



Detektion von anormalen Zeitreihen an Persistent-Scatterer - Punkten im Zusammenhang mit der Ableitung flächenhafter Bodenbewegungen

DISSERTATION

**zur Erlangung des Doktorgrades
der Ingenieurwissenschaften**

**vorgelegt von
M. Sc. Furui Xi
aus China**

**genehmigt von der Fakultät für Energie- und
Wirtschaftswissenschaften der Technischen
Universität Clausthal**

Clausthal-Zellerfeld 2017

Furui Xi

Detektion von anormalen Zeitreihen an Persistent-Scatterer - Punkten im Zusammenhang mit der Ableitung flächenhafter Bodenbewegungen

Dissertation, TU Clausthal, 2017

Mitglieder der Promotionskommission:

Univ.-Prof. Dr.-Ing. Norbert Meyer (Vorsitzender)

Univ.-Prof. Dr.-Ing. Wolfgang Busch (Hauptberichterstatter)

Univ.-Prof. Dr. Sven Hartmann (Mitberichterstatter)

Eidesstattliche Erklärung

Hiermit erkläre ich an Eides Statt, dass ich die bei der Fakultät für Energie- und Wirtschaftswissenschaften der Technischen Universität Clausthal eingereichte Dissertation selbständig und ohne unerlaubte Hilfe angefertigt habe. Die benutzten Hilfsmittel sind vollständig angegeben.

Unterschrift

Kurzfassung

Zur Erfassung von Bodenbewegungen hat sich die radarinterferometrische Methode der Persistent Scatterer Interferometrie (PSI) aufgrund der erreichbaren hohen Qualität und der zunehmenden Verfügbarkeit von Radar-Satellitendaten in vielen Bereichen der Wissenschaft, Wirtschaft und Behörden bewährt. Im Vergleich zu herkömmlichen Vermessungsverfahren erzeugt die PSI-Methode eine hohe Punktdichte mit Informationen über die zeitlich variierenden Bodenbewegungen bzw. Höhenänderungen an der Tagesoberfläche und mit einer Genauigkeit im Sub-Zentimeterbereich. Aufgrund verschiedener, auch verfahrenstechnischer Ursachen können die ermittelten PSI-Höhenänderungszeitreihen an einzelnen PS-Punkten jedoch von dem erwarteten flächenhaften Bodenbewegungsverhalten abweichen. Sie sind somit für dieses Bodenbewegungsverhalten nicht repräsentativ oder im messtechnischen Sinne grob falsch. Vor einer weiteren Nutzung dieser Ergebnisse sollten diese auffallenden „Anomalien“ deshalb in den vorliegenden Massendaten automatisiert erkannt und von einer Interpretation des Bodenbewegungsverhaltens ausgeschlossen werden.

Im Rahmen der vorliegenden Arbeit wird ein raumzeitliches Clusterverfahren vorgestellt, das die automatisierte Detektion von anormalen Zeitreihen an PS-Punkten in flächenhaft vorliegenden PSI-Ergebnissen ermöglicht. Zur Auswahl eines für diese Aufgabenstellung geeigneten Verfahrens wurden verschiedene Clusterverfahren auf ihre Anwendbarkeit hin untersucht und letztlich das Verfahren „Local-Moran's-Index“ auf Grundlage lokaler räumlicher Autokorrelationen ausgewählter Attribute zwischen den PS-Punkten an praktischen Beispielen getestet. Es zeigte sich, dass grundsätzlich mit Hilfe des Local-Moran's-Index - Clusterverfahrens signifikante „räumliche Anomalien“ detektiert werden können, dabei aber der Wahl des Attributes eine wichtige Bedeutung zukommt. Im Vergleich zur Clusteranalyse mit den Attributen „Polynomgrad“ und „Höhenänderung“ (abgeleitet aus einer Trendpolynombestimmung) konnte die Clusteranalyse mit dem Attribut „Höhenänderungsrate“ ein deutlich besseres Ergebnis erzielen. Darüber hinaus ergab die Untersuchung, dass zur Beschreibung des charakteristischen Verlaufs einer Höhenänderungszeitreihe ein einziges Attribut offensichtlich nicht ausreicht. Da bei der Local-Moran's-Index - Clusteranalyse jedoch nur ein einziges Attribut berücksichtigt werden kann, erfolgt eine Modifikation dieses Verfahrens durch Einführung eines Gewichtungsfaktors zur Beurteilung und Quantifizierung der Ähnlichkeit der Höhenänderungsverläufe benachbarter PS-Punkte. Des Weiteren wurde ein Kriterium abgeleitet, das auf einer gemeinsamen Betrachtung von räumlichen Autokorrelationen und der Anzahl benachbarter PS-Punkte basiert, und zu einer Modifikation der Local-Moran's-Index - Clusteranalyse führte. Erweitert wurde das Vorgehen um eine kleinräumige Clusteranalyse mit spezifischen, angepassten Parametern.

Abschließend wurde am Beispiel eines umfangreichen Testdatensatzes das erweiterte und modifizierte Local-Moran's-Index - Verfahren zur Identifizierung von anormalen Zeitreihen an PS-Punkten erprobt und das Analyseergebnis mit einem entwickelten statistischen Validationsverfahren, basierend auf berechneten Höhenänderungsdifferenzen, überprüft. Zusammenfassend zeigte sich, dass eine automatisierte Detektion von anormalen Zeitreihen an massenhaft vorhandenen PS-Punkten mittels des modifizierten Local-Moran's-Index - Verfahrens zuverlässig erfolgen kann. Die hierdurch ermöglichte Datenbereinigung liefert einen wichtigen Beitrag zur Verbesserung der Qualität einer flächenhaften Modellierung von Bodenbewegungen.

Abstract

Due to the achievable high quality and the increasing availability of radar satellite data, the radar interferometric method of persistent scatterer Interferometry (PSI) has proved of great value for the detection of ground movements in many fields of science, economics and public authorities. Compared to conventional surveying methods, the PSI method produces a high point density with information about the temporally varying ground movements or height changes on the ground surface with an accuracy in the sub-centimeter range. However, due to various, also procedural reasons, the determined PSI time series of height changes at individual PS-points can deviate from the expected areal ground movement behavior. They are therefore not representative to describe this ground movement behavior or are grossly wrong in the metrological sense. Before any further use of these results, these striking "anomalies" should therefore be recognized automatically in the present mass data and excluded from an interpretation of the ground movement behavior.

In the present work, a space-time clustering method is presented, which enables the automated detection of abnormal time series at PS points in areal PSI results. In order to select a suitable method for this task, various clustering methods have been investigated for their applicability and finally the "Local-Moran's-Index" method on the basis of local spatial autocorrelation of selected attributes between the PS points has been tested in practical examples. It was found that in principle significant "spatial anomalies" can be detected with the help of the Local-Moran's-Index clustering method, but the choice of attribute is of great importance. In comparison to the cluster analysis with the attributes "polynomial degree" and "height change" (derived from a trend polynomial determination), the cluster analysis with the attribute "height change rate" was able to achieve a significantly better result. Furthermore, the investigation showed that a single attribute is obviously not sufficient to describe the characteristic course of a height change time series. However, since only a single attribute can be taken into account in the Local-Moran's-Index cluster analysis, a modification of this method is carried out by introducing a weighting factor for judging and quantifying the similarity of the height changes of adjacent PS points. Moreover, a criterion was derived based on a joint consideration of spatial autocorrelation and the number of neighboring PS points and led to a modification of the Local-Moran's-Index cluster analysis. The approach was then extended by a small-scale cluster analysis with specific, adapted parameters.

Finally, using the example of a comprehensive test data set, the extended and modified Local-Moran's Index method for the identification of abnormal time series at PS points was tested, and the analysis result was checked with a developed statistical validation method based on calculated differences of height changes. In summary, it was found that automated detection of abnormal time series at mass PS points can be reliably performed by means of the modified Local-Moran's-Index method. The data cleansing provided thereby makes an important contribution to the improvement of the quality of areal modeling of ground movements.

Inhaltsverzeichnis

Eidesstattliche Erklärung	I
Kurzfassung	II
Abstract	III
Inhaltsverzeichnis	IV
Abbildungsverzeichnis	VII
Tabellenverzeichnis	X
1. Einleitung	1
2. Persistent Scatterer Interferometrie	3
2.1. Allgemeine Beschreibung der Radarinterferometrie	3
2.2. Grundlagen der PSI-Technologie.....	8
2.3. Datenaufbereitung und Postprocessing der PSI-Ergebnisse	10
2.4. Anwendungsbeispiele	13
3. Anomalien in den PSI-Ergebnissen	18
3.1. Definition von Anomalien.....	18
3.2. Grund für die Detektion von Anomalien	20
3.3. Klassifikation von Anomalien	22
3.3.1. Räumliche Anomalien	22
3.3.2. Temporale Anomalien	22
3.3.3. Spatio-Temporale Anomalien	23
3.4. Anomalien im raumzeitlichen Verhalten benachbarter PS-Punkte und ihre Detektion	28
3.4.1. Anomalien durch Phasensprünge	29
3.4.2. Anomalien infolge verrauschter Signale.....	31
3.4.3. Anomalien durch kurzfristige anthropogene Veränderungen oder Messfehler.	33
4. Clusteranalyse	36
4.1. Prinzip der Clusteranalyse	36
4.2. Ablauf und Arbeitsschritte einer Clusteranalyse	38

4.3.	Festlegung von Ähnlichkeiten bzw. des Distanzmaßes	40
4.3.1.	Minkowski-Metriken	42
4.3.2.	Korrelationskoeffizient zur Bestimmung der Ähnlichkeit.....	46
4.4.	Auswahl des Algorithmus zur Gruppierung	49
4.4.1.	Hierarchische Verfahren	50
4.4.2.	Partitionierende Verfahren	54
4.4.3.	Graphentheoretische Verfahren	57
4.4.4.	Alternative Optimierungsverfahren	57
4.5.	Bestimmung einer optimalen Clusteranzahl.....	57
5.	Moran's-Index - Clusterverfahren	59
5.1.	Clusterverfahren mit Raumbezug	59
5.2.	Clusterverfahren mit Anomalie-Detektion	63
5.3.	Prinzip des Local-Moran's-Index.....	66
5.3.1.	Räumliche Autokorrelation	66
5.3.2.	Globale räumliche Autokorrelation.....	68
5.3.3.	Lokale räumliche Autokorrelation.....	70
5.3.4.	Signifikanzprüfung	74
5.3.5.	Bestimmung/Bildung von Nachbarschaft.....	78
5.3.6.	Gewichtung der räumlichen Nachbarn	81
5.4.	Beispiel zur Berechnung des Local-Moran's-Index.....	84
6.	Modifizierung des Local-Moran's-Index - Clusterverfahrens zur praktischen Anwendung.....	87
6.1.	Allgemeiner Verfahrensablauf des Local-Moran's-Index - Clusterverfahrens zur Identifizierung von Anomalien.....	87
6.2.	Beschreibung der Testdaten.....	88
6.2.1.	TerraSAR-X Testdaten	89
6.2.2.	Radarsat-2 Testdaten	90
6.2.3.	Envisat Testdaten	91
6.3.	Optimierte Festlegung der Nachbarschaft	92
6.4.	Local-Moran's-Index - Clusteranalyse mit Zeitreihen-Attributen	104
6.4.1.	Clusteranalyse mit Attribut „Höhenänderungsrate“	105
6.4.2.	Clusteranalyse mit Attribut „Polynomgrad“	106
6.4.3.	Clusteranalyse mit Attribut „Höhenänderung“	107

6.5.	Modifizierte Local-Moran's-Index - Clusteranalyse mit zeitlicher Ähnlichkeitsgewichtung	109
6.6.	Kleinräumige Clusteranalyse	111
6.7.	Verfahrensablauf der modifizierten Local-Moran's-Index - Clusteranalyse	114
7.	Validierung des modifizierten Local-Moran's-Index - Verfahrens	117
7.1.	Musteranalyse	117
7.2.	Bestimmung eines optimalen Suchradius.....	118
7.3.	Detektion von anormalen Zeitreihen an PS-Punkten	122
7.4.	Validation des Analyseergebnisses	129
8.	Zusammenfassung	136
9.	Anhang	139
10.	Literaturverzeichnis	149

Abbildungsverzeichnis

Abbildung 1: Vereinfachte Geometrie einer Radarmessung [VERÄNDERT NACH BAMLER 2008]	3
Abbildung 2: Vereinfachte Aufnahme-Geometrie der DinSAR [VERÄNDERT NACH BAMLER 2008]	4
Abbildung 3: Vereinfachter Prozessierungsablauf zur Bestimmung deformationsbedingter Phasenanteile (Bodenbewegungen) [VERÄNDERT NACH RABUS 2002]	5
Abbildung 4: Zusammensetzung der Phasenanteile	6
Abbildung 5: Arten von Dekorrelationen des Radarsignals	7
Abbildung 6: Wesentliche Bearbeitungsschritte der PSI-Auswertung [VERÄNDERT NACH ATHAYDE PINTO 2015].....	9
Abbildung 7: Höhenänderungszeitreihe eines PS-Punktes	11
Abbildung 8: Ausgleichendes Trendpolynom 2. Grades der Höhenänderungszeitreihe eines PS- Punktes	11
Abbildung 9: Polynomgrade von PS-Punkt-Zeitreihen im Großraum Leipzig [BUSCH 2014].....	12
Abbildung 10: Bewegungsrate in LOS (mm/a) der ENVISAT und Radarsat-2 PSI-Zeitreihen im Bereich des Bergwerk Walsum [BUSCH 2016]	14
Abbildung 11: Zeitreihen nivellitisch und radarinterferometrisch bestimmter Höhenänderungen an einem ausgewählten Punkt des Leitnivelements im Bereich des Bergwerks Walsum [BUSCH 2016] ..	15
Abbildung 12: Bewegungsrate in LOS von PS-Punkten in Düsseldorf (mm/Jahr) [VERÄNDERT NACH MARK 2012].....	16
Abbildung 13: Höhenänderungsraten von PS-Punkten über einem Aquifer-Gas-Speicher im westlichen Teil Berlins [KNOSPE, ET AL. 2011]	17
Abbildung 14: Typische Anomalien im Histogramm	18
Abbildung 15: Vereinfachte Darstellung einer Anomalie (rotes Quadrat) im Vergleich zu den Messwerten in seiner Nachbarschaft.....	19
Abbildung 16: Vereinfachte Darstellung einer temporären Anomalie und ihre Auswirkung	23
Abbildung 17: Georeferenzierte Variablen [VERÄNDERT NACH HADI 2012].....	25
Abbildung 18: Georeferenzierte Zeitreihen [VERÄNDERT NACH HADI 2012]	25
Abbildung 19: Bewegte Punkte [VERÄNDERT NACH HADI 2012]	26
Abbildung 20: Trajektorien [VERÄNDERT NACH HADI 2012]	27
Abbildung 21: Typen spatio-temporaler Daten [VERÄNDERT NACH MAIMON 2010].....	27
Abbildung 22: Phasensprung in einer PSI-Zeitreihe [VERÄNDERT NACH YIN 2015].....	29
Abbildung 23: Detektion eines Phasensprunges anhand eines vordefinierten Schwellenwertes [VERÄNDERT NACH YIN 2015]	30
Abbildung 24: Flussdiagramm zur Detektion von sprunghaften Höhenänderungen [YIN 2015]	31
Abbildung 25: Stark verrauschte PSI-Zeitreihe.....	32
Abbildung 26: Abweichendes Höhenänderungsverhalten einer Anomalie gegenüber umgebenden PS- Punkten [VERÄNDERT NACH XI 2015]	34
Abbildung 27: Lage der Untersuchungsobjekte in einem zweidimensionalen Merkmalsraum [VERÄNDERT NACH FLETLING 2010]	37
Abbildung 28: Verfahrensschritte einer Clusteranalyse [VERÄNDERT NACH BACKHAUS 2016]	40
Abbildung 29: Vergleich der Höhenänderungsverläufe von „PS-Punkt1“ und „PS-Punkt5“	48

Abbildung 30: Vereinfachte Profilverläufe der Merkmale von drei Untersuchungsobjekten [VERÄNDERT NACH FLETLING 2010].....	48
Abbildung 31: Einteilung von ausgewählten Cluster-Algorithmen [VERÄNDERT NACH BACKHAUS 2016]...	50
Abbildung 32: Dendrogramm der hierarchischen Verfahren [VERÄNDERT NACH BERGS 1981]	51
Abbildung 33: Schematische Darstellung der Fusionierungsverfahren [CLEFF 2015]	53
Abbildung 34: Ablaufschritte beim iterierten Minimaldistanzverfahren [LANGE 2013]	56
Abbildung 35: „Ellbogen-Kriterium“ zur Ermittlung der optimalen Clusteranzahl bei den partitionierenden Verfahren [VERÄNDERT NACH FLETLING 2010].....	58
Abbildung 36: Übersicht und Einteilung von ausgewählten räumlichen Cluster-Algorithmen [VERÄNDERT NACH KOLATCH 2001]	60
Abbildung 37: Drei Arten von Objekten in DBSCAN.....	64
Abbildung 38: Kettenverbindung von Kernobjekten und Dichte-erreichbare-Objekten.....	65
Abbildung 39: Typen räumlicher Autokorrelation [VERÄNDERT NACH LO 2007]	68
Abbildung 40: Signifikanzniveau gemäß Z-Ergebnis und p-Wert im Nullhypotesentest [VERÄNDERT NACH ESRI ONLINE].....	76
Abbildung 41: Queen’s Case (links) und Rook’s Case (rechts)	79
Abbildung 42: Nachbarschaft im TIN-Modell mit Raumschrittweiten 1 und 2.....	80
Abbildung 43: Nachbarschaft im Rastermodell mit Raumschrittweiten 1 (R1) und 2 (R2) (links: Queen’s Case Rechts: Rook’s Case).....	81
Abbildung 44: Nachbarschaft mit fester Distanz.....	81
Abbildung 45: Eigenschaften bekannter Gewichtungsmethoden [VERÄNDERT NACH ESRI ONLINE].....	83
Abbildung 46: Räumliche Lage und Variablenwerte der fünf Untersuchungsobjekte.....	84
Abbildung 47: Allgemeiner Verfahrensablauf des Local-Moran’s-Index - Clusterverfahrens zur Identifizierung von Anomalien [VERÄNDERT NACH XI 2015].....	88
Abbildung 48: Räumliche Lage der TerraSAR-X PS-Punkte	90
Abbildung 49: Räumliche Lage und Verteilung der Radarsat-2 PS-Punkte	91
Abbildung 50: Räumliche Lage und Verteilung der Envisat PS-Punkte	92
Abbildung 51: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung	94
Abbildung 52: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung (20 m - 50 m) für die Envisat-PSI-Daten	95
Abbildung 53: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung (50 m – 110 m) für die Envisat-PSI-Daten	96
Abbildung 54: Fünf ausgewählte Teilgebiete im Testgebiet der Radarsat-2-Daten	97
Abbildung 55: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 1 (Abbildung 54) mit Radarsat-2-Testdaten	98
Abbildung 56: Fünf ausgewählte Teilgebiete im Testgebiet mit den Envisat-Daten	99
Abbildung 57: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 1 (Abbildung 56) mit Envisat-Testdaten	100
Abbildung 58: Zehn ausgewählte Teilgebiete im Testgebiet mit den TerraSAR-X-Daten.....	101
Abbildung 59: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 1 (Abbildung 58) mit TerraSAR-X-Testdaten	102
Abbildung 60: Ausgewählte Eigenschaftswerte zur Charakterisierung einer Höhenänderungszeitreihe eines PS-Punktes	104
Abbildung 61: Detektierte räumliche Anomalien mit Attribut „Höhenänderungsrate“ (dargestellt durch rote Punkte) und vergleichende Darstellung ausgewählter Zeitreihen (in grün dargestellt der Zeitreihenverlauf eines als Anomalie detektierten PS-Punktes).....	106

Abbildung 62: Detektierte räumliche Anomalien mit Attribut „Polynomgrad“ (dargestellt durch rote Punkte) und vergleichende Darstellung ausgewählter Zeitreihen (in grün dargestellt der Zeitreihenverlauf eines als Anomalie detektierten PS-Punktes).....	107
Abbildung 63: Detektierte räumliche Anomalien mit Attribut „Höhenänderung“ (dargestellt durch rote Punkte) und vergleichende Darstellung ausgewählter Zeitreihen (in grün dargestellt der Zeitreihenverlauf eines als Anomalie detektierten PS-Punktes).....	108
Abbildung 64: Mit Attribut „Höhenänderungsrate“ potentiell falsch detektierte räumliche Anomalie (grüner Zeitreihenverlauf) am Beispiel von TerraSAR-X Testdaten	109
Abbildung 65: Höhenänderungs-Zeitreihe eines ausgewählten Clusterpunktes im Vergleich zur mittleren Zeitreihe benachbarter PS-Punkte	112
Abbildung 66: Häufigkeitsverteilung der Höhenänderungsraten der Nachbarpunkte des untersuchten Clusterpunktes	114
Abbildung 67: Flussdiagramm des Verfahrensablaufs bei Anwendung des modifizierten Local-Moran's-Index - Clusterverfahren zur Detektion von Anomalien in PS-Punkt-Daten.....	116
Abbildung 68: Musteranalyse von Radarsat-2-Testdaten.....	118
Abbildung 69: Anzahl der PS-Punkte mit min. 8 Nachbarn in Abhängigkeit einer intervallweisen Vergrößerung des Suchradius	120
Abbildung 70: PS-Punkte (orange Punkte), die nicht über min. acht benachbarte PS-Punkte innerhalb des Suchradius verfügen	121
Abbildung 71: Räumliche Lage der detektierten PS-Punkte mit anormalen Zeitreihen (rot dargestellt)	123
Abbildung 72: Gegenüberstellung der detektierten anormalen Zeitreihe A mit der gemittelten Zeitreihe benachbarter PS-Punkte	124
Abbildung 73: Gegenüberstellung der detektierten anormalen Zeitreihe B mit der gemittelten Zeitreihe benachbarter PS-Punkte	124
Abbildung 74: Gegenüberstellung der detektierten anormalen Zeitreihe C mit der gemittelten Zeitreihe benachbarter PS-Punkte	125
Abbildung 75: Gegenüberstellung der detektierten anormalen Zeitreihe D mit der gemittelten Zeitreihe benachbarter PS-Punkte	125
Abbildung 76: Räumliche Lage der detektierten Clusterpunkte mit abweichenden Höhenänderungsraten (rot dargestellt).....	127
Abbildung 77: Abweichende Zeitreihe des Clusterpunktes E und gemittelte Zeitreihe benachbarter PS-Punkte	128
Abbildung 78: Abweichende Zeitreihe des Clusterpunktes F und gemittelte Zeitreihe benachbarter PS-Punkte	128
Abbildung 79: Gegenüberstellung der detektierten anormalen Zeitreihe G mit der gemittelten Zeitreihe benachbarter PS-Punkte	132
Abbildung 80: Gegenüberstellung der detektierten anormalen Zeitreihe H mit der gemittelten Zeitreihe benachbarter PS-Punkte	133
Abbildung 81: Gegenüberstellung der detektierten anormalen Zeitreihe I mit der gemittelten Zeitreihe benachbarter PS-Punkte	133
Abbildung 82: Gegenüberstellung der detektierten anormalen Zeitreihe J mit der gemittelten Zeitreihe benachbarter PS-Punkte	134
Abbildung 83: Häufigkeitsverteilung der Höhenänderungsdifferenzen der Nachbarpunkte des detektierten anormalen PS-Punktes G.....	134

Tabellenverzeichnis

Tabelle 1: Spatio-temporale Ereignisse	24
Tabelle 2: Vereinfachte Darstellung des Aufbaus einer Rohdatenmatrix	41
Tabelle 3: Vereinfachte Darstellung des Aufbaus einer Ähnlichkeits- oder Distanzmatrix	42
Tabelle 4: Ausgangsdatenmatrix von PS-Punkten mit metrischer Merkmalsstruktur	42
Tabelle 5: Distanzmatrix von PS-Punkten anhand der City-Block Metrik	44
Tabelle 6: Distanzmatrix für PS-Punkten mit Euklidischen Distanzen.....	45
Tabelle 7: Reihenfolge der Ähnlichkeiten zwischen den PS-Punkten anhand der City-Block Distanz (jeweils erste Zahl) und der Euklidischen Distanz (Werte in Klammern)	46
Tabelle 8: Ähnlichkeitsmatrix von PS-Punkten mit den Korrelationskoeffizienten	47
Tabelle 9: Wesentliche Charakteristika der Fusionierungsverfahren [VERÄNDERT NACH ECKSTEIN 2014].	53
Tabelle 10: Bewertung einiger ausgewählter räumlicher Clusterverfahren [VERÄNDERT NACH KOLATCH 2001].....	62
Tabelle 11: Signifikanzniveaus und entsprechende kritische Z-Ergebnisse und p-Werte im Nullhypothesentest [VERÄNDERT NACH ESRI ONLINE]	77
Tabelle 12: Entfernungen (Suchradius), in denen der erste Peak in den fünf Teilgebieten der Radarsat- 2-Testdaten auftritt	98
Tabelle 13: Entfernungen (Suchradius), in denen der erste Peak in den fünf Teilgebieten der Envisat- Testdaten auftritt	100
Tabelle 14: Entfernungen (Suchradius), in denen der erste Peak in den zehn Teilgebieten der TerraSAR-X-Testdaten auftritt.....	102
Tabelle 15: Anzahl der PS-Punkte mit min. 8 Nachbarn in Abhängigkeit einer intervallweisen Vergrößerung des Suchradius	120
Tabelle 16: Charakteristische Merkmale der in Abbildung 72 – 75 dargestellten anomalen Zeitreihen	126
Tabelle 17: Zusammenhang zwischen Signifikanzniveau und k	130

1. Einleitung

In den vergangenen Jahren finden die Methoden der Radarinterferometrie aufgrund der technologischen Weiterentwicklung und der kontinuierlichen Verfügbarkeit der Satellitensensoren in vielen Zweigen der Wissenschaft und Wirtschaft immer mehr Anwendung. Im Vergleich zu herkömmlichen Vermessungsverfahren ermöglicht die Radarinterferometrie, wirtschaftlich und mit hoher zeitlicher Auflösung flächenhafte Aussagen über Bodenbewegungen bzw. Höhenänderungen an der Tagesoberfläche mit einer Genauigkeit im Sub-Zentimeterbereich zu gewinnen.

Das Verfahren der Persistent Scatterer Interferometry (PSI) ist ein punktbezogenes radarinterferometrisches Auswerteverfahren, bei dem die zur Ableitung von Höhenänderungen benötigten interferometrischen Phasen an Objekten bestimmt werden, die sich über eine Zeitreihe aufeinanderfolgender Satelliten-Radaraufnahmen als langzeitstabile Rückstreuer erwiesen haben. Die mit diesem Verfahren ermittelten Persistent Scatterer Punkte (PS-Punkte) sind in der Realität meist künstliche Objekte mit stabilen Rückstreuereigenschaften über den gesamten Untersuchungszeitraum, z.B. Bauwerke unterschiedlicher Art, Form und Größe. Wegen der hohen Punktdichte und der damit verbundenen hohen Informationsdichte über die zeitlich variierenden Bodenbewegungen bzw. Höhenänderungen an der Tagesoberfläche und wegen der Genauigkeit im Sub-Zentimeterbereich eignet sich das Verfahren sehr gut für die Erfassung von natürlichen sowie anthropogen verursachten Bodenbewegungen, insbesondere in bebauten Gebieten.

Obwohl mit dem PSI-Verfahren das Problem der Dekorrelation, an dem das DinSAR (Differential Interferometry Synthetic Aperture Radar) -Ergebnis üblicherweise leidet, umgeht, können „Anomalien“ in den an den PS-Punkten gewonnenen Ergebnissen nicht ausgeschlossen werden. Eine derartige Anomalie weist eine deutliche, anormale Zeitreihe an einem PS-Punkt auf, die im Vergleich zu den Zeitreihen der benachbarten PS-Punkte unplausibel erscheint. Solche Anomalien können z.B. durch Messfehler oder verfahrenstechnische Berechnungsfehler oder Atmosphäreneinflüsse verursacht sein. Sie können aber auch reale Bodenbewegungen wiedergeben, die jedoch nur an dieser lokalen Stelle auftreten, und nicht repräsentativ für ein üblicherweise flächenhaftes Bodenbewegungsverhalten sind. Eine Feststellung der Ursachen für diese anormalen Zeitreihen an PS-Punkten ist nur in einem fachlichen Zusammenhang und mit zusätzlichem Fachwissen möglich, aber schwierig und mitunter aufwändig. Treten solche Anomalien bei PSI-Zeitreihen auf, werden bei der Ableitung der Höhenänderungsraten oder der Trendmodellierung der gesamten Zeitreihe falsche bzw. verfälschte Ergebnisse erzeugt, welche zu fehlerhaften Interpretationen des Bodenbewegungsverhaltens führen können.

Deshalb sollten diese auffälligen „Anomalien“ in den heutzutage vorliegenden Massendaten automatisiert erkannt, dokumentiert und von einer weiteren Nutzung ausgeschlossen werden.

Ziel dieser Arbeit ist es, mithilfe einer raumzeitlichen Clusteranalyse anormale Zeitreihen an PS-Punkten zu detektieren, die sich von der Mehrheit der Nachbarzeitreihen der umgebenden PS-Punkte unterscheiden und somit nicht repräsentativ für die Beschreibung des Bodenbewegungsverhaltens sind. Ein Ausschluss (nicht Elimination aus dem ursprünglichen Datenbestand) dieser detektierten PS-Punkte verbessert die Qualität und Aussagefähigkeit abgeleiteter (generalisierter) flächenhafter Bodenbewegungen.

Zum besseren Verständnis erfolgt in Kapitel 2 eine Einführung in die Grundlagen der Radarinterferometrie und der Persistent Scatterer Interferometry (PSI) sowie eine Darstellung exemplarischer Beispiele ihrer Anwendung zur Erfassung flächenhafter Bodenbewegungen. Da in den PSI-Ergebnissen anormale Zeitreihen an PS-Punkten auftreten können, erfolgt in Kapitel 3 eine ausführliche Betrachtung der Eigenschaften solcher Anomalien im Zusammenhang mit der Ableitung flächenhafter Bodenbewegungen. Hinsichtlich des raumzeitlichen Charakters der Anomalien von PS-Punkten wird im Rahmen der Arbeit für deren Detektion die Verwendung einer Clusteranalyse untersucht. Ausgehend von einer Einführung in die theoretischen Grundlagen einer Clusteranalyse in Kapitel 4 erfolgt in Kapitel 5 eine spezifische Untersuchung von Verfahren zur räumlichen Clusteranalyse und ihre Eignung zur Detektion von Anomalien. Des Weiteren wird in Kapitel 5 auf das Moran's-Index – Clusterverfahren, insbesondere das Local-Moran's-Index Clusterverfahren näher eingegangen. Darauf aufbauend erfolgt in Kapitel 6 eine Modifizierung des Local-Moran's-Index - Clusterverfahrens zur Detektion von anormalen Zeitreihen an PS-Punkten und eine Ergänzung dieses Clusterverfahrens durch eine kleinräumige Clusteranalyse mit lokalen Parametern. Die Anwendbarkeit des modifizierten und ergänzten Verfahrensablaufes wird in Kapitel 7 untersucht. Hierzu dienen PSI-Daten aus einem großräumigen Untersuchungsbeispiel. Abschließend erfolgt eine automatisierte Überprüfung und Validierung des Analyseergebnisses mit einem neu erstellten Kriterium. Im Kapitel 8 werden die erzielten Ergebnisse der vorliegenden Arbeit zusammengefasst und Empfehlungen für eine Anwendung gegeben.

2. Persistent Scatterer Interferometrie

2.1. Allgemeine Beschreibung der Radarinterferometrie

Die Radarinterferometrie ist ein aktives, fernerkundliches Messverfahren, d.h. satellitengetragene Radarsensoren können, im Gegensatz zu optischen Sensoren, Tag und Nacht sowie auch bei Bewölkung Messungen zur Erdoberfläche durchführen. Während sich der Radarsatellit entlang seiner Umlaufbahn bewegt, sendet der Radarsensor Mikrowellenstrahlung mit einer Wellenlänge von einem Millimeter bis zu 30 Zentimetern (entsprechend Frequenzen von 300 bis 1 GHz) seitlich der Flugbahn zur Erdoberfläche (Range-Richtung) aus und nimmt anschließend die von der Erdoberfläche zurückgestreuten Radarsignale wieder auf (Abb. 1) [SCHÄFER 2012].

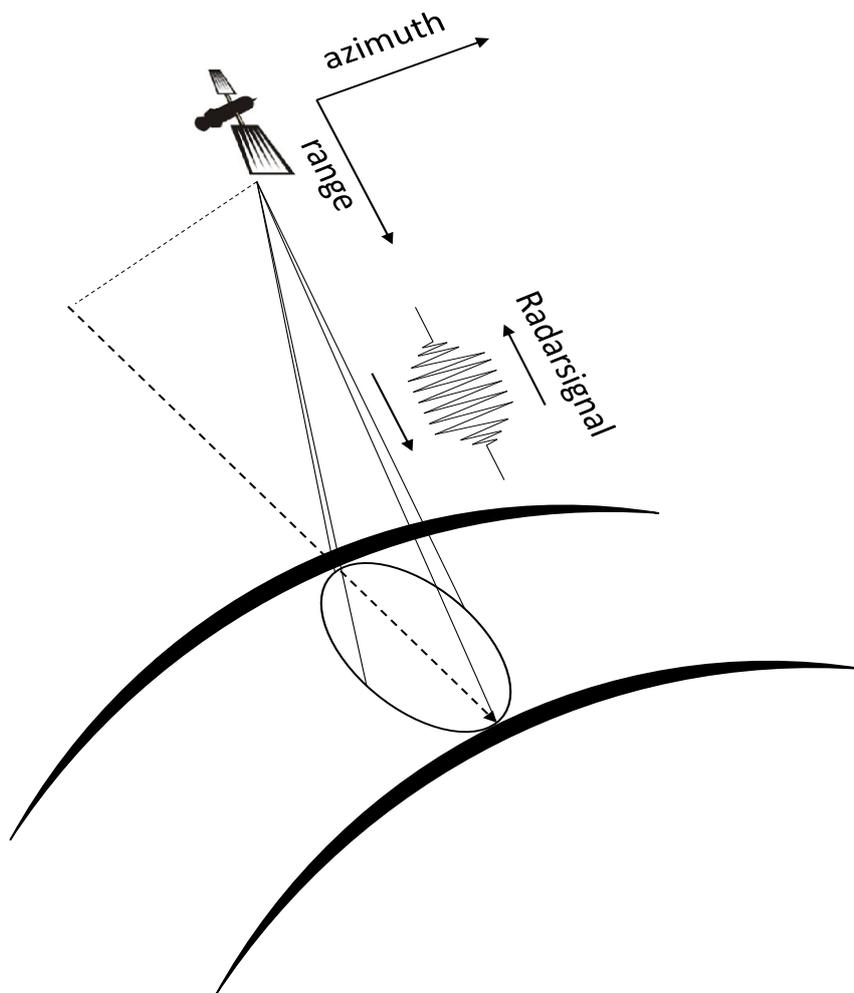


Abbildung 1: Vereinfachte Geometrie einer Radarmessung [VERÄNDERT NACH BAMLER 2008]

Gemessen werden die Rückstreuintensität und die Phasenlage, mit der der Schwingungszustand der reflektierten Radarsignale im Wertebereich von $-\pi$ bis $+\pi$ dokumentiert wird. Daraus lässt sich auch die Entfernung zwischen dem Sensor und der rückstreuenden Erdoberfläche bzw. den sich darauf befindlichen, reflektierenden Objekten bestimmen [WOODHOUSE 2006] [MASSONNET 2008].

Die differentielle Radarinterferometrie (DinSAR) ist das grundlegende Messverfahren der Radarinterferometrie zur Bestimmung von Bodenbewegungen. Hier wird das Interferogramm aus jeweils zwei Radarbildern erzeugt, das die Phasendifferenz $\Delta\varphi$ beider Aufnahmen (SAR1 und SAR2) beinhaltet und mit deren Hilfe kleine relative Abstandsänderungen ΔA und, unter Annahmen, daraus abgeleitete Höhenänderungen zwischen den beiden Aufnahmezeitpunkten t_1 und t_2 bestimmt werden können (s. Abb. 2).

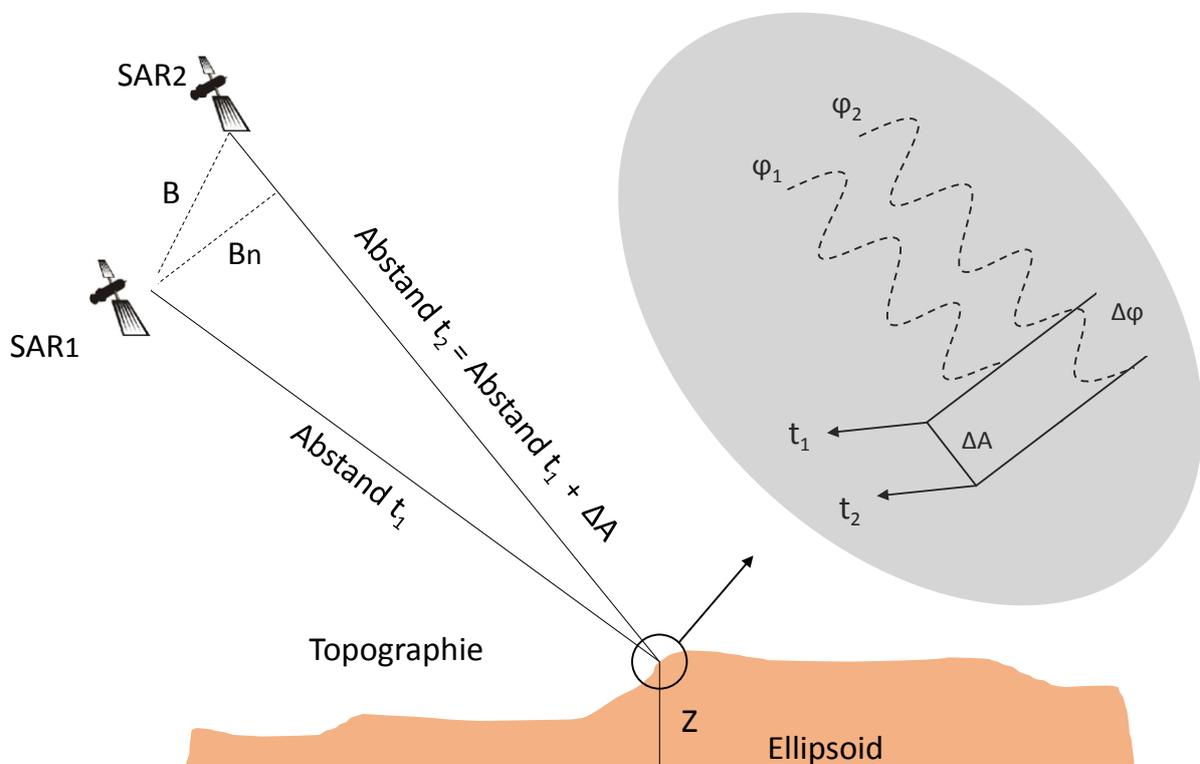


Abbildung 2: Vereinfachte Aufnahme-Geometrie der DinSAR [VERÄNDERT NACH BAMLER 2008]

Die wesentlichen Prozessschritte zur Erstellung eines differentiellen Interferogramms für die Bestimmung von Bodenbewegungen sind in Abbildung 3 dargestellt. Da die zwei Radarbilder durch unterschiedliche Aufnahmepositionen zu verschiedenen Zeitpunkten von dem Radarsensor aufgenommen wurden, besitzt ein betrachtetes Objekt in den Aufnahmen nicht die identischen Bildkoordinaten. Um daraus ein Interferogramm erstellen zu können, müssen zuerst die Radarbilder in ein einheitliches Koordinatensystem überführt werden, d.h. es muss

eine Koregistrierung der Radarbilder erfolgen. Hierfür wird das Amplitudenbild, das die Intensität des zurückgestreuten Radarsignals aufzeichnet, zur Identifizierung des gleichen Objektes (Pixel auf dem Radarbild) verwendet. Anschließend wird das beim Überlagern beider Radarbilder durch Frequenzen ausgelöste Phasenrauschen durch eine spektrale Filterung reduziert, so dass die Radarbilder über ein gemeinsames Frequenzspektrum verfügen [PRATI 1993]. Im Anschluss daran erfolgt die Erstellung des Interferogramms. Es beinhaltet ein Abbild der Phasendifferenzen, die sich aus verschiedenen Phasenanteilen (s. Abb. 4) zusammensetzen. Nachdem die nicht deformationsbedingten Phasenanteile (z.B. topographie- sowie atmosphärenbedingte Phasenanteile und Rauschen) eliminiert und die Phasenmehrdeutigkeit durch Phase Unwrapping-Verfahren gelöst wurden, lässt sich letztendlich die deformationsbedingte Phasendifferenz durch verschiedene Verfahren (z.B. Multilook-Verfahren) zwischen den beiden Radarbildern berechnen [JUST 1994] [SCHWÄBISCH 1995] [KIRCHER 2004] [ROTT 2001].

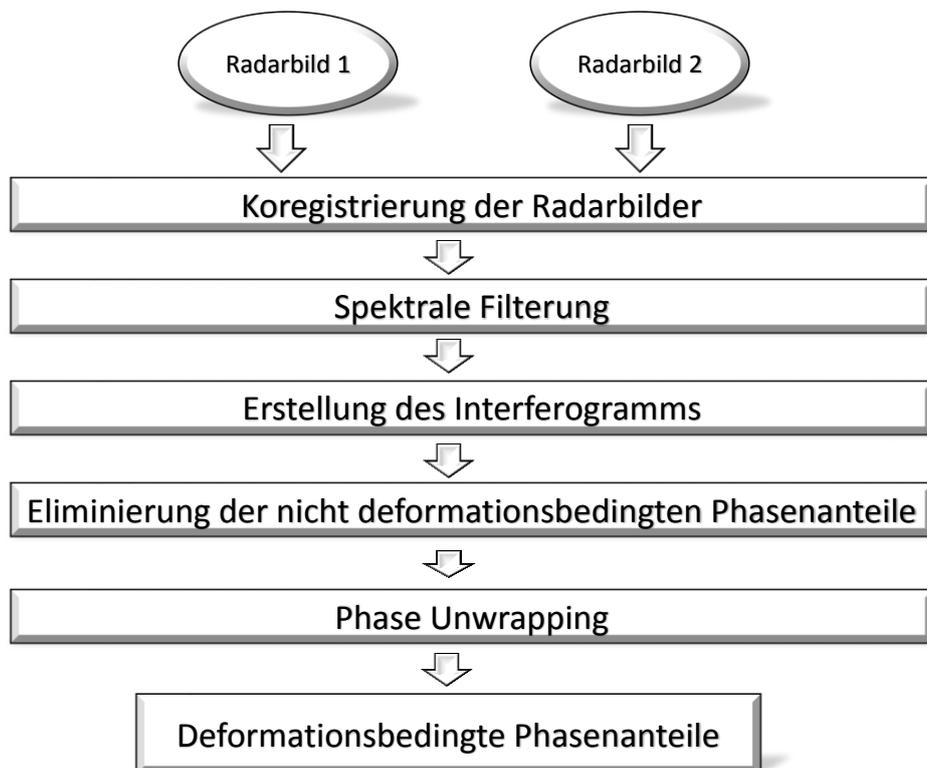


Abbildung 3: Vereinfachter Prozessierungsablauf zur Bestimmung deformationsbedingter Phasenanteile (Bodenbewegungen) [VERÄNDERT NACH RABUS 2002]

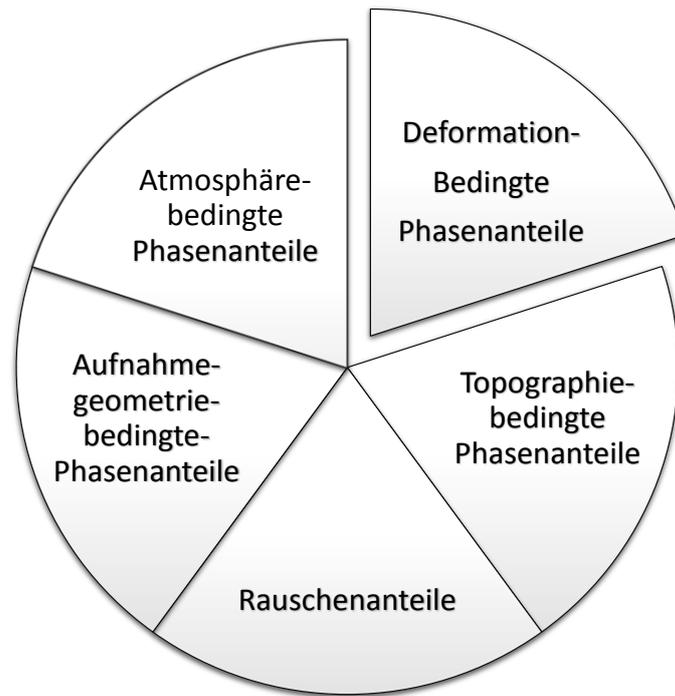


Abbildung 4: Zusammensetzung der Phasenanteile

Da der Abstand zwischen den beiden Aufnahmepositionen viel kleiner als die Entfernung zwischen dem Satellit und der Erdoberfläche ist, kann angenommen werden, dass die Blickrichtungen des Sensors nahezu parallel sind. Die dadurch entstandene Phasenkomponente wird als aufnahmegeometriebedingte Phasendifferenz gekennzeichnet. Durch die Verwendung einer bestangepassten Referenzfläche und genauer Orbitinformationen können die Phasenanteile korrigiert werden [GEUDTNER 1995]. Die topographiebedingten Phasenanteile können unter Zuhilfenahme eines digitalen Höhenmodells (DHM, DOM) eliminiert werden [MASSONNET 1993]. Da sich höhenmäßige Abweichungen von dem DHM als Verschiebungsfehler in topographiebedingten Phasenanteilen auswirken, sollte das verwendete DHM möglichst aktuell und hochauflösend sein [SCHWÄBISCH 1995] [WALTER 2012]. Die großräumigen atmosphärischen und lokalen troposphärischen Einflüsse führen zu zusätzlichen phasenwirksamen Signalverzögerungen und die entsprechenden Phasenanteile können nur bedingt durch statistische Analysen unter Verwendung mehrerer Radarbilder und räumlicher Filterungen aus der differentiellen Phase eliminiert werden [SCHÄFER 2012]. Die Rauschanteile der Phase, die meistens auf Veränderungen der Rückstreueigenschaften der betrachteten Objekte (Signaldekorrelation) zurückzuführen sind, werden unter Nutzung von geeigneten Phasenfiltern teilweise unterdrückt [HANSEN 2001] [SCHÄFER 2012].

Entscheidend für die DinSAR-Auswertungen ist die räumliche und auch zeitliche Kohärenz der auszuwertenden Signale zwischen zwei Radarbildern des gleichen Gebietes. Der Verlust der Kohärenz (Dekorrelation) in den Interferogrammen führt zu systematischen Fehlern der abgeleiteten Phasendifferenz und damit der entsprechenden Entfernungsänderung bis hin zur Unbrauchbarkeit des Interferogramms, da aus ihm keine Phaseninformationen mehr bestimmt werden können [LEIJEN 2004] [HANSEN 2001].

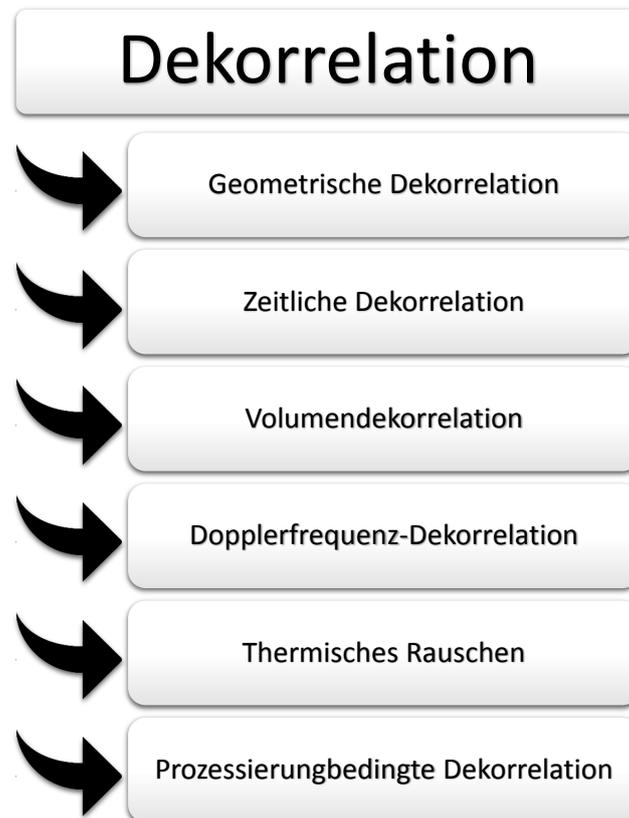


Abbildung 5: Arten von Dekorrelationen des Radarsignals

Ursachen für eine Dekorrelation (s. Abb. 5) können sein [WALTER 2012] [SCHÄFER 2012]:

- Geometrische Dekorrelation durch die leicht voneinander abweichenden Blickwinkel zweier Radaraufnahmen [SCHEIBER 2004].
- Zeitliche Dekorrelation aufgrund der jahreszeitabhängigen physikalischen Veränderungen der Geländeoberfläche und dadurch bedingter Änderungen der Rückstreuungseigenschaft (z.B. vegetationsbedingte Veränderung im Sommer, wetterbedingte Veränderung durch Schneefall im Winter) [WALTER 2012] [SCHÄFER 2012].
- Volumendekorrelation infolge der Eindringtiefe und Volumenstreuung der Radarstrahlung in unterschiedlichen Medien [ZEBKER 2000].

- Dopplerfrequenz-Dekorrelation aufgrund von Unterschieden der Dopplerschwerpunkte zwischen zwei Radaraufnahmen, verursacht durch die leichte Differenz des seitlichen „Schielens“ des Sensors gegenüber der Azimut-Richtung [MIRANDA 2004].
- Thermisches Rauschen durch die Charakteristika des elektronischen Sensorsystems und der Antenne [HANSSEN 2001].
- Prozessierungsbedingte Dekorrelation aufgrund von Fehlern bei der interferometrischen Datenakquisition, Prozessierung und Weiterverarbeitung der SAR-Daten (z.B. Fehler bei der Koregistrierung zweier Radaraufnahmen und bei der Fokussierung der SAR-Daten) [JUST 1994].

Eine erfolgreiche DinSAR - Auswertung setzt eine stabile Kohärenz zwischen den zu unterschiedlichen Zeitpunkten aufgenommenen Radarbildern voraus. Verschlechtert wird die Qualität der DinSAR-Analyse allerdings durch die oben angesprochene Dekorrelation. Die Signaldekorrelationen können räumliche Lücken in den Interferogrammen und somit teilweise große Unterbrechungen von Beobachtungszeitreihen in bestimmten Gebieten verursachen. Daher kann die Nutzung der DinSAR - Auswertung für die Erfassung von Bodenbewegungen über längere Zeiträume eingeschränkt werden. Um dieses Problem zu umgehen, bzw. zu mindern und möglichst umfangreiche Datenstapel für die Erfassung von Bodenbewegungen effizient nutzen zu können, bietet die Methode der Persistent Scatterer Interferometry (PSI) eine geeignete Lösung [KAMPES 2006].

2.2. Grundlagen der PSI-Technologie

Das Verfahren der Persistent Scatterer Interferometry (PSI) wurde von FERRETTI entwickelt. Ziel des punktbezogenen PSI-Verfahrens ist die Ableitung von vollständigen Bodenbewegungszeitreihen für jeden Persistent Scatterer (PS) Punkt. Die PS-Punkte sind in der Realität meist künstliche Objekte mit stabilen Rückstreueigenschaften über den gesamten Untersuchungszeitraum, z.B. Bauwerke unterschiedlicher Art, Form und Größe. Daher eignet sich das Verfahren sehr gut für die Erfassung von Bodenbewegungen vor allem in bebauten Gebieten [FERRETTI 2001].

Zu Beginn der PSI-Auswertung werden die Satelliten-Rohdaten durch eine Vorprozessierung zu Single Look Complex (SLC) Daten umgewandelt. Dabei werden die empfangenen Radarsignale zu einem Radarbild fokussiert und die darin enthaltenden Phasen und Intensitäten in ein komplexes Zahlenformat konvertiert. Die PSI-Auswertung erfordert einen SLC-Stapel mit mindestens 30 aufeinander folgenden Radarbildern gleicher Aufnahmegeometrie des zu untersuchenden Gebietes [COLESANTI 2003]; je mehr Szenen zur Verfügung stehen, desto qualitativ hochwertiger kann das Ergebnis einer PSI-Auswertung sein.

Nach der Vorprozessierung findet die Koregistrierung aller Radarszenen des Datenstapels jeweils zu einer Masterszene statt. Die auszuwählende Masterszene sollte über eine geringe atmosphärische Beeinflussung verfügen. Darüber hinaus sind sowohl die effektive Basislinie des Radarsensors als auch der Aufnahmezeitpunkt der Radarbilder als weitere Auswahlkriterien zu berücksichtigen [KIRCHER 2004]. Anschließend erfolgt die Identifizierung von PS-Punkten. Dabei werden potenzielle PS-Punkte anhand ihrer Rückstreuintensität in verschiedenen Szenen detektiert. Für diese erfolgt dann die Erstellung von differentiellen Punktinterferogrammen. Darauf aufbauend wird für jeden PS-Punkt ein individuelles Modell der Phasenkomponenten erstellt und über eine iterative zweidimensionale Regressionsanalyse parametrisiert. Danach wird ein Referenzpunkt ausgewählt, bezogen auf den die Bodenbewegungen jedes PS-Punktes berechnet werden. Bei der Auswahl des Referenzpunktes muss sichergestellt werden, dass er stabil, also nicht von Bodenbewegungen betroffen ist. Im Anschluss daran erfolgt die Eliminierung der atmosphärenbedingten Phasenanteile und der Rauschteile. Letztlich kann die relative Bewegung eines jeden PS-Punkt zum Referenzaufnahmezeitpunkt in Blickrichtung des Sensors (LOS, Line of Sight) ermittelt werden. Außerdem werden die Deformationsrate und die Höhenkorrektur, die durch die Nutzung eines ungenauen oder nicht aktuellen DHM (oder DOM) verursacht wird, bestimmt [KAMPES 2006].

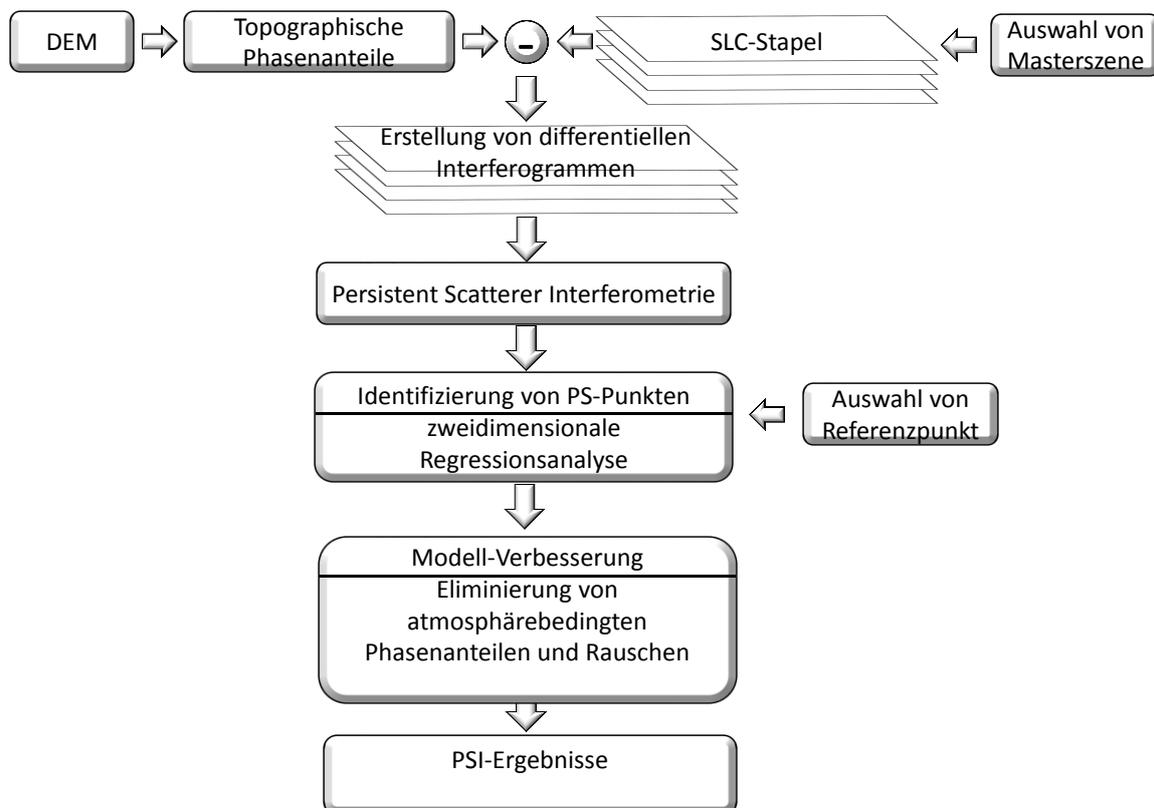


Abbildung 6: Wesentliche Bearbeitungsschritte der PSI-Auswertung [VERÄNDERT NACH ATHAYDE PINTO 2015]

Das Flussdiagramm der wesentlichen Arbeitsschritte einer PSI-Auswertung ist in Abbildung 6 dargestellt.

2.3. Datenaufbereitung und Postprocessing der PSI-Ergebnisse

Nach der PSI-Datenprozessierung werden die detektierten PS-Punkte in einem kartesischen Koordinatensystem georeferenziert. Dadurch verfügt jeder PS-Punkt über eine bestimmte geographische Position innerhalb des Untersuchungsgebietes. Für jede ausgewertete Radarszene wird einem PS-Punkt zunächst ein Abstandsänderungswert in LOS zugewiesen, der unter Berücksichtigung des Einfallswinkels und der Annahme, dass keine Horizontalbewegungen aufgetreten sind, in einen Höhenänderungswert umgerechnet werden kann. Er repräsentiert die relative Höhenänderung des PS-Punktes zwischen dem jeweiligen Aufnahmezeitpunkt und dem ausgewählten Referenzzeitpunkt und bezogen auf einen lagestabilen Referenzpunkt. Die Höhenänderungswerte aller PS-Punkte zum Referenzzeitpunkt besitzen deshalb den Wert 0. Unter Berücksichtigung aller ausgewerteten Radarszenen erhält man somit die zeitlich geordnete Folge von Höhenänderungen des PS-Punktes, eine sogenannte Höhenänderungszeitreihe. In Abbildung 7 ist eine typische Höhenänderungszeitreihe eines PS-Punktes (Lage des PS-Punktes wurde mit der Farbe hellblau eingefärbt) dargestellt. Jedes Kreuz kennzeichnet einen Höhenänderungswert zwischen dem jeweiligen Aufnahmezeitpunkt und dem gewählten Referenzzeitpunkt. Man könnte deshalb auch von einer Zeitreihe von Höhenänderungsdifferenzen sprechen, auch weil sie sich auf einen Referenzpunkt beziehen. Diese Höhenänderungszeitreihen dienen als wesentliche Grundlage zur Beschreibung von Deformationen der Erdoberfläche.

Darauf aufbauend kann außerdem auch der Höhenänderungstrend mittels passender Regressionsmodelle abgeleitet werden, womit die langzeitige Analyse der zeitlichen Entwicklung des Höhenänderungsverhaltens beschrieben werden kann.

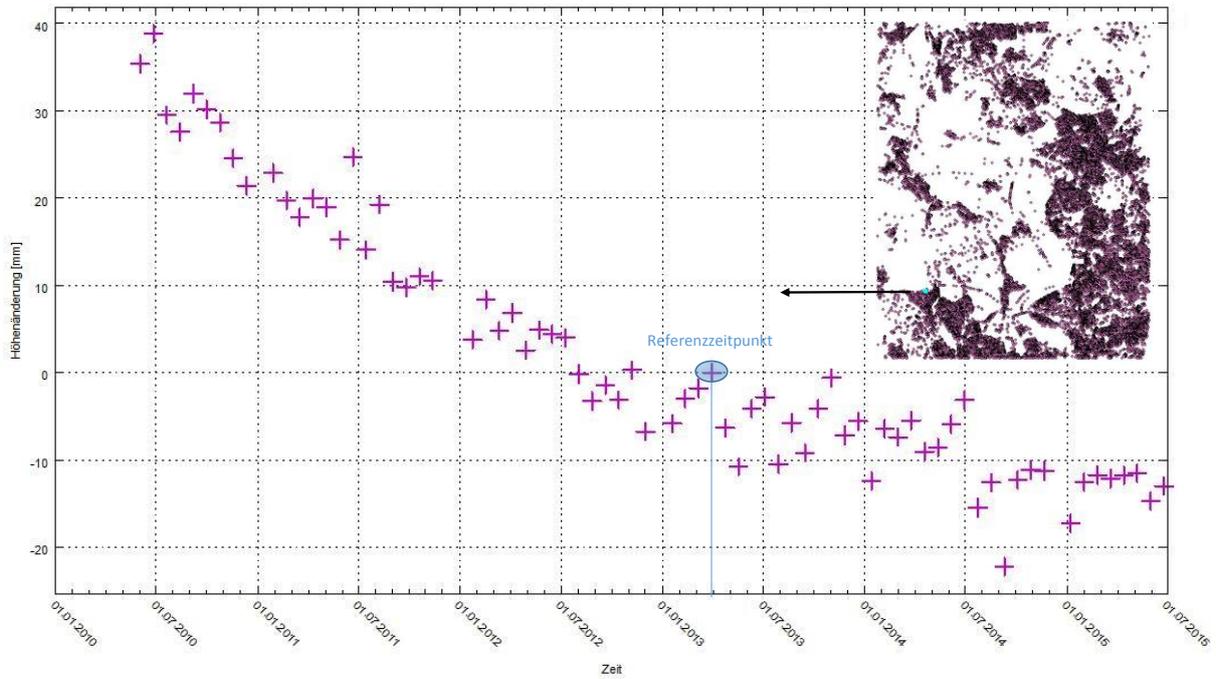


Abbildung 7: Höhenänderungszeitreihe eines PS-Punktes

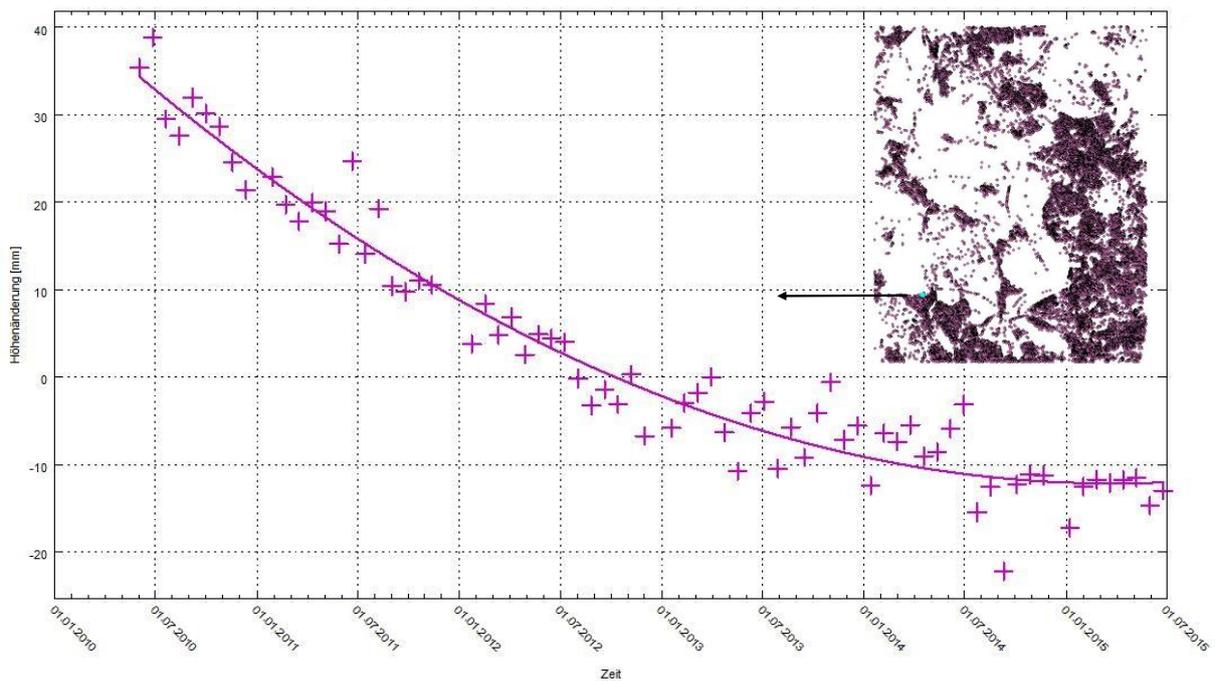


Abbildung 8: Ausgleichendes Trendpolynom 2. Grades der Höhenänderungszeitreihe eines PS-Punktes

Das PSI-Verfahren „Interferometric Point Target Analysis (IPTA)“ der Gamma SAR and Interferometry Software der Firma Gamma Remote Sensing and Consulting AG bietet beispielsweise die Möglichkeit, anhand einer zweidimensionalen Regressionsanalyse einen linearen Trend der LOS-Werte für jeden PS-Punkt zu schätzen [WERNER 2003]. Da die Bodenbewegung, die z.B. durch Stilllegung bergbaulicher Aktivitäten hervorgerufen ist, öfters

einen nichtlinearen Trend aufweist, wurde am Institut für Geotechnik und Markscheidewesen der TU Clausthal ein funktionales orthogonales Polynommodell zur automatischen Bestimmung des bestangepassten Polynommodells für die ermittelten LOS- bzw. Höhenänderungszeitreihen von PS-Punkten entwickelt [BUSCH 2014]. Das angewendete orthogonale Polynommodell ermöglicht, automatisiert den bestangepassten Polynomgrad für die Zeitreihe jedes durch PSI-Auswertung detektierten PS-Punktes zu bestimmen und so auch das wegen der zeitlichen Variabilität einer oder mehrerer Ursachen aufgetretene nichtlineare Bodenbewegungsverhalten aus den Höhenänderungszeitreihen zu beschreiben (s. Abb. 8). Dadurch können nicht nur die durch fehlende Szenen vorhandenen Datenlücken in den Höhenänderungszeitreihen gefüllt, sondern auch die Genauigkeiten der ermittelten Höhenänderungen anhand der Standardabweichungen angegeben werden. Mit Hilfe der farblichen Darstellung von klassifizierten Polynomgraden lassen sich intuitiv die Bodenbewegungsbereiche von den Bereichen ohne Bodenbewegung unterscheiden (s. Abb. 9). Ferner ist es auch möglich, anhand der Standardabweichungen Gebiete mit hoher oder niedriger Genauigkeit identifizieren [BUSCH 2014]. Die Ergebnisse der Polynomausgleichung können weitere Analysen und Interpretationen sinnvoll unterstützen.

