarität ist der Wert des MORAN's Index normiert und liegt zwischen -1 und +1. Werte kleiner als null zeigen negative, Werte größer als null positive Autokorrelation an, d. h. dass bei der Raumschrittweite h ähnliche Ausprägungen wie in der Ausgangsregion zu erwarten sind.

Zur Bestimmung des GEARY's C-Index dagegen wird Ähnlichkeit über die quadrierte Differenz der Punktpaare ausgedrückt:

$$C = \frac{(n-1) \left[\sum_{i=1}^n \sum_{j=1}^n \lambda_{ij} (x_i - x_j)^2 \right]}{2 \sum_{i=1}^n (x_i - \bar{m})^2 \left(\sum_{i=1}^n \sum_{j=1}^n \lambda_{ij} \right)} \quad (3.35)$$

Typischerweise variieren die Werte zwischen 0 und 2. Der theoretische Wert von C ist 1, das bedeutet, dass die Raumeinheiten unabhängig sind. Werte kleiner als 1 zeigen positive räumliche Autokorrelation, Werte größer als eins negative. Der C-Index ist somit in inverser Beziehung zum MORAN's I-Index. Der MORAN's I-Index stellt einen mehr globalen Indikator da, während der GEARY's C-Index sensibler auf kleine Nachbarschaften reagiert. Die Gewichte werden meistens für kleine Distanzen

angepasst, sodass das maximale Gewicht keinen Wert größer als eins annehmen kann. Ansonsten würden Distanzen gegen null ein sehr großes Gewicht erhalten, wodurch die Indizes verzerrt werden.

3.2.3.2 Empirisches Variogramm

Eine andere Möglichkeit der Analyse von Autokorrelation ist die Analyse mithilfe des Variogramms. In praktischen Fällen kann das Variogramm nur bis zu einer bestimmten Distanz verwendet werden. Daher muss das Phänomen nur innerhalb dieser Distanz stationär sein. Das Problem liegt darin ob eine Serie von Nachbarnschaften gefunden werden kann in welcher der Erwartungswert und das Variogramm als konstant angesehen werden können und genug Daten für eine aussagekräftige Schätzung vorhanden sind. Die Annahme der Quasi-Stationarität ist ein Kompromiss zwischen dem Umfang der Homogenität des Phänomens und der Stichprobendichte. Es kann ein Phänomen lokal stationär oder zumindest intrinsisch sein, obwohl es nicht stationär über größere Reichweiten ist.

Die im Zuge der Einführung der Theorie der regionalisierten Variable unterstellte Autokorrelationsstruktur nach Gl. (3.32) bleibt unbekannt und kann lediglich geschätzt werden. Hierzu erfolgt die Bestimmung der (Semi-)Varianz von Punktpaaren, die innerhalb einer bestimmten Entfernung voneinander liegen nach

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} (Z(x_i) - Z(x_j))^2$$

bzw.

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} (Z(x_i) - Z(x_i + h))^2 \quad (3.36)$$

mit $n(h)$ Anzahl der Punkte mit Entfernung h
 h Entfernung zwischen i und j
 $Z(x_i)$ (Mess-)Wert der Variablen X am Ort i
 $Z(x_j)$ (Mess-)Wert der Variablen X am Ort j

Im ersten Schritt werden also für alle Datenpunkte die Distanzen zu allen anderen Datenpunkten im Untersuchungsgebiet über die quadratische euklidische Distanz bestimmt. Daran wird unmittelbar ersichtlich, dass mit zunehmender Anzahl an Datenpunkten der Prozess der Distanzbestimmung rasch zunimmt und eine Interpretation der entstandenen Punktwolke im Variogramm zunehmend schwieriger oder gar unmöglich wird. Aus diesem Grund werden Entfernungsintervalle h gebildet und die Punktpaare gruppiert. Anschließend werden die entfernungsabhängigen

Variogrammwerte (Varianzen) gegen die jeweilige Entfernung (Schrittweite) h aufgetragen und ergeben so das empirische oder experimentelle Variogramm (Abb. 3.6).

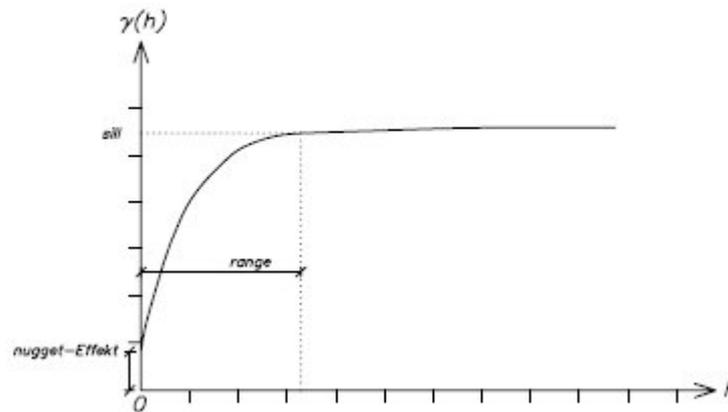


Abb. 3.6: (Semi-)Variogramm

Beim eigentlichen Variogramm werden keine richtungsspezifischen Unterschiede dargestellt, sondern nur das Verhältnis zwischen Semivarianz und Distanz. Allgemein steigt das Variogramm im Ursprung relativ steil an, d. h. die Messwerte benachbarter Datenpunkte haben eine starke räumliche Autokorrelation. Ab einer bestimmten Entfernung, der **Reichweite a (range)**, flacht es ab und erreicht einen **Schwellwert c (sill)**. Dieser entspricht theoretisch der Gesamtvarianz σ_0^2 der Daten. Erreicht das Variogramm keinen Schwellwert, kann dies darauf hindeuten, dass die Daten trendbehaftet sind. Der range a kennzeichnet den maximalen Einflussbereich der Korrelation von Datenpunkten. Definitionsgemäß muss die Variogrammfunktion durch den Ursprung verlaufen ($\gamma(0) = 0$), oft ergibt sich aber ein Wert verschieden von null, der als **Nugget-Effekt (c_0)** bezeichnet wird. Dies resultiert entweder aus Messfehlern der Datenpunkte oder deutet auf eine Mikrovariabilität der Datenpunkte hin, die mittels der Beprobungsabstände nicht zu erfassen ist.

Bevor das Variogramm im Detail betrachtet wird, seien noch drei wichtige Eigenschaften der räumlichen Kovarianz und ihrer Beziehung zum Variogramm stationärer Zufallsfunktionen angeführt:

$$\begin{aligned}
 C(0) &= \sigma^2 \\
 C(h) &= C(-h) \\
 |C(h)| &\leq C(0)
 \end{aligned}
 \tag{3.37}$$

Im Fall der Erfüllung von zumindest Stationarität zweiter Ordnung sind Kovarianz C und Variogramm γ gleichwertige Möglichkeiten der Strukturbeschreibung und können nach Gl. (3.38) ineinander übergeführt werden:

$$\begin{aligned}\gamma(h) &= C(0) - C(h) = \sigma^2 - C(h) \text{ oder} & (3.38) \\ C(h) &= \sigma^2 - \gamma(h)\end{aligned}$$

Das zeigt, dass die korrespondierende Kovarianz durch Spiegelung des Variogramms erhalten wird (Abb. 3.7).

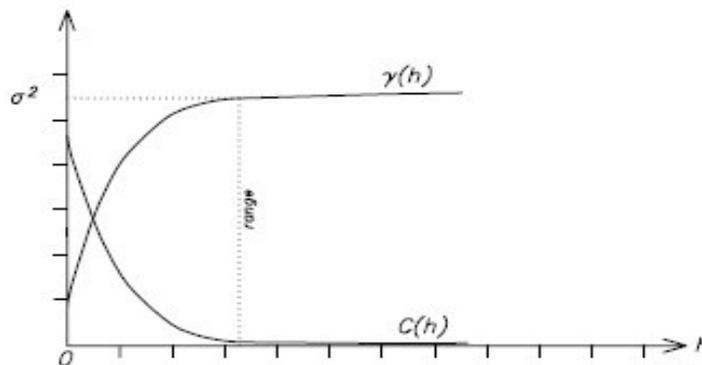


Abb. 3.7: Kovarianz/Variogramm

Es kann mathematisch gezeigt werden, dass Variogramme mit einer Obergrenze von stationären regionalisierten Variablen stammen. Genauer ist zu sagen nur stationäre regionalisierte Variablen haben begrenzte Variogramme. Daher stammen unbegrenzte Variogramme von nicht stationären oder intrinsischen regionalisierten Variablen. Zuletzt kann noch gezeigt werden, dass für stationäre und intrinsische Variablen gilt:

$$\lim_{h \rightarrow \infty} \frac{\gamma(h)}{h^2} = 0 \quad (3.39)$$

Das bedeutet falls das Variogramm mehr als quadratisch für große h steigt, dann ist die Variable nicht stationär. Andernfalls kann sie als stationär oder intrinsisch betrachtet werden. Das ist hilfreich um zu entscheiden ob eine Variable als stationär oder intrinsisch oder nicht stationär angesehen werden kann.

3.2.3.3 Theoretisches Variogramm

Da das experimentelle Variogramm nur an diskreten Punkten mit dem Entfernungsintervall h bekannt ist muss zur Ermittlung der räumlichen Struktur und der Schätzung an einem beliebigen Punkt eine kontinuierliche Funktion modelliert werden. Dazu wird das empirische Variogramm an ein theoretisches Variogramm angepasst. Die mathematische Funktion muss gewissen Bedingungen genügen, um

eine nicht negative Varianz zu erhalten. Positiv definite Funktionen erfüllen bei stationären Variablen diese Bedingung, die Situation ist bei intrinsischen Variablen leicht verschieden. Die am meisten verbreiteten Schätzer sind Linearkombinationen der Daten und im Fall der intrinsischen Variablen muss die Varianz einer beliebigen Linearkombination nicht existieren, sondern nur die Linearkombination von Inkrementen. Kombinationen werden als zulässig bezeichnet, wenn die Summe ihrer Gewichte gleich null ist:

$$\sum \lambda_i = 0. \quad (3.40)$$

Diese Bedingung ist schwächer als die vorhergehende für Kovarianzen, welche für alle möglichen Gewichte halten muss. Daher sind die zulässigen Variogrammodelle für die intrinsische Hypothese umfangreicher, sie beinhalten begrenzte Variogramme verbunden mit Kovarianzen als auch unbegrenzte Variogramme, welche keine Kovarianz als Gegenstück besitzen. Für die stationäre Hypothese ist die Anzahl der Variogramme mehr beschränkt, dafür können beliebige Gewichtungsfaktoren verwendet werden.

Da es nicht leicht ist Funktionen zu erkennen, die die notwendigen Eigenschaften besitzen oder nach diesen zu testen, ist es das beste Variogrammodelle aus einer verfügbaren Auswahl von brauchbaren Funktionen zu wählen. Generell kann man intransitive und transitive Modelle unterscheiden, wobei letztere auf die Erfüllung der Stationarität zweiter Ordnung hinweisen, während für die intransitiven Modelle lediglich die Erfüllung der intrinsischen Hypothese als gegeben angenommen werden darf. Als wichtigste intransitive Modelle sind das *Potenzmodell* (Abb. 3.8g) nach Gl. (3.41)

$$\gamma(h) = c \cdot h^\alpha \quad \text{mit } c > 0 \text{ und } 0 < \alpha \leq 2, \quad (3.41)$$

wobei sich für $\alpha = 1$ das *lineare Modell* (Abb. 3.8f) ergibt, oft modifiziert durch einen Bruch im sill

$$\gamma(h) = \begin{cases} c \cdot h & \text{für } h < a \\ c & \text{für } h \geq a \end{cases}, \quad (3.42)$$

da sie keinen Schwellwert aufweisen und das *logarithmische Modell* nach Gl. (3.43) zu nennen.

$$\gamma(h) = 3\alpha \log h \quad (3.43)$$

Einige Autoren kritisieren das lineare Modell in Hinblick auf Kriging, da dort das Variogramm eine durchgehend glatte Funktion sein muss. Vorausgesetzt das Variogramm wird nur für Distanzen sehr viel geringer als die Reichweite berechnet, dann ist das lineare Modell eine ausgezeichnete Approximation.

Die Gruppe der transitiven Modelle ist ungleich umfangreicher und der Aufbau der Funktionen unterscheidet sich deutlich voneinander. Zu den gebräuchlichsten Modellen, die auch in ArcGIS® Anwendung finden, zählen das *sphärische* Modell (Gl. 3.44), das *exponentielle* Modell (Gl.3.45) und das *Gaußsche* Modell (Gl. 3.46):

$$\gamma(h) = \begin{cases} c \cdot \left(\frac{3h}{2a} - \frac{h^3}{2a^3} \right) & \text{für } 0 \leq h < a \\ c & \text{für } h \geq a \end{cases} \quad (3.44)$$

$$\gamma(h) = c \cdot (1 - e^{-3h/a}) \quad (3.45)$$

$$\gamma(h) = c \cdot (1 - e^{-3h^2/a^2}) \quad (3.46)$$

Das sphärische Modell (Abb. 3.8a) wird üblicherweise als ideale Form für das Variogramm bezeichnet. Das Modell erreicht in der Reichweite a den Schwellwert c und zeigt lineares Verhalten nahe dem Ursprung, also für kleine Distanzen. Das exponentielle Modell nähert sich dem Schwellwert c asymptotisch, der die praktische Reichweite ist $3a$, das ist die Distanz wenn die Semivarianz 95% des Schwellwertes erreicht. Das Modell zeigt ebenfalls lineares Verhalten für kleine Distanzen. Das exponentielle und Gaußsche Modell (Abb. 3.8b bzw. 3.8c) leiten sich als Spezialfall aus der Familie der robusten (*stable*) Modelle (Abb. 3.8d) mit $p=1$ bzw. $p=2$ (Gl. 3.47) ab.

$$\gamma(h) = c \cdot (1 - e^{-3h^p/a^p}) \quad \text{mit } 0 < p \leq 2 \quad (3.47)$$

Das Gaußsche Modell nähert sich dem Schwellwert ebenfalls asymptotisch, die praktische Reichweite ist $1.73a$. Nahe dem Ursprung zeigt es quadratisches Verhalten und das Modell repräsentiert sehr kontinuierliche Phänomene. Eine rein zufällig Variable ohne räumlich Autokorrelation besitzt keine Kontinuität und ihr Variogramm wäre eine horizontale Linie, die der Gesamtvarianz σ_0^2 entspricht und entspricht dem *Nugget*-Effekt (Abb. 3.8e):

$$\gamma(0) = \begin{cases} 0 & h = 0 \\ c & \text{sonst} \end{cases} \quad (3.48)$$

Zu diesen auf beliebigen Datensätzen anwendbaren Modellen sind eine Reihe weiterer Modelle bekannt, die im Rahmen von geologischen und geophysikalischen Untersuchungen Anwendung finden (vgl., ARMSTRONG 1998). Ohne auf diese näher einzugehen, seien die folgenden Modelle angeführt:

Kubisch:

$$\gamma(h) = \begin{cases} c \cdot \left(7 \frac{h^2}{a^2} - 8.75 \frac{h^3}{a^3} - 3.5 \frac{h^5}{a^5} - 0.75 \frac{h^7}{a^7} \right) & \text{für } \frac{h}{a} < 1 \\ c & \text{sonst} \end{cases} \quad (3.49)$$

2D-Loch Effekt (Abb. 3.8h):

$$\gamma(h) = c \cdot \left(1 - e^{-\frac{h}{a}} \right) J_0 \left(2\pi \frac{h}{\lambda} \right) \quad \text{mit } J_0 \text{ als Besselfunktion} \quad (3.50)$$

Kardinal Sinus (Abb.3.8i):

$$\gamma(h) = c \cdot \left(1 - \frac{\sin \frac{h}{a}}{\frac{h}{a}} \right) \quad (3.51)$$

Prismato-Magnetisch:

$$\gamma(h) = c \cdot \left(1 - \frac{1}{\left(1 + \frac{h^2}{a^2} \right)^{1.5}} \right) \quad (3.52)$$

Prismato-Gravimetrisch:

$$\gamma(h) = c \cdot \left(1 - \frac{1}{\left(1 + \frac{h^2}{a^2} \right)^{0.5}} \right) \quad (3.53)$$

Die angepasste kontinuierliche Variogrammfunktion liefert für jede beliebige Distanz einen Variogrammwert, der bei der optimalen Bestimmung der Schätzwichte für Schätzverfahren Berücksichtigung findet. Die Modelle können addiert werden, weil dies äquivalent der Addition von unabhängigen Zufallsfunktionen ist. Subtraktion ist jedoch nicht erlaubt, weder dürfen sie stückweise kombiniert werden, womit gemeint ist, dass ein Modell bis zur einer gewissen Distanz gewählt wird und ein weiteres von dieser Distanz an.

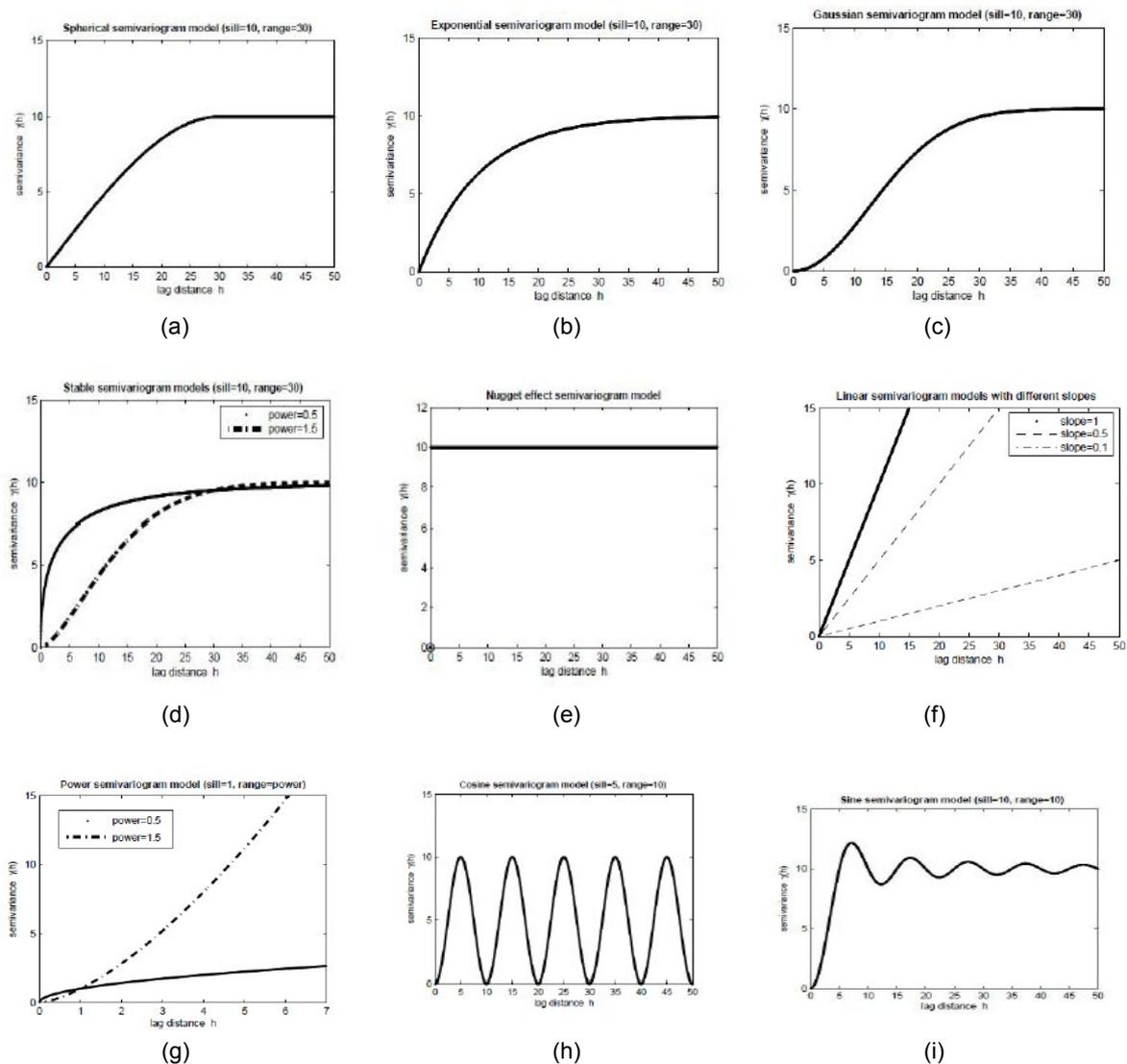


Abb. 3. 8: Variogrammmodelle - (a) Sphärisches Modell, (b) Exponentielles Modell, (c) Gaußsches Modell, (d) Robustes Modell, (e) Nugget Modell, (f) Lineares Modell, (g) Potenz Modell, (h) 2D-Loch Effekt, (i) Kardinal-Sinus Modell (alle aus FUCHS, 2011)

3.2.3.4 Eigenschaften von Variogrammen

3.2.3.4.1 Verhalten im Ursprung

Die verschiedenen theoretischen Variogrammmodelle unterscheiden sich hinsichtlich ihres Verhaltens im Ursprung, in der Nähe der Reichweite a , durch den Schwellwert c und bei sehr großen Abständen h . Der Kurvenverlauf kann als Erhaltungseigung interpretiert werden, der Anstieg kann als Stetigkeit bzw. Glattheit der Variablen im Raum aufgefasst werden und die Reichweite a als diejenige Schrittweite gedeutet werden, bei der der Übergang von strukturell bedingter zu zufälliger Variabilität

auftritt. Noch wichtiger als das Verhalten der Variogramme für große Entfernungen ist die Kenntnis des Verhaltens für kleine Distanzwerte h , weil sich dies auf die Kontinuität und die räumliche Regelmäßigkeit der Variablen bezieht. Man kann vier wichtige Eigenschaften nahe des Ursprunges unterscheiden:

1. *Quadratisch*: Zeigt eine hohe Kontinuität der regionalisierten Variablen.
2. *Linear*: Die regionalisierte Variable ist kontinuierlich, aber weniger regelmäßig wie die quadratische Eigenschaft.
3. *Diskontinuität*: $\gamma(h)$ geht nicht gegen null mit h gegen null
4. *Flach*: Reine Zufälligkeit, die regionalisierten Variablen $Z(x+h)$ und $Z(x)$ sind nicht korreliert für alle Werte von h , egal wie nahe sie aneinanderliegen. Dies ist der begrenzte Fall einer fehlender Struktur und ist das angenommene Modell in der Trendoberflächenanalyse.

3.2.3.4.2 Anisotropie

Wenn das Variogramm in verschiedene Richtungen berechnet wird, kann es sich verschieden in diesen verhalten. Falls dies nicht auftritt, hängt das Variogramm nur zwischen den zwei Punkten ab und wird isotrop genannt. Bezüglich der Anisotropie können die geometrische und die zonale Anisotropie unterschieden werden.

Geometrische Anisotropie:

Anisotropie herrscht vor, wenn die Reichweite (begrenzte Variogramme) oder die Steigung (unbegrenzte Variogramme) eine Funktion der Richtung ist. Wenn das Diagramm eine Ellipse ergibt (in 2D), dann wird die Anisotropie als geometrisch bezeichnet. Eine einfache Transformation kann die Ellipse in einen Kreis umrechnen und die Anisotropie beseitigen. Diese ist besonders einfach, falls die Hauptachsen der Ellipse mit den Koordinatenachsen zusammenfallen. Wenn die Gleichung des Variogramms in die erste Richtung $\gamma_1(h)$ ist, dann ergibt sich für das gesamte Variogramm nach Korrektur der Anisotropie die Form

$$\gamma(h) = \gamma_1\left(\sqrt{h_1^2 + k^2 h_2^2}\right) \quad (3.54)$$

wobei h_1 und h_2 die zwei Komponenten von h und k das Anisotropieverhältnis bezeichnet ($k=a_1/a_2$ oder *Steigung₁/Steigung₂*.)

Zonale Anisotropie:

Ist eine komplexere Form der Anisotropie. Zwischen Zonen ist die Variation oftmals größer als innerhalb dieser. In diesen Fällen ist der Schwellwert der Variogramme nicht in allen Richtungen derselbe. In der Praxis wird das Variogramm in zwei Kom-

ponenten unterteilt, in eine isotrope und eine weitere, die nur die wirkende Richtung berücksichtigt:

$$\begin{aligned}
 \text{isotrope Komponente:} & \quad \gamma_0 = \sqrt{h_1^2 + h_2^2} \\
 \text{wirkende Komponente:} & \quad \gamma_1(h_2) \\
 \text{Gesamtvariogramm:} & \quad \gamma(h) = \gamma_0(h) + \gamma_1(h)
 \end{aligned} \tag{3.55}$$

3.2.3.4.3 Drift

Wie bereits behandelt, zeigt die Theorie, dass das Variogramm für intrinsische und stationäre Variablen weniger schnell als quadratisch für große Distanzen anwächst (Gl. 3.39). In der Praxis ergibt sich aber oft, dass Variogramme schneller als h^2 anwachsen, dadurch nicht begrenzt sind und Änderungen des Mittelwerts in Anhängigkeit vom Ort anzeigen. Dies deutet auf eine vorhandene Drift bzw. einen Trend in den Daten hin, wodurch das empirische Variogramm das zugrunde liegende Variogramm überschätzt (ARMSTRONG, 1998). Die (Semi-)Varianz Gl. (3.36) kann auch in der Form

$$\gamma(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} \left\{ (Z(x_i) - Z(x_i + h)) - \frac{\sum_{i=1}^{n(h)} Z(x_i) - Z(x_i + h)}{n(h)} \right\}^2 \tag{3.56}$$

formuliert werden. Falls die regionalisierte Variable stationär oder intrinsisch ist, ist der Mittelwert am Ort x_i gleich dem Mittelwert am Ort $(x_i + h)$, d. h.

$$\frac{\sum Z(x_i)}{n(h)} = \frac{\sum Z(x_i + h)}{n(h)} \quad \text{oder} \quad \frac{\sum Z(x_i) - \sum Z(x_i + h)}{n(h)} = 0 \tag{3.57}$$

Eingesetzt in Gl. (3.56) verschwindet der zweite Teil, da dieser Null ergibt und man erhält Gl. (3.36). Falls die regionalisierte Variable jedoch nicht stationär oder intrinsisch ist, ergibt sich für den zweiten Term ein Wert. Faktisch kann die regionalisierte Variable als aus zwei Teilen zusammengesetzt angesehen werden, nämlich der Drift und den Residuen. Wenn man die Drift subtrahiert, dann sind die Residuen

$$R_i = Z(x_i) - \bar{m}_i \tag{3.58}$$

selbst regionalisierte Variablen mit einem lokalen Mittelwert null (bzw. der Mittelwert der Inkremente), mit anderen Worten, die Residuen sind stationär bzw. intrinsisch und damit ist es möglich das Variogramm zu berechnen.

3.2.3.4.4 Proportionale Effekte

Ein Variogramm besitzt einen proportionalen Effekt, wenn der Maximalwert proportional zum Quadrat des lokalen Mittelwertes ist. Das ist oft bei lognormal verteilten Daten der Fall. Das Variogramm hat für verschiedene Zonen dieselbe Form, aber der Maximalwert ist in reichhaltigen Zonen viel größer als in armen. Das zugrunde liegende Variogramm kann dann gefunden werden, indem jedes lokale Variogramm durch das Quadrat des lokalen Mittelwertes dividiert wird und dann diese gemittelt werden, bevor ein Variogrammmodell angepasst wird. Anomalien in der lokalen Variabilität haben einen großen Einfluss auf die Genauigkeit einer Interpolation bzw. Schätzung. Jede Schätzung irgendeiner vernünftigen Interpolationsmethode profitiert von geringer und leidet von starker Variabilität (ISAACS & SRIVASTAVA 1989). In der Statistik werden dafür die Begriffe Homo- bzw. Heteroskedastizität verwendet. Um diese zu untersuchen, wird das Untersuchungsgebiet in mehrere gleichgroße Fenster unterteilt und darin statistische Maßzahlen berechnet (engl. *moving window statistics*). Für ungleichmäßig verteilte Daten empfiehlt sich eine Überlappung der Fenster. Neben der Berechnung statistischer Maßzahlen wie Mittelwert, Standardabweichung oder Quartildistanz (vgl. Kap. 3.1.1) sind bekannte Verfahren zur Aufdeckung von Heteroskedastizität der *Goldfeld/Quandt-Test* oder das *Verfahren von Glesjer* (BACKHAUS et al., 2008).

3.2.3.4.5 Loch-Effekt und Periodizität

In manchen Fällen steigt das Variogramm über den Maximalwert (sill) und fällt dann wieder zurück. Dies kann darauf hindeuten, dass Bereiche mit sehr hohen und niedrigen Werten nebeneinander liegen oder dass zu wenig Punktpaare für die Berechnung des experimentellen Variogramms verwendet wurden. Bei vorhandenem Loch-Effekt können sinusförmige Variogrammmodelle (3.50 und 3.51) eingesetzt werden oder die verschiedenen Bereiche müssen gesondert behandelt werden. Variogramme können auch periodisches Verhalten anzeigen. Als natürliches Phänomen findet man dieses bei Zeitreihen, mit anderen Variablen ist dieses seltener und schwieriger zu erklären.

3.2.3.4.6 Verschachtelte Strukturen

Manchmal können aus dem Variogramm verschachtelte Strukturen abgeleitet werden. Dies ist der Fall, wenn zwei Reichweiten aufscheinen. Die lange Reichweite ist ersichtlich ab dem Punkt wo das Variogramm ausflacht, die kurze Reichweite kann aus einer charakteristischen Änderung der Variogrammkurve herausgelesen werden

(Abb. 3.8). Die Änderung ist offensichtlich, wenn die zwei Reichweiten ziemlich verschieden sind. Falls nicht, ist die Änderung eher graduell und nicht eindeutig.

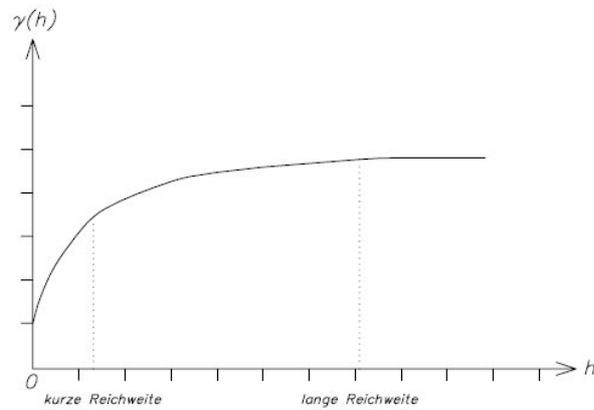


Abb. 3.9: Verschachtelte Struktur

Verschachtelte Strukturen zeigen das Vorhandensein von Prozessen, die auf verschiedenen Maßstäben wirken, an. Das heißt, es wirken verschiedene Variabilitäten in verschiedenen Maßstäben.

4 RÄUMLICHE INTERPOLATION

4.1 Theorie der räumlichen Interpolation

Unter dem Begriff räumliche Interpolation versteht man im GIS-Bereich die Berechnung bzw. Schätzung von Werten an nicht beprobten Orten auf Grundlage punktuell gemessener bzw. erfasster Datenpunkte. Um quantifizierte und genaue Informationen über räumliche Variationen eines zu untersuchenden Phänomens zu erhalten, wird mit räumlich kontinuierlich variierenden Oberflächen gearbeitet, die jedoch nicht flächendeckend beobachtet oder gemessen werden können. Aus meist wenigen punktuellen Stichproben-Messungen sollen Einzelwerte an anderen nicht beprobten Orten über das Untersuchungsgebiet geschätzt werden (Abb. 4.1). Während die Geländeoberfläche ein anschauliches Beispiel für den Typ einer zumeist kontinuierlich variierenden Oberfläche ist, liegen bei vielen thematischen Werteoberflächen (Klima- und Wetterdaten, Schadstoffkonzentrationen, Grundwasseroberfläche, usw.) kompliziertere Verhältnisse vor. Aufgabe der Interpolation ist es, Näherungswerte für einen unbekanntem Ort zu ermitteln bzw. aus den wenigen punktuellen Messungen ein möglichst gutes Bild und Verständnis der räumlichen Variation thematischer Oberflächen in einem Untersuchungsgebiet zu gewinnen.

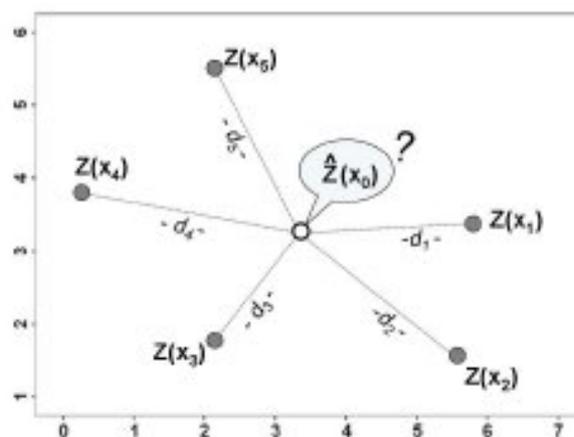


Abb. 4.1: Grundaufgabe der Interpolation (PETERS, 2008)

Die Grundannahme der verschiedenen Interpolationsverfahren ist eine räumliche Ähnlichkeit zwischen den einzelnen benachbarten Werten in dem Sinn, dass bei räumlich naheliegenden Messpunkten eine größere Ähnlichkeit zu verzeichnen ist, als bei Messpunkten, die weit voneinander entfernt sind. Die Qualität der interpolierten Ergebnisse hängt von der Genauigkeit, der Anzahl und von der Verteilung der in die Berechnung eingehenden Messpunkte ab. Günstig ist es, wenn

die Messpunkte relativ gleichmäßig über den Untersuchungsraum verteilt sind, aber auch eine unregelmäßige Anordnung ist für eine Interpolation möglich. Grundsätzlich ist zu berücksichtigen, dass Interpolationen nicht besser sein können als die zugrunde liegenden Daten und jede Schätzung nur eine mehr oder weniger gelungene Annäherung an den „wahren“ Wert ist.

