

MANUAL OF INDICATORS AND METHODS FOR ASSESSING FISH STOCKS USING ONLY FISHERY-INDEPENDENT, SURVEY BASED INFORMATION

Contribution to EC research project:

Fishery Independent Survey Based Operational Assessment Tools (FISBOAT),

DG-Fish, STREP n° 502572 (2004-2007)

Co-ordinator: P. Petitgas, IFREMER, Nantes, France

compiled by

*J. Cotter, CEFAS, Lowestoft, Suffolk, NR33 0HT, UK
email: john.cotter@cefas.co.uk*

from contributions by :

*B Mesnil, IFREMER, Nantes, France
V Trenkel, IFREMER, Nantes, France
M-J Rochet, IFREMER, Nantes, France
P Petitgas, IFREMER, Nantes, France
M Woillez, Ecoles des mines, Fontainebleau, France
J Rivoirard, Ecoles des mines, Fontainebleau, France
J Cotter, CEFAS, Lowestoft, England
A Uriarte, AZTI, Pasaia, Spain
P Witthames, CEFAS, Lowestoft, England
M T Spedicato, COISPA Tecnologia & Ricerca, Bari, Italy
G Lembo, COISPA Tecnologia & Ricerca, Bari, Italy*

Contents

Contents	3
Abstract	5
Keywords:	5
1. Introduction	9
2. Surveys and estimation	13
2.1 Introduction	13
2.2 Potential limitations of survey data as the unique source of information for stock assessment and management	14
2.3 Estimation of indicator values from surveys	18
3. Biological indicators	21
3.1 Introduction	21
3.2 Intrinsic population growth rate, r	22
3.3 Total mortality, Z	23
3.4 Numbers-at-length, numbers-at-age: NaL , NaA	25
3.5 Spawning Stock in number: SSN	27
3.6 Length statistics: $Lbar$, $L25$, $L50$, $L75$	29
3.7 Survey catch weight: W	32
3.8 Condition: C	35
3.9 Gonadosomatic index: GSI	37
3.10 Length and age at maturity: $LaM50$, $AaM50$	39
4. Spatial indicators	41
4.1 Introduction	41
4.2 Centre of gravity: CG	43
4.3 Inertia: I	45
4.4 Anisotropy and isotropy: An , Is	47
4.5 Global index of collocation: GIC	49
4.6 Number of spatial patches: NOP	51
4.7 Positive area: PA	53
4.8 Spreading area: SA	55
4.9 Equivalent area: EA	57
4.10 Microstructure index: MI	59
4.11 References for spatial indicators	60
5. Methods	62
5.1 Introduction	62
5.2 Simulating population dynamics	64
5.2.1 ALADYM(v 08)	64
5.2.2 Estimating indicators and reference points	75
5.3 Indicator time-series methods	78
5.3.1 Nonparametric method for determining recent trends	78
5.3.2 Assessing the power to detect future trends	86
5.3.3 Statistical Process Control (SPC) schemes	89
5.3.4 Nonparametric statistical methods for assessing trends	95
5.4 Construction of multivariate indicators	118
5.4.1 Principal Components Analysis (PCA) and biological indicators	118
5.4.2 Multi-factorial analysis (MFA) and spatial indicators	122
5.4.3 Min/Max autocorrelation factors (MAFs) and time continuity	126
5.5 Diagnosing stock status from indicator series	129
5.5.1 Combining trend signals using a cause-effects table	129

5.5.2 A 'traffic light' procedure based on Cusum out-of-control tables.....	133
5.5.3 A multivariate statistical procedure.	136
Appendix 1.....	140
Format for documenting indicators.....	140

Abstract

This manual discusses the use of indicators for assessing the state of a fish stock. The indicators envisaged are estimated from trawl surveys, or trawl-supplemented surveys such as egg and acoustic surveys, in which catchability is maintained constant so far as practically possible. A preliminary section considers factors that determine the appropriateness of a survey for each species and indicator, and makes suggestions for statistical estimators for quantitative and descriptive indicators. A widely applicable selection of biological and spatial indicators is then documented in a standardised format that includes references to examples of their use and modifications. The biological indicators include those relating to quantity of fish, size, and reproduction. The spatial indicators characterise the geographic distribution of a stock and make allowance for low or zero densities of fish at some stations. The final section of the manual presents several presentational and statistical methods for assessing and interpreting trends in indicators. A stock simulation model is described that may assist with determination of reference points for biological indicators. Since indicators tend to be highly specific and normally many would be used to assess the state of a stock, multivariate methods form an important part of this section. Indicators offer valuable biological and geographic information for supplementing existing model-based stock assessments. They are also likely to form an important component of an ecosystem approach to fishery management, or they could be used pragmatically to tune harvest control rules in a form of adaptive management.

Keywords:

fish stock assessment; trawl survey; fish survey; biological indicator; spatial indicator; stock status;

1. Introduction

Indices are measurable quantities that can be used to characterise a fish stock, for example, the mean length of the fish or the centre of gravity of its geographic distribution. When variation in a particular index relates to a process in the stock that one wants to follow, the index becomes an indicator of that process. This manual discusses indicators that are estimated from consistent time-series of results from scientific surveys conducted by research vessels (Anonymous 2004), or by other fishing vessels being operated purely for the ends of a survey and not for the pursuit of catches for commercial gain (since that goal would almost certainly cause bias). The surveys envisaged use a standardised trawl (Anonymous 2006) with small-mesh codends such that selectivity can be assumed to be reasonably constant for all fish larger than the selection range of the codend mesh. The trawling may only have a supporting role in the survey, for example for acoustic or egg surveys. For surveys using other catching or sampling methods, the applicability or otherwise of each indicator should be carefully considered in relation to selectivity. The fish species envisaged breed and recruit annually; some of the indicators may not be suitable for species that do otherwise, e.g. tropical species. A single species is assumed unless stated.

The relevance of indicators to the management of fisheries has become more widely acknowledged in recent years. In Northern Europe at least, the scientific community advising on fisheries management has mostly relied on quite complicated models to assess the state of fish resources and make recommendations. This approach entails significant costs in terms of data volume and quality, and of expertise as well, and is only affordable for the top-valued species and fisheries. Nowadays, other species that were of secondary interest in the past have become the primary targets of fishers, and scientific advice on the conditions for their sustainable exploitation is also sought by managers. In most instances, however, the rich data bases required for the conventional assessment models are lacking; other routes have to be considered, such as indicator-based methods. Moreover, States have embraced the "ecosystems approach to fisheries management" which, however defined, basically implies that the interactions between ecosystems' states or processes and fisheries have to be accounted for when decisions are made for fisheries. There is a broad consensus within the scientific community that building detailed quantitative models, continuing the traditional fishery science approach, to advise ecosystems-based management is simply not an option given the gaps in knowledge and the huge cost of getting the appropriate data. The alternative route is to identify and monitor a suite of indicators that reflect the state of, and possibly the human pressure on, the marine systems; management action is then advised based on observed changes in the indicators. Considerable scientific efforts are underway concerning indicators, as evidenced by a burgeoning literature.

There are at least two potential difficulties in this context. Firstly, like other environmental issues, fisheries issues are highly controversial. It can be expected that when systems of indicators – or indicator-based assessments – eventually get included in the management decision framework, they will be exposed to strong reactions, perhaps more so than traditional fish stock assessments, if only because they will be novel. A prime concern, therefore, is that the process to infer the state of fish stock, ecosystems, etc. from indicators should be formalised, in the sense of being rational, objective, defensible, and replicable by others; it should be amenable to non-ambiguous descriptions that are intelligible to stakeholders and managers.

Secondly, marine ecosystems and fish stocks are notoriously subject to high variability. Yet, it is quite likely that several indicators will be based on relatively limited samples (in terms of size, frequency, seasonal or area coverage). This will be particularly true for indicators collected during scientific sea surveys, which are already an important recourse for monitoring fish stocks when commercial fishery data are lacking or deteriorating (the topic dealt with by the FISBOAT project), and will be the prime – or even the sole – source for several ecosystems indicators. Surveys mobilise costly vessels and typically involve a limited number of stations and samples of moderate size, thus survey indicators often have large CVs. In any case, the signal-to-noise ratio is likely to be low. Procedures therefore have to be found to avoid casting measurement noise straight into the advice, and triggering undue action with all the political fuss that may follow.

For the present, indicators could be used to supplement existing methods of model-based, single-species stock assessment and management (Demaré 2006). This would incorporate some additional biology into what is otherwise mainly a computational exercise. In due course, indicators are likely to form a fundamental part of an ‘ecosystem approach to fisheries’ (EAF) (Garcia and Cochrane 2005), Jennings (2005), Cury and Christensen (2005). Indicators might also be used pragmatically to inform a management system based on harvest or effort control rules negotiated relatively from year to year between management and industry. A suite of well-chosen indicators are envisaged to tune such a system so that, after an initial period of trial and error, indicators relevant to the health of the stock respond somewhat predictably to management actions. This would be a form of adaptive management (Walters 1986).

This manual is intended as a contribution to the necessary formal structure for using indicators to assist the management of fisheries. A preliminary, general section, section 2, discusses the appropriateness of the surveys used, and some possible estimators that can give different results from the same survey. Section 3 summarises a small selection of potentially useful indicators of the biological state of a stock of fish. Section 4 considers indicators of the spatial state of a fish stock, most of which allow for the occurrence of low or zero densities of a species of fish at some stations, making them very generally applicable. The last section, section 5, presents methods to assist with the interpretation of time-series of different indicators, singly and in combination. This aspect is important because indicators tend to be highly specific, so that use of many is often necessary to gain a full picture of a stock (Rice and Rochet 2005). Additionally, the interpretation of indicator series often depends more on their trends up or down over time than on their absolute values (Jennings and Dulvy 2005; Trenkel *et al.* 2007). The general problem is how to assemble all the different results in a way that is informative and suitable for justifying decisions about management of the fishery.

Table 1.1 is a list of the state indicators described in this manual, the marine environmental processes and population characteristics that they relate to, and the primary authors from the FISBOAT project in each case. The indicators are described briefly using a standardised format; each relates to a specific biological or spatial characteristic of the stock, or of selected age or length classes within it. In several cases, the indicator described is one example from a suite of related indicators; where possible, references are provided to allow further information to be followed up. The manual includes but is not limited to indicators and methods trialled during the FISBOAT project, 2005 to 2007. The methods trialled were limited to those applicable to most of the project case studies with the available data and within the available time. The FISBOAT project also considered non-indicator based

methods of stock assessment using surveys; these methods are described in a related project manual entitled 'Review of fishery-independent assessment methods' prepared by Mesnil (ICES CM2007/O:04).

References

Anonymous (2004) Report of the workshop on survey design and data analysis (WKSAD). *International Council for the Exploration of the Sea*, Copenhagen. ICES CM 2004/B:07, Ref. D.G., 261pp.

Anonymous (2006) Report of the study group on survey trawl standardisation (SGSTS). *International Council for the Exploration of the Sea*, Copenhagen. ICES CM 2006/FTC:05, 67 pp.

Cury, P.M. and Christensen, V. (2005) Quantitative ecosystem indicators for fisheries management. *ICES Journal of Marine Science* **62**, 307-310.

Demaré, W. (2006) Fisheries-based management and advice in Europe. In: *The knowledge base for fisheries management*. ed Motos, L. and Wilson, D.C., Elsevier: 295-305.

Garcia, M. and Cochrane, K.L. (2005) Ecosystem approach to fisheries: a review of implementation guidelines. *ICES Journal of Marine Science* **62**, 311-318.

Jennings, S. (2005) Indicators to support an ecosystem approach to fisheries. *Fish and Fisheries* **6**, 212-232.

Jennings, S. and Dulvy, N.K. (2005) Reference points and reference directions for size-based indicators of community structure. *ICES Journal of Marine Science* **62**, 397-404.

Rice, J.C. and Rochet, M.-J. (2005) A framework for selecting a suite of indicators for fisheries management. *ICES Journal of Marine Science* **62**, 516-527.

Trenkel, V.M., Rochet, M.-J. and Mesnil, B. (2007) From model-based prescriptive advice to indicator-based interactive advice. *ICES Journal of Marine Science* **64**, in press.

Walters, C. (1986) *Adaptive management of renewable resources.*, Macmillan Publishing Co., New York

Table 1.1 List of the indicators documented in this manual, the marine environmental processes and population characteristics that they are thought to relate to, and the primary authors of text.

Indicator and abbreviation	Processes affecting	Population characteristics	Contributors
Intrinsic population growth rate, r	Fishing, natural mortality, reproduction	Numerical abundance summed over all ages	Trenkel
Total mortality, Z	Fishing, natural mortality, migrations with age into or out of survey area	Rate of dying, migrations related to age	Rochet, Trenkel, Cotter
Spawning stock in number, SSN	Maturation, fishing, natural mortality, nutrition	Abundance of potentially breeding fish, sustainability of the stock	Mesnil, Uriarte, Witthames
Length statistics, L_{bar} , L_{25} , L_{50} , L_{75}	Recruitment, growth, fishing, natural mortality	Growth, length frequency distribution, recruitment	Trenkel, Mesnil, Cotter
Total weight caught, W	Fishing, natural mortality, growth, feeding	Numerical abundance, age composition, growth	Cotter
Condition, C	feeding, growth	Nutritional status of individuals, reproductive fitness	Cotter, Witthames
Gonadosomatic index, GSI	feeding, maturation	Nutritional status, reproductive fitness	Cotter, Witthames
Length and age at maturity, $LaM50$, $AaM50$	maturation, fishing mortality, evolutionary selection	Size and age of potentially breeding fish.	Rochet, Trenkel, Witthames, Cotter
N-at-length, N-at-age, NaL , NaA	Recruitment, growth, fishing, natural mortality	Length and age frequency distribution	Cotter
Centre of gravity, CG	Migrations, climate change, fishing, population size	geographic location of the whole population or of concentrations of fish	Wuillez, Rivoirard, Petitgas
Inertia, I	Dispersal, environmental change, migrations	Changing population size, migrations, climate and environmental changes	Wuillez, Rivoirard, Petitgas
Anisotropy, An , Isotropy, Is	Depth, currents, proximity to shore	Alignment of the population in relation to environmental gradients	Wuillez, Rivoirard, Petitgas
Global index of collocation, GIC	Competition, genetic differences, dispersal	Geographical overlap of two populations	Wuillez, Rivoirard, Petitgas
Number of spatial patches, NOP	Dispersal, common attractants, lack of mixing	Patterns of movement, foraging strategies, population size	Wuillez, Rivoirard, Petitgas
Positive area, PA	Dispersal without regard to variations of abundance	Population size, habitat preferences, food availability	Wuillez, Rivoirard, Petitgas
Spreading area, SA	Dispersal with regard to variations of abundance	Population size, habitat preferences, food availability	Wuillez, Rivoirard, Petitgas
Equivalent area, EA	Dispersal assuming uniform abundance	Population size	Wuillez, Rivoirard, Petitgas
Microstructure index, MI	Small-scale variability of habitat, abundance	Relationship of population to environment	Wuillez, Rivoirard, Petitgas

2. Surveys and estimation

2.1 Introduction

This preliminary section is intended to be general to the use and interpretation of indicators from fish surveys. In section 2.2, Trenkel warns of dangers arising from placing excessive reliance on survey results for the estimation of indicators without first carefully considering whether the survey will in fact provide appropriate data for the species and indicators of interest. In section 2.3, Cotter discusses alternative estimators that can give widely differing results for an indicator using the same survey data.

2.2 Potential limitations of survey data as the unique source of information for stock assessment and management

V. Trenkel

IFREMER, Nantes, France

Introduction

There are several issues that need to be considered when using survey data as the unique source for carrying out stock assessments and providing management recommendations. The main issues leading to bias or uncertainty regarding evolution of the stock are

1. the surveyed area does not encompass the stock area;
2. the size/age classes sampled are not representative of the stock;
3. variation in survey catchability.

Survey area \neq stock area

There are various reasons why survey areas might not encompass stock areas, in addition to the problem of stock boundaries not being well known, or survey areas varying between years. The simplest reason is that part of the stock, or certain age classes, are outside the survey area. No single survey will cover the whole stock area for geographically wide spread species such as northern hake. For other species the problem might be that certain age groups are too deep to be caught by the survey gear, or too shallow for the survey vessel to access them, or they are not accessible to the survey gear because their habitat is for example not trawlable. Anchovy in the Bay of Biscay is an example of a species with a variable proportion of recruits too close to the coast and thus in too shallow water for the survey vessel being used. The visible effect of this is that numbers at age 2 are higher than numbers at age 1 in the previous year (Fig 2.2.1a). Similarly for cod in the North Sea IBTS survey as shown by the negative mortality rates for age 1 in many years (Fig 2.2.1b). The negative Z for cod age 5 are probably an effect of small sample sizes. The same effect can of course be caused by gear selectivity so that in each case it is necessary to find the most plausible explanation.

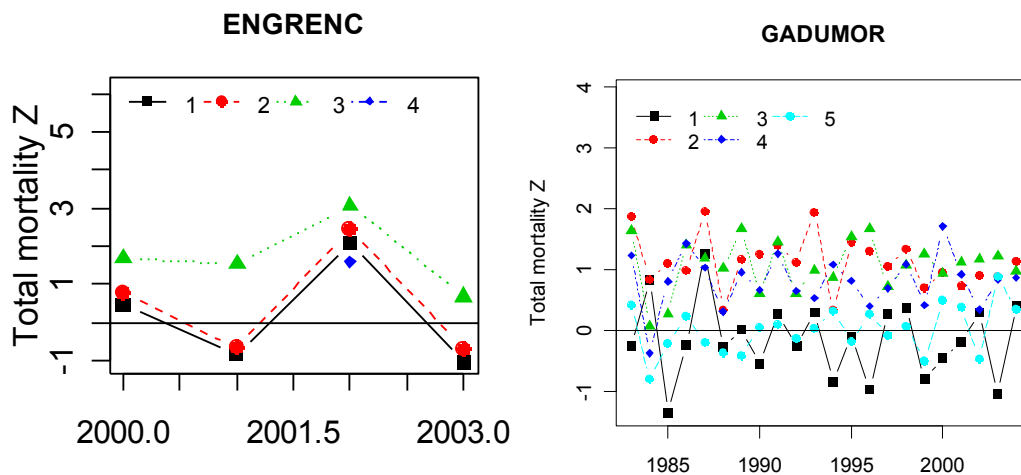


Figure 2.2.1. Total mortality estimates Z by age and year derived from survey numbers-at-age. Z estimates for age a in year t refers to the total mortality between age a in t and age $a+1$ in $t+1$. a) anchovy in Bay of Biscay; b) cod in North Sea.

In addition to the above issues, stocks might move out of, or into the survey area in response to changing environmental conditions. Diel migrations or other activity patterns can also lead to variability in availability to the survey gear.

Size classes sampled not representative of stock

Many scientific surveys have been designed as young fish surveys. For example, what is called today the *International Bottom Trawl* survey (IBTS) started off as the *International Young Herring Survey* (IYHS) in the North Sea, then became the *International Young Fish Survey* (IYFS) before finally obtaining its current name. The change in objectives reflected in the varying names did not imply any change in design, rather a modification of the list of species for which information was collected. Hence in response to the initial objectives, a sampling trawl designed for catching young fish is still used today (GOV 36/47). The time of year of the survey was decided similarly. The initial survey took place in the first quarter as herring juveniles are then available. Currently a third quarter survey is also carried out.

When the IBTS survey was extended to the Bay of Biscay and Celtic Sea (French EVHOE survey), the same GOV trawl was adopted despite the fact that substrates are often more difficult for trawling and the GOV is more suited for soft bottoms. The GOV was slightly adapted by removing the exocet Kite and replacing it by 6 additional floats. As the main target species are hake, megrim and monkfish, the survey is carried out in the fourth quarter when the recruits of those species become accessible.

The consequence of designing surveys to target recruits is that there can be the problem of exploited size classes not being well represented in the survey catches. This can be due to the survey gear being used (selectivity), the vessel speed, or of course an area mismatch dealt with above.

As an example, consider the length distribution of hake in landings and in the EVHOE autumn groundfish survey (Fig 2.2.2). In the survey the bulk of individuals is smaller than 20 cm while in the landings the distribution is rather flat between 30 and 60 cm. Note that the legal landing size is 27 cm. The survey catches very few individuals in the size range targeted by the fishery. Thus, for northern hake, it seems that the EVHOE survey might be suitable to provide recruitment (age 0) and perhaps age 1 estimates, but it is unlikely that variations in total stock abundance or other indices relating to the adults will be captured reliably.

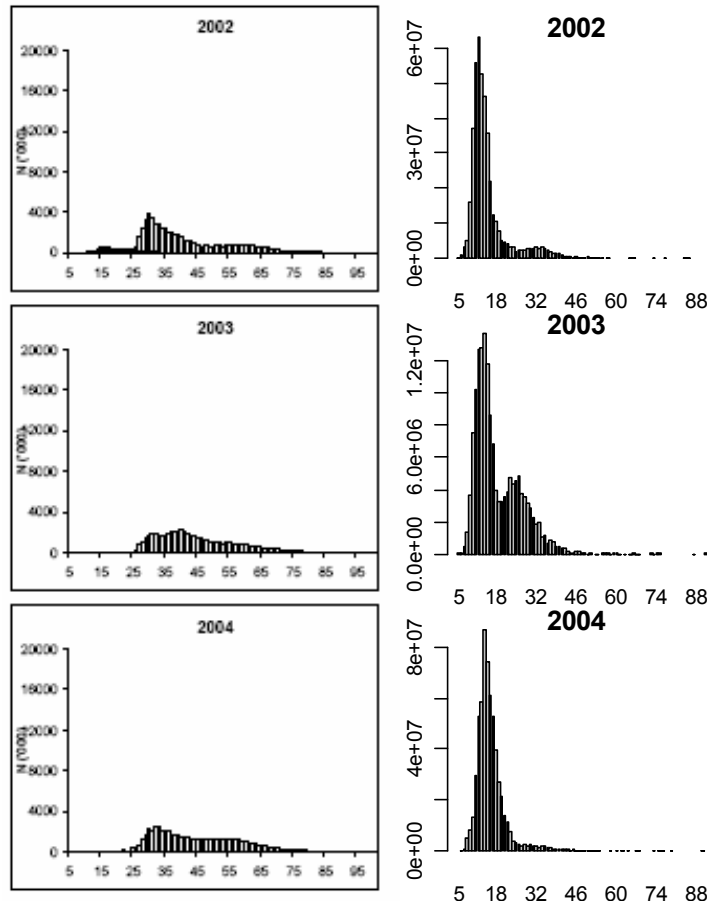


Figure 2.2.2. Length distribution of hake landings for the northern stock (left) and hake caught in the autumn groundfish survey in the Bay of Biscay and Celtic Sea (right).

Variations in survey catchability

A range of factors can make survey catchability vary between hauls and interannually. Between-haul variability will most likely reduce the precision of survey indices while interannual variation might bias estimates and affect time trends. The latter might be called a year-effect in survey catches. A study of the potential year-effect in survey catches for the EVHOE autumn groundfish survey taking place in the Bay of Biscay showed that, on average, 20% of interannual variation in abundance indices could be explained by survey conditions for benthic species, 11% for demersal, and none for pelagic species (Poulard and Trenkel, submitted) (Fig. 2.2.3). In contrast, survey conditions explained a smaller and decreasing part of the interannual variability in the coefficients of variations of these abundance indices and in species mean weight. Thus survey conditions might bias survey indices. In the same study it was found that taking account of survey conditions could alter time trends in species' abundance indices and, as a consequence, influence the stock assessment based on survey information.

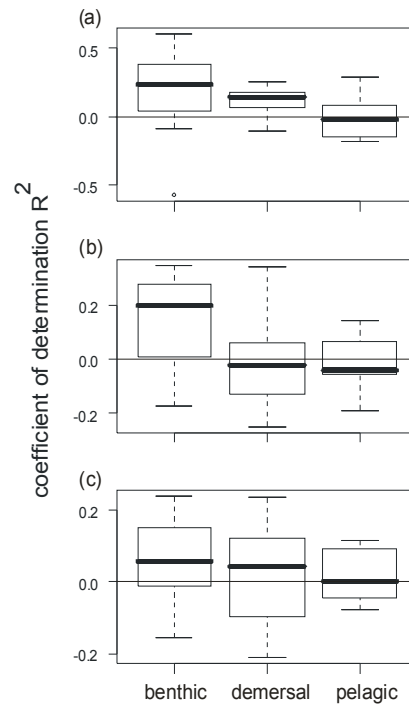


Figure 2.2.3. Box plots with coefficients of determination R^2 by species for best fitting models explaining survey indices by environmental (wind) conditions and survey design: (a) average survey density; (b) CV of average survey density and (c) individual mean weight. Results are grouped by habitat type: benthic, demersal (near sea floor) and pelagic.

References

Poulard, J.-C. & Trenkel, V. M. (submitted). Do survey design and wind conditions influence survey indices ? *Canadian Journal of Fisheries and Aquatic Sciences*.

2.3 Estimation of indicator values from surveys

J. Cotter

Cefas, Lowestoft, UK.

Introduction

When using a survey to estimate average values of an indicator over a stock or a sampling stratum of a survey, the statistical method of estimation used can significantly affect the computed results. In this regard, indicators of quantity and attributes present slight differences in theory. This is most easily seen with simple numerical examples.

Indicators of quantities of fish

Firstly, consider indicators of abundance (or weight) and suppose for simplicity that there are just three fishing stations in the survey area, yielding 1, 3, and 0 fish. The mean density per station over the survey area is $4/3 = 1.33$. This estimate is responsive both to the number of zero-yielding stations and to the densities of fish observed when fish are found to be present. On the other hand, the domain occupied by the stock implicitly includes only the two non-zero values, so that the mean density per station over the stock domain is $4/2 = 2$. This different estimate is only responsive to the densities of fish when found to be present, while reduction of stock abundance is estimated by geographic contraction of the stock domain. Of course, the catching of zero fish does not confirm that fish are absent, so the estimated stock domain is subject to sampling error as for the estimated mean density. We could distinguish the two estimates by calling them the *survey mean density per station* and the *stock mean density per station*. Indicators of the area occupied by a stock are presented in the section on spatial indicators.

Indicators of attributes of the fish

Next, consider indicators based on measured attributes of the individual fish caught, e.g. their lengths, and suppose for simplicity, using the same example, that the single fish at the first station was 50 cm, and the three fish at the second station were 10, 12, and 15 cm for which the station average is 12.33 cm. Including the third station where zero fish were caught makes no sense when averaging attributes, so the ‘survey mean length’ is unimportant. However, we can estimate a *stock mean length* as (1) a mean for the fish or (2) a mean for the stations, i.e.

1. as $(50 + 10 + 12 + 15)/4 = 21.75$, or

2. as $(50 + 12.33)/2 = 31.17$.

Note that the first estimate is weighted towards values observed at the station yielding most fish; it is a ratio estimator because it uses the number of fish at a station as a covariate of length to improve the precision of the estimate. The second estimate gives equal weight to the average value at each station without regard to how many fish were caught (given that at least one was caught). Choice between the two estimators may depend on the degree of within-haul correlation (Pennington and Vølstad 1994). A low degree, i.e. a good mix of lengths (in this example) at each station, suggests that the first estimator will be best because more fish implies more information about the stock. A high degree, i.e. long fish predominate at some stations, short at others, suggests the second because, by contrast, more stations implies more information about the stock. We could distinguish (1) and (2) as the *ratio estimator*, and the *station estimator*, respectively, of stock mean length.

