ECOLOGICAL MODELLING 203 (2007) 119–131

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/ecolmodel

Macroinvertebrate assemblages in glacial stream systems: A comparison of linear multivariate methods with artificial neural networks

Valeria Lencioni[a], *Bruno Maiolini*[a], *Laura Marziali*[b], *Sovan Lek*[c], *Bruno Rossaro*[b,*]

[a] *Section of Invertebrate Zoology and Hydrobiology, Museo Tridentino di Scienze Naturali, Via Calepina 14, 38100 Trento, Italy*
[b] *Department of Biology, Section of Ecology, University of Milano, Via Celoria 26, 20133 Milan, Italy*
[c] *LADYBIO, Université Paul Sabatier Toulouse-III, 118 Route De Narbonne, 31062 Toulouse Cedex 04, France*

ARTICLE INFO

Article history:
Received 31 December 2004
Received in revised form
16 April 2006
Accepted 26 April 2006
Published on line 13 December 2006

Keywords:
Multivariate analysis
Artificial neural networks
Sensitivity analysis
Chironomidae
Glacial streams

ABSTRACT

The distribution of 19 macroinvertebrate taxa was related to 36 environmental variables in 3 Alpine glacial streams. Principal component analysis (PCA) and a self-organising map (SOM) were used to ordinate sample sites according to community composition. Multiple linear regression (MLR) was carried out with the environmental variables as predictors and each macroinvertebrate taxon as criterion variable, a multilayer perceptron (MLP) used the environmental variables as input neurons and each taxon as output neuron. The contribution of each environmental variable to macroinvertebrate response was quantified examining MLR regression coefficients and compared with partial derivative (Pad) and connection weights approach (CW) methods. PCA and SOM emphasized a difference between glacial (kryal) and non-glacial (krenal and rhithral) stations. Canonical correlation analysis (CANCOR) confirmed this separation, outlining the environmental variables (altitude, distance from source and water temperature) which contributed most with macroinvertebrates to site ordination. SOM clustered kryal, rhithral and krenal in three well separated group of sites. MLR and MLP detected the best predictors of macroinvertebrate response. Pad sensitivity analysis and CW method emphasized the importance of water chemistry and substrate in determining the response of taxa, the importance of substrate was overlooked by linear multivariate analysis (MLR).

1. Introduction

The response of macroinvertebrates to environmental factors in running waters was analyzed in different habitats. Glacial stream biotopes as extreme habitats (kryal) and their biocoenoses (kryon) (Steffan, 1971) were analyzed within an EC project (AASER, Milner et al., 2001). Ward (1994) identified three main types of high-altitude and latitude streams, between the permanent snowline and the tree line on the basis of their origin: glacial-fed (kryal), spring-fed (krenal) and rainfall/snowmelt dominated (rhithral) streams.

Milner and Petts (1994) and Milner et al. (2001) suggested that benthic macroinvertebrates in glacial streams are determined mainly by maximum water temperature and channel stability, with both variables expected to increase downstream of the glacial margin. When the maximum water temperature is <2 °C and the channel is unstable, only chironomids of the genus *Diamesa* are expected (Brittain and Milner, 2001). The

* *Corresponding author*. Tel.: +39 02 50314722; fax: +39 02 50314713.
E-mail address: bruno.rossaro@unimi.it (B. Rossaro).
0304-3800/$ – see front matter
doi:10.1016/j.ecolmodel.2006.04.028

abundance and diversity of macroinvertebrates is predicted to increase with distance from the glacial snout. However, other factors may be critical in determining the distribution and abundance of the different taxa in such streams e.g. food availability, discharge, turbidity, duration of ice/snow cover (Moore, 1979), but the relationship between the environmental variables and the biota in glacial habitats needs further study.

Multivariate statistics have been extensively used in ecology and glacier-fed streams were analysed with a general linear model (GLM) (Rossaro and Lencioni, 1999) and a generalized additive model (GAM) (Castella et al., 2001). Artificial neural networks (ANN) were more recently used to map community structure in a few dimensions (Céréghino et al., 2001) and to understand communities with respect to environmental features (Park et al., 2003). For example ANN were used to predict the species richness of four major orders of aquatic insects (Ephemeroptera, Plecoptera, Trichoptera, and Coleoptera) using four environmental variables (Céréghino et al., 2003; Park et al., 2003). ANN are able to treat non-linear relationships between variables and are less affected by outliers (Park et al., 2004).

In the present work the relationships between benthic macroinvertebrates and environmental factors have been analyzed for three Alpine glacial systems, comparing different linear multivariate analyses: (a) principal component analysis (PCA), (b) canonical correlation analysis (CANCOR), (c) multiple linear regression (MLR), (d) stepwise regression (STMLR) with (e) self-organizing map (SOM) (Giraudel and Lek, 2001)—an unsupervised neural network to pattern the community structure, (f) multilayer perceptron with backward–forward propagation algorithm (MLP) (Gevrey et al., 2004)—a supervised neural network to predict such community, (g) partial derivatives sensitivity analysis (PaD) (Dimopoulos et al., 1995) and (h) connection weights method (CW) (Olden et al., 2004) able to quantify environmental variable importance in ANNs.

The aims of the present paper were an attempt to understand the main factors driving macroinvertebrate response in glacial habitat and to develop a general scenario able to predict glacial fauna distribution in Alpine headwaters.

1.1. The study area

The study area is in the Trentino region (NE Italy), in the Southern Alps, within the Adamello-Brenta Natural Park (Fig. 1). Three glacial systems were investigated: Conca (abbr. C), Niscli (abbr. N) and Vedretta Cornisello (abbr V). The Conca and Niscli glaciers are located on two opposite sides of mount Carè Alto in the Adamello massif (46°6′N, 10°36′E), the Cornisello is located in the Presanella mountain group (46°13′N, 10°41′E). The altitude of the glacial snouts ranged from 2610 m (Niscli) to 2800 m (Cornisello) and 3000 (Conca) m a.s.l.

Two streams were selected in the Conca basin (Maiolini and Lencioni, 2001): the SE facing Conca glacial stream and its NE facing non-glacial tributary. With the exception of station C8, all sampling sites were above the tree line. Eight stations were investigated during the melting season in 1996 and 1997: C0–C1, C2 and C3 on the glacial reach; C6 and C7 on its non-glacial tributary; C4, C5 and C8 downstream of their confluence. All sites were sampled during 5-day periods at six occasions (end of June, beginning of August and mid September in both 1996 and 1997), except C0–C1, sampled in August in 1996, C2, sampled in August and September in 1966.

The Niscli glacial stream has a higher discharge, channel instability and turbidity than the Conca stream (mean suspended sediment concentration $9.8 \pm 13.5\,\mathrm{mg\,l^{-1}}$ in Conca and $21.4 \pm 7.1\,\mathrm{mg\,l^{-1}}$ in Niscli, Lencioni, 2000). Three stations (N0, N1, N2) were investigated within 720 m from the snout. Niscli N0 and N1 were sampled in August and September 1997, N2 only in August.

The Cornisello glacial stream includes four stations (V0, V1, V2, V3), a fifth station (V4) is immediately below lake Vedretta (2605 m a.s.l.). The glacial stream has a moderate instability and high turbidity ($36.1 \pm 25.0\,\mathrm{mg\,l^{-1}}$ of suspended sediments, Lencioni, 2000). Cornisello was sampled in June and September in 1997–1998. Only one sample was collected in August 1997 from station V5, situated about 1 km downstream of the lake, it was excluded from data analysis (Fig. 1).

