

REDE DE KOHONEN APLICADA À CLASSIFICAÇÃO TRÓFICA DE PEIXES

CLÁUDIA A. CAVALHEIRO FRANCISCO¹ E MARIA TERESINHA ARNS STEINER²

¹Universidade Federal do Rio Grande do Norte – Departamento de Engenharia de Produção - Campus Universitário Lagoa Nova – CP 1524 - Natal – RN – CEP: 59072-970 – e-mail: claudia@ct.ufrn.br. ²Universidade Federal do Paraná – Programa de Pós-Graduação em Métodos Numéricos em Engenharia - CP:19081 – Curitiba – PR – CEP: 81531-990

Resumo

As redes neurais artificiais têm sido aplicadas com sucesso em variados problemas de análise exploratória de dados multivariados relacionados à ecologia. Todavia, poucos são os trabalhos utilizando a rede de Kohonen (mapas auto-organizáveis) para classificação de padrões de dados ecológicos. A maior dificuldade encontra-se na obtenção de um conjunto adequado de dados para o treinamento e validação da rede. Este trabalho utiliza a rede de Kohonen (SOM)[6] como uma metodologia para a classificação das espécies de peixes em categorias tróficas do reservatório de Corumbá, GO, de acordo com a dieta dos mesmos, envolvendo duas fases distintas: fase rio (anterior ao represamento), com a classificação de 1845 estômagos de 33 espécies de peixes, e a fase reservatório (posterior ao represamento), com a classificação de 5492 estômagos de 36 espécies de peixes. As classificações foram obtidas nas duas fases distintamente, resultando em ambas, em uma separação de dois grandes grupos: o dos generalistas e o dos especialistas. Os peixes especialistas foram divididos em quatro grupos: insetívoros, herbívoros, piscívoros e detritívoros. Dentre estas, os insetívoros apresentaram um grande número de espécies, sendo necessário modelar uma nova rede, visando a separação desta categoria, em quatro sub-grupos. A rede de Kohonen mostrou-se uma ferramenta robusta para a classificação dos dados, apresentando resultados rápidos, com uma clara visualização dos agrupamentos, facilitando sobremaneira a interpretação dos resultados.

Palavras-chave: Redes Neurais Artificiais; Redes de Kohonen; peixes e categorias tróficas.

I. Introdução

A motivação deste trabalho foi encontrar uma ferramenta para a identificação da estrutura trófica de peixes no reservatório de Corumbá, Estado de Goiás (bacia do rio Paranaíba), que fornecesse informações confiáveis, de forma que os ecólogos pudessem considerar nos estudos da ecologia trófica. O conhecimento acerca das assembléias de peixes fornece embasamento sobre condições dos ecossistemas aquáticos, ou seja, são informações bioindicadores fornecendo assim, por exemplo, informações sobre a qualidade da água [1]. Desta forma é de fundamental importância a utilização de ferramentas matemáticas eficientes para avaliação de dados ecológicos, sendo que os métodos mais comumente

utilizados nos estudos ecológicos são: regressão linear, regressão linear múltipla, correspondência canônica, análise de componentes principais, entre outras [5]. Especificamente a classificação trófica utiliza-se de métodos subjetivos com o auxílio da estatística multivariada, que no caso, indica o número de agrupamentos, mas não informa claramente quais espécies pertencem a cada grupo. Estes métodos conduzem freqüentemente a uma perda de informações e realidade ecológica, pois não consideram a complexidade e a não-linearidade do conjunto de dados. Uma ferramenta que vem sendo amplamente utilizada nos estudos ecológicos são as redes neurais artificiais que têm se mostrado eficientes para trabalhar com conjuntos de dados com estas características [9,12]. São muitos os tipos ou modelos de redes neurais, sendo que a principal diferença entre os mesmos está na forma de treinamento, que pode ser supervisionado ou não supervisionado. As redes neurais artificiais mais utilizadas em ecologia são as de múltiplas camadas, com algoritmo de *backpropagation* [3,10,8], normalmente utilizadas para predição e as não supervisionadas, como os mapas auto-organizáveis (SOM – *Self Organizing Maps*) [2,5,11] para ordenação e classificação.