A complication arises when attributes such as length or weight are not measured on all individuals caught by a survey but only on samples of the catch at some stations. It is assumed that the fish in the samples are chosen randomly from the fish in the catch. The best estimate of the average length or weight (etc.) at the station comes directly from the catch sample without raising. However, ratio estimates require all fish caught at each station to be included in the calculations. Consequently, catch sampling may rule out the use of ratio estimates, or it may require some sort of approximate fix, e.g. by raising the frequency distribution in the sample to one for the catch. Of course, the statistical properties of the fix may be poor and difficult to establish leading possibly to bias of the station mean, or a downward bias of the variance causing excessive confidence in the mean.

Standard errors

Standard errors for the ratio estimator are available from (Pennington and Vølstad 1994) or (Thompson 1992). Standard errors for the mean of haul results weighted equally depend on whether the stations can be considered as random sampling locations over the stock distribution. If 'yes', the standard deviation uses the usual formula for simple random sampling. If 'no', a spatial surface could be fitted to the haul results and the standard error of the integral derived from the model. Geostatistical estimates may also be feasible using a variogram, and bootstrapping (Beare *et al.* 2002) is another possibility. Standard errors are more complicated when catch sampling has taken place but an analytic approach is offered by Cotter (1998).

For a review of the wide range of estimation methods including variances used for fish surveys, see section 2.6 in Anonymous (2004).

References

Anonymous (2004). Report of the workshop on survey design and analysis (WKSAD). International Council for the exploration of the sea report ICES CM 2004/B:07, Ref. D,G. 72pp.

Beare, D., Castro, J., Cotter, J., van Keeken, O., Kell, L., Laurec, A., *et al.* (2002) Evaluation of research surveys in relation to management advice (EVARES). Final report. *DGXIV Fisheries, European Commission, Brussels*. **FISH/2001/02 - Lot 1**. Available from john.cotter@cefas.co.uk

Cotter, A.J.R. (1998) Method for estimating variability due to sampling of catches on a trawl survey. *Canadian Journal of Fisheries & Aquatic Sciences* **55**, 1607-1617.

Halliday, R.G. and Mohn, R. (2001) Proceedings of the Fisheries Management Studies working group, 8-11 January 2001., Canadian Science Advisory Secretariat, Proceedings Series.

Jennings, S. (2005) Indicators to support an ecosystem approach to fisheries. *Fish and Fisheries* **6**, 212-232.

Pennington, M. and Vølstad, J.H. (1994) Assessing the effect of intra haul correlation and variable density on estimates of population characteristics from marine surveys. *Biometrics* **50**, 725-732.

Thompson, S.K. (1992) *Sampling*, John Wiley & Sons, Inc., New York

3. Biological indicators

3.1 Introduction

Biological indicators are available to measure most aspects of the health of a fish stock in addition to its numerical abundance. These include growth, age composition, fecundity, recruitment, and total mortality. Such indicators are ‘state’ indicators in the ‘pressure-state-response’ (PSR) system for classifying indicators (Jennings 2005). A selection of commonly considered biological indicators is presented here according to a format given by Halliday and Mohn (2001) for which see Appendix 1 of this report. The advice given for each indicator comes from discussions among the participants in the FISBOAT project. It is not intended to replace careful consideration of the relevance and value of each indicator to a particular stock.

3.2 Intrinsic population growth rate, r

Description

The slope of log total abundance against time.

Stock attribute

Numerical abundance summed over all ages.

Derivation

The indicator is Lotka's intrinsic population growth rate, r . It has been suggested for fish populations by (Quinn and Szarzi 1993).

Reference points

Taking $r = 0$ as the target reference point assumes that, without any noticeable impact of fishing, the population would be stable in the long term, even though it varies from year to year.

Interpretability

If r is significantly lower (respectively higher) than 0, the population is decreasing (increasing). The expected and undesired effect of fishing is to decrease r , although many other factors might have the same effect. A long term decline in r suggests that both recruitment and standing stock numbers are declining and implies that remedial action is necessary.

Measurability

r is readily and directly estimable from the time-series of abundance indices produced by a survey for all ages combined. The model is $N_t = N_{t-1} \exp(r)$. It can be estimated by fitting the mixed model, $\log N_t = \beta_0 + rt + \omega(t) + e_t$ where $\omega(t) \sim \text{Normal}(0, \sigma^2)$ represents the year to year variance, and $e_t \sim \text{Normal}(0, \sigma_e^2)$ represents random error. The estimate of r will depend on the time-window chosen since the linear slope measured by r is merely an average of fluctuations over time. The fitted linear slope acts as a smoothing function which may give a more stable indication of stock abundance trends than the time-series of raw abundance indices. Deviations from linearity could affect measurability adversely.

Sensitivity

Since r is a measure for all ages combined it is likely to be affected by large recruitment pulses, particularly if numbers of adults are low.

Example

Not available.

References

Quinn, T.J. and Szarzi, N.J. (1993) *Determination of sustained yield in Alaska's recreational fisheries*. International symposium on management strategies for exploited fish populations., Alaska sea grant college program, University of Alaska, Fairbanks, Alaska.

3.3 Total mortality, Z

Description

The coefficient of total mortality averaged over a given age range, i.e. the average mortality rate between years $t-1$ and t of all individuals aged a_{\min} to $a_{\max-1}$.

Stock attribute

Z measures mortality due to fishing and natural causes. It is also affected by loss or gain of fish from a survey area as a result of a net migration occurring in relation to age.

Derivation

Z is the sum of natural and fishing mortalities: $Z = F + M$. It comes from the exponential model of mortality in population dynamics. The coefficient of total mortality, Z , is defined by $\frac{dN}{dt} = -ZN$ where N is abundance, t is time, and Z is conventionally regarded as positive. This solves to $N_t = N_0 \exp(-Zt)$ from which (i) $\log N_t = \log N_0 - Zt$, or (ii) $Z = -\log(N_t/N_{t-1})$.

Reference points

Z during a period of acceptable fishing mortality.

Interpretability

Z has been suggested as a robust indicator for exploited populations (Die and Caddy 1997). Different Z over different age ranges could be caused by less than full selectivity of the survey gear for some ages, or by migrations related to age particularly if the survey only covers part of the known range of the stock (Cotter *et al.* 2007). Interpretation of Z requires that M be assumed constant if, as is usually the case, it is not known accurately.

Measurability

Z may be estimated by fitting linear regressions to log abundance indices by year-class over age, equation (i) above, the so-called year-class curve method (Cotter *et al.* 2007). Removal of the youngest, and possibly the oldest ages may be necessary to find a satisfactory linear fit. Standard errors are available from the linear modelling. Alternatively, Z may be estimated separately for each (age, year) to (age+1, year+1) using equation (ii) above (Beare *et al.* 2002), and averaged over consecutive ages.

Sensitivity

Changes in Z over time are only likely to be discerned by surveys when commercial fishing effort changes substantially (Cotter, 2001). Changes in Z regionally resulting from net migration from one region to another as the fish grow older can be detected for plaice (Cotter *et al.* 2007). It is seldom possible to discern effects of changing M on Z except for year classes that are not susceptible to the fishery.

Example

See references.

References

Beare, D., Castro, J., Cotter, J., van Keeken, O., Kell, L., Laurec, A., *et al.* (2002) Evaluation of research surveys in relation to management advice (EVARES). Final report. DGXIV Fisheries, European Commission, Brussels. *FISH/2001/02 - Lot 1*. Available from john.cotter@cefas.co.uk

Cotter, A.J.R. (2001) Intercalibration of North Sea International Bottom Trawl Surveys by fitting year-class curves. *ICES Journal of Marine Science* **58**, 622-632 [Erratum, Ibid. 58:1340].

Cotter, A.J.R., Mesnil, B. and Piet, G. (2007) Estimating stock parameters from trawl cpue-at-age series using year-class curves. *ICES Journal of Marine Science* **64**, 234-247.

Die, D.J. and Caddy, J.F. (1997) Sustainable yield indicators from biomass: are there appropriate reference points for use in tropical fisheries? *Fisheries Research* **32**, 69-79.

3.4 Numbers-at-length, numbers-at-age: NaL, NaA

Description

The length or age frequency distribution.

Stock attribute

Age structure and growth of the population. Relative numerical strengths of different year classes.

Derivation

Frequency distributions may be estimated for individual stations, strata, or for the survey as a whole.

Reference points

A sustainable stock should have a reasonable proportion of larger or older individuals capable of breeding, as well as to allow the commercial fishery to sustain itself when recruitment is poor. Protection of size or age classes just in the breeding category may not be sufficient if young fish tend not to be successful at breeding, or if the stock is poorly nourished such that individuals may not breed even though they have reached an age when they could.

Interpretability

A stock lacking large or old fish is likely to be over-fished and under-productive economically. However, similar effects could also occur due to predation or disease for example. Knowing indices of the abundance of predator-, and perhaps prey-species could be useful for interpretation of trends in frequency distributions. Increasing proportions of large or old fish implies better survival with age and may signal recovery of a stock.

Measurability

- NaA is harder to measure than NaL since otoliths have to be read.
- In general, RV surveys use gear with very small mesh so that most size and age groups will be fully selected. However, allowances must be made to frequency distributions if this cannot safely be assumed. Variability of the period when young fish settle to the bottom (and become vulnerable to a demersal trawl) relative to the survey period could affect estimated frequency distributions, as could the escape of large, fast-swimming individuals.
- The following comments follow from the note on estimation in the preliminary section of this manual. An age- or length-frequency distribution compiled from all fish of a species caught on a survey (or in a sampling stratum) will be most influenced by the frequency distributions prevailing at the stations where most fish were caught. [This estimate is to frequency distributions as the ratio estimate is to means.] This estimate would be preferred if the highest yielding stations are thought to provide the best indication of the frequency distribution for the whole stock.
- The alternative is to estimate a probability density for numbers-at-length or -at-age at each station where fish were caught, i.e. a histogram scaled to integrate to unity. Then the probability density for the stock (or a stratum) is obtained by averaging the station densities for each length or age class. [This estimate is to frequency distributions as the station estimate is to means.] This estimate would be preferred if length frequencies

3. BIOLOGICAL INDICATORS

differ noticeably between stations and no single station is thought to be more representative of the stock than another.

Sensitivity

In general, length and age composition of a stock change rapidly in response to fishing pressure, then numbers of older, larger fish remain low. Both indices are affected by pulses of recruitment, and this may be a more influential factor than slight changes of fishing when fishing pressure is high.

Example

Not available

References

3.5 Spawning Stock in number: SSN

Description

Mean catch in number of mature fish per tow (or per standard time or distance unit) realised during a survey.

Stock attribute

Abundance of fish with potential to breed. Sustainability of the stock.

Derivation

For trawl surveys, SSN is probably most simply computed in the same way as the usual survey abundance index but with numbers-mature at each fishing station substituted for numbers, or numbers-at-age.

Egg surveys directly provide an index of spawning stock biomass (SSB). If appropriate corrections are applied for mortality to the estimates of egg abundance, and if estimates of daily specific fecundity are available, then the index can be taken as an estimate of absolute SSB. Otherwise, it will only estimate relative SSB. SSB is converted to SSN by dividing by the mean weight of the mature fish in the population. The latter parameter, as for the fecundity parameters, is obtained by averaging the estimates from the individual fishing hauls, using weighting factors proportional to the egg abundance divided by the mean weight of the anchovies in the haul.

Reference points

Possibly, lowest historically observed estimates known to sustain a satisfactory recruitment.

Interpretability

- The index will probably be most closely related to numbers of the younger age classes that are mature since their abundance will usually be much higher than that of larger, older fish, depending on rates of total mortality, and annual recruitments.
- Changes in SSN are likely to be due to fishing but could also be caused by natural events.
- Further understanding of the health of a stock may be acquired by looking at indices for first-time and for replicate spawners separately: stocks subject to high F or high M generally show higher proportions of first-time spawners.

Measurability

Accurate assessment of maturity is crucial (Kjesbu *et al.* 2003; Murua *et al.* 2003) particularly as the proportion mature is likely to be most hard to estimate accurately for the most numerous age classes just coming into maturity. See also www.ices.dk/datacentre/datras/NSIBTSmanualRevVIIIdraft.pdf. Careful standardisation of the maturity assessments across years and across survey crews is essential to avoid step changes in time or space purely as a result of inconsistent technique. Objective histological methods can be helpful for quality control of maturity assessments based on external morphology. The seasonal timing of the survey is also very important. Summer surveys can be especially poor for estimating maturity of species that spawn in winter or spring unless histological criteria are used to hindcast or predict maturation in the previous or next spawning season.

Sensitivity

SSN indices are likely to reflect changes in the age composition and reproductive potential of a stock over time, though precision will depend on standardisation of the techniques of maturity staging.

Example

Not available.

References :

Kjesbu, O.S., Hunter, J.R. and Witthames, P.R. (2003) Report of the working group on modern approaches to assess maturity and fecundity of warm- and cold-water fish and squids. *Institute of Marine Research*, Bergen, Norway. **Fisken og havet**, **12**, 140 pp.

Murua, H., Kraus, G., Saborido-Rey, F., Witthames, P.R., Thorsen, A. and Junquera, S. (2003) Procedures to estimate fecundity of marine fish species in relation to their reproductive strategy. *Journal of northwest Atlantic Fisheries Science* **33**, 33-54.

3.6 Length statistics: L_{bar} , L_{25} , L_{50} , L_{75}

Description

Mean (L_{bar}) or percentiles (L_{25} , etc) of fish lengths found in survey catches.

Stock attribute

Growth, length frequency distribution of stock, recruitment success annually.

Derivation

Descriptive statistics for the length frequency distribution.

Reference points

Options are:

- Length at 50% maturity.
- Lengths associated with biological events such as migrations.
- Lengths associated with the fishery, e.g. length at 50% selectivity, minimum landing size (MLS).
- Historic values when the stock was deemed to be at a satisfactory level.

Interpretability

- In general, high fishing intensity reduces the relative abundance of large compared to small fish for two reasons: (a) large individuals are mostly older and therefore have had more exposure to fishing gear, and (b) because trawlers tend to catch large fish more effectively than small.
- The L_{25} , L_{50} , and L_{75} percentiles characterise the smaller, middle sized, and older fish, respectively, and are therefore expected to respond differently to recruitment pulses, growth factors, and to changes in abundance and spawning stock biomass.
- A short selection of references relevant to interpretation of fish length are by Ault et al. (2005), Kvamme and Froyso (2004), Deriso and Parma (1988), Piet and Jennings (2004), Jennings et al. (1999), and Rochet and Trenkel (2003).
- L_{bar} for mixtures of species has been shown to decrease in exploited communities (Jennings et al. 1999).

Measurability

Please refer to the note on estimation in the Introduction to Biological indicators. Care should be taken to standardise the bodily extremities to be used to measure each species, particularly across national surveys, and across years (Beckett 1983). Estimation of L_{bar} with omission of fish shorter than a given length, e.g. the length best separating 0 and 1-groups, or the MLS, is likely to improve its precision considerably when annual recruitments are very variable.

Sensitivity

L_{bar} is most influenced by the smallest, youngest year classes included in the estimate because they are usually the most numerous. Use of a ratio estimator, rather than a station estimator (see section on Estimation of indicators from surveys, above) would enhance this effect. It follows that entry of a new year class into the estimate could alter L_{bar} over only one year if that year class had been affected by a step-change of conditions affecting growth

or abundance. However, the full effect on L_{bar} would not be manifest until all living ages had experienced the change. L_{25} is affected by the number and size of young fish; thus low recruitment of young, small fish would be expected to cause L_{25} to increase in the same year, and vice versa. L_{50} would be expected to respond in a similar way to L_{bar} . Fish longer than L_{75} are likely to belong to several age classes. Therefore, L_{75} is likely to decrease gradually over years in response to selective removal of predominantly large individuals by fishing faster than they are replaced by growth.

Example

Fig. 3.6.1 illustrates length-based indicator series derived from the EVHOE demersal trawl survey of the Bay of Biscay, one for anchovy and one for hake. No trends are evident in the L_{bar} or L quantile series for either species but, curiously, a linear decline in length variance could be fitted to the variance of length for hake. The explanation for this is not known.

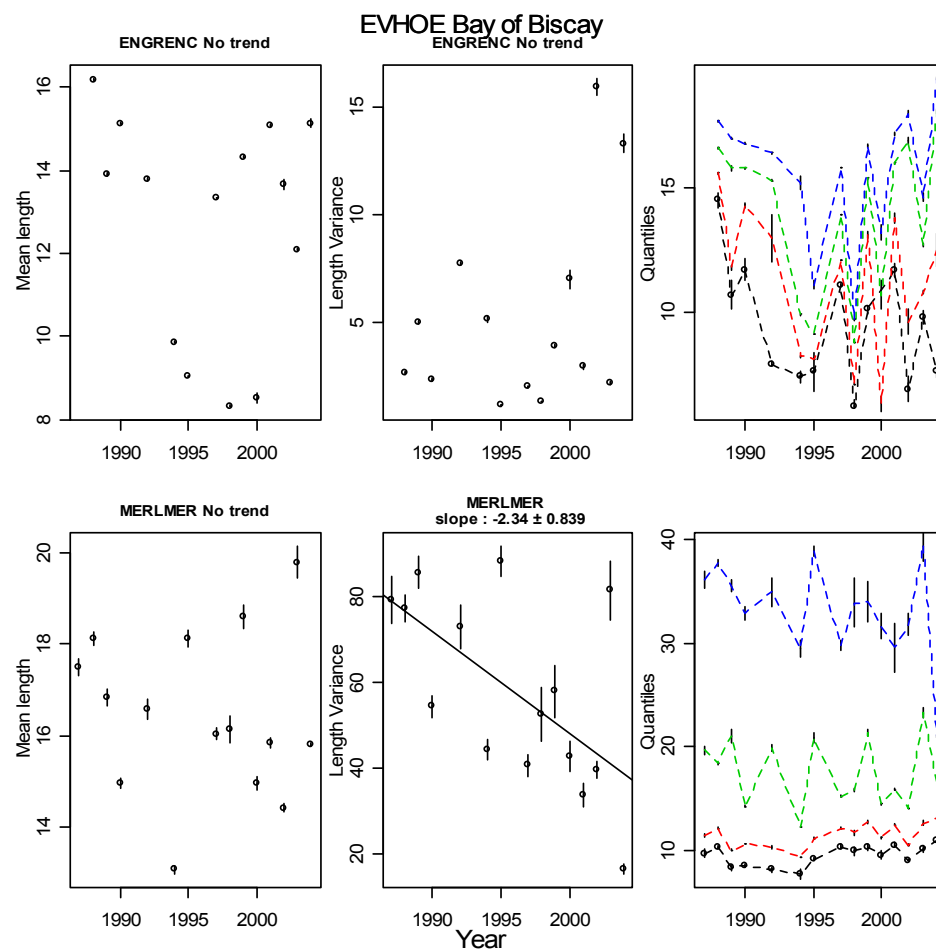


Figure 3.6.1: Time series of indicators for Bay of Biscay anchovy (code= ENGRENC) and hake (code= MERLMER) with long term trends. Data from the IFREMER EVHOE demersal trawl survey of the Bay of Biscay.

References

Ault, J. S., Smith, S. G., and Bohnsack, J. A. 2005. Evaluation of average length as an estimator of exploitation status for the Florida coral-reef fish community. *ICES Journal of Marine Science*, 62: 417-423.

3. BIOLOGICAL INDICATORS

- Beckett, J.S. 1983. Standards used for the sampling of commercial catch under ICNAF/NAFO. In *Sampling commercial catches of marine fish and invertebrates*. Ed. Doubleday, W.G. and Rivard, D. Canadian Special Publications in Fisheries and Aquatic Science, 66: 251-254.
- Deriso, R.B. & A.M. Parma. 1988. Dynamics of age and size for a stochastic population model. *Can. J. Fish. Aquat. Sci.* 45: 1054-1068.
- Kvamme, C. & K.G. Froyso. 2004. Assessing the effects on stocks of selectivity changes in a fishery. *Fish. Res.* 69: 283-292.
- Jennings, S., S.P.R. Greenstreet & J.D. Reynolds. 1999. Structural change in an exploited fish community: a consequence of differential fishing effects on species with contrasting life histories. *J. Anim. Ecol.* 68: 617-627.
- Piet, G. J., and Jennings, S. 2005. Response of potential fish community indicators to fishing. *ICES Journal of Marine Science*, 62: 214-225.
- Rochet, M.J. & V. M. Trenkel. 2003. Which community indicators can measure the impact of fishing? A review and proposals. *Can. J. Fish. Aquat. Sci.* 60: 86-99.

3.7 Survey catch weight: W

Description

The total weight of one (or more) species caught on a survey with constant fishing effort.

Stock attribute

Stock biomass, size composition.

Derivation

Many surveys routinely weigh catches by species. If such data are not available, weights may be estimated for each species using the allometric formula: $W = \sum_l n_l a l^b$ where l is length class, n_l is the numbers caught during the entire survey in that length class, and a and b are constants for the species, the first being a scaling factor for units, the second being a factor relating to change of shape with increasing size. a and b are determined by plotting $\log(W_l/n_l)$ against $\log l$ using a sample of weighed and measured fish – more details are given for the Condition index, below. The estimates of a and b should be up-to-date. An even simpler weight index derived from length data may be adequate for some species; i.e. assume that $W \propto \sum_l n_l l^3$. In this formula, $b = 3$ implies no change of shape with size.

Reference points

A possibility is to use historical estimates of total survey weight when the stock was considered to display a satisfactory age composition, i.e. one having sufficient mature age groups present to provide resilience to occasional poor recruitment or temporary, heavy fishing pressure.

Interpretability

- Condition of the fish (weight/length) may be important for interpreting W but only if fish weights are measured directly, rather than being estimated from lengths.
- The quality of information that W provides about the weight of fish in the stock depends partly on the relative catchabilities of different sized fish by the survey; e.g. catchability of large individuals of strongly swimming species may be low on surveys with short tows.
- Calculation of W might omit fish shorter than the minimum landing size (MLS) for the fishery so that W becomes of direct relevance to the legal yield of the fishery. It might then be serviceable as guidance for a harvest control rule.
- Standardising W as a total weight for the entire survey area is proposed because high yielding stations then have most influence on the result. This could make the index of immediate relevance to commercial fishers who tend to target such localities. On the other hand, comparison of the survey mean weight per station with the stock mean weight per station (see section on Estimation of indicators from surveys, above) while also noting any changes in the estimated domain of the stock may be more relevant for ecological studies.
- W could also be translated into a mean individual weight. This would be akin to a condition index (see C below) but without allowing for body length. The ratio and the station estimator of the stock mean individual weight might then have different interpretations due to the different weightings given to stations yielding different

numbers of fish (see section on Estimation of indicators from surveys, above). The ratio estimator could be more relevant for a fishery that focuses on aggregations of fish, while the station estimator could be more relevant for regional, ecological studies.

Measurability

W can be expected to vary a lot for migratory stocks that are not satisfactorily enclosed by the survey domain, or whose distribution in relation to fishing stations varies with abundance. Variation from year to year of stations fished, or in the duration of tows should be allowed for by raising to a standard level of effort equivalent to the complete, standard survey while recognising that bias might have been caused by omission of certain stations. The season of the survey should be kept constant because W will be affected by seasonal changes in availability of food, and by enlargement of gonads for breeding. A constant selectivity of the survey trawl and fishing method with respect to size is crucial. For some ground-loving species, e.g. certain flatfish, distance towed over the ground may be more relevant as a standardising measure for survey CPUE than the duration of the tow in minutes.

Sensitivity

A merit of W is that it is less influenced by varying recruitment from year to year than raw length indices such as L_{bar} because young fish, although very numerous, are very small, so their abundance and weight tend to cancel in the overall index. W would be influenced by changing abundance and average condition of individuals.

Example

Heessen et al. (1997) state that the International Bottom Trawl Survey of the North Sea and Skagerrak covered the whole of the North Sea, Skagerrak and Kattegat from 1974 onwards. Catch-at-length data for quarter 1 were obtained for cod from the ICES secretariat and average catch weight per fishing station per rectangle was estimated using a ratio estimator for each of the 138 rectangles in the standard area used to estimate the abundance index for cod. Weight indices were estimated using $W \propto \sum_i n_i l^3$. The results were summed over all rectangles to estimate W for the whole survey. The time series from 1974 to 2005 is shown in fig. 3.7.1, below, together with the numbers of rectangles where positive catches of cod were taken. The obvious downward trend in W from approximately 1980 is consistent with the well known decline in the North Sea cod stock in recent years. Variable results in the 1970s may possibly have been due to inconsistent survey practices in the early days of this survey. However, no attempt has been made here to remove inconsistencies at any time in the series. See Heessen et al. (1997) for details of inconsistencies over the period.

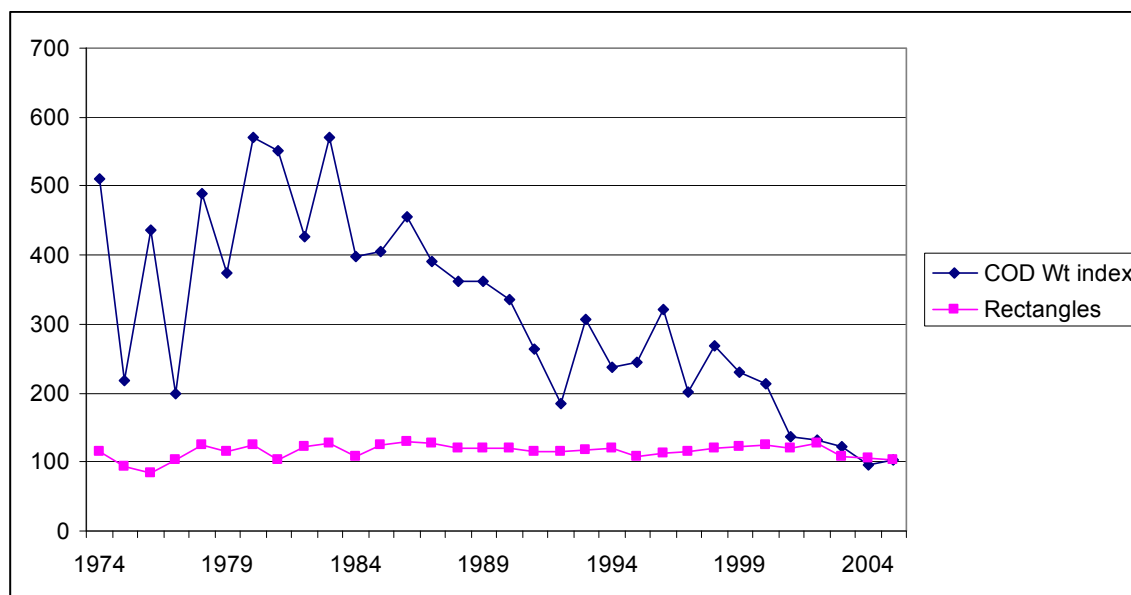


Figure 3.7.1. International Bottom Trawl Survey, quarter 1: cod weight indices, W , computed from numbers-at-length by ICES rectangle assuming isometric growth. Data supplied to FISBOAT by ICES secretariat.

References

Heessen, H.J.L. (1997) The International Bottom Trawl Survey in the North Sea, the Skagerrak and Kattegat. *International Council for the Exploration of the Sea*. ICES CM 1997/Y:31, 25 pp.

