



Mag. Dr. Armin Heller, studierte zunächst in Tübingen und ist als Assistent am Institut für Geographie in Innsbruck zuständig für das GIS-Labor. Als diplomierter Mathematiker findet er ungewöhnliche Zugänge zu seiner neuen Wahlheimat.

Der Vergleich von Raumgliederungsmodellen mit Klassifikationsbäumen am Beispiel der Stadtregion Innsbruck

von Armin Heller

Einführung

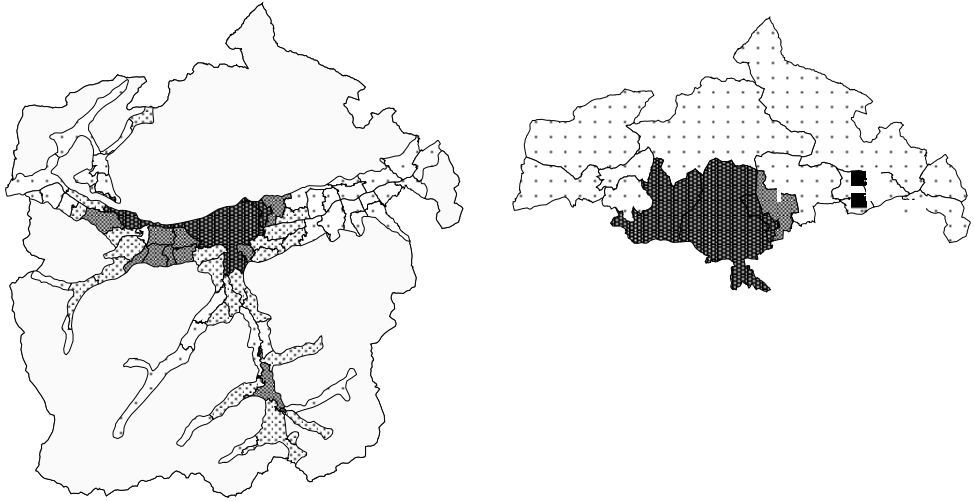
Raumtypisierungen und Regionalisierungen werden in der Geographie schon seit langem in fast allen Bereichen eingesetzt. Bodentypen, Städtetypen, Landschaftstypen, Klimatypen etc. haben stets das Ziel, die jeweils spezifische Beobachtungswelt zu gliedern und zu strukturieren, um sie dadurch letztendlich mit all ihren relevanten Einflussfaktoren besser zu verstehen. Die Abgrenzung räumlicher Einheiten mit Hilfe verschiedener Raumgliederungsmodelle führt fast immer zu unterschiedlichen Ergebnissen. Damit stellt sich die Frage wie Raumgliederungsmodelle verglichen werden können.

Im Rahmen des Forschungsprojektes „Virtual Project - Urbanisation as a Geographical Effective Structure Between Conurbations and Rural Countryside“ (<http://business.carinthia.com/bgolob/virp/>) wurden in einer Pilotstudie für den Raum Innsbruck und Umgebung insgesamt drei verschiedene deduktive Raumgliederungsmodelle der Stadtregion, jeweils für die Jahre 1971, 1981 und 1991, eingesetzt. Für eine ausführliche Diskussion der Modelle und inhaltliche Interpretation der Ergebnisse wird auf Borsdorf (2000) verwiesen. Im folgenden Beitrag soll gezeigt werden, wie diese Raumgliederungsverfahren ohne wesentliche Beeinflussung der Ergebnisse mit Hilfe von Klassifikationsbäumen vereinfacht werden können, wie die Wichtigkeit der verschiedenen Eingangsparameter für das jeweilige Ergebnis beurteilt werden kann und wie damit das gesamte Modell bewertet werden kann.

