



TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

Diplomarbeit

WELCHES INTERPOLATIONSVERFAHREN IST FÜR DIE PRÄDIKTION VON
DIGITALEN GELÄNDEHÖHENDATEN NACH DEN
ICAO-QUALITÄTSANFORDERUNGEN OPTIMAL?

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines

Diplom-Ingenieurs

unter der Leitung von

Privat Doz. Dr. Gerhard Navratil

(Gußhausstr. 27-29 / E120 (CD0342) Wien, 1040)

eingereicht an der Technischen Universität Wien

Fakultät für Mathematik und Geoinformation

von

Christoph Weichselbaum

1126113

Wien, im März 2020

Gerhard Navratil

Christoph Weichselbaum



TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

Ich habe zur Kenntnis genommen, dass ich zur Drucklegung meiner Arbeit unter der Bezeichnung

Diplomarbeit

nur mit Bewilligung der Prüfungskommission berechtigt bin. Ich erkläre weiters Eides statt, dass ich meine Diplomarbeit nach den anerkannten Grundsätzen für wissenschaftliche Abhandlungen selbstständig ausgeführt habe und alle verwendeten Hilfsmittel, insbesondere die zugrunde gelegte Literatur, genannt habe.

Weiters erkläre ich, dass ich dieses Diplomarbeitsthema bisher weder im In- noch Ausland (einer Beurteilerin/einem Beurteiler zur Begutachtung) in irgendeiner Form als Prüfungsarbeit vorgelegt habe und dass diese Arbeit mit der vom Begutachter beurteilten Arbeit übereinstimmt.

Wien, im März 2020

Christoph Weichselbaum

Danksagung

Diese Diplomarbeit entstand im Rahmen meiner Arbeitstätigkeit bei der Fa. FREQUENTIS AG und bildet den Abschluss meines Studiums der Geodäsie und Geoinformation an der Fakultät für Mathematik und Geoinformation der Technischen Universität Wien. An dieser Stelle möchte ich mich bei all denjenigen bedanken, die mich während der Anfertigung dieser Diplomarbeit unterstützt und motiviert haben.

Zuerst möchte ich mich bei meinen Kolleginnen und Kollegen der Fa. FREQUENTIS AG bedanken, die mich zum Thema dieser Diplomarbeit inspiriert und während der Anfertigung stets wertvollen Input gegeben haben.

Ebenfalls gebührt mein Dank Herrn Dr. Navratil, der meine Diplomarbeit betreut und begutachtet hat. Für die hilfreichen Anregungen und das konstruktive Feedback bei der Erstellung dieser Arbeit möchte ich mich herzlich bedanken.

Ich möchte mich bei meinen beiden Brüdern Jürgen und René bedanken, die mich während meines Studiums oft mit wertvollem Rat unterstützt haben.

Abschließend möchte ich mich ausdrücklich bei meinen Eltern Wilma und Josef bedanken, die mir mein Studium durch ihre Unterstützung ermöglicht haben und mir stets zur Seite gestanden sind.

Kurzfassung

In Zeiten der Digitalisierung steht man in der Flugsicherung vor der Herausforderung, große Mengen an Geodaten unterschiedlicher Qualität und Referenzsysteme in einem einheitlichen System zu integrieren und trotzdem zu gewährleisten, dass Mindeststandards eingehalten werden. Zentrale Werkzeuge zur Erstellung und Bearbeitung digitaler Geländehöhenmodelle (DHM) sind Interpolationsverfahren und die damit verbundenen Schätzungen von Höhenwerten. Ziel dieser Diplomarbeit ist die Evaluierung des geeignetsten Verfahrens unter Anbetracht der Anforderungen der internationalen Zivilluftfahrtorganisation (ICAO). Dazu wird die Forschungsfrage gestellt: Welches Interpolationsverfahren ist für die Prädiktion von digitalen Geländehöhendaten nach den ICAO-Qualitätsanforderungen optimal?

Zur Beantwortung dieser Frage wurde der aktuelle Stand der Forschung für räumliche Interpolationsverfahren evaluiert. Aufbauend wurde ein Anforderungsprofil anhand der ICAO-Qualitätsanforderungen aufgestellt. Unter der Prämisse, dass das Interpolationsergebnis aus sicherheitskritischen Gründen das reale Gelände möglichst genau annähern soll, wurden die Kriging- und Spline-Interpolation, aufgrund ihrer fortgeschrittenen und flexiblen Modelle, für weiterführende Analysen ausgewählt. Ein zweiter wesentlicher Punkt ist die Forderung nach Genauigkeitsinformationen und Konfidenzbereichen auf Datenebene (pro Feature bzw. Rasterzelle). Die Evaluierung der optimalen Interpolationsmethode erfolgte auf Basis von Referenzdaten und Berechnung verschiedener Interpolationsmodelle ausgewählter Gebiete für das Land Niederösterreich. Durch Vergleich der Interpolationsergebnisse wurde die optimale Interpolationsmethode ausgewählt.

Allgemein lässt sich sagen, dass es keine eindeutige Antwort dafür gibt, welche Interpolationsmethode generell am genauesten arbeitet. Die Wahl der Interpolationsmethode ist stark vom Einsatzgebiet abhängig. Richtet man sich streng nach den ICAO-Anforderungen, ist die Angabe eines Grenzwertes mit zugehöriger Konfidenz aus den Daten einer Kreuzvalidierung für die Qualitätsbeurteilung ausreichend, solange dieser Grenzwert den Bestimmungen des jeweiligen Anforderungsgebietes genügt. Allgemein zeigte sich, dass Universal Kriging die besten Resultate liefert. Zudem ermöglicht Kriging die Ableitung von punktuellen Genauigkeitsinformationen. Da eine sorgfältige Datenanalyse maßgeblich für die Genauigkeit des Endresultats ist und berechnete Genauigkeitsinformationen auf Datenebene weitere regionale und vor allem statistisch belastbare Aussagen ermöglichen, ist Kriging für die Prädiktion von digitalen Geländehöhendaten im Kontext der ICAO-Qualitätsanforderungen *optimal* geeignet.

Liegen auf Basis statistischer Interpolationsmethoden Schätzungen für die Genauigkeiten von Interpolationspunkten vor, können diese variabel und je nach Anwendungsfall im Qualitätsbeurteilungsprozess berücksichtigt werden. Allgemein kann gefolgert werden, dass sich ein stochastisches Modell und die Kreuzvalidierung im Rahmen des Anwendungsfalls der Flugsicherung komplementieren und somit die besten Mittel für einen möglichst *sicheren* Qualitätsbeurteilungsprozess bereitstellen.

Abstract

In times of digitization, air traffic control is confronted with the challenge of integrating large amounts of geodata of different quality and reference systems in a uniform system and still ensuring that minimum standards are met. Spatial interpolation and respective prediction of elevation values is a central tool for creating and editing digital elevation models (DEM). The aim of this diploma thesis is the evaluation of the most suitable method considering the requirements of the international civil aviation organization (ICAO). Hence, the research question is formulated: Which interpolation method is optimal for the prediction of digital elevation models according to the ICAO quality requirements?

To answer this question, the current state of research for spatial interpolation methods was evaluated. Furthermore, a requirement profile based on the ICAO quality requirements was established. Based on the premise that the interpolation result should approximate the real terrain as accurately as possible due to safety-critical reasons, the kriging and spline interpolation, based on their advanced and flexible models, were selected for further analysis. A second essential point is the requirement for accuracy information and confidence levels at the data level (per feature or grid cell). The optimal interpolation method was evaluated on the basis of reference data and the calculation of various interpolation models for selected areas for the state of Lower Austria. The optimal interpolation method was selected by comparing the interpolation results.

In general, it can be stated that there is no clear answer as to which interpolation method generally works most precisely. The choice of interpolation method depends heavily on the area of application. If you strictly follow the ICAO requirements, the specification of a threshold value with associated confidence, derived from a cross-validation, is sufficient for the quality assessment, as long as this threshold value meets the requirements of the respective area of interest. In general, it turned out that Universal Kriging provides the best results. In addition, Kriging enables the derivation of selective accuracy information. Since careful data analysis is crucial for the accuracy of the end result and calculated accuracy information at the data level enables further regional and especially statistically reliable statements, Kriging provides the *optimal* means for the prediction of digital elevation models in the context of the ICAO quality requirements.

If estimates for the accuracy of interpolation points are available based on statistical interpolation methods, these can be taken into account variably and depending on the application in the quality assessment process. In general, it can be concluded that a stochastic model and cross-validation complement each other in the context of the use in air traffic control and thus provide the best means for a *safe* quality assessment process.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Relevanz der Fragestellung	1
1.1.1	Geländehöhendaten in der Flugsicherung	2
1.1.2	Anwendungsfälle der Interpolation in der Flugsicherung	3
1.2	Lösungsansatz	3
2	Grundlagen	5
2.1	Explorative Datenanalyse	5
2.1.1	Räumliche Struktur	5
2.1.2	Datenverteilung	11
2.1.3	Ausreißeranalyse	17
2.2	Räumliche Interpolation	19
2.2.1	Deterministische Interpolationsmethoden	20
2.2.2	Statistische Interpolation	28
2.2.3	Beurteilung der Interpolationsqualität	36
3	Stand der Forschung	41
3.1	Analyse von Interpolationsmethoden bei unterschiedlicher Topographie	41
3.2	Analyse von Interpolationsmethoden anhand von GPS-Referenzmessungen	47
3.3	Analyse der Eigenschaften von Interpolationsmethoden	49
3.4	Elektronische Gelände- und Hindernisdaten nach ICAO-Standards	52
3.5	Zusammenfassung und Bewertung des Forschungsstandes	55
4	ICAO Anforderungsanalyse	56
4.1	Qualitätsanforderungen der ICAO Spezifikation	56
4.1.1	Anforderungen von ICAO Annex 15	57
4.1.2	Anforderungen von PANS AIM	63
4.2	Vorschlag geeigneter Interpolationsverfahren	69
5	Evaluierung von Interpolationsverfahren	72
5.1	Datengrundlage	72
5.2	Explorative Datenanalyse	74
5.3	Interpolation und Vergleich der Ergebnisse	81
5.3.1	Interpolationsergebnisse	82
5.3.2	Qualitätsbeurteilung	86
5.4	Bewertung der Interpolationsverfahren	89

6	Abschließende Bewertung	91
6.1	Diskussion	91
6.2	Ausblick	93

Die approbierte gedruckte Originalversion dieser Diplomarbeit ist an der TU Wien Bibliothek verfügbar.
The approved original version of this thesis is available in print at TU Wien Bibliothek.

1 Einleitung

1.1 Relevanz der Fragestellung

Die internationale Luftfahrtbehörde (International Civil Aviation Authority, ICAO) fordert von allen Mitgliedsstaaten mit den Dokumenten 'Annex 15' (vgl. ICAO, 2018a) und 'PANS AIM' (vgl. ICAO, 2018b) die Bereitstellung von elektronischen Gelände- und Hindernisdaten in einer Datenbank (electronic Terrain and Obstacle Database, eTOD). Dabei werden klare Anforderungen an Abdeckungsgebiete, Datenbereitstellung sowie Inhalt, numerische Anforderungen und Struktur von Geländedaten in der Flugsicherung gestellt. Für Staaten stellt sich daher die Frage, wie man diese Anforderungen kosteneffizient erfüllen kann. Ein flächendeckendes Geländehöhenmodell könnte beispielsweise durch die Integration vorhandener Daten mit globalen Modellen erreicht werden. Die Ableitung eines qualitätsgesicherten Gesamtmodells ist aber auch für zentrale Instanzen, wie die Europäische Organisation zur Sicherung der Luftfahrt (European Organisation for the Safety of Air Navigation, EUROCONTROL) interessant. Die von der EUROCONTROL betriebene Europäische Luftfahrt-datenbank (European Aeronautical Information Service Database, EAD) dient als zentraler Provider für nationale Flugsicherungen (Air Navigation Service Provider, ANSPs) weltweit. ANSPs können Luftfahrt-daten bereitstellen und abfragen. Dabei gilt es seitens der EAD die eingehenden Daten von unterschiedlichen Quellen zu prüfen, in das System zu integrieren und dem globalen Netz für Luffahrtinformationen (Aeronautical Information Management, AIM) zur Verfügung zu stellen. Diese Diplomarbeit wird in Zusammenarbeit mit der FREQUENTIS AG erstellt. Das strategische Geschäftsfeld Aeronautical Information Management (AIM) entwickelt und erweitert sowohl die EAD als auch ein eTOD-System für nationale und internationale Flugsicherungen. Das EAD-System ist operationell im Einsatz. Das eTOD-System ist als server- und datenbankgestützte WebGIS-Anwendung zum Erstellungszeitpunkt dieser Diplomarbeit bereits implementiert und ermöglicht das Management von georeferenzierten Raster- und Vektordaten, sowie zusätzlichen Metadaten. Da Interpolation als zentrales Werkzeug für die Erstellung und Bearbeitung von digitalen Geländehöhendaten in eine sicherheitskritischen Anwendung integriert werden soll, hat die Implementierung konform internationaler Standards der Flugsicherung hohe Priorität.

Generell steht man vor der Herausforderung, Datensätze unterschiedlicher Qualität und Referenzsysteme in einem einheitlichen System zu integrieren und trotzdem zu gewährleisten, dass Mindeststandards der Flugsicherung eingehalten werden. Zentrale Werkzeuge zur Erstellung und Bearbeitung digitaler Geländehöhenmodelle (DHMs) sind Interpolationsverfahren und die damit verbundenen Schätzungen von Höhenwerten. Es existiert eine Vielzahl an Interpo-

lationsmethoden, wobei jede für sich wiederum bestimmte Vor- und Nachteile aufweist. Eine *optimale* Interpolationsmethode würde nicht nur die internationalen Qualitätsanforderungen erfüllen, sondern auch entsprechend flexibel einsetzbar sein um verschiedenste Aufgabenstellungen mit maximaler Sicherheit und Qualität in der Flugsicherung zu lösen. Folglich leitet sich die zentrale Fragestellung dieser Diplomarbeit ab: Welches Interpolationsverfahren ist für die Prädiktion von digitalen Geländehöhendaten nach den ICAO-Qualitätsanforderungen optimal?

1.1.1 Geländehöhendaten in der Flugsicherung

Die International Civil Aviation Organization (ICAO) hat mit den Dokumenten 'ICAO Annex 15' und 'PANS AIM' die Anforderungen zur Bereitstellung von digitalen Gelände- und Hindernisdaten an alle Mitgliedsstaaten gestellt. Dabei werden Daten in Gebiete mit unterschiedlichen numerischen Anforderungen rund um einen Flughafen eingeteilt. Ein eTOD-System übernimmt das Datenmanagement und leitet Folgeprodukte ab. In solchen geographischen Informationssystemen (GIS) und Geodatenbanken (Spatial DB) werden eine Vielzahl unterschiedlicher Datensätze gespeichert und analysiert. Diese Daten unterscheiden sich in technischer (z.B. Format, Größe) und semantischer Hinsicht (z.B. Thema, Darstellung, Bezugsrahmen).

Digitale Gelände- und Hindernisdaten sind essentiell für die Sicherheit in der Luftfahrt. Beispielsweise ist bei Abflug- und Landemanövern ein detailliertes Wissen über das umliegende Gelände und den darauf befindlichen Hindernissen von großer Bedeutung. Die Speicherung von Gelände- und Hindernisdaten in digitaler Form ist ein wesentliches Anliegen der ICAO, um die Luftfahrtsicherheit in Zeiten vernetzter IT-Systeme zu optimieren. Konkrete Anwendungsfälle für ein eTOD-System sind:

- Geländewarnsysteme
- Procedure Design (Planung von Strecken, Wartemustern, Luftraumstrukturen, etc.)
- Notlandungen
- Movement Guidance and Control System (Systeme zur Steuerung und Kontrolle von Bewegungen auf der Erdoberfläche)
- Erstellung von Luftfahrtkarten und On-Board Datenbanken
- Detektion, Eingrenzung und Beseitigung von Hindernissen
- Radio-Altimeter Höhenbestimmung
- Synthetic Vision
- Flugsimulation

1.1.2 Anwendungsfälle der Interpolation in der Flugsicherung

Geländehöhendaten sind in der Flugsicherung die Grundlage vieler Anwendungsfälle. Die Interpolation wiederum ist ein wesentliches Werkzeug zur Erstellung und Bearbeitung von Geländehöhendaten. Alle Anwendungsfälle haben gemein, dass Höhenwerte an Orte gesucht werden, wo keine Daten vorhanden sind. Entweder sind die Basisdaten unregelmäßig in einer Punktwolke verstreut oder es liegen für ein räumliches Gebiet verschiedene DHM-Datensätze (mit möglicherweise unterschiedlicher Qualität) vor. Das Endprodukt wird üblicherweise im Raster-Dateiformat bereitgestellt. Rasterdaten unterscheiden sich möglicherweise in räumlicher Ausdehnung, Auflösung, usw. Zur Darstellung in einem einheitlichen Format, müssen die Daten zuerst durch Resampling unter Verwendung eines Interpolationsverfahrens in ein einheitliches Modell transformiert werden. Beim Resampling wird die räumliche Auflösung des DHM-Datensatzes verändert. Dabei kann entweder ein engmaschigeres oder ein breitmaschigeres Modell erzeugt werden. Ein engmaschigeres Modell kann in der Flugsicherung nötig sein, um die geforderte Auflösung in speziellen Gebieten um den Flughafen zu erreichen. Zusätzlich kann das bestehende Modell durch qualitativ hochwertigere Modelle verbessert werden. Die Erstellung eines breitmaschigeren Datensatzes kann zum Zwecke der Speicheroptimierung sinnvoll sein, solange die notwendigen Anforderungen noch eingehalten werden. Das bestehende Modell kann durch Hinzufügung weiterer DHMs geographisch ausgedehnt werden um auch andere geographische Gebiete abzudecken. Auf Basis der bestehenden Daten sind auch weiterführende, regionale Analysen möglich. Beispielsweise können alle Geländehöhen interessant sein, die eine sicherheitskritische Begrenzungsfläche durchstoßen, um das Gefahrenpotential abzuklären. Letztendlich liegt im Kontext der Flugsicherung aber bei allen Anwendungsfällen spezielles Augenmerk auf der verlässlichen und aussagekräftigen Qualitätsbeurteilung.

1.2 Lösungsansatz

Für die Beantwortung der Fragestellung dieser Diplomarbeit werden im ersten Schritt notwendige Grundlagen, nämlich die explorative Datenanalyse und räumliche Interpolation erarbeitet. Es wird auf rein mathematische (z.B. Variationsansatz) und statistische (z.B. Kriging) Verfahren sowie Methoden zur Beurteilung der Ergebnisse eingegangen. Anschließend wird Bezug auf den Stand der Forschung genommen. Dabei werden gängige Interpolationsverfahren und deren Ergebnisse bei unterschiedlichen Geländetypen analysiert. Zusätzlich werden relevante Abschnitte einer Anleitung der EUROCONTROL, mit Empfehlungen zur Implementierung eines eTOD-Systems evaluiert. Die nötigen Qualitätsanforderungen an digitale Geländehöhendaten werden auf Basis der ICAO Spezifikationen 'ICAO Annex 15' und 'PANS AIM' erhoben. Für die Erstellung eines Anforderungsprofils werden die geforderten Qualitätsparameter identifiziert. Auf Basis des definierten Anforderungsprofils erfolgt die Eingrenzung geeigneter Interpolationsverfahren durch Abwägung ihrer jeweiligen Vor- und

Nachteile. Anschließend wird die Eignung von als qualifiziert erachteten Interpolationsmethoden anhand von praktischen Berechnungen verifiziert. Aufbauend erfolgt der Vorschlag der optimalen Interpolationsmethode für die Prädiktion von digitalen Geländehöhen nach den ICAO-Qualitätsanforderungen. Abschließend werden die Ergebnisse dieser Diplomarbeit zusammengefasst, analysiert und ein Ausblick gegeben.

2 Grundlagen

In diesem Kapitel werden die Grundlagen zur Bearbeitung der Fragestellung dieser Diplomarbeit erarbeitet. Das erste Themenfeld der explorativen Datenanalyse (EDA) beschäftigt sich mit der Untersuchung und Aufbereitung vorhandener Daten für die Interpolation. Dabei wird speziell auf die räumliche Struktur, Datenverteilung und Ausreißeranalyse eingegangen. Das zweite Themenfeld der räumlichen Interpolation beschreibt deterministische und statistische Interpolationsmethoden. Abschließend wird auf Verfahren zur Beurteilung der Interpolationsqualität Bezug genommen.

2.1 Explorative Datenanalyse

Interpolationsverfahren arbeiten auf Basis mathematischer und statistischer Modellannahmen. Die Hauptaufgabe der Datenanalyse ist die Qualitätsbeurteilung der Eingabedaten, um möglichst genaue Endergebnisse zu bekommen. Essenziell für die Analyse räumlicher Daten ist, dass diese nicht nur zu ihrer Verteilung, sondern auch zu ihren Nachbarn passen. Wichtige Untersuchungen müssen daher bezüglich räumlicher Struktur, Datenverteilung und Ausreißeranalyse vorgenommen werden (vgl. Navratil, 2006a, S. 197).

2.1.1 Räumliche Struktur

Ein wichtiger Grundsatz der Geographie wird durch *Tobler's First Law of Geography* begründet: ‚Alle Dinge hängen zusammen, aber nahe beieinander liegende Dinge hängen mehr zusammen als weiter voneinander entfernte.‘ Dieser Grundsatz ist das Fundament vieler Interpolationsverfahren, deren Ziel es ist, die wahre Erdoberfläche möglichst genau anzunähern. Bei der Interpolation kann zwischen deterministischen und statistischen Methoden unterschieden werden.

Bei deterministischen Interpolationsmethoden (siehe Abschnitt 2.2.1) werden globale oder lokale Annahmen getroffen, wie z.B.:

- Anzahl miteinbezogener Punkte oder Suchradius für die Interpolation
- Glatter Verlauf durch die Messpunkte

Diese Methoden werden in der Regel direkt auf die Daten im Untersuchungsgebiet angewendet und das Ergebnis in Abhängigkeit des zu modellierenden Phänomens verfeinert.

Bei statistischen Interpolationsmethoden (siehe Abschnitt 2.2.2) werden die räumlichen Beziehungen zwischen allen Messpunkten vor der eigentlichen Interpolation modelliert. Der Zusammenhang zwischen räumlich verteilten Punkten wird durch die räumliche Autokorrelation beschrieben. Dabei werden Zusammenhänge zwischen allen Punktepaaren im Untersuchungsgebiet erhoben. Folgende Funktionen finden dafür Verwendung:

- Kovarianzfunktion
- Variogrammfunktion

Beide Funktionen sind unabhängig voneinander durch die Entwicklung der Interpolation nach kleinsten Quadraten (Kovarianzfunktion) und Kriging (Variogrammfunktion) entstanden. Zudem sind beide Interpolationsmethoden praktisch ident, wobei Kovarianz- und Variogrammfunktion ineinander überführt werden können (vgl. Kraus, 1998, S. 47).

Eine fundamentale Annahme statistischer Methoden ist die *Stationarität*. Diese geht davon aus, dass Punkte gleichen Abstandes und Richtung eine ähnliche Korrelation aufweisen. Ist die Korrelation nur vom Abstand abhängig, so spricht man von *Isotropie*. Anisotropie, also richtungsabhängige Einflüsse, sollte für optimale Interpolationsergebnisse in der Kovarianz- oder Variogrammfunktion berücksichtigt werden.

Stationarität

Stationarität beschreibt generell eine zeitliche Unveränderlichkeit eines stochastischen Prozesses. Dabei wird zwischen folgenden Stufen der Stationarität unterschieden (vgl. Niemeier, 2008, S. 48)

- Mittelwertstationär: Mittelwertfunktion ist zeitlich konstant
- Varianzstationär: Varianzfunktion ist zeitlich konstant
- Kovarianzstationär: Kovarianzfunktion ist nur vom zeitlichen Abstand abhängig

Sind alle drei Forderungen erfüllt, spricht man von 'schwacher' Stationarität. Von 'starker' Stationarität wird gesprochen, wenn die Verteilungsfunktion der Zufallsvariablen im Untersuchungsgebiet zeitlich unveränderlich ist. In der Geostatistik wird anstelle einer zeitlichen Änderung einer Zufallsvariable $Z(t)$ von einer räumlichen Änderung einer regionalisierten Zufallsvariable $Z(x)$ ausgegangen. Analog lässt sich das Konzept der Stationarität für räumliche, stochastische Prozesse anwenden. In der Geostatistik geht man von der intrinsischen Stationarität aus, welche die folgenden Eigenschaften hat:

- Mittelwertstationär: $E(Z(x)) = \mu$ und $E(Z(x) - Z(x+h)) = 0$
- Varianzstationär: $Var(Z(x)) = \sigma^2$
- Kovarianzstationär: $E(Z(x_1) - Z(x_2))^2 = C(x_1 - x_2) = C(h)$

Intrinsisch stationäre Felder werden allgemein als homogen und isotrop bezeichnet (vgl. Navratil, 2006b, S. 191-192).

Kovarianzfunktion

Die Kovarianzfunktion ist Bestandteil der Interpolation nach der Methode der kleinsten Quadrate. Messwerte werden bei der statistischen Interpolation als Realisierungen eines stochastischen Prozesses aufgefasst. In der Praxis weisen stochastische Prozesse (z.B. Hangrutschung) oft einen langfristigen Trend, zyklische Schwankungen und zufällige Effekte auf. Dabei erfolgt meist eine Zerlegung in folgende Anteile nach dem Kollokationsansatz (vgl. Niemeier, 2008, S. 49/420)

$$x(t) = \text{Trend}(t) + \text{Signal}(t) + \text{Rauschen}(t) \quad (2.1)$$

bzw. in der Schreibweise der Regression (vgl. Abschnitt 2.2.2)

$$y = Xb + s + \varepsilon \quad (2.2)$$

wobei

- $Xb \dots$ Trend: systematische Veränderung der Mittelwertfunktion μ_t
- $s \dots$ Signal: Schwankungen der Mittelwertfunktion μ_t
- $\varepsilon \dots$ Rauschen: zufällige Restabweichungen

Der Trend wird nach einem Regressionsansatz bestimmt. Das Signal wird stochastisch aufgefasst und enthält eine *Erhaltensneigung* (gleiches oder ähnliches Verhalten von Residuen) die durch eine Autokovarianzfunktion ausgedrückt wird. Diese beschreibt die Korrelation des Signals mit sich selbst zu einem früheren Zeitpunkt. Das Rauschen wird durch zufällige Varianzen modelliert. Die Erfassung der räumlichen Struktur erfolgt durch die Bestimmung einer Kovarianzfunktion $C(\tau)$ für das Signal s . Diese Funktion soll analog zur Autokovarianzfunktion die Erhaltensneigung des zu modellierenden Prozesses widerspiegeln. Die Kovarianzfunktion wird durch zwei Parameter charakterisiert (vgl. Niemeier, 2008, S. 421):

- Die *Korrelationslänge* beschreibt, bis zu welchem Abstand τ eine Erhaltensneigung wirksam ist.
- Die *Form* beschreibt, wie stark benachbarte Werte korreliert sind und wie sich die Korrelation mit dem Abstand ändert.

Bei der empirischen Bestimmung von $\hat{C}(\tau)$ werden die Varianzen aller Beobachtungen und Kovarianzen für alle Punktepaare aus dem Feld der stationären Residuen z nach Abzug des Trends ermittelt:

$$z = yb - Xb = s + e \quad (2.3)$$

Dem Feld der Residuen z wird Homogenität und Isotropie unterstellt und hängt damit nur vom Abstand zwischen den Punkten ab. Bei vielen Daten erfolgt oft auch eine zusätzliche Unterteilung in Abstandsklassen mit anschließender Mittelwertbildung. Daraufhin wird eine ausgleichende Funktion ermittelt, die den mittleren Verlauf der Daten bestmöglich approximiert. Häufig verwendete Kovarianzfunktionen sind:

- Exponentiell: $C_1(\tau) = ae^{-b\tau}$
- Linear: $C_2(\tau) = a\tau$
- Periodisch: $C_3(\tau) = \frac{\sin(a\tau)}{a\tau}$

Die Parameter a und b steuern die Korrelationslänge und Form (vgl. Niemeier, 2008, S. 422).

Variogramm

Das Variogramm wird bei Verfahren der Geostatistik eingesetzt. Beim Kriging geht man von der intrinsischen Hypothese aus. Diese fordert, dass der Erwartungswert aller Zufallsvariablen Z im Untersuchungsgebiet konstant ist (vgl. Niemeier, 2008, S. 426-427)

$$E(Z(x)) - E(Z(x+h)) = 0 \quad (2.4)$$

und der räumliche Zusammenhang zweier Variablen nur von ihrem Abständen h abhängt

$$E[Z(x) - Z(x+h)]^2 = 2\gamma(h) \quad (2.5)$$

Die Semivarianz $\gamma(h)$ beschreibt den räumlichen Zusammenhang und kann empirisch berechnet werden:

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^n [Z(x_i) - Z(x_i+h)]^2 \quad (2.6)$$

Dabei wird für jeden gegebenen Messwert Z_i die Differenz zu allen übrigen Z_i im Abstand h gebildet. Bei der empirischen Berechnung des Variogramms wird analog zur Kovarianzfunktion häufig in Abstandsklassen unterteilt. Die Werte des Semivariogramms werden mit zunehmendem Abstand bis zu einem gewissen Grenzwert größer. Der empirische Verlauf kann wieder durch eine Funktion approximiert werden (vgl. Abbildung 2.1).

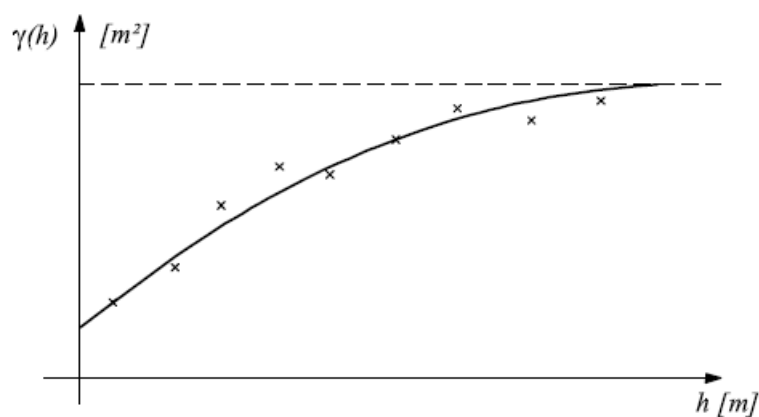


Abbildung 2.1: Empirisches und theoretisches Variogramm (Niemeier, 2008, S. 427)

Das theoretische Variogramm wird auf Basis des empirischen Variogramms erstellt. Das empirische Variogramm ist keine eindeutig bestimmte Funktion und unterscheidet sich je nach Wahl der einbezogenen Daten. Als theoretisches Variogramm wird jene Funktion gewählt, die das jeweilige Phänomen (z.B. DHM) am besten repräsentiert. Die Qualität der Interpolation ist dabei vom gewählten Variogrammmodell abhängig. Das Variogramm kann nun dazu genutzt werden, den Einfluss jedes einzelnen Punktes auf das Ergebnis der räumlichen Interpolation zu bestimmen. Damit entscheidet nicht eine deterministische Funktion (globale Interpolationsvorschrift), sondern eine aus den Daten abgeleitete und durch die Daten gesicherte Funktion über das Interpolationsergebnis. Das theoretische Variogramm ist für jedes h gültig. Häufig verwendete Modelle sind (vgl. Niemeier, 2008, S. 428):

- Lineares Modell: $\gamma(h) = c_0 + bh$
- Sphärisches Modell: $\gamma(h) = c_0c_1\left\{\frac{3h}{2a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right\}$
- Exponentielles Modell: $\gamma(h) = c_0 + c_1\{1 - e^{-\frac{h}{a}}\}$

Der Parameter c_0 bezeichnet eine Grundvarianz und wird beim Kriging als *Nugget-Effekt* bezeichnet. Dieser Effekt beschreibt das Verhalten im Bereich des Ursprungs $\gamma(0) = 0$ und kann als Messgenauigkeit interpretiert werden. Analog zur Kovarianzfunktion beschreiben die Parameter a und b die Korrelationslänge und Form des räumlichen Zusammenhangs und werden beim Kriging als *Range* und *Sill* bezeichnet. Die *Range* gibt die Korrelationsweite an, bis wohin ein räumlicher Zusammenhang besteht. Der dazugehörige Schwellwert wird *Sill* genannt.

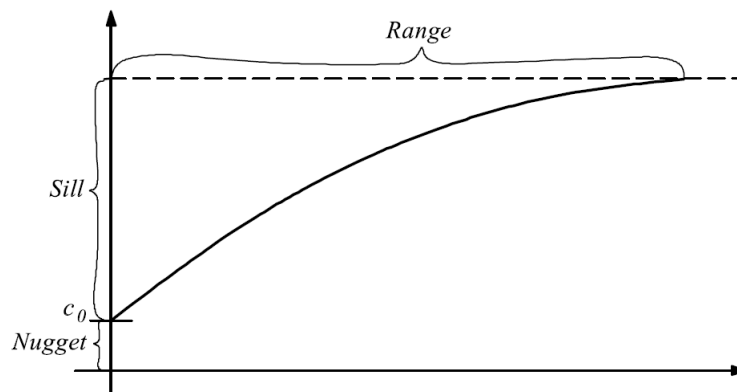


Abbildung 2.2: Variogramm-Parameter (Niemeier, 2008, S. 428)

Richtungsabhängige Einflüsse

Statistische Interpolationsmethoden setzen allgemein Homogenität und Isotropie im Untersuchungsgebiet voraus. Dies kann einerseits durch einen globalen Trend, der alle Messwerte betrifft, oder zusätzliche richtungsabhängige Einflüsse verursacht werden.

Trend

Ein allgemeiner Trend in den Daten wirkt sich auf alle Messpunkte aus und widerspricht der Annahme vieler statistischer Verfahren, dass der Mittelwert über das Untersuchungsgebiet konstant ist. Globale Trends können durch mathematische Funktionen (Polynome) beschrieben, entfernt und während der Interpolation berücksichtigt werden.

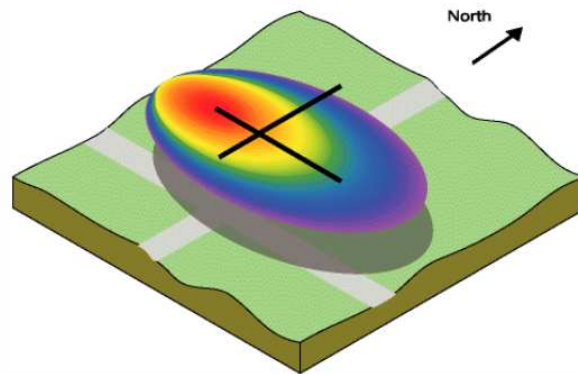


Abbildung 2.3: Globaler Trend (ESRI, 2020c)

Anisotropie

Anisotropie kann in den Daten vorhanden sein, auch wenn kein globaler Trend vorliegt oder dieser bereits entfernt wurde. Im Unterschied zum Trend wird Anisotropie für gewöhnlich als Zufallsprozess in der Autokorrelation mitmodelliert, wobei der eigentliche Grund oft nicht bekannt ist.

