

ISTITUTO SUPERIORE DI SANITÀ

**Alghe tossiche:
tecniche di campionamento e identificazione,
e valutazioni statistiche di uso comune**

A cura di
Milena Bruno e Serena Melchiorre
Laboratorio di Igiene Ambientale

ISSN 1123-3117
Rapporti ISTISAN
03/32

Istituto Superiore di Sanità

Alghe tossiche: tecniche di campionamento e identificazione, e valutazioni statistiche di uso comune.

A cura di Milena Bruno e Serena Melchiorre

2003, iii, 33 p. Rapporti ISTISAN 03/32

Le fioriture di alghe tossiche sono un fenomeno ormai consueto e generalizzato in Italia, ma la gestione di questo problema da parte degli organi tecnici locali è spesso non semplice né facilmente realizzabile, sia per la mancanza di competenze tecniche specifiche nel riconoscimento del fenomeno, che per la oggettiva difficoltà nella valutazione dei dati su campo per prevedere l'estensione del danno all'ambiente e del rischio per la salute umana. Il rapporto fornisce delle indicazioni pratiche per la strategia concettuale da seguire nel riconoscimento di specie, nei metodi per l'identificazione delle tossine e per la problematica rappresentata dalla valutazione statistica dei dati su campo. A questo scopo, viene presentato anche l'utilizzo dell'analisi multivariata nelle sue diverse applicazioni.

Parole chiave: Alghe tossiche, Cianofitee, Acqua dolce, Fioriture algali

Istituto Superiore di Sanità

Toxic algae: sampling and identification methods, and common statistical evaluations.

Edited by Milena Bruno and Serena Melchiorre

2003, iii, 33 p. Rapporti ISTISAN 03/32 (in Italian)

The toxic algal blooms are now a common and extended occurrence in Italy, but the management of such a new problem by the local technical authorities is often neither simple nor easy, due to lacking of specific expertise for the detection of this phenomenon, and to objective difficulties in the evaluation of field data, to foresee the extension of the environmental damage and of the health risk. This report gives practical directions for the conceptual strategy in identification of species, for methods in toxins detection, and for the problems represented by statistical evaluations of field data. In this aim, the use of multivariate analysis is proposed, in its different applications.

Key words: Toxic algae, Cyanobacteria, Freshwater, Algal blooms

Per informazioni su questo documento scrivere a: mbruno@iss.it

Il rapporto è accessibile online dal sito di questo Istituto: www.iss.it/pubblicazioni.

Presidente dell'Istituto Superiore di Sanità e Direttore responsabile: *Enrico Garaci*
Registro della Stampa - Tribunale di Roma n. 131/88 del 1° marzo 1988

Redazione: *Paola De Castro e Sandra Salinetti*

La responsabilità dei dati scientifici e tecnici è dei singoli autori.

© 2003 Istituto Superiore di Sanità (Viale Regina Elena, 299 - 00161 Roma)

INDICE

Premessa	iii
-----------------------	-----

Valutazioni statistiche di uso comune in campo ambientale

<i>Mauro Mecozzi</i>	1
Distribuzioni dei dati.....	1
Analisi della varianza.....	4
Analisi di regressione.....	5
Analisi delle serie temporali.....	10
Analisi multivariata.....	13
Cluster analysis.....	14
Analisi delle componenti principali.....	16
Bibliografia di riferimento	19

Ecologia e tecniche di campionamento del fitoplancton e delle specie algali potenzialmente tossiche

<i>Carmela Caroppo</i>	21
Ecologia del fitoplancton	21
Ecologia del fitoplancton potenzialmente tossico	23
Bibliografia	24

Metodi per l'estrazione e la rilevazione delle principali tossine algali

<i>Milena Bruno</i>	26
Analisi strumentale.....	26
Saxitossine.....	26
Anatossina-a	27
Microcistine e nodularine	28
Tossine DSP	28
Metodi enzimantici.....	29
Saxitossine.....	29
Microcistine.....	29
Tossine DSP	30
Biosaggi	30
Bibliografia	32

PREMESSA

Il Piano Sanitario Nazionale italiano esprime con forza l'esigenza di un maggiore e specifico controllo della qualità delle acque e dei rischi sanitari acuti e cronici che possono derivare dal loro consumo, sia in ottemperanza delle leggi e allegati sulla balneazione e sulle acque potabili che per rispondere alla crescente domanda di indicazioni tecniche per affrontare i numerosi eventi portatori di rischio per la salute umana che vedono le acque come principali vettori o protagonisti.

Le fioriture algali tossiche nelle acque adibite ad usi potabili e lungo le coste del territorio nazionale costituiscono un evento sempre più ricorrente e importante, ma problematico nella rilevazione e nella gestione.

Questo rapporto si propone di fornire cognizioni del rischio rappresentato dalle fioriture algali tossiche, criteri per l'identificazione delle specie algali relative, conoscenza dei sistemi di rilevazione e quantificazione della presenza delle stesse nelle acque e cognizioni delle principali strategie statistiche per la gestione dei dati di campo relativi alla dinamica di popolazioni algali tossiche.

Le numerose richieste ricevute dagli enti di controllo territoriale, per specifici corsi sull'argomento, sono la dimostrazione dell'importanza e urgenza di queste tematiche metodologiche e della necessità di trattazioni specifiche e agili che le riguardino.

VALUTAZIONI STATISTICHE DI USO COMUNE IN CAMPO AMBIENTALE

Mauro Mecozzi

Istituto Centrale per la Ricerca Scientifica e Tecnologica Applicata al Mare, Roma

Lo studio della statistica è normalmente demandato agli opportuni corsi universitari per gli studenti e alla consultazione dei testi più appropriati per coloro che essendo “utilizzatori della statistica” si trovano, a causa del proprio lavoro, a dover utilizzare solo alcuni degli strumenti della statistica. Tra i due approcci esiste ovviamente una notevole differenza in quanto se lo studente, per il conseguimento del titolo di studio, necessita di conoscere “tutta la statistica”, “l'utilizzatore della statistica” ha invece un approccio meno didattico e più pragmatico in quanto necessita di conoscere solo alcuni strumenti statistici. Gli studiosi in campo ambientale appartengono alla categoria degli “utilizzatori della statistica” in quanto la conoscenza e l'utilizzo dello strumento statistico deve semplicemente rispondere alla richiesta di fornire un'informazione specifica per verificare una ipotesi relativa al significato dei dati ambientali esaminati.

Questo capitolo è rivolto agli “utilizzatori della statistica” e pertanto il suo scopo è quello di chiarire alcuni aspetti relativi al corretto utilizzo di alcuni test statistici sottolineando il significato interpretativo connesso alla elaborazione dei dati.

Pertanto verranno introdotti i concetti relativi alla distribuzione dei dati, all'analisi della varianza e della regressione, all'analisi delle serie temporali e all'analisi multivariata.

Tutti gli utilizzatori di questo testo sono comunque invitati a riesaminare gli argomenti illustrati in questa sezione mediante la consultazione di specifici testi perché il confronto è sempre il miglior metodo di apprendimento e verifica delle conoscenze acquisite.

Distribuzioni dei dati

Il primo e più semplice test statistico, che si applica ad una serie di dati, è il calcolo del valore medio e della deviazione standard (DS).

Data la serie x_1, x_2, \dots, x_n costituita da n - numeri, il valore medio x_m e la DS saranno rispettivamente:

$$x_m = (x_1 + x_2 + \dots + x_n) / n$$

$$DS = (\sum (x_m - x_i)^2 / (n-1))^{1/2}$$

Queste funzioni possono essere facilmente calcolate con il foglio elettronico Excel mediante le apposite funzioni statistiche incluse:

$$=media(a1:an)$$

$$=dev.st(a1:an)$$

dove a1:an rappresenta la sezione della colonna del foglio elettronico dove sono stati collocati i valori numerici della serie x .

Ma dietro questa apparente semplicità si nasconde un tranello. Siamo sicuri di utilizzare correttamente questi due test? La domanda non è oziosa in quanto media e DS sono grandezze legate a serie di dati aventi distribuzioni normali (gaussiane). Le distribuzioni normali sono

quelle distribuzioni in cui tutti i valori della serie $x_1, x_2 \dots x_n$ sono equidistribuiti intorno al valore medio x_m . Se infatti noi avessimo una serie di n-dati normalmente distribuiti e costruiamo un diagramma riportando sulle ascisse i valori della serie e sull'asse delle ordinate quello del numero di volte (frequenze) con cui ogni i-esimo valore si ripete all'interno della serie, avremmo un diagramma finale come quello di Figura 1; dove vediamo due serie di dati aventi il medesimo valore medio ma deviazioni standard diverse e quindi diverse distribuzioni. Infatti nelle distribuzioni normali la simmetrica ampiezza (distribuzione) dei dati stessi è data dalla relazione:

$$\Lambda_{1/2} = 2 \Lambda * (\text{LN}(2))^{1/2}$$

con Λ pari all'ampiezza della curva, $\Lambda_{1/2}$ pari alla ampiezza della curva a mezza altezza e $\text{LN}(2)$ è il logaritmo naturale di 2. In questa ampiezza noi avremo il 99.7% di probabilità di trovare il "vero" valore medio della distribuzione compreso nell'intervallo $x_m \pm 3 * \text{dev. std.}$

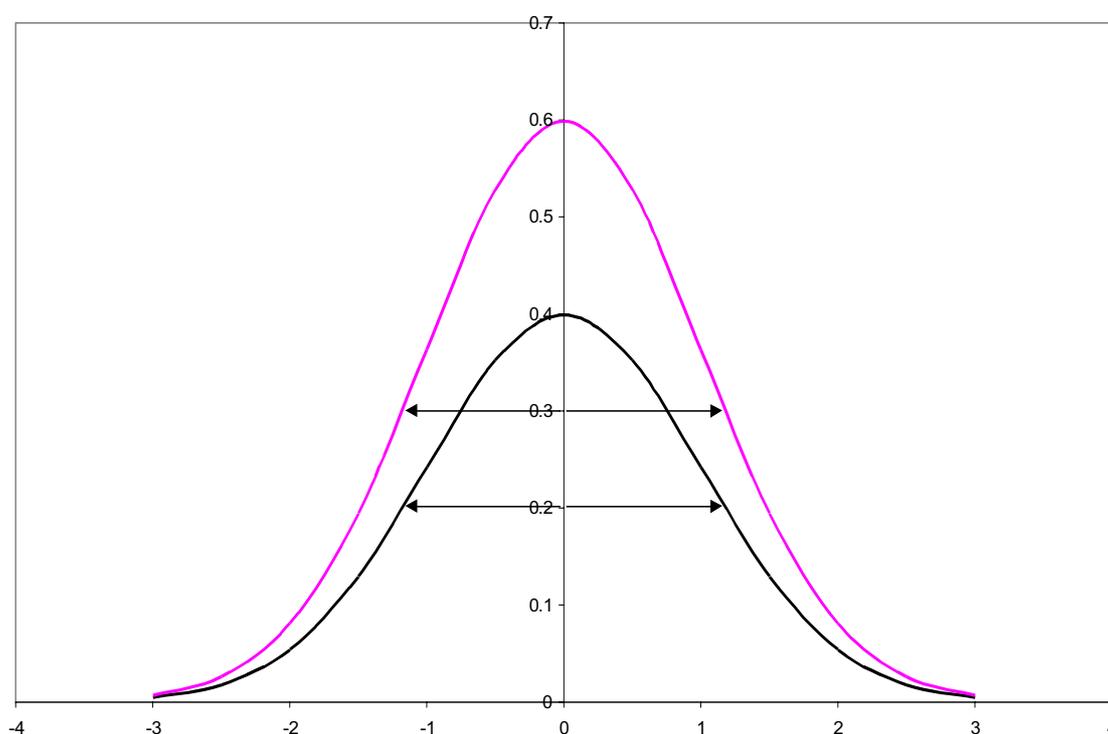


Figura 1. Distribuzioni normali per serie di dati aventi uguale valore medio e deviazioni standard diverse

Occorre riflettere sul fatto che simili distribuzioni in dati ambientali sono piuttosto rare per non dire inesistenti; basti pensare ad un caso tipico come il monitoraggio di una grandezza chimica (es. il fosfato) misurato in una stazione di campionamento in un lago in periodi diversi dell'anno quando sul valore di fosfato incidono contemporaneamente fattori diversi quali apporti esterni, temperatura, produzione primaria, ecc.; in tali situazioni si verificano talvolta casi estremi con valori di DS maggiori del valore medio della serie numerica dei dati.

È evidente che in questi casi limite media e DS non sono rappresentativi della distribuzione dei dati. Occorre, allora, chiedersi perché tutti usano la media e la DS per caratterizzare la serie

di dati in esame e non usano test statistici più specifici e attinenti alla distribuzione dei dati. In realtà le distribuzioni dei dati sono molteplici, basti pensare alle distribuzioni di Poisson, di Lawrence, alle distribuzioni esponenziali, quadrate e così via.

L'analista ambientale non ha, comunque, un reale interesse pratico a conoscere la distribuzione dei dati ma ha viceversa interesse ad individuare un valore "centrale" della sua serie numerica e un altro valore che sia rappresentativo della dispersione dei dati. Un aiuto ci viene dato dalla "statistica non parametrica", la statistica per lo studio delle distribuzioni non normali. La statistica non parametrica è stata sviluppata proponendo test alternativi a tutti quelli previsti dalla statistica delle distribuzioni normali.

L'alternativa della statistica non parametrica per la media e la DS è la mediana e la deviazione assoluta della mediana. Data una serie numerica x_1, x_2, \dots, x_n , la mediana viene calcolata nel seguente modo:

1. I valori sono dapprima ordinati in senso crescente dal più piccolo al più grande.
2. Si considera il valore di "n"; se è dispari la mediana sarà uguale al valore $n/2 + 0.5$ della serie. Ad esempio per una serie di tre valori la mediana corrisponde al secondo valore ($3/2 + 0.5 = 2$), per una serie di cinque valori la mediana corrisponde al terzo valore ($5/2 + 0.5 = 3$). Per sette valori la mediana corrisponde al quarto valore.
3. Se "n" è pari la mediana sarà uguale al valore medio dato dalla somma dei valori rispettivamente $n/2$ ed $n/2 + 1$ della serie. Ad esempio per $n = 4$, la mediana corrisponde al valore medio dato della somma del secondo e del terzo valore della serie ordinata in senso crescente; se $n = 6$ la mediana corrisponde al valore medio del terzo e del quarto valore.

