Bayesian analysis of the multivariate dependence of three transitional water ecosystem classifications.

A. Pollice, S. Arima, G. Jona Lasinio, A. Basset, I. Rosati

Abstract The Water Framework Directive (WFD) recognizes benthic macroinvertebrates as a good biological quality element for transitional waters as they are the most exposed to natural variability patterns characteristic of these ecosystems, due to their life cycles and space-use behavior. Here, we address the ecological status classification issue for three lagoons in Apulia, using benthic macroinvertebrates and three proposed multimetric indices (namely M-AMBI, BITS and ISS), likely to respond differently to different sources of stress and natural variability. Lagoon classification is based on discretization by standard classification boundaries with only partial consideration of the natural variability of ecosystem properties and possible inaccuracies of the classification procedures. In order to investigate the possible contrasting behavior of the three classifications, we propose Bayesian hierarchical models in which the multimetric indices and their discrete counterparts are jointly modeled as function of abiotic covariates, external anthropogenic pressures indicators and spatio-temporal effects.

Key words: Multivariate ordinal data, Classification, Bayesian inference, Ecosystem status

---

A. Pollice
Dipartimento di Scienze Economiche e Metodi Matematici, Università degli Studi di Bari Aldo Moro, Largo Abbazia Santa scolastica 53, 70124 Bari, e-mail: alessio.pollice@uniba.it

S. Arima
Dipartimento di Metodi e Modelli per l'Economia, il Territorio e la Finanza, Sapienza Università di Roma, via del Castro Laurenziano 9, 00161 Roma, e-mail: serena.arima@uniroma1.it

G. Jona Lasinio
Dipartimento di Scienze Statistiche, Sapienza Università di Roma, via del Castro Laurenziano 9, 00161 Roma, e-mail: giovanna.jonalasinio@uniroma1.it

A. Basset, I. Rosati
Dipartimento di Scienze e Tecnologie Biologiche ed Ambientali, Università del Salento, Prov.le Lecce-Monteroni, 73100 Lecce, e-mail: {alberto.basset, ilaria.rosati}@unisalento.it

1 Introduction

Lagoons represent important and fragile ecosystems in the coastal landscape, however their geographic position along the coast and their close relation with terrestrial ecosystems make these environments especially vulnerable to anthropogenic pressures [16]. The need to act has been acknowledged by politicians and legislation has been adopted to stop further deterioration and restore lagoons healthy state. The Water Framework Directive (WFD, 2000/60/EC) requires EU Member States to assess the ecological status of each water body in Europe and to ensure a sustainable management such that good ecological quality of all water bodies will be obtained by 2015. Lagoons can be clustered into types [3, 9], but display several different internal gradients of physical conditions and hence of the biota associated with them. The ecological status of aquatic ecosystems is defined in terms of the quality of the biological community, as well as the systems' hydrological and chemical characteristics. Several simple indicators as the Shannon-Wiener index, the Margalef index and the AMBI index account for the composition and abundance of biological communities and are widely used in the ecological literature. Multimetric indices, combining simple indices as multiple sources of information, focus on benthic macroinvertebrates which are known to be sensitive to both natural and anthropogenic pressures [8, 10]. In this paper, we will focus of three multimetric indices: M-AMBI [13, 7], BITS [12] and ISS [4]. The so called *a priori* approach to lagoon classification by multimetric indices was introduced in [6] and used in [7, 4]. According to some reference samples, the authors choose boundary values of the multimetric indices to define ecological status classes and classify the lagoons according to these values. However the proposed indices are likely to respond differently to different sources of stress and natural variability components, adding uncertainty to resulting classifications. As the *a priori* approach does not take into account the sensitivity of benthic macroinvertebrate taxa to different sources of abiotic heterogeneity in lagoon ecosystems, the same lagoon can be classified in a different ecological status according to the different indices. To properly understand what drives multimetric indices disagree in classifying lagoons' ecological status, statistical models linking values of the indices with abiotic information and models relating ecological status categories with the same explanatory variables are suitable tools. In particular Bayesian hierarchical models, allowing for the inclusion of multiple sources of information and external prior knowledge, are adopted in what follows.

2 Materials and methods

Data on benthic macroinvertebrates colonizing various habitat types were collected in 3 transitional water ecosystems in Apulia: Alimini, Lesina and Varano. Seasonal field sampling campaigns were performed in 2008 and 2009 in the three ecosystems. Overall 15 sites were sampled pooling the three lagoons. For each sampling site three replicates were collected by manual Reineck box-corer (0.03 m2) in Alimini and an Ekman-Birge grab in Lesina and Varano. In the laboratory, benthic samples were sorted under a stereomicroscope, identified to the lowest possible taxonomic level, counted, measured individually (total length for most taxa) and weighted. Chemical and physical water parameters (water temperature, dissolved oxygen, salinity, pH and chlorophyll) were monitored at each station. For each lagoon, the macrobenthic community was examined in order to build two of the most common indices proposed in the Water Framework Directive and a new index for benthic assessment. The multimetric indices M-AMBI [13], BITS [12], and ISS [4] were calculated at the replicate level. Expert's opinion evaluations of four external anthropogenic pressures (agricultural diffuse inputs, domestic discharges, industrial discharges, fin fisheries) were also available for each monitoring station.

In this paper, in order to investigate the possible contrasting behavior of the three indices, we propose to analyze both their values and the three classifications obtained by the *a priori* approach (section 1) building two independent Bayesian hierarchical models. First the values of the multimetric indices are jointly modeled as functions of abiotic covariates, experts opinion on external anthropogenic pressures, spatial location of the monitoring sites and temporal sequence of the samples. Then an analogous model for multivariate ordinal responses investigates the relation of the three classifications with the same quantities. Here the following notation is adopted: $r = 1,\ldots,R$ is a replicate of a biotic record taken at time $t = 1,\ldots,T$ and location $s = 1,\ldots,n_l$ within lagoon $l = 1,\ldots,L$. Here specifically we have $R = 3$ replicates, $T = 8$ times, $L = 3$ lagoons with $n_l = 6,6,3$ monitoring stations respectively for Lesina, Varano and Alimini.

First we present a model for the multivariate continuous response. For each record the three-dimensional response vector $\mathbf{z}_{rtsl}$ contains values of the three multimetric indices M-AMBI, BITS and ISS. For each monitoring station and every time point the P-dimensional vector $\mathbf{x}_{tsl}$ contains values of P continuous and ordinal explanatory variables.

$$\mathbf{z}_{rtsl} = \boldsymbol{B}\mathbf{x}_{tsl} + w_{tsl}\mathbf{1}_3 + \boldsymbol{\varepsilon}_{rtsl} \tag{1}$$

where $\boldsymbol{B}$ is a $3 \times P$ matrix of index-specific regression coefficients that measure the effect of each abiotic covariate on every multimetric index, $\mathbf{1}_3$ is a unit vector, w_{tsl} is a latent Gaussian process describing the space-time variation common to the three multimetric indices within each lagoon ($l = 1,\ldots,L$) and $\boldsymbol{\varepsilon}_{rtsl}$ is a 3-dimensional correlated random error vector. At the second level each element of $\boldsymbol{B}$ is independently normally distributed and the space-time variation is specified as follows:

$$\boldsymbol{B}(j,p) \sim N(\mathbf{0},\sigma^2_{\boldsymbol{B}})$$

$$\mathbf{w}_l \sim N_{Tn_l}(\mathbf{0},\sigma^2_{\mathbf{w},l}\mathbf{H}(\phi_l))$$

$$\boldsymbol{\varepsilon}_{rtsl} \sim N_3(\mathbf{0},\Sigma_{\boldsymbol{\varepsilon}})$$

where $j = 1,2,3$ for M-AMBI, BITS and ISS and $\mathbf{w}_l = (w_{11l},\ldots,w_{Tn_l l})'$. Here $\sigma^2_{\mathbf{w},l}$ and $\mathbf{H}(\phi_l)$ are the variance component and the space-time correlation matrix of the

l-th lagoon, $l = 1, \dots, L$. Notice that, the correlation between the three indices is only accounted for by the unstructured covariance matrix $\Sigma_{\boldsymbol{\varepsilon}}$.

A multivariate ordinal response $\mathbf{y}_{srtl}$ is obtained from the observed continuous values of the three indices in $\mathbf{z}_{srtl}$ letting $\boldsymbol{\gamma}$ be a matrix of cutpoints, one row for each index, six columns corresponding to standard boundary values defining five ecosystem status classes:

$$\gamma_{j,c-1} < z_{rtslj} \leq \gamma_{j,c} \quad \Rightarrow \quad y_{rtslj} = c \quad c = 1, \dots, 5$$

Let the best ecological status (category 5) be the baseline category and $\boldsymbol{\ell}_{rtslj} = \text{logit}\left(\Pr(y_{rtslj} \leq c)\right)$ be the 4-dimensional vector of the cumulative logits for the j-th index at record $rtsl$ (i.e. replicate r, time t, site s, lagoon l), with $c = 1, \dots, 4$. If $\boldsymbol{L}_{rtsl} = (\boldsymbol{\ell}'_{rtsl1}, \boldsymbol{\ell}'_{rtsl2}, \boldsymbol{\ell}'_{rtsl3})'$ denotes the 3×4 matrix corresponding to the three indices, then the multivariate ordinal response of a generic record can be given the following *cumulative proportional odds logit model* [1] [11] in matrix notation:

$$\boldsymbol{L}_{rtsl} = \mathbf{1}_3 \boldsymbol{\alpha} + \boldsymbol{B}\mathbf{x}_{tsl} \mathbf{1}_4 + w_{tsl} \mathbf{1}_{3\times 4} \tag{2}$$

with $\boldsymbol{B}$ and $\boldsymbol{w}_l$ specified as before and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_4)$, $\alpha_c \sim N(0, \sigma^2_\alpha)$ for $c = 1, \dots, 4$. Notice that here α_c's are increasing in c, since $\Pr(y_{rtslj} \leq c)$ increases in c and the logit is an increasing function of this probability.

In this model the effect of the p-th covariate on the j-th index is the same for each logit, i.e. for $c = 1, \dots, 4$. The cumulative logit model (2) satisfies:

$$\begin{aligned} &\text{logit}\left(\Pr(y_{rtslj} \leq c) | \boldsymbol{x}'_{tsl}\right) - \text{logit}\left(\Pr(y_{rtslj} \leq c) | \boldsymbol{x}''_{tsl}\right) \\ &\quad = \log \frac{\Pr\left(y_{rtslj} \leq c) | \boldsymbol{x}'_{tsl}\right) / \Pr\left(y_{rtslj} > c) | \boldsymbol{x}'_{tsl}\right)}{\Pr\left(y_{rtslj} \leq c) | \boldsymbol{x}''_{tsl}\right) / \Pr\left(y_{rtslj} > c) | \boldsymbol{x}''_{tsl}\right)} = \boldsymbol{B}_j \left(\boldsymbol{x}'_{tsl} - \boldsymbol{x}''_{tsl}\right) \end{aligned} \tag{3}$$

The odds of ecological status $\leq c$ for the j-th index at $\boldsymbol{x}'_{tsl}$ are $\exp\left[\boldsymbol{B}_j \left(\boldsymbol{x}'_{tsl} - \boldsymbol{x}''_{tsl}\right)\right]$ times the odds at $\boldsymbol{x}''_{tsl}$, where $\boldsymbol{B}_j$ is the j-th row of $\boldsymbol{B}$. The log cumulative odds ratio is proportional to the distance between $\boldsymbol{x}'_{tsl}$ and $\boldsymbol{x}''_{tsl}$ and the same proportionality constant applies to each logit ($c = 1, \dots, 4$), i.e. no matter how the cutpoints in $\boldsymbol{\gamma}$ divide the scale of the three indices. The effect parameters in $\boldsymbol{B}$ are thus invariant to the choice of categories for the ordinal response. This feature makes it possible to compare estimates for the three indices using different response scales (see [1] page 278). Notice that when $B_{jp} > 0$, as the covariate $\boldsymbol{x}_p$ increases each cumulative logit and each corresponding cumulative probability increase. Thus relatively more probability mass falls at the low end of the scale of the ordinal variable $\boldsymbol{y}_j$. As a consequence when $B_{jp} > 0$ the j-th index tends to assign a lower ecologic class with higher values of $\boldsymbol{x}_p$.

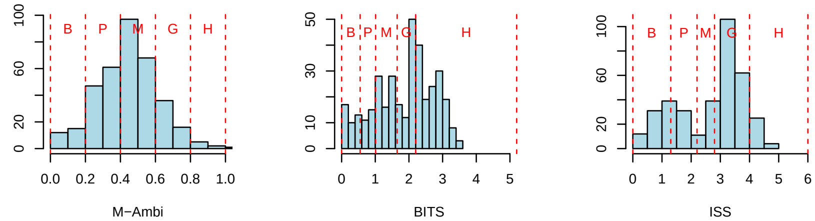

Fig. 1 Distributions of the biotic multimetric indices with standard classification boundaries.

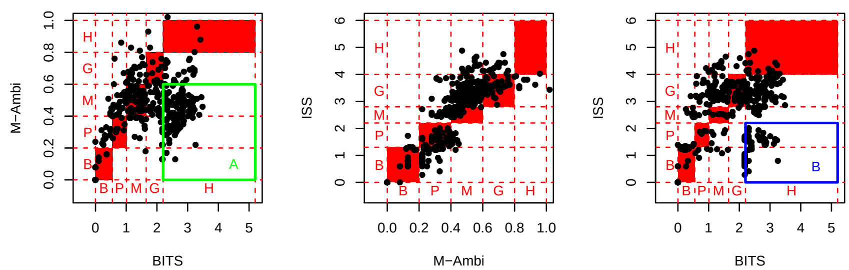

Fig. 2 Joint distributions of pairs of multimetric indices with standard classification boundaries. Areas A and B contain clusters of observation with contrasting behavior of two indices.

3 Results

The three biotic indices have different theoretical range and are usually considered in association with boundaries defined in the literature. Within the mentioned *a priori* approach boundaries allow to classify lagoon ecosystems in five quality categories (bad, poor, moderate, good, high), though they do not necessarily correspond to discontinuities in the data, as Fig. 1 shows. Exploratory data analysis (not reported) does not provide evidence for appreciable one-to-one relations between multimetric indices, abiotic variables and the effects of external pressures. In some circumstances M-AMBI, BITS and ISS can be in disagreement and lead to contrasting assessments of the ecological status of the same ecosystem. The relation among the three indices is expressed in Fig. 2. Clearly M-AMBI and ISS have stronger linear correlation ($R \simeq 0.84$), M-AMBI and BITS have marked nonlinear dependence with high variability for increasing values of BITS ($R \simeq 0.25$), finally BITS and ISS show a weaker dependence ($R \simeq 0.25$) and a cluster of records with high values of BITS and low values of ISS (area B in Fig. 2). In general terms we can confirm that while M-AMBI and ISS convey similar information, BITS does not. Due to the stochastic relation among the three multimetric indices, their use in association with standard boundaries can lead to contrasting ecosystem classifications. In Fig. 2 areas in red correspond to both indices assigning the same ecologic class to the ecosystems, but the majority of observed records fall outside these areas. It is very evident that the “high” BITS category corresponds to a very variable behavior of both M-AMBI and ISS. In Fig. 2 points/records within areas A and B correspond to discrepant values for M-AMBI and BITS and for BITS and ISS respectively.

Inferences on the previous ecological issues regarding indices variability and lagoon classification are respectively obtained by hierarchical models (1) and (2), properly accounting for all sources of natural and external variability. Notice that a simplification of the general spatio-temporal term w_{tsl} in (1) and (2) was necessary, due to data availability. Indeed the three lagoons are geographically well separated and most likely independent and the small number (3 to 6) of monitoring stations within each lagoon does not allow the estimation of a spatial covariance. Simple forms of temporal autocorrelation or seasonality are even hardly detectable with only 8 time points and exploratory data analysis didn’t show any form of longitudinal trend. Then only the fixed effects of the lagoons were considered as spatial effects and no time effects were taken into account in the model specification.

Here we adopt a Bayesian approach and estimate model parameters by a Markov chain Monte Carlo (McMC) posterior simulation algorithm. For actual implementation we use the WinBUGS [15] and JAGS [14] softwares with 2 chains with different starting points (code can be obtained upon request to the authors). For each chain, we allow a 50000 samples burn-in and estimate the posterior distribution of quantities of interest by 50000 iterations thinned by 100. Chain convergence was ascertained by visual inspection of standard convergence diagnostic tools, such as trace plots and autocorrelation plots.

Tables 1 and 2 show the posterior estimates of index-specific regression coefficients and the 95% credibility intervals obtained with the multivariate continuous response model (Table 1) and with the multivariate ordinal response model (Table 2). In both models abiotic indices and expert’s opinion evaluation of four external anthropogenic pressures (agricultural, domestic, industrial and due to the presence of fisheries) have been considered as explanatory variables. In Table 1, pH and domestic discharges positively influence the average value of the three indices. On the other hand the average value of the indices significantly decreases when industrial discharges and the amount of chlorophyll increase. Fisheries negatively influences the average value of M-AMBI and ISS but not BITS which is inversely proportional to the amount of dissolved oxygen. Values in Table 2 allow to draw conclusions similar to those obtained for Table 1, where coefficients with opposite sign should be interpreted as explained in Section 2. In addition model (2) shows that the amount of dissolved oxygen positively affects M-AMBI and ISS. Salinity significantly acts on both M-AMBI and BITS increasing their values, while it has an opposite effect on ISS according to model (1). Water temperature negatively influences the BITS index but it is not significant for the other two indices. Domestic and diffuse agricultural discharges, as far as mainly characterized by organic and inorganic inputs, are likely to have idiosyncratic impacts at the community and ecosystem levels in lagoon ecosystems, being strongly affected by the overall abiotic context (e.g., oxygen concentration and hydrodynamics) and differing among biological quality elements. In the previous tables the expected effects of agricultural pressure are probably masked by the presence of other informative variables (as domestic pressure and dissolved oxygen) in the predictors. The significance and the sign of the effects of all other pressures and abiotic variables are supported by sound ecological knowledge and have plausible interpretation. The two models confirm a known higher correlation between M-AMBI and ISS and a lower correlation of these two with BITS. Non significant lagoon effects are finally to be reported, although lagoon consideration improves the overall fit of both models.

Table 1 Posterior estimates of index-specific regression coefficients and 95% credibility intervals for model (1). In red 95% significant estimates.

	M-AMBI		BITS		ISS	
Water temperature	-0.000	(-0.004, 0.003)	-0.015	(-0.032, 0.002)	0.008	(-0.014, 0.029)
Dissolved oxygen	0.008	(-0.001, 0.017)	-0.083	(-0.126, -0.040)	0.044	(-0.010, 0.099)
Salinity	-0.001	(-0.006, 0.005)	-0.001	(-0.019, 0.018)	-0.034	(-0.057, -0.011)
pH	0.088	(0.042, 0.136)	0.375	(0.274, 0.478)	0.399	(0.276, 0.522)
Chlorophyll	-0.008	(-0.012, -0.004)	-0.052	(-0.072, -0.033)	-0.055	(-0.080, -0.031)
Agricultural inputs	-0.003	(-0.028, 0.022)	0.001	(-0.115, 0.118)	0.093	(-0.054, 0.240)
Domestic discharges	0.046	(0.022, 0.069)	0.170	(0.058, 0.282)	0.249	(0.107, 0.390)
Industrial discharges	-0.079	(-0.116, -0.041)	-0.248	(-0.420, -0.076)	-0.245	(-0.463, -0.028)
Fisheries	-0.054	(-0.084, -0.024)	0.092	(-0.000, 0.184)	-0.205	(-0.321, -0.089)

Table 2 Posterior estimates of index-specific regression coefficients and 95% credibility intervals for model (2). In red 95% significant estimates.

	M-AMBI		BITS		ISS	
Water temperature	0.021	(-0.011, 0.052)	0.082	(0.043, 0.124)	0.001	(-0.034, 0.033)
Dissolved oxygen	-0.104	(-0.183 , -0.025)	0.267	(0.163, 0.315)	-0.143	(-0.213, -0.051)
Salinity	-0.058	(-0.091, -0.013)	-0.065	(-0.111, -0.012)	0.001	(-0.044, 0.052)
pH	-1.011	(-1.376 , -0.625)	-1.507	(-2.102 , -0.973)	-1.076	(-1.501, -0.688)
Chlorophyll	0.073	(0.035, 0.105)	0.172	(0.133, 0.227)	0.113	(0.076, 0.151)
Agricultural inputs	0.102	(-0.115, 0.327)	0.034	(-0.212, 0.283)	-0.117	(-0.339, 0.127)
Domestic discharges	-0.347	(-0.543, -0.146)	-0.541	(-0.803, -0.302)	-0.466	(-0.707, -0.244)
Industrial discharges	0.664	(0.337, 0.968)	0.935	(0.534, 1.327)	0.789	(0.431, 1.120)
Fisheries	0.541	(0.295, 0.774)	-0.061	(-0.312, 0.194)	0.700	(0.471, 0.944)

References

1. Agresti A.: Categorical Data Analysis - Second Edition. Wiley, Hoboken NJ (2002)
2. Barbone, E., Rosati, I., Reizopoulou, S., Basset A.: Linking classification boundaries to sources of natural variability in transitional waters: a case study of benthic macroinvertebrates. Ecological Indicators. **12**, 105–122 (2012)

3. Basset, A., Sabetta, L., Fonnesu, A., Mouillot, D., Do Chi, T., Viaroli, P., Reizopoulou, S., Carrada, G.C.: Typology in Mediterranean transitional waters: new challenges and perspectives. Aquatic Conservation: Marine and Freshwater Ecosystems. **16**, 441–445 (2006)
4. Basset, A., Barbone, E., Borja, A., Brucet, S., Pinna, M., Quintana, X.D., Reizopoulou, S., Rosati, I., Simboura,N.: A benthic macroinvertebrate size spectra index for implementing the Water Framework Directive in coastal lagoons in Mediterranean and Black Sea. Ecological Indicators. **1240**, 72–83 (2012)
5. Basset, A., Barbone, E., Borja, A., Jona Lasinio, G., Marques, J.C., Mazik, K., Muxika, I., Neto, J.M., Teixeira, H., Rosati, I.: Natural variability and reference conditions: setting type-specific classification boundaries for lagoon macroinvertebrates in the Mediterranean and Black Seas. Hydrobiologia. Doi: 10.1007/s10750-012-1273-z (2012)
6. Borja, A., Franco, J., Perez, V.J.: A marine biotic index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. Marine Pollution Bulletin **40**, 1100–1114 (2000)
7. Borja, A., Franco, J., Valencia, V., Bald, J., Muxika, I., Belzunce, M.J., Solaun, O.: Implementation of the European Water Framework Directive from the Basque Country (northern Spain): a methodological approach. Marine Pollution Bulletin **48**, 209–218 (2004)
8. Borja, A., Barbone, E., Basset, A., Borgersen, G., Brkljacic, M., Garmendia, J.M., Marques, J.C., Mazik, K., Muxika, I., Neto, J.M., Norling, K., Rodriguez, J.G., Rosati, I., Rygg, B., Texteira, H., Trayanova, A.: Response of single benthic metrics and multi-metric methods to anthropogenic pressure gradients, in five distinct Europen coastal and transitional ecosystems. Marine Pollution Bulletin, **62(3)**, 499–513 (2011)
9. Lucena-Moya, P., Pardo, I., Alvarez, M.: Development of a typology of transitional waters in Mediterranean ecoregion: the case of the islands. Estuarine, Coastal and Shelf Science. **82**, 61–72 (2009)
10. Marchini, A., Marchini, C.: A fuzzy logic model to recognise ecological sectors in the lagoon of Venice based on the benthic community. Ecological Modelling. **193(1-2)**, 105–118 (2006)
11. McKay Curtis S.: BUGS Code for Item Response Theory. Journal of Statistical Software. **36**, Code Snippet 1 (2010) http://www.jstatsoft.org/
12. Mistri, M. and Munari, C.: BITS: a SMART indicator for self-bottom, non-tidal lagoon. Marine Pollution Bulletin. **56**,587–599 (2008)
13. Muxika, I., Borja, A. and Bald, J.: Using historical data, expert judgement and multivariate analysis in assessing reference conditions and benthic ecological status, according to the European Water Framework Directive. Marine Pollution. **55**, 16–29 (2007)
14. Plummer, M.: JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling, Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 2022, Vienna, Austria (2003)
15. Spiegelhalter, D.J., Thomas, A. and Best, N.: WinBUGS Version 1.2 User's Manual. MRC Biostatistics Unit (1999). Software available at http://www.mrcbsu.cam.ac.uk/bugs/winbugs/contents.html
16. Viaroli, P., Mistri, M., Trousselier, M., Gierzoni, S., Cardoso, A.C.: Structure, functions and ecosystem alterations in Southern European coastal lagoons. Hydrobiologia. **550**, 57–71 (2005)