



Indici di idoneità d'habitat: un approccio neurale

Gianandrea La Porta^{*}, Maria Pia Spigonardi, Massimo Lorenzoni, Mario Mearelli

Dipartimento di Biologia Cellulare ed Ambientale, Via Elce di Sotto, Perugia 06123, Italia

Abstract

Scopo della ricerca è quello di mettere a punto un metodo per stimare le esigenze d'habitat per alcune specie ittiche caratterizzanti i corsi d'acqua delle rete idrografica del bacino del fiume Tevere. Il livello di gradimento delle condizioni ambientali è stato espresso attraverso un indice di idoneità (HSI) che si basa su osservazioni sperimentali pluriennali su 18 corsi d'acqua. La tecnica sfrutta le potenzialità delle reti neurali artificiali (ANN) che sono in grado di superare alcuni limiti dei modelli HSI riportati in letteratura. L'approccio neurale prescinde, infatti, dagli assunti di indipendenza tra le grandezze d'habitat utilizzate per il calcolo dell'indice di idoneità e dal rapporto di linearità tra le variabili indipendenti e quelle dipendenti. I risultati conseguiti permettono di stimare la validità del percorso metodologico adottato e, attraverso l'analisi di sensibilità, i contributi di ciascuna variabile ambientale all'indice di idoneità. © 2005 SItE. All rights reserved

Keywords: ANN; reti neurali; modelli non lineari; zonazione ittica; densità; indici di idoneità, bacino del Tevere

1. Introduzione

Gli ecologi hanno impegnato molte energie per cercare di correlare l'abbondanza delle popolazioni animali con gli attributi d'habitat. L'interesse per la ricerca di questo legame è data sia da motivazioni di tipo teorico che da necessità pratiche rilevanti per coloro che si occupano di pianificazione e gestione della fauna selvatica. La stima delle densità e della biomassa animale sono tematiche centrali in ecologia e in modo particolare per l'itticoltura. In passato sono state utilizzate numerose tecniche per spiegare le connessioni tra habitat e abbondanza faunistica: analisi canoniche, regressioni logistiche, lineari, multiple (Chon *et al.* 1996; Recknagel 1997; Manel *et al.* 1999; Scardi & Harding 1999; Dimopoulos *et al.* 1999). Queste tecniche sono però limitate dalla necessità di assumere relazioni concettualmente e

formalmente semplici tra le grandezze esaminate (relazioni lineari, polinomiali, logaritmiche, potenza). Al contrario, le reti neurali artificiali (ANN) sono in grado di superare queste difficoltà (Lek & Guégan 1999) e sono conosciute per la loro abilità nella ricerca di pattern non lineari. Le ANN sono uno strumento di analisi ampiamente utilizzato in varie discipline: fisica, chimica, ingegneria, economia, ma ancora poco sfruttate in ambito ecologico. In questo contributo è stato sfruttato lo studio del comportamento delle reti neurali per interpretare l'influenza di alcune variabili d'habitat sulle comunità ittiche del bacino del fiume Tevere.

2. Materiali e Metodi

Settantasei stazioni appartenenti a 18 differenti corsi d'acqua sono state esaminate nella rete idrografica umbra del bacino del fiume Tevere

^{*} Corresponding author. Tel.: +39-075-585-5711; fax: +39-075-585-5733; e-mail: mearelli@unipg.it.

durante una campagna di ricerca pluriennale terminata nel 2005. Le stazioni di campionamento sono state scelte cercando di assicurare la massima rappresentatività dei differenti ambienti lotici presenti nel bacino e tenendo conto delle diverse facies fluviali. Questo ha consentito di costituire una base dati in cui i dati ambientali risultano appaiati con le informazioni biologiche (Tab. 1). I campionamenti ittici sono stati condotti mediante pesca elettrica, con elettroscorditori di potenza pari a 2.5 e 4 kW, dotati di commutatore per corrente continua pulsata o non pulsata, e alternati in base alla conducibilità dell'acqua ed alla tipologia fluviale riscontrate.

