

Datenstrukturen zur Analyse archäologischer Schichten

Irmela Herzog

Zusammenfassung – Die Harris-Matrix ist eine bewährte Methode, um archäologische Schichten zu analysieren. Dabei ist es das Ziel, die relative chronologische Abfolge der Schichten zu rekonstruieren. Mathematisch gesehen handelt es sich bei den Schichten mit den zugehörigen stratigraphischen Beziehungen um eine Halbordnung. Das bedeutet für den Archäologen, daß alle zusätzlich vorhandenen Datierungsinformationen berücksichtigt werden sollten, um die chronologische Reihenfolge der Schichten so genau wie möglich zu rekonstruieren. In diesem Aufsatz werden Datenstrukturen für die stratigraphischen Daten und die Zusatzinformationen diskutiert. Verschiedene Ansätze werden vorgestellt, die nicht-stratigraphische Daten beim Layout des Harris-Diagramms berücksichtigen.

Schlüsselwörter – Harris Matrix, Graph-Layout, Datenstrukturen, monotone Regression

Abstract – The Harris matrix is a well established method to analyse archaeological stratigraphy, the aim is to reconstruct the relative chronology of the stratigraphic units. Mathematically speaking, the contexts and their stratigraphic relationships form a partially ordered set, that means for the archaeologist that all available additional dating information should be taken into account in order to reconstruct the chronological sequence of the contexts as exactly as possible. This paper discusses data structures for dealing with stratigraphic data including the additional information which might be available. Several approaches to use non-stratigraphic chronological data in Harris diagram layout are presented.

Keywords – Harris Matrix, automatic graph drawing, data structures, monotone regression

Einführung

In der Archäologie ist die eingehende chronologische Untersuchung die Grundlage von allen weitergehenden Analysen. Das bedeutet bei Stadtkerngrabungen, daß der Dokumentation und Auswertung von archäologischen Schichten eine zentrale Rolle zukommt. Dazu hat sich seit Jahren die sogenannte Harris-Matrix bewährt (HARRIS 1989).

Nun ist es relativ schwierig, bei mehreren hundert oder gar mehreren tausend Schichten die Harris-Matrix von Hand zu zeichnen. Deshalb gab es schon 1976 (BISHOP & WILCOCK 1976) erste Versuche, das Diagramm mit Hilfe des Computers zu erstellen. Inzwischen ist viel Zeit vergangen, und man könnte meinen, das Problem sei gelöst. Doch dieser Aufsatz wird zeigen, daß es noch einiges zu tun gibt.

In der zweiten Hälfte der 80er Jahre wurde erkannt, daß man die Harris-Matrix mathematisch weniger mit Konzepten der Matrixrechnung als mit denen der Graphentheorie in den Griff bekommen kann. Denn mathematisch gesehen bilden die Schichten und die „ist später als“-Beziehungen eine Halbordnung, die als Graph dargestellt wird. Mit der zunehmenden Verbreitung von grafikfähigen Rechnern wurde auch das Zeichnen von Graphen zu einem immer populärerem Forschungsgebiet bei Informatikern. Seit 1992 gibt es alljährlich eine internationale Tagung mit dem Titel „Graph Drawing“.

Dabei ist die Harris-Matrix ein Graph mit speziellen Eigenschaften, die beim Layout berücksichtigt werden sollten (d.h. gerichtet, orthogonal). Für solche Graphen wurde 1995 auf dem „Graph Drawing“-Symposium in Passau ein Layout-Verfahren vorgestellt (SANDER 1996), das ich in meinem Programm Stratify (www.stratify.privat.t-online.de) mit ein paar Abweichungen implementiert habe.

Speicherung der Daten

Schon bei meinem ersten Harris-Matrix-Programm (BRIDGER & HERZOG 1991) hatte ich anfänglich den Versuch unternommen, die stratigraphischen Daten in einem Datenbankformat zu speichern. Der Vorteil ist, daß damit die Datenverwaltung erheblich vereinfacht wird: Die Daten werden automatisch auf der Festplatte gesichert, Sortieren und Suchen müssen nicht erst mühsam programmiert werden, Filter-Funktionen erlauben es, nur auf die gerade benötigten Daten zuzugreifen. 1990 scheiterte die Einbindung des damals zur Verfügung stehenden Datenbank-Tools daran, daß sich dieses als unzuverlässig erwies. Mehr als zehn Jahre später sah die Situation anders aus: Mit Delphi steht eine Entwicklungsumgebung zur Verfügung, in der auch Daten im Paradox-Datenbank-Format abgespeichert werden können. Mit diesem Format habe ich mehr als zehn Jahre Erfahrung im Rheinischen Amt für Bodendenkmalpflege: Es kommt selten zu Problemen, wenn man sicherstellt, daß man das Datenbankprogramm grundsätzlich ordnungsgemäß verläßt. Da dies nicht immer gewährleistet werden