Neste contexto, tornou-se importante a implementação de um modelo matemático capaz de agrupar e identificar as espécies pertencentes a cada categoria trófica. Optou-se por um modelo de redes neurais competitivo, conhecido como rede de Kohonen ou mapa auto-organizável (SOM) para realizar tal tarefa [6].

II. Material e Métodos

Para o estudo, foi analisado o espectro alimentar das espécies de peixes presentes em amostras obtidas em rios, ribeirões e riachos, influenciados em graus variados pelo represamento da Usina Hidrelétrica de Corumbá (UHE), GO. Foram analisados de forma distinta dois períodos: fase rio (março de 1996 a agosto de 1996) e fase reservatório (março de 1997 a fevereiro de 1999). As amostras foram separadas, analisadas e organizadas por pesquisadores do projeto “Estudos ictiológicos na área de influência do Aproveitamento Hidrelétrico (AHE) Corumbá”, do Núcleo de Pesquisas em Limnologia, Ictiologia e Aqüicultura (Nupélia), da Universidade Estadual de Maringá (UEM). Para análise da dieta, foram utilizados os volumes dos itens alimentares encontrados nos estômagos dos peixes de acordo com o método volumétrico, onde se registra o volume de cada item ou

recurso alimentar, obtendo-se a porcentagem em relação ao volume total de todos os conteúdos estomacais [7]. Os recursos alimentares explorados pelos peixes foram separados em onze classes: microcrustáceos, insetos aquáticos, insetos terrestres, outros invertebrados aquáticos, outros invertebrados terrestres, peixes, outros vertebrados, algas, vegetais aquáticos, vegetais terrestres e detrito/sedimento [4]. A caracterização da dieta foi realizada para conteúdos estomacais de 7337 peixes pertencentes a 52 espécies no total, pois espécies que apresentaram um número de estômagos analisados com alimento passível de identificação menor que dez foram descartadas.

Os dados foram normalizados, utilizando a razão do volume de cada recurso alimentar pela soma dos volumes de todos os recursos alimentares da amostra, sendo na seqüência separados em dois arquivos: treinamento com 70% dos dados e validação com 30%.

O SOM [6] consiste de duas camadas de neurônios conectadas por pesos; a camada de contribuição (entrada) é conectada com um vetor do padrão de entrada do conjunto de dados e a camada de produção (saída) consiste em um mapa bidimensional disposto numa grade retangular. No caso aqui estudado, a rede apresenta onze neurônios na camada de entrada, onde cada um representa a quantidade (volume-ml) de um recurso alimentar, encontrado no estômago do peixe, e a rede realizará os agrupamentos de acordo com estes recursos (Figura 1).

Durante o processo de aprendizagem os pesos são modificados, visando minimizar a distância euclidiana entre os vetores peso (neurônios no mapa de saída) e de contribuição (padrão de entrada). Isto resulta na classificação dos dados de contribuição de acordo com suas semelhanças.

Para a inicialização do treinamento, valores aleatórios devem ser designados para os pesos iniciais, sendo que estes pesos, preferencialmente, devem estar no mesmo intervalo de valores das entradas. A taxa de aprendizagem deve decrescer no decorrer do processo, assim como o raio de vizinhança que deve estar de acordo com o tamanho do mapa desejado; neste trabalho utilizou-se um SOM de dimensão 8X6. O treinamento, utilizando o algoritmo de Kohonen, é realizado em duas fases: a 1ª fase é relativa à ordenação (ou auto-organização), onde a taxa de aprendizagem e o raio topológico apresentam valores maiores no início do processo e aos poucos vão diminuindo (por exemplo, a taxa de aprendizagem pode iniciar em 0,8 para finalizar em 0,5 e o raio topológico pode iniciar de tal forma a abranger todos os neurônios do mapa e finalizar na vizinhança mais próxima do neurônio vencedor); já a 2ª fase, relativa a convergência, é necessária para o “ajuste fino” do mapa (a taxa de aprendizagem deve ser mantida em um valor pequeno, da ordem de 0,01, observando que nunca deverá cair a zero e o raio topológico deve conter somente os vizinhos mais próximos do neurônio vencedor, ou até mesmo somente o neurônio vencedor) [6].