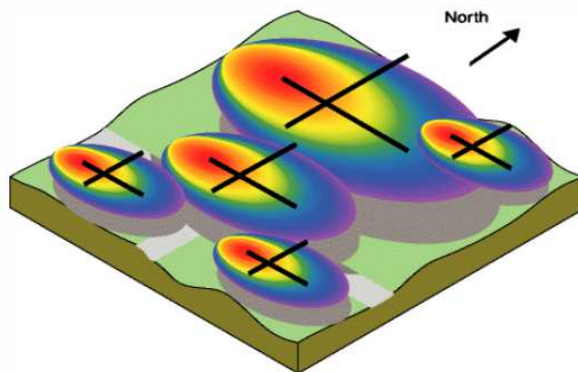


Abbildung 2.4: Richtungsabhängige Einflüsse (ESRI, 2020c)

Der Nachweis von Anisotropie erfolgt durch Einschränkung der Messpunkte nach bestimmten Himmelsrichtungen und Untersuchung der Kovarianz- oder Variogrammfunktion. Ergibt sich eine höhere Variabilität in der Autokorrelation bei bestimmten Himmelsrichtungen, so liegt der Verdacht von Anisotropie nahe. Zur Berücksichtigung von Anisotropie werden Kovarianz-

und Variogrammodelle entsprechend vor der Interpolation angepasst. Auch deterministische Methoden ermöglichen die Berücksichtigung von Anisotropie durch eine entsprechende Gestaltung der Form des Suchradius (z.B. Ellipse).

2.1.2 Datenverteilung

Die Verteilung der Daten im Untersuchungsgebiet ist ein wichtiges Kriterium für eine erfolgreiche Interpolation. Bei der Datenerhebung ist sowohl auf die Aufnahme strategisch wichtiger Punkte zur möglichst genauen Modellierung von prägenden Geländeeigenschaften als auch eine möglichst homogene Verteilung für die genaue Modellierung des Gesamtmodells zu achten. Beispielsweise sind einige Interpolationsmethoden auf den Wertebereich der Eingabedaten beschränkt. Dabei werden keine Punkte außerhalb dieses Bereichs interpoliert, auch wenn es für das zu modellierende Phänomen von Vorteil wäre. Bei ungünstiger Konfiguration der Daten neigen wiederum andere Interpolationsmethoden zu einer starken Überschwingung des Wertebereichs oder erzeugen lokale Maxima und Minima. Statistische Interpolationsmethoden setzen auf ein stochastisches Modell und liefern erst bei einer Normalverteilung der Fehlerresiduen optimale Ergebnisse für Genauigkeitsabschätzungen. In den nächsten Abschnitten werden Grundlagen und Methoden zur Untersuchung von Verteilungen erläutert.

Wahrscheinlichkeitsverteilung

Eine Wahrscheinlichkeitsverteilung (Verteilung einer Zufallsgröße) beschreibt die Angabe der Wahrscheinlichkeit, mit der eine Zufallsgröße bestimmte Werte annimmt. Die Verteilungsfunktion $F(x)$ wird als Summe der Einzelwahrscheinlichkeiten $f(x_i)$ definiert

$$F(x) = \sum_{i: x_i < x} f(x_i) \quad (2.7)$$

bzw. im stetigen Fall

$$F(x) = \int_{-\infty}^x f(t) dt \quad (2.8)$$

wobei $f(x)$ die Dichtefunktion der Verteilung ist (vgl. Navratil, 2006a, S. 83). Wichtige Parameter der Wahrscheinlichkeitsverteilung sind:

Kenngröße	Beschreibung
Erwartungswert	Zahl, welche die Zufallsvariable im Mittel annimmt
α -Quantil	Schwellwert eines Wertebereichs α
Varianz	Maß für die Streuung einer Zufallsvariablen
Schiefe	Symmetrieabweichung von der Normalverteilung

Kurtosis	Gruppierung von Daten um die Mitte der Verteilung
Exzess	Kurtosis vermindert um den Normalwölbungswert 3

Diese Parameter können anhand eines Datensatzes empirisch bestimmt werden.

Normalverteilung

Die Normalverteilung ist eine symmetrische Verteilung und wird durch den Erwartungswert μ und Varianz σ^2 definiert. Die Dichtefunktion ist durch

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x - \mu)^2}{2\sigma^2}} \quad e \dots \text{Euler'sche Zahl} \quad (2.9)$$

gegeben und wird allgemein auch als *Glockenkurve* bezeichnet. Die Parameter Schiefe und Exzess der Normalverteilung haben beiden den Wert 0, die Kurtosis den Wert 3. Der Erwartungswert μ (gleichzeitig der Median) bildet das Zentrum und die Standardabweichung σ gibt den Abstand zu den Wendepunkten an. Allgemein zeigt sich, dass große Abweichungen (vom Erwartungswert) weniger wahrscheinlich als kleine Abweichungen sind (vgl. Navratil, 2006a, S. 89).

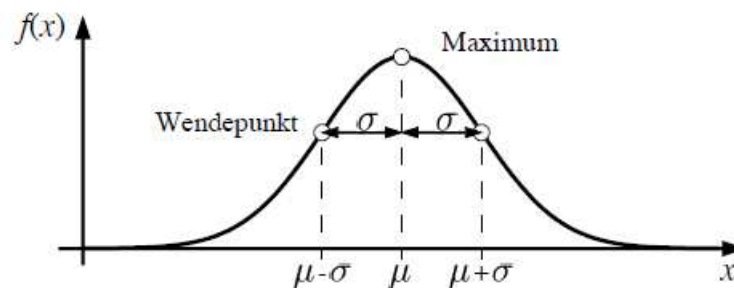


Abbildung 2.5: Normalverteilung und deren Parameter (Niemeier, 2008, S. 36)

Durch Integration der Dichtefunktion $f(x)$ erhält man wiederum die Verteilungsfunktion der Normalverteilung $F(x)$. Zudem wird oft eine Standardisierung der Verteilung vorgenommen:

$$Z = \frac{X - \mu}{\sigma} \quad X \dots \text{Zufallsgröße} \quad (2.10)$$

Für die *Standard-Normalverteilung* gilt dann:

$$\mu = 0 \quad (2.11)$$

$$\sigma^2 = 1 \quad (2.12)$$

Allgemein gilt für eine normalverteilte, eindimensionale Zufallsgröße (vgl. Navratil, 2006a, S. 91)

- im Intervall $\mu \pm 1\sigma$ liegen $\approx 68\%$ aller Realisierungen
- im Intervall $\mu \pm 2\sigma$ liegen $\approx 95\%$ aller Realisierungen
- im Intervall $\mu \pm 3\sigma$ liegen $\approx 99\%$ aller Realisierungen

Bei geostatistischen Verfahren ist speziell darauf zu achten, dass die (beispielsweise durch Kreuzvalidierung bestimmten) Abweichungen zwischen geschätzten und wahren Werten einer Normalverteilung folgen, um vertrauenswürdige Genauigkeitsinformationen zu produzieren.

Statistischer Test auf Normalverteilung

Die Prüfung auf Normalverteilung kann auf Basis eines statistischen Tests erfolgen. Dies wird etwa durch einen Anpassungstest der empirischen Verteilung auf die theoretische Normalverteilung realisiert. Als Prüfgröße kann die χ^2 -Verteilung verwendet werden. Eine χ^2 -Verteilung der Zufallsgröße Y ergibt sich durch Transformation der standard-normalverteilten Zufallsgrößen X_i

$$Y = \sum_{i=1}^m X_i^2 \quad m \dots \text{Freiheitsgrade} \quad (2.13)$$

mit

$$k = m - r \quad k \dots \text{Redundanz} \quad (2.14)$$

Als Prüfgröße wird (vgl. Navratil, 2006a, S. 199)

$$\chi^2 < \sum_{m=1}^r \frac{(h_m - np_m)^2}{np_m} \quad (2.15)$$

verwendet, wobei

$r \dots$ Klassenzahl in die der gesamte Wertebereich der Zufallsgröße eingeteilt wird

$m \dots$ Klassennummer

$h_m \dots$ Empirische, absolute Häufigkeit der m -ten Klasse

$n \dots$ Gesamtanzahl aller Merkmalswerte

$p_m \dots$ Theoretische Wahrscheinlichkeit der m -ten Klasse

Unter Berücksichtigung einer statistischen Sicherheitsgrenze der χ^2 -Verteilung und Redundanz k , kann der entsprechende Wert aus der χ^2 -Verteilung entnommen und mit der berechneten Prüfgröße verglichen werden. Ist die Prüfgröße größer-gleich dem Wert der χ^2 -Verteilung, so ist anzunehmen, dass die Verteilung der Stichprobe von einer Normalverteilung abweicht. Neben diesem Chi-Quadrat-Test existieren aber noch weitere Testverfahren wie (vgl. Dehdari, 2014, S. 62):

- Kolmogorov-Smirnov (KS-)Test: Effektiver als obig beschriebener χ^2 -Test bei der Evaluierung der Güte der Anpassung.
- Lilliefors-Test: Sensitivere Erweiterung des KS-Tests.
- Anderson-Darling: Sensitivere Erweiterung des KS-Tests.
- Cramér-von-Mises-Test: Sensitivere Erweiterung des KS-Tests.
- Shapiro-Wilk Test: Sehr effektiver Test in Bezug auf unterschiedliche Verteilungen und Stichprobengröße.

Visualisierung

Datenvisualisierungen sind für einen ersten Eindruck sehr hilfreich. Die räumliche Struktur der Daten kann etwa in 2D- oder 3D-Darstellungen kartographisch festgestellt werden. Häufig verwendete Werkzeuge sind jedoch das Histogramm und QQ-Plots, welche die rohen Daten ohne direkten Bezug zur räumlichen Verteilung darstellen. Anhand dieser Werkzeuge lassen sich in den Daten oft schon Anomalien wie Ausreißer oder eine Abweichung zur Normalverteilung erkennen.

Histogramm

Histogramme stellen die Häufigkeitsverteilung grafisch dar. Auf der horizontalen Achse werden die Werte in Klassen des Wertebereichs eingeteilt. Auf der vertikalen Achse werden die Klassenbalken entsprechend ihren absoluten und relativen Häufigkeiten aufgetragen.

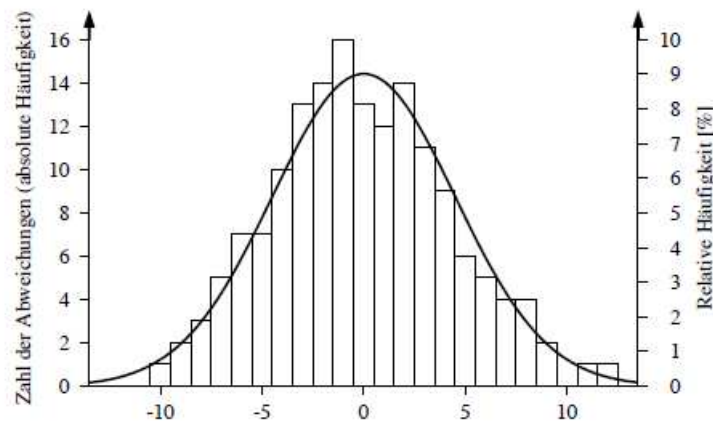


Abbildung 2.6: Histogramm von Messwerten mit Dichtefunktion einer Normalverteilung (Niemeier, 2008, S. 15)

Die Gegenüberstellung zwischen Histogramm und Normalverteilung ist insbesondere für die Analyse der Abweichungen der geschätzten von den wahren Werten sinnvoll. Bei ausreichen-

der Normalverteilung der Residuen kann auf eine ausreichend gute Qualität der geschätzten Genauigkeitsinformationen geschlossen werden.

QQ-Diagramm

Quantil-Quantil-Diagramme sind Visualisierungen, die für die Gegenüberstellung der Quantile zweier statistischer Variablen verwendet werden. Für jede Variable werden auf der horizontalen Achse die geordneten Daten und auf der vertikalen Achse die kumulierte Häufigkeit dargestellt. Anschließend werden jene Datenwerte, die bestimmten Quantilen zuordenbar sind, in einem gemeinsamen Diagramm auf der horizontalen und vertikalen Achse dargestellt.

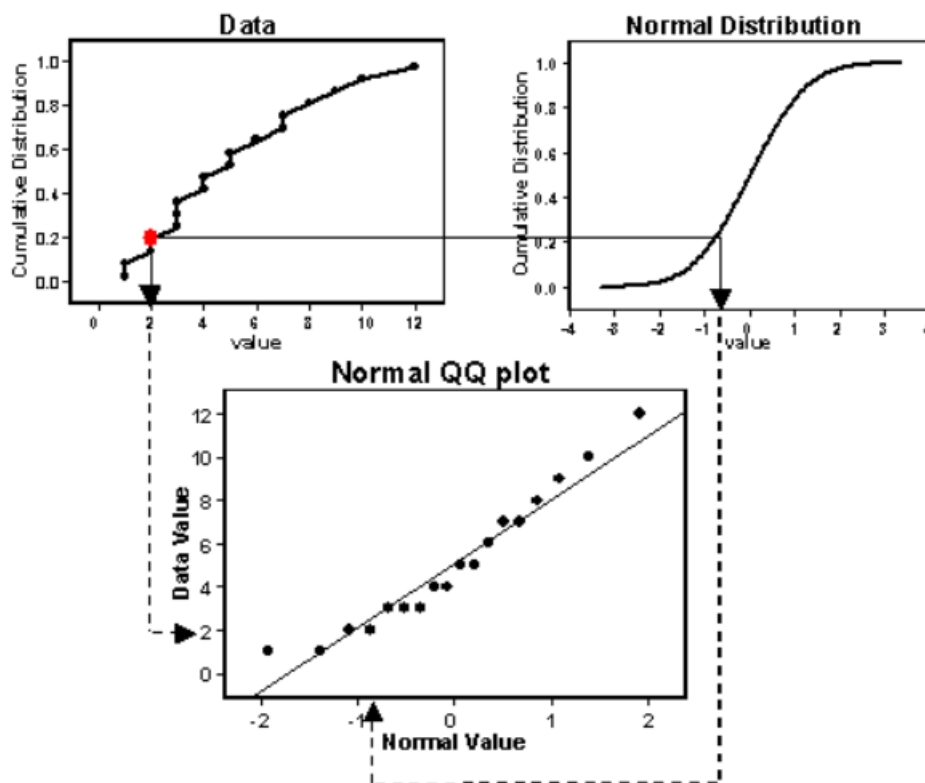


Abbildung 2.7: Erstellung von QQ-Diagrammen (ESRI, 2020f)

Liegen die im QQ-Diagramm dargestellten Punkte auf einer Geraden, so ist dies ein Indiz, dass den Daten dieselbe Verteilung zugrunde liegt. Werden die Verteilungen unterschiedlicher Datensätze verglichen, spricht man von *generellen* QQ-Diagrammen. Wird ein Datensatz mit der Normalverteilung verglichen, spricht man von einem *normalisierten* QQ-Diagramm (vgl. ESRI, 2020f). Letzteres ist ein Mittel, um eine Normalverteilung in den Daten nachzuweisen.

Kenngrößen

Auf Basis empirischer Daten werden Kennwerte empirischer Häufigkeitsverteilungen definiert (vgl. Navratil, 2006a, S. 64-70):

Kenngröße	Definition
Minimalwert	$x_{min} = x_1$
Maximalwert	$x_{max} = x_n$
Arithmetischer Mittelwert	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$
Median	$m = x_{1/2}$
Quantile	$r_\alpha = \alpha(n - 1) + 1$
Modalwert	Häufigster Wert
Spannweite	$\Delta = x_{max} - x_{min}$
Varianz	$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
Kovarianz	$s_{xy}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$
Korrelation	$r_{xy} = \frac{s_{xy}}{s_x s_y}$
Standardabweichung	$s = \sqrt{s^2}$
Variationskoeffizient	$v_x = \frac{s}{ \bar{x} } 100\%$
Zentrierter Beobachtungswert	$x_i - \bar{x}$
Standardisierter Beobachtungswert	$\frac{x_i - \bar{x}}{s}$
Schiefe	$\gamma_1 = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s}\right)^3$
Kurtosis	$\gamma_2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s}\right)^4$
Exzess	$\beta_2 = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{s}\right)^4 - 3$

Diese Kenngrößen erlauben eine quantitative Analyse der gegebenen Daten. Im Rahmen der explorativen Datenanalyse interessiert die Datenverteilung vor allem in Hinblick auf die Detektion von Trends, Ausreißer und räumliche Struktur. Bei der Schätzung von statistischen Genauigkeitsinformation wird auf normalverteilte Residuen geachtet. *Lagekennwerte* (Minimal-, Maximal- und Mittelwert, Median und Quantile) sind im Rahmen der Trend- und Ausreißeranalyse interessant. Ein extremer Minimal- oder Maximalwert deutet auf einen Ausreißer hin. Ein räumlich variabler Mittelwert lässt auf einen allgemeinen Trend oder andere richtungsabhängige Einflüsse schließen. *Streuungskennwerte* (Spannweite, Varianzen, Standardabweichung Kovarianzen und Korrelation) geben Aufschluss über die räumliche Variabilität und Struktur. *Formkennwerte* (zentrierter und standardisierter Beobachtungswert, Schiefe, Kurtosis und Exzess) geben Aufschluss über Symmetrieabweichungen zur Normalverteilung.

Datentransformationen

Für geostatistische Interpolationsverfahren gilt die Annahme der intrinsischen Hypothese. Diese fordert, dass Mittelwert, Varianz und Kovarianz räumlich konstant sind und der räumliche Zusammenhang zweier Zufallsvariablen nur vom Abstand abhängt. Diese Forderungen werden bei vorhandenen Ausreißern, einem allgemeinen Trend oder richtungsabhängigen Einflüssen nicht mehr streng eingehalten. Eine Datentransformation der Messwerte kann zur Minimierung der durch derartige Anomalien entstehenden Effekte genutzt werden. Die Wahl der Transformation ergibt sich anhand der Messdaten und deren Verteilung und kann in den Interpolationsprozess mitaufgenommen werden. Dabei werden die Daten vor Ausgabe der Interpolationsergebnisse zurück-transformiert. Bekannt Transformationen sind (vgl. ESRI, 2020g):

- Box-Cox Transformation
- Log Transformation
- Arcsine Transformation
- Normal Score Transformation

2.1.3 Ausreißeranalyse

Ausreißer sind Beobachtungen, die eine signifikant hohe Abweichung vom Mittelwert aufweisen (vgl. Niemeier, 2008, S. 9). Prinzipiell kann zwischen globalen und lokalen Ausreißern unterschieden werden. Globale Ausreißer unterscheiden sich grob von allen Punkten im Datensatz, wohingegen sich lokale Ausreißer nur in einem bestimmten Gebiet abgrenzen. Ausreißer führen zu einer Verzerrung des räumlichen Modells, was sich besonders bei wenig resistenten Kennwerten (z.B. Mittelwert) auswirkt und das Ergebnis von deterministischen und statistischen Interpolationsverfahren beeinflusst. Im Bereich der Rohstoffsuche, in der auch die statistische Interpolationsform Kriging ihre Ursprünge hat, wird aber oft genau nach solchen Spitzen in den Datensätzen gesucht. Gängige Mittel zur Ausreißeranalyse sind die visuelle Überprüfung, die räumliche Strukturanalyse und statistische Ausreißertests.

Visuelle Überprüfung

Oft können Ausreißer schon durch einen visuellen Abgleich der Verteilung in einem Histogramm erkannt werden. Dieser zeichnet sich durch einen verhältnismäßig großen Maximal- oder Minimalwert mit entsprechendem Abstand vom Mittelwert aus. Nach Abschnitt 2.1.2 liegen 99% aller Realisierungen innerhalb der 3-fachen Standardabweichung. Demnach ist eine Überschreitung dieses Grenzwertes bereits ein starkes Indiz für einen Ausreißer. Histogramme eignen sich gut zur Identifikation von globalen Ausreißern. Lokale Ausreißer fallen hingegen in der allgemeinen Verteilung nicht auf.

Räumliche Strukturanalyse

Im Bereich statistischer Interpolationsmethoden ist die Kovarianzfunktion bzw. das Semivariogramm für das Interpolationsergebnis essenziell. Diese Funktionen beschreiben den räumlichen Zusammenhang aller Messpunkte. Demnach werden alle mit einem globalen Ausreißer in Verbindung gebrachten Messwerte entsprechend hohe Residuen aufweisen und in der Variogrammanalyse auffallen:

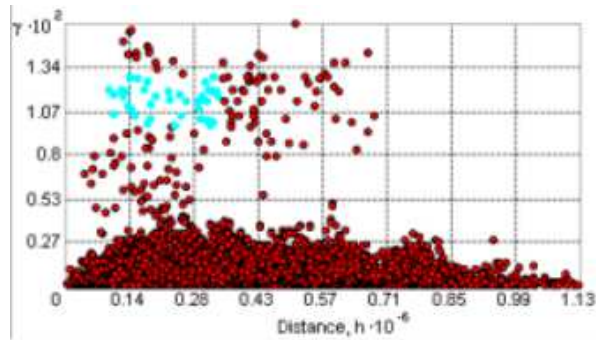


Abbildung 2.8: Globale Ausreißer im Variogramm (ESRI, 2020e)

Lokale Ausreißer lassen sich vor allem durch hohe Residuen bei gleichzeitig geringem Abstand lokalisieren

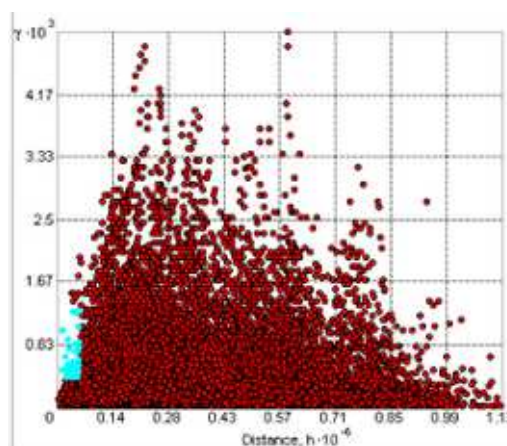


Abbildung 2.9: Lokale Ausreißer im Variogramm (ESRI, 2020e)

Statistischer Ausreißertest

Ein Ausreißertest wird zum Nachweis der statistischen Signifikanz von Ausreißern verwendet. Ein einfacher Test ergibt sich unter Verwendung des Mittelwertes \bar{x} , der Extremwerte

x_{min} , x_{max} und der Standardabweichung s' , wobei Extremwerte nach oben

$$\hat{\Delta} = \frac{x_{max} - \bar{x}}{s'} \quad (2.16)$$

oder nach unten

$$\hat{\Delta} = \frac{\bar{x} - x_{min}}{s'} \quad (2.17)$$

identifiziert werden können, wobei die $\hat{\Delta}$ mit den Sicherheitsgrenzen der Δ -Verteilung abgeglichen werden (vgl. Navratil, 2006a, S. 196/205).

2.2 Räumliche Interpolation

Grundsätzlich versteht man unter räumlicher Interpolation die Schätzung von Werten an räumlichen Positionen, an denen keine Messdaten vorliegen. Die Interpolationsverfahren beziehen sich dabei auf die Annahme eines räumlichen Zusammenhangs. Ziel der Interpolation ist es, aus einer begrenzten Anzahl an diskreten Stützstellen bzw. Messpunkten, den kontinuierlichen Verlauf einer interessierenden Größe abzuleiten. Dabei kann grundsätzlich zwischen deterministischen und statistischen Interpolationsverfahren unterschieden werden. Bei den deterministischen Verfahren sind die Stützpunkte als konkrete Bedingungen gegeben. Statistische Verfahren interpretieren Messdaten als Beobachtungen, denen ein stochastischer Prozess zugrunde liegt. Liegen Höhenwerte als zu interpolierendes Merkmal vor, dann spricht man von einem digitalen Höhenmodell. Ausgangssituation im 2D-Fall sind n in einer XY-Ebene verteilt liegende Punkte P_1, P_2, \dots, P_n mit gemessenem oder gegebenem Höhenwerten. Speziell liegt hier ein 2,5D Modell vor, wo jedem Punkt exakt ein Höhenwert zugeordnet ist (vgl. Niemeier, 2008, S. 410):

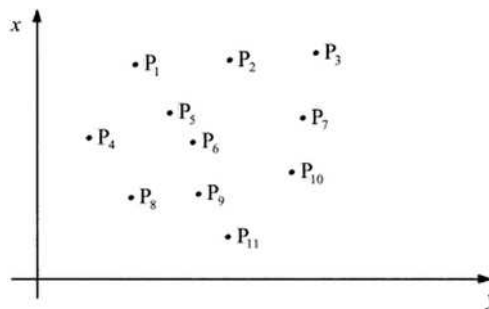


Abbildung 2.10: Räumliche Verteilung von Messwerten (Niemeier, 2008, S. 410)

2.2.1 Deterministische Interpolationsmethoden

Diese Kategorie von Interpolationsverfahren schätzt Punkte mittels mathematischer und geometrischer Modelle und der Verteilung umliegender Punkte (vgl. Niemeier, 2008, S. 410-411). Dabei werden anhand des zu modellierenden Phänomens Interpolationsvorschriften und die Art der Einbeziehung umliegender Punkte definiert.

Interpolation mit Dreiecksvermaschung (TIN)

Grundlage dieses Interpolationsansatzes ist ein unregelmäßiges Netz von Dreiecksflächen, auch *triangulated irregular network* (TIN) genannt. Die Dreiecksflächen werden durch Triangulation bestimmt. Auf Basis der aufgespannten Dreiecksflächen wird anschließend interpoliert. Ein gängiges Verfahren zur Dreiecksvermaschung ist die Delaunay-Triangulation.

Delaunay-Triangulation und Voronoi-Diagramme

Eine Triangulation von Messpunkten muss nicht eindeutig sein, d.h. für ein und dasselbe Gebiet können sich geometrisch unterschiedliche Modelle ergeben. Die korrekte Wiedergabe der Oberfläche durch die Triangulation ist essenziell, da die Interpolation auf Basis der erstellten Dreiecke durchgeführt wird. Je nach Anlage der Dreiecke ergeben sich somit vollständig unterschiedliche Ergebnisse. Das Konstruktionsprinzip der Delaunay-Triangulation beruht darauf, dass jeweils jene drei Punkte ein Dreieck bilden, in deren Umkreis kein weiterer Messpunkt liegt. Das entstehende Modell liefert immer ein eindeutiges und reproduzierbares Ergebnis (vgl. Niemeier, 2008, S. 416):

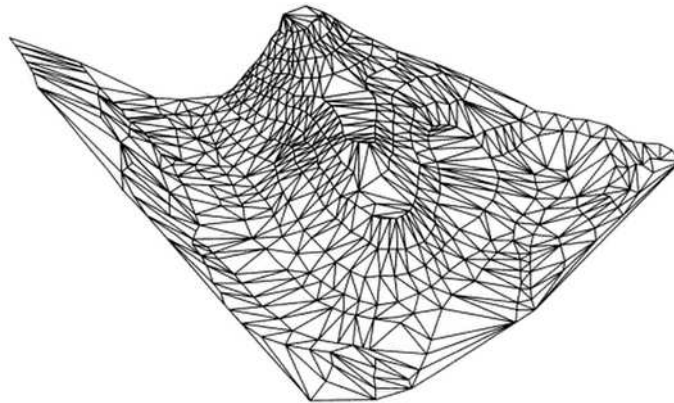


Abbildung 2.11: Dreiecksvermaschung als Basis eines Oberflächenmodells (Niemeier, 2008, S. 418)

Eine fundamentale Eigenschaft der Delaunay-Triangulation ist die Dualität zu den Voronoi-Diagrammen. Dabei wird um jeden Triangulationspunkt jene Region bestimmt, die näher zu diesem Punkt als zu allen umliegenden Punkten liegt:

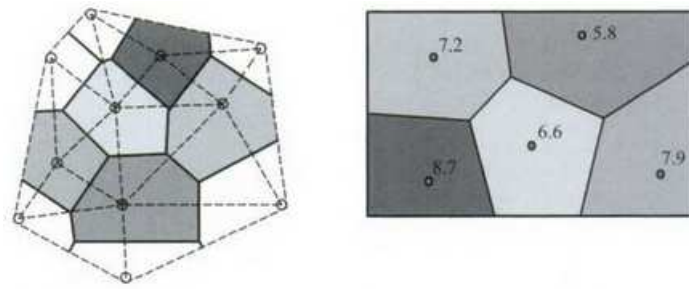


Abbildung 2.12: Delaunay-Triangulation und Voronoi-Diagramme (Niemeier, 2008, S. 420)

Eine Interpolation kann grundsätzlich auf Basis der Dreiecksvermaschung durchgeführt werden (z.B. linear oder kubisch). Voronoi-Diagramme ermöglichen die Berücksichtigung einer räumlichen Gewichtung anhand der Distanz des Interpolationspunktes zu den umliegenden Messpunkten (z.B. nächster Nachbar oder natürlicher Nachbar).

Bilineare-und Bikubische Interpolation

Dieser Ansatz verwendet eine Dreiecksvermaschung der gegebenen Punkte, um eine bivariate Funktion für jedes Dreieck abzuleiten und Interpolationswerte zu bestimmen. Zu den Interpolationsansätzen im Dreieck zählen z.B. der lineare und kubische Ansatz. Beim linearen Ansatz wird die einzelne Dreiecksmasche als Ebene betrachtet (vgl. Niemeier, 2008, S. 417):

$$z(x, y) = a_0 + a_1x + a_2y \quad (2.18)$$

Innerhalb der Masche wird linear durch die Bestimmung der Koeffizienten interpoliert. Diese werden eindeutig über die drei z-Koordinaten der Eckknoten berechnet. Beim kubischen Ansatz ist die einzelne Dreiecksmasche als kubisches Polynom wesentlich komplizierter aufgebaut:

$$z(x, y) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ij}x^i y^j \quad (2.19)$$

Der wesentliche Vorteil dieser Methode ist, dass regionale Besonderheiten wie Gefällenwechsel, Böschungen und Gräben gut miteinbezogen werden und die Interpolation keiner global definierten Interpolationsvorschrift folgt (vgl. Niemeier, 2008, S. 416). Die TIN-Triangulation gehört jedoch trotzdem, besonders bei einfachen linearen Ansätzen, zu den am wenigsten genauen Methoden (vgl. Mitas und Mitasova, 1999, S. 483). Ein wesentlicher Nachteil ist auch, dass nur drei Messpunkte bei der Interpolation miteinbezogen werden. Hinzukommt, dass sich die resultierende Oberfläche an den Kanten der Dreiecke abrupt ändert und so kein glattes Erscheinungsbild zulässt (vgl. Webster und Oliver, 2007, S. 38-39).

Gleitende Mittelbildung

Dieser Interpolationsansatz nützt umliegende Datenpunkte des zu interpolierenden Wertes. Durch gewichtete Mittelbildung der Nachbarn wird die Interpolation des Schätzwertes ermöglicht, wobei der Einfluss umliegender Punkte gesteuert werden kann (vgl. Niemeier, 2008, S. 411).

- **Nächster Nachbar**

Bei dieser Interpolation wird allen Punkten innerhalb eines Voronoi-Polygons der Wert des in der Triangulation verwendeten Messpunktes zugewiesen. Ein großer Nachteil dieser Methode ist, dass die Interpolation nur auf einem Messwert beruht. Dieser Umstand kann speziell bei großen Messfehlern starke Auswirkungen auf das Ergebnis haben. Die resultierende Oberfläche ist generell stufig aufgebaut (vgl. Webster und Oliver, 2007, S. 38) und gibt die Ausgabeoberfläche daher ungenau und unnatürlich wieder.

- **Natürlicher Nachbar**

Dieser Interpolationsansatz sucht zuerst das nächstgelegene Subset an Nachbarn innerhalb eines Voronoi-Diagramms und bestimmt dann den zu interpolierenden Wert. Die Berechnung erfolgt als gewichteter Durchschnitt bezüglich aufgespannter Flächen zwischen dem zu interpolierenden Punkt und den nächstgelegenen Nachbarn. Für die Gewichte gilt (vgl. Webster und Oliver, 2007, S. 39)

$$\lambda_i = \frac{A_i}{\sum_{k=1}^N A_k} \quad i = 1, 2, \dots, N. \quad (2.20)$$

wobei A_k die Teilflächen des Voronoi-Diagramms des zu interpolierten Wertes und A_i die von A_k eingeschlossenen Teilflächen der Voronoi-Diagramme um alle Messwerte sind. Der Interpolationswert z ergibt sich aus einer Linearkombination der Gewichte λ_i mit den Stützpunkten z_i (vgl. Webster und Oliver, 2007, S. 37):

$$z = \sum_{i=1}^n \lambda_i z_i \quad (2.21)$$

Die Interpolation mit natürlichen Nachbarn führt zu einer kontinuierlichen und glatten Interpolationsfläche und wird bevorzugt bei topografischen, bathymetrischen, geophysikalischen und Bodendaten angewandt (vgl. Mitas und Mitasova, 1999, S. 483).

- **Inverse Distance Weighted Interpolation (IDW)**

Bei der IDW-Interpolation wird der Messwert z an den Interpolationspunkten x_0 durch ein gewichtetes Mittel der benachbarten Stützwerte $z(x_i)$ geschätzt (vgl. Niemeier,

2008, S. 411):

$$z(x_0) = \frac{\sum_{i=0}^n \left\{ \frac{1}{d_i^k} z(x_i) \right\}}{\sum \frac{1}{d_i^k}} \quad (2.22)$$

Die Gewichte sind dabei umgekehrt proportional zum Abstand d_i^k zwischen Stützpunkten und Interpolationspunkten. Mit dem Exponenten k wird der Einfluss des Nachbarschaftsbereichs gesteuert. Dieser Ansatz ist einer der einfachsten und verbreitetsten Methoden. Ein Nachteil dieser Methode ist, dass spezielle Konfigurationen der Messwerte (z.B. Clustering) nicht ausreichend berücksichtigt werden und lokale Minima oder Maxima auftreten können ('bull-eye' Effekt). Weiters ist der Interpolationswertebereich bezüglich der Minima und Maxima der Eingabedaten beschränkt. Dies kann dazu führen, dass Geländecharakteristiken bei fehlenden Messdaten nicht abgebildet werden (vgl. Philip und Watson, 1982, S. 205–212). Eine spezielle Variante ist die Methode von Shepard, mit der zusätzliche Bedingungen an den Exponenten k gestellt werden, wodurch sich die Form der Ausgabeoberfläche steuern lässt. Die multiquadratische Interpolation nach Hardy (vgl. Hardy, 1972) ist ebenso ein verwandtes Verfahren. Diese Methode berücksichtigt zusätzlich einen Glättungsfaktor, basiert aber auf stochastischen Konzepten und ist daher eher den statistischen Interpolationsverfahren zuzuordnen (vgl. Navratil, 2006b, S. 170).

