Indices for capturing spatial pattern and change across years of fish population: an application on European hake (*Merluccius merluccius*) in the Bay of Biscay

Mathieu Woillez[a,*], Pierre Petitgas[b], Jacques Rivoirard[a], Jean Charles Poulard[b] and Nicolas Bez[c]

[a] *Centre de Géostatistique de l'Ecole des Mines de Paris, 35 Rue St Honoré, 77300 Fontainebleau, France*
[b] *IFREMER, Laboratoire d'Ecologie Halieutique, rue de l'île d'Yeu, BP21105, F-44311 cedex 03, Nantes, France*
[c] *IRD, Centre de Recherche Halieutique Méditerranéenne et Tropicale, rue Jean Monnet, BP171, 34203 Sète Cedex, France*

Abstract

Collapse of important fish stocks have revealed that fish stock assessment based on fishery landings at age suffer from a number of limitations. Probably the most serious limitation is that the indication of population collapse is only perceived very late with such data. The present work aims at looking for indices, based on research survey data only, that could capture the spatial pattern of populations and changes across years. A series of candidate statistical indices has been selected or developed for that purpose. In order to handle diffuse population limits, these indices have been designed so that they do not depend on an arbitrary delineation of the domain. The spatial pattern of fish population is described by a variety of indices characterizing location (centre of gravity, inertia and anisotropy, spatial patches), spatial extension (positive area, spreading and equivalent area) and microstructure. Collocation between different ages and years is summarized by a global index of collocation. The indices are estimated on the hake data series of bottom trawl surveys performed by IFREMER from 1987 to 2003 in the Bay of Biscay. Behaviour of the spatial indices is analysed, including interaction between the different indices, their relationship with abundance and age. The ability of the indices to capture patterns and provide early warning on incoming changes is discussed.

Keywords: Spatial indices; European Hake; Geostatistics; Bay of Biscay.

1. Introduction

Fishery-independent data such as research surveys data are often misused; they are aggregated to form indices of relative abundance for stock assessments purposes and all spatial information is lost. However, the spatial dynamic of a stock *i.e.* the way that fish occupy the space from a cohort to another, is a fundamental component for a good understanding and management of the stock. The collapse of northern cod, *Gadus morhua*, off Newfoundland and Labrador was associated with clearly defined spatial and temporal changes in density and biomass. Although this fishery was followed, this change was perceived very late with the methods used for the stock assessment.

In this study, the idea is to look for indices, based on research survey data only, that could capture the spatial patterns of fish resources and see if a change can be linked to a change of abundance. Thus, spatial indices have been developed on raw data of scientific surveys (fish densities measured locally by trawl haul). For that purpose, data from IFREMER (Institut Français de Recherche pour l'Exploitation de la MER) surveys on European hake have been used. The behaviour of the spatial indices is analysed through their pairwise interaction and their relationship with abundance and age.

* Corresponding author. Tel: + 33 (0) 1 64 69 47 76; fax: + 33 (0) 1 64 69 47 05. *E-mail address:* mathieu.woillez@ensmp.fr

The ability of these indices to capture spatial patterns and provide early warning on incoming changes is discussed.

2. Materials and methods

2.1 Data

The European hake (*Merluccius merlucius*) of Bay of Biscay is a commercially important species and biologically well known although there are some uncertainties about is growth. In addition to its economic importance, this species drew the attention of the managers and the researchers because of the problem raised by the simultaneous capture of nephrops and juvenile of this species.

Since 1987, groundfish surveys are carried out in the Bay of Biscay in autumn with the R/V "Thalassa" (the EVHOE survey). During each cruise, IFREMER samples between 139 and 70 stations, distributed according a stratified (in latitude and depth) random sampling design located over the shelf between 15 and 600 m of depth. Fish were sampled with a 'Grande Ouverture Verticale' GOV bottom trawl during day time. Haul duration lasted 30 min with a towing speed of 4 knots. All hake caught were counted, sexed, weighted, and measured on board. Age was estimated back to the laboratory by reading otoliths. Thus, data used in this study are densities of hake, expressed in a number of individuals per hour of trawling (ind/h) and disaggregated by age classes over the period 1987 to 2003 (gaps in 1991, 1993 and 1996). For the calculation of index of abundance, one will use the fact that the average size of the support of the densities (i.e. the trawling surface in a half hour) is of 0.02 square nautical miles (nm²).

2.2 Spatial indices

In the context of populations with diffuse limits, spatial indicators have been built to avoid the problem of the delineation of the area of presence, which may be variable from one year to another. This is achieves by letting zero sample values having a null contribution to these indices (Bez and Rivoirard, 2001).

Centre of gravity and inertia

The spatial distribution of population can be easily summarized by tools such as centre of gravity and inertia (Bez, 1997). The centre of gravity (*CG*) is the mean location of the population, also the mean location of an individual taken at random in the field, and the inertia is the mean square distance between such an individual and the centre of gravity.

Let x be a point in 2D (short for usual 2D notation (x,y)), $z(x)$ the fish density of population at location x and the total abundance of the population is :

$$Q = \int z(x)dx$$

and the probability density function of the location x_I of a random individual I is $\frac{z(x)}{Q}$.

Then, the centre of gravity is:

$$CG = E(x_I) = \int x \frac{z(x)}{Q} dx = \frac{\int x.z(x)dx}{\int z(x)dx}$$

and the inertia is:

$$Inertia = Var(x_I) = \frac{\int (x - CG)^2 .z(x)dx}{\int z(x)dx}$$

In the case of an irregular sampling, surfaces of influence affected to samples are used as weighted factors. Practically, from sample values z_i at locations x_i, with surfaces of influence s_i, we have:

$$CG = \sum_{i=1}^{N} x_i.s_i z_i \Big/ \sum_{i=1}^{N} s_i z_i$$

$$Inertia = \sum_{i=1}^{N} (x_i - CG)^2 . s_i z_i \; / \sum_{i=1}^{N} s_i z_i$$

Global index of collocation

The global index of collocation looks how close two populations are geographically, by comparing the distance between their *CG* to the mean distance between individuals taken at random and independently from each population (Bez and Rivoirard, 2000). Let us consider two populations, $z_1(x)$ and $z_2(x)$, their density at point x, ΔCG the distance between their centres of gravity and I_1 and I_2 their inertia. Then the global index of collocation is:

$$I_g = 1 - \frac{\Delta CG^2}{\Delta CG^2 + I_1 + I_2}$$

I_g = 1 – ratio between the square distance between the two centres of gravity and the square distance between individuals from the two different populations. It is a spatial statistics that ranges between 0, in the extreme case where each population is concentrated on a single but different location (inertia equal to 0), and 1, when the two centres of gravity coincide.

Anisotropy

In 2D, the total inertia of a population can be decomposed on its two principal axes, orthogonal to each other, and explaining respectively the maximum and the minimum of the overall inertia. The square root of the inertia along a given axis gives the standard deviation of the projection of the location of the population along this axis. An anisotropy exists when there is a difference in inertia between the directions. This is summarized by the square root ratio between maximum and minimum of inertia.

$$A = \sqrt{\frac{I \max}{I \min}}$$

Number of spatial patches

The spatial distribution of a fish population in a given area may not be homogeneous. Local aggregations of fish, i.e. spatial patches which are bigger than a fish school, may be present. To identify spatial patches, an algorithm has been written, based on a distance limit to attribute sample values to patches. The algorithm starts from the richest value and considers each sample in decreasing order. The richest value forms the first patch. Then, the current sample value is attributed to the nearest patch, provided that the distance to its centre of gravity is smaller than the chosen distance limit. Otherwise, the current sample value defines a new patch. The indicator obtains is, for each survey, the number of patches whose biomasses is superior to 10 % of overall biomass. The chosen distance limit has been fixed at 100 nm for the hake.

Positive area

An area of influence is attributed to each sample. The positive area is the sum of the surfaces of influence of the positive values.

$$PA = \sum_i s_i 1_{z_i \geq 0}$$

Spreading Area

This index comes from the selectivity curves which have been developed in mining geostatistics to characterize probability distributions and their dispersion (Matheron, 1981). These curves are in particular useful to handle the effect of the support on which the variable is measured or defined. They have been used in fisheries to look at the aggregation of values when the abundance changes (Petitgas, 1998).

Let Q be the total abundance (in number of individuals), $Q(T)$ the abundance corresponding to the area T occupied by the highest density values (in nm² or percentage of a maximal area). A selectivity or aggregation index can be defined as:

$$2\int_0^1 \left(\frac{Q(T)}{Q} - T \right) dT$$

when expressing *T* as a proportion of a maximal area. However this index is dependent on this maximal area and on the zero values. In order to have statistics in which the contribution of zero sample values is null, curves have been reversed bottom up, leading to the spreading area. The spreading area is then proposed instead, equal to:

$$2\int_0^1 \frac{R(T)}{Q} dT$$

with *R(T) = Q-Q(T)* being the abundance remaining off the richest area *T*.

Equivalent area

The transitive geostatistical approach (Matheron, 1970) can be used to describe the spatial distribution of a fish population when this includes a few large density values, and when delimitating a domain with homogeneous variations is difficult (Bez and al, 1995, Bez. and al., 1997). The (transitive) covariogram, function of the distance between two locations, is a tool for description of the spatial structure and can be used for mapping:

$$g(h) = \int z(x)z(x+h)dx$$

Here, the equivalent area is defined as the integral range of the covariogram:

$$EA = \frac{\int g(h)dh}{g(0)} = \frac{Q^2}{g(0)}$$

As we can write:

$$EA = \frac{Q}{\int z(x)\frac{z(x)}{Q}dx}$$

And it represents the area that would be covered by the population, if all individuals had the same density, equal to the mean density per individual.

Microstructure index

This index comes from the covariogram. It measures the relative decrease between distance h=0 and a distance h0 chosen to represent the mean lag between samples. It measures the relative importance of the structural components with scale smaller than h0 (including random noise):

$$MI = \frac{(g(0) - g(h0))}{g(0)}$$

Some of the indices are spatial statistics in the sense that their values would be changed by inverting densities between two locations. This is the case of the centre of gravity, the inertia, the global index of collocation, the anisotropy and the microstructure index. The abundance and the other indices (positive area, spreading area and equivalent area) depend only on the statistical distribution of values over the region.

For comparison between surveys, we have decided to fix the maximal area of presence which has been surveyed along the time series (polygon restriction); the surfaces of influence are computed within this delineated domain from year to year. This domain limits the surface of influence of positive sample values when the spatial population is not closed by zero values. Using such weighted factors when computing spatial indices such as centre of gravity, inertia, and global index of collocation, reduces the effects of the variations in sampled area along a time series. So the centre of gravity of trawl haul is fixed over the period.

2.3 Data analysis

Bubbles plots of fish densities samples, plots of the time series of spatial indices and abundance, plots of centre of gravity and tables of global index of collocation have been used to explore the data. Then we have been interested in examining the relationship among variables by looking at pairwise scatter plots of the spatial indices, and also scatter plots of the log10 of the abundance with each spatial index. The level of abundance of each cohort has also been examined by plotting the log10 of abundance for each age in order to identify trend or change in the abundance over the studied period.

2 Results

For the fourteen studied years (years 1991, 1993 and 1996 are missing), spatial indices and abundance have been computed. The spatial occupation of hake along the age is described by the centres of gravity (Fig. 1); the young hake present centres of gravity that are rather stable over the time. This comes from the fact that they have a preferential habitat (nourricery grounds) around a place named “la grande vasière” (Dardignac, 1988). Distribution of centres of gravity of age 1 and age 2 are also relatively stable. The older ages have a fluctuating average position which tends to move away from the coast. However if one is interested by displaying the age classes per year or per cohort, the previous schema is not so constant (Fig. 2-3). The global index of collocation completes this approach by quantifying the observed charts. Thus, the years 1988, 2000 and 2002 present three different situations (Fig. 3): one where the ages are close together, another where the young ages are more distant compared to the old ages and a last one where most ages are distant from each other. In addition, global index of collocation allows to identify particular years. For example, age 0 in year 2000 is distant from the other ages 0 (Fig. 4). Indeed, the global index of collocation calculated between age 0 of this year and the others always presents lower values. It appears that its centre of gravity have a position more northern than the others.

For the other indicators, the analysis of the time series allows to identify the main trend (increase, decrease, stability or none) and the years which are different. The number of patch increases with the age. There are 1 to 2 patches for age 0, 1 to 3 patches for age 1, then 2 to 4 patches for the older ages except 1997 (Fig. 5). The number of spatial patches at the age 0 shows three years (1987, 2000 and 2002) where there is only one patch instead of two. Fig. 6 shows the patches surimposed to the bubble plots of hake densities for two typically years of age 0. The 2nd patch corresponds to a southern part of population, absent for the year 2000. It can be interpreted as a bad recruitment on this region of the Bay of Biscay and it explains the northern position of the centre of gravity. For age 5, the single patch above 10% of biomass for year 1997 is due to uncommon cluster of high values. In addition, 400 to 600 m depth strata, which are the preferential habitat of age 5, have not been sampled for years 1997 and 1998. So this index may not be relevant here. The anisotropy is a relatively stable indicator within age classes (Fig. 7) it decreases for the ages 0 to 3 and then it increases for age 4 and 5. The age 5 is mostly concentrated in deep zone along the slope, which explains the large values of anisotropy. The exceptionally low anisotropy for 1987 age 0 is due to the absence of patch on south and larger values being more regularly concentrated into a circle than the other years without the south patch (Fig. 8). One can note that the spreading area decreases in the recent years for the old ages (Fig. 9) where the abundance is mainly due to one or few high values (Fig. 10).

The next step has been to examine the level of abundance of the cohorts (Fig. 11). It seems that the cohorts have comparable levels of abundance. For each age, the abundance varies by a factor 10 (one unit in log10). However there is no cohort being lower or higher than the others at all ages. Most of the time, there is only one age for each cohort which is far from the mean. Some curves also intersect with higher values for the young ages and lower values for the old ages. So the follow-up of the cohorts did not show outliers, but raises the problem of sampling comparison.

The most interesting pairwise scatter plots between the spatial indices have been plotted on Fig. 12. First, we have the centre of gravity which moves toward the West with the age and also North for some age 5 (Fig. 12a). Indeed, for the age 5, the ordinate of centre of gravity is sometimes larger. It is explained by large value of densities in the North West. One can note that inertia and number of patches move in the same sense; there is more patches when the inertia is larger (Fig. 12b). In addition, large inertia corresponds to old ages which are mostly on the West side of the studied zone (Fig. 12c). Pairwise scatter plot between positive area and inertia is very interesting (Fig 12d). The positive area decreases with the age whereas the inertia raises. So, age 4 and 5 which have a great inertia with a small positive area are opposed to the young ages which have a great positive area and a small inertia. Otherwise, positive area goes in the sense of spreading and equivalent area (Fig 12e). Moreover, spreading area and equivalent area are close to each other (Fig. 12f). Indeed, the value of these area indices (*i.e.* positive area, spreading area and equivalent area) is low for the old ages and large for the young. To summarize, young hake are concentrated on the East side of the studied zone, and they usually present two patches for age 0. Then, hake move offshore and scatter over a larger habitat with the age, but with a reduced density.

Scatter plot of spatial indices and log10 of the abundance have been examined and are given on Fig. 13. The most interesting plot is the one with the positive area (Fig. 13a). There is a strong link between these two variables. Moreover, this link remains at age except for the age 0. One can note that there is an overlap between ages in the continuity of the observed link. Age could not be the best discriminator; investing such relationship with the length could be interesting. Indeed, there is a problem of interpretation of the otoliths. Studies are in development and hake have been tagged. Links with abundance at age were not found with the other spatial indices, except for the spreading area at age 4 and age 5 but to a smaller extent (Fig. 13f). Some graphs do not bring more than the previous pairwise scatter plot. Fig.13b shows that the inertia decreases with the abundance. Age 0, 4 and 5 present large values of anisotropy (Fig. 13c). The number of patches increases when the abundance decreases (Fig. 13d). The centre of gravity moves West when the abundance decreases (Fig. 13g), whereas the ordinate of the centre of gravity is rather stable with abundance, except for some values (Fig.13h). One can also observe that the microstructure index is rather stable with the abundance (Fig. 13i). Finally, abundance is not linked to the majority of indices except the positive area, which might be very interesting because of its robustness to the high values.

Decrease of spreading area in the recent years for age 5 has been previously noted. Positive area, which have a similar behaviour, also decreases with abundance on age 5 along the time series. The question is to see if we can predict such levels of abundance. We are also looking for low values of abundance of age 5 in the case of age 4 would not be low in abundance, but the positive area would be. However no result appears in that sense. Indeed, it goes in the opposite direction of the observed relationship between abundance and the positive area. So explanations should be search with the other indices. But it appears that other indices did not clearly explain the low values of abundance for age 5.

3 Conclusion

Spatial indices that have been selected or developed in this work allow us to draw a rather complete situation of the spatial pattern of the European hake. We observed that the older ages of the population moves toward the west of the Bay of Biscay (centre of gravity). However, when the age increases, the population is more scattered (The number of patch increases and the inertia is larger) and it occupies a smaller area but on a larger space (the positive area is smaller). Moreover, at fixed age, the abundance raises with the positive area. One can remark that spreading and equivalent area are relatively close to each other and they go in the sense of the positive area, which appears to be the most interesting index of these. For the others indices, the microstructure index is stable for all ages and the anisotropy is larger for the age 0 and 5.

The most interesting result is the link between abundance and positive area, with continuity between successive ages. It would be interesting to see if the partial overlapping between different ages is not reduced when looking at length, which would then explain more closely the variation of abundance and positive area.

We have not been able to see if these indices could produce an early warning in incoming change. But these indices still remain interesting in helping to make the link between abundance and spatial patterns. More than searching a relationship with the present abundance, which reveals to be irrelevant to predict a change, it is better to look for relationship differed in time to identify such early warning. Looking at other data sets could improve our knowledge and especially for a stock with a clearly identified decline.

Finally, resampling should be undertaken to test the sensibility of these indices to high values and their robustness in general.

4 Acknowledgement

These developments have been done within the European program FISBOAT (Fisheries Independent Survey Based Operational Assessment Tools).

5 References

Bez, N., and Rivoirard, J. 2001. Transitive geostatistics to characterize spatial aggregation with diffuse limits: an application on mackerel ichtyoplankton. Fish. Res. 50: 41-58.

Bez, N., and Rivoirard, J. 2000. MS Indices of collocation between populations. In: Chekley, D.M., J.R., Hunter, L. Motos, and C.D. van der Lingen (eds). Report of a workshop on the use of Continuous Underway Fish Egg Sampler (CUFES) for mapping spawning habitat of pelagic fish. GLOBEC Report 14, 1-65.

Bez, N., Rivoirard, J., Guiblin, Ph., and Walsh, M., 1997. Covariogram and related tools for structural analysis of fish survey data. E.Y. Baafi and N.A. Schofield (eds), Geostatistics Wollongong`96, Volume 2, 1316-1327.

Bez, N., 1997. Statistiques individuelles et géostatistique transitive en écologie halieutique. Thèse Dr. En géostatistique. ENSMP, France, 276p.

Bez, N., Rivoirard, J., and Poulard, J.C. 1995. Approche transitive et densité de poissons. Compte rendu des journées de géostatistique, 15-16 juin 1995, Fontainebleau, France. Cah. Géostat. 5: 161-177.

Dardignac, J., 1988. Les pêcheries du Golfe de Gascogne – Bilan des connaissances. Rapports scientifiques et techniques de l'IFREMER, n) 9.

Petitgas P.1998. Biomass dependent dynamics of fish spatial distributions characterized by geostatistical aggregation curves. ICES J. Mar. Sci., 55, 443-453.

Matheron, G., 1981. La sélectivité des distributions. Note N-686. Rapport du CGMM, ENSMP, Fontainebleau, France, 45p.

Matheron, G., 1970. La théorie des variables régionalisées et ses applications. Les cahiers du Centre de Morphologie Mathématique, Fascicule 5, ENSMP, Fontainebleau, France, 212p.

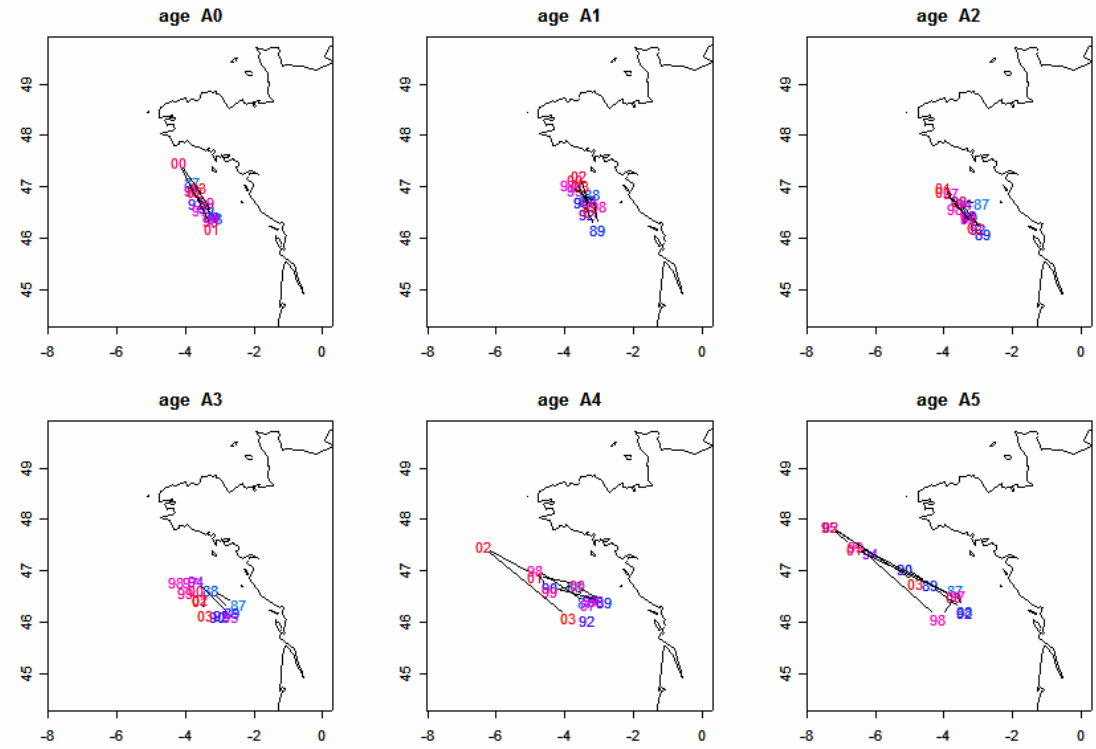

Fig. 1. Centre of gravity of the location of hake densities per age.

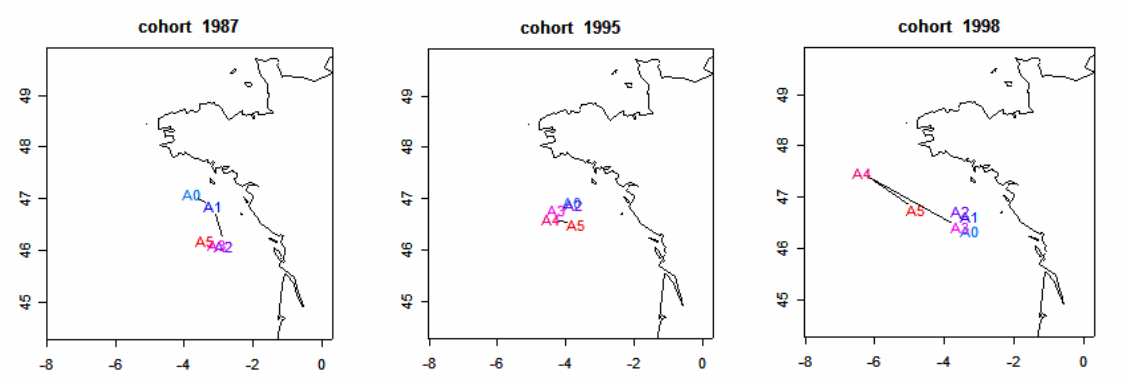

Fig. 2. Centre of gravity of the location of hake densities per cohort.

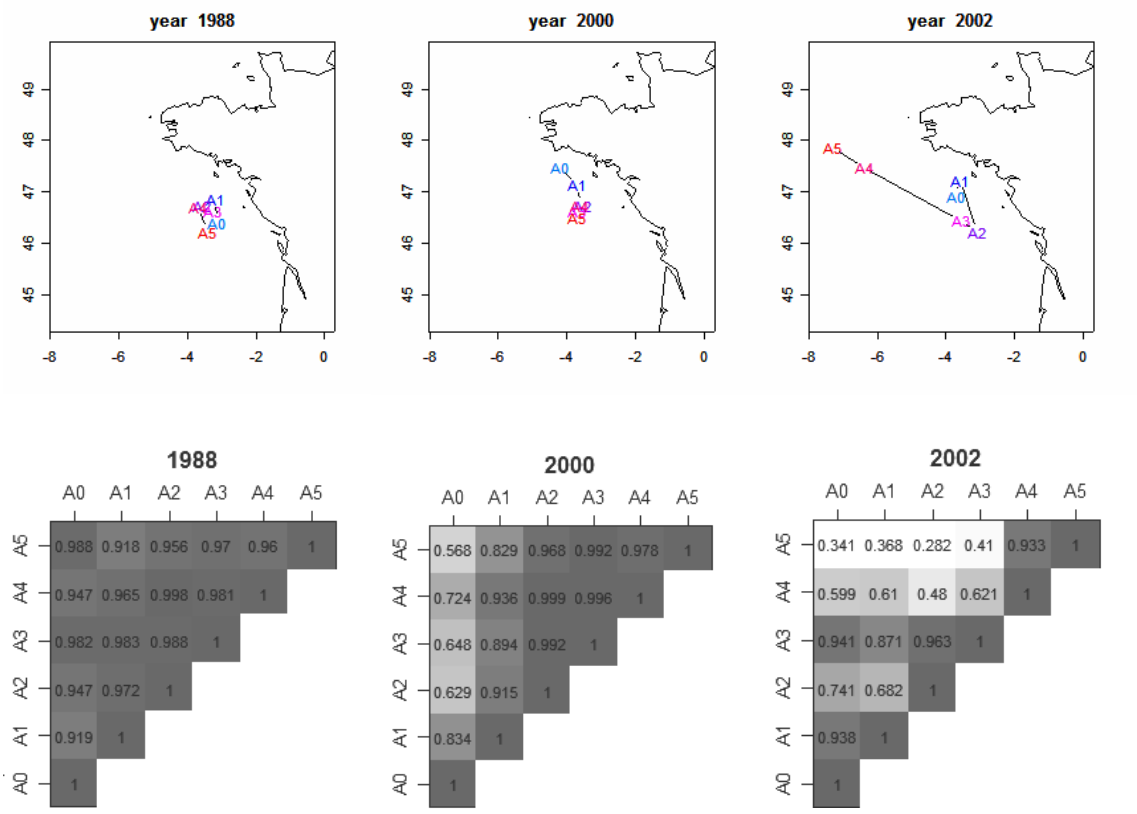

Fig. 3. Centre of gravity of the location of hake densities and global index of collocation computed between age classes for the years 1988, 2000 and 2002.

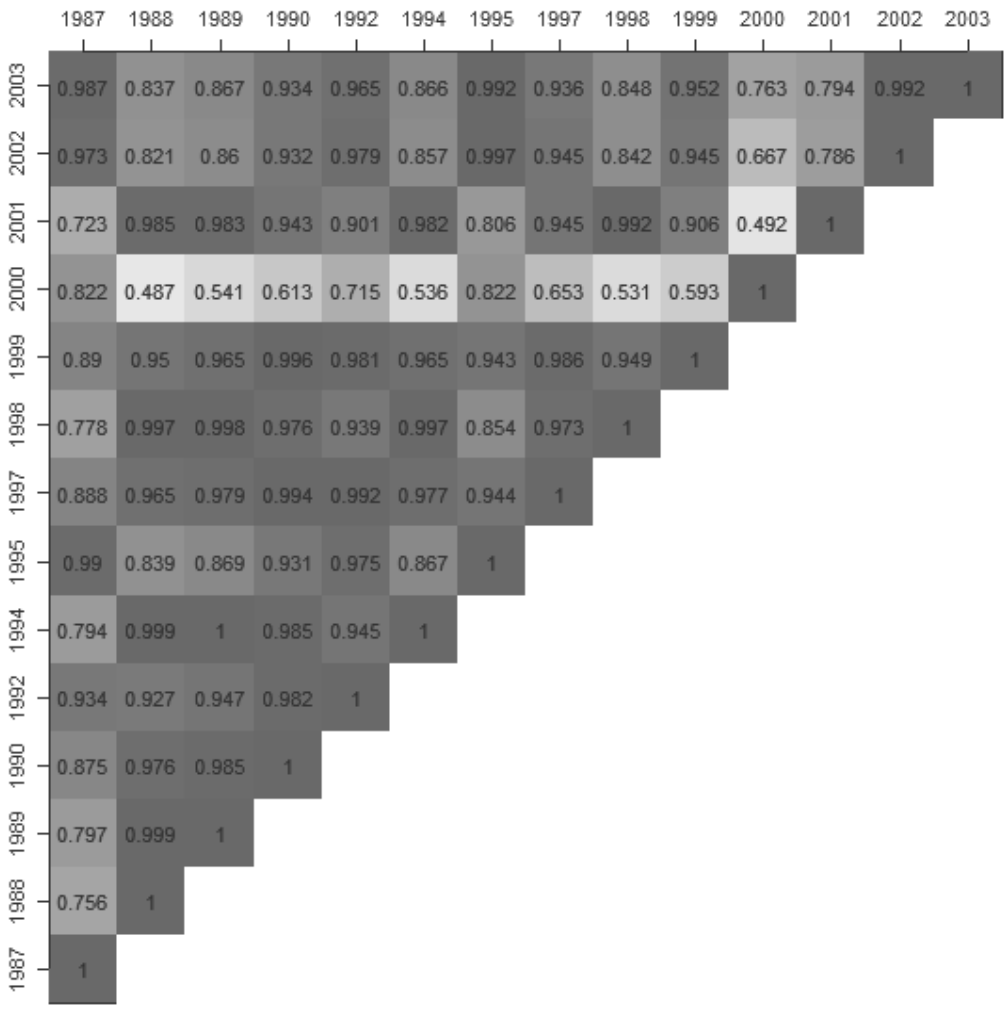

Fig. 4. Global index of collocation computed between years of age 0.

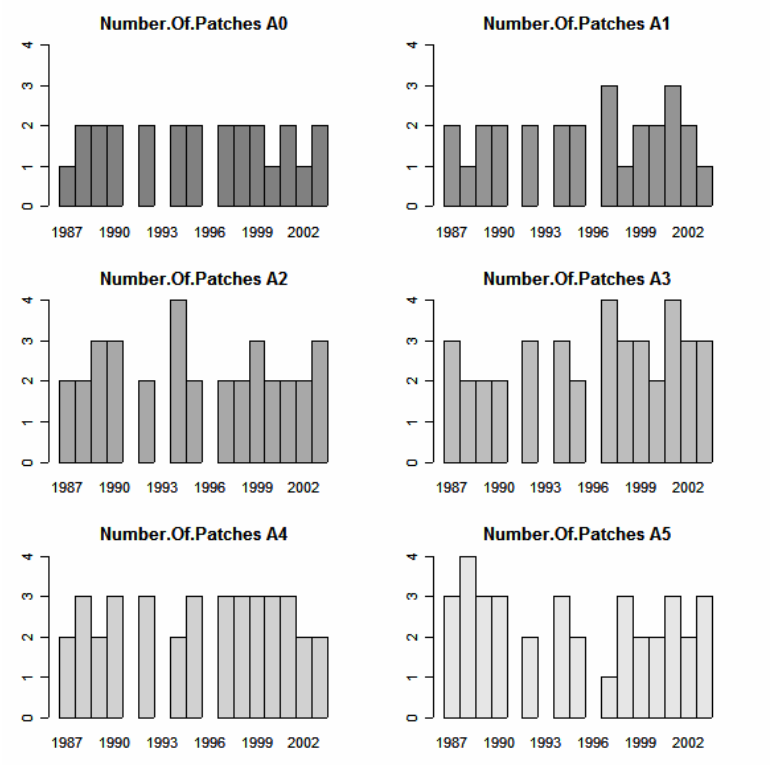

Fig. 5. Time series of number of patches for the different age classes.

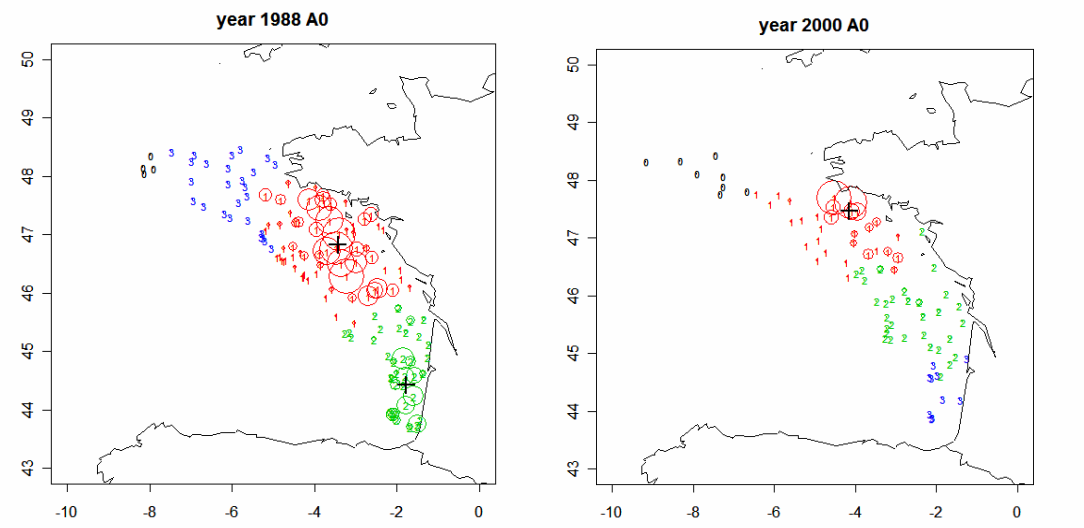

Fig. 6. Bubble plot of hake densities and the corresponding patches (marked by the black cross) defined by a limit distance of 100 nm and a biomass superior to 10 % of the total biomass.

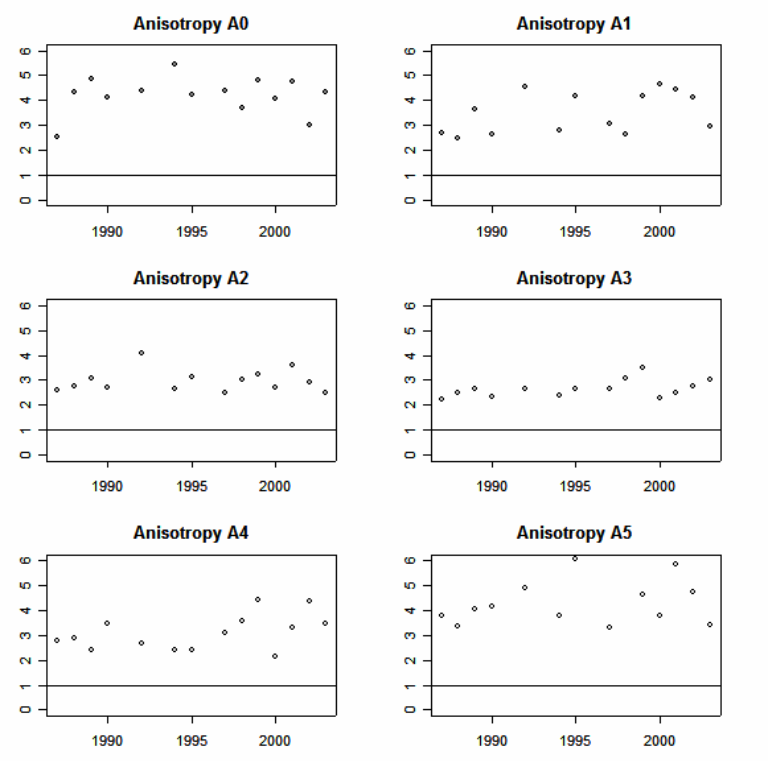

Fig.7. Time series of anisotropy for the different age classes.

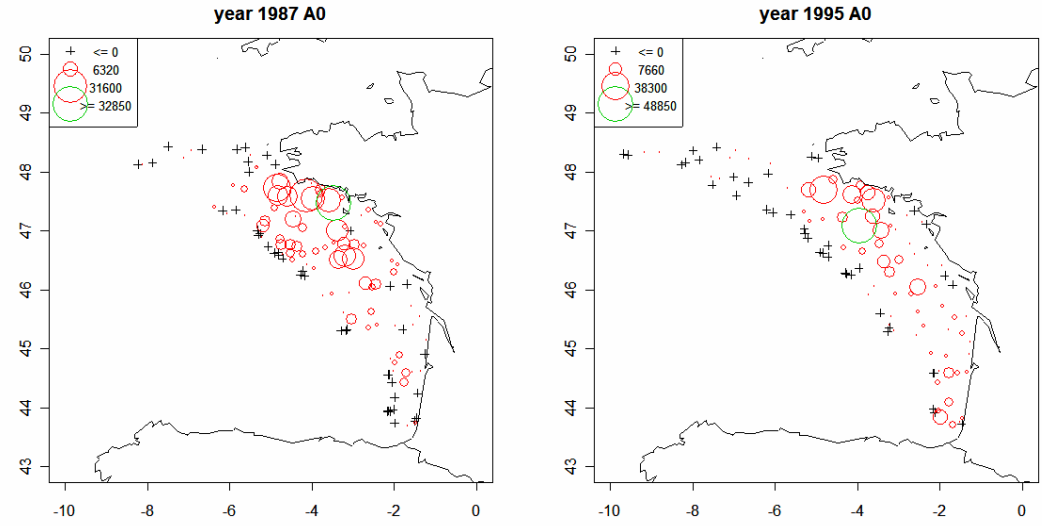

Fig. 8. Bubble plot of hake densities at age 0 for the years 1987 and 1995.

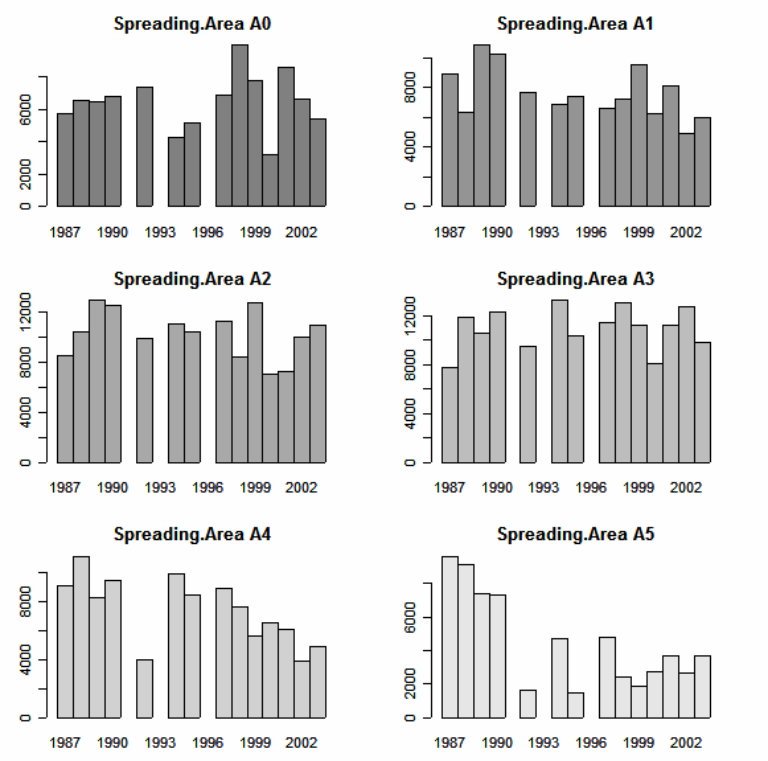

Fig. 9. Time series of spreading area for the different age classes.

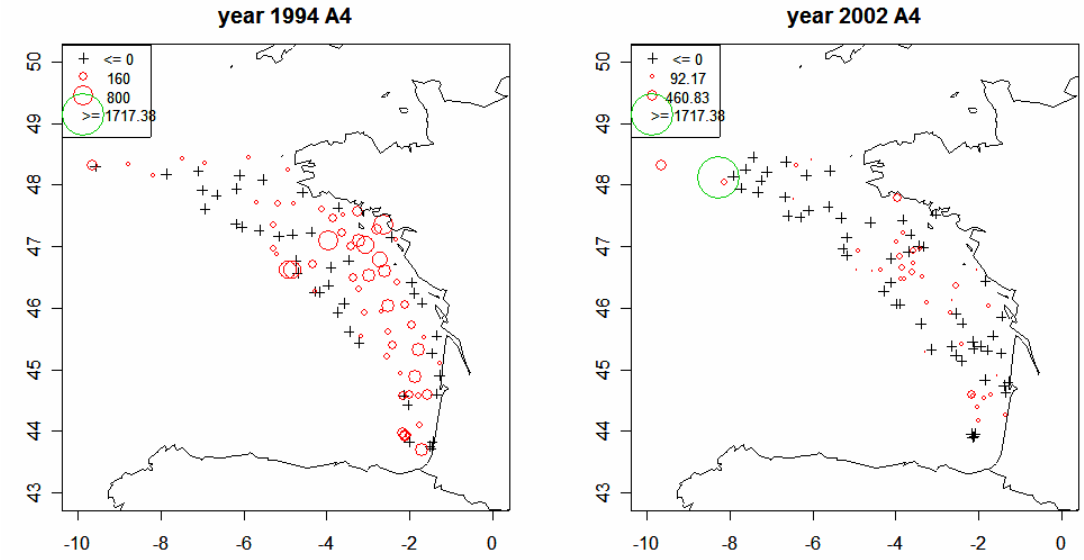

Fig. 10. Bubble plot of hake densities at age 4 for the years 1994 and 2002.

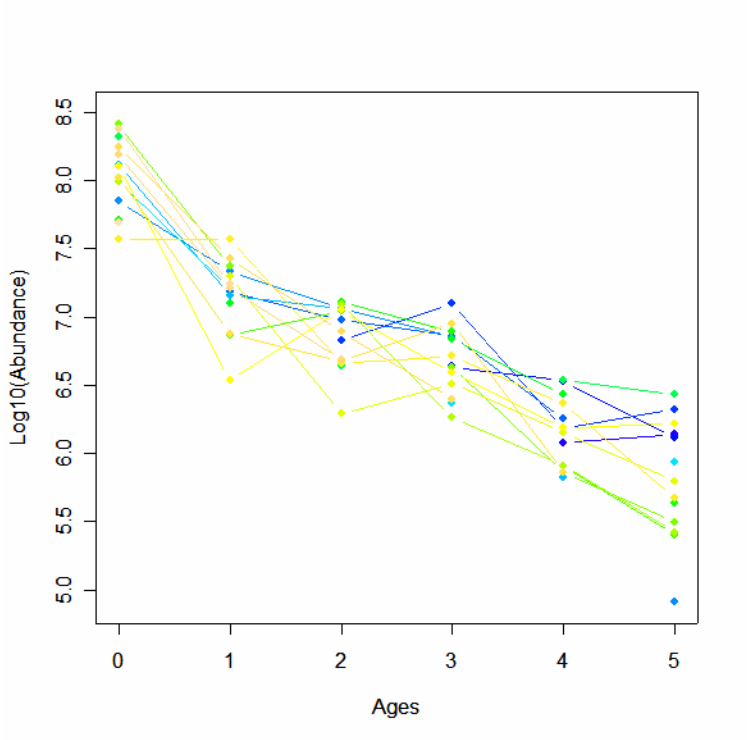

Fig. 11. Follow up of the log10 of the abundance for the ages of a cohort.

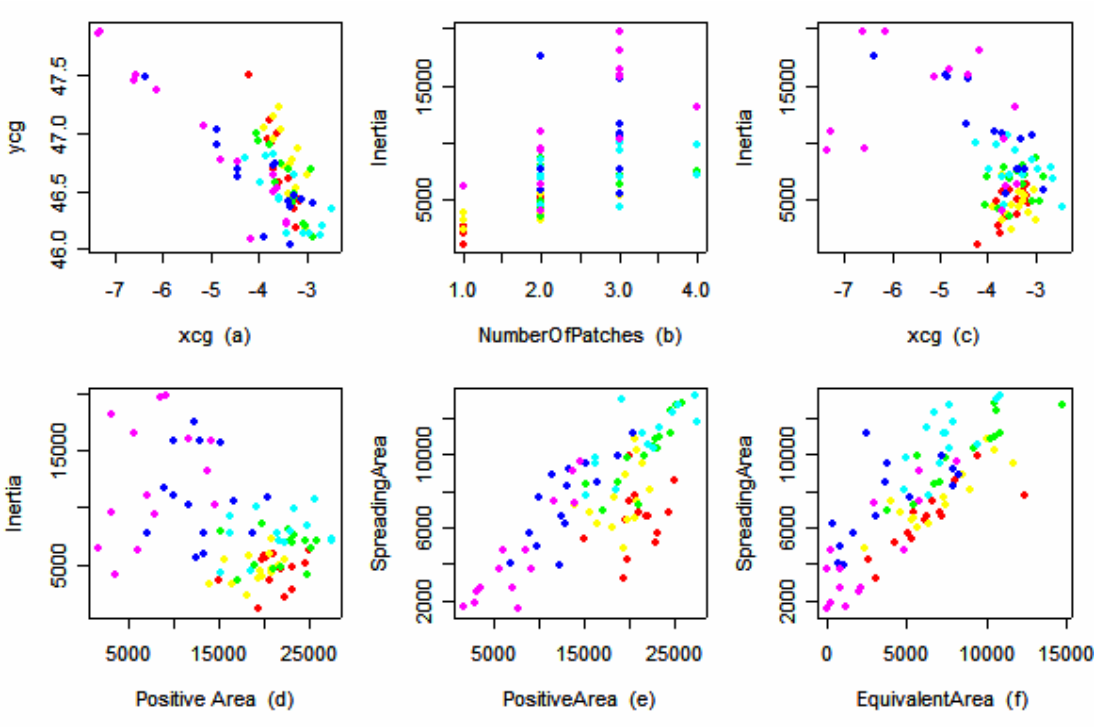

Fig. 12. Most interesting pairewise scatter plot between the spatial indices. Color code corresponds to red for the age 0, yellow for age 1, green for age 2, light blue for age 3, dark blue for age 4 and purple for age 5.

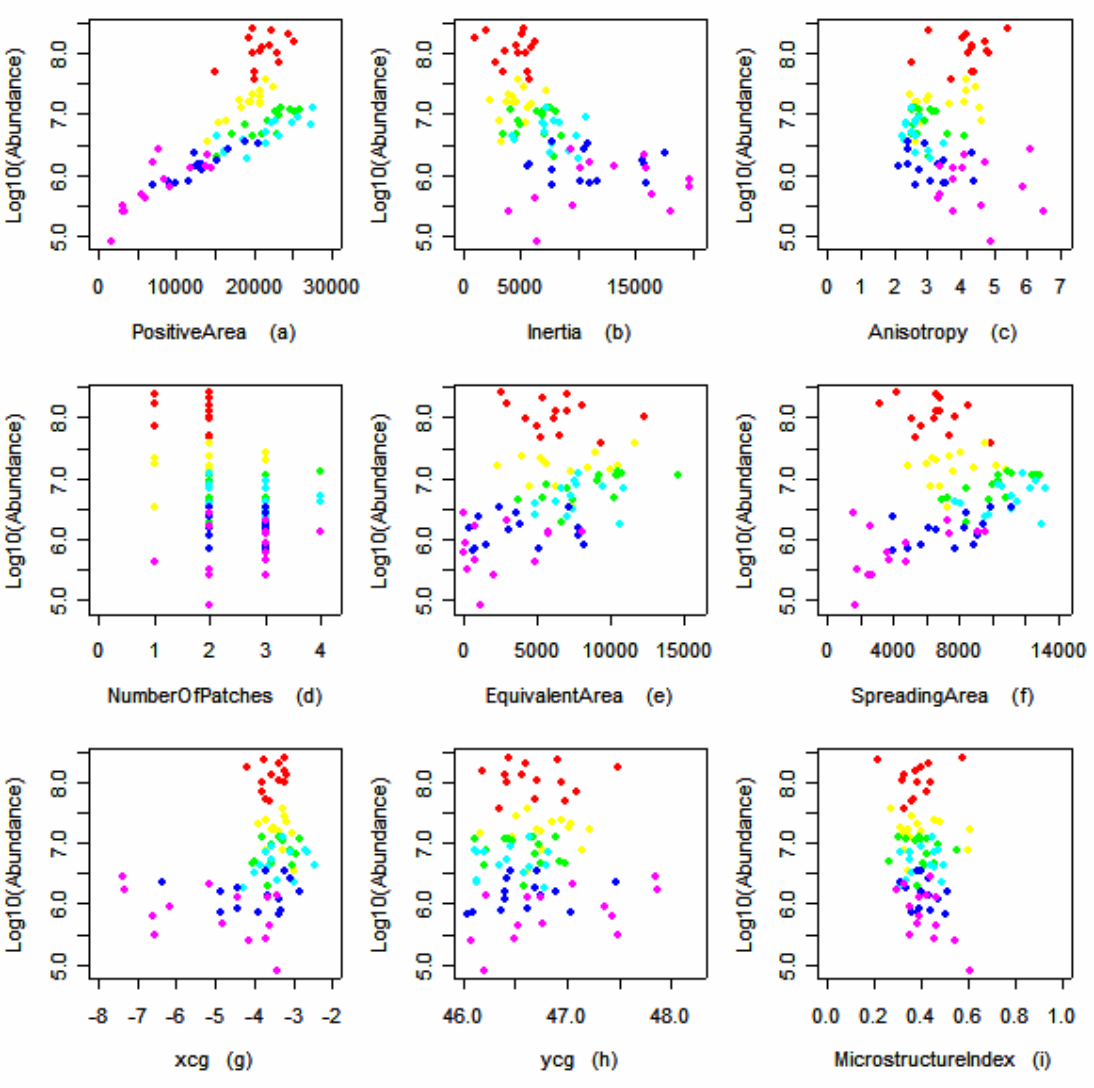

Fig. 13. Scatter plot between the log10 of the abundance and the spatial indices. Color code corresponds to red for the age 0, yellow for age 1, green for age 2, light blue for age 3, dark blue for age 4 and purple for age 5.