

# Visualisierung

- Visualisierung = Optimierung der Schnittstelle Rechner → Gehirn
  - Fokussierung des Gehirns auf Verarbeitung visueller Reize ( $75 \text{ MB s}^{-1}$ )
  - effiziente Extraktion von Information:  
250 Mio Sinneszellen → 2 Mio Ganglien → Erkennen weniger Objekte
- => Visualisierung großer Datensätze als Basis  
entscheidungsunterstützender Systeme

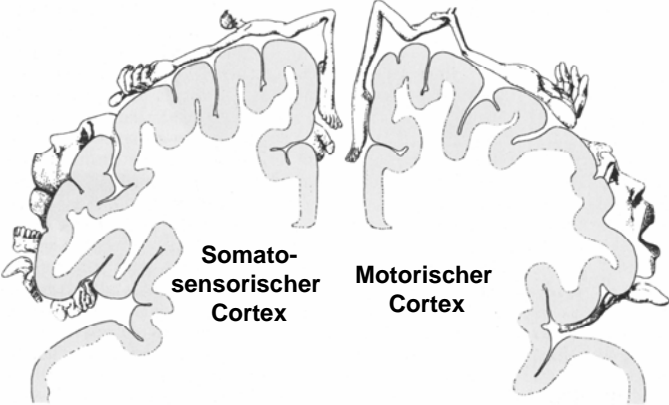
## Selbstorganisierende Karte

(Kohonen 1982)

Visualisierung großer Datensätze durch

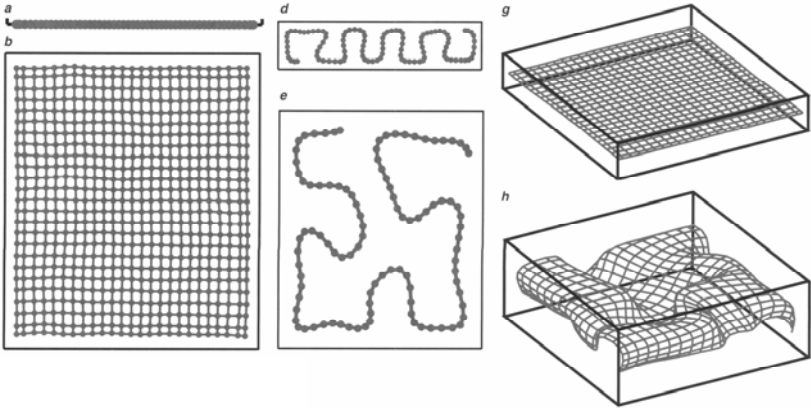
- Reduzierung der Dimensionalität
- Wahrung von Nachbarschaftsbeziehungen

# Repräsentation im Gehirn



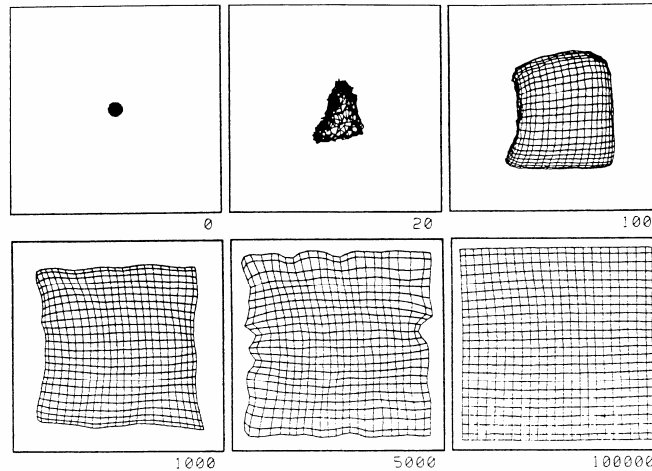
(Bauer et al. 1997)

# Dimensionsreduktion



(Bauer et al. 1997)

# Anpassung der Struktur



(Kohonen 2001)

WEBSOM map - Million documents - Netscape

File Edit View Go Bookmarks Tools Window Help

http://websom.hut.fi/websom/millondemo/html/root.html

WEBSOM map - Million documents

music - music  
 nt - comp.os.ms-windows.nt.setup.hardware  
 pc.cdrom - comp.sys.ibm.pc.hardware.cd-rom  
 pc.chips - comp.sys.ibm.pc.hardware.chips  
 pc.comun - comp.sys.ibm.pc.hardware.comun  
 pc.storage - comp.sys.ibm.pc.hardware.storage  
 pc.video - comp.sys.ibm.pc.hardware.video  
 philosophy - philosophy  
 plant - bionet.biology.plant  
 prolog - comp.lang.prolog  
 sci.lang - sci.lang  
 smalltalk - comp.lang.smalltalk  
 speech - comp.speech  
 sun - comp.sys.sun.hardware

[Full list of newsgroups](#)

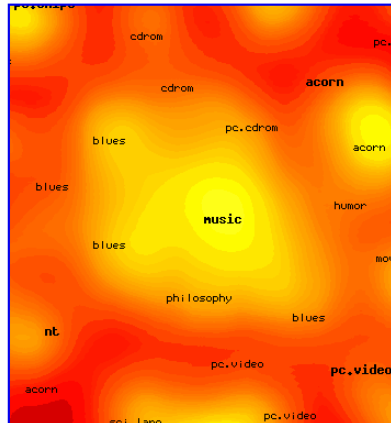
**Web-SOM**

Click any area on the map to get a zoomed view! Color denotes the *density* or the *clustering tendency* of the documents. Light (yellow) areas are clusters and dark (red) areas empty space, "ravines", between the clusters.

Document: Done (0.093 sec)

13:49

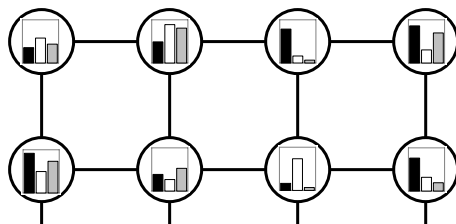
# Web-SOM



acorn - comp.sys.acorn.hardware  
blues - rec.music.bluenote  
cdrom - comp.publish.cdrom.hardware  
humor - rec.humor  
music - music  
nt - comp.os.ms-windows.nt.setup.hardware  
pc.cdrom - comp.sys.ibm.pc.hardware.cd-rom  
pc.video - comp.sys.ibm.pc.hardware.video  
philosophy - sci.philosophy.meta  
sci.lang - sci.lang

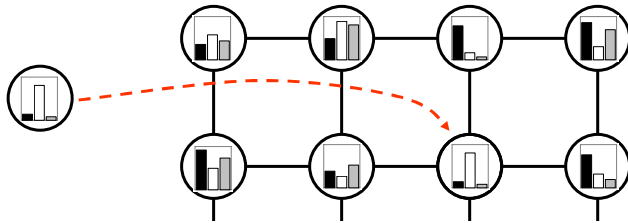
Click any area on the map to get a zoomed view!