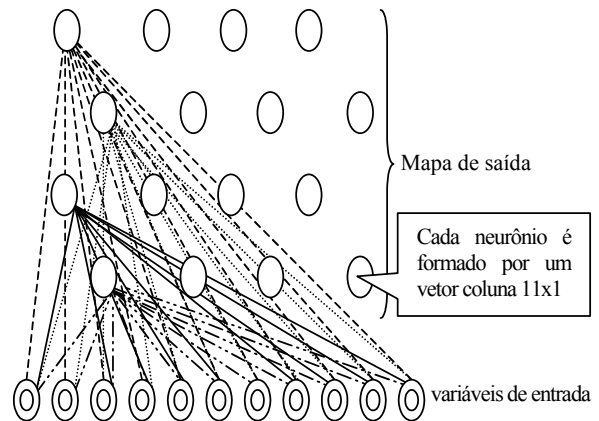


Fig. 1: Ilustração do SOM apresentando a conexão entre as i variáveis de entrada e os neurônios $j=1$ do mapa.

Este trabalho foi realizado no ambiente computacional Matlab 6.0, implementado com a caixa de ferramentas SOM para Matlab 5 (versão 2) disponível no site <http://www.cis.hut.fi/projects/somtoolbox/>.

III. Resultados e Discussões

A rede foi treinada com a apresentação dos padrões (distintamente da fase rio e da fase reservatório), durante 50 épocas para a fase de ordenação e 300 épocas para a fase de convergência.

As redes utilizadas nos dois treinamentos apresentaram a mesma configuração com 11 neurônios na camada de entrada e 48 neurônios no mapa de saída. Para atualização da taxa de aprendizado foi utilizada a função

$$\eta(t) = \frac{\eta_0}{\left(1 + \frac{100t}{T}\right)}, \text{ onde } t \text{ é o número de iterações, } \eta_0 \text{ é a}$$

taxa de aprendizagem inicial e T é uma constante. O raio de vizinhança foi atualizado através da função gaussiana. A fase de ordenação foi inicializada com uma taxa de aprendizado igual a 0,5 e um raio topológico igual a 2. Para a fase de convergência a taxa de aprendizado foi inicializada com 0,05 e o raio topológico com 1, decrescendo até 0 (onde somente o neurônio vencedor era atualizado).

Desta forma, identificou-se quatro grupos especialistas: A-insetívoros, B-piscívoros, C-herbívoros e D-detritívoros. A figura 2 apresenta o mapa resultante da classificação trófica na fase rio, separando os quatro grupos obtidos.

A figura 3 apresenta a localização da ocorrência dos agrupamentos dos recursos alimentares no mapa.

Neste treinamento algumas espécies ocorreram em todos os agrupamentos, ou seja, utilizavam-se de diversos recursos alimentares conjuntamente sendo consideradas generalistas.

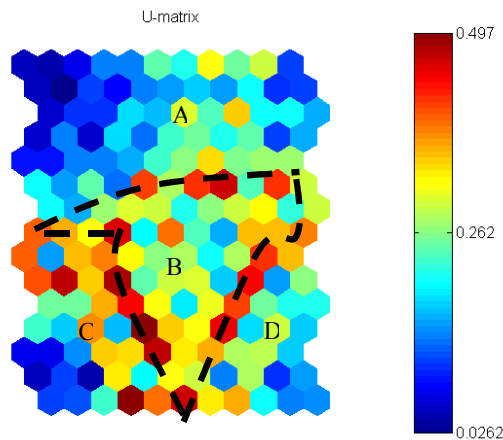


Fig. 2: SOM com separação dos agrupamentos iniciais – fase rio.