kann (z.B. bei Stromausfall), hat es sich bewährt, wenn die Daten zunächst in einer Eingabe-Datenbank erfaßt und erst nach Abschluß der Plausibilitäts- und Dublettenkontrolle in die Hauptdatenbank übertragen werden. Dann ist die Hauptdatenbank nur sehr kurze Zeit im Zugriff; eine fehlerhafte Eingabe-Datenbank kann ohne Datenverlust schnell ersetzt werden. Diese bewährte Strategie verwendete ich auch bei Stratify.

Die Datenerfassung ließ sich so relativ einfach programmieren. Basis bilden zwei Tabellen, und zwar für die archäologischen Schichten und die Beziehungen. Hinzu kommt eine optionale Tabelle für Begriffe. Damit wird die Verwendung einer einheitlichen Terminologie unterstützt, neben dem Begriff selbst kann ein Kürzel und eine Erklärung bzw. Definition gespeichert werden. Bei der Dateneingabe dient diese Tabelle als Nachschlage-liste, wobei dann nur die Begriffe gezeigt werden, die für das aktuelle Feld definiert wurden. Gibt man nur das Kürzel ein, wird dieses beim Verlassen des Feldes durch den vollständigen Begriff ersetzt. Neben dem Kürzel, dem Begriff und der Erklärung kann in der Begriffstabelle auch die gewünschte Farbe und der Linienstil für den Rand des entsprechenden Kästchens im Harris-Diagramm eingetragen werden. Diese Stileinstellungen dienen als Vorgaben bei der sogenannten thematischen Darstellung des Diagramms, bei der aufgrund der Werte in einem Feld der Schichten-Datenbank den Kästchen entsprechende Farben und Randstile zugewiesen werden.

Auch bei den Abfragen zur Plausibilitätskontrolle leistet die Datenbankstruktur gute Dienste: Man kann so recht einfach eine Liste erstellen, die alle Schichten mit einer vorgegebenen Befundansprache oder Datierung zeigt. Auch die Kombination von mehreren Auswahlkriterien ist möglich.

Schließlich war auch das Layout-Verfahren von Sander fertig und funktionierte. Doch die Enttäuschung war groß: Die ersten Layouts etwas größerer Datensätze wurden nur quälend langsam berechnet. In der Tuning-Phase kehrte ich daher zu den früher genutzten Datenstrukturen zurück: Verkettete Listen bzw. Arrays variabler Länge. Die meiste Zeit wird nun weniger für das Layout selbst, sondern für das Auslesen der Datenbanken und das Speichern der Rechenergebnisse in den Datenbanktabellen benötigt. Aber noch sind nicht alle Layoutschritte auf die schnellere Datenstruktur umgestellt, dies bleibt eine Aufgabe für die Zukunft.

Layout fertig – alles fertig?

Ist damit die Aufgabe gelöst, die Harris-Matrix PC-gestützt zu zeichnen? Ja und nein. Ja, das Ergebnis ist

eine graphische Darstellung. Nein, das ist aber noch nicht das Ziel. Zur Erinnerung: Die Harris-Matrix wird gezeichnet, um die relative chronologische Abfolge der archäologischen Schichten einer Grabung zu ermitteln. Schon früh wurde das Problem der sogenannten „floating sequences“ erkannt, d.h. von Schichten, die sich aufgrund ihrer stratigraphischen Beziehungen allein nicht eindeutig chronologisch einordnen lassen (HARRIS 1984, 128). Ein Beispiel ist Schicht 035 in Abb. 1, die aufgrund der stratigraphischen Beziehungen allein einer beliebigen Phase zugeordnet werden könnte. Das Problem ist mathematisch gesehen darin begründet, daß das stratigraphische Beziehungsgeflecht nur eine Halbordnung darstellt, gewünscht wird jedoch eine totale Ordnung, wie sie z.B. die Seriation liefert.

