

Diskussion des Artikels

**„Non-linear Approach
to Grouping, Dynamics and Organizational Informatics
of Benthic Macroinvertebrate Communities in Streams
by Artificial Neural Networks“**

von Tae-Soo CHON, Young-Seuk PARK, Inn-Sil KWAK und Eui Young CHA

Hauptseminar
Prof. Dr. Michael Matthies

Inhaltsverzeichnis

Einleitung	2
Das Verfahren der Self-Organizing Map (SOM)	2
Anwendung der SOM	4
Das Verfahren der Adaptive Resonance Theory (ART)	7
Anwendung der ART	7
Anwendung von verschiedenen Netzarchitekturen zur Vorhersage	8
Anwendung eines Counterpropagation Networks zur Nachbildung von Biozönosen	13
Anwendungen zum Exergie-Begriff	14
Formale Mängel und inhaltliche Fehler	16
Literaturverzeichnis	17
Anhang: „Schrebergärten“	18

Einleitung

Die Arbeit steht im Zusammenhang mit der Entwicklung nachhaltiger Managementstrategien für Gewässer. An Wasser werden vielfältige Nutzungsansprüche gestellt: Ernährung (Trinkwasser), Landwirtschaft (Bewässerung), Industrie, Energieerzeugung (Kühlwasser), Erholung (Wassersport, Angeln).

Gewässergütebestimmung ist mit physikalischen, chemischen oder biologischen Methoden möglich [6]. Mit biologischen Methoden lassen sich auch kurzfristige Kontaminationen erfassen, da Populationen träge auf Änderungen von physikalischen und chemischen Parametern reagieren. Während man sich beim Saprobienindex die Beziehung zwischen stenöken Spezialisten und der Verschmutzung ihres Lebensraums zunutze macht (Ubiquitisten, deren Ansprüche nicht so eingeschränkt sind, haben hier nur eine geringe Aussagekraft), wurde in dieser Arbeit dieses Expertenwissen außen vor gelassen und durch Anwendung neuronaler Netze auf ganze Lebensgemeinschaften ersetzt. Die Gaia-Hypothese wird kurz erwähnt, welche besagt, dass Lebensgemeinschaften tiefgreifend auf abiotische Faktoren des Ökosystems zurückwirken und erst so ihr Dasein ermöglichen [7,S.38].

Es wurden Flüsse in den urbanen Räumen Seoul (10 Mio. Einwohner) und Pusan (3 Mio. Einwohner) untersucht hinsichtlich der Besiedlung durch die benthische (Bodenbewohner) wirbellose Fauna. Hierzu gehören Arten wie *Chironomus* (rote Zuckmückenlarve, auch adulte betrachtet), *Hirudinea* (Egel), *Gastropoda* (Schnecken), *Oligochaeta* (Wenigborster, z.B. Regenwürmer). Ihnen wird eine zentrale Rolle zugeschrieben, da sie in der Nahrungskette die Produzenten mit den Top-Räubern verbinden. Von Vorteil ist ihre Sesshaftigkeit. Ferner wurden in einigen Kapiteln physiko-chemische Umweltdaten (Fließgeschwindigkeit, Tiefe, sedimentierte organische Masse, Substratvolumen < 0,5 mm) hinzugezogen.

Der Han River, der in Seoul eine respektable Breite von einem Kilometer aufweist, wäre aber auf dem Weg der Erholung, da die Stadtverwaltung Therapie-Maßnahmen vornähme [2,S.104].

Konventionelle statistische Verfahren sind meist auf eindimensionale Daten beschränkt und bereiten Probleme, wenn die Daten lückenhaft sind. Nun wurden für verschiedene Fragestellungen unterschiedliche Netzarchitekturen angewendet.

Das Verfahren der Self-Organizing Map (SOM)

Die SOM geht auf KOHONEN zurück und ist in der Lage, mehrdimensionale Cluster in eine sogenannte topologische Karte mit weniger Dimensionen zu projizieren. Werden weniger als vier Dimensionen verwendet, hat dieses Verfahren den Vorteil, sehr anschaulich zu sein. Bei einer eindimensionalen Karte ist auch die Anordnung in einem Ring denkbar, bei einer zweidimensionalen Karte entsprechend mit Torus-Randbedingungen. Es kann auch ein hexagonales Gitter vorliegen. Eine geeignete Größe des Gitters wird empirisch bestimmt, je nachdem wie schnell der Algorithmus konvergiert bzw. die Daten getrennt werden können [3,S.154]. Liegen die Daten nicht numerisch vor, müssen sie vorher chiffriert werden.

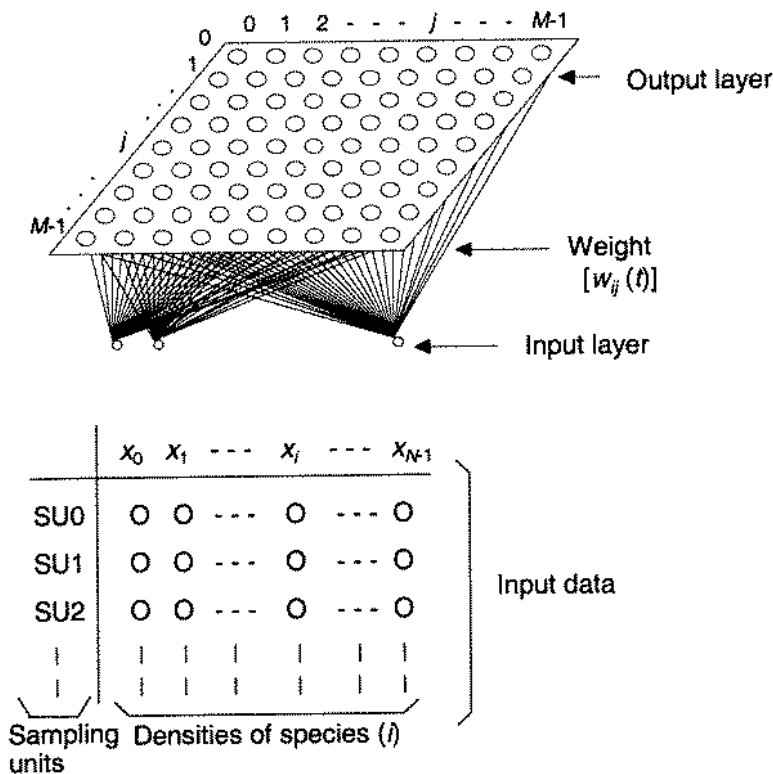


Abb. 8.2

In diesem Netzwerk ist jedes Neuron i des Input-Layers über eine Wichtung w mit jedem Neuron j des Output-Layers verbunden. Diese Wichtungen werden mit Zufallszahlen initialisiert. Darum, und wegen der iterativen Vorgehensweise, können sich bei mehreren Läufen leicht auch andere (oft nur verdrehte, gespiegelte) Anordnungen ergeben. Dieser Umstand macht eine Objektivierbarkeit schwierig [1,S.73].

Jetzt wird der SOM ein erstes Cluster x präsentiert und für alle Neuronen im Output-Layer die Distanz d berechnet:

$$d_j(t) = \sum_{\text{über alle Inputs } i=0 \text{ bis } N-1} (x_i - w_{ij}(t))^2$$

Das Neuron mit der geringsten Distanz „gewinnt“, d.h. es bildet x am besten ab. Damit sich x beim Gewinner-Neuron noch heimischer fühlt, werden in seiner Umgebung die Wichtungen an x angepasst:

$$w_{ij}(t+1) = w_{ij}(t) + \eta(t) * (x_i - w_{ij}(t)) * Z_j$$

Dabei ist η die Lernrate (wird meist mit 0,1-0,4 angesetzt) und Z_j die Umgebung (1 für das Gewinner-Neuron und dessen Nachbarn, sonst 0). Die Änderung der Wichtungen kann auch in Abhängigkeit von der Distanz zum Gewinner-Neuronen erfolgen (etwa kegelförmig). Schließlich werden Lernrate und Umgebung noch etwas verkleinert. Manchmal geht man auch zwei Phasen durch (ordering and tuning phase).

Nun wird mit anderen Clustern genauso verfahren.

In [1,S.70] wird noch eine Erweiterung mit dem „conscience algorithm“ (eine Art Gedächtnis) vorgestellt. Um die Gewinnchancen der Neuronen anzugleichen, werden jene bevorzugt, die lange Zeit kein Pattern mehr abbekommen haben:

$$d'_j(t) = d_j(t) - b_j(t)$$

$$\text{mit } b_j(t) = \gamma * (1/M^2 - f_j(t))$$

$$\text{und } f_j(t) = f_j(t-1) + \beta * (Z_j - f_j(t-1))$$

β wurde mit 0,0001 und γ mit 10 angesetzt, f_j zufällig initialisiert. Z_j bezieht sich hier auf den letzten Gewinner, der mit seinen Nachbarn mit einem hohen f_j -Term gestraft wird (wird zu einem niedrigen b_j -Term und einen höheren d'_j -Term). Insgesamt kommt so mehr Abwechslung zustande.

Im Anhang steht ein frei erfundenes Beispiel, das den Ablauf illustriert.