La mediana di una serie numerica di dati può essere calcolata direttamente in Excel tramite l'apposita istruzione mediana:

=mediana(a1:an)

La deviazione assoluta della mediana non è inserita nelle funzioni di Excel come è il caso della DS ma può essere facilmente ottenuta mediante il calcolo di:

$$\text{deviazione assoluta della mediana} = \text{mediana} (|x_i - \text{mediana}|) / 0.6745$$

calcolando cioè la mediana dei i valori assoluti delle differenze tra la mediana della serie x e i valori i -esimi della serie stessa (scarti) e dividendo infine il risultato per il coefficiente 0.6745; tale coefficiente costituisce un fattore di scala per adattare la distribuzione dei dati ad una distribuzione normale.

Ma perché per visualizzare la distribuzione di una serie di dati ambientali la mediana è spesso più rappresentativa della media? La risposta a questa domanda è data dalla maggiore capacità della mediana rispetto alla media di non essere influenzata dalla distribuzione dei dati.

Osserviamo il seguente esempio: la serie numerica 2.5, 2.6, 2.9, 3.4 ha un valore medio 2.85 e un valore di mediana pari a 2.75. Se per un banale errore di trascrizione dei dati, il valore 3.4 viene riportato come 34, il valore medio diviene 10.5, un valore assolutamente non rappresentativo della distribuzione della serie poiché nessun valore della serie originale è vicino a questo valore medio.

Tuttavia, in base alle caratteristiche del calcolo della mediana è facile e allo stesso tempo stupefacente notare che il suo valore rimane comunque 2.75, un valore quindi ancora rappresentativo della distribuzione dei dati nonostante l'errore di trascrizione.

Questo è un eccellente esempio della caratteristica di "robustezza", tipica della mediana e di tutta la statistica non parametrica, di fornire valori effettivamente rappresentativi della distribuzione anche in caso di forti fluttuazioni dei dati. Una discussione sui vari test della statistica non parametrica esula dagli scopi di questa nota introduttiva sull'utilizzo di comuni test statistici, per particolari esigenze si rimanda alla lettura di specifici testi di statistica non

parametrica. È importante sottolineare che ciascun test della statistica per distribuzioni normali dei dati ha un suo equivalente nella statistica non parametrica come riportato in Tabella 1.

Tabella 1. Equivalenze tra test della statistica parametrica e non parametrica

Statistica parametrica (normale)	Statistica non parametrica
Valore medio	Valore della mediana
t-student	Mann-Whitney U-test
ANOVA (analisi della varianza)	Kruskal Wallis test
Regressione lineare	Regressione di Theil

L'ultimo spunto di riflessione è se nell'esame di dati ambientali è più corretto abbandonare l'uso del valore medio e della DS sostituendoli con il calcolo della mediana e della deviazione assoluta della mediana. In assoluto questo approccio è il più corretto; l'analista ambientale ha comunque sempre la possibilità di compiere una scelta calcolando per la propria serie di dati tutti i parametri tipici di una distribuzione e verificando l'esistenza di deviazioni standard paragonabili al valore medio.

In questo caso l'uso del valore medio deve essere scartato poiché non rappresentativo della distribuzione dei dati. Solo nei casi di elevate misure ($n > 500$) l'uso di media e mediana è praticamente indifferente ai fini della caratterizzazione della distribuzione dei dati.

Analisi della varianza

Quando si vogliono comparare serie ripetute di dati, ad esempio n-misurazioni di fosfato presenti in c-gruppi diversi di misurazioni per verificare se i gruppi sono omogenei tra loro, si ricorre all'analisi della varianza (*ANalysis Of VAriance*, ANOVA).

La varianza di una serie di misure ripetute è il quadrato della DS ed è quindi rappresentativa della dispersione dei valori della serie rispetto al valore medio.

L'analisi della varianza è detta ad una via (*one-way ANOVA*) se è uno solo il fattore rispetto al quale si vuole testare la presenza di significative variazioni. Un esempio di applicazione di ANOVA è costituito da una serie di misurazioni di una grandezza chimica, ad esempio il fosfato, misurata in un lago in cinque punti diversi di campionamento per tutti i dodici mesi dell'anno. Se siamo interessati a verificare la presenza di una variabilità stagionale, ipotizziamo che le concentrazioni di fosfato varieranno solo in funzione del periodo di campionamento e pertanto applicheremo la *one-way ANOVA*.

Se volessimo verificare anche l'esistenza di una variabilità dovuta ai differenti luoghi di campionamento, i fattori di influenza della variabilità dei dati saranno allora tempo e spazio, applicheremmo in questo caso la *two-way ANOVA*.

In campo ambientale i dati vengono spesso esaminati rispetto ad una sola fonte di variazione e pertanto in questo testo tratteremo soltanto la *one-way ANOVA*.

La matematica dell'ANOVA è piuttosto semplice in quanto è basata sul calcolo della varianza totale del sistema e sul calcolo della varianza per ogni ciclo di misure replicate.

Ricordiamo che la varianza è il quadrato della DS di una serie di misure.

Pertanto, data una tabella di dati di r-righe (cicli di misurazioni) per c-colonne (repliche per ogni ciclo), l'analisi della varianza calcolerà la varianza derivante dalle $r \cdot c$ misurazioni totali e

la varianza calcolata all'interno dei campioni definita dal valore medio delle varianze osservate per tutti gli r-cicli di misurazioni.

Il rapporto di queste due varianze, definito come il valore sperimentale di F-Snedekor, viene comparato con quello critico tabulato per r-1 ed r*c-c gradi di libertà. Se il valore sperimentale è maggiore di quello critico tabulato ad una coda, per un designato livello di significatività statistica, ad esempio al 95%, allora ANOVA ci indicherà che esistono significative differenze tra i nostri campioni.

In caso contrario, quando il valore sperimentale di F-Snedekor è minore di quello critico tabulato allora le differenze non sono significative ma rientrano nella fluttuazione dei dati.

Come si può vedere, ANOVA è basata su una matematica concettualmente semplice e anche in assenza di uno specifico software essa può essere calcolata con l'ausilio di un foglio elettronico od una piccola calcolatrice, consultando successivamente le tabelle dei valori critici di F-Snedekor ad una coda.

Occorre comunque considerare che ANOVA ci indicherà la presenza di differenze tra i campioni ma non ci dirà dove sono posizionate queste differenze; in parole più semplici ANOVA non ci mostra quali campioni sono significativamente diversi tra loro e quali no.

Per rispondere a questo quesito introduciamo il test della Minima Differenza Significativa (MDS). L'applicazione di questo test è anch'essa piuttosto semplice; si calcola il valore MDS in accordo all'espressione:

$$\text{MDS} = \sqrt{(\text{Varianza all'interno dei campioni}) * (2/c) * t}$$

con "c" numero delle repliche e "t" valore del t-student per r*(c-1) gradi di libertà al 95% di significatività.

Successivamente si calcolano e si confrontano i valori medi degli r-campioni verificando quelli le cui differenze sono maggiori della quantità MDS.

Le differenze tra campioni superiori al valore MDS indicheranno quali campioni sono significativamente diversi tra loro.

Ovviamente il test della "Minima Differenza Significativa" non ha senso quando ANOVA non mostra differenze significative tra i campioni.

Analisi di regressione

L'analisi di regressione è una tecnica matematico-statistica che si pone lo scopo di quantificare una relazione causa-effetto tra una variabile X e un'altra Y; quest'ultima per convenzione viene definita variabile "dipendente" in quanto i suoi valori numerici dipendono dai valori numerici della variabile X che viene invece definita "indipendente".

L'analisi di regressione è comunque una serie piuttosto eterogenea di tecniche matematiche, aventi in comune soltanto il principio dell'algoritmo di calcolo detto dei "minimi quadrati", dovuto al grande matematico e fisico tedesco Karl Friedrich Gauss.

L'analisi di regressione infatti assume differenti nomi in base al tipo di relazione matematica che si vuole stabilire tra le grandezze X e Y. Vediamo in dettaglio: date due serie x_1, x_2, \dots, x_n e y_1, y_2, \dots, y_n , l'analisi di regressione che stabilisce tra le serie "x" e "y" una relazione matematica $y = f(x)$ è detta :

- *lineare* se $y = ax + \text{cost}$
- *polinomiale* se $y = ax^n + bx^{n-1} + \dots + x + \text{cost}$
- *esponenziale* se $y = e^{(bx)} + \text{cost}$

Se infine la serie $y_1, y_2 \dots y_n$ della variabile dipendente Y deve essere esaminata rispetto a più serie di variabili indipendenti quali $x_1, x_2 \dots x_n$ e $z_1, z_2 \dots z_n$ la regressione che stabilirà la relazione $y = f(x, z)$ è detta Regressione Lineare Multipla.

In campo ambientale, si tende, generalmente, ad utilizzare esclusivamente l'analisi di regressione lineare rendendola quasi un uso forzato e indipendente dalla bontà della relazione trovata.

Si verificano quindi casi come quello di Figura 2, dove è osservabile un buon adattamento dei punti sperimentali alla retta e quello di Figura 3 dove l'adattamento della retta non è soddisfacente.

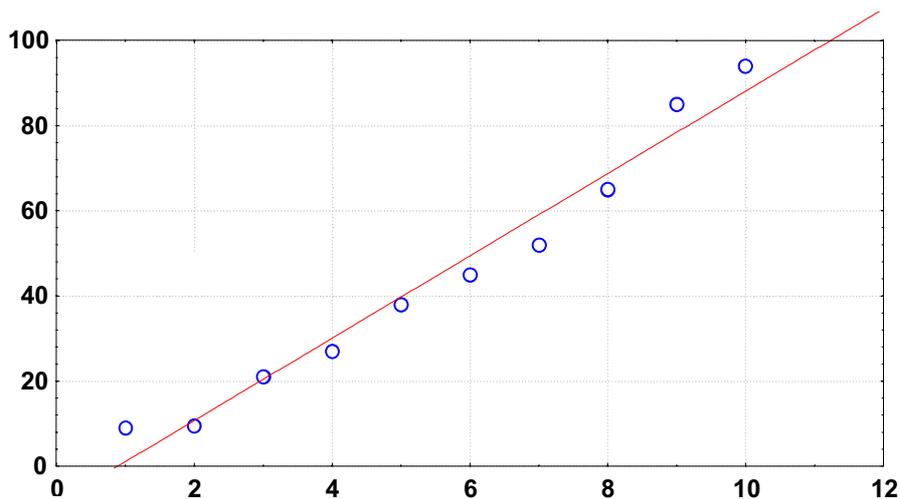


Figura 2. Grafico di una retta di equazione $y = - 8.633 + 9.67x$ (esempio di un buon adattamento di una retta ai punti X e Y sperimentali)

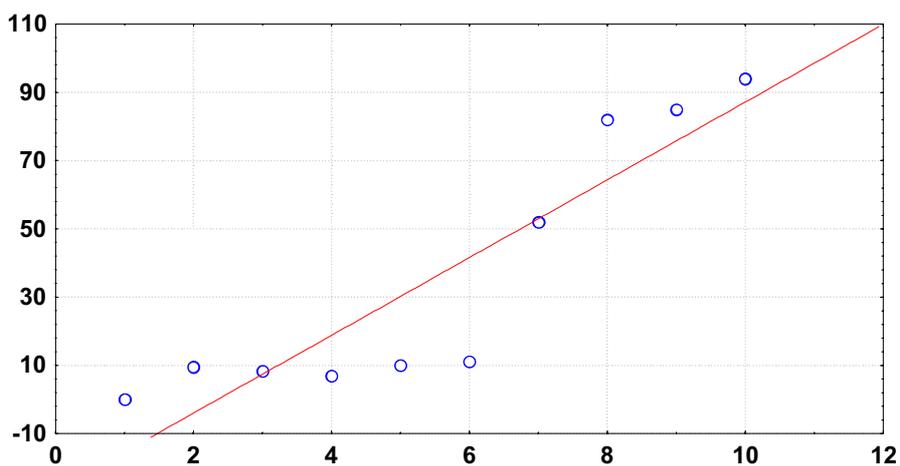


Figura 3. Retta di regressione di equazione $y = - 26.747 + 11.388 x$ (esempio di uso "forzato" della regressione lineare: nel plot l'adattamento della retta ai punti sperimentali X e Y è piuttosto scarso)

In realtà non c'è nessun vincolo che impedisca di applicare ai dati in esame anche altri tipi di regressione oltre la lineare. Viceversa, la ricerca di relazioni matematiche più complesse è un lavoro che rende più rigoroso l'approccio statistico.

Ma come si opera per giudicare la qualità di un'analisi di regressione? Per rispondere a questa domanda introduciamo le definizioni dei parametri caratteristici di tutti i tipi di regressione:

- *Coefficiente di Correlazione (r)*
Stima della significatività della regressione
- *Errore standard della stima*
Stima della dispersione dei valori sperimentali rispetto a quelli teorici previsti dal modello di regressione. È un analogo della DS per i valori medi di una serie numerica.
- *Adattamento della regressione ai valori sperimentali*
Stima della possibilità di migliorare la regressione attraverso l'applicazione di un'altra tecnica di regressione.

L'esame dei tre punti ci dirà come e quanto la regressione si adatta alle coppie x_i e y_i sperimentali. Cominciamo dal calcolo del coefficiente di correlazione "r" che, date le serie x_1, x_2, \dots, x_n e y_1, y_2, \dots, y_n , è definito come:

$$r = \frac{\sum_i \{(x_i - x_m)(y_i - y_m)\}}{\sum_i \{(x_i - x_m)^2\} \sum_i \{(y_i - y_m)^2\}}^{1/2}$$

con x_m e y_m rispettivamente i valori medi delle serie con $i = 1, 2, \dots, n$; è importante sottolineare che $-1 \leq r \leq 1$ dove con il segno negativo si esprimono le correlazioni di inversa proporzionalità tra le variabili X e Y, mentre con il positivo si esprimono le relazioni di diretta proporzionalità.

Infine "r" può essere direttamente calcolato da Excel mediante l'impostazione della funzione:

$$=\text{correlazione}(a1:an,b1:bn)$$

dove a1:an e b1:bn indicano rispettivamente le colonne del foglio elettronico dove sono collocati i valori delle serie X e Y.

Il coefficiente di correlazione è spesso l'unico test associato all'analisi di regressione, rendendola per questo motivo il test statistico più usato e abusato poiché viene spesso associato alle serie X e Y di dati senza verificare se la relazione matematica trovata è statisticamente significativa oppure no.

In altre parole, il computer utilizzato per l'analisi di regressione fornisce sempre e comunque una risposta calcolando il valore del coefficiente di regressione, ma tale risposta deve essere verificata per vedere se la relazione matematica trovata è statisticamente (realmente) significativa.

La verifica che si può effettuare conoscendo il valore di "r" e il numero "n" delle coppie x_i e y_i non richiede necessariamente l'utilizzo di uno specifico software statistico in quanto si può facilmente impiegare Excel.

