

## 10. Das Programm "Analytische Werkzeuge der Aggregation und Generalisierung von Bodenkarten"

### 10.1 Einleitende Bemerkungen

Das Programm entstand über einen Zeitraum von zwei Jahren, mit dem Anspruch die Methoden - verbal beschrieben in Kapitel 9 - zusammenzuführen, die mir wesentlich erschienen, um den Prozess der geometrisch-begrifflichen Generalisierung von bodenkundlichen Karten zu unterstützen. Bodenkarten stellen höchst komplexe, häufig nur dem Experten erschließbare fachbezogene Sachverhalte dar. Sie nehmen unter den thematischen Karten eine Sonderstellung ein, da sich das darzustellende Medium Boden schon in größten Maßstäben der exakten Verortung entzieht. Der Zwang, mit dem Fachgebiet eigenen Begriffsinventar zur räumlichen Dimension von Bodenkörpern, zu einem Ergebnis zu gelangen, führt zu Heterogenitäten in der Darstellung die durch das lokale Expertenwissen bewältigt werden.

Das Programm fasst Methoden, analytische Rechnungen, Ermittlung von Kennwerten und kontextbasierte Regeln zusammen, die sich in anderen Fachgebieten etabliert haben und teilweise zum Standard in dort durchgeführten Auswertungen gehören.

Die Anwendung des Programms auf eine Bodenkarte ergibt zunächst einmal eine beträchtliche Anzahl an Indizes und Lösungsgruppen die bewertet werden müssen. Diese Bewertung soll dem Bodenkundler in die Hand gegeben werden, um dem Ziel der adäquaten Darstellung in einem Folgemaßstab mit Transparenz und Objektivität zu genügen. Dabei können die Indizes und Lösungsgruppen den Entscheidungsprozess der Zusammenfassung unterstützen, die entscheidende Instanz bleibt jedoch immer noch der bodenkundliche Experte.

Je einfacher der Sachverhalt wird; d. h. es werden Teilinventare oder Merkmalskomplexe die mit den Flächen assoziiert sind betrachtet, desto erfolgreicher kann mit den Indizes und Lösungsgruppen umgegangen werden.

Es gibt eine Fülle von Ansätzen, die Generalisierung in Kartenwerken zu automatisieren. In Teilen ist dies auch gelungen, nämlich dort wo eine klare hierarchische Struktur zwischen Objekten in Katalogen zusammen mit den Regeln der Verarbeitung abgebildet wird. Die weitere Automatisierung der Generalisierung ist bei den topografischen Karten nachvollziehbar. Die Objekte sind hier ihrem Wesen nach monotypisch fassbar. Das scheint derzeit, eine unabdingbare Voraussetzung der automatischen Generalisierung zu sein. Die assoziative Aufnahmemethode und die damit verbundene Darstellung des Bodens steht dem Anliegen der automatischen Generalisierung im Wege.

## 10.2 Systemvoraussetzungen

Plattform:	Workstation-ArcInfo, ab ARC Version 7.2.1 Patch 2 auf NT4.0 oder Windows 2000
Datenbank und Anbindung:	Microsoft Access 97 oder höher, ODBC 32bit Administrator für die Einrichtung der ODBC-Benutzerdatenquelle(n)

## 10.3 Neue Softwareversion

Esri lieferte im Mai/Juni 2001 ArcGis 8.1 aus. Neben dem neuen Desktop-ArcInfo wird weiterhin die Möglichkeit des Workstation-ArcInfo vorgehalten. In dieser Umgebung ist das Programm uneingeschränkt lauffähig.

## 10.4 Inhaltliche Voraussetzungen

ArcInfo:	vollständige Polygoncoverage(s) mit Polygon-Attributtabelle (.PAT) und Arc-Attributtabelle (.AAT). Das Item %covername%-ID der PAT ist eindeutig belegt und diese Infotabelle enthält zumindest ein Verknüpfungssitem, bspw. die Kartier-/Legendeneinheiten für die Tabellen/Sichten der externen Datenbank. Die Maßeinheit im Koordinatensystem des Covers ist Meter.
Datenbank:	Flächendatenbank(en) mit n Relationen. Eine Tabelle enthält das Verknüpfungsfeld mit den Objekten (bspw. Kartier-/Legendeneinheiten).

## 10.5 Installation

Die Archivdatei tag.zip wird in einem leeren Verzeichnis entpackt (z. B. e:\Mydir\tag).

Der Programm-Verzeichnispfad wird in den Systemeigenschaften über die Ikone System der Systemsteuerung von NT 4.0 oder Windows 2000 vergeben (Systemsteuerung, System, Umgebung, Benutzervariablen für ...).

Variable: fpath

Wert: Programm-Verzeichnispfad (z. B. e:\Mydir\tag)

Der Name der Variablen muss **fpath** lauten.

## 10.6 Programmstart

1. ArcInfo starten und in einen Workspace wechseln

Arc: &workspace <workspace>

2. Aufruf des Programms auf der ARC-Ebene

Arc: &run <Programm-Verzeichnispfad>tag

**tag**, der Name der Routine zum Aufruf des Programms, steht für tools of aggregation and generalization.

Das Hauptmenü des Programms meldet sich (Abb. 28).

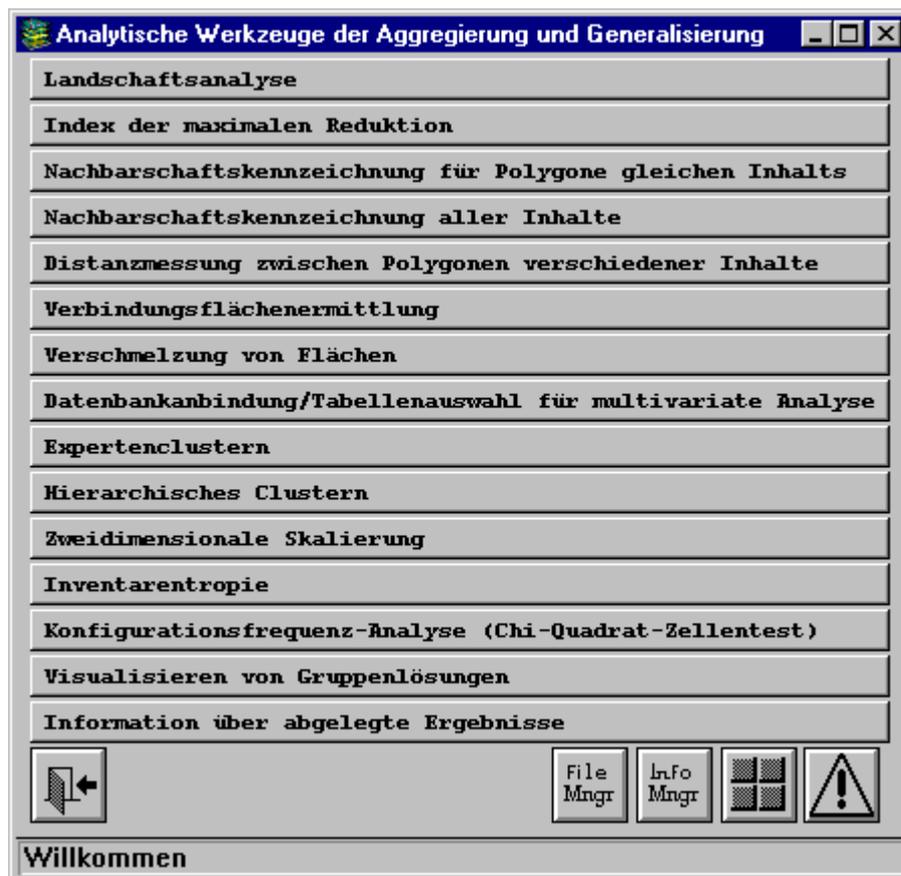


Abb. 28: Bildschirmabzug des Hauptmenüs

*Anmerkung:* Unter Windows 2000 muss ArcInfo vom DOS-Prompt gestartet werden (siehe ArcInfo Help Funktion TASK). Die Funktion TASK arbeitet nicht beim Start aus der Shortcut-Leiste.

## 10.7 Allgemeine Angaben

Das Programm setzt sich aus AML-Code, Form Menu und kompiliertem Fortran77-Code zusammen. Die Daten, die verarbeitet/analysiert werden sollen, bezieht das Programm aus der ArcInfo-Workspace Umgebung als auch aus der über ODBC angebotenen externen Flächendatenbank (Abb. 29).

AML-Notation: <> Zeichenketten in spitzen Klammern bezeichnen Pflichtparameter  
% % In Prozentzeichen eingefasste Zeichenketten bezeichnen Variablenamen  
{ } In geschweifte Klammern eingefasste Zeichenketten bezeichnen optionale Parameter bzw. Platzhalter

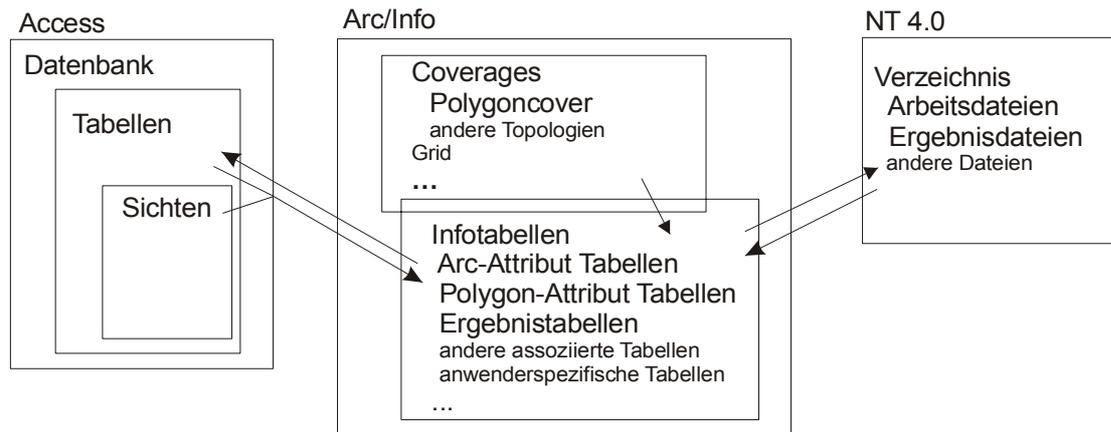


Abb. 29: Datenhaltung und Datenfluss im Programmablauf

Die Prüfung auf Vollständigkeit der Programmdateien kann mit der Routine checklist.aml vorgenommen werden, die folgendermaßen aufgerufen wird:

```
&run <Programm-Verzeichnispfad>checklist <Programm-Verzeichnispfad>.
```

Die Routine vergleicht die im Programm-Verzeichnis vorhandenen Dateien mit der Liste in der Textdatei komponenten.txt.

Aktionen, die Ergebnisdateien produzieren, werden in der Log-Datei result-tag.txt im aktuellen Workspace in der Form:

```
* Komponente: Teilkomponente
  Eingangsgrößen, Parameter
  Ergebnisdatei(en)
```

dokumentiert und können über den button 'Information über abgelegte Ergebnisse' gelistet werden.

Externe Ergebnisdateien werden im aktuellen Workspace abgelegt, die Info-Tabellen im aktuellen Info-Verzeichnis. Existiert eine Ergebnisdatei und es wird eine nochmalige Berechnung angestoßen, wird diese nicht durchgeführt, sondern die alte Ergebnisdatei mit dem Vermerk, dass diese bereits existiert, angezeigt.

Die Liste der abgesetzten Kommandos und Argumente kann man den beiden ESRI log Dateien, log im Workspace und log im Coverage, entnehmen.

Der grundsätzliche Aufbau des Programms und das Zusammenspiel, skizziert in folgender Abbildung (Abb.30), basiert auf der von ESRI in ArcTools realisierten Programmierweise.

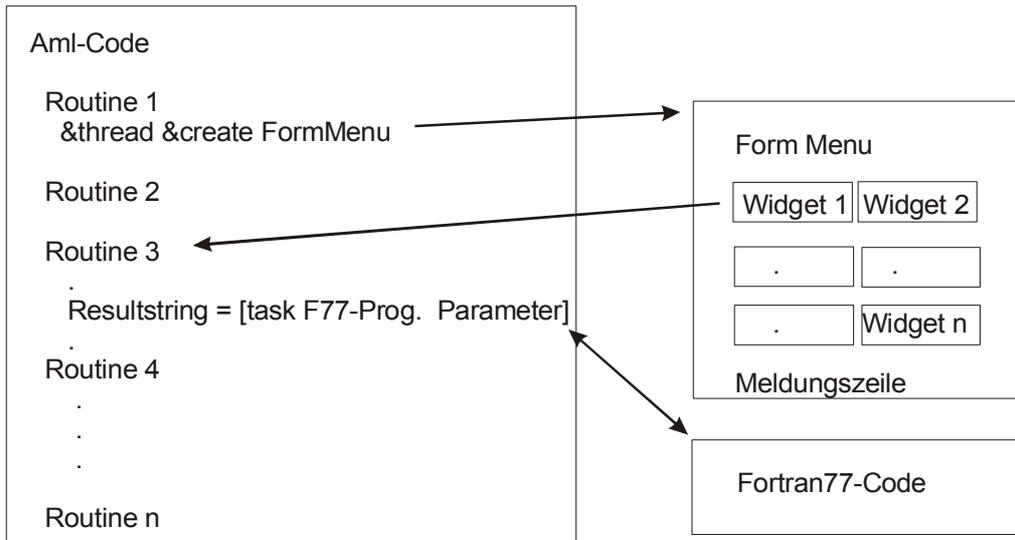


Abb. 30: Grundsätzlicher Aufbau des Programms und das Zusammenspiel von AML, Menu und externem F77-Code

In den Aml-Code-Dateien befinden sich Routinen, die über ihren Namen angesteuert werden und den, ihnen zugehörigen, Aml-Code abarbeiten. Eine Routine initialisiert ein Form-Menü, welches Widgets enthält (Abb. 30). Das sind grafische Steuerelemente zur Parametergewinnung und Aktionsausführung. Die aktionsausführenden Widgets rufen Routinen in den AML-Code-Dateien auf.

Komponenten des Programms, die nicht mit Aml-Code realisiert werden konnten, sind in Fortran77-Standard geschrieben und mit dem Lahey-Fujitsu Fortran95 v.5.5 Compiler kompiliert (Tab. 17). Diese externen Programme werden über die task-Funktion aufgerufen und geben ihre Ergebnisse in Form einer Zeichenkette an den Aml-Prozessor zurück.

Externes Programm in Komponente, Teilkomponente	externer Programmname
Distanzmessung zwischen Polygonen verschiedener Inhalte	dist.exe
Verbindungsflächenermittlung (Linienfindung im vorgegebenen Abstand zwischen Polygonen)	tfe.exe
Tabellenauswahl für multivariate Statistik (Datenmatrix transponieren)	inv.exe
Tabellenauswahl für multivariate Statistik (Daten standardisieren)	tra.exe
Berechnung der F- und t- Werte für alle multivariaten Verfahren	ftw.exe
Hierarchisches Clustern	hcl.exe
Zweidimensionale Skalierung	tds.exe
Zweidimensionale Skalierung (Merkmale, Eigenschaften skalieren)	fea.exe
Zweidimensionale Skalierung (Gruppenbildung)	mtds.exe
Konfigurationsfrequenzanalyse	kfa.exe

Tab. 17: Tabelle der über die task-Funktion eingebundenen externen F77-Programme

Der Quellcode ist in den .f-Dateien gleichen Namens im Programmverzeichnis verfügbar.

Die Liste aller Programmdateien ist in der Textdatei komponenten.txt - Reihenfolge: AML-Code -, Form-Menü -, Execute - und Bitmap Dateien - enthalten.

AML-Notation: <> Zeichenketten in spitzen Klammern bezeichnen Pflichtparameter  
 %% In Prozentzeichen eingefasste Zeichenketten bezeichnen Variablenamen  
 {} In geschweifte Klammern eingefasste Zeichenketten bezeichnen optionale Parameter bzw. Platzhalter

Aktionsausführende Widgets sind - ausgenommen sind die Widgets des Hauptmenüs, die sich selbst erläutern - durch eine Zeile hinterlegt, die beim Klick mit der rechten Maustaste auf dem entsprechenden Widget in der Meldungszeile erscheint. Widgets für die Auswahl von Parametern in Listen bieten bei dieser Aktion ein vom Menü abgelöstes Auswahl-Widget an, welches eine bessere Sicht auf die Liste gibt.

Die Initialisierung eines Folgemenüs erfolgt, bis auf wenige Ausnahmen, mit der Option des Deaktivierens des Vorgängermenüs. Den Fokus erhält das aufgerufene Menü. Die Ausnahmen bestehen bei der Sicht auf die ausgewählten Objekte und Variablen als Bestandteil der Tabellenauswahl aus der Flächendatenbank und den darauf aufbauenden multivariaten Methoden und der Anzeige verschiedener Ergebnisdateien. Bei der Anwendung des Programms erfolgt in vielen Komponenten bei Berechnungen ein Ebenenwechsel zwischen Arc und Arcplot, der zum Flackern der Menüs führt. Um diesen Effekt zu mindern, kann zumindest das inaktive Hauptmenü minimiert werden.

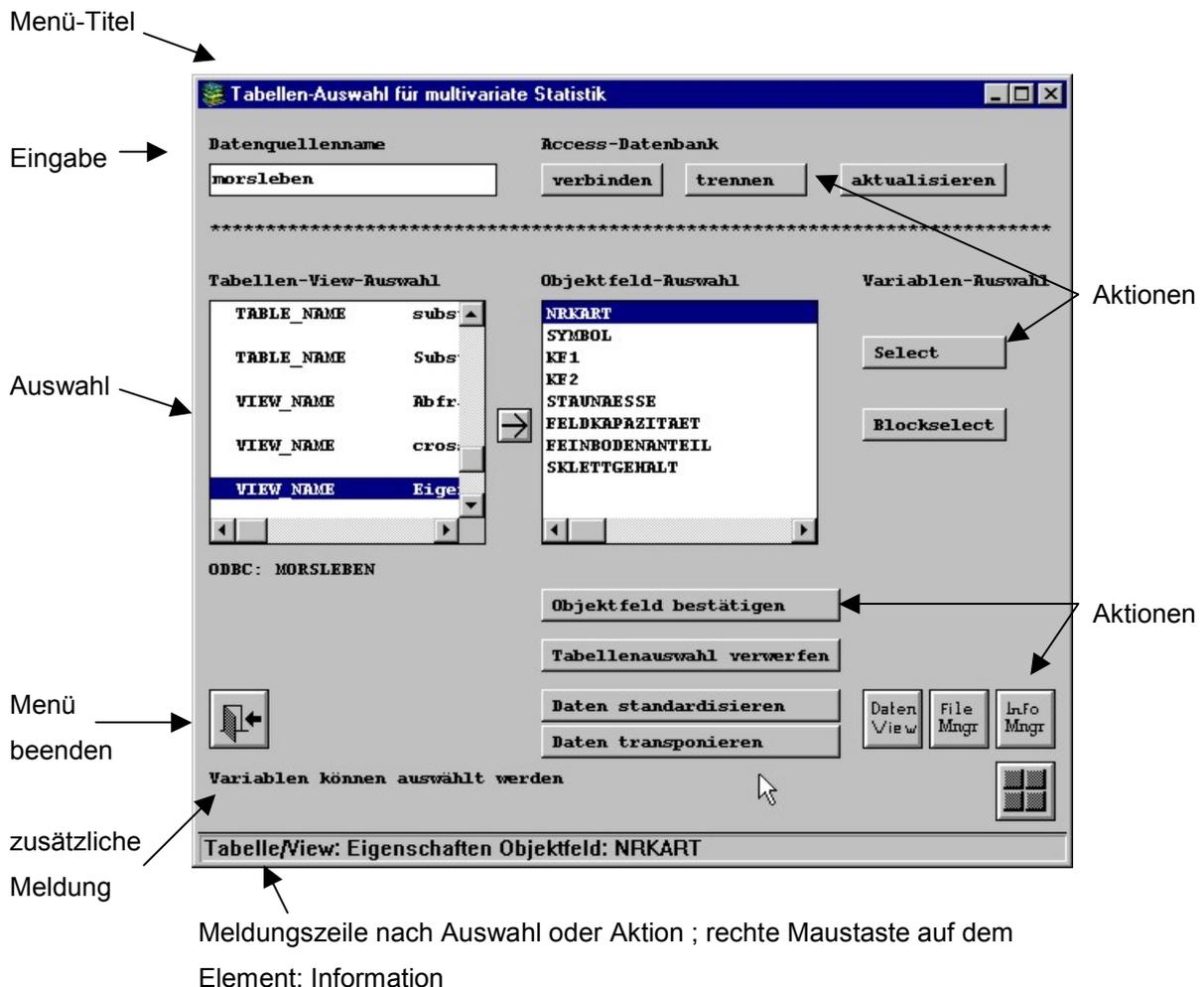


Abb. 31: Aufbau der Oberfläche, erläutert am Beispiel des Menüs zur Datenbankanbindung und Tabellenauswahl

AML-Notation: <> Zeichenketten in spitzen Klammern bezeichnen Pflichtparameter  
 %% In Prozentzeichen eingefasste Zeichenketten bezeichnen Variablenamen  
 {} In geschweifte Klammern eingefasste Zeichenketten bezeichnen optionale Parameter bzw. Platzhalter

Nach Aussage von ESRI (Hotline), ist es nicht möglich, das Flackern der Menüoberflächen bei einem Ebenenwechsel zu verhindern. Begründung aus ArcInfo Help: Why do menus flash? Menus flash when you change between modules. One module needs to write the state of ARC/INFO to the disk so the next module can read that state and restore ARC/INFO.

## 10.8 Programmkomponenten

### 10.8.1 Gliederung der Programmkomponenten

- Allgemein verfügbare Komponenten (Dienstprogramme):
  - Externer Dateimanager
  - Info-Dateimanager
  - Kommandoeingabe
  
- Komponenten, die auf der Polygontopologie und den damit assoziierten Attributen aufbauen:
  - Landschaftsanalyse
    - Berechnung von Heterogenität und Diversität für attributierte Polygone auf der Ebene des Covers mit Entropieformeln
    - Berechnung des Durchdringungs- und Einsprengungsindex für attributierte Polygone
    - Berechnung der Gestaltindizes für die Einzelpolygone und attributierte Polygone
    - Berechnung der Doppelt Logarithmischen Fraktalen Dimension; Gestaltindex für attributierte Polygone
    - Berechnung und Darstellung der Standardisierten Entropie des Vergleichsindex Prozent der Gesamtfläche (Goldener Schnitt) für attributierte Polygone
    - Join-Count Statistik für attributierte Polygone auf Grundlage der Nachbarschaftsmatrix
    - Erstellen von Streudiagrammen (Scatter-Plots mit linearer Regression)
  - Index maximaler Reduktion für einen Maßstab, bezogen auf attributierte Polygone und die Gesamtfläche des Covers
  - Nachbarschaftskennzeichnung für Polygone eines Attributs im Cover; Berechnung der Grenzlänge zu den benachbarten Polygonen anderer Attribute im Cover
  - Grenzlinien-Längen Berechnung; Berechnung der gemeinsamen Kantenlängen zwischen allen attributierten Polygonen
  - Distanzmessung zwischen attributierten Polygonen im Cover
  - Verbindungsflächenermittlung mit den Ergebnissen der Distanzmessung
  - Verschmelzung von Flächen mit den Ergebnissen der Distanzmessung
- Komponenten die auf der Flächendatenbank aufbauen und der Erarbeitung inhaltlicher und landschaftlicher Ähnlichkeit dienen:

- Datenbankanbindung über ODBC an eine Access-Datenbank und Tabellenauswahl für die multivariaten Verfahren
- Expertengruppierung
- Hierarchische Gruppierung (Hierarchisches Clustern)
- Zweidimensionale Skalierung
- Inventar-Entropie, innere Areal-Heterogenität bezogen auf die Flächenanteile des Inventars (inhaltliche Beschreibung) von Objekten
- Konfigurationsfrequenz-Analyse; typisieren kreuztabellierter Häufigkeitsverteilungen
- Visualisieren von Lösungs-Gruppen; grafische Darstellung von Gruppen, die durch die multivariaten Verfahren gewonnen wurden

### 10.8.2 Allgemein verfügbare Komponenten (Dienstprogramme)

Aus allen Menüs können der *externe Dateimanager*, der *Info-Dateimanager* und die *Kommando-eingabe* gestartet werden. Der Info-Dateimanager und die Kommando-eingabe wurden Arctools entnommen und modifiziert. Der externe Dateimanager wurde dem Info-Dateimanager nachgestaltet.

Der *externe Dateimanager* (Abb. 32) startet im aktuellen Workspace und zeigt alle Dateien eines Unterverzeichnisses an. Die Sicht auf ausgewählte Dateien ist über die Eingabe von eingrenzenden Kriterien im Feld 'Suche' möglich. Die jeweils selektierte Datei kann kopiert, gelöscht, aufgelistet oder umbenannt werden. Das Info-Verzeichnis und die Coververzeichnisse sollten damit **nicht** bearbeitet werden.

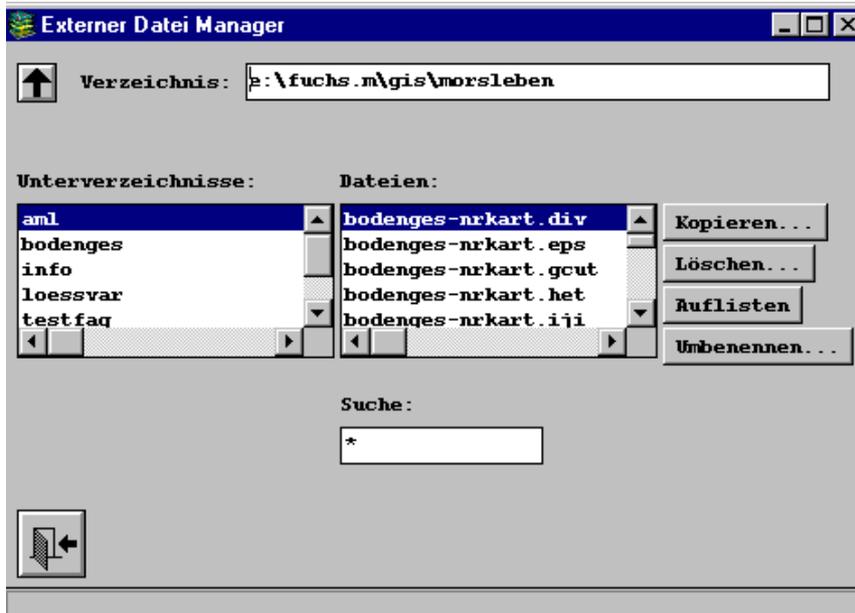


Abb. 32: Bildschirmabzug der Oberfläche der Komponente "Externer Datei Manager"

Der *Info-Dateimanager* (Abb. 33) startet im aktuellen Infoverzeichnis und zeigt alle Infotabellen an. Die Sicht auf ausgewählte Tabellen ist über die Eingabe von eingrenzenden Kriterien im Feld 'Suche' möglich. Die jeweils selektierte Tabelle kann kopiert, gelöscht, aufgelistet, umbenannt oder in die Datenbank exportiert werden. Die Export-Funktion setzt eine Verbindung mit der Datenbank voraus (Komponente 'Datenbankanbindung/...'). Zur Benennung der Access-Tabelle werden im vorgeschlagenen Exportnamen Bindestrich und Punkt durch Unterstrich ersetzt. Der Export schlägt fehl, wenn in der Infotabelle Feldnamen (Items) mit # enthalten sind. Für die selektierte Tabelle ist die Definition der Items abrufbar und für jedes Item können die einmaligen Attributwerte abgefragt werden.

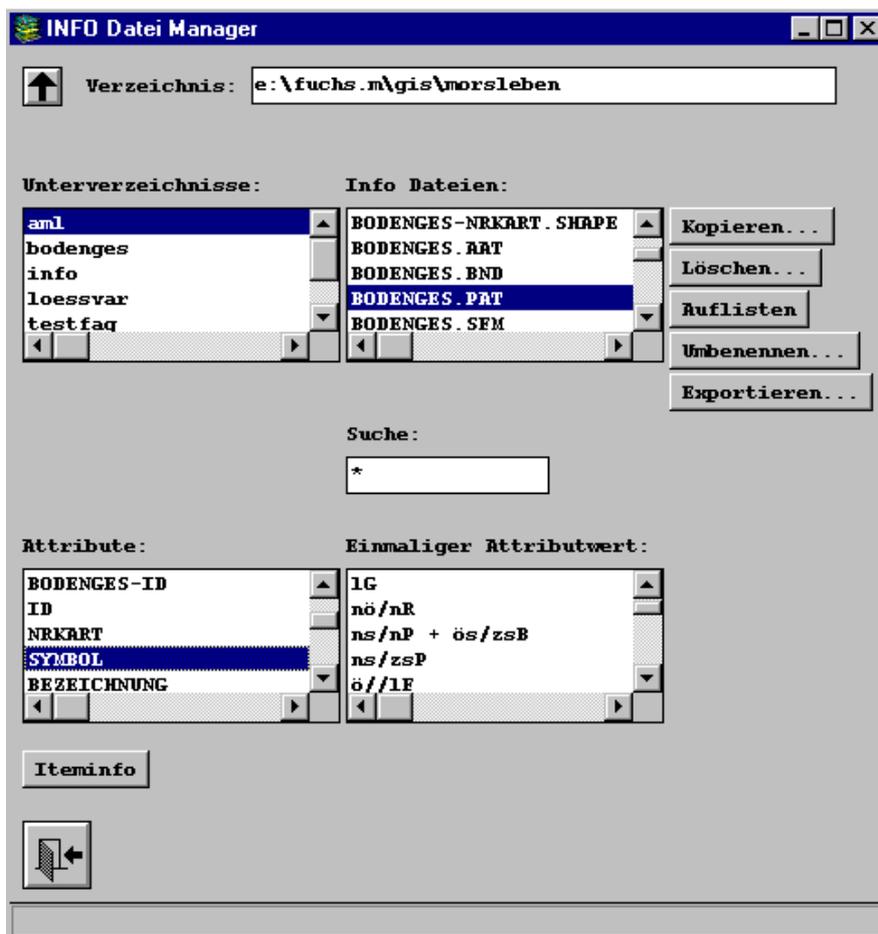


Abb. 33: Bildschirmabzug der Oberfläche der Komponente "INFO Datei Manager"

Die *Kommandoeingabe* (Abb. 34) ermöglicht das Absetzen aller gültigen Befehle auf der jeweiligen Ebene. Befehle, die erfolgreich ausgeführt wurden, bleiben im Pufferfenster erhalten und können über Mausclick wieder geholt und ausgeführt werden. Der Kommandopuffer kann gelöscht werden. Ein Wechsel zum Kommandomode im Arc-Fenster ist möglich. Die Rückkehr zum Menü erfolgt dann mit **&return**. Der Echo-Schalter dient dazu, Aml-Code im Arc-Fenster mit anzuzeigen, eine Möglichkeit beim Auftreten von Fehlern, diese besser einzugrenzen. Zum Deaktivieren der Ausgabe im Arc-Fenster kann der Prompt-Schalter genutzt werden. Das ist z. B. bei den Modulen zur Flächenverschmelzung sinnvoll, die viele Meldungen im Arc-Fenster erzeugen. Für Datenlisting und interaktive Information, die in verschiedenen Komponenten angeboten wird, muss der Prompt-Schalter wieder aktiviert werden.

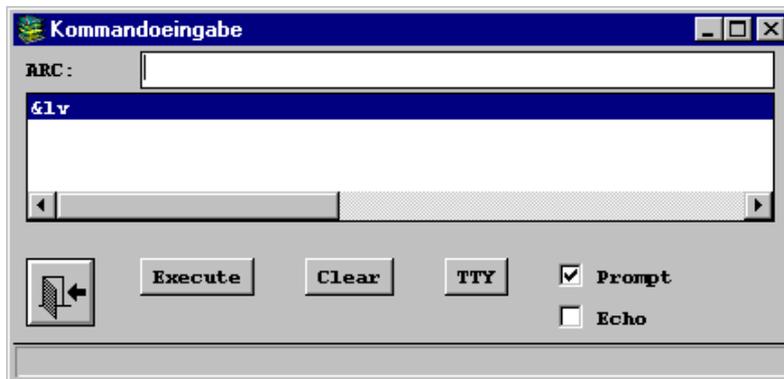


Abb. 34: Bildschirmabzug der Oberfläche der Komponente "Kommandoeingabe"

### 10.8.3 Komponente "Landschaftsanalyse"

Die Landschaftsanalyse (Abb. 35) startet in der aktuellen Workspace-Umgebung und bietet zur Auswahl die Polygoncover und die Datenfelder der zugehörigen Polygon-Attribut-Tabellen an. Die gewählten Parameter werden in der Meldungszeile des Menüs angezeigt. Die Berechnungen sind ausschließlich in Aml realisiert.

Zur weiteren Verarbeitung der Ergebnisse der Landschaftsanalyse, können die Info-Ergebnistabellen über die export-Funktion des Info-Dateimanagers in die Flächendatenbank übernommen werden.

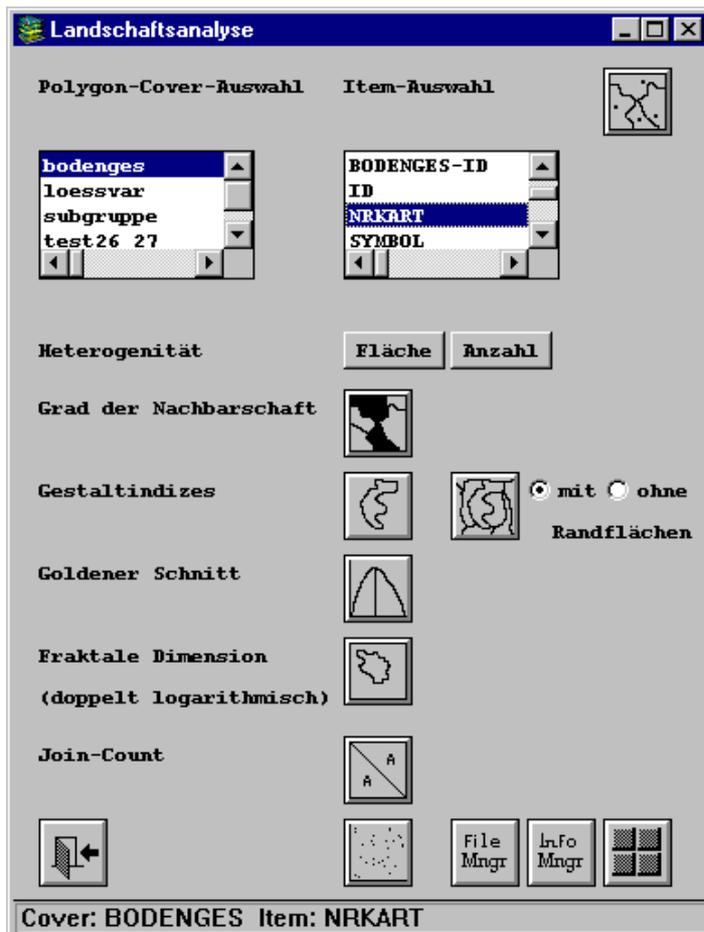


Abb. 35: Bildschirmabzug der Oberfläche der Komponente "Landschaftsanalyse"

*Beschreibung der Teilkomponenten der Landschaftsanalyse:*

Die Teilkomponenten der Landschaftsanalyse greifen auf die Polygontopologie eines Covers und die mit diesem assoziierten Inhalte der zugehörigen Polygon-Attribut-Tabelle zu.

Eingangsgrößen: Polygoncover, Item

Ergebnis: externe Textdateien und Infotabellen mit den berechneten Größen

*Die Teilkomponente "Heterogenität"* basiert auf den Formeln der Entropie. Der Betrag an Information, für  $n$  Inhalte die  $m$  Flächen zugeordnet sind ( $n \ll m$ ), wird für das Cover quantifiziert. Die berechneten Indizes eignen sich zum Vergleich der Heterogenität/Diversität von beschreibenden Teilinventaren und Merkmalskomplexen, die in den Flächen Gültigkeit besitzen. 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 In-

AML-Notation: <> Zeichenketten in spitzen Klammern bezeichnen Pflichtparameter  
 % % In Prozentzeichen eingefasste Zeichenketten bezeichnen Variablenamen  
 {} In geschweifte Klammern eingefasste Zeichenketten bezeichnen optionale Parameter bzw. Platzhalter

formation unterstützen dabei in der Abschätzung des Informationsverlustes und in der Entscheidung über die Schranke der Darstellbarkeit eines Themas.

Die Heterogenität kann in zwei Varianten berechnet werden, einmal mit der Proportion der Flächen-  
größen (Heterogenität) und zum anderen mit der Proportion der Häufigkeiten (Diversität) im ausge-  
wählten Item. Die Ergebnisse werden in jeweils eine externe Datei mit dem Namen %covername%-  
%itemname%.het bzw. .div geschrieben. Folgende Größen werden berechnet:

#### Shannons Diversitäts-Index

$$SHDI = -\sum_{i=1}^m (p_i * \ln p_i) \quad (2) \quad \text{Wertebereich: } 0 \leq SHDI \leq \ln m$$

$p_i$  - Proportion von Flächenanzahl des Flächentyps zur Gesamtanzahl oder Flächengröße des Flächentyps zur Gesamtfläche  
m - Anzahl der Flächentypen (Klassen)

SHDI = 0, wenn die Karte aus einem Flächentyp besteht. SHDI wächst mit der Anzahl verschiedener  
Flächentypen und /oder deren Annäherung an die Gleichverteilung. Eine maximale Entropie (H) liegt  
vor, wenn alle Flächentypen die gleiche Proportion besitzen ( $H_{\max} = \ln m$ ). SHDI > 3.5, die Schranke  
für die Darstellbarkeit eines Themas ist erreicht. Der Absolutbetrag des Wertes wird ausgegeben.

#### Simpsons Diversitäts-Index

$$SIDI = 1 - \sum_{i=1}^m p_i^2 \quad (3) \quad \text{Wertebereich: } 0 \leq SIDI < 1$$

$p_i$  - Proportion von Flächenanzahl des Flächentyps zur Gesamtanzahl oder Flächengröße des Flächentyps zur Gesamtfläche  
m - Anzahl der Flächentypen (Klassen)

SIDI = 0, wenn die Karte aus einem Flächentyp besteht. SIDI nähert sich 1, wenn die Anzahl ver-  
schiedener Flächentypen zunimmt und /oder deren Gleichverteilung, ( $H_{\max} = 1$ ). Der Simpson Diver-  
sitäts-Index repräsentiert die Wahrscheinlichkeit, dass einige zufällig ausgewählte Flächen verschie-  
denen Typen angehören.

#### Modifizierter Simpsons Diversitäts-Index

$$MSIDI = 1 - \ln \sum_{i=1}^m p_i^2 \quad (4) \quad \text{Wertebereich: } MSIDI \geq 0$$

$p_i$  - Proportion von Flächenanzahl des Flächentyps zur Gesamtanzahl oder Flächengröße des Flächentyps zur Gesamtfläche  
m - Anzahl der Flächentypen (Klassen)

MSIDI = 0, wenn die Karte aus einem Flächentyp besteht. MSIDI wächst mit der Anzahl verschiedener Flächentypen und /oder deren Annäherung an die Gleichverteilung. Der Modifizierte Simpson Diversitäts-Index ist in seiner Anwendbarkeit, mit dem Shannon-Maß gleichzusetzen.

Shannons Ebenheits-Index

$$SHEI = \frac{-\sum_{i=1}^m (p_i * \ln p_i)}{\ln m} \quad (5) \quad \text{Wertebereich: } 0 \leq SHEI < 1$$

$p_i$  - Proportion von Flächenanzahl des Flächentyps zur Gesamtanzahl oder Flächengröße des Flächentyps zur Gesamtfläche  
 $m$  - Anzahl der Flächentypen (Klassen)

SHEI = 0, wenn die Karte aus einem Flächentyp besteht und nähert sich 0, wenn die Aufteilung der Fläche in den verschiedenen Flächentypen ungleichmäßiger wird. SHEI = 1, wenn die Aufteilung der Fläche in den verschiedenen Flächentypen genau gleich ist.

Simpsons Ebenheits-Index

$$SIEI = \frac{1 - \sum_{i=1}^m p_i^2}{1 - \left(\frac{1}{m}\right)} \quad (6) \quad \text{Wertebereich: } 0 \leq SIEI < 1$$

Beschreibung siehe SHEI

$p_i$  - Proportion von Flächenanzahl des Flächentyps zur Gesamtanzahl oder Flächengröße des Flächentyps zur Gesamtfläche  
 $m$  - Anzahl der Flächentypen (Klassen)

Modifizierter Simpson Ebenheits-Index

$$MSIEI = \frac{1 - \sum_{i=1}^m p_i^2}{\ln m} \quad (7) \quad \text{Wertebereich: } 0 \leq MSIEI < 1$$

Beschreibung siehe SHEI

$p_i$  - Proportion von Flächenanzahl des Flächentyps zur Gesamtanzahl oder Flächengröße des Flächentyps zur Gesamtfläche  
 $m$  - Anzahl der Flächentypen (Klassen)

Anzahl von Flächentypen

$$PR = m \quad (8) \quad \text{Wertebereich: } PR \geq 1$$

$m$  - Anzahl der Flächentypen (Klassen)

Dichte der Flächentypen

$$PRD = \frac{m}{A}(10000)(X) \quad (9)$$

Anzahl pro X Hektar

$$[X \in (10, 100, 1000, 10000, 100000)]$$

Wertebereich: PRD > 0

m - Anzahl der Flächentypen (Klassen)

A - Gesamte Flächengröße in der Karte/Landschaft [m<sup>2</sup>]

Aufbau der Ergebnisdatei %covername%- %itemname%.het bzw. %covername%- %itemname%.div:

Ergebnisse der Heterogenitäts-Messung für Cover %cover%  
Patch Typen sind im Item %itemname% enthalten

Shannons Areal-Heterogenitäts Index	: %shdi%
Simpsons Areal-Heterogenitäts Index	: %sidi%
Modifizierter Simpson Areal-Heterogenitäts Index:	%msidi%
Shannons Areal-Ebenheits Index	: %shei%
Simpsons Areal-Ebenheits Index	: %siei%
Modifizierter Simpson Areal-Ebenheits Index	: %msiei%
Patch-Anzahl im Cover	: %n%
Patch-Typen-Anzahl im Cover	: %pr%
Patch-Typen-Dichte im Cover	: %prd% [auf 100 ha]

Die Teilkomponente " Durchdringungs- und Einsprengungsindex" basiert auf dem Entropiemodell, die Grundformel nach Shannon ist in der Formel bestimmend. Gerechnet wird mit den gemeinsamen Kantenlängen zwischen den benachbarten Flächentypen, wobei für jede Kombination die Proportion der Kantenlänge der benachbarten Typen zur gesamten Kantenlänge eingeht. Die eingeflossenen Inhalte zur Flächenbeschreibung werden im Hinblick auf ihre Einsprengung im Cover quantifiziert.

Die Berechnung des Durchdringungs- und Einsprengungsindex (IJI) für das gewählte Item eines Polygoncovers gibt zwei Ergebnisdateien zurück. Die externe Ergebnisdatei - Name: %covername%- %itemname%.iji - gibt den Index und alle Nachbarn des jeweiligen Attributs wieder. Die Info-Tabelle gleichen Namens beinhaltet die Attribute des gewählten Items und die ihnen zugehörigen IJI-Werte.

$$IJI = \frac{\sum_{k=1}^{m'} \left[ \left( \frac{e_{ik}}{\sum_{k=1}^{m'} e_{ik}} \right) \ln \left( \frac{e_{ik}}{\sum_{k=1}^{m'} e_{ik}} \right) \right]}{\ln(m' - 1)} \quad (10)$$

Wertebereich: 0 ≤ IJI ≤ 100

e<sub>ik</sub> - Gesamtlänge der Kanten zwischen den Flächentypen (Klassen) i und k [m]

m' - Anzahl der benachbarten Flächentypen (Klassen)

Der IJI wächst mit der Anzahl an verschiedenen Nachbarn und der Länge der Kanten zu diesen.

Hohe Indizes für Flächentypen sind ein Hinweis auf eine ausgeprägte Einsprengung im Cover, dabei ist jedoch zu beachten, dass der Index von zwei Größen abhängt, einmal ist das die Anzahl an verschiedenen Nachbarn und zum anderen die gemeinsame Kantenlänge der benachbarten Typen.

Aufbau der Ergebnisdatei %covername%-itemname%.iji:

Einsprengungs- und Nebeneinanderstellungs-Index

Interspersion and Juxtaposition Index (IJI)

für Cover: %cover% und Item: %item%

-1 - nicht berechnet

Patch-Type %attribut% hat folgende Nachbarn:

att 1 , ... , att n

IJI: %iji%

.....

.....

Struktur der Infotabelle %covername%-itemname%.iji:

ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC
%itemname%		%itemdefinition%		
IJI	4	12	F	4

Die Teilkomponente " Gestaltindizes" bietet zwei Verfahren, um die Gestalt von Polygonen zu quantifizieren. Die Berechnungen basieren auf Formeln die zum Einen die Abweichung der Flächenform von einer einfachen geometrischen Form quantifizieren. Zum Anderen wird die Fraktale Dimension der Flächen ermittelt. Die Indizes ermöglichen damit die Analyse des Formeninventars im Polygoncover.

Die Gestaltindizes können für jede Fläche des Polygoncovers und für das ausgewählte Item als mittlere Gestaltindizes berechnet werden. Die Ergebnisse werden im Info-Tabellensystem abgelegt, Gestaltindizes jeder Fläche in der Tabelle %covername%.sfm, mittlere Gestaltindizes in der Tabelle %covername%-itemname%.shape.

Folgende Größen werden für jede Fläche berechnet:

Gestaltindex

$$SHAPE = \frac{P_{ij}}{2\sqrt{\pi * a_{ij}}} \quad (11) \quad \text{Wertebereich: } SHAPE \geq 1$$

$a_{ij}$  - Größe der Einzelfläche [m<sup>2</sup>];  $p_{ij}$  - Umfang der Einzelfläche [m]

Dieser Index kennzeichnet eine Fläche im Hinblick auf ihre Abweichung von der Kreisform. SHAPE = 1 für Kreisflächen. Mit Zunahme der Flächenkomplexität steigt der Gestaltindex.

Fraktale Dimension

$$FRACT = \frac{2 \ln p_{ij}}{\ln a_{ij}} \quad (12) \quad \text{Wertebereich: } 1 \leq FRACT \leq 2$$

$a_{ij}$  - Größe der Einzelfläche [m<sup>2</sup>];  $p_{ij}$  - Umfang der Einzelfläche [m]

Für einfach gestaltete Flächen nähert sich FRACT dem Wert 1 und für Flächen mit hoher Komplexität dem Wert 2.

Die folgenden vier Indizes (Formeln 13 - 16) sind in der Kennzeichnung der Flächenkomplexität und ihrer Anwendbarkeit, dem SHAPE-Index (Formel 11) gleichzusetzen.

Kompaktheit

$$COMPACTNESS = \frac{2\sqrt{\pi * a_{ij}}}{p_{ij}} \quad (13) \quad \text{Wertebereich: } 0 < COMPACTNESS \leq 1$$

$a_{ij}$  - Größe der Einzelfläche [m<sup>2</sup>];  $p_{ij}$  - Umfang der Einzelfläche [m]

Kreisförmigkeit

$$CIRCULARITY = \frac{4 * a_{ij}}{p_{ij}^2} \quad (14) \quad \text{Wertebereich: } 0 < CIRCULARITY \leq 1/\pi$$

$a_{ij}$  - Größe der Einzelfläche [m<sup>2</sup>];  $p_{ij}$  - Umfang der Einzelfläche [m]

Gestrecktheit

$$ELONGATENESS = \frac{P_{ij}}{\sqrt{a_{ij}}} \quad (15) \quad \text{Wertebereich: } ELONGATENESS \geq 2 * \pi^{0,5}$$

$a_{ij}$  - Größe der Einzelfläche [m<sup>2</sup>];  $p_{ij}$  - Umfang der Einzelfläche [m]

Dünneverhältnis

$$THINNESS\ RATIO = \frac{4 * \pi * a_{ij}}{p_{ij}^2} \quad (16) \quad \text{Wertebereich: } 0 < THINNESS\ RATIO \leq 1$$

$a_{ij}$  - Größe der Einzelfläche [m<sup>2</sup>];  $p_{ij}$  - Umfang der Einzelfläche [m]

Struktur der Infotabelle %covername%.sfm:

ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC
AREA	4	12	F	3
PERIMETER	4	12	F	3
%cover%-ID	4	5	B	-
SHAPE	8	18	F	5
FRACTAL	8	18	F	5
COMPACT	8	18	F	5
CIRCULAR	8	18	F	5
ELONGATE	8	18	F	5
THINRATIO	8	18	F	5

Folgende Größen werden für die Attribute des Items berechnet (mittlere Gestaltindizes):

Flächenanzahl

$$NP = n_i \quad (17)$$

Vergleichsindex Prozent der Gesamtfläche  
(Prozent Flächenanteil)

$$LSIM = \frac{\sum_{j=1}^m a_{ij}}{A} (100) \quad (18)$$

$a_{ij}$  - Größe der Einzelfläche [m<sup>2</sup>]

A - Gesamte Flächengröße in der Karte/Landschaft [m<sup>2</sup>]

Mittlerer Gestaltindex

$$MSI = \frac{\sum_{j=1}^n \left( \frac{p_{ij}}{2\sqrt{\pi * a_{ij}}} \right)}{n_i} \quad (19)$$

Wertebereich: MSI ≥ 1

$a_{ij}$  - Größe der Einzelfläche [m<sup>2</sup>];  $p_{ij}$  - Umfang der Einzelfläche [m];  $n_i$  - Flächenanzahl

Ein Schalter ermöglicht, die Randpolygone nicht einfließen zu lassen. Diese Option ist für Cover, deren Begrenzung durch Blattsschnitte erfolgt, gedacht. Für Inselkarten kann mit Randpolygonen gerechnet werden.

Variationskoeffizient des Mittleren Gestaltindex

$$SCV = \frac{STD * 100}{MSI} \quad (20)$$

STD -Standardabweichung in der Attributklasse, MSI - Mittlerer Gestaltindex

Mittlere Fraktale Dimension

$$MPFD = \frac{\sum_{j=1}^n \left( \frac{2 \ln p_{ij}}{\ln a_{ij}} \right)}{n_i} \quad (21) \quad \text{Wertebereich: } 1 \leq MPFD \leq 2$$

$a_{ij}$  - Größe der Einzelfläche [m<sup>2</sup>];  $p_{ij}$  - Umfang der Einzelfläche [m];  $n_i$  - Flächenanzahl

Struktur der Infotabelle %covername%- %itemname%.shape (Option: mit Randpolygonen):

ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC
FREQUENCY	4	5	B	-
%itemname%	%itemdefinition%			
LSIM	8	18	F	5
SHAPE	8	18	F	6
SHAPE-VARKOE	8	18	F	5
FRACTAL	8	18	F	6

Struktur der Infotabelle %covername%- %itemname%.shape (Option: ohne Randpolygone):

ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC
FREQUENCY	4	5	B	-
%itemname%	%itemdefinition%			
SHAPE	8	18	F	6
SHAPE-VARKOE	8	18	F	5
FRACTAL	8	18	F	6

Der prozentuale Anteil an Fläche (LSIM) kann auch durch die Teilkomponente 'Goldener Schnitt' gewonnen werden.

Die *Doppelt Logarithmische Fraktale Dimension* basiert ebenfalls auf der Beziehung Umfang (P) - Fläche (A), die durch  $A = k * P^{2/D}$  beschrieben werden kann und ist ein weiterer Index, der zur Analyse des Formeninventars der Flächen im Cover dient.

Der Anstieg der Regressionsgeraden  $\ln(P)$  zu  $\ln(A)$  beträgt  $2/D$  und somit wird  $D$  nach folgender Formel (22) berechnet. Dafür ist eine ausreichende Anzahl an Flächen notwendig. Berechnet wird die DLFD für die Attribute des Items mit einer Anzahl von  $n \geq 10$  und in einer Info-Tabelle mit dem Namen %covername%-itemname%.dlfd abgelegt.

$$DLFD = \frac{2 \left[ n_i \sum_{j=1}^n (\ln p_{ij} * \ln a_{ij}) \right] - \left[ \left( \sum_{j=1}^n \ln p_{ij} \right) \left( \sum_{j=1}^n \ln a_{ij} \right) \right]}{\left( n_i \sum_{j=1}^n \ln p_{ij}^2 \right) - \left( n_i \sum_{j=1}^n \ln p_{ij} \right)^2} \quad (22)$$

Wertebereich:  $1 \leq DLFD \leq 2$

$a_{ij}$  - Größe der Einzelfläche [m<sup>2</sup>];  $p_{ij}$  - Umfang der Einzelfläche [m];  $n_i$  - Flächenanzahl

Für einfach geformte Flächen nähert sich DLFD dem Wert 1 und für Flächen mit hoher Komplexität dem Wert 2.

Struktur der Infotabelle %covername%-itemname%.dlfd.

ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC
FREQUENCY	4	5	B	-
%itemname%	%itemdefinition%			
DLFD	8	18	F	5

Die Teilkomponente "Goldener Schnitt" beruht auf der Tatsache, dass das Maximum des Grundterms der Entropie und der Goldene Schnitt nahezu identische Größen liefern. Die der Berechnung zugrunde liegende Größe ist der prozentuale Anteil an Fläche der jeweiligen Flächentypen (LSIM).

Im funktionellen Verlauf des Grundterms der Entropie  $-p * \ln(p)$  (Abb. 36) liegt das Maximum der Funktion bei

$$1/e \approx 0,37 \quad (e = 2,71.. = EulerscheZahl)$$

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 und findet sich in vielen Formen der Natur. Das Maximum der Entropie wird hier als Überraschungswert bezeichnet und kennzeichnet eine Auffälligkeit. Die Auffälligkeit einer Fläche oder eines Flächentyps in der Karte kann durch den Verlauf der Entropie gemessen werden.

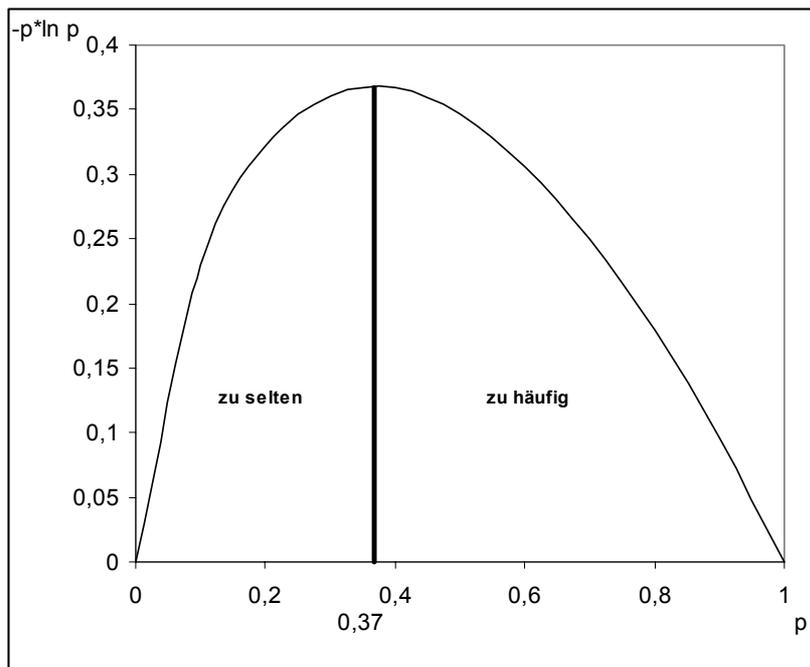


Abb. 36: Funktioneller Verlauf des Grundterms der Entropie

Die Berechnung der Standardisierten Entropie liefert für das ausgewählte Item in einer externen Datei mit dem Namen %covername%-%itemname%.gcut die Attribute, die nahe dem Maximum der Funktion mit einer Abweichung von 1% liegen. Die Info-Tabelle gleichen Namens beinhaltet alle Attribute des gewählten Items, und besitzt folgende Struktur:

ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC	
FREQUENCY	4	5	B	-	Häufigkeit des Attributs
%itemname%	%itemdefinition% - gewähltes Item				
PI	8	9	F	6	Proportion von Flächenanzahl des Flächentyps zur Gesamtanzahl
LNPI	8	9	F	6	Logarithmierte Proportion von Flächenanzahl des Flächentyps zur Gesamtanzahl
PILNPI	8	9	F	6	Shannon Diversitäts-Index
LSIM	8	9	F	6	Vergleichsindex Prozent der Gesamtfläche
SLSIM	8	9	F	6	Standardisierter Vergleichsindex Prozent der Gesamtfläche
STDE	8	9	F	6	Standardisierte Entropie
SFILNPI	8	9	F	6	Standardisierter Shannon Diversitäts-Index

AML-Notation: <> Zeichenketten in spitzen Klammern bezeichnen Pflichtparameter  
 %% In Prozentzeichen eingefasste Zeichenketten bezeichnen Variablenamen  
 {} In geschweifte Klammern eingefasste Zeichenketten bezeichnen optionale Parameter bzw. Platzhalter

Im Verlauf der Berechnung kann die Funktion mit Kennzeichnung der Attribute nahe dem Maximum und dem Verlauf des Shannon Diversitäts-Index als Postscript-Datei %covername%-%itemname%.eps oder frei wählbaren Namens abgelegt werden.

Die *Join-Count Teilkomponente* bewertet die Matrix aller nachbarschaftlichen Kombinationen der Attribute des gewählten Items und gibt die Ergebnisse in einer externen Datei %covername%-%itemname%.jcs aus. Damit ist eine Aussage bezüglich der Heterogenität möglich, nämlich, handelt es sich um strukturierte oder unstrukturierte Heterogenität. Zur Betrachtung der räumlichen Korrelation sollten nur grobe Abstufungen in der inhaltlichen Kennzeichnung der Flächen herangezogen werden - wenige Attribute -, denn zu viele Abstufungen führen zu einer Verwischung der Muster.

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.

Für die gefundene Anzahl räumlicher Nachbarschaft jeder Kombination wird der Erwartungswert berechnet. Die Formeln lauten:

$$E(rein) = \frac{m * n_1 * (n_1 - 1)}{n * (n - 1)} \quad (23)$$

$$E(hybrid) = \frac{2 * m * n_1 * n_2}{n * (n - 1)} \quad (24)$$

$E(rein)$ ,  $E(hybrid)$  - Erwartungswerte

$m$  - Gesamtzahl der Kombinationen

$n$  - Gesamtzahl der Flächen

$n_1$ ,  $n_2$  - Anzahl der Flächen mit jeweiligem Attribut

Die Beurteilung, ob eine signifikante Abweichung der gefundenen Anzahl räumlicher Nachbarschaft der Kombination von ihrem Erwartungswert vorliegt, beruht auf dem Vergleich von  $\chi^2$  mit einer festgelegten Schranke. Überschreitet  $\chi^2$  die vereinbarte Schranke kann eine signifikante Abweichung vermutet werden. Ist  $\chi_{ijk}$  positiv handelt es sich um einen Typ, im Fall einer negativen Abweichung um einen Antityp.

Formel zur Berechnung von  $\chi^2$

$$\chi^2 = \sum_{i=1}^k \frac{(n_{ijk} - e_{ijk})^2}{e_{ijk}} \quad (25)$$

Formel zur Berechnung von  $\chi_{ijk}$

$$\chi_{ijk} = \frac{(n_{ijk} - e_{ijk})}{\sqrt{e_{ijk}}} \quad (26)$$

$n_{ijk}$  - Zellenhäufigkeit;  $e_{ijk}$  - erwartete Zellenhäufigkeit (Erwartungswert),  $k$  - Objektanzahl

Für den  $\chi^2$ -Komponententest wird folgende Schranke genutzt: Freiheitsgrad (f) 1;  $\alpha = 0,1\%$ ;  $\chi^2 = 10,8$   
Diese Abweichung wird verbal gekennzeichnet (Typ, Antityp).

Aufbau der Ergebnisdatei %covername%-%itemname%.jcs:

Join-Count Statistik

für Cover: %cover% und Item: %item%

%m% Prozent der Patches (n= %n%) sind hybrid (Grenz-Länge= %length%)

%m% Prozent der Patches (n= %n%) sind rein (Grenz-Länge= %length%)

von der indizierten Partition der Matrix (n= %n%)

Reine Kombinationen:

Record L%item% R%item% %FREQUENCY% %Erwartungswert% %Abweichung% % $\chi^2$ % % $\chi$ %

.....

Hybride Kombinationen:

Record L%item% R%item% %FREQUENCY% %Erwartungswert% %Abweichung% % $\chi^2$ % % $\chi$ %

.....

Chi-Quadrat: % $\chi^2$ %

Freiheitsgrad: %f%

*Anmerkung:* Attribute die einen Wert von 0 (Zahlenkodierung) besitzen oder leer sind (Zeichenkodierung) werden nicht in die Berechnung einbezogen.

### *Erstellen von Streudiagrammen (Scatter-Plots mit linearer Regression)*

Mit dieser Komponente können die gewonnenen landschaftsanalytischen und informationstheoretischen Indizes grafisch auf korrelativen Zusammenhang geprüft werden.

In der Oberfläche (Abb. 37) werden die Info-Tabellen und die ihnen zugehörigen Items, die in Zahlenformaten kodiert sind, angeboten. Sollen Werte aus zwei verschiedenen Info-Tabellen dargestellt werden, muss das Relate-Item in beiden vorhanden sein. Dieses Item kann als Text im Scatter-Plot dargestellt (Schalter Relate-Item anzeigen). Die aktuelle Auswahl der Parameter wird in der Meldungszeile angezeigt. Nach erfolgter Auswahl kann über den Button 'execute' der Scatter-Plot erzeugt werden. In der Darstellung werden die beiden Regressionsgeraden eingezeichnet. Die beiden Gleichungen für x und y, der Korrelationskoeffizient r und das Bestimmtheitsmaß  $r^2$  werden im Plot und im Arc-Fenster angezeigt. Der Scatter-Plot (Abb. 38) kann als Postscript-Datei abgelegt werden.

Es handelt sich um eine lineare Regression. Für die Regressionsgeraden gelten - wie für jede Gerade die Gleichungen:

$$y = a_1 + b_1 * x \quad (27) \quad \text{und} \quad x = a_2 + b_2 * y \quad (28)$$

Für die Regressionskoeffizienten b gilt:

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) * (y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad (29) \quad \text{und} \quad b_2 = \frac{\sum_{i=1}^n (x_i - \bar{x}) * (y_i - \bar{y})}{\sum (y_i - \bar{y})^2} \quad (30)$$

Für die Schnittpunkte a gilt:

$$a_1 = \bar{y} - b_1 * \bar{x} \quad (31) \quad \text{und} \quad a_2 = \bar{x} - b_2 * \bar{y} \quad (32)$$

Für den Korrelationskoeffizienten gilt:

$$r = \sqrt{b_1 * b_2} \quad (33)$$

Für das Bestimmtheitsmaß gilt:

$$r^2 = b_1 * b_2 \quad (34)$$

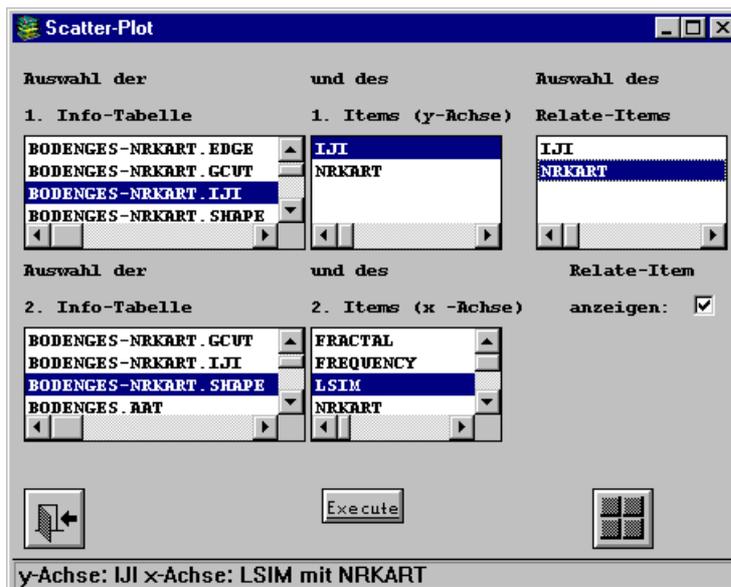


Abb. 37: Bildschirmabzug des Menüs zum Erstellen von Scatter-Plots

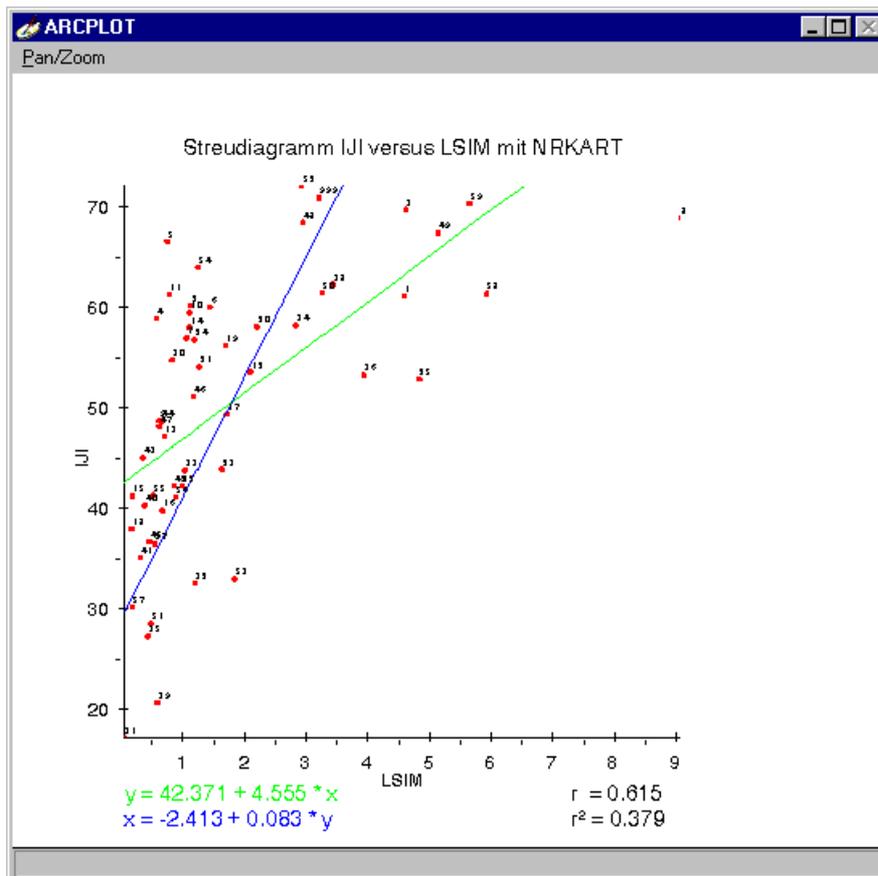


Abb. 38: Bildschirmabzug der Ergebnisdarstellung "Scatter-Plots"

#### 10.8.4 Komponente "Index maximaler Reduktion"

Eingangsgrößen: Polygoncover, Item, Zielmaßstab, MLD-Konstante

Ergebnis: externe Textdatei mit den berechneten Größen

Die hier berechneten Größen sind in zweierlei Hinsicht nützlich. Sie geben einmal 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. Zweitens unterstützen die Indizes im Prozess der Überführung einer Karte in einen kleineren Folgemaßstab. Die Komponente 'Index maximaler Reduktion' (Abb. 39) startet in der aktuellen Workspace-Umgebung und bietet zur Auswahl die Polygoncover und die zugehörigen Items der PAT an. Die gewählten Parameter werden in der Meldungszeile des Menüs angezeigt. Die Berechnung ist ausschließlich in Aml realisiert. Weitere Parameter die zur Berechnung notwendig sind, sind die Konstanten der minimalen lesbaren Zeichengröße und der angestrebte Zielmaßstab.

- Minimale lesbare Zeichengröße (MLD) = 40 mm<sup>2</sup>, Cornell University group on Adequacy of Soil Resource Inventories
- Minimale lesbare Zeichengröße (MLD)= 25 mm<sup>2</sup>, Vink, soil survey programs in some countries, e.g. Netherlands

Die Minimale lesbare Flächengröße wird nach folgender Formel berechnet:

$$MLA = \frac{\left(\frac{SN}{1000}\right)^2}{MLD} \quad (35)$$

Der Index maximaler Reduktion gibt Auskunft darüber, ob die Flächen in einer Karte den kartografischen Anforderungen entsprechen. Die Formel lautet:

$$IMR = \sqrt{\frac{ASA}{MLA}} \quad \text{Wertebereich: } IMR > 0 \quad (36)$$

MLD <sub>c</sub> = 250	<i>Cornell-Konstante</i>
MLD <sub>v</sub> = 400	<i>Vink-Konstante</i>
MLA =	<i>Minimale lesbare Flächengröße [m<sup>2</sup>], [ha]</i>
MLD =	<i>Minimale lesbare Zeichengröße</i>
SN =	<i>Maßstabszahl des Zielmaßstabs</i>
IMR =	<i>Index der Maximalen Reduktion</i>
ASA =	<i>Durchschnittliche Flächengröße (für Klassen und für gesamte Karte) [m<sup>2</sup>]</i>

Eine Definition des optimalen Wertes von **2** für den **IMR** beruht darauf, dass in diesem Fall die durchschnittliche Flächengröße der optimal lesbaren Flächengröße entspricht.

IMR > 2 - Der Maßstab (SN) kann reduziert werden

IMR < 2 - Die Karte enthält zu viele Flächen, der Maßstab sollte erhöht werden.

IMR < 1 - Die Karte wird unleserlich.

Die externe Ergebnisdatei `%covername%-%itemname%%scalenummer%.imr` (Abb. 40) gibt die Attribute des gewählten Items wieder, die in ihrer mittleren Flächengröße die MLA unterschreiten. Die Anzahl aller Einzelflächen mit Unterschreitung der MLA wird ausgegeben. Der IMR für die gesamte Karte, bezogen auf die gewählte Konstante und den gewünschten Zielmaßstab, folgt. Zur Distanzmessung zwischen ausgewählten Attributen eines Items wird ein Vorschlag gegeben, der auf der Formel für das eingeschriebene regelmäßige Vieleck im Kreis basiert:

$$h_n = r * \cos \frac{180^\circ}{n} \quad (37)$$

h - Bogenhöhe; n - Anzahl der Ecken; r - Radius

Ausgehend von der MLA wird der Radius des zugehörigen Kreises bestimmt. Das einfachste eingeschriebene regelmäßige n-Eck ist das gleichseitige Dreieck für das gilt:  $h = \frac{r}{2}$ . Diese Distanz entspricht auf der Karte einer Länge von 1,41 mm.

Aufbau der Ergebnisdatei %covername%- %itemname%%scalenummer%.imr:

Mittelwert(e) von m BL\_NR <= %mla% m² (MLA)

Record BL\_NR FREQUENCY MEANAREA IMR

.....

%n% Einzelflächen <= %mla% m² (MLA)

-----

Ergebnis für Cover %cover% und Item %itemname%

Test für Zielmaßstab 1 : %scale% mit %const% Definition

Minimale darstellbare Größe (MLA) : %mla% m² | %mla% ha

Index der Maximalen Reduktion (IMR) : %imr%

Vorschlag zur Distanzmessung : %distanz% m

Vorschlag zur Distanzmessung (Vink) : %vdistanz% m

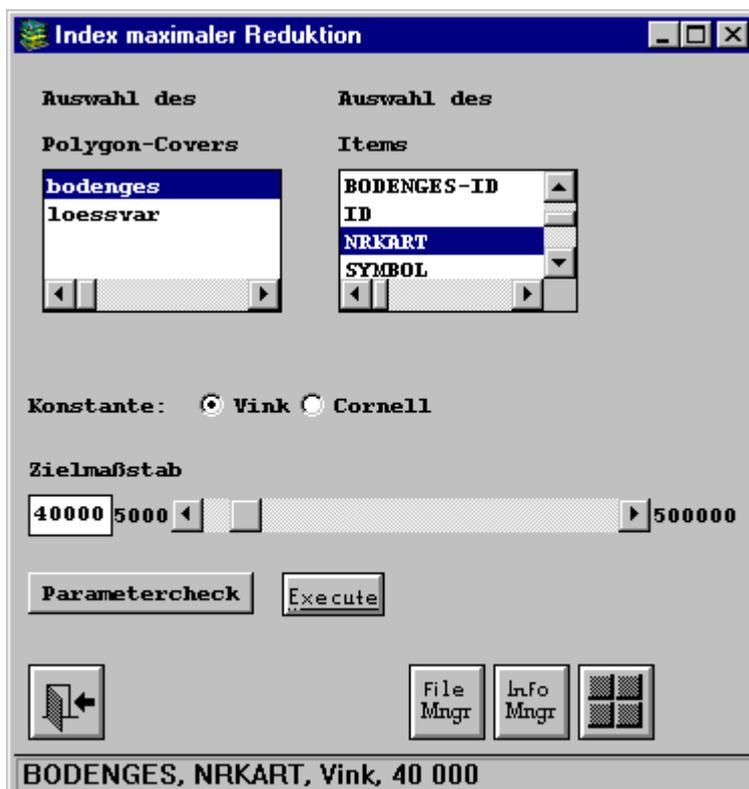


Abb. 39: Bildschirmabzug der Oberfläche der Komponente "Index maximaler Reduktion"

AML-Notation: < > Zeichenketten in spitzen Klammern bezeichnen Pflichtparameter  
 % % In Prozentzeichen eingefasste Zeichenketten bezeichnen Variablenamen  
 {} In geschweifte Klammern eingefasste Zeichenketten bezeichnen optionale Parameter bzw. Platzhalter

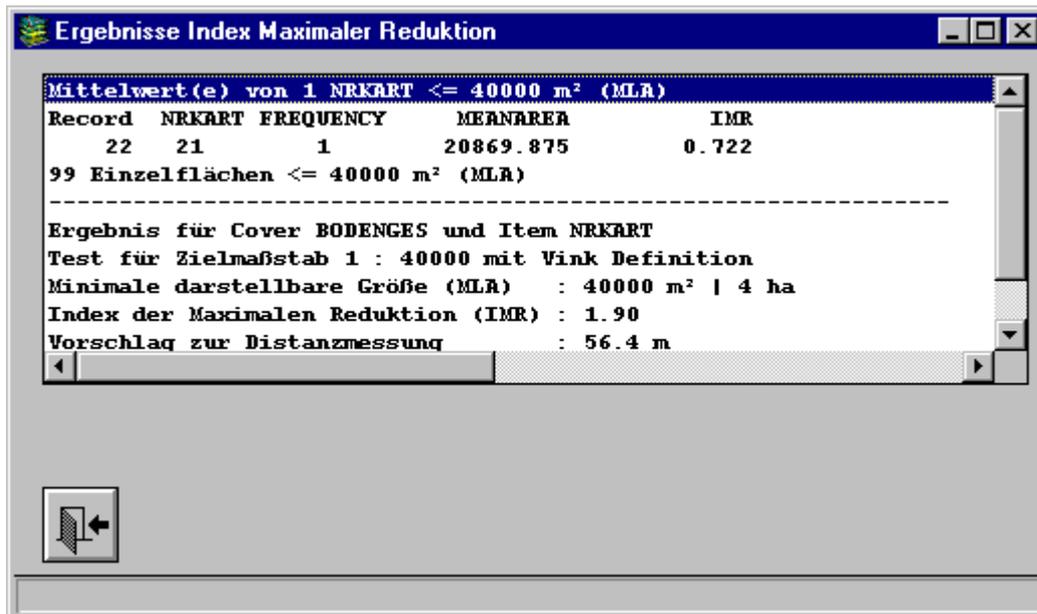


Abb. 40: Bildschirmabzug der Ergebnisdarstellung der Komponente 'Index maximaler Reduktion'

#### 10.8.5 Komponente "Nachbarschaftskennzeichnung für ein Attribut"

Eingangsgrößen: Polygoncover, Item, Attribut

Ergebnis: externe Textdatei mit den berechneten gemeinsamen Grenzlängen

Das Menü der Komponente 'Nachbarschaftskennzeichnung für ein Attribut' (Abb. 41) bietet zur Auswahl die Polygoncover und die zugehörigen Items der Polygon-Attribut-Tabellen und deren Attribute an. Die gewählten Parameter werden in der Meldungszeile des Menüs angezeigt. Die externe Ergebnisdatei %covername%-%itemname%\_%attribut%.edge gibt für das gewählte Attribut alle Nachbarn anderer Attribute und die zugehörigen Grenzlängen [m] wieder. Im Verlauf der Berechnung erfolgt eine Darstellung im Grafik-Fenster. Der button 'Koordinaten-Verortung' im Ausgabefenster für die Ergebnisdatei, ermöglicht die Verortung von Flächen. Dabei werden die Koordinaten des gewählten Punktes auf dem Grafikbildschirm im ARC-Fenster ausgegeben. Über den button 'Polygon-Information' im Ausgabefenster sind die Informationen über die einzelnen Polygone abrufbar.

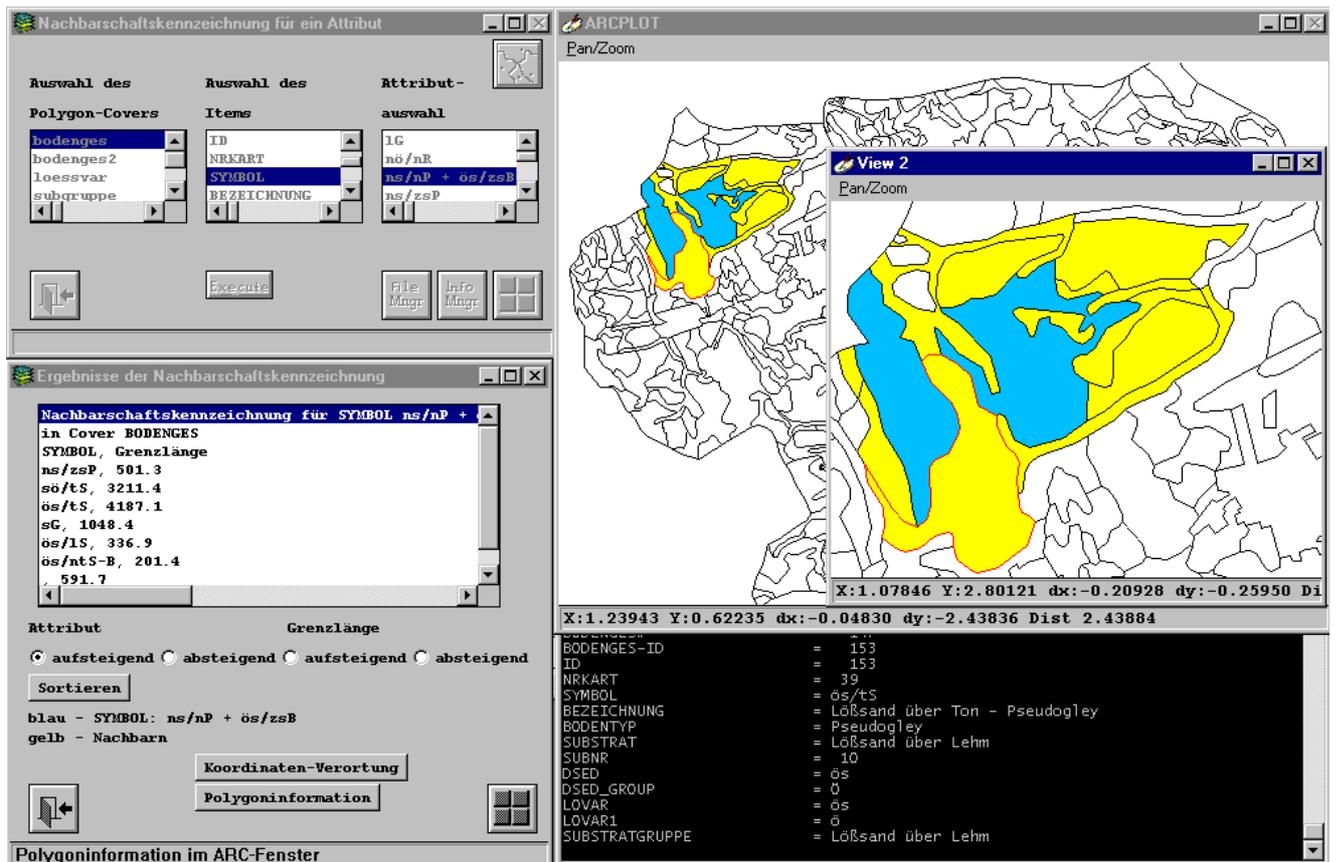


Abb. 41: Bildschirmabzug der Oberfläche zur Ermittlung der Nachbarschaft für ein Attribut

Aufbau der Ergebnisdatei %covername%-%itemname%\_%attribut%.edge:

```
Nachbarschaftskennzeichnung für %itemname% %attribut%
in Cover %cover%
%itemname%, Grenzlänge
%attribut 1% %length%
.....
%attribut n% %length%
```

**Anmerkungen:**

- Existieren Nachbarschaften zwischen gleichen Attributen, werden diese mit ausgegeben. Auf dem Grafikbildschirm werden dann bereits eingefärbte Flächen mit der Farbe zur Kennzeichnung von Nachbarn überzeichnet.
- Grenzen Polygone des Attributs an das Außenpolygon, wird die gemeinsame Grenzlänge dafür ausgegeben.
- Existiert die Ergebnisdatei bereits, erfolgt keine Darstellung auf dem Grafikbildschirm.
- Im Namen der Ergebnisdatei werden auftretende Sonderzeichen (/ \ ( ) + ! ; ^) im Attribut durch Bindestriche ersetzt.

AML-Notation: <> Zeichenketten in spitzen Klammern bezeichnen Pflichtparameter  
 %% In Prozentzeichen eingefasste Zeichenketten bezeichnen Variablenamen  
 {} In geschweifte Klammern eingefasste Zeichenketten bezeichnen optionale Parameter bzw. Platzhalter

### 10.8.6 Komponente "Nachbarschaftskennzeichnung aller Inhalte"

Eingangsgrößen: Polygoncover, Item

Ergebnis: Info-Datei mit den berechneten Grenzlängen, der Häufigkeit, dem Längenindex und dem Längenindex der Kombination

Das Menü der Komponente bietet zu Auswahl die Polygoncover mit den zugehörigen Items der Polygon-Attribut-Tabellen an. Die gewählten Parameter werden in der Meldungszeile des Menüs angezeigt. Die Berechnung hat eine Infotabelle zur Folge, die alle Kombinationen der Attribute mit ihren zugehörigen Grenzlinien-Längen [m], der Anzahl der Nachbarschaft und einem Längenindex beinhaltet. Die Attribute sind aufsteigend sortiert. Die Anzahl n der Kombinationen entspricht maximal Größe der halben Matrix, nach der Formel:  $n_{\text{Kombinationen}} = n_{\text{Attribute}} * (n_{\text{Attribute}} - 1) / 2$

Die Längenindizes LI und LIK berechnen sich nach folgender Formel:

$$LI, LIK = \frac{length}{\left( \frac{\sum_{i=1}^n (length_i * m_i)}{\sum_{i=1}^n m_i} \right)}$$

Der Klammerausdruck im Nenner entspricht der Formel zur Berechnung des gewogenen arithmetischen Mittels.

(38)

Wertebereich: LI , LIK  $\geq 0$

LI - Längenindex

LIK - Längenindex-Kombination

length, length<sub>i</sub> - gemeinsame Grenzlänge der Kombination

m<sub>i</sub> - Anzahl der Nachbarschaft

n - Anzahl der Kombinationen

LI > 1 ist Ausdruck für dominante Nachbarschaft im Hinblick auf alle Kombinationen. Zur Wichtung fließen alle Grenzlängen und Häufigkeiten ein.

Die Berechnung von LIK erfolgt mit der selektierten Teilmenge. Zur Wichtung fließen nur die Grenzlängen dieser Teilmenge und deren Häufigkeiten ein.

Aufbau der Infotabelle %covername%-%itemname%.edge%

%Attribut 1%	%Attribut 2%	%Länge%	%Anzahl%	%LI%	%LIK%
%Attribut 1%	%Attribut 3%	%Länge%	%Anzahl%	%LI%	%LIK%
.	.	%Länge%	%Anzahl%	%LI%	%LIK%
%Attribut 1%	%Attribut n%	%Länge%	%Anzahl%	%LI%	%LIK%
.	.	%Länge%	%Anzahl%	%LI%	%LIK%
.	.	%Länge%	%Anzahl%	%LI%	%LIK%
%Attribut n-1%	%Attribut n%	%Länge%	%Anzahl%	%LI%	%LIK%

Struktur der Infotabelle %covername%-%itemname%.edge:

ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC	
L%itemname%		%itemdefinition%			
R%itemname%		%itemdefinition%			
LENGTH	8	18	F	6	
ANZAHL	8	18	F	6	
LI	8	18	F	6	Längenindex
LIK	8	18	F	6	Längenindex-Kombination

*Anmerkung:* Räumliche Nachbarschaften zwischen gleichen Attributen führen zu einem falschen Längenindex (LI), da die Anzahl (m) dabei größer, als die der wirklichen Nachbarschaften ist. In diesem Fall ist eine Verschmelzung der Flächen gleicher Attribute mit dem Dissolve-Befehl vor der Ermittlung der Grenzlängenindizes notwendig.

### 10.8.7 Komponente "Distanzmessung zwischen Polygonen verschiedener Inhalte"

Eingangsgrößen: Polygoncover, Item, Attribute [ $n_{\max} = 3$ ], Distanz h

Ergebnis: externe Textdatei mit den Polygon Identifikations-Nummern %cover%-ID aus der selektierten Teilmenge, die der Bedingung Polygonzwischenabstand  $\leq h$  genügen

Aus der aktuellen Workspace-Umgebung wird die Auswahl der Polygoncover, ihrer zugehörigen Items der Polygon-Attribut-Tabelle und deren Attribute angeboten (Abb. 42). Cover- und Itemnamen werden in der Meldungszeile des Menüs angezeigt. Die selektierten Attribute müssen über den button 'Att bestätigen' bestätigt werden. Dabei sind maximal drei Attribute zulässig. Die Distanz h, der gesuchte maximale Abstand zwischen den Polygonen der gewählten Attribute, wird eingegeben. Hierfür kann der Vorschlag zur Distanzmessung aus der Komponente 'Index maximaler Reduktion' herangezogen werden. Die Berechnung erfolgt über die aus ArcInfo im Polygon-Ungenerate Format erzeugte externe Datei. Zwar besitzt ArcInfo eine Vielzahl Kommandos zur Distanzmessung, diese bieten jedoch nicht den Zugriff auf jeden Punkt eines Linienzuges (Vertex). Die Selektion wird durch ein externes Programm vorgenommen. Von den Polygonen die der Bedingung, Polygonzwischenabstand  $\leq h$ , entsprechen, werden die ID's %cover%-ID in eine externe Datei geschrieben. Der Dateiname setzt sich aus %covername%%itemname%\_%attribute% .%distanz% zusammen. Die erste Zeile in der Datei enthält den Cover- und Itemnamen, gefolgt von der Anzahl an gewählten Attributen, den Attributen und der genutzten Distanz. Die Cover-ID's werden paarweise geschrieben. Über den button 'Ergebnisse anzeigen' (Abb. 43) können alle Ergebnisse der Abstandsmessung auf dem Grafikbildschirm dargestellt werden. Die Ausgabe auf dem Grafikbildschirm zeigt alle Polygone der gewählten Attribute. Hervorgehoben sind die Polygone, deren räumliche Nachbarschaft durch die vorgegebene Distanz

bestimmt ist. Koordinaten können verortet, Inhalte von Polygonen bestimmt und Längen gemessen werden.

Die Namens-Konvention für die Ergebnisdateien sollte beibehalten werden, da die Auswahl zur Darstellung als auch die Auswahl für die Komponenten 'Verbindungsflächenermittlung' und 'Verschmelzung von Flächen' auf \*.\*.0 Dateien basiert.

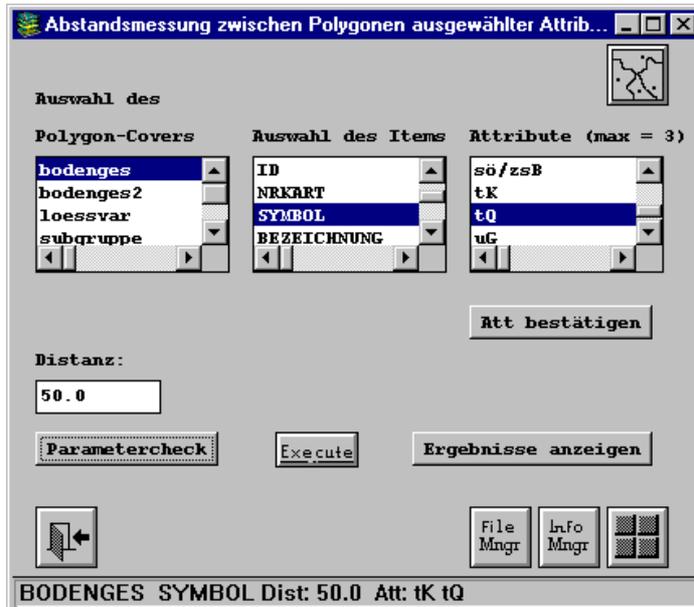


Abb. 42: Bildschirmabzug der Oberfläche zur Distanzmessung.

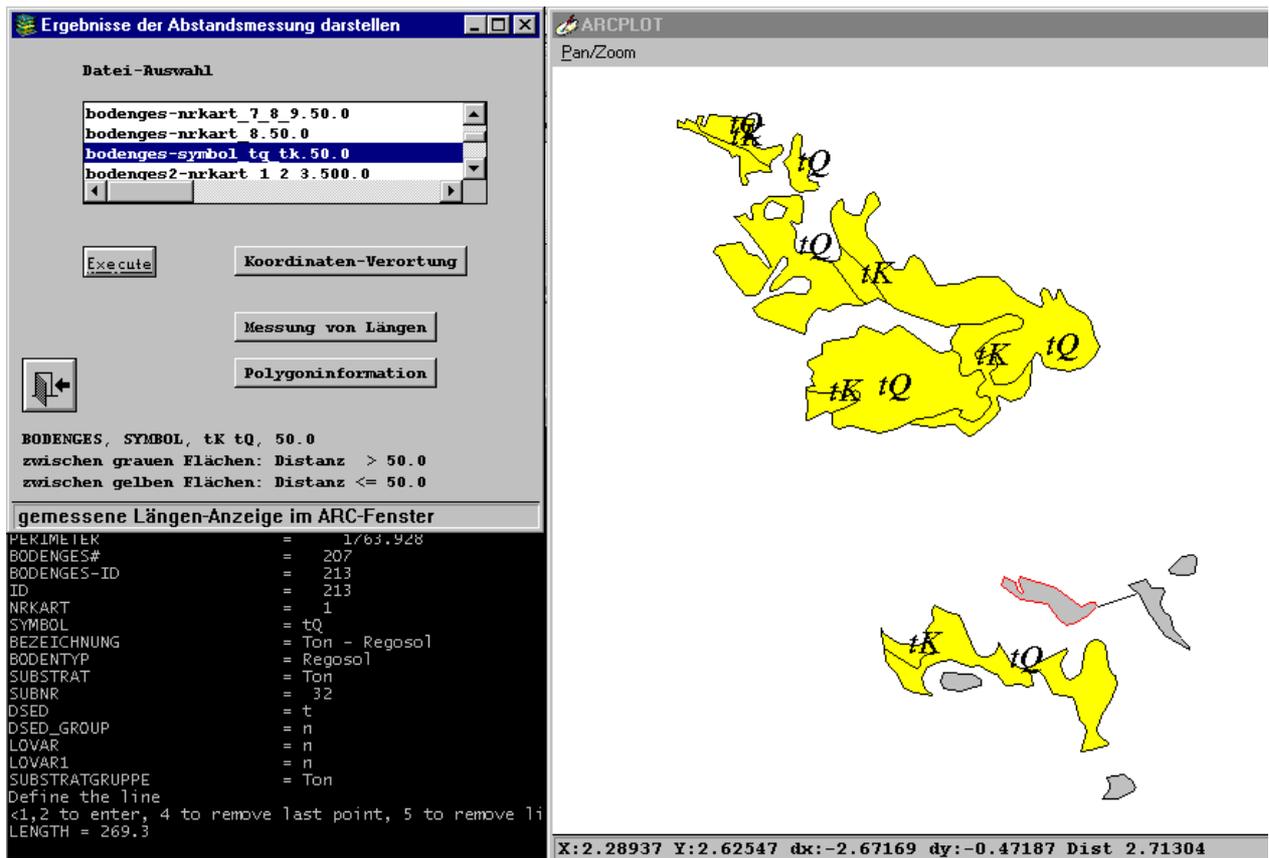


Abb. 43: Ein Beispiel für die grafische Darstellung der Ergebnisse der Distanzmessung. Alle Polygone der gewählten Attribute werden gezeichnet. Die Polygone, die der Bedingung Polygonzwischenabstand  $\leq h$  genügen, werden hervorgehoben und mit den Attributen bezeichnet.

Aufbau der Ergebnisdatei %covername%- %itemname%\_ %attribute%. %distanz%:

```
%cover% %itemname% %attributanzahl% %attrinut 1% .{%attribut 2%. %attribut 3%} %distanz%
%cover%-id %cover%-id
...
%cover%-id %cover%-id
```

**Anmerkungen:**

- Die Berechnungszeit hängt von der Anzahl der durch die Auswahl erzeugten Stützpunkte ab. Große Distanzen und geglättete Umrisslinien der Flächen führen zu langen Rechenzeiten.
- Im Namen der Ergebnisdatei werden auftretende Sonderzeichen (/ \ ( ) + ! ; ^) in den Attributen durch Bindestriche ersetzt. Attribute dürfen keine Leerzeichen enthalten.

**10.8.8 Komponente "Verbindungsflächenermittlung"**

Eingangsgrößen: externe Ergebnisdatei aus der Distanzmessung

AML-Notation: <> Zeichenketten in spitzen Klammern bezeichnen Pflichtparameter  
 %% In Prozentzeichen eingefasste Zeichenketten bezeichnen Variablenamen  
 {} In geschweifte Klammern eingefasste Zeichenketten bezeichnen optionale Parameter bzw. Platzhalter

Ergebnis: externe Textdatei mit Flächenanteilskennzeichnung und optional Verbindungsflächencover

Aufbauend auf der Abstandsmessung zwischen Polygonen ausgewählter Inhalte, können mit der Komponente die trennenden Flächen zwischen diesen Polygonen ausgewertet werden. Die Auswahl greift auf die in der Distanzmessung erzeugten Ergebnisdateien zu (Abb.45). Nach Angabe eines neuen Namens für das Verbindungsflächencover kann die Ermittlung mit oder ohne das Aufheben dieses Covers durchgeführt werden. In diesem aufwendigen Prozess mit Ermittlung der Linien durch ein externes Programm (s. Komponente Distanzmessung), Linien- und Polygonbildung, Verschneidung und Identifizierung kann die Linienbildung zum Abbruch führen (Software Limitation, zu viele Punkte bzw. Linien). Die Meldung 'keine Ergebnisse produziert' erscheint in diesem Fall. Eine Möglichkeit dieses Problem zu umgehen, wird durch den Schalter 'Fuzzy-Toleranz erhöhen' angeboten. Durch das Aktivieren dieses Schalters wird für die Linienbildung im temporären Cover mit einer Fuzzy-Toleranz gearbeitet, die sich aus der Fuzzy-Toleranz des Eingangscovers mal 10 ergibt.

Die Verbindungsflächen werden nur zwischen Polygonen im angegebenen Mindestabstand gebildet, die filigranen Formen von einzelnen Polygonen, bei der Kanten in den Bereich des Mindestabstands gelangen, werden nicht betrachtet.

Nach erfolgreicher Ermittlung wird die externe Ergebnisdatei %verbindungsflächencover%.vfe angezeigt. Folgende Informationen stehen zur Verfügung:

1. Prozent der Verbindungsfläche aus %cover%-ID mit Item %itemname%
2. Prozent der Verbindungsfläche aus Item %itemname%
3. Prozent der Flächenanteile in der Zusammenfassung REGION-ID

Die Polygone werden über die REGION-ID des Verbindungsflächencovers zusammengefasst (Abb. 44). Eine Region - Das Regionen Konzept von ArcInfo wird nicht genutzt. - beinhaltet die im angesetzten Mindestabstand zueinander liegenden Polygone der gewählten Attribute und die dazwischen befindlichen Verbindungsflächen anderer Inhalte. Die Abbildung zeigt in einem Ausschnitt die Flächen der gewählten Attribute, gelb hinterlegt sind die Polygone, die über die roten Verbindungsflächen miteinander verbunden sind.

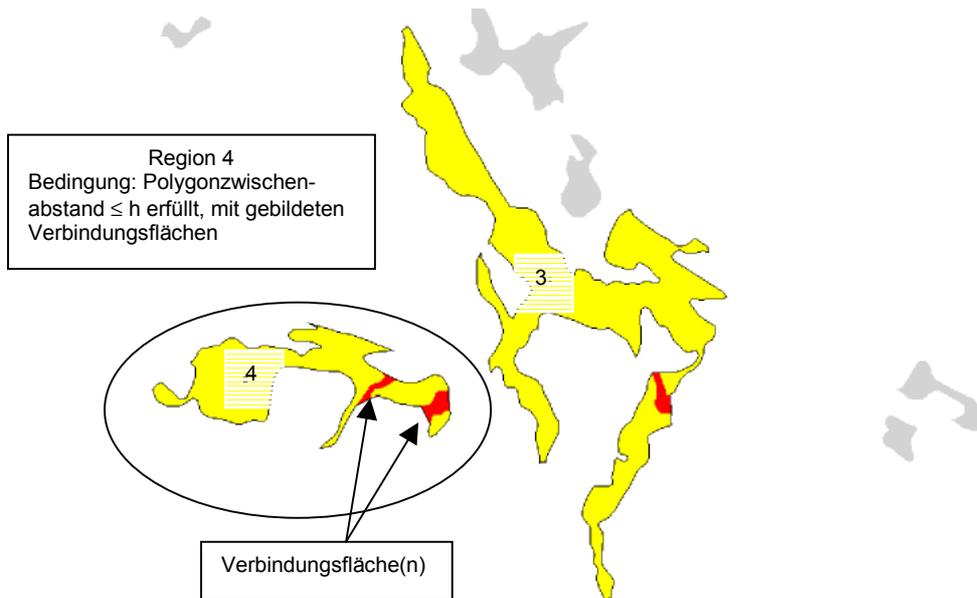


Abb. 44: Erläuterung der grafischen Darstellung der Ergebnisse der Verbindungsflächenermittlung

Ist die Option 'und Verbindungsflächencover' gewählt, wird zusätzlich zur externen Ergebnisdatei das Verbindungsflächencover abgelegt. Die Polygon-Attribut Tabelle des Verbindungsflächencovers führt zur REGION-ID auch das Item aus der Distanzmessung.

Aufbau der Ergebnisdatei %verbindungsflächencover%.vfe:

Verbindungsflächenermittlung für Cover %cover%

Item %itemname% Attribute %attribut 1% {%attribut 2% %attribut3%} Distanz <= %distanz%

Prozent der Verbindungsfläche aus %cover%-ID mit Item %itemname%

Record %cover%-ID %attribut% %PROZENT%

Prozent der Verbindungsfläche aus Item %itemname%

Record %attribut% %PROZENT%

Prozent der Flächenanteile in der Zusammenfassung REGION-ID

Record REGION-ID %attribut% %PROZENT%

Struktur der Polygon-Attribut Tabelle %Verbindungsflächencover%.pat:

ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC
AREA	8	18	F	5
PERIMETER	8	18	F	5
%verbindungsflächencover%#	4	5	B	-
%verbindungsflächencover%-ID	4	5	B	-
%cover%-ID	4	5	B	-
%itemname%			%itemdefinition%	
REGION-ID	4	5	B	-
INSIDE	4	5	B	-

**Anmerkungen:**

- Ist trotz erhöhter Fuzzy-Toleranz eine Verbindungsflächenermittlung nicht möglich, kann auf der ArcEdit-Ebene die Dichte der Stützpunkte verringert werden. Das geschieht durch die Kombination des Grain-Faktors mit dem Spline-Befehl.
- Das Ergebniscover enthält nur die Flächen der ausgewählten Attribute und die Zwischenflächen anderer Attribute, die mit der Bedingung Distanz  $h \leq$  gewählter Distanz gebildet wurden.

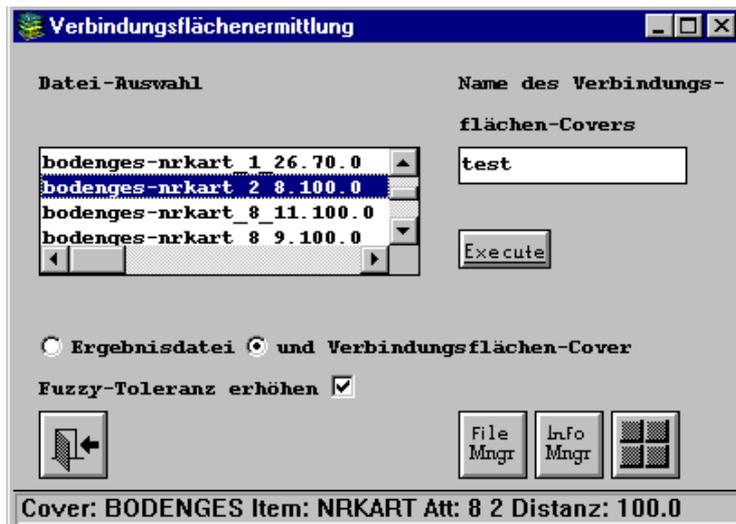


Abb. 45: Bildschirmabzug der Oberfläche zur Verbindungsflächenermittlung

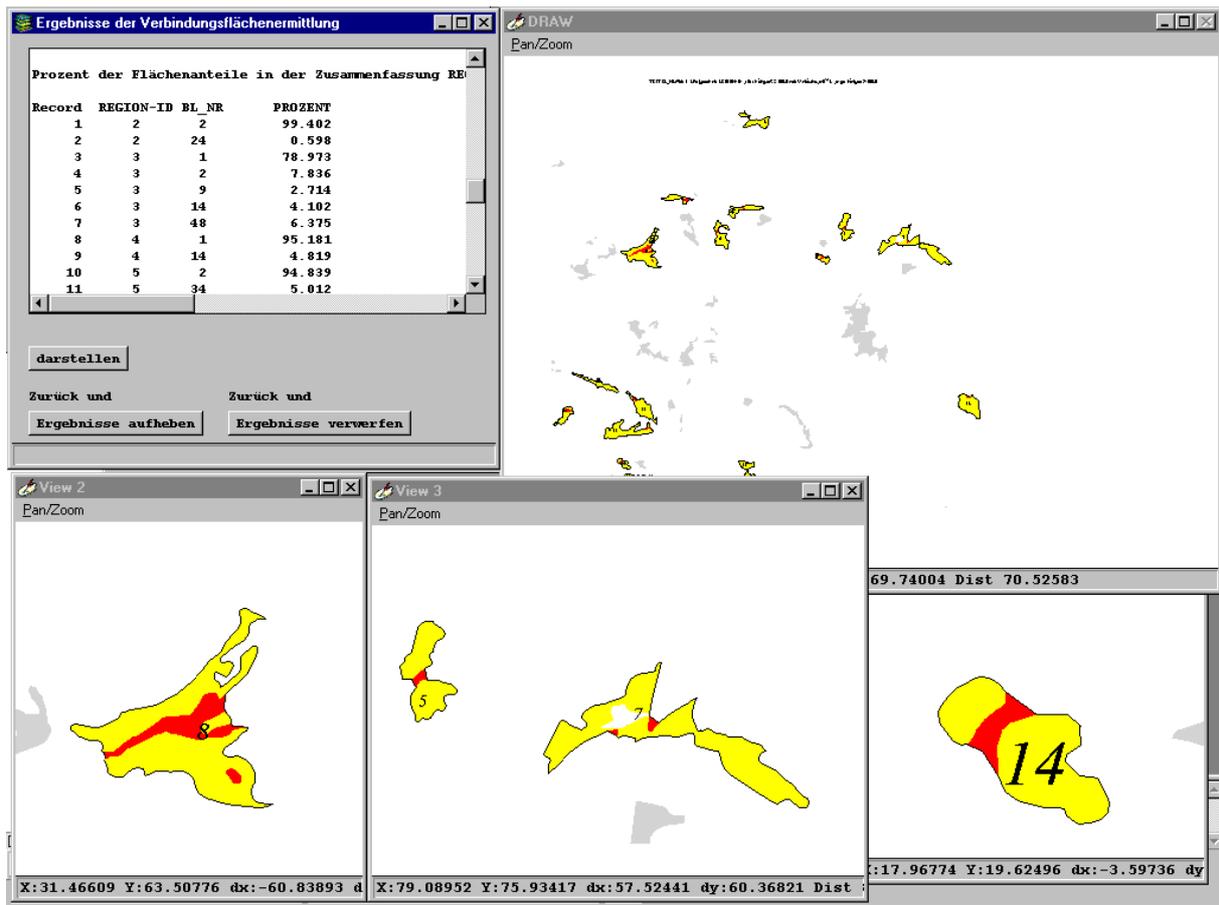


Abb. 46: Bildschirmabzug der Oberfläche zur Darstellung der Ergebnisse der Verbindungsflächenermittlung

### 10.8.9 Komponente "Verschmelzung von Flächen"

Eingangsgrößen: externe Ergebnisdatei aus der Distanzmessung

Ergebnis: externe Textdatei mit Flächenanteilskennzeichnung und Flächencover

Aufbauend auf der Abstandsmessung zwischen Polygonen ausgewählter Inhalte, können mit dieser Komponente Flächen verschmolzen werden (Abb. 48). Die Auswahl greift auf die in der Distanzmessung erzeugten Ergebnisdateien zu (Abb. 47). Ein Name für das Verschmelzungscover muss vergeben werden.

Die Routine nutzt den AreaAggregate-Befehl der ArcInfo Version 8.1. Die Flächen werden über Grid-Funktionen verschmolzen. Für die hierfür erforderliche Rasterzellengröße (cellsize) wird ein Vorschlag gegeben: cellsize = genutzte Distanz aus Eingabedatei / 10. Dieser Wert kann im Menü geändert werden.

Aufbau der externen Ergebnisdatei % Verschmelzungscover%.fag:

Flächen-Verschmelzung aus Cover %cover%

Item %itemname% Attribute %attribut 1% {%attribut 2% %attribut3%} Distanz <= %distanz% Cellsize %cellsize%

Prozent der Flächenanteile in der Zusammenfassung %Verschmelzungscover%#

in Cover %Verschmelzungscover%

Record %Verschmelzungscover%# %attribut% %FREQUENCY% %PROZENT%

Struktur der Polygon-Attribut Tabelle %Verschmelzungscover%.pat:

ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC
AREA	8	18	F	5
PERIMETER	8	18	F	5
%Verschmelzungscover%#	4	5	B	-
%Verschmelzungscover%-ID	4	5	B	-

Anmerkungen:

- Die Verschmelzung über Grid-Funktionalität führt zu neuen Flächen, deren Umrisslinien-Verlauf nicht mehr identisch mit dem Verlauf dieser im Ausgangscover ist. Dabei hängt die Abweichung vom Cellsize-Parameter ab. Kleine Cellsize-Werte führen zu langen Rechenzeiten. Bei weiterer Nutzung des Verschmelzungscovers für Verschneidungen und Überlagerungen muss eine Bereinigung der Zwickelflächen (Sliver-Polygone) erfolgen.
- Das Ergebniscover enthält nur die Flächen der ausgewählten Attribute, die mit der Bedingung Distanz h <= gewählter Distanz neu gebildet wurden.