Die Interpolationsmethoden können in deterministische und stochastische Verfahren unterteilt werden. Deterministische Verfahren zeichnen sich dadurch aus, dass sie entsprechend der räumlichen Nähe der Messpunkte diese zueinander gewichten. Im Unterschied dazu berücksichtigen stochastische Verfahren zusätzlich die korrelativen Beziehungen der Messpunkte, also die Existenz räumlicher Autokorrelationen. Ein weiterer wesentlicher Unterschied ist, dass deterministische Verfahren die Ähnlichkeit benachbarter Punkte, rein mithilfe mathematischer Funktionen berechnen, wohingegen stochastische Verfahren eine auf statistischen Annahmen beruhende Ableitung der Interpolationsflächen vornehmen. Als großer Vorteil der stochastischen Methoden ist zu nennen, dass mit jeder ermittelten Schätzoberfläche auch eine Qualitätsaussage mit den dazugehörigen Fehleroberflächen verbunden ist.

4.2 Räumliche Anordnung der Stichproben

Die räumliche Verteilung der Datenpunkte (Stichproben) hat entscheidenden Einfluss auf die Qualität und Wahl der Interpolation. Die Stichproben repräsentieren die Grundgesamtheit und sollten deshalb einen hinreichend großen Umfang und räumlich gleichmäßig verteilt sein. Oft spielen aber finanzielle oder organisatorische Aspekte sowie die Zugänglichkeit zu den Örtlichkeiten eine Rolle bzw. ist in vielen Anwendungen die Anzahl und Lage der Stichproben bereits vorgegeben.

Das Verteilungsmuster der Datenpunkte im zweidimensionalen Untersuchungsgebiet kann zufällig (random), gleichmäßig (regular) oder gruppiert (clustered) erscheinen (vgl. Abb. 4.2).

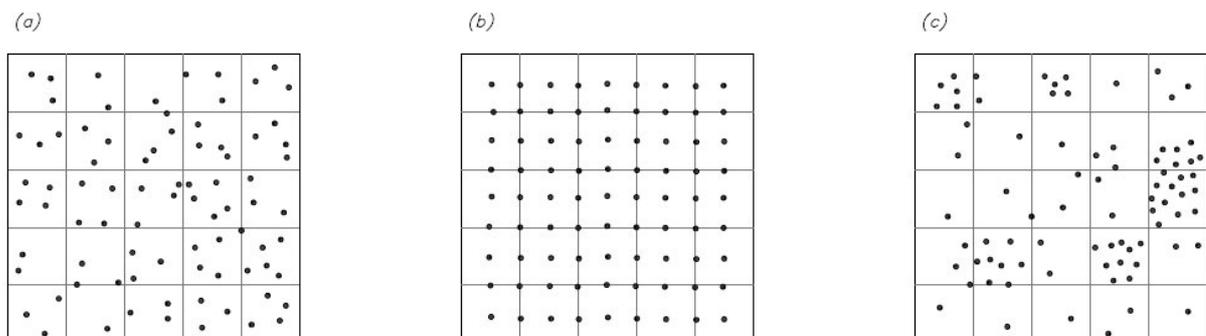


Abb. 4.2: Verteilungsmuster (a) zufällig-random, (b) gleichmäßig-regular, (c) gruppiert-clustered

Unter dem Begriff Point Pattern Analysis (PPA) werden alle Methoden zusammengefasst, die die räumliche Verteilung innerhalb des Interessengebietes untersuchen. Das grundlegende Prinzip beruht auf einem Vergleich des empirischen Verteilungsmusters mit einem theoretischen. Das theoretische Zufallsmuster wird über die Poisson-Verteilung ermittelt. Eine Zufallsvariable heißt poissonverteilt, wenn ihre Wahrscheinlichkeitsfunktion nach

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda} \quad (4.1)$$

mit

λ Anzahl der Punkte

$p(x)$ Wahrscheinlichkeit, dass 0,1,2, ... Punkte im Untersuchungsgebiet theoretisch erwartet werden können.

beschrieben werden kann. Die Verfahren der PPA können in zwei Gruppen unterteilt werden, nämlich in solche die erstrangige Effekte (first-order analysis), und solche, die zweitrangige Effekte (second-order analysis) behandeln. Erstrangige Effekte beziehen sich auf die Intensität, das ist die mittlere Anzahl der Datenpunkte pro Flächeneinheit. Zweitrangige Effekte oder räumliche Abhängigkeit nehmen Bezug auf die Beziehung von gepaarten Datenpunkten, indem die Entfernung zwischen diesen als Information verwendet wird. Die Art der Verteilung (zufällig, gruppiert, gleichmäßig) kann mit verschiedenen Methoden wie der

- Quadrat Methode
- Kernel Estimation (KE)
- Nearest Neighbour Analysis (NNA)
- Kernel Density Function (K-function)

untersucht werden, wobei die zwei erstgenannten der Gruppe der first-order analysis zuzuordnen sind und NNA bzw. K-function der second-order analysis. Eine einfache Methode den Grad der Gleichmäßigkeit oder Gruppiertheit einer Punktverteilung abzuschätzen, ist die Quadratmethode. Dabei wird über die Punktverteilung ein regelmäßiger Raster (Quadrate) gelegt. Die mittlere Anzahl pro Quadrat ist die Anzahl der Datenpunkte dividiert durch die Anzahl der Quadrate:

$$\bar{m}_Q = \frac{n}{n_Q} \quad (4.2)$$

mit

n Anzahl der Stichproben (Datenpunkte)

n_Q ... Anzahl der Quadrate

Die Varianz kann nach Gl. (3.6) berechnet werden und damit das Verhältnis Varianz zur mittleren Anzahl pro Quadrat (variance/mean ratio):

$$VMR = \frac{s^2}{\bar{m}_Q} \quad (4.3)$$

VMR zeigt den Grad an mit welchem ein Punktverteilungsmuster zu einer Poisson-Verteilung abweicht. Für ein zufälliges Verteilungsmuster ist der erwartete Wert eins, ein Wert größer als eins zeigt Gruppiertheit an, ein Wert kleiner als eins deutet auf eine zunehmende Tendenz zu gleichmäßiger Verteilung hin.

Eine ausführliche Beschreibung der weiteren Methoden NNA, KE und K-function ist in BAILEY & GATRELL (1995) oder LLOYD (2010) zu finden.

4.3 Deterministische Interpolationsverfahren

Deterministische Verfahren unterscheiden sich ihrerseits nach zwei Kriterien: global – lokal und exakt – approximativ.

Bei globalen Methoden gehen alle Daten in die Interpolation mit ein, es wird ein Algorithmus für die gesamte Fläche angewendet. Lokale Methoden dagegen beziehen nur ein Teilgebiet, das sind Punkte der näheren Umgebung, in die Berechnung ein. Es ist hierbei zu erwähnen, dass nicht alle Verfahren als global oder lokal gekennzeichnet werden können. Bei Veränderung bestimmter Parameter der Methoden können manche der Verfahren beiden Kategorien zugeordnet werden. Als exakt werden Verfahren bezeichnet, bei denen die Interpolationsfläche exakt durch die Messpunkte verläuft, als approximativ, wenn die Oberfläche nicht durch die vorliegenden Messpunkte verläuft. Immer wenn die an den Messpunkten ermittelten Werten relativ genau, wiederholbar und keine großen Fehler aufweisen, wird ein exakter Interpolator eingesetzt. Sind Messwerte nur ungenau und variieren von Messung zu Messung stark, so wird es sinnvoll sein eine Oberfläche nicht exakt durch diese Punkte zu zwingen, sondern sich den empirischen Messwerten nur anzunähern. Aus der Vielzahl von deterministischen Interpolationsmethoden

- **Polygon-Methoden:**
Thiessen - Polygone, Voronoi - Interpolation oder Natural Neighbour Interpolation (NNI)
- **Triangulation:**
Delauny - Triangulation, TIN (Triangulated Irregular Network)
- **Polynom Interpolation (PI)**
- **Inverse Distance Weighted (IDW)**
- **Spline Interpolation (SI)**

werden im folgenden die, für die Aufgabenstellung verwendeten Methoden PI, IDW und SI beschrieben.

Bei Polygon-Methoden wird nur der nächstgelegene Datenpunkt zur Schätzung an nicht beprobten Orten herangezogen. Daher liefert das Verfahren abrupte Merkmalsänderungen zwischen den einzelnen Polygonen der interpolierten Oberfläche und es kann keine stetige Wertoberfläche erzeugt werden. Es kommt zur Entstehung von Sprungstellen (Artefakten), die nicht den realen Vorgängen entsprechen. Anwendbar ist diese Methode jedoch im Rahmen der explorativen räumlichen Datenanalyse (ERDA) um Nachbarschaftsbeziehungen (Trends, Ausreißer) schnell zu visualisieren (Voronoi - Diagramm).

Die Triangulation (Dreiecksvermaschung) wird durch das Aufspannen von Dreiecken durch drei benachbarte Datenpunkte die Wertoberfläche der Beobachtungsvariablen gebildet. Der Schätzwert an nicht beprobten Orten ist daher nur von diesen drei Datenpunkten anhängig und von den anderen unabhängig. Im Unterschied zur Polygon-Methode werden stetige Wertoberflächen erzeugt, was im Allgemeinen den realen Vorgängen näher kommt. Ein Nachteil dieser Methode kann eine ungünstige Anordnung der Datenpunkte sein, die schmale, lange und spitzwinkelige Dreiecke entstehen lassen. Abhilfe schaffen Nebenbedingungen, die von den Dreiecken erfüllt werden müssen, wodurch die Auswahl der Datenpunkte gesteuert wird. Die bekannteste Methode ist die *Delauny* – Triangulation. Probleme entstehen aber auch in datenarmen Gebieten und sog. kritischen Flächen, die dort auftreten, wo alle Punkte eines Elementardreiecks dieselben Werte aufweisen (BARTELME, 2005). Die Triangulation findet vor allem für Nachbarschaftsanalysen und die Erstellung von Geländemodellen Anwendung.

4.3.1 Polynom Interpolation

Deuten explorative Untersuchungsmuster auf relativ wenig räumliche Unterschiede hin oder zeigt sich ein klarer Trend für das Untersuchungsgebiet, bieten sich polynomische Verfahren (Regressionsverfahren) an. Die Polynom Interpolation passt eine glatte Oberfläche, die durch eine Polynomfunktion definiert ist, an die vorgegebenen Datenpunkte an. Für die Berechnung werden multiple regressionsanalytische Verfahren nach der „Methode der kleinsten Quadrate“ (Minimierung der Abweichungsquadrate) verwendet. Für ein Polynom erster Ordnung lautet das Modell

$$Z(x_i, y_i) = \beta_0 + \beta_1 x_i + \beta_2 y_i + \varepsilon(x_i, y_i) \quad (4.4)$$

mit

$Z(x_i, y_i)$	Wert am Ort (x_i, y_i)
β_0	Regressionskonstante
β_i	Regressionskoeffizienten
$\varepsilon(x_i, y_i)$	Zufallsfehler

und entspricht einer linearen Gleichung bzw. einer Ebene. Ein Polynom zweiter Ordnung entspricht einer quadratischen Gleichung

$$Z(x_i, y_i) = \beta_0 + \beta_1 x_i + \beta_2 y_i + \beta_3 x_i^2 + \beta_4 y_i^2 + \beta_5 x_i y_i + \varepsilon(x_i, y_i) \quad (4.5)$$

ein Polynom dritter Ordnung einer kubischen Gleichung

$$Z(x_i, y_i) = \beta_0 + \beta_1 x_i + \beta_2 y_i + \beta_3 x_i^2 + \beta_4 y_i^2 + \beta_5 x_i y_i + \beta_6 x_i^3 + \beta_7 x_i^2 y_i + \beta_8 x_i y_i^2 + \beta_9 y_i^3 + \varepsilon(x_i, y_i) \quad (4.6)$$

und so weiter. Im Geostatistical Analyst von ArcGIS® sind Polynome bis zur Ordnung zehn auswählbar. Geschätzt werden die Regressionskoeffizienten β_i nach der „Methode der kleinsten Quadrate“. Unterscheiden kann man zwischen einer globalen und lokalen polynominalen Interpolation. Die globale polynomiale Interpolation bezieht alle Datenpunkte in die Berechnung ein, während die lokale polynomiale Interpolation nur Daten aus entsprechenden Teilflächen des Untersuchungsgebietes verwendet und darin Polynome unterschiedlichen Grades berechnet. Im Ergebnis erhält man eine an die lokalen Verhältnisse angepasste Schätzoberfläche. Zu beachten ist, dass es zu Oszillationen zwischen den Punkten bei der Berechnung von Polynomen höherer Ordnung kommen kann, was als *Runges* Phänomen bekannt ist. Ebenso kann das Ergebnis durch unvorteilhaft liegende Datenpunkte und Ausreißer bis zur Unkenntlichkeit oszillieren, weshalb sich dann ein Wechsel des Interpolationsverfahrens, etwa die Spline Interpolation empfiehlt. Die bevorzugte Anwendung findet das Verfahren in der sog. Trendoberflächenanalyse (vgl. Kap. 3.2.3.4.3) für das Abschätzen und Entfernen von räumlichen Trends und dient daher eher zur Vorbereitung auf andere Interpolationsverfahren wie das Kriging.

4.3.2 Inverse Distance Weighted Interpolation

Die Methode der inversen Distanzgewichtung (IDW) zählt zu den flexibelsten und am häufigsten verwendeten deterministischen Interpolationstechniken. IDW ist ein lokales und exaktes Interpolationsverfahren. Es berücksichtigt jeweils eine zu definierende Umgebung des zu schätzenden Punktes. Es wird nicht versucht eine

mathematisch beschreibbare Teilfläche zu erstellen, sondern diese dynamisch aus den vorhandenen Informationen zu berechnen. Die inverse Distanzgewichtung ist die einfachste Interpolationsmethode, bei der das Gewicht jedes in die Berechnung eingehenden Punktes umgekehrt proportional zu seiner Entfernung zum geschätzten Punkt gesetzt wird. Die Berechnung erfolgt nach der Formel

$$\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i Z_i(x_i) = \frac{\sum_{i=1}^n \frac{1}{d_i^p} Z_i}{\sum_{i=1}^n \frac{1}{d_i^p}} \quad (4.7)$$

mit

$\hat{Z}(x_0)$...	zu schätzender Wert an der Stelle x_0
$Z(x_i)$...	bekannte Messwerte am Punkt x_i
λ_i ...	Gewichte ($\sum \lambda_i = 1$)
d_i^p ...	p exponentierte euklidische Distanzen der n Datenpunkte zum geschätzten Punkt

Je niedriger der Exponent gesetzt wird, desto gleichförmiger gehen alle Nachbarn in die Berechnung ein und desto glatter wird die Schätzoberfläche. Je höher der Exponent wird, desto unruhiger wird die Oberfläche, da nur mehr das Gewicht der nächstgelegenen Nachbarn in die Interpolation einfließt. Die Gewichte die jedem Messpunkt entsprechend der Gewichtungsfunktion zugeordnet werden, sind einer Summe von eins angepasst. Somit ordnet die Gewichtungsfunktion proportionale Gewichte zu und drückt den relativen Einfluss jeden Messpunktes aus.

Verfeinerte IDW-Interpolationsalgorithmen berücksichtigen einen Glättungsfaktor und Anisotropie. Aus obiger Grundformel ergibt sich die neue Gleichung (TOMCZAK, 1998):

$$\hat{Z}(x_0) = \frac{\sum_{i=1}^n \frac{Z_i}{(d_i + \delta)^p}}{\sum_{i=1}^n \frac{1}{(d_i + \delta)^p}} \quad (4.8)$$

mit δ als Glättungsparameter. Für isotrope Datensätze wird wie zuvor die euklidische Distanz verwendet. Für anisotrope Datensätze (z. B. durch Beeinflussung von Wind oder Topographie) wird die euklidische Distanz d_i durch die effektive Distanz d_{i-eff} ersetzt:

$$d_{i-eff} = \sqrt{A_{xx}(\Delta x)^2 + A_{xy}\Delta x\Delta y + A_{yy}(\Delta y)^2} \quad (4.9)$$

mit

$$\begin{aligned}
 A_{xx} &= T_{xx}^2 + T_{yy}^2 & T_{xx} &= \frac{\cos \theta}{\rho}, & T_{yy} &= \cos \theta \\
 A_{xy} &= 2(T_{xx}T_{xy} + T_{yx}T_{yy}) & T_{xy} &= \frac{-\sin \theta}{\rho}, & T_{yx} &= -\sin \theta \\
 A_{yy} &= T_{yy}^2 + T_{xy}^2
 \end{aligned}$$

wobei θ die Richtung der Anisotropie Achse angibt und ρ das Anisotropie Verhältnis ($\rho = 1$, für Isotropie). Das Konzept ist in Abbildung 4.3 illustriert.

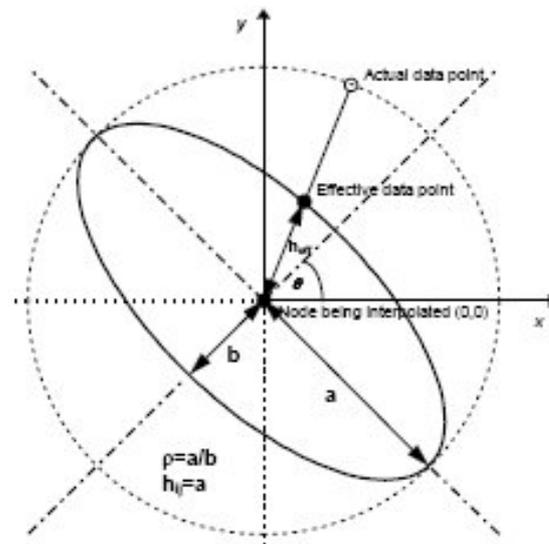


Abb. 4.3: Illustration des Konzeptes der bzgl. Anisotropie korrigierten effektiven Distanz (TOMCZAK, 1998)

Ein Nachteil der IDW ist, dass die Schätzfunktion gezwungen ist, lokale Maxima oder Minima in den Datenpunkten zu bilden. Dadurch werden kreisförmige Bereiche (sog. bull eyes) gleicher Werte um die bekannten Datenpunkte gebildet (vgl. Abb. 4.4).

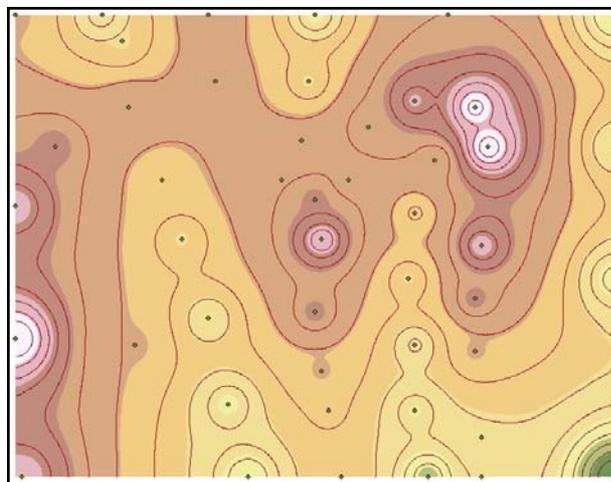


Abb. 4.4: IDW „bull-eyes“ Effekt

Die Shepard's Methode ist eine Variation der inversen exponierten Distanzgewichtung mit zwei verschiedenen Gewichtsfunktionen für zwei separate Umgebungen. In der inneren Umgebung wird ein Exponent von zwei für die Gewichtung verwendet, für die äußere Umgebung ein Gewicht von vier. Die innere Umgebung wird mit einem Drittel des Radius der äußeren Umgebung festgelegt.

Unterschiede gibt es auch in den Suchalgorithmen, die die Messpunkte selektieren, die in der Umgebung des geschätzten Punktes liegen. Der einfachste Algorithmus ist die nearest-neighbour Suche. Dabei wird eine vorgegebene Anzahl von Messpunkten, welche am nächsten zum Schätzpunkt liegen, identifiziert (Abb. 4.5a). Ein Nachteil der Methode ist, dass die Suche Punkte ermitteln kann, die alle nur auf einer Seite des Schätzpunktes liegen. Mit der Quadranten-Suche ist eine gleichmäßigere Verteilung der Messpunkte gewährleistet. Eine Mindestanzahl von Punkten muss in jedem der vier Quadranten um den Schätzpunkt verwendet werden (Abb. 4.5b). Eine Erweiterung der Quadranten-Suche ist die Oktanten Suche (Abb. 4.5c).

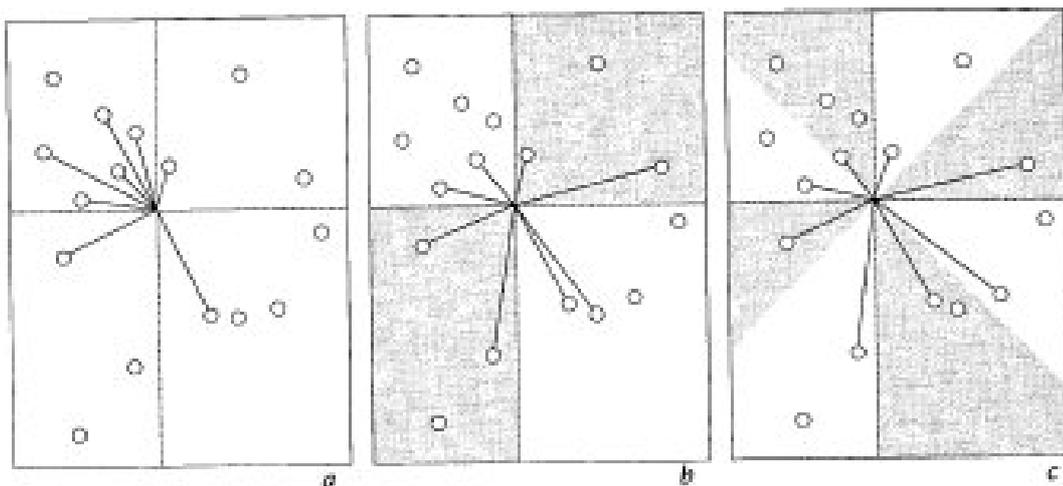


Abb. 4.5: (a) Suchtechnik die n nächste Nachbarn um den Stützpunkt identifizieren
 (b) Quadrantensuche die eine gleichmäßige Verteilung der Nachbarn um den Schätzpunkt gewährleistet
 (c) Oktantensuche für die Nachbarn um den Schätzpunkt (DAVIS, 2002)

Festzuhalten ist, dass tendenziell mit größerer Zahl von Datenpunkten die Glättung der resultierenden Oberfläche zunimmt.

4.3.3 Spline Interpolation

Spline ist ein Begriff aus der numerischen Mathematik der erstmals von SCHOENBERG (1946) für glatte, harmonische zusammengesetzte mathematische Kurven dritten Grades verwendet wurde. Der Begriff stammt aus dem Schiffsbau und bezeichnet die sog. Straklatte (engl. *spline*) die mit Gewichten, den sog. Molchen,

belegt wird, um harmonische Kurven ohne plötzliche Änderung des Krümmungsradius zu entwerfen. Die Form der Straklatte ist von ihrer Elastizität abhängig und wird von den Gewichten bestimmt.

Ein Spline n -ten Grades ist eine mathematische Funktion, die stückweise aus Polynomen mit maximal Grad n zusammengesetzt ist. Dabei sind an den Übergangsstellen der Teilintervalle bestimmte Bedingungen gestellt, etwa dass der Spline $(n-1)$ -mal stetig differenzierbar ist. Eine Spline-Funktion dritten Grades hat die Eigenschaft, dass die Gesamtkrümmung der interpolierenden Kurve minimal wird, eine Eigenschaft die eine mit Gewichten belegte Straklatte ebenfalls realisiert (BRONSTEIN & SEMENDJAJEW, 1987). Nach der Elastizitätstheorie nehmen Materialien immer die energieärmste Materialstruktur ein, d. h. die Biege- oder Spannungsenergie ist ein Minimum.

Durch die stückweise Definition sind Splines flexibler als Polynome und dennoch relativ einfach und durch die vorgeschriebene Stetigkeit glatt, sodass der Kurvenverlauf beim Übergang der Teilintervalle beibehalten wird. Dadurch ergeben sich bei der Spline Interpolation nicht die Nachteile, die durch starke Oszillation von Polynomen höherer Ordnung bei der Polynom Interpolation entstehen. Im GIS-Anwendungsbereich werden radiale Basisfunktionen (RBF) als mathematische Funktion verwendet, sodass die geschätzte Oberfläche exakt durch die Datenpunkte verläuft und die Oberflächenkrümmung minimiert wird. Im Unterschied zum ebenfalls lokalen und exakten IDW - Verfahren ist eine Überschreitung der lokalen Maxima bzw. eine Unterschreitung der lokalen Minima zulässig. Dies führt zu einem geglätteten Verlauf zwischen den Datenpunkten (vgl. Abb. 4.6).



Abb. 4. 6: Grafischer Vergleich zwischen IDW (links) und Spline (rechts) Interpolation
(JOHNSTON et al., 2002)

Eine radiale Basisfunktion ist eine reelle Funktion deren Wert nur vom Abstand zu einem Punkt c abhängt, der Zentrum genannt wird, sodass

$$\phi(x, c) = \phi(\|x - c\|), \quad (4.10)$$

wobei gewöhnlicherweise die euklidische Norm gewählt wird. Für die Oberflächeninterpolation, wo an die zwei-dimensionale Lageinformation (x,y) der Datenpunkte ein nicht räumliches Attribut geknüpft wird, werden kubische Splines herangezogen. Für

die Spline Interpolation wird eine Menge von N Basisfunktionen verwendet, nämlich eine für jeden Datenpunkt unter der Bedingung, dass die mittlere quadratische Krümmung (Glättung) ein Minimum wird. Nach MITASOVA & MITAS (1993) oder LLOYD (2010) kann die Lösung dieses Problems als die Summe von zwei Komponenten ausgedrückt werden:

$$\hat{Z}(\bar{x}_0) = T(\bar{x}_0) + \sum_{j=1}^N \lambda_j \phi(\|\bar{x}_j - \bar{x}_0\|) \quad (4.11)$$

mit

N Anzahl der Datenpunkte

λ_j zu schätzende Gewichte

$\phi(r)$... radiale Basisfunktion mit $r = \|\bar{x}_j - \bar{x}_0\|$ als euklidische Distanz zwischen dem geschätzten Ort \bar{x}_0 und jeden Datenpunkt \bar{x}_j

$T(\bar{x}_0)$ wird als Trendfunktion bezeichnet und ist eine lokale Untermenge M (die M nächsten Nachbarn) linear unabhängiger Funktionen

$$T(\bar{x}_0) = \sum_{l=1}^M a_l f_l(\bar{x}_0) \quad (4.12)$$

Mit σ als Bezeichnung für den Tensionsparameter werden als RBF folgende Funktionen verwendet:

Multiquadratische Funktion:

$$\phi(r) = \sqrt{r^2 + \sigma^2} \quad (4.13)$$

Inverse multiquadratische Funktion:

$$\phi(r) = 1/\sqrt{r^2 + \sigma^2} \quad (4.14)$$

Thin Plate Spline Funktion (TPS):

$$\phi(r) = (\sigma \cdot r)^2 \ln(\sigma \cdot r) \quad (4.15)$$

Spline mit Tension Funktion:

$$\phi(r) = \ln\left(\sigma \cdot \frac{r}{2}\right) + K_0(\sigma \cdot r) + C_E \quad (4.16)$$

mit

K_0 ... modifizierte Besselfunktion (BRONSTEIN & SEMENDJAJEW, 1987, S. 633).

C_E ... Eulersche Konstante (BRONSTEIN & SEMENDJAJEW, 1987, S. 332)

($C_E = 0,577215665\dots$)

Completely Regularized Spline Funktion (CRS):

$$\phi(r) = \ln\left(\sigma \cdot \frac{r}{2}\right)^2 + E_1\left(\sigma \cdot \frac{r}{2}\right)^2 + C_E \quad (4.17)$$

mit

E_1 ... Integralexponentialfunktion (BRONSTEIN & SEMENDJAJEW, 1987, S. 61).

C_E ... Eulersche Konstante ($C_E = 0,577215665\dots$)

Von den fünf vorgestellten Spline - Interpolationsfunktionen hat sich die mit „completely regularized spline with tension and smoothing“ erzeugte Oberfläche im Allgemeinen als genaueste und flexibelste erwiesen (MITASOVA & MITAS, 1993 bzw. MITASOVA et al., 1995). Bei raschem Anstieg oder abrupten Änderungen der Datenwerte in lokaler Nachbarschaft kann es bei den vier erst genannten Spline – Interpolationsfunktionen zu übertriebenen Überschreitungen der lokalen Maximal- bzw. Minimalwerte kommen. Mit dieser Technik kann das Ausgreifen der Oberfläche über den Wertebereich der Datenpunkte reguliert werden. Ein weiterer Vorteil ist, dass auch die Ableitungen der Ordnung größer als zwei existieren und simultan mit der Interpolation berechnet werden können, um sie direkt für die Analyse der Oberflächengeometrie zu verwenden. Mithilfe des Tensionsparameters wird das lokale Verhalten der Oberfläche gesteuert. Je größer der Tensionsparameter, desto lokaler erfolgt die Formung der Oberfläche. In ArcGIS® wird der optimale Tensionsparameter durch Minimierung des RMSPE mithilfe der Kreuzvalidierung (siehe Kap. 4.5) automatisch ermittelt. Eine ausführlichere Beschreibung zu RBF und deren Beziehung zu Splines und Kriging (Kap. 4.4) kann in CHILÈS & DELFINER (1999) gefunden werden.

4.4 Stochastische Interpolationsverfahren

Im Unterschied zu und in Erweiterung der deterministischen Verfahren des vorigen Kapitels handelt es sich bei stochastischen Interpolationsverfahren um eine Kombination mathematischer und statistischer Verfahren, die auf Basis der Theorie der regionalisierten Variablen entwickelt wurden. Der Interpolationsprozess, d. h. die Schätzung der unbekanntenen Werte, beruht auf Wahrscheinlichkeiten. Die aus der Variographie gewonnene Information – räumliche Kontinuität (Variogrammmodell) und ihre Reichweite, großräumige (Schwellwert) und kleinräumige Variabilität (Nuggeteffekt) – wird im Zuge eines geostatistischen Schätzprozesses weiterverwendet. Unter dem Oberbegriff Kriging, benannt nach D.G. KRIGE, einem südafrikanischen Pionier in der Anwendung statistischer Techniken für Berechnungen im Minenabbau, werden eine Reihe von geostatistischen Interpolationsverfahren bezeichnet, die sich anhand der verschiedenen stochastischen Ansätze

unterteilen lassen. Entsprechend den Voraussetzungen können diese als stationär - nicht stationär, linear – nicht linear, uni- oder multivariat bezeichnet werden. Gemeinsam ist allen Verfahren, dass sie auf Regressionstechniken beruhen, die eine Minimierung der Schätzvarianz zum Ziel haben. Der Unterschied zu den deterministischen Interpolationsmethoden besteht darin, dass die Gewichtungsfaktoren nicht allein aufgrund der Lage der Messpunkte im Raum bestimmt werden, sondern ein, aus den Eigenschaften der Variablen in ihrer Anordnung im Raum abgeleiteter, stochastischer Anteil hinzugenommen wird. Die Basis für das eigentliche Krigingverfahren bildet der Betrag der räumlichen Varianzänderung zwischen den jeweiligen einzelnen Werten. Erhalten wird dieses Maß über den räumlichen Zusammenhang der Messpunkte durch das (Semi-)Variogramm (*Kap. 3.2.3*), welches zur Bestimmung der optimalen Interpolationsgewichte verwendet wird. Somit entscheidet eine durch die Daten gesicherte und aus ihnen abgeleitete Funktion über das Ergebnis der Interpolation und nicht eine willkürlich gewählte deterministische Funktion. Analog der IDW (Gl. 4.7) wird der Schätzwert \hat{Z}_0 durch Bildung eines gewichteten Mittelwertes aus den vorhandenen Messwerten Z_i berechnet:

$$\hat{Z}(x_0) = \sum_{i=1}^n \lambda_i^K Z(x_i) \quad (4.18)$$

mit

$\hat{Z}(x_0) \dots$	zu schätzender Wert an der Stelle x_0
$Z(x_i) \dots$	bekannte Messwerte am Punkt x_i
$\lambda_i^K \dots$	Kriginggewichte

Kriterien für die Bestimmung der Gewichte λ_i^K sind die Erwartungstreue der Schätzung sowie die Minimierung des mittleren quadratischen Schätzfehlers:

1. Erwartungstreue: $E[\hat{Z}(x_0) - Z(x_0)] = 0$.

Der Erwartungswert des Schätzfehlers, also die Differenz zwischen dem wahren Wert und dem geschätzten Wert an einem beliebigen Ort ist null. Damit sollen systematische Über- oder Unterschätzungen ausgeschlossen werden. Dies gilt wenn der Mittelwert m konstant über das Untersuchungsgebiet ist, d. h. es liegt Stationarität der 2. Ordnung vor (*Kap. 3.2.2.2*) oder zumindest die abgeschwächte Stationarität, die sog. intrinsische Hypothese (*Kap. 3.2.2.3*) muss erfüllt sein, d. h. das Mittel und die Varianz der Inkremente $Z(x+h)-Z(x)$ existieren und sind unabhängig vom Punkt x .

2. Varianzminimum: $Var(\hat{Z}(x_0) - Z(x_0)) = E[\hat{Z}(x_0) - Z(x_0)]^2 \rightarrow \min.$

Die Varianz zwischen Schätzwert und tatsächlichen Wert (Schätzfehler) soll ein Minimum ergeben.