Variationsansatz (Splines)

Der Variationsansatz für räumliche Interpolation basiert auf der Definition einer Funktion, für die folgende Eigenschaften Gültigkeit haben (vgl. Mitas und Mitasova, 1999, S. 484):

- Die Funktion soll durch die Datenpunkte, oder nahe an den Datenpunkten vorbei verlaufen

Die Definition einer Funktion durch gegebene Daten wird im zweidimensionalen Fall durch die Polynominterpolation erreicht. Bei diesem Interpolationsverfahren werden Polynome n -ten Grades verwendet (vgl. Niemeier, 2008, S. 412):

$$y = f(x) = a_0 + a_1x + a_2x^2 + \dots + a_kx^k \quad (2.23)$$

Die Darstellung solcher Polynome kann anstatt der Monombasis

$$1, x, x^2, \dots, x^n \quad (2.24)$$

auch anhand anderer Basen, wie Lagrange oder Newton, erfolgen (vgl. Schwarz und Köckler, 2009, S. 95). Die Bestimmung der Koeffizienten ermöglicht die Lösung des Gleichungssystems. Ein bekanntes Problem der Polynominterpolation ist das häufige

auftreten starker Ausschwingungen bei Polynomen höheren Grades und unvorteilhafter Datenkonfiguration (vgl. Navratil, 2006b, S. 158). Eine entscheidende Eigenschaft ergibt sich jedoch bei der Interpolation in höheren Dimensionen. Nach dem Theorem von Mairhuber (1956) gibt es bei der Anwendung von Polynominterpolation bei mehr als zwei Dimensionen immer eine räumliche Verteilung von Daten, für die man kein Polynom (oder Basisfunktion) finden kann, das die Daten interpoliert und aus der gewählten Basis stammt. Die Lösung dafür ist eine Funktion, die das Set aller Polynome (oder Basisfunktionen) jeder Dimension approximiert (vgl. Buhmann und Jaeger, 2019, S. 5).

- *Die Funktion soll so glatt wie möglich sein*

Diese Bedingung wird im Rahmen der Spline-Interpolation durch eine mathematische Funktion zur Minimierung der Krümmung der Oberfläche realisiert. Dieses Verfahren stützt sich auf die Polynominterpolation, verwendet jedoch stückweise Polynome und spezielle Bedingungen, um diese zusammenzufügen. Nach Navratil (2006b, S. 159) wird die eindimensionale Spline-Interpolierende $s(x)$ als Lösung der Variationsaufgabe mit folgenden Nebenbedingungen definiert:

- "1. Die Funktion $s(x)$ muss die Interpolationseigenschaften $s(x_i) = y_i$ erfüllen, also durch die Stützwerte gehen
2. Die Funktion $s(x)$ soll an allen inneren Stützstellen x_1 bis x_{n-1} mindestens einmal stetig differenzierbar sein.
3. Zwischen den Stützstellen soll die Funktion (mindestens) viermal stetig differenzierbar sein.
4. $s(x)$ soll das Integral $E = \frac{1}{2} \int_{x_0}^{x_n} s''(x)^2$ minimieren."

Bei der Spline-Interpolation werden mehrere Polynome niedrigen Grades zu einem Gesamtpolynom zusammengesetzt. Diese Vorgehensweise löst das Problem, dass Polynome hohen Grades dazu neigen stark auszuschwingen (vgl. Niemeier, 2008, S. 414). Die Berechnung der kubischen Spline-Interpolierenden basiert auf kubischen Polynomen:

$$s_i(x) = a_i(x - x_i)^3 + b_i(x - x_i)^2 + c_i(x - x_i) + d_i \quad (2.25)$$

Die Lösung erhält man durch Berechnung der Koeffizienten unter Einhaltung der Variationsbedingungen und Aufstellung der kubischen Polynome der Teilintervalle aus Gleichung 2.25. Es gibt einige Arten von Splines, die bei bestimmten Interpolationsszenarien Vorteile mit sich bringen oder versuchen bestimmten Nachteilen entgegenzuwirken (vgl. kubische Spline, periodische Spline und Akima-Interpolation bei Navratil, 2006b, S. 163-165). Analog zur herkömmlichen Polynominterpolation stößt man auch bei stückweise Polynomen auf das Problem, dass sich in höheren Dimensionen nicht zwangsläufig für jede räumliche Verteilung eine Interpolationsfunktion finden lässt (vgl. Buhmann und Jaeger, 2019, S. 5-6).

Die Einschränkung der Interpolation durch das Theorem von Mairhuber bei mehr als zwei Dimensionen führt zum Konzept der radialen Basisfunktion.

Radiale Basisfunktionen (RBF)

Die Interpolation mittels radialer Basisfunktionen ist eine allgemeinere Methode zur (multivariaten) Interpolation in höheren Dimensionen (vgl. Buhmann und Jaeger, 2019, S. 6). Radiale Basisfunktionen sind reelle Funktionen, bei denen für jedes $x = (x_1, \dots, x_n)$ in \mathbb{R}^n gilt

$$\Phi(x) = \varphi(\|x\|_2) \quad (2.26)$$

und $\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \dots + x_n^2}$ die euklidische Distanz des Punktes x vom Ursprung ist. Die Funktion $\Phi(x)$ hängt also nur von der Distanz jedes Punktes x zum Ursprung ab und ist daher radial-symmetrisch. Die Interpolationsfunktion für eine 'wahre' Funktion f , die an den Datenpunkten bekannt ist, ergibt sich aus einer Linearkombination von radialen Basisfunktionen an allen Datenpunkten x_j

$$F(x) = \sum_j \lambda_j \varphi(\|x - x_j\|_2) \quad (2.27)$$

mit Koeffizienten $\lambda_j \in \mathbb{R}$. Die interpolierende Funktion F soll f an den Datenpunkten k interpolieren, sodass $F(x_k) = f(x_k)$ gilt. Die Koeffizienten λ_j können durch Lösung eines linearen Gleichungssystems gefunden werden

$$\lambda = A^{-1} f \quad (2.28)$$

wobei f die Datenpunkte enthält und für A gilt

$$A = r(|x_i - x_j|)_{i,j} \quad (2.29)$$

Es gibt in Abhängigkeit des zu interpolierenden Phänomens verschiedene Ansätze für die Wahl der RBF mit $r = \|x - x_j\|$. Zusätzlich kann noch ein Form- oder Glattheitsfaktor ε in die Interpolation integriert werden. Bekannte Vertreter sind (vgl. Mitas und Mitasova, 1999, S. 486; Fornberg und Flyer, 2015, S. 80; De Smith et al., 2018, Kapitel 6.6.4):

- Polyharmonische RBF (Spline):

$$\varphi(r) = \varepsilon^n r^n \text{ für } n \text{ ungerade}$$

$$\varphi(r) = \varepsilon^n r^n \ln(\varepsilon r) \text{ für } n \text{ gerade}$$

- Multiquadratische RBF: $\varphi(r) = \sqrt{\varepsilon^2 + r^2}$
- Inverse multiquadratische RBF: $\varphi(r) = \frac{1}{\sqrt{\varepsilon^2 + r^2}}$
- Inverse quadratische RBF: $\varphi(r) = \frac{1}{\varepsilon^2 + r^2}$
- Gaußsche RBF: $\varphi(r) = e^{-(\varepsilon r)^2}$

Analog zu eindimensionalen Splines, stützt man sich bei höheren Dimensionen wiederum auf die Annahme, dass die Interpolationsfunktion so glatt wie möglich durch die Datenpunkte

gehen soll. Dies wird durch die Kombination aus Minimierung der Summe der quadratischen Abweichungen der gemessenen Punkte von den RBFs und einem Glattheitsfaktor erreicht. Dabei wird die beste Approximationsfunktion durch Lösung des folgenden Ansatzes gesucht (vgl. Mitas und Mitasova, 1999, S. 484):

$$\sum_{j=1}^n |z_j - F(r_j)|^2 \lambda_j + \lambda_0 I(F) \rightarrow \min. \quad (2.30)$$

Die Koeffizienten λ_j und λ_0 sind positive Gewichte und $I(F)$ die Glattheits-Halbnorm. Die Glattheits-Halbnorm beschreibt, analog zu Splines, die minimale Biegeenergie. Die Lösung dieser Gleichung gelingt mittels RBFs :

$$F(r) = T(r) + \sum_{j=1}^n \lambda_j \varphi(r_j) \quad (2.31)$$

Dabei ist $T(r)$ eine mögliche Trendfunktion und $\varphi(r_j)$ eine RBF mit r_j als euklidische Distanz zwischen dem geschätzten Ort und dem Datenpunkt. Die Summe ergibt die interpolierende Funktion $F(r)$. Die Funktionswerte an den Stützstellen werden exakt eingehalten. Für die Interpolation von Geländemodellen wurden spezielle RBFs entworfen (vgl. De Smith et al., 2018, Kapitel 6.6.4 und Mitas et al., 2002/3, S. 138):

- Thin Plate Spline: $\varphi(r) = \varepsilon^2 r^2 \ln(\varepsilon r)$
- Thin Plate Spline mit Tension: $\varphi(r) = \ln\left(\frac{\varepsilon r}{2}\right) + l_0\left(\frac{\varepsilon r}{2}\right) + \gamma$
- Regularisierte Thin Plate Spline: $\varphi(r) = \ln\left(\left(\frac{\varepsilon r}{2}\right)^2\right) + E_1\left(\left(\frac{\varepsilon r}{2}\right)^2\right) + \gamma$

mit

l_0 ... modifizierte Besselfunktion

E_1 ... Integralexponentialfunktion

γ ... Eulersche Konstante

Die jeweiligen Interpolationsmethoden ergeben sich in Abhängigkeit der Wahl der Glattheits-Halbnorm (vgl. Mitas und Mitasova, 1999, S. 485). Eine *Thin Plate Spline* wird in zweiter Dimension analog zu den kubischen Splines in erster Dimension definiert. Diese Funktion minimiert die Krümmung der Oberfläche durch Forderung der geringsten Biegeenergie durch die gegebenen Punkte. Dabei wird folgende Glattheits-Halbnorm minimiert:

$$I(F) = \int [F_{xx}^2 + F_{yy}^2 + 2F_{xy}^2] dx \quad (2.32)$$

Ein Nachteil der Thin Plate Spline ist einerseits das Überschwingen der Funktion in Regionen großer Steigung und andererseits das Divergieren der zweiten Ableitungen in den Datenpunkten. Die *Thin Plate Spline mit Tension* löst das Überschwingen durch Einbezie-

hung der ersten Ableitungen und eines Tensionsparameter ϕ in der Seminorm:

$$I(F) = \int [\phi^2 [F_x^2 + F_y^2] + [F_{xx}^2 + \dots]] dx \quad (2.33)$$

Die analytischen Eigenschaften können nochmals mit der *Regularisierten Thin Plate Spline* verbessert werden, wobei diese die Summe aller Ableitungen enthält:

$$I(F) = \int [F_{xx}^2 + \dots] + \tau^2 [F_{xxx}^2 + \dots]] dx \quad (2.34)$$

Wird zusätzlich die Tension ϕ berücksichtigt, ergibt sich die Halbnorm zu:

$$I(F) = \sum_{mn} c_{mn}(\phi) \int [F_{xy}^{nm}]^2 dx \quad (2.35)$$

Die Parameter der Splines können empirisch durch das zu modellierende Phänomen oder automatisch auf Basis der Minimierung von Prädiktionsfehlern durch Kreuzvalidierung gewählt werden. Die Berechnung der Interpolation bei großen Datensätzen stellt sich bei diesem Verfahren jedoch als aufwendig heraus. Dem kann durch Implementierung eines automatischen Segmentierungsprozesses entgegengewirkt werden. Der Variationsansatz kann zudem um einige Eigenschaften erweitert werden (vgl. Mitas und Mitasova, 1999, S. 486):

- *Partielle Splines* beziehen zusätzliche Variablen ein.
- *Modellerweiterungen* beziehen topographische Features ein (z.B. ANUDEM-Algorithmus).
- *Räumlich variable Tension* berücksichtigt unterschiedliche Geländeeigenschaften.
- *Anisotropie* berücksichtigt richtungsabhängige Einflüsse durch Erweiterung des Tensionsparameters.

Diese Erweiterungen des gewöhnlichen Variationsansatzes finden ihre Analogien in erweiterten geostatistischen Ansätzen (siehe Abschnitt 2.2.2).

ANUDEM

ANUDEM ist ein von Hutchinson (1989) entwickelter Algorithmus. Das Verfahren wurde für die Erstellung einer Oberfläche entworfen, die eine natürliche Drainageoberfläche genauer darstellt. Dabei werden sowohl Bergkämme als auch Wasserlaufnetze durch Eingabekonturliniendaten besser beibehält. Die Interpolationsmethode basiert auf der Methode der iterativen finiten Differenzen. Der Vorteil liegt darin, dass die Effizienz lokaler Interpolationsmethoden (Natürlicher Nachbar oder IDW) mit der Aufrechterhaltung einer zusammenhängenden Oberfläche der globalen Methoden (z.B. Kriging oder Spline) kombiniert wird. Dabei wird eine gekapselte Thin Plate Spline-Interpolation durchgeführt. Die Thin Plate Spline-Interpolation ist ein Variationsansatz und basiert auf Gleichung 2.30. Bei der ANUDEM Methode werden die Parameter der Thin Plate Spline modifiziert, um die Wiedergabe von abrupten Geländeübergängen, wie Wasserläufen, Bergrücken und Klippen, möglichst genau

zu berücksichtigen (vgl. ESRI, 2020a).

2.2.2 Statistische Interpolation

Statistische Interpolationsmethoden verwenden stochastische Modelle zur Schätzung von Interpolationswerten. Dabei werden die Messwerte als Realisierungen von Zufallsvariablen (Beobachtungen) mit zugehörigen Varianz- und Kovarianzinformationen aufgefasst. Ein wesentlicher Schritt für möglichst optimale Ergebnisse ist die Analyse und Ableitung von Korrelationen im Untersuchungsgebiet. Bei der Interpolation nach der Methode der kleinsten Quadrate wird dies durch die Kovarianzfunktion erreicht. Beim Kriging erfolgt die Evaluierung durch das Variogramm (siehe Abschnitt 2.1.1). Für diese Modelle gilt die Annahme der (intrinsischen) Stationarität. Dabei wird gefordert, dass ein Erwartungswert und eine Varianz existieren, die unabhängig vom Ort sind. Zusätzlich darf die Varianz der Differenz zweier Zufallsvariablen nur von der (horizontalen) Entfernung abhängen und es sollte kein systematischer Trend in den Daten vorhanden sein. Ein Trend ist ein richtungsabhängiger Einfluss und lässt sich in geschlossener Form analytisch durch Regressionsansätze modellieren.

Regression

Diese Methodik befasst sich mit der Evaluierung von funktionalen Zusammenhängen zwischen Variablen. Sofern eine funktionale Abhängigkeit vorliegt, erfolgt die Regression durch die Wahl eines approximierenden Funktionstyps und der anschließenden Parameterschätzung der Funktion. Der allgemeine Ansatz der linearen Regression ergibt sich aus (vgl. Niemeier, 2008, S. 382)

$$y = Xb + \varepsilon \quad (2.36)$$

wobei

- $y \dots$ Beobachtungen
- $b \dots$ Parameter
- $X \dots$ Designmatrix
- $\varepsilon \dots$ Residuenvektor

und stochastischem Modell

$$\sigma_{yy} = \sigma_{\varepsilon\varepsilon} = \sigma_0^2 I \quad (2.37)$$

Dieser Ansatz ist die Grundlage des Kollokationsansatzes aus Abschnitt 2.2 mit dem Unterschied, dass in diesem Fall kein Signal mitmodelliert wird. Je nach Anwendungsfall stehen unterschiedliche Regressionsmodelle zur Verfügung. *Lineare Regression* bezeichnet lineare funktionale Zusammenhänge bezüglich der zu schätzenden Parameter b_i (vgl. Niemeier, 2008, S. 381). Ein einfacher Ansatz zur Lösung von 2.36 ist durch eine Geradengleichung gegeben:

$$y = b_0 + x_1 b_1 + \varepsilon_i \quad (2.38)$$

Man spricht von *multipler linearer Regression*, wenn es mehrere Einflussparameter b_i gibt

$$y = b_0 + x_1 b_1 + x_2 b_2 + \dots + x_n b_n + \varepsilon_i \quad (2.39)$$

wobei für jeden Messwert y_i ein Residuum ε_i vorliegt. Unter Annahme gleichgenauer und unkorrelierter Messwerte, wird die Kovarianzmatrix zur Einheitsmatrix mit den Varianzen in der Diagonalen:

$$\Sigma_{yy} = \Sigma_{\varepsilon\varepsilon} = \sigma_0^2 I \quad (2.40)$$

Dabei ergibt sich als Lösung für den Parametervektor durch die Ausgleichung:

$$\hat{b} = (X^T X)^{-1} X^T y \quad (2.41)$$

Das Ergebnis der Regression wird allgemein als Trend bezeichnet. Die Trendfunktion wird auf Basis einer globalen Ansatzfunktion (dem Regressionsmodell) abgeleitet. Derartige Modelle werden häufig vor der eigentlichen Interpolation in Form einer Trendanalyse eingesetzt (vgl. Abschnitt 2.1.1) um das Kriterium der Stationarität zu erfüllen. Unter der Annahme, dass sich nahe beieinander liegende Punkte ähnlicher sind als weiter auseinander liegende Punkte, kann man den Messwerten und Interpolationspunkten eine räumliche Korrelation unterstellen. Diese kann unter Einführung einer Kovarianzmatrix bzw. Gewichtsmatrix für die Messwerte modelliert werden. Liegt die Annahme eines nicht-linearen Zusammenhangs nahe, spricht man von *Nichtlinearer Regression*. Bei diesem Ansatz werden polynomiale, trigonometrische oder hyperbolische Funktionen verwendet, die jedoch im Zuge der Regression linearisiert werden müssen.

Interpolation nach der Methode der kleinsten Quadrate

Diese Interpolationsmethode, auch lineare Regression genannt, schätzt die zu interpolierenden Werte \hat{s}_P an den Positionen P anhand korrelierter Beobachtungen z_i an den Positionen i . Der Interpolationsansatz geht von der Annahme aus, dass sich das Signal \hat{s}_P über eine Linearkombination bestimmen lässt. Dabei handelt es sich um einen linearen Regressionsansatz (vgl. Niemeier, 2008, S. 423):

$$\hat{s}_P = a_{p1} z_1 + a_{p2} z_2 + \dots + a_{pn} z_n \quad (2.42)$$

Der interpolierte Wert \hat{s}_P entspricht allein dem Signalanteil. Die Gewichte sollen nur vom Abstand abhängen. Die Herleitung der Methode kann über die Abweichung ε_P zwischen 'wahrem' Signal s_P und dem zu interpolierenden Wert \hat{s}_P erfolgen:

$$\varepsilon_P = s_P - \hat{s}_P = s_P - \sum_{i=1}^n a_{pi} z_i \quad (2.43)$$

Zu beachten ist, dass in diesem Modell kein allgemeiner Trend berücksichtigt wird. Sofern ein Trend vorhanden ist, müssen die Messwerte vor Durchführung der Interpolation um diesen reduziert werden (vgl. Kraus, 1998, S. 45). Für den wahren Wert s_P wird eine Varianz C_0 und Kovarianz C_{pi} zu den tatsächlich gemessenen Werten eingeführt. Weiters wird eine Kovarianzmatrix C_{ik} der gemessenen Werten definiert. Durch Anwendung des Fehlerfortpflanzungsgesetzes ergibt sich:

$$\sigma_P^2 = C_0 - 2 \sum_{i=1}^n a_{pi} C_{pi} + \sum_{i=1}^n \sum_{k=1}^n a_{pi} a_{pk} C_{ik} \quad (2.44)$$

Die Koeffizienten a_{pk} erhält man durch Minimierung der Varianz der Abweichung ε nach Anwendung des Fehlerfortpflanzungsgesetzes:

$$\frac{\partial(\sigma_P^2)}{\partial_{pi}} = 0 \quad (2.45)$$

woraus folgt

$$\sum_{k=1}^n a_{pk} C_{ik} = C_{pi} \quad (2.46)$$

bzw.

$$a_{pk} = \sum_{i=1}^n C_{ik}^{-1} C_{pi} \quad (2.47)$$

Die Schätzung der Signale ergibt sich dann zu

$$\hat{s}_P = \sum_{k=1}^n \sum_{i=1}^n C_{ik}^{-1} C_{pi} z_i \quad (2.48)$$

bzw. in Matrizenform

$$\hat{s}_P = \begin{pmatrix} C_{p1} & C_{p2} & \dots & C_{pn} \end{pmatrix} \begin{pmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} \\ \vdots & \vdots & & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} \end{pmatrix}^{-1} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix} \quad (2.49)$$

oder

$$\hat{s}_P = C_P^T C_{ij}^{-1} z \quad (2.50)$$

mit C_P als Kovarianzvektor zwischen Interpolationswerten und Messwerten sowie C_{ij} als Kovarianzmatrix der Messwerte. Die Elemente von C_P und C_{ij} werden durch eine Kovarianzfunktion (vgl. Abschnitt 2.1.1) ermittelt. Für die gleichzeitige Interpolation an mehreren Stellen wird C_P zur Kovarianzmatrix erweitert. Zur Berücksichtigung des zufälligen Rauschens kann zusätzlich eine Diagonalmatrix D mit den entsprechenden Varianzen angesetzt

und zu C_{ij} addiert werden (vgl. Niemeier, 2008, S. 426):

$$\bar{C} = C_{ij} + D \quad (2.51)$$

Daraus ergibt sich die Prädiktion zu

$$\hat{s}_P = C_P \bar{C}^{-1} z \quad (2.52)$$

mit Prädiktionsvarianz

$$\sigma_p^2 = C_0 - C_P^T C_{ij}^{-1} C_P \quad (2.53)$$

Die Interpolation nach der Methode der kleinsten Quadrate liefert Interpolationswerte mit Varianzinformationen. Im Unterschied zum Kriging (vgl. Abschnitt 2.2.2) ist die Anforderung im Allgemeinen nicht, dass die Interpolationsfläche exakt durch die Datenpunkte verläuft. Das Ziel bei der Interpolation nach kleinsten Quadraten ist die Minimierung der Summe der quadratischen Abweichungen der geschätzten Punkte von den wahren Werten. Die Eigenschaft der Erwartungstreue, also der Verlauf der Interpolationsfläche durch die Datenpunkte, kann mittels einer Trendabspaltung im gleichen Rechenschritt mit der Prädiktion herbeigeführt werden. Diese erweiterte Interpolation wird als Kollokation bezeichnet (vgl. Kraus, 1998, S. 47). Eine detaillierte Betrachtung der Kollokation findet man bei Navratil, 2006b, S. 171).

Multiquadratische Interpolation nach Hardy

Die multiquadratische Interpolation nach Hardy (1972) ist mit der Kollokation verwandt. Das stochastische Modell ist bei der multiquadratischen Interpolation jedoch vereinfacht und behandelt nur den Trend und das Rauschen (vgl. Navratil, 2006b, S. 183). Das Verfahren beruht auf der Grundlage einer Interpolationsfläche, die durch Überlagerung von n Flächen 2. Grades (Hyperboloiden) entsteht. Die Idee ist, dass jede unregelmäßige, glatte Fläche durch viele kleine regelmäßige Flächen approximiert werden kann. Dies ist auch das Grundkonzept radialer Basisfunktionen, wo multiquadratische Ansätze beschrieben sind (vgl. Abschnitt 2.2.1). Zur Berechnung der multiquadratischen Interpolation wird eine Kernfunktion mit Glättungsfaktor G verwendet. Die Werte der Kernfunktion werden in der Funktionalmatrix $C = [s_{ij}]_{n,n}$ für alle Stützstellen zusammengefasst. Die gesuchten Höhen z_P im Interpolationspunkt P ergeben sich durch das Gleichungssystem (vgl. Bretterbauer und Schuh, 2013, S. 15-16):

$$z_P = z_m + c^T C^{-1} dz \quad \text{mit} \quad z_m \dots \text{Mittelwert} \quad (2.54)$$

$$C = \begin{bmatrix} \sqrt{G} & s_{12} & \dots & s_{1n} \\ s_{21} & \sqrt{G} & \dots & s_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \dots & \sqrt{G} \end{bmatrix} \quad (2.55)$$

$$c^T = [s_{P1}, s_{P2}, \dots, s_{Pn}] \quad (2.56)$$

$$s_{ij} = [(x_j - x_i)^2 + (y_j - y_i)^2 + G]^{\frac{1}{2}} \quad (2.57)$$

$$dz_i = z_i - z_m \quad (2.58)$$

Dabei bezeichnet z_m den Mittelwert aller n gegebenen Funktionswerte z_i . Der Vektor $c = [s_i]_{n,1}$ wird aus den Distanzen des Interpolationspunktes zu allen Stützpunkten (mit Berücksichtigung von G) gebildet. Nach Göpfert (1977) wird ein Wert für $G = 0.6 \cdot s_{min}^2$ vorgeschlagen, wobei s_{min} den kleinsten Abstand zwischen den Stützstellen beschreibt.

Kriging

Diese Interpolationsmethode minimiert, analog zur Interpolation nach kleinsten Quadraten, den mittleren quadratischen Fehler der räumlichen Prädiktion und wird allgemein als 'Best Linear Unbiased Estimator' (BLUE-Schätzer) bezeichnet. *Linear* bedeutet, dass die Schätzfunktion linear ist und *Unbiased* bedeutet, dass Erwartungstreue gefordert ist. Nach Kraus (1998, S. 45) sind Kriging und die Interpolation nach kleinsten Quadraten bis auf die Forderung der Erwartungstreue ident.

Interpolationsansatz

Der Wert einer Prädiktion $Z(x_0)$ auf Basis der Daten $Z = (Z(x_1), \dots, Z(x_n))$ wird mit Hilfe eines Prädiktors $p(Z; x_0)$ bestimmt. Weiters wird eine Verlustfunktion (loss function) definiert, welche die Abweichung zwischen wahren und geschätztem Wert beschreibt (vgl. Navratil, 2006b, S. 203):

$$L(Z(x_0), p(Z; x_0)) \quad (2.59)$$

Prädiktoren werden als *optimal* bezeichnet, wenn sie das Bayes'sche Risiko

$$E\{L(Z(x_0), p(Z; x_0))\} \quad (2.60)$$

minimieren. Der Prädiktor kann linear oder nicht-linear angesetzt werden. Ein Ansatz für

einen linearen Prädiktor ist

$$p(Z; x_0) = k + \sum_{i=1}^n l_i Z(x_i) \quad (2.61)$$

und kann wieder als gewichteter Mittelwert aller Nachbarwerte aufgefasst werden, wobei an die Bestimmung der Gewichte l_i statistische Bedingungen geknüpft sind.

Simple Kriging

Diese Form des Kriging erhält man durch Minimierung des Bayes'schen Risikos:

$$E(Z(x_0) - p(Z; x_0))^2 \quad (2.62)$$

Der beste lineare Prädiktor ergibt sich dann zu

$$p(Z; x_0) = c^T \sum^{-1} (Z - \mu) + \mu(x_0) \quad (2.63)$$

mit Kovarianzen $c = (C(x_0, x_1), \dots, C(x_0, x_n))^T$ zwischen Interpolations- und Messwerten, der Matrix \sum mit Kovarianzen $C(x_i, x_j)$ zwischen den Messpunkten sowie den Mittelwerten $\mu = (\mu(x_1), \dots, \mu(x_n))^T$. Der minimalisierte mittlere quadratische Prädiktionsfehler ergibt sich zu:

$$\sigma_{SK}^2 = C(x_0, x_0) - c^T \sum^{-1} c \quad (2.64)$$

Diese Form des Kriging setzt das Wissen der konstanten Mittelfunktion $\mu(\cdot)$ im Untersuchungsgebiet voraus, die durch vorhandene Messungen geschätzt werden kann. Oft ist die Annahme, über die Mittelfunktion genau Bescheid zu wissen, jedoch unrealistisch. Dies wäre beispielsweise der Fall, wenn Residuen einer Trendoberfläche vorliegen und der Mittelwert mit Null angenommen werden kann. *Ordinary Kriging* und *Universal Kriging* betrachten die Mittelfunktion als linear in einer fixen Anzahl von unbekanntem Parametern. Dabei werden die Parameter optimal bestimmt, was aber zu höheren Prädiktionsfehlern wie beim Simple Kriging führt. Im Unterschied zu Ordinary und Universal Kriging, ist Simple Kriging nicht erwartungstreu (vgl. Navratil, 2006b, S. 203).

Ordinary Kriging

Diese Kriging-Methode wird durch die Einführung zweier Bedingungen definiert (vgl. Navratil, 2006b, S. 204)

- Modell-Annahme (Mittelwert ist für alle Werte gleich):

$$Z(x) = \mu + \delta(x) \quad (2.65)$$

- Prädiktions-Annahme (Erwartungstreue):

$$p(Z; x_0) = \sum_{i=1}^n \lambda_i Z(X_i) \quad \sum_{i=1}^n \lambda_i = 1 \quad (2.66)$$

Diese Methode geht davon aus, dass der konstante Mittelwert μ unbekannt ist. Die Minimierung von

$$E(Z(x_0) - p(Z; x_0))^2 \quad (2.67)$$

mit der Nebenbedingung der Erwartungstreue erfolgt mittels Lagrange'scher Multiplikatoren m . Als Ergebnis ergibt sich das Gleichungssystem (vgl. Navratil, 2006b, S. 204)

$$\begin{pmatrix} \gamma(x_1 - x_1) & \dots & \gamma(x_1 - x_n) & 1 \\ \gamma(x_2 - x_1) & \dots & \gamma(x_2 - x_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_n - x_1) & \dots & \gamma(x_n - x_n) & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ m \end{pmatrix} = \begin{pmatrix} \gamma(x_0 - x_1) \\ \gamma(x_0 - x_2) \\ \vdots \\ \gamma(x_0 - x_n) \\ 1 \end{pmatrix} \quad (2.68)$$

bzw.

$$\Gamma_0 \lambda_0 = \gamma_0 \quad (2.69)$$

und aufgelöst nach λ_0

$$\lambda_0 = \Gamma_0^{-1} \gamma_0 \quad (2.70)$$

Der minimierte mittlere quadratische Prädiktionsfehler für das Ordinary Kriging ergibt sich zu

$$\sigma_k^2 = \sigma_0^T \gamma_0 \quad (2.71)$$

Die mit Formel 2.70 geschätzten Werte beinhalten jedoch noch Störeinflüsse. Dies ergibt sich aus Annahme 2.65, da sich $\delta(x)$ aus einem stationären Prozess $\eta(x)$ und dem Messrauschen, bzw. Nugget-Effekt $\varepsilon(x)$ ergibt. Richtiger ist es eine messfehlerfreie Prädiktion

$$Z(x_0) = S(x_0) + \varepsilon(x_0) \quad (2.72)$$

mit

$$S(x_0) = \mu + \eta(x) \quad (2.73)$$

zu bestimmen. Der Prädiktor ergibt sich dann aus:

$$p(S; x_0) = \sum v_i Z(x_i) \quad \text{mit} \quad \sum v_i = 1 \quad (2.74)$$

Nach Minimierung des Bayes'schen Risikos ergibt sich (vgl. Navratil, 2006b, S. 206):

$$v_0 = \Gamma_0^{-1} \gamma_0^* \quad (2.75)$$

Dabei ist Γ_0 gleich wie bei 2.70 und für die anderen Größen gilt

$$v_0 = (v_1, \dots, v_n, m)$$

$$\gamma_0^* = (\gamma(x_0 - x_1), \dots, \gamma(x_0 - x_n), 1) \quad \text{wenn } x_0 \neq x_i$$

$$\gamma_0^* = \sigma_F^2 \quad \text{wenn } x_0 = x_i$$

mit mittlerem Prädiktionsfehler

$$\tau_k^2 = \sum_{i=1}^n v_i \gamma^*(x_0 - x_i) + m - \sigma_I^2 \quad (2.76)$$

wobei σ_F der Fehleranteil des stationären Prozesses und σ_I der Fehleranteil des Messrauschens ist. Verwendet man den Prädiktor $p(Z; x_0)$, bekommt man die exakten Messwerte geschätzt. Bei Verwendung von $p(S; x_0)$ erhält man geglättete Werte. Ordinary Kriging setzt analog zum Simple Kriging voraus, dass die Mittelfunktion μ konstant ist. Das muss in der Realität aber nicht unbedingt der Fall sein. Ein Vorteil ist jedoch, dass die Mittelfunktion nicht bekannt sein muss, was einen flexiblen Einsatz der Methode ermöglicht.

Universal Kriging

Diese Kriging-Methode wird bei gleichzeitiger Berücksichtigung eines (globalen) Trends oder (lokalen) Drifts verwendet (z.B. Windrichtung). Diese Trend- und Drift-Einflüsse können durch eine deterministische Funktion, wie z. B. einem Polynom, modelliert werden. Im Ansatz erweitert diese Methode die Modell-Annahme 2.65 von Ordinary Kriging durch die Annahme (vgl. Navratil, 2006b, S. 206)

$$Z(x) = \sum_{j=1}^{p+1} f_{j-1}(x) \beta_{j-1} + \delta(x) \quad (2.77)$$

bzw. in Vektorschreibweise

$$Z = X\beta + \delta \quad (2.78)$$

Die Summe ist eine Linearkombination aus einer Anzahl von p bekannten Funktionen f mit unbekanntem Parametern β , die den Mittelwert beschreibt. Für eine lineare Prädiktion erfolgt der Ansatz

$$p(Z; x_0) = \sum_{i=1}^n \lambda_i Z(s_i) \quad \text{mit} \quad \lambda^T X = x^T \quad (2.79)$$

wobei $\lambda^T X = x^T$ die Erwartungstreue sichert. Als Lösung der Minimierung des Bayes'schen Risikos ergibt sich die Gleichung (vgl. Navratil, 2006b, S. 207)

$$\lambda_u = \Gamma_u^{-1} \gamma_u \quad (2.80)$$

mit

$$\begin{aligned} \lambda_u &= (\lambda_1, \dots, \lambda_n, m_0, \dots, m_p) \\ \gamma_u &= (\gamma(x_0 - x_1), \dots, \gamma(x_0 - x_n), 1, f_1(x_0), \dots, f_p(x_0)) \\ \Gamma_u &= \begin{cases} \gamma(x_i - x_j), & i = 1, \dots, n & j = 1, \dots, n \\ f_{j-i-n}(x_i), & i = 1, \dots, n & j = n + 1, \dots, n + p + 1 \\ 0, & i = n + 1, \dots, n + p + 1 & j = n + 1, \dots, n + p + 1 \end{cases} \end{aligned}$$

und mittlerem Prädiktionsfehler

$$\sigma_{ku}^2 = \lambda_u^T \gamma_u \quad (2.81)$$

Ordinary Kriging kann zwar bei vorhandenem Trend eingesetzt werden, liefert dann aber keine optimalen Ergebnisse. Bei Universal Kriging muss der Mittelwert im Untersuchungsgebiet nicht mehr konstant sein, da der Trend mitmodelliert wird. Die Evaluierung des Trends erfolgt gewöhnlicherweise vor Durchführung der Interpolation in einer Trendanalyse des Untersuchungsgebiets.