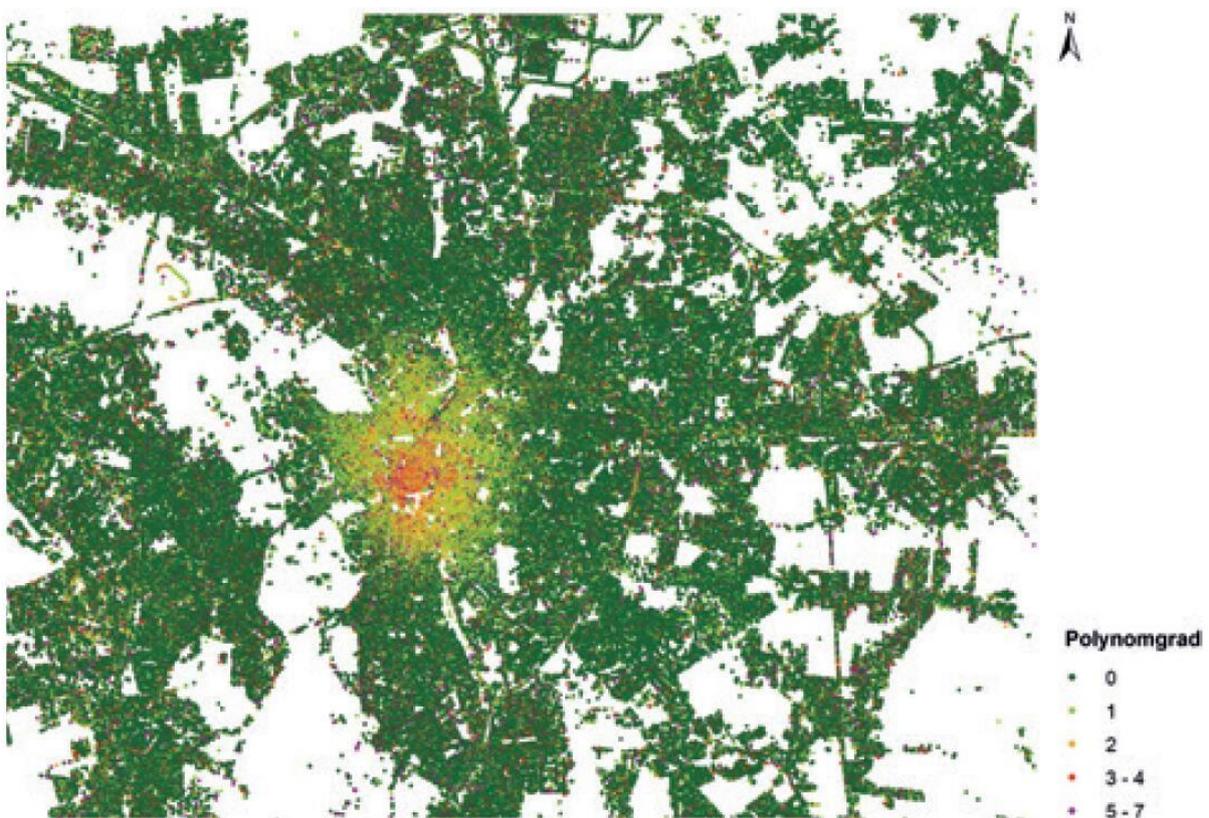


Abbildung 9: Polynomgrade von PS-Punkt-Zeitreihen im Großraum Leipzig [BUSCH 2014]

Die orthogonale Polynom-Trendmodellierung von PSI-Zeitreihen wurde in Form einer funktionalen Erweiterung in ArcGIS implementiert. Neben der grafischen Darstellung des Trendpolynoms erhält man als Ergebnis für jeden PS-Punkt den Polynomgrad sowie die

Polynomparameter eines bestangepassten Polynoms, die verbesserten Höhenänderungen (Funktionswerte), die Verbesserungen (Differenz zwischen den ursprünglich gemessenen Höhenänderungen und den Funktionswerten) sowie weitere Genauigkeitsangaben (Standardabweichung der gemessenen Höhenänderung, der Polynomparameter und sonstiger abgeleiteter Größen). Einige dieser Größen werden für die nachfolgende Clusteranalyse angewendet.

2.4. Anwendungsbeispiele

Zum besseren Verständnis sollen nachfolgend einige typische Anwendungsfälle der PSI-Technologie beispielhaft vorgestellt werden.

Mit dem PSI-Verfahren lassen sich Bodenbewegungen großflächig mit einer Genauigkeit im Bereich von einigen Millimetern erfassen. Im Vergleich zum DinSAR-Verfahren begegnet man beim PSI-Verfahren dem Problem der Dekorrelationen, worunter die Genauigkeit des DinSAR-Verfahrens oftmals leidet, sodass nur die Punktstreuer mit einem stabilen Phasenverhalten über lange Zeiträume, die sogenannten Persistent Scatterer (PS-Punkte), für die Auswertung genutzt werden können.

Angesichts des Charakters des PSI-Verfahrens eignet sich dieses punktbezogene radarinterferometrische Verfahren für markscheiderische und vermessungstechnische Aufgabenstellungen unter Berücksichtigung der fachlichen Anforderungen an das markscheiderisch-geodätische Bodenbewegungsmonitoring [DMV 2013].

Das Ruhrgebiet ist durch Bergbau- und Montanindustrie geprägt. Die intensiven Abbautätigkeiten im Rahmen der großindustriellen Rohstoffgewinnung führen zu fortschreitenden Höhenänderungen der Tagesoberfläche. Um die dadurch verursachten Veränderungen bzw. die Folgen des anthropogenen Eingriffs auf die Natur, Infrastruktureinrichtungen und Bauwerke jeglicher Art zu überwachen, hat die Bezirksregierung Arnsberg die Abteilung Markscheidewesen und Geoinformation des Instituts für Geotechnik und Markscheidewesen der TU Clausthal beauftragt, die Senkungserscheinungen außerhalb der in den Rahmenbetriebsplänen dargestellten prognostizierten Einwirkungsbereiche aktiver und in jüngerer Zeit stillgelegter Steinkohlebergwerke zu analysieren. Dabei kommt das PSI-Verfahren für die radarinterferometrische Ermittlung der räumlichen Ausdehnung und Größenordnung von Senkungen zum Einsatz. Genutzt werden die verfügbaren Radardaten des europäischen Satelliten ENVISAT und des kanadischen Satelliten Radarsat-2. So standen beispielsweise für das Steinkohlenbergwerk Walsum 71 ENVISAT ASAR Satellitenaufnahmen zwischen 21.12.2003 und 10.10.2010 mit einer Wiederholungsrate von ca. 35 Tagen und 77 Radarsat-

2 Standardmode Satellitenaufnahmen vom 03.06.2010 bis zum 25.06.2015 mit einer Wiederholungsrate von ca. 24 Tagen zur Verfügung. Die Envisat ASAR Daten besitzen eine Bodenauflösung von etwa 23 m x 8 m und die Radarsat-2 Daten von etwa 23 m x 5 m verfügen [BUSCH 2016].

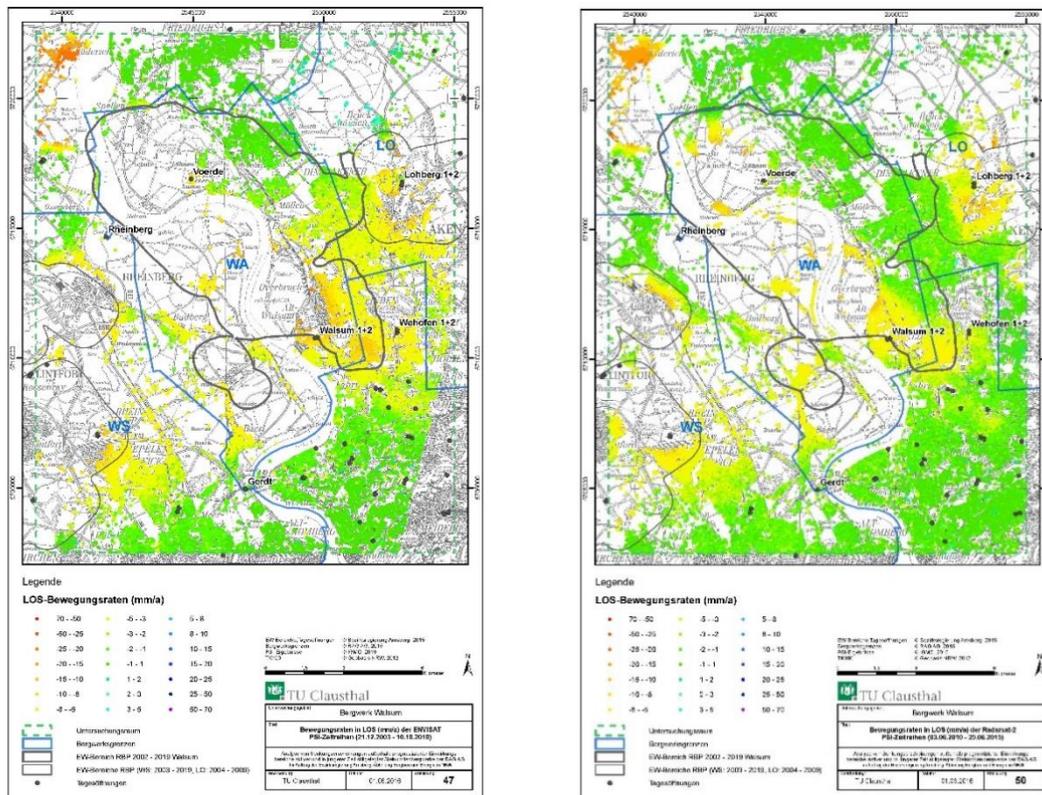


Abbildung 10: Bewegungsrate in LOS (mm/a) der ENVISAT und Radarsat-2 PSI-Zeitreihen im Bereich des Bergwerk Walsum [BUSCH 2016]

Die radarinterferometrischen Auswertungen erfolgten mit dem PSI Verfahren „Interferometric Point Target Analysis (IPTA)“ der Gamma SAR und Interferometry Software (Firma Gamma Remote Sensing Research and Consulting AG, Schweiz) [WERNER 2003]. Im Untersuchungsraum konnten nach einer Datenbereinigung 30124 PS-Punkte für den ENVISAT-Datenstapel mit 69 Szenen und 63253 PS-Punkte für den Radarsat-2-Datenstapel mit 74 Szenen detektiert werden. Für jeden detektierten PS-Punkt wird als Ergebnis einer PSI-Auswertung der zeitliche Bewegungsverlauf in LOS-Richtung zwischen der ersten und letzten Radarszene eines Stapels berechnet. Anschließend werden die Bewegungsraten der ENVISAT und Radarsat-2 PSI-Zeitreihen in LOS-Richtung mittels einer Klasseneinteilung und Farbskala auf einer topographischen Karte dargestellt (s. Abb. 10) [BUSCH 2016].

In der Abbildung 10 ist deutlich zu erkennen, dass in städtischen Bereichen die räumliche Dichte der detektierten PS-Punkte deutlicher höher als in ländlichen Gebieten ist, da in vegetationsreichen Gebieten nur wenige langzeitstabile Punktstreuer zu finden sind. In

Gebieten mit geringer PS-Punktdichte sind Aussagen über Bodenbewegungen nur eingeschränkt möglich. Dagegen sind in den städtischen Gebieten, z.B. hier in der Stadt Duisburg, sehr viele PS-Punkte detektiert worden [BUSCH 2016]. Mit Hilfe der PSI-Methode sowie generell mit der Radarinterferometrie lassen sich keine flächendeckenden Aussagen erzielen, jedoch in bebauten Gebieten (u.a. abhängig von der Bodenauflösung des verwendeten Sensors) sehr wohl eine hohe Punktdichte.

Da sich unter der Annahme, dass keine Horizontalbewegungen aufgetreten sind, mit Hilfe der Einfallswinkel auch der Höhenänderungsverlauf und die Größe der gesamten Höhenänderungen innerhalb des Untersuchungszeitraums ableiten lassen, können zur Verifizierung der radarinterferometrisch ermittelten Höhenänderungen zeitlich vergleichbare Ergebnisse klassischer geodätischer Messungen genutzt werden. Jedoch ist dies nur eingeschränkt möglich, da z.B. Nivellementsergebnisse nur an sehr wenigen Stellen zur Verfügung stehen und häufig in deutlich größeren Zeitabständen gemessen werden. Hinzu kommt, dass ein direkter Vergleich nur mit dem nächsten benachbarten PS-Punkt möglich ist, also ein räumlicher Unterschied besteht. Eine beispielhafte Gegenüberstellung der mit beiden Verfahren gewonnenen Ergebnisse ist in Abbildung 11 dargestellt und zeigt eine sehr gute Übereinstimmung der mit beiden Verfahren erzielten Ergebnisse.

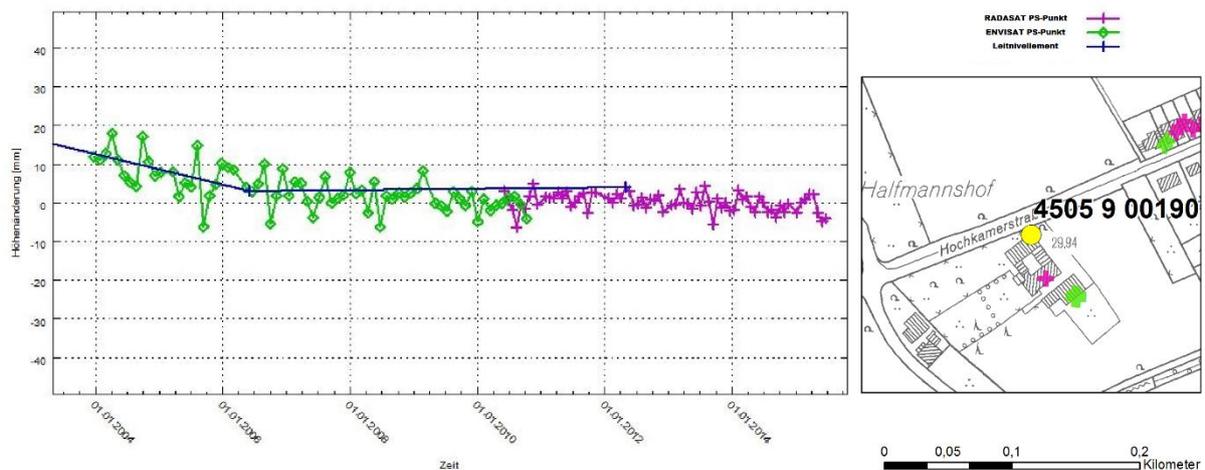


Abbildung 11: Zeitreihen nivellitisch und radarinterferometrisch bestimmter Höhenänderungen an einem ausgewählten Punkt des Leitnivelements im Bereich des Bergwerkes Walsum [BUSCH 2016]

Zum Setzungsmonitoring bei großen geotechnischen Baumaßnahmen findet die PSI-Methode ebenso ihre Einsatzmöglichkeit (s. auch [BUSCH 2014]). In [MARK 2012] wird ein Beispiel über die Anwendung des PSI-Verfahrens zum flächendeckenden Setzungsmonitoring beim innerstädtischen Tunnelbau im Rahmen des Bauprojektes „Wehrhahn-Linie“ in Düsseldorf vorgestellt. Dabei wurden 24 Szenen des hochauflösenden deutschen Radarsatelliten TerraSAR-X im Zeitraum von Januar bis Dezember 2011 mit einer Bodenauflösung von 3 m x

3 m verwendet. Es konnten ca. 16000 PS-Punkte detektiert werden. Die für jeden PS-Punkt abgeleiteten Bewegungsraten in LOS-Richtung sind auf einer Luftbildkarte farbig dargestellt (s. Abb. 12). Das radarinterferometrisch abgeleitete Ergebnis verfügt über eine Genauigkeit von einigen Millimetern und stimmte in den meisten Gebieten mit den nivellitischen Vermessungen sehr gut überein. Als vorteilhaft erwies sich im Vergleich zur terrestrischen Vermessungsmethode, mit der nur eine geringe Anzahl von Messungen in größeren Zeitintervallen durchgeführt werden kann, dass das PSI-Verfahren flächenhafte Vermessungen ohne aufwändige Punktvermarkungskosten über einen langen Zeitraum und mit kurzen Zeitabständen (z.B. 11 Tage für TerraSAR-X) ermöglicht [MARK, ET AL. 2012].



Abbildung 12: Bewegungsraten in LOS von PS-Punkten in Düsseldorf (mm/Jahr) [VERÄNDERT NACH MARK 2012]

Bei der Überwachung von Höhenänderungen der Tagesoberfläche, die durch den Betrieb eines untertägigen Erdgasspeichers verursacht werden, liefert die PSI-Methode auch zufriedenstellende Ergebnisse [KNOSPE, ET AL. 2011]. Der als Beispiel betrachtete Aquiferspeicher befindet sich in einer Tiefe von ca. 800 m unter NN westlich des Großraums Berlin. Durch die Druckänderungen bei der Befüllung des Erdgasspeichers traten Hebungen an der Tagesoberfläche auf, die mittels der PSI-Methode sowohl räumlich als auch in ihren zeitlichen Entwicklungsverläufen festgestellt werden konnten. Dabei sind insgesamt 65 Szenen der Satelliten ERS-1 und ERS-2 über einen Zeitraum von mehr als acht Jahren

ausgewertet worden. Als Ergebnis sind die abgeleiteten Bewegungsraten für jeden PS-Punkt in der Abbildung 13 dargestellt.

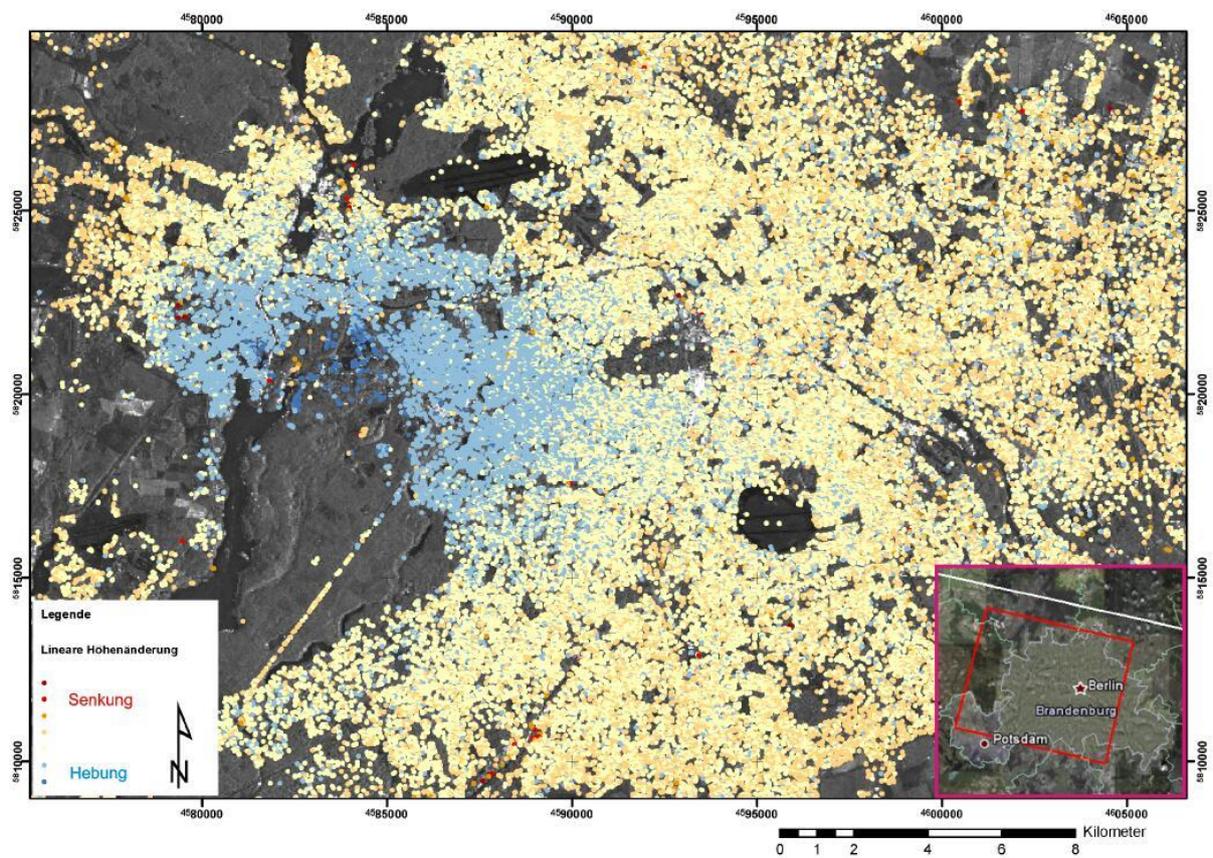


Abbildung 13: Höhenänderungsraten von PS-Punkten über einem Aquifer-Gas-Speicher im westlichen Teil Berlins [KNOSPE, ET AL. 2011]

3. Anomalien in den PSI-Ergebnissen

Es hat sich bereits in mehreren Anwendungsgebieten erwiesen, dass die PSI-Methode sehr gut für die Überwachung natürlicher sowie anthropogen verursachter Veränderungen geeignet ist. Obwohl diese Methode nicht von dem Problem der Dekorrelationen, an dem das DinSAR-Ergebnis üblicherweise leidet, betroffen ist, können „Anomalien“ in den Messwerten nicht ausgeschlossen werden, welche zu fehlerhaften Interpretationen der gesamten PSI-Ergebnisse führen können. Um ein realitätsnahes Analyseergebnis erzielen zu können, sind solche Anomalien zu detektieren und ggf. zu korrigieren oder zu eliminieren. In den beiden folgenden Kapiteln soll zunächst eine Definition solcher „Anomalien“ und ihr Verständnis im Rahmen einer Anwendung in dieser Arbeit erfolgen.

3.1. Definition von Anomalien

Im Allgemeinen versteht man unter „Anomalien“ extreme oder ungewöhnliche Messwerte, die stark von den anderen Messwerten abweichen [BARNETT 2000]. Eine Anomalie stellt also eine Beobachtung dar, die aufgrund einer vordefinierten Modellvorstellung im Vergleich zu anderen Beobachtungen so extrem auffällig ist, dass diese von dem durch die Mehrheit der Beobachtungen charakterisierten Datenverlauf abweicht und somit aus dieser ausgeschlossen werden kann [HAWKINS 2004]. Dabei handelt es sich immer um ein relatives Verhältnis, welches nur im Zusammenhang mit einem global oder lokal vorliegenden Datensatz zu bewerten ist [CLARK 1989]. Statistisch gesehen sind sie deutlich isolierte Werte, die sich auffallend von den anderen Werten unterscheiden (s. Abb. 14) [HASLETT 1991] [RAMBOLD 1999].

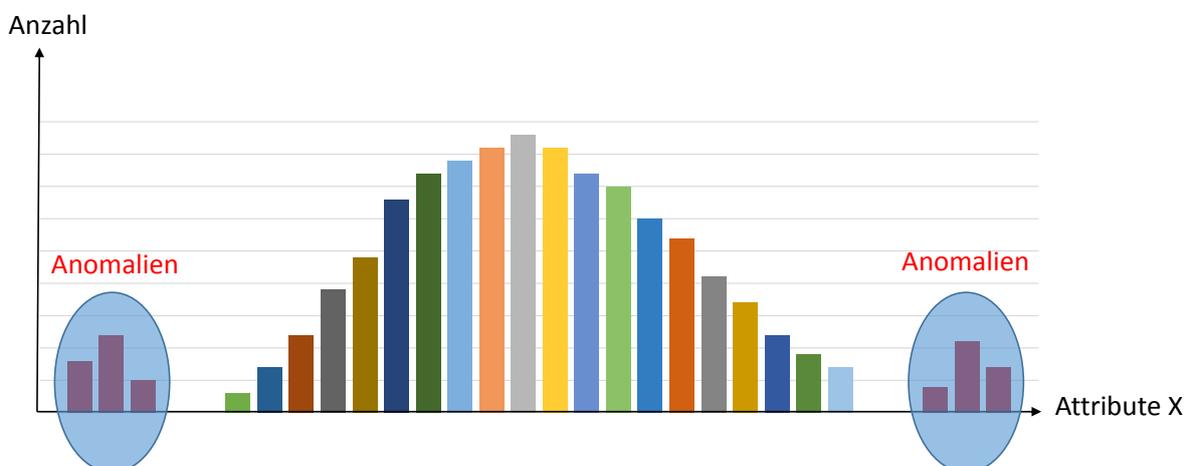


Abbildung 14: Typische Anomalien im Histogramm

In den Geowissenschaften sind die Anomalien durch ihre räumlichen Eigenschaften geprägt. Es gilt generell, dass je näher sich die räumlichen Phänomene zueinander befinden, desto ähnlicher sind sie. So hat Waldo Tobler im Jahr 1970 das berühmte erste geographische Gesetz eingebracht. Es lautet: *“Everything is related to everything else, but near things are more related than distant things”* [TOBLER 1970]. Unter diesem Gesichtspunkt besitzt eine Anomalie in der Geowissenschaft eine deutliche, unerwartete Merkmalsausprägung im Vergleich zu den anderen Objekten in ihrer Nachbarschaft, z.B. eine geologische Störung, die eine deutliche lokale Veränderung in der großräumigen Struktur der geologischen Schichten hervorruft oder ein hoher gemessener lokaler Schwermetallgehalt im Boden, der im Vergleich zu Messwerten in der Nachbarschaft unplausibel erscheint. Ein vereinfachtes Beispiel zeigt Abbildung 15, in dem der Messwert (28, rotes Quadrat) einen gemessenen Schwermetallgehalt bezeichnet, der sich deutlich von den benachbarten Messwerten (schwarze Kreise) mit durchschnittlichen Werten von kleiner als 5 unterscheidet und somit als Anomalie in Erscheinung tritt.

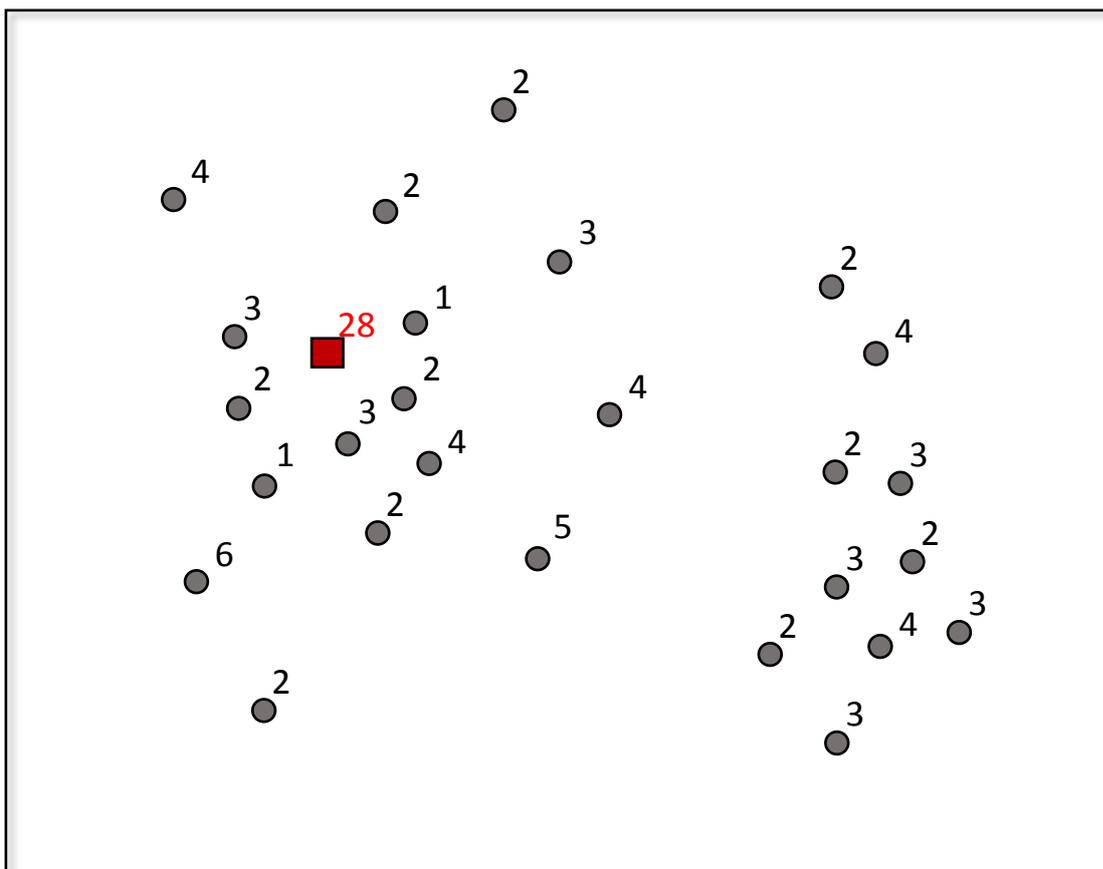


Abbildung 15: Vereinfachte Darstellung einer Anomalie (rotes Quadrat) im Vergleich zu den Messwerten in seiner Nachbarschaft

Hierbei sollte auch bedacht werden, dass solche, sogenannten Anomalien nicht „automatisch“ mit fehlerhaften Beobachtungen gleichgesetzt werden dürfen. Anomalien können entweder falsche Beobachtungen sein, die durch Mess-, Erfassungs- oder Rechenfehler entstanden sind, oder korrekte Beobachtungen sein, die den Zustand an dieser lokalen Stelle richtig beschreiben, jedoch nicht im Einklang mit den anderen Beobachtungen stehen oder im Vergleich zu einer Modellannahme extrem auffallen. Eine Ermittlung der Ursachen für diese auffälligen Werte ist schwierig und mitunter aufwändig, in Abhängigkeit vom jeweiligen fachlichen Zusammenhang und dem vorhandenen fachlichen Kenntnisstand. In der Regel sind hierfür zusätzliche Informationen sowie weitere, mitunter komplexe Ursache-Wirkungsanalysen erforderlich.

Somit kann festgestellt werden, dass die Detektion von Anomalien (im Kontext zu anderen bekannten Messwerten oder einer Modellvorstellung) zu einem Interpretationsproblem führt. Sind die erfassten Anomalien Ausdruck einer fehlerhaften Messung oder einer unerwarteten Realität? Fehlen eindeutige und belegbare Gründe für eine Antwort, so kann diese Frage, insbesondere im Rahmen einer Analyse von Massendaten, pragmatisch nur im Hinblick auf die (fachliche) Aufgabenstellung (hier die Ermittlung von flächenhaften Bodenbewegungen) und eine Modellvorstellung beantwortet werden.

Im Rahmen dieser Arbeit sollen deshalb einzelne, anormale Beobachtungswerte, die sich räumlich „isoliert“ von der Mehrheit der umgebenden Messwerte unterscheiden, detektiert werden, so dass, nach Ihrem Ausschluss (nicht Elimination aus dem ursprünglichen Datenbestand), die Modellierung (generalisierter) flächenhafter Bodenbewegungen ermöglicht und verbessert wird. Dafür ist von geringer Bedeutung, ob die detektierten Anomalien fehlerhafte Einzelmesswerte sind oder Messwerte, die das lokale Bewegungsverhalten eines Einzelobjektes beschreiben und damit nicht repräsentativ für flächenhafte Bodenbewegungen sind.

Ergänzend, aber auch einschränkend, soll darauf hingewiesen werden, dass Beobachtungen, die im Zuge dieser Datenanalyse nicht als Anomalien detektiert wurden, trotzdem sich im Sinne der Aufgabenstellung als „Ausreißer“ oder sogar als fehlerhaft erweisen können [Xi 2015]. Eine Ursache für die mit diesem Ansatz nicht detektierbare Anomalien besteht z.B. im flächenhaften Auftreten systematischer Prozessierungsfehler.

3.2. Grund für die Detektion von Anomalien

Wie in Kapitel 3.1 erwähnt, kann eine Anomalie entweder als grober Messfehler oder als Ausdruck eines realen aber ungewöhnlichen Ereignisses interpretiert werden. Falls es sich um

eine fehlerhafte Messung handelt, wird eine solche Beobachtung einen Einfluss auf die Qualität der Analyseergebnisse ausüben, indem die fehlerhafte Anomalie die Parameter für die Modellierung des gesamten Datensatzes verzerrt abbildet. Jedoch sollte (in alter geodätischer Tradition) ein als Anomalie detektierter Wert nur dann eliminiert werden, wenn er als grober Messfehler interpretiert bzw. erklärt werden kann [Xi 2015].

Andererseits kann eine Anomalie auch nützliche Informationen beinhalten, wenn sie reale Datenmerkmale repräsentiert. Durch solche Anomalien lassen sich lokale Inhomogenitäten entdecken, die z.B. bei der Untersuchung und Detektion sich abzeichnender Tagesbrüche oder Erdfälle als Indikatoren von großer Bedeutung sein können.

Treten bei PSI-Zeitreihen Anomalien, wie Phasensprünge, verrauschte Messwerte oder Messfehler auf, werden bei der Ableitung der Höhenänderungsraten oder der Trendmodellierung der gesamten Zeitreihe falsche bzw. verfälschte Ergebnisse erzeugt, welche zu fehlerhaften Interpretationen des Bodenbewegungsverhaltens führen können. Daher müssen derartige Anomalien zuerst erkannt und aus dem PSI-Datensatz entfernt werden.

Falls es zu Anomalien infolge realer, unerwarteter Ereignisse, wie z.B. Baumaßnahmen an oder vor einem Haus, die zu Veränderungen der reflektierten Eigenschaften der sich darauf befindlichen PS-Punkte sowie der daraus abgeleiteten Bodenbewegungen führen können, besteht der Bedarf, diese zunächst zu identifizieren. Da solche Situationen zu einer Veränderung im PSI-Zeitreihenverlauf führen können, repräsentieren sie das flächenhafte Bodenbewegungsverhalten nicht mehr oder nur „eingeschränkt“. Daher wäre die Streichung solcher PS-Punkte empfehlenswert.

Besitzt ein reflektierendes Einzelobjekt ein eigenes Setzungs- oder Sackungsverhalten, dann repräsentieren die Zeitreihen von PS-Punkten an einem solchen Objekt nicht die großräumigen Bodenbewegungen, da sie von dem Objekteigenverhalten überlagert sind. Auch in diesen (nachgewiesenen) Fällen ist eine Streichung solcher PS-Punkte empfehlenswert. Diese Beispiele zeigen auch, dass detektierte Anomalien nicht generell als mangelhafte Beobachtungen beurteilt werden dürfen, sondern sie vielmehr auch Hinweise auf interessante lokale Phänomene enthalten können, welche für anders ausgerichtete Analysen (z.B. Ursache-Wirkung-Analyse) relevant sind [Xi 2015].

Eine Entscheidung, ob die identifizierten Anomalien zu groben Messfehlern oder realen unerwarteten Phänomenen gehören, wird meist nur durch Nutzung von weiterem anwendungsspezifischem Fachwissen möglich sein.

3.3. Klassifikation von Anomalien

Anhand verschiedener Kriterien können Anomalien in verschiedene Klassen eingeteilt werden. Beispielsweise in Bezug auf die Dimensionalität können sie im eindimensionalen Fall als univariate Anomalien (extreme Werte in einer einzelnen Variablen) und im mehrdimensionalen Fall als multivariate Anomalien (auffällige Wertekombinationen mehrerer Variablen) auftreten. Wenn sie über räumliche bzw. zeitliche Merkmale verfügen, wie es bei den PSI-Daten der Fall ist, können sie wiederum in räumliche Anomalien, zeitliche Anomalien und raumzeitliche Anomalien untergliedert werden.

3.3.1. Räumliche Anomalien

Es gilt eine grundlegende Regel für die räumlichen Daten, dass räumlich nahe beieinander liegende Werte ähnlicher als weiter voneinander entfernte Werte sind. Darauf basierend ist eine räumliche Anomalie ein georeferenziertes Objekt, dessen nicht-räumliche Attributwerte sich deutlich von denen anderer georeferenzierter Objekte in ihrer räumlichen Nachbarschaft unterscheiden [SHEKHAR 2003A]. Abbildung 15 zeigt beispielhaft eine räumliche Anomalie in vereinfachter Darstellung. Hierbei spielt die räumliche Abhängigkeit eine wichtige Rolle bei der Identifizierung von räumlichen Anomalien [SHEKHAR 2003B].

Für die räumlichen Daten sind zwei Arten von Attributen zu unterscheiden. Sie enthalten nicht nur die räumlichen Attribute, wie beispielsweise die Koordinaten, welche die räumliche Lage der Objekte beschreiben, sondern auch die nicht räumlichen Attribute, die auf der Lokation, wo das Objekt sich befindet, gemessen sind. Typische nicht räumliche Attribute sind z.B. Temperatur, Niederschlagsmenge, Schwermetallgehalt sowie die aus PSI-Daten abgeleiteten Höhenänderungen. Die abrupten Änderungen solcher Attribute im Vergleich zu den Attributwerten naheliegender Objekte, infolgedessen die räumlichen Anomalien zu identifizieren sind, bieten normalerweise nützliche Informationen in den entsprechenden Untersuchungsgebieten, wenn sie nicht durch Messfehler verursacht wurden.

3.3.2. Temporale Anomalien

Wie bereits in Kapitel 2.3 erwähnt, ist eine Zeitreihe eine Folge von zeitlich geordneten Beobachtungen (z.B. Höhenänderungen von PS-Punkten), die in gleichen Zeitabständen (z.B. 11 Tage bei TerraSAR-X) vorliegen. Im Kontext von Zeitreihen werden diejenigen Beobachtungen, die nicht dem der Zeitreihe zugrundeliegenden statistischen Prozess

entsprechen, als temporale Anomalien identifiziert (s. Abb. 16). Sie können entweder ein zufälliges (spezielles) seltenes Phänomen aufdecken, oder durch einen Messfehler verursacht worden sein [XI 2015]. Bei Vorhandensein einer Anomalie in der Zeitreihe könnte die Schätzung der Prozessparameter und ferner die zugrundeliegende Charakteristik der Zeitreihe verfälscht werden. Wie Abbildung 16 zeigt, weicht durch den Einfluss einer temporären Anomalie der Regressionsgrad bei einer linearen Regressionsanalyse deutlich nach oben ab.

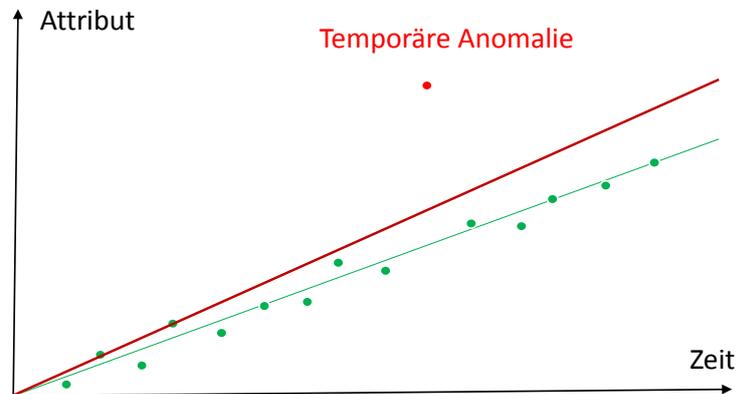


Abbildung 16: Vereinfachte Darstellung einer temporären Anomalie und ihre Auswirkung

Die Ähnlichkeit zwischen der Erkennung von räumlicher und temporaler Anomalie besteht darin, dass in beiden Fällen die Abhängigkeit (Autokorrelation) der Beobachtungen in Bezug auf ein oder mehrere Attribute mitberücksichtigt wird. Jedoch ist eine Zeitachse eindimensional und unidirektional, auf der nur die Beobachtungen in der Vergangenheit bekannt sind, während die räumlichen Beobachtungen mehrdimensional und in den verschiedenen Richtungen aller Achsen bekannt sind.

3.3.3. Spatio-Temporale Anomalien

Bevor wir die spatio-temporalen Anomalien definieren, sollten zuerst die Typen von spatio-temporalen Daten erläutert werden. Sie sind die raumzeitlichen Daten, die neben den beobachteten Attributen zusätzlich einen Raum- und Zeitbezug besitzen. Im Vergleich zu den herkömmlichen Daten werden bei den spatio-temporalen Daten unterschiedliche Attribute an einer bestimmten räumlichen Lage zu unterschiedlichen Zeitpunkten erhoben. Abhängig von der Art und Weise, wie die zeitliche Dimension mit der räumlichen Dimension kombiniert wird, können die spatio-temporalen Daten in folgende Typen eingeteilt werden [MAIMON 2005] [HADI 2012] [KISILEVICH 2010].

- Spatio-temporale Ereignisse: der einfachste Typ der spatio-temporalen Daten ist das spatio-temporale Ereignis, wie z. B. ein Erdbeben oder ein Epidemie-Ausbruch. Tritt das zu beobachtende Ereignis auf, wird das Ereignis einer räumlichen Lage, an der es aufgenommen wurde, und einem entsprechenden Aufnahmezeitpunkt zugeordnet. Außer den räumlichen und zeitlichen Informationen, die mit den beobachteten Ereignissen verbunden sind, werden keine zusätzlichen Attribute (Eigenschaftswerte) erfasst [MAIMON 2010] [HADI 2012]. In Tabelle 1 sind beispielhaft insgesamt 5 solcher spatio-temporalen Ereignisse aufgelistet. Jedes Ereignis besitzt ein (X, Y) Koordinaten-Paar und einen Aufnahmezeitpunkt. So wurde beispielsweise für das Ereignis 3 zum Zeitpunkt 2 das Koordinaten-Paar (X3, Y3) aufgenommen. Zwischen den Ereignissen gibt es keine räumlichen oder zeitlichen Zusammenhänge.

Tabelle 1: Spatio-temporale Ereignisse

ID	X	Y	Zeitpunkt
1	X1	Y1	1
2	X2	Y2	1
3	X3	Y3	2
4	X4	Y4	2
5	X5	Y5	3

- Georeferenzierte Variablen: Unter einer georeferenzierten Variable versteht man ein Attribut eines Objektes, dessen zeitliche Entwicklung einer festen räumlichen Lage zugeordnet ist und ständig beobachtet wurde; es wird jedoch nur der aktuelle (letzte) Wert der Zeitreihe des Attributes erfasst. Im Vergleich zum spatio-temporalen Ereignis wird außer den raumzeitlichen Informationen auch ein Attributwert mit dokumentiert [MAIMON 2010] [HADI 2012]. Wie in Abbildung 17 dargestellt, wurde für jedes beobachtete Objekt jeweils neben der Lage- und Zeitinformaton nur der aktuellste Eigenschaftswert aufgezeichnet, so z.B. für das Objekt 1 nur der aktuelle Eigenschaftswert $A_{1,3}$ zum Zeitpunkt 3. Die grau eingefärbten Eigenschaftswerte der Zeitpunkte 1 und 2 sind nicht gespeichert worden.
- Georeferenzierte Zeitreihen: Wenn die gesamte zeitliche Entwicklung von den georeferenzierten Variablen erfasst werden kann, ergibt sich eine sogenannte georeferenzierte Zeitreihe [MAIMON 2010][HADI 2012][GIANNOTTI 2008]. Wie in Abbildung 18 dargestellt, könnte jedes zu beobachtende Objekt eine Wetterstation sein, an der der Eigenschaftswert, z.B. Regenmenge, zu unterschiedlichen Zeitpunkten und in gleichen Zeitabständen aufgenommen wurde, wie z.B. für die Wetterstation 1 die Regenmengen $A_{1,1}$, $A_{1,2}$ und $A_{1,3}$ zu den Zeitpunkten 1, 2 und 3.

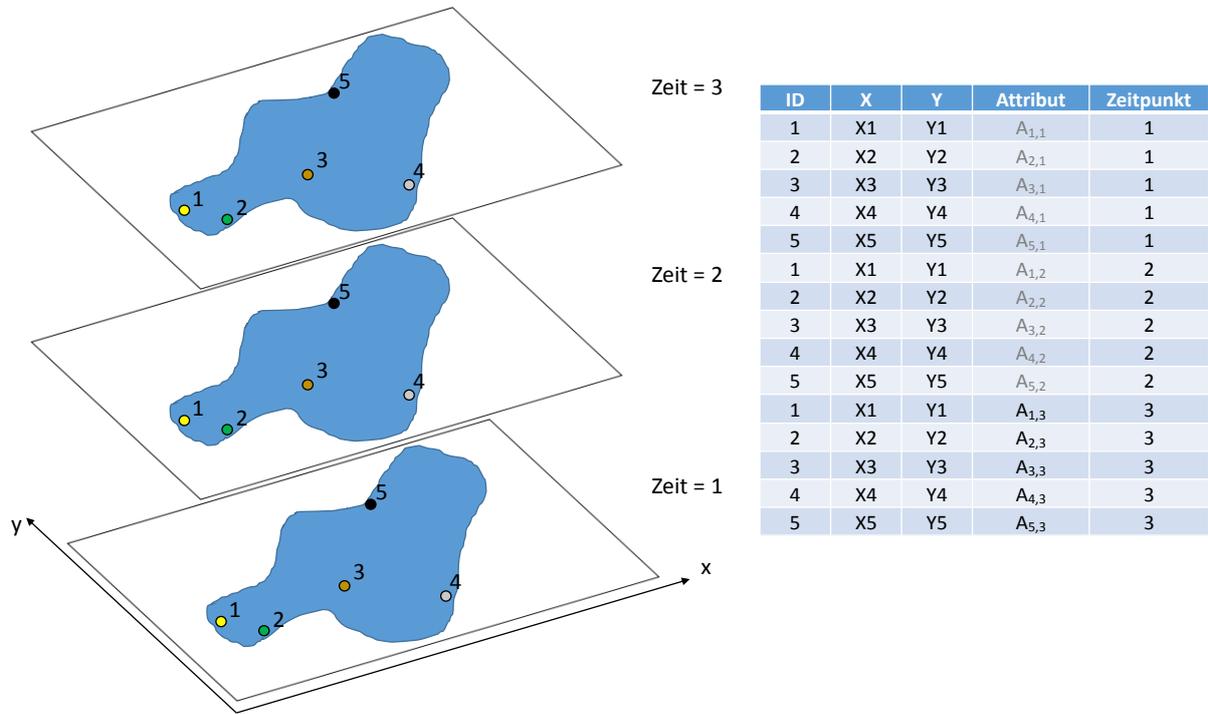


Abbildung 17: Georeferenzierte Variablen [VERÄNDERT NACH HADI 2012]

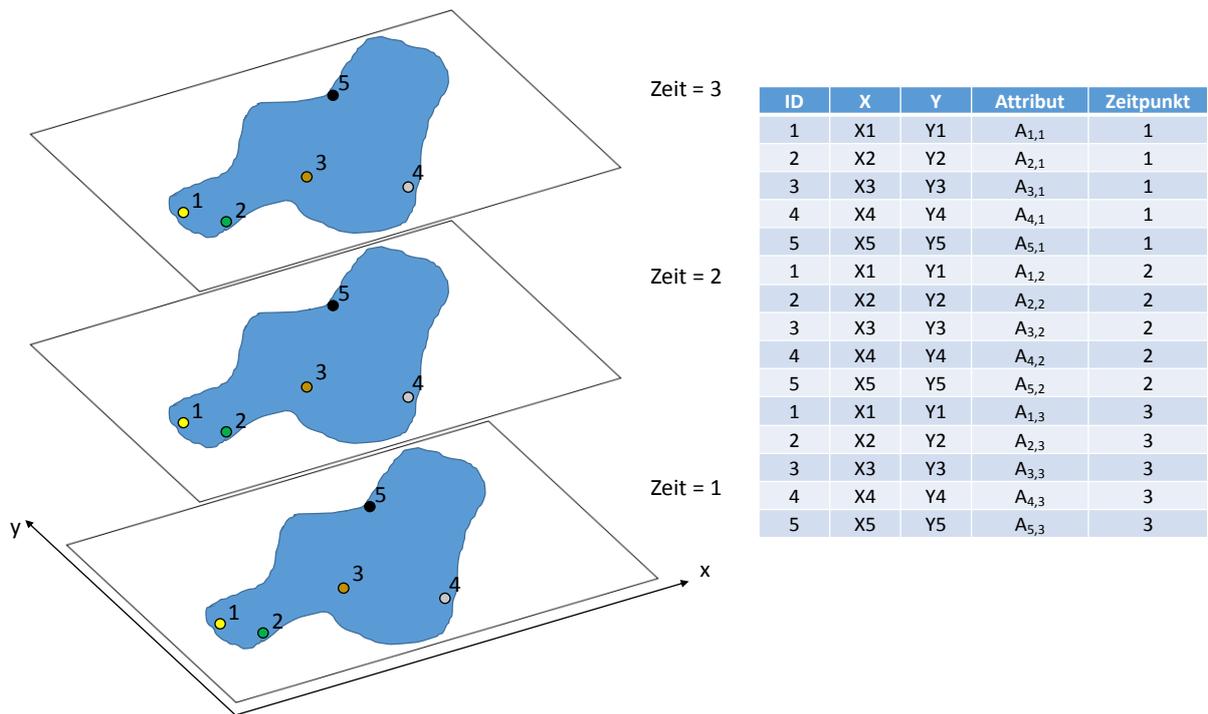


Abbildung 18: Georeferenzierte Zeitreihen [VERÄNDERT NACH HADI 2012]

- Bewegte Punkte: Wenn sich die räumliche Lage eines beobachteten Objekts zeitlich ändert und als Information nur die aktuellste Position des Objekts aufgezeichnet wurde, handelt es sich um einen bewegten Punkt. Außer der räumlichen Information (und dem Zeitpunkt) wurde kein Attribut erfasst [MAIMON 2010][HADI 2012]. Wie in Abbildung 19

zu erkennen, ist z.B. im Rahmen der Echtzeitüberwachung von Fahrzeugen für Sicherheitsanwendungen nur die aktuelle Position ($X_{1,3}$, $Y_{1,3}$) des Fahrzeuges 1 registriert worden.

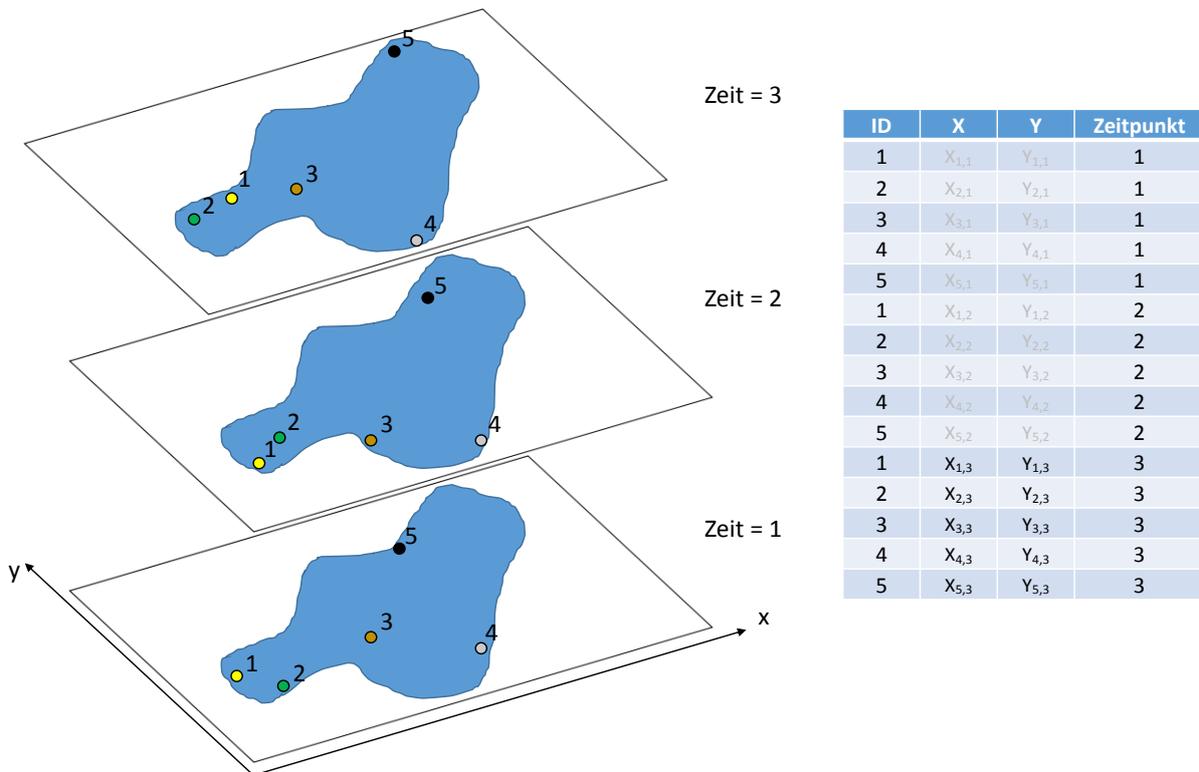


Abbildung 19: Bewegte Punkte [VERÄNDERT NACH HADI 2012]

- Trajektorien: Wenn die gesamte zeitliche Bewegungshistorie eines Punktes gespeichert wird und für die weiteren Analysen zur Verfügung steht, bildet sich aus einer zeitlichen Abfolge der veränderten räumlichen Orte des Objekts zusammen mit den entsprechenden Zeitpunkten eine sogenannte Trajektorie, die zur Beschreibung des Bewegungsverhaltens eines Objektes dient [MAIMON 2010] [HADI 2012]. So zeigt Abbildung 20 die komplette Bewegungshistorie des sich bewegenden Punkt 1.

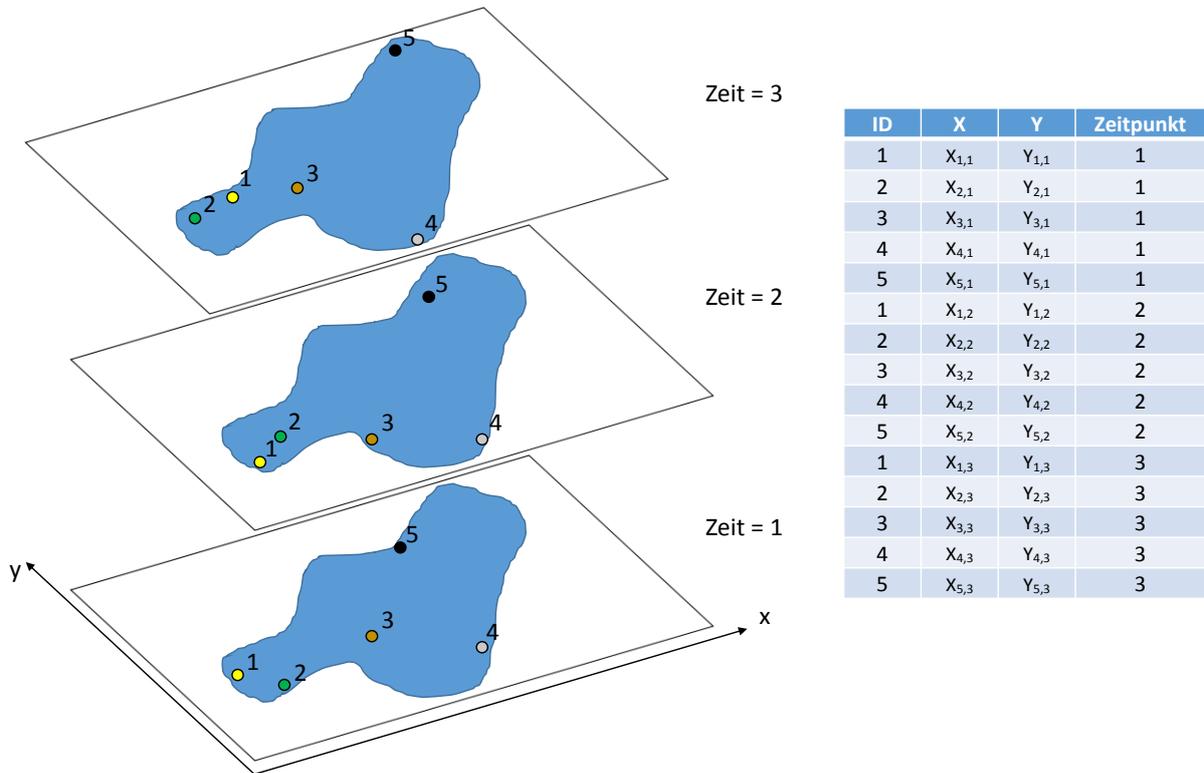


Abbildung 20: Trajektorien [VERÄNDERT NACH HADI 2012]

Die beschriebenen Typen spatio-temporalen Daten sind anhand ihrer raumzeitlichen Charakteristika in Abbildung 21 zusammengefasst dargestellt.

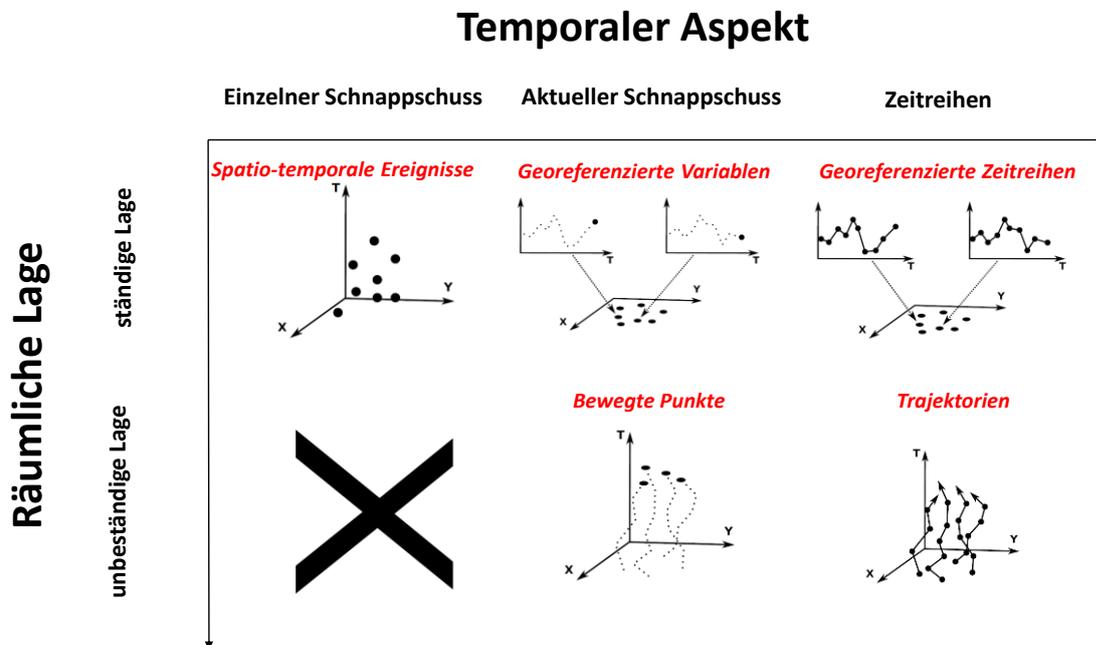


Abbildung 21: Typen spatio-temporalen Daten [VERÄNDERT NACH MAIMON 2010]

Wie in Kapitel 3.2 erwähnt, besitzt jeder PS-Punkt außer seiner räumlichen unveränderten Lage noch die Höhenänderungszeitreihe zur Beschreibung seines Bewegungsverhaltens nach der Zeit. Anhand der Datenstruktur können die PSI-Daten als spatio-temporale Daten definiert und dem Typ von georeferenzierten Zeitreihen zugeordnet werden.

Unter dieser Annahme ist eine spatio-temporale Anomalie (STO) für den Typ von georeferenzierten Zeitreihen ein raumzeitliches georeferenziertes Objekt, dessen Attribut deutlich von denen anderer raumzeitlicher georeferenzierter Objekte in ihrer räumlichen und/oder zeitlichen Nachbarschaft abweicht. Durch die Identifizierung von STOs können unerwartete implizite Erkenntnisse über den Datensatz sowie eine lokale Instabilität von dem Datensatz entdeckt werden.

3.4. Anomalien im raumzeitlichen Verhalten benachbarter PS-Punkte und ihre Detektion

Obwohl das PSI-Verfahren nicht von dem Problem der Dekorrelation betroffen ist, können Anomalien in den Messwerten in Abhängigkeit von der Qualität und der Anzahl der angewendeten Radardaten, der Oberflächennutzung des Untersuchungsgebietes, der Geschwindigkeit und Richtung der aufgetretenen Bodenbewegungen sowie dem ausgewählten Auswerteverfahren nicht ausgeschlossen werden. Während „Fehler“ als systematische, grobe und zufällige Abweichungen des Messergebnisses vom (unbekannten) wahren Wert definiert werden können, lassen sich „Anomalien“ im statistischen Sinn als „Extremwerte“ verstehen, die „weit außerhalb“ der anderen Messwerte in den räumlichen und/oder zeitlichen Nachbarschaften im Datensatz liegen. Solche statistisch definierten Anomalien lassen sich entweder als Messfehler oder als Ausdruck eines realen (aber ungewöhnlichen) Ereignisses interpretieren [Xi 2015]; anders ausgedrückt, sie gehören nicht zur Grundgesamtheit der anderen Daten.

Ursachen für solche Anomalien können z.B. ein ungenaues digitales Geländemodell zur Abspaltung des topographischen Phasenanteils, atmosphärische Einflüsse sowie schnelle nichtlineare Bewegungsverläufe (die zu einem Phasenabwicklungsfehler führen können) sein. Hinzu kommen zufällige Abweichungen (Rauschen), die aufgrund der inneren Messgenauigkeit (u.a. abhängig vom verwendeten Radarsensor) unvermeidbar sind sowie die kurzfristigen anthropogen verursachten Bodenbewegungen, die als unerwartete Ereignisse kenngezeichnet sind [BUSCH 2016] [Xi 2015].

3.4.1. Anomalien durch Phasensprünge

Unter dem Begriff Phasensprung versteht man eine abrupte Änderung der Messwerte in einer PSI-Zeitreihe, welche einen großen Unterschied zwischen zwei benachbarten Messwerten vor und nach dem Sprungdatum zeigt [BERTI 2013][BOROWSKI 2013]. Die sprunghafte Änderung, die einem Phasensprung zuzuordnen ist, ist deutlich größer als die durchschnittliche Schwankung zwischen zwei benachbarten Messwerten in der PSI-Zeitreihe. Infolgedessen kann die PSI-Zeitreihe in zwei oder mehrere schwankungsstabile Teilbereiche zerlegt werden (s. Abb. 22) [YIN 2015]. Die Phasensprünge können somit als temporale Anomalien in den PSI-Zeitreihen verstanden werden.

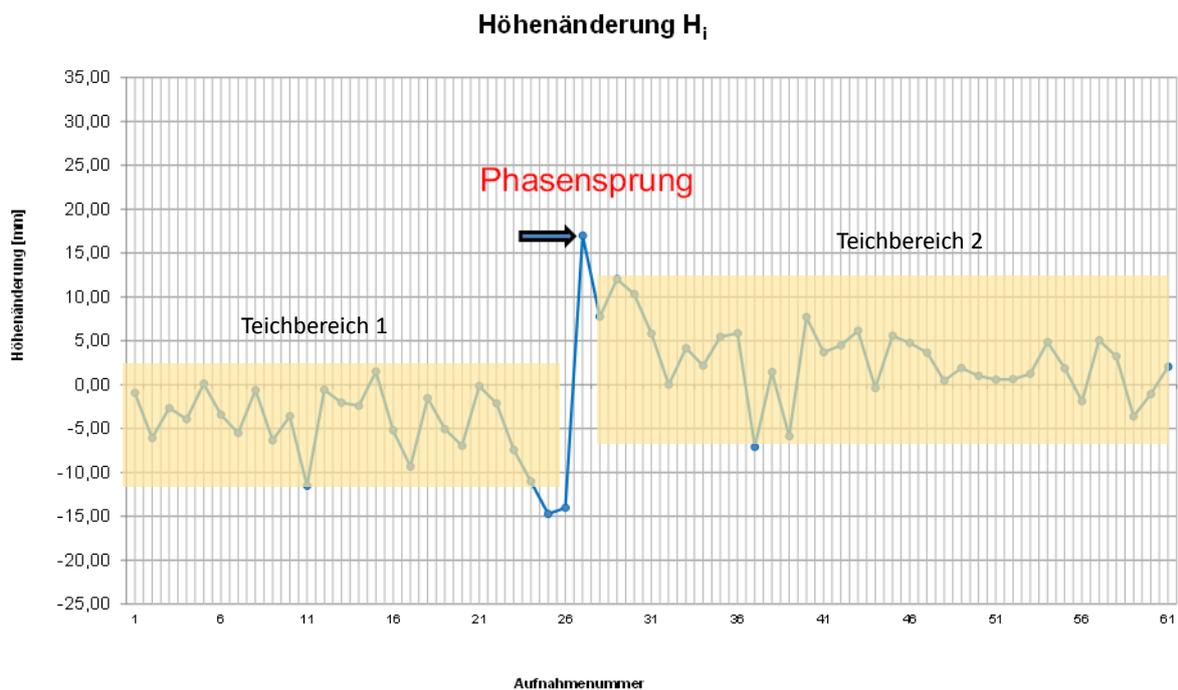


Abbildung 22: Phasensprung in einer PSI-Zeitreihe [VERÄNDERT NACH YIN 2015]

Eine typische Ursache von Phasensprüngen ist der Phasenabwicklungsfehler, der normalerweise wegen der Phasenmehrdeutigkeit bei der PSI-Auswertung zu falschen Ergebnissen führt. Tritt z.B. eine abrupte oder schnelle Bodenbewegung innerhalb kurzer Zeit oder Phasenrauschen infolge der Signalverluste auf, werden die Phasendiskontinuitäten ausgelöst, welche die Phasenmehrdeutigkeit hervorrufen und somit die weitere Phasenabwicklung beeinflussen. Außerdem spielt der angewendete Algorithmus zur Phasenabwicklung sowie dessen Eingangsparameter eine Rolle. Ebenfalls können durch Modellfehler bei der PSI-Regressionsanalyse auch Phasenabwicklungsfehler entstehen [YIN 2015]. Unabhängig von den benachbarten PS-Punkten, sind die PS-Punkte mit solchen Phasensprüngen als temporäre Anomalien zu bezeichnen.

Am IGMC der TU Clausthal wurde ein Ansatz zur automatischen Erkennung von Phasensprüngen in PSI-Zeitreihen entwickelt und in einigen Projekten zur Ableitung von bergbaubedingten Bodenbewegungen praktisch umgesetzt. Dabei werden die Messwertänderungen zwischen zeitlich benachbarten Aufnahmezeitpunkten einer Zeitreihe sukzessive einem vom Benutzer vordefinierten Schwellenwert für die Höhenänderungsdifferenz gegenübergestellt. Wird der vorgegebene Schwellenwert überschritten, erhält dieses Datum eine vorläufige Sprungdatum-Marke. Im Anschluss wird geprüft, ob vor und nach dem vorläufigen Sprungdatum nur ein Phasensprung vorkommt, ist dies der Fall, kann das vorläufige Sprungdatum als endgültig bezeichnet und die betrachtete Zeitreihe als „sprunghaft“ beurteilt werden [YIN 2015].

Der Benutzer hat die Möglichkeit, die Anzahl der zeitlich aufeinanderfolgenden Zeitpunkte für die Ermittlung von Höhenänderungsdifferenzen sowie den Schwellenwert für die Abgrenzung von Phasensprüngen selbst einzustellen, wobei der Schwellenwert nicht kleiner als die Standardabweichung der Höhenänderungsdifferenzen darf (s. Abb. 23) [YIN 2015].

