

9. Die Methoden zur Unterstützung des Aggregierungsprozesses in der Bodenkunde

Die in dieser Arbeit vorgeschlagenen und exemplarisch angewandten Methoden zur Unterstützung des Aggregierungsprozesses sind in die Verfahrensgruppen der Landschaftsanalyse und Entropiemodellierung, der heuristischen multivariaten Verfahren und der kontext-basierten Regelanwendung einordenbar. Der vom Bodenkundler geführte Prozess der Aggregation wird durch diese Methoden in seinen wesentlichen Teilschritten objektiviert. Die mit diesen Methoden erarbeiteten Kriterien zur Ähnlichkeit von Legenden-, Kartiereinheiten und deren Inventar bieten sich als reproduzierbare Entscheidungsgrundlage für Zusammenfassungen an. Sie wurden für GIS ArcInfo in handhabbare Form umgesetzt. Das Programm ist Grundlage, Hauptwerkzeug dieser Arbeit. Im Leitfaden zum Programm "Analytische Werkzeuge der Aggregation und Generalisierung von Bodenkarten" (Kap. 10) werden die einzelnen Komponenten und deren Zusammenspiel beschrieben, die zugrunde liegenden Formeln wiedergegeben und die aus den Berechnungen resultierenden Ergebnisse und die Formate der Ergebnisdateien beschrieben. Bei der Arbeit mit dem Programm werden nur die Gruppierungsvorschläge, Indizes und Parameter bereitgestellt. Geometrieändernde Arbeitsschritte werden dann durch den Bearbeiter auf der entsprechenden ArcInfo-Ebene vorgenommen.

9.1 Landschaftsanalyse und Entropiemodellierung

Dahinter verbergen sich Werkzeuge zur Exploration topologischer Strukturen, die besonders für Fragestellungen in Ökologie, Tier- und Pflanzensoziologie etc. entwickelt wurden. Dabei richtet sich die Analyse auf Struktur, Funktion und Wechsel in der Landschaft. Eine frühe Entwicklung für GIS stellt FRAGSTATS - Spatial pattern analysis program for quantifying landscape structure - dar (MCGARIGAL & MARKS 1994), das mittlerweile als kommerzielles Produkt für ArcInfo angeboten wird. Darauf basieren auch Entwicklungen für ArcView in Form von Avenue-Scripten, wie z. B. der Patch-Analyst (ELKIE et al. 1999).

Bodenkarten liegen in Polygontopologie vor. Hierfür wurden die zur Unterstützung des Aggregierungsprozesses wesentlichen Indizes ausgewählt und programmiert. Das sind:

- Indizes zur Kennzeichnung der Diversität / Heterogenität
- Indizes zur Kennzeichnung der Gestalt von Einzelflächen und Flächen gleicher Attribute
- Indizes zur Kennzeichnung der räumlichen Struktur von Flächen gleicher Attribute.

9.1.1 Indizes zur Kennzeichnung der Diversität / Heterogenität

Der populärste Index zur Messung der Diversität / Heterogenität ist der Shannon-Index. Er basiert auf der Informationstheorie, wird auch als Negentropie bezeichnet und repräsentiert den Betrag an Information für eine definierte Menge.

Entropie ist ein Spezialfall der Information (PALM 1985). Die dem Maß inne wohnende Eigenschaft, dass bei Gleichverteilung der höchste Grad der Unbestimmtheit vorliegt, führt zu weiteren Formeln, durch die wahrscheinlichkeitstheoretisch der Grad der Gleichverteilung dargestellt werden kann. Die hieraus berechneten Zahlen geben die relative Entropie wieder.

Das Entropie-Modell fand Ende der 70er Jahre Anwendung in der Bodengeographie. Als Arealheterogenität bezeichnet (ALTMANN & HAASE 1984), wird das Maß für die Unbestimmtheit nach der Shannonschen Formel berechnet und in einem Ansatz zur Kennzeichnung der Landschaftsstruktur diskutiert. Während in dieser Arbeit die Anwendungsmöglichkeit des Entropiemodells als begrenzt angesehen wird, postulieren IBÁÑEZ et al. (1995) eine Vielzahl möglicher Ziele für die Bodenaufnahme und im Bodeninformationssystem, wobei auch die Abschätzung des Informationsverlustes im Prozess der Kartengeneralisierung genannt ist.

Das Maß für die Unbestimmtheit einer Informationsmenge bildet in der Humangeographie die Grundlage für entropiebasierte Bewertungen (PAULOV 1989) und wird in der Kartographie als ein Ansatz zur Unterstützung der Generalisierung vorgestellt (BJØRKE 1996). Zur Messung der Unbestimmtheit, hier als Mannigfaltigkeit betrachtet, existieren Modifikationen des Shannonschen Maßes nach Simpson, die nicht der Informationstheorie entstammen (MCGARIGAL & MARKS 1994).

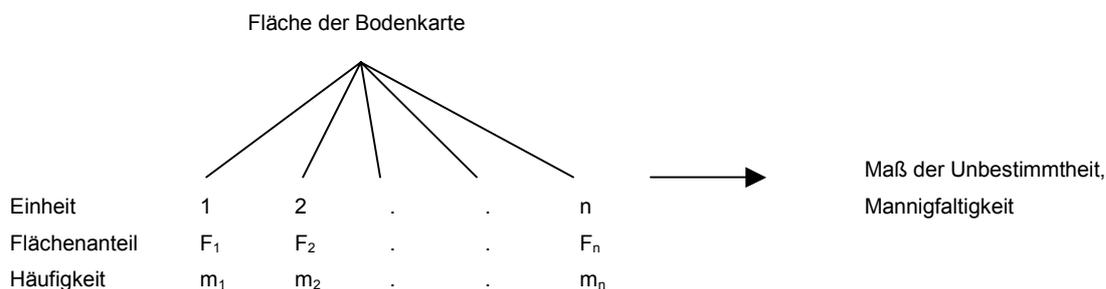


Abb. 19: Eingangsgrößen zur Bestimmung von Diversität/Heterogenität auf der Ebene der Bodenkarte. Während in der physikalischen Messtechnik in der Unordnung eines Systems durch die Entropiemessung Ausgleichsprozesse in der Zeit quantifiziert werden und der informationstheoretische Begriff der Entropie als Maß der Unbestimmtheit für Wahrscheinlichkeiten des Eintretens von Ereignissen Geltung erhält, wird durch die Anwendung dieser Berechnungen auf eine Bodenkarte die kognitive Abgrenzung von Kartiereinheiten, die Mannigfaltigkeit dieser Einheiten quantifiziert. Die Mannigfaltigkeit von n Einheiten die m Flächen zugeordnet sind ($n \ll m$), wird für die Bodenkarte quantifiziert.

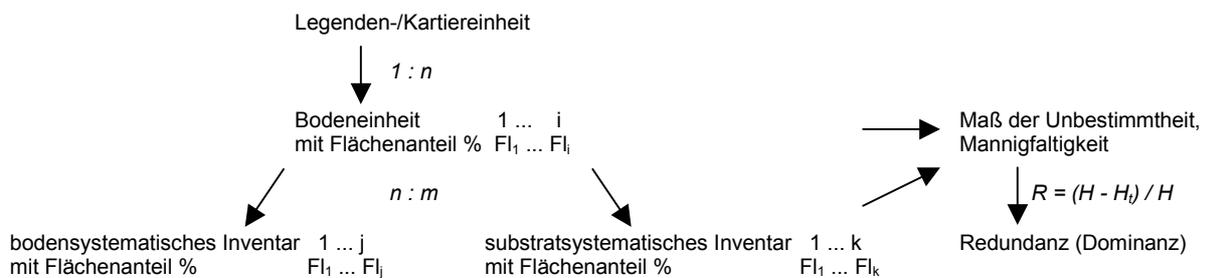
Begriffsabgrenzung Diversität und Heterogenität, wie sie hier einfließen:

- Diversität bezeichnet die quantifizierten Größen, die mit der Proportion der Häufigkeiten berechnet werden.
- Heterogenität bezeichnet die quantifizierten Größen, die mit der Proportion der Flächengrößen berechnet werden.