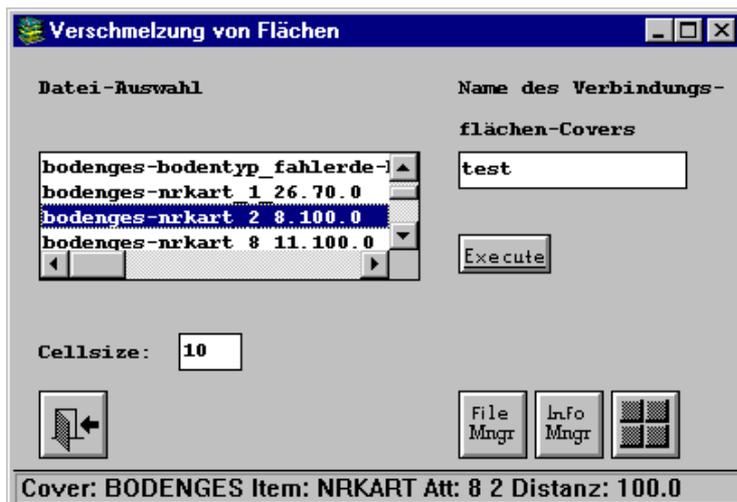


Abb. 47: Bildschirmabzug der Oberfläche zur Flächen-Verschmelzung

AML-Notation: <> Zeichenketten in spitzen Klammern bezeichnen Pflichtparameter  
 %% In Prozentzeichen eingefasste Zeichenketten bezeichnen Variablenamen  
 {} In geschweifte Klammern eingefasste Zeichenketten bezeichnen optionale Parameter bzw. Platzhalter

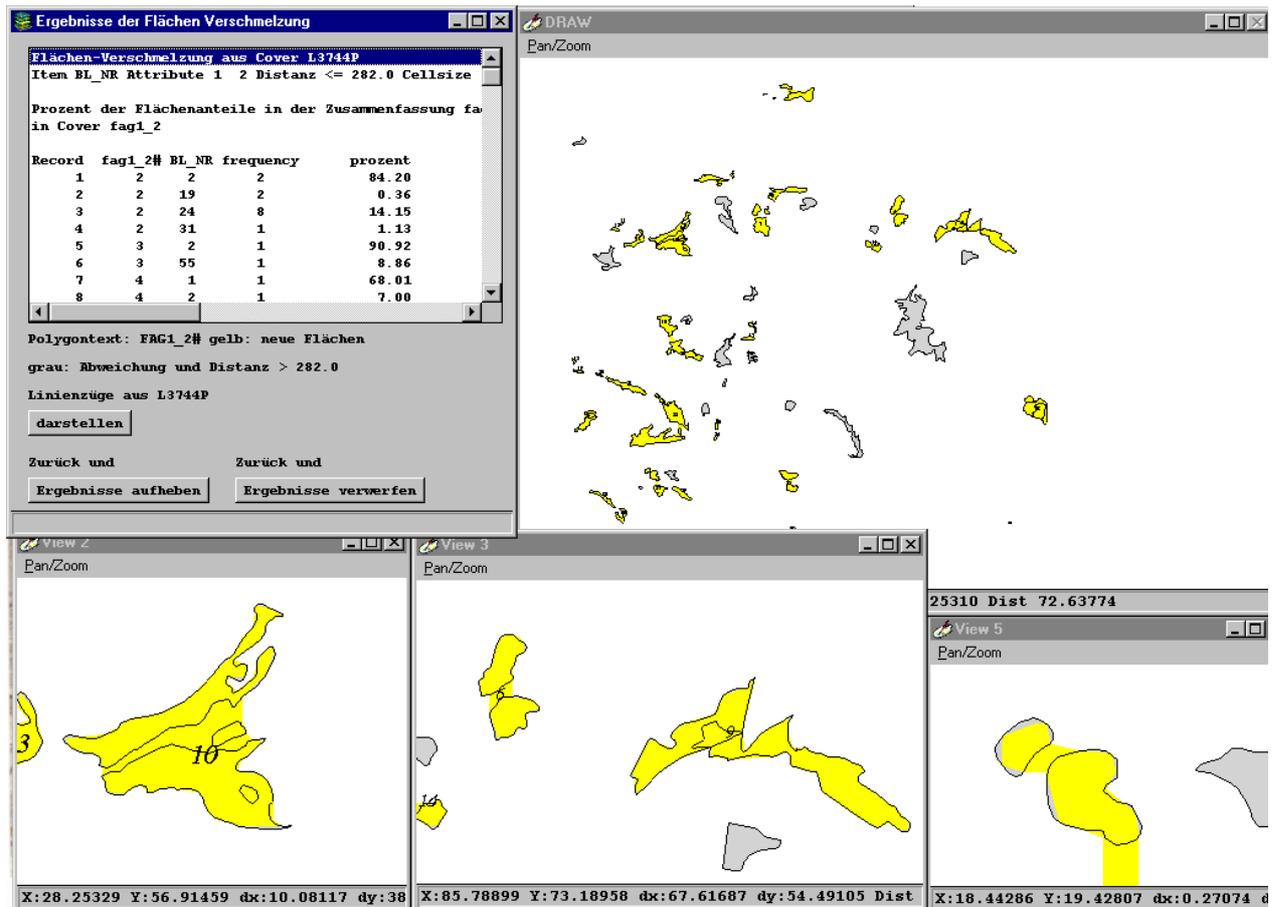


Abb. 48: Bildschirmabzug der Oberfläche zur Darstellung der Ergebnisse der Flächen-Verschmelzung

#### 10.8.10 Komponente "Datenbankanbindung/Tabellenauswahl für multivariate Statistik"

Über den in ODBC vergebenen Datenquellennamen wird die gewünschte Access-Datenbank verbunden.

Die ODBC-Anbindung erfolgt über die Ikone 'ODBC-Datenquellen' der Systemsteuerung von NT 4.0.

Folgende Aktionen sind für den erfolgreichen Zugriff auf Tabellen und Sichten der betreffenden Datenbank notwendig: Es werden zwei Tabellen mit den Namen user\_tables und user\_views, die im Feld table\_name (text 255) bzw. view\_name (text 255) die Namen der Datenbankobjekte beinhalten, angelegt. Diese Namenskonventionen wurden der Bezeichnung der Systemvariablen in der Oracle-Umgebung nachgestaltet. Die Aktion erfolgt mit dem Verbindungsaufbau. Bei Trennung der Verbindung werden die beiden Tabellen wieder gelöscht.

Um die beiden Tabellen bei der Verbindung erstellen zu können, sind folgende Eintragungen einmal für die jeweilige Datenbank unter Access vorzunehmen:

1. Unter Extras, Optionen, Ansicht werden die Systemobjekte sichtbar gemacht.
2. Unter Extras, Sicherheit, Benutzer- und Gruppenberechtigungen, Berechtigungen wird für die Tabelle MsysObjects 'Daten lesen' aktiviert.
3. Die Sichtbarkeit der Systemobjekte kann wieder zurückgesetzt werden.

Ist die Verbindung hergestellt, erscheint im Tabellen-View Auswahlfenster die Liste der Datenbankobjekte in der Form:

TABLE\_NAME Name des ersten Datenbank(Tabelle-)-Objektes

...

VIEW\_NAME Name des letzten Datenbank(Sicht)- Objektes

alphabetisch geordnet.

Die aktuell selektierte Tabelle oder Sicht wird mittels Pfeilbutton in das Info-Tabellensystem übernommen. Enthalten die Feldnamen der gewählten Tabelle bzw. Sicht Sonderzeichen, schlägt die Aktion fehl.

Nach erfolgreichem Übertragen kann im nächsten Schritt das Objektfeld gewählt und bestätigt werden, das Feld, welches die Objekte enthält die multivariat analysiert und gruppiert werden sollen.

Die Variablen können nachfolgend gewählt werden. Das ist in zwei verschiedenen Modi möglich:

1. Select - Die Variablen werden einzeln im über den Button angebotenen Auswahlfenster selektiert und bestätigt. Zur Auswahl stehen nur Felder mit Zahlenformaten. Die ausgewählten Variablen werden fortlaufend im Arc-Fenster angezeigt.
2. Blockselect - Die Variablen werden im Block gewählt (erstes Feld /letztes Feld). Es werden alle Felder angeboten. Eine Prüfung auf gültige Zahlenformate erfolgt nicht.

Ist die Variablenauswahl abgeschlossen wird die Info-Tabelle gelöscht und die Daten sind in Form einer externen ASCII-Datei für die multivariate Analyse vorbereitet. Die zusätzliche Meldung: 'Datei für multivariate Analyse vorbereitet' erscheint. Die Daten können über den button 'Datenview' gesichtet werden.

Eine Transformation der Daten kann über den button 'Daten standardisieren' vorgenommen werden. Es handelt sich hier um eine Standardisierung, d. h. gleichzeitige Zentrierung und Normierung, eine Verteilung der Daten mit Mittelwert 0 und Standardabweichung 1. Die Formel lautet:

$$u_i = \frac{x_i - \bar{x}}{s_x} \quad \langle \bar{u} = 0, s_u = 1 \rangle \quad (39)$$

$u_i$  - standardisierter Wert der Variablen  $i$

$x_i$  - Wert der Variablen  $i$

$\bar{x}$  - Mittelwert

$s_x$  - Standardabweichung

Eine weitere in Form eines externen Programms zur Verfügung stehende Möglichkeit, ist die des Transponierens der Daten. Die Zeilen und Spalten einer Datenmatrix werden getauscht.

$$\mathfrak{S} \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \alpha_{14} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \alpha_{24} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \alpha_{34} \end{pmatrix} \Rightarrow \mathfrak{S}' \begin{pmatrix} \alpha_{11} & \alpha_{21} & \alpha_{31} \\ \alpha_{12} & \alpha_{22} & \alpha_{32} \\ \alpha_{13} & \alpha_{23} & \alpha_{33} \\ \alpha_{14} & \alpha_{24} & \alpha_{34} \end{pmatrix}$$

Ein Name für das neue Objektfeld wird abgefragt, dabei wird der Name des alten Objektfeldes angeboten. Diese Funktionalität über den button 'Daten transponieren' erlangt dann Bedeutung, wenn:

1. die Abfragen auf Tabellen mehr als 255 Spalten-Datenfelder produzieren würden,
2. die Abfragen auf Tabellen zu komplex werden,
3. die Feldnamen in einer Sicht oder Tabelle Sonderzeichen enthalten.

*Ein Vorschlag zur Bereitstellung der Daten unter MS Access:*

Die Flächendatenbank für bodenkundliche Kartenwerke kann sehr komplex und mit (n) Relationen ausgestattet sein. Deshalb ist es sinnvoll, die zu analysierenden Daten möglichst über Sichten bereitzustellen.

<pre> TRANSFORM Count(Kartiereinheiten.KARTEINHNR) AS [Anzahl von KARTEINHNR] SELECT Kartiereinheiten.BL_NR FROM Kartiereinheiten GROUP BY Kartiereinheiten.BL_NR PIVOT "Hgg" &amp; [HGG] &amp; "PedDom" &amp; [PED_DOM]; </pre>	<pre> TRANSFORM Count([KARTEINHNR])/Count([KARTEINHNR]) AS Ausdr2 SELECT Kartiereinheiten.BL_NR FROM Kartiereinheiten GROUP BY Kartiereinheiten.BL_NR PIVOT "Hgg" &amp; [HGG] &amp; "PedDom" &amp; [PED_DOM]; </pre>
--	--

Beispiele für Sichten: SQL-Statements die in Form der Kreuztabellierung Legendeneinheiten durch die Häufigkeiten ihrer pedologischen Dominanz (links) und als vorhandene/nicht vorhandene pedologische Dominanz (rechts) vorbereiten.

#### Anmerkungen:

- Wurden neue Tabellen oder Sichten in der Datenbank erstellt oder Tabellen aus dem Info-Tabellensystem übernommen, auf die zugegriffen werden soll, muss die Tabellen-View-Auswahl über den button 'aktualisieren' erneuert werden.

- Beim Verlassen des Menüs muss die Verbindung zur Datenbank bestehen bleiben, wenn die ausgewählten Daten mit den multivariaten Werkzeugen bearbeitet werden sollen.
- Auf die Umlaute Ä, ä, Ö, ö, Ü und ü und Zeichen ß und - in der Benennung der Tabellen, Sichten und Feldnamen muss verzichtet werden.

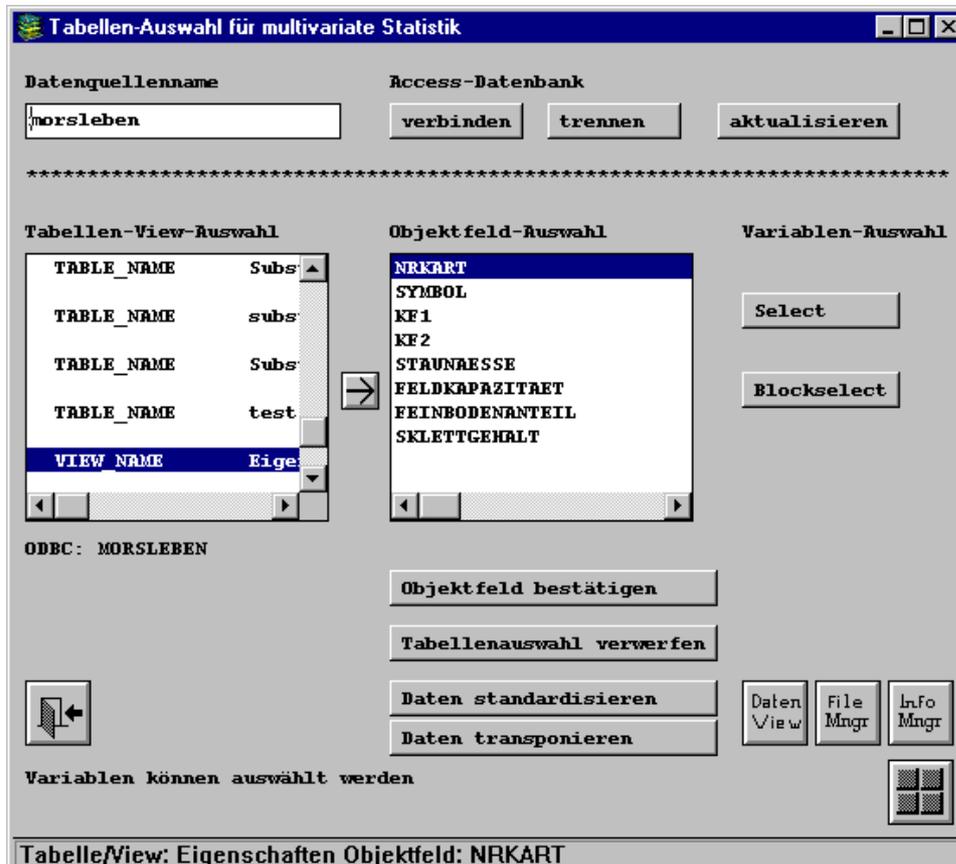


Abb. 49: Bildschirmabzug der Oberfläche der Komponente "Datenbankanbindung/Tabellenauswahl für multivariate Statistik"

### 10.8.11 Die gemeinsamen Teilkomponenten der Gruppierungsverfahren

Die multivariaten Gruppierungsverfahren greifen immer auf die zuletzt in der Komponente '.../Tabellenauswahl für multivariate Statistik' ausgewählten und bereitgestellten Daten zu. Der Dateiname der gültigen Auswahl wird im unteren Bereich der Menüs zusammen mit Objekt- und Variablenanzahl angezeigt. Die Daten können über den button 'Datenview' gelistet werden. Den drei Komponenten gemeinsam ist die Berechnung der F- und t- Werte über den button 'F- und t-Werte' (angezeigt werden die laufenden Nummern der Objekte und Variablen, assoziiert mit den F- und t-Werten und Gruppen-Mittelwerten) und die Möglichkeit der Sicherung von bestimmten Clusterlösungen. Die F- und t-Werte berechnen sich wie folgt:

$$F = \frac{V(J,G)}{V(J)} \quad t = \frac{\bar{X}(J,G) - \bar{X}(J)}{S(J)} \quad (40, 41)$$

$V(J,G)$  - Varianz der Variablen J in der Gruppe G;  $V(J)$  - Varianz der Variablen J in der Erhebungsgesamtheit

$\bar{X}(J,G)$  - Mittelwert der Variablen J über die Objekte der Gruppe G;  $\bar{X}(J)$  - Mittelwert der Variablen J in der Erhebungsgesamtheit

$S(J)$  - Standardabweichung der Variablen J in der Erhebungsgesamtheit

Die F-Werte (40) stellen ein Kriterium zur Beurteilung der Homogenität der jeweiligen Gruppe dar. Überschreitet der F-Wert 1, liegt die Varianz der Variablen in den Gruppen über der in der Erhebungsgesamtheit. Wird mit dem Distanzmaß Tanimoto (Jaccard) gerechnet, ergeben sich bereits in den ersten Heterogenitätssprüngen F-Werte  $> 1$ .

Die t-Werte (41) dienen der Charakterisierung der jeweiligen Gruppe. Positive t-Werte kennzeichnen die Variable in der Gruppe im Vergleich zur Erhebungsgesamtheit als überrepräsentiert, negative t-Werte weisen auf Unterpräsenz der Variablen bei Bildung der Gruppe hin.

Die Gruppen-Mittelwerte der Variablen werden mit ausgegeben.

Wird eine bestimmte Gruppenlösung gesichert, entstehen zwei Dateien, eine externe Ergebnisdatei und eine Infotabelle. Der Name ist für beide gleich und setzt sich aus dem Präfix %objectname% gefolgt von einer eindeutigen fünfstelligen Zahl und dem Suffix .clul und der Gruppenanzahl zusammen.

Die externe Ergebnisdatei besitzt folgenden Aufbau:

```

Clusterergebnisse für Clusterlösung %group%
Eingangsdaten: %Name der Tabelle bzw. Sicht%.txt
%m% Objekte aus %objectname%
%n% Variablen
Methode: %method%
Abstandsmaß: %proximity%
Gruppenzuordnung in Info-Datei %objectname%XXXXX.clul%group%
F- und T-Werte:
%Gruppe% %Mitgliederanzahl%
%Mitgliedernummern%
Variable 1 %F-Wert% %T-Wert% %Gruppen-Mittelwert%
...
Variable n %F-Wert% %T-Wert% %Gruppen-Mittelwert%
...
Objekte:
Objnr 1 %Objektname 1%
...
Objnr m %Objektname m%
Variable:

```

Varnr 1 %Variablenname 1%  
 ...  
 Varnr n %Variablenname n%

Die zugehörige Infotabelle enthält die Objektnummern, die Objekte und im Item Cluster die Zuordnung der Fälle in die Gruppenlösung.

Struktur der Infotabelle %objectname%XXXXX.clul%group%

ITEM NAME	WIDTH	OUTPUT	TYPE	N.DEC
OBJNR	4	5	B	-
CLUSTER	4	5	B	-
%itemname%				%itemdefinition%

*Anmerkung:* Die Namens-Konvention für die Ergebnisdateien sollte beibehalten werden, da die Komponente zum Visualisieren der Gruppenlösungen auf der Auswahl von Dateien mit dem Suffix \*.clul\* basiert.

### 10.8.12 Komponente "Expertengruppierung"

Die Objekte werden über Pfeilbutton gruppiert. Nach vollständiger Übernahme und Zuordnung der Objekte in Gruppen muss über den button 'Gruppendatei schreiben' der Vorgang abgeschlossen werden (Abb.50). Dann ist es möglich, die F- und t-Werte für die Variablen in den entsprechenden Gruppen berechnen zu lassen.

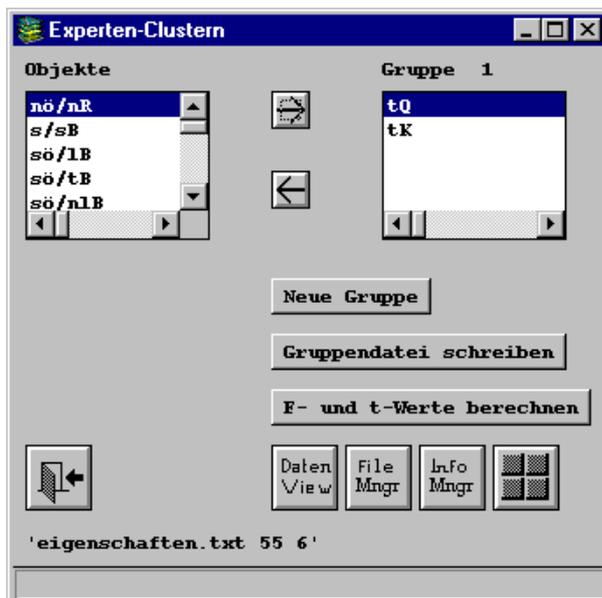


Abb. 50: Bildschirmabzug der Oberfläche der Komponente "Experten-Clustern"

AML-Notation: <> Zeichenketten in spitzen Klammern bezeichnen Pflichtparameter  
 %% In Prozentzeichen eingefasste Zeichenketten bezeichnen Variablennamen  
 {} In geschweifte Klammern eingefasste Zeichenketten bezeichnen optionale Parameter bzw. Platzhalter

### 10.8.13 Komponente "Hierarchische Gruppierung (Clustern)"

Die verfügbaren Distanzmaße (Proximitäten) sind:

Euklidische Distanz

(für metrische Variablen)

$$d_{k,l} = \sqrt{\sum_{j=1}^j |x_{kj} - x_{lj}|^2} \quad (42)$$

Quadrierte Euklidische Distanz

(für metrische Variablen)

$$d_{k,l}^2 = \sum_{j=1}^j |x_{kj} - x_{lj}|^2 \quad (43)$$

$d_{k,l}$  - Distanz der Objekte k und l;  $x_{kj}$ ,  $x_{lj}$  - Wert der Variablen j bei Objekt k, l

Das euklidische Distanzmodell (42, 43) spiegelt die Unähnlichkeiten zwischen den Objekten wider. Zwei Objekte sind umso ähnlicher, je kleiner die Distanz zwischen ihnen ist.

Ein Distanzmaß für binäre Variablen ist der

Tanimoto (Jaccard)-Koeffizient

(für binäre Variablen)

$$r_{k,l} = \frac{a}{a + b + c} \quad (44)$$

$r_{k,l}$  - Ähnlichkeit der Objekte k und l

a - Anzahl der Eigenschaften bei Objekt k und l gleich

b - Anzahl der Eigenschaften die nur bei Objekt k auftreten

c - Anzahl der Eigenschaften die nur bei Objekt l auftreten

Q-Korrelationskoeffizient

(für metrische Variablen)

$$r_{k,l} = \frac{\sum_{j=1}^j (x_{jk} - \bar{x}_k) * (x_{jl} - \bar{x}_l)}{\sqrt{\sum_{j=1}^j (x_{jk} - \bar{x}_k)^2 * \sum_{j=1}^j (x_{jl} - \bar{x}_l)^2}} \quad (45)$$

$x_{jk}$  - Ausprägung der Eigenschaft j bei Objekt k

$x_{jl}$  - Ausprägung der Eigenschaft j bei Objekt l

$\bar{x}_k$  - Durchschnittswert aller Eigenschaften bei Objekt k

$\bar{x}_l$  - Durchschnittswert aller Eigenschaften bei Objekt l

Wertebereich für beide Distanzmaße:  $0 \leq r_{k,l} \leq 1$ ; 0 - die Objekte sind vollkommen unähnlich; 1 - die Objekte sind in bezug auf die beschreibenden Variablen vollkommen ähnlich.

Der Q- Korrelationskoeffizient (45) gibt Ähnlichkeiten der Objekte in bezug auf gleiche Profilverläufe wieder. Der Tanimoto (Jaccard) Koeffizient (44) misst den relativen Anteil gemeinsamer Eigenschaften für binär skalierte Variablen.

Chi-Quadrat

$$d_{k,l} = \frac{\sum_{j=1}^j (fb_j - fe_j)^2}{fe_j} \quad (46)$$

$d_{k,l}$  - Distanz der Objekte k und l

$fb_j$  - Beobachtete Häufigkeit der Variablen in Objekt k und l

$fe_j$  - Erwartete Häufigkeit der Variablen in Objekt k und l

Das Chi-Quadrat Wert (46) misst die Unähnlichkeiten für Häufigkeitsverteilungen (Histogramme).

Die verfügbaren Fusionierungsalgorithmen sind:

Nächster Nachbar (Single-Linkage)

$$D(R;P+Q) = 0,5 * \{D(R,P) + D(R,Q) - |D(R,P) - D(R,Q)|\} \rightarrow D(R;P+Q) = \min \{D(R,P); D(R,Q)\} \quad (47)$$

Single-Linkage ist ein kontrahierendes Verfahren, welches zur Kettenbildung neigt.

Entferntester Nachbar (Complete-Linkage)

$$D(R;P+Q) = 0,5 * \{D(R,P) + D(R,Q) + |D(R,P) - D(R,Q)|\} \rightarrow D(R;P+Q) = \max \{D(R,P); D(R,Q)\} \quad (48)$$

Complete-Linkage ist ein dilatierendes Verfahren, d. h. es neigt dazu, kleine Gruppen zu bilden

$D(P,Q)$  - Distanz zwischen den Gruppen P und Q

$D(R,P)$  - Distanz zwischen den Gruppen R und P

$D(R,Q)$  - Distanz zwischen den Gruppen R und Q

Nach erfolgreicher Berechnung erscheint die Meldung 'Fertig! Ergebnisse können ausgewertet werden'. Der button 'Elbow' ermöglicht jetzt, den Verlauf des Fehlerkoeffizienten in Form eines Diagramms zu betrachten, die Gruppenlösungen, an denen Sprünge des Koeffizienten zu verzeichnen sind, für die weitere Betrachtung zu lokalisieren. Die Zahl der Cluster ist auf der X-Achse abgetragen, der Koeffizient auf der Y-Achse. Der Verlauf des Koeffizienten erfolgt für die Abstandsmaße (Euklid, Chi-Quadrat) von rechts nach links aufsteigend, da die Heterogenität in den gebildeten Gruppen zunimmt. Für die Ähnlichkeitsmaße (Tanimoto-, Q-Korrelationskoeffizient) fällt der Koeffizient im Diagramm mit fortschreitender Fusionierung, die vereinigten Objekte werden unähnlicher. Nach Eintrag der interessierenden Gruppenlösung können die F-, t-Werte und Gruppen-Mittelwerte der Variablen für diese Lösung angezeigt und bei Bedarf gesichert werden.

*Anmerkungen:* Die Wahl der Verfahrenskombination nimmt direkten Einfluss auf die Interpretationswürdigkeit der Ergebnisse. So sollen beispielsweise Ähnlichkeiten für Objekte binär skalierten Datenmatrizen ausschließlich mit dem Jaccard-Koeffizienten berechnet werden. Fehlt in solcherart Daten in einer Variable (Spalte) der Wert 1, macht sich dies bei der Auswertung der F- und t-Werte bemerkbar, für diese Variable liegt keine Berechnung vor (Kennzeichnung durch N. A.). Die Lösung sollte verworfen und eine neue Sicht auf die zu analysierenden Daten aufgebaut werden.

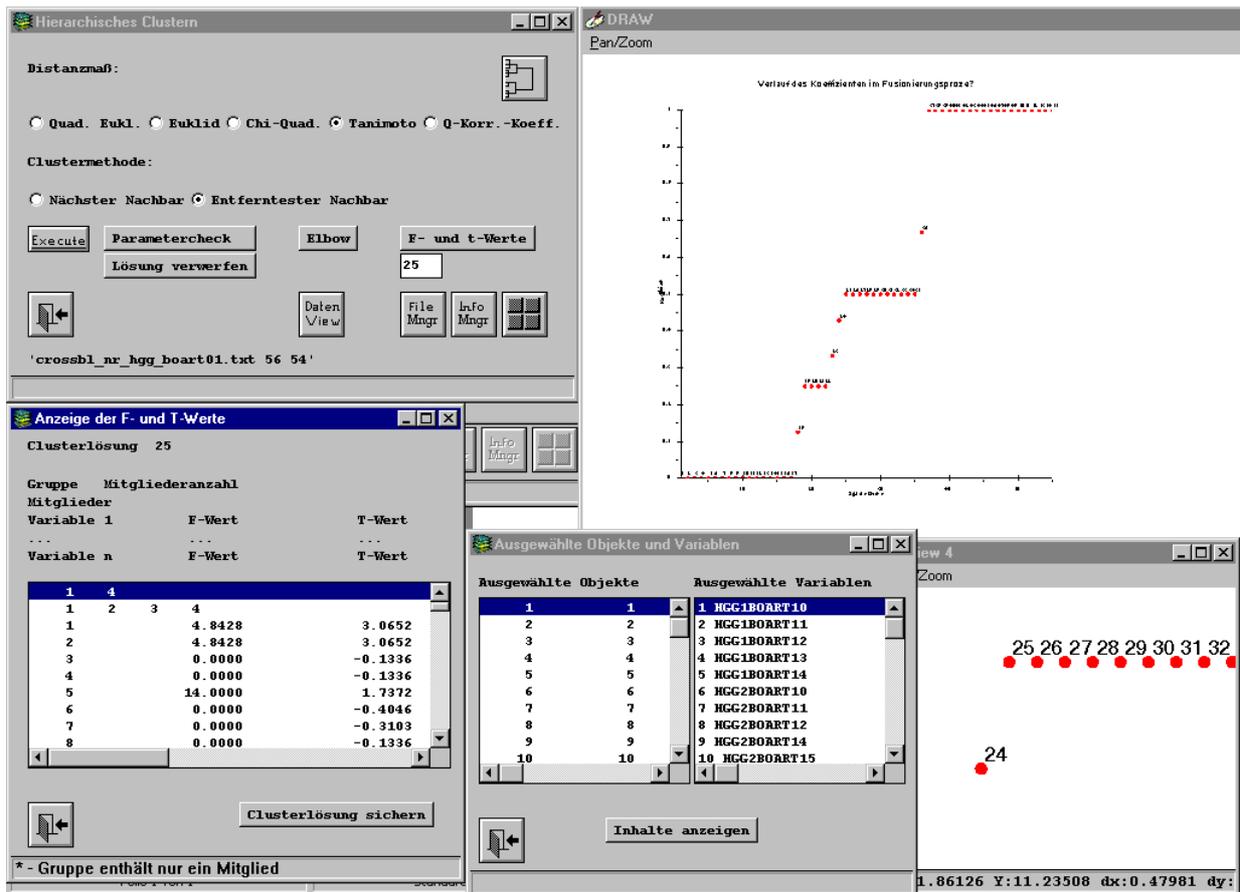


Abb. 51: Bildschirmabzug der Oberfläche der Komponente "Hierarchisches Clustern"

Die Berechnung der Proximitäten zwischen den Objekten und die Fusionierung der Objekte ist ein rechentechnisch aufwendiger Prozess. Vor allem die Fusionierung erfordert CPU-Leistung, denn sie verlangt eine ständig erneuerte Berechnung der Abstände zwischen den Gruppen bei deren Bildung. Das Clusterprogramm kann 200 Objekte mit 1000 Variablen verarbeiten. Handelt es sich um Integer Variablen kann diese Zahl auch praktisch verarbeitet werden. Die Begrenzung der Variablenanzahl liegt einerseits bei Access. Dort sind 255 Datenfelder möglich. Durch die Möglichkeit des Transponierens der Datenmatrix nach ihrer Auswahl in der Komponente '!.../Tabellenauswahl für ...' lässt sich diese Zahl erhöhen. Die Info System-Variable \$MAXREC = 4096 Byte ermöglicht die Haltung von

AML-Notation: <> Zeichenketten in spitzen Klammern bezeichnen Pflichtparameter  
 %% In Prozentzeichen eingefasste Zeichenketten bezeichnen Variablenamen  
 {} In geschweifte Klammern eingefasste Zeichenketten bezeichnen optionale Parameter bzw. Platzhalter

1024 Integer Items, dementsprechend für Long Integer 512 Items, usw.. Mit der zu verarbeitenden Objektanzahl, steigt die Rechenzeit exponentiell.

#### 10.8.14 Komponente "Zweidimensionale Skalierung"

Die zur Verfügung stehenden Distanzmaße, sind denen der Hierarchischen Clusteranalyse analog (42 - 46). Die berechneten Proximitäten werden jedoch ranggereiht, wobei das ähnlichste Objektpaar in Abhängigkeit von der Art des Maßes der Proximität den Rang 1 erhält und das Unähnlichste den letzten Rang. Den Objekten werden Zufallskoordinaten zugeordnet und die Abstände zwischen den Objekten, repräsentiert durch x,y-Koordinaten, mittels einem iterativen Optimierungsverfahren (Gradientenverfahren, Algorithmus von Kruskal) so angepasst, dass die Lage der Objekte in der Ebene ihre Ähnlichkeit in Form des Abstands zueinander widerspiegeln. Dabei wird solange verbessert bis ein minimaler Stress erreicht ist. Zugrunde liegt hier die Euklidische Metrik für zwei Dimensionen.

Die Formel zur Anpassung der Koordinatenwerte in der Ebene an die Unähnlichkeiten zwischen den Objekten lautet:

$$x_{kr}^+ = x_{kr} + \frac{\alpha}{K-1} \sum_{l=1}^K \left(1 - \frac{\delta_{kl}}{d_{kl}}\right) * (x_{lr} - x_{kr}) \quad (r = 1, 2)$$

(49)

Die Formel zur Bestimmung des Stressmaßes lautet:

$$STRESS = \sqrt{\frac{\sum_k \sum_l (d_{kl} - \delta_{kl})^2}{\sum_k \sum_l d_{kl}^2}}$$

(50)

$x_{kr}^+$  - verbesserter Koordinatenwert für Objekt k in Dimension r

$x_{kr}$  - Koordinatenwert für Objekt k in Dimension r;  $x_{lr}$  - Koordinatenwert für Objekt l in Dimension r

$\alpha$  - Schrittweite; K - Anzahl der Objekte

$d_{kl}$  - Distanzen zwischen Objekten k und l (Euklidische Metrik);  $\delta_{kl}$  - Disparitäten für Objekte k und l

Die Schrittweite  $\alpha$  kann variiert werden, um möglichst die beste Lösung zu finden.

Die Eigenschaften in Form der Variableninhalte können für die Lösungsebene skaliert werden; d. h. die Minima bzw. Maxima der Variableninhalte werden der Lösung über die gemittelten Koordinatenwerte der Objekte die diese Extremeigenschaften besitzen zugeordnet.

Ist die Rechnung erfolgt, werden im unteren Teil des Menüs die Namen der externen Ergebnisdateien angezeigt und in der Meldungszeile das Stressmaß.

Anpassungsgüte	STRESS
gering	0,2
ausreichend	0,1
gut	0,05
ausgezeichnet	0,025

Gute und ausgezeichnete Anpassungsgüten sind anstrebenswert. Ist der STRESS = 0, existieren mehrere Lösungen.

Die Anzahl der Objekte sollte 9 nicht unterschreiten (Verdichtungskoeffizient  $\geq 2$  für Dimension 2)

Tab. 18: Beurteilung des STRESS; Güte der Lösung

Die Lösung kann nun angezeigt, gruppiert oder verworfen werden. Die Anzeige auf dem Grafikbildschirm bildet im oberen Teil die Objekte, repräsentiert durch die laufenden Objektnummern, in der Ebene ab. Ist die Rechnung mit der Skalierung der Eigenschaften erfolgt, sind die laufenden Nummern der Variablen am Punkt ihres Minimums und Maximums eingetragen. Der untere Teil des Grafikbildschirms zeigt das Streudiagramm der linearen Anpassung zwischen Distanzen und Disparitäten. Da es sich bei der Berechnung der Disparitäten um eine schwach monotone Transformation der Unähnlichkeiten handelt, ist die Diagrammdarstellung eine weitere Möglichkeit die Lösungsgüte einzuschätzen. Die Gruppierung der Objekte wird durch das Hierarchische Clustern (Methode: Quadrierte Euklidische Distanz, Entferntester Nachbar) der Objektkoordinaten vorbereitet. Eine gewählte Gruppenlösung kann nun angezeigt, in Form der Gruppenmittelpunkte in der Lösungsebene dargestellt werden. Die zugehörigen F- und t-Werte der Variablen in den Gruppen können angezeigt und bei Bedarf gesichert werden.

#### Anmerkungen:

- Die Begrenzung der zu verarbeitenden Datenmenge liegt bei 200 Objekten mit 1000 Variablen. Mit der Anzahl der Objekte steigt auch hier die Rechenzeit exponentiell, da in jedem Schritt die Koordinatenwerte für alle Objekte neu berechnet werden.
- Die Zweidimensionale Skalierung bietet sich für Objekte mit Variablen metrischer Skala an.
- Um die bestmögliche Lösung (STRESS  $\gg$  min) zu finden, d. h. das globale Lösungsoptimum zu erreichen, kann die Schrittweite  $\alpha$  variiert werden.

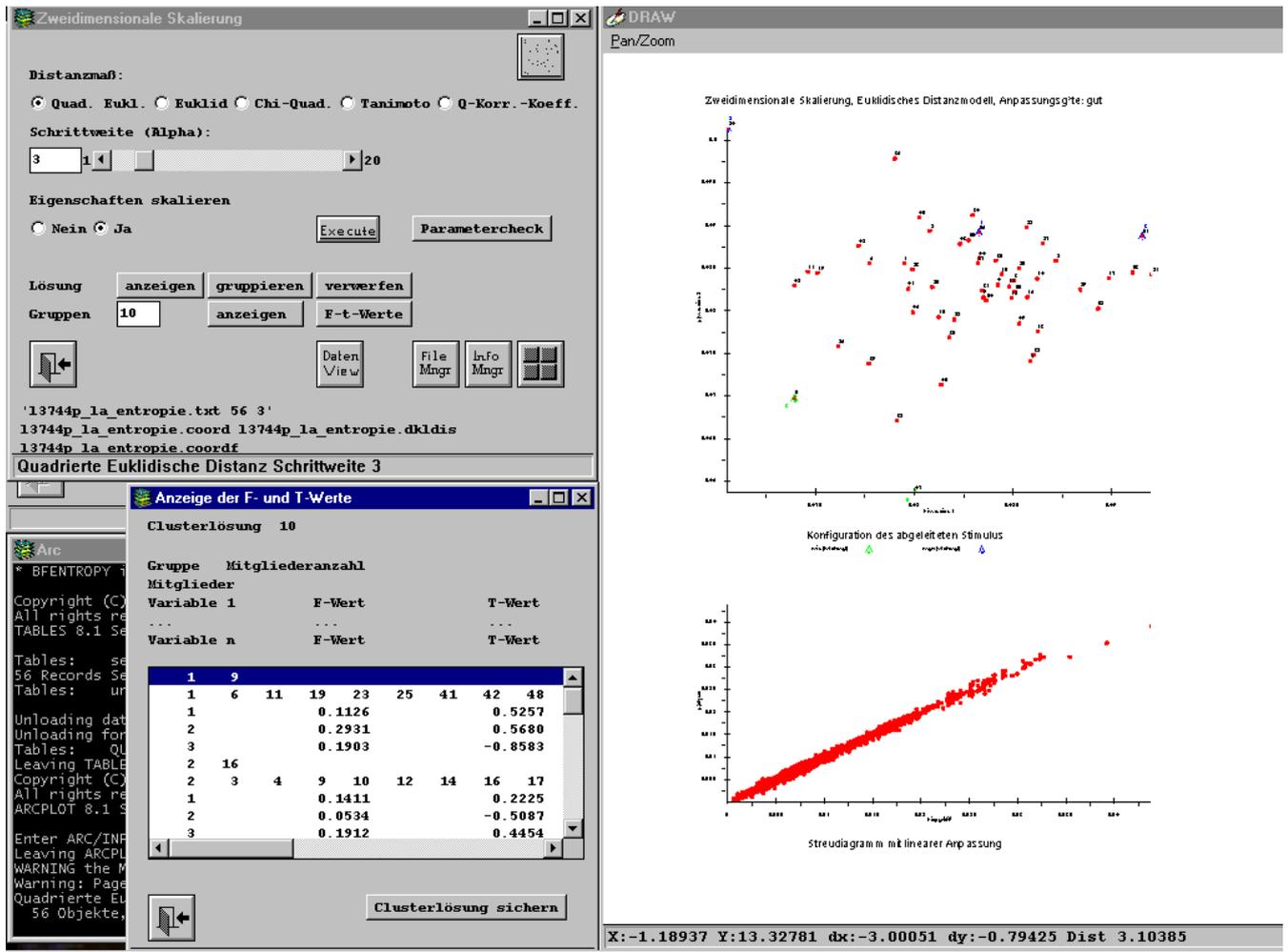


Abb. 52: Bildschirmabzug der Oberfläche der Komponente "Zweidimensionale Skalierung"

### 10.8.15 Komponente "Inventar-Entropie"

Die Komponente dient zur Kennzeichnung der inneren Heterogenität (Inventar-Areal-Heterogenität) mittels Entropieformel. Bezogen auf die Flächenanteile des Inventars der inhaltlichen Beschreibung von Objekten, werden für diese die inneren Heterogenitäten quantifiziert.

Grundlage bildet der Shannon Diversitäts Index.

$$SHDI = -\sum_{i=1}^m (p_i * \ln p_i) \quad (51) \quad \text{Wertebereich: } 0 \leq SHDI \leq \ln m$$

$p_i$  - Proportion von prozentualem Flächenanteil des Inventars mit 100 % des Objektes

$m$  - Anzahl der Inventartypen im Objekt

AML-Notation: <> Zeichenketten in spitzen Klammern bezeichnen Pflichtparameter  
 % % In Prozentzeichen eingefasste Zeichenketten bezeichnen Variablenamen  
 {} In geschweifte Klammern eingefasste Zeichenketten bezeichnen optionale Parameter bzw. Platzhalter

Mit der Formel (51) wird die Entropie von Objekten bezüglich ihres Inventars berechnet. Die Komponente greift auf die durch die Datenbankanbindung vorhandenen Tabellen/Sichten zu und listet diese in einem Auswahlfenster (Abb. 53). Über Pfeilbutton wird das gewählte Datenbank-Objekt in das Info-Tabellensystem übernommen. Hier werden Gruppierungsfeld, Inventarfeld und Flächenanteilsfeld abgefragt, die ausgewählt jeweils bestätigt werden müssen. Die Berechnung ist für folgenden Aufbau von Tabellen/Sichten gedacht (Tab. 19):

<Gruppierungsfeld>	<Inventarfeld 1>	{ }	{Inventarfeld n}	<Flächenanteilsfeld 1>	{ }	{Flächenanteilsfeld n}
Objekt 1	Inventar 1		Inventar 1	Flächenanteil		Flächenanteil
...	...		...	...		...
Objekt 1	Inventar n		Inventar n	Flächenanteil		Flächenanteil
...	...		...	...		...
Objekt m	Inventar 1		Inventar 1	Flächenanteil		Flächenanteil
...	...		...	...		...
Objekt m	Inventar n		Inventar n	Flächenanteil		Flächenanteil

Tab. 19: Aufbau der Eingangstabelle zur Berechnung der Inventar-Entropie. Feldreihenfolge und Sortierung sind unerheblich. Die Flächenanteile liegen in Prozent vor, wobei die Summe für das jeweilige Inventar eines Objektes 100 % ergibt.

Als Ergebnis entsteht eine Infotabelle mit der Bezeichnung: %tabellen/sichten\_name%-inventarfeld%.ien. Sie enthält die Objekte des Gruppierungsfeldes, deren Häufigkeiten und die berechneten Shannon-Indizes. Zur weiteren Bearbeitung können die Ergebnistabellen in die Datenbank übernommen werden (Exportfunktion im Info Datei-Manager).

Struktur der Infotabelle %tabellen/sichten\_name%-inventarfeld%.ien:

```

ITEM NAME  WIDTH OUTPUT  TYPE N.DEC
FREQUENCY   4   5   B   -
%itemname%  %itemdefinition%
ENTROPY      8  18   F   6

```

**Anmerkung:** Nicht vollständig gruppierte Eingangsdaten, d. h. gleiches Inventar tritt mehrfach für ein Objekt auf, führen zu einem fehlerbehafteten Ergebnis.

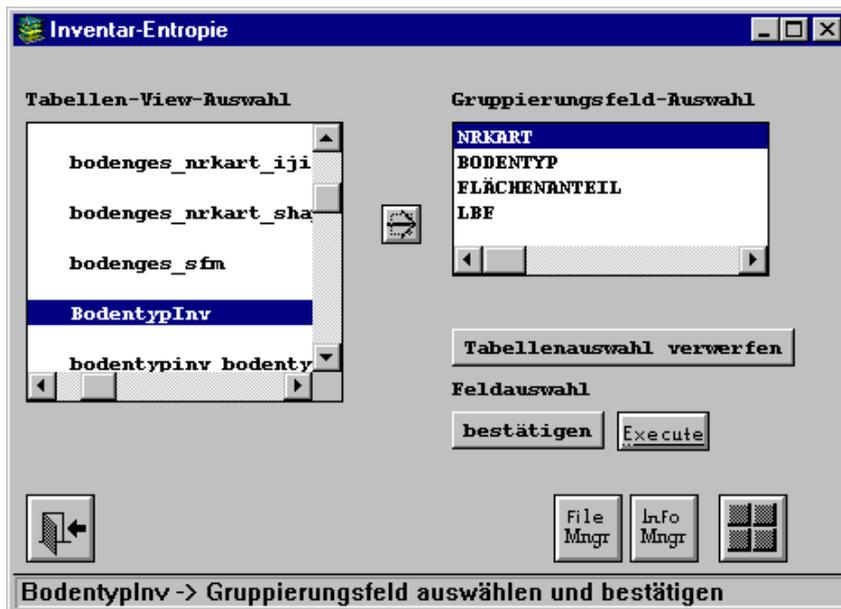


Abb. 53: Bildschirmabzug der Oberfläche der Komponente "Inventar-Entropie"

#### 10.8.16 Komponente "Konfigurationsfrequenzanalyse"

Die Komponente dient zur Typisierung in kreuztabellierten Sachverhalten. Die Komponente greift auf die in der Datenbankanbindung vorhandenen Tabellen/Sichten zu und listet diese in einem Auswahl-fenster (Abb. 54). Für ein gewähltes Datenbank-Objekt in diesem Fenster kann die Berechnung durchgeführt werden. Die Analyse baut auf folgender Tabellen/Sichten-Struktur auf (Tab. 20):

Objekt/Attribut (Typ Character)	Attribut/Objekt				
	1 (Typ Integer)	2 (Typ Integer)	...	.. (Typ Integer)	j (Typ Integer)
1	$n_{11k}$	$n_{12k}$			
2	$n_{21k}$	$n_{22k}$			
..					
..					
i	$n_{i1k}$				$n_{ijk}$

$i \times j$  Kreuztabelle,  $n$  - Zellenhäufigkeit

Tab. 20: Tabellen/Sichten-Struktur für die Konfigurationsfrequenzanalyse

Das Datenbank-Objekt besitzt ein Textfeld mit Objekten oder Attributen. Die anderen Spalten-Attribute/Objekte sind Felder des Typs Integer. Die Zellen  $n_{ij,k}$  beinhalten Häufigkeitswerte. Andere Strukturen werden zurückgewiesen.

Für die Kreuztabelle wird der  $\chi^2$ -Test (52) und für die Zellenkombinationen der  $\chi^2$ -Komponententest von Lienert (53) durchgeführt. Die  $\chi^2$ -(Global-)Testentscheidung beruht auf dem Vergleich von  $\chi^2$  mit dem kritischen Wert  $\chi^2_{\alpha;f}$  (wenn  $\chi^2 < \chi^2_{\alpha;f}$  - totale Unabhängigkeit der Merkmale; wenn  $\chi^2 \geq \chi^2_{\alpha;f}$  - wechselseitige Abhängigkeiten zwischen den Merkmalen). Die Beurteilung, ob eine signifikante Abweichung der Zellenhäufigkeit von ihrem Erwartungswert vorliegt, beruht auf dem Vergleich von  $\chi_{ijk}$  mit einer festgelegten Schranke. Überschreitet  $\chi_{ijk}$  die vereinbarte Schranke kann ein Typ vermutet werden, im Fall einer negativen Abweichung ein Antityp.

Formel zur Berechnung von  $\chi^2$

$$\chi^2 = \sum_{i=1}^k \frac{(n_{ijk} - e_{ijk})^2}{e_{ijk}} \quad (52)$$

Formel zur Berechnung von  $\chi_{ijk}$

$$\chi_{ijk} = \frac{(n_{ijk} - e_{ijk})}{\sqrt{e_{ijk}}} \quad (53)$$

$n_{ijk}$  - Zellenhäufigkeit;  $e_{ijk}$  - erwartete Zellenhäufigkeit (Erwartungswert), k - Objektanzahl

Kontingenzkoeffizient C

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} \quad (54)$$

Kontingenzkoeffizient K

$$K = \sqrt{\frac{\chi^2}{n(q-1)}} \quad (55)$$

n = Spaltensumme = Zeilensumme; q = min(Spalten-,Zeilenzahl)

Wertebereich:  $0 \leq C; K \leq 1$

Die Kontingenzkoeffizienten (54, 55) können herangezogen werden, um eine Aussage zur Ausprägung des Zusammenhangs (wenn  $\chi^2 \geq \chi^2_{\alpha;f}$ ) zu geben.

Für den  $\chi^2$ -Komponententest werden folgende Schranken genutzt:

Freiheitsgrad (f) 1  $\alpha = 5\%; \chi^2 = 3,84$

$\alpha = 1\%; \chi^2 = 6,64$

$\alpha = 0,1\%; \chi^2 = 10,8$

Aufbau der Ergebnisdatei %tabellen/sichten\_name%-inventarfeld%.kfa:

\*Chi-Quadrat-Zellenest (Signifikante Zellen)

T - Typ, AT - Antityp, alpha - Irrtumswahrscheinlichkeit

01 - alpha = 0.1%, 1 alpha = 1%, 5 alpha 5%

%Objekt i, %      Attribut j<sub>1</sub>      % $\chi^2$  %      % $\chi$ %      {T/AT%alpha%}

...

...

...

...

...

AML-Notation:

<> Zeichenketten in spitzen Klammern bezeichnen Pflichtparameter

% % In Prozentzeichen eingefasste Zeichenketten bezeichnen Variablenamen

{ } In geschweifte Klammern eingefasste Zeichenketten bezeichnen optionale Parameter bzw. Platzhalter

%Objekt  $i_n$       Attribut  $j_n$        $\% \chi^2$  %     $\% \chi$  %    {T/AT%alpha%}

\*Chi-Quadrat-Test:

Chi-Quadrat            :  $\% \chi^2$  -Global%

Freiheitsgrad         : %f%

Anzahl gültiger Zellen    : %Zellenwert > 0%

Anzahl signifikanter Zellen : %Anzahl von Typen und Antitypen%

Kontingenzkoeffizient (K) : %K%

Kontingenzkoeffizient (C) : %C%

*Anmerkung:* Der Anteil der Zellen in der Kreuztabelle mit erwarteten Häufigkeiten < 5 sollte 20% nicht überschreiten.

The screenshot shows a window titled 'ARC' with a list of soil types and their corresponding Chi-square values, expected frequencies, and significance levels. The list includes soil types like Lehm, Löss, Lösssand, Sand, and Schluff, along with their attributes (e.g., GLEY, FAHLERDE, KOLLUVIUM). The results are presented in a table format with columns for the soil type, Chi-square value, expected frequency, and significance level (e.g., T01, AT1, T5).

Soil Type	$\chi^2$	Expected Freq	Significance
Lehm GLEY	69.09	8.31	T01
Löss FAHLERDE	27.08	5.20	T01
Löss FAHLERDE-GLEY	150.48	12.27	T01
Löss über Lehm FAHLERDE	46.93	6.85	T01
Löss über Lehm FAHLERDE-PSEUDOGLEY	63.39	7.96	T01
Löss über Lehm PARABRAUNERDE-PSEUDOGLEY	63.39	7.96	T01
Löss über Schutt REDZINA	320.19	17.89	T01
Lösssand über Lehm BRAUNERDE	6.84	-2.62	AT1
Lösssand über Lehm KOLLUVIUM	6.25	2.50	T5
Lösssand über Lehm PSEUDOGLEY	61.55	7.85	T01
Lösssand über Sand BRAUNERDE	49.34	7.02	T01
Lösssand über Sand FAHLERDE-BRAUNERDE	20.01	4.47	T01
Lösssand über Sand PARABRAUNERDE-BRAUNER	10.01	3.16	T1
Lösssand über Sand PSEUDOGLEY	5.21	-2.28	AT5
Lösssand über Sandlöss KOLLUVIUM	57.14	7.56	T01
Lösssand über Ton PSEUDOGLEY-BRAUNERDE	57.15	7.56	T01
Sand GLEY	117.33	10.83	T01
Sand über Schutt PODSOL	130.43	11.42	T01
Sand über Ton PSEUDOGLEY-BRAUNERDE	31.73	5.63	T01
Sandlöss über Lehm FAHLERDE	7.61	2.76	T1
Sandlöss über Lehm PARABRAUNERDE	31.73	5.63	T01
Sandlöss über Löss FAHLERDE	7.70	2.78	T1
Sandlöss über Sand BRAUNERDE	7.28	2.70	T1
Sandlöss über Ton PSEUDOGLEY	26.79	5.18	T01
Schluff GLEY	23.03	4.80	T01
Schuttsand über Sand PODSOL	195.64	13.99	T01
Ton KOLLUVIUM	36.01	6.00	T01
Ton REGOSOL	190.48	13.80	T01
Torf über Mudde KALKNIEDERMOOR	332.01	18.22	T01

\*Chi-Quadrat-Test:

Buttons: Continue, Pause, Quit

Abb. 54: Bildschirmabzug der Ergebnisdarstellung der Komponente "Konfigurationsfrequenzanalyse"

AML-Notation:      <> Zeichenketten in spitzen Klammern bezeichnen Pflichtparameter  
 % % In Prozentzeichen eingefasste Zeichenketten bezeichnen Variablenamen  
 {} In geschweifte Klammern eingefasste Zeichenketten bezeichnen optionale Parameter bzw. Platzhalter

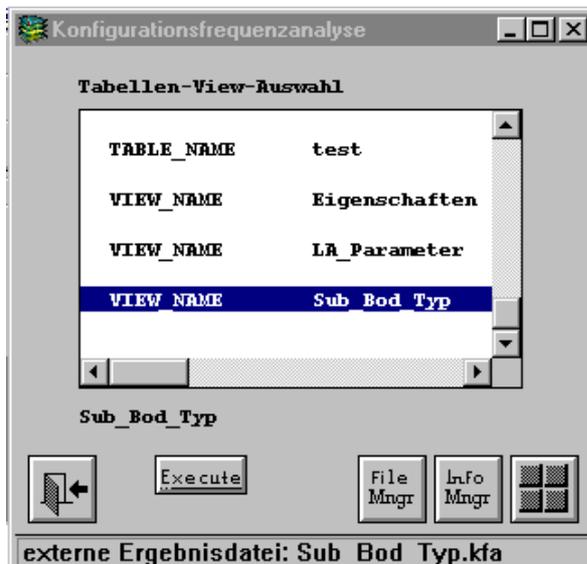


Abb. 55: Bildschirmabzug der Oberfläche der Komponente "Konfigurationsfrequenzanalyse"

#### 10.8.17 Komponente "Visualisieren von Gruppenlösungen"

Hiermit können die durch die multivariaten Verfahren gewonnenen Gruppierungen in ihrem geografischen Bezug dargestellt werden.

Aus der aktuellen Workspace-Umgebung wird die Auswahl der Polygoncover, ihrer zugehörigen Items der Polygon-Attribut Tabellen und der Infotabellen mit den gespeicherten Gruppenlösungen angeboten. Polygoncover, Verknüpfungsim (Item in der PAT = Objektfeld aus der multivariaten Analyse) und Gruppenlösungsdatei werden gewählt. Die externe Ergebnisdatei für die Gruppenlösung wird bei Wahl angezeigt (Dateiaufbau: siehe Gemeinsame Teilkomponenten der multivariaten Gruppierungsverfahren). Existiert eine gültige Verknüpfung, können die Gruppen auf dem Grafikbildschirm angezeigt werden. Hier kann zwischen zwei Darstellungsvarianten gewählt werden, der Gesamtdarstellung und der Darstellung einzelner Gruppen. In der Gesamtdarstellung werden die Polygone entsprechend ihrer Gruppenzugehörigkeit, unter Nutzung der Standardfarbtabelle colornames.shd, eingefärbt. Werden einzelne Gruppen aus der Lösung gewählt, können diese in einer bestimmten Farbe gezeichnet werden. Im ARC-Fenster erscheinen Informationen zur Gruppe, das sind: Flächenanzahl, Mitgliederanzahl, Mitglieder und entscheidende Variablen (wenn  $|t\text{-Wert}| \geq 1$ ) (41).

Die ausgewählte Gruppenlösung kann über den Button 'Ergebnisdateien entfernen' gelöscht werden, das bedeutet, externe Ergebnisdatei und Infotabelle %objectname%XXXXX.clul%group% werden gelöscht.

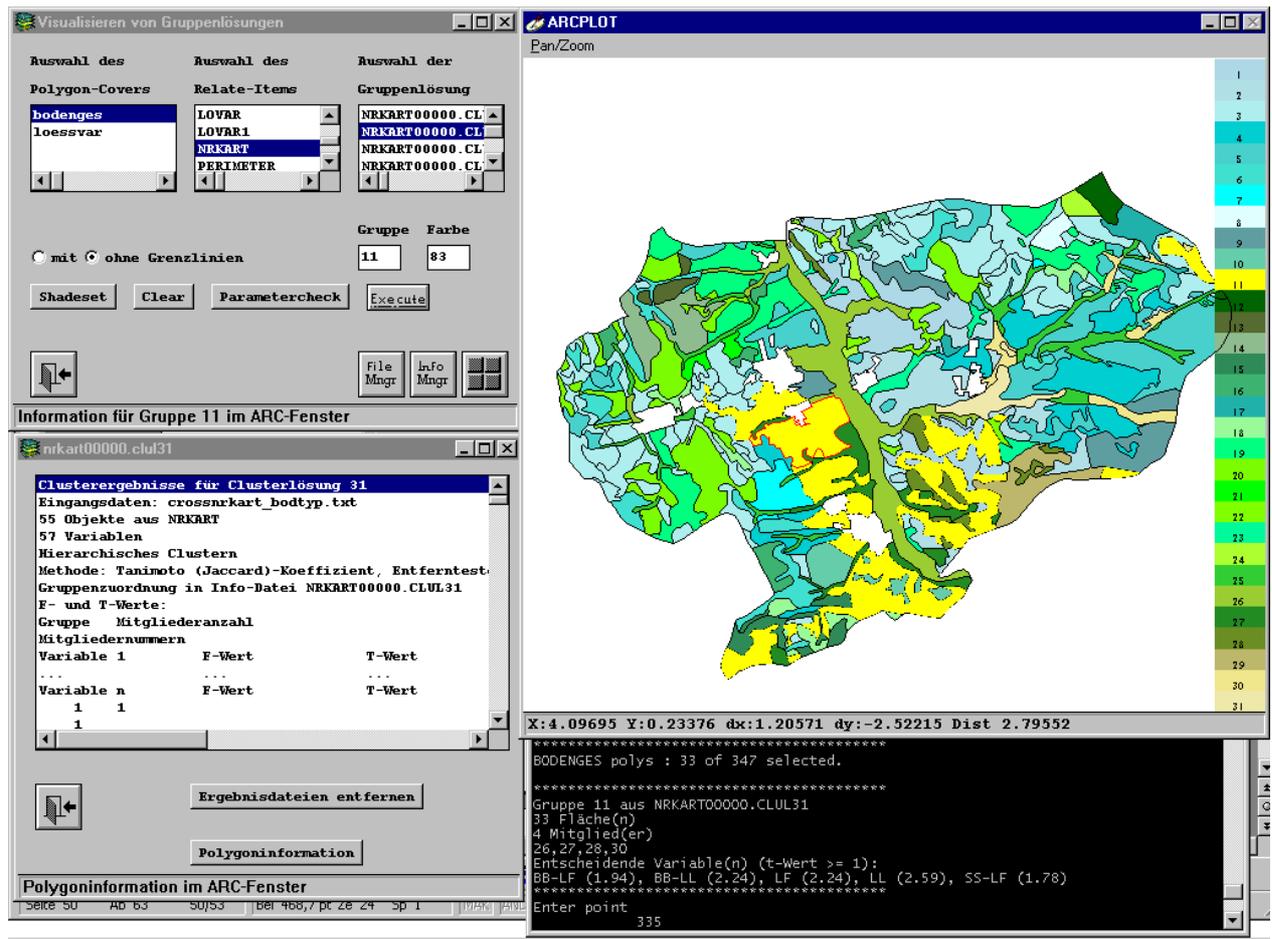


Abb. 56: Bildschirmabzug der Oberfläche zum Visualisieren von Gruppenlösungen

## 10.9 Literatur zur Herkunft der Formeln und Verfahren

In Klammern sind die jeweiligen Programmkomponenten, in denen die Formeln und Verfahren umgesetzt sind, vermerkt.

- BACKHAUS, K.; ERICHSON, B.; WULFF, P.; WEIBER, R.(1996): Multivariate Analysemethoden: eine anwendungsorientierte Einführung. Berlin; Heidelberg; New York; London; Paris; Tokyo: Springer.
- (Hierarchisches Clustern und Zweidimensionale Skalierung, F- und t-Werte)
- BORG, E.; FICHELTMANN, B. (1998): "Vergleichende Analyse von Formindizes zur Charakterisierung von Landschaftsobjekten unter ökologischen Aspekten". Zeitschrift für Photogrammetrie und Fernerkundung, 4, 108-119. (Landschaftsanalyse Teil: Gestaltindizes von Einzelflächen)
- CLAUß, E. (1982): Statistik für Soziologen, Pädagogen, Psychologen und Mediziner. Frankfurt am Main: Harri Deutsch. (Basisformeln und Konfigurationsfrequenzanalyse)

- Göhler, W. (1983): Höhere Mathematik -Formeln und Hinweise- Kleiner Wissensspeicher. Leipzig: VEB Deutscher Verlag für Grundstoffindustrie. (*Index maximaler Reduktion Teil: Vorschlag zur Distanzmessung und darauf aufbauender Verbindungsflächenermittlung*)
- LAUTSCH, E.; VON WEBER, ST. (1995): Methoden und Anwendungen der Konfigurationsfrequenzanalyse (KFA). Weinheim: Belz, Psychologie-Verl. Union. ( *Konfigurationsfrequenzanalyse*)
- LI, H...; REYNOLDS, F. (1993): A new contagion index to quantify spatial pattern of landscapes. Landscape Ecology 8, 155-162. (*Landschaftsanalyse Teil: Durchdringungs- und Einsprengungsindex*)
- LOHSE, H.; LUDWIG, R.; RÖHR, M. (1982): Statistische Verfahren für Psychologen, Pädagogen und Soziologen. Berlin: Volk und Wissen. (*Datentransformation*)
- MANDELBROT, B. B. (1987): Die fraktale Geometrie der Natur. Basel, Boston: Birkhäuser Verlag. (*Landschaftsanalyse Teil: Gestaltindizes*)
- MCGARIGAL, K.; MARKS, B. J. (1994): FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure. Oregon: Forest Science Department. (*Landschaftsanalyse*)
- ROSSITER, D. G. (1998): Methodology for Soil Resource Inventories. International Institute for Aerospace Survey & Earth Sciences (ITC). (*Index maximaler Reduktion*)
- VÖLZ, H. (1990): Computer und Kunst. Leipzig, Jena, Berlin: Urania. (*Landschaftsanalyse Teil: Standardisierte Entropie/Goldener Schnitt*)

#### 10.10 Dokumentationen der eingesetzten Softwareprodukte

- ESRI (1994): ArcInfo Version 7.0 Handbücher und ArcInfo Online Help.
- Lahey Computer Systems (1994-98): Lahey/Fujitsu Fortran 95, Express User's Guide, *Revision A*.
- Lahey Computer Systems (1994-98): Lahey/Fujitsu Fortran 95, Language Reference, *Revision D*.
- Microsoft Access 97 Online Hilfe.
- Microsoft ODBC Online Hilfe.

#### 10.11 Glossar der wichtigsten verwendeten ArcInfo-Ausdrücke

Arc-Attribut Tabelle (.AAT) - Enthält die mit der topologischen Struktur (Linie) des Covers assoziierten Attribute.

AML - Arc Macro Language; Interpretersprache im ArcInfo

Cover(age) - Georeferenziertes Vektordatenmodell in ArcInfo, in Form eines Unterverzeichnisses in der gleichen Verzeichnishierarchie mit dem Info-Verzeichnis abgelegt.

Erläuterung aus dem ArcInfo Glossar:

1. A digital version of a map forming the basic unit of vector data storage in ArcInfo. A coverage stores geographic features as primary features (such as arcs, nodes, polygons, and label points) and secondary features (such as tics, map extent, links, and annotation). Associated feature attribute tables describe and store attributes of the geographic features.
2. A set of thematically associated data considered as a unit. A coverage usually represents a single theme such as soils, streams, roads, or land use.

Fuzzy-Toleranz - Diese Größe gibt die interne Bearbeitungsgenauigkeit für topologiebearbeitende ArcInfo Kommandos vor. Voreingestellt ist eine Fuzzy-Toleranz von 1/10000 der maximalen Ausdehnung des Covers. Die Fuzzy-Toleranz wird mit der AML-Direktive &describe <geo\_dataset> in die Systemvariable DSC\$FUZZY geschrieben.

Info - Datenhaltungs- und Datenverwaltungssystem als Bestandteil von ArcInfo. Führt unter anderem in internen Tabellenformaten die mit den Topologien assoziierten Attribute.

Item - Feld mit definierten Eigenschaften (Name, Typ, interne Größe) in einer Info-Tabelle

Polygon-Attribut Tabelle (.PAT) - Enthält die mit der topologischen Struktur (Fläche) des Covers assoziierten Attribute. Der Mindestumfang der Infotabelle beinhaltet für jedes Polygon die Flächengröße, den Umfang, die interne - und externe Polygon-ID. Hinzu kommen von Anwender definierte und eingefügte Items.

Widget - grafisches Steuerelement in einem Form Menü, aktionsausführend oder parametergewinnend; Im Programm genutzte Widgets sind :

button -	Knopf
slider -	Schieber
checkbox -	Schalter
choice -	Auswahl
input -	Eingabe

Workspace - Bezeichnet das aktuelle Arbeitsverzeichnis, in dem die topologischen Strukturen in Unterverzeichnissen abgelegt sind und ein Infoverzeichnis mit den zugehörigen Attributtabelle existiert.

Esri bietet unter [http://www.esri.com/library/glossary/i\\_1.html](http://www.esri.com/library/glossary/i_1.html) ein GIS Glossar an, in dem allgemeine und ArcInfo spezifische Ausdrücke erläutert werden.

## 10.12 Glossar der verwendeten Datenbank-Ausdrücke

ODBC - Object DataBase Connect ist eine Programmierschnittstelle mit der Anwendungen auf Daten in Datenbankmanagementsystemen zugreifen können, die SQL als Standardsprache für den Zugriff unterstützen.

Sichten (views) - in der Entwurfsansicht von MS Access oder durch SQL-Anweisungen erstellte virtuelle Tabellen

SQL - Structured Query Language, standardisierte strukturierte Datenbank-Abfrage-Sprache