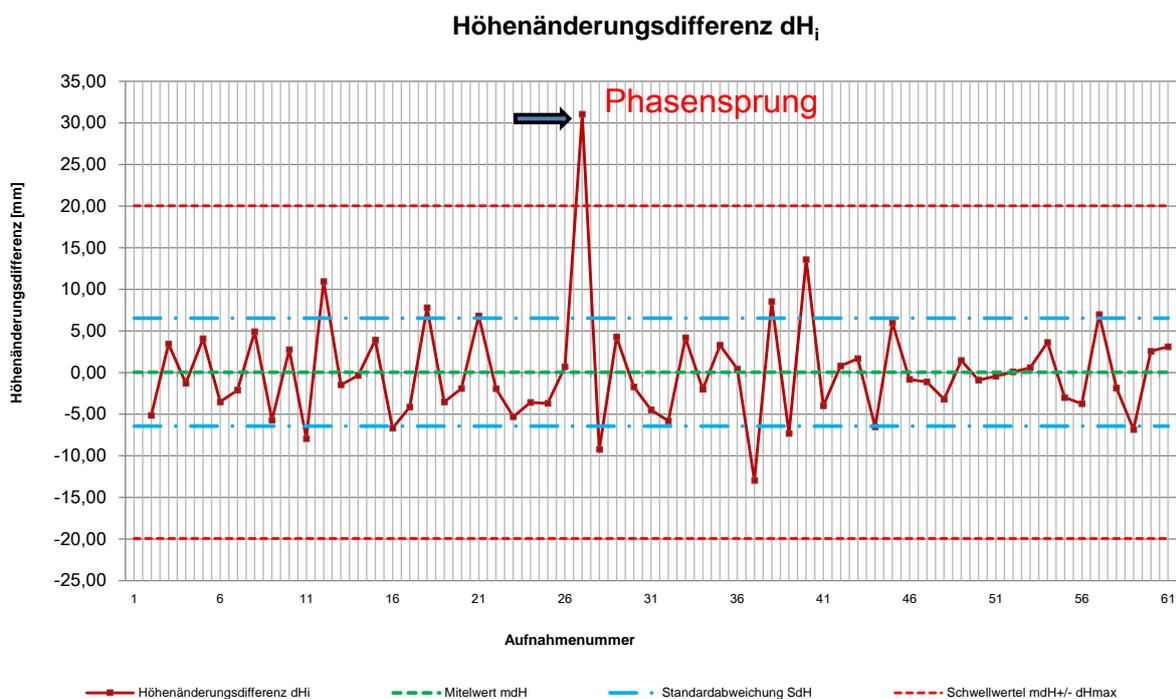


Abbildung 23: Detektion eines Phasensprunges anhand eines vordefinierten Schwellenwertes
[VERÄNDERT NACH YIN 2015]

Die Vorgehensweise zur Detektion von sprunghaften Höhenänderungen wird in Abbildung 24 verdeutlicht.

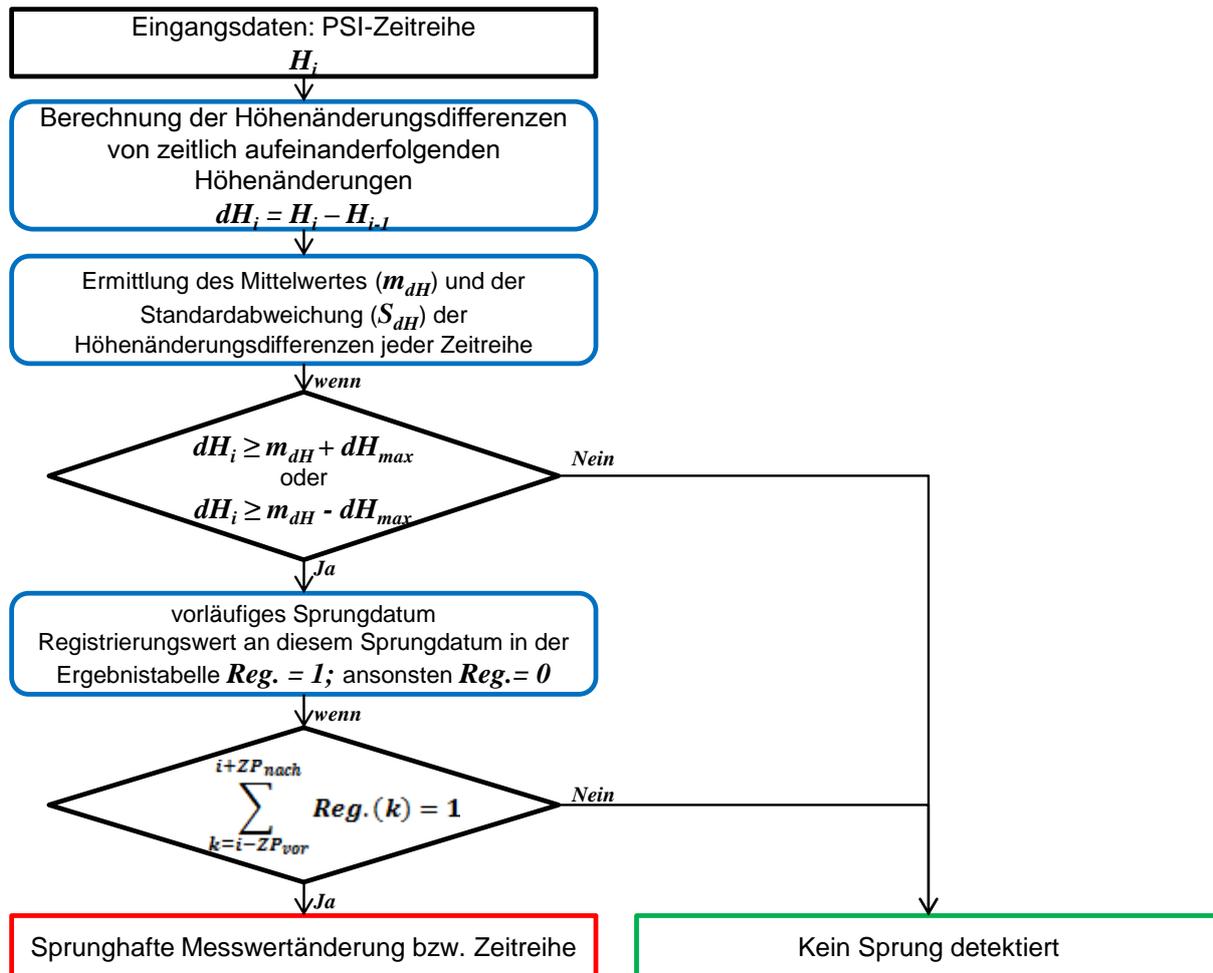


Abbildung 24: Flussdiagramm zur Detektion von sprunghaften Höhenänderungen [YIN 2015]

3.4.2. Anomalien infolge verrauschter Signale

Das von dem Radarsensor ausgesandte und empfangene Signal ist mehr oder weniger verrauscht. Das Rauschen ist ein statistisch zufälliges Signal, aus dem sich auch die innere Messgenauigkeit ableiten lässt. Es ist dem Nutzsignal überlagert und wird meist als Störeffekt angesehen, insbesondere wenn es schwachen Radarsignalen überlagert ist, so dass es aufgrund der instabilen Rückstreuung die Messergebnisse enorm verschlechtert [BUSCH 2016]. Ein stark schwankendes Rauschsignal führt nach der PSI-Auswertung zu einer deutlich größeren Schwankung der Messwerte in der PSI-Zeitreihe (s. Abb. 25), deren Standardabweichung größer als die durchschnittliche Standardabweichung aller PS-Punkte in dem Untersuchungsgebiet ist [BUSCH 2016]. Statistisch gesehen können sie deshalb auch als Anomalien verstanden werden. Da es kaum möglich sein dürfte, das unbekannte Nutzsignal aus einer stark verrauschten PSI-Zeitreihe abzuleiten, sind solche mit Messrauschen verunreinigten Zeitreihen und damit die entsprechenden PS-Punkte für die weiteren Analysen meist unbrauchbar.

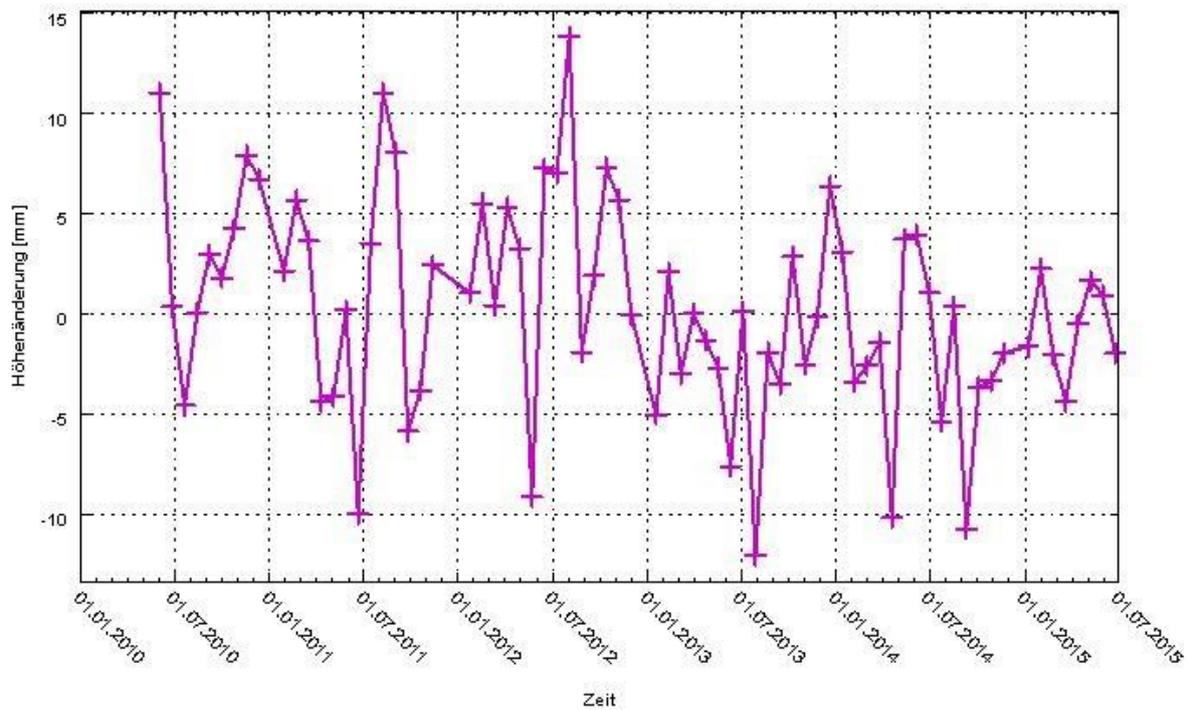


Abbildung 25: Stark verrauschte PSI-Zeitreihe

Die PS-Punkte mit einem stark verrauschten Höhenänderungsverhalten liefern keine vertrauenswürdige Höhenänderungsinformation. Daher wird die statistische Analyse einer Höhenänderung nach einer Polynomausgleichung zur Beurteilung der Qualität der Zeitreihen und darauf aufbauend zur Identifizierung stark verrauschter PS-Punkte als temporäre Anomalien benutzt [BUSCH 2014].

Die aus den erhaltenen Residuen berechneten Standardabweichungen einer Höhenänderung (für jede Zeitreihe eines PS-Punktes) werden in einem Histogramm dargestellt. Eine kleine Standardabweichung weist auf einen niedrigen Streuungsgrad der Höhenänderungen in der PSI-Zeitreihe hin. Extremwerte dieser Standardabweichungen, die ca. 1% aller Daten entsprechen und im oberen Wertebereich dieser Häufigkeitsverteilung liegen, werden anschließend als Kandidaten für eine Elimination im Sinne temporärer Anomalien identifiziert [BUSCH 2016]. Untersuchungen dieser detektierten PS-Punkte mit ihren Zeitreihen haben in verschiedenen Projekten ergeben, dass das stark schwankende zeitliche Höhenänderungsverhalten dieser PS-Punkte z.B. durch bauliche Veränderungen an den Objekten oder in ihrem nahen Umfeld verursacht sind. Die Ergebnisse sind somit nicht repräsentativ (genug) zur Beschreibung eines flächenhaften Bodenbewegungsverhaltens und eine Elimination dieser PS-Punkte kann unter diesem Gesichtspunkt erfolgen [XI 2015].

3.4.3. Anomalien durch kurzfristige anthropogene Veränderungen oder Messfehler

Das Ziel der Anwendung von PSI-Verfahren ist die flächenhafte Überwachung von großräumigen Bodenbewegungen, wie sie z.B. durch Bergbauaktivitäten und Grundwasserentnahme im Untergrund verursacht werden. Dabei werden die sog. persistenten Streuer (z.B. Hausdächer, Strommasten, Industrieanlagen) mit langzeitstabilen Rückstreueigenschaften zur Ermittlung von Bodenbewegung der Erdoberfläche eingesetzt. Durch eine kurzfristige kleinräumige anthropogene Veränderung solcher persistenten Streuer, z.B. durch die Renovierung eines Hausdaches, kann die Rückstreueigenschaft der betroffenen PS-Punkte beeinträchtigt werden, welche zu großen Messwertänderungen in Zeitreihen oder sogar zu einem umkehrenden Bewegungstrend führen können. Außerdem könnten auch z.B. systematische Fehler durch atmosphärische und topographische Einflüsse, zufällige Fehler infolge Messrauschen sowie grobe Messfehler durch Modellfehler ein abweichendes Höhenänderungsverhalten der PS-Punkte verursachen. Somit sind solche PS-Punkte nicht repräsentativ für die Bestimmung von großräumigen Bodenbewegungen und stellen dadurch, im Vergleich zu anderen benachbarten PS-Punkten, die nicht von den kurzfristigen menschlichen Eingriffen betroffen sind, lokale Anomalien dar [Xi 2015].

Für solche Anomalien können angesichts ihres Höhenänderungsverhalten in einer Zeitreihe keine Auffälligkeiten festgestellt werden, jedoch zeigen solche PS-Punkte im Vergleich zu benachbarten Objekten unerwartete oder sogar widersprüchliche Höhenänderungsverläufe, welche sie als räumliche oder spatio-temporale Anomalien einstufen lässt. Die Identifizierung derartiger Anomalien erfordert meistens eine Gegenüberstellung des Höhenänderungsverhaltens im nachbarschaftlich-räumlichen Kontext.

Beispielsweise zeigen alle PS-Punkte in Abbildung 26 in ihren Höhenänderungszeitreihen unauffällige Verläufe, die nicht von Phasensprüngen oder starkem Rauschen beeinträchtigt sind. Somit können sie hinsichtlich ihrer Höhenänderungszeitreihe nicht als temporäre Anomalien erkannt werden. Wenn diese PS-Punkte jedoch in ihrer räumlichen Nachbarschaft betrachtet werden, weist im Gegensatz zu den benachbarten Punkten der PS-Punkt 25055 einen deutlichen Senkungstrend auf. In diesem Sinne kann der Punkt 25055 als eine „räumliche Anomalie“ detektiert werden.

Des Weiteren ist in Abbildung 26 ersichtlich, dass der Höhenänderungswert des PS-Punktes 25055 am 18.02.2002 etwa 35 mm beträgt (markiert als „roter Stern“ in Abbildung 26) und sich deutlich von den Höhenänderungswerten (markiert mit schwarzen Rechtecken) in seiner raumzeitlichen Nachbarschaft (blauer rechteckiger Bereich in Abbildung 26) mit einem durchschnittlichem Höhenänderungswert von ca. 10 mm abweicht. Somit kann auch dieser Höhenänderungswert als „spatio-temporale Anomalie“ identifiziert werden.

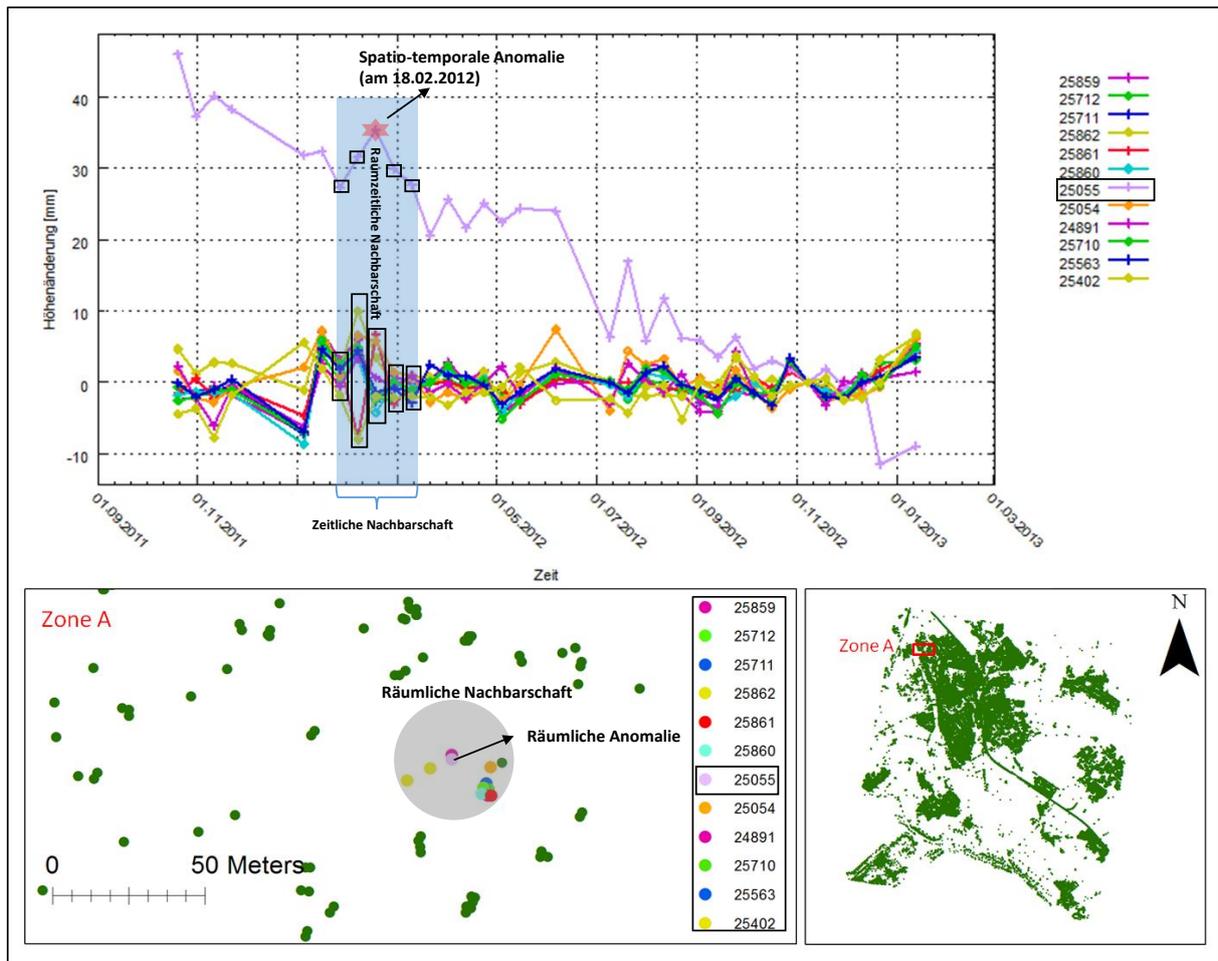


Abbildung 26: Abweichendes Höhenänderungsverhalten einer Anomalie gegenüber umgebenden PS-Punkten [VERÄNDERT NACH XI 2015]

Die Identifizierung von solchen räumlichen bzw. spatio-temporalen Anomalien ist nur unter Berücksichtigung von PS-Punkten in ihren räumlichen bzw. raumzeitlichen Nachbarschaften möglich. Zu ihrer Detektion erscheinen räumliche Clusterverfahren als besonders zielführend [XI 2015].

Clusterverfahren werden in vielen Anwendungsbereichen, z.B. der Biologie, der Physik oder den Wirtschaftswissenschaften, eingesetzt. Ziel dieser Verfahren ist es, die Untersuchungsobjekte im Hinblick auf die Ähnlichkeit der betrachteten Eigenschaften (Merkmale) zu Gruppen zusammen zu fassen, wobei die Objekte innerhalb einer Gruppe möglichst homogen und die Gruppen untereinander möglichst heterogen sind [BACKHAUS 2016].

Bei räumlichen Clusterverfahren spielt außerdem die räumliche Beziehung zwischen den untersuchten Objekten eine große Rolle. Sie basieren auf der Grundannahme einer räumlichen Abhängigkeit, d.h. je näher die Objekte beieinander liegen, desto ähnlicher sind sie. Mit Distanzbeziehungen lassen sich Objekte von Objekten in anderen Clustern abgrenzen.

Die räumlichen Clusterverfahren lassen sich wiederum im Hinblick auf die Wahl der Eigenschaften, mit der die Ähnlichkeit der Objekte bewertet wird, in zwei Kategorien unterscheiden:

- räumliche Clusterverfahren, die lediglich Nachbarschaftsbeziehungen bei der Gruppenbildung berücksichtigen und
- raumzeitliche Clusterverfahren, die außer der Nachbarschaftsbeziehung auch nicht-räumliche Eigenschaften des Objektes in Betracht ziehen.

Grundgedanke bei der Detektion von Anomalien in flächenhaft vorliegenden PSI-Ergebnissen ist, dass beim Auftreten von flächenhaften Bodenbewegungen alle eine deformierte Fläche repräsentierenden Punkte in ihrem räumlichen Umfeld eine ähnliche Höhenänderung (oder allgemein ähnliche Werte eines anderen Merkmals) aufweisen. Einzelne Punkte mit davon stark abweichendem Höhenänderungsverhalten können entweder auf ein lokales objektspezifisches Bewegungsverhalten hinweisen (und sind damit nicht repräsentativ für die zu detektierende flächenhafte Bodenbewegung) oder sind Ausdruck verfahrenstechnisch bedingter fehlerhafter Anomalien. Somit könnten solche detektierten Punkte, als nicht repräsentativ für Bodenbewegungen oder als messtechnisch grob fehlerhaft, mithilfe der Clusteranalyse erkannt und bzw. eliminiert werden [Xi 2015].

Eine Gemeinsamkeit der beiden vorgestellten Verfahren zur Erkennung von Phasensprüngen und stark verrauschten Messwerten ist, dass sie auf einer Zeitreihenanalyse basieren, wobei (nur) die zeitvariablen Höhenänderungen jedes einzelnen PS-Punktes für die Detektion von temporären Anomalien herangezogen werden. Im Gegensatz dazu sollen bei dem Clusterverfahren nicht nur die Eigenschaft der Zeitreihen (z.B. Deformationsrate) sondern auch die räumlichen Beziehungen der PS-Punkte berücksichtigt werden. Wie bereits in Kapitel 3.2 erwähnt, handelt es sich bei dem PSI-Datensatz um eine Datenstruktur von georeferenzierten Zeitreihen, in denen die zeitlichen und räumlichen Eigenschaften miteinander so eng gekoppelt sind, dass sie nicht voneinander getrennt betrachtet werden dürfen. Im Rahmen dieser Arbeit werden die Möglichkeiten verschiedener Clusteranalysen für die Erkennung unterschiedlicher Arten von Anomalien in flächenhaft vorliegenden PSI-Ergebnissen untersucht. Darüber hinaus werden die Anwendbarkeit sowie die Vor- und Nachteile anhand von praktischen Beispielen mit Messdaten veranschaulicht.

4. Clusteranalyse

Die Clusteranalyse findet in vielen Gebieten, beispielsweise in der Archäologie, der Soziologie oder den Wirtschaftswissenschaften eine breite Anwendung. Oft wird man mit der Forderung konfrontiert, ob zwischen den untersuchten Objekten Ähnlichkeiten bestehen. Das Ziel der Clusteranalyse ist es, eine Menge von betrachteten Untersuchungsobjekten, die durch verschiedene Merkmale (Eigenschaftswerte) charakterisiert sind, so zu gruppieren, dass diejenigen Objekte, die sich im Hinblick auf die Ausprägungen der Merkmale am ähnlichsten sind, in Cluster zusammengefasst werden. Dabei sollen innerhalb eines Clusters die Objekte, bezogen auf die betrachteten Merkmale, möglichst homogen sein und gleichzeitig die Cluster untereinander eine möglichst große Heterogenität aufweisen [BACKHAUS 2016] [ECKSTEIN 2016].

Als Beispiel können die PS-Punkte als Untersuchungsobjekte dienen, die als Merkmale (Eigenschaftswerte) die Höhenänderungen, die Höhenänderungsraten, die Polynomgrade usw. besitzen. Ausgehend von diesen Daten wird untersucht, ob sich die PS-Punkte hinsichtlich der erhobenen Merkmale so gruppieren lassen, dass die PS-Punkte innerhalb eines Clusters eine möglichst hohe Ähnlichkeit bezüglich des Bewegungsverhaltens aufweisen und sich unterscheiden von dem Bewegungsverhalten der PS-Punkte in den anderen Clustern.

4.1. Prinzip der Clusteranalyse

Um eine Clusteranalyse durchführen zu können, benötigt man zunächst eine Menge von Untersuchungsobjekten; jedes Untersuchungsobjekt besitzt hierbei einige Merkmale, die für die Zuordnung der Untersuchungsobjekte in verschiedene Cluster charakteristisch sind. Werden z.B. n Untersuchungsobjekte $O = \{O_i, j = 1, 2, \dots, n\}$ durch p Merkmale $X = \{X_i, i = 1, 2, \dots, p\}$ beschrieben, dann können die Eigenschaftswerte $X_i(O_j)$ als x_{ij} umgeschrieben und in einer $(p \times n)$ -Merkmalmatrix $\mathbf{X} = [x_{ij}]$ zusammengefasst werden (s. Formel 1) [NIEMANN 1974].

Formel 1:

$$X_{(p,n)} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{p,1} & \cdots & x_{p,n} \end{bmatrix}$$

Die Merkmale eines beliebig ausgewählten Untersuchungsobjektes O_i können in Vektorform mit Formel 2 beschrieben werden [FLETTLING 2010]:

Formel 2:

$$x(O_i) = \begin{bmatrix} x_{1,i} \\ \vdots \\ x_{p,i} \end{bmatrix}$$

Die Untersuchungsobjekte lassen sich somit geometrisch als Punkte in einem p-dimensionalen Merkmalsraum darstellen, der als p-dimensionaler euklidischer Raum angesehen werden kann [FLETING 2010]. Dadurch kann jedes Untersuchungsobjekt durch die individuelle Lage hinsichtlich seiner Merkmale in dem Merkmalsraum lokalisiert werden (s. Abb. 27). Für die Clusterbildung dient die Berechnung der Distanzen zwischen den Untersuchungsobjekten im Merkmalsraum als grundlegendes Maß zur Bestimmung der Ähnlichkeiten der Untersuchungsobjekte. Eine geringe Distanz deutet darauf hin, dass die betrachteten Untersuchungsobjekte ähnlich (homogen) sind und sich in einem Cluster befinden. Im Gegensatz dazu werden die Untersuchungsobjekte als unähnlich (heterogen) bezeichnet und in unterschiedlichen Clustern gruppiert, wenn sie vergleichsweise weit voneinander entfernt sind. In diesem Sinne werden die Untersuchungsobjekte den verschiedenen Clustern zugeordnet mit dem Ziel, dass innerhalb der Cluster eine Homogenität und zwischen den Clustern eine Heterogenität besteht [ECKSTEIN 2014].

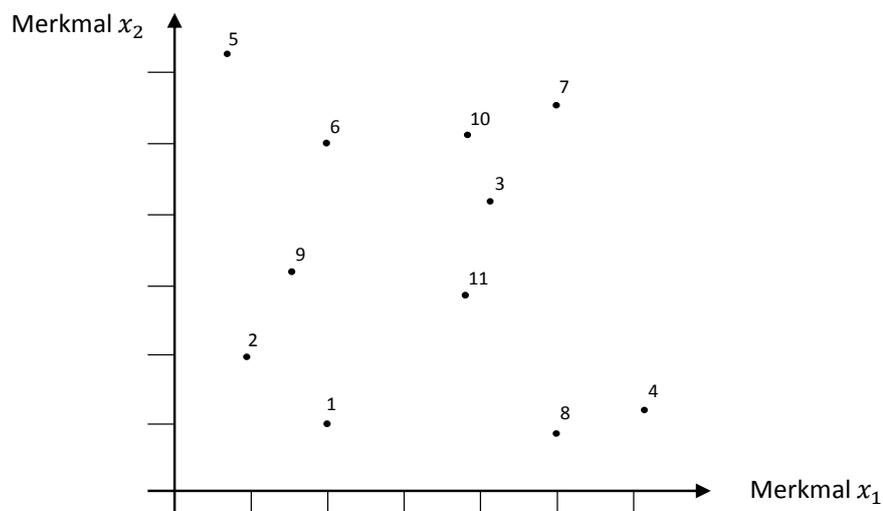


Abbildung 27: Lage der Untersuchungsobjekte in einem zweidimensionalen Merkmalsraum [VERÄNDERT NACH FLETING 2010]

Die Clusterbildung der gesamten Untersuchungsobjekte $O = \{O_j, j = 1, 2, \dots, n\}$ lässt sich mathematisch als die Partition von X in $c \in \{2, 3, \dots, n-1\}$ disjunkte Cluster C_1, \dots, C_c so definieren [RUNKLER 2010], dass gilt:

Formel 3:

$$O = C_1 \cup C_2 \cup \dots \cup C_c$$

Formel 4: $C_i \neq \{ \}$ für alle $i=1, \dots, c$

Formel 5: $C_i \cap C_j = \{ \}$ für alle $i, j=1, \dots, c, i \neq j$

4.2. Ablauf und Arbeitsschritte einer Clusteranalyse

Die wesentlichen Arbeitsschritte zur Durchführung einer Clusteranalyse sind in Abbildung 28 dargestellt [BACKHAUS 2016].

Bevor eine Clusteranalyse ausgeführt wird, sollte zuerst die Problemstellung sowie das Ziel der Untersuchung konkretisiert werden. Beruht es auf einer statistischen Analyse, ist zusätzlich noch zu überlegen, welche Hypothesen getestet werden müssen.

Im Anschluss daran findet die Auswahl und Aufbereitung der Ausgangsdaten statt. Dabei soll das Ausmaß der zu klassifizierenden Objekte bestimmt werden. Wenn die Clusteranalyse nur im Sinne einer Stichprobe durchgeführt wird, muss garantiert werden, dass die ausgewählte Stichprobe die gesamten Untersuchungsobjekte vertreten kann und sich genügend Objekte in den einzelnen Clustern ergeben, um repräsentative Rückschlüsse auf die Gesamtheit der Untersuchungsobjekte ziehen zu können. Dann erfolgt die Auswahl der Merkmale. Es gibt keine eindeutigen Vorschriften für die Anzahl der in einer Clusteranalyse zu berücksichtigenden, qualitativen oder quantitativen Merkmale. Grundsätzlich sollte darauf geachtet werden, dass nur die Merkmale der Untersuchungsobjekte in der Clusteranalyse herangezogen werden, die, ausgehend von theoretischen Überlegungen, für die Untersuchung von Bedeutung sind. Irrelevante Merkmale können zu falschen Clusterergebnissen führen und müssen von der Clusteranalyse ausgenommen werden. Falls die Merkmale auf unterschiedlichen Skalen erhoben sind, ist es sinnvoll, eine Homogenisierung der Merkmalseinheiten vorzunehmen, um die Vergleichbarkeit zwischen den betrachteten Merkmalen sicherzustellen. Nachdem die passenden Merkmale für die Clusteranalyse ausgewählt worden sind, werden die Ähnlichkeiten bzw. Distanzen für jeweils zwei Untersuchungsobjekte bestimmt, indem die Unterschiede bzw. Übereinstimmungen der Ausprägungen der betrachteten Merkmale durch Zahlenwerte berechnet werden. Mit Berücksichtigung der Ähnlichkeits- bzw. Distanzmaße werden die Untersuchungsobjekte mit dem ausgewählten Clusterverfahren so zu Clustern geordnet, dass sich diejenigen Objekte mit weitgehend ähnlich ausgeprägten Merkmalen in einem Cluster zusammenfügen [BACKHAUS 2016]. Bei der Auswahl des Fusionierungsalgorithmus von Clusterverfahren stehen verschiedene Algorithmen zur Verfügung, worauf später noch näher eingegangen wird.

Sind Cluster gebildet, ist anschließend, je nach dem angewendeten Clusteralgorithmus, manchmal noch die optimale Clusteranzahl zu bestimmen. Auch hierfür bestehen verschiedene Methoden, mit denen vor allem die Balance zwischen Handhabbarkeit mit geringer Clusterzahl und Homogenitätsanforderung mit großer Clusterzahl gewahrt werden soll.

Nach der Durchführung des Gruppierungsvorganges erfolgt abschließend die Interpretation der Clusterergebnisse. Im Falle einer nicht sinnvollen Interpretation des Clusterergebnisses kann die Analyse und Interpretation der Ergebnisse durch einen wiederholten Durchlauf einzelner Stufen erneut durchgeführt werden. Außerdem kann ein Vergleich der Auswirkungen einer anderen Clusterungsvariante auf die Clusterergebnisse zu einem wiederholten Durchlauf einzelner Stufen führen [BACKHAUS 2016].

Es sei noch darauf hingewiesen, dass die Qualität der Clusteranalyse von vielen Einflussfaktoren abhängt. Dabei spielt die breite Auswahl der zur Verfügung stehenden Clusterverfahren und der verschiedenen Ähnlichkeitsmaße und Algorithmen eine entscheidende Rolle. Für viele Fragestellungen ergeben sich mehr als zwei Clusterverfahren, die als erfolgversprechend angesehen werden können. Im Hinblick auf die subjektive Beurteilung eines einzelnen Anwenders besteht manchmal die Gefahr, die Eingangsdaten und die Verfahrensparameter so anzupassen, dass die gewünschten Ergebnisse auch eintreffen; dies setzt allerdings voraus, dass der Anwender so viel Spielraum bei der Clusteranalyse besitzt. Deshalb empfiehlt es sich, die Stabilität der erzielten Clusterergebnisse durch die Veränderung des Ähnlichkeitsmaßes und einen Wechsel des Algorithmus zu validieren.

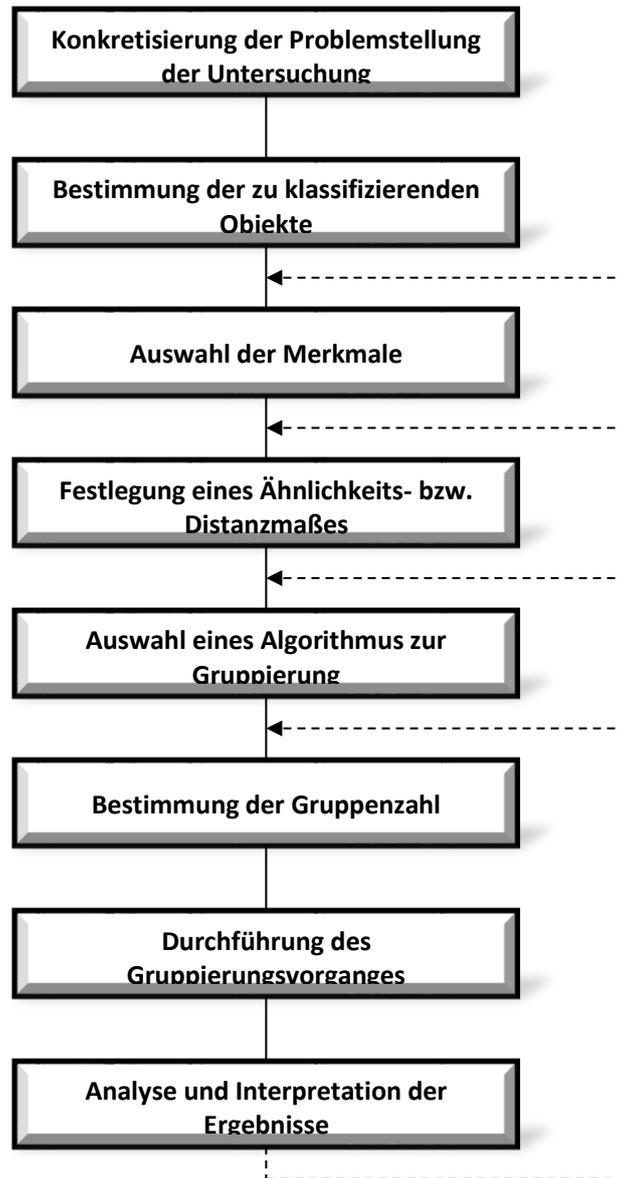


Abbildung 28: Verfahrensschritte einer Clusteranalyse [VERÄNDERT NACH BACKHAUS 2016]

4.3. Festlegung von Ähnlichkeiten bzw. des Distanzmaßes

Basierend auf den Ähnlichkeiten der Untersuchungsobjekte werden verschiedene Cluster zusammengefasst. Die Festlegung von Ähnlichkeiten stellt einen grundlegenden Arbeitsschritt für die Clusteranalyse dar, wobei für jeweils zwei Untersuchungsobjekte die Ausprägungen der ausgewählten Merkmale miteinander verglichen und die Unterschiede bzw. Übereinstimmungen durch einen Zahlenwert berechnet werden. Während die Maße zur Quantifizierung der Ähnlichkeit von Merkmalen in der Clusteranalyse als Ähnlichkeitsmaße bezeichnet werden, nennt man die Maße zur Quantifizierung der Unähnlichkeit von Merkmalen Unähnlichkeits- oder Distanzmaße [HARTUNG 2007]. Ähnlichkeitsmaße geben die Ähnlichkeit

zwischen zwei Objekten wieder; größere Werte eines Ähnlichkeitsmaßes bedeuten, dass zwei Objekte ähnlich sind. Im Gegensatz dazu wird die Unähnlichkeit zwischen zwei Objekten durch die Distanzmaße gemessen; bei größerer Distanz sind zwei Objekte unähnlicher. Wenn sich zwischen zwei Objekten eine Distanz von Null ergibt, können sie als identische Objekte angesehen werden.

Die beiden Maße werden unter dem Oberbegriff der Proximitätsmaße zusammengefasst. Unter Berücksichtigung der Merkmalsstruktur ist die Proximitätsermittlung in vier Kategorien zu unterscheiden [BACKHAUS 2016], mit:

- binärer,
- nominaler,
- metrischer und
- gemischt skaliertes Merkmalsstruktur.

Da die in der vorliegenden Arbeit betrachteten PSI-Daten hauptsächlich über metrische Merkmale verfügen, erfolgt eine Beschränkung auf die Proximitätsermittlung bei metrischer Merkmalsstruktur.

Vor der Bestimmung von Proximitäten wird zuerst eine Rohdatenmatrix mit I Objekten gebildet, die durch J Merkmale charakterisiert werden. Die objektspezifischen Merkmalswerte werden in einer Matrix eingetragen, die als Ausgangspunkt für die Clusteranalyse dient (s. Tabelle 2).

Tabelle 2: Vereinfachte Darstellung des Aufbaus einer Rohdatenmatrix

	Merkmal 1	Merkmal 2	...	Merkmal J
Objekt 1				
Objekt 2				
...				
...				
Objekt I				

Zur Quantifizierung der Proximitäten zwischen den Objekten erfolgt zunächst auf Basis der Rohdatenmatrix die Erstellung einer Ähnlichkeits- oder Distanzmatrix (s. Tabelle 3). Dazu werden unter Verwendung der objektbezogenen Merkmalswerte (Eigenschaftswerte) aus der Rohdatenmatrix die statistischen Maßzahlen in Form von Ähnlichkeits- oder Unähnlichkeitswerten (Distanzwerte) zwischen den betrachteten Objekten berechnet, und diese für die Quantifizierung der Ähnlichkeit oder Distanz zwischen den Objekten in die quadratische Ähnlichkeits- oder Distanzmatrix eingetragen.

Tabelle 3: Vereinfachte Darstellung des Aufbaus einer Ähnlichkeits- oder Distanzmatrix

	Objekt 1	Objekt 2	Objekt I
Objekt 1				
Objekt 2				
...				
...				
Objekt I				

Hinsichtlich des Skalenniveaus der betrachteten Merkmale ergibt sich eine Vielzahl an Proximitätsmaßen. Im Folgenden werden deshalb die gebräuchlichen Maße zur Ermittlung der Ähnlichkeit bzw. Distanz für die Kategorie mit metrischem Skalenniveau näher erörtert.

Die Berechnung von Proximitätsmaßen bei metrischem Skalenniveau der Merkmale von Untersuchungsobjekten wird am Beispiel von PS-Punkten verdeutlicht. Es sind 5 PS-Punkte anhand der Merkmale, Höhenänderungswerte (HÄW) am Datum1, Datum2 und Datum3, auf einer standardisierten Skala bestimmt worden (s. Tabelle 4).

Tabelle 4: Ausgangsdatenmatrix von PS-Punkten mit metrischer Merkmalsstruktur

	Merkmale		
PS-Punkte	HÄW am Datum1	HÄW am Datum2	HÄW am Datum3
Punkt 1	2	4	2
Punkt 2	2	2	1
Punkt 3	3	1	4
Punkt 4	6	4	2
Punkt 5	5	7	5

4.3.1. Minkowski-Metriken

Es gibt eine ganze Reihe von Distanzmaßen für metrische Merkmale. Davon besitzen die weit verbreiteten Minkowski-Metriken die größte Bedeutung, die sich allgemein gemäß Formel 6 berechnen lassen [BOCK 1974][BACKHAUS 2016]:

Formel 6:
$$d_{i,k} = \left(\sum_{j=1}^p |x_{k,j} - x_{i,j}|^r \right)^{1/r}$$

Hieraus lassen sich durch Wahl von r (eine positive ganze Zahl) verschiedene Metriken (Distanzen) ableiten. Für $r = 1$ (Formel 7) ergibt sich die City-Block-Metrik (L1-Norm) und für $r = 2$ die Euklidische Distanz (L2-Norm) (Formel 8 und in Matrizenschreibweise Formel 9). Als

ein weiterer Spezialfall für $r = \infty$ wird in Formel 10 die Distanz als größte der absoluten achsparallelen Distanzen zwischen den Objekten x_i und x_k berechnet.

Formel 7:
$$d_{i,k} = \sum_{j=1}^p |x_{k,j} - x_{i,j}|$$

Formel 8:
$$d_{i,k} = \left(\sum_{j=1}^p (x_{k,j} - x_{i,j})^2 \right)^{1/2}$$

Formel 9:
$$d_{i,k} = \sqrt{(x_k - x_i)^T (x_k - x_i)}$$

Formel 10:
$$d_{i,k} = \max\{|x_{k,j} - x_{i,j}|, 1 \leq j \leq p\}$$

Werden die individuellen Gewichtungen q_j für jedes Merkmal bei der Berechnung von Distanzen zwischen den Objekten mitberücksichtigt, kann der Einfluss von einzelnen Merkmalen verstärkt oder abgeschwächt werden (s. Formel 11) [BACKHAUS 2016].

Formel 11:
$$d_{i,k} = \left(\sum_{j=1}^p q_j |x_{k,j} - x_{i,j}|^r \right)^{1/r}$$

mit:

$r \geq 1$: Minkowski-Konstante

$r = 1$: City-Block Distanz

$r = 2$: Euklidische Distanz

$r = \infty$: Tschebyscheff Distanz

$d_{i,k}$: Distanz der Objekte i und k

$x_{k,j}$: Wert des j-ten Merkmals des Objektes k

p : Dimension des Merkmalsraumes

q_j : Gewichtung des j-ten Merkmals

Als Beispiel wird im Folgenden die City-Block-Metrik (auch Manhattan- oder Taxifahrer-Metrik genannt) anhand der PS-Punkte illustriert, die bei praktischen Aufgabenstellungen, z.B. bei der Clusterbildung von Standorten in Städten, eine wichtige Rolle spielt. Die Distanz lässt sich berechnen, indem die Differenz zwischen jedem Merkmal für ein Objektpaar gebildet wird und die sich ergebenden absoluten Differenzwerte addiert werden (s. Formel 7) [BACKHAUS 2016]. Die Berechnung dieser Distanz (d) sei beispielhaft für das PS-Punkt-paar „PS-Punkt 1“ ($P1$) und „PS-Punkt 2“ ($P2$) durchgeführt, wobei die erste Zahl bei der Differenzbildung jeweils den Merkmalswert (Eigenschaftswert) von „PS-Punkt 1“ ($P1$) darstellt.

$$d_{p1,p2} = |2 - 2| + |4 - 2| + |2 - 1| = 3$$

Zwischen den PS-Punkten 1 und 2 ergibt sich unter Verwendung der L1-Norm eine Distanz von 3. In der gleichen Weise werden für alle anderen PS-Punkt-paare die Distanzen bestimmt. Das Ergebnis der Berechnungen enthält Tabelle 5:

Tabelle 5: Distanzmatrix von PS-Punkten anhand der City-Block Metrik

	Punkt 1	Punkt 2	Punkt 3	Punkt 4	Punkt 5
Punkt 1	0				
Punkt 2	3	0			
Punkt 3	6	5	0		
Punkt 4	4	7	8	0	
Punkt 5	9	12	9	7	0

Da ein Objekt zu sich selbst keine Distanz besitzt, besteht die Hauptdiagonale einer Distanzmatrix aus Nullen. Aus der Tabelle 5 ist deutlich zu sehen, dass mit einer Distanz von 3 das PS-Punkt-paar „PS-Punkt 1“ und „PS-Punkt 2“ die kleinste Distanz und somit die größte Ähnlichkeit aufweist. Im Gegensatz dazu besteht zwischen „PS-Punkt 2“ und „PS-Punkt 5“ die größte Distanz (12) und somit die geringste Ähnlichkeit.

Da das Merkmal mit dem größten Wertebereich einen höheren Einfluss auf die Ergebnisse als die anderen Merkmale hat, kann es bei unterschiedlichen oder nicht vergleichbaren Maßeinheiten auf den Merkmalsachsen zu verzerrten Ergebnissen kommen. Daher muss bei der Anwendung der Minkowski-Metriken besonders darauf geachtet werden, dass diese

Distanzmetrik nur vergleichbare Maßeinheiten erfordert [BACKHAUS 2016]. Das ist in dem vorliegenden Beispiel erfüllt, da alle Merkmale der PS-Punkte auf eine vereinfachte normierte Skala konvertiert wurden. Ist diese Bedingung nicht erfüllt, so müssen die Ausgangsdaten zuerst mit Hilfe einer Normierungsmethode umgewandelt werden.

Aufgrund der leichten Verständlichkeit zählt in vielen Anwendungsgebieten die Euklidische Distanz in der Praxis zu den weit verbreiteten Distanzmaßen. Dazu werden für jedes Objektpaar gemäß dem Satz von Pythagoras die Differenzwerte jedes Merkmales quadriert, addiert und aus der Summe die Quadratwurzel gezogen [BACKHAUS 2016].

Für die PS-Beispieldaten ergibt sich z.B. die Euklidische Distanz für das Punktpaar „PS-Punkt 1“ (P_1) und „PS-Punkt 2“ (P_2) zu:

$$d_{p_1,p_2}^2 = \sqrt{(2 - 2)^2 + (4 - 2)^2 + (2 - 1)^2} \approx 2,24$$

Zur Quantifizierung der Unähnlichkeit bzw. Distanz zwischen Objekten können als Maß sowohl die quadrierte Euklidische Distanz als auch die Euklidische Distanz berücksichtigt werden [BACKHAUS 2016]. Da viele Clusterungsalgorithmen auf der Euklidischen Distanz basieren, wird im Folgenden beispielhaft ebenfalls diese verwendet. Die Euklidischen Distanzen für die 5 PS-Punkte enthält Tabelle 6.

Tabelle 6: Distanzmatrix für PS-Punkten mit Euklidischen Distanzen

	Punkt 1	Punkt 2	Punkt 3	Punkt 4	Punkt 5
Punkt 1	0				
Punkt 2	2,24	0			
Punkt 3	3,74	3,32	0		
Punkt 4	4,00	4,58	4,69	0	
Punkt 5	5,20	7,07	6,40	4,36	0

Sowohl mit der Euklidischen Distanz als auch mit der City-Block-Metrik erhält man bezüglich der ähnlichsten und unähnlichsten Paare das gleiche Ergebnis. Wird die Reihenfolge der Ähnlichkeiten zwischen den PS-Punkten anhand beiden Metriken in einer Tabelle zusammengefasst (s. Tabelle 7), wird deutlich, dass es bei den Punktpaaren „PS-Punkt 1“ und „PS-Punkt 3“, „PS-Punkt 1“ und „PS-Punkt 4“, „PS-Punkt 2“ und „PS-Punkt 3“, „PS-Punkt 2“ und „PS-Punkt 4“ sowie „PS-Punkt 3“ und „PS-Punkt 5“ eine Änderung der Reihenfolge der Ähnlichkeiten gibt. Ursache hierfür ist, dass im Vergleich zur Euklidischen Distanz alle Differenzwerte der Merkmale bei der City-Block-Metrik gleichgewichtig in der Berechnung beurteilt worden sind. Somit ist ersichtlich, dass die Wahl des Distanzmaßes einen Einfluss auf die Berechnung der Ähnlichkeiten (Ähnlichkeitsreihenfolge) der Untersuchungsobjekte besitzt und demzufolge zu unterschiedlichen Ergebnissen führt.

Tabelle 7: Reihenfolge der Ähnlichkeiten zwischen den PS-Punkten anhand der City-Block Distanz (jeweils erste Zahl) und der Euklidischen Distanz (Werte in Klammern)

	Punkt 1	Punkt 2	Punkt 3	Punkt 4	Punkt 5
Punkt 1	0				
Punkt 2	1(1)	0			
Punkt 3	4(3)	3(2)	0		
Punkt 4	2(4)	5(6)	7(7)	0	
Punkt 5	8(8)	10(10)	8(9)	5(5)	0

4.3.2. Korrelationskoeffizient zur Bestimmung der Ähnlichkeit

Außer der oben dargestellten Minkowski-Metriken erweist sich der Korrelationskoeffizient, der auch „Element zu Element Korrelationskoeffizient“ oder „Q- Korrelationskoeffizient“ genannt wird [FLETTLING 2010] [BACKHAUS 2016], als ein weiteres gebräuchliches Ähnlichkeitsmaß zur Bestimmung der Ähnlichkeiten zwischen Objekten mit metrischer Merkmalsstruktur. Während das Distanzmaß der Minkowski-Metriken als ein Maß für eine Distanz zwischen den Objekten im Merkmalsraum zu verstehen ist, dient der Korrelationskoeffizient als ein Maß zur Bestimmung der linearen Abhängigkeit der Merkmalsvektoren verschiedener Untersuchungsobjekte untereinander. Der Korrelationskoeffizient berechnet sich gemäß Formel 12 [BACKHAUS 2016]:

$$\text{Formel 12: } r_{i,k} = \frac{\sum_{j=1}^p (x_{i,j} - \bar{x}_i)(x_{k,j} - \bar{x}_k)}{\sqrt{\sum_{j=1}^p (x_{i,j} - \bar{x}_i)^2 \cdot \sum_{j=1}^p (x_{k,j} - \bar{x}_k)^2}}$$

mit:

$r_{i,k}$: Korrelationskoeffizient

\bar{x}_i : Merkmalsdurchschnitt (Mittelwert) für Objekt i

Da für den Korrelationskoeffizienten $-1 \leq r_{i,k} \leq 1$ gilt, ist bei der praktischen Anwendung von Korrelationskoeffizienten als Ähnlichkeitsmaß noch zu entscheiden, ob $r_{i,k} = -1$ oder $r_{i,k} = 0$ die größte Unähnlichkeit widerspiegelt. Während im Falle von $r_{i,k} = 0$ zwischen den Merkmalsvektoren der Untersuchungsobjekte keine lineare Beziehung besteht, bedeutet $r_{i,k} = -1$ oder $r_{i,k} = 1$ eine starke negative oder positive Abhängigkeit, Diese Werte können als die größte Ähnlichkeit zwischen den Untersuchungsobjekten angesehen werden [FLETTLING 2010].

Bei der Berechnung der Ähnlichkeit zwischen zwei Objekten i und k mit Hilfe des Korrelationskoeffizienten werden alle Merkmale eines Objektes berücksichtigt. So ergibt sich

z. B. für „PS-Punkt1“ ein Merkmalsdurchschnitt von $(2+4+2)/3 = 8/3$ und für „PS-Punkt2“ ein Merkmalsdurchschnitt von $(2+2+1)/3 = 5/3$.

Mit Hilfe der Formel 12 lässt sich die Ähnlichkeit zwischen „PS-Punkt1“ ($P1$) und „PS-Punkt2“ ($P2$) beispielsweise wie folgt ermitteln:

$$r_{p1,p2} = \frac{6/9}{\sqrt{24/9 * 6/9}} = 0,5$$

Die Berechnungen der Korrelationskoeffizienten für alle Punktpaare führen für das Beispiel zu der in Tabelle 8 dargestellten Ähnlichkeitsmatrix.

Tabelle 8: Ähnlichkeitsmatrix von PS-Punkten mit den Korrelationskoeffizienten

	Punkt 1	Punkt 2	Punkt 3	Punkt 4	Punkt 5
Punkt 1	1				
Punkt 2	0,5	1			
Punkt 3	-0,94	-0,76	1		
Punkt 4	0	0,87	-0,33	1	
Punkt 5	1	0,5	-0,94	0	1

Vergleicht man diese Ähnlichkeitsmatrix mit der Distanzmatrix aus Tabelle 6, so zeigt sich, dass sich die Beziehungen zwischen den PS-Punkten stark verändert haben. Entsprechend der Euklidischen Distanz besitzen „PS-Punkt1“ und „PS-Punkt5“ geringe Ähnlichkeit, jedoch werden sie mittels der Q-Korrelationskoeffizienten als das ähnlichste Punktpaar angesehen. Ebenfalls wurden „PS-Punkt1“ und „PS-Punkt4“ mit einer Distanz von 4 nach der Euklidischen Distanz als ähnlich erkannt, während sie gemäß einem Korrelationswert von 0 in Tabelle 8 überhaupt keine Ähnlichkeit aufweisen. Es ist durch diese Vergleiche deutlich zu erkennen, dass je nach Aufgabenstellung zunächst inhaltliche Überlegungen für die Auswahl des Proximitätsmaßes von großer Bedeutung sind. Zu diesem Zweck werden in Abbildung 29 die Höhenänderungsverläufe von „PS-Punkt1“ und „PS-Punkt5“ dargestellt.

Die Höhenänderungsverläufe zeigen, dass sich die Merkmalswerte von „PS-Punkt1“ und „PS-Punkt5“ zwar deutlich unterscheiden, ihre Höhenänderungsverläufe jedoch völlig gleich sind. Dadurch lässt sich erklären, warum sie im Hinblick auf die Distanzmaße als vollkommen unähnlich und im Hinblick auf die Korrelationskoeffizienten als vollkommen ähnlich zu bewerten sind.

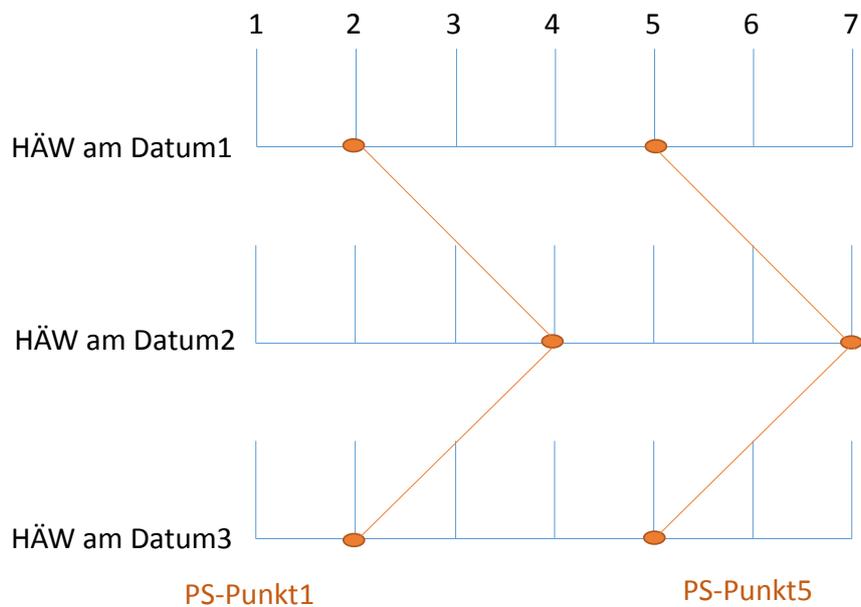


Abbildung 29: Vergleich der Höhenänderungsverläufe von „PS-Punkt1“ und „PS-Punkt5“

Den Unterschied zwischen euklidischer Distanz und Korrelationskoeffizient als Ähnlichkeitsmaß erkennt man beispielsweise in der vereinfachten Abbildung 30.

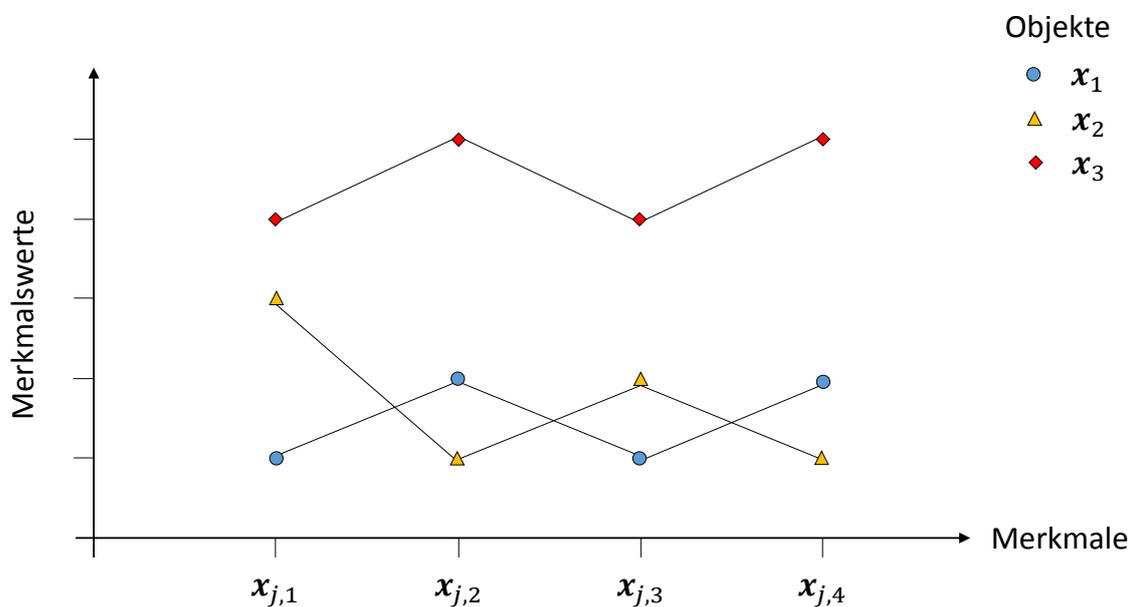


Abbildung 30: Vereinfachte Profilverläufe der Merkmale von drei Untersuchungsobjekten [VERÄNDERT NACH FLETLING 2010]

In Abbildung 30 sind die Profilverläufe der Merkmale von drei Untersuchungsobjekten dargestellt. Es wird deutlich, dass die Objekte 1 und 2, deren Merkmalsprofile ziemlich dicht

nebeneinander verlaufen und somit auf eine relative kleine Distanz dazwischen hinweisen, sich jedoch erheblich hinsichtlich ihrer Profilverläufe unterscheiden, während die Objekte 1 und 3 einen fast gleichen Profilverlauf , allerdings mit einer relativ großen Distanz zueinander aufweisen. Das führt dazu, dass das Objektpaar 1/2 ähnlich entsprechend dem Distanzmaß und unähnlich entsprechend dem Korrelationskoeffizienten eingestuft wird, während das Objektpaar 1/3 unähnlich entsprechend dem Distanzmaß und ähnlich entsprechend dem Korrelationskoeffizient beurteilt wird [FLETILING 2010].

Die Anwendung von Distanzmaßen oder Korrelationskoeffizienten zur Bestimmung der Ähnlichkeit zwischen Untersuchungsobjekten lässt sich im Allgemeinen wie folgt beurteilen:

- Wenn der absolute Abstand zwischen Objekten von Bedeutung ist, eignen sich Distanzmaße am besten.
- Wenn aber mehr Gewicht auf die Ähnlichkeit (also den Gleichlauf) zweier Höhenänderungsverläufe gelegt wird, eignet sich der Korrelationskoeffizient besser. Dann spielt es keine Rolle mehr, auf welchem Niveau sich die Merkmalswerte der Objekte befinden [FLETILING 2010].

4.4. Auswahl des Algorithmus zur Gruppierung

Nachdem die Distanz- oder Ähnlichkeitsmatrix, die als Ausgangspunkt der Clusteralgorithmen für die Zuordnung von Objekten in Clustern dient, mit Hilfe von Proximitätsmaßen aus den Ausgangsdaten berechnet worden ist, erfolgt der nächste Arbeitsschritt für die Clusteranalyse, nämlich die Auswahl des Cluster-Algorithmus.

Es gibt ein breites Methodenspektrum an veröffentlichten Cluster-Algorithmen zur Gruppierung von Objekten. Entsprechend ihrer Ähnlichkeit lassen sie sich in Gruppen zusammenfassen. Zunächst können, unter Berücksichtigung der Anzahl der betrachteten Merkmale, die monothetischen und polythetischen Cluster-Algorithmen unterschieden werden. Bei den monothetischen Algorithmen wird jeweils nur ein Merkmal zur Gruppierung berücksichtigt, während die polythetischen Cluster-Algorithmen dadurch charakterisiert sind, dass alle relevanten Merkmale zur Gruppierung der Objekte herangezogen werden. Da sich die meisten Aufgabenstellungen zur Clusterung auf mehrere Merkmale beziehen, erfüllen nur die polythetischen Cluster-Algorithmen diese Voraussetzung.

Eine weitere Einteilung der Clusteralgorithmen lässt sich anhand der Vorgehensweise bei der Clusterung vornehmen. Ein entsprechender Überblick über die verschiedenen Clustermethoden ist in Abbildung 31 dargestellt [BACKHAUS 2016].

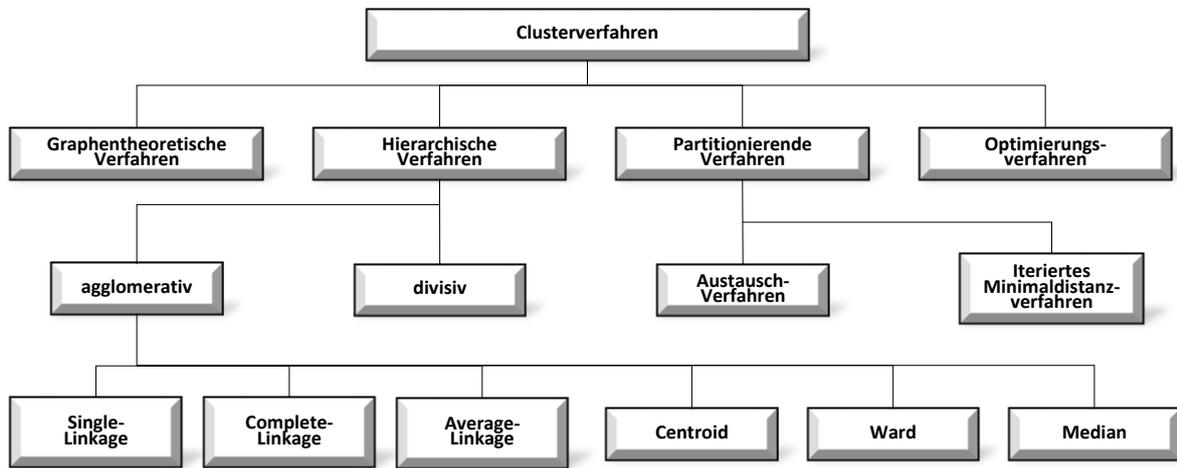


Abbildung 31: Einteilung von ausgewählten Cluster-Algorithmen [VERÄNDERT NACH BACKHAUS 2016]

Aus der Vielzahl von existierenden Verfahren wird im Folgenden auf die folgenden vier Arten von Clusteralgorithmen näher eingegangen: hierarchische Verfahren, partitionierende Verfahren, graphentheoretische Verfahren und Optimierungsverfahren, wobei die ersten zwei Verfahren aufgrund ihrer großen Bedeutung und breiten Anwendungsmöglichkeiten ausführlich erläutert werden.

4.4.1. Hierarchische Verfahren

Die hierarchischen Verfahren lassen sich in agglomerative und divisive Verfahren unterteilen. Bei den agglomerativen Verfahren wird jedes Objekt in der Ausgangssituation für sich selbst als einzelnes Cluster zu betrachten. Es bildet somit die feinste Partition, d.h. die Anzahl der Cluster entspricht der Anzahl der Objekte (bei n Objekten n Cluster). Bei den divisiven Verfahren hingegen wird von der größten Partition ausgegangen, d.h. alle Untersuchungsobjekte werden zuerst in einer Gruppe zusammengefasst. Somit lässt sich der Ablauf der agglomerativen Verfahren durch die Fusionierung (Zusammenfassung von Clustern) und der der divisiven Verfahren durch die Zerlegung (Aufteilung einer Gesamtheit in Clustern) fortsetzen [BACKHAUS 2016]. Dies kann häufig mithilfe eines Dendrogramms (baumartiges Diagramm) beschrieben werden, wobei die zwei umgekehrten Konstruktionsprinzipien durch die Richtung der Pfeile verdeutlicht sind (s. Abb. 32). Die Höhe eines Querbalkens in dem Dendrogramm drückt die Distanz aus, ab der Cluster zusammengefügt oder zerlegt werden. Somit kann das Dendrogramm dazu dienen, eine optimal erscheinende Clusteranzahl festzulegen [FLETTLING 2010] [BACKHAUS 2016].

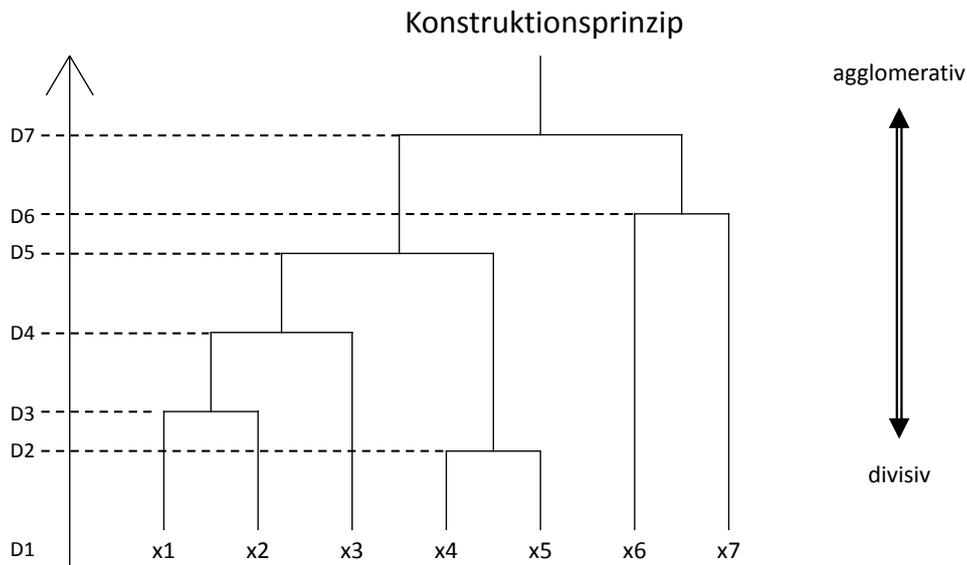


Abbildung 32: Dendrogramm der hierarchischen Verfahren [VERÄNDERT NACH BERGS 1981]

Die wesentlichen Ablaufschritte der agglomerativen Verfahren können wie folgt beschrieben werden [BACKHAUS 2016]:

1. Jedes Untersuchungsobjekt stellt zuerst ein eigenes Cluster dar.
2. Für alle Objektpaare der gesamten Untersuchungsobjekte wird die Ähnlichkeit bzw. Distanz berechnet und eine Ähnlichkeits- bzw. Distanzmatrix erstellt.
3. Die beiden Cluster, die die größte Ähnlichkeit bzw. geringste Distanz zueinander besitzen, werden gesucht und zu einem neuen Cluster fusioniert. Dadurch verringert sich die Anzahl der Cluster um 1.
4. Es werden die Distanzen zwischen dem neuen fusionierten Cluster und den übrigen Clustern berechnet, wodurch eine reduzierte Distanzmatrix entsteht.
5. Der Ablauf der Schritte 3 bis 5 wird so lange wiederholt, bis alle betrachteten Untersuchungsobjekte in einem Cluster zusammengefasst sind.