Die berechneten Indizes eignen sich zum Vergleich der Mannigfaltigkeit/Unbestimmtheit von beschreibenden Teilinventaren und Merkmalskomplexen. Die Indizes bieten sich auch zum Vergleich von thematisch gleichen Kartenwerken unterschiedlicher Maßstäbe und von Kartenblättern gleichen Blattschnitts in thematisch verwandten Kartenwerken an. Die quantifizierten Beträge an Information

erlauben die Abschätzung des Informationsverlustes (IBÁÑEZ et al. 1995) und unterstützen in der Entscheidung über die Schranke der Darstellbarkeit eines Themas (FINKE et al. 1998).

Die Bestimmung der Inventar-Areal-Heterogenität nach der Shannon-Formel für die Bodeneinheiten und deren Teilinventare in den Legendeneinheiten quantifiziert die innere Heterogenität von Legendeneinheiten/Kartiereinheiten.



H - Entropiemaß für die Bodeneinheit; H_i - Entropiemaß für den betrachteten Teilkomplex; 1 : n, n : m - Beziehungstypen

Abb. 20: Eingangsgrößen zur Bestimmung von innerer Heterogenität der Einheiten. Die Legendeneinheiten werden durch den Shannon-Index abstrakt ohne inhaltliche Betrachtung quantifiziert. Die Größe dient zum Vergleich der Heterogenität der Legendeneinheiten. Typische Flächenanteilskombinationen können ermittelt werden. Das Redundanzmaß R eignet sich zur Aussage, welche der betrachteten Teilkomplexe den Aufbau der Legendeneinheiten dominieren, steuern.

Diese Größe ist Bestandteil der Eingangsdaten zur Gruppierung der Flächenstruktur (Durchdringung, Flächenanteil und Inventar-Areal-Heterogenität der Legendeneinheiten) und bietet sich auch zur Bewertung der Dominanz von Teilinventaren an.

9.1.2 Gestaltindizes

Fläche und Umfang eines Polygons sind Grundgrößen in einem GIS. Gestaltindizes kennzeichnen die Ausprägung der Gestalt von Einzelflächen und Gruppen von Flächen (z. B. Legendeneinheiten). Darauf basieren Entscheidungen für die Eliminierung oder Zusammenfassung aufgrund von Mindestflächengrößen und räumlichem Nebeneinander. Um typische Ausprägungen von Flächenformen zu erfassen, z. B. sind Gleye und Auenböden an langgestreckten komplizierten Flächengebilden kenntlich, werden Indizes zur Kennzeichnung und zum Formenvergleich von Einzelflächen und attribuierten Flächen interessant. Die hier eingesetzten Formeln basieren auf dem Umfang-Fläche Verhältnis und stammen aus der Landschaftsanalyse (MCGARIGAL & MARKS 1994). Die Tatsache, dass Gestaltindizes in der Landschaftsanalyse eingesetzt werden, beruht darauf, dass die Form von Flächen, die in diesem Fachgebiet als Landschaftsobjekte bezeichnet werden, die Austauschbeziehungen mit ihrer Umgebung beeinflusst. Die quantitative Erfassung dieser geometrischen Eigenschaft (Abb. 21) fließt dort zur Bewertung ökologischer Prozesse ein. Für den Teilschritt der Flächenaggregation sind Gestaltindizes für Legendeneinheiten, die durch diese Größen gruppiert werden, einsetzbar. Die Gruppen sind verbal fassbar und bieten mit weiteren gruppierten landschaftlichen Indizes eine Entscheidungsgrundlage für Zusammenfassungen.

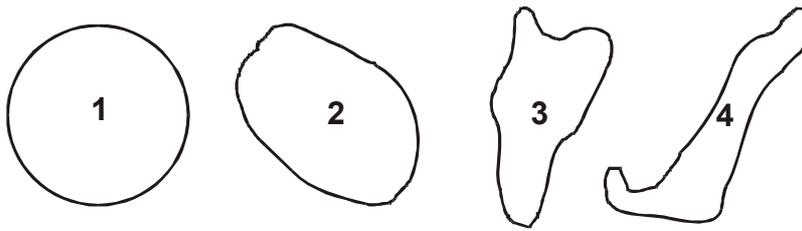


Abb. 21: Verschieden gestaltete Flächen (2-4) in ihrer Abweichung vom Kreis (1). Der Gestaltindex wächst mit zunehmender Abweichung vom Kreis.

Der Gestaltindex und dessen Variation in der Legendeneinheit oder Teilinventareinheit kann durch das Hierarchische Clustern oder die Zweidimensionale Gruppierung gruppiert werden. Die Gruppen eines interessierenden Lösungsschnitts sind dann, als Kennzeichner der Ausprägung der Gestalt, Teil der bewertenden Zusammenfassung.

Weitere Indizes zur Charakterisierung von Landschaftsobjekten sind der Zusammenstellung in BORG & FICHTELMANN (1998) entnommen. Die Formeln orientieren sich am Kreis gleichen Flächeninhalts. Die hier einbezogenen Größen sind Kompaktheit, Kreisförmigkeit, Gestrecktheit und Dünneverhältnis. Ihre Anwendbarkeit ist mit dem Gestaltindex vergleichbar.

Eine weitere Möglichkeit zur Quantifizierung der Gestalt von Flächenobjekten bietet der fraktale Ansatz. Auf den Begründer der fraktalen Geometrie B. B. Mandelbrot geht der Ausdruck der Fraktalen Dimension zurück. Er schlägt die Umfang-Fläche Methode zur Berechnung der Fraktalen Dimension (D) vor (dargestellt in MANDELBROT 1987), die den Grad der Komplexität eines Polygons charakterisiert. Das Verhältnis Umfang (P) : Fläche (A) ist bei fraktalen Objekten durch die Beziehung $\ln P \approx \frac{1}{2}D \ln A$ darstellbar.

9.1.3 Indizes zur Kennzeichnung der räumlichen Struktur

- Einsprengung und Nebeneinanderstellung
- Kennzeichnung von Auffälligkeit
- Räumliche Korrelation

Die Formel zur Berechnung des Index zur Kennzeichnung der Einsprengung und Nebeneinanderstellung (IJI) der Flächen basiert auf dem Entropie-Modell. Die Grundformel der Entropie nach Shannon ist Teil der summarischen Berechnung (s. Leitfaden). In die Berechnung des IJI gehen die gemeinsamen Kanten zwischen den Flächen gleicher Attribute ein, wobei für jede Kombination die Proportion der Kantenlänge der benachbarten Attribute zur gesamten Kantenlänge berechnet wird. Die Formel stammt aus der Landschaftsanalyse und ist durch LI & REYNOLDS (1993) korrigiert und für verschiedene simulierte Landschaften gerechnet und bewertet worden. Hohe Indizes für Flächen des gleichen Attributs sind ein Hinweis auf eine ausgeprägte nachbarschaftliche Einsprengung in der Karte/Landschaft, d. h. die entsprechenden Flächenattribute besitzen Nachbarn vieler verschiedener

Flächen anderer Attribute. Der Index hängt von zwei Parametern ab, einmal ist das die Anzahl an Nachbarn und zum anderen die gemeinsame Kantenlänge der Nachbarn.

Zusammen mit Flächengröße und innerer Heterogenität eignet sich der Einsprengungs- und Nebeneinanderstellungs-Index zur Kennzeichnung der Komplexität der einzelnen Einheiten in der Karte. Auf diese Größen aufgesetzte Gruppierungsverfahren erbringen in der Regel Lösungsgruppen, die bewertet und verbal benannt werden können.