3.8 Condition: C

Description

Condition refers to average body weight for a given body size. It is often normalised as Fulton's condition index (Anderson and Neumann 1996), W/kL^3 , where k is chosen for appropriate scaling.

Stock attribute

Nutritional status of individuals, reproductive fitness.

Derivation

C may be estimated by a in the allometric equation $W = \sum_l n_l a l^b$ where l is length class, n_l is the numbers caught in that length class, a measures condition as well as serving to scale the units of measurement, and b represents the changing shape of the species with increasing length. Partitioning this equation by length gives $W_l = n_l a l^b$. Taking logs and substituting C for a gives $\log(W_l/n_l) = \log C + b \log l$. Therefore a regression of \log (average weight per individual at length l) on $\log l$ will allow estimation of $\log C$ as the intercept and b as the slope. In many studies, b is standardised at 3, as for the Fulton index, implying growth without change of shape.

Reference points

Comparison with historic data might reveal when problems are occurring.

Interpretability

- Low condition implies too much competition for available food. This in turn implies that some mature individuals may not mature reproductively for the coming spawning season, or that their fecundity may be reduced by follicular atresia (Thoresen *et al.* 2006; Kennedy *et al.* in press). Low condition can also increase the age of first maturity, and possibly increase natural mortality of post spawning individuals.
- Condition varies in males and females according to season, for example after a winter fast, and over the reproductive cycle especially for species with a capital spawning strategy (Stearns 1992). See also (Lambert *et al.* 2003).
- Low condition implies reduced incomes for fishers because they will not attain high prices for the fish they catch.

Measurability

The regression could be carried out at each station and the estimated C averaged over all stations; i.e. a station-based estimator, giving equal weight to each station yielding fish. Alternatively, the regression could be carried out for all stations at once; this is a regression analogue of a ratio estimator because the results at each station are automatically weighted in relation to the number of fish caught at each length. The weights assigned to each station will therefore vary from length class to length class. Please refer to the section on Estimation of indicators from surveys, above.

Standard errors for a and b estimated from a regression are likely to over-estimate confidence in the result if the individual fish in the sample were not independently and randomly

selected from the stock, as they would not be if taken with a trawl survey for example. This should be kept in mind when making comparisons from year to year. A spatial model could be a good way to summarise station-by-station results because condition may vary with location, particularly with latitude if accompanied by changing temperatures. Kriging is another possibility.

Sensitivity

Not known.

Example

Not available.

References

Anderson, R.O. and Neumann, R.M. (1996) Length, weight, and associated structural indices. In: *Fisheries Techniques*. ed Murphy, B.R. and Willis, D.W. 2nd edn, American Fisheries Society: 447-482.

Kennedy, J., Witthames, P.R. and Nash, R.D.M. (in press) The concept of fecundity regulation in plaice *Pleuronectes platessa* L. tested on three Irish Sea spawning populations. *Canadian Journal of Fisheries & Aquatic Sciences*.

Lambert, Y., Yaragina, N.A., Kraus, G., Marteinsdottir, G. and Wright, P.J. (2003) Using environmental and biological indices as proxies for egg and larval production of marine fish. *Journal of northwest Atlantic Fishery Science* **33**, 115-159.

Thoresen, A., Marshall, C.T. and Kjesbu, O.S. (2006) Comparison of various potential fecundity models for north-east Arctic cod *Gadus morhua* L. using oocyte diameter as a standardizing factor. *Journal of Fish Biology* **69**, 1709-1730.

Stearns, S.C. (1992) *The evolution of life histories.*, Oxford University Press, Oxford, UK

3.9 Gonadosomatic index: GSI

Description

The gonadosomatic index (GSI) is the ratio of gonad weight to body weight.

Stock attribute

Nutritional status, reproductive fitness.

Derivation

Average gonad weight/body weight, or gonad weight/length³.

Reference points

Comparison with historic data might reveal when problems are occurring with regard to stock reproductive potential.

Interpretability

- GSI depends on fish size and will obviously be low for immature individuals. It will also increase rapidly towards the start of the spawning season. It is likely to be highest after ovulation when the first batch of eggs is ready for spawning, and lowest just after spawning when fish have not only lost their reproductive material but are also likely to show low bodily condition.
- GSI may vary with location, particularly with latitude if accompanied by changing temperatures. Spatial modelling or Kriging could be helpful for seeing these effects.
- A low GSI in fish of mature age at the start of the spawning season may imply skipped spawning (Rideout *et al.* 2005), suggesting too much competition for available food, as well as low reproductive success in the coming spawning season, either through lack of fertile adults or through reduced egg production.

Measurability

Surveys intended to estimate GSI should be timed to coincide with the onset of spawning, or, if that is not possible, to avoid the post-spawning period. They should occur in the same season each year. At least a sample of fish from every fishing station should be measured individually for length, body weight, and gonad weight. Stratification by depth bands may be helpful for estimation if GSI is related to depth, as it is for some species (Rijnsdorp 1989). A cut-off length for exclusion of immature individuals would save much pointless dissection work on deck.

Weighing gonads requires that they be removed from the fish and weighed on a balance capable of resolving down to 1% of the gonad mass. Motion compensated balances capable of resolving to 0.1g are required for small fish such as sprat or anchovy weighed at sea. An advantage of GSI over maturity indices is that it does not require accurate maturity staging.

Sensitivity

The GSI is very sensitive to the maturity stage and the timing of the spawning cycle.

Example

Not available.

References

Rideout, R.M., Rose, G.A. and Burton, M.P.M. (2005) Skipped spawning in female iteroparous fishes. *Fish and Fisheries* **6**, 50-72.

Rijnsdorp, A.D. (1989) Maturation of male and female North Sea plaice (*Pleuronectes platessa* L.). *Journal du Conseil international Exploration du Mer* **46**, 35-51.

3.10 Length and age at maturity: LaM50, AaM50

Description

The length or age at which 50% of the individuals in a fish stock are estimated to have reached reproductive maturity.

Stock attribute

Reproductive capability, spawning stock biomass, nutritional status.

Derivation

The estimated proportion mature at each length is plotted against length and the length at which 50% of individuals are mature is the LaM50. Fitting a logistic or other function may be the best way to estimate this statistic using reaction norms (Heino *et al.* 2002). If otoliths are removed at the same time as gonads are examined, and subsequently read without breaking the link with the observation of maturity, the AaM50 can be estimated similarly with a plot over age.

Reference points

Reaction norms are used to describe the phenotypic range of LaM50 and AaM50, and to identify evolutionary selection of alleles favouring early maturation. They may already be in the genetic structure of the population, or they may be introduced by genetic drift and/or mutation (Heino *et al.* 2002).

Interpretability

- AaM50 has been found to decrease under the effect of fishing (Trippel 1995; Rochet *et al.* 2000). If individual growth remains similar under the impact of fishing, LaM50 will decrease in a similar manner. Compensatory growth might, to some degree, reduce the impact of fishing on the observed reduction in length at maturity but strong signals should still be detectable. The reaction norm method is used to investigate evidence for evolution of maturation stage with respect to size and age after several generations of high fishing mortality.
- LaM50 and AaM50 may both vary with location, particularly with latitude if accompanied by changing temperatures.

Measurability

Please refer to Measurability of Spawning stock number (SSN) for comments on estimating maturity and the timing of surveys. Since maturity staging requires that fish be opened and the gonads examined carefully, it is time-consuming on deck.

Sensitivity

Problems with achieving consistent maturity staging from year to year when using only external morphology may seriously reduce the sensitivity of LaM50 and AaM50 to fishery and environmental factors. Better sensitivity could be expected if histological examinations are carried out for each fish but this is obviously much more time-consuming.

Example

Not available.

References

Heino, M., Deickmann, U. and Godø, O.R. (2002) Measuring probabilistic reaction norms for age and size at maturation. *Evolution* **56**, 669-678.

IBTS w.g. (revision VII in draft in 2007). Manual for the international bottom trawl surveys. Appendices VII Finfish maturity key, and VIII Four stage maturity key for skates and rays (Rajidae). www.ices.dk/datacentre/datras/NSIBTSmanualRevVIIIdraft.pdf

Rochet, M-J., P.A. Cornillon, R. Sabatier & D. Pontier. 2000 Comparative analysis of phylogenetic and fishing effects in life history patterns of Teleost fishes. *Oikos* 91: 255-270.

Trippel, E.A. 1995 Age at maturity as a stress indicator in fisheries. *Bioscience* 45: 759-771.

4. Spatial indicators

4.1 Introduction

Spatial indicators are statistics aimed at describing and summarizing the spatial distribution of populations, in terms of location, fish density or possibly an environmental variable e.g. depth. A list of 10 geostatistical indices are here proposed (Woillez et al., 2007) to characterise occupation, aggregation, location, dispersion, correlation and overlap. These notions are somewhat related, e.g., aggregation, dispersion and occupation, and formal relationships exist between indices. The centre of gravity of a population with a measure of dispersion around it had been proposed already (Swain and Sinclair 1994, Atkinson et al. 1997, Bez and Rivoirard 2001). The occupation and aggregation indices are not truly spatial in the sense that they are sensitive to the histogram and not to the spatial location of values. Various indices to characterise aggregation have been suggested (area coverage: Swain and Sinclair 1994, Gini index: Myers and Cadigan 1995, spatial selectivity index: Petitgas 1998) which all relate to the area coverage of highest values. But the spreading index is more general in the sense that the amount of zeroes do not affect this index. Therefore in the calculation of the spreading index the delineation of the data positive domain is not necessary. The spatial indices are useful in characterising the spatial organisation of the life cycle. It can be evidenced that young immature fish, young matures and older matures differ in some aspects of their spatial distributions, in particular for location, aggregation and dispersion (Woillez et al., 2007). The spatial indices have the potential to be used in a monitoring system so as to detect changes in the spatial distributions, which could be helpful in relating the spatial distribution properties of fish stocks to their dynamics, climate change or habitat conservation.

When selecting spatial indicators for this manual, care has been taken to avoid the problem of zero density values by excluding statistics that would depend on the inclusion or exclusion of zero density values according to their belonging to a more or less arbitrarily delineated domain. For instance, the mean of the density values within a given domain is not considered here, nor the variance or the Gini index of these values to measure their statistical dispersion. On the contrary, the contribution of the zero density values is zero in the statistics that have been selected here. For instance, the center of gravity, or mean location, of a population will depend on whether the density value at a sampled location is zero or not, but if it is zero, its numerical contribution to the center of gravity will be zero.

In particular the contribution of zero density values is zero in all statistics based on the individuals of the population: e.g. the mean location of the population, which is the mean location of the individuals that constitute this population (in such a case the statistics are weighted by the fish density). By contrast, the Positive Area looks at where the fish density values are strictly positive but does not depend on the level of these density values, that is, on each individual.

Some of the selected statistics (e.g. the center of gravity) would change if the fish density values were permuted between sampled locations (even assuming a regular sample grid). The other statistics would be unchanged: for instance the Positive Area measures the domain covered by the non-zero density values, not its shape, and it would be unchanged by permuting density values. Similarly the Spreading Area or the Equivalent Area will depend on the histogram of density values, not on their location (at least assuming a regular sample grid). As a consequence these statistics are not dependent on the large-scale spatial structure.

4. SPATIAL INDICATORS

However they do depend on the fine scale structure through the 'support', i.e. the surface in size and geometry (e.g. the trawled area) on which each fish density is measured.

Note: references for the following indicators are collected together at the end of section 4.

4.2 Centre of gravity: CG

Description

The center of gravity is the mean location of the population, that is, the mean of the location of its individuals (Bez, 1997).

Stock attribute

Mean location of the population.

Derivation

Let x be a point in two-dimensional space (short for the usual two-dimension notation (x, y)), and $z(x)$ be the density of population at location x . Then, the total abundance of the population (Q) is calculated from:

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

and the probability density function of the location \underline{x} of a random individual is $z(x)/Q$. The centre of gravity (CG) is:

$$CG = E(\underline{x}) = \int \underline{x} \frac{z(\underline{x})}{Q} d\underline{x} = \frac{\int \underline{x} z(\underline{x}) d\underline{x}}{\int z(\underline{x}) d\underline{x}},$$

In practice, this statistic is estimated from the data through discrete summations over sample locations. In the case of irregular sampling, areas of influence around samples are used as weighting factors (**Figure 4.2.1**). Practically, from sample values z_i at locations x_i , with areas of influence s_i , we have:

$$CG = \frac{\sum_{i=1}^N x_i s_i z_i}{\sum_{i=1}^N s_i z_i},$$

The area of influence of a sample location is defined as the area made up of the points in space that are closer to this sample than to others. It can be evaluated by overlaying a very fine regular grid and counting the grid points closer to the sample. Known or supposed boundaries (e.g. land, a limit distance of influence from a sample location) of the sampled population may be used.

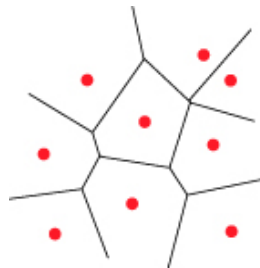


Figure 4.2.1. Areas of influence (delimited by the black lines) of sample points (in red).

Reference points

CG during a period of acceptable state of the population.

Interpretability

The CG indicates the mean location of the surveyed population. Note that fish may not be present at the CG location (which may be on land, e.g. within an island). Note also that part of the population may be not represented, when not covered geographically by the sampling. To check that movements of CG when following a series are not due to changes in the sampling design (e.g. due to bad weather), the CG of the sample locations (that is, unweighted by fish density) can also be produced.

Measurability

The estimated CG is sensitive to high fish density values. It may differ from the true unknown CG, particularly when high density values exist (whether sampled or not).

Sensitivity

Despite the possible difference between true and estimated CG, a shift during a series is likely to represent an actual shift, when it is gradual. On the other hand, an eccentric estimated CG requires a visual inspection of the fish density to detect the causes (e.g. unusual presence of high density values in some remote area, or disappearance of usually high values in some region; see the indicators Inertia and Number Of Patches).

Examples

Several authors have used the center of gravity, also referred to as the distributional centroid or as the center of an ellipse to describe the distribution of a population or a life stage of a population such as Walleye pollock eggs and larvae (Kendall and Picquelle, 1990), Pacific hake larvae (Hollowed, 1992), cod off Newfoundland (Atkinson et al., 1997), yellowtail flounder off the Grand Bank (Brodie et al., 1998) or European hake in the Bay of Biscay, eggs and larvae (Alvarez et al., 2001) as well as fish at age (Woillez et al., 2007). For example, the CGs have been used to describe the distribution of the strong year classes on the Pacific hake late stage larvae and also the systematical shift towards the south east of cod off Newfoundland from 1987 to 1993. Distributions have been also described temporally along the season, e.g. for eggs and larvae of the European hake in the Bay of Biscay and of Walleye Pollock.

4.3 Inertia: *I*

Description

The inertia is the variance of the location of the individuals of the population, that is, the mean square distance between an individual fish and the centre of gravity of the population (Bez, 1997).

Stock attribute

It describes the dispersion of the population around its centre of gravity.

Derivation

With the notations used for CG, the inertia (*I*) is

$$I = \text{Var}(\underline{x}) = \frac{\int (x - CG)^2 z(x) dx}{\int z(x) dx}.$$

and is estimated as:

$$I = \frac{\sum_{i=1}^N (x_i - CG)^2 s_i z_i}{\sum_{i=1}^N s_i z_i}.$$

Reference points

I during a period when the population was in an acceptable state.

Interpretability

The inertia *I* indicates how dispersed the population is around its center of gravity.

Measurability

I is sensitive to high density values. *I* is homogeneous to square nautical miles (in 2D). The square root of *I*, that is, the root mean square distance between individuals and their CG, may be preferred, being homogeneous to nautical miles.

Sensitivity

An increase in *I*, for instance, indicates a population more dispersed around its CG, i.e. high density values are more scattered. While the population is then scattered over a larger region, the actual area covered by the population may be smaller (see the different Area indicators).

Examples

Most of the authors cited for the CGs also described the studied population in terms of inertia, or they refer to the size of its graphical representation through an ellipse (see next section on anisotropy). Brodie et al. (1998) showed a drop in the area of the ellipse for the yellowtail flounder of the Grand bank in late 1980 after a period of stability. Atkinson et al. (1997) showed that a shift of the CG is accompanied by a decrease of the size of the ellipse, i.e. the inertia. For the European hake in the Bay of Biscay, concerning the eggs, the size of

the ellipses increased somewhat from February to May in both directions, N-S and W-E (Alvarez et al., 2001), while the inertia increased with the age for fish stage (Woillez et al., 2007).

4.4 Anisotropy and isotropy: An, Is

Description

When the dispersion of the population around its center of gravity is the same along every direction, the spatial distribution is said to be isotropic. In general, the dispersion of a population around its centre of gravity is not identical in every direction of space: there is an anisotropy. The root mean square distance to the center of gravity is maximal along the first principal axis, and minimal along the second principal axis, orthogonal to the first one (in 2D). The anisotropy index is taken as the ratio (≥ 1) between these distances, and the isotropy index as the inverse ratio (≤ 1).

Stock attribute

Anisotropy measures the elongation of the spatial distribution of the population.

Derivation

In two dimensions, the total inertia of a population can be decomposed on its two principal axes, orthogonal to each other, explaining respectively the maximum and the minimum of the inertia. These two principal axes and their inertia can be obtained as the eigen vectors and values of a principal component analysis of the coordinates of the individuals of the population (i.e. the coordinates of the samples weighted by the fish densities) (Bez, 1997). The square root of the inertia along a given axis (or root mean square distance to CG) gives the standard deviation of the projection of the location of the population along that axis. These can be represented conveniently on a map with a cross depicting the two principal directions (**Figure 4.4.1**), or with an ellipse (with area proportional to the total inertia). The anisotropy index (≥ 1) is the square root ratio between the maximum and the minimum of the inertia. Similarly, an index of isotropy can be defined as the inverse of anisotropy, ranging more conveniently from 0 to 1:

$$\text{Isotropy} = \sqrt{\frac{I_{\min}}{I_{\max}}} \text{ and } \text{Anisotropy} = \sqrt{\frac{I_{\max}}{I_{\min}}}.$$

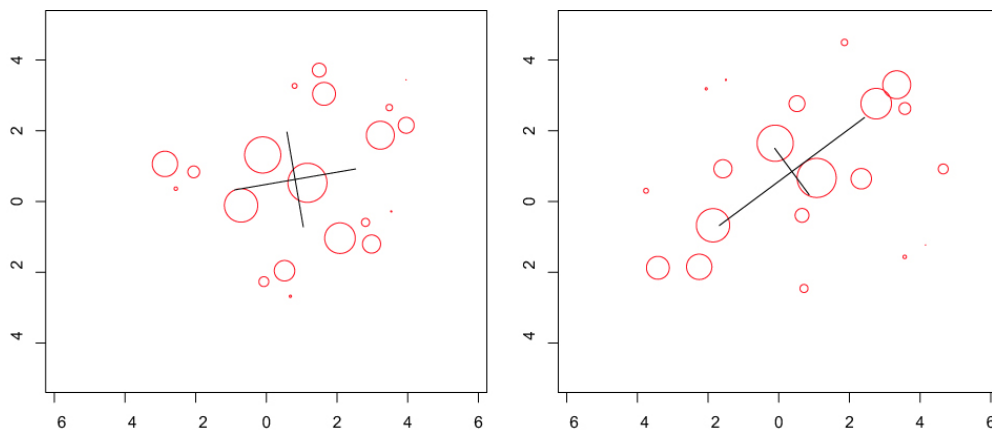


Figure 4.4.1. Two examples of spatially distributed data sets, with anisotropy being more marked in the second case. The black cross is located on the center of gravity, from which it represents the square root of inertia along the two principal directions.

Reference points

The indicators during a period of acceptable state of the population.

Interpretability

The anisotropy index gives roughly the elongation of the spatial population. It does not take into account the actual shape of the distribution, which may be different from being elliptical or may be constituted by different patches.

Measurability

The anisotropy and isotropy indices are equivalent, being the inverse of each other. However, since the anisotropy index is unbounded above 1, the isotropy index is more robust and may be more conveniently used, e.g. in correlation or regression analyses.

In case of isotropy, that is, when the anisotropy and isotropy indices approach 1, the principal axes, orthogonal to each other, become arbitrary.

Sensitivity

Sudden changes in anisotropy index may be due to the disappearance or, on the contrary, the appearance, of patches of fish in some areas.

Examples

Few authors have discussed the anisotropy even if it was available in the representation mode. In Alvarez et al., 2001, the direction of the principal axis of the hake egg distribution in the Bay of Biscay corresponds to that of the shelf break, i.e. NW-SE, along the whole sampling period. Woillez et al., 2007, completes the description for the fish stage, showing a preferential direction, more marked for age 0 and age 5+. The direction for age group 0 corresponds roughly to muddy sediment off Brittany. For ages 4 and 5+, the direction corresponds to the shelf edge, where older hake are mainly concentrated. For the intermediate ages, the population is still anisotropic, probably because of the general shape of the continental shelf, but the anisotropy is less marked.

4.5 Global index of collocation: GIC

Description

The global index of collocation looks at how geographically distinct two populations are by comparing the distance between their CGs and the mean distance between individual fish taken at random and independently from each population (Bez and Rivoirard, 2000).

Stock attribute

Overlap of two spatial populations.

Derivation

Let us consider two populations with densities $z_1(x)$ and $z_2(x)$ at point x , with ΔCG being the distance between their centers of gravity and I_{z_1} and I_{z_2} , their respective inertias (

Figure 4.5.1). Then the mean square distance between individuals taken at random and independently from each population is $\Delta CG^2 + I_{z_1} + I_{z_2}$. The global index of collocation (GIC) is:

$$GIC = 1 - \frac{\Delta CG^2}{\Delta CG^2 + I_{z_1} + I_{z_2}}$$

or 1 if $\Delta CG^2 = I_{z_1} = I_{z_2} = 0$. The spatial index ranges between 0, in the extreme case where each population is concentrated on a single but different location (inertia = 0), and 1, where the two CGs coincide.

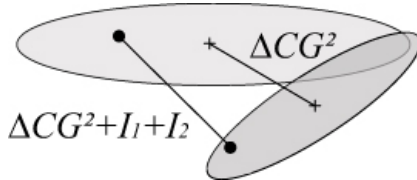


Figure 4.5.1. Collocation of two spatial populations, represented by two ellipses showing their center of gravity and their inertia, is measured with the GIC through specific distances.

Reference points

GIC between acceptable states of the populations.

Interpretability

Collocation is considered here in a global meaning, the populations being, grossly, in the same place. This is not to say that the two populations are present at the same locations. A spatial population that would be distributed all around a first one, with the same CG, would give a GIC equal to 1, even if not overlapping the first population.

Local overlapping between two populations would rather be addressed using the local index of collocation, that is, the noncentered correlation between their fish densities.

Measurability

Alternative indices, also between 0 and 1, may be $\sqrt{\text{GIC}}$ for collocation, or $\sqrt{1-\text{GIC}}$ for separation (ratio between distance between the CGs and distance between individuals from the two populations).

Sensitivity

Unusually high GIC requires inspection of the fish density data.

Examples

In Bez and Rivoirard, 2000, global and local collocation indices have been measured on pelagic species in the Bay of Biscay. Local collocation appears very small between mackerel and the other species (anchovy, sardine and horse mackerel). In Woillez et al., 2007, GIC is used to detect outliers in the age time series of the hake population in the Bay of Biscay. The year 2000 appears to be particular for age 0.

References

See the main references of the topics about the spatial indicators.

4.6 Number of spatial patches: NOP

Description

A spatial population of fish may be distributed into several spatial patches, with size much larger than a fish school. An algorithm has been written to identify patches (Woillez et al., 2007) by attributing each sample to the nearest patch, with respect to a maximal threshold distance to its CG. The Number of Patches then includes all patches that include more than a given part (e.g. 10%) of the total abundance.

Stock attribute

Patchiness.

Derivation

The algorithm starts from the value displaying the maximum density $z(x)$, and considers every other sample in decreasing order of density (**Figure 2.6.1**). The maximum value initiates the first patch (1). Then, the current sample value is attributed to the nearest patch, if the distance to its CG is smaller than the threshold distance $dlim$ (2). Otherwise, the current sample value defines a new patch (3). Spatial patches whose abundance is $>10\%$ of overall abundance are retained (N). The summary index is then the number of patches.

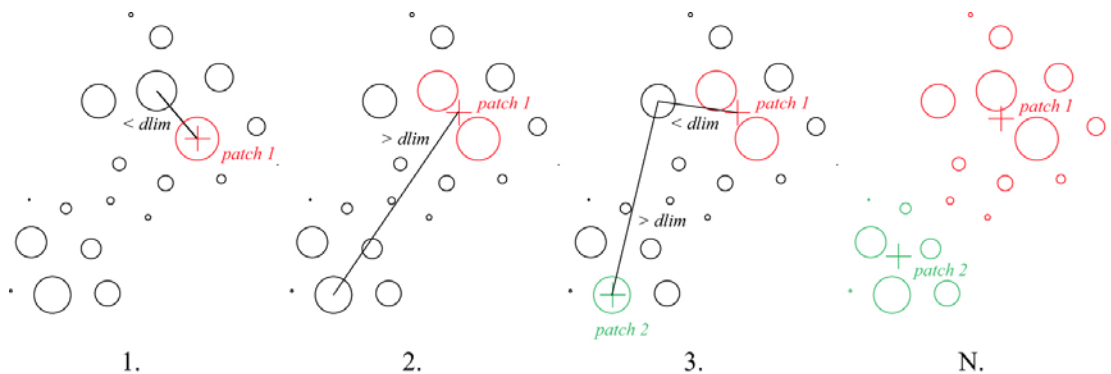


Figure 2.6.1. Main steps of the algorithm used to determine the number of patches of a spatial population.

Reference points

Number of patches during a period of acceptable state of the population.

Interpretability

The identification of patches is dependent on the threshold distance, typically some fraction of the diameter of the sampled domain, chosen by the user.

Measurability

The Number of patches is very sensitive to the location of the highest fish density values, but this makes sense.

Sensitivity

In a series, the location of patches is likely to present some stability. Hence a change in the number of patches is likely to reveal the disappearance, or the appearance of fish in some areas.

Examples

In Woillez et al., 2007, the Number of Patches have been illustrated on the hake population in the Bay of Biscay. According to ages, it increases slightly up to age 3 then decreases. Disappearance of patches has been observed and localised for age 0 hake, in particular for the year 2000.

References

See the main references of the topics about the spatial indicators.

4.7 Positive area: PA

Description

The positive area is the measure, in square nautical miles, of the space occupied by fish densities strictly above zero (Woillez et al., 2007).

Stock attribute

Area of presence, in square nautical miles, occupied by the stock, even with a low density.