Die Fusionierungsverfahren, mit denen Distanzen zwischen Clustern zur Zusammenfügung ermittelt werden, werden im Folgenden erläutert:

Single Linkage: Zuerst werden sämtliche Distanzen zwischen den Objekten des einen Clusters und den Objekten des anderen Clusters berechnet. Im Anschluss daran werden die zwei Cluster zu einem Cluster zusammengefasst, das die geringste Distanz zwischen jeweils am nächsten beieinanderliegenden Objekten zweier Cluster aufweist. Deshalb nennt man das Verfahren auch „nearest neighbour“ oder „Minimummethode“. Da der Fusionierungsprozess bei diesem Verfahren nur über jeweils zwei benachbarte Objekte bestimmt wird, kann es

manchmal zu unerwünschten kettenförmigen Clusterbildungen führen. Andererseits ist dieses Verfahren sehr gut geeignet, wenn tatsächlich längliche Cluster im Datensatz zu erwarten ist [FLETTLING 2010] [CLEFF 2015].

Complete Linkage: Wie beim Single Linkage lassen sich mit dem Complete Linkage ebenfalls sämtliche Distanzen zwischen den Objekten des einen Clusters und den Objekten der anderen Cluster berechnen. Zusammengefasst werden die beiden Cluster, deren Distanz zwischen am weitesten auseinanderliegenden Objekten zweier Cluster minimal ist. Das Verfahren wird deshalb auch als „furthest neighbour“ oder Maximummethode bezeichnet. Im Gegensatz zum Single Linkage führt es zur Ausbildung von breiten und kompakten Clustern [FLETTLING 2010] [CLEFF 2015].

Average Linkage: Wie bei dem Single Linkage und Complete Linkage werden hier zunächst auch sämtliche Distanzen zwischen den Objekten des einen Clusters und den Objekten der anderen Cluster berechnet. Zusammengefasst werden nur die beiden Cluster, deren durchschnittliche Distanz (arithmetisches Mittel) zwischen den Objekten zweier Cluster am kleinsten ist [FLETTLING 2010] [CLEFF 2015].

Centroid: Für jeden Cluster getrennt wird zunächst der Clusterschwerpunkt (Centroid) als arithmetischer Mittelwert aus den zugehörigen Objekten berechnet. Im Anschluss daran wird die Distanz zwischen den Clusterschwerpunkten zweier Cluster berechnet. Zusammengefasst werden die beiden Cluster, deren Distanz zwischen den Clustercentroiden am kleinsten ist. Der Centroid des neuen Clusters lässt sich dann aus dem gewichteten arithmetischen Mittel der beiden neuen vereinigten Ausgangscentroide berechnen, dabei wird die Anzahl der Objekte in den einzelnen Clustern als Gewichtung herangezogen [FLETTLING 2010] [CLEFF 2015].

Median: Ähnlich wie die Centroid-Clusterung erfolgt die Berechnung des Centroids des neuen Clusters als ungewichtetes statt gewichtetes arithmetisches Mittel aus den Centroiden der neuen vereinigten Ausgangscluster [FLETTLING 2010].

Ward: Bei der Ward Methode werden zunächst wie bei der Centroid-Clusterung für jeden Cluster die Clusterschwerpunkte als arithmetisches Mittel aus den zugehörigen Objekten berechnet. Im Anschluss daran werden die quadrierten euklidischen Distanzen der einzelnen Objekte aller Cluster zu allen Clusterschwerpunkten gemessen. Summiert werden diese für jede Clusterkombination. Dabei wird diejenige Clusterbildung durchgeführt, durch welche die Vereinigung die Quadratsumme innerhalb aller Cluster minimal zunimmt [FLETTLING 2010] [CLEFF 2015].

Die einzelnen Fusionierungsverfahren sind in Abbildung 33 schematisch dargestellt [CLEFF 2015].

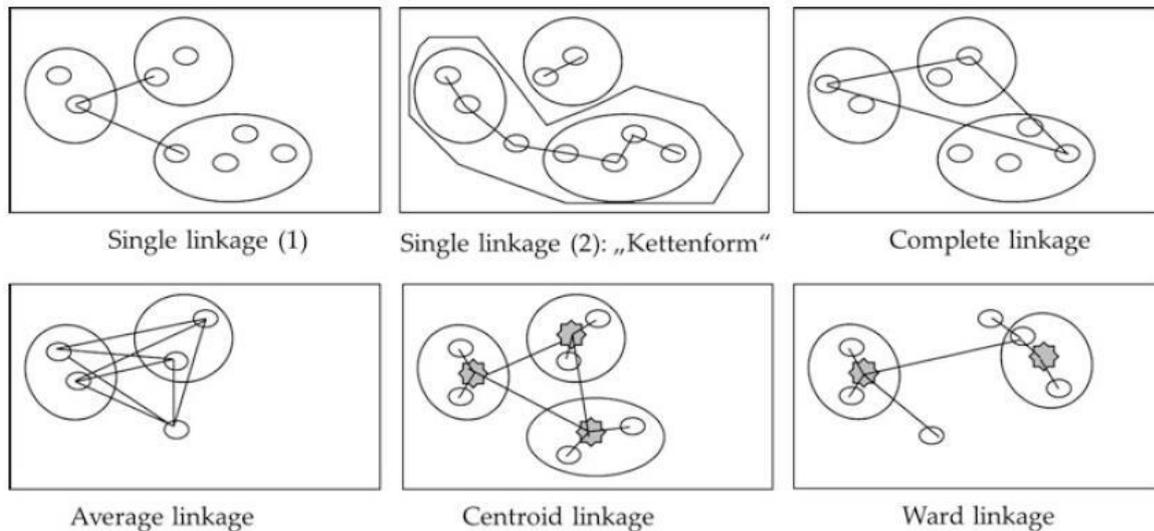


Abbildung 33: Schematische Darstellung der Fusionierungsverfahren [CLEFF 2015]

Die wesentlichen Charakteristiken der obengenannten Fusionierungsverfahren werden in Tabelle 9 aufgelistet [ECKSTEIN 2014].

Tabelle 9: Wesentliche Charakteristika der Fusionierungsverfahren [VERÄNDERT NACH ECKSTEIN 2014]

Verfahren	Charakteristik
Linkage-Methode (<i>Linkage</i> → Verknüpfung)	
<i>single linkage</i> (Nächstgelegener Nachbar)	Kleinste Distanz zwischen zwei Objekten zweier Cluster; beliebiges Proximitätsmaß
<i>complete linkage</i> (entferntester Nachbar)	Größte Distanz zwischen zwei Objekten zweier Cluster, beliebiges Proximitätsmaß
<i>average linkage</i>	Mittlere Distanz zwischen den Objekten zweier Cluster, beliebiges Proximitätsmaß
Zentroid-Methode (<i>Centroid</i> → Schwerpunkt)	
<i>Centroid</i>	Distanz zwischen den Zentroiden zweier Cluster; metrische Clustermerkmale, quadriertes euklidisches Distanzmaß
Median	Modifiziertes Zentroid-Verfahren, quadriertes euklidisches Distanzmaß
Varianz-Methode	
Ward	Kleinsten Zuwachs der Fehlerquadratsumme bei Clusterfusionierung, metrische Clustermerkmale, quadriertes euklidisches Distanzmaß

In Gegensatz zu den agglomerativen Verfahren stellen bei den divisiven Verfahren zunächst alle Objekte ein gemeinsames Cluster dar, welches im Anschluss daran durch schrittweise Reduzierung des Distanzniveaus so lange zerlegt wird, bis die maximale Distanz von zwei sich innerhalb eines Cluster befindenden Objekten kleiner ist als eine vordefinierte Grenzdistanz [FLETLING 2010]. Wegen des extrem hohen Rechenaufwands finden divisive Verfahren in der Praxis im Vergleich zu den weit verbreiteten agglomerativen Verfahren kaum Anwendung [TAN 2009].

4.4.2. Partitionierende Verfahren

Die Grundidee partitionierender Verfahren besteht darin, dass ausgehend von einer vorgegebenen Clustereinteilung versucht wird, durch schrittweise Verlagerung der Objekte von einem Cluster zu einem anderen ein vorgegebenes Gütemaß oder eine Zielfunktion zu optimieren und so eine bessere Lösung der Clusterung zu finden [FLETLING 2010] [BACKHAUS 2016]. Der wesentliche Unterschied zu den hierarchischen Verfahren liegt darin, dass die Anzahl der Cluster zu Beginn der Clusteranalyse festgelegt werden muss. Hinzu kommt noch, dass eine erste Zuordnung der Untersuchungsobjekte zu den einzelnen Clustern als Anfangspartition vorgegeben werden muss. Hierzu spielt es zunächst keine Rolle, ob die Anfangspartition tatsächlich sinnvoll oder sachlogisch vordefiniert wurde, oder die einzelnen Untersuchungsobjekte einfach nur beliebig oder sogar willkürlich zugeordnet worden sind. Das Verfahren beinhaltet eine sukzessive Zuordnung der Untersuchungsobjekte zu den einzelnen Clustern, um so die Qualität der Clusterbildung iterativ zu verbessern. Die Qualität der Anfangspartition besitzt nur einen Einfluss auf die Rechenzeit; je besser die Anfangspartition für die Clusterbildung ausgewählt wird, desto schneller nähert sich die Clusteranalyse dem Endergebnis.

Eine optimale Lösung der Clusterbildung mit dem partitionierenden Verfahren könnte idealerweise gefunden werden, indem man sämtliche möglichen Clusterkombinationen (Partitionen) der Untersuchungsobjekte durchführt und jeweils ein vorgegebenes Gütemaß berechnet. Die Clusterkombination mit dem minimalen bzw. maximalen Wert des Gütemaßes ist dann als optimale Clusterbildung anzusehen. Praktisch wäre diese Vorgehensweise in der Regel fast unmöglich, weil die Anzahl möglicher Clusterkombinationen exponentiell mit der Anzahl der Untersuchungsobjekte ansteigt und die benötigten Rechenzeiten schon bei mittelgroßen Datensätzen mit geringer Clusteranzahl sehr groß werden [FLETLING 2010]. Aus wirtschaftlichen Gründen besteht die Notwendigkeit, die Anzahl aller möglichen Aufteilungen der Objekte einzuschränken. Es droht allerdings die Gefahr, dass die optimale Clusterbildung so nicht entdeckt werden kann [FLETLING 2010].

Für die Bewertung der Qualität der partitionierenden Clusterbildung dient als ein sehr gebräuchliches Gütemaß das Varianzkriterium g_{Var} (s. Formel 13).

Formel 13:

$$g_{Var} = \sum_{j=1}^c \sum_{i=1}^{n_j} d^2(x_i, \bar{x}_j)$$

mit:

g_{Var} : Varianzkriterium

x_i : i -tes Objekt

\bar{x}_j : Clusterzentrum des j -ten Clusters

n : Anzahl der Objekte im j -ten Cluster

d : Distanz

c : Anzahl der Cluster

Dazu wird zunächst für jedes Objekt die euklidische Distanz zum jeweiligen Clusterzentrum berechnet. Ähnlich wie bei der Berechnung der empirischen Varianz werden im Anschluss daran alle Distanzen innerhalb eines Clusters quadratisch aufsummiert. Dann wird die Summe aller Clustersummen gebildet. Das kleinere Varianzkriterium entspricht einer besseren Qualität der Clusterbildung. Beispielsweise ergibt sich bei dem Ausgangsdatensatz mit relativ dichten Punktwolken von Objekten im Merkmalsraum ein minimaler Wert für das Varianzkriterium, wenn sich die Clusterzentren idealerweise um den Schwerpunkten der Punktwolken befinden und damit die Distanzen der Objekte innerhalb eines Clusters zu den Clusterzentren relativ klein sind [FLETTLING 2010].

In Bezug auf das Kriterium zur Suche nach einer optimalen Clusterbildung unterscheidet man im Kontext eines partitionierenden Clusterverfahrens zwischen optimierenden Austauschverfahren und iterativen Minimal-Distanz-Verfahren. Im Vergleich zu den optimierenden Austauschverfahren, die unmittelbar auf einem Optimierungskriterium basieren, machen die iterativen Minimal-Distanz-Verfahren im Allgemeinen von einem „mittelbaren“ Optimierungskonzept Gebrauch, das wie folgt beschrieben werden kann [ECKSTEIN 2014]:

Eine Verbesserung einer vorgegebenen Clusterbildung wird iterativ dadurch intendiert, dass durch einen Objektaustausch die quadrierte Euklidische Distanz des ausgetauschten Objekts

zum Centroid der neu entstandenen Cluster kleiner als zum Centroid der ursprünglichen Cluster ist [ECKSTEIN 2014].

Im Folgenden werden die wesentlichen Ablaufschritte des Austauschverfahrens kurz erläutert [STEINHAUSEN 1977]:

1. Eine Anfangspartition wird ausgewählt.
2. Für jede Gruppe wird das arithmetische Mittel je Merkmal bestimmt.
3. Die Fehlerquadratsumme wird jeweils für die Clusterbildung über alle Cluster ermittelt.
4. Es wird durch Verlagerung von Objekten sukzessiv angestrebt, das Varianzkriterium zu vermindern.
5. Das Objekt, das eine maximale Verringerung des Varianzkriteriums ermöglicht, wird in den entsprechenden Cluster verlagert.
6. Für die durch Verlagerung von Objekten veränderten Cluster ist das neue arithmetische Mittel zu berechnen.

Die Schritte 3 bis 6 sollen solange wiederholt werden, bis alle Objekte angesichts ihrer Verlagerung untersucht wurden und keine Verbesserung des Varianzkriteriums zu erwarten ist [BACKHAUS 2016].

Der schematische Ablauf der Clusterbildung von acht Objekten in zwei Clustern mit dem iterierten Minimaldistanzverfahren wird in Abbildung 34 dargestellt [LANGE 2013].

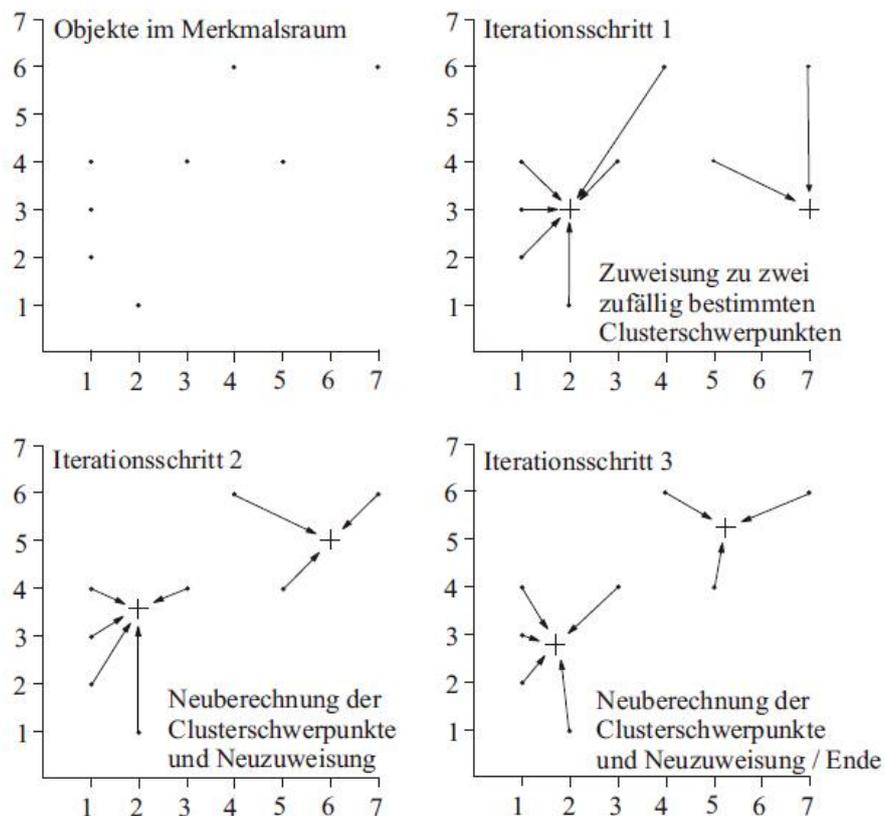


Abbildung 34: Ablaufschritte beim iterierten Minimaldistanzverfahren [LANGE 2013]

4.4.3. Graphentheoretische Verfahren

In Abhängigkeit der Topologie der Objektmenge werden bei den graphentheoretischen Verfahren nur Distanzen zwischen Objekten, die über eine gemeinsame Kante in dem zugehörigen Graphen verbunden sind, berechnet. Sowohl im Objektraum als auch im Merkmalsraum kann der Graph erstellt werden. Die Graphentheorie liefert bei speziellen Aufgabenstellungen sehr effektive Ergebnisse. Besonders gut eignet sich das graphentheoretische Verfahren für die Bestimmung von Single Linkage Clustern und für die Clusterbildung von räumlichen Daten, bei denen die geographischen Nachbarschaftsbeziehungen von großer Bedeutung sind [FLETTLING 2010].

4.4.4. Alternative Optimierungsverfahren

Zu den alternativen Optimierungsverfahren werden alle anderen Verfahren gezählt, die nicht zu den hierarchischen-, partitionierenden- und graphentheoretischen Verfahren gehören. z.B. die künstlichen neuronalen Netze, die in der vorliegenden Arbeit nicht weiter verfolgt werden [FLETTLING 2010].

4.5. Bestimmung einer optimalen Clusteranzahl

Im vorherigen Kapitel wurde gezeigt, dass es verschiedene Clusteranalysealgorithmen für eine Fusionierung von Einzelobjekten zu Clustern gibt. Dabei beginnen alle agglomerativen Verfahren bei der feinsten Partition (jedes Objekt bildet jeweils ein eigenständiges Cluster) und enden mit einer Fusionierung aller betrachteten Objekte in einer großen Gruppe [BACKHAUS 1996]. Hierbei ist noch zu entscheiden, welche Anzahl von Clustern als „optimal“ gilt. Da die Clusterverfahren nur die lokalen Optima für die Clusterbildung finden, gibt es im engeren Sinne keine optimale Clusteranzahl [FLETTLING 2010]. Basierend auf subjektiven Kriterien ist es oft unmöglich oder zumindest sehr schwierig die scheinbar optimale Clusteranzahl zu bestimmen. Besonders wenn der Anwender über keine Vorkenntnisse zu den vorliegenden Datensätzen und über keine sachlogisch begründbaren Vorstellungen zur Gruppierung der Untersuchungsobjekte verfügt, empfiehlt es sich, sich bei der Bestimmung der Clusterzahl an statistischen Kriterien zu orientieren. Diese sind, aus theoretischer Sicht, objektive Ansätze bei der Suche nach einer optimalen Clusteranzahl für die hierarchischen und partitionierenden Clusterverfahren [BACKHAUS 2016].

Die wesentliche Aufgabe bei den partitionierenden Clusterverfahren ist es üblicherweise, ein Gütemaß zu optimieren. Dabei erweist sich der Wert des Gütemaßes (zum Beispiel des Varianzkriteriums) in der Regel als eine mit zunehmender Clusteranzahl c fallende monotone

Funktion $f(c)$. Bei der wiederholten Durchführung von partitionierenden Verfahren mit jeweils unterschiedlichen Clusteranzahlen erhält man dazu unterschiedliche erreichte Gütemaße. Solche Werte trägt man dann in einem Diagramm über der entsprechenden Clusteranzahl auf (s. Abb. 35). Als optimaler Wert c^* für die Clusteranzahl wird der angesehen, wo der Kurvenverlauf einen erkennbaren Knick aufzeigt. Dieses Verfahren wird in der Literatur auch als „Ellbogen-Kriterium“ bezeichnet [BOCK 1994] [FLETTLING 2010].

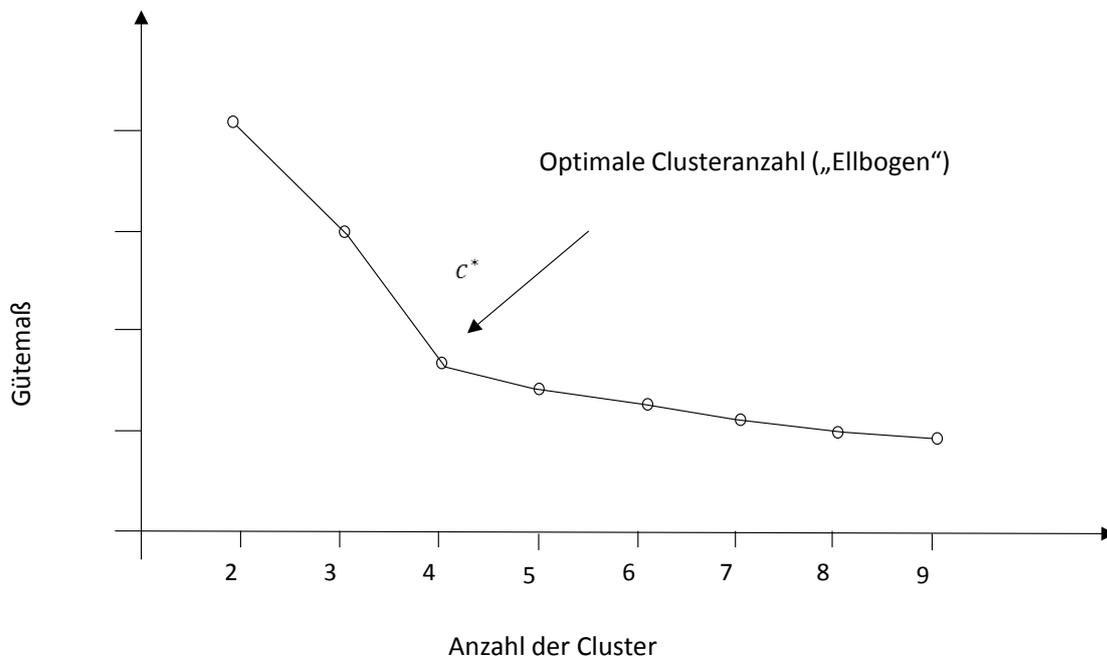


Abbildung 35: „Ellbogen-Kriterium“ zur Ermittlung der optimalen Clusteranzahl bei den partitionierenden Verfahren [VERÄNDERT NACH FLETTLING 2010]

Ähnlich wird für die Bestimmung der optimalen Clusteranzahl bei den hierarchischen Clusterverfahren angenommen, dass relativ große Intervalle des Distanzniveaus im Dendrogramm, in denen jedoch keine Zusammenfügen oder Zerlegungen von Clustern stattfinden, auf eine optimale Clusteranzahl hindeuten [FLETTLING 2010].

5. Moran's-Index - Clusterverfahren

In vielen wissenschaftlichen Disziplinen, wie z.B. in der Ökologie, der Astronomie, der Epidemiologie, den Geowissenschaften sowie auch den Sozialwissenschaften werden Daten mit räumlichen Bezügen erfasst und analysiert. Solche räumlichen Daten besitzen eine bestimmte räumliche Lage auf der Erdoberfläche, die durch Koordinaten in verschiedenen Koordinatensystemen definiert ist. Außer den geometrischen Informationen (Form der Objekte als Punkte, Linien oder Polygone) werden oft auch die zugehörigen Sachinformationen von und an den Orten erhoben. Aufbauend auf den herkömmlichen Datenbeständen erweitern sich räumliche Daten um eine weitere räumliche Dimension und sind durch die räumliche Abhängigkeit und Heterogenität charakterisiert, wobei die räumlichen Nachbarschaftseffekte, bei denen sich räumlich nahe beieinander liegende Objekte ähnlich verhalten, von großer Bedeutung sind. Aufgrund dessen werden bei den räumlichen Clusterverfahren die gegenseitigen Abhängigkeiten von Beobachtungsobjekten basierend auf ihrer räumlichen Lage zueinander berücksichtigt und die räumlichen Abstände zwischen den Objekten als Kriterien zur Abbildung der Ähnlichkeiten angewendet.

5.1. Clusterverfahren mit Raumbezug

Obwohl sich Ähnlichkeiten zwischen räumlichen und nicht-räumlichen Clusterverfahren ergeben, stellen räumliche Daten aufgrund ihrer Multidimensionalität und der großen Datenmengen besondere Anforderungen an räumliche Clusterverfahren. Es besteht die Notwendigkeit, schnell und effizient große räumliche Datenmengen zu verarbeiten, ohne dass die Qualität der Clusterergebnisse beeinträchtigt wird. Außerdem müssen die räumlichen Clusterverfahren in der Lage sein, unempfindlich gegenüber Datenlücken und die Reihenfolge der Eingangsdaten, unregelmäßige Formen von Clustern zu identifizieren. Darüber hinaus ist es im Vergleich zu nicht-räumlichen Clusterverfahren nicht erforderlich, die Clusteranzahl vorab anzugeben, so dass räumliche Clusterverfahren durchgeführt werden können, auch wenn der Anwender nicht über Vorkenntnisse über räumlichen Daten verfügt [HAN 2001] [KOLATCH 2001].

In Anlehnung an die mögliche Einteilung von ausgewählten allgemeinen Cluster-Algorithmen (s. Abb. 31) ist ein Überblick der verschiedenen Arten von ausgewählten räumlichen Cluster-Algorithmen in Abbildung 36 dargestellt [KOLATCH 2001].

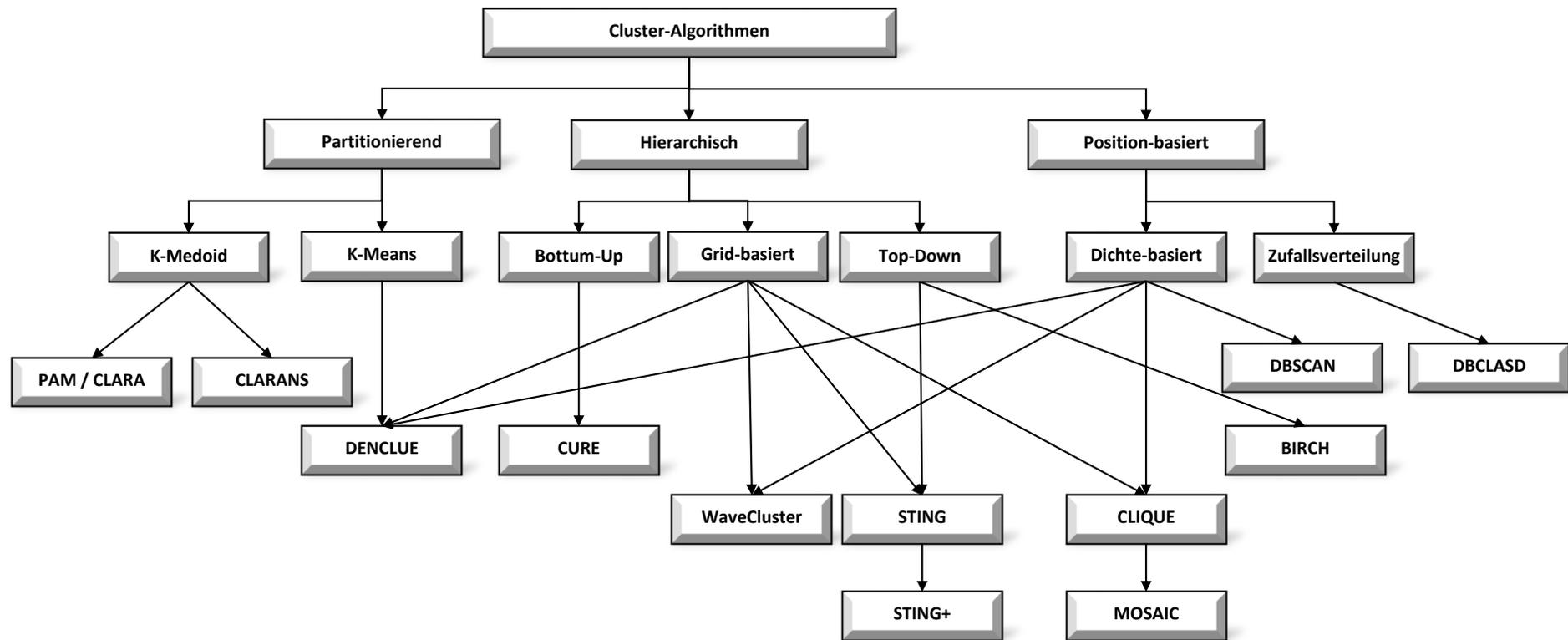


Abbildung 36: Übersicht und Einteilung von ausgewählten räumlichen Cluster-Algorithmen [VERÄNDERT NACH KOLATCH 2001]

Zusammenfassend sind die räumlichen Clusterverfahren in drei Kategorien einzuteilen: partitionierende, hierarchische und Position-basierte Verfahren. Zu den partitionierenden Clusterverfahren zählen zunächst die häufig angewendeten K-Medoid und K-Means Verfahren. Darauf basierend wurde das PAM (Partitioning around Medoids) [KAUFMAN 1987], CLARA (Clustering Large Applications) [KAUFMAN 1990] sowie CLARANS (A Clustering Algorithm based on Randomized Search) [NG 1994] zur Identifizierung von räumlichen Clustern entwickelt. Im Vergleich zum PAM, das besonders bei großen Datensätzen an seinem Grenzen gestoßen ist, hat CLARA den Vorteil, große räumliche Datensätze effizient zu verarbeiten. Beruhend auf dem K-Means Clusterverfahren zählt CLARANS zu einem der ersten räumlichen Clusterverfahren, das besonders für die Clusterung von räumlichen Daten in räumlichen Datenbanken entwickelt wurde. Zu den hierarchischen Clusterverfahren gehören beispielsweise die CURE (Clustering Using Representatives) [GUHA 1998] und STING (Statistical Information Grid-based method), STING+ [WANG 1997] sowie BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies) [ZHANG 1996] Clusterverfahren. Im Vergleich dazu sind DBSCAN (Density Based Spatial Clustering of Applications with Noise) [ESTER 1996] und DBCLASD (Distribution Based Clustering of Large Spatial Databases) [XU 1998] den Position-basierten räumlichen Clusterverfahren zugeordnet, wobei die Beobachtungsobjekte anhand ihrer räumlichen Dichten zu verschiedenen Clustern zusammengefasst werden; allerdings erlaubt das DBCLASD die zufällige Verteilung von Objekten. Die DENCLUE (DENSity basted CLUstEring) [HINNEBURG 1998], WaveCluster [SHEIKHOESLAMI 1998], CLIQUE (Clustering In QUEst) [AGRAWAL 2005] und MOSAIC (Agglomerative Clustering with Gabriel Graphs) [JIAMTHAPTHAKSIN 2010] sind sogenannte Hybrid-Clusterverfahren, die auf einer kombinierten Anwendung von Clusteralgorithmen verschiedener Kategorien basieren.

Angesichts der Leistungsfähigkeit, Behandlung von hoher Dimensionalität, Identifizierung von irregulären Clusterformen, Empfindlichkeit gegen Anomalien, Unabhängigkeit von der Reihenfolge der Inputdaten sowie Notwendigkeit der Vorkenntnisse, die als wichtige Kriterien zur Bewertung eines räumlichen Clusterverfahren dienen, lassen sich die obengenannten Clusterverfahren gemäß Tabelle 10 charakterisieren [KOLATCH 2001].

Tabelle 10: Bewertung einiger ausgewählter räumlicher Clusterverfahren [VERÄNDERT NACH KOLATCH 2001]

Algorithmus	Effektivität	Behandlung von hohen Dimensionalität	Behandlung von irregulären Clusterform	Robustheit gegen Anomalien	Abhängigkeit von der Reihenfolge der Inputdaten	Vorkenntnisse
CLARANS	$O(KN)^2$	Nein	nicht absolute	nicht absolute	Ja	2 Parameter benötigt
DBSCAN	$O(N \log N)$	Nein	nicht absolute	Ja	Ja	2 Parameter benötigt
DBCLASD	Laufzeit: 1.5-3x DBSCAN	Nein	Besser als DBSCAN	Ja	Nein	Ja
STING	$O(K)$ K:Clusteranzahl	Nein	Clusters sind ähnlich	Ja	Ja	Ja
BIRCH	$O(N)$	Nein	Nicht einfach und sehr aufwendig	Teilweise	Nein	Parameter benötigt
WaveCluster	$O(N)$ Nur für niedrige Dimensionalität	Nicht gut	Ja	Ja	Ja	Ja
DENCLUE	$O(D \log D)$ D:Anzahl der Datensätze	Kaum möglich	Ja	Ja	Ja	2 Parameter benötigt
CLIQUE	Quadratisch an der Dimensionalität	Ja	Kaum möglich	Teilweise	Ja	2 Parameter benötigt
CURE	$O(n^2)$ für niedrige Dimensionalität $O(n^2 \log n)$ für hohe Dimensionalität n:Anzahl der Proben	Ja	Ja	Ja	Ja	1 Parameter benötigt

5.2. Clusterverfahren mit Anomalie-Detektion

Ziel der Arbeit ist die Identifikation von Anomalien in den PSI-Ergebnissen mittels Clusteranalyse, die zur Ableitung von Bodenbewegungen (Höhenänderungen) dienen. In Kapitel 5.1 sind die typischen räumlichen Clusterverfahren kurz vorgestellt worden, davon haben allerdings nur einige die Fähigkeit, während des Clusterungsvorganges auch Anomalien in den räumlichen Datensätzen zu erkennen. Beispielsweise bieten manche agglomerativen hierarchischen Clusterverfahren (Vorgehensweise s. Kapitel 4.4.1) die Möglichkeit, Beobachtungsobjekte, die sich extrem von allen anderen Objekten unterscheiden, erst am Ende zu fusionieren und als Anomalien auszuweisen, da immer die nächsten Nachbarn mit ähnlichen Ausprägungen zuerst identifiziert und in einem Cluster zusammengelegt werden.

Darüber hinaus sind die Dichte-basierten Clusterverfahren auch in der Lage, räumliche isolierte Objekte zu identifizieren. Daher zählt DBSCAN [ESTER 1996] zu einer der meisten benutzten Algorithmen in den räumlichen Datenanalysen. Im Folgenden wird das häufig angewandte DBSCAN Clusterverfahren kurz erläutert.

Grundidee des DBSCAN Clusterverfahrens ist es, die Objekte anhand ihrer Dichten in Clustern zuzuordnen, wobei die Objektdichte innerhalb eines Clusters deutlich höher als die außerhalb des Clusters sein sollte. Die Objekte, die keinem Cluster zugeteilt werden können, werden als Anomalien erkannt.

Zur Bemessung der Objektdichte sind zwei Eingangsparameter, nämlich ϵ -Umgebung und *MinPts*, erforderlich, die vor der Durchführung des Verfahrens vom Anwender angegeben werden müssen. Die ϵ -Umgebung definiert ein Suchbereich, in dem sich eine Mindestanzahl von Objekten (*MinPts*) befinden müssen, damit die Objekte in einem Cluster zusammengefasst werden können. Je nach der verwendeten Distanz-Metrik variiert die Form der ϵ -Umgebung. Beispielsweise wird die Form bei der Anwendung der üblichen euklidischen Distanz als eine Hypersphäre festgelegt (s. Abb. 37) [ESTER 1996].

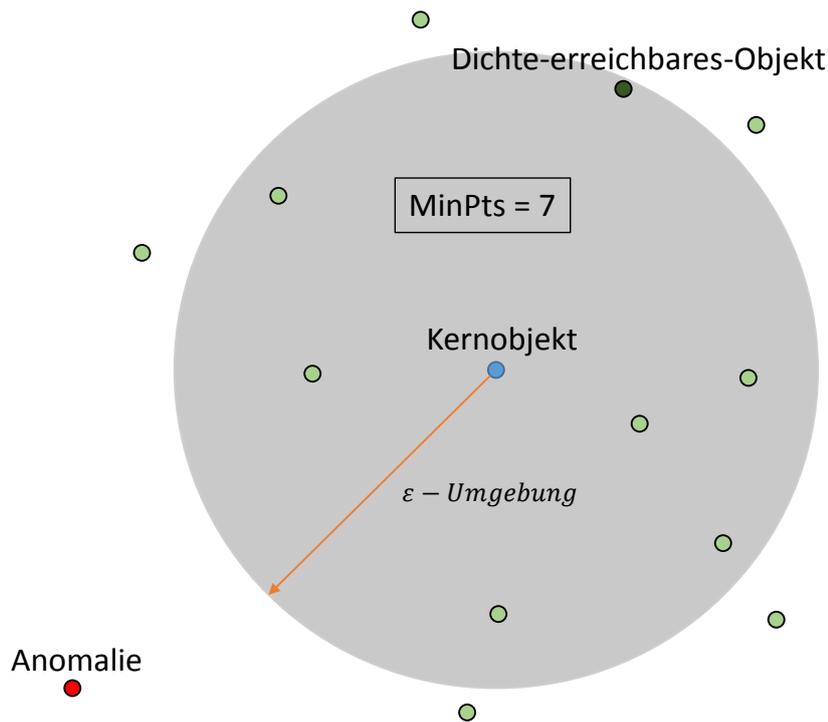


Abbildung 37: Drei Arten von Objekten in DBSCAN

Mathematisch wird die ϵ -Umgebung eines Objektes p , gekennzeichnet als $U_\epsilon(p)$, wie folgt definiert:

Formel 14:
$$U_\epsilon(p) = \{q \in O \mid d(p, q) \leq \epsilon\}$$

mit:

O : Gesamtheit aller Objekte

$d(p, q)$: Distanz zwischen Objekt p und q

Die Objekte lassen sich in drei Arten unterteilen [ESTER 1996]:

- Kernobjekt: Ein Kernobjekt befindet sich im Inneren eines dichten Bereiches; in der ϵ -Umgebung des Kernobjektes befinden sich mindestens n Objekte, wobei n größer als oder gleich $MinPts$ ist. Mathematisch kann ein Objekt p als Kernobjekt angesehen werden, wenn gilt:

Formel 15:
$$|U_\epsilon(p)| \geq MinPts$$

- Dichte-erreichbares-Objekt (Randobjekt): Ein Dichte-erreichbares-Objekt liegt innerhalb der ϵ -Umgebung des Kernobjektes; mathematisch kann ein Objekt q als Dichte-erreichbares-Objekt vom Kernobjekt p bezeichnet werden, wenn gilt:

Formel 16: $q \in U_\epsilon(p) \text{ und } q \neq p$

- Anomalie: ein Objekt a kann als Anomalie bezeichnet werden, wenn es sich nicht in der ϵ -Umgebung irgendeines Kernobjektes p befindet :

Formel 17: $a \notin \forall U_\epsilon(p)$

Ein Kernobjekt $p1$ kann mit einem anderen Kernobjekt $p2$ direkt verbunden werden, dadurch sind die Dichte-erreichbaren-Objekte von $p2$ auch von $p1$ erreichbar. Durch die Kettenverbindung von Kernobjekten und Dichte-erreichbare-Objekten wird somit ein Cluster abgebildet. Alle übrigen Objekte, die keinem Cluster zuzuordnen sind, werden als Anomalien erkannt.

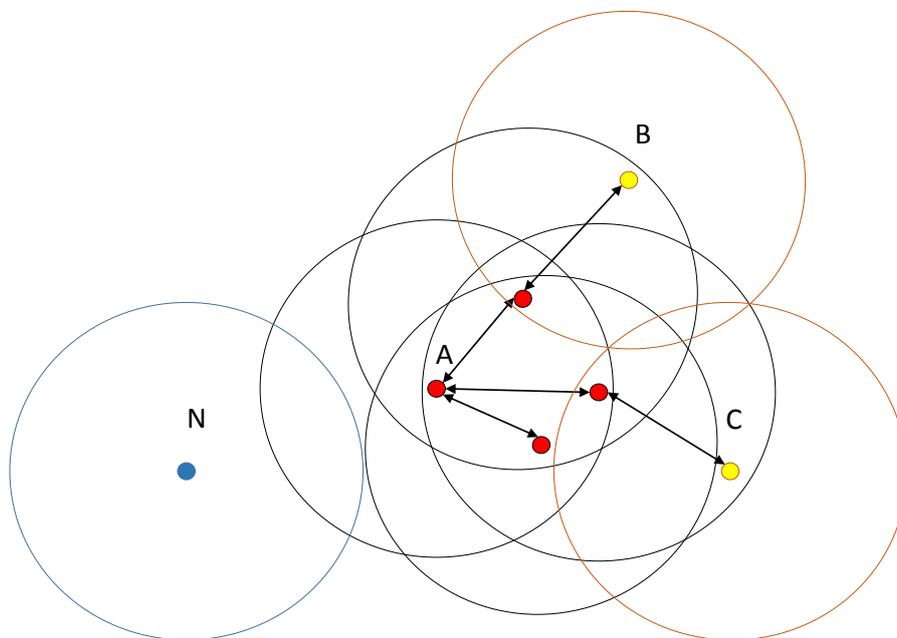


Abbildung 38: Kettenverbindung von Kernobjekten und Dichte-erreichbaren Objekten

Dies ist in Abbildung 38 ersichtlich: Die ϵ -Umgebung eines Objektes ist durch einen farbigen Kreis gekennzeichnet, während die rot markierten Objekte mit $MinPts = 3$ als Kernobjekte zu verstehen sind. Die Objekte B und C sind durch eine direkte Verbindung der Kernobjekte von A mit anderen Kernobjekten erreichbar, somit bilden sie zusammen mit den Kernobjekten ein Cluster. Objekt N ist eine Anomalie, da es weder ein Kern- noch ein Dichte-erreichbares-Objekt ist.

Die durch das DBSCAN identifizierten Anomalien unterscheiden sich jedoch von den in Kapitel 3.3.1 definierten räumlichen Ausreißern. Es handelt sich hier lediglich um die Erkennung von räumlichen isolierten Objekten, wobei nur ihre räumliche Lage bzw. Nachbarschaftsbeziehungen in Betracht gezogen wurden. Die nicht-räumlichen Eigenschaften des Objektes wurden nicht miteinbezogen. Dies gilt auch für die meisten in Kapitel 5.1 genannten Clusterverfahren. Infolgedessen eignen sie sich nicht für die Erkennung von räumlichen bzw. spatio-temporalen Anomalien in flächenhaft vorliegenden PSI-Ergebnissen, da nicht nur die räumlichen Beziehungen zwischen den Beobachtungsobjekten sondern auch die raumzeitlichen Abhängigkeiten (Autokorrelationen) der nicht-räumlichen Eigenschaften des Objektes gleichzeitig berücksichtigt werden müssten.

Basierend auf der räumlichen Autokorrelation bietet jedoch die Local-Moran's-Index - Clusteranalyse (LISA) [ANSELIN 1995] die Möglichkeit, derartige Anomalien zu identifizieren. Nachfolgend wird das Grundprinzip der Local-Moran's-Index - Clusteranalyse vorgestellt und ihre Anwendbarkeit für die Anomalieerkennung in PSI-Daten anhand praktischer Beispiele dargestellt.

5.3. Prinzip des Local-Moran's-Index

5.3.1. Räumliche Autokorrelation

Aus der Zeitreihenanalyse ist bekannt, dass die zeitlichen benachbarten Werte einer beobachteten Variable untereinander korreliert sind, d. h. die Werte einer Variable in Periode t haben eine zeitliche Regelmäßigkeit mit den Werten derselben Variable in den Vorperioden $t-1$, $t-2, \dots$ und die Reihenfolge der Beobachtungen unterliegt einer logischen Ordnung. Die zeitliche Autokorrelation bedeutet, dass die Werte derselben beobachteten Variable zeitlich mit sich selbst korreliert sind. Dies führt dazu, dass die zeitliche Abhängigkeit der Werte in der Zeitreihe durch eine stochastische Funktion mit einer Wahrscheinlichkeitsaussage eingeschätzt werden kann [WINKER 2010]. Analog hierzu zeichnen sich die beobachteten

räumlichen Variablen durch räumliche Autokorrelation aus, die sich ebenfalls durch stochastische Methoden modellieren lassen. Die räumliche Autokorrelation (Abhängigkeit) besteht in den meisten Fällen in mehreren Richtungen.

Die räumliche Autokorrelation wurde von Tobler's erstem Gesetz der Geographie, nämlich „Everything is related to everything else, but near things are more related than distant things“ beschrieben [TOBLER 1970]. Das bedeutet, dass zwischen den beobachteten Variablen der räumlichen Objekte distanzabhängige Relationen bestehen: je näher sie zueinander liegen, desto ähnlicher sind die Werte der beobachteten Variablen der Objekte. In der Praxis bedeutet die Existenz der räumlichen Autokorrelation folgendes: Wenn A und B nahe beieinander sind, bezieht sich das, was bei A passiert, auf das, was bei B geschieht, und umgekehrt [ABLER 1992]. Unter diesen Gesichtspunkten kann räumliche Autokorrelation als eine Erweiterung der klassischen Korrelation um den räumlichen Kontext angesehen werden, wobei „auto“ darauf hindeutet, dass Beobachtungen sich auf ein und dieselbe Variable, nicht aber auf zwei unterschiedliche Variablen beziehen. Griffith [GRIFFITH 1991] definiert die räumliche Autokorrelation als „...the relationship among values of a single variable that is attributable to the geographic arrangement ...“ [PICHLER 1998].

Beim Vorhandensein der räumlichen Autokorrelation sind die Ähnlichkeiten oder die Unterschiede der von einer bestimmten räumlichen Distanz getrennten Beobachtungen größer, als dies bei zufälligen Prozessen zu vermuten wäre [LEGENDRE 1993]. Diese Eigenschaft führt in der Folge dazu, dass traditionelle statistische Methoden, bei denen von der Grundannahme ausgegangen wird, dass jede Beobachtung zufällig und unabhängig von anderen Beobachtungen ist, für die Analyse von räumlichen Datensätzen nicht geeignet sind, da Toblers erstes Gesetz der Geographie missachtet und die räumliche Autokorrelation vernachlässigt wird [CLIFF 1973] [ANSELIN 1989]. Deshalb gewinnt die räumliche Datenanalyse (spatial data analysis) in den letzten Jahren immer mehr an Bedeutung, da sie versucht, Effekte, Beziehungen und Auswirkungen zwischen räumlichen Objekten zu berücksichtigen und zu ermitteln.

Die räumliche Autokorrelation kann als eines der wichtigsten Bewertungskriterien für raumbezogene Fragenstellungen betrachtet werden, da ihre Ermittlung eine notwendige Vorbedingung und Grundlage für eine Vielzahl von räumlichen statistischen Verfahren und Analysen ist.

Bei der räumlichen Autokorrelation sind drei Typen zu unterscheiden:

- Positive räumliche Autokorrelation: Wenn Objekte, die sich in räumlicher Nähe zueinander befinden, ähnlichere Werte derselben Variablen besitzen als weiter voneinander entfernte Objekte.

- Negative räumliche Autokorrelation: Wenn Objekte, die sich in räumlicher Nähe zueinander befinden, unähnlichere Werte derselben Variablen besitzen.
- Keine räumliche Autokorrelation: Wenn die beobachtete Variable von Objekten nicht ortsabhängig ist (in Bezug auf die beobachtete Variable verteilen sich die Objekte zufällig im Untersuchungsraum)

In Abbildung 39 sind diese drei Typen von räumlichen Autokorrelation vereinfacht dargestellt.

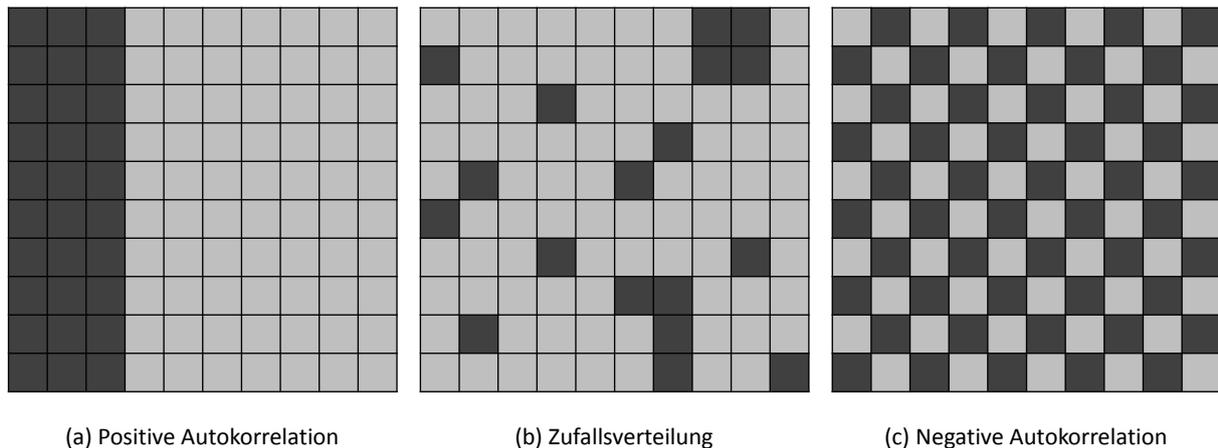


Abbildung 39: Typen räumlicher Autokorrelation [VERÄNDERT NACH LO 2007]

Mit Hilfe der Autokorrelationskoeffizienten ist es möglich, die Stärke der räumlichen Abhängigkeit einer beobachteten Variablen zwischen den Objekten abzuschätzen. Dabei kann man zwischen dem globalen Autokorrelationskoeffizienten und dem lokalem Autokorrelationskoeffizienten unterscheiden. Anhand des globalen Autokorrelationskoeffizienten kann aufgezeigt werden, ob eine räumliche Autokorrelation in dem gesamten Untersuchungsgebiet (in der Gesamtheit der Untersuchungsobjekte) besteht [CLIFF 1973], während mittels des lokalen Autokorrelationskoeffizienten die räumliche Autokorrelation für jedes Untersuchungsobjekt untersucht werden kann [GETIS 1992] [ANSELIN 1995].

5.3.2. Globale räumliche Autokorrelation

Die Untersuchung der globalen räumlichen Autokorrelation nahm ihren Anfang in den 40er Jahren des 20. Jahrhunderts. Seitdem wurden verschiedene globale Autokorrelationsmaße vorgestellt. Hiervon sind Maße Geary's c [GEARY 1954], Moran's-I [MORAN 1948] und General G-Statistik [GETIS 1992] in der geographischen Forschung besonders verbreitet und in vielen

raumbezogenen Fragestellungen, z.B. der Epidemie-Forschung und soziodemographischen Untersuchungen, eingesetzt worden sind [GOODCHILD 1986][GRIFFITH 1987]. Im Folgenden werden sie anhand mathematischer Formeln vorgestellt.

- Geary's c

Die Korrelation wird nach Geary's c mit Hilfe des Koeffizienten c wie folgt definiert [CLIFF 1973] [CLIFF 1981]:

Formel 18:

$$c = \frac{n - 1}{2 \sum_{i=1}^n \sum_{j=1, j \neq i}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1, j \neq i}^n w_{ij} (x_i - x_j)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

mit:

x_i : Wert der beobachteten Variable vom Objekt i

x_j : Wert der beobachteten Variable vom Objekt j

\bar{x} : Mittelwert der beobachteten Variable aller Objekte,

w_{ij} : räumliche Gewichtung zwischen Objekt i und j

n : Anzahl der Objekte

Wenn die benachbarten Objekte ähnliche Werte bezüglich der betrachteten Variablen besitzen, führt dies zu einem kleinen Wert des Koeffizienten. Umgekehrt erzeugen große Unterschiede in den Werten einen großen Wert des Koeffizienten c . Der Wert c besitzt einen Wertebereich zwischen 0 und 2. Liegt eine hohe räumliche Korrelation vor, nimmt der Koeffizient den Wert 0 an. Je höher der Variablenwert, desto niedriger ist die räumliche Korrelation; ein Wert um 1 bedeutet eine zufällige Verteilung der Variable in dem gesamten Untersuchungsraum und es gibt keine räumliche Abhängigkeit zwischen den räumlich benachbarten Objekten. Wenn sich c dem Wert 2 annähert, deutet dies auf eine negative räumliche Autokorrelation hin [CLIFF 1973][CLIFF 1981].

Am häufigsten wird das globale Moran's-I Maß zur Bewertung der globalen räumlichen Autokorrelation eingesetzt. In Analogie zu dem Koeffizient nach Geary weist der Moran's-I eine ähnliche Algorithmus-Struktur auf, jedoch wird statt der quadrierten Differenz zwischen den Variablenwerten von benachbarten Objekten die Kovarianz (das Produkt der Abweichungen der Werte vom Mittelwert) berechnet, die die Abhängigkeit zweier Variablenwerte an den Orten i und j modelliert [LINDER 2013].

Der globale Moran's-Index berechnet sich nach folgender Formel [CLIFF 1973][CLIFF 1981]:

Formel 19:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Der Wertebereich des Parameters I liegt zwischen -1 und 1. Der Wert 1 bedeutet eine extrem positive Autokorrelation, d.h. die Variablenwerte benachbarter Objekte sind sehr eng miteinander korreliert und es bildet sich ein homogenes Cluster. Nähert sich der Wert 0, liegt eine zufällige Verteilung der Variablenwerte und somit keine räumliche Autokorrelation vor. Folglich weist ein Wert von -1 auf eine extrem negative räumliche Autokorrelation hin [CLIFF 1973][CLIFF 1981].

Unter Zuhilfenahme des globalen Moran's-Index kann festgestellt werden, ob grundsätzlich im gesamten Untersuchungsraum eine räumliche Autokorrelation vorliegt und darauf aufbauend sich Cluster bilden. Jedoch besitzt er nicht die Fähigkeit festzulegen, um welche Art von Cluster es sich handelt.

Diese Lücke wird durch die General G-Statistik [GETIS 1992] gefüllt, indem im Falle von positiver Autokorrelation die Cluster mit hohen und niedrigen Werten unterschieden werden.

Die General G-Statistik wird durch den Koeffizienten G gemäß Formel 20 beschrieben:

Formel 20:

$$G = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} x_i x_j}{\sum_{i=1}^n \sum_{j=1}^n x_i x_j}$$

5.3.3. Lokale räumliche Autokorrelation

Anhand der im vorigen Kapitel vorgestellten, globalen räumlichen Autokorrelationskoeffizienten kann festgestellt werden, ob und in welchem Ausmaß eine räumliche Autokorrelation im Untersuchungsraum vorhanden ist. Es ist hingegen nicht möglich, durch einen einzigen Wert dieser Koeffizienten aufzuzeigen, ob in Teilgebieten eine stärkere Autokorrelation als in anderen Teilgebieten vorliegt. Bei lokaler Heterogenität wird die räumliche Stationarität verletzt, die als Grundvoraussetzung für die Bestimmung von globalen räumlichen Autokorrelationen gilt, so dass es unmöglich ist, besonders in einem großen Untersuchungsraum, die räumliche Variabilität in Datensätzen aufzudecken, wenn sich Daten deutlich vom globalen Trend unterscheiden und somit unter Umständen verdeckt sind [BAILEY 1995]. Durch den globalen räumlichen Autokorrelationskoeffizienten ist die räumliche

Information eher generalisiert, was zu einem unerwünschtem Ergebnis führen kann, wenn an solchen Teilgebieten ein besonderes Forschungsinteresse besteht [CRESSIE 1993]. Dies kann aber mithilfe einer lokalen räumlichen Autokorrelation untersucht werden, indem für jedes Objekt in dem Untersuchungsraum ein entsprechender individueller Koeffizient berechnet wird. Er gestattet die Identifikation von Teilgebiet-Clustern (bei positiver Autokorrelation) oder Anomalien (bei negativer Autokorrelation) [ANSELIN 1995].

Da die Bestimmung von lokaler räumlicher Autokorrelation deutlich aufwändiger als die von globaler räumlicher Autokorrelation ist und viel Rechenkapazität benötigt, wurden verschiedene Verfahren erst Ende der 80er Jahre mit fortschreitender Verbesserung der elektronischen Datenverarbeitung entwickelt [GETIS 1992] [ORD 1995] [ANSELIN 1995] [PICHLER 1998]. Davon sind der Getis-Ord G_i^* Index und der Local-Moran's-Index besonders erwähnenswert, da sie in vielen Gebieten, z.B. den Geowissenschaften, Sozialwissenschaften sowie Wirtschaftswissenschaften, ihre Anwendungen finden können.

- Getis-Ord G_i^* Index

Die G_i^* Index lässt sich wie folgt berechnen [GETIS 1992]:

Formel 21:

$$G_i^* = \frac{\sum_{j=1}^n w_{ij} x_j}{\sum_{j=1}^n x_j}$$

mit:

x_j : Wert der beobachteten Variable vom Objekt j , $j \neq i$

w_{ij} : räumliche Gewichtung zwischen Objekt i und j

Mithilfe des G_i^* Index werden die gewichteten, benachbarten Variablenwerte in einem lokalen Teilgebiet vom Untersuchungsobjekt i mit dem globalen Durchschnitt verglichen, indem die Summe der Variablenwerte des Teilgebietes in einer durch die Gewichtungsmatrix w_{ij} definierten Nachbarschaft durch die Summe der Variablenwerte aller anderen Objekte in dem gesamten Untersuchungsraum dividiert wird. Der Index gibt somit den Anteil der Summe der Variablenwerte an den benachbarten Objekten an, die sich in dem Teilgebiet des Untersuchungsobjektes i befinden.

Wenn ein Objekt von Objekten mit überdurchschnittlichen Variablenwerten umgeben ist, nimmt der G_i^* Index einen relativ hohen Wert an, welcher ein potenzielles Cluster mit hohen Variablenwerten („hot-spot“) andeutet. Andererseits wird ein potenzielles Cluster mit niedrigen

Variablenwerten („cold-spots“) durch einen unterdurchschnittlichen G_i^* Indexwert angedeutet [GETIS 1992].

- Local-Moran's-Index

Das Local-Moran's-Index - Verfahren, das zu einem der am häufigsten verwendeten Verfahren zur Bestimmung der lokalen räumlichen Autokorrelation gehört, wurde von Luc Anselin im Jahr 1988 entwickelt [ANSELIN 1988]. Basierend auf Nachbarschaftsverhältnissen zwischen den gemessenen Untersuchungsobjekten dient es der Erkennung von Clustern und Anomalien. Dadurch können nicht nur Bereiche von Objekten mit ähnlichen Variablenwerten als lokale Cluster abgegrenzt sondern auch Objekte mit auffällig andersartigen Variablenwerten als Anomalien identifiziert werden. Zur Berechnung des Local-Moran's-Index dient Formel 22 [ANSELIN 1995]:

Formel 22:
$$I_i = \frac{(x_i - \bar{x}) \sum_{j=1}^n w_{ij} (x_j - \bar{x})}{S_i^2}$$

in der S gemäß Formel 23 bestimmt wird:

Formel 23:
$$S_i^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

mit

x_i : Wert der beobachteten Variable des Objektes i ,

x_j : Wert der beobachteten Variable des Objektes j (j ist der Nachbar von i),

\bar{x} : Mittelwert der beobachteten Variable aller Objekte,

w_{ij} : räumliche Gewichtung zwischen Objekt i und j ,

S_i^2 : Varianz des Variablenwertes

Der Zähler der Formel 22 beinhaltet das Produkt aus der Abweichung des Variablenwertes des Untersuchungsobjektes i vom Mittelwert und der Summe aller n gewichteten

Abweichungen der benachbarten Objekte j vom Mittelwert. Der Nenner entspricht der Varianz der Variablen aller n Objekte in dem gesamten Untersuchungsraum. Da die Varianz der Variablen (S_i^2) für ein bestimmtes Untersuchungsgebiet als positive Konstante angesehen werden kann, besitzt das Produkt $((x_i - \bar{x}) \sum_{j=1}^n w_{ij}(x_j - \bar{x}))$ in der Gleichung eine entscheidende Rolle für die Bestimmung der lokalen räumlichen Autokorrelation. Drei Fälle sind zu unterscheiden:

1. Das Produkt besitzt einen positiven Wert: Dieser Fall liegt vor, wenn das Untersuchungsobjekt i und die benachbarten Objekte gleichzeitig überdurchschnittliche oder unterdurchschnittliche Variablenwerte aufweisen.
2. Das Produkt besitzt einen negativen Wert: Dies ergibt sich, wenn das Untersuchungsobjekt i und überdurchschnittlich viele benachbarte Objekte von i abweichende Variablenwerte vom Mittelwert besitzen.
3. Das Produkt hat einen Wert nahe bei null: Dies tritt ein, wenn der Variablenwert des Untersuchungsobjektes i oder/und das gewichtete arithmetische Mittel aus den Variablenwerten benachbarter Objekte von i dem Mittelwert entspricht.

Der Local-Moran's-Index I_i ist normiert, da das Produkt durch die Varianz dividiert wird. Da die Varianz immer einen positiven Wert besitzt, erhält der Local-Moran's-Index immer das gleiche Vorzeichen wie das Produkt $((x_i - \bar{x}) \sum_{j=1}^n w_{ij}(x_j - \bar{x}))$. Positive Index-Werte entsprechen einer positiven räumlichen Autokorrelation, d.h. das Untersuchungsobjekt hat einen ähnlichen Variablenwert wie der seiner benachbarten Objekte. Somit bildet sich ein Cluster mit hohen oder niedrigen Variablenwerten. Da sie positiv miteinander autokorreliert sind, ist das betrachtete Objekt i Teil eines Clusters. Negative Index-Werte deuten an, dass es eine negative räumliche Autokorrelation in dem Lokalebereich gibt und das Untersuchungsobjekt einen abweichenden Variablenwert im Vergleich zu seinen benachbarten Objekten besitzt; somit ist das Objekt in diesem Sinne eine Anomalie. Wenn der Index-Wert in der Nähe von Null liegt, bedeutet dies, dass es keine räumliche Autokorrelation gibt und sich die Variablenwerte im lokalen Bereich zufällig verteilen [ANSELIN 1995].

Somit kann der Local-Moran's-Index durch folgende Eigenschaften charakterisiert werden [ANSELIN 1995]:

- Für jedes Untersuchungsobjekt ergibt sich ein Local-Moran's-Index, mit dem festgestellt werden kann, ob das untersuchte Objekt Teil eines lokalen Clusters oder eine lokale Anomalie im Vergleich zu seiner Nachbarschaft ist.
- Die Summe aller Local-Moran's-Indizes der Untersuchungsobjekte ist proportional zum globalen Moran's-Index.

Die Zielsetzungen des Getis-Ord G_i^* Index und des Local-Moran's-Index zur Bestimmung der lokalen räumlichen Autokorrelation sind leicht unterschiedlich. Während der Getis-Ord G_i^* Index Auskunft über das Ausmaß der räumlichen Cluster sowohl mit besonders hohen als auch mit besonders niedrigen Variablenwerten gibt, können, bezogen auf die Variablenwerte der Untersuchungsobjekte, mittels dem Local-Moran's-Index nicht nur lokale Cluster gebildet sondern auch lokale Anomalien identifiziert werden.

Bei der radarinterferometrischen Methode der Persistent Scatterer Interferometrie (PSI) können aufgrund verschiedener Ursachen die ermittelten PSI-Höhenänderungszeitreihen an einzelnen PS-Punkten von dem erwarteten flächenhaften Bodenbewegungsverhalten abweichen. Solche PS-Punkte können als „räumliche Anomalien“ angesehen werden, deren Identifizierung durch viele räumliche Clusterverfahren nicht möglich ist, da sie bei dem Clusterungsvorgang lediglich die räumliche Lage der untersuchten Objekte berücksichtigen. Da eine räumliche Autokorrelation bei flächenhaften Bodenbewegungen in lokalen Bereichen besteht, kann jedoch der Local-Moran's-Index, der sich mit lokaler räumlicher Autokorrelation auseinandersetzt, zur Detektion von solchen Anomalien eingesetzt werden. Dieser praktikable Ansatz zur automatisierten Detektion von Anomalien in massenhaft vorhandenen PS-Punkten wird im Rahmen dieser Arbeit an praktischen Beispielen getestet und untersucht.

5.3.4. Signifikanzprüfung

In diesem Kapitel soll untersucht werden, wie geprüft werden kann, ob die, mit den verschiedenen Koeffizienten, festgestellten Ergebnisse statistisch signifikant sind. Hierzu soll eine statistische Signifikanzprüfung erfolgen. Da die Vorgehensweise der Signifikanzprüfung mehr oder wenig ähnlich erfolgt, wird im Folgenden beispielhaft die Signifikanzprüfung für den Local-Moran's-Index erläutert.

Die Signifikanzprüfung des Local-Moran's-Index - Wertes basiert auf dem Nullhypothesentest, wobei als Nullhypothese „keine lokale räumliche Autokorrelation“ gegen die Alternativhypothese „es gibt lokale räumliche Autokorrelation“ aufgestellt wird. Ob die Nullhypothese abgelehnt werden kann oder nicht, hängt von der angewendeten Teststatistik und der vordefinierten Irrtumswahrscheinlichkeit ab. Die Nullhypothese kann verworfen werden, wenn die beobachteten Werte aus der Teststatistik in das Konfidenzintervall fallen, dessen Ober- und Untergrenze unter Annahme eines vordefinierten Signifikanzniveaus festgelegt wird.

Die Teststatistik zur Signifikanzprüfung der durch den Local-Moran's-Index - Wert identifizierten Cluster oder Anomalien kann entweder durch die Gegenüberstellung mit einer Zufallsverteilung oder durch Z-Statistik mit der Annahme einer Normalverteilung ausgeführt werden [FISCHER 2010]. Die Zufallsverteilung erfolgt normalerweise durch eine Monte-Carlo-Simulation. Dabei werden die Variablenwerte der Untersuchungsobjekte zufällig über das gesamte Untersuchungsgebiet neu verteilt und basierend auf den dadurch generierten Erwartungswerten erfolgt eine Neuberechnung des Local-Moran's-Index. Durch Vergleich der Ergebnisse kann mithilfe eines vordefinierten Signifikanzniveaus aufgezeigt werden, ob die Nullhypothese abgelehnt werden kann. Je öfter die Permutation wiederholt wird, desto geringer ist die Irrtumswahrscheinlichkeit.

Bei der Z-Statistik sind die Z-Ergebnisse und p-Werte für die Signifikanzprüfung zu berechnen. Der p-Wert gibt an, mit welcher Wahrscheinlichkeit die beobachtete räumliche Beziehung der Untersuchungsobjekte einem zufälligen Prozess entstammt, während die Z-Ergebnisse die Standardabweichungen darstellen.

Das Z-Ergebnis wird gemäß der nachfolgenden Gleichung berechnet [ANSELIN 1995]:

Formel 24:

$$Z_{I_i} = \frac{I_i - E[I_i]}{\sqrt{Var[I_i]}}$$

mit:

$E[I_i]$: Erwartungswert von I_i gemäß Formel 25:

Formel 25:

$$E[I_i] = - \frac{\sum_{j=1, j \neq i}^n w_{ij}}{n - 1}$$

mit

$Var[I_i]$: Varianz von I_i , die mit Formel 26 berechnet wird:

Formel 26:

$$Var[I_i] = E[I_i^2] - E[I_i]^2$$

Der Erwartungswert E ergibt sich gemäß Formel 27:

Formel 27:

$$E[I_i^2] = \frac{(n - b_{2i}) \sum_{j=1, j \neq i}^n w_{ij}^2}{n - 1} + \frac{(2b_{2i} - n) \sum_{k=1, k \neq i}^n \sum_{h=1, h \neq i}^n w_{ik} w_{ih}}{(n - 1)(n - 2)}$$

und dem Parameter b gemäß Formel 28:

Formel 28:

$$b_{2i} = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{(S_i^2)^2}$$

Der Zusammenhang zwischen den Z-Ergebnissen und den P-Werten lässt sich in der Normalverteilung veranschaulichen (s. Abb. 40):

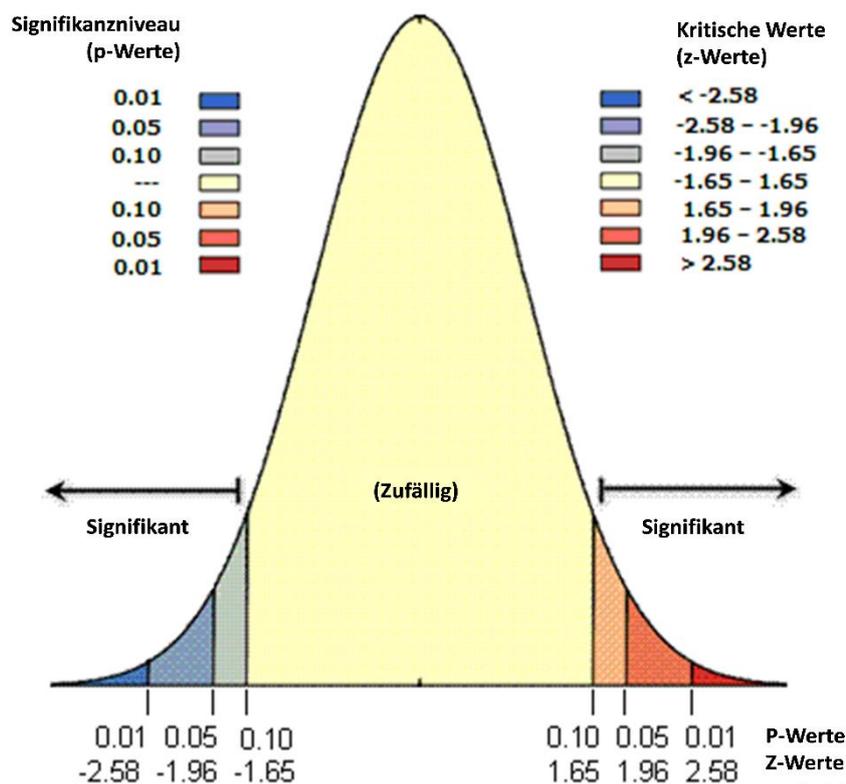


Abbildung 40: Signifikanzniveau gemäß Z-Ergebnis und p-Wert im Nullhypotesentest [VERÄNDERT NACH ESRI ONLINE]

In der Abbildung 40 ist dargestellt, wie das Signifikanzniveau mithilfe der Z- und p-Werte festgelegt wird. Je größer der absolute Betrag des Z-Ergebnisses ist, desto kleiner ist der p-Wert, und somit ist es wahrscheinlicher, die Nullhypothese abzulehnen und das Ergebnis als signifikant zu bewerten [CHESTER 1987].