Anwendung der SOM

Die SOM wurde hier verwandt, um Flussabschnitte zu verschiedenen Jahreszeiten anhand der aufgefundenen Artenzusammensetzung zu gruppieren. Über einen Zeitraum von zwei Jahren wurden an elf verschiedenen Probenahmeorten aus dem Pusan-Gebiet 99 Spezies betrachtet. Für eine andere Untersuchung am Han River werden die Daten dargestellt:

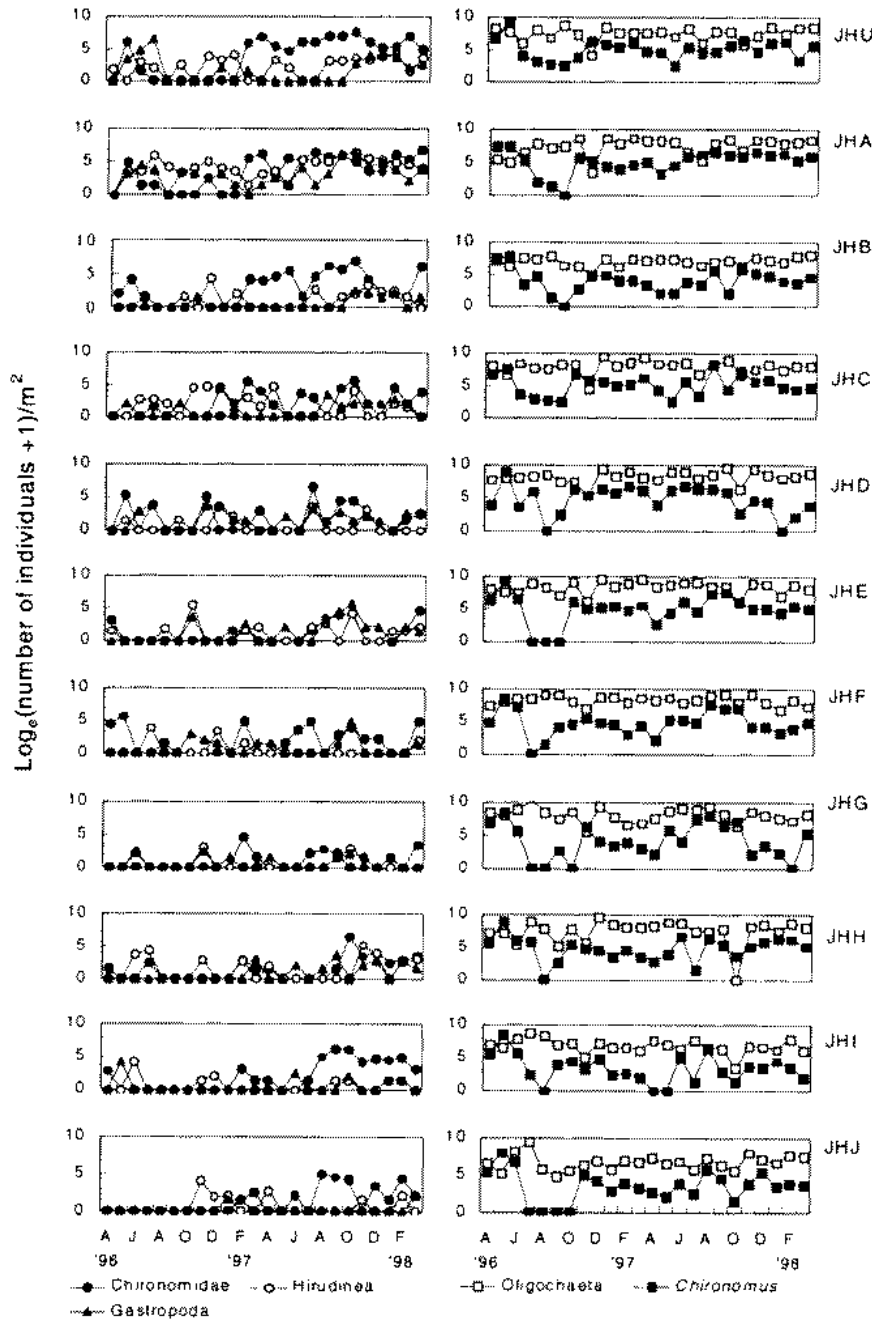


Abb. 8.1

In den Spalten die einzelnen Probenahmeorte, die Aufteilung der fünf Arten in zwei Graphen soll der Übersichtlichkeit dienen. Eine jahreszeitliche Dynamik ist auf den ersten Blick nicht erkennbar. Ob ein egales/neutrales Nebeneinander, Konkurrenz oder eine andere Wechselwirkung zwischen den Arten vorliegt, wird nicht gesagt.

In der Vorverarbeitung wurden 4 Ausreißer aussortiert, die 56 verbleibenden Proben log-transformiert (um geringe Dichten besser sichtbar zu machen) und auf das Intervall $[0,01;0,99]$ normalisiert. Das Input-Layer setzt sich jetzt aus Knoten für die Artendichten der einzelnen Spezies zusammen.

Konvergenz wurde nach 500-1000 Iterationen erreicht. Danach bot sich folgendes Bild auf dem 9x9-Gitter:

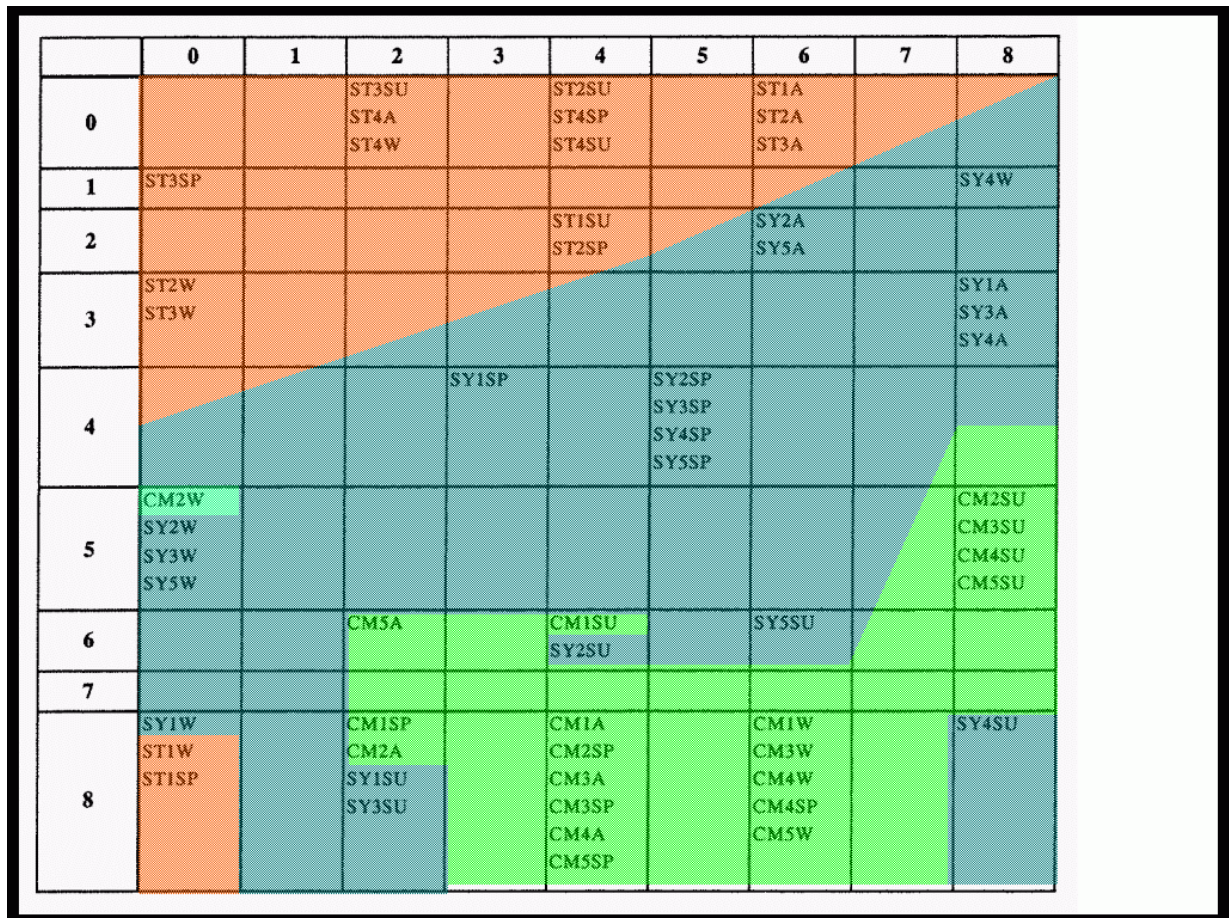


Abb. 8.3 farblich nach Flüssen unterteilt

Im Probencode bezeichnen die ersten beiden Buchstaben den Fluß (CholMa, SokTae und SuYong), die Ziffer den Flussabschnitt und der/die letzte/n Buchstaben die Jahreszeit (SPring, SUMmer, Autumn, Winter).

Die grobe Ordnung erfolgte gemäß der einzelnen Flüsse, die sich auch durch unterschiedliche Grade der Verschmutzung auszeichnen. Dabei wurden Gemeinschaften mit einem höheren Saprobienindex (z.B. der Soktae-Fluß in der linken oberen Ecke) etwas besser versammelt als jene mit einem geringeren Saprobienindex. Auf einer kleineren Skala fand sich eine Ordnung der Jahreszeiten (siehe die Neuronen (8,5) und (0,5)). Eine hierarchische Ordnung erscheint also machbar.

Die Autoren warnen vor dem Trugschluss, aus der Distanz auf der SOM eine adäquate Ähnlichkeit ableiten zu können, da die Daten in einen Raum geringerer Dimension hineingepresst wurden. Genauso fraglich erscheint der Versuch, zeitliche und räumliche Pfade auf der SOM auszumachen [1,S.75].

In einem zweiten Schritt wurden der fertigen SOM neue Proben zur Einordnung gegeben, die sich entsprechend ihrer Präferenz auf dem Gitter verteilten. Die SOM erweist sich hier als Alternative zur Principal Component Analysis, die im Kapitel 7 vorgestellt wird. Im Gegensatz zum Dendrogramm beim „Classifying“ werden Überschneidungen von Gruppen zugelassen. Für diesen Vorgang prägen die Autoren die Vokabel „Patternizing“ (liefert unter www.google.de 157 Ergebnisse statt 408000 für „Patterning“).

Dann wurde die SOM benutzt, verschiedene Spezies zu gruppieren, die nach ihrer Häufigkeit und bevorzugten Saprobität klassifiziert worden waren (S.136). Mit ihren zwei Dimensionen hatte die SOM leichtes Spiel, nach diesen zwei Kriterien zu ordnen. Auf der horizontalen Achse ergab sich ein Gradient für die Saprobität, auf der vertikalen Achse einer für die Abundanz. (Es wäre hier denkbar, die SOM von vornherein mit entsprechenden Gradienten auszustatten – also keine zufällige Initialisierung.) Nur für die häufigsten Arten wird die Kennziffer angegeben.