1.2. Database

Topographical, physical and chemical data were produced according to the AASER protocol (Brittain and Milner, 2001). Water temperature was recorded using digital loggers (Gemini TinyTalk II, Gemini Data Loggers Ltd., Chichester, UK) during the sampling periods (5 days). Point measures of temperature were also recorded during faunal surveys (T_{survey}) using a field multiprobe (Hydrolab). Temperature values are coded in Tables 1–3 and in Figs. 2, 4 and 6 as T_{mean}, T_{min} and T_{max} corresponding to the mean, minimum and maximum water temperature recorded during the 5-day sampling period. Channel stability was evaluated using the bottom component of the Pfankuch index (Pfankuch, 1975): six variables (rock angularity, bed-surface brightness, particle packing, percentage of stable materials, scouring, presence and type of aquatic vegetation) were evaluated summing the scores to provide an overall index of channel stability with a potential range of 15–70, high scores representing unstable and low scores stable channel reaches. Beside quantitative variables, binary coded factors were also introduced in the model: glacial influence, snow cover, riffle, pool, sheer rocks (shrock), presence of *Hydrurus* mats and moss. Code 1 meant presence, 0 absence. The substrate composition was coded as percentage of the different fractions by visual assessment: boulder (>20 cm), cobbles (5–20 cm), gravel (0.2–5 cm), sand (0.01–0.2 cm) and silt-mud (<0.01 cm). Coarse particulate organic matter (BPOM) was separated from the invertebrate samples during sorting and ashed at 500 °C.

Other variables included in data analysis (in parenthesis the abbreviations used) were: water velocity (velocity), altitude, source distance, slope, chlorophyll *a*, river discharge (discharge), pH, conductivity, hardness, alkalinity, SO_4, Cl, SiO_2, N-NO_3, N-NH_4, soluble P-PO_4 and total phosphorous (TP). Other details about environmental factors/variables are in Lencioni and Rossaro (2005).

Within each 15-m long station, benthic macroinvertebrates were collected by kick sampling in 4–10 replicate areas (each $0.1\,\mathrm{m^2}$) using a 30 cm × 30 cm pond net with a 250-μm mesh. Prior to benthic sampling, in each $0.1\,\mathrm{m^2}$ areas, velocity was

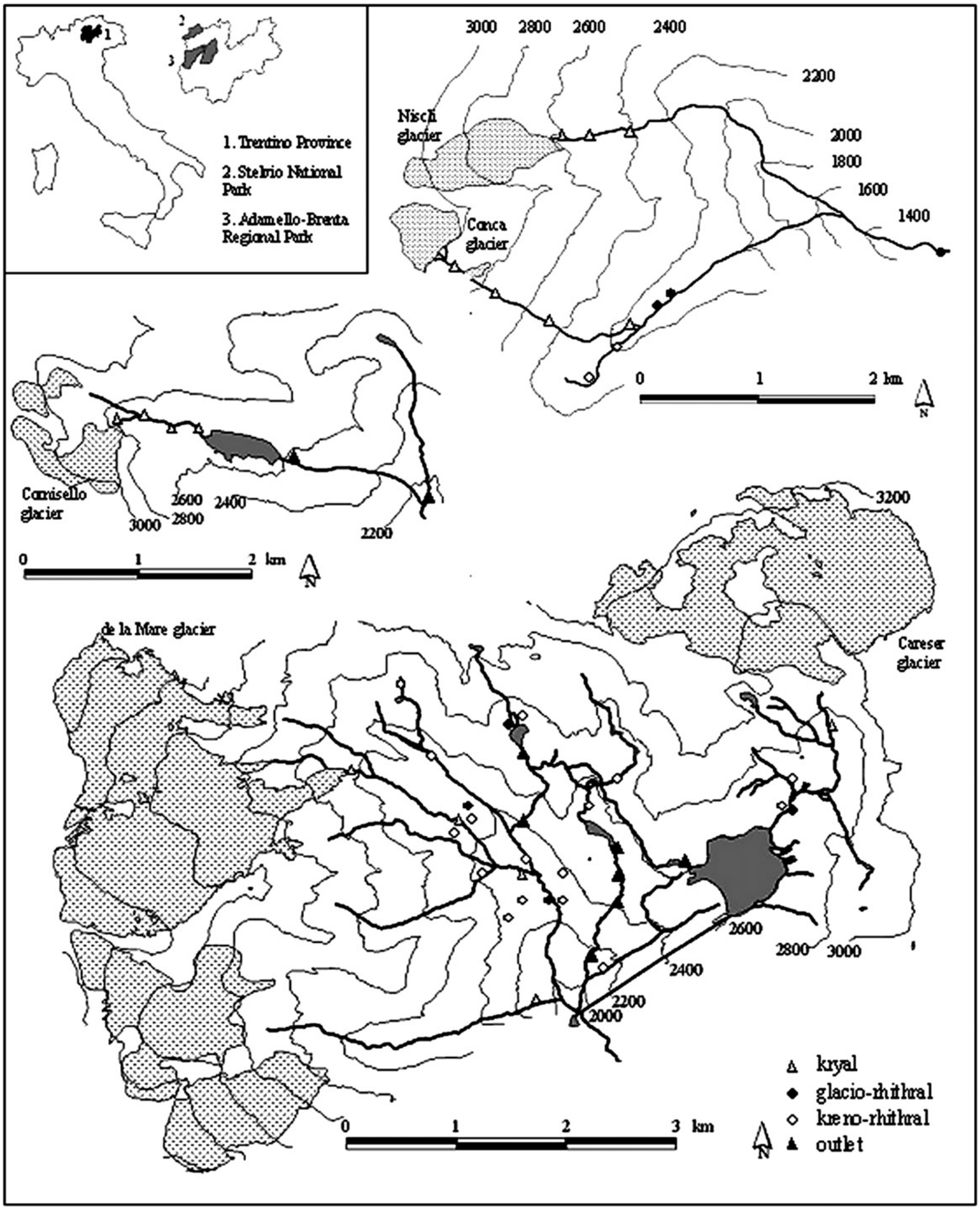

Fig. 1 – Location of sampling sites in the Italian Alps (Trentino, 46°N, 10°E).

measured using a current meter at 0.6 m depth from the bottom. Benthic macroinvertebrates were preserved in 75% ethanol.

Chironomidae, Simuliidae, Ephemeroptera, Plecoptera and Trichoptera were identified to species level, Diptera (except Chironomidae), Oligochaeta, Nematoda and Hydracarina to higher taxonomic levels.

Four hundred and thirty-nine samples were analysed: 16 stations were sampled in 1–3 months for 1–2 years with 4–10 replicates; 36 environmental factors/variables and 72 species were used for data analysis. The 72 species were aggregated into 19 higher taxonomic units of different rank (order, family, genus) to avoid the estimation of a too large number of parameters in performing the analysis, having experienced similar preliminary results in two PCA performed with all the separated 72 species or the 19 aggregated taxa. Within aquatic Insects the taxonomic units considered were the orders of Plecoptera, Ephemeroptera and Trichoptera, the families of Diptera and the subfamilies of Chironomidae (Tanypodinae, Diamesinae, Orthocladiinae and Chironomi-

Table 1 – Stepwise multiple regression results

	Diamesinae	Orthocladiinae	*Eukiefferiella*	Plecoptera	Ephemeroptera	Tanytarsini	Oligochaeta	Limoniidae	Trichoptera
Velocity	–	–	–	+	+	–	+	–	–
BPOM	–	–	–	–	–	–	–	–	–
Boulder	–	–	–	–	–	–	–	–	–
Cobbles	–	–	–	–	–	–	–	–	–
Gravel	–	–	–	–	–	–	–	–	–
Sand	–	–	–	–	–	+	–	+	–
Silt mud	–	–	–	–	–	–	–	–	–
Riffle	–	–	–	–	–	–	–	–	–
Pool	–	–	–	–	+	–	–	–	–
Shrock	–	–	–	–	–	–	–	–	–
Hydrurus	–	–	–	–	–	–	–	–	–
Moss	+	+	–	+	–	–	–	–	–
Altitude	+	–	–	–	–	–	–	–	–
Source distance	–	–	–	+	+	–	–	–	–
Slope	–	+	–	–	–	–	–	–	+
Glacial influence	–	–	–	+	–	+	–	–	–
Snow cover	–	+	–	–	–	–	+	–	–
Chlorophyll *a*	+	+	+	–	–	–	–	+	–
T_{survey}	–	–	–	–	–	–	–	–	–
T_{mean}	+	–	–	+	+	–	–	–	+
T_{min}	–	–	–	–	–	–	–	–	–
T_{max}	+	–	–	–	–	–	–	–	–
Pfankuch	+	–	–	–	–	–	–	+	–
T_{range}	–	+	+	–	–	+	+	–	–
Discharge	–	–	–	–	+	–	–	–	–
pH	+	+	+	+	–	–	–	–	–
Conductivity	+	–	–	–	–	–	+	–	–
Hardness	–	–	+	+	–	–	–	–	–
Alkalinity	+	+	–	–	–	+	–	–	–
SO_4	+	+	+	+	–	–	–	–	–
Cl	+	–	–	+	+	–	–	+	+
SiO_2	+	–	+	+	+	+	+	+	+
N-NO_3	+	–	–	–	+	+	–	–	+
N-NH_4	+	–	–	–	+	–	–	+	–
P-PO_4	+	–	–	–	+	–	–	–	–
TP	+	+	–	+	+	+	+	–	+

	Simuliidae	Empididae	Tricladida	Blephariceridae	Athericidae	Nematoda	Hydracarina	Thaumaleidae	Tipulidae	Tanypodinae
Velocity	+	–	–	–	–	+	–	–	–	–
BPOM	–	–	–	–	–	–	–	–	–	–
Boulder	–	–	–	+	–	+	–	–	–	–
Cobbles	–	–	–	–	–	–	–	–	–	–
Gravel	–	–	–	–	–	–	–	–	–	–
Sand	–	–	–	–	–	–	–	–	–	–
Silt mud	+	–	–	–	–	–	–	–	–	–
Riffle	–	–	–	–	–	–	–	–	+	–
Pool	–	–	–	–	+	–	–	–	–	+
Shrock	+	–	–	–	–	–	–	–	–	–
Hydrurus	–	–	–	–	–	–	–	–	–	–
Moss	–	+	–	–	–	–	–	–	–	–
Altitude	–	–	–	+	–	–	–	–	–	–
Source distance	–	–	–	–	+	+	–	–	–	–
Slope	–	+	+	–	–	+	–	–	–	–
Glacial influence	+	–	–	–	–	–	–	–	–	–
Snow cover	–	–	–	–	–	+	+	–	–	–
Chlorophyll *a*	+	+	–	–	–	+	+	–	–	–
T_{survey}	–	–	–	+	–	–	–	–	–	–
T_{mean}	–	–	–	–	–	–	+	–	–	–
T_{min}	–	–	–	–	+	–	–	+	+	–
T_{max}	–	–	–	–	–	–	–	–	–	–
Pfankuch	–	+	–	–	–	+	–	+	–	–
T_{range}	–	–	+	–	–	–	–	–	–	–
Discharge	+	–	–	+	–	–	–	–	–	–
pH	+	–	–	–	–	–	–	–	–	–
Conductivity	–	–	–	–	–	–	–	–	–	–
Hardness	–	–	–	+	–	+	–	–	–	–
Alkalinity	+	–	–	–	–	–	–	–	–	–
SO_4	–	–	–	–	–	–	–	–	–	–
Cl	+	–	–	+	+	–	–	–	+	–
SiO_2	–	–	–	–	–	–	–	–	–	–
N-NO_3	+	+	+	–	–	–	–	–	–	–
N-NH_4	–	+	–	+	–	+	–	–	–	–
P-PO_4	–	–	–	–	–	–	+	–	–	–
TP	+	–	+	–	–	–	–	–	–	–

+ = variable included in the model and – = not included.

Table 2 – Results of MLR and MLP analysis: correlation coefficients between observed and predicted criterion (target) variable

Criterion/target	MLR				MLP			
	All samples	Training	Testing	Validating	All samples	Training	Testing	Validating
Diamesinae	0.829	0.866	0.726	0.756	0.657	0.743	0.581	0.553
Orthocladiinae	0.876	0.870	0.853	0.805	0.474	0.613	0.425	0.381
Eukiefferiella	0.838	0.848	0.684	0.760	0.511	0.804	0.248	0.485
Plecoptera	0.883	0.906	0.725	0.819	0.626	0.695	0.469	0.586
Ephemeroptera	0.870	0.906	0.643	0.820	0.620	0.658	0.590	0.574
Tanytarsini	0.711	0.772	0.507	0.586	0.439	0.509	0.379	0.456
Oligochaeta	0.659	0.700	0.598	0.489	0.468	0.552	0.148	0.502
Limoniidae	0.679	0.717	0.553	0.572	0.402	0.503	0.356	0.317
Trichoptera	0.700	0.784	0.504	0.534	0.440	0.551	0.322	0.321
Simuliidae	0.672	0.721	0.495	0.404	0.390	0.699	0.184	0.227
Empididae	0.639	0.718	0.408	0.572	0.416	0.717	0.262	0.407
Tricladida	0.598	0.651	0.518	0.265	0.397	0.695	0.227	0.325
Blephariceridae	0.747	0.808	0.583	0.468	0.551	0.607	0.517	0.552
Athericidae	0.762	0.806	0.460	0.667	0.639	0.699	0.457	0.672
Nematoda	0.575	0.665	0.380	0.610	0.298	0.390	0.233	0.363
Hydracarina	0.476	0.596	0.082	0.092	0.312	0.440	0.249	0.257
Thaumaleidae	0.528	0.603	0.258	0.495	0.301	0.551	0.125	0.253
Tipulidae	0.403	0.353	−0.042	−0.024	0.364	0.542	0.065	0.268
Tanypodinae	0.484	0.758	−0.105	0.001	0.021	0.284	0.004	0.000

The criterion variable in MLR (target in MLP) is each of the 19 taxa. The predictors are the 36 environmental variables. Training was carried out with 220 = 439/2 samples, testing and validating with 109 and 110 (=439/4) samples, respectively.

nae). Within Orthocladiinae the genus *Eukiefferiella* (including *Tvetenia*) was separated from the other ones. Non-Insects taxa included were: Oligochaeta, Hydracarina and Nematoda. The pooling of data into 19 taxa of different taxonomic rank was justified, because of the different importance of each taxon in terms of abundance and frequency in the glacial streams. Diptera (especially Chironomids) were particularly abundant and diversified in glacial habitat and required major taxonomical detail.

1.3. *Data analysis*

Taxa abundances were $\log_{10}(x+1)$ transformed to carry out principal component (PCA), canonical correlation (CANCOR),

Table 3 – Criterion or target: taxon used as criterion variable in STMLR, as output neuron in PaD and CW; predictor: environmental variable with the lowest *p*-value in STMLR, where *p*-val is the *p*-value

Criterion or target	Predictor	*p*-Val	Input neuron PaD	PaD	Input neuron CW	CW
Diamesinae	$N\text{-}NO_3$	0.00E + 00	$N\text{-}NH_4$	0.058	$N\text{-}NH_4$	11.152
Orthocladiinae	pH	0.00E + 00	pH	0.025	pH	11.010
Eukiefferiella	pH	8.39E − 10	SiO_2	0.104	SiO_2	12.196
Plecoptera	pH	0.00E + 00	Cl	0.019	Cl	9.339
Ephemeroptera	Source distance	0.00E + 00	Source distance	0.015	Source distance	12.793
Tanytarsini	SiO_2	0.00E + 00	Sio_2	0.005	SiO_2	10.267
Oligochaeta	SiO_2	0.00E + 00	Silt mud	0.014	Boulder	8.782
Limoniidae	Pfankuch	7.00E − 12	Shrock	0.009	Discharge	6.544
Trichoptera	Slope	1.97E − 06	Discharge	0.005	Altitude	10.157
Simuliidae	pH	0.00E + 00	BPOM	0.016	Alkalinity	6.281
Empididae	Pfankuch	2.96E − 08	Cobbles	0.009	T_{min}	5.974
Tricladida	Slope	2.64E − 08	Cobbles	0.018	Gravel	6.324
Blephariceridae	Altitude	0.00E + 00	Silt mud	0.017	Source distance	9.235
Athericidae	Source distance	0.00E + 00	Cl	0.009	Cl	5.454
Nematoda	Snow cover	5.08E − 09	BPOM	0.009	BPOM	5.775
Hydracarina	T_{mean}	4.04E − 05	BPOM	0.008	BPOM	9.416
Thaumaleidae	T_{min}	2.88E − 07	Hardness	0.004	Hardness	6.017
Tipulidae	Cl	2.31E − 08	BPOM	0.008	BPOM	7.750
Tanypodinae	Pool	1.85E − 07	BPOM	0.003	BPOM	14.822

Input neuron PaD: environmental variable more contributing in prediction of target neuron in Pad sensitivity analysis; PaD: sum of the square partial derivative of each environmental variable. Input neuron CW: environmental variable more contributing in prediction of target neuron using the connection weight method; CW: connection weight value.

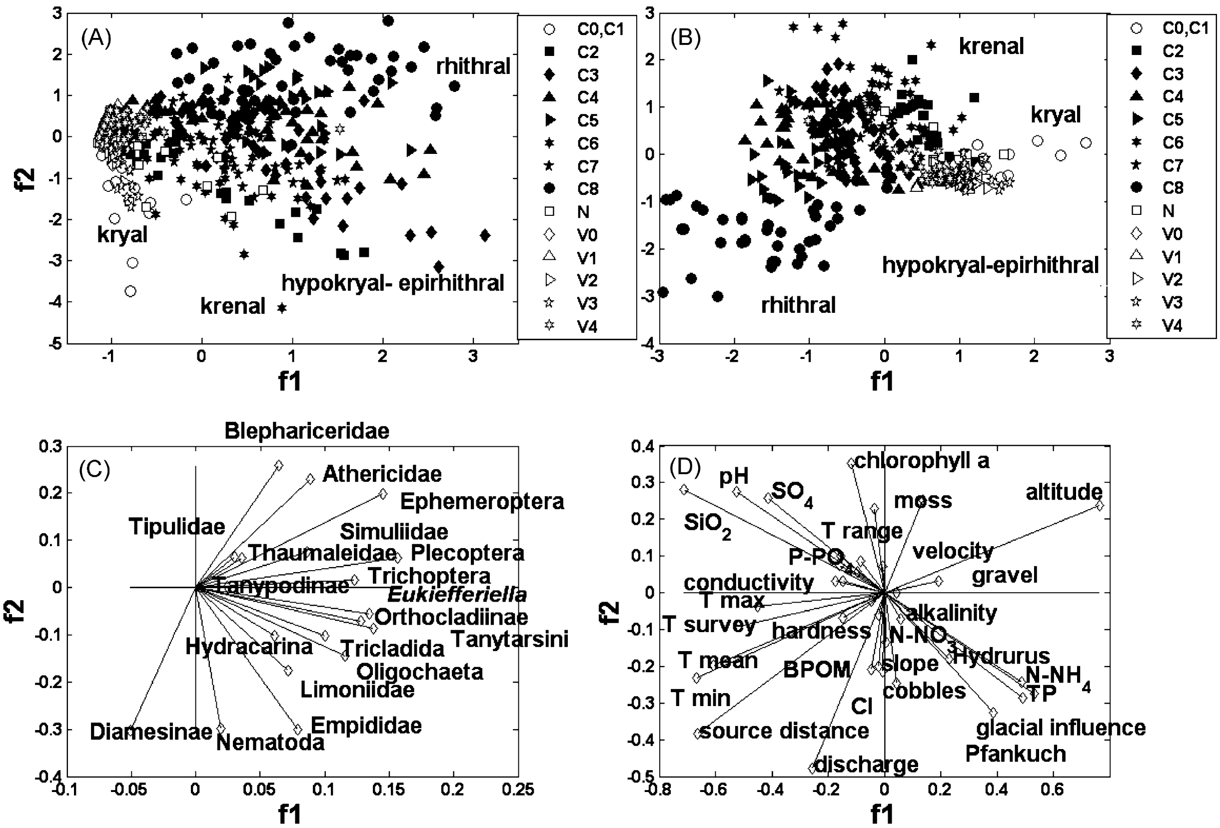

Fig. 2 – (A) Plot of sites factor scores in the first and second PCA axis; (B) plot of sites factor scores in the first and second CANCOR axis (biological set); (C) plot of factor scores coefficients of taxa in the plan of the first and second PCA axis; (D) plot of canonical correlation coefficients of environmental variables in the plan of the first and second CANCOR axis. Stations code: C = Conca, N = Niscli and V = Cornisello.

multiple linear (MLR), stepwise regression (STMLR) and self-organizing map (SOM) analysis. In multi-layer perceptron analysis (MLP) the taxa abundances were rescaled to fall within the −1+1 range; this scaling was requested by automated regularization algorithm (MacKay, 1992).

Environmental variables were standardized subtracting mean and dividing by standard deviation.

PCA was carried out calculating eigenvalues and eigenvectors of the between taxa correlation matrix; eigenvectors were rescaled to factors scores coefficients and to factor structure matrix (Cooley and Lohnes, 1971) to plot taxa; sites scores were then obtained multiplying taxa factor scores coefficients matrix by rescaled original data matrix, to have a low dimension plot of samples. Canonical correlation coefficients of both taxa and environmental variables sets were estimated (Cooley and Lohnes, 1971; Gittins, 1979), canonical variates of both sets were then calculated multiplying the canonical correlation coefficients by rescaled environmental and biological data matrix.

A self-organizing map analysis (SOM) allowed to map sites and species into two dimensions. Environmental variables were then included in two dimension maps (Park et al., 2003). PCA and SOM were able to pattern sampling sites according to community structure and to detect the taxa accounting for the largest proportion of variation. SOM was used as a site ordination and a classification method applied to biological data; the map size is critical: if too small some important information can be lost, if too large a detailed pattern of no ecological significance can appear. The optimum size was established examining the quantization and the topographic error (Park et al., 2003).

The 36 environmental data were related to all 19 taxa in CANCOR, maximizing correlations between linear combinations of the 2 sets of variables. CANCOR allowed plot of taxa, environmental variables and sites.

MLR was carried out including each of the 19 taxa as criterion variable and all the 36 environmental variables as predictors. The significance of predictors in MLR was estimated examining the fiducial limits of regression coefficients, which were considered significant if both lower and upper values of the 0.01 p fiducial limit had the same sign. In stepwise regression the predictors were considered significant when retained in the model with a max p-value to be added = 0.01 and a min p-value to be removed = 0.10.

In MLP all the 36 environmental variables were included as input neurons and each of the 19 taxa as a single target output neuron.

A Bayesian regularization (MacKay, 1992) was used to improve generalization avoiding overfitting in MLP, five hidden neurons used in the model gave the best performance reaching a stable number of parameters after 15–20 epochs. The Matlab ® “trainbr” routine was used for computations.

Models were trained using one half of the samples data (220) and were tested and validated using the other two quarters (109 and 110 samples, respectively).

A sensitivity analysis was carried out to determine the influence of each input (=environmental variable) and to evaluate its contribution to the output (=macroinvertebrate taxon). The partial derivatives (PaD) and the connection weights (CW) methods were used to calculate the relative influence of the environmental variables in the prediction of macroinvertebrates. If there are q input vectors x and one output vector y of N observations, the partial derivative method calculates the gradient vector d_j, $j = 1, 2, \ldots, q$ of the output vector y with respect to the input vectors x_j and sums of the square d_j for all the N observations (Dimopoulos et al., 1995; Gevrey et al., 2003). The connection weights are the products of the input–hidden and the hidden–output connections weights, summed across all the hidden neurons (Olden et al., 2004).

Calculations were performed in Matlab® environment, version 7.2, using Statistic Toolbox to perform MLR and stepwise MLR, Neural Network Toolbox to perform MLP, SOM Toolbox by Vesanto et al. (2000) to perform SOM analysis; the Eco-ANN tool developed by Park and Lek was used to perform Pad sensitivity Analysis and the inclusion of environmental variables in SOM maps; PCA and CANCOR were implemented by the last author adapting Cooley and Lohnes (1971) routines to run in Matlab environment.

2. Results

In the glacial streams, water temperature ranged from a minimum of −0.1 °C at C0–C1 in June 1997 to a maximum of 15.4 °C at C3 in August 1996; in the non-glacial stations temperature ranged from a minimum of 2.0 °C at C5 in September 1996 to a maximum of 13.6 °C at C4 in August 1996. In the tributary of the Conca stream, temperature ranged from a minimum of 1.1 °C at C6 in September 1996 to a maximum of 9.1 °C at C6 in September 1997. Mean temperature increased downstream.

The Pfankuch index indicated a moderate channel instability close to the glacier (C0–C1, C2, C3) and at station C8 (values higher than 30). High stability characterized the stations located on the tributary (C6, C7) and after the confluence with the glacial stream (C4 and C5) with values below 30. In Niscli and Cornisello the highest instability (>30) was measured.

Very low conductivity (maximum <10 μS cm^{-1}), low pH (4.5–6.6) and low alkalinity (−16.9–75.5 μeq l^{-1}) were recorded in the upstream glacial stations of Conca and Niscli. Higher values were observed in Cornisello. Discharge was highest in Niscli (600–1410 l s^{-1}).

In the Conca stream as well as in the tributary, a longitudinal succession of the invertebrate fauna was evident. The uppermost Conca station (C0–C1), all the stations sampled in the Niscli (N0, N1, N2) and the four upstream stations sampled in Cornisello (V0, V1, V2, V3) were characterized by a kryal fauna, dominated by Diamesinae. The dominance of Diamesinae was particularly evident in the Niscli and Cornisello upstream stations, even if the number of specimens in Cornisello was lower. The downstream station of Cornisello (V5) had a transitional fauna between kryal, krenal and rhithral with Orthocladiinae, Simuliidae, Empididae, Limoniidae, Nematoda, Oligochaeta and Ephemeroptera. This station was not included in data analysis because only one replicate sample was available.

A rhithral fauna was found in the downstream Conca stations. At stations C4, C5 and C8 Plecoptera and Ephemeroptera were abundant. Simuliidae, followed by Athericidae and Blephariceridae, prevailed within the Diptera. Among the Chironomidae, Orthocladiinae prevailed downstream along with Chironominae, Tanypodinae were rare in all stations.

The tributary of the Conca stream had a fauna dominated by Chironomidae (Diamesinae and Orthocladiinae at C6 and Orthocladiinae at C7) followed by Plecoptera, other Diptera (mostly Simuliidae) and Ephemeroptera. Simuliidae were more abundant in C6, Empididae in C7.

2.1. *Ordination and classification of community*

The two highest principal components, accounted for the 40% of the total variation. The first two PCA axes emphasized a separation of glacial (kryal, C0, C1-N0-1-2,V0-1-2-3-4) from non-glacial stations (rhithral, C8) in the first axis, less evident was the separation of krenal (tributary C6 and C7) in the second axis (Fig. 2A). A transition zone represented by Conca stations C2–C5 was also observed and labeled hypokryal–epirhithral.

The first principal component separated Diamesinae in the kryal from all the other taxa, the second separated Diamesinae, Nematoda and Empididae in the krenal from Blephariceridae, Athericidae, Simuliidae, Plecoptera, Ephemeroptera and Trichoptera (Fig. 2C).

The CANCOR analysis gave a similar separation of sites and a good agreement was observed between scores calculated from environmental and biological data. The scatter plot of the first two canonical variates calculated from biological data gives a pattern similar to scatter plot of PCA (Fig. 2B). Within kryal the Conca stations (C0–C8) were better separated to each other than Niscli and Cornisello ones.

Both the first and the second axis (Fig. 2D) separated high altitude and low temperature sites from downstream low altitude and high water temperature sites. The first two axes separated also stations characterized by high pH, dissolved SiO_2 and SO_4 from stations characterized by high Pfankuch index (>40), high total phosphorous (TP) and N-NH_4 concentration. The second axis separated sites with high chlorophyll *a* from stations with high discharge (Fig. 2D). Kryal stations (C0–C1, N0-3, V0-4) were well separated because of high Pfankuch index (>40, meaning low stability), TP, N-NH_4 and presence of *Hydrurus* mats. Rhithral downstream station (C8) was separated by high temperature, krenal stations (C6–C7) were separated by high chlorophyll *a* concentration and presence of moss (Fig. 2B–D).

SOM maps of different size were calculated, the minimum quantization error (1.96) and a low topographic error (0.009) was evidenced at size 16 × 7, so this map was selected. The clusters of sites were arranged in the cells of the SOM map (Gevrey et al., 2004). It was again possible to separate different clusters of sites, according to the fauna composition (Fig. 3) and environmental variables (Fig. 4). The most evident sep-

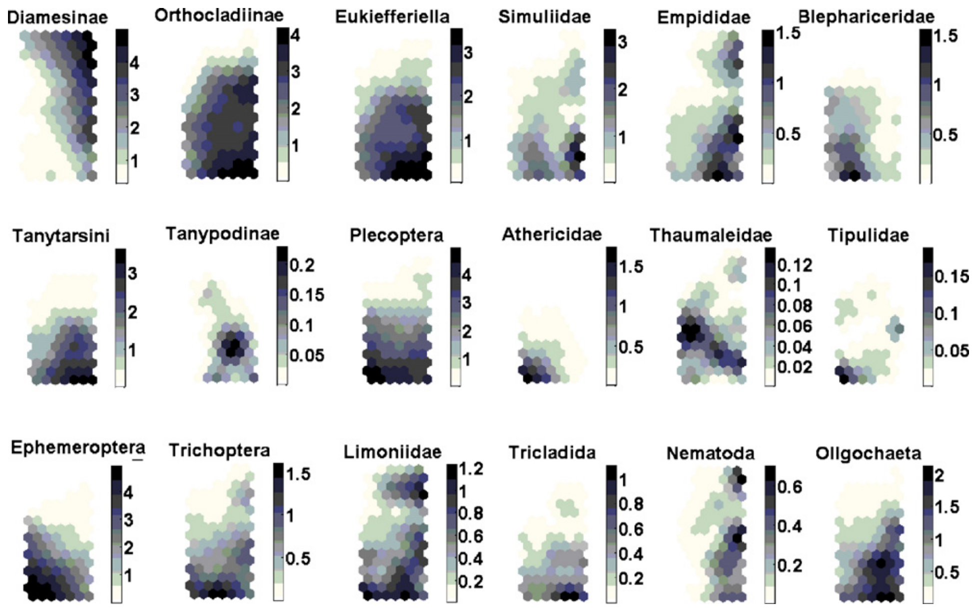

Fig. 3 – SOM map of 18 taxa, values near the bars are the codebook values of each taxon.

aration was between glacial (kryal) and non-glacial (krenal, rhithral) habitats. The glacial sites were all grouped in the upper part of the map (Fig. 5). The sites in Conca (C0, C1, C2), Niscli (N0, N1, N2) and Cornisello (V0, V1, V2, V3) have a fauna dominated by Diamesinae (Fig. 3) and were mapped in the upper part of the map. Differences among the three kryon communities were explainable by differences in abundance of Diamesinae. Many Cornisello sites were mapped in the upper left side, mostly due to the low densities of Diamesinae, which in any case was the dominant group in the sites considered. Conca sites were clustered on the upper right, due to the presence of high numbers of Diamesinae; some sites of Niscli and Cornisello, rich in Diamesinae, were also mapped in the upper right.

Another separation was possible within the non-glacial habitats, all belonging to the Conca system.

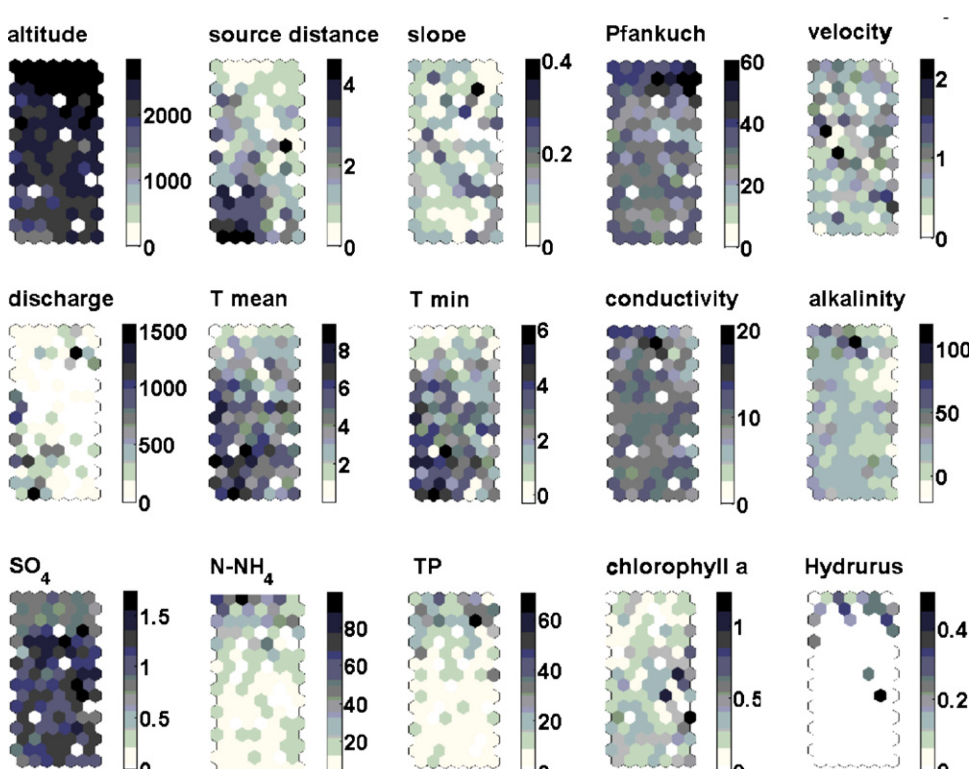

Fig. 4 – SOM map of 15 environmental variables, values near the bars are the values of the environmental variables.

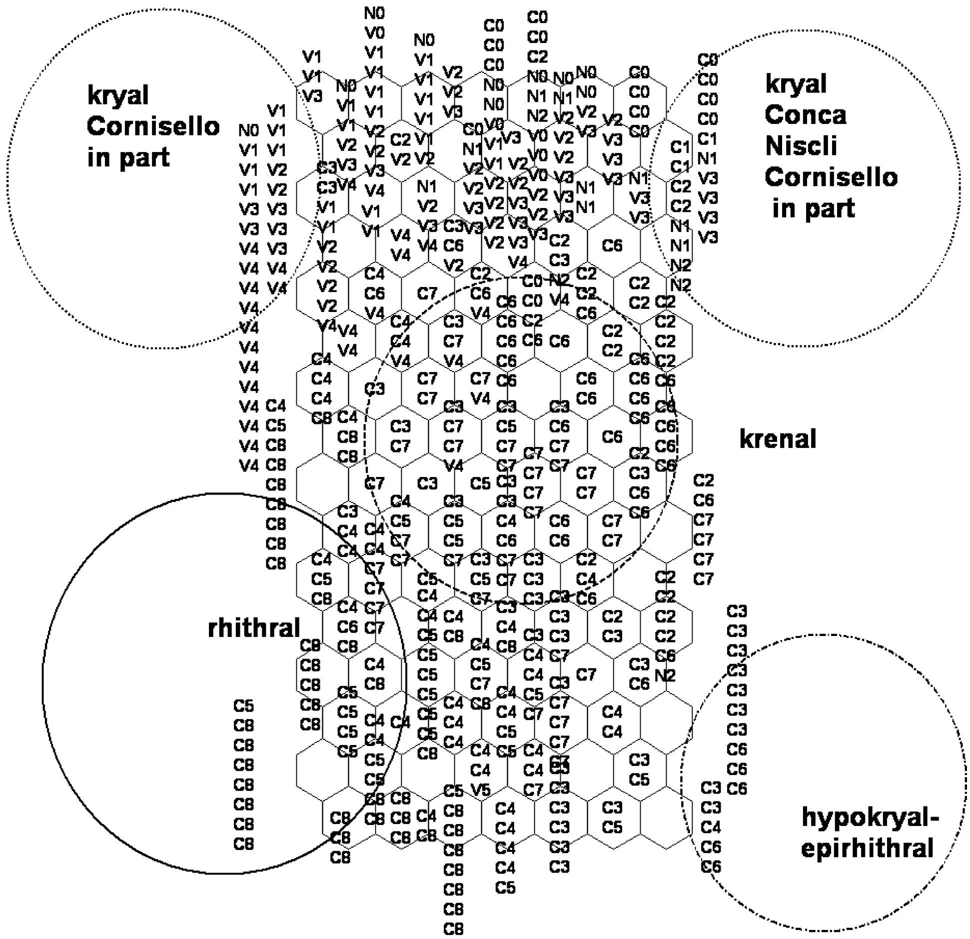

Fig. 5 – Map of the trained SOM units, the codes in each unit of the map represent the sampling sites (see Fig. 1).

The most downstream station (C8), a typical rhithral station, occupied the lower left cells of the map. C8 was characterized by the presence of Plecoptera, Ephemeroptera, Trichoptera, Diptera Athericidae and Blephariceridae, as represented in the taxa maps in Fig. 3. The non-glacial stations in the Conca tributary (C6–C7) were characterized by chironomids Orthocladiinae, Tanytarsini, Tanypodinae and by Nematoda: the corresponding cells were mapped in the central part of the map and represent krenal stations.

An hypokryal–epirhithral zone was also evidenced in the lower right part of the SOM map and can be interpreted as a transition zone.

The SOM maps of the biological variables can be compared with the SOM maps of the environmental variables (Figs. 3 and 4) (Park et al., 2003). Glacial influence, altitude, Pfankuch index, source distance and water temperature well separated kryal from non-kryal sites. NH_4, TP, conductivity had higher values in kryal zone. *Hydrurus* was more abundant in kryal. It was more difficult to evidence environmental variables responsible of the separation between krenal and the other zones. Low distance from source, slope, discharge, temperature, high chlorophyll *a* content characterized krenal, whereas higher distance from source, water temperature, discharge and Cl characterized rhithral.

2.2. *Prediction of taxa*

The environmental variables with the lowest *p*-values (best predictors) in stepwise multiple regression (STMLR) (Tables 1 and 3) were: N-NO_3, pH, source distance, SiO_2, Pfankuch, slope, altitude, source distance, snow cover, T_{mean}, T_{min}, Cl. Maximum water temperature was included as a significant predictor in STMLR only when Diamesinae were the criterion variable (Table 1).

In Table 2 the correlation coefficients between the expected and observed values calculated carrying out a MLR with each the 19 taxa as criterion variables are reported. Minimum and maximum correlation coefficients obtained for training were 0.353 and 0.906, respectively, the values for validation were 0.001 and 0.820. MLP gave lower values: minimum and maximum for training were 0.284 and 0.743, for validation the values ranged from 0.000 to 0.586.

Connection weights (CW) and partial derivative (PaD) sensitivity analysis were carried out to quantify the contribution of each predictive variable to explain each macroinvertebrate taxon. The results were compared with *p*-values of the regression coefficients calculated in STMLR. In was observed a good agreement between CW and PaD analysis. According to both CW and PaD results, N-NH_4, pH, SiO_2, Cl, source

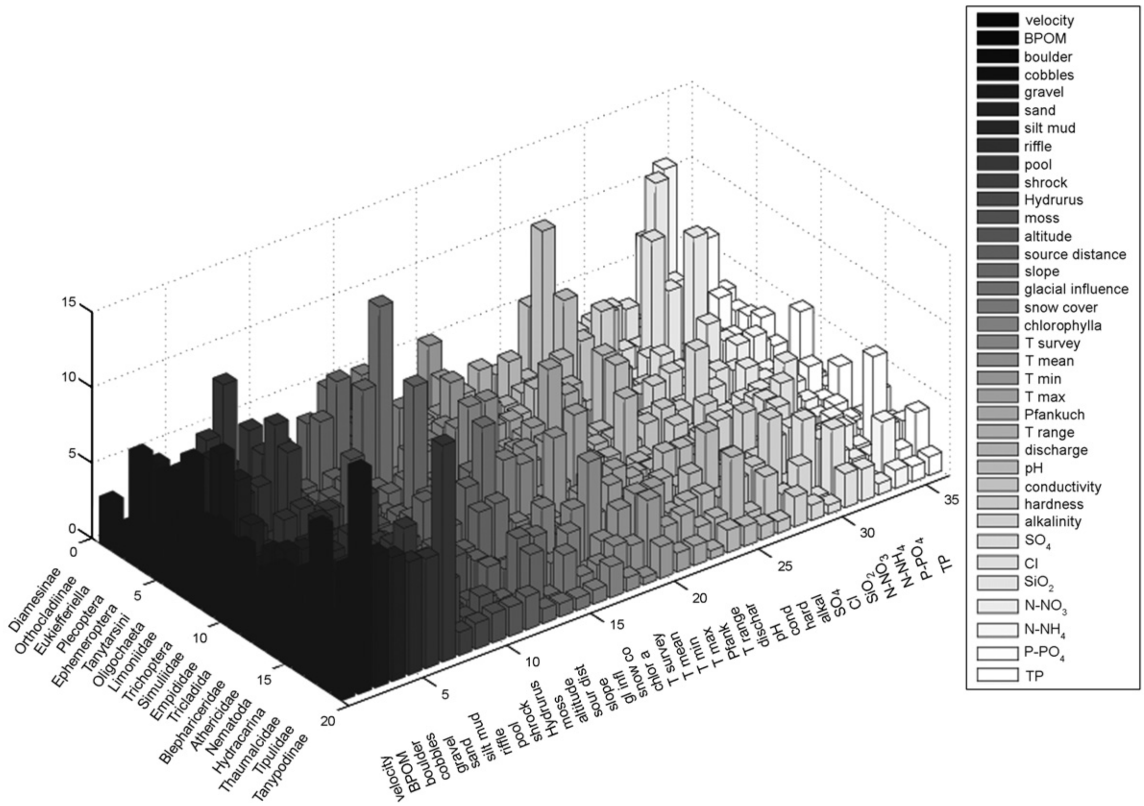

Fig. 6 – Connection weights (CW) between the 36 environmental variables and the 19 taxa.

distance, discharge, BPOM and hardness were the best contributors to the prediction of Diamesinae, Orthocladiinae, Eukiefferiella, Plecoptera and Ephemeroptera (Table 3). Different variables were the most significant predictors in STMLR. Substrate (boulder, cobbles, silt, mud proportion) was a significant predictor in PaD and CW; CANCOR, MLR and STMLR did not detect the importance of substrate. The contribution of each environmental variable to the prediction of each taxon using connection weights (CW) method is summarized in Fig. 6.

3. Discussion

Milner and Petts (1994) proposed a conceptual model to predict the gradient of macroinvertebrate communities in rivers downstream of glacial margins as determined by two principal variables: maximum water temperature and channel stability. Low water temperature and low channel stability create extreme conditions where larvae of the genus *Diamesa* are the sole inhabitants, when conditions are less extreme other taxa colonize the glacial stream. The AASER project (Milner et al., 2001) was planned with the aim of validating Milner and Petts model. An interdisciplinary research carried out by different European research groups emphasized that other factors interact with water temperature and channel stability determining different conditions in different glacier-fed streams. A generalized additive model (GAM) applied to seven glacier-fed European streams, among which the Conca stream, emphasized that maximum temperature, Pfankuch index, suspended solids and tractive force were the explanatory variables most frequently incorporated in the model (Castella et al., 2001).

In the present research non-linear neural networks analysis enforced and expanded the conclusions suggested by GAM. Both techniques (SOM and MLP) applied to the data collected in the Conca, Cornisello and Niscli glacial systems confirmed that other environmental variables of interest were able to predict the response of benthic macroinvertebrates. Maximum water temperature and channel stability (measured as Pfankuch index) were sometime included in the models, but less as expected, other variables characterizing substrate, water chemistry (conductivity, SiO_2, SO_4, $N\text{-}NH_4$, $N\text{-}NO_3$, TP, pH) and variables influenced by biological processes (BPOM and chlorophyll *a*) were more often included in the predictive models.

The glacial/near-glacial condition is a complex of different factors, all effective in modelling benthic fauna and linear models revealed unable to detect some important non-linear relationships.

The aggregation of species into a lower number of taxa (19) was needed to avoid the estimation of a too large number of parameters, but it must be emphasized that the community of glacial habitat was very rich in species composition; the kryal zone of Conca, Niscli and Cornisello was dominated by the chironomid subfamily Diamesinae, but other genera (*Pseudokiefferiella*) besides *Diamesa* were present and different species of *Diamesa* appeared to colonize reaches at different distance from the glacial snout (Lencioni and Rossaro, 2005). These details were not modelled in the present analysis, they will be investigated in the future.

A well separated cold water krenal zone appeared less evident than the kryal and rhithral zones. Krenal was characterized by the dominance of Orthocladiinae, including *Eukiefferiella*, Tanytarsini and Nematoda. Kryal colonized by Diamesinae was well separated from rhithral where Plecoptera, Ephemeroptera and different families of Diptera dominated. Between kryal, krenal and rhithral a transition zone, here named hypokryal–epirhithral zone, was suggested from PCA, CANCOR and SOM results; in this zone many taxa were present, but none was dominant or characteristic of the reach.

The three glacial systems revealed peculiar environmental features. Some characteristics were peculiar of the Conca basin, as a relatively low Pfankuch index in kryal and a relatively high index in rhithral, high slope in krenal and high water temperature in kryal. The moderate instability observed suggested that Conca should not be considered a typical kryal habitat, as Niscli and Cornisello were, characterized by high instability. This must be taken into account in developing a model for glacial streams, to be validated with data from other areas.

The importance of the Pfankuch index and water temperature was confirmed, but other variables as BPOM, chlorophyll *a* and water chemistry (Robinson et al., 2001) must be taken into account in future research in glacial streams.

The advantages of non-linear ANNs respect to linear models were here stressed: the separation between kryal, krenal and rhithral was better evidenced in SOM maps than in PCA and CANCOR plots; it was observed that MLP gave lower correlation coefficient between observed and expected values than MLR, but the automated regularization algorithm used in MLP ensured an improving in model generalization, at the expense of accuracy.

Sensitivity analysis results allowed to extend predictions to non-linear relations, confirming ANN as a powerful data analysis tool: the influence of substrate particle size and of biological factors (BPOM and chlorophyll *a*) on dependent variables, not detected by linear analysis, were emphasized as important predictors by PaD and CW analysis.

Future need is a validation of the present results, analysing samples in other glacial habitat. A larger database will allow subsampling for model testing and the use of information from other glaciers for validating (Lek and Guégan, 2000). The impressive withdrawal of glaciers in the Alps makes these studies more urgent.

Acknowledgements

This work was supported by: (i) the European Union Environment and Climate Programme (contract no. ENV4-CT95-0164), (ii) the Italian MURST n. 2002058154_002 within the Project "An analysis of spatio-temporal distribution of species and populations belonging to critical groups living in inland waters, with morphological and molecular characterization"; (iii) the Italian MURST FIRST 2001-2002-2003: "Taxonomy, ecology, biogeography of Diptera Chironomidae".

REFERENCES

Brittain, J.E., Milner, A.M., 2001. Ecology of glacier-fed rivers: current status and concepts. Fresh. Biol. 46, 1571–1578.

Castella, E., Adalsteinsson, H., Brittain, J.E., Gislason, G.M., Lehmann, A., Lencioni, V., Lods-Crozet, B., Maiolini, B., Milner, A.M., Olafsson, J.S., Saltveit, S.J., Snook, D.L., 2001. Macrobenthic invertebrate richness and composition along a latitudinal gradient of European glacier-fed streams. Fresh. Biol. 46, 1811–1831.

Céréghino, R., Giraudel, J.L., Compin, A., 2001. A Spatial analysis of stream invertebrates distribution in the Adour-Garonne drainage basin (France), using Kohonen self organizing maps. Ecol. Model. 146, 167–180.

Céréghino, R., Park, Y.S., Compin, A., Lek, S., 2003. Predicting the species richness of aquatic insects in streams using a limited number of environmental variables. J. North Am. Benthol. Soc. 22, 442–456.

Cooley, W.W., Lohnes, P.R., 1971. Multivariate Data Analysis. J. Wiley, New York.

Dimopoulos, Y., Bourret, P., Lek, S., 1995. Use of some sensitivity criteria for choosing networks with good generalization ability. Neural Process. Lett. 2, 1–4.

Gevrey, M., Dimopoulos, I., Lek, S., 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecol. Model. 160, 249–264.

Gevrey, M., Rimet, F., Park, Y.S., Giraudel, J.L., Ector, L., Lek, S., 2004. Water quality assessment using diatom assemblages and advanced modelling techniques. Fresh. Biol. 49, 208–220.

Giraudel, J.L., Lek, S., 2001. A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination. Ecol. Model. 146, 329–339.

Gittins, R., 1979. Ecological applications of canonical analysis. In: Orloci, L., Rao, C.R., Stiteler, W.M. (Eds.), Multivariate Analysis in Ecological Work. International Co-operative Publishing House, Fairland, Maryland, USA, pp. 309–535.

Lek, S., Guégan, J.F., 2000. Artificial Neuronal Networks. Springer, Berlin.

Lencioni, V., 2000. Chironomid (Diptera: Chironomidae) assemblages in three alpine glacial systems. PhD Thesis, University of Innsbruck, Innsbruck, Austria.

Lencioni, V., Rossaro, B., 2005. Microdistribution of chironomids (Diptera: Chironomidae) in Alpine streams: an autoecological perspective. Hydrobiologia 533, 61–67.

MacKay, D.J.C., 1992. Bayesian interpolation. Neur. Comput. 4, 415–447.

Maiolini, B., Lencioni, V., 2001. Longitudinal distribution of macroinvertebrate assemblages in a glacially influenced stream system in the Italian Alps. Fresh. Biol. 46, 1625–1639.

Milner, A.M., Petts, G.E., 1994. Glacial rivers: physical habitat and ecology. Fresh. Biol. 32, 295–307.

Milner, A.M., Brittain, J.E., Castella, E., Petts, G.E., 2001. Trends of macroinvertebrate community structure in glacier-fed rivers in relation to environmental conditions: a synthesis. Fresh. Biol. 46, 1833–1847.

Moore, J.W., 1979. Some factors influencing the distribution, seasonal abundance and feeding of subarctic Chironomidae (Diptera). Arch. Hydrobiol. 83, 303–325.

Olden, J.D., Joy, M.K., Russell, G.D., 2004. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. Ecol. Model. 178, 389–397.

Park, Y.S., Céréghino, R., Compin, A., Lek, S., 2003. Applications of artificial neural networks for patterning and predicting

aquatic insect species richness in running waters. Ecol. Model. 160, 265–280.
Park, Y.S., Chon, T.S., Kwak, I.S., Lek, S., 2004. Hierarchical community classification and assessment of aquatic ecosystems using artificial neural networks. Sci. Tot. Environ. 327, 105–122.
Pfankuch, D.J., 1975. Stream Reach Inventory and Channel Stability Evaluation. U.S. Department of Agriculture. Forest Service, Region 1, Missoula, MO, USA.
Robinson, C.T., Uehlinger, U., Hieber, M., 2001. Spatio-temporal variation in macroinvertebrate assemblages of glacial streams. Fresh. Biol. 46, 1663–1672.
Rossaro, B., Lencioni, V., 1999. Analysis of relationships between chironomid species (Diptera, Chironomidae) and environmental factors in an Alpine glacial stream system using a general linear model. Studi Trent. Sci. Nat., Acta Biol. 76, 17–27.
Steffan, A.W., 1971. Chironomid (Diptera) biocoenoses in Scandinavian glacier brooks. Can. Entomol. 103, 477–486.
Vesanto, J., Himberg, J., Alhoniemi, J., Parhankangas, J., 2000. SOM Toolbox for Matlab 5. Helsinki University of Technology.
Ward, J.V., 1994. Ecology of alpine streams. Fresh. Biol. 32, 277–294.