In der Tabelle 11 sind die p-Werte und Z-Ergebnisse für unterschiedliche Signifikanzniveaus gegenübergestellt:

Tabelle 11: Signifikanzniveaus und entsprechende kritische Z-Ergebnisse und p-Werte im Nullhypothesentest [VERÄNDERT NACH ESRI ONLINE]

Signifikanzniveau	p-Wert	Z-Ergebnis
10%	< 0,10	< -1,65 oder > +1,65
5%	< 0,05	< -1,96 oder > +1,96
1%	< 0,01	< -2,58 oder > +2,58

Bei der Berechnung des Local-Moran's-Index ist jedoch zu beachten, dass die räumlich benachbarten Objekte über gemeinsame Nachbarn verfügen, die somit mehrfach in der Berechnung ihrer Index-Werte auftauchen; damit sind jedoch die Index-Werte miteinander korreliert und dadurch ist die Signifikanz solcher Werte überbewertet. Um dieses Problem zu umgehen, ist es oft empfehlenswert, durch die Anwendung einer Bonferroni-Korrektur die Irrtumswahrscheinlichkeit um die Anzahl der Untersuchungsobjekte zu reduzieren [ORD 1995]. Die korrigierte Irrtumswahrscheinlichkeit α^* lässt sich gemäß Formel 29 berechnen:

Formel 29:
$$\alpha^* = \frac{\alpha}{n}$$

mit:

α : ursprüngliche Irrtumswahrscheinlichkeit,

n : Anzahl der Untersuchungsobjekte.

Nach der Signifikanzprüfung werden die betrachteten Objekte klassifiziert. Bei positiver signifikanter Autokorrelation ($I_i > 0$) gehört das Objekt zu einem Cluster, während bei negativer signifikanter Autokorrelation ($I_i < 0$) das Objekt als „Anomalie“ bewertet wird.

5.3.5. Bestimmung/Bildung von Nachbarschaft

Wie in Kapitel 4.3 erwähnt, ist es bei der Durchführung der Clusteranalyse von großer Relevanz, die Ähnlichkeitsmaße zwischen den Untersuchungsobjekten zu ermitteln. Ein wesentlicher Unterschied zwischen räumlichen und herkömmlichen, lageunabhängigen Clusterverfahren besteht darin, dass bei den räumlichen Clusterverfahren die räumlichen Beziehungen direkt im Algorithmus integriert sind. Dies drückt sich bei der Berechnung von räumlichen Autokorrelationskoeffizienten besonders durch die Bestimmung von räumlichen Nachbarschaften und einer Gewichtungsmatrix aus. Die räumliche Nachbarschaftsmatrix w_{ij} spielt dabei eine wichtige Rolle, denn sie sollte möglichst die tatsächlichen räumlichen Verhältnisse und Abhängigkeiten widerspiegeln.

Die Nachbarschaft zwischen den räumlichen Objekten ist abhängig von dem beobachteten Phänomen sowie dem dahinter verborgenen Erprobungsverfahren (z.B. PS-Punkte). Es können durch die Abbildung der Nachbarschaft unterschiedliche räumliche Autokorrelationen in verschiedenen Richtungen angezeigt werden, so dass es nicht möglich ist, ein allgemeingültiges Kriterium für die Nachbarschaftsbildung zu definieren. Vielmehr sollte dies je nach Aufgabestellung dem individuellen räumlichen Prozess angepasst werden. Dazu lässt sich die Bestimmung der räumlichen Nachbarschaft durch folgende Aspekte charakterisieren:

- **Anzahl der Nachbarn:** Die räumliche Nachbarschaft kann durch eine Anzahl von Nachbarn definiert werden, so dass für das Untersuchungsobjekt nur die angegebene Anzahl seiner nächsten Nachbarn innerhalb des räumlichen Kontextes zur Clusteranalyse berücksichtigt wird. Dadurch variiert das räumliche Ausmaß der Nachbarschaft hinsichtlich der Objektdichte. In Bereichen mit hoher Objektdichte ist das räumliche Ausmaß der Nachbarschaft kleiner als in Bereichen, in denen eine niedrige Objektdichte besteht. Damit wird sichergestellt, dass sich für jedes Untersuchungsobjekt Nachbarn ergeben, unabhängig davon, ob sich die Objekte gleichmäßig über das gesamte Untersuchungsgebiet verteilen. Es gibt für viele räumliche Clusterverfahren keine Obergrenze für der Anzahl der Nachbarn. Jedoch ist zu beachten, dass, je mehr Nachbarn für die Berechnung der Gewichtungsmatrix herangezogen werden, desto höher sind die benötigte Rechenkapazität und der Rechenaufwand. Die Wahl einer angemessenen Anzahl von Nachbarn für große räumliche Datensätze ist daher, außer den räumlichen Verhältnissen, immer ein Kompromiss zwischen der Genauigkeit der Ergebnisse und dem Rechenaufwand (bzw. der Rechenzeit). Es darf aber andererseits für viele räumliche Clusterverfahren eine bestimmte Untergrenze für die Anzahl der Nachbarn nicht unterschritten werden. Ein erfolgreicher Clusterungsvorgang setzt voraus, dass eine genügende Anzahl von

Nachbarn in der Nähe von dem Untersuchungsobjekt gewährleistet wird und damit die lokalen räumlichen Zusammenhänge durch die begrenzte Anzahl von Objekten möglichst realitätsnah repräsentiert werden können.

Beispielsweise haben Getis und Ord [GETIS 1992] eine Mindestanzahl von acht Nachbarn für das Local-Moran's-Index - Clusterverfahren zur Analyse von polygonbasierten Datensätzen vorgeschlagen. Dies entspricht dem Queen's Case [LLOYD 2010] (s. Abb. 41), wobei die Objekte, die eine gemeinsame Kante oder Ecke mit dem Untersuchungsobjekt haben, als Nachbarn angesehen werden. Darüber hinaus existiert noch ein Rook's Case [LLOYD 2010]. In Anlehnung an den Queen's Case beinhaltet der Rook's Case nur die Objekte, die eine gemeinsame Kante mit dem Untersuchungsobjekt besitzen, und somit zur Nachbarschaft eines Untersuchungsobjekt zählen dürfen (s. Abb. 41).

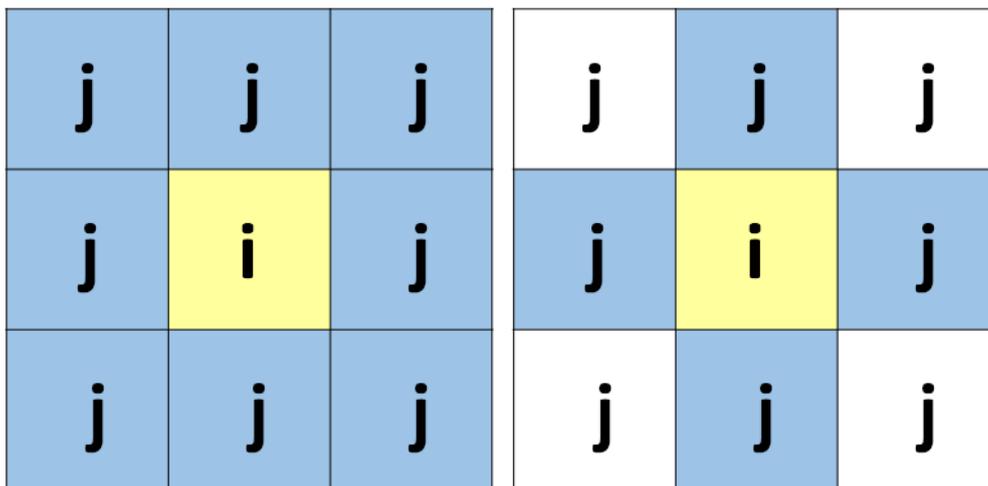


Abbildung 41: Queen's Case (links) und Rook's Case (rechts)

- **Ausmaß der Nachbarschaft:** Die räumlichen Abhängigkeitsbeziehungen können nicht nur Objekte in der nahen Umgebung betreffen, sondern auch über mehrere räumliche Objekte in einem bestimmten Bereich hinweg existent sein. Dies kann üblicherweise durch eine Raumschrittweite (z.B. für ein topologisches Verbindungsnetz) oder eine Distanzfunktion (z.B. für punktbezogene Daten) angegeben werden. Im Vergleich zur eindimensionalen Zeitschrittweite, die eindeutig durch Jahr, Monat oder Tag definiert werden kann, besitzt die Raumschrittweite wegen der höheren Dimensionalität keinen einheitlichen Maßstab [BAHRENBERG 2003]. Daher kann die Raumschrittweite nur in Abhängigkeit von dem ausgewählten räumlichen Modell aus der (topologischen) Nachbarschaftsbeziehung der Objekte abgeleitet werden.

Bei einem TIN-Modell von Objekten, das durch Delaunay-Triangulation erstellt wurde, kennzeichnet eine Raumschrittweite von 1 die unmittelbar durch Dreieckskanten mit dem Untersuchungsobjekt (Eckpunkt) verbundenen Objekte (direkte Nachbarn) und eine Raumschrittweite 2 die durch Dreieckskanten mit den direkten Nachbarn des Untersuchungsobjektes verbundenen Objekte (s. Abb. 42).

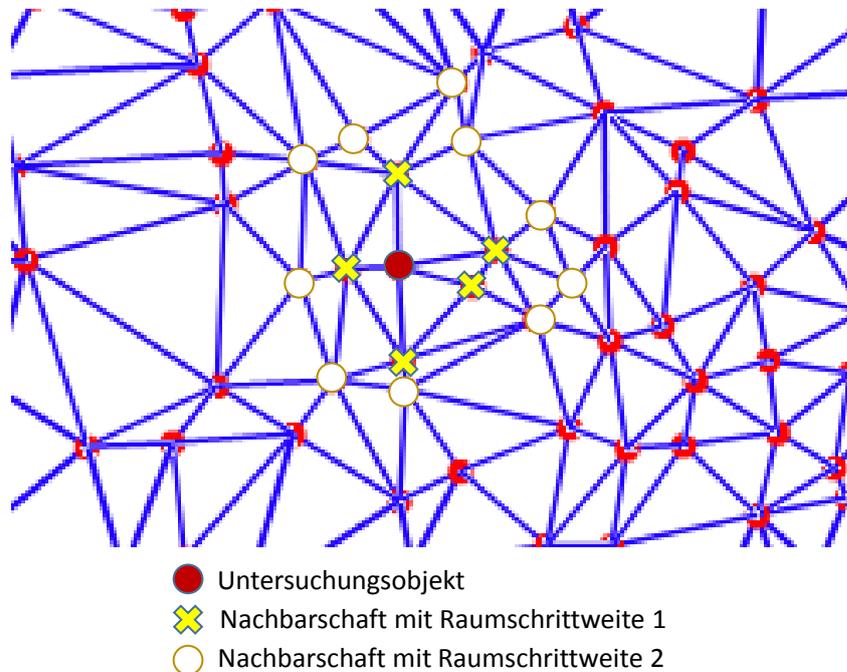


Abbildung 42: Nachbarschaft im TIN-Modell mit Raumschrittweiten 1 und 2

In ähnlicher Weise kann die Raumschrittweite in einem Rastermodell darauf hinweisen, dass die Objekte, die die Nachbarschaft vom Untersuchungsobjekt (Zelle) bilden, sich eine gemeinsame direkte/indirekte Kante (bei Rook's Case) oder Kante/Ecke (bei Queen's Case) mit dem Untersuchungsobjekt teilen (s. Abb. 43).

Mit den durch die Raumschrittweite definierten Nachbarschaften kann sichergestellt werden, dass jedes Untersuchungsobjekt mindestens einen Nachbarn besitzt, unabhängig davon, ob es Datenlücken oder stark variierende Objektdichten im Datensatz gibt.

Außer durch Raumschrittweiten kann das Maß von Nachbarschaften durch eine feste Distanz definiert werden (s. Abb. 44). Die Objekte, die innerhalb einer Distanz um das Untersuchungsobjekt liegen, können als Nachbarn festgelegt werden. Damit wird ein einheitlicher Analysemaßstab für alle Untersuchungsobjekte festgelegt. Jedoch könnte es dadurch möglich sein, dass für manche Untersuchungsobjekte überhaupt keine Nachbarn innerhalb einer vordefinierten Distanz gefunden werden.

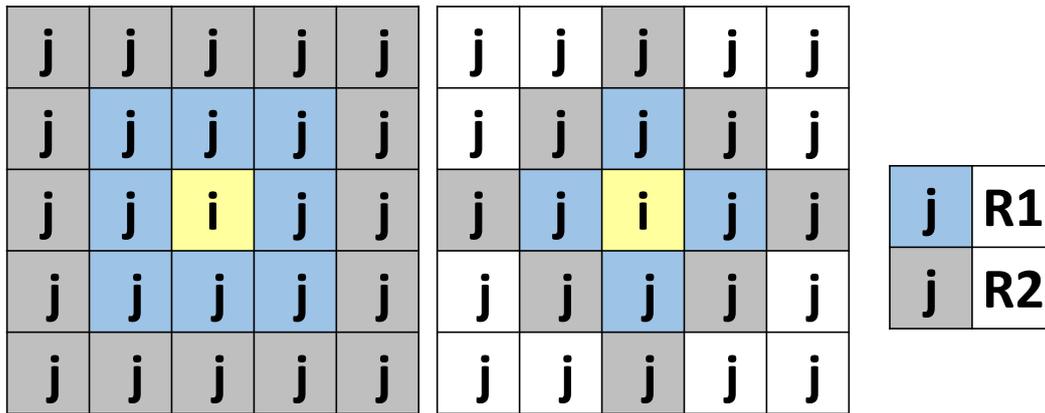


Abbildung 43: Nachbarschaft im Rastermodell mit Raumschrittweiten 1 (R1) und 2 (R2) (links: Queen's Case Rechts: Rook's Case)

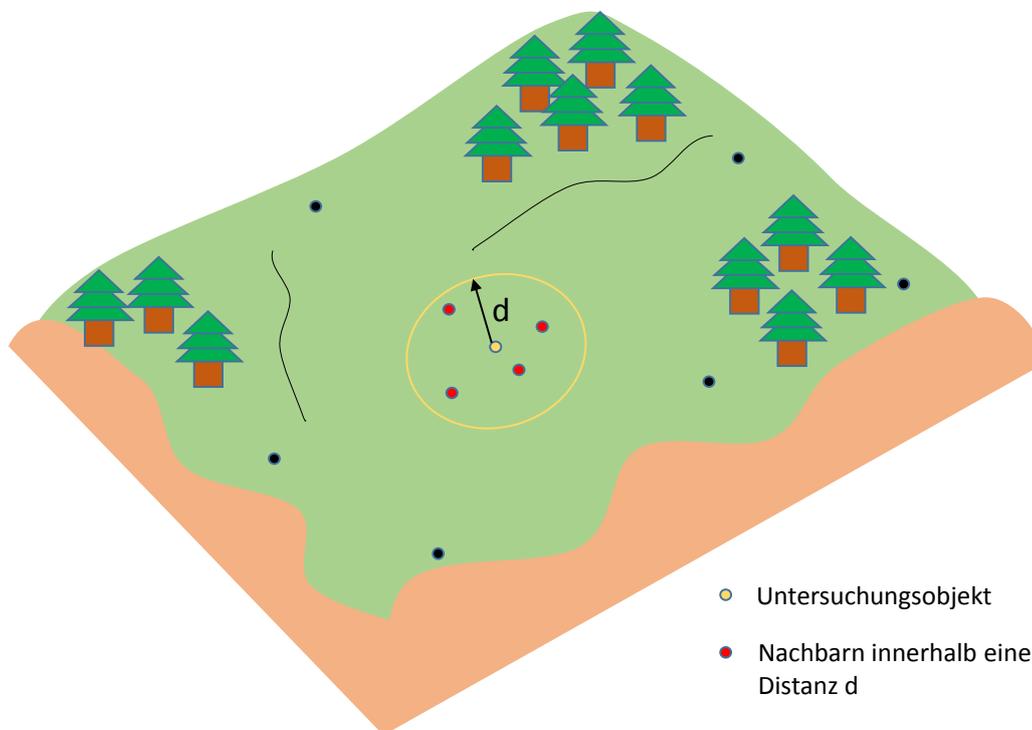


Abbildung 44: Nachbarschaft mit fester Distanz

5.3.6. Gewichtung der räumlichen Nachbarn

Nachdem Festlegung der Nachbarschaft von Untersuchungsobjekten ist noch die Gewichtung zur Quantifizierung der lagebedingten Abhängigkeit zwischen den räumlichen Objekten zu bestimmen. Die Gewichtungsmatrix, in der die räumlichen Gewichtungen in Abhängigkeit ihrer Nachbarschaft enthalten sind, spiegelt die räumliche Struktur der Daten in der Nachbarschaft eines Untersuchungsobjektes wider. Zur Auswahl einer Gewichtungsmethode, die am besten

die gegenseitige Beeinflussung der Objekte widerspiegelt, stehen folgende bekannten Gewichtungsmethoden zur Verfügung [MITCHELL 2005]:

- **Inverse Distanz**

Es wird angenommen, dass die Abhängigkeit zwischen zwei Objekten mit zunehmender Distanz abnimmt: Je größer die Distanz zwischen den Objekten, desto geringer sind die gegenseitige Einflüsse. Die Gewichtung beruht auf der Inversen einer potenzierten Distanz. Es gilt:

Formel 30:
$$w_{ij} = d_{ij}^{-p}$$

Wobei:

w_{ij} : räumliche Gewichtung zwischen Objekt i und j

d_{ij} : Distanz zwischen Objekt i und j

p : wählbare Potenz

Die Potenz ist eine positive reelle Zahl, mit deren Hilfe die gegenseitigen Einflüsse auf Basis der Distanz gesteuert werden können. Durch einen hohen Potenzwert werden die Einflüsse von am nächsten liegenden Objekten verstärkt, während weiter entfernt liegende Objekte schneller an Einfluss verlieren.

- **Einflusszone**

Bei dieser Gewichtungsmethode besitzen alle Objekte, die Nachbarn eines Untersuchungsobjektes innerhalb einer vordefinierten Distanz sind, gleiche Werte. Die Objekte außerhalb der vorgegebenen Distanz haben hingegen überhaupt keinen Einfluss auf das Untersuchungsobjekt, d. h. Ihre Gewichte sind null. Es gilt:

$$w_{ij} = \begin{cases} \text{Konstante}, & d_{ij} \leq d \\ 0 & , d_{ij} > d \end{cases}$$

mit:

d : vordefinierte Distanz.

- **Indifferenzzone**

Bei dieser Gewichtungsmethode handelt es sich um eine Kombination der Methoden „inverse Distanz“ und „Einflusszone“. Ähnlich wie bei der Gewichtungsmethode „Einflusszone“ wird den

innerhalb einer vordefinierten Distanz liegenden Nachbarn die gleiche Gewichtung gegeben. Sobald Objekte die vordefinierte Distanz überschreiten, wird ihr Einfluss mittels der „inversen Distanz“ gewichtet; die Gewichtung fällt mit der Distanz schnell ab. Es gilt:

$$w_{ij} = \begin{cases} \text{Konstante} & , d_{ij} \leq d \\ \text{nach Inverse Distanz} & , d_{ij} > d \end{cases}$$

mit:

d : vordefinierte Distanz

Eine graphische Gegenüberstellung der charakteristischen Eigenschaften der genannten Gewichtungsmethoden zeigt Abbildung 45.

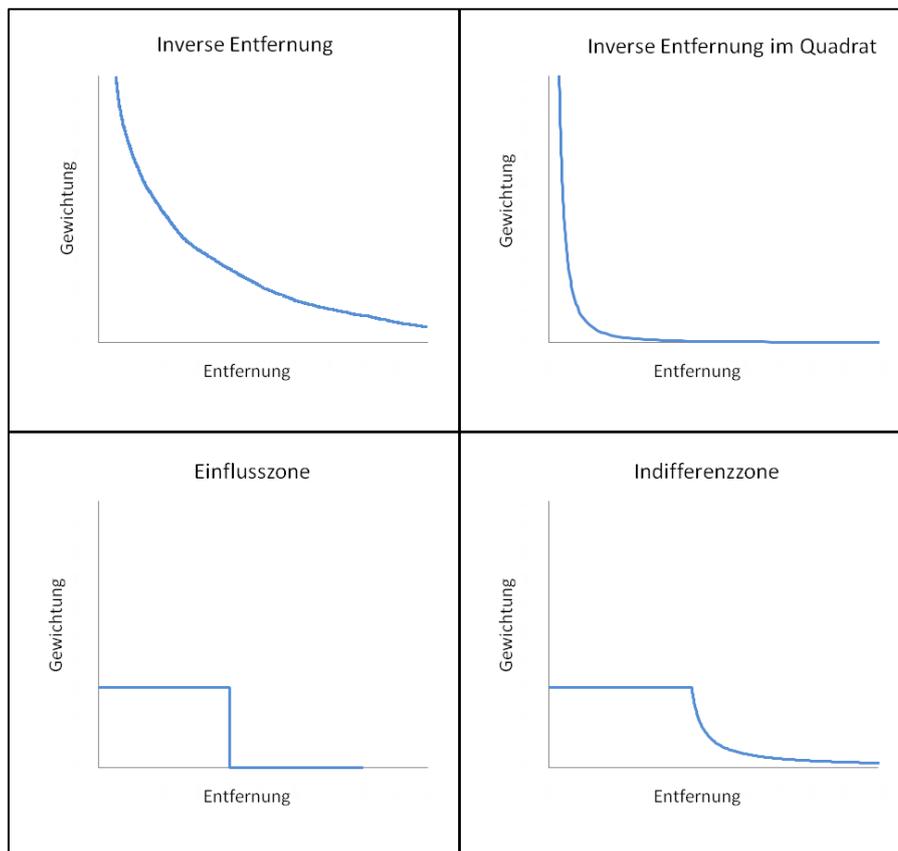


Abbildung 45: Eigenschaften bekannter Gewichtungsmethoden
[VERÄNDERT NACH ESRI ONLINE]

Nach Festlegung der Gewichtungen zwischen den benachbarten Objekten lässt sich die $n \times n$ Gewichtungsmatrix w_{ij} in folgender Form erstellen, wobei die Zeilen der Matrix durch i ($i = 1, \dots, n$) und die Spalten durch j ($j = 1, \dots, n$) indiziert wurden [PICHLER 1998].

$$W_{ij} = \begin{bmatrix} W_{11} & \cdots & W_{1n} \\ \vdots & \ddots & \vdots \\ W_{n1} & \cdots & W_{nn} \end{bmatrix}$$

Diese Gewichtungsmatrix kann normiert werden, so dass die Zeilensumme aller Gewichtungen eins ergibt. Damit wird sichergestellt werden, dass für jedes Objekt die Summe aller Gewichte seiner Nachbarn gleich groß (= 1) ist. Die standardisierte Gewichtung W_{ij}^* kann wie folgt berechnet werden:

Formel 31:

$$W_{ij}^* = \frac{W_{ij}}{\sum_{j=1}^n W_{ij}}$$

5.4. Beispiel zur Berechnung des Local-Moran's-Index

Die Anwendung des Local-Moran's-Index - Verfahrens zur Berechnung des lokalen räumlichen Autokorrelationskoeffizienten soll an einem einfachen Beispiel verdeutlicht werden.

Im Untersuchungsgebiet befinden sich fünf Objekte, deren räumliche Lage sowie Eigenschaftswerte x_i in Abbildung 46 dargestellt sind.

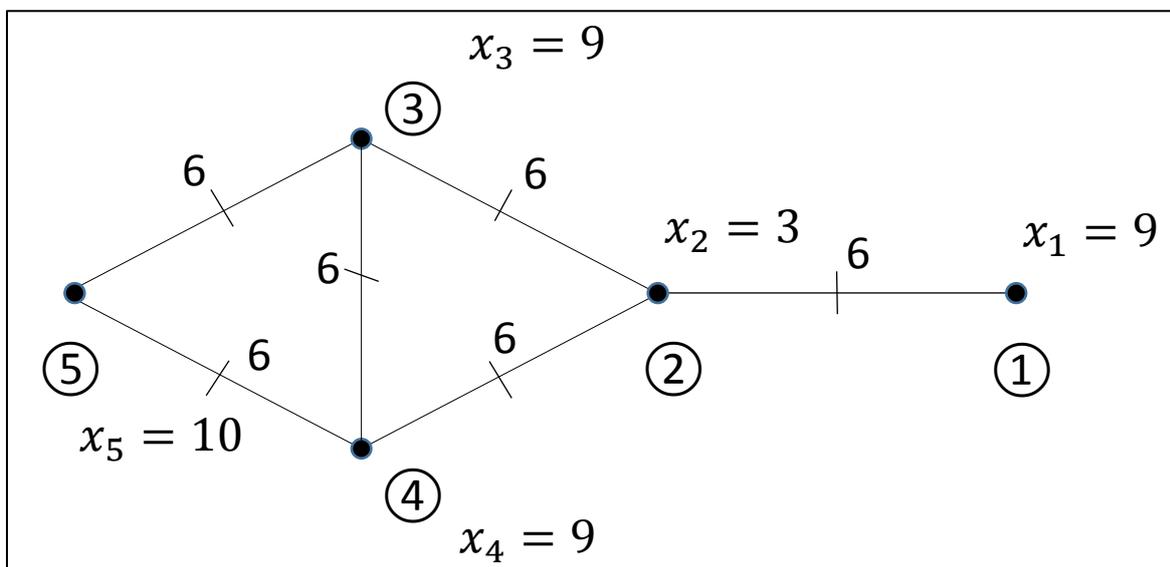


Abbildung 46: Räumliche Lage und Variablenwerte der fünf Untersuchungsobjekte

Es wird davon ausgegangen, dass die Objekte innerhalb einer Distanz von 8 m mit der gleichen Gewichtung untereinander korreliert sind. Es gilt somit die Gewichtungsmethode

„Einflusszone“. Die Objekte außerhalb der Distanzzone werden nicht gewichtet. Darauf aufbauend wird zuerst die Gewichtungsmatrix erstellt. Beispielsweise hat das Objekt 1 innerhalb einer 8 m Einflusszone nur einen Nachbarn (Objekt 2) und die Gewichte W_{12} sowie W_{21} erhalten daher jeweils den Wert 1. Entsprechend ergeben sich die Gewichte für die anderen Objekte, wobei die Gewichte auf der Hauptdiagonalen der Matrix (die Gewichtung für sich selbst) immer gleich Null sind.

$$W = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Nach der Normierung der Gewichte lässt sich die Gewichtungsmatrix wie folgt darstellen:

$$W^* = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{bmatrix}$$

Das arithmetische Mittel der Eigenschaftswerte der Untersuchungsobjekte beträgt:

$$\bar{x} = \frac{\sum_{i=1}^5 x_i}{5} = \frac{9 + 3 + 9 + 9 + 10}{5} = 8$$

und die Varianz der Eigenschaftswerte:

$$S_i^2 = \frac{\sum_{j=1}^n (x_j - \bar{x})^2}{n - 1} = \frac{(9 - 8)^2 + (3 - 8)^2 + (9 - 8)^2 + (9 - 8)^2 + (10 - 8)^2}{5 - 1} = 8$$

Anhand der Formel 22 können beispielsweise die Local-Moran's-Indizes von Objekt 2 und 5 berechnet werden:

$$\begin{aligned} I_2 &= \frac{(x_2 - \bar{x}) \sum_{j=1}^5 w_{2j} (x_j - \bar{x})}{S_i^2} \\ &= \frac{(3 - 8) \times [1/3 \times (9 - 8) + 0 \times (3 - 8) + 1/3 \times (9 - 8) + 1/3 \times (9 - 8) + 0 \times (10 - 8)]}{8} \\ &= -0,625 \end{aligned}$$

$$\begin{aligned} I_5 &= \frac{(x_5 - \bar{x}) \sum_{j=1}^5 w_{5j} (x_j - \bar{x})}{S_i^2} \\ &= \frac{(10 - 8) \times [0 \times (9 - 8) + 0 \times (3 - 8) + 1/2 \times (9 - 8) + 1/2 \times (9 - 8) + 0 \times (10 - 8)]}{8} \\ &= 0,25 \end{aligned}$$

Da das Untersuchungsobjekt 2 einen negativen Index-Wert besitzt, ergibt sich, dass es, im Vergleich zu seinen Nachbarn ($x_1, x_3, x_4 = 9$), einen unterschiedlichen Eigenschaftswert ($x_2 = 3$) besitzt. Somit kann das Objekt 2 als „Anomalie“ erkannt werden. Für das Objekt 5 liegt eine positive Autokorrelation mit seinen Nachbarn vor, d.h. das Objekt mit seinem überdurchschnittlichen Eigenschaftswert ($x_5 = 10$) grenzt auch an Objekte ($x_3, x_4 = 9$) an, deren Eigenschaftswerte ebenfalls überdurchschnittlich sind.

Anschließend sollte in der Praxis noch eine Signifikanzprüfung durchgeführt werden. Wegen der im Beispiel überschaubaren Untersuchungsobjekte wird jedoch hier darauf verzichtet.

6. Modifizierung des Local-Moran's-Index - Clusterverfahrens zur praktischen Anwendung

Wie im vorherigen Kapitel dargelegt, erweist sich das Local-Moran's-Index - Clusterverfahren, im Vergleich zu den anderen Clusterverfahren, als eine geeignete Methode zur Erkennung von Anomalien in räumlichen Datensätzen mit Eigenschaftswerten. In diesem Kapitel soll am Beispiel realer Datensätze untersucht werden, ob eine Identifizierung von Anomalien in flächendeckend vorliegenden PSI-Ergebnissen mit diesem Verfahren tatsächlich erfolgreich ist oder ob Erweiterungen notwendig sind.

6.1. Allgemeiner Verfahrensablauf des Local-Moran's-Index - Clusterverfahrens zur Identifizierung von Anomalien

Der allgemeine Verfahrensablauf des Local-Moran's-Index - Clusterverfahrens zur Identifizierung von Anomalien ist in Abbildung 47 dargestellt.

Mithilfe des Verfahrens sollen räumliche Cluster mit hohen oder niedrigen Eigenschaftswerten gebildet und darauf aufbauend räumliche Anomalien identifiziert werden. Als Inputdaten dient ein räumlicher Datensatz, in dem jedes Untersuchungsobjekt einen lagebezogenen sachgemäßen Eigenschaftswert besitzt. Anhand der räumlichen Beziehungen zwischen den Untersuchungsobjekten wird zuerst die räumliche Gewichtungsmatrix berechnet. Sie fließt später zusammen mit den Eigenschaftswerten der Untersuchungsobjekte als Eingangsparameter in die Berechnung ein. Als Ergebnis wird danach ein Local-Moran's-Index - Wert, ein Z-Ergebnis und ein p-Wert für jedes Untersuchungsobjekt erhalten. Dabei dient der I-Wert zur Einordnung der betrachteten Objekte in Clustern und zur Identifizierung von Anomalien, während die Z-Ergebnisse und p-Werte zur Prüfung der statistischen Signifikanz der berechneten Indexwerte dienen [Xi 2015].

Ein positiver und signifikanter Wert von I_i ($I_i > 0$) bedeutet, dass das Untersuchungsobjekt i einen ähnlichen Eigenschaftswert wie seine benachbarten Objekte besitzt. Da die Objekte positiv miteinander autokorreliert sind, ist das Untersuchungsobjekt Teil eines Clusters. Im Gegensatz dazu bedeutet ein negativer signifikanter Wert von I_i ($I_i < 0$) dass das Untersuchungsobjekt i einen deutlich größeren oder kleineren Eigenschaftswert als die benachbarten Objekte besitzt. Da sie negativ miteinander autokorreliert sind, kann das Untersuchungsobjekt als eine „Anomalie“ bezeichnet werden. $I_i = 0$ bedeutet, dass keine Autokorrelation zwischen dem Objekt i und den benachbarten Objekten vorliegt [Xi 2015].

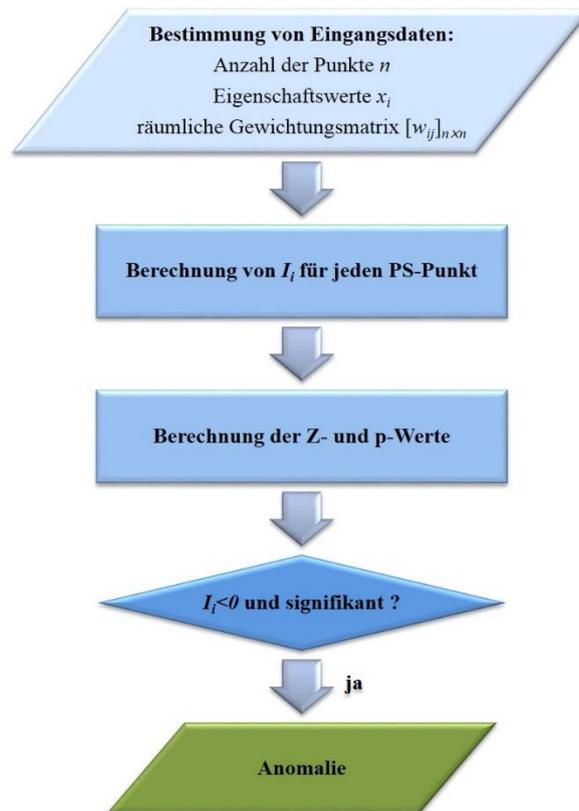


Abbildung 47: Allgemeiner Verfahrensablauf des Local-Moran's-Index - Clusterverfahrens zur Identifizierung von Anomalien [VERÄNDERT NACH XI 2015]

6.2. Beschreibung der Testdaten

Für praktische Untersuchungen stehen drei PSI-Datensätze aus zwei Untersuchungsgebieten mit unterschiedlichen Datendichten zur Verfügung. Grundlage sind Datenstapel der Radarsatelliten TerraSAR-X, Radarsat-2 sowie Envisat. Für die weiteren Untersuchungen wurden die ursprünglichen PSI-Ergebnisse, Abstandsänderungen (Bodenbewegungen) in LOS-Richtung, mittels des Einfallswinkels in Höhenänderungen umgerechnet. Bezogen auf die Höhenänderungen zu einem ausgewählten Referenzzeitpunkt und einen stabilen Referenzpunkt ergibt sich für jeden Messzeitpunkt und PS-Punkt die Höhenänderungsdifferenz, aus der sich für den gesamten Untersuchungszeitraum eine Höhenänderungszeitreihe bildet. Darauf aufbauend lassen sich die linearen Höhenänderungsraten (mm/Jahr) basierend auf einem im Rahmen der PSI-Auswertung verwendeten linearen Regressionsmodell berechnen. Da Höhenänderungen, die durch Bergbauaktivitäten verursacht wurden, über den Untersuchungszeitraum sowohl lineare als auch nicht lineare Höhenänderungstrends aufweisen können, wird zur Ergänzung noch eine statistische Analyse auf Grundlage eines best-angepassten Polynom-Trendmodelles (s.

Kapitel 2.3) für die Höhenänderungszeitreihe durchgeführt. Als Ergebnisse dieser Polynomausgleichung ergeben sich für jeden PS-Punkt:

- der optimale Polynomgrad
- die Parameter (Polynomkoeffizienten) des orthogonalen Polynom-Trendmodells und ihre Standardabweichungen
- die Verbesserungen der Höhenänderungswerte zu jedem Messzeitpunkt
- die Standardabweichung eines einzelnen Höhenänderungswertes.

Einige dieser Daten werden später als Merkmale (Eigenschaften) für die Clusteranalyse zur Anomaliedetektion verwendet und auf ihre Eignung hin diskutiert.

6.2.1. TerraSAR-X Testdaten

Der 2007 gestartete deutsche Radarsatellit TerraSAR-X arbeitet im X-Band Frequenzbereich und liefert alle 11 Tage Radardaten [PITZ 2010]. Für dieses Testbeispiel stand ein Szenenstapel mit einer Bodenauflösung von ca. 2 m * 2 m für den Zeitraum Oktober 2011 bis Januar 2013 zur Verfügung. Eine am IGMC (Institut für Geotechnik und Markscheidewesen der TU Clausthal) durchgeführte PSI-Datenprozessierung mit dem Softwarepaket der Fa. GAMMA für dieses Testgebiet (Ausdehnung ca. 6 km × 7 km) erzeugte etwa 232.000 PS-Punkte, deren räumliche Verteilung (Punktdichte) sich aufgrund der unterschiedlichen Rückstreuungseigenschaften in städtischen und ländlichen Gebieten deutlich unterscheidet (Abbildung 48).

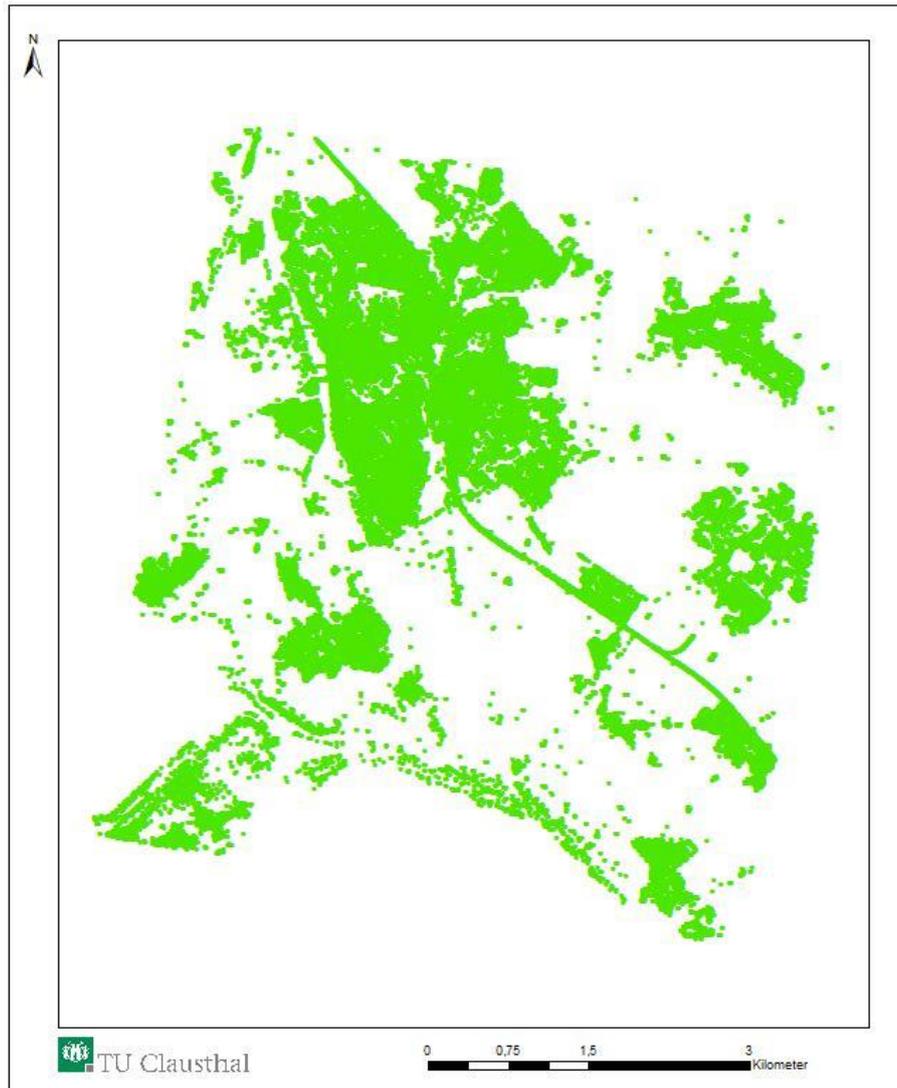


Abbildung 48: Räumliche Lage der TerraSAR-X PS-Punkte

6.2.2. Radarsat-2 Testdaten

Der Radarsat-2 Erdbeobachtungssatellit wird von dem kanadischen Unternehmen MDA Space Missions betrieben und liefert seit 2007 im Abstand von 24 Tagen radarinterferometrisch auswertbare Daten. Radarsat-2 arbeitet im C-Band Frequenzbereich. Die zur Verfügung gestellten Daten besitzen eine Bodenauflösung von etwa 23 m * 8 m. Für das Testgebiet (Ausdehnung ca. 16 km × 20 km) konnten durch das IGMC nach der PSI-Auswertung ca. 70.000 PS-Punkte aus 74 Szenen im Zeitraum von Juni 2010 bis Juni 2015 ermittelt werden [BUSCH 2016], deren Lage und Verteilung in Abbildung 49 dargestellt ist.

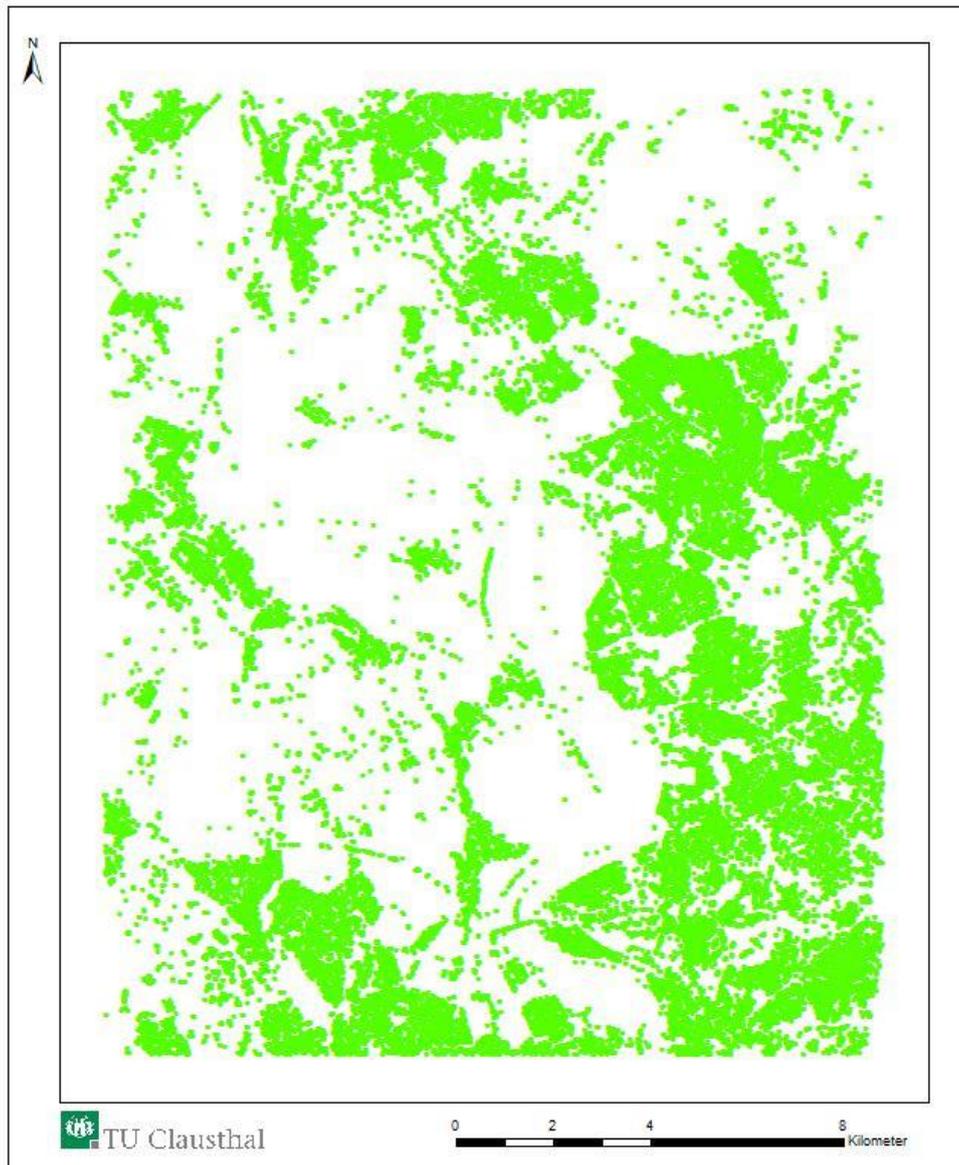


Abbildung 49: Räumliche Lage und Verteilung der Radarsat-2 PS-Punkte

6.2.3. Envisat Testdaten

Für dasselbe Testgebiet standen auch Daten des Radarsensors ASAR (Advanced Synthetic Aperture Radar) des von der europäischen Weltraumorganisation ESA im Jahr 2002 gestarteten Umweltsatelliten Envisat zur Verfügung. Die Envisat-Radardaten, die im C-Band Frequenzbereich alle 35 Tage aufgenommen wurden, besitzen eine Bodenauflösung von etwa 23 m * 5 m. Insgesamt konnten für dieses Testgebiet (Ausdehnung ca. 16 km × 20 km) etwa 34.000 PS-Punkte aus dem Envisat-Datenstapel für den Zeitraum Dezember 2003 bis Oktober 2010 mit 69 Messzeitpunkten bestimmt werden [BUSCH 2016]. Die räumliche Lage und Verteilung der ermittelten PS-Punkte der Envisat-Daten ist in Abbildung 50 dargestellt.



Abbildung 50: Räumliche Lage und Verteilung der Envisat PS-Punkte

6.3. Optimierte Festlegung der Nachbarschaft

Wie bereits im Kapitel 5.3.5 ausgeführt, besitzt die Festlegung von Nachbarschaftsverhältnissen eine wichtige Rolle bei der Abbildung von räumlichen Beziehungen zwischen den Untersuchungsobjekten. Daher setzt eine erfolgreiche Durchführung von räumlichen Clusteranalysen für die Detektion von Anomalien, die auf den räumlichen Beziehungen von Untersuchungsobjekten beruht, eine geeignete Nachbarschaftsdefinition voraus. Zur Entdeckung lokaler Phänomene soll das Local-Moran's-Index - Clusterverfahren zum Einsatz kommen. Seine Anwendung soll das Ausmaß lokaler Abhängigkeiten und Inhomogenitäten in Form von Anomalien mit Signifikanzausgagen quantifizieren. Hierbei kann die Definition eines lokalen Untersuchungsbereiches nur durch die

Festlegung einer Nachbarschaft, die meistens durch ein festes Entfernungsband (Suchradius) gekennzeichnet ist, ermöglicht werden.

Angesichts der Charakteristik der verwendeten PSI-Daten sowie des Local-Moran's-Index - Clusterverfahrens sind folgende Restriktionen für die Festlegung eines geeigneten Suchradius zu beachten:

- Prinzipiell sollte der Suchradius so gewählt werden, dass er die geographische Ausdehnung eines lokalen räumlichen Prozesses wiedergibt. Da die Bodenbewegungen als räumlicher Prozess, der sich in den PSI-Daten widerspiegelt, jedoch in seiner geographischen Ausdehnung ohne detaillierte fachliche Hintergrundkenntnisse nicht abgegrenzt werden kann, gibt es keinen allgemein gültigen und anwendbaren Erfahrungswert über die Größe des Suchradius.
- Für die Durchführung des Local-Moran's-Index - Clusterverfahrens muss sichergestellt werden, dass alle PS-Punkte mindestens einen Nachbarn besitzen. Es empfiehlt sich jedoch, dass jedem PS-Punkt mindestens acht Nachbarn für die Clusteranalyse zur Verfügung stehen [GETIS 1992]. Dies kann infolge der üblichen inhomogenen Verteilung von PS-Punkten dazu führen, dass manche PS-Punkte Tausende von Nachbarn haben, während andere räumlich isolierte Einzelpunkte keine erforderliche Anzahl von Nachbarn besitzen. Daher ist es empfehlenswert, räumlich isolierte PS-Punkte zu detektieren und bei der Festlegung eines optimalen Suchradius vorübergehend auszuschließen.
- Der Suchradius sollte weder zu klein noch zu groß sein. Im Falle eines kleinen Suchradius besteht die Gefahr, dass die Mindestanzahl von Nachbarn zur erfolgreichen Ausführung des Local-Moran's-Index - Clusterverfahrens nicht erreicht wird, während bei einem großem Suchradius PS-Punkte sehr viele Nachbarn haben können; ein Zustand, der die Signifikanzprüfung durch unzuverlässige Z-Werte verfälschen könnte.
- Der Indikator zur Quantifizierung der räumlichen Autokorrelation gibt Aufschluss darüber, inwieweit die zugrunde liegenden räumlichen Prozesse, die durch die Attribute der Untersuchungsobjekte beschrieben werden, noch eine signifikante Abhängigkeit besitzen. Mithilfe solcher Indikatoren (Index) kann der optimale Suchradius bestimmt werden. Dazu wird die räumliche Autokorrelation (s. Kapitel 5.3.2 globaler Moran's-Index) für eine Reihe zunehmender Entfernungen (Suchradien) berechnet. Der optimale Suchradius stellt die Entfernung dar, bei der die räumlichen Abhängigkeiten der zugrunde liegenden räumlichen Prozesse (Bodenbewegungen) am größten sind. Dieses Ergebnis wird typischerweise in einem Diagramm verdeutlicht (s. Abb. 51), in

dem die Veränderung der Z-Werte in Relation zur Erhöhung der Entfernungswerte dargestellt ist.

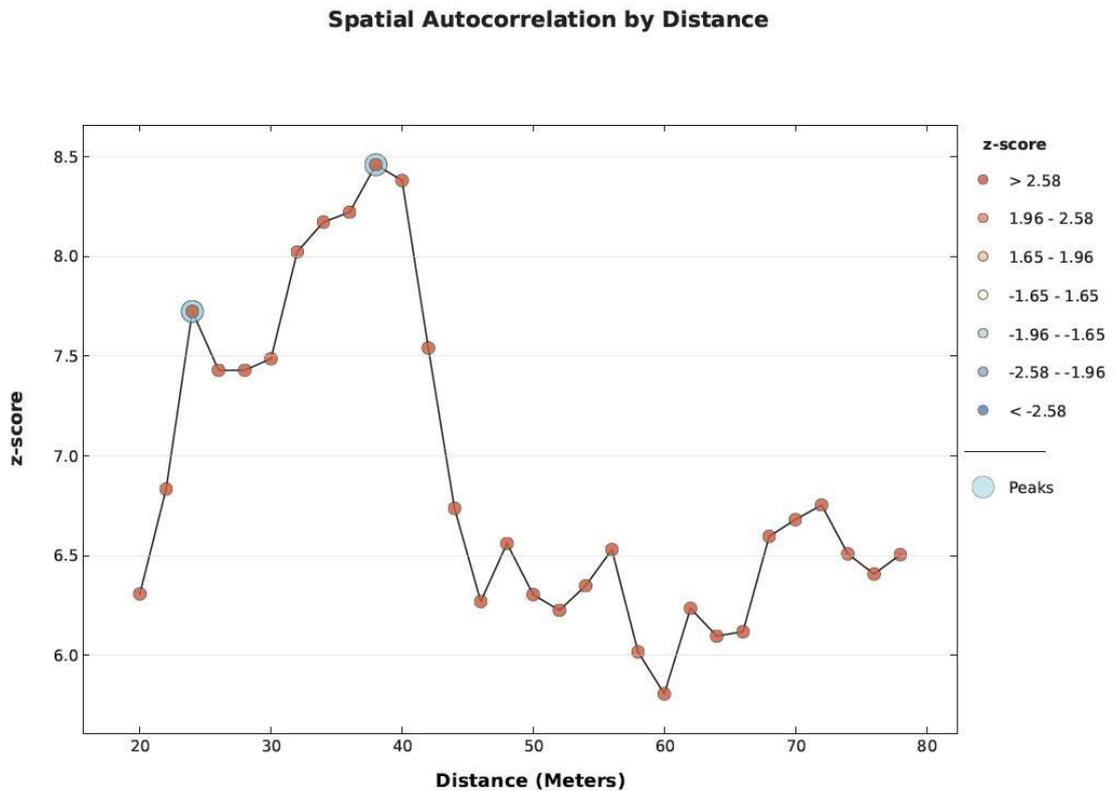


Abbildung 51: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung

Die Stärke der räumlichen Autokorrelation zwischen den betrachteten Untersuchungsobjekten wird durch den berechneten Z-Wert repräsentiert. Normalerweise erhöht sich am Anfang der Z-Wert mit ansteigender Entfernung, bis er bei einer bestimmten Entfernung seinen Höhepunkt (blauer Peak bei etwa 24 m in Abbildung 51) erreicht. Ab dieser Entfernung verringert sich der Z-Wert und somit die Stärke der räumlichen Autokorrelation. Die Entfernung, bei der ein Maximalwert auftritt, kann als optimaler Suchradius interpretiert werden, da die zugrunde liegenden räumlichen Prozesse in dieser Entfernung am stärksten korreliert sind. Es kann aber auch vorkommen, dass es mehrere Peaks des Z-Wertes gibt. Wie in Abbildung 51 dargestellt, erreicht hier der Z-Wert einen zweiten Höhepunkt bei einer Entfernung von ca. 38 m. Im Allgemeinen gibt der mit größerer Entfernung entstandene Maximalwert Aufschluss darüber, dass Trends über einen größeren Bereich des Datensatzes vorliegen. Für die lokalen Clusteranalysen ist hauptsächlich der erste Maximalwert des Z-Wertes mit einer kleineren Entfernung von Interesse [ESRI ONLINE]. Je nach vordefiniertem Signifikanzniveau gibt die Farbe jedes Punktes in dem Diagramm die statistische Signifikanz der berechneten Z-Werte an. Die orangefarbenen Punkte mit

den Z-Werten größer als 2,58 deuten darauf hin, dass sie bei einem Signifikanzniveau von 1% statistisch signifikant sind. Dies trifft im obigen Beispiel für alle Entfernungen zu.

Im Folgenden wird diese Methode zur Festlegung eines optimalen Suchradius beispielhaft mit den Testdaten erläutert und untersucht. Grundlage hierfür sind alle Daten innerhalb des Untersuchungsgebietes, ohne Differenzierung in lokale Teilbereiche.

In Abbildung 52 ist die Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung am Beispiel der Envisat-PSI-Daten dargestellt.

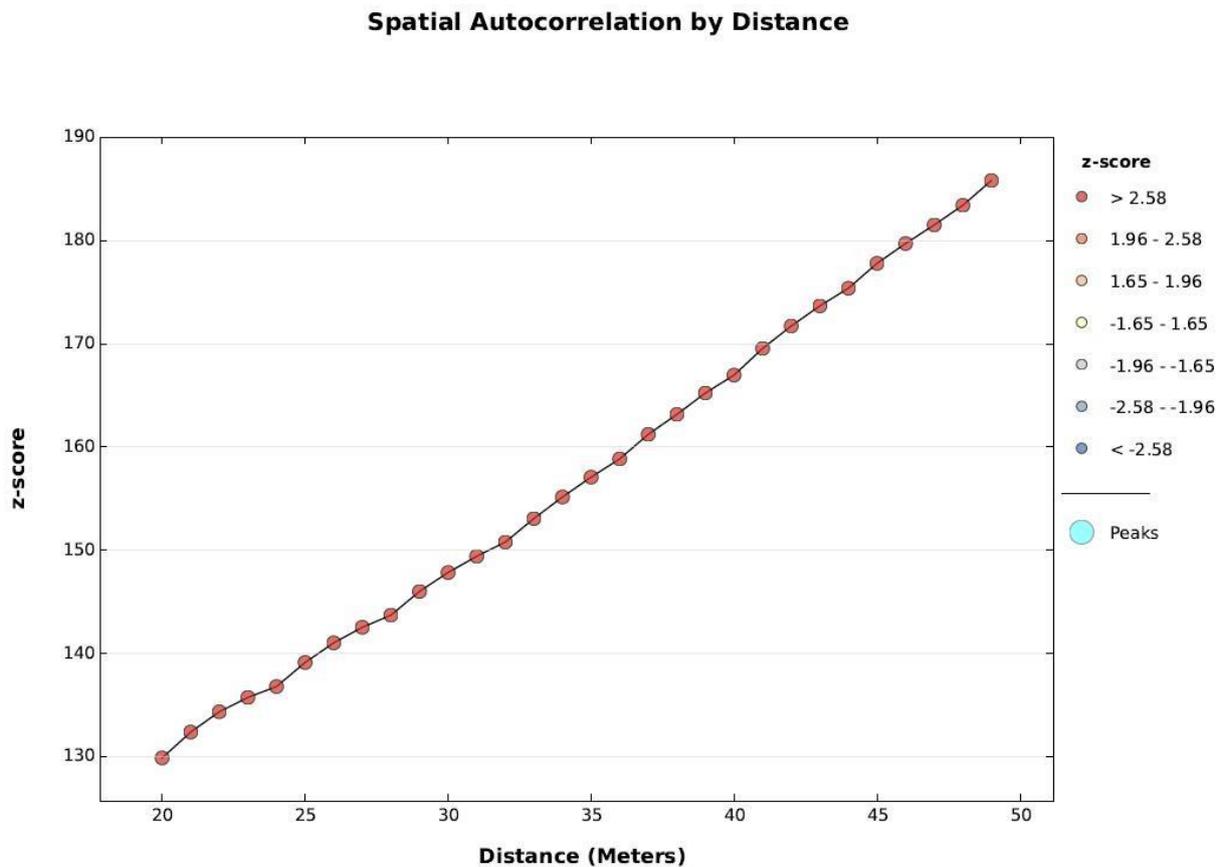


Abbildung 52: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung (20 m - 50 m) für die Envisat-PSI-Daten

Im Gegensatz zu Abbildung 51 und den theoretischen Darlegungen ergibt sich mit zunehmender Entfernung kein Peak im Z-Wert, sondern eine ununterbrochene Vergrößerung des Z-Wertes. Ein möglicher Grund hierfür könnte sein, dass die Anfangsentfernung und das Entfernungsintervall zu klein für die Berechnung der räumlichen Autokorrelation gewählt worden sind.

Deshalb erfolgte eine Neuberechnung mit einem größeren Start- und Endwert. Das Ergebnis in Abbildung 53 zeigt jedoch ebenfalls keinen Peak im Z-Wert Verlauf. Ein vergleichbares Ergebnis ergibt sich auch für die Radarsat-2 und TerraSAR-X PSI-Daten.

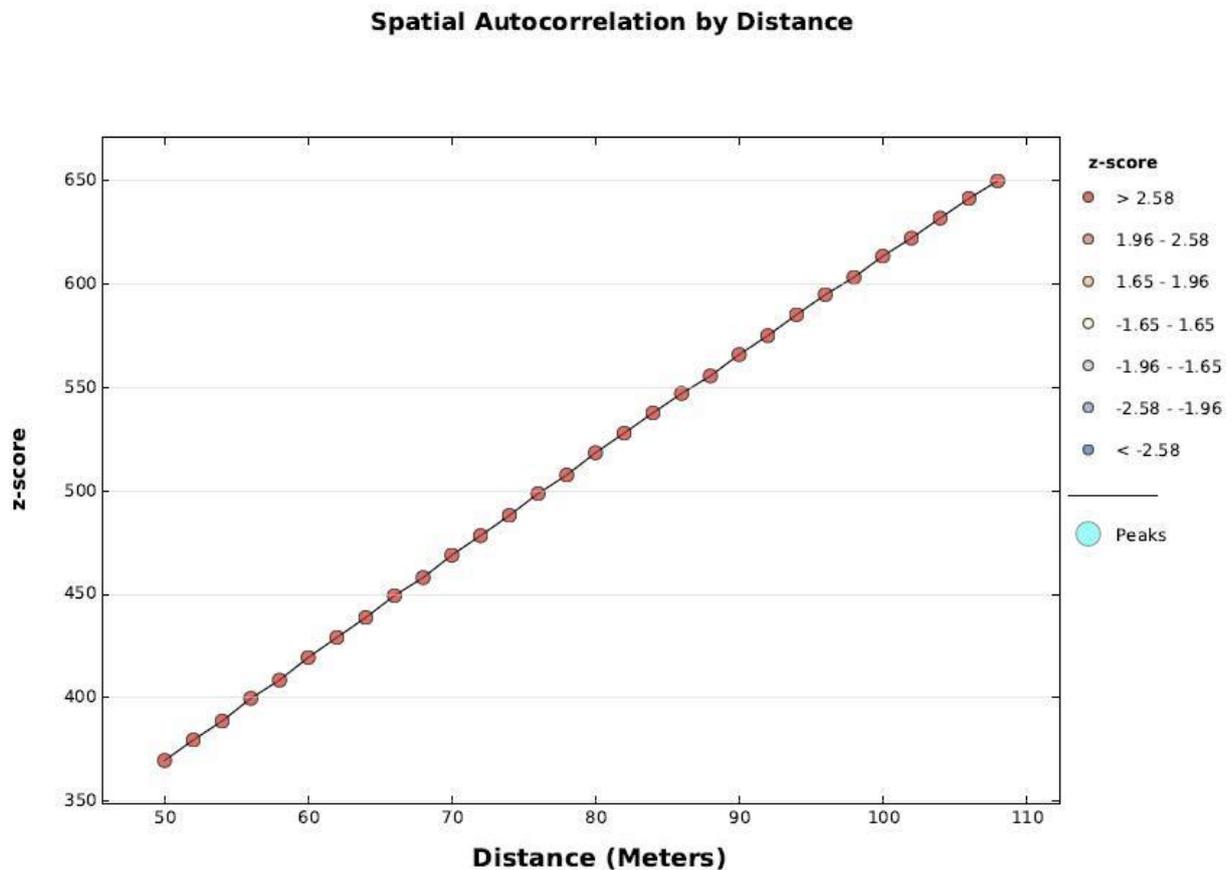


Abbildung 53: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung (50 m – 110 m) für die Envisat-PSI-Daten

Dieses Ergebnis deutet darauf hin, dass offensichtlich mehrere räumliche Prozesse in unterschiedlichen räumlichen Maßstäben im gesamten Untersuchungsgebiet vorliegen und eine gesamtheitliche Betrachtung des gesamten Untersuchungsgebietes die tatsächlichen räumlichen Beziehungen in lokalen Teilbereichen nicht aufdecken kann. Da die Punktdichte und –Verteilung der PS-Punkte aus verfahrenstechnischen Gründen in hohem Maße mit der Bebauungsdichte im Untersuchungsgebiet korreliert ist, wird die PS-Punktverteilung ebenfalls sehr unregelmäßig sein. In den Städten werden deutlich mehr PS-Punkte als in den ländlichen Gebieten detektiert werden, weshalb, siehe auch Abbildung 48 bis Abbildung 50, in einem Untersuchungsgebiet mit Bereichen unterschiedlicher Bebauungsdichte sich eine inhomogene PS-Punktverteilung (im Regelfall) einstellen wird. Infolge einer hohen Bodenauflösung des verwendeten Radarsatelliten können sich, bedingt durch z.B. einzeln stehende Gebäude, zudem auch relativ viele räumlich isolierte PS-Punkte ergeben.

Aufgrund dieser aufnahmebedingten Zusammenhänge bietet sich an, Teilgebiete mit relativ gleichmäßiger Punktverteilung (Bereiche homogener Bebauungsdichte) zu selektieren, und räumlich isolierte PS-Punkte durch z.B. DBSCAN-Clusterverfahren (s. Kapitel 5.2) vorab zu identifizieren und von den weiteren Untersuchungen und Berechnungen auszuschließen.

Zur Demonstration und Validierung des Verfahrens wurden exemplarisch je fünf Teilgebiete des Testgebietes mit den Radarsat-2 und Envisat Daten ausgewählt. Für das Testgebiet mit den TerraSAR-X Daten erfolgte eine Auswahl von zehn Teilgebieten, da dieser Sensor über eine sehr hohe Bodenauflösung verfügt und somit eine große Anzahl von PS-Punkten mit relativ gleichmäßiger Punktverteilung existiert.

Die fünf ausgewählten Teilgebiete der **Radarsat-2 Testdaten**, für die die räumliche Autokorrelation berechnet wurde, sind in Abbildung 54 dargestellt.

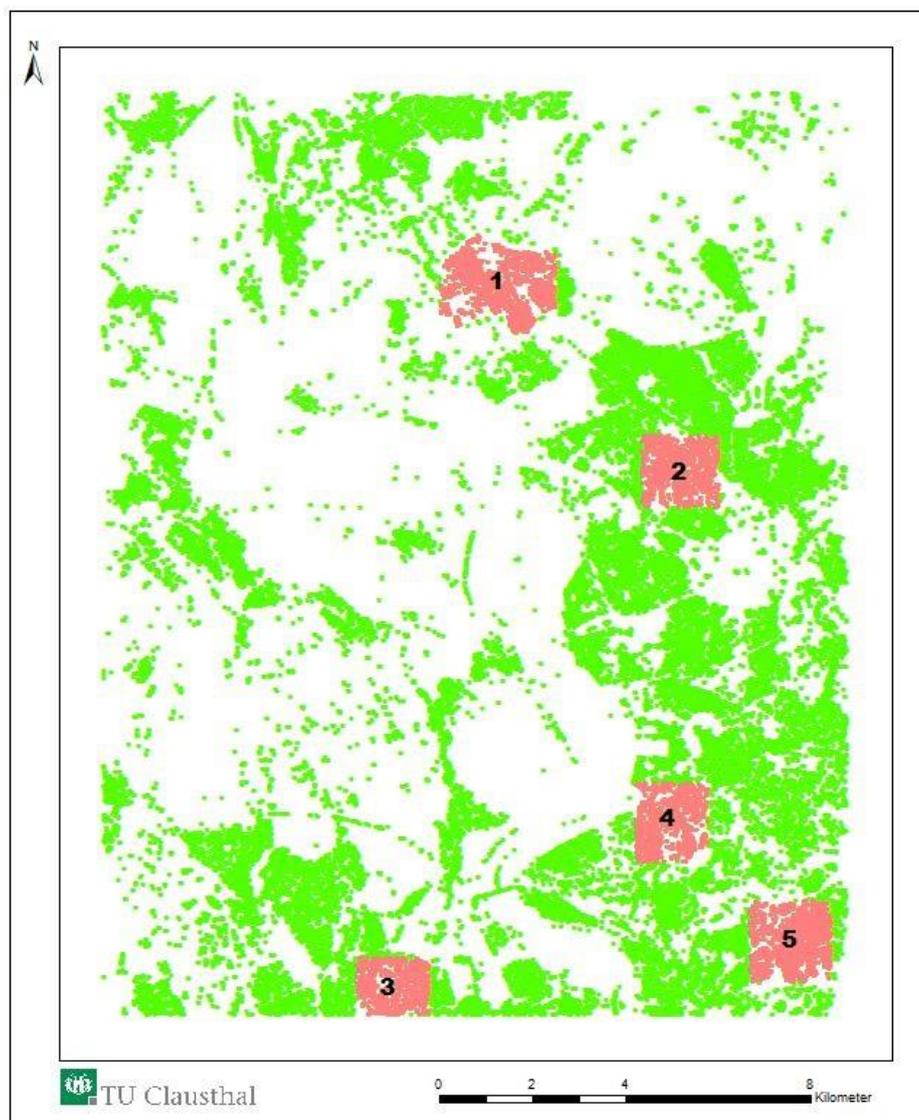


Abbildung 54: Fünf ausgewählte Teilgebiete im Testgebiet der Radarsat-2-Daten

Das Ergebnis der entfernungsbezogenen Autokorrelationsberechnungen ist exemplarisch für das Teilgebiet 1 in Abbildung 55 dargestellt. Die entsprechenden Diagramme für die anderen Teilgebiete finden sich im Anhang A2 - A5.