Klassifikation nach Boustedt/Nellner

Das Raumgliederungsmodell nach Boustedt für die Abgrenzung von Stadtregionen 1950/1961 und die Modifikation von Nellner für das Jahr 1970 der Bundesrepublik Deutschland klassifiziert - ausgehend von der Kernstadt - insgesamt vier Zonen: Kerngebiet (1) – Verstädterte Zone (2) – Randzone (3) und Umland (4). Dabei werden insgesamt 5 Variablen betrachtet: Bevölkerungsdichte (BD), Einwohner-Arbeitsplatz-Dichte (EAD), Agrarquote (AQ), Anteil der in das Kerngebiet auspendelnden Erwerbspersonen an den Auspendlern insgesamt (PP) und Anteil der in das Kerngebiet aus-

pendelnden Erwerbspersonen an den Erwerbspersonen insgesamt (PE). Mit den bei Boustedt festgelegten Schwellenwerten kann schließlich jede der 69 Gemeinde im Bereich Innsbruck und Umgebung einer Zone zugeordnet werden (siehe Karte 1). Dieses Raummodell, das vorrangig in der Bundesrepublik Deutschland eingesetzt wird, wurde im Rahmen des „Virtual Project - Urbanisation as a Geographical Effective Structure Between Conurbations and Rural Countryside“ unverändert für Innsbruck übernommen. (Boustedt 1953, 1970, 1975, Nellner 1975)



Zonen nach Boustedt/Nellner:

-  1 = Kerngebiet
-  2 = verstärkte Zone
-  3 = Randzone
-  4 = Umland

Datengrundlagen: Völkzählung 1971, 1981 und 1991

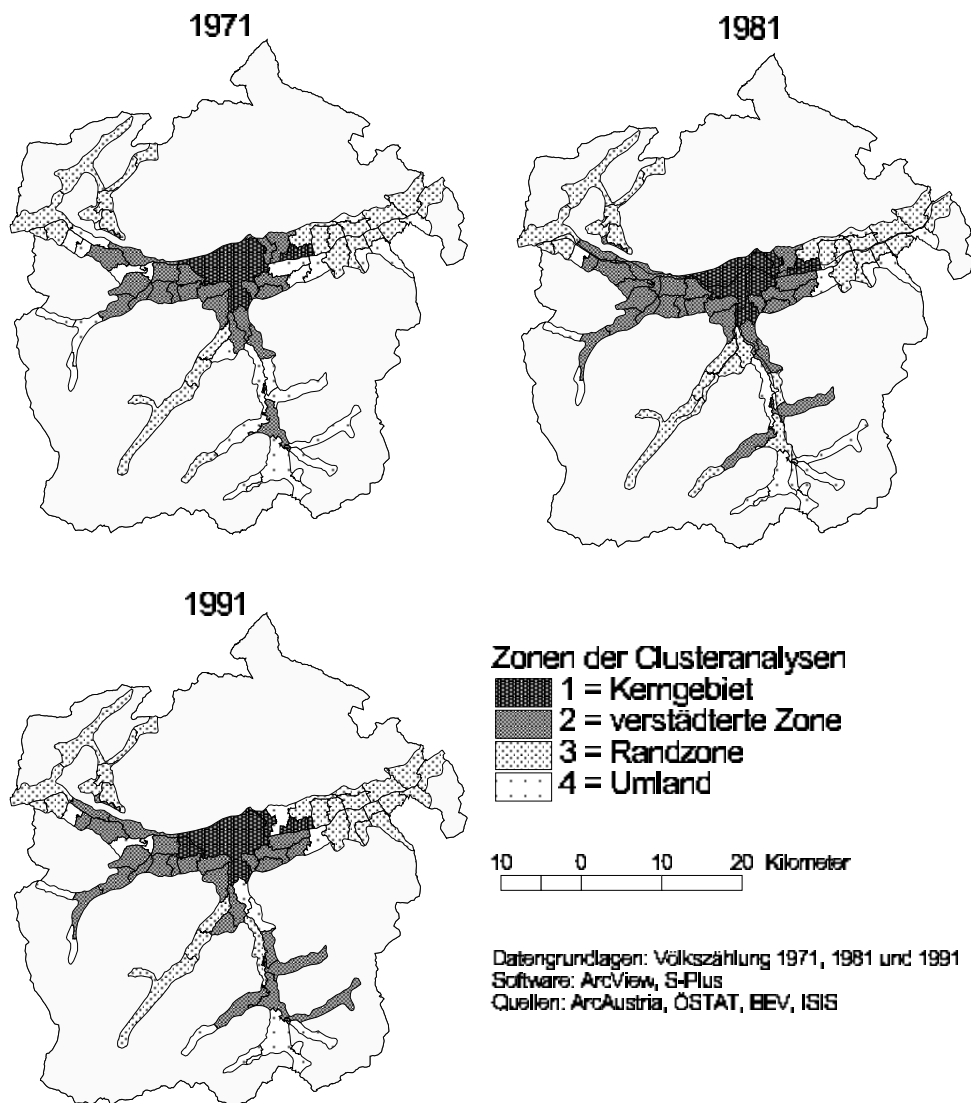
Software: ArcView, S-Plus

Quellen: ArcAustria, ÖSTAT, BEV, ISIS

Karte 1: Klassifikation nach Boustedt/Nellner für die Jahre 1971, 1981 und 1991

Automatische Klassifikation

Bei dieser Methode wurden die Gemeinden im Raum Innsbruck und Umgebung als räumliche, unabhängige Einheiten betrachtet. Sie wurden mit denselben Variablen wie bei Boustedt/Nellner charakterisiert, um eine größtmögliche Vergleichbarkeit der Ergebnisse mit diesem Modell zu gewährleisten. Mit Methoden der automatischen Klas-

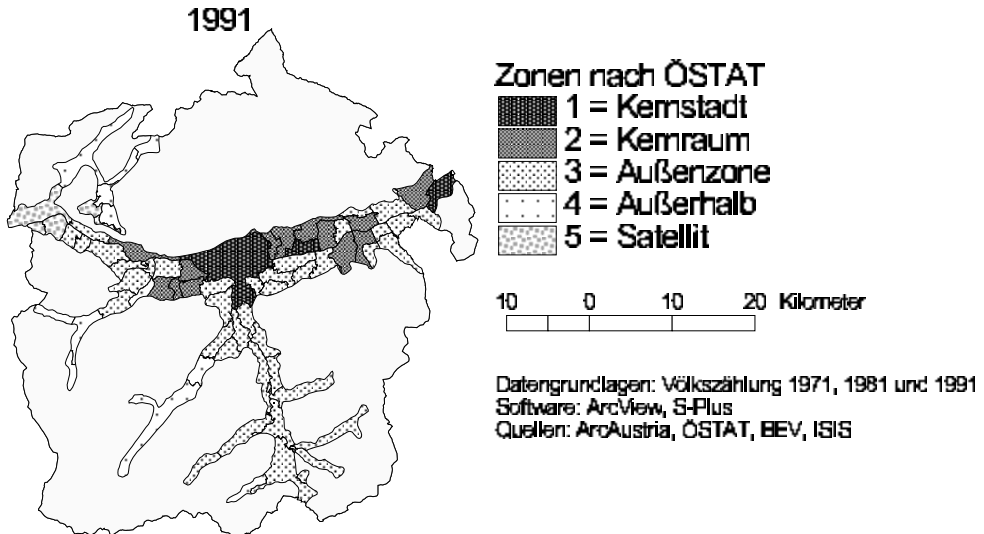


Karte 2: Automatische Klassifikation für die Jahre 1971, 1981 und 1991

sifikation (Clusteranalysen) wurden dann die Gemeinden in vier Gruppen gegliedert (siehe Karte 2). Diese multivariate statistische Methode geht nach folgendem Grundprinzip vor: Gemeinden innerhalb einer Gruppe sollen sich möglichst ähnlich sein – die verschiedenen Gruppen untereinander möglichst unähnlich. Da verschiedenste Algorithmen nahezu identische Ergebnisse lieferten, kann davon ausgegangen werden, dass die gefundene Gruppierung (Complete Linkage Methode) die strukturellen Eigenschaften des Datensatzes wiedergibt. (Heller 1997, 1999)

