



**TECHNISCHE
UNIVERSITÄT
DRESDEN**



Technische Universität Dresden Institut für Kartographie

Alternative Analysemöglichkeiten geographischer Daten in der Kartographie mittels Self-Organizing Maps

**Ralf Klammer
Mat.Nr.: 3018143
Mylau, 01.06.2010**

Inhaltsverzeichnis

Inhaltsverzeichnis	1
1.) Einleitung	2
2.) Aufbau und Funktionsweise von SOM	5
2.1.) Was sind Self-Organizing Maps?	5
2.2.) Funktionsweise	7
2.3.) Visualisierung des trainierten Kohonen-Netz	11
2.4.) Software	12
3. Möglichkeiten für die Kartographie	14
3.1 Geowissenschaftliches Data Mining	15
3.2 Visualisierung von Daten	17
4. explorative Datenanalyse geographischer Daten	19
4.1 SOM als Geovisualisierung	19
4.1.1 U-Matrix-Darstellung	22
4.1.2 Projektionen (Netzdarstellungen)	26
4.1.3 2D & 3D-Plots	28
4.1.4 Komponentenebenen	29
4.2 Geo-SOM & andere Möglichkeiten zur Verarbeitung von geowissenschaftlichen Daten	32
4.2.1 Hierarchische SOMs	33
4.2.2 Geo-enforced SOM	34
4.2.3 Geo-SOM	35
4.3 SOM & GIS	38
5. Datenverarbeitende Anwendungen	40
5.1 Klassifizierung von Fernerkundungsdaten	40
5.2 Kantendetektion in Satellitenbildern	43
5.3 Auswertung von Zeitreihen & Monitoring	47
5.4 Klassifikation von SAR-Daten	49
5.5 Generalisierung	50
5.6 Problem des Handlungsreisenden (Travelling Salesman Problem)	52
6. SOM als Kartenmetapher zur Visualisierung nicht-geographischer Daten	54
7. Zusammenfassung	62
X. Quellenverzeichnis	63
X.I Literaturnachweise	63
X.II Lehrinhalte aus dem Internet	69
X.III Softwarelösungen	69

1.) Einleitung

Die Kartographie ist eine Wissenschaft, die in ihrem Charakter starke interdisziplinäre Züge aufweist. Sie zeigt sich in den verschiedensten Facetten und wird darum in den unterschiedlichsten Wissenschaften angewandt. Markantester Charakter ist, schon per Definition, die Modellierung von geowissenschaftlichen Ereignissen und Sachverhalten.

„A unique facility for the creation and manipulation of visual or virtual representations of geospace – maps – to permit the exploration, analysis, understanding and communication of information about that space.“ (ICA 2003)

Aus dieser Definition wird die Charakteristik einer Kommunikationswissenschaft (Brassel) deutlich. Gerade seit dem Paradigmenwechsel der 1970er Jahre fließen zahlreiche weitere Aspekte wie Informatik, Semiotik und Psychologie in das Verständnis von Kartographie ein. Dadurch wird die Karte nicht mehr als reines graphisches Mittel verstanden, sondern als Träger und Übermittler von Informationen verstanden. Der Kartennutzer und dessen Verständnis von Karten rücken dabei immer weiter in den Vordergrund und werden „Ziel“ der kartographischen Verarbeitung. Aus diesem Verständnis heraus, möchte ich in der folgenden Arbeit einen relativ neuen Einfluss und Aspekt der Kartographie vorstellen. Es handelt sich um das Modell der Self-Organizing Maps (SOM), welches erstmalig Anfang der 1980er Jahre von Teuvo Kohonen vorgestellt wurde und deshalb auch, von einigen Autoren, als Kohonenmaps bezeichnet wird.

Dem Typus nach, handelt es sich dabei um künstliche neuronale Netze, welche dem Nervensystem des menschlichen Gehirns nachempfunden sind und damit allgemein als eine Art selbständiger, maschineller Lernvorgang angesehen werden können. Im Speziellen sind Self-Organizing Maps ein unüberwachtes Lernverfahren, das in der Lage ist völlig unbekannte Eingabewerte zu erkennen und zu verarbeiten. Durch diese Eigenschaft eignen sie sich als optimales Werkzeug für Data Mining sowie zur Visualisierung von hochdimensionalen Daten. Eine Vielzahl von Wissenschaftlern hat diesen Vorteil bereits erkannt und das Modell in ihre Arbeit einbezogen oder auf dessen Verwendbarkeit analysiert. Deshalb möchte ich im Folgenden, einige dieser Verwendungsmöglichkeiten und den daraus resultierenden Vorteil für die Kartographie aufzeigen.

Zum Einstieg in die Thematik der Self-Organizing Maps muss man zunächst ganz klar differenzieren.

Laut ICA (1996) ist eine „Karte“ eine „graphische Repräsentationsform“, bei welcher der Unterstützungscharakter „zur visuellen Aufnahme und kognitiven Repräsentation von georäumlichen Daten impliziert wird.“

Diese Definition ist für den Begriff der „Map“ in Self-Organizing Maps nicht zulässig. Wie bereits erwähnt, handelt es sich vielmehr um ein Netz aus künstlichen Neuronen, die sich während des Lernvorgangs selbständig anordnen. Daraus ergibt sich die Möglichkeit, hochdimensionale Datenbestände zu verarbeiten und zur Informations- & Erkenntnisgewinnung zu veranschaulichen.

Diese visuelle Repräsentation abstrakter Informationen ist Inhalt der relativ neuen Wissenschaft der Informationsvisualisierung und ist laut André Skupin auch ein interessantes Phänomen für Geographie und Kartographie. In diesem Zusammenhang führt er den Begriff der „map-like-visualization“ ein, welche durch Self-Organizing Maps realisiert werden können.

Durch die Eigenschaft der Dimensionsreduzierung der Self-Organizing Maps ist dieses Modell ebenfalls ein sehr beliebtes Werkzeug in der Knowledge Discovery of Databases (KDD). Nach Fayyad et al.¹ ist KDD der Gesamtprozess des Erforschens von nützlichem Wissen aus Datenbeständen. Im Speziellen finden Self-Organizing Maps Anwendung im Data Mining, einem Teilbereich des KDD. Durch Anwendung bestimmter Algorithmen, können dabei Muster aus Daten extrahiert werden. Als bedeutenden Einfluss muss man in diesem Bezug die Abhandlungen von Fernando Baçao nennen, welcher eine Weiterentwicklung zur Geo-SOM veröffentlicht hat.

Durch diese Charakteristiken lassen sich die Self-Organizing Maps in die Kategorie der technischen Methoden (Koch²) der Kartographie einordnen. Im Speziellen als Verfahren der Datenverwaltung und der graphischen Datenverarbeitung.

Ich werde im Folgenden zunächst, in Kapitel zwei, die Definition von Self-Organizing Maps, sowie deren Algorithmus genauer erklären. Im dritten Kapitel möchte ich darauf eingehen, in welcher Form diese Technik in der Kartographie genutzt werden könnte. Daraufhin widme ich mich der Analyse von Veröffentlichungen, welche sich bereits mit der Verwendung des SOM-Modells in den

¹ Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (1996) – “From Data Mining to Knowledge Discovery in Databases” - S.3

² Koch, Wolf Günther (2007) – Vorlesungsunterlagen “Theoretische Kartographie”

Geowissenschaften beschäftigt haben. Dazu analysiere ich im vierten Kapitel Methoden zur explorativen Datenanalyse mittels SOM, gehe im fünften Kapitel auf datenverarbeitende Anwendungen ein und stelle abschließend, im sechsten Kapitel, die Visualisierung von nicht-geographischen Daten mittels SOM vor.

Ich kann und möchte mit dieser Arbeit keinen Anspruch auf Vollständigkeit erheben. Das Ausmaß der Anwendungen und Erweiterungen ist enorm, weshalb es mir nur möglich ist, einen Teil der für die Kartographie interessanten Anwendungen vorzustellen. Teuvo Kohonen erweitert ständig die Liste der Arbeiten, die auf seinem Algorithmus basieren, und veröffentlicht diese auf seiner Institutshomepage³. Der aktuelle Stand sind 7719 Arbeiten, die sich mit SOM beschäftigen.

³ <http://www.cis.hut.fi/research/som-bibl/>

2.) Aufbau und Funktionsweise von SOM

2.1.) Was sind Self-Organizing Maps?

Das Modell der Self-Organizing Maps wurde in den 1980er Jahren von Teuvo Kohonen entwickelt, weshalb es auch als Kohonenmaps bezeichnet wird. Kohonen ist Leiter des Laboratory of Computer and Information Science an der technischen Universität von Helsinki. Seine Arbeiten befassten sich in den 1970er und frühen 1980er Jahren mit „selbständigen assoziativen Speichern“ wodurch er diesen Algorithmus entwickelte. Seit die Öffentlichkeit, Ende der 1980'er Jahre, auf seine Arbeit aufmerksam wurde gab es zahlreiche Projekte und Arbeiten, die sich damit beschäftigt und den Algorithmus teilweise weiterentwickelt haben. Allerdings gelten die 1995⁴ und 2001 von Kohonen veröffentlichten Bücher bis heute als wesentliche Referenz zu diesem Thema. Hier werden die grundlegenden mathematischen Prinzipien, der Grundaufbau von SOM, physiologische Interpretationen sowie Varianten und Anwendungen vorgestellt.

Die wesentliche Idee werde ich kurz anhand eines Beispiels erläutern. Wenn man eine unbekannte Karte das erste Mal betrachtet, dann benötigt man zunächst etwas Zeit, um sich darauf zu orientieren. Welches Thema soll dargestellt werden, was für ein Gebiet wird abgebildet, wie ordnet sich dieses überregional ein oder welcher Maßstab wird verwendet? Über die Legende erfährt man mit welchen Signaturen die Objekte veranschaulicht werden und viele andere Eindrücke muss man zunächst verarbeiten. Umso öfter man diese Karte nun benutzt, desto leichter fällt anschließend die Orientierung. Man sammelt Erfahrungen und kann diese, bei Bedarf, abrufen. Der dargestellte Lernprozess des menschlichen Gehirns wird mittels künstlicher neuronaler Netze nachempfunden, wodurch ein selbständiger Lernvorgang für Computer ermöglicht wird.

Die Methode der Self-Organizing Maps ordnet sich in die Systematik der künstlichen neuronalen Netze ein, die mit unüberwachtem Lernverfahren arbeiten und ist dabei das Bekannteste. Unüberwacht bedeutet, dass keinerlei Regeln vorgegeben werden müssen, sondern das Netz rein über den Algorithmus in der Lage ist, selbständig neue Eingabewerte zu erkennen und zu verarbeiten.

Inspiriert ist dies von der menschlichen Großhirnrinde. Dort sind die Regionen für Sinneswahrnehmungen topologisch ähnlich

⁴ Kohonen, Teuvo (1995) – “Self-Organizing Maps”

angeordnet, wie die korrespondierenden Körperteile. Betrachtet man zum Beispiel den somatosensorischen Kortex (Abb.1), welcher empfindlich gegenüber Berührung, Druck, Vibration, Temperatur und teilweise Schmerz ist, dann wird der genannte Zusammenhang deutlich. Zum Beispiel befinden sich die Regionen der Sinneswahrnehmung der Finger direkt neben denen der Hand, des Unterarms und des Ellenbogens. Kohonen bezeichnet diese benachbarte Anordnung als „Brain Maps“⁵. Ein Schema, das auf viele weitere Bereiche der Großhirnrinde zu trifft.



Abb. 1 - somatosensorischer Kortex
 (Quelle: Kranz & Burkhard (2007) - Visualisierung von Self Organizing Maps mittels Java-Applets und HTML-Tutorial sowie Analyse ihrer Anwendbarkeit im Data- und Text-Mining-Kontext – S.32)

Ebenso interessant ist der Fakt, dass häufig genutzte Areale sehr stark ausgeprägt sind. Der für den Tastsinn der Finger verantwortliche Bereich ist deutlich größer als der des Oberschenkels. Dadurch ist eine detailliertere Wahrnehmung mit den Fingern möglich. Der Algorithmus der Self-Organizing Maps ist genau an diese Funktionsweise angelehnt und bildet sie nach.

Die Verarbeitung ähnlicher Werte in topologisch benachbarten Regionen zeigt in Verbindung mit dem „First Law of Geography“⁶ einen Hinweis inwiefern ein solcher Algorithmus für die Kartographie oder allgemein die Geowissenschaften von Bedeutung sein kann. Denn dieses sagt aus, dass alles miteinander verbunden ist, jedoch ist nahe beieinander liegendes stärker miteinander verbunden als weit entferntes. Auf den eben dargestellten Sachverhalt werde ich in den folgenden Kapiteln näher eingehen.

⁵ Kohonen, Teuvo (1995) – „Self-Organizing Maps“ (S.99)

⁶ Tobler (1970) – „A computer movie simulating urban growth in the Detroit region“ (S.3 der Onlineversion)

2.2.) Funktionsweise

Wie bereits erwähnt ist der Algorithmus der Self-Organizing Map ein Werkzeug zur Visualisierung von hochdimensionalen Daten, indem Ähnlichkeitsvektoren, auch als Gewichtsvektoren bezeichnet, von Eingabewerten erstellt werden. Diese Vektoren haben immer die gleiche Dimension wie die Eingabewerte. Sie werden normalerweise in ein zweidimensionales Gitter abgebildet und somit wesentlich vereinfacht. Zwar wird der Informationsgehalt dabei stark komprimiert, aber die wichtigen topologischen Beziehungen der Datenelemente bleiben bei der Darstellung erhalten. Somit ergibt sich eine Art Abstraktion oder auch Generalisierung. Kohonen beschreibt den SOM-Algorithmus formal als „nonlinear, ordered, smooth mapping of high-dimensional input data manifolds onto the elements of a regular, low-dimensional array.“⁷

Man unterscheidet im SOM-Modell zwischen dem Kohonen-Netz, in welchem der gesamte Eingaberaum durch Neuronen abgebildet wird und der eigentlichen Self-Organizing Map. Diese wird im Anschluss zur Visualisierung der Ergebnisse verwendet und kann mittels quadratischer oder hexagonaler Topologie realisiert werden. Der Unterschied und Vorteil einer jeweiligen Topologie liegt dabei im Nachbarschaftsverhältnis mit vier oder sechs Nachbarn, sowie dem Zweck der Anwendung des Algorithmus. Die Anzahl der Neuronen der SOM entspricht immer exakt der des Kohonennetzes wobei jedes Neuron im Netz durch eines in der SOM abgebildet wird.

Im SOM-Algorithmus wird zwischen Lernmodus und Ausführmodus unterschieden. Der Lernmodus wird mit einem zufällig generierten Netz gestartet. Über das Anlegen von Eingabewerten wird dieses dann trainiert. Im Ausführmodus ist dieses Netz anschließend in der Lage neue Eingabewerte zu erkennen und zu verarbeiten. Wie schon angesprochen ist dies ein unüberwachtes Verfahren der künstlichen neuronalen Netze. Somit können im trainierten Zustand neue, völlig unbekannte Eingabewerte verarbeitet werden.

Der Lernmodus beginnt damit, dass eine Neuronenkarte initialisiert wird, bei der alle Neuronen einen Zufallswert für deren Ähnlichkeitsvektor zugewiesen bekommen. Ein Neuron ist dabei die kleinste informationsverarbeitende Einheit des gesamten Systems. Durch die Vernetzung mit anderen Neuronen ist eine effektive Informationsverarbeitung möglich. Für jeden Eingabewert wird nun ein Lernschritt durchgeführt, wobei Neuronen ausgewählt werden deren Ähnlichkeitsvektoren an den Vektor des Eingabewertes, den

⁷ Kohonen (1995) – “Self-Organizing Maps” (S.106)

Trainingsvektor, angepasst werden. Diese Anpassung, also der Lernvorgang, erfolgt in 4 Schritten⁸:

1. Finden des Neurons, dessen Ähnlichkeitsvektor die meisten Übereinstimmungen mit dem Trainingsvektor aufweist
2. Anpassung des Ähnlichkeitsvektors des Gewinnerneurons nach Kohonen-Lernregel
3. Finden aller Neurons, die sich in topologischer Nachbarschaft zu dem Gewinnerneuron befinden (bezogen auf Blockradius)
4. Anpassung der Ähnlichkeitsvektoren dieser benachbarten Neurons

Für jeden Eingabewert wird zunächst das Abbruchkriterium überprüft. Üblich ist dafür eine maximale Anzahl an Lernschritten, und in deren Folge wird der Lernvorgang wiederholt. Mit der Zeit ordnen sich Neurons mit vergleichbaren Ähnlichkeitsvektoren, innerhalb des Netzes, nahe beieinander an. Entstehen stark ausgeprägte Bereiche des Netzes, so lässt dies darauf schließen, dass bestimmte Trainingsvektoren sehr häufig angelegt wurden.

3	3	3	5	5	6	6	6
3	3	5	5	8	8	6	6
3	1	1	5	8	4	8	6
1	1	1	4	4	4	4	4
2	2	4	4	4	4	4	4
2	2	9	4	4	4	4	7
2	9	9	4	4	4	7	7
2	9	9	9	4	4	7	7

Abb. 2 - Beispiel für ein Neuronennetz
(Quelle: Kranz (2007) – "Self Organizing Maps Tutorial" – S.5)

In Abbildung 2 ist zu erkennen, dass es zu einer starken Häufung der Zahl 4 kommt. Dieser Wert wurde offensichtlich überdurchschnittlich oft durch einen Trainingsvektor repräsentiert.

Der Lernmodus beginnt zunächst mit einer Grobsortierung. Mit steigender Anzahl an Lernschritten tritt zunehmend eine Art Feinjustierung ein. Visualisiert man die einzelnen Schritte, wie dies in verschiedenen onlineverfügbaren Applets⁹ der Fall ist, kann man erkennen, dass das Netz in der ersten Phase zunächst kollabiert und sich dann „grob“ ordnet. „Grob“ soll bedeuten, dass zunächst keine klare Struktur zu erkennen ist. Zu Beginn der zweiten Phase zeichnet

⁸ Kranz, Sabine (2007) – "Self Organizing Maps Tutorial" (S.5)

⁹ z.B.: <http://www.hiskp.uni-bonn.de/gruppen/edda/analysis/SOM.html> (05.02.2010)

sich bereits eine Struktur ab, welche im weiteren Verlauf zunehmend detaillierter wird.

Hervorgerufen wird dieser Effekt durch die veränderlichen Parameter Blockradius und Lernrate, welche in der Kohonen-Lernregel enthalten sind.

Die einfachste Form des Algorithmus setzt sich wie folgt zusammen:

$$\vec{w}_k^{neu} := \vec{w}_k^{alt} + \alpha \times (\vec{x} - \vec{w}_k^{alt})$$

w_1, \dots, w_m	Neuronen des Netz mit Anzahl m
$\vec{w}_1, \dots, \vec{w}_m$	Ähnlichkeitsvektoren, die Neuronen repräsentieren
α	Lernrate, mit $0 < \alpha < 1$ (wird problemabhängig gewählt & meist durch Test bestimmt)
\vec{w}_k	Neuron mit „maximaler Erregung“
\vec{x}	Trainingsvektor

In dieser einfachen Version wird lediglich das betreffende Neuron trainiert, also angepasst. Alle anderen bleiben unverändert. Somit wird die topologische Nachbarschaft zunächst in keiner Weise berücksichtigt. Ein kurzes Beispiel soll die daraus resultierenden Konsequenzen verdeutlichen. Nehmen wir an, der Eingaberaum ist eine Weltkarte und alle Anfangsgewichte (Ähnlichkeiten) werden mit Koordinaten von Dresden initialisiert. Die Wahrscheinlichkeiten der Trainingsvektoren werden jedoch mit Werten aus Europa und Südamerika belegt. Wenn nun ein südamerikanischer Eingabewert angelegt wird, dann wird das Neuron mit dem kleinsten euklidischen Abstand, zum Beispiel in der Südvorstadt, maximal erregt. Dieses Neuron wird in Richtung Südamerika, vielleicht bis zu den Kanaren, verschoben. Werden weitere südamerikanische Eingabewerte angelegt, so wird natürlich das Neuron auf den Kanaren erregt und weiter verschoben. Statt ein anderes, welches zu Dresden gehört. Der resultierende Effekt daraus ist, dass ein einziges Neuron Südamerika repräsentiert und der Rest in Europa verbleibt. Dieser Effekt wird in Abbildung 3 angedeutet.

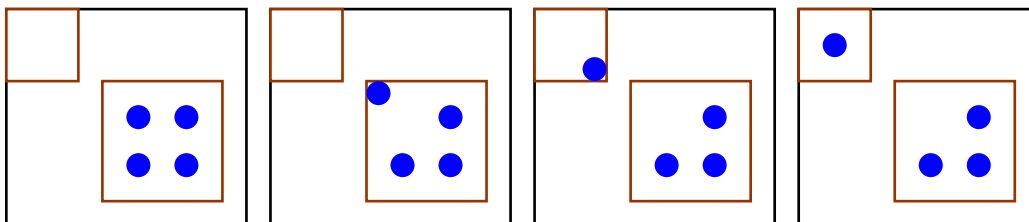


Abb.3 – Topologiefehler (nach Lippe)

(Quelle: Lippe (2007) – „Interaktive Einführung in Neuronale Netze - Kohonen-Netze“ – Kapitel „Der Algorithmus (SOM)“)

Allerdings soll der Algorithmus die Topologie erhalten. Die Konsequenz daraus ist, dass der Algorithmus um eine Nachbarschaftsfunktion erweitert werden muss. Dadurch wird nicht nur das Gewinnerneuron bewegt, sondern auch alle Neuronen, die sich innerhalb eines bestimmten Radius befinden. Kohonen schlägt dafür verschiedene Funktionen vor. Ich möchte hier jedoch nur die einfachste Nachbarschaftsfunktion veranschaulichen, welche ausschließlich den euklidischen Abstand ermittelt:

$$\varphi(w_i, w_k) = \begin{cases} 1: d(w_i, w_k) \leq R \\ 0: \text{sonst.} \end{cases}$$

Weitere Möglichkeiten sind beispielsweise Treppen-, Gauss-, Cosinus- oder Kegel-Funktion.

Durch Einbindung der Nachbarschaftsfunktion erhält man die erweiterte Kohonen-Lernregel:

$$\vec{w}_i^{neu} := \vec{w}_i^{alt} + \alpha \times \varphi(w_i, w_k) \times (\vec{x} - \vec{w}_i^{alt})$$

Wie beschrieben, läuft der Algorithmus in zwei Phasen ab. Zu Beginn ist die Lernrate relativ groß, um die Neuronen grob zu ordnen. Im weiteren Verlauf wird diese immer kleiner, um die Neuronen in ihrer Anordnung lediglich fein zu justieren. Dazu wird die Lernrate als Funktion der Zeit aufgefasst. Ebenso sollte die Nachbarschaftsfunktion zeitabhängig implementiert werden. Denn der Einfluss auf Nachbarneuronen sollte, im Sinne der Feinjustierung, zum Ende hin geringer sein. Somit ergibt sich die endgültige Formel:

$$\vec{w}_i^{neu} := \vec{w}_i^{alt} + \alpha(t) \times \varphi(w_i, w_k, t) \times (\vec{x} - \vec{w}_i^{alt})$$

t steht dabei für die Anzahl der schon durchgeführten Trainingsschritte.

Für die Berechnung der Lernrate gibt es wiederum verschiedene Ansätze. Diese kann in Abhängigkeit von der Problematik gewählt werden. Hier zum Beispiel, wie vorab beschrieben, ein regressiver Ansatz:

$$\alpha(t) = \alpha_0 \left(1 - \frac{t}{t_{\max}} \right)$$

Der Ablauf des Lernmodus wird in Abbildung 4 noch einmal graphisch verdeutlicht.

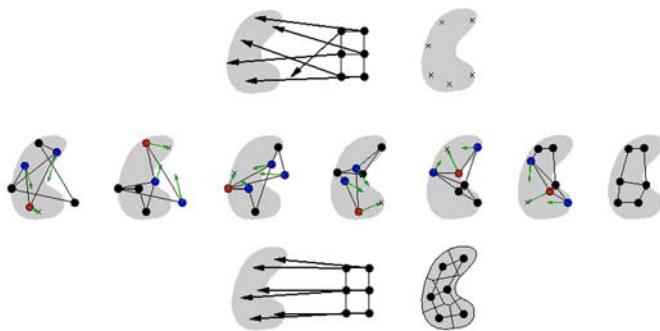


Abb.4 – Funktionsweise der Lernregel
 (Quelle: Lippe (2007) – „Interaktive Einführung in Neuronale Netze - Kohonen-Netze“ – Kapitel „Der Algorithmus (SOM)“)

2.3.) Visualisierung des trainierten Kohonen-Netz

Die wesentlichen Visualisierungsmethoden des SOM-Netzes sind laut Agarwal und Skupin¹⁰ Komponentenebenen (component planes), Verzerrungsmuster (distortion patterns) sowie Klassendarstellungen (clustering).

Komponentenebenen sind eine Darstellung der Neuronengewichte jeder einzelnen Variable des Eingangsdatensatzes. Nimmt man beispielsweise Klimadaten, dann würde für jede der Variablen, wie Luftfeuchtigkeit, Temperatur oder Windgeschwindigkeit die relative Verbreitung dargestellt. Somit ergibt sich die Möglichkeit, die einzelnen Ebenen auf ihre Ähnlichkeit zu untersuchen und damit Vergleiche einzelner Variablen anzustellen.

Wie bereits beschrieben liegt der wesentliche Vorteil der SOM darin, dass topologische Beziehungen erhalten bleiben. Dies gilt allerdings nicht für die geometrischen Lagebeziehungen. Aus der Visualisierung der auftretenden Verzerrungsgrade ergibt sich die Möglichkeit Verzerrungsmuster zu generieren und daraus Clustergrenzen abzuleiten. Die am weitesten verbreitete Vorgehensweise ist, die von Ultsch¹¹ entwickelte U-Matrix-Repräsentation.

Abbildung 5 zeigt einen Vergleich der beiden bereits beschriebenen Methoden. Die Komponentenebenen geben die 8 einzelnen sozioökonomische Indizes eines Datensatzes wieder, während mittels U-Matrix der gesamte Datensatz visualisiert wird.