Weiterführende Kriging-Methoden

Es gibt eine Reihe von Erweiterungen, welche es ermöglichen weiterführende Problemstellungen zu lösen:

- *Cokriging* bezieht Information mehrerer korrelierter Werte für die Verbesserung der Qualität der Interpolation mit ein.
- *Robustes Kriging* schätzt korrekte Werte, auch wenn es sich nicht um einen normalverteilten Prozess handelt. Es wird ein robuster Schätzer für die Ermittlung des Variogramms verwendet.
- *Wahrscheinlichkeits-, Indikator und Disjunktives Kriging* werden eingesetzt, wenn die Wahrscheinlichkeit, dass ein gemessener Wert eine gewisse Grenze überschreitet, von Interesse ist.
- *Empirical Bayesian Kriging* berechnet die Modellparameter anhand von Teilmengenbildung und Simulationen anhand eines Maximum Likelihood-Verfahrens.
- *Block-Kriging* wird für die Aufteilung des Schätzproblems in Blockform verwendet. Dabei wird mit Mittelwerten für diese Blöcke anstelle der Originalwerte gearbeitet.
- *Spatio-temporal Kriging* wird eingesetzt, wenn Aussagen über die Entwicklung eines modellierten Phänomens über die Zeit getroffen werden müssen.

2.2.3 Beurteilung der Interpolationsqualität

Die Qualität von Interpolationsergebnissen lässt sich daran messen, wie gut stochastische Bedingungen erfüllt werden und wie genau Interpolationsmodelle die Realität annähern.

Gängige Werkzeuge für die Qualitätsbeurteilung sind daher ein stochastisches Modell und die Kreuzvalidierung:

- Stochastisches Modell

Bei einem stochastischen Modell werden Genauigkeitsinformationen aus der räumlichen Datenverteilung aller Messpunkte für jeden Interpolationspunkt berechnet. Fehler bei der Modellierung der zugrundeliegenden Autokorrelation übertragen sich jedoch direkt auf die geschätzten Genauigkeitsinformationen. Aus diesem Grund ist nach Anwendung eines statistischen Interpolationsverfahrens jedenfalls auch eine Kreuzvalidierung durchzuführen. Die Erstellung von Konfidenzbereichen ist auf Basis einer einzigen geostatistischen Interpolation möglich, liefert aber nur bei einer Normalverteilung der Fehlerresiduen zuverlässige Ergebnisse. Dies kann wiederum im Zuge einer Kreuzvalidierung überprüft werden.

- Kreuzvalidierung

Bei der Kreuzvalidierung wird anhand des gewählten Modells immer ein Messpunkt aus den Daten entfernt und an derselben Stelle ein Schätzwert interpoliert. Aus den Residuen können Statistiken abgeleitet werden, die Aufschluss über die Qualität der Interpolation geben. Die Kreuzvalidierung liefert eine Einschätzung darüber, wie gut das Interpolationsmodell die Referenzpunkte annähert und erlaubt so eine allgemeine Beurteilung der Qualität des Interpolationsmodells. Im Unterschied zum stochastischen Modell werden keine Genauigkeiten für die Interpolationspunkte geschätzt. Konfidenzbereiche können anhand der Abweichungen zwischen Messwerten und geschätzten Werten berechnet werden.

Folgende Metriken werden oft zur Qualitätsbeurteilung von Interpolationsergebnissen herangezogen (vgl. ESRI, 2020b):

Kennwert	Definition
Mittlere Abweichung (Mean Error, ME)	$\frac{\sum_{i=1}^n (Z(s_i) - z(s_i))}{n}$
Wurzel der mittleren quadratischen Abweichung (Root Mean Square Error, RMSE)	$\sqrt{\frac{\sum_{i=1}^n (Z(s_i) - z(s_i))^2}{n}}$
Durchschnittliche Standardabweichung (Average Standard Error, ASE)	$\sqrt{\frac{\sum_{i=1}^n \sigma(s_i)^2}{n}}$

Kennwert	Definition
Mittlere standardisierte Abweichung (Mean Standardized Error, MSE)	$\frac{\sum_{i=1}^n (Z(s_i) - z(s_i)) / \sigma(s_i)}{n}$
Wurzel der mittleren quadratischen standardisierten Abweichungen (Root Mean Square Standardized Error, RMSSE)	$\sqrt{\frac{\sum_{i=1}^n [(Z(s_i) - z(s_i)) / \sigma(s_i)]^2}{n}}$
Allgemeine Anpassung (Overall Fitness)	$1 - \frac{\sum_{i=1}^n (Z(s_i) - z(s_i))^2}{\sum_{i=1}^n (Z(s_i) - \bar{z})^2}$

Im Bereich der Flugsicherung ist sowohl ein qualitätsgesichertes Gesamtmodell als auch die lokale Qualitätsbeurteilung von Bedeutung.

Globale Qualitätsbeurteilung

Eine für das Untersuchungsgebiet global geltende Qualitätsbeurteilung kann beispielsweise bei Start- und Landemanövern, sowie bei Flugsimulationen nötig sein. Für diese Anwendungsfälle wird ein möglichst realitätsnahes Gesamtmodell benötigt. Ein sich darauf stützendes Geländewarnsystem sollte dabei nicht zu früh, aber schon gar nicht zu spät alarmieren. Die aussagekräftigsten Indikatoren für die allgemeine Qualität der Interpolation sind wohl der RMSE und ASE. Der RMSE benötigt kein stochastisches Modell und berechnet sich im Rahmen der Kreuzvalidierung auf Basis der Abweichungen zwischen Messwerten und geschätzten Werten an den Positionen der Messwerte. Ähnlich zur mittleren Abweichung (ME), beschreibt der RMSE einen Durchschnittswert für die Abweichungen, wohingegen aufgrund der Quadrierung der Fehlerresiduen größeren Abweichungen eine höhere Gewichtung zukommt. Dies macht den RMSE insbesondere in der Flugsicherung, wo hohe Abweichungen besonders unerwünscht sind, zu einem sensitiveren Kontrollparameter für das Gesamtmodell. Ein ungewöhnlich hoher RMSE deutet auf in den Daten vorhandene Trends oder Ausreißer hin. Aus diesem Grund sollten in diesem Fall sowohl die Messdaten auf Anomalien, aber auch das Interpolationsmodell auf dessen Eignung geprüft werden. Der ASE wird bei Verwendung eines stochastischen Modells als Durchschnittswert der geschätzten und quadrierten Standardabweichungen berechnet. Der RMSE stellt mit den Abweichungen zu den Messpunkten einen direkten Bezug zur Realität her, wohingegen sich der ASE auf Basis stochastischer Annahmen und der räumlichen Korrelation ergibt. Der Vergleich zwischen RMSE und ASE kann für den Nachweis verwendet werden, ob das Interpolationsmodell zu einer Über- oder Unterschätzung der Genauigkeiten von Schätzwerten neigt.

Lokale Qualitätsbeurteilung

Lokale Qualitätsbeurteilungen können im Untersuchungsgebiet im Rahmen der Planung von Flugstrecken notwendig sein. Beispielsweise soll ein Mindestabstand von gefährlichen Topographien eingehalten oder die Gefährlichkeit eines Hindernisses eingeschätzt werden. Lokale Genauigkeitsinformationen können durch Schätzung in einem stochastischen Modell oder Referenzpunktmessungen abgeleitet werden. Die Schätzung von Genauigkeiten lässt sich mittels statistischer Interpolationsmethoden für jeden Interpolationspunkt erreichen. Die Verlässlichkeit der Ergebnisse ist aber ausführlich zu prüfen. Die Vermessung von Referenzpunkten ist vergleichsweise aufwändig, stellt jedoch einen direkten Bezug zwischen Realität und Interpolationsmodell her.

Qualitätsbeurteilung in der Flugsicherung

Für den Ausschluss potentieller Gefahrenquellen ist in der Flugsicherung der maximal zu erwartende Fehler mit zugehörigen Sicherheitsgrenzen von Bedeutung. Nach EUROCONTROL (2019, S. 164) wird vorgeschlagen, die globale Validierung der Genauigkeit von Datensätzen anhand eines Konfidenzbereiches durchzuführen. Dabei sollen die Abweichungen zwischen wahren Referenzpunkten und gemessenen Schätzwerten verglichen und ein Konfidenzgrenzwert erreicht werden:

$$100 - \frac{d_e}{d_n} \cdot 100 \geq c \quad (2.82)$$

mit

d_e ... Anzahl der Messungen, die eine maximal zulässige Abweichung für einen Datensatz überschreiten

d_n ... Anzahl aller Kontrollmessungen

c ... Konfidenzgrenzwert

Diese Vorgehensweise kann im Zuge einer Kreuzvalidierung zur Definition von Konfidenzintervallen angewandt werden. Die Qualität der Validierung hängt jedoch hauptsächlich von der Anzahl der Referenzmessungen (Größe der Stichprobe) ab und liefert eine datensatzweite Aussage. Somit könnten regionale Untersuchungen nur mit äußerster Unsicherheit vorgenommen werden, da nicht garantiert ist, dass die dortigen Daten dem geforderten Genauigkeitsniveau entsprechen. Für die lokale Beurteilung werden wiederum ein stochastisches Modell oder Referenzpunktmessungen benötigt. Anhand von lokal geschätzten Standardabweichungen können ebenso Konfidenzbereiche angegeben werden. Beispielsweise können anhand der σ -Regel (vgl. 2.1.2), unter der Annahme eines normalverteilten Prozesses, die Wahrscheinlichkeit angegeben werden, dass ein Grenzwert nicht überschritten wird. Mittels Kriging lassen sich solche Konfidenzbereiche im Zuge der Interpolation für jeden Interpolationspunkt herleiten.

Aufgrund der variablen Ansprüche an die Beurteilung der Sicherheit des Interpolationsmodells, wäre es nur begrenzt sinnvoll einfach einen konstanten Sicherheitspuffer an das Modell anzubringen. Liegen auf Basis statistischer Interpolationsmethoden Schätzungen für die Genauigkeiten von Interpolationspunkten vor, können diese variabel und je nach Anwendungsfall im Qualitätsbeurteilungsprozess berücksichtigt werden. Allgemein kann gefolgert werden, dass sich ein stochastisches Modell und die Kreuzvalidierung im Rahmen des Anwendungsfalls der Flugsicherung komplementieren und somit die besten Mittel für einen möglichst *sicheren* Qualitätsbeurteilungsprozess bereitstellen.

3 Stand der Forschung

Im letzten Abschnitt wurde die Theorie sowie allgemeine Vor- und Nachteile von Interpolationsmethoden besprochen. In diesem Abschnitt wird näher auf wissenschaftliche und praxisnahe Untersuchungen eingegangen. Für die Erarbeitung des aktuellen Standes der Forschung für die Interpolation von digitalen Geländehöhendaten wird folgende Fachliteratur herangezogen:

- a) Comparative Analysis of Spatial Interpolation Methods: an Experimental Study (Tan und Xu, 2014)
- b) A comparative analysis of different DEM interpolation methods (Arun, 2013)
- c) Spatial interpolation (Mitas und Mitasova, 1999)
- d) Terrain and Obstacle Data Manual (EUROCONTROL, 2019)

Die ersten drei Publikationen vergleichen bekannte DHM-Interpolationsmethoden und diskutieren Vor- und Nachteile. Tan und Xu (2014) und Arun (2013) vergleichen die Genauigkeiten von Interpolationsmethoden anhand von Feldmessungen und Gegenüberstellung der Ergebnisse. Mitas und Mitasova (1999) erläutern die jeweiligen Eigenschaften der Interpolationsverfahren. EUROCONTROL (2019) beschreibt allgemeine Richtlinien zur Implementierung einer elektronischen Gelände- und Hindernisdatenbank. In den folgenden Abschnitten dieses Kapitels wird die genannte Fachliteratur evaluiert und deren Ergebnisse diskutiert. In der abschließenden Zusammenfassung wird ein konsolidiertes Bild über die zur Verfügung stehenden Methoden zur Beantwortung der Fragestellung dieser Diplomarbeit bereitgestellt.

3.1 Analyse von Interpolationsmethoden bei unterschiedlicher Topographie

Tan und Xu (2014) beschäftigten sich mit der Analyse verschiedener Interpolationsmethoden in Hinblick auf unterschiedlicher Landschaftsformen. Studien von Zimmerman et al. (1991) weisen darauf hin, dass Kriging unabhängig vom Landschaftstyp bessere Schätzungen als IDW liefert. In anderen Studien von Guarneri und Weih (2012) wurde wiederum gezeigt, dass IDW oder Radiale Basisfunktionen ähnliche oder bessere Ergebnisse wie Kriging liefern. Die Untersuchungen von Tan und Xu (2014) erweitern bestehenden Studien und werden in diesem Abschnitt näher erläutert.

Evaluierung

Die Analyse von Tan und Xu (2014) erfolgte auf Basis einer topographischen Karte im Maßstab 1:50.000. Von dieser wurden für verschiedene Landschaftstypen Referenzpunkte abgeleitet und DHMs mit einer Zellengröße von 25m erstellt:

- Hüggelland (<300m relativer Höhenunterschied, 32 Referenzpunkte)

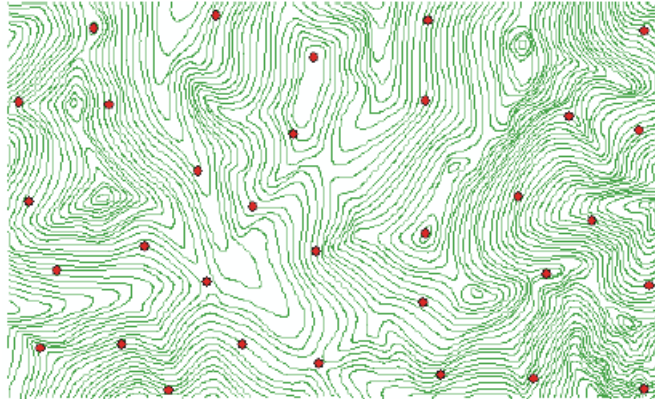


Abbildung 3.1: Hüggelland (Tan und Xu, 2014, S. 158)

- Bergland (<390m relativer Höhenunterschied, 26 Referenzpunkte)

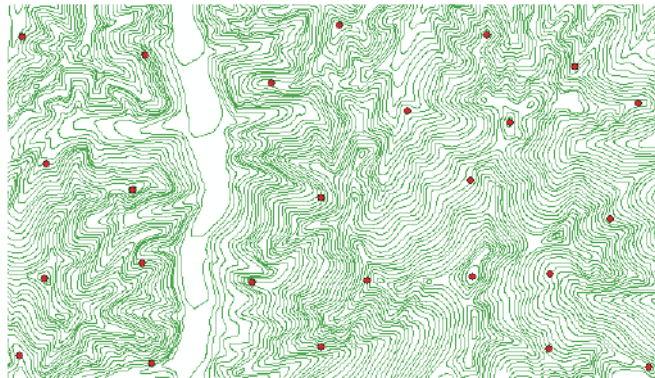


Abbildung 3.2: Bergland (Tan und Xu, 2014, S. 160)

- Gebirgsland (<760m relativer Höhenunterschied, 26 Referenzpunkte)

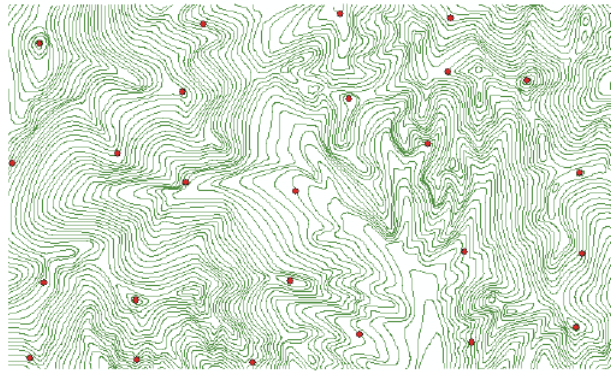


Abbildung 3.3: Gebirgsland (Tan und Xu, 2014, S. 161)

Für das Hügelland, Bergland und Gebirgsland wurden die mittlere (absolute) Abweichung (MAE), die Wurzel der Mittleren quadratischen Abweichung (RMSE) und die allgemeine Anpassung (R^2) für die Interpolationsmethoden IDW, Kriging, Natürlicher Nachbar, TIN, Spline und ANUDEM gegenübergestellt.

Hügelland

Algorithms	MAE(m)	RMSE(m)	R^2
ANUDEM	2.704	5.864	0.98781
TIN	5.512	6.438	0.98333
IDW	4.894	6.153	0.98499
Kriging	5.517	6.436	0.98340
Natural Neighbor	5.495	6.435	0.98336
Spline method	5.857	6.939	0.98053

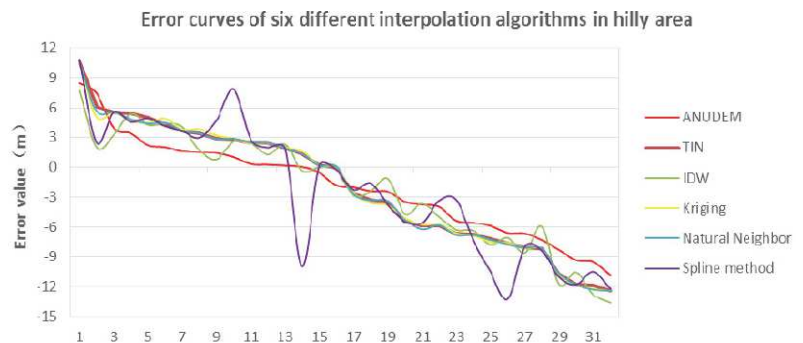


Abbildung 3.4: Fehler im Hügelland (Tan und Xu, 2014, S. 159)

Im Hügelland weichen die Interpolationsergebnisse mit einem RMSE von bis zu 6,9 Metern von den Referenzpunkten ab. Die besten Ergebnisse liefert die eigens für hydrologische Anwendungen konzipierte Methode ANUDEM mit einem RMSE von 5,9 Metern. Die nächstbesten Resultate erzielt IDW. TIN, Kriging und Natürlicher Nachbar liefern sehr ähnliche Ergebnisse. Am schlechtesten schneidet die Spline-Interpolation ab. In der Fehlerkurve ist ersichtlich, dass die Spline-Interpolation in manchen Referenzpunkten große Abweichungen aufweist. Der Vergleich von Konturlinien zwischen Referenz- und Interpolationsmodell zeigt allgemein eine große Überlappung.

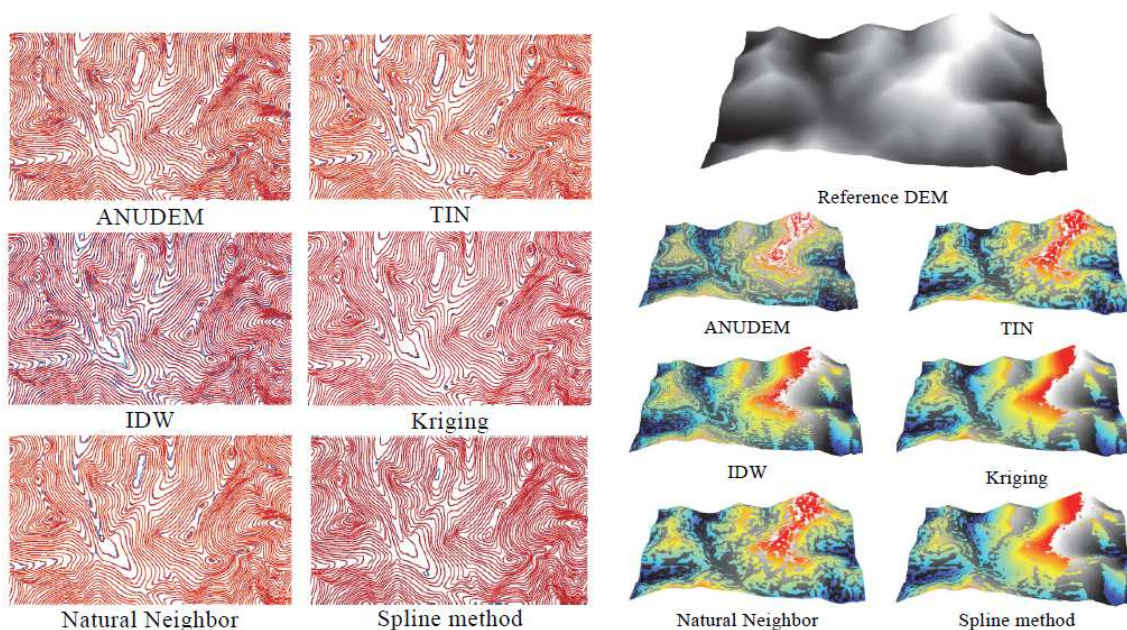


Abbildung 3.5: Hügelland: Vergleich von Konturlinien (rot=Original, blau=Interpolation) und 3D-Visualisierungen (vgl. Tan und Xu, 2014, S. 159)

Bergland

Algorithms	MAE(m)	RMSE(m)	R ²
ANUDEM	5.918	7.083	0.99977
TIN	7.733	9.250	0.99962
IDW	7.231	8.862	0.99965
Kriging	7.479	9.111	0.99963
Natural Neighbor	7.643	9.150	0.99963
Spline method	8.118	9.985	0.99956

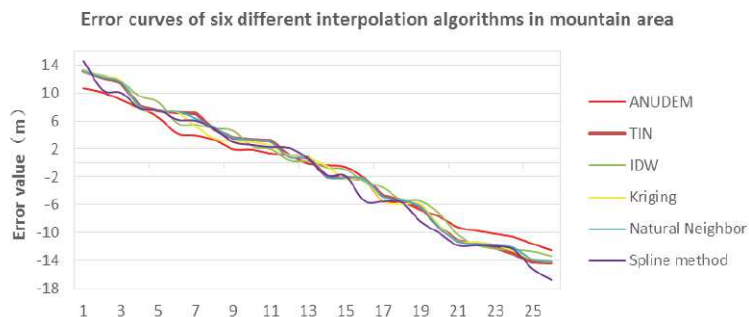


Abbildung 3.6: Fehler im Bergland (Tan und Xu, 2014, S. 160)

Im Bergland weichen die Interpolationsergebnisse mit einem RMSE von bis zu maximal 9,9 Metern von den Referenzpunkten ab. Die besten Ergebnisse verzeichnet wieder die ANUDEM Methode mit einem RMSE von 7,1 Metern. IDW liefert die zweitbesten Ergebnisse. Kriging, Natürlicher Nachbar und TIN haben analog zum Hügelland ähnlich gute Resultate im Mittelfeld. Die Spline-Interpolation schneidet erneut am schlechtesten ab. Bei der Interpolation im Bergland fällt allgemein eine größere Abweichung der Konturlinien aus dem Interpolati-

onsergebnis, zu denen des Referenzmodells auf. Vor allem bei der Spline-Interpolation können im Bereich des Bergkammes größere Abweichungen festgestellt werden.

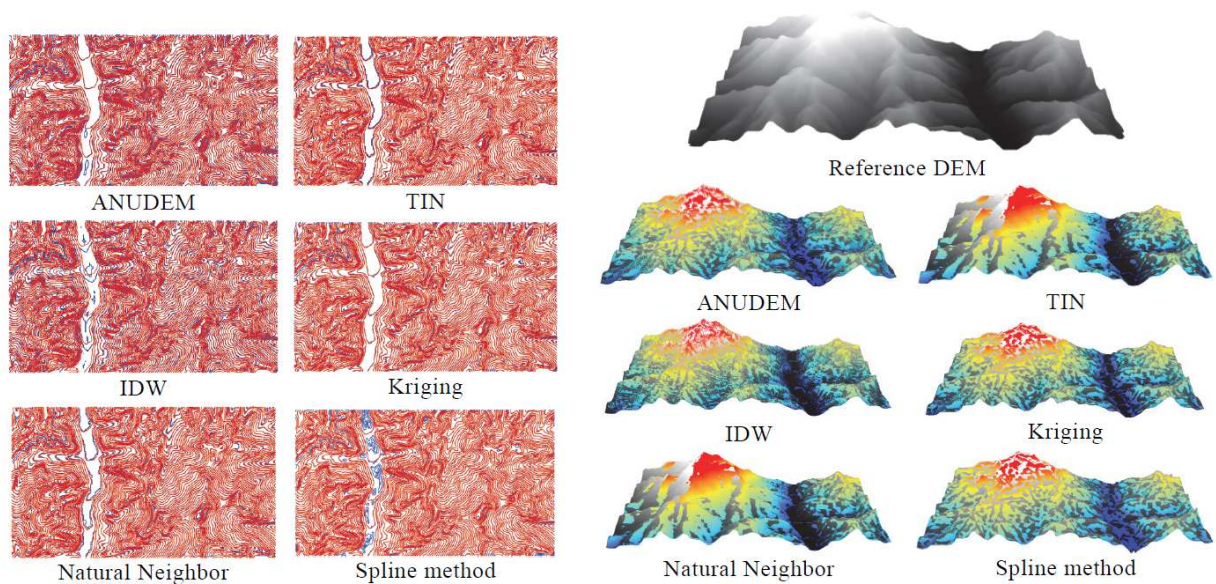


Abbildung 3.7: Bergland: Vergleich von Konturlinien (rot=Original, blau=Interpolation) und 3D-Visualisierungen (Tan und Xu, 2014, S. 160-161)

Gebirgsland

Algorithms	MAE (m)	RMSE	R ²
ANUDEM	6.661	8.799	0.99983
TIN	9.021	10.732	0.99969
IDW	8.612	10.571	0.99970
Kriging	8.997	10.764	0.99969
Natural Neighbor	8.999	10.700	0.99970
Spline method	9.158	12.315	0.99959

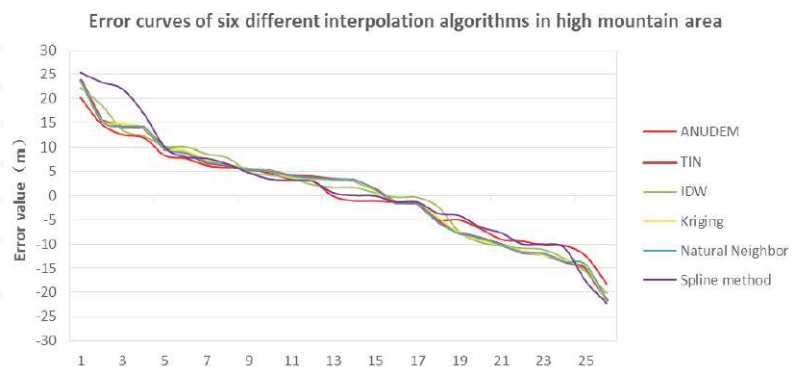


Abbildung 3.8: Fehler im Gebirgsland (Tan und Xu, 2014, S. 161)

Im Gebirgsland weichen die Interpolationsergebnisse mit einem RMSE von bis zu 12,3 Metern von den Referenzpunkten ab. Analog zu den vorherigen Ergebnissen liefert ANUDEM mit

einem RMSE von 8,8 Metern das beste Ergebnis. Bemerkenswert ist, dass die Methode auch in Relation zu den Ergebnissen vom Bergland sehr gute Resultate liefert. Analog zu den Untersuchungen im Hügel- und Bergland liefert IDW die zweitbesten Ergebnisse. TIN, Kriging und Natürlicher Nachbar liegen wieder im Mittelfeld. Die Spline-Interpolation erzielt mit Abstand die schlechtesten Ergebnisse. Dies bestätigt auch der visuelle Abgleich mit den Konturlinien.

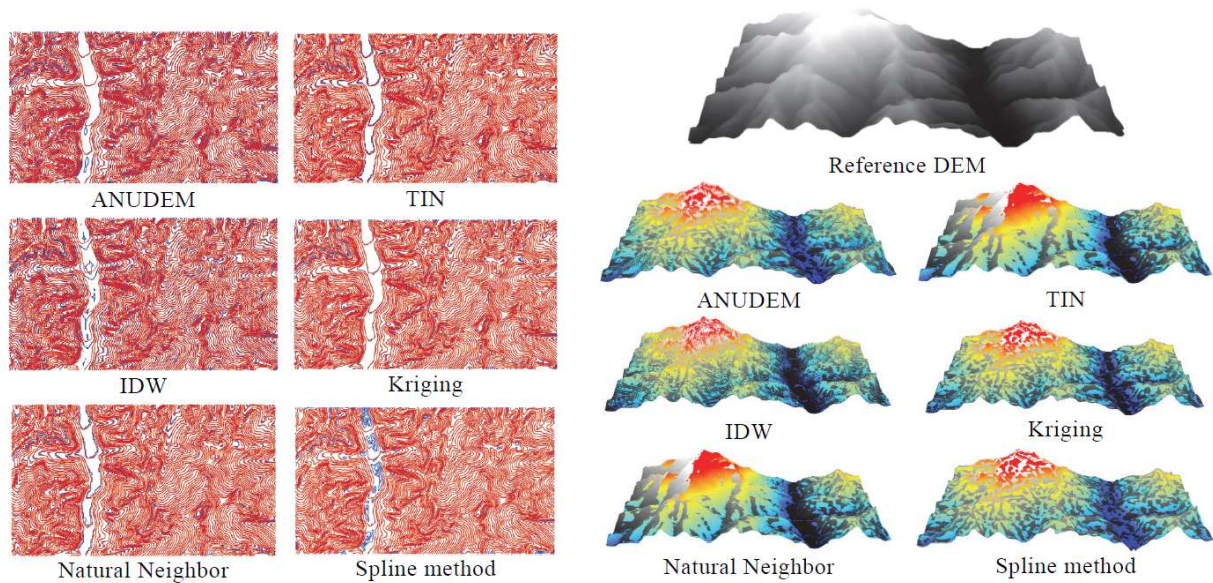


Abbildung 3.9: Gebirgsland: Vergleich von Konturlinien (rot=Original, blau=Interpolation) und 3D-Visualisierungen (Tan und Xu, 2014, S. 161-162)

Ergebnisse

Bei Tan und Xu (2014) werden die Ergebnisse gängiger Interpolationsmethoden bei drei Landschaftsformen evaluiert. Allgemein lässt sich ableiten, dass der mittlere Fehler mit zunehmender Variabilität der Höhenwerte im Untersuchungsgebiet zunimmt. Die ANUDEM Methode lieferte in allen Versuchsgruppen die besten Ergebnisse. IDW erzielt die zweitbesten Ergebnisse. Die lokalen Methoden TIN und Natürlicher Nachbar weisen zusammen mit Kriging vergleichbare Resultate auf. Am schlechtesten schneidet die Spline-Interpolation ab. Da bei Tan und Xu (2014) keine Trendanalysen oder spezielle Konfigurationen der Interpolationsmethoden genannt werden, ist davon auszugehen, dass die jeweiligen Verfahren in ihrer einfachsten Form angewandt wurden. Speziell globale Methoden sind eher auf eine vorangehende Datenanalyse angewiesen, um möglichst optimale Ergebnisse zu liefern. Deshalb und aufgrund der Eigenschaft von lokalen Methoden, sich gut an variables Gelände anzupassen, ist es auch nicht verwunderlich, dass die lokalen Methoden im Schnitt bessere

Ergebnisse liefern. Es lässt sich folgern, dass lokale Interpolationsmethoden in Gebieten hoher Variabilität bessere Ergebnisse liefern als Methoden mit globaler Interpolationsvorschrift ohne Feinabstimmung.

3.2 Analyse von Interpolationsmethoden anhand von GPS-Referenzmessungen

Arun (2013) untersuchte die Genauigkeit einer Reihe von gängigen Interpolationsmethoden anhand eines GPS-Referenzpunktnetzes.

Evaluierung

Im Rahmen der Feldmessung von Arun (2013) wurden mittels DGPS rund 1000 Testpunkte im WGS84-Datum bestimmt. Rund 680 Testpunkten wurden zur Generierung der DHMs verwendet. Die restlichen Punkte dienten zur Schätzung der Interpolationsgenauigkeiten. Es wurden die Interpolationsmethoden IDW, Kriging, ANUDEM, Natürlicher Nachbar und Spline verglichen. Im Artikel wurde Natürlicher Nachbar mit 'Nearest Neighbor (NN)' bezeichnet. Aus der Beschreibung geht jedoch hervor, dass es sich um eine gewichtete Berechnung anhand der Flächenproportionalitäten handelt, was der Natürliche Nachbar Interpolation entspricht. Die Interpolationsergebnisse wurden durch Vergleich mit Referenzmessungen evaluiert. Abbildung 3.10 zeigt die Gegenüberstellung der Resultate verschiedener Interpolationsmethoden mit dem ermittelten Wert der DGPS-Beobachtung.

Table 2 Ellipsoidal heights at test GCPs from interpolated and DGPS observed values.

Control points ID	Ellipsoidal heights in meter					DGPS observed value
	IDW Value	ANUDEM	Nearest neighbor	Spline	Kriging	
FID-24	474.97	474.89	476.93	476.33	476.02	476.72
FID-81	476.64	476.78	476.72	477.54	476.68	478.54
FID-39	479.42	478.63	479.22	479.37	478.90	477.60
FID-7	476.78	477.30	475.59	475.66	475.83	478.40
FID-11	477.75	477.30	477.78	477.30	477.29	478.24
FID-14	478.06	477.48	478.59	480.76	479.63	479.58
FID-17	479.27	477.85	479.32	480.36	479.44	479.37
FID-71	477.38	476.26	477.18	477.65	477.64	476.68
FID-64	477.39	478.41	478.05	478.18	477.93	477.28
FID-61	479.30	480.06	479.59	477.79	479.03	479.69
FID-56	477.72	478.11	478.40	477.96	478.20	475.72
FID-45	477.93	478.93	477.57	477.36	477.43	477.97
FID-39	479.450	478.32	479.22	479.123	478.75	477.60
FID-91	473.11	471.95	473.94	474.50	474.18	475.83
FID-87	473.02	471.95	473.56	473.25	473.39	476.42
FID-30	474.32	474.83	473.67	471.40	472.58	473.19
FID-28	473.80	475.57	473.41	473.32	473.47	471.82
FID-89	473.14	471.95	472.31	472.33	472.22	469.89
FID-95	471.08	471.95	471.27	471.15	471.11	472.14
FID-34	477.07	478.01	477.90	477.14	477.43	477.82

Abbildung 3.10: Vergleich von interpolierten Punkten mit DGPS-Beobachtung (Arun, 2013, S. 135)

Weiters wurde die Genauigkeit der Interpolationsmethoden in Hinblick auf die Steilheit des Geländes untersucht. Für die Evaluierung der Genauigkeit wurde der RMSE angegeben.