Die Stadtregionen nach ÖSTAT

Die Raumgliederung des Österreichischen Statistischen Zentralamtes (ÖSTAT) zur Abgrenzung von Stadtregionen 1991 ist eine Überarbeitung der Konzepte aus den Jahren 1971 und 1981. Kleinste räumliche Einheit bilden die auf Gebäudeebene abgegrenzten Siedlungseinheiten. Mit einem ähnlichen Gravitationskonzept wie in der Bundesrepublik Deutschland, aber etwas anders definierten Parametern und Schwellenwerten wurden für 1991 die amtlichen Regionen festgelegt, die als Regionalkriterium für alle Datenabfragen der Ergebnisse der Volkszählungen in Österreich auf Gemeindeebene herangezogen werden können (siehe Karte 3). Dieses Raumgliederungsmodell wurde für den Raum Innsbruck und Umgebung unverändert übernommen, die Anzahl und Bezeichnungen der Raumeinheiten weichen allerdings etwas von Boustedt/Nellner ab: Kernstadt (1) – Kernraum ohne Kernstadt (2) – Außenzone (3) – Außerhalb der Stadtregion (4) – Satelliten (5). (Fuchs 1997)



Karte 3: Klassifikation nach ÖSTAT für das Jahr 1991

Klassifikationsbäume

Die Methode der Klassifikationsbäume geht davon aus, dass zu den betrachteten Objekten eine Klassifikation bereits vorliegt. Wie diese Klassifikation zustande gekommen ist, ist völlig unerheblich. Mit Hilfe der gegebenen Variablen und der bekannten Klassifikation wird unter allen möglichen Partitionen diejenige algorithmisch gesucht, die der gegebenen Gruppierung am nächsten kommt und gleichzeitig mit einem binären Entscheidungsbaum beschrieben werden kann. Wird ein solcher Entscheidungsbaum gefunden, können gegebene Klassifikationen damit stark vereinfacht werden. Zusätzlich werden für die einzelnen Variablen statistische Kenngrößen berechnet mit deren Hilfe dann beurteilt werden kann wie wichtig die jeweilige Variable für die Klassifikation ist. Mit dieser statistischen Methode wurden im folgenden die oben beschriebenen Raumgliederungsmodelle untersucht. Für eine detaillierte mathematische Darstellung und algorithmische Umsetzung dieser Methode wird auf folgende Literatur verwiesen: Breiman (1984), Statistical Science (1997), Steinberg (1998), Clark (1993).

Ergebnisse

Im folgenden werden die Klassifikationsbäume vorgestellt, die aus den oben beschriebenen Ergebnissen der verschiedenen Raumordnungsmodelle berechnet wurden. Die Auswertungen beziehen sich alle auf die statistischen Kenngrößen aus dem Jahr 1991.

Als erstes stellt sich die Frage nach der Genauigkeit, mit der die Klassifikationsbäume die ursprünglichen Gruppierungen wiedergeben. Im Falle des Modells von Boustedt/Nellner werden insgesamt 67 von 69 Gemeinden (97.10 %) völlig identisch klassifiziert (siehe Abb. 1).

Actual Class	Total Cases	Percent Correct	Class 1 8	Class 2 12	Class 3 10	Class 4 39
1	6	100,000	6	0	0	0
2	13	92,308	1	12	0	0
3	11	90,909	1	0	10	0
4	39	100,000	0	0	0	39

Abb. 1: Gleichklassifikation Modell Boustedt/Nellner

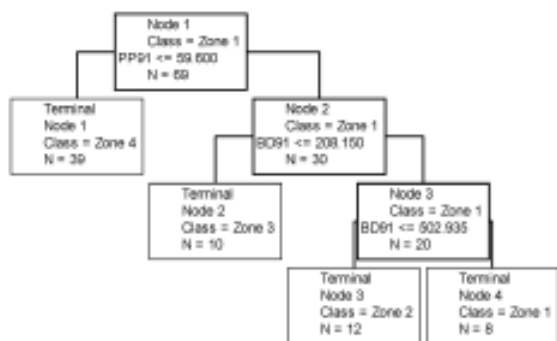


Abb. 2:
Klassifikationsbaum
- Modell Boustedt/Nellner

In der Zone 1 (Kerngebiet) und Zone 4 (Umland) beträgt die Gleichklassifikation sogar 100 % - nur in Zone 2 (verstädtertes Gebiet) und Zone 3 (Randzone) wird jeweils eine Gemeinde anders gruppiert. Der entsprechende Klassifikationsbaum besteht aus genau vier Endknoten (siehe Abb.2):