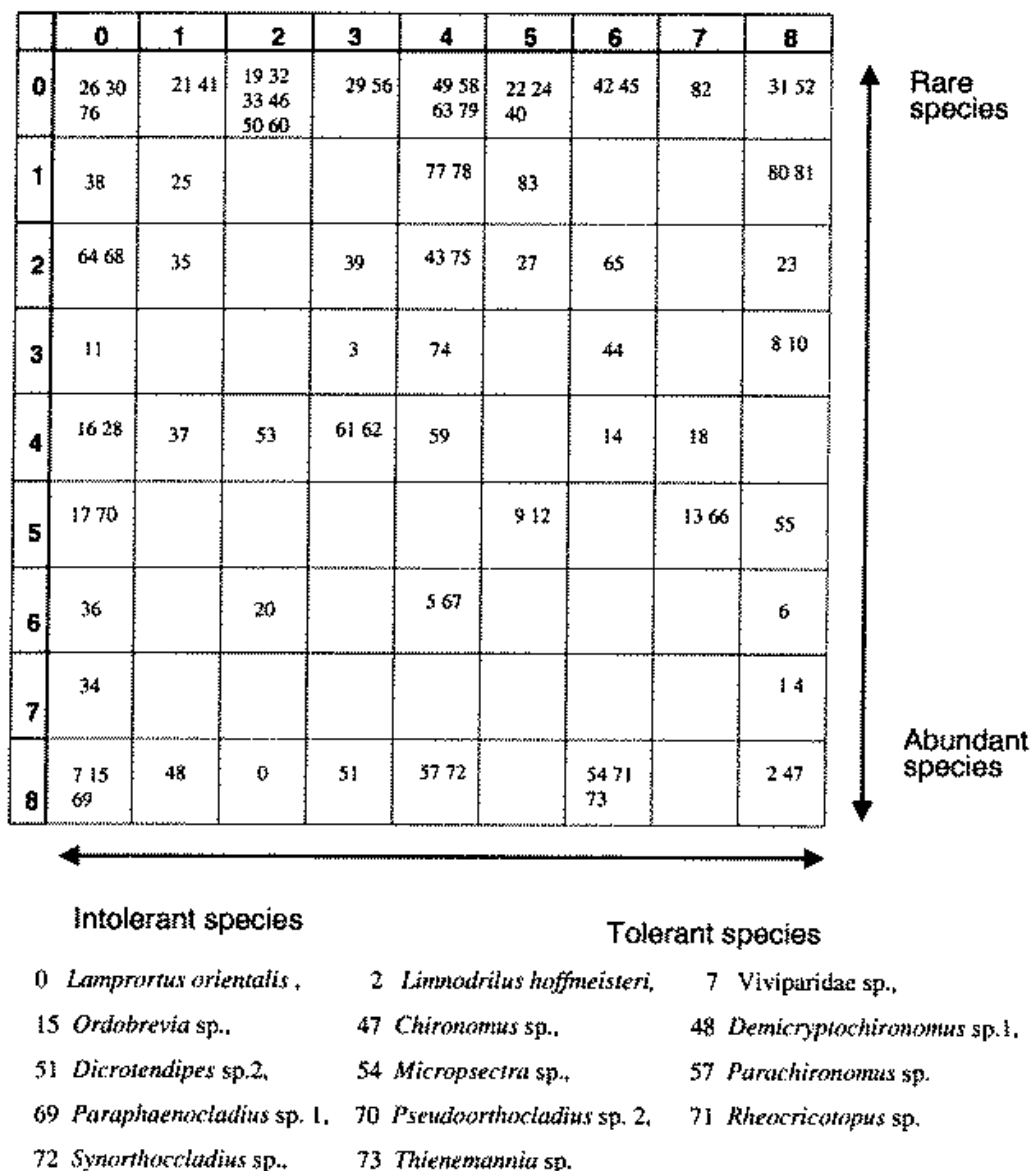


Abb. 8.6

Später im Artikel wird noch eine, hinsichtlich Raum und Zeit großskalige, Klassifizierung durchgeführt (S.142). Aus 14 Artikeln wurden für 23 Nebenflüsse aus ganz Südkorea Daten entnommen, die einen Zeitraum von 12 Jahren abdecken. Besonders hier bietet sich die Einordnung neuer Datensätze an, um ihren ökologischen Status mit schon vorhandenen Daten zu vergleichen.

Von MELSEN et al. gäbe es eine Implementierung der SOM für parallel processing [1,S.77].

Das Verfahren der Adaptive Resonance Theory (ART)

Dieses Verfahren wird immer mit anderen Netzen kombiniert, hier in der Reihenfolge ART-SOM als auch umgekehrt. Die Wichtungen b_{ji} zwischen Knoten i im Input-Layer und Knoten j im Output-Layer werden wiederum anfangs mit Zufallszahlen bestückt. Der Output-Knoten j mit der geringsten euklidischen Distanz

$$d_j(t) = \sqrt{(\sum \text{über alle Inputs } i = 0 \text{ bis } N-1 (b_{ji} - x_i)^2)}$$

wird mit einem Asteriskus * ausgezeichnet. Nun wird $d_{j^*}(t)$ mit einem Schwellwert ρ verglichen (hier 0,61, wurde empirisch ermittelt [3,S.154]). Im Fall $d_{j^*}(t) < \rho$ kann das Muster einem schon bekannten Knoten zugeordnet werden, die Wichtung wird folgendermaßen angepasst:

$$b_{j^*i}(t+1) = (c * b_{j^*i}(t) + x_i) / (c+1)$$

andernfalls wird ein neuer Knoten erzeugt:

$$b_{j^*i}(t+1) = x_i$$

So soll sichergestellt werden, daß bereits trainierte Knoten nicht wieder zerstört werden.

Anwendung der ART

Auch hier wurden Datensätze gruppiert, diesmal nach Monaten unterschieden. Da die Zahl der Spezies mit 132 zu aufwendig für eine Berechnung war, wurden sie zu 7 Taxa zusammengefaßt. Der Input umfaßte nun die Artendichte und die Zahl der gefundenen Arten im jeweiligen Taxon und zusätzlich die gesamte Populationsdichte und die Gesamtzahl der Arten, also $7*2+2=16$ Knoten im Input-Layer. So verteilten sich die 84 Datensätze auf 11 Neuronen, welche wiederum in eine SOM gepackt wurden.

a)

Neuron	Number	Sampling units
0	14	YIG3-9 YIG3-10 YIG3-11 YIG4-7 YCK3-9 YCK3-10 YCK4-3 YSC3-9 YSC3-10 YSC3-12 YSC4-7 TSD3-9 TKC3-9 TKC4-8
1	12	YIG3-12 YIG4-1 YIG4-2 YIG4-3 YIG4-4 YIG4-5 YCK3-12 YSC4-2 YSC4-4 TKC4-1 TKC4-4 TKC4-5
2	12	YIG4-6 YIG4-8 YCK4-1 YCK4-2 YCK4-4 YCK4-5 YCK4-6 YCK4-7 YCK4-8 TSD4-1 TSD4-4 TSD4-7
3	6	YCK3-11 YSC3-11 YSC4-6 YSC4-8 TSD4-5 TKC3-10
4	2	YSC4-1 YSC4-3
5	3	TSD3-10 TSD3-11 TSD3-12
6	4	TSD4-2 TSD4-3 TSD4-6 TSD4-8
7	2	TKC4-2 TKC4-3
8	2	TKC4-6 TKC4-7
9	8	TKC3-11 THP3-9 THP3-10 THP3-11 THP4-6 TCL3-11 TCL3-12 TCL4-6
10	19	YSC4-5 TKC3-12 THP3-12 THP4-1 THP4-2 THP4-3 THP4-4 THP4-5 THP4-7 THP4-8 TCL3-9 TCL3-10 TCL4-1 TCL4-2 TCL4-3 TCL4-4 TCL4-5 TCL4-7 TCL4-8

Abb. 8.8

Hier stehen der zweite und dritte Buchstabe für den Ort der Probenahme (SaDeungkol, KoChon, HaPansong, ChungLi, ImGog, ChangKi, ShinChon), die erste Zahl für das Jahr 199x und die zweite für den Monat.

Das Verfahren umgedreht – erst SOM, dann ART – wäre weniger erfolgreich verlaufen. Weitere Untersuchungen wären hier notwendig.

Jetzt sollten auch Änderungen in Lebensgemeinschaften erfasst werden. Dazu werden jedem Datum ein oder zwei Vormonate angefügt, dieses Zeitfenster verschiebt sich über den Beobachtungszeitraum.

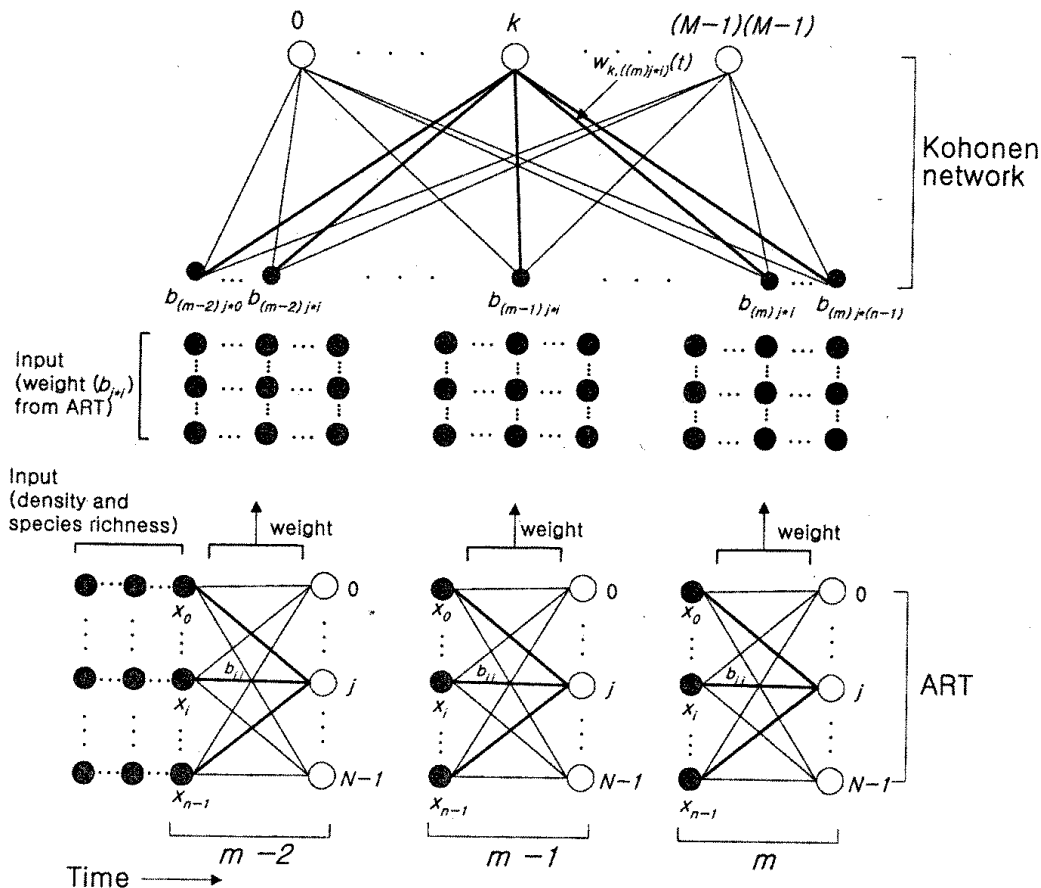


Abb. 8.7

Wenn Daten fehlten, wurde der Mittelwert aus dem vorhergehenden und dem nachfolgenden Monat gebildet [3,S.155]. Auch hier gelang die Einordnung neuer Daten. Die SOM hatte ihre Stärken beim Vorliegen vieler Nullen, wohingegen das ART-Verfahren Vorteile bei gleichmäßigen Daten hatte.