¹⁰ Agarwal; Skupin (2008) – „Introduction: What is a Self-Organizing Map?“ (S.12)

¹¹ Ultsch (1993) – „Self-Organizing Neural Networks for Visualization and Classification“

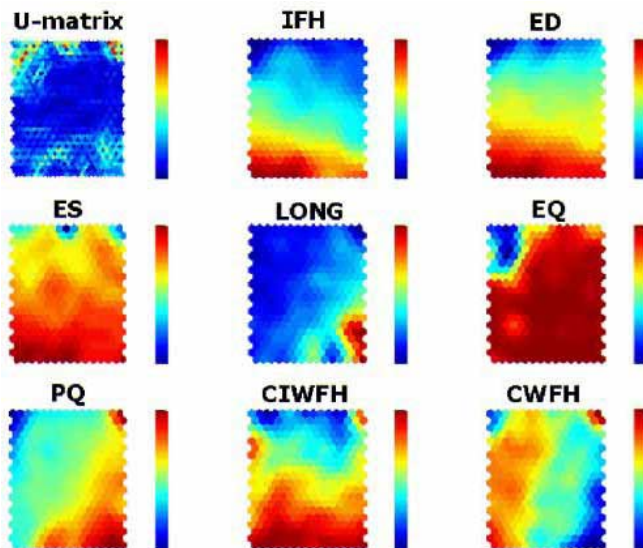


Abb. 5 – Vergleich U-Matrix und Komponentenebenen für ein 20 x15 hexagonales SOM
 (Quelle: da Silva.; Monteiro; Medeiros (2004) – “Visualization of Geospatial Data by Component Planes and U-Matrix” – S.12)

Eine eingehende Beschreibung dieser Visualisierungstechniken werde ich im Kapitel 4.1 „Geovisualisierung“ geben und mit anschaulichen Beispielen verdeutlichen.

2.4.) Software

Das vollständige Softwarepaket zu dem von Kohonen entwickelten Algorithmus ist öffentlich zugänglich und kann von dessen Institutshomepage heruntergeladen werden. SOM_PAK¹² hat keinerlei graphische Oberfläche und wird über die DOS-Eingabeaufforderung angesteuert. Somit ist der universelle Einsatz des Softwarepaketes sichergestellt. Nachteilig ist, dass man einige Zeit investieren muss, abhängig von den eigenen Programmierkenntnissen, bis man ein zufrieden stellendes Ergebnis erhält.

Das SOM_PAK besteht aus einer losen Sammlung von C-Programmen, deren Quelltexte mitgeliefert werden und somit einsehbar sind. Bação und Lobo stellen im Internet einen E-Learning-Kurs¹³ zur Verfügung, in dessen Verlauf anschaulich verdeutlicht wird wie das SOM_PAK ganz simpel anwendbar ist.

Es gibt eine große Anzahl von weiteren, öffentlich verfügbaren Softwarepaketen. Durch die sehr einfache Gestaltung von SOM_PAK liegt es auf der Hand, dass jeder, der sich intensiver mit SOM beschäftigt, eine Benutzeroberfläche gestaltet, um einen förderlicheren Umgang damit zu ermöglichen. Vorstellen möchte ich,

¹² Kohonen; Hynninen; Kangas; Laaksonen (1996) – „SOM_PAK: The Self-Organizing Map Program Package“

¹³ Bação; Lobo (2007) – “Geospatial Data Mining (E-Learning)” - Project 1

in diesem Zusammenhang, die SOM Toolbox¹⁴ sowie das Geo-SOM suite tool¹⁵.

Die SOM toolbox¹⁴ ist ein Erweiterungspaket für die Software Matlab5. Dadurch werden Training, Visualisierung und Auswertung der SOM wesentlich vereinfacht. Dementsprechend ist die Anwendung von SOM sehr vielseitig. Benötigt allerdings auch ein hohes Maß an Nutzereingriffen.¹⁶

Das Geo-SOM suite tool¹⁵ basiert auf der Toolbox für Matlab und stellt den von Bação vorgestellten Ansatz der Geo-SOM¹⁷ direkt als Softwarelösung zur Verfügung. Doch darauf werde ich erst in Kapitel 4.2 näher eingehen.

Ebenso stehen zahlreiche weitere Softwarelösungen zur Verfügung, die sich entweder rein auf SOM spezialisiert haben (z.B.: SOMine von Viscovery) oder teilweise in ihren vorhandenen Softwareanwendungen SOM-Ansichten mitliefern (z.B.: Synapse von Peltarion). Diese sind jedoch in jedem Fall kostenpflichtig.

¹⁴ Alhoniemi; Himberg; Parhankangas; Vesanto (2000-2005) – „SOM Toolbox“

¹⁵ Bação (2004) – “GeoSOM - Clustering geo-referenced data”

¹⁶ Vesanto; Himberg; Alhoniemi; Parhankangas (2000) – “SOM Toolbox for Matlab 5”

¹⁷ Bação (2005b) – “Geo-SOM and it’s Integration with Geographic Information Systems”

3. Möglichkeiten für die Kartographie

Der traditionelle Weg der Erkenntnisgewinnung aus Daten besteht darin, diese manuell zu analysieren und zu interpretieren. Der Einsatz von Computertechnik macht es heutzutage möglich, sehr große Datenmengen über die Welt zu erheben und zu sammeln. Der damit verbundene drastische Anstieg von Datenmengen führt dazu, dass die manuelle Interpretation dieser, in vielen Bereichen, zu aufwendig wird. Datenbanken können heute schnell mehr als 10^9 Einträge besitzen. Dadurch wäre es einfach zu teuer und zu langsam diese manuell zu analysieren. In Folge dessen ist es notwendig geworden, solche Analysen, zumindest teilweise automatisiert und computergestützt ablaufen zu lassen. Mit dieser Problematik setzt sich die Knowledge Discovery in Databases (KDD) auseinander und stellt Lösungsansätze zur Verfügung.

Zweierlei Vorteile ergeben sich aus einer automatisierten Datenanalyse. Zum einen steht, wie schon beschrieben, der ökonomische Faktor im Vordergrund, da die Zeit- & Kostenersparnis, gerade bei sehr großen Datenbanken, exorbitant sein kann. Zum anderen erschließen sich auch wissenschaftliche Vorteile daraus. Denn es wird möglich, Muster und Strukturen in Daten zu erkennen, die man sonst nicht hätte identifizieren können.

Die Bezeichnung „Knowledge Discovery in Databases“ wurde erstmalig auf einem Workshop 1989¹⁸ verwendet und sollte verdeutlichen, dass Erkenntnisse (knowledge) das Ergebnis von an Daten orientierten Untersuchungen ist. Somit wurde es in Forschungsbereichen der künstlichen Intelligenz bekannt. Laut Definition ist KDD ein „nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data“¹⁹.

Die Verifizierung von Hypothesen sowie die Entdeckung von neuem, nützlichem Wissen, indem Regelmäßigkeiten und Strukturen erkannt und visualisiert werden, können als wesentliche Ziele der KDD herausgestellt werden. Diese Disziplin besteht aus zahlreichen untergeordneten Prozessen, wie Datenvorverarbeitung, Datentransformation, Datenauswertung (Data Mining) oder Dateninterpretation und -visualisierung. Diese können alle unter dem Oberbegriff der KDD zusammengefasst werden.

Für eine Anwendung der SOM in der Kartographie sind die Teilbereiche Data Mining sowie Datenvisualisierung im Wesentlichen

¹⁸ Piatetsky-Shapiro (1991) – “Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop”

¹⁹ Fayyad; Piatetsky-Shapiro; Smyth (1996) – “From Data Mining to Knowledge Discovery in Databases” (S.5)

die interessantesten und nützlichsten Bereiche der KDD. Aus diesem Grund werde ich auf diese im Folgenden näher eingehen.

3.1 Geowissenschaftliches Data Mining

Wie bereits beschrieben ist Data Mining ein Teilschritt von KDD. Dabei werden, unter Anwendung bestimmter Algorithmen, verschiedene Muster der Daten erkannt und extrahiert. Nach Koua & Kraak²⁰ kann man diesen Prozess in drei Hauptkategorien unterscheiden:

- Erklären (Explanatory)
- Bestätigen von Hypothesen (Confirmatory)
- Erforschen (Exploratory)

Die Anwendungen reichen von klassischen, statistischen Methoden bis hin zu maschinengesteuerten Methoden, wie es die SOM ist.

Das Interesse der Kartographie liegt in diesem Bereich natürlich in der Anwendung auf geographische Daten. Da Geoinformationen bestimmte Eigenschaften besitzen, die sie von anderen Informationen unterscheiden, ist es notwendig spezielle Verfahren zu verwenden, welche den individuellen Charakter von Geoinformationen berücksichtigen. In der Literatur findet man für diese Methoden verschiedene Begriffe (GeographicalDM²¹; GeospatialDM²²). Aufgrund dieser Heterogenität möchte ich diese Methoden hier unter dem Begriff des Geowissenschaftlichen Data Mining (GDM) zusammenfassen.

Was macht also den individuellen Charakter von Geodaten aus?

Nach dem Lexikon der Kartographie sind Geodaten „Daten, die einen Raumbezug aufweisen, über den ein Lagebezug zur Erdoberfläche hergestellt werden kann.“²³ Diese dienen dazu, reale Objekte durch ihre geometrischen und inhaltlichen Eigenschaften zu beschreiben.

Nach dem 1. Gesetz der Geografie ist „Alles mit Allem verbunden aber Näheres ist mehr verbunden als weiter Entferntes.“²⁴ Dementsprechend sind räumliche Daten durch die Existenz von räumlichen Abhängigkeiten charakterisiert, was auch deren

²⁰ Koua; Kraak (2008) – „An Integrated Exploratory Geovisualisation Environment Based on Self-Organizing Map“ (S. 46)

²¹ Openshaw, Stan (1999) – „Geographical data mining: key design issues“

²² Bação (2007) – „Geospatial Data Mining“

²³ Bollmann; Koch (2002) – „Lexikon der Kartographie und Geomatik“

²⁴ Tobler (1970) – „A computer movie simulating urban growth in the Detroit region“ (S.3 der Onlineversion)

wichtigste Eigenschaft ist. Bação, Lobo & Painho konkretisieren dies. „Ein Phänomen an einer bestimmten Stelle ist eine Funktion der zugrunde liegenden Faktoren, aber auch der Intensität des gleichen Phänomens an einer benachbarten Stelle.“²⁵ Das bedeutet, dass räumliche Daten, immer miteinander korrelieren. Diesen Effekt beachten klassische statistische Methoden nicht, weshalb spezielle GDM-Methoden notwendig sind. Räumliche Heterogenität ist ebenfalls eine sehr wichtige Eigenschaft von Geodaten, denn jedes natürliche Objekt der realen Welt, variiert in seinen Eigenschaften. Diese Mannigfaltigkeit der Attribute von geographischen Objekten führt dazu, dass Geodaten überall unterschiedlich sind. Betrachtet man GIS-Datenbanken, so bestehen diese im Wesentlichen aus drei großen Klassen von Datentypen:

1. geographische Koordinate
2. zeitliche Koordinate
3. geographischen Entitäten mit multiplen Attributen²⁶

Diese liegen in jeweils unterschiedlichen Einheiten vor und können nicht direkt miteinander verglichen werden. Ein praktikables GDM-Werkzeug sollte damit umgehen können.

Daten können auch in Bezug auf deren Aufnahme unterschieden werden. So kann man zwischen primären und sekundären Daten unterscheiden²⁵. Primäre Daten werden in Bezug auf ein bestimmtes Analyseziel und dementsprechend, methodisch auf die speziellen Anforderungen gesammelt. Sie sind meist abfragefertig vorhanden und können direkt genutzt werden. Sekundäre Daten dagegen, wurden für irgendeine Analyse gesammelt und werden im Nachhinein zur Datenanalyse verwendet. Diese Daten können unzuverlässig sein, wenn deren Herkunft und Aufnahmebedingungen nicht klar nachvollziehbar sind. Des Weiteren könnten sie inkompatible Dateiformate oder Maßeinheiten aufweisen. Es liegt jedoch auf der Hand, dass für Data Mining zum Großteil sekundäre Daten verwendet werden. Die Digitalisierung der Welt nimmt immer weiter zu und so werden zahlreiche Daten gespeichert, die man im Nachhinein für Auswertungen verwenden kann, welche bei der Aufnahme noch nicht bekannt oder beabsichtigt waren.

Diese verschiedenen Eigenschaften stellen bestimmte Ansprüche an ein GDM-Werkzeug. Openshaw²⁷ stellt diese speziellen Ansprüche als

²⁵ Bação; Lobo; Painho (2005a) – “On the particular characteristics of spatial data and its similarities to secondary data used in data mining” (S.3)

²⁶ Openshaw (1999) – “Geographical data mining: key design issues” (S.6)

Basisfunktionen zusammen, die nützlich sind, aber nicht alle erfüllt werden müssen. Ein GDM sollte:

- grundlegende Bedürfnisse an räumliche Datenerforschung erfüllen
- das Potential besitzen, neue Einblicke, Ideen & Hypothesen für die Datenanalyse bereitzustellen
- künstlerische Darstellung von Musterstrukturen bereitstellen um damit die Vorstellungen anzuregen
- allgemein unübliche Datenbankmuster beleuchten und ortsabhängige Regelmäßigkeiten entdecken
- einfach zu benutzen sein und den Kriterien für Einsatz in GIS entsprechen

Inwiefern bereits durchgeführte Projekte und Arbeiten diesen Anforderungen entsprechen, werde ich im 5. Kapitel analysieren.

3.2 Visualisierung von Daten

Der Teilbereich des KDD, welcher sich mit der Überprüfung von Hypothesen beschäftigt ist die explorative Datenanalyse²⁸. Hier werden die Daten visualisiert um diese zu verstehen, zu klassifizieren und zusammenzufassen. Tukey²⁸ beschreibt dies als Detektivarbeit und bezieht sich dabei auf das menschliche, visuelle Augen-Gehirn-System. Demnach sind große Datenmengen mittels Illustrationen besser verständlich, als in einer tabellarischen Darstellung.

Koua & Kraak²⁹ unterteilen den Prozess, der explorativen Datenanalyse, in zwei Ebenen. Am Beginn steht die allgemeine Visualisierung der Daten, mit ihren Strukturen und Regelmäßigkeiten. Zum Zweiten das Entdecken von Mustern und Generieren von Hypothesen aus den dargestellten Mustern.

Rein auf die Visualisierungsregeln reduziert, ist das auch ein wichtiges Feld der Kartographie. Denn es müssen verschiedenste Möglichkeiten und Techniken beachtet werden, bei der Visualisierung einer Karte wie auch von Daten. Demzufolge wird

²⁷ Openshaw (1999) – “Geographical data mining: key design issues” (S.7)

²⁸ Tukey (1977) – “Exploratory Data Analysis”

²⁹ Koua; Kraak (2008) – “An Integrated Exploratory Geovisualisation Environment Based on Self-Organizing Map” (S. 48)

einem Betrachter das Auffinden von gesuchten Informationen in optimierter Form ermöglicht. Für eine Visualisierung im Sinne der explorativen Datenanalyse lassen sich vier grundlegende Ziele herausstellen, welche die Visualisierung bereitstellen sollte:

- Muster erkennen
- Beziehungen erforschen, um Hypothesen zu generieren
- Erforschung der Verteilung von Daten auf der Karte
- Erkennung von Unregelmäßigkeiten in den Daten³⁰

Im Sinne der explorativen Datenanalyse kann die SOM ebenfalls genutzt werden, um Datenbestände in einer visuellen Darstellung unter Verwendung einer Kartenmetapher zu analysieren. Dieses Verfahren wird rein auf nicht-geographische Daten angewandt. Es werden dafür die semiotischen Anforderungen einer Karte auf die kartenähnliche Darstellung von nicht-geographischen Daten angewendet. Deswegen kann der Eindruck von Bergen und Tälern oder auch von Inseln im Meer entstehen. Auf die Anwendungen der SOM in diesem Teilbereich werde ich im sechsten Kapitel näher eingehen.

Über eine kombinierte Darstellung von Karte und den Attributwerten der geographischen Daten in einer SOM erschließt sich ein weiteres Feld der explorativen Datenanalyse. Auf diesen Bereich werde ich im vierten Kapitel näher eingehen.

³⁰ Koua; Kraak (2008) – “An Integrated Exploratory Geovisualisation Environment Based on Self-Organizing Map” (S. 48)

4. explorative Datenanalyse geographischer Daten

4.1 SOM als Geovisualisierung

Der Grundgedanke der Geovisualisierung besteht im Wesentlichen darin, verschiedene Muster und Beziehungen in Daten zu finden, welche bei der Lösung bestimmter geowissenschaftlicher Probleme hilfreich sein könnten. Gerade bei großen Datenmengen können konventionelle, räumliche Analysewerkzeuge schnell überfordert sein. Deshalb möchte ich im Folgenden einige Anwendungen aufzeigen, die mittels SOM realisiert wurden.

Allgemein beschreibt Ware³¹ die Visualisierung von Informationen als den „Einsatz von interaktiven visuellen Repräsentationen, um die Kognition zu verstärken und zu unterstützen“, wodurch Sachverhalte sichtbar gemacht werden können. Die Geovisualisierung umfasst im Speziellen die „Entwicklung von Theorien, Methoden und Werkzeugen für die visuelle Exploration, Analyse, Synthese und Präsentation von räumlichen Daten für bestimmte Anwendungen und Nutzer.“³² Sie beinhaltet verschiedenste Techniken zur wissenschaftlichen Visualisierung in der Geoinformatik sowie Kartographie. Zwar verwischen die Grenzen zwischen Karten und anderen geo-räumlichen Darstellungen immer häufiger³³, doch während die klassische Kartographie eher statisch auf die Präsentation von geographischen Inhalten ausgerichtet ist, wird hier die Erforschung von Daten, Überprüfung von Hypothesen sowie die Entwicklung von Lösungen ins Zentrum gerückt. Durch diese Kombination, von Methoden aus Informationsvisualisierung und Kartographie, wird es möglich ein besseres Verständnis für Strukturen und geographische Prozesse aus großen Datenmengen zu generieren.

Eines der bekanntesten Beispiele und sogleich einfachste Art einer Geovisualisierung ist die Arbeit von Dr. John Snow, der zur Zeit der großen Choleraepidemie (1854) in London lebte. Seine These war, dass sich diese Epidemie nicht durch „Dünste“ sondern über das Trinkwasser ausbreitete. Er stellte zunächst eine Karte mit allen Straßen her und visualisierte darin jeden Toten mit einem Punkt (Abb.6). Daraufhin veranschaulichte er auch die Standorte aller

³¹ Ware (2004) – “Information Visualization: Perception for Design”

³² Prof. Dr. Schiewe – „geovisualisierung.net“

³³ Meng (2003a) – “Rahmenbedingungen beim Einsatz von Geovisualisierungsmethoden und –
techniken”(S.3)

öffentlichen Wasserpumpen. Dadurch konnte Snow eine starke Konzentration von Todesfällen im Einzugsbereich einer bestimmten Wasserpumpe nachweisen und diese als Hauptursache der Epidemie bestimmen.

Man kann also sehen wie eine funktionale Darstellung einzelner Geobjekte dazu beitragen kann, Hypothesen zu bestätigen.

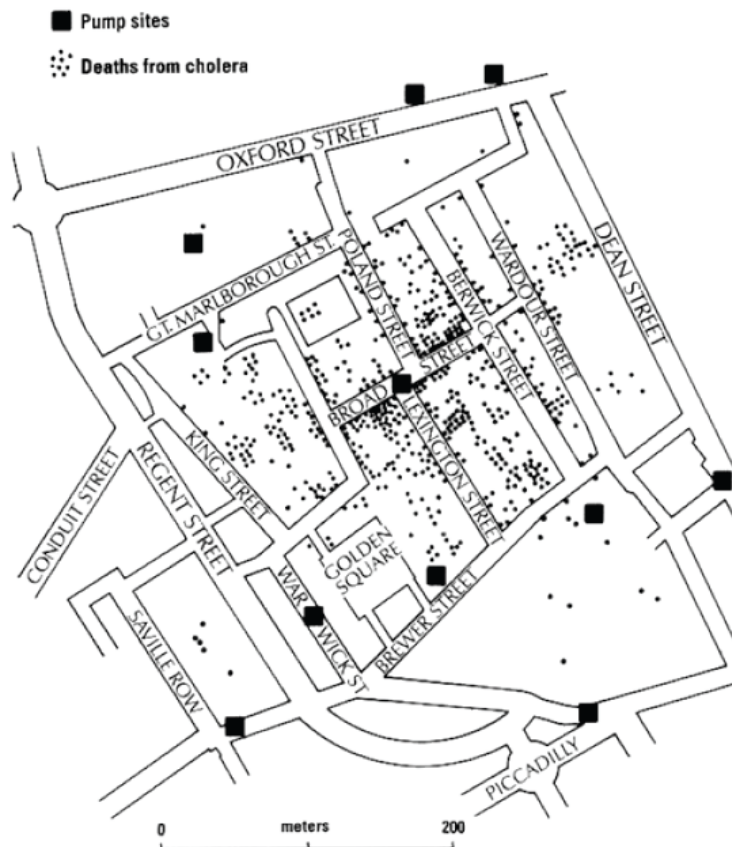


Abb. 6 – Darstellung von Choleraopfern und Standorten von Wasserpumpen (Quelle: Bação (2007) – „Geospatial Data Mining“ – S.7)

Ein weiteres Beispiel ist die Verdeutlichung von Sachverhalten als anamorphote Darstellung. Eine „kartenverwandte Darstellung, deren Geometrie proportional zur Ausprägung eines Merkmals verzerrt wird.“³⁴ Dadurch können verschiedene Sachverhalte eindrucksvoll veranschaulicht, aber auch bestimmte Probleme leichter erschlossen werden. Die Verwendung von SOM für diesen Einsatz wurde bereits von Henriques³⁵ vorgeschlagen und durchgeführt.

In Abbildung 7 wird zum Beispiel die Größe eines Landes verändert, indem der prozentuale Anteil der Einwohner dieses Landes im Vergleich zur Gesamtbevölkerung in die Berechnung einbezogen

³⁴ Bollmann; Koch (2002) – „Lexikon der Kartographie und Geomatik“

³⁵ Henriques; Bação; Lobo (2009) – „Carto-Som – Cartogram creation using self-organizing maps“

wird. Damit ist es möglich Zusammenhänge nicht nur statistisch sondern auch visuell zu interpretieren.

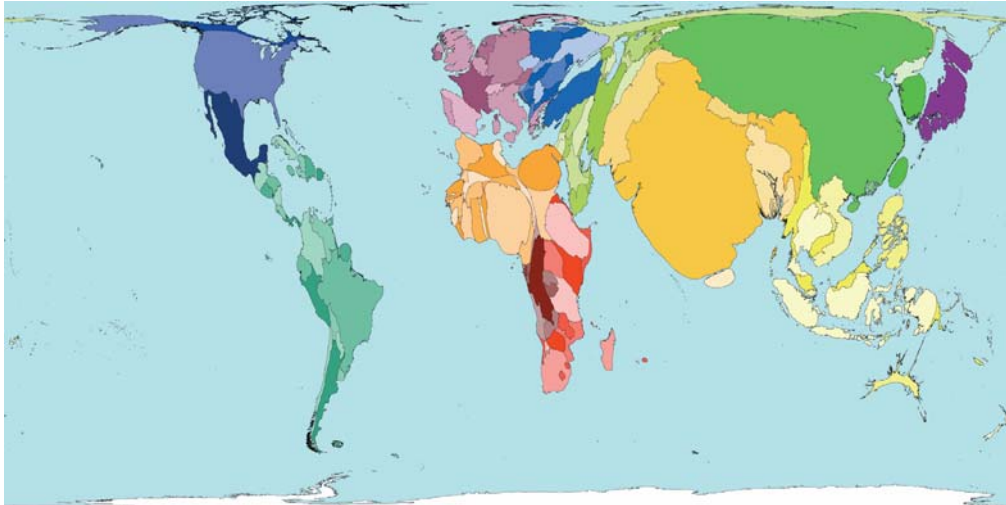


Abb.7 – Beispiel einer Kartenanamorphose zur Darstellung der Gesamtbevölkerung (Quelle: SASI Group; Newman (2006) – „www.worldmapper.org“)

Eine Geovisualisierung ist also nicht die pure objektive Beschreibung der realen Umwelt, wie es bei einer Karte der Fall ist. Vielmehr ist es die Darstellung von Geobjekten in Verbindung mit dem Wissen, welches man über diese gewonnen hat. Bei einer Karte ist die Interpretation des Dargestellten immer die Aufgabe des Nutzers. Bei einer Geovisualisierung wird dies meist vom Darsteller übernommen, da stets ein bestimmter Sachverhalt veranschaulicht werden soll.

Die grundlegenden Möglichkeiten, um SOM in diesem Sinne als Geovisualisierungswerkzeug zu verwenden, hat Koua³⁶ in seiner Arbeit „Using Self-Organizing Maps for Information Visualisation and Knowledge Discovery in Complex Geospatial Datasets“ bereits beschrieben. Diese möchte ich deshalb im Folgenden vorstellen und durch verschiedene Beispiele ergänzen.

Im Allgemeinen kann man vier grundlegende Darstellungstechniken herausstellen, die eine anschauliche Visualisierung der Self-Organizing Map ermöglichen:

- U-Matrix-Darstellung
- Projektionen (Netzdarstellungen)
- zwei- oder dreidimensionale Oberflächenplots der Distanzmatrizen
- Darstellung der Komponentenebenen

³⁶ Koua (2003) – “Using Self-Organizing Maps for Information Visualisation and Knowledge Discovery in Complex Geospatial Datasets”

4.1.1 U-Matrix-Darstellung

Die Unified Distance Matrix visualisiert, wie bereits beschrieben, die Entfernung jedes einzelnen Neuron zu jedem seiner Nachbarn des SOM-Netzes. Eine räumliche Metapher entsteht, indem die Neuronen, als Quadrat oder Hexagon dargestellt, hell und dunkel (z.B.: kleine und große Entfernungen) eingefärbt werden. Dadurch ergeben sich helle Cluster und dunkle Abgrenzungsbereiche, die man auch mit Bergen und Tälern assoziieren kann.

Koua³⁵ zeigt ein sehr anschauliches Beispiel der Overijssel Region in den Niederlanden (Abb.8). Ein Datensatz bestehend aus den Gemeinden der Region in Verbindung mit 29 sozioökonomischen Variablen (z.B.: Durchschnittseinkommen pro Einwohner).