Zunächst werden alle Gemeinden mit einem Wert der Variablen PP unter 59.600 % (Anteil der in das Kerngebiet auspendelnden Erwerbspersonen an den Auspendlern insgesamt) in der Zone 4 (Umland) gruppiert (Knoten 1). Knoten 2 selektiert aus den restlichen Gemeinden alle diejenigen mit einer Bevölkerungsdichte (BD) unter 208.150 (bezogen auf den Dauersiedlungsraum) in die Zone 3 (Randzone). Mit der Bedingung $BD \leq 502.935$ werden schließlich die verbleibenden 20 Gemeinden der Zone 2 (verstärkte Zone) und der Zone 1 (Kerngebiet) zugeordnet.

Variable	Score	
BD91	100,00	
PP91	58,80	
EAD91	0,00	
AQ91	0,00	
PE91	0,00	

Abb. 3: Einfluss der Variablen Modell Boustedt/Nellner

Hier fällt bereits auf, dass nur zwei Variablen (BD und PP) wirklich relevant sind. Die Einwohner-Arbeitsplatz-Dichte (EAD), Agrarquote (AQ) und der Anteil der in das Kerngebiet auspendelnden Erwerbspersonen an den Erwerbspersonen insgesamt (PE) sind völlig unerheblich (siehe Abb. 3).

Der Klassifikationsbaum der aus dem Ergebnis der Clusteranalysen abgeleitet werden kann ergibt für die Gleichklassifikation ein nahezu identisches Ergebnis: Nur drei Gemeinden werden anders klassifiziert (siehe Abb. 4).

Actual Class	Total Cases	Percent Correct	Class 1	Class 2	Class 3	Class 4
1	5	100,000	5	0	0	0
2	26	92,308	0	24	0	2
3	26	100,000	0	0	26	0
4	12	91,667	0	0	1	11

Abb. 4: Gleichklassifikation der automatischen Klassifikation

Der Klassifikationsbaum gestaltet sich etwas komplexer, er besteht aus fünf Endknoten. Zunächst werden die Gemeinden des Kerngebietes (Zone 1) mit Hilfe der Bevölkerungsdichte ($BD > 1359.405$) abgespalten. Die Gemeinden der Zone 3 (Rand-

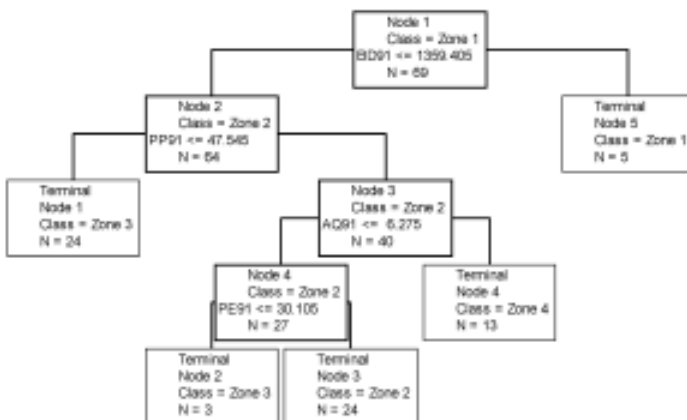


Abb. 5: Klassifikationsbaum - Modell automatische Klassifikation

Variable	Score	
BD91	100,00	
AQ91	77,28	
PP91	72,78	
PE91	20,51	
EAD91	0,00	

Abb. 6: Einfluss der Variablen der automatischen Klassifikation

zone) bilden zwei Endknoten die entweder über die Pendlervariable PP oder über die Agrarquote (AQ) und die Pendlervariable PE erreicht werden können (siehe Abb. 5).