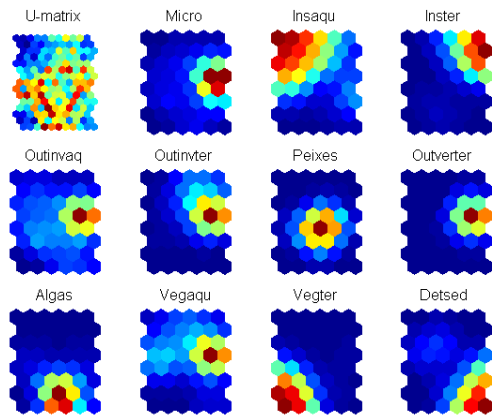


Fig. 3: SOM e recursos alimentares – agrupamento inicial da fase rio.

Observou-se ainda que espécies classificadas como insetívoras ocorreram em outros agrupamentos, motivada pela sua dieta não ser específica de insetos aquáticos, envolvendo outros recursos alimentares associados. Para uma melhor identificação destas espécies, foi treinada uma nova rede gerando o SOM apresentado na figura 5, onde se visualiza a separação das espécies que ingerem insetos aquáticos das que se alimentam deste alimento associado a outros recursos alimentares, ocasionando assim uma divisão deste grupo em quatro sub-grupos: insetívoros aquáticos, insetívoros aquáticos/piscívoros, insetívoros aquáticos/detrítivos e insetívoros aquáticos/terrestres.

A tabela 1 apresenta a classificação trófica das espécies de peixes do reservatório de Corumbá - fase rio, indicando as espécies pertencentes a cada agrupamento, representando cada espécie por um número. Os números utilizados para identificação das espécies podem ser conferidos na tabela 3.

A fase do reservatório foi analisada de maneira análoga à fase rio, apresentando a classificação em grupos similares,

conforme mostra a figura 4, indicando as espécies pertencentes a cada agrupamento.

Como ocorrido na fase rio anteriormente descrita, algumas espécies (nesta fase em maior número) apareceram em todos os agrupamentos, indicando serem generalistas. Espécies do grupo insetívoro apareceram ocasionalmente em outros agrupamentos, tornando-se necessário o treinamento de uma nova rede para uma melhor separação.

Tabela 1: Classificação trófica das espécies de peixes no reservatório de Corumbá- fase rio.

	GRUPOS	ESPECIES
Generalistas		2, 3, 4, 29, 38
Especialistas	I – Insetívoros aquáticos	7
	II – Insetívoros aquáticos/piscívoros	35
	III – insetívoros aquáticos/detrítivos	9, 10, 11, 12, 15, 16, 17, 24, 25, 31, 40, 44, 52
	IV – insetívoros aquáticos/terrestres	42
	B – Piscívoros	20, 45
	C – Herbívoros	27, 33, 50
D - Detritívoros	1, 6, 18, 22, 23, 41, 46	

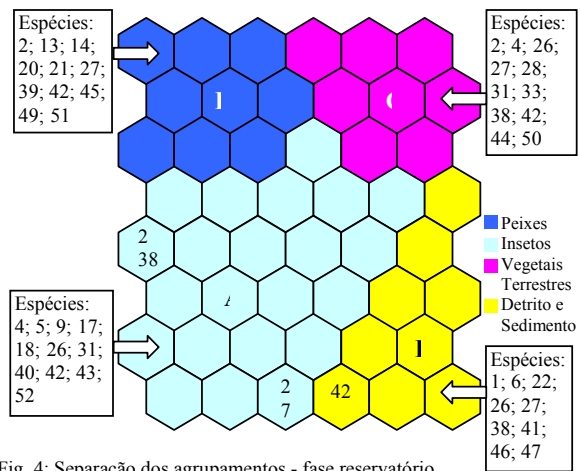


Fig. 4: Separação dos agrupamentos - fase reservatório.

A figura 5 ilustra a separação do grupo insetívoro em quatro sub-grupos: I–insetívoros aquáticos, II–insetívoros aquáticos/piscívoros, III–insetívoros aquáticos/detrítivos e IV–insetívoros aquáticos/terrestres, encontrados na fase reservatório.

