Un approccio alla valutazione della qualità ambientale ai sensi della Direttiva 2000/60/CE basato su metodi di Intelligenza Artificiale

Michele Scardi e Lorenzo Tancioni

Dipartimento di Biologia, Università di Roma 'Tor Vergata'

mscardi@mclink.it **tancioni@uniroma2.it**

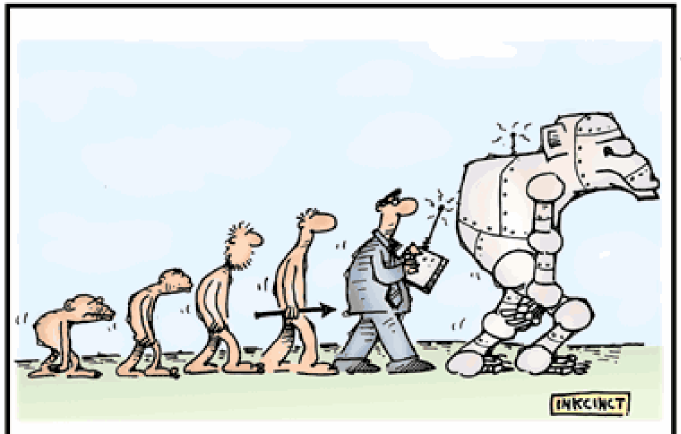
Integrità Biotica *vs.* Stato Ecologico

- **Integrità Biotica**: la capacità di sostenere e mantenere una comunità di organismi bilanciata, integrata, adattativa, con composizione in specie, diversità e organizzazione funzionale comparabile con quella degli ambienti naturali della regione (Karr & Dudley, 1981).

- **Stato Ecologico**: espressione della qualità della struttura e del funzionamento degli ecosistemi acquatici, associati ai corpi d'acqua superficiali... (WFD, 2000)

Available methods

- Biological indicators
- Biotic indices
- Multimetric (biotic) indices
- Expected vs. observed community structure
- Expert systems (A.I.)



The ascent of man.

Why not just a better index? (1)

- Biotic indices are the most obvious solution to the problem of ecological status assessment, but they are not the <u>only</u> solution.
- Most ecologists are not familiar with optimization techniques, so they stick with computationally simple methods.
- Developing and testing different methods or indices is certainly important: preserving diversity of methods is as important as preserving diversity of fish fauna.

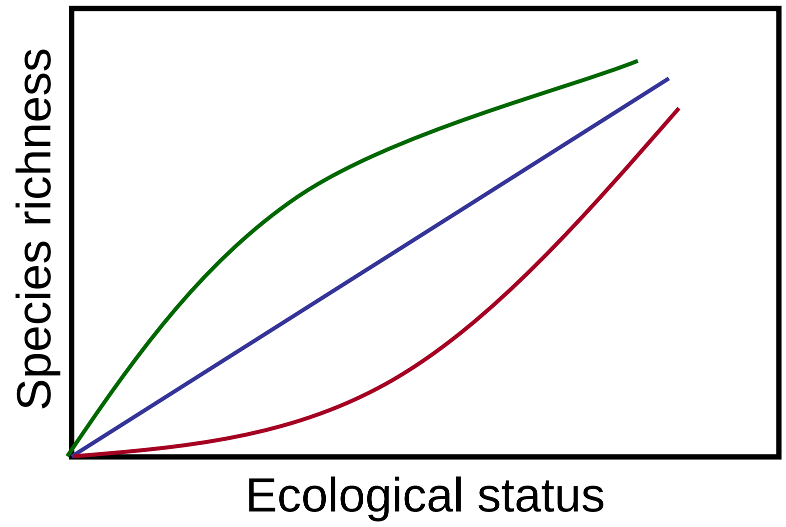
Why not just a better index? (2)

- Biotic indices (e.g. IBI, EFI, etc.) are based on metrics that are supposed to be linearly or at least monotonically related to ecosystem quality.
- From an ecological point of view, it is very clear that biotic responses are seldom linear and very often not monotonic (species abundance along an environmental gradient, intermediate disturbance hypothesis, effects of inter-species competition, etc.).

In most biotic indices:

species richness $\propto$ ecological status

(monotonic relationship)

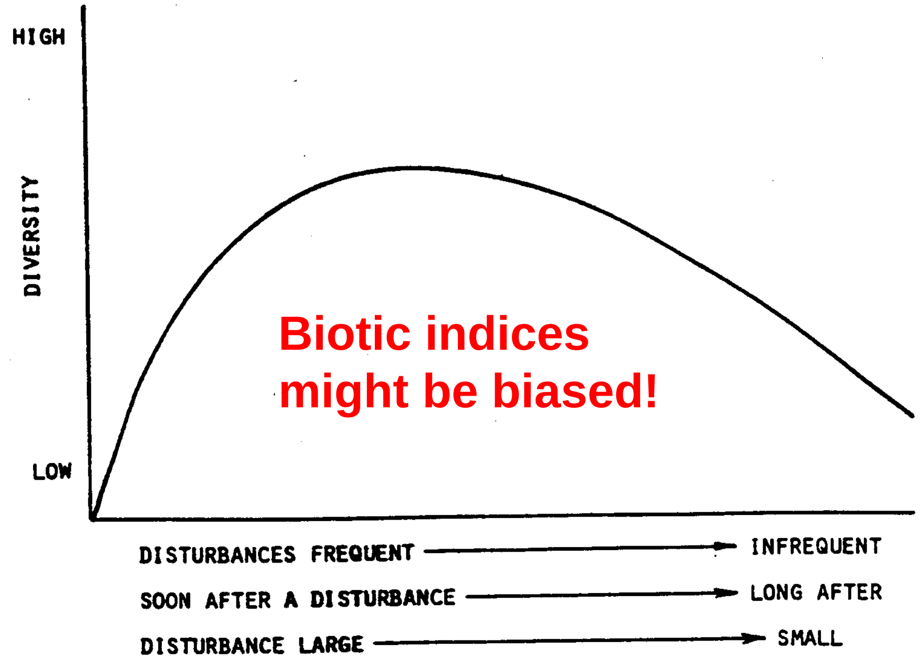


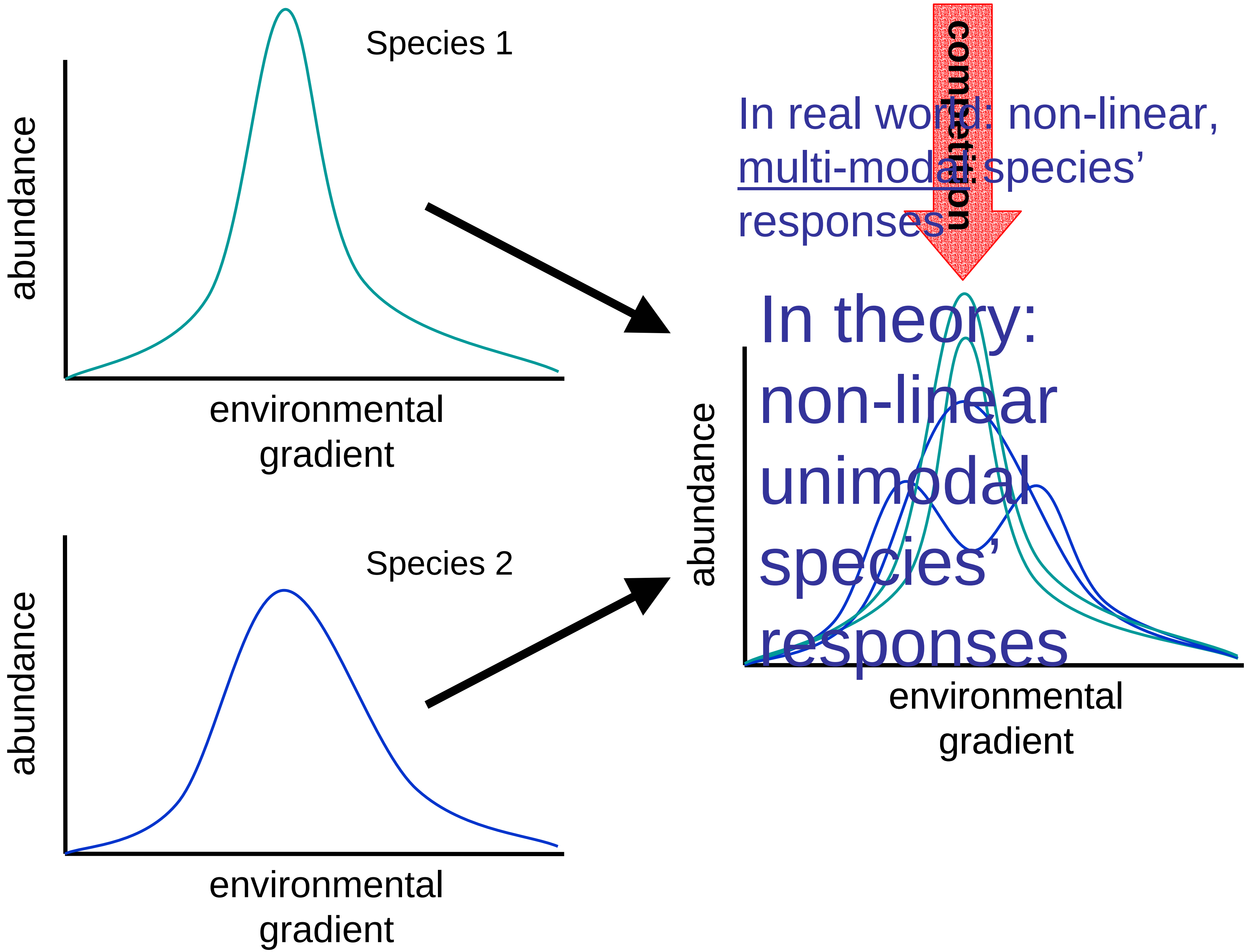
But:

intermediate disturbance hypotesis

Connell, J. H. (1978): Diversity in Tropical Rain Forests and Coral Reefs. Science 199: 1302-1310.

(non-monotonic relationship)

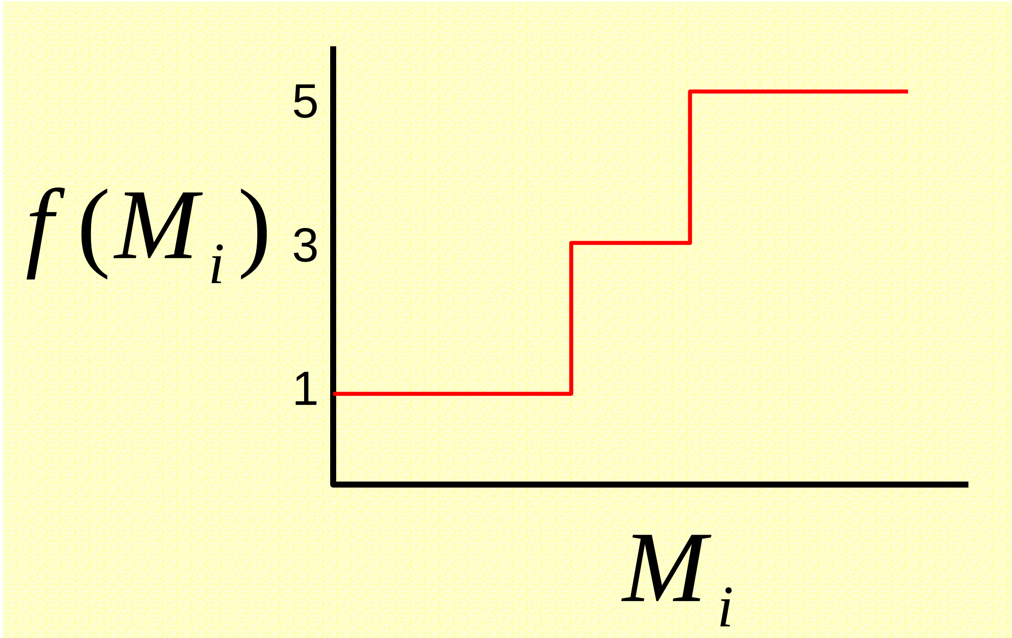




Species 1
abundance
environmental gradient
Species 2
abundance
environmental gradient
competition
In real world: non-linear, multi-modal species' responses
In theory: non-linear unimodal species' responses
abundance
environmental gradient

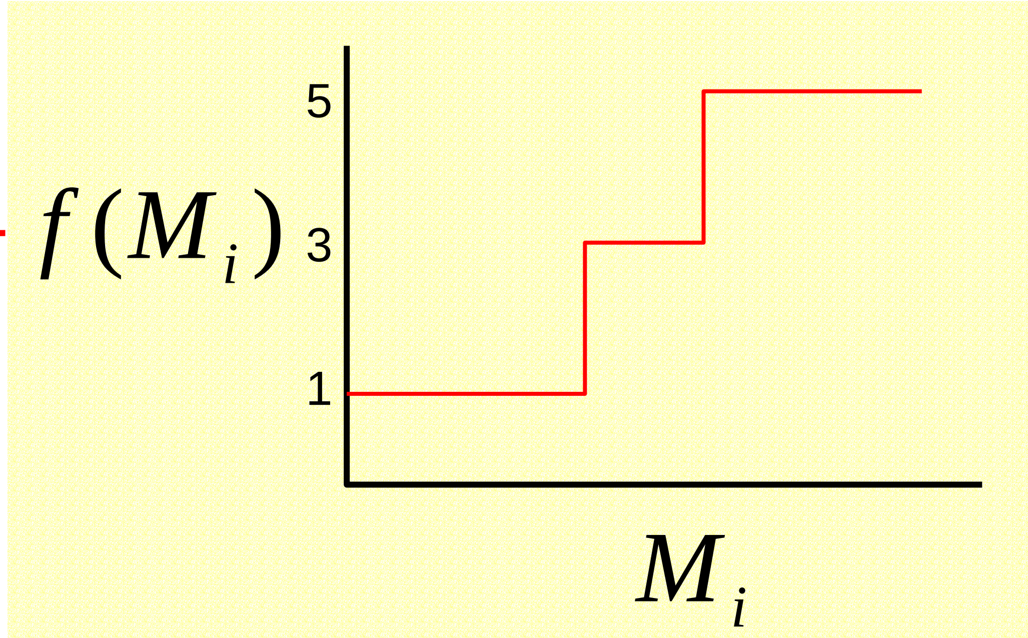
A multimetric index

$$I = \sum_{i=1}^{n} \boxed{f(M_i)}$$



$$\boxed{M_i \propto \text{Ecological Status}}$$

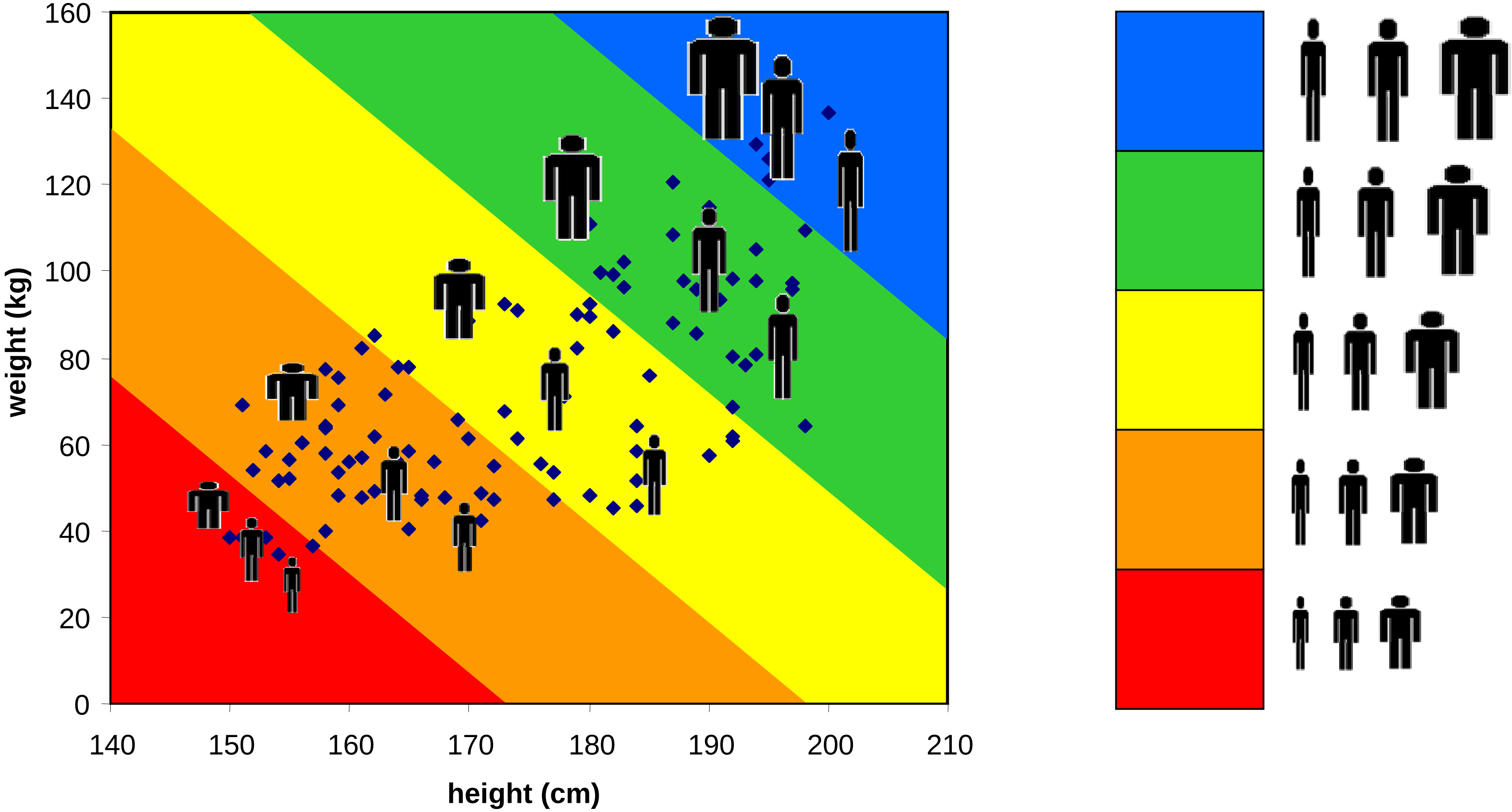
$$I = \sum_{i=1}^{n} \boxed{f(M_i)}$$



M_1				
	5	*6*	*8*	*10*
	3	*4*	*6*	*8*
	1	*2*	*4*	*6*
		1	**3**	**5**
			M_2	

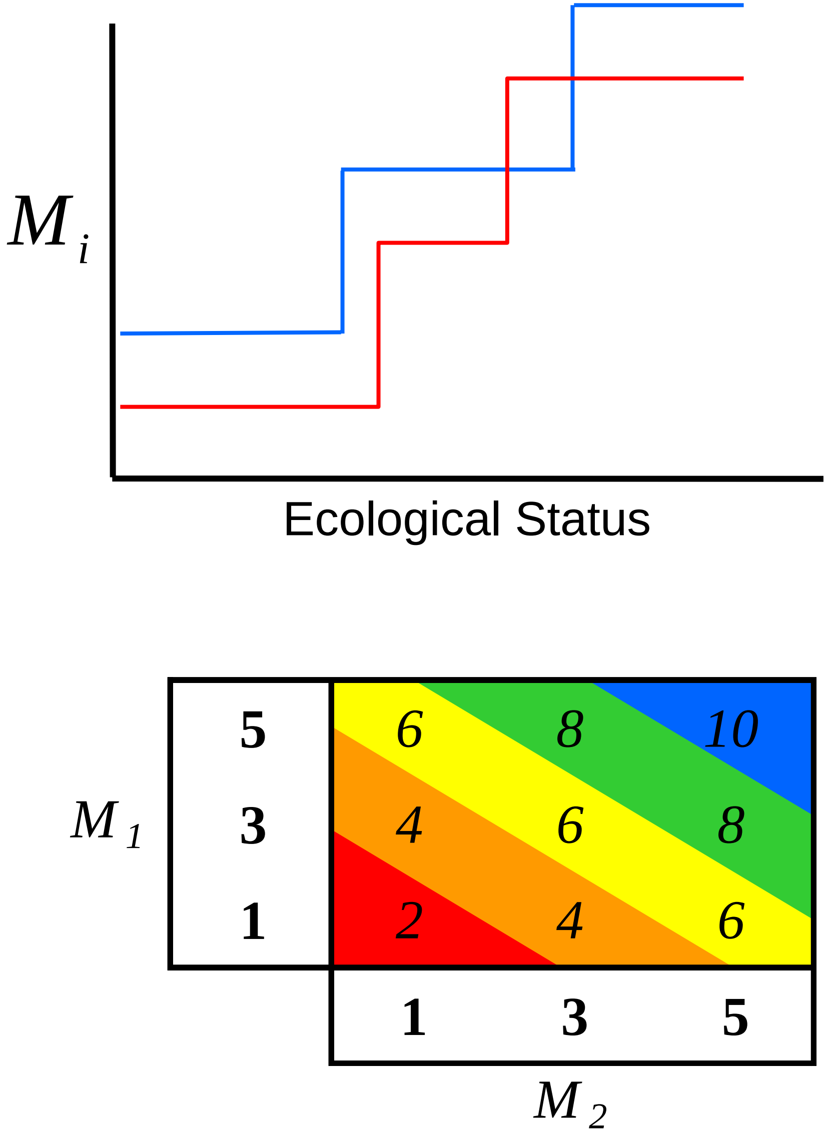
if $n = 2$

$$I = f(M_1) + f(M_2)$$

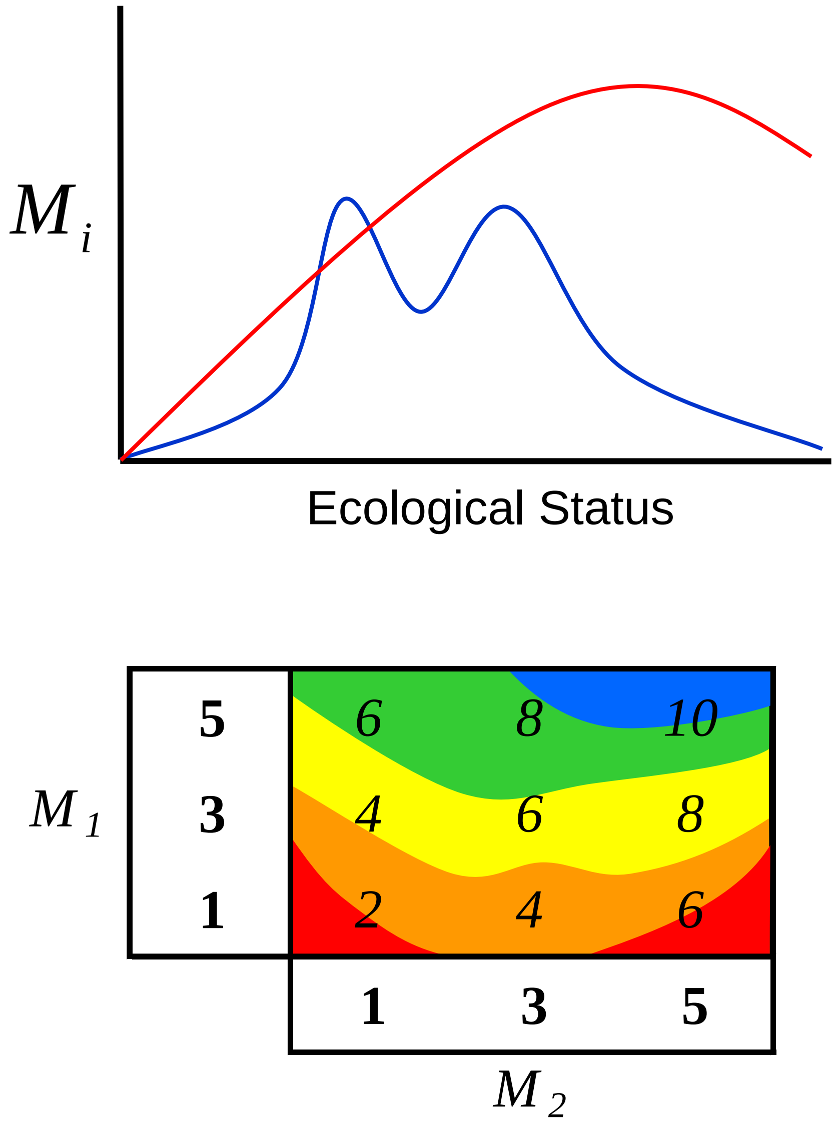


- Metrics depend on each other (correlation)
- Both metrics are causally linked to other variables (age, food availability, health, etc.)