Anwendung von verschiedenen Netzarchitekturen zur Vorhersage

Es ist nützlich, die Reaktion einer aquatischen Lebensgemeinschaft auf eine Kontamination vorhersagen zu können. Die meisten Modelle waren bisher statisch, d.h. die Eingangsdaten stammten vom Zeitpunkt der „Vorhersage“. Erst in letzter Zeit werden dynamische neuronale Netze benutzt, so z.B. für Zeitreihen über die Dicke von Jahresringen.

In diesem Artikel wurden vier Architekturen eingesetzt, um ein Datum mit einem vorhergehenden Zeitfenster von ein bis fünf Monaten Länge zu matchen:

1. Multilayer Perceptron
2. rekurrentes neuronales Netz nach ELMAN
3. Real-Time Recurrent Network (RTRN)
4. RTRN mit Umweltdaten

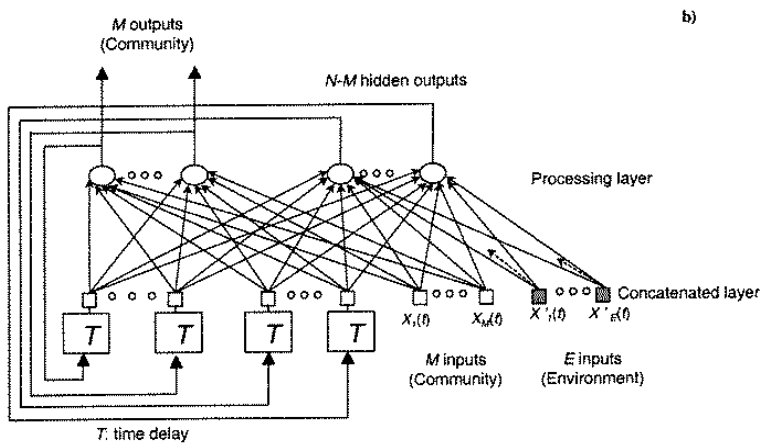
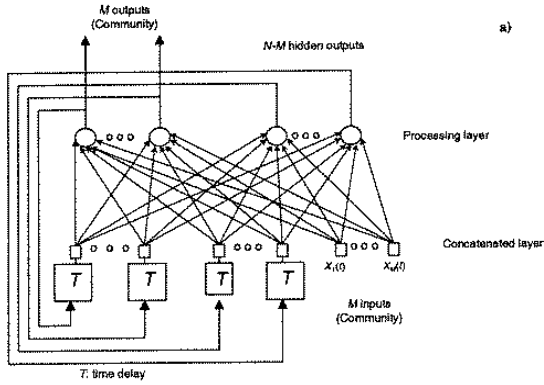
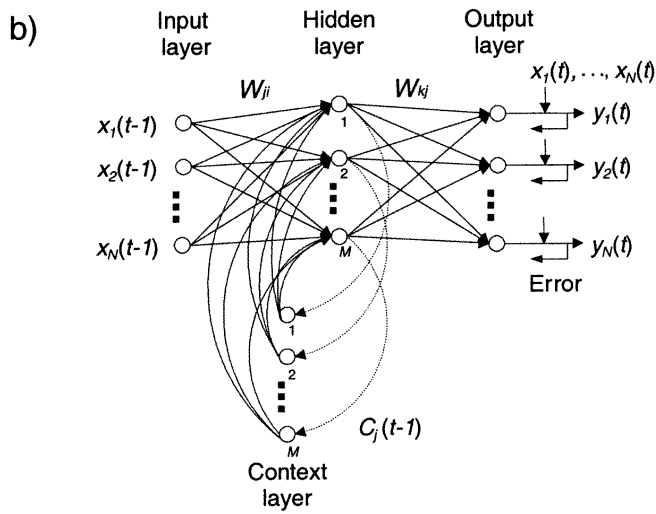
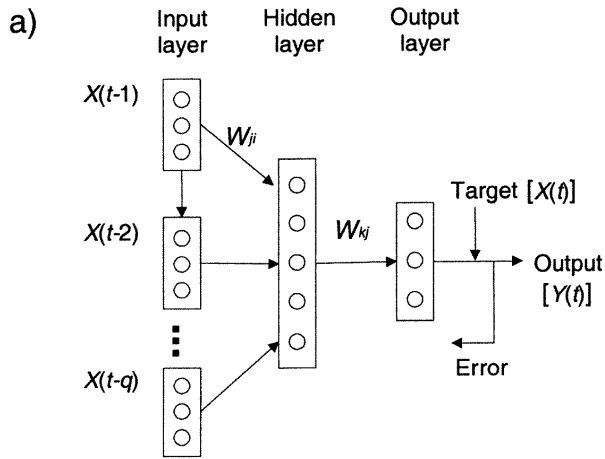


Abb. 8.12 und 8.15

1. In der Hidden Layer wurden 8-30 Knoten benutzt, die Aktivierungsfunktion logistisch, Backpropagation-Algorithmus.

Als Trainingsmenge dienten Proben aus einem 200 m langen Abschnitt des Yangjae, monatsweise für zwei Jahre, auf die fünf häufigsten Arten beschränkt. Konvergenz wurde erst nach 20000-30000 Iterationen erreicht.

Die Ergebnisse waren wenig überzeugend. Der Korrelationskoeffizient lag für die Testmenge mit einem zweimonatigen Zeitfenster bei 0,556 und mit einem viermonatigen bei 0,489. Dies ist mir unverständlich, da bei einem viermonatigen Zeitfenster mehr Informationen zur Verfügung stehen und die „unbrauchbaren“ Monate über geringe Wichtungen ja einfach ausgeblendet werden können. Schon durch bloßes Abzählen der Fälle, in denen das zweimonatige Fenster näher an die Felddaten herankommt, als das viermonatige, erahnt man das Scheitern dieses Verfahrens.

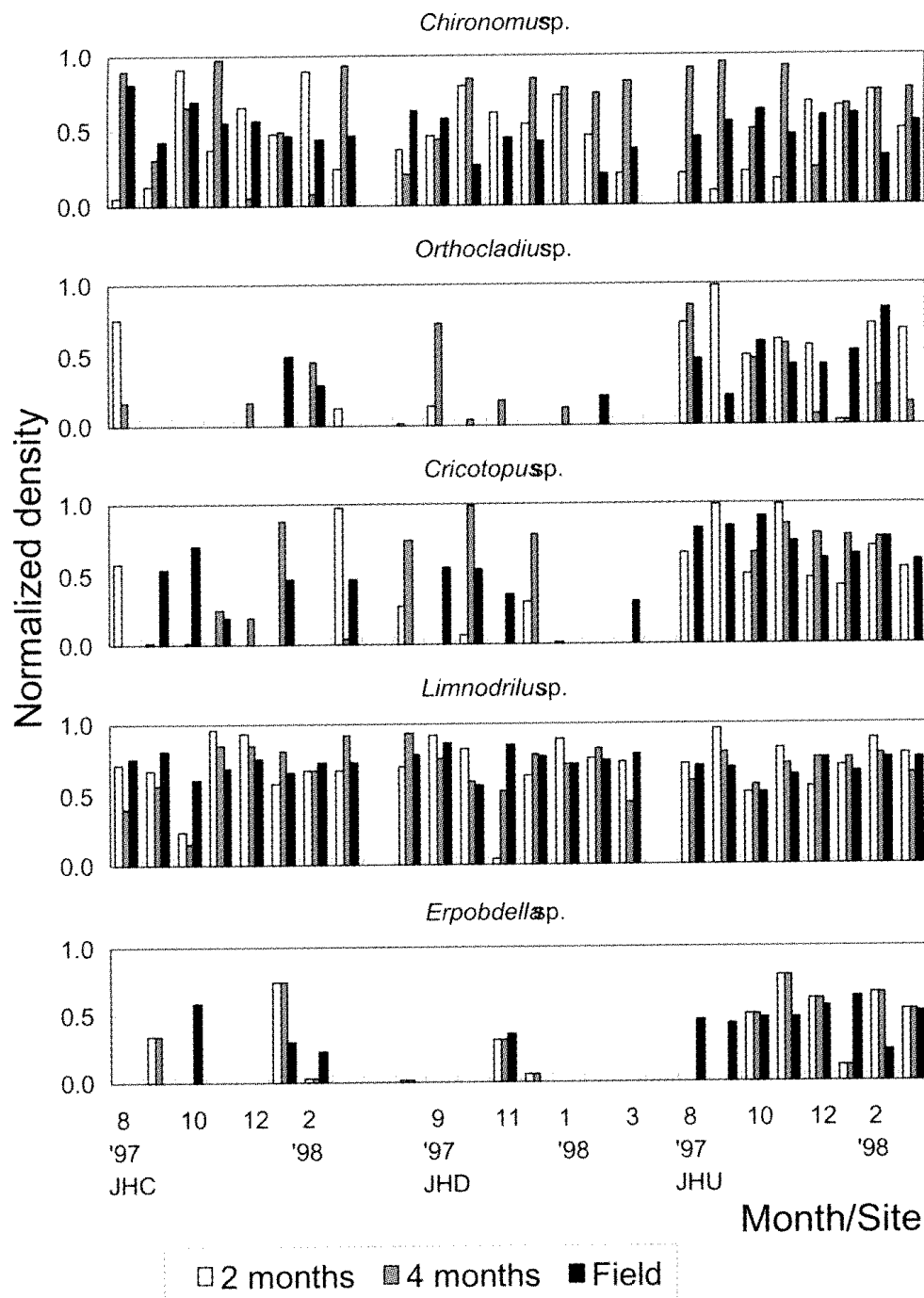


Abb. 8.13

2. Besser schnitt das rekurrente neuronale Netz nach ELMAN ab, das mit einer Context Layer zum Abspeichern des vorherigen Zustands ausgestattet war (wie die Hidden Layer mit 30 Knoten). Es kam auf einen Korrelationskoeffizienten von 0.675, besonders die Übereinstimmung bei *Erpobdella* sp. gefällt.