Si determina il valore del t-student associato alla regressione mediante la formula:

$$t = \frac{|r| * (n-2)^{1/2}}{(1-r^2)^{1/2}}$$

con $|r|$ = valore assoluto di "r" e n = numero di coppie "x", "y".

Per il calcolo del valore t- student per la regressione lineare si imposta in Excel la formula:

$$= r*\text{radq}(n-2)/\text{radq}(1-r^2)$$

Tale valore deve essere comparato con il valore del t-student critico tabulato per n-2 gradi di libertà mostrato in Tabella 2.

Tabella 2. Valori del t-test per regressione lineare ai vari livelli di significatività

Gradi di libertà	Livelli di significatività statistica			
	90%	95%	98%	99%
	0.1	0.05	0.02	0.01
1	6.31	12.71	31.82	63.66
2	2.92	4.30	6.96	9.92
3	2.35	3.18	4.54	5.84
4	2.13	2.78	3.75	4.60
5	2.02	2.57	3.36	4.03
6	1.94	2.45	3.14	3.71
7	1.89	2.36	3.00	3.50
8	1.86	2.31	2.90	3.36
9	1.83	2.26	2.82	3.25
10	1.81	2.23	2.76	3.17
12	1.78	2.18	2.68	3.05
14	1.76	2.14	2.62	2.98
16	1.75	2.12	2.58	2.92
18	1.73	2.10	2.55	2.88
20	1.72	2.09	2.53	2.85
30	1.70	2.04	2.46	2.75
50	1.68	2.01	2.40	2.68
∞	1.64	1.96	2.33	2.58

Esempio: consideriamo una serie $x_1, x_2 \dots x_n$ e una serie $y_1, y_2 \dots y_n$ con $n=6$; al 95% di significatività statistica il valore di t-student tabulato per un test a due code per 4 (n-2) gradi di libertà è pari a 2.78. Pertanto se il t-student calcolato con la formula vista precedentemente è maggiore di 2.78 la correlazione è significativa al livello del 95%. Nel caso opposto, la correlazione non è statisticamente significativa.

Il secondo e significativo parametro associato alla regressione è l'errore standard, detto anche errore standard delle stime che, come già detto, ha una espressione di calcolo analoga a quella della DS:

$$S_{x/y} = \frac{[\sum(y_i - \hat{y}_i)^2]^{1/2}}{(n-2)^{1/2}}$$

dove: y_i = elemento i-esimo della serie y;
 \hat{y}_i = valore i-esimo determinato dalla regressione;
n = numero delle coppie x, y.

Excel non permette direttamente il calcolo di $S_{x/y}$ che deve essere impostato calcolando preliminarmente tutti i valori di \hat{y} . Con Excel si può calcolare l'errore massimo associato ai valori di y mediante la funzione:

$$= \text{err.std.yx}(a1:an,b1:bn)$$

La maggior parte dei comuni software per l'analisi di regressione consentono, comunque, il calcolo di $S_{x/y}$. È pertanto possibile sottoporre un data set X e Y a differenti analisi di regressione, come nell'esempio della Figura 4, diagrammare i risultati e confrontare gli errori standard della stima.

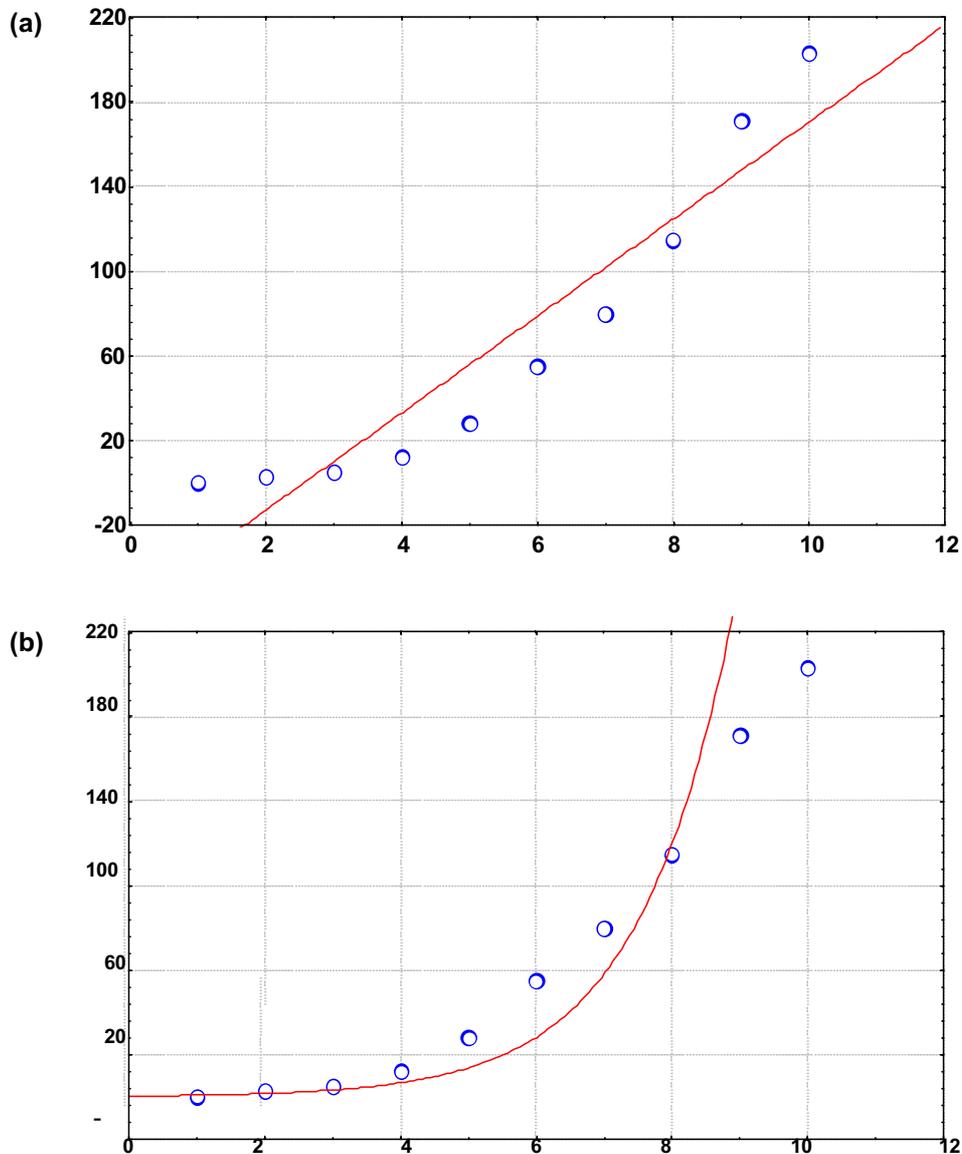


Figura 4. Esempio di applicazione di regressione lineare (a) e polinomiale (del secondo ordine) (b) allo stesso set di dati X-Y

L'esempio della Figura 4 è un caso dove un esame visivo è sufficiente per stabilire la miglior relazione tra i dati X e Y; l'approccio visivo può comunque essere non sempre chiarificante.

L'approccio più rigoroso per stabilire la migliore relazione tra i dati consiste nell'applicare differenti tecniche di regressione, ad esempio lineare e polinomiale, lineare ed esponenziale oppure lineare e multilineare e comparare i valori di errore standard dei residui ottenuti perché $S_{x/y}$ è fondamentale anche per valutare la più idonea tecnica di regressione.

Date quindi due serie "X" e "Y" con n-elementi ciascuna e analizzati mediante regressione lineare e un'altra qualsiasi regressione (polinomiale, esponenziale, lineare multipla se si introducono nuove variabili), detti S_a ed S_b gli errori standard delle stime associate a ciascuna regressione, si determina il rapporto tra i due valori ponendo il valore maggiore al numeratore in modo che il risultante valore del test F sia sempre maggiore di 1.

$$F = \frac{S_a^2}{S_b^2}$$

Questo valore deve essere comparato con quello critico, tabulato ad una coda per $n-2/n-p-2$ gradi di libertà, con "p" pari al numero di variabili introdotte nell'applicazione dell'ulteriore tecnica di regressione.

In un confronto tra regressione lineare ed esponenziale, "p" equivale a zero poiché non c'è introduzione di una nuova variabile, mentre in un confronto tra regressione lineare e polinomiale del secondo ordine "p" equivale ad 1 perché abbiamo introdotto una nuova variabile rispetto alla regressione lineare.

Un valore sperimentale "F" maggiore di quello critico tabulato ci indicherà che gli errori standard delle stime associati alle due regressioni sono significativamente diversi. Pertanto il modello con l'errore minore è quello statisticamente più accurato nel descrivere la relazione matematica che lega le due serie di dati.

Analisi delle serie temporali

Lo studio della variazione di una grandezza chimico-fisica o biologica rispetto al tempo per identificarne ciclicità o trend costituisce l'analisi delle serie temporali. Poiché la variazione di tale grandezza rispetto al tempo è nella maggior parte dei casi una variazione puramente qualitativa, per visionare tale variazione si usa spesso un approccio semplicistico che consiste nel costruire un grafico avente in ascissa il tempo e in ordinata i valori della grandezza in esame (Figura 5).

Questo approccio ha però parecchie limitazioni di cui occorre tener conto, tutte le grandezze sperimentali hanno un *noise* (disturbo) che può creare oscillazioni dei valori. In un esame visivo e qualitativo come quello di Figura 5a, come possiamo essere certi che una variazione temporale sia reale e indipendente dal *noise*?

Un approccio usato per ridurre gli effetti del *noise* consiste nell'effettuare lo *smoothing* o "allisciamento" dei dati.

Esistono differenti metodi di *smoothing* basati sulla sostituzione dei valori presenti nella serie con nuovi valori ottenuti calcolando il valore medio o il valore della mediana per piccoli intervalli della serie stessa, ma senza entrare nel dettaglio delle tecniche di *smoothing*, è sufficiente dire che questo tipo di tecniche comporta sempre l'impossibilità di utilizzare una frazione dei valori della serie. Pertanto se ne sconsiglia l'utilizzo nell'analisi delle serie temporali dove tutti i valori della serie numerica sono importanti per descrivere l'andamento di una grandezza rispetto al tempo.

Un altro caso di errate valutazioni nelle serie temporali, nell'utilizzo del diagramma variabile contro tempo, si presenta quando l'intervallo dei valori della variabile in esame è percentualmente piuttosto piccolo.

In questo caso il suddetto diagramma produrrà una serie di valori oscillanti in un range forzatamente ridotto come è il caso dell'esempio in Figura 5a dove non si avvertono andamenti ciclici. Pertanto, poiché l'esame di una serie temporale è fondamentalmente visivo, ne consegue che l'occhio umano sarà soggetto a visualizzare oscillazioni fittizie ma non reali.

Si potrebbe allora pensare che per meglio evidenziare l'andamento temporale dei valori della variabile si possa restringere il range dei valori delle ordinate come nel caso del diagramma di Figura 5b.

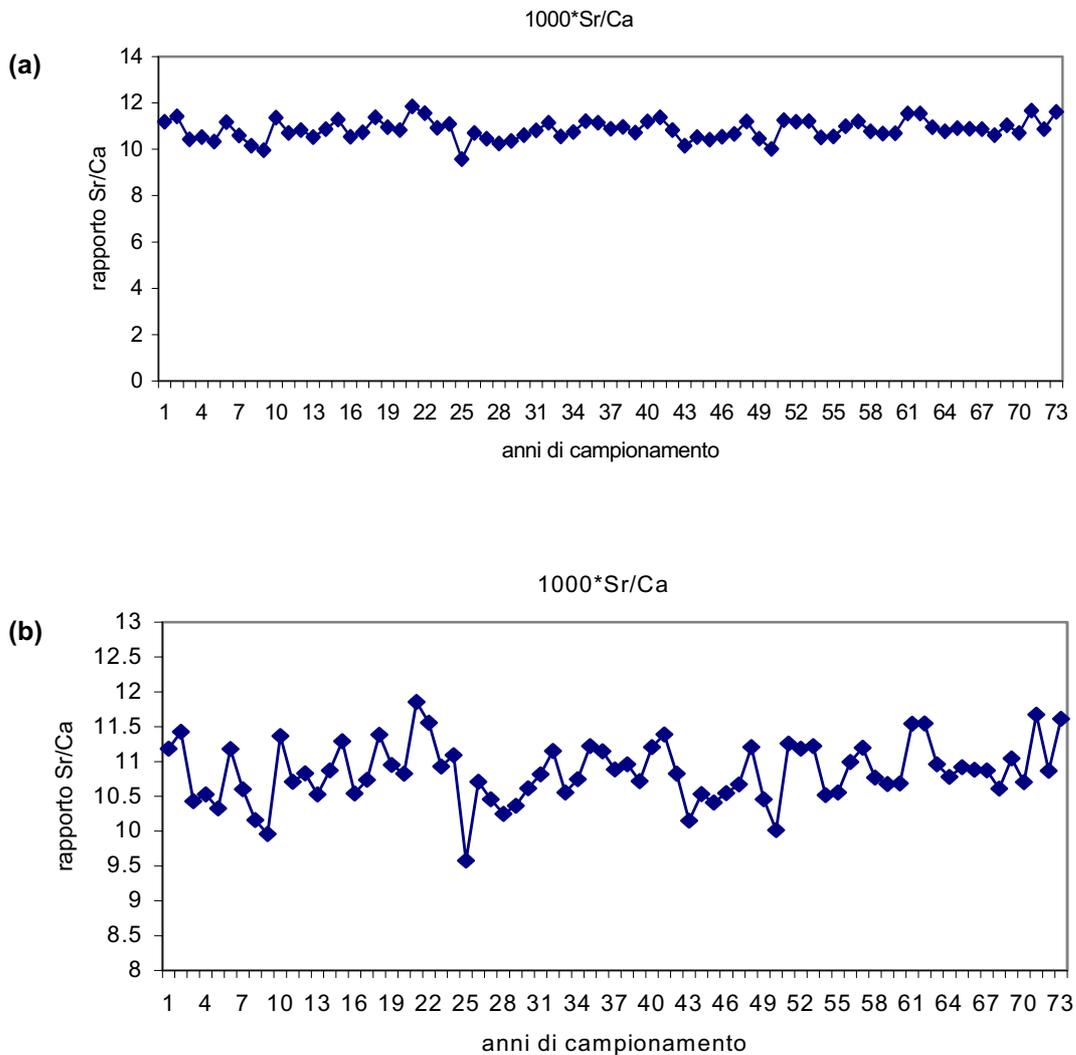


Figura 5. Esempi di diagramma temporale del valore del rapporto Sr/Ca misurato nelle 73 sezioni di un corallo e corrispondenti ad altrettanti anni di crescita del corallo stesso con scale diverse: scala delle ordinate normale (a); scala delle ordinate con restrizione (b)

La restrizione della scala delle ordinate nella Figura 5b è di fatto una amplificazione e corrisponde al caso in cui nella registrazione sonora di un brano musicale a basso volume, noi amplifichiamo il volume aumentando però contemporaneamente anche il *noise*. Infatti innalzando il volume aumentano i fruscii di sottofondo.