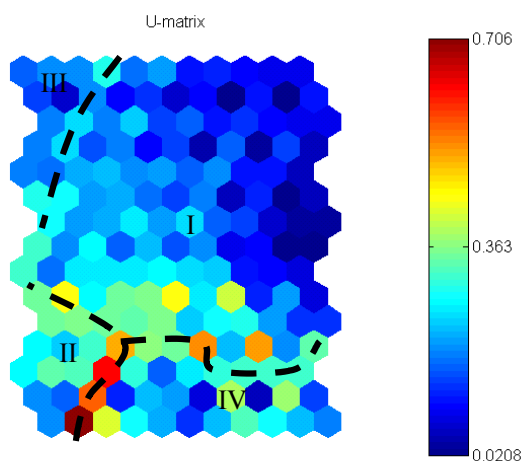


Fig. 5: Separação do grupo insetívoro em quatro sub-grupos – fase reservatório.

A classificação trófica das espécies de peixes do reservatório de Corumbá - fase reservatório é apresentada na tabela 2.

Tabela 2: Classificação trófica das espécies de peixes do reservatório de Corumbá –fase reservatório.

		GRUPOS	ESPÉCIES
Generalistas			2, 4, 26, 27, 31, 38, 42
	Especialistas	A-Insetívoros	I – Insetívoros aquáticos
II – Insetívoros aquáticos/piscívoros			37, 40, 43
III – insetívoros aquáticos/detrítívoros			18, 24, 25
IV – insetívoros aquáticos/terrestres			9
B – Piscívoros		13, 14, 20, 21, 39, 45, 49, 51	
C – Herbívoros	28, 33, 44, 50		
D - Detritívoros	1, 6, 22, 41, 46, 47,		

Tabela 3: Código, espécies de peixes e nome popular.

Nº	Espécies	Nome popular
01	<i>Apareiodon affinis</i>	“canivete”
02	<i>Astyanax bimaculatus</i>	“tambú”
03	<i>Astyanax eigenmanniorum</i>	“lambari”
04	<i>Astyanax fasciatus</i>	“lambari”
05	<i>Apareiodon ibitiensis</i>	“canivete”
06	<i>Apareiodon piracicabae</i>	“canivete”

07	<i>Astyanax scabripinnis paranae</i>	“lambari”
08	<i>Brycon nattereri</i>	“pirapitinga”
09	<i>Bryconamericus stramineus</i>	“pequira”
10	<i>Bryconamericus</i> sp.	“pequira”
11	<i>Crenicichla haroldoi</i>	“joaninha”
12	<i>Cetopsorhamdia iheringi</i>	“mandizinho”
13	<i>Cichla monoculus</i>	“tucunare”
14	<i>Cichlasoma paranaense</i>	“cará”
15	<i>Characidium zebra</i>	“canivete”
16	<i>Corydoras</i> sp.	“coridoras”
17	<i>Eigenmannia</i> sp.	“tuvira”
18	<i>Gymnotus carapo</i>	“morenita”
19	<i>Glanidium cesarpintoi</i>	“jundiá”
20	<i>Galeocharax knerii</i>	“peixe-cadela”
21	<i>Hoplias malabaricus</i>	“traíra”
22	<i>Hypostomus regani</i>	“cascudo”
23	<i>Hypostomus</i> sp.	“cascudo”
24	<i>Iheringichthys labrosus</i>	“mandi-beiçudo”
25	<i>Leporinus amblyrhynchus</i>	“canivete”
26	<i>Leporinus elongatus</i>	“piapara”
27	<i>Leporinus friderici</i>	“piauí”
28	<i>Leporinus octofasciatus</i>	“ferreirinha”
29	<i>Lebistes reticulatus</i>	“guaru”
30	<i>Leporellus vittatus</i>	“solteira”
31	<i>Moenkhausia intermedia</i>	“pequi”
32	<i>Megalonema platanus</i>	“pati”
33	<i>Myloplus tiete</i>	“pacu-prata”
34	<i>Metynnis</i> sp.	“pacu”
35	<i>Odontostilbe</i> sp.	“pequira”
36	<i>Oreochromis niloticus</i>	“tilápia”
37	<i>Oligosarcus planaltinae</i>	“cadela”
38	<i>Piabina argentea</i>	“pequira”
39	<i>Pseudoplatystoma corruscans</i>	“pintado”
40	<i>Pimelodus fur</i>	“mandi”
41	<i>Prochilodus lineatus</i>	“curimbatá”
42	<i>Pimelodus maculatus</i>	“mandi”
43	<i>Pimelodus paranensis</i>	“mandi”
44	<i>Pimelodella</i> sp.	“mandi-chorão”
45	<i>Rhamdia quelen</i>	“bagre”
46	<i>Steindachnerina corumbae</i>	“sagüiru”
47	<i>Steindachnerina insculpta</i>	“sagüiru”
48	<i>Sternopygus macrurus</i>	“tuvira”
49	<i>Salminus maxillosus</i>	“dourado”
50	<i>Schizodon nasutus</i>	“ximborê”
51	<i>Serrasalmus spilopleura</i>	“piranha”
52	<i>Tatia neivai</i>	“jundiá”