In the multimetric world…



In the real world…



Why not just a better index? (3)

- Biotic indices produce scores, which in turn have to be interpreted and discretized (thresholds for high, good, moderate, etc.)
- This process is inherently subjective, as it is based on calibration procedures involving expert judgement
- “Ecological status” is not an emergent property of ecological data, so we must accept subjectivity!

Therefore…

- Ecological complexity requires more effective approaches (making it too simple today will not pay back tomorrow).
- Biotic indices are never optimized from a computational point of view (would you fit a regression line by hand?).
- Biotic indices only take into account a subset of the available information (metrics).
- Using all the available data and optimization techniques may provide better results…

A few steps back…

The roots of our approach

Machine Learning techniques for the implementation of the Water Framework Directive: a perspective from the **PAEQANN** project.

Scardi M.[1], Lek S.[2], Coste M.[3], Descy J.P.[4], Ector L.[5], Jorgensen S.[6], Knoflacher M.[7], Verdonschot P.[8]

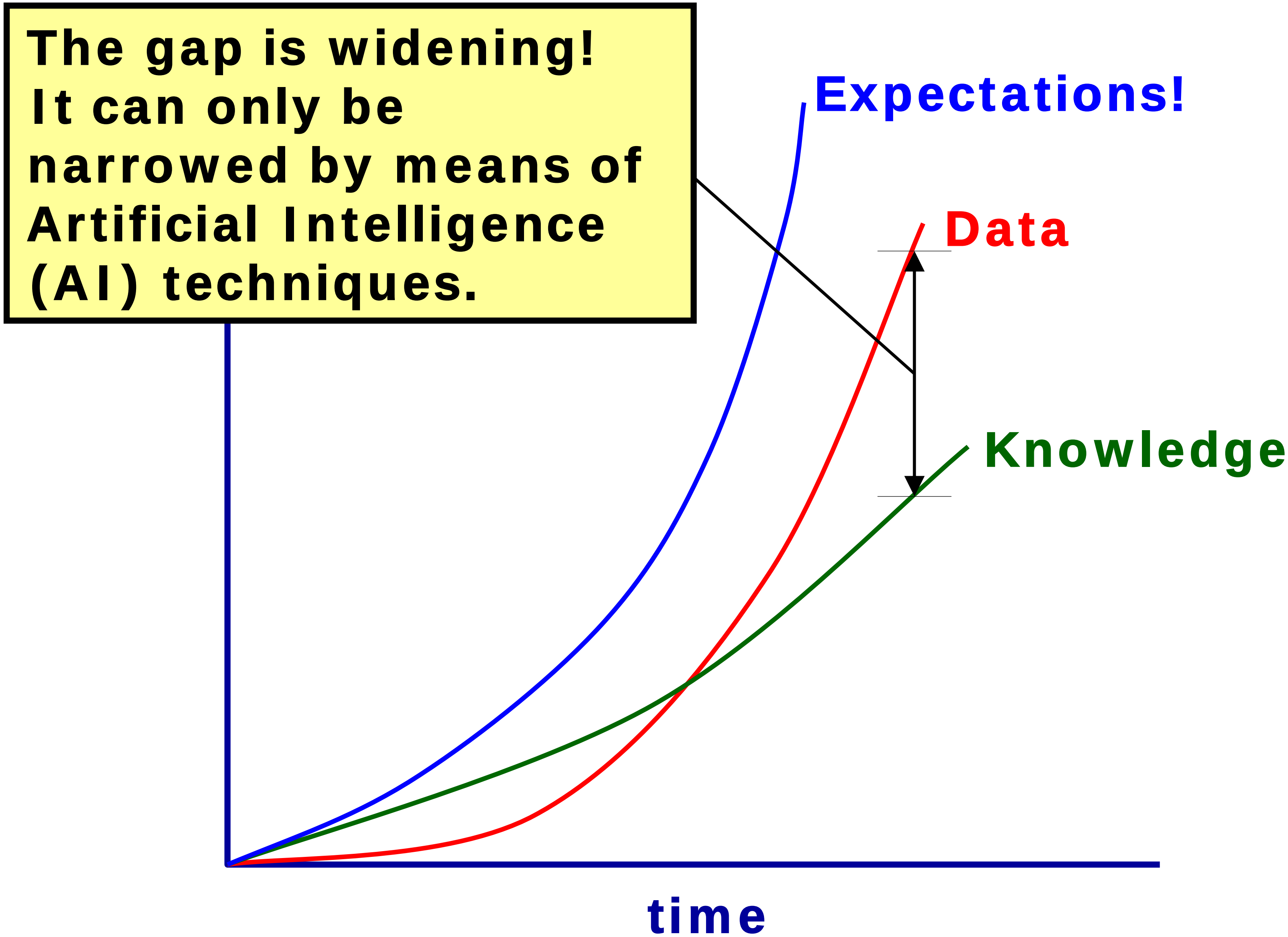
[1]UNIROMA2, Rome, Italy (email: mscardi@mclink.it)
[2]CESAC, Toulouse, France
[3]CEMAGREF, Cestas, France
[4]LFE, FUNDP, Namur, Belgium
[5]CRP-GL, Luxembourg
[6]DFH, University Copenhagen, Denmark
[7]ARCS, Seibersdorf, Austria
[8]ALTERRA, Wageningen, The Netherlands

Presented at the Third Symposium for European Freshwater Sciences (**SEFS3**)
Edinburgh, 13–17 July 2003

Machine Learning includes many different methods:

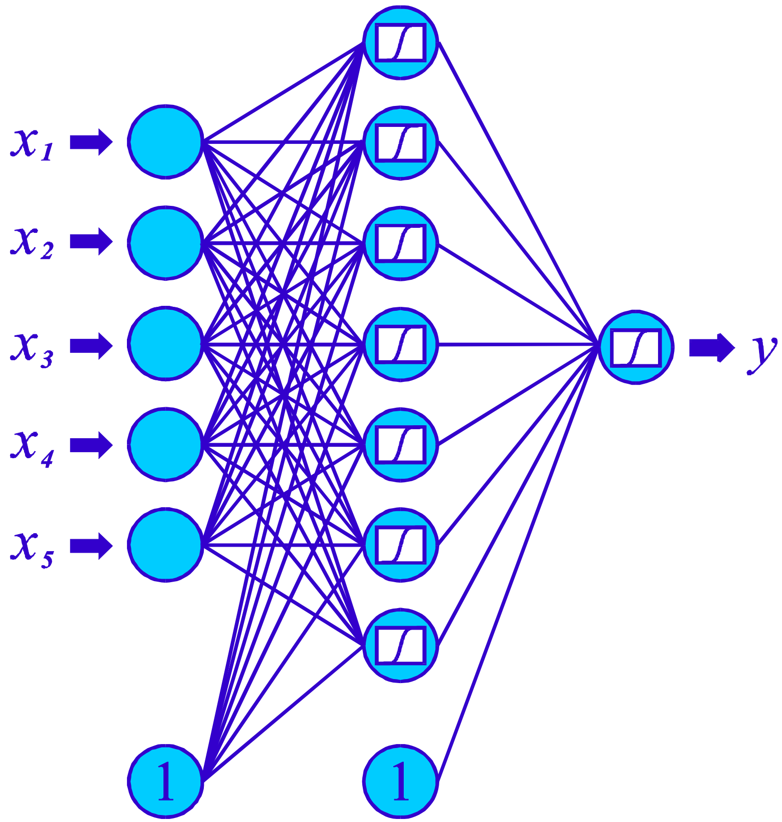
- Artificial Neural networks
- Genetic algorithms
- Classification and regression trees
- Fuzzy logic applications
- Cellular automata
- Agent based models
- Etc.

PAEQ**ANN** project focused on **A**rtificial **N**eural **N**etworks, even though other methods have been tested.

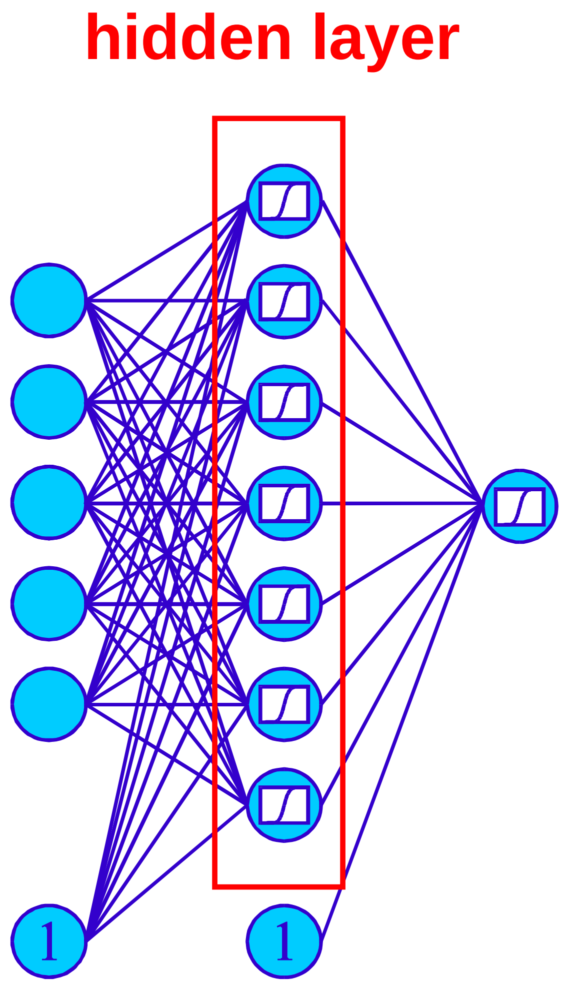

The gap is widening!
It can only be
narrowed by means of
Artificial Intelligence
(AI) techniques.
Expectations!
Data
Knowledge
time

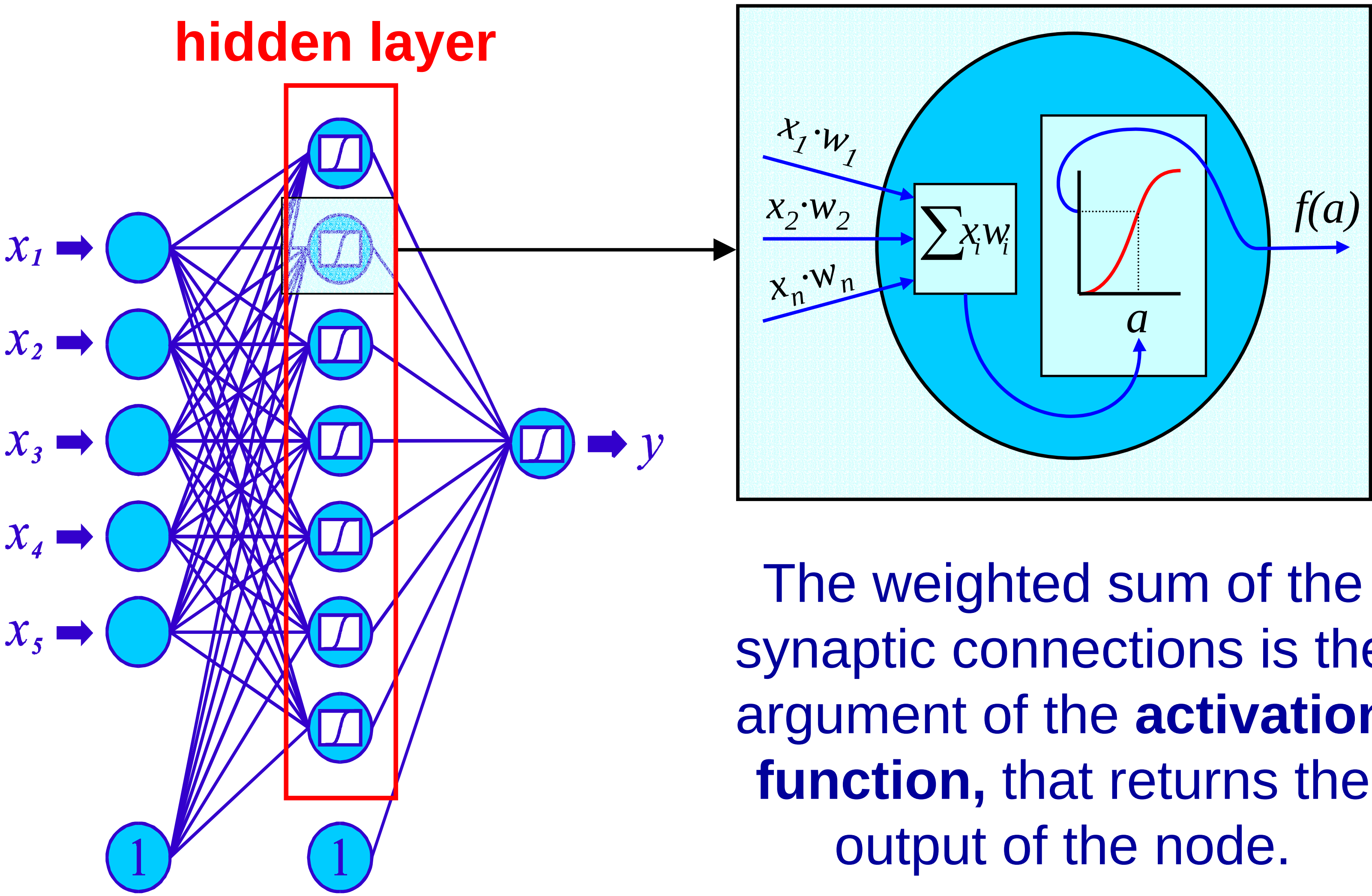
The most popular AI tool

"...a **neural network** is a system composed of many simple processing elements operating in parallel whose function is determined by network structure..."



The multilayer perceptron (the most popular NN)

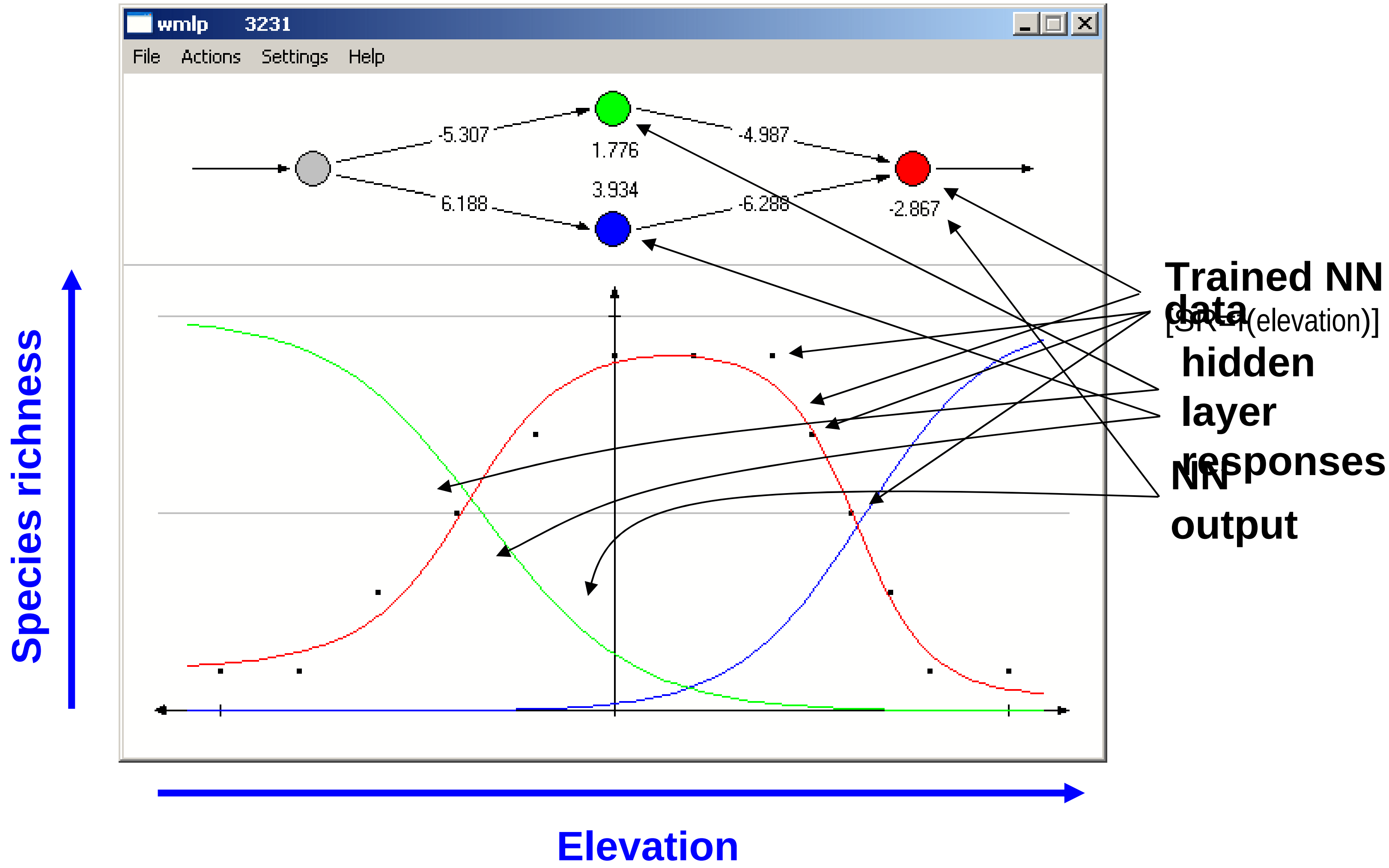


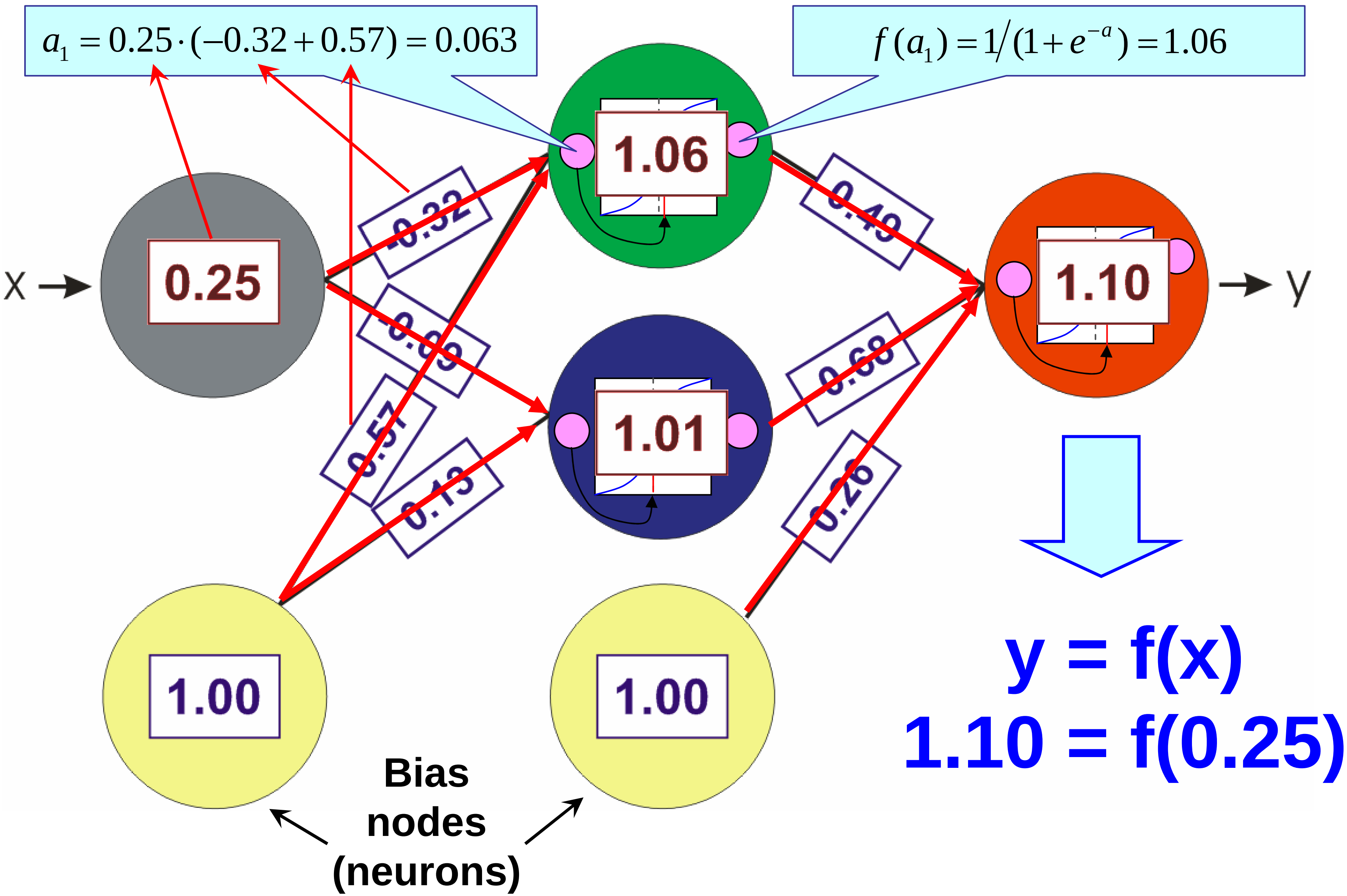


The weighted sum of the synaptic connections is the argument of the **activation function,** that returns the output of the node.