Tecniche alternative che consentono di esaminare l'andamento temporale di una serie riducendo il *noise* e senza causare perdita di dati sono l'analisi di autocorrelazione e l'analisi spettrale di Fourier.

Nel nostro caso illustreremo l'analisi di autocorrelazione che ha il duplice vantaggio di poter essere eseguita tramite un comune foglio elettronico quale Excel, utilizzando inoltre quanto già appreso con il paragrafo precedente in merito all'analisi di regressione.

Il principio dell'analisi di autocorrelazione è quello che per evidenziare le vere variazioni di una grandezza rispetto al tempo, si può studiare come una serie temporale $x_1, x_2, x_3, x_4, \dots, x_n$ correla con se stessa; in questo modo, se nella serie temporale sono presenti variazioni temporali regolari (ciclicità), l'analisi di regressione può evidenziarle poiché in presenza di una ciclicità, una parte dei dati deve correlarsi con un'altra parte dei dati stessi. Ma come si attua l'analisi di autocorrelazione e cosa vuol dire correlare una serie con se stessa?

Data la serie numerica $x_1, x_2, x_3, x_4, \dots, x_n$, si scala la serie di una quantità (detta lag) corrispondente ad un arbitrario intervallo temporale.

Per la serie numerica x , avente "n" misurazioni effettuate ad "n" (mesi) tempi diversi si sposta di una quantità pari ad un mese la serie ripetendo l'operazione n-3 volte (Figura 6a). Successivamente si calcoleranno i coefficienti di correlazione della serie x_1-x_7 con la serie x_2-x_8 , della serie x_1-x_6 con la serie x_3-x_8 , della serie x_1-x_5 con la serie x_4-x_8 , della serie x_1-x_4 con la serie x_5-x_8 fino a giungere all'ultima serie che nel nostro esempio è la serie x_1-x_3 con la serie x_6-x_8 . In questo caso, poiché abbiamo ipotizzato una serie di campionamenti mensili, usiamo un lag (intervallo) pari ad 1, ma possiamo prendere intervalli superiori dipendenti dalle nostre esigenze relative; l'esempio successivo con lag = 2 è riportato in Figura 6b.

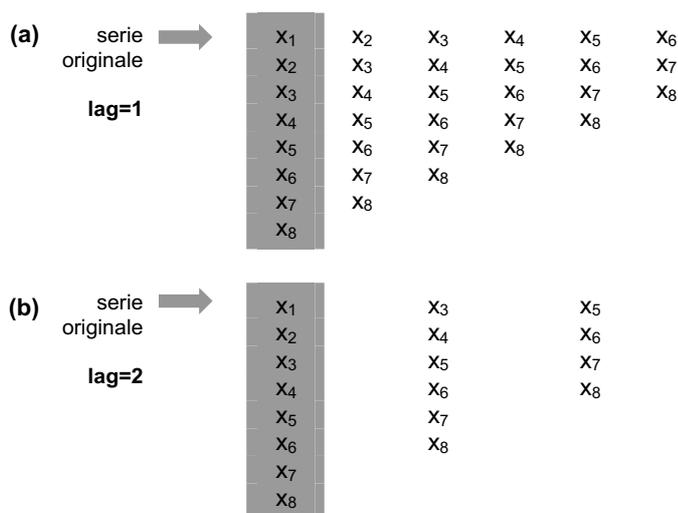


Figura 6. Esempi d'impostazione di una serie numerica per il calcolo dell'autocorrelazione con lag=1 (la serie originale è quella presente nella prima colonna) (a) e con lag=2 per la stessa serie (b)

I vantaggi dell'approccio basato sul calcolo dei coefficienti di autocorrelazione sono:

1. riduzione del *noise* dei dati in quanto ogni punto della serie numerica originale viene sostituito da un valore di autocorrelazione che tiene conto di un insieme di più valori;

2. poiché un coefficiente di correlazione deve per sua stessa definizione, essere compreso tra -1 e 1, tutti i plot (autocorrelogrammi) esaminati saranno compresi tra -1 e 1 e potranno quindi essere messi in un unico diagramma facilitando la comparazione dell'andamento temporale di più variabili;
3. non ci sono perdite di dati come nel caso dello *smoothing*.

In Figura 7 riportiamo l'autocorrelogramma ottenuto per la stessa serie di valori riportati nella Figura 5 ed è immediato vedere che esistono differenze rilevanti. La serie in questione presenta, infatti, una ciclicità che ricorre ogni 5-6 anni (Figura 7) che non risulta invece evidente dai diagrammi delle precedenti figure.

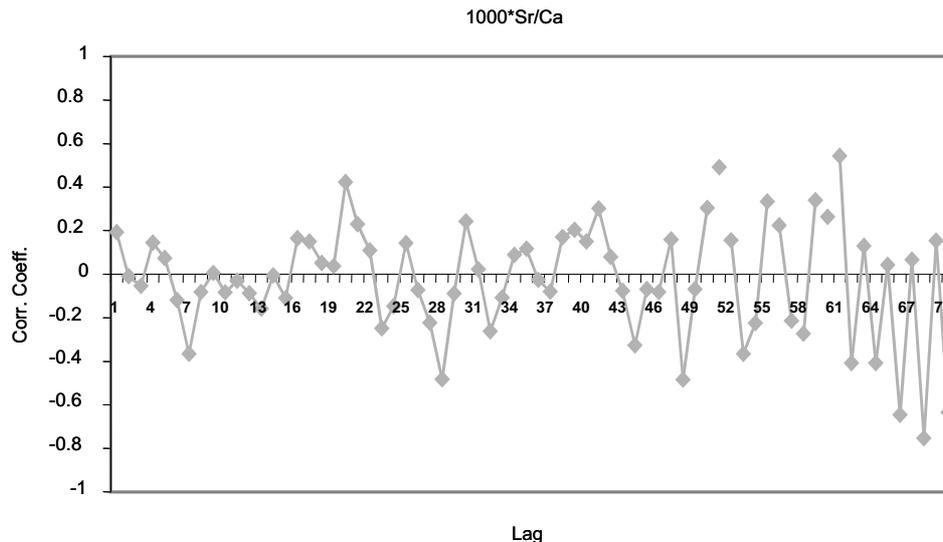


Figura 7. Autocorrelogramma della stessa serie temporale delle figure precedenti

Analisi multivariata

L'analisi multivariata raccoglie una grande famiglia di tecniche matematico-statistiche che come dice il nome "multivariata", consentono l'esame simultaneo di tutte le variabili chimico-biologiche caratteristiche di un set di campioni. Pertanto, se i nostri campioni sono contraddistinti da dieci variabili (es. nitrati, cloruri, clorofilla, ecc.), essi potranno essere esaminati non più per i singoli andamenti di ciascuna variabile ma simultaneamente per tutte le dieci variabili. In altre parole grazie all'esame simultaneo di tutte le variabili, l'analisi multivariata consente una completa classificazione dei campioni.

Ora è bene precisare che l'analisi multivariata non ha soltanto scopi di classificazione poiché sono possibili anche utilizzazioni di tipo modellistico, analoghe per finalità all'analisi di regressione multipla, ma in campo ambientale è intuitivo che l'approccio di classificazione è quello più applicato in quanto consente di esaminare similitudini tra i campioni.

La famiglia delle tecniche multivariate è molto ampia ma per gli scopi di questo corso esamineremo soltanto la *Cluster Analysis* (CA) e l'analisi delle componenti principali (*Principal Component Analysis*, PCA); queste tecniche presentano tra loro analogie di impiego e come tali sono complementari per la classificazione di dati ambientali.

Occorre comunque premettere che la matematica dell'analisi multivariata è basata sull'algebra delle matrici, certamente più complessa della matematica incontrata finora nell'analisi univariata. Pertanto non ci soffermeremo nell'illustrazione della matematica associata a queste due tecniche ma viceversa focalizzeremo gli scopi e le applicazioni di CA e PCA nel campo dell'analisi ambientale.

Cluster analysis

La parola *cluster* significa agglomerato o raggruppamento e la CA ha, infatti, lo scopo di raggruppare i dati per similitudine tenendo conto di tutte le variabili incluse nel data set. Per lo studio della similitudine esistono metodi detti "gerarchici" e metodi detti "non gerarchici" e la CA appartiene al gruppo dei metodi gerarchici. Esaminiamo ora alcuni aspetti fondamentali della CA per capire con quale criterio essa stabilisce similitudine tra campioni.

Dato un insieme di punti in uno spazio bidimensionale come quello della Figura 8 è intuitivo che i dati (campioni) più simili sono quelli più vicini tra loro; pertanto la distanza geometrica tra due punti, descritta dal teorema di Pitagora, è un criterio di similitudine. Si veda l'esempio dei punti di coordinate x_1, y_1 e x_2, y_2 .

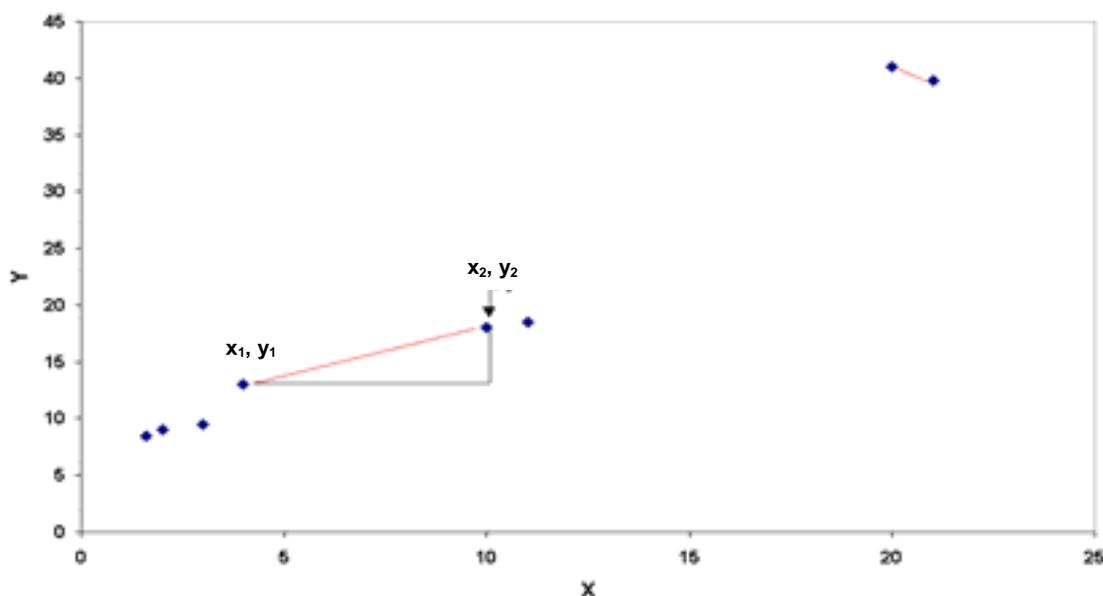


Figura 8. Distanza tra i due punti di coordinate x_1, y_1 e x_2, y_2 come ipotenusa del triangolo rettangolo costruito con le proiezioni delle coordinate x e y sugli assi

La distanza tra i punti 1 e 2 in uno spazio bidimensionale è quindi calcolata mediante il teorema di Pitagora come:

$$D_{12} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

e in uno spazio n -dimensionale la distanza sarà analogamente calcolata tenendo conto della nuove variabili:

$$D_{12} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2 + \dots + (w_2 - w_1)^2}$$

Esistono altri tipi di distanze per calcolare la similitudine tra campioni ma la loro definizione esula dallo scopo introduttivo di questo lavoro e pertanto continueremo a trattare della distanza euclidea, in quanto i concetti delle applicazioni della CA sono indipendenti dal tipo di distanza utilizzata.

Una volta definita la distanza geometrica si può successivamente stabilire la similitudine relativa tra due oggetti “a” e “b” calcolandola come il rapporto:

$$S_{ab} = \frac{D_{ab}}{D_{max}} \cdot 100$$

dove D_{ab} è la distanza multivariata tra i due oggetti calcolata in accordo al teorema di Pitagora e D_{max} è la massima distanza determinata calcolando e paragonando tutte le possibili $n \cdot (n-1)/2$ distanze all'interno del data set contenente n-oggetti.

È evidente, quindi, che una similarità prossima al 100% indicherà elevata similitudine tra gli oggetti e deviazioni dal valore 100% indicheranno invece differenze (Figura 9). Questo appena illustrato è il primo step della CA; successivamente gli oggetti più simili tra loro sono “fusi” in un singolo oggetto e il calcolo della similitudine è ripetuto iterativamente. Il calcolo si interrompe una volta stabilito la massima dissimilarità tra gli oggetti.

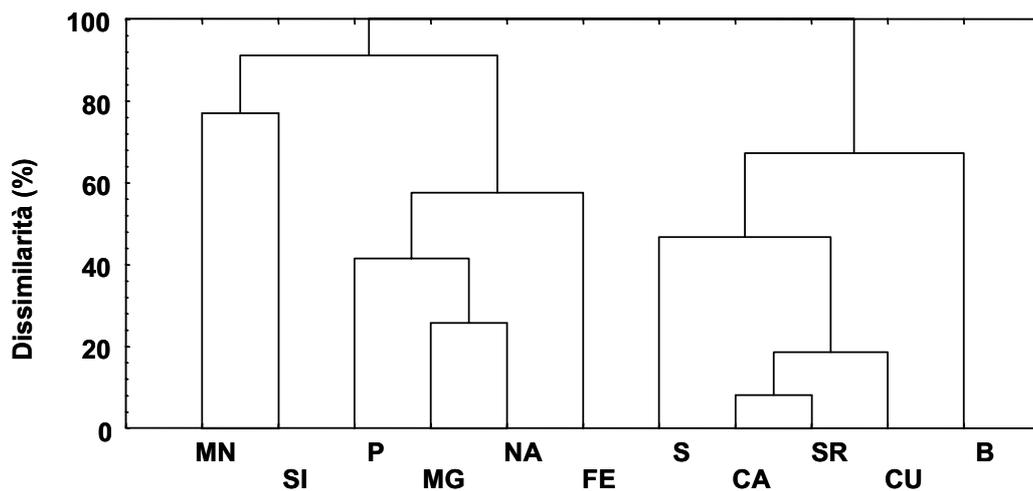


Figura 9. Esempio di CA (dendrogramma) eseguito sulle variabili di un data set sperimentale relativo alla concentrazione dei metalli presenti nelle sezioni di un corallo – le altezze dei segmenti indicano la dissimilarità (es. Ca e Sr sono dissimili al 10%, simili al 90%)

La CA può essere eseguita sui campioni per vedere quelli che sono più simili tra loro; e sulle variabili per vedere quelle che sono tra loro correlate. In ogni caso è bene puntualizzare che la CA è semplicemente una tecnica di screening dei dati con il compito di fornire indicazioni preliminari sul blocco di dati in esame come nell'esempio di Figura 10.