Tabella 1

Elenco delle variabili utilizzate. Con i sono indicate le variabili di input e o le variabili di output.

Variabile	Tipo	Descrizione
Bacino	i	Superficie di bacino sottesa
Quota	i	Quota altimetrica
Pendenza	i	Pendenza % del tratto
Dens di drenaggio	i	Densità di drenaggio
Superficie	i	Superficie stimata del tratto
Riffle	i	Frazione di <i>facies</i> di riffle
Run	i	Frazione di <i>facies</i> di run
Pool	i	Frazione di <i>facies</i> di pool
Sup Ombreggiata	i	Superficie ombreggiata del tratto
Granulometria	i	Tipologia prevalente
N granulometria	i	Numero di tipologie
Cop vegetale	i	Copertura vegetale del fondo
Tirante idrico	i	Tirante idrico medio
Dev tirante	i	Deviazione standard del tirante
Velocità	i	Velocità di corrente media
Dev velocità	i	Deviazione standard della velocità
Portata	i	Portata
Per bagnato monte	i	Perimetro bagnato di sezione
Per bagnato valle	i	Perimetro bagnato di sezione
Prof media monte	i	Profondità media di sezione
Prof media valle	i	Profondità media di sezione
Larghezza monte	i	Larghezza media di sezione

continua

Larghezza valle	i	Larghezza media di sezione
VH	i	Rapporto velocità profondità
LP	i	Rapporto larghezza perimetro bagnato
LH	i	Rapporto larghezza profondità
DBarbo	o	Densità <i>Barbus tyberinus</i>
DCavedano	o	Densità <i>Leuciscus cephalus</i>
DRovella	o	Densità <i>Rutilus rubilio</i>
DTrota	o	Densità <i>Salmo (trutta) trutta</i>

La modellizzazione delle reti neurali è stata eseguita utilizzando l'algoritmo di *backpropagation* (Abdi 1994). Una rete di questo tipo include tre tipi di livelli cellulari: un livello (*layer*) di input, uno di output ed uno nascosto (hidden layer) di connessione. Ogni layer è composto da una serie di elementi, che prendono il nome di neuroni. Ogni neurone è connesso con i neuroni dello strato successivo mediante assoni, ma è indipendente rispetto a quelli del suo stesso livello (modello *feed-forward*) (Fig. 1). Il legame tra un livello (*i*) e il successivo (*j*) è espresso attraverso una serie di segnali che vengono amplificati o contratti da fattori ponderali (W_{ji}). L'espressione matematica dell'attivazione di un neurone è riassunta dall'equazione:

$$a_j = \sum X_i W_{ji} \quad (1)$$

dove:

a_j indica il coefficiente di attivazione del j^{mo} neurone
 X_i il valore assunto dall' i^{ma} cellula
 W_{ji} il fattore ponderale di connessione tra la i^{ma} e la j^{ma} cellula.

L'output di un neurone è invece determinato da una funzione di trasferimento e dal suo valore di input. La funzione di trasferimento utilizzata in questo lavoro è la classica sigmoide:

$$f(a_j) = 1 / (1 + e^{-a_j}) \quad (2)$$

Poiché i tratti fluviali indagati appartengono a due zone ittiche (barbo e trota) sono state messe a punto due diverse reti neurali indipendenti. La rete della zona del barbo è composta da 26 elementi di input, le variabili d'habitat, e 3 di output, relativi alle densità ittiche delle principali specie guida: barbo, cavedano

e rovello (Fig. 1). Quindici neuroni hanno composto lo strato nascosto e, per evitare situazioni di overfitting, il numero di iterazioni possibili è stato limitato a 500, che nella modellistica con rete neurale è un valore piuttosto basso. Per l'individuazione dei pesi di propagazione del segnale sono stati impostati dei valori iniziali compresi tra -0.01 e +0.01 in modo random.

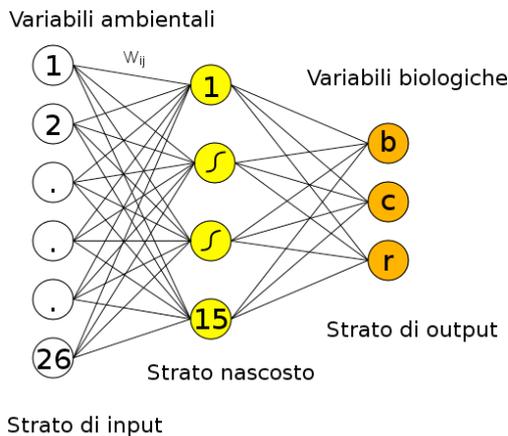


Fig. 1. Rete neurale artificiale utilizzata per la zona del barbo. Le variabili ambientali sono state utilizzate come variabili di input e prima del loro ingresso nell'ANN state normalizzate in una scala 0-1. Anche le variabili di output, rappresentate dalle densità ittiche, sono state trasformate in un intervallo 0-1, in associazione con l'utilizzo della funzione logistica usata per la trasmissione del segnale.

Per la fase di addestramento della rete è stata utilizzata una variante della classica tecnica leave one out (Recknagel 1997) in cui, per essere ancor più restrittivi, si è scelto di utilizzare 100 osservazioni per la formulazione della previsione e di utilizzarne 10 per la verifica. L'individuazione di queste 10 osservazioni è stata condotta attraverso un campionamento randomizzato e il ciclo di calcolo della rete ripetuto per 1000 volte.

Utilizzando una metodica del tutto simile a quanto fatto per la zona del barbo è stato calcolato un modello ANN anche per la stima delle densità di *Salmo (trutta) trutta*. Il set di dati si compone di 78 osservazioni, delle quali 68 utilizzate in modo random per la fase di addestramento e previsione delle altre 10. In questo caso la rete è stata elaborata con 2 neuroni nello strato nascosto.

Per ottenere indicazioni sul contributo che ciascuna variabile indipendente porta all'output della rete neurale è stata eseguita un'analisi di sensibilità (Lek *et al.* 1996; Ozesmi & Ozesmi 1999). I risultati di questa analisi, oltre a fornire un utile approfondimento per la comprensione del modello neurale, sono stati utilizzati per la determinazione degli indici (HSI) delle 4 specie. Per la sua conduzione sono state "congelate" tutte le variabili indipendenti meno una, quella soggetta ad osservazione, al loro valore medio ed esaminata la conseguente variazione di densità ittica.

3. Risultati

L'esecuzione del processo iterativo neurale in una prima fase ha prodotto un addestramento delle ANN per il confronto di set selezionati con la tecnica modificata del leave one out. I coefficienti di correlazione tra i valori stimati dal modello della zona del barbo e quelli osservati sperimentalmente nei 1000 cicli della fase di addestramento sono rappresentati in figura 2.

L'elevato valore del coefficiente di correlazione lineare, che nella classe modale è di oltre lo 0.9 per il barbo, compreso tra 0.7 e 0.8 per il cavedano e dello 0.9 per la rovello, indicano che l'ANN è in grado di fornire stime previsionali in linea con i dati riscontrati sul campo. Terminato l'addestramento è stato prodotto il modello finale che include l'intero set di dati disponibili.

Il coefficiente di determinazione per il barbo è risultato essere altamente significativo ($p < 0.001$) pari a 0.912 (R^2 aggiustato 0.911). Inoltre, al *t-test* non è stata rilevata alcuna differenza statisticamente significativa tra i dati osservati ed attesi ($df=213$; $p=0.99$). L'analisi dei residui ha permesso di rilevare una totale indipendenza tra gli scarti della regressione e i valori stimati, mostrando un buon adattamento all'ipotesi di normalità.

Anche per il cavedano si è registrata una correlazione altamente significativa tra valori attesi ed osservati ($R^2=0.914$; R^2 aggiustato=0.913), senza rilevare tra di essi differenze statisticamente significative (*t-test* $df=189$; $p=0.52$). Lo studio delle relazioni tra residui e valori stimati ha evidenziato

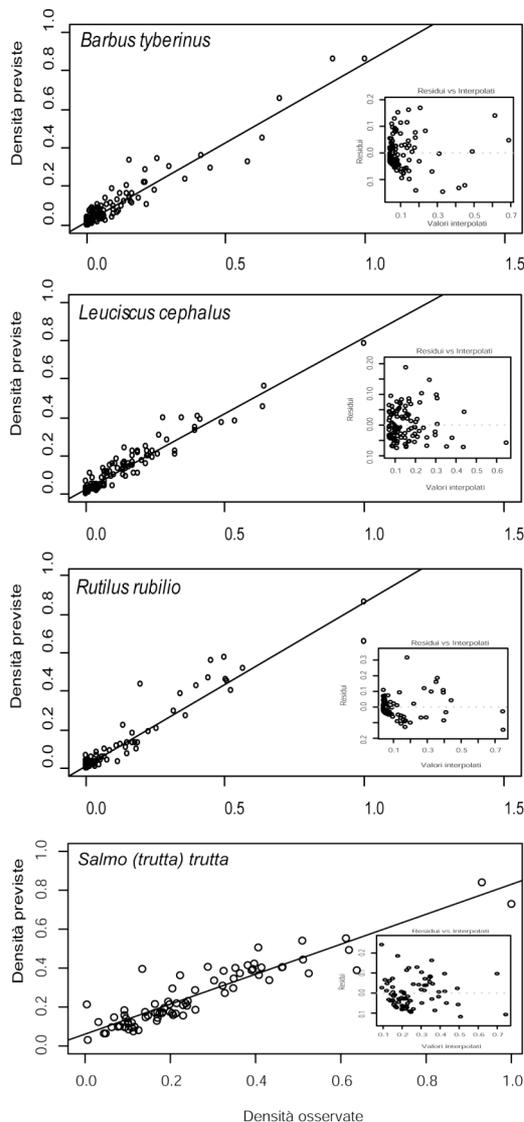


Fig. 2. Analisi di regressione tra valori di densità ricavati sperimentalmente e valori stimati dall'ANN sulla base delle condizioni morfo-idrologiche dei corsi d'acqua. A lato di ciascun grafico sono riportate le analisi dei residui. Per i dettagli si rimanda alla lettura del testo.

una totale indipendenza, con una distribuzione molto prossima alla normalità.

Per la rovello l' R^2 tra densità attese ed osservate è pari a 0.92 ($p < 0.001$) con un t-test sulle differenze non significativo ($t = -0.536$; $df = 215$; $p = 0.992$). Anche in questo caso i residui sono normalmente distribuiti ed indipendenti dal valore stimato.

Il coefficiente di determinazione mostra un range di variazione piuttosto ampio, ma le frequenze maggiori sono da associare alle classi 0.6-0.9 (72.3%) con un valore modale pari all'intervallo 0.8-0.9 (22.4%).

Per la rete della zona della trota il coefficiente di correlazione è risultato altamente significativo ($p < 0.001$), pari a 0.845, (R^2 aggiustato 0.843). Inoltre, i risultati del t-test hanno confermato la sovrapposizione tra i dati sperimentali e quelli forniti dall'ANN ($t = -0.031$, $df = 148$, $p = 0.975$). L'analisi dei residui ha evidenziato una indipendenza tra dati previsti e scarti, i quali mostrano una distribuzione prossima alla normalità.

L'analisi di sensibilità ha permesso di rappresentare nello spazio delle fasi la variazione di densità, normalizzata in una scala 0-1, in funzione di quella specifica variabile. Così facendo sono stati ottenuti una serie di rappresentazioni, associate con funzioni matematiche, che non solo possono essere lette come contributi relativi al modello neurale, ma che rappresentano in senso biologico l'espressione degli indici di idoneità di ciascuna specie rispetto alla variazione dei diversi parametri ambientali. In particolare in figura 3 sono riportati esempi di curve di preferenza univariate per le tre specie guida della zona del barbo. La limitazione al modello univariato è stata scelta solo per consentire un loro confronto rispetto a quelle riportate in letteratura (Bovee 1982; Crosa *et al.* 1988), ma non esistono impedimenti alla stima di più interessanti funzioni bi o multivariate.

4. Conclusioni

La stima della densità animale attraverso metodi indiretti è una tematica che si incontra spesso in ecologia e la tecnica presentata in questo lavoro sembra fornire un percorso metodologico valido ed accurato. Alcune indagini condotte in ambito ittologico avevano già evidenziato tali capacità (Lek & Baran 1997; Brosse *et al.* 1999; Lae *et al.* 1999), derivanti soprattutto dalla possibilità di superare i modelli interpretativi convenzionali e di adottare relazioni non lineari (Guègan *et al.* 1998; Lek *et al.* 1996). Questa tecnica può essere messa a frutto per la sua capacità predittiva in circostanze in cui si attua una variazione delle caratteristiche delle variabili

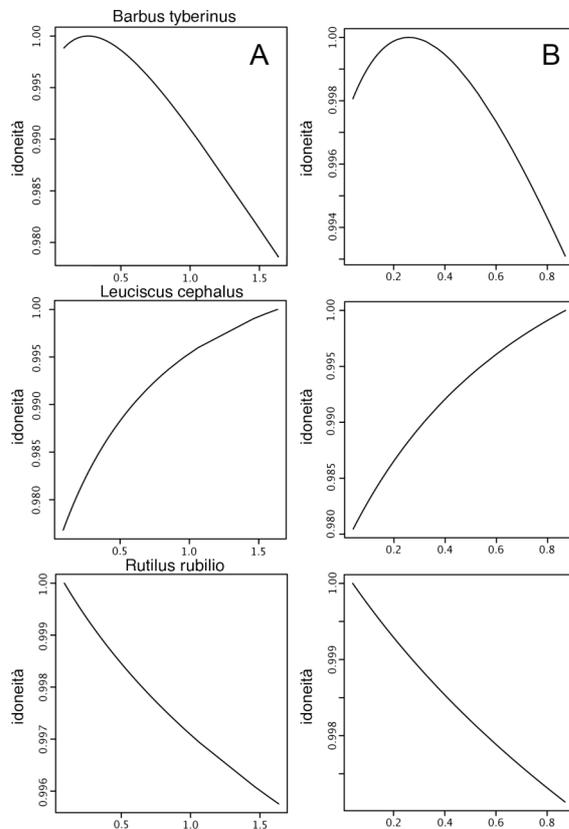


Fig. 3. Esempi di curve di idoneità univariate. La sequenza verticale contraddistinta con la lettera A si riferisce alle variazioni di densità in relazione alla variazione del tirante idrico, mentre la B rispetto alla variazione della velocità di corrente.

d'habitat e si vogliono stimare gli effetti ottenuti sulle comunità ittiche. In particolare, grazie all'analisi di sensibilità è possibile individuare quali forzanti ambientali giocano un ruolo determinante per l'espressione del livello di gradimento che le specie ittiche mostrano al variare delle condizioni ambientali. Le curve di idoneità ottenute con questa tecnica possono essere confrontate con i dati di letteratura, tuttavia occorre tener conto del fatto che, mentre con i metodi tradizionali le variabili ambientali sono considerate indipendenti una dall'altra, nel caso neurale le grandezze in gioco agiscono in modo sinergico rispetto al risultato finale. L'artificio del congelamento di tutte le variabili al valore medio, tranne quella oggetto di studio, rappresenta un utile via per superare la limitazione dell'impossibilità di

comprensione delle dinamiche interne alla rete stessa. È altresì opportuno rilevare che alcune variabili difficilmente possono essere ritenute indipendenti tra loro e che il modello rappresentativo univariato è una forzatura rispetto alla complessità dei sistemi naturali. Per tal motivo le indicazioni che sono state ottenute in questa indagine possono essere considerate soltanto preliminari e ulteriori sforzi dovranno essere fatti verso l'elaborazione di più efficaci modelli descrittivi multivariati.

Inserite in un contesto più ampio le analisi di sensibilità offrono inoltre un'interessante alternativa alle tecniche conosciute per la stima degli indici HSI che nell'ambito dell'ecologia delle acque interne, rappresentano un punto centrale per la valutazione della disponibilità di habitat e dei minimi deflussi necessari per garantire il perpetuarsi delle comunità biologiche. I risultati ottenuti si integrano infatti perfettamente con i metodi di stima IFIM/PHABSIM o RHYABSIM riguardanti le portate di minimo vitale.

Bibliografia

- Abdi, H. (1994) A neural network primer, *Journal of Biological Systems*, **2**(3), 247-283.
- Bovee, K.D. (1982) A guide to stream habitat analyses using the instream flow incremental methodology. *Instream Flow Information Paper 12*. US Fish and Wildlife Service, Fort Collins.
- Brosse, S., Guégan, J., Tourenq, J. & Lek, S. (1999) The use of artificial neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake. *Ecological modelling*, **120**, 299-311.
- Chon, T., Park, Y., Moon, K. & Cha, E. (1996) Patternizing communities by using an artificial neural network. *Ecological modelling*, **90**, 69-78.
- Crosa, G., Cotta-Ramusino, M. & Marchetti, R. (1988) Determinazione delle portate minime necessarie per la tutela della vita acquatica in corsi d'acqua soggetti a derivazioni o ritenute. *Acqua e Aria*, **7**, 839-850.
- Dimopoulos, I., Chronopoulos, J., Chronopolou-Sereli, A. & Lek, S. (1999) Neural network models to study relationships between lead concentration in grasses and permanent urban descriptors in Athens city (Greece). *Ecological modelling*, **120**, 157-165.

Guégan, J., Lek, S. & Oberdorff, T. (1998) Energy availability and habitat heterogeneity predict global riverine fish diversity. *Nature*, **391**, 382-384.

Lae, R., Lek, S. & Moreau, S. (1999) Prediction fish yield of african lakes using neural networks. *Ecological modelling*, **120**, 325-335.

Lek, S. & Baran, P. (1997) Estimations of trout density and biomass: a neural networks approach. *Nonlinear Analysis, Theory, Methods and Applications*, **30**, 4985-4990.

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J. & Aulagnier, S. (1996) Application of neural networks to modelling nonlinear relationships in ecology. *Ecological modelling*, **90**, 39-52.

Lek, S. & Guégan, J. (1999) Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological modelling*, **120**, 65-73.

Manel, S., Dias, J.M. & Ormerod, S. (1999) Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird. *Ecological modelling*, **120**, 337-347.

Ozesmi, S. & Ozesmi, U. (1999) An artificial neural network approach to spatial habitat modelling with interspecific interaction. *Ecological modelling*, **116**, 15-31.

Recknagel, F. (1997) ANNA - Artificial Neural Network model for predicting species abundance and succession of blue-green algae. *Hydrobiologia*, **349**, 47-57.

Scardi, M. & Harding, L. (1999) Developing an empirical model of phytoplankton primary production: a neural network case study. *Ecological modelling*, **120**, 213-223.