# Initialisierung des Netzes



- Anordnung der Knoten in einem regelmäßigen Gitter
- Knoten:  $n$ -dimensionaler Vektor ( $n$ : Dimensionalität des Datensatzes)
- zufällige Initialisierung der Knotenvektoren

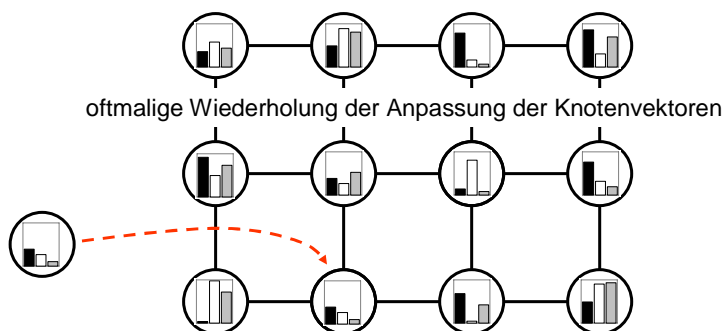
## Training



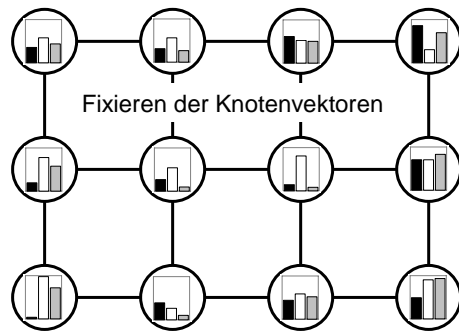
Für den ersten Datenvektor:

- Suche des ähnlichsten Knotenvektors
- Anpassung des ähnlichsten Knotenvektors an den Datenvektor
- Anpassung der Nachbarn des ähnlichsten Knotenvektors

## Training



## Trainiertes Netz

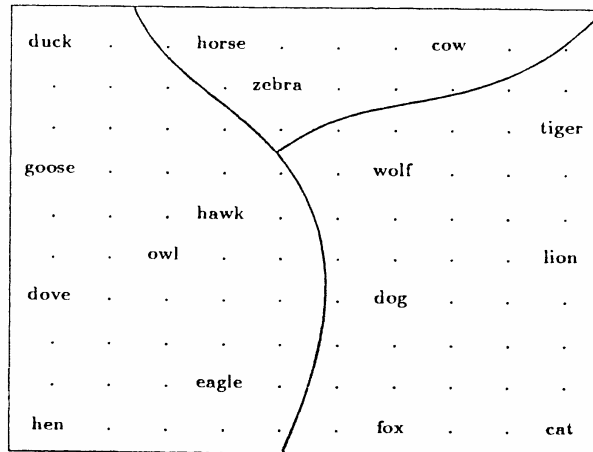


## Beispiel: Tiere

		d	h	d	g	o	h	e	w	t	h	z
		o	e	u	o	w	a	a	o	c	o	c
		v	n	c	s	l	w	g	l	a	r	b
		e	k	e	e	k	e	o	f	t	n	r
								f	d			a
								x	o			r
								o	g			s
												e
												a
												w
is	small	1	1	1	1	1	1	0	0	0	0	0
	medium	0	0	0	0	0	0	1	1	1	1	0
	big	0	0	0	0	0	0	0	0	0	1	1
has	2 legs	1	1	1	1	1	1	1	0	0	0	0
	4 legs	0	0	0	0	0	0	0	1	1	1	1
	hair	0	0	0	0	0	0	0	1	1	1	1
	hooves	0	0	0	0	0	0	0	0	0	0	1
	mane	0	0	0	0	0	0	0	0	0	1	1
feathers	1	1	1	1	1	1	1	0	0	0	0	
likes to	hunt	0	0	0	0	1	1	1	0	1	1	0
	run	0	0	0	0	0	0	0	1	1	0	1
	fly	1	0	1	1	1	1	0	0	0	0	0
	swim	0	0	1	1	0	0	0	0	0	0	0

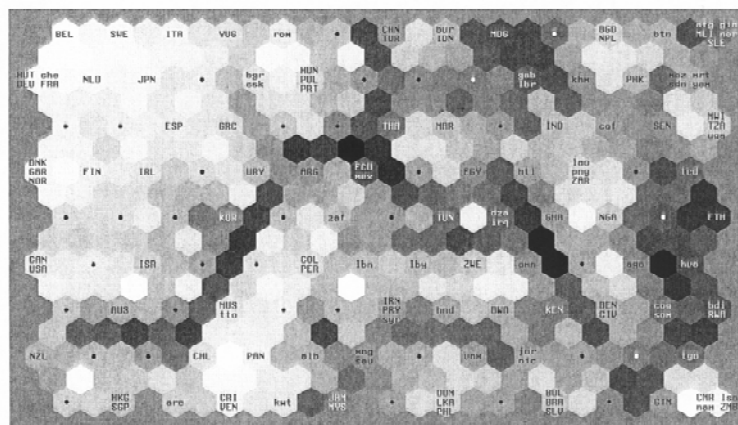
(Kohonen 2001)

## Beispiel: Tiere



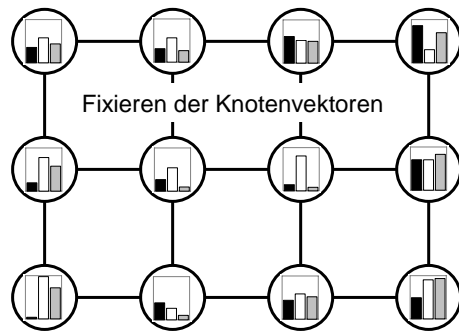
(Kohonen 2001)

## SOM Ländervergleich - U-Matrix



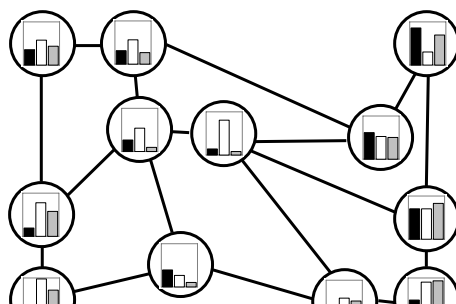
(Kohonen 2001)