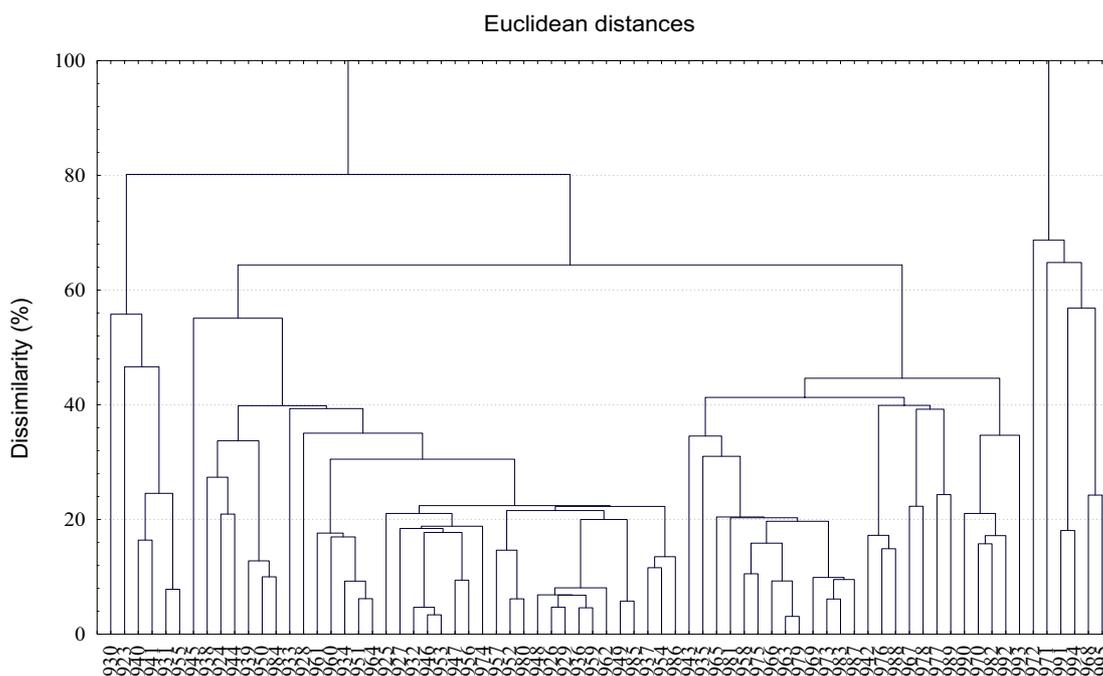


Figura 10. Dendrogramma sul data set della concentrazione dei metalli determinato in differenti sezioni di un corallo (le sigle in ascissa corrispondono alle sezioni)

Infatti la CA fornisce solo la similitudine (o dissimilitudine) tra i campioni come riportato per le sezioni del corallo in Figura 10. La CA mostra che ad esempio che le sezioni 1963 e 1979 hanno una dissimilarità del 5% circa e analogamente abbiamo le coppie 1932, 1946 e 1953, ma non ci mostra quali variabili determinano tale similitudine. Inoltre essa non ci indica l'importanza relativa (significatività) delle variabili che caratterizzano il data set in quanto alcune variabili potrebbero non essere importanti per descrivere il nostro data set. Questa capacità di determinare similitudini tra campioni indicando simultaneamente anche le variabili che determinano similitudini o dissimilitudini è invece la caratteristica portante dell'analisi delle componenti principali (PCA).

Analisi delle componenti principali

L'algoritmo della PCA fu sviluppato nei primi anni del 19° secolo ma la tecnica ha cominciato ad essere intensivamente applicata soltanto dopo gli anni 70 a seguito della diffusione dei personal computer; infatti la matematica connessa alla PCA richiede calcoli piuttosto complicati anche per matrici di dati aventi dimensioni ridotte. Per questa ragione, a differenza di quanto fatto per la *Cluster Analysis*, non ci soffermeremo nel descriverla, ma viceversa puntualizzeremo quelli che sono gli aspetti operativi della PCA e come essa trasforma la matrice dei dati.

Cominciamo col fissare la nostra tabella di dati "M" aventi dimensioni $R * C$ dove "R" sono le righe (o campioni) inclusi nel set e "C" sono le variabili misurate per ciascun campione. L'applicazione della PCA determina che la matrice "M" sarà decomposta nel prodotto di due nuove matrici:

$$M = S * L$$

dove S (matrice degli *score*) è la matrice di dimensione avente “R” righe pari ai campioni originali e “P” colonne corrispondenti ad un set di nuove variabili definite dalla tecnica e dette componenti principali o fattori ed “L” (matrice dei *loading*) è una matrice di dimensioni “P” righe, corrispondenti alle nuove variabili già definite nella matrice S e “C” colonne corrispondenti alle originali variabili della matrice M.

Geometricamente le componenti principali sono assi che attraversano un set di dati multivariati come è il caso di Figura 11, minimizzando la varianza delle vecchie variabili cioè la distanza che si ottiene proiettando gli oggetti sugli assi delle nuove variabili. Il primo asse (o componente principale) giustificherà la porzione più grande di varianza e gli assi successivi saranno quelli che giustificheranno porzioni sempre più piccole della varianza. Un importantissimo aspetto è che ogni asse è comunque indipendente (ortogonale) dagli altri, quindi in termini strettamente matematici, ogni asse apporta un contributo “personale e indipendente” alla varianza e quindi alla descrizione del sistema.

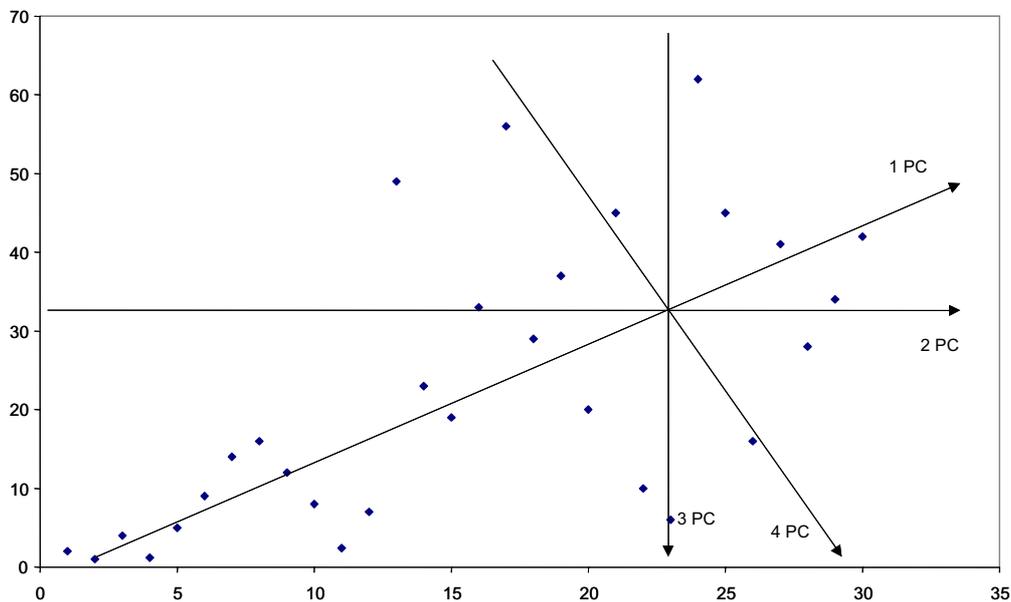


Figura 11. Spiegazione geometrica della PCA. Ogni nuovo asse (componente principale) spiega una porzione della varianza totale del sistema

Definita in questo modo il ruolo della PCA può non essere chiaro ma occorre precisare che il numero “P” delle nuove variabili (componenti principali) è sempre molto più piccolo del numero “C” delle vecchie variabili. Inoltre la matrice S (matrice degli *score*) che rappresenta come già detto il nuovo set di dati, contiene la medesima informazione qualitativa e quantitativa della matrice originale “M” contenente le variabili misurate sperimentalmente.

Per meglio comprendere questo aspetto della PCA esaminiamo due casi pratici.

Il primo caso è quello dei dati di 11 metalli determinati in 73 sezioni di un corallo, esempio già utilizzato per la CA. Volendo esaminare e determinare i campioni di corallo aventi simili (e diverse) concentrazioni degli undici metalli dovremmo costruire un plot 11-dimensionale, cosa ovviamente impossibile.

Applicando però la PCA otteniamo la riduzione delle originali 11 variabili a tre sole nuove variabili (componenti principali). Questo ci consente di ottenere un plot in due o tre dimensioni dei campioni che ci permetterà di esaminare più efficacemente le similitudini di composizione, ma soprattutto la matrice dei *loading* (Tabella 3) ci dà la possibilità di identificare quale delle variabili originali è realmente significativa nello spiegare le caratteristiche dei campioni in esame.

Tabella 3. Loading risultanti dalla PCA. I loading statisticamente significativi al 95% hanno un valore superiore a 0.7 e sono riportati in neretto. “% V.” è la percentuale di varianza giustificata da ogni fattore

Metallo	Factor 1	Factor 2	Factor 3
B	.364177	-.845862	-.071383
Cu	.823376	-.400171	-.064377
Fe	.638446	.255556	-.486781
Na	.633264	.615482	.177357
Mg	.700189	.546366	.016291
S	.748001	.179907	.128223
Sr	.897583	-.258203	.136586
P	.207013	.905543	-.066087
Si	.197573	-.265912	-.816943
Ca	.883746	-.299677	.061958
Mn	-.240032	.295133	-.699441
%V.	39.9	25.3	13.4

Varianza totale spiegata dai tre fattori = 78.6 %

Vediamo pertanto come alcune variabili (Fe e Na) non sono significative dal punto di vista statistico; in altre parole esse non variano significativamente nei campioni esaminati e pertanto non sono importanti per la loro caratterizzazione.

Il secondo caso è relativo alla matrice dei dati degli assorbimenti nell'infrarosso di un set di trenta campioni di mucillagini marine; poiché ogni lunghezza d'onda corrisponde ad una variabile, dato il range di lunghezze d'onda comprese nell'infrarosso tra i 2 e 100 micron, avremo un totale di circa 2500 variabili e pertanto la nostra matrice di dati avrà dimensioni 30 righe (campioni) * 2500 colonne (variabili costituite dalle lunghezze d'onda).

Un semplice esame per verificare quali campioni sono da ritenersi simili tra loro per la presenza di assorbimenti simili è pressoché impossibile poiché il “data handling” simultaneo di tutte le variabili di una simile matrice comporta il maneggiare grafici di 2500 dimensioni. Tale matrice non può quindi essere tabulata e occorre ridurre la dimensionalità prima di qualsiasi esame. Dopo l'applicazione della PCA assistiamo ad uno stupefacente risultato. Le componenti principali individuate dalla PCA sono soltanto due e di conseguenza la matrice degli *score*, che ripetiamo essere la nuova matrice dei campioni in esame, diventa una matrice 30 * 2; i campioni di mucillagini potranno così essere diagrammati su un comune grafico bidimensionale ricordando comunque che l'informazione contenuta nella matrice originale 30*2500 è interamente contenuta anche nella matrice degli *score*.

In Figura 12 riportiamo il plot relativo che ci consente di confrontare per similitudine tutti i campioni di mucillagini mediante l'informazione dei loro spettri infrarossi.

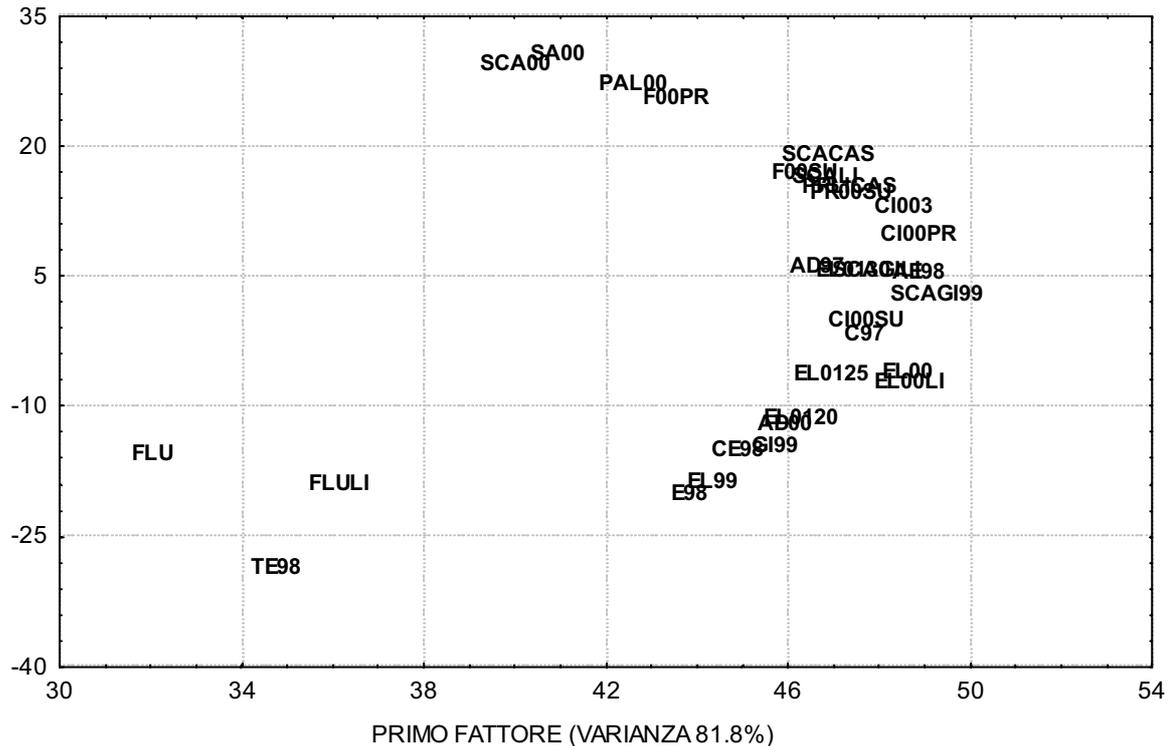


Figura 12. Diagramma delle due prime componenti principali che permettono di visualizzare le similitudini tra la composizione della sostanza organica tra i campioni esaminati (le sigle sono relative a campioni di mucillagini marine prelevate dal 1997 al 2000)

Ovviamente tutta l'applicazione della PCA su dati ambientali presenta differenti aspetti che devono essere valutati caso per caso.

