Coastal waters monitoring data: frequency distributions of the principal water quality variables

Franco GIOVANARDI*, Maria Grazia FINOIA, Simone RUSSO, Marina AMORI and Bianca DI LORENZO

ICRAM – Central Institute for Marine Research, via di Casalotti, 300, 00166 Roma, Italy
*e-mail corresponding author: f.giovanardi@icram.org

ABSTRACT

Examining the results of the Italian national programme of marine coastal monitoring, the old problem has arisen about the choice of the most appropriate procedures and methods to validate data and screen preliminary data. Therefore, statistical distributions of water quality parameters have been taken into consideration, in order to assign appropriate frequency distributions to all the routinely measured variables. Each sample distribution has been analysed and defined by a probability density function (p.d.f.), by means of a powerful method of data analysis (Johnson 1949) that allows for the computation of statistical parameters of a wide variety of non-normal distributions. The resulting Johnson distributions are then classified depending on four fundamental categories of frequency distributions: normal, log-normal, bounded and unbounded. Theoretical aspects of the method are explained and discussed in an adequate way, so as to allow for practical applications. The shape and nature of these curves require further consideration, in order to understand the behaviour of water quality variables and to make comparison among different coastal zones. To this end, two coastal systems were considered in this work: the Emilia-Romagna coastal area of the NW Adriatic Sea and the Tuscany littoral of the Northern Tyrrhenian Sea. There are notable advantages to the adopted approach. First it offers the possibility to overcome severe constraints requested by the normality assumption, and avoids the troublesome search for the most appropriate transformation function (i.e. for ensuring normality). Second, it avoids searching for other kinds of theoretical distributions that are appropriate for the data. In our approach, the density functions are opportunely integrated, in such a way that, for whatever value assumed by a given variable, the corresponding expected percentage point value under the respective frequency curve, can be calculated, and vice versa. We believe that the Johnson method, although tested with coastal monitoring data, can be usefully adopted whenever we have to analyse environmental data and try to understand how an aquatic system works (e.g. large lakes). In the Appendix specific details about the Johnson classification criterion are reported and highlight the case of bimodal distributions. Finally, an example of data analysis is provided, by using the R (V. 2.11) software, with both graphical and numerical outputs.

Key words: Coastal waters, monitoring data processing, frequency distributions, Johnson's distributions, normal law approximation, coastal eutrophication, TRIX Index.

1. INTRODUCTION

In general, the goal of the monitoring data analysis is to study the behaviour of single variables and then evaluate the functional interrelationships among them, e.g. by means of correlation and regression. This represents the so-called "empirical statistical" approach, that has been essential in the study of Eutrophication.

Richard Vollenweider is well known for his fundamental contributions to developing these procedures. The OECD global programme [OECD International Programme on Eutrophication, Vollenweider & Kerekes (Eds), 1982] covered 150 lakes. Preliminary statistical analysis sought to characterize each of the most representative parameters taken into consideration (in number of 40, including chemical, physical, biological, lake morphological and functional factors) by means of their geometric mean. The maximum and minimum value was used to define the range of variation.

In the OECD report, the utilisation of geometric means instead of the arithmetic means, was because of the very practical necessity of adjusting average values to make them truly representative of variables that, for their own nature, vary in a non-linear way (i.e. exponential). The OECD criterion that classifies the trophic status of lakes ("Open boundary system" Cfr. the cited OECD report), is also based on geometric means. Group means and standard deviations for each selected parameter are computed following log-transformation, "which was found to be necessary for normalisation". (Note: geometric means in fact may be expressed as the antilog of arithmetic means of log-transformed data.). The decimal log-transformation was therefore considered realistic and more than adequate for the comprehensive scopes of the OECD programme.

The "normalising" transformations aim at stabilising variances. When this objective is achieved, it is possible to utilise "the powerful arsenal offered by parametric statistics". A general transformation that is specifically aimed at homogenizing variances is represented by the Taylor Power Law (Taylor 1961; Legendre & Legendre 1984). This transformation function is derived from the relationship between variances and related means. The data, transformed following this rule, are likely to com-

ply with the conditions requested by parametric statistics (i.e. the independence of the means and the homogeneity of variances).

Another empirical method involves evaluating the optimum transformation function, based on the value of the parameter λ that maximises the Likelihood Function (Box & Cox 1964; Sokal & Rohlf 1995). This procedure has been widely used (Giovanardi & Vollenweider 2004), with the aim of testing the goodness of fit for log-decimal transformations that have been applied to the original components of TRIX Index. This Index (Vollenweider *et al.* 1998), and the related trophic scale, is currently used in Italy and adopted by UNEP-MEDPOL for the trophic classification of coastal waters (UNEP 2003).

Other authors (Ignatiades *et al.* 1992; Heyman *et al.* 1984.), have faced the problem of rough data distributions in a rigorous way, by comparing the observed data frequencies with the expected theoretical frequencies, based on some type of distribution with a known *p.d.f.* In such cases, the most appropriate procedure is the χ^2 Goodness-of-Fit test, usually reported in statistical handbooks and available in several statistics packages for computers. If we accept the null hypothesis (i.e. not significant differences at an assigned probability), we can conclude that the sample values have been drawn from a distribution that has the same *p.d.f.* of the theoretical distribution that best fits our data.

On the other hand, "if sample distributions can be fitted to a theoretical distribution, probability for occurrence of, e.g. chlorophyll-*a* concentration above a certain level could be calculated." The already cited paper of Heyman *et al.* (1984), referred to a data-set of 950 samples related to 25 Swedish lakes and focused on the effects of improved waste water treatment interventions and sewage diversion on water bodies. The results demonstrated that the frequency distributions, which were similar for a great number of water quality variables in different lakes, belonged to the log-normal, beta and gamma distribution families. For several parameters of interest it was then possible to provide estimates on the maximum value attainable for a given lake, starting from average values, with a big enough approximation to plan for correct policies of control and sanitation, and to verify the effects of such policies during the time period.

This is a "theoretical statistical"approach, because it is based on the assumption that a given variable is distributed following density functions of theoretical distributions chosen a priori. It has been successfully utilised e.g. in Pluviometry and Hydrology. By applying the Extreme Values Theory and the related Gumbel distribution to the time-series of the maximum values recorded annually for rain data and/or flow rate of a river, it is possible to describe exceptional events, by assigned "times of return" (occurrence), expressed as number of years (e.g. the centennial flood of a river, the rainstorm of the century, and so on; cfr. e.g. Maione & Moisello 1974). The study of the eutrophication phenomenon that affected the NW Adriatic Sea in the 1980s, followed the same methodological approach, and evaluated exceptional chlorophyll peaks, with the related probability of occurrence (Giovanardi & Tromellini 1992).

Often, there are no known theoretical distributions that fit our data, even though they are properly transformed. Hypotheses and significance tests currently in use (the already cited χ^2 Goodness-of-Fit test, the Kolmogorov-Smirnov test, etc.), are quite sensitive to non-normal conditions. If heterogeneous conditions remain, with a strong residual dependence of the error variance on the means, despite the most appropriate transformations of the rough data and the removal of outliers, one must decide whether to insist on finding other theoretical distributions or, instead, adopting robust statistics (distribution free) procedures.

Nevertheless, the type of distribution, the shape of the frequency curves and their nature are essential attributes of the systems. Consequently, we went to the root of the problem, and focused our attention directly on statistical distribution theory, and related mathematical treatments, with the awareness that making normal approximations of empirical data distributions has "…played a dominant role in both theoretical and applied statistics since the time of Laplace".

Johnson (1949) described a system of frequency curves that represents transformations of the standard normal curve. By applying these transformations to a standard normal variable, a wide variety of non-normal distributions can be approximated, including distributions which are bounded on either one or both sides. Once a particular Johnson curve has been fitted, the normal integral can be used to compute the expected percentage points under the respective curve.

The main goal of the present paper is, therefore, to apply the above approach to the analytical data and measurements retrieved from the coastal monitoring activity. We believe that this approach can be usefully adopted whenever we seek to understand how aquatic systems work, including lakes. Furthermore, the application of the method can be extended to matrices other than seawater or lake water such as sediment and biota. Finally, we emphasize the great opportunity revealed by defining predictive models and scenarios, and the undoubted advantages offered by the Johnson method for management purposes.

2. MATERIAL AND METHODS

The procedures developed were tested using the data of the Italian National Monitoring Programme, and in particular a time series from 1997 to 1999, which has been considered and presented in a previous paper published in this journal (Giovanardi & Vollenweider 2004). The coastal systems taken into consideration are:
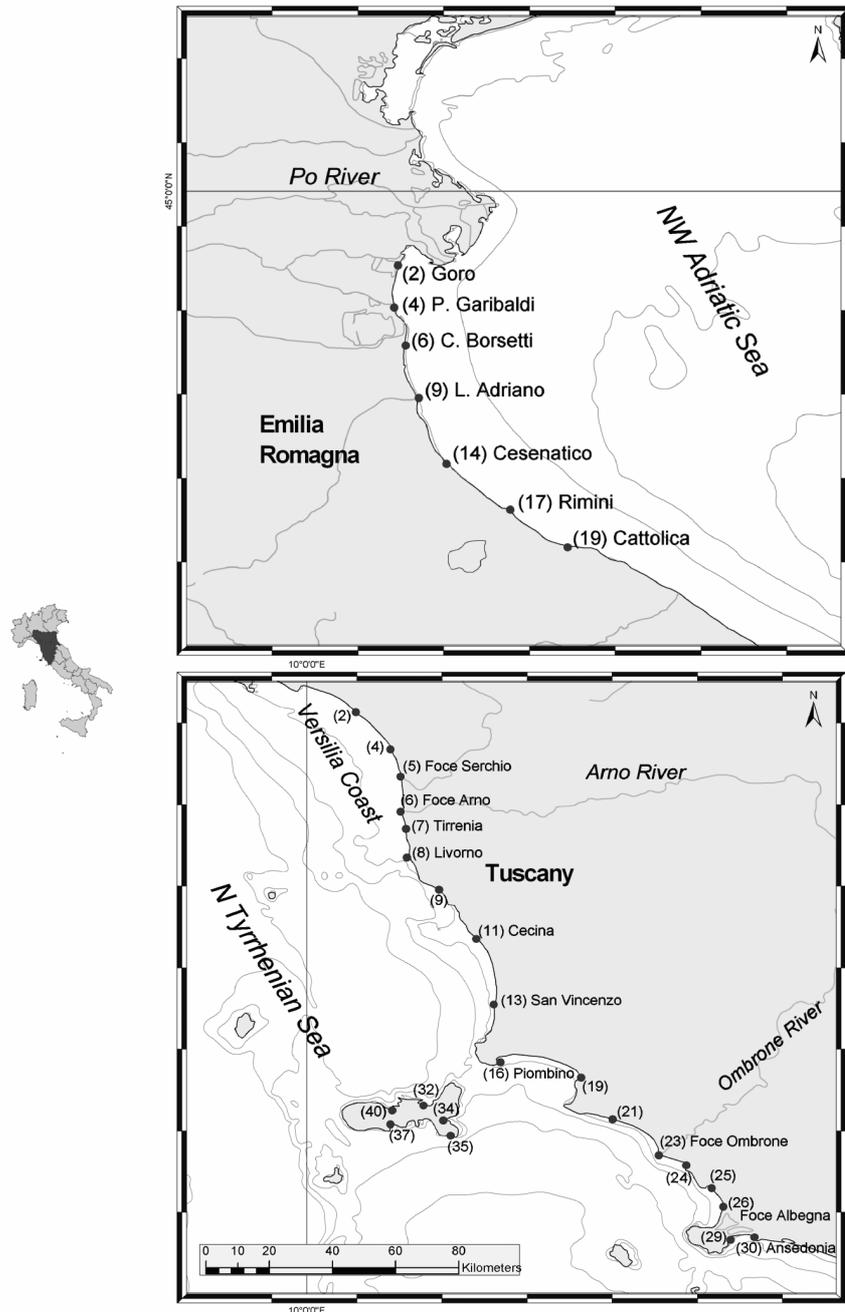
Fig. 1. Emilia Romagna and Tuscany coastal waters. Transects location.
(Italian Ministry of the Environment – SDM. & ICRAM. National monitoring programme)

1) the Emilia-Romagna Adriatic coastal belt (around 600 records of data), and 2) the Tuscany coast (N Tyrrhenian Sea - *ca* 1200 records). Monitoring campaigns were carried out along transects perpendicular to the coast, covering a belt of about 3000 m from on-shore to off-shore. Figure 1 depicts the geographical location of the transects. Periodical measurements of physical-chemical parameters along the water column, surface values of chlorophyll-*a*, nutrient concentrations and phytoplankton biomass, were taken.