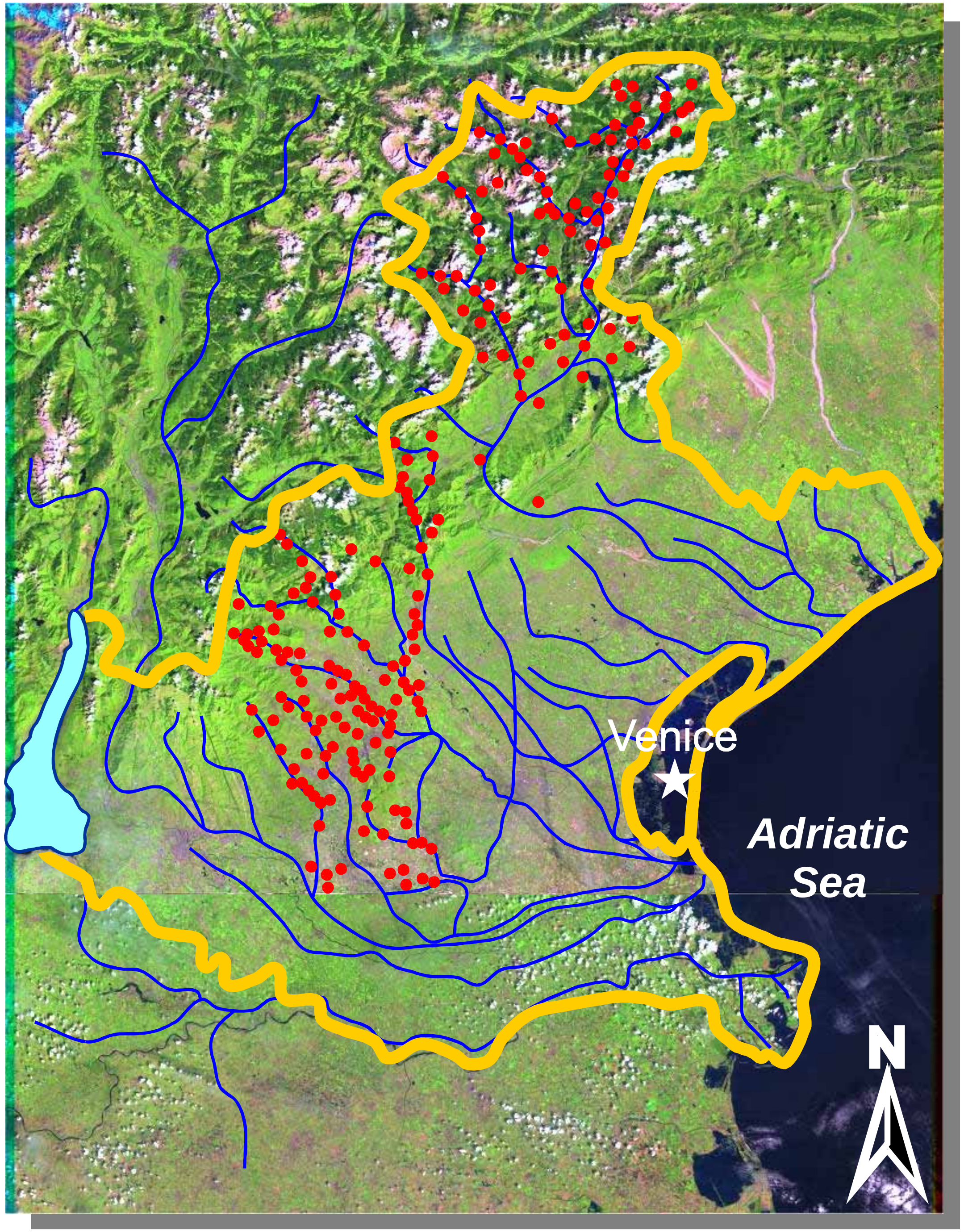
Example: $$f(a) = \frac{1}{1+e^{-a}}$$

Learning in a very simple NN





$a_1 = 0.25 \cdot (-0.32 + 0.57) = 0.063$
$f(a_1) = 1/(1+e^{-a}) = 1.06$
x
0.25
-0.32
0.57
0.13
1.06
1.01
0.68
0.26
1.10
y
1.00
1.00
Bias
nodes
(neurons)
y = f(x)
1.10 = f(0.25)



Data source:
Bioprogramm and
Aquaprogram

Predictive variables (NN inputs)

1	elevation (m)
2	mean depth (m)
3	runs (surface, %)
4	pools (surface, %)
5	riffles (surface, %)
6	mean width (m)
7	boulders (surface, %)
8	rocks and pebbles (surface, %)
9	gravel (surface, %)
10	sand (surface, %)
11	silt and clay (surface, %)
12	stream velocity (score, 0-5)
13	vegetation covering (surface, %)
14	shade (%)
15	anthropic disturbance (score, 0-4)
16	pH
17	conductivity (µS/cm)
18	gradient (%)
19	catchment area surface (km^2)
20	distance from source (km)

Fish community composition (NN outputs)

	Species name	English name
1	*Salmo (trutta) trutta (Linnaeus, 1758)*	Sea Trout
2	*Leuciscus cephalus (Linnaeus, 1758)*	Chub
3	*Padogobius martensii (Günther, 1861)*	*(Italian name: Ghiozzo di fiume)*
4	*Scardinius erythrophthalmus (Linnaeus, 1758)*	Rudd
5	*Esox lucius (Linnaeus, 1758)*	European Pike
6	*Rutilus erythrophthalmus (Zerunian, 1982)*	*(Italian name: Triotto)*
7	*Alburnus alburnus alborella (De Filippi, 1844)*	Bleak
8	*Cottus gobio (Linnaeus, 1756)*	Bullhead
9	*Tinca tinca (Linnaeus, 1758)*	Tench
10	*Cobitis taenia (Linnaeus, 1758)*	Spined loach
11	*Phoxinus phoxinus (Linnaeus, 1758)*	Minnow
12	*Anguilla anguilla (Linnaeus, 1758)*	European Eel
13	*Orsinigobius punctatissimus (Canestrini, 1864)*	*(Italian name: Panzarolo)*
14	*Salmo (trutta) marmoratus (Cuvier, 1817)*	Marble Trout
15	*Sabanejewia larvata (DeFilippi, 1859)*	Italian Loach
16	*Ictalurus melas (Rafinesque, 1820)*	Black Bullhead
17	*Lepomis gibbosus (Linnaeus, 1758)*	Pumpkinseed
18	*Barbus plebejus (Bonaparte, 1839)*	Italian Barbel
19	*Chondrostoma genei (Bonaparte, 1839)*	South Europe Nase
20	*Gasterosteus aculeatus (Linnaeus, 1758)*	Three-spined Stickleback
21	*Carassius carassius (Linnaeus, 1758)*	Crucian Carp
22	*Gobio gobio (Linnaeus, 1758)*	Gudgeon
23	*Leuciscus souffia (Risso, 1826)*	Blageon
24	*Thymallus thymallus (Linnaeus, 1758)*	Grayling
25	*Lampetra planeri (Bloch, 1784)*	Brook Lamprey
26	*Gambusia holbrooki (Girard, 1859)*	Eastern mosquitofish
27	*Barbus meridionalis*	Meriditerranean Barbel
28	*Micropterus salmoides (Lacepede, 1802)*	Large-Mouthed Bass
29	*Perca fluviatilis (Linnaeus, 1758)*	Perch
30	*Abramis brama (Linnaeus, 1758)*	Common Bream
31	*Cyprinus carpio (Linnaeus, 1758)*	Common Carp
32	*Salvelinus fontinalis M.*	Brook Char
33	*Oncorhynchus mykiss (Walbaum, 1792)*	Rainbow Trout
34	*Salmo (trutta) hybr. trutta/marmoratus*	Sea Trout - Marble Trout hybrid

Neural network model structure: 20-17-32

- 264 patterns (i.e. samples) $\left\{ \begin{array}{l} \text{131 training patterns} \\ \text{66 validation patterns} \\ \text{67 test patterns} \end{array} \right.$

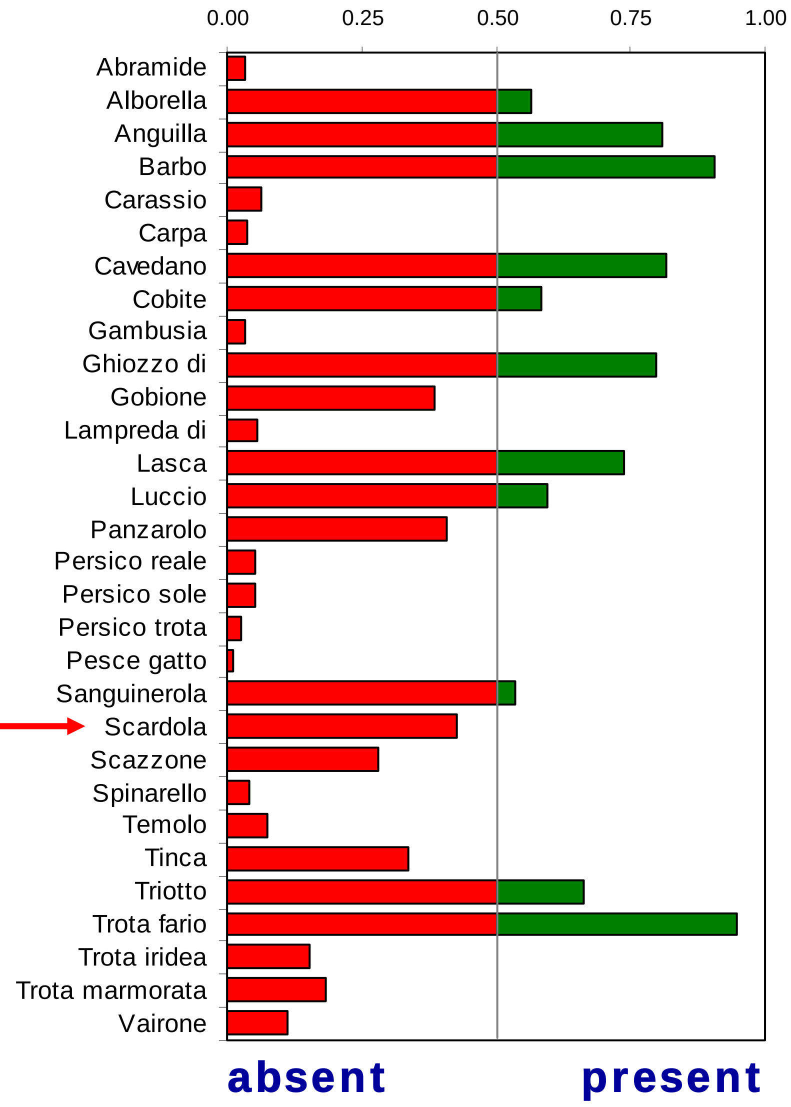
- 20 predictive environmental variables

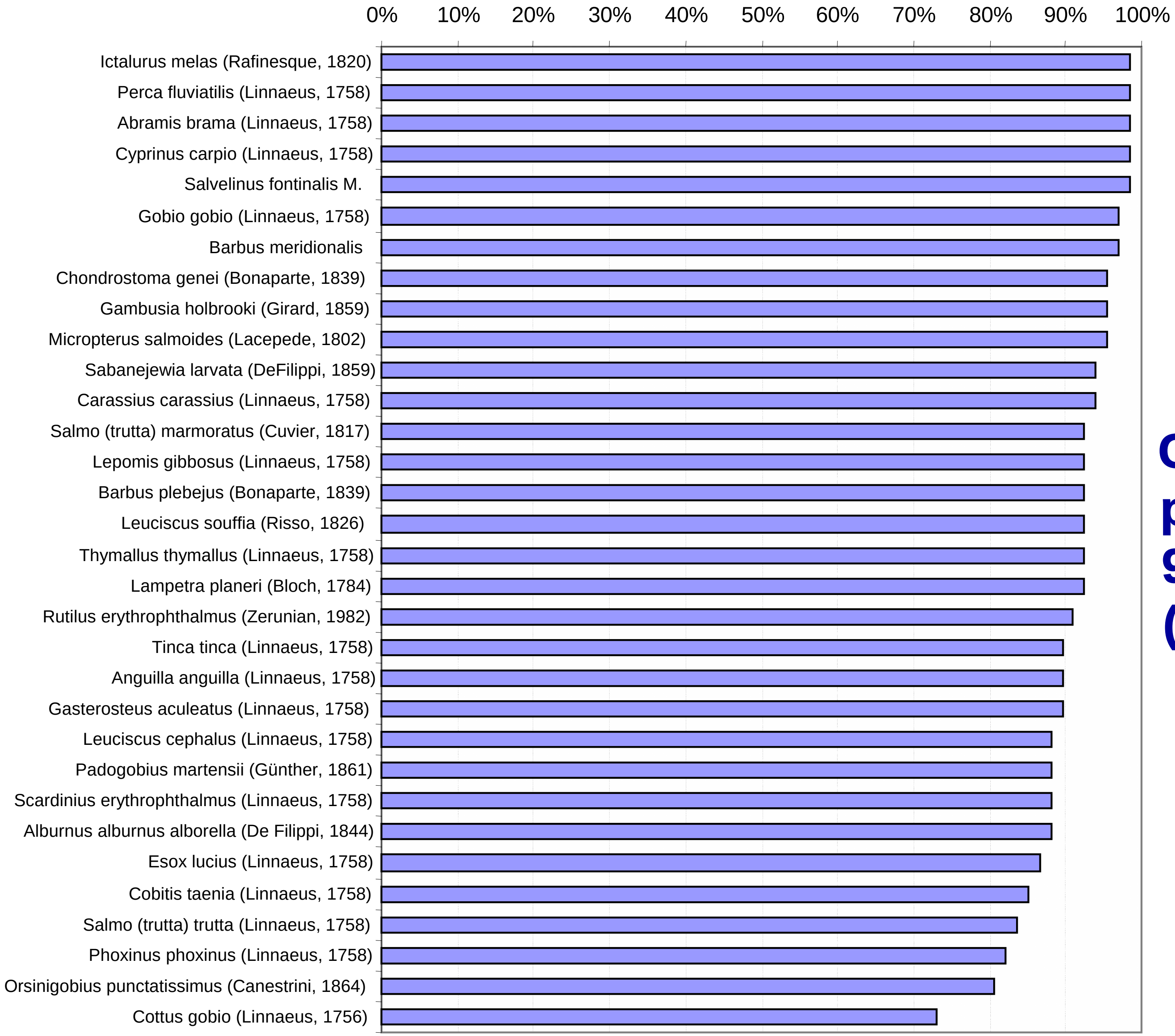
- 32 species (binary data)

- NN training:error back-propagation algorithm
  with early stopping based on
  validation set mean square error (MSE)

An example of NN output

Taxon	NN output	>0.5?	osservato	ok?
Abramide	0.032	0	0	1
Alborella	0.565	1	1	1
Anguilla	0.807	1	1	1
Barbo	0.905	1	1	1
Carassio	0.064	0	0	1
Carpa	0.038	0	0	1
Cavedano	0.817	1	1	1
Cobite	0.584	1	1	1
Gambusia	0.036	0	0	1
Ghiozzo di fiume	0.798	1	1	1
Gobione	0.384	0	0	1
Lampreda di ruscello	0.057	0	0	1
Lasca	0.739	1	1	1
Luccio	0.597	1	1	1
Panzarolo	0.407	0	0	1
Persico reale	0.053	0	0	1
Persico sole	0.054	0	0	1
Persico trota	0.026	0	0	1
Pesce gatto	0.011	0	0	1
Sanguinerola	0.536	1	1	1
Scardola	0.427	0	1	0
Scazzone	0.281	0	0	1
Spinarello	0.040	0	0	1
Temolo	0.074	0	0	1
Tinca	0.337	0	0	1
Triotto	0.663	1	1	1
Trota fario	0.948	1	1	1
Trota iridea	0.154	0	0	1
Trota marmorata	0.182	0	0	1
Vairone	0.111	0	0	1
		previsioni corrette:		**29 su 30**





Correct predictions: 91.6 % (overall)

The K statistics

		model output	
		presence	*absence*
target	*presence*	**1 - 1**	**1 - 0**
	absence	**0 - 1**	**0 - 0**

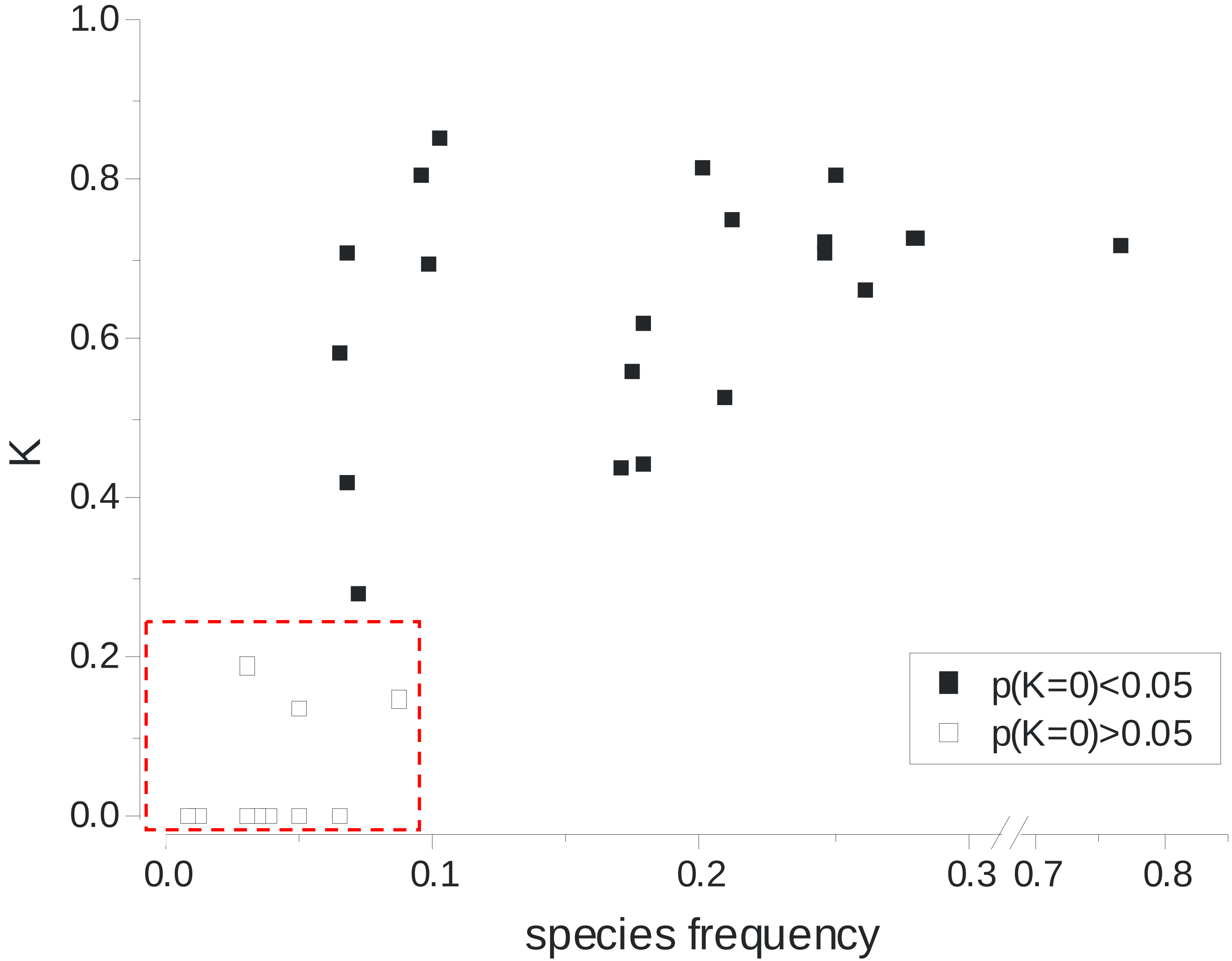
H_0 = modeled and observed data are independent of each other

$$K = \frac{Oa - Ea}{N - Ea}$$

Oa = observed count of matches

Ea = expected count of matches

N = total number of cases



1.0
0.8
0.6
0.4
0.2
0.0
K
0.0
0.1
0.2
0.3
0.7
0.8
species frequency
p(K=0)<0.05
p(K=0)>0.05

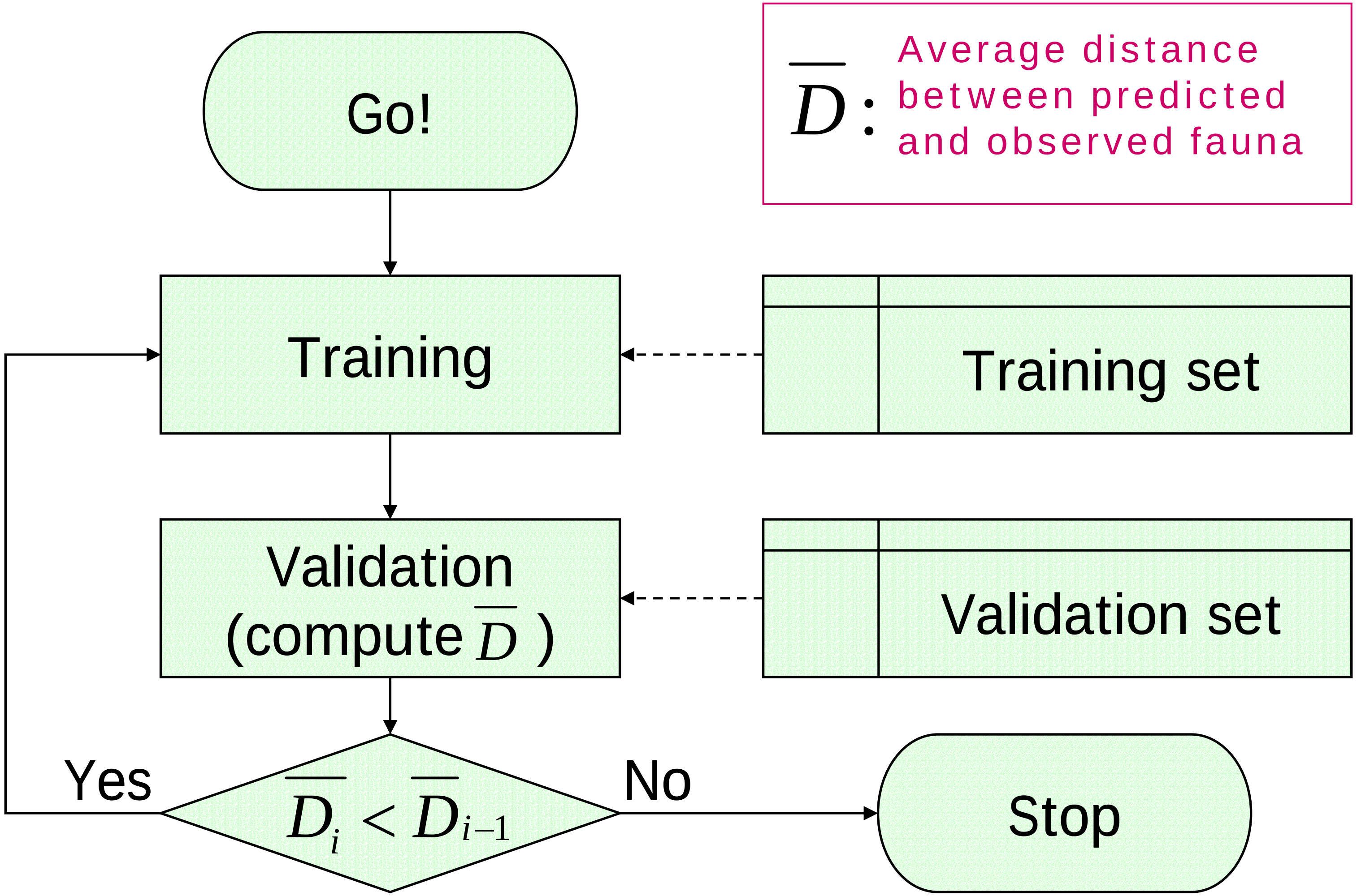
How to get more relevant info

- Regular, homogeneous sampling strategies are useful, but not suffient
- Sampling must take into account more than a single spatial scale
- Exploit alternate sources of info (experts, historical records, fisheries, etc.)
- Keep collecting more and more data!