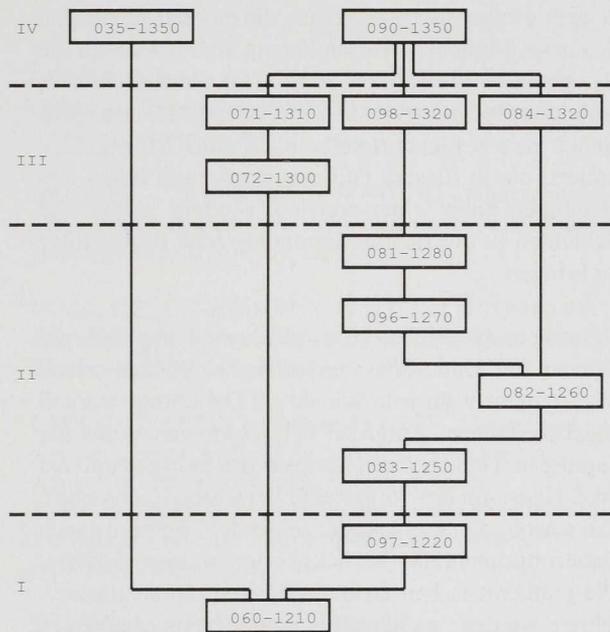
Welche Schritte unternehmen Archäologen normalerweise, um die archäologischen Schichten in einer Harris-Matrix in die richtige Zeitstufe einzuordnen? Neben den stratigraphischen Beziehungen gibt es je nach Grabung unterschiedliche Informationen, die bei der chronologischen Einstufung verwendet werden können: Aus der Struktur der Befunde ergeben sich oft Gleichzeitigkeitsbeziehungen, beispielsweise kann man davon ausgehen, daß die Pfostenlöcher für ein Gebäude annähernd gleichzeitig entstanden sind. Weiter erscheint es sinnvoll, Befunde aus einem Befundkomplex im Harris-Diagramm als eine Gruppe darzustellen, umrandet von einem Rechteck, ein Beispiel sind die Schichten einer Grubenverfüllung. Solche Gruppen können in mehreren Stufen verschachtelt sein (siehe z. B. ROSKAMS 2001, 257-261). Ein Verfahren zur Darstellung dieser Gruppen wurde bereits vorgestellt (HERZOG 2004) und ist daher nicht mehr Thema dieses Aufsatzes.

Weitere Informationen liefern naturwissenschaftliche Verfahren zur absoluten Datierung, z.B. Dendrochronologie oder ¹⁴C-Analysen. Hinzu kommen die eingemessenen Höhenwerte und – am häufigsten vorhanden – die archäologischen Funde.

Mit der Kombination von stratigraphischen Beziehungen und den Ergebnissen von ¹⁴C-Analysen beschäftigen sich BUCK, CAVANAGH & LITTON (1996). Dabei kommen Methoden der Bayes-Statistik zum Einsatz. Beim Kalibrierungsprozess mit stratigraphischen Randbedingungen erhält man im Ergebnis deutlich kleinere Datierungsintervalle. Nachteil dieses Verfahrens ist, daß zum einen meist nicht besonders viele ¹⁴C-Daten von einer Grabung vorliegen, zum anderen, daß die Rechenmethode recht aufwendig ist.

Eine Sonderstellung unter den Funden nehmen Teile von Zusammenpassungen ein. Dabei handelt es sich um Keramik-, Knochen- oder Steinfragmente, die

Abb. 1
Kleines Beispiel für eine Harris-Matrix mit Jahreszahlen hinter den Schichtnummern, außerdem mit Phasen.



man wieder so zusammensetzen kann, daß der Nachweis erbracht ist, daß sie von einem Ausgangsobjekt stammen. In der Regel wird man die zugehörigen Schichten als gleichzeitig ansehen. Bei Steinzusammenpassungen kann man berücksichtigen, daß die äußersten Objekte zuerst und die inneren zuletzt abgetrennt wurden, woraus sich nicht-stratigraphische „ist später als“-Beziehungen ergeben.

Berücksichtigung der Funddatierungen

Mit Hilfe der archäologischen Funde kann man versuchen, jeder archäologischen Schicht eine Phase zuzuordnen. Damit wird der chronologische Spielraum der Schichten eingeschränkt, es ist also ein Schritt in Richtung von der Halbordnung zur totalen Ordnung. Phasengrenzen werden normalerweise in dem Harris-Diagramm als gestrichelte horizontale Linien dargestellt (siehe Abb. 1). Wenn für jede Schicht eine Phase angegeben ist und diese Zuordnungen nicht den stratigraphischen Beziehungen widersprechen, kann Stratify diese Phasengliederung berücksichtigen und im Diagramm einzeichnen.