Tropho-dynamic structure of those coastal systems is strongly influenced by freshwater inputs coming from two of the major river basins of the Italian peninsula. However, the systems have different trophic responses: Tyrrhenian coastal waters, although affected by the nutrient rich Arno River discharge, present generally low productivity, while the coastal waters of the NW Adriatic Sea have much higher risk of Eutrophication, due to the nutrient loads discharged by the Po River. The bottom waters of the NW Adriatic Sea exhibit fre-

quent anoxic episodes during summertime. These different behaviours have been interpreted, respectively, in terms of low and high "efficiency" of nutrient utilisation (Giovanardi & Vollenweider 2004).

In addition, we performed temporal comparisons over a longer period, by processing over twenty years series of analytical data related to the NW Adriatic Sea, collected by the Environment Agency of the Emilia Romagna Government (ARPA-ER: Daphne Oceanographic Structure), as part of its regional control and monitoring programmes.

Our results would not have been reached without the R (V. 2.11) software. R is a free software easily downloadable via the Internet, (we have followed, with pleasure, the guidelines provided by the Core Group of the R foundation regarding the citation of R). We analysed sample distributions using the translation method of Johnson (1949). In order to fit frequency curves, we largely utilised R package SuppDist (Bob Wheeler - 2005. SuppDists: Supplementary distributions. R package version 1.0-13. http://www.bobwheeler.com/stat.), which allows the computation of the following functions: the density function [*dJohnson()*], the distribution function [*pJohnson()*], its inverse [*qJohnson()*], the random number generator [*rJohnson()*] and the summary function *[sJohnson()]*, and gives a list of the whole statistical parameters of the distributions. The command *JohnsonFit()* generates a list containing the Johnson parameters (γ, δ, ξ, λ and *type*), where type is one of the four types of Johnson curves: S_N (normal), S_L (log-normal), S_B (bounded) or S_U (unbounded). The calculation algorithms for the estimate of the Johnson parameters are those of Wheeler (1980: quantiles method), and Hill, Hill & Holder (1976: moments method).

3. THE JOHNSON SYSTEM OF FREQUENCY CURVES

The work of Johnson (1949), starts from the general function of transformation to the normal law, of a continuous random variable *x*, whose distribution is not known. This function, represented in a formal way as: $z = f(x)$, is then re-defined by introducing 4 new parameters, in the following way:

$$Z = \gamma + \delta \cdot g\left(\frac{x - \xi}{\lambda}\right) \quad (1$$

where *Z* is the normal standard variate, distributed with mean = 0 and STD = 1.

By setting:

$$y = (x - \xi) / \lambda, \quad (2$$

the equation (1) becomes simpler:

$$Z = \gamma + \delta g(y).$$

The probability density function (*p.d.f.*) of the y variable, can therefore be written as:

$$p(y) = \delta\, g'(y) p(Z) = \frac{\delta}{\sqrt{(2\pi)}} g'(y) exp\left(-1/2\,[\gamma + \delta\, g(y)]^2\right) \quad (3$$

From the linear relationship:

$$x = \xi + \lambda\ y, \quad (4$$

we observe that the distribution of *x* will have the same shape of *y*. By definition, the STD of *x* will be λ times the STD of *y*, and the variations of ξ determine a simple translation that will only affect the expected values of the *x* distribution.

Accordingly, the λ and ξ parameters in (1) are simple scale and location factors, respectively, while the γ and δ parameters will determine the shape of the distributions, which is given, in general, by the equation (3).

The process provided by the *R SuppDist* package for fitting a Johnson distribution on sample data utilises the procedure of quantiles (Wheleer 1980), and requires, first, a choice of translation function and, second, an estimate of the values to be assigned to the four parameters γ, δ, ξ, λ. They may also be estimated from the moments. Applied Statistics algorithm 99 (Hill *et al.* 1976), has been translated into C for this implementation and is currently available in the *R SuppDist*. We should add that the appropriate translation function *g()* is uniquely identified by the third and fourth order moments of the sample distribution of *x*, i.e. skewness and kurtosis.

The Johnson system classifies sample distributions according to four families of frequency curves that can be plotted in the plan identified by β_1 (i.e.: skewness to the square) and by β_2 (i.e. kurtosis +3). Diagrams in figure 2 illustrate an example of classification. The simplest case is normal distributions (S_N) identified by the point $\beta_1 = 0$ and $\beta_2 = 3$.

Log-normal distributions lie exactly on the S_L line. This line separates the domain of the unbounded curves (S_U = System Unbounded) from that of the bounded curves (S_B = System Bounded). Inside this last sector, the lower boundary line is drawn, beyond which S_B curves become bimodal. Finally, the impossible area extends beyond the straight line labelled as: $\beta_2 - \beta_1 - 1 = 0$

The example reported in figure 2 is related to sample quality variables distributions of the marine coastal waters of the Adriatic Sea, close to the Po River mouth: sampling stations of Goro and Porto Garibaldi. More details of the procedures used for computing, and plot Johnson diagrams are provided in the Appendix.

Table 1 was generated using the *sJohnson ()* function. Each variable is identified by the whole statistical parameters of its own distribution. The processed dataset, with a sample size of 180 records, is the same as the representation in figure 2.

We then calculated, for each of the water quality variables, the exact translation function (1) and, using proper adjustments, the optimum transformation to normality:
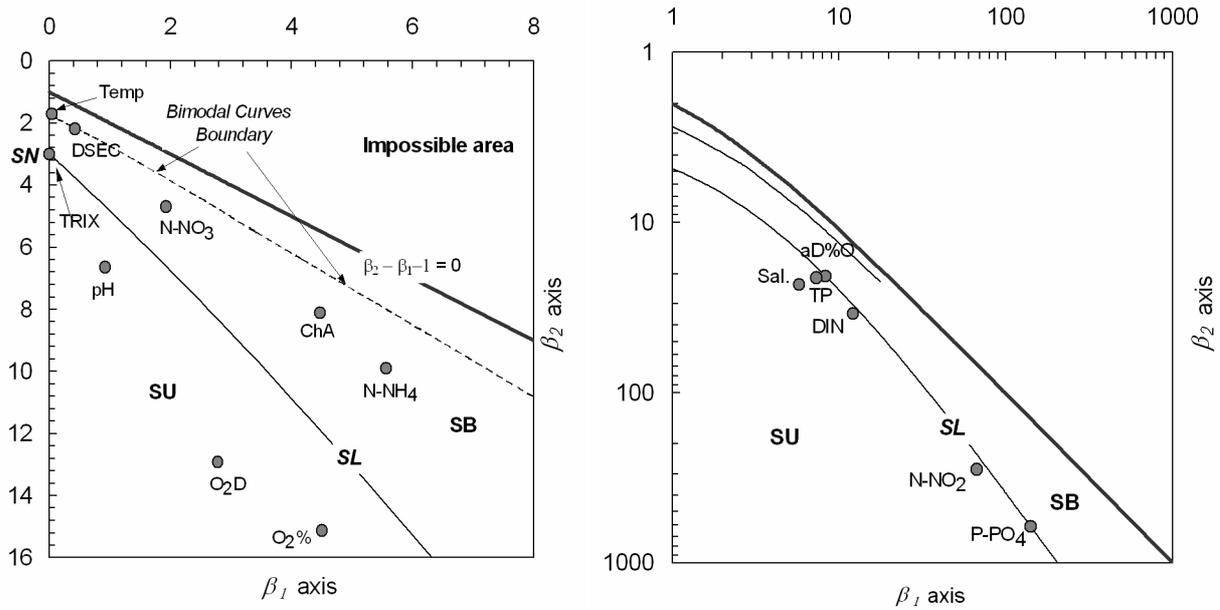
Fig. 2. The Johnson scheme of frequency curves classification, depending on moments β_1 and β_2.

Tab. 1. NW Adriatic coastal waters. Goro and Porto Garibaldi transects. (From on-shore to 3 km off-shore. Sample size =180). List of the variables processed, type of distribution and related Johnson's parameters. (*): Mean minus mode div. STD. (**): Oxygen as absolute deviation from 100% saturation. (Parameter requested for the TRIX Index formulation).