No decorrer do trabalho foram elaborados dezenas de mapas, com diferentes estruturas e funções de vizinhança.

A escolha do “melhor” mapa foi baseada nos erros obtidos em cada treinamento. Para esta análise utilizou-se a função *som_quality* da caixa de ferramentas *SOM* para Matlab 5 (versão 2) [13], que fornece os erros quantitativo e topográfico. O erro quantitativo (*Eq*) mede a resolução do mapa, ou seja, a média da distância euclidiana entre cada vetor do conjunto de dados e seu neurônio vencedor, conforme a expressão abaixo:

$$Eq = \frac{\sum \sqrt{(\text{neurônio vencedor} - \text{vetor de entrada})^2}}{n}$$

onde *Eq* é o erro quantitativo e *n* o número de amostras do conjunto de dados.

O erro topográfico apresenta a proporção de todos os vetores dados para os quais o primeiro e o segundo neurônio vencedor não são unidades adjacentes.

Neste trabalho, o erro quantitativo não foi considerado representativo, pois utiliza a matriz peso dos neurônios vencedores e a matriz dos dados para o cálculo da distância. Como a matriz peso dos neurônios vencedores não é normalizada, este valor depende da dimensão dos dados e, desta forma, não é possível avaliar a qualidade de um determinado mapa comparativamente com mapas obtidos em outros treinamentos. Por esta razão foi calculado o erro geral, que apresenta a porcentagem de amostras que foram alocadas fora da região de seu agrupamento. Para o cálculo foram utilizados os mapas individuais de cada espécie de peixe, observando-se a região destinada a cada agrupamento. O valor foi obtido pela razão do número de amostras alocadas erroneamente pelo número total na amostra. O cálculo deste erro foi possível devido a uma grande porcentagem de cada amostra estar localizada numa determinada região do mapa, tornando clara a visualização do agrupamento. Como se trata de uma medida adimensional, pode ser utilizada para uma comparação entre mapas obtidos em outros treinamentos.

A tabela 4 apresenta um resumo do treinamento realizado, onde se observa os valores dos erros obtidos nas duas fases analisadas: rio e reservatório.

Tabela 4: Resumo treinamento *SOM*.

	Padrões	Erro		
		Quant.	Topog.	Geral
TREINAMENTO				
Rio	1318	0,1379	2,30%	9,25%
Reservat.	3872	0,0995	1,85%	3,05%
VALIDAÇÃO				
Rio	527	0,1495	4,80%	8,73%
Reservat.	1620	0,0929	1,68%	3,09%

IV. Conclusões

Algumas espécies foram classificadas em categorias tróficas distintas na fase rio e na fase reservatório. Isto pode ser explicado pelo fato das espécies de peixes serem

dotadas de grande plasticidade na dieta, adaptando-se aos recursos alimentares que se apresentem mais disponíveis.

As análises apresentaram resultados satisfatórios para o espaço amostral estudado, representando o *SOM* um recurso muito útil na discriminação dos grupos e na classificação trófica deste estudo.

Pode-se apontar outros pontos positivos para a utilização do *SOM* como ferramenta na classificação trófica das espécies de peixes, como a facilidade encontrada na visualização dos agrupamentos, a rapidez do modelo e principalmente a facilidade na interpretação dos resultados, pois cada célula do mapa formado após o treinamento informa quais espécies foram agrupadas.