How to improve NN learning

- Exclude species, taxa, classes, coenotypes, etc. whose frequency in training, validation and test sets is very high or very low (no info there!)
- Use alternate criteria for error measurements (no MSE, please!)
- Use ecological rules to constrain NN learning

EBP NN training based on ecological distance



Measuring ecological distance

- Both presence and absence of species are relevant, so we need a symmetrical index.
- E.g. Rogers & Tanimoto dissimilarity:

$$D = 1 - \frac{a+d}{a+2b+2c+d}$$

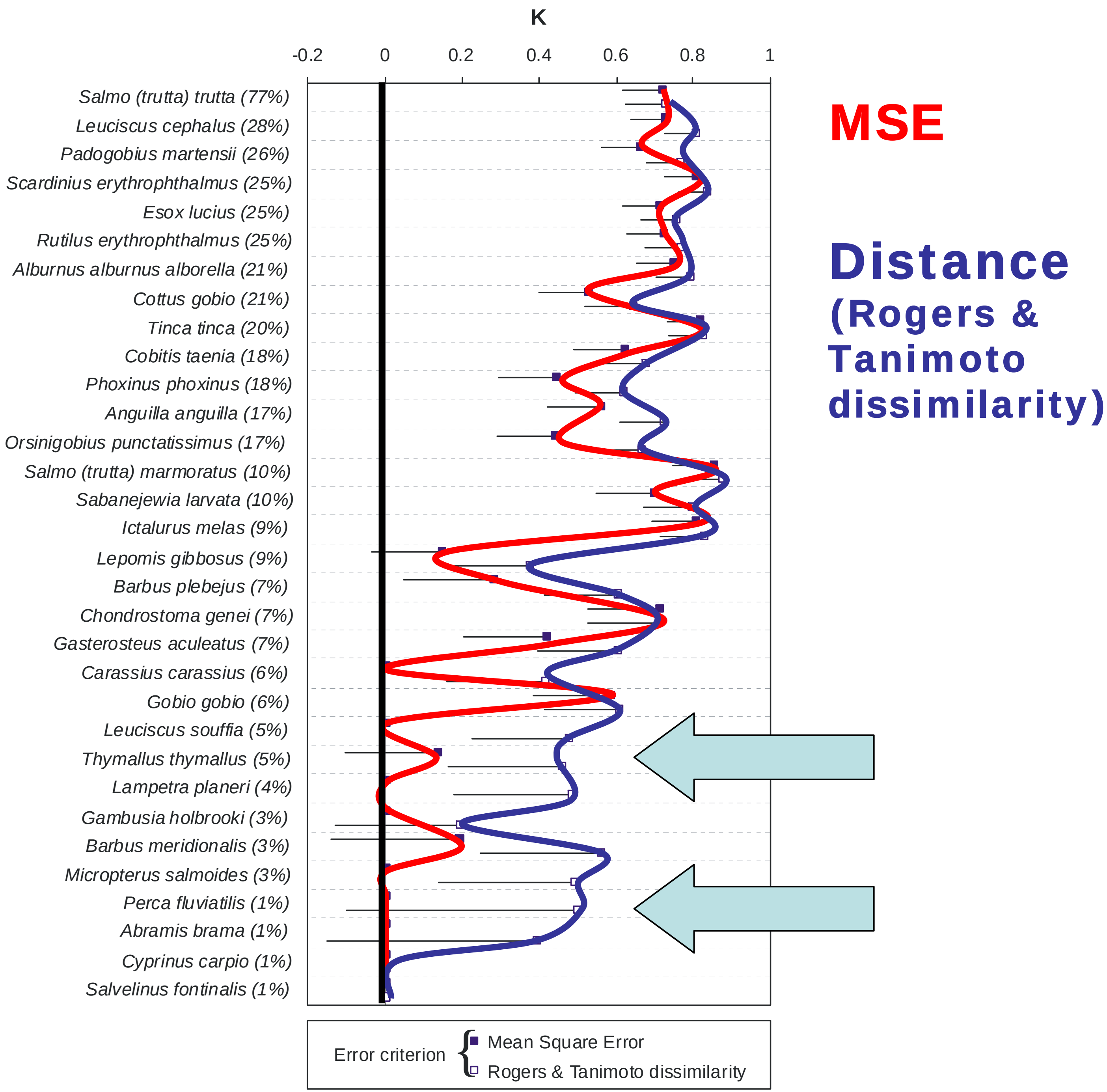
N.B. Discordancies have a double weight in this index (useful in case absence data are much more frequent than presence data or vice versa)

Distance vs. MSE NN training

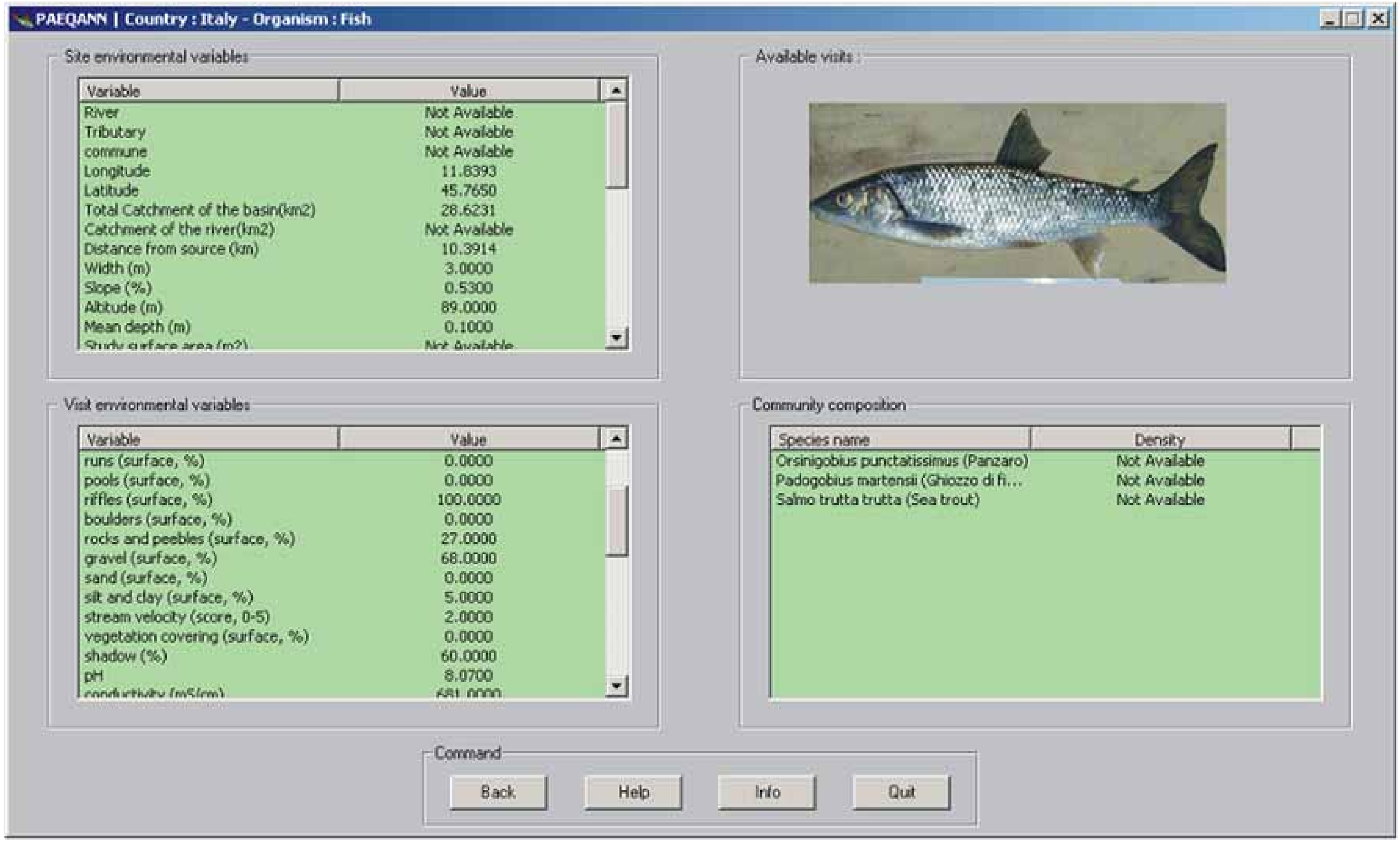
Correct predictions	*Distance*	*MSE*
overall	**94.4%**	**93.1%**
training+validation	**95.4%**	**93.6%**
test	**91.8%**	**91.6%**

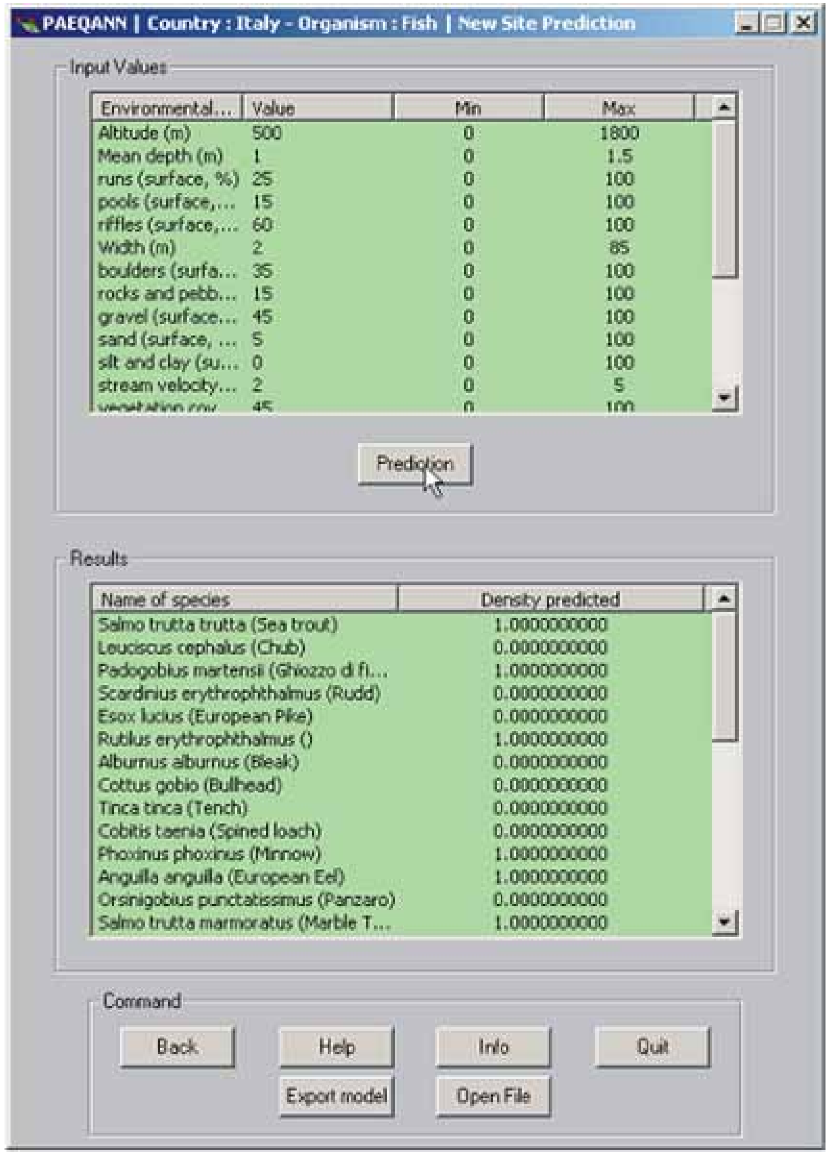
Distance training: only the two rarest species are always absent in NN predictions (2 known records each)

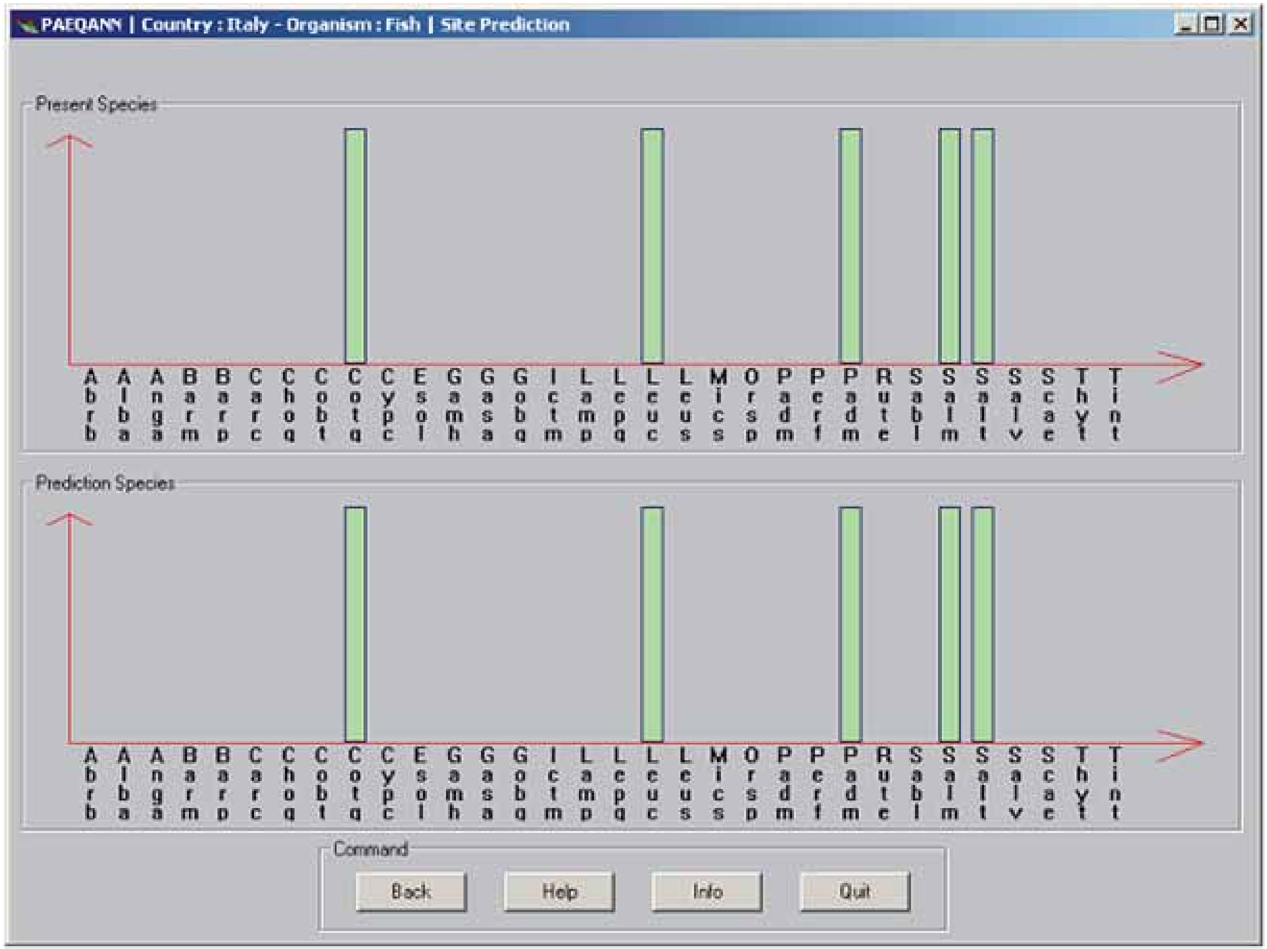
MSE training: nine rare species are always absent in NN predictions

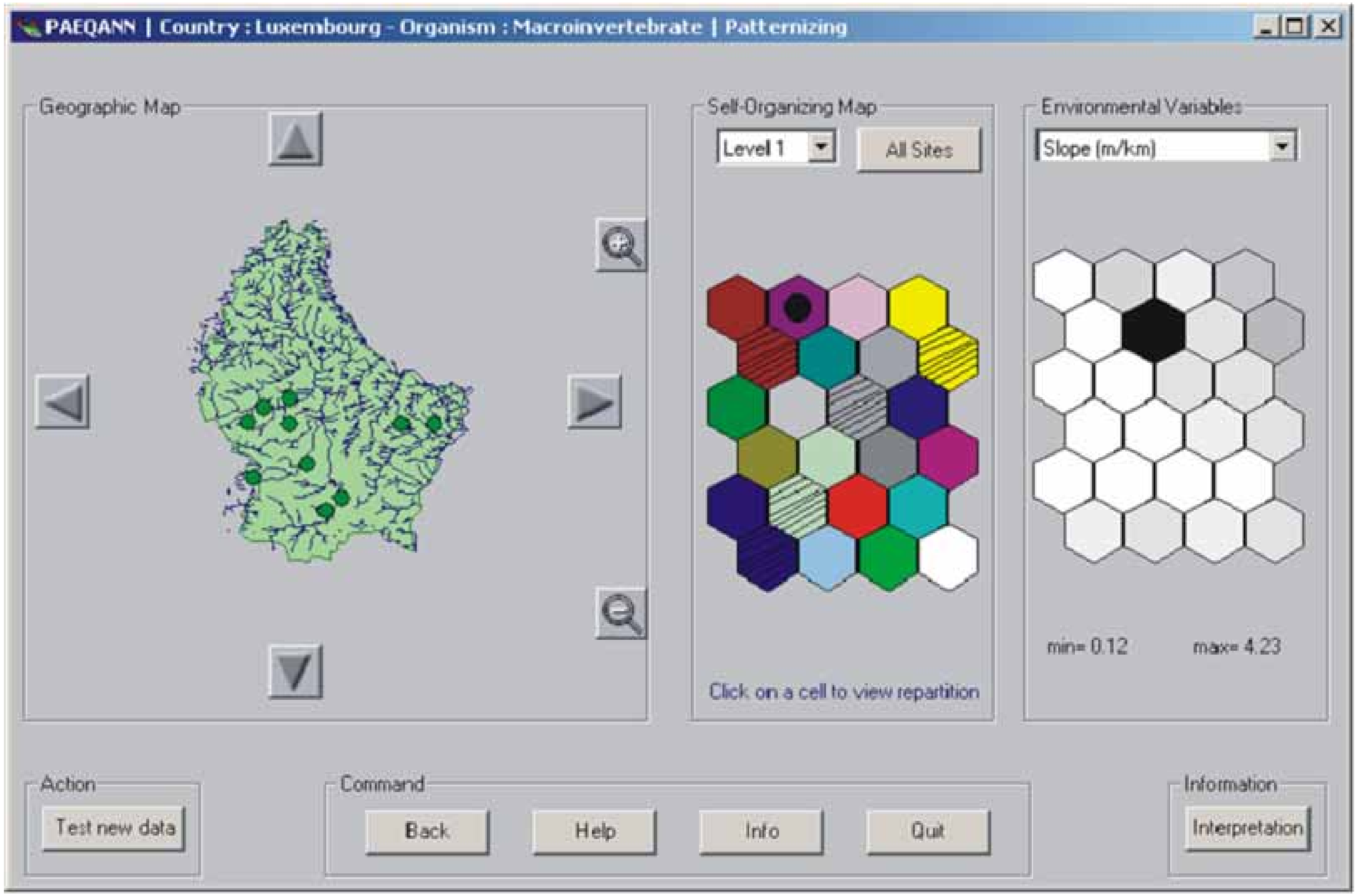

K
-0.2
0
0.2
0.4
0.6
0.8
1
Salmo (trutta) trutta (77%)
Leuciscus cephalus (28%)
Padogobius martensii (26%)
Scardinius erythrophthalmus (25%)
Esox lucius (25%)
Rutilus erythrophthalmus (25%)
Alburnus alburnus alborella (21%)
Cottus gobio (21%)
Tinca tinca (20%)
Cobitis taenia (18%)
Phoxinus phoxinus (18%)
Anguilla anguilla (17%)
Orsinigobius punctatissimus (17%)
Salmo (trutta) marmoratus (10%)
Sabanejewia larvata (10%)
Ictalurus melas (9%)
Lepomis gibbosus (9%)
Barbus plebejus (7%)
Chondrostoma genei (7%)
Gasterosteus aculeatus (7%)
Carassius carassius (6%)
Gobio gobio (6%)
Leuciscus souffia (5%)
Thymallus thymallus (5%)
Lampetra planeri (4%)
Gambusia holbrooki (3%)
Barbus meridionalis (3%)
Micropterus salmoides (3%)
Perca fluviatilis (1%)
Abramis brama (1%)
Cyprinus carpio (1%)
Salvelinus fontinalis (1%)
Error criterion
Mean Square Error
Rogers & Tanimoto dissimilarity
MSE
Distance
(Rogers &
Tanimoto
dissimilarity)

http://aquaeco.ups-tlse.fr


PAEQANN | Country : Italy - Organism : Fish
Site environmental variables
Variable
Value
River
Not Available
Tributary
Not Available
commune
Not Available
Longitude
11.8393
Latitude
45.7650
Total Catchment of the basin(km2)
28.6231
Catchment of the river(km2)
Not Available
Distance from source (km)
10.3914
Width (m)
3.0000
Slope (%)
0.5300
Altitude (m)
89.0000
Mean depth (m)
0.1000
Available visits :
Visit environmental variables
Variable
Value
runs (surface, %)
0.0000
pools (surface, %)
0.0000
riffles (surface, %)
100.0000
boulders (surface, %)
0.0000
rocks and peebles (surface, %)
27.0000
gravel (surface, %)
68.0000
sand (surface, %)
0.0000
silt and clay (surface, %)
5.0000
stream velocity (score, 0-5)
2.0000
vegetation covering (surface, %)
0.0000
shadow (%)
60.0000
pH
8.0700
Community composition
Species name
Density
Orsinigobius punctatissimus (Panzaro)
Not Available
Padogobius martensii (Ghiozzo di fi...
Not Available
Salmo trutta trutta (Sea trout)
Not Available
Command
Back
Help
Info
Quit


PAEQANN | Country : Italy - Organism : Fish | New Site Prediction
Input Values
Environmental... Value Min Max
Altitude (m) 500 0 1800
Mean depth (m) 1 0 1.5
runs (surface, %) 25 0 100
pools (surface,... 15 0 100
riffles (surface,... 60 0 100
Width (m) 2 0 85
boulders (surfa... 35 0 100
rocks and pebb... 15 0 100
gravel (surface... 45 0 100
sand (surface, ... 5 0 100
silt and clay (su... 0 0 100
stream velocity... 2 0 5
Prediction
Results
Name of species Density predicted
Salmo trutta trutta (Sea trout) 1.0000000000
Leuciscus cephalus (Chub) 0.0000000000
Padogobius martensii (Ghiozzo di fi... 1.0000000000
Scardinius erythrophthalmus (Rudd) 0.0000000000
Esox lucius (European Pike) 0.0000000000
Rutilus erythrophthalmus () 1.0000000000
Alburnus alburnus (Bleak) 0.0000000000
Cottus gobio (Bullhead) 0.0000000000
Tinca tinca (Tench) 0.0000000000
Cobitis taenia (Spined loach) 0.0000000000
Phoxinus phoxinus (Minnow) 1.0000000000
Anguilla anguilla (European Eel) 1.0000000000
Orsinigobius punctatissimus (Panzaro) 0.0000000000
Salmo trutta marmoratus (Marble T... 1.0000000000
Command
Back
Help
Info
Quit
Export model
Open File


PAEQANN | Country : Italy - Organism : Fish | Site Prediction
Present Species
Prediction Species
Command
Back
Help
Info
Quit


PAEQANN | Country : Luxembourg - Organism : Macroinvertebrate | Patternizing
Geographic Map
Self-Organizing Map
Level 1
All Sites
Click on a cell to view repartition
Environmental Variables
Slope (m/km)
min= 0.12
max= 4.23
Action
Test new data
Command
Back
Help
Info
Quit
Information
Interpretation

From a neural network to the WFD

Predicted fish assemblage structure

+

Observed fish assemblage structure

+

A suitable similarity or distance coefficient

=

Estimate of the 'ecological status' (*sensu* WFD)