Folgende Eigenschaften kennzeichnen das Krigingverfahren:

- *Kriging* liefert den besten, unverzerrten Schätzwert, d. h. es gibt keine systematischen Verzerrungen (BLUE = **B**est **L**inear **U**nbiased **E**stimator).
- *Kriging* ist ein exakter Interpolator, d. h. die Messwerte an den Datenpunkten werden immer exakt wiedergegeben.
- *Kriging* berücksichtigt die Lage von Datenpunkten und bezieht die räumliche Struktur einer Messgröße über das Semivariogramm mit ein. Damit ist *Kriging* den klassischen distanzgewichteten Verfahren überlegen, da es das Potenzial besitzt, Informationen über die Kovarianz von stark beprobten Gebieten in spärlich beprobte zu übertragen (CHILÈS & DELFINGER, 1999). Es ist daher für ein unregelmäßiges Verteilungsmuster der Datenpunkte besser geeignet.
- *Kriging* ermöglicht die Beurteilung der Zuverlässigkeit der Schätzung für jeden Schätzwert durch Angabe des Krigingfehlers. Der Krigingfehler (Schätzfehler) ist dabei nur abhängig von der Lage der Datenpunkte und vom Semivariogrammmodell, nicht jedoch von den Messwerten.

Wie bereits am Anfang des Kapitels beschrieben, wird der Schätzwert \hat{Z}_0 durch Bildung einer gewichteten Linearkombination der Messwerte gebildet. Der Erwartungswert des Schätzfehlers ist somit:

$$E(\hat{Z}_0 - Z_0) = E(\sum \lambda_i^K Z(x_i) - Z_0) = \sum \lambda_i^K m - m = m(\sum \lambda_i^K - 1) \quad (4.19)$$

Damit der erwartete Fehler null ist, muss entweder $m=0$ sein, d. h. der Mittelwert ist bekannt oder die Kriginggewichte in Summe eins ($\sum \lambda_i^K = 1$) ergeben. Kriging umfasst verschiedene Methoden der Bestimmung gewichteter Mittelwerte, unterschiedliche Verfahren und die einzelnen Schritte sind etwa bei CHILÈS & DELFINGER (1999), CRESSIE (1993), ISAACS & SRIVASTAVA (1989) oder WACKERNAGEL (2003) beschrieben.

4.4.1 Simple Kriging

Simple Kriging (SK) ist die mathematisch am wenigsten komplizierte Form des Kriging. Diese räumliche Regressionstechnik setzt einen bekannten Mittelwert, welcher als Durchschnittswert aus den Daten berechnet werden kann, über das

Untersuchungsgebiet voraus. Die Definition der Simple Kriging Schätzung für einen Wert am Ort x_0 lautet:

$$\hat{Z}(x_0) = m + \sum_{i=1}^n \lambda_i^{SK} [Z(x_i) - m] \quad (4.20)$$

mit den zu den Residuen $Z(x_i) - m$ dazugehörigen Simple Kriginggewichten λ_i^{SK} . Da der Mittelwert gleich null ist und damit die Schätzung automatisch erwartungstreu, benötigt es keine Bedingung bzgl. der Summe der Kriginggewichte (vgl. Gl.4.19). Die Varianz des Schätzfehlers kann damit unter Annahme der Stationarität zweiter Ordnung einfach ausgedrückt werden als

$$\begin{aligned} s_E^2 &= Var(\hat{Z}(x_0) - Z(x_0)) = E[\hat{Z}(x_0) - Z(x_0)]^2 = & (4.21) \\ &= E\left[\left(\hat{Z}(x_0)\right)^2 + \left(Z(x_0)\right)^2 - 2\hat{Z}(x_0)Z(x_0)\right] = \\ &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i^{SK} \lambda_j^{SK} C(x_i - x_j) + C(x_0 - x_0) - 2 \sum_{i=1}^n \lambda_i^{SK} C(x_i - x_0) = \\ &= C(x_0 - x_0) + \sum_{i=1}^n \sum_{j=1}^n \lambda_i^{SK} \lambda_j^{SK} C(x_i - x_j) - 2 \sum_{i=1}^n \lambda_i^{SK} C(x_i - x_0) \end{aligned}$$

(a) (b) (c)

mit

$$C(x_i - x_j) = C[Z(x_i), Z(x_j)]$$

Die Gl. (4.21) besteht aus drei Termen: (a) der Varianz der Daten $C(0) = \sigma^2$, (b) der Redundanz der Information infolge der Datenkonfiguration und (c) der Distanz zwischen den Daten und dem geschätzten Ort. Die Varianz des Schätzfehlers wächst mit der Variabilität der Daten und den redundanten Informationen, reduziert sich aber durch Verwendung von Daten die zum geschätzten Ort eine Korrelation aufweisen.

Die Bedingung des Varianzminimums für den Schätzfehler führt zu einem analogen Problem wie die Ausgleichung nach der Methode der kleinsten Quadrate. Die Minimierung wird durch Bildung der partiellen Ableitungen nach den einzelnen λ_i^{SK} und deren Gleichsetzung zu null erreicht. Explizit ergibt sich für alle $i = 1, \dots, n$

$$2 \sum_{j=1}^n \lambda_j^{SK} C(x_i - x_j) - 2C(x_i - x_0) = 0 \quad (4.22)$$

und man erhält damit das Gleichungssystem für Simple Kriging wie folgt:

$$\sum_{j=1}^n \lambda_j^{SK} C(x_i - x_j) = C(x_i - x_0) \quad (4.23)$$

Die zugehörige Varianz für SK ergibt sich nach Gl. (4.21) durch Ersetzen der linken Seite der Simple Kriging Gleichung (4.23) durch die rechte Seite:

$$s_{SK}^2 = C(0) - \sum_{i=1}^n \lambda_i^{SK} C(x_i - x_0) \quad (4.24)$$

Anstatt der Kovarianz kann auch das (Semi-)Variogramm (Gl. 3.38), unter Annahme der zumindest erfüllten intrinsischen Hypothese, verwendet werden. Die Lösung des SK Gleichungssystems ergibt die Simple Kriginggewichte und damit die Schätzung von $\hat{Z}(x_0)$ nach Gl. (4.20):

$$\begin{aligned} \hat{Z}(x_0) &= m + \sum_{i=1}^n \lambda_i^{SK} [Z(x_i) - m] = \\ &= \sum_{i=1}^n \lambda_i^{SK} Z(x_i) + m \cdot \left(1 - \sum_{i=1}^n \lambda_i^{SK}\right) = \\ &= \sum_{i=1}^n \lambda_i^{SK} Z(x_i) + m \cdot \lambda_M^{SK} \end{aligned} \quad (4.25)$$

Der Term λ_M^{SK} wird als Gewicht des Mittelwertes bei SK bezeichnet.

Heutzutage wird SK selten angewandt, da der Mittelwert meist nicht bekannt ist. Es gibt aber Umstände wo der Mittelwert im Vorhinein bekannt ist. Dies ist der Fall bei standardisierten Variablen, wo dieser null beträgt. Ebenfalls null ist der Mittelwert bei Residualvariablen (z. B. Residuen einer Trendoberfläche oder eines Regressionsmodells). Deshalb wird SK manchmal auch als Kriging mit bekannten Mittelwert bezeichnet. Ein wichtiger Grund aber SK anzuwenden ist, dass das Gewicht des Mittelwertes λ_M^{SK} eines der besten Kriterien zum Test der Kriging-Qualität bildet (siehe Kap. 4.5.2). Wenn die Stichproben unregelmäßig im Raum verteilt sind, ist es sinnvoll eine Karte der SK-Varianz als Ergänzung zu den Schätzwerten zu erstellen. Man erhält dadurch eine Abschätzung der variierenden Genauigkeit der Schätzwerte infolge der unregelmäßigen Anordnung der Datenpunkte (WACKERNAGEL, 2003).

4.4.2 Ordinary Kriging

Ordinary Kriging ist das meist verbreitete Kriging Verfahren. Mit einer Ausnahme sind die Voraussetzungen für das Ordinary Kriging (OK) dieselben wie für das Simple Kriging. Der Mittelwert der regionalisierten Variable ist hypothetisch konstant über das Untersuchungsgebiet, jedoch unbekannt. Die Aufhebung dieser Einschränkung vergrößert die Anwendungsmöglichkeiten des Kriging beträchtlich. Die Schätzung aus dem SK kann auch in folgenderweise umgeschrieben werden:

$$\hat{Z}(x_0) = m (1 - \sum_{i=1}^n \lambda_i) + \sum_{i=1}^n \lambda_i Z(x_i) \quad (4.26)$$

Wenn man die Summe der Gewichte gleich eins setzt, dann wird die Größe innerhalb der Klammern gleich null und der erste Term, der den Mittelwert enthält, verschwindet. Die Kriging Schätzung ist dann unabhängig vom Wert des Mittels. Unter Verwendung des Lagrange Multiplikators kann für das Problem der Varianzminimierung die Bedingung $\sum \lambda_i^{OK} = 1$ hinzugenommen werden. Die einzelnen Schritte sind etwa bei ISAAKS & SRIVASTAVA (1989) beschrieben und führen unter Verwendung des (Semi-)Variogramms zur Bildung des bekannten erweiterten Gleichungssystems für OK:

$$\begin{pmatrix} \gamma(x_1 - x_1) & \dots & \gamma(x_1 - x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n - x_1) & \dots & \gamma(x_n - x_n) & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \lambda_1^{OK} \\ \vdots \\ \lambda_n^{OK} \\ \mu_{OK} \end{pmatrix} = \begin{pmatrix} \gamma(x_1 - x_0) \\ \vdots \\ \gamma(x_n - x_0) \\ 1 \end{pmatrix} \quad (4.27)$$

mit den zu den Datenwerten zugehörigen OK-Gewichten λ_i^{OK} und dem Lagrange Parameter μ_{OK} . Die linke Seite von Gl. (4.27) beschreibt die Unähnlichkeiten zwischen den Datenpunkten, während die rechte Seite die Unähnlichkeiten zwischen jedem Datenpunkt und der Schätzung am Ort (x_0) zeigt. Die Varianz des Schätzfehlers von OK ergibt sich zu

$$s_{OK}^2 = \mu_{OK} - \gamma(x_0 - x_0) + \sum_{i=1}^n \lambda_i^{OK} \gamma(x_i - x_0). \quad (4.28)$$

Die lokale Schätzvarianz $\sigma_{OK}^2(x_0)$ kann ebenfalls bestimmt werden:

$$s_{OK}^2(x_0) = - \sum_{j=1}^n \sum_{i=1}^n \lambda_j \lambda_i \gamma(x_i - x_j) + 2 \sum_{i=1}^n \lambda_i \gamma(x_i - x_0) \quad (4.29)$$

Interessant sind noch die zwei folgenden Beziehungen:

$$\mu_{OK} = \lambda_M^{SK} \mu_{KM} \quad (4.30)$$

$$s_{OK}^2 = s_{SK}^2 + (\lambda_M)^2 s_{KM}^2 = s_{SK}^2 + (\lambda_M)^2 \mu_{KM} \quad (4.31)$$

Gl. (4.30) liefert einen Ausdruck des Lagrange Parameters von OK in Termen des Gewichtes des Mittelwertes in SK und des Lagrange Parameters für Kriging des Mittelwertes $\mu_{KM} = s_{KM}^2$. Gl. (4.31) zeigt, dass die OK Varianz in zwei Teile getrennt werden kann. Sie ergibt sich als Summe der SK-Varianz und der Varianz der Schätzung des Mittelwertes multipliziert mit dem Quadrat des Gewichtes des Mittelwertes in SK. Der zweite Teil gibt ein Maß des Genauigkeitsverlustes infolge des unbekanntes Mittelwertes an.

Im Fall der abgeschwächten Stationarität (intrinsic Hypothese) wird mit Inkrementen anstatt der Variablen selbst gearbeitet. Unter der zugrunde liegenden Annahme der Gl. (3.30) und (3.31) ist der Schätzfehler ($\hat{Z}(x_i) - Z(x_i)$) ein Inkrement, vorhergesehen dass die Summe der Gewichte gleich eins ist, und die Erwartung und die Varianz existieren und können berechnet werden. Dies führt unter Verwendung des Variogramms zu denselben Kriging-Gleichungen wie zuvor, weshalb im Allgemeinen die intrinsic Hypothese anstatt der Stationarität verwendet wird.

4.4.3 Universal Kriging

Die bisherigen Verfahren zur Ermittlung der Schätzwerte gehen davon aus, dass die Messwerte keinem räumlichen Trend unterliegen. In vielen Fällen zeigen Messwerte jedoch eine systematische Zu- oder Abnahme im Raum, die als Trend (global) oder Drift (lokal) bezeichnet wird. In diesen Fall ist das OK nicht optimal, da die Schätzungen im Vergleich zum wahren Wert zu hoch oder niedrig ausfallen. Damit wird auch der Schätzfehler übermäßig erhöht und ein Varianzminimum ist nicht mehr möglich.

Universal Kriging (UK) ist eine weitere Generalisierung des Krigingverfahren, bei welchem ein konstanter Mittelwert über den Untersuchungsraum nicht mehr gegeben sein muss. Das Verfahren leistet in einer simultanen Operation das, was ansonsten in drei Schritten durchgeführt werden muss:

- Eine Drift wird geschätzt und beseitigt um Stationarität bzgl. den Residuen in den Messpunkten zu erhalten.
- Ein Kriging wird mit den stationären Residuen durchgeführt, um geschätzte Residualwerte an den nicht beprobten Orten zu erhalten.
- Die geschätzten Residualwerte werden mit der Drift kombiniert, um die aktuelle Oberfläche zu erhalten.

Anstatt diese komplizierten Schritte durchzuführen, können die Kriging Gleichungen um weitere Lagrange Parameter erweitert werden, die eine Drift repräsentieren. Diese kann als Linearkombination der Koordinaten am Messpunkt innerhalb einer lokalen Umgebung modelliert werden. In der Geostatistik werden üblicherweise Polynome erster oder zweiter Ordnung verwendet:

$$m(x_0) = \beta_0 + \sum_{i=1}^n (\beta_1 x_i + \beta_2 y_i) \quad (4.32)$$

$$m(x_0) = \beta_0 + \sum_{i=1}^n (\beta_1 x_i + \beta_2 y_i + \beta_3 x_i^2 + \beta_4 y_i^2 + \beta_5 x_i y_i) \quad (4.33)$$

mit β_i als unbekannte Drift Koeffizienten.

Im Unterschied zum OK muss das (Semi-)Variogramm von den Residuen der regionalisierten Variable stammen. Das bedeutet das zuvor eine Strukturanalyse durchgeführt werden muss, um die beste Kombination von Nachbarschaftsgröße und Repräsentation der Drift zu erhalten. Dies ist kein triviales Unterfangen, weil das Drift-Modell und die Nachbarschaftsgröße in gegenseitiger Wechselwirkung stehen. Durch die Hinzunahme weiterer Lagrange Parameter für die Drift müssen mehr unbekannte Koeffizienten berechnet werden. Dazu wird eine größere Anzahl von Messpunkten benötigt, um genügend Überbestimmung für die Lösung des Gleichungssystems zu haben. Die Varianz des Schätzfehlers beinhaltet die Unsicherheit der geschätzten Drift und den Schätzfehler der Residuen, der wiederum abhängig von der Wahl des Driftmodells ist.

4.4.4 Lognormal Kriging

Mit Lognormal Kriging werden Schätzungen von Zufallsvariablen $Z(x_i)$ vorgenommen, die log-normalverteilt vorliegen, insbesondere bei positiver Schiefe (Gl. 3.9). Die Datenwerte werden zuerst durch Logarithmieren mit dem natürlichen Logarithmus (\ln) oder dem Logarithmus zur Basis 10 (\log) in die normalverteilte Zufallsvariable

$$Y(x_i) = \log Z(x_i) \quad (4.34)$$

transformiert (vgl. 3.17), mit welchen das Variogrammmodell und die Interpolation durch SK, OK oder UK vorgenommen wird:

$$\hat{Y}(x_0) = \sum_{i=1}^n \lambda_i^K \log Z(x_i) = \sum_{i=1}^n \lambda_i^K Y(x_i) \quad (4.35)$$

Die geschätzten Werte können nach der Interpolation auf die ursprüngliche Verteilung erwartungstreu durch

$$\hat{Z}(x_0) = \exp \left[\hat{Y}(x_0) + \frac{1}{2} \sigma_{Y,SK}^2(x_0) \right] \quad (4.36)$$

rücktransformiert werden (BURROUGH & McDONELL, 1998), wobei $\sigma_{Y,SK}^2$ die Lognormal Kriging Varianz aus dem Simple Kriging ist. Mit $\sigma_{Y,OK}^2$ und dem Lagrange-Parameter Kriging λ_M aus dem Ordinary Kriging der logarithmierten Werte ergibt sich für die Rücktransformation (CRESSIE, 1993):

$$\hat{Z}(x_0) = \exp \left[\hat{Y}(x_0) + \frac{1}{2} \sigma_{Y,OK}^2(x_i) - \lambda_M \right] \quad (4.37)$$

CHILÈS & DELFINGER (1999) weisen weiters darauf hin, dass das Verfahren sehr sensitiv gegenüber Variogrammfluktuation und relativ schwierig in der Anwendung ist. Notwenig ist eine feine Abstimmung des Schwellwertes (sill) des Variogramms. Bei der Transformation der Werte ist darauf zu achten, dass die logarithmierten Werte tatsächlich eine Normalverteilung aufweisen bzw. die Abweichung von dieser vernachlässigbar ist.

4.4.5 Trans-Gaussian Kriging

Der Ansatz des Lognormal Kriging kann für eine Zufallsvariable mit einer beliebigen Transformation $Y = \varphi(Z)$ verallgemeinert werden, wobei $\varphi(Z)$ einen normalverteilten stochastischen Prozess darstellt, der intrinsisch stationär ist. Die Funktion φ ist zweifach differenzierbar und invertierbar. Die transformierten Y -Werte lassen sich mit einem linearen Schätzer interpolieren und die interpolierten Werte werden rücktransformiert. Mithilfe der Taylor-Reihenentwicklung und unter der Annahme, dass der Kriging Fehler $(Y - \hat{Y})$ - vgl. Gl (4.39) - relativ klein sind, kann eine erwartungstreue Rücktransformation der Zufallsvariable Z erreicht werden:

$$\begin{aligned} \varphi(Y) &= \varphi(\hat{Y}) + (Y - \hat{Y})\varphi'(\hat{Y}) + \frac{1}{2}(Y - \hat{Y})^2\varphi''(\hat{Y}) \\ E[\varphi(Y)] &\cong \varphi(\hat{Y}) + \frac{1}{2} \sigma_{SK}^2\varphi''(\hat{Y}), \text{ da } E[(Y - \hat{Y})] = 0 \\ \hat{Z} &= \varphi(\hat{Y}) + \frac{1}{2} \sigma_{SK}^2\varphi''(\hat{Y}) \end{aligned} \quad (4.38)$$

für Simple Kriging (CHILÈS & DELFINGER 1999) bzw.

$$\hat{Z} = \varphi(\hat{Y}) + \varphi''(\mu_{\hat{Y}}) \left\{ \frac{1}{2} \sigma_{OK}^2 - \lambda_M \right\} \quad (4.39)$$

für Ordinary Kriging (CRESSIE, 1993).

Im Lognormalen Fall erhält man aus Gl. (4.38) die Beziehung nach Gl. (4.36)

$$\hat{Z} = \exp \left[\hat{Y} + \frac{1}{2} \sigma_{Y,SK}^2 \right].$$

Im Fall einer Box-Cox Transformation (Gl. 3.17) kann das Trans-Gaussian Kriging in ArcGIS® mit OK und UK verwendet werden.

4.4.6 Co-Kriging

Das Co-Kriging ist eine multidimensionale Erweiterung des Krigingverfahrens, d. h. es lassen sich zwei oder mehrere Variablen geostatistisch analysieren. Zusätzliche Variablen (sekundäre Variablen), die mit der primären Variable eine räumliche Kreuzkorrelation aufweisen, können in das Modell integriert werden. Bei einer Kreuzkorrelation der beiden Variablen werden die in der sekundären Variablen gespeicherten Informationen dazu verwendet, die primäre Variable in ihrer räumlichen Verteilung zu beschreiben. Damit kann die Interpolation in der primären Variable verbessert und die Varianz des Schätzfehlers verringert werden. Voraussetzung für die sekundären Daten ist eine höhere Erfassungsdichte und eine Korrelation der Daten mit der primären Schätzvariablen. Das Co-Kriging modelliert die gemeinsame Co-Varianz beider Variablen in einem Co-Semivariogramm, in dem es Informationen aus den Semivariogrammen beider Daten verwendet.

4.5 Validierung von Schätzergebnissen

4.5.1 Kreuzvalidierung

Unverzichtbar für das sinnvolle Konstruieren von Oberflächen ist eine nachvollziehbare Prüfung der Qualität der Interpolationen bzw. der Methoden, mit denen diese erstellt wurden (ISAACS & SRIVASTAVA, 1989). Eine der am häufigsten verwendeten Methoden zur Prüfung der Qualität einer Interpolation stellt die sog. Kreuzvalidierung (engl. *cross validation*) dar. Die Qualität der interpolierten Punkte hängt zum einen von der Anzahl und dem Verteilungsmuster der Datenpunkte ab, zum anderen ist das mathematische Modell (PI, IDW, SI, SK, OK, UK), das das zu untersuchende Phänomen beschreibt, entscheidend. Mithilfe der Kreuzvalidierung lassen sich verschiedenen Annahmen über das Modell (z. B. Parameterwahl, Variogrammmodell, Nachbarschaftsgröße) oder den Daten (z. B. Ausreißer, punktweise Anomalien) vergleichen.

Im Kreuzvalidierungsverfahren wird jeder Messwert $Z(x_i)$ der Reihe nach aus den n Beobachtungen entfernt und mit den übrigen $(n-1)$ Beobachtungen an der entfernten Stelle eine Schätzung $Z^*(x_{[i]})$ mit den konstant gehaltenen Interpolationsparametern erneut vorgenommen. Die eckigen Klammern beim Index i zeigen an, dass die Schätzung an der Stelle x_i ohne den Messwert Z_i durchgeführt wird. Die Differenz zwischen dem Messwert und dem geschätzten Wert

$$SF(x_i) = Z(x_i) - Z^*(x_{[i]}) \quad (4.39)$$

wird als lokaler Schätzfehler am Ort x_i bzw. Kreuzvalidierungsfehler bezeichnet. Die Schätzfehler SF werden auch als Residuen bezeichnet. Die Ergebnisse der Kreuzvalidierung können mithilfe der Berechnung verschiedener statistischer Maßzahlen analysiert werden, welche Aufschluss über die Güte eines Interpolationsverfahrens geben. Ist der mittlere lokale Schätzfehler SF einer Stichprobe (n) und damit das arithmetische Mittel (engl. *mean prediction error - MPE*) annähernd null

$$MPE = \frac{1}{n} \sum_{i=1}^n (Z(x_i) - Z^*(x_{[i]})) \cong 0, \quad (4.40)$$

dann liegt offensichtlich keine Verzerrung und damit Erwartungstreue vor. Dagegen kann ein signifikanter negativer (oder positiver) mittlerer Schätzfehler eine systematische Überschätzung (bzw. Unterschätzung) repräsentieren (WACKERNAGEL, 2003).

Ein weiteres aussagekräftiges Maß ist die Schiefe (Gl. 3.9), da eine schiefe Verteilung ebenfalls eine Verzerrung der geschätzten Oberfläche impliziert. Weiterhin erhofft man sich eine geringe Streuung der lokalen Schätzfehler, die durch die Standardabweichung s (Gl. 3.7) angegeben wird. Eine Schätzung mit einer geringen Abweichung von der Erwartungstreue und einer kleinen Standardabweichung der Residuenwerte ist einer erwartungstreuen Schätzung mit einer großen Standardabweichung vorzuziehen (ISAAKS & SRIVASTAVA, 1989).

Der mittlere quadrierte Schätzfehler (engl. *mean squared prediction error - MSPE*) ist eine weitere hilfreiche statistische Maßzahl für den Vergleich verschiedener Schätzungen:

$$MSPE = \frac{1}{n} \sum_{i=1}^n (Z(x_i) - Z^*(x_{[i]}))^2 \quad (4.41)$$

In Analogie zur Standardabweichung kann zudem der *root mean squared prediction error (RMSPE)* aus den quadrierten Schätzfehlern nach Gl. (4.37) berechnet werden, dessen Optimum ebenfalls null ist.

$$RMSPE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Z(x_i) - Z^*(x_{[i]}))^2} \quad (4.42)$$

Wegen des Proportionalitätseffektes bei hohen Eingangswerten gehen besonders hohe Differenzen in entscheidendem Maße in die Berechnung des *MSPE* bzw. *RMSPE* ein. Daher sind bei Modellen mit höheren Wertebereichen zwangsläufig höhere Fehler zu erwarten. Vor diesem Hintergrund ist die Größe von *MSPE* bzw. *RMSPE* nicht allein von der Qualität des Modells, sondern auch von den Eingangsdaten abhängig.

Die Kriging Standardabweichung σ_K^2 (vgl. Gl. (4.24) bzw. (4.29)) gibt den Schätzfehler des Modells für die Schätzung an der Stelle x_0 , bei Entfernung der Stichprobe x_0 , an. Dividiert man den Kreuzvalidierungsfehler aus Gl. (4.34) durch die Standardabweichung σ_K^2 , dann kann man den aktuellen und geschätzten Fehler vergleichen:

$$\frac{Z(x_i) - Z^*(x_{[i]})}{\sigma_K} \quad (4.43)$$

Falls der mittlere quadrierte, standardisierte Schätzfehler aus Gl. (4.38) ungefähr eins ergibt,

$$\frac{1}{n} \sum_{i=1}^n \frac{(Z(x_i) - Z^*(x_{[i]}))^2}{\sigma_K^2} \cong 1, \quad (4.44)$$

dann ist der aktuelle Schätzfehler gleich dem Durchschnitt des Schätzfehlers aus dem Modell. Diese Größe zeigt die Angemessenheit des Modells und seinen Parametern (WACKERNAGEL, 2003).

Die Beschreibung der Anpassungsgüte des Modells an die zugrunde liegenden Daten geschieht auch durch den empirischen Korrelationskoeffizienten r mittels einer Normierung der empirischen Kovarianz durch das Produkt der Standardabweichungen:

$$r = \frac{C(Z^*, Z)}{s(Z^*) \cdot s(Z)} \quad (4.45)$$

Für die Beurteilung der Stärke des Zusammenhanges zwischen beobachteter und geschätzter Variable an den Stützstellen kann der Rangkorrelationskoeffizient ρ_S nach Spearman ($-1 < \rho_S < 1$) verwendet werden:

$$\rho_S = 1 - \frac{6 \cdot \sum_{i=1}^n (rg(Z_i^*) - rg(Z_i))^2}{n(n^2 - 1)} \quad (4.46)$$

mit $rg(Z_i)$ Rang von Z_i

Der Wert $\rho_S = -1$ bedeutet funktional entgegengesetzter, der Wert $\rho_S = 0$ kein und der Wert $\rho_S = 1$ vollständiger Zusammenhang.

Ein weiterer interessanter Ansatz ist das Jackknifing, besonders mit weniger als 50 bis 100 Datenpunkten, wenn sich bereits in der Variographie kaum ausreichend Datenpunkte bilden lassen. Dabei wird ein Punkt aus den bekannten Stützpunkten geschätzt, wobei jeweils einer dieser Punkte der Reihe nach entfernt wird. Sämtliche Schätzungen - auch Pseudowerte genannt - werden letztlich gemittelt und ergeben die Jackknife Schätzung ZJ .

Die Beurteilung des Interpolationsergebnisses kann zunächst auch visuell in Form von Darstellung der Wertepaare in einem Streudiagramm erfolgen. Ideal ist hierfür die Lage aller Wertepaare auf einer 45°-Geraden ($Z^*(x_i) = Z(x_i)$) als Nachweis der bedingten Unverzerrtheit. Dieser Idealzustand wird sich im Regelfall nicht ergeben, sondern zumeist eine Punktwolke in Form einer lang gestreckten Ellipse. Diese ist mit ihrer Längsachse genau dann an der 45°-Geradenorientiert, wenn keine systematischen Über- oder Unterschätzungen auftreten und folglich der Erwartungswert der Schätzung dem Mittelwert der Stichprobenpopulation entspricht.

4.5.2 Testen einer Kriginganordnug

Die Erfahrung hat gezeigt, dass sich die Krigingvarianz als ziemlich unsensibler Parameter für die Beurteilung der Qualität einer Kriginganordnung erweist (ARMSTRONG, 1998). Die Gleichungen (4.30) und (4.31) sind hierfür nützlicher. Gleichung (4.30) zeigt, falls das Gewicht des Mittelwertes in SK λ_M^{SK} gering ist, hängt die Schätzung hauptsächlich von den lokalen Werten $Z(x_i)$ ab, d. h. von den Daten der Kriginganordnung und nicht vom geschätzten Mittelwert. Der Grad der erforderlichen Stationarität ist daher geringer. Gl. (4.31) zeigt, falls λ_M^{SK} gering ist, ist die Varianz von OK nahe zur Varianz von SK, d. h. geringe Genauigkeit wird verloren durch Schätzung des Mittelwertes. RIVOIRARD (1987) hat gezeigt, dass je weniger Information aus der Krigingnachbarschaft erhältlich ist, desto wichtiger ist der Mittelwert. Dieser Parameter zeigt daher an, wie verstreut die Daten in Bezug auf das Variogrammodell sind und bis zu welchem Ausmaß sich die Hypothese der Stationarität stützt.

5 AUSWERTUNG UND ERGEBNISSE

5.1 Datengrundlage

5.1.1 Immobiliendatenbank immo netZT

Die Immobiliendatenbank „*immo netZT*“ ist ein Service der Firma ZT-Datenforum registrierte Genossenschaft mit beschränkter Haftung in Graz, die von den Genossenschaftsmitgliedern gesammelten und erhobenen Daten über Grund und Boden via Internet zugänglich zu machen (Abb. 5.1).



Abb. 5.1: Internetportal des ZT-Datenforum (<http://www.zt.co.at>) [30.04.2011]

In der Datenbank sind zur Zeit über 170.000 Vergleichswerte für die Steiermark und Kärnten verfügbar. Insgesamt sind aus 1593 Katastralgemeinden in der Steiermark und 746 Katastralgemeinden in Kärnten Datensätze enthalten, die jeweils die Tagebuchzahl, das Verkaufsdatum, den Verkäufer, die Katastralgemeinde, die Grundstücksnummer, die Nutzung, den Gesamtpreis, den Preis pro m², geografische Koordinaten, Gauß-Krüger Koordinaten und andere relevante Informationen (z. B. Dienstbarkeiten, Zubehör) umfassen. Durch die Erfassung der Koordinaten kann „*immo netZT*“ auch als Geodatenbank bezeichnet werden.

Über die Webplattform des ZT-Datenforum (Abb. 5.2) können nach Registrierung die Datensätze abgefragt und mithilfe von Karten übersichtlich präsentiert werden. Die Datenbank „*immo netZT*“ beinhaltet die Kaufpreise für Bauland, Landwirtschaft und Wald in der Steiermark seit 1995. Für die vorliegende Masterthesis standen die Daten aus dem Gerichtsbezirk Schladming zur Verfügung, die von der Firma ZT-Datenforum registrierte Genossenschaft mbH zur Verfügung gestellt wurden.

ZT datenforum

WIR ÜBER UNS | SUCHE | PLÄNE | ORTSPÄNE | 3D | PRODUKTE | WARENKORB | LOGIN | KONTAKT

>> HOME

Wanderkarte
Naturbestand
Flächenwidmung
Kataster
Luftbild
Ortsplan
Rad-/Laufstrecke

Willkommen auf der Webplattform des ZT-Datenforum

Das ZT-Datenforum ist eine Genossenschaft von Ziviltechnikern, mit der Zielsetzung von den Genossenschaftsmitgliedern erhobene und bereits existierende Informationen über Grund und Boden via Internet zugänglich zu machen.

Diese Webseite erleichtert Ihnen die Suche nach Karten und Plänen von allen Gebieten Österreichs. Bei uns finden Sie Informationen unterschiedlichster Themenbereiche (Digitale Orthophotos, Luftbilder, Katasterpläne, Naturbestandsdaten, Freizeitkarten etc.) unabhängig vom Ersteller.

immo netZT ist die schlaue **Online-Datenbank für Sachverständige, Ziviltechniker, Banken und die Immobilienbranche.**
Vergleichswerte bzw. Vergleichspreise von Immobilien in Österreich

Pläne in Österreich zum Thema **Wanderkarte** gibt es in den **Bundesländern**

[Burgenland](#)
[Kärnten](#)
[Niederösterreich](#)
[Oberösterreich](#)
[Salzburg](#)
[Steiermark](#)
[Tirol](#)
[Vorarlberg](#)
[Wien](#)

map4you

ECG - Information

© Copyright

Abb. 5.2: Webplattform des ZT-Datenforum (<http://www.map4you.at>) [30.04.2011]

5.1.2 Digitale Katastralmappe

Die digitale Katastralmappe (DKM) enthält den grafischen Datenbestand des Grundsteuer- und Grenzkatasters, kurz Kataster, im Koordinatensystem der österreichischen Landesvermessung in digitaler Form und ist mit den Daten der Grundstücksdatenbank (GDB) und Koordinatenbank (KDB) konsistent. Sie liegt seit 2005 flächendeckend für ganz Österreich vor und ersetzt damit die analoge Katastralmappe. Die DKM dokumentiert in Verbindung mit dem Grundbuch Eigentumsverhältnisse an Grund und Boden und beinhaltet Grundstücksgrenzen, Grundstücksnummern, Nutzungsarten, Grenzpunkte und amtliche Vermessungspunkte. Sie wird vom Bundesamt für Eich- und Vermessungswesen (BEV) geführt und laufend aktualisiert. Die DKM kann als DXF-Datei für CAD-Systeme oder als Shape-Datei für GIS bezogen werden. Für die gegenständlichen Untersuchungen wurden die sog. Stichtagsdaten in Form der Shape-Dateien verwendet, die als kostengünstigstes Produkt über das BEV bezogen werden können.