Derivation

The positive area is estimated from data as the sum of the areas of influence around samples where there are fish densities >0 (Fig. 4.7.1):

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

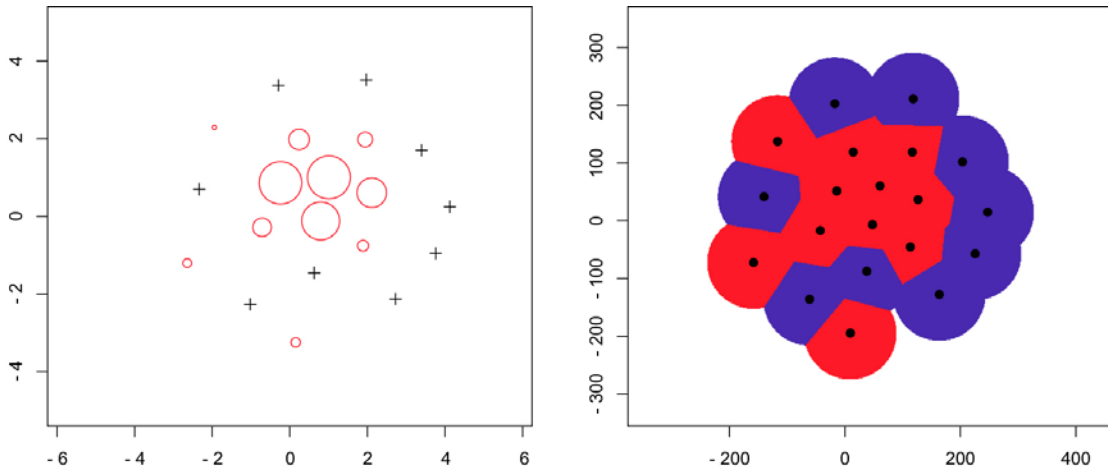


Figure 4.7.1. Bubble plot of the sample values and corresponding positive area shaded in red (with a limit to the area of influence of each sample).

Reference points

Positive Area during a period of acceptable state of the population.

Interpretability

The positive area measures the area of effective presence, in square nautical miles. It does not include zero density areas possibly existing between positive density areas, and it may correspond to a small fraction of the geographical envelope of fish presence, in particular when the dispersion (inertia) is high.

Measurability

Zero values of density make no contribution to the positive area. However, the positive area is sensitive to the low values of density, because the contribution of a small density value is the same as that of a high density value.

Sensitivity

Changes in the Positive Area may reveal changes in the way the population occupies the space, including the small fish density values that are usually numerous even though they contribute poorly to the global abundance.

Examples

In Woillez et al., 2007, Positive Area of the hake population in the Bay of Biscay was relatively stable until age 3 then dropped. It was shown also that whereas Positive Area decreased with age, inertia increased with age, the older hake occupying a smaller but more dispersed area.

References

See the main references of the topics about the spatial indicators.

4.8 Spreading area: SA

Description

The spreading area is a measure, in square nautical miles, of how the population is distributed in space, taking into account the variations in fish density (Woillez et al., 2007).

Stock attribute

A measure of the area occupied by the stock, based on how the abundance is spreading in space.

Derivation

Let T be the cumulative area occupied by the density values, ranked in decreasing order; let $Q(T)$ be the corresponding cumulative abundance, and Q be the overall abundance. The SA (expressed in square nautical miles) is then simply defined as twice the area below the curve expressing $(Q-Q(T))/Q$ as a function of T (Fig. 4.8.1):

$$SA = 2 \int \frac{Q - Q(T)}{Q} dT$$

As $(Q - Q(T))/Q$ decreases from 1 to 0 and is convex, the SA is less than the PA. It equals the PA when the population is evenly spread with a constant density.

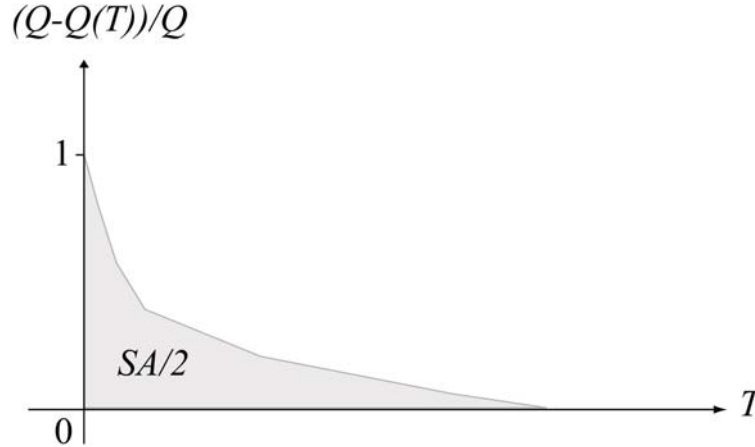


Figure 4.8.1. SA is defined as twice the area below the curve expressing $(Q-Q(T))/Q$ as a function of T .

The curve above is a derivation of the Lorenz curve representing the histogram of fish density values, but having the advantage of receiving no contribution from zero density values. The spreading area can be related to the area occupied by the positive fish density values and their Gini index of dispersion G_0 through $\frac{SA}{PA} + G_0 = 1$ (Woillez et al. 2007).

Reference points

Spreading Area during a period of acceptable state of the population.

Interpretability

A spatial abundance is generally distributed in space into highly varying fish density values, spreading over its positive area. The spreading area index has been designed to describe this spreading, or equivalently the lack of aggregation or variation, while satisfying the condition of having no contribution from zero density values. Despite its name, the spreading area depends exclusively on the amount and histogram of positive fish density values.

Measurability

Zero values of density make no contribution to the spreading area. The spreading area depends on the variation in density values (and not on the absolute abundance) and is much less sensitive to low values of density than the positive area.

Sensitivity

Changes in SA are likely to reveal changes in the way the abundance is split into low and high density values.

Examples

In Woillez et al., 2007, the hake population of the Bay of Biscay has been described through the SA. This showed a better spread of the 3 year-old hake. In addition, a decrease of SA through the time series was detected for hake age 4 and 5+.

4.9 Equivalent area: EA

Description

The Equivalent Area represents the area, in square nautical miles, that would be covered by the population if all individuals had the same density, equal to the mean density per individual (Bez and Rivoirard, 2001).

Stock attribute

An individual-based measure of the area occupied by the stock, in square nautical miles.

Derivation

The transitive geostatistical approach (Matheron, 1971) can be used to describe the spatial distribution of a fish population when it includes a few large values of density, and when it is difficult to delimit a domain with homogeneous variations. The spatial structure is then represented by a (transitive) covariogram, a function of the distance between two locations:

$$g(h) = \int z(x)z(x+h)dh.$$

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

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

It can also be written:

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

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 (the denominator in the previous equation (Fig. 4.9.1)).

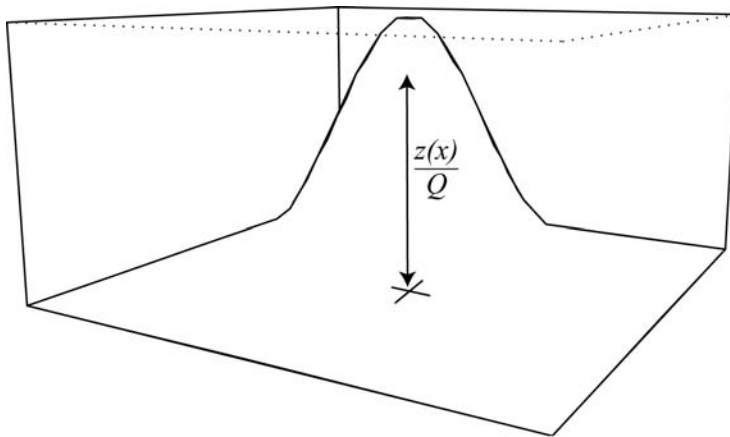


Figure 4.9.1. The probability density function for a random individual to be at x is given by $z(x)/Q$.

Practically, in the discrete case with sample values z_i and areas of influence s_i , it gives:

$$EA = \left(\sum_{i=1}^N s_i z_i \right)^2 / \sum_{i=1}^N s_i z_i^2 .$$

The EA ranges from 0 to the PA. It would be equal to the PA if all strictly positive values of density were the same. The EA can be related to the area occupied by the positive fish density values and their coefficient of variation CV_0 through $\frac{PA}{EA} = 1 + CV_0^2$. The EA and SA are related through inequalities, in particular $EA \leq \frac{9}{8} SA$ (Woillez et al. 2007).

Reference points

Equivalent Area during a period of acceptable state of the population.

Interpretability

The positive area describes the area of presence of fish, with a low density value being equivalent to a high one. The spreading area describes the area occupied by the stock, taking into account the variations in fish density. Now, the equivalent area is still another way to do this, while being individual-based (it gives the same weight to each individual, that is, the weight of a sample is proportional to its fish density). Just like the spreading area, the equivalent area depends exclusively on the amount and histogram of positive fish density values.

Measurability

The equivalent area is independent of the absolute abundance. Being individual-based, it is very sensitive to the highest density values. The inverse of the equivalent area can be considered as an index of aggregation (Bez and Rivoirard, 2001).

Sensitivity

Changes in EA are likely to reveal changes in the contribution of high density values to the total abundance.

Examples

In Woillez et al., 2007, the Equivalent Area on the hake population of the Bay of Biscay was shown to be larger for hake aged 3 years.

4.10 Microstructure index: *MI*

Description

The Microstructure Index (Woillez et al., 2007) measures the relative importance of structural components having a scale smaller than the sample lag (including random noise).

Stock attribute

The fine-scale variability of the fish density surface.

Derivation

The microstructure index (MI) is taken as the relative decrease of the transitive covariogram (Matheron, 1971; Bez et al., 1997) between distance zero and a distance h_0 chosen to represent the mean lag between samples (Fig. 4.10.1):

$$MI = \frac{(g(0) - g(h_0))}{g(0)}$$

It lies between 0 and 1. Values close to 0 correspond to a very regular, well-structured density surface, and values close to 1 correspond to a highly irregular, poorly structured, density surface.

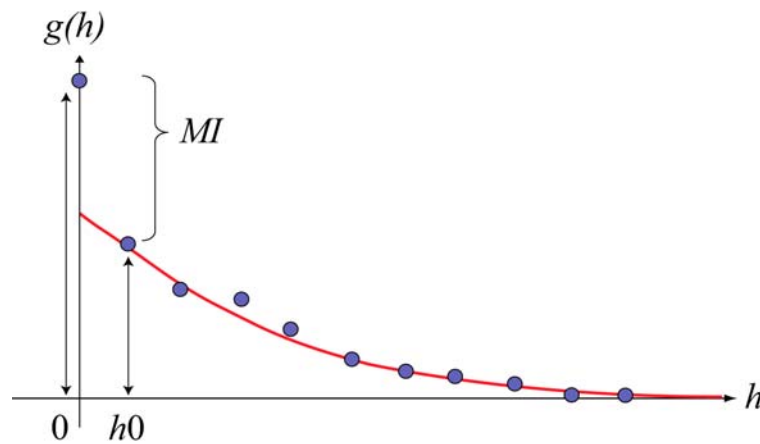


Figure 4.10.1. Real (dashed line), experimental (blue points) and modelled covariogram (red line) with the representation of the microstructure index.

Reference points

Microstructure Index during a period of acceptable state of the population.

Interpretability

The Microstructure Index does not make the distinction between spatial variability with a range less than the chosen lag but positive, and purely random variability (e.g. due to noise or sampling error).

Measurability

The Microstructure Index, as obtained from the transitive covariogram, is very sensitive to high fish density values (but it is more robust than its equivalent feature that would be obtained from the more traditional variogram or covariance).

Sensitivity

A high Microstructure Index is likely to correspond to fine-scale aggregations.

Examples

In Woillez et al., 2007, Microstructure Index has been followed through age and time on the hake population in the Bay of Biscay. It showed a relative stability for the younger ages, then it rose markedly from age 4.

4.11 References for spatial indicators

Alvarez P., Motos L., Uriarte A., and Egana J. 2001. Spatial and temporal distribution of European hake, *Merluccius merluccius* (L.), eggs and larvae in relation to hydrographical conditions in the Bay of Biscay. *Fisheries Research*, 50: 111–128.

Atkinson, D. B., Rose, G. A., Murphy, E. F., and Bishop, C. A. 1997. Distribution and abundance of northern cod, 1981–1993. *Canadian Journal of Fish and Aquatic Sciences*, 54 (Suppl. 1): 132–138.

Bez, N. 1997. Statistiques individuelles et géostatistique transitive en écologie halieutique. Thèse de Docteur en Géostatistique, Ecole Nationale Supérieure des Mines de Paris, France. 276 pp.

Bez, N., and Rivoirard, J. 2000. Indices of collocation between populations. In Report of a Workshop on the Use of Continuous Underway Fish Egg Sampler (CUFES) for Mapping Spawning Habitat of Pelagic Fish, pp. 48–52. Ed. by D. M. Checkley, J. R. Hunter, L. Motos, and C. D. von der Lingen. GLOBEC Report, 14.

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

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.

Brodie, W. B., Walsh, S. J., and Atkinson, D. B. 1998. The effect of stock abundance on range contraction of yellowtail flounder (*Pleuronectes ferruginea*) on the Grand Bank of Newfoundland in the Northwest Atlantic from 1975 to 1995. *Journal of Sea Research*, 39: 139–152.

Hollowed, A. B. 1992. Spatial and temporal distribution of Pacific hake, *Merluccius productus*, larvae and estimates of survival during early life stages. *CalCOFI Report*, 33: 100–123.

Kendall, A. W., and Picquelle, S. J. 1989. Egg and larva distributions of walleye pollock *Theragra chalcogramma* in Shelikof Strait, Gulf of Alaska. Fishery Bulletin US, 88: 133–154.

Matheron, G. 1971. The theory of regionalized variables and its applications. Les cahiers du Centre de Morphologie Mathématique, Fascicule 5. Ecole Nationale Supérieure des Mines de Paris, Paris. 212 pp.

Myers, R. and Cadigan, N. (1995) Was an increase in natural mortality responsible for the collapse of northern cod? Can. J. Fish. Aquat. Sci. 52: 1274-1285.

Petitgas P (1998) Biomass dependent dynamics of fish spatial distributions characterized by geostatistical aggregation curves. ICES Journal of Marine Science 55: 443-453

Swain D, Sinclair, A (1994) Fish distribution and catchability: what is the appropriate measure of distribution? Can. J. Fish. Aquat. Sci. 51:1046-1054

Woillez, M., Poulard, J-C., Rivoirard, J., Petitgas, P., and 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 Journal of Marine Science, 64: 537–550.

5. Methods

5.1 Introduction

The following methods for integrating and interpreting time-series of indicators as estimated from research vessel surveys were developed and discussed during the FISBOAT project. In general, they attempt to ease the problem of deciding whether an indicator or a suite of indicators is signalling, or will signal, a change in the state of a fish stock that calls for an adjustment to the controls on a fishery.

Section 5.2 by Lembo et al. presents an age-length based simulation model, Aladym, for predicting the effects of different fishing pressures on a single population of fish. Section 5.3, also by Lembo et al. describes how a Monte Carlo approach can be used with Aladym in order to help create reference points for selected indicators.

Section 5.4 by Trenkel et al. points out the importance of relating current trends in indicator values with a previous reference period when indicator values were agreed to be either satisfactory or not. They offer a simple system for integrating different types of biological and fishery information provided by indicator time series, and by other sources if available, in a way that can be discussed by stock managers, stake-holders, and scientists in order to decide what, if any, measures should be taken to control the fishery. This would often form part of an adaptive management scheme.

Mesnil and Petitgas, section 5.5, describe how the quality control schemes that originated in manufacturing industry can be adapted to monitor fishery and environmental qualities derived from time series of indicators. The CUSUM control-chart method offers considerable potential for rapid detection of changes of state.

Bogaards et al., section 5.6, describe a simple, hypothesis-testing approach for deciding how long will be needed before an indicator series is expected to reveal a given linear trend. This could be helpful for deciding whether a survey is sufficiently sensitive to detect a response to new controls on a fishery within a reasonable time frame.

In section 5.7, Trenkel proposes a GAM and bootstrap-based solution for the long standing problem of deciding whether recent changes in an important time-series represent a valid signal, or just sampling noise. Such methods can greatly assist the provision of rapid, confident advice for managing a fishery.

In sections 5.8 and 5.9, Petitgas describes and illustrates two complementary approaches to making a single assessment from multiple time-series of indicators. In the first part, the well-known 'traffic light' method is set out. In the second, multivariate methods based on principal components and multi-factorial analysis (MFA) are described and illustrated with an example. The multivariate analysis is likely to provide complementary interpretation of results in the traffic light table.

Petitgas and Poulard in section 5.10 describe a multivariate statistical method for visualising groupings of indicator variables in space and time. They applied this method during the FISBOAT project to examine the changing spatial distributions of fish with age, as signalled by the spatial indicators described in the first part of this manual.

In section 5.12, Woillez and Rivoirard describe a multivariate statistical method applicable to parallel time series of values for many indicators when continuity of some trend in time is of interest, possibly in relation to a study of cause and effect. MAFs are linear combinations of variables that are conceptually similar to principal components (PCs) but, whereas a fitted series of PCs explains independent components of variance the magnitude of which decreases from first to last, a series of MAFs explains independent components of autocorrelation, the first of which displays the highest, and the last of which displays the lowest autocorrelation at lag 1 observation in the time series. Thus MAFs offer a way of finding the combination of variables that present maximal continuity in time.

Finally, in section 5.13, Cotter reviews and illustrates with example analyses a collection of nonparametric statistical methods that allow assessment of indicator trends whilst avoiding the assumptions and other uncertainties of modelling. Suggestions are made for improving the objectivity of statistical inference.

Several of the methods use especially written software. The FISBOAT project team agreed at the outset to use the R programming language because it is freely available from <http://www.r-project.org/>, it is highly versatile and because, by doing so, the portability of software and ideas is maximised. The software can be downloaded freely from <http://www.ifremer.fr/drvecohal/fisboat/index.htm>. Data sets and spreadsheets referred to in the following sections should also be available from this site.

5.2 Simulating population dynamics.

5.2.1 ALADYM(v 08)

Lembo G., A. Abella, F. Fiorentino, S. Martino and M.T. Spedicato
COISPA Tecnologia & Ricerca, Bari, Italy

Introduction

ALADYM (Age-Length Based Dynamic Model) is an age-length based simulation model developed within the conceptual framework of dynamic pool models, following the predictive Thompson & Bell (1934) approach. The model is designed to predict, through simulations, the effects of different fishing pressure scenarios on a single population, in terms of different metrics and indicators. Removals are simulated on the basis of the total mortality rate modulated using harvesting pattern and a fishing activity coefficient. *Aladym* can work in absence of fishery-dependent data, although its predictive capability of real catch levels can be verified using information on commercial catches or fishing activity per month.

From the *Aladym* core model three complementary, but independent, tools have been derived:

- the quasi-deterministic dynamic tool named *Aladym-r*;
- the tuning tool *Aladym-z*;
- the stochastic dynamic tool named *Aladym-q*.

The core *Aladym* model is described in this chapter together with *Aladym-r* and *Aladym-z*, while *Aladym-q* is described in the following one.

General assumptions

The basic assumptions of the model are:

- natural mortality as estimated reflects the rate of decline of a population from all causes excluding fishing;
- total mortality Z reliably reflects the decline of ages/sizes in the population, including the effects of different fishing gears;
- the growth, the natural mortality, and the maturity parameters are assumed constant over time;
- given the small time interval (1 month) between cohorts the effect of the spreading of the lengths with respect to the ages can be neglected.

Derivations

The quasi-deterministic dynamic tool named Aladym-r

General framework

The model is designed to simulate population dynamics of a given species accounting for differences by sex in growth, maturity and mortality. All the quantities are calculated as vectors with a time step Δt (time slice=1 month). An operational framework of the *Aladym-r* model is shown in fig. 5.2.1. Step A) regards the *input and initialization*. In order to generate an unbiased initial population, the number of runs specified by the user (e.g. 100) is performed in this step, randomly varying the recruitment, the growth and the size-at-maturity parameters according to the values and distributions specified by the user. The user can choose among the following distribution type: log-normal, normal, gamma and uniform. For

the parameter t_0 a uniform distribution is associated by default. Two populations are generated: the exploited (where total mortality is acting) and the unexploited one (where natural mortality only is acting).

The obtained initial populations enter in the *start loop* (or *seed run*) (step B in fig. 5.2.1), where the dynamics are formulated to follow the evolution of several cohorts over a monthly scale. Here the number of recruits entering in the population is generated from a stock-recruitment relationship. Alternatively, it is given as an input vector. In both cases, a uniform variability for the obtained number of recruits can be set by the user. The *start loop* runs for a number of years that is a multiple of the two sex life-spans. This step aims to eliminate the artefacts in the initial population due to the use of an equilibrium model in the initialization step. After this phase, the *simulation loop* starts and runs over the period required by the user (step C in fig. 5.2.1) generating the outputs (step D in fig. 5.2.1).

Model components

Growth

The growth process is modelled using a von Bertalanffy growth function:

$$L_{age} = L_{\infty} \cdot \left(1 - e^{-K \cdot (age - t_0)}\right).$$

For each age (time step $\Delta t = 1$ month) length is calculated using the input parameters L_{∞} , K and t_0 . The average length in the time interval $(t, t + \Delta t)$ is calculated as:

$$\bar{L}_{age} = L_{\infty} + \frac{(L_{age} - L_{age + \Delta t})}{K \cdot \Delta t}$$

The weight at average length, for each age, is calculated from the length-weight relationship in the form: $W_{age} = a \bar{L}_{age}^b$; with a and b as input parameters.

Population

The population dynamics is formulated following the simultaneous evolution of several cohorts at monthly scale through the exponential population decline model, both in absence (1) and in presence (2) of fishing mortality:

$$\frac{dN}{dt} = -MN \quad (1)$$

$$\frac{dN}{dt} = -ZN \quad (2)$$

used respectively in the form (3) and (4):

$$N_{(t+\Delta t),j} = N_{t,j} e^{-M_{t,j} \cdot \Delta t} \quad (3)$$

$$N_{(t+\Delta t),j} = N_{t,j} e^{-(F_{t,j} + M_{t,j}) \cdot \Delta t} \quad (4)$$

where j indicates the cohort, t the time, Z , M and F the total, natural and fishing mortality respectively. (Notice that in any formula where j , age and t are present, it is assumed that age represents the age of the cohort j at time t).

Maturity

Maturity Mat is a function of the length L and is calculated following an ogive model (Quinn and Deriso, 1999):

$$Mat(L) = \frac{1}{1 + e^{-r(L-L_{m50\%})}}$$

where r is the ogive slope and $L_{m50\%}$ is the length at which 50% of fish mature. The proportion of mature fish at age is computed as:

$$Mat_{age} = \frac{1}{1 + e^{\left(\frac{2 \cdot \ln 3}{L_{m75\%} - L_{m25\%}}\right)(L_{m50\%} - \bar{L}_{age})}}$$

where the maturity range $L_{m75\%} - L_{m25\%}$, is related to the ogive slope.

Biomass

The biomass (B_j) and the spawning stock biomass (SSB_j) of the cohort j at time t are respectively computed as:

$$B_{t,j} = N_{t,j} \cdot w_{age};$$

$$SSB_{t,j} = N_{t,j} \cdot w_{age} \cdot Mat_{age}$$

Analogously, the unexploited biomass (UB_j) and the unexploited spawning stock biomass ($USSB_j$) of the cohort j at time t are calculated as:

$$UB_{t,j} = UN_{t,j} \cdot w_{age};$$

$$USSB_{t,j} = UN_{t,j} \cdot w_{age} \cdot Mat_{age}$$

Initial recruitment and stock recruitment relationship

During the step A) (fig. 5.2.1) the initial number of individuals in the population are from estimates of recruitment independently obtained from e.g. trawl surveys or other sources. These numbers randomly selected for each of the e.g. 100 runs (see also the *general framework* paragraph) are used to initialize the population. Successively (step B and C in fig. 5.2.1), the number of individuals entering in the population can be a vector or is estimated from one of the following user selected stock-recruitment relationships:

Beverton & Holt (1957):

$$R = \frac{S}{(a + bS)};$$

Ricker (1954):

$$R = a \cdot S \cdot e^{(-bS)};$$

Shepherd (1982):

$$R = a \cdot S / [1 + (S/c)^b];$$

Barrowman & Myers (2000):

$$R = \alpha \cdot \min(S, S^*) = \begin{cases} \alpha \cdot S & \text{if } S < S^* \\ \alpha \cdot S^* & \text{if } S \geq S^* \end{cases}$$

$$R = \begin{cases} \alpha \cdot S & \text{if } S \leq S^* \cdot (1 - \delta) \\ \alpha \cdot \left(S - \frac{(S - S^* \cdot (1 - \delta))^2}{4\delta \cdot S^*} \right) & \text{if } S^* \cdot (1 - \delta) < S < S^* \cdot (1 + \delta) \\ \alpha \cdot S^* & \text{if } S \geq S^* \cdot (1 + \delta) \end{cases}$$

R and S represent the number of recruits and spawners respectively, whilst a , b , c , α , δ , S^* are the model's parameters. Uniformly distributed random variations can be applied by the user to the number of offspring (from the vector or stock-recruitment relationship).

The number of the events (on monthly basis) generating the offspring is an input of the model. The population of spawners generating the recruits is calculated by summing up the number of individuals of the different age classes of the different cohorts occurring in the population one or more (depending on the biological features of the species) months before the offsprings are produced. Thus this quantity is calculated as follows:

$$SSN_t = \sum_j SSN_{t,j};$$

where $SSN_{t,j}$ represents the number of mature females at time t , of the cohort j :

$$SSN_{t,j} = N_{t,j} \cdot Mat_{age}.$$

Mortality

The natural mortality can be constant for each age/length, or a vector by age/length calculated outside the model and used as input. Alternatively, it is estimated inside the model from the Chen and Watanabe equations (1989):

$$M(t) = \begin{cases} \frac{K}{1 - e^{-K \cdot (t - t_0)}} & t \leq t_M \\ \frac{K}{a_0 + a_1 \cdot (t - t_M) + a_2 \cdot (t - t_M)^2} & t \geq t_M \end{cases}$$

where:

$$t_M = -\frac{1}{K} \ln|1 - e^{K t_0}| + t_0$$

$$a_0 = 1 - e^{-K \cdot (t_M - t_0)}$$

$$a_1 = K \cdot e^{-K \cdot (t_M - t_0)}$$

$$a_2 = -\frac{1}{2} K^2 \cdot e^{-K \cdot (t_M - t_0)}$$

The two parameters of the Chen and Watanabe model are t_0 and K . The asymptotic length (L_∞) is not necessary, but t_0 cannot be equal to 0 (otherwise the parameter t_M cannot be defined). The quantities a_0 , a_1 , a_2 and t_M cannot be strictly considered as parameters of the model, as they depend on t_0 and K . The parameter t_M represents the age beyond which the

contribution of the fish of a given cohort can be considered negligible. If parameters are consistent the relationship between age and natural mortality shows a “bath tube” shape.

The fishing mortality rate $F(L)$ is modelled for each cohort using the following general equation (Sparre and Venema, 1998):

$$F(L) = F_{max} \cdot S(\bar{L})$$

where F_{max} is the maximum fishing mortality and $S(\bar{L})$ the proportion of retained fish. In *Aladym* the fishing mortality rate is calculated as

$$F(L) = F_{max} \cdot S(\bar{L}) \cdot f_{act}$$

where maximum fishing mortality (F_{max}) is calculated as

$$F_{max} = QZ_{input} - M_{min}$$

using the input values of QZ (a Z proxy) and where M_{min} represents the minimum value that the M vector assumes. As an alternative option, F_{max} can also be a user selected input to be set for each month. In addition, a fishing activity coefficient (f_{act}) is introduced in order to consider the possibility of a fishing ban or changes in fishing effort throughout time.