Die Frage ist, wie man mit Hilfe der Funde zu einer Phasengliederung kommen kann. Dabei ist zu berücksichtigen, daß man archäologische Schichten in der Regel nicht als geschlossene Fundkomplexe ansehen kann. Wichtig ist hier, sekundär verlagerte Funde zu erkennen und diese von der chronologischen Analyse auszuschließen. Schließlich bleibt das Problem der fundleeren Schichten: Im Idealfall ist es so, daß solche Schichten von Schichten einer einzigen Phase

geklammert werden, d.h. es gibt eine sicher ältere und eine jüngere Schicht, die zu ein und derselben Phase gehören. Dann kann man die fundleere Schicht genau dieser Phase zuordnen.

Im einfachsten Fall kommt man mit dem Leitfund-Konzept zu einer Phasengliederung: Gibt es für jede Phase einen oder mehrere Leitfundtypen, können die Phasen danach getrennt werden. Probleme bereiten hier fließende Übergänge und die Tatsache, daß nur ein Bruchteil der Funde als Leitfunde Berücksichtigung finden können, alle anderen Funde werden ignoriert.

Wenn man aufgrund der Funde Datierungen schätzen kann, sind Methoden wie *terminus post quem*, *terminus ante quem* und *mean ceramics date* gebräuchlich. Doch wie kann man vorgehen, um diese Datierungsinformationen nutzen, wenn man dabei den stratigraphischen Beziehungen Vorrang einräumt?

Wie ein glücklicher Zufall es wollte, hielt Prof. Dr. Dr. Jürgen Hansohm von der Universität der Bundeswehr München auf der Tagung der Gesellschaft für Klassifikation in Dortmund einen Vortrag zum Thema „Monotone Regression on Partially Preordered Sets“. Mit den dort dargestellten Methoden scheint es möglich, die oben genannte Aufgabe zu lösen: Prof. Hansohm beschäftigt sich mit Halbordnungen und Funktionswerten (also in unserem Beispiel Jahreszahlen). Wenn es in der Halbordnung keine Widersprüche zu den Funktionswerten gibt, ist es ja ein Leichtes, die untersuchten Objekte (in unserem Fall die Schichten) nach den Funktionswerten zu ordnen. Gibt es jedoch Widersprüche, so ist das Ziel der monotonen Regression,

neue Funktionswerte zu finden, die möglichst nahe bei den ursprünglichen Werten liegen, aber bei denen die Beziehungen der Halbordnung nicht verletzt werden. Abbildung 1 zeigt ein kleines künstliches Beispiel: Neben den Schichtbezeichnungen sind Jahreszahlen notiert, die in diesem Fall den stratigraphischen Beziehungen nicht widersprechen, sondern helfen, die Schichten in die richtige chronologische Reihenfolge zu bringen.

Nun ist es plausibler, statt mit einer Jahreszahl pro Schicht mit Datierungsintervallen zu arbeiten. Auch wird man nicht für jede Schicht ein Datierungsintervall angeben können. Aufgrund der Diskussionen auf der Tagung und einem regen Mailverkehr im Anschluß hat Prof. Hansohm das vorgestellte Verfahren so erweitert, daß sowohl das Problem der fehlenden Werte als auch Datierungsintervalle berücksichtigt werden können. Die mathematischen Grundlagen und das Rechenverfahren werden voraussichtlich im Tagungsband der Dortmunder Konferenz publiziert. Probleme bereitet noch, daß bei einer naiven Umsetzung des Verfahrens mit zunehmender Zahl der Schichten der Rechenaufwand sehr stark ansteigt.

Mit dem Verfahren von Hansohm wird jeder Schicht ein Datierungsintervall zugewiesen, auch denjenigen, für die ursprünglich die Datierungswerte fehlten. Damit kann man nicht nur Phasen definieren, sondern beim Layout in jeder Zeile des Diagramms die Datierungen entsprechend berücksichtigen. Normalerweise verwendet man folgende Vorgehensweise, um die Schichten im Diagramm in Zeilen anzuordnen: In die oberste Zeile werden alle Schichten ohne Vorgänger eingetragen, d.h. die Schichten, für die es nur stratigraphische Beziehungen zu späteren Schichten gibt (In Abbildung 1 sind dies die Schichten 035 und 090). Die archäologischen Schichten, die unmittelbar älter sind als die Schichten der ersten Zeile, sind Kandidaten für die zweite Zeile (Kandidaten für die zweite Zeile sind in Abbildung 1 zunächst die folgenden Schichten: 060, 071, 098 und 084). Dabei wird aber nur dieser Kandidat in die zweite Zeile aufgenommen, bei dem alle Vorgänger bereits der ersten Zeile zugeordnet wurden. Allgemein gesagt, besteht die Kandidatenliste für Zeile n aus all den Nachfolgern von Schichten in Zeilen 1 bis $n-1$, die bis jetzt noch nicht berücksichtigt werden konnten. Nach einer Überprüfung der Vorgänger dieser Schichten wird entschieden, ob sie in die Zeile aufgenommen werden können. Die Idee mit der Kandidatenliste habe ich bei dem Phasenlayout weiter ausgebaut: Nur die Schichten werden aus der Kandidatenliste für die aktuelle Zeile gewählt, die auch zur gerade angesagten Phase gehören, erst wenn alle Schichten einer Phase zugeordnet sind, kommen die Kandidaten aus der nachfolgenden Phase zum Zuge. Dieses Konzept kann man schließlich so erweitern, daß die Datierungen der

Schichten einer Zeile möglichst wenig variieren: Schichten, die deutlich älter sind als der Durchschnitt der anderen Schichten in der Kandidatenliste, werden bei der aktuellen Zeile nicht berücksichtigt, sondern bleiben so lange in der Kandidatenliste, bis sie chronologisch in die Zeile passen.

Ein anderer interessanter Ansatz versucht, zusätzliche „ist älter als“-Beziehungen aus den Fundtypen herzuleiten. Wenn für zwei unterschiedliche Fundtypen gilt, daß die Schicht des ersten Fundtyps immer stratigraphisch älter ist als die des zweiten Fundtyps (falls es eine stratigraphische Beziehung zwischen den Schichten gibt), so liegt es nahe anzunehmen, daß der erste Fundtyp älter ist als der zweite. Gibt es nun zwei stratigraphisch unverbundene Schichten, die diese beiden Fundtypen enthalten, folgt mit der obigen Argumentation, daß diese Schichten entsprechend den Fundtypen chronologisch anzuordnen sind. Hierbei ist natürlich einige Vorsicht geboten, umgelagerte Funde müssen sicher erkannt und von dieser Betrachtung ausgeschlossen werden. Auch wäre es kaum zu verantworten, aufgrund eines Fundtyps, der nur einmal vorkommt, vielleicht mit nur einer einzigen Scherbe, solche Schlüsse zu ziehen. Schwieriger als die Anwendung des Verfahrens selbst ist deshalb vermutlich die Festlegung von Schwellenwerten: Ab welcher Anzahl, welchem anteiligen Gewicht etc. sollen Fundtypen in einer Schicht Berücksichtigung finden, wie groß muß die Anzahl der stratigraphischen Beziehungen sein, um die Übertragung der Beziehung auf die zugehörigen Fundtypen rechtfertigen zu können? Schließlich: Wie viele solcher Übertragungen sind überhaupt möglich, wenn man vernünftige Schwellenwerte wählt?

Funde-Datenbank

Die Datierungsintervalle werden in der Regel aufgrund der Funde geschätzt, oder genauer, aufgrund der den Funden zugeordneten Fundtypen. Deshalb liegt es nahe, in Stratify eine weitere Tabelle einzubinden, die die Funddaten speichert. Der Aufbau von mehreren, relational verknüpften Tabellen wäre auch möglich, ist aber mit mehr Aufwand verbunden, der in diesem Fall nicht gerechtfertigt erscheint.

Neben den Fundtypen sollte diese Tabelle Informationen zu Zusammenpassungen erfassen können. Eine Möglichkeit wäre, alle Objekte einer Zusammenpassung einem künstlichen Fundtyp zuzuordnen, somit verwendet man die gleichen Datenstrukturen für Fundtypen und Zusammenpassungen. Ein Nachteil wäre, daß dann ein Fund evtl. mehrfach eingegeben werden müßte, als Teil einer Zusammenpassung und mit seinem Fundtyp. Damit wäre die Gesamtzahl der Fundobjekte in einem Befund nicht mehr genau zu ermitteln. Ähnliches passiert auch, wenn

man die Fundtypen aufgrund von Verzerrungen definiert, aber ein Fundobjekt mehrere Verzerrungstypen aufweist. Die Frage wird deshalb hier zur Diskussion gestellt, ob es sinnvoll ist, Zusammenpassungen als Fundtypen zu betrachten. Alternativ könnte jeder Zusammenpassung eine Nummer zugeordnet und diese bei Bedarf im Funddatensatz eingetragen werden.

Bei der Eingabe von Fundtypen bietet sich die Nutzung der bereits oben erläuterten Begriffstabelle an: Diese Tabelle kann nicht nur als Nachschlagetabelle der Namen von Fundtypen dienen, sondern auch die zugehörige Beschreibung und das Datierungsintervall speichern.

Zusammenfassung und Ausblick

Beim Layout der sogenannten Harris-Matrix reichen die stratigraphischen Beziehungen normalerweise nicht aus, um eine eindeutige relativchronologische Reihenfolge zu erhalten. Zusätzliche Datierungsinformationen sind in der Regel vorhanden und sollten auch beim Layout des Harris-Diagramms berücksichtigt werden. Eine besonders wichtige Rolle spielen hier die Funde und ihre Datierungen. Bei der Einbeziehung dieser Zusatzinformationen ist die Methode der monotonen Regression von Hansohm ein erfolgversprechender Ansatz. Wenn Sie, liebe Leserin bzw. lieber Leser, entsprechende Daten haben, also eine größere Anzahl Schichten mit stratigraphischen Beziehungen und relativ gut datierbarem Fundmaterial, wäre ich sehr dankbar, wenn Sie sich mit mir in Verbindungen setzen würden. Denn ich würde gerne das oben kurz vorgestellte Verfahren anhand von Daten aus der Praxis testen.

Umgekehrt gibt es auch Datensätze, bei denen relativ wenige stratigraphische Beziehungen beobachtet wurden und das Fundmaterial bei der chronologischen Analyse im Vordergrund steht. In einer solchen Situation bietet sich das Verfahren von GROENEN & POBLOME (2003) an, das in einem anderen Aufsatz des vorliegenden Heftes vorgestellt wird. Vorteil ist hier, daß die Datierungen bzw. Datierungsintervalle der Funde nicht vorgegeben werden müssen, sondern für jeden Fundkomplex eine Jahreszahl aus dem Verfahren errechnet wird.

Literatur

- BISHOP, S. & J.D. WILCOCK (1976) Archaeological context sorting by computer: The strata program. *Science and Archaeology* 17, 3-12.
- BRIDGER, C. & I. HERZOG (1991) Die stratigraphische Methode und ein neues PC-Programm zur Erstellung der Harris-Matrix, *Arch. Korr.* 21, 133-144.
- BUCK, C.E., CAVANAGH, W.G. & C.D. LITTON (1996) Bayesian Approach to Interpreting Archaeological Data. Chichester 1996, insbesondere 190-192.
- GROENEN, P.J.F. & J. POBLOME (2003) Constrained Correspondence Analysis for Seriation in Archaeology Applied to Sagalassos Ceramic Tablewares. In: SCHWAIGER M. & O. OPITZ (eds.) *Exploratory Data Analysis in Empirical Research*. Berlin/Heidelberg/ New York 2003, 90-97.
- HARRIS, E.C. (1984) The analysis of multilinear stratigraphic sequences. *Scottish Arch. Review* 3 (2), 127-133.
- (1989) *Principles of Archaeological Stratigraphy*. London 1989.
- HERZOG, I. (2002) Possibilities of Analysing Stratigraphic Data. In: CD "Workshop 6: Archäologie und Computer". Wien 2002.
- (2004) Group and conquer – a method for displaying large stratigraphic data sets. In: *STADTARCHÄOLOGIE WIEN (ed.) Enter the Past. CAA 2003. B.A.R. Int. Ser. 1227*. Oxford 2003, 423-425.
- ROSKAMS, S. (2001) *Excavation. Cambridge Manuals in Archaeology*. Cambridge 2001.
- SANDER, G. (1996) A Fast Heuristic for Hierarchical Manhattan Layout. In: BRANDENBURG, F. J. (ed.) *Graph Drawing. Symposium on Graph Drawing, GD '95, Passau, Germany September 1995, Proceedings*. Berlin/Heidelberg/New York 1996, 447-458.
- Das Programm Stratify kostenlos zum Download:
<http://www.stratify.privat.t-online.de>
- Irmela Herzog
Landschaftsverband Rheinland
Rheinisches Amt für Bodendenkmalpflege
Endenicher Str. 133
D-53115 Bonn
i.herzog@lvr.de