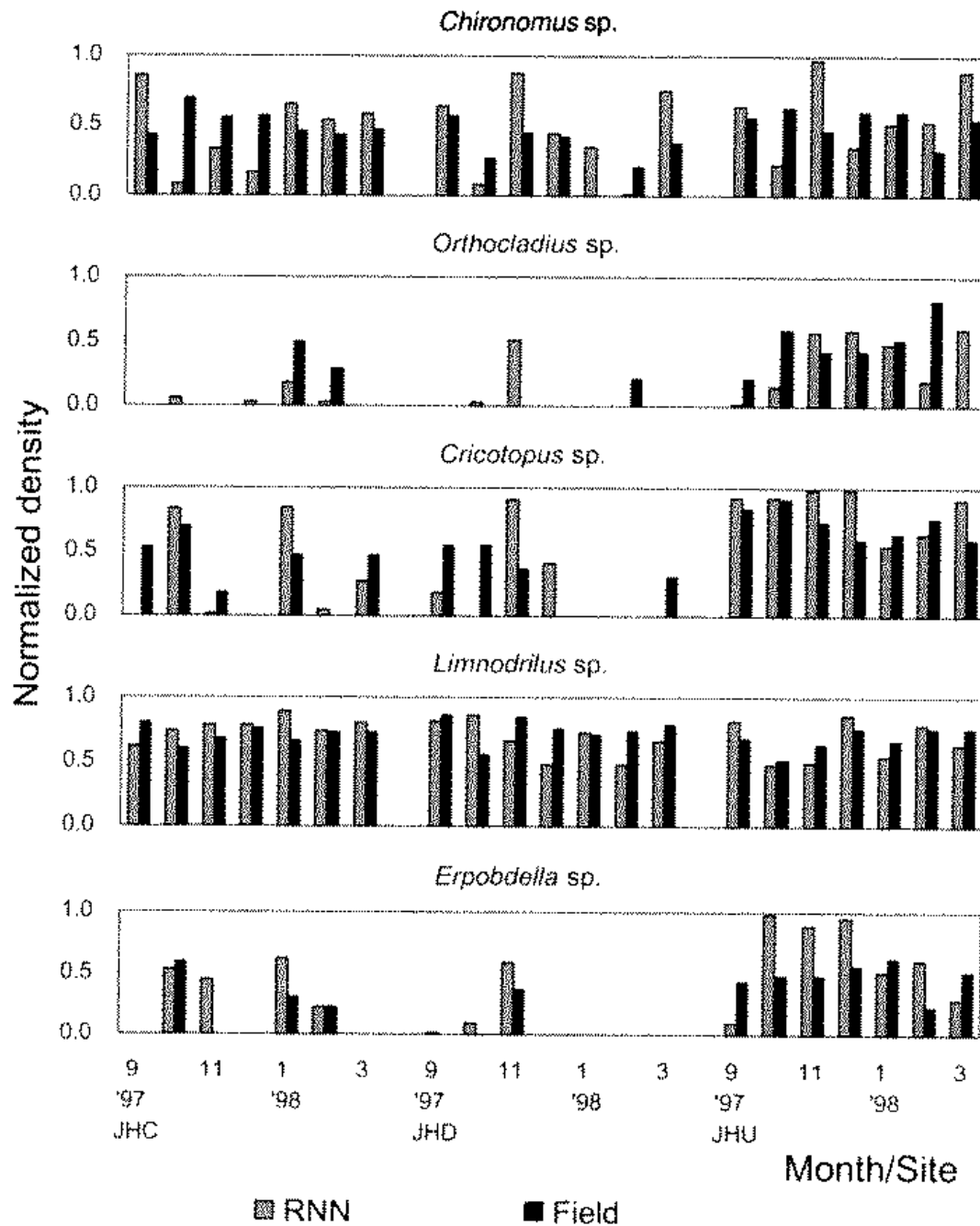


Abb. 8.14

3. Das Real-Time Recurrent Network wird auch Fully Connected Recurrent Network genannt. Diesmal 7 Arten, im Hidden Layer 13 Knoten, das Zeitfenster umfasste drei Monate. Konvergenz trat nach 5000-10000 Iterationen ein. Der Korrelationskoeffizient lag zwischen 0,55 und 0,80, die vorherige Architektur kam auf 0,3-0,4 (in der Übersicht an dritter Stelle).

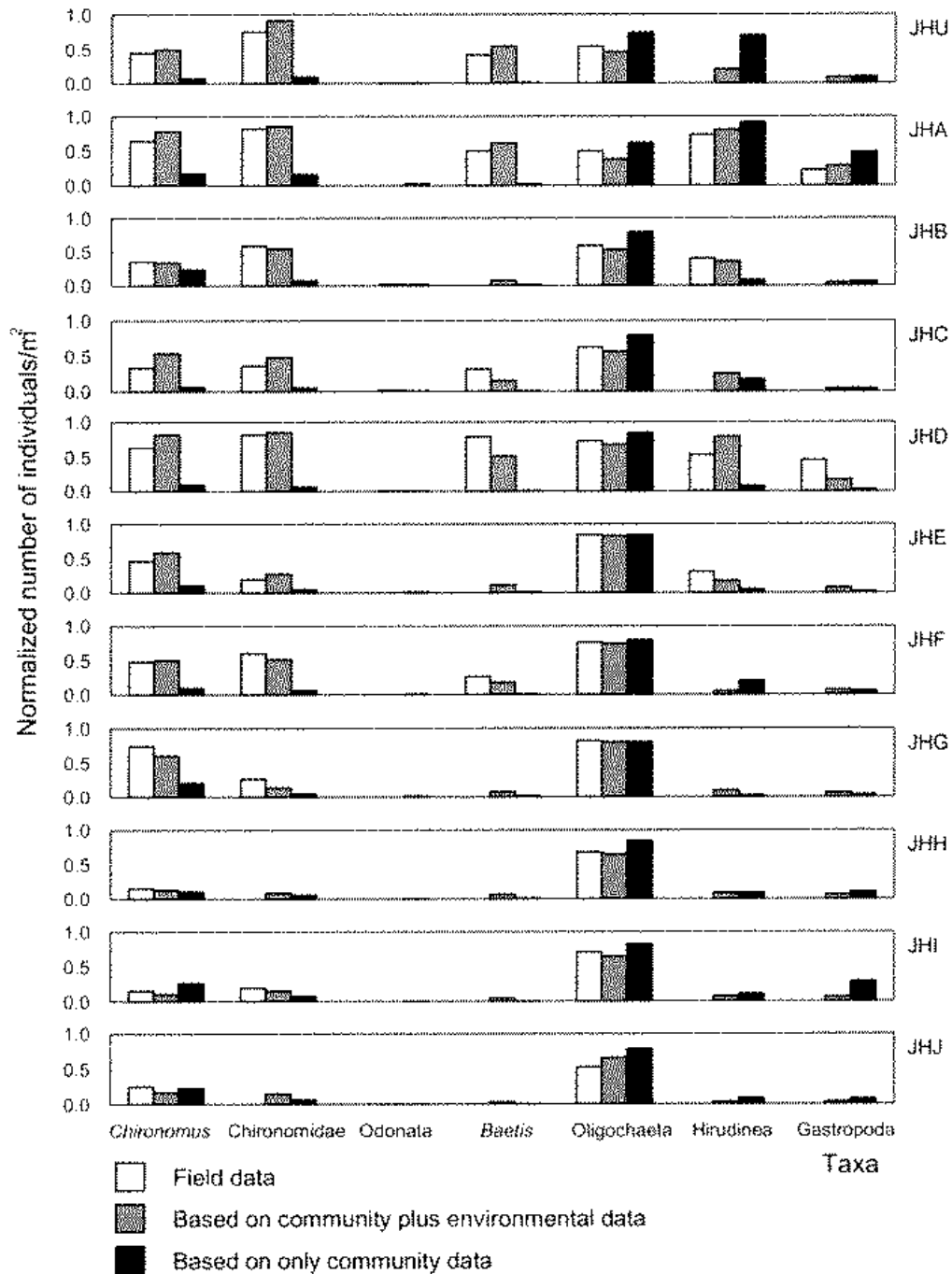


Abb. 8.16.a)

4. Der Nutzen von Umweltdaten (Fließgeschwindigkeit, Tiefe, sedimentierte organische Masse und Substratvolumen kleiner 0,5 mm; ohne Rückkopplung in der Architektur) offenbarte sich in Zeiten starker Umwelteinflüsse. So erhöht sich der Korrelationskoeffizient im Juli während eines Flutereignisses von 0,55 auf 0,94, im November aber nur von 0,59 auf 0,60. Ungewöhnlich niedrige Temperaturen im Winter verzögerten die Entwicklung der Lebensgemeinschaften – die Temperatur wurde aber nicht als Parameter mit aufgenommen, da sie im 200-m-Abschnitt keine Unterschiede aufzeigte [4,S.191].