The value of QZ by sex can be assumed, as a first order approximation, numerically equal to the value of Z observed that is obtained from estimations outside the simulation model (e.g. from trawl-survey). A better approximation of QZ is obtained using the tool *Aladym-z* (see a later paragraph).

In the model, the probability of selection $S(\bar{L})$ of the cohort j is calculated at time t from one of the two following user-selected relationships:

$$S(\bar{L}) = \frac{1}{1 + e^{\left(\frac{2 \cdot \ln 3}{L_{75\%} - L_{25\%}}\right) \cdot (L_{50\%} - \bar{L}_{age})}};$$

or

$$S(\bar{L}) = \frac{1}{1 + e^{\left(\frac{2 \cdot \ln 3}{L_{75\%} - L_{25\%}}\right) \cdot (L_{50\%} - \bar{L}_{age})}} \cdot \frac{1}{1 + e^{\left(\frac{-2 \cdot \ln 3}{D_{25\%} - D_{75\%}}\right) \cdot (D_{50\%} - \bar{L}_{age})}};$$

where $L_{50\%}$, $L_{75\%}$ and $L_{25\%}$ are the selectivity parameters and $D_{50\%}$, $D_{25\%}$, $D_{75\%}$ the de-selection parameters of the model. The total mortality Z at time t for the cohort j is thus computed as

$$Z_{t,j} = F_{t,j} + M_{t,j}$$

that is the value acting on the population in the model computations.

The biomass of individuals of the cohort j at time t death for all causes ($BP_{t,j}$) is computed as

$$D_{t,j} = N_{t,j} - N_{t+\Delta t,j} = N_{t,j} \cdot (1 - e^{-Z_{t,j} \cdot \Delta t}),$$

$$BP_{t,j} = D_{t,j} \cdot w_{age};$$

while the biomass of those dead from all causes excluding fishing ($BND_{t,j}$) is computed as

$$BND_{t,j} = \frac{M_{t,j}}{Z_{t,j}} \cdot N_{t,j} \cdot (1 - e^{-(F_{t,j} + M_{t,j}) \cdot \Delta t}) \cdot w_{age}.$$

Harvest control rules

The simulation approach can be used as a tool to convert survey biological information and relative assessment into quantitative HCRs. The options implemented in the simulation model are based on the following aspects: QZ , gear selectivity (size at first capture $L_{50\%}$ and selection range) and fishing activity (alone or in combination). These three are inputs that can be used to simulate different exploitation scenarios. The effects of HCRs (selectivity and fishing activity) are then analysed in terms of the sustainability of the population in the long-term. For example, the ratio between the mean spawning stock biomass and the mean unexploited spawning stock biomass ($SSB/USSB$, output) is also estimated for each harvesting scenario.

A vector of yield (Y) by time is also simulated, estimating the catch (C) according to the following general equation (Gulland, 1969):

$$C_{\Delta t} = \int_0^{\Delta t} F \cdot N_0 \cdot e^{-Z \cdot \tau} d\tau = \frac{F}{Z} N_0 \cdot (1 - e^{-Z \cdot \Delta t})$$

where Δt is the time to which the catch is referred. Thus the catch (Yield) in the time interval $(t, t+\Delta t)$ is computed in *Aladym* as (Sparre and Venema, 1998):

$$Y_{t,j} = \frac{F_{t,j}}{Z_{t,j}} \cdot N_{t,j} \cdot (1 - e^{-(F_{t,j} + M_{t,j}) \cdot \Delta t}) \cdot w_{age}.$$

Software

Aladym is written in the R language and licensed as open source under GPL2. The data and parameters feeding the model can be easily entered using an excel data sheet. The results of the simulation are stored into three Export files (.din for inputs, .dou for outputs, .RData for the R workspace) and saved in the same directory where R is started using the basename of the input sheet.

To give an idea of the running time, *Aladym-r* requires about 25 seconds (assuming 40 years of start loop and 20 years of simulation) with a Intel (R) Pentium (R) personal computer with a processor of 1.70 GHz and 1 GB RAM. The tool *Aladym-z* requires about 2.6 hours (assuming 40 years of start loop and 20 years of simulation) with the same computer. The

software can be downloaded from the Fisboat web-site, where also a detailed description of the input sheet for user help is available.

Inputs

Input parameters to the *Aladym-r* model are:

- von Bertalanffy growth parameters by sex with associated variability,
- length-weight relationship parameters by sex;
- maturity ogive parameters by sex ($L_{m50\%}$ and $L_{m25\%}$ - $L_{m75\%}$ range);
- natural mortality by sex (a constant value or a vector);
- seed values (minimum, maximum, \ln -mean and \ln -standard deviation) of recruitment by sex;
- proportion of offsprings entering in the stock by month;
- stock-recruitment relationship parameters or a vector of recruit numbers by month both with associated variability;
- time elapsing from spawning to birth;
- sex-ratio (female/total) of offsprings;
- F_{\max} by month (option 2) or from the model (option 1);
- QZ by sex;
- selection ogive parameters (2 options) of the gear used by the fleet ($L_{50\%}$ and $L_{25\%}$ - $L_{75\%}$ range, $D_{50\%}$ in case of the selectivity option 2);
- fishing activity coefficient by month (0, in case of absence of fishing activity).

Outputs

The outputs automatically produced by the simulations of *Aladym-r* can be summarised in the following items:

- Export data file (.dou):
- exploited and unexploited population by sex, per month and age;
- exploited and unexploited biomass by sex, per month and age;
- exploited and unexploited population of females belonging to the spawning stock per month;
- total mortality Z calculated by the model for females, males and the whole population in each month and year of the simulation as follows (Sinclair, 2001):

$$Z_t = \frac{1}{\Delta t} \ln \left(\frac{\sum_{j=1}^{\infty} N_{t,j}}{\sum_{j=2}^{\infty} N_{t+\Delta t,j}} \right);$$

- exploited and unexploited biomass per month;
- exploited and unexploited spawning stock biomass per month;
- ratio between exploited and the unexploited spawning stock biomass per month;
- average length and age of exploited and unexploited populations per month;
- average length and age of exploited and unexploited spawning populations per month;
- yield in tonnes per month;
- average length and age of catches per month;
- fishing mortality per month calculated as;

$$F_t = \frac{1}{\Delta t} \ln \left(\frac{\sum_{j=1}^{\infty} N_{t,j}}{\sum_{j=2}^{\infty} N_{t+\Delta t,j}^F} \right),$$

where $N_{t+\Delta t,j}^F$ is the number of survivors at the time $t+\Delta t$ under the hypothesis that only fishing mortality is acting;

- biomass of natural losses and total biological production per month.

Plots per year of the outputs listed from items 4 to 13 are also produced. Some other outputs are also made available to the user:

- average length at age and age by sex;
- natural mortality at age/length by sex;
- weight at age/length by sex;
- proportion of mature individuals at age/length by sex.

These outputs help the user to check the results obtained from the sub-models, in particular those related to the VBGF, the length-weight relationship, the natural mortality, and the maturity.

Practical guidelines

The *Aladym* core model does not make any fixed or hidden (from the user) assumption about the values of the parameters describing the behaviour of the equations on which the model itself is built. The user is allowed to (and needs to) input all the parameters involved: whilst this makes the model highly flexible in adapting to different species/environments it loads the user with the responsibility to validate each single value and to assess the coherence as a whole. Very few checks are foreseen at the moment to supervise the consistency of the data supplied: often it is a critical analysis of the results which spots such consistency. The checks guarantee the positivity of F_{max} , of length at t_0 and a sex ratio between 0 and 1.

The model is extensively based on a closed form solution to the dynamical equations it solves. Thus two key options, both related to the early phase, are available for tuning: the ‘Multiplier of Life-span’ which controls the amount of years that must be simulated in order to cancel the artefacts from the equilibrium model used to initialise the population; and the ‘Number of Run for seed randomization’ which sets the number of samples to be taken in order to derive the average values for the growth and population parameters. For both parameters the rule is: bigger is better, however the default values (1, 100) are a reasonable choice.

One of the parameters highly influencing the behaviour of the model is QZ which, however, does not have an immediate counterpart but can be naively associated to the total mortality Z . A specific tool (*Aladym-z*) has been developed which, starting from the observed values of Z and the description of the life and population traits, is able to calculate values of QZ which better approximate the given scenario. Starting from the $Z_{observed}$, *Aladym-z* iterates the model modifying, in each run, the amplitudes of the QZ waveforms. It stops when the Least Square convergence criteria are met.

Sensitivity

The extensive number of simulation runs performed has shown that the model behaviour is influenced by the consistency between the set of life-history parameters and the population

dynamics. The model results are thus expected to be particularly sensitive to the stock-recruitment relationship and natural mortality.

Strengths/weaknesses

In *Aladym* the following points can be considered the strengths:

- the model is designed to work in the absence of fishery-dependent information;
- the model is built using separated components that give it enough flexibility to account for the use of different equations;
- the model allows the population dynamics to evolve in a very detailed time scale, thus permitting analysis of fluctuations within the year;
- the detailed time scale allows modelling the effects of the harvest control over the year;
- the model allows input of natural mortality varying by age/length, and is thus able to allow for species exploited at an early phase.

The following points can be considered as the weak ones:

- the model does not account for environmental changes, such as those related for example to temperature variations, or food availability;
- the life-history traits that are used for modelling the population dynamics (e.g. growth, natural mortality, maturity) are assumed stable along the time and not to be density dependent; only direct effects of the fishery on the population are considered;
- the model does not include components of spatial behaviour;
- harvesting scenarios based on the control of the total catches are not foreseen;
- the user should be aware of the range of validity of the sub-model parameters such as those related to the stock-recruitment relationships.

References

- Beverton R.J.H., Holt S.J. (1957). On the dynamics of exploited fish populations. U.K. Min. Agric. Fish., *Fish. Invest.* (Ser. 2), 19: 533 pp.
- Barrowman N.J., Myers R.A. (2000). Still more spawner-recruitment curves: the hockey stick and its generalizations. *Can. J. Fish Aquat. Sci.*, 57: 665-676.
- Chen S., Watanabe S. 1989. Age Dependence of Natural Mortality Coefficient in Fish Population Dynamics. *Nippon Suisan Gakkaishi*, 55(2): 205-208.
- Gulland J.A. (1969). Manuel des methods d'évaluation des stocks d'animaux aquatiques. Première partie-Analyse des populations. *Manuels FAO de science halieutique*, N. 4 : 160 pp.
- Quinn T.J.II, R.B. Deriso (1999). *Quantitative fish dynamics*. Oxford University Press, Oxford, U.K.: 542 pp.
- Ricker W.E. (1954). Stock and recruitment. *J. Fish. Res. Bd. Can.*, 11: 559-623.
- Shepherd J.G. (1982). A versatile new stock-recruitment relationship for fisheries, and the construction of sustainable yield curves. *Journal du Conseil, Conseil Internationale pour l'Exploration de la Mer*, 40 (1): 67-75.

Sinclair, A. F. (2001). Natural mortality of cod (*Gadus morhua*) in the Southern Gulf of St Lawrence. *ICES Journal of Marine Science*, 58: 1-10.

Sparre P., Venema S.C. (1998) - Introduction to tropical fish stock assessment. Part 1, manual. *FAO Fish. Techn. Pap.*, 306 (1) Rev. 2: 407 pp.

Thompson W.F., Bell F.H. 1934. Biological statistics of the Pacific halibut fishery. 2. Effect of changes in intensity upon total yield and yield per unit of gear. *Rep. Int. Fish. (Pacific Halibut) Comm.*, (8): 49 p.

Working documents

Lembo G, M.T. Spedicato (2006a). ALADYM – Age-Length based dynamic Model. In: Petitgas P. (coordinator). *Fisboat Activity Report*, Appendices: 113-114 p.

Lembo P. and M.T. Spedicato (2006b). ALADYM Age-Length Based Dynamic Model. MEDITS Coordination Meeting, Kavala, Greece, April 2006.

Lembo G., Spedicato M.T. (2006c). Red mullet assessment in the GSA 10 using Aladym simulation model. Sub-Committee Stock Assessment SAC-GFCM, FAO Headquarters, 11-14 September 2006. 6 pp.

Lembo G., Spedicato M.T., Abella A., Fiorentino F, Martino S. (2006d). Aladym description (version 06). *Fisboat Meeting*, IJmuiden 06-10 November 2006, 11 pp.

Lembo G., Spedicato M.T. (2006e). Aladym: un modello previsionale di valutazione e gestione per le risorse della pesca nel Mediterraneo. 5° Convegno Nazionale delle Scienze del Mare, promosso ed organizzato da CoNISMa in collaborazione con A.I.O.L., S.I.B.M. e S.It.E. Viareggio, 14-18 Novembre 2006. abstract.

Lembo G., Martino S., Abella A.J., Fiorentino F. and M.T. Spedicato. 2007. *ALADYM* (Age-Length Based Dynamic Model): a stochastic simulation tool to predict population dynamics and management scenarios using fishery-independent information. Scientific Advisory Committee-Sub-Committee Stock Assessment. Workshop on trawl survey based monitoring fishery system in the Mediterranean, Rome, Italy, 26-28 march 2007. 6 pp.

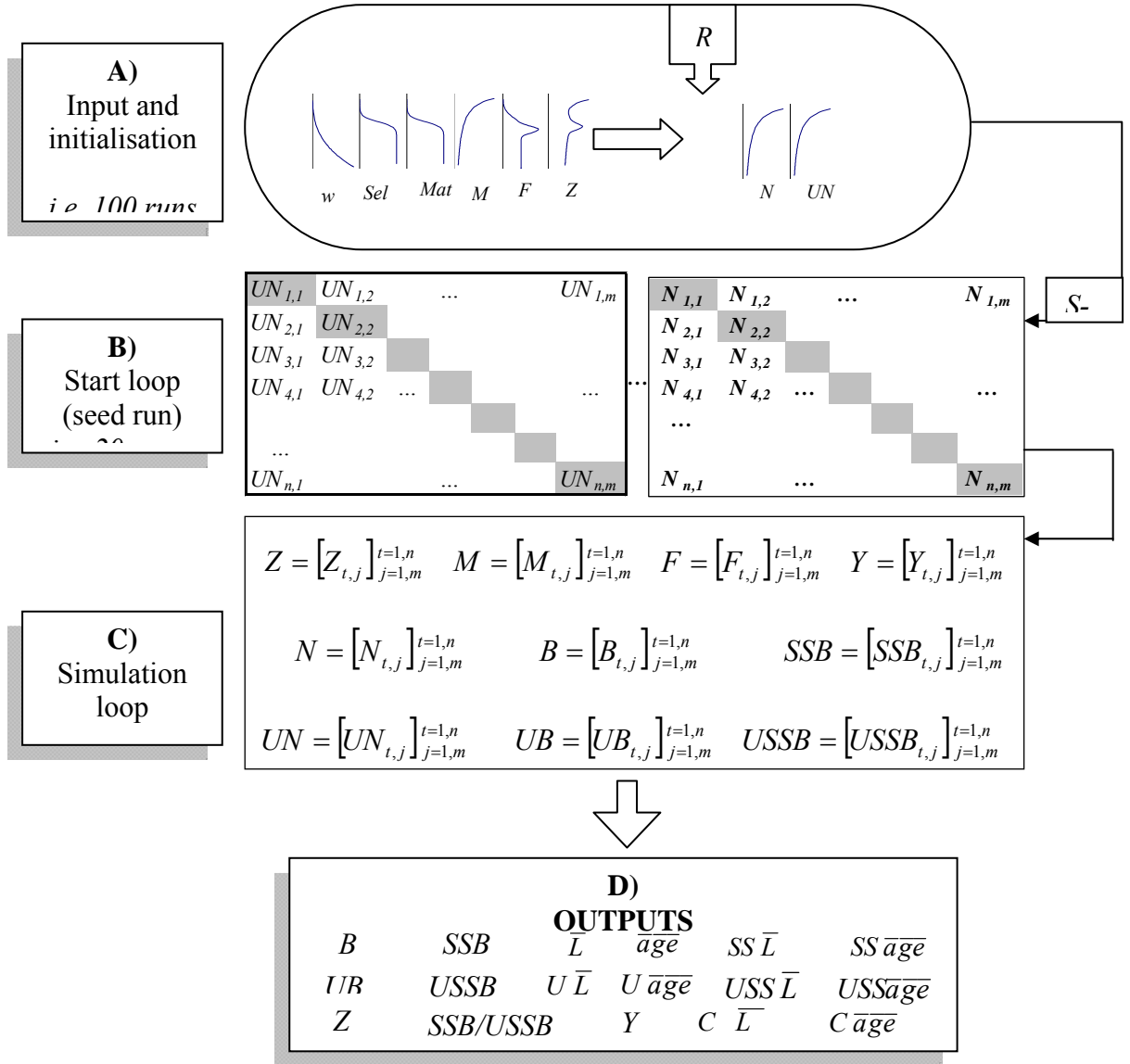


Figure 5.2.1. Scheme of the *Aladym-r* tool. R =recruitment; w =individual weight; Sel =selectivity; Mat =maturity; M =natural mortality; F =fishing mortality; Z =total mortality; N =exploited population, UN =unexploited population, B =exploited biomass, SSB =exploited spawning stock biomass, UB =unexploited biomass, $USSB$ =unexploited spawning stock biomass, $S-R$ =stock-recruitment relationship, Y =yield, t =time, j =cohort.

5.2.2 Estimating indicators and reference points

Lembo G., A. Abella, F. Fiorentino, S. Martino, and M.T. Spedicato
COISPA Tecnologia & Ricerca, Bari, Italy

Introduction

Aladym-q adds to the same mathematical model of *Aladym-r* the capability to deal with the stochastic representation of some input parameters, in order to evaluate the corresponding distribution functions of the output variables using a MonteCarlo approach. This feature aims to build a procedure to help identification of indicators and/or reference points, associating a confidence interval with them.

Derivation

The stochastic dynamic tool Aladym-q

The stochastic dynamic model defined as *Aladym-q* follows the same basic formulations as *Aladym-r*. The main difference consists in modelling the uncertainty of estimates related to the initial recruitment, growth and maturity traits of the population through stochastic processes. Moreover, a uniform distribution is applied to the number of recruits generated by the stock-recruitment relationship. In addition, probability distribution functions (*pdfs*) selected by the user are applied to the growth parameters K and L_∞ , and to the maturity parameters. This makes *Aladym-q* more adaptable for estimating the probability associated to metrics, indicators and reference points.

An operational framework of the *Aladym-q* is in fig. 5.3.1. The step AA) concerns the *input and initialization*. Given the parameters of the identified *pdfs* a first random realization is made in this step. Then the population evolves in the steps BB) and CC). These steps are reiterated for a number of realizations, sampling at each run a new set of parameters from the *pdfs*. In the output step *pdfs* and cumulative *pdfs* are generated, the latter calculated according the following general formulation:

$$f(X) = P(X < x) = \int_{-\infty}^x pdf(\chi) d\chi$$

Software

Aladym is written in the R language and licensed as open source under GPL2. The data and parameters feeding the model can be easily entered using the same excel data input sheet as *Aladym-r*. The differences regard the number of realizations to be performed (user selected and mandatory for *Aladym-q*) and the parameters of the *pdfs* associated with growth and maturity, that for *Aladym-q* operate also in the simulation loop. The results of the simulation are stored into three Export files (.din for inputs, .dou for outputs, .RData for the R workspace) and saved in the same directory where R is started using the basename of the input sheet.

To give an idea of the running time, using a Intel (R) Pentium (R) personal computer with a processor of 1.70 GHz and 1 GB RAM, *Aladym-q* might requires 572 seconds for 100 realizations, ~1.5 hours for 1000 realizations and about 17 hours for 10000 realizations (assuming 40 years of start loop and 20 years of simulation). The software can be downloaded from the Fisboat web-site, where also a detailed description of the input sheet for user help is available.

Inputs

As regards the *inputs*, besides those already mentioned for *Aladym-r*, *Aladym-q* requires:

- the number of realizations;
- the parameters of the *pdfs*.

Outputs

The *outputs* automatically produced by the simulations of *Aladym-q* can be summarised in the following items:

- Export data file (the quantities are related to each realization):
 1. exploited and unexploited biomass in tons per month;
 2. exploited and unexploited biomass of spawners in tons per month;
 3. ratio between exploited and unexploited spawning stock biomass per month;
 4. Z calculated by the model combined for sex per month and by sex per year;
 5. QZ (the input values) by sex;
 6. average length and age of exploited and unexploited populations per month;
 7. average length and age of exploited and unexploited spawner populations per month;
 8. F per month;
 9. yield in tons per month;
 10. average length and age of the catches per month;
 11. biomass of natural losses and total biological production in tons per month.
- Plots of the *pdfs* and the cumulative (*cpdfs*) are interactively produced per year for the same items listed above.
- Some other outputs are also made available to the user:
 12. average number of recruits at each realization;
 13. growth and maturity parameters by sex at each realization.

These outputs help the user to check the results from the sub-models related to the VBGF, the maturity, and the recruitment. In addition, they also allow the outputs at each realization to be tracked with the related key-inputs.

Practical guidelines

The same considerations that were developed for *Aladym-r* hold for *Aladym-q*. A new parameter is introduced for tuning the quality of the output *pdfs*: the number of realizations. This parameter should be set so as to account for a trade-off between the running time and the target confidence level. Experiments showed that values in the range from 1000 to 10000 give an error level varying from about 6-7 to ~1%. These confidence levels are well below the precision by which most of the input parameters are known.

As regards sensitivity and the strengths/weaknesses of the models, similar consideration as were developed for *Aladym-r* can be applied to *Aladym-q*, although the latter tool has the advantage of including stochastic effects in some of the key life-history traits. This stochasticity masks the effects due to uncertainty on the knowledge of input data and of their relationships.

References

See section 5.2: Simulating population dynamics. Aladym model.

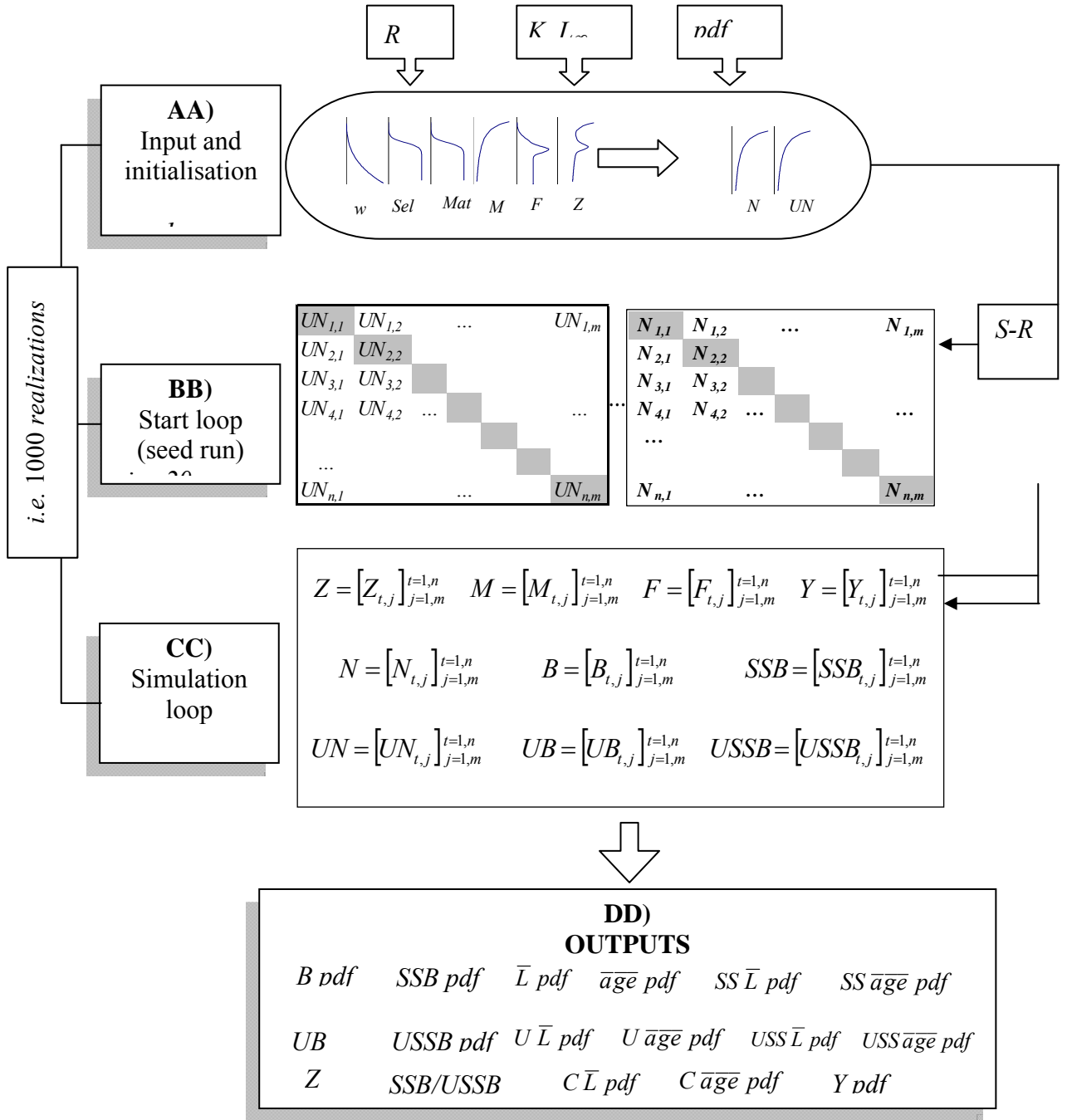


Figure 5.3.1. Scheme of the *Aladym-q* tool. *pdf*=probability distribution function; K , L_{∞} growth parameters, R =recruitment; w =individual weight; *Sel*=selectivity; *Mat*=maturity; M =natural mortality; F =fishing mortality, Z =total mortality; N =exploited population, UN =unexploited population, B =exploited biomass, SSB =exploited spawning stock biomass, UB =unexploited biomass, $USSB$ =unexploited spawning stock biomass, $S-R$ =stock-recruitment relationship; \bar{L} = average length ; \bar{age} = average age ; SS =exploited spawner's population; USS =unexploited spawner's population; C =capture in numbers; Y =yield, t =time, j =cohort.

5.3 Indicator time-series methods.

5.3.1 Nonparametric method for determining recent trends

Verena Trenkel
Ifremer, Nantes, France

Introduction

The most commonly used method for determining the direction of changes in estimated indicator time series is fitting linear models and then using the sign of the slope if it is significantly different from zero ((e.g. Trenkel and Rochet 2003). This method is reliable for determining long term time trends. However, it is less satisfactory for short term time trends. Trends might not be linear and inter-annual variability in estimated indicators can be strong enough to mask short term changes. Furthermore, if only the trend over the most recent years is considered, the overall evolution is not taken into account, for example whether the indicator values are among the lowest or highest of the available series or whether the indicator has a tendency to fluctuate randomly with a certain phase. Consider the fluctuations of plaice ln-abundance estimated for the Southern North Sea (Fig. 5.7.1, left panel). The population seems to have been fluctuating randomly over the course of the 23 years. So looking at this picture one would probably conclude that the recent years are not any different from the whole time series. However, depending on how many recent years are used for estimating a linear trend, a positive, negative or no trend will be found. Similarly for dab, although ln-abundance has been decreasing in the most recent years, overall the population levels remain well above that at the beginning of the series. So, it seems desirable to include the whole time series in the assessment of the dynamics of the most recent years.