Bei den Variablen kann auf die Einwohner-Arbeitsplatz-Dichte (EAD) verzichtet werden. Sie hat keinen Einfluss auf die Klassifikation (siehe Abb. 6).

Actual Class	Total Cases	Percent Correct	Class 5	Class 13	Class 31	Class 4	Class 15	Class 5
1	3	100,000	3	0	0	0	0	0
2	13	76,923	1	10	0	0	0	2
3	40	72,500	1	1	29	9	0	0
4	10	60,000	0	2	2	6	0	0
5	3	100,000	0	0	0	0	0	3

Abb. 7: Gleichklassifikation Modell ÖSTAT

Auch für die Klassifikation des ÖSTAT war es möglich eine Gruppierung mit einem sehr hohen Übereinstimmungsgrad zu berechnen. Nur in

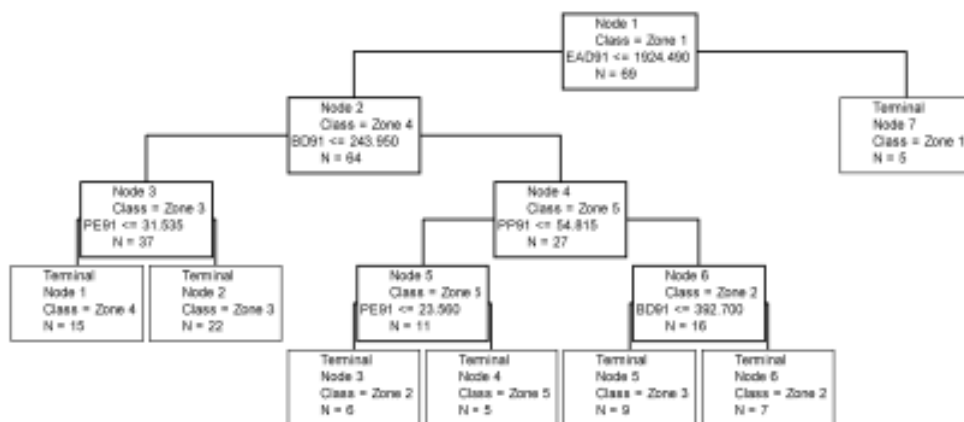


Abb. 8: Klassifikationsbaum - Modell ÖSTAT

Variable	Score	
EAD91	100,00	
BD91	68,64	
PE91	66,16	
PP91	53,45	
AQ91	0,00	

Abb. 9: Einfluss der Variablen Modell ÖSTAT

der schwach besetzten Klasse 4 liegt die Anzahl der Fehlklassifikationen bei vier Gemeinden. Insgesamt wurden 18 Gemeinden anders klassifiziert (siehe Abb. 7). Der Klassifikationsbaum besteht aus sechs Knoten und sieben Endknoten (siehe Abb. 8).

Bei den Variablen kann auf die Agrarquote (AQ) verzichtet werden. Sie hat keinen Einfluss auf die Klassifikation (siehe Abb. 9).

Zusammenfassung

Die Nachbildung der Klassifikation des ÖSTAT mit Hilfe von Entscheidungsbäumen zeigt eine durchschnittliche Übereinstimmung von 82 % und liegt damit etwas unter den beiden anderen Klassifikationen (jeweils 96 % Übereinstimmung). Das Klassifikationsverfahren des ÖSTAT unterscheidet sich von Boustedt/Nellner und den Clusteranalysen dadurch, dass für dieses Raummodell zum Teil andere Variablen herangezogen werden, als kleinste räumliche Einheit die sogenannte Siedlungseinheit verwendet wird und schließlich nur in diesem Modell ein Distanzkriterium eingetht (Fuchs 1997). Die Raumgliederungsmodelle von Boustedt/Nellner und der automatischen Klassifikation können erstaunlich gut mit Klassifikationsbäumen nachgebildet werden. Auf die Variable der Einwohner-Arbeitsplatz-Dichte (EAD) kann beim Entscheidungsbaum der Clusteranalysen verzichtet werden. Die Irrelevanz von drei der insgesamt fünf Variablen bei Boustedt/Nellner (EAD, AQ, PE) ist beachtlich. Daraus kann gefolgert werden, dass die Raumstrukturen im Stadt-Land-Verbund Innsbruck und Umgebung allein mit der Bevölkerungsdichte (BD) und dem Anteil der nach Innsbruck auspendelnden Erwerbepersonen an den Auspendlern insgesamt (PP) bereits vollständig beschrieben werden kann. Insbesondere die Agrarquote (AQ) und die Einwohner-Arbeitsplatz-Dichte (EAD) sind für die Klassifikation nach Boustedt/Nellner - zumindest im hier betrachteten Untersuchungsgebiet (Innsbruck und Umgebung, Jahr 1991) - bedeutungslos.