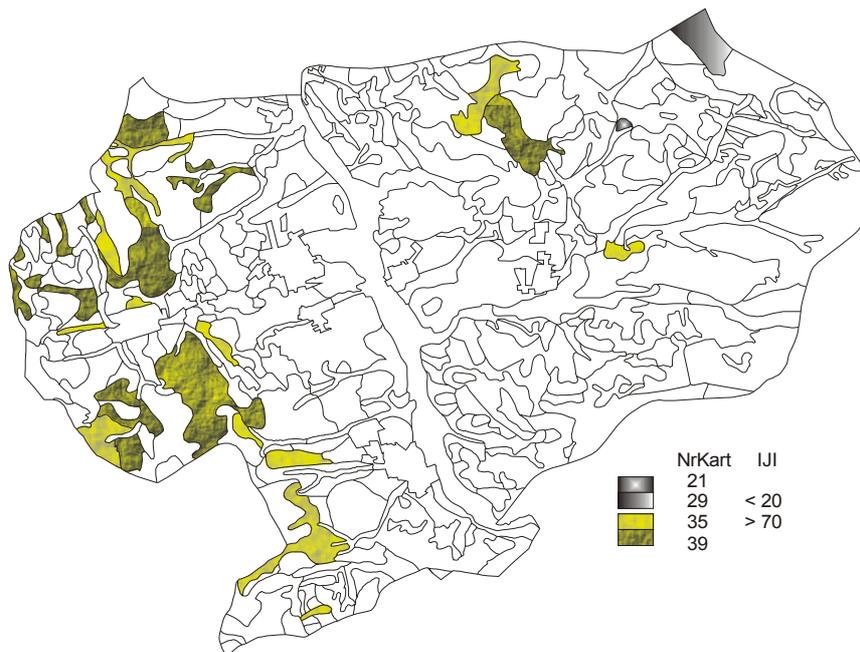


Abb. 22: Geringe und hohe Einsprengungsindizes für Kartiereinheiten einer Inselkarte. Die Bodenkarte wird von 55 Kartiereinheiten aufgebaut. Die Kartenskizze verdeutlicht die Ausprägung des Einsprengungsindex. Die beiden Kartiereinheiten geringer Einsprengung sind deutlich unterschieden in ihrer Flächengröße. Sie besitzen aufgrund ihres einmaligen Auftretens nur wenige Nachbarn. Die beiden Kartiereinheiten hoher Einsprengungsindizes konzentrieren sich in ihrer Lage auf Teilbereiche in der Inselkarte. Die hohen Indizes ergeben sich durch die innige Nachbarschaft mit vielen verschiedenen Kartiereinheiten. Der Index ist nicht von der Flächengröße abhängig, er quantifiziert Flächeneinheiten nach ihren Nachbarschaften. Hohe Indizes sind ein Ausdruck für komplexe Nachbarschaftsverhältnisse.

Im funktionellen Verlauf des Grundterms der Entropie $-p \cdot \ln(p)$ liegt das Maximum der Funktion bei: $1 / e \approx 0,37$ ($e = 2,71.. =$ Eulersche Zahl). Eine Eigenschaft mit dieser Wahrscheinlichkeit ist besonders auffällig. Für den Goldenen Schnitt gilt: Die kürzere Strecke verhält sich zur längeren wie die längere zur ganzen, ungeteilten Strecke. Daraus folgt das Streckenverhältnis: $1 - (\sqrt{5} - 1) / 2 \approx 0,38$.

Das Maximum der Entropie und der Goldene Schnitt liegen dicht beieinander. Der Goldene Schnitt bestimmte, über Epochen hinweg, wesentlich die Gestaltung von Kunstwerken (VÖLZ 1990). Das Maximum der Entropie wird hier als Überraschungswert bezeichnet und kennzeichnet eine Auffälligkeit. Die Auffälligkeit einer Fläche oder von Flächen gleicher Attribute in der Karte kann durch den Verlauf der Entropie gemessen werden.

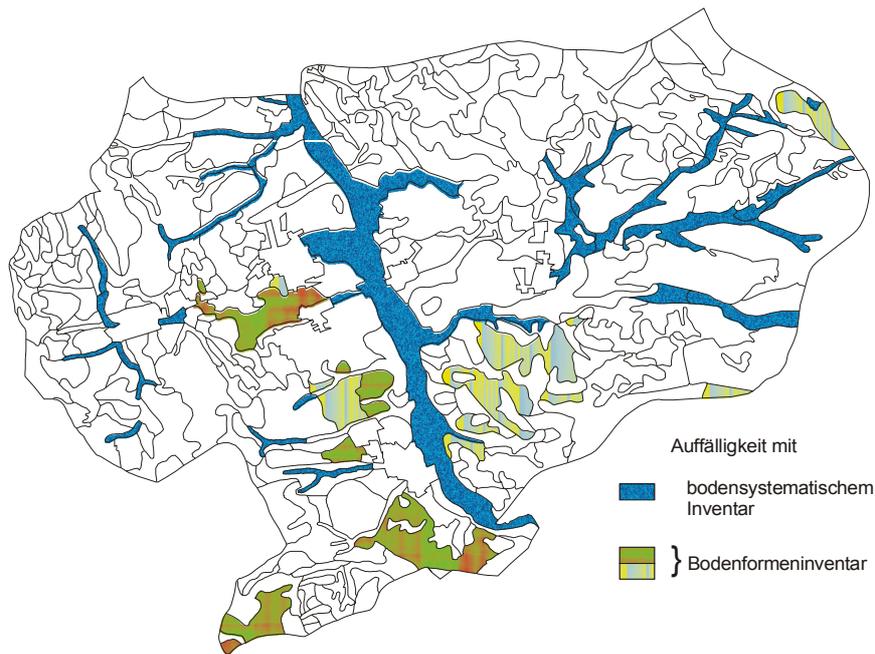


Abb. 23: Auffällige Flächeneinheiten einer Inselkarte. Die Auffälligkeit bezieht sich auf den standardisierten Anteil an Fläche, einer zu charakterisierenden Attributmenge, unabhängig von Farb- und Signaturgebung, Kontrast und Dichte. Gezeichnet sind in der Kartenskizze die Einheiten, die nahe dem Maximum der Entropie von 37% liegen; ein Hinweis auf die Erhaltungswürdigkeit solcher Elemente im Folgemaßstab.

Weitere Maße zur Kennzeichnung räumlicher Verteilung stellen die Räumliche Autokorrelation und die Join-Count-Methode bereit. Aus der Zeitreihenanalyse stammend (CHATFIELD 1982) wird Autokorrelation in den Raum übertragen. Die Formel zur Berechnung geht auf MORAN (1948) zurück. Da die Formel für metrische Variable definiert ist, kann sie nur bedingt für die Analyse von Karten eingesetzt werden. Hier setzt die Join-Count-Statistik ein, die sich als Matrix aller nachbarschaftlichen Kombinationen darstellt. Zur Betrachtung sollten nur wenige Abstufungen herangezogen werden, denn zu viele Abstufungen führen zu einer Verwischung der Muster. Dabei kann es sich um übergreifende Beschreibungen oder grobe Unterteilungen von Merkmalen handeln, die in den Flächen Gültigkeit besitzen. Als ein Beispiel kann die Verteilung der Lössarten in einem Kartierungsgebiet oder einer Karte der Join-Count-Statistik unterzogen werden (Tab. 10). Damit ist eine Aussage bezüglich der Struktur der Heterogenität möglich.

Erläuterung des Ausdrucks Kombination für die Join-Count Statistik:

- Reine Kombination bedeutet, dass benachbarte Flächen den gleichen Inhalt besitzen.
- Hybride Kombination bedeutet, dass benachbarte Flächen unterschiedlichen Inhalts sind.