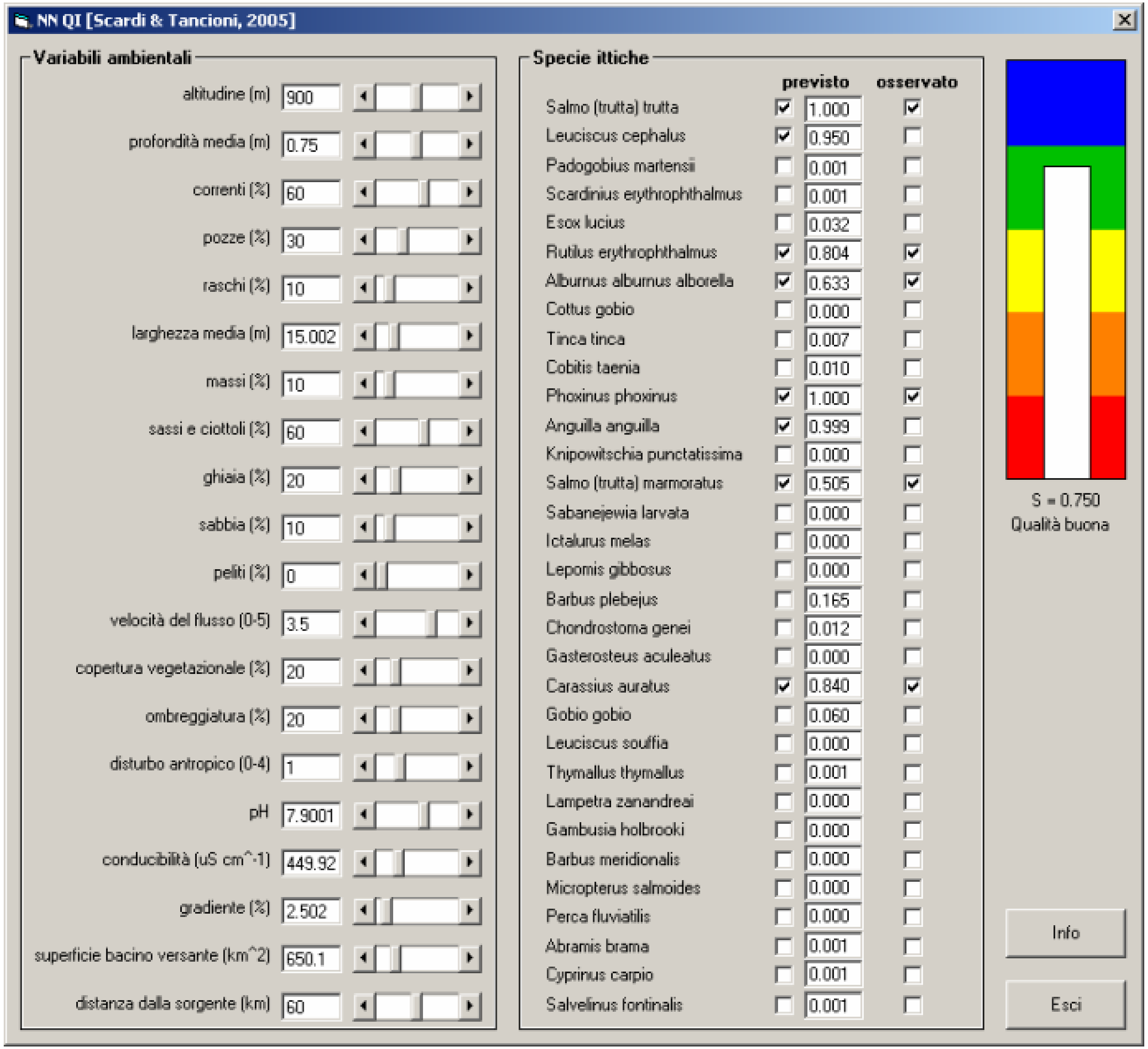
How to measure deviation from expected fish fauna

cfr. Moss *et al*. (1987) O/E

Sokal & Michener (1958) $$S_{jk} = \frac{a+d}{a+b+c+d}$$

Rogers & Tanimoto (1960) $$S_{jk} = \frac{a+d}{a+2b+2c+d}$$

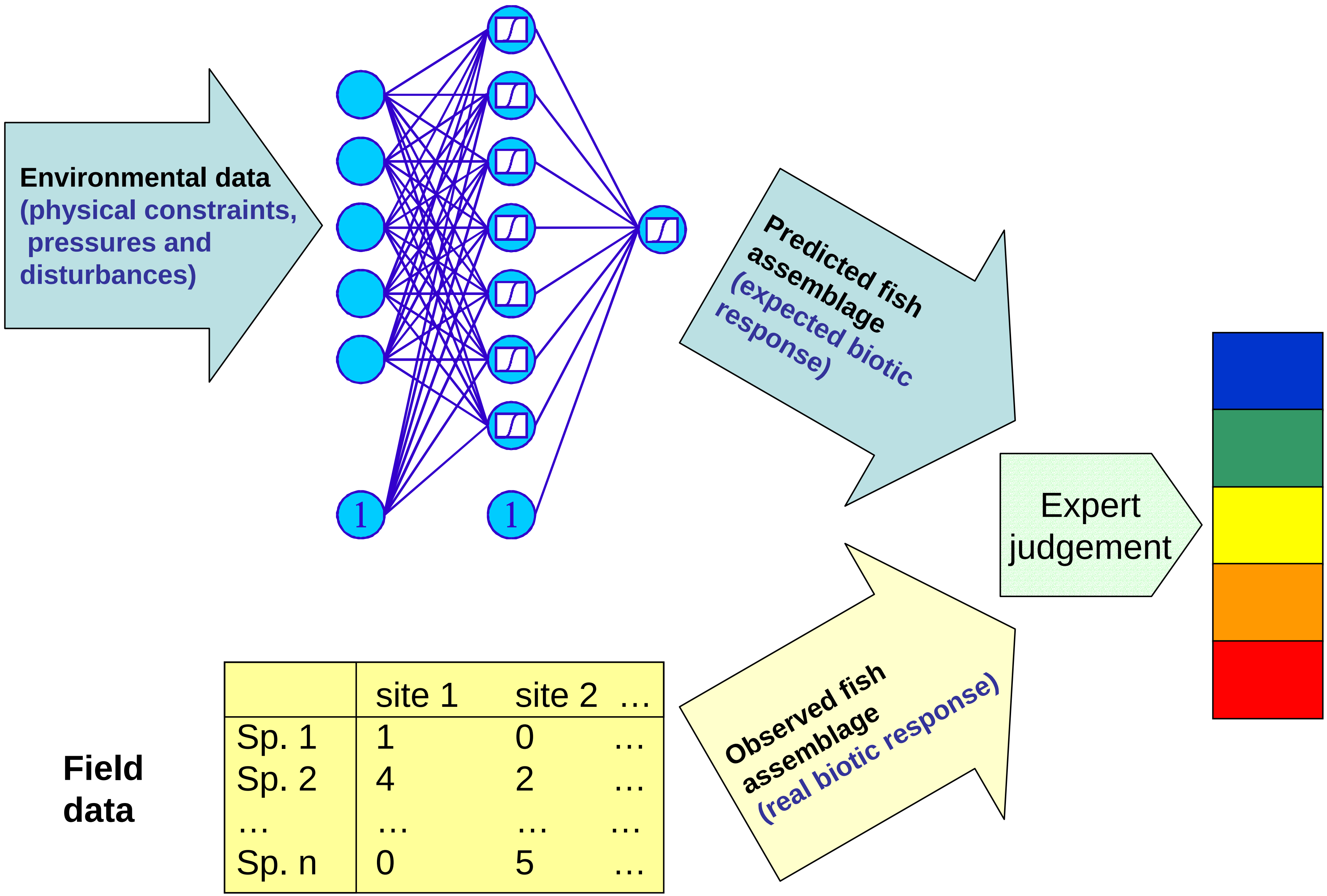
Jaccard (1900) $$S_{jk} = \frac{a}{a+b+c}$$



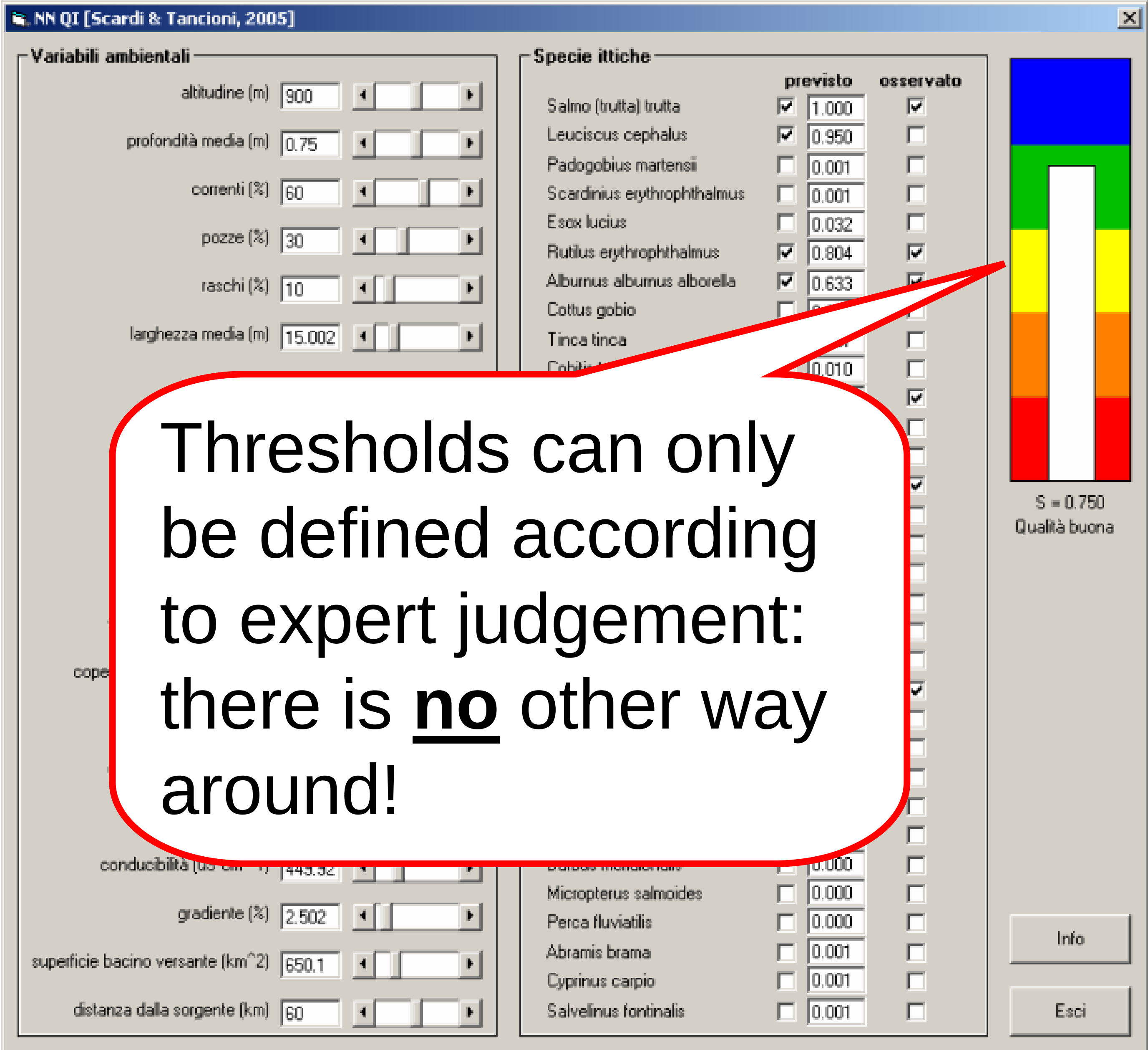
http://www.mare-net.com/mscardi/work/wfd/nnqi.htm

A conventional approach

Expert judgement is only used for defining (*ex post*) thresholds in the final score

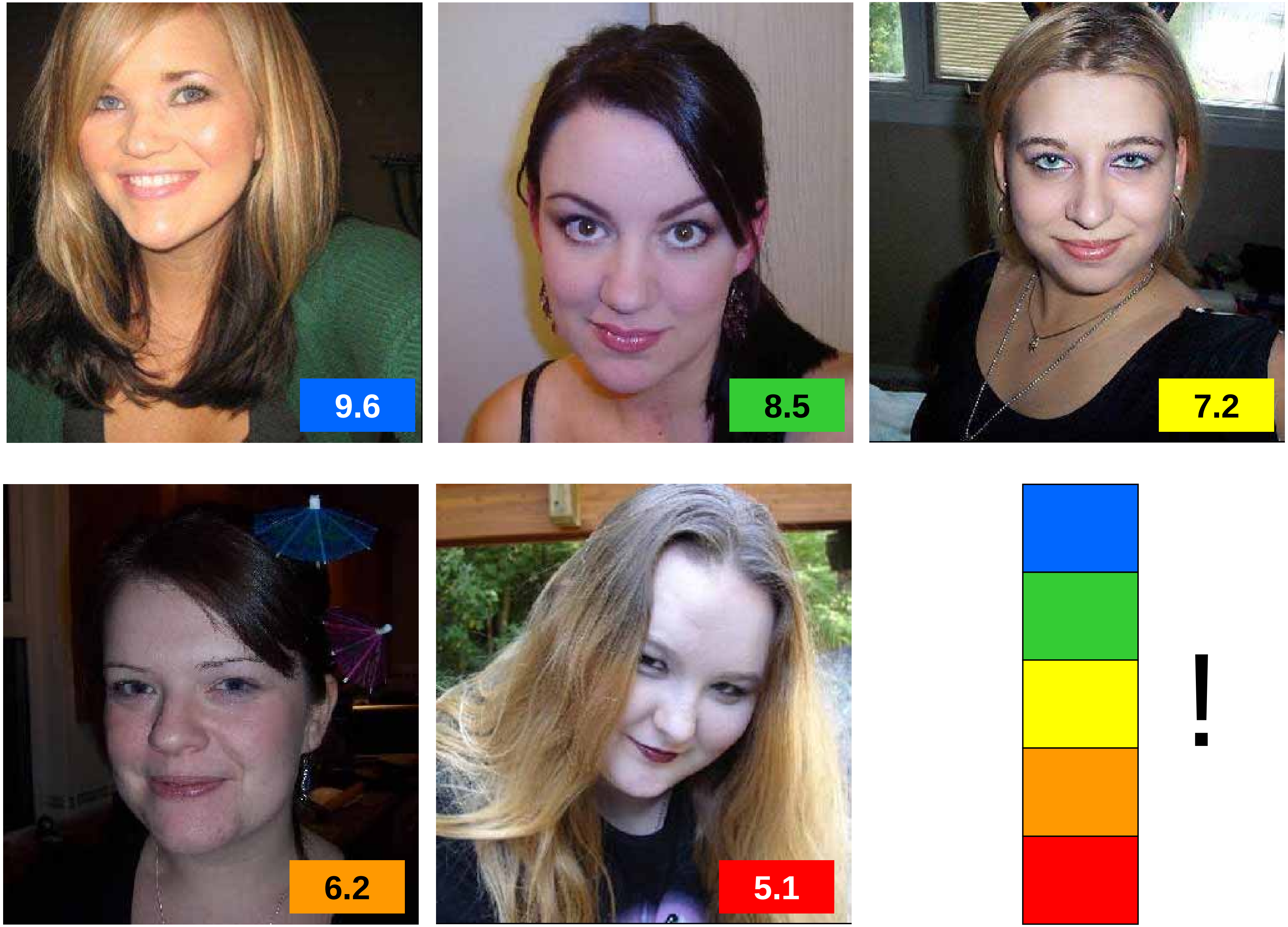


Environmental data
(physical constraints, pressures and disturbances)
1
1
Predicted fish assemblage
(expected biotic response)
Expert judgement
Observed fish assemblage
(real biotic response)
Field data
site 1 site 2 …
Sp. 1 1 0 …
Sp. 2 4 2 …
… … … …
Sp. n 0 5 …



http://www.mare-net.com/mscardi/work/wfd/nnqi.htm

Can we trust expert judgement?


9.6
8.5
7.2
6.2
5.1
!

Ok, it works.

(Ranking based on thousands of records on www.hotornot.com)

BTW: does a **multimetric** approach to facial recognition/evaluation work?

Vol.42 , No.4, December 1997, pp. ***–***. / Journal of Mathematical Psychology

Principal Component and Neural Network Analyses of Face Images: What Can Be Generalized in Gender Classification?

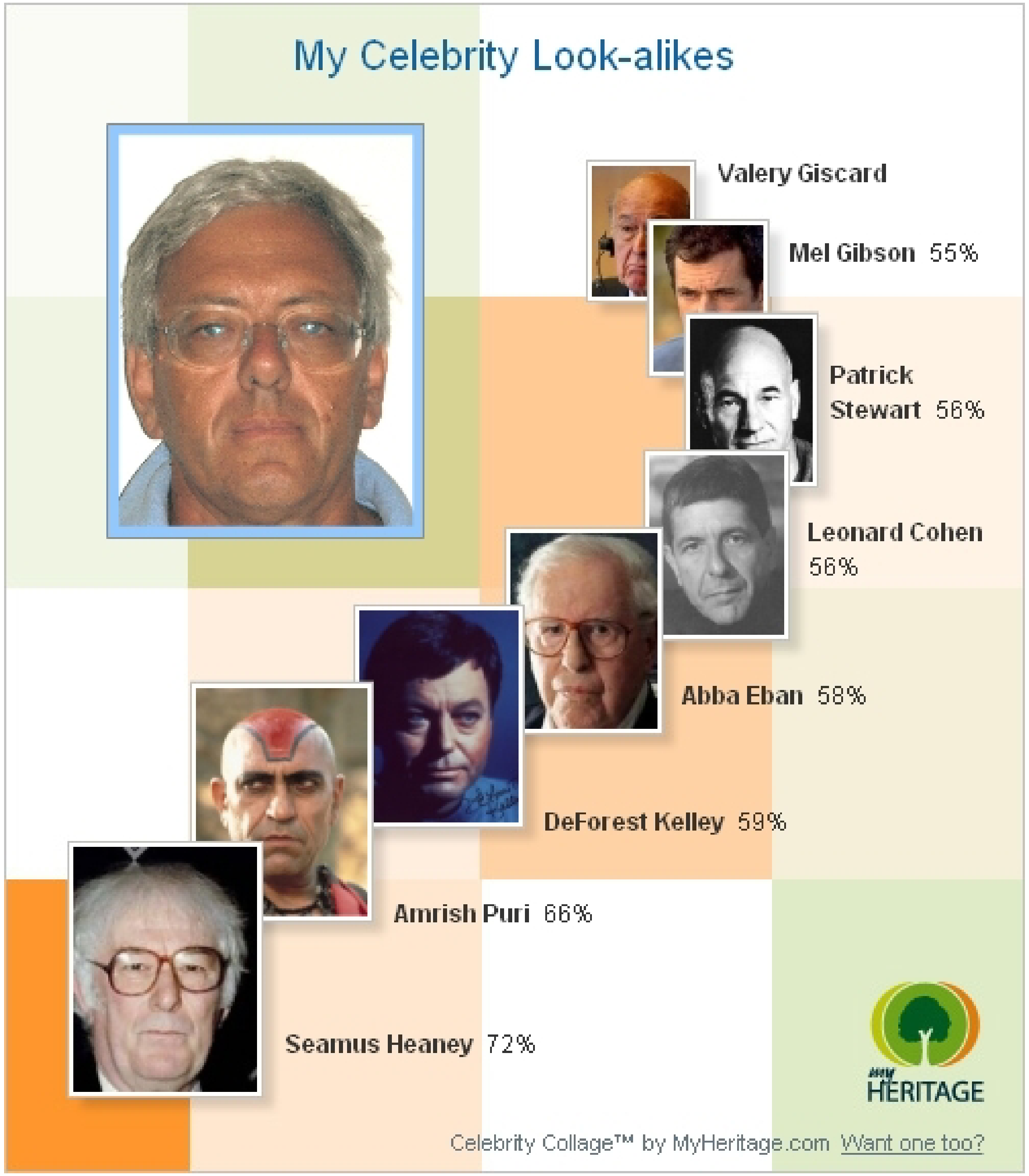
Dominique Valentin*†, Hervé Abdi*†, Betty Edelman* and Alice J. O'Toole*
* *The University of Texas at Dallas,* † *Université de Bourgogne à Dijon*

We present an overview of the major findings of the principal component analysis (PCA) approach to facial analysis. In a neural network or connectionist framework this approach is known as the linear autoassociator approach. Faces are represented as a weighted sum of macrofeatures (eigenvectors or eigenfaces) extracted from a cross-product matrix of face images. Using gender categorization as an illustration, we analyze the robustness of this type of facial representation. We show that eigenvectors representing general categorical information can be estimated using a very small set of faces and that the information they convey is generalizable to new faces of the same population and to a lesser extent to new faces of a different population.

1. Introduction

One of the major problems in modeling face processing is to find a way of representing faces that allows & Kidode, 1971) or in terms of template parameters (Yuille, 1991), or isodensity lines (Nakamura, Mathur & Minami, 1991). Although these approaches economically represent faces in a way that is relatively insensitive to variations in scale, tilt, or rotation of the faces, they are not without problems (for a review, see Samal & Iyengar, 1992).

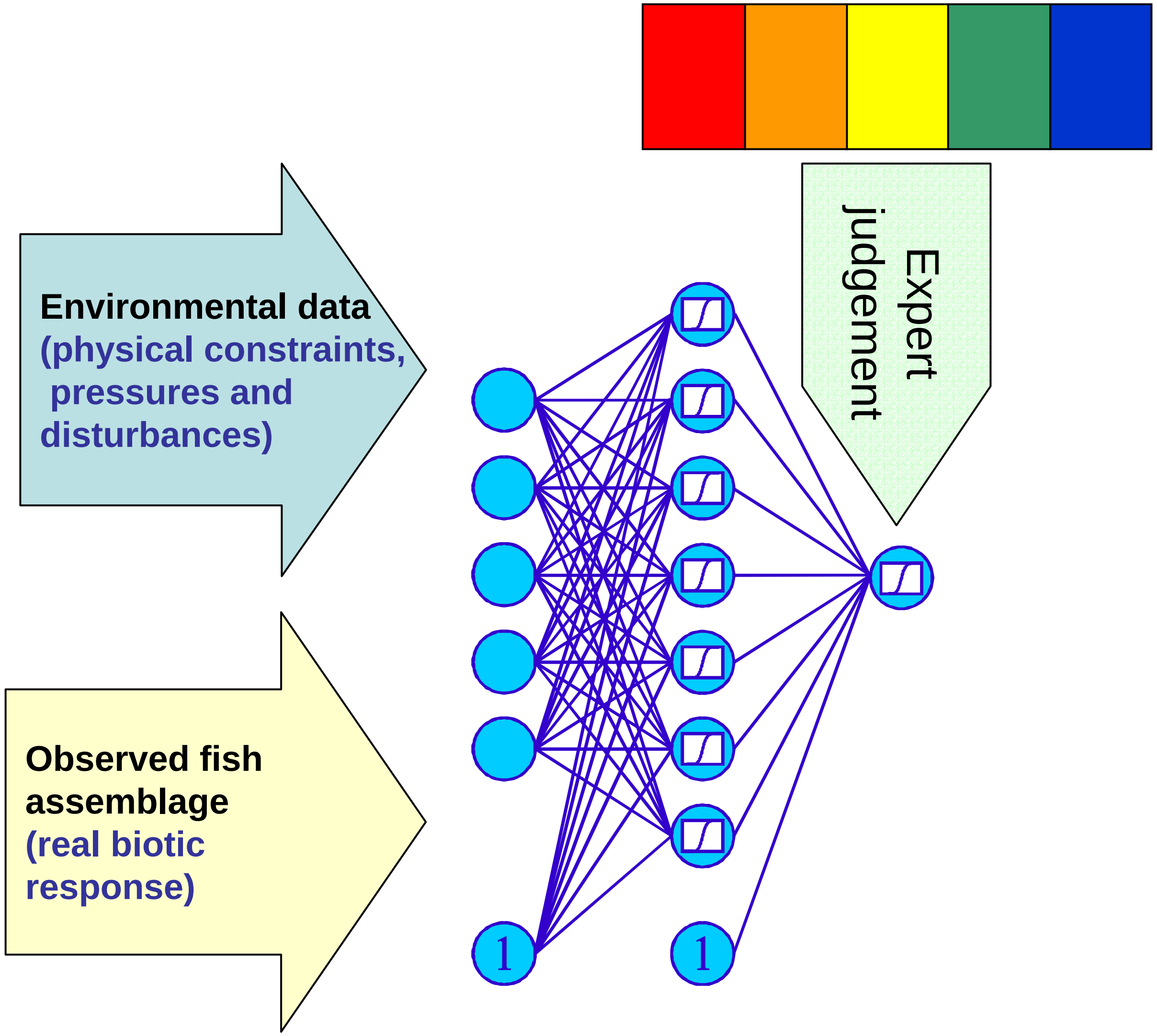
The major difficulty with representing faces as a set of features is that it assumes some *a priori* knowledge about what are the features and/or what are the relationships between them that are essential to the task at hand. Burton, Bruce, and Dench (1993), for example, showed the difficulty of finding a set of features useful in discriminating accurately between male and female faces. In a series of five experiments, they examined the usefulness of different kinds of feature measures for predicting the gender of a set of faces. The measures they used ranged from simple raw distances between facial landmarks (*e.g.*, pupils)


My Celebrity Look-alikes
Valery Giscard
Mel Gibson 55%
Patrick Stewart 56%
Leonard Cohen 56%
Abba Eban 58%
DeForest Kelley 59%
Amrish Puri 66%
Seamus Heaney 72%
my HERITAGE
Celebrity Collage™ by MyHeritage.com Want one too?