5.2 Datenaufbereitung

Die Daten der Grundstückspreise wurden als MS®-Excel Datei zur Verfügung gestellt und beinhalten für Bauland 296 Datensätze mit Grundstückspreisen im Gerichtsbezirk (GB) Schladming und 465 Datensätze im ehemaligen Gerichtsbezirk Gröbming für den Zeitraum von 1995 bis 2009. Mit 1.7.2002 wurde der GB Gröbming aufgelöst und dem GB Schladming zugeordnet, in der Arbeit wird aber weiterhin die Unterteilung aufgrund der unterschiedlichen Charakteristika beibehalten. Der alte GB Schladming ist touristisch weit mehr erschlossen und besonders in Tourismusgebieten steht oft nur mehr wenig Bauland zur Verfügung. Die Grundstücke werden hinsichtlich Ihrer Lage und Beschaffenheit unterschiedlicher, woraus mehr Preisstreuung resultiert (PÖCHINGER, 2009). Die Unterteilung in die alten GB Schladming und Gröbming erscheint daher sinnvoll.

5.2.1 Explorative Datenanalyse

5.2.1.1 Fehlerbereinigung und Valorisierung

Der Grundstückspreis pro Quadratmeter, welcher dem Grundstück zugeordnet ist, ist Ausgangsbasis für die späteren Interpolationsmodelle. Als erster Schritt wurden auffallende und grobe Datenfehler allein anhand der größten, völlig unplausiblen Extremwerte eliminiert oder korrigiert (vgl., Tab. 5.1).

Preis pro m ²		GB Schladming	GB Gröbming
Größte Werte	1	90000,00	35000,00
	2	29069,13	588,44
	3	25000,00	338,55
	4	18168,21	122,58
	5	965,38	112,68
Kleinste Werte	1	2,75	1,00
	2	3,60	3,50
	3	7,00	3,56
	4	7,00	3,96
	5	7,27	4,00

Tab. 5.1: Extremwerte für den Grundstückspreis pro m² aus den Originaldaten

Der Grund für die außergewöhnlich hohen Werte liegt einerseits in einer fehlerhaften Angabe der Grundstücksgröße wodurch der Grundstückspreis pro Quadratmeter falsch berechnet wird und andererseits in der Tatsache, dass auch der Verkauf von

Liegenschaftsanteilen und der Verkauf von Grundstücken inklusive Nutzungsrechten oder Inventar erfasst ist. Des weiteren sind manchmal Kompensationsgeschäfte im Kaufpreis enthalten. Die niedrigen Kaufpreise kommen durch eine falsche Zuordnung der Nutzung zustande, indem Bauland anstatt landwirtschaftlicher Fläche vergeben wurde. Datensätze mit unter fünf Euro pro Quadratmeter werden gelöscht, weil diese nach Expertenwissen nicht der Realität entsprechen (PÖCHINGER, 2009). Aus diesen Gründen mussten zehn Datensätze eliminiert und sieben korrigiert werden.

Da sich die Verkaufspreise auf einen Zeitraum von 15 Jahren beziehen, müssen alle Grundstückspreise auf eine vergleichbare Basis gestellt werden. Die Basis des Vergleichs bildet der Geldwert mit Stand 2009. Die in der Vergangenheit entrichteten Kaufpreise lassen sich grundsätzlich mittels Valorisierung

$$K_n = K_0 \left(1 + \frac{p}{100}\right)^n \quad (5.1)$$

mit

K_0 ... Preis im Verkaufsjahr

K_n ... Preis im Vergleichsjahr (2009)

p ... jährliche Preisänderung in Prozent

n ... Anzahl der Jahre

umrechnen. Die jährliche Preisänderung der Grundstückspreise im Bauland für den jetzigen GB Schladming (inkl. ehemaligen GB Gröbming) wurde im Rahmen der Masterthesis von PÖCHINGER (2009) mit 2.5% ermittelt. Die statistischen Maßzahlen und das Histogramm der Fehler bereinigten und valorisierten Grundstückspreise pro Quadratmeter für den Gerichtsbezirk zeigen die Tab. 5.2 und Abb. 5.3.

Pv_m2	GB Schladming	GB Schladming
Mittelwert	76,78	48,24
Median	69,58	46,52
Varianz	2546,11	549,67
Standardabweichung	50,46	23,45
Minimum	7,92	5,52
Maximum	313,40	238,10
Spannweite	305,48	232,58
Interquartilbereich	51,10	31,60
Schiefe	1,411	1,499
Kurtosis	3,073	9,182

Tab. 5.2: Statistische Maßzahlen der valorisierten Grundstückspreise pro m2

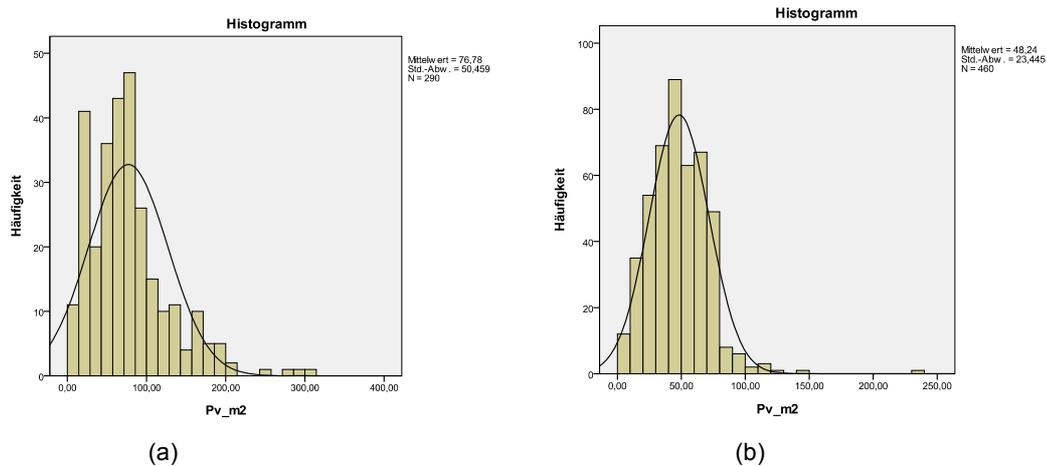


Abb. 5.3: Histogramm der valorisierten Grundstückspreise pro m² im GB Schladming (links) und Gröbming (rechts)

Die Daten liegen auf Ebene der Katastralgemeinde (KG) vor. Da in manchen Katastralgemeinden die Anzahl der Datenwerte gering ist (KG Ennsling, Weißenbach, und Untertal im GB Schladming und KG Diemlern, Lengdorf und Sonnberg in GB Gröbming) werden diese auf Ebene der politischen Gemeinde zusammengefasst. Aber auch für die größeren Katastralgemeinden spricht nichts gegen das Zusammenfassen, da einerseits die Mikrostruktur eines Ortes (Verkehrssystem, Siedlungsstruktur) einheitlich für die gesamte Gemeinde zu betrachten ist und andererseits weist die deskriptive Statistik der größeren zu einer Gemeinde gehörenden Katastralgemeinden (KG Rohrmoos und KG Untertal sowie KG Ramsau und KG Leiten) ohnehin ähnliche Werte auf.

Die statistischen Maßzahlen und Histogramme für jede Katastralgemeinde sind im Anhang 2, für jede politische Gemeinde in Anhang 3 und für die zwei Gerichtsbezirke in Anhang 4 angefügt. Obwohl diese Daten bereits von PÖCHINGER (2009) im Rahmen seiner Masterthesis untersucht und fehlerbereinigt wurden, zeigt dies wiederholt die Wichtigkeit einer vorhergehenden Datenanalyse.

5.2.1.2 Ausreißeranalyse

Der Begriff Ausreißer bezeichnet im gegenständlichen Fall Kaufpreise, die für einen Preisvergleich oder eine wissenschaftliche Analyse nicht geeignet sind. Kaufpreise, die nicht im üblichen, gewöhnlichen Geschäftsverkehr zustande gekommen sind, können u. a. folgendermaßen auftreten (PÖCHINGER, 2009):

- Grundstückspreise, die von der öffentlichen Hand aus siedlungspolitischen Gründen zu Vorzugspreisen abgegeben werden,

- Preise aus einem Zwangsversteigerungsverfahren, wenn nur der Ausrufungspreis (Hälfte des Schätzwertes) erzielt werden konnte,
- Liebhaberpreise oder Preise, die Spekulationen zuzuordnen sind,
- Preise unter Ausschluss anderer Bewerber, z. B. Arrondierungsverkäufe,
- Preise aufgrund überhöhter Angebote zur beschleunigten Abwicklung eines Kaufes (z. B. im Interesse eines Mobilfunkbetreibers),
- Gefälligkeitspreise unter Verwandten oder Freunden.

Die Detektion von Ausreißern kann visuell einfach und rasch mithilfe eines Box-Plots (Gl. 3.19) erfolgen (Abb. 5.4).

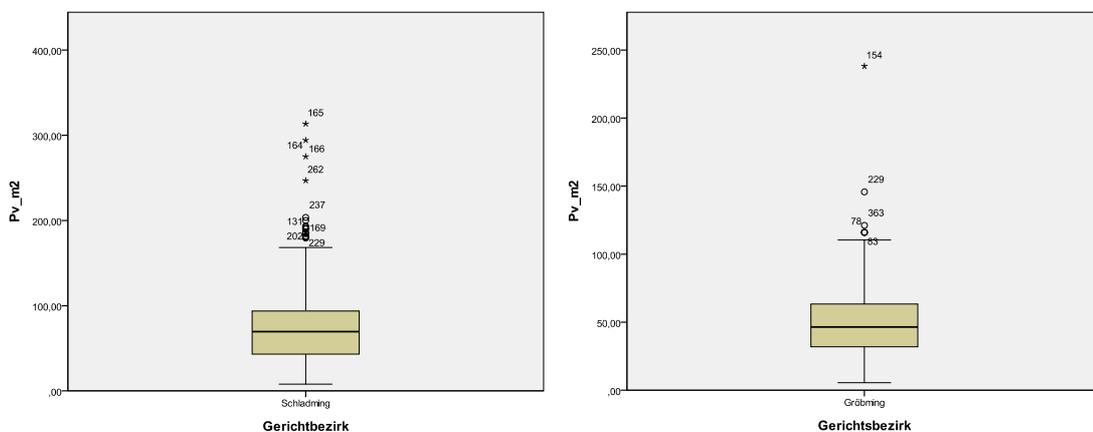


Abb. 5.4: Box-Plot der valorisierten Grundstückspreise pro m² im GB Schladming (links) und Gröbming (rechts)

Auf Ebene der Katastralgemeinde werden für den GB Schladming drei Preiswerte (zwei in der KG Leiten, einer in der KG Haus), auf Gemeindeebene keiner und auf Gerichtsbezirksebene vier (zwei in der KG Leiten, einer in der KG Ramsau einer in der KG Klaus) als Ausreißer detektiert (vgl. Anhang 5). Im GB Gröbming werden auf Ebene der Katastralgemeinde vier Ausreißer (je einer in den KG Lengdorf, Michaelerberg, Öblarn, Gröbming), auf Gemeindeebene keiner und auf Gerichtsbezirksebene einer (KG Gröbming) ermittelt.

Die Verwendung des 4-Sigma-Bereichs (Gl. 3.18) zeigt auf Katastralgemeindeebene die zwei Werte in der KG Leiten und den jeweils einen Wert in den KG Gröbming, Michaelerberg und Öblarn, auf Gemeindeebene dieselben Werte in Gröbming, Michaelerberg und Öblarn und auf Gerichtsbezirksebene vier Preiswerte (zwei Werte in der KG Leiten, ein Wert in der KG Ramsau und ein Wert in der KG Gröbming) an.

Der Walsh-Test nach den Gl. (3.20) bis (3.22) hingegen detektiert auf Gerichtsbezirksebene für Schladming keinen Preiswert als Ausreißer, für Gröbming jeweils einen in der KG Gröbming und in der KG Michaelerberg (Abb. 5.5).

AUSREISSERTEST nach WALSH GB Schladming	AUSREISSERTEST nach WALSH GB Gröbming
Datensätze: 290	Datensätze: 460
Mittelwert: 76.78	Mittelwert: 48.24
Standardabweichung: 50.46	Standardabweichung: 23.44
1 kleinster Ausreißer falls <0 : 8.45	2 kleinste Ausreißer falls <0 : 3.97
1 größter Ausreißer falls >0 : -84.63	2 größte Ausreißer falls >0 : 8.30

Abb. 5.5: Walsh-Test für valorisierte Grundstückspreise pro m² auf Gerichtsbezirksebene

Auf Ebene der Gemeinde allerdings detektiert auch der Walsh-Test zwei Preiswerte in der KG Leiten und einen in der KG Ramsau als Ausreißer.

Damit ergeben die einfachen Methoden Box-Plot und 4-Sigma-Bereich die Meinung der Immobilienexperten wieder, die Grundstückspreise über 250 € pro Quadratmeter als nicht repräsentativ sehen (PÖCHINGER, 2009). Unter Berücksichtigung dieses Expertenwissens wurden daher diese Preiswerte eliminiert, sodass letztendlich 287 Datenwerte im GB Schladming für die nachfolgenden Interpolationen verbleiben. Im GB Gröbming ist das Preisniveau niedriger, da hier starke Tourismusgemeinden wie im GB Schladming fehlen. Die drei Ausreißer werden jedoch ebenfalls eliminiert, da Nachforschungen ergeben haben, dass diese Preise tatsächlich nicht repräsentativ sind. Der hohe Preiswert in der KG Gröbming inkludiert ein Gebäude, jener in der KG Michaelerberg war eine familieninterne Transaktion und jener in der KG Öblarn eine Einbringung des Grundbesitzers in seine private Stiftung. Somit verbleiben 457 Datensätze für den GB Gröbming. Die weiteren Untersuchungen auf Ebene der Katastralgemeinde werden aus den bereits angeführten Gründen nicht weiter ausgeführt, sondern nur mehr auf Gemeinde- und Gerichtsbezirksebene.

5.2.1.3 Test auf Normalverteilung

Eine erste grafische Überprüfung erfolgt anhand des QQ-Diagramms. Im QQ-Diagramm werden nicht die theoretischen Quantile aufgetragen, sondern die Quantile („erwarteter Wert von ...“) derjenigen Verteilung, die aus der Standard-Verteilung (für die Zufallsvariable Z) durch Transformation mit alternativen Schätzern für μ und σ (z. B. empirischer Mittelwert, empirische Standardabweichung) erhalten werden. Für die Anpassung an das Modell einer Normalverteilung können als Approximation

für die Wahrscheinlichkeitswerte p_i zu den nach der Größe geordneten Beobachtungen $x_{(i)}$ ($i=1, \dots, n$) folgende Möglichkeiten, die aus unterschiedlichen theoretischen Überlegungen resultieren, verwendet werden:

Methode	$(x_{(i)}, p_i)$	
Blom (Voreinstellung):	$(i - 3/8)/(n + 1/4)$	
Tukey:	$(i - 1/3)/(n + 1/3)$	
Rankit:	$(i - 1/2)/n$	(5.2)
van der Waerden:	$i/(n + 1)$	

Im PP-Diagramm (**P**robability-**P**lot) werden anstatt der Quantilfunktionen die Verteilungsfunktionen direkt verwendet, wobei Abweichungen im Wesentlichen in der Mitte des PP-Plotes feststellbar sind. Die QQ-Diagramme für die einzelnen Gemeinden bzw. die gesamten Gerichtsbezirke sind in Anhang 6 und die PP-Diagramme in Anhang 7 enthalten. Die Preiswerte des GB Gröbming weisen eine bessere Annäherung an die Normalverteilung infolge der geringeren Streuung auf. Für den ganzen GB Gröbming zeigt das PP-Diagramm sogar eine gute Annäherung an eine Normalverteilung an, während für den GB Schladming doch erkennbare Abweichungen auftreten.

Als weiteres Maß kann das Verhältnis zwischen Spannweite und Standardabweichung nach Gl. (3.15) untersucht werden. Mit den Werten aus der Tabelle der kritischen Quotienten B/s nach PEARSON & STEPHENS (1964) aus Anhang 1 ist die Prüfgröße B/s für ein Signifikanzniveau von 0.05 und 0.10 für die Gemeinden der beiden Gerichtsbezirke in den Tab. 5.3 und 5.4 angegeben.

Gemeinde	N	B	s	B/s	u.Q.		o.Q.	
					0.05/0.10	0.10/0.05		
Aich	34	70,90	21,63	3,278	3.56/3.68	4.81/5.01		
Gössenberg	10	13,45	5,04	2,681	2.67/2.76	3.57/3.68		
Haus	58	143,27	27,07	5,293	3.94/4.06	5.27/5.48		
Oberhaus	9	89,53	26,91	3,327	2.59/2.68	3.45/3.55		
Pichl	16	73,34	22,49	3,261	3.01/3.12	4.09/4.24		
Preunegg	3	90,02	50,62	1,778	1.758/1.782	1.997/1.999		
Ramsau	68	181,86	50,35	3,612	4.04/4.17	5.39/5.61		
Rohrmoos	35	184,31	48,91	3,768	3.58/3.70	4.84/5.04		
Schladming	54	231,87	54,77	4,233	3.89/4.01	5.20/5.41		
GERICHTSBEZIRK	287	238,94	45,49	5,253	4.95/5.07	6.32/6.55		

Tab. 5.3: Statistik der Prüfgröße B/s für die Gemeinden im GB Schladming

Gemeinde	N	B	s	B/s	u.Q. 0.05/0.10	o.Q. 0.10/0.05
Gröbming	173	110,45	20,28	5,446	4.68/4.81	6.05/6.27
Großsölk	25	58,40	13,10	4,458	3.34/3.45	4.53/4.71
Kleinsölk	15	56,21	15,91	3,533	2.97/3.07	4.02/4.17
Michaelerberg	31	50,91	14,69	3,466	3.49/3.61	4.73/4.92
Mitterberg	92	102,56	20,43	5,020	4.25/4.38	5.62/5.84
Öblarn	40	59,81	14,08	4,249	3.67/3.79	4.96/5.16
Pruggern	28	42,85	12,16	3,524	3.41/3.52	4.63/4.81
St. Martin	38	90,45	15,34	5,896	3.63/3.75	4.91/5.11
St. Nikolai	15	18,52	4,79	3,866	2.97/3.07	4.02/4.17
GERICHTSBEZIRK	457	110,45	21,00	5,260	5.28/5.40	6.64/6.86

Tab. 5.4: Statistik der Prüfgröße B/s für die Gemeinden im GB Gröbming

Die in den Tabellen 5.3 und 5.4 blau gekennzeichneten Werte erfüllen die Annahme einer Normalverteilung nach der Prüfmethode von PEARSON & STEPHENS. Interessanterweise wird die Annahme der Normalverteilung für den GB Gröbming knapp abgelehnt, obwohl für fünf von sieben Gemeinden diese bestätigt wird. Umgekehrt wird für den GB Schladming diese bestätigt, obwohl die Prüfgrößen der einzelnen Katastralgemeinden nicht so gut passen wie im GB Gröbming und das PP-Diagramm ebenfalls schlechter passt.

Die Ergebnisse des statistischen Tests von Kolmogorov-Smirnov (vgl. Anhang 8) ergeben eine Ablehnung der Hypothese auf Normalverteilung für den GB Schladming und keine Ablehnung für den GB Gröbming. Für den GB Schladming kann auch nach einer Box-Cox Transformation (Gl. 3.17) keine Verbesserung der Signifikanzwerte zur Annahme einer Normalverteilung erreicht werden. Durch weitere Untersuchungen dieser Daten konnten bimodale, multimodale und andere Verteilungsformen nicht festgestellt werden. Unter Berücksichtigung, dass nur für drei Gemeinden im GB Schladming die Hypothese der Normalverteilung abgelehnt wird, wird für die weiteren geostatistischen Untersuchungen von annähernder Normalverteilung ausgegangen.

5.2.1.4 Test auf Autokorrelation

Da die bereitgestellten tabellarischen Daten bereits die Information der Örtlichkeit beinhalten – das Grundstück wird durch einen koordinativ angegebenen Punkt repräsentiert – wird in einem ersten Schritt die Autokorrelation mithilfe des

Programms *crimestat*®, einem Programm für die räumliche Analyse der Örtlichkeiten von Kriminalvorfällen, bestimmt. Mit diesem kann der MORAN's I-Index und der GEARY's C-Index berechnet werden. Beide Indizes zeigen leichte positive Autokorrelation (Tab. 5.5), die detaillierten Ergebnisse für jede Katastralgemeinde und sind in Anhang 9 dargestellt. Mit Verwendung der Distanzkorrektur wird keine Autokorrelation angezeigt, allerdings ist diese auf die Entfernung von einer Meile, eine typische Distanz für Kriminalereignisse, vorgegeben. Des weiteren können die Daten einem Trend unterliegen, der zuvor eliminiert werden muss. Dies wird im nächsten Kapitel mithilfe der Variographie untersucht.

Gemeinde	n	Moran's I	Geary's C	Moran's I Distanzkorrektur	Geary's C Distanzkorrektur	Nearest Neighbour Index
GB Schladming	287	0.22	0.74	-0.01	1.02	0.39
GB Gröbming	457	0.20	0.75	0.07	0.95	0.24

Tab. 5.5: Moran's Index, Geary's C-Index und NN-Index für die Gerichtsbezirke

Mit *Crimestat*® kann auch die Punktverteilung mit der Methode NNA untersucht werden. Falls die empirische Nearest-Neighbour Distanz (NN-Distanz) in etwa gleich der theoretischen (zufälligen) NN-Distanz ist, ergeben sich für den *NNI* -Werte um 1. Werte von *NNI* kleiner 1 indizieren eine zunehmende gruppierte Verteilung (clustered pattern), Werte von *NNI* größer als 1 eine zunehmende Tendenz zu gleichmäßiger Verteilung (regular pattern). Im Gerichtsbezirk Schladming weist von den neun Gemeinden nur Oberhaus eine zufällige Verteilung (random pattern) auf, im Gerichtsbezirk Gröbming zeigen die zwei Gemeinden Pruggern und Öblarn eine zufällige und Kleinsölk eine gleichmäßige Verteilung an. Die restlichen Gemeinden mit Ausnahme von Preunegg zeigen eine gruppierte Verteilung, die einzelnen Ergebnisse sind in Anhang 9 angeführt. Die Gemeinde Preunegg muss wegen der sehr geringen Anzahl von Grundstückspreisen aus den Betrachtungen ausgeschlossen werden, weil mit drei Verkaufspreisen keine signifikanten Berechnungen bzw. Aussagen möglich sind. Die Betrachtung der ganzen Gerichtsbezirke ergibt aufgrund des großen Untersuchungsgebietes erwartungsgemäß eine gruppierte Verteilung (Tab. 5.5). Die Ortschaften liegen überwiegend im von Westen nach Osten ausgerichteten Ennstal, zwischen diesen liegen große landwirtschaftliche Flächen.

5.2.1.5 Übernahme der Daten in ArcGIS

Obwohl in den tabellarischen Daten der Grundstückspreise geografische Koordinaten bzw. Gauß-Krüger Koordinaten im Landesvermessungssystem enthalten, wurden diese unabhängig von dieser Information mit den Grundstückspolygonen des Katas-

ters verbunden. Die Verbindung konnte durch Zusammenfassen der in beiden Datensätzen enthaltenen Katastralgemeindenummer (KGNR) und Grundstücksnummer (GNR) zu einer eindeutigen KGNR_GNR bewerkstelligt werden. Der Grund für diesen Schritt ist die Möglichkeit der Überprüfung der Datensätze zufolge fehlerhafter Eingaben, die durch die bisherigen Untersuchungen (Ausreißeranalyse) nicht bewerkstelligt werden konnten, wie fehlerhafte Angaben der Grundstücksnummer oder Koordinaten.

Die Verbindung (join) der 744 tabellarischen Datensätze ergab tatsächlich eine geringere Anzahl von 663 Datensätzen in ArcGIS. Den Hauptgrund für diese Abweichung bilden 41 Grundstücke, die im gegebenen Zeitraum von 1995 bis 2009 zweimal verkauft wurden. Die Bereinigung erfolgte durch Elimination des älteren Datensatzes. Die verbleibenden Datensätze enthielten fehlerhafte Grundstücksnummern (Ziffernstürze, fehlende oder falsche Unterteilungsnummern) und idente Datensätze. Diese wurden korrigiert bzw. idente eliminiert, sodass für die Untersuchungen in ArcGIS 696 Datensätze mit Verkaufspreisen zur Verfügung stehen.

Da im Gegensatz zu den deterministischen Interpolationsmethoden Kriging ausschließlich mit Punkthemen (point features) als Input arbeitet, musste das durch die Verbindung erhaltene Polygonthema (polygon feature) auf ein Punkthema rückgeführt werden. Mit dem Werkzeug 'Feature to Point' und der Option 'inside' wird ein neues Punkthema erzeugt, bei welchem alle Attribute der Tabelle übernommen und einem Punkt innerhalb des entsprechenden Polygons zugeordnet werden.

5.2.2 Variographie

Bis jetzt wurden die Daten anhand des Verkaufspreises mit Ausnahme der Untersuchungen in Crimestat® eindimensional untersucht. Mithilfe der Variographie werden die Daten einer zweidimensionalen Betrachtung unterzogen. Wenn die Ergebnisse der Variographie für eine ortsabhängige Variable räumlichen Zusammenhang zeigen, kann für die Verkaufspreise eine Oberfläche mittels Kriging erstellt und damit Schätzwerte für Bereiche ohne Stützstellen berechnet werden. Weiters besteht die Möglichkeit, räumliche Strukturen visuell leichter zu erfassen. Die Regionalisierung birgt jedoch auch die Gefahr einer Fehlinterpolation infolge geringer räumlicher Zusammenhänge und/oder unzureichender Datendichte. Deshalb wird zuvor die Eignung für eine Regionalisierung geprüft, welche von mehreren Faktoren die untereinander in Zusammenhang stehen abhängt. Diese Faktoren sind die Anzahl der Datenpunkte (Stützstellen) und deren Verteilung im Raum sowie die räumliche Kontinuität der betrachteten Variable. Die räumliche Kontinuität lässt sich mit der Reichweite (range) und dem Nugget-Anteil am Schwellwert des Variogrammes der untersuchten Variable beschreiben.

5.2.2.1 Räumliche Statistik

Mithilfe des Werkzeuges ‚räumliche Statistik‘ wurde in ArcGIS nochmals Auswertungen der endgültig fehlerbereinigten Daten durchgeführt. Ein Vergleich der deskriptiven Statistik zeigt erwartungsgemäß eine Verbesserung (Anhang 10, 11), beim Normalverteilungstest erhöht sich leicht die Signifikanz für den Shapiro-Wilk Test auf Gemeindeebene, auf Gerichtsbezirksebene wird weiterhin keine Signifikanz angezeigt (Anhang 12). Abbildung 5.6 zeigt das Untersuchungsgebiet für die Gerichtsbezirke Schladming und Gröbming.

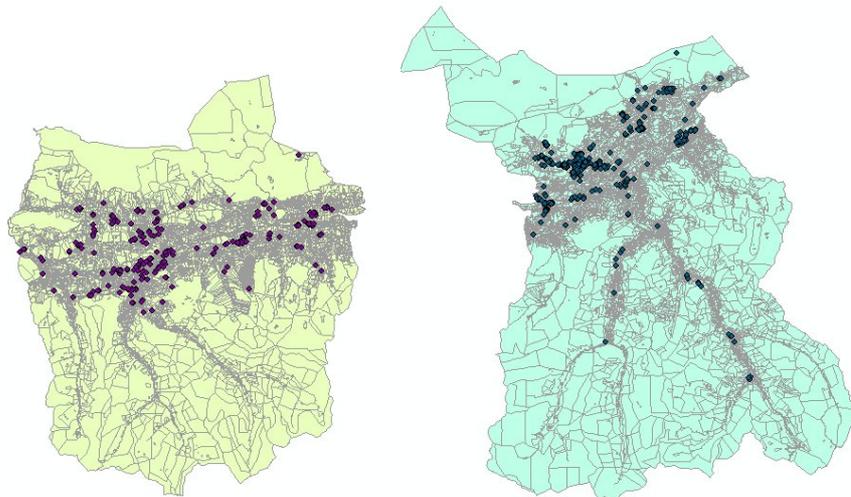


Abb. 5.6: Untersuchungsgebiet und Datenwerte der Grundstückspreise
links GB Schladming, rechts GB Gröbming

Der Test auf räumliche Autokorrelation mittels des Moran's I-Index in ArcGIS zeigt einen erhöhten Wert für positive Autokorrelation. Weiters wird das Verteilungsmuster für die Gemeinden des GB Schladming Aich, Oberhaus*, Schladming und die Gemeinden des GB Gröbming Großsölk*, Kleinsölk*, Pruggern*, St. Martin mit einem Signifikanzniveau von 10% als zufällig ausgewiesen (siehe Anhang 13 und 14). Zu beachten ist allerdings, dass die mit (*) gekennzeichneten Gemeinden einen Stichprobenumfang von weniger als 30 aufweisen. Dies gilt allgemein als zu klein, um Repräsentativität zu erreichen (BAHRENBERG et al., 1999).

5.2.2.2 Test auf Drift

Das Wissen um das Vorhandensein einer räumlichen Drift bzw. eines räumlichen Trends für das zu untersuchende Phänomen, d. h. dass richtungsabhängige Unterschiede in der Merkmalsausprägung der untersuchenden Variable auftreten ist für geostatistische Analysen essenziell. Bei einer klaren räumlichen Drift der Daten ist

die Bedingung der Stationarität verletzt, weshalb die Kriging Verfahren einen trendbereinigten Datensatz benötigen.

Die Untersuchungen wurden in ArcGIS mit dem Werkzeug Trendanalyse, einer dreidimensionalen Darstellung der Daten durchgeführt. Der Vorteil dieser Methode liegt darin, sofort den (globalen) Trendverlauf – linear, quadratisch – visuell zu erkennen und damit für das nachfolgende Kriging gezielt zu berücksichtigen. Die grafischen Ergebnisse der Trendanalyse sind in Anhang 15 und 16 zusammengefasst. Als einzige Gemeinde ist Gröbming nicht Trend behaftet (vgl. Tab. 5.6).

Trend GB Schladming		
	Ost-West	Nord-Süd
AICH	quadratisch von Mitte fallend	quadratisch von Mitte fallend
GÖSSENBERG	quadratisch nach Ost fallend	quadratisch von Mitte fallend
HAUS	quadratisch von Mitte fallend	linear nach Süd steigend
OBERHAUS	linear nach Ost steigend	quadratisch nach Süd steigend
PICHL	linear nach Ost steigend	linear nach Süd fallend
PREUNEGG	---- zu wenig Daten	---- zu wenig Daten
RAMSAU	quadratisch nach Ost fallend	quadratisch von Mitte fallend
ROHRMOOS	linear nach Ost steigend	quadratisch von Mitte steigend
SCHLADMING	gering linear nach Ost steigend	gering linear nach Süd fallend
Trend GB Gröbming		
	Ost-West	Nord-Süd
GRÖBMING	nicht vorhanden	nicht vorhanden
GROßSÖLK	quadratisch von Mitte steigend	quadratisch von Mitte steigend
KLEINSÖLK	quadratisch nach Ost steigend	quadratisch nach Süd fallend
MICHAELERBERG	quadratisch von Mitte steigend	quadratisch von Mitte fallend
MITTERBERG	linear nach Ost fallend	linear nach Süd steigend
ÖBLARN	linear nach Ost fallend	linear nach Süd steigend
PRUGGERN	quadratisch nach Ost fallend	linear nach Süd steigend
ST. MARTIN	quadratisch nach Ost steigend	quadratisch von Mitte steigend
ST. NIKOLAI	gering linear nach Ost steigend	gering linear nach Süd fallend

Tab. 5.6: Ergebnis der Trendanalyse

Abbildung 5.7 zeigt die drei-dimensionale Darstellung der Trendanalyse für den gesamten Datensatz. Demnach liegt ein linear fallender Trend von West nach Ost und ein quadratisch fallender von der Mitte in Nord-Süd Richtung vor. Dies lässt sich mit den touristischen Gemeinden im Westen und den im Norden und Süden vom Ennstal weniger besiedelten Regionen erklären.

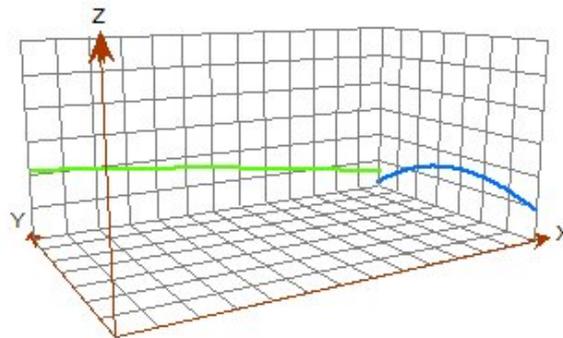


Abb. 5.7: Trendanalyse des gesamten Datensatzes

Mit der Methode der Voronoi-Diagramme, welche die Verteilung der Werte in den Vordergrund stellen, können ebenfalls Trends visualisiert werden. Beispielhaft ist in Abb. 5.7 das Voronoi-Diagramm für die Gemeinde Mitterberg angegeben, aus welchem ein fallender Trend nach Osten erkennbar ist.

Anstatt der Datenwerte können auch andere Parameter für verschiedene Zwecke verwendet werden, wie z. B. der Mittelwert der die lokale Glättung anzeigt oder die Standardabweichung, welche die lokale Variation anzeigt. Der Datenwert selbst gibt den lokalen Einflussbereich wieder.