Substratbegriff		Kombination	Grenz- länge [m]	Häufigkeit	erwartete Häufigkeit	Abweichung*
Lösssand	Lösssand	rein	73726	181	150	
andere Substrate	Lösssand	hybrid	76591	172	149	
Lösssand	Sandlöss	hybrid	42603	111	170	Antityp
andere Substrate	Sandlöss	hybrid	40475	99	84	
andere Substrate	andere Substrate	rein	30992	72	36	Typ
Löss	Sandlöss	hybrid	23695	66	50	
Sandlöss	Sandlöss	rein	24519	52	47	
Löss	Löss	rein	22796	51	12	Typ
andere Substrate	Löss	hybrid	12435	26	43	
Löss	Lösssand	hybrid	3209	14	88	Antityp
Schuttlöss	Sandlöss	hybrid	7361	13	9	
Schuttlöss	Löss	hybrid	3289	8	5	
Schuttlöss	andere Substrate	hybrid	2331	5	8	
Schuttlöss	Lösssand	hybrid	871	3	17	Antityp

*verbale Kennzeichnung nach χ^2 -Komponententest mit Schranke $\chi^2_{0,1;1} = 10,8$

Tab. 10: Join-Count-Statistik für die Lössarten in einer Inselkarte. 59% der Kombinationen der Kartiereinheiten sind bezogen auf den umgelegten Substratbegriff in ihrer Kombinationsart hybrid. 41% der Kombinationen der Kartiereinheiten sind in ihrer Kombinationsart rein. Eine Struktur in der räumlichen Heterogenität dieser Merkmale ist erkenntlich. Zwei reine Kombinationsarten überschreiten mit ihrer berechneten Häufigkeit signifikant die erwartete Häufigkeit. Eine beobachtete Häufigkeit nahe der erwarteten Häufigkeit ist in ihrer räumlichen Verteilung als zufällig zu bewerten.

Dieses Verfahren ermöglicht, den räumlichen Zusammenhang von Flächen, die durch wenige Merkmale gekennzeichnet sind, zu charakterisieren.

9.2 Multivariate Gruppierungsverfahren

Aus den multivariaten Verfahren wurden solche ausgewählt, die im weiteren Sinne eine Eignung zur explorativen Gruppenfindung aufweisen und sich zur Gruppierung der Inhalte der Legenden- und Kartiereinheiten und ihrer geometrie- und strukturbeschreibenden Eigenschaften eignen:

- Clusterverfahren
- Multidimensionale Skalierung
- Formale Begriffsanalyse
- Konfigurationsfrequenzanalyse.

9.2.1 Gruppierung durch Experten

Hierunter ist nichts weiter zu verstehen, als dass der Fachbearbeiter eine Gruppierung vornimmt, die dann im Hinblick auf ihre Güte nach mathematischen Kriterien geprüft wird; d. h. es folgt die Berechnung der F- und t-Werte mit den ausgewählten Eigenschaften (Variablen) in den neuen Gruppen.

9.2.2 Hierarchische Clusteranalyse

Die Clusteranalyse ist ein allgemein bekanntes und in diversen Problemstellungen zur Strukturierung, Gruppierung eingesetztes Verfahren. In der Bodenkunde wurde dieses Verfahren anfänglich zur Feststellung des Übereinstimmungsgrades zwischen konventioneller und numerischer Taxonomie eingesetzt (SCHEFFER & SCHACHTSCHABEL 1997). Das Verfahren dient zur Gruppierung von Objekten (Horizonten, Schichten, Profilen, Kartiereinheiten usw.), um auf Basis von Primärdaten Grundlagen für Klassifikationssysteme zu legen. Sind geeignete Klassifikationssysteme nicht gegeben, so kann das Verfahren eingesetzt werden, um für konkrete Fragestellungen Gruppen zu finden. Für die Clusteranalyse sind zwei Schritte wichtig, die wesentlich auf das Ergebnis Einfluss haben:

- 1) die Wahl des Proximitätsmaßes (Ähnlichkeit/Distanz) und
- 2) die Wahl des Fusionierungsalgorithmus.

Für die Unterstützung des Aggregierungsprozesses wurden mehrere Proximitätsmaße und Fusionierungsalgorithmen programmiert:

- Proximitätsmaße
 - Euklidische Distanz
 - Quadrierte Euklidische Distanz
 - Tanimoto (Jaccard)-Koeffizient
 - Q-Korrelationskoeffizient
 - Chi-Quadrat
- Fusionierungsalgorithmen
 - Nächster Nachbar (Single-Linkage)
 - Entferntester Nachbar (Complete-Linkage).

Sie sind beliebig kombinierbar, vorausgesetzt, die entsprechende Kombination ist für die Ausgangsdaten sinnvoll.

Während das euklidische Distanzmodell die Unähnlichkeiten zwischen den Objekten widerspiegelt, gibt der Q-Korrelationskoeffizient Ähnlichkeiten der Objekte in Bezug auf gleiche Profilverläufe wieder. Der Tanimoto-Koeffizient misst den relativen Anteil gemeinsamer Eigenschaften für binär skalierte Variablen. Das Chi-Quadrat-Maß gibt die Unähnlichkeiten für Häufigkeitsdaten wieder.

Die Ergebnisse der beiden Fusionierungsalgorithmen unterscheiden sich deutlich. BACKHAUS et al. (1996) charakterisieren diese agglomerativen Clusterverfahren. Sie unterscheiden in kontrahierende, konservative und dilatierende Verfahren, wobei das Single-Linkage ein kontrahierendes Verfahren darstellt, welches zur Kettenbildung neigt und das Complete-Linkage als dilatierend eingeschätzt wird, d. h. es neigt dazu, kleine Gruppen zu bilden.

9.2.3 Zweidimensionale Skalierung

Die Multidimensionale Skalierung (MDS) bezeichnet eine Gruppe von Methoden, die Ähnlichkeiten räumlich repräsentieren. Die MDS dient unter anderem in der sozialwissenschaftlichen Forschung dazu, die subjektive Bewertung von Objekten durch Personen zu bestimmen.

Zur Unterstützung des Aggregierungsprozesses wurde eine Nichtmetrische Zweidimensionale Skalierung programmiert. Aufbauend auf den Proximitätsmaßen der Clusteranalyse, welche die Unähnlichkeiten bzw. Ähnlichkeiten zwischen den Objekten widerspiegeln, wird hier eine Rangreihung der Objekte vorgenommen. Die Rangordnung der Objekte wird durch ein iteratives Optimierungsverfahren (Algorithmus von KRUSKAL) in die Wahrnehmungsebene projiziert. Ähnlichkeit zwischen den Objekten drückt sich in der Ebene durch ihre Entfernung zueinander aus. Die Auswahl einer Gruppenlösung kann dann durch einen auf die Koordinaten aufgesetzten Clusterprozess vorbereitet werden.

9.2.4 Bewertung von Gruppierungen, F- und t-Werte

Zur Bewertung von Gruppierungen dienen die F-, t-Werte und Gruppen-Mittelwerte der Variablen. Die Berechnung dieser Größen ist Bestandteil der drei Gruppierungsmethoden Expertenclustern, Hierarchisches Clustern und Zweidimensionale Skalierung.