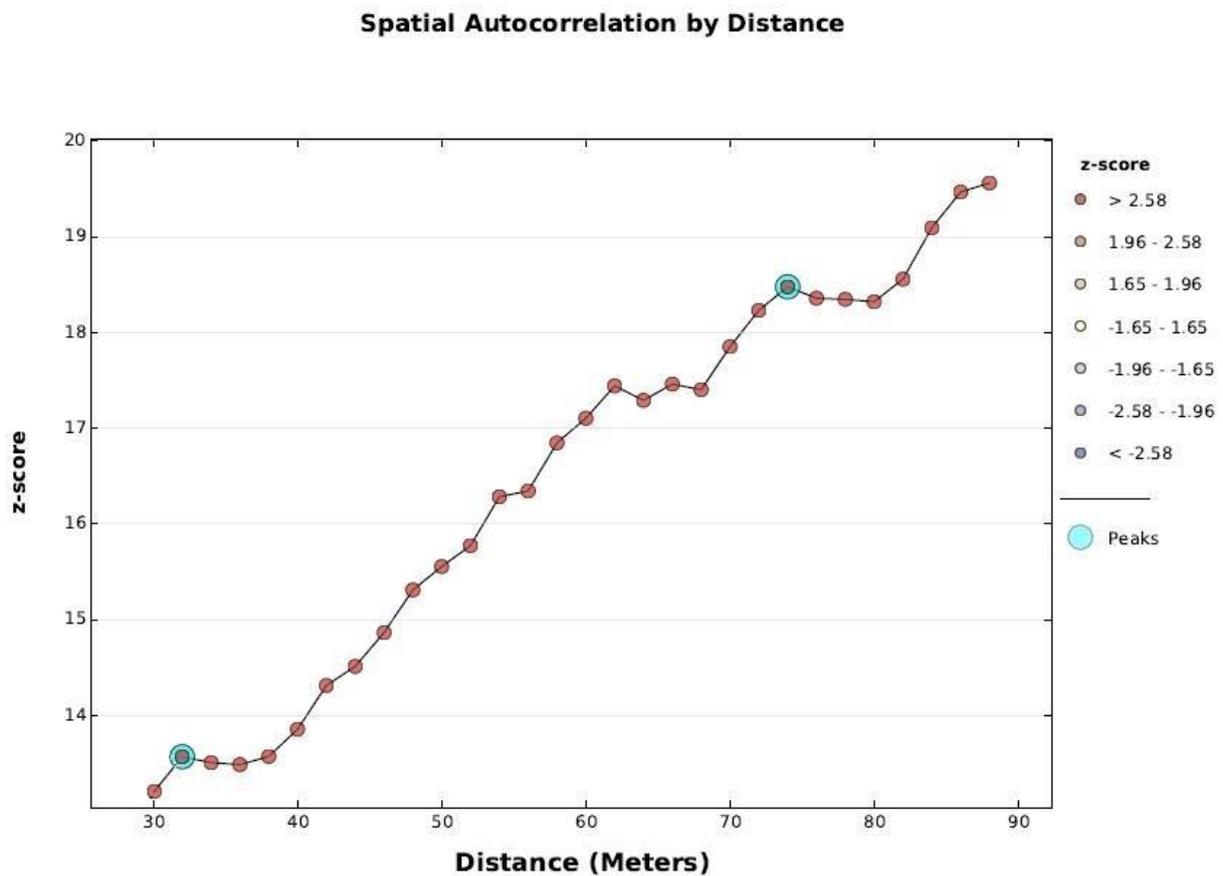


Abbildung 55: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 1 (Abbildung 54) mit Radarsat-2-Testdaten

Für das Teilgebiet 1 ergab sich der erste Peak bei ca. 32 m und der höchste Peak im Z-Wert bei ca. 76 m (als blaue Punkte in Abbildung 55 markiert). Da nur der erste Peak des Z-Wertes (bei kleiner Entfernung) für die lokalen Clusteranalysen von Bedeutung ist, sind die in allen 5 Teilgebieten ermittelten Entfernungen für den ersten Peak im Z-Wert in Tabelle 12 zusammengestellt.

Tabelle 12: Entfernungen (Suchradius), in denen der erste Peak in den fünf Teilgebieten der Radarsat-2-Testdaten auftritt

Teilgebiet	1	2	3	4	5
Entfernung am ersten Peak [m]	32	34	40	34	34

Das Ergebnis dieser Berechnungen für die 5 Radarsat-2-Teilgebiete führt zu einem optimalen Suchradius in einer Größenordnung von 34 m (Median) bei einer Bandbreite von 32 m bis 40 m.

Für die Untersuchungen mit den **Envisat-Testdaten** wurden die in Abbildung 56 dargestellten 5 Teilgebiete ausgewählt.

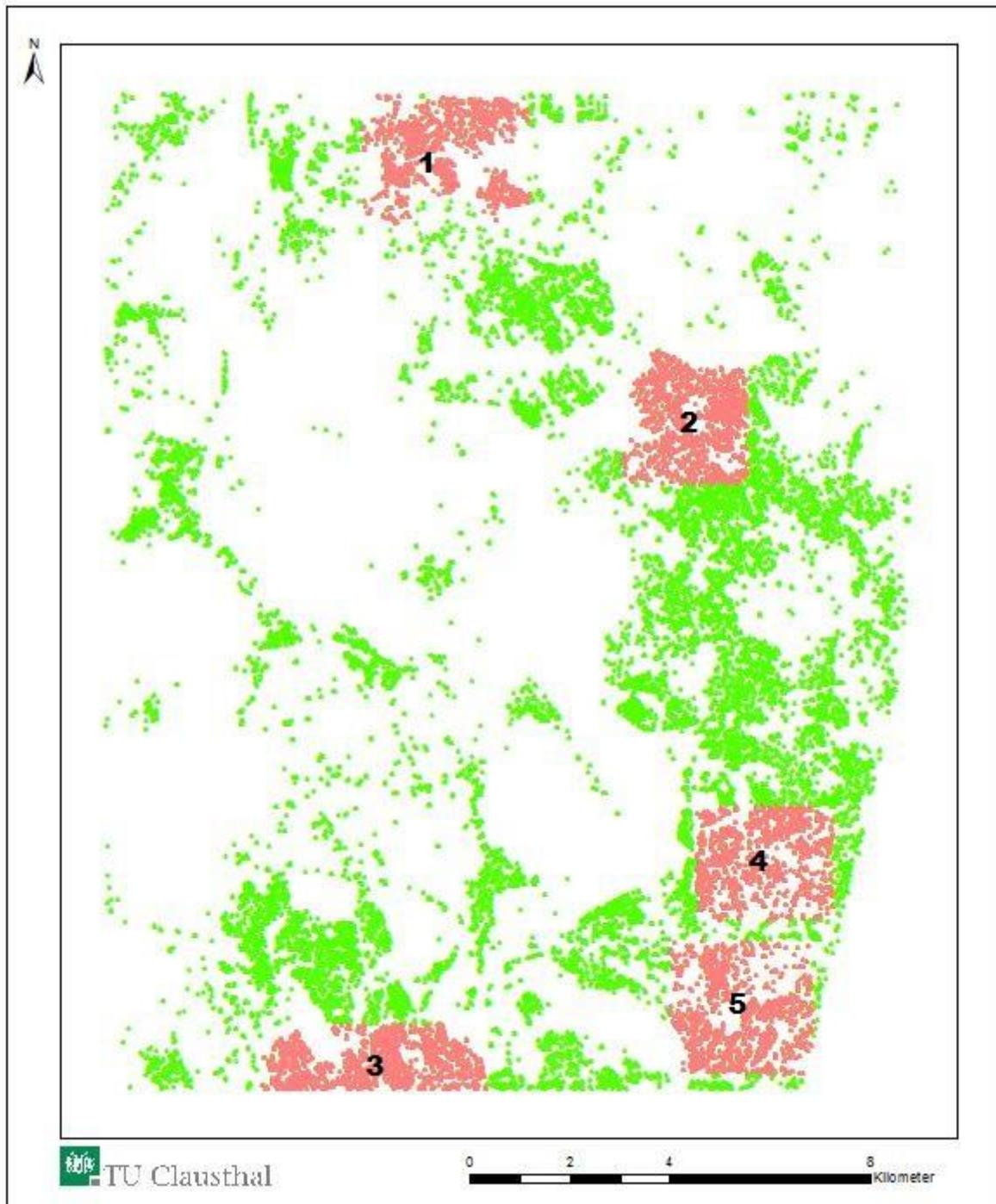


Abbildung 56: Fünf ausgewählte Teilgebiete im Testgebiet mit den Envisat-Daten

Exemplarisch ist in Abbildung 57 für das Teilgebiet 1 der Envisat-Testdaten die Veränderung der räumlichen Autokorrelation gegenüber der zunehmenden Entfernung dargestellt. Die entsprechenden Diagramme für die anderen Teilgebiete finden sich im Anhang A7 - A10.

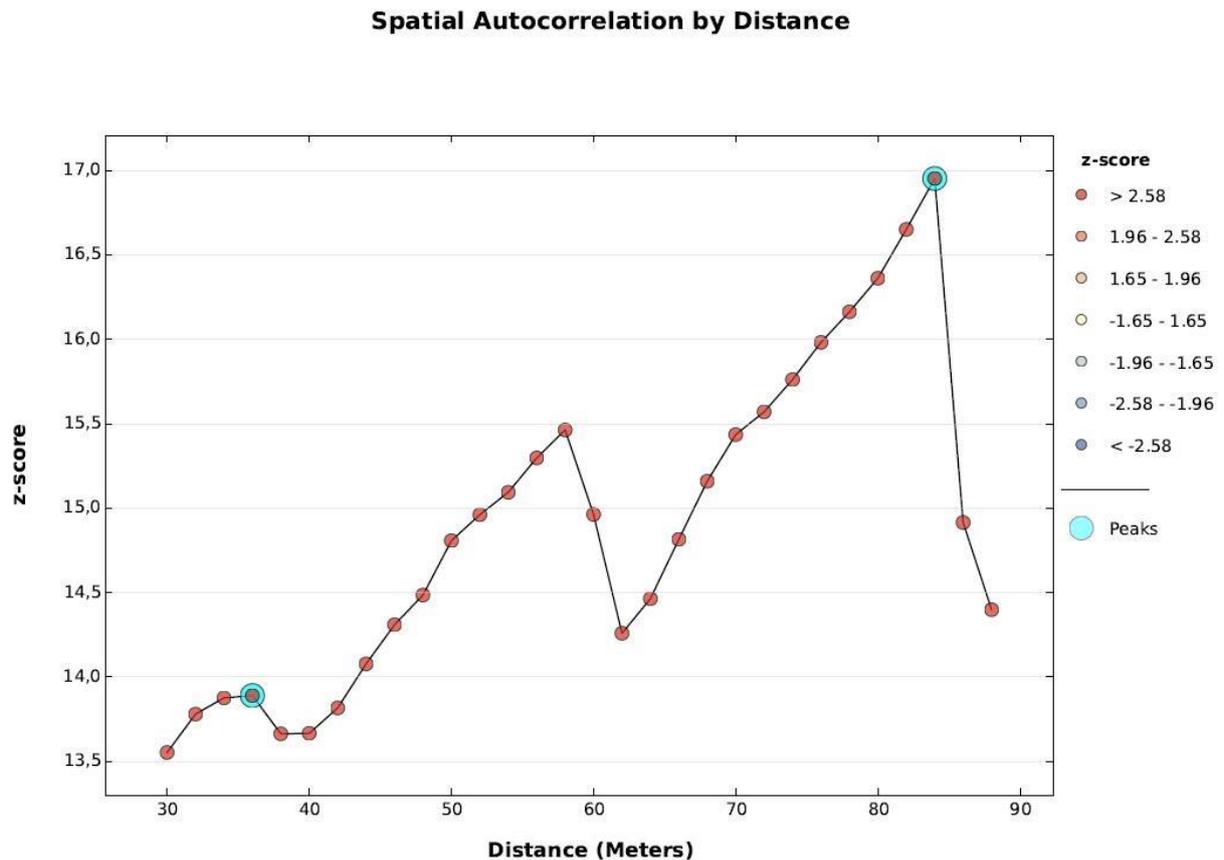


Abbildung 57: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 1 (Abbildung 56) mit Envisat-Testdaten

In Abbildung 57 ist deutlich erkennbar, dass für das Teilgebiet 1 der erste Peak im Z-Wert bei einer Entfernung von ca. 36 m auftritt. In Tabelle 13 sind die erzielten Ergebnisse für alle Teilgebiete mit den Envisat-Testdaten zusammengestellt.

Tabelle 13: Entfernungen (Suchradius), in denen der erste Peak in den fünf Teilgebieten der Envisat-Testdaten auftritt

Teilgebiet	1	2	3	4	5
Entfernung am ersten Peak [m]	36	40	32	36	38

Das Ergebnis dieser Berechnungen für die 5 Envisat-Teilgebiete führt zu einem optimalen Suchradius in einer Größenordnung von 36 m (Median) bei einer Bandbreite von 32 m bis 40 m.

Eine entsprechende Untersuchung erfolgte für die **TerraSAR-X-Daten** in 10 ausgewählten Teilgebieten, deren Lage in Abbildung 58 dargestellt ist.



Abbildung 58: Zehn ausgewählte Teilgebiete im Testgebiet mit den TerraSAR-X-Daten

Die exemplarische Darstellung des Ergebnisses für das Teilgebiet 1 der TerraSAR-X Testdaten in Abbildung 59 zeigt, dass der erste Peak im Z-Wert bei einer Entfernung von ca. 5,1 m auftritt. Die entsprechenden Diagramme für die anderen Teilgebiete finden sich im Anhang A11 - A20.

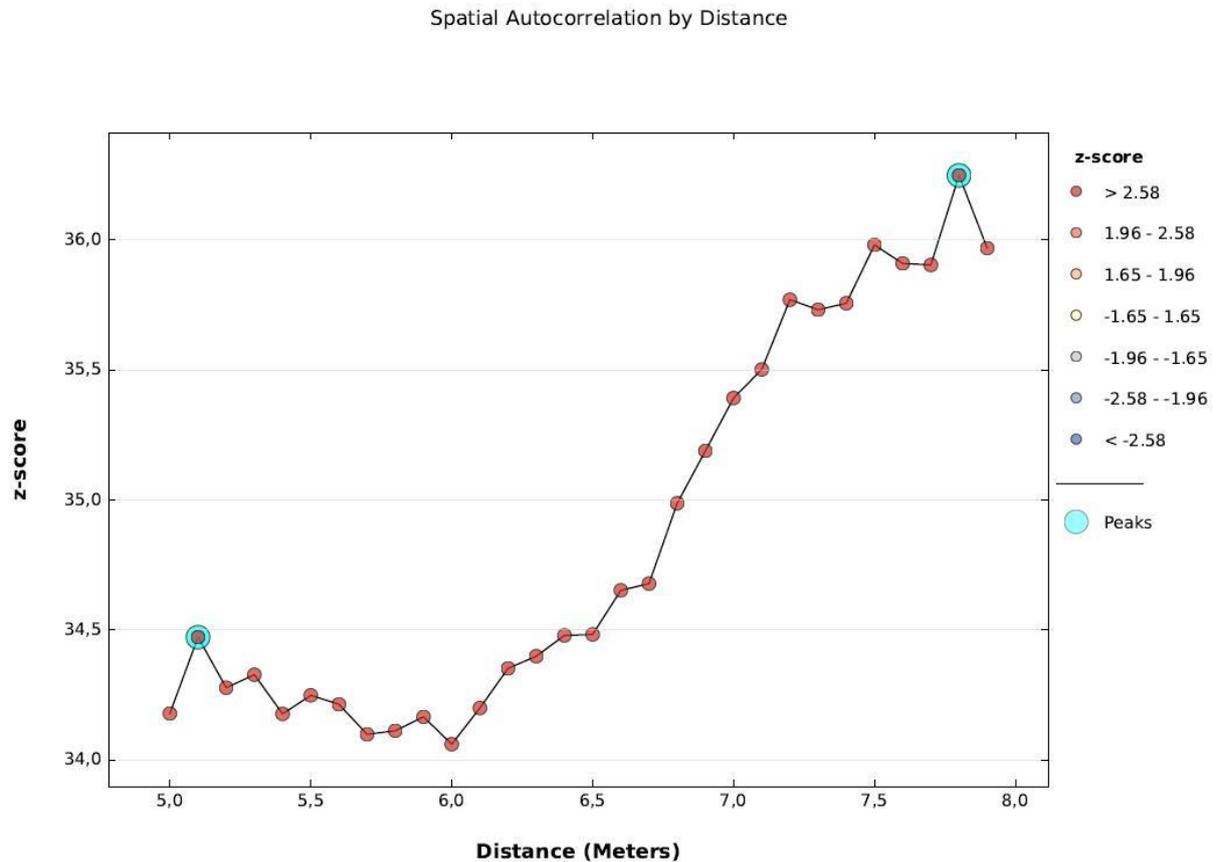


Abbildung 59: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 1 (Abbildung 58) mit TerraSAR-X-Testdaten

In Tabelle 14 sind die Entfernungen für die ersten Peaks im Z-Wert in den zehn ausgewählten Teilgebieten der TerraSAR-X-Testdaten aufgelistet.

Tabelle 14: Entfernungen (Suchradius), in denen der erste Peak in den zehn Teilgebieten der TerraSAR-X-Testdaten auftritt

Teilgebiet	1	2	3	4	5	6	7	8	9	10
Entfernung am ersten Peak [m]	5,1	5,4	5,1	5,3	5,2	5,2	5,4	5,1	5,3	5,3

Das Ergebnis dieser Berechnungen für die 10 TerraSAR-X-Teilgebiete führt zu einem optimalen Suchradius in einer Größenordnung von 5,2 m (Median) bei einer Bandbreite von 5,1 m bis 5,4 m.

Aus den Ergebnissen dieser gebiets- und sensorspezifischen Untersuchungen kann folgendes abgeleitet werden:

- Die Entfernungen, für die ein erster Peak im Verlauf der Z-Werte auftritt, variieren für die ausgewählten Teilgebiete eines Sensors nur geringfügig. Für die Radarsat-2 Testdaten liegt die Entfernung im Bereich zwischen 32 und 40 m (Median 34 m), für die Envisat-Testdaten zwischen 32 und 40 m (Median 36 m) und für die TerraSAR-X-Testdaten zwischen 5,1 m und 5,4 m (Median 5,2 m).
- Die Entfernungen, für die ein erster Peak im Verlauf der Z-Werte auftritt, sind offensichtlich sensorabhängig. Dies zeigt sich besonders deutlich durch einen Vergleich der TerraSAR-X-Ergebnisse mit denen der beiden anderen Sensoren. Zu vermutende Ursache hierfür dürfte die deutlich unterschiedliche Bodenauflösung der verwendeten Sensoren sein.
- Die Entfernungen, für die ein erster Peak im Verlauf der Z-Werte auftritt, sind offensichtlich abhängig vom Untersuchungsgebiet und dem Bodenbewegungsgeschehen im Untersuchungszeitraum. Diese Aussage lässt sich aus einem Vergleich der im gleichen Untersuchungsgebiet jedoch für unterschiedliche Zeiträume ermittelten Ergebnisse der Radarsat-2- und Envisat-Daten ableiten.

Zusammenfassend lässt sich feststellen, dass, im Gegensatz zu einer gesamtheitlichen Betrachtung des Untersuchungsgebietes zur Ableitung eines optimalen Suchradius, durch die Aufteilung in ausgewählte Teilgebiete der vorhandenen PSI-Daten ein optimaler Suchradius, basierend auf der lokalen räumlichen Autokorrelation, zur Bildung von räumlichen Beziehungen für die Local-Moran's-Index - Clusteranalyse ermittelt werden kann. Je nach Anzahl ausgewählter Teilgebiete kann der optimale Suchradius als Mittelwert oder Medianwert der Entfernungen, an denen jeweils der Z-Wert seinen ersten Höhepunkt erreicht hat, angesehen werden. Die Auswahl der Teilgebiete sollte möglichst in zusammenhängend bebauten Gebieten mit relativ regelmäßiger Punktverteilung erfolgen, da ansonsten die Gefahr besteht, dass überhaupt kein Maximalwert für den Z-Wert gefunden werden kann. Ferner spielt die wählbare Anfangsentfernung auch eine wichtige Rolle bei der Bestimmung des optimalen Suchradius. Für ihre Wahl kann die Bodenauflösung des verwendeten Sensors sowie eine Kenntnis über das Bodenbewegungsgeschehen herangezogen werden.

6.4. Local-Moran's-Index - Clusteranalyse mit Zeitreihen-Attributen

Da das Local-Moran's-Index - Clusterverfahren für eine raumzeitliche Detektion von Anomalien verwendbar ist und die PS-Punkte eine raumzeitliche Datenstruktur in Form von georeferenzierten Zeitreihen besitzen, sind nach der Festlegung des Nachbarschaftswertes (s. Kapitel 6.3) in einem weiteren Schritt geeignete Attribute zur Charakterisierung der PS-Punkte auszuwählen. Diese Attribute sollen die wesentlichen Charakteristika der Höhenänderungszeitreihe von PS-Punkten beschreiben und zur Autokorrelationsberechnung benachbarter Punkte benutzt werden.

Exemplarisch sind in Abbildung 60 ausgewählte Eigenschaftswerte für eine Höhenänderungszeitreihe eines PS-Punktes aus dem Radarsat-2 Testdatensatz dargestellt.

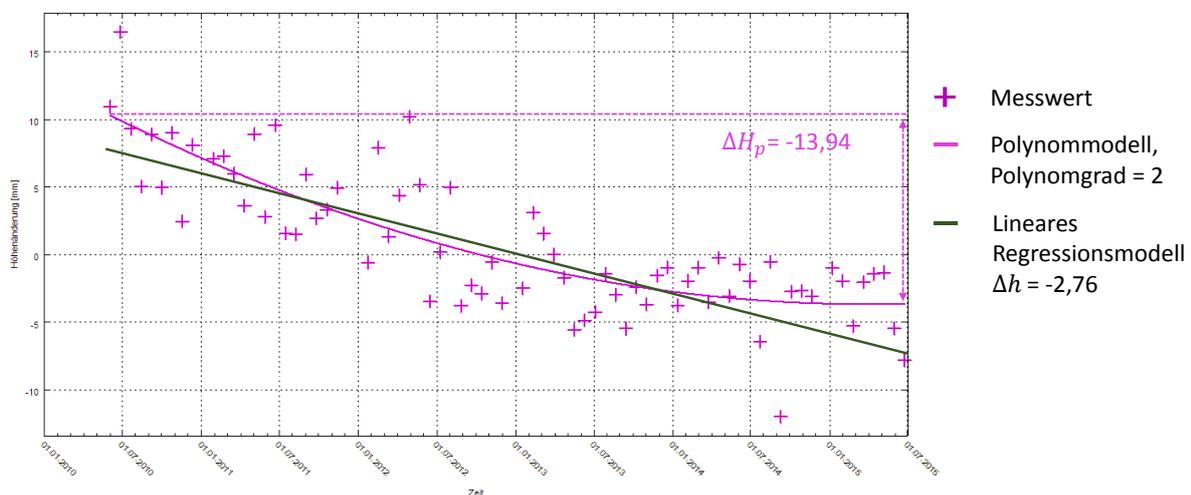


Abbildung 60: Ausgewählte Eigenschaftswerte zur Charakterisierung einer Höhenänderungszeitreihe eines PS-Punktes

Wie die Abbildung zeigt, können zur Charakterisierung eine lineare Höhenänderungsrate, der Polynomgrad eines bestangepassten Trendpolynoms sowie die Höhenänderung, abgeleitet aus den Funktionswerten des Trendpolynoms, zwischen dem ersten und letzten Messdatum verwendet werden.

- Lineare Höhenänderungsrate (mm/Jahr):

Der Eigenschaftswert wurde mittels zweidimensionaler Regressionsanalysen der Zeitreihen durch das GAMMA Modul „Interferometric Point Target Analysis“ (IPTA) als Deformationsrate (in LOS-Richtung) generiert [WERNER 2003] und in eine Höhenänderungsrate umgewandelt. In Abbildung 60 beträgt die lineare Höhenänderungsrate beispielsweise ca. 2,76 mm/Jahr (markiert durch die dunkelgrüne Linie).

- Polynomgrad eines optimalen Trendpolynoms:

Die im Rahmen der PSI-Auswertung ermittelten Höhenänderungszeitreihen werden über eine Trendmodellierung mittels eines orthogonalen Polynommodells analysiert (s. Kapitel 2.3). Das über eine Signifikanzprüfung ermittelte bestangepasste Trendpolynom ist durch den Polynomgrad charakterisiert. So konnte in dem Beispiel in Abbildung 60 der Höhenänderungstrend mit einem Polynom zweiten Grades beschrieben werden (gekennzeichnet durch die pinke Kurve).

- Höhenänderung (mm) über die gesamte Zeitreihe, abgeleitet aus der Trendfunktion:

Zur Bestimmung der über die gesamte Zeitreihe (zwischen dem ersten und letzten Messdatum) erfolgten Höhenänderung wird die Differenz der Funktionswerte zum Anfangsdatum und zum Enddatum für jeden PS-Punkt berechnet. In Abbildung 60 beträgt die so berechnete gesamte Höhenänderung ca. 13,94 mm.

Im Folgenden wird am Beispiel der TerraSAR-X Daten die Anwendbarkeit der drei oben genannten Attribute für die Durchführung der Local-Moran's-Index - Clusteranalyse zur Detektion räumlicher Anomalien in flächenhaft vorliegenden PSI-Daten geprüft.

6.4.1. Clusteranalyse mit Attribut „Höhenänderungsrate“

In der Abbildung 61 ist das Ergebnis der Clusteranalyse mit dem Attribut „Höhenänderungsrate“ zu sehen. Insgesamt wurden ca. 1.000 Punkte von etwa 230.000 PS-Punkten als „räumliche Anomalien“ detektiert. Obwohl die meisten Anomalien erwartungsgemäß in einer Gegenüberstellung der PS-Punktzeitreihen zu erkennen sind (s. Abb. 61, Diagramm r. o.), zeigen jedoch auch einige detektierte „räumliche Anomalien“ in ihren Zeitreihen eine offensichtlich gute Übereinstimmung mit den benachbarten Punkten (s. Abb. 61, Diagramm r. u.). Auf dieses Ergebnis wird in Kapitel 6.5 näher eingegangen und zum Anlass für eine Modifikation des Moran's-Index - Clusteralgorithmus genommen.

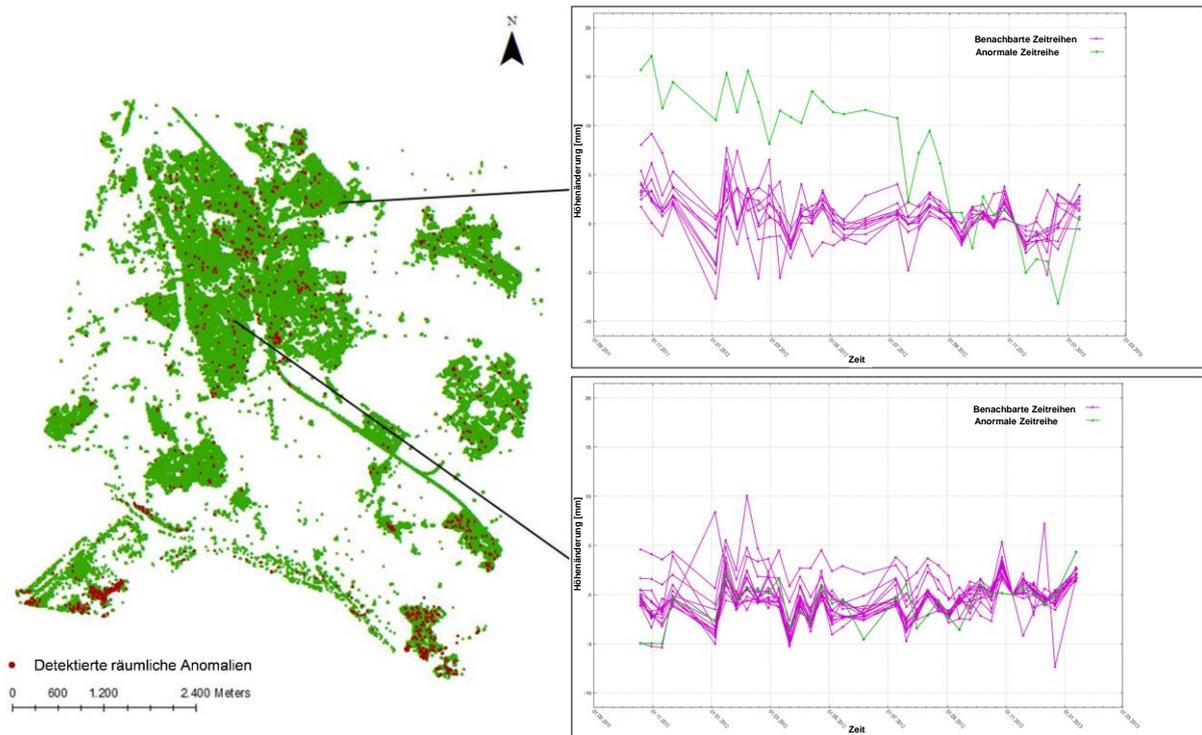


Abbildung 61: Detektierte räumliche Anomalien mit Attribut „Höhenänderungsrate“ (dargestellt durch rote Punkte) und vergleichende Darstellung ausgewählter Zeitreihen (in grün dargestellt der Zeitreihenverlauf eines als Anomalie detektierten PS-Punktes)

6.4.2. Clusteranalyse mit Attribut „Polynomgrad“

Wie das Ergebnis in Abbildung 62 zeigt, ergeben sich nach der Clusteranalyse mit dem Attribut „Polynomgrad“ deutlich mehr Ausreißer (ca. 12.000 Punkte) als mit dem Attribut „Höhenänderungsrate“ (s. Abb. 61). Ursache hierfür ist, dass im Gegensatz zur Höhenänderungsrate der Polynomgrad lediglich durch ganzzahlige (Integer-) Werte (im Beispiel von 0 bis 6) beschrieben wird. Infolgedessen zeigen viele der als „räumliche Anomalien“ identifizierten Zeitreihen keine bzw. nur geringfügige, durch den Rauschanteil in den Messwerten begründbare, Abweichungen (s. Abb. 62, Diagramm r. u.) in ihren Höhenänderungen, obwohl der Polynomgrad „zufällig“ unterschiedlich ist (s. Abb. 62, unten links).

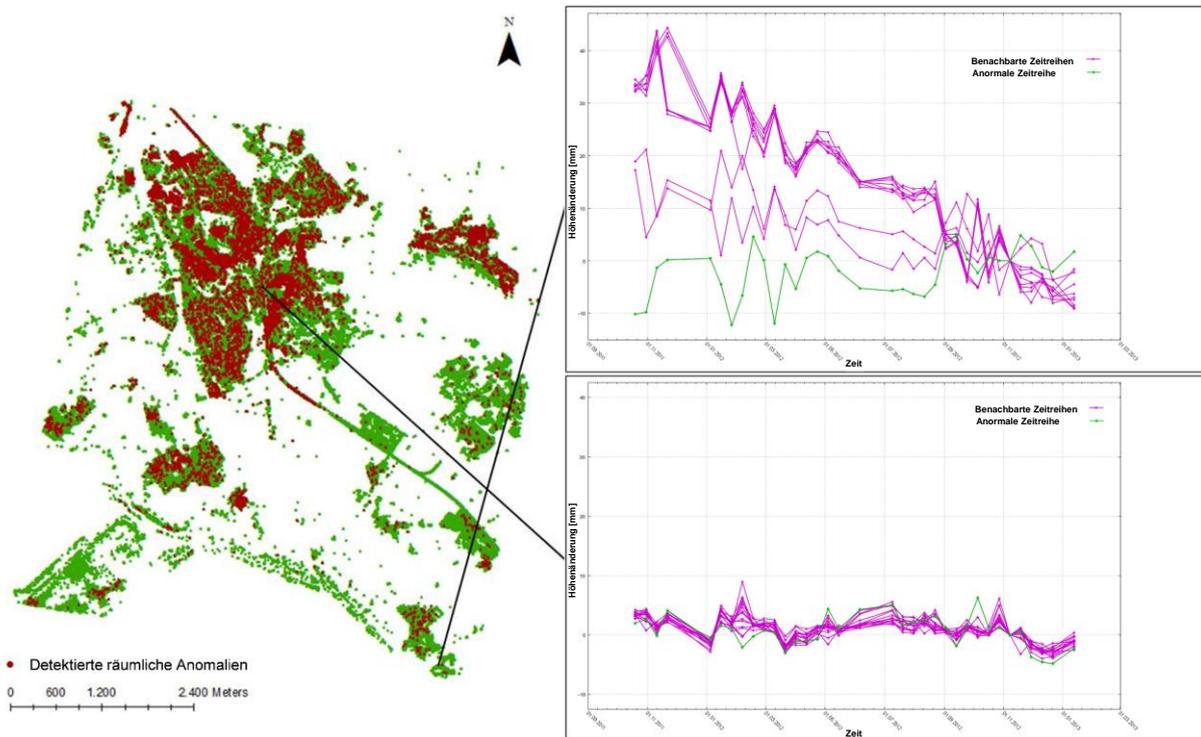


Abbildung 62: Detektierte räumliche Anomalien mit Attribut „Polynomgrad“ (dargestellt durch rote Punkte) und vergleichende Darstellung ausgewählter Zeitreihen (in grün dargestellt der Zeitreihenverlauf eines als Anomalie detektierten PS-Punktes)

6.4.3. Clusteranalyse mit Attribut „Höhenänderung“

Eine Clusteranalyse mit dem Attribut „Höhenänderung“ (abgeleitet aus den Funktionswerten der Polynomfunktion) ergab ca. 2000 PS-Punkte als „räumliche Anomalien“.

Wie der Abbildung 63 (Diagramm r. u.) zu entnehmen ist, sind auch hier PS-Punkte als „räumliche Anomalien“ detektiert worden, obwohl sich ihr Höhenänderungsverlauf kaum vom Verlauf an benachbarten Punkten unterscheidet. Dies liegt vorwiegend daran, dass die berechnete Höhenänderung nur die Differenz der Funktionswerte vom Anfangsdatum und dem Enddatum berücksichtigt und demzufolge den gesamten Höhenänderungsverlauf der zugrunde liegenden Zeitreihe nicht widerspiegeln kann. Außerdem ist das Attribut sehr empfindlich gegen Extremwerte in der Höhenänderungszeitreihe; falls Extremwerte am Anfang oder Ende der Zeitreihe auftreten, kann die Höhenänderungsfunktion durch einen „Hebeleffekt“ beeinflusst werden. Dies gilt mitunter auch für den abgeleiteten Polynomgrad (insbesondere beim Polynomgrad 1).

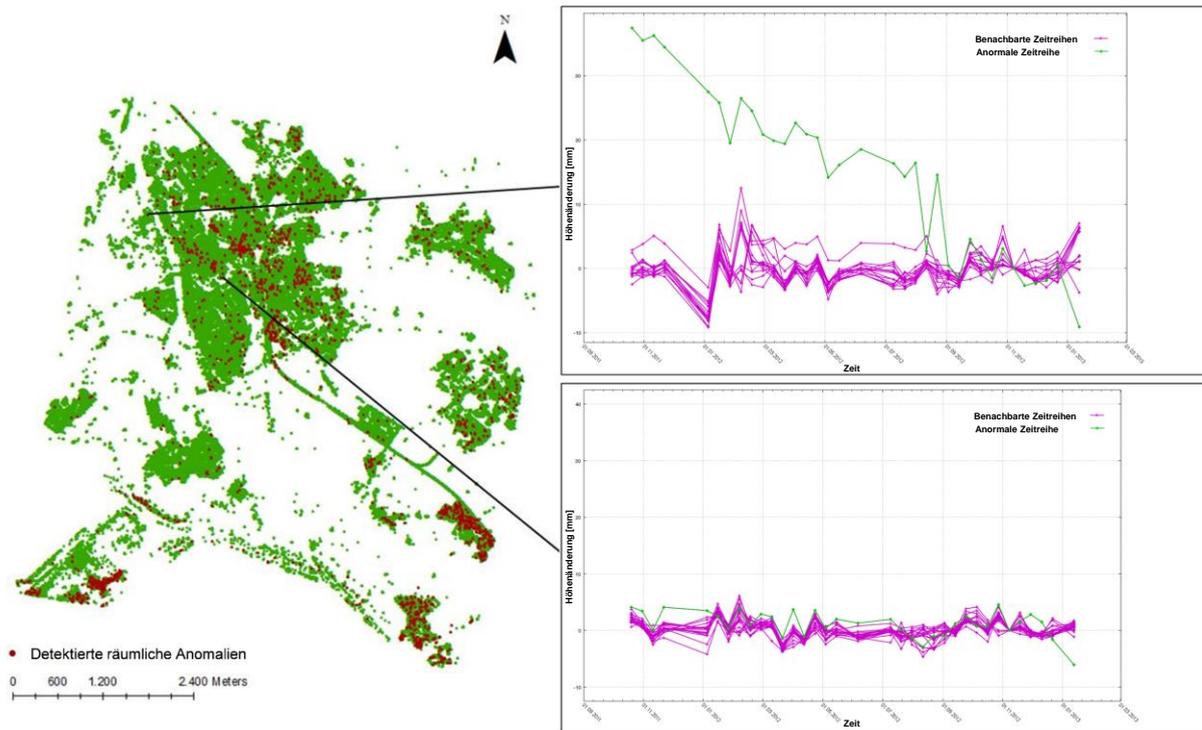


Abbildung 63: Detektierte räumliche Anomalien mit Attribut „Höhenänderung“ (dargestellt durch rote Punkte) und vergleichende Darstellung ausgewählter Zeitreihen (in grün dargestellt der Zeitreihenverlauf eines als Anomalie detektierten PS-Punktes)

Das Testbeispiel zeigt, dass grundsätzlich mit Hilfe des Local-Moran's-Index - Clusterverfahrens signifikante „räumliche Anomalien“ detektiert werden können, dabei aber der Wahl des Attributes (Eigenschaftswertes) eine wichtige Bedeutung zukommt. Im Vergleich zur Clusteranalyse mit dem Attribut „Polynomgrad“ und „Höhenänderung“ kann die Clusteranalyse mit dem Attribut „Höhenänderungsrate“ ein deutlich besseres Ergebnis mit weniger nicht auffälligen „Anomalien“ erzielen. Dies kann durch eine Einzelbetrachtung der Zufallsstichproben von detektierten Anomalien bestätigt werden. Dabei wurden jeweils 100 mit dem Attribut „Polynomgrad“, „Höhenänderung“ sowie „Höhenänderungsrate“ detektierte Anomalien zufällig in dem gesamten Untersuchungsgebiet ausgewählt und ihr Höhenänderungsverhalten untersucht. Während ca. 80% der mit dem Attribut „Polynomgrad“ und ca. 35% der mit dem Attribut „Höhenänderung“ detektierten Anomalien anhand ihres Höhenänderungsverhaltens nicht auffällig gegenüber dem ihrer benachbarten PS-Punkte sind, beträgt die Prozentzahl der nicht auffälligen Anomalien mit dem Attribut „Höhenänderungsrate“ nur knapp 10%.

6.5. Modifizierte Local-Moran's-Index - Clusteranalyse mit zeitlicher Ähnlichkeitsgewichtung

Mittels der Clusteranalyse mit dem Attribut "Höhenänderungsrate" wurden auch PS-Punkte detektiert, die sich in ihrem zeitlichen Höhenänderungsverhalten kaum von dem Verhalten benachbarter Punkte unterscheiden (s. Kapitel 6.4.1). Eine solche, möglicherweise falsch identifizierte räumliche Anomalie ist beispielhaft in Abbildung 64 (TerraSAR-X Testdaten) graphisch dargestellt.

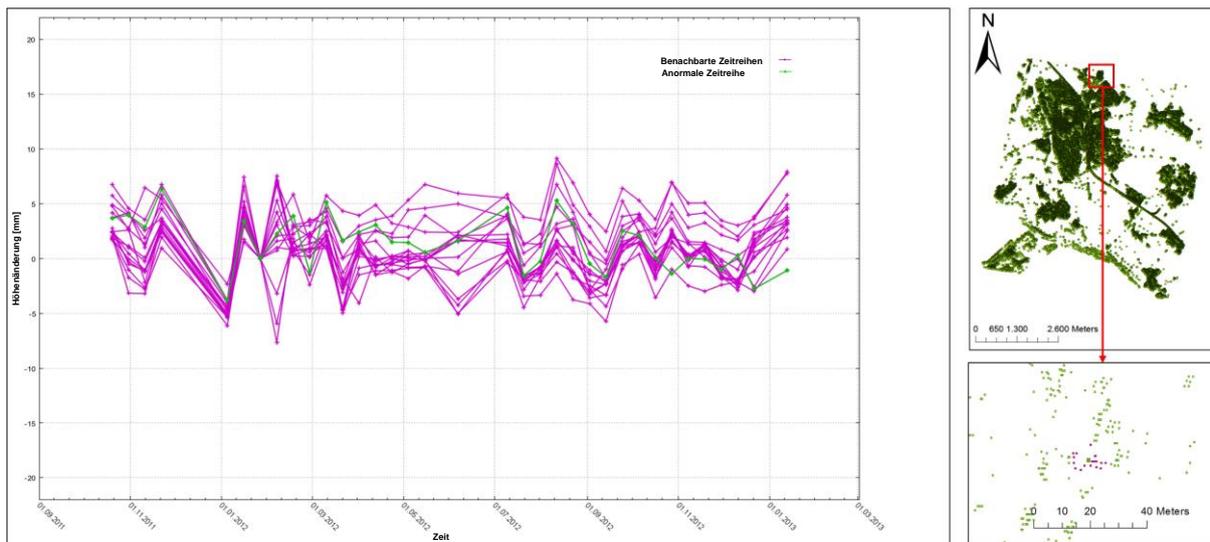


Abbildung 64: Mit Attribut „Höhenänderungsrate“ potentiell falsch detektierte räumliche Anomalie (grüner Zeitreihenverlauf) am Beispiel von TerraSAR-X Testdaten

In Abbildung 64 ist zu erkennen, dass sich der Höhenänderungsverlauf (gekennzeichnet durch grüne Linie in Abbildung 64 l.) des detektierten PS-Punktes (markiert durch grünes Rechteck in Abbildung 64 r. u.) nicht deutlich von dem der benachbarten PS-Punkte (gekennzeichnet durch pink-farbene Linien in Abbildung 64 l.) unterscheidet. Ursache hierfür kann sein (wie auch in anderen Fällen), dass die durch Bergbauaktivitäten verursachten Höhenänderungen im gesamten Untersuchungszeitraum zeitweise und punktbezogen sowohl lineare als auch nicht lineare Höhenänderungstrends aufweisen können, und deshalb im Falle eines nichtlinearen Höhenänderungsverhalten durch eine generalisierende, lineare Höhenänderungsrate, die sich aus einem linearen Regressionsmodell ergibt, verfälscht wiedergegeben werden. Offensichtlich reicht zur Beschreibung des charakteristischen Verlaufs einer Höhenänderungszeitreihe ein einziges Attribut (Merkmal) nicht aus. Da bei der Local-Moran's-Index - Clusteranalyse jedoch nur ein einziges Attribut (Eigenschaftswert) berücksichtigt werden kann, erfolgt eine Modifikation dieses Verfahrens durch Einführung

eines Gewichtungsfaktors zur Beurteilung und Quantifizierung der Ähnlichkeit der Höhenänderungsverläufe zweier PS-Punkte.

Zur Ableitung der Ähnlichkeit des Höhenänderungsverlaufes verschiedener PS-Punkte kann der Pearson Korrelationskoeffizient verwendet werden, der durch folgende Formel bestimmt wird [HARTUNG 1999]:

Formel 32:

$$r_{ij} = \frac{\sum_{t=1}^n (i_t - \bar{i})(j_t - \bar{j})}{\sqrt{\sum_{t=1}^n (i_t - \bar{i})^2 * \sum_{t=1}^n (j_t - \bar{j})^2}}$$

mit:

(i_t, j_t) : Eigenschaftswerte am gleichen Datum t der Zeitreihen an den Punkten i und j

\bar{i}, \bar{j} : Mittelwerte aller Eigenschaftswerte der Zeitreihen an den Punkten i und j

In der Formel entsprechen die Wertepaare (i_t, j_t) bezogen auf die PSI-Daten den Höhenänderungswerten der PS-Punkte i und j am gleichem Datum t . Der Pearson Korrelationskoeffizient besitzt einen Wertebereich von $[-1, 1]$; je größer der Wert ist, desto ähnlicher sind die Zeitreihen der Punkte i und j . Um den Koeffizient als Gewichtungsfaktor in die Formel zur Berechnung vom Local-Moran's-Index zu integrieren, wird der Pearson Korrelationskoeffizient vorher noch mittels Formel 33 normiert, so dass die Werte im Bereich von $[0, 1]$ liegen.

Formel 33:

$$\tilde{r}_{ij} = \frac{r_{ij} + 1}{2}$$

Die modifizierte Formel zur Berechnung des Local-Moran's-Index ergibt sich entsprechend:

Formel 34:

$$\tilde{I}_i = \frac{(x_i - \bar{x}) \sum_{j=1}^n \tilde{r}_{ij} * w_{ij} (x_j - \bar{x})}{\frac{\sum_{j=1}^n (x_j - \bar{x})^2}{n - 1}}$$

mit:

x_i : Wert der beobachteten Variable am PS-Punkt i ,

x_j : Wert der beobachteten Variablen am PS-Punkt j (j ist ein Nachbar von i),

\bar{x} : Mittelwert der beobachteten Variablen aller PS-Punkte (ohne Punkt i),

w_{ij} : die räumliche Gewichtung zwischen PS-Punkt i und j ,

\widetilde{r}_{ij} : die zeitliche Gewichtung zwischen PS-Punkt i und j

Damit können nun auch die zeitlichen Korrelationen (Ähnlichkeiten) zwischen dem betrachteten PS-Punkt und seinen benachbarten PS-Punkten bei der Berechnung des Local-Moran's-Indizes berücksichtigt, und so eine raumzeitliche Clusteranalyse basierend auf raumzeitlicher Autokorrelation für die Suche von Anomalien ermöglicht werden.

Das Ergebnis scheint sehr vielversprechend zu sein: Anhand des modifizierten Local-Moran's-Index nach Formel 34 tritt die in Abbildung 64 detektierte Anomalie nicht mehr auf, da der Indexwert durch die Berücksichtigung der zeitlichen Gewichtung kleiner wurde und im Vergleich zu seinen Nachbarn nun nicht mehr als signifikant unterschiedlich angesehen werden kann.

6.6. Kleinräumige Clusteranalyse

Da gemäß Formel 34 zur Berechnung des Local-Moran's-Index der Mittelwert (\bar{x}) der Variablenwerte aller PS-Punkte im Untersuchungsgebiet eingeht, wurde untersucht, ob ggf. durch diese „Globalisierung“ lokale Abweichungen unerkannt bleiben können. Hierzu wurden PS-Punkte, die nach der Clusteranalyse positive Indexwerte erhielten und demnach als Punkte eines Clusters einzustufen waren, näher untersucht. Es zeigte sich, dass einzelne dieser Punkte unerwarteter Weise deutlich bezüglich der Höhenänderungsraten von PS-Punkten in ihrer Umgebung abweichen. In Abbildung 65 ist exemplarisch die Höhenänderungszeitreihe eines solchen PS-Punktes, der nach der Clusteranalyse einen positiven Indexwert erhielt, mit den Zeitreihen der PS-Punkte in seiner Nachbarschaft gegenübergestellt. Da dieser PS-Punkt mehr als 500 Nachbarpunkte besitzt, wurden für jeden Messzeitpunkt die Zeitreihenwerte der benachbarten PS-Punkte gemittelt und ein mittlerer Zeitreihenverlauf dargestellt, um den Unterschied zwischen den Höhenänderungszeitreihen graphisch besser erkennen zu können.

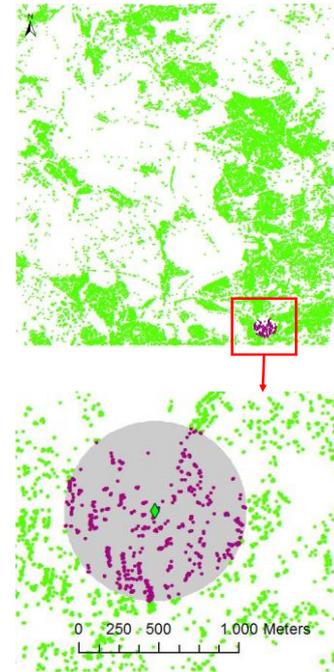
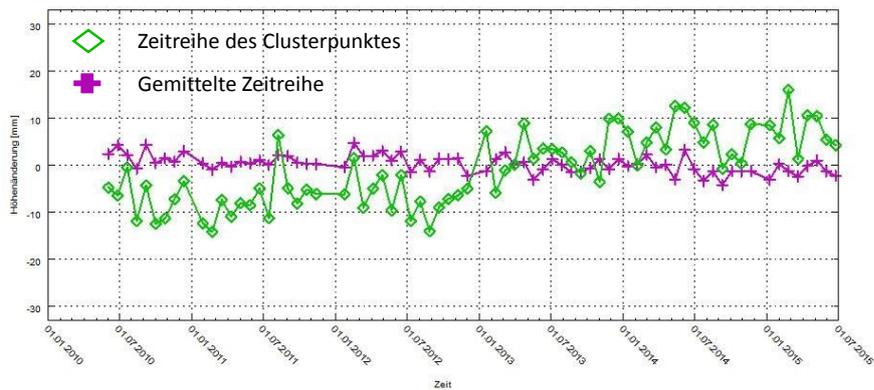


Abbildung 65: Höhenänderungs-Zeitreihe eines ausgewählten Clusterpunktes im Vergleich zur mittleren Zeitreihe benachbarter PS-Punkte

Wie die Abbildung zeigt, lassen sich über den gesamten Betrachtungszeitraum deutlich Abweichungen zwischen der Zeitreihe des ausgewählten Clusterpunktes und der gemittelten Zeitreihe seiner Nachbarn erkennen. Die Höhenänderungs-Zeitreihe des Clusterpunktes beinhaltet eine deutliche Hebung über den Untersuchungszeitraum (etwa 15 mm), während die gemittelte Zeitreihe nahezu keine Höhenänderung (bzw. tendenziell eine Senkung) aufweist. Somit müsste diese Zeitreihe aufgrund des abweichenden Höhenänderungsverhaltens als anormale Zeitreihe deklariert werden.

Maßgeblich für die Entscheidung, ob der betrachtete PS-Punkt eine anormale Zeitreihe besitzt, ist das Vorzeichen des Local-Moran's-Index. Da der Nenner in Formel 34 immer positiv ist, bestimmt der Zähler das Vorzeichen des Index. Ein positives Vorzeichen ergibt sich, wenn $(x_i - \bar{x})$ und $(\sum_{j=1}^n \tilde{r}_{ij} * w_{ij}(x_j - \bar{x}))$ gleiches Vorzeichen besitzen, bzw. die Abweichung des untersuchten PS-Punktes vom Mittelwert und die Summe aller Abweichungen der benachbarten PS-Punkte vom Mittelwert gleiches Vorzeichen besitzen (da \tilde{r}_{ij} und w_{ij} größer Null sind). Hieraus ergibt sich die Vermutung, dass der globale Mittelwert der Höhenänderungsrate (\bar{x}) eine entscheidende Rolle spielt. Dieser globale Mittelwert als Mittelwert der Höhenänderungsraten aller PS-Punkte in dem gesamten Untersuchungsgebiet beträgt in dem angeführten Beispiel -2,03, während die Höhenänderungsrate für den untersuchten Clusterpunkt einen Wert von +3,84 ($x_i = 3,84$) besitzt und somit die Differenz $(x_i - \bar{x})$ positiv ist. Da lediglich 21 von insgesamt 564 Nachbarpunkten kleinere

Höhenänderungsraten (in einem Wertebereich von -5,36 bis -2,09) als der Mittelwert besitzen, ist der Faktor $(\sum_{j=1}^n \widetilde{r}_{ij} * w_{ij}(x_j - \bar{x}))$ im Zähler auch positiv. Daher erhält das Produkt $((x_i - \bar{x}) \sum_{j=1}^n \widetilde{r}_{ij} * w_{ij}(x_j - \bar{x}))$ und somit der Local-Moran's-Index einen positiven Wert und der untersuchte PS-Punkt wird gemäß der modifizierten Local-Moran's-Index – Clusteranalyse nicht als Anomalie detektiert.

Da nicht zuletzt sich durch dieses Beispiel feststellen lässt, dass die Anwendung eines globalen Mittelwertes des betrachteten Attributes in Einzelfällen PS-Punkte mit anormalen Zeitreihen unentdeckt lässt, erfolgt zur Verbesserung des Analyseergebnisses eine Ergänzung des Verfahrens durch eine kleinräumige Clusteranalyse. Hierzu werden die Höhenänderungsraten der PS-Punkte, die nach der modifizierten Local-Moran's-Index – Clusteranalyse als signifikante Clusterpunkte (mit positiven Indexwerten) identifiziert wurden, mit dem Werteintervall, das durch den Mittelwert \bar{x} und die Standardabweichung σ der Höhenänderungsraten ihrer Nachbarpunkte (ohne den zu untersuchenden Clusterpunkt) innerhalb des festgelegten Suchradius definiert ist, verglichen. Ein Clusterpunkt kann als eine Anomalie erkannt werden, wenn er nicht in das Werteintervall $[\bar{x} - 4\sigma, \bar{x} + 4\sigma]$ fällt, da sich bei Annahme einer Normalverteilung nahezu 99,99% aller Höhenänderungsraten im dem 4σ -Bereich befinden sollten. Die zu den 0,01% gehörenden Werte unterscheiden sich stark vom Mittelwert und somit deutlich von den anderen Werten; sie gehören nicht zu der statistischen Verteilung der übrigen Werte ihrer benachbarten PS-Punkte und können dementsprechend als Anomalien deklariert werden.

Für das obengenannte Beispiel ist in Abbildung 66 die Häufigkeitsverteilung der Höhenänderungsraten der Nachbarpunkte des untersuchten Clusterpunktes graphisch dargestellt. Sie zeigt, dass die Höhenänderungsraten der Nachbarpunkte annähernd normalverteilt sind. Bei einem Mittelwert von -0,65 und einer Standardabweichung von 0,83 zeigt das Beispiel, dass die Höhenänderungsrate (3,84) des untersuchten Clusterpunktes nicht zu dem Werteintervall $[\bar{x} - 4\sigma = -3,97, \bar{x} + 4\sigma = 2,67]$ gehört, und somit der Clusterpunkt als Anomalie erkannt werden kann.

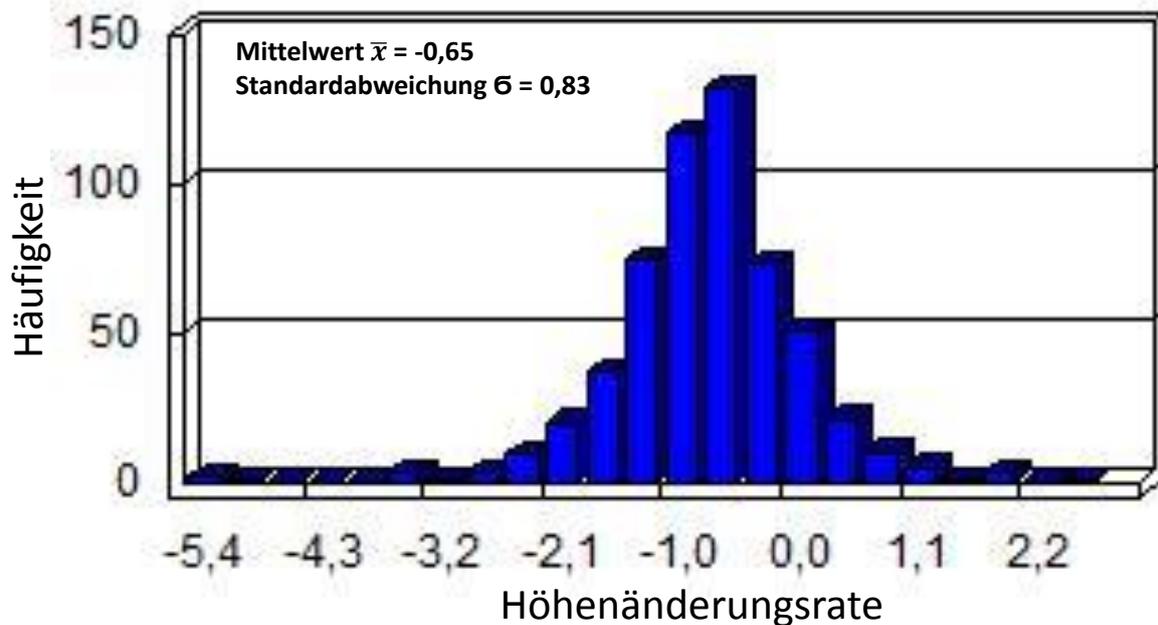


Abbildung 66: Häufigkeitsverteilung der Höhenänderungsraten der Nachbarpunkte des untersuchten Clusterpunktes

6.7. Verfahrensablauf der modifizierten Local-Moran's-Index - Clusteranalyse

Zusammenfassend ist der Verfahrensablauf der modifizierten Local-Moran's-Index - Clusteranalyse zur Identifizierung von räumlichen Anomalien in Abbildung 67 dargestellt.

Zuerst wird anhand der globalen Moran's Autokorrelation eine Musteranalyse vorgenommen um festzustellen, ob in dem gesamten Untersuchungsgebiet grundsätzlich eine räumliche Autokorrelation besteht. Eine erfolgreiche Durchführung der Clusteranalyse setzt die Existenz der räumlichen Autokorrelation voraus. Im Falle einer vorhandenen räumlichen Autokorrelation kann abschließend eine Local-Moran's-Index - Clusteranalyse ausgeführt werden. Ansonsten muss das gesamte Untersuchungsgebiet angepasst bzw. in Teiluntersuchungsgebiete unterteilt werden, solange bis die Voraussetzung erfüllt ist. Im Vergleich zu dem allgemeinen Verfahrensablauf wird hierbei der optimale Suchradius für die Nachbarschaftsbildung mittels inkrementeller räumlicher Autokorrelation ermittelt. Dabei kann die mittlere Entfernung zum nächsten Nachbarn als ein möglicher Referenzwert für die Bestimmung der Anfangsentfernung angesehen werden. Zugleich muss noch eine genügende Anzahl von Nachbarn sichergestellt werden. Wenn es nicht möglich ist, einen optimalen Suchradius für das gesamte Untersuchungsgebiet zu ermitteln, kann dieser als Mittelwert (oder Medianwert)

der ermittelten Suchradien der ausgewählten Teilgebiete festgelegt werden. Hierbei sollte das ausgewählte Teilgebiet über eine relativ gleichmäßige Punktverteilung verfügen und möglichst nicht isolierte PS-Punkte beinhalten. Ferner erwies sich die Höhenänderungsrate als charakteristisches Suchmerkmal (Attribut) geeignet und die Einführung des Pearson Korrelationskoeffizienten als normierter Gewichtungsfaktor zur Modellierung der zeitlichen Autokorrelationen zwischen den benachbarten PS-Punkten als notwendig.

Darüber hinaus wurde die Auswahlmöglichkeit geschaffen, die abgeleiteten räumlichen und zeitlichen Gewichtungen zu normieren, indem jede Gewichtung des Nachbarpunktes durch die Summe aller Gewichtungen der benachbarten PS-Punkte innerhalb des Suchradius dividiert wird. Damit kann sichergestellt werden, dass die Summe aller Gewichtungen für jeden untersuchten PS-Punkt „Eins“ ergibt. Die so berechnete proportionale Gewichtung wird als „normierte Gewichtung“ für die nachfolgende Clusteranalyse eingesetzt.

Nach der Durchführung der modifizierten Local-Moran's-Index – Clusteranalyse wird für die detektierten signifikanten Clusterpunkte noch eine kleinräumige Clusteranalyse vorgenommen, um die Nachteile einer Anwendung eines globalen Parameterwertes bei der Detektion von anormalen PS-Punkten in lokalen Untersuchungsteilbereichen zu beheben.

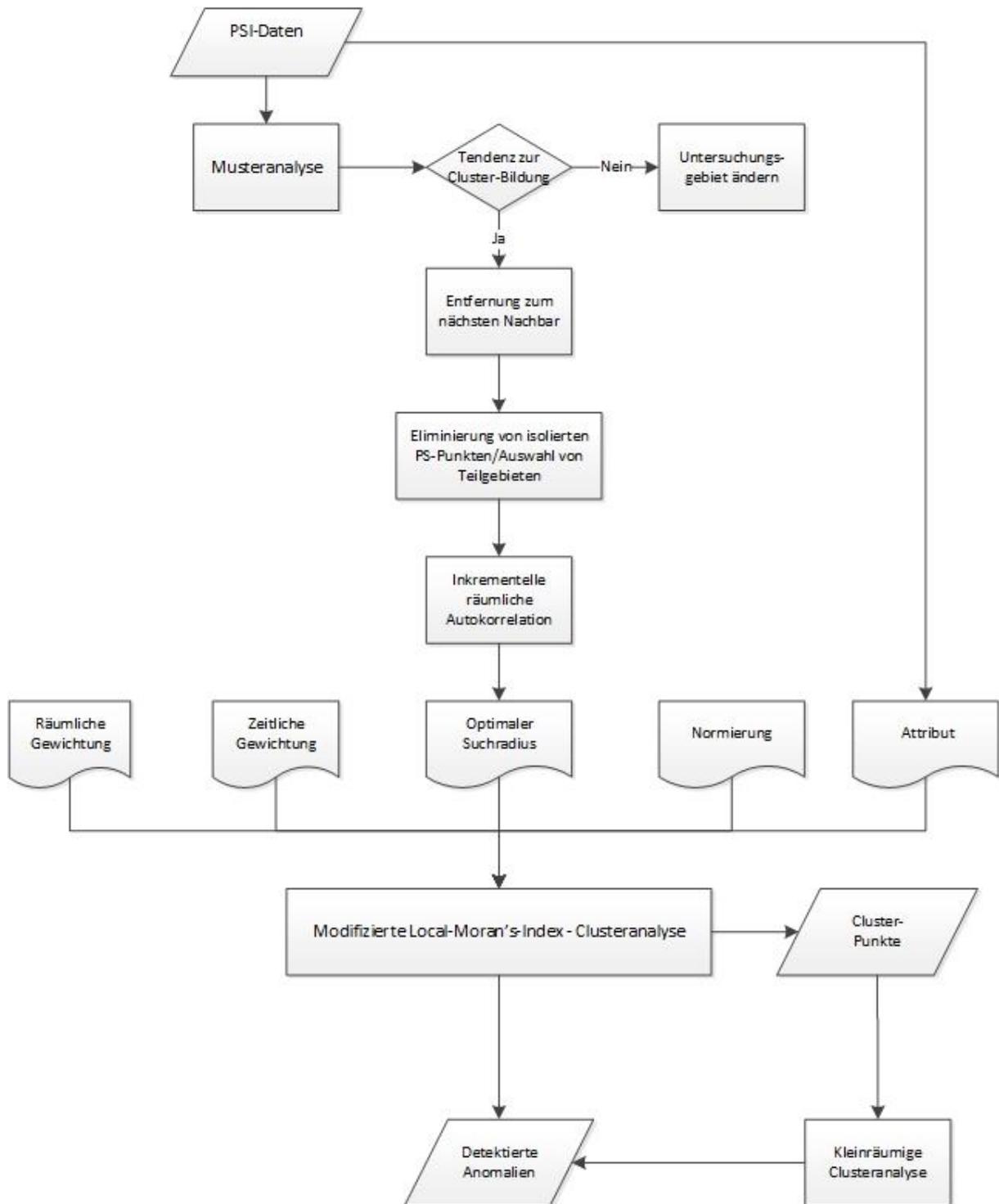


Abbildung 67: Flussdiagramm des Verfahrensablaufs bei Anwendung des modifizierten Local-Moran's-Index - Clusterverfahrens zur Detektion von Anomalien in PS-Punkt-Daten

7. Validierung des modifizierten Local-Moran's-Index - Verfahrens

Im Folgenden wird das modifizierte Local-Moran's-Index - Verfahren zur Identifizierung von anormalen Zeitreihen an PS-Punkten in flächenhaft vorliegenden PSI-Ergebnissen beispielhaft mit den Radarsat-2-Testdaten (s. Kapitel 6.2.2) erprobt. Hierfür sind alle PS-Punkte innerhalb des gesamten Untersuchungsgebietes für die Untersuchung berücksichtigt worden.

7.1. Musteranalyse

Für die Verwendung des modifizierten Local-Moran's-Index - Verfahrens wird grundsätzlich angenommen, dass die betrachteten Attribute der Untersuchungsobjekte in dem Untersuchungsgebiet autokorreliert sein müssten. Zur Überprüfung der Existenz von Autokorrelationen zwischen den Untersuchungsobjekten kommt im Grunde genommen eine Musteranalyse zum Einsatz, die oft auf einer globalen Autokorrelation basiert. Mit der Musteranalyse kann ausgehend von der Nullhypothese festgestellt werden, ob, bezüglich dem betrachteten Attribut, die Untersuchungsobjekte statistisch zufällig oder entsprechend einem räumlichen Prozess, z.B. einer „Clusterbildung“, in dem gesamten Untersuchungsgebiet verteilt sind. Sie bildet somit den Ausgangspunkt für die weitergehende Clusteranalyse und sollte prinzipiell zuerst durchgeführt werden.

Die Null-Hypothese bei der Musteranalyse geht davon aus, dass die Untersuchungsobjekte in Bezug auf das untersuchte Attribut zufällig im Untersuchungsgebiet verteilt sind; es gibt somit grundsätzlich keinen signifikanten räumlichen Prozess (Trend) im Datensatz und das beobachtete Muster entspricht lediglich einer der vielen möglichen Realisationen entsprechend einer zufälligen räumlichen Verteilung. Nach Durchführung der Musteranalyse werden letztlich ein Indexwert, ein Z-Wert und ein p-Wert, zur Quantifizierung der globalen räumlichen Autokorrelation ausgegeben. Mit ihnen wird es möglich zu untersuchen, ob die Null-Hypothese abgelehnt werden kann und sich die Untersuchungsobjekte in einem räumlichen Prozess befinden.

Das Ergebnis einer Musteranalyse am Beispiel der Radarsat-2-Testdaten ist in Abbildung 68 dargestellt. Hierbei werden die Höhenänderungsraten der PS-Zeitreihen als Attribut zur Analyse herangezogen (s. Kapitel 6.4).

