Aquatic Living Resources

A multivariate indicator to monitor changes in spatial patterns of age-structured fish populations

Pierre Petitgas[a] and Jean-Charles Poulard

IFREMER, Dép. Ecologie et modèles pour l'halieutique, EMH, BP 21105, 44311 Nantes Cedex 3, France

Abstract – Quantitative analysis of the spatial patterns in age-structured fish populations provides a useful complement to stock assessment methods. The spatial distribution of an age-structured population is characterized by indicators for location, dispersion and aggregation, which are estimated from survey data. The times series of spatial indicators are organised in a 3D data structure with dimensions along indicators, age classes and years. Multi factor analysis (MFA) is applied to quantify the reproducibility in time of the multivariate structure between age classes and indicators. MFA computes the mean annual pattern over the age classes. The deviation around that pattern in each year is also quantified. A multivariate indicator that characterise the spatial pattern over all age classes is developed. Its variation in time is evaluated with quality control methods. The multivariate indicator is shown to be χ^2 distributed and a CUSUM control chart is used to detect any significant changes in time. Software for implementing the method is presented and described. To illustrate this new method, North Sea cod data are analysed, and changes in the spatial patterns of the age-structured population are detected and interpreted.

Key words: Multi factor analysis / CUSUM / Monitoring / Spatial distribution

1 Introduction

Fisheries research survey programs provide time series of a wide range of indices measuring attributes of fish populations. The indices may serve as indicators to monitor the state of fish stocks with a broad biological basis, increasing the reliability of assessments. In particular, surveys provide information on the spatial distribution of different age groups in the fish populations. Geographical organisation of life cycles being related to population dynamics (e.g., Sinclair 1988; Bakun 1996), indicators of spatial distributions provide information complementary to the classical (non spatial) biological indicators such as abundance at age or length at maturity. It has long been acknowledged that the range of a population's spatial occupation can vary with its abundance (Paloheimo and Dickie 1964; Fisher and Frank 2004) and different models have been proposed (Petitgas 1998). But spatial distributions are not controlled by density-dependence only. They also result from the interactions between environmental factors controlling the suitability of the habitats (e.g., temperature or food), and internal factors in the population controlling the capacity to occupy these habitats (e.g., demography, learning: Corten 2002). Further, some population parameters (e.g., recruitment, mortality) relate to particular characteristics of the spatial distributions of age groups (e.g., year class strength and the spatial distribution of spawners: Woillez et al. 2006). Therefore,

it is important to consider the spatial patterns in all age groups when assessing the status of a population. The spatial organisation of different age groups can be characterized using for each age group a list of spatial indicators that characterize the different aspects of its distribution (e.g., aggregation, occupation, location: Woillez et al. 2007, 2009). Here we suggest combining all these indicators for all age groups into one multivariate indicator summarising the spatial pattern over all age groups. We further suggest a method for monitoring deviations around the average pattern.

When many correlated indicators are used for monitoring purposes, they may show conflicting signals that are not interpretable, or over emphasize the seriousness of the situation. A multivariate monitoring procedure is then more efficient than procedures based on the analysis of a collection of univariate monitoring charts, as the multivariate analysis will make coherent use of the relationships between the many indicators (Montgomery 2005). Here, we used a multivariate distance-based approach adapted to ecological monitoring applications (Anderson and Thompson 2004). We constructed a multivariate index to measure the departure in each year from the average spatial pattern, and monitored these departures with an appropriate cumulative sum monitoring scheme (CUSUM: Hawkins and Olwell 1998; Mesnil and Petitgas 2009). The construction of the multivariate spatial indicator was based on multi factor analysis (MFA: Escoffier and Pagès 1994; Dazy and Le Barzic 1996), which is a multi-table method based on

---

[a] Corresponding author: pierre.petitgas@ifremer.fr

principal components analysis (PCA) and designed to quantify the reproducibility in time of a multivariate structure. MFA has been applied already to hake in the Bay of Biscay to provide an overview of the spatial patterns in different age groups (Woillez et al. 2007). Following further this approach, a multivariate indicator is here developed to measure the departure from the average spatial pattern and monitor these departures in time. The procedure developed here is original and provides the means to statistically monitor changes in the spatial distribution of age-structured fish populations using research survey data.

2 Method

2.1 Multi factor analysis (MFA)

MFA is a multi-table method that extends principle components analysis (PCA) to the analysis of 3D structured data (Escoffier and Pagès 1994; Dazy and Le Barzic 1996). MFA is designed for situations where the same variables (data matrix columns) are measured on the same individuals (data matrix rows) at various times (or for different experiments). In the MFA analysis, variability in time and correlation structure between variables are separated. This permits to estimate an average correlation structure and quantify variations in time. In contrast in single PCA, all sources of variability are pooled and thus the structures are less clearly identified. When using MFA, the data are structured as a time series of elementary matrices. MFA computes (using a double PCA) a factorial space that is a compromise between all elementary matrices, in which the 3D structure of the data is represented. Each elementary matrix, each variable and each individual is geometrically positioned in the MFA space. In particular, each individual (matrix row) is represented by n points (n repetitions in time). Further, the principal components of the MFA are interpretable using their correlation with the variables. MFA has been applied in fisheries science to characterise seasonal and inter-annual variations in fish community structures (Gaertner et al. 1998), fishing activities (Poulard and Léauté 2002) as well as identify a common structure between trophic levels (Petitgas et al. 2006).

To characterise the average spatial pattern of the age classes as well as the variability between years, MFA requires the following data and structure. For each age group in each year, the spatial indicators defined in Woillez et al. (2007, 2009) are first estimated from the survey data. The variables in the study are the spatial indicators and the individuals are the age groups. Thus the yearly elementary matrices contain the spatial indicators as columns and the age groups as rows. The columns of the yearly elementary matrices are centred and normed. The MFA is then applied on the time series of the yearly elementary matrices (3D data structure). In the MFA factorial space, the spatial distributions of each age group are represented by Nt points (Nt years) and their gravity centre represents the age group average spatial pattern. The average spatial pattern over all age classes is then represented by the Ng gravity centres of the Ng age groups. To interpret the principal axes of the MFA, their correlation with each variable in each year are estimated, resulting in Nt correlation coefficients

for each variable along each axis. For each variable and axis we counted the number of times the correlation was equal or greater than 0.5 or equal or lower than –0.5. The axes were interpreted using these counts. In general, structure in the data is strong, meaning that there is a marked change with age in the spatial distribution characteristics of the fish (e.g., location, aggregation).

2.2 Multivariate index and its statistical distribution

The deviation in each year around the average spatial pattern is here quantified in the MFA space by a square distance between the points representing year and age and the age-specific gravity centres. In a monitoring approach the current year is assessed as being inside or outside in-control limits. Therefore the monitoring scheme requires a reference to be defined beforehand. In-control mean and variance parameters is here estimated (Mesnil and Petitgas 2009) in a reference period defined as a list of years. The age-specific gravity centres are then computed using the reference years only and so were the age-specific variances around the gravity centres.

Let $X_{k,a,t}$ denote the coordinate on principal axis k for age a in year t, $C_{k,a}$ the gravity centre on axis k for age a, and $\sigma^2_{k,a}$ the variance in time of $X_{k,a,t}$ around $C_{k,a}$. Let N_{ref} be the number of reference years and Ref the list of reference years. The reference age-specific gravity centre is estimated during the reference period:

$$C_{k,a} = \frac{1}{N_{\mathrm{ref}}} \sum_{t \in \mathrm{Ref}} X_{k,a,t}.$$

The set of points $C_{k,a}$ represent the reference average spatial distribution over all ages. During the reference period, the reference variance for each age is:

$$\sigma^2_{k,a} = \frac{1}{N_{\mathrm{ref}}} \sum_{t \in \mathrm{Ref}} (X_{k,a,t} - C_{k,a})^2.$$

In year t, the elementary square distance on axis k between point $X_{k,a,t}$ and the age-specific gravity centre for age a is:

$$d^2_{k,a}(t) = (X_{k,a,t} - C_{k,a})^2 / \sigma^2_{k,a}.$$

The multivariate spatial index in year t over all ages is then:

$$D^2(t) = \sum_{k=1}^{Nc} \sum_{a=1}^{Ng} d^2_{k,a}(t)$$

where Nc is the number of principal axes considered and Ng the number of age groups.

The multivariate index $D^2(t)$ measures the departure in year t from the average spatial pattern over the age classes. The (reference) average spatial pattern is defined during the reference period. All age classes are here given similar importance in the construction of $D^2(t)$.

Assuming that the elementary coordinate $(X_{k,a,t} - C_{k,a})$ / $\sigma_{k,a}$ along each principal axis is a Gaussian with zero mean and unit variance in the reference period, and that variability in the coordinates along the different axes are uncorrelated,

the multivariate index, $D^2(t)$ $t \in Ref$, has a χ^2 distribution $\chi^2(df, 0)$ with non-centrality parameter 0 and df degrees of freedom ($df = Nc(Ng - 1)$). The scatter of individual points being centred in the MFA space, $Ng-1$ is logically considered in calculating df.

2.3 CUSUM control chart for a χ^2 distributed indicator

Once the time series of $D^2(t)$ is estimated, significant shifts from a mean reference value can be detected with a defined statistical performance by applying a CUSUM monitoring scheme (Hawkins and Olwell 1998; Mesnil and Petitgas 2009). In-control limits for the monitoring chart are estimated after the acceptable probabilities of false positive and false negative signals have been defined. When the CUSUM chart trespasses these limits a significant out-of-control shift is detected. Limits are based on the statistical distribution that the indicator is assumed to follow. The indicator $D^2(t)$ being χ^2 distributed the Gaussian transform $\sqrt{2D^2(t)}$ could be applied and a CUSUM monitoring scheme for a Gaussian variable used (Petitgas 2009). Alternatively here, a CUSUM monitoring scheme for a χ^2 variable is implemented and practical guide lines given for doing so. It is expected useful to detail a procedure adapted to multivariate indicators and based on quadratic measures, as these are often encountered in ecological studies.

The CUSUM monitoring scheme for a χ^2 distributed variable is here designed to detect a deviation in the non centrality parameter ncp over time, the degrees of freedom being fixed. The monitoring scheme will signal when values of $D^2(t)$ are expected to belong to $\chi^2(df, ncp)$ rather than $\chi^2(df, 0)$ due to a shift in the mean of $D^2(t)$ of magnitude ncp. A shift in ncp may originate from a shift in the mean or the variance of the elementary distances $d^2_{k,a}(t)$. The CUSUM monitoring sequence S_t for the χ^2 variable $D^2(t)$ will accumulate positive only deviations:

$$S_0 = 0; S_t = \max(0, S_{t-1} + D^2(t) - ck).$$

The scheme will signal when: $S_t > ch$. The parameter ck (allowance) filters inherent variability as only the positive deviations from ck are accumulated. The parameter ch (decision interval) defines the out-of-control limit for S_t. The CUSUM monitoring scheme is tuned to detect deviations above ch with set probabilities of false-positive and false-negative signals (Mesnil and Petitgas 2009). First the parameter ncp (shift in the mean of the distribution) is estimated outside the reference period by: $ncp = m_1 - df$, where df is the number of degrees of freedom ($df = Nc(Ng-1)$) and m_1 is the reference (in-control) mean of $D^2(t)$. Then the following steps are followed iteratively until a compromise is achieved between a low enough probability of false alarm and a short enough detection time:

- choose ck and ch;
- estimate the average run lengths (ARLs) of the CUSUM monitoring scheme: *ARL(0)* the time to signal a change when none has happened (time to a false alarm or inverse of the probability of a false positive signal); *ARL(ncp)* the time to signal a change when one has effectively happened

(time to detection or inverse of the probability of a false negative signal).

For a first choice of ck we suggest to take the optimal ck value given by Hawkins and Olwell (1998, chapter 6) to detect a shift in the scale parameter b of a Gamma distribution when its shape parameter a is fixed. The χ^2 distribution $\chi^2(df, 0)$ is a special case of the Gamma distribution $\Gamma(a = df/2, b_0 = 2)$. A change in ncp with fixed df for the χ^2 is thus paralleled to a change from b_0 to b_1 with a fixed a for the Γ. The value of b_1 can be estimated by considering that the means of $\chi^2(df, ncp)$ and $\Gamma(df/2, b_1)$ are equal. In general, this first choice for ck is satisfactory. The ARLs of the CUSUM for a χ^2 variable were estimated by simulations (Jun and Choi 1993). The procedure used a variance reduction technique with the cumulative distribution of the CUSUM as the control variate. Variance of the ARLs were estimated using bootstrap (Jun and Choi 1993). Monitoring outside the reference period can then proceed with the parameters ck and ch corresponding to satisfactory ARLs.

3 Application to North Sea cod

To illustrate, the procedures were applied to the North Sea cod stock for which indices of the spatial distribution were estimated (Woillez et al. 2007, 2009) using data from the North Sea quarter 1 International Bottom Trawl Survey series (DATRAS data base maintained by ICES). The survey time series spanned 21 years, 1985-2005. The ages recorded in the survey ranged from 1 to 6. Eight spatial indicators for each age were considered, which are available from the EU project FISBOAT website at http://www.ifremer.fr/drvecohal/fisboat/. The eight spatial indicators as defined by Woillez et al. (2007, 2009) were: longitude and latitude of the gravity centre, inertia, anisotropy, positive area, spreading area, equivalent area, microstructure. They characterised location, dispersion, aggregation, occupation and correlation in the spatial distributions of the different age groups.

The data were organised in yearly tables containing the indicator values in columns and the age groups in rows. MFA was applied to that series of matrices. The age-specific reference gravity centres were positioned in the MFA sub-space of the first two principal components. The gravity centres represented the reference age-structured spatial organisation of North Sea cod (Fig. 1). Marked and progressive differences existed across ages. The two principal components explained 69% of the data variance. The inter-annual variability for each age was represented by an ellipse. Each point represented the position of a given age in a given year. The principal axes were interpreted based on the reproducibility in time of their correlation with the spatial indicators (Table 1). The first axis distinguished spatial distributions that showed larger occupied areas, more northerly gravity centres, and higher inertia (dispersion) to distributions showing more westerly gravity centres and a smaller correlation range (microstructure). The second axis separated distributions that showed stronger anisotropies and more northerly gravity centres to distributions showing smaller positive areas. Marked and progressive differences in the spatial distributions of the different age groups thus characterized the spatial pattern of the age-structured
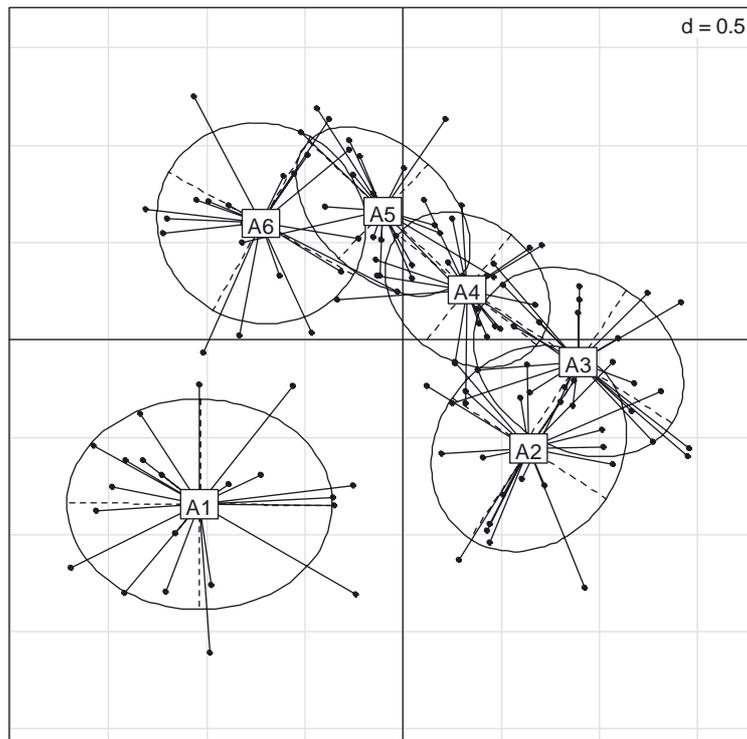
Fig. 1. MFA representation of the North Sea cod age-structured spatial pattern. The points represent the spatial distribution of each age in each year. The age labels represent the age-specific gravity centres, materialising the reference age-structured spatial pattern.

Table 1. Number of times in the series 1985-2005 that the spatial indicators for North Sea cod are correlated with the MFA first two principal axes. Correlation is counted when it is greater or equal to 0.5 in absolute value. The sign of the correlation is indicated after the count number. When the count is greater than 11, correlation occurred for more than half of the years, and the count is marked in bold.

	Principal axis 1	Principal axis 2
Positive Area	**12+** | 0–	0+ | **16**–
Inertia	**12+** | 2–	1+ | 2–
Anisotropy	0+ | 2–	**15+** | 0–
Longitude of gravity Centre	0+ | **17**–	4+ | 4–
Latitude of gravity Centre	**13+** | 0–	**15+** | 0–
Microstructure	1+ | **11**–	3+ | 3–
Equivalent Area	**12+** | 0–	1+ | 2–
Spreading Area	**18+** | 0–	0+ | 7–

North Sea cod population. Young (A1) and old ages (A5-6) differed on the first axis from intermediate ages (A2-4). Spatial distributions of young and old ages tended to be more to the east, less dispersed, and occupying smaller areas as well as showing higher small-scale heterogeneity. In contrast, spatial distributions of intermediate ages were more to the west, more dispersed, occupied larger areas and showed smoother correlation. Age 1 and Ages 5-6 differed on the second principal axis because of their gravity centres and anisotropies. The spatial distribution of old ages was more to the north, more anisotropic, and occupied a smaller area than that of the age-1 fish.

The multivariate distances to the age-specific reference gravity centres were calculated in the plane of the first two

principal axes ($Nc = 2$, $Ng = 6$). The distribution of the elementary square distances $d^2_{k,a}(t)$ compared well with a χ^2 distribution $\chi^2(1, 0)$ (Fig. 2). This confirmed that the variations in the different years and ages around the reference age-specific gravity centres could be assumed Gaussian. As a result, the distribution of the multivariate spatial index $D^2(t)$ was assumed $\chi^2(df = 2(6 - 1) = 10, ncp = 0)$.

A trend was visually apparent on the time series of the multivariate spatial index $D^2(t)$ (Fig. 3). The departure from the reference spatial pattern increased during the period 1985-2005 except for the years 1996-2000. The highest values in the series were in the period 2001-2005. Did these high values belong to a $\chi^2(df, ncp > 0)$? With what statistical performance (i.e., ARLs) could a shift from 0 to ncp be detected? To answer these questions, a monitoring scheme was tuned and applied. The reference (in-control) period was the first 10 years of the series, 1985-1994, which corresponded to a relatively high abundance in the survey data series with no particular trend in abundance nor recruitment. The reference period was agreed after discussions that took place during an international workshop of the FISBOAT project.

The CUSUM monitoring scheme was tuned using simulations. We generated $nr = 5000$ Run Lengths from which the Run Length (RL) distribution, the average RL (ARL) and the quartiles were estimated. Variance of an ARL was estimated by resampling. A 100 RLs were resampled with replacement in the set of simulated RLs to estimate $(RL\text{-}ARL)^2$ and this was repeated 100 times to estimate $E[(RL\text{-}ARL)^2]$.

Tuning the CUSUM scheme resulted in the following parameters: $ncp = 11.0$, $ck = 14.2$, $ch = 5.0$, $ARL(0) = 23$ *years*, $ARL(ncp) = 2$ *years*. A shift in ncp from 0 (in-control) to
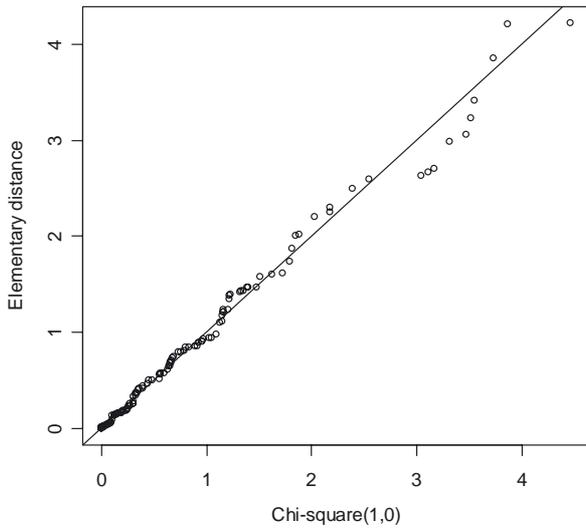
Fig. 2. Q-Q plot between the elementary square distances $d^2_{k,a,t}$ and the χ^2 distribution with one degree of freedom, for the reference period (1985-1994).
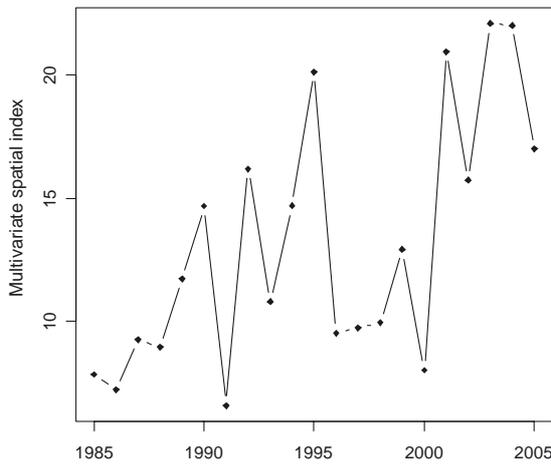
Fig. 3. Time series of the MFA-based multivariate spatial index $D^2(t)$. The index is a square distance in the MFA factorial subspace of the two first principal components measuring the departure in each year from the reference average spatial pattern over all age groups.

Fig. 4. Distribution of the (in-control) Run Length RL(0) of the CUSUM sequence for the multivariate spatial index $D^2(t)$ that is χ^2 with 10 degrees of freedom $\chi^2(10,0)$. The run length RL(0) is the time to a false alarm when no change in the distribution parameters has occurred.
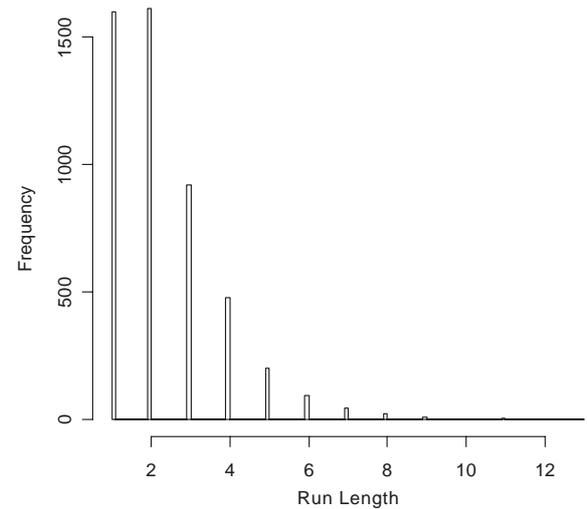
Fig. 5. Distribution of the (out-of-control) Run Length RL(*ncp*) of the CUSUM sequence for the multivariate spatial index $D^2(t)$ that is χ^2 with 10 degrees of freedom $\chi^2(10,0)$. The run length RL(*ncp*) is the time to detect a change in the non centrality parameter *ncp* when it has effectively happened.

11(out-of-control) could be detected by using parameters *ck* and *ch* as defined. A false alert will occur on average every 23 years (false alarm probability of 0.04) and it will take on average 2 years to detect a change once it has happened. The distributions of the RLs were skew (Figs. 4 and 5). The first quartile of *RL(0)* was 7 years (the chance was 0.25 that a false alarm occurred every 7 years) and that of the *RL(ncp)* was 1 year. The ARLs were estimated with good precision as their standard deviations were: $\sigma(ARL(0)) = 0.23$ and $\sigma(ARL(ncp)) = 0.01$ years.

The application of the monitoring scheme (Fig. 6) detected out-of-control values in the year 1995 and since 2001. Although the 2002 and 2005 values were not extreme, the accumulation above the threshold *ch* did not allow the CUSUM scheme to return into the in-control area below the threshold. The values in the period 1996-2000 were low enough for the CUSUM to return into the in-control zone in that period after signalling out-of-control in 1995. This result compared with the alerts raised when using another distribution for the control chart. Petitgas (2009) used the Gaussian transform $\sqrt{2D^2(t)}$ on the same data and applied a CUSUM scheme in the case of a Gaussian indicator. Both CUSUM schemes, when considering a gaussian or a χ^2, signalled changes in the same years (1995 and 2001 to 2005). The time to detection was similar (2 years) in both cases. But the probability of false-alarm in the case of the χ^2 was higher: the ARL(0) was 23 years in the case of the χ^2 instead of 52 for the Gaussian. When considering a skewer distribution for the indicator the performance of the corresponding CUSUM scheme is thus lower.
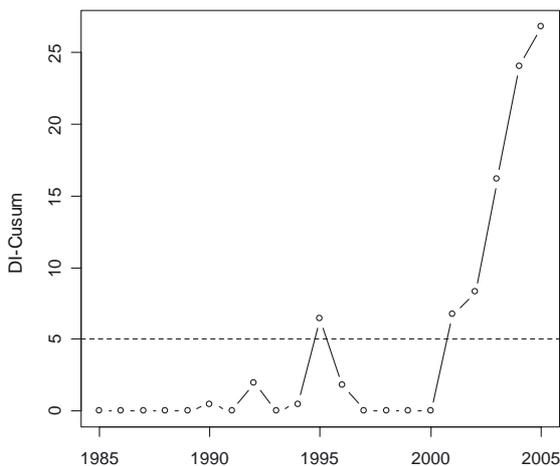
Fig. 6. Decision-interval CUSUM (DI-CUSUM) control chart for monitoring the multivariate spatial indicator $D^2(t)$. The in-control distribution of the indicator is a χ^2 with 10 degrees of freedom $\chi^2(10, 0)$. The values above the threshold indicate an out-of-control state in which the non centrality parameter has shifted from 0 to 11. The detection is achieved with an average false alarm rate of 0.04 and an alarm detection time of 2 years. The parameters of the CUSUM scheme are $ck = 14.2$ (allowance) and $ch = 5$ (decision interval).

4 Software

To help readers implement an MFA-based multivariate spatial index and its CUSUM monitoring scheme for a χ^2 distributed variable, attached to this paper on-line are two scripts in the R language (R Development Core Team 2005) that were developed for the purpose of the EU project FISBOAT (see Supporting information, only available in electonic form at www.alr-journal.org):

- FB_Mfa_Spatial_D2.R is designed to apply MFA, represent the age-structured spatial pattern and calculate the MFA-based spatial index $D^2(t)$. The data input is the table of indicators with the format specific to the FISBOAT project.
- FB_CUSUM_Chisq_Simu.R is designed to simulate the run length of a CUSUM control scheme for a χ^2 distributed variable and estimate its average run length with a variance reduction technique. The data input is the time series of the $D^2(t)$ MFA-based multivariate spatial index.
- The data files input to the above R scripts, which are used in this paper: codNS_tab1_indicspa.txt and codNS_mfa_d2spa.txt.

The table of spatial indicators for North Sea cod is also available from the project website at http://www.ifremer.fr/drvecohal/fisboat/. In that table, the columns are the spatial indicators and the rows are the series of years for each age. The table containing the time series of $D^2(t)$ has two columns one for the years and the other for $D^2(t)$.

The script *FB_Mfa_Spatial_D2.R* proceeds as follows. The R library *ade4* (Chessel et al. 2006) is first sourced as the MFA is performed using that library. The FISBOAT table of spatial indicators is read and its format modified to match the format needed for the MFA computation. The MFA is performed using the *ade4* routine *mfa*(). To interpret the MFA

principal axes the number of times the correlations of the indicators with the axes are greater than 0.5 in absolute value are counted. These counts are gathered into two separate tables for positive and negative correlation. The data are represented in the MFA sub-space of the two first principal axes as in Figure 1. Then the elementary distances to the age-specific gravity centres are computed for each age and year and the multivariate distance $D^2(t)$ is calculated. Results are saved in a R object as elements of a list.

The script *FB_Cusum_Chisq_Simu.R* is made of three functions and commands. The function *dicusum1.f( )* calculates the CUSUM sequence given the indicator series and the allowance parameter *ck*. The function *ck.chisq.f( )* calculates the optimal allowance *ck* for a CUSUM scheme designed to detect a change in the scale parameter of a Gamma distributed variable. The function *arl.chisq.f( )* simulates run lengths of the CUSUM scheme designed to detect a change in the non centrality parameter of a χ^2 variable. It estimates the average run length (ARL) using a procedure of variance reduction detailed in Jun and Choi (1993). It further estimates the variance of the ARL by bootstrap. The function also plots the distribution of the run lengths and saves the ARL, its variance, the first and third quartile of the distribution of run lengths. In the part of the script with the commands lines, the indicator series is first read. The degrees of freedom for the χ^2 is input. Also input is the list of years outside the reference period, which serve to estimate the shift to be detected in the non centrality parameter *ncp*. Then the function *ck.chisq.f( )* is used to give a first choice for the parameter *ck*. The decision interval *ch* is input. The function *arl.chisq.f( )* is used to calculate the ARLs. The commands can be iterated using different values for *ck*, *ch* or *ncp* until a CUSUM control chart is achieved with good statistical performance (high ARL(0) and low ARL(*ncp*)).

5 Discussion

The MFA-based multivariate spatial indicator $D^2(t)$ and its CUSUM control chart are a simple and practical procedure for monitoring spatial distributions of age-structured populations. When a change is detected, the natural next step is to understand the cause of that change. For that purpose, it may be helpful to analyse in the MFA space the directivity in the age-specific scatters of points around their gravity centres as well as the relative contributions of the ages to the indicator $D^2(t)$. Dazy and Lebarzic (1996) suggest to analyse the age-specific scatters of points as trajectories in time in the MFA space. Directivity along the principal axes and time may indicate systematic change that can be interpreted using the correlation of the variables on the principal axes. Here in the application to North Sea cod, the scatters for age 4 and 5 showed some directivity in the factorial space (Fig. 1) and in time (not shown), meaning that these ages tended to show in recent years a reduction in area occupied and a more northerly centre of gravity. Also in the years in which $D^2(t)$ was greatest (Fig. 3), these high values were generated by several age groups (not just one) which showed important departures from their gravity centre: Ages 1, 3 and 6 in 1995, and ages 1, 4 and 5 in 2001, 2003 and 2004.

Recent studies on changes in the spatial distribution of North Sea cod have showed a northerly shift in the gravity

centre of the distribution (Rindorf and Lewy 2006; Perry et al. 2005; Hedger et al. 2004). Our study considered here more spatial indicators, which allowed for an extensive description of the spatial distributions. The use of MFA on the spatial indicators for different age groups across the survey series enabled to summarize and yet describe extensively the spatial patterns over the age groups, as well as quantify their inter-annual variations. In the MFA analysis, the gravity centres of young ages were more southerly located than those of older ages, which agreed with Rindorf and Lewy (2006). The CUSUM control chart of the multivariate spatial index signalled significant deviations in the spatial distributions in the mid-90s and in the recent period since 2001, which was consistent with previous findings. Significant changes in the spatial distributions for the young and old age groups as detected in this study agreed with the results by Rindorf and Lewy (2006), which showed that the shifts in the gravity centres for young and old age groups were correlated.

The MFA method involves performing two PCAs. If the number of indicators is large in comparison to the number of age groups, the results of the PCAs will depend on a few data values and therefore be less reliable. In such situation, which applies to short-lived species, it is advised either to reduce the number of spatial indicators or to perform a single PCA instead of an MFA. In a single PCA approach, the data are structured in a single matrix where the indicators are in the columns and each row represents a given age for a given year. Similarly as in the MFA space, the age-specific reference gravity centres can be represented in the PCA space and the above procedures applied. The drawback of using a single PCA is that all variability is pooled and the results are therefore less structured. In effect in the MFA procedure, variability in time and correlation structure between indicators are separated in the analysis, leading to a well structured representation of the spatial pattern and its variation in time.

Supporting information

Appendix S1. FB_Mfa_Spatial_D2.R is designed to apply multi factor analysis.
Appendix S2. FB_CUSUM_Chisq_Simu.R simulates the run length of a CUSUM control scheme.
Appendix S3. Data files (txt) input to the above R scripts.

Supporting information is only available in electronic form at www.alr-journal.org

References

Anderson M., Thompson A., 2004, Multivariate control charts for ecological and environmental monitoring. Ecol. Applic. 14, 1921–1935.

Bakun A., 1996, Patterns in the ocean. California Sea Grant College System.

Chessel D., Dufour A., Dray S., 2006, ade4: analysis of ecological data. Package in R language. http://cran.r-project.org/

Corten A., 2002, The role of conservatism in herring migrations. Rev. Fish Biol. Fish. 11, 339–361.

Dazy F., Le Barzic J.-F., 1996, L'analyse des données évolutives méthodes et applications. Editions Technip, Paris.

Escoffier B., Pagès J., 1994, Multi factor analysis (AFMULT package). Comput. Stat. Data Anal. 18, 121–140.

Fisher J., Frank K., 2004, Abundance-distribution relationships and conservation of exploited marine fishes. Mar. Ecol. Prog. Ser. 279, 201–213.

Gaertner J.-C., Chessel D., Bertrand J., 1998, Stability of spatial structures of demersal assemblages. Aquat. Living Resour. 11, 75–85.

Hawkins D., Olwell D., 1998, Cumulative sum charts and charting for quality improvement. Springer Verlag, New York.

Hedger R., McKenzie E., Heath M., Wright P., Scott B., Gallego A., Andrews J., 2004, Analysis of the spatial distribution of mature cod (*Gadus morhua*) and haddock (*Melanogrammus aeglefinus*) abundance in the North Sea (1980-1999) using generalised additive models. Fish. Res. 70, 17–25.

Jun C.-H., Choi M., 1993, Simulating the average run length for CUSUM schemes using variance reduction techniques. Commun. Stat. Simul. Comput. 22, 877–887.

Mesnil B., Petitgas P., 2009, Detection of changes in time-series of indicators using CUSUM control charts. Aquat. Living Resour. 22, 187–192.

Montgomery D., 2005, Introduction to statistical quality control. 5th edn. Wiley, New York.

Paloheimo J., Dickie L., 1964, Abundance and fishing success. Rapp. P.-v. Cons. Int. Explor. Mer 155,152–163.

Perry A., Low P., Ellis J., Reynolds J., 2005, Climate change and distribution shifts in marine fishes. Science 308, 1912–1915.

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

Petitgas P., 2009, The CUSUM out-of-control table to monitor changes in fish stock status using many indicators. Aquat. Living Resour. 22, 201–206.

Petitgas P., Massé J., Bourriau P., Beillois P., Bergeron J.-P., Delmas D., Herbland A., Koueta N., Froidefond J.-M., Santos M., 2006, Hydro-plankton characteristics and their relationship with sardine and anchovy distributions on the French shelf of the Bay of Biscay. Scient. Mar. 70(S1), 161–172.

Poulard J.-C., Léauté J.-P., 2002, Interaction between marine populations and fishing activities: temporal patterns of landings of La Rochelle trawlers in the Bay of Biscay. Aquat. Living Resour. 15, 197–210.

Rindorf A., Lewy P., 2006, Warm windy winters drive cod north and homing keeps them there. J. Appl. Ecol. 43, 445–453.

Sinclair M., 1988, Marine Populations. University of Washington Press, Seattle.

Woillez M., Petitgas P., Rivoirard J., Fernandes P., terHofstede R., Korsbrekke K., Orlowski, A., Spedicato M.T., Politou C.Y., 2006, Relationships between population spatial occupation and population dynamics. ICES CM 2006/O:05

Woillez M., Poulard J.-C., Rivoirard J., Petitgas P., Bez N., 2007, Indices for capturing spatial patterns and their evolution in time, with application to European hake (*Merluccius merluccius*) in the Bay of Biscay. ICES J. Mar. Sci. 64, 537–550.

Woillez M., Rivoirard J., Petitgas P., 2009, Notes on survey-based spatial indicators for monitoring. Aquat. Living Resour. 22, 155–164.