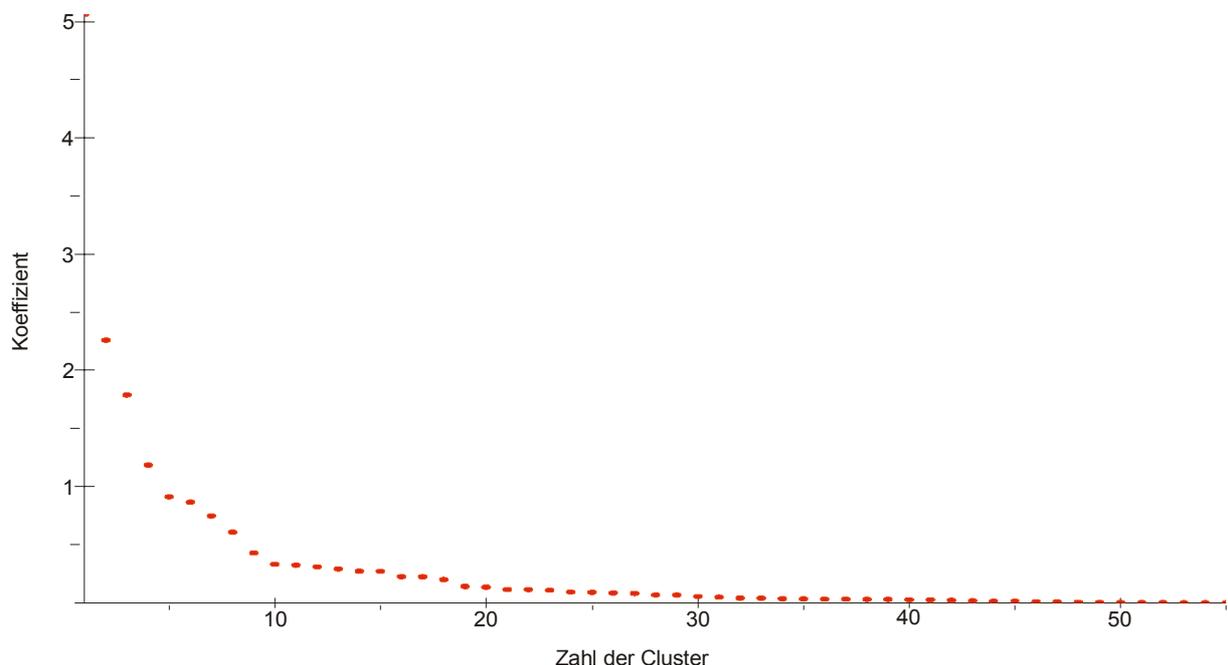


Abb. 24: Beispiel für den Anstieg des Koeffizienten im Fusionierungsprozess auf der Grundlage eines Distanzmaßes. Der Koeffizient steigt kontinuierlich bis zur Clusterlösung 18 an. Die weitere Fusionierung der Objekte führt zu Heterogenitätssprüngen. Sucht man in diesem Bereich (Gruppenzahl \ll 18) nach der optimalen Gruppenlösung, sollten an den Sprüngen Lösungen geschnitten und analysiert werden.

Im hierarchischen Clusterverfahren erfolgt die Festlegung auf eine Gruppierung in Lösungsschnitten. Gruppenlösungen werden dort geschnitten, wo durch den agglomerativen Prozess Heterogenitäts-

sprünge entstehen (Abb. 24). So wird herausgearbeitet, welche Objekte die sprunghafte Zunahme der Heterogenität verursachen. Das ermöglicht Gruppen mit hoher Varianz, abgebildet in den F-Werten der Variablen, im Hinblick auf ihre Ausprägung richtig zu beurteilen.

Das Ergebnis der Zweidimensionalen Skalierung ist die Anordnung der Objekte in der Ebene. Der auf die Koordinaten dieser Objekte aufgesetzte Clusterprozess, ermöglicht die Wahl von Gruppenlösungen (Abb. 25). Die Gruppenzahl wird vorab visuell abgeschätzt. So ist auch hier eine schrittweise Verringerung der Gruppenzahl mit Betrachtung der Heterogenitätssprünge gegeben.

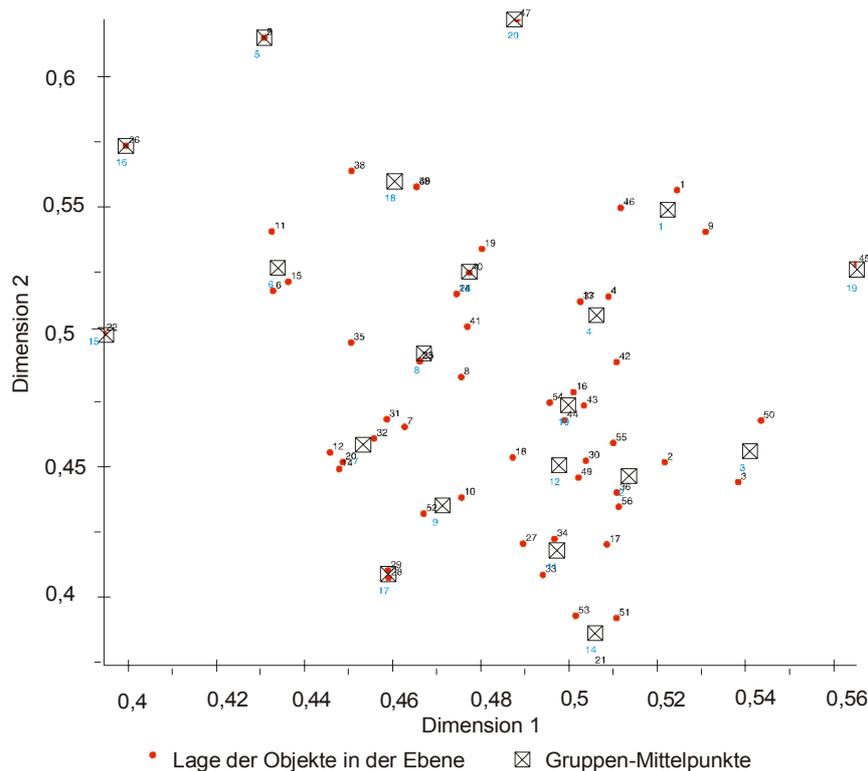


Abb. 25: Beispiel für die Anordnung der Objekte in der Ebene (Konfiguration des abgeleiteten Stimulus aus der Zweidimensionalen Skalierung). Die Zahl der Gruppen wird visuell abgeschätzt und schrittweise verringert.

Durch das Expertenclustern wird eine Gruppierungsvariante erarbeitet. Für diese erfolgt die Berechnung der F-, t-Werte und Gruppen-Mittelwerte der Variablen.

Aussagen über die Homogenität der gefundenen Gruppen einer Lösung sind in Form der F (FISHER) -Werte gegeben. Der Quotient der Varianzen zwischen Gruppen und Erhebungsgesamtheit wird berechnet. F-Werte nahe 0 weisen auf homogene Gruppenlösungen hin. Übersteigt der F-Wert 1, liegt die Varianz für die Variable in der Gruppe über der in der Erhebungsgesamtheit. Das ist ein Kriterium um für metrisch skalierte Variable, die Güte der Lösung zu überprüfen. Homogene Lösungen sind anstrebenswert (Tab. 11). Für binär skalierte Variablen ist die Homogenitätsbetrachtung mit dem Wert 1 nicht anwendbar. Doch können die Variablen mit hoher Varianz in den gebildeten Gruppen im Hinblick auf ihre Ausprägung beurteilt werden.

Die t- (STUDENT) Werte dienen der Charakterisierung der jeweiligen Gruppen einer Lösung. Hier wird die Abweichung zwischen den Mittelwerten der Variable in der Gruppe und der Erhebungsgesamtheit bestimmt. Durch die t-Werte wird angezeigt, welche Eigenschaften (Variablen) vorrangig zur Bildung der entsprechenden Gruppe geführt haben (Tab. 11). Positive t-Werte zeigen, dass der Gruppen-Mittelwert der betrachteten Variablen größer ist als der Mittelwert in der Erhebungsgesamtheit. Negative t-Werte sind entsprechend zu bewerten. Sie bieten die Möglichkeit, die Gruppen eines Lösungsschnitts verbal zu fassen. Als Faustregel kann gelten, dass die Abweichung vom Gesamtmittel signifikant (auf dem 95%-Niveau) ist, wenn der Betrag von $t > 2$ ist.