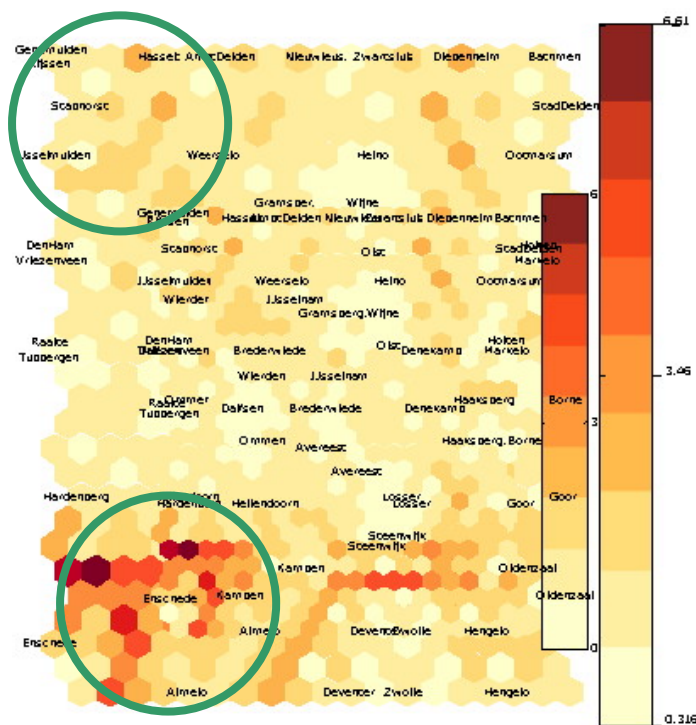


Abb.8 – U-Matrix-Darstellung des Overijseldatensatzes
(Quelle: Koua (2003) – “Using Self-Organizing Maps for Information Visualisation and Knowledge Discovery in Complex Geospatial Datasets” – S.5 der pdf-Version)

Man kann in diesem Beispiel (Abb.8) sehr gut erkennen, wie sich die Gemeinde Enschede klar von allen anderen abgrenzt. Dies liegt eindeutig darin begründet, dass sie, die am stärksten entwickelte und besiedelte sowie die größte Gemeinde der Region ist. Koua³⁷ ist ein weiterer sehr interessanter Fakt aufgefallen. Die Gemeinden, welche in der oberen Ecke nahe beieinander liegen, sind sich offensichtlich in ihren sozioökonomischen Faktoren sehr ähnlich. Doch obwohl im Datensatz keinerlei Variable vorhanden war, die

³⁷ Koua (2003) – “Using Self-Organizing Maps for Information Visualisation and Knowledge Discovery in Complex Geospatial Datasets” (S.5 der pdf-Version)

eine Aussage über die Religionszugehörigkeit der Einwohner vorgibt, sind alle diese Orte stark protestantisch geprägt.³⁶ Es ist also erkennbar, dass durch die Methode der SOM Zusammenhänge aufgedeckt werden können, die im Vorhinein nicht offensichtlich waren. Sie ergeben sich erst aus der Vielzahl von Variablen.

In einem weiteren Beispiel (Abb.9), sollen über den Vergleich 30 grundlegender Indikatoren des Gesundheits- und Lebensstandards von insgesamt 152 Ländern, Rückschlüsse auf den Entwicklungsstand dieser Länder gezogen werden.

Es handelt sich dabei um Daten des „human development index“ (HDI).

Die Gruppierung bestimmter, in ihrem Entwicklungsstand sehr ähnlicher Länder, ist hier wieder klar erkennbar. Sämtliche hoch entwickelten Industrienationen verdichten sich im untersten (rechten) Bereich der Darstellung. Wohingegen der Entwicklungsstand nach oben kontinuierlich abnimmt.

Zwei interessante Dinge sind mir in dieser Visualisierung aufgefallen. Zum einen befindet sich Kuba in einem gemeinsamen Cluster mit europäischen Ländern, wie Slowakei, Ungarn und Tschechien. Wirtschaftlich sind diese Länder nicht miteinander vergleichbar. Doch in diesem Kontext haben sie offensichtlich einen ähnlichen Entwicklungsstand. Die vorliegende Arbeit ist aus dem Jahre 2004. Betrachtet man den korrespondierenden „human development report“ (Abb.10), so liegen Kuba und die benachbarten Ländern tatsächlich nicht sehr weit auseinander.

Ebenso interessant ist meiner Meinung nach, die Einordnung von Kasachstan. Es befindet sich innerhalb eines sehr dunklen Bereiches, also einem Areal das auf große Entfernungen der benachbarten Neuronen schließen lässt.

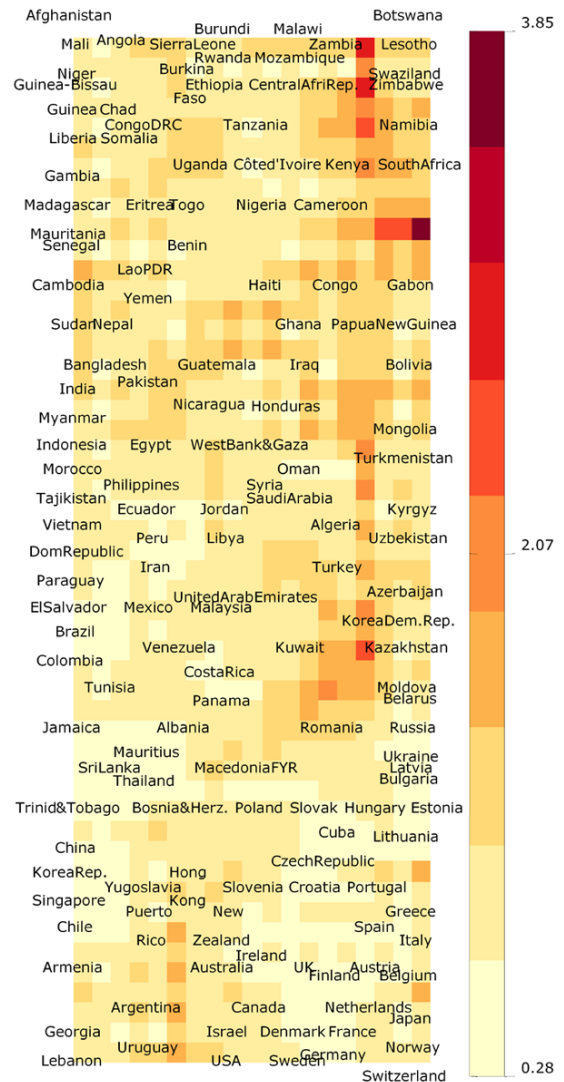


Abb.9 – U-Matrix-Darstellung zu Entwicklungsständen (Quelle: Koua (2004) – „Geovisualization to support the exploration of large health and demographic survey data“ –S.3 der pdf-Version)

Es liegt dementsprechend im Grenzbereich zwischen Moldawien und der Demokratischen Volksrepublik Korea. Warum sich dieser Sachverhalt so gestaltet, ist etwas das, meiner Meinung nach, weiter hinterfragt werden und somit einen Ansatz zur Findung neuer Erkenntnisse liefern könnte. In diesem Zusammenhang sehe ich einen Nachteil solch einer kombinierten Darstellung der U-Matrix und der direkten Verortung der Länder.

- 31 Malta
- 32 Czech Republic
- 33 Brunei Darussalam
- 34 Argentina
- 35 Seychelles
- 36 Estonia
- 37 Poland
- 38 Hungary
- 39 Saint Kitts and Nevis
- 40 Bahrain
- 41 Lithuania
- 42 Slovakia
- 43 Chile
- 44 Kuwait
- 45 Costa Rica
- 46 Uruguay
- 47 Qatar
- 48 Croatia
- 49 United Arab Emirates
- 50 Latvia
- 51 Bahamas
- 52 Cuba
- 53 Mexico
- 54 Trinidad and Tobago
- 55 Antigua and Barbuda

Abb.10 – HDI von 2004
(Quelle: Fukuda-Parr (2004) – „PressKit- Human Development Report 2004“ – S.2)

Im Vergleich zur Größe der Neuronen (Quadraten) sind die Ländernamen teilweise stark überdimensioniert. Betrachtet man z.B. die Platzierung von Finnland, so liegt es genau zwischen Großbritannien und Österreich. Doch worauf bezieht sich diese Signatur? Auf den ersten oder den letzten Buchstaben oder auf das Zentrum? Je nachdem wie man es betrachtet, würde Finnland eher dem einem oder dem anderen Land ähneln oder eben exakt dazwischen liegen.

Wie bereits beschrieben, deuten helle Bereiche in der U-Matrixdarstellung auf Cluster (Häufungen) hin und dunkle Areale grenzen diese voneinander ab. Somit kann man automatisiert Klassen bestimmen, und diese mit unterschiedlichen flächenhaften Farbcodierungen versehen, wie es in Abbildung 11 verdeutlicht wird.

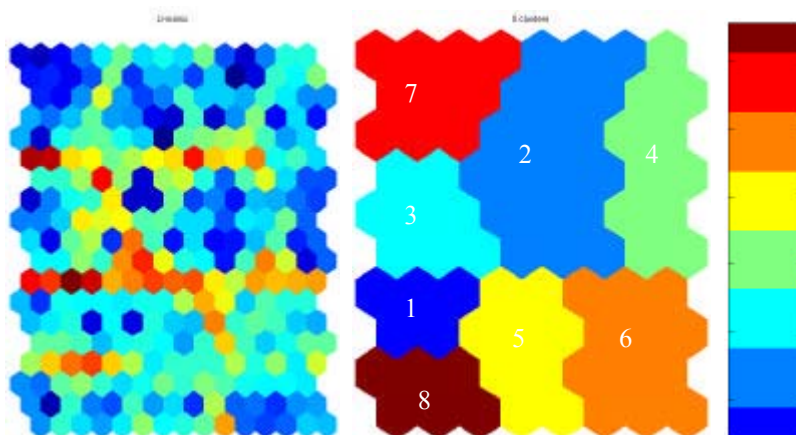


Abb.11 – Clusterdarstellung der U-Matrix
(Quelle: Oyana (2005) – „Exploration Of Geographic Information Systems-Based Medical Databases With Self Organizing Maps (Som): A Case Study Of Adult Asthma“ – S.16)

Auf den genauen Verlauf der Clusterbildung möchte aufgrund des Umfangs an dieser Stelle nicht eingehen. Vielmehr möchte ich nur das damit verbundene Potential der SOM für diesen Bereich verdeutlichen. Für die Clusteranalyse des in Abbildung 11 dargestellten Beispiels wurde K-means angewandt und die Cluster mittels „Davies-Bouldin validity index“ identifiziert. Es ist klar zu erkennen, dass ein erheblicher Interpretationsspielraum bei der Bildung der Cluster möglich ist. Sodass man davon ausgehen muss, dass ein Anwender gezwungen ist, durch empirische Tests das jeweils am besten, für sein gesetztes Ziel, geeignete Klassifikationsergebnis zu ermitteln.

Aus dem Klassifikationsansatz heraus, ist es möglich, ermittelte Werte für ein Flächenkartogramm zu nutzen und das Ergebnis der Ähnlichkeitsuntersuchung direkt in einer thematischen Karte darzustellen. In Abbildung 12 wird diese Möglichkeit für den bereits angesprochenen HDI-Datensatzes von Koua & Kraak angewandt.

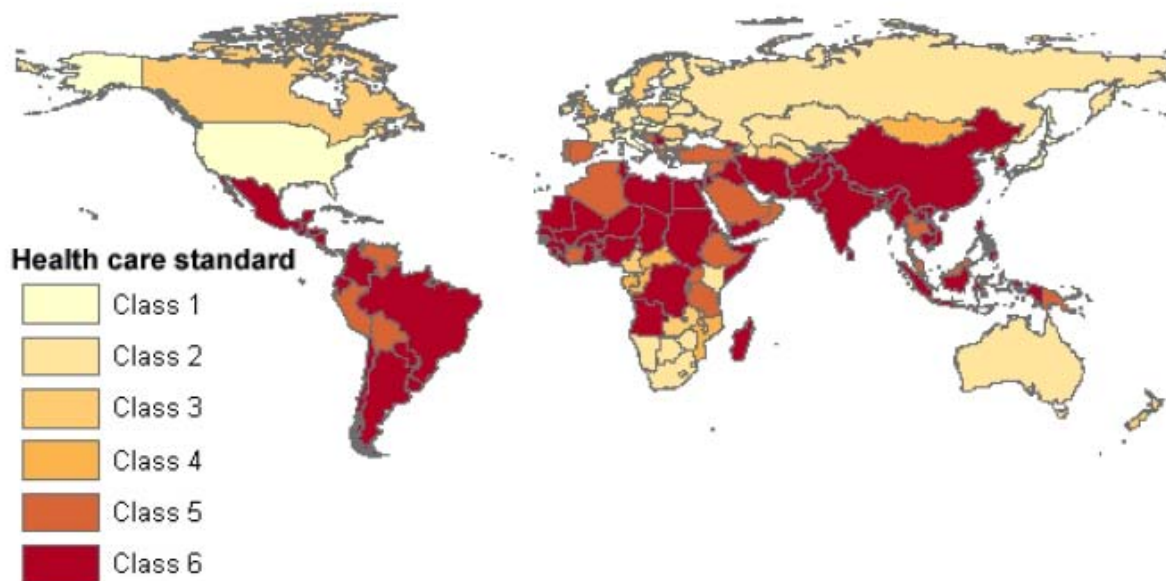


Abb.12 – Darstellung der aus U-Matrix bestimmten Klassen als Flächenkartogramm
(Quelle: Koua (2004) – „Geovisualization to support the exploration of large health and demographic survey data“ –5.4 der pdf-Version)

In einem weiteren Projekt der Universität von Lausanne wurde dieses Verfahren für einen Datensatz von 212 Gemeinden der schweizerischen Region Graubünden mit jeweils 26 demographischen Variablen angewandt. Deren Ergebnis wird in Abbildung 13 dargestellt.

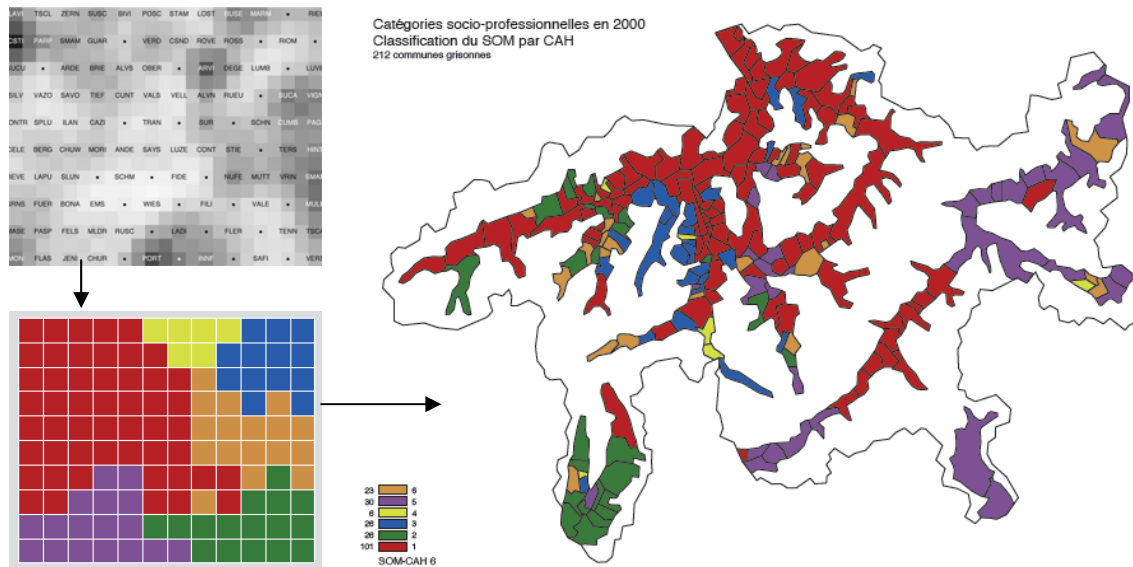


Abb.13 – Klassifizierung mittels U-Matrix und Visualisierung in einer Karte
(Quelle: Kanevski (2007) – „Classification and Visualization of High-Dimensional Socio-Economic Data Using Self-Organizing Maps“ – S.28 & S.30)

In Abbildung 13 wird der Interpretationsspielraum einer solchen Klassifikation sehr stark deutlich. Auf Anhieb ist es kaum möglich den Weg von der SOM zur Bildung der 6 Klassen nachzuvollziehen. Es hätten sich ebenso auch völlig andere Areale zu Clustern zusammenschließen können. Die Dokumentation zu diesem Beispiel ist leider sehr knapp, weshalb ich zur exakten Durchführung dieser Clusteranalyse keine weiteren Aussagen geben kann. Allerdings ist Abbildung 13 lediglich eine von vier Lösungen, welche von Kanevski ermittelt wurden. Bei einer Clusteranalyse von SOMs gibt es also offensichtlich keine ultimative Lösung.

4.1.2 Projektionen (Netzdarstellungen)

Die U-Matrix-Methode ist sehr gut geeignet, um Cluster abzuleiten. Sie liefert jedoch keinerlei Aussage über die Struktur des Datenraumes, da sie fest an das Gitter der SOM gebunden ist. Um die Form der SOM darstellen zu können, müssen deren Werte projiziert werden. Die am häufigsten verwendeten Projektionen sind dabei PCA (principle component analysis)(Abb.15) sowie „Sammon’s mapping“ (Abb.14). Bei der Darstellung des Netzes wird jedes Neuron durch einen Punkt abgebildet, wodurch ein „informatives Abbild der globalen Struktur der SOM“³⁸ gegeben werden soll. Die Abbildungen 14 und 15 zeigen jeweils Projektionen bezogen auf den bereits angesprochenen Overijssel-Datensatz von Koua. Die

³⁸ Koua (2003) – „Using Self-Organizing Maps for Information Visualisation and Knowledge Discovery in Complex Geospatial Datasets“ (S.5 der pdf-Version)

Farbgebung entspricht, wie im vorangegangenen Beispiel, der Zugehörigkeit zu einem bestimmten Cluster.

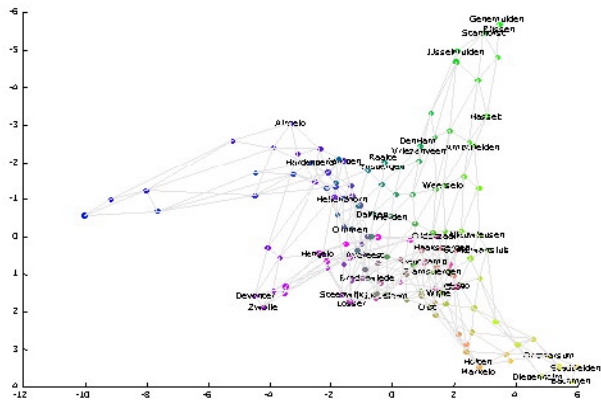


Abb.14 – Netzdarstellung mit „Sammon’s mapping“ des Overijseldatensatzes (Quelle: Koua (2003) – “Using Self-Organizing Maps for Information Visualisation and Knowledge Discovery in Complex Geospatial Datasets” - S.6 der pdf-Version)

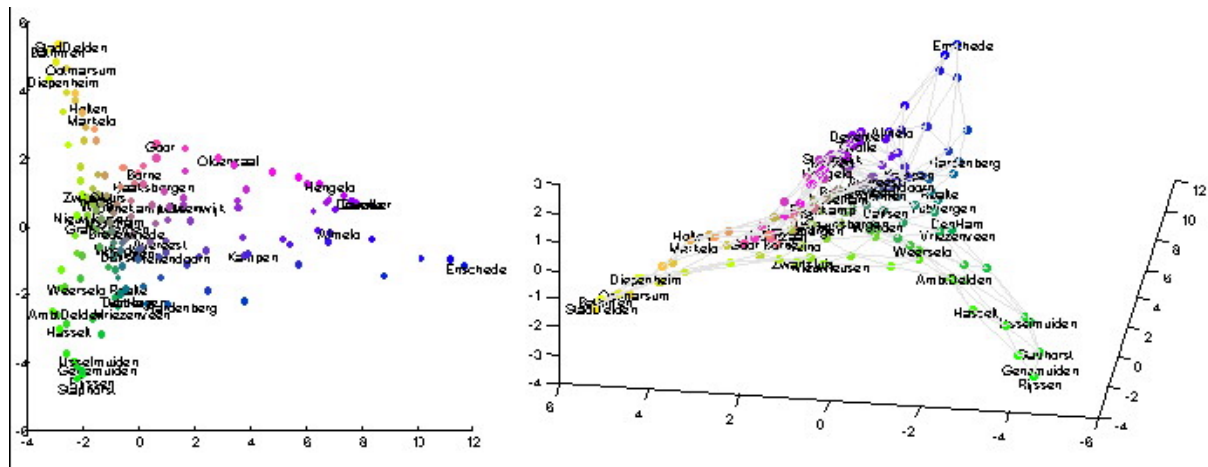


Abb.15 – Netzdarstellung mit PCA des Overijseldatensatzes (in 2D- und 3D-Ansicht) (Quelle: Koua (2003) – “Using Self-Organizing Maps for Information Visualisation and Knowledge Discovery in Complex Geospatial Datasets” - S.6 der pdf-Version)

Der Vorteil dieser Darstellung ist, dass hier dem Nutzer Möglichkeiten der Interaktion zur Verfügung gestellt werden können. Wie zum Beispiel Maßstabsfaktor, Projektionsart, Größe und Farbe der Punkte sowie Variationen in den Eigenschaften der Verbindungslinien. So ist diese Methode zur reinen, visuellen Analyse eines Datensatzes sehr viel besser geeignet als die Darstellung der U-Matrix. Die Lage eines Neurons ist hier exakt über die Lage des Punktes bestimmbar und der konkrete Abstand zu den Nachbarneuronen ist, durch eine Darstellung der Netzlinien, ebenfalls sehr gut abschätzbar. Genauso ist die Veränderung des Maßstabsfaktors ein sehr hilfreiches Mittel, sodass Bereiche mit sehr hoher Punktdichte in einem größeren Maßstab abgebildet und somit besser differenziert werden können.

4.1.3 2D & 3D-Plots

Eine weitere Möglichkeit die Form der SOM darzustellen, ist die Netzdarstellung, um eine Landschaftsmetapher zu erweitern. Statt lediglich die Position und Clusterzugehörigkeit (Farbe) der einzelnen Einheiten darzustellen, werden Oberflächen dazwischen interpoliert. Diese können anschließend, in Abhängigkeit vom Distanzwert, eingefärbt werden. So ergibt sich das Bild einer homogenen, zweidimensionalen „Landschaft“. Dadurch kann zusätzlich zur Form, auch die Dichte und Größe von Clustern (Klassen) visuell abgeschätzt werden. Das ist in der Netzdarstellung nicht ohne weiteres möglich. In Abbildung 16 kann man sehen, wie sich das Bild einer „Landschaft“ mit Bergen und Tälern assoziieren lässt.

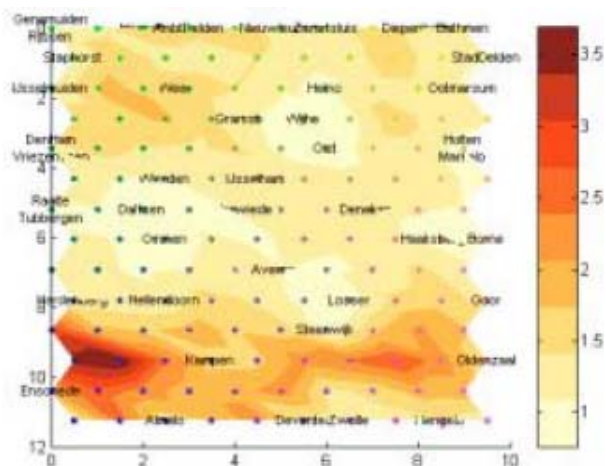


Abb.16 – Darstellung des Overijseld Datensatzes als 2D-Plotoberfläche (Quelle: Koua (2003) – “Using Self-Organizing Maps for Information Visualisation and Knowledge Discovery in Complex Geospatial Datasets” - S.7 der pdf-Version)

Des Weiteren ist es ebenso möglich, den Intensitätswert (Färbung) als z-Koordinate zu verwenden, um somit eine dreidimensionale Darstellung zu erhalten (Abb.17).

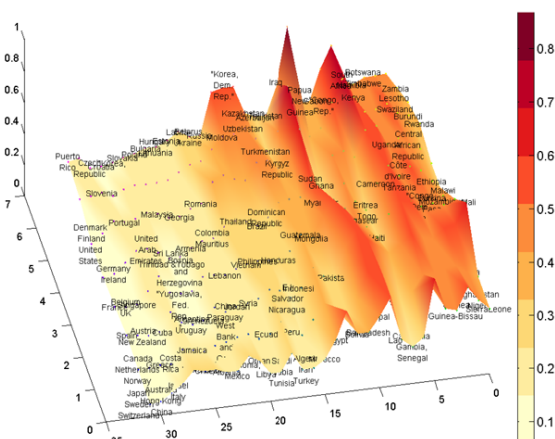


Abb.17 – Darstellung des HDI-Datensatzes als 3D-Plotoberfläche (Quelle: Koua (2004) – „Geovisualization to support the exploration of large health and demographic survey data“ -S.7 der pdf-Version)

4.1.4 Komponentenebenen

Die vorangegangenen Darstellungsmethoden der SOM sind eher auf das Erforschen von allgemeinen Mustern und Klassen (Clustern) ausgerichtet. Alternativ dazu ist es möglich, die Werte für jede einzelne Variable des Ausgangsdatensatzes darzustellen. Allgemein kann man sagen, dass die Komponentenebenen die Verteilung der Werte der einzelnen Variablen über das SOM-Gitter aufzeigen. Diese Ebenen, der einzelnen Komponenten, erhalten hier eine Farbkodierung für die Werte der einzelnen Neuronen (Abb.18). Im Gegensatz zur U-Matrix, wo die Ähnlichkeitswerte dargestellt werden. Dazu werden Minimum und Maximum der jeweiligen Komponente ermittelt und mit einem bestimmten Farbwert versehen. Im Beispiel werden dabei hohe Werte dunkel und niedrige hell eingefärbt. Dazwischen liegende Werte werden mittels einer gleichverteilten Farbskala kodiert. In Folge entsteht eine qualitative Einfärbung der einzelnen Elemente. Deren Lage ist durch die SOM festgelegt.

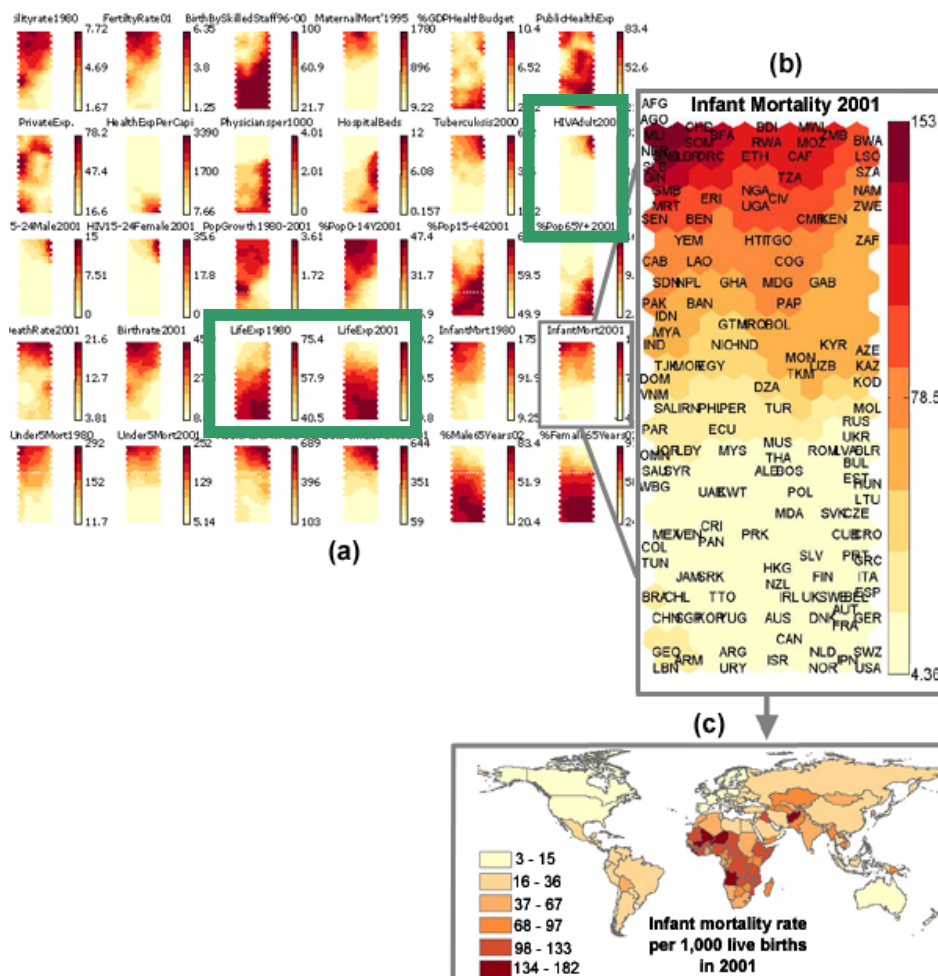


Abb.18 – Ableitung von Informationen aus Komponentenebenen für HDI-Datensatz (Quelle: Koua (2004) – „Geovisualization to support the exploration of large health and demographic survey data“ –5.8 der pdf-Version)

Aus der Position und Färbung der einzelnen Einheiten können Korrelationen und Beziehungen einzelner Neuronen erforscht werden.

So können zum Beispiel bestimmte Areale auf deren Färbung in unterschiedlichen Komponentenebenen untersucht und damit Assoziationen zwischen einzelnen Attributen gezogen oder Hypothesen angestellt werden. Um den Effekt der visuellen Erschließung von Korrelationen verschiedener Variablen zu verstärken, kann die Anordnung der Ebenen variiert und somit Komponenten, welche in Bezug zueinander stehen, nebeneinander dargestellt werden.

Im Beispiel des bereits vorgestellten HDI-Datensatzes (Abb.18-a) kann man diese Eigenschaft gut erkennen. Betrachtet man die Komponenten innerhalb der beiden grün umrandeten Felder, so fällt ein deutlicher Zusammenhang auf. Es handelt sich um die Rate der HIV-Infizierungen im Jahr 2001 sowie die durchschnittliche Lebenserwartung in den Jahren 1980 und 2001. Es ist ersichtlich, dass der bei der Komponente für HIV-Infektionen stark ausgeprägte Bereich, in den beiden Lebenserwartungskomponenten völlig unterschiedliche Ausprägungen aufweist. Die Lage der jeweiligen Einheiten, hier also der Länder, ist in den einzelnen Komponentenebenen immer dieselbe, wie auch in der U-Matrix-Darstellung (Abb.9). Somit kann man eindeutig eine Korrelation zwischen der Verbreitungsrate des HIV-Virus und der Lebenserwartung in den am stärksten betroffenen Ländern deutlich erkennen.

Ebenso ist es möglich, die Darstellung einer einzelnen Komponente zu analysieren und zu klassifizieren, um diese Ergebnisse für weitere statistische Auswertungen zu nutzen und in einer Karte darzustellen. Im Beispiel wird die Komponente der Kindersterblichkeit im Jahr 2001 zur Auswertung genutzt (Abb.18-b) und die Ergebnisse in einer Karte als Flächenkartogramm dargestellt (Abb.18-c).

Die Ableitung von Korrelationen ist in einem weiteren Beispieldatensatz ebenfalls gut erkennbar. Dieser besteht aus 8 sozioökonomischen Indizes³⁹, bezogen auf die Verwaltungsbezirke von São Paulo: familiar income (IFH), educational development (ED), educational stimulus (ES), longevity (LONG), environmental quality (EQ), home quality (PQ), concentration of family headed by women (CIWFH) und concentration of family headed by illiterate women (CWFH) (Abb.19).

³⁹ da Silva (2004) – "Visualization of Geospatial Data by Component Planes and U-Matrix" (S.7)

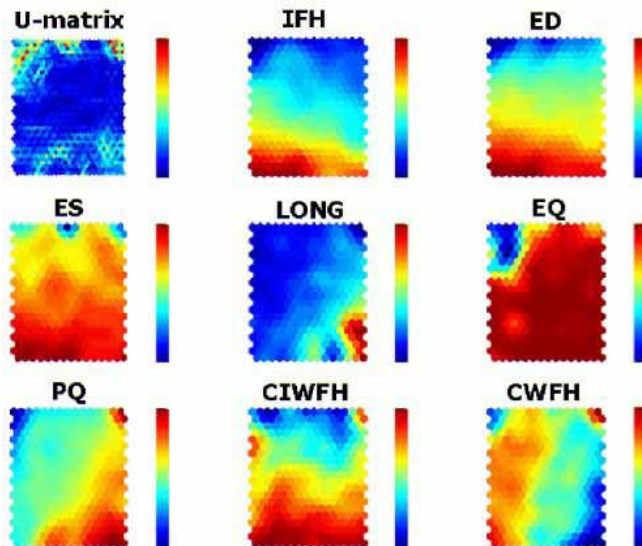


Abb.19 – Darstellung von U-Matrix & Komponentenebenen der 8 Indizes
 (Quelle: da Silva (2004) – “Visualization of Geospatial Data by Component Planes and U-Matrix” – S.12)

Die relativ ähnliche Ausprägung der Komponenten IFH und ED lässt, rein visuell, darauf schließen, dass sie miteinander korrelieren müssen. Das bedeutet, dass in Regionen mit hohem Durchschnittseinkommen auch der Bildungsstand höher ist als in anderen Regionen. Dies scheint durchaus plausibel zu sein. Die Komponente EQ (Umweltqualität) sticht durch ihre homogene Ausprägung deutlich heraus. Daraus lassen sich folgende Schlussfolgerungen ziehen. Zum einen kann deren Einfluss auf die Berechnung der Ähnlichkeitswerte nicht sehr hoch sein. Zum anderen wird aber auch deutlich, dass die Umweltqualität in nahezu allen Bezirken gleich hoch ist.

Eine exaktere qualitative oder quantitative Aussage lässt sich aus dieser Abbildung leider nicht ableiten. Allerdings habe ich im vorangegangenen Beispiel gezeigt, inwiefern das daraus gezogene Wissen weiter verwendet werden könnte.

Es handelt sich bei den vorgestellten Beispielen um sehr kleine Datensätze, die von den Autoren jeweils zur Verdeutlichung und zum Testen verwendet wurden. Doch indem bereits hier nützliches Wissen und Anregungen gefunden werden konnten, wird deutlich, dass in wesentlich größeren und unübersichtlicheren Datensätzen das Potential dafür noch höher sein dürfte.

Es lässt sich also zusammenfassend feststellen, dass die Methode der SOM ein geeignetes Werkzeug für die Visualisierung von geowissenschaftlichen Datenbanken ist. Je nachdem inwiefern diese Datenbanken untersucht werden sollen, stehen verschiedene Mittel zur Verfügung. Möchte man Cluster (Klassen) in den Datensätzen identifizieren, so ist die Darstellung der U-Matrix am Besten

geeignet. Für reine Visualisierungen des Datensatzes oder der SOM sind Netzdarstellung (Projektion) sowie die Oberflächenplots optimal. Zusätzlich ist die Möglichkeit der Abgrenzung von Clustern ist, bei den Oberflächenplots, sehr vorteilhaft.

Möchte man Korrelationen und Beziehungen in Datensätzen entdecken so empfiehlt sich in jedem Fall eine Visualisierung und Verarbeitung über die einzelnen Komponentenebenen der SOM. Auf eine weitere alternative Darstellungsform der Komponentenebenen werde ich in Kapitel 6 noch einmal eingehen.

4.2 Geo-SOM & andere Möglichkeiten zur Verarbeitung von geowissenschaftlichen Daten

Im vorangegangenen Kapitel wurde bereits deutlich, dass SOM ein gutes Werkzeug zur explorativen Datenanalyse ist und somit die Möglichkeit zur Verfügung stellt ein erweitertes Wissen über vorhandene Datenbestände zu erlangen. Doch bisher wurde SOM lediglich verwendet, um geowissenschaftlich interessante Daten zu verarbeiten ohne ihren jeweiligen Raumbezug direkt in die Analyse einzubeziehen. Dementsprechend möchte ich im folgenden Kapitel auf verschiedene Varianten der SOM eingehen, wo der geographische Raumbezug der Daten in die Kalkulation einbezogen wurde. Eine umfassende Analyse der vorhandenen Möglichkeiten wurde bereits von Bação, Lobo & Painho⁴⁰ durchgeführt, weshalb ich mich auf deren Ergebnisse beziehen und ihre Beispiele vorstellen werde.

Der grundlegende Gedanke, hinter der Einbindung von geographischen Koordinaten in die Berechnungen der SOM, ist die Beachtung des „1st law of geography“⁴¹ von Tobler. Sodass geographische Nachbarn in benachbarten Bereichen der SOM klassifiziert werden. Gerade in den geographischen Informationswissenschaften ist Klassifikation eine kombinierte Suche von Ähnlichkeiten im geographischen sowie im Merkmalsraum. Nicht nur der Charakter einer Entität ist dabei von Belang, sondern auch deren Interaktionen im geographischen Raum. Der allgemeine Ansatz ist, die geographischen Koordinaten als regulären Attributwert einer Entität in die Berechnung der SOM einzubeziehen.

⁴⁰ Bação (2004) – “The self-organizing map, the Geo-SOM, and relevant variants for geosciences”

⁴¹ Tobler (1970) – “A computer movie simulating urban growth in the Detroit region” (S.3 der Onlineversion)

Die bedeutendsten von Bação, Lobo & Painho³⁸ vorgeschlagenen SOM-Varianten sind:

- Hierarchische SOMs
- Geo-enforced SOM
- Geo-SOM

4.2.1 Hierarchische SOMs

Hierarchische SOMs können immer angewendet werden wenn, der betreffende Datensatz oder auch das gestellte Problem, eine Zerlegung in untergeordnete Teilebenen oder –probleme es möglich oder notwendig machen. Ein sehr anschauliches Beispiel ist die Spracherkennung mittels SOM. Hier wird die Sprache in unterschiedliche Ebenen, wie Phoneme, Silben und Wörter unterteilt. Dabei werden zunächst die unteren Ebenen verarbeitet, die Ergebnisse an die jeweils höhere Ebene übergeben und weiterverarbeitet. Bação, Lobo & Painho schlagen die hierarchischen SOMs zur Verwendung bei demographischen Problemen vor.⁴² Die unteren Ebenen können, nach unterschiedlichen Themen sortiert und klassifiziert werden, um die Gesamtklassifikation dann in einer übergeordneten Ebene zu veranschaulichen. Die Struktur solcher hierarchischer SOMs wird in Abb.20 dargestellt.

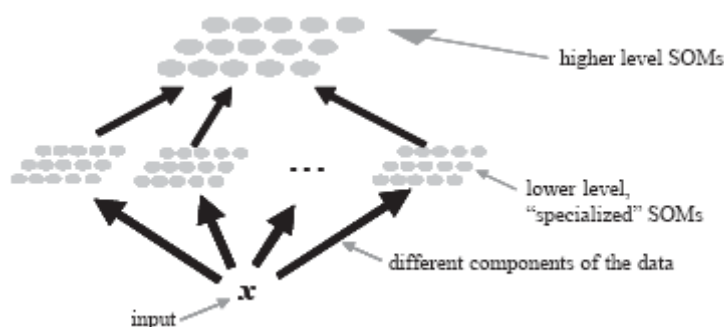


Abb.20 – Struktur von hierarchischen SOMs
(Quelle: Bação (2004) – "The self-organizing map, the Geo-SOM, and relevant variants for geosciences" – S.3)

Man könnte beispielsweise aus einem sozioökonomischen Datensatz verschiedene Komponenten zusammenfassen, um geordnete Klassifikationen für Themen wie Altersverteilung, Wohnbedingungen oder Bildungsniveau zu erhalten. Eine übergeordnete SOM entnimmt verschiedene Eingabeattribute aus den untergeordneten SOMs und gibt eine einzelne übergeordnete

⁴² Bação (2004) – "The self-organizing map, the Geo-SOM, and relevant variants for geosciences" (S.3-4)

Klassifikation dieser aus. Somit ist es zum einen möglich, sich einen allgemeinen Überblick über die Klassifikation eines Datensatzes zu verschaffen. Zum anderen kann man aber auch spezifische geographische Merkmale in einem untergeordneten thematischen Kontext betrachten. Gerade in einer computergestützten Umgebung, wo es möglich ist, die verschiedenen Ansichten dynamisch und interaktiv miteinander zu verbinden, ist eine Veranschaulichung von Datensätzen dann optimal möglich.

Es zeigt sich, dass die Verwendung von hierarchischen SOMs für die Geowissenschaften nur indirekt nützlich ist. Das Interesse liegt eher in einer zielorientierten Verarbeitung und Darstellung, während eine direkte Einbindung der geographischen Koordinaten jedoch nicht vorgesehen ist. Tests haben ergeben, dass gerade eine auf einzelne Ziele aufgeteilte Klassifikation von demographischen Geodaten wesentlich effektiver ist, als sehr allgemein gehaltene Klassifikationen solcher Daten.⁴³

4.2.2 Geo-enforced SOM

Im Gegensatz zur zuvor gezeigten Methode werden bei der Geo-enforced SOM die geographischen Koordinaten der Daten bei der Erstellung der SOM nicht außen vor gelassen. Vielmehr werden die Koordinaten mit einem bestimmten, vom Anwender festgelegten, Gewicht direkt in die Berechnung der SOM eingebunden. Sie werden quasi als eine weitere Variable verarbeitet. Je nachdem wie stark das Gewicht der geographischen Information angegeben wird, desto stärker ist auch ihr Einfluss auf das Ergebnis.

Baçaõ, Lobo & Painho zeigen dafür beispielhaft zwei mögliche Anwendungen auf.⁴⁴ Zum einen könnte man, wenn sich die geographischen Merkmale auf Gebiete oder Areale beziehen, die Koordinaten des korrespondierenden Centroids mit den anderen Variablen kombinieren. Dadurch wird dann die Lage eines Elementes innerhalb der SOM auch wesentlich durch deren Lage im reellen Raum beeinflusst. Wodurch dann die Zuordnung bei einer anschließenden Klassifikation indirekt beeinflusst wird.

Zum anderen ist es möglich die reellen Distanzwerte von Merkmalen zu weiteren bedeutenden geographischen Merkmalen mit einzubeziehen. Würde man zum Beispiel demographische Daten deutscher Landkreise verarbeiten, so hätte eine Variable, welche deren Abstand zu bedeutenden Ballungszentren wiedergibt,

⁴³ Openshaw; Wymer (1995) – „Classifying and regionalizing census data“ (S.4)

⁴⁴ Baçaõ (2004) – “The self-organizing map, the Geo-SOM, and relevant variants for geosciences” (S.4)

sicherlich einen starken Einfluss auf das ermittelte Klassifikationsergebnis.

Bei dieser Methode handelt es sich um keinerlei Modifikation des SOM-Algorithmus. Es ist vielmehr eine Möglichkeit, während der Vorverarbeitung der Daten, das Endergebnis zu beeinflussen.

4.2.3 Geo-SOM

Die Methode der Geo-SOM stellt eine verallgemeinerte Kombination von zwei zuvor entwickelten SOM-Architekturen dar. Zum einen der Hypermap⁴⁵, die von Kohonen zur Spracherkennung entwickelt wurde. Dazu werden ausschließlich bestimmte Teile der Eingabewerte zum Finden der BMU verwendet. Zum anderen die Kangasarchitektur⁴⁶, bei der lediglich eine geringe Anzahl von Nachbarn des Gewinnerneurons bei der Berechnung beachtet wird. Es wird also deutlich, dass es sich bei dieser Methode um eine direkte Modifikation des originalen SOM-Algorithmus handelt.

Abgeleitet aus Toblers „1st law of geography“⁴⁷, wird bei dieser Methode das Gewinnerneuron bestimmt, indem lediglich Neuronen betrachtet werden, die sich im geographischen Raum innerhalb einer bestimmten Nachbarschaft des Eingabevektors befinden. Um dies zu realisieren haben Bação, Lobo & Painho⁴⁸ eine zusätzliche Variable t eingeführt. Diese gibt den geographischen Toleranzbereich wieder. In Abbildung 21 wird dieses Auswahlverfahren verdeutlicht.

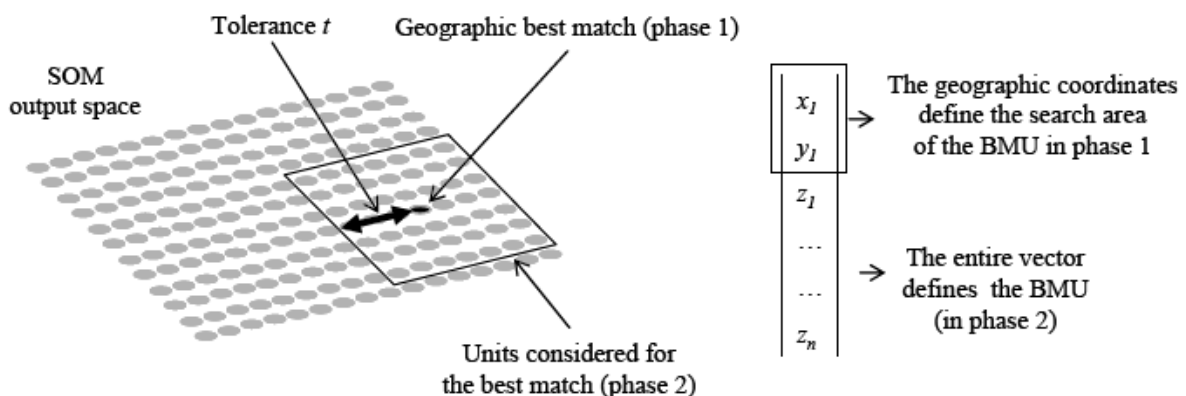


Abb.21 – Auswahlverfahren der Geo-SOM

(Quelle: Bação (2005b) – "Geo-SOM and it's Integration with Geographic Information Systems" – S.2)

⁴⁵ Kohonen (1991) – "The hypermap architecture"

⁴⁶ Kangas (1992) – "Temporal knowledge in locations of activations in a self-organizing map"

⁴⁷ Tobler (1970) – "A computer movie simulating urban growth in the Detroit region" (S.3 der Onlineversion)

⁴⁸ Bação (2004) – "The self-organizing map, the Geo-SOM, and relevant variants for geosciences" (S.4)

Dieses Auswahlverfahren des Gewinnerneurons läuft in zwei Stufen ab. Als erstes wird ein vorläufiges Gewinnerneuron ausschließlich nach den geographischen Koordinaten des Eingabevektors ausgewählt. Anschließend wird innerhalb des gewählten Tolleranzbereiches t das endgültige Gewinnerneuron bestimmt. Die Anpassung der umliegenden Neuronen erfolgt nach den normalen Regeln der SOM. Damit werden Einheiten, welche im reellen Raum nahe beieinander liegen, gezwungen, sich ebenfalls im Ausgaberaum in unmittelbarer Nähe zueinander zu befinden. Durch den Einsatz dieser Methode können Klassen entstehen, die sich geographisch sehr nahe sind. In Abbildung 22 wird dieser Zusammenhang deutlich.

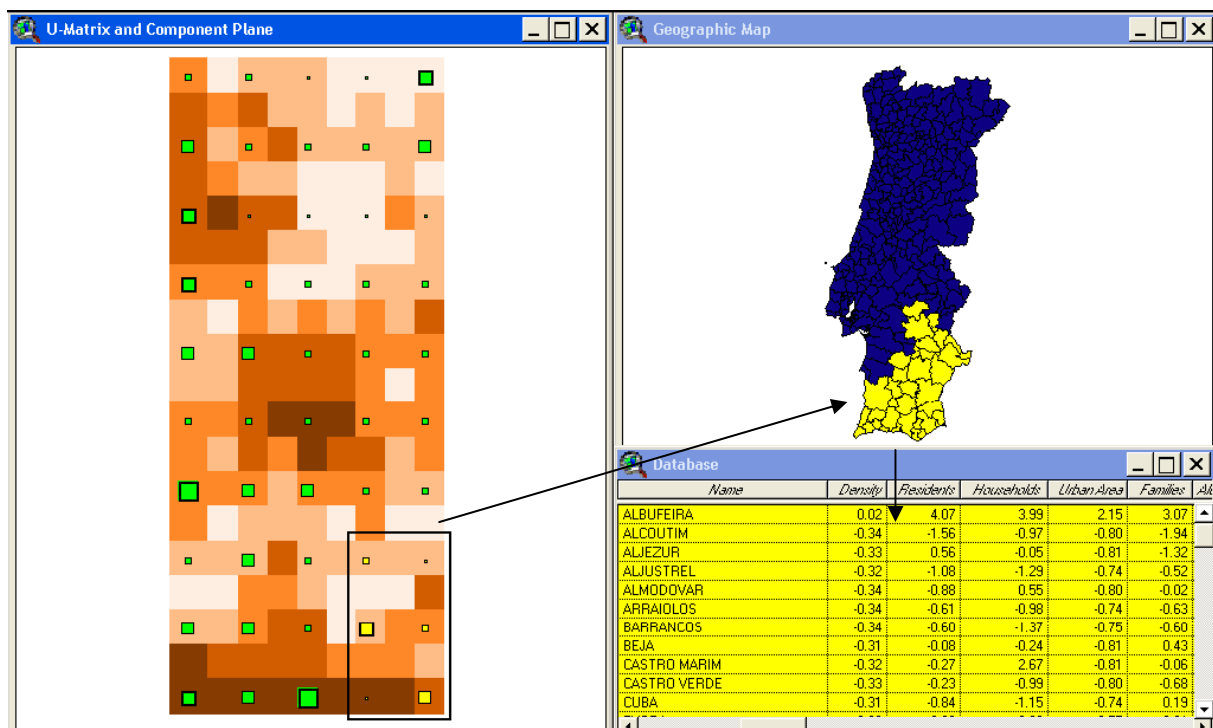


Abb.22 – Anwendungsumgebung für Geo-SOM
(Quelle: Bação (2005b) – “Geo-SOM and it’s Integration with Geographic Information Systems” – S.5)

Es handelt sich hierbei um eine kombinierte Darstellung der Geo-SOM aus einer U-Matrix, einer Vektordarstellung sowie einer Datenbankansicht der Ausgangsdaten. Man kann deutlich erkennen, wie die Auswahl eines Areals in der Geo-SOM sich ebenfalls auf benachbarte Gebiete in der Kartendarstellung bezieht.

In Abbildung 23 ist der Unterschied zwischen den verschiedenen vorgestellten Methoden erkennbar. Links eine Geo-enforced SOM, in der Mitte eine Geo-SOM mit $t=0$ und rechts eine Geo-SOM mit $t=1$.

Der Beispieldatensatz beinhaltet 250 Regionen mit jeweils 65 charakteristischen Variablen. Zusätzlich wurden die x- & y-Koordinaten des geometrischen Centroids, von jedem Element, in den Datensatz eingefügt. In allen drei Klassifikationen ist die Region Braga (rot) in einer Gruppe mit verschiedenen anderen Regionen (gelb).

Es ist klar zu erkennen, welchen Einfluss die unterschiedlichen Algorithmen auf das Ergebnis haben. Aus dem Standardalgorithmus entstehen Klassen, die deutlich auf den Attributwerten basieren. Denn die anderen Elemente der Gruppe sind vergleichbare Metropolregionen. Für das Ergebnis der Geo-SOM mit $t=0$ sieht man, dass die geographische Lage, den Schwerpunkt bei der Klassenbildung setzt. Abschließend kann man sagen, dass die Geo-SOM mit $t=1$ eine Art Kompromiss aus Attributwerten und geographischer Lage darstellt. Es gibt lediglich zwei weitere Regionen in der Klasse. Diese sind, innerhalb eines bestimmten geographischen Toleranzbereiches, die ähnlichsten Regionen.⁴⁹

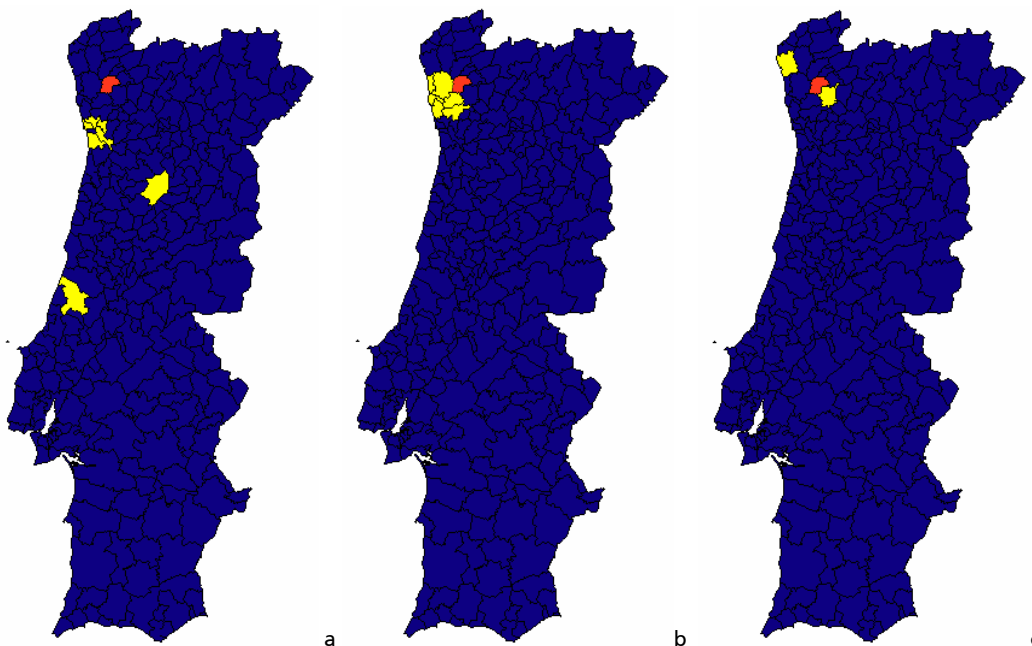


Abb.23 – Vergleich von Geo-enforced SOM (a), Geo-SOM mit $t=0$ (b) & $t=1$ (c)
(Quelle: Bação (2005b) – “Geo-SOM and it’s Integration with Geographic Information Systems” – S.7)

Damit kann man die Geo-SOM als eine Methode verstehen, die multidimensionale Daten in den geographischen Raum projiziert.⁵⁰ Der Anwender kann über eine Variierung der Toleranz t , das am besten für sein Problem angepasste Ergebnis erzielen. Der herausragende Vorteil, der sich aus diesem Algorithmus ergibt, ist die Möglichkeit geographische Merkmale in Klassen

⁴⁹ Bação (2005b) – “Geo-SOM and it’s Integration with Geographic Information Systems” (S.7)

⁵⁰ Bação (2004) – “The self-organizing map, the Geo-SOM, and relevant variants for geosciences” (S.8)

zusammenzufassen, welche in ihren Attributen ähnlich sind und sich gleichzeitig innerhalb eines bestimmten geographischen Raums befinden.

4.3 SOM & GIS

Die Methoden zur Auswertung der SOM in Kapitel 4.1 deuten bereits zahlreiche Möglichkeiten an, inwiefern sich das Modell in ein GIS einbinden lässt, um damit das Erforschen von Daten auf unentdeckte Informationen zu erleichtern. Ebenso sieht man in Abbildung 22 wie es möglich ist SOM und GIS zu kombinieren. Deshalb möchte ich im Folgenden auf die Möglichkeiten zur Implementierung der SOM nur noch einmal kurz eingehen.

Laut Definition wird ein Geoinformationssystem „als Software verstanden, mit dessen Hilfe Geodaten erfasst, verwaltet und ausgegeben werden können“⁵¹. SOM kann dabei mit seinem dimensionsreduzierenden Charakter zum Aufspüren von Ähnlichkeiten und Beziehungen eingesetzt werden. Um demzufolge ein verbessertes Verständnis bei der Ausgabe von Geodaten zu ermöglichen.

Eine verknüpfte Visualisierung einzelner Darstellungen ist, meiner Meinung nach, das beste Mittel um diese Anwendung in ein GIS einzubinden. Ähnlich zu Abbildung 22 ist eine Darstellung der Daten als Karte die Grundlage für weitere Repräsentationen. Darauf basierend kann man nun eine Darstellungsform der SOM, abhängig vom Ziel der Datenerforschung, damit verknüpfen. Dadurch wird der Überblick über die Attributwerte der Geodaten sowie deren Ähnlichkeiten und Relationen untereinander möglich. Die zusätzliche Ausgabe des Datenblattes ist eine optionale Möglichkeit, welche der Anwender anfügen kann, wenn ein detaillierter Einblick in die Attributwerte gewünscht ist. In Abbildung 22 sieht man sehr gut, wie man einen Bereich in der auswählen kann und den direkten Bezug dazu in der Karte eingebildet bekommt. Ebenso ist dies auch in die entgegengesetzte Richtung vorstellbar. Sodass man Bereiche in der Karte auswählt und diese auf deren Ähnlichkeitswerte hin überprüft.

Gleichermaßen könnte ich mir auch vorstellen, dass man hierarchische SOMs in dieser Form visualisiert. In Folge erhält man einen detaillierten Einblick in verschiedene, selbst erstellte Themenbereiche der Daten und kann diese weiter erforschen.

⁵¹ Bollmann; Koch (2002) – „Lexikon der Kartographie und Geomatik“

Überdies ist erkennbar wie in den Abbildungen 12, 13 und 18 Klassifizierungsergebnisse für eine statistische Darstellung in einer thematischen Karte eingesetzt werden können.

Es ist vorstellbar, jede Art der aufgezeigten Geovisualisierungen in ein GIS einzubinden und daraus Erkenntnisse zu gewinnen. Sodass Geodaten besser klassifiziert oder visuell erschlossen werden können. Auf die technische Umsetzung möchte ich nicht im Detail eingehen. Erste Versuche wurden bereits von Fincke, Lobo & Bação⁵² veröffentlicht und ausgewertet. Sie haben die Softwarelösungen SOM_PAK und SOM toolbox verwendet und deren Ergebnisse, mithilfe von Java, in ein von ArcGIS lesbares Textformat umgewandelt. Ebenso haben Bação, Lobo & Painho⁵³ die Implementierung der Geo-SOM-Methode vorgestellt (vgl. Abb.22).

⁵² Fincke (2008) – “Visualizing self-organizing maps with GIS”

⁵³ Bação (2005b) – “Geo-SOM and it’s Integration with Geographic Information Systems”

5. Datenverarbeitende Anwendungen

Im vorangegangenen Kapitel habe ich Möglichkeiten aufgezeigt, in wiefern Visualisierungen im Bereich der Geowissenschaften umgesetzt werden können. Dazu habe ich diese möglichen Methoden eher formell vorgestellt und zusätzlich einzelne Beispiele aus der Literatur aufgezeigt.

In diesem Kapitel möchte ich jedoch keine theoretischen Methoden vorstellen. Vielmehr werde ich praktisch durchgeführte Projekte zur Datenverarbeitung mittels SOM aufzeigen und analysieren.

5.1 Klassifizierung von Fernerkundungsdaten

Die Dimensionalität, Menge und Heterogenität von Fernerkundungsdaten ist ein Bereich, der immer innovative Techniken erfordert, um effektiv Informationen extrahieren zu können. Eine auf neuronalen Netzen basierende Klassifikation ermöglicht dabei die Verwendung von spektralen sowie räumlichen Merkmalen als Eingangsvariablen.

Einleiten möchte ich diesen Themenbereich mit der Arbeit von Evangelou⁵⁴, in der das Ergebnis aus einer klassischen Klassifizierung mit dem Ergebnis einer SOM-Klassifizierung verglichen wird. Genutzt wurde dafür der Bildausschnitt eines Landsat 4-TM Band4-Bildes vom April 1984, mit der Aufnahme des unteren Themsetales südwestlich von Heathrow Airport. Eine Szene, die laut Autoren, von zahlreichen Inlandwasserflächen, Autobahnen sowie Industriegebieten geprägt ist.⁵⁵ Diese Variation unterschiedlicher urbaner und suburbaner Oberflächenbedeckungen enthält damit aussagekräftige Beispiele für anthropogene, geologische und hydrogeologische Merkmale. Das Bild wurde klassisch vorbereitet. Das bedeutet, es wurde geometrisch und radiometrisch korrigiert und im Anschluss in 624 x 516 Pixel große Ausschnitte zerlegt, die dann weiter bearbeitet wurden.

Zur Bearbeitung wurde die Standardversion des SOM-Algorithmus verwendet und die Klassifizierung als „Segmentierung des Bildes in aussagekräftige Regionen“⁵⁶ verstanden. Eine detailliertere Erklärung zum Ablauf der Segmentierung wird in dieser Arbeit nicht gegeben. Zum Vergleich wurde zusätzlich ein Bildausschnitt mit einem konventionellen Algorithmus klassifiziert. Eine exaktere

⁵⁴ Evangelou (2001) – „Data Mining and Knowledge Discovery in Complex Image Data using Artificial Neural Networks“

⁵⁵ Ebd. (S.2)

⁵⁶ Ebd. (S.4)

Aussage dazu, als dass es sich um einen klassischen Algorithmus handelt, wurde ebenfalls nicht angegeben.

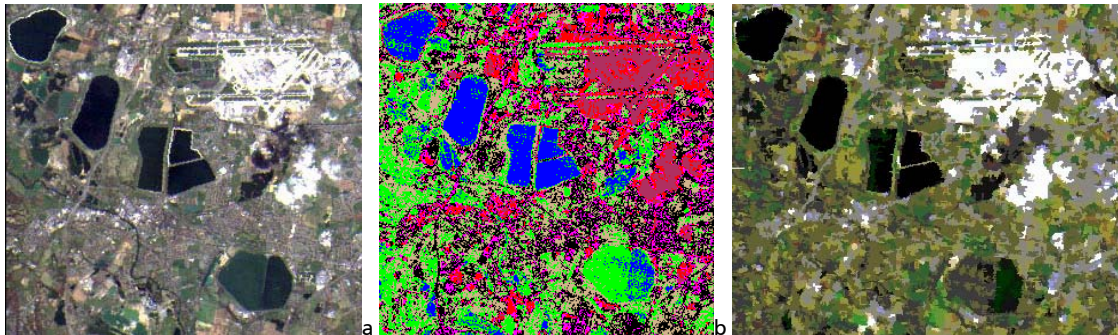


Abb.24 - Klassifikationsergebnisse eines Landsat-4TM Band4-Bildausschnittes des unteren Themseales vom 12. April 1984

a – Bildausschnitt nach Vorverarbeitung
b – Ergebnis konventioneller Klassifikation
c – Ergebnis einer SOM-Klassifikation

(Quelle: Evangelou (2001) – „Data Mining and Knowledge Discovery in Complex Image Data using Artificial Neural Networks“ – S.7-8)

Im Ergebnis liefern beide Techniken 7 Klassen. Diese sind, mittels SOM, vollständig ohne vordefiniertes Ziel und jegliche Art von Nutzereingriff entstanden. Lediglich die Benennung der einzelnen Klassen wurde nachträglich durchgeführt.

Die Methode der SOM ist bei der Interpretation von Satellitenbildern sehr hilfreich, da ausdrucksstarke Informationen durch Vereinfachung des Bildes geliefert werden. In den Ergebnissen treten leichte Klassifizierungsfehler auf, weshalb eine weitere Feinjustierung des Algorithmus notwendig ist.⁵⁷ Eine exakte Aussage über den Grad der Genauigkeit der erzielten Klassifikationsergebnisse wird leider nicht geliefert. Eine visuelle Interpretation der gezeigten Abbildungen ist jedoch auch nicht ohne weiteres möglich, da leider die beiden Klassifikationsergebnisse mit unterschiedlichen Farbcodierungen versehen sind. Man könnte lediglich sagen, dass das Gebiet des Heathrow Airport durch die SOM-Klassifikation deutlicher wiedergegeben wird als durch die konventionelle Klassifikation.

Allgemein geben, laut Evangelou, die positiven Ergebnisse eine Basis für weitere Forschungen, wodurch eine höhere Genauigkeit und Effektivität erreicht werden könnte.⁵⁸

Dagegen hat der Ansatz zur Landnutzungsklassifikation von Mielens⁵⁹, ebenfalls basierend auf Landsat TM- Daten, eine vergleichsweise sehr geringe Klassifikationsrate von 74% ergeben. Wodurch er zu dem Schluss gelangt, dass klassische Methoden, wie

⁵⁷ Evangelou (2001) – „Data Mining and Knowledge Discovery in Complex Image Data using Artificial Neural Networks“ (S.8-9)

⁵⁸ Ebd. (S.8-9)

⁵⁹ Mielens (2008) – „Self-Organizing Maps for Land Cover Classification“

MLP und Entscheidungsbäume, zur Klassifizierung besser geeignet sind.⁶⁰

Ähnliche Ergebnisse erhält Carrão⁶¹ in seiner Studie zur Landnutzungsklassifikation von MERIS-Daten mittels SOM. Er stellt fest, dass SOM ein sehr leistungsfähiger Algorithmus ist, der jedoch durch die enorme Anzahl an einstellbaren Parametern einen sehr erfahrenen Anwender benötigt, um akzeptable Ergebnisse zu liefern. Des Weiteren schlägt er die Verwendung mehrerer SOMs vor, wodurch die Identifizierung einzelner Klassen verbessert werden könnte.⁶²

Diese Erweiterungen liefern Neagoe & Ropot⁶³ mit der „Concurrent Self-Organizing Map“(CSOM). Diese Methode basiert auf einem überwachten Ansatz, indem eine Sammlung von kleinen SOMs aus zuvor ausgewählten Trainingsgebieten erstellt wird. Somit wird für jedes auftretende spektrale Muster eine spezielle SOM trainiert. Zur Erkennung dieser Muster wird ein Testmuster jeder trainierten SOM präsentiert. Es wird die SOM mit dem geringsten Quantisierungsfehler ermittelt. Abschließend wird deren Index, basierend auf einem globalem „winner-takes-all-Prinzip“⁶⁴, diesem Muster zugeordnet. Die Klassifizierungsrate von multispektralen Satellitenbildern soll dadurch noch höher sein als bei der Verwendung einer einzelnen SOM.

Die Autoren haben zum Testen und Trainieren ein Landsat TM-Bild mit 7 spektralen Bändern genutzt und dabei 7 thematische Kategorien mittels Trainingsgebieten vorklassifiziert. Somit wird jedes multispektrale Pixel, in diesem Beispiel, durch einen 7-dimensionalen Vektor charakterisiert, welcher die Pixelwerte aus jedem Band beinhaltet (Abb.25).

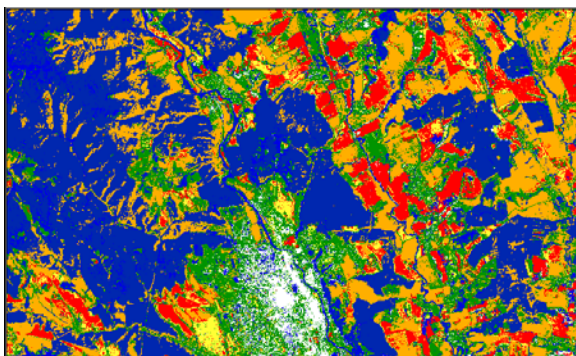


Abb.25 - Klassifikationsergebnis unter Verwendung einer circularen CSOM-Architektur (Klassifizierungsrate: 95,29%)
(Quelle: Neagoe (2006) – „Pattern Recognition in Multispectral Satellite Images Using Concurrent Self-Organizing Modular Neural Networks“ – S.5)

⁶⁰ Mielens (2008) – „Self-Organizing Maps for Land Cover Classification“ (S.7)

⁶¹ Carrão (2006) – „Meris Based Land Cover Classification with Self-Organizing Maps: Preliminary Results“

⁶² Ebd. (S.7)

⁶³ Neagoe (2006) – „Pattern Recognition in Multispectral Satellite Images Using Concurrent Self-Organizing Modular Neural Networks“

⁶⁴ Ebd. (S.3)

Zur Einordnung der Ergebnisse dieser neuen Methode, haben die Autoren⁶⁰ ebenfalls das multispektrale Bild mithilfe des einfachen SOM-Algorithmus klassifiziert. Im Ergebnis stellen sie fest, dass das beste Ergebnis (95,29% Klassifikationsrate) für das CSOM-Modell unter Verwendung von 7 circularen SOMs mit jeweils 112 Neuronen erreicht wurde. Zwar liefert das beste Ergebnis der Klassifikation mit einer einfachen SOM ebenfalls ein akzeptables Ergebnis von 94,31%, allerdings benötigt die CSOM eine wesentlich geringere Trainingszeit.

Es zeigt sich, dass Neagoe & Ropot eine sehr effektive Erweiterung der Methode zur Klassifizierung von Satellitenbildern entwickelt haben.

Der ursprüngliche SOM-Algorithmus bietet bereits sehr viel Potential für eine Anwendung zur Klassifikation. Doch sieht man, dass dessen Ergebnisse noch nicht zufriedenstellend sind. Darum komme ich zu dem Schluss, dass man zur Auswertung „simpler“ Satellitenbilder, mit klassischen Klassifikationsverfahren bessere Ergebnisse erzielt. Wenn man jedoch mit sehr hochdimensionalen Daten arbeitet, dann bietet sich die Erweiterung des CSOM-Modells auf jeden Fall an. Ich könnte mir vorstellen, dass man zusätzlich zu den reinen spektralen Merkmalen auch räumliche und radiometrische Merkmale mit einbezieht. Ähnlich dem Modell der Geo-SOM⁶⁵. Dadurch könnte man sicherlich noch höhere Genauigkeiten erzielen.

5.2 Kantendetektion in Satellitenbildern

Die Identifizierung von Kanten ist ein klassischer Teilbereich der Bildverarbeitung. Der Anspruch liegt darin, flächenhaft verbreitete Bereiche eines Bildes voneinander abzutrennen. Im Detail ist ein Pixel Teil einer Kante, wenn die Grauwerte in dessen direkter Nachbarschaft eine hohe Variation aufweisen. Sind diese Pixel entlang eines Pfades miteinander verbunden so, deutet dies auf eine Kante hin. Klassisch ist dies möglich indem ein lokaler Operator (Filtermatrix) über das Bild geführt und somit jedes Pixel neu berechnet wird.

Ein völlig neuer Ansatz ist die Erkennung von Kanten mittels neuronalen Netzwerken. Dazu möchte ich nun die Arbeit von Marques & Chen⁶⁶ vorstellen. Ziel dieser Arbeit ist es, mediterrane Wasserwirbel zu erkennen, die mit geringer Vermischung teilweise

⁶⁵ Bação (2004) – “The self-organizing map, the Geo-SOM, and relevant variants for geosciences”

⁶⁶ Marques (2003) - „Border detection on remote sensing satellite data using self-organizing maps“

tausende kilometerweit in den Atlantik eindringen und damit Effekte auf die Hydrologie und Dynamik des Atlantiks haben. Dafür nutzen sie thermale Satellitenbilder, welche die Oberflächentemperatur in Form einer Matrizendarstellung der einzelnen Messpunkte abbilden. Jedes Element gibt die Temperatur der Oberfläche einer 1,1km x 1,1km großen Region wieder. Die Identifizierung der Wirbel soll, nahezu automatisiert, über eine Erkennung der Kanten mittels SOM-Algorithmus stattfinden.

Es gibt bereits einige Arbeiten zu diesem Ansatz. Diese sind jedoch teilweise nur für eine spezielle Anwendung geeignet (Reyes-Aldasoro⁶⁷) oder setzen ein bestimmtes Maß an vorgegebenem Wissen voraus (Brejl⁶⁸). Marques & Chen stellen jedoch eine Möglichkeit vor, Kanten ohne jegliches a-priori-Wissen zu erkennen. Diese ist, laut eigener Aussage, allgemein gehalten und kann somit auch in anderen Bereichen eingesetzt werden.⁶⁹

Die Grundidee besteht in der Verwendung von SOM, um Pixel in einem Cluster zusammenzufassen. Die Pixel am Rande dieser Cluster deuten dann auf die gesuchten Kanten hin. Dazu wird zunächst für jedes Pixel ein Merkmalsvektor erstellt und daraus eine SOM initialisiert und trainiert. Die Eingabewerte werden anschließend dem Gewinnerneuron zugeordnet. Durch die Topologieerhaltung der SOM stellen die Entfernungen der Neuronen die Ähnlichkeiten der Eingabewerte dar, wodurch sich Cluster ausbilden. In Abhängigkeit eines vom Nutzer eingestellten Schwellenwertes können die Neuronen an den Clustergrenzen erkannt werden. Die entsprechenden Pixel dieser Neuronen können folglich als Kantenpositionen im Bild identifiziert werden. Durch Isolierung und Einzeldarstellung dieser Pixel erhält man in Folge eine Abbildung der Kanten. Der vollständige Ablauf wird in Abbildung 26 verdeutlicht.

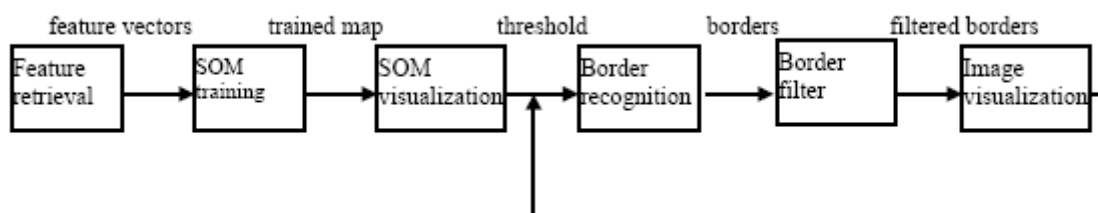


Abb.26 - Ablauf der Kantendetektion mittels SOM nach Marques & Chen
(Quelle: Marques (2003) - „Border detection on remote sensing satellite data using self-organizing maps“ – S.4)

⁶⁷ Reyes-Aldasoro (2000) – „Image segmentation and compression using neural networks“

⁶⁸ Brejl (1999) – „Medical image segmentation: Automated design of border detection criteria from examples“

⁶⁹ Marques (2003) - „Border detection on remote sensing satellite data using self-organizing maps“ (S.2)

Die Auswahl des Schwellenwertes spielt eine entscheidende Rolle. Um den besten Schwellenwert empirisch zu ermitteln, sind deshalb die letzten drei Schritte iterativ angelegt. Dieser darf nicht zu klein oder zu groß sein. Ansonsten könnten die Kanten zu stark (bis hin zu falschen Kanten) beziehungsweise zu schwach dargestellt werden oder gänzlich fehlen. Dadurch ist der Prozess nicht vollständig automatisiert. Denn der Nutzer muss die Teilschritte ständig wiederholen, um das Kantenbild mit der höchsten Aussagekraft zu ermitteln. Marques & Chen schlagen deshalb, für zukünftige Arbeiten, eine automatisierte Bestimmung des optimalen Schwellenwertes vor.

Ebenso haben die Initialisierungswerte der SOM Einfluss auf das Ergebnis der Kantenerkennung. Deshalb wurden viele Varianten auf bestimmte Kriterien getestet und deren Ergebnisse in einer Tabelle verdeutlicht (Tab.1). Es zeigt sich, dass für den Zweck der Kantenerkennung eine SOM mit einer kreuzförmigen Nachbarschaftsfunktion und einem Gewicht des zentralen Pixels zwischen 6 und 20, in Bezug auf die gestellten Kriterien, die besten Ergebnisse liefert. In Bezug auf die Größe der Karte, die Anzahl der Trainingsschritte oder auch die Lernrate werden in dieser Arbeit leider keinerlei Aussagen geliefert. So kann man nicht sagen, ob diese ebenfalls variiert wurden und inwiefern deren Werte einen Effekt auf das Endergebnis hätten. Das finde ich sehr schade, da diese ebenfalls einen großen Einfluss ausüben.

Neighborhood region		Noise robustness	Continuity	Definition	Threshold detection
Point		very bad	very bad	good	bad
Square		good	good	bad	bad
Cross	weight < 6	bad	bad	good	bad
	6 ≤ weight ≤ 20	good	good	good	good
	weight > 20	good	good	bad	good

Tab.1 - Auswertung der Ergebnisse

(Quelle: Marques (2003) - „Border detection on remote sensing satellite data using self-organizing maps“ – S.10)

Zur Einordnung der Resultate aus der Kantendetektion mittels SOM wurde eine Weitere mit der klassischen „Canny“-Methode⁷⁰ durchgeführt. Laut Autoren eine sehr zuverlässige und erfolgreiche Methode um Kanten zu erkennen.⁷¹ In Abbildung 27 sieht man das Ausgangsbild (a), die U-Matrix-Visualisierung der SOM (b), die Ergebnisse der Kantendetektion mittels SOM für 2 verschiedene Schwellenwerte (c & d) sowie die Ergebnisse der Kantendetektion mittels „Canny“-Operator für 2 verschiedene Schwellenwerte (e & f).

⁷⁰ Canny (1986) – “A Computational Approach to Edge Detection”

⁷¹ Marques (2003) - „Border detection on remote sensing satellite data using self-organizing maps“ (S.7)

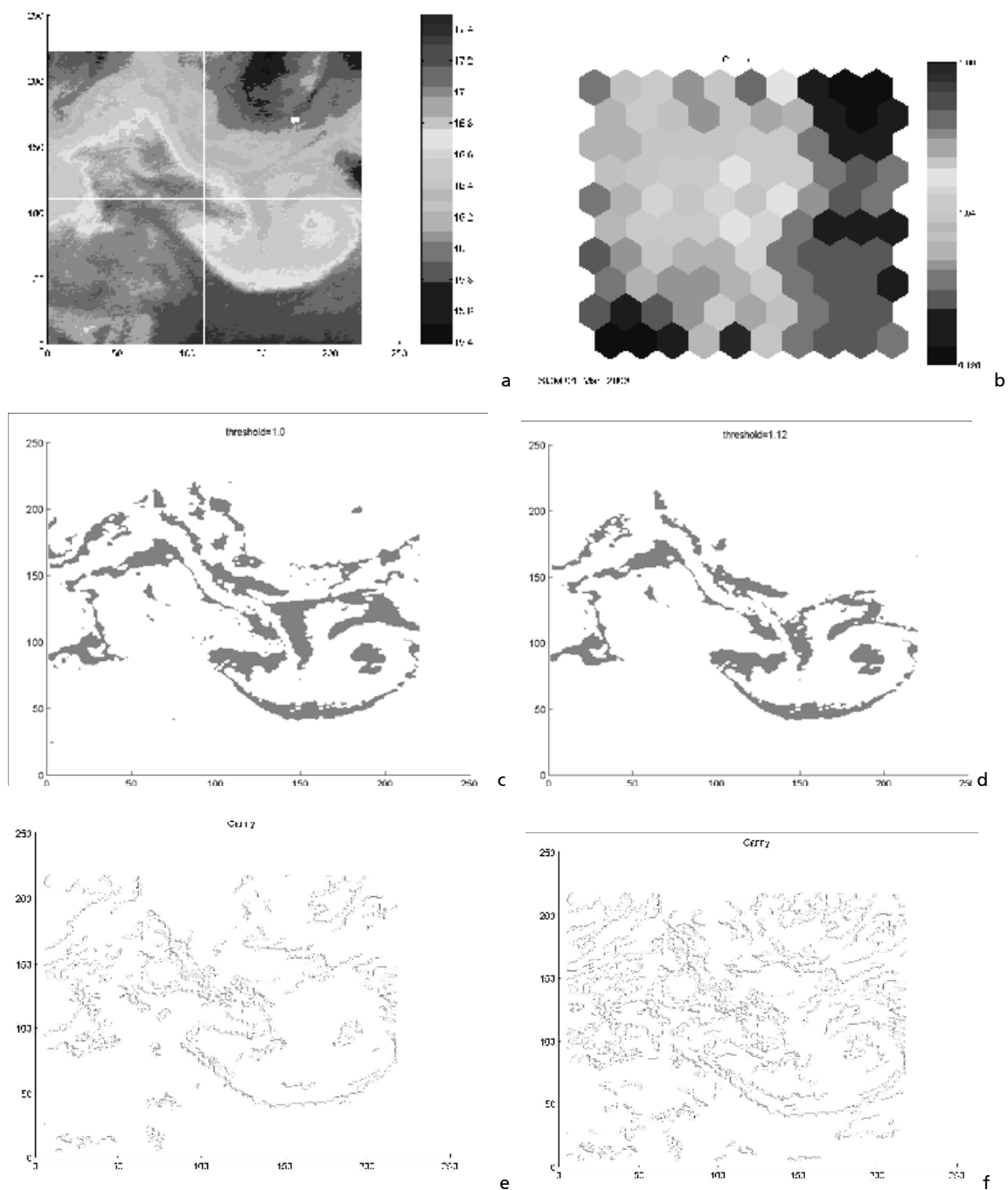


Abb.27 - Ausgangsbild (a), die U-Matrix-Visualisierung der SOM (b), Ergebnisse der Kantendetektion mittels SOM für 2 verschiedene Schwellenwerte (c & d) sowie die Ergebnisse der Kantendetektion mittels „Canny“-Operator für 2 verschiedene Schwellenwerte (e & f)
 (Quelle: Marques (2003) - „Border detection on remote sensing satellite data using self-organizing maps“ – S.8)

Es ist zu erkennen, dass die Kantenerkennung mittels SOM im Allgemeinen ein wesentlich signifikanteres und kontinuierlicheres Kantenbild liefert. Die „Canny“-Methode versagt im direkten Vergleich, durch ein zu starkes Rauschen und Vernachlässigen einiger Kanten. Es fällt jedoch auf, dass der Kantenverlauf nicht mittels Linien dargestellt wird. Es werden vielmehr Flächen abgebildet. In

der klassischen Bildverarbeitung werden Kantenverläufe extrahiert, indem das Bild auf Helligkeitsunterschiede hin untersucht wird. Dort wo starke Differenzen auftreten wird dann eine Kante vermutet. Bei der Kantenerkennung mittels SOM werden allerdings, wie zuvor beschrieben, nicht die Übergangsbereiche im Bild sondern in der U-Matrix mittels Clusteranalyse ermittelt. Dementsprechend werden alle Pixel, die sich in dem Übergangsbereich zwischen Clustern befinden als Kantenpixel identifiziert. So kommt es, dass hier flächenhafte Bereiche als Kanten identifiziert werden. Man müsste diese eher als Randbereiche bezeichnen. Im Vergleich mit dem Ausgangsbild wird deutlich, dass es sich tatsächlich um die Randbereiche der Wirbel handelt. Diese zeichnen sich durch eine relativ homogene Farbinformation aus. Der Vorteil ist, dass man so „Ausreißer“ gut identifizieren kann, da diese deutlich abseits der Randbereiche liegen. Damit scheint diese Methode sich sehr robust gegenüber von Rauschen zu verhalten.

Zusammenfassend kann man sagen, dass diese vorgestellte Methode sehr beeindruckende Ergebnisse liefert und damit, wie schon Marques & Chen feststellen, eine Basis für weitere Forschungen liefert. Gerade als vollständig selbständig arbeitendes Verfahren, mittels automatisierter Schwellenwertermittlung, würde die Kantendetektion mit SOM, für die digitale Bildverarbeitung enorme Möglichkeiten bieten.

5.3 Auswertung von Zeitreihen & Monitoring

Die Auswertung von Bildern einer Zeitreihe zur Überwachung bestimmter Oberflächen über einen längeren Zeitraum ist ebenfalls ein Bereich, in dem die Technik der SOM bereits angewendet wurde. Beginnen möchte ich mit einem Beispiel, in welchem der SOM-Algorithmus nicht direkt zur Auswertung, sondern vielmehr zur Vorverarbeitung der Ausgangsdaten genutzt wurde.

Abdel Latif & Mercier⁷² haben in ihrer Arbeit SOM genutzt, um die Auswertung von Zeitreihen aus MODIS-Bildern zu verbessern. Sie wollten saisonale Veränderungen der Vegetationsbedeckung einer bestimmten Region untersuchen. Dazu verwendeten sie Aufnahmen mit einer zeitlichen Auflösung von 24 Stunden. Problematisch ist, dass eine Vielzahl der Aufnahmen durch unregelmäßig auftretende Wolken und deren Schatten verunreinigt sind. Dadurch sind Datenlücken vorhanden.

⁷² Abdel Latif (2006) – “Self-Organizing Map for erroneous data processing in time series analysis”

Dementsprechend entstand die Idee, mittels SOM, die Wolken zu erkennen, zu entfernen und die fehlenden Bereiche zu ersetzen. Zur Charakterisierung des regulären, zeitlichen Verhaltens der überwachten Landschaft wird die SOM anhand von Gebieten trainiert, welche nicht durch Wolken gestört werden. Das eigentliche Untersuchungsgebiet wird darauf projiziert und daraus die Werte der fehlerhaften Gebiete ermittelt.

Im Ergebnis erhalten sie einen Datensatz, der frei von Wolken ist und die zeitliche Auswertung der Vegetationsbedeckung besser ermöglicht. Eine vollständige Auswertung über die Qualität der Korrekturen findet sich leider in dieser Arbeit nicht. In Abbildung 28 sieht man den Vergleich zwischen einer stark von Wolken bedeckten Aufnahme (a) sowie dem korrigierten Bild (b).

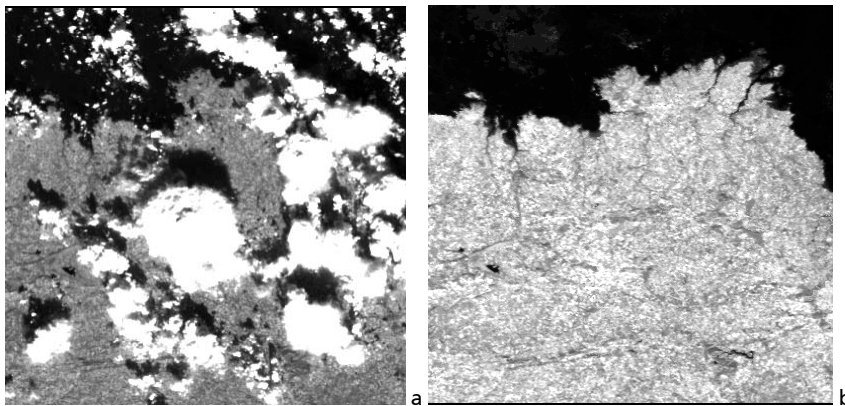


Abb.28 – Vergleich von Ausgangsbild (mit 50x20 SOM) und korrigierter Aufnahme
(Quelle: Abdel Latif (2006) – "Self-Organizing Map for erroneous data processing in time series analysis" – 5.3-4)

In einer weiteren Arbeit zeigen Abdel Latif u.a.⁷³ dahingegen die direkte Charakterisierung der Oberfläche über Zeitreihen mittels SOM. Sie verwenden dazu MODIS-Aufnahmen und zusätzlich hochauflösende SPOT- & Landsat-Aufnahmen zur Auswertung. Untersucht wurde die Vegetationsbedeckung während der Wintermonate. Mithilfe des SOM-Algorithmus wurden diese anschließend klassifiziert. Um das Potential der SOM einschätzen zu können hat man die Ergebnisse mit einem einfachen NDVI verglichen. Im Ergebnis zeigt sich, dass mittels SOM eine wesentlich höhere Gesamtgenauigkeit erreicht werden kann.⁷⁴ Diese liegt für hochauflösende Aufnahmen bei 90%. Die Autoren sind der Meinung, dass diese weiter gesteigert werden könnte, wenn noch mehr Bilder der Zeitreihen verwendet würden.

Die Ausführungen zu dieser Thematik sind sehr knapp gehalten, weshalb sich leider keine detaillierteren Aussagen treffen lassen.

⁷³ Abdel Latif (2007) – "Self-Organizing Map for surface characterization in time series"

⁷⁴ Ebd. (S.4)

Jedoch ist bereits die Vorverarbeitung von mit Wolken bedeckten Aufnahmen eine sehr eindrucksvolle Präsentation des Potentials der SOM.

5.4 Klassifikation von SAR-Daten

Zu beiden folgenden Arbeiten liegen mir lediglich die Zusammenfassungen vor, weshalb ich an dieser Stelle deren Inhalt nur kurz aufzeigen möchte. Für eine Auswertung der Ergebnisse liegen mir nicht genügend Informationen vor.

SOM besitzt auch Potential für die Auswertung von Daten des Synthetic Aperture Radar. So haben Nagao u.a.⁷⁵ in ihrer Studie SOM genutzt, um Treibeis in SAR-Aufnahmen zu klassifizieren und vom umgebenden Meerwasser zu unterscheiden. Da hierbei noch Fehlklassifikationen aufgetreten sind haben sie die Erweiterung zur „multistep SOM“ entwickelt, um diese fehlerhaften Daten zu korrigieren. Abbildung 29 zeigt das Ergebnis aus deren Klassifikation.

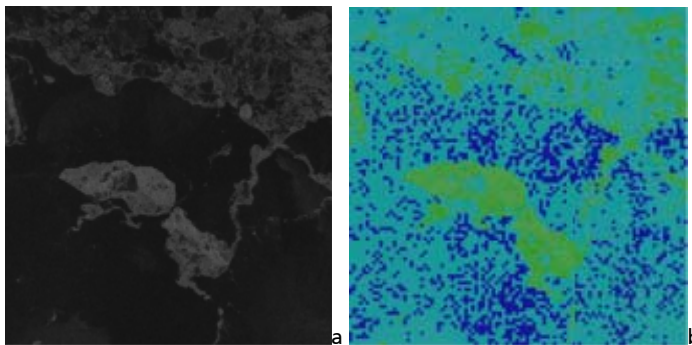


Abb.29 - a.) original SAR-Aufnahme b.) Klassifikationsergebnis mit dickem & dünnem Eis sowie umgebendes Wasser (Quelle: Nagao (2005) – „Drift Ice Classification Using SAR Image Data by a Self-Organizing Neural Network“ – S.1(Abtract))

Del Frate u. a.⁷⁶ zeigen in ihrer Arbeit die Klassifikation von Oberflächenbedeckung aus SAR-Aufnahmen mithilfe von SOM. Zur Auswertung vergleichen sie die Ergebnisse einer Wishart-Klassifikation⁷⁷ und kommen zu dem Schluss, dass SOM vergleichbare Ergebnisse mit einer wesentlich feineren Abgrenzung verschiedener Anbauarten liefert.

⁷⁵ Nagao (2005) – „Drift Ice Classification Using SAR Image Data by a Self-Organizing Neural Network“

⁷⁶ del Frate (2007) – “Self-organizing maps of polarimetric SAR imagery”

⁷⁷ Ebd. (S.1)

5.5 Generalisierung

Die kartographische Generalisierung ist im Allgemeinen ein Forschungsgebiet, das sich mit Klassifikation, Zusammenfassung und Auswahl von Daten zur vereinfachten Darstellung von Objekten in unterschiedlichen Maßstäben beschäftigt. Der allgemeine Vorgang, ist die Identifizierung von Regionen mit sehr hoher Dichte und die anschließende Eliminierung einiger Objekte innerhalb dieser Regionen. Da auch hier ständig nach Automatisierungsmöglichkeiten geforscht wird, möchte ich an dieser Stelle zwei Ansätze vorstellen, bei denen die SOM zur computergestützten Erkennung und Auswahl dieser Objekte verwendet wurde.

Zunächst möchte ich die Arbeit von Jiang und Harrie⁷⁸ vorstellen, welche sich der automatisierten Generalisierung von Straßenkarten widmen. Sie verwenden einen Datensatz von 785 Straßen mit 7 verschiedenen Attributwerten, wie Geschlossenheit, Länge oder Geschwindigkeitsbeschränkungen. Damit erzeugen sie Eingabevektoren, die anschließend mittels SOM visualisiert werden. Durch die Auswahl verschiedener Neuronengruppen werden Straßen ausgewählt. Diese werden daraufhin entfernt oder bleiben erhalten (Abb.32).

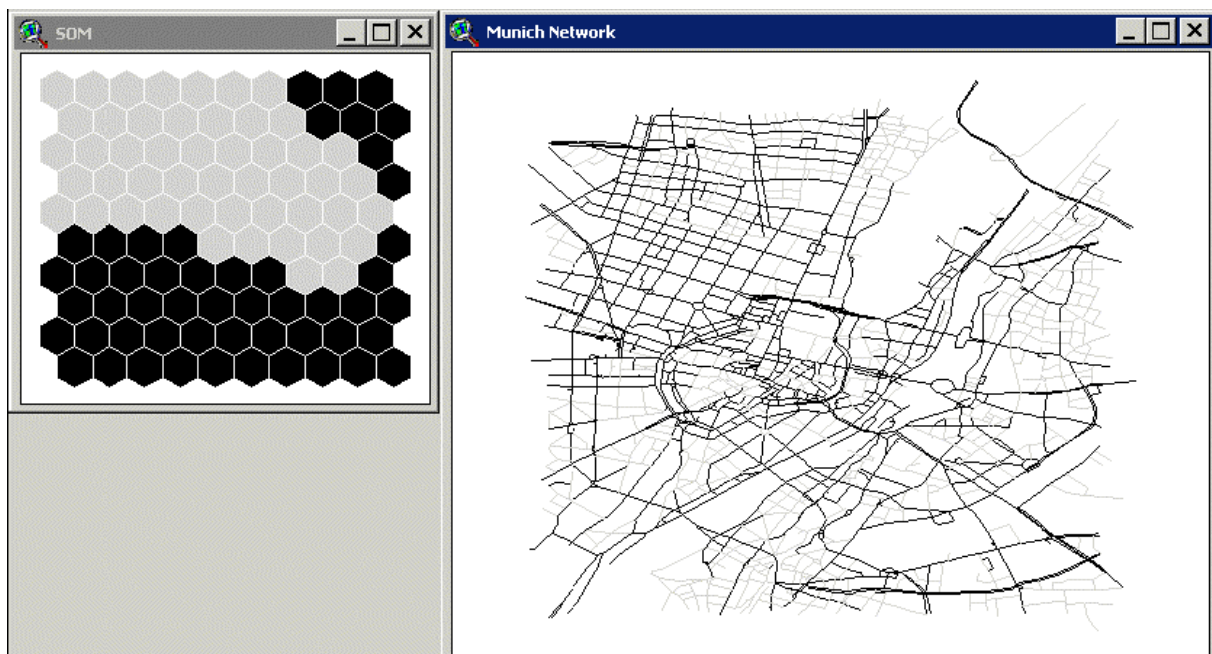


Abb.32 – Generalisierung von Straßenkarten mittels SOM
(Quelle: Jiang (2003) – „Cartographic Selection Using Self-Organizing Maps“ – S.5)

⁷⁸ Jiang (2003) – „Cartographic Selection Using Self-Organizing Maps“

Im Ergebnis sieht man, auf den ersten Blick, eine sehr gute und schnelle Vereinfachung des Straßennetzes. Doch bei näherem Betrachten fällt auf, dass es Straßen gibt, die im Nichts enden oder die mitten in einem freien Bereich liegen. Somit wird deutlich, dass es sich auf keinen Fall um eine vollständige Automatisierungsmethode zur Generalisierung handeln kann. Über empirische Tests und die Auswahl unterschiedlicher der Neuronen lassen sich sicherlich bessere Ergebnisse erzielen. Jedoch würde ich diese Methode als unterstützendes Werkzeug verwenden. Sodass man zum Beispiel eine Vorselektion der zu entfernenden Straßen durch die SOM erzeugt, um dann im Nachgang die Feinheiten manuell herauszuarbeiten und die Generalisierung abzuschließen. Des Weiteren führt dieses Verfahren lediglich zu einer Vereinfachung von Straßen. Die wesentlich komplexere Vereinfachung von Gebäuden wird noch nicht berücksichtigt, weshalb ich ein weiteres Beispiel vorstellen möchte.

Sester hat sich eingehender mit der „automatisierten Generalisierung von Gebäuden für kleinere Maßstäbe“⁷⁹ beschäftigt und in diesem Zusammenhang das Programm „Typify“⁸⁰ entwickelt. Bei ihrem Ansatz verwendet sie die SOM zur „dichteerhaltenden Filterung von Objekten.“⁸¹ Auf Basis des Straßennetzes werden dazu Maschen angelegt, innerhalb derer die Dichte der Objekte reduziert werden kann. Anschließend werden die zu reduzierenden Objekte ausgewählt und durch eine neue, an den Maßstab angepasste, Gebäudedarstellung ersetzt. Auf eine Darstellung der SOM wird hier vollständig verzichtet. Die Technik wird lediglich im Hintergrund angewandt, um die Objekte auszuwählen und deren neue Lage zu bestimmen (Abb.33).

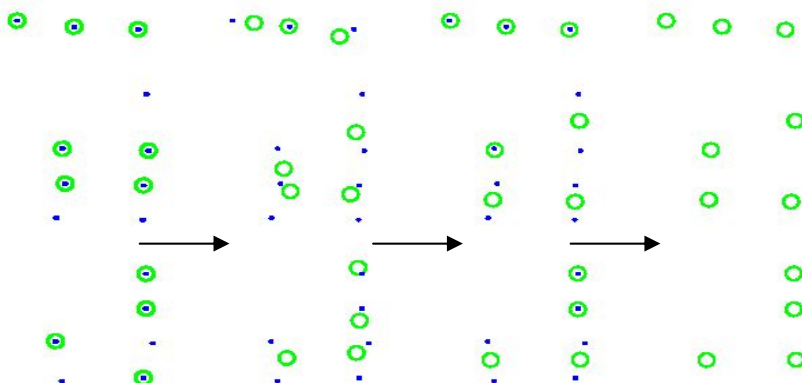


Abb.33 – Anpassung der neuen Objekte
(Quelle: Sester, Monika (2005a) – „TYPIFY“)

⁷⁹ Sester (2005b) – „TYPIFY - Ein Programm zur automatischen Generalisierung von Gebäuden für kleinere Maßstäbe“

⁸⁰ Sester, Monika (2005a) – „TYPIFY“

⁸¹ Sester (2005b) – „TYPIFY - Ein Programm zur automatischen Generalisierung von Gebäuden für kleinere Maßstäbe“ (S.3)

Zwar müssen im Nachgang noch einige Fehler und Überschneidungen mit Straßen korrigiert werden. Doch sieht das Endergebnis der automatisierten Generalisierung schon sehr beeindruckend aus (Abb.34).

Diese beiden Ansätze zeigen, dass SOM durchaus ein geeignetes Mittel ist, um eine automatisierte Generalisierung zumindest zu unterstützen. Ich schätze dies deshalb als eine geeignete Anwendung für die Methode der SOM ein.

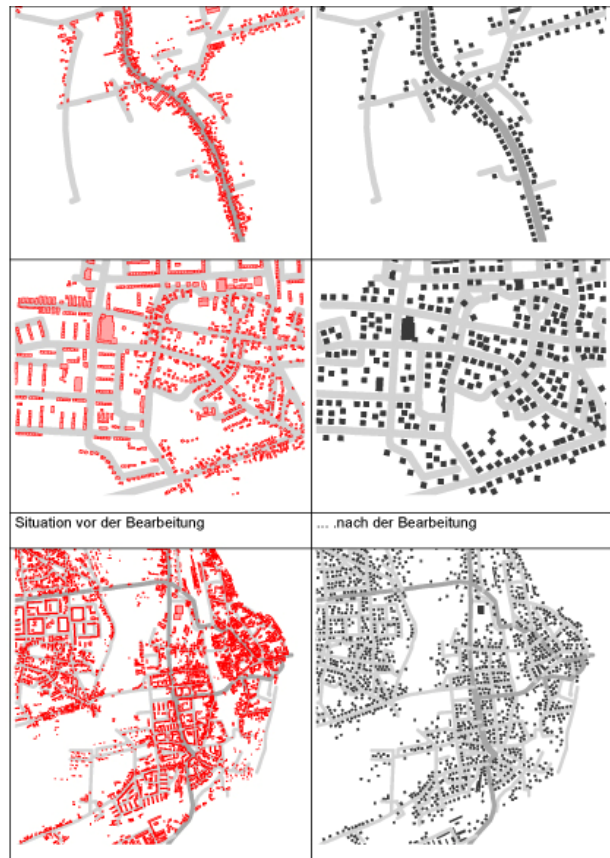


Abb.34 - Generalisierungsergebnis von „Typify“
(Quelle: Sester (2005b) – „TYPIFY - Ein Programm zur automatischen Generalisierung von Gebäuden für kleinere Maßstäbe“ – S.9)

5.6 Problem des Handlungsreisenden (Travelling Salesman Problem)

Abschließen möchte ich dieses Kapitel mit der Anwendung des SOM-Algorithmus für das Problem des Handlungsreisenden, auch Travelling Salesman Problem (TSP) genannt. Es handelt sich um ein bekanntes Optimierungsproblem, bei dem ein Handlungsreisender eine Reihe von Kunden in jeweils verschiedenen Städten, mit der möglichst kürzesten Route erreichen möchte. Für eine geringe Anzahl von Städten erscheint die Lösung noch recht einfach, doch der Berechnungsaufwand steigt schnell ins unermessliche. Denn die Anzahl der notwendigen Berechnungen steigt exponentiell mit der Anzahl der Städte. Für n Städte gibt es demnach $n!$ verschiedene mögliche Routen.

Da das Verhalten des exponentiellen Anstiegs der Möglichkeiten eine Vielzahl von Problemen betrifft, haben sich bereits sehr viele Forschungsgruppen mit diesem Problem beschäftigt.

Dadurch verwundert es nicht, dass der SOM-Algorithmus bereits erfolgreich darauf angewendet wurde (Bai⁸², Brocki⁸³, Budinich⁸⁴, Jin⁸⁵). Auf die technischen Umsetzungen werde ich, aufgrund der Komplexität, jedoch verzichten. Allerdings möchte ein Projekt der HTW-Dresden erwähnen, in dessen Zusammenhang ein Java-Applet entwickelt wurde. Dort wird die Streckenoptimierung sehr gut veranschaulicht (Abb.35).

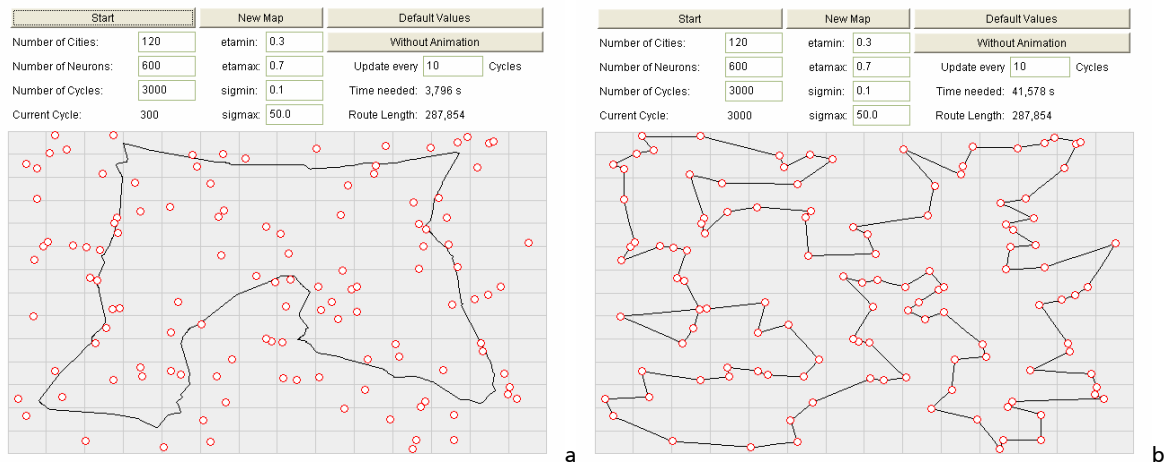


Abb.35 – Berechnung der optimalen Route für TSP mittels SOM a.) nach 300 Trainingsschritten b.) nach 3000 Trainingsschritten
(Quelle: Börner – „Das Problem des Handlungsreisenden“)

⁸² Bai (2005) – „An new self-organizing maps strategy for solving the traveling salesman problem“

⁸³ Brocki (2007) – „Kohonen Self-Organizing Map for the Traveling Salesperson Problem“

⁸⁴ Budinich (1996) – „A Self-Organising Neural Network for the Travelling Salesman Problem that is Competitive with Simulated Annealing“

⁸⁵ Jin (2003) – „An Efficient Self-Organizing Map Designed by Genetic Algorithms for the Traveling Salesman Problem“

6. SOM als Kartenmetapher zur Visualisierung nicht-geographischer Daten

In den vorangegangenen Ausführungen ist bereits deutlich geworden, dass nicht einzig der Begriff der „Map“ direkt eine Anwendung der SOM-Methode in der Kartographie impliziert. Vielmehr ist es die Möglichkeit der Anwendung der kartographischen Zeichentheorie auf die Visualisierung von hochdimensionalen Daten. Demzufolge kann die Kognition dieser Datenmengen erheblich gesteigert werden. Durch Anwendung einer „Kartenmetapher“⁸⁶ bei der Visualisierung von Daten, wird endgültig deutlich, dass durch die Betrachtung von Datenzusammenhängen in einer landschaftsähnlichen Form das kartographische Verständnis vom Raum hilfreich sein kann. Einige Wissenschaftler haben bereits begonnen das kartographische Wissen mit anderen modernen Techniken der Informationsvisualisierung zu kombinieren. Old beschreibt diesen Trend als „Information Cartography.“⁸⁷ Dies bezeichnet die Entwicklung und Nutzung von kartenähnlichen Darstellungen zur Visualisierung von nicht-geographischen Informationen.

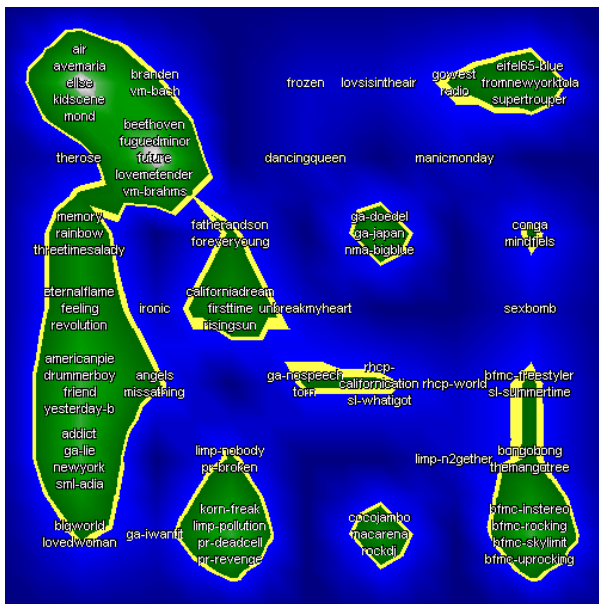
Ein sehr eindrucksvolles und anschauliches Beispiel ist das Projekt „Islands of Music.“ Dabei werden Musikdatenbanken mittels SOM auf Ähnlichkeiten und Beziehungen untersucht, Cluster generiert und diese in Form von „Inseln“ innerhalb eines „Meeres“ veranschaulicht. So ergibt sich das Bild einer landschaftsähnlichen Darstellung innerhalb derer sich musikalisch ähnliche Künstler gemeinsam auf einer „Insel“ befinden (Abb.36). Man muss in diesem Zusammenhang von einer landschaftsähnlichen Darstellung sprechen, da es sich um eine dreidimensionale Ansicht aus der Vogelperspektive handelt. Im Gegensatz zu kartenähnlichen Darstellungen (Abb.37 und 38).

⁸⁶ Old (2002) – „Information Cartography: Using GIS for Visualizing Non-Spatial Data“ (S.2)

⁸⁷ Ebd.



Abb.36 - landschaftsähnliche Darstellung der „Islands of Music“
 (Quelle: Knees (2006) – „An Innovative ThreeDimensional User Interface for Exploring Music Collections Enriched with MetalInformation from the Web“ – S.3)



air avemaria elise kidszene mond	branden vm-bach		frozen	lovsisintheair	goveest radio	eifel66-blue fromnewyorktola supertrouper
therose beethoven fuguedminor future lovemetender vm-brahms			dancingqueen		manicomonday	
memory rainbow threetimesalady			fatherandson foreveryoung		ga-doedel ga-japan nma-bigblue	conga mindfiels
eternallflame feeling revolution	ironic	californiadream firsttime risingsun	unbreakmyheart			sexbomb
americanple drummerboy friend yesterday-b	angels missathing		ga-nospeech tom	rhop- californication rhop-world sl-whatgot	bfmc-freestyler sl-summertime	
addict ga-lie newyork sml-adia	limp-nobody pr-broken				limp-n2gether	bongobong themangotree
blaworld lovedwoman	ga-ivawantit	korn-freak limp-pollution pr-deadocell pr-revenge		cocojambo macarena rockdj		bfmc-instereo bfmc-rocking bfmc-skylimit bfmc-uprocking

Abb.37 – a.) kartenähnliche Darstellung der „Islands of Music“ b.) Basis 7x7 SOM
 (Quelle: Pampalk (2001) – „Islands of Music - Analysis, Organization, and Visualization of Music Archives“ – Bilder von Homepage⁸⁸)

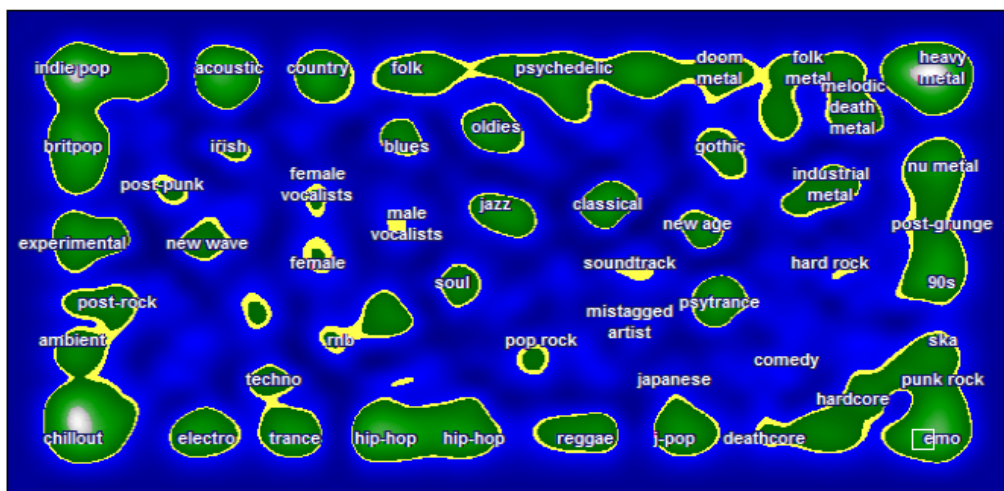


Abb.38 - kartenähnliche Darstellung der „Islands of Music“ nach Stilrichtungen
 (Quelle: <http://playground.last.fm/iom>)

⁸⁸ <http://www.ofai.at/~elias.pampalk/music/index.html>

Es wird deutlich, wie einfach es ist die Musikdatenbank visuell zu erschließen. Synchron kann man erkennen ob die Kompositionen oder Stilrichtungen sich ähneln oder stark differieren.

Gerade dieses Navigieren im Informationsraum stellt den Vorteil der kartographischen Informationsvisualisierung mittels räumlicher Metaphern⁸⁹ dar. Fields beschreibt drei grundlegende Informationstopologien.⁹⁰ Über die *Relation* einzelner Punkte kann man auf deren gegenseitige Beziehung schließen. Vergleichbar mit Toblers „1st law of geography“⁹¹, stehen alle Punkt miteinander in irgendeiner Form in Beziehung. Denn Näheres ist enger verbunden als weit Entferntes. In Abbildung 38 kann man gut erkennen, dass die Musikrichtungen „Chillout“ und „Heavy Metal“ stilistisch sehr wenig miteinander verbindet. Ebenso gibt die Gruppierung von Subjekten eine Verbindung dieser innerhalb einer bestimmten *Thematik* wieder. Zusätzlich lassen sich mithilfe des allgemeinen Überblicks der *Struktur* konzeptionelle Korrespondenzen innerhalb der Daten verdeutlichen.

Der vorangegangenen Grundidee schließt sich auch, das wohl populärste Projekt, die „Map of Mozart“⁹², an. Diesem wurde bereits in nichtwissenschaftlichen Magazinen wie „GEO“ und „Spiegel-Online“ Aufmerksamkeit geschenkt.

Wie auch bei „Islands of Music“⁹³ wird eine Musikdatenbank mittels landschaftsähnlicher Visualisierungen veranschaulicht. Im Unterschied dazu wird dies jedoch nicht in einer rechteckigen Grundform abgebildet, sondern auf die Silhouette von Mozarts Kopf projiziert (Abb.39). Damit lehnen sie sich an das Konzept der Mnemonic SOM⁹⁴ an, in welchem durch Projektion der Daten in eine simple und bekannte Form, die Erfassung der Daten nochmals wesentlich erleichtert werden soll.

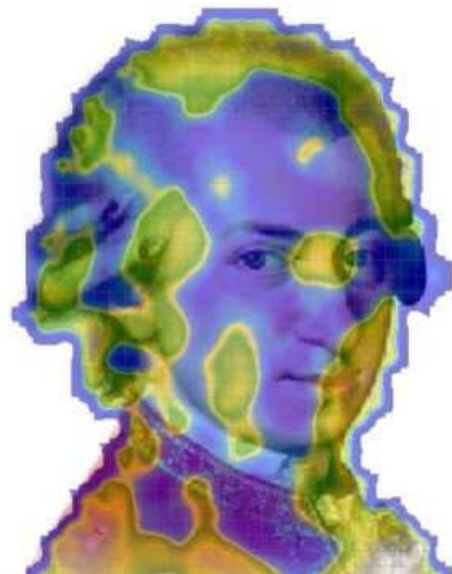


Abb.39 – Visualisierung des Gesamtwerkes von Mozart
(Quelle: Mayer (2006) – „The Map of Mozart“ – S.2)

⁸⁹ Old (2002) – „Information Cartography: Using GIS for Visualizing Non-Spatial Data“

⁹⁰ Fields (2004) – „The Map is the Territory“ (S.4)

⁹¹ Tobler (1970) – „A computer movie simulating urban growth in the Detroit region“ (S.3 der Onlineversion)

⁹² Mayer (2006) – „The Map of Mozart“

⁹³ Pampalk (2001) – „Islands of Music - Analysis, Organization, and Visualization of Music Archives“

⁹⁴ Mayer (2005) – „Mnemonic SOMs: Recognizable shapes for self-organizing maps“

Indem man sich markante Bereiche des Gesichts als Anhaltspunkte nimmt, ist es somit möglich die einzelnen Cluster innerhalb der Darstellung sehr leicht zuzuordnen. Meiner Meinung nach, eine sehr eindrucksvolle Umsetzung.

In seiner Veröffentlichung „The world of geography: Visualizing a knowledge domain with cartographic means“, hat Skupin⁹⁵ den Ansatz der kartenähnlichen Visualisierung genutzt, um einen Datensatz von 2.200 Kurzbeschreibungen auf deren Struktur hin zu untersuchen. Diese wurden für ein Treffen der „Association of American Geographers“ eingereicht. Das Ziel war die „Übermittlung der semantischen Aspekte der geographischen Arbeitsgebiete in Bezug auf maßstabsabhängige Ansichten der globalen, regionalen beziehungsweise lokalen Strukturen“⁹⁶. Zum Beispiel physikalische, städtische oder Standortgeographie. Ein Großteil der geometrischen und topologischen Transformationen wird direkt auf die rohe Konfiguration der, durch SOM-Training entstandenen, Daten angewendet. Dazu hat er die Zusammenfassungen in verschiedene Komponenten, wie Titel oder Autoreninformationen, zerlegt. Auf Basis dieser Stichwörter wurde anschließend die SOM trainiert. Die Symbolisierung wird ausschließlich über Standard-GIS-Software realisiert und dabei klassische kartographische Prinzipien angewendet. Eine visuelle hierarchische Ordnung realisiert er, durch abgestufte qualitative Flächenfärbung sowie Variation der Linienstärke und Schriftgröße (Abb.40).

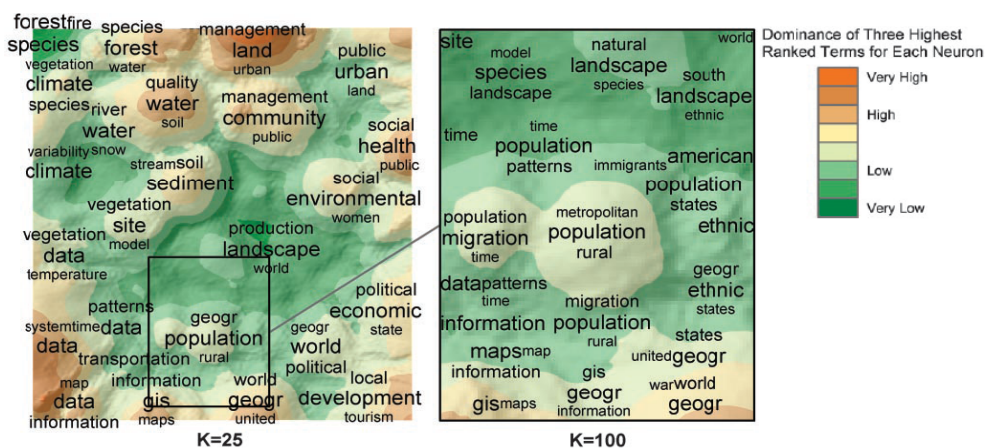


Abb.40 – kartenähnliche Darstellung als „term dominance landscape“
(Quelle: nach Skupin (2004) – „The world of geography: Visualizing a knowledge domain with cartographic means“ - S.3&5)

⁹⁵ Skupin (2004) – „The world of geography: Visualizing a knowledge domain with cartographic means“

⁹⁶ Ebd. (S.1 der pdf-Version)

Durch die damit erzeugte Hierarchie stellen lokale Minima (grüne Bereiche) das Fehlen eines eindeutigen Fokus auf ein spezielles Thema dar. Gleichzeitig deuten Maxima (braune Bereiche) Bereiche an, in welchen ein bestimmter Inhalt sehr stimmig zusammengefasst wird. Durch die unterschiedlichen Schriftgrößen wird parallel angezeigt, welches Thema den entsprechenden Bereich am stärksten dominiert. Im Unterschied zu Beispielen aus den vorangegangenen Kapiteln drückt diese „Landschaftsdarstellung“ nicht die Dichte der vorhandenen Elemente in einem Bereich aus. Vielmehr werden „Berge“ durch dominante Themenkombinationen und „Täler“ durch eine starke Streuung der Themen gebildet. Skupin bezeichnet dies als „term dominance landscape.“⁹⁷

Auf eine Liniendarstellung wurde in dieser Abbildung verzichtet. Allerdings wäre es möglich die Clustergrenzen so noch besser zu verdeutlichen. Es würde jedoch einen Nachteil im Sinne des landschaftsähnlichen Charakters der Visualisierung darstellen, da diese dann zerklüftet wirkt. Lediglich wenn man die beiden Einzeldarstellungen aus Abbildung 40 zusammenfassen möchte, ist die Verwendung von Linien notwendig. Wie man in Abbildung 41 sieht, können somit die unterschiedlichen „Maßstabsbereiche“ (10 – 800 Klassen) in einer Visualisierung dargestellt werden. Die Schriftfarbe wurde ebenso angepasst und die Flächenfärbung, zur besseren Übersicht, entfernt.

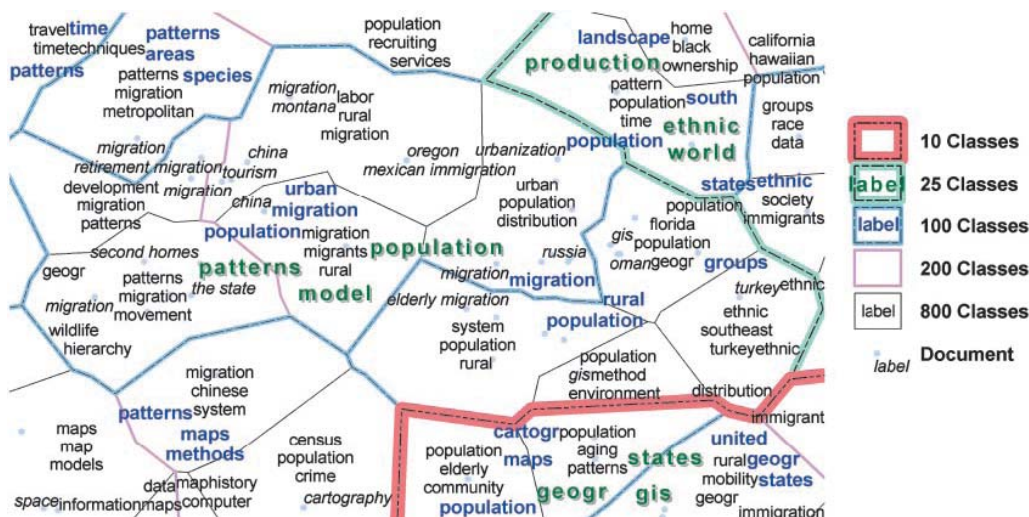


Abb.41 – Darstellung unterschiedlicher Maßstabsbereiche
(Quelle: Skupin (2004) – „The world of geography: Visualizing a knowledge domain with cartographic means“ - S.2)

⁹⁷ Skupin (2004) – „The world of geography: Visualizing a knowledge domain with cartographic means“ (S.3)

Betrachtet man den Ausdruck "migration" in der rechten Abbildung. So ist, laut Skupin, dessen Position, zwischen „population“ und den Berechnungsmethoden der unteren linken Ecke, signifikant. Denn eine Vielzahl von Berechnungsmethoden wird in Migrationsstudien verwendet. Im Gegensatz dazu liegt „ethnic“ auf der rechten Seite von „population“, da hier eher ein qualitativ beschreibendes Paradigma herrscht, wie in den umliegenden Themen dieses Bereichs.⁹⁸

Damit entspricht diese Visualisierung exakt dem, was man sich von der Methode der SOM erhofft. Es kann auf einfache Art und Weise eine sehr große Datenmenge mit einem Blick überschaut werden. Muster und Zusammenhänge, die sonst unbemerkt geblieben wären, sind erkennbar.

Während die vorangegangenen Beschreibungen die Metapher einer Karte zur Informationsvisualisierung eher als landschaftsähnliche Darstellung verstehen, möchte ich abschließend noch eine vollständig andere Idee vorstellen.

Neumayer u.a.⁹⁹ verfolgen den Ansatz der kartenähnlichen Darstellung im Sinne eines Liniennetzplanes der U-Bahn, wie ihn Harry Beck 1931 erstmalig vorgestellt hat.¹⁰⁰ Eine Visualisierung, die heute jedem vertraut ist. Darin wird das vollständige Liniennetz eines Verkehrsbetriebes in einer kompakten, rechteckigen Form verdichtet. Der Vorteil liegt in der einfachen Veranschaulichung eines eigentlich komplexen Inhalts, während die dargestellten Distanzen mitunter stark von den reellen abweichen können.

Die Metapher eines Liniennetzplanes wurde von den Autoren als Erweiterung zur Darstellung der SOM mittels Komponentenebenen erdacht. Diese habe ich bereits in Kapitel 4.1.4 vorgestellt. Zunächst wird standardmäßig die SOM trainiert. Die einzelnen Komponentenebenen werden anschließend klassifiziert, das bedeutet in separate Cluster aufgeteilt (Abb.42-a). Die gravitativen Zentren, der Cluster, werden ermittelt, aneinander grenzende miteinander verbunden und deren Distanzen ermittelt (Abb.42-b).

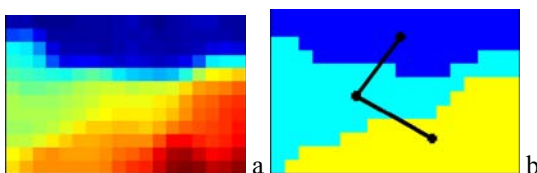


Abb.42 – a.) Komponentenebene einer 12x18 SOM b.) Cluster und verbundene Grenzen
(Quelle: Neumayer (2007) – „The Metro Visualisation of Component Planes for Self-Organising Maps“ – S.2)

⁹⁸ Quelle: Skupin (2004) – „The world of geography: Visualizing a knowledge domain with cartographic means“ (S.4)

⁹⁹ Neumayer (2007) – „The Metro Visualisation of Component Planes for Self-Organising Maps“

¹⁰⁰ Ebd. (S.2)

Aus den Verbindungslinien der einzelnen Komponenten wird nun der Liniennetzplan visualisiert. Der Hauptaspekt ist es, die Darstellung der unsortiert verlaufenden Komponentenslinien (Abb.43-a) an die Metapher eines Liniennetzplanes anzupassen. Als Erstes werden die Linien so angepasst, dass sie nur noch horizontal, vertikal oder diagonal verlaufen (Abb.43-b). Dieses „Snapping“ sorgt für eine grobe Sortierung und drückt besser die angestrebte Metapher aus. Daraufhin werden verschiedene Punkte, die innerhalb von Intersektionsbereichen liegen zusammengefasst und somit der Eindruck einer großen Haltestelle erzeugt (Abb.43-c). Im Hintergrund erkennt man ein zusätzliches Feature. Die Clustergrenze der SOM, welche zuvor aus der U-Matrix ermittelt wurde. Dadurch wird der Eindruck eines Flusses oder Sees erzeugt, wodurch die Visualisierung stärker als „Liniennetzplan“ wirken soll.¹⁰¹

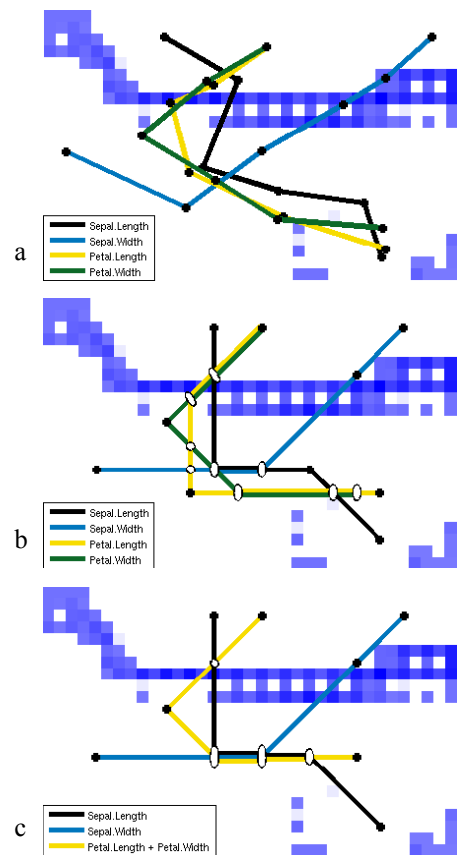


Abb.43 – Darstellungsphasen von Komponentenebenen als Liniennetzplan (Quelle: Neumayer (2007) – „The Metro Visualisation of Component Planes for Self-Organising Maps“ – S.4)

Mit zunehmender Dimensionalität des Datensatzes wird eine übersichtliche Darstellung natürlich zunehmend schwieriger, deshalb empfiehlt es sich, sehr ähnliche Komponentenslinien zu einem Vertreter zusammenzufassen (Abb.43-c).

Der visuelle Eindruck, dass man einen Liniennetzplan irgendeiner Stadt sieht, ist eindeutig gegeben. Gerade bei einer Visualisierung mit hoher Anzahl von Komponenten (Abb.44) ist dieser Effekt wirklich beeindruckend. Inwiefern es möglich ist, sich daraus neues Wissen zu entnehmen, kann ich nicht beurteilen, da die Autoren nicht genügend konkrete Informationen zu den Daten und den einzelnen Komponenten liefern. Man kann jedoch sagen, dass sich dadurch verschiedene Beziehungen und Relationen zwischen einzelnen Komponenten erkennen lassen. In Bereichen, wo mehrere Komponentenslinien verlaufen, tritt eine Häufung von bestimmten Eigenschaften auf. Würde man zum Beispiel einen Datensatz verschiedener Gemeinden betrachten, so könnte dementsprechend eine hohe Dichte von Linien, auf ein Ballungszentrum hinweisen.

¹⁰¹ Neumayer (2007) – „The Metro Visualisation of Component Planes for Self-Organising Maps“ (S.4)

Diesbezüglich würde ich an dieser Stelle zusätzlich eine Beschriftung der Einheiten anfügen. Ähnlich wie es auch im vorangegangenen Beispiel durchgeführt wurde. Dies würde die Analyse erleichtern und zu der Metapher des Liniennetzplanes passen.

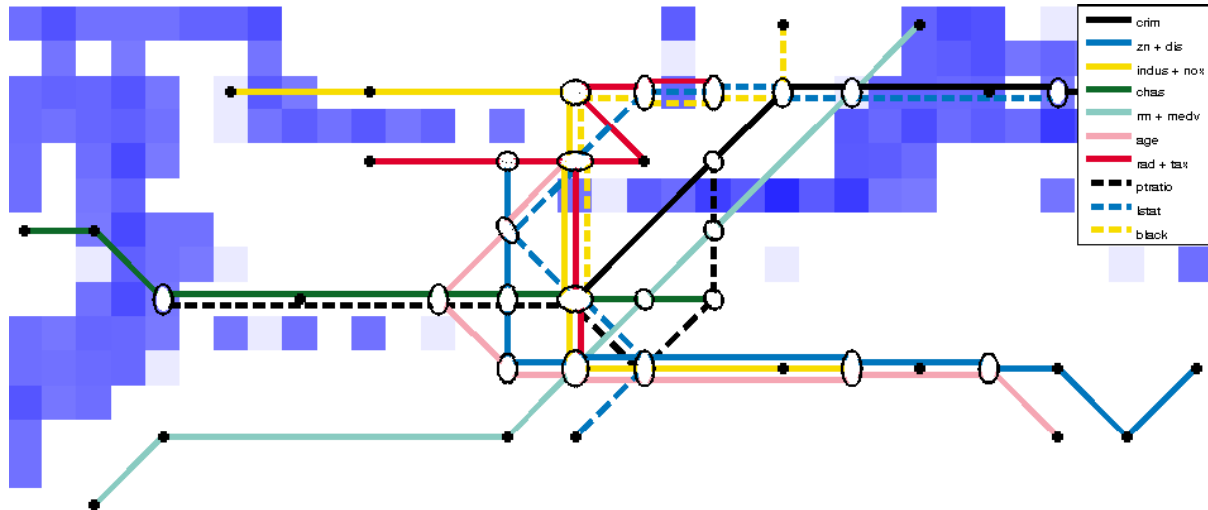


Abb.44 - Darstellung von Komponentenebenen im Sinne eines Liniennetzplanes
(Quelle: Neumayer (2007) – „The Metro Visualisation of Component Planes for Self-Organising Maps“ – S.5)

Man kann also sehen, dass selbst bei der Visualisierung von nicht-geographischen Daten die Methode der SOM hilfreich und interessant für die Kartographie ist. Die Orientierung innerhalb des Datenraums fällt dem Nutzer wesentlich leichter, wenn dieser in einer vertrauten, kartenähnlichen Form dargestellt wird.

Durch eine äußere Begrenzung in Form einer bekannten Silhouette wird das kognitive Verständnis immer weiter verbessert. Man muss jedoch sagen, dass man auf eine zusätzliche Erklärung dessen, was betrachtet wird, nie verzichten kann. Gerade die Visualisierung über die Metapher des Liniennetzplanes ist zwar sehr eindrucksvoll jedoch bei Weitem nicht autoplausibel.

7. Zusammenfassung

Ich habe in dieser Arbeit versucht, einen Überblick über die vielfältigen Einsatzmöglichkeiten der Self-Organizing Map zu geben. Ich habe aufgezeigt, dass dieser, in der Grundidee sehr simple, Algorithmus für eine Vielzahl von Teilbereichen der Kartographie hilfreich sein kann und bereits zahlreiche Wissenschaftler zu neuen Ideen inspiriert hat.

In diesem Zusammenhang habe ich zunächst den allgemeinen Aufbau sowie die Funktionsweise der SOM verdeutlicht. Daraufhin habe ich festgestellt, dass Self-Organizing Maps gerade als Werkzeug des geowissenschaftlichen Data Mining und als Visualisierungsmöglichkeit im Sinne der explorativen Datenanalyse für die Kartographie geeignet ist. Dies habe ich im Folgenden durch Beispiele aufgezeigt und detaillierter beschrieben. Dabei ist zum einen deutlich geworden, dass die Methode sich sehr gut zur Geovisualisierung und zur Einbindung in Geoinformationssysteme eignet. Zum anderen bin ich aber auch auf die Verarbeitung von geographischen Daten eingegangen, wobei eine Eignung des Algorithmus zur Datenverarbeitung und -vorbereitung deutlich wurde. Abschließend bin ich auf die Visualisierung von nicht-geographischen Daten eingegangen. In diesem Zusammenhang wird die Visualisierung dieser Daten als Metapher einer Karte betrachtet. Dadurch wird eine visuelle Erschließung von nicht-geographischen Daten möglich. Durch Anwendung von kartographischen Regeln kann zusätzlich das kognitive Verständnis solcher Visualisierungen deutlich verbessert werden.

Man kann also erkennen, dass bestimmte Anwendungen der ursprünglichen Self-Organizing Map bereits in der Kartographie genutzt werden. Durch verschiedene Erweiterungen, wie der Geo-SOM, kann und muss der Algorithmus noch weiter für eine Anwendung in der Kartographie optimiert werden.

X. Quellenverzeichnis

X.1 Literaturnachweise

- Abdel Latif, Bassam; Mercier, Grégoire (2006) – “Self-Organizing Map for erroneous data processing in time series analysis”*
- IEEE International Geoscience and Remote Sensing Symposium Vol.1, Denver, Colorado, USA, 2006
 - Und Online: <http://public.enst-bretagne.fr/~mercierg/articles/2006/alatif.igarss06.pdf> (03.09.2009)
- Abdel Latif, Bassam; Mercier, Grégoire; Solaiman, Basel (2007) – “Self-Organizing Map for surface characterization in time series”*
- International Geoscience and Remote Sensing Symposium (IGARSS), Barcelona, Spain, 2007
 - Und Online: <http://public.enst-bretagne.fr/~mercierg/articles/2007/som.igarss07.pdf> (03.09.2009)
- Agarwal, Pragma; Skupin, André (2008) - “Self-Organising Maps: Applications in Geographic Information Science”
- John Wiley & Sons Ltd., The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England
- Agarwal, Pragma; Skupin, André (2008) – “Introduction: What is a Self-Organizing Map?”
- Agarwal; Skupin (2008) - “Self-Organising Maps: Applications in Geographic Information Science” – Kapitel 1 (S. 1-20)
- Bação, Fernando Luças; Lobo, Victor; Painho, Marco (2004) – “The self-organizing map, the Geo-SOM, and relevant variants for geosciences”*
- Computers and Geosciences 31 (2005) 155-163
 - Und Online: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.59.4698> (25.01.2010)
- Bação, Fernando Luças; Lobo, Victor; Painho, Marco (2005a) – “On the particular characteristics of spatial data and its similarities to secondary data used in data mining”*#
- http://edugi.uni-muenster.de/eduGI/downloads/03/teaching_materials_NU_Lisbon_update/bacaolobopainho2005.pdf (14.09.2009)
- Bação, Fernando Luças; Lobo, Victor; Painho, Marco (2005b) – “Geo-SOM and its Integration with Geographic Information Systems”*#
- <http://www.isegi.unl.pt/labnt/geosom/GeoSOM.Publ.htm> (12.08.2009)
- Bação, Fernando Luças (2007) – „Geospatial Data Mining“*#
- erschienen für: Bação; Lobo (2007) – “Geospatial Data Mining (E-Learning)”
 - http://edugi.uni-muenster.de/eduGI/downloads/03/teaching_materials_NU_Lisbon_update/GeospatialDataMining.pdf (04.09.2009)
- Bai, Yanping; Zhang, Wendong; Jin, Zhen (2005) – „An new self-organizing maps strategy for solving the traveling salesman problem“*
- Chaos, Solitons & Fractals Volume 28, 2006, S.1082-1089
 - Und Online: <http://7y.nuc.edu.cn/jinzhen/paper/An%20new%20self-organizing%20maps%20strategy%20for%20solving%20the%20traveling%20salesman%20problem.pdf> (24.02.2010)
- Bertin, Jacques (1974) – “Graphische Semiologie – Diagramme, Netze, Karten“
- Verlag Walter de Gruyter, Berlin, New York
- Bollmann, Jürgen; Koch, Wolf Günther (2002) – „Lexikon der Kartographie und Geomatik“
- Spektrum Akademischer Verlag, Heidelberg, Berlin

- Börner, Sven – „Das Problem des Handlungsreisenden“^{**#}
- HTW Dresden, Fakultät Informatik, Beleg „Neuronale Netze“
 - <http://www.htw-dresden.de/~iwe/Belege/Boerner/>
- Brejzl, Marek; Sonka, Milan (1999) – „Medical image segmentation: Automated design of border detection criteria from examples“[~]
- Special Issue of Journal of Electronic Imaging Vol.8, 1999, S.54–64
- Brocki, Lucas (2007) – „Kohonen Self-Organizing Map for the Traveling Salesperson Problem“^{*}
- Recent Advances in Mechatronics Part1, Springer Berlin, Heidelberg, 2007, S.116-119
 - Und Online: <http://www2.in.tu-clausthal.de/~hammer/lectures/heursem/tsp.pdf> (03.09.2009)
- Budinich, Marco (1996) – „A Self-Organising Neural Network for the Travelling Salesman Problem that is Competitive with Simulated Annealing“^{*}
- Neural Computation Volume 8, February 15, 1996
 - Und Online: http://wwwusers.ts.infn.it/~mbh/TSP_final_paper.pdf (03.09.2009)
- Canny, John (1986) – “A Computational Approach to Edge Detection”^{*}
- IEEE Transactions on Pattern Analysis and Machine Intelligence archive, Volume 8, 1986, S.679 - 698
 - Und Online: <http://www.icst.pku.edu.cn/course/ImageProcessing/2009/resource/canny1986.pdf> (24.02.2010)
- Carrão, Hugo; Capão, Lois; Bação, Fernando Luças; Caetano, Mário (2006) – “Meris Based Land Cover Classification with Self-Organizing Maps: Preliminary Results”^{*}
- Proceedings of the 2nd Workshop of the EARSeL SIG on Land Use and Land Cover, Center for Remote Sensing of Land Surfaces, Bonn, 28-30 September 2006
 - Und Online: http://www.zfl.uni-bonn.de/earsel/papers/16-24_carrao.pdf (22.08.2009)
- da Silva, Marcos A. S.; Monteiro, Antônio M. V.; Medeiros, José S. (2004) – “Visualization of Geospatial Data by Component Planes and U-Matrix”^{**#}
- Beitrag bei: GeoInfo 2004 – VI Brazilian Symposium on Geoinformatics – „Session 2- Spatial Analysis and Statistics“ (<http://www.geoinfo.info/geoinfo2004/program.html>) (10.02.2010)
 - Dokument: <http://www.geoinfo.info/geoinfo2004/papers/6419.pdf> (10.02.2010)
- del Frate, Fabio; del Greco, Marco; Schiavon, Giovanni; Solimini, Domenico; Putignano, Cosimo (2007) – “Self-organizing maps of polarimetric SAR imagery”^{*}
- POLinSAR 2007, 3rd International Workshop on Science and Applications of SAR Polarimetry and Polarimetric Interferometry, Frascati, Italy, 22-26 January 2007
 - Und Online (Abstract): http://earth.esa.int/workshops/polinsar2007/abstracts/148_delfrate.pdf (24.02.2010)
- Evangelou, Iordanis E.; Hadjimitsis, Diofantos G.; Lazakidou, Athina A.; Clayton, Chris (2001) – „Data Mining and Knowledge Discovery in Complex Image Data using Artificial Neural Networks“
- Workshop on Complex Reasoning an Geographical Data, Cyprus, 2001
 - Und Online: http://freenet-homepage.de/lazakid/paper_proceedings11.pdf (15.10.2009)
- Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (1996) – “From Data Mining to Knowledge Discovery in Databases”ⁱ
- AI Magazine (S.37-54) - American Association for Artificial Intelligence
 - Und Online: http://edugi.uni-muenster.de/eduGI/downloads/03/teaching_materials_NU_Lisbon_update/fayyadetal1996.pdf (09.09.2009)
- Fields, Kenneth (2004) – „The Map is the Territory“^{*}
- Conference Consciousness Reframed 2004: QI and COMPLEXITY, Beijing, China, 24-27 November, 2004
 - Und Online: <http://cenc.com.edu.cn/docs/MapTerritory.pdf> (23.09.2009)

- Fincke, Tonio; Lobo, Victor; Bação, Fernando Luças (2008) – “Visualizing self-organizing maps with GIS”*
- GITage 2008, Münster, Deutschland, 16-17 Juni 2008
 - Und Online: <http://www.gitage.de/archive/2008/downloads/acceptedPapers/Papers/Fincke.Lobo.Bacao.pdf> (03.09.2009)
- Fukuda-Parr, Sakiko (2004) – „Human Development Report 2004“*
- United Nations Development Programme, New York, USA, Hoechstetter Printing Co.
 - Und Online: http://hdr.undp.org/en/media/hdr04_complete.pdf (22.02.2010)
 - PresseKit: http://www.hdr.undp.org/en/media/HDR_2004_Presskit_EN.pdf (22.02.2010)
- Giacco, Ferdinando; Scarpetta, Silvia; Pugliese, Luca; Marinaro, Maria; Thiel, Christian (2009) – “Application of Self Organizing Maps to multi-resolution and multi-spectral remote sensed images”*
- Frontiers in Artificial Intelligence and Applications Vol. 193, Proceeding of the 2009 conference on New Directions in Neural Networks: 18th Italian Workshop on Neural Networks: WIRN 2008, 245-253
 - Und Online: <http://www.fisica.unisa.it/NeuralGroup/wirn2008-giaccoscarpetta.pdf> (03.09.2009)
- Henriques, Roberto; Bação, Fernando Luças; Lobo, Victor (2009) – “Carto-Som – Cartogram creation using self-organizing maps“*
- International Journal of Geographical Information Science, Volume 23, Number 4, April 2009, pp. 483-511(29), Taylor and Francis Ltd
 - Und Online: http://www.isegi.unl.pt/ensino/docentes/fbacao/henriques_coloquium_paper.pdf (12.08.2009)
- Jiang, Bin; Harrie, Lars (2003) – „Cartographic Selection Using Self-Organizing Maps“*
- Fifth workshop on progress in automated map generalisation, ICA, commission on map generalisation, Paris, France, 2003
 - Und Online: http://www.nateko.lu.se/Personal/Lars.Harrie/ICA_mgws_03.pdf (11.08.2009)
- Jiao, Limin; Liu, Yaolin (2008) – „Knowledge Discovery in Spatial Clustering Based on Self-Organizing Feature Map and a Composite Distance Measure“*
- The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences Vol. 37, Part B2, Beijing, China, 2008
 - Und Online: http://www.isprs.org/congresses/beijing2008/proceedings/2_pdf/2_WG-II-2/09.pdf (09.10.2009)
- Jin, Hui-Dong; Leung, Kwong-Sak; Wong, Man-Leung; Xu, Zong-Ben (2003) – „An Efficient Self-Organizing Map Designed by Genetic Algorithms for the Traveling Salesman Problem“*
- Systems, Man, and Cybernetics, Part B: Cybernetics Vol.33, IEEE Transactions, 2003
 - Und Online: <http://cptra.ln.edu.hk/~mlwong/journal/smc2003.pdf> (03.09.2009)
- Kanevski, Mikhail; Kaiser, Christian (2007) – „Classification and Visualization of High-Dimensional Socio-Economic Data Using Self-Organizing Maps“*#
- Vortrag bei Spatial Econometrics 2007, Cambridge, 11-14 July 2007
 - pdf-Version: <http://www.clusterville.org/index.php?page=som-classification> (25.01.2010)
- Kangas J. (1992) – “Temporal knowledge in locations of activations in a self-organizing map”
- Artificial Neural Networks Vol. 2, Elsevier, Amsterdam, The Netherlands, 117–120
- Knees, Peter; Schedl, Markus; Pohle, Tim; Widmer, Gerhard (2006) – „An Innovative ThreeDimensional User Interface for Exploring Music Collections Enriched with MetaInformation from the Web“*
- Proceedings of the ACM Multimedia 2006, Santa Barbara, California, USA, October, 2006, S.17-24
 - Und Online: http://www.cp.jku.at/research/papers/Knees_etal_ACM_2006.pdf (21.09.2009)
- Koch, Wolf Günther (2007) – Vorlesungsunterlagen “Theoretische Kartographie”
- TU-Dresden, Fakultät Forst-, Geo- und Hydrowissenschaften, Institut für Kartographie

- Kohonen, Teuvo (1995) – “Self-Organizing Maps”
 - Springer Series in Information Sciences, Vol. 30, Springer, Berlin, Heidelberg
- Kohonen, T.; Makisara, K.; Simula, O.; Kangas, J. (1991) – “The hypermap architecture”
 - Artificial Neural Networks Vol. 1, Elsevier, Amsterdam, The Netherlands, 1357–1360
- Koua, Etien L. (2003) – “Using Self-Organizing Maps for Information Visualisation and Knowledge Discovery in Complex Geospatial Datasets”
 - Proceedings of the 21st International Cartographic Conference (ICC), Durban, South Africa, 10-16 August 2003
 - Und Online: http://www.itc.nl/library/Papers_2003/art_proc/koua.pdf (03.09.2009)
- Koua, Etien L.; Kraak, Menno-Jan (2004) – „Geovisualization to support the exploration of large health and demographic survey data“
 - International Journal of Health Geographics 2004, 3:12
 - Und Online: <http://www.ij-healthgeographics.com/content/3/1/12> (22.02.2010)
- Koua, Etien L.; Kraak, Menno-Jan (2008) – “An Integrated Exploratory Geovisualisation Environment Based on Self-Organizing Map”
 - Agarwal; Skupin (2008) - “Self-Organising Maps: Applications in Geographic Information Science” – Kapitel 3 (S. 45-66)
- Kranz, Sabine; Lenze, Burkhard (2007) – “Visualisierung von Self Organizing Maps mittels Java-Applets und HTML-Tutorial sowie Analyse ihrer Anwendbarkeit im Data- und Text-Mining-Kontext”
 - Diplomarbeit - Fachhochschule Dortmund (Fachbereich Informatik) - <http://www.kn-netze.de/files/Diplomarbeit.pdf> (22.10.2009)
- Kranz, Sabine (2007) – “Self Organizing Maps Tutorial”
 - Fachhochschule Dortmund (Fachbereich Informatik) - http://www.kn-netze.de/files/SOM_Tutorial.pdf
- Marques, Nuno C.; Chen, Ning (2003) - „Border detection on remote sensing satellite data using self-organizing maps“
 - Progress in Artificial Intelligence, 11th Portuguese Conference on Artificial Intelligence, EPIA 2003, Beja, Portugal, December 4-7, 2003, Proceedings 2003
 - Und Online: <http://ssdi.di.fct.unl.pt/~nmm/MyPapers/MN03.pdf> (03.11.2009)
- Mayer, Rudolf; Merkl, Dieter; Rauber, Andreas (2005) – „Mnemonic SOMs: Recognizable shapes for self-organizing maps“
 - Proceedings of the Workshop on Self-Organizing Maps, Paris, France, 2005
 - Und Online: http://www.ifs.tuwien.ac.at/~andi/publications/pdf/may_wsom05.pdf (02.10.2009)
- Mayer, Rudolf; Lidj, Thomas; Rauber, Andreas (2006) – “The Map of Mozart”
 - Proceedings of the 7th International Conference on Music Information Retrieval, Victoria, Canada, October 8-12, 2006
 - Online-Pdf: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.96.7598&rep=rep1&type=pdf> (03.03.2010)
 - Homepage mit Online-Demo: http://www.ifs.tuwien.ac.at/mir/mozart/index_de.html
- Meng, Liqiu (2003a) – “Rahmenbedingungen beim Einsatz von Geovisualisierungsmethoden und –techniken”
 - Kartographische Schriften, Band 7, Visualisierung und Erschließung von Geodaten, Kirschbaum Verlag Bonn, 3-12
 - Und Online: http://www.geovisualisierung.net/geovis_refs/meng_2003_paper.pdf (12.02.2010)
- Meng, Liqiu (2003b) - “Rahmenbedingungen beim Einsatz von Methoden und Techniken der Geovisualisierung”
 - Vortragsfolien GEOVIS Hannover (2003)
 - http://www.geovisualisierung.net/geovis_refs/Meng_2003.pdf (12.02.2010)

- Mielens, Jason (2008) – „Self-Organizing Maps for Land Cover Classification“^{**#}
- University of Wisconsin-Madison, College of Engineering, Department of Electrical & Computer Engineering, Introduction to Artificial Neural Network and Fuzzy Systems, Fall 2008 Semester Class Project
 - http://homepages.cae.wisc.edu/~ece539/fall08/project/mielens_rpt.pdf (03.08.2009)
- Nattkemper, Tim W. (2005) – Vorlesungsunterlagen „Information Visualisation“^{**+}
- Universität Bielefeld, Technische Fakultät, Angewandte Neuroinformatik
 - Und Online: <http://www.techfak.net/ags/ni/lectures/lectures-w06/infvis/media/InfoVis-01.pdf> (24.09.2009)
- Nagao, Taketsugu; Mitsukura, Yasue; Fukumi, Minoru; Akamatsu, Norio (2005) – „Drift Ice Classification Using SAR Image Data by a Self-Organizing Neural Network“^{**}
- IEEJ Transactions on Electronics, Information and Systems Vol. 125, S.800-806, 2005
 - Und Online (Abstract): http://www2.iee.or.jp/ver2/honbu/14-magazine/log/2005/2005_05c_13.pdf (03.09.2009)
- Neagoe, Victor-Emil; Ropot, Armand-Dragos (2006) – „Pattern Recognition in Multispectral Satellite Images Using Concurrent Self-Organizing Modular Neural Networks“^{**#}
- Conference Paper, University of Politehnica, Bucharest, Romania, Faculty of Electronics and Telecommunications
 - <http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA479800&Location=U2&doc=GetTRDoc.pdf> (03.09.2009)
- Neumayer, Robert; Mayer, Rudolf; Pözlbauer, Georg; Rauber, Andreas (2007) – „The Metro Visualisation of Component Planes for Self-Organising Maps“^{**}
- Proceedings of the 20th International Joint Conference on Neural Networks (IJCNN'07), Orlando, Florida, USA
 - Und Online: http://publik.tuwien.ac.at/files/pub-inf_4745.pdf (07.10.2009)
- Old, L. John (2002) – “Information Cartography: Using GIS for Visualizing Non-Spatial Data”^{**}
- Proceedings, ESRI International Users' Conference, San Diego, CA, July 2002
 - Und Online: <http://www.dcs.napier.ac.uk/~cs171/LJOld/papers/InformationCartographyESRI.pdf> (22.09.2009)
- Openshaw, Stan; Wymer, Colin (1995) – „Classifying and regionalizing census data“^{**}
- Census User Handbook Geoinformation International, Cambridge, 239-268
- Openshaw, Stan (1999) – “Geographical data mining: key design issues”^{**}
- GeoComputation CD-Rom (1999) - J.Diaz, R.Tynes, D.Caldwell, J.Ehlen
 - Textversion: http://www.geocomputation.org/1999/051/gc_051.htm (11.02.2010)
 - Pdf-Version: http://edugi.uni-muenster.de/eduGI/downloads/03/teaching_materials_NU_Lisbon_update/Openshaw1999.pdf (04.09.2009)
- Oyana, Tonny J., Boppidi, Deepthi; Yan, Jun; Lwebuga-Mukasa, Jamson S. (2005) – „Exploration Of Geographic Information Systems-Based Medical Databases With Self Organizing Maps (Som): A Case Study Of Adult Asthma“^{**}
- GeoComputation CD-Rom (2005) – Y.Xie, D.G.Brown
 - Und Online: <http://www.geocomputation.org/2005/Oyana.pdf> (22.02.2010)
- Piatetsky-Shapiro, Gregory (1991) – “Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop”^{**}
- AI Magazine 11(5) (S.68-70) - American Association for Artificial Intelligence
- Pöllä, Matti; Honkela, Timo; Kohonen, Teuvo (2009) – „Bibliography of SOM papers”^{**}
- Laboratory of Computer and Information Science – Helsinki University of Technology – <http://www.cis.hut.fi/research/som-bibl/> (05.02.2010)

- Pampalk, Elias (2001) – “Islands of Music - Analysis, Organization, and Visualization of Music Archives”*
- Technische Universität Wien, Institut für Softwaretechnik und interaktive Systeme, Diplomarbeit
 - Und Online: <http://www.ofai.at/~elias.pampalk/music/pampalk.pdf> (02.10.2009)
 - zusätzlich Homepage zur Vorstellung des Themas: <http://www.ofai.at/~elias.pampalk/music/index.html> (24.02.2010)
- Reyes-Aldasoro, Constantin Carlos; Aldeco, Ana Laura (2000) – „Image segmentation and compression using neural networks“**
- Advances in Artificial Perception and Robotics CIMAT, Guanajuato, Mexico, 2000
 - Und Online: <http://www.cs.jhu.edu/~cis/cista/446/papers/SegAndCompression.pdf> (24.02.2010)
- Sester, Monika (2008) – “Self-Organizing Maps for Density-Preserving Reduction of Objects in Cartographic Generalization”
- Agarwal; Skupin (2008) - “Self-Organising Maps: Applications in Geographic Information Science” – Kapitel 6 (S. 107-120)
- Sester, Monika (2005b) – „TYPIFY - Ein Programm zur automatischen Generalisierung von Gebäuden für kleinere Maßstäbe“**
- Universität Hannover, Institut für Kartographie und Geoinformatik, Dokumentation zu „TYPIFY“
 - Und Online : <http://www.ikg.uni-hannover.de/fileadmin/ikg/staff/research/documents/typify.pdf> (01.09.2009)
- Skupin, André (2004) – „The world of geography: Visualizing a knowledge domain with cartographic means“**
- Proceedings of the National Academy of Sciences of the United States of America (PNAS) Vol.101, 2004, S.5274-5278
 - Und Online: <http://www.pnas.org/content/101/suppl.1/5274.full.pdf+html> (23.09.2009)
- Tobler, Waldo (1970) – "A computer movie simulating urban growth in the Detroit region"*
- Economic Geography 46 (S.234-240)
 - Und Online: http://www.geog.ucsb.edu/~tobler/publications/pdf_docs/geog_analysis/ComputerMovie.pdf (05.02.2010)
- Tukey, John Wilder (1977) – “Exploratory Data Analysis”
- Addison-Wesley, Reading
- Ultsch, Alfred (1993) – „Self-Organizing Neural Networks for Visualization and Classification“
- Opitz; Lausen; Klar (1993) - “Information and Classification: Concepts, Methods and Applications” (S.307-313)
- Vesanto, Juha; Himberg, Johan; Alhoniemi, Esa; Parhankangas, Juha (2000) – “SOM Toolbox for Matlab 5” **
- Helsinki University of Technology, Laboratory of Computer and Information Science - <http://www.cis.hut.fi/projects/somtoolbox/package/papers/techrep.pdf> (10.02.2010)
- Ware, Colin (2004) – “Information Visualization: Perception for Design”
- Morgan Kaufmann (Elsevier), San Francisco

X.II Lehrinhalte aus dem Internet

- Prof. Dr. Lippe, Wolfram (2007) – „Interaktive Skripte, Interaktive Einführung in Neuronale Netze, Kohonen-Netze“ - Institut für Informatik, Westfälische Wilhelms-Universität *#
- Hauptseite: <http://cs.uni-muenster.de/Studieren/Scripten/Lippe/> (10.02.2010)
 - Skript: <http://cs.uni-muenster.de/Professoren/Lippe/lehre/skripte/wwwnscript/strfx/Kohonen.html> (10.02.2010)
- Prof. Dr. Bação, Fernando Luças; Prof. Dr. Lobo, Victor (2007) – “Geospatial Data Mining (E-Learning)” - Institute of Statistics and Information Management, Universidade Nova de Lisboa *#
- Hauptseite: <http://edugi.uni-muenster.de/eduGI/downloads/03/eduGI%20elearning.htm> (10.02.2010)
 - pdf-Version liegt der CD bei
- Prof. Dr. Schiewe, Jochen – „geovisualisierung.net“ *#
- HafenCity Universität Hamburg, Department Geomatik
 - <http://www.geovisualisierung.net/index.htm> (12.02.2010)
 - pdf-Version liegt der CD bei
- SASI Group; Newman, Mark (2006) – „www.worldmapper.org“
- SASI Group – University of Sheffield
 - Mark Newman – University of Michigan
 - Beispiel einer Kartenanamorphose zur Darstellung der Gesamtbevölkerung:
http://www.worldmapper.org/posters/worldmapper_map2_ver5.pdf (21.04.2010)*#

X.III Softwarelösungen

- Kohonen, Teuvo; Hynninen, Jussi; Kangas, Jari; Laaksonen, Jorma (1996) – „SOM_PAK: The Self-Organizing Map Program Package“
- Technical Report A30 - Helsinki University of Technology, Laboratory of Computer and Information Science, Neural Networks Research Centre
 - Hauptseite: <http://www.cis.hut.fi/research/som-research/nnrc-programs.shtml> (10.02.2010)
 - Software: http://www.cis.hut.fi/research/som_pak/ (10.02.2010)
- Alhoniemi, Esa; Himberg, Johan; Parhankangas, Juha; Vesanto, Juha (2000-2005) – „SOM Toolbox“
- Helsinki University of Technology, Laboratory of Computer and Information Science
 - Hauptseite: <http://www.cis.hut.fi/projects/somtoolbox/> (10.02.2010)
 - Software: <http://www.cis.hut.fi/projects/somtoolbox/download/> (10.02.2010)
- Bação, Fernando Luças; Lobo, Victor; Painho, Marco (2004) – “GeoSOM - Clustering geo-referenced data”
- Instituto Superior de Estatística e Gestão de Informação, Universidade Nova de Lisboa, Campus de Campolide
 - Hauptseite: <http://www.isegi.unl.pt/labnt/geosom/index.htm> (10.02.2010)
 - Software: http://www.isegi.unl.pt/labnt/geosom/GeoSOM_downloads.htm (10.02.2010)
- Sester, Monika (2005a) – „TYPIFY“
- Universität Hannover, Institut für Kartographie und Geoinformatik
 - Hauptseite: <http://www.ikg.uni-hannover.de/de/dienstleistungen/typify/> (24.02.2010)
 - Software: nicht frei erhältlich

* Dokument befindet sich auf beigelegter CD-Rom

+ bei Abgabe bereits nicht mehr abrufbar

~ Dokument nicht frei erhältlich

reine Internetquelle