Variables	type	Johnson's parameters				Distribution parameters							
		γ	δ	ξ	λ	mean	median	mode	variance	STD	Pearson's Schewness (*)	Schewness	Kurtosis
Temp.	°C	-0.195	0.509	5.62	21.6	17.67	18.49	26.89	44.45	6.67	-1.383	-0.213	1.701
Sal.	psu	0.652	1.199	31.06	4.55	27.39	28.47	29.80	44.88	6.70	-0.360	-2.405	23.204
DSEC	m	0.547	0.537	0.837	2.37	1.65	1.47	0.86	0.45	0.67	1.177	0.652	2.191
O_2sat	%	-1.041	1.397	84.95	17.45	103.32	99.17	93.90	497.46	22.30	0.422	-2.122	15.133
aD%O(**)	%	0	1.430	-4.39	15.0	14.77	10.61	4.81	231.34	15.21	0.654	2.883	20.713
O_2D	mg l^{-1}	-0.678	1.370	7.10	1.99	8.44	8.13	7.81	5.03	2.24	0.282	-1.668	12.921
pH	pH units	-0.751	1.741	8.18	0.28	8.33	8.31	8.28	0.05	0.21	0.247	-0.960	6.633
ChA	µg l^{-1}	2.023	0.761	1.297	124.5	15.02	9.45	3.02	235.39	15.34	0.782	2.115	8.114
N-NH$_4$	µg l^{-1}	2.381	0.815	-1.167	477.5	39.50	23.25	4.70	2110.4	45.94	0.758	2.360	9.903
N-NO$_3$	µg l^{-1}	1.583	0.857	-26.82	1194.4	193.2	135.7	26.2	34886.3	186.78	0.894	1.387	4.696
N-NO$_2$	µg l^{-1}	-0.749	0.867	5.94	3.66	12.88	9.51	7.24	223.42	14.95	0.377	-8.195	282.285
DIN	µg l^{-1}	-1.288	1.197	48.99	108.45	248.4	189.6	129.3	58063	240.96	0.494	-3.483	34.288
P-PO$_4$	µg l^{-1}	0	0.820	1.005	2.31	5.87	3.32	1.53	80.87	8.99	0.482	11.885	611.702
TP	µg l^{-1}	-1.332	1.331	14.56	11.32	32.21	27.87	22.64	370.29	19.24	0.497	-2.711	21.085
TRIX	TRIX units	0	1	5.921	0.748	5.921	5.921	5.921	0.560	0.748	0	0	3

$$(Z-\gamma)/\delta = g(y) \qquad (5$$

If the x variable is already normally distributed, by definition we will have: $g(y) = y$. In such a case, parameters γ and δ do not affect the shape of the y distribution, which will also be normal. Therefore, by setting $\gamma = 0$ and $\delta = 1$, we obtain $Z = y$ in equation (5). Accordingly, equation (2) represents the well known standardisation procedure of a variable, with $\bar{x} = \xi$ and STD of $x = \lambda$. This can be easily verified with the TRIX parameter distribution, which is the unique distribution belonging to the S_N family in the example reported in table 1.

For log-normal distributions, only the δ is important in determining the shape of the function $g()$, (Wheeler

1980). By setting $\gamma = 0$, the transformation function to achieve normality will be:

$$Z/\delta = log(y). \qquad (6$$

The different forms that the normalising function can assume, depending on the type of distribution, and the related domain of existence, are presented in table 2, which also reports characteristic parameters of the normal resulting distributions, i.e. means and STDs of the derived variables x' and y', expressed as different combinations among the four Johnson parameters.

It should always be possible to return to the original variables, even if the normalising function is complex. After a normal standard variate Z has been generated, a backwards procedure can be adopted to generate a cor-

Tab. 2. Optimum transformation functions to the Normal Law of the variate: $y = (x - \xi)\lambda$, for the 4 Johnson categories and characteristic parameters of the resulting normal distributions.

	S_N (normal)		S_L (log-normal)		S_U (unbounded)		S_B (bounded)	
$y' = g(y)$	$y' = y$		$y' = ln(y)$		$y' = ln[y+\sqrt{y^2+1}]$ i.e.: $arcsinh(y)$		$y' = ln[y/(1-y)]$	
Domain of the function	y $-\infty, +\infty$	x $-\infty, +\infty$	y $0,...+\infty$	x $\xi, +\infty$	y $-\infty, +\infty$	x $-\infty, +\infty$	y $0,...1$	x $\xi,...\xi+\lambda$
	Transfrormed variables							
	y'	x'	y'	x'	y'	x'	y'	x'
Mean	0	ξ	0	ξ	$-\gamma/\delta$	$(-\gamma/\delta)\lambda+\xi$	$-\gamma/\delta$	$(-\gamma/\delta)\lambda+\xi$
STD	1	λ	$1/\delta$	$(1/\delta)\lambda$	$1/\delta$	$(1/\delta)\lambda$	$1/\delta$	$(1/\delta)\lambda$

responding Johnson variate, by applying the inverse translation to (1):

$$x = \xi + \lambda \cdot g^{-1}\left(\frac{Z-\gamma}{\delta}\right) \quad (7$$

The appropriate $g^{-1}()$ function, for each of the four Johnson distribution families, is the inverse function of the corresponding $g()$ reported in table 2.

By setting $z = (Z - \gamma)/\delta$, for each family of distribution the related inverse functions: $g^{-1}(z)$, are reported in table 3.

Tab. 3. Table of the inverse functions: $\tilde{g}^{-1}(z) =$

	Inverse function
S_N (normal family)	z
S_L (log-normal family)	e^z
S_U (unbounded family)	$(e^z - e^{-z})/z$ i.e.: $sinh(z)$
S_B (bounded family)	$1/(1+e^{-z})$

These inverse functions permit the generation of Johnson distributions with the same characteristic parameters of the original distributions (i.e. the same frequency curves). Once the value $y = g^{-1}(z)$ has been calculated, we can derive the corresponding x value from $x = \xi + \lambda\ y$.

It remains now to discuss the procedure to be adopted for computing the *p.d.f.*

In order to compute the *p.d.f.* we can evaluate the corresponding value of the probability density, for each of the x generated with the translation function (7), using the *dJohnson ()* function of *the R SuppDist* package.

Alternatively, we can directly compute the *p.d.f.* value by re-writing the equation (3) as density function in relation to x:

$$f_x(x) = \frac{\delta}{\lambda\sqrt{(2\pi)}} g'\left(\frac{x-\xi}{\lambda}\right) \cdot exp\left(-\frac{1}{2}\left[\gamma + \delta\ g\left(\frac{x-\xi}{\lambda}\right)\right]^2\right) (8$$

In order to simplify the necessary calculations, we note that the expression between the square brackets is equal to Z as shown in equation (1), while the value of

the $g'()$ function must be computed using the derivative of each of the four $g()$ functions reported in table 2.

By setting, as usual, $y = (x - \xi)/\lambda$, the value of the derivative functions $g'(y)$ is presented in table 4.

Tab. 4. Table of the derivative functions: $g'(y) =$

	Derivative function
S_N (normal family)	1
S_L (log-normal family)	$1/y$
S_U (unbounded family)	$1/\sqrt{y^2+1}$
S_B (bounded family)	$1/[y\ (1-y)]$

Using the integration of equation (8), we may achieve a fundamental result, namely, the precise area under the frequency curve, to the left of an assigned value x. It is also possible to estimate x as the limit value for a given probability, fixed *a priori*. The first result is provided by *R SuppDist* as an output of the command *pJohnson()*, which gives the value of the distribution function depending on each of the four families of frequency curves. The second result is gained with the command *qJohnson()*, which gives the x value for an assigned quantile. By this method, we can precisely estimate the probability of occurrence of maximum (and/or minimum) values of a water quality variable, for each of the four types of Johnson distributions. Table 5 addresses this approach.

4. RESULTS

In this section, variables characterising coastal water quality are represented by their own frequency curve. The first series of figures (Fig. 3 to Fig. 8) show comparisons between the NW Adriatic and N Tyrrhenian coastal systems. Each coastal system has been divided into three sub-areas, each identified with the following codes:

- NW Adriatic Sea:
  G. – PG.: transects of Goro and Porto Garibaldi,
  CB. – LAdr.: transects of Casalborsetti and Lido Adriano.
  C. – Ctt.: transects of Cesenatico and Cattolica.

Tab. 5. Lower and upper boundary values, corresponding to 99%, 95% and 80% of the area under the curve of the related p.d.f. (Data samples and measurement units are the same as in Tab. 1). Note: the negative concentration values, when occurred for lower limits of some parameters, have been reported as zero.