Die Gruppen-Mittelwerte der Variablen unterstützen die Benennung der Gruppen. Sie fließen bereits in die Berechnung der t-Werte ein. Handelt es sich in den Eingangsdaten um metrisch skalierte, nicht standardisierte Variablen, ist eine direkte Benennung der Gruppen durch die Bewertung der Gruppen-Mittelwerte der Variablen gegeben (Tab. 11). Dabei können vorhandene Klassifikationssysteme genutzt werden (Bodenartendiagramme, Humusklassen, Einstufungen ...).

Gruppe	Mitgliederzahl	Variablennummer*	F-Wert	t-Wert	Gruppen-Mittelwert
1	32	I	0,80	-0,01	3,7
		II	0,25	-0,63	1,7
		III	0,17	0,74	3,8
		IV	0,17	0,25	3,1
2	7	I	0,49	0,89	4,3
		II	0,21	1,73	4,4
		III	0,05	-1,35	0,1
		IV	0,12	-1,56	1,9
3	4	I	0,51	0,87	4,3
		II	0,69	1,14	3,8
		III	0,09	-0,14	2,3
		IV	0,59	-1,01	2,3
4	9	I	0,35	-0,39	3,4
		II	0,27	0,40	2,9
		III	0,17	-1,03	0,7
		IV	0,29	-0,06	2,9
5	3	I	0,68	-1,90	2,3
		II	0,25	-0,08	2,3
		III	0,00	-1,45	0,0
		IV	0,65	2,54	4,5

*Einteilung nach AG Boden (1994):

- I - Stufen der Wasserdurchlässigkeit im Oberboden (gering [2] - sehr hoch [5])
- II - Stufen der Wasserdurchlässigkeit im Unterboden (sehr gering [1] - sehr hoch [5])
- III - Vernässungsgrad in Stufen (nicht vernässt [0] - sehr stark vernässt [5])
- IV - Feldkapazität (sehr gering [1] - sehr hoch [5])

Tab. 11: Beispiel für die Darstellung der Bewertungskriterien eines Lösungsschnitts. Dieses sehr einfache Beispiel illustriert die Anwendung der Bewertungskriterien. Es wurden 55 Kartiereinheiten mit ihren mittleren Wasserbindungs- und -bewegungseigenschaften hierarchisch gruppiert. Dargestellt ist der Lösungsschnitt 5. Die F-Werte weisen die Lösung als homogen aus, da sie kleiner als 1 sind. Hohe absolute t-Werte zeigen an, welche der eingeflossenen Variablen für die Bildung der jeweiligen Gruppe entscheidend waren. Die Gruppen-Mittelwerte stehen in diesem Fall für den Stufenwert in der entsprechenden Variable.

9.2.5 Formale Begriffsanalyse

Die Formale Begriffsanalyse (Formal Concept Analysis, FCA) gründet sich auf die Theorie geordneter Mengen (GANTER, WILLE 1996; VOIGT 1996) und wird mittlerweile auch als kommerzielles Softwareprodukt angeboten. RÖDER (1997) nutzt die FCA zur Analyse forstökologischer Daten, einschließlich ihrer Quellen. Durch die Darstellung der Daten in Begriffsnetzen und ihrer inne liegenden Eigenschaft, dass kein Informationsverlust zu verbuchen ist, eignet sich die FCA zur Recherche und Strukturierung von Wissen in Datenbanken. Der Prozess des Erkundens zur Präzisierung einer Fragestellung kann auch in bodenkundlichen Datenbanken sinnvoll das Expertenwissen begleiten. Die Begriffsnetze werden bereits für einfache Zusammenhänge komplex. So ist die FCA mehr als ein Recherche-Werkzeug geeignet und weniger zur Ableitung von sinnvollen Gruppierungen einsetzbar. Da die FCA als kommerzielles Softwareprodukt vorliegt ist sie nicht Bestandteil des vorliegenden Programms. Ihrer Bedeutung wegen werden die notwendigen Erläuterungen zur FCA in diesem Abschnitt gegeben.

Durch Kreuztabellierung wird die formale Definition eines Begriffsverbands gegeben. Das zugehörige Mengenpaar setzt sich aus einer Gegenstandsmenge als Begriffsumfang und einer Merkmalsmenge als Begriffsinhalt zusammen. Die Ordnung des gegebenen Kontextes in der Kreuztabelle wird als gerichteter Graph mit bestehenden Ober- Unter- Begriff-Relationen wiedergegeben. So lassen sich Mengenpaare klassifizieren und Abhängigkeiten zwischen Begriffen ausweisen (Abb. 26).

Hauptgenesegruppe	gefiedert	gefleckt	gestaffelt	gestaffelt am Hang	gestaffelt im Tal	kompakt	konzentrisch
äolische Sedimente	0	49	1	0	0	23	8
Fluss- und Seesedimente (mit Urstromtalsanden)	1	68	29	0	1	257	43
glazigene, glazifluviatile Sedimente einschl. periglaziärer Überprägungen und Bildungen	1	69	20	6	0	419	16
organogene Sedimente	0	22	6	0	0	34	58
anthropogen um-/abgelagerte Sedimente und Materialien	0	22	1	0	0	152	2

Definition von Kontext und Begriff:

- Formaler Kontext $K := (G, M, I)$ ist als Kreuztabelle darstellbar
G, M - Mengen; I - Relation
- Formaler Begriff des Kontextes (G, M, I) ist ein Paar (A, B)
mit $A \subseteq G, B \subseteq M, A' = B, B' = A$
A - Umfang, B - Inhalt des Begriffes (A, B)
 $B(G, M, I)$ - Menge aller Begriffe des Kontextes (G, M, I)

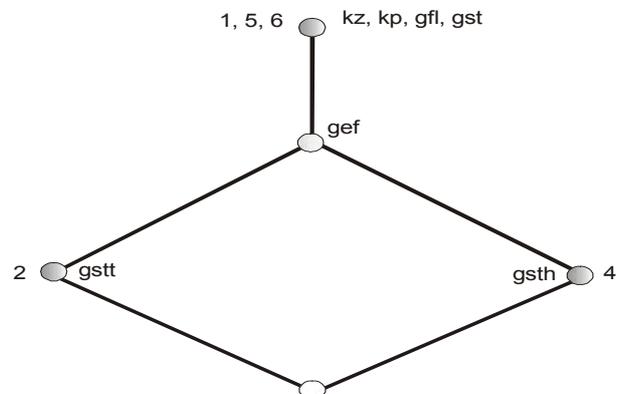


Abb. 26: Beispiel für den Aufbau eines einfachen Begriffsverbands. Der Begriffsverband "Hauptgenesegruppe - Verteilungsmuster" wurde aus den Kartiereinheiten der BK50 L3744, Blatt Potsdam abgeleitet; Verteilungsmuster: AG Boden (1994: 281). Der Begriffsverband besteht aus fünf formalen Begriffen, von denen drei durch Gegenstände (Hauptgenesegruppe) und Merkmale (Verteilungsmuster) gekennzeichnet sind. Ein formaler Begriff ist nur durch ein

Merkmal belegt. Ein leerer Begriff wird zur Vervollständigung des Verbandsgraphen eingeführt. Umfang (Gegenstandsmenge) und Inhalt (Merkmalsmenge) eines Begriffs sind durch den Verlauf der Linien im geordneten Graphen des Begriffsverbands ablesbar. So ist bspw. der Begriff mit dem Merkmal 'gefiedert' (gef) Oberbegriff für die beiden Hauptgenesegruppen 2 und 4.

9.2.6 Konfigurationsfrequenzanalyse

Die Konfigurationsfrequenzanalyse (KFA) beinhaltet Methoden und Verfahren, Merkmalsträger mathematisch-statistisch zu klassifizieren (zu typologisieren). Die verschiedenen Testverfahren zur Typenfindung sind bei LAUTSCH & WEBER (1995) beschrieben. Die Verfahren der KFA wurden in der psychologischen Forschung entwickelt. Dabei ging es um das Auffinden von Syndromen, wie beispielsweise im Konflikt- und Drogenverhalten.

