

Modellistica ambientale: applicazioni e prospettive

Prof. Michele Scardi

*Dipartimento di Biologia
Università di Roma "Tor Vergata"*

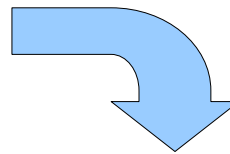
<http://www.mare-net.com/mscardi>

mscardi@mclink.it

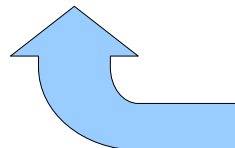
sistema reale



astrazione



modello



interpretazione

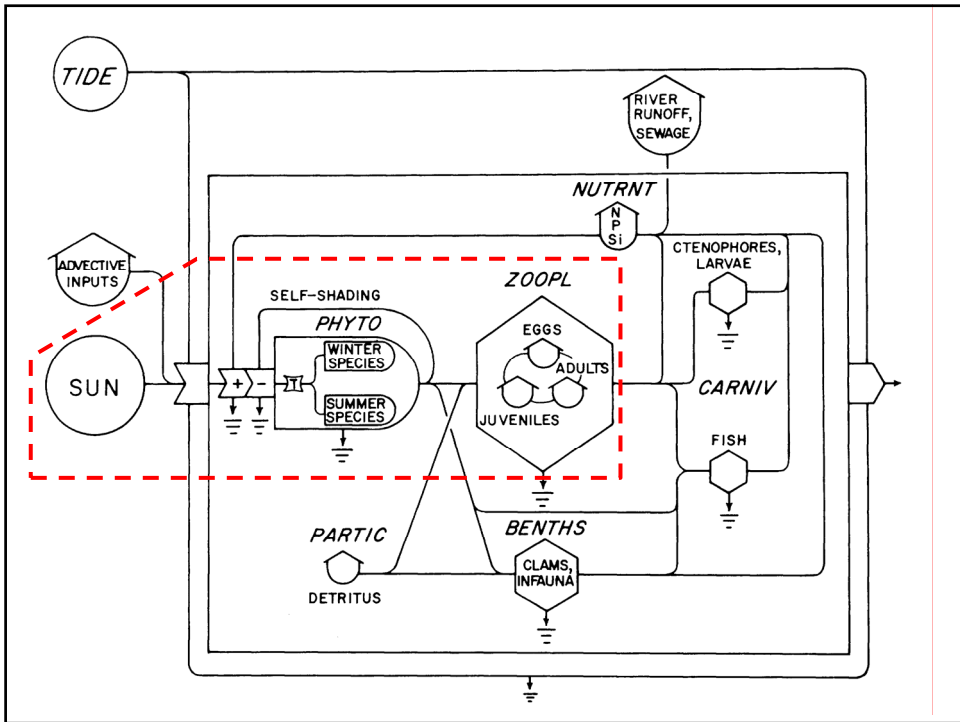
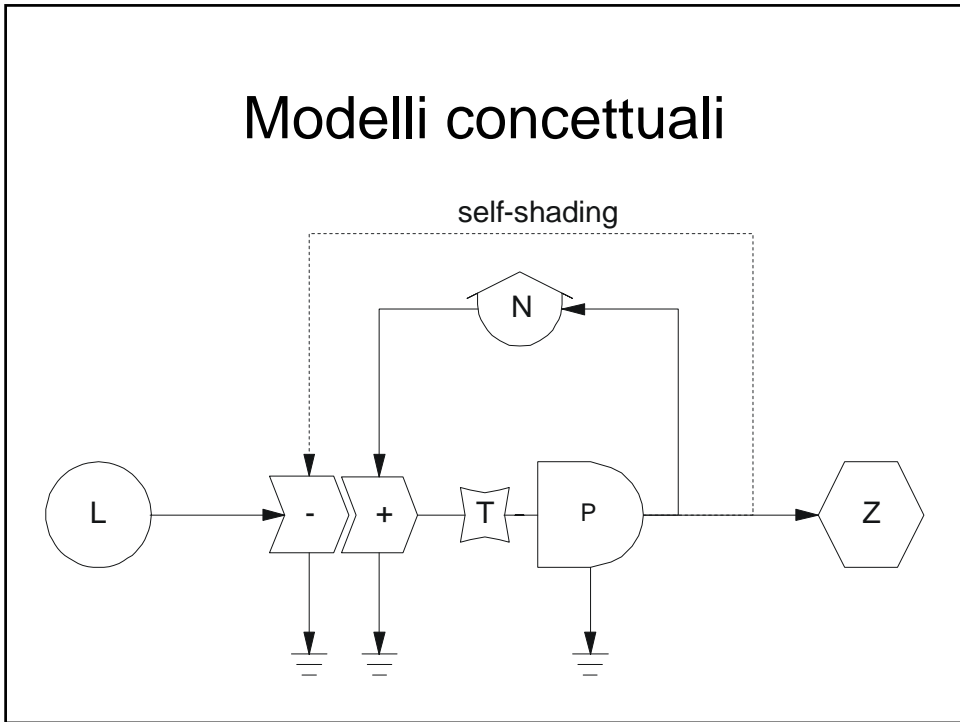
Tipologie di modelli

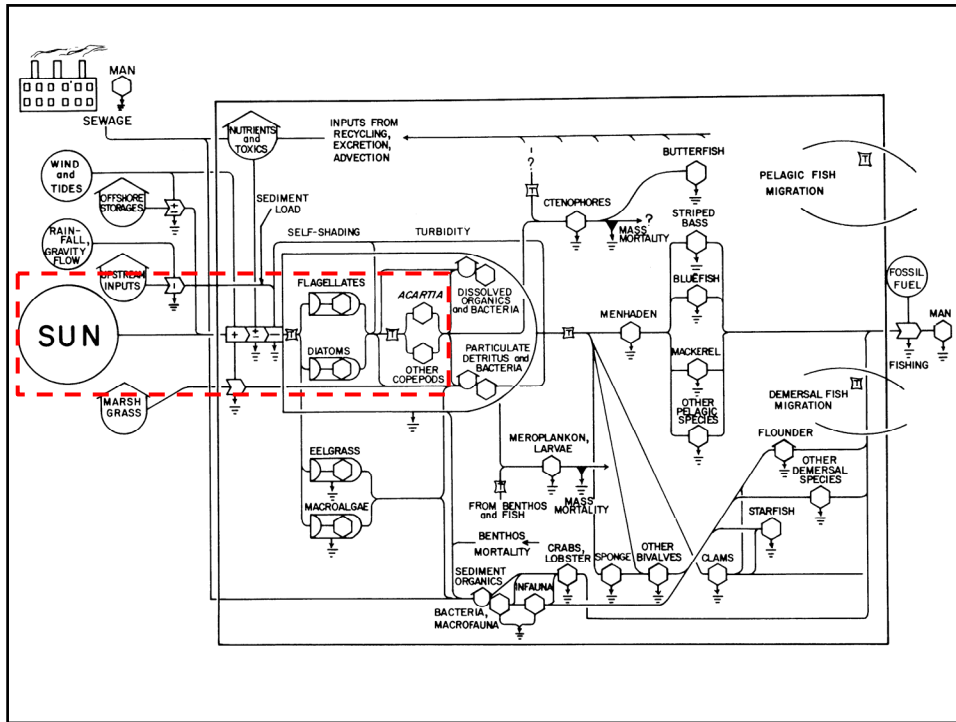
- **Modelli fisici**
(es. mesocosmi)
- **Modelli concettuali**
(es. modello di Tilman)
- **Modelli analitici**
(es. modello di Lotka-Volterra)
- **Modelli empirici**
(es. modello di Von Bertalanffy)

Modelli fisici

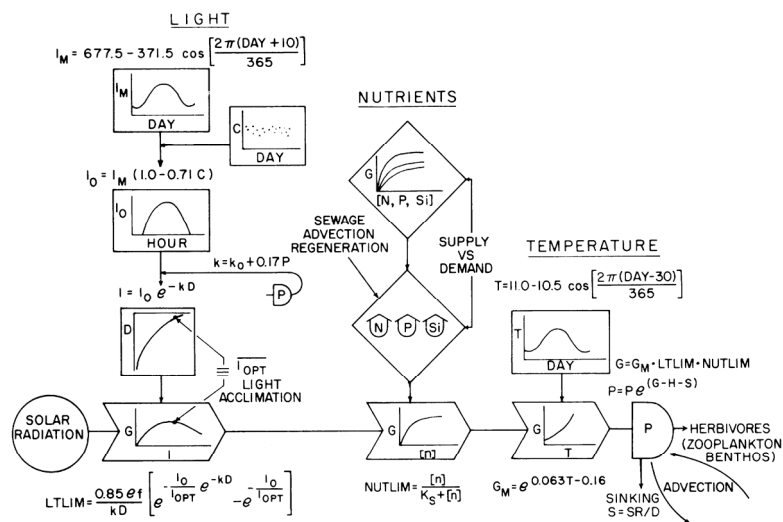


Modelli concettuali





Modelli analitici



$$\frac{dP}{dt} = f(I, T, N)P - g(P, T)Z - c_{mp}P \pm \text{scambi}$$

$$\frac{dZ}{dt} = g(P, OM, T)Z - h(T)Z - c_{na}g(P, T)Z - c_{mz}Z \pm \text{scambi}$$

$$\frac{dOM}{dt} = c_{np}P + c_{na}g(P, T)Z - k(T)OM - c_sOM \pm \text{scambi}$$

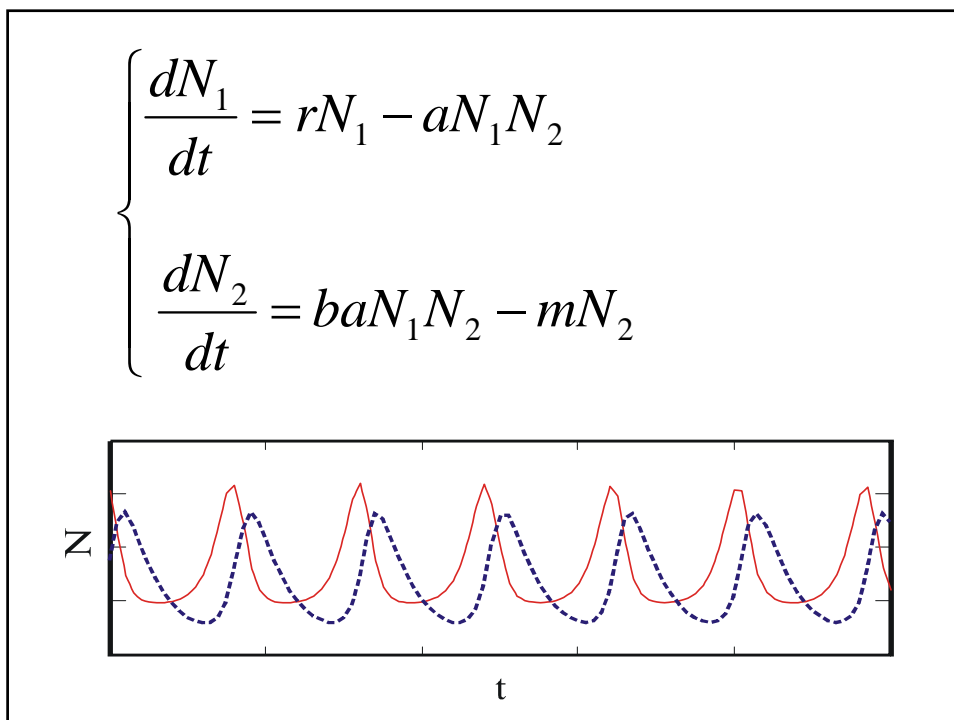
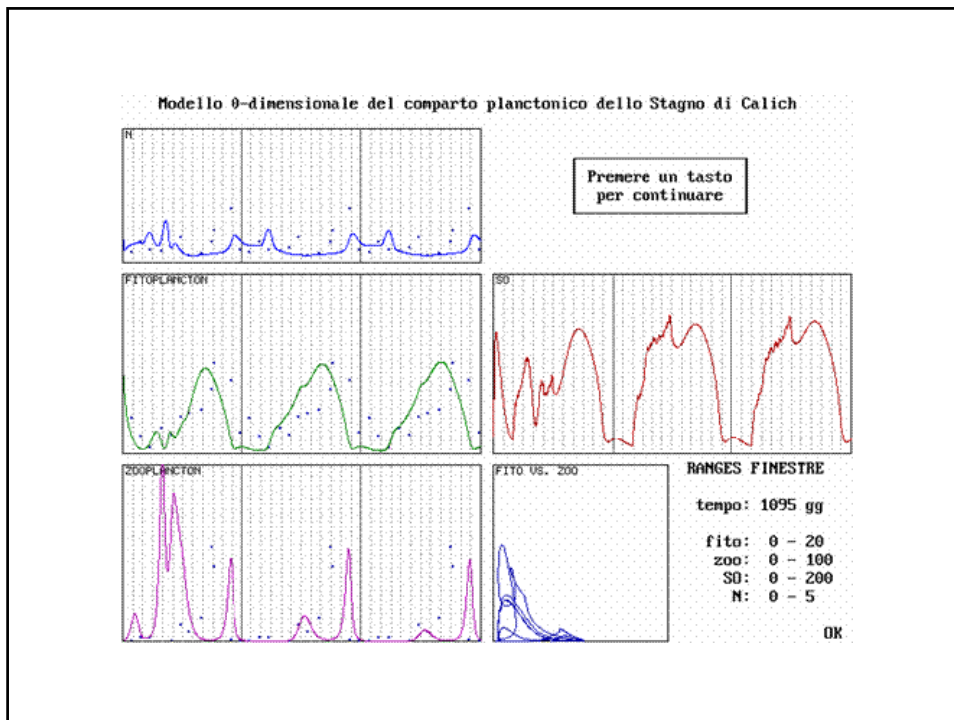
$$\frac{dN}{dt} = 12 \dots P, T)Z + k(T)OM - f(I, T, N)P \pm \text{scambi}$$

$$g(P, OM, T) = g_{\max} \cdot [g_1(P) + g_2(OM)] \cdot g_3(T)$$

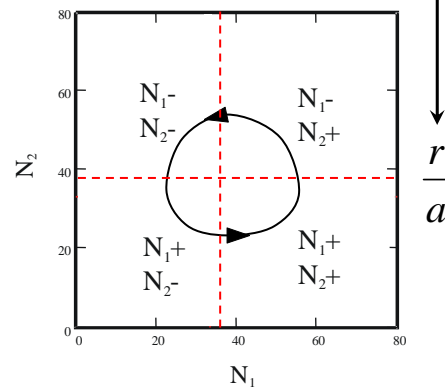
$$g_3(T) = \frac{T}{T_{opt}} e^{1 - \frac{T}{T_{opt}}}$$

Parametri del modello (assunti come costanti)

densità iniziale fitoplancton (Chl mg m ⁻³)	10
densità iniziale zooplancton (DW mg m ⁻³)	0.2
concentrazione iniziale SOM (mg C m ⁻³)	20
conc. iniziale azoto inorg. (mM N dm ⁻³)	1
numero di giorni simulati dall'1 gennaio	1460
torbidità relativa (componente inorganica)	2
temperatura massima acqua (°C)	26
temperatura minima acqua (°C)	10
tasso di crescita massimo del fitoplancton	0.59
pendenza della curva di crescita come f(T)	0.075
costante di semisaturazione per l'azoto	1.1
tasso di mortalità naturale fitoplancton	0.15
cost. conversione mg C dm ⁻³ -> mM Nxy dm ⁻³	0.0126
tasso di crescita max zooplancton erbivoro	1.2
costante di semisaturazione per il grazing	25
coeff. preferenza grazing su fitoplancton	0.66
coeff. preferenza grazing su part. org.	0.33
tasso di respirazione a 0 °C	0.06
pendenza curva respirazione [f(T)]	0.085
razione di fitoplancton non assimilata	0.33
tasso di mortalità dello zooplancton	0.21
tasso di decomposizione della SOM a 0 °C	0.05
pendenza funzione decomposizione [f(T)]	0.06
tasso di sedimentazione della SOM	0.33



$$\begin{cases} \frac{dN_1}{dt} = rN_1 - aN_1N_2 = 0 \text{ per } N_2 = \frac{r}{a} \\ \frac{dN_2}{dt} = baN_1N_2 - mN_2 = 0 \text{ per } N_1 = \frac{m}{ba} \end{cases}$$



```

Sub MODELLO(f(),Y(),r1,p1,r2,m2)
'prede
f(1) = r1 * Y(1) - p1 * Y(1) * Y(2)
'predatori
f(2) = r2 * p1 * Y(1) * Y(2) - m2 * Y(2)
End Sub

Static Sub PREDCORR(x0, x9, k1, n%, k5, e, z%, fl%)
ReDim w(1 To 9) As Single
ReDim h2(n%, 22) As Single
On Error GoTo ErrorHandler
'verifica le possibili condizioni di errore alla
chiamata
If x0 >= x9 Then fl = 1: Exit Sub
If (n% <= 0) Or (n% <> Int(n%)) Then fl = 2: Exit Sub
If (k5 <= 0) Or (k5 <> Int(k5)) Then fl = 3: Exit Sub
If (k1 <= 0) Or (k1 <> Int(k1)) Then fl = 4: Exit Sub
If e < 0.000000001 Then fl = 5: Exit Sub
If (z - 1) * (z - 2) <> 0 Then fl = 6: Exit Sub
'definizione delle costanti
w(1) = 0.2071067811865
w(2) = 0.2928932188135
w(3) = 0.1
'inizializzazione
For i% = 1 to n%
    For j% = x0 to x9
    ...

```

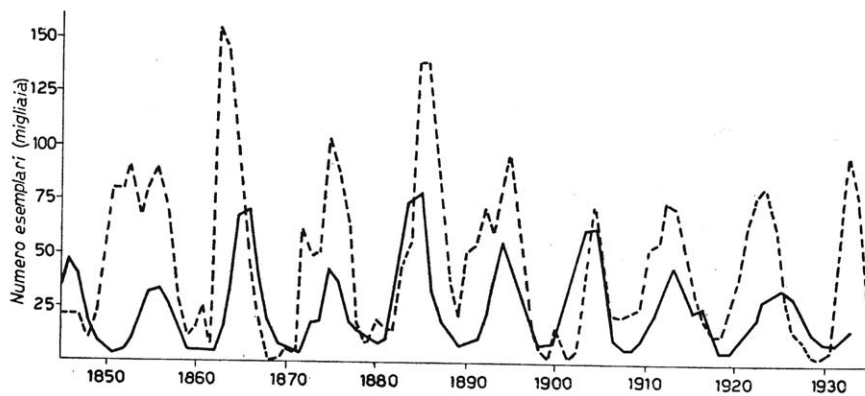
$$\frac{dN_1}{dt} = rN_1 - aN_1N_2$$

Skip

$$\frac{dN_2}{dt} = baN_1N_2 - mN_2$$

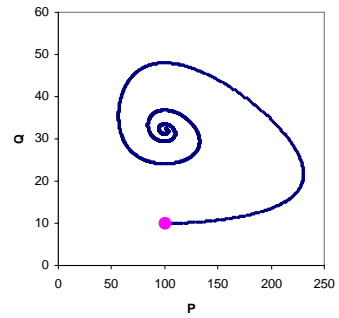
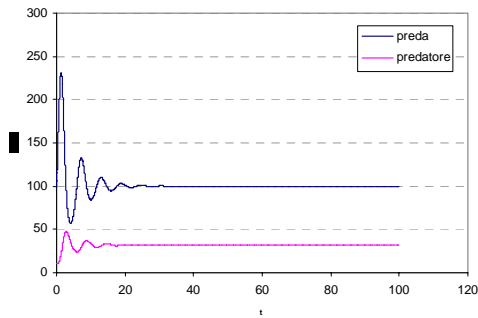
	A	B	C	D	E	F
1	P(0)	100	t	P(t)	Q(t)	
2	Q(0)	20		0	100	20
3	a	0.05		0.01	101	20.05
4	r	2		0.02	102.007	20.1016
5	b	0.15		0.03	103.022	20.1549
6	m	0.5		0.04	104.045	20.2099
7				0.05	105.074	20.2665
8	passo	0.01		0.06	106.111	20.3249
9				0.07	107.155	20.385
10				0.08	108.206	20.4469
11				0.09	109.264	20.5106
12				0.1	110.328	20.5762
13				0.11	111.4	20.6435
14				0.12	112.478	20.7128
15				0.13	113.563	20.784

http://www.mare-net/mscardi/work/ecologia/mod_preda_predatore.xls



$$\begin{cases} \frac{dN_1}{dt} = rN_1 \left(1 - \frac{N_1}{K} \right) - aN_1N_2 \\ \frac{dN_2}{dt} = baN_1N_2 - mN_2 \end{cases}$$

competizione
intraspecifica
fra prede



$$\frac{dN_1}{dt} = rN_1 \left(1 - \frac{N_1}{K} \right) - a_1N_1N_2$$

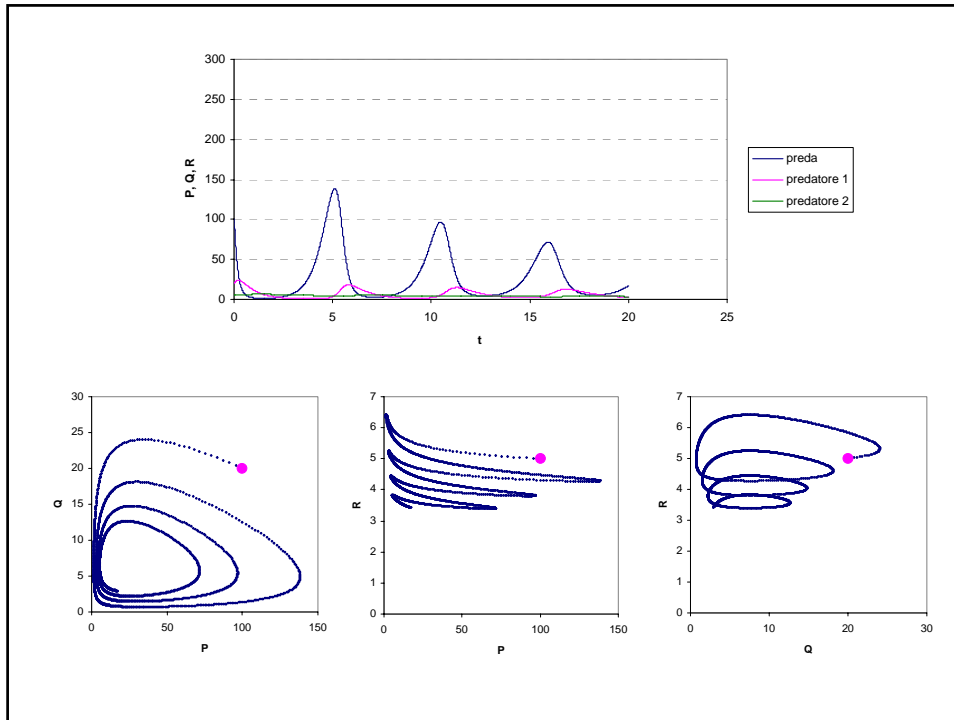
competizione
intraspecifica
fra prede

$$\frac{dN_2}{dt} = b_1a_1N_1N_2 - a_2N_2N_3$$

la mortalità dei
predatori dipende solo
da predatori di livello
superiore

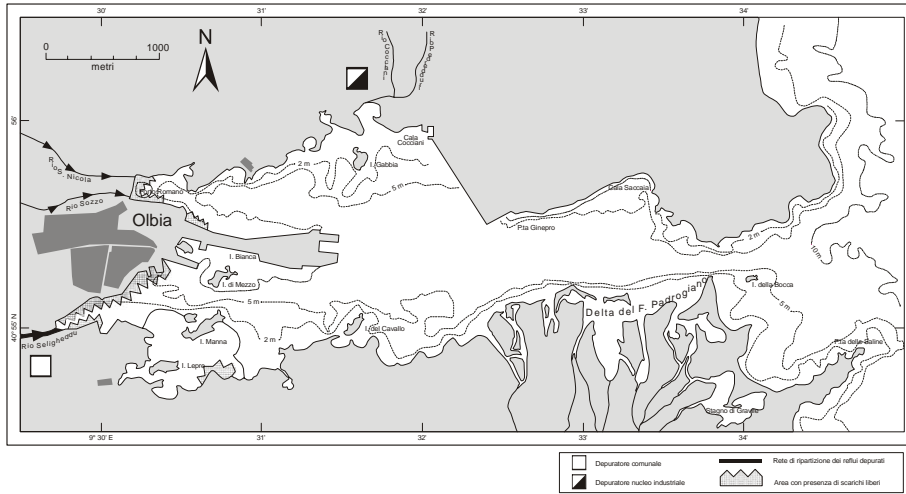
$$\frac{dN_3}{dt} = b_2a_2N_2N_3 - mN_3$$

al vertice della rete trofica,
la mortalità è solo per
senescenza (in questo
caso si assume il tasso di
mortalità come una
costante)



$$\left\{ \begin{array}{l} \frac{dN_1}{dt} = rN_1 \left(1 - \frac{N_1}{K} \right) - (a_1N_2 + a_2N_3)N_1 \\ \frac{dN_2}{dt} = b_1a_1N_1N_2 - m_1N_2 \\ \frac{dN_3}{dt} = b_2a_2N_1N_3 - m_2N_3 \end{array} \right.$$