Per comprendere a fondo la potenza applicativa di una tale tecnica occorre studiare di pari passo gli step fondamentali di applicazione ed esaminare casi pratici riportati nella letteratura scientifica.

Per questa ragione riportiamo per la PCA una approfondita bibliografia.

Bibliografia di riferimento

Miller JM, Miller JN (Ed.). *Statistics for analytical chemists*. Chichester (UK): Ellis Horwood Series; 1989.

Per la distribuzione dei dati:

Analytical Methods Committee. Robust statistics- How not to reject outliers. *Analyst* 1989;114:1693-7.

Conover WJ. *Practical non-parametric statistics*. New York: Wiley; 1971.

Han JH, Ward JI, Lavine BK. The problem of adequate sample size in patten recognition studies. *Journal of Chemometrics* 1989;4:91-6.

Machuca I, Cortes JD, Fuentes-Arderiu X. Consequences of applying parametrics methods to non-gaussian distribution in the validation of analytical results. *Clinica Chimica Acta* 1992;209:215-7.

Per l'analisi della varianza (ANOVA):

Hirsch RH. Analysis of variance in analytical. *Chemistry Analytical Chemistry* 1977;49:691A-700A.

Mecozzi M, Cicero AM. Descrizione di un metodo di confronto per la valutazione di omogeneità spaziale e temporale di dati di monitoraggio ambientale. *Acqua Aria* 1995;9:923-6.

Per l'analisi di regressione:

Clementi S, Fringuelli F, Savelli G. L'analisi di regressione lineare applicata a problemi di chimica fisica organica: un esame critico. *La Chimica e l'Industria* 1978;60:598-610.

Per l'analisi delle serie temporali:

Brereton R. *Chemometrics, the application of mathematics and statistics to laboratory systems*. Chichester (UK): Ellis Horwood Series; 1990.

Per lo smoothing dei dati:

Mecozzi M. Tecniche di *smoothing* per l'eliminazione del *noise* strumentale in spettroscopia. *Rassegna Chimica* 2000;50: 7-12.

Per la Cluster Analysis:

Massart DL, Kaufman L (Ed.). *The interpretation of analytical chemical data by the use of cluster analysis*. Malabar (Florida): Edizioni R.E. Krieger Publishing Company; 1989.

Brereton R. *Chemometrics, the application of mathematics and statistics to laboratory systems*. Chichester (UK): Ellis Horwood Series; 1990.

Per la PCA:

Ausili A, Mecozzi M, Gabellini M, Ciuffa G, Mellara F. Physical chemical characteristics and multivariate analysis of contaminated harbour sediments. *Water Science & Technology* 1998;37:131-9.

Bisani ML, Clementi S, Wold S. Chemiometria II: analisi delle componenti principali (PCA) ed analisi delle variabili latenti (PLS). *La Chimica e l'Industria* 1982;64:655-65.

Brereton R. *Chemometrics, the application of mathematics and statistics to laboratory systems*. Chichester (UK): Ellis Horwood Series; 1990.

Cripps GC. Baseline levels of hydrocarbons in seawater of the Southern Ocean: natural variability and regional patterns. *Marine Pollution Bulletin* 1992 ; 24: 109-14

Massart DL, Kaufman L (Ed.). *The interpretation of analytical chemical data by the use of cluster analysis*. Malabar (Florida): Edizioni R.E. Krieger Publishing Company; 1989.

Mecozzi M, Cardarilli D, Pietrantonio E, Amici M. Estimation of the similarity in the qualitative composition of the humic substance in marine sediments by means of an UV spectroscopic library. *Chemistry & Ecology* 2001;17:239-54.

Meglen RR. Examining large databases: a chemiometric approach using principal component analysis. *Marine Chemistry*, 1992;39:217-37.

Todeschini R. *Introduzione alla chemiometria*. Napoli: EdiSes srl Editore; 1998.

ECOLOGIA E TECNICHE DI CAMPIONAMENTO DEL FITOPLANCTON E DELLE SPECIE ALGALI POTENZIALMENTE TOSSICHE

Carmela Caroppo

Consiglio Nazionale delle Ricerche, Istituto Ambiente Marino Costiero, Sezione di Taranto

Ecologia del fitoplancton

Gli studi più recenti sull'ecologia del fitoplancton hanno evidenziato l'opportunità di affiancare allo studio tassonomico in senso lato anche quello diretto a considerare le diverse classi dimensionali. È così emerso il contributo offerto dalla frazione picofitoplanctonica alla composizione delle comunità.

Il picofitoplancton include tutti gli organismi pigmentati di taglia inferiore ai 2 μm , quali i cianobatteri, le proclorofite e i piccoli eucarioti (1, 2).

Negli ambienti oligotrofici il picofitoplancton domina alla base delle catene alimentari, essendo responsabile del 50-90% della produzione primaria, pertanto prevale il microbial loop che rappresenta così una forza biologica che guida la distribuzione spazio-temporale degli elementi biogeni fondamentali (3).

Al contrario negli ambienti eutrofizzati prevale la componente nano- (di dimensioni comprese tra 2 e 20 μm) e micro- (di dimensioni comprese tra 20 e 200 μm) autotrofica rappresentata da nanoflagellati, dinoflagellati e diatomee. Quest'ultima componente è utilizzata dagli organismi dello zooplancton e quindi, in queste condizioni, domina la classica catena alimentare (3).

I fattori fondamentali che condizionano la produzione primaria negli ambienti acquatici sono l'irradiazione luminosa, il rimescolamento delle acque, la concentrazione e dinamica dei nutrienti e il *grazing* (4). Rivestono una notevole importanza anche fattori quali temperatura e salinità, condizioni meteomarine, *life cycle* delle diverse specie e la velocità di sedimentazione degli individui (5).

L'influenza di tali fattori fa sì che la distribuzione delle specie fitoplanctoniche e la loro produttività sia frammentata nel tempo e nello spazio. La variabilità temporale è dovuta prevalentemente alla successione delle stagioni e ai ritmi nictemerali. Infatti la produzione fotosintetica è generalmente maggiore prima di mezzogiorno, con fitoplancton rappresentato da individui giovani, in via di crescita mentre nel pomeriggio le alghe sono a crescita lenta e senescenti.

Il fitoplancton è distribuito in modo variabile anche nello spazio: si rinvencono *patch*, zone di aggregazione del fitoplancton alternate a zone in cui esso è praticamente assente. La variabilità spaziale è legata a processi idrodinamici quali la circolazione di Langmuir e la formazione di fronti (acque con diverse caratteristiche chimico-fisiche che si mescolano tra di loro), ma rivestono il loro ruolo anche processi riproduttivi e trofici (pascolo).

Il campionamento del fitoplancton deve tenere in considerazione questi aspetti, quindi, oltre all'analisi della comunità fitoplanctonica devono essere monitorati i fattori ambientali condizionanti la dinamica della popolazione studiata.

Il campionamento del fitoplancton (6) può essere effettuato mediante diversi strumenti e metodologie, che prevedono sia la raccolta dell'acqua tal quale sia la filtrazione di grandi volumi d'acqua.

Il primo metodo si effettua prevalentemente per mezzo delle bottiglie tipo Niskin o Van Dorn e delle pompe, il secondo con i retini.

La bottiglia viene solitamente utilizzata per un'analisi di tipo quantitativo e l'aliquota da raccogliere varia in relazione al livello di trofia dell'ecosistema studiato, ma solitamente si preferisce raccogliere un litro di campione, visto che il conteggio delle cellule algali prevede altri due sub-campionamenti.

Il retino invece viene utilizzato per una stima qualitativa della comunità fitoplanctonica. Le reti sono fabbricate con tessuti di nylon, perlon o altri materiali sintetici. La dimensione della maglia è in relazione alle dimensioni degli individui da campionare. La rete può essere trainata verticalmente, orizzontalmente od obliquamente con una velocità ottimale di traino inferiore a 2 nodi (1 nodo = $0,514 \text{ m s}^{-1}$).

Una volta raccolti i campioni devono essere immediatamente fissati e le modalità di trattamento variano a seconda delle classi dimensionali che si vogliono studiare:

– **Picofitoplancton**

- Il campione appena prelevato viene conservato in provette sterili contenenti formalina tamponata prefiltrata (0,2 mm), ad una concentrazione finale del 2%.
- I campioni (50-100 ml) vengono conservati in frigo al buio.
- Il conteggio, realizzato mediante microscopio a fluorescenza, si realizza previa filtrazione su filtri di policarbonato (rispettivamente da 0,2 e 0,45 μm di diametro dei pori)?

– **Nano e microfitoplancton**

- I campioni vengono fissati in bottiglie di vetro scuro per mezzo di una soluzione di Lugol al 5% o formalina al 2%.
- Il conteggio viene effettuato in laboratorio dopo sedimentazione in camera Utermöhl con microscopio rovesciato.

La fissazione del nano e microfitoplancton prevede l'uso di:

- **Formalina al 40%:**

- La formalina si neutralizza con carbonato di sodio (aggiunto in eccesso) o esametiltetramina (100 g per dm^3 di soluzione), e viene filtrata su carta da filtro prima dell'uso.
- Si aggiungono 20-40 cm^3 per dm^3 di campione.
- La formalina va versata nella bottiglia preferibilmente prima del campionamento, allo scopo di rendere istantanea la fissazione del campione.
- Le bottiglie vanno riempite fino all'orlo.
- I campioni vanno mantenuti in un luogo fresco e scarsamente illuminato e agitati il meno possibile.

I vantaggi si possono così schematizzare:

- La formalina neutralizzata rende i diversi gruppi fitoplanctonici ben riconoscibili (eccetto alcuni dinoflagellati nudi).
- I campioni fissati con la formaldeide possono essere conservati per anni, senza particolari attenzioni, se conservati correttamente.

Invece, gli svantaggi sono:

- La fissazione con la formaldeide è alquanto selettiva in quanto distorce la forma cellulare delle specie 'nude' e provoca la perdita dei flagelli nei fitoflagellati.
- Il contenuto cellulare perde il suo colore, rendendo praticamente impossibile distinguere le forme autotrofe da quelle eterotrofe.
- *Lugol*
 - Si aggiunge una quantità tale (0,4-0,8 ml per 200 ml di campione) da rendere il campione leggermente colorato.

Questo tipo di fissativo si usa solo per i campioni d'acqua e non per le retinate. Tale metodo ha il vantaggio di conservare sia i flagelli dei fitoflagellati sia gli altri organismi.

Gli svantaggi possono essere così sintetizzati:

- La soluzione acida del Lugol può dissolvere i coccoliti e dopo un certo tempo, anche i frustuli silicei.
- Molti organismi risultano troppo colorati e, come nel caso dei dinoflagellati, le piastre della teca non sono evidenziate.

Ecologia del fitoplancton potenzialmente tossico

In tutto il mondo le tossine fitoplanctoniche rappresentano un problema serio e ricorrente associato con i *bloom* di specie tossiche microalgali. Tali *bloom* consistono di specie tossiche che contaminano alimenti marini interessando la salute pubblica.

Così è importante stabilire adeguati programmi di sorveglianza e di controllo della qualità dei prodotti eduli marini. *Paralytic Shellfish Poisoning* (PSP), costituite da un numero di derivati delle saxitossine, sono prodotte dai dinoflagellati appartenenti al genere *Alexandrium*.

La diatomea *Pseudonitzschia* spp., un componente comune del fitoplancton nei mari, produce l'acido domoico (DA), un aminoacido neurotossico responsabile di intossicazione del tipo *Amnesic Shellfish Poisoning* (ASP).

Altre specie tossiche (es. *Dinophysis* spp.) associate alle *Diarrheic Shellfish Poisoning* (DSP), sono ugualmente implicate in problemi di salute. Specie di dinoflagellati appartenenti al genere *Dinophysis* sono state identificate nelle aree costiere del mare Adriatico dove sono condotte attività di monitoraggio algale. Ricorrenti osservazioni di contaminazione del tipo DSP sono state osservate negli ultimi anni nei molluschi delle acque costiere del mare Adriatico.

Negli ambienti di acqua dolce, la tossicità algale è dovuta ad alcuni taxa di cianobatteri.

Dopo una prima segnalazione di tossicità di *Nodularia spumigena*, l'elenco delle specie capaci di produrre ceppi tossici si è progressivamente allungato ed è destinato ad ampliarsi ulteriormente (8).

Le tossine prodotte dalle alghe tossiche sono accumulate dagli animali filtratori e da altri *grazer* e poi raggiungono l'uomo e gli altri animali attraverso la rete alimentare. Sebbene la natura chimica delle tossine sia molto differente, esse generalmente non cambiano né si riducono significativamente dopo la cottura degli alimenti; inoltre queste tossine solitamente non influenzano il gusto dell'alimento. Sfortunatamente la detenzione di alimenti contaminati non è facile da rilevare, e né i pescatori né i consumatori possono solitamente determinare se i prodotti eduli sono idonei al consumo umano.

Nonostante la disponibilità di conoscenze sulle tossine e sulle specie potenzialmente produttrici di tossine, molto poco si sa circa i meccanismi che controllano la dinamica delle

single specie tossiche e le condizioni idrografiche ed ecologiche che provocano le fioriture delle specie algali tossiche.

Nel monitoraggio delle specie fitoplanctoniche potenzialmente tossiche è necessario porre in relazione la dinamica spazio-temporale di tali specie con le caratteristiche fisiche, chimiche e biologiche della colonna d'acqua. Per esempio, la turbolenza delle acque mediante il trasporto di nutrienti e il trasferimento del fitoplancton attraverso un gradiente luminoso, può avere un effetto significativo sulla crescita e il declino di una popolazione potenzialmente tossica (9).

Inoltre, il ruolo dei nutrienti sugli eventi tossici è ancora oggi in discussione, infatti, solo in alcuni casi i *bloom* tossici appaiono correlati all'eutrofizzazione degli ecosistemi costieri. L'antropizzazione degli ambienti costieri può determinare un'alterazione del ciclo dei nutrienti, che possono stimolare l'impatto delle specie potenzialmente tossiche in modo differente. Le specie tossiche possono aumentare di numero insieme alle altre specie fitoplanctoniche, rappresentando sempre in termini percentuali, la stessa frazione della densità fitoplanctonica totale. Una situazione simile è stata riscontrata lungo le coste del mare Adriatico meridionale dove, durante il periodo invernale, si è osservato un incremento della densità fitoplanctonica e contemporaneamente anche un picco di numerose specie appartenenti al genere *Dinophysis*. (10). In altri casi, invece l'incremento dei nutrienti può favorire una specie potenzialmente tossica rispetto ad altre non produttrici di biotossine.

Il ruolo dei nutrienti è stato investigato anche per ciò che riguarda la produzione delle tossine per specie di dinoflagellati e diatomee. Per esempio, c'è un'ampia variabilità, non solo nell'ambito della produzione delle tossine, ma anche nell'ambito di una stessa specie e persino di un clone.

Il *Working Group on Harmful Algal Bloom Dynamics* (WGHABD) dell'ICES-IOC (*International Council for the Exploration of the Sea – Intergovernmental Oceanographic Commission*) è stato fondato nel 1994 sulla base del *Study Group on the Dynamics of Algal Blooms*. Una pianificazione della ricerca scientifica e del coordinamento per lo sviluppo delle capacità internazionali per la determinazione, predizione e mitigazione di *bloom* algali tossici è stata incoraggiata dallo IOC e dallo SCOR (*Scientific Committee on Oceanic Research*) alla fine del 1998 (11).

Attualmente in Italia esistono solo ricerche individuali su questo argomento.

Bibliografia

1. Li WKW, Dickie PM, Irwin BD, Wood AM. Biomass of bacteria, cyanobacteria, prochlorophytes and photosynthetic eukaryotes in the Sargasso Sea. *Deep Sea Res* 1992;39(3-4):501-19.
2. Sherr EB, Sherr BF. Bacterivory and herbivory: key roles of phagotrophic protists in pelagic food webs. *Microb Ecol* 1994; 28: 223-35.
3. Fenchel T. Marine plankton food chains. *Ann Rev Ecol Syst* 1988;19:19-38.
4. Barnes RSK, Hughes RN. Il sistema planctonico nelle acque superficiali. In: *Introduzione alla ecologia marina*. Padova: Piccin Nuova Libreria; 1990. p. 41-71.
5. Della Croce N, Cattaneo Vietti R, Danovaro R. Ecologia e protezione dell'ambiente marino costiero. Torino: UTET Libreria; 1997.
6. Maugeri TL, Acosta Pomar L, Bruni V. Picoplancton. In: Innamorati M, *et al.* (Ed.). Metodi nell'ecologia del plancton marino. *Nova Thalassia* 1990;11:199-205.
7. Zingone A, Honsell G, Marino D, Montresor M, Social G. Fitoplancton. In: Innamorati M, *et al.* (Ed.). Metodi nell'ecologia del plancton marino. *Nova Thalassia* 1990;11:183-98.

8. Sechi N. Le alghe tossiche d'acqua dolce: identificazione e classificazione. In: Bruno M (Ed.). *Le alghe tossiche marine e d'acqua dolce: impatto sanitario e strategie di controllo*. Roma: Istituto Superiore di Sanità; 2000. (Rapporti ISTISAN 00/31). p. 43-6.
9. Smayda TJ. Turbulence, watermass stratification and harmful algal blooms: an alternative view and frontal zones as 'pelagic seed banks'. *Harmful Algae* 2002;1:95-112.
10. Caroppo C, Congestri R, Bruno M. Dynamics of *Dinophysis* sensu lato species (Dinophyceae) in a coastal Mediterranean environment (Adriatic Sea). *Continental Shelf Research* 2001;21(16-17):1839-54.
11. Glibert PM, Pither G (Ed.). GEOHAB Global Ecology and Oceanography of Harmful Algal Blooms. Science plan. Baltimore and Paris: SCOR and IOC; 2001.

METODI PER L'ESTRAZIONE E LA RILEVAZIONE DELLE PRINCIPALI TOSSINE ALGALI

Milena Bruno

Laboratorio di Igiene Ambientale, Istituto Superiore di Sanità, Roma

I sistemi di rilevazione delle tossine algali comprendono:

- *Analisi strumentale*
come la cromatografia liquida ad alta pressione (*High Pressure Liquid Chromatography*, HPLC) accoppiata alla spettrometria di massa, la gas-cromatografia (*Gas Chromatography*, GC) accoppiata alla spettrometria di massa, la cromatografia ad alta pressione su strato sottile (*High-Performance Thin Layer Chromatography*, HPTLC);
- *Metodi enzimatici*
come il biosaggio delle proteinfosfatasi e l'immunosaggio ELISA.
- *Biosaggi*
su organismi animali o batterici.

Analisi strumentale

Le analisi strumentali suggerite dai diversi laboratori di ricerca chimica che in tutto il mondo hanno elaborato metodi di analisi per queste sostanze, sono spesso legate al patrimonio culturale del peculiare gruppo di ricerca. Si possono trovare così protocolli di estrazione decisamente complessi, a fronte di sistemi più semplici ma ugualmente efficienti.

Di seguito vengono riportate le varie procedure di analisi strumentale applicate alle diverse tossine.

Saxitossine

Le saxitossine, famiglia di tossine che comprende una quarantina di composti, sono state analizzate in HPLC (1) con un metodo che permette la separazione della saxitossina, della neo-saxitossina, di tutte le gonyautossine dalla 1 alla 6, e delle C-tossine 1 e 2. Il metodo prevede l'uso di due fasi mobili, la prima (A) composta di 1,5 mM di acido sulfonico di esano e 1,5 mM di acido sulfonico di eptano in 1,5 mM di tampone ammonio fosfato a pH 6,7; e la seconda (B) composta al 25% di acetonitrile, di acidi sulfonici di esano ed eptano come sopra, il tutto in 6,25 mM di tampone ammonio fosfato a pH 7. Viene predisposto un gradiente da (A) a (B) per 19 minuti attraverso una colonna Hamilton PRP-1 di 15 cm per mm 4,1 di spessore: al termine tutte le tossine sono risolte.

Oshima nel 1993 (2) predispose un metodo che comprendeva una corsa isocratica di tre fasi mobili:

1. tetrabutylammonio fosfato 1 mM portato a pH 5 con acido acetico;
2. sodio 1-eptansulfonato 2 mM in ammonio fosfato 10 mM a pH 7,2,
3. tampone eptansulfonato come sopra in proporzione 9:1 con acetonitrile.

La corsa veniva svolta attraverso una colonna C-8 (ad esempio, quella scelta dall'autore era una Develosil C8-5 da 250 mm per 4,6 mm, della Nomura Chemical). Le tossine venivano eluite nel seguente ordine:

- C-tossine da 1 a 4;
- Gonyautossine da 1 a 6; dcgonyautossine 2 e 3;
- Saxitossina, neo-saxitossina, desaxitossina.

Entrambi i metodi utilizzano una procedura sviluppata da Bates e Rapoport nel 1975 (3) modificata, che comprende un'ossidazione basica con periodato per produrre derivati fluorescenti delle tossine, rilevati per emissioni a 390-400 nm dopo eccitazione a 330-340 nm.

A queste procedure può essere eventualmente fatta seguire una identificazione delle tossine in spettrometria di massa FAB (*Fast Atom Bombardment*), per ulteriore sicurezza di identificazione.

Anatossina-a

Per l'anatossina-a sono stati proposti i seguenti due metodi per analisi GC:

- *Primo metodo*

Il primo metodo prevede l'estrazione di 5 mg di cellule liofilizzate sciolte in soluzione fisiologica, mediante la sonicazione in ghiaccio.

Successivamente, va effettuata la iniezione di 2 µl (*splitless* per 0,6 minuti) in un gascromatografo tipo Varian modello 3500, dotato di autocampionatore. L'iniezione viene effettuata divisa in due colonne capillari di silicio di 30 m ciascuna, di 0,25 mm di diametro e 0,25 µm di spessore di film (nell'esempio vengono utilizzate due colonne DB-5 e DB-1701 della J e W Scientific). Le due colonne sono fuse da un connettore Glasseal Y e monitorate su un detector a cattura di elettroni separati. Le temperature di iniezione e dei detector sono rispettivamente di 250 °C e di 300 °C; il flusso di idrogeno va mantenuto ad 1,5 ml/min misurato ai detector; il flusso di azoto va mantenuto a 25 ml/min.

La temperatura delle rampe è la seguente:

- temperatura iniziale della colonna: 100 °C;
- tempo di attesa iniziale: 0 minuti.

Il metodo prevede tre fasi:

1. temperatura della colonna: 200 °C; tasso di rampa: 20 °C/min; tempo di attesa: 0 min;
2. temperatura della colonna: 226 °C; tasso di rampa: 2.0 °C/min; tempo di attesa: 0 min;
3. temperatura della colonna: 265 °C; tasso di rampa : 20 C°/min; tempo di attesa 3 min.

- *Secondo metodo*

Il secondo metodo prevede che gli estratti vengano dissolti in 1-2 ml di acido cloridrico 1 mM, diluiti con 20 ml di acqua e alcalinizzati con idrossido di sodio 1 N fino a pH 11. A questo punto vengono effettuate 4 estrazioni successive, ciascuna con 50 ml di cloroformio. La fase organica viene filtrata su una colonna da 25 cm di lunghezza per 1,5 cm di diametro, impaccata con 15 grammi di solfato di sodio anidro, e raccolta in flask da 300 ml. Il lavaggio finale della colonna viene eseguito con circa 25 ml di cloroformio.

La fase organica viene quindi ridotta a 2-3 ml in un rotavapor a temperatura di 40 °C, ed evaporata a secco sotto blanda corrente di azoto.

L'estratto viene dissolto in 2 ml di acetone. Vengono aggiunti 60 µl di idrossido di potassio al 30% in acetone e 100 µl di pentafluorobenzilbromuro all'1% in acetone, e la soluzione viene riscaldata per 3 ore ad 80 °C. Dopo il raffreddamento a temperatura ambiente, il solvente viene attentamente evaporato e l'estratto dissolto in 1-10 ml di isottano, per l'analisi in gas-massa (4).

Microcistine e nodularine

Anche per le microcistine e le nodularine vengono qui descritti due metodi, che permettono un limite di rilevazione fino ad 1 mg/l.

Il primo prevede l'analisi in HPLC in fase inversa con Photo-Diode Array Detector a 240 nm. La fase mobile impiegata è composta da acetonitrile al 27-30% in tampone fosfato 0,1 M a pH 7, con flusso di 1 ml/minuto; la colonna è tipo Brownlee ODS 5 µm sferi-analitica, da 220 mm di lunghezza per 4,7 mm di diametro. Il secondo metodo prevede l'uso di una fase mobile al 12% di acetonitrile in tampone potassio di idrogeno fosfato 0,1 M a pH 6,8 con flusso di 1 ml/min e lettura a 238 nm, 0,005 AUFS e volume campionato di 50 µl. La colonna è tipo Regis ISRP GFF S5-80, oppure tipo C18 RP Lichrochart. Questo sistema permette la risoluzione entro 5 minuti delle tossine -RR; -YR ed -LR (5).

Tossine DSP

Per l'analisi del gruppo delle tossine DSP (*Diarrhetic Shellfish Poisoning*), diverse procedure di estrazione vengono suggerite da Marr *et al.* (6), a seconda che si operi su campioni da plancton o da molluschi.

Nel primo caso campioni da 200 ml di plancton vengono raffreddati in ghiaccio e sonicati per 2 minuti, quindi estratti 2 volte con un ugual volume di cloroformio. Il cloroformio viene quindi ridotto a 4 ml in rotovapor, e sottoposto a lavaggio con ugual volume di acqua, poi trasferito in una flask volumetrica da 5 ml ed essiccato sotto corrente di azoto. Il residuo viene disciolto in 5 ml di dimetilformamide, e 250 µl di questo soluto vengono derivatizzati.

In caso di esame su tessuti di molluschi, vengono omogeneizzati 2 g di epatopancreas in 8 ml di metanolo all'80% in acqua per 2 minuti, utilizzando una provetta in plastica da centrifuga da 50 ml.

L'omogenato si centrifuga per 10 minuti a 4.000 rpm e 5 ml del surnatante risultato, trasferiti in una provetta di vetro, ed estratti 2 volte con 5 ml di esano miscelando su vortex per 30 secondi.

Dopo aver scartato gli strati di esano, si aggiunge 1 ml di acqua allo strato residuo di metanolo acquoso, che viene poi estratto 2 volte con 6 ml di cloroformio. I due estratti in cloroformio vengono quindi riuniti ed essiccati sotto corrente di azoto, e il residuo disciolto in 1 ml di dimetilformamide. Una aliquota da 25 a 100 µl del soluto viene quindi derivatizzata.

La derivatizzazione richiede che 100 µl di reagente cumarinico (soluzione di 4-bromo-metil-6,7-dimetossicumarina 10 mM in dimetilformamide) e 10 ml di N,N-diisopropilettilammina 950 mM in acetonitrile vengano aggiunti a 10 µl di campione estratto; il tutto viene lasciato reagire a 45 °C per 2 ore.

La soluzione viene quindi fatta evaporare sotto corrente di azoto, il residuo disciolto in 1 ml di esano/cloroformio 1:1 e passato su una cartuccia SPE con 500 mg di silice, condizionata precedentemente con 5 ml di cloroformio più 5 ml di esano/cloroformio 1:1. Dopo alcuni lavaggi con il secondo composto e poi con il primo, le tossine vengono eluite con 5 ml di metanolo al 5% in cloroformio, quindi evaporate e ridissolte in 50 µl di acetonitrile.

L'estratto derivatizzato viene passato su colonna LC ODS2 o RP18 LC da 250 mm di lunghezza per 4 mm di diametro interno mantenuta a 40°C, con un gradiente lineare di eluizione da 60 a 100% di metanolo con 0,1% di acido trifluoroacetico e velocità di flusso di 1 ml/minuto. L'eluizione delle tossine avviene in 15 minuti, con il detector a 346 nm di eccitazione e 430 nm di emissione.

Metodi enzimatici

I diversi metodi elaborati comprendono il saggio enzimatico per la rilevazione di attività anti-proteinfosfatasica per le tossine DSP e per le microcistine; l'immunodosaggio ELISA per le microcistine e un saggio ELISA a sandwich per triplo anticorpo contro la proteina S.I.P. indotta dalla saxitossina per la rilevazione di quest'ultima.

Saxitossine

Anticorpi monoclonali sono stati preparati da splenociti di topo immunizzato contro la proteina (S.I.P.) indotta dalla saxitossina, tossina della famiglia PSP, nei granchi *Hemigrapsus oregonensis* e nei molluschi che vengono in contatto con la saxitossina. Il sistema è stato articolato in un saggio ELISA a sandwich per triplo anticorpo contro la S.I.P. Questi anticorpi consentono di discriminare i prodotti ittici contaminati da saxitossina, e hanno il limite di rilevazione di S.I.P. indotta da una dose di tossina superiore a 5 ng.

Un metodo di estrazione da noi elaborato permette di disporre di estratti di PSP in acqua distillata, utilizzabili contemporaneamente sia nei saggi enzimatici e immunologici che nelle analisi strumentali e nei biosaggi. Il metodo prevede la sospensione mediante vortex di 10 mg di alghe in peso fresco in 1 ml di acqua sterile bidistillata. La sospensione viene riscaldata per 5 minuti a 70°C, quindi sonicata per altrettanti minuti a 30-40 °C, con una risospensione sul vortex per 20 secondi, dopo 2 minuti e mezzo di sonicazione. La sospensione viene quindi centrifugata per 10 minuti ad 11.000 rpm, il surnatante raccolto, il pellet risospeso in 1 ml di acqua sterile bidistillata e l'intera procedura replicata, fino all'unione dei due surnatanti ricavati. A questo punto si può procedere ai vari test, o conservare l'estratto a -30 °C. Poiché ripetuti scongelamenti per il prelievo di aliquote causano la distruzione delle tossine estratte, è consigliabile stoccare l'estratto in diverse quote (es. in eppendorf), e prelevare una quota intera quando serve. Una variante in macro di questo metodo parte da 200 mg di alghe in peso fresco da sospendere in 10 ml di H₂O sterile bidistillata, con la centrifugazione per 20 minuti ad 8.000 rpm. Per il resto il metodo è sovrapponibile.

Microcistine

Il saggio enzimatico per la rilevazione dell'attività anti-proteinfosfatasica utilizzato per le saxitossine viene usato anche per la rilevazione delle microcistine, data la identità di bersaglio di queste tossine con il gruppo DSP. Nell'ambito delle microcistine, il saggio permette la rilevazione di -LR, la più diffusa, fino al limite di 0,1 µg/l in acqua.

Metodi immunologici, come l'immunodosaggio ELISA, sono stati messi a punto e commercializzati per le microcistine, raggiungendo soglie di sensibilità fino a 0,25-0,5 µg di tossina per g di alghe liofilizzate, e fino a 6 ng/g di acqua. Il limite di questo saggio è la possibilità di rilevare solo le quattro microcistine principali (-LR; -RR; -YR e -LA) e in blocco,

senza capacità di discriminare tra le stesse. Il metodo di estrazione è quello elaborato per le saxitossine con l'esclusione del passaggio di riscaldamento a 70 °C.

Tossine DSP

Le tossine DSP possono essere rilevate per mezzo di saggi enzimatici per la rilevazione di attività anti-proteinfosfatasi, la specifica azione patogena che le distingue. Un sistema messo a punto per consentire la rilevazione da tessuti di pesci, molluschi, spugne, oppure da fitoplancton o acqua potabile, comprende l'esame dell'estratto metanolico del campione mediante il saggio delle proteinfosfatasi, successivamente mediante cromatografia liquida seguita da saggio proteinfosfatasi, quindi su elettroforesi capillare seguita nuovamente da saggio proteinfosfatasi, e infine in spettrometria di massa. L'estratto che dopo il primo saggio delle proteinfosfatasi abbia dato risultato negativo, viene estratto con un solvente organico, ed esaminato in cromatografia su colonna Sephadex LH-20, poi in cromatografia liquida SPRP a due passaggi, rispettivamente a pH 6,5 e pH 2. A questo va fatto seguire l'esame in elettroforesi capillare, in spettrometria di massa, eventualmente in NMR (*Nuclear Magnetic Resonance*) e per analisi degli aminoacidi.

Il metodo di estrazione è quello elaborato per la saxitossine con la variante di utilizzare metanolo al 90% (invece dell'acqua bidistillata), e la diluizione 1:1 con acqua sterile bidistillata dei due surnatanti riuniti, in modo da avere una concentrazione finale di metanolo al 45%. Questa consente, con specifica perizia, di utilizzare l'estratto anche per il biosaggio su MICROTOX, sistema che non tollera l'uso di solventi puri.

Biosaggi

Tra i numerosi biosaggi esistenti per le tossine algali, vengono annoverati quelli su vari insetti (moscerini, ecc.), crostacei (gamberetti, copepodi, cladoceri), pesci, roditori (coniglio, ratto, topo), maiali e batteri. Tra questi ultimi, recentemente il più praticato è certamente quello con il *Vibrio fischeri*.

La via seguita nell'elaborazione di questo test è particolarmente interessante.

Dopo gli anni quaranta le implicazioni tossiche dovute all'introduzione quotidiana nell'ambiente di miriadi di composti chimici estranei al ciclo naturale divennero evidenti, e crebbe l'esigenza di utilizzare sistemi di monitoraggio rapidi ed economici. Una delle numerose linee di ricerca in tal senso partì dall'acquisizione della elevata permeabilità ai tossici da parte delle membrane cellulari dei batteri gram-negativi.

Dal 1960 al 1990 vennero sviluppati studi su questi batteri miranti ad accrescere la permeabilità cellulare della membrana esterna a composti idrofili, idrofobi e surfactanti, con l'obiettivo di selezionare specie batteriche per test di ecotossicità.

I criteri con cui si scelse di operare furono la semplicità di mantenimento e conservazione, la sensibilità alle sostanze, la rapidità di esecuzione e l'economicità di costo; altro problema da affrontare fu quello della quantificazione degli effetti tossici prodotti sui batteri. Dal 1970 al 1980 l'attenzione dei laboratori si concentrò sul *Photobacterium phosphoreum*, poi rinominato *Vibrio fischeri*. Quest'ultimo è un batterio marino di profondità, dotato di bioluminescenza. La bioluminescenza batterica è legata direttamente alla respirazione cellulare, di cui è un effetto collaterale. Ci si accorse rapidamente che ogni inibizione dell'attività cellulare, mediante ad es. una sostanza tossica, si traduceva in una diminuzione del tasso di respirazione, con un

corrispondente calo del tasso di luminosità: il calo era proporzionale all'intensità tossica della sostanza con cui venivano in contatto i batteri.

Il meccanismo fisiologico che determina la bioluminescenza nel *Vibrio fischeri* utilizza l'ossidazione del NADH a NAD, data dall'azione di vari enzimi e coenzimi (tra cui il NAO) su substrati di base, aminoacidi e minerali presenti nel citoplasma. In questo modo si libera energia che riduce il flavinmononucleotide, il quale insieme ad un atomo di ossigeno ossida l'enzima luciferasi, che riducendosi emette un quanto luminoso a 490 nm di emissione (luce blu-verde).

Il ceppo di *Vibrio fischeri* NRRL B-11177 fu scelto come quello dotato di più alta sensibilità ad una vasta gamma di composti chimici della più varia origine.

Queste acquisizioni portarono per la prima volta alla realizzazione, da parte della ditta americana Microbics Inc., di uno strumento cyborg, il MICROTOX, in cui l'attività di un organismo vivente veniva misurata e trasformata in un coefficiente tossicologico: l'EC 50 (*Effective Concentration resulting in a 50% reduction in light emission*).

Il test si basa su cinque pozzetti con un milione di batteri ciascuno, che ricevono 4 diluizioni scalari della sostanza sospetta, più il bianco in cui è diluita. I batteri vengono lasciati in contatto 5 minuti (tossicità acuta), 15 minuti (tossicità cronica) o più. La luminosità prima e dopo del contatto viene misurata da due luminometri presenti nello strumento, e quantificata secondo una scala interna. La variazione di luminosità (I_t) rispetto a prima del contatto (I_o) viene composta con quest'ultima nella formula:

$$\frac{I_o - 1}{I_t} = \gamma(\text{gamma})$$

I valori del gamma vengono processati da un software statistico, per costruire una curva dose-risposta su scala bilogaritmica con calcolo dei limiti di confidenza.

Da questa si giunge all'EC 50, cioè alla concentrazione efficace del tossico alla temperatura dell'esperimento e per la durata dell'esposizione. Lo strumento esibisce trenta pozzetti da utilizzare in coppie di file per tre test contemporanei; la misura dell'intensità luminosa rilevata dai luminometri viene espressa sul display a diodi dello strumento; a parte c'è un ulteriore pozzetto a 7 °C per il mantenimento dei batteri da utilizzare nei test. Il sistema fornisce operativamente vari vantaggi rispetto agli altri biosaggi: i batteri sono conservati liofilizzati; i reagenti sono forniti pronti dalla casa costruttrice; lo strumento è corredato da un software di facile gestione, ma in grado di valutare le caratteristiche dei risultati e di dare consigli sull'impostazione degli esperimenti; il tutto ha un ingombro estremamente contenuto, tale da poter essere tranquillamente ospitato su di un bancone.

La sensibilità del sistema è stata attualmente vagliata per più di 1.200 diverse sostanze chimiche, ed esistono più di 500 pubblicazioni scientifiche su studi che lo utilizzano.

I campi di intervento del test comprendono l'inquinamento dell'aria, le tossine biologiche, i biomateriali, le correlazioni tra sensibilità tossiche, la compilazione di database su sostanze tossiche, l'acqua potabile, l'ecotossicologia, gli effluenti industriali, quelli municipali, la genotossicità, i rifiuti rischiosi, la petrolchimica, i reflui di cartiere, la regolazione degli effluenti, la regolazione dei rifiuti rischiosi, i sedimenti, i sistemi a membrane semipermeabili, gli inquinanti tossici urbani nell'acqua di precipitazione, la variabilità o la riduzione della tossicità.

Attualmente il biosaggio con *Vibrio fischeri* è stato approvato dalla normativa internazionale ISO/FDIS 11348 parte 1, 2 e 3 ed è approvato come test ufficiale in Italia (DL 152/1999 poi 258/2000); il sistema MICROTOX è approvato come test ufficiale in USA (ASTM Standard D-5660) e nello Stato dell'Alberta, in Canada; il corrispondente sistema LUMISTOX è approvato come test ufficiale in Germania.

A tutt'oggi le ditte produttrici di sistemi biosaggio con *Vibrio fischeri* sono la AZUR Environmental (USA, MICROTOX), la MERCK Eurolab (Germania, TOXALERT), la Scelvo (Italia, Test Scelvo), la Cyttox (USA, CYTOX TEST) e la Lange (Germania, LUMISTOX). A parte gli apparecchi da laboratorio, esistono dei biosaggi con *Vibrio fischeri* portatili, il DELTATOX, della AZUR Environmental, e il TOXALERT portatile della MERCK Eurolab. Il secondo fornisce la risposta solo in percentuale (%) di sostanza tossica che causa la riduzione del 50% della luminosità batterica, mentre il primo, come il suo omologo di laboratorio, fornisce la risposta anche in quantità (peso) della sostanza tossica necessaria.

Entrambi gli strumenti rendono possibile l'esecuzione del saggio su campo, con disponibilità immediata del risultato; possono inoltre immagazzinare i dati per scaricarli in un secondo momento ed elaborarli con un software statistico; in conclusione sono versatili come i modelli da laboratorio e nello stesso tempo sono leggeri e maneggevoli.

Vari biosaggi sono stati confrontati con il MICROTOX per testarne la comparabilità: tra questi quelli con *Pimephales promelas* (fathead minnow), *Salmo gairdneri*, *Salmo trutta*, *Poecilia reticulata* (guppy), *Lepomis macrochirus*, *Daphnia magna*, *Selenastrum capricornutum*, *Tetrahymena piriformis*, *Pseudomonas fluorescens*, *Ceriodaphnia dubia*, *Vibrio ginghaiensis* e anche il test di irritazione endoculare nel coniglio. Con tutti è stato trovato un ottimo accordo, e per alcuni dei primi è nota la correlazione riportata in Tabella 1.

Tabella 1. Correlazione tra i risultati di biosaggi che utilizzano organismi diversi

Organismo	Coefficiente di correlazione (r)
<i>Pimephales promelas</i>	1,00
<i>Salmo gairdneri</i>	0,89
<i>Salmo trutta</i>	0,92
<i>Poecilia reticulata</i>	0,89
<i>Lepomis macrochirus</i>	0,77
	Percentuale di accordo
<i>Daphnia magna</i>	86%
<i>Selenastrum capricornutum</i>	90%

Numerose tossine algali sono state testate con successo sul MICROTOX, che posta la limitazione della aspecificità del test (non è infatti possibile identificare le singole sostanze utilizzando un biosaggio), ha evidenziato una sensibilità pienamente paragonabile alle rilevazioni in HPLC. Finora sono state testate le seguenti tossine: microcistine (EC 50 limite: 59 µg/ml), saxitossina (EC 50 limite: 1,28 µg/ml), neo-saxitossina, nodularina, anatossina-a (EC 50 limite: 10 µg/ml), acido ocaidaico, acido domoico, ciguatossina, maitotossina, pectentossina, brevetossine, fibrocapsina e dinofisittossine. Sotto questo aspetto alcuni biosaggi rapidi sono stati confrontati con *Vibrio fischeri* (sistema LUMISTOX della Lange) per quanto riguarda le tossine delle Cianofeece: questo, a parità di sensibilità, è risultato il più rapido (5, 10 o 15 minuti) rispetto ad *Artemia salina* (22-24 ore) e a *Pseudomonas putida* (16 ore), con la complicazione che l'estrazione di tossine in fase solida con uso di solventi ed evaporazione a secco dà luogo ad interferenze e falsi positivi nei test, mentre il metodo sopra illustrato, messo a punto nel nostro laboratorio, permette di evitare completamente questi inconvenienti.

Bibliografia

1. Sullivan JJ, Wekell MM. The application of high performance liquid chromatography in a paralytic shellfish poisoning monitoring program. In: Kramer DF, Liston J (Ed.). *Seafood quality determination*. Anchorage: University of Alaska; 1987. p. 357-71.
2. Oshima Y, Itakura H, Lee KC, Yasumoto T, Blackburn S, Hallegraef G. Toxin production by the dinoflagellate *Gymnodinium catenatum*. In: Smayda TJ, Shimizu Y (Ed.). *Toxic phytoplankton blooms in the sea*. Amsterdam: Elsevier Science Publishers B.V.; 1993. p. 907-912.
3. Bates H A, Rapoport H.A chemical assay for saxitoxin, the paralytic shellfish poison. *J Agric Food Chem* 1975;23:237-9.
4. Bruno M, Barbini Attard D, Pierdominici E, Serse A, Ioppolo A. Anatoxin-a and a previously unknown toxin in *Anabaena planctonica* from blooms found in lake Mulargia (Italy). *Toxicon* 1994;32(3):369-73.
5. Meriluoto JAO, Eriksson JE. Rapid analysis of peptide toxins in cyanobacteria. *Journal of Chromatography* 1988;(438):93-9.
6. Marr JC, McDowell L, Quillian MA. Investigations of derivatization reagents for the analysis of diarrhetic shellfish poisoning toxins by liquid chromatography with fluorescence detection. *Natural Toxins* 1994;2:302-11.

*La riproduzione parziale o totale dei Rapporti e Congressi ISTISAN
deve essere preventivamente autorizzata.*

*Stampato da Ditta Grafiche Chicca & C. snc
Via di Villa Braschi 143, 00019 Tivoli (Roma)*

Roma, dicembre 2003 (n. 4) 4° Suppl.