Das für diese Arbeit programmierte Testverfahren basiert auf der zweidimensionalen Kontingenztafel und dem χ^2 -Komponententest von LIENERT (1969). In Form einer Kreuztabellierung wird die Häufigkeit von Merkmalen abgebildet, um in diesen nominal skalierten Variablen Zusammenhänge aufzudecken. Der χ^2 -Global-Wert wird zum Test von Abhängigkeiten zwischen den Merkmalsausprägungen (Nullhypothese: keine Abhängigkeiten) herangezogen. Ist dieser Wert größer als der Vergleichswert in der χ^2 -Tabelle, ist von Abhängigkeiten zwischen den Merkmalsausprägungen auszugehen und die Nullhypothese wird zurückgewiesen. Auch wenn der globale χ^2 -Wert nicht signifikant ist, kann die Häufigkeit in einzelnen Zellen signifikant vom Erwartungswert abweichen. Deshalb werden die Zellen einzeln mit den Schranken in der χ^2 -Verteilung verglichen und Typen und Antitypen entsprechend ihrer Signifikanz gekennzeichnet (Tab. 12). Für den Zellentest können drei Signifikanzniveaus genutzt werden ($\alpha = \{0,1\%, 1\%, 5\%\}$).

Hauptgenesegruppe	Verteilungsmuster	χ^2	Zellentest	Typen und Antitypen
äolische Sedimente	gefleckt	84,82	9,21	T01
	kompakt	18,46	-4,30	AT01
Fluss- und Seesedimente (mit Urstromtalsanden)	gestaffelt	7,76	2,78	T1
glazigene, glazifluviatile Sedimente einschließlich ihrer periglaziärer Überprägungen und Bildungen	gefleckt	6,36	-2,52	AT5
	gestaffelt am Hang	5,22	2,28	T5
	kompakt	9,93	3,15	T1
	konzentrisch	24,52	-4,95	AT01
organogene Sedimente	kompakt	27,43	-5,24	AT01
	konzentrisch	184,37	13,58	T01
anthropogen um-/abgelagerte Sedimente und Materialien	gestaffelt	5,84	-2,42	AT5
	kompakt	8,68	2,95	T1
	konzentrisch	13,42	-3,66	AT01

Irrtumswahrscheinlichkeit α : $\alpha = 0,1\%$ - T01, AT01; $\alpha = 1\%$ - T1, AT1; $\alpha = 5\%$ - T5, AT5

Tab. 12: Signifikante Zellen mit Testentscheidung gemäß χ^2 -Zellentest für die Zuordnung von Hauptgenesegruppe und Verteilungsmuster der Kartiereinheiten der BK50 L3744, Blatt Potsdam

Aufbauend auf dem Begriffsverband (Abb. 26), ist es von Interesse, hinsichtlich der aus dem geordneten Graphen ablesbaren Begriffe eine Typisierung dieser Begriffe vorzunehmen. Der globale χ^2 -Wert = 404,90 übersteigt wesentlich den entsprechenden Vergleichswert $\chi^2_{0,001,24} = 51,18$. So kann damit gerechnet werden, dass Typen auftreten. Typen und Antitypen kennzeichnen die über- bzw. unterrepräsentierten Tabellenfelder in der Kreuztabelle. Der lokale χ^2 -Zellentest weist bei einer Irrtumswahrscheinlichkeit $\alpha = 0,1\%$ zwei Typen (T01) und vier Antitypen (AT01) aus. Hier lässt sich die Aussage wiederfinden, dass die Bodenformen in den Kartiereinheiten auf Dünenfeldern, bedingt durch die variierende Mächtigkeit der Flugsanddecke, in großen Arealen auftreten, die von kleineren Arealen durchsetzt sind (geflecktes Verteilungsmuster). Ein weiterer signifikanter Typ entsteht durch die Kombination organogener Sedimente und konzentrischem Verteilungsmuster, herrührend aus dem zonalen Aufbau von Verlandungsmooren. Alle zwölf signifikanten Kombinationen (Typen als auch Antitypen) sind plausibel und gut interpretierbar. Das Testverfahren weist in diesem Beispiel allerdings nur den Zusammenhang von Zuordnung eines Verteilungsmusters zu einer Kartiereinheit durch den Bodenkundler nach und bestätigt die Plausibilität der Zuordnung.

9.3 Kontextbasierte Regeln

9.3.1 Darstellbarkeitsindizes

Die Festlegung von Mindestflächengrößen in der praktischen Kartographie sind abhängig von Zielmaßstab und Signaturgebung. Dieses Prinzip gilt auch für Bodenkarten, wobei für die absoluten Mindestflächengrößen unterschiedliche Vorschläge existieren (Tab. 13). Die Indizes zur Bestimmung der maximalen Reduktion sind in zweierlei Hinsicht nützlich: Sie geben Auskunft darüber, ob der für die Karte angestrebte Darstellungsmaßstab den Anforderungen nach Harmonie zwischen der natürlichen Variation der kartierten Flächen und den Nutzeranforderungen an die Bodenkarte/Bodenübersichtskarte genügt und sie unterstützen im Prozess der Überführung einer Karte in einen kleineren Folgemaßstab.

ROSSITER (1998) beschreibt die Ansätze, mit denen die minimal lesbare, erkennbare Flächengröße bestimmt wird. Auf der Grundlage der gewählten Konstanten lässt sich die Mindestflächengröße, zum Vergleich mit den realen Flächengrößen von Einzelflächen, für den Zielmaßstab berechnen (Tab.13). Der Index zur Abschätzung der möglichen Reduktion gibt Auskunft darüber, ob die thematische Darstellung der Karte den kartografischen Anforderungen entspricht (Tab. 14).

Maßstab der Darstellung	Mindestflächengröße [ha]	
	Cornell-Ansatz	VINK-Ansatz
1 : 25 000	2,5	1,56
1 : 50 000	10	6,25
1 : 100 000	40	25
1 : 200 000	250	100

Tab. 13: Mindestflächengrößen in Bodenkarten für ausgewählte Darstellungsmaßstäbe nach der Berechnungsgrundlage der Cornell University group on Adequacy of Soil Resource Inventories und VINK. Der Umrechnungsfaktor vom VINK- zum Cornell-Ansatz beträgt 1,6.

Der Cornell-Ansatz beruht auf der Beobachtung, dass in publizierten Bodenkarten Flächen, deren Flächeninhalt kleiner $0,4 \text{ cm}^2$ ist, nur selten vorkommen. Nach dem VINK-Ansatz wird auch die Darstellung kleinerer Flächen vorgesehen (Mindestgröße $0,25 \text{ cm}^2$). Diesem Ansatz folgen die meisten Bodenaufnahmeprogramme in den verschiedenen Ländern. So gibt die Anleitung zur Erarbeitung der Bodenübersichtskarte 1 : 200 000 (HARTWICH et al. 1995: 9) den Hinweis, dass Flächen $< 100 \text{ ha}$ nur in Ausnahmefällen dargestellt werden sollen.

Die Quadratwurzel aus dem Quotienten von durchschnittlicher Flächengröße und Mindestflächengröße, wird als Index der maximalen Reduktion (IMR) bezeichnet. Der optimale Wert für den IMR beträgt 2.

IMR	Bewertung der Größe
> 2	Der Maßstab kann reduziert werden
< 2	Die Karte enthält zu viele Flächen, der Maßstab sollte erhöht werden.
< 1	Die Karte wird unleserlich.

Tab. 14: Bewertung des Index der maximalen Reduktion für einen Darstellungsmaßstab. Erfolgt die Bewertung zur Lesbarkeit für die einzelnen darzustellenden Legendeneinheiten, wird ein IMR von 1 zugrunde gelegt.