|			Inverse of the distribution function for assigned values of α					
	Type	0.005	0.025	0.1	median (0.5)	0.9	0.975	0.995
Temperature	SB	5.82	6.27	7.90	18.49	26.14	26.96	27.17
Salinity	SU	0	11.23	20.11	28.47	33.57	37.07	41.93
DSEC	SB	0.84	0.86	0.91	1.47	2.73	3.05	3.15
O_2sat	SU	61.68	72.62	81.92	99.17	129.25	158.60	200
aD%O	SL	0	0	1.73	10.61	32.36	54.67	86.46
O_2D	SU	3.37	4.95	6.19	8.13	11.02	13.78	17.71
PH	SU	7.83	7.97	8.10	8.31	8.59	8.82	9.11
ChA	SB	1.59	1.96	2.89	9.45	35.43	61.01	85.26
$N-NH_4$	SB	0	1.15	4.12	23.25	97.26	177.29	266.01
$N-NO_3$	SB	0	0	13.89	135.72	466.32	699.66	882.28
$N-NO_2$	SU	0	0	3.54	9.51	24.80	47.50	90.64
DIN	SU	0	0	49.59	189.63	506.98	863.92	1416.28
$P-PO_4$	SL	1.10	1.22	1.49	3.32	12.03	26.22	54.45
TP	SU	2.37	9.02	14.99	27.87	54.07	81.19	120.85
TRIX	SN	3.993	4.454	4.962	5.921	6.880	7.388	7.848
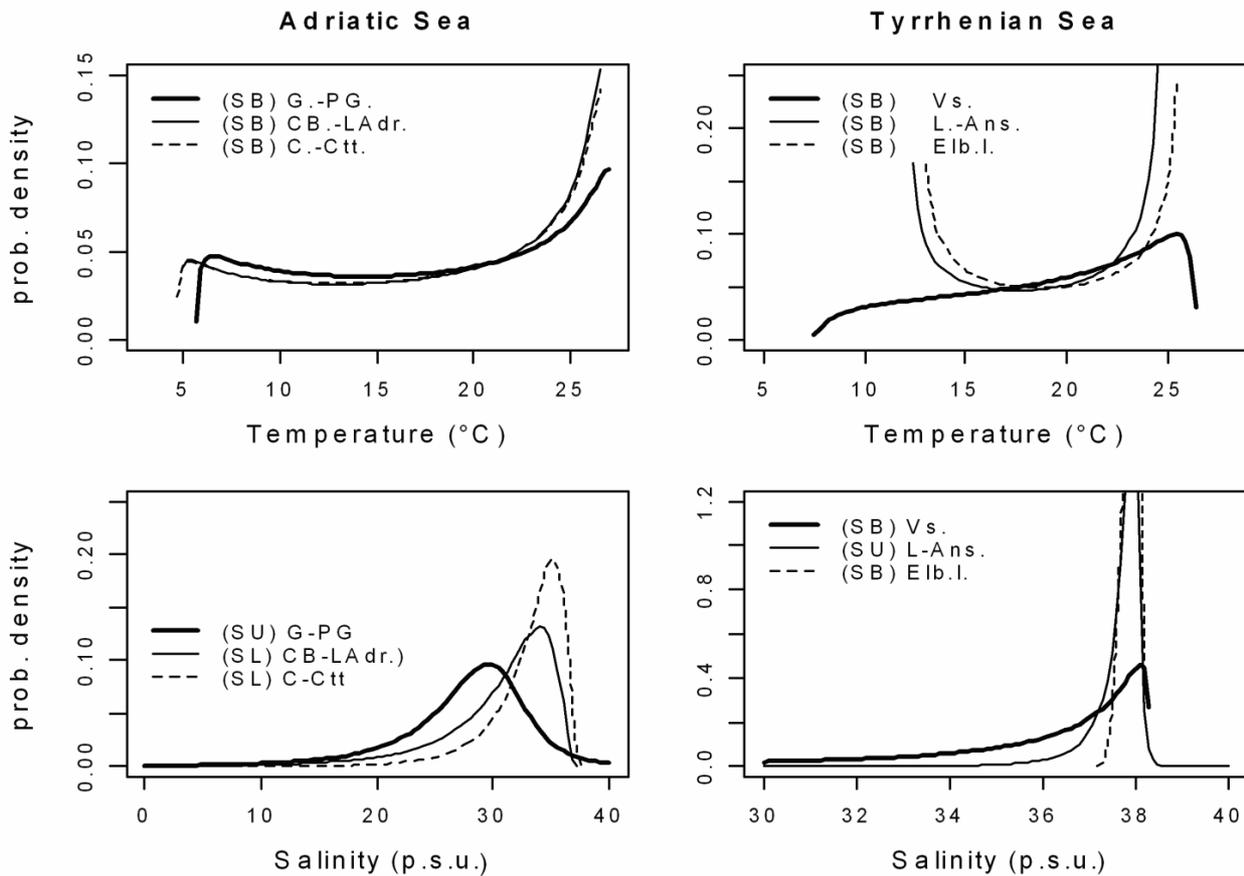
Fig. 3. Frequency curves of hydrological parameters.
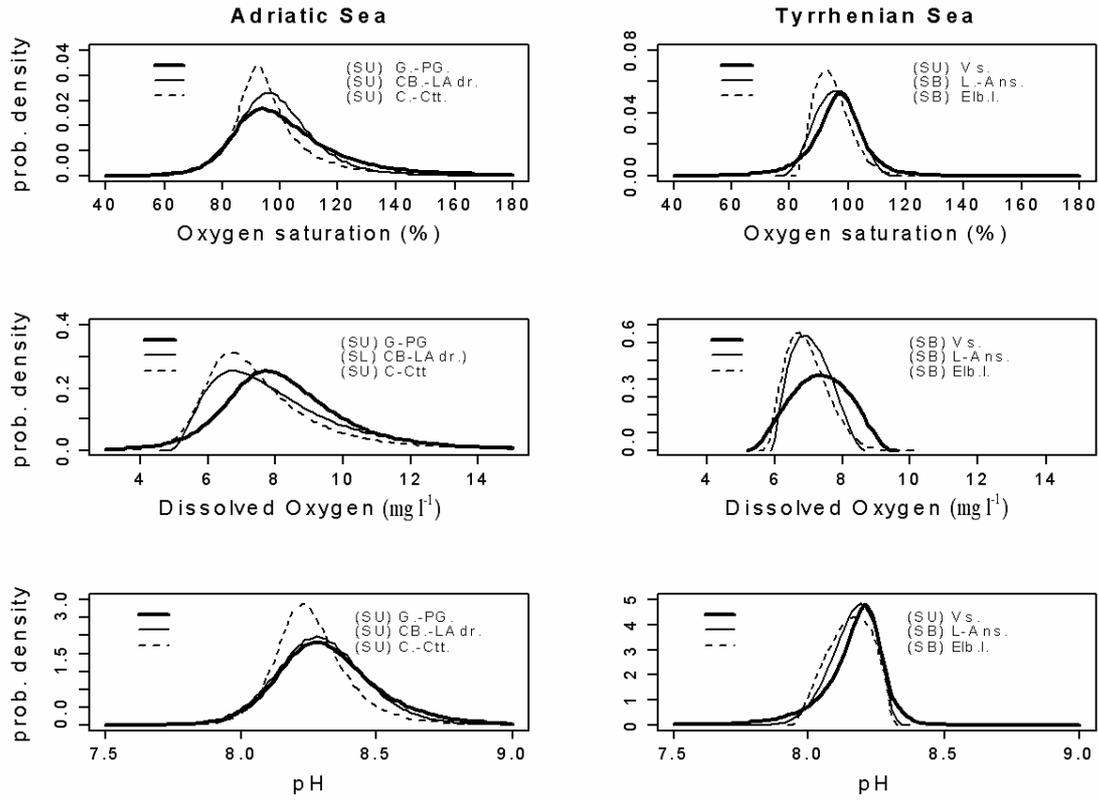
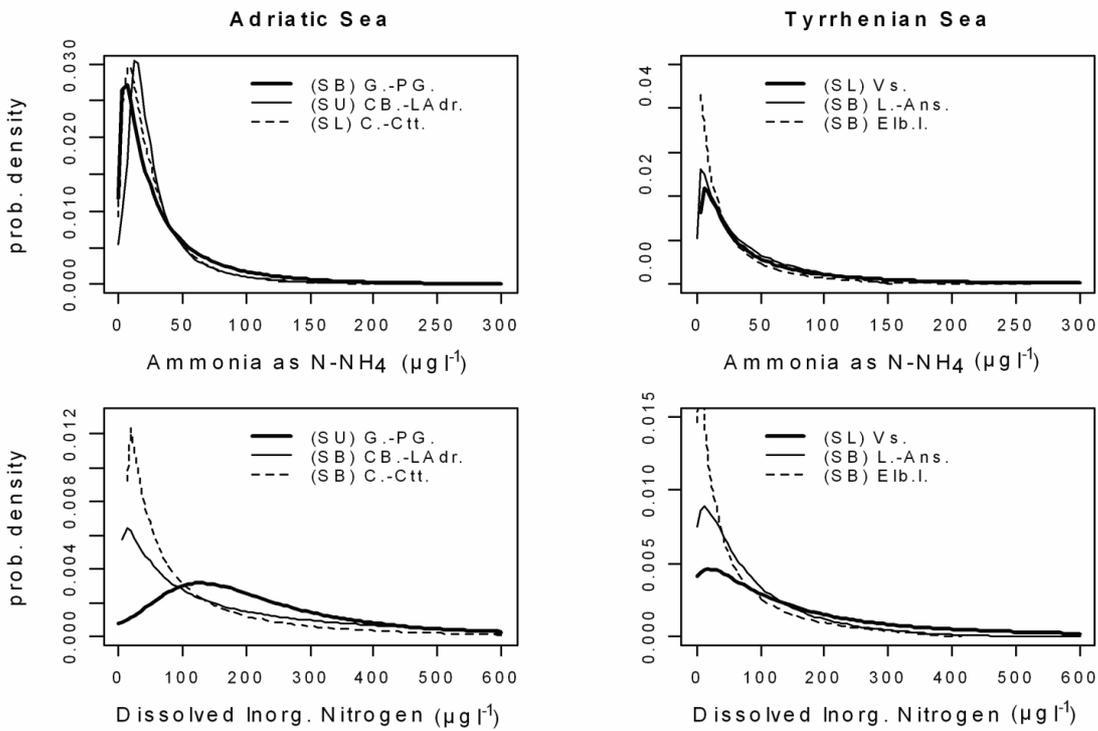**Fig. 4**. The principal chemical-physical parameters.
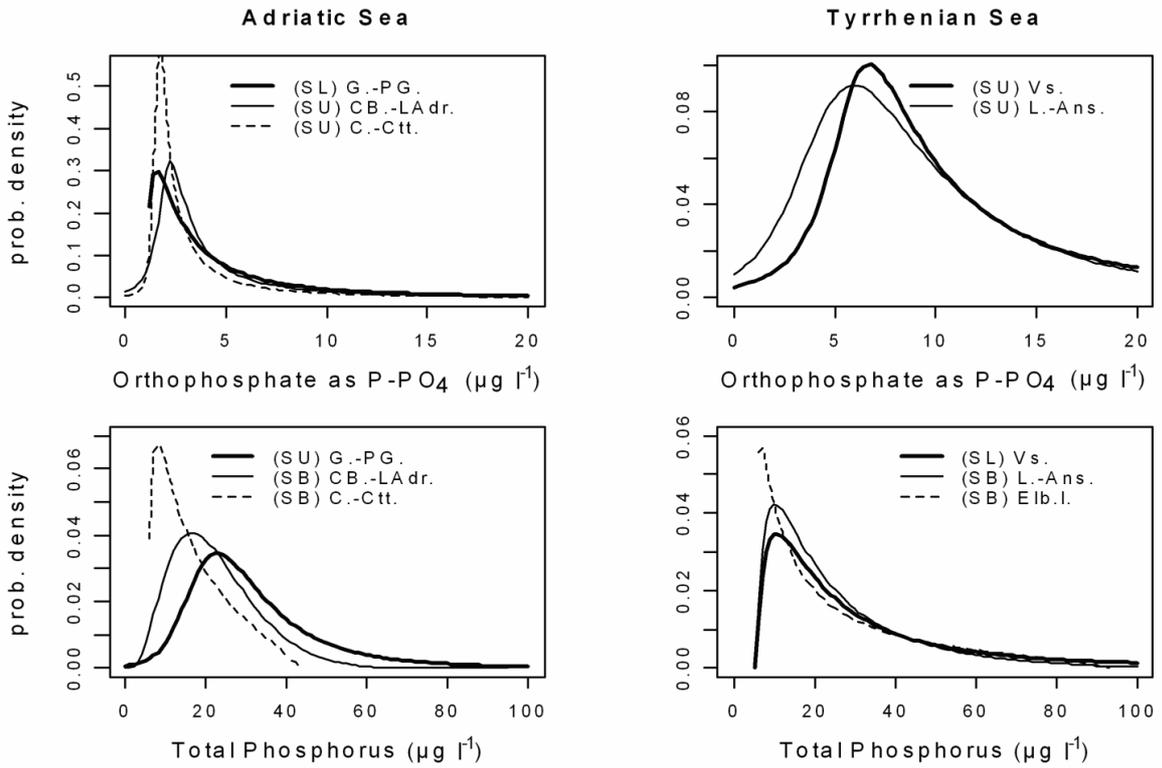

Fig. 5. Inorganic Nitrogen compounds. DIN as N-(NO_3+NO_2+NH_4).

