

DATENANALYSE UND KLASSIFIKATION

H. Goebel · M. Schader (Hrsg.)

Datenanalyse, Klassifikation und Informations- verarbeitung

Methoden und Anwendungen
in verschiedenen Fachgebieten

Physica-Verlag Heidelberg

Ein Unternehmen des Springer-Verlags



SERION — Programmpaket zur chronologischen und chorologischen Auswertung archäologischer Daten

P. Stadler

Naturhistorisches Museum Wien, Prähistorische Abteilung
Burgring 7, Postfach 417, 1014 Wien, AUSTRIA

Inhalt: Es wird das Programmpaket SERION vorgestellt, das auf dem Großrechner IBM 3090 läuft und von dem es in naher Zukunft eine kompatible PC-Version geben wird. Dabei werden spezielle auf archäologische Fragestellungen angepaßte Verfahren verwendet. Neben einer kurzen Darstellung dieser Prozeduren und der Erklärung der Syntax zur Erstellung von Jobs wird anhand von praktischen Beispielen die Flexibilität des Programmes demonstriert.

1 Zweck des Programmpaketes

Das Programmpaket SERION wurde für den Zweck entwickelt, geschlossene archäologische Fundkomplexe aufgrund ihrer Typenkombinationen relativchronologisch auszuwerten. Diese Ergebnisse können nach verschiedenen Verfahren erhalten und auch grafisch dargestellt werden. Ferner werden Kartierungen für Gräberfelder, Siedlungen oder Großräume vorgenommen, wobei die Möglichkeit besteht, zufällige Verteilungen von notwendigen zu unterscheiden. Schnittstellen zu den Programmen SPSSX, SAS und CLUSTAN ermöglichen eine große Flexibilität. Dazu müssen die Daten gewissen Regeln entsprechend erfaßt werden. Mithilfe der SERION Control Language (SCL) können entsprechende Jobs zusammengestellt werden, die unterschiedliche Auswertungen durch parametergesteuerte Prozeduren ermöglichen. Schnittstellen zur Paläozoologie und -botanik, zur Anthropologie und Chemie, ermöglichen es, komplexe Fragestellungen - an denen verschiedene Wissenschaften teilnehmen - zu lösen und in bisher unbekannte Bereiche vorzustoßen (z.B.: Vergleich Soziologie - Familien- oder Verwandtschaftstruktur der Bestatteten eines Gräberfeldes mit der Relativchronologie der Fundkomplexe)

2 Benutzerkreis

Derzeit kann von allen Nodes, die im EARN (European Academic Research Network) integriert sind (also ca. 3200 Rechenzentren), durch Absenden eines Jobs mit Daten an eine bestimmte Server-ID an der Universität Wien (AWIUNI11) SERION kostenlos benutzt werden. Dem Interessenten wird auf Anfrage an A7101DAA AT AWIUNI11 diese ID und eine Benutzerberechtigung mit PassWort bekannt gegeben. Die gewünschten Ergebnisse werden automatisch an den Absender zurückgesandt. Über die Aktivitäten der Benutzer wird Protokoll geführt. Dieser Job wird über EARN an den Server an der Universität Wien (AWIUNI11) gesandt. Die zugehörigen Daten müssen in diesem Job inkludiert sein.

Derzeit wird an einer daten- und resultatportablen PC-Version dieses Programmpaketes gearbeitet.

3 Syntax von SCL

Jeder Job mit beliebigem FileName muß den FileType "SERION" haben. Der Job muß in SCL geschrieben sein. SCL ist eine parameterorientierte Sprache. Dabei wird zuerst immer eine Prozedur angegeben. In der nächsten Zeile folgen die Parameter für diese Prozedur. In Tab. 1 folgt ein Beispiel für einen Job (ohne Daten) in SCL. Im folgenden werden die Prozeduren und die Parameternamen und -werte besprochen. Die Prozedurnamen sind im allgemeinen 8 Großbuchstaben lang (nur SERIATION ist länger), die Parameter setzen sich aus Parameternamen und -wert zusammen. Die Parameternamen sind bis zu 8 Buchstaben lang, in Groß- und Kleinschreibung, sie werden durch ein "="-Zeichen abgeschlossen, die Parameterwerte befinden sich auf der rechten Seite des "="-Zeichens und können bis zu 20 Zeichen (Buchstaben und/oder Ziffern) lang sein. Zur Abgrenzung zum nächsten Parameter folgt mindestens ein Leerzeichen. Die Parameter besitzen Defaultwerte und sonstige erlaubte Werte: Es dürfen Buchstaben, REAL-Zahlen und INTEGER-werte verwendet werden, was aber für jeden Parameter genau definiert wird. Es können Intervalle angegeben werden, in denen die Parameterwerte liegen dürfen. Bei INTEGER-zahlen z.B.: 1 - n; bei REAL-zahlen z.B.: 0.0 - r. Eine ausführliche Beschreibung der Parameter ist im Manual enthalten.

4 Prozeduren in alfabetischer Anordnung

Die Parameter besitzen immer Defaultwerte, für den Fall, daß im Job kein Wert angegeben wird. Im Allgemeinen sind diese Defaults die ersten bei den jeweiligen Parametern aufgelisteten Werte.

4.1 AVERGRAF

AVERGRAF liefert die grafische Ausgabe der besten Ordnung der Matrix nach Verfahren, die Eigenvektoren liefern. Es werden Funde und Typen entsprechend ihren Eigenvektoren geplottet. Dabei kommt es zu einer automatischen Abstimmung auf die maximale Ausgabegröße. Das Ausgabefile ist "FileName" GKSM und wird mithilfe des Erlanger Grafiksystems erstellt. In Folge kann es dann auf dem Bildschirm oder einem Plotter ausgegeben werden.

4.2 CONTVARI

Dieses Prozedur dient hauptsächlich zum Einlesen der Daten und zum Abzählen der Varianten in den Funden bzw. der Funde, in denen die Varianten auftreten. Die Daten werden auf dem Jobfile erwartet. Ansonsten auf dem File "FileName" SERINP: Für den Aufbau der Datenbank stehen zwei Möglichkeiten zur Verfügung:

1. Die Dateneingabe kann auf zwei Weisen erfolgen: durch Angabe von Typen oder Varianten. Die Typen werden vom Benutzer bestimmt, so wie das für gewöhnlich der Fall ist. Soll jedoch eine bildorientierte Datenbank geschaffen werden, so erfolgt das mithilfe von Varianten. Diese Varianten sind die "Adressen" der Abbildungen in einem Typenkatalog. So bedeutet zum Beispiel K5981.01 ein Objekt, das im Band K des Typenkatalogs auf Seite (=Typentafel) 5981 an erster Stelle abgebildet

SERIATION/D
 KATALOGE
 KatMode=F
 REDUDATA
 InList=1 Document=1 TypeLim=2 FindLim=2 ClusIn=0&
 TypeExcl=5 FindExcl=23 JoinFind=2
 PETRIFIC
 OrderN=3 miSteps=30 Document=1 RunTiDoc=0 firsTyp=10050
 GRAFIRES
 PlotSel=4 Legend=0 Orient=2 PlotMean=0 MaxDiv=20
 RECIAVER
 OrderN=1 Epsilon=3.0 miSteps=30 Document=1 RunTiDoc=0 k
 firsTyp=10050 RankEige=0 Stress=3 BreakCek=1 StepWise=0&
 StepNumb=0
 STATISTI
 Switch=41 CorrPlot=1
 LPRESULT
 GRAFIRES
 PlotSel=4 Legend=0 Oriente PlotMean=0 MaxDiv=20
 AVERGRAF
 Otient=2 PlotMean=0 DiaShow=0
 TOPOSERI
 FindV/T=10 NumbKart=20
 TOPOKART
 A-Format-4 ClusIn=1 Color=1 Legend=0 LinePrMa=1 Map=0 k
 NextNeig=0 NumbMode=0 Origin=RE OverPrnt=:0 PlotLimi=2 k
 Plotter=1 PlotNumb^1 ProbLimi=0.1 PublSize=1 RepHead=0 k
 UpLimit=200 Weight=1.0 WindowX1=-29.42 WindowX2=40.0 k
 WindowY1=0. WindowY2=48.9
 VARISEQU
 ListFind=1
 FINDSEQU
 FindSort=0 FindShow=0 FindStat=1
 KONDENSA
 Origin=RE
 STATISTI
 Switch=42 CorrPlot=1
 LPRESULT
 GRAFIRES
 PlotSel=4 Legend=0 Orient=2 PlotMean=0 MaxDiv=20
 TOPOSERI
 FindV/T=10 NumbKart=20
 VARISEQU
 ListFind=1
 FINDSEQU
 FindSort=0 FindShow=0 FindStat=1
 TRIANGUL
 Origin=RE
 STATISTI
 Switch=43 CorrPlot=:1
 LPRESULT
 GRAFIRES
 PlotSel=4 Legend=0 Orient=2 PlotMean=0 MaxDiv=20
 TOPOSERI
 FindV/T=10 NumbKart=20

Tabelle 1: Beispiel eines Jobs in SCL ohne Daten

ist oder sich in einer Bilddatenbank an entsprechender Stelle befindet. Wie dieser Typenkatalog erstellt werden soll, zeigt beigelegter Ausschnitt aus einer awarischen Datenbank. Liegt nur ein Band vor, so erübrigt sich die Angabe einer Bandnummer. 53.01 bedeutet also Abbildung 1 auf Typentafel 53.

Abgesehen von diesem variablen Aufbau der Datenstruktur, können die Daten auf zwei Weisen eingegeben werden:

- (a) fundorientiert: Das heißt zu aller erst müssen sämtliche Abbildungen aus der Literatur kopiert, ausgeschnitten (oder eingescannt und auf einer Bildplatte abgespeichert werden) und/oder typologisch geordnet werden. Denn nur wenn die Typologie dynamisch ständig an den Abbildungen überprüft und korrigiert werden kann, kann die zunächst subjektive Typologie objektiviert werden. Erst in einem zweiten Verfahren erfolgt die eigentliche Dateneingabe. In Tab. 2 gebe ich ein Beispiel für ein Datenfile, das fundorientiert erfaßt wurde.

Erklärende Kommentare sind normal gedruckt, die Daten, wie sie eingegeben werden sollen, fett..

- (b) Typorientiert: Genauso wie man eine Typentafel komplettiert, gibt man auch die Daten ein, Typentafel für Typentafel. Voraussetzung ist, daß jedes Objekt exakt beschriftet wurde. Damit nicht ständig der gesamte Fundort geschrieben werden muß, erstellt man eine Abkürzungsliste. Jede Typentafeleingabe erfolgt durch Angabe einer Abkürzung für den Typ, eine ausführliche Typbezeichnung oder -beschreibung. Danach werden - jeweils durch Leerzeichen getrennt - die Vertreter dieses Typs aufgelistet. Tab. 3 gibt diese Datenstruktur wieder.

2. Bei fundorientierter Eingabe wird eine Fundortliste erstellt mit der Angabe sämtlicher weiterer in der Literatur üblicher, oft anderssprachiger Bezeichnungen unter Verweis auf die in der Datenbank verwendete. Für die weitere Verarbeitung stehen für die Fundnummern folgende Möglichkeiten zur Verfügung: Aus dem Fundortnamen wird - wenn keine Abkürzungen angegeben sind - automatisch eine eindeutige, bis zu 10 Zeichen lange Abkürzung gebildet. Diese Abkürzung kann auch eine laufende Nummer sein. Auf dem File "FileName" NAMES befinden sich die Fundortnamen, auf dem File "FileName" NAMCORT deren Abkürzungen.

Entsprechend dem Vorliegen von Fundortabkürzungen werden die Fundnummern gebildet. z.B.: Aus "Abony" wird "Abo", so bedeutet Abo. 527 Grab 527 aus Abony. Bei typorientierter Eingabe muß zusätzlich ein File "FileName" NAMES zur Verfügung gestellt werden, auf dem sich die Abkürzungen und die ungekürzt geschriebenen Fundortnamen befinden sollen. Automatisch wird hierbei für die Fundnummern die Abkürzung verwendet. CONTVARI ist die unmittelbare Voraussetzung für die Benutzung von KATALOGE.

4.3 FINDSEQU

FINDSEQU liefert eine nach ihrer zeitlichen Abfolge gereichte Liste der Funde nach dem Seriationsergebnis unter Angabe der darin vorkommenden Typen, ihre Maximal- sowie die Standardabweichungen vom berechneten Mittelwert. Die Prozedur dient also

»Awaren 4.11.1990

Überschrift mit Datum der letzten Bearbeitung

\$Abo # Abony

Nach dem "\$"-Zeichen steht die Abkürzung des Fundortes Abony, der nach dem "^"-Zeichen ausgeschrieben angeführt wird.

Im K2970+05

der ler bedeutet Fundnummer 1, als Fundnummer können bis zu 10 Zeichen, also auch Buchstaben verwendet werden. Daran anschließend können bis zu zwei Zeichen zur Geschlechtsangabe eingegeben werden, m bedeutet Mann, / Frau, k Kind, p Pferd etc. m? bedeutet vielleicht Mann. Achten Sie auf die rechtsbündige Eingabe der Fundnummern.

2970.05 bedeutet: K = Band K, 2970 = Seite 2970, 05 = Abbildung 5 auf dieser Seite. "*" bedeutet, der Gegenstand ist auf der Typentafel abgebildet. "." bedeutet, daß dieser Gegenstand durch die Abbildung eines anderen vertreten wird, "#" zeigt, daß die Abbildung noch fehlt.

Am	GF060*02				
PferdeGrabp	K3110.03	20510*03	22400*08	24470*01	41510*02
	50640.01	50690.01	60750.01		
12m	K1010+02	12050.02	12050.04	12060.03	53390.01
13m	60280.02	63830.01			
16m	10630*01	12190*02	15280*02	40950*01	50580*01
20m	WP420*01				
»geschlossener	Fund??				
Ein Kommentar	wird durch *	in Spalte 1	begonnen		
21m	TK120*02	10660*02	60300*01	60380*01	
30m	63850*03				
88m	26340*02	41560*02			
89m	24920*06				
112m	15900*02	15900*03	16530*01	25060*01	43450*01
	50580*01				
116m	K2050*10	21910*04	24910*01	50560*01	
124m	44760*02				
184m	65850*02				
336m	22800*04	25070*03	31630*02	40580*01	45210*07
Strf1	GG030*01				
* Keramikkombination					
Strf2	GF160*01				
Strf3	GF040*04				
Strf4	GF030+03				
\$Ach	#Achmim (=Achmim-Panopolis=)				
	(=Ägypten=)				
Zweitnamen können durch "(= ;...=)"	gekennzeichnet werden.				
Strf01	100607				

Tabelle 2: Fundorientierte Eingabe

zur übersichtlicheren Ausgabe der von LPRESULT und GRAFIRES gedruckten Matrix. Zusätzlich werden die Ergebnisse auf für weitere Auswertungen verwendbare Files ausgegeben. Auf "FileName" LISTING wird ausgegeben: die einzelnen Funde in der Reihenfolge des Seriationsergebnisses oder alfabetisch sortiert mit folgenden Angaben:

1. Laufende Rangordnungszahl des Fundes im Seriationsergebnis,

Awaren 4.11.1990

Überschrift mit Datum der letzten Bearbeitung

A130#Topf, Prager Typus I

GA130 ist die Abkürzung des Typs, wobei hier GA einen Band bezeichnet und 130 die Seite angibt. Danach folgt ein kurzer Typname oder eine kurze Beschreibung; das entspricht dem Fundortnamen in der fundorientierten Eingabe. Diese Aufteilung auf Bände und Seiten kann jedoch innerhalb der maximalen Typlänge von 10 beliebig gestaltet werden.

+ Abo.Strf6 Abo*25m Abo.26m

Abo = die Abkürzung des Fundortes Abony, Strf6 = Streufund 6, 25m s Männergrab 25. "♦" bedeutet, daß der entsprechende Topf im Katalog abgebildet ist. An und für sich entspricht jeder abgebildeten Variante eine Zeile, die mit "+" beginnt. Manche nicht abgebildete Varianten können durch eine andere vertreten sein und befinden sich dann an derselben Zeile. "." zwischen Fundort und Fundnummer bedeutet, dieser Gegenstand ist nicht abgebildet, "#" zeigt, daß die Abbildung noch fehlt.

Tabelle 3: Typorientierte Eingabe

2. Nummer des Fundes in einer alphabetischen Liste,
3. Fundnummer,
4. Mittelwert (absolut),
5. Differenz der aufeinanderfolgenden Mittelwerte,
6. Standardabweichung in Sequenzdaten,
7. die maximale Abweichung eines einzelnen Typs vom Mittelwert,

8. Angabe des Typs mit dieser Abweichung in Typen und Sequenzprozent,
9. Anzahl der bei der Seriation verwendeten Typen,
10. Auflistung dieser Typen.

Bei Ausgabe der Funde in der Seriationsabfolge wird in einer weiteren mehrspaltigen Liste eine alphabetische Auflistung der Funde mit der Stellung in der Seriation zur leichteren Auffindung eines beliebigen Fundes in der vorhergehenden Liste angegeben.

4.4 GRAFIRES

GRAFIRES liefert die Ausgabe der besten Ordnung der Matrix nach Verfahren, die Rangordnungen liefern. Um beim Ausdruck Papier und Platz zu sparen, werden mehrere Zeilen bzw. Spalten übereinander geplottet. Somit kann aus der Matrix allein nicht mehr gesehen werden, in welchen Funden welche Typen vorkommen. Dabei wird automatisch auf das maximale Ausgabeformat abgestimmt. Das Ausgabeformat ist "FileName" GKSM und wird mithilfe des Erlanger Grafiksystems erstellt. In Folge kann es dann auf einem Plotter ausgegeben werden.

4.5 KATALOGE

Diese Prozedur dient zum Umschalten zwischen fund- und typorientierten Daten. Somit können Daten zuerst nach dem einen Format erstellt und korrigiert werden, dann in das andere übergeführt und weiter bearbeitet werden. Oft muß man die Fundkomplexe überprüfen, dann aber wieder die Typologie. KATALOGE erstellt also aus "FileName" FINDSERI "FileName" TYPESERI und umgekehrt. Die Zusatzinformationen zu den Daten wie Abkürzungen von Fundortnamen sowie ihre Langformen oder Kommentare werden in den Files "FileName" FINDLABS sowie "FileName" TYPELABS abgespeichert. Der genaue Ablauf sieht also folgendermaßen aus: "FileName" FINDSERI + "FileName" TYPELABS ==> "FileName" TYPESERI + "FileName" FINDLABS oder umgekehrt. In den Ergebnis-Files befinden sich die Daten alphabetisch sortiert: bei den fundorientierten Daten nach den Abkürzungen der Fundorte, bei den typorientierten Daten nach den Abkürzungen der Typenbezeichnungen.

4.6 KONDENSA

KONDENSA versucht so in den Goldmannschen Algorithmus einzugreifen, daß die Durchläufertypen bzw. -funde mit einer hohen Standardabweichung bei der Mittelwertberechnung normaler Typen bzw. Funde nicht stark berücksichtigt werden. Dabei werden die Standardabweichungen der Besetzungen in Funden und Typen als Gewichte für weitere Mittelwertberechnungen herangezogen. Nach kurzem iterativen Prozeß kommt dieses Verfahren von selbst zum Stillstand. Das Ergebnis nennen wir Kondensation der Matrix. Diese Prozedur ist momentan nur für Rangordnungen, nicht aber für Eigenvektoren implementiert.

4.7 LPRESULT

LPRESULT liefert die Ausgabe der besten Ordnung der Matrix. Um beim Ausdruck Papier und Platz zu sparen, werden mehrere Zeilen bzw. Spalten übereinander gedruckt. Somit kann aus der Matrix allein nicht mehr ersehen werden, in welchen Funden welche Typen vorkommen. Dabei wird automatisch eine Abstimmung auf die maximale Ausgabegröße vorgenommen. In den Zeilen sind die Funde angeordnet, in den Spalten die Typen. In einem rechteckigen Feld vor und nach der Matrix sind die Fundnummern - entsprechend KATALOGE - und in ebensolchen Feldern oberhalb und unterhalb der Matrix die Typen - diesmal senkrecht zu lesen - angeordnet. In der Matrix selbst befinden sich Buchstabensymbole, die die Besetzungsdichte in einem von mehreren Funden und Typen gebildeten Quadrat angeben. Dabei bedeuten "*" eine Besetzung, "A" zwei, "B" drei usw.

4.8 PETRIFIC

Diese Prozedur benutzt den Goldmann-Wilkinson-Legoux'schen (Goldmann (1980), Wilkinson (1974) und Legoux (1980)) Algorithmus und minimiert das Stresskriterium. Dazu sind unterschiedliche Dokumentationsmöglichkeiten vorgesehen. Es kann jeder Einzelschritt oder nur das Ergebnis jedes Laufes dokumentiert werden. In der ersten Spalte steht die Anzahl der bisher notwendigen Matrixumstellungen, sodann folgt die Nummer des Laufes, ferner folgen die Anzahlen der Einzelschritte für diesen Lauf, der Abbrüche durch Stillstand des Seriationsalgorithmus sowie der Schritte, bei denen sich eine Verbesserung des Stresskriteriums ergab, ferner Zeilen- und Spaltenstress und ihr geometrischer Mittelwert als Gesamtstress, jeweils in Prozent angegeben.

4.9 RECIAVER

RECIAVER benutzt den von Peter Ihm (Ihm (1983)) beschriebenen, als Reciprocal Averaging bekannten Algorithmus der Korrespondenzanalyse unter Maximierung des Korrelationskoeffizienten, der unter den Bedingungen, daß Mittelwerte und Varianzen der Variablen 0 bzw. 1 sind, in diesem Fall gleich der Kovarianz ist. Diese Prozedur sollte anstelle von PETRIFIC verwendet werden. Vorteil dieses Verfahrens gegenüber PETRIFIC liegt darin, daß unabhängig von der (zufälligen) Ausgangsordnung bei ungleichen Singulärwerten immer dasselbe Resultat erhalten wird oder aber etwas mehr mathematisch ausgedrückt: "die Korrespondenzanalyse ist eine Singulärwertzerlegung und das Reciprocal Averaging ein Algorithmus zur Berechnung der dominanten (eindeutigen), nichttrivialen Lösung."

In dieser Prozedur kann auch getestet werden, ob in den Eigenvektoren "Sprünge" auftreten. Solche bedeuten immer Inhomogenitäten oder Diskontinuitäten, die das Ergebnis beträchtlich verzerren können. Derartig störende Daten können vor einem weiteren Lauf mit der Prozedur REDUDATA eliminiert werden, wodurch diese "Sprünge" vermieden werden können. Dadurch erhält man ein "ungestörtes" Resultat.

4.10 REDUDATA

Sie dient zur Datenreduktion auf die Funde und Typen, für die gilt: jeder Fund, der mehr als FundLim Typen enthält und jeder Typ, der in mehr als TypeLim Funden auf-

tritt. Zumeist werden TypeLim und FundLim=2 gewählt. Das ist auch der Default. Die Elimination ist ein iterativer Prozeß, der so lange anhält, wie noch Typen oder Funde vorhanden sind, die obigen Bedingungen noch nicht gehorchen. Dieser Prozeß kann durch Auflisten der im jeweiligen Schritt eliminierten Funde und Typen dokumentiert werden. Außerdem kann das Ergebnis im Format von binären CLUSTAN-datenfiles ausgegeben werden. Auf "FileName" LISTING können erhalten werden: Eine Liste aller Typen vor der Elimination, eine Dokumentation des Eliminationsprozesses, die reduzierten Matrices Typen/Funde und Funde/Typen.

Durch Benutzung von Files mit Varianten-, Typen- oder Fundlisten ist es möglich, die angeführten Typen oder Funde aus den Daten zu eliminieren. Dabei ist die Verwendung von WildCards "*" erlaubt.

4.11 SERIATION

In dieser Hauptprozedur werden für den gesamten Programmverlauf geltende Parameter festgelegt. So wird bestimmt, ob der Benutzer - sofern er connected ist - an seinem Terminal Messages über den Fortgang des Jobs erhält. Wenn ja, so wird der Start einer jeden Prozedur mitgeteilt. Ferner wird bestimmt welcher von derzeit 2 implementierten Sprachen (Deutsch und Englisch) die Messages und der Output-Text abgefaßt sein sollen. Übersetzungen in andere Sprachen sind leicht zu bewerkstelligen, da sämtliche Messages nicht im Programm enthalten sind, sondern auf einem externen File vorliegen. Ferner wird bestimmt, ob die Daten im Job inkludiert sind oder auf zusätzlichen Files zur Verfügung gestellt werden, die gleich nach dem Job-File abgeschickt werden müssen.

4.12 SIMUDATA

SIMUDATA ermöglicht die Erstellung simulierter Daten, wobei unterschiedliche Modellvorstellungen verwendet werden können.

4.13 STATISTI

Diese Prozedur bereitet eine Statistik der Seriationverläufe von Petrification und Reciprocal Averaging vor. Dabei werden die notwendigen Ordnungsschritte, Mittelwerte, Minima, Maxima und Standardabweichungen der jeweiligen Stresskriterien gelistet etc. Ferner kann die Bestordnung - entsprechend dem Stresskriterium - mit den anderen Ergebnissen, den Ergebnissen von Kondensation, Triangulation sowie diese untereinander verglichen werden. Die Vergleiche erfolgen über Spearmankoeffizienten und einen weiteren Koeffizienten, der die mittlere Abweichung der Rangordnungen bzw. Eigenvektoren berechnet und in Sequenzprozent ausdrückt. Diese Vergleiche verschiedener Rangordnungen oder Eigenvektoren können auch grafisch als Korrelationsplot ausgegeben werden.

4.14 TOPOKART

TOPOKART ermöglicht das Ausdrucken von inter- und intrasite spatialen Typenverteilungen. Gleichzeitig können diese Verteilungen auf Zufälligkeit oder Notwendigkeit getestet werden. Als Grundlagen dafür können digitalisierte Pläne oder Karten verwendet werden.

4.15 TOPOSERI

Diese Prozedur teilt die Funde in Sequenzgruppen, deren Größe bestimmt werden kann. Es wird eine Art Sequenzkatalog geliefert, der als Input für Kartierungsprogramme dient. Somit kann bei Untersuchung nur eines Fundortes eine Toposeriation in verschiedenen Phasen dargestellt werden. Bei einer Großraumdatenbank kann der Besiedlungsvorgang in diesem Gebiet dargestellt werden.

4.16 TRIANGUL

Sein Ausgangspunkt kann beliebig festgelegt werden, die Matrix wird so umgeordnet, daß die Typen nach ihrem ersten Auftreten und die Funde nach ihrem jüngsten Typ gereiht werden. Es wird also das Ergebnis von RECIAVER oder PETRIFIC in die triangulierte Form übergeführt.

4.17 VARISEQU

VARISEQU berechnet für Varianten und Typen die Sequenzdaten (SD) - das ist eine Zahl zwischen 1 und 1000, die angibt, in welchem Bereich der relativen Skala der Seriation sich der entsprechende Fund befindet. Ferner werden Mittelwerte, Standardabweichungen und Konfidenzintervalle aus den Positionen der Varianten und Typen in der Seriation errechnet. Auf einen Blick ist erkennbar, in Kombination mit dem Abbildungskatalog oder einer Bilddatenbank, ob die Zuordnung einer Variante zu einem Typ chronologisch verantwortlich ist oder nicht. Für die Typen wird ebenfalls ein Mittelwert berechnet und als Gesamtmittelwert ausgegeben. (GESMW) Die Prozedur dient also zur übersichtlicheren Auswertung der von LPRESULT und GRAFIRES gedruckten Matrix. Zusätzlich werden die Ergebnisse auf für weitere Auswertungen verwendbaren Files ausgegeben.

Auf "FileName" LISTING kann ausgegeben werden: Die Varianten unter Angabe der Sequenzdaten (=SD) derjenigen Funde, in denen sie auftreten, oder die Funde zusammen mit ihren Sequenzdaten. Zunächst werden die Variantennummern, ihre Häufigkeit sowie die SD angegeben. Die Angabe eines "-" bedeutet, daß der Fund in der Seriation nicht vertreten ist, weil er möglicherweise Typen enthält, die sonst nur einmal auftreten oder aus anderen Gründen die Bedingungen zur Aufnahme in die Seriation nicht erfüllt haben. Tritt eine Variante häufiger als einmal auf, so wird ein Mittelwert (MW), durch \pm angedeutet das zugehörige Konfidenzintervall für diese Variante zusammen mit der Standardabweichung (STAV) in Sequenzprozent angegeben. Von mehreren Varianten, die zu einem Typ gehören, wird ein Gesamtmittelwert GESMW, ein Konfidenzintervall und die Standardabweichung angegeben, der nächste Typ ist durch eine Leerzeile abgetrennt.

5 Anwendungen

Folgende Ergebnisse mit SERION - die hier jedoch aus Platzmangel nur ausschnittsweise abgebildet werden können - konnten auf der Tagung vorgeführt werden:

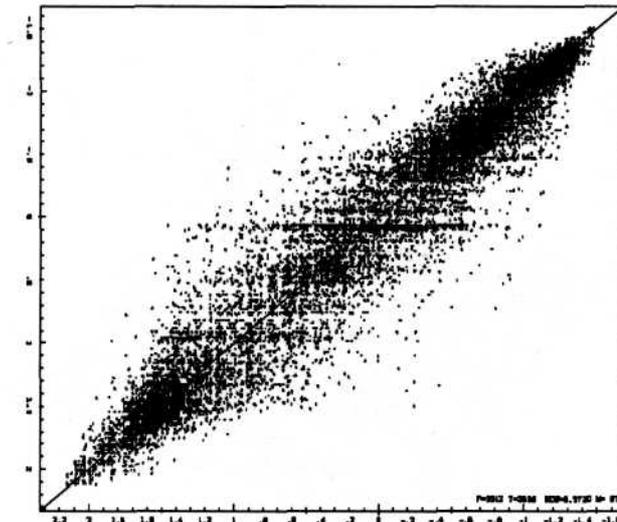


Abbildung 1: Reciprocal Averaging mit Eigenvektordarstellung

5.1 Verlauf des Reciprocal Averaging bei kleinem Datenmaterial

Am Beispiel des hallstattzeitlichen Gräberfeldes von Volders in Tirol zeigte sich, daß schwach assoziiertes Fundmaterial das Reciprocal Averaging erheblich stören kann. Deshalb empfiehlt es sich, dieses nach einem ersten Lauf ermittelte Fundmaterial in einem zweiten Lauf zu eliminieren. Das Ergebnis wird dann viel besser!

5.2 Verschiedene Seriationen zu einem großen Datensatz

Hier folgen die Abbildungen zur Seriation des großen Datensatzes meiner awarischen Datenbank.

5.2.1 Reciprocal Averaging

Reciprocal Averaging mit Darstellung der Rangordnungen.

- Reciprocal Averaging mit Eigenvektoren (siehe Abb. 1).
- 3-Dimensionale Darstellung der Eigenvektoren (siehe Abb. 2). Diese Grafik wurde über eine Schnittstelle mit dem Statistikpaket SAS erzeugt.
- Darstellung des Reciprocal Averaging mit Höhenschichtlinien (hier nicht abgebildet).

5.2.2 Petrifikation

Petrifikation (bestes Ergebnis von 3 Läufen).

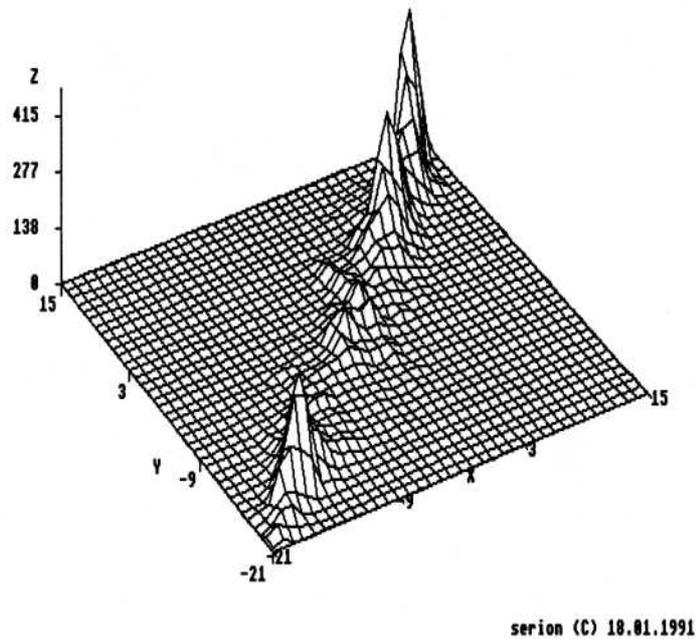


Abbildung 2: 3-Dimensionale Darstellung der Eigenvektoren

Vergleich von Reciprocal Averaging mit bester Petrifikation (von 3).

- bezüglich der Funde (siehe Abb. 3).
- bezüglich der Typen (nicht abgeb.).

5.2.3 Kondensation

- Kondensation (siehe Abb. 4).
- Vergleich von Reciprocal Averaging mit Kondensation.
 - bezüglich der Funde (siehe Abb. 5).
 - bezüglich der Typen (nicht abgeb.).

5.2.4 Triangulation

- Triangulation (siehe Abb. 6).
- Vergleich von Reciprocal Averaging mit Triangulation.
 - bezüglich der Funde (nicht abgeb.).
 - bezüglich der Typen (siehe Abb. 7).

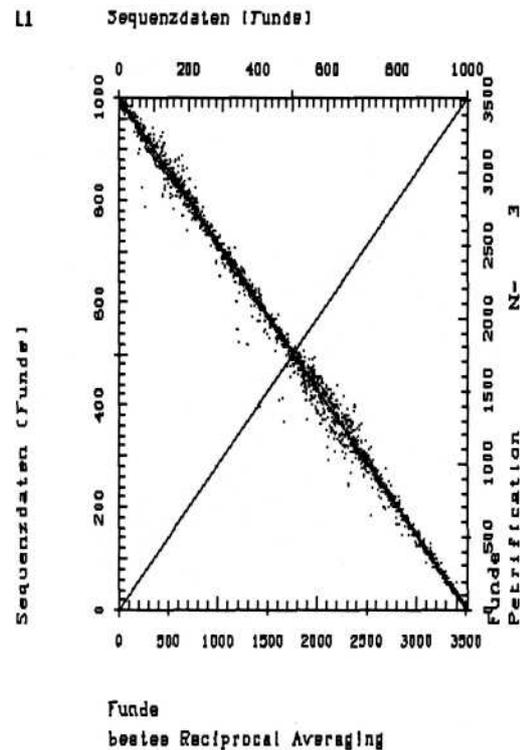


Abbildung 3: Vergleich von Reciprocal Averaging mit bester Petrifikation (von 3), bezüglich der Funde

5.3 Darstellung der chronologischen Stellung gewisser Typen

Zur Verdeutlichung der Typenabfolge kann eine Grafik erstellt werden, die die Typen, mit ihren eingescannten Abbildungen, in einer Sequenztabelle darstellt (hier jedoch nicht abgebildet).

5.4 Spatial Analysis für chronologische Sequenz

Die awarische Besiedlungstätigkeit (568-ca. 830 n. Chr.) kann in 20 relativ gleich langen Phasen dargestellt werden.

Dabei werden die Funde in 10 Gruppen von A-J unterteilt, wobei A die ältesten, J die jüngsten auf einer Karte sind. Die Karte wurde digitalisiert, in Zukunft ist auch Einscannen möglich. Hier kann nur die letzte Phase dargestellt werden (siehe Abb. 8). Als interessantes Detail der awarischen Besiedlungsgeschichte ergibt sich dabei, daß in der letzten Phase eine verstärkte Besiedlung an der Theiß um Szeged und Szentés erfolgte, daß also dort nach den Awarenkriegen Karls des Großen offensichtlich ein Rückzugsgebiet verblieb.

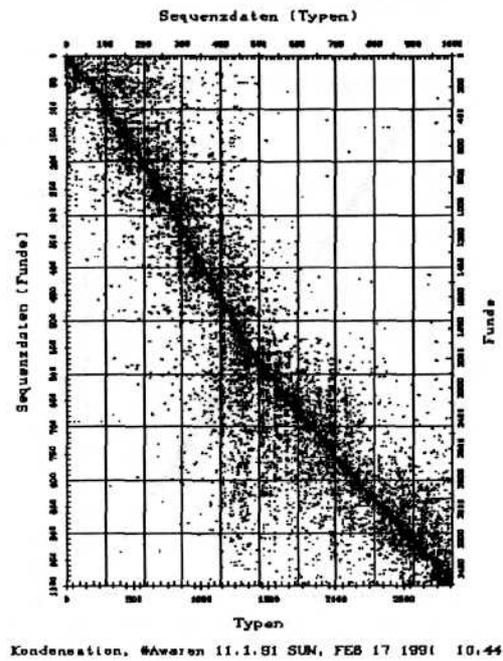


Abbildung 4- Kondensation

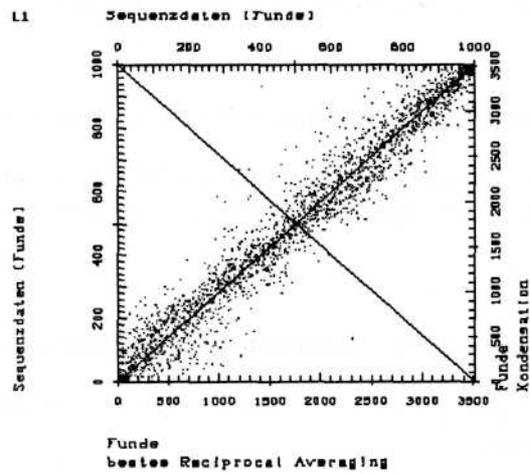


Abbildung 5: Vergleich von Reciprocal Averaging mit Kondensation, bezüglich der Funde

5.5 Intersite Spatial Analysis

Die geografische Verbreitung awarischer Typen muß dahingehend untersucht werden, ob sie zufällig ist oder nicht. Erst dieser Test ermöglicht es, Aussagen über Herkunft aus

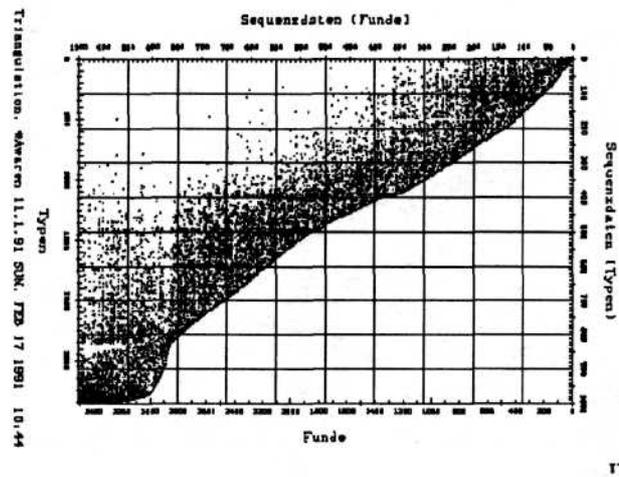


Abbildung 6: Triangulation

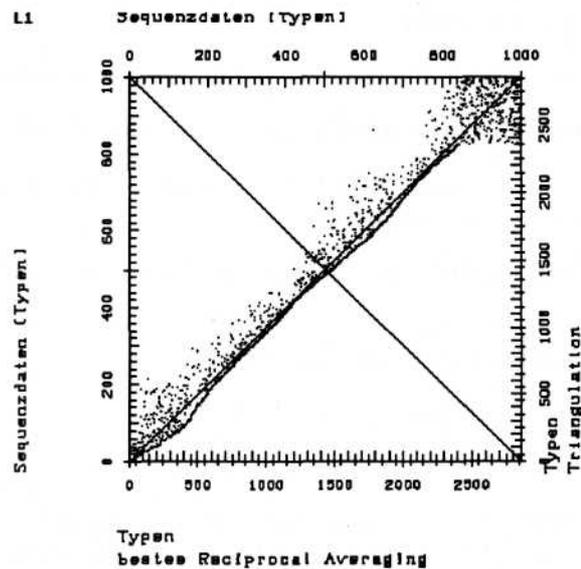


Abbildung 7: Vergleich von Reciprocal Averaging mit Triangulation, bezüglich der Typen

lokalen Werkstätten zu treffen. Derzeit liegen fast 3000 Verbreitungskarten vor, an ihrer Auswertung wird gearbeitet.

5.6 Intrasite Spatial Analysis

Auswertung des frühbronzezeitlichen Gräberfeldes von Franzhausen archäologisch und anthropologisch:

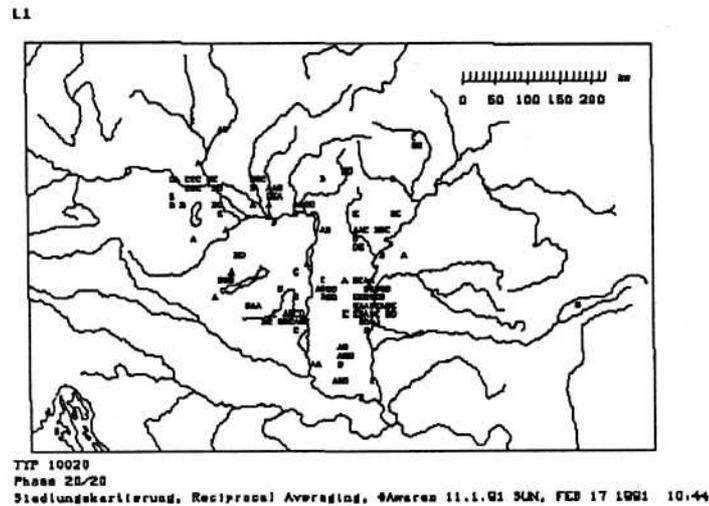


Abbildung 8: Letzte von 20 Phasen der Awarischen Siedlungsgeschichte

1. archäologische Typenkartierung.
2. Kartierung epigenetischer Merkmale bzw. metrischer Klassifikationsgruppen.
3. Untersuchung von Merkmalskombinationen von mehr als 250 Merkmalen.
4. Ausscheidung der Verteilungen auf dem Gräberfeldareal, die nach einem hypergeometrischen Verteilungsmodell zufällig sind.
5. Erstellung einer Datenmatrix aus "Beziehungen" (zwischen den Individuen) und Individuen.
6. Reciprocal Averaging mit dieser Kontingenztafel.
7. Clusteranalyse der Eigen Vektoren, (mit CLUSTAN)
8. Kartierung der Verwandtschaftsgruppen im Gräberfeldplan.

Die Interpretation des Resultates der sogenannten "Familiengruppen" ist momentan noch nicht möglich, da noch eine detaillierte Auswertung des archäologischen Materials aussteht. Möglicherweise wurden hier nicht die Familien erfaßt, sondern einfach die zeitliche Veränderung des genetischen Pools. Sollte das zutreffen, so könnte man also mithilfe der Anthropologie ebenfalls eine Chronologie erstellen.

5.7 Simulation des Reciprocal Averaging mit dem awarischen Datensatz

Es ist mit SERION möglich, die einzelnen Schritte eines Seriationsvorganges - ausgehend von einer Zufallsanordnung - darzustellen. Diese Einzelschritte des Ordnungsprozesses können auf einem PC in einer Präsentation schnell hintereinander dargestellt werden. Somit kann dieser Vorgang besser verständlich gemacht werden.

Literatur

- FLINDERS P.W.M. (1899), Sequences in prehistoric remains, *Journal of the Anthropological Institute* 29, 295-301.
- GOLDMANN K. (1980), *Die Seriation chronologischer Leitfunde der Bronzezeit*, Berliner Beiträge zur Vorgeschichte 1.
- IHM P. (1983), Korrespondenzanalyse und Seriation, *Archäologische Informationen*, 6/1, 8-21.
- LEGOUX R. (1980), Le recours à l'informatique: La Chronologie relative par permutation matricielle automatiques, in: *La Datation des tombes Mérovingiennes*, Hrsg. P. Perin, *Hautes Etudes Médiévales et modernes* 39, 138-155.
- STADLER, P. (1984), Überlegungen zur computerunterstützten Seriation, in: *Das awarische Gräberfeld von Sommerein am Leithagebirge*, Hrsg. F. Daim und A. Lippert, *NÖ, Studien zur Archäologie der Awaren I*, 171-179.
- STADLER, P. (1985a), Die Riemenzunge aus dem slawischen Hügelgrab 36 von Wimm, *Archaeologia Austriaca* 68, 227-233.
- STADLER, P. (1985b), *Die Seriation awarischer Gürtelgarnituren*, Masch. Diss. Wien.
- STADLER, P. (1985a), Seriation awarischer Gürtelgarnituren aus Nove Zámky und Zelovce, in: *Die Bayern und ihre Nachbarn*, Hrsg. F. Daim und H. Friesinger, *Österr. Ak. Wiss. Veröff. Komm. f. Frühmittelalterfor.* 9, 1, 27-132.
- STADLER, P. (1985b), Was ist "Spatial Analysis" in der Archäologie? *Mitteilungen der Anthropologischen Gesellschaft in Wien CXV*, 163-168.
- STADLER, P. (1986), Ausgewählte awarische Bronzegüsse als Parallelen zu Gürtelbeschlägen von Vrap und Erseke, in: *Der Schatzfund von Vrap in Albanien*, Hrsg. J. Werner, *Studien zur Archäologie der Awaren* 2, 105-118.
- STADLER, P. (1987a), Möglichkeiten statistischer Untersuchungen im Vergleich Archäologie, Anthropologie und Zoologie, in: *Die Bronzezeit im Osten Österreichs*, Hrsg. J.-W. Neugebauer, 95-101.
- STADLER, P. (1987b), Mathematical and Archaeological investigation of seriation, Paper für den World Archaeological Congress, *Section Data Management and Mathematical Methods in Archaeology*, Mainz 1987, im Druck.
- STADLER, P. (1987c), Statistische Auswertung verschiedener Befundparameter des Gräberfeldes von Gemeinlebarn, in: *Die Nekropole F vom Gemeinlebarn*, Hrsg. J.-W. Neugebauer, *NÖ, Römisch- Germanische Forschungen* 47, im Druck.
- STADLER, P. (1988a), Quantitative Methoden in der Archäologie, *Historikum* 12, 32-35.
- STADLER, P. (1988b), *SERION, a program package for the chronological evaluation of archaeological data*, Manual, Vienna, EARN-distribution, English edition.
- STADLER, P. (1988c), *SERION, ein Programmpaket zur chronologischen Auswertung archäologischer Daten*, Manual, Wien im Eigenverlag, EARN-Versand, Deutsche Ausgabe,
- STADLER, P. (1988d), Statistische Auswertung der Analysenergebnisse awarischer Bronzebeschläge aus Leobersdorf, in: *Das awarische Gräberfeld v. Leobersdorf*, Hrsg. F. Daim, *Studien zur Archäologie der Awaren* 3/2, 57-73.
- STADLER, P. (1989a), *Methodos quantitativos en la arqueologia*, *A Distancia*, Junio, 82-87.
- STADLER, P. (1989b), Nuovi argomenti per la datazione della tomba principesca di Bocsa, In Ist van Bóna, *Die Geschichte der Awaren im Lichte der archäologischen Quellen*, *Settimane di studio del Centro italiano di studi sull'alto medioevo* XXXV, 465-476.

STADLER, P. (1990a), Die Werkstätten awarischer Riemenbeschläge mit Greifendarstellung, Typen der Ethnogenese unter besonderer Berücksichtigung der Bayern, Veröff. *Komm. f. Frühmittelalterforschung* 13, 305-350.

STADLER, P. (1990b), Kombinierte Auswertung archäologischer und anthropologischer Daten eines Gräberfeldes, *Archäologische Informationen* 12/2, 249-250.

STADLER, P. (1990c), *La Chronologie de l'armement des Avars du VIe au VIIe siècle*, Paris, im Druck.

STADLER, P. (1991), Argumente für die Echtheit des "Avar Treasure", *Mitteilungen der Anthropologischen Gesellschaft in Wien, Festschrift Angeli*, 118/119, 193-217.

WILKINSON E.M. (1974), Techniques of Data Analysis - Seriation Theory, *Archaeo-Physika* 5, 1-142.