Observou-se ainda que o erro encontrado na fase rio foi aproximadamente três vezes maior que o da fase reservatório. Não é possível generalizar que os erros serão menores quando o conjunto de dados forem maiores, pois não foi possível realizar análises adicionais buscando esta comprovação, mas tudo indica que sim, pois um dos princípios para a aprendizagem do *SOM* é que o conjunto de dados deva apresentar redundância de informações para que a rede possa apresentar um treinamento satisfatório.

Agradecimentos: FURNAS Centrais Elétricas S.A

Referências

- [1] AGUILAR IBARRA, A.; GEVREY, M.; PARK, Y.-S.; LIM, P.; LEK, S. Modelling the factors that influence fish guilds composition using a back-propagation network: assessment of metrics for indices of biotic integrity. *Ecological Modelling*, v.160, p. 281-290, 2003.
- [2] BROSSE, S.; GIRAUEDEL, J.L.; LEK, S. Utilisation of non-supervised neural networks and principal component analysis to study fish assemblages. *Ecological Modelling*, v.146, p.159-166, 2001.
- [3] BROSSE, S.; GUEGAN, J.-F.; TOURENQ, J.-N.; LEK, S. The use of artificial neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake. *Ecological Modelling*, v.120, p.299-311, 1999.
- [4] GASPARD DA LUZ, K.D. *Espectro alimentar e estrutura trófica da ictiofauna do reservatório da UHE Corumbá-GO*. Maringá, 2000. 25 f., il. Dissertação (Mestrado em Ecologia de Ambientes Aquáticos Continentais) – Departamento de Biologia, Universidade Estadual de Maringá, 2000.
- [5] GEVREY, M.; RIMET, F.; PARK, Y.S.; GIRAUEDEL, J.-L.; ECTOR, L.; LEK, S. Water quality assessment using diatom assemblages and advanced modelling techniques. *Freshwater Biology*, v. 49, p.208-220, 2004.
- [6] HAYKIN, S. *Redes neurais: princípios e prática*. 2.ed. Tradução de, Paulo Martins Engel. Porto Alegre: Bookman, 2001. 900 p., il. Obra originalmente publicada sob o título: *Neural Networks: a comprehensive foundation*, 2/E.
- [7] HYSLOP, E.J. Stomach contents analysis, a review of methods and their application. *Journal Fish Biology*, v.17, p.411-422, 1980.
- [8] LEE, J.H.W.; HUANG, Y.; DICKMAN, M.; JAYAWARDENA, A.W. Neural network modelling of coastal algal blooms. *Ecological Modelling*, v.159, p.179-201, 2003.
- [9] LEK, S.; BELAUD, A.; DIMOPOULOS, I.; LAUGA, J.; MOREAU, J. Improved estimation, using neural networks, of the food consumption of fish populations. *Marine and Freshwater Research*, v.46, no.8, p.1229-1236, 1995.
- [10] MAIER, H.R.; DANDY, G.C. Modelling cyanobacteria (blue-green algae) in the River Murray using artificial neural networks. *Mathematics and Computers in Simulation*, v.43, p.377-386, 1997.

- [11] PENCZAK, T.; LEK, S.; GODINHO, F.; AGOSTINHO, A. A. Patterns of fish assemblages in tropical streamlets using SOM algorithm and conventional statistical methods. *Ecohydrology & Hydrobiology*, v. 4, no.2, p.139-146, 2004.
- [12] RECKNAGEL, F.; FRENCH, M.; HARKONEN, P.; YABUNAKA, K.-I. Artificial neural network approach for modelling and prediction of algal blooms. *Ecological Modelling*, v.96, p.11-28, 1997.
- [13] VESANTO, J.; HIMBERG, J.; ALHONIEMI, E.; PARHANKANGAS, J. *SOM Toolbox for Matlab 5*: report A57, April 2000. Libella Oy: Finland: SOM Toolbox Team, Helsinki University of Technology, 2000b. 59 p., ill. Disponível em: <<http://www.cis.hut.fi/projects/somtoolbox>> . Acesso em: 06 Aug. 2004.