Type of test GCPs used	RMSE values				
	IDW	ANUDEM	NN	Spline	Kriging
Mild slope areas	0.93	0.87	0.72	0.91	0.70
Steep slope area	1.45	1.82	1.34	1.37	1.31
Combined slope area	1.73	2.02	1.53	1.62	1.49

Abbildung 3.11: Vergleich von RMSE-Werten bei verschiedenen Geländetypen (Arun, 2013, S. 135)

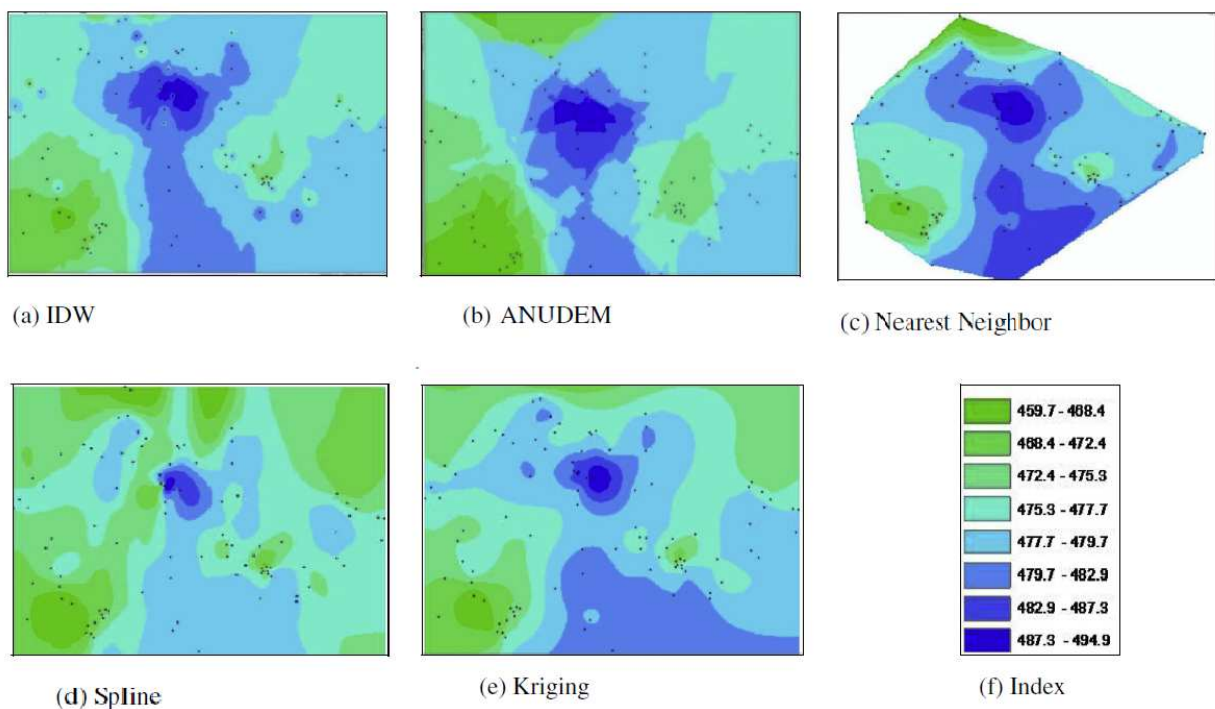


Abbildung 3.12: Graphische Visualisierung der Interpolationsergebnisse (Arun, 2013, S. 136)

Kriging und die Natürliche Nachbar Interpolation erreichen die niedrigsten RMSE-Werte und scheinen sich daher auf unterschiedliche Geländebeschaffenheiten am besten anzupassen. Bei leichten Steigungen schneidet die IDW-Methode am schlechtesten ab. Bei steileren Steigungen liefert die ANUDEM Interpolation mit den höchsten RMSE-Werten die schlechtesten Ergebnisse. Bei Arun (2013) wurde jedoch darauf hingewiesen, dass ANUDEM bei Bergrücken und Flussnetzwerken überzeugende Resultate liefert. Die Spline-Interpolation landet, im Vergleich zu den anderen Methoden, im unteren Mittelfeld.

Ergebnisse

Durch Vergleich der niedrigsten RMSE-Werten hat sich gezeigt, dass Kriging, auch bei unterschiedlichen Geländetypen, die besten Ergebnisse liefert. Auch wenn ANUDEM nicht die niedrigsten RMSE-Werte erzielt, wird es doch für die bestmögliche Bestimmung von Interpolationsmodellen für Flussnetzwerken und Bergkämmen empfohlen. Arun (2013) bestätigt die von Zimmerman et al. (1991) durchgeführten Studien, dass Kriging unabhängig vom Landschaftstyp und Abtastmuster bessere Schätzungen für Höhenwerte als IDW oder Radiale Basisfunktionen liefert.

3.3 Analyse der Eigenschaften von Interpolationsmethoden

Mitas und Mitasova (1999) haben sich generell mit dem Problem der räumlichen Interpolation aus verstreuten Daten befasst. Dabei ist je nach Aufgabenstellung das entsprechende Interpolationskonzept zu wählen:

- Geostatistisches Konzept → Kriging
- Lokalität → Nächster Nachbar, oder finite Elemente Methode
- Glattheit und Spannung → Splines
- Funktionale Form → Polynome oder Radiale Basisfunktion

Im Kontext von Geographischen Informationssystemen (GIS) wurde auf die Komplexität der zu modellierenden Daten (heterogene Verteilung, unvorteilhaftes Sampling, signifikante Diskontinuitäten) eingegangen und wünschenswerte Kriterien formuliert:

- Hohe Genauigkeit und Vorhersagekraft
- Robustheit und Flexibilität zur Beschreibung verschiedener Phänomene
- Glättung von verrauschten Daten
- Mehrdimensionale Formulierung
- Direkte Schätzung von Ableitungen (Steigungen, Krümmungen)
- Anwendbarkeit für große Datensätze
- Recheneffizienz
- Benutzerfreundlichkeit

Mitas und Mitasova (1999) betonten, dass es schwer ist eine Methode zu finden, die allen oben genannten Anforderungen entspricht. Die Auswahl der *optimalen* Interpolationsmethode ist dabei je nach Aufgabenstellung zu unterscheiden. Im Unterschied zu Tan und Xu (2014) sowie Arun (2013) wo die Genauigkeit der Interpolationsverfahren im Vordergrund stand, konzentrierten sich Mitas und Mitasova (1999) auf die Eigenschaften der Interpolationsmethoden.

Evaluierung

Mitas und Mitasova (1999) beschrieben die Eigenschaften der Interpolationsverfahren IDW, Natürlicher Nachbar, TIN, Kriging und Splines. Die Methoden wurden anhand verfügbarer GIS-Implementierungen und Beispieldatensatz gegenübergestellt:

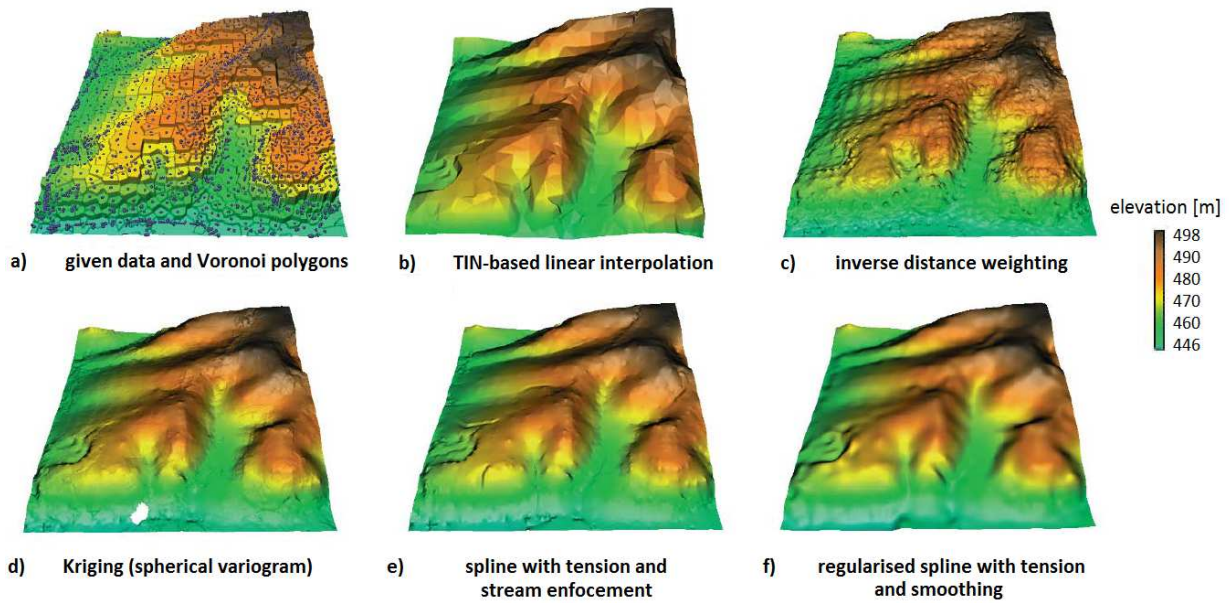


Abbildung 3.13: Vergleich von Interpolationsmethoden anhand unregelmäßig verstreuter Daten (Mitas und Mitasova, 1999, S. 482-483)

Die bei Mitas und Mitasova (1999) genannten Vor- und Nachteile sind in folgender Tabelle zusammengefasst und gegenübergestellt:

Methoden	Vorteile	Nachteile
Inverse Distance Weighting	<ul style="list-style-type: none"> • Einfach anzuwenden • Performant • Suchradius konfigurierbar 	<ul style="list-style-type: none"> • Produziert lokale Minima und Maxima (Bull-Eye Effekt)
Natürlicher Nachbar	<ul style="list-style-type: none"> • Bezieht lokale Begebenheiten gut ein • Um Glattheits- und Tensionseigenschaften erweiterbar 	<ul style="list-style-type: none"> • Pre-Prozessierung für Glattheits- und Tensionseigenschaften erforderlich

Methoden	Vorteile	Nachteile
TIN	<ul style="list-style-type: none"> • Performant • Bezieht lokale Begebenheiten gut ein 	<ul style="list-style-type: none"> • Verhältnismäßig ungenau
Kriging	<ul style="list-style-type: none"> • Basiert auf räumlicher Korrelation der Messwerte • Liefert Kovarianzinformationen • Anisotropie kann berücksichtigt werden • Durch viele Erweiterungen sehr flexibel einsetzbar 	<ul style="list-style-type: none"> • Erfasst lokale Geometrien weniger gut • Weniger für möglichst glatte Ergebnisse geeignet • Explorative Datenanalyse notwendig
Variationsansatz	<ul style="list-style-type: none"> • Basiert auf physikalischem Modell • Glattheit und Tension steuerbar • Anisotropie kann berücksichtigt werden • Durch viele Erweiterungen sehr flexibel einsetzbar • Gut für Geometrieanalysen geeignet 	<ul style="list-style-type: none"> • Aufwendiges Berechnungsverfahren

Ergebnisse

Mitas und Mitasova (1999) widmeten sich speziell dem Vergleich zwischen Kriging und dem Variationsansatz und stellten fest, dass beide Methoden ähnlich einsetzbar und erweiterbar sind. Dabei wurde betont, dass die Wahl der Methode stark vom Einsatzgebiet und von dem zu modellierenden Phänomen abhängt. Kriging und Splines verfügen beide über komplexe, aber auch flexibel einsetzbare Modelle, mit denen die genauesten Ergebnisse erreicht werden können. Anhand der Komplexität ist jedoch auch ein erweitertes Theorieverständnis nötig, um die optimale Konfiguration für das Interpolationsmodell aufzustellen.

3.4 Elektronische Gelände- und Hindernisdaten nach ICAO-Standards

EUROCONTROL (2019) hat alle nötigen Grundlagen, Anforderungen und Vorgehensweisen zur Implementierung einer elektronischen Gelände- und Hindernisdatenbank beschrieben. Dabei wurde unter anderem auf Aspekte der Datenakquise und Zusammenführung von Datensätzen eingegangen, die in diesem Abschnitt näher betrachtet werden.

Evaluierung

EUROCONTROL (2019) beschreibt Maßnahmen zur Erstellung von Gelände- und Hindernisdaten unter Berücksichtigung der ICAO-Qualitätsanforderungen. Geländedaten können dabei grundsätzlich aus unterschiedlichsten Quellen stammen (vgl. EUROCONTROL, 2019, S. 74):

- Militär
- Flugsicherungen
- Geodätische Institute
- Kartierungsagenturen
- Energieversorgungsunternehmen
- Windparkbetreiber
- Mobilfunkbetreiber
- Behörden

Dabei existieren unterschiedliche Verfahren zur qualitativ hochwertigen Erfassung von Geländehöhendaten. Die Art der Datenaufnahme ist immer in Hinblick auf den kompletten Datenerstellungsprozess zu wählen. Daher können sich einzelne Messmethoden, aber auch kombinierte Ansätze, als vorteilhaft für die Erstellung von Geländedaten im Kontext von eTOD erweisen (vgl. EUROCONTROL, 2019, S. 155-156):

- *Terrestrische Vermessung*: Wird für die Erfassung von Hindernis- und Geländedaten sowie für die Aufnahme von Referenzpunkte zur fluggestützten Vermessung und deren Validierung verwendet.
- *Luftbild-Photogrammetrie*: Wird für die großräumige Erfassung von Hindernisdaten, Geländedaten und Validierung von ALS-Daten verwendet.
- *Airborne Laser Scanning (ALS)*: Wird für die großräumige Erfassung von Gelände- und Hindernisdaten eingesetzt. Ein wesentlicher Vorteil dieser Methode ist der hohe Grad an Automatisierung.

- *Interferometric Synthetic Aperture Radar (IfSAR)*: Wird für die Geländedatenerfassung verwendet.
- *Sensorfusion*: Wird für die Erfassung von Hindernis- und Geländedaten verwendet. Generell hat sich die Kombination aus ALS und Luftbild-Photogrammetrie bewährt.

Die Umgebung rund um einen Flughafen wird in sicherheitskritische Gebiete (Area 1-4) mit unterschiedlichen Genauigkeitsanforderungen eingeteilt. Dabei werden folgende Vermessungsmethoden für das jeweilige Gebiet empfohlen:

	ALS	IfSAR	Photo-grammetry	Terrestrial Survey
Area 1	+	++	+	o
Area 2	++	++	++	o
Area 3	+ / ++	-	+ / ++	+
Area 4	+ / ++	-	+ / ++	++

Abbildung 3.14: Empfohlene Vermessungsmethoden für Geländedaten (EUROCONTROL, 2019, S. 157-158)

- ++ ... technisch sehr gut geeignet und sehr kostengünstig
- + ... technisch sehr gut geeignet, aber nicht die kostengünstigste Variante
- o ... technisch geeignet, aber sehr schlechtes Kosten-Nutzen-Verhältnis
- ... technisch ungeeignet und sehr schlechtes Kosten-Nutzen-Verhältnis

Nach EUROCONTROL (2019, S. 196) können datenlose Bereiche bis zu 9 Rasterzellen durch eine Spline-Interpolation gefüllt werden. Zusätzlich sollte die Interpolation vor Erstellung des Rasters in einer Kartenprojektion durchgeführt werden, um das Problem von ungleichen Zellengrößen mit zunehmender geographischer Breite zu vermeiden. Datentransformationen von Kartenprojektionen zu geographischen Koordinatensystemen dürfen dabei keine Fehler einführen. Datenlose Bereiche größer 9 Rasterzellen müssen dokumentiert werden. Datenlose Bereiche größer 36 Rasterzellen sind inakzeptabel. Interpolierte Punkte müssen als solche gekennzeichnet werden. Die Erstellung des Rasters sollte der letzte Prozessierungsschritt sein.

Zusammenführung von sicherheitskritischen Geländedaten

Eine kostengünstigere Variante zur Erstellung von geeigneten Geländedaten ist die Verwendung von bestehenden Datensätzen. Dabei müssen die Daten jedoch immer einer gründlichen Qualitätsprüfung unterzogen werden. Bestehen Zweifel an der Tauglichkeit der Daten für Anwendungsfälle der Flugsicherung, ist dies durch den Datenanbieter eindeutig zu kennzeichnen (vgl. EUROCONTROL, 2019, S. 91). Allgemein sollte die Qualität der Daten und deren Konformität zu den ICAO-Anforderungen durch Referenzpunkte verifiziert werden.

Sofern die ausreichende Qualität der Datensätze feststeht, können diese zusammengeführt werden, um die nötige Abdeckung in den verschiedenen Sicherheitsgebieten rund um einen Flughafen zu erreichen. Bei geographisch überlappenden Bereichen stellt sich die Frage, welche Datenpunkte nun für die Geländeoberfläche repräsentativ sind. Da die Sicherheit in der Flugsicherung an oberster Stelle steht, sollte auch das Endresultat der Interpolation den größtmöglichen Grad an Sicherheit aufweisen. Nach EUROCONTROL (2019, S. 110) wurde die Wahl des höchsten zur Verfügung stehenden Datenpunktes als *sicherste* Methode vorgeschlagen. Ein wichtiges Kriterium für diese Vorgehensweise ist jedoch, dass alle verwendeten Daten den ICAO-Qualitätsanforderungen entsprechen, ansonsten würde dies zu signifikanten Qualitätseinbußen führen.

Ergebnisse

EUROCONTROL (2019) beschreibt allgemeine Leitlinien und Überlegungen, die für die Implementierung einer Datenbank für Gelände- und Hindernisdaten besonders zu berücksichtigen sind. Es wurde aufgezeigt, dass Geländedaten für ein eTOD-System sowohl durch Vermessungs-Kampagnen als auch durch frei zugängliche Daten erstellt werden können. Aus Sicht des Datenanbieters ist darauf zu achten, dass die Daten entsprechend ihrer Qualität geprüft und für den Endnutzer gekennzeichnet werden. In Hinblick auf die Erstellung der Datensätze wird gefordert, dass bei mehreren qualitätsgesicherten Daten, immer der höchste Wert für die jeweilige Zelle verwendet wird. In Abhängigkeit der Genauigkeitsanforderungen wurde empfohlen, dass datenlose Bereiche bis zu 9 Rasterzellen durch eine Spline-Interpolation gefüllt und gekennzeichnet werden können. Die Spline-Interpolation wurde empfohlen, da sie nicht auf den Wertebereich der Messpunkte beschränkt ist. Das Ausschwingen der Interpolation aus dem Wertebereich bei Maxima (z.B. Bergspitze) führt zu einem zusätzlichen Sicherheitspuffer, sollten die Daten nicht genau erfasst worden sein. EUROCONTROL (2019) begründete die Empfehlung, eine Interpolation nur im Ausmaß der Ausdehnung von 9 Rasterzellen vorzunehmen, nicht weiter. Es wird die Annahme vermutet, dass sich die Interpolationsfehler in diesen Zellbereichen mit ziemlicher Sicherheit im Rahmen der ICAO-Anforderungen bewegen werden. Die Überprüfung der Genauigkeit wurde durch Vergleiche mit Referenzpunkten vorgeschlagen. Sofern gewisse Fehlergrenzwerte mit einer bestimmten Konfidenz nicht überschritten werden, gilt ein Datensatz als genau genug. Die Referenzdaten sollten dabei jedoch um den Faktor 3 genauer sein (vgl. EUROCONTROL, 2019, S. 164). Ein Nachteil dieses Verifizierungsansatzes ist, dass dabei die datensatzweite Qualität beschrieben wird und keine regionalen Aussagen gemacht werden können. Weiters ist die Qualität der Beurteilung von der Anzahl der vorgenommenen Kontrollmessungen abhängig. Auf statistische Interpolationsverfahren wurde bei EUROCONTROL (2019) nicht eingegangen. Dabei stellen gerade diese Verfahren Methoden bereit, um stochastisch begründbare Aussagen darüber zu liefern, ob jeder geschätzte Punkte noch den Qualitätsanforderungen entspricht und mit welcher Wahrscheinlichkeit kritische Grenzwerte überschritten werden. Die Kontrolle anhand von Referenzpunkten sollte in jedem Fall trotzdem durchgeführt werden. Die geschätzten Genauigkeitsinformationen auf Datenebene

würden dazu beitragen den Qualitätsbeurteilungsprozess flexibler (regionale Untersuchungen), effizienter (ad-hoc Genauigkeitsinformationen nach Interpolation) und sicherer (Genauigkeitsinformationen auf Datenebene) zu machen.

3.5 Zusammenfassung und Bewertung des Forschungsstandes

Alle in diesem Kapitel beschriebenen Untersuchungen haben gezeigt und betont, dass der Einsatz verschiedener Interpolationsmethoden stark vom Anwendungszweck abhängt. Zudem steigt die Ungenauigkeit der Interpolationsergebnisse mit zunehmender Variabilität in den Daten. Bei Tan und Xu (2014) hatte sich ANUDEM bei allen Landschaftstypen als genaueste Methode herausgestellt. Die zweitbesten Ergebnisse lieferte die IDW-Interpolation. Natürlicher Nachbar, Kriging und TIN erzielten für alle Landschaftstypen ähnliche Ergebnisse. Die Spline-Interpolation lieferte die ungenauesten Ergebnisse. Arun (2013) kam zum selben Ergebnis wie Zimmerman et al. (1991) und beschrieb Kriging bei verschiedenen Steigungstypen als genaueste Methode. Die ANUDEM-Methode schnitt bei diesen Untersuchungen, gerade in Bereichen steiler und variabler Steigungen am schlechtesten ab. Bei Mitas und Mitasova (1999) wurden der Spline- und Kriging-Ansatz als genaueste Interpolationsmethoden hervorgehoben. Kriging wurde speziell zur Schätzung von statistischen Unsicherheiten empfohlen. Splines sind dann am geeignetsten, wenn Geometrieanalysen von Interesse sind. Allgemein kann festgestellt werden, dass sich anhand der Feldmessungen kein eindeutiges Bild darüber ergibt, welche Methoden nun am genauesten arbeiten. Dabei bestätigt sich wiederum der Schluss, dass die zu wählende Interpolationsmethode stark vom Verwendungszweck abhängt. Nach EUROCONTROL (2019) wurde der Spline-Ansatz für die Interpolation von Geländehöhendaten in beschränkten geographischen Bereichen vorgeschlagen. Statistische Interpolationsverfahren könnten jedoch durch die direkte Berechnung von Fehlerschätzungen und Konfidenzbereichen eine vertretbare Ausweitung der Interpolation auf größere Gebiete ermöglichen und damit eine kostengünstige Alternative zu Vermessungen oder Datenanschaffungen darstellen. Allgemein würden die durch statistische Interpolationsmethoden berechneten Genauigkeitsinformationen auf Datenebene dabei helfen, den Qualitätsbeurteilungsprozess flexibler, effizienter und sicherer zu machen.

4 ICAO Anforderungsanalyse

Für den Validierungsprozess der *optimalen* Interpolationsmethode im Kontext elektronischer Gelände- und Hindernisdatenbanken werden konkrete und messbare Kriterien benötigt. Für die Erhebung dieser Kriterien wird eine Anforderungsanalyse durchgeführt. Nach Ebert (2019, Kapitel 2.1) kann eine Anforderung als eine Eigenschaft oder Bedingung definiert werden, die ein System erfüllen muss, um einen Vertrag, eine Norm oder andere, formell vorgegebene Dokumente zu erfüllen. Bei Ebert (2019, Kapitel 2.5) findet man eine allgemeine Vorgehensweise zur Erstellung eines konkreten Anforderungsprofils:

- **Anforderungen ermitteln:** Recherche von allgemeinen Standards und Normen
- **Anforderungen dokumentieren:** Extraktion der relevanten Anforderungen im Rahmen der Aufgabenstellung dieser Diplomarbeit
- **Anforderungen analysieren:** Interpretation der Auswirkung der Anforderung auf das Ergebnis dieser Diplomarbeit
- **Anforderungen prüfen:** Validierung der Lösung und Erfüllung der Anforderung

Speziell im Bereich der Flugsicherung kommen eine Vielzahl an Regelwerken und Normen zum Einsatz, um die geforderte Sicherheit zu gewährleisten. Für die Evaluierung einer geeigneten Interpolationsmethode wird eine Anforderungsanalyse basierend auf den Dokumenten 'ICAO Annex 15' (vgl. ICAO, 2018a) und 'PANS AIM' (vgl. ICAO, 2018b) durchgeführt. Diese beschreiben Datenqualität im Kontext von Gelände- und Hindernisdaten und stützen sich auf die 'ISO 19100 series of geospatial standards' als generelles Framework zur Datenmodellierung. Die Implementierung dieser Standards soll die richtige Datenqualität im Kontext der Flugsicherung gewährleisten.

4.1 Qualitätsanforderungen der ICAO Spezifikation

ICAO (2018a) und ICAO (2018b) beschreiben notwendige Anforderungen der Flugsicherung an digitale Geländehöhendaten. Die ICAO unterscheidet zwischen verpflichtenden Anforderungen und Empfehlungen. Der Vollständigkeit halber werden beide Anforderungstypen interpretiert und die Konsequenz abgeleitet. In der Lösungsfindung für die Fragestellung dieser Diplomarbeit werden jedoch nur die verpflichtenden Anforderungen als zwingend einzuhalten erachtet. Beide Anforderungsdokumente referenzieren aufeinander und behandeln

digitale Geländehöhendaten im jeweils gleichnamigen Kapitel 5 'Aeronautical information products and services'. Die für digitale Geländehöhendaten relevanten Anforderungen finden sich in den Abschnitten 'Terrain and obstacle data sets' und 'Terrain data sets'. Die Anforderungen werden in den folgenden Abschnitten im Originaltext mit zugehöriger ICAO Anforderungsnummer angegeben, interpretiert und Erfüllungskriterien zur Validierung von Datenanforderungen und Interpolationsmethoden erarbeitet.

4.1.1 Anforderungen von ICAO Annex 15

In diesem Abschnitt werden Anforderungen zu sicherheitskritischen Abdeckungsgebieten und der allgemeinen Datenbereitstellung evaluiert.

Anforderung 5.3.3.1

"The coverage areas for terrain and obstacle data sets shall be specified as:

- *Area 1: the entire territory of a State;*
- *Area 2: within the vicinity of an aerodrome, subdivided as follows:*
 - *Area 2a: a rectangular area around a runway that comprises the runway strip plus any clearway that exists;*
 - Note.— See Annex 14, Volume I, Chapter 3, for dimensions for runway strips.*
 - *Area 2b: an area extending from the ends of Area 2a in the direction of departure, with a length of 10 km and a splay of 15 per cent to each side;*
 - *Area 2c: an area extending outside Area 2a and Area 2b at a distance of not more than 10 km from the boundary of Area 2a; and*
 - *Area 2d: an area outside Areas 2a, 2b and 2c up to a distance of 45 km from the aerodrome reference point, or to an existing terminal control area (TMA) boundary, whichever is nearest;*
- *Area 3: the area bordering an aerodrome movement area that extends horizontally from the edge of a runway to 90 m from the runway centre line and 50 m from the edge of all other parts of the aerodrome movement area; and*
- *Area 4: the area extending 900 m prior to the runway threshold and 60 m each side of the extended runway centre line in the direction of the approach on a precision approach runway, Category II or III. "*

Interpretation

Jeder Staat **muss** für sein jeweiliges Hoheitsgebiet Geländedaten bereitstellen (Gebiet 1). Weiters gibt es ausgewiesene Gebiete um einen Flughafen, für die Geländedaten vorliegen müssen (Gebiet 2-4). ICAO (2018b, Appendix 8) lässt sich folgende Visualisierung für diese

Gebiete entnehmen:

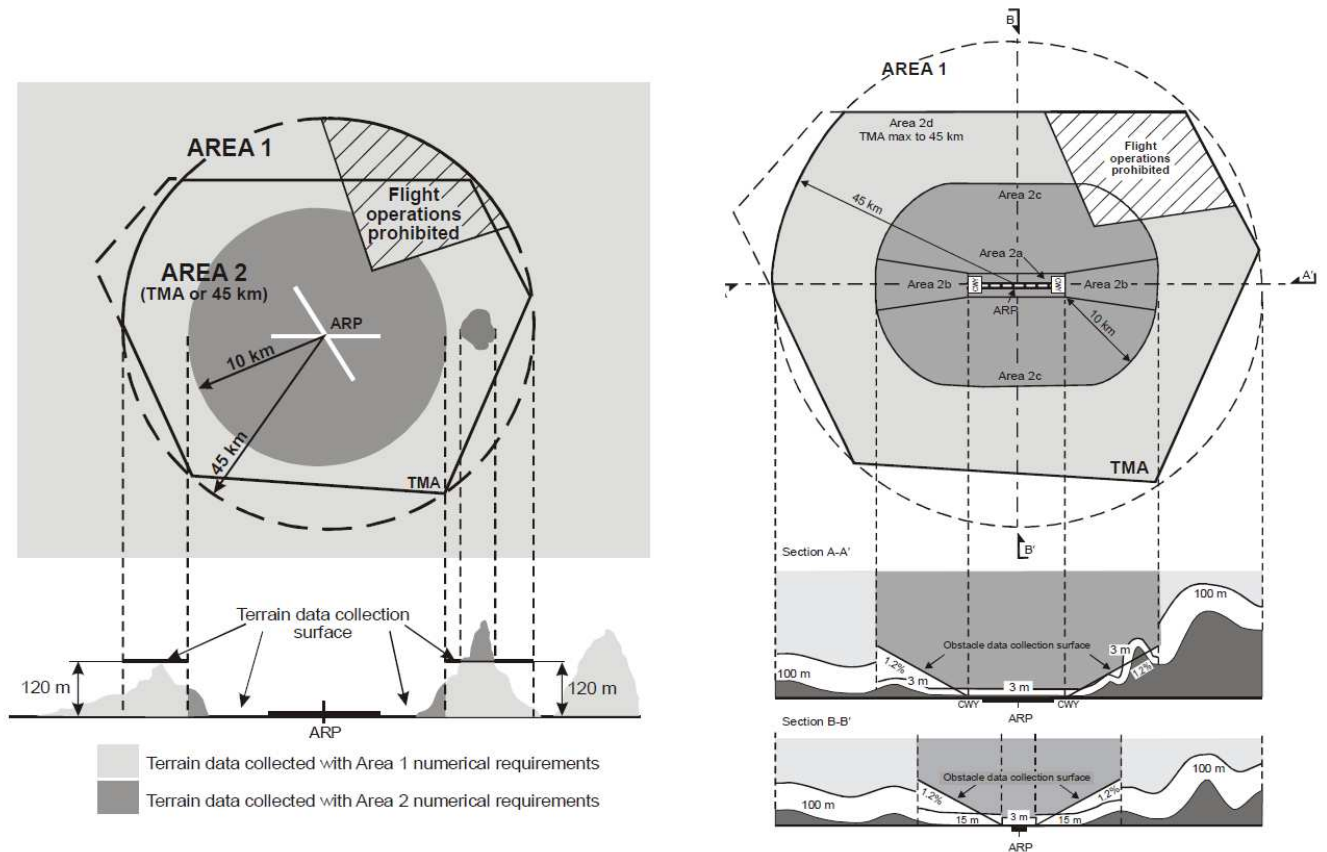


Abbildung 4.1: Gebiete 1 und 2 (ICAO, 2018b, Appendix 8)

- * ARP: Aerodrome Reference Point, Flugplatzbezugspunkt
- * TMA: Terminal Area, Nahverkehrsbereich

Interpolation kann generell dafür eingesetzt werden um aus globalen, regionalen und lokalen digitalen Geländehöhenmodellen ein flächendeckendes, einheitliches Modell zu erstellen.

Erfüllungskriterium

Die Interpolation muss in einem bestimmten geographischen Bereich und aus unterschiedlichen Datenquellen ein einheitliches Modell generieren.

Anforderung 5.3.3.2

"Recommendation.— Where the terrain at a distance greater than 900 m (3 000 ft) from the runway threshold is mountainous or otherwise significant, the length of Area 4 should be extended to a distance not exceeding 2 000 m (6 500 ft) from the runway threshold."

Interpretation

Für Gebiet 4 **sollte** deren Länge bis zu (aber nicht mehr) als 2000m erweitert werden, sofern es signifikante Geländehöhen gibt, die mehr als 900m von der Landebahnschwelle entfernt liegen. Interpolation stellt dabei die Mittel zur Verfügung, bestehende Datensätze zu erweitern.

Erfüllungskriterium

Die Interpolation muss in einem bestimmten geographischen Bereich und aus unterschiedlichen Datenquellen ein einheitliches Modell generieren.

Anforderung 5.3.3.3.1

"Terrain data sets shall contain the digital representation of the terrain surface in the form of continuous elevation values at all intersections (points) of a defined grid, referenced to common datum."

Interpretation

Ein Geländedatensatz **muss** digitale Höhenwerte für alle Punkte eines definierten Gitters in einem gemeinsamen Datum darstellen.

Erfüllungskriterium

Die Interpolation muss in einem bestimmten geographischen Bereich und aus unterschiedlichen Datenquellen ein einheitliches Modell generieren.

Anforderung 5.3.3.3.2

"Terrain data shall be provided for Area 1."

Interpretation

Für Gebiet 1 **müssen** Geländedaten vorliegen. Dies kann durch Interpolation aus mehreren Datenquellen erreicht werden.

Erfüllungskriterium

Die Interpolation muss in einem bestimmten geographischen Bereich und aus unterschiedlichen Datenquellen ein einheitliches Modell generieren.

Anforderung 5.3.3.3.3

"For aerodromes regularly used by international civil aviation, terrain data shall be provided for:

- a Area 2a;*

b the take-off flight path area; and

c an area bounded by the lateral extent of the aerodrome obstacle limitation surfaces."

Interpretation

Für Gebiet 2a, die Startbahnfläche und einen durch die Hindernisdatenerfassungsfläche abgedeckten Bereich **müssen** Geländedaten vorliegen. Graphik 4.1 visualisiert das Konzept der Gelände- und Hindernisdatenerfassungsfläche für die Gebiete 1 und 2. Die Hindernisdatenerfassungsfläche identifiziert signifikante Hindernisse um die Start- und Landebahnen. Generell kann durch Interpolation aus mehreren Datenquellen ein flächendeckendes Modell generiert werden.

Erfüllungskriterium

Die Interpolation muss in einem bestimmten geographischen Bereich und aus unterschiedlichen Datenquellen ein einheitliches Modell generieren.

Anforderung 5.3.3.3.4

"Recommendation.— For aerodromes regularly used by international civil aviation, additional terrain data should be provided within Area 2 as follows:

a in the area extending to a 10-km radius from the ARP; and

b within the area between 10 km and the TMA boundary or a 45-km radius (whichever is smaller), where terrain penetrates a horizontal terrain data collection surface specified as 120 m above the lowest runway elevation."

Interpretation

Für Gebiet 2 **sollten** Geländedaten vorliegen:

a Innerhalb eines Umkreises von 10km um den Flugplatzbezugspunkt

b Innerhalb eines Umkreises von 45km um den Flugplatzbezugspunkt oder der Nahverkehrsgrenze, sofern eine horizontale Geländedatenerfassungsfläche 120m über der niedrigsten Landebahnhöhe überschritten wird.

Das Konzept der Geländedatenerfassungsfläche ist in Graphik 4.1 dargestellt. Die Geländedatenerfassungsfläche identifiziert signifikante Geländehöhen um einen Flughafen und markiert den Übergang zu einer anderen Datenqualitätsstufe. Für Geländedaten, die sich innerhalb von Gebiet 1 befinden und die Geländedatenerfassungsfläche durchstoßen, gelten die Genauigkeitsanforderungen von Gebiet 2. Da die Datenerfassungsfläche sicherheitskritische Grenzwerte markiert, sollte die Interpolation verlässliche Schätzwerte liefern.

Erfüllungskriterium

Die Interpolation muss in einem bestimmten geographischen Bereich und aus unterschiedlichen Datenquellen ein einheitliches Modell generieren. Die Interpolation muss bei gegebenen Gelände- und/oder Hindernisdatenerfassungsflächen eine Aussage ermöglichen, mit welcher

Wahrscheinlichkeit die betreffende Fläche durch Schätzwerte durchstoßen wird.