9.3.2 Grenzlängenindex

Sind im angestrebten Folgemaßstab weitere Zusammenfassungen notwendig, so führt dies zur Betrachtung der Vergesellschaftung. Der Bodenkundler wird -bezogen auf das jeweilige Niveau der bodengeografischen Dimension- typische Vergesellschaftungen definieren und Legendeneinheiten zusammenführen. Dieser fachliche Prozess wird durch den Grenzlängenindex unterstützt.

Der Grenzlängenindex wird für alle nachbarschaftlichen Kombinationen bereitgestellt und bietet sich als normierter Wert (Grenzlänge der Kombination / gewogenes arithmetisches Mittel) zur Unterstützung der Flächenaggregation an. An einem Beispiel zur Abbildung von Vergesellschaftung zwischen inhaltlich aggregierten Legendeneinheiten durch den Grenzlängenindex wird dessen Handhabung dargestellt (Tab. 15). Die Legendeneinheiten 18 (Gleye aus Sand mit Niedermooren ...) und 19 (Anmoorgleye aus Sand und Niedermoore ...) der BK50, Blatt Potsdam sind in ihrer Substratkennzeichnung als homogen und in der pedologischen Dominanz als quasihomogen gewertet und bilden eine neue Legendeneinheit. Diese ist in den Maßstab 1 : 200 000 nicht zu transportieren. Eine Flächenaggregation wird notwendig.

Nicht immer ist der dominante Grenzlängenindex dem Prinzip plausibler Vergesellschaftung adäquat, z. B. in anthropogen gestörten Gebieten. Zusätzlich ist es notwendig, die realen Flächengrößen bei der Zusammenfassung zu betrachten.

Ein Grenzlängenindex > 1 ist ein Ausdruck für dominante Nachbarschaft im Hinblick auf alle Kombinationen.

Grenzlängenindex	Benachbarte inhaltlich zusammengefasste Legendeneinheiten
4,15	18+19 mit 45+48
2,90	18+19 mit 10+11
0,42	18+19 mit 14+15
0,42	18+19 mit 6+7
0,40	18+19 mit 46+49
0,35	18+19 mit 16+17
0,25	18+19 mit 12+13
0,22	18+19 mit 20
0,17	18+19 mit 50
0,09	18+19 mit 40

Tab. 15: Grenzlängenindizes inhaltlich aggregierter Legendeneinheiten der BK50, Blatt Potsdam am Beispiel der LE 18+19 (Gleye aus Sand mit Niedermooren ... + Anmoorgleye aus Sand und Niedermoore ...) und deren relevante Nachbareinheiten. In dieser Liste bildet der Index die weiteren Zusammenfassungsmöglichkeiten dem bodenkundlichen Sachverhalt entsprechend ab (LE 45 und 48 - Niedermoore aus Torf über Sand ...; LE 10 und 11 - Gleye und Humusgleye aus Sand ...).

9.3.3 Abstandsmessung, Verbindungsflächenermittlung und Flächenverschmelzung

Im Schritt geometrischer Generalisierung sind Entscheidungen, über den Verbleib von Kleinstflächen zu treffen. Dafür bieten sich die Abstandsmessung und die darauf aufbauenden räumlichen Zusammenführungen von Flächen an.

Ein hier vorgeschlagenes Maß h zur Abstandsmessung zwischen Kleinstflächen basiert auf der Kreisfläche der darstellbaren Mindestflächengröße für den jeweiligen Maßstab (Tab. 16). Die Berechnung ist im Leitfaden des Programms erläutert. Flächen mit bestimmten Inhalten, die den Mindestabstand zueinander unterschreiten, können zu größeren Flächen zusammengefasst werden (Abb. 27).

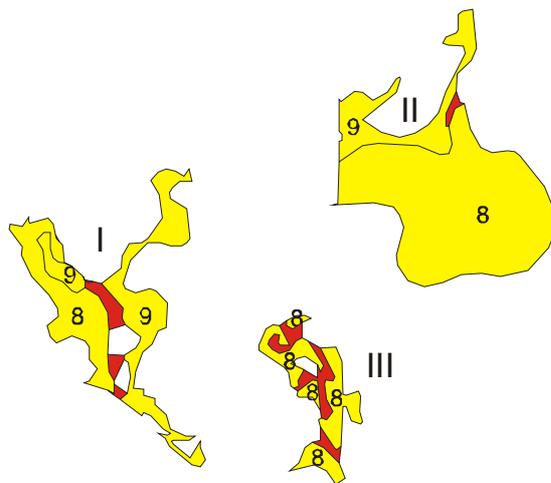
Maßstab der Darstellung	Mindestabstand [m]	
	h	VINK (1975)
1 : 25 000	35,3	50
1 : 50 000	70,5	100
1 : 100 000	141	200
1 : 200 000	282	400

Die Mindestdistanzen zwischen Flächen werden in der geometrischen Generalisierung herangezogen. Sind Zwischenräume zu klein, so werden sie durch Verbindungsflächen räumlich gruppiert. Für Flächen deren Größe unter der erforderlichen Mindestflächengröße liegt, wird über die Verbindungsflächenermittlung oder Flächenverschmelzung eine Flächenvergrößerung erreicht.

Tab. 16: Mindestabstände für ausgewählte Darstellungsmaßstäbe

Kleinstflächen mit bestimmten Inhalten, die durch die Abstandsmessung als räumlich isoliert erkannt werden, können eliminiert werden.

Die Abstandsmessung mit einer Distanz, die größer als der hier vorgeschlagene Mindestabstand ist, eignet sich, um die Teilmengen von Legendeneinheiten herauszufiltern, die in ihrem Auftreten auf Bereiche in der Karte begrenzt sind und räumliche Cluster bilden. So lassen sich Gebiete aufzeigen, in denen durch die Vorgänge der Auswählens und Betonens (Kapitels 6) generalisiert werden kann. Beispiel: Sind durch die Abstandsmessung mit einer Distanz $2 \cdot h$ für bestimmte Legendeneinheiten auf Dünenanden räumliche Cluster festgestellt worden, ist es in diesen Teilbereichen möglich, die Flächen auszuwählen, die den richtigen Eindruck über Verbreitung und strukturellen Zusammenhang vermitteln können (freie Generalisierung ARNBERGER 1987; intuitive Generalisierung HAKE & GRÜNREICH 1994).



Flächenanteile der Legendeneinheiten in den neuen Flächen:

- I - LE 9 mit 53%, LE 8 mit 39%, Le 11 mit 7%, LE6 mit 0,5% und LE 13 mit 0,5%
- II - LE 8 mit 82%, LE 9 mit 17% und Le 19 mit 1%
- III - LE 8 mit 73% und LE 19 mit 27%

Abb. 27: Verbindungsflächenenermittlung zwischen Polygonen mit einer definierten Distanz h , die auf der Abstandsmessung aufbaut.

Die Polygone der Legendeneinheiten 8 (Gleye aus Sand) und 9 (Reliktgleye aus Sand) der BK50, Blatt Potsdam, als Teilmenge einer für die Zusammenfassung interessierenden Polygonmenge, erfüllen die Bedingung: Polygon-Zwischenabstand $\leq h$. Für diese Bereiche werden Verbindungsflächen gebildet, die in anderen Legendeneinheiten liegen. Die zu einer neuen Flächeneinheit verbundenen Polygone bilden zwar nicht in jedem Fall, wie hier im Beispiel, eine kartografisch befriedigende Lösung für einen Zielmaßstab, jedoch sind die Flächeneinheiten im Hinblick auf Inventarverschiebungen auswertbar und die neuen Geometrien für den Editiervorgang nutzbar. Die andere Möglichkeit der Flächenverschmelzung führt über Grid- Funktionalität. Dabei entstehen neue Flächen, die im Verlauf ihrer Umrisslinien von den Eingangsflächen abweichen (siehe Kap. 10)