

SEARCHING FOR COMMON DIVERSITY TRENDS IN THE DEEP DEMERSAL ASSEMBLAGE OFF PORTUGAL, 1989-1999: AN ESSAY WITH DYNAMIC FACTOR ANALYSIS

Pedro Sousa*, Manuela Azevedo**, Manuel C. Gomes***

Resumo

Estimou-se o índice de diversidade de Shannon em cada estação de amostragem pertencente à comunidade demersal profunda da costa continental Portuguesa, com base em dados obtidos numa série de 22 cruzeiros demersais efectuados no período 1989-1999. As comunidades são definidas por grupos de espécies que ocupam uma área geográfica cujos contornos foram identificados e cuja coesão se mantém ao longo do tempo. Aplicou-se um método multivariado denominado Análise Factorial Dinâmica (*Dynamic Factor Analysis*) para investigar a existência de tendências temporais comuns de diversidade ao longo da última década. Foram detectadas duas tendências comuns (CT) mas apenas uma delas explicou a maioria da variabilidade observada. De acordo com esta CT não se verificaram grandes alterações temporais na diversidade. No entanto, detectou-se um ligeiro aumento em meados dos anos 90. Os resultados obtidos são discutidos com base na natureza dos métodos utilizados e também na abundância relativa das espécies mais importantes da comunidade.

Abstract

The Shannon index of diversity was estimated at each sampling station of the deep demersal assemblage off Portugal, using data from 22 research surveys in the 1989-1999 time period. Assemblages are spatially distinct and temporally persistent groups of species. A multivariate method called Dynamic Factor Analysis was applied to the diversity time series in order to search for common trends (CTs) along the nineties. Two common trends were detected

* Mestre em Matemática Aplicada às Ciências Biológicas, docente no INUAF.

** Investigadora Auxiliar do Instituto de Investigação das Pescas e do Mar, Lisboa.

*** Professor Associado da Faculdade de Ciências da Universidade de Lisboa.

with a single one accounting for the majority of variation. According to the main CT, no clear temporal shifts occurred, although a relative peak in diversity could be identified by the mid nineties. Observed patterns are discussed both in light of the methods used and with respect to the relative proportion of the most important species in the assemblage.

Keywords: common trends, demersal assemblages, Dynamic Factor Analysis, Shannon diversity (H).

1. Introduction

The definition of fish assemblages as being persistent groups of species that occupy geographically defined areas, was first introduced by Fager and Longhurst (1968). The identification of groups of species that tend to co-occur persistently, however, dates back to Fager (1957) and Day and Percy (1968). The fish assemblage concept appeared as a result of a change in the way scientists searched for the comprehension of the response of marine ecosystems to fisheries exploitation, shifting from the traditional single-species approach to the more holistic inclusion of interactions among species. The identification and description of fish assemblages has been conducted mostly for the northeastern and northwestern Atlantic shelves (e.g., Tyler *et al.*, 1982; Gabriel, 1989; Mahon and Smith, 1989; Gomes, 1993) but also for shelves all over the world, e.g., the southeastern Atlantic margin (Bianchi, 1992), the Indian Ocean (Adjeroud *et al.*, 1998), southwestern Atlantic (Araújo and Costa de Azevedo, 2001), Australia (Ayvazian and Hyndes, 1995), Alaska (Barber *et al.*, 1997), Arabian Gulf (Coles and Tarr, 1990), French Polinesia (Galzin, 1987), north-central Pacific (Jennings *et al.*, 1994), Yellow Sea (Jin, 1995), South Africa (Russel, 1997) and southeastern Pacific (Varas and Ojeda, 1990).

Sousa *et al.* (2005) have identified five distinct demersal¹ assemblages on the Portuguese continental margin (Figure 1), namely, (i) shallow northern and (ii) southern assemblages, (iii) intermediate northern and

¹ Demersal species are those that either dwell near the ocean bottom or its life cycle depends, somehow, on it. Conversely, pelagic species, such as sardine (*Sardina pilchardus*), for example, inhabit the water column, i.e., the open ocean, during all its ontogeny.

(iv) southern and (v) deep assemblages, inhabiting the Portuguese shelf and upper slope. The geographical boundaries were drawn for each of the assemblages, the most important species were identified, and seasonal and temporal patterns were described regarding their persistence and the most important species therein.

The identification of such persistent communities appears as a natural first step towards the assessment of marine biological diversity at the regional scale of continental shelves. Accordingly, several methods have been developed for measuring biodiversity [see Gray (2000) for an extensive and updated review and comparative approach] and authors have also shown the usefulness of ecological indices in evaluating the general biological patterns of fish assemblages (e.g., Seldberry and Van Dolah, 1984; Barber *et al.*, 1997).

Searching for patterns in biodiversity indicators over time is a natural way to assess the status of marine communities, in particular those targeted by the fisheries. The traditional approach to time series analysis (e.g., auto-regressive integrated moving average models, ARIMA) requires a relatively high number of points in time and a stationary process (Murteira *et al.*, 1993). However, besides that long time series of diversity indices are seldom available, the stationary process assumption will not be met if overall shifts in time are likely to occur.

In such a context in Ecology, Zuur *et al.* (2003a) introduced in ecology a multivariate time-series analysis method called Dynamic Factor Analysis (DFA), which estimates underlying common trends from a set of different time series. This technique, appropriate to analyse short and non-stationary time series, was initially reported in psychology (Molenaar, 1985) and econometry (Harvey, 1989 *in* Zuur, 2003b) and has been used recently in fisheries and ecology (e.g., Zuur *et al.* 2003b, Zuur and Pierce, 2004; Erzini, 2005; Erzini *et al.* 2005).

In this paper we illustrate the application of DFA to analyse whether common patterns exist in time series of species diversity in the deep demersal assemblage off Portugal. Computations were conducted using R (R development core team, 2005) and Brodgar (Highland Statistics Ltd.).

2. Dynamic Factor Analysis

Dynamic Factor Analysis bears a relationship to Factor Analysis in the sense that new variables (factors) are created that, although in much lesser number than the original variables, are expected to explain the majority of the variation. However, in DFA, the new axes (common trends in time series) are smooth functions (curves) over time (Zuur *et al.*, 2003a). The smoothing is part of the first of two steps of the iterative algorithm used. In fact, the joint log-likelihood function and the trend components are unknown but mutually dependent, thus neither can be analytically obtained without knowing the other (Zuur and Pierce, 2004). This is solved using an iterative process conducted with a two-step EM (expectation-maximization) algorithm (details in Zuur *et al.*, 2003a). In the first step the conditional expectation of the log-likelihood function is calculated and in the second step the function is maximized over the entire set of iterations.

The model, with M common trends, is defined by,

$$(1) \quad y_t = \mathbf{c} + \mathbf{Z}\alpha_t + e_t$$

where y_t is a $N \times 1$ vector of the values of the N time series at time t , \mathbf{c} is a constant level parameter of dimensions $N \times 1$, \mathbf{Z} is a $N \times M$ matrix containing factor loadings, α_t is a vector containing the M common trends and e_t is a vector of error terms at time t . The trends at time t are modelled as a function of the trends at time $t-1$ plus an assumed normally distributed error term with mean zero and diagonal covariance matrix \mathbf{Q} .

Model (1) can also be modified to include the effect of explanatory variables in a way that the N time series are modelled as the effects of those variables, plus a linear combination of the common trends and an error term which accounts for the remaining non-explained variation (Zuur *et al.*, 2003b):

$$(2) \quad y_t = \mathbf{c} + \mathbf{Z}\alpha_t + \mathbf{D}x_t + e_t$$

where \mathbf{D} is an $N \times L$ matrix with regression coefficients, L being the number of explanatory variables and x_t is a vector containing the values of the explanatory variables at time t .

In Factor Analysis, some subjectivity exists on selecting the number of factors to retain in the model (Jolliffe, 1986). In DFA, measures of parsimony such as Akaike's information criterion (AIC) (Akaike, 1974) can be used objectively in selecting the number of existent common trends, the lowest value indicating the best solution. The AIC is defined as twice the difference between the log likelihood function (measure of fit) and the number of parameters (penalty). In cases where no substantial difference exists between AIC values of different models (different number of common trends) additional methods, such as plots of residuals and fitted values, can be used for validation. Loadings are used to assess both the relationships between variables and between variables and common trends.

3. The essay: trends of biodiversity in the deep assemblage

3.1. Data

Demersal research surveys have been regularly carried out by IPIMAR (Portuguese Institute for Fisheries and Sea Research) since 1979, covering the Portuguese continental coast from Caminha (N) to Vila Real de Santo António (SE). Data from 22 such surveys were used in the present work, covering the 1989 to 1999 time period, and taking place mostly in autumn (11 surveys) and summer (9) but also in the wintertime (2). The sampling scheme used up to 1989 by IPIMAR selected fishing stations randomly, within strata defined by depth and geographic area, but from 1990 onwards, a fixed grid of 97 sampling stations was adopted (Figure 1). For a complete description of the Portuguese demersal survey procedures and methodology refer to Cardador *et al.* (1997). Figure 1 also shows the geography of the demersal assemblages, i.e., the belonging of each sampling station to one of the five assemblages, as defined by Sousa *et al.* (2005). Different symbols are used for each assemblage and sampling stations are numbered (not shown graphically) from the north (sampling station 1) down to the extreme SE of Portugal (sampling station 97).

3.2. Measuring diversity

Diversity at each sampling station (Figure 1) was computed using the Shannon index (H),

$$(3) \quad H = - \sum_{i=1}^S \left[\left(\frac{n_i}{n} \right) \ln \left(\frac{n_i}{n} \right) \right]$$

where n_i is the number of individuals of the i -th species from a total of S , n is the total number of individuals ($n = \sum n_i$), and \ln is the natural logarithm (Ludwig and Reynolds, 1988).

3.3. Preliminaries

The H_{ij} entry in the working matrix for DFA, was the estimated diversity at sampling station j during survey i . The matrix should thus have 22 lines (surveys) by 97 columns (stations). However, throughout the 1989-1999 period, the survey sampling effort varied, with the number of stations effectively sampled ranging from 57 to 96, out of the 97 included in the original fixed grid (Sousa *et al.*, 2005). Consequently, of the possible 22 points in time, several sampling stations presented much fewer than that. Therefore, fifty two sampling stations were selected out of 97 (~54%) for having no more than one consecutive missing value. The majority of the excluded stations included 5 to 8 missings. DFA was applied to the diversity working matrix (H), with each sampling station regarded as a single time series. The main objective was to detect common trends in diversity within assemblages during the 1989-1999 time period. Diagonal covariance matrices were used to make up for the missing values at sampling stations (Zuur *et al.*, 2003a). Although the method was applied by assemblage, here we focus on the results obtained for the deep assemblage alone. This assemblage accounted for 18 sampling stations, after the exclusions above described.

The starting point of the analysis is a joint time plot with all the univariate (sampling stations) series (Figure 2) to depict the likely number of common trends. Standardization was conducted previously to all plots and DFA analyses.

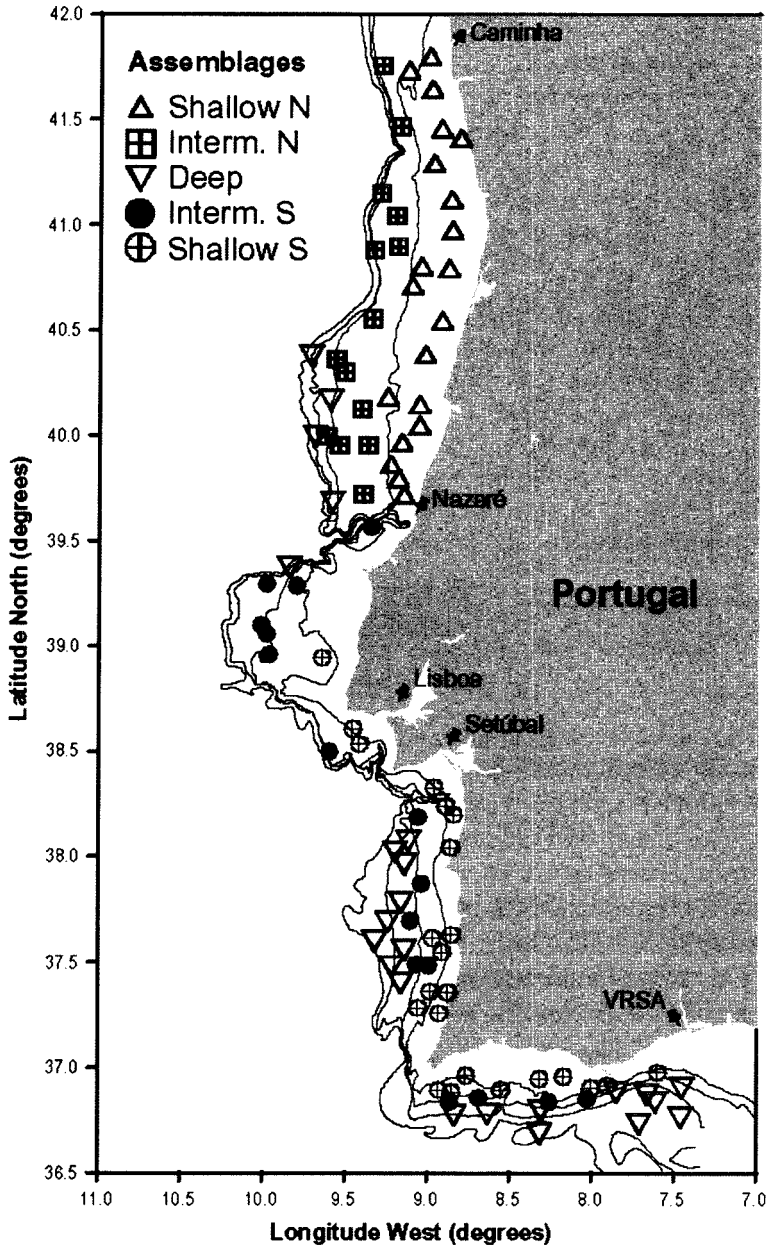


Figure 1 – Geography of Portuguese demersal assemblages, according to Sousa *et al.* (2005). The bathymetrics of 100, 200, 500 and 700 meters deep are shown as contour lines. All symbols are positioned at sampling stations and each symbol associates a station to one of five assemblages, as in the legend.

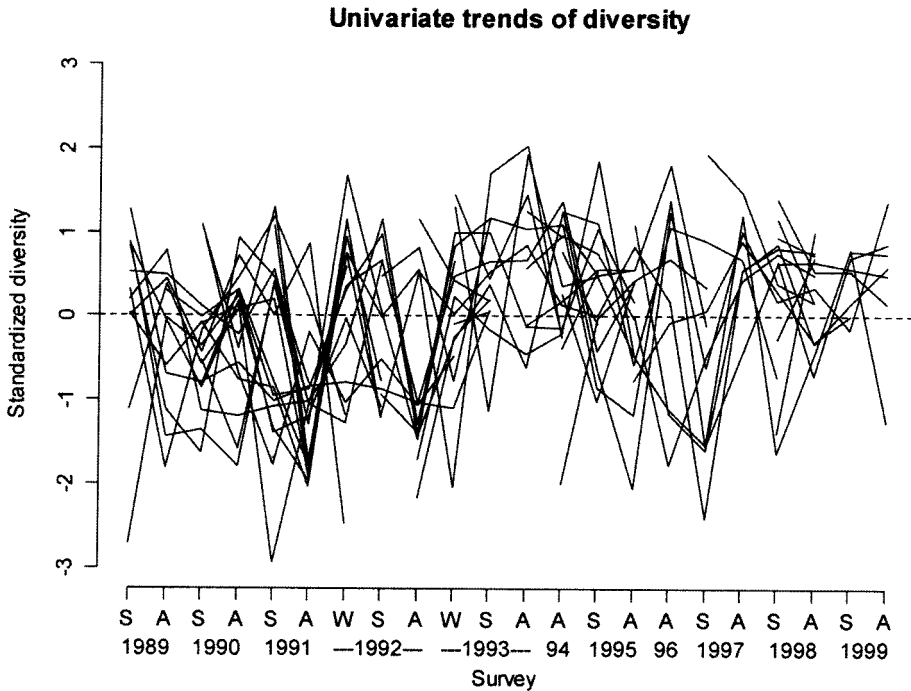


Figure 2 – Standardized diversity series measured at each of eighteen sampling stations of the deep assemblage, throughout 22 surveys. Surveys are referred to as “S” for summer, “A” for autumn and “W” for winter, and by the corresponding year, e.g., “1989”.

Figure 2 suggests that a relative amount of variability exists between sampling stations. However, a slight, overall increase and subsequent decrease is observed at the middle of the series, i.e., between the winter of 1993 and the summer of 1995.

3.4. DFA results and discussion

DFA was fitted assuming either of 1 to 5 common trends. The values of AIC were, respectively, 973, 963, 972, 975 and 992, thus suggesting a solution with two common trends (lowest AIC). These trends are presented in Figure 3a, whereas Figure 3b shows the bi-dimensional (common trends 1 × 2) plot of factor loadings.

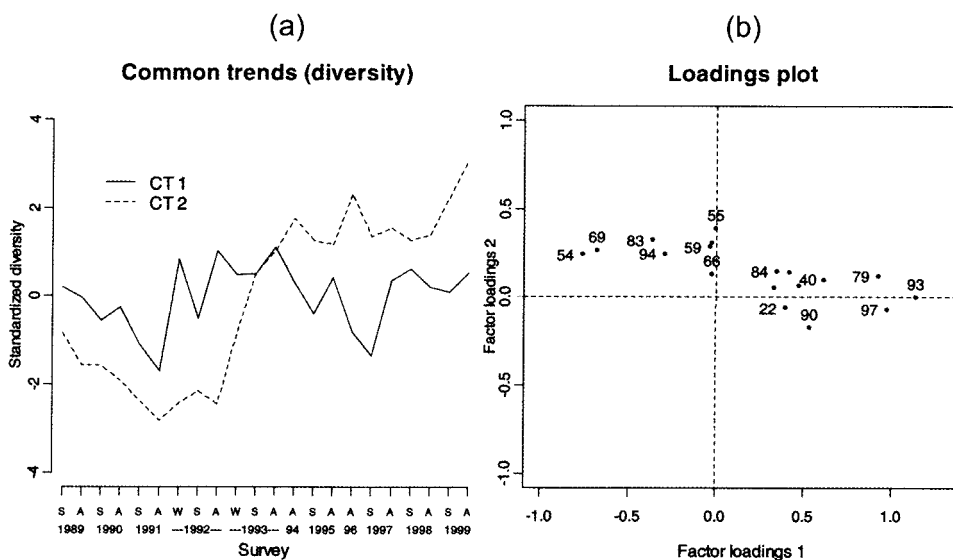
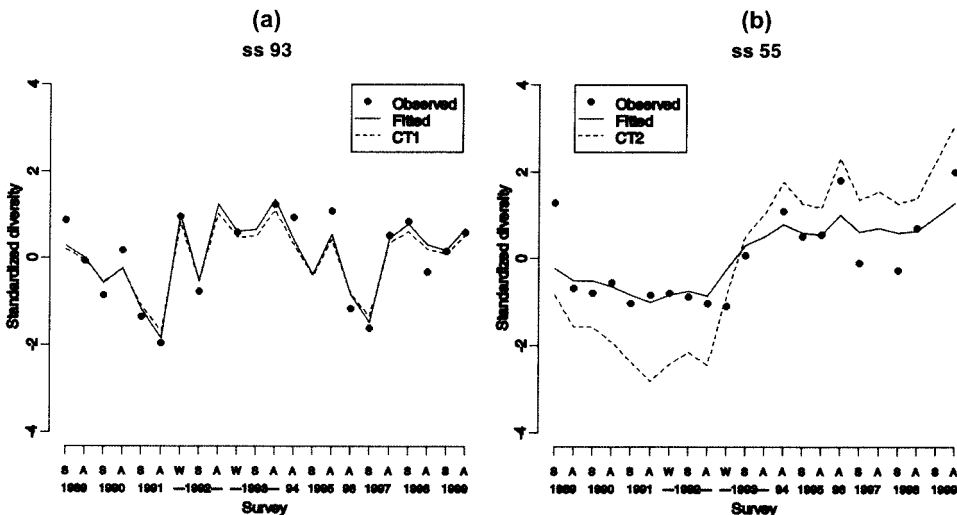


Figure 3 – Common trends estimated for diversity series of the deep demersal assemblage (a) and corresponding factor loadings with numbers indicating the sampling stations with loadings greater than 0.1 (b).

The majority of sampling stations are distributed along the common trend 1 factor (Figure 3b), while loadings to the second common trend are, overall, very small, i.e., very few univariate trends (sampling stations) can be depicted by common trend 2 (Figure 3a). The small importance of the second trend could have been anticipated from the small difference in the AIC criterion between solutions with one (AIC=973) and with two (AIC=963) common trends. Therefore, common trend 1 (CT1) is enough to describe the overall pattern of diversity within the individual series, since a substantial proportion of loadings are relatively high, i.e., greater than ~0.2. The analysis of the factor loadings, presented in Figure 3b, also indicate that the diversity trend in sampling station (ss) 93 will be better fitted by CT1, while a worst fit is expected for, e.g., ss 22. Difference in fitting individual sampling stations are expected since the solution accounts for a small number of common trends and, especially, because in this case the variability explained by the second CT is very low (Figure 3b). Figure 4 presents examples of better fits of individual series to CT1 (Figure 4a – ss 93) and to CT2 (Figure 4b – ss 55) and worse fits to CT1 (4c – ss 22 and 4d – ss 40).

Better fit (examples):



Worse fit (examples):

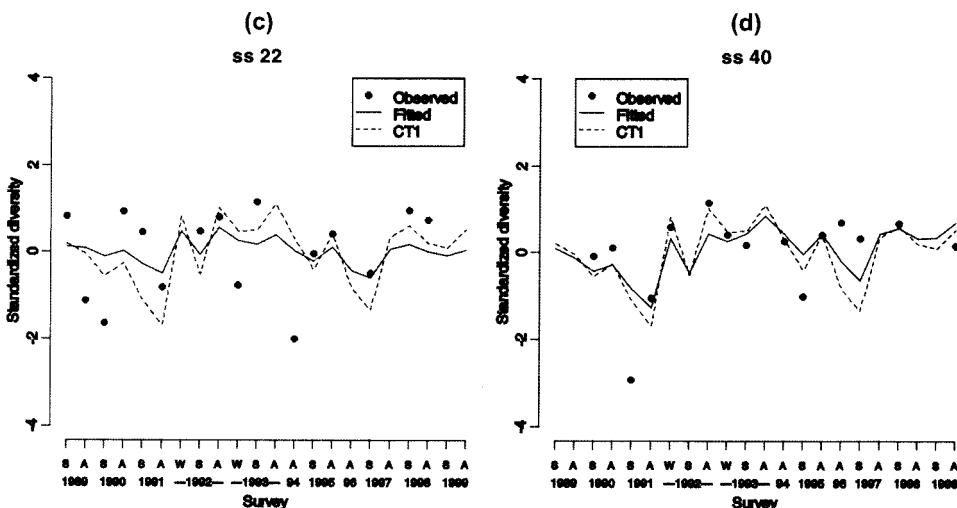


Figure 4 – Examples of better fits of individual series to CT1 (a – ss 93) and to CT2 (b – ss 55) and worse fits to CT1 (c – ss 22 and d – ss 40).

The overall pattern described by CT1 suggests that diversity was relatively higher in the middle nineties, from the winter of 1993 until the summer of 1995, with relatively low diversity before and after this period

(Figure 3a). Figure 5 overlaps CT1 with all univariate time series presented in Figure 2.

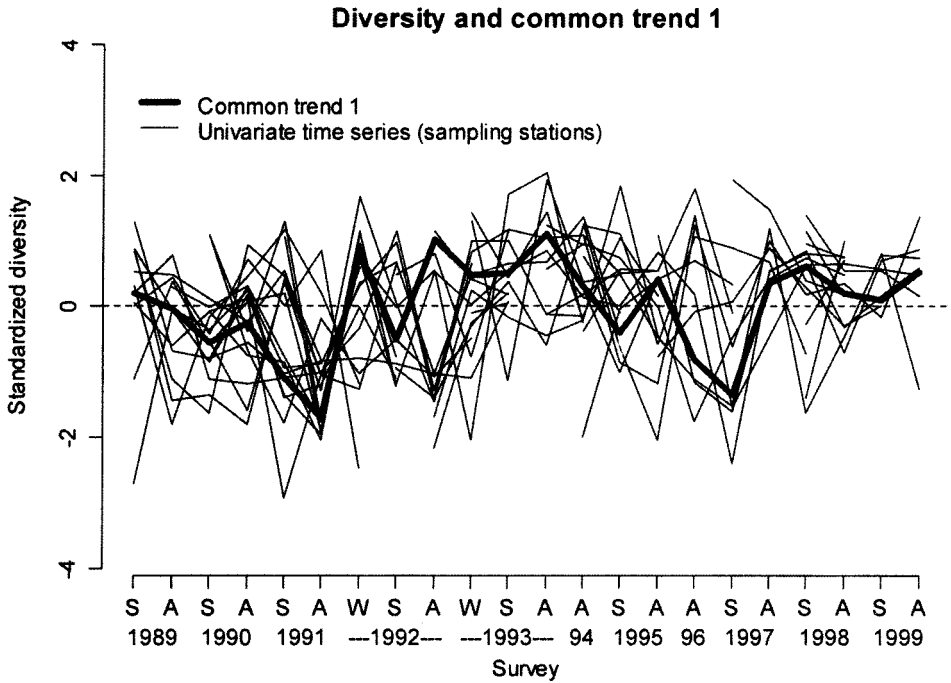


Figure 5 - Standardized diversity series and common trend 1 (overlapped, bold).

The trends of some sampling stations were quite different than CT1 (Figure 5). Such is the case of ss54 and ss69, with the highest negative loadings on CT1 (Figure 3b), due to a behaviour which was quite the opposite than CT1. Sampling stations 54 and 69 are, at the southwestern coast (Alentejo) of Portugal, whereas sampling stations 93, 79 and 97, that had maximum positive loadings on CT1 (Figure 3b), are located at the southern Portuguese margin (Algarve). It thus appears that diversity determinants somehow differed between these two deep areas. The Shannon index of diversity (Equation 3) depends both on the number of species in a sample, i.e., species richness (S), and on the relative proportion of each species. Low diversity occurs either when one species or a small group dominates the sample and/or the total number of species is low. For illustration purposes, consider Figure 6 where four scenarios are simulated, varying the number of species in a sample (S) and the relative proportion of each.

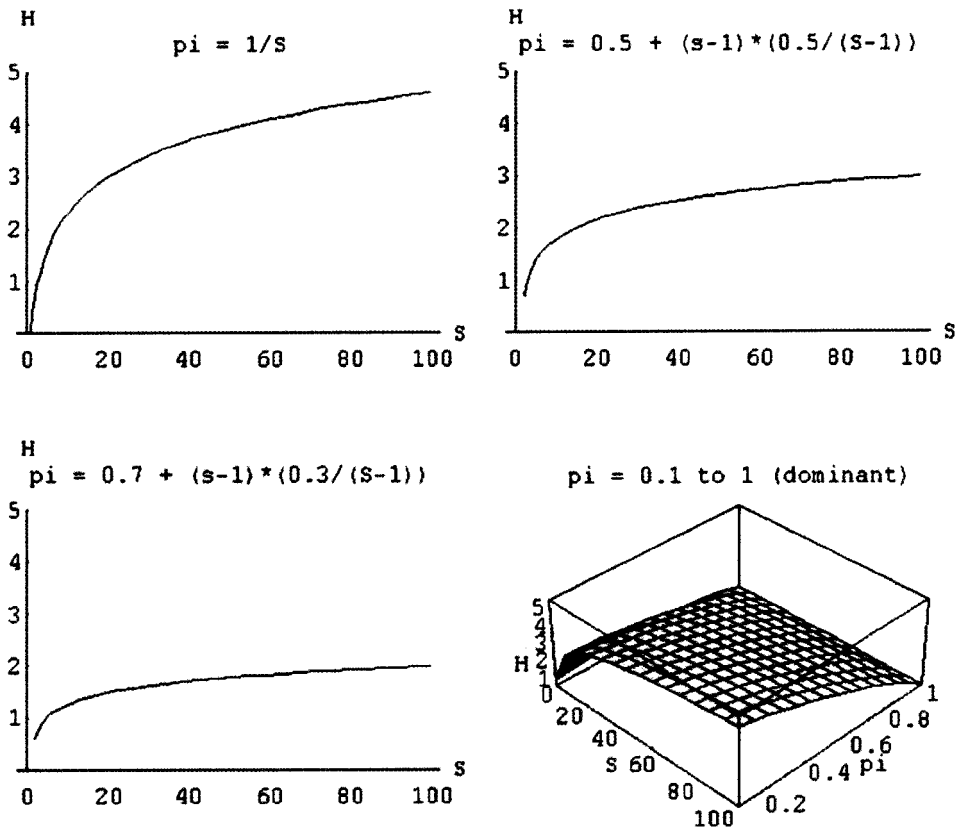


Figure 6 – Simulation of diversity (H) variation against richness, (number of different species, S), and some patterns of dominance, p_i being the proportion of abundance of species i . In the top-left dominance is fixed at minimum (0), i.e., all species occur in the same proportion ($p_i=1/S$) and H is only a function of S. In the top-right and bottom-left graphs, a dominant species accounts for, respectively, 50% and 70% of total abundance whereas the complementary 50% and 30% are equally distributed by S: the diversity curve is flattened by increasing dominance. The bottom-right graph generalizes this feature showing a surface of H values depending on S and on the degree of dominance ($p_i = 0.1$ to 1) of a given species.

Between the winter of 1993 and the summer of 1995 (higher diversity), the mean proportion of the most important species in the southern sampling stations (Algarve) never reached 50%, whereas in southwestern stations the corresponding mean proportion was, at least, 70%. Thus, higher dominance in this last group of stations induced lower diversity.

Figure 7 shows that the relative proportion of blue whiting in the assemblage has been consistently decreasing from the beginning of the series, as shown by the linear model adjusted. Boarfish increased its proportion from the mid nineties being the most important species by the end of the decade. This was accompanied by the increase of other low-abundant species, such as the crustaceans blue and red shrimp (*Aristeus antennatus*) and deep-water rose shrimp (*Parapenaeus longirostris*) (not presented graphically) both species having a very high market price. The rise of low-abundant species induces, consequently, higher diversity in the assemblage.

3.5. Concluding

No clear trends in diversity were detected along the nineties in the deep demersal assemblage. However, a relatively high diversity period was observed by the middle of the decade, on account of the increased abundance of species, other than the dominant one (blue whiting), in the assemblage. The relative abundance of blue whiting seems to have consistently decreased from the beginning of the nineties, but other species, like boarfish, might have occupied its niche, not allowing for an increase in overall diversity. The analysis of demersal species diversity through time could, however, be revisited in the future, eventually with the addition of explanatory variables to DFA, such as fishing effort from the commercial fleet or other environmental factors suspected to have an impact on the marine community.

Acknowledgements

Financial support for P. Sousa has been provided by the Fundação para a Ciência e Tecnologia (Programa PRAXIS XXI), Lisbon, Portugal, in a PhD scholarship programme. We are indebted to the IPIMAR staff, as their overall planning and sustained fieldwork made this work possible. In particular, we express our gratitude to Dr. Fátima Cardador (IPIMAR) for assistance with data handling. We also acknowledge the helpful comments of an anonymous referee.

References

- Adjeroud, M., Letourneur, Y., Porcher, M. and Salvat, B., 1998. Factors influencing spatial distribution of fish communities on a fringing reef at Mauritius, S.W. Indian Ocean. *Environmental Biology of Fishes*, **53**: 169-182.
- Akaike, H., 1974. A new look at the statistical model identification. *IEEE Trans. Auto. Cntl.*, AC-19, 716-723.
- Araújo, F. G. and Costa de Azevedo, M. C., 2001. Assemblages of Southeast-South Brazilian Coastal Systems Based on the Distribution of Fishes. *Estuarine, Coastal and Shelf Science*, **52**: 729-738.
- Ayvazian, S. G. and Hyndes, G. A., 1995. Surf-zone fish assemblages in South-Western Australia: do adjacent nearshore habitats and the warm Leeuwin Current influence the characteristics of the fish fauna? *Marine Biology*, **122**: 527-536.
- Barber, W. E., Smith, R. L., Vallarino, M. and Meyer, R. M., 1997. Demersal fish assemblages of the northeastern Chukchi sea, Alaska. *Fishery Bulletin*, **95**: 195-209.
- Bianchi, G., 1992. Demersal assemblages of the continental shelf and upper slope of Angola. *Marine Ecology Progress Series*, **81**: 101-120.
- Cabral, H. N. and Murta A. G., 2002. The diet of blue whiting, hake, horse mackerel and mackerel off Portugal. *Journal of Applied Ichthyology*, **18**, 14-23.
- Cardador, F., Sánchez, F., Pereiro, F. J., Borges, M. F., Caramelo, A. M., Azevedo, M., Silva, A., Pérez, N., Martins, M. M., Olaso, I., Pestana, G., Trujillo, V. and Fernandez, A., 1997. Groundfish surveys in the Atlantic Iberian waters (ICES divisions VIIIc and IXa): history and perspectives. *ICES C.M.*, (CM 1997/Y:08): 29 p.
- Coles, S. L. and Tarr, A. B., 1990. Reef fish assemblages in the western arabian gulf: a geographically isolated population in an extreme environment. *Bulletin of Marine Science*, **47** (3): 696-720.
- Day, D. S. and Percy, W. G., 1968. Species Associations of Benthic Fishes on the Continental Shelf and Slope off Oregon. *Journal of the Fisheries Research Board of Canada*, **25** (12): 2665-2675.

- Erzini, K., 2005. Trends in NE Atlantic landings (southern Portugal): identifying the relative importance of fisheries and environmental variables. *Fisheries Oceanography*, **14** (2): 1-15.
- Erzini, K., Inejih, C.A.O. and Stobberup, K.A., 2005. An application of two techniques for the analysis of short, multivariate non-stationary time-series of Mauritanian trawl survey data. *ICES Journal of Marine Science*, **62** (3): 353-359.
- Fager, E. W., 1957. Determination and analysis of recurrent groups. *Ecology*, **38** (4): 586-595.
- Fager, E. W. and Longhurst, A. R., 1968. Recurrent Group Analysis of Species Assemblages of Demersal Fish in the Gulf of Guinea. *Journal of the Fisheries Research Board of Canada*, **25** (7): 1405-1421.
- Gabriel, W. L., 1989. Persistence in Northwestern Demersal Fish Assemblages. *NAFO SCR Doc.*, **89/77**: 17 pp.
- Galzin, R., 1987. Structure of fish communities of French Polynesian coral reefs. I. Spatial scales. *Marine Ecology Progress Series*, **41**: 129-136.
- Gomes, M. C., 1993. *Predictions Under Uncertainty, Fish Assemblages and Food Webs on the Grand Banks of Newfoundland*. Institute of Social and Economical Research. Social and Economical Studies, 51. Memorial University of Newfoundland, Newfoundland, Canada. 205 p.
- Gray, J. S., 2000. The measurement of marine species diversity, with an application to the benthic fauna of the Norwegian continental shelf. *Journal of Experimental Marine Biology and Ecology*, **250**: 23-49.
- Jennings, S., Greenstreet, S. P. R., Hill, L., Piet, G. J., Pinnegar, J. K. and Warr, K. J., 2002. Long-term trends in the trophic structure of the North Sea fish community: evidence from stable-isotope analysis, size-spectra and community metrics. *Marine Biology* **141** (6): 1085-1097.
- Jin, X., 1995. Seasonal changes of the demersal fish community of the Yellow Sea. *Asian Fisheries Science*, **8**: 177-190.
- Jolliffe, I. T., 1986. *Principal Component Analysis*. Springer-Verlag. New York. 271 p.
- Ludwig, J. A. and Reynolds, J. F., 1988. *Statistical Ecology - a primer on methods and computing*. John Wiley & Sons. New York. 337 p.

- Mahon, R. and Smith, R. W., 1989. Demersal Fish Assemblages on the Scotian Shelf, Northwest Atlantic: Spatial Distribution and Persistence. *Canadian Journal of Fisheries and Aquatic Sciences*, **46**: 134-152.
- Molenaar, P. C., 1985. A Dynamic Factor Model for the Analysis of Multivariate Time Series. *Psychometrika*, **57**: 333-349.
- Murteira, B.J.F, Müller, D.A. and Turkman, K.F., 1993. *Análise de Sucessões Cronológicas*. McGraw-Hill Portugal. 300 p.
- 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>.
- Russell, I. A., 1997. Spatial variation in the structure of fish assemblages in the Vaalbos National Park, South Africa. *Koedoe*, **40** (1): 113-123.
- Sedberry, G. R. and Van Dolah, R. F., 1994. Demersal fish assemblages associated with hard bottom habitat in the South Atlantic Bight of the U.S.A. *Environmental Biology of Fishes*, **11** (4): 241-258.
- Silva, A., Azevedo, M., Cabral, H., Machado, P., Murta, A. and Silva, M. A., 1997. Blue whiting (*Micromesistius poutassou*) as a forage fish in Portuguese waters. In Proceedings of the International Symposium on the role of Forage Fishes in Marine Ecosystems, pp. 127-146. Alaska Sea Grant College Program, USA.
- Sousa, P., Azevedo, M. and Gomes, M.C., 2005. Demersal assemblages off Portugal: mapping, seasonal, and temporal patterns. *Fisheries Research*, **75**: 120-137.
- Trondsen, T., 1998. Blue whiting surimi: new perspectives on the market value. *Fisheries Research*, **34**: 1-15.
- Tyler, A. V., Gabriel, W. L. and Overholtz, W. J., 1982. Adaptive management based on structure of fish assemblages of northern continental shelves. In Mercer, M. C.(ed.). *Multispecies approach to fisheries management advice*, pp. 149-156. Canadian Special Publication of Fisheries and Aquatic Sciences, **59**.
- Varas, E. and Ojeda, F. P., 1990. Intertidal fish assemblages of the central Chilean coast: diversity, abundance and trophic patterns. *Revista de Biología Marina*, **25** (2): 59-70.

- Zuur, A. F., Fryer, R. J., Jolliffe, I. T., Dekker, R. and Beukema, J. J., 2003a. Estimating common trends in multivariate time series using dynamic factor analysis. *Environmetrics*, **14** (7): 665-685.
- Zuur, A. F., Tuck, I. D. and Bailey, N., 2003b. Dynamic factor analysis to estimate common trends in fisheries time series. *Canadian Journal of Fisheries and Aquatic Sciences*, **60**: 542-552.
- Zuur, A. F. and Pierce G. J., 2004. Common trends in Northeast Atlantic Squid time series. *Journal of Sea Research*, **52**: 57-72.