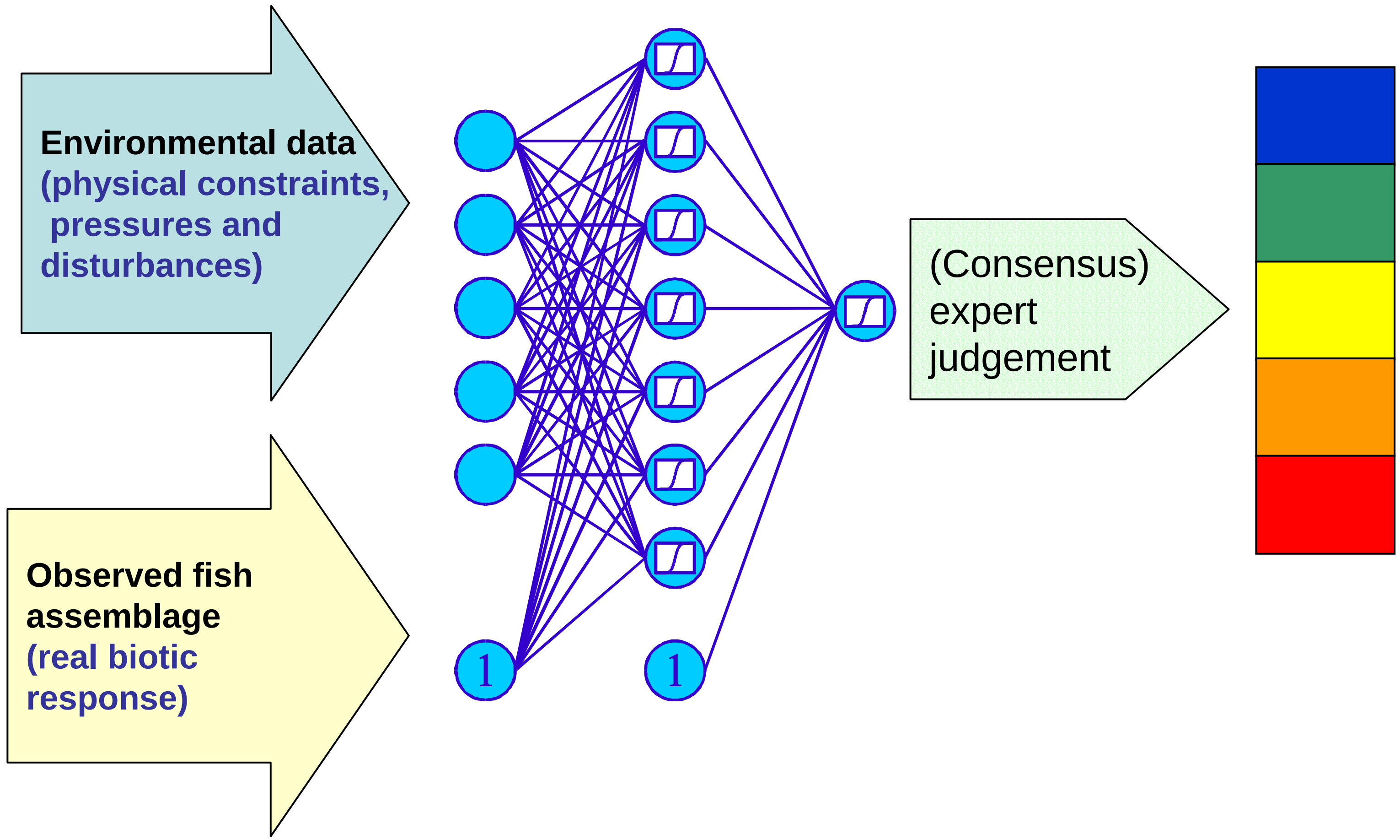
An alternate approach

Expert judgement is considered (*ex ante*) as the target and the system is trained to reproduce it

Training phase



Operational phase



Evaluating the ‘ecological status’
(expert system based on a neural network)

- Input:
  - Environmental data
  - Fish assemblage composition and simplified population structure (are juveniles present?)
  - Expert judgement about overall ecosystem quality (more than one per site, if possible!)
- Output:
  1. Consensus expert judgement (best estimate for ecological status)
  2. Sensitivity analyses for environmental management

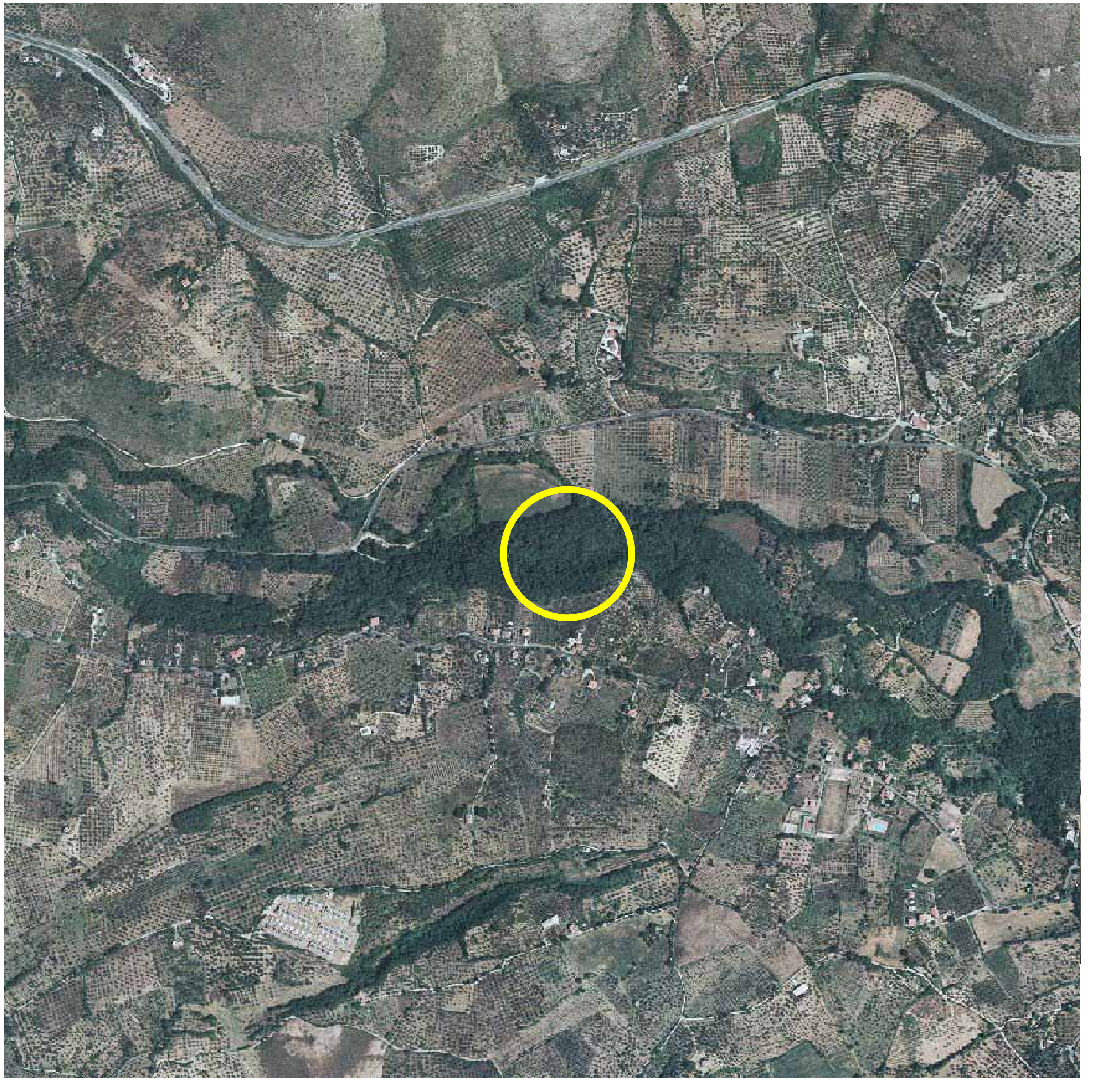
	tev02-2	cre1-3	sac1	fsv1-3	ani2-2	far5-3
Altitudine (m)	23	162	250	76.13	494	55
Profondità (m)	3	0.3	0.4	0.3	1.2	1.2
Correntini (% sup)	70	80	50	85	60	60
Pozze (% sup)	25	20	30	10	30	35
Raschi (% sup)	5	0	20	5	10	5
Pattern indistinto alla superficie (% sup)	0	0	0	0	0	0
presenza zone umide connesse (0-1)	1	0	0	0	0	0
barre di meandro e puntiformi, isole (0-1)	1	0	0	0	0	0
Massi (% sup)	0	0	0	0	0	0
Sassi e ciottoli (% sup)	20	10	5	0	65	10
Ghiaia (% sup)	30	30	25	23	25	20
Sabbia (% sup)	40	40	70	64	20	70
Limo e argilla (% sup)	10	20	0	13	0	0
Velocità (0 - 5)	2	1	2	1.5	2	2
Copertura vegetale in alveo (% sup)	10	20	10	10	0	30
Ombreggiamento (% sup)	10	60	60	15	15	65
Disturbo antropico (0 - 4)	1.5	3	2	3	2	0.5
Sbarramento a monte entro (Km, 100=max=no)	100	2	100	100	5	3
Sbarramento a valle entro (0/1)	0	1	0	0	1	1
Lago a monte entro (Km, 50=max=no)	50	50	50	50	50	50
Temperatura (°C)	21.8	23.26	18.52	13.93	12.4	16.5
Torbidità (NTU)	10.64	14	11	2	2	8
pH	7.4	7.8	7.93	8.35	8.23	7.8
Conducibilità specifica	1200	1388	579	860	372	480
O2 %	73	75	69.45	67	85	96
SQRT(Area sottesa (kmq))	119.337	4.590207	5.611595	8.11172	14.12232	15.11952
Distanza sorgente	300.558	5.81	7.211	15.182	24.042	26

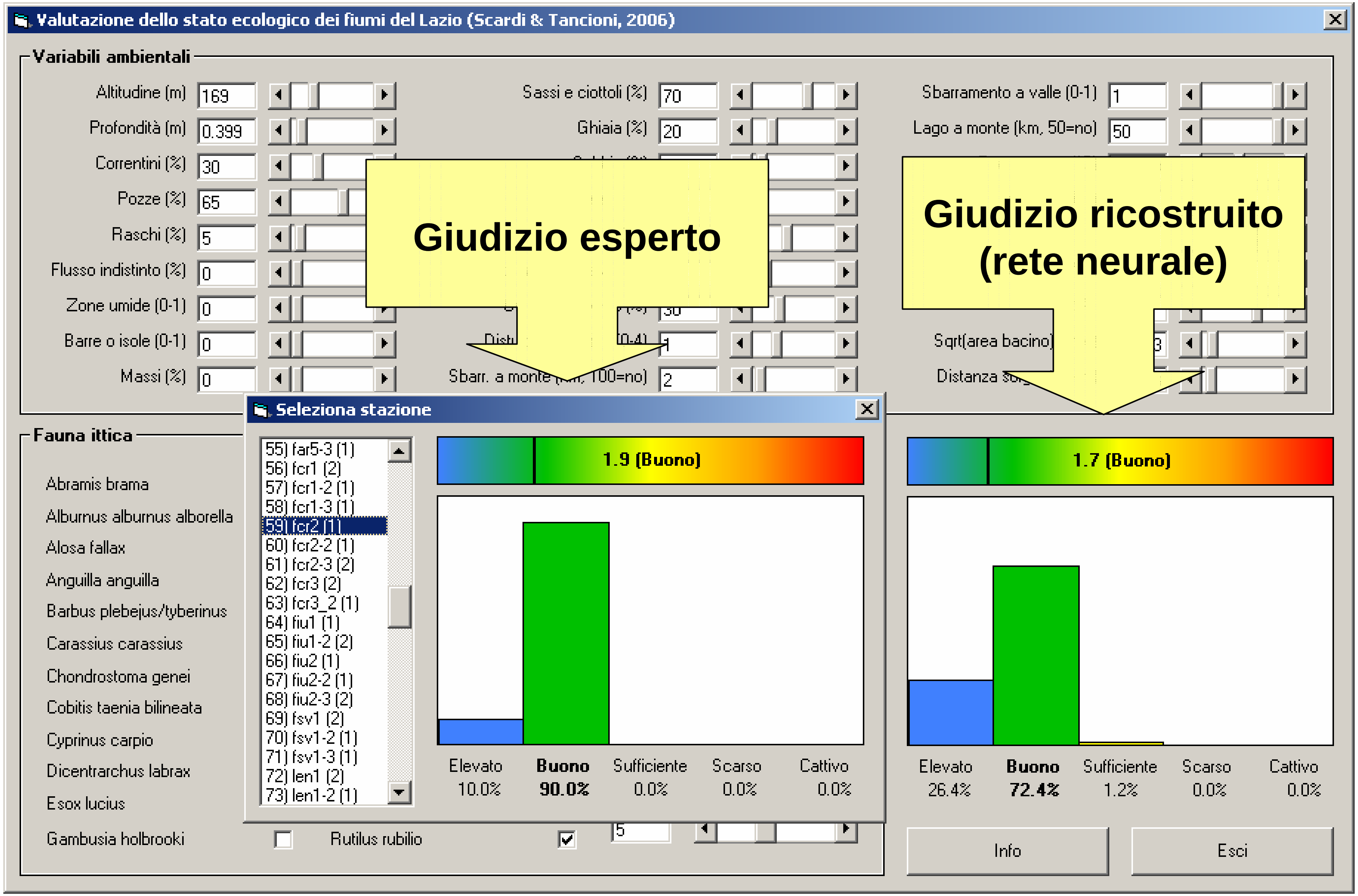
	tev02-2	cre1-3	sac1	fsv1-3	ani2-2	far5-3
Abramis brama	0	0	0	0	0	0
Alburnus alburnus alborella	1	0	0	0	0	0
Alosa fallax	1	0	0	0	0	0
Ameiurus melas	0	0	0	0	0	0
Anguilla anguilla	1	0	0	0	0	1
Barbus barbus	0	0	0	0	0	0
Barbus plebejus	0	0	1	0	0	0
Barbus tyberinus	1	1	1	1	0	1
Carassius auratus	0	0	0	0	0	0
Carassius carassius	0	0	0	0	0	0
Chondrostoma genei	0	0	0	0	0	0
Clarias gariepinus	0	0	0	0	0	0
Cobitis taenia bilineata	0	1	1	0	0	0
Cyprinus carpio	1	0	0	0	0	0
...	...	...	...	...	...	...
Rutilus rutilus	1	0	0	0	0	0
Salaria fluviatilis	1	0	0	0	0	0
Salmo trutta	0	0	0	0	1	0
Sander lucioperca	1	0	0	0	0	0
Scardinius erythrophthalmus	1	0	0	0	0	0
Silurus glanis	0	0	0	0	0	0
Tinca tinca	1	0	0	0	0	0
RS	16	4	7	3	1	7
RS(juv)	6	0	3	0	0	6
Stato elevato (%)	0	0	0	0	0	20
Stato buono(%)	80	0	0	0	40	80
Stato sufficiente(%)	20	20	70	40	60	0
Stato insufficiente(%)	0	80	30	60	0	0
stato pessimo (%)	0	0	0	0	0	0

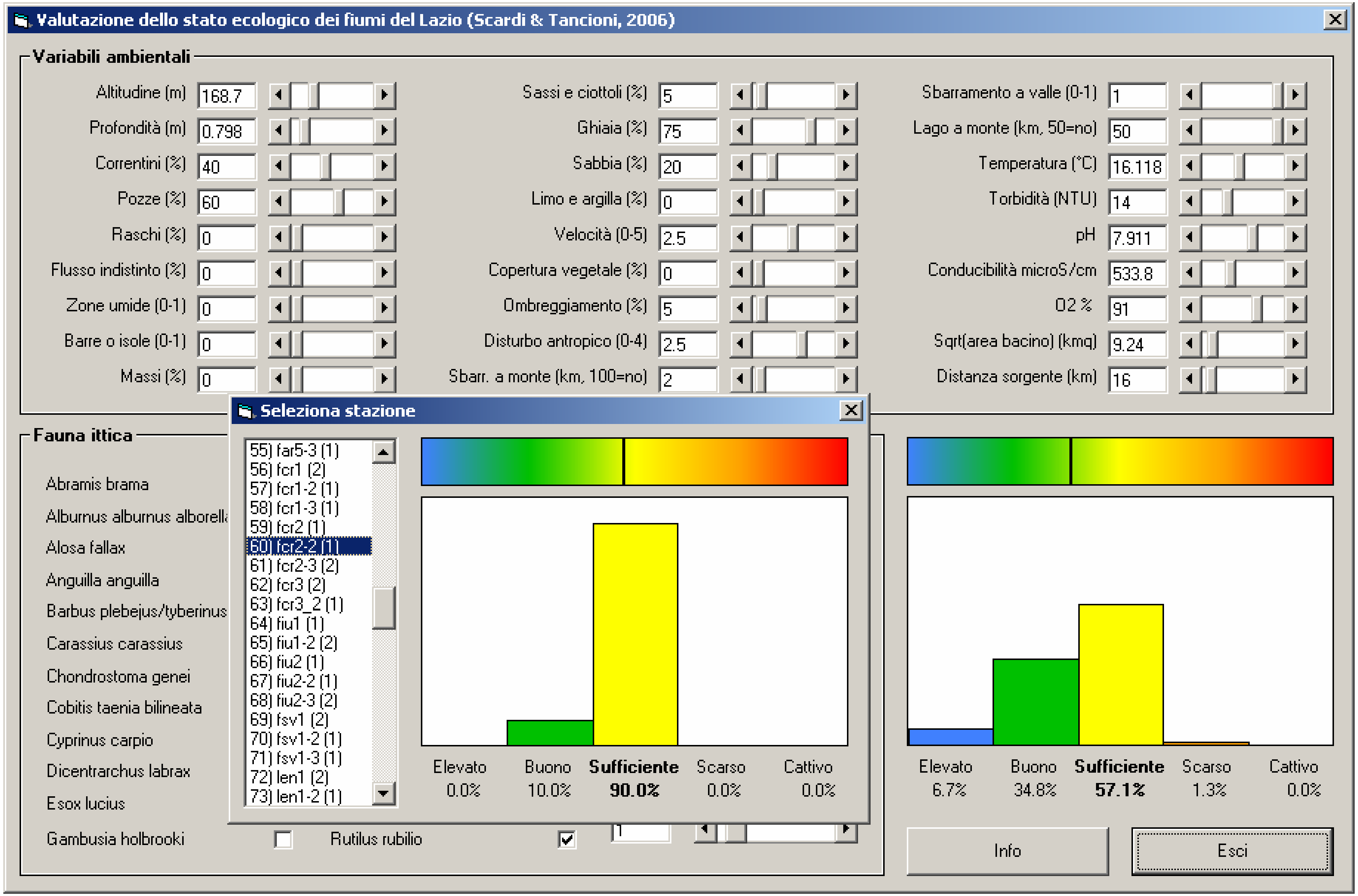
	real	simulated	
Sampling site	**ani1**	**ani1-2**	**ani1-3**
Elevation (m)	535	535	535
Depth (m)	1	1	**1.5**
Runs (% surface)	70	**60**	**100**
Pools (% surface)	25	**35**	**0**
Riffles (% surface)	5	5	**0**
Rocks (% surface)	5	5	0
Stones and pebbles (% surface)	70	**50**	**0**
Gravel (% surface)	25	**45**	**80**
Sand (% surface)	0	0	**20**
Silt and clay (% surface)	0	0	0
Water flow (0 - 5)	3	**2**	3
Vegetational cover (% surface)	0	**10**	0
Shade (% surface)	70	70	**0**
Anthropic disturbance (0 - 4)	0.5	0.5	**3**
Lampetra planeri	0	**1**	0
Salmo trutta	1	1	1
RS	1	**2**	1
RS(juv)	0	**2**	0
High (fuzzy membership, %)	10	**70**	0
Good (fuzzy membership, %)	90	**30**	**10**
Moderate (fuzzy membership, %)	0	0	**90**
Poor (fuzzy membership, %)	0	0	0
Bad (fuzzy membership, %)	0	0	0

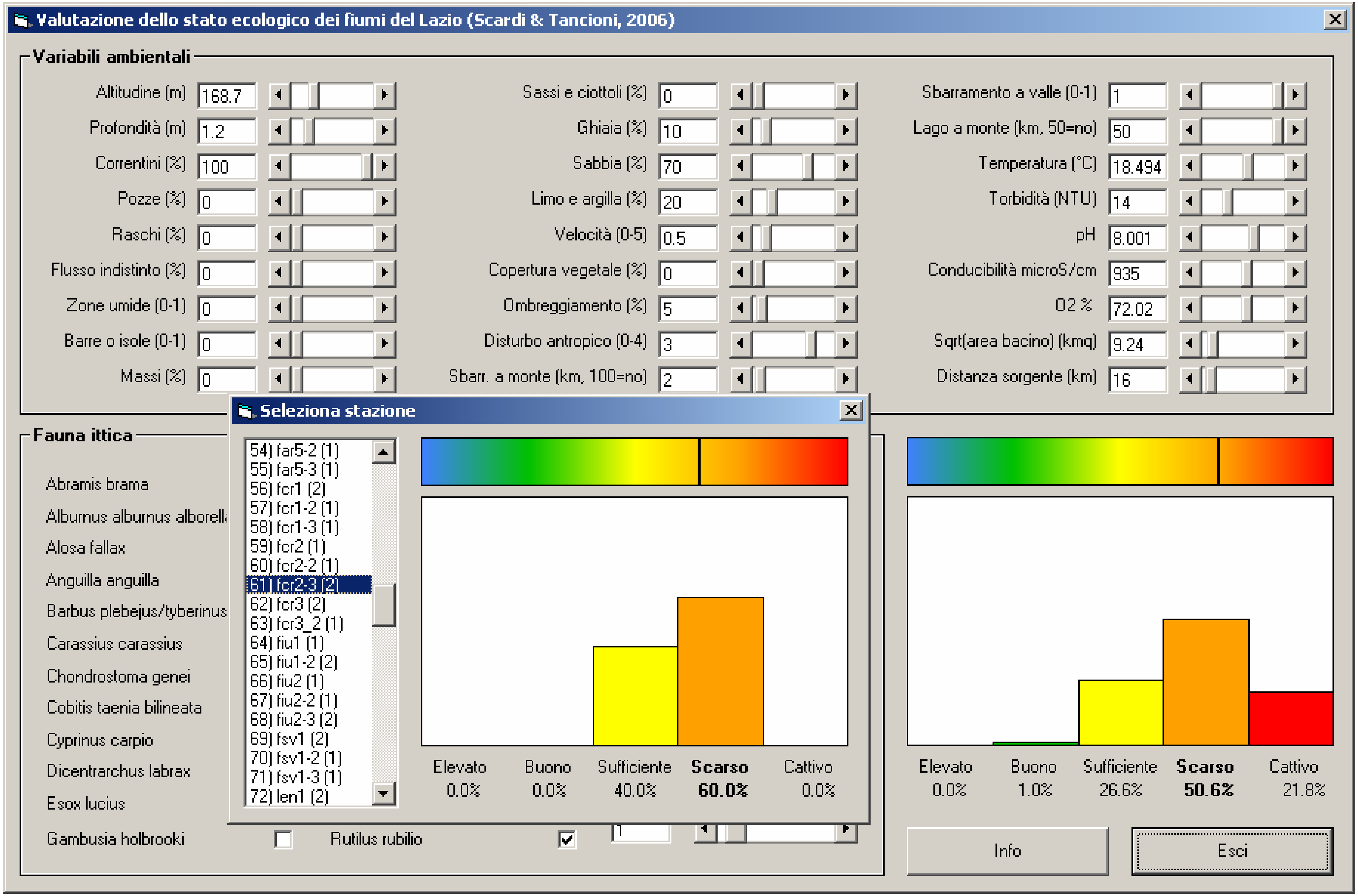
Our very preliminary data set (but growing!):