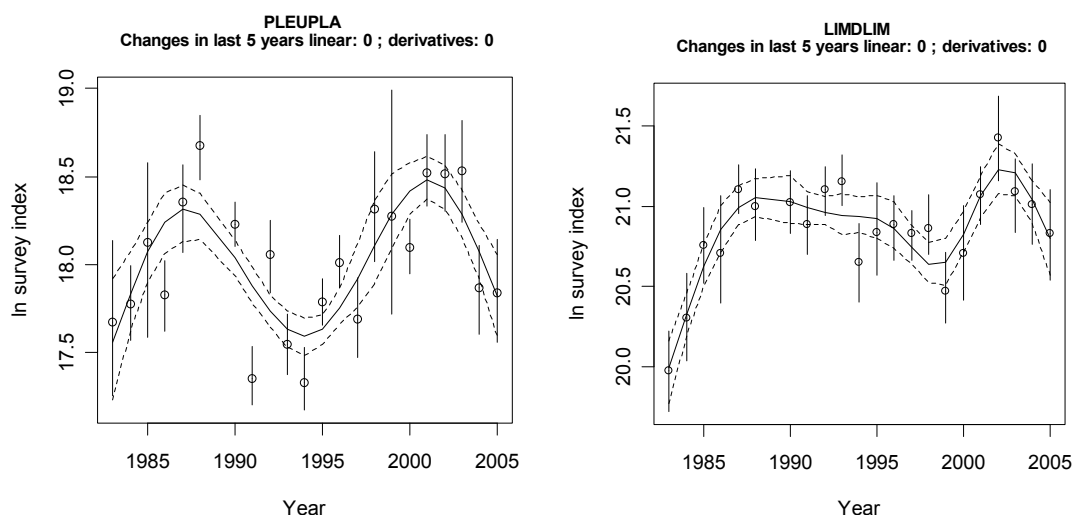


Figure 5.7.1. Time series of ln-abundance for plaice and dab in Southern North Sea based on IBTS data. The continuous line is generalised additive model (GAM) fit. The broken lines are 95% confidence bands for this fit based on a parametric bootstrap of annual indicator estimators.

In this document a method is proposed to estimate the direction of recent changes making use of first and second derivatives of smoothed indicator time series and the position of the most recent years with respect to the full time series. The first derivative, which is actually the

local slope (tangent) at each point of a function, here each year, describes the dynamics of the indicator changes. In contrast, the second derivative describes the changes in the slope. A positive second derivative indicates that the slope is increasing while a negative second derivative means that the slope is decreasing. A location at which the second derivative is zero is called a change point as at this point the dynamics change from accelerating to decelerating, i.e. the slope gets smaller from this point onwards, or vice versa. The slope will be zero when either a maximum or minimum is reached. Figure 5.7.2 illustrates the signs of the first and second derivatives. The proposed method is described in details in the next section.

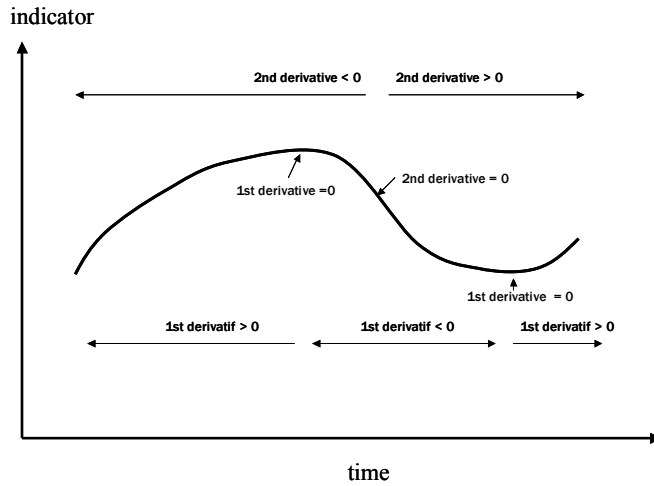


Figure 5.7.2. Diagram showing the signs of the first and second derivative of an indicator time series used for determining the direction of change.

Method

The proposed method consists of several steps which are

1. fit a generalised additive model to the time series in order to obtain a smoothed series;
2. calculate first and second derivatives for the smoothed time series for all years (including years with no data);
3. determine direction of change in recent years using a combination of criteria for the smoothed series as well as the first and second derivatives of the smoothed series.

To obtain smoothed indicator series, generalised additive models (GAM) are fitted with *year* as a cubic regression spline and automatic selection of the degrees of freedom (minimum 3) using the *mgcv* package in R (R development Core Team 2003) developed by Wood (2000).

As spline models are twice differentiable, first and second derivatives of the smooth series can be calculated for every year of the time series using an approximation based on first and second differences, as used by Fewster et al. (2000).

The first derivative of indicator I in year t is approximated by the first difference

$$\hat{I}'(t) = I\left(t + \frac{1}{2}\right) - I\left(t - \frac{1}{2}\right)$$

The second derivative is approximated by the sixth difference

$$\hat{I}''(t) = \frac{1}{180} \{ 2I(t+3) - 27I(t+2) + 270I(t+1) - 490I(t) + 270I(t-1) - 27I(t-2) + 2I(t-3) \}.$$

For the third ($t=3$) and two before last years the fourth difference is used to approximate the second derivative

$$\hat{I}''(t) = \frac{1}{12} \{ -I(t+2) + 16I(t+1) - 30I(t) + 16I(t-1) - I(t-2) \}$$

Finally, for the second ($t=2$) and one before last the second difference is calculated for the estimating the second derivative

$$\hat{I}''(t) = I(t+1) - 2I(t) + I(t-1)$$

For obvious reasons the first and second derivatives cannot be estimated for the first and final year of the time series.

In order to determine whether the estimated first and second derivatives are significantly different from zero for a given year, i.e. $\hat{I}'(t) \neq 0$ and $\hat{I}''(t) \neq 0$ for $i=1, \dots, T$, a parametric bootstrap is carried out. For this, indicator time series are created by resampling each data point (year) $I(t)^b \sim N(I(t), \sigma(t))$ from a normal distribution with, as mean, the estimated indicator value for year t , $I(t)$ and, as standard deviation, its estimated standard deviation $\sigma(t)$. A separate GAM is then fitted to each bootstrap series $I(1)^b \dots I(T)^b$, $b=1, \dots, B$, using the same degrees of freedom (degree of smoothness) as was found optimal for the original indicator time series. Subsequently, for each bootstrap sample, first and second derivatives are estimated by year. This provides the distribution of first and second derivatives for each year based on which the 2.5 and 97.5 percentiles are calculated. If the value zero is included in the interval formed by the 2.5 and 97.5 percentiles, which is actually a 95% confidence interval, the derivative of the given year is not significantly different from zero and the indicator variable for the derivative is set to zero, otherwise the sign of the derivative is either positive (1) or negative (-1) depending on whether the values within the confidence interval are all negative or positive. The result of this test is a time series of an indicator variable for the first derivative which is either 0, 1 or -1. Similarly for the second derivative.

In order to determine the direction of recent changes in indicator time series, the indicator variables with the signs of the first and second derivatives are combined in a decision rule (Table 5.7.1). In addition, the location of the minimum and maximum value in the time series is used in order to put the most recent years into the perspective of the whole time series. If the maximum is not found within the last three years and the annual slopes (first derivative) are predominantly negative and annual second derivatives are negative or zero in the last five years (no change point appears with sign of second derivative passing from -1 to $+1$), the direction of change is declared as recently decreasing. The second derivative is used to establish whether an improvement has already taken place most recently. Similarly for a recently increasing series, the minimum should not be within the last three years, the average of the annual slopes should be positive (apart from one year) and no change for a decreasing trend (sign of second derivatives positive) should have occurred during the last five years. For all other cases there is no indication for a change. These decision rules are proposed based on empirical tests, however they are by no means prescriptive. The important point is the principle, i.e. the combination of different measures of the dynamics of a time series,

minimum, slope and change points. In particular the time spans considered, which is five years for the first and second derivatives and three years for the location of the maximum and minimum, are easily adapted for a particular study.

Table 5.7.1. Decision rules used to determine direction of recent changes for an indicator time series based on first and second derivatives of smoothed indicator time series. y_m = number of years to consider for minimum/maximum; e.g. $y_m = 3$. y_c = number of years to consider for first and second derivatives, e.g. $y_c = 5$

Decrease	Increase
1. Maximum value before final y_m years	1. Minimum value before the final y_m years
AND	AND
2. Signs of annual slopes for final y_c years negative or at most 0 for 1 year	2. Signs of annual slopes during at least final y_c years positive or at most 0 for 1 year
AND	AND
3. Sign of annual second derivatives during final y_c years negative or zero (persistence of decrease)	3. Sign of annual second derivatives during final y_c years positive (persistence of increase)

For comparison purposes, linear time trends over the whole data series and the last five years are also calculated.

Example: cod in North Sea

As an example the method was applied to the indicator table for cod in the North Sea based on IBTS data. Fig. 5.7.3 gives the smoothed indicator time series. The direction of recent change assessed by two methods is indicated in the header of each figure. The proposed method is referred to as ‘derivatives’. Thus the diagnosis obtained with the proposed method is that ln-abundance and L50 maturity are decreasing, while mean length and the length quartiles are all increasing. Thus all signs points towards a deterioration of this cod stock. In contrast to the proposed method, linear time trends over the most recent five years were only significant ($\alpha=0.05$) for the ln-abundance time series.

The total mortality estimates Z and the length quartiles L_{25} and L_{75} in Fig. 5.7.3, are varying interannually more than seems plausible biologically. As a consequence the smooth function fits (cubic splines) might not be considered representative for the temporary evolution of these indicators and the resulting diagnoses might be considered unreliable. This example points out the need to carefully select the indicators and to evaluate their reliability before using them for any assessment purposes because the results obtained with the proposed method will entirely depend on the suitability of the GAM fits.

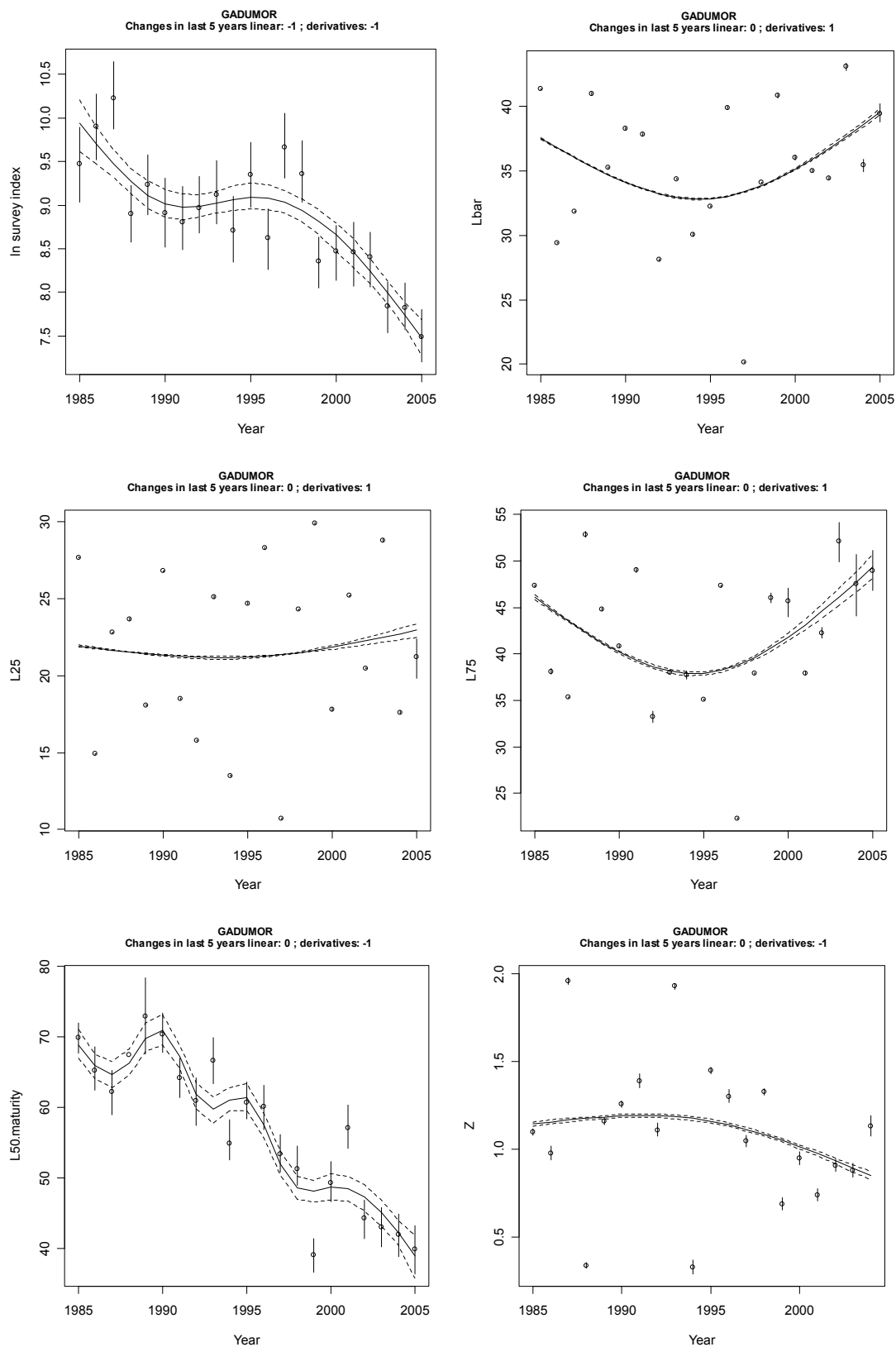


Figure 5.7.3. Indicator time series of North Sea cod with cubic spline model. Assessment of recent direction of changes in figure headers using the proposed method ('derivatives') and linear trend estimation for the final five years.

Software

The software for the non-linear estimation procedure based on first and second derivatives can be downloaded from <http://www.ifremer.fr/drvecohal/fisboat/index.htm>. To run it:

1. Copy the function “FunctionsTimeChangeDerivatives.R” and script “ScriptTimeChange.R” into same folder as the data table.
2. Edit line 10 in script replacing file name,
e.g. `indicest<-read.table("codNS_tab2_wp2A.txt",header=T,sep="\t",as.is=T)`
3. Edit line 28 to select the time horizon for first and second derivatives
e.g. `lastn=5`
4. Edit line 31 for time horizon for maximum and minimum values
e.g. `lastnmin=3`
5. Run edited script.
6. Results are obtained as a table called `Trendestimates.txt` and as smoothed time series plots for each indicator

Example results for North sea cod

1. Trendestimates.txt

Area	Species	Indicator	LinearSlope	PvalueAll	LinSlopeLastYears	PvalueLast	DiagnosLinearRecent	DiagnosNonLinearRecent
NorthSea	GADUMOR	ln_survey_index	-0.067	0.00014	-0.219	0.04792	-1	-1
NorthSea	GADUMOR	Lbar	0.101	0.61485	0.996	0.47505	0	1
NorthSea	GADUMOR	L25	0.051	0.80099	-1.09	0.51169	0	1
NorthSea	GADUMOR	L75	0.161	0.55643	2.73	0.13226	0	1
NorthSea	GADUMOR	L50.maturity	-1.525	0	-3.657	0.06658	0	-1

Explanation of column names:

LinearSlope: Linear slope over whole time series

PvalueAll: p-value for linear slope over whole time series

LinSlopeLastYears: Linear slope for most recent years

PvalueLast: p-value for above linear slope

DiagnosLinearRecent: sign of slope for most recent years if significant (p-value ≤ 0.05)

DiagnosNonLinearRecent: direction of change using proposed method (decrease =-1, increase=1, no change =0)

2. Figures

e.g. `ln_survey_index.wmf` (fig. 5.7.4).

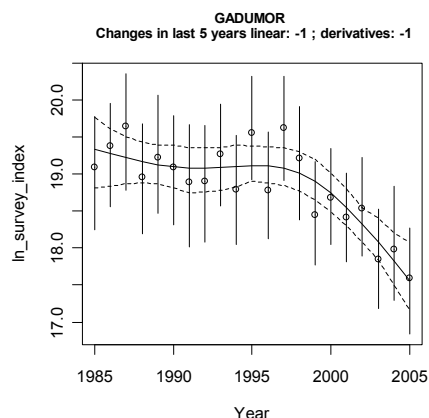


Figure 5.7.4. Ln-transformed survey time series with fitted smoothed model (GAM). Header provides direction of changes as found by fitting a linear slope over the 5 most recent years, and by the proposed method, referred to as ‘derivatives’. This information is repeated in the table.

References

- Fewster, R.M., Buckland, S.T., Siriwardena, G.M., Baillie, S.R. and Wilson, J.D. (2000) Analysis of population trends for farmland birds using generalized additive models. *Ecology (New York)* **81**, 1970-1984.
- Trenkel, V.M. and Rochet, M.-J. (2003) Performance of indicators derived from abundance estimates for detecting the impact of fishing on a fish community. *Canadian Journal of Fisheries & Aquatic Sciences* **60**, 67-85.
- Wood, S.N. (2000) Modelling and smoothing parameter estimation with multiple quadratic penalties. *Journal of the Royal Statistical Society B* **62**, 413-428.

5.3.2 Assessing the power to detect future trends

Hans Bogaards, Charlotte Deerenberg, Gerjan Piet
Wageningen Institute for Marine Resources and Ecosystem Studies
IJmuiden, The Netherlands

Introduction

An assessment of the power of a survey to detect future trends in candidate indicators is obviously an important task before those indicators are selected for use in some kind of fishery management programme. Formally, the power of a statistical test is the probability of not making a type II error, where a type II error is defined as accepting the null hypothesis when the alternative is true. Within the context of a monitoring program, power can be interpreted as the probability that a particular trend change will be detected. This note describes a method for estimating the power of a survey to detect future, linear trends given standard modelling assumptions. Software in R available from the FISBOAT website is also described.

Method

Power calculation first requires a specification of the testing procedure to be used, together with the significance level α (the probability of making a type I error: rejecting the null hypothesis when it is true) of the test. Furthermore, it requires the definition of a null hypothesis H_0 and an alternative hypothesis H_1 .

To provide a generic power calculator for the evaluation of candidate indicators, it is here assumed that the time series of an indicator can be described by a stochastic linear model with an additive normally distributed error term. Observations are assumed to be derived from annual sampling schemes. In the case of missing observations, the analysis is restricted to the longest consecutive stretch of non-missing values in the time series. The slope of the historic trend line is estimated by simple linear regression analysis, which implies that the residual variance is assumed to be constant and has no autocorrelation. The extent to which these assumptions are violated should be judged by the user. Visual inspection of the time series with its trend line and residuals is imperative.

The testing procedure concerns the slope of the trend line for a specified number of future years of follow-up. The test statistic T is defined as the difference between the observed slope of the future trend line and its anticipated value under H_0 , divided by the standard error of the estimated slope parameter. If the variance of observations about the linear trend will remain constant, T will follow a non-central t -distribution, the non-centrality parameter being equal to the slope parameter under H_1 minus its value under H_0 , divided by its standard deviation under H_0 . From this, it follows that the power of a one-sided test is the probability that T is more extreme than some critical value c . The power of a two-sided test is the overall probability that T is more extreme than either c or $-c$. Critical values are determined by the significance level of the test. Typically, critical values in a two-sided test with $\alpha = 0.05$ correspond to the 2.5th and 97.5th quantiles of a central t -distribution.

Final comment

It is assumed that the residual variance is constant throughout, not only for the duration of the historic time series, but also for the interval over which the power of an indicator is to be evaluated. Transformations to stabilize the variance and to make the trend linear may be

considered in advance, after which power should be evaluated on the transformed time series. In addition, the power of a particular indicator will be underestimated if intervention strategies tend to reduce its randomness, or overestimated if the opposite is true.

References

Gerrodette, T. 1987. A power analysis for detecting trends. *Ecology*, 68: 1364–1372.

Nicholson, M.D. and Jennings, S. 2004. Testing candidate indicators to support ecosystem-based management: the power of monitoring surveys to detect temporal trends in fish community metrics. *ICES Journal of Marine Science*, 61: 35–42.

Piet, G.J. and Jennings, S. 2005. Response of potential fish community indicators to fishing. *ICES Journal of Marine Science*, 62: 214–225.

Software

After the function templates have been downloaded from <http://www.ifremer.fr/drvecohal/fisboat/index.htm> and sent to R, the power of a candidate indicator can be obtained via a call to the function `linear.trend()`. This function requires at least two arguments:

<i>file</i>	tab-delimited text file to be used, e.g. “codNS_tab2_wp2a.txt”
<i>var</i>	candidate indicator to be evaluated, e.g. “L50.maturity”

If the candidate indicator has been calculated separately for each age-class (for example, the wp2a spatial indicators), a third argument is required:

<i>age</i>	age-class to be evaluated, e.g. “A4”
------------	--------------------------------------

The name for *var* should correspond to the variable name as provided in the header of the requested text file, whereas the name for *age* should correspond to a value of the variable Age.

Additional arguments that may be set optionally by the user are:

<i>dir</i>	directory where <i>file</i> is located (default: working directory)
<i>h0</i>	slope under H_0 (default: continuation of trend line)
<i>h1</i>	slope under H_1 (default: stabilization of trend line)
<i>alpha</i>	significance level of the test (default: $\alpha = 0.05$)
<i>horizon</i>	interval over which power is to be evaluated (default: 25 years)

Arguments acting as character strings should be enclosed in quotation marks, e.g.

```
> linear.trend(file="codNS_tab2_wp2a.txt", var="L50.maturity", h1=0, alpha=0.01)
```

Power is calculated both for one-sided and two-sided tests. In one-sided tests, an increased slope parameter is anticipated if the historic trend line was decreasing and vice versa. By doing so, ecosystem-based management objectives are evaluated more efficiently than in two-sided tests, as the latter are more conservative.

Two graphs are output by the R script. The first shows the time series that is used, together with the best-fitting linear trend line, and residuals which are plotted separately. The second shows the results of the power calculations, for one-sided and two-sided tests. Figures are also printed to the screen, with oc1 and oc2 denoting one-sided and two-sided test results, respectively.

5.3.3 Statistical Process Control (SPC) schemes

Benoit Mesnil and Pierre Petitgas
IFREMER, Nantes

Introduction

Control charts are part of the statistical process control (SPC) tools routinely used over decades to monitor manufacturing processes and signal anomalies in performance. The process has some inherent variability and is said to be 'in-control' as long as it remains within acceptable bounds. If an anomaly occurs causing a deterioration in quality beyond the baseline variability the system is said to be 'out-of-control'. Control charts are graphical displays of some summary statistic of the observation data (e.g. an indicator) against the order index of the sample (e.g. time), together with reference 'marks' based on the in-control mean and variance, that are designed to detect whether a worrisome change in process output is indicated by the current data and a fix is required. Since there are costs associated with both false alarms and quality losses, the charts' parameters are tuned to achieve a desired trade-off between the risk of false alarm and the power to detect changes promptly.

The cumulated sum type of control chart in its 'decision interval' form (DI-Cusum) has been selected for this project because it is advocated in SPC textbooks (e.g. Montgomery, 1991; Hawkins & Olwell, 1998) for the type of data considered in the survey indicators context. Control charts can be designed to monitor changes in mean level (location charts) or in variance (scale charts) of process outputs; explanations are only given for location charts here. More details can be found in the literature digest made for this project by Mesnil & Petitgas (WD) and in the papers cited therein.

Derivation

Suppose a suite of observations (individual or group means) x_i collected at time $i = 1, \dots, m$ and assume that their in-control mean μ and standard deviation σ_x are known from a pilot study or for a reference period. In the following, it is considered that the data are first standardised through the transformation $z_i = (x_i - \mu) / \sigma_x$.

The decision-interval Cusum works by recursively accumulating positive and negative deviations separately with two statistics:

$$S_i^+ = \max[0, S_{i-1}^+ + z_i - k]$$

for positive deviations ('one-sided upper Cusum'), and

$$S_i^- = \min[0, S_{i-1}^- + z_i + k]$$

for negative deviations ('one-sided lower Cusum'), with starting values normally set as $S_0^+ = S_0^- = 0$. A Cusum chart is obtained by plotting these statistics against i .

The parameter k is usually called the *reference value*, or the *allowance*, and is related to the size of the smallest shift in the level of x that one is wishing to detect quickly. Note that deviations smaller than k are ignored in the recursions above. The decision rule is to declare an out-of-control state whenever S^+ exceeds the *decision interval* h or S^- falls below $-h$. The values chosen for the parameters h and k (in standard deviation units) determine the performance of the control chart; there is no theoretical objection against setting different h - k pairs for upper and lower Cusum's if changes in one direction matter more than in the other.

The performance of control charts is generally evaluated in terms of their run length. A run is the number of sampling events that elapse between the start of the monitoring and the first alarm. Run length is a random variable whose probability distribution depends on the process and the chart parameters, and it is its expectation – called Average Run Length (ARL) – that is commonly used as a summary measure of performance. In many instances the run length distribution is very broad and skewed, and it may be misleading to only consider its mean; the experts recommend to also look at other percentiles, whenever the distribution can be computed. The notation $ARL(\delta)$ is used to designate the ARL of an SPC scheme for detecting a change of size δ (in σ_x units) occurring in the process mean level. Thus, $ARL(0)$ is the ARL of a scheme when the process actually stays in-control all the time (in-control, or IC ARL); yet, due to its inherent variability, an alarm may be raised by chance alone when the chart is updated with a new datum. In other words, $ARL(0)$ is the average time until a false alarm is raised, which should ideally be large. Conversely, if the mean of the process distribution shifts from μ to $\mu+\delta$, due to an anomaly the chart should detect this quickly, implying a short $ARL(\delta)$. Chart parameters can be tuned to achieve the desired compromise, as explained in the guidelines below.

Software

Two R scripts have been developed to implement a Cusum monitoring scheme: *CusumTutorial.r* is generic, for exploring Cusum charts with 'free-format' time series vectors; *FBCusumCharts.R* is designed to automate the production of standard tables of results for the report ('traffic light template'). Both use a set of functions stored in the separate file *CusumFuncs.r* that must be sourced into the user's R workspace (on first use) as instructed in the scripts. The scripts are meant to be run in a stepwise fashion (highlight a line or a block and submit to R) and are amply commented to guide the user.

The top part of *FBCusumCharts.R* deals with each indicator in turn. Note that a logarithm transformation is applied to the Survey and Recruit indices (columns 5 and 6); the reference period for each case study is 'hard-wired' but can be edited if needed; an indication of an appropriate value for the allowance k , based on the mean deviation from the reference mean outside the reference period, is proposed but is not coded as a default value. Once the full set of indicators has been processed, the bottom part of the script gathers the individual *resnam.#* objects to produce the table of alarms (signed Cusum values above h or below $-h$) and the table of Cusum parameters and saves them to files.

This implementation includes functions to compute in-control or out-of-control ARLs and run length distributions of one-sided Cusum for normal data, adapted from a Fortran code by F.F. Gan (1993) found on the StatLib JQT archive. They have been checked against the values tabulated in various SPC textbooks and articles, and the results match very well. They need to be optimised for R, to speed up the computation of RL distributions which requires some patience at the moment.

Practical guidelines

A) Cusum design: tuning the chart parameters k and h

With fisheries survey data, we in general have to analyse time series of one or several indicators of population status (control variables). We have one value per indicator per year (individual data) with perhaps the precision on the indicator in each year. We distinguish 2

phases: Phase I for defining the in-control (IC) period; and Phase II for designing the Cusum to signal change from the in-control state with desired performance.