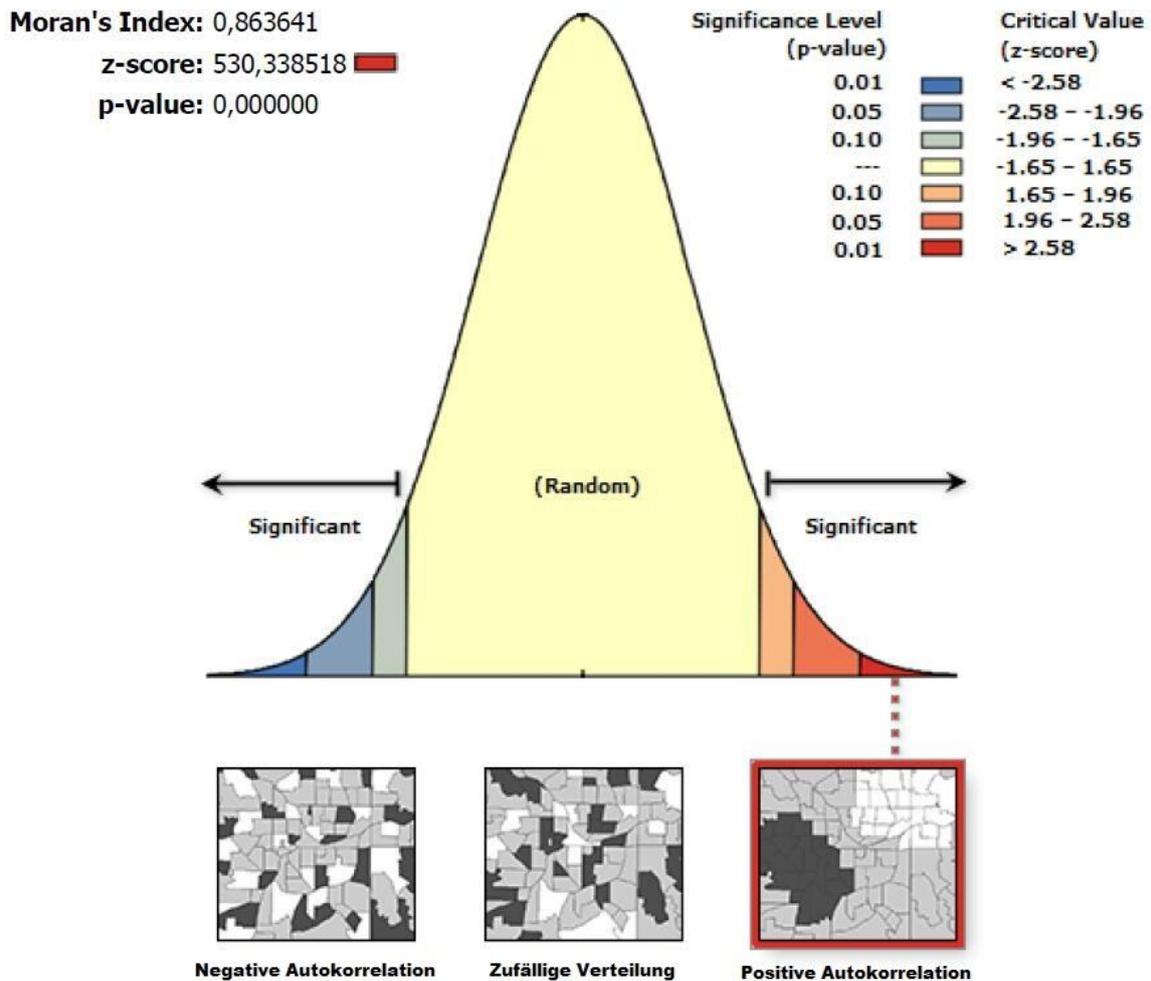


Abbildung 68: Musteranalyse von Radarsat-2-Testdaten

Wie der Abbildung 68 zu entnehmen ist, besitzt der globale Moran's-Index einen Wert größer als 0,86. Dies deutet darauf hin, dass eine starke positive Autokorrelation der Höhenänderungsraten der PS-Punkte in dem gesamten Untersuchungsgebiet besteht (s. Kapitel 5.3.2). Mit einem Z-Wert von ca. 530 und einem p-Wert von ungefähr 0 erweist sich dieses Ergebnis als statistisch hoch signifikant bei einem Signifikanzniveau von 1%. Somit kann die Null-Hypothese abgelehnt werden; die Höhenänderungsraten der PS-Punkte in dem Untersuchungsgebiet sind miteinander autokorreliert. Somit kann die Höhenänderungsrate als Attribut für die weitere Clusteranalyse verwendet werden.

7.2. Bestimmung eines optimalen Suchradius

Wie bereits in Kapitel 6.3 beschrieben, kann man mittels der lokalen räumlichen Autokorrelation den optimalen Suchradius zur Modellierung von räumlichen Beziehungen für

die Local-Moran's-Index - Clusteranalyse ermitteln. Eine praktische Untersuchung ergab jedoch, dass es bei gesamtheitlicher Betrachtung des Untersuchungsgebietes nicht möglich war, einen optimalen Suchradius abzuleiten. Durch Auswahl zusammenhängender Teilgebiete konnte jedoch in diesen für die Radarsat-2 Testdaten ein optimaler Suchradius zwischen 32 und 40 m (Median 34 m) ermittelt werden.

Unter Nutzung dieses Suchradius von 34 m erfolgte eine Anwendung auf das gesamte Untersuchungsgebiet und die anschließende modifizierte Local-Moran's-Index - Clusteranalyse. Als Ergebnis dieser Clusteranalyse ergab sich jedoch nur ein einziger anomaler PS-Punkt. Ursache für dieses unrealistische Ergebnis kann der, für das gesamte Untersuchungsgebiet zu klein gewählte Suchradius sein. So zeigte sich, dass 11526 Punkte weniger als die geforderten 8 benachbarten PS-Punkte und 3830 Punkte keine Nachbarpunkte innerhalb des Suchradius besitzen. . Damit fielen ca. 17,5% aller PSI Punkte (ca. 70.000) für die Anwendung des modifizierten Local-Moran's-Index - Clusterverfahrens aus.

Obwohl der Medianwert von 34 m durch die lokale Autokorrelation als optimaler Suchradius angesehen werden durfte, ist dieser Suchradius unter Beachtung der beiden Kriterien (lokale Autokorrelation und min. acht Nachbarn) jedoch für eine Anwendung auf das gesamte Untersuchungsgebiet nicht geeignet. Daher besteht die Notwendigkeit, den Suchradius für eine Anwendung auf das gesamte Untersuchungsgebiet zu vergrößern, und damit möglichst viele PS-Punkte in die Clusteranalyse einzubeziehen.

Um einen in diesem Sinne für das gesamte Untersuchungsgebiet optimalen Suchradius zu bestimmen, erfolgte bei intervallweiser Vergrößerung des Suchradius die Bestimmung von PS-Punkten, die mindestens 8 Nachbarn besitzen. In Tabelle 15 und Abbildung 69 ist das Ergebnis dieser Untersuchung dargestellt.

Erwartungsgemäß wächst mit zunehmendem Suchradius die Anzahl der PS-Punkte mit min. acht Nachbarn. Bei Verwendung eines Suchradius von beispielsweise 100 m erfüllen ca. 90% aller PS-Punkte das Nachbarschaftskriterium (s. Tabelle 15). Ab dieser Entfernung verringert sich deutlich die Zunahme der Punktzahl. Bei zu einem Suchradius von 570 m verfügen nahezu 99,9% aller PS-Punkte über min. acht Nachbarn; lediglich 65 Punkte erfüllen dieses Kriterium nicht. Die räumliche Verteilung dieser Punkte ist in Abbildung 70 (orangene Punkte) dargestellt. Es handelt sich hauptsächlich um räumlich isolierte PS-Punkte, für die aufgrund der räumlichen Distanz eine sichere Identifizierung von anomalen PS-Zeitreihen nicht möglich ist. Um einerseits möglichst viele Punkte in die Analyse mit einzubeziehen und zu prüfen, andererseits wegen eines zu großen Punktabstandes die Gefahr einer Fehlinterpretation zu reduzieren, wurde als Kompromiss für dieses Untersuchungsgebiet ein Suchradius von 570 m für die nachfolgende modifizierte Local-Moran's-Index - Clusteranalyse festgelegt.

Da diese restlichen ca. 0,1% aller PS-Punkte durch die Clusteranalyse nicht auf Anomalien getestet werden, sind sie als nicht geprüfte Daten durch ein entsprechendes Attribut zu kennzeichnen.

Tabelle 15: Anzahl der PS-Punkte mit min. 8 Nachbarn in Abhängigkeit einer intervallweisen Vergrößerung des Suchradius

Suchradius	Anzahl	Kumuliert %
0-35	16480	25,04%
36-50	15204	48,14%
51-70	16661	73,45%
71-100	10936	90,07%
101-133	3216	94,95%
134-150	702	96,02%
151-170	560	96,87%
171-210	662	97,88%
211-280	708	98,95%
281-570	625	99,90%
> 571	65	100,00%

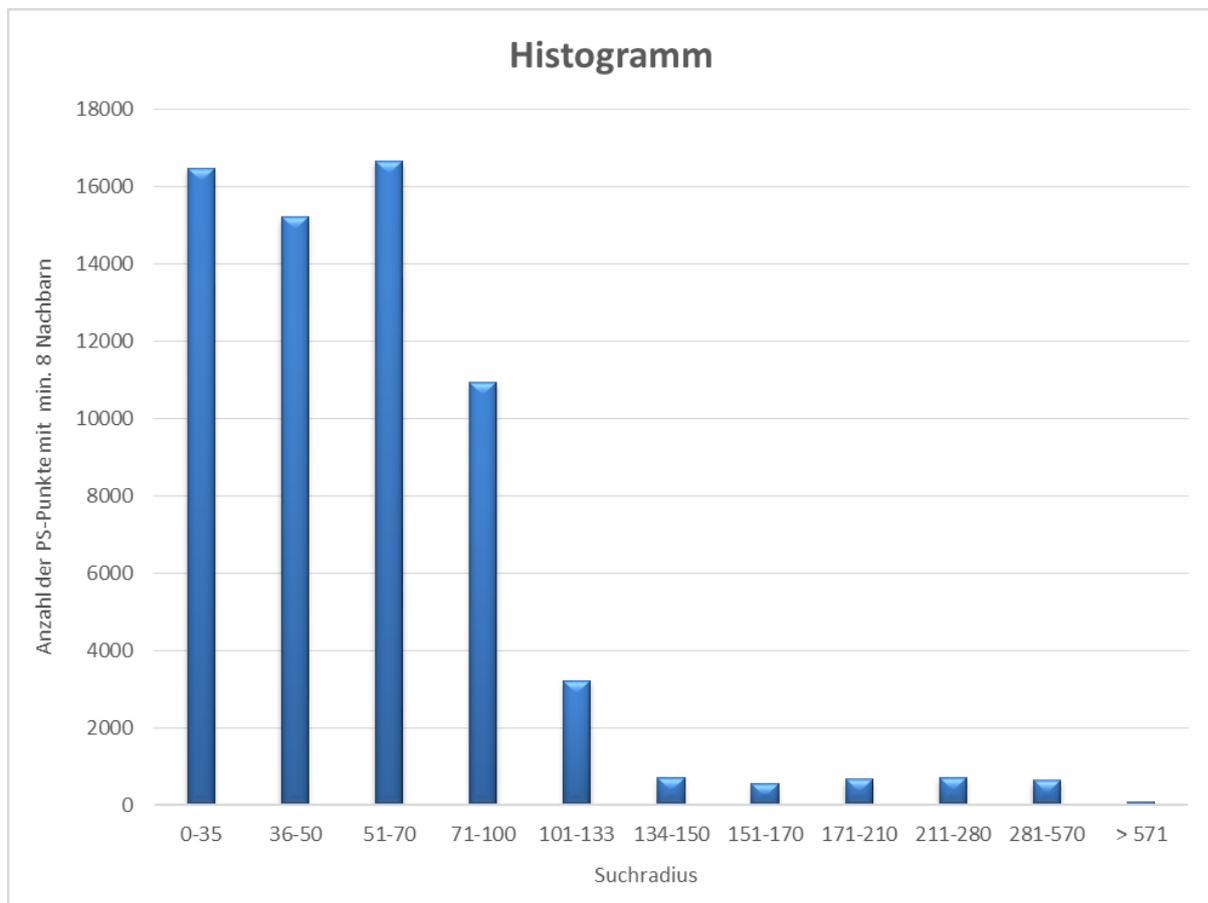


Abbildung 69: Anzahl der PS-Punkte mit min. 8 Nachbarn in Abhängigkeit einer intervallweisen Vergrößerung des Suchradius

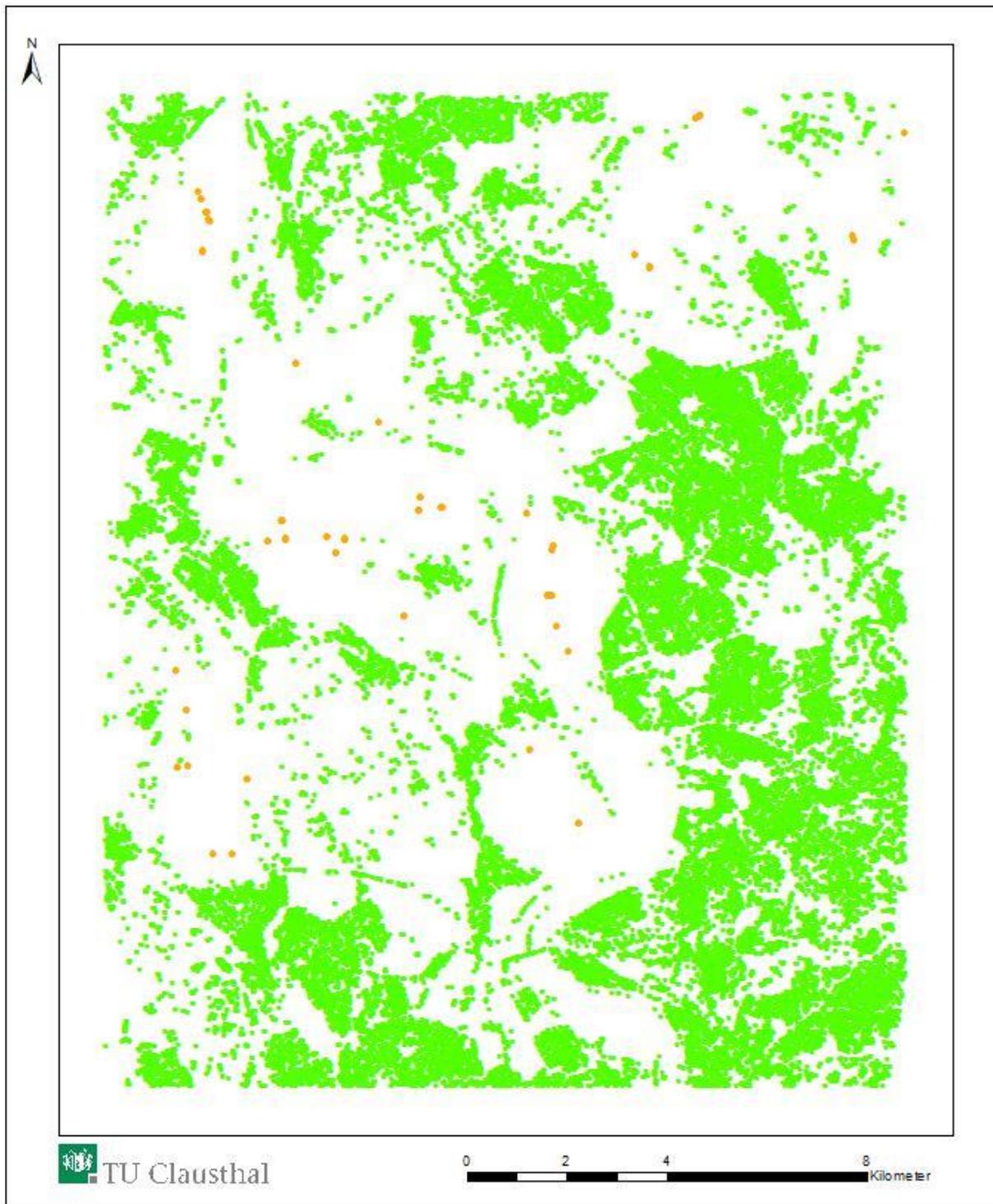


Abbildung 70: PS-Punkte (orange Punkte), die nicht über min. acht benachbarte PS-Punkte innerhalb des Suchradius verfügen

7.3. Detektion von anormalen Zeitreihen an PS-Punkten

Für die Detektion von anormalen Zeitreihen an PS-Punkten am Beispiel der Radarsat-2-Testdaten wird das modifizierte Local-Moran's-Index - Clusterverfahren mit den folgenden Eingangsparametern durchgeführt:

- **Suchradius:** 570 m
- **Betrachtetes Attribut:** Höhenänderungsrate
- **Gewichtungsmethode :** räumliche und zeitliche Gewichtung
- **Normierung:** aller Gewichte
- **Signifikanzniveau:** 5%

Mittels der modifizierten Local-Moran's-Index - Clusteranalyse konnten insgesamt 236 PS-Punkte mit anormalen Zeitreihen identifiziert werden, deren räumliche Lage in Abbildung 71 (rot markierte Punkte) dargestellt ist. Die detektierte Punktzahl entspricht etwa 0,36% aller PS-Punkte im Untersuchungsgebiet, die ein statistisch signifikant abweichendes Höhenänderungsverhalten im Vergleich zu dem ihrer räumlichen Nachbarpunkte innerhalb des Suchradius aufweisen.

Exemplarisch sind die Höhenänderungszeitreihen von vier detektierten PS-Punkten in Abbildung 72 – 75 (Zeitreihen A, B, C und D) mit den Zeitreihen der PS-Punkte in ihrer Nachbarschaft gegenübergestellt. Um die Abweichungen der Höhenänderungszeitreihen graphisch besser erkennen zu können, wurden die Zeitreihenwerte der benachbarten PS-Punkte für jeden Messzeitpunkt gemittelt dargestellt.

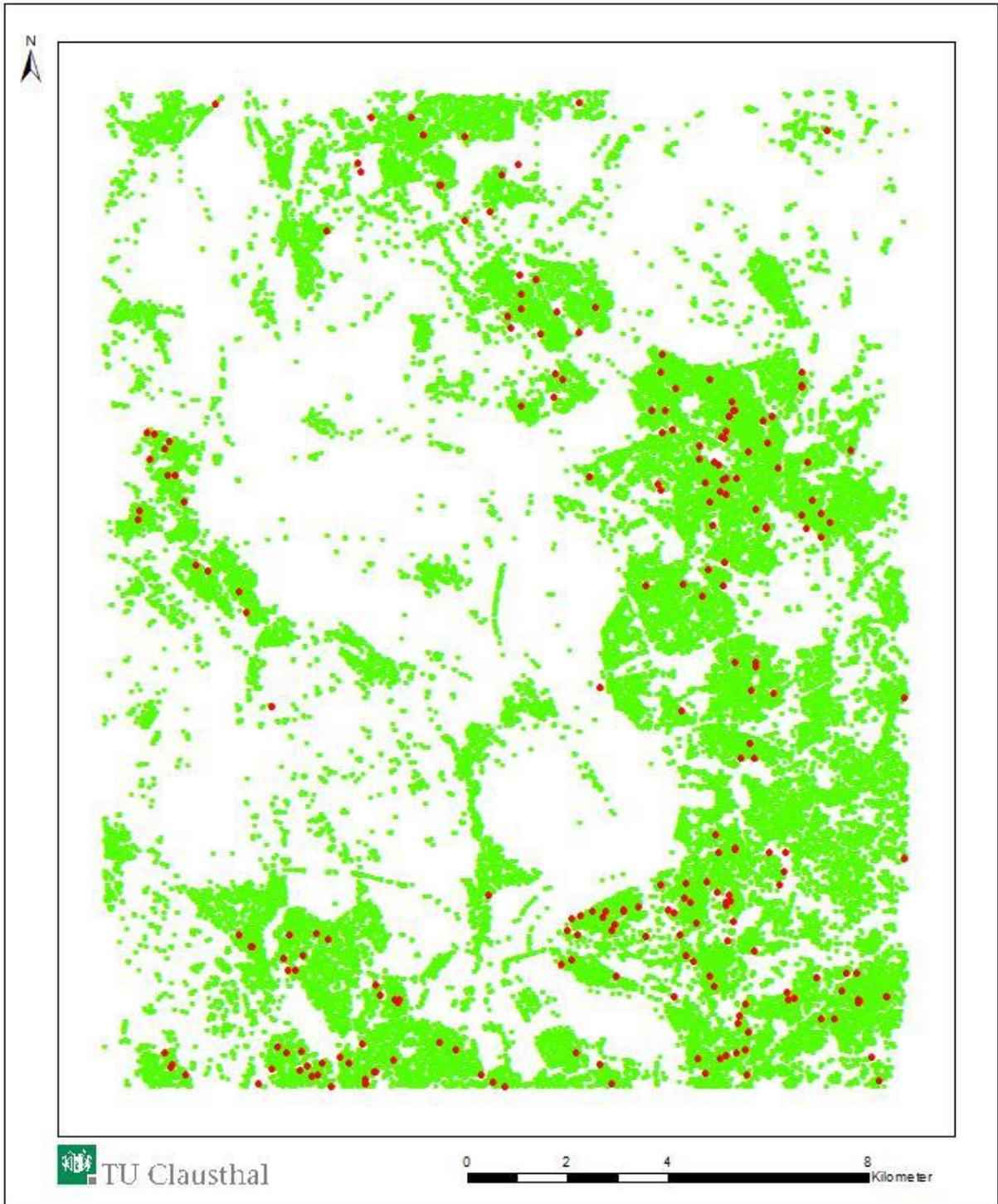


Abbildung 71: Räumliche Lage der detektierten PS-Punkte mit anormalen Zeitreihen (rot dargestellt)

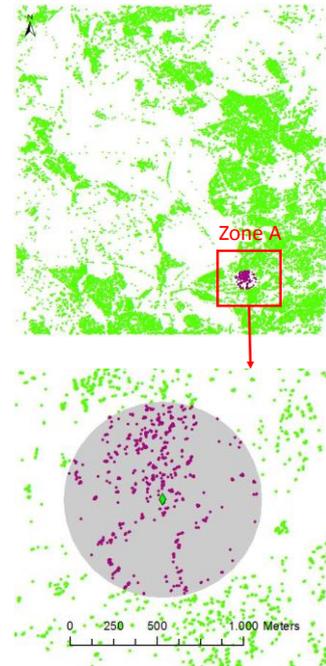
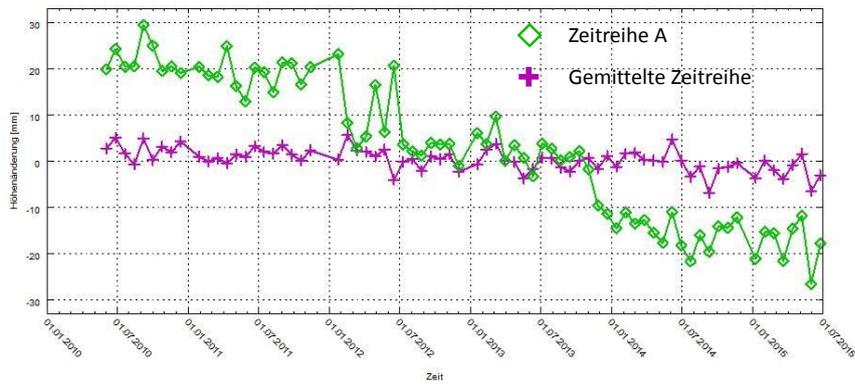


Abbildung 72: Gegenüberstellung der detektierten anomalen Zeitreihe A mit der gemittelten Zeitreihe benachbarter PS-Punkte

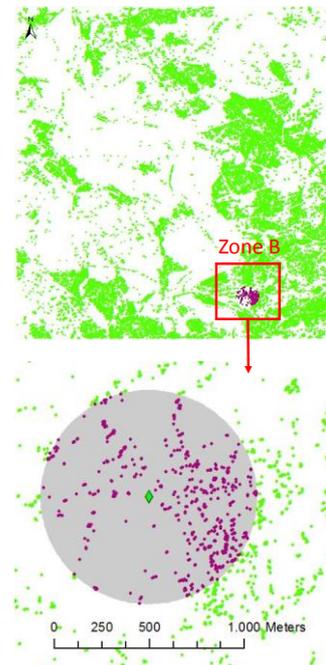
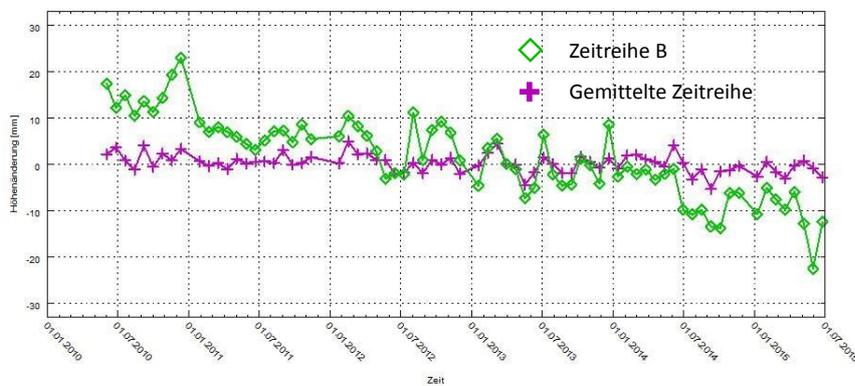


Abbildung 73: Gegenüberstellung der detektierten anomalen Zeitreihe B mit der gemittelten Zeitreihe benachbarter PS-Punkte

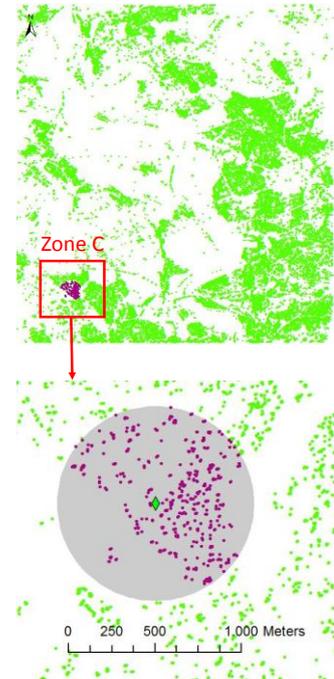
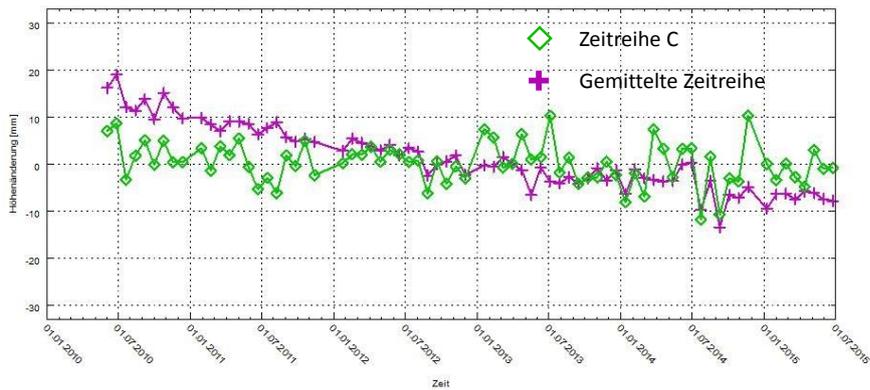


Abbildung 74: Gegenüberstellung der detektierten anomalen Zeitreihe C mit der gemittelten Zeitreihe benachbarter PS-Punkte

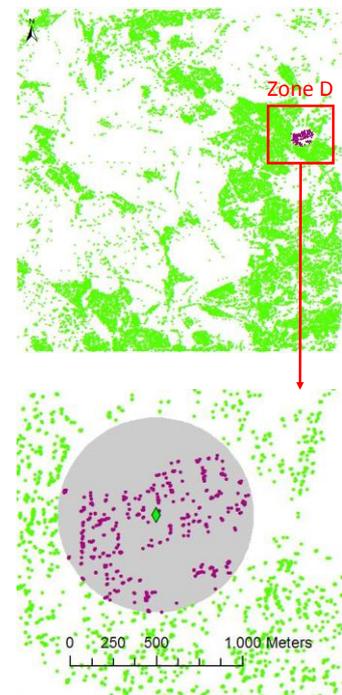
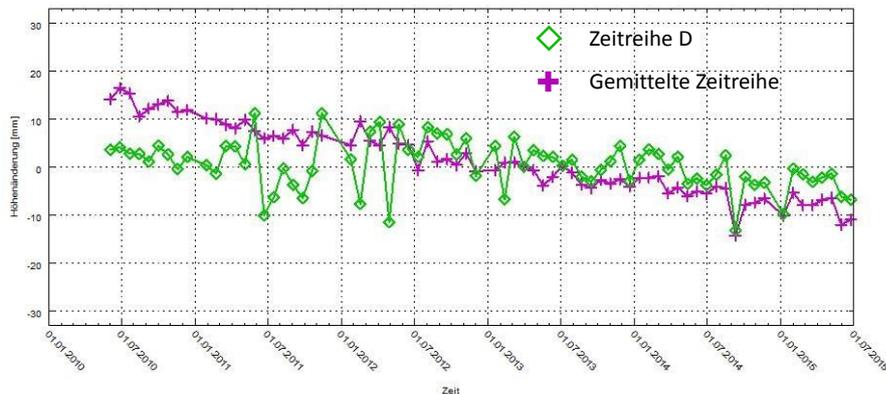


Abbildung 75: Gegenüberstellung der detektierten anomalen Zeitreihe D mit der gemittelten Zeitreihe benachbarter PS-Punkte

Die wesentlichen charakteristischen Merkmale, Höhenänderungsrate, Anzahl der benachbarten PS-Punkte, der in Abbildung 72 – 75 dargestellten anomalen Zeitreihen sowie

ihre Analyseergebnisse aus der modifizierten Local-Moran's-Index - Clusteranalyse (Local-Moran's-Index - Wert, Z-Wert und P-Wert) sind in Tabelle 16 zusammengestellt.

Tabelle 16: Charakteristische Merkmale der in Abbildung 72 – 75 dargestellten anormalen Zeitreihen

Anormale Zeitreihe	Höhenänderungsrate	Anzahl der Nachbarn	Index-Wert	Z-Wert	P-Wert
Zeitreihe A	-9,63	562	-0,75	-17,69	≈ 0
Zeitreihe B	-5,02	577	-0,44	-13,09	≈ 0
Zeitreihe C	-0,75	455	-0,29	-5,28	≈ 0
Zeitreihe D	-1,16	392	-0,25	-3,25	≈0,001

Innerhalb eines Suchradius von 570 m verfügen sie alle über mehr als 390 benachbarte PS-Punkte und erfüllen damit ausreichend die Anforderung an eine erfolgreiche Durchführung der Clusteranalyse. Obwohl sie unterschiedlich starke Höhenänderungsraten besitzen, ergaben sich für alle negative Local-Moran's-Index - Werte, was eindeutig darauf hindeutet, dass sie ein abweichendes Höhenänderungsverhalten im Vergleich zu dem ihrer benachbarten PS-Punkte aufweisen; sie sind negativ mit ihren Nachbarn autokorreliert. Darüber hinaus kann dieses Analyseergebnis mit einem Z-Wert kleiner als -3 und einem p-Wert kleiner als 0,05 als „statistisch signifikant“ (bei einem Signifikanzniveau von 5%) bezeichnet werden (vgl. Kap. 5.3).

Der visuelle Vergleich der Zeitreihen in Abbildung 72 - 75 bestätigt dieses Ergebnis; die Abweichungen zwischen den anormalen Zeitreihen der Punkte A, B, C, D und den gemittelten Zeitreihen ihrer Nachbarn sind erheblich. So zeigen die Zeitreihen A und B eine deutliche Senkung über den Untersuchungszeitraum (etwa -40 mm bzw. -30 mm), während die gemittelten Zeitreihen nahezu keine Höhenänderung aufweisen. Die Vergleiche für die Zeitreihen C und D zeigen hinsichtlich Zeitreihenverlauf und quantitativen Unterschieden ein etwas anderes Bild; beide Zeitreihen C und D beinhalten nahezu keine Senkung, während sich durch den kontinuierlichen Senkungsverlauf der gemittelten Nachbarzeitreihen eine Senkung von ca. 25 mm ergibt.

Nach der Durchführung der modifizierten Local-Moran's-Index – Clusteranalyse erfolgte anschließend die kleinräumige Clusteranalyse für die detektierten Clusterpunkte. Hierdurch konnten weitere insgesamt 114 PS-Punkte mit abweichenden Höhenänderungsraten als Anomalien entdeckt werden, deren räumliche Lage in Abbildung 76 (rot markierte Punkte) dargestellt ist.

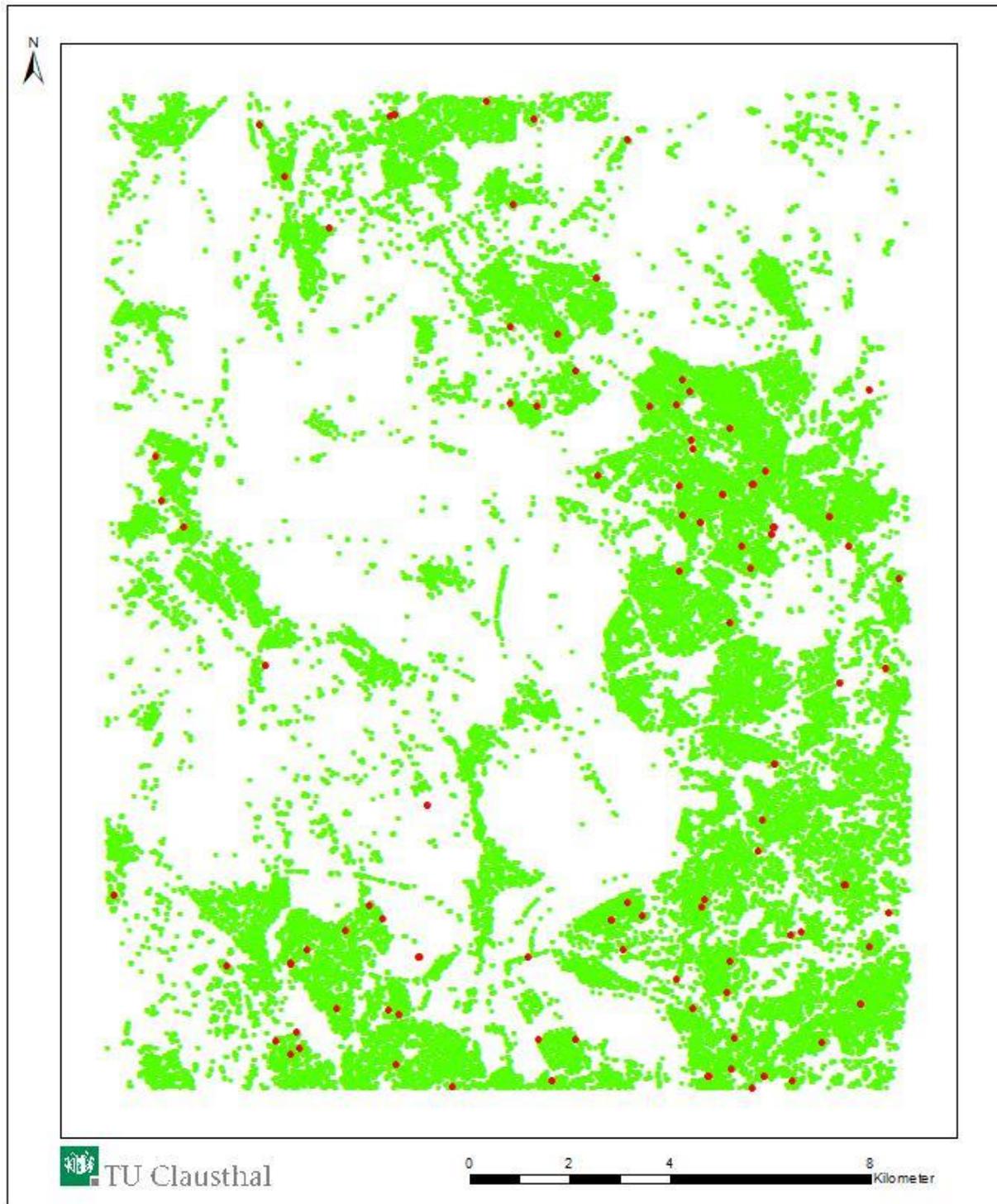


Abbildung 76: Räumliche Lage der detektierten Clusterpunkte mit abweichenden Höhenänderungsraten (rot dargestellt)

Exemplarisch werden die Höhenänderungszeitreihen von zwei detektierten anomalen Clusterpunkten mit extrem auffälligen Höhenänderungsraten in Abbildung 77 – 78 (Zeitreihen E und F) mit den gemittelten Zeitreihen der PS-Punkte in ihren Nachbarschaften gegenübergestellt.

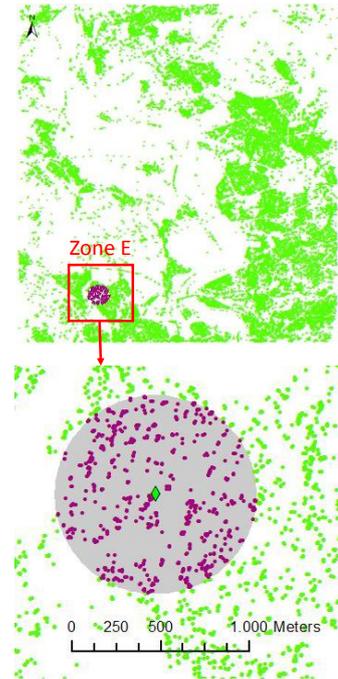
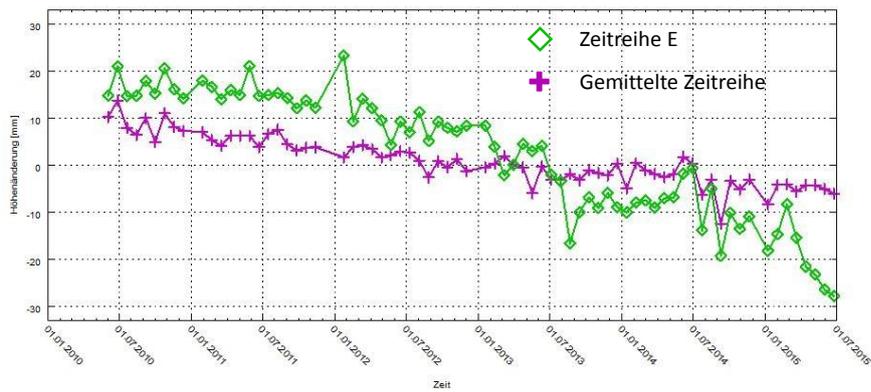


Abbildung 77: Abweichende Zeitreihe des Clusterpunktes E und gemittelte Zeitreihe benachbarter PS-Punkte

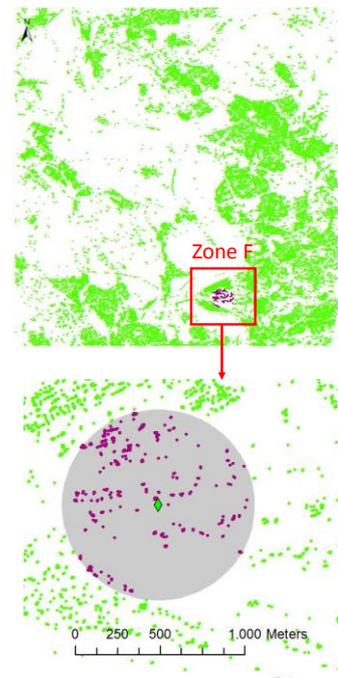
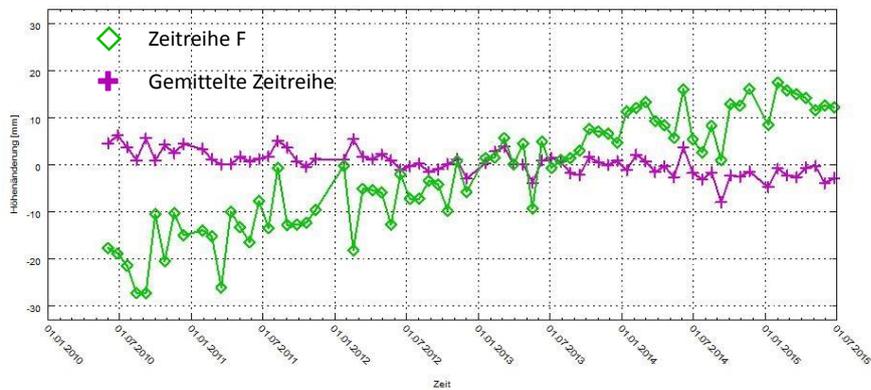


Abbildung 78: Abweichende Zeitreihe des Clusterpunktes F und gemittelte Zeitreihe benachbarter PS-Punkte

Die Abbildungen 77 und 78 bestätigen die detektierten Unterschiede der Höhenänderungszeitreihen der Clusterpunkte E und F gegenüber den gemittelten Zeitreihen

ihrer benachbarten PS-Punkte. So zeigt die Zeitreihe E eine starke Senkung von etwa -45 mm über den gesamten Untersuchungszeitraum, während die gemittelte Zeitreihe eine Senkung von nur ca. -15 mm aufweist. Die Zeitreihe F und die gemittelte Zeitreihe der Nachbarn besitzen einen entgegengesetzten Höhenänderungstrend über den gesamten Untersuchungszeitraum; die Zeitreihe F weist einen kontinuierlichen Hebungstrend von über 30 mm auf, während die gemittelte Zeitreihe der Nachbarn eine leichte Senkung von ca. -5 mm beinhaltet.

7.4. Validation des Analyseergebnisses

Obwohl die exemplarischen, visuellen Vergleiche in Abbildung 72 – 75 und Abbildung 77 - 78 eine Bestätigung für die deutlichen Abweichungen der detektierten Zeitreihen gegenüber ihren Nachbarzeitreihen aufzeigen, sollte die „Richtigkeit“ des Ergebnisses der Clusteranalyse für alle 350 detektierten anormalen Zeitreihen (236 aus der modifizierten Local-Moran's-Index – Clusteranalyse und 114 aus der nachfolgenden kleinräumigen Clusteranalyse) geprüft werden. Da eine visuelle Überprüfung angesichts der Punktzahl zu aufwändig und auch zu subjektiv erscheint, ist eine automatisierte und objektivere Prüfmethode zur Validation des Analyseergebnisses zu entwickeln.

Hierzu bietet sich als Prüfkriterium, analog zu einem visuellen Vergleich, die Höhenänderungsdifferenz zwischen dem ersten und letzten Datum der Zeitreihe an. Dabei sind jedoch die statischen Unsicherheiten der Messwerte zu berücksichtigen. Deshalb erfolgt die Berechnung der Höhenänderungsdifferenzen auf Grundlage der Funktionswerte der für jeden PS-Punkt mittels Ausgleichung bestimmten Trendpolynome (s. Abb. 60) und durch Differenzbildung der Funktionswerte für das erste und letzte Datum einer Zeitreihe. Diese Berechnungen werden für jeden detektierten, anormalen PS-Punkt und seine Nachbarn durchgeführt.

Bildet man den gewogenen arithmetischen Mittelwert aus allen so berechneten Höhenänderungsdifferenzen der benachbarten PS-Punkte, kann der Quotient aus seinem Unterschied gegenüber der Höhenänderungsdifferenz des detektierten, anormalen PS-Punktes und der Standardabweichung der Differenz als Kriterium für eine Validierung herangezogen werden. Ist der Quotient gemäß Formel 35 größer oder gleich k , deutet dies darauf hin, dass die Höhenänderungsdifferenz der als anormal detektierten Zeitreihe eine signifikante Abweichung gegenüber den Höhenänderungsdifferenzen der Nachbarn aufweist und somit der identifizierte anormale PS-Punkt anhand des Validationskriteriums als

„richtig“ detektiert bezeichnet werden kann. Der Wert von k hängt von einem gewählten Signifikanzniveau ab (s. Tabelle 17).

Tabelle 17: Zusammenhang zwischen Signifikanzniveau und k

Signifikanzniveau	10%	5%	1%
k	1,65	1,96	2,58

Formel 35:

$$\frac{|\Delta h_o - \overline{\Delta h_j}|}{s_D} \geq k$$

mit:

Δh_o : Höhenänderungsdifferenz am detektierten anormalen PS-Punkt o ,

$\overline{\Delta h_j}$: gewogener arithmetischer Mittelwert der Höhenänderungsdifferenzen aller benachbarten PS-Punkte j des Punktes o ,

s_D : Standardabweichung der Differenz ($\Delta h_o - \overline{\Delta h_j}$)

k : Konstante zur Charakterisierung des Signifikanzniveaus

Da die aus den Trendfunktionen abgeleiteten Höhenänderungsdifferenzen der benachbarten PS-Punkte j unterschiedliche Standardabweichungen aufweisen können, wird ihr Mittelwert als gewogener Mittelwert unter Nutzung der Formel 36 berechnet.

Formel 36:

$$\overline{\Delta h_j} = \frac{\sum_{j=1}^n \Delta h_j * p_j}{\sum_{j=1}^n p_j}$$

mit:

Δh_j : Höhenänderungsdifferenz eines benachbarten PS-Punktes j ,

p_j : Gewicht der Höhenänderungsdifferenz eines benachbarten PS-Punktes j

Die Gewichte p_j der n Höhenänderungsdifferenzen Δh_j ergeben sich mit Hilfe des Varianzfortpflanzungsgesetzes und den aus der Polynomgleichung ableitbaren Varianzen der Funktionswerte für das erste und letzte Messdatum der jeweiligen PSI-Zeitreihe.

Die Standardabweichung s_D der Differenz ($\Delta h_o - \overline{\Delta h_j}$) lässt sich mit Hilfe des Varianzfortpflanzungsgesetzes auf Grundlage der Varianz der Höhenänderungsdifferenz Δh_o des detektierten anomalen PS-Punktes o und der Varianz des gewogenen arithmetischen Mittelwertes der Höhenänderungsdifferenzen aller benachbarten PS-Punkte mittels Formel 37 berechnen.

Formel 37:
$$s_D = \sqrt{s_{\Delta h_o}^2 + s_{\overline{\Delta h_j}}^2}$$

mit:

$s_{\Delta h_o}$: Standardabweichung der Höhenänderungsdifferenz des detektierten anomalen PS-Punkt o

$s_{\overline{\Delta h_j}}$: Standardabweichung des gewogenen arithmetischen Mittelwertes der Höhenänderungsdifferenzen aller benachbarten PS-Punkte j

Die Varianz $s_{\overline{\Delta h_j}}^2$ kann mittels Formel 38 berechnet werden:

Formel 38:
$$s_{\overline{\Delta h_j}}^2 = \frac{\sum_{j=1}^n p_j * v_j^2}{(n - 1) \sum_{j=1}^n p_j}$$

mit:

p_j : Gewicht der Höhenänderungsdifferenz eines benachbarten PS-Punktes j

v_j : Verbesserung der Höhenänderungsdifferenz eines benachbarten PS-Punktes j

Die Verbesserung der Höhenänderungsdifferenz eines benachbarten PS-Punktes j ergibt sich gemäß Formel 39:

Formel 39:
$$v_j = \overline{\Delta h_j} - \Delta h_j$$

Mit diesem Ansatz wird anders als bei der Clusteranalyse, bei der die Höhenänderungsrate von PS-Punkten als Attribut benutzt wird, die ausgeglichene Höhenänderungsdifferenz herangezogen. Die nach den Clusteranalysen identifizierten anomalen Zeitreihen, die auch das Validationskriterium erfüllt haben, können letztendlich als wirkliche Anomalien bezeichnet werden, da sie anhand sowohl der Höhenänderungsrate als auch der

Höhenänderungsdifferenz signifikant von dem Verhalten ihrer benachbarten PS-Punkte abweichen. Ergänzend sei darauf hingewiesen, dass bei Anwendung dieses Validationskriteriums die Validation basierend auf den ausgeglichenen Höhenänderungsdifferenzen ausschließlich innerhalb des vordefinierten Suchradius stattfindet; sie verlangt auch, gemäß dem Grundprinzip einer Clusteranalyse, dass für eine erfolgreiche Durchführung des Validationsvorgangs eine genügende Anzahl von PS-Punkten in der Nachbarschaft des zu validierenden PS-Punktes befinden müssen.

Anhand des genannten Validationskriteriums (s. Formel 35) sind die 350 detektierten, anormalen Zeitreihen mit einem Signifikanzniveau von 5% ($k = 1,96$) überprüft worden. Insgesamt ergaben sich 4 von 350 detektierten anormalen Zeitreihen als nicht signifikant. In den Abbildungen 79 - 82 sind die Zeitreihen (Zeitreihe G, H, I und I) dieser nicht signifikanten Anomalien im Vergleich mit den gemittelten Zeitreihen der benachbarten PS-Punkte dargestellt.

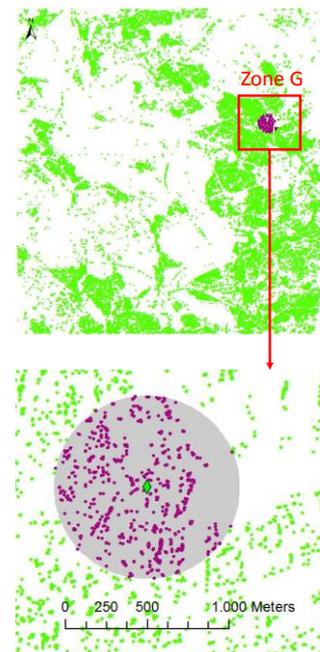
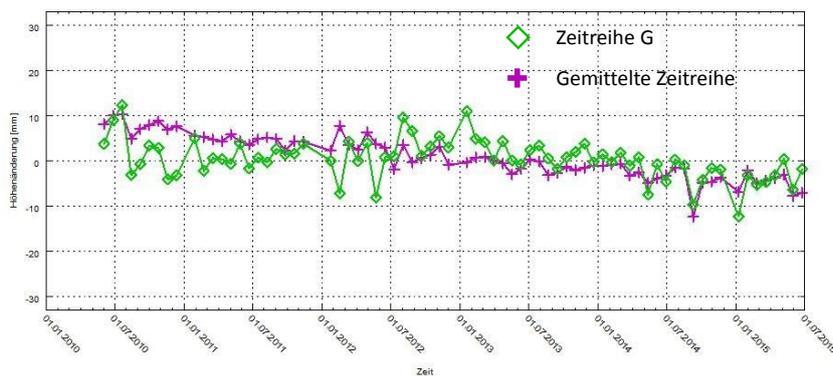


Abbildung 79: Gegenüberstellung der detektierten anormalen Zeitreihe G mit der gemittelten Zeitreihe benachbarter PS-Punkte

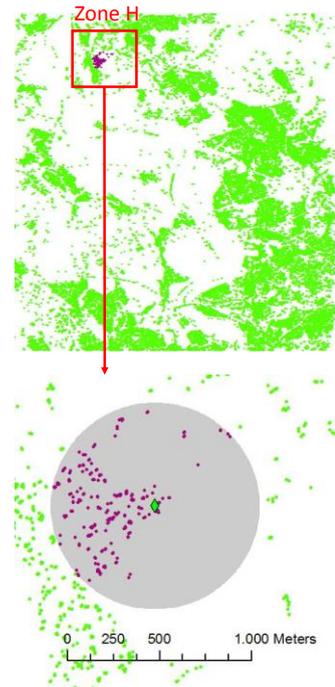
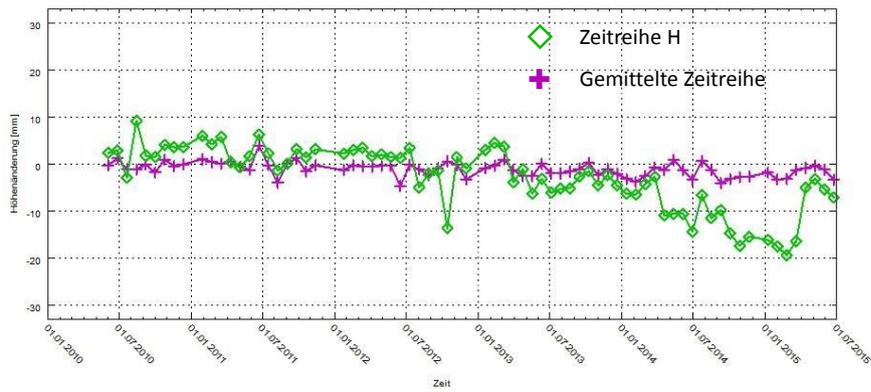


Abbildung 80: Gegenüberstellung der detektierten anomalen Zeitreihe H mit der gemittelten Zeitreihe benachbarter PS-Punkte

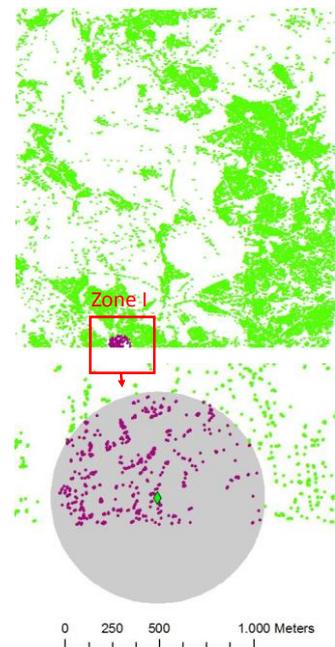
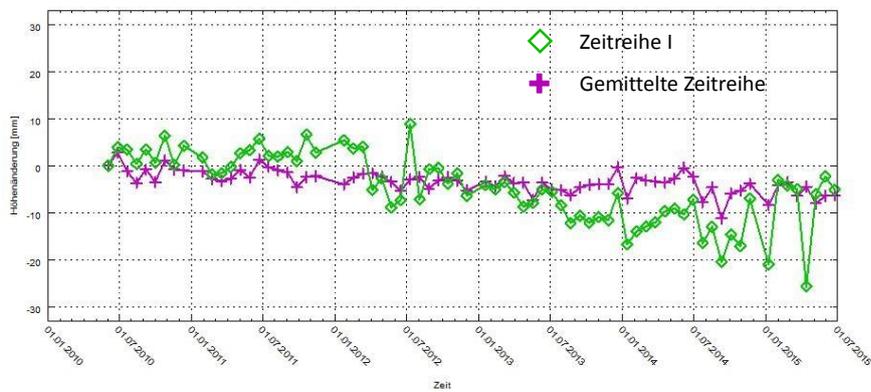


Abbildung 81: Gegenüberstellung der detektierten anomalen Zeitreihe I mit der gemittelten Zeitreihe benachbarter PS-Punkte

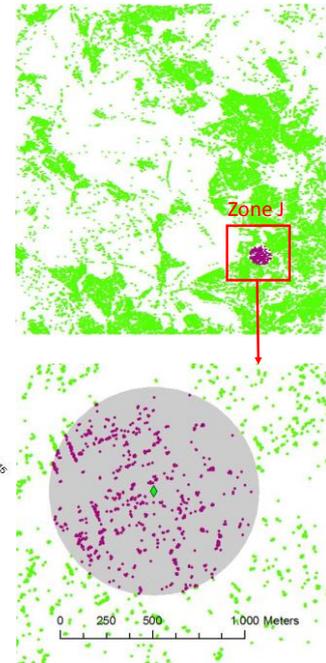
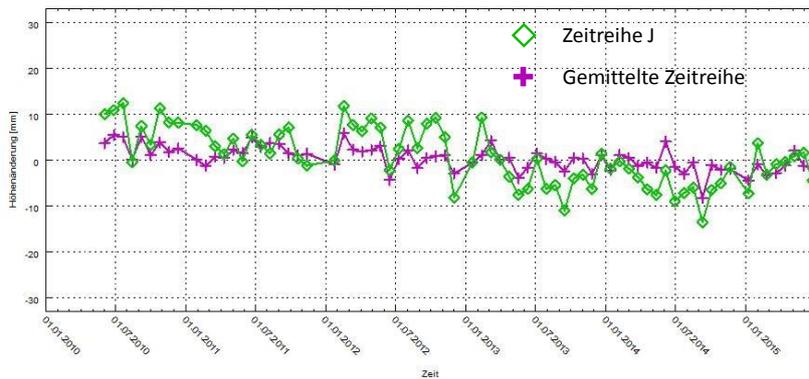


Abbildung 82: Gegenüberstellung der detektierten anomalen Zeitreihe J mit der gemittelten Zeitreihe benachbarter PS-Punkte

Ein visueller Vergleich der Höhenänderungszeitreihen in den Abbildungen 79 - 82 belegt die Richtigkeit dieser durchgreifenden Prüfung. Das Ergebnis dieser visuellen Vergleiche wird ebenfalls bestätigt durch Abbildung 83, in der exemplarisch die Häufigkeitsverteilung der Höhenänderungsdifferenzen der benachbarten PS-Punkte des PS-Punktes G dargestellt ist.

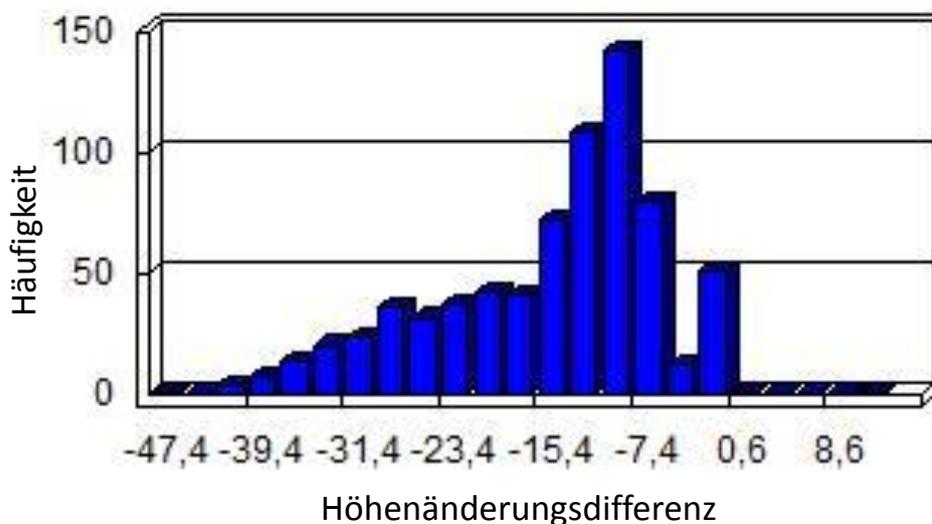


Abbildung 83: Häufigkeitsverteilung der Höhenänderungsdifferenzen der Nachbarpunkte des detektierten anomalen PS-Punktes G

Die für den PS-Punkt G berechnete Höhenänderungsdifferenz von -9,7 mm gehört eindeutig zur Grundgesamtheit dieser Werte.

Somit lässt sich feststellen, dass mit Hilfe des entwickelten Kriteriums die Qualität der Ergebnisse der Clusteranalysen zuverlässig geprüft werden kann. Die sich als nicht signifikant anormal erwiesenen PS-Punkte werden deshalb nicht ausgesondert.

Aufgrund dieses Ergebnisses wird die Validierung als letzter Schritt des Prozessablaufs (s. Abb. 67) empfohlen.

8. Zusammenfassung

Die anormalen Zeitreihen an PS-Punkten, die sich aufgrund verschiedener natürlicher, anthropogener, und auch verfahrenstechnischer Ursachen deutlich von dem erwarteten Bodenbewegungsverhalten unterscheiden, sind nicht repräsentativ für die Ableitung flächenhafter Bodenbewegungen oder im messtechnischen Sinne sogar grob falsch. Vor einer weiteren Nutzung der PSI-Ergebnisse sollten daher zuerst diese auffälligen „Anomalien“ in den vorliegenden Massendaten automatisiert erkannt und anschließend von einer nachfolgenden Interpretation des Bodenbewegungsverhaltens ausgeschlossen werden.

Da die Zeitreihen von PS-Punkten dem Typ georeferenzierter Zeitreihen zugeordnet werden können, wird im Rahmen der vorliegenden Arbeit ein raumzeitliches Clusterverfahren entwickelt, das die automatisierte Detektion von anormalen Zeitreihen an einzelnen PS-Punkten in flächen- und massenhaft vorliegenden PSI-Ergebnissen zur Ableitung flächenhafter Bodenbewegungen ermöglicht.

Zur Auswahl eines für diese Aufgabenstellung geeigneten Verfahrens wurden verschiedene Clusterverfahren auf ihre Anwendbarkeit hin untersucht. Das erwünschte Clusterverfahren sollte in der Lage sein, mit räumlichen Daten umzugehen und zugleich auch die Fähigkeit haben, Anomalien hinsichtlich ihres untersuchten Attributes zu detektieren. Letztlich wurde das „Local-Moran's-Index – Clusterverfahren“ auf Grundlage lokaler räumlicher Autokorrelationen ausgewählter Attribute zwischen den PS-Punkten ausgewählt und an praktischen Beispielen getestet.

Es zeigte sich, dass grundsätzlich mit Hilfe des Local-Moran's-Index - Clusterverfahrens signifikante „räumliche Anomalien“ detektiert werden können, dabei aber der Wahl des Attributes eine wichtige Bedeutung zukommt. Im Vergleich zur Clusteranalyse mit den Attributen „Polynomgrad“ und „Höhenänderung“ (abgeleitet aus einer Trendpolynombestimmung) konnte die Clusteranalyse mit dem Attribut „Höhenänderungsrate“ ein deutlich besseres Ergebnis erzielen.

Darüber hinaus ergab die Untersuchung, dass zur Beschreibung des charakteristischen Verlaufs einer Höhenänderungszeitreihe ein einziges Attribut offensichtlich nicht ausreicht. Da bei der Local-Moran's-Index - Clusteranalyse jedoch nur ein einziges Attribut berücksichtigt werden kann, erfolgt eine Modifikation dieses Verfahrens durch Einführung eines Gewichtungsfaktors zur Beurteilung und Quantifizierung der Ähnlichkeit der Höhenänderungsverläufe benachbarter PS-Punkte. Damit können nun auch die zeitlichen Korrelationen (Ähnlichkeiten) zwischen dem betrachteten PS-Punkt und seinen benachbarten PS-Punkten bei der Berechnung des Local-Moran's-Indizes berücksichtigt und so eine

raumzeitliche Clusteranalyse basierend auf raumzeitlicher Autokorrelation für die Suche von anormalen Zeitreihen an PS-Punkten ermöglicht werden. Am Beispiel von Testdaten wurde das so modifizierte Local-Moran's-Index – Clusterverfahren erfolgreich getestet.

Des Weiteren wurde ein Kriterium für die Festlegung eines optimalen Suchradius abgeleitet, das auf einer gemeinsamen Betrachtung von räumlichen Autokorrelationen und der Anzahl benachbarter PS-Punkte basiert, und zu einer Modifikation der Local-Moran's - Clusteranalyse führte. Dazu wurde die räumliche Autokorrelation für eine Reihe zunehmender Entfernungen (Suchradien) berechnet. Der optimale Suchradius stellt die Entfernung dar, bei der die räumlichen Abhängigkeiten der zugrunde liegenden räumlichen Prozesse (Bodenbewegungen) am größten sind. Anhand der praktischen Ergebnisse der gebiets- und sensorspezifischen Untersuchungen wurde festgestellt, dass grundsätzlich der optimale Suchradius von dem Untersuchungsgebiet und dem Bodenbewegungsgeschehen im Untersuchungszeitraum sowie der unterschiedlichen Bodenauflösung der verwendeten Sensoren abhängig ist. Manchmal ist es jedoch nicht möglich, anhand der lokalen räumlichen Autokorrelation einen einheitlichen, globalen Suchradius für das gesamte Untersuchungsgebiet festzulegen. In diesem Fall empfiehlt es sich, einen optimalen Suchradius in ausgewählten Teilgebieten zu ermitteln, und anschließend zu prüfen, ob deren Mittelwert oder Medianwert auf das gesamte Untersuchungsgebiet übertragbar ist. Die Auswahl der Teilgebiete sollte möglichst in zusammenhängend bebauten Gebieten mit relativ regelmäßiger PS-Punktverteilung erfolgen. Für die erfolgreiche Durchführung des Local-Moran's-Index – Clusterverfahrens muss sichergestellt werden, dass jedem PS-Punkt eine genügende Anzahl von Nachbarn zur Verfügung steht. Um somit möglichst viele PS-Punkte in die Clusteranalyse einzubeziehen, besteht ggf. die Notwendigkeit, den Suchradius für eine Anwendung des Clusterverfahrens zu vergrößern.

Darüber hinaus wurde anhand eines praktischen Beispiels ermittelt, dass Anomalien trotz deutlich unterschiedlichen Attributwerten bei Anwendung des globalen Mittelwertes des betrachteten Attributes in einem lokalen Untersuchungsbereich anhand des Local-Moran's-Index – Clusterverfahrens nicht entdeckt werden können. Um solche Anomalien auch erkennen zu können, wurde das modifizierte Local-Moran's-Index – Clusterverfahren um eine weitere kleinräumige Clusteranalyse, die auf der statistischen Häufigkeitsverteilung des betrachteten Attributs von zuvor ermittelten Clusterpunkten innerhalb der Nachbarschaft beruht, ergänzt.

Abschließend wurde am Beispiel eines umfangreichen Testdatensatzes die erweiterte und modifizierte Local-Moran's-Index – sowie kleinräumige Clusteranalyse zur Identifizierung von anormalen Zeitreihen an PS-Punkten erprobt und das Analyseergebnis mit einem entwickelten statistischen Validationsverfahren, basierend auf berechneten Höhenänderungsdifferenzen,

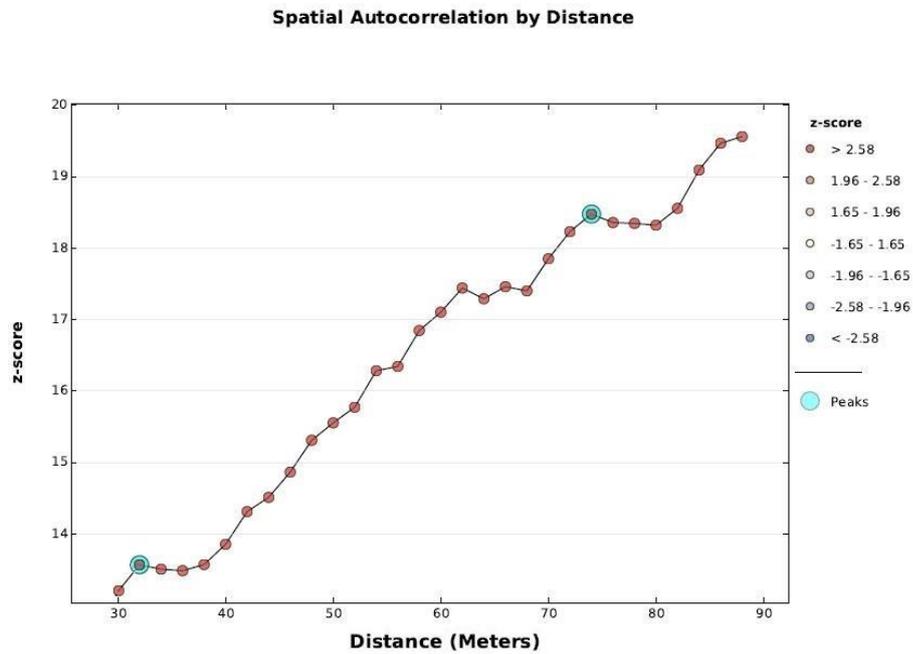
überprüft. Zusammenfassend zeigte sich, dass eine automatisierte Detektion von anormalen Zeitreihen an massenhaft vorhandenen PS-Punkten mittels des modifizierten Local-Moran's-Index – Verfahrens und einer Erweiterung um eine kleinräumige Clusteranalyse, zuverlässig erfolgen kann. Die hierdurch ermöglichte Datenbereinigung liefert einen wichtigen Beitrag zur Verbesserung der Qualität einer flächenhaften Modellierung von Bodenbewegungen.