In einer Sensitivitätsanalyse wurden diese Umweltdaten im letzten Monat um $\pm 50\%$ variiert. Wie erwartet, ergaben sich im Juli für alle Parameter größere Änderungen als im November. Einige Arten erwiesen sich aber als unempfindlicher gegenüber der Parametervariation.

eine Übersicht:

Architektur	# Knoten im Hidden Layer	Zeitfenster in Monaten	Datensatz	Iterationen in Tausend	mittlerer quadratischer Fehler	Korrelationskoeffizienten
1.	8-30	2 und 4	I	20-30	0,05	0,556; 0,489
2.	30	quasi 1	I		< 0,05	0,675
2.	30	quasi 1	II			0,3 - 0,4
3.	13	3	II	5-10	0,006	0,55 - 0,8
4.	13	3	II + Umweltdaten			0,6 - 0,94

Anwendung eines Counterpropagation Networks zur Nachbildung von Biozöosen

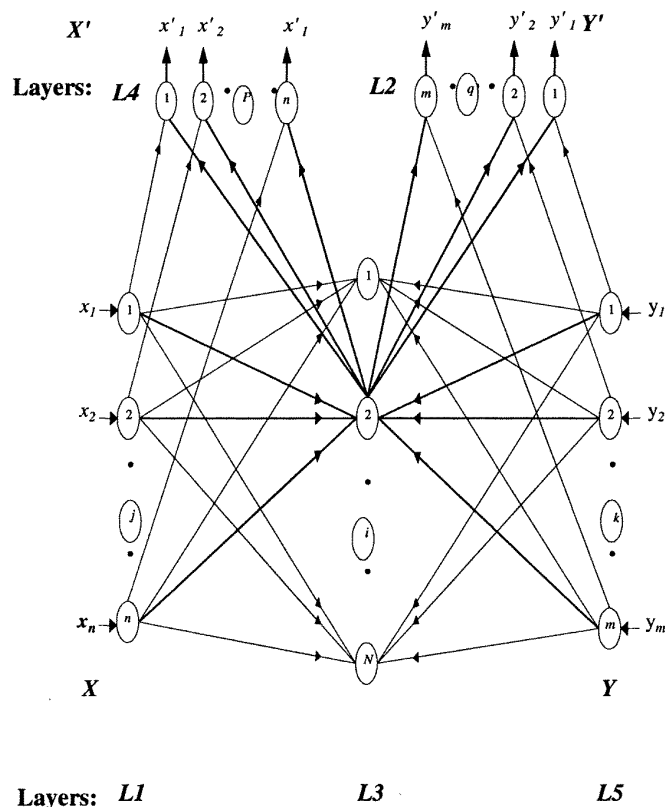


Abb. 8.18

trainiert wurde, werden umgekehrt die Wichtungen des Gewinnerknoten zu einem gewünschten Ergebnis im GROSSBERG-Layer angepasst.

Die Organisation einer Lebensgemeinschaft lässt sich taxonomisch und funktionell strukturieren. Die taxonomische Hierarchie setzte sich hier aus 7 Klassen, 19 Ordnungen, 48 Familien und 105 Arten bzw. Rassen (nicht unterschieden) zusammen. Weiter konnte man 5 funktionelle Gruppen (wie Filtrierer und Destruenten) ausmachen.

Das Counterpropagation Network sollte diese Zusammenhänge lernen. Es ist eine hybride Architektur aus einer SOM (mittlere Schicht L3) und dem sogenannten GROSSBERG Outstar (L2 und L4). Die Eingangsdaten werden gemäß ihrer hierarchischen Stufe in zwei Gruppen aufgeteilt (hier Familien und Arten/Rassen) und verschiedenen Layern (L1 und L5) zugeführt. Nachdem aus den Eingangsdaten ein Gewinnerknoten in L3 ermittelt und in einer bestimmten Nachbarschaft

Die Daten stammen aus einem Zwei-Jahreszeitraum und könnten die Daten, die bei ART verwendet wurden, mit beinhalten. Dabei nahm man aber nur die Flussabschnitte des Suyong Rivers, die sich durch ähnliche mesosaprobe Verhältnisse auszeichnen. Im Resultat fiel eine Kontrastverstärkung auf („averaging effect“): Dominante Gruppen bildeten sich noch stärker heraus, seltenere verschwanden etwas.

Anwendungen zum Exergie-Begriff

Der Exergie-Begriff wird als integratives Konzept verstanden, mit dem sich das Entwicklungsstadium einer Lebensgemeinschaft beschreiben lässt (also ob etwa im Reifestadium, gestört oder auf dem Wege der Erholung). Exergie wird definiert als die Menge an Arbeit, die ein System zu leisten vermag, bis es sich im thermodynamischen Gleichgewicht mit seiner Umwelt befindet. In einer Lebensgemeinschaft werden die einzelnen Arten aufsummiert (eigentlich eine isolierte Betrachtung - ohne „Synergie“):

$$\text{Exergie} = \sum (C_i * W_i)$$

mit C_i der Konzentration der i-ten Art (Biomasse) und W_i dem dazugehörigen „Informationsgehalt“ (dessen Einheit J/Liter erweckt bei mir den Eindruck, dass die Tiere einer Feuerbestattung zugeführt werden).

Die Werte für den Informationsgehalt mussten abgeschätzt werden. Detritus erhielt 1, sonst lagen die Arten zwischen 29,6 und 43,9; hier wurde den zehn Arten „willkürlich“ 30 oder 35 zugewiesen. Wie eine Messung aussehen könnte, wurde mir nicht klar.

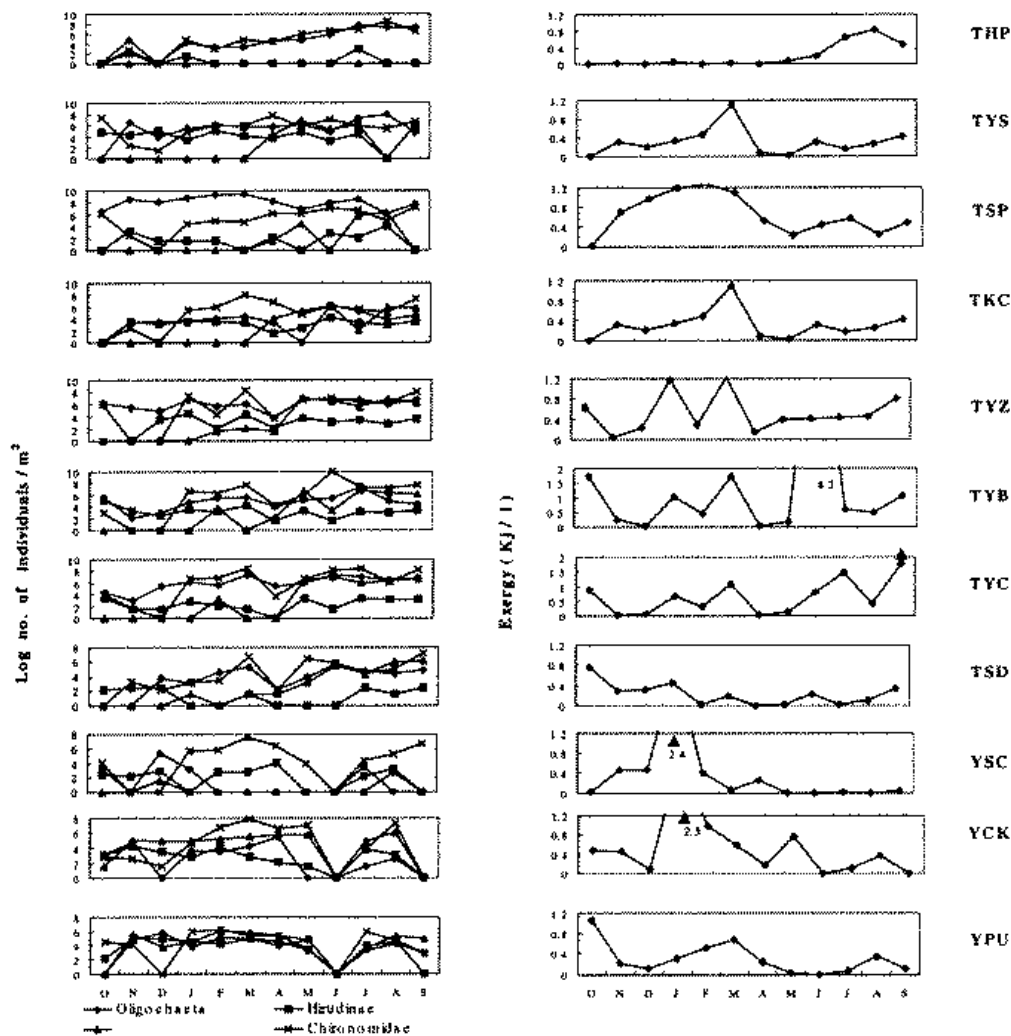


Abb. 8.22

Obwohl die Flussabschnitte nah beieinander lagen (genaue Angaben in [5,S.146-148]), wichen sie im Verlauf der Exergie stark voneinander ab. Dies ist auf diverse Einleitungen zurückzuführen, wie Schlamm aus einer Baustelle, Gülle aus Schweine- und Geflügelfarmen, Ackerflächen, Handwerksbetriebe und häusliche Abwässer [5,S.146]. Auch innerhalb eines Abschnitts war die Dynamik beträchtlich. So stieg einmal im Laufe zweier Monate die Exergie von nahe Null auf über 8 kJ/l. Hier hätte eine ungleichmäßige Verteilung der Organismen vorgelegen und die Beprobung erschwert [5,S.148]. Ein Zusammenhang zwischen 8.22 a) und b) ist nicht immer offensichtlich, was daran liegen könnte, dass nicht alle Arten in 8.22 a) aufgeführt wurden [5,S.148]. So hebt sich die Exergiekurve kaum von der Nulllinie, obwohl mittlere Dichten angezeigt werden (am deutlichsten in YSC). Im flachen Wasser (12-18 cm) von TCD, TYB und TYC trat im Winter Eisbildung auf [5,S.150], ferner heftige Regenfälle im April (250 mm durchschnittlich, [5,S.152]).

Es wurde noch einmal das Multilayer Perceptron herangezogen, mit fünf Knoten in der Hidden Layer. Die Artenzusammensetzung traf im Netz auf die Exergie mit bzw. ohne einen Monat Verzögerung, wobei letztere Variante im Training besser abschnitt (der Korrelationskoeffizient wird aber mit 0,97 gegenüber 0,98 angegeben [5,S.153]). (Ist eine Berechnung ohne Verzögerung nicht trivial? Man statte die Kanten mit 30 oder 35 aus. Ich habe den Verdacht, dass die Felddaten in obiger Abbildung deswegen nicht erkennbar sind, weil sie von den Daten ohne Verzögerung einfach überdeckt werden.)

Skeptisch bin ich bei der impliziten Annahme, dass man einer Artenzusammensetzung ansehen kann, welchen zukünftigen Umwelteinflüssen sie ausgesetzt sein wird (siehe auch die vorherigen Vorhersagen ohne Beachtung der Umweltdaten).