Anforderung 5.3.3.3.5

"Recommendation.— Arrangements should be made for coordinating the provision of terrain data for adjacent aerodromes where their respective coverage areas overlap to assure that the data for the same terrain is correct."

Interpretation

Für alle Gebiete **sollten** Vorkehrungen für benachbarte Flugplätze getroffen werden, bei denen sich die jeweiligen Abdeckungsgebiete überschneiden. Dabei soll sichergestellt werden, dass die Daten für dasselbe Hindernis oder Gelände korrekt sind. Die Interpolation muss die Zusammenführung von Datensätzen aus unterschiedlichen Quellen ermöglichen.

Erfüllungskriterium

Die Interpolation muss in einem bestimmten geographischen Bereich und aus unterschiedlichen Datenquellen ein einheitliches Modell generieren.

Anforderung 5.3.3.3.6

"Recommendation.— For those aerodromes located near territorial boundaries, arrangements should be made among States concerned to share terrain data."

Interpretation

Für alle Gebiet **sollten** Vorkehrungen für Flugplätze in der Nähe von Staatsgrenzen getroffen werden, sodass Gelände- und Hindernisdaten geteilt werden. Die Interpolation muss die Zusammenführung von Datensätzen aus unterschiedlichen Quellen ermöglichen.

Erfüllungskriterium

Die Interpolation muss in einem bestimmten geographischen Bereich und aus unterschiedlichen Datenquellen ein einheitliches Modell generieren.

Anforderung 5.3.3.3.7

"Recommendation.— For aerodromes regularly used by international civil aviation, terrain data should be provided for Area 3."

Interpretation

Für Gebiet 3 **sollten** Geländedaten vorliegen. Gebiet 3 und 4 beschreiben das unmittelbare Umfeld der Flug- und Landebahnen. Gelände- und Hindernisdatenerfassungsflächen sind hier ident:

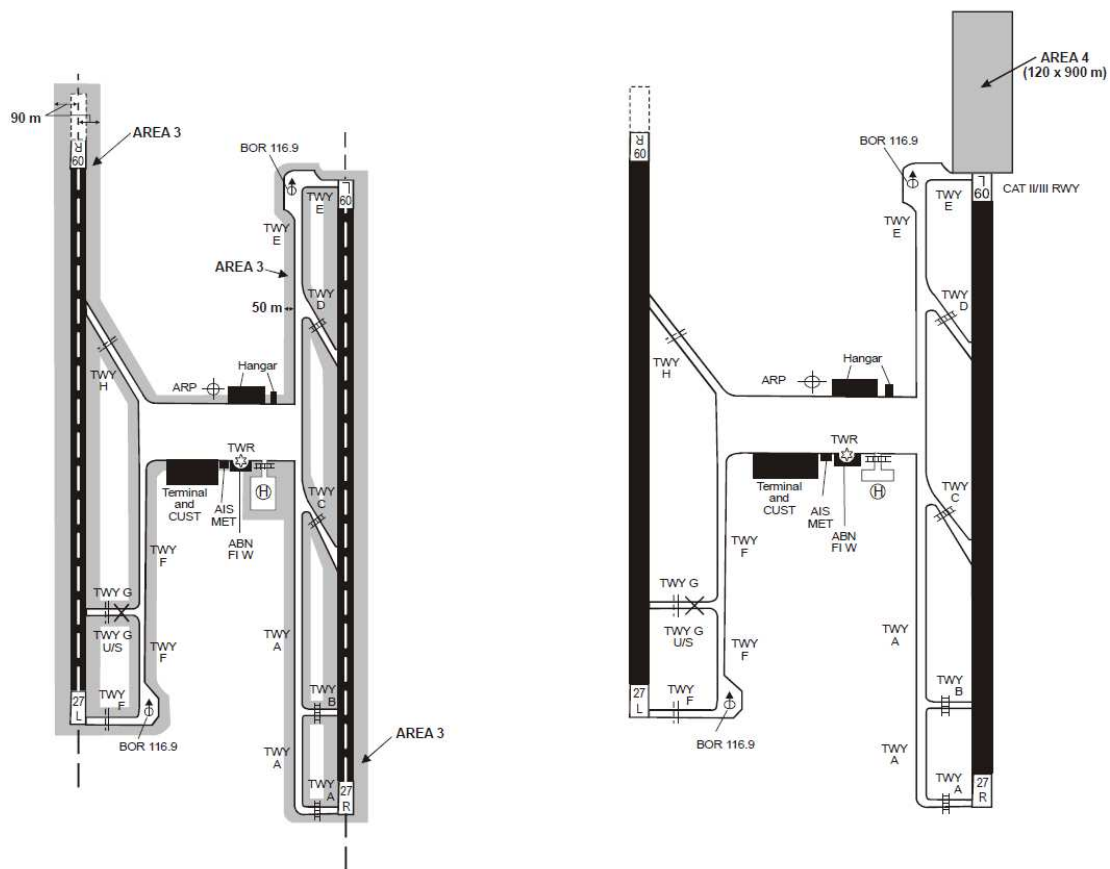


Abbildung 4.2: Gebiete 3 und 4 (ICAO, 2018b, Appendix 8)

Erfüllungskriterium

Die Interpolation muss in einem bestimmten geographischen Bereich und aus unterschiedlichen Datenquellen ein einheitliches Modell generieren.

Anforderung 5.3.3.3.8

"For aerodromes regularly used by international civil aviation, terrain data shall be provided for Area 4 for all runways where precision approach Category II or III operations have been established and where detailed terrain information is required by operators to enable them to assess the effect of terrain on decision height determination by use of radio altimeters."

Interpretation

Für Gebiet 4 **müssen** Geländedaten vorliegen, sofern es aus sicherheitskritischen Bestimmungen notwendig ist. Die Interpolation kann bei adäquater Datengrundlage als allgemeines Werkzeug verwendet werden, um Datensätze entsprechend den geltenden Bestimmungen zu erstellen.

Erfüllungskriterium

Die Interpolation muss in einem bestimmten geographischen Bereich und aus unterschiedlichen Datenquellen ein einheitliches Modell generieren.

Anforderung 5.3.3.3.9

"Recommendation.— Where additional terrain data is collected to meet other aeronautical requirements, the terrain data sets should be expanded to include this additional data."

Interpretation

Geländedaten **sollten** überall dort vorliegen, wo sie gebraucht werden, um aeronautischen Anforderungen gerecht zu werden. Interpolation kann zu Erweiterung bestehender Datensätze verwendet werden.

Erfüllungskriterium

Die Interpolation muss in einem bestimmten geographischen Bereich und aus unterschiedlichen Datenquellen ein einheitliches Modell generieren.

4.1.2 Anforderungen von PANS AIM

In diesem Abschnitt werden Anforderungen zu Inhalt, numerischen Anforderungen und Struktur von Geländedatensätzen evaluiert.

Anforderung 5.3.3.2.1.1

"A terrain grid shall be angular or linear and shall be of regular or irregular shape."

Interpretation

Ein Geländedatensatz **muss** winkelförmig oder linear und **muss** regelmäßig oder unregelmäßig geformt sein. Durch Interpolation können Punkte in beliebiger Anordnung geschätzt werden. Meistens werden Punkte in einem Raster mit einheitlicher Zellengröße bestimmt und in einem gängigen Rasterdateiformat zur Verfügung gestellt.

Erfüllungskriterium

Die Interpolation muss in einem bestimmten geographischen Bereich und aus unterschiedlichen Datenquellen ein einheitliches Modell generieren.

Anforderung 5.3.3.2.1.2

"Sets of terrain data shall include spatial (position and elevation), thematic and temporal

aspects for the surface of the Earth containing naturally occurring features such as mountains, hills, ridges, valleys, bodies of water, and permanent ice and snow, and exclude obstacles. Depending on the acquisition method used, this shall represent the continuous surface that exists at the bare Earth, the top of the canopy or something in-between, also known as “first reflective surface”.

Interpretation

Ein Geländedatensatz **muss** räumliche (Position und Höhe), thematische und temporale Aspekte der Erdoberfläche enthalten.

Erfüllungskriterium

Die Interpolation muss das reale Gelände möglichst genau annähern.

Anforderung 5.3.3.2.1.3

“In terrain data sets, only one feature type, i.e. terrain, shall be provided. Feature attributes describing terrain shall be those listed in Appendix 6, Table A6-1. The terrain feature attributes listed in Appendix 6, Table A6-1 represent the minimum set of terrain attributes, and those annotated as mandatory shall be recorded in the terrain data set.”

Interpretation

Ein Geländedatensatz **muss** nur den Feature-Typ Gelände enthalten. Der Datensatz muss nach ICAO (2018b, Appendix 6) Anforderungen bezüglich notwendiger Geländeattribute erfüllen. EUROCONTROL (2019, Kapitel 3.7.5) enthält nähere Beschreibungen der Attribute. Die Attributwerte können auf drei Datenebenen angegeben werden (vgl. EUROCONTROL, 2019, S. 45)

- Datensatz-Ebene: Ein Datensatz besteht potenziell aus mehreren Geländegebieten (Gebiet 1, 2, 3, 4).
- Geländegebiet-Ebene: Beschreibt Daten für ein einziges Geländegebiet.
- Daten-Ebene: Beschreibt die Summe der gegebenen Messwerte für ein Gebiet.

Die notwendigen und optionalen Informationen für Geländedatensätze sind in folgender Tabelle zusammengefasst (vgl. ICAO, 2018b, Appendix 6):

Geländeattribut	Notwendig/Optional	Beschreibung
Area of Coverage	Notwendig	Angabe von Bounding-Boxes für den gesamten Datenbereich und enthaltener Gebiete.

Geländeattribut	Notwendig/Optional	Beschreibung
Data originator identifier	Notwendig	Angabe der Entität (z.B. Ingenieurkonsulent), welche für die Erstellung des Datensatzes und enthaltener Gebiete verantwortlich ist.
Data source identifier	Notwendig	Angabe der Entität(z.B. Behörde), die für die Bereitstellung des Datensatzes und enthaltener Gebiete verantwortlich ist.
Acquisition method	Notwendig	Angabe über verwendete Methoden zur Erstellung des Datensatzes.
Post spacing	Notwendig	Angabe der horizontalen Distanz zwischen Höhenwerten in beide Richtungen des Rasters.
Horizontal reference system	Notwendig	Angabe des horizontalen Referenzsystems.
Horizontal resolution	Notwendig	Angabe der Anzahl der Nachkommastellen der Lagekoordinaten.
Horizontal accuracy	Notwendig	Angabe der maximal erlaubten horizontalen Abweichung zwischen gemessenen und wahren Höhenwerten mit einer vorgegebenen Wahrscheinlichkeit.
Horizontal confidence level	Notwendig	Angabe der Wahrscheinlichkeit, dass sich die wahren Werte der Position innerhalb eines bestimmten Bereichs um den geschätzten Wert befinden.
Horizontal position	Notwendig	Angabe der Lagekoordinaten.
Elevation	Notwendig	Angabe der Höhenwerte.

Geländeattribut	Notwendig/Optional	Beschreibung
Elevation reference	Notwendig	Angabe des Referenzpunktes des Höhenwertes (z.B. Mittelpunkt einer Rasterzelle).
Vertical reference system	Notwendig	Angabe des vertikalen Referenzsystems.
Vertical resolution	Notwendig	Angabe der Anzahl der Nachkommastellen der Höhenwerte.
Vertical accuracy	Notwendig	Angabe der maximal erlaubten vertikalen Abweichung zwischen gemessenen und wahren Höhenwerten mit einer vorgegebenen Wahrscheinlichkeit.
Vertical confidence level	Notwendig	Angabe der Wahrscheinlichkeit, dass sich die wahren Höhenwerte innerhalb eines bestimmten Bereichs um den geschätzten Wert befinden.
Surface type	Optional	Angabe über den Geländetyp an den gemessenen Punkten (z.B. Wasser, Permafrost, Stein, Sand, etc.).
Recorded surface	Notwendig	Angabe über den Typ von digitalem Höhenmodell (z.B. bare-earth, Oberflächenmodell, etc.).
Penetration level	Optional	Angabe über die Eindringtiefe der jeweiligen Vermessungsmethoden (z.B. Baumkronen, Waldboden, etc.).
Known Variations	Variations	Angabe über vorhersagbare Änderungen der Daten (z.B. jahreszeitliche Änderungen).

Geländeattribut	Notwendig/Optional	Beschreibung
Integrity	Notwendig	Angabe einer Risiko-Klassifizierung für fehlerhafte Daten. Die Klassifizierung als 'routine' und 'essential' beschreiben eine geringe Wahrscheinlichkeit, dass unter Verwendung fehlerhafter Daten eine starke Gefährdung für Abflug, Flug und Landung besteht.
Date and time stamp	Notwendig	Angabe über den Zeitpunkt der Datenerstellung.
Unit of measurement used	Notwendig	Angabe über die verwendete Einheit (z.B. Meter, Fuß, etc.).

Anhand dieser Tabelle kann zwischen numerischen und nicht-numerischen Metadaten unterschieden werden. Da der Interpolationsalgorithmus selbst keine qualitative Aussage über nicht-numerische Attribute des geschätzten Modells erlaubt, müssen diese im Rahmen der Datenbereitstellung vom Analysten evaluiert werden. Soll eine Interpolation bei der Erstellung eines Datenprodukts zum Einsatz kommen, muss die Ableitung der numerischen Attribute ermöglicht werden. Da die Datensätze aus unterschiedlichen Quellen stammen können, sollten Position, Messwert, Genauigkeiten und Konfidenzen auf Datenebene (pro Rasterzelle) angegeben werden. Streng genommen können die von der ICAO geforderten Genauigkeitsanforderungen durch eine Kreuzvalidierung berechnet werden. Dabei wird jedoch jedem Rasterpunkt ein datensatzweites Genauigkeitsmaß zugeordnet. Aus diesem Grund ist es sinnvoll, die Interpolation mit einem stochastischen Modell zu erweitern, um punktuelle Aussagen liefern zu können.

Erfüllungskriterium

Die Interpolation muss Aussagen über notwendige numerische Attribute des generierten Datensatzes ermöglichen:

- Geltungsbereich
- Stützwerteabstand
- Horizontale Auflösung
- Horizontale Genauigkeit
- Horizontales Konfidenzintervall
- Horizontale Position

- Höhenwert
- Vertikale Auflösung
- Vertikale Genauigkeit
- Vertikales Konfidenzintervall

Anforderung 5.3.3.2.1.4

"Terrain data for each area shall conform to the applicable numerical requirements in Appendix 1."

Interpretation

Geländedaten **müssen** für jedes Gebiet den numerischen Anforderungen von ICAO (2018b, Appendix 1) entsprechen. Weiters wird von EUROCONTROL (2019, Kapitel 3.7.5) auch eine Empfehlung für die horizontale Genauigkeit gegeben. In folgender Tabelle sind alle numerischen Anforderungen zusammengefasst (vgl. ICAO, 2018b, Appendix 1 und EUROCONTROL, 2019, Kapitel 3.7.5):

	Gebiet 1	Gebiet 2	Gebiet 3	Gebiet 4
Post spacing	3 Bogensekunden (ca. 90m)	1 Bogensekunden (ca. 30m)	0,6 Bogensekunden (ca. 20m)	0,3 Bogensekunden (ca. 9m)
Vertical accuracy	30m	3m	0,5m	1m
Vertical resolution	1m	0,1m	0,01m	0,1m
Horizontal accuracy	50m	5m	0,5m	2,5m
Horizontal resolution	1m	0,1m	0,01m	0,01m
Confidence Level	90%	90%	90%	90%
Integrity classification	routine	routine	essential	essential
Maintenance period	as required	as required	as required	as required

Diese Tabelle gibt Grenzwerte für numerische Attribute in verschiedenen geographischen Gebieten an. Prinzipiell ist es Sache des Analysten zu entscheiden, in welchem Bereich und auf Basis welcher Daten eine Interpolation durchgeführt werden soll. Idealerweise erlaubt die Interpolationsmethode eine Aussage darüber, ob Gefahr besteht, dass die entsprechenden

Grenzwerte überschritten werden. Der Nachweis sollte statistische Aussagekraft haben.

Erfüllungskriterium

Die Interpolation muss ausweisen, wenn das Ergebnis gewisse numerische Grenzwerte überschreitet:

- Stützwerteabstand
- Vertikale Genauigkeit
- Vertikale Auflösung
- Horizontale Genauigkeit
- Horizontale Auflösung
- Konfidenzgrad

4.2 Vorschlag geeigneter Interpolationsverfahren

Durch die Zuweisung von Erfüllungskriterien an die ICAO Anforderungen lässt sich für die Wahl der optimalen Interpolationsmethode ein eindeutiges Anforderungsprofil zuweisen.

Erfüllungskriterium	Anforderung	Empfehlung	Interpolation
Die Interpolation muss in einem bestimmten geographischen Bereich und aus unterschiedlichen Datenquellen ein einheitliches Modell generieren.	5.3.3.1, 5.3.3.3.1, 5.3.3.3.2, 5.3.3.3.3, 5.3.3.3.8, 5.3.3.3.9, 5.3.3.2.1.1	5.3.3.2, 5.3.3.3.4, 5.3.3.3.5, 5.3.3.3.6, 5.3.3.3.7	<ul style="list-style-type: none"> • TIN • Gleitende Mittelbildung • Variationsansatz • ANUDEM • Regression • Interpolation nach kleinsten Quadraten • Multiquadratische Interpolation • Kriging

Erfüllungskriterium	Anforderung	Empfehlung	Interpolation
Die Interpolation muss das reale Gelände möglichst genau annähern.	5.3.3.2.1.2	-	<ul style="list-style-type: none"> • Gleitende Mittelbildung • Variationsansatz • ANUDEM • Interpolation nach kleinsten Quadraten • Multiquadratische Interpolation • Kriging
Die Interpolation muss bei gegebenen Gelände- und/oder Hindernisdatenerfassungsflächen eine Aussage darüber treffen können, mit welcher Wahrscheinlichkeit die betreffende Fläche durch Schätzwerte durchstoßen wird.	-	5.3.3.3.4	<ul style="list-style-type: none"> • Interpolation nach kleinsten Quadraten • Kriging
<p>Die Interpolation muss Aussagen über notwendige numerische Attribute des generierten Datensatzes ermöglichen:</p> <ul style="list-style-type: none"> • Geltungsbereich • Stützwerteabstand • Horizontale Auflösung • Horizontale Genauigkeit • Horizontales Konfidenzintervall • Horizontale Position • Höhenwert • Vertikale Auflösung • Vertikale Genauigkeit • Vertikales Konfidenzintervall 	5.3.3.2.1.3	-	<p>Mit Kreuzvalidierung:</p> <ul style="list-style-type: none"> • TIN • Gleitende Mittelbildung • Variationsansatz • ANUDEM • Regression • Interpolation nach kleinsten Quadraten • Multiquadratische Interpolation • Kriging <p>Ohne Kreuzvalidierung:</p> <ul style="list-style-type: none"> • Interpolation nach kleinsten Quadraten • Kriging

Erfüllungskriterium	Anforderung	Empfehlung	Interpolation
<p>Die Interpolation muss ausweisen, wenn das Ergebnis gewisse numerische Grenzwerte überschreitet:</p> <ul style="list-style-type: none"> • Stützwerteabstand • Vertikale Genauigkeit • Vertikale Auflösung • Horizontale Genauigkeit • Horizontale Auflösung • Konfidenzgrad 	5.3.3.2.1.4	-	<p>Mit Kreuzvalidierung:</p> <ul style="list-style-type: none"> • TIN • Gleitende Mittelbildung • Variationsansatz • ANUDEM • Regression • Interpolation nach kleinsten Quadraten • Multiquadratische Interpolation • Kriging <p>Ohne Kreuzvalidierung:</p> <ul style="list-style-type: none"> • Interpolation nach kleinsten Quadraten • Kriging

Auf Basis der Zuordnung der ICAO-Anforderung zu konkreten Erfüllungskriterien und geeigneten Interpolationsmethoden wurde ein klares Anforderungsprofil erstellt. Die entscheidenden Kriterien zur Eingrenzung der optimalen Interpolationsverfahren sind die Forderung nach möglichst hoher Genauigkeit sowie die Schätzung von Genauigkeitsinformationen und Konfidenzintervallen. Zur Erfüllung der ICAO-Anforderungen wird streng genommen kein stochastisches Modell benötigt. Das Verfahren der Kreuzvalidierung liefert eine datensatzweite Einschätzung der Qualität des Interpolationsmodells anhand von Referenzmessungen, die auf Datenebene angegeben werden kann. Im Zuge der Evaluierung wurde jedoch deutlich, dass statistische Genauigkeitsinformationen auf Datenebene viele Vorteile mit sich bringen. Daraus empfiehlt sich ein kombinierter Ansatz aus stochastischem Modell und Kreuzvalidierung. Eine vorteilhafte Eigenschaft ist speziell die Möglichkeit der Ableitung von punktwisen Wahrscheinlichkeitsgrenzen für die Überschreitung von Grenzwerten.

Generell werden Kriging und Splines als vielversprechendste Methoden für den Anwendungsfall dieser Diplomarbeit erachtet und in weiterer Folge praktisch evaluiert.

5 Evaluierung von Interpolationsverfahren

Im letzten Abschnitt 4 wurde zusammengefasst, dass Splines und Kriging die wohl *optimalen* Mittel zur Erfüllung des ICAO-Anforderungsprofils bereitstellen. Daher wird in diesem Abschnitt die praktische Eignung zur Erfüllung der ICAO-Anforderungen evaluiert. Im Rahmen dieser Untersuchungen wird die Software ArcGIS Pro 2.4.3 verwendet. Die Strategie zur Evaluierung der Interpolationsverfahren setzt auf die Durchführung einer explorativen Datenanalyse (EDA) sowie der Evaluierung von Interpolationsmethoden anhand von Beispieldaten.

5.1 Datengrundlage

Im ICAO-Anforderungsprofil erfolgt eine Einteilung in verschiedene Genauigkeitsstufen für Geländedaten (Gebiet 1, 2, 3, 4). Die Erstellung und Qualitätssicherung der jeweiligen Gebiete ist in erster Linie von den Basisdaten abhängig. Um die Genauigkeitsanforderungen der Gebiete erreichen zu können, müssen qualitätsgesicherte Referenzdaten vorliegen. Diese sollten mindestens um den Faktor 3 genauer als das zu erstellende DHM sein. Eine optimale Interpolationsmethode sollte jedoch unabhängig vom Basisdatensatz, die besten Ergebnisse im Vergleich zu andern Interpolationsmethoden liefern. Aus diesem Grund wird ein vom Bundesland Niederösterreich frei zugänglich gemachter 10x10m Rasterdatensatz verwendet:

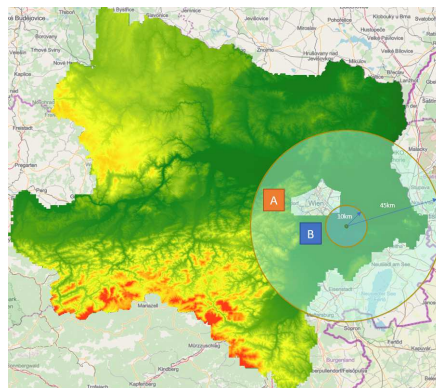


Abbildung 5.1: Digitales Geländemodell: Niederösterreich, 10x10m

In Abbildung 5.1 ist der Referenzpunkt des Flughafens Wien-Schwechat eingezeichnet. Dieser Punkt markiert das Zentrum sicherheitskritischer Bereiche (vgl. Abbildung 4.1):

- Umkreisradius 10km: Anforderungen für ICAO Gebiet 2
- Umkreisradius 45km: Anforderungen für ICAO Gebiet 1, wenn Daten unter einer Geländedatenerfassungsfläche (120m über dem niedrigsten Punkt der Landebahn) liegen und ICAO Gebiet 2, wenn sie darüber liegen.

Für die praktische Evaluierung werden zwei Untersuchungsgebiete mit unterschiedlicher Topographie im Ausmaß von 10x10km gewählt. Aus Gründen der Datenprozessierung wird das Gebiet innerhalb des relevanten Umkreisradius von 45km ausgeschnitten:

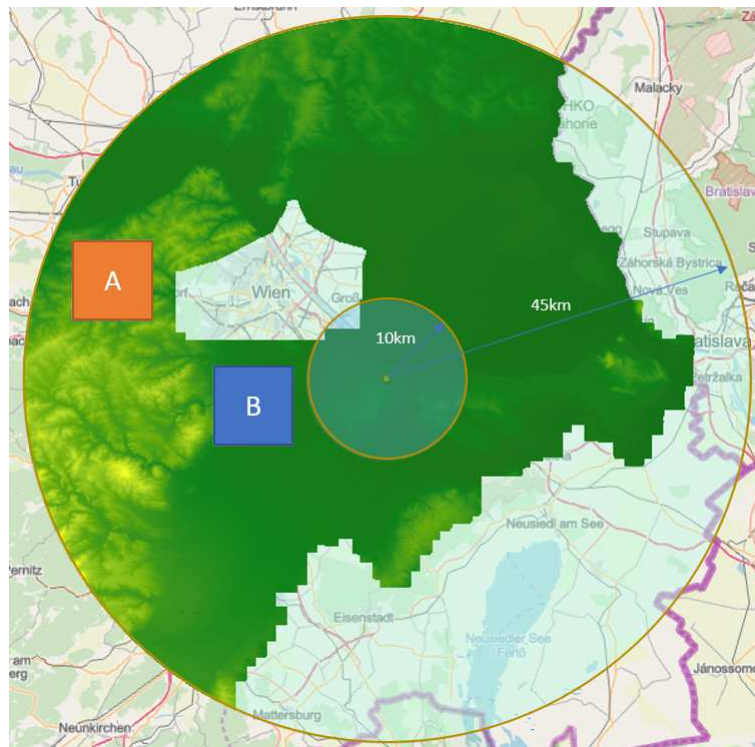


Abbildung 5.2: DHM Wien-Schwechat: Radius 45km

Für jedes Untersuchungsgebiet werden per Zufallsgenerator 1000 Referenzpunkte mit einem Mindestabstand von 100m generiert. Diese Punkte werden im Zuge dieser Analyse als Referenzpunkte zur Ableitung von Interpolationsflächen und für den Genauigkeitsvergleich der Methoden verwendet:

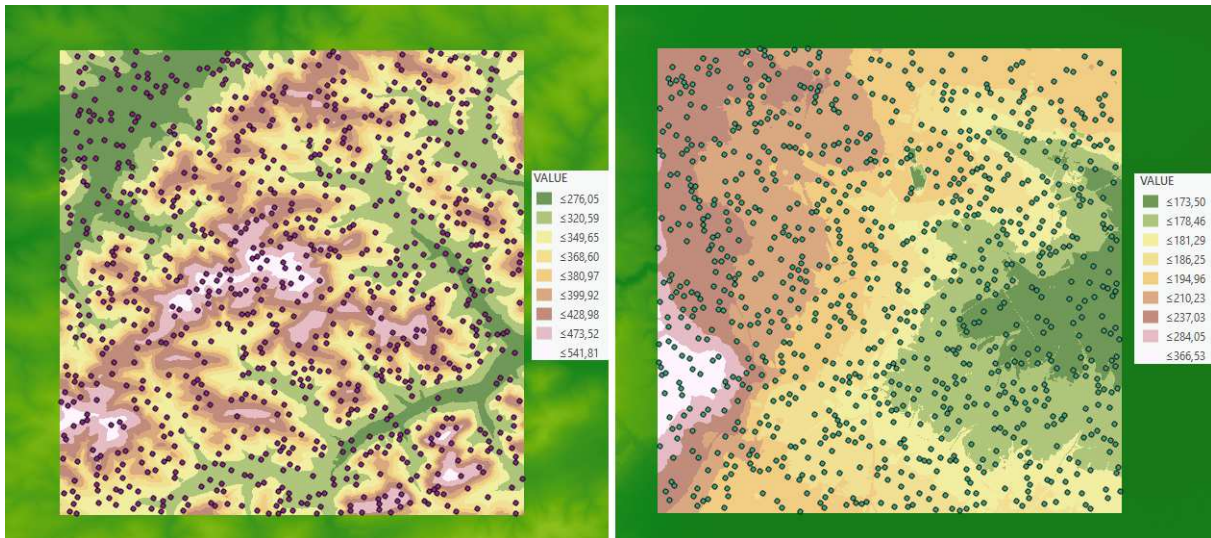


Abbildung 5.3: Untersuchungsgebiete: A (links) und B (rechts)

In Gebiet A ist die Topographie allgemein sehr variabel und die relative Höhe schwankt bis zu 334m. In Gebiet B zeigt sich eine regelmäßigeren Topographie mit einer relativen Höhenschwankung von bis zu 202m. Letztere ist vor allem durch eine im Bild westlich gelegene Erhöhung bedingt. Der Großteil der Höhenwerte hat in Gebiet B eine Schwankungsbreite von lediglich 40m. Allgemein liegt für beide Untersuchungsgebiete eine gut verteilte Punktwolke vor. Da die Referenzpunkte direkt aus einem qualitätsgesicherten Datensatz stammen, wird auf eine Ausreißeranalyse verzichtet.

5.2 Explorative Datenanalyse

Interpolationsverfahren arbeiten auf Basis mathematischer und statistischer Annahmen und liefern nur unter Einhaltung dieser Bedingungen optimale Ergebnisse. Daher werden die Daten in Hinblick auf ihre Verteilung, vorliegende Trends und die allgemeine räumliche Struktur untersucht.

Datenverteilung

Im ersten Schritt wird die allgemeine Verteilung der Daten analysiert, um Information über das Lage- und Streuungsverhalten zu bekommen. Die Visualisierung eines Histogramms und eines QQ-Plots zeigt, dass die Datenverteilung in Gebiet A näherungsweise normalverteilt ist (siehe Abbildung 5.4). Die meisten Daten liegen zwischen 290 und 410 Höhenmetern. Die symmetrische Gruppierung um den Mittelwert zeigt, dass generell eine ausgeglichene Variabilität in den Daten vorhanden ist. Der kontinuierliche Verlauf der Datenverteilung

bis hin zu den Extremwerten lässt gleichzeitig auf runde Übergänge zu den Extremwerten schließen.

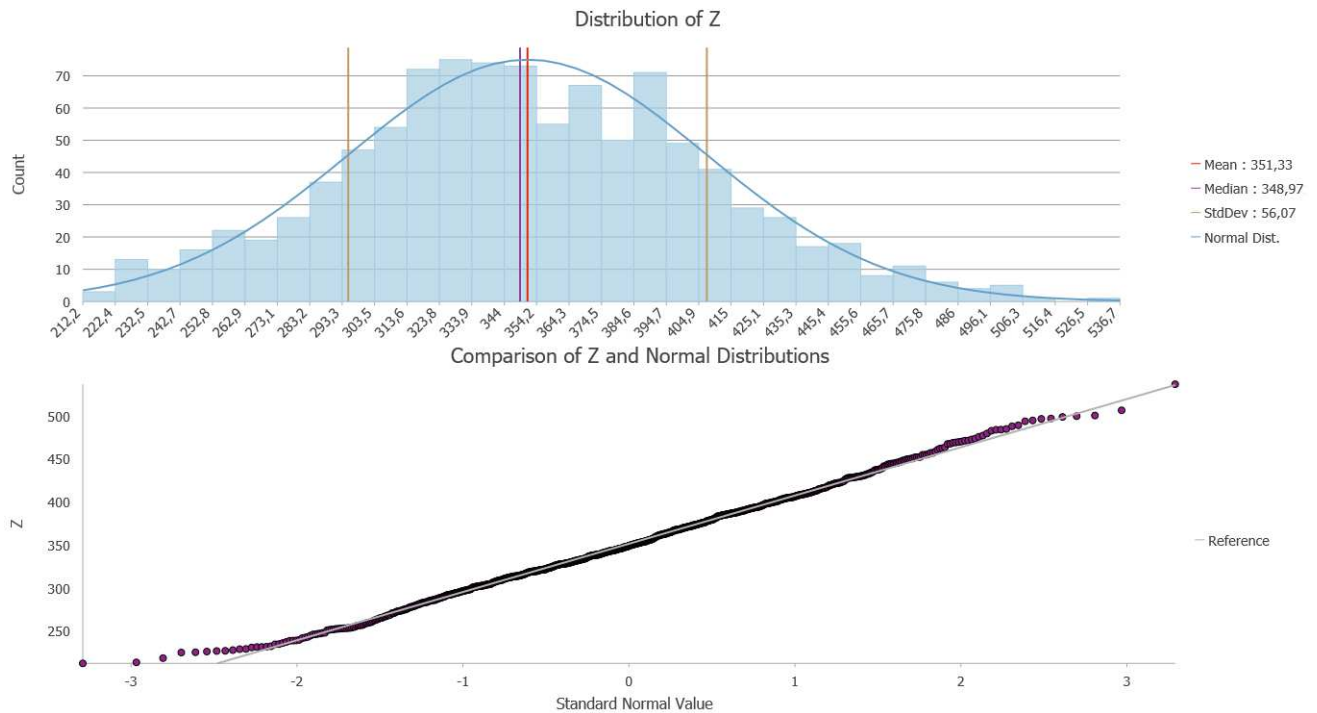


Abbildung 5.4: Gebiet A: Datenverteilung

Für Gebiet B ergibt sich eine größere Abweichung zur Normalverteilung (vgl. Abbildung 5.5). Mittelwert und Median liegen im Vergleich zu Gebiet A verhältnismäßig weit auseinander. Auch der QQ-Plot verdeutlicht vor allem bei höheren Werten eine größere Abweichung zur Normalverteilung. Diese asymmetrische, rechts-schiefe Verteilung zeigt, dass sich die meisten Daten im Bereich von 170-200 Höhenmetern befinden. Dies lässt auf eine allgemein regelmäßige Topographie mit geringer Variabilität schließen. Auffällig ist jedoch der Maximalwert von 358 Höhenmetern, der auf eine Anomalie im Untersuchungsgebiet schließen lässt. Der Verlauf bis hin zum Maximalwert deutet auf einen steilen Anstieg hin.

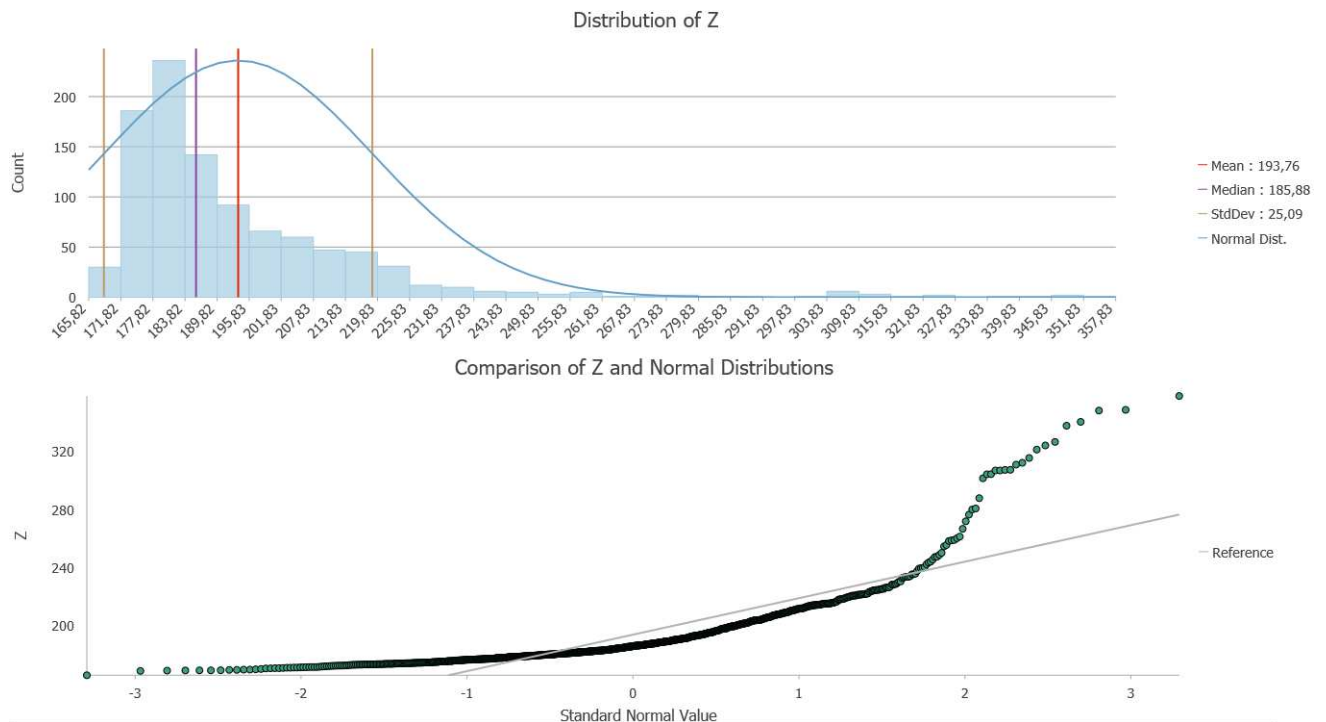


Abbildung 5.5: Gebiet B: Datenverteilung

Trendanalyse

Geostatistische Verfahren arbeiten unter der Annahme eines konstanten Mittelwertes im Untersuchungsgebiet. Gebiet A zeigt generell eine sehr variable Topographie mit Tälern und Hügeln:



Abbildung 5.6: Gebiet A: 3D-Ansicht

Eine Profilsicht der Rasterhöhenwerte mit zugehöriger Regression zeigt in Ost-West-Richtung keinen Trend. In Nord-Süd Richtung macht sich ein leichter Trend bemerkbar:

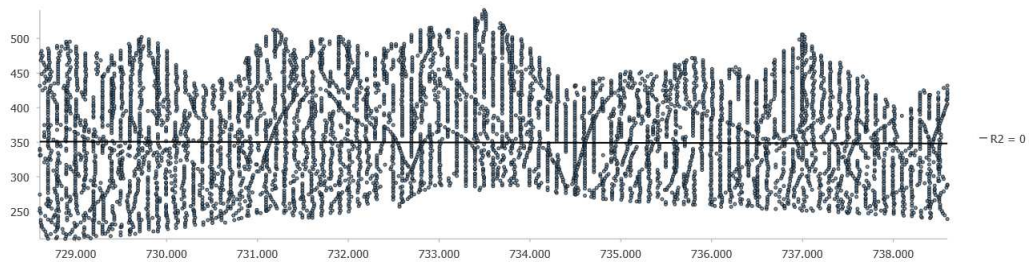


Abbildung 5.7: Gebiet A: Profil Ost-West

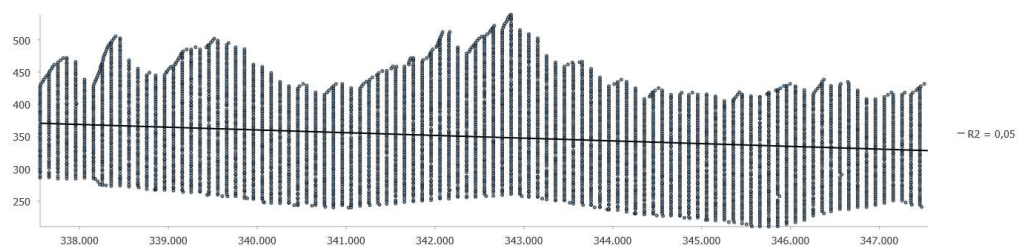


Abbildung 5.8: Gebiet A: Profil Süd-Nord

Der leichte Trend-Abstieg kann als Trendfunktion 1. Grads modelliert und in der Interpolation berücksichtigt werden:

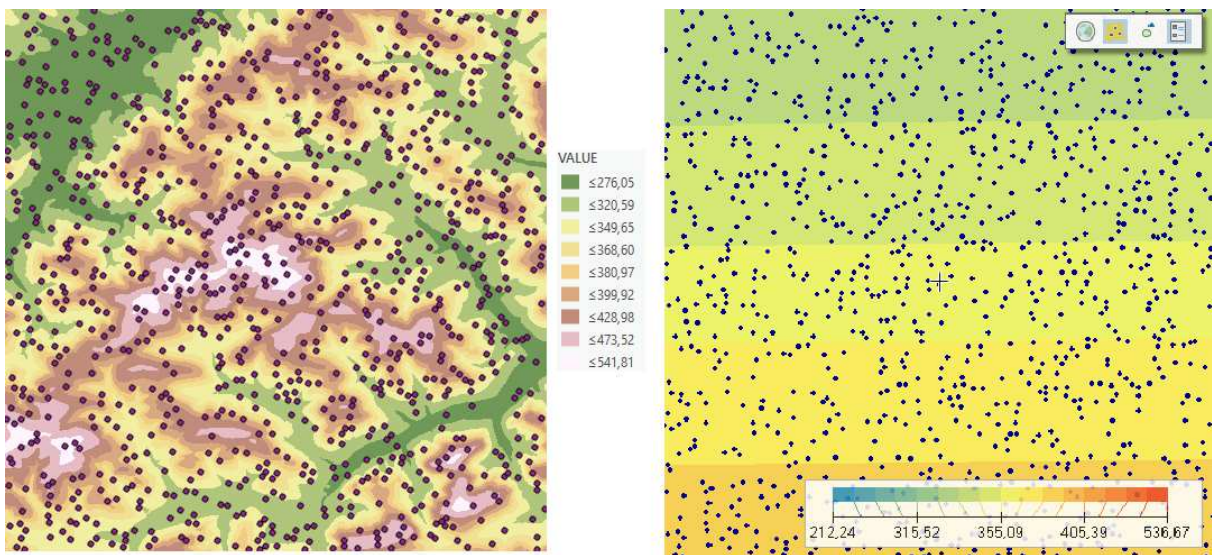


Abbildung 5.9: Gebiet A: Trend

Die 3D-Ansicht von Gebiet B zeigt einen gleichmäßigen Verlauf mit moderatem bis steilem Anstieg:

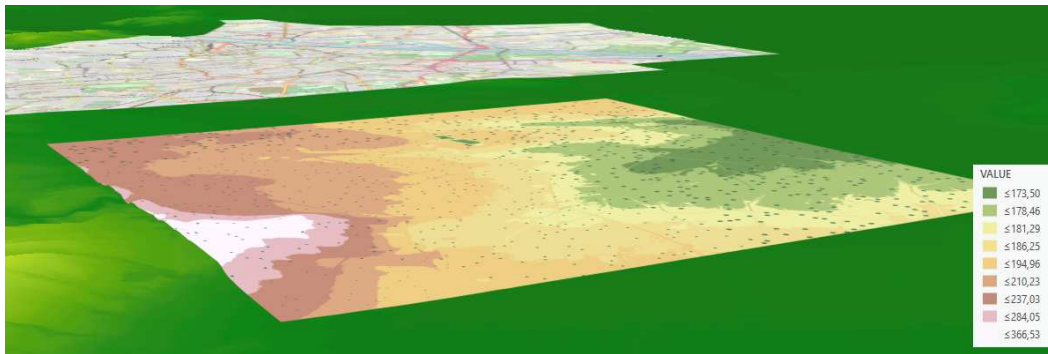


Abbildung 5.10: Gebiet B: 3D-Ansicht

Dies zeigt sich auch in der Profilansicht mit Regressionsgerade vom Ost-West-Profil. Dabei ist auch der verhältnismäßig steile Anstieg im Osten gut zu erkennen. Im Süd-Nord-Profil zeichnet sich kein klarer Trend ab:

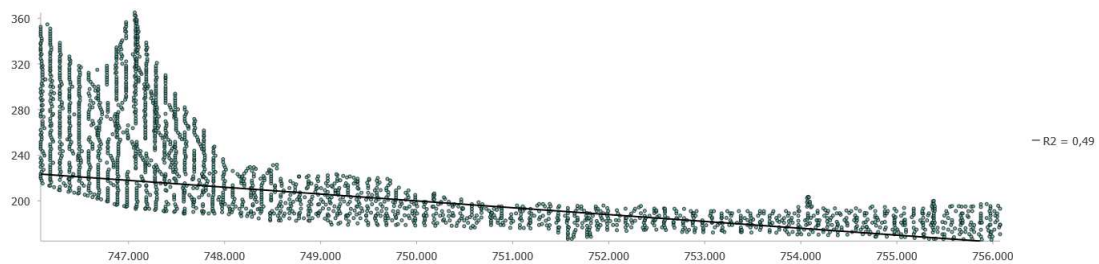


Abbildung 5.11: Gebiet A: Profil Ost-West

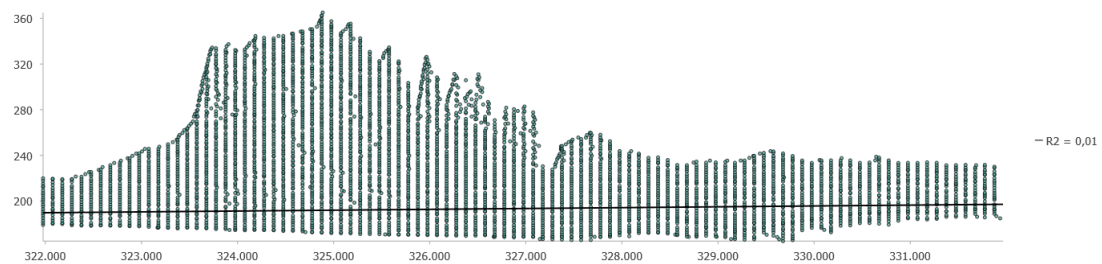


Abbildung 5.12: Gebiet A: Profil Süd-Nord

Der Trend-Abstieg im Ost-West-Profil wird ebenso durch eine Trendfunktion 1. Grades modelliert:

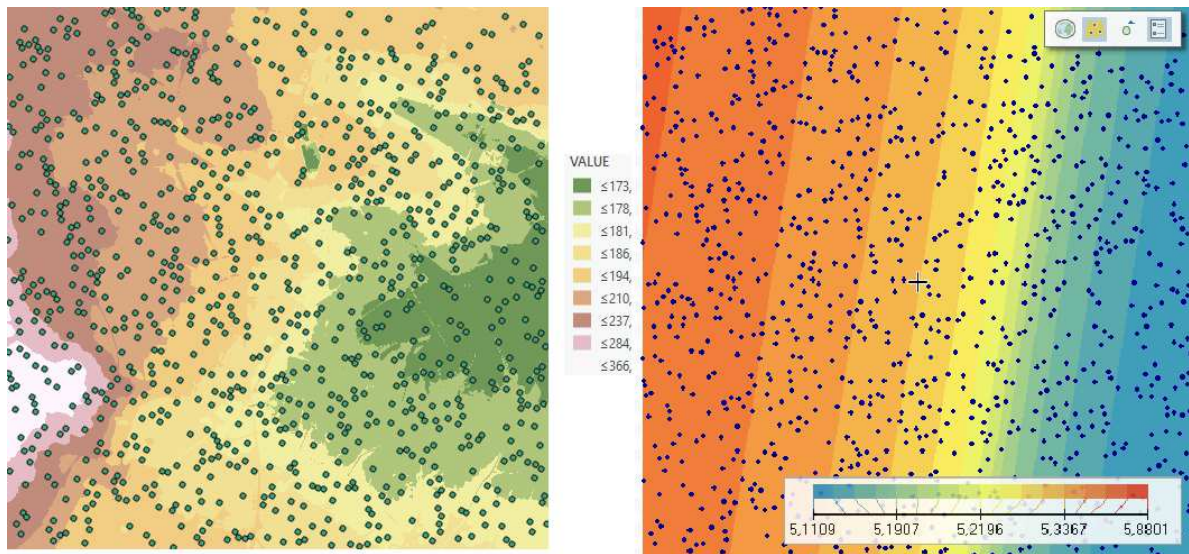


Abbildung 5.13: Gebiet B: Trend

Die berechneten Trendflächen für Gebiet A und Gebiet B werden für möglichst optimale Ergebnisse im Zuge des Interpolationsprozesses berücksichtigt.

Räumliche Struktur

Für die räumliche Strukturanalyse wird das Semivariogramm betrachtet. ArcGIS bietet Unterstützung bei der Wahl der Variogramm-Parameter zur Ableitung des theoretischen Variogramms. Für Gebiet A wird das Semivariogramm nach folgender Konfiguration erstellt:

Parameter	Wert
Lag Größe	202m
Anzahl der Lags	12
Modell	Stable
Modell-Parameter	1,5
Nugget	0,15m
Range	1612m
Sill	2535m
Messfehler	100% des Nuggets
Anisotropie	Nein

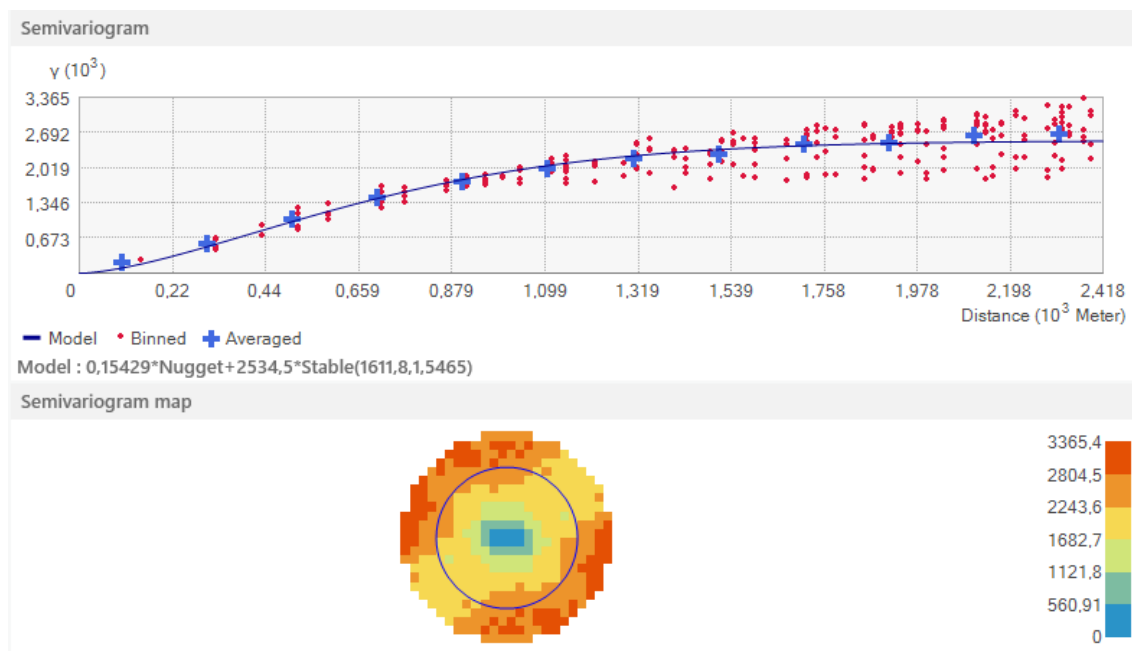


Abbildung 5.14: Gebiet A: Semivariogramm

Die Semivariogrammwerte werden in 12 Gruppen (lags) eingeteilt. Es wird das 'Stable' Modell (vgl. ESRI, 2001, S. 258) mit Parameter 1,5 verwendet. Ein Parameter von 1 kommt einem Exponential-Modell gleich, ein Parameter von 2 einem Gauß-Modell. Nugget, Range und Sill werden auf Basis der Datengrundlage angenommen. Es wird konfiguriert, dass 100% des Nugget-Effekts durch den Fehleranteil der Messfehler und nicht durch den Fehleranteil des stationären Prozesses (Micro-Scale Variation) verursacht wird. Dies hat ein geglättetes Endresultat zur Folge, das nicht durch die Datenpunkte verläuft, aber allgemein kleinere Standardabweichungen produziert (vgl ESRI, 2020d). Auf die Modellierung von Anisotropie wird verzichtet.

Für Gebiet B wird das Semivariogramm nach analogem Vorgehen erstellt. Aufgrund der unterschiedlichen Topographie sind Lag Größe, Modell-Parameter, Nugget, Range und Sill anders gewählt:

Parameter	Wert
Lag Größe	179m
Anzahl der Lags	12
Modell	Stable
Modell-Parameter	2
Nugget	3,0E-06m
Range	1434m

Parameter	Wert
Sill	3,0E-03m
Messfehler	100% des Nuggets
Anisotropie	Nein

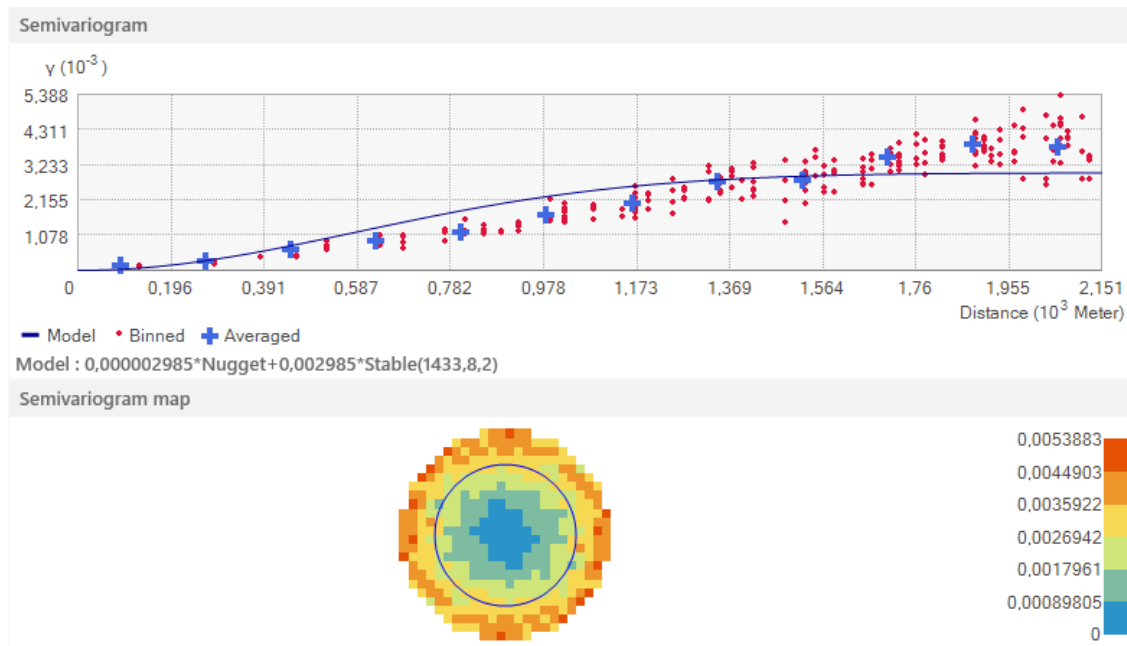


Abbildung 5.15: Gebiet B: Semivariogramm

Der Modellparameter 2 (entspricht Gauß-Modell) bewirkt, dass nahe beieinander liegende Nachbarn weniger stark korreliert und die Einflüsse weiter voneinander entfernter Punkte früher verringert werden.

5.3 Interpolation und Vergleich der Ergebnisse

Für den Vergleich der Interpolationsergebnisse werden die originalen Messdaten (aus denen die Referenzpunkte abgeleitet wurden) mit den Ergebnissen von Universal Kriging, Empirical Bayesian Kriging und Regularisierter Spline-Interpolation verglichen. Universal Kriging wird auf Basis der explorativen Datenanalyse durchgeführt. Für eine möglichst gute lokale Anpassung wird zusätzlich ein in 4 Sektoren eingeteilter Suchradius definiert, wobei pro Sektor 2-5 Punkte enthalten sein müssen. Empirical Bayesian Kriging (EBK) ist ein von ArcGIS

implementierter Algorithmus und automatisiert den Großteil der Parameterfindung für geostatistische Verfahren anhand von Teilmengenbildung und Semivariogramm-Simulationen. EBK liefert zudem genauere Standardfehler als andere Kriging-Methoden. Durch den Vergleich mit EBK soll die adäquate Durchführung der explorativen Datenanalyse verifiziert werden. Da Splines im Vergleich zu Kriging ähnlich gute Ergebnisse liefern sollen, wird eine Regularisierte Spine-Interpolation für eine Gegenüberstellung durchgeführt.

5.3.1 Interpolationsergebnisse

Gebiet A (Unregelmäßige Topographie)

Die Interpolationen mit Universal Kriging, EBK und Regularisierter Spline liefern folgende Ergebnisse bei unregelmäßiger Topographie:

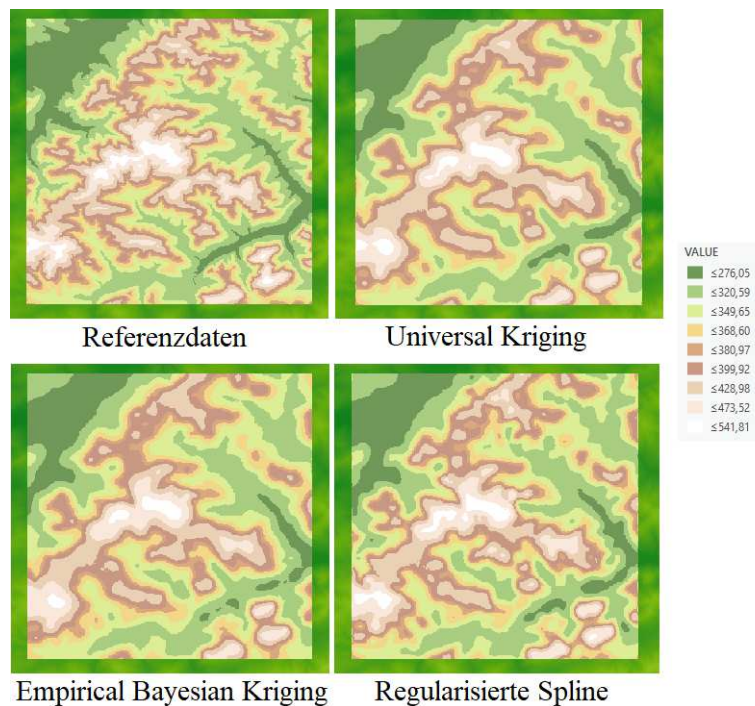


Abbildung 5.16: Gebiet A: Interpolationsergebnisse

Bei visueller Betrachtung lassen sich in den Ergebnissen nur minimale Unterschiede ausmachen. Einen quantitativen Vergleich liefert die Kreuzvalidierung:

Methode	RMSE [m]	Durchschnittlicher Standardfehler [m]
Universal Kriging	15,33	14,57
EBK	16,00	17,10
Regularisierte Spline	15,80	-

Der Vergleich zeigt, dass Universal Kriging mit vorangegangener explorativen Datenanalyse die vergleichsweise geringsten Abweichungen im RMSE zu den Referenzpunkten liefert. Der durchschnittliche Standardfehler unterschätzt die Abweichungen geringfügig. EBK liefert den vergleichsweise höchsten RMSE und überschätzt diesen auch im durchschnittlichen Fehler. Die Regularisierte Spline reiht sich mit dem RMSE zwischen Universal Kriging und EBK ein. Da der Spline-Interpolation kein stochastisches Modell zugrunde liegt, ist kein durchschnittlicher Standardfehler verfügbar.

Gebiet B (Regelmäßige Topographie mit Anstieg)

Die Interpolationen mit Universal Kriging, EBK und Regularisierter Spline liefern folgende Ergebnisse bei regelmäßiger Topographie mit Anstieg:

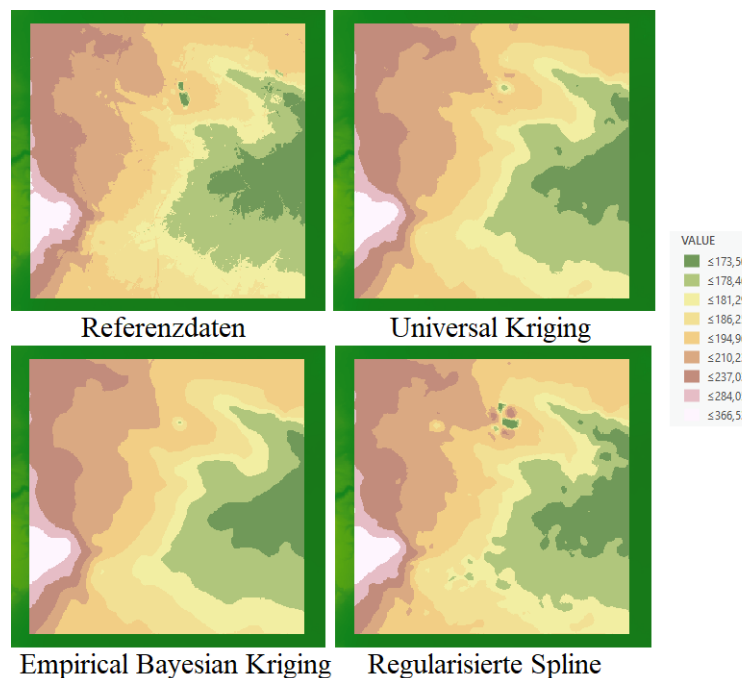


Abbildung 5.17: Gebiet B: Interpolationsergebnisse

Rein visuell betrachtet liefern wieder alle Methoden gute Resultate. Die stärksten Unterschiede lassen sich an den Übergängen zwischen den Höhenklassen erkennen. Einen quantitativen Vergleich liefert wiederum die Kreuzvalidierung:

Methoden	RMSE [m]	Durchschnittlicher Standardfehler [m]
Universal Kriging	2.57	0.69
EBK	2.82	2.97
Regularisierte Spline	3.91	-

Universal Kriging liefert wieder den geringsten RMSE, gefolgt von EBK und der Regularisierten Spline. Beim durchschnittlichen Standardfehler zeigt sich bei Universal Kriging eine Unterschätzung der tatsächlichen Abweichungen. Dem könnte mit einem verfeinerten Trendmodell entgegengewirkt werden. Die EBK-Methode überschätzt hingegen die tatsächlichen Abweichungen leicht. Dies wird dadurch verursacht, dass EBK anhand von vielen regionalen Untersuchungen und Semivariogramm-Simulationen versucht, die räumliche Korrelation bestmöglich zu erfassen. Die Regularisierte Spline liefert den verhältnismäßig größten RMSE. Allgemein fällt der große Unterschied im RMSE zwischen unregelmäßiger (Gebiet A $\approx 15\text{m}$) und regelmäßiger (Gebiet B $\approx 3\text{m}$) Topographie auf.

Varianzinformationen

Geostatistische Interpolationsverfahren liefern für jede Schätzung auch Genauigkeitsinformationen. Die Anforderungsanalyse in Abschnitt 4 hat gezeigt, dass Informationen bezüglich Genauigkeit auf Datenebene, also pro Höhenwert, mit einer Konfidenz von 90% angegeben werden müssen. Dies entspricht ungefähr dem 1.65-fachen der Standardabweichung. Diese Informationen basieren jedoch auf einem stochastischem Modell, nachdem nur dann aussagekräftige Ergebnisse zu erwarten sind, wenn die Interpolationsfehler einer Normalverteilung folgen. Aufschluss darüber geben beispielsweise die QQ-Plots der Gebiete A und B aus den Abweichungen der Kreuzvalidierung nach Universal Kriging (vgl. Abbildung 5.18):

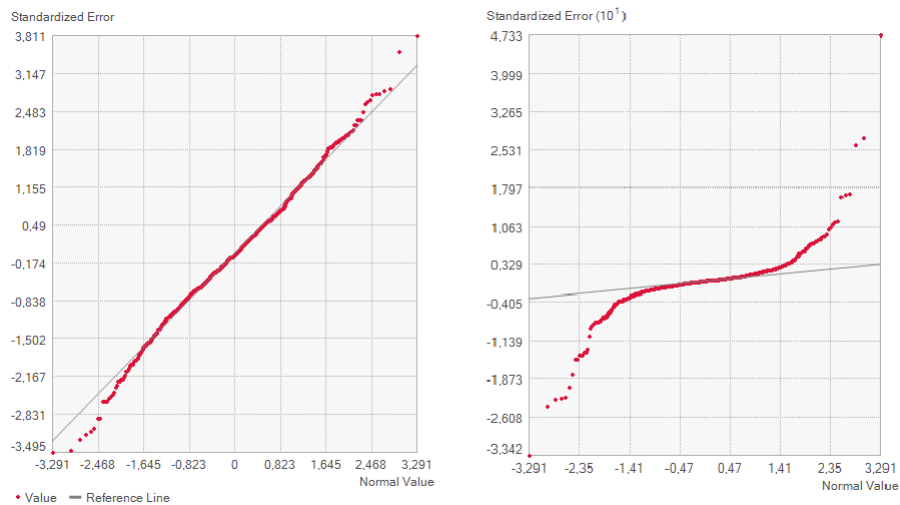


Abbildung 5.18: QQ-Plots für den standardisierten Fehler für UK: Gebiet A (links) und B (rechts)

Die Fehler in Gebiet A nähern eine Normalverteilung ganz gut an. In Gebiet B gibt es Abweichungen zur Normalverteilung, daher ist dort bei Betrachtung der geschätzten Standardabweichungen Vorsicht geboten. Allgemein fällt wiederum auf, dass in unregelmäßigem Gelände viel höherer Abweichungen geschätzt werden als bei moderaten Übergängen:



Abbildung 5.19: Geschätzte Standardabweichungen für UK: Gebiet A (links) und B (rechts)

In Gebiet A wird für Universal Kriging ein durchschnittlicher Standardfehler von 14,57m und für EBK von 17,10m geschätzt. In Gebiet B beträgt die Schätzung bei Universal Kriging 0,69m und bei EBK 2,97m. Generell ist die Gegenüberstellung von RMSE und durchschnittlichem Standardfehler stark zu empfehlen, da sich anhand dessen Modellverbesserungen ableiten lassen.

5.3.2 Qualitätsbeurteilung

Die ICAO fordert die Angabe von Genauigkeitsinformationen auf Datenebene (pro Höhenwert). Nach EUROCONTROL (2019) wird vorgeschlagen, die Eignung eines Datensatzes anhand von Stichproben und Konfidenzintervallen durchzuführen. Beispielsweise ergeben sich bei Universal Kriging anhand der Kreuzvalidierung für Gebiet A folgende Grenzwerte:

Kennwert	Grenzwert [m]
Mittlere Abweichung	11,33
0,5-Perzentil (Median)	8,43
0,7-Perzentil	13,70
0,9-Perzentil	25,40

Dies würde beispielsweise bedeuten, dass in 90% aller Fälle keine größere Abweichung als 25,40m auftritt. Umgekehrt kann durch den Eignungstest laut Formel 2.82 bei einem vorgegebenen Grenzwert von beispielsweise 30m (Anforderung für Gebiet 1) eine Konfidenz von 93,6% berechnet werden. Diese Berechnungen werden (ähnlich wie die Kreuzvalidierung selbst) mittels Referenzpunkten durchgeführt. Das Subset der Referenzpunkte ist im Grunde aber eine Stichprobe aus der Grundgesamtheit des 'wahren' Geländes. Da die Referenzpunkte einem Referenzraster entnommen wurden, kann mittels Rasterkalkulation die 'wahre' Abweichung der Interpolationsmodelle zum Ursprungsraster (Raster, aus dem die Messpunkte für die Interpolation abgeleitet wurden) berechnet werden. Dabei wird jeweils der Referenzraster vom Interpolationsraster subtrahiert und Absolutwerte der 'wahren' Abweichungen gebildet. Aus dem Fehlerraster mit den Werten Δ_i werden 'wahre' RMSE-Werte nach

$$RMSE_{wahr} = \sqrt{\frac{\sum \Delta_i^2}{n}} \quad (5.1)$$

berechnet und mit den RMSE-Werten der Kreuzvalidierung verglichen. Dabei wird evaluiert, ob die alleinige Qualitätsprüfung durch Kreuzvalidierung Probleme mit sich bringen kann.

Gebiet A (Unregelmäßige Topographie)

In der folgenden Visualisierung sind die Abweichungsraster für Universal Kriging, EBK und Regularisierter Spline mit zugehöriger Verteilung dargestellt:

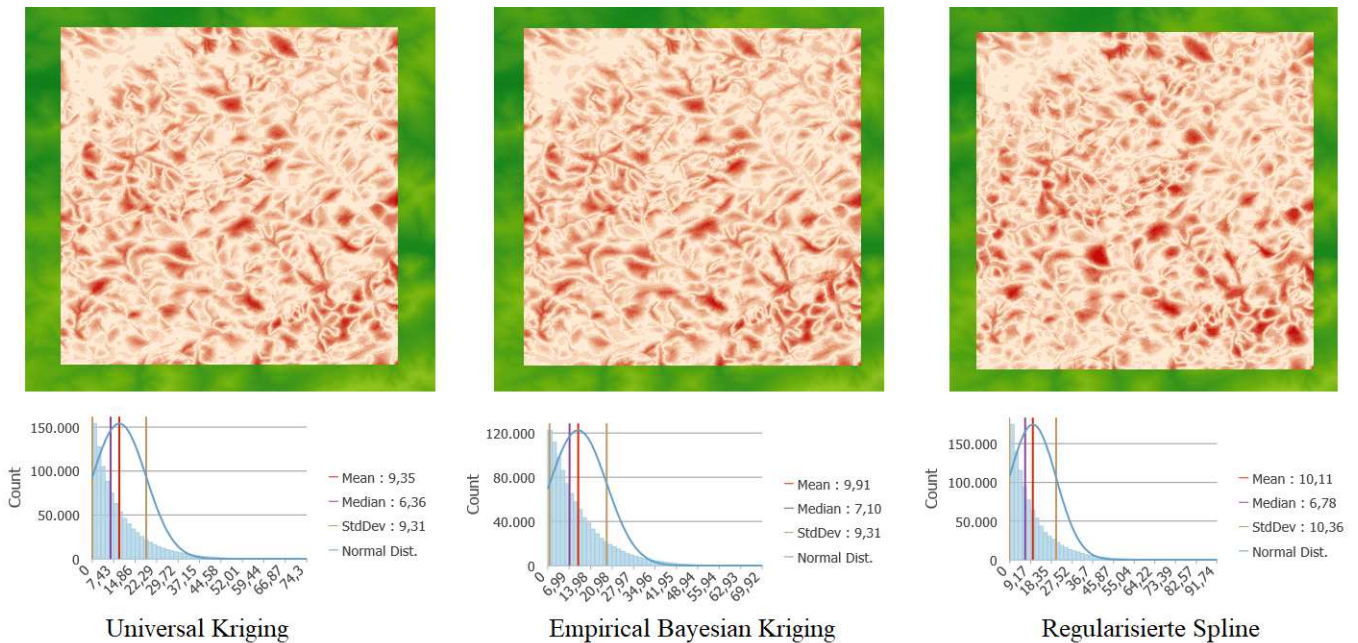


Abbildung 5.20: Gebiet A: Globale Abweichungen der Interpolationsmethoden

Bei dieser globalen Fehlerbetrachtung bestätigt sich, dass Universal Kriging mit explorativer Datenanalyse den geringsten mittleren Fehler verursacht. EBK erreicht den geringsten maximalen Fehler. Die Spline-Interpolation zeigt den größten mittleren und maximalen Fehler. Anhand des Fehlerrasters wird der 'wahre' RMSE dem RMSE der Kreuzvalidierung und dem durchschnittlichen Standardfehler gegenübergestellt:

Method	Wahrer RMSE [m]	Kreuzvalidierungs RMSE [m]	Durchschnittlicher Standardfehler [m]
Universal Kriging	13,20	15,33	14,57
EBK	13,60	16,00	17,10
Regularisierte Spline	14,48	15,80	-

Die Gegenüberstellung zeigt, dass die wahren RMSE in diesem Fall tendenziell niedriger als die der Kreuzvalidierung sind. Somit liegt bei Universal Kriging schlussendlich eine leichte Überschätzung der wahren Abweichungen vor. Weiters ergibt sich anhand des wahren RMS,

dass das EBK-Modell das Gelände etwas genauer annähert wie die Regularisierte Spline.

Gebiet B (Regelmäßige Topographie mit Anstieg)

Die Ableitung der Fehlerraster aus den Interpolationen von Universal Kriging, EBK und Regularisierter Spline ergibt für Gebiet B folgende Resultate:

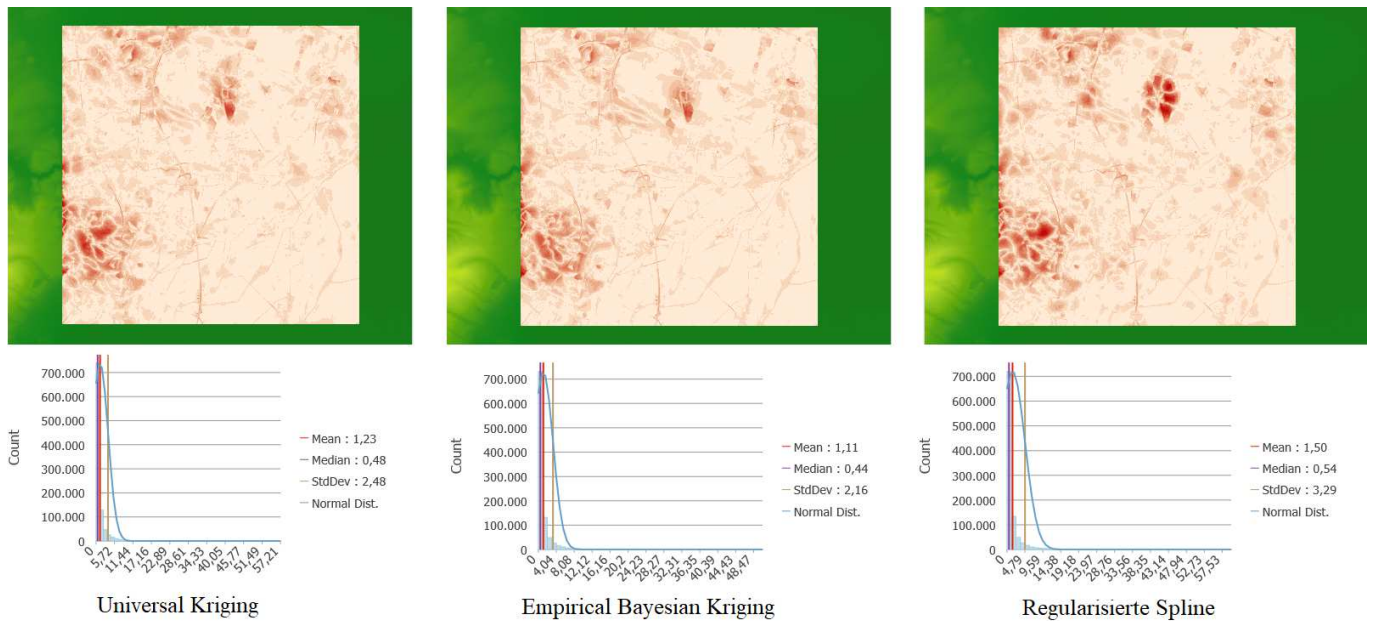


Abbildung 5.21: Gebiet B: Globale Abweichungen der Interpolationsmethoden

Bei den Visualisierungen und Histogrammen bemerkt man sofort die verhältnismäßig großen Abweichungen in Gebieten mit starkem Anstieg. Der Großteil des Untersuchungsgebiets weist nur geringe Abweichungen auf, was sich in den rechts-schiefen Histogrammen widerspiegelt. Es zeigt sich auch, dass EBK die geringste mittlere und maximale Abweichung hat. Dies ist speziell im westlichen Bereich der Visualisierung, mit weniger stark ausgeprägten Farben in der Hügelregion zu beobachten. Dies kann durch den aufwändigeren Modellierungsprozess der räumlichen Korrelation von EBK erklärt werden. Universal Kriging hat einen ähnlichen Maximalwert wie die Regularisierte Spline, schneidet aber im Mittel besser ab. Der Vergleich der RMSE der Kreuzvalidierung mit dem RMSE aller Abweichungsraster ergibt folgende Aufstellung:

Methode	Wahrer RMSE [m]	Kreuzvalidierungs RMSE [m]	Durchschnittlicher Standardfehler [m]
Universal Kriging	2,77	2,57	0,69

Method	Wahrer RMSE [m]	Kreuzvalidierungs RMSE [m]	Durchschnittlicher Standardfehler [m]
EBK	2,43	2,82	2,97
Regularisierte Spline	3,62	3,91	-

Anhand dieser Auswertung ergibt sich, dass doch EBK das Untersuchungsgebiet am besten annähert. Anders als bei Gebiet A, zeigt sich für Universal Kriging eine leichte Unterschätzung des RMSE der Kreuzvalidierung. Universal Kriging liefert die zweitbesten Ergebnisse. Die Spline Interpolation weist für beide RMSE die schlechtesten Ergebnisse auf. Diese Untersuchungen unterstreichen, dass die Qualitätsbeurteilung durch Kreuzvalidierung von der Anzahl und Konfiguration der Referenzpunkte abhängig ist.

5.4 Bewertung der Interpolationsverfahren

Anhand der Abweichungen des Interpolationsmodells vom Referenzraster (wahrer RMSE) und den Referenzpunkten (RMSE der Kreuzvalidierung) hat sich Kriging als genaueste Methode herausgestellt. Die Abweichungen zwischen wahren RMSE und RMSE der Kreuzvalidierung bewegen sich je nach Interpolationsmethode in Gebiet A im Bereich von 1,3-2,1m und in Gebiet B im Bereich von 0,2-0,4m. Diese Schwankungsbreiten könnten bei grenzwertigen Datenqualitäten und alleiniger Verifikation durch Kreuzvalidierung zu Fehlern in der Qualitätsprüfung führen. Die ICAO-Anforderungen verlangen Höhengenaugigkeiten auf bis zu 0,5m mit 90-prozentiger Wahrscheinlichkeit. Da in der Realität für gewöhnlich keine flächendeckenden und oft nicht ausreichend genaue Referenzmessungen zur Verfügung stehen, bietet ein stochastisches Modell starke Indikatoren dafür, ob ein Geländedatensatz mit einer bestimmten Konfidenz klassifiziert werden darf. Die ICAO fordert auch, dass Genauigkeitsinformationen pro Feature (auf Datenebene) angegeben werden müssen. Wird also lediglich eine Kreuzvalidierung durchgeführt, kann den einzelnen Höhenwerten nur ein über den gesamten Datensatz gültiger Genauigkeitsparameter zugewiesen werden. Dies kann beispielsweise bei regionalen Untersuchungen zu Problemen führen. Interpolationsergebnisse variieren in Bezug auf Genauigkeit stark mit der vorherrschenden Topographie. Diese Variationen werden ignoriert, wenn für alle Punkte eines Datensatzes nur der datensatzweite RMSE (mit Konfidenzbereich) angegeben wird. Möchte man beispielsweise regionale Untersuchungen im Datensatz anstellen, hat man keinerlei Indikation, ob man sich hier außerhalb des Konfidenzbereiches bewegt. Ein stochastisches Modell erlaubt hingegen regionale Aussagen mit statistisch belegbarer Signifikanz. Eine derartige Aufgabestellung ist beispielsweise bei der Identifikation von Geländehöhendaten gegeben, die eine sicherheitskritische Begrenzungsfläche überragen. Nach den ICAO-Anforderungen sind für Geländedaten, die weiter als 10km vom Flugplatz-Bezugspunkt entfernt sind und eine Begrenzungsfläche 120m über dem niedrigsten Punkt der Landebahn nicht überragen, die Genauigkeitsanforderungen für ICAO-Gebiet 1 ausreichend. Liegen die Geländedaten über dieser Begrenzungsfläche, sind

die Genauigkeitsanforderungen für ICAO-Gebiet 2 anzuwenden. Möchte man eine statistisch belastbare Aussage darüber treffen, für welche Geländedaten genauere Informationen benötigt werden, so stellt Wahrscheinlichkeits-Kriging (siehe Abschnitt 2.2.2) die notwendigen Mittel bereit. In folgender Visualisierung wurden beispielsweise alle Wahrscheinlichkeiten berechnet, bei denen ein Höhengrenzwert von 306m überschritten wird:

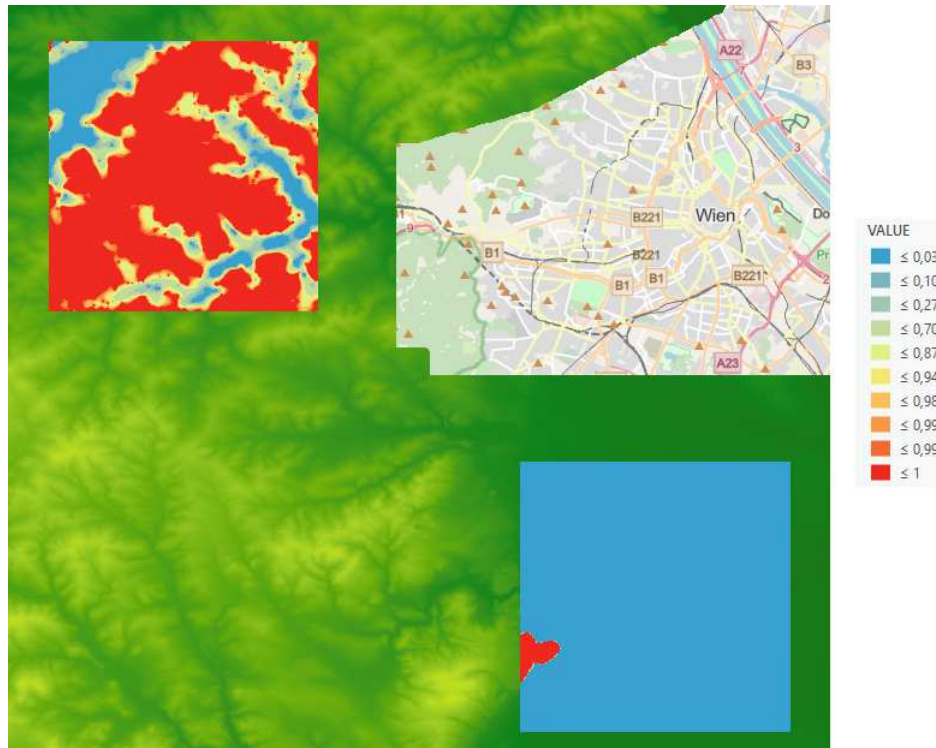


Abbildung 5.22: Wahrscheinlichkeit der Überschreitung eines Höhengrenzwertes von 306m

Für Gebiet A bedeutet dies, dass bei allen Grün bis Gelb gekennzeichneten Bereichen eine reelle Chance besteht den Grenzwert zu überschreiten. Bei den Gelb bis Rot gekennzeichneten Arealen ist eine Unterschreitung fast auszuschließen. Bei Gebiet B können annähernd alle Daten mit den Anforderungen von ICAO-Gebiet 1 klassifiziert werden. Lediglich eine Erhebung (westlich im Bild) liegt mit statistischer Signifikanz über dem Grenzwert. Dieses Beispiel unterstreicht, dass Kriging aufgrund der

- hohen Genauigkeit
- Schätzung von Genauigkeitsinformationen für Interpolationspunkte
- flexiblen Einsatzfähigkeit

in Verbindung mit der Kreuzvalidierung, die optimalen Mittel zur Interpolation von Geländehöhendaten in der Flugsicherung bereitstellt.

6 Abschließende Bewertung

6.1 Diskussion

Diese Diplomarbeit behandelt die Fragestellung: 'Welches Interpolationsverfahren ist für die Prädiktion von digitalen Geländehöhendaten nach den ICAO-Qualitätsanforderungen optimal?'

Zur Beantwortung dieser Frage wurden in Kapitel 2 die Grundlagen zur explorativen Datenanalyse und räumlichen Interpolation erarbeitet. Dabei ist spezielles Augenmerk auf die Analyse der räumlichen Struktur, Datenverteilung und Ausreißeranalyse gelegt worden. Im Kontext der räumlichen Interpolation wurde zwischen deterministischen und statistischen Interpolationsmethoden unterschieden. Für den Vergleich unterschiedlicher Interpolationsverfahren wurden in diesem Kapitel abschließend Methoden zur Qualitätsbeurteilung bearbeitet.

Aufbauend behandelt Kapitel 3 den Stand der Forschung und stellt zwei Feldstudien, eine mathematische Analyse sowie die Empfehlungen der EUROCONTROL gegenüber. Allgemein lässt sich sagen, dass es keine eindeutige Antwort dafür gibt, welche Interpolationsmethode nun generell am genauesten arbeitet. Die Wahl der Interpolationsmethode ist stark vom Einsatzgebiet abhängig. Dabei steigt die Ungenauigkeit der Interpolationsergebnisse mit zunehmender Variabilität in den Daten.

Mit einer Anforderungsanalyse in Kapitel 4 wurden konkrete Kriterien an die optimale Interpolationsmethode im Kontext der Flugsicherung gestellt. Dabei wurden die für Geländedaten relevanten Anforderungen aus ICAO (2018a) und ICAO (2018b) bezüglich Abdeckungsgebiete, Datenbereitstellung sowie Inhalt, numerische Anforderungen und Struktur von Geländedatensätzen analysiert. Unter der Prämisse, dass das Interpolationsergebnis aus sicherheitskritischen Gründen das reale Gelände möglichst genau annähern soll, wurden die Kriging- und Spline-Interpolation, aufgrund ihrer fortgeschrittenen und flexiblen Modelle, für weiterführende Analysen ausgewählt. Ein zweiter wesentlicher Punkt ist die Forderung nach Genauigkeitsinformationen und Konfidenzbereichen auf Datenebene (pro Feature bzw. Rasterzelle). Die EUROCONTROL empfiehlt Splines aufgrund deren Eigenschaft, den Wertebereich der Referenzpunkte verlassen zu können, rät aber von Interpolationen in zu großen Datenlücken ab. Dieser Limitierung könnte mittels eines stochastischen Modells und der Möglichkeit zur Schätzung von Genauigkeiten entgegengewirkt werden. Es ist auch ohne stochastischem Interpolationsmodell, nämlich durch Kreuzvalidierung, möglich Genauigkeitsinformationen konform der ICAO-Anforderungen abzuleiten. Der Nachteil die-

ser Methode ist jedoch, dass dabei (beispielsweise durch RMSE) die allgemeine Qualität des Interpolationsmodells bezüglich der Referenzpunkte berechnet wird. Richtet man sich streng nach den ICAO-Anforderungen, ist die Angabe eines Grenzwertes mit zugehöriger Konfidenz aus den Daten einer Kreuzvalidierung jedoch ausreichend, solange dieser Grenzwert den Bestimmungen des jeweiligen Anforderungsgebietes genügt. Die alleinige Qualitätsbeurteilung der Ergebnisse der Kreuzvalidierung kann aber problematisch sein. Einerseits handelt es sich dabei um eine Auswertung, die auf einer Stichprobe von Referenzmessungen aus der Grundgesamtheit des wahren Geländes basiert. Andererseits werden dadurch weiterführende, regionale und statistisch belegbare Analysen eingeschränkt.

Für die Evaluierung von geeigneten Interpolationsmethoden wurden in Kapitel 5 zwei Testgebiete mit unterschiedlicher Topographie im Ausmaß von 10x10km mit einer Zellengröße von 10m definiert. Anhand eines Referenzdatensatzes wurden Referenzpunkte in den Untersuchungsgebieten simuliert und eine explorative Datenanalyse durchgeführt. Beim ersten Gebiet handelt es sich um ein topographisch sehr variables und unregelmäßiges Gelände. Das zweite Gebiet variiert sehr viel weniger, weist jedoch eine verhältnismäßig starke Erhebung am Rand des Untersuchungsgebietes auf. Nach der explorativen Datenanalyse wurden die Ergebnisse von Universal Kriging, Empirical Bayesian Kriging und Regularisierter Spline-Interpolation miteinander verglichen. Universal Kriging wurde mit einer Trendmodellierung kombiniert. Da Empirical Bayesian Kriging viele Schritte der Datenanalyse automatisiert, wurde es als Referenz zum Universal Kriging mitbestimmt. Die Spline-Interpolation wurde nach den Empfehlungen von Mitas und Mitasova (1999) als (komplett) regularisierte Spline mit Hilfe der ArcGIS Parameteroptimierung definiert. Allgemein zeigte sich, dass Universal Kriging in beiden Untersuchungsgebieten bessere Ergebnisse als die Spline-Interpolation aufweist. Zudem ermöglicht Kriging die Ableitung von punktuellen Genauigkeitsinformationen. Die Qualität dieser Information ist jedoch von der Annahme normalverteilter Messfehler abhängig, kann aber durch QQ-Plots oder durch Vergleich des durchschnittlichen Standardfehlers mit dem RMSE der Kreuzvalidierung evaluiert werden. Da die Kreuzvalidierung im Grunde nur eine Stichprobe ist, kann es vor allem bei ungünstigen Konfigurationen von Messpunkten zu nicht-repräsentativen Ergebnissen kommen. Dabei besteht beispielsweise bei qualitativ grenzwertigen Datensätzen die Gefahr, dass diese falsch klassifiziert werden. Zwar liefern auch statistische Interpolationsverfahren bei einer ungünstigen Verteilung der Referenzpunkte keine guten Ergebnisse, dies wird aber zumindest durch entsprechend hohe Standardabweichungen in den betreffenden Gebieten angezeigt. Zudem erfolgt vor Anwendung statistischer Interpolationsmethoden meist eine explorative Datenanalyse, bei der die topographischen Eigenschaften genau geprüft werden. Weiters bieten statistische Interpolationsverfahren die Möglichkeit weiterführende, regionale Analysen im Datensatz durchzuführen. So wurde beispielsweise in den Untersuchungsgebieten evaluiert, welche Datenpunkte nach andern ICAO-Qualitätsanforderungen klassifiziert werden dürften.

Für eine abschließende Bewertung wird zusammengefasst, dass beim statistischen Interpolationsverfahren Kriging, anhand der Berechnungen in Abschnitt 5, die genauesten Ergebnisse zu erwarten sind. Die Nachvollziehbarkeit und qualitative Beurteilung der Daten hat speziell im sicherheitskritischen Anwendungsbereich hohen Stellenwert. Da eine sorgfältige

Datenanalyse maßgeblich für die Genauigkeit des Endresultats ist und berechnete Genauigkeitsinformationen auf Datenebene weitere regionale und vor allem statistisch belastbare Aussagen ermöglichen, ist Kriging für die Prädiktion von digitalen Geländehöhendaten im Kontext der ICAO-Qualitätsanforderungen optimal geeignet.

6.2 Ausblick

Digitale Geländehöhenmodelle werden im Grunde bei allen Anwendungsfällen mit Raumbegzug benötigt und entsprechendes gilt für die räumliche Interpolation. Weitere Entwicklungen werden sich durch die Optimierung von Interpolationsverfahren und den implementierten Algorithmen ergeben. Das Kriging-Framework ist bereits sehr flexibel einsetzbar und kann durch die Konfiguration unterschiedlichster Parameter gesteuert werden. Potential gibt es hier bei der Automatisierung der explorativen Datenanalyse (z.B. Empirical Bayesian Kriging) und beim Umgang mit großen Datenmengen. Dies wird insbesondere durch die zunehmende Automatisierung und Vernetzung entlang der digitalen Datenkette, von der Datenaufnahme bis hin zum Endprodukt, wichtig sein. Dadurch ergeben sich für internationale Organisationen neue Möglichkeiten Standards zu setzen und sich auch als zentraler Datenprovider und Qualitätssicherer zu positionieren. Beispielsweise könnte das eTOD-Datenmanagement zentral von der EUROCONTROL durchgeführt werden, anstatt diese Aufgabe allein den nationalen Flugsicherungen zu überlassen. Dabei könnten durch die Weiterentwicklung bestehender Interpolationsverfahren eine Vielzahl maßgeschneiderter Endprodukte für unterschiedlichste eTOD-Anwendungsfälle auf internationaler und nationaler Ebene bereitgestellt werden. Im Rahmen dieser Diplomarbeit wurde verifiziert, dass die Genauigkeit von Interpolationsergebnissen von der vorherrschenden Topographie abhängig ist. Entsprechend könnten Optimierungen in der Wahl und Konfiguration der Interpolationsverfahren bezüglich der Topographie und Geologie entwickelt werden. Dabei könnte im Sinne der Laufzeitoptimierung jener Interpolationsalgorithmus gewählt werden, der ausreichend genaue Ergebnisse liefert. Gerade aufgrund der Eigenschaft, dass die optimale Interpolationsmethode je nach Anwendungsgebiet zu wählen ist, gibt es sehr viel Potential für weiterführende Aufgabenstellungen in unterschiedlichsten Disziplinen.

Abbildungsverzeichnis

2.1	Empirisches und theoretisches Variogramm (Niemeier, 2008, S. 427)	8
2.2	Variogramm-Parameter (Niemeier, 2008, S. 428)	9
2.3	Globaler Trend (ESRI, 2020c)	10
2.4	Richtungsabhängige Einflüsse (ESRI, 2020c)	10
2.5	Normalverteilung und deren Parameter (Niemeier, 2008, S. 36)	12
2.6	Histogramm von Messwerten mit Dichtefunktion einer Normalverteilung (Niemeier, 2008, S. 15)	14
2.7	Erstellung von QQ-Diagrammen (ESRI, 2020f)	15
2.8	Globale Ausreißer im Variogramm (ESRI, 2020e)	18
2.9	Lokale Ausreißer im Variogramm (ESRI, 2020e)	18
2.10	Räumliche Verteilung von Messwerten (Niemeier, 2008, S. 410)	19
2.11	Dreiecksvermaschung als Basis eines Oberflächenmodells (Niemeier, 2008, S. 418)	20
2.12	Delaunay-Triangulation und Voronoi-Diagramme (Niemeier, 2008, S. 420)	21
3.1	Hügelland (Tan und Xu, 2014, S. 158)	42
3.2	Bergland (Tan und Xu, 2014, S. 160)	42
3.3	Gebirgsland (Tan und Xu, 2014, S. 161)	43
3.4	Fehler im Hügelland (Tan und Xu, 2014, S. 159)	43
3.5	Hügelland: Vergleich von Konturlinien (rot=Original, blau=Interpolation) und 3D-Visualisierungen (vgl. Tan und Xu, 2014, S. 159)	44
3.6	Fehler im Bergland (Tan und Xu, 2014, S. 160)	44
3.7	Bergland: Vergleich von Konturlinien (rot=Original, blau=Interpolation) und 3D-Visualisierungen (Tan und Xu, 2014, S. 160-161)	45
3.8	Fehler im Gebirgsland (Tan und Xu, 2014, S. 161)	45
3.9	Gebirgsland: Vergleich von Konturlinien (rot=Original, blau=Interpolation) und 3D-Visualisierungen (Tan und Xu, 2014, S. 161-162)	46
3.10	Vergleich von interpolierten Punkten mit DGPS-Beobachtung (Arun, 2013, S. 135)	47
3.11	Vergleich von RMSE-Werten bei verschiedenen Geländetypen (Arun, 2013, S. 135)	48
3.12	Graphische Visualisierung der Interpolationsergebnisse (Arun, 2013, S. 136)	48
3.13	Vergleich von Interpolationsmethoden anhand unregelmäßig verstreuter Daten (Mitas und Mitasova, 1999, S. 482-483)	50
3.14	Empfohlene Vermessungsmethoden für Geländedaten (EUROCONTROL, 2019, S. 157-158)	53

4.1	Gebiete 1 und 2 (ICAO, 2018b, Appendix 8)	58
4.2	Gebiete 3 und 4 (ICAO, 2018b, Appendix 8)	62
5.1	Digitales Geländemodell: Niederösterreich, 10x10m	72
5.2	DHM Wien-Schwechat: Radius 45km	73
5.3	Untersuchungsgebiete: A (links) und B (rechts)	74
5.4	Gebiet A: Datenverteilung	75
5.5	Gebiet B: Datenverteilung	76
5.6	Gebiet A: 3D-Ansicht	76
5.7	Gebiet A: Profil Ost-West	77
5.8	Gebiet A: Profil Süd-Nord	77
5.9	Gebiet A: Trend	77
5.10	Gebiet B: 3D-Ansicht	78
5.11	Gebiet A: Profil Ost-West	78
5.12	Gebiet A: Profil Süd-Nord	78
5.13	Gebiet B: Trend	79
5.14	Gebiet A: Semivariogramm	80
5.15	Gebiet B: Semivariogramm	81
5.16	Gebiet A: Interpolationsergebnisse	82
5.17	Gebiet B: Interpolationsergebnisse	83
5.18	QQ-Plots für den standardisierten Fehler für UK: Gebiet A (links) und B (rechts)	85
5.19	Geschätzte Standardabweichungen für UK: Gebiet A (links) und B (rechts)	85
5.20	Gebiet A: Globale Abweichungen der Interpolationsmethoden	87
5.21	Gebiet B: Globale Abweichungen der Interpolationsmethoden	88
5.22	Wahrscheinlichkeit der Überschreitung eines Höhengrenzwertes von 306m	90

Literaturverzeichnis

- P.V. Arun. A comparative analysis of different DEM interpolation methods. *The Egyptian Journal of Remote Sensing and Space Science*, 16, 12 2013. URL https://www.researchgate.net/publication/259143611_A_comparative_analysis_of_different_DEM_interpolation_methods/fulltext/02a5a2580cf276f46e58360b/A-comparative-analysis-of-different-DEM-interpolation-methods.pdf. Online; Stand 04. Februar 2020.
- K. Bretterbauer und H. Schuh. Mathematische Methoden der Geowissenschaften. Skriptum, 2013.
- M. Buhmann und J. Jaeger. On radial basis functions. *Snapshots of modern mathematics from Oberwolfach*, (50):16, March 2019. URL <https://d-nb.info/118173214X/34>. Online; Stand 04. Februar 2020.
- M.J. De Smith, M.F. Goodchild, und P. Longley. *Geospatial Analysis: A Comprehensive Guide*. Winchelsea Press, 2018. ISBN 9781912556038. URL <https://books.google.at/books?id=-guAuAEACAAJ>. Online; Stand 04. Februar 2020.
- P. Dehdari. *Measuring the Impact of Lean Techniques on Performance Indicators in Logistics Operations*. Wissenschaftliche Berichte des Institutes fuer Foerdertechnik und Logistiksysteme des Karlsruher Instituts fuer Technologie. KIT Scientific Publishing, 2014. ISBN 9783731500964. URL <https://books.google.at/books?id=gFdBAwAAQBAJ>. Online; Stand 20. Februar 2020.
- C. Ebert. *Systematisches Requirements Engineering: Anforderungen ermitteln, dokumentieren, analysieren und verwalten*. dpunkt.verlag, 2019. ISBN 9783960884538. URL <https://books.google.at/books?id=7-uEDwAAQBAJ>. Online; Stand 04. Februar 2020.
- ESRI. *Using ArcGIS Geostatistical Analyst*. ESRI, 2.1 edition, 2001. URL http://downloads2.esri.com/support/documentation/ao_/Using_ArcGIS_Geostatistical_Analyst.pdf. Online; Stand 04. Februar 2020.
- ESRI. Topo zu Raster, 2020a. URL <https://pro.arcgis.com/de/pro-app/tool-reference/3d-analyst/how-topo-to-raster-works.htm>. Online; Stand 11. Jänner 2020.
- ESRI. Performing cross-validation and validation, 2020b. URL <https://desktop.arcgis.com/de/arcmap/10.3/guide-books/extensions/geostatistical-analyst/performing-cross-validation-and-validation.htm>. Online; Stand 05. Jänner 2020.

- ESRI. Accounting for directional influences, 2020c. URL <https://desktop.arcgis.com/de/arcmap/latest/extensions/geostatistical-analyst/accounting-for-directional-influences.htm>. Online; Stand 05. Jänner 2020.
- ESRI. Understanding Measurement Error, 2020d. URL <https://pro.arcgis.com/de/pro-app/help/analysis/geostatistical-analyst/understanding-measurement-error.htm>. Online; Stand 03. Februar 2020.
- ESRI. Looking for global and local outliers, 2020e. URL <https://desktop.arcgis.com/de/arcmap/latest/extensions/geostatistical-analyst/looking-for-global-and-local-outliers.htm>. Online; Stand 04. Jänner 2020.
- ESRI. Normal QQ plot and general QQ plot, 2020f. URL <https://desktop.arcgis.com/de/arcmap/latest/extensions/geostatistical-analyst/normal-qq-plot-and-general-qq-plot.htm>. Online; Stand 04. Jänner 2020.
- ESRI. Box-Cox, arcsine, and log transformations, 2020g. URL <https://desktop.arcgis.com/de/arcmap/latest/extensions/geostatistical-analyst/box-cox-arcsine-and-log-transformations.htm>. Online; Stand 04. Jänner 2020.
- EUROCONTROL. *Terrain and Obstacle Data Manual*. EUROCONTROL, 2.2 edition, 2019. URL <https://www.eurocontrol.int/sites/default/files/2019-12/eurocontrol-terrain-obstacle-data-manual-v2-2.pdf>.
- B. Fornberg und N. Flyer. *A Primer on Radial Basis Functions with Applications to the Geosciences*. CBMS-NSF Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM, 3600 Market Street, Floor 6, Philadelphia, PA 19104), 2015. ISBN 9781611974041. URL <https://books.google.at/books?id=-aqmCgAAQBAJ>.
- W. Göpfert. Interpolationsergebnisse mit der multiquadratischen Methode, 1977.
- Jeffrey C. Guarneri und Robert C. Weih. Comparing methods for interpolation to improve raster digital elevation models. 2012. URL https://pdfs.semanticscholar.org/b4f4/32006def7949297093b3df013cce163e9ba2.pdf?_ga=2.124540362.1110635131.1582642893-892593240.1582642893. Online; Stand 25. Februar 2020.
- R. Hardy. Geodetic applications of multiquadratic analysis, 1972.
- M.F. Hutchinson. A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. 1989. URL <https://www.sciencedirect.com/science/article/pii/0022169489900735?via%3Dihub>. Online; Stand 01. Februar 2020.
- ICAO. *ICAO Annex 15, Amendment 40 (Sixteenth Edition)*, 2018a.
- ICAO. *PANS-AIM, Doc 10066*, 2018b.

- K. Kraus. Interpolation nach kleinsten Quadraten versus Krige-Schätzer. *VGI – Österreichische Zeitschrift für Vermessung und Geoinformation*, 86(1):45–48, 1998. URL <https://www.ovg.at/de/vgi/files/pdf/4734>. Online; Stand 04. Februar 2020.
- J. C. Mairhuber. On Haar’s theorem concerning Chebychev approximation problems having unique solutions. *American Mathematical Society* 7, (4):609–615, 1956. URL <https://www.math.umanitoba.ca/hat/fpapers/mairhuber.pdf>. Online; Stand 04. Februar 2020.
- L. Mitas und H. Mitasova. *Geographical information systems: Principles and technical issues, Chapter 34*. Number Bd. 1. John Wiley, 1999. ISBN 9780471321828. URL <https://books.google.at/books?id=gfNOAAAAMAAJ>. Online; Stand 04. Februar 2020.
- L. Mitas, H. Mitasova, J. Hofierka, und J. Parajka. Multivariate interpolation of precipitation using regularized spline with tension. *Transactions in GIS*, (6):135–150, 2002/3. URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.71.4351&rep=rep1&type=pdf>. Online; Stand 04. Februar 2020.
- G. Navratil. Ausgleichsrechnung I. Skriptum, November 2006a.
- G. Navratil. Ausgleichsrechnung II. Skriptum, Dezember 2006b.
- W. Niemeier. *Ausgleichsrechnung: Statistische Auswertemethoden*. De Gruyter Lehrbuch. De Gruyter, 2008. ISBN 9783110190557.
- G.M. Philip und D.F. Watson. A Precise Method for Determining Contoured Surfaces. *Australian Petroleum Exploration Association Journal*, (22):205–212, 1982. URL <https://www.publish.csiro.au/aj/AJ81016>. Online; Stand 04. Februar 2020.
- H.R. Schwarz und N. Köckler. *Numerische Mathematik*. Vieweg Studium. Vieweg+Teubner Verlag, 2009. ISBN 9783834892829. URL <https://books.google.at/books?id=REcwk9DUpowC>. Online; Stand 04. Februar 2020.
- Q. Tan und X. Xu. Comparative Analysis of Spatial Interpolation Methods: an Experimental Study. 165:155–163, 02 2014. URL https://www.researchgate.net/publication/279902313_Comparative_Analysis_of_Spatial_Interpolation_Methods_an_Experimental_Study/fulltext/59e768cea6fdcc6433a946b7/Comparative-Analysis-of-Spatial-Interpolation-Methods-an-Experimental-Study.pdf. Online; Stand 04. Februar 2020.
- R. Webster und M.A. Oliver. *Geostatistics for Environmental Scientists*. Statistics in Practice. Wiley, 2007. ISBN 9780470517260. URL <https://books.google.at/books?id=WBwSyvIvNY8C>. Online; Stand 04. Februar 2020.
- D. Zimmerman, C. Pavlik, A. Ruggles, und et al. Comparison of Ordinary and Universal Kriging and Inverse Distance Weighting. *Mathematical Geology*, (31):375–390, 1991. URL <https://doi.org/10.1023/A:1007586507433>. Online; Stand 04. Februar 2020.