Literatur:

- Bock, H.H. (1974): Automatische Klassifikation. Theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten. Göttingen.
- Borsdorf, A., Heller, A., Bogner, D., Bartl, K. (2000): Das Stadt-Land-Kontinuum im Alpenraum – Methodenvergleich zur Abgrenzung von Stadtregionen in verfingerten Raumsystemen. In: Die „Alpine Stadt“ zwischen lokaler Verankerung und globaler Vernetzung – Beiträge zur regionalen Stadtforschung im Alpenraum. ISR-Forschungsberichte – 20, S. 59-75, Wien.
- Boustedt, O. (1953): Die Stadtregion. Ein Beitrag zur Abgrenzung städtischer Agglomerationen. Allgemeines statistisches Archiv 37, S.13-26.
- Boustedt, O. (1970): Stadregionen. In: Akademie für Raumforschung und Landesplanung (Hrsg). Handwörterbuch der Raumforschung und Raumordnung. 2. Aufl. S.3207-3237, Hannover.
- Boustedt, O. (1975): Grundriß der empirischen Regionalforschung – Teil 1. (= Taschenbücher zur Raumplanung Band 4), Veröff. Akademie f. Raumforschung u. Landesplanung, Hannover.
- Breiman, L., Friedman, J.H., Olsen, R.A., Stone, C.J. (1984): Classification and Regression Trees. Pacific Grove, Wadsworth.
- Clark, L.A., Pregibon, D. (1993): Tree-Based Models. In: Statistical Models in S, New York.
- Fraley, C., Raftery, A.E. (1991): MCLUST: Software for Model-Based Cluster and Discriminant Analysis. University of Washington - Department of Statistics, Technical Report No. 342, Seattle.
- Fuchs, I. (1997): Stadtregionen 1991 – Das Konzept. Statistische Nachrichten 1, S.76-83, Wien.
- Heller, A. (1997): Neue Typen der Agrarstruktur Österreichs – Automatische Klassifikation mit Clusteranalyse und GIS. (= Innsbrucker Geographische Studien, 27), Innsbruck.
- Heller, A. (1997): Clusteranalyse und GIS: Methoden zur Klassifikation großer Objektmengen mit S-PLUS und ArcView. In: Salzburger Geographische Materialien, 25 – Angewandte Geographische Informationstechnologie IX, S. 269-274, Salzburg.
- Heller, A. (1999): Virtual Project: Urbanisation as a Geographical Effective Structure Between Conurbations and Rural Countryside – Results of Cluster Analysis. http://business.carinthia.com/bgolob/viirp/results_clusterhtml, Klagenfurt.
- Kaufman, L., Rousseeuw, P.J. (1990): Finding Groups in Data: An Introduction to Cluster Analysis. New York.
- Nellner, W. (1975): Das Konzept der Stadtregionen und ihre Neuabgrenzung 1970. In: Veröff. Akademie f. Raumforschung u. Landesplanung (Hrsg): Forschungs- u. Sitzungsberichte, Bd. 103, S. 1-26, Hannover.
- Statistical Science (Hrsg.) (1997): S-Plus. Guide to Statistics Version 4.5 MathSoft Inc., Seattle.
- Steinberg, D., Colla, P. (1998): CART – Classification and Regression Trees. Salford Systems, San Diego.
- Venables, W.N., Ripley, B.D. (1996): Modern Applied Statistics with S-Plus. (= Statistics and Computing), Berlin.