Bei der Vorhersage von Daten stellte sich nur ein schlechter Zusammenhang ein (Korrelationskoeffizient zwischen 0,45 und 0,65).

Zuletzt wurden Änderungen der Exergie in einem Vektor aus z.B. 3 oder 4 Monaten einer SOM (hier ohne ART) übergeben. Bei längerem Zeitfenster standen entsprechend weniger Vektoren zur Verfügung, so dass auch die Größe der SOM angepasst wurde. Gleichermäßen die räumliche wie die zeitliche Komponente führten eine Ordnung herbei.

Noch einige Anmerkungen zum Exergie-Begriff: Einen leichten Einstieg bietet [9]. Hier wird Energie in wertvolle Exergie und nutzlose Anergie unterteilt. Der Wirkungsgrad eines Kraftwerks beeinflusst dieses Verhältnis. Quelle [10] ist für den (ambitionierten?) Schulunterricht gedacht und grenzt den Begriff von dem der Entropie ab (Maß für die Unordnung, entspricht der Wahrscheinlichkeit für eine Realisierung; Einheit in der Informationstheorie bit, in der Physik J/K; ferner negative Entropie: Negentropie). Im biologischen Kontext ist es vielleicht unüblich, vom thermodynamischen Gleichgewicht zu reden. JØRGENSEN schreibt deshalb auch [5,S.144]: „The environment or reference state could be represented as the inorganic soup of the system without life.“ WALL wirbt schließlich für einen Dreiklang aus Exergie, Ökologie und Demokratie [11].

Formale Mängel und inhaltliche Fehler

(Die große Häufung lässt es nicht zu, diesen Punkt auszusparen. Aus diesen Gründen habe ich auch auf die Abbildung der ein oder anderen Abbildung verzichtet.)

Die Nummerierung der Kapitel weist Lücken auf (nach 8.1.1 folgt 8.2.2 – wo war 8.2.1?).

Evtl. gab es Formatierungsprobleme bei Sonderzeichen – an einigen Stellen sieht man im Text bloß „●“ (S. 139, 149, 154, 168, 171).

S. 164: X' und Y' wurden anscheinend vertauscht, andererseits habe ich hier Verständnisprobleme.

S. 165: ein X^{ccc} in Abb. 8.20 gibt es nicht.

S. 171: ein falscher Verweis auf Abb. 8.3 (statt 8.2)

überflüssige Wörter:

- S. 133: „Groupings in other cases [...] than groupings by the clustering analysis.“
- S. 158: “The previously mentioned [...] fully connected recurrent network, ...”
- S. 173: “...networks were ~~networks~~ useful...”

Rechtschreibfehler, die wiederum sinnvolle Wörter ergeben und darum irritieren:

- S. 168: „Multilayer Perceptron with Time-Delay“
- S. 170: „...were explainable in many cases.“
- S. 172: „...alpha codes stand for the name“
- S. 173 “... groupings were efficiently conducted...”

In der Beschreibung zu Abb. 8.7 wurden a) und b) vertauscht. Das Semikolon anstatt eines Doppelpunktes stört. Diese Abb. gehört besser auf S. 143, da sonst das Zeitfenster (Time: m-2, m-1, m) zusammenhangslos ist.

Einige SOMs sind nicht quadratisch. Womöglich gingen unbesetzte Randspalten verloren. In diesem Fall hätte man das Output-Layer von Anfang an kleiner machen können, um den Wettbewerb zwischen den Items bei der Besetzung der Neuronen zu fördern.

Für die Vergleichbarkeit der Netzarchitekturen wäre eine einheitliche Herangehensweise praktisch gewesen. Die Abb. 8.16 weicht so in der Darstellungsweise von den Abb. 8.13 u. 8.14 ab.

Einer Aussage auf S. 159 muss widersprochen werden: „..., while the network trained only with the community data consistently underestimated the densities of Chironomidae and Chironomus“. So wurde für Chironomus am Ort JHI eine höhere Dichte als tatsächlich vorhanden geschätzt.

Abb. 8.17 ist zu klein ausgefallen, um die Punkte für einzelne Arten unterscheiden zu können. In Abb. 8.21 fehlt womöglich c); und d) wurde fälschlicherweise mit „c)“ deklariert („c)“ bezieht sich im Gegensatz zur Legende auf Familien). b) und „c)“ hätte man gleich stauchen können. Die Musterbildung finde ich zu suggestiv, da die Aneinanderreihung der Arten/Rassen bzw. Familien willkürlich ist. In vertikaler Richtung besteht nur innerhalb eines Flussabschnittes ein zeitlicher Zusammenhang.

In Abb. 8.22 fehlt die Bezeichnung einer Art, für gewöhnlich schreibt man kJ statt Kj.

In Abb. 8.24 hätte man die Ordinate bei 0,4 begrenzen können, im Hinblick auf Abb. 8.23 vielleicht sogar eine logarithmische Darstellung (hier fehlten auch die Felddaten).

In der Legende zu Abb. 8.25 wird nur auf eine der beiden SOMs eingegangen. Der Textteil liefert andere Größenangaben der SOMs (8*8 bzw. 7*7 statt 8*9 bzw. 7*9).

In den Papern [1] und [3] wird Dank für's Proof-Reading ausgesprochen. Für diesen Artikel ist dies wohl unterblieben.

Redundanz ließe sich vermeiden, wenn man Gemeinsamkeiten voranstellen würde (so liest man häufig „...as previously mentioned...“). Die Sätze „..., while data collected from April 1997 to March 1998 were used as new data for testing the trained network.“ und “In order to

verify the predictability of the trained network, we further provided new community data from April 1997 to March 1998.” sind keine vier Zeilen voneinander entfernt (S.158). Die zehnte SOM hat nicht wirklich einen Neuigkeitswert, da nicht das Resultat an sich, sondern die Verfahrensweise interessiert. Bei soviel Selbstähnlichkeit und verkleinerten Abbildungen könnte man ja mal die fraktale Dimension berechnen...

$\log(\text{Anzahl der Kopien}) / \log(\text{Skalierungsfaktor}) \dots$

Einen roten (oder andersfarbigen) Faden konnte ich nicht erkennen, der Artikel schien mehr Stückwerk zu sein, aus mehreren Papern zusammengetragen. Die Wahl der Netzarchitektur wirkte manchmal beliebig, oft verblieb die Diskussion der Methodik bei einer oberflächlichen Beschreibung der Ergebnisse.

Literaturverzeichnis

- Tae-Soo CHON, Young-Seuk PARK, Inn-Sil KWAK und Eui Young CHA (2003): Non-linear Approach to Grouping, Dynamics and Organizational Informatics of Benthic Macroinvertebrate Communities in Streams by Artificial Neural Networks. In: RECKNAGEL, Friedrich: Ecological Informatics. Heidelberg. S.127-178
- [1] Tae-Soo CHON, Young-Seuk PARK, Kyong Hi MOON und Eui Young CHA (1996): Patternizing communities by using an artificial neural network. Ecol. Model. S.69-78
- [2] Tae-Soo CHON, Young-Seuk PARK und Eui Young CHA (2000): Patterning of Community Changes in Benthic Macroinvertebrates Collected from Urbanized Streams for the Short Time Prediction by Temporal Artificial Neuronal Networks. In: LEK, SOVAN und GUÉGAN, Jean-François (Eds.): Artificial Neuronal Networks – Application to Ecology and Evolution. Berlin. S.99-114
- [3] Tae-Soo CHON, Young-Seuk PARK und June Ho PARK (2000): Determining temporal pattern of community dynamics by using unsupervised learning algorithms. Ecol. Model. S.151-166
- [4] Tae-Soo CHON, Inn-Sil KWAK, Young-Seuk PARK, Tae-Hyung KIM und YooShin KIM (2001): Patterning and short-term predictions of benthic macroinvertebrate community dynamics by using a recurrent artificial neural network. Ecol. Model. S.181-193
- [5] Young-Seuk PARK, Inn-Sil KWAK, Tae-Soo CHON, Jwa-Kwan KIM und Sven Erik JØRGENSEN (2001): Implementation of artificial neural networks in patterning and predictio of exergy in response to temporal dynamics of benthic macroinvertebrate communities in streams. Ecol. Model. S.143-157
- [6] Werner H. BAUR (1998): Gewässergüte beurteilen und bestimmen. Berlin
- [7] Euan George NISBET (1994): Globale Umweltveränderungen: Ursachen, Folgen, Handlungsmöglichkeiten, Klima, Energie, Politik. Heidelberg
- [8] Barbara HAMMER (2000): Skript zur Vorlesung Neuronale Netze. Osnabrück
- [9] Udo LEUSCHNER: Energie = Exergie + Anergie
(<http://www.udo-leuschner.de/energie/e07exergie.htm> am 31.04.04)
- [10] Udo BACKHAUS, H. Joachim SCHLICHTING (1984): Entropie und Exergie – Zwei Größen zur Beschreibung von Irreversibilität und Energieentwertung. In: Der Physikunterricht 18/3 S.41-57
- [11] Göran WALL: Exergy, Ecology and Democracy – Concepts of a Vital Society or A Proposal for An Exergy Tax (<http://exergy.se/> am 31.04.04)

Anhang: „Schrebergärten“

Man stelle sich eine Kleingartenanlage mit 3*3 Parzellen vor, auf die 6 Gartennutzer losgelassen werden. Jeder Gärtner hat andere Präferenzen hinsichtlich Bepflanzung und Wegbeschaffenheit. Diese beiden Parameter werden wie folgt in einer Skala codiert (sie muss nicht unbedingt äquidistant sein, die Anordnung erfolgt hier intuitiv):

Bepflanzung B		Wegbeschaffenheit W	
Ackerbau	0	Asphalt	0
Obstbäume mit Wiese	1	Rasengittersteine	1
barocke Gartenanlage	2	Kies	2
		Rasen	3
		kein Weg	4

Bei den Gärtnern sind diese Parameter folgendermaßen ausgeprägt:

a) praktischer Landwirt	0	0
b) extensiver Nutzer	0	3
c) Apfelsaftgenießer	1	1
d) ordentlicher Autofahrer	2	0
e) Naturtyp	2	4
f) Versaille-Spaziergänger	2	2

Die Lernrate η wird auf 0,25 festgesetzt, für Z_{ij} die von-Neumann-Nachbarschaft genommen (auf Zäune wird weitgehend verzichtet). Die zufällige Initialisierung der self-organizing Kleingartenanlage sieht so aus:

Jede schwarzumrandete Parzelle enthält die aktuellen Werte für **Bepflanzung** und **Wegbeschaffenheit**.