Simulazione di scenari di gestione ambientale nel Golfo di Olbia



$$\frac{dP}{dt} = f(T, I, N)P - g(P, T)Z - c_{np}P \pm \text{scambio}$$

$$\frac{dZ}{dt} = (1 - c_{na}) \dots - h(T)Z - c_{mz}Z \pm \text{scambio}$$

$$\frac{dOM}{dt} = c_{mp}P + c_{\dots} \dots - k(T)OM - c_{\dots}OM \pm \text{scambio} + \text{reflui} + \text{escrezione}$$

$$\frac{dN}{dt} = c_{CN} \cdot (\dots) \pm \text{scambio} + \text{reflui} + \text{escrezione}$$

$$f(T, I, N) = \mu(T) \cdot f_1(I) \cdot f_2(N)$$

$$\mu(T) = \mu_{\max} \frac{e^{k_i T}}{e^{k_i T_{\max}}}$$

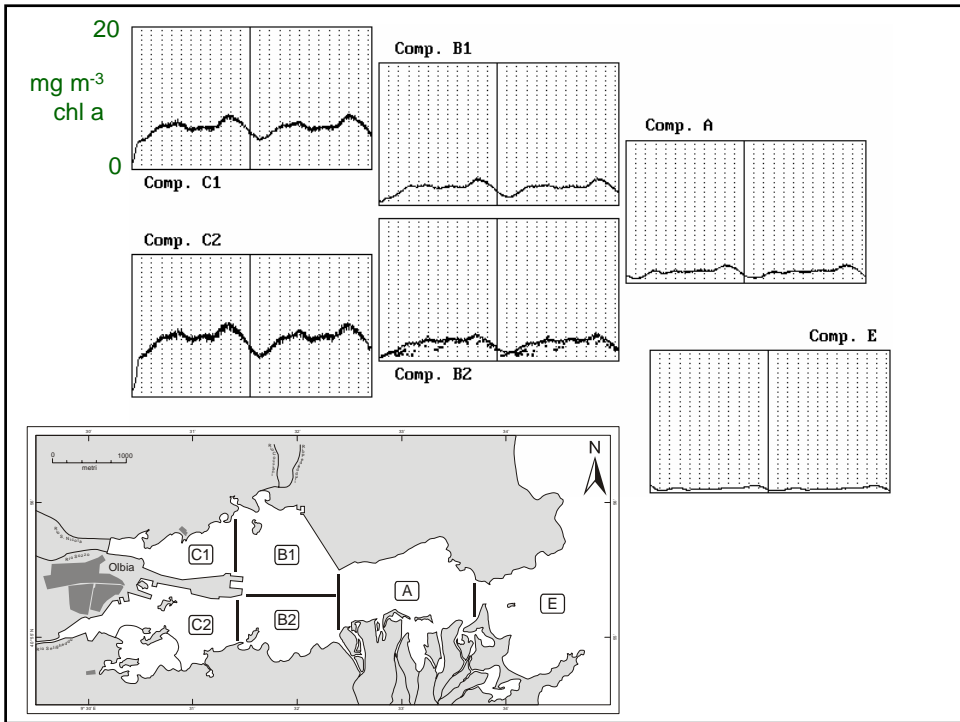
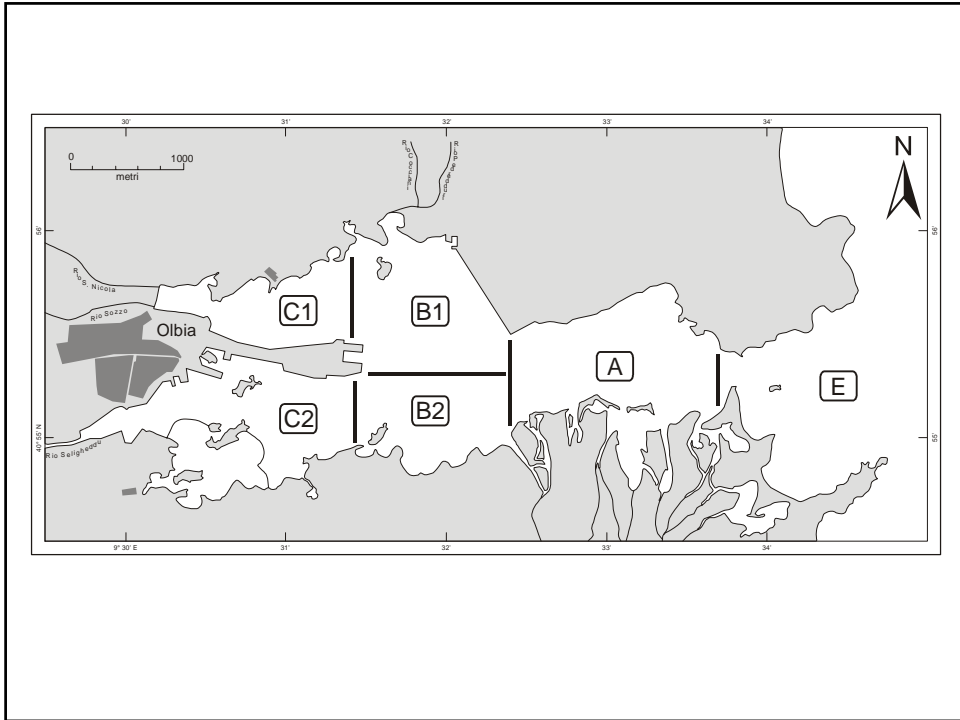
$$f_1(I) = \frac{\bar{I}}{I_{opt}} e^{1 - \frac{\bar{I}}{I_{opt}}}$$

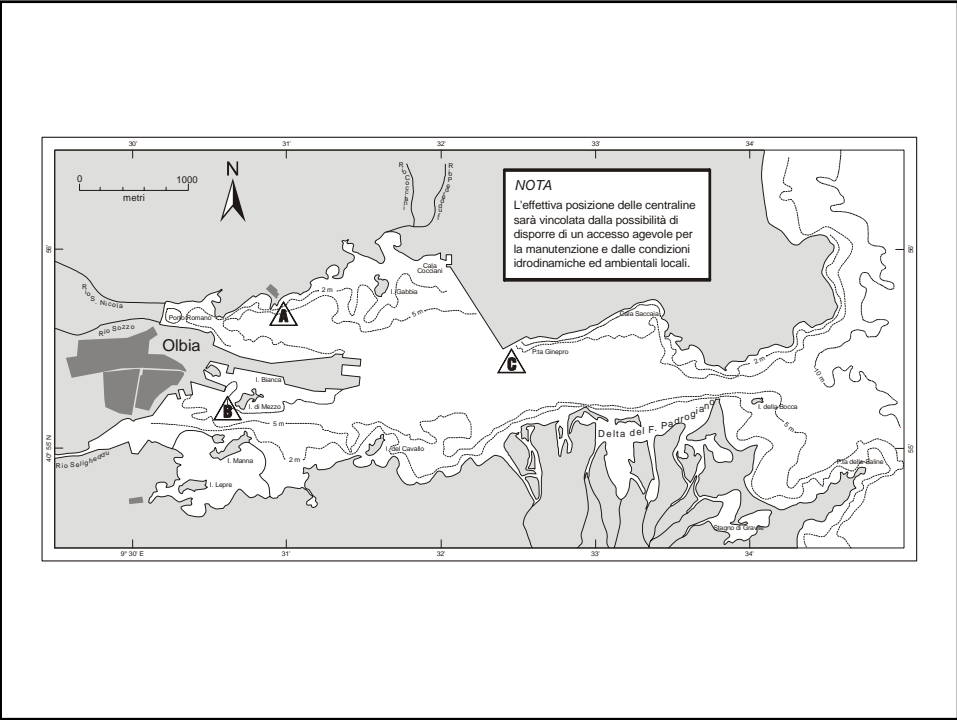
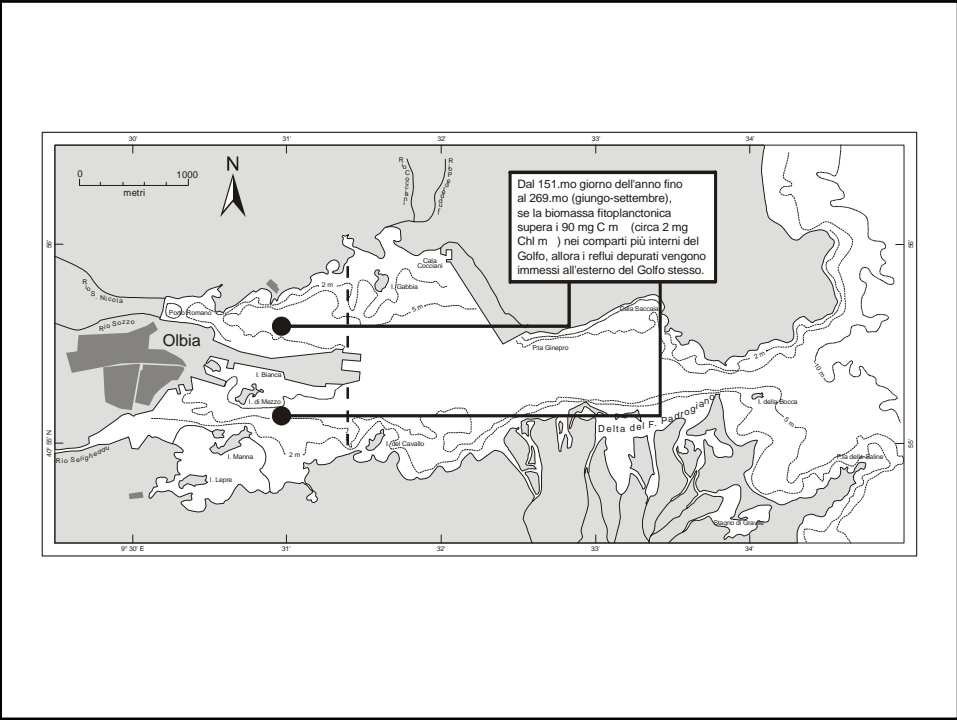
$$f_2(N) = \frac{N}{k_N + N}$$

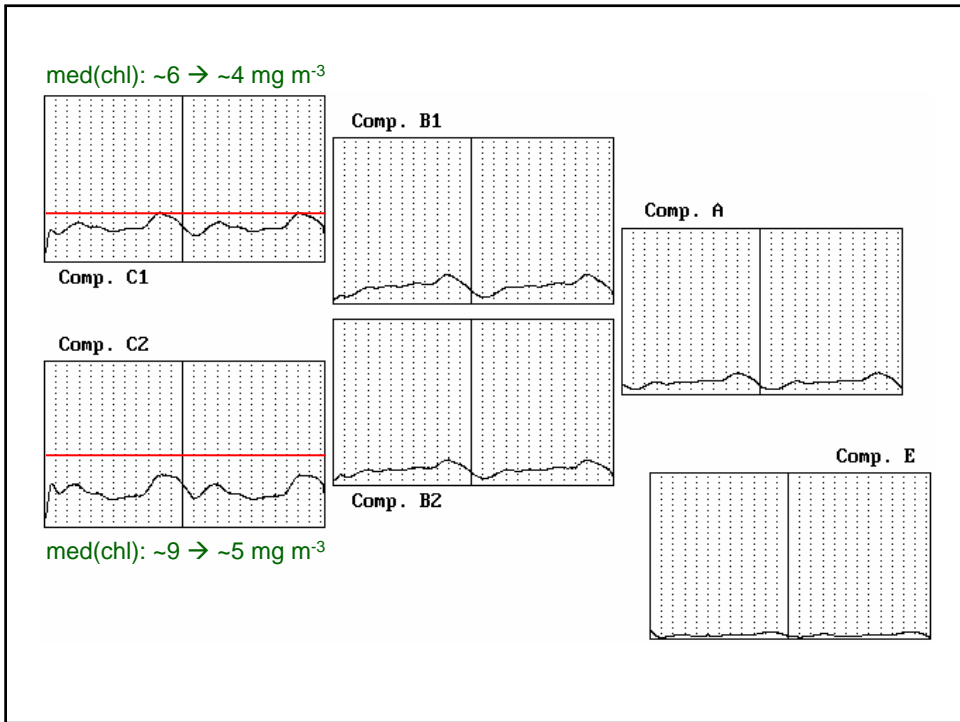
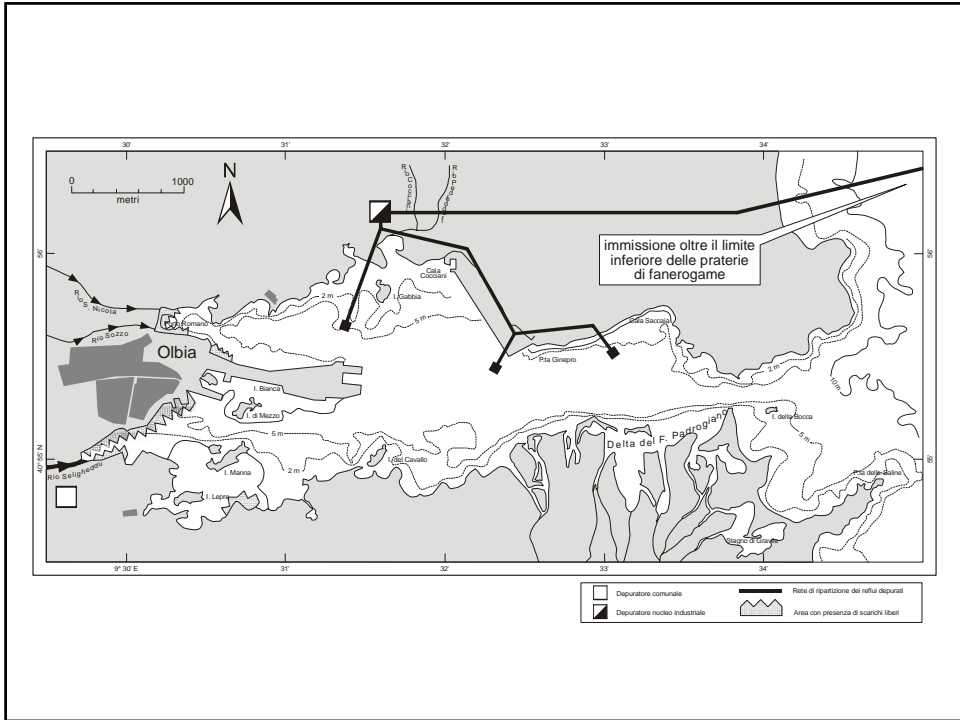
$$\bar{I} = \frac{1}{z} \int_0^z I_0 e^{-cz} dz$$

$$I_{opt} = 0.7 \cdot \bar{I}_{t-1} + 0.2 \cdot \bar{I}_{t-2} + 0.1 \cdot \bar{I}_{t-3}$$

$$c = 0.35 + 0.05 \cdot P_{chl}$$





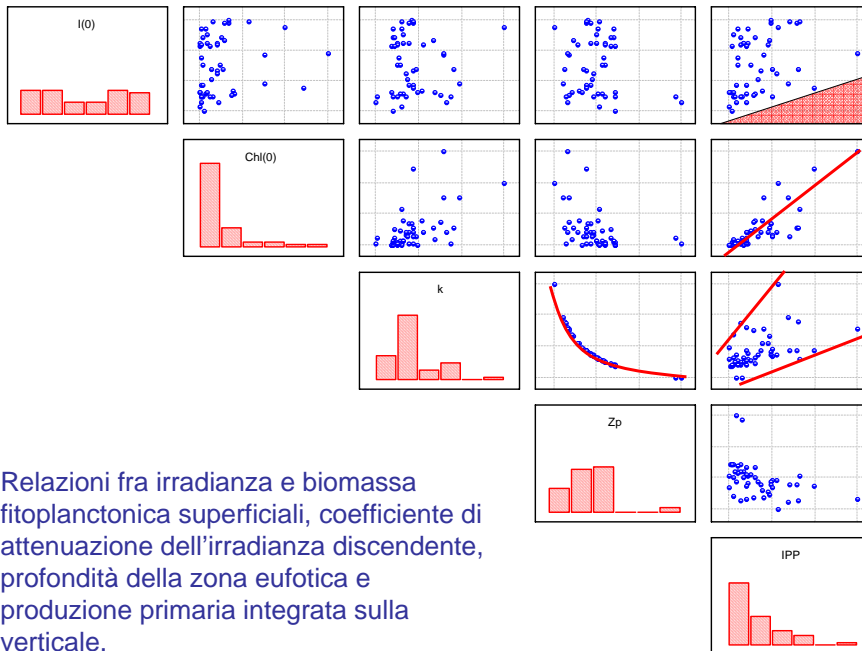


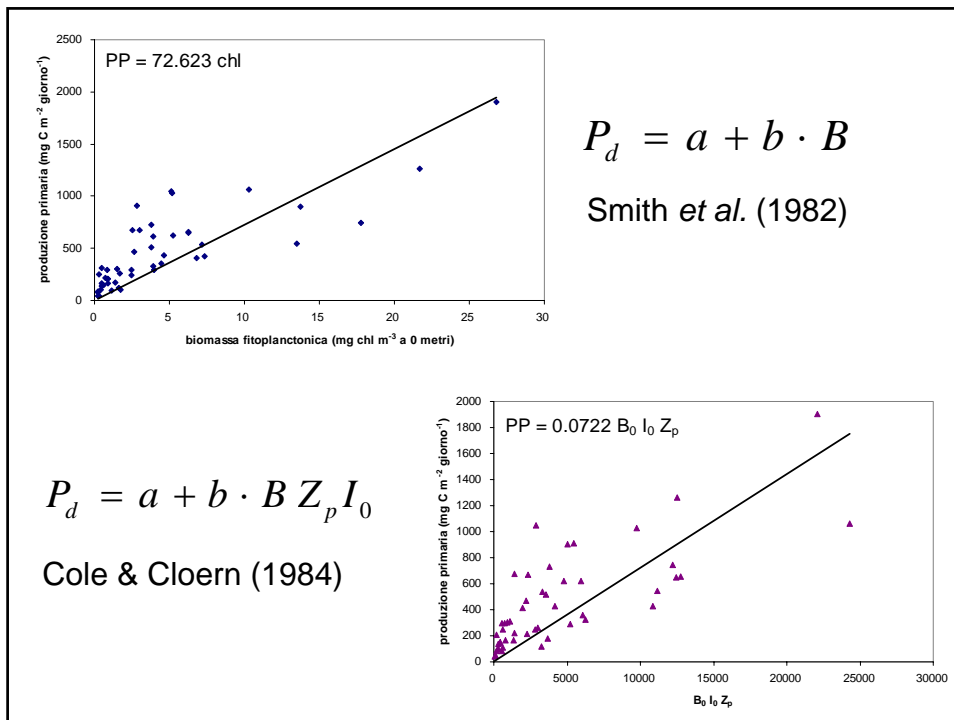
Modelli empirici

Stima della produzione primaria

- La produzione primaria dipende da:
 - Biomassa fitoplanctonica
 - Irradianza
 - Temperatura
 - Efficienza fotosintetica del fitoplancton
- La biomassa fitoplanctonica dipende da:
 - Produttività primaria
 - Disponibilità di nutrienti
 - Scambi advettivi

data	$I_0(\text{PAR})$ $Em^{-2}d^{-1}$	$CHL_m(0-2)$ $mg\ m^{-3}$	k m^{-1}	z_p m	$P_T(0-60)$ $mg\ C\ m^{-2}\ d^{-1}$
02/09/84	11.831	3.04	0.113	40.8	678
03/01/84	15.110	0.25	0.087	53.0	80
03/15/84	30.966	0.93	0.100	46.1	168
03/29/84	13.783	7.33	0.112	41.2	426
04/12/84	37.119	13.52	0.127	36.3	541
04/26/84	44.005	3.93	0.127	36.3	325
05/09/84	45.180	5.27	0.128	36.0	620
05/24/84	55.343	6.26	0.109	42.3	657
06/06/84	12.762	6.83	0.123	37.5	410
06/28/84	57.094	6.25	0.126	36.6	646
07/10/84	55.333	5.20	0.119	38.7	1027
07/24/84	52.449	0.51	0.111	41.5	313
08/08/84	47.360	0.92	0.089	51.8	210
08/21/84	43.527	0.32	0.094	49.0	247
09/07/84	42.333	0.25	0.087	53.0	85
09/18/84	15.649	7.20	0.158	29.2	536
...





Vertically Generalized Production Model (VGPM)

(Behrenfeld & Falkowski, 1997)

$$\Sigma PP := 0.66125 \cdot P_{b_opt} \cdot \frac{I_0}{I_0 + 4.1} \cdot Chl \cdot Z_{eu} \cdot DI$$

Where:

$$Chl_{tot}(Chl) := \text{if}(Chl < 1.0, 38.0 \cdot Chl^{0.42}, Chl)$$

$$Z_{eu}(Chl) := \text{if}\left[568.2 \cdot \left(\frac{Chl_{tot}}{100}\right)^{-0.746}, 100\right]$$

$$P_{b_opt}(SST) := \text{if}(SST < -1, 0, \text{if}(SST < 10, 0.01, \text{if}(SST < 13.83, 0.1, 13.21)))$$

VGPM can be redefined as:

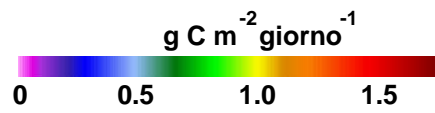
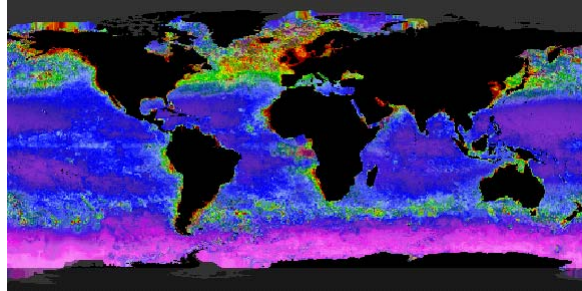
$$\Sigma PP(Chl, I_0, SST, DI) := 0.66125 \cdot P_{b_opt}(SST) \cdot \frac{I_0}{I_0 + 4.1} \cdot Chl \cdot Z_{eu}(Chl) \cdot DI$$

E.g.: $\Sigma PP(.601, 0.1, 13.83, 13.21) = 29.095$

$PP = f(B_0, Z_p, \text{fotoperiodo}, P_B^{opt})$
 $Z_p = f(B_0)$
 $P_B^{opt} = f(SST)$

More info about VGPM at <http://warrior.rutgers.edu>

PP media in giugno



Modelli empirici di nuova
generazione: reti neurali

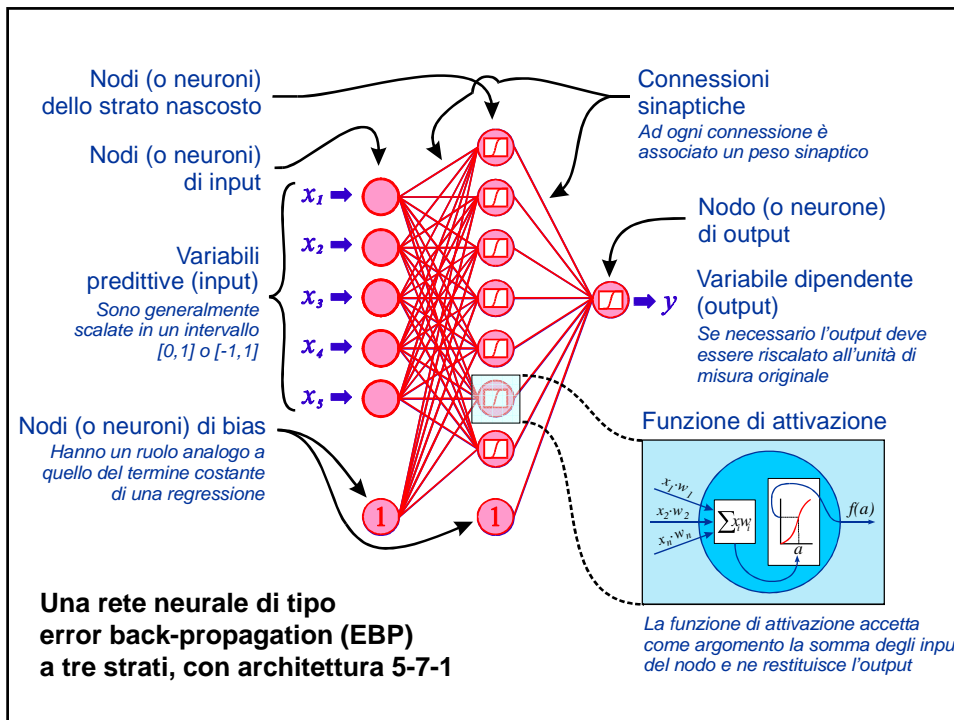
“...a neural network is a system composed of many simple processing elements operating in parallel whose function is determined by network structure, connection strengths, and the processing performed at computing elements or nodes.”

“...una rete neurale è un sistema composto da molti elementi di processo semplici che operano in parallelo, la cui funzione è determinata dalla struttura della rete, dall'intensità delle connessioni e dai processi attuati negli elementi di calcolo o nodi.”

*(DARPA Neural Network Study, 1988,
AFCEA International Press, p. 60)*

**COME FUNZIONANO
LE RETI NEURALI?**

SOLO UN PIZZICO DI TEORIA...



L'algoritmo EBP

L'algoritmo EBP (Rumelhart et al., 1986) è di gran lunga il più diffuso fra gli algoritmi di training per le reti neurali e può essere schematizzato come segue:

1. le connessioni sinaptiche sono inizializzate in maniera casuale

$$w_{jk} = \text{rnd}[0,1] \quad \begin{cases} j = 1, \dots, n_{i+1} \\ k = 1, \dots, n_h \end{cases}$$

$$z_{kl} = \text{rnd}[0,1] \quad \begin{cases} k = 1, \dots, n_{h+1} \\ l = 1, \dots, n_o \end{cases}$$

2. un training pattern è immesso nella rete neurale

$$h_k = \sum_{j=1}^{n_{i+1}} i_j w_{jk} \quad (k = 1, \dots, n_h) \quad h_k = f(h_k) \quad h_{n_h+1} = 1$$

$$h_k = \sum_{j=1}^{n_{i+1}} i_j w_{jk} \quad (k = 1, \dots, n_h) \quad h_k = f(h_k) \quad h_{n_h+1} = 1$$

3. tutti i pesi sinaptici vengono quindi modificati in funzione dello scarto rilevato tra outputs e valori noti (error-backpropagation)

$$z_{kl} = z_{kl} + \eta \delta_l^o h_k \quad (k = 1, \dots, n_{h+1}; l = 1, \dots, n_o)$$

$$\delta_l^o = (y_l - o_l) \cdot f'(o_l) = (y_l - o_l) \cdot o_l (1 - o_l)$$

$$w_{jk} = w_{jk} + \eta \delta_k^h i_j \quad (k = 1, \dots, n_h + 1; j = 1, \dots, n_{i+1})$$

$$\delta_k^h = f'(h_k) \cdot \sum_{l=1}^{n_o} \delta_l^o z_{kl} = h_k (1 - h_k) \cdot \sum_{l=1}^{n_o} \delta_l^o z_{kl}$$

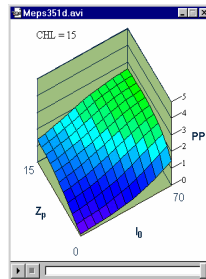
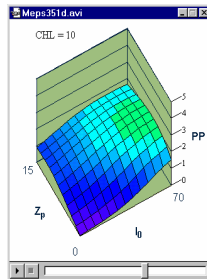
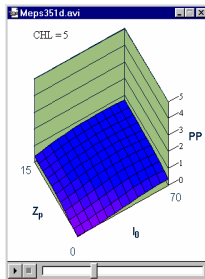
4. l'output della rete è confrontato con i valori noti del set di validazione

$$E = \frac{1}{n_o} \sum_{l=1}^{n_o} (y_l - o_l)^2$$

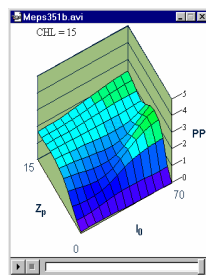
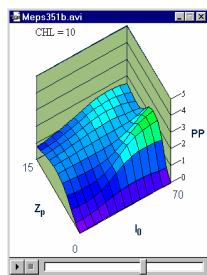
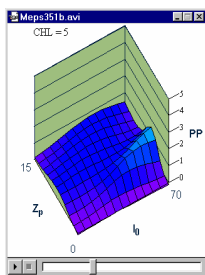
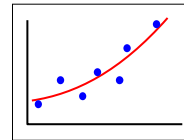
5. se le condizioni di convergenza sono raggiunte, si termina il training, altrimenti si torna al punto 2

Demo

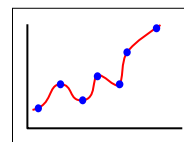
$$PP = f(l_0, Z_p, CHL)$$



generalizzazione



overfitting



PER EVITARE L'OVERFITTING:

- *early stopping*
- *jittering* (aggiunta di rumore agli inputs)
- *weight decay*
- *training patterns in ordine casuale*
- etc.

INOLTRE, PER UNA BUONA GENERALIZZAZIONE:

- gli inputs devono contenere abbastanza informazione predittiva in rapporto agli outputs desiderati (targets)
- la relazione da modellizzare deve essere preferibilmente regolare (cioè piccola variazione in input --> piccola variazione in output)
- il training set deve essere abbastanza grande e comunque deve essere un sottoinsieme rappresentativo del sistema reale

skip

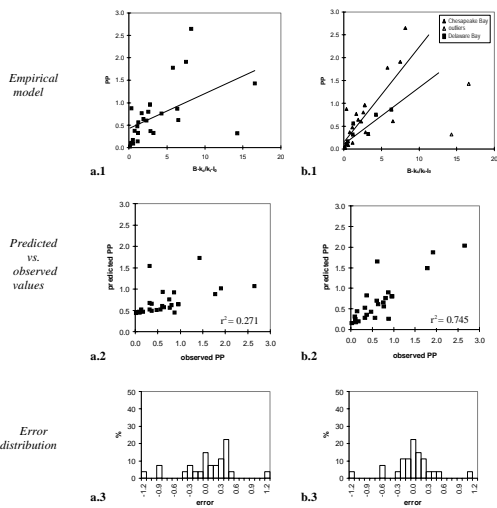
A PHYTOPLANKTON PRIMARY PRODUCTION MODEL FOR CHESAPEAKE BAY

Michele Scardi¹ & Lawrence W. Harding, Jr.²

1. Dept. of Biology, Univ. of Roma "Tor Vergata", Roma, Italy

2. Horn Point Lab., University of Maryland, USA

Chesapeake Bay



Phytoplankton primary production in Chesapeake Bay



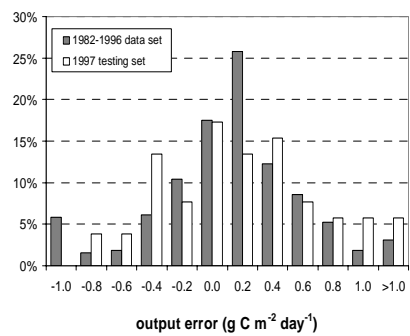
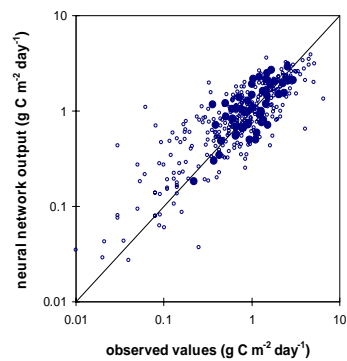
Predictive variables

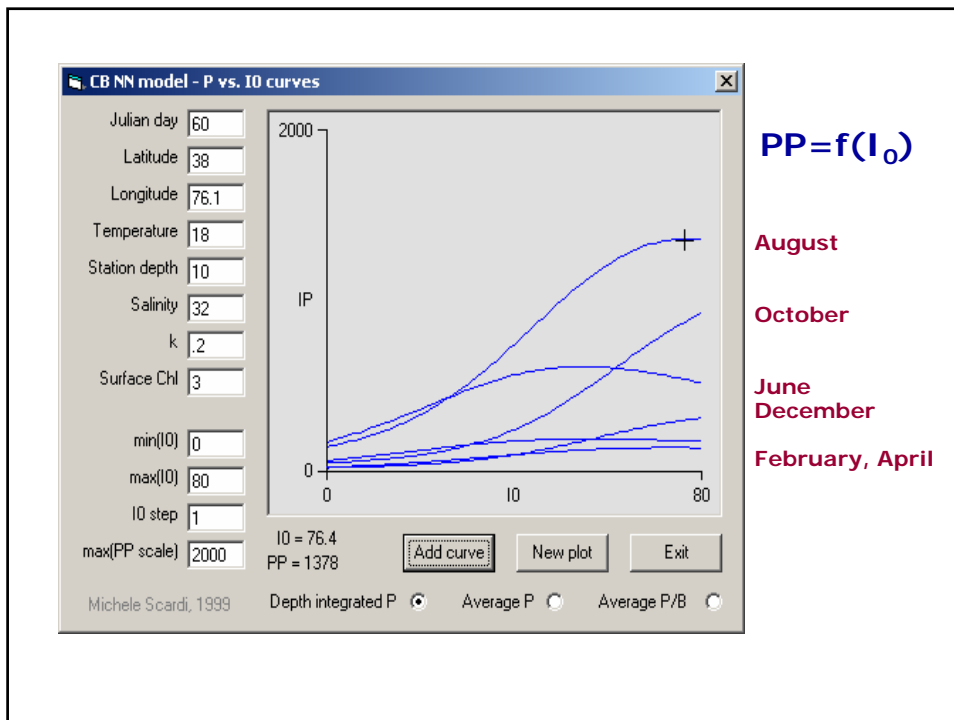
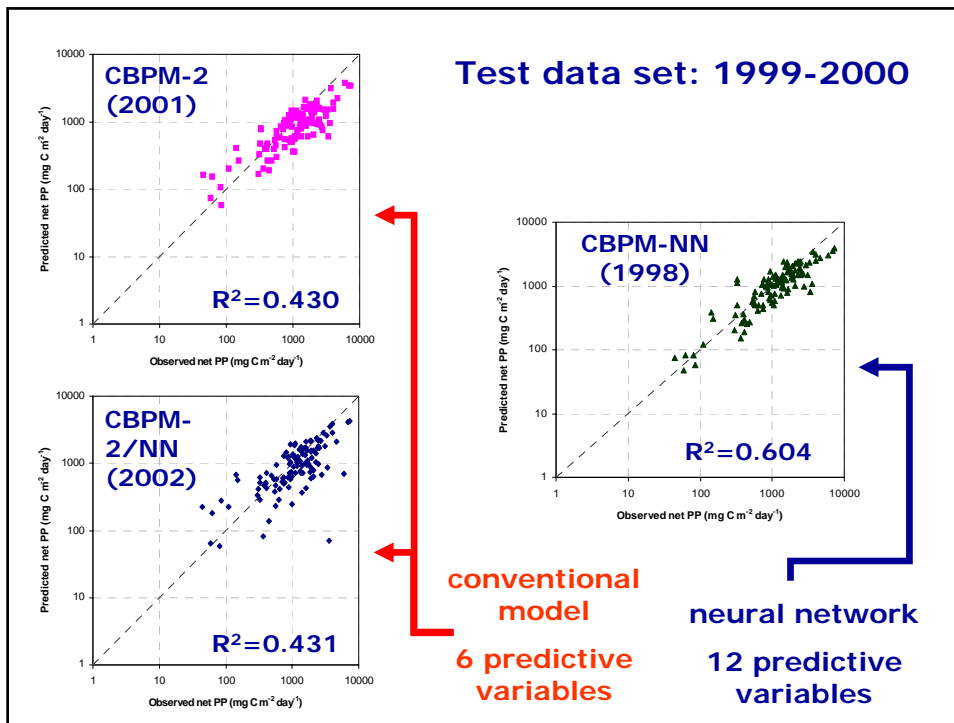
- $\frac{1}{2} \left[\cos \left(\frac{JulianDay}{2 \cdot 365 \cdot \pi} \right) + 1 \right]$
- $\frac{1}{2} \left[\sin \left(\frac{JulianDay}{2 \cdot 365 \cdot \pi} \right) + 1 \right]$
- latitude
- longitude
- depth
- water temperature
- salinity
- log chlorophyll
- log chlorophyll (Z_p)
- I_0 (PAR)
- k
- Z_p

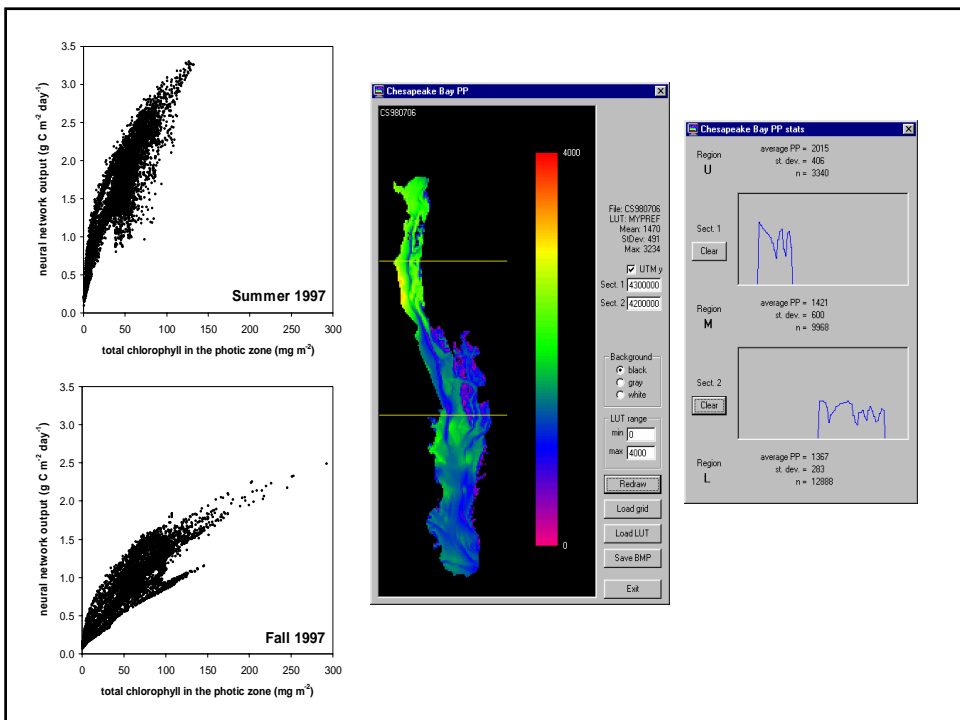
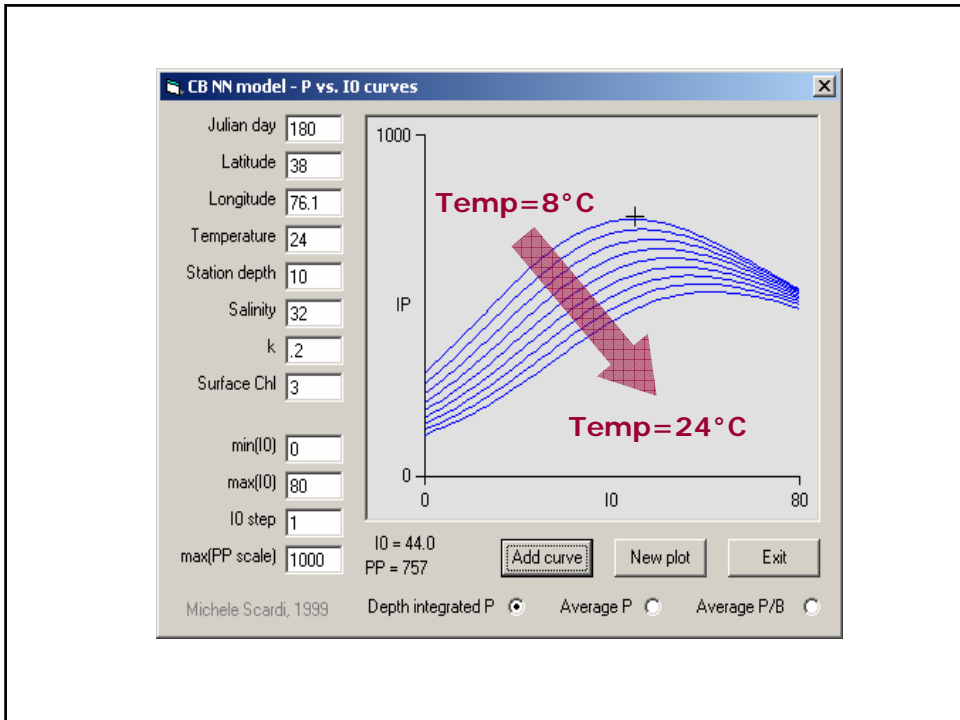
NN structure: 12 - 5 - 1

Output variable: primary production ($\text{mg C m}^{-2} \text{ day}^{-1}$)

- training and validation sets (1982-96)
- testing set (1997)







PREVISIONE DELLA STRUTTURA DELLA FAUNA ITTICA MEDIANTE RETI NEURALI ARTIFICIALI

Michele Scardi¹, Stefano Cataudella¹, Paola Di Dato¹,
Giuseppe Maio², Enrico Marconato², Stefano Salviati²,
Lorenzo Tancioni¹, Paolo Turin³ e Marco Zanetti³

1. Dip. di Biologia, Univ. di Roma "Tor Vergata", Roma

2. Aquaprogram s.r.l., Vicenza

3. Bioprogramm s.c.r.l., Padova

Problema applicativo #1:

- 1. Previsione della composizione della fauna ittica sulla base di **predittori** ambientali**
- 2. Comparazione della composizione **attesa** con quella **osservata****
- 3. Formulazione di un giudizio di **qualità ambientale** (es. *sensu* Direttiva 2000/60/CE)**

Problema applicativo #2:

- 1. Analisi di sensibilità** del modello previsionale
- 2. Identificazione dei fattori ambientali** che determinano le maggiori variazioni nella risposta del modello
- 3. Definizione di possibili strategie** per la **gestione** dell'ambiente

Previsione della composizione di una comunità

- Caso 1:** numero limitato di specie
- Caso 2:** molte specie, ma un singolo cenocline (in teoria, tutte le risposte sono unimodali)
- Case 3:** molte specie e più cenoclini (interazioni di ordine superiore fra variabili ambientali)
- Case 4:** elevatissimo numero di specie, cenospazio estremamente complesso con discontinuità

Caso 1 o 2: poche specie o singolo cenocline

- Buoni risultati si possono ottenere sia con metodi convenzionali (es. regressione logistica), sia con le **reti neurali**.
- Le **relazioni implicite** (non note) fra specie **riducono** significativamente la **dimensionalità reale** del problema.
- Solo le **reti neurali** possono trarre vantaggio da ciò.

Specie presenti nel data set:

32

Combinazioni possibili:

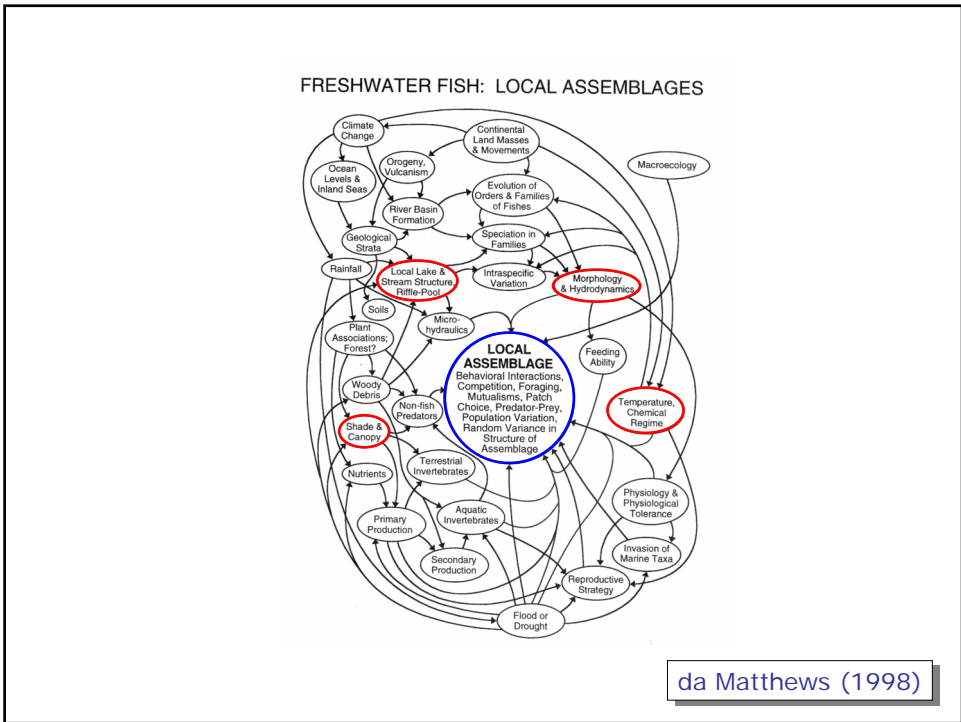
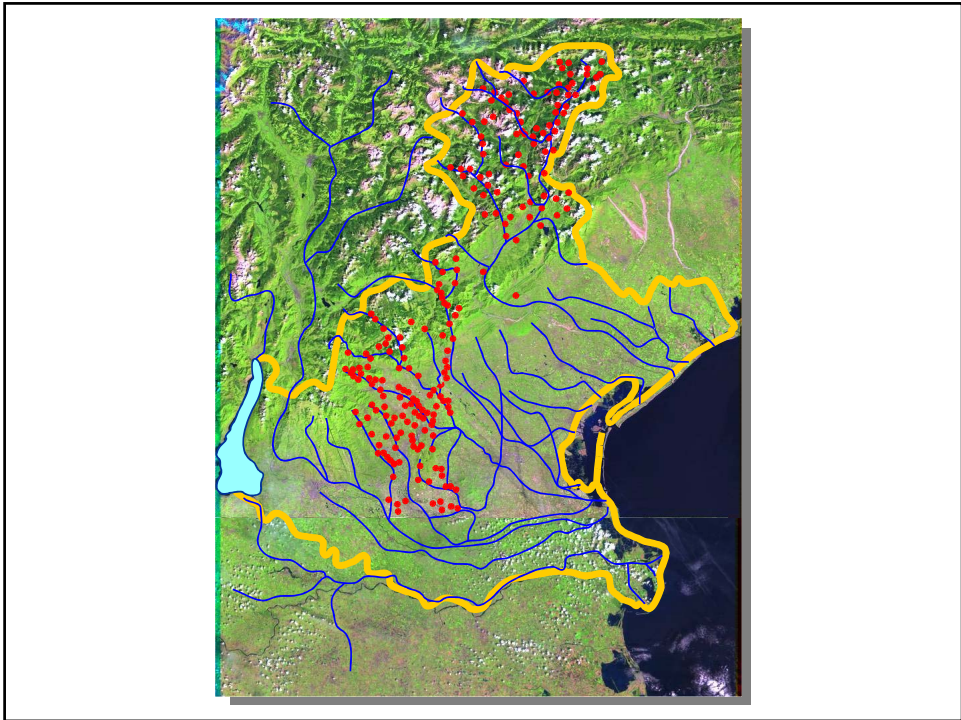
$2^{32} = 4294967296$

Combinazioni osservate:

131 su 264 casi

Le specie non sono indipendenti le une dalle altre (meno male, se no, poveri ecologi!)

Una rete neurale può "catturare" l'informazione relativa alle relazioni interspecifiche (es. competizione), migliorando la sua capacità predittiva



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 ($\mu\text{S}/\text{cm}$)
- 18 gradiente (%)
- 19 bacino versante (km^2)
- 20 distanza dalla sorgente (km)

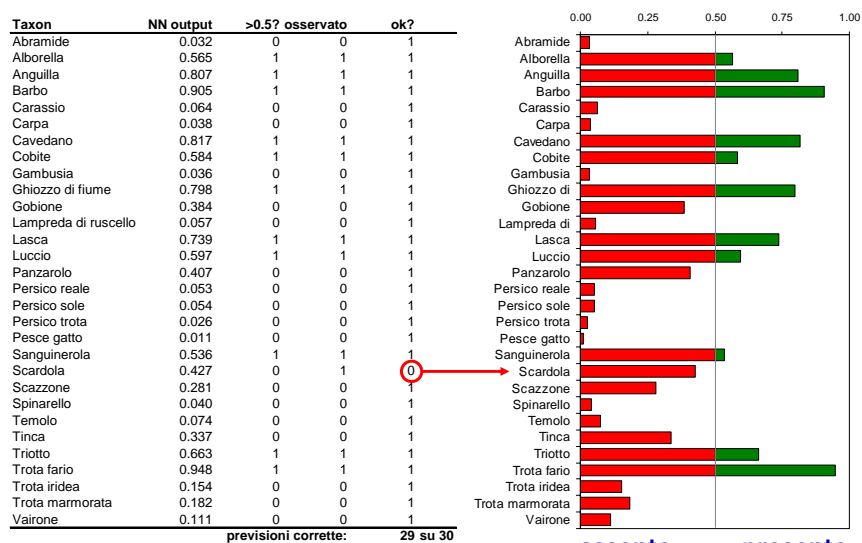
Taxa considerati

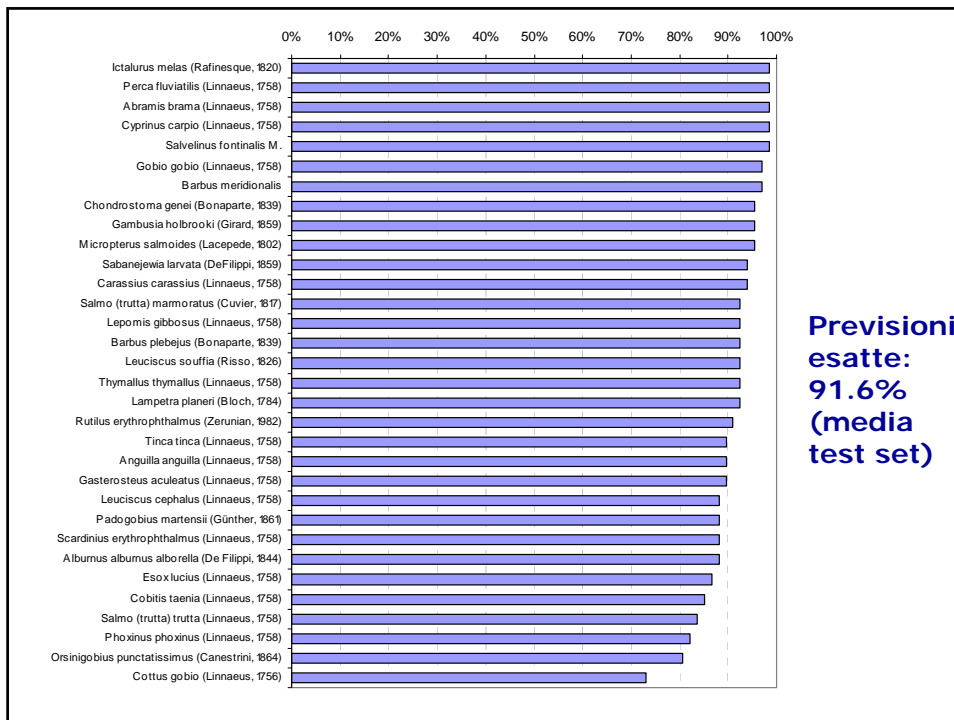
<i>Abramis brama</i>	<i>Leuciscus cephalus</i>
<i>Alburnus alburnus alborella</i>	<i>Leuciscus souffia</i>
<i>Anguilla anguilla</i>	<i>Micropterus salmoides</i>
<i>Barbus meridionalis</i>	<i>Orsinigobius punctatissimus</i>
<i>Barbus plebejus</i>	<i>Padogobius martensii</i>
<i>Carassius carassius</i>	<i>Perca fluviatilis</i>
<i>Chondrostoma genei</i>	<i>Phoxinus phoxinus</i>
<i>Cobitis taenia</i>	<i>Rutilus erythrophthalmus</i>
<i>Cottus gobio</i>	<i>Sabanejewia larvata</i>
<i>Cyprinus carpio</i>	<i>Salmo (trutta) marmoratus</i>
<i>Esox lucius</i>	<i>Salmo (trutta) trutta</i>
<i>Gambusia holbrooki</i>	<i>Salvelinus fontinalis</i>
<i>Gasterosteus aculeatus</i>	<i>Scardinius erythrophthalmus</i>
<i>Gobio gobio</i>	<i>Thymallus thymallus</i>
<i>Ictalurus melas</i>	<i>Tinca tinca</i>
<i>Lampetra planeri</i>	<i>Salmo (trutta) hybr. trutta/marmoratus</i>
<i>Lepomis gibbosus</i>	<i>Oncorhynchus mykiss</i>

Struttura del modello: 20-17-32

- 264 patterns (campioni)
 - training, n=131
 - validazione, n=66
 - test, n=67
- 20 variabili ambientali predittive
- 32 specie (dati binari, presenza/assenza)
- training della rete neurale: algoritmo di error back-propagation con early stopping basato sull'errore del set di validazione

Un esempio di output





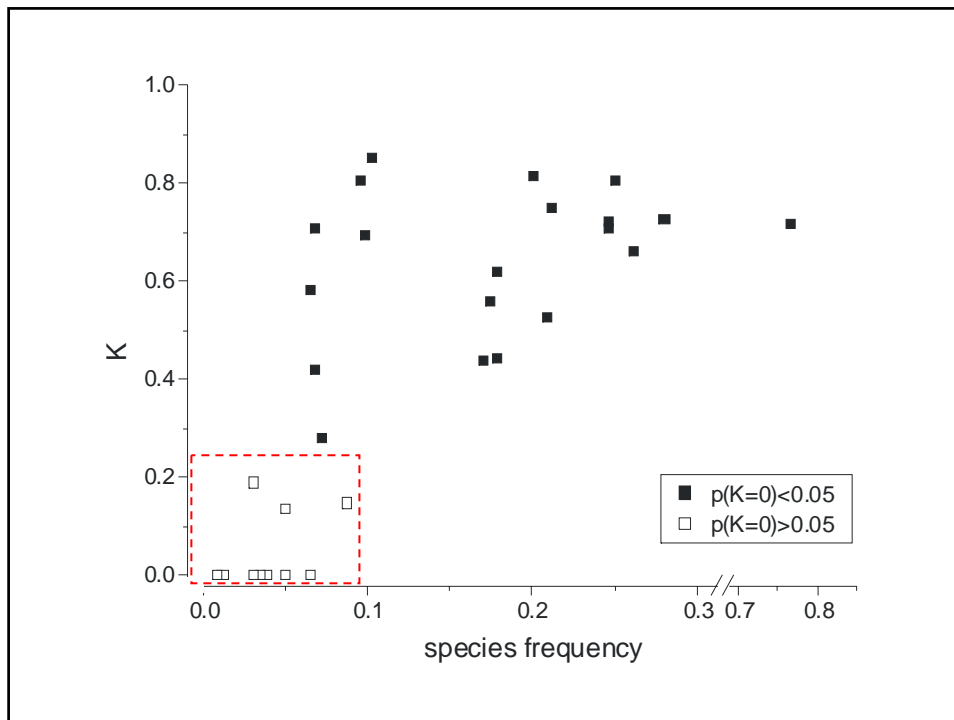
The K statistics

		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

$$K = \frac{O_a - E_a}{N - E_a}$$

O_a = observed count of matches
 E_a = expected count of matches
 N = total number of cases



Come migliorare le basi di dati

- Una copertura spaziale **omogenea** è **utile**, ma non **sufficiente**
- Il piano di campionamento deve considerare **più scale spaziali**
- Occorrono più osservazioni
- Occorrono più osservazioni
- Occorrono più osservazioni

Come migliorare l'apprendimento

- Escludere specie, taxa, classi, cenotipi, etc. la cui **frequenza** nei set di training, validazione e test sia **troppo alta o troppo bassa**
- Usare **criteri alternativi** all'MSE per misurare l'errore.
- Utilizzare **regole ecologiche** per vincolare l'apprendimento

Come misurare l'errore

Con dati binari (presenza/assenza):

◆ Per singoli output, su tutti i patterns:

- **numero (or percentuale) di previsioni esatte**
- **coefficiente Kappa**
- **correlazione tetracorica**

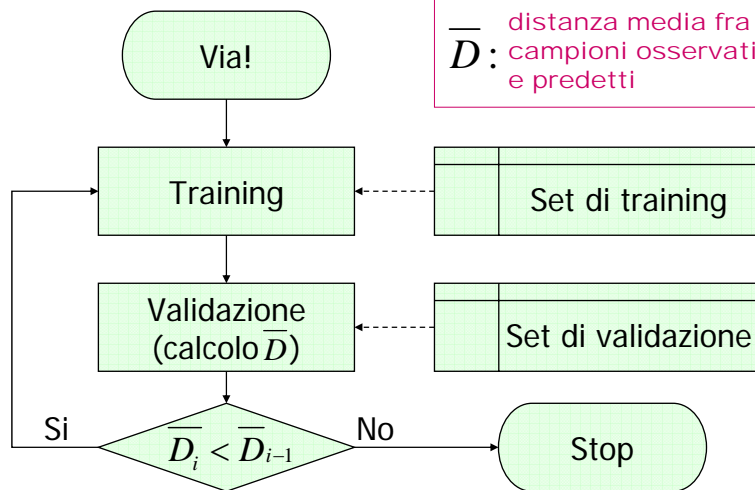
◆ Tutti gli output per singolo pattern:

- **indici di distanza/similarità**

◆ Tutti gli outputs su tutti i patterns (totale):

- ▶ **distanza/similarità media**
- **statistica di Mantel**

Training di una EBP NN basato su misure di distanza



Misurare la distanza fra patterns (campioni)

- Sia la presenza che l'assenza predette sono rilevanti, quindi è necessario un **indice simmetrico**.
- es. la dissimilarità di **Rogers & Tanimoto**:

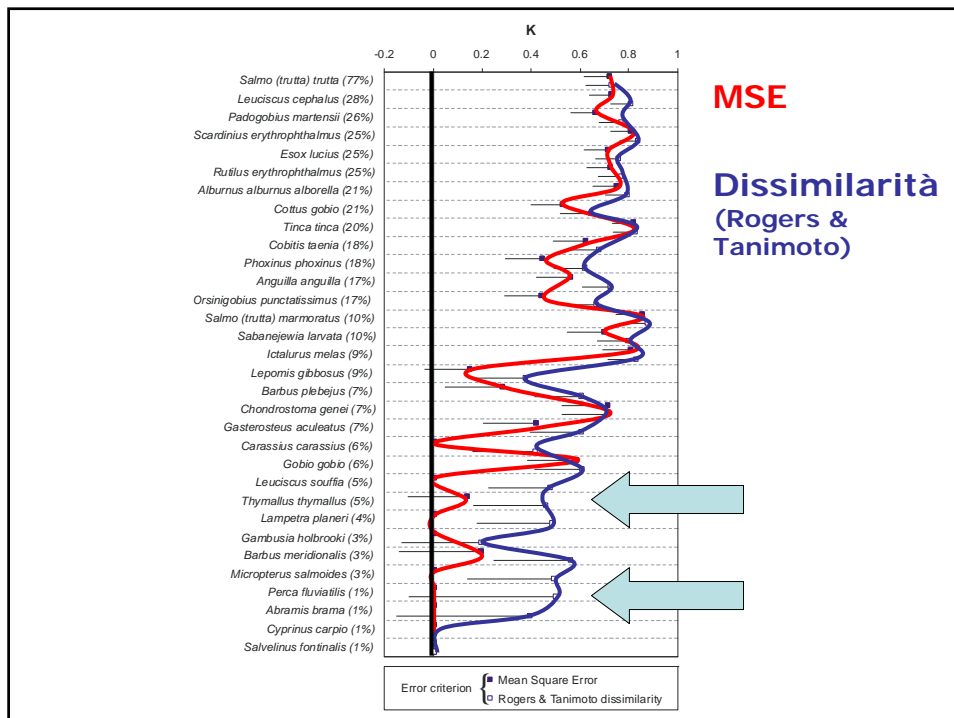
$$D = 1 - \frac{a + d}{a + 2b + 2c + d}$$

N.B. Le discordanze pesano più delle concordanze in questo indice (particolarità utile se le assenze sono molto più frequenti delle presenze nei dati o viceversa)

Risultati di un training basato su una misura di distanza

<i>Predizioni esatte</i>	<i>Distanza</i>	<i>MSE</i>
totale	94.4%	93.1%
training+validazione	95.4%	93.6%
test	91.8%	91.6%

Inoltre, le specie la cui **presenza** non è stata **mai predetta** dal modello sono **solo 2 e molto rare** (carpa e salmerino, presenti due sole volte ciascuna nel test set). Con un training convenzionale, basato sull'**MSE**, sono ben **9**.



UNO STRUMENTO PIU' COMPLESSO

skip

EU 5th Framework Programme

PAEQANN

Contract n°:

EVK1-CT1999-00026

Proposal number:

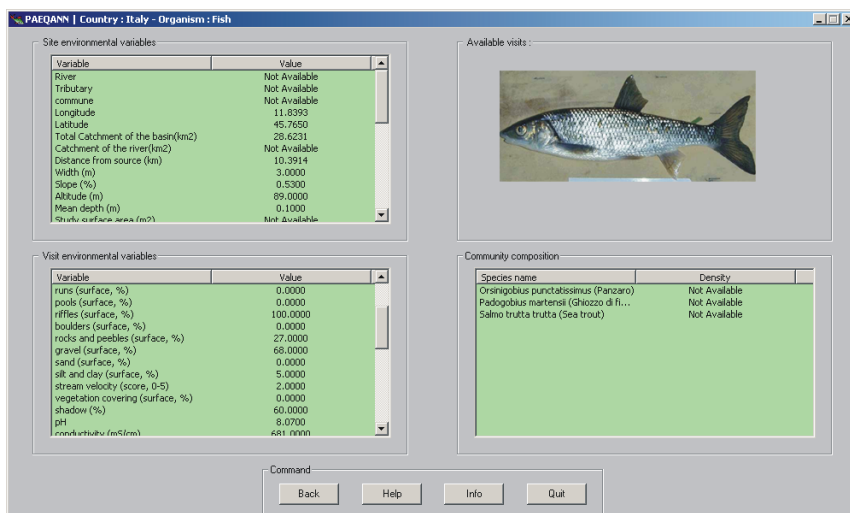
EVK1-1999-00125

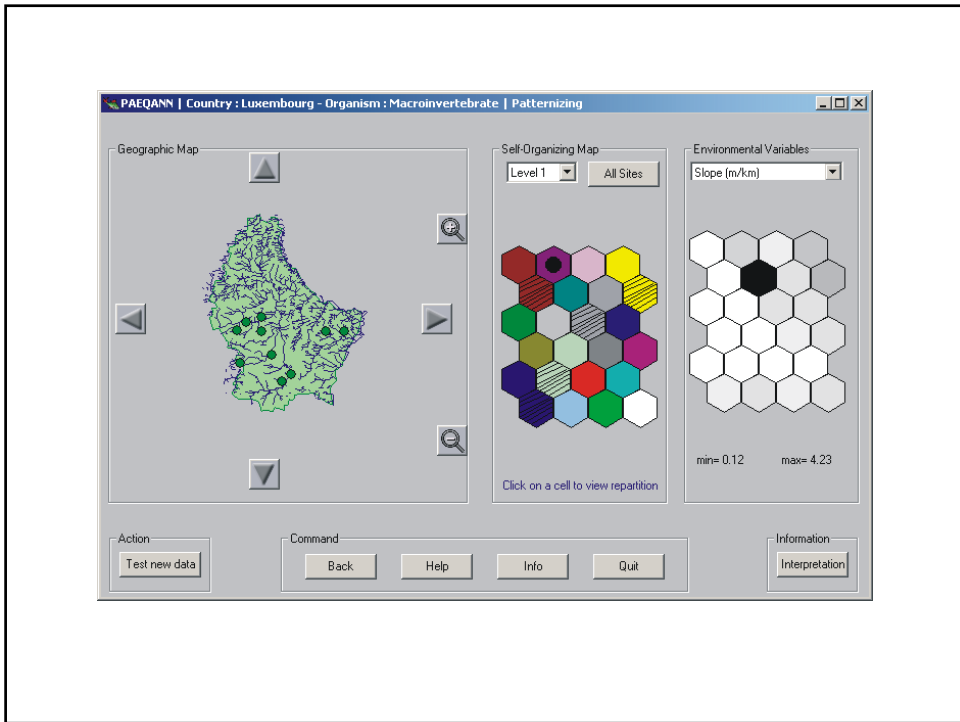
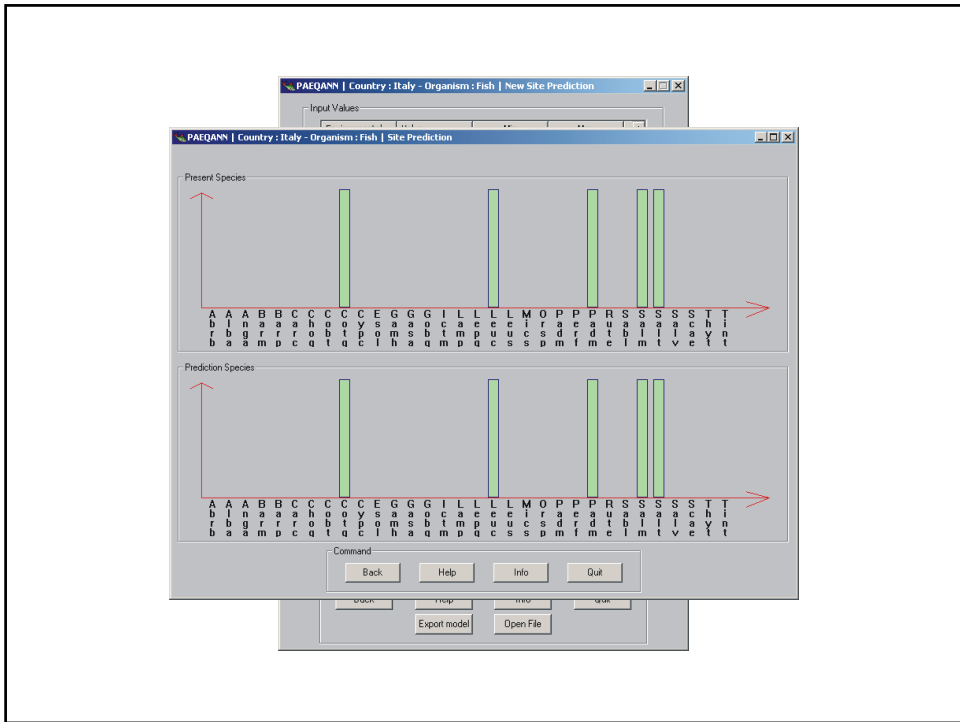
Predicting Aquatic Ecosystem Quality using Artificial Neural Networks: Impact of Environmental characteristics on the Structure of Aquatic Communities (Algae, Benthic and Fish Fauna)

- | | | |
|---|---|----------------------|
| 1. France, Univ. Paul Sabatier, CESAC lab | → | S. Lek (ccord.) |
| 2. Denmark, Univ. Copehargen, DFH, Env. Chemistry lab | → | S.E. Jorgensen |
| 3. Italy, Univ. Rome Tor Vergata, Dept. Biology | → | M. Scardi |
| 4. Belgium, Univ. Namur, Freshwater Ecology lab (LFE) FUNDP | → | J.P. Descy |
| 5. France, CEMAGREF, Bordeaux | → | F. Delmas / M. Coste |
| 6. Luxemburg, Cell. Recherche Environ. Biotechnol. | → | L. Ector |
| 7. Netherlands, ALTERRA Institute, Freshwater Ecol. team | → | P.F.M. Verdonschot |
| 8. Austria, Austrian Research Center, Seibersdorf | → | H.M. Knoflacher |



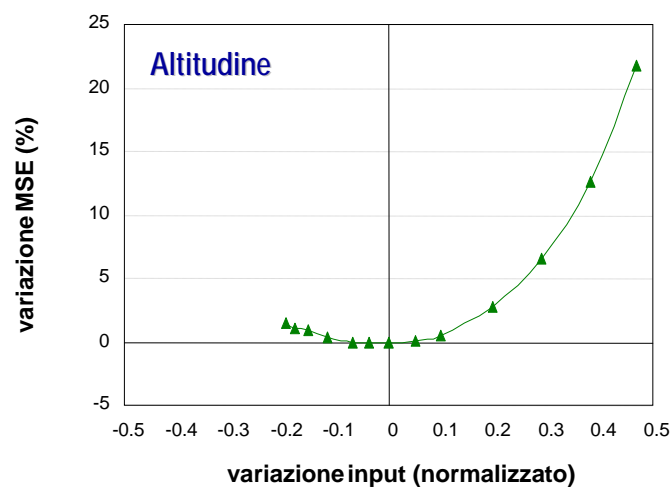
<http://aquaeco.ups-tlse.fr>

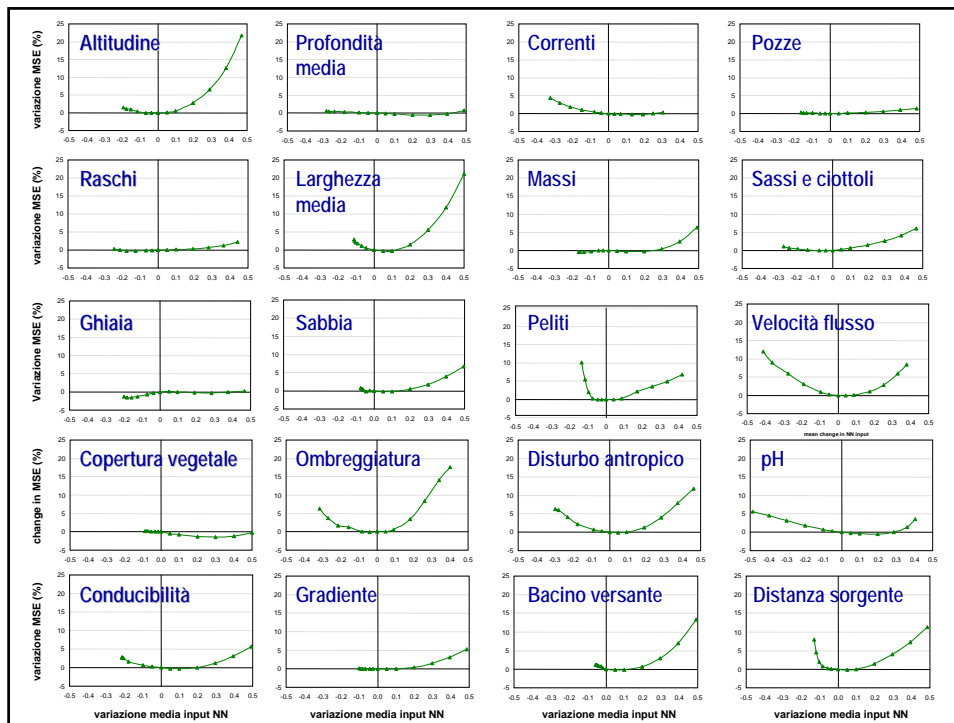




APRIAMO IL *BLACK BOX*: COME ESTRARRE INFORMAZIONE DAI NOSTRI MODELLI

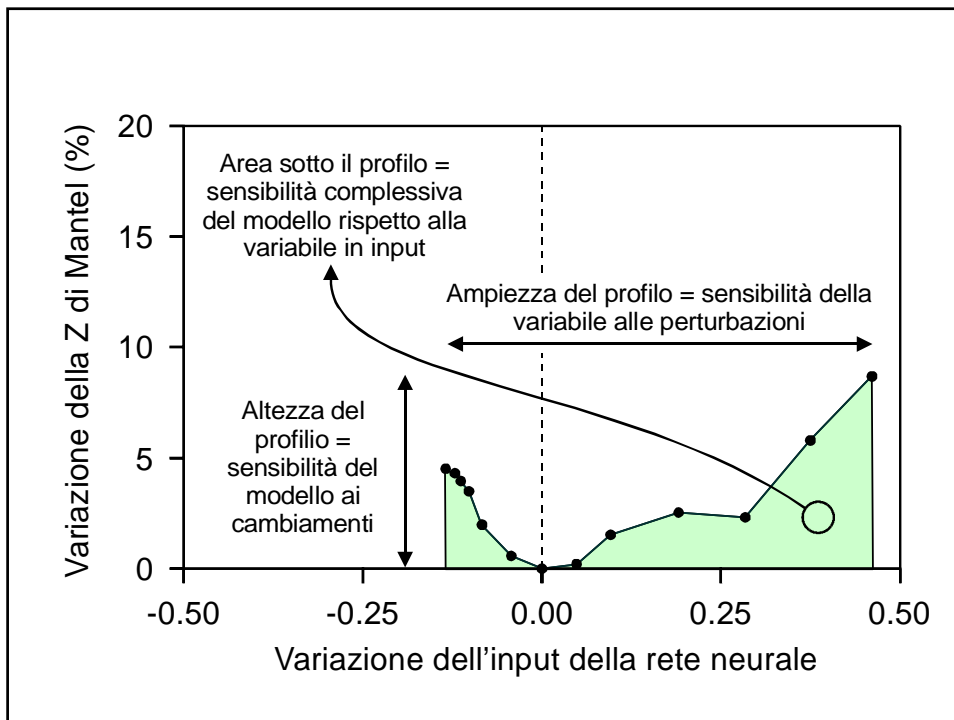
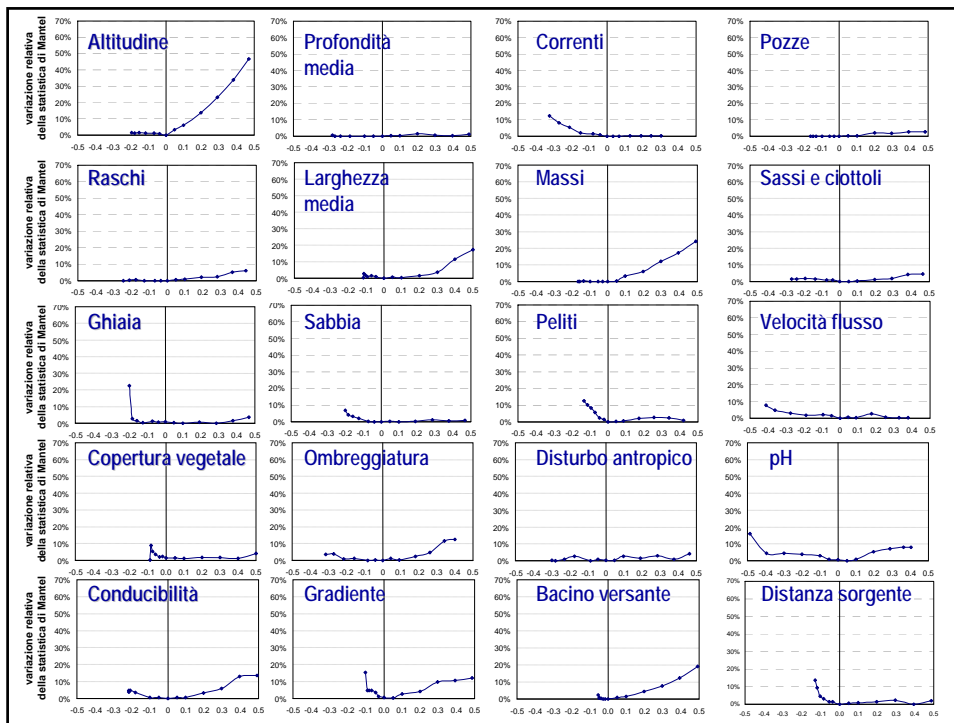
Analisi di sensibilità

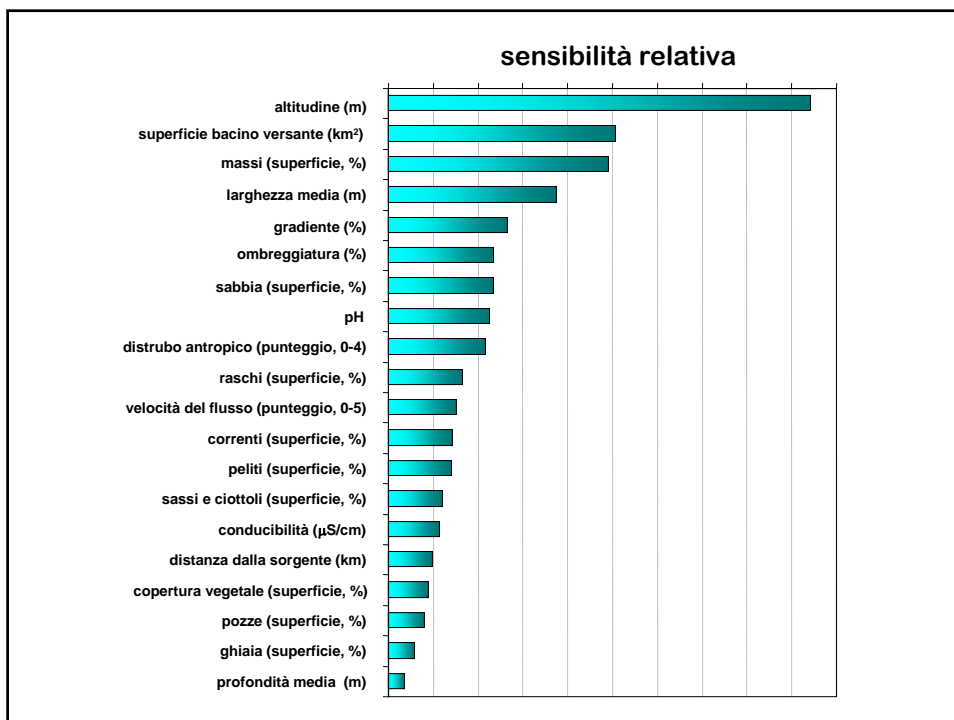
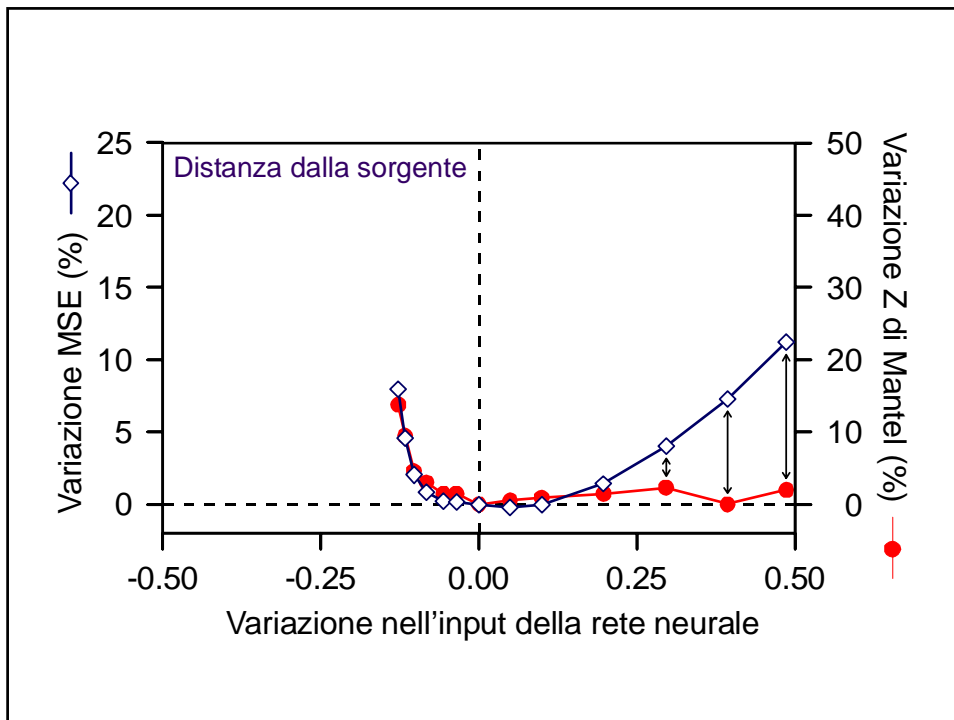




Questi risultati sono ecologicamente significativi?

- Una procedura di **analisi di sensibilità** di questo tipo è uno standard applicato comunemente ad un ampio spettro di problemi.
- Questa procedura, in particolare, implica che il dato di **presenza (1)** o di **assenza (0)** siano **numeri comparabili** e trattabili in maniera **quantitativa**.
- Tuttavia, il significato della **presenza** o dell'**assenza** di una specie varia in funzione del **contesto ecologico**.
- Quindi, è necessario utilizzare **altri criteri** per valutare modelli che prevedano la presenza di specie.





E LA DIRETTIVA SULLE ACQUE?

Problema

- Una volta definita la struttura della **comunità di riferimento**, date le condizioni dell'ambiente fisico al contorno, come si misura lo **scarto** da essa della **comunità osservata**?
- Ovvero, come si formula un **giudizio di qualità**?

Alcune misure di scarto dall'assetto atteso della comunità

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

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

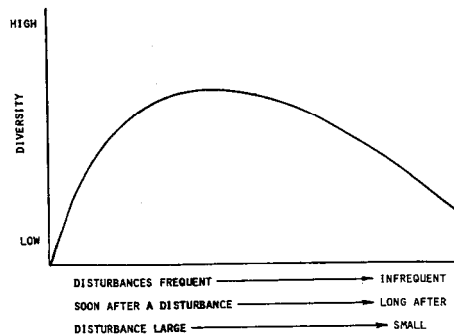
Rogers & Tanimoto (1960) $S_{jk} = \frac{a+d}{a+2b+2c+d}$

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

O/E {
 $< 1 \rightarrow$ comunità degradata
 $\approx 1 \rightarrow$ comunità integra
 $> 1 \rightarrow$ comunità arricchita

Teoria del disturbo intermedio

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



	osservato		atteso (rete neurale)		
	n	p	n	d	
Abramide	0	0.030	0	d	
Alborella	0	0.528	1	c	
Anguilla	0	0.751	1	c	
Barbo	1	0.853	1	a	
Carassio	0	0.074	0	d	
Carpa	0	0.033	0	d	
Cavedano	1	0.718	1	a	
Cobite	1	0.634	1	a	
Gambusia	0	0.040	0	d	
Ghiozzo di fiume	1	0.818	1	a	
Gobione	0	0.345	0	d	
Lampreda di ruscello	0	0.042	0	d	
Lasca	1	0.552	1	a	
Luccio	0	0.589	1	c	
Panzarolo	0	0.360	0	d	
Persico reale	0	0.053	0	d	
Persico sole	0	0.060	0	d	
Persico trota	0	0.028	0	d	
Pesce gatto	0	0.019	0	d	
Sanguinerola	0	0.391	0	d	
Scardola	1	0.521	1	a	
Scazzone	0	0.245	0	d	
Spinarello	0	0.026	0	d	
Temolo	0	0.069	0	d	
Tinca	0	0.364	0	d	
Trotto	0	0.646	1	c	
Trota fario	1	0.897	1	a	
Trota iridea	0	0.119	0	d	
Trota marmorata	0	0.202	0	d	
Vairone	1	0.090	0	b	

a = 7
b = 1
c = 4
d = 18

O = 8
E_p = 10.099
E_n = 11

O/E_p = 0.792
O/E_n = 0.727

Concordanza semplice = 0.833
Rogers & Tanimoto = 0.714
Jaccard = 0.583

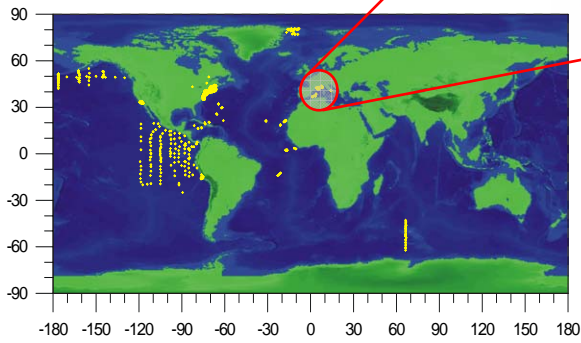
skip

AN EMPIRICAL MODEL CONSTRAINED BY A BIOLOGICAL RULE

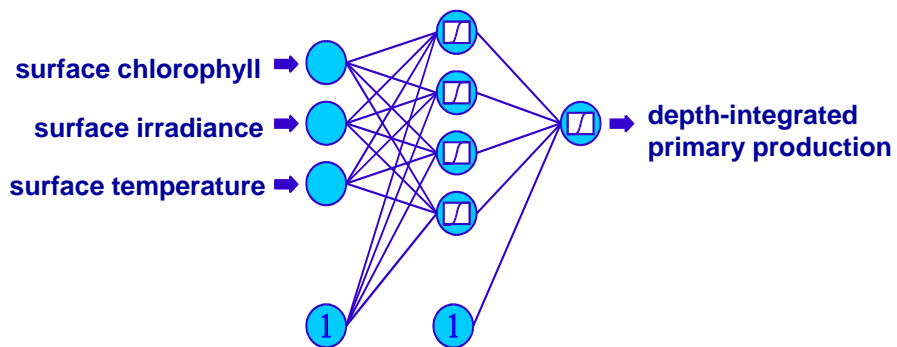
Michele Scardi

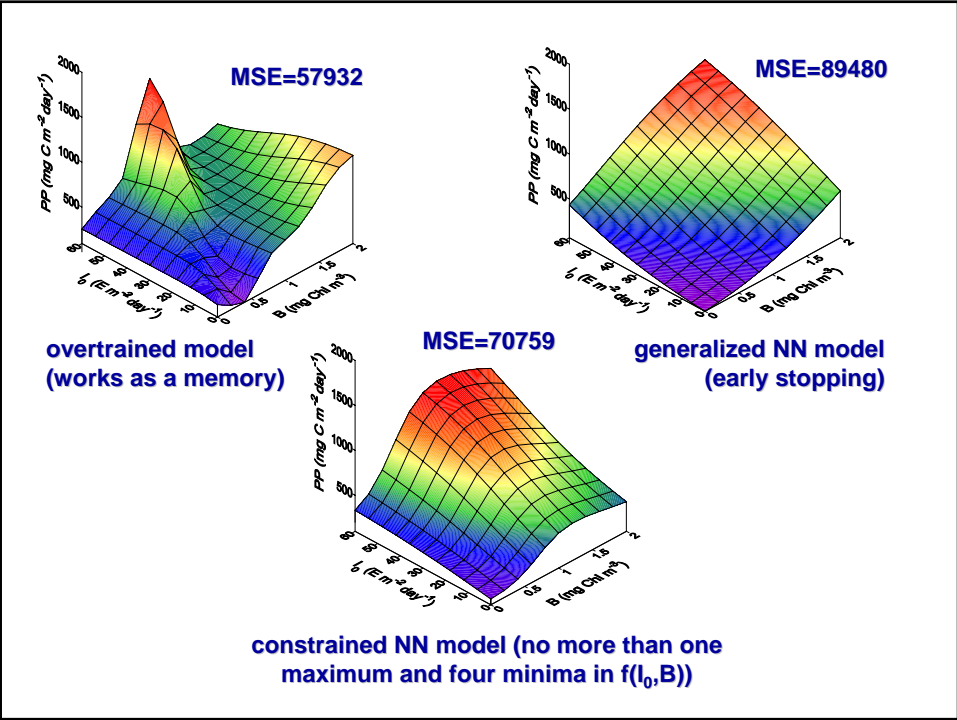
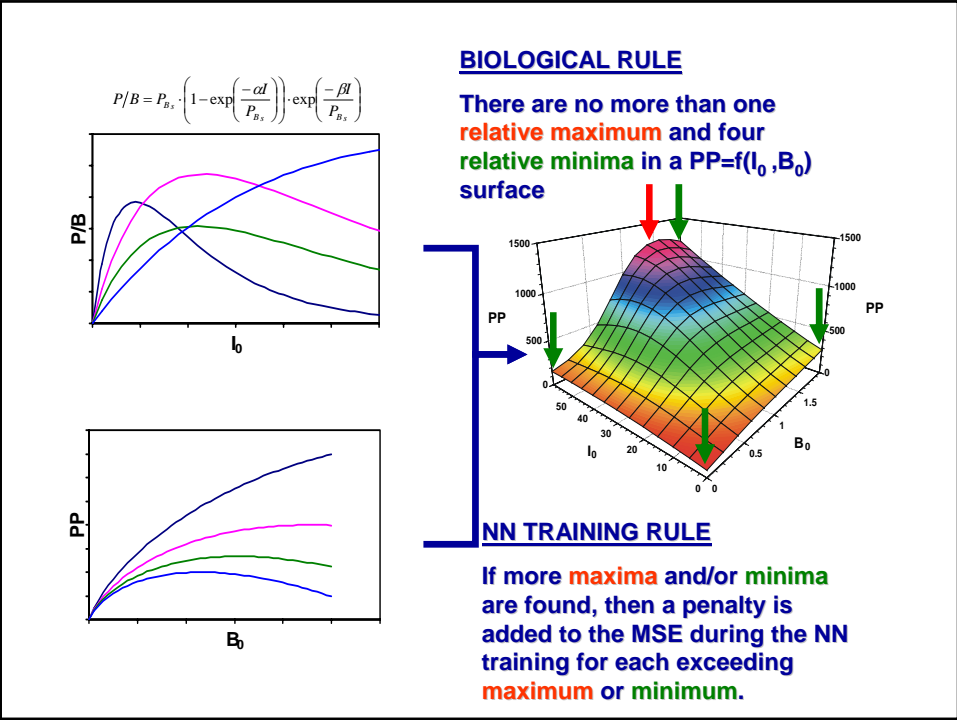
Dept. of Biology, Univ. of Roma "Tor Vergata", Roma, Italy

**Phytoplankton primary production
sampling sites in Western
Mediterranean Sea**



**A simple 3-4-1 neural network model
of phytoplankton primary production**





skip

MODELLING EEL CATCHES IN THE TEVERE RIVER

Eleonora Ciccotti, Stefano Cataudella & Michele Scardi

Dept. of Biology, Univ. of Roma "Tor Vergata", Roma, Italy

Eel catches in the Tevere river

Predictive variables

- julian day
- no fishing
- moon
- river flow (day₀)
- river flow (day₋₁)
- river flow (day₋₂)
- cloud coverage (day₀)
- cloud coverage (day₋₁)
- cloud coverage (day₋₂)
- rainfall (day₀)
- rainfall (day₋₁)
- rainfall (day₋₂)

NN structure

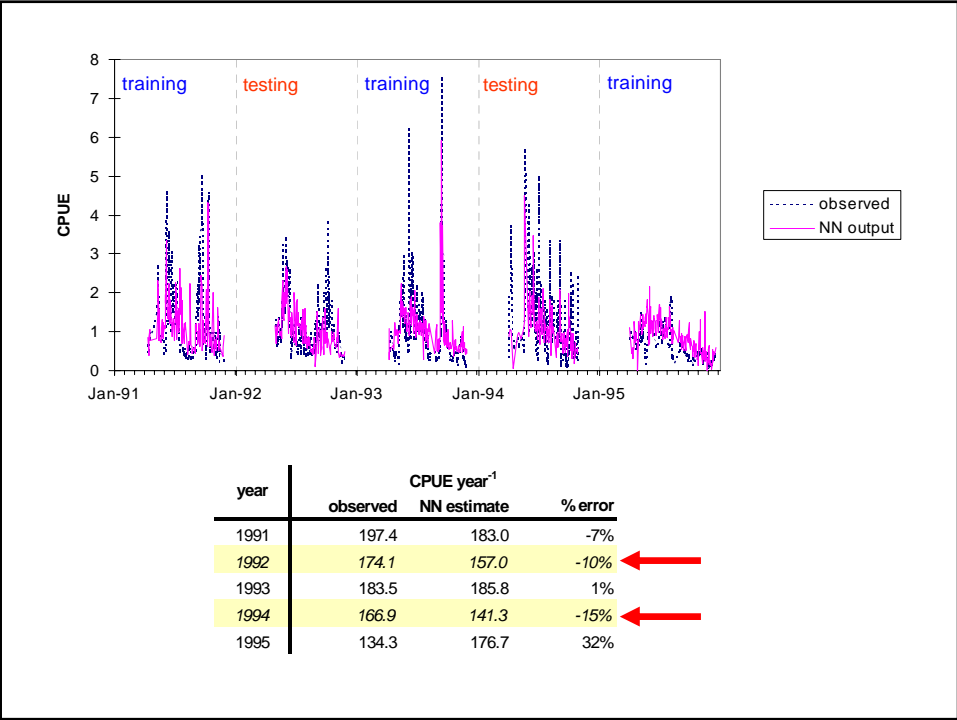
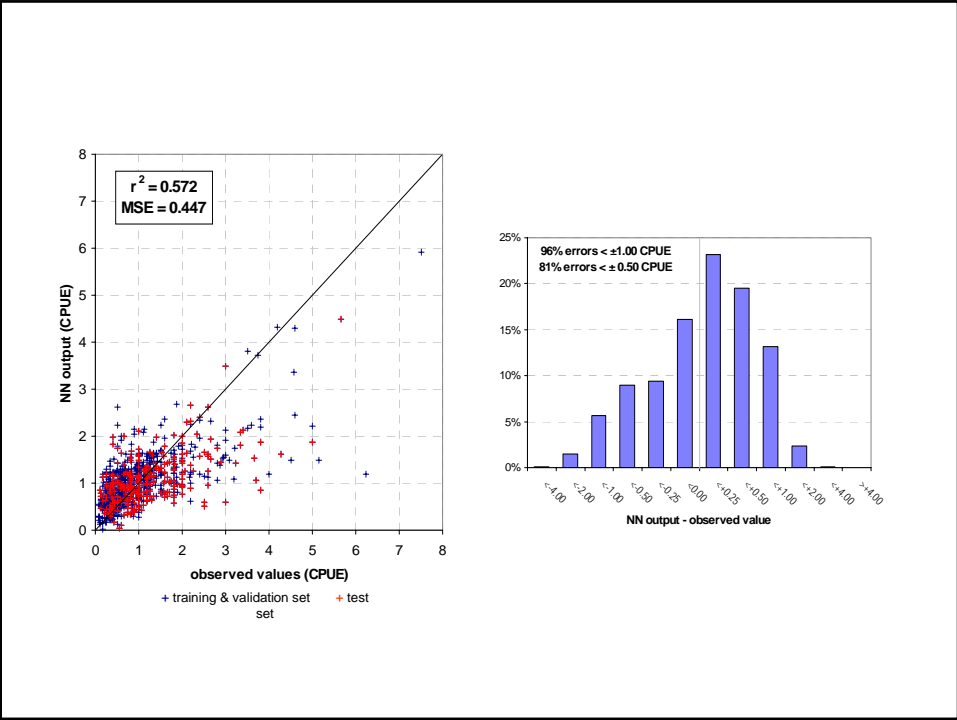
- 12 input nodes
- 8 hidden nodes
- 1 output node

Output variable

- eel catches (CPUE)

Data sets

- 1991, 1993 and 1995: training (n=532)
- 1992 and 1994: validation (n=283)



skip

MODELLING HEAVY METAL AND HC CONCENTRATIONS IN MARINE SEDIMENTS

Michele Scardi & Eugenio Fresi

Dept. of Biology, Univ. of Roma "Tor Vergata", Roma, Italy

Variabili predittive

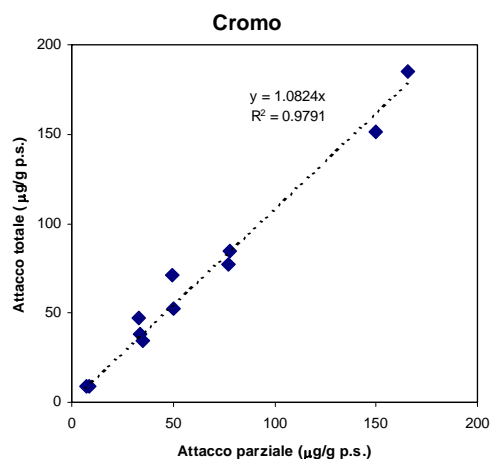
- Latitudine
- Longitudine
- Profondità (log)
- Sabbie (%)
- Distanza dalla costa
- Distanza della batimetrica dei 200 m dal punto più vicino della linea di costa
- Frequenza ponderata delle diverse tipologie del profilo di costa (3 variabili)
- Composizione delle argille (4 variabili)

skip

Correzione dei dati di concentrazione

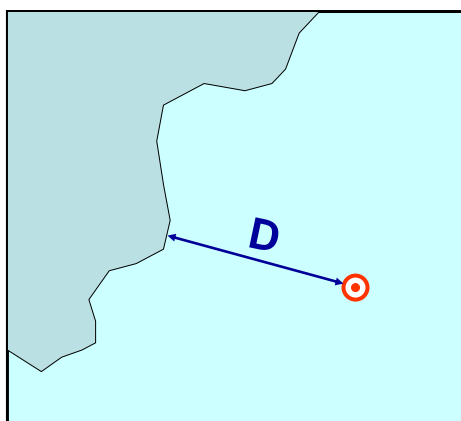
I dati per i quali il tipo di attacco era riportato in maniera esplicita sono stati corretti secondo le indicazioni prodotte dalla campagna sperimentale.

Ad esempio 



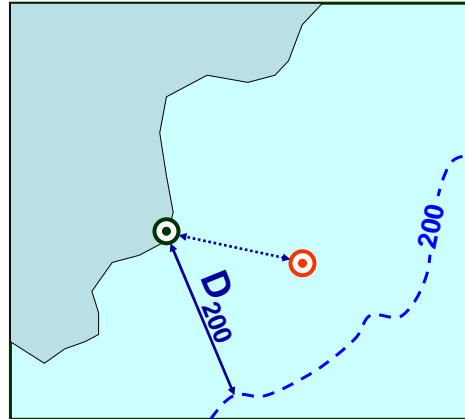
Distanza minima del punto dalla costa

E' una misura indiretta (se rapportata alla profondità del punto, dunque) dell'andamento batimetrico medio.



Distanza del punto più prossimo della costa dalla batimetrica dei 200 m

Esprime l'andamento batimetrico medio dell'area indipendentemente dalla profondità del punto considerato.

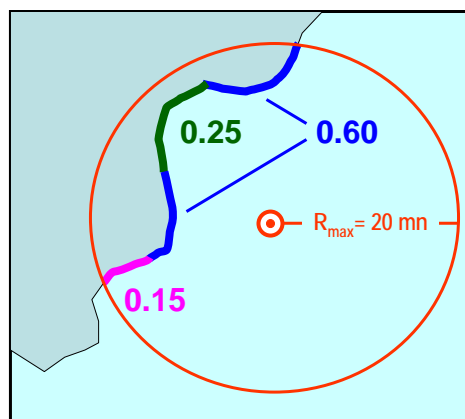


Frequenza delle tipologie del profilo di costa nell'intorno del punto

Le tre tipologie riconosciute sono funzione delle differenze fra profilo emerso e sommerso:

- **concordante**
- **poco discordante**
- **molto discordante**

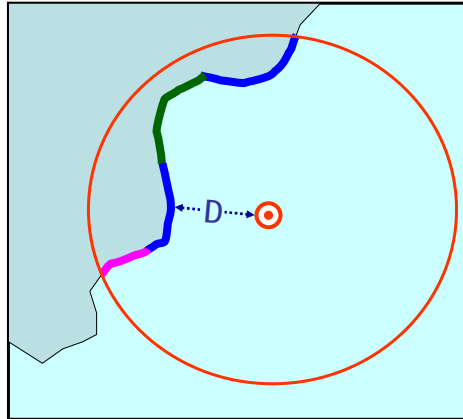
$$R(D) = D \frac{20 + D}{20} \text{ mn}$$



Influenza della tipologia del profilo di costa

E' stata definita una semplice relazione empirica che esprime l'influenza della tipologia del profilo di costa come funzione della distanza dalla costa.

$$W(D) = \frac{20 - D}{20} mn$$



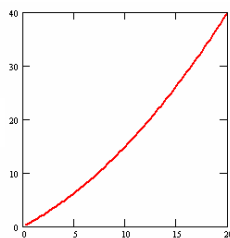
Influenza della tipologia del profilo di costa

Quanto più è ampio il raggio di ricerca...

...tanto più è bassa l'influenza della tipologia del profilo di costa.

Quindi la massima influenza si ha per distanze intermedie.

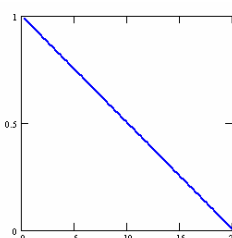
$$R(D) = D \frac{20 + D}{20} mn$$



D

X

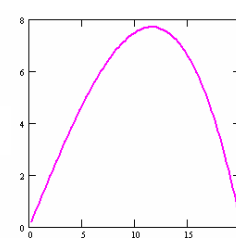
$$W(D) = \frac{20 - D}{20} mn$$



D

=

$$R(D) \cdot W(D)$$

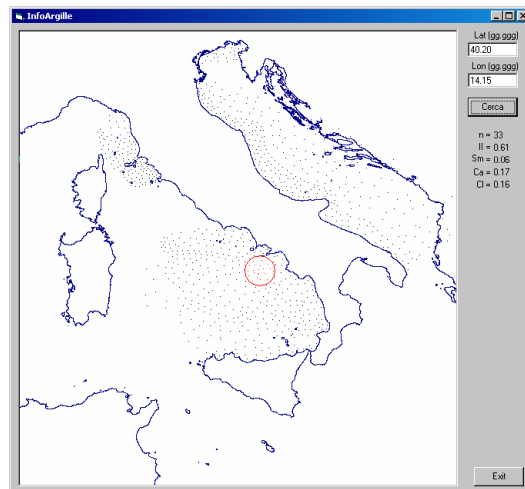


D

Origine del sedimento superficiale

Sono state utilizzate come descrittrici dell'origine dei sedimenti le proporzioni medie in un raggio di 50 km di **illite, smectite, caolinite e clorite** nelle argille (dati derivati dalla letteratura).

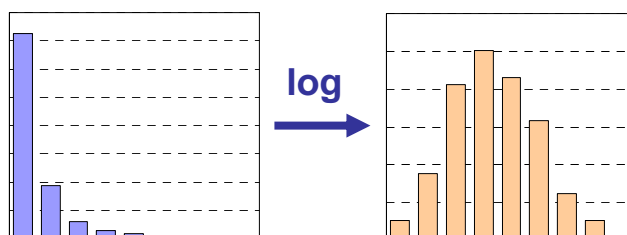
Laddove non sono disponibili dati vengono utilizzate le medie generali (0.56, 0.16, 0.15, 0.13 nell'ordine).



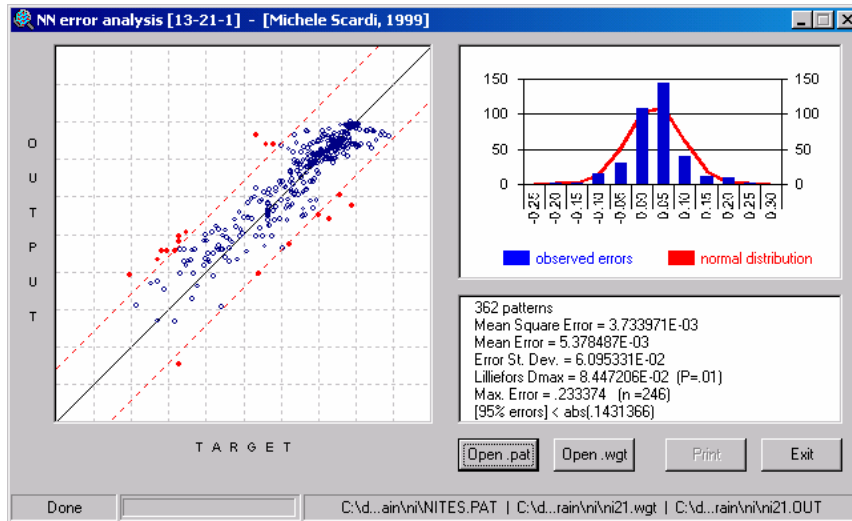
Trasformazione dei dati

Per:

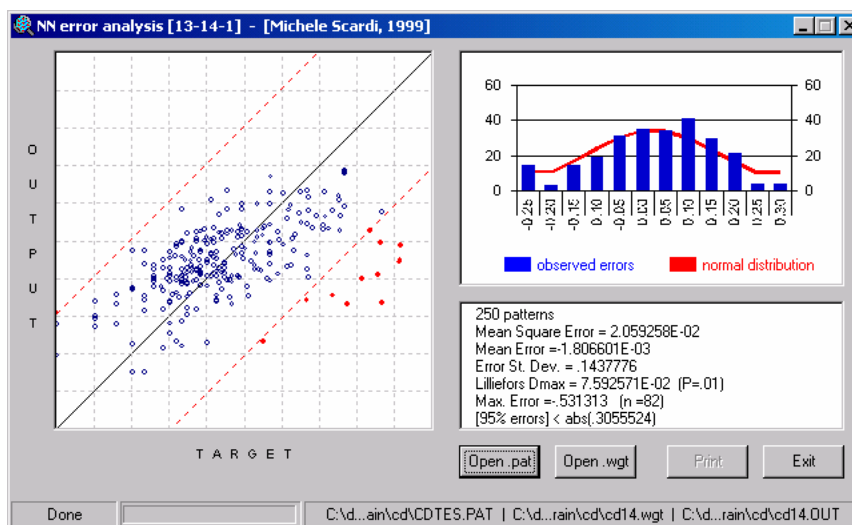
- profondità
- concentrazioni degli inquinanti



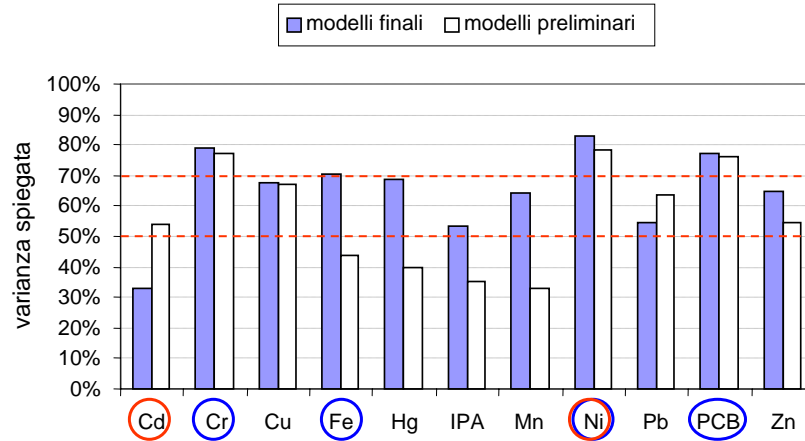
Ni ($r^2=0.82$)



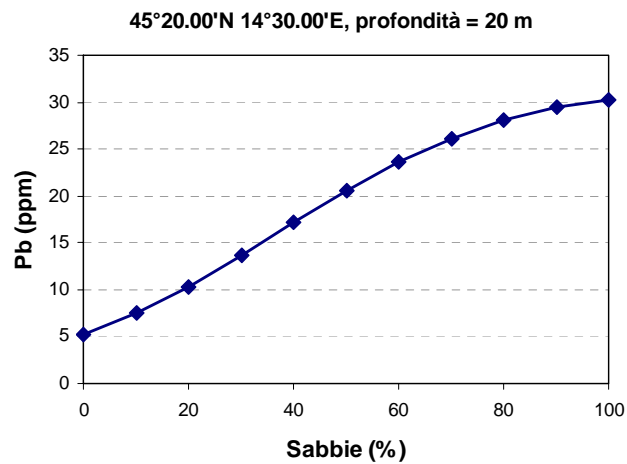
Cd ($r^2=0.32$)



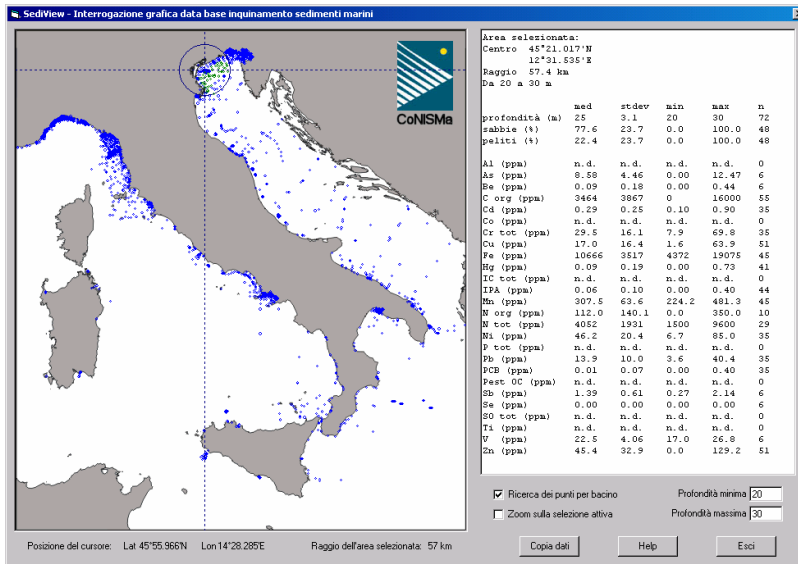
Dai modelli preliminari a quelli finali



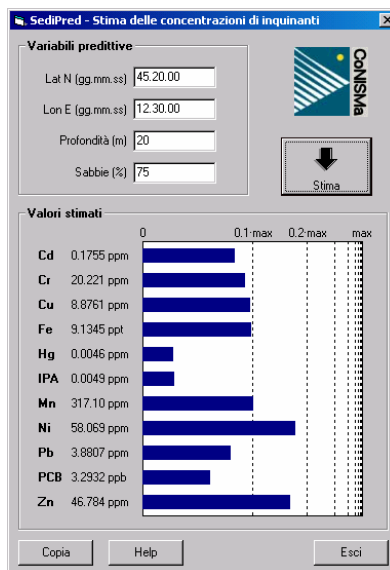
Variazione di un profilo tipo in funzione della frazione di sabbie



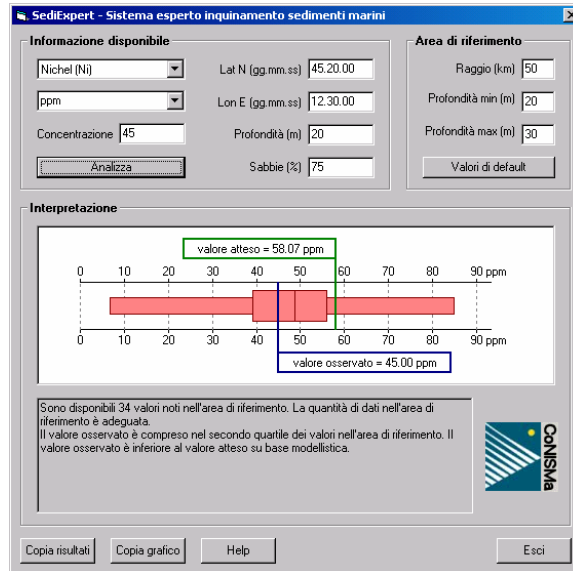
Data base SediView



Modelli SediPred



Data base + modelli + regole = SediExpert



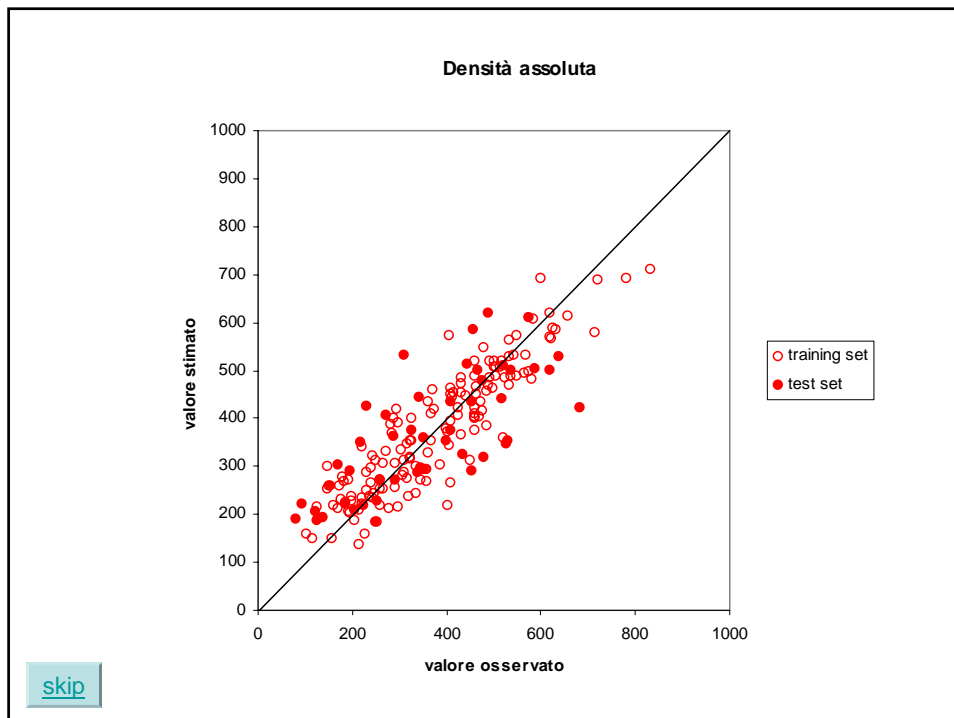
skip

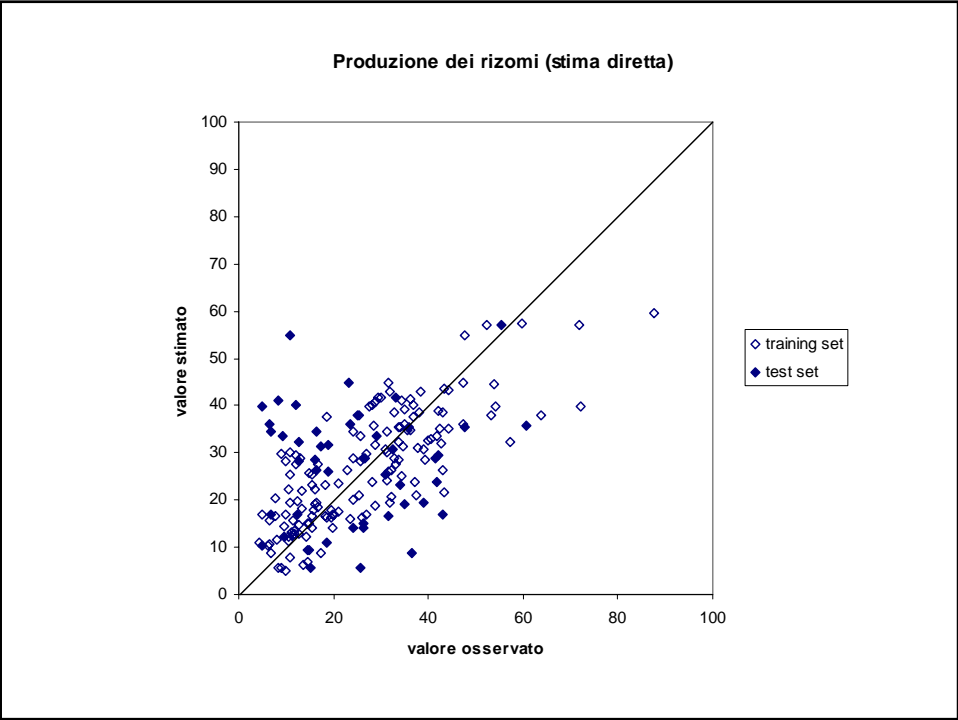
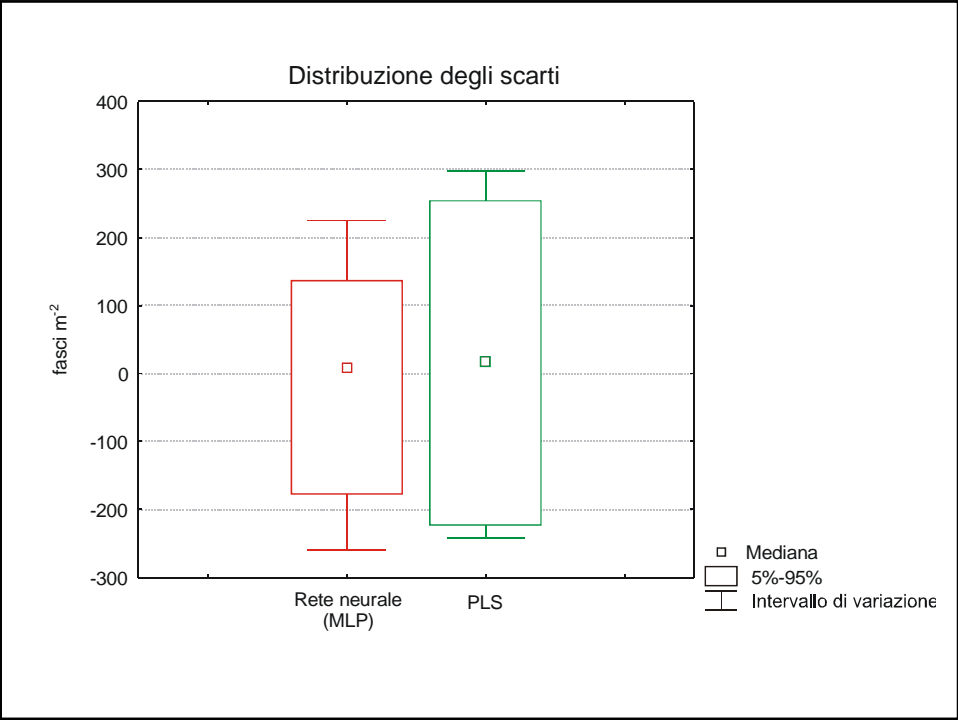
Stima della densità e della produzione attese in praterie di *Posidonia oceanica* (L.) Delile

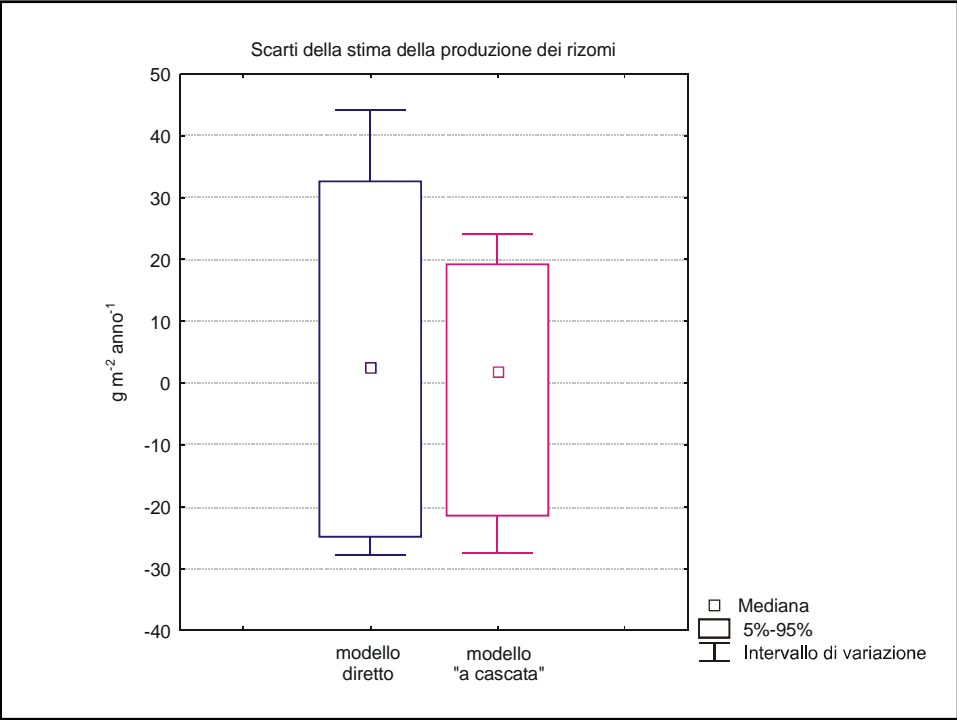
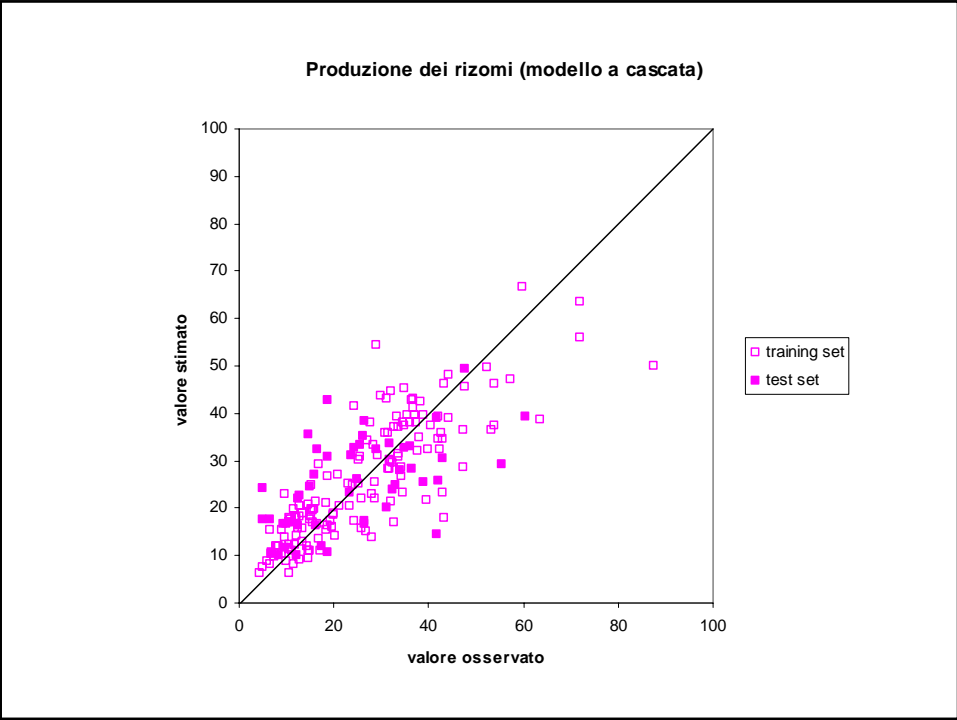
Michele Scardi & Eugenio Fresi

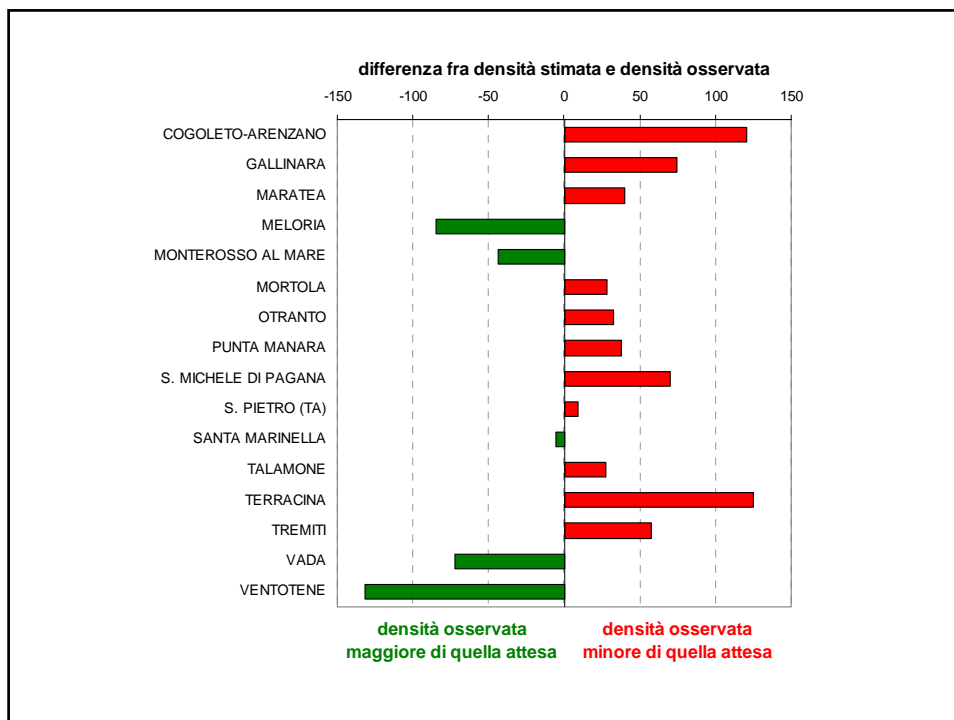
Dept. of Biology, Univ. of Roma "Tor Vergata", Roma, Italy

variabile	ID	minimo	massimo	
lat. (gradi decimali)	LAT	39.9	44.4	
long. (gradi decimali)	LON	7.5	18.6	
profondità (m)	PROF	0	35	
gradiente	GRAD	0	0.3	
concordanza esposizione	CONCESP	0	1	
concordanza venti dominanti	CONCVENT	0	1	
conformazione profilo batimetrico	concavo	CONC	0	1
	piano	PIANO	0	1
	convesso	CONV	0	1
runs variazione profilo	RUNS	0	4	
"apertura" litorale	APERTLIT	1	5	
esposizione ai venti dominanti	ESPVENTI	0	1	
tipologia fondale	sabbia	SABBIA	0	1
	roccia	ROCCIA	0	1
	matte	MATTE	0	1
sorgenti di disturbo	ancoraggio	ANCOR	0	1
	inquinamento	INQUIN	0	1
	immissioni	IMMISS	0	1
variabili da predire	dens ass media	DENSASS	0	900
	prod rizoma	PRODRIZ	0	100

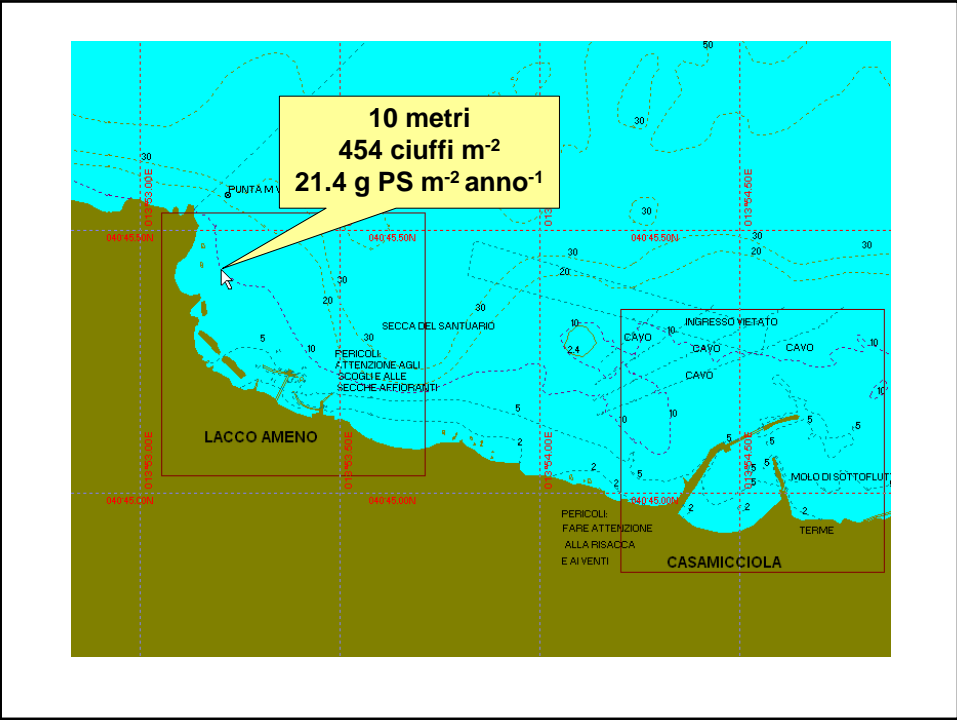








prateria	<10 m	10-15 m	15-20 m	>20 m	Totale
COGOLETO-ARENZANO		104.6	136.9		120.7
GALLINARA			74.2		74.2
MARATEA	13.1			66.3	39.7
MELORIA		-131.2		53.9	-84.9
MONTEROSSO AL MARE		-32.3		-54.5	-43.4
MORTOLA		-51.6		107.6	28.0
OTRANTO	-3.5	133.4	18.9		32.9
PUNTA MANARA	37.5				37.5
S. MICHELE DI PAGANA	70.2				70.2
S. PIETRO (TA)	-44.7	36.4			9.3
SANTA MARINELLA	-106.1	135.8	-45.8		-5.4
TALAMONE		27.5			27.5
TERRACINA	198.0			88.3	124.9
TREMITI	224.6		-25.5		57.9
VADA	-74.1		-70.4		-72.3
VENTOTENE	-116.2	-176.9			-131.4
<i>Totale</i>	-1.8	-9.0	14.5	59.5	8.8



PosiPred - Stime di densità e produzione in praterie di Posidonia oceanica

Variabili predittive

latitudine in gradi e decimali (39.9-44.4)	40.7572
longitudine in gradi e decimali (7.5-18.6)	13.8865
profondità in metri (0.0-35.0)	10
gradiente (0.0-0.3)	0.15
concordanza gradiente-esposizione (0/1)	0.5
concordanza gradiente-venti dominanti (0.0-1.0)	0.5
profilo batimetrico concavo (0/1)	0.5
profilo batimetrico piano (0/1)	0.5
profilo batimetrico convesso (0/1)	0.5
runs variazione profilo (0-4)	2
apertura del litorale (1-5)	3
esposizione ai venti dominanti (0.0-1.0)	0.5
substrato con sabbia (0/1)	0.5
substrato con roccia (0/1)	0.5
substrato con matite (0/1)	0.5
danni da ancoraggio (0/1)	0.5
sorgenti di inquinamento (0/1)	0.5
immissioni al contorno (0/1)	0.5
densità assoluta misurata (opzionale, 0.0-900.0)	

Cancella tutti i campi

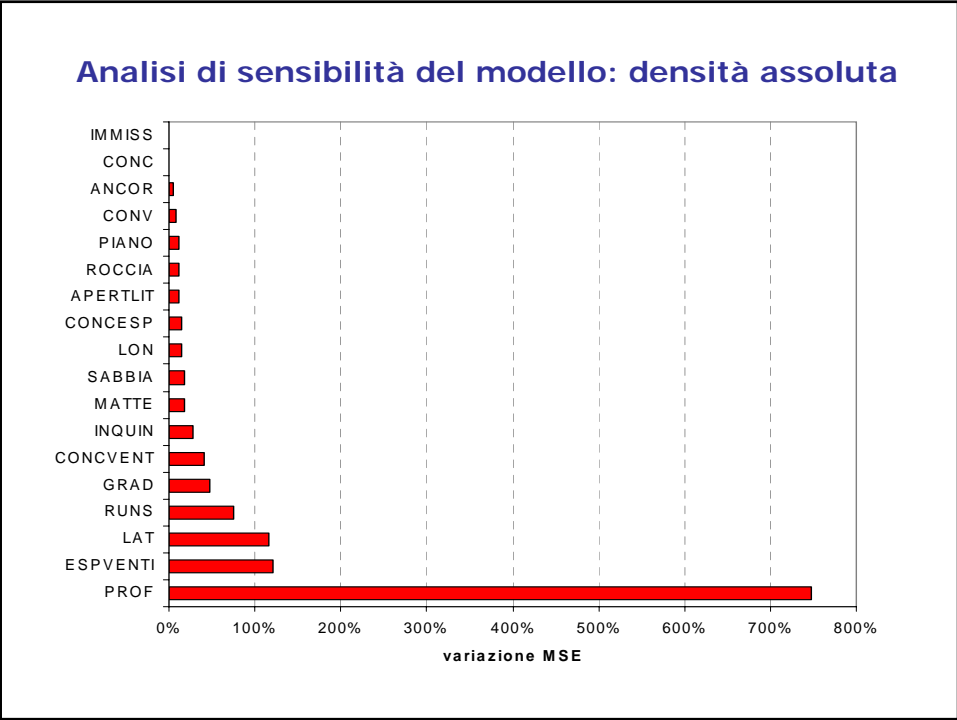
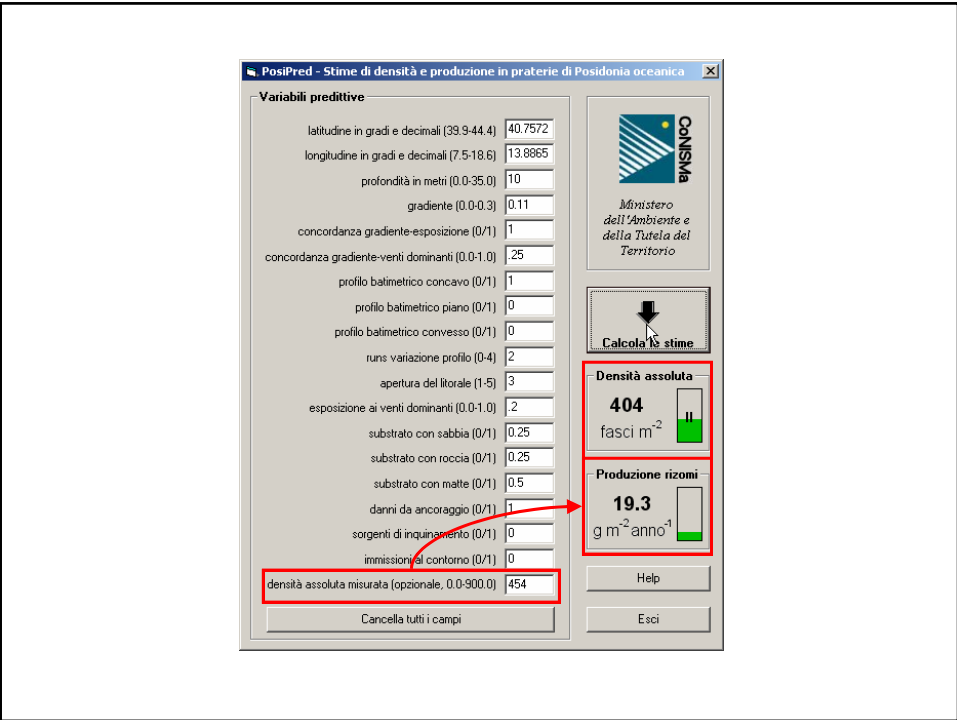
CAISM
Ministero dell'Ambiente e della Tutela del Territorio

Calcola le stime

Densità assoluta
286 fasci m⁻² IV

Produzione rizomi
12.7 g m⁻² anno⁻¹

Help
Esci



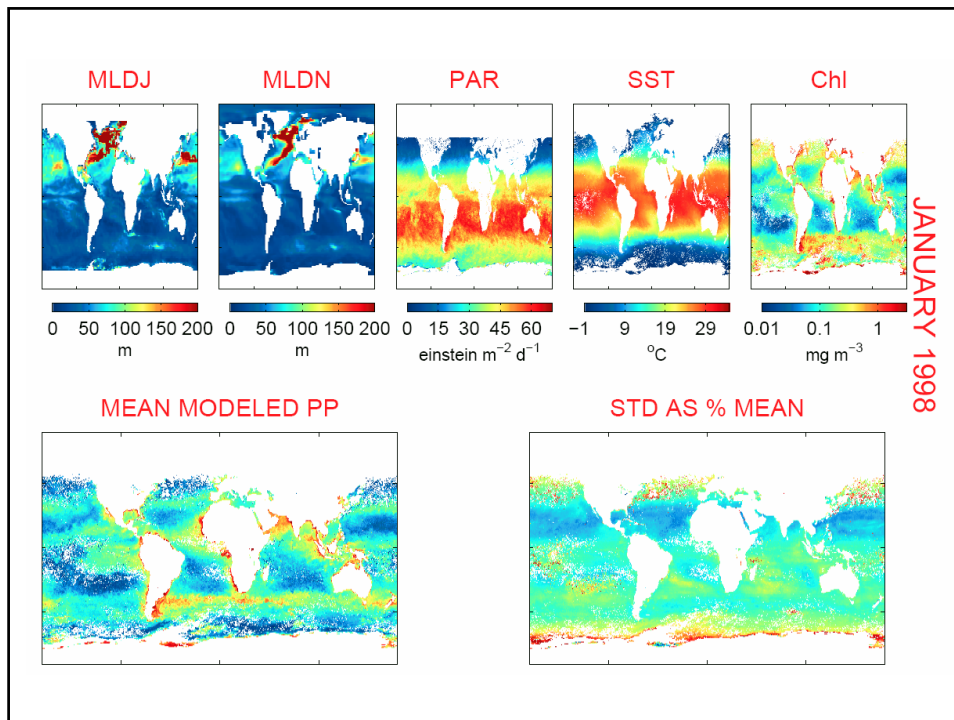
The Third Primary Production Algorithm Round-Robin (PPARR3)

Un'esercizio di comparazione dei modelli globali di produzione primaria fitoplanctonica

<http://oceans-www.jpl.nasa.gov/bio/PrimaryProd/index.html>

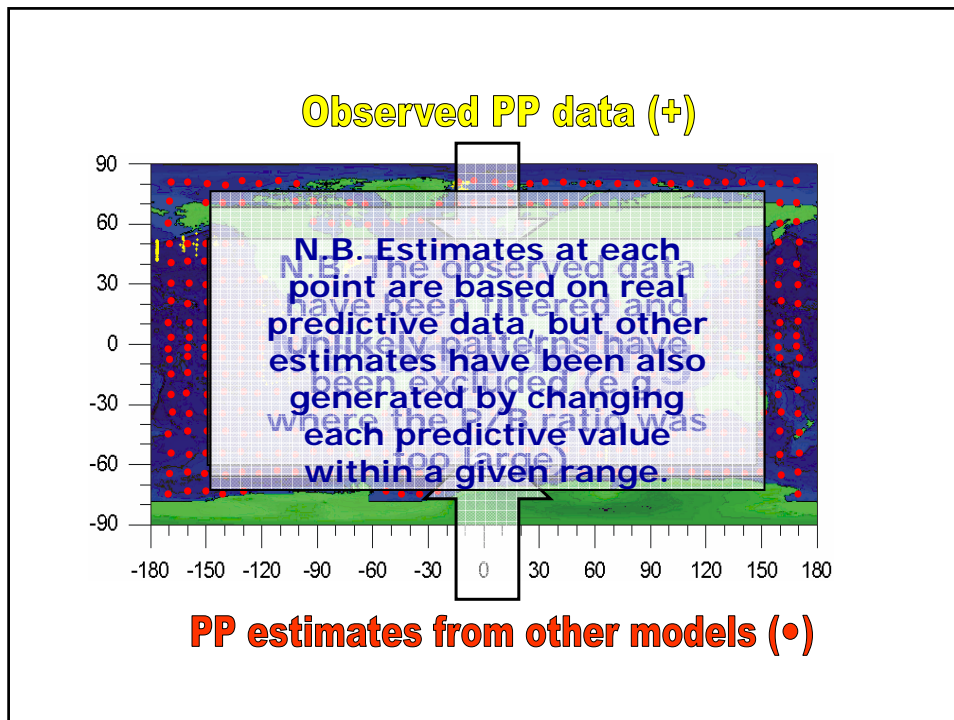
Partecipanti:

- 1'. David. IPCF. D. Antoine, B. Gentili and A. Morel.
2. Nick. BIO variant. N. Hoepffner and F. Melin.
3. Kirk. K. Waters and B. Bidigare.
4. Tim. Modified IPCF. T. Smyth and S. Groom.
5. Mike. VGPM. M. Behrenfeld.
6. Mike2. VGPM (Eppley P_{Bopt}). M. Behrenfeld.
7. ModisBF. VGPM. K. Turpie and W. Esaias.
8. Aurea. VGPM. A. Ciotti.
9. Joji. J. Ishizaka and Mr. Kameda.
10. Keith. K. Moore.
11. Heidi. Southern Ocean. H. Dierssen.
12. Heidi2. Southern Ocean chlorophyll. H. Dierssen.
13. Ichio. I. Asanuma.
14. Mark: Province-based. M. Dowell.
- 15'. ModisHYR. HoYoRy. K. Turpie and W. Esaias.
- 16'. RyYo. HoYoRy variant. J. Ryan.
17. HYRZe. HoYoRy variant (Z_e). M-E Carr.
18. Michele. Neural network. M. Scardi.
19. John. J. Marra.
20. SteveB. Hybrid WIM. VGPM P_{Bopt} . S. Lohrenz.
21. SteveB2. Hybrid WIM, VGPM P_{Bopt} . S. Lohrenz.
22. SteveA. Hybrid WIM, IPCF P_{Bmax} . S. Lohrenz.
23. OliCor. Ecosystem model. O. Aumont.
- 24'. KeithE. Ecosystem model. K. Moore.



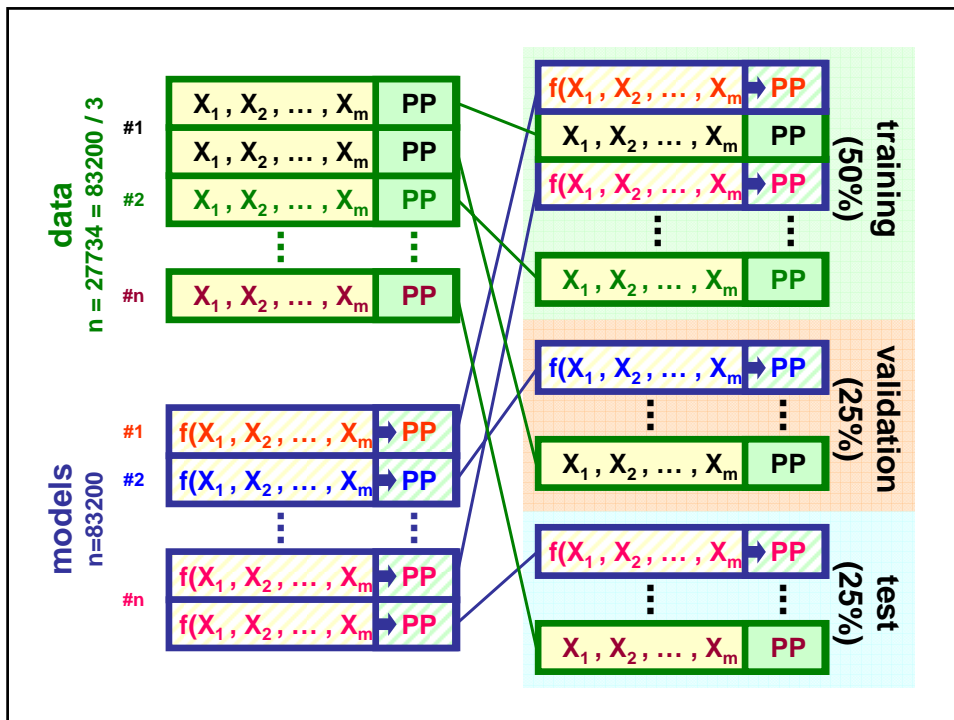
The NN PP model is based on:

- **The OPPWG global data set from OPPWG**
(Institute of Marine and Coastal Sciences at Rutgers University, New Brunswick, NJ, USA)
- **The F029 and F049 data sets**
(N.O.A.A. National Oceanographic Data Center, Silver Spring, MD, USA)
- **A time series from the Gulf of Naples, Italy**
(Stazione Zoologica 'A. Dohrn' of Naples)
- **Other models**
(VGPM by B&F, Howard-Yoder, MOD-27 by Esaias)

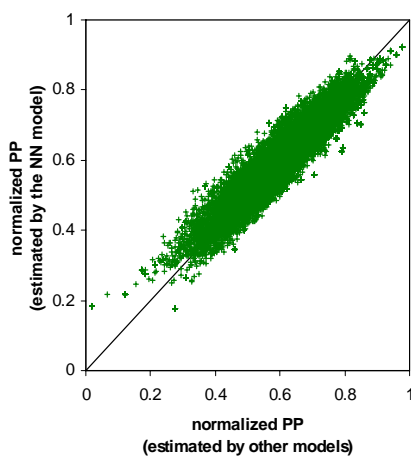


Predictive (input) variables:

- Lat latitude
- sinLon $1/2[\sin(2\pi \cdot (\text{Longitude}+180)/360)+1]$
- cosLon $1/2[\cos(2\pi \cdot (\text{Longitude}+180) /360)+1]$
- sinDay $1/2[\sin(2\pi \cdot \text{JulianDate}/365)+1]$
- cosDay $1/2[\cos(2\pi \cdot \text{JulianDate}/365)+1]$
- DL day length
- I_0 surface downwelling PAR
- log(MLD) log-transformed mixed layer depth
- SST sea surface temperature
- P_{opt}^b measured or estimated after B&F
- log(Chl) log-transformed surface chlorophyll

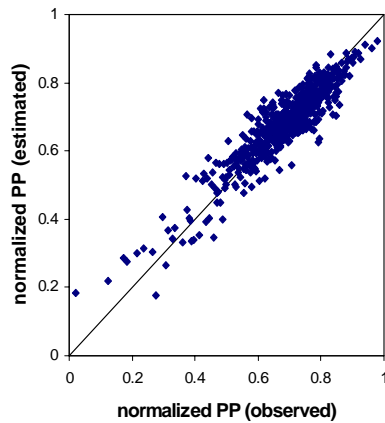


Metamodeling, i.e. modeling other models

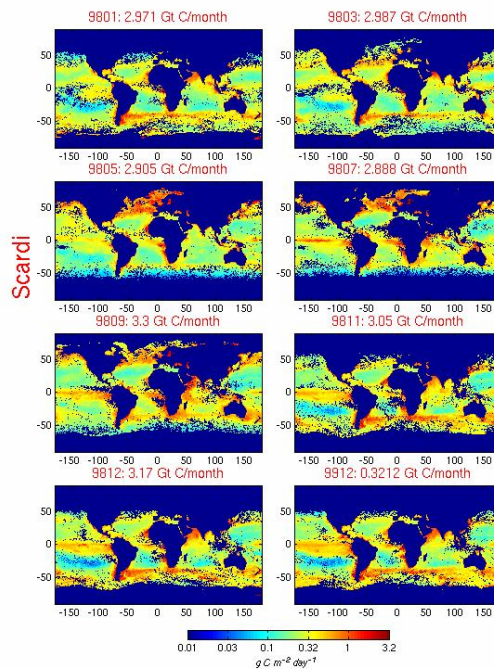


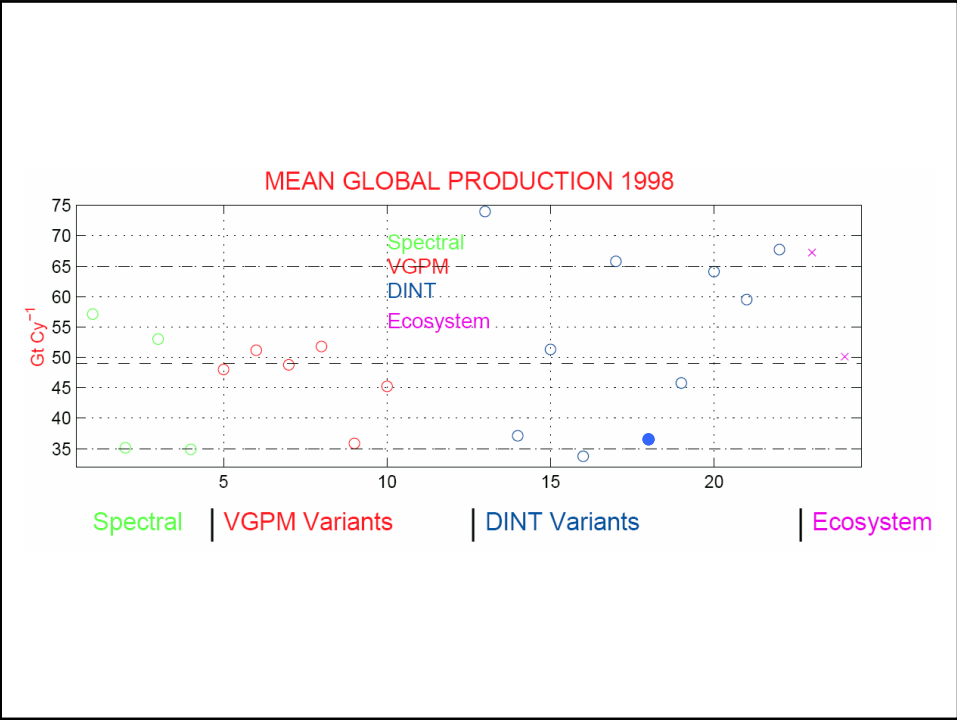
$n=227732$
 $MSE=0.00156$
 $R^2=0.88$
 mean error=-0.0003
 median error=-0.0001
 $-0.025 < 50\%$ of errors < 0.023
 $-0.063 < 90\%$ of errors < 0.064

Checking the NN metamodel using an independent test data set



n=552
MSE=0.00248
R²=0.85
mean error=-0.004
median error=-0.005
-0.034 < 50% of errors < 0.026
-0.082 < 90% of errors < 0.084





Sensitivity analysis

+/- change in input	Lat	sinLon	cosLon	sinDay	cosDay	DL	I0	log(MLD)	SST	PBopt'	log(Chl)
1.0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	7%	4%
2.0%	0%	0%	0%	0%	0%	0%	1%	0%	0%	29%	17%
5.0%	1%	0%	0%	0%	0%	2%	5%	3%	1%	184%	105%
10.0%	4%	1%	0%	1%	0%	9%	18%	12%	4%	493%	417%
15.0%	8%	2%	1%	2%	1%	21%	36%	26%	8%	581%	921%
20.0%	13%	3%	1%	4%	1%	40%	59%	47%	13%	639%	1598%
25.0%	18%	5%	1%	5%	1%	68%	89%	75%	19%	686%	2415%
30.0%	25%	7%	2%	7%	2%	108%	121%	107%	26%	722%	3334%
40.0%	38%	10%	2%	10%	3%	221%	201%	172%	41%	763%	5246%
50.0%	52%	12%	3%	14%	4%	323%	302%	216%	56%	780%	6771%

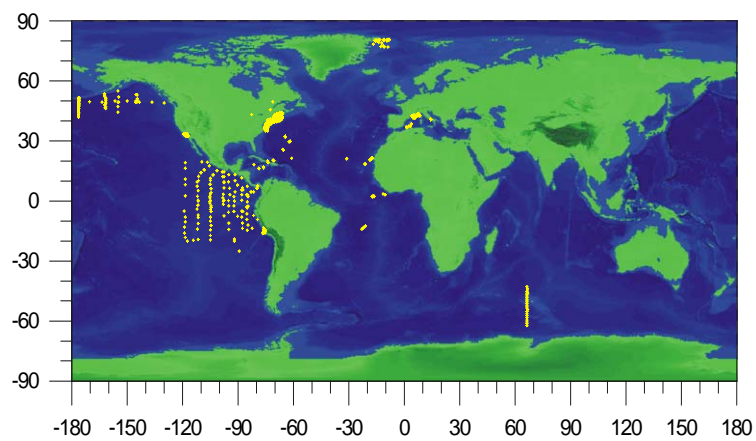
skip

IMPROVING A GLOBAL MODEL OF PHYTOPLANKTON PRIMARY PRODUCTION

Michele Scardi

Dept. of Biology, Univ. of Roma "Tor Vergata", Roma, Italy

Phytoplankton primary production sampling sites



A global model of phytoplankton primary production (Scardi, 2000)

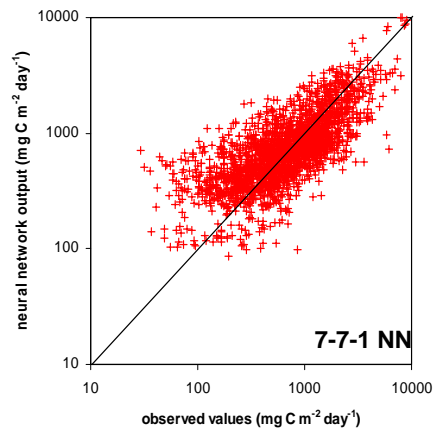
Predictive variables:

- surface biomass
- surface irradiance
- surface temperature
- date*
- longitude
- latitude

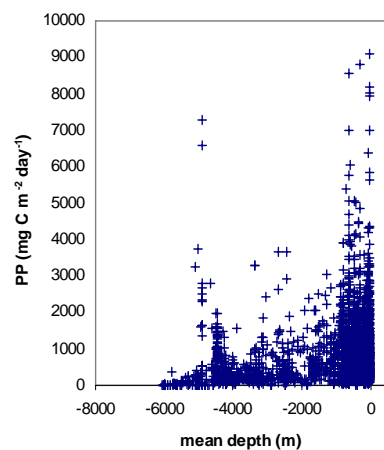
* 2 variables, i.e.

$$date_1 = \frac{1}{2} \left[\cos\left(\frac{2\pi \cdot day}{365}\right) + 1 \right]$$

$$date_2 = \frac{1}{2} \left[\sin\left(\frac{2\pi \cdot day}{365}\right) + 1 \right]$$

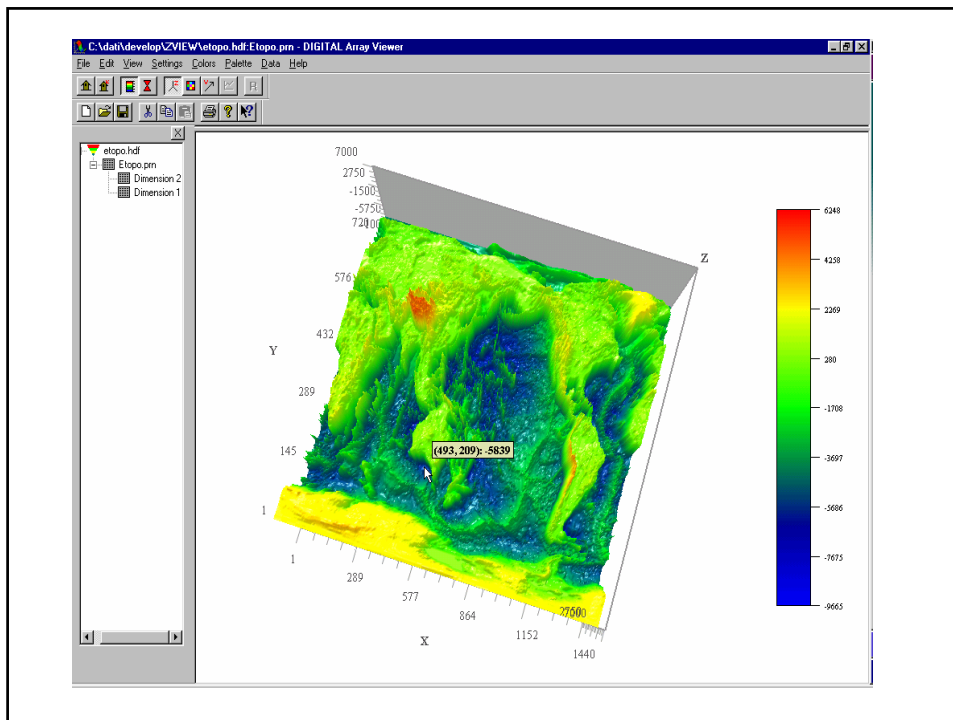


Water column depth as a co-predictor for primary production



Depth affects PP via:

- water column dynamics
- upwelling regions
- coastal fronts
- nutrient dynamics
- freshwater run-off
- etc.



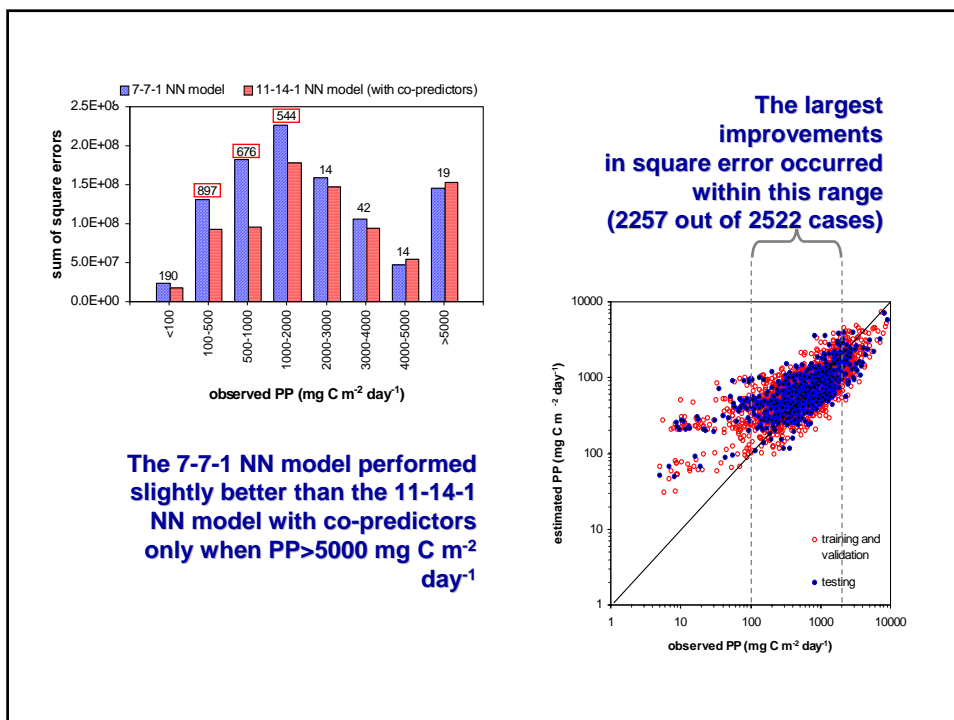
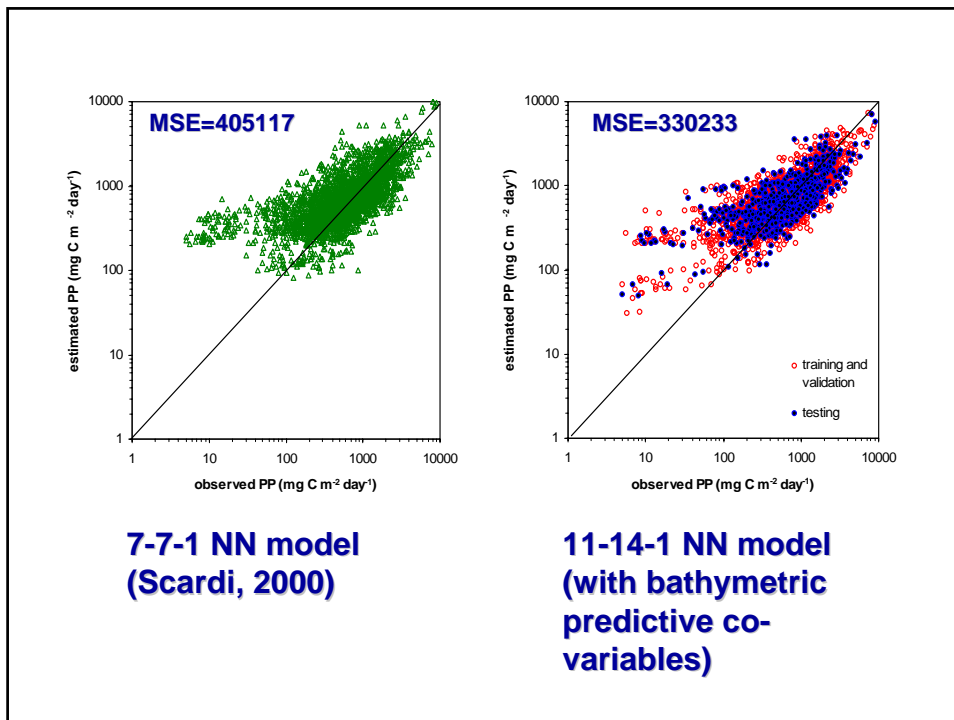
Primary production predictors and co-predictors:

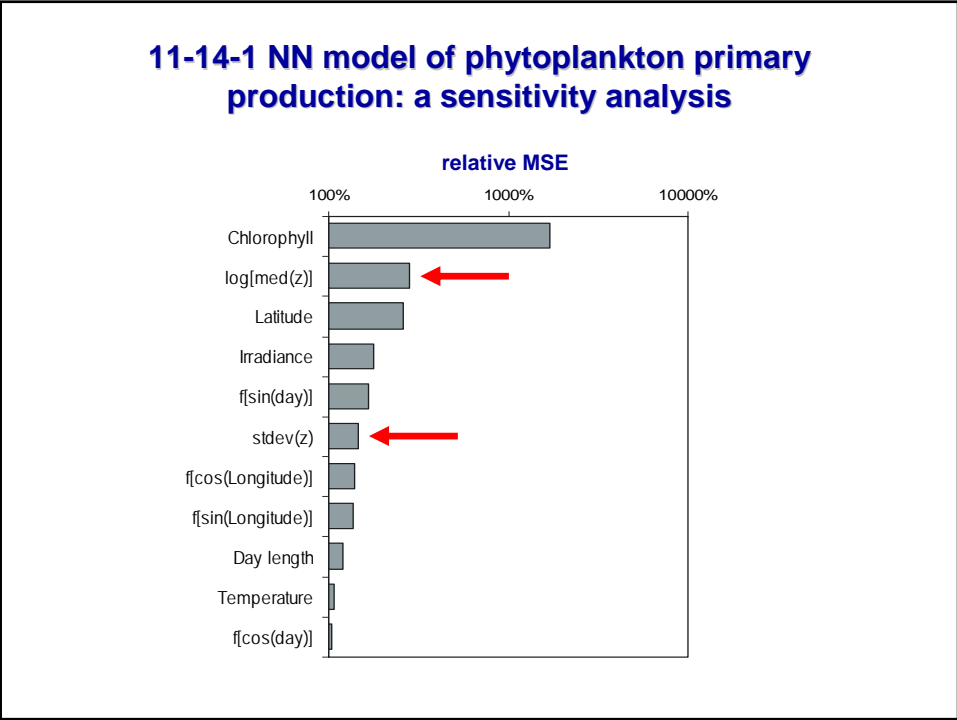
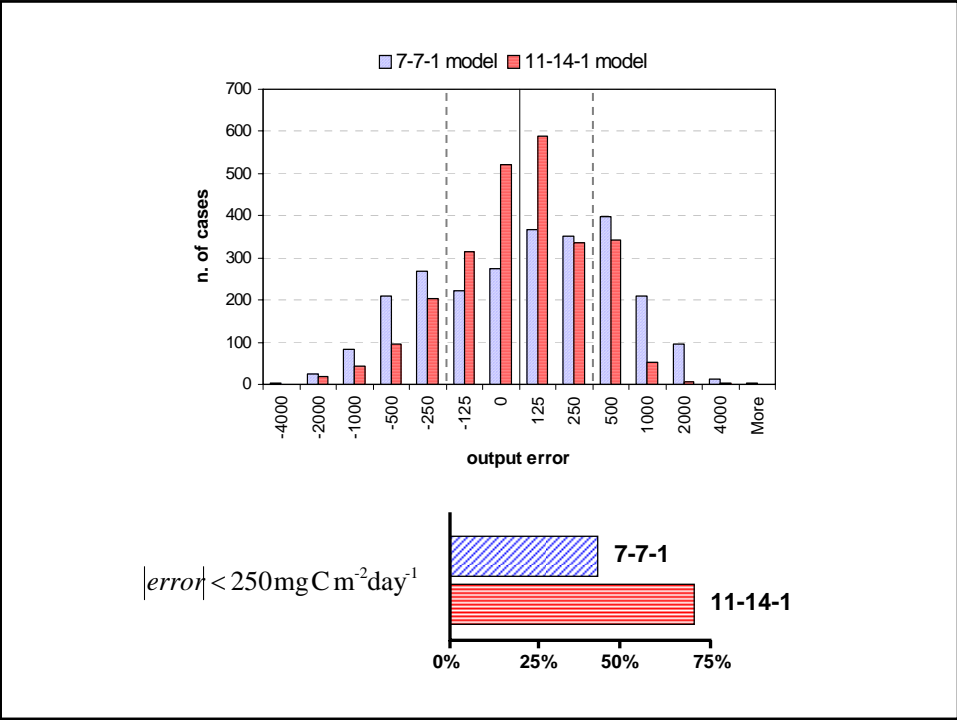
- Phytoplankton surface biomass (as Chl concentration)
- Surface irradiance
- Surface temperature
- Latitude

• Longitude →
$$\begin{cases} longitude_1 = \frac{1}{2} \left[\cos\left(\frac{2\pi \cdot (longitude + 180)}{365}\right) + 1 \right] \\ longitude_2 = \frac{1}{2} \left[\sin\left(\frac{2\pi \cdot (longitude + 180)}{365}\right) + 1 \right] \end{cases}$$

• Date →
$$\begin{cases} date_1 = \frac{1}{2} \left[\cos\left(\frac{2\pi \cdot day}{365}\right) + 1 \right] \\ date_2 = \frac{1}{2} \left[\sin\left(\frac{2\pi \cdot day}{365}\right) + 1 \right] \end{cases}$$

- • Average depth } → 1.25° Lon x 0.75° Lat window
→ • St. dev. of depth }
 • Day length

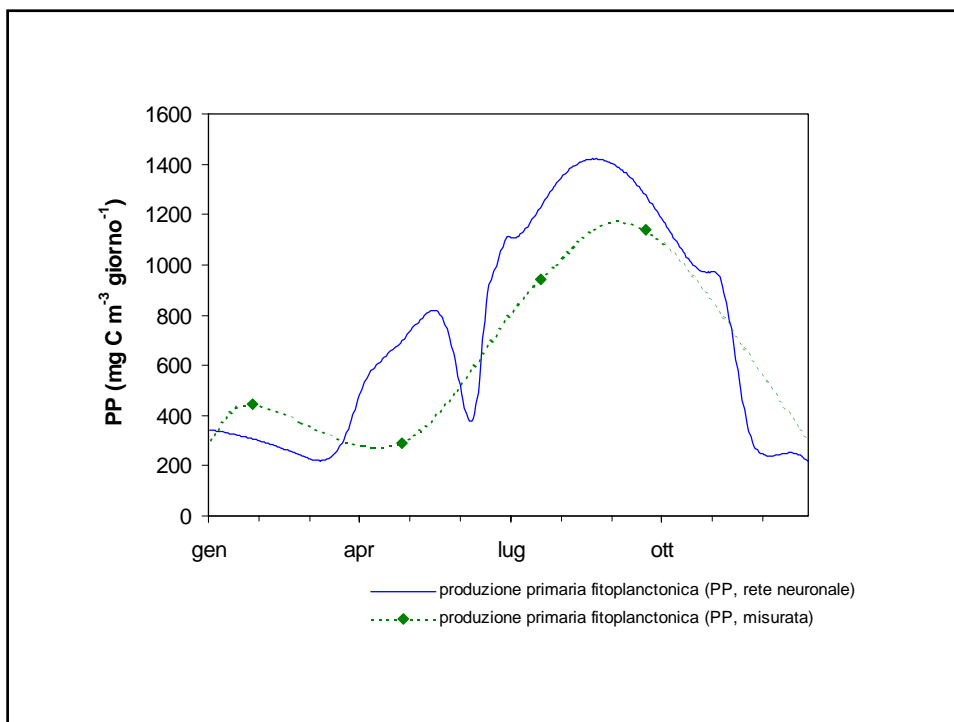
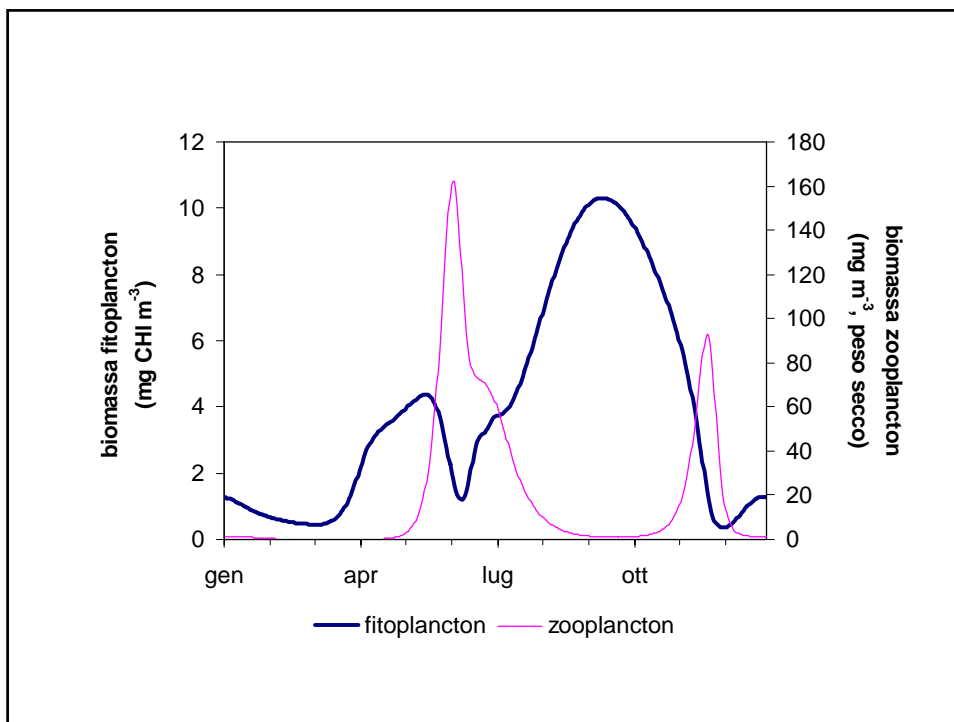


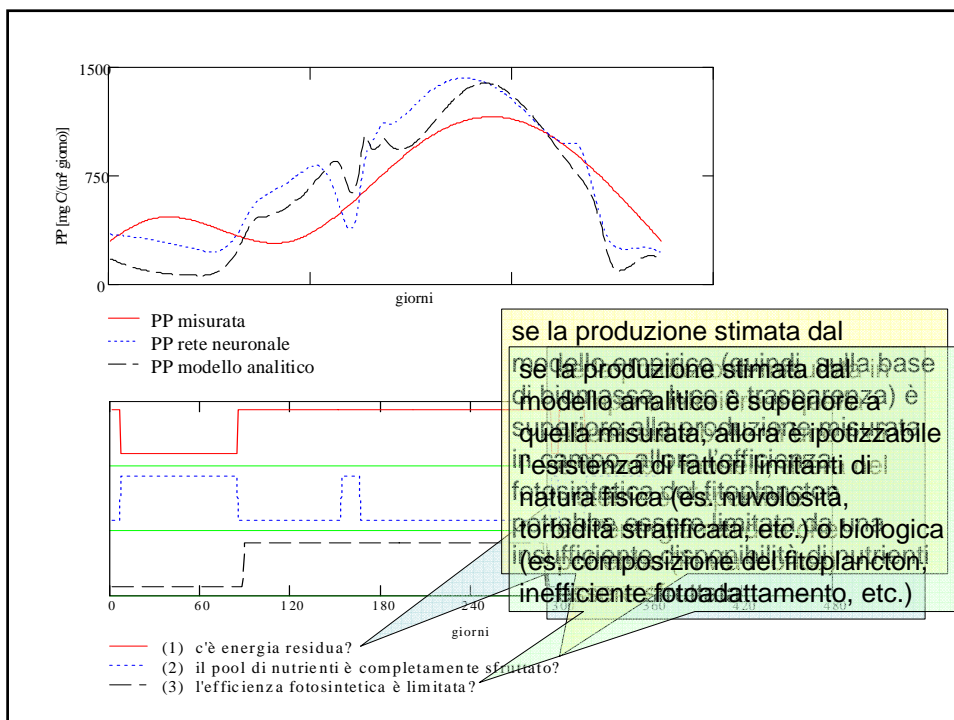
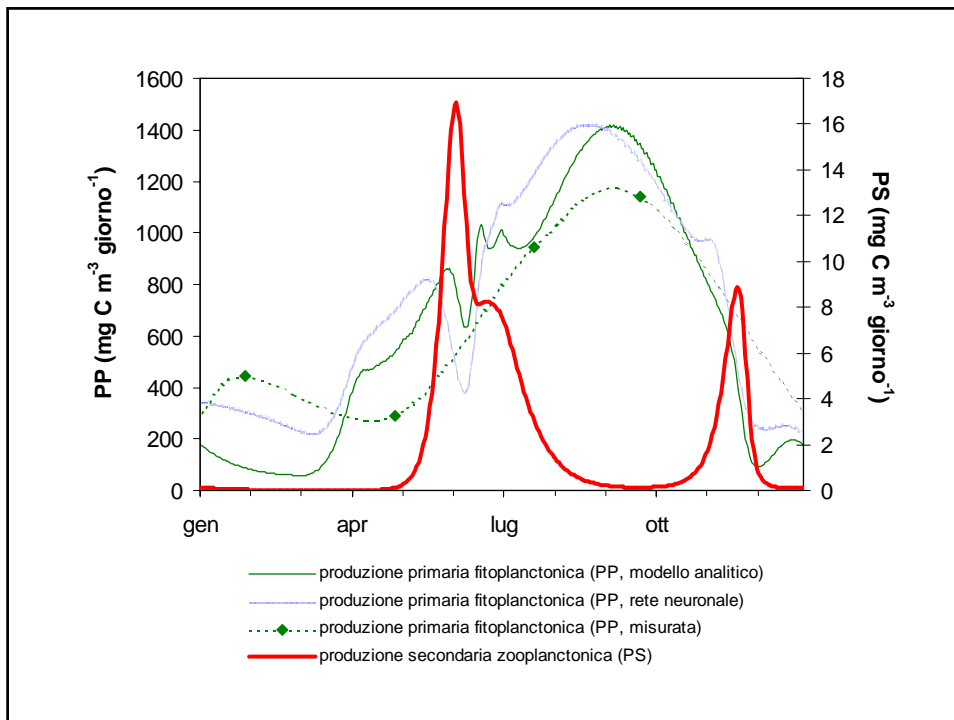


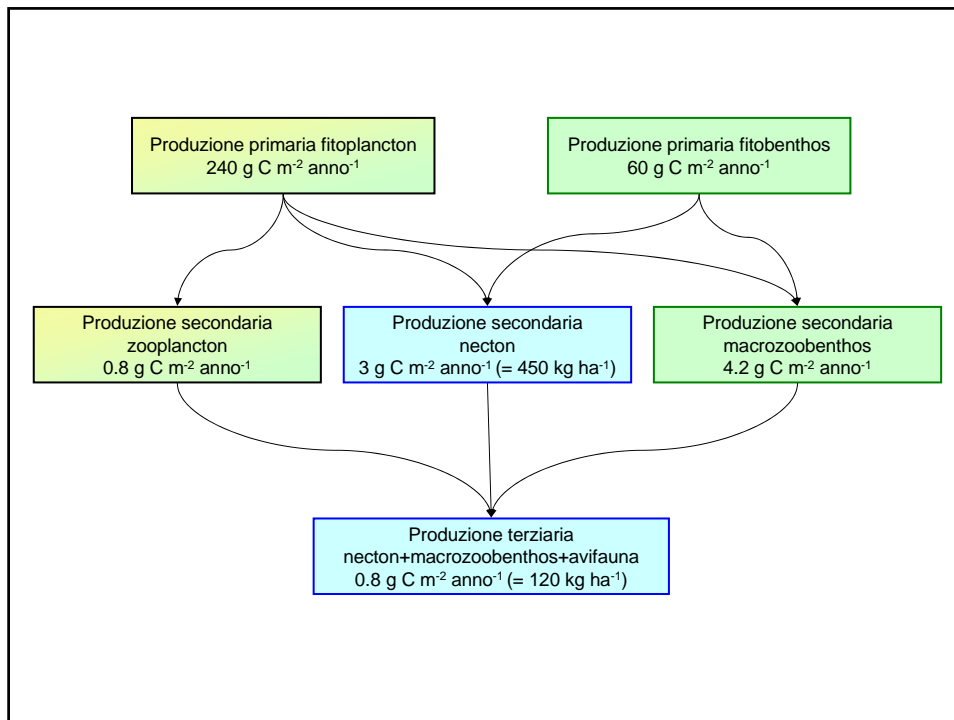
1. **Co-variables** may help to improve neural network output (i.e. predictions)
2. Biological **constraints** damp variability in real systems as well as in neural network models
3. Data and knowledge (i.e. theoretical assumptions, other models, etc.) can be merged into neural network **metamodels**

USO COMBINATO DI DIVERSE TIPOLOGIE DI MODELLI

**Analisi della produttività dello
Stagno del Calich (Sardegna)**







Links utili

- Il mio sito web (reti neurali, analisi dati, etc.)
<http://www.mare-net.com/mscardi>
- Progetto PAEQANN
<http://aquaeco.ups-tlse.fr/>
- Modelling Community Structure in Freshwater Ecosystems
<http://www.springeronline.com/sgw/cda/frontpage/0,11855,4-40109-22-36347264-0,00.html>
- Congressi International Society for Ecological Informatics
<http://www.isei3.org> e <http://www.isei4.org>