Es soll darauf hingewiesen werden, dass mittels der Clusteranalyse nicht alle PS-Punkte auf Anomalien getestet werden können, da manche, stark räumlich isolierte PS-Punkte nicht über genügend Nachbarn verfügen können, und somit eine grundlegende Voraussetzung für die erfolgreiche Durchführung einer Clusteranalyse nicht erfüllen. Sie sind als nicht geprüfte Daten durch ein entsprechendes Attribut zu kennzeichnen. Es wäre deswegen von Interesse, auch solche PS-Punkte später durch andere Methoden zu überprüfen.

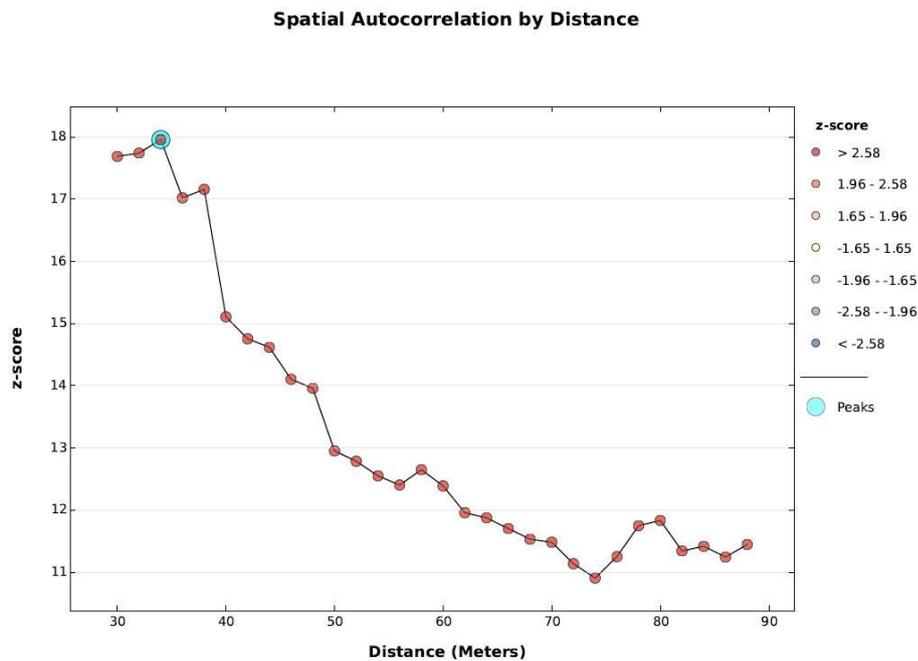
Des Weiteren sei ausdrücklich darauf hingewiesen, dass die detektierten Anomalien nicht „automatisch“ mit fehlerhaften Beobachtungen gleichgesetzt werden dürfen. So könnten je nach fachlichem Zusammenhang, vorhandenem fachlichen Kenntnisstand und der Aufgabenstellung in Verbindung mit einer Ursache-Wirkungsanalyse diese „anormalen“ Zeitreihen an PS-Punkten Hinweise auf ein lokal auftretendes Bodenbewegungsphänomen (z.B. an geologischen Störungen) geben. Somit sollten diese als anormal detektierten PS-Punkte nicht eliminiert sondern für weitere Untersuchungen vorgehalten werden.

9. Anhang

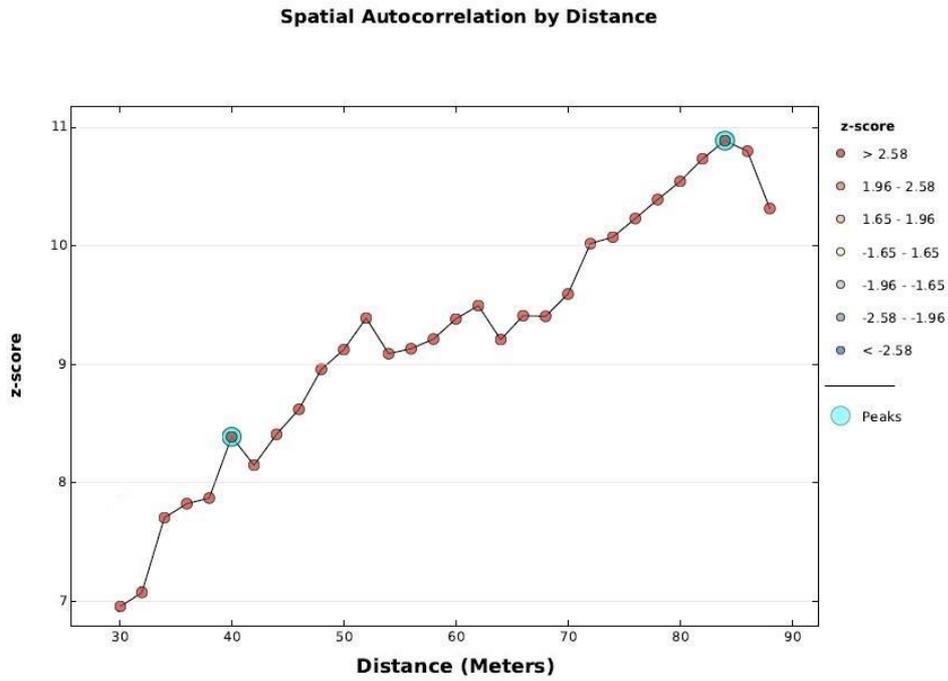
- Radarsat-2-Testdaten



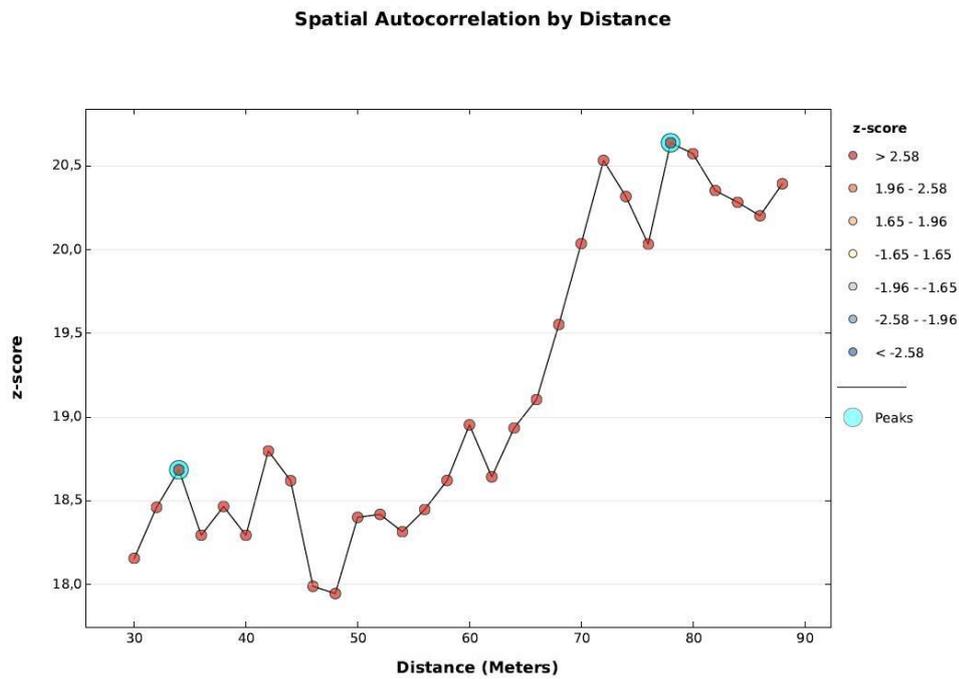
A 1: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 1 (Abbildung 54) mit Radarsat-2-Testdaten



A 2: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 2 (Abbildung 54) mit Radarsat-2-Testdaten

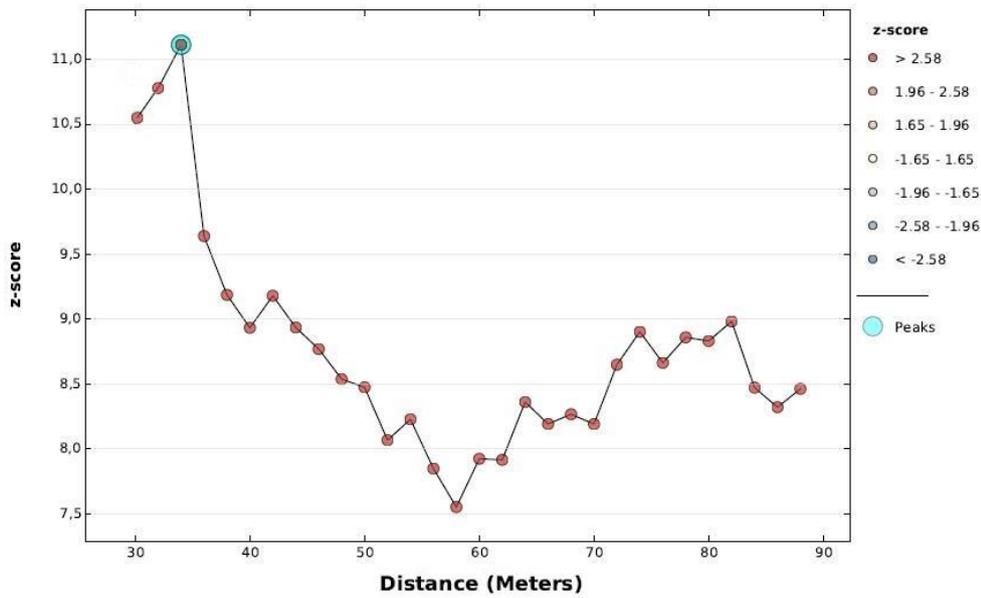


A 3: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 3 (Abbildung 54) mit Radarsat-2-Testdaten



A 4: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 4 (Abbildung 54) mit Radarsat-2-Testdaten

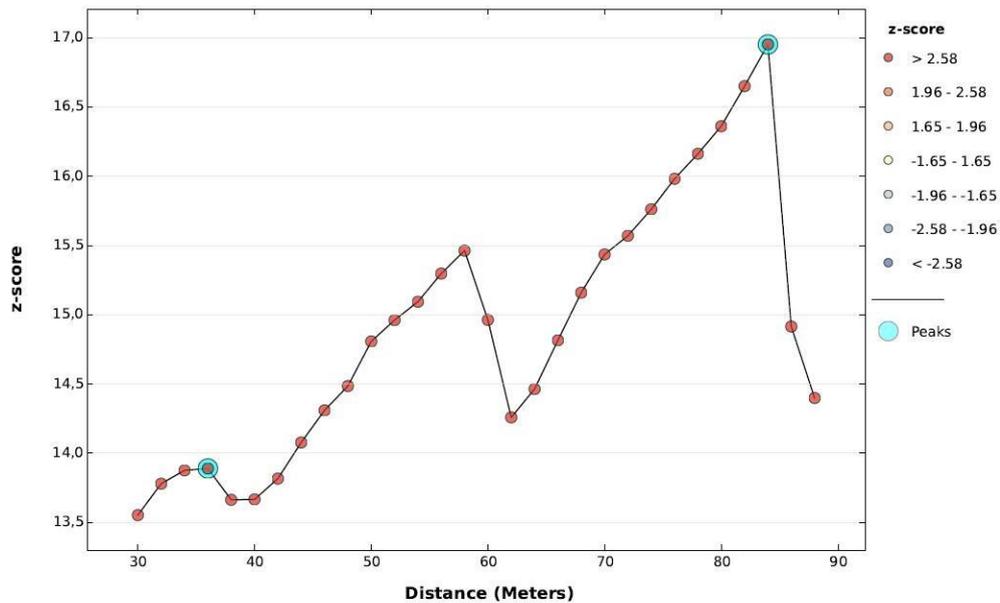
Spatial Autocorrelation by Distance



A 5: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 5 (Abbildung 54) mit Radarsat-2-Testdaten

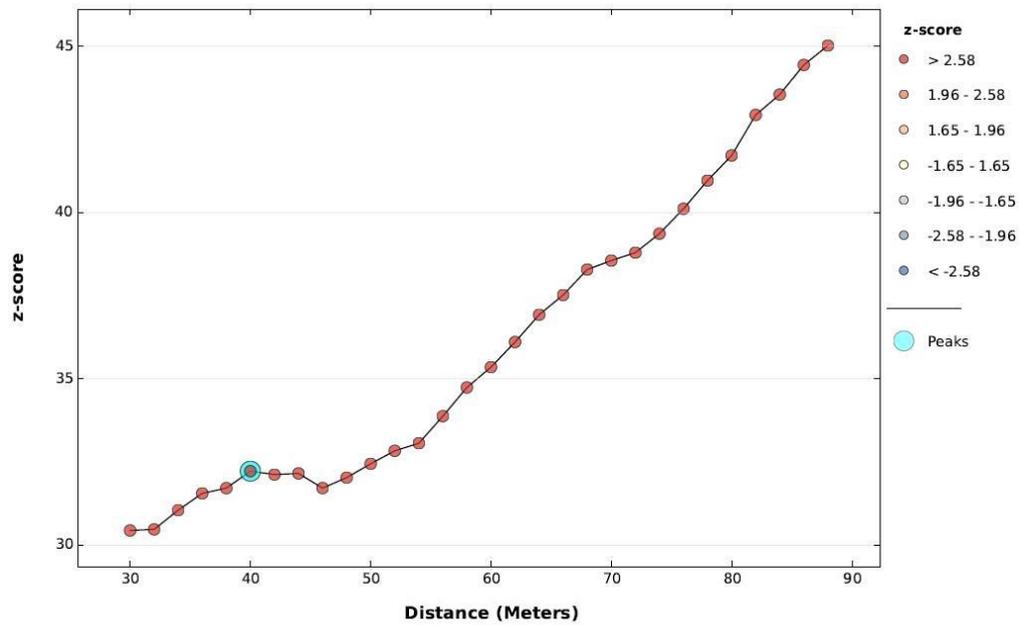
- **Envisat-Testdaten**

Spatial Autocorrelation by Distance



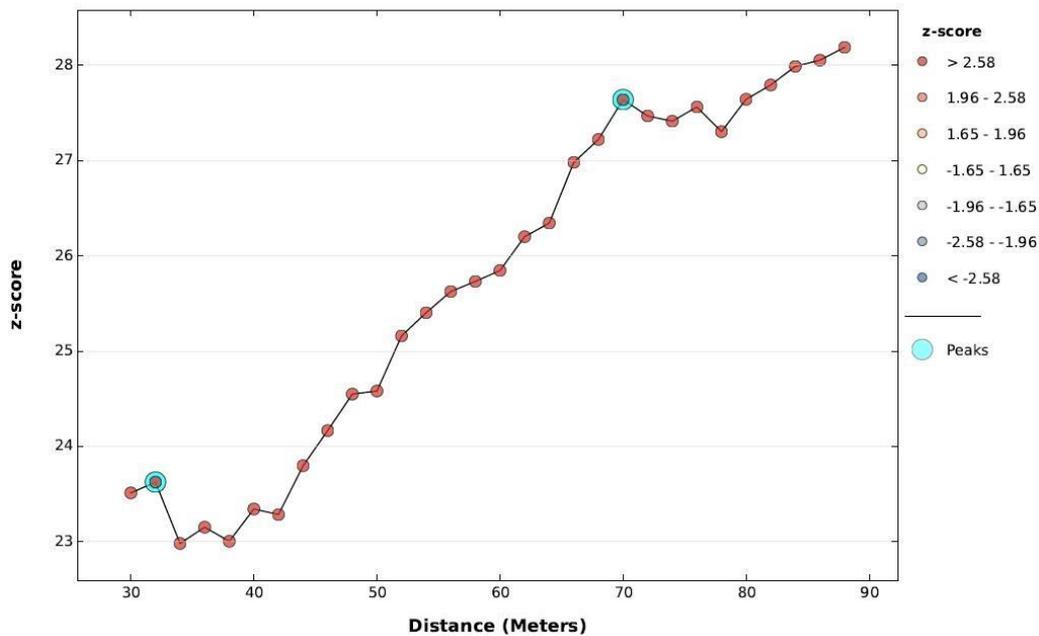
A 6: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 1 (Abbildung 56) mit Envisat-Testdaten

Spatial Autocorrelation by Distance



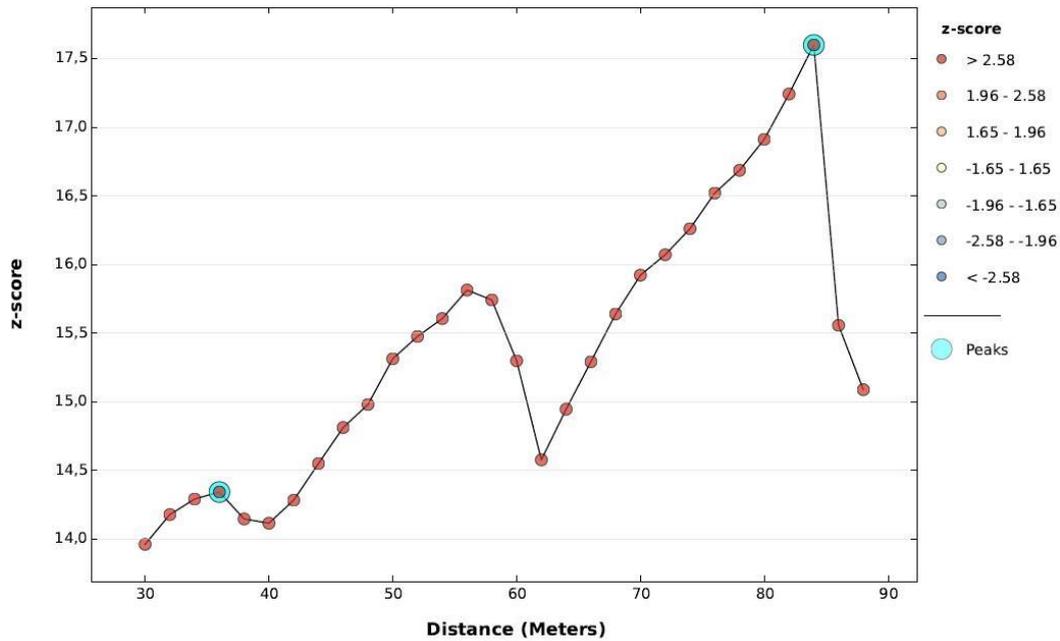
A 7: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 2 (Abbildung 56) mit Envisat-Testdaten

Spatial Autocorrelation by Distance



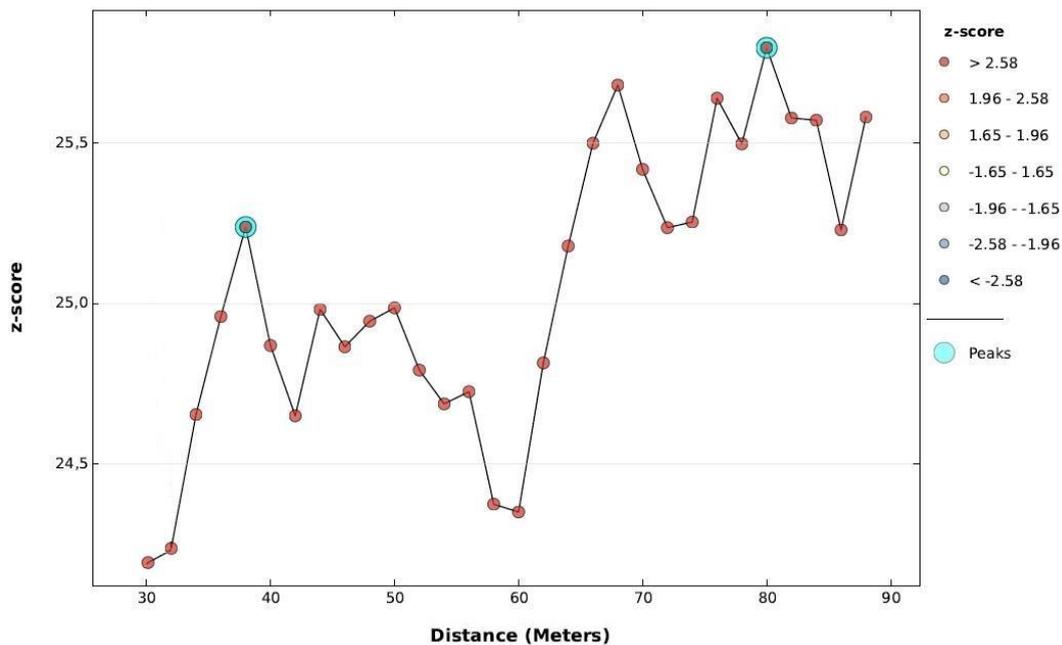
A 8: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 3 (Abbildung 56) mit Envisat-Testdaten

Spatial Autocorrelation by Distance



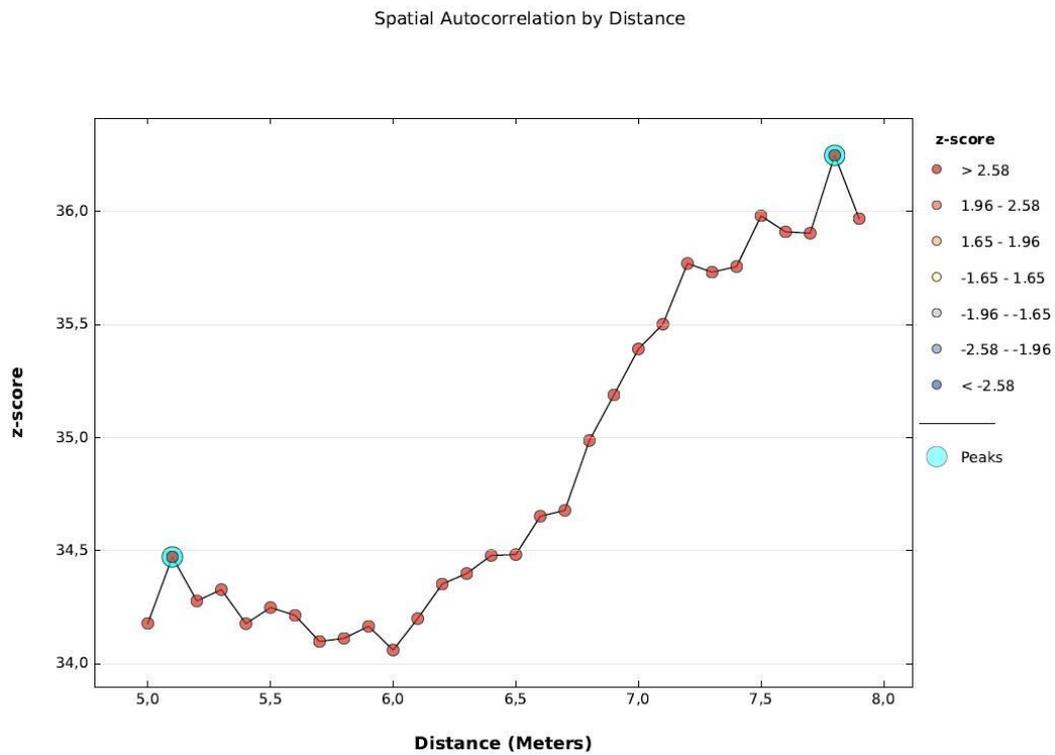
A 9: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 4 (Abbildung 56) mit Envisat-Testdaten

Spatial Autocorrelation by Distance

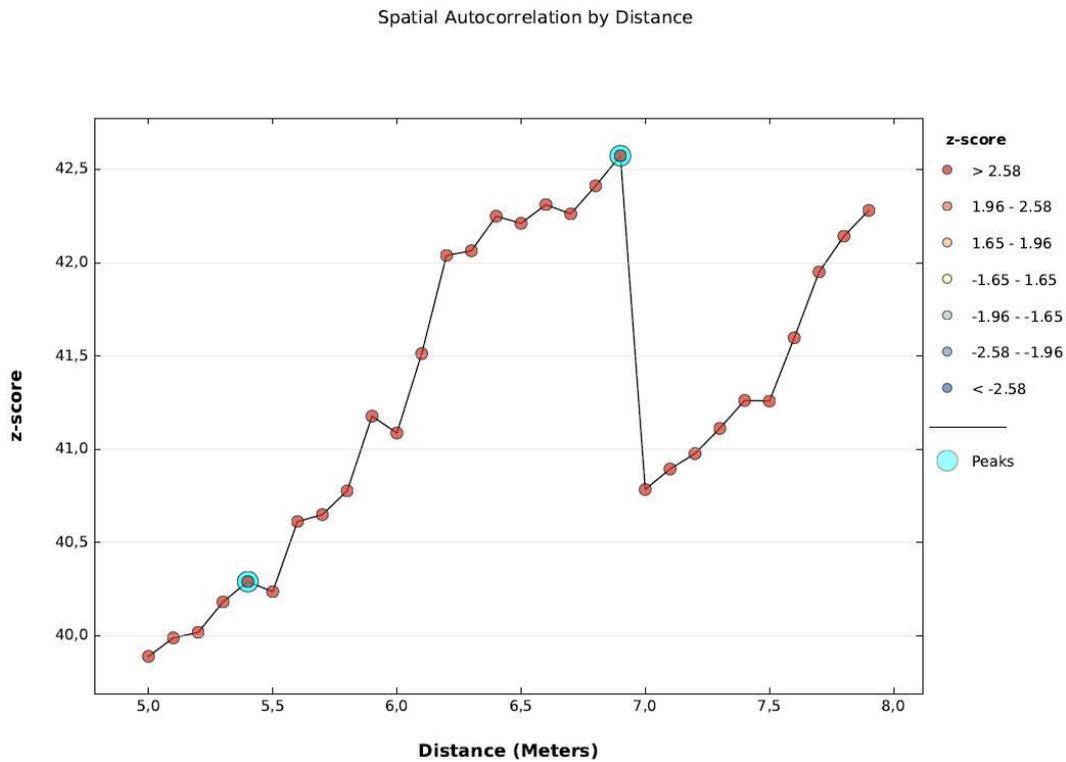


A 10: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 5 (Abbildung 56) mit Envisat-Testdaten

- TerraSAR-X-Testdaten

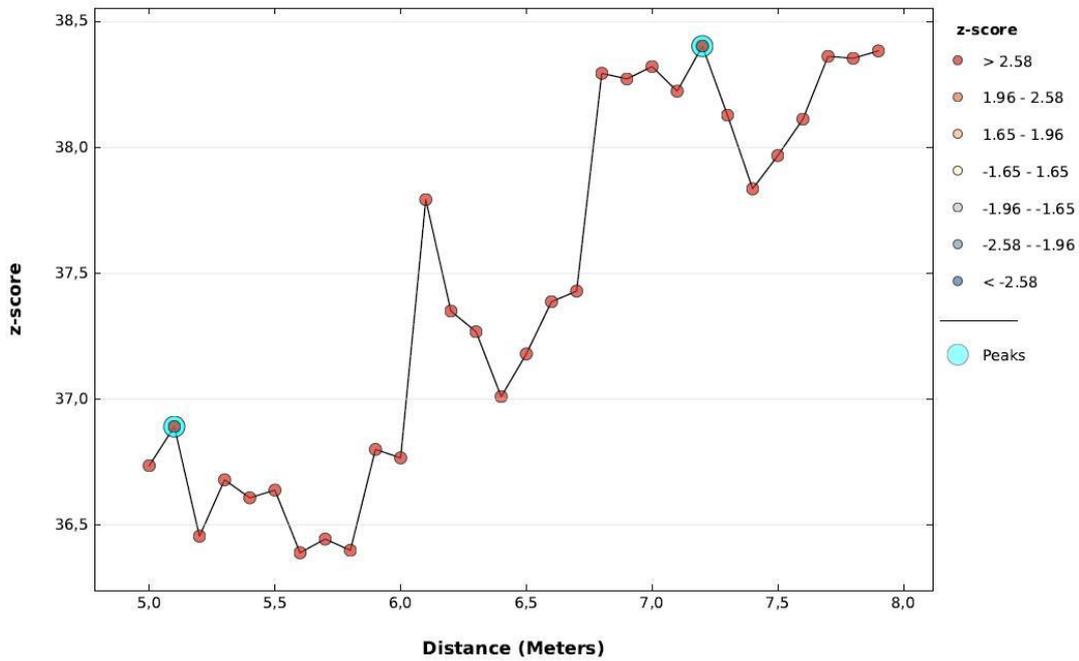


A 11: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 1 (Abbildung 58) mit TerraSAR-X-Testdaten



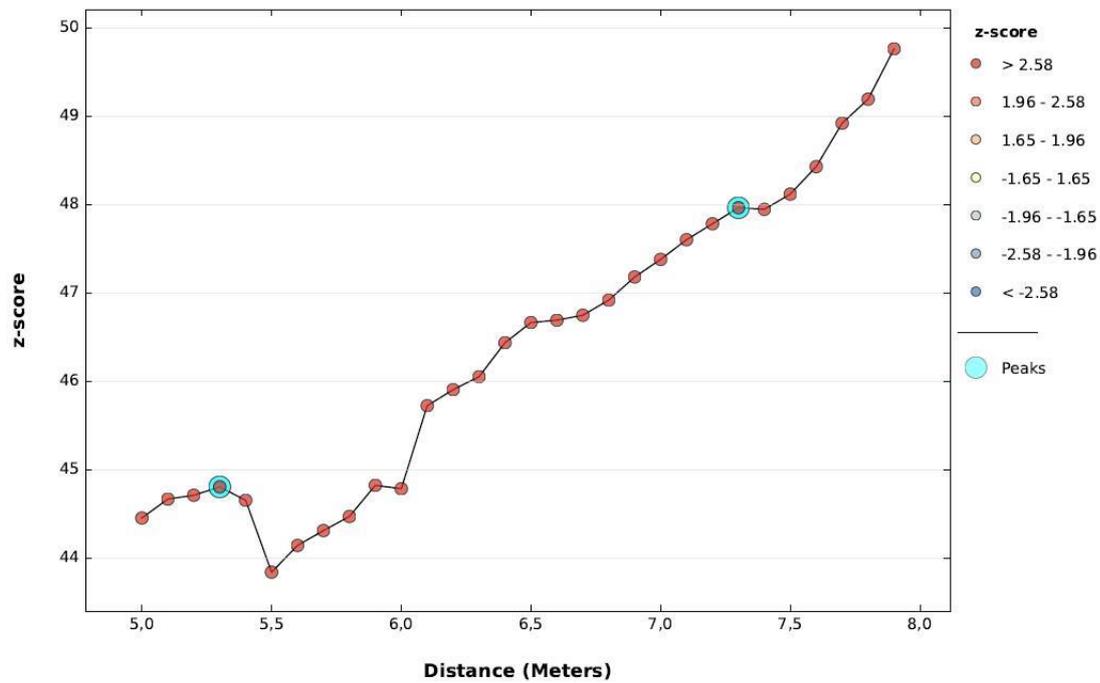
A 12: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 2 (Abbildung 58) mit TerraSAR-X-Testdaten

Spatial Autocorrelation by Distance



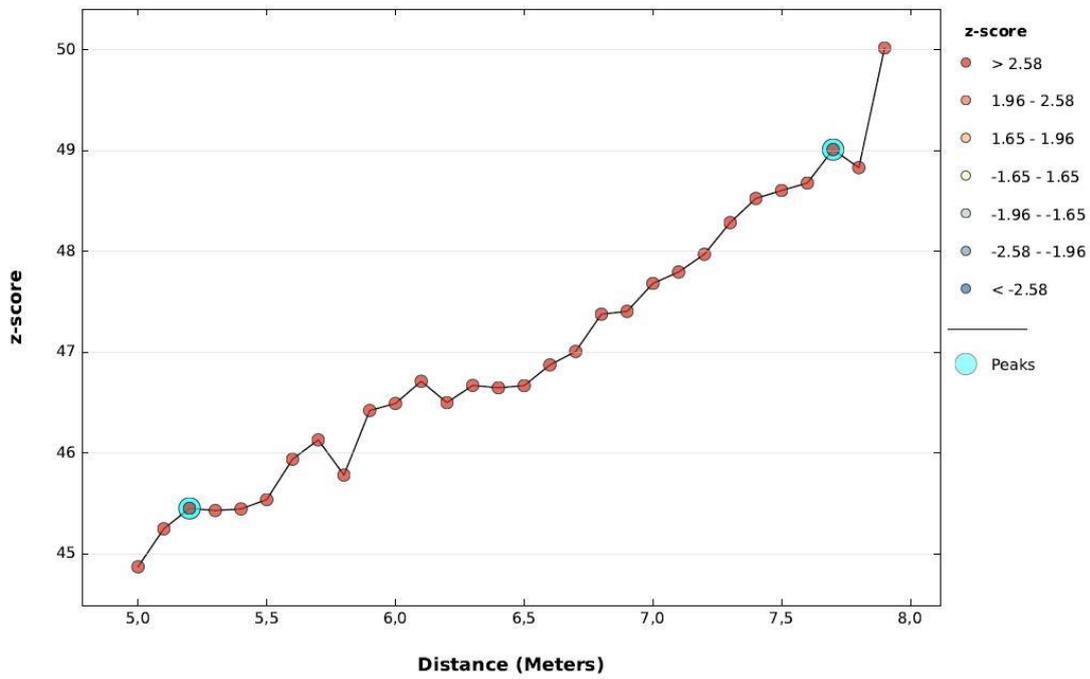
A 13: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 3 (Abbildung 58) mit TerraSAR-X-Testdaten

Spatial Autocorrelation by Distance



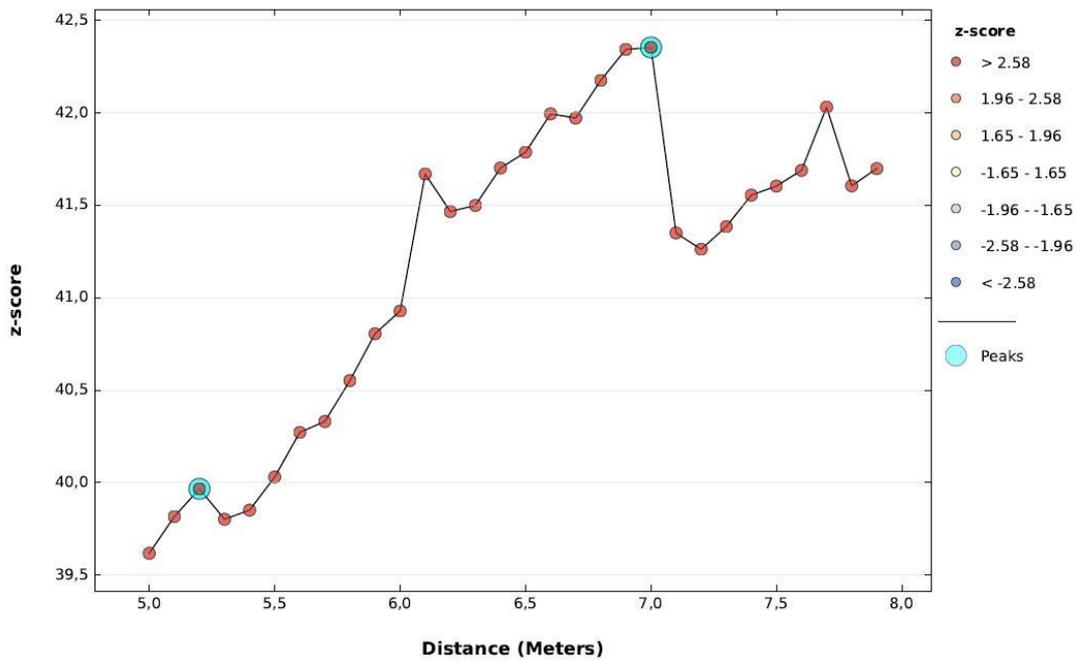
A 14: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 4 (Abbildung 58) mit TerraSAR-X-Testdaten

Spatial Autocorrelation by Distance



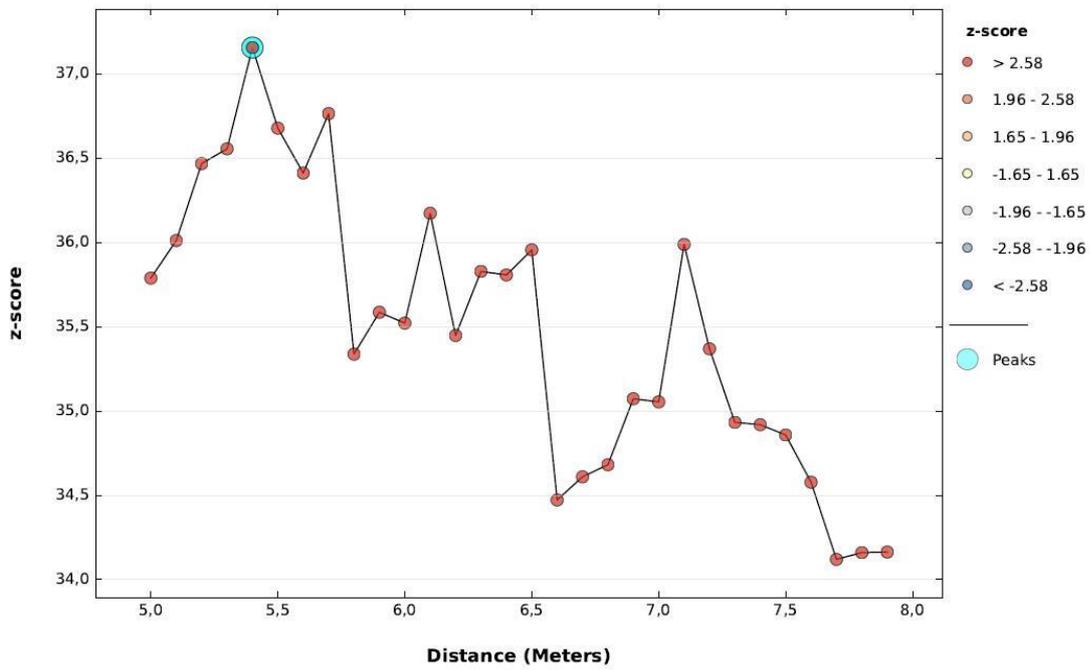
A 15: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 5 (Abbildung 58) mit TerraSAR-X-Testdaten

Spatial Autocorrelation by Distance



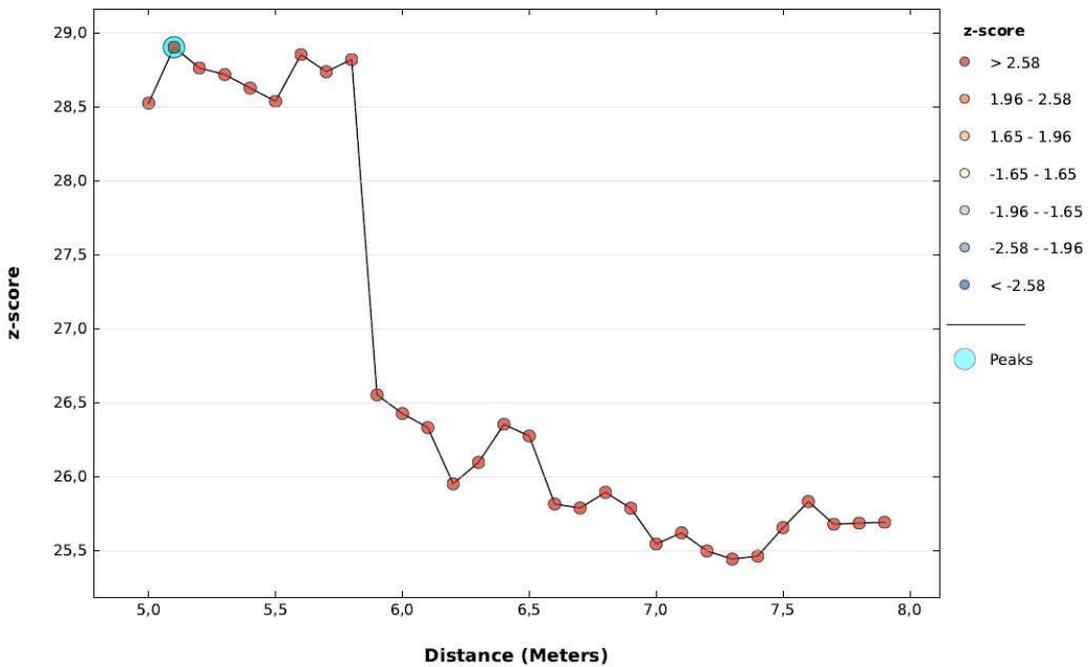
A 16: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 6 (Abbildung 58) mit TerraSAR-X-Testdaten

Spatial Autocorrelation by Distance



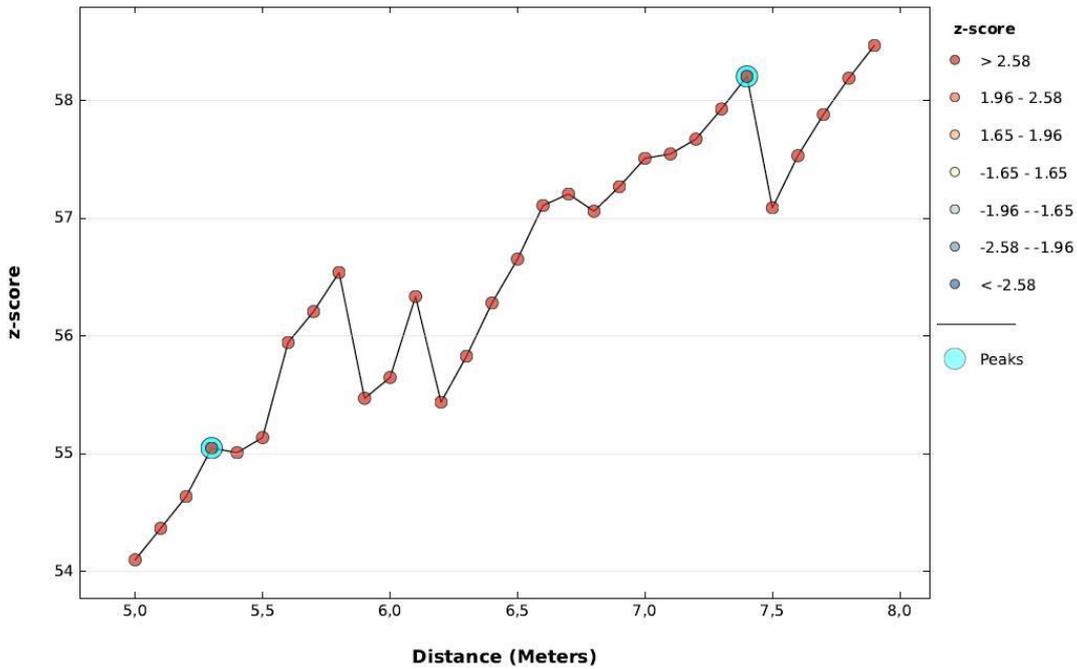
A 17: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 7 (Abbildung 58) mit TerraSAR-X-Testdaten

Spatial Autocorrelation by Distance



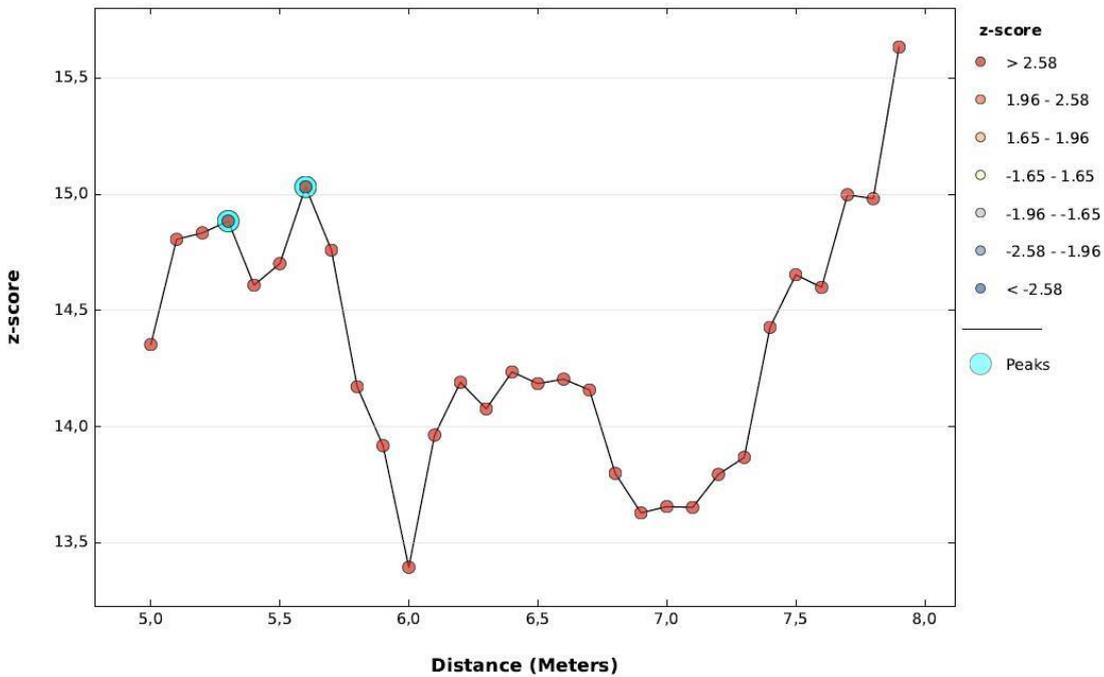
A 18: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 8 (Abbildung 58) mit TerraSAR-X-Testdaten

Spatial Autocorrelation by Distance



A 19: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 9 (Abbildung 58) mit TerraSAR-X-Testdaten

Spatial Autocorrelation by Distance



A 20: Veränderung der räumlichen Autokorrelation mit zunehmender Entfernung für das Teilgebiet 10 (Abbildung 58) mit TerraSAR-X-Testdaten

10. Literaturverzeichnis

- ABLER, R. F. (HG.) (1992): *Geography's inner worlds. Pervasive themes in contemporary American geography*. New Brunswick, NJ: Rutgers Univ. Press (Occasional publications of the Association of American Geographers, 2).
- AGRAWAL, R.; GEHRKE, J.; GUNOPULOS, D.; RAGHAVAN, P. (2005): *Automatic Subspace Clustering of High Dimensional Data*.
- ANSELIN, L. (1988): *Spatial Econometrics. Methods and Models*. Dordrecht: Springer Netherlands.
- ANSELIN, L. (1989): *What is Special About Spatial Data? Alternative Perspectives on Spatial Data Analysis*. In: NCGIA Technical Reports Spring.
- ANSELIN, L. (1995): *Local Indicators of Spatial Association-LISA*. In: *Geographical Analysis* 27 (2), S. 93–115.
- ATHAYDE PINTO, C.; PARADELLA, W. R.; MURA, J. C.; GAMA, F. F.; RIBEIRO DOS SANTOS, A.; SILVA, G. G.; HARTWIG, M. E. (2015): *Applying persistent scatterer interferometry for surface displacement mapping in the Azul open pit manganese mine (Amazon region) with TerraSAR-X StripMap data*. In: *J. Appl. Remote Sens* 9 (1).
- BACKHAUS, K. (1996): *Multivariate Analysenmethoden. Eine anwendungsorientierte Einführung*. Berlin, Springer-Verlag.
- BACKHAUS, K.; ERICHSON, B.; PLINKE, W.; WEIBER, R. (HG.) (2016): *Multivariate Analysemethoden*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- BAHRENBURG, G.; GIESE, E.; NIPPER, J. (2003): *Statistische Methoden in der Geographie. 2. Aufl.* Berlin, Stuttgart: Gebr. Borntraeger (Statistische Methoden in der Geographie, 02).
- BAILEY, T. C.; GATRELL, A. C. (1995): *Interactive spatial data analysis*. Harlow: Longman Scientific & Technical.
- BAMLER, R.; ADAM, N.; HINZ, S.; EINEDER, M. (2008): *SAR-Interferometrie für geodätische Anwendungen*. In: *Allgemeine Vermessungs-Nachrichten AVN 7/2008*, S. 243–252.
- BARNETT, V.; LEWIS, T. (2000): *Outliers in statistical data*. 3. ed., reprinted. Chichester: Wiley (Wiley series in probability and mathematical statistics).
- BERGS, S. (1981): *Optimalität bei Clusteranalysen. Experimente zur Bewertung numerischer Klassifikationsverfahren*. Dissertation. Universität Münster, Münster. Fachbereich Wirtschafts- und Sozialwissenschaften.

- BERTI, M.; CORSINI, A.; FRANCESCHINI, S.; IANNAcone, J. P. (2013): *Automated classification of Persistent Scatterers Interferometry time series*. In: Nat. Hazards Earth Syst. Sci. 13 (8), S. 1945–1958.
- BOCK, H. H. (1974): *Automatische Klassifikation. Theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten ; (Cluster-Analyse)*. Göttingen: Vandenhoeck & Ruprecht (Studia mathematica, 24).
- BOCK, H. H. (1994): *Classification and clustering*. Problems for the future, S. 3–24.
- BOROWSKI, M. (2013): *Echtzeit-Extraktion relevanter Information aus multivariaten Zeitreihen basierend auf robuster Regression*. Technische Universität Dortmund.
- BUSCH, W.; LINKE, J. (2014): *Räumliche Höhenänderungsanalyse auf Grundlage einer automatisierten Ausgleichung massenhaft vorliegender PSI-Zeitreihen mittels orthogonaler Polynome*. In: AVN Allgemeine Vermessungs-Nachrichten (Heft 8/9), S. 298–310.
- BUSCH, W.; WALTER, D.; COLDEWEY, W. G.; HEJMANOWSKI, R. (2016): *Analyse von Senkungserscheinungen außerhalb des prognostizierten Einwirkungsbereiches. Bergwerk Walsum der RAG AG*. Gutachten im Auftrag der Bezirksregierung Arnsberg. Institut für Geotechnik und Markscheidewesen, TU Clausthal. Clausthal-Zellerfeld.
- CHESTER, D. K. (1987): *Statistics in geography by David Ebdon*, Basil Blackwell, 1985. No. of pages. 232. In: Geol. J. 22 (1), S. 61–62.
- CLARK, T. (1989): *Managing Outliers: Qualitative Issues in the Handling of Extreme Observations in Marketing Research*. In: Marketing Research (Vol. 1 Issue 2), S. 31–48.
- CLEFF, T. (2015): *Clusteranalyse*. In: Thomas Cleff (Hg.): *Deskriptive Statistik und Explorative Datenanalyse*. Wiesbaden: Gabler Verlag, S. 189–215.
- CLIFF, A. D.; ORD, J. K. (1973): *Spatial autocorrelation*. London: Pion (Monographs in spatial and environmental systems analysis, 5).
- CLIFF, A. D.; ORD, J. K. (1981): *Spatial processes. Models & applications*. London: Pion.
- COLESANTI, C.; FERRETTI, A.; PRATI, C.; ROCCA, F. (2003): *Monitoring landslides and tectonic motions with the Permanent Scatterers Technique*. In: Engineering Geology 68 (1-2), S. 3–14.
- CRESSIE, N. A. C. (1993): *Statistics for Spatial Data*. Hoboken, NJ, USA: John Wiley & Sons, Inc.
- DMV (DEUTSCHER MARKSCHEIDER-VEREIN E.V.), ARBEITSGRUPPE RADARINTERFEROMETRIE (2013): *Grundsätze zum Einsatz von satellitengestützten Verfahren der Radarinterferometrie zur Erfassung von Höhenänderungen*. Deutscher Markscheider-Verein e.V. Herne.

- ECKSTEIN, P. P. (2014): *Clusteranalyse*. In: Peter P. Eckstein (Hg.): Statistik für Wirtschaftswissenschaftler. Wiesbaden: Springer Fachmedien Wiesbaden, S. 407–426.
- ECKSTEIN, P. P. (HG.) (2016): *Angewandte Statistik mit SPSS*. Wiesbaden: Springer Fachmedien Wiesbaden.
- ESRI ONLINE: Online verfügbar unter <http://resources.arcgis.com/en/help/>.
- ESTER, M.; KRIEGEL, H.-P.; SANDER, J.; XU, X. (1996): *A density-based algorithm for discovering clusters in large spatial databases with noise*. In: AAAI Press, S. 226–231.
- FERRETTI, A.; PRATI, C.; ROCCA, F. (2001): *Permanent scatterers in SAR interferometry*. In: IEEE Trans. Geosci. Remote Sensing 39 (1), S. 8–20.
- FISCHER, M. M.; GETIS, A. (HG.) (2010): *Handbook of applied spatial analysis. Software tools, methods and applications*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.
- FLETLING, R. (2010): *Methodische Ansätze zur unscharfen Mustererkennung bei Deformationsmessergebnissen*. Zugl.: Braunschweig, Techn. Univ., Diss., 2010. Braunschweig: Techn. Univ. Inst. für Geodäsie und Photogrammetrie (Geodätische Schriftenreihe der Technischen Universität Braunschweig, 27).
- GEARY, R. C. (1954): *The Contiguity Ratio and Statistical Mapping*. In: The Incorporated Statistician 5 (3), S. 115.
- GETIS, A.; ORD, J. K. (1992): *The Analysis of Spatial Association by Use of Distance Statistics*. In: Geographical Analysis 24 (3), S. 189–206.
- GEUDTNER, D. (1995): *Die interferometrische Verarbeitung von SAR-Daten des ERS-1*. Online verfügbar unter <http://elib.dlr.de/23721/>.
- GIANNOTTI, F.; PEDRESCHI, D. (2008): *Mobility, data mining and privacy*. Geographic knowledge discovery; with 5 tables. Berlin [u.a.]: Springer.
- GOODCHILD, M. F. (1986): *Spatial autocorrelation*. Norwich: Geo Books (Concepts and techniques in modern geography, 47).
- GRIFFITH, D. A. (1987): *Spatial autocorrelation. A primer*. Washington, DC: Assoc. of American Geographers (Resource publications in geography).
- GRIFFITH, D. A.; AMRHEIN, C. G. (1991): *Statistical analysis for geographers*. Englewood Cliffs, NJ: Prentice-Hall.
- GUHA, S.; RASTOGI, R.; SHIM, K. (1998): *CURE. An Efficient Clustering Algorithm for Large Databases*. In: SIGMOD Rec. 27 (2), S. 73–84.

- HADI, F.-T (2012): *Spatio-Temporal Clustering Methods Classification*. In: Doctoral Symposium on Informatics Engineering Conference 2012, S. 199–209.
- HAN, J.; KAMBER, M.; TUNG, A. K.H. (2001): *Spatial Clustering Methods in Data Mining: A Survey*. Geographic Data Mining and Knowledge Discovery. London, New York: Taylor & Francis (Research monographs in geographic information systems).
- HANSEN, R. F. (2001): *Radar interferometry. Data interpretation and error analysis*. Dordrecht, Boston: Kluwer Academic (Remote sensing and digital image processing, v. 2).
- HARTUNG, J.; ELPELT, B.; KLÖSENER, K.-H. (1999): *Statistik. Lehr- und Handbuch der angewandten Statistik ; mit zahlreichen, vollständig durchgerechneten Beispielen*. 12. Aufl. München: Oldenbourg.
- HARTUNG, J.; ELPELT, B. (2007): *Multivariate Statistik. Lehr- und Handbuch der angewandten Statistik*. 7. Aufl. München: Oldenbourg.
- HASLETT, J.; BRADLEY, R.; CRAIG, P.; UNWIN, A.; WILLS, G. (1991): *Dynamic Graphics for Exploring Spatial Data with Application to Locating Global and Local Anomalies*. In: The American Statistician 45 (3).
- HAWKINS, D. M. (2004): *Outliers. Band 6, Encyclopedia of statistical sciences*. New York, Chichester: John Wiley & Sons (1).
- HINNEBURG, A.; KEIM, D. A. (1998): *An Efficient Approach to Clustering in Large Multimedia Databases with Noise*. In: KDD'98 Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, S. 58-65.
- JIAMTHAPTHAKSIN, R.; CHOO, J.; CHEN, C.; CELEPCIKAY, O. U.; GIUSTI, C.; EICK, C. F. (2010): *MOSAIC*. In: Tho Manh Nguyen (Hg.): *Complex Data Warehousing and Knowledge Discovery for Advanced Retrieval Development*: IGI Global, S. 231–250.
- JUST, D.; BAMLER, R. (1994): *Phase statistics of interferograms with applications to synthetic aperture radar*. In: Applied optics 33 (20), S. 4361–4368.
- KAMPES, B. M. (2006): *Radar interferometry. Persistent scatterer technique*. Dordrecht, the Netherlands: Springer (Remote sensing and digital image processing, v. 12).
- KAUFMAN, L.; ROUSSEEUW, P. J (1987): *Clustering by means of medoids*. Delft: Faculty of Mathematics and Informatics (Reports of the Faculty of Mathematics and Informatics, no. 87-03.).
- KAUFMAN, L.; ROUSSEEUW, P. J. (1990): *Finding Groups in Data*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

- KIRCHER, M. (2004): *Analyse flächenhafter Senkungserscheinungen in sedimentären Gebieten mit den neuen Techniken der Radarfernerkundung am Beispiel der Niederrheinischen Bucht*. Mathematisch-Naturwissenschaftliche Fakultät, Universität Bonn. Bonn.
- KISILEVICH, S.; MANSMANN, F.; NANNI, M.; RINZIVILLO, S. (2010): *Spatio-temporal clustering*. In: Oded Maimon und Lior Rokach (Hg.): *Data Mining and Knowledge Discovery Handbook*. Boston, MA: Springer US, S. 855–874.
- KNOSPE, S.; HEBEL, H.-P.; SCHÄFER, M.; SCHÄFER, T.; WALTER, D.; BUSCH, W. (2011): *Die Anwendung der satellitengestützten Radarinterferometrie zur großräumigen Erfassung von Höhenänderungen*. In: Geomonitoring Tagung 2011 – Ein Paradigmenwechsel zur Beherrschung von Georisiken, S. 73–90.
- KOLATCH, E. (2001): *Clustering Algorithms for Spatial Databases: A Survey*. Online verfügbar unter <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.28.1145&rep=rep1&type=pdf>.
- LANGE, N. (2013): *Geoinformatik. In Theorie und Praxis*. 3., vollst. überarb. u. akt. Aufl. 2013. Berlin, Heidelberg: Springer.
- LEGENDRE, P. (1993): *Spatial Autocorrelation. Trouble or New Paradigm?* In: *Ecology* 74 (6), S. 1659–1673.
- LEIJEN, F.; HANSEN, R. F. (2004): *Interferometric Radar Meteorology: Resolving the acquisition ambiguity*. In: Proc. of CEOS SAR Workshop, 6 S., 27.-28.05.2004, Ulm.
- LINDER, S. (2013): *Räumliche Diffusion von Photovoltaik-Anlagen in Baden-Württemberg*. Dissertation. Universität Würzburg, Würzburg. Institut für Geographie und Geologie.
- LLOYD, C. D. (2010): *Spatial data analysis. An introduction for GIS users*. Oxford: Oxford Univ. Press.
- LO, C. P.; YEUNG, A. K. W. (2007): *Concepts and techniques of geographic information systems*. 2. ed. Upper Saddle River, NJ: Pearson Prentice Hall (PH series in geographic information science).
- MAIMON, O.; ROKACH, L. (HG.) (2005): *Data mining and knowledge discovery handbook*. New York: Springer.
- MAIMON, O.; ROKACH, L. (HG.) (2010): *Data Mining and Knowledge Discovery Handbook*. Boston, MA: Springer US.
- MARK, P.; NIEMEIER, W.; SCHINDLER, S.; BLOME, A.; HEEK, P.; KRIVENKO, A.; ZIEM, E. (2012): *Radarinterferometrie zum Setzungsmonitoring beim Tunnelbau*. In: *Bautechnik* 89 (11), S. 764–776.

- MASSONNET, D.; ROSSI, M.; CARMONA, C.; ADRAGNA, F.; PELTZER, G.; FEIGL, K.; RABAUTE, T. (1993): *The displacement field of the Landers earthquake mapped by radar interferometry*. In: *Nature* 364 (6433), S. 138–142.
- MASSONNET, D.; SOUYRIS, J.-C. (2008): *Imaging with synthetic aperture radar*. 1st ed. Lausanne, Switzerland, Boca Raton, FL: EPFL Press (Engineering sciences Electrical engineering).
- MIRANDA, N.; ROSICH, B.; SANTELLA, C.; GRION, M. (2004): *Review of the Impact of ERS-2 Piloting Modes on the SAR Doppler Stability*. In: *Proceedings of the 2004 Envisat & ERS Symposium (ESA SP-572)*. 6-10 September 2004, Salzburg, Austria.
- MITCHELL, A. (2005): *The ESRI guide to GIS analysis. Volume 2: Spatial Measurements and Statistics*. 1st ed. Redlands, Calif.: ESRI.
- MORAN, P. A. P. (1948): *The Interpretation of Statistical Maps*. In: *Journal of the Royal Statistical Society. Series B (Methodological)* (Vol. 10, No. 2), S. 243–251.
- NG, R. T.; HAN, J. (1994): *Efficient and Effective Clustering Methods for Spatial Data Mining*. In: *Proceedings of the 20th International Conference on Very Large Data Bases*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc (VLDB '94), S. 144–155.
- NIEMANN, H. (1974): *Methoden der Mustererkennung*. Frankfurt/M.: Akad. Verl.-Ges (Informationsverarbeitung in technischen, biologischen und ökonomischen Systemen, Bd. 2).
- ORD, J. K.; GETIS, A. (1995): *Local Spatial Autocorrelation Statistics. Distributional Issues and an Application*. In: *Geographical Analysis* 27 (4), S. 286–306.
- PICHLER, W. (1998): *Räumliche Autokorrelationsanalyse und geographische Informationssysteme*. [Diplomarbeit]. Vollst. zugl.: Wien, Univ., Dipl.-Arb., 1998. Wien: Inst. für Wirtschafts- u. Sozialgeographie Wirtschaftsuniv. Wien (WSG-Forschungsberichte, 13).
- PITZ, W.; MILLER, D. (2010): *The TerraSAR-X Satellite*. In: *IEEE Trans. Geosci. Remote Sensing* 48 (2), S. 615–622.
- PRATI, C.; ROCCA, F. (1993): *Use of the spectral shift in SAR interferometry*. In: *Proceedings of the 2nd ERS-1 Symposium - Space at the Service of our Environment*, Hamburg, Germany, 11-14 October 1993, ESA SP-361, S. 691-696.
- RABUS, B.T.; LANG, O. (2002): *On the representation of ice-shelf grounding zones in SAR interferograms*. In: *Journal of Glaciology* 48 (162), S. 345–356.
- RAMBOLD, A. (1999): *Ausgewählte Verfahren zur Identifikation von Ausreißern und einflußreichen Beobachtungen in multivariaten Daten und Verfahren. Theoretische*

Überlegungen und empirische Befunde. Zugl.: Passau, Univ., Diss., 1999. München: Utz Wiss (Statistik).

ROTT, H.; MAYER, C.; NAGLER, T. (2001): *Einsatzmöglichkeiten der Radarinterferometrie zur Überwachung von Hangbewegungen und Absenkungen*. In: VGI - Österreichische Zeitschrift für Vermessung und Geoinformation 89 (3-4), S. 167–172.

RUNKLER, T. A. (2010): *Data Mining. Methoden und Algorithmen intelligenter Datenanalyse*. Wiesbaden: Vieweg + Teubner (Computational intelligence). Online verfügbar unter <http://dx.doi.org/10.1007/978-3-8348-9353-6>.

SCHÄFER, M. (2012): *Atmosphäre als Phasenbestandteil der differentiellen Radarinterferometrie und ihr Einfluss auf die Messung von Höhenänderungen*. Clausthal, Techn. Univ., Diss., Clausthal-Zellerfeld: Univ.-Bibliothek Clausthal.

SCHEIBER, R. (2004): *Hochauflösende Interferometrie für Radar mit synthetischer Apertur*. In: Deutsches Zentrum für Luft- und Raumfahrt (DLR), Forschungsbericht; 12; 1-167; Deutsches Zentrum für Luft- und Raumfahrt (DLR), Forschungsbericht.

SCHWÄBISCH, M. (1995): *Die SAR-Interferometrie zur Erzeugung digitaler Geländemodelle*. Online verfügbar unter <http://elib.dlr.de/73/>.

SHEKHAR, S.; LU, C.-T.; ZHANG, P. (2003A): *A Unified Approach to Detecting Spatial Outliers*. In: Geoinformatica 7 (2), S. 139–166.

SHEKHAR, S.; ZHANG, P.; HUANG, Y.; VATSAVAI, R. (2003B): *Trends in Spatial Data Mining (Data Mining: Next Generation Challenges and Future Directions)*.

SHEIKHOESLAMI, G.; CHATTERJEE, S.; ZHANG, A. (1998): *Wavecluster. A multi-resolution clustering approach for very large spatial databases*.

STEINHAUSEN, D.; LANGER, K. (1977): *Clusteranalyse. Einführung in Methoden und Verfahren der automatischen Klassifikation ; mit zahlreichen Algorithmen, FORTRAN-Programmen, Anwendungsbeispielen und einer Kurzdarstellung der multivariaten statistischen Verfahren*. Berlin: De Gruyter (De Gruyter Lehrbuch).

TAN, P.-N.; STEINBACH, M.; KUMAR, V. (2009): *Introduction to data mining*. [Nachdr.]. Boston: Pearson Addison-Wesley.

TOBLER, W. R. (1970): *A Computer Movie Simulating Urban Growth in the Detroit Region*. In: Economic Geography 46, S. 234.

WALTER, D. (2012): *Systematische Einflüsse digitaler Höhenmodelle auf die Qualität radarinterferometrischer Bodenbewegungsmessungen*. Clausthal-Zellerfeld: Papierflieger-Verl.

- WANG, W.; YANG, J.; MUNTZ, R. R. (1997): *STING. A Statistical Information Grid Approach to Spatial Data Mining*. In: Proceedings of the 23rd International Conference on Very Large Data Bases. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc (VLDB '97), S. 186–195.
- WERNER, C.; WEGMUELLER, U.; STROZZI, T.; WIESMANN, A. (2003): *Interferometric Point Target Analysis for Deformation Mapping*. IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Toulouse, France, 21-25 July 2003.
- WINKER, P. (2010): *Empirische Wirtschaftsforschung und Ökonometrie*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- WOODHOUSE, I. H. (2006): *Introduction to microwave remote sensing*. Boca Raton, Fla.: CRC Press Taylor & Francis.
- XI, F.; BUSCH, W.; ZHANG, C. (2015): *Multikriterielle Clusteranalysen zur Detektion von Ausreißern in flächenhaft vorliegenden PSI-Ergebnissen zur Bestimmung von Bodenbewegungen*. In: Bergbau, Energie und Rohstoffe 2015, 7.-9. Oktober 2015, Freiberg, Wagner Digitaldruck, Nossen, S. 209–220.
- XU, X.; ESTER, M.; KRIEGEL, H.-P.; SANDER, J. (1998): *A distribution-based clustering algorithm for mining in large spatial databases*. In: Proceedings 14th International Conference on Data Engineering. 14th International Conference on Data Engineering. Orlando, FL, USA, 23-27 Feb. 1998: IEEE Comput. Soc, S. 324–331.
- YIN, X.; BUSCH, W.; WALTER, D. (2015): *Detektion von sprunghaften Messwertänderungen in PSI-Zeitreihen zur Bestimmung von Bodenbewegungen*. In: Tagungsband Bergbau, Energie und Rohstoffe 2015, 7.-9. Oktober 2015, Freiberg, Wagner Digitaldruck, Nossen, S. 181–195.
- ZEBKER, H. A.; WEBER HOEN, E. (2000): *Penetration depths inferred from interferometric volume decorrelation observed over the Greenland Ice Sheet*. In: IEEE Trans. Geosci. Remote Sensing 38 (6), S. 2571–2583.
- ZHANG, T.; RAMAKRISHNAN, R.; LIVNY, M. (1996): *BIRCH. An Efficient Data Clustering Method for Very Large Databases*. In: Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data. New York, NY, USA: ACM (SIGMOD '96), S. 103–114.