Adriatic Sea

Tyrrhenian Sea

Fig. 6. Phosphorus as Orthophosphate and Total P.
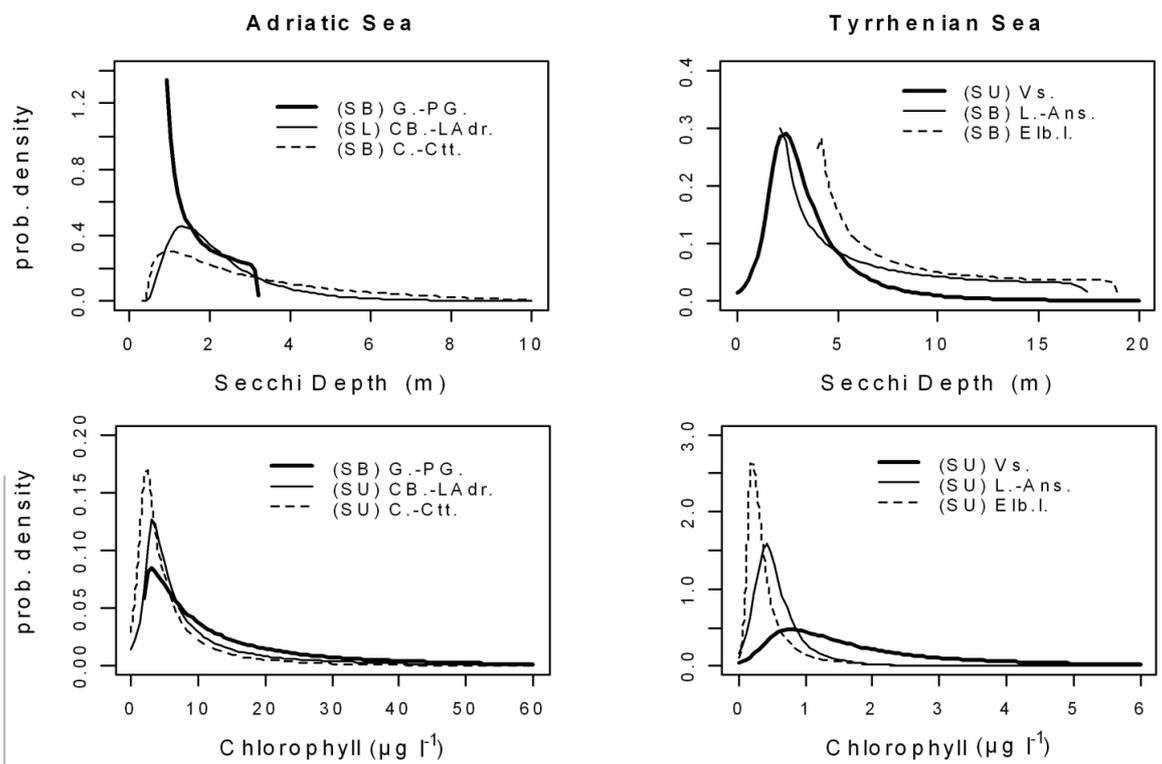
Adriatic Sea

Tyrrhenian Sea

Fig. 7. Trophic indicator parameters, as a direct expression of phytoplankton biomass.
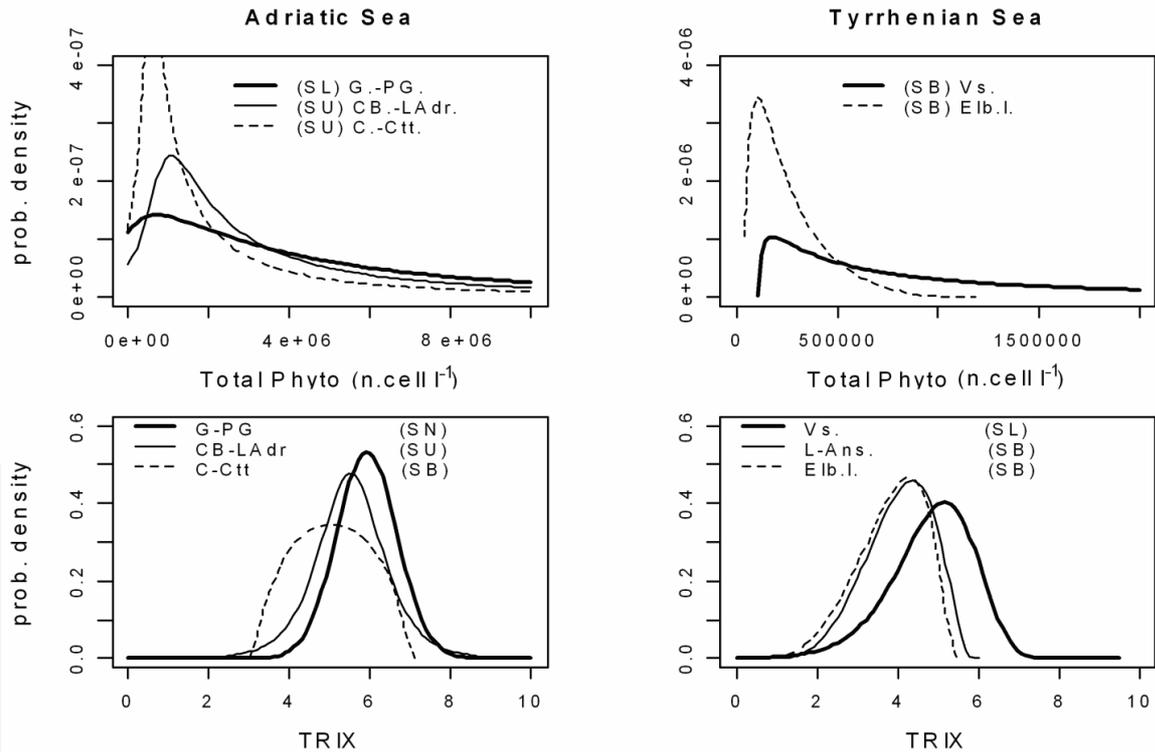
Fig. 8. Total phytoplankton biomass data and related TRIX curves.

- N Tyrrhenian Sea:
    Vs.: Transects of the Versilia littoral
    L. – Ans.: Transects from Livorno to Ansedonia.
    Elb.I.: Coastal waters of Elba Island.

Variables are grouped per typology, namely a) hydrological: temperature and salinity, b) chemical-physical: dissolved oxygen and pH, c) chemical: nutrients in their different forms, and d) variables directly related to trophic levels: Secchi depth, chlorophyll and total phytoplankton density. Lastly, TRIX Index distributions are reported as a direct measure of trophic levels.

Each frequency curve is then labelled with the corresponding acronym for classification according to the Johnson scheme.

The second series of figures (Fig. 9 to Fig. 11) provides a temporal comparison between 1990 and 2002. The data distributions are from a particular Adriatic sampling station (st. 314), located 3 km off-shore Cesenatico. Therefore, in the same diagrams, we have also represented the *p.d.f.* variations, accounting for the same variable in the two distinct periods.

Thus, by means of the above procedures (used for calculations and the successive graphical representation of the frequency curves), each water quality parameter is "certified" with its own range of variation and associated probabilities. This variation interval is characteristic to the coastal area under examination and is intended to be an intrinsic property of the related statistical distributions, fully described by:

-   central values: mean, median and mode,
-   dispersion around average values: STD, variance,
-   moments of higher order: skewness, kurtosis
-   shape of the frequency curves and their classification.

5. DISCUSSION

The emerging framework is remarkably heterogeneous, due to the high variation in curve shapes and types. This framework can help explain the behaviour of single variables and their evolution as one moves away from the major freshwater sources affecting coastal areas.

Statistical distributions of temperature (Fig. 3) were always of the S_B type. The maximum variation range, λ appears to be wider in the Adriatic coastal waters, with a difference of 22 °C between winter minimum and summer maximum value. (Note: the entire field of existence of the S_B distributions goes from ξ to $\xi + \lambda$, as shown in Tab. 2).

In the Tyrrhenian coastal system, maximum variation occurs in the areas near the mouth of the Arno River (Versilia), where the range is 19 °C. In this area, the annual temperature distribution is almost uniform, with a slight tendency to bimodality.
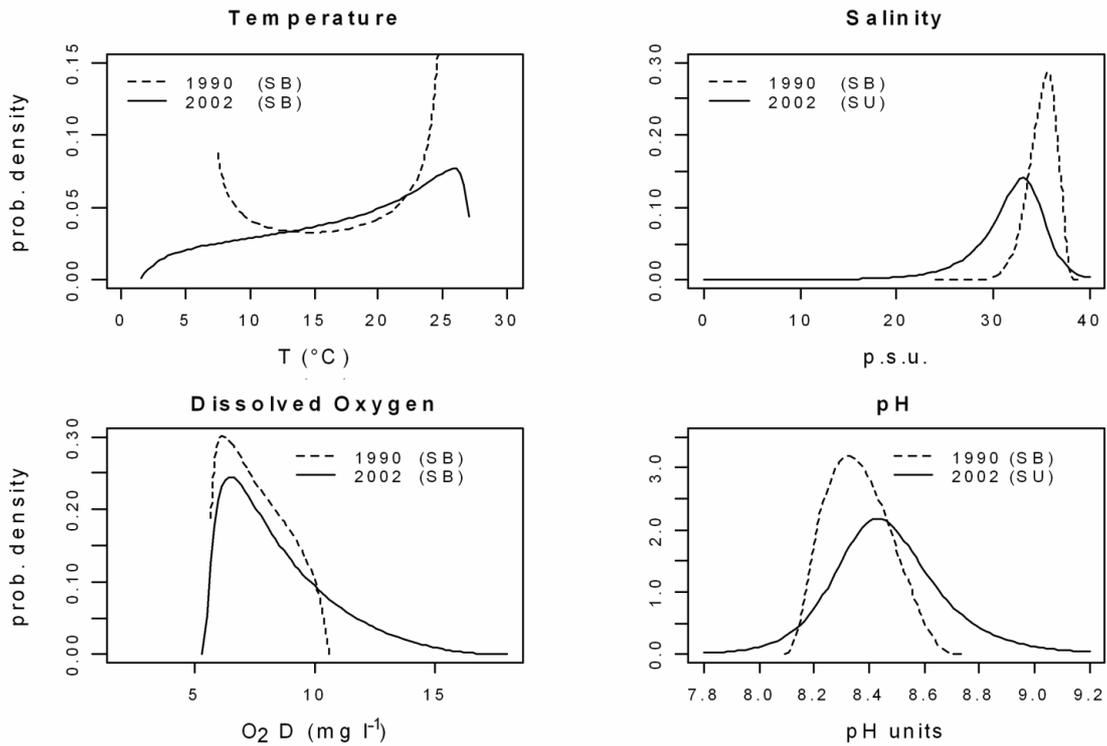
Fig. 9. Adriatic coastal waters: sampling station 314 – Hydrological, physical and chemical data distributions.
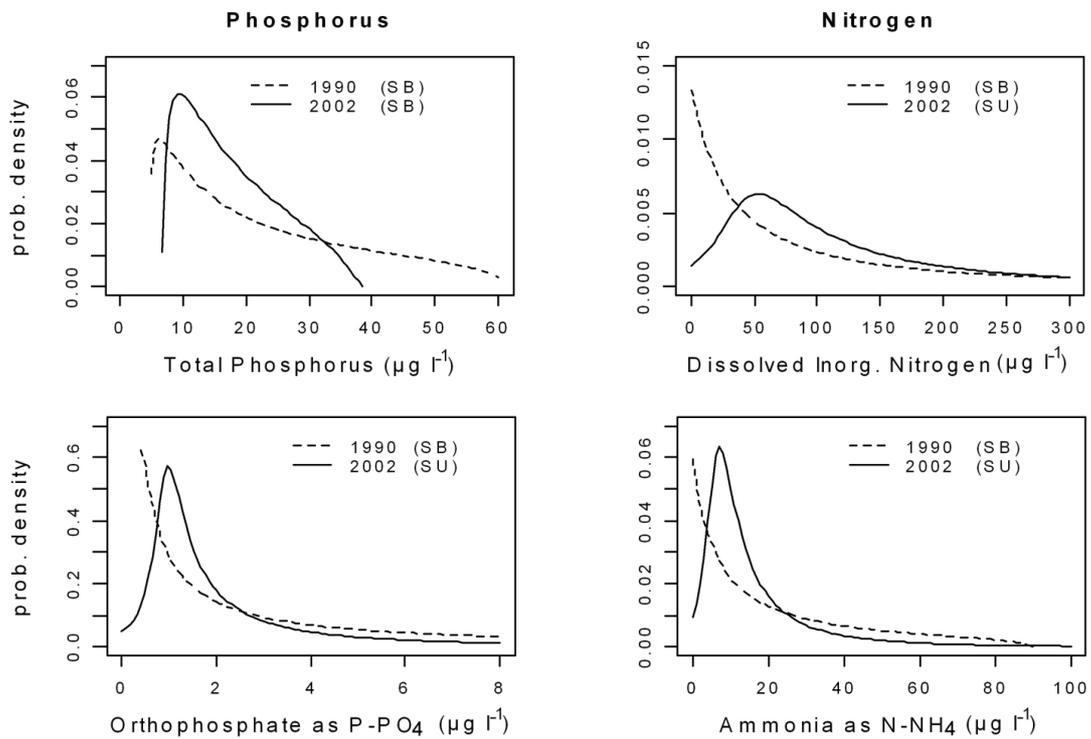
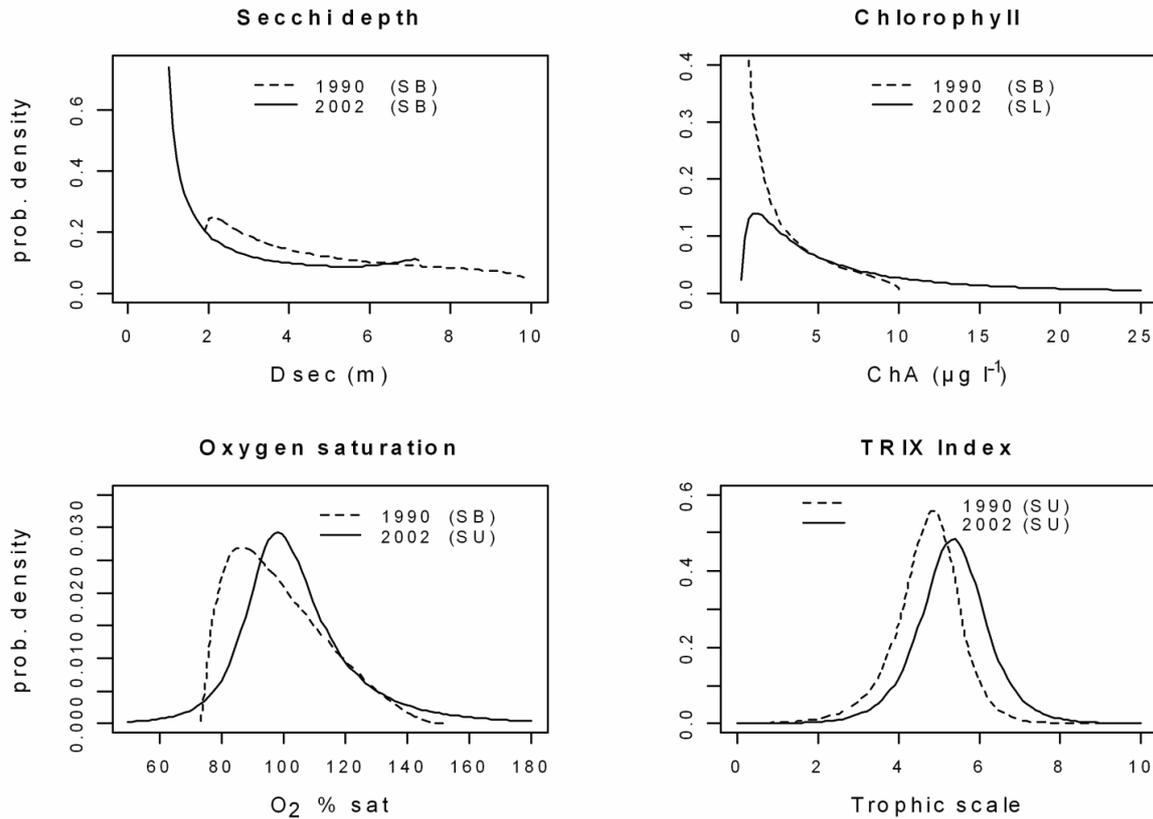**Fig. 10**. Nutrient data distributions.

Fig. 11. Trophic indicator parameters and TRIX data.

When we consider the data from Elba Island and from Tuscany littoral, south of Livorno, the tendency to bimodality becomes very marked, with winter minimum values never falling below 10 °C, and a maximum seasonal interval of only 12.6 °C (the peculiarity of the annual temperature data distributions and the characteristic anomaly in their frequency curves, which is highlighted in figure 3, is discussed, in detail, in the Appendix).

Salinity appears to drive the variability in both coastal systems (Fig. 3). Low salinity values are synonymous with high, nutrient rich, freshwater inputs from the continent. Salinity determines the scale of the variation in a considerable number of the examined parameters. The related frequency curves in both coastal systems, show salinity distributions that are strongly skewed, with a tail to the left side, towards low salinity values, depending on the effects of the river inputs. The salinity *p.d.f.* for Elba Island is very narrow and peaked, and has a notably reduced range of variation (S_B type, with: $\xi = 37.14$, $\lambda = 1.12$). We obtained opposite results in the coastal waters of the Adriatic, near the mouth of the Po River (distribution of S_U type, with an interval from 11.2 to 37.1 p.s.u., for $\alpha = 0.025$ and $\alpha = 0.975$ respectively).

A significant example of functional differentiation between the Adriatic and Tyrrhenian coastal systems is provided by the behaviour of the dissolved oxygen (Fig. 4).

The Adriatic coastal belt presented distributions of the S_U or S_L type, with a high range in variation in the concentration and saturation of O2 However, there were some some differences between the Northern zone of Goro-Po River delta and the Southern coastal reaches. Looking at the O_2% saturation values, cases of over-saturation are prevalent in the Northern area (Goro-Porto Garibaldi transects), with a tail more extended to the right hand, exceeding 140% sat. values, indicating that high biomass production activities are under way. Oxygen concentrations in that area are typically around 8 mg l^{-1}, compared to further South, where the typical values are around 6-6.5 mg l^{-1}. These results support a conceptual model of the coastal belt affected by the Po River inputs, that considers the Northern area as a zone of massive utilisation of nutrients to produce biomass (with generation of photosynthetic oxygen), and Southern areas (Cesenatico and Cattolica transects) as a zone of organic substance decomposition and mineralization , with consequent demand for and consumption of oxygen (Vollenweider 1995).

The diagrams for the Tyrrhenian coastal system show that the variation ranges of oxygen saturation are visibly less extended; the frequency curves are nearly always of S_B type, with an elevated degree of symmetry. In any case, the Arno River effect, of maintaining the high variability of the dissolved oxygen concentrations,

is clear from the large variation range in oxygen saturation in that area, from 5 to 9.5 mg l^{-1}, as compared to an average range of 6 to 9 mg l^{-1} for the remaining stretch of the Tuscany coast.

The pH values (Fig. 4) confirm the different behaviour of the two coastal systems.

We know that seawaters in equilibrium with the CO_2 in the atmosphere, typically have surface pH values of between 8.1 and 8.3 (see e.g. Sverdrup *et al.* 1942). In the Tyrrhenian system, beyond the upper limit of 8.3 pH units, we observe an immediate decay of the related frequencies. The Tuscany littoral shows pH distributions of S_B type, with more frequent values around 8.18. Along the Versilia coast, the presence of the Arno River does not alter the results; the mode of the related distribution (SU type) is around 8.2, with a maximum value of 8.35 ($\alpha = 0.975$). Therefore, we can infer that, despite freshwater inputs from the Arno, the Tyrrhenian coastal system is in hydro-chemical equilibrium.

When the photosynthetic activity of phytoplankton (and of plants in general) has reduced the amount of CO_2 in the water column, an increase of the pH over the upper limit of the "natural" range may occur. However, pH is also linked to the amount of dissolved oxygen. In eutrophied coastal areas, pH may also fall below the typical range. In fact, when the breakdown of organic matter exceeds synthesis activities, we see not only oxygen consumption, but also a consequent increase in CO_2 along the water column. In the Adriatic system, the buffer capacity of the seawater may often be at risk. The pH distributions are always of the unbounded type (S_U), with an evident platykurtic shape. Along the Goro and Port Garibaldi transects, the pH ranged from a minimum value of 7.97 ($\alpha = 0.025$) to a maximum of 8.82 ($\alpha = 0.975$), with a possibility of exceeding 9 units.

A comparison of nutrient distribution highlights the different trophic regimes in the two coastal systems. In order to simplify the comparison, we used the same scale for the concentration axes in each diagram. With respect to nitrogen and its dissolved compounds (Fig. 5), frequency curves are quite similar, both in terms of distribution shapes, strongly skewed toward the right side, and in terms of modal values. The frequency curves tend to rise very quickly on the left, in correspondence of the nitrogen lower concentrations, with higher peaks as the distance from the freshwater sources increases, due to the progressive decay of the nitrogen through physical dilution, uptake by phytoplankton, etc.

In contrast, the phosphorus behaviour (Fig. 6) is clearly distinct between the Adriatic and Tyrrhenian systems. Total form TP was quite different between the three Adriatic sub-areas, with a regular decrease of the modal values, from 23 µg l^{-1} in the Northern zone (G.-PG.), to 16 µg l^{-1} (CB.-L.Adr.), and to 8 µg l^{-1} in the Southern zone (C.-Ctt.). Conversely, the Tyrrhenian system shows markedly skewed curves that appear to

overlap, with modal values around 10 µg l^{-1} for Vs. and L.Ans., and around 7 µg l^{-1} for Elba Island.

The modal values for orthophosphate in the Adriatic system are very close to the analytical detection limit (*ca* 2 µg l^{-1} as P-PO$_4$), in the whole coastal area, with distributions of the S_L or S_U type. The variation range is very narrow and 10 µg l^{-1} of P-PO$_4$, chosen as a limit concentration, would be exceeded only in 10% of cases. The two Tyrrhenian sub-groups show much higher modal values (*ca* 7 µg l^{-1} as P-PO$_4$), and the probability of exceeding the limit of 10 µg l^{-1} is around 40-50%.

These results indicate that the amount of bioavailable phosphorus remains continually elevated in the Tyrrhenian system, but not in the Adriatic system. For further consideration of the peculiarity of the two trophic regimes and the possible causes of the different trophic responsessee Giovanardi & Vollenweider (2004).

In terms of productivity, figures 7 and 8 summarize the significance of different trophic responses to nutrient availability. Adriatic waters are much less transparent, with a maximum Secchi depth of only 3.2 m in the G.-P.G. transect. Chlorophyll and biomass density can reach maximum values of 50 µg l^{-1} and 10 millions cells l^{-1}, respectively. TRIX Index distributions are always highly symmetric, of a normal distribution, or very close to normal, and show a nearly constant variation range of *ca* 2-2.5 TRIX units. Average annual values of 5.5-6 TRIX units are typical of the Adriatic eutrophied coastal area, and confirm the suffering status of the Adriatic-Po River system (Vollenweider *et al.* 1998).

Conversely, Tuscany littoral results indicate a scarcely productive system, with values of Secchi depth and chlorophyll concentration which confirm the oligotrophic status of Tyrrhenian coastal waters. In practice, TRIX Index distributions were similar in both systems, with an average value around 4 units for Elba Island waters and for the Tuscany coast from Livorno to Ansedonia. Despite the Arno River inputs, the Versilia coastal waters show low phytoplankton densities, around 200,000 cells l^{-1} (as a modal value), and an annual TRIX mean around 5 units indicating a meso-trophic coastal area.. (Note: following the criterion of the TRIX trophic scale, the Italian coastal waters are currently classified into "elevated status" with annual TRIX mean <4, "good status", between 4 and 5, "mediocre status" between 5 and 6, and "bad status" with annual TRIX means >6).

The comparison of distributions and frequency curves may be very useful to the analysis of system evolution over time. The examples reported in figures 9 to 11 use data from the same Adriatic sampling station (st. 314, 3 km off-shore Cesenatico). The years under comparison are 1990 and 2002. Over 12 years, we would expect some improvement in the water quality parameter of trophic status, as a consequence of sanitation policies of nutrient control over the entire Po River
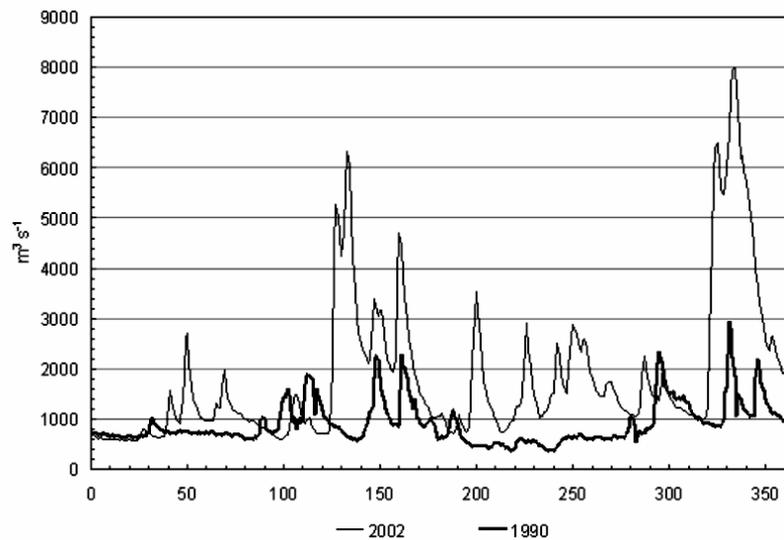
Fig. 12. Po River flow rates.

basin. In the last two decades, we have seen a notable commitment of the Emilia-Romagna regional government to adopting adequate regulations for fertiliser spreading and animal manure disposal, and completing the waste water collection and improving treatment plants along the coast, using tertiary treatment for phosphorus removal. However, our results do not meet these expectations. In 2002, the trophic levels are still "mediocre", with an annual TRIX mean of 5.34 units. Statistical distribution of TRIX data points indicates that the limit of 5 (the lower boundary of "mediocre" status), may be exceeded in 66% of cases, and the limit of 6 ("bad" status) in the 23% of the cases. In 1990, trophic conditions of the Adriatic coastal area appeared to be quite different. Trophic levels on average were "good" (TRIX annual mean = 4.7) and the probability of exceeding the critical levels of 5 and 6 TRIX units was 37% and 5%, respectively. This situation was surely more acceptable.

By analysing the frequency curves of the other water quality variables, we observe that, in 2002, high variability of salinity affects the behaviour of nitrogen (DIN and nitrogen-ammonia), orthophosphate and chlorophyll. In particular, the modal values of these parameters appear notably high, in comparison to the 1990 distributions. In terms of shape, the frequency curve from 1990 has no left tail.

The primary cause of this apparent deterioration is likely the different freshwater inputs coming from the Po River. Figure 12 shows the daily variation in flow rate in the two years under comparison. During 1990, there were no signfcant floods. The annual average flow rate was 903 m^3 s^{-1}, compared to an average of 1870 m^3 s^{-1} in 2002, a year characterised by large spring and autumn floods.

These differences in volumes entering the sea, necessarily contribute to strong disparities in the contribu-

tions and nutrient loads entering the Adriatic coastal area.

The evidence of a high inter-annual variability caused by the Po River flows leads us to conclude that the effects of the strategies of nutrient control and removal cannot be measured solely through simple comparisons of annual mean values. The above example, which shows extreme variability due only to "natural causes" indicates that the situation is much more complex. The normalisation methods described in this work provide a road map not only for applying best fits to monitoring data, but also for properly planning the analyses of variances, that would permit discrimination of the innumerable effects on the system behaviour induced by the different sources of variation (internal and/or external).

REFERENCES

Box, G.E.P. & D.R. Cox. 1964. An analysis of transformations. *J. Roy. Stat. Soc.* Ser. B, 26: 211-243.
Giovanardi, F. & E. Tromellini. 1992. Statistical Assessment of Trophic Conditions. Application of the OECD Methodology to the Marine Environment. Proc. Int. Conf. Marine Coastal Eutrophication. *Sci. Total Environ.* Suppl. 1992: 211-233.
Giovanardi, F., & R.A. Volleinweider. 2004. Trophic conditions of marine coastal waters: experience in applying the Trophic Index TRIX to two areas of the Adriatic and Tyrrhenian seas. *J. Limnol.*, 63: 199-218.
Heyman, U., S.O. Ryding & C. Forsberg. 1984. Frequency distributions of water quality variables. Relationships between mean and maximum values. *Water Res.*, 18: 787-794.
Hill, I.D., R. Hill & R.L. Holder. 1976. Fitting Johnson curves by moments. *Applied Statistics*. AS99. (Algorithm can be downloaded from: http://lib.stat.cmu.edu/apstat/99)
Ignatiades, L., M. Karydis & P. Voumatsou. 1992. A possibile method for evaluating oligotrophy and eutrophication based on nutrient concentration scales. *Mar. Poll. Bull.*, 24(5): 238-243.
Johnson, N.L. 1949. Systems of frequency curves generated by methods of translation. *Biometrika*, 36: 149-176.

Legendre, L. & P. Legendre. 1984. *Ecologie numerique. 1. Le traitement des données écologiques. 2. La structure des données écologiques.* Masson-Presse de l'Université du Quebec; deuxième édition;: 1° Vol. 260 pp., 2° Vol. 335 pp.

Maione, U. & U. Moisello. 1974. *Appunti di idrologia: introduzione alle elaborazioni statistiche.* La Goliardica Pavese.

R Development Core Team. 2005. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Sokal, R.R. & F.J. Rohlf. 1995. *Biometry – The Principles and Practice of Statistics in Biological Research.* 3rd ed. W.H. Freeman & Co. New York: 887 pp.

Sverdrup, H.U., M.W. Johnson & R.H. Fleming. 1942. *The Oceans. Their Physics, Chemistry, and General Biology.* Prentice-Hall, Inc.: 1087 pp.

Taylor, L.R. 1961. Aggregation, variance and the mean. *Nature*, 189: 732-735.

UNEP(DEC)/MED WG 231/14. 2003. *Eutrophication monitoring Strategy of MED POL.*

Vollenweider, R.A. & J.J. Kerekes (Eds). 1982. *Eutrophication of Waters, Monitoring, Assessment and Control.* O.E.C.D. Paris: 154 pp.

Vollenweider, R.A. 1995. L'Eutrofizzazione del bacino nord dell'Adriatico: considerazioni e valutazioni critiche sulle condizioni trofiche e sulla gestione dei sistemi marini costieri, In: Evoluzione dello stato trofico in Adriatico: analisi degli interventi attuati e future linee di intervento. *Atti del Convegno di Marina di Ravenna. 28-29 Settembre 1995.* Regione Emilia Romagna – Provincia di Ferrara – Autorità di Bacino del Fiume Po.

Vollenweider, R.A., F. Giovanardi, G. Montanari & A. Rinaldi. 1998. Characterization of the trophic conditions of marine coastal waters, with special reference to the NW Adriatic Sea: proposal for a trophic scale, turbidity and generalized water quality index. *Environmetrics*, 9: 329-357.

Wheeler, R.E. 1980. Quantile estimators of Johnson curve parameters. *Biometrika*, 67: 725-728.

APPENDIX

1. THE JOHNSON DIAGRAM FOR FREQUENCY CURVE CLASSIFICATION

The reference system for distribution types is built along the (β_1, β_2) plane, where β_1 = skewness to the square and β_2 = kurtosis +3. By inspecting the diagrams in figure 2, we observe that the log-normal system is defined by the points lying on the S_L line, which could be considered as a transition form between two more general systems of curves, one with variation ranges bounded at both extremities (System Bounded: S_B), the other unbounded at either extremity (System Unbounded: S_U). The S_B system covers the region between the log-normal line and the straight line where $\beta_2 - \beta_1 - 1 = 0$. The impossible area extends beyond this line, where the requested combinations between β_1 and β_2 will never occur. The remaining part of the (β_1, β_2) plane, below the S_L line, is covered by the S_U system.

1.1. The S_L line

The coordinates of the points lying on the SL line, have to be computed by the following parametric equations:

$$\beta_1 = (\omega - 1)(\omega + 2)^2 \qquad (\text{with } \sqrt{\beta_1} > 0)$$

$$\beta_2 = \omega^4 + 2\omega^3 + 3\omega^2 - 3,$$

where: $\omega = exp(\delta^{-2})$.

If the β_2 value of a given distribution is lower than the corresponding value, calculated as above, then that distribution will be classified as S_B, otherwise, if β_2 is higher, as S_U.

β_1, β_2, and ω values are reported in table 6, that allows us to plot the S_L curve, depending on significant values of δ.

Tab. 6. Johnson scheme of frequency curves classification: characteristic β_1 and β_2 values for plotting S_L curve, depending on significant values of δ.

δ	ω	β_1	β_2
$+\infty$	1	0	3
10	1.010	0.091	3.162
5	1.041	0.38	3.68
2	1.284	3.06	8.90
1.5	1.560	7.09	17.80
1	2.718	38.3	113.9
0.9	3.44	72.0	253.2
0.7	7.7	629.7	4596.3
0.5	54.6	171,694	9,220,560

1.2. The Bimodal Curves Boundary in the S_B region

In the Johnson system, bimodal frequency curves are a particular case of the S_B family.

The limit beyond which the (β_1, β_2) points identify this type of curve, is indicated in figure 2 with a dotted line. Necessary and sufficient conditions for bimodality, (regardless the sign of γ), are:

$$\delta < 1/\sqrt{2}$$

$$|\gamma| < \delta^{-1}\sqrt{(1-2\delta^2)} - 2\delta \cdot \operatorname{arctanh}\sqrt{(1-2\delta^2)}$$

The related β_1 and β_2 coordinates are provided in table 7, and are calculated for some combinations of δ and γ, in agreement with the above conditions.

Tab. 7. Bimodal curves boundary line: characteristic β_1 and β_2 values for assigned δ and γ combination.

δ	$|\gamma|$	β_1	β_2
0.707107	0	0	1.866122
0.70	0.003	6.9E-06	1.86
0.68	0.021	0.0004	1.84
0.65	0.065	0.004	1.81
0.60	0.175	0.035	1.79
0.55	0.330	0.14	1.86
0.50	0.533	0.42	2.13
0.45	0.793	1.08	2.86
0.40	1.125	2.75	4.75
0.35	1.553	7.71	10.5
0.33	1.759	12.5	16.0
0.30	2.117	29.4	35.8

Temperature data provide a particular example of bimodality. The resulting *p.d.f.* curves, for an entire annual cycle, sometimes do not match the corresponding frequency bars. For example, in figure 3, which shows the data of L-Ans. and Elb.I, such correspondence is definitively lost, and the related frequency curves assume an unlikely "U" shape.

This strange behaviour is clear from figure 13. Panel (A) of this figure shows the temperature data from on shore sampling stations situated close to the mouth of the Po River. The measurements are from the year 2004. The frequency curve is slightly bimodal and tends to faithfully follow the histogram bars, using both the moments and the quantiles method to fit the *p.d.f.* Numerical values of the related parameters of such a distribution are:

$$\delta = 0.628 \; ; \; |\gamma| = 0.165; \quad \beta_1 = 0.024; \qquad \beta_2 = 0.1.816.$$

In the β_1, β_2 plane, this distribution falls more or less exactly on the line identifying the lower boundary of the bimodality. Such behaviour is often found in marine areas directly affected by river inputs, with quite constant frequencies for the whole annual variation in temperature. The final result is a typical S_B distribution, with a strong tendency to the uniformity.

Panel (B) shows a quite different situation. This panel reports the frequency class distribution of temperatures in Tyrrhenian coastal waters, from Livorno transect to Ansedonia. The parameter values of the related *p.d.f.* are:

$$\delta = 0.354; \quad |\gamma| = 0.119; \quad \beta_1 = 0.018; \qquad \beta_2 = 1.468.$$

In the Johnson system of classification, this distribution lies beyond the bimodality boundary, and falls dangerously close to the straight line $\beta_2 - \beta_1 - 1 = 0$, which fixes the limit of the impossible area. The same relationship for the curve in panel (A) is $\beta_2 - \beta_1 - 1 = 1.816 - 0.024 - 1 = 0.79$; for the curve in panel (B), it is worth $1.468 - 0.018 - 1 = 0.45$. As this value approaches zero, *p.d.f.* curves will increasingly present as two ascending branches at the extremities, until they are represented by two simple ordinates. In the paper of Hill *et al*. (1976), these limit-distributions are classified in the S_T family, where T represents "two-ordinates" and the related Johnson parameters, ξ and λ, are fixed depending on the values of the x at which ordinates occur, δ is set to the proportion of values at λ, and γ is set, arbitrarily, to zero.

Generally, in our analyses, we observed accentuated bimodality in temperature distributions, but they may also appear, if in a more modest version, in Secchi depth and nutrients. In any case, those parameters are subjected to variation sources of a markedly seasonal character.

In practice, it would be more correct to consider those particular distributions as a sum of two distinct sub-sets of data, as represented in panel (C), where the same annual statistical sample in (B) has been divided into winter and summer temperature data sets. We obtain two contiguous distributions, the first of normal kind (S_N) and the second close to normality.

2. CONCLUDING REMARKS

Further exploration of these topics would exceed the object of the present work. The reader more interested in the theoretical aspects of translation, may consult the original paper of Johnson, which includes a graphical representation of the normalisation method, that clarifies the meaning of the γ and δ parameters, and their role in determining the shape of frequency curves. We have traced the Johnson paper to the Library of the "Istituto di Statistica" of the Rome University "La Sapienza".

Finally, the calculation algorithms (Moments method and Quantiles method) may be downloaded *via* the Internet, from the library of Applied Statistics, from the the e-mail address in bibliography and from the site: http://www.bobwheeler.com/stat.
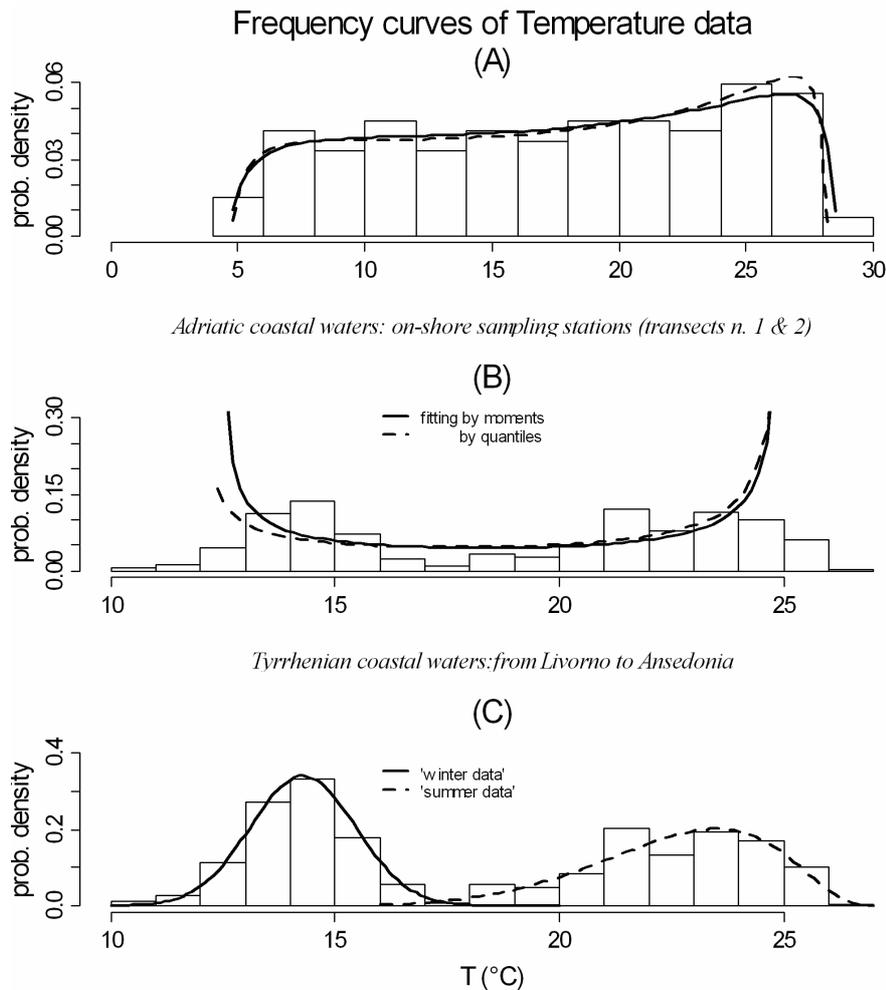
Frequency curves of Temperature data

(A)

Adriatic coastal waters: on-shore sampling stations (transects n. 1 & 2)

(B)

Tyrrhenian coastal waters:from Livorno to Ansedonia

(C)

T (°C)

Fig. 13. The case of the Temperature data distributions.

To complete the methodological framework, an example of elaboration on rough data, is reported using a script prepared for R, and providing both graphical and numerical outputs. The input data-files are text-files with .txt or .csv extension. In this example, the input-file "ChA_data.txt" contains rough chlorophyll data (sample size = 135), from some on-shore sampling stations of the Emilia-Romagna coastal area.

INPUT:
>library(SuppDists) ## Loads package 'Supplementary Distributions'
>data=read.table("ChA_data.txt",header=TRUE) ## Read a file in table format and ##creates a data frame from it, with cases corresponding to lines and variables ##to fields in the file.
>attach(data) ## The database is attached to the R search path.
>fit_ChA=JohnsonFit(ChA) ## (ChA) is the header of the column data
>sJohnson(fit_ChA) ## Outputs a list containing the Johnson parameters, the type ##of curve and the other distribution parameters.
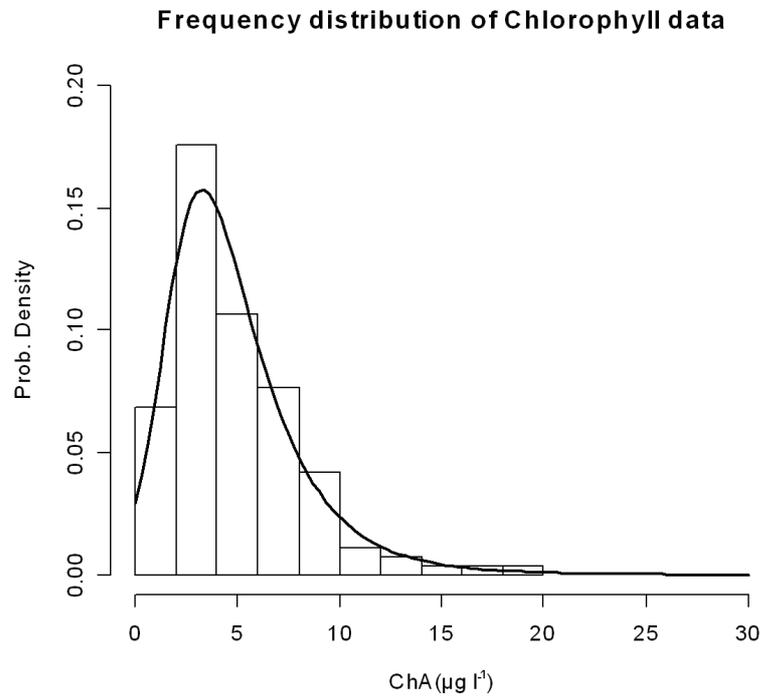
OUTPUT:
"Johnson Distribution" $gamma [1] -1.513282 $delta [1] 1.696650 $xi [1] 0.8646296 $lambda [1] 3.243419 $type SU 3 $Mean [1] 4.781085 $Median [1] 4.156621 $Mode [1] 3.310031 $Variance [1] 11.64819
$SD [1] 3.412944 $ThirdCentralMoment [1] -66.50013 $FourthCentralMoment [1] 1303.414 $PearsonsSkewness...mean.minus.mode.div.SD [1] 0.4310219 $Skewness...sqrtB1 [1] -1.672765 $Kurtosis...B2.minus.3 [1] 6.606508
INPUT:> pJohnson(15,fit_ChA) ## gives the distribution function value corresponding to x=15 μg l-1 ##of ChA.
OUTPUT: [1] 0.9854529
INPUT:> qJohnson(.99,fit_ChA) ## gives the resulting x value, from an assigned quantile. (inverse ## of pJohnson( ) function).
OUTPUT: [1] 16.28462
INPUT:
>hist(ChA,prob=T,nclass=10,xlim=c(0,30),ylim=c(0,0.2)) ## computes a histogram of ## the given data values.

```
>plot(function(ChA)dJohnson(ChA,fit_ChA),0,30,add=T,lwd=2,col=2)
## "plot( )": function for plotting of R objects.  "dJohnson( )": gives the density function values.
```

OUTPUT:

Frequency distribution of Chlorophyll data