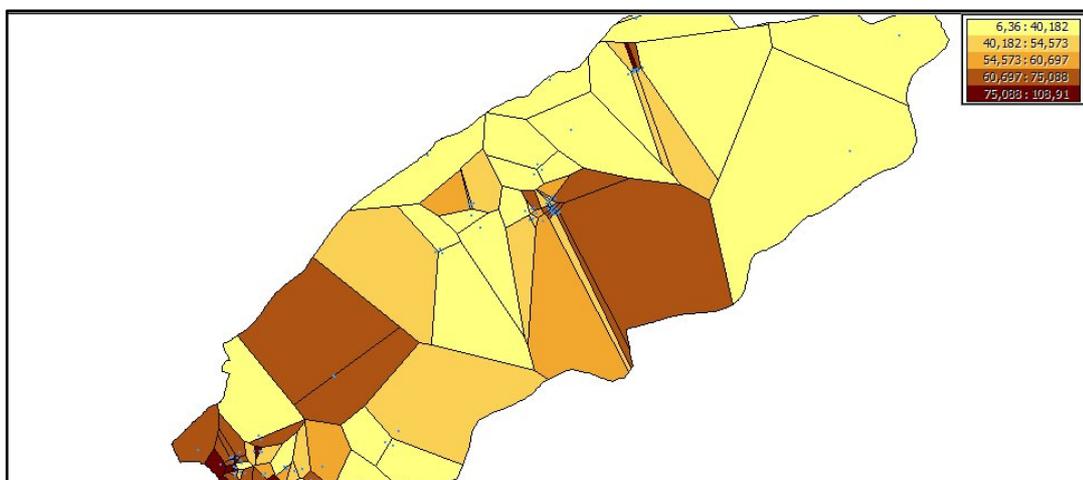


Abb. 5.8: Voronoi-Diagramm der Gemeinde Mitterberg

5.2.2.3 Variogrammanalyse

5.2.2.3.1 Toleranzkriterien

Bei der empirischen Variographie werden Differenzen von Wertepaaren $Z(x)$ und $Z(x+h)$ mit dem Abstand h gebildet (vgl. Kap. 3.2.3.2). Können bei einer gewählten Schrittweite h keine Wertepaare gebildet werden, weil im Untersuchungsgebiet keine zwei Datenpunkte den vorgegebenen Abstand aufweisen, oder können nur wenige Wertepaare gebildet werden, deren geringe Anzahl die Berechnung des empirischen Variogramms nicht ermöglicht, müssen Toleranzkriterien eingeführt und verwendet werden. Insbesondere bei gruppierten Verteilungen, wie im vorliegenden Fall, ist die Verwendung von Toleranzen erforderlich. Dazu können eine Schrittweitentoleranz Δh und eine Winkeltoleranz $\Delta\varphi$ angegeben werden, die die Verwendung auch außerhalb der Schrittweite bzw. der vorgegebenen Suchrichtung liegender Punkte erlauben. Zusätzlich kann in ArcGIS eine Bandbreite b angegeben werden, die bei größeren Schrittweiten zu einer Beschränkung der Datenpunkte auf die in unmittelbarer Nähe zur Suchrichtung φ gelegenen Punkte führt. Dadurch kann bei großen Schrittweiten und Winkeltoleranzen ein allzu großes Anwachsen der Toleranzbereiche verhindert werden. Mit der Verwendung von Toleranzen kann dem Problem einer zu geringen Datenmenge entgegengewirkt werden, wenn diese so gewählt werden, dass sie sich bei benachbarten Paarbildungen überlappen (z. B. $\Delta h > \frac{1}{2} h$). Damit gehen die Datenpunkte bei der empirischen Variographie mehrmals in die Berechnung verschiedener $\gamma(h)$ ein. Eng verbunden mit der Einführung von Toleranzen ist auch die Frage nach einer Mindestanzahl der Wertepaare $[Z(x_i), Z(x_i+h)]$ bei einer bestimmten Schrittweite, bis zu der die Berechnung der jeweiligen $\gamma(h)$ noch sinnvoll erscheint. Als Faustregel wird in der Literatur eine Anzahl von mehr als 50 Wertepaaren für die Berechnung des empirischen Variogramms angegeben (CHILÈS & DELFINER, 1999, Seite 38 oder BURROUGH & McDONELL, 1998, Seite 137), um zuverlässige Variogrammwerte zu erhalten.

Zu klein gewählte Toleranzbereiche führen zu einem stärker ausgeprägten erratischen Verhalten des empirischen Variogramms, wodurch die Auswahl und Anpassung der Variogrammfunktion erschwert werden kann. Im Gegensatz dazu führen zu groß gewählte Toleranzbereiche zu einer stärkeren Glättung, wodurch zwar eine einfachere Auswahl einer theoretischen Variogrammfunktion möglich ist, dies jedoch mit einem größeren Nugget-Effekt und einer größeren Reichweite verbunden ist. Die Festlegung optimaler Toleranzbereiche ist daher ein Kompromiss zwischen der Möglichkeit der Aufdeckung von Mikrostrukturen und anisotroper Verhältnisse (kleine Toleranzen), verbunden mit einer geringen Vertrauenswürdigkeit dieser Erkenntnisse (wenige Datenpunkte) und der Ermittlung eines strukturell verallgemeinerten Variogramms. Letzteres wird zwar in Teilbereichen des Untersuchungsgebietes

möglicherweise starke Abweichungen erzeugen, jedoch durch die Vielzahl von Wertepaaren eine statistisch abgesicherte Basis darstellen und damit im Durchschnitt aller Schätzwerte eine realistischere Prognose erlauben (GAU, 2010).

5.2.2.3.2 Variogrammodell

Nach Ermittlung des empirischen Variogramms ist dieses durch ein theoretisches Variogramm anzupassen. Die am häufigsten verwendeten Variogrammodelle sind das sphärische, das exponentielle und das Gaußsche Modell (vgl. Kap. 3.2.3.3). Im Regelfall wird ohne konkreten Bezug zur Datengrundlage dasjenige Variogrammodell ausgewählt, welches das empirische Variogramm am besten approximiert. Studien haben gezeigt, dass die Auswahl der Variogrammfunktion nur einen geringen Einfluss auf die durch Kriging berechneten Schätzwerte ausübt (GAU, 2010, CHILÈS & DELFINER, 1999). Die Auswahl des konkreten Variogrammodells ist hiernach als zweitrangig einzustufen, es benötigt einer groben falschen Angabe des Variogrammodells, um einen signifikanten Einfluss auf die Schätzungen mittels Kriging zu erhalten. Ein wichtiger Faktor ist jedoch, das Verhalten nahe des Ursprungs korrekt zu repräsentieren (vgl. Kap. 3.2.3.4.1), die spätere Steigung hat wenig Einfluss auf die Kriginggewichte. Das Verhalten im Ursprung (Nugget-Effekt und Steigung) haben einen sehr großen Einfluss auf die Kriging-Ergebnisse und spielen eine entscheidende Rolle für die Variogrammanpassung (ARMSTRONG, 1998).

Nach Auswahl eines theoretischen Variogrammodells sind seine Parameter zu ermitteln, d. h. die Anpassung der theoretischen Variogrammfunktion an das empirische Variogramm. Dies kann entweder auf subjektiver Basis in einem grafisch interaktiven Prozess oder unter Verwendung verschiedener mathematischer Kriterien erfolgen. Diesem Teil des Modellierungsprozesses wird übereinstimmend die größte Bedeutung zugemessen, weil hierdurch Schlussfolgerungen von den Eingangsdaten auf unbekannte Werte erlaubt werden sollen. Hauptargument gegen eine subjektive Anpassung ist die fehlende Objektivität oder die fehlende Nachvollziehbarkeit der gewählten Lösung. Die Kritik an einer automatisierten Anpassung bezieht sich darauf, dass diese unabhängig von einer kritisch zu prüfenden Datenmenge und hinreichend räumlicher Verteilung immer durchgeführt werden kann und je nach Dichte und Zuverlässigkeit der Daten mehr oder weniger Sinn ergibt (GAU, 2010). Die Variogramapproximation ist daher als Zwiespalt zwischen einer visuell möglichst guten und sachlich plausiblen Anpassung einerseits, sowie der mathematisch basierenden Anpassung andererseits zu betrachten. Das am häufigsten verwendete Anpassungsverfahren stellt die Kleinste-Quadrate-Methode dar.

5.2.2.3.3 Variogramparameter

Der Einfluss der Variogramparameter eines bestimmten Variogrammmodells lässt sich bereits aus den Grundgleichungen des Kriging-Schätzverfahrens ableiten. Eine größere Reichweite führt zu einem kontinuierlicheren Modellergebnis, da weiter vom Schätzpunkt entfernt liegende Datenwerte größere Gewichte für die Schätzung erhalten und etwaige klein- und mittlräumliche Variabilität durch Überbetonung der großräumlichen Variabilität tendenziell überdeckt werden. Ein größerer Schwellwert führt demgegenüber nicht zu einer Änderung des Modells. Hinsichtlich des Nugget-Anteils ist davon auszugehen, dass steigende Werte zu einer Vergleichmäßigung der Variabilität führen, weil die Gewichte für die Schätznachbarschaft ähnlicher werden. Parameterstudien zeigen, dass die Auswahl der Variogramparameter, d. h. die Anpassung einer konkreten Variogrammfunktion, von untergeordneter Bedeutung ist. Sicherzustellen ist lediglich, dass die Variogrammfunktion innerhalb der Schwankungsbreite des empirischen Variogramms liegt. Für ein gegebenes empirisches Variogramm zwischen diesem und der theoretischen Variogrammfunktion besteht keine Möglichkeit bessere Modelle durch eine genauere Anpassung zu erzeugen (GAU, 2010). Aus Gründen der visuellen Erscheinung des späteren Kriging-Modells ist die Wahl des Nugget-Wertes von größtem Interesse.

Für die Berechnung des Variogramms ist es wichtig, zumindest vier Richtungen zu verwenden (ARMSTRONG, 1998). Wenn das Variogramm nur in zwei senkrecht aufeinander stehenden Richtungen berechnet wird, ist es möglich, die Anisotropie völlig zu übersehen. Das wäre der Fall, wenn die Hauptachse 45° zur Richtung des berechneten Variogramms steht.

5.2.2.3.4 Auswahl des Schätzverfahrens

Erfolgt die Schätzung der zu untersuchenden Variable durch Anwendung geostatistischer Methoden, ist zunächst eine konkrete Kriging Methode auszuwählen. Hierbei kann auf eine Vielzahl verschiedener Verfahren zurückgegriffen werden (vgl., Kap. 4.4). In der überwiegenden Mehrzahl der publizierten Anwendungen wird das Ordinary Kriging und damit das ursprünglichste der geostatistischen Schätzverfahren herangezogen. Von den übrigen Verfahren weisen lediglich das Universal Kriging und das Indikator-Kriging eine häufigere Anwendung auf. In mehreren auf den qualitativen Vergleich verschiedener Kriging Methoden abzielenden Untersuchungen konnte nachgewiesen werden, dass bereits Ordinary Kriging gute bis sehr gute Schätzergebnisse liefert im Vergleich zu anderen z. T. weit komplexeren Verfahren. Dieser Nachweis resultiert einerseits auf der theoretischen Eigenschaft des OK, bei der Annahme einer bekannten Kovarianzstruktur und der Konstanz von Mittelwert und Varianz im Untersuchungsgebiet das beste Schätzmodell zu liefern (BLUE), und andererseits daraus, dass bei komplexeren Verfahren ein umfangreicher

Anwendereinfluss erforderlich ist, der das Modellergebnis unsicherer werden lassen kann.

Vor allem in Fällen, in denen starke Abweichungen von den theoretischen Voraussetzungen vorliegen, z. B. extrem schiefe Wertverteilungen können andere Kriging Methoden bessere Ergebnisse als OK liefern, häufig scheitern jedoch selbst nicht-lineare Verfahren (GAU, 2010). Durch das Einbeziehen externer Informationen können mittels Co-Kriging die Ergebnisse des einfachen OK übertroffen werden, aufgrund der Korrelation zu den primären Daten (vgl. Kap. 4.4.6).

Im Allgemeinen kann der grundsätzlichen Empfehlung gefolgt werden, die Ergebnisse zunächst mittels OK zu berechnen und erst dann zu prüfen, ob die Notwendigkeit besteht, komplexere Verfahren zu verwenden.

5.2.2.3.5 Suchbereich der Schätzung

Nach der Auswahl eines Schätzverfahrens ist durch den Anwender festzulegen, innerhalb welchen Bereichs um den jeweiligen Datenpunkt der Stichprobe vorhandene Messwerte für die Schätzung herangezogen werden sollen. Außerhalb liegende Werte werden für die Schätzung nicht verwendet. Die Größe und Form des Suchbereichs, die Mindest- und Maximalanzahl sowie eine Einteilung des Suchbereichs in Sektoren sind in ArcGIS® frei wählbar. Mit der Festlegung eines Suchbereichs sollen nur diejenigen Messwerte für die Schätzung herangezogen werden, die auf gleichen oder möglichst ähnlichen Ursachen basieren, wie sie auch für den Schätzwert angenommen werden. Dabei kann von der Grundannahme, dass bei räumlich naheliegenden Messpunkten eine größere Ähnlichkeit zu verzeichnen ist, ausgegangen werden. Andererseits ist der Suchbereich hinreichend groß zu wählen, um eine ausreichende Anzahl von Messwerten für die Schätzung zu gewährleisten und damit deren rechnerische Genauigkeit zu erhöhen. Zu kleine Suchbereiche können zu einer Überbewertung der kleinräumigen Variabilität führen, zudem ist dies mit einem verstärkten Screeneffekt verbunden, der zur Abschwächung der Gewichte für unmittelbar hinter von in Nachbarschaft des Schätzwertes liegenden Punkten führt (WACKERNAGEL, 2003). Eine Interpolation die auf Gewichten aus der inversen Distanz der Messwerte zum geschätzten Punkt beruhen zeigt keinen Screeneffekt. Die Wahl des Suchbereiches ist daher ein Kompromiss zwischen der angestrebten rechnerischen Genauigkeit des Schätzwertes und der notwendigen Richtigkeit des Schätzvorganges (GAU, 2010).

Obwohl dem Suchbereich eine hohe Bedeutung für das Kriging Ergebnis zugeschrieben wird, fehlen konkrete Grundsätze zur dessen Festlegung. Viele Empfehlungen gehen dahin, den Suchbereich hinsichtlich seiner Größe und Ausrichtung der Anisotropieellipse anzupassen. GAU (2010) kommt in seinen Studien

ebenfalls zu diesem Schluss und beurteilt die Zahl der Sektoren, die den Suchbereich einteilen, als zweitrangig.

5.3 Räumliche Interpolation

Nachdem sämtliche vorausgehenden notwendigen Untersuchungen durchgeführt wurden, erfolgt die Berechnung der einzelnen deterministischen und stochastischen Interpolationsmodelle und deren Gegenüberstellung im Rahmen der Validierung. Die Interpolationen werden für die gesamten Gerichtsbezirke als auch für einzelne Gemeinden berechnet. Aufgrund allgemeiner Anhaltspunkte zum Stichprobenumfang (BAHRENBURG, et al. 1999)

- je stärker die Werte der untersuchten Variable streuen, desto größer sollte der Stichprobenumfang sein,
- Stichprobenumfänge von weniger als 30 gelten allgemein als zu klein,
- für die Repräsentativität ist weniger der relative Anteil an der Größe der Grundgesamtheit von Bedeutung, als vielmehr die absolute Größe der Stichprobe

und unter Berücksichtigung einer ausreichenden Anzahl von Wertepaaren für die Erstellung der empirischen Variogramme (10 Datenpunkte ergeben 45 Wertepaare, 15 Datenpunkte ergeben 105 Wertepaare), werden Gemeinden unter 30 Datenwerte a priori von der Interpolation ausgeschlossen. Von den verbleibenden neun Gemeinden werden für jeden Gerichtsbezirk die drei am besten geeigneten als Auswahl für die Kriging-Interpolation verwendet. Die Bewertung erfolgte hinsichtlich der Stichprobengröße (B_1), der Maßzahlen der deskriptiven Statistik (B_2), der Verteilungsform der zu untersuchenden Variable Grundstückspreis (B_3), der Punktverteilung (B_4) und Autokorrelation (B_5).

	B_1	B_2	B_3	B_4	B_5	$B_1 \cdot B_2 \cdot B_3 \cdot B_4 = \mathbf{B}$
AICH	1	4	3	4	3	$1 \cdot 4 \cdot 3 \cdot 4 \cdot 3 = \mathbf{144}$
HAUS	4	5	1	2	2	$4 \cdot 5 \cdot 1 \cdot 2 \cdot 2 = 80$
RAMSAU	5	3	5	1	5	$5 \cdot 3 \cdot 5 \cdot 1 \cdot 5 = \mathbf{375}$
ROHRMOOS	2	1	2	3	4	$2 \cdot 1 \cdot 2 \cdot 3 \cdot 4 = 48$
SCHLADMING	3	2	4	5	1	$3 \cdot 2 \cdot 4 \cdot 5 \cdot 1 = \mathbf{120}$
GRÖBMING	4	3	1	2	3	$4 \cdot 3 \cdot 1 \cdot 2 \cdot 3 = \mathbf{72}$
MITTERBERG	3	2	3	1	4	$3 \cdot 2 \cdot 3 \cdot 1 \cdot 4 = \mathbf{72}$
ÖBLARN	2	4	4	3	1	$2 \cdot 4 \cdot 4 \cdot 3 \cdot 1 = \mathbf{96}$
ST. MARTIN	1	1	2	4	2	$1 \cdot 1 \cdot 2 \cdot 4 \cdot 2 = 16$

Tab. 5.7: Bewertung der Gemeinden für die Kriging-Interpolation

5.3.1 Deterministische Interpolation

Für die Datensätze wurden jeweils die globale Polynominterpolation (GPI), die lokale Polynominterpolation (LPI), Inverse Distance Weighted (IDW) und die Spline Interpolation (CRS) mithilfe des Geostatistical Analyst in ArcGIS® durchgeführt. GPI verwendet als globale Methode den kompletten Datensatz, LPI, IDW und die Spline Interpolation als lokale Methoden die definierte Nachbarschaftsumgebung, deren Datenpunkte in die Berechnung eingehen. Zur Verfügung stehen die nächsten Nachbarn, Quadranten und Oktanten Suchmethode in Form eines Kreises bzw. einer Ellipse mit Hilfe derer Anisotropie berücksichtigt werden kann. LPI, IDW und die Spline Interpolation sind sensitiv gegenüber der Nachbarschaftsdefinition.

Die Nachbarschaftsumgebung – das sind die Maximal- und Minimalanzahl der zu verwendenden Datenpunkte, sowie die Suchrichtung und Umgebung wurde durch so gewählt, dass der resultierende RMSPE minimiert wird. Dies konnte nur durch manuelles „Versuchen und Probieren“ bewerkstelligt werden. IDW und die Spline Interpolation reagierten dabei am empfindlichsten gegenüber der Form und Orientierung der Nachbarschaftsumgebung (Kreis, Ellipse, Ausrichtung) und weniger auf die Anzahl der Datenpunkte, während sich die LPI am empfindlichsten gegenüber der Anzahl der Datenpunkte erwies.

Die optimalen Modellparameter „power“ (vgl. Gl. 4.7 – Exponent) bei IDW und Kernparameter bei der Spline Interpolation können in ArcGIS® über eine entsprechende Optimierungsfunktion berechnet werden und wurden unverändert übernommen. Grundlage ist auch hier die Minimierung des RMSPE. Für verschiedene Modellparameter werden automatisch mithilfe der Kreuzvalidierung die entsprechenden RMSPE berechnet und diejenigen als optimal identifiziert, welchen den geringsten RMSPE ergeben.

Der Vergleich der Interpolationen erfolgt ebenfalls anhand des RMSPE Wertes aus der Kreuzvalidierung, die Ergebnisse sind in Tab. 5.8 angeführt. Vorab sei noch erwähnt, dass für die GPI aller Daten bzw. die beiden Gerichtsbezirke getrennt nur geringfügige Unterschiede hinsichtlich der verwendeten Ordnung im RMSPE auftreten. Für den kompletten Datensatz und den GB Gröbming ergeben die GPI der Ordnung ein bis fünf, für den GB Schladming der Ordnung eins bis drei praktisch gleiche RMSPE. Hinsichtlich der Spline Interpolation ergibt die Basisfunktion ‘completely regularized spline’ (CRS) das beste Ergebnis, der Unterschied zu ‘spline with tension’ ist allerdings gering.

In sieben von neun Fällen, insbesondere für die kompletten Gerichtsbezirke bzw. allen Datenpunkten erweist sich unter den deterministischen Interpolationen die Spline Interpolation mit CRS in Hinblick auf den Schätzfehler am besten. In den zwei anderen Fällen ist sie einmal knapp gefolgt die zweit beste Methode, im anderen Fall

ergibt sie das schlechteste Ergebnis, wobei hier alle Ergebnisse am engsten beieinander liegen. In diesem Fall ist der Grund offensichtlich ein räumlich isolierter Datenpunkt, der von allen anderen weit entfernt liegt.

	GPI		LPI		IDW		CRS
	Ordnung	RMSPE	Ordnung	RMSPE	power	RMSPE	RMSPE
Aich	2	23.78	2	23.12	1.18	20.62	19.91
Ramsau	2	50.69	2	50.11	1.06	42.61	41.41
Schladming	1	56.94	2	60.77	1	58.38	57.15
GB-Schladming	3	42.74	2	42.34	1	39.62	39.42
Gröbming	3	19.35	2	18.92	1.84	17.82	17.67
Mitterberg	2	18.08	2	16.96	1.03	17.16	16.42
Öblarn	1	12.90	2	12.53	1	13.27	13.58 ¹
GB-Gröbming	3	18.40	2	17.66	1.28	16.05	15.71
Alle Datenpunkte	4	31.06	1	29.74	1	27.97	27.81

Tab. 5.8: Vergleich der deterministischen Interpolationsverfahren
¹spline with tension

IDW liefert in den meisten Fällen das zweit beste Ergebnis. In Abbildung 5.8 ist das Untersuchungsgebiet mit den 18 Gemeinden und der Grundstückspreisoberfläche dargestellt. Die dick umrahmten Polygone kennzeichnen die aus Tab. 5.7 sechs ausgewählten Gemeinden (Ramsau, Schladming, Aich, Gröbming, Mitterberg und Öblarn von West nach Ost) für den Interpolationsvergleich.

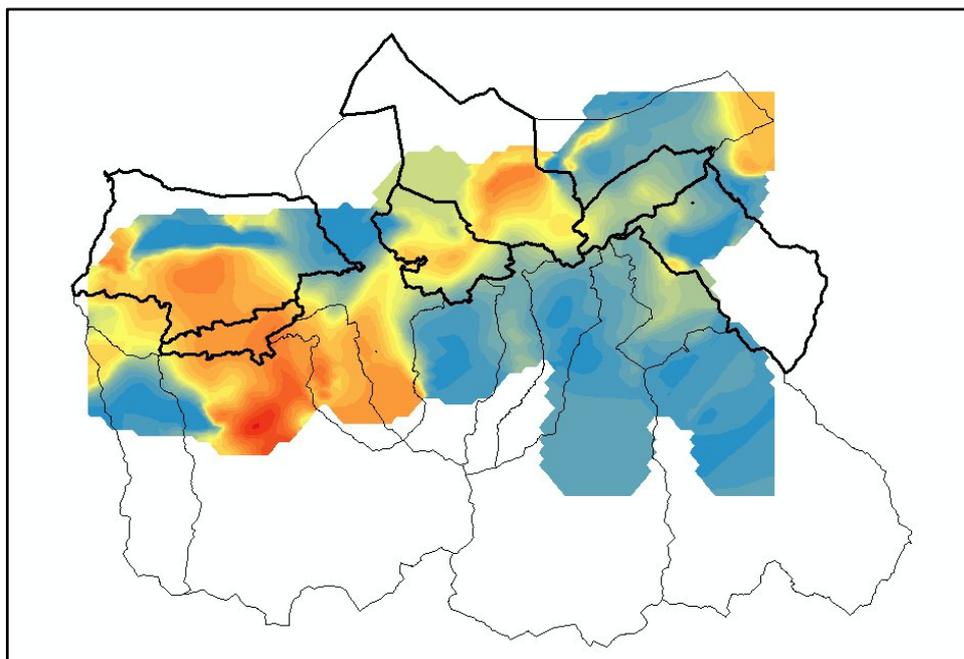


Abb. 5.9: Interpolationsoberfläche der Grundstückspreise

Im Anhang 17 sind beispielhaft im Bereich Gröbming Interpolationsoberflächen aus LPI, IDW und CRS gegenübergestellt.

5.3.2 Kriging Interpolation

Als Kriging Methoden wurden Ordinary Kriging und Universal Kriging verwendet, welche in ArcGIS® ebenfalls im Geostatistical Analyst integriert sind. Der Arbeitsablauf wird durch einen Assistenten gesteuert (Abb. 5.10), die erstellten Interpolationsmodelle werden als Datenebene in ArcMap® eingefügt und bleiben editierbar, was eine einfache und nachträgliche Änderung und Feinabstimmung der Kriging-Parameter ermöglicht. In der Literatur wird überwiegend für eine empirische Auswahl der Variogrammfunktion, basierend auf Referenzprojekten, plädiert. Da für die gegenständlichen Untersuchungen keine Vergleichsprojekte vorlagen, erfolgte die Auswahl der Variogrammfunktion und die Ermittlung der Variogrammparameter automationsunterstützt durch ArcGIS®, welches das optimale Variogrammmodell und dessen Parameter über Minimierung des RMSPE berechnet.

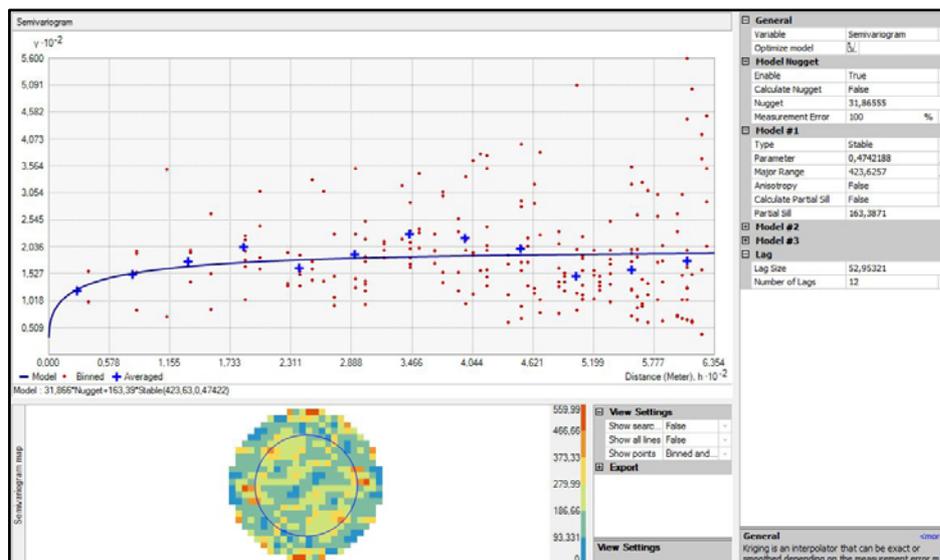


Abb. 5.10: Assistent des Geostatistical Analyst in ArcGIS®

Für die sechs Gemeinden, die zwei Gerichtsbezirke und den kompletten Datensatz wurden als Variogrammmodell durchwegs das robuste Modell (vgl., Gl. 3.47) gewählt, die (Semi-)Variogramme steigen im Ursprung steil an und erreichen schnell den Schwellwert oder zeigen einen reinen Nugget-Effekt. Dies bedeutet, dass Autokorrelation nur in der nahen Umgebung oder überhaupt nicht nachgewiesen werden kann. Keine große Autokorrelation zeigte bereits der im Rahmen der

explorativen Datenanalyse durchgeführte Test (Moran's I-Index, Kap. 5.2.1.4 bzw. Kap. 5.2.2.1), ein leicht positiver Wert ist jedoch vorhanden, sodass die Kriging Interpolation a priori nicht ausgeschlossen werden kann. Die fehlende Autokorrelation muss jedoch nicht darin begründet sein, dass die Verteilung der Grundstückspreise nicht grundsätzlich autokorreliert ist, sondern kann auch darin liegen, dass der Punktabstand zu groß ist, um eine räumliche Autokorrelation festzustellen.

Die Ergebnisse der verschiedenen OK und UK Interpolationsmodelle sind in Tab. 5.9 zusammengefasst. Da OK einen Trend bereinigten Datensatz benötigt, muss dieser zuvor eliminiert werden. Für die Berechnung stehen Polynome der Ordnung null (Konstante) bis drei zur Verfügung. Nach erfolgter OK Schätzung wird der berechnete Trend wieder angebracht. UK berücksichtigt den Trend in den Kriging Gleichungen, wobei wiederum dieselben Trendmodelle wie bei OK zur Verfügung stehen. Die Trendberücksichtigung kann mithilfe eines Kontrollparameters gesteuert werden, sodass diese mehr einer GPI oder LPI entspricht. Die Auswirkung im resultierenden RMSPE-Wert ist eher gering, doch interessanterweise hat sich gezeigt, dass sowohl für OK als auch UK die GPI geringfügig bessere Ergebnisse aufweisen. Bedeutend wichtiger ist die Berücksichtigung einer Anisotropie, die sehr wohl einen markanten Einfluss auf den berechneten RMSPE-Wert ausübt. Bezüglich der Größe des Suchbereiches können die in Kap.5.2.2.3.5 getroffenen Aussagen bestätigt werden. Die optimalen Werte der verwendeten Maximal- bzw. Minimalpunktanzahl wurden durch „Probieren“ ermittelt, die Minimalanzahl liegt zwischen vier für eine eher kleine und acht für eine größere Punktanzahl. Die Anzahl der Sektoren spielt eine untergeordnete Rolle (vgl. Kap. 5.2.2.3.5). Allgemein lässt sich eher festhalten, dass komplexere Modelle (z. B. hochgradige LPI bei der Trendbereinigung, Berechnung von zusätzlichen Parametern, erhöhte Sektorenanzahl) numerisch zu schlechteren RMSPE-Werten führen.

Trend	OK [RMSPE]			UK [RMSPE]		
	konstant	linear	quadratisch	konstant	linear	quadratisch
Aich	18.00	19.44	20.67	17.83	20.04	n.b.
Ramsau	41.87	42.18	43.31	41.35	44.72	43.20
Schladming	55.68	56.85	58.50	55.14	57.17	59.79
GB-Schladming	40.32	40.01	40.22	39.59	39.55	42.59
Gröbming	17.86	17.84	17.89	17.65	17.58	19.18
Mitterberg	16.46	16.68	16.34	16.59	17.21	16.36
Öblarn	12.96	12.43	n.b.	13.08	13.79	n.b.
GB-Gröbming	16.01	15.75	16.04	15.69	15.70	17.63
Alle Datenpunkte	28.09	28.02	28.07	28.03	27.82	30.14

Tab. 5.9: Vergleich der Kriging Interpolationsverfahren
n.b. nicht berechnet

Für drei mit "n.b." bezeichneten Gemeinden wurde die Interpolationsoberfläche mit quadratischer Trendberücksichtigung nicht berechnet, da sich kein optimales Modell berechnen lässt und die manuellen Berechnungen weit schlechtere Resultate ergeben. Infolge eines einzelnen räumlich isolierten und relativ weit entfernten Datenpunktes kann offensichtlich keine gute und sinnvolle quadratische Trendbereinigung berechnet werden. Als Vergleich werden ohnehin die Interpolationsmodelle mit linearer Trendberücksichtigung herangezogen, da dies den Gegebenheiten am meisten entspricht. Aus der Gegenüberstellung ist zu sehen, dass UK erst mit größerer Datenmenge (Gemeinde Gröbming mit 158, GB Schladming mit 275 und GB Gröbming mit 421 Datenpunkten) geringfügig bessere Ergebnisse in Hinblick auf den Schätzfehler liefert, ansonsten ergibt OK das bessere Ergebnis, wobei die Unterschiede teilweise marginal sind. Dies bestätigt die in Kap. 5.2.2.3.4 aufgestellte Behauptung, dass OK in den meisten Fällen ausreichend ist.

Als wichtig hat sich die Optimierung des Trendmodells herausgestellt, um möglichst kleine RMSPE Werte zu erhalten. Die Verwendung von lokalen Polynomen für die Trendbereinigung ergibt deutlich kleinere RMSPE Werte

Tabelle stellt 5.10 stellt schließlich die besten Ergebnisse der deterministischen und Kriging Interpolation gegenüber. Daraus ist ersichtlich, dass für die vorliegende Untersuchung der Grundstückspreise im GB Schladming und GB Gröbming beide Methoden gleichwertig sind.

	DETERMINISTISCH [RMSPE]	KRIGING [RMSPE]
Aich	19.91	19.44
Ramsau	41.41	42.43
Schladming	56.94	56.85
GB-Schladming	39.42	39.55
Gröbming	17.67	17.81
Mitterberg	16.42	16.85
Öblarn	12.53	12.43
GB-Gröbmimg	15.71	15.70
Alle Datenpunkte	27.81	27.82

Tab. 5.10: Vergleich beste deterministische Interpolation - Kriging

Im Anhang 18 sind für denselben Bereich in Gröbming die Interpolationsoberflächen aus OK und UK dargestellt. Eine Verbesserung des Kriging Ergebnisses ließe sich nur durch zusätzliche Informationen, die als weitere Datenebene in ein Co-Kriging Modell eingehen, erzielen.

6 SCHLUSSBEMERKUNGEN

6.1 Zusammenfassung der verwendeten Methoden

Grundstückspreise und deren Entwicklung sind im privaten als auch öffentlichen Bereich von großem Interesse, aber in Österreich wenig transparent. Im Wesentlichen wird weitgehend noch auf Schätzungen von Experten oder hedonischen Preiskonzepten zurückgegriffen. Diese Arbeit behandelt und vergleicht verschiedene Interpolationsmethoden auf Grundlage tatsächlich getätigter Grundstücksverkäufe, mithilfe derer in weiterer Folge Bodenpreiskarten abgeleitet werden können.

Die Ausgangsdaten der Grundstückspreise liegen im Zeitraum von 1995 bis 2009 auf Parzellenniveau vor. Da nur eine sehr beschränkte Anzahl von tatsächlich verkauften Grundstücken als Stichprobe zur Verfügung steht, müssen von den bekannten Datenpunkten ausgehend, die unbekanntes Preiswerte der dazwischen liegenden Parzellen mithilfe eines geeigneten Interpolationsverfahrens berechnet werden. Die Anzahl der innerhalb eines Jahres getätigten Grundstücksverkäufe als Interpolationsstützpunkte reicht bei den Beispielgemeinden nicht aus, weshalb die Verkäufe mehrerer Jahre zusammengefasst werden mussten. Dazu ist die Hochrechnung der älteren Preise mit einem Preisindex auf das Preisniveau des zu interpolierenden Jahres 2009 erforderlich. Für diese Valorisierung wurde nicht der Verbraucherpreisindex verwendet, sondern ein für diesen Datensatz berechneter Bodenpreisindex (PÖCHINGER, 2009).