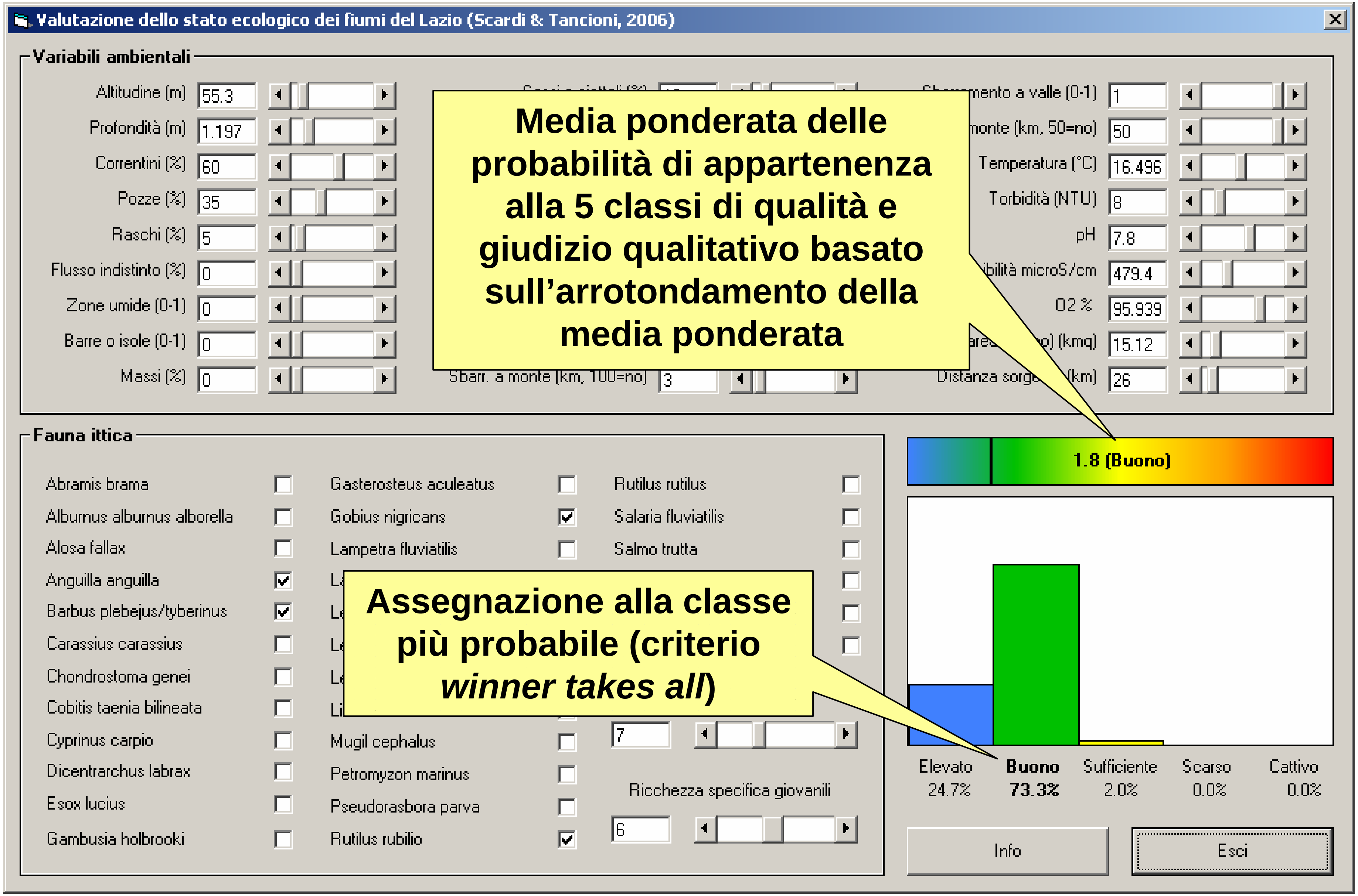
219 records
(both real and simulated)




Valutazione dello stato ecologico dei fiumi del Lazio (Scardi & Tancioni, 2006)
Variabili ambientali
Altitudine (m) 169
Profondità (m) 0.399
Correntini (%) 30
Pozze (%) 65
Raschi (%) 5
Flusso indistinto (%) 0
Zone umide (0-1) 0
Barre o isole (0-1) 0
Massi (%) 0
Sassi e ciottoli (%) 70
Ghiaia (%) 20
Sbarr. a monte (km, 100=no) 2
Sbarramento a valle (0-1) 1
Lago a monte (km, 50=no) 50
Sqrt(area bacino)
Giudizio esperto
Giudizio ricostruito (rete neurale)
Seleziona stazione
55) far5-3 (1)
56) fcr1 (2)
57) fcr1-2 (1)
58) fcr1-3 (1)
59) fcr2 (1)
60) fcr2-2 (1)
61) fcr2-3 (2)
62) fcr3 (2)
63) fcr3_2 (1)
64) fiu1 (1)
65) fiu1-2 (2)
66) fiu2 (1)
67) fiu2-2 (1)
68) fiu2-3 (2)
69) fsv1 (2)
70) fsv1-2 (1)
71) fsv1-3 (1)
72) len1 (2)
73) len1-2 (1)
1.9 (Buono)
Elevato 10.0%
Buono 90.0%
Sufficiente 0.0%
Scarso 0.0%
Cattivo 0.0%
1.7 (Buono)
Elevato 26.4%
Buono 72.4%
Sufficiente 1.2%
Scarso 0.0%
Cattivo 0.0%
Fauna ittica
Abramis brama
Alburnus alburnus alborella
Alosa fallax
Anguilla anguilla
Barbus plebejus/tyberinus
Carassius carassius
Chondrostoma genei
Cobitis taenia bilineata
Cyprinus carpio
Dicentrarchus labrax
Esox lucius
Gambusia holbrooki
Rutilus rubilio
Info
Esci


Valutazione dello stato ecologico dei fiumi del Lazio (Scardi & Tancioni, 2006)
Variabili ambientali
Altitudine (m) 168.7
Profondità (m) 0.798
Correntini (%) 40
Pozze (%) 60
Raschi (%) 0
Flusso indistinto (%) 0
Zone umide (0-1) 0
Barre o isole (0-1) 0
Massi (%) 0
Sassi e ciottoli (%) 5
Ghiaia (%) 75
Sabbia (%) 20
Limo e argilla (%) 0
Velocità (0-5) 2.5
Copertura vegetale (%) 0
Ombreggiamento (%) 5
Disturbo antropico (0-4) 2.5
Sbarr. a monte (km, 100=no) 2
Sbarramento a valle (0-1) 1
Lago a monte (km, 50=no) 50
Temperatura (°C) 16.118
Torbidità (NTU) 14
pH 7.911
Conducibilità microS/cm 533.8
O2 % 91
Sqrt(area bacino) (kmq) 9.24
Distanza sorgente (km) 16
Fauna ittica
Abramis brama
Alburnus alburnus alborell
Alosa fallax
Anguilla anguilla
Barbus plebejus/tyberinus
Carassius carassius
Chondrostoma genei
Cobitis taenia bilineata
Cyprinus carpio
Dicentrarchus labrax
Esox lucius
Gambusia holbrooki
Rutilus rubilio
Seleziona stazione
55) far5-3 (1)
56) fcr1 (2)
57) fcr1-2 (1)
58) fcr1-3 (1)
59) fcr2 (1)
60) fcr2-2 (1)
61) fcr2-3 (2)
62) fcr3 (2)
63) fcr3_2 (1)
64) fiu1 (1)
65) fiu1-2 (2)
66) fiu2 (1)
67) fiu2-2 (1)
68) fiu2-3 (2)
69) fsv1 (2)
70) fsv1-2 (1)
71) fsv1-3 (1)
72) len1 (2)
73) len1-2 (1)
Elevato 0.0%
Buono 10.0%
Sufficiente 90.0%
Scarso 0.0%
Cattivo 0.0%
Elevato 6.7%
Buono 34.8%
Sufficiente 57.1%
Scarso 1.3%
Cattivo 0.0%
Info
Esci


Valutazione dello stato ecologico dei fiumi del Lazio (Scardi & Tancioni, 2006)
Variabili ambientali
Altitudine (m) 168.7
Profondità (m) 1.2
Correntini (%) 100
Pozze (%) 0
Raschi (%) 0
Flusso indistinto (%) 0
Zone umide (0-1) 0
Barre o isole (0-1) 0
Massi (%) 0
Sassi e ciottoli (%) 0
Ghiaia (%) 10
Sabbia (%) 70
Limo e argilla (%) 20
Velocità (0-5) 0.5
Copertura vegetale (%) 0
Ombreggiamento (%) 5
Disturbo antropico (0-4) 3
Sbarr. a monte (km, 100=no) 2
Sbarramento a valle (0-1) 1
Lago a monte (km, 50=no) 50
Temperatura (°C) 18.494
Torbidità (NTU) 14
pH 8.001
Conducibilità microS/cm 935
O2 % 72.02
Sqrt(area bacino) (kmq) 9.24
Distanza sorgente (km) 16
Fauna ittica
Abramis brama
Alburnus alburnus alborell
Alosa fallax
Anguilla anguilla
Barbus plebejus/tyberinus
Carassius carassius
Chondrostoma genei
Cobitis taenia bilineata
Cyprinus carpio
Dicentrarchus labrax
Esox lucius
Gambusia holbrooki
Rutilus rubilio
1
Seleziona stazione
54) far5-2 (1)
55) far5-3 (1)
56) fcr1 (2)
57) fcr1-2 (1)
58) fcr1-3 (1)
59) fcr2 (1)
60) fcr2-2 (1)
61) fcr2-3 (2)
62) fcr3 (2)
63) fcr3_2 (1)
64) fiu1 (1)
65) fiu1-2 (2)
66) fiu2 (1)
67) fiu2-2 (1)
68) fiu2-3 (2)
69) fsv1 (2)
70) fsv1-2 (1)
71) fsv1-3 (1)
72) len1 (2)
Elevato 0.0%
Buono 0.0%
Sufficiente 40.0%
Scarso 60.0%
Cattivo 0.0%
Elevato 0.0%
Buono 1.0%
Sufficiente 26.6%
Scarso 50.6%
Cattivo 21.8%
Info
Esci


Valutazione dello stato ecologico dei fiumi del Lazio (Scardi & Tancioni, 2006)
Variabili ambientali
Altitudine (m) 55.3
Profondità (m) 1.197
Correntini (%) 60
Pozze (%) 35
Raschi (%) 5
Flusso indistinto (%) 0
Zone umide (0-1) 0
Barre o isole (0-1) 0
Massi (%) 0
Sbarr. a monte (km, 100=no) 3
Temperatura (°C) 16.496
Torbidità (NTU) 8
pH 7.8
O2 % 95.939
Distanza sorg (km) 26
Media ponderata delle probabilità di appartenenza alla 5 classi di qualità e giudizio qualitativo basato sull'arrotondamento della media ponderata
Fauna ittica
Abramis brama
Alburnus alburnus alborella
Alosa fallax
Anguilla anguilla
Barbus plebejus/tyberinus
Carassius carassius
Chondrostoma genei
Cobitis taenia bilineata
Cyprinus carpio
Dicentrarchus labrax
Esox lucius
Gambusia holbrooki
Gasterosteus aculeatus
Gobius nigricans
Lampetra fluviatilis
Mugil cephalus
Petromyzon marinus
Pseudorasbora parva
Rutilus rubilio
Rutilus rutilus
Salaria fluviatilis
Salmo trutta
Assegnazione alla classe più probabile (criterio *winner takes all*)
Ricchezza specifica giovanili
1.8 (Buono)
Elevato 24.7%
Buono 73.3%
Sufficiente 2.0%
Scarso 0.0%
Cattivo 0.0%
Info
Esci

All records

CCI=73.5%

	1	2	3	4	5	
1	**19**	21				*40*
2	2	**40**	4			*46*
3		14	**35**	2		*51*
4			9	**31**		*40*
5				6	**36**	*42*
	21	*75*	*48*	*39*	*36*	***219***

Validation records only

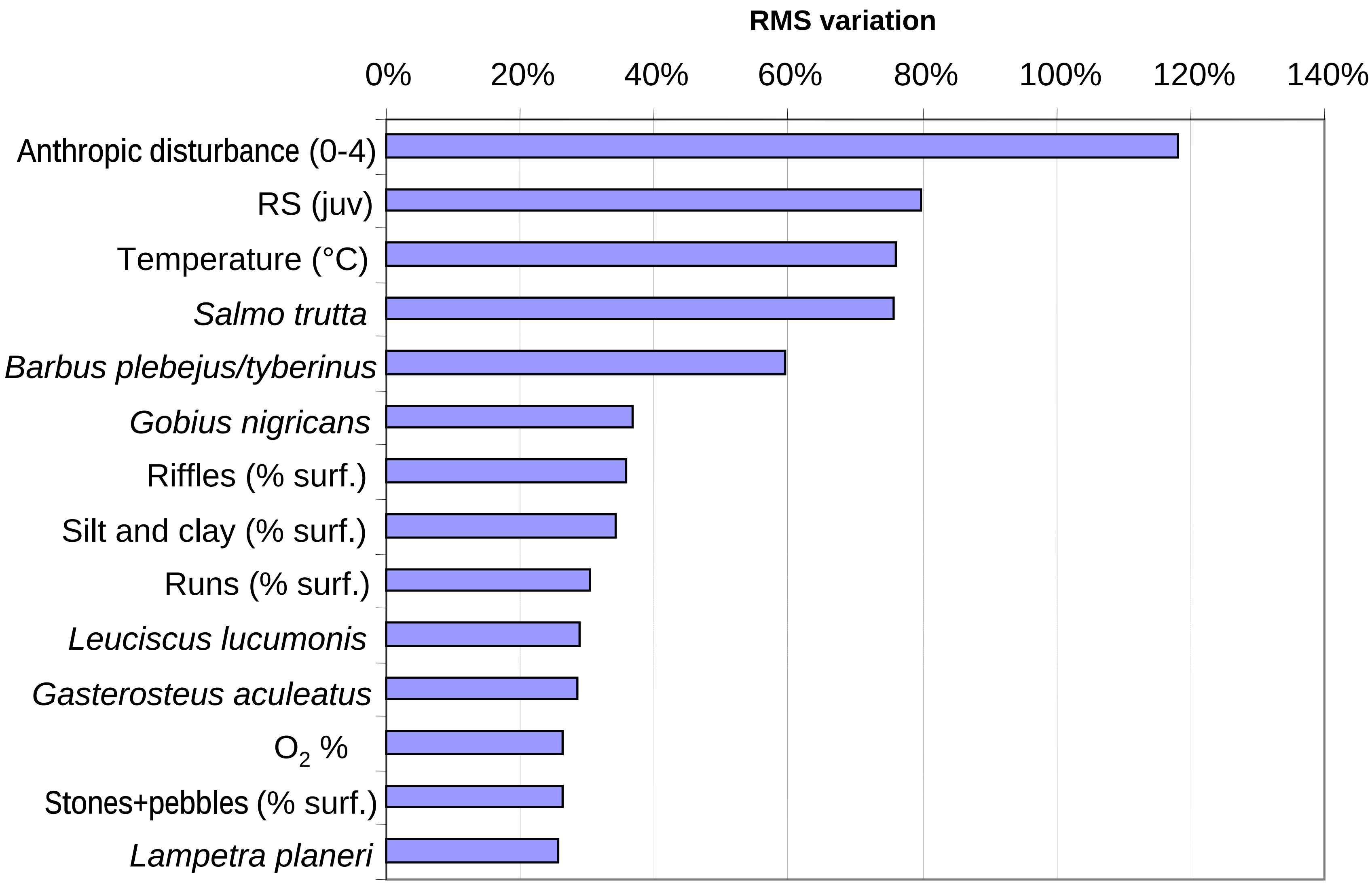
CCI=66.7%

	1	2	3	4	5	
1	**5**	7				*12*
2	2	**15**	1			*18*
3		5	**6**	2		*13*
4			2	**11**		*13*
5				4	**9**	*13*
	7	*27*	*9*	*17*	*9*	***69***

Worst misclassification: previous or next class!

Sensitivity analysis

(most relevant variables only)

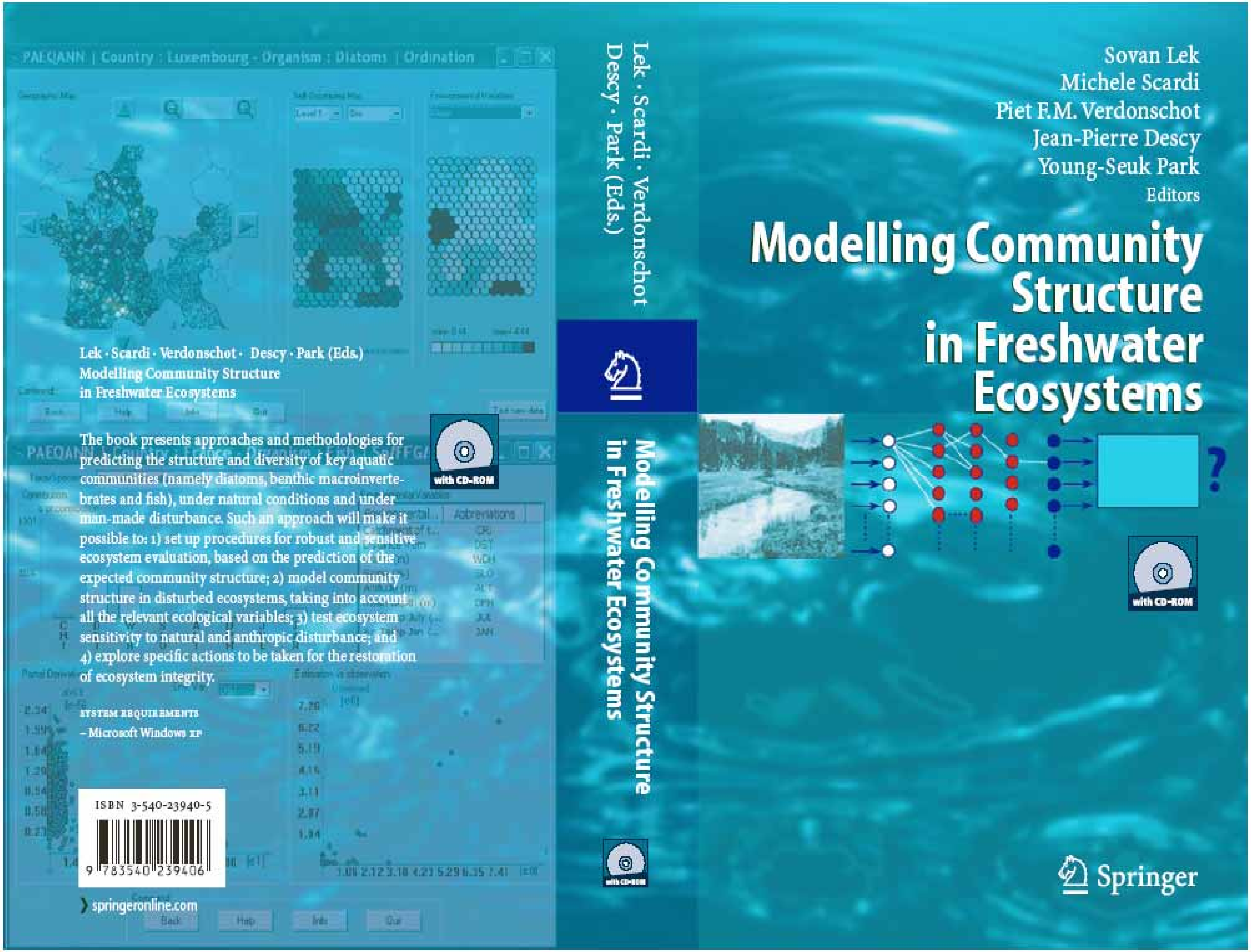


The bottom line

- There is nothing like **the** optimal method for evaluating ecosystem quality (ecological status, sensu WFD): all methods imply some subjectivity!
- So, let’s use (subjective) expert judgements as the basis for the evaluation of ‘ecological status’
- Using biotic and abiotic information and Artificial Intelligence methods, we’re able to reproduce consensus expert judgements
- However, we are going:
  - to collect new data and expert judgements
  - to validate our results in the real world
  - to improve the expert system on a routine basis
  - to use advanced computational tools, but always working, observing and thinking as ecologists!

The <u>very</u> bottom line

- Indices and other tools (ours included!) must not substitute expert judgement: environmental issues need real ecologists exactly like we need doctors
- Do you think that an automatic diagnosis system would be better than your doctor?



http://www.springeronline.com (search for: “modelling community”)

Looking for some more info?

- My home page and our email addresses:
  - **http://www.mare-net.com/mscardi**
  - **mscardi@mclink.it**
    tancioni@uniroma2.it
- 5th Conference of the International Society for Ecological Informatics
  - **http://www.isei5-conference.elsevier.com**
- Former Conferences in the same series
  - **http://www.isei3.org**
    http://www.isei4.org
- A new journal: Ecological Informatics
  - **http://www.elsevier.com/locate/ecolinf**

Popolamento atteso vs. osservato (RIVPACS)

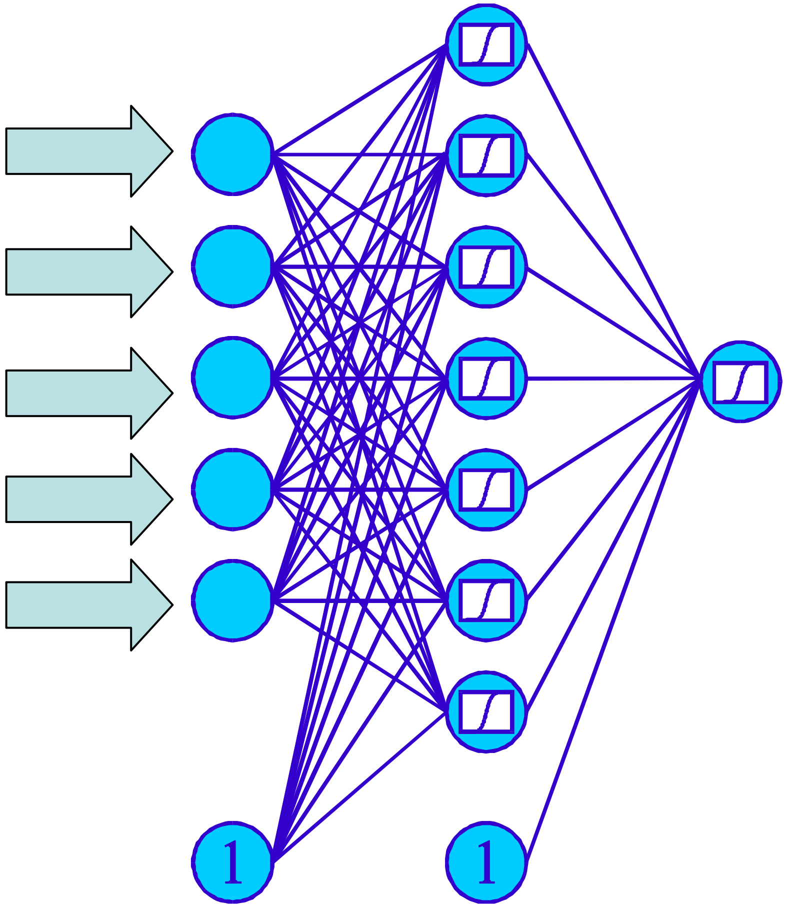
Stima del numero di taxa presenti date le condizioni al contorno = E

Numero di taxa effettivamente osservati = O

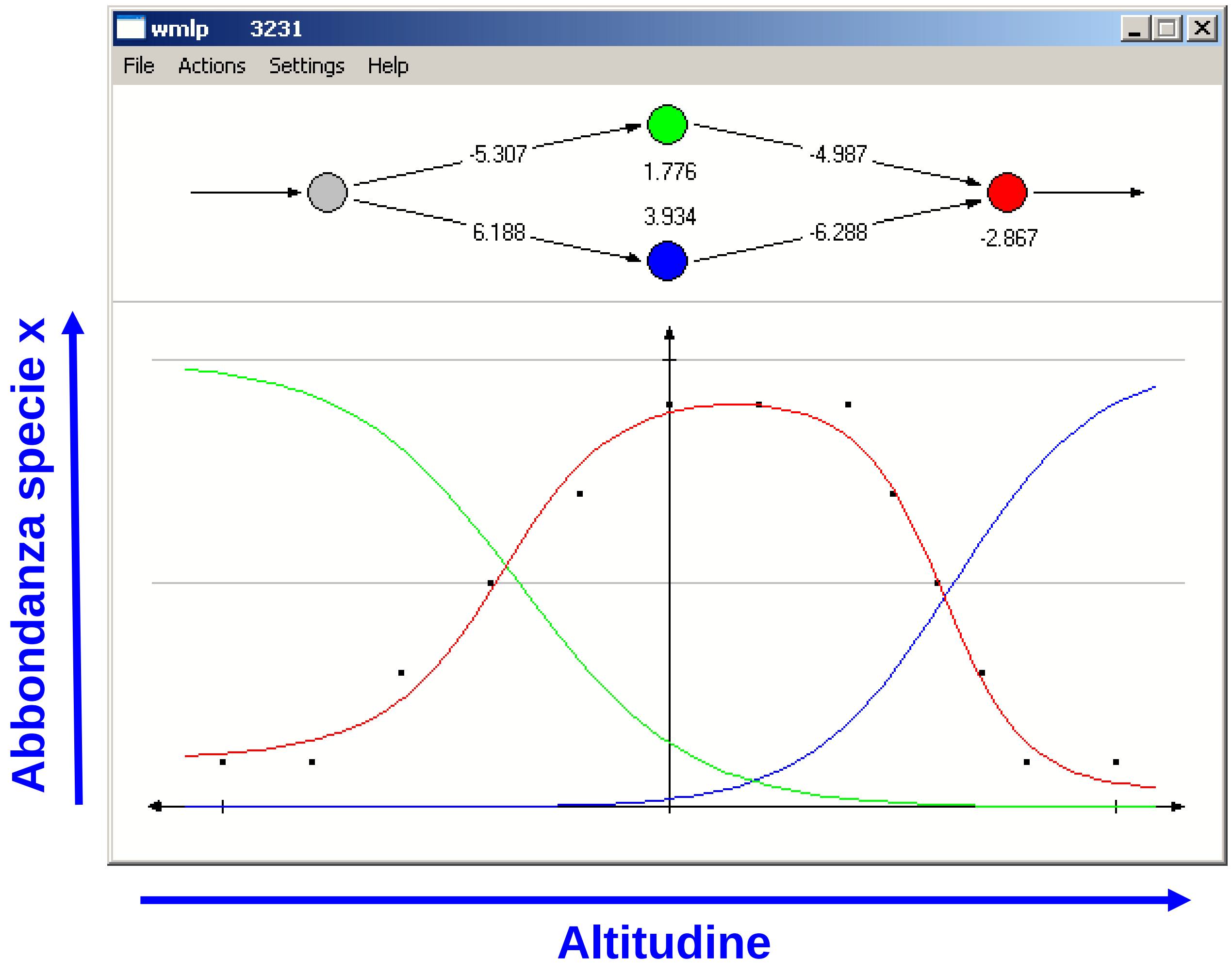
Il criterio di valutazione è basato sulla proporzione dei taxa attesi effettivamente presenti = O/E

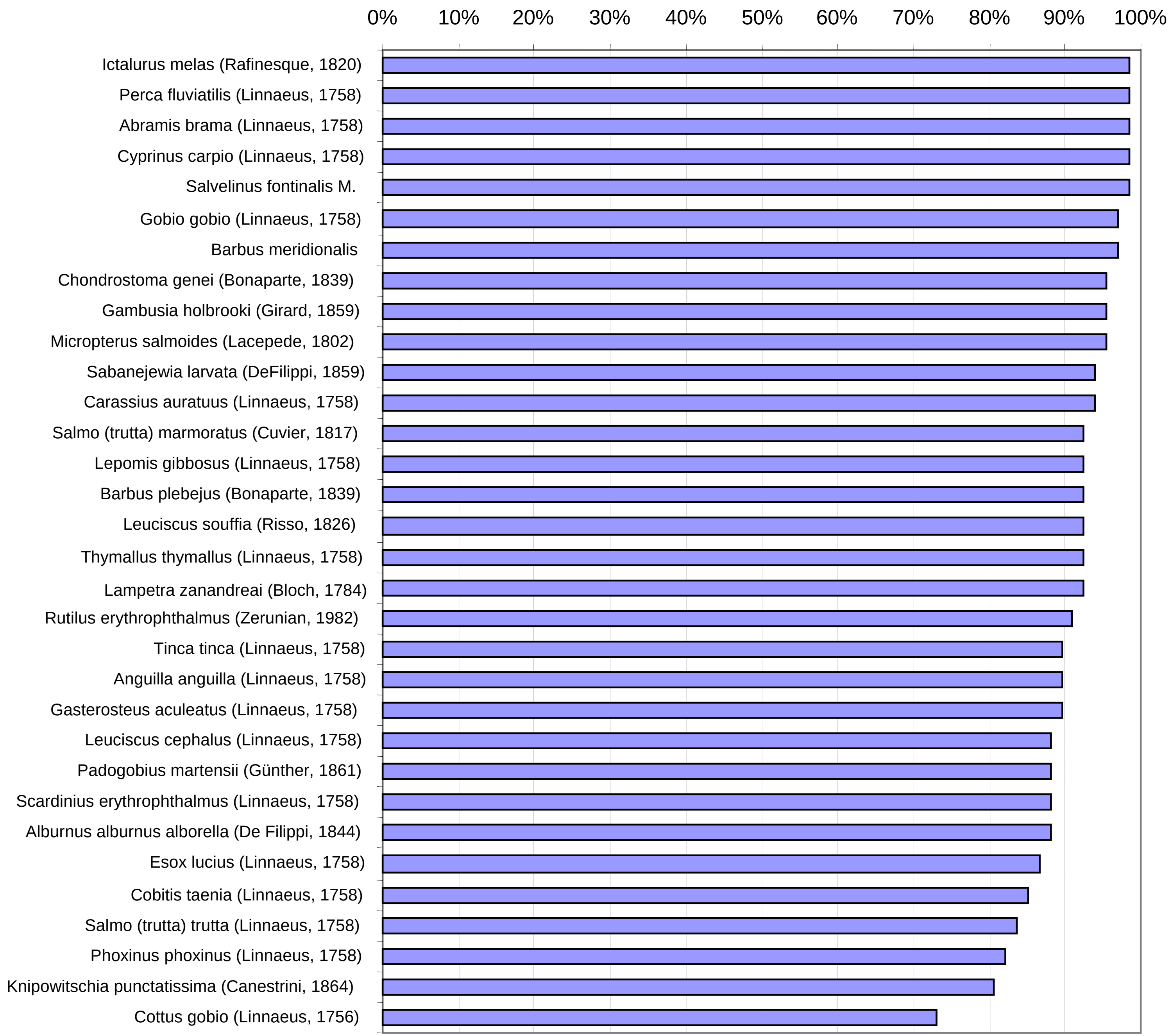
Variabili predittive (inputs NN)

1	altitudine (m)
2	profondità media (m)
3	correnti (superficie, %)
4	pozze (superficie, %)
5	raschi (superficie, %)
6	larghezza media (m)
7	massi (superficie, %)
8	sassi e ciottoli (superficie, %)
9	ghiaia (superficie, %)
10	sabbia (superficie, %)
11	peliti (superficie, %)
12	velocità flusso (punteggio, 0-5)
13	copertura vegetale (superficie, %)
14	ombreggiatura (%)
15	disturbo antropico (punteggio, 0-4)
16	pH
17	conducibilità (μS/cm)
18	gradiente (%)
19	bacino versante (km^2)
20	distanza dalla sorgente (km)



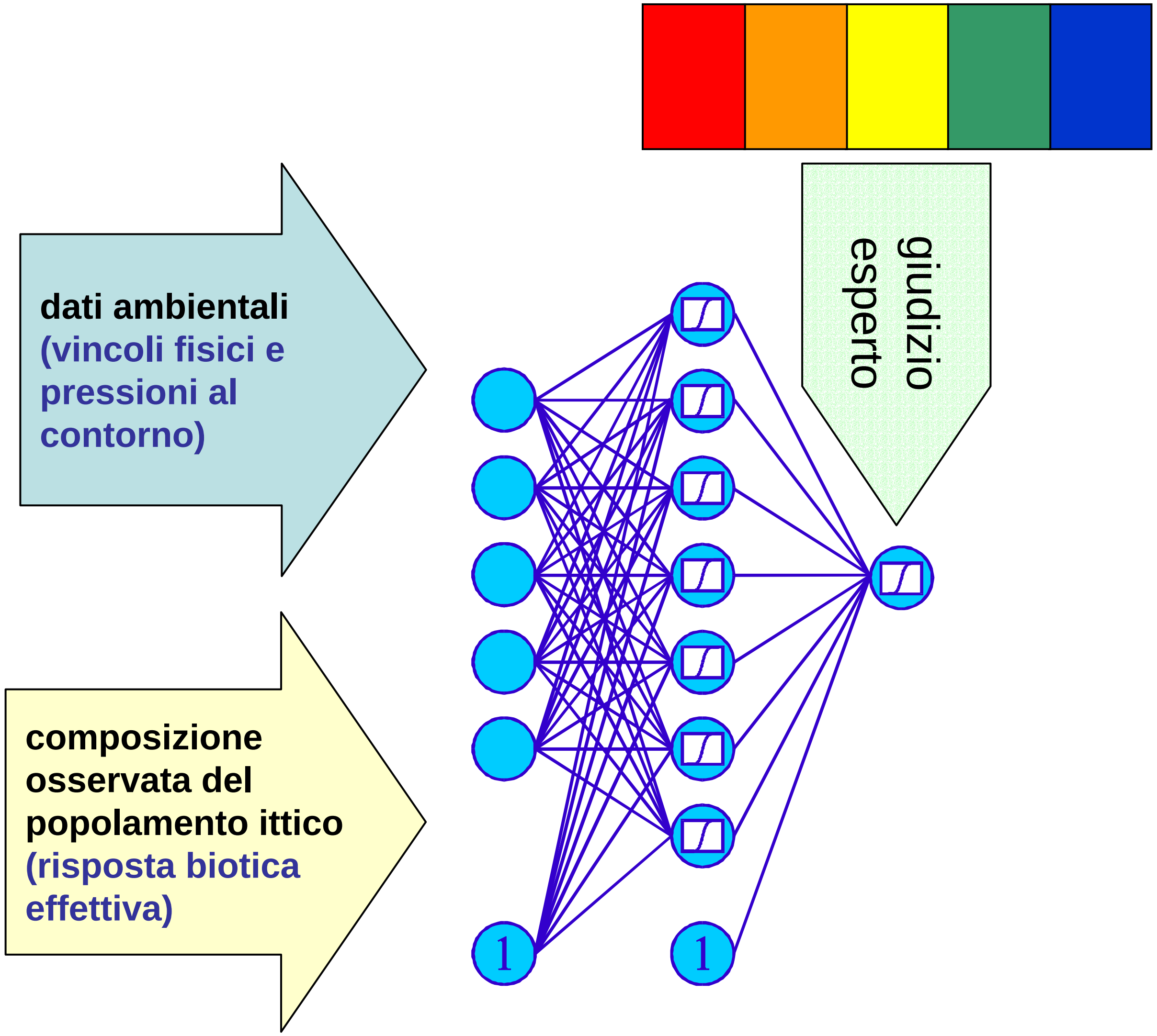
Un esempio di apprendimento in una rete neurale molto semplice



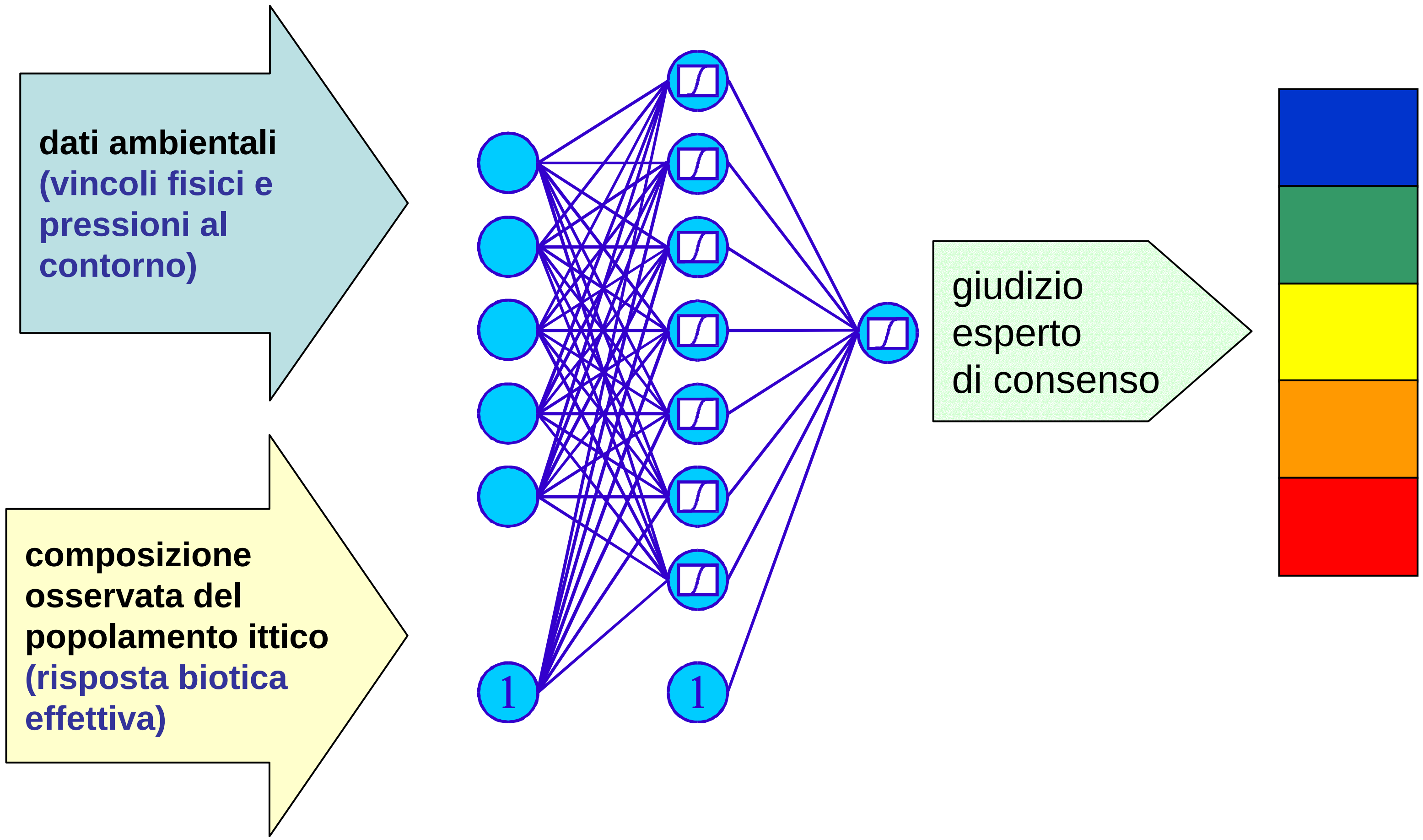


Previsioni esatte: 91.6% (media test set)

Fase di training



Fase operativa



Valutazione dello stato ecologico

(sistema esperto basato su rete neurale)

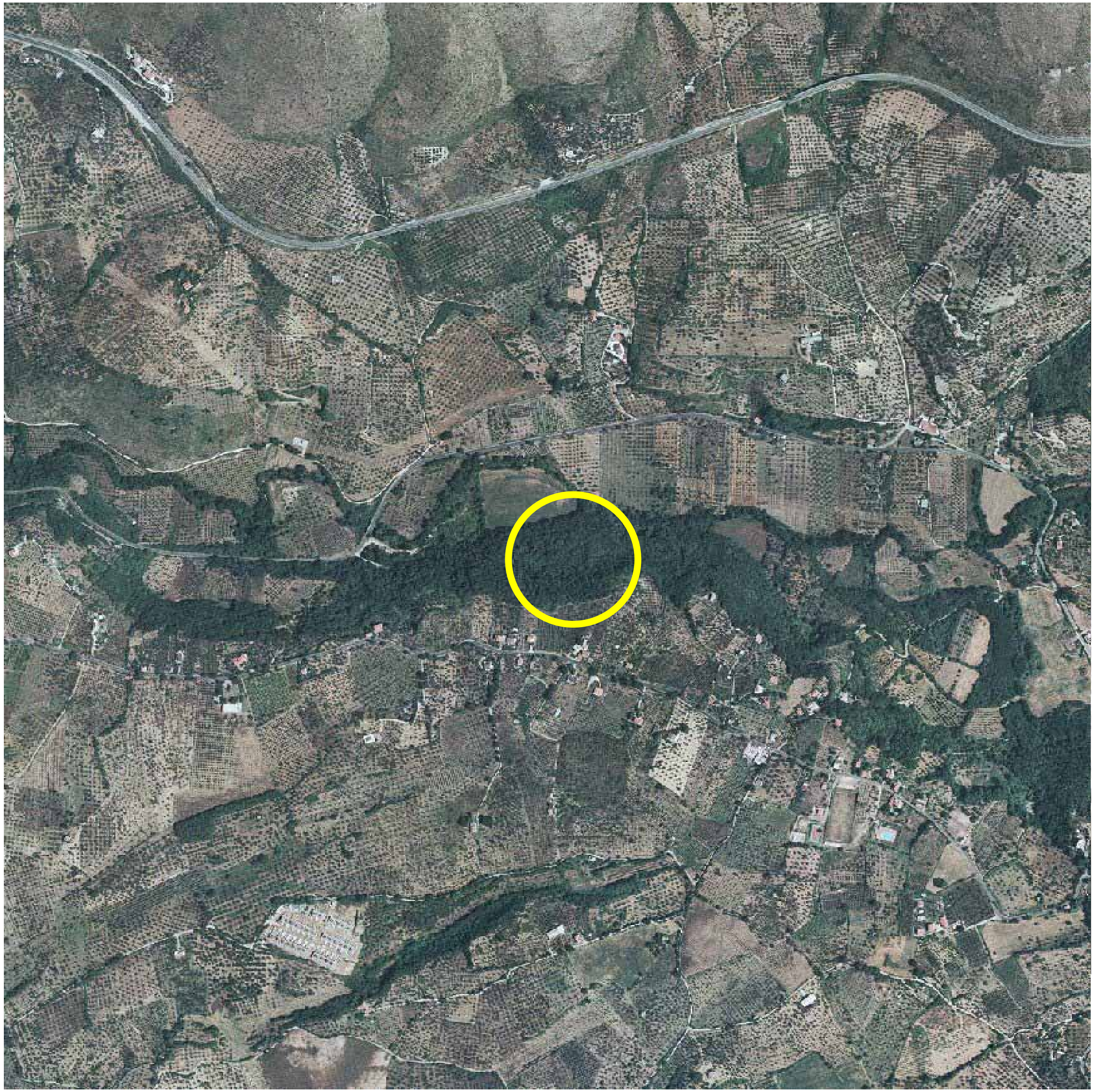
- Informazioni necessarie:
  - Dati ambientali
  - Composizione popolamento
  - Giudizio esperto (da più fonti)
- Risultati:
  1. Giudizio esperto di consenso (migliore approssimazione dell'insieme dei giudizi)
  2. Analisi del giudizio esperto di consenso (attraverso analisi di sensibilità o estrazione di regole)

Dati sui fiumi del Lazio

- Fiume Tevere (asta principale, fiume Aniene, torrente Simbrivio, torrente Fiumicino, torrente Licenza, Fosso San Vittorino, Fosso Corese, torrente Farfa, Fosso Cremera)
- Fiume Mignone (asta principale, torrente Lenta)
- Fiume Marta

Totale: 219 osservazioni (reali+simulate)

Fosso Corese



Tutti i dati

CCI=73.5%

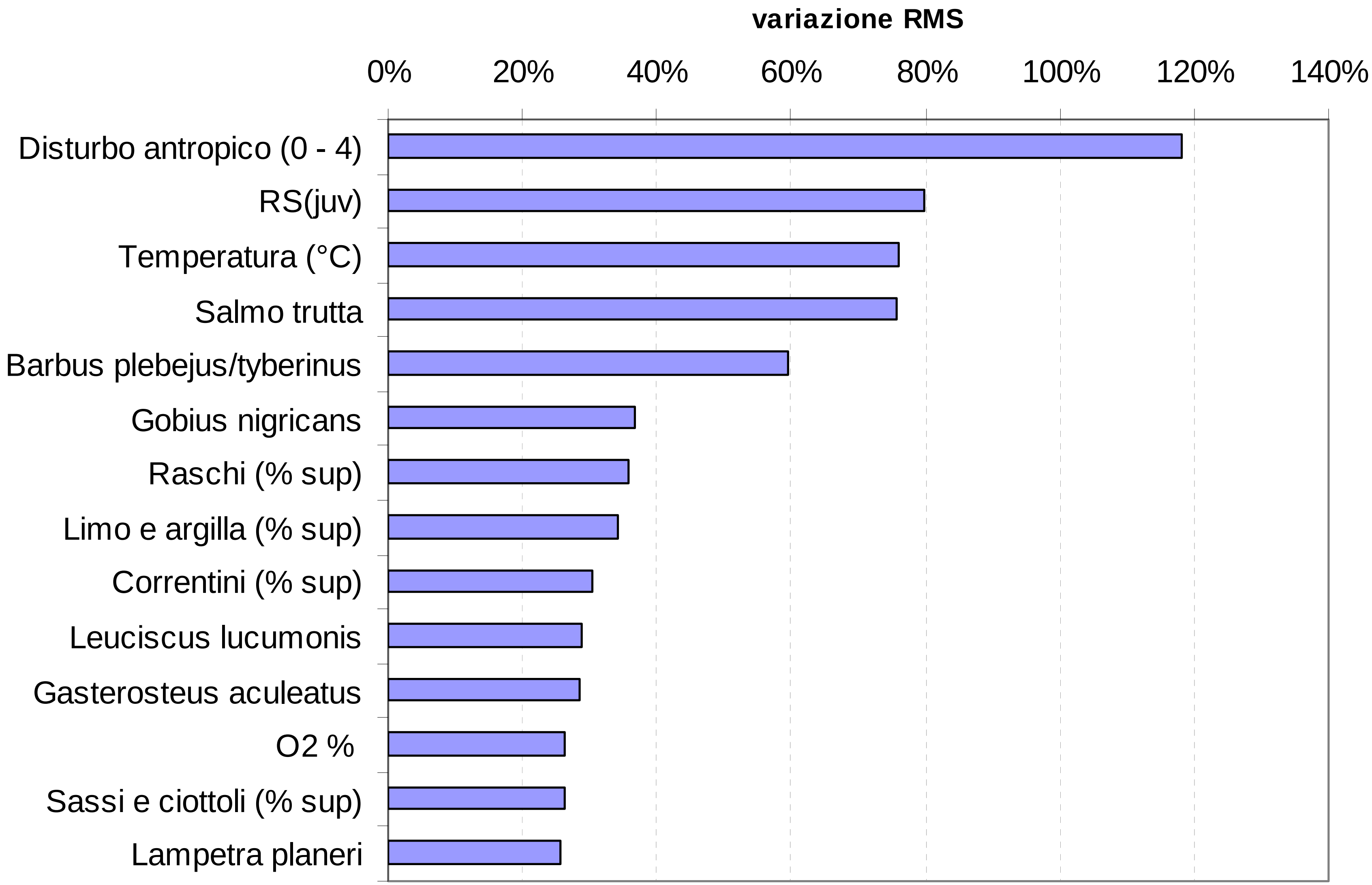
	1	2	3	4	5	
1	**19**	21				*40*
2	2	**40**	4			*46*
3		14	**35**	2		*51*
4			9	**31**		*40*
5				6	**36**	*42*
	21	*75*	*48*	*39*	*36*	***219***

Solo validazione

CCI=66.7%

	1	2	3	4	5	
1	**5**	7				*12*
2	2	**15**	1			*18*
3		5	**6**	2		*13*
4			2	**11**		*13*
5				4	**9**	*13*
	7	*27*	*9*	*17*	*9*	***69***

Analisi di sensibilità



In conclusione…

- Il problema dei problemi: l’informazione disponibile in materia ambientale è drammaticamente carente.
- Non soluzioni “chiavi in mano”, ma cornici metodologiche che facilitano la collaborazione ed ottimizzano l’uso dei dati disponibili.
- Strumenti di nuova generazione, ma risultati sempre e costantemente a validati sul campo.
- Sovrasemplificare la complessità dei problemi ambientali **oggi** significa pagare prezzi scientifici, sociali ed economici **domani**.

mscardi@mclink.it
tancioni@uniroma2.it

URL: **http://www.mare-net.com/mscardi/**

Sei interessato alle applicazioni di Intelligenza Artificiale e Machine Learning in Ecologia?

- **www.isei3.org**
- **www.isei4.org**
- **http://www.isei5-conference.elsevier.com**
- **Ecological Informatics (Elsevier)**