2	0	1	3	1	1
0	1	2	2	2	4
1	2	0	4	1	3

Nun darf sich der erste Gärtner auf die Kleingartenanlage stürzen, es ist zufällig der praktische Landwirt a). Er untersucht jede Parzelle, ob sie seinem Wunschenken entspricht. Für die linke obere Zelle mit den Koordinaten [1,1] rechnet er zum Beispiel

$$d_{[1,1]}(t) = (B_a - B_{[1,1]})^2 + (W_a - W_{[1,1]})^2 = (0 - 2)^2 + (0 - 0)^2 = 4$$

und trägt diese Werte der Übersicht halber in eine Matrix ein:

4	10	2
1	8	20
5	16	10

Die Zelle [2,1] weist den niedrigsten Wert auf und gefällt ihm darum am besten. Er müsste hier nur noch den Weg asphaltieren und fängt damit schon mal an:

$$W_{[2,1]}(t+1) = W_{[2,1]}(t) + \eta * (W_a - W_{[2,1]}(t)) = 1 + 0,25 * (0 - 1) = 0,75$$

Genauso verfährt er in seiner Nachbarschaft (kursiv dargestellt):

1,5	0	1	3	1	1
0	0,75	1,5	1,5	2	4
0,75	1,5	0	4	1	3

Als nächstes ist der Versaille-Spaziergänger f) an der Reihe und begutachtet die schon veränderte Kleingartenanlage:

4,25	2	2
5,56	0,5	4
1,81	8	2

Die zentrale Zelle sagt ihm am meisten zu.

Und auch sein Tun erstreckt sich über die Nachbargärten.

1,5	0	1,25	2,75	1	1
0,5	1,06	1,63	1,63	2	3,5
0,75	1,5	0,5	3,5	1	3

1,25	3,13	0
0,25	0,78	7,25
0,31	6,5	4

Der Apfelsaftgenießer hat sein Paradies schon gefunden.

Um seinen Herrschaftsanspruch zu festigen, werden die Gärten [1,2] und [2,3] angepasst.

1,5	0	1,19	2,31	1	1
0,5	1,06	1,63	1,63	1,75	2,88
0,75	1,5	0,5	3,5	1	3

16,25	3,51	10
10,88	5,78	1,33
7,81	2,5	2

Nun tritt der Naturtyp erstmals auf den Plan.

1,50	0,00	1,19	2,31	1,25	1,75
0,50	1,06	1,72	2,22	1,81	3,16
0,75	1,50	0,50	3,50	1,25	3,25

11,25	1,88	3,13
4,00	3,56	3,31
2,81	0,50	1,63

Der extensive Nutzer ist von der Zelle [3,2] angetan.

1,50	0,00	1,19	2,31	1,25	1,75
0,50	1,06	1,29	2,41	1,81	3,16
0,56	1,88	0,38	3,38	0,94	3,19

0,25	6,01	3,63
3,38	6,33	10,00
5,58	14,03	11,29

Der ordentliche Autofahrer...

1,63	0,00	1,39	1,73	1,25	1,75
0,88	0,80	1,29	2,41	1,81	3,16
0,56	1,88	0,38	3,38	0,94	3,19

0,14	3,38	3,63
1,90	6,33	10,00
5,58	14,03	11,29

...darf gleich ein zweites Mal...

...und bleibt natürlich bei seiner eben bearbeiteten Zelle.

1,72	0,00	1,54	1,30	1,25	1,75
1,16	0,60	1,29	2,41	1,81	3,16
0,56	1,88	0,38	3,38	0,94	3,19

2,95	4,07	4,63
1,69	7,49	13,25
3,83	11,53	11,04

Obwohl sich in der Zwischenzeit schon viel getan hat, nimmt der Landwirt keinen Standortwechsel vor.

1,29	0,00	1,54	1,30	1,25	1,75
0,87	0,45	0,97	1,81	1,81	3,16
0,42	1,41	0,38	3,38	0,94	3,19

1,08	0,39	0,63
0,32	0,66	5,31
0,50	6,03	4,79

Der Apfelsaftgenießer kommt anscheinend mit dem Landwirt ganz gut aus, er war eben noch in einer ganz anderen Ecke.

1,22	0,25	1,54	1,30	1,25	1,75
0,90	0,59	0,98	1,61	1,81	3,16
0,57	1,30	0,38	3,38	0,94	3,19

14,68	7,49	5,63
12,86	6,77	0,75
9,32	3,03	1,79

der Naturtyp

1,22	0,25	1,54	1,30	1,44	2,31
0,90	0,59	1,23	2,21	1,86	3,37
0,57	1,30	0,38	3,38	1,20	3,39

1,54	4,07	7,41
1,15	6,38	14,80
2,02	11,53	12,94

der Landwirt

0,91	0,19	1,54	1,30	1,44	2,31
0,68	0,44	0,92	1,65	1,86	3,37
0,42	0,98	0,38	3,38	1,20	3,39

4,47	0,70	0,41
4,19	1,28	1,89
3,52	4,53	2,57

der Spaziergänger...

0,91	0,19	1,66	1,48	1,58	2,23
0,68	0,44	0,92	1,65	1,89	3,03
0,42	0,98	0,38	3,38	1,20	3,39

4,47	0,39	0,23
4,19	1,28	1,06
3,52	4,53	2,57

...verweilt noch etwas länger.

0,91	0,19	1,74	1,61	1,68	2,18
0,68	0,44	0,92	1,65	1,92	2,77
0,42	0,98	0,38	3,38	1,20	3,39

1,22	2,65	4,83
1,95	3,90	7,67
3,44	14,03	12,13

Seit seinem letzten Besuch hat sich die Lage für den Autofahrer ziemlich verschlechtert.

1,18	0,14	1,81	1,21	1,68	2,18
1,01	0,33	0,92	1,65	1,92	2,77
0,42	0,98	0,38	3,38	1,20	3,39

9,58	6,49	3,51
8,14	2,66	3,74
4,27	0,28	1,60

der Nutzer

1,18	0,14	1,81	1,21	1,68	2,18
1,01	0,33	0,69	1,99	1,92	2,77
0,32	1,48	0,28	3,28	0,90	3,29

4,12	0,67	0,13
3,78	1,71	0,60
3,09	4,60	2,88

der Spaziergänger

1,18	0,14	1,86	1,40	1,76	2,13
1,01	0,33	0,69	1,99	1,94	2,58
0,32	1,48	0,28	3,28	0,90	3,29

1,42	5,41	7,65
1,12	4,44	10,41
2,30	10,85	11,66

der Landwirt

0,89	0,11	1,86	1,40	1,76	2,13
0,75	0,25	0,52	1,49	1,94	2,58
0,24	1,11	0,28	3,28	0,90	3,29

0,81	0,89	1,86
0,63	0,47	3,37
0,59	5,72	5,27

Der Apfelsaftgenießer überlässt die linke Zelle doch dem Landwirt. Mit seinen mittleren Werten passt er auch gut in die Mitte.

Hier ist aber die Gefahr größer, dass seine Zelle von den anderen verändert wird. Vielleicht laufen diese Veränderungen aber auf plus-minus-Null hinaus.

0,89	0,11	1,64	1,30	1,76	2,13
0,82	0,44	0,64	1,37	1,71	2,18
0,24	1,11	0,46	2,71	0,90	3,29

0,81	0,50	1,86
0,35	0,27	1,90
0,59	3,22	5,27

noch mal fällt das Los auf ihn

0,89	0,11	1,48	1,23	1,76	2,13
0,86	0,58	0,73	1,28	1,53	1,89
0,24	1,11	0,60	2,28	0,90	3,29

9,17	5,34	3,86
6,62	3,50	3,58
3,62	0,87	0,90

der Nutzer...

0,89	0,11	1,48	1,23	1,76	2,13
0,86	0,58	0,55	1,71	1,53	1,89
0,18	1,58	0,45	2,46	0,68	3,22

9,17	5,34	3,86
6,62	1,97	3,58
2,04	0,49	0,51

...auch gleich zweimal

0,89	0,11	1,48	1,23	1,76	2,13
0,86	0,58	0,41	2,03	1,53	1,89
0,13	1,94	0,34	2,60	0,51	3,16

1,25	1,78	4,60
1,63	6,65	3,78
7,24	9,52	12,24

der Autofahrer

1,17	0,08	1,61	0,92	1,76	2,13
1,15	0,43	0,41	2,03	1,53	1,89
0,13	1,94	0,34	2,60	0,51	3,16

1,37	3,44	7,65
1,50	4,29	5,90
3,78	6,86	10,27

Überraschung: Landwirt und Autofahrer machen gemeinsame Sache.

0,87	0,06	1,21	0,69	1,76	2,13
0,86	0,32	0,41	2,03	1,53	1,89
0,13	1,94	0,34	2,60	0,51	3,16

Der untere Grenzwert der Wegbeschaffenheits-Skala wird fast wieder erreicht.

5,03	2,34	0,07
4,11	2,53	0,23
3,48	3,13	3,58

der Spaziergänger

0,87	0,06	1,41	1,02	1,82	2,10
0,86	0,32	0,41	2,03	1,65	1,92
0,13	1,94	0,34	2,60	0,51	3,16

16,80	9,25	3,65
14,81	6,40	4,47
7,73	4,74	2,92

Da der Naturtyp von der Gartengestaltung lange Zeit ausgeschlossen war, passt ihm die Landschaft nicht so recht (über 2,9 Distanz).

Als Außenseiter verzieht er sich in die Ecke, die der des Landwirtes gegenüberliegt. Ein Konflikt mit dem extensiven Nutzer nebenan ist absehbar, da unterschiedliche Bepflanzungen angestrebt werden.

...

...

So kann ein einfacher Algorithmus eine ganze Geschichte erzählen. Und wenn die Kleingartenanlage nicht vergrößert wird, dann graben sie noch heute (bei 4*3 Parzellen kämen sie sich nicht mehr ins Gehege). Typische Laufzeiten gehen in die Hunderte, oft Tausende.