Nach Detektion und Elimination von fehlerhaften Datenwerten bzw. Ausreißern und einer nicht räumlichen statistischen Analyse der Verkaufspreise für Baulandgrundstücke erfolgte die Verknüpfung dieser mit dem digitalen Kataster (DKM). Ziel ist die Berechnung einer bestmöglichen Schätzoberfläche auf dieser minimalen unabdingbaren Datengrundlage. Für die Schätzung der Grundstückspreise pro Quadratmeter stehen prinzipiell eine Reihe von Interpolationsverfahren zu Verfügung, die tatsächlichen Möglichkeiten sind aber durch die verwendete GIS-Software (ArcGIS® 10 Desktop) beschränkt. Neben den eher einfachen Verfahren, wie Polynom Regression (GPI, LPI) und IDW stehen auch die vergleichsweise komplexeren Methoden SPLINE und KRIGING zur Verfügung. Im Unterschied zu den deterministischen Verfahren (GPI, LPI, IDW und SPLINE) werden bei dem stochastischen Interpolationsverfahren KRIGING die Gewichtungsfaktoren nicht allein aufgrund der Lage der Messpunkte im Raum bestimmt, sondern ein, aus den Eigenschaften der Variablen in ihrer Anordnung im Raum abgeleiteter, stochastischer Anteil hinzugenommen wird. Die Basis für das eigentliche Krigingverfahren bildet der Betrag der räumlichen Varianzänderung zwischen den jeweiligen einzelnen Werten.

6.2 Zusammenfassung der Ergebnisse

Das vorhandene Punktdatennetz ist unregelmäßig und weist größtenteils eine gruppierte Struktur auf. Dies sind für die nachfolgende Interpolation nicht die idealen Voraussetzungen. Die exakten und lokalen Interpolationsmethoden ergeben erwartungsgemäß die besseren Ergebnisse. Als Kriterium für den Vergleich der verschiedenen Interpolationsoberflächen wurde der RMSPE-Wert aus der Kreuzvalidierung herangezogen.

Unter den deterministischen Verfahren wird das beste Resultat in Hinblick auf den Schätzfehler durch die Spline Interpolation mittels CRS erreicht. Dieses Ergebnis bestätigt ähnliche Untersuchungen von DUMFARTH (1996) für die Bodenpreislandschaft der Stadt Salzburg.

Als geostatistische Modelle wurden Ordinary Kriging und Universal Kriging verwendet. Dabei hat sich die allgemeine Empfehlung bestätigt, dass man mit OK ausreichend genaue Ergebnisse erhält. Der Vergleich zwischen OK und UK zeigt, dass beide Verfahren nahe beieinander liegen und UK erst bei einer größeren Datenpunktzahl Vorteile, da eine größere Punktzahl notwendig ist, um mit der integrierten Trendberücksichtigung gleichwertige oder evtl. bessere Ergebnisse zu erzielen. Bezüglich der Suchumgebung konnten bereits bestehenden Studien von GAU (2010) auf die gegenständliche Anwendung bestätigt und übertragen werden. Es hat sich herausgestellt, dass die Kriging Interpolationsmodelle unter Berücksichtigung der wichtigen Parameter (z. B. Anisotropie) so einfach wie möglich gehalten werden können.

Obwohl wenig Autokorrelation festzustellen war, liefert die geostatistische Schätzung die gleich guten Ergebnisse wie die beste deterministische Interpolationsmethode. Die schwache Autokorrelation muss jedoch nicht darin begründet sein, dass die Verteilung der Grundstückspreise nicht grundsätzlich autokorreliert ist, sondern kann auch darin liegen, dass der Punktabstand zu groß ist, um eine räumliche Autokorrelation festzustellen. Die Hinzunahme von sekundär oder tertiär Information für ein erweitertes Co-Kriging war wegen fehlender Daten nicht möglich.

6.3 Ausblick

Die Untersuchungen haben gezeigt, dass die geostatistischen Interpolationsmethoden nur auf Grundlage des Verkaufspreises pro Quadratmeter vergleichbar gute Ergebnisse wie die besten deterministischen Methoden ergeben. Unter Verwendung weiterer Informationen liegt ein großes Potenzial für die Erstellung von Grundstückspreisoberflächen mit Kriging Verfahren vor. Nützliche Informationen hierfür sind Angaben, die beeinflussende Merkmale des Bodenwertes berücksichtigen, wie Form und Höhenlage des Grundstückes, Bodenverhältnisse, Ver- und Entsorgungslei-

tungen. Zu diesen individuellen Merkmalen können noch die Eigenschaften der Mikrolage des Grundstückes, wie Wohnlage (unmittelbare Verbauung), Verkehrslage, Infrastrukturlage, Klimalage oder Lage in Gefahrenzonen hinzugezogen werden. Diese könnte durch hedonische Preisindizes, welche die qualitativen Unterschiede beschreiben berücksichtigt werden. Mit Hilfe von Co-Kriging ist es möglich, diese weiteren Informationen in das Modell zu integrieren und eine bessere Beschreibung der räumlichen Verteilung zu erhalten.

Neue Forschungen beschäftigen sich mit raumzeitlichen Interpolationsmethoden, da in zunehmenden Maße in GIS Anwendungen Daten, die Raum und Zeit miteinander verbinden, sog. raumzeitliche Daten (engl. *spatiotemporal data*) verwendet werden. Die gegenständliche Arbeit ist ebenfalls ein Beispiel dafür, wo für die Schaffung ausreichender Interpolationsstützpunkte die Grundstückspreise der verschiedenen Zeitepochen auf eine einheitliche Basis umgerechnet werden mussten. Infolge des zusätzlichen Zeitattributes sind die traditionellen räumlichen Interpolationsmethoden oft unzureichend für raumzeitliche Daten und daher wurden neue raumzeitliche Interpolationsmethoden entwickelt (LIXIN, 2009). Für raumzeitliche Interpolationen sind generell zwei Methoden möglich: **Reduktion** oder **Erweiterung**.

Die Reduktionsmethode behandelt die Zeit unabhängig von der räumlichen Dimension. Diese Methode reduziert das raumzeitliche Interpolationsproblem zu einer gewöhnlichen räumlichen Interpolation und umfasst zwei Stufen. Zuerst wird der gesuchte Wert über die Zeit für alle Datenpunkte interpoliert und dann eine räumliche Interpolation durchgeführt. Diese Methode wurde in der vorliegenden Arbeit angewendet.

Die Erweiterungsmethode behandelt die Zeit als äquivalent zur räumlichen Dimension. Diese Methode behandelt die Zeit als weitere Dimension im Raum und erweitert das raumzeitliche Interpolationsproblem in ein eindimensionales höheres räumliches Interpolationsproblem. Für den Fall eines zweidimensionalen Raumbereichs würde der Erweiterungsansatz eine dreidimensionale Interpolationsmethode für den Bereich von (x, y, t) benötigen.

Als räumliche Interpolationsmethode untersucht LIXIN (2009) zusätzlich zu IDW und Kriging die sog. **Form-Funktionen** (engl. *shape functions*) anhand von Testdaten aus tatsächlich verkauften Immobilien. Die Studien zeigen, dass die Erweiterungsmethode mit Form-Funktionen die genaueste raumzeitliche Interpolationsmethode in Hinblick auf dem RMSPE-Wert ist. Dabei ist zu beachten, dass es unter Umständen nicht leicht ist, zeitliche Einheiten mit räumlichen Einheiten zu vergleichen. Wichtig ist, dass Form-Funktion und Kriging basierende Interpolationsmodelle invariant gegenüber einer Koordinatenskalierung sind, während IDW dies nicht ist. Zuletzt wird noch gezeigt, dass sich die Aggregation von Daten in Zeitintervallen eine negative Auswirkung auf die Genauigkeit der Reduktionsmethode hat.

Anhand von Testdaten für Ozonwerte zeigt LIXIN (2009), dass die Erweiterungsmethode die genauesten Ergebnisse in Bezug auf den RMSPE-Wert ergeben. In diesem Fall allerdings erreicht Kriging das leicht bessere Ergebnis.

Bestehende GIS-Softwareprodukte haben üblicherweise noch nicht die Möglichkeit raumzeitliche Daten darzustellen, zu interpolieren und zu visualisieren. Raumzeitliche Interpolationen werden im Allgemeinen reduziert auf eine Sequenz von räumlichen Interpolationen. Diese Reduktionsmethode ist im Allgemeinen geeignet für Daten, die regelmäßig und hinreichend über den Zeitbereich gesammelt werden. Diese Erweiterungsmethode benötigt nicht die Annahme von genügenden Daten über den Zeitbereich. Sie funktioniert auch gut für unregelmäßig vorhandene Daten, sowohl im Raum als auch in der Zeit (LIXIN, 2009). Für die oben vorgestellten Studien wurde ein auf universitärer Ebene entwickeltes Datenbanksystem (MLPQ – Management of Linear Programming Queries) verwendet. Man darf gespannt sein, ob diese Entwicklungen auch in kommerziellen GIS-Softwareprodukten Einzug halten werden.

7 LITERATURVERZEICHNIS

- ARMSTRONG, M. (1998): Basic Linear Geostatistics. Springer: Berlin.
- BACKHAUS, K., B. ERICHSON, W. PLINKE, R. WEIBER (2008¹²): Multivariate Analysemethoden. Eine anwendungsorientierte Einführung. Springer: Berlin.
- BAHRENBURG, G., E. GIESE, J. NIPPER (1999⁴): Statistische Methoden in der Geographie . Band 1: Univariate und bivariate Statistik. Teubner: Stuttgart.
- BAHRENBURG, G., E. GIESE, N. MEVENKAMP, J. NIPPER (2008⁴): Statistische Methoden in der Geographie. Band 2: Multivariate Statistik. Gebr. Borntraeger: Berlin.
- BAILEY, T.C., A.C. GATRELL (1995): Interactive Spatial Data Analysis. Prentice Hall: Harlow.
- BARTELME, N. (2005⁴): Geoinformatik. Modelle · Strukturen · Funktionen. Springer: Berlin.
- BRONSTEIN, I. N., K. A. SEMENDJAJEW (1987²³): Taschenbuch der Mathematik. BSB Teubner: Leipzig.
- BÜHL, A. (2010¹²): SPSS 18. Einführung in die moderne Datenanalyse. Pearson Studium: München.
- BURROUGH, P. A., R. A. McDONELL (1998): Principles of Geographical Information Systems. Spatial Information Systems and Geostatistics. Oxford University Press. New York.
- CHILÈS, J. P., P. DELFINGER (1999): Geostatistics. Modeling Spatial Uncertainty. Wiley Series in Probability and Statistics. Wiley-Interscience: New York.
- CRESSIE, N. A. C. (1993): Statistics for spatial data. Wiley: New York.
- DAVIS, J.C. (2002³): Statistics and Data Analysis in Geology. Wiley: New York.
- DUMFARTH, E. (1996): Interpolation und Analyse von Bodenpreisoberflächen für die Stadt Salzburg. Diplomarbeit, Universität Salzburg

- FUCHS, C. (2011): Overview of Statistical Analysis of Spatial Data. Modeling Semivariograms. UC Santa Barbara, Department of Geography.
<http://www.geog.ucsb.edu/~chris/Lecture13_210C_Spring2011_ModelSemivariograms.pdf>
(Zugriff: 2011-03-06).
- GAU, C. (2010): Geostatistik in der Baugrundmodellierung. Die Bedeutung des Anwenders im Modellierungsprozess. Dissertation, Technische Universität Berlin.
- GERLACH, N. (2001): Ein Vergleich räumlicher Interpolationsverfahren für Windgeschwindigkeiten im komplexen Gelände. Diplomarbeit, Universität Münster.
- GRAMS, S. (2000): Einsatz geostatistischer Verfahren zur Charakterisierung der Grundwasserbeschaffenheit im Bereich der Rieselfelder südlich Berlins. Dissertation, Technische Universität Berlin.
- HUNSAKER, C.T., M.F. GOODCHILD, M.A. FRIEDL, T.J. CASE (2001): Spatial Uncertainty in Ecology. Implications for Remote Sensing and GIS Applications. Springer: New York.
- ISAAKS, E.H., R.M. SRIVASTAVA (1989): An Introduction to Applied Geostatistics. Oxford University Press: New York.
- JOHNSTON, K., J. M. VER HOEF, K. KRIVORUCHKO, L. NEIL (2003): Using ArcGIS Geostatistical Analyst. ESRI: Redlands
- KRIGE, D. G. (1951): A statistical approach to some basic mine valuation problems on the Witwatersrand". In: Journal of the Chemical, Metallurgical and Mining Society of South Africa, 52 (6), 119–139.
- LLOYD, C.D. (2010): Spatial Data Analysis. An Introduction for GIS Users. Oxford University Press: New York.
- LIXIN, L. (2009): Spatiotemporal Interpolation Methods in GIS. Exploring Data for Decision Making. VDM Verlag Dr. Müller: Saarbrücken.
- MATHERON, G. (1963): Principals of geostatistics. In: Economic Geology, Vol. 58, 1246-1266
- MATHERON, G. (1965): Les variables régionalisées et leur estimation: une application de la théorie des fonctions aléatoires aux sciences de la nature. Thèses a la Faculté des Sciences de l'université Paris, Masson et Cie, Paris.

- MCKILLUP, S., M. DARBY DYAR (2010): Geostatistics Explained. An Introductory Guide for Earth Scientists. Cambridge University Press: Cambridge.
- MITASOVA, H., L. MITAS (1993): Interpolation by Regularized Spline with Tension: I. Theory and Implementation. In: Mathematical Geology, Vol. 25, no. 6, 641-655
- MITASOVA, H., L. MITAS, W. M. BROWN, D.G. GERDES, I. KOSINOSKY, T. BAKER (1995): Modelling spatially and temporally distributed phenomena: new methods and tools for GRASS GIS. In: International Journal of Geographical Information Science, Vol. 9, Issue 4., 443-446
- PEARSON, E.S., M.A. STEPHENS (1964): The ratio of range to standard deviation in the same normal sample. In: Biometrika, Vol. 51, 484-487.
- PETERS, S. (2008): Ein Vergleich räumlicher Interpolationsverfahren für Ertragswerte im Weinbau. Diplomarbeit, Technische Universität München.
- PÖCHINGER, M. (2009): GIS-gestützte Analyse in der Immobilienbewertung. Masterarbeit, Technische Universität Graz.
- RAHMAN, M. M., GOVINDARAJULU, Z. (1997): A modification of the test of Shapiro and Wilk for normality. Journal of Applied Statistics 24, 219-236.
- ROYSTON, P. (1982): An extension of Shapiro and Wilk's W test for normality to large samples. In: Applied Statistics 31, 115-124
- RIVOIRARD, J. (1987): Two key parameters when choosing the kriging neighborhood. In: Mathematical Geology, Vol.19, no. 8, 851-856
- SACHS, L., J. HEDDERICH (2009¹³): Angewandte Statistik. Methodensammlung mit R. Springer: Berlin.
- SCHAUPPENLEHNER, T. (2008): Geostatistische Analysen zur Integration von Gelände-modellen und Bodenschätzungsdaten für verbesserte digitale Bodenkarten am Beispiel einer alluvialen Landschaft. Dissertation, Universität für Bodenkultur, Wien.

- SCHOENBERG, I. J. (1946): Contributions to the problem of approximation of equidistant data by analytic functions. In: Quaterly of Applied Mathematics, vol. 4, 45-99, 112-141
- SPICKERMANN, G. (2005): GIS-basierte Erfassung und Prognose der räumlichen Verteilung von Pflanzenkrankheiten. Dissertation, Universität Bonn.
- TOMCZAK, M. (1998): Spatial Interpolation and its Uncertainty Using Automatic Inverse Distance Weighted (IDW) – Cross Validation/Jackknife Approach. In: Journal of Geographic Information and Decision Analysis, vol. 2(2), 18-33.
- WACKERNAGEL, H. (2003³): Multivariate Geostatistics. An Introduction with Applications. Springer: Berlin.
- WALSH, J.E. (1951): Some nonparametric test of whether the largest observations of a set are too large or too small. In: Annals of Mathematical Statistics, vol. 21, number 4, 583-592
- WALSH, J.E. (1953): Correction to “Some nonparametric test of whether the largest observations of a set are too large or too small“. In: Annals of Mathematical Statistics, vol. 24, number 1, 134-135

8 ANHANG

Anhang 1:	Tabelle des kritischen Quotienten B/s auszugsweise aus PEARSON & STEPHENS (1964)	93
Anhang 2:	Deskriptive Statistik und Histogramme der valorisierten Grund- stückspreise auf Katastralgemeindeebene im GB Schladming und GB Gröbming	94
Anhang 3:	Deskriptive Statistik und Histogramme der valorisierten Grund- stückspreise auf Gemeindeebene im GB Schladming und GB Gröbming	98
Anhang 4:	Deskriptive Statistik und Histogramme der valorisierten Grund- stückspreise auf Gerichtsbezirksebene	100
Anhang 5:	Ausreißeranalyse mittels Box-Plot auf Katastralgemeindeebene, Gemeindeebene und Gerichtsbezirksebene	101
Anhang 6:	Test auf Normalverteilung QQ-Diagramme GB Schladming und GB Gröbming	102
Anhang 7:	Test auf Normalverteilung PP-Diagramme GB Schladming und GB Gröbming	104
Anhang 8:	Test auf Normalverteilung – GB Schladming und GB Gröbming Kolmogorov-Smirnov und Shapiro-Wilk Test	106
Anhang 9:	Test auf Autokorrelation mittels Crimestat® – GB Schladming und GB Gröbming	107
Anhang 10:	Deskriptive Statistik der valorisierten Grundstückspreise auf Gemeindeebene nach Fehlerbereinigung und Übernahme in ArcGIS	108
Anhang 11:	Deskriptive Statistik und Histogramm der valorisierten Grundstücks- preise auf Gerichtsbezirksebene nach Fehlerbereinigung und Über- nahme in ArcGIS	109
Anhang 12:	Test auf Normalverteilung – GB Schladming und GB Gröbming Nach Fehlerbereinigung und Übernahme in ArcGIS Kolmogorov-Smirnov und Shapiro-Wilk Test	110
Anhang 13:	Test auf Autokorrelation GB Schladming – Moran's I-Index nach Fehlerbereinigung und Übernahme in ArcGIS	111
Anhang 14:	Test auf Autokorrelation GB Gröbming – Moran's I-Index nach Fehlerbereinigung und Übernahme in ArcGIS	112

Anhang 15: Trendanalyse – GB Schladming	113
Anhang 16: Trendanalyse – GB Gröbming	114
Anhang 17: Deterministische Interpolationsoberflächen im Bereich Gröbming	115
Anhang 18: Kriging Interpolationsoberflächen im Bereich Gröbming	116

Anhang 1: Tabelle des kritischen Quotienten B/s auszugsweise aus PEARSON & STEPHENS (1964)

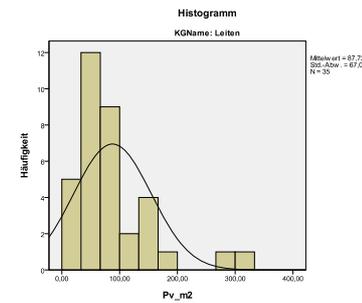
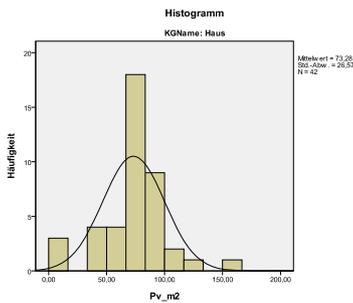
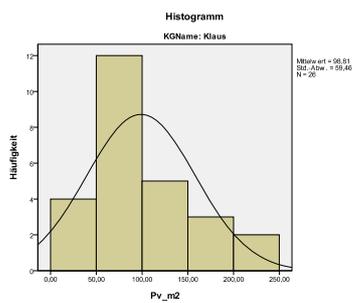
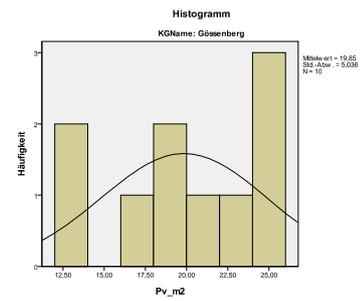
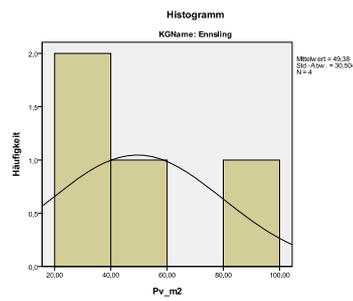
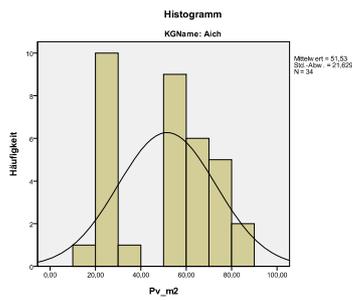
n	untere Quantile			obere Quantile		
	Signifikanz-Niveau			Signifikanz-Niveau		
	0,01	0,05	0,10	0,10	0,05	0,01
3	1.737	1.758	1.782	1.997	1.999	2.000
4	1.87	1.98	2.04	2.409	2.429	2.445
5	2.02	2.15	2.22	2.712	2.753	2.803
6	2.15	2.28	2.37	2.949	3.012	3.095
7	2.26	2.40	2.49	3.143	3.222	3.338
8	2.35	2.50	2.59	3.308	3.399	3.543
9	2.44	2.59	2.68	3.449	3.552	3.720
10	2.51	2.67	2.76	3.57	3.685	3.875
11	2.58	2.74	2.94	3.68	3.80	4.012
12	2.64	2.80	2.90	3.78	3.91	4.134
13	2.70	2.86	2.96	3.87	4.00	4.244
14	2.75	2.92	3.02	3.95	4.09	4.34
15	2.80	2.97	3.07	4.02	4.17	4.44
16	2.84	3.01	3.12	4.09	4.24	4.52
17	2.88	3.06	3.17	4.15	4.31	4.60
18	2.92	3.10	3.21	4.21	4.37	4.67
19	2.96	3.14	3.25	4.27	4.43	4.74
20	2.99	3.18	3.29	4.32	4.49	4.80
25	3.15	3.34	3.45	4.53	4.71	5.06
30	3.27	3.47	3.59	4.70	4.89	5.26
35	3.38	3.58	3.70	4.84	5.04	5.42
40	3.47	3.67	3.79	4.96	5.16	5.56
45	3.55	3.75	3.88	5.06	5.26	5.67
50	3.62	3.83	3.95	5.14	5.35	5.77
55	3.69	3.90	4.02	5.22	5.43	5.86
60	3.75	3.96	4.08	5.29	5.51	5.94
65	3.80	4.01	4.14	5.35	5.57	6.01
70	3.85	4.06	4.19	5.41	5.63	6.07
75	3.90	4.11	4.24	5.46	5.68	6.13
80	3.94	4.16	4.28	5.51	5.73	6.18
85	3.99	4.20	4.33	5.56	5.78	6.23
90	4.02	4.24	4.36	5.60	5.82	6.27
95	4.06	4.27	4.40	5.64	5.86	6.32
100	4.10	4.31	4.40	5.68	5.90	6.36
150	4.38	4.59	4.72	5.96	6.18	6.64
200	4.59	4.78	4.90	6.15	6.39	6.84
500	5.13	5.37	5.49	6.72	6.94	7.42
1000	5.57	5.79	5.92	7.11	7.33	7.80

Anhang 2: Deskriptive Statistik und Histogramme der valorisierten Grundstückspreise auf Katastralgemeindeebene im Gerichtsbezirk Schladming

Deskriptive Statistik

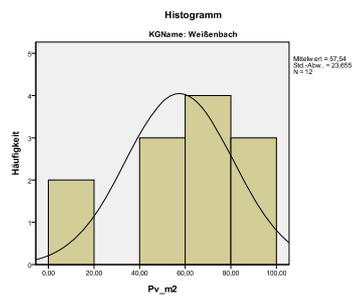
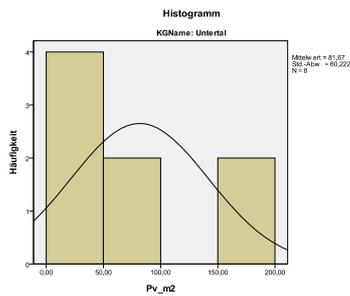
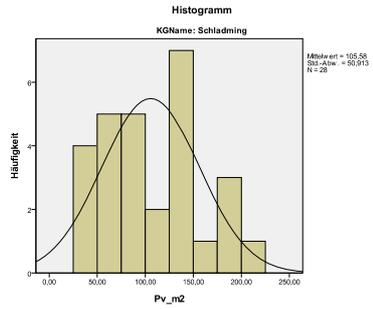
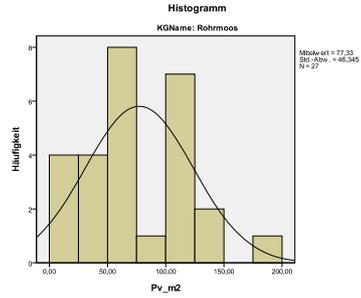
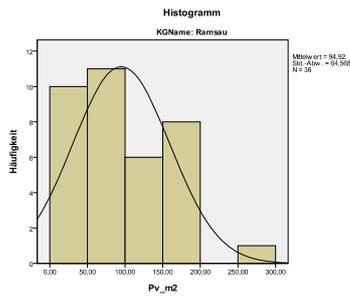
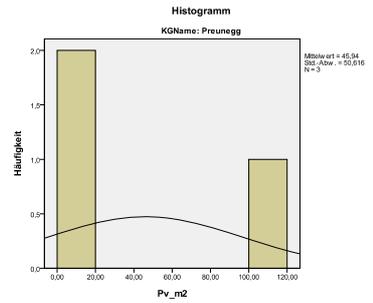
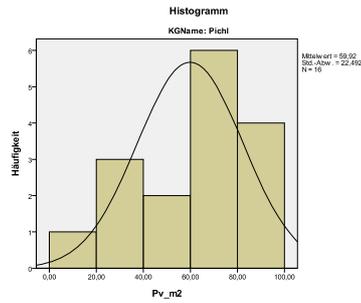
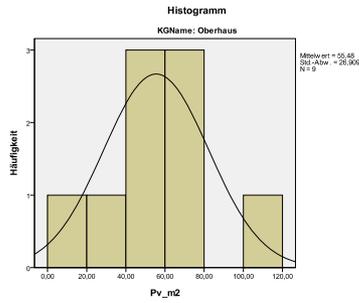
Katastralgemeinde		N	Min	Max	\bar{m}	s	Schiefe	Kurtosis
Aich	Pv_m2	34	17,41	88,31	51,53	21,63	-0,133	-1,290
Ennsling ¹	Pv_m2	4	27,23	93,64	49,38	30,50	1,644	2,620
Gössenberg	Pv_m2	10	12,18	25,62	19,85	5,04	-0,393	-1,092
Haus ¹	Pv_m2	42	8,43	151,70	73,28	26,57	-0,137	2,136
Klaus ⁴	Pv_m2	26	14,99	246,86	98,81	59,46	0,899	0,285
Leiten ²	Pv_m2	35	9,08	313,40	87,72	67,00	1,782	3,918
Oberhaus	Pv_m2	9	16,44	105,96	55,48	26,91	0,292	0,592
Pichl	Pv_m2	16	14,96	88,31	59,92	22,49	-0,703	-0,752
Preunegg	Pv_m2	3	14,30	104,31	45,94	50,62	1,714	
Ramsau ²	Pv_m2	36	8,33	294,18	94,92	64,57	0,797	0,856
Rohrmoos ³	Pv_m2	27	7,92	192,22	77,33	46,34	0,384	-0,208
Schladming ⁴	Pv_m2	28	27,23	200,00	105,58	50,91	0,282	-1,066
Untertal ³	Pv_m2	8	30,41	184,50	81,67	60,22	1,086	-0,434
Weißbach ¹	Pv_m2	12	16,85	88,25	57,54	23,66	-0,560	0,637

Histogramme



Anhang 2: Deskriptive Statistik und Histogramme der valorisierten Grundstückspreise auf Katastralgemeindeebene im Gerichtsbezirk Schladming

Histogramme

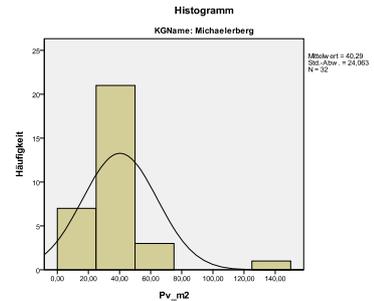
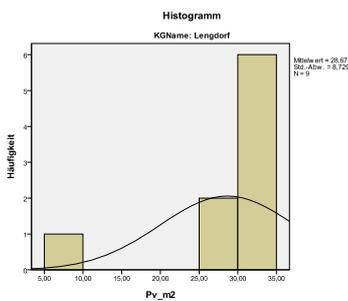
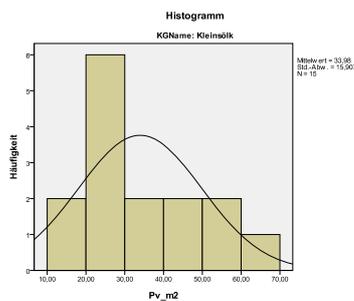
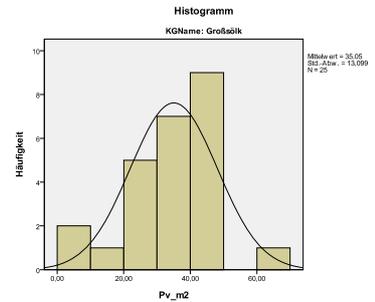
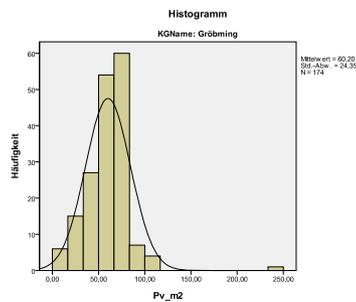
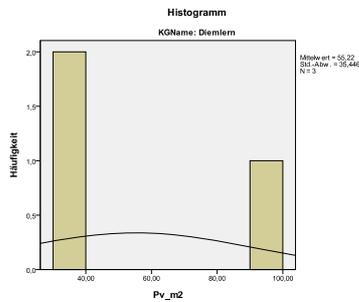


Anhang 2: Deskriptive Statistik und Histogramme der valorisierten Grundstückspreise auf Katastralgemeindeebene im Gerichtsbezirk Gröbming

Deskriptive Statistik

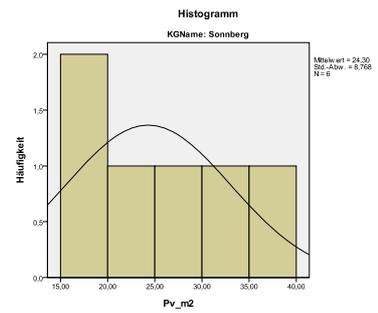
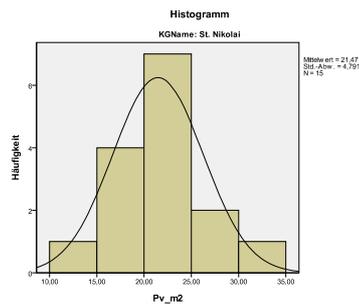
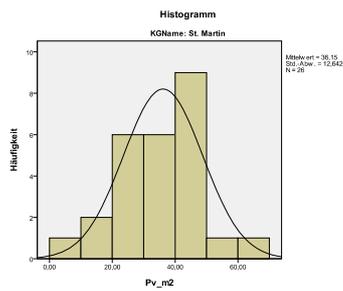
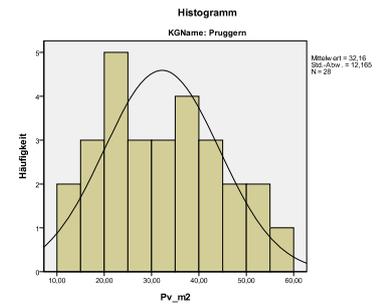
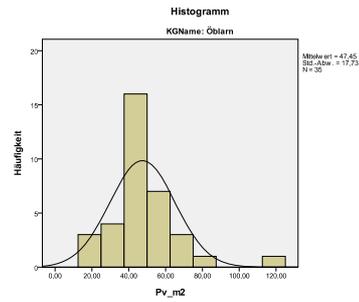
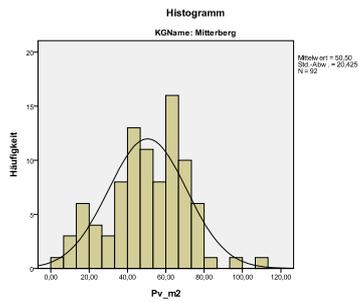
Katastralgemeinde		N	Min	Max	\bar{m}	s	Schiefe	Kurtosis
Diemlern ⁵	Pv_m2	3	32,57	96,07	55,22	35,45	1,70	
Gröbming	Pv_m2	174	5,52	238,10	60,20	24,35	2,059	15,778
Großsölk	Pv_m2	25	9,31	67,70	35,05	13,10	-0,117	0,934
Kleinsölk	Pv_m2	15	10,42	66,63	33,98	15,91	0,511	-0,166
Lengdorf ⁵	Pv_m2	9	5,62	33,38	28,67	8,73	-2,890	8,512
Michaelberg	Pv_m2	32	7,73	145,71	40,29	24,06	2,533	11,703
Mitterberg	Pv_m2	92	6,36	108,91	50,50	20,43	-0,161	0,002
Öblarn ⁶	Pv_m2	35	19,95	121,14	47,45	12,73	2,145	8,064
Pruggern	Pv_m2	28	12,76	55,61	32,16	12,16	0,281	-0,956
St. Martin ⁵	Pv_m2	26	9,28	65,94	36,16	12,64	0,013	0,210
St. Nikolai	Pv_m2	15	12,03	30,55	21,47	4,79	0,027	0,104
Sonnberg ⁶	Pv_m2	6	15,20	37,34	24,30	8,77	0,535	-1,210

Histogramme



Anhang 2: Deskriptive Statistik und Histogramme der valorisierten Grundstückspreise auf Katastralgemeindeebene im Gerichtsbezirk Gröbming

Histogramme

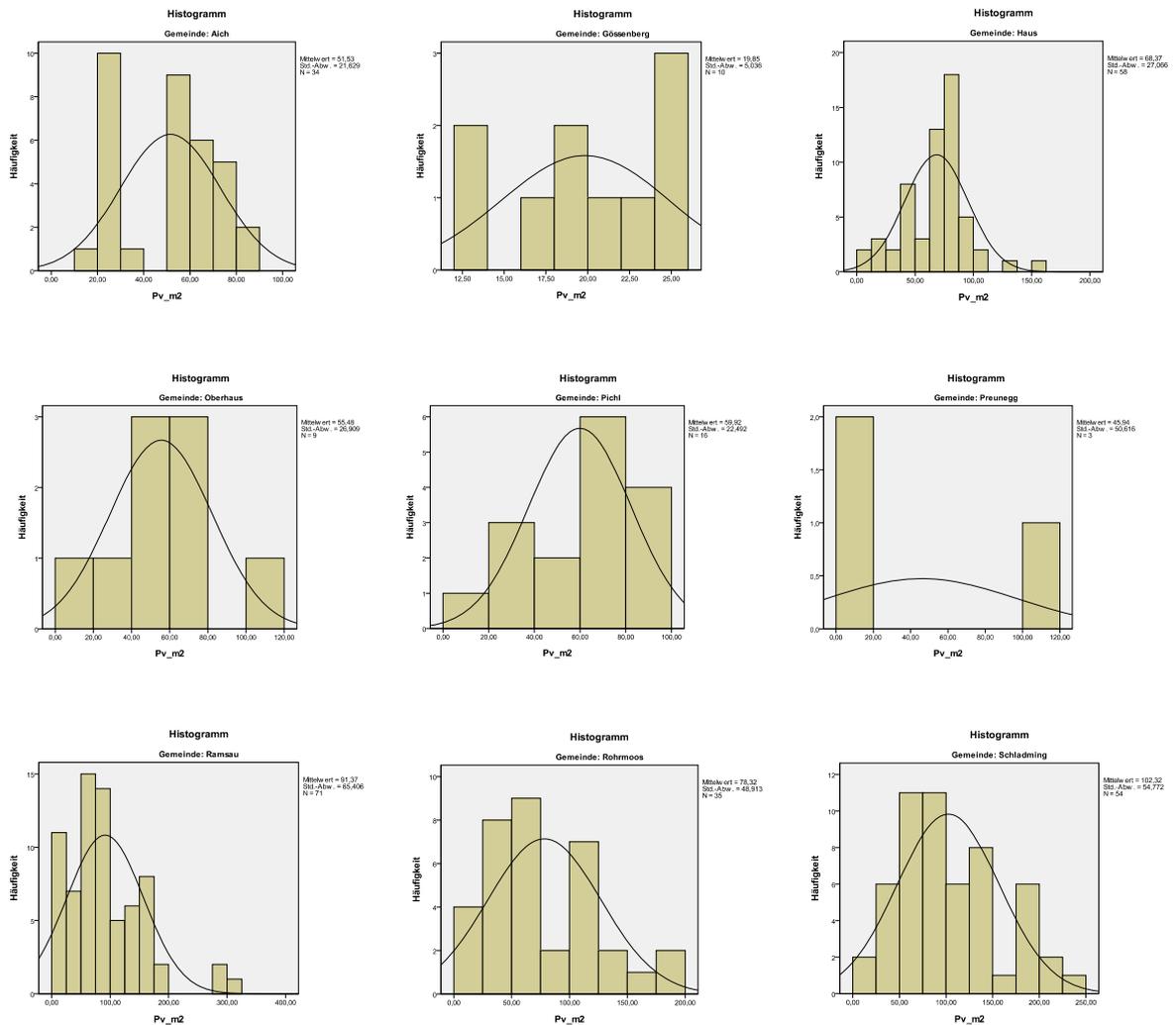


Anhang 3: Deskriptive Statistik und Histogramme der valorisierten Grundstückspreise auf Gemeindeebene im GB Schladming

Deskriptive Statistik

Gemeinde		N	Min	Max	\bar{m}	s	Schiefe	Kurtosis
Aich	Pv_m2	34	17,41	88,31	51,53	21,63	-,133	-1,290
Gössenberg	Pv_m2	10	12,18	25,62	19,85	5,04	-,393	-1,092
Haus	Pv_m2	58	8,43	151,70	68,37	27,07	-,063	1,070
Oberhaus	Pv_m2	9	16,44	105,96	55,48	26,91	,292	,592
Pichl	Pv_m2	16	14,96	88,31	59,92	22,49	-,703	-,752
Preunegg	Pv_m2	3	14,30	104,31	45,94	50,62	1,714	
Ramsau	Pv_m2	71	8,33	313,40	91,37	65,41	1,269	2,077
Rohrmoos	Pv_m2	35	7,92	192,22	78,32	48,91	,601	-,290
Schladming	Pv_m2	54	14,99	246,86	102,32	54,77	,602	-,359

Histogramme

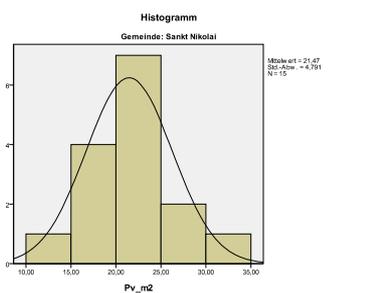
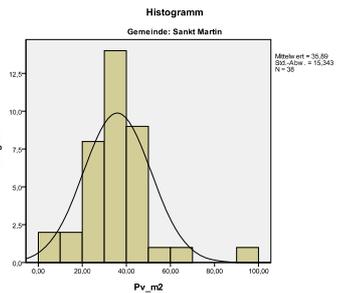
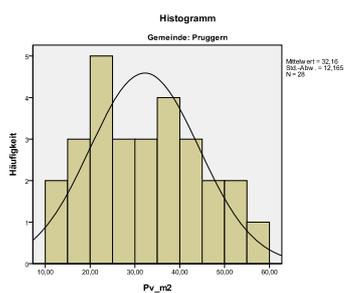
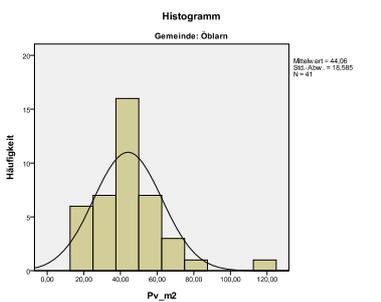
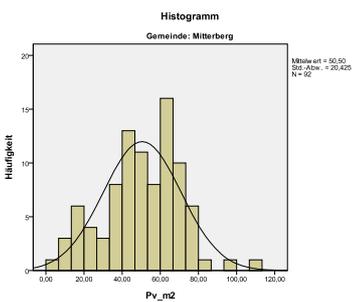
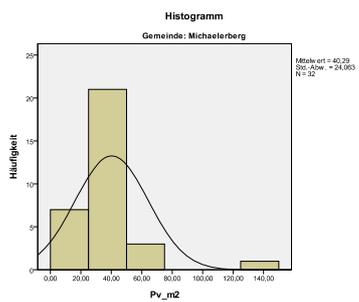
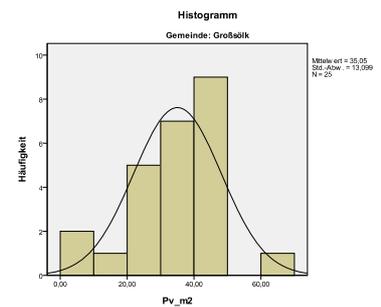
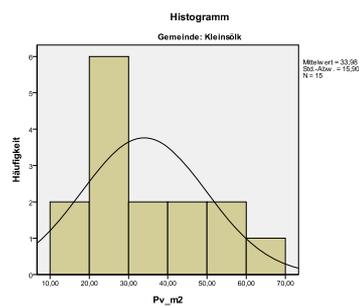
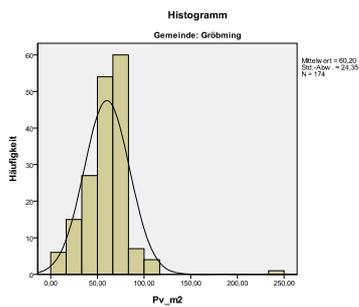


Anhang 3: Deskriptive Statistik und Histogramme der valorisierten Grundstückspreise auf Gemeindeebene im GB Gröbming

Deskriptive Statistik

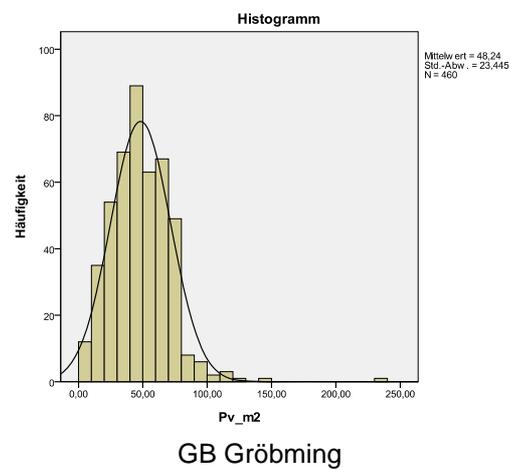
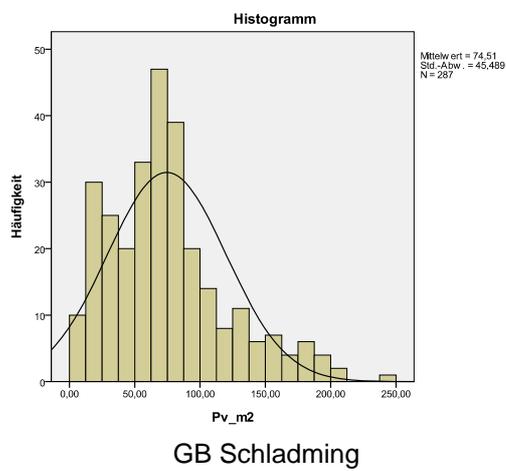
Gemeinde		N	Min	Max	\bar{m}	s	Schiefe	Kurtosis
Gröbming	Pv_m2	174	5,52	238,10	60,20	24,35	2,059	15,778
Großsölk	Pv_m2	25	9,31	67,70	35,05	13,10	-0,117	0,934
Kleinsölk	Pv_m2	15	10,42	66,63	33,98	15,91	0,511	-0,166
Michaelaberg	Pv_m2	32	7,73	145,71	40,29	24,06	2,533	11,703
Mitterberg	Pv_m2	92	6,36	108,91	50,50	20,43	-0,161	0,002
Öblarn	Pv_m2	41	15,20	121,14	44,06	18,59	1,742	6,415
Pruggern	Pv_m2	28	12,76	55,61	32,16	12,16	0,281	-0,956
St. Martin	Pv_m2	38	5,62	96,07	35,89	15,34	1,526	5,789
St. Nikolai	Pv_m2	15	12,03	30,55	21,47	4,79	0,027	0,104

Histogramme



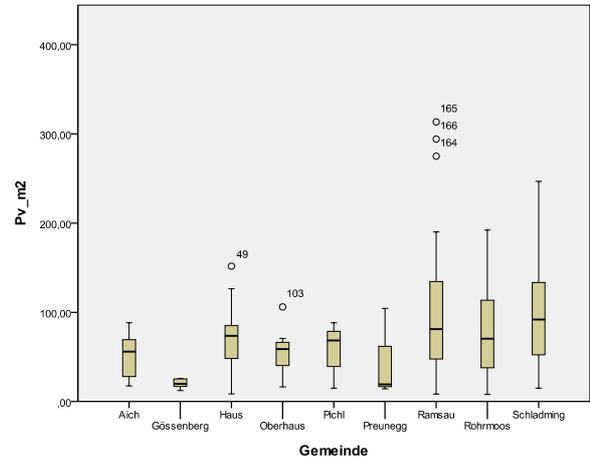
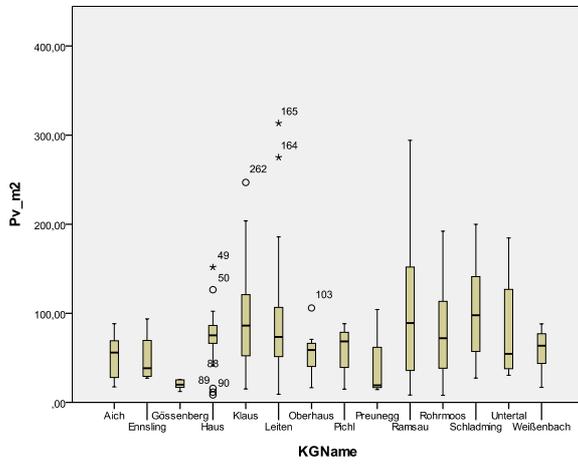
Anhang 4: Deskriptive Statistik und Histogramm der valorisierten Grundstückspreise auf Gerichtsbezirksebene

Pv_m2	N	Min	Max	\bar{m}	s	Schiefe	Kurtosis
GB Schladming	290	7,92	313,40	76,78	50,46	1,411	3,073
GB Gröbming	460	5,52	238,10	48,24	48,24	1,499	9,182

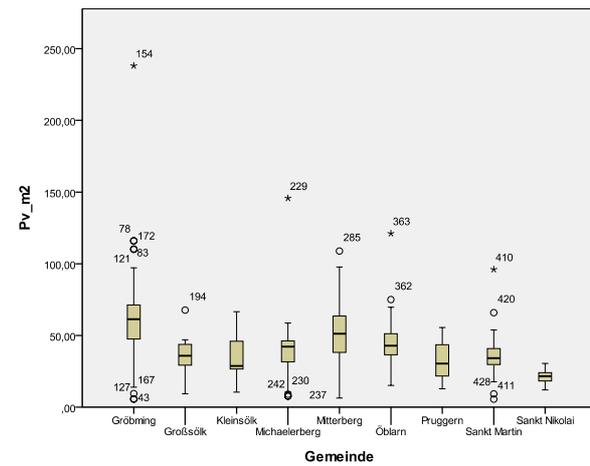
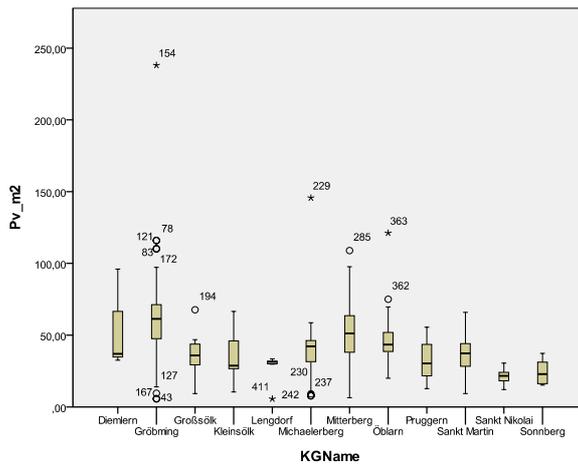


Anhang 5: Ausreißeranalyse mittels Box-Plot auf Katastralgemeindeebene, Gemeindeebene und Gerichtsbezirksebene

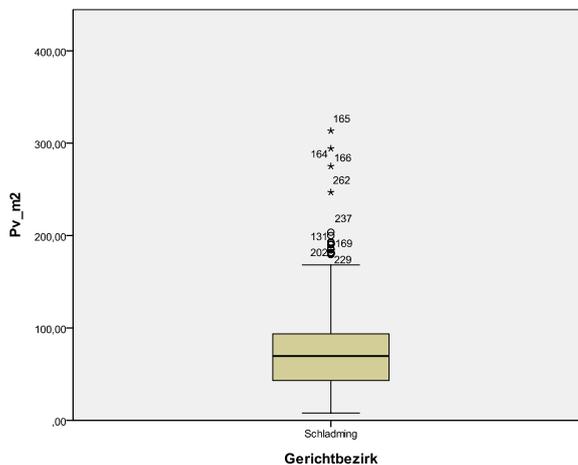
GB Schladming



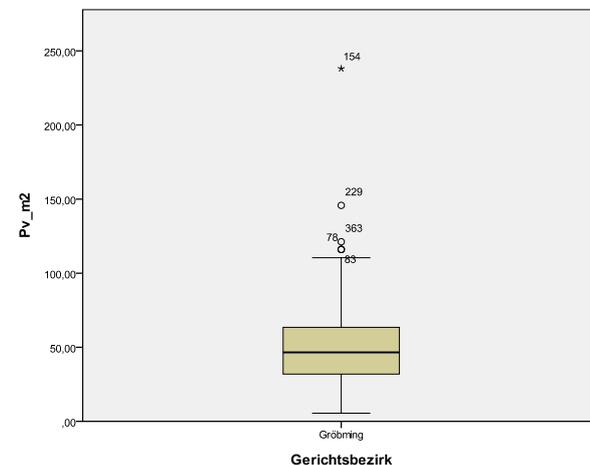
GB Gröbming



GB Schladming

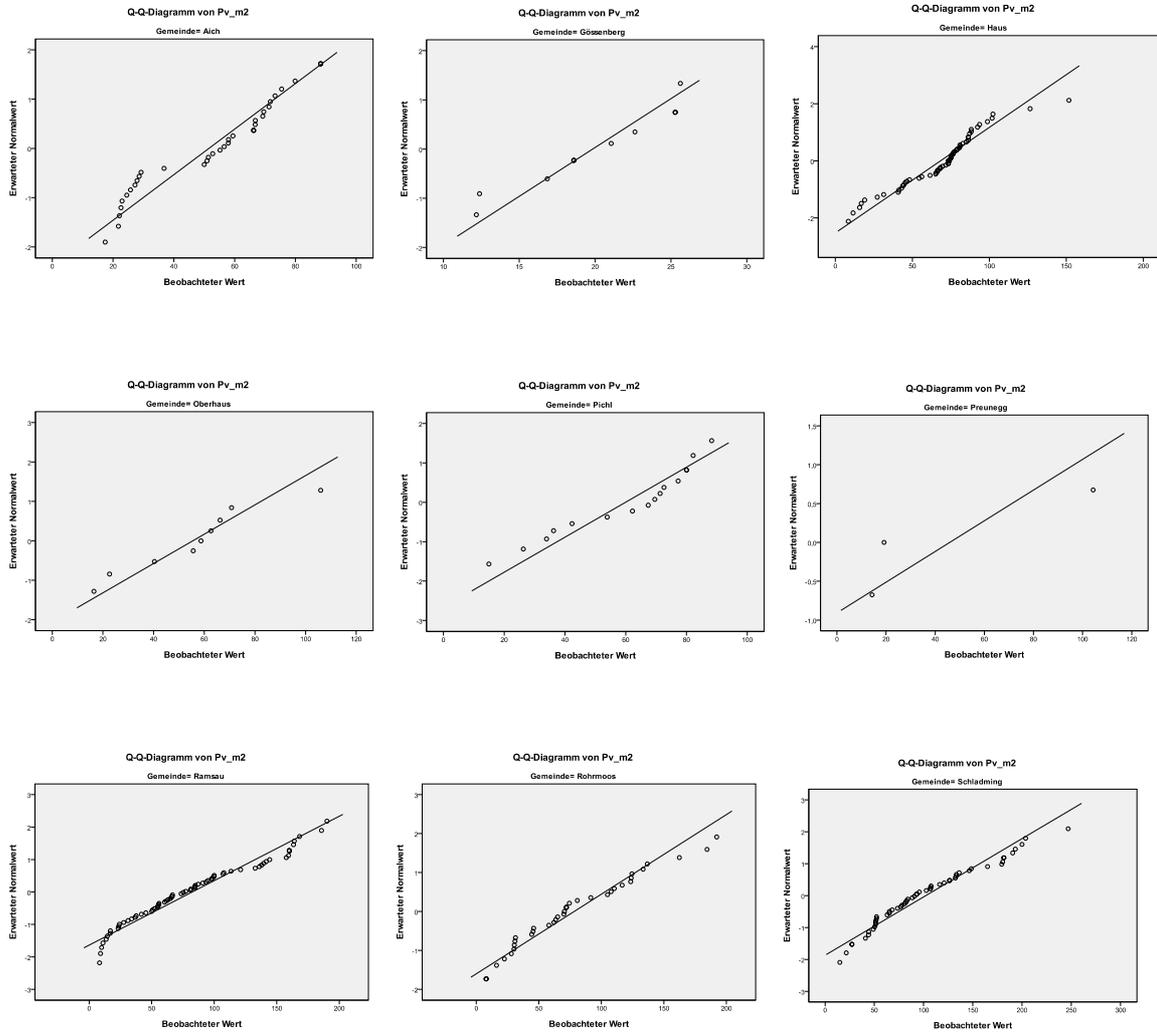


GB Gröbming

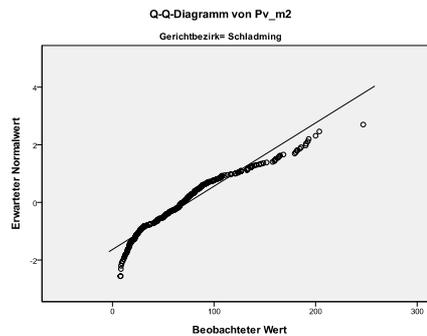


Anhang 6: Test auf Normalverteilung – QQ-Diagramme GB Schladming

QQ-Diagramme der Gemeinden

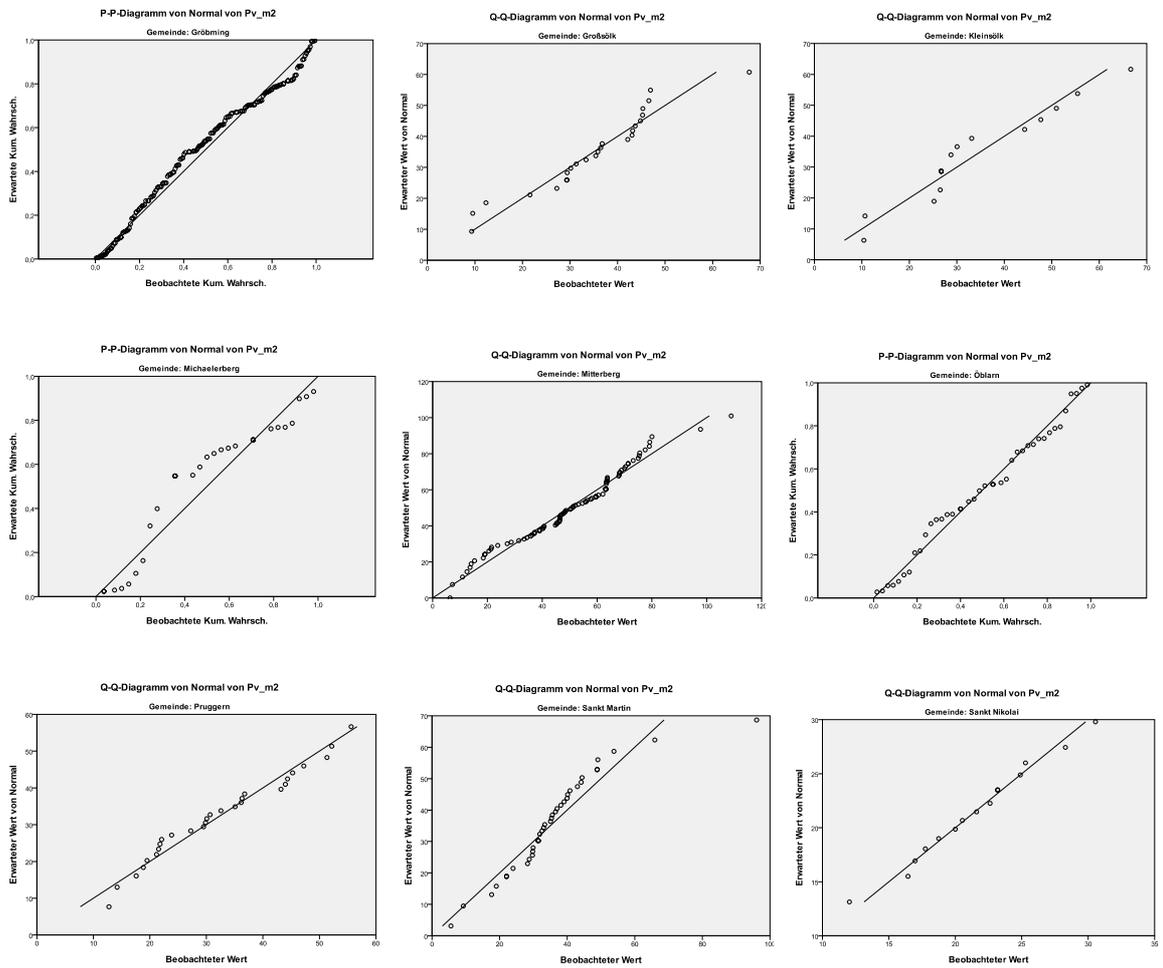


QQ-Diagramm des Gerichtsbezirks

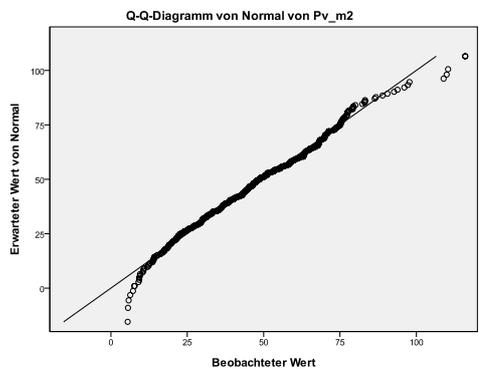


Anhang 6: Test auf Normalverteilung – QQ-Diagramme GB Gröbming

QQ-Diagramme der Gemeinden

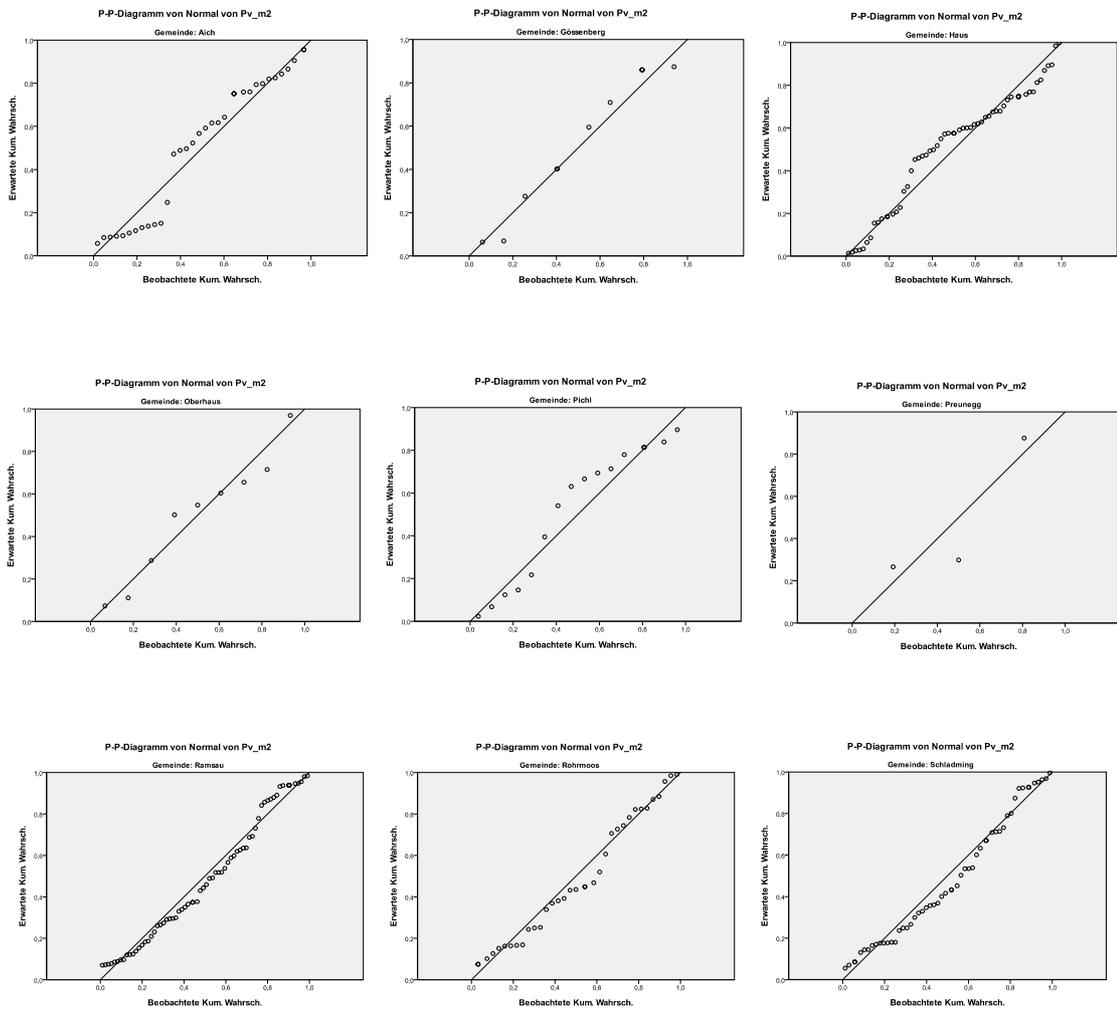


QQ-Diagramm des Gerichtsbezirks

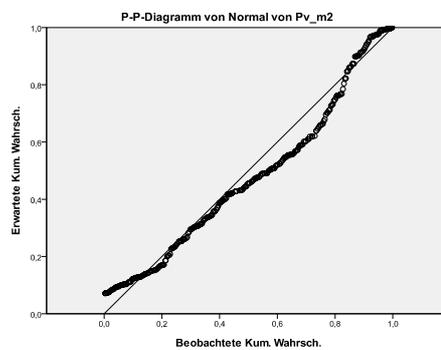


Anhang 7: Test auf Normalverteilung – PP-Diagramme GB Schladming

PP-Diagramme der Gemeinden

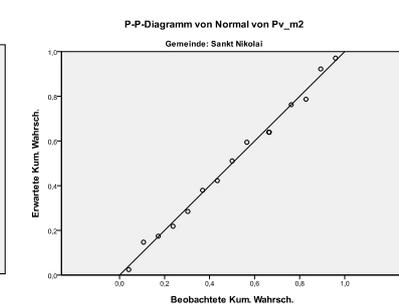
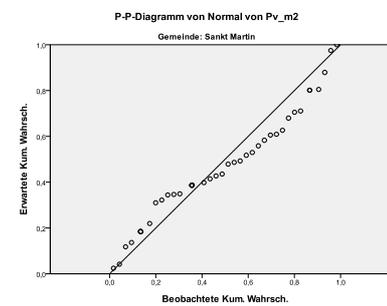
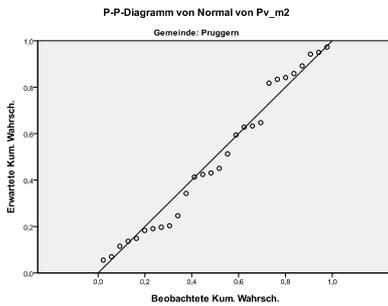
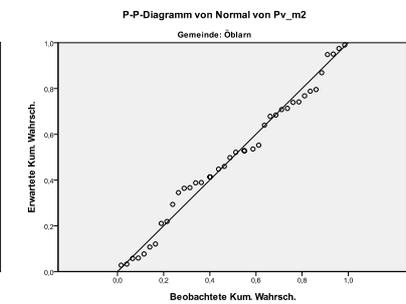
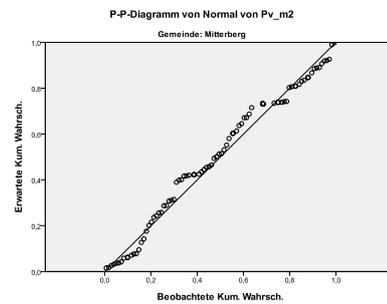
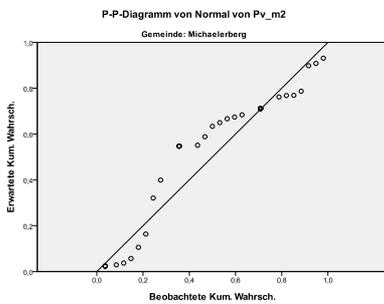
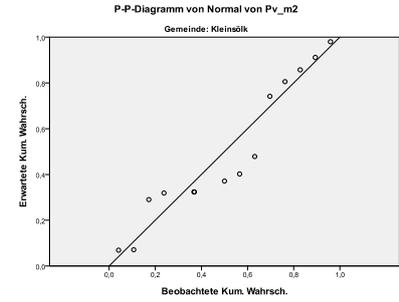
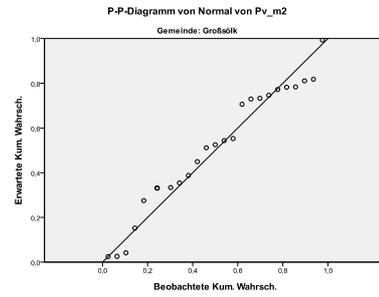
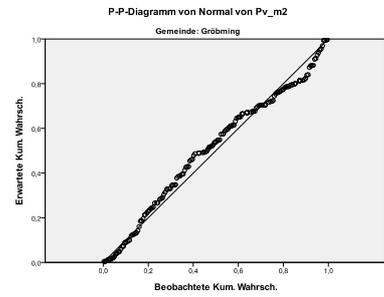


PP-Diagramm des Gerichtsbezirks

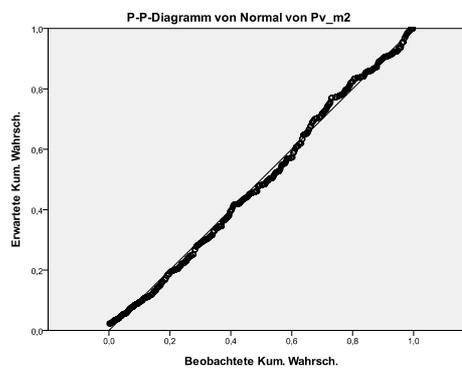


Anhang 7: Test auf Normalverteilung – PP-Diagramme GB Gröbming

PP-Diagramme der Gemeinden



PP-Diagramm des Gerichtsbezirks



Anhang 8: Test auf Normalverteilung – GB Schladming und GB Gröbming Kolmogorov-Smirnov und Shapiro-Wilk Test

Test auf Normalverteilung für die Katastralgemeinden im GB Schladming

Gemeinde	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Aich Pv_m2	,173	34	,012	,917	34	,014
Gössenberg Pv_m2	,160	10	,200*	,903	10	,237
Haus Pv_m2	,142	58	,005	,948	58	,014
Oberhaus Pv_m2	,174	9	,200*	,948	9	,671
Pichl Pv_m2	,193	16	,113	,910	16	,116
Preunegg Pv_m2	,368	3		,791	3	,092
Ramsau Pv_m2	,094	68	,200*	,949	68	,008
Rohrmoos Pv_m2	,132	35	,130	,948	35	,102
Schladming Pv_m2	,104	54	,200*	,953	54	,033

a. Signifikanzkorrektur nach Lilliefors

*. Dies ist eine untere Grenze der echten Signifikanz.

Test auf Normalverteilung für die Katastralgemeinden im GB Gröbming

Gemeinde	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Gröbming Pv_m2	,082	173	,006	,975	174	,004
Großsölk Pv_m2	,142	25	,200*	,937	25	,124
Kleinsölk Pv_m2	,198	15	,118	,925	15	,230
Miachealaberg Pv_m2	,257	31	,000	,871	31	,001
Mitterberg Pv_m2	,090	92	,063	,973	92	,056
Öblarn Pv_m2	,155	40	,200*	,977	40	,573
Pruggern Pv_m2	,118	28	,200*	,958	28	,320
St. Martin Pv_m2	,137	38	,071	,881	38	,001
St. Nikolai Pv_m2	,094	15	,200*	,991	15	1,000

a. Signifikanzkorrektur nach Lilliefors

*. Dies ist eine untere Grenze der echten Signifikanz.

Test auf Normalverteilung für den Gerichtsbezirk

Gerichtsbezirk	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
Schladming Pv_m2	,111	287	,000	,935	287	,000
Gröbming Pv_m2	,041	457	,061	,986	457	,000

a. Signifikanzkorrektur nach Lilliefors

Anhang 9: Test auf Autokorrelation – Moran's I-Index und Geary's C-Index
 Test der Punktverteilung mittels Nearest Neighbour Analysis (NNA)

Gemeinde	n	Moran's I	Geary's C	Moran's I Distanzkorrektur	Geary's C Distanzkorrektur	NNA Index
Aich	34	0.02	0.95	-0.03	1.01	0.60
Gössenberg	10	0.44	0.36	0.04	0.85	0.59
Haus	58	0.13	0.66	0.00	0.97	0.51
Oberhaus	9	-0.19	1.16	-0.11	0.96	0.89
Pichl	16	-0.19	1.20	-0.05	0.95	0.55
Preunegg	3	-0.36	0.86	-0.43	0.93	2.44
Ramsau	68	0.75	0.65	0.01	0.99	0.44
Rohrmoos	35	0.19	0.85	-0.03	1.00	0.66
Schladming	54	-0.07	1.06	-0.02	0.99	0.53
GB Schladming	287	0.22	0.74	-0.01	1.02	0.39

Gemeinde	n	Moran's I	Geary's C	Moran's I Distanzkorrektur	Geary's C Distanzkorrektur	NNA Index
Gröbming	173	0.04	0.85	0.00	0.99	0.42
Großsölk	25	-0.01	0.79	-0.04	1.00	0.18
Kleinsölk	15	0.13	0.51	0.03	0.93	1.03
Michaelaberg	31	0.23	0.52	0.01	0.93	0.97
Mitterberg	92	0.14	0.61	0.03	0.91	0.30
Öblarn	40	0.06	0.65	--0.01	0.94	0.53
Pruggern	28	-0.02	0.99	-0.03	1.01	0.81
St. Martin	38	0.04	0.48	-0.01	0.80	0.42
St. Nikolai	15	0.76	0.36	0.07	0.77	0.16
GB Gröbming	457	0.20	0.75	0.07	0.95	0.24

Anhang 10: Deskriptive Statistik der valorisierten Grundstückspreise auf Gemeindeebene nach Fehlerbereinigung und Übernahme in ArcGIS

GB Schladming

Gemeinde		N	Min	Max	\bar{m}	s	Schiefe	Kurtosis
Aich	Pv_m2	31	17,41	88,31	52,76	21,43	-,213	-1,163
Gössenberg	Pv_m2	10	12,18	25,62	19,85	5,04	-,394	-1,092
Haus	Pv_m2	57	8,43	151,70	67,95	27,11	-,029	1,105
Oberhaus	Pv_m2	9	16,44	105,96	55,48	26,91	,292	,592
Pichl	Pv_m2	15	14,96	88,31	58,58	22,62	-,601	-,867
Preunegg	Pv_m2	3	14,30	104,31	45,94	50,61	1,714	
Ramsau	Pv_m2	65	8,33	190,19	83,59	50,75	,373	-,908
Rohrmoos	Pv_m2	33	7,92	192,22	76,47	49,31	,690	-,157
Schladming	Pv_m2	52	14,99	246,86	100,79	54,70	,654	-,250

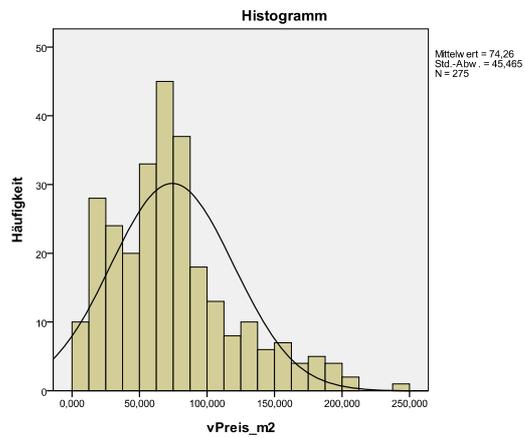
GB Gröbming

Gemeinde		N	Min	Max	\bar{m}	s	Schiefe	Kurtosis
Gröbming	Pv_m2	158	5,52	115,97	59,16	20,12	-0,333	0,526
Großsölk	Pv_m2	22	9,31	67,70	36,54	12,69	-0,110	1,280
Kleinsölk	Pv_m2	15	10,42	66,63	33,98	15,91	0,511	-0,166
Michaelaberg	Pv_m2	27	7,73	58,64	35,89	15,47	-0,694	-0,638
Mitterberg	Pv_m2	86	6,36	108,91	50,96	20,66	-0,170	-0,031
Öblarn	Pv_m2	37	16,24	75,01	42,84	13,93	0,187	-0,062
Pruggern	Pv_m2	26	12,76	55,61	31,89	12,59	0,343	-1,065
St. Martin	Pv_m2	35	5,62	96,07	35,89	15,36	1,719	6,799
St. Nikolai	Pv_m2	15	12,03	30,55	21,47	4,79	0,028	0,103

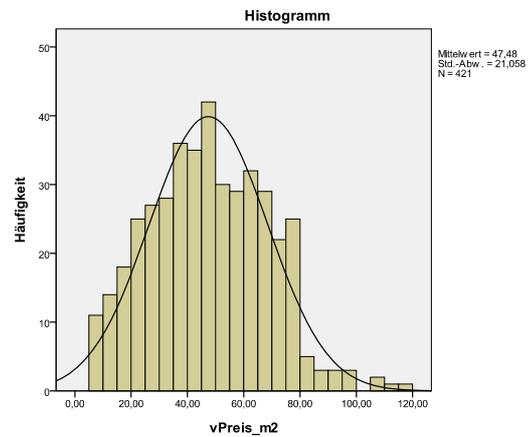
Anhang 11: Deskriptive Statistik und Histogramm der valorisierten Grundstückspreise auf Gerichtsbezirksebene nach Fehlerbereinigung und Übernahme in ArcGIS

Gerichtsbezirk

Pv_m2	N	Min	Max	\bar{m}	s	Schiefe	Kurtosis
GB Schladming	275	7,92	313,40	76,26	45,47	0,94	0,72
GB Gröbming	421	5,52	115,97	47,48	21,06	0,232	-0,256



GB Schladming



GB Gröbming

**Anhang 12: Test auf Normalverteilung – GB Schladming und GB Gröbming nach Fehlerbereinigung und Übernahme in ArcGIS
Kolmogorov-Smirnov und Shapiro-Wilk Test**

Tests auf Normalverteilung für die Katastralgemeinden im GB Schladming

Gemeinde		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistik	df	Signifikanz	Statistik	df	Signifikanz
Aich	Pv_m2	,155	31	,057	,927	31	,037
Gössenberg	Pv_m2	,160	10	,200*	,903	10	,236
Haus	Pv_m2	,143	57	,006	,947	57	,014
Oberhaus	Pv_m2	,174	9	,200*	,948	9	,671
Pichl	Pv_m2	,185	15	,175	,926	15	,236
Preunegg	Pv_m2	,368	3	.	,791	3	,092
Ramsau	Pv_m2	,092	65	,200*	,949	65	,010
Rohrmoos	Pv_m2	,153	33	,048	,939	33	,063
Schladming	Pv_m2	,114	52	,088	,950	52	,029

a. Signifikanzkorrektur nach Lilliefors

*. Dies ist eine untere Grenze der echten Signifikanz.

Tests auf Normalverteilung für die Katastralgemeinden im GB Gröbming

Gemeinde		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistik	df	Signifikanz	Statistik	df	Signifikanz
Groebming	Pv_m2	,082	158	,011	,975	158	,006
Grosssoelk	Pv_m2	,161	22	,143	,939	22	,192
Kleinsoelk	Pv_m2	,198	15	,118	,925	15	,230
Michaelerberg	Pv_m2	,237	27	,000	,891	27	,008
Mitterberg	Pv_m2	,095	86	,053	,973	86	,072
Oeblarn	Pv_m2	,082	37	,200*	,979	37	,700
Pruggern	Pv_m2	,128	26	,200*	,944	26	,170
Sankt Martin	Pv_m2	,164	35	,018	,846	35	,000
Sankt Nikolai	Pv_m2	,094	15	,200*	,991	15	1,000

a. Signifikanzkorrektur nach Lilliefors

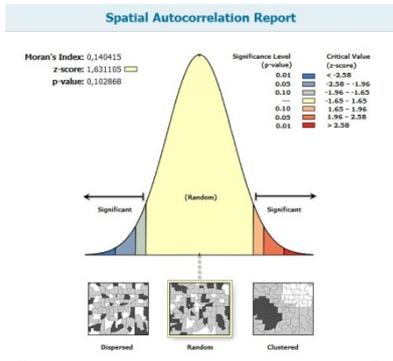
*. Dies ist eine untere Grenze der echten Signifikanz.

Test auf Normalverteilung für den Gerichtsbezirk

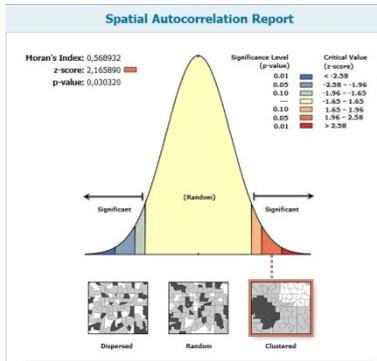
Gerichtsbezirk		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistik	df	Signifikanz	Statistik	df	Signifikanz
Schladming	Pv_m2	,111	287	,000	,935	287	,000
Gröbming	Pv_m2	,041	457	,061	,986	457	,000

a. Signifikanzkorrektur nach Lilliefors

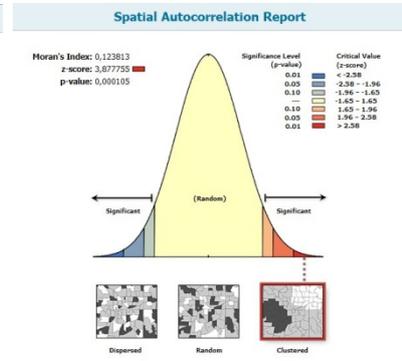
Anhang 13: Test auf Autokorrelation GB Schladming - Moran's I-Index nach Fehlerbereinigung und Übernahme in ArcGIS



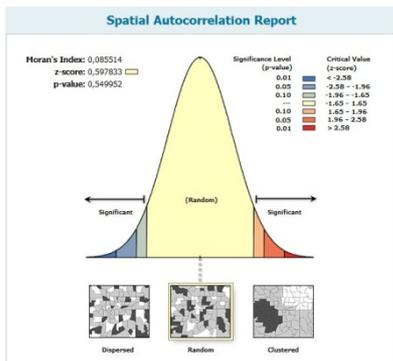
Aich



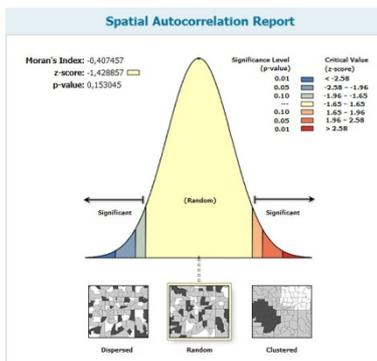
Gössenberg



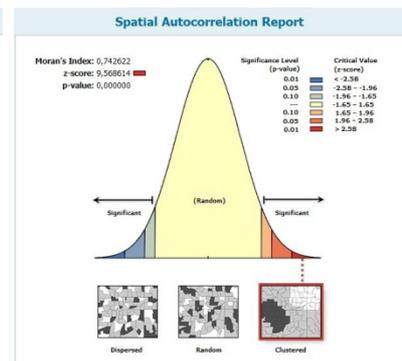
Haus



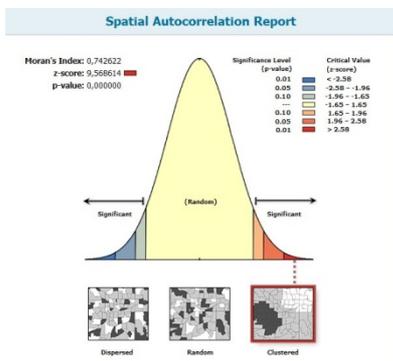
Oberhaus



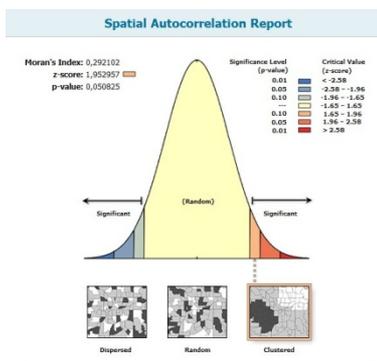
Pichl



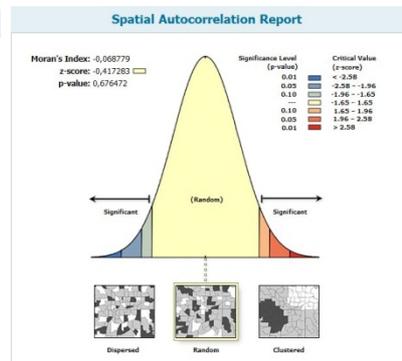
Preunegg



Ramsau

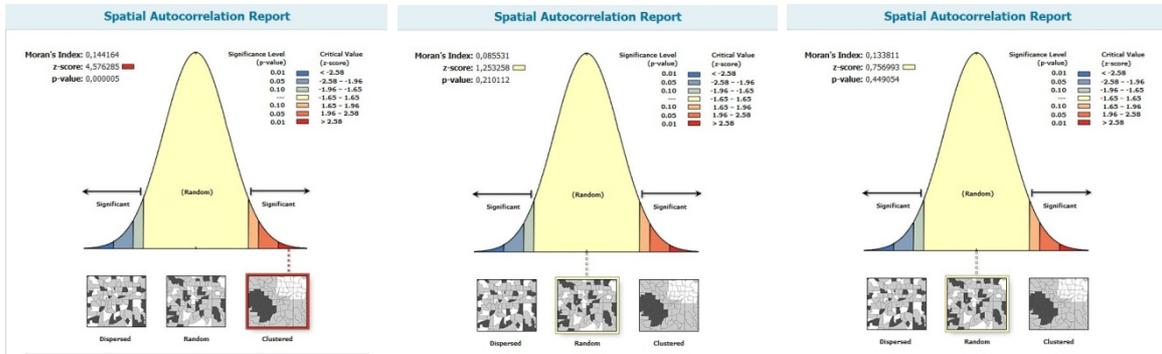


Rohrmoos



Schladming

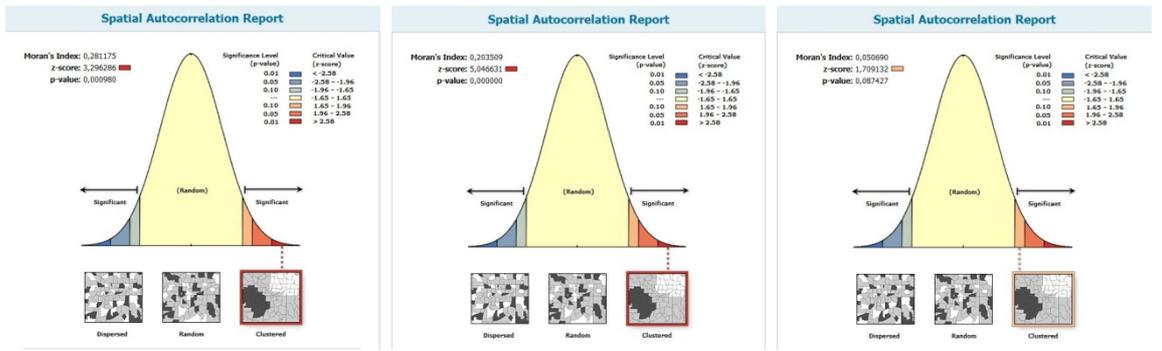
Anhang 14: Test auf Autokorrelation GB Gröbming - Moran's I-Index nach Fehlerbereinigung und Übernahme in ArcGIS



Gröbming

Großsölk

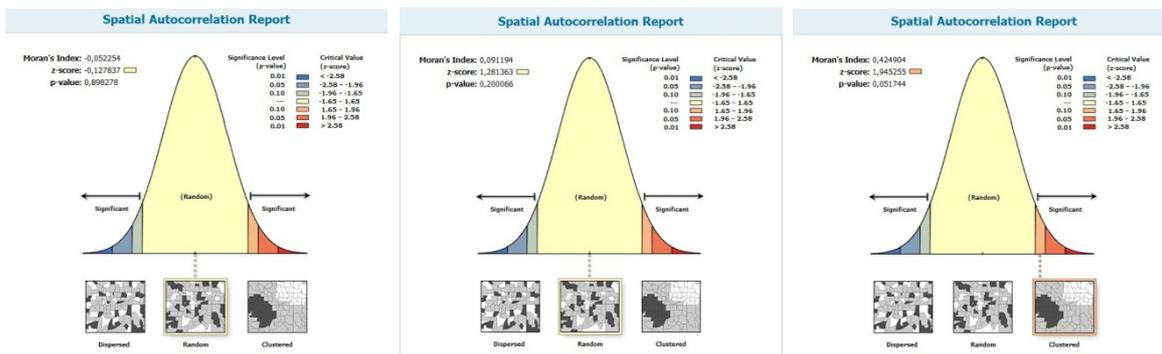
Kleinsölk



Michaelerberg

Mitterberg

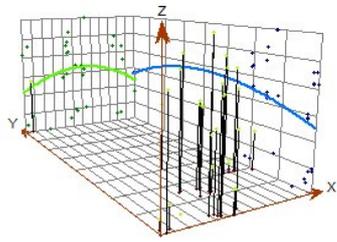
Öblarn



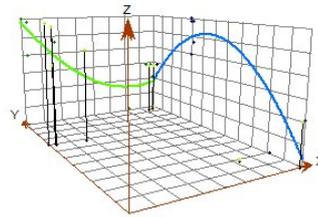
Pruggern

St. Martin

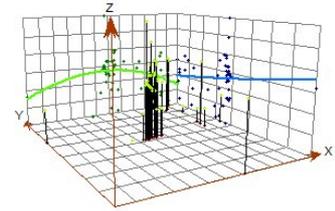
St. Nikolai

Anhang 15: Trendanalyse – GB Schladming

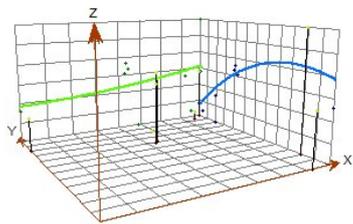
Aich



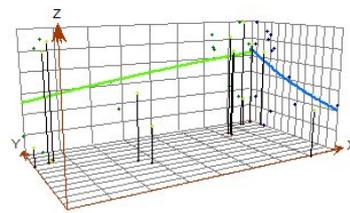
Gössenberg



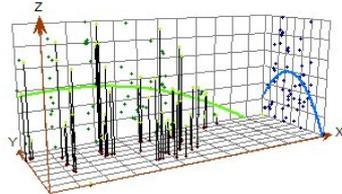
Haus



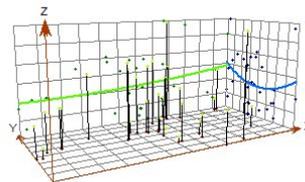
Oberhaus



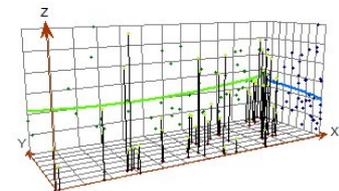
Pichl



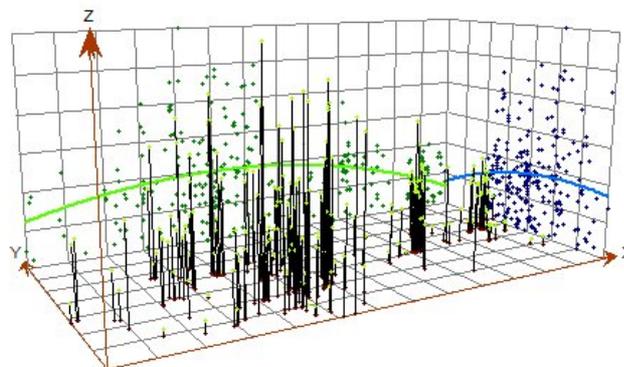
Ramsau



Rohrmoos

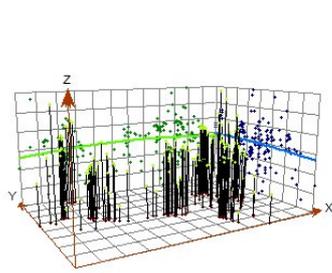


Schladming

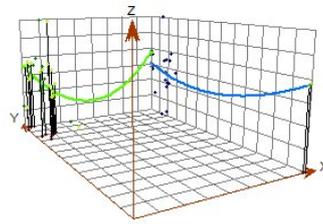


GB Schladming

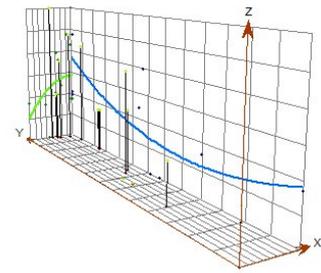
Anhang 16: Trendanalyse – GB Gröbming



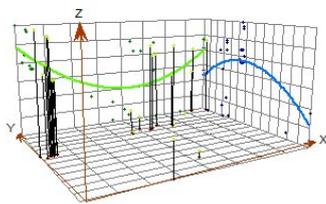
Gröbming



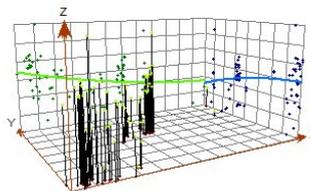
Großsölk



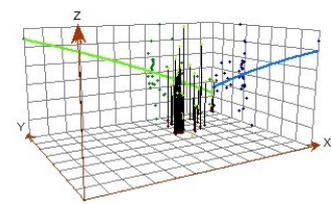
Kleinsölk



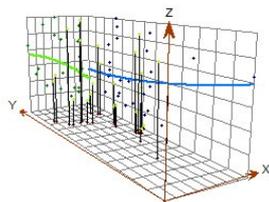
Michaelerberg



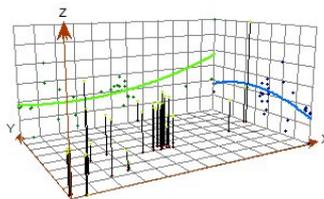
Mitterberg



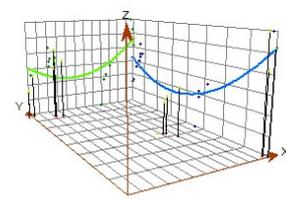
Öblarn



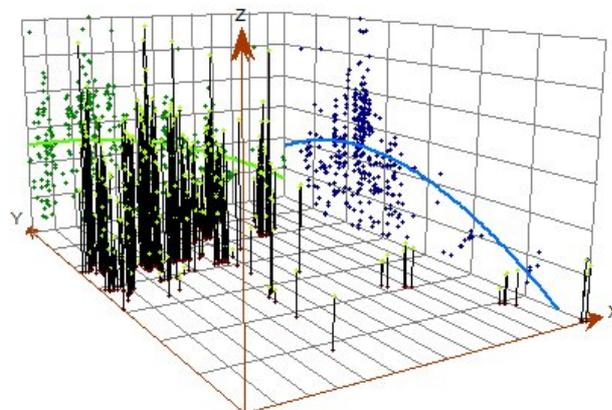
Pruggern



St. Martin

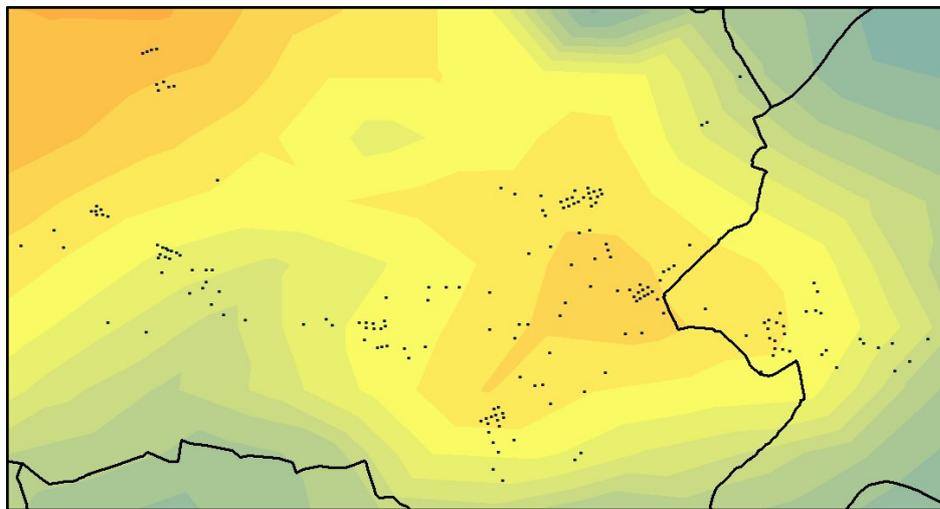


St. Nikolai

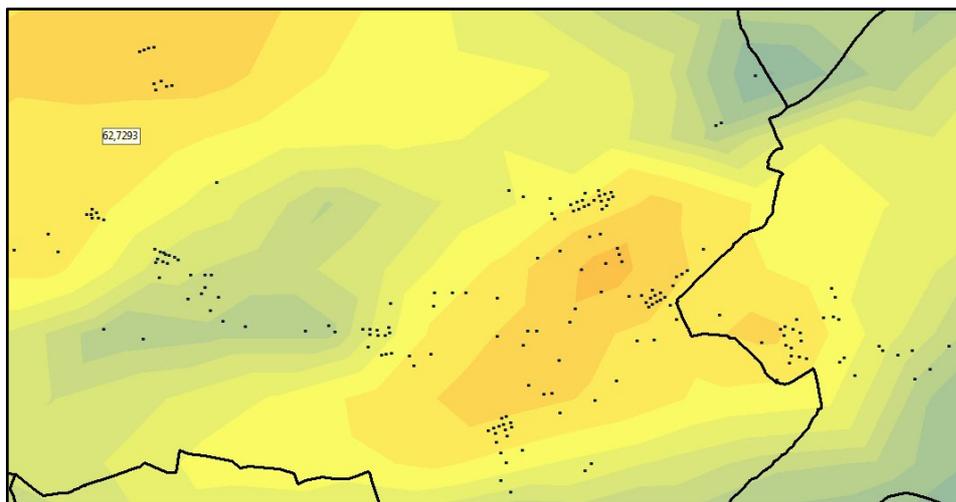


GB Gröbming

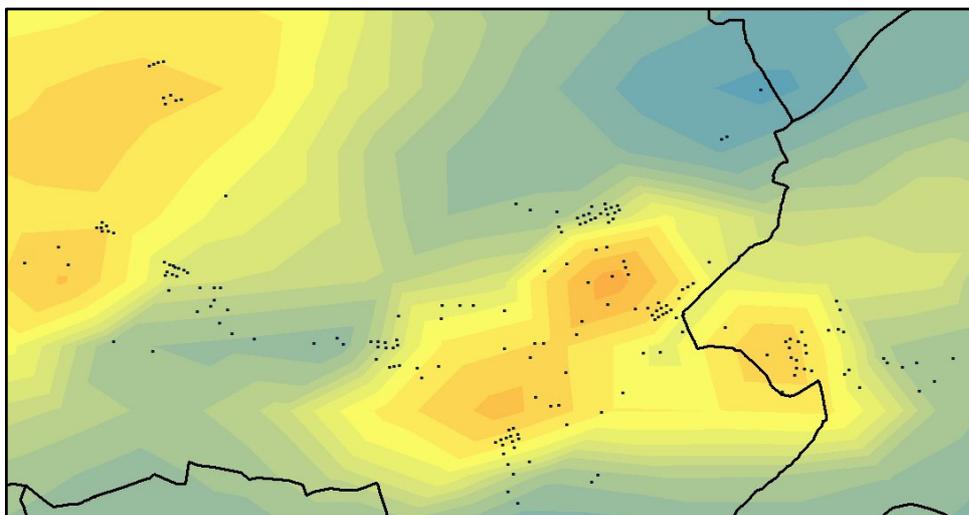
Anhang 17: Deterministische Interpolationsoberflächen im Bereich Gröbming



LPI

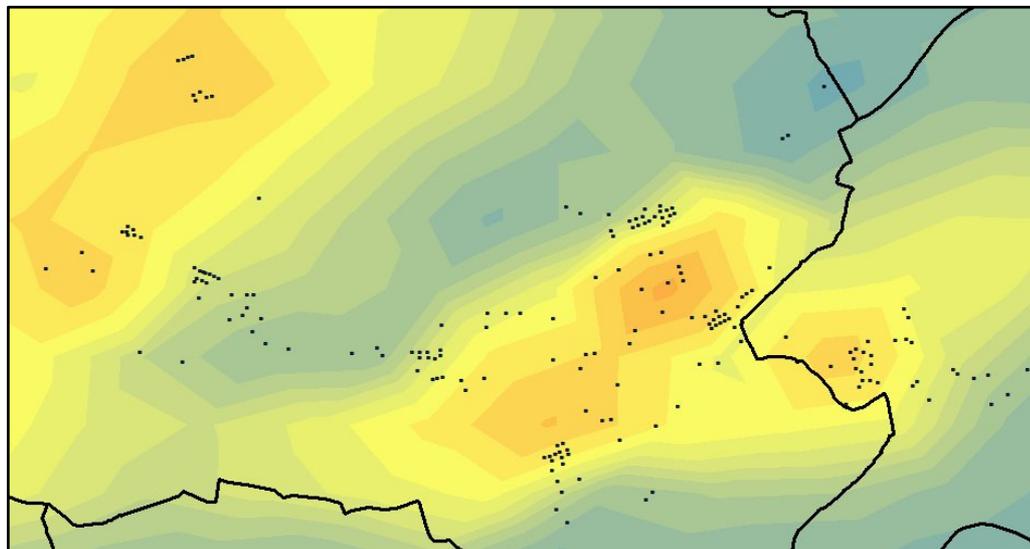


IDW

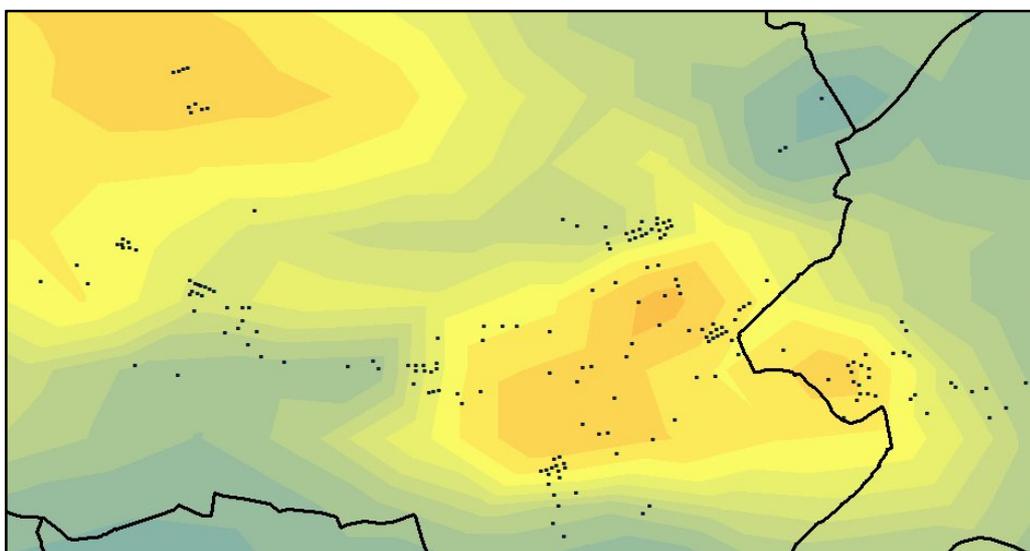


CRS

Anhang 18: Kriging Interpolationsoberflächen im Bereich Gröbming



Ordinary Kriging



Universal Kriging