Phase I

The task in Phase I is to set the IC or reference process parameters μ and σ . It is a critical phase in that the values adopted for these parameters will condition the diagnostic that will be made later. Normally, this is an experimental phase where the state of the system is closely checked, many measurements are taken and scrutinised, to retain only those that can be safely assumed to correspond to a well behaved process. In our case, we will often start with existing data collected in the past, and Phase I will essentially consist in the definition/choice of an in-control (or reference) period, and using the data in the selected to years to estimate in-control parameters μ and σ . The IC period can be defined on various criteria, including an analysis of the times series. The IC period is best defined collectively, on expert knowledge, as the period in which the population was in a satisfactory state and/or showing satisfactory dynamics. For example, within the Fisboat project, the IC period was defined collectively during a workshop as the period when the indicator value showed "satisfactory" values with low variation (no obvious outlier). Thus, the IC period may not necessarily comprise consecutive years. Sensitivity to the IC period should be analysed and the IC period may be also re-defined *a posteriori*. This is consistent with the iterative and rejection procedures described in SPC textbooks.

Phase II

In Phase II the task is to design (or tune) the Cusum scheme to signal a specified deviation from the IC mean with a desired performance, i.e. this is where the chart parameters k (allowance) and h (interval) are determined. The choice of k is based on the magnitude of the shift δ in the mean that makes "a meaningful impact" on the system, driving it out of control. The value of h determines whether an alarm is raised or not (an alarm is triggered when the cusum plot crosses the horizontal line at h , or $-h$ or $+h$ for a two sided Cusum). The rationale for choosing h is primarily based on minimising the risk of false alarm, but the ability to promptly detect shifts that matter should also be preserved. Setting h , once k is chosen, involves Run Length considerations. A four-step procedure is suggested¹:

1. Regarding k , if δ is the shift of interest (in sd units), there is broad support in the literature for setting k at half the value of that shift (formal demonstration in Chap. 6 of Hawkins & Olwell), and this rule can be safely adopted. The "meaningful" shift δ can be set after analysing the deviations from μ outside the IC period. For instance, the shift to be detected can be set to a percentile of these deviations or to their mean. For fisheries survey based population indicators, k will take in general a value between 0.5 and 1.5; too small values of k should be avoided (Hawkins & Olwell, p. 33).
2. Using tables or software with a zero value for the shift δ , search for an h that gives desirably large IC ARL(0) given k , and thus a low risk of false alarm. Larger values of h (and k) lead to larger ARLs.
3. Because the RL distribution may be quite skewed, consideration of the average RL alone may be misleading and, using the function `arldis.f` in the Fisboat R scripts, the full distribution of the in-control RL should be checked. For example, if you choose k and h to aim for a "large" IC ARL of 100, and observe a "small" value of 10 samples or less for the 25th percentile, it is likely in the actual application of the scheme that more false alarms will occur than the large ARL(0) makes you think. If so increase h .

¹ Reminder: in all this, we assume the indicator series have first been standardised.

4. The h value arrived at in the previous steps may have made you reasonably content with the false alarm risk, but you then need to ascertain that the scheme is able to quickly detect the shifts you are interested in, i.e. that its out-of-control ARL is small enough. Return to the ARL tables or software with a value of $2k$ for the shift δ , and check that the OC ARL(δ) corresponding to the envisaged h is adequately small. In general, it is possible to find values of k and h so that the OC ARL does not exceed 3 years.

As pointed out earlier the basic challenge of using and tuning SPC schemes is to find an acceptable compromise between the risk of false alarm and the power to detect shifts that matter in the state of the system, and it is often necessary to iterate through these 4 steps to arrive at that compromise. The notions of "meaningful impact", "acceptable risk" and "desired performance" are very much policy issues and have to be decided in partnership with managers and stakeholders.

B) Assumptions and effects of violations

The main assumptions underlying the statistical properties of Cusum charts are (i) that the monitored variable has a distribution from the exponential family; in particular, the run length characteristics commonly tabled in textbooks or computed with the R software coded for this project are only valid for normally distributed data; (ii) that the in-control process parameters are known rather than estimated; and (iii) that the time series of residual variation has no correlation in time. Violations of these assumptions all go in the same direction: the in-control ARL(0) experienced in practice is shorter than the value computed for the perfect case, i.e. the chances of false alarms are larger than expected (e.g. Section 3.7 in Hawkins & Olwell; Jones et al., 2004; Lu & Reynolds, 1999; Reynolds & Stoumbos, 2004). Smaller values of k (also large h) enhance the robustness to non-normality, but increase the impact of estimating the reference mean and sd from the data. An encouraging note: even though a Cusum tuned with a given k is optimal for detecting shifts of $2*k$ standard deviations, its performance remains high for actual shifts that are 'not too far' (Hawkins & Olwell, p. 54). Time series of survey data population indicators are often short (< 20 years) with marked deviations and sometimes show correlation or trend. It is advised to check the distribution of the indicator variable as well as its correlation in time. It may be necessary in some cases to transform the variable into a Gaussian or to detrend the time series. The reference period is even shorter, and we use noisy data to estimate the IC process parameters. Since all departures from the assumptions will result in effective RLs being very different (in general shorter) than values publicised for the "clean" case, an ad hoc remedy is to take relatively large h values. Conservative advice is to use (k, h) parameters giving large IC RLs: ARL > 20 years and 25th percentile of RL distribution > 10 years. When some deviations from μ outside the reference period are large in comparison to σ , it may be telling that the variance has changed or that the indicator variable is skewed. In that case, starting Phase II with a large value of k is advisable. When the value of h is small in comparison to an increasing (decreasing) Cusum deviation, it may be telling that there is correlation in time in the indicator series.

C) Strengths and weaknesses

Control charts have been in operation in many branches of industry since the 1930's and their statistical bases have been thoroughly investigated in a huge body of literature (the references below are just a tiny sample). They are still a recurrent topic of specialised journals such as the Journal of Quality Technology or Technometrics. Applications have been extended to environmental surveillance, biomedicine, clinical tests, and public health. The strengths in

these domains are that the in-control state is well defined, the monitoring involves numerous samples taken at high frequency through rigorous sampling designs, and measurement errors are often small.

In contrast, this defines the weaknesses for fisheries applications. Perhaps the main limitation is our poor ability to characterise the reference state of fisheries (or of ecosystems) with survey data that just span the recent decade(s) in a background of large variability compounded by substantial sampling variance (i.e. we do not have a proper Phase I). Keep in mind, however, that the reference state does not imply perfect stability; the goal of control charts is to spot those events where the state of the system jumps beyond the domain of its inherent variability.

A virtue of the Cusum approach is that it does not presume the nature of the change (linear, trend or otherwise) and treats positive and negative deviations equally. Cusum charts are best suited to detecting small, persistent changes. Anomalies in the system can take the form of shifts in the mean and/or changes in the variance of the distribution. Specific control charts can deal with both situations. Actually, it is common to combine location and scale charts to enhance the detection performance for both small and large shifts (Reynolds & Stoumbos, 2004).

It has been demonstrated that, among the procedures that have similar in-control ARL(0), the Cusum has the smallest expected time until a change is detected when it occurs. This is the basis of the rationale for tuning the chart, with priority given to achieving large ARL(0). The emphasis on low risk of false alarm has some practical advantage in our application to fisheries management and its overly controversial atmosphere; we have learnt to know that casting assessment noise straight into fisheries regulations has damaged our credibility and our relations with the industry, and a method that explicitly aims to avoid this should help. Lastly, the biggest advantage of the Cusum is that it is so simple to implement. Yet, it provides a formal framework to establish diagnostics in an objective and replicable way.

References

The oft-quoted textbooks

Derman, C. & S.M. Ross. 1997. *Statistical aspects of quality control*. Academic Press, London.

Hawkins, D.M. & D.H. Olwell. 1998. *Cumulative sum charts and charting for quality improvement*. Springer-Verlag, New York.

Manly, B.F.J. 2001. *Statistics for environmental science and management*. Chapman & Hall/CRC, Boca Raton.

Montgomery, D.C. 1991. *Introduction to statistical quality control*. Wiley, New York. (5th Edition 2005).

Whetherill, G.B. & D.W. Brown. 1991. *Statistical process control: theory and practice*. Chapman & Hall, London.

Articles on SPC methods

- Gan, F.F. 1993. The run-length distribution of a cumulative sum control chart. *J. Quality Technology* 25: 205-215.
- Jones, L.A., C.W. Champ & S.E. Rigdon. 2004. The run length distribution of the CUSUM with estimated parameters. *J. Quality Technology* 36: 95-108.
- Lu, C.-W. & M.R. Reynolds Jr. 1999. Control charts for monitoring the mean and variance of autocorrelated processes. *J. Quality Technology* 31: 259-274.
- Luceño, A. & J. Puig-Rey. 2002. Computing the run-length probability distribution for CUSUM charts. *J. Quality Technology* 34: 209-215.
- Mesnil, B. & Petitgas, P. (WD). Statistical Process Control (SPC) schemes for indicator-based assessments. A literature review and manual for the Fisboat project, 2006. 30 pp.
- Page, E.S. 1961. Cumulative sum control charts. *Technometrics* 3: 1-9.
- Reynolds, M.R. & Z.G. Stoumbos. 2004. Control charts and the efficient allocation of sampling resources. *Technometrics* 47: 409-424.
- Scandol, J.P. 2003. Use of cumulative sum (CUSUM) control charts of landed catch in the management of fisheries. *Fish. Res.* 64: 19-36.
- Woodall, W.H. & B.M. Adams. 1993. The statistical design of Cusum charts. *Quality Engineering* 5: 559-570.

5.3.4 Nonparametric statistical methods for assessing trends.

John Cotter
CEFAS, Lowestoft

Introduction

This note describes a selection of nonparametric statistical methods thought to be useful for assessing trends in fishery statistics or indicators, e.g. abundance-at-age, mean length, geostatistical indices, or just about any continuous variable. The trends referred to here are assumed to relate to time but they could also relate to a transect over a spatial dimension. The literature on trends is extensive so this note can only provide a modest introduction to it. Loftis et al. (1991b) point out that formal statistical methods do not usually reveal trends that are not apparent from inspection of the data but they are useful for allowing different data analysts to reach similar conclusions from the same data and assumptions. Several of the references cited come from the literature on monitoring of water pollution where the sporadic and chaotic nature of variation combined with frequent gaps in the time-series has stimulated development of nonparametric methods because of their minimal assumptions. Fisheries scientists typically prefer modelling, i.e. parametric methods for assessing trends in fish stocks and have exploited nonparametric methods relatively lightly. All the same, trends in fish stocks could be established with less reliance on assumptions about the data and models if nonparametric methods were used. Furthermore, interest nowadays is shifting from estimation of quantities of fish in a single commercial stock to assessment of whole ecosystems, a task for which well established, structural models are not always available.

There has been little discussion in the literature of statistical inference in relation to trends. This note therefore begins by proposing some points thought to be important. A variety of nonparametric statistical tests tailored for assessing trends is then introduced, some of which are easy to calculate with a spreadsheet but limited to providing only the most general statements, e.g. the binomial test with the median, and others which are more elaborate and specific, e.g. Mann-Kendall's Tau which finds monotonic trends. Multivariate tests presented include Cochran's Q and the Dietz-Killeen test. A spreadsheet accompanying the paper illustrates application of each method to a single test set of data, namely a set of abundance-at-age figures for cod from the North Sea IBTS quarter 1 survey. There were few problems in Excel for the univariate tests once the relevant functions had been discovered (e.g. RANK(), MEDIAN(), and BINOMDIST()) although some methods were quite labour intensive. The spreadsheet can be downloaded from <http://www.ifremer.fr/drvecohal/fisboat/index.htm>. Alternatively, the methods could easily be implemented in R (and several already are). R code is available from the same site for the Dietz and Killeen multivariate trend test.

Inferring about trends

A distinction is acknowledged here between the true, unknown trend, called the signal, and the measures of it made with error, called observations. Most analyses of trends have to be based on the following assumptions: at all times, t ,

- $E(\text{measurement error}) = 0$
- $E[(\text{measurement error at } t) * (\text{measurement error at } t+\Delta)] = 0$ where Δ is any lag interval, and
- $E(\text{measurement error} * \text{signal}) = 0$.

$E()$ is the operator for statistical expectation. In words, measurement errors should average to zero, have no serial correlation, and be independent of the level of the signal. Failure of any of these assumptions could lead to spurious trends unrelated to the signal.

“Trend” is hard to define more specifically than our intuitive understanding of a general movement up or down of an observed variable. A trend can occur in the observations, in the signal, or in both. Usually we think of a trend as monotonic upwards (or downwards), i.e. with every observed value at time t equal to, or higher (or lower) than that at $t-1$. However, turning points, either real or error-based, are likely to occur too. Their occurrence in an observed series is not necessarily, by itself, an accurate indicator of the position of the turn in the signal, or of its magnitude at that point. For an explanatory analogy of this, consider flying over a mountain range and dropping weights at fixed intervals without looking where; some might fall on high ground but few, if any, will fall exactly on the turning points of height, e.g. the mountain peaks. Step changes can also occur in time series and may look like smooth trends when obscured by observation errors. Nonparametric methods for inferring the location and magnitude of a step are discussed by Pettitt (1979). The Mann-Whitney nonparametric test is another option when the location of the step is known (Lettenmaier 1976). The binomial test, see below, would be even simpler.

Statistical tests for trend are affected by the statistical approach adopted. There are two accepted ways of thinking about time-series:

- (i) *design-based*: the signal is assumed to be unique and fixed over any defined interval of time; the results of a survey depend on its design.
- (ii) *model-based*: the signal is assumed to be one of many possible realisations over that interval; results of a survey depend on the model fitted to the data.

Under design-based thinking, the null hypothesis of no trend, meaning exactly equal values of the signal at all observation points, would seldom be plausible for fisheries data unless the locations of observation were extremely close, or all possible causes of variation temporarily ceased to exist. The analogy of the rocky mountain range is again applicable – no two observation points along a transect are likely to be at exactly the same height. Provided that there are enough observations, and measurements are made accurately enough, statistically significant differences in value will be discovered even though these may not be significant in practical terms (Loftis *et al.* 1991b). Model-based thinking comes from the other conceptual direction by assuming that a signal should be assumed to be horizontal until evidence indicates otherwise. The analogy here is of a randomised experiment in which subjects from one defined population are assigned randomly to treatments so that, if the treatments have no effect, the null hypothesis of equal means in each experimental group is readily plausible. This brings in the concept of the statistical power of a test for trend (Lettenmaier 1976; Nicholson and Fryer 1992).

The implications of serial correlation are also affected by whether the approach is design- or model-based. In the first case, the distinction between serial correlation and trend is undefined. Serially correlated values can look like a trend when observed through a narrowed time window, and, vice versa, a trend can look like serial correlation when observed through a widened window. Either situation could cause rejection of the “no trend” hypothesis. Under model-based inference, serial correlation invalidates nonparametric tests that are based on the null assumption that all permutations of values around a horizontal signal are equally likely. Serial correlation can be decreased by increasing the time intervals

between observations, or by modelling the serial correlation and subtracting the estimates from the series (Lettenmaier 1976). Seasonal or other cyclical trends add to the complications since the seasonal trends may themselves move independently over years (Van Bell and Hughes 1984), and measurement errors could be serially correlated from one season to the next (Zetterqvist 1991; El-Shaarawi and Niculescu 1992). Several nonparametric methods for seasonal trends are available (Hirsch *et al.* 1982; El-Shaarawi 1993; Esterby 1993; Yu *et al.* 1993) but they are not considered in detail here since most fish survey data are annual.

Monotonic trends might appear *linear* or *curvilinear*. These are easily modelled, of course, but, with the design-based approach, any structural model of the pattern could only be postulated as a rough approximation to the signal from a natural system.

The design-based approach is preferred here for its plausibility, because avoidance of modelling is consistent with the simplicity underlying nonparametric statistics, and because it relieves the analyst of many assumptions associated with model identification and fitting, thereby offering a genuine alternative to modelling. To be consistent with the design-based approach, I suggest replacing the term “hypothesis” with the word “notion” when describing the signal as having no trend or a specific type of trend so as to be clearer about the informality of a test in these circumstances. Trends can be estimated together with confidence limits without testing the usually untenable notion of ‘no trend’. Alternatively, the more reasonable null hypothesis (H) : “trend ≤ 0 ” can be tested with nonparametric methods against the alternative (A) : “trend > 0 ”. This one-sided H encompasses a region of probability, not a point. [Technically, it is a ‘composite hypothesis’ (Brownlee 1965).] If true, it would not be rejected by a sample, however large or precise, except by chance with probability α , as expected for a statistical test. Usually, this null hypothesis would be the most sensible choice for a test because a one-sided test is consistent with a prior concern that the trend is in one direction. Applying a two-sided test for either a positive or negative trend could suggest that the data are being mined unscientifically for any detectable feature.

A special problem with assessing trends is that they are often noticed in graphical plots before they are tested statistically or confidence limits are fitted. Bearing in mind that trends are often visible in series of random numbers (Kendall 1976), the application of statistical methods *a posteriori* could be misleading. Equally risky is when the terminal points of a trend are decided by inspection. Questions of the type “Is this variable going up or down?”, e.g. for the purposes of controlling environmental quality, should be completed with “since when” before assessing statistically because the probability of a trend is likely to depend on the chosen starting point, as well as the end point if not the final observation. If the interest lies in cause and effect, the recommended plan is to decide by prior reasoning when a trend might occur and in which direction, then to apply statistical methods to test whether the trend is present. If it is, linking it with a putative cause in a matching time-frame might be reasonable as a cautious, on-going hypothesis. Loftis *et al* (1991b) point out that trend analysis cannot establish cause and effect relationships.

Nonparametric statistical methods for trends

1. Example data

Table 5.13.1 shows abundance (N per hour) indices for North Sea cod as found by the ICES International Bottom Trawl quarter 1 surveys from 1976 to 2004 at ages 1 to 6, except that results for ages 3 to 6 were missing in the earlier years. These values were taken from a

report of the ICES working group on fish stocks of the North Sea and Skagerrak. They will be used to illustrate application of various nonparametric statistical methods. Some use the full time-series; others have to use only 1983 to 2004 when all age classes were determined.

Table 5.13.1. International bottom trawl survey (IBTS) quarter 1: time series of abundance indices (numbers caught per hour) for cod in the North Sea in 6 age classes. -1 = missing value.

Year	Age1	Age2	Age3	Age4	Age5	Age6
1976	7.9	19.9	-1	-1	-1	-1
1977	36.7	3.2	-1	-1	-1	-1
1978	12.9	29.3	-1	-1	-1	-1
1979	9.9	9.3	-1	-1	-1	-1
1980	16.9	14.8	-1	-1	-1	-1
1981	2.9	25.5	-1	-1	-1	-1
1982	9.2	6.7	-1	-1	-1	-1
1983	3.9	16.6	2.7	1.8	0.8	1.5
1984	15.2	8	3.9	0.9	1	0.9
1985	0.9	17.6	3.5	1.7	0.5	1
1986	17	3.6	6.8	2.3	1.3	1.1
1987	8.8	28.8	1.4	1.7	0.6	0.9
1988	3.6	6.1	5.8	0.6	0.9	1.1
1989	13.1	6.3	5	2.3	0.4	1
1990	3.4	15.2	2	1	1	0.8
1991	2.4	4.1	3.4	0.8	0.4	0.8
1992	13	4.5	1.2	1	0.3	0.5
1993	12.7	19.9	2	0.7	0.6	0.4
1994	14.8	4.4	3	0.8	0.5	0.5
1995	9.7	22.1	2.8	1.1	0.3	0.3
1996	3.5	8	6	0.7	0.6	0.4
1997	40	6.9	2.3	1.1	0.4	0.4
1998	2.7	26.4	2	0.9	0.5	0.4
1999	2.1	1.6	8.1	0.8	0.5	0.5
2000	6.6	3.8	0.7	2	0.4	0.5
2001	2.8	8.7	1.7	0.2	0.4	0.3
2002	7.8	3.4	4.3	0.5	0.1	0.2
2003	0.6	3	1	1.4	0.4	0.3
2004	7.5	1.3	1.2	0.30	0.4	0.01
Median	7.9	8.0	2.75	0.95	0.5	0.5

2. Quantiles and binomial methods

A time-series may be characterised most basically by its median value +/- binomial confidence limits. The latter are found by firstly ranking the observed values, then finding the ranks, conventionally shown in brackets as (a) and (b) , with cumulative binomial probabilities nearest to the required confidence limits, e.g. 2.5% and 97.5% for the case of limits of approximately 95%. Limits exactly at some preset, rounded percentage are seldom possible with the binomial distribution. Binomial confidence limits for the median, $\tilde{\mu}$, of a variable X are obtained with

$$\Pr\{X_{(a)} \leq \tilde{\mu} < X_{(b)}\} = \sum_{i=0}^{b-1} B\{N, 0.5, i\} - \sum_{j=0}^{a-1} B\{N, 0.5, j\}$$

where $B\{\}$ is the binomial probability function for sample size N and probability of 'success'=0.5 (for the median). See Conover (1971) for more details. The binomial

probabilities are based on an assumption that the observations fall independently and randomly to either side of the median for the tested time period, a questionable assumption if a trend is present. [They can be computed in an MicroSoft Excel spreadsheet with the BINOMDIST function.] Applying these formulae to an example subset of the abundance indices shown in table 5.13.1, the median index for age 1 cod from 1976 to 2004 was 7.9 with 93.86% confidence limits of 3.5 and 12.7 if there was no trend. These correspond to ranks $a = 9$ and $b = 20$ with cumulative binomial probabilities of 0.0307 and 0.969 respectively. Binomial confidence limits can also be estimated for more than one percentile simultaneously, e.g. the 10, 50, and 90 percentiles (Cotter 1985).

The binomial distribution can be used to test H : “trend ≥ 0 ” against A : “trend < 0 ” by assuming only that the estimated median is close to the true median for the whole tested period. This is a very simple test to carry out but it would often miss trends that would be detected by more elaborate methods. Four quadrants are formed by intersection of the estimated median observed value with the median of the observation times, the latter being the vertical line half way through the observed series. The test could also be applied to look for a step change; the vertical line would then be located at the time when the step change is expected. Each observed value and its associated time of observation is then classified by quadrant. The null hypothesis, H , implies that observations will fall equally into each quadrant or that there will be more in the lower left and top right quadrants. Suppose that x out of N observations fall in either the top left or bottom right quadrants, implying A , a downward trend. The probability that H is true is

$$\Pr\{H = \text{true}\} \leq 1 - \sum_{i=0}^x B\{N, 0.5, i\}.$$

[The ' \leq ' would be '=' if H only represented independence of the four quadrants.] Calculations for all ages are shown in table 5.13.2. With the probability of rejecting H given that it is true being $\alpha=0.05$, we would accept downward trends in abundance for cod of ages 2, 5, and 6. Note however that this is a set of univariate tests, so use of a lower value of α might be preferred to allow for the increased possibilities of type 1 errors in multiple tests. A simple, if conservative, way to achieve this is with the Bonferroni inequality (Prins 2006); when conducting $m=1, \dots, k$ tests, set $\alpha_m = \alpha/k$. In this case, $\alpha_m = 0.0083$ implying that downward trends should only be accepted for the 5 and 6 year-olds. Binomial tests can also be applied to assess compliance with an ecological quality objective set as a quantile other than the median though larger sample sizes tend to be necessary to find statistical significance (Cotter 1985). Compliance testing with multiple objectives set as quantiles is further discussed by Cotter (1994).

Table 5.13.2. Binomial test for a trend relative to the median value for the abundance indices shown in table 5.13.1. x = observed value, t = observation time; 0 indicates index \leq median ;1 indicates index $>$ median.

Year	Age1	Age2	Age3	Age4	Age5	Age6
1976	0	1				
1977	1	0				
1978	1	1				
1979	1	1				
1980	1	1				
1981	0	1				
1982	1	0				
1983	0	1	0	1	1	1
1984	1	0	1	0	1	1
1985	0	1	1	1	0	1
1986	1	0	1	1	1	1
1987	1	1	0	1	1	1
1988	0	0	1	0	1	1
1989	1	0	1	1	0	1
1990 = median (1976-2004)	0	1	0	1	1	1
1991	0	0	1	0	0	1
1992	1	0	0	1	0	0
1993 = median (1983-2004)	1	1	0	0	1	0
1994	1	0	1	0	0	0
1995	1	1	1	1	0	0
1996	0	0	1	0	1	0
1997	1	0	0	1	0	0
1998	0	1	0	0	0	0
1999	0	0	1	0	0	0
2000	0	0	0	1	0	0
2001	0	1	0	0	0	0
2002	0	0	1	0	0	0
2003	0	0	0	1	0	0
2004	0	0	0	0	0	0
Number of $x >$ median and $t \leq \text{med}(\text{year})$	9	9	6	7	7	9
Number of $x <$ median and $t > \text{med}(\text{year})$	9	10	6	7	10	11
Number of x	29	29	22	22	22	22
Binomial probability (H : trend ≥ 0)	0.068	0.031	0.262	0.067	0.002	<0.001

3. Cochran's Q test

Cochran's Q (Cochran 1950) tests the notion that the probabilities of response are the same (H) in different groups or, alternatively (A) detectably different. The test can be adapted to look at multiple trends with the aim of avoiding the problem of multiple univariate tests mentioned in connection with binomial tests, above. This is illustrated in table 5.13.3 using the abundance indices for 1 to 6 year-olds from 1983 to 2004. Age classes of cod are treated as groups, and each observed value is marked as a response, i.e. with a 1, if above the median value and located on or before the median time, or if below the median value and after the median time. Otherwise it is marked as a non-response, i.e. with a 0. In other words, each value gets 1 if it is consistent with a downward trend, and 0 otherwise. The markings are shown in table 5.13.3. Let T_j be the column sums in the j 'th age class, $j = 1, \dots, C$, and \bar{T} the mean of them. Let u_i be the i 'th row sum, $i = 1, \dots, R$. Then Cochran's statistic is defined as

$$Q = \frac{C(C-1) \sum_j (T_j - \bar{T})^2}{C \sum_i u_i - \sum_i u_i^2}.$$

Q is distributed as χ^2 with $(C-1)$ degrees of freedom (Brownlee 1965, section 7.10) under H . Note that Q is not sensitive to the total number of responses (since many of the u_i may be zero) hence, for our purposes, it does not by itself establish whether or not an overall trend is present. For the example, $Q = 3.56$ which is much less than $\chi^2(5) = 11.07$ implying that the different age classes are not showing detectably different trends, given that a general downward trend exists. Cochran's Q appears to have similarities with Friedman's rank test for blocked data (Brownlee 1965), and with van Belle and Hughes' test for homogeneity of seasonal trend (Van Bell and Hughes 1984). El-Shaarawi (1993) suggests ways of extending the latter method to testing the notions that linear or quadratic patterns exist in the seasonal trends.

Table 5.13.3. Cochran's Q test applied to age classes 1 to 6 from 1983 to 2004 for the abundance indices shown in table 5.13.1 illustrating how observed abundance indices are marked. 0 means index \leq median and 1 means index $>$ median if year \leq 1993, and vice versa if year $>$ 1993.