## Trainiertes Netz



## Sammon's Mapping

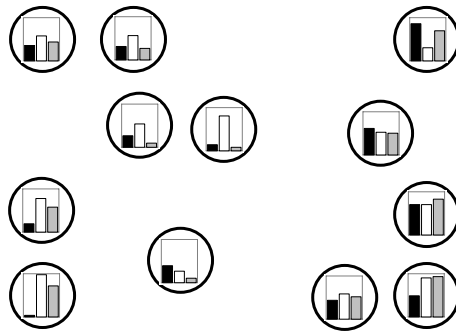
(Sammon 1969)



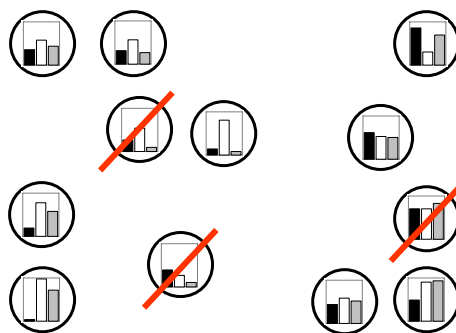
- Veränderung der Koordinaten der Knotenvektoren, so dass die Abstände im 2-dimensionalen Raum möglichst proportional der euklidischen Distanz



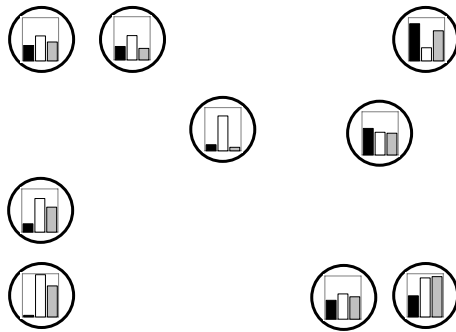
## Löschen der Verbindungen



## Löschen nicht genutzter Knotenvektoren



## Grafische Darstellung

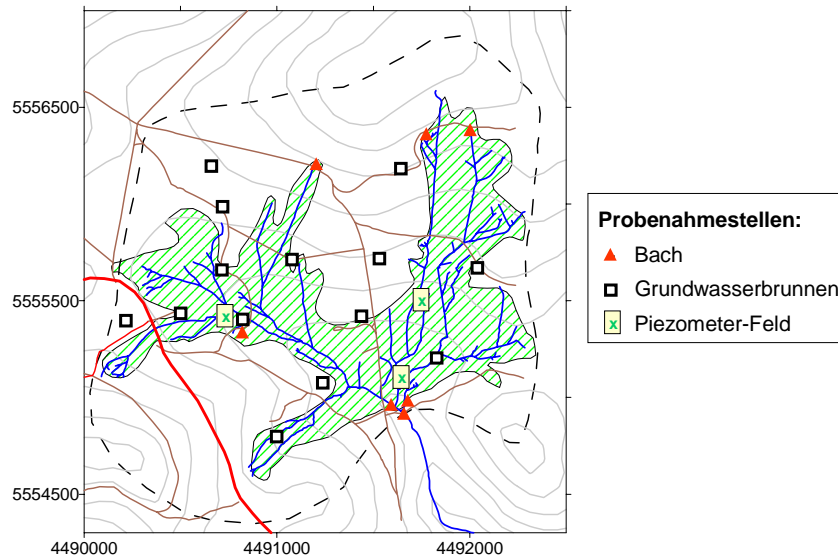


## Beispiel

**Daten:** Bach- und Grundwasserproben aus einem kleinen Einzugsgebiet  
(Bayerisches Landesamt für Wasserwirtschaft; BITÖK, Universität Bayreuth)

- 2641 Wasserproben
- 13 Parameter (*pH, Ca, Mg, Na, K, Al, Mn, Fe, Si, NO<sub>3</sub>, SO<sub>4</sub>, Cl, DOC*)
- Zeitraum: 1987 - 2002
- 38 Beprobungsstellen:
  - 7 Bäche
  - 13 "tiefe" Grundwasserbrunnen (10 - 24 m tief)
  - 18 Piezometer (max. 2 m tief)

## Lage der Messstellen

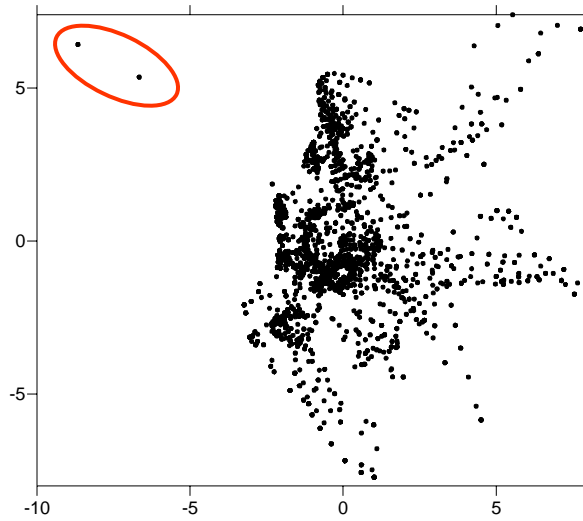


## Zum Vergleich: Hauptkomponentenanalyse (PCA)

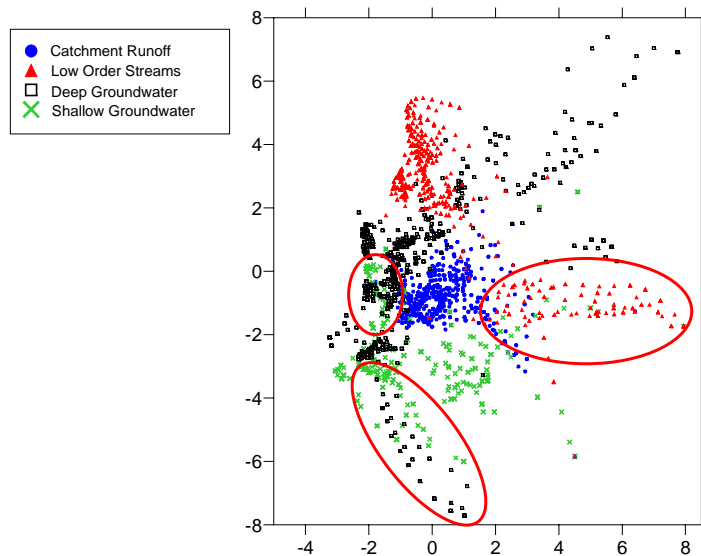
### Anteil der erklärten Varianz:

PCA, 4 Faktoren:	71.8%
PCA, 2 Faktoren:	50.8%
SOM, 2 Dimensionen:	77.4%

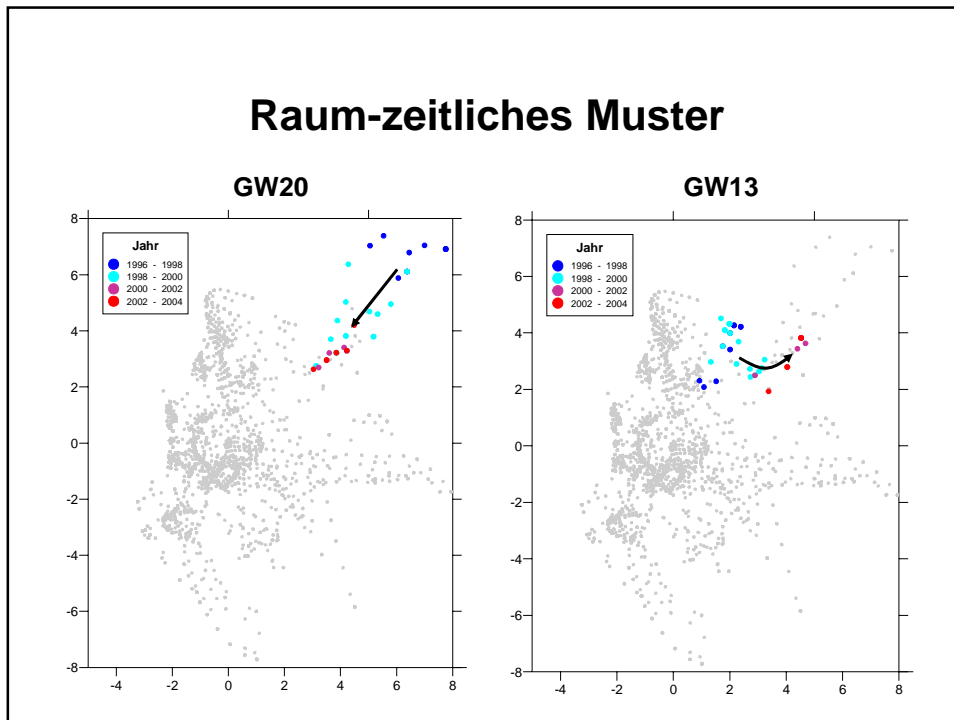
## Ausreißer



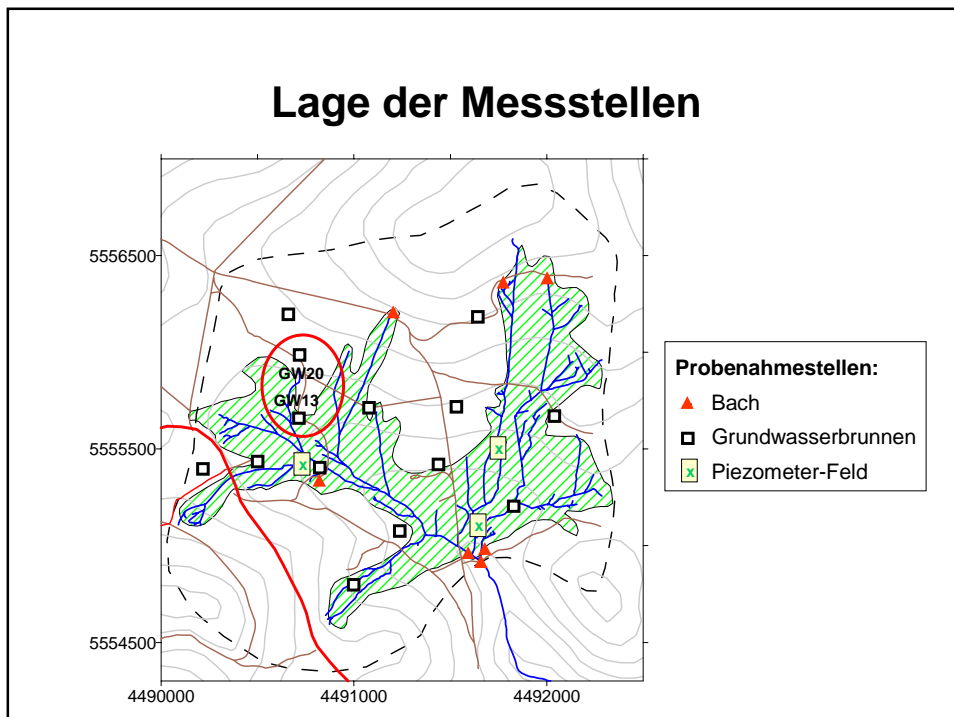
## Beprobungsstellen



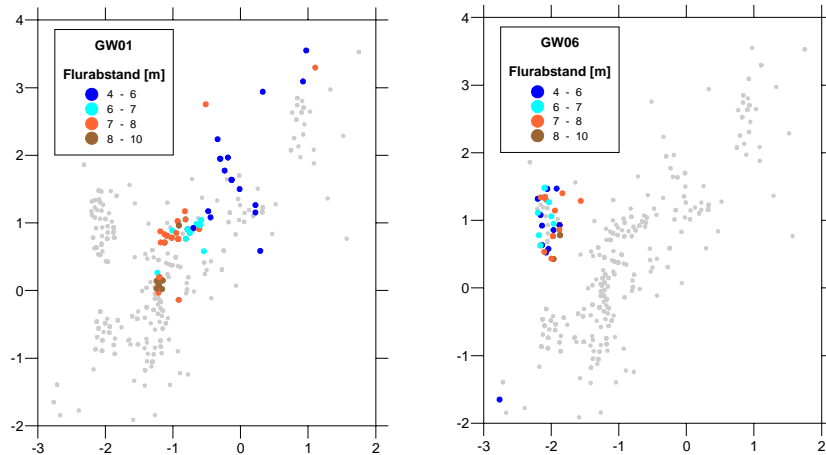
## Raum-zeitliches Muster



## Lage der Messstellen



## Kurzfristige Dynamik



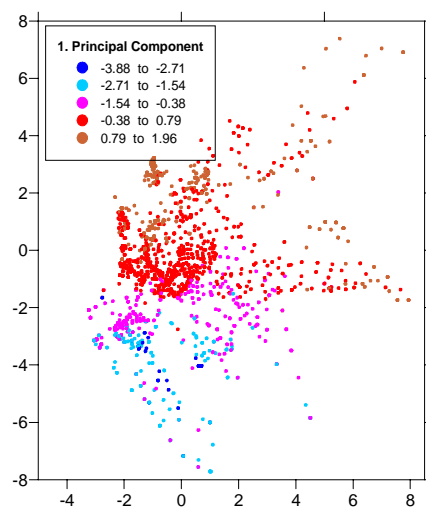
## Vergleich SOM - HKA

- Makrostruktur: Anordnung der Datenpunkte in der SOM oft ähnlich den ersten Hauptkomponenten
- Mikrostruktur: SOM bestimmt "lokale Hauptkomponenten"
- SOM berücksichtigt nicht-lineare Zusammenhänge
- Interpretation für einzelne Variablen bei SOM schwieriger

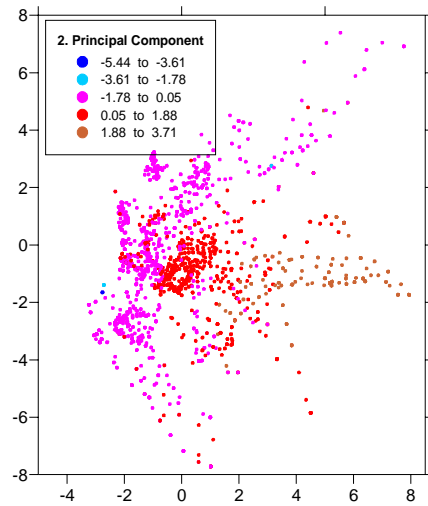
# Hauptkomponentenanalyse

	Principal Component			
	1	2	3	4
<b>Expl. Variance</b>	33.5%	17.3%	13.1%	7.9%
<b>Loading</b>				
Al	0.29	0.01	0.80	-0.18
Ca	0.80	0.16	0.22	0.08
Cl	0.19	0.91	-0.04	-0.10
DOC	-0.41	0.55	0.44	-0.22
Fe	-0.67	0.38	0.40	0.07
K	0.79	0.17	0.05	0.03
Mg	0.86	0.11	0.29	-0.01
Mn	0.07	-0.09	0.70	0.14
Na	0.05	0.90	-0.18	0.16
NO <sub>3</sub>	0.89	-0.11	0.14	-0.17
pH	0.07	-0.05	0.08	0.82
Si	-0.44	0.12	-0.14	0.46
SO <sub>4</sub>	0.56	-0.11	0.58	0.04

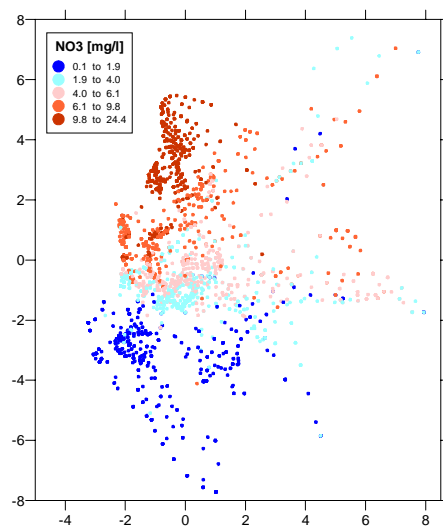
## 1. Hauptkomponente



## 2. Hauptkomponente

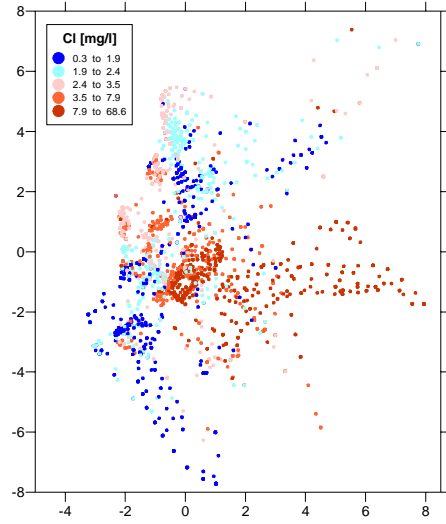


## NO<sub>3</sub>

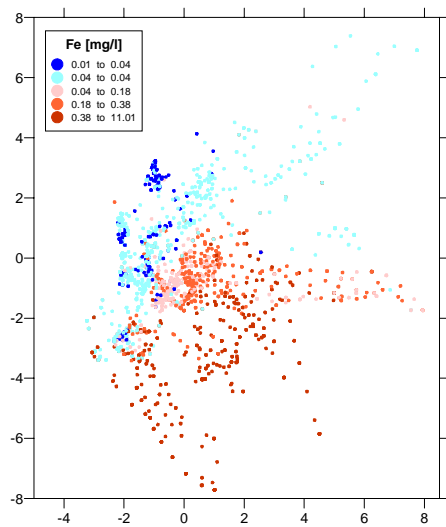




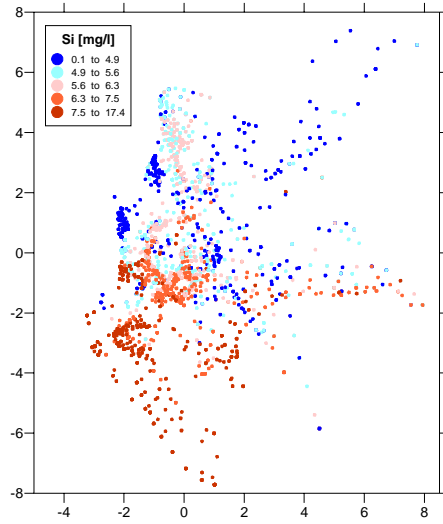
# Cl



# Fe

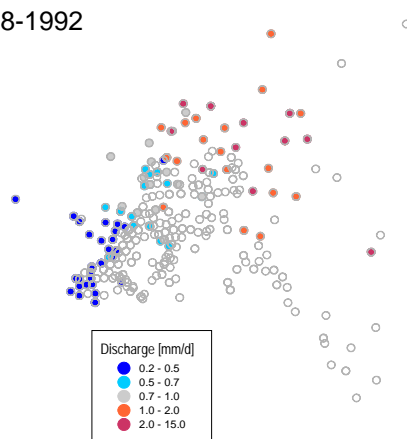


# Si



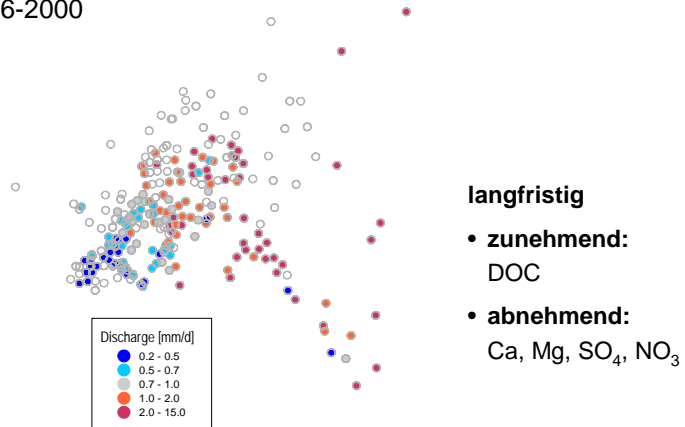
# Kurzfristige Dynamik

1988-1992



## Langfristiger Trend

1996-2000

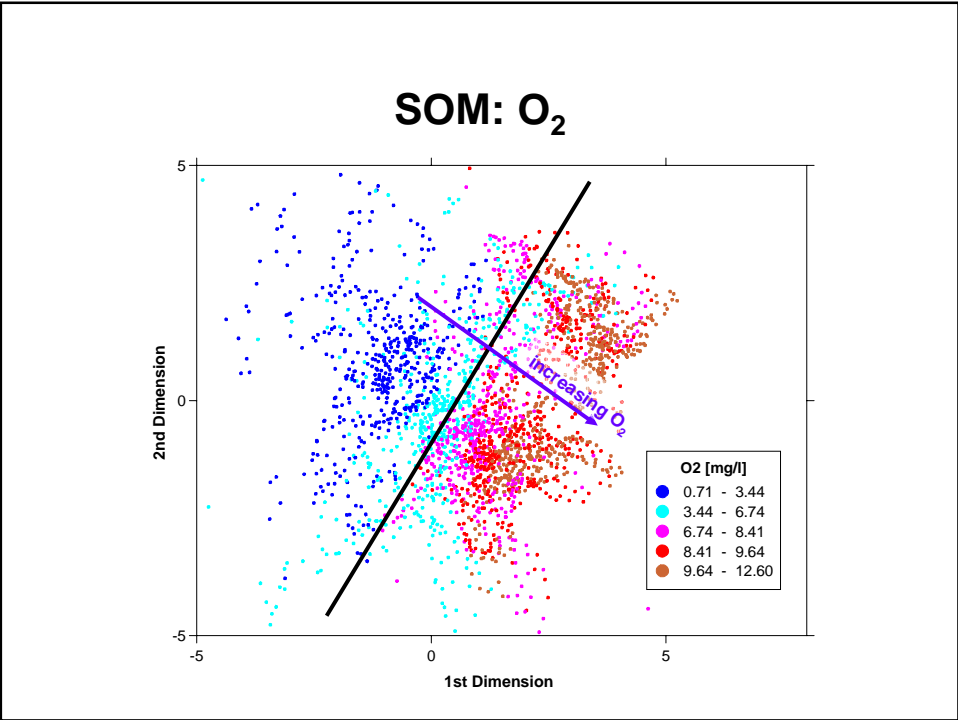
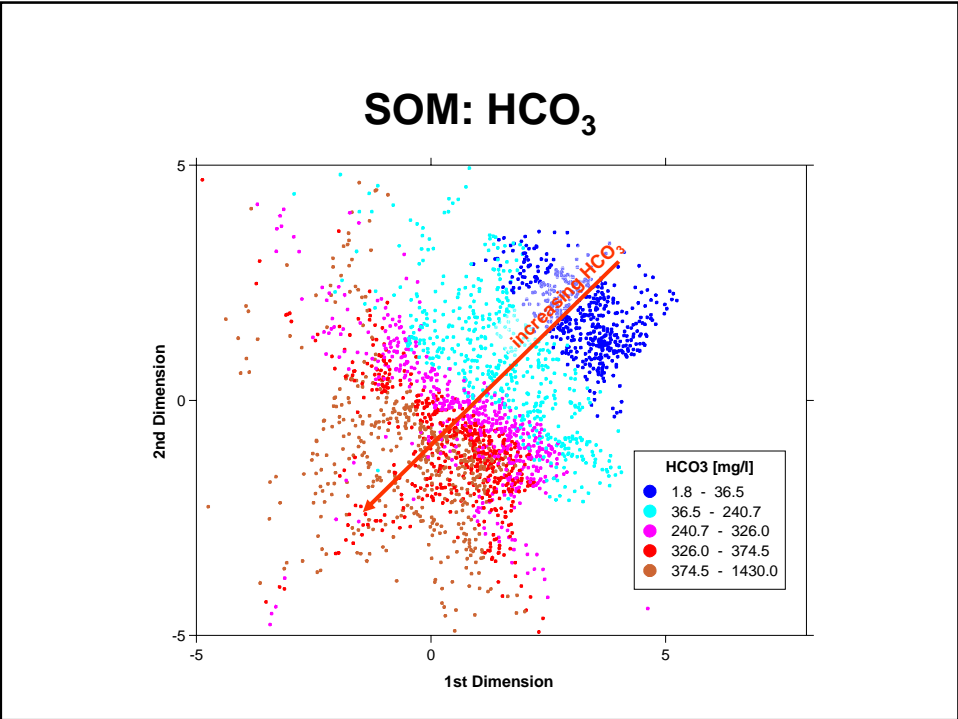


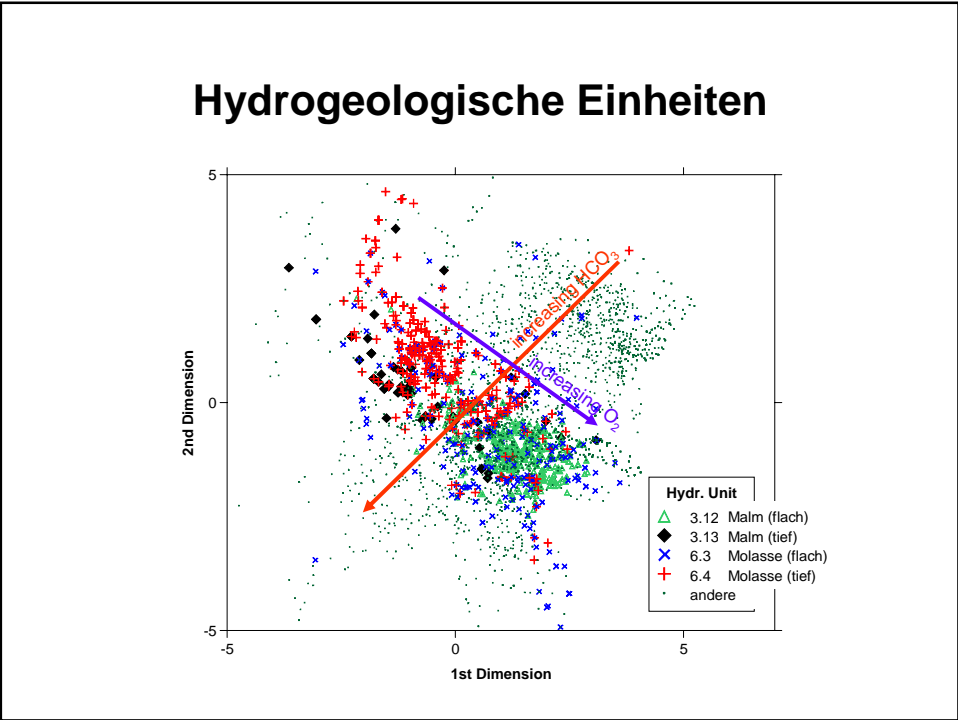
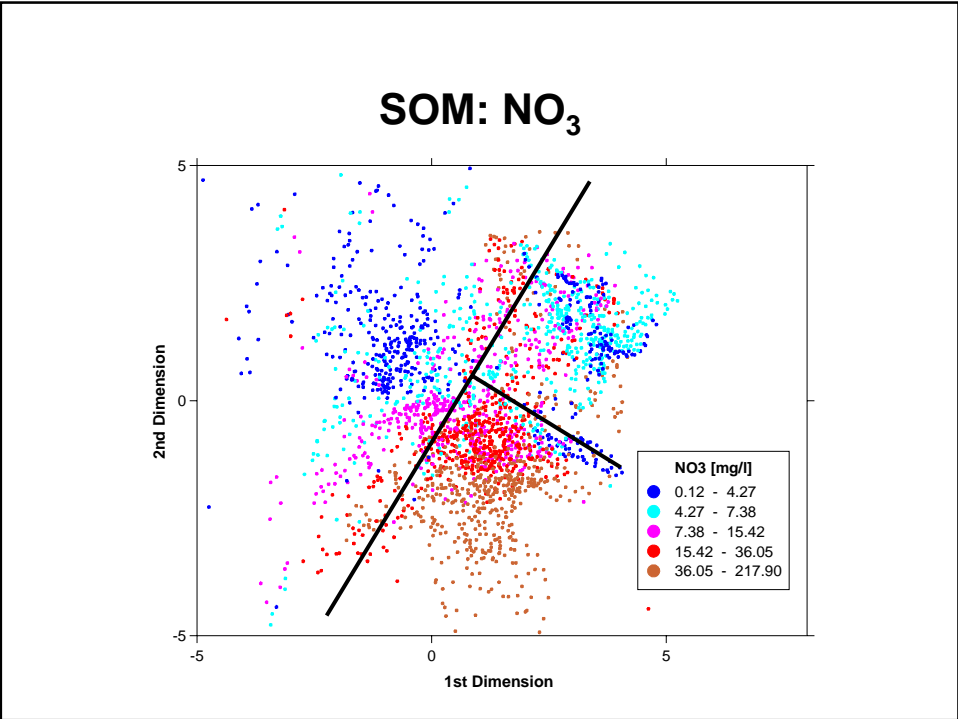
## Grundwasserbeschaffenheit in Bayern

**Daten:** Grundwasser-Monitoring (Bayerisches Geologisches Landesamt)

→ 5441 Probenahmestellen, 17 Parameter:

- Hintergrund-Daten (EU-WFD)
- 32 hydrogeologische Einheiten
- pH, Ca, Mg, Na, K, NH<sub>4</sub>, Fe, Mn, Cl, NO<sub>3</sub>, HCO<sub>3</sub>, SO<sub>4</sub>, NO<sub>2</sub>, Si, O<sub>2</sub>, CO<sub>2</sub>, Temperatur





# Lernende Vektor-Quantisierung (LVQ)

(Kohonen 1989)

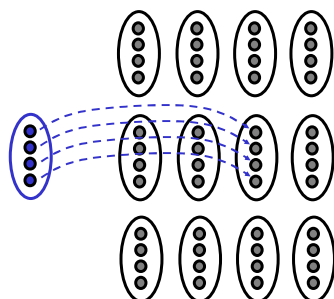
→ Verfahren der überwachten Klassifizierung (vergl. Diskriminanzanalyse)

Prinzip ähnlich dem der Selbstorganisierenden Karte, jedoch:

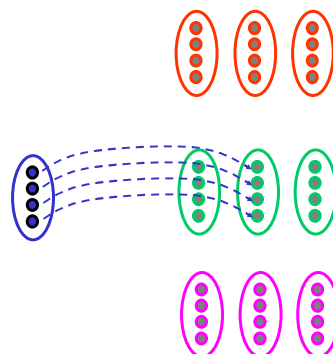
- Knotenvektoren = "Codebook Vectors" für jede Gruppe
- Anpassung der Codebook-Vektoren der zugehörigen Gruppe an die Datenvektoren
- (dem Datenvektor ähnliche Codebook-Vektoren anderer Gruppen werden so verändert, dass sie dem Datenvektor unähnlicher werden)
- Zuordnung zur Gruppe nach Fixieren des Netzes durch Ermittlung des ähnlichsten Codebook-Vektors

## Vergleich SOK - LVQ

Selbstorganisierende Karte (SOK)



Lernende Vektorquantisierung (LVQ)



# Hydrogeologische Einheiten (Bayern)

## Diskriminanzanalyse (50%)

		zugeordnete Einheit																				Anzahl							
		1.1	1.3	1.4	3.1	3.3	3.4	3.5	3.6	3.7	3.8	3.10	3.11	3.12	3.13	3.14	3.15	5	6.1	6.2	6.3		6.4	7.1	7.2	7.3	7.4	7.5	7.6
Hydrogeologische Einheit	1.1	50	11	11	4	6	0	0	0	0	0	0	0	9	2	0	0	0	4	0	0	1	1	0	0	1	0	0	91
	1.3	4	76	7	1	5	0	0	0	0	0	0	0	0	0	0	4	0	1	0	0	0	1	0	0	0	0	0	730
	1.4	7	5	75	5	0	0	0	2	0	0	0	0	0	0	4	0	0	2	0	0	0	0	0	0	0	0	0	57
	3.1	0	0	4	59	7	0	0	0	0	4	7	0	0	0	7	0	0	7	0	4	0	0	0	0	0	0	0	27
	3.3	8	10	1	5	47	3	1	1	1	9	1	1	1	0	1	1	4	2	0	3	0	0	0	0	0	0	0	294
	3.4	0	0	0	1	0	37	1	27	2	2	5	1	6	0	0	1	8	4	0	5	0	0	0	0	0	0	0	146
	3.5	0	0	0	0	0	8	84	0	7	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	76
	3.6	0	0	1	1	1	13	1	61	4	4	0	0	2	1	0	0	3	3	0	6	0	0	0	0	0	0	0	163
	3.7	0	0	0	0	0	5	19	0	70	2	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	57
	3.8	0	1	0	6	5	6	1	4	2	36	5	2	3	5	4	1	6	4	1	4	6	1	0	0	0	0	0	510
	3.10	0	0	0	7	0	3	0	0	0	0	53	13	0	0	0	3	10	3	0	0	0	0	0	0	0	0	0	30
	3.11	0	2	0	2	3	0	0	0	0	2	14	48	10	4	3	3	3	2	0	1	0	2	0	0	0	0	0	249
	3.12	0	0	0	0	0	1	0	4	0	2	4	12	36	1	0	15	1	15	4	5	0	0	0	1	0	0	0	655
	3.13	0	0	0	0	0	0	0	0	2	0	2	3	0	0	62	2	7	0	4	0	7	12	0	0	0	0	0	68
	3.14	0	9	0	3	0	0	0	0	3	3	0	6	0	0	74	0	3	0	0	0	0	0	0	0	0	0	0	35
	3.15	0	0	0	4	0	0	0	2	0	0	0	0	0	4	4	0	67	0	7	4	7	0	0	0	0	0	0	45
	5	0	4	0	4	6	8	2	3	3	3	4	5	0	1	3	9	40	2	0	1	3	0	0	1	0	0	0	103
	6.1	0	0	0	0	0	4	0	5	1	4	2	0	4	3	0	3	7	37	12	6	0	5	0	2	1	2	2	667
	6.2	0	0	0	0	0	1	0	0	0	1	1	4	1	0	2	0	17	35	4	1	12	3	2	2	8	7	1	260
	6.3	0	1	1	5	2	4	0	7	0	3	1	2	2	5	0	3	4	4	2	36	12	3	1	0	0	0	0	263
6.4	0	0	0	0	0	1	0	3	0	1	0	0	0	0	9	0	1	2	0	23	57	1	0	0	0	0	0	453	
7.1	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	17	0	2	67	5	3	0	2	58	
7.2	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	6	0	0	9	53	21	9	0	34	
7.3	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	3	0	0	3	15	67	0	5	0	60	
7.4	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	1	0	2	0	3	1	8	3	54	25	3	1	137	
7.5	0	0	0	0	0	0	0	0	0	0	2	0	2	0	0	0	0	0	0	0	2	3	0	2	5	79	8	66	
7.6	0	0	2	0	0	0	2	0	7	0	0	0	0	3	0	0	0	0	2	0	0	0	0	3	3	0	23	55	60

# Hydrogeologische Einheiten (Bayern)

## Lernende Vektorquantisierung (56%)

		zugeordnete Einheit																				Anzahl							
		1.1	1.3	1.4	3.1	3.3	3.4	3.5	3.6	3.7	3.8	3.10	3.11	3.12	3.13	3.14	3.15	5	6.1	6.2	6.3		6.4	7.1	7.2	7.3	7.4	7.5	7.6
Hydrogeologische Einheit	1.1	32	27	1	0	26	0	0	0	0	4	0	3	0	0	0	0	2	0	0	0	1	0	1	0	1	0	91	
	1.3	0	90	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	730	
	1.4	4	67	23	0	4	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	57
	3.1	4	22	7	11	22	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0	27
	3.3	0	19	0	0	64	1	0	0	1	7	0	0	3	0	0	0	0	0	0	3	0	0	0	0	0	0	0	294
	3.4	0	0	0	0	0	35	3	22	1	12	1	1	5	1	0	1	4	8	0	3	2	0	0	0	0	0	0	146
	3.5	0	0	0	0	0	4	88	1	4	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	76
	3.6	0	0	1	1	0	1	2	4	72	1	1	7	0	0	1	1	0	0	1	2	0	3	2	0	0	0	0	163
	3.7	0	0	0	0	0	2	19	4	67	4	0	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	57
	3.8	2	5	1	1	4	1	0	1	1	58	0	2	3	1	1	0	0	4	0	1	14	0	0	0	0	0	0	510
	3.10	0	0	3	0	7	0	3	0	3	7	30	17	13	0	0	0	0	7	0	0	7	0	0	0	0	0	0	30
	3.11	1	5	0	0	4	0	0	0	0	4	0	49	18	1	1	0	0	4	1	0	4	0	1	3	0	0	0	249
	3.12	0	0	0	0	2	1	0	0	0	15	0	8	57	0	0	0	1	10	2	0	4	0	0	1	0	0	0	655
	3.13	0	1	0	0	1	0	0	1	0	0	0	3	3	56	0	0	0	7	0	0	25	0	0	1	0	0	0	68
	3.14	3	23	0	0	6	0	0	0	0	17	0	3	0	0	0	43	0	0	0	3	0	0	0	0	0	3	0	35
	3.15	0	0	0	2	0	0	0	2	0	2	0	11	38	7	0	18	2	9	0	0	9	0	0	0	0	0	0	45
	5	2	11	0	1	5	2	3	2	0	5	2	4	6	2	0	0	40	4	0	2	6	0	0	0	2	2	0	103
	6.1	0	0	0	0	1	5	1	3	0	6	0	3	18	1	0	1	2	33	10	2	4	1	0	1	2	1	2	667
	6.2	1	0	0	0	3	3	0	0	0	3	1	3	13	1	0	1	1	12	28	1	3	3	3	3	3	9	4	260
	6.3	0	5	1	0	3	1	2	8	0	13	1	2	4	0	0	1	0	6	2	17	30	0	0	0	2	0	0	453
6.4	0	0	0	0	1	0	0	0	0	4	0	0	1	2	0	0	0	2	0	2	86	0	0	0	0	0	0	0	0
7.1	2	0	0	0	0	0	0	0	0	2	0	5	22	2	0	0	2	7	10	0	7	33	2	2	3	2	0	58	
7.2	6	0	0	0	0	0	0	0	0	3	0	6	0	0	0	0	3	0	3	0	0	0	71	6	0	0	0	34	
7.3	2	0	0	0	0	0	0	0	0	0	3	2	0	0	0	0	0	3	0	2	0	8	62	17	2	0	0	60	
7.4	1	1	1	1	1	0	0	0	0	3	1	4	1	0	0	0	0	1	1	2	1	4	4	64	7	1	1	137	
7.5	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	2	0	0	3	0	2	2	18	62	2	66	
7.6	0	0	0	0	0	2	8	0	5	0	0	0	0	0	0	0	2	0	0	0	0	0	2	3	5	8	7	58	60

## Aufgabe

1. Speichern Sie den Datensatz z-normiert als ASCII-Datei ab. Fügen Sie eine erste Zeile hinzu, die die Zahl der Parameter angibt.
2. Führen Sie mit dem Datensatz Sammon's Mapping durch und bestimmen Sie die Dimension des Netzes (s. *SOM\_Muster.txt*).
3. Erstellen Sie eine Selbstorganisierende Karte. Setzen Sie die Anzahl der Knoten ca. auf das Doppelte der Zeilenzahl des Datensatzes.
4. Führen Sie mit dieser Karte Sammon's Mapping durch.
5. Importieren Sie die Ergebnisse in Excel, Statistica oder Surfer und vergleichen Sie die Ergebnisse mit denen der Hauptkomponentenanalyse oder Clusteranalyse.