Year	Age1	Age2	Age3	Age4	Age5	Age6
1983	0	1	0	1	1	1
1984	1	0	1	0	1	1
1985	0	1	1	1	0	1
1986	1	0	1	1	1	1
1987	1	1	0	1	1	1
1988	0	0	1	0	1	1
1989	1	0	1	1	0	1
1990	0	1	0	1	1	1
1991	0	0	1	0	0	1
1992	1	0	0	1	0	0
1993	1	1	0	0	1	0
1994	0	1	0	1	1	1
1995	0	0	0	0	1	1
1996	1	1	0	1	0	1
1997	0	1	1	0	1	1
1998	1	0	1	1	1	1
1999	1	1	0	1	1	1
2000	1	1	1	0	1	1
2001	1	0	1	1	1	1
2002	1	1	0	1	1	1
2003	1	1	1	0	1	1
2004	1	1	1	1	1	1

4. Runs test

A 'run' is defined as any sequence of 1 or more like elements from two classes. In the present context, this could mean above or below a level line, or a notional trend line. The runs test examines the notion of randomness in a series by looking at the number of runs of observed values above and below the median and comparing with the expected number which, along with variance, can be computed from theory. Non-randomness is usually represented by positive serial correlation of the observations, i.e. fewer than the expected number of runs, hence the test is usually one-sided. Serial correlation may be of interest in itself, e.g. as an interfering factor in a model-based test of trend (Loftis *et al.* 1991b), but could also arise from non-monotonic trends in the underlying signal.

Sources on the runs test are texts by Brownlee (1965, section 6.3) and Conover (1971, p. 349). Let the two classes of elements be a or b for 'above' or 'below' the sample median. Values equal to the median are ignored. Let the number of a 's be m . Then the number of b 's turns out also to be m , assuming no tied values. The expected number, u , of runs is

$$E(u) = 1 + m$$

and the variance is

$$V(u) = \frac{m(m-1)}{2m-1}.$$

Exact probabilities of runs are available (Swed and Eisenhart 1943) but for series of reasonable length (?), it is easier, and justifiable under the Central Limit Theorem, to assume that the statistic

$$\frac{u + 0.5 - (1 + m)}{\sqrt{m(m-1)/(2m-1)}}$$

is approximately a standard normal variate. The 0.5 in the numerator is a correction for continuity. For a one-sided test, compare the test statistic with the standard normal deviate having cumulative probability of α . If the probability of the observed number of runs is less, serial correlation is detected.

The runs test is illustrated using the abundance indices for 1 to 2 year-olds from 1976 to 2004, and for 3 to 6 year-olds from 1983 to 2004 in table 5.13.4. Scoring of the runs above and below the median is shown in table 5.13.4. It is similar to the markings in table 5.13.2 except that values equal to the median must also be marked and ignored when counting the runs. Some tied values prevented n and m from being equal in each of age classes 2 and 5; m and n were adjusted to the minimum of the pair. The probabilities that the series were random, shown at the bottom of table 5.13.4, indicates that only the 6 year-olds were non-random by this test. The runs test is noticeably less sensitive to pure trend than the binomial test in relation to the median (section 2 above) because the continuity of runs above and below the median is frequently broken by variant observations.

Table 5.13.4. Runs test applied to age classes 1 to 6 from 1983 to 2004 for the abundance indices shown in table 5.13.1 illustrating how runs are obtained. 0 means index \leq median and 1 means index $>$ median; M means = median (observation ignored).

Year	Age1	Age2	Age3	Age4	Age5	Age6
1976	M	1				
1977	1	0				
1978	1	1				
1979	1	1				
1980	1	1				
1981	0	1				
1982	1	0				
1983	0	1	0	1	1	1
1984	1	M	1	0	1	1
1985	0	1	1	1	M	1
1986	1	0	1	1	1	1
1987	1	1	0	1	1	1
1988	0	0	1	0	1	1
1989	1	0	1	1	0	1
1990 = Median (1976-2004)	0	1	0	1	1	1
1991	0	0	1	0	0	1
1992	1	0	0	1	0	M
1993 = Median (1983-2004)	1	1	0	0	1	0
1994	1	0	1	0	M	M
1995	1	1	1	1	0	0
1996	0	M	1	0	1	0
1997	1	0	0	1	0	0
1998	0	1	0	0	M	0
1999	0	0	1	0	M	M
2000	0	0	0	1	0	M
2001	0	1	0	0	0	0
2002	0	0	1	0	0	0
2003	0	0	0	1	0	0
2004	0	0	0	0	0	0
N runs=>	14	18	13	16	8	2
N>median: <i>m</i>	14	13	11	11	8	9
N<=median: <i>n</i>	14	14	11	11	10	9
Adjusted <i>m,n</i>	14	13	11	11	8	9
E(runs)=	15	14	12	12	9	10
V(runs)=	6.74	6.24	5.24	5.24	3.73	4.24
Normal d.f.	-0.19	1.80	0.66	1.97	-0.26	-3.64
Probability that series is random	0.42	0.96	0.74	0.98	0.40	<0.01

5. Mann-Kendall's K

Kendall's tau is used to measure concordance or correlation (Hollander and Wolfe 1973). Slightly adapted (Mann 1945), Kendall's method is considered to be very sensitive to monotonic trends (Esterby 1993). Every observed value is paired with every value observed after it and the pair scored 1 if the first is greater than the second, 0 if the same, and -1 if the first is less than the second. The Mann-Kendall test statistic, K , is the sum of these values². The null hypothesis is H : "trend ≥ 0 " against A : "trend is monotonic, negative", or vice versa. Note that, if a turning point is definitely expected at a certain time, the observations could be re-ordered in time to conform with monotonicity under A .

K takes on large positive or negative values when a monotonic trend is present. One-tail probabilities of observing K under H are tabulated by Hollander and Wolfe (1973, Appendix A.21) but, for large samples of size n ,

$$K^* = K / \sqrt{[n(n-1)(2n+5)/18]}$$

is distributed as a standard normal variate if no data are tied. If there are ties, the square root denominator representing the standard error of K has to be inflated (Hollander and Wolfe 1973, p. 187). According to a citation in Yu et al. (1993), 10 observations are adequate for "large sample".

Scoring of observations for Kendall's K is illustrated in table 5.13.5 for age 1 abundance indices for cod from 1976 to 2004. The full table has 29 columns, so only 3 years of scores are shown. The sum of all the scores from 1976 to 2004 (K) was -106, and $n=29$, giving a large-sample standard normal approximation of -1.988. The corresponding probability of no trend is 0.02, implying here that a monotonic, downward trend was present in the signal. Note that the test is more sensitive to trend than the median test for trend where the probability of no trend was found to be 0.068 (table 5.13.2). On the other hand, the Mann-Kendall test is much more work to carry out on a spreadsheet. Kendall's test is available in R [`cor.test(. . . method="kendall". . .)`]. The observed values are correlated with times of observation, or their ranks, to achieve the Mann-Kendall test.

² Kendall's $\tau = 2K/n(n-1)$ is that used in correlation studies.

Table 5.13.5. Kendall's Tau test applied to age class 1 from 1976 to 2004 for the abundance indices shown in table 5.13.1 illustrating how the sequence of observed indices is scored for 1976 with subsequent years (column 1), and similarly for 1977 and 1978 (columns 2 and 3). Other years to 2004 not shown. x denotes observed value, k and i are years.

Year	sign($x(k) - x(i)$), $i < k$		
	1976	1977	1978
1977	1		
1978	1	-1	
1979	1	-1	-1
1980	1	-1	1
1981	-1	-1	-1
1982	1	-1	-1
1983	-1	-1	-1
1984	1	-1	1
1985	-1	-1	-1
1986	1	-1	1
1987	1	-1	-1
1988	-1	-1	-1
1989	1	-1	1
1990	-1	-1	-1
1991	-1	-1	-1
1992	1	-1	1
1993	1	-1	-1
1994	1	-1	1
1995	1	-1	-1
1996	-1	-1	-1
1997	1	1	1
1998	-1	-1	-1
1999	-1	-1	-1
2000	-1	-1	-1
2001	-1	-1	-1
2002	-1	-1	-1
2003	-1	-1	-1
2004	-1	-1	-1

6. Thiel's or Sen's slope estimator

Thiel's slope estimator is used for a notional linear trend:

$$Y_i = \alpha + \beta x_i + e_i, \quad i = 1, \dots, n$$

The e 's must be mutually independent and from the same continuous population (Hollander and Wolfe 1973). Thiel's estimator for β is similar in construction to that for Mann-Kendall's K . Every observed value is paired with every value observed after it, and the slope, $S_{ij} = (Y_j - Y_i)/(x_j - x_i)$, $i < j$, calculated. Thiel's estimator is the median of these values. Sen's estimator, as described by Yu et al. (1993), appears to be exactly the same. Small sample confidence limits are available using Hollander and Wolfe (1973, chapter 9 and table A.21). For large samples, use the rounded integer value of

$$C_{\alpha} \approx z_{\alpha/2} \left\{ \frac{n(n-1)(2n+5)}{18} \right\}^{0.5}.$$

The $1-\alpha$ confidence interval is obtained from the ranked slope values. Use $\{S_{(L)}, S_{(U)}\}$ where rank $L = (N - C_{\alpha})/2$ and rank $U = (N + C_{\alpha})/2$.

As an illustration, Thiel's slope estimator and 95% confidence limits were calculated for the age 1 abundance indices for cod from 1976 to 2004. The median slope, -0.23, is drawn through the intersection of the median value of Age 1 indices, 7.9 fish per hour, and the median observation time, 1990, in fig. 5.13.1 below. I am not aware of a method for estimating confidence limits for Y that takes into account the covariance of estimated α and β .

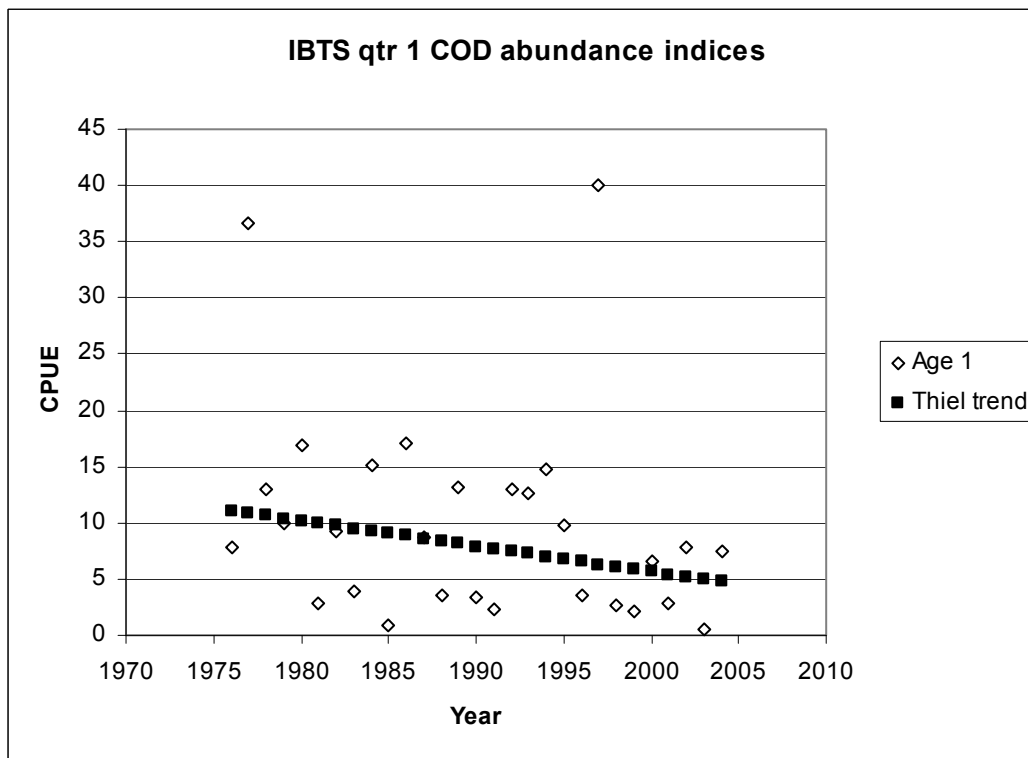


Figure 5.13.1 Thiel's slope estimator for the age 1 abundance indices, 1976 to 2004, shown in table 5.13.1.

7. Spearman's rho

Spearman's rho is the product-moment correlation between the ranks of paired data, the ranking being carried out separately for each variable of the pair. To test for trend, one member of the pair is the time of observation, the other is the observed variable. In practice, the arithmetic needed to calculate rho can be avoided by simply using

$$T = \sum_{i=1}^n (R(X_i) - i)^2$$

where i indexes the observation times and $R(X_i)$ is the rank of the corresponding observation (Lettenmaier 1976). This is also known as the Hotelling-Pabst test. T is small when $R(X_i)$ and i are positively correlated, and large when negatively correlated. Conover (1971, p389) gives quantiles for T for series up to 30 observations. Alternatively, use

$$w_p \approx \frac{1}{6} \left(n(n^2 - 1) + \frac{x_p(n(n^2 - 1))}{6\sqrt{n-1}} \right)$$

where x_p is the p 'th quantile of a standard normal deviate. T should be less than w_p (and x_p on the negative side of the normal distribution) for an upward trend, i.e. positive correlation with time, and T should be greater than w_p (and x_p on the positive side of the normal distribution) for a downward trend, i.e. negative correlation (Conover 1971).

Calculation of Spearman's Rho is illustrated in table 5.13.6 for age 1 abundance indices for cod from 1976 to 2004. $T = 5586$ which is greater than $w_p = 5564$ with $x_{0.975} = 1.96$, implying that the downward trend for 1-year olds is significant at $\alpha = 0.025$. This is not as significant as was found with Mann-Kendall's K which gave $\alpha = 0.018$ but it is more significant than was found with the median test, $\alpha = 0.07$. The sequence of probabilities is roughly matched inversely by the work required to carry out the tests on a spreadsheet. Conover states that the normal approximation is better for Kendall's tau than for Spearman's rho with small sample sizes.

Table 5.13.6. Spearman's Rho test applied to age class 1 from 1976 to 2004 for the abundance indices shown in table 5.13.1 illustrating the ranking (R) of years (y) and observed indices (a), and the computation of T .

Year	Age 1 index	Year rank	Data rank	$(R(y)-R(a))^2$
1976	7.9	1	15	196
1977	36.7	2	28	676
1978	12.9	3	21	324
1979	9.9	4	19	225
1980	16.9	5	26	441
1981	2.9	6	7	1
1982	9.2	7	17	100
1983	3.9	8	11	9
1984	15.2	9	25	256
1985	0.9	10	2	64
1986	17	11	27	256
1987	8.8	12	16	16
1988	3.6	13	10	9
1989	13.1	14	23	81
1990	3.4	15	8	49
1991	2.4	16	4	144
1992	13	17	22	25
1993	12.7	18	20	4
1994	14.8	19	24	25
1995	9.7	20	18	4
1996	3.5	21	9	144
1997	40	22	29	49
1998	2.7	23	5	324
1999	2.1	24	3	441
2000	6.6	25	12	169
2001	2.8	26	6	400
2002	7.8	27	14	169
2003	0.6	28	1	729
2004	7.537	29	13	256
			Sum, T	5586

8. Jonckheere's test

Jonckheere's test is a nonparametric version of a one-way analysis of variance with unequal sample sizes, except that it tests H : 'no treatment effect' versus the special alternative, A : 'the treatments are ordered in effect' (Hollander and Wolfe 1973, p. 120). This can be applied to trends by equating observation times to treatments and then arranging them in the order implied by the possible trend so as to make a one-sided test. Hypothesis A is then equivalent to a monotonic trend, as for Mann-Kendall's K . For a time-series of un-replicated observations, Mann-Kendall's K is the conventional test to use. However, if observations are independently replicated at some time points, the Mann-Kendall test is not clearly applicable and Jonckheere's test can be applied instead.

Let the number of observation times be k . It is necessary first to find $k(k-1)/2$ Mann-Whitney counts U_{uv} where u and v are times of observation and $1 \leq u < v \leq k$:

$$U_{uv} = \sum_{i=1}^{n_u} \sum_{j=1}^{n_v} \phi(X_{iu}, X_{jv}).$$

n_u, n_v are the numbers of observations at times u and v respectively, while $\phi(a, b) = 1$ if $a < b$, and 0 otherwise. Be warned, this is an extensive job for large k using a spreadsheet. The test statistic, $J = \sum_{u < v}^k U_{uv}$, can be compared with table A.8 in Hollander and Wolfe (1973), or, when the minimum number of replicate observations at any point is large (?), can be transformed to J^* which is approximately normally distributed:

$$J^* = \frac{J - \left\{ \left(N^2 - \sum_{j=1}^k n_j^2 \right) / 4 \right\}}{\sqrt{\left(N^2(2N + 3) - \sum_{j=1}^k n_j^2(2n_j + 3) \right) / 72}}.$$

$N = \sum_{j=1}^k n_j$. As for Kendall's test, this would be one-tailed. H is rejected if J^* exceeds the standard normal deviate at the required level of significance.

Table 5.13.7 presents a fabricated set of data to illustrate computations for Jonckheere's test. Abundance indices for 1-year olds were arbitrarily re-assigned to a shortened series of 10 years so as to create a set of replicate observations. J was found to be 209 which together with the totals in the right 3 columns of table 5.13.7a gave $J^* = 0.8318$ which is less than 1.645, the standard normal deviate corresponding to 95% of the area under the normal curve. Not surprisingly, the arbitrarily re-arranged data did not show a significant monotonic trend.

Table 5.13.7. Jonckheere's test applied to abundance indices for cod of age class 1 from 1976 to 2004 taken from table 5.13.1 and arbitrarily assigned to a shortened series of 10 years with variable replication annually for illustrative purposes. a) resulting test data and computation of terms for the test statistic; b) illustrating scoring for $u = 1976$ and $v = 1977, 1978$.

a)

Year	Dummy replicate observations					nj	nj^2	nj^2*(2nj+3)
1976	7.9	17	3.5			3	9	81
1977	36.7	8.8				2	4	28
1978	12.9	3.6	2.7			3	9	81
1979	9.9	13.1	2.1	40	7.8	5	25	325
1980	16.9	3.4	6.6			3	9	81
1981	2.9	2.4				2	4	28
1982	9.2	13				2	4	28
1983	3.9	12.7	0.6	2.8		4	16	176
1984	15.2	14.8	7.537			3	9	81
1985	0.9	9.7				2	4	28
					Sum=>	29	93	937

b)

$u = 1976$	Replicate 1	Replicate 2	Replicate 3
$v = 1977$			
Replicate 1	1	1	1
Replicate 2	1	0	1
$v = 1978$			
Replicate 1	1	0	1
Replicate 2	0	0	1
Replicate 3	0	0	0

9. Permutation (randomisation) and bootstrapping methods

Permutation tests are described in connection with benthic studies by Bell et al. (1981). They are alleged to be more sensitive to trend than rank tests. Let Y be the observed value, and t the time of observation, $t = 1, \dots, T$. Calculate

$$h = \sum_t tY_t$$

and compare with the $T!$ values of h^* computed with the t values permuted. Bell et al. state that if there is no change in Y over time, h is likely to fall near the middle of the range of h^* , otherwise near one of the extremes. Probabilities of h or values more extreme can be found because each permutation is equally likely. Computation of all permutations may be an onerous and unnecessary task. The `sample()` function in the R programming language readily produces random permutations that may suffice for building up a reference set for assessing probabilities. The case for a randomisation test is likely to be strong when the

number of observations available is very small, e.g. less than 10. Many other nonparametric methods then need access to special tables to estimate significance because the Central Limit Theorem can not be invoked satisfactorily to argue that test statistics are approximately normally distributed, as for Spearman's rho and Mann-Kendall statistics for example. Ties can further complicate application of competing nonparametric methods.

Edgington (1995, p. 15) points out that "Parametric tests of all kinds, including relatively complex tests, . . . become distribution-free when the significance is determined by a randomization test procedure." Consider for example a least squares estimator of slope. Rather than assume that the observations at each time point were a random sample from some population and looking up the value of t or F for the slope based on that assumption, the observations and times are permuted to create a reference set of estimated t or F statistics, then the observed value is compared with the reference set and its statistical significance judged from the proportion of the reference set having a more extreme value. Edgington describes several randomisation tests applicable to trends over treatment levels in randomised experiments but states (p. 217) ". . . randomization trend tests do *not* test hypotheses about trends; they simply utilize test statistics sensitive to trend which test the null hypothesis of no differential treatment effect." His view is possibly related to the focus on experiments in his book. On the other hand, finding that an observed trend over time is unlikely by a randomisation test, even outside a randomised experiment, seems to be no less useful than finding it by means of other nonparametric tests.

Bootstrapping observations could be another way of assessing the significance of observed trends, particularly for fisheries survey data for which abundance indices can be bootstrapped to estimate sampling and measurement errors (Beare *et al.* 2002) even though analytical formulae for variances are hard to derive or non-existent. The trend would be fitted to each bootstrapped series and the distribution determined. To be nonparametric, the estimates should be based on sampling theory and no model assumed. The bootstrap is no less vulnerable than most other statistical methods to small numbers of observation times.

There are subtle differences between bootstrapping and randomisation tests. Bootstrapping uses a sample as a surrogate for the population and (re-)samples the sample *with* replacement. A randomisation test of trend permutes observations to time points *without* replacement and without reference to the sampling process generating the observations. Bootstrapping empirically estimates the distribution of statistics assuming that the observed sample looks like the true population. On the other hand, randomisation tests provide a form of statistical inference when the sample itself is assumed to be the total population of interest. When testing for trend in an environmental context, the quality of the sample appears to be just as important as for the bootstrap.

10. Dietz and Killeen test for multivariate monotone trend (turquoise box)

Dietz and Killeen (1981) derived a formula for the covariance matrix of Mann-Kendall statistics estimated from a multivariate monitoring programme, and proposed a test statistic based on it that is asymptotically distributed as χ^2 . Use of multivariate methods "controls the overall significance level" when multiple univariate tests are made with covarying observations, and there were very few cases in a study of lake water quality "where the univariate methods perform better" (Loftis *et al.* 1991b). Multivariate methods appear to be especially applicable when monitoring groups of indicators expected to respond similarly to environmental changes of concern. The Dietz and Killeen test has the same strengths as the Mann-Kendall test for discovering trends that are specifically monotonic. An alternative but

related test when the aggregate trend of several variables is of interest, e.g. when correlated measures of one variable are made seasonally, was given by Hirsch and Slack (Hirsch and Slack 1984). Comparisons of nonparametric methods with linear models in a multivariate context were made using simulation by Loftis *et al.* (Loftis *et al.* 1991a).

Code in R (`Mvar.trend.r`) for the Dietz and Killeen test is available from <http://www.ifremer.fr/drvecohal/fisboat/index.htm>; it was trialled with data given by Dietz and Killeen and found to give the same results. Fig. 5.13.2 shows selected output from `Mvar.trend.r` applied to the cod abundance indices, ages 1 to 6, 1983 to 2004. (About 2 minutes were required for running; times increase roughly in proportion to the product of $n^3 p$.) The first matrix shows the Mann-Kendall statistics. Those along the diagonal are computed as for the univariate statistics; see section 5 above. Those off the diagonal are

$$K_{XY} = \sum_{i < j} \text{sign}[(X_j - X_i)(Y_j - Y_i)]$$

where $\text{sign}(x) = 1, 0, -1$ for positive, zero, and negative values respectively. K_{XY} is high and positive when both the X and Y variables are showing monotonic trends in the same direction, and high and negative when in opposite directions. The second matrix, **S**, in box 10 shows the covariances of these statistics calculated using formulae given in the appendix of Dietz and Killeen. The third matrix shows the Spearman rank correlations of the observed values. The highest correlations occur among the older age groups. Note that although values along the diagonal of a correlation matrix are normally 1, some here are less than 1 due to tied values within the time-series. The test statistic, $\mathbf{K}^T \mathbf{S}^{-1} \mathbf{K} = 25.09$ is compared with χ^2 with 6 d.f. (all age classes contribute to d.f. since no parameters are estimated) and found to be highly significant, $P < 0.005$; $\chi^2 = 18.548$. The standardised K for each age class and their standard errors are shown below. This is the same information that would come from testing each age class separately. Examination of them when the multivariate null hypothesis is rejected should indicate which of the age classes was responsible (Loftis *et al.* 1991b). For 1-sided tests, standardised K greater in magnitude than 1.64 are likely to be contributing to the significance of the multivariate result. The results suggest that all except 1-year olds are contributing.

Figure 5.13.2. Output from Mvar.trend.r applied to cod abundance indices for ages 1 to 6 from 1983 to 2004 given in table 5.13.1. \$Kxy is Matrix of Kendall stats (K), age1 to 6; \$S is covariance matrix of K; \$Spearman.corr is the rank correlation matrix of the abundance indices; \$Standardised.K is K from diagonal of S (above), divided by \$St.error.K (last row). Some rows of output are omitted to save space.

```
>Mvar.trend(IBTS.indices)
$Kxy
      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
[1,]      0         0         0         0         0         0         0
[2,]      0        -45        -2         12         32         11        22
[3,]      0         -2        -72        -31         27         34         27
[4,]      0         12        -31        -60        -13         52         66
[5,]      0         32         27        -13        -74         16         91
[6,]      0         11         34         52         16        -97         95
[7,]      0         22         27         66         91         95       -166

$S
      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
[1,]      0         0         0         0         0         0         0
[2,]      0 1257.667 29.33333 108.6667 249.3333 61.66667      140
[3,]      0 29.33333 1256.667      -155 206.3333      309 232.3333
[4,]      0 108.6667      -155      1254 -90.3333      394 475.3333
[5,]      0 249.3333 206.3333 -90.3333      1248 106.6667      662
[6,]      0 61.66667      309      394 106.6667      1215 724.6667
[7,]      0      140 232.3333 475.3333      662 724.6667 1232.667

$Spearman.corr
      [,1]      [,2]      [,3]      [,4]      [,5]      [,6]      [,7]
[1,]      0         0         0         0         0         0         0
[2,]      0         1 0.025409 0.08865 0.202146 0.049125 0.112366
[3,]      0 0.025409 0.999435 -0.12253 0.167137 0.252117 0.189159
[4,]      0 0.08865 -0.12253 0.997741 -0.07284 0.319029 0.383964
[5,]      0 0.202146 0.167137 -0.07284 0.994353 0.085827 0.535009
[6,]      0 0.049125 0.252117 0.319029 0.085827 0.971203 0.586957
[7,]      0 0.112366 0.189159 0.383964 0.535008 0.586957 0.98419

$Standardised.K
[1]      -1.26891 -2.03106 -1.69435 -2.09471 -2.78281 -4.72809

$St.error.K
[1]      35.4636 35.44949 35.41186 35.32704 34.85685 35.10935
```

References

- Beare, D., Castro, J., Cotter, J., van Keeken, O., Kell, L., Laurec, A., *et al.* (2002) Evaluation of research surveys in relation to management advice (EVARES). Final report. *DGXIV Fisheries, European Commission, Brussels*. **FISH/2001/02 - Lot 1**. Available from a.j.cotter@cefas.co.uk
- Bell, C.B., Conquest, L.L., Pyke, R. and Smith, E.P. (1981) Some nonparametric statistics for monitoring water quality using benthic species counts. *Environmetrics, selected papers, SIAM, Philadelphia* **8**, 100-120.
- Brownlee, K.A. (1965) *Statistical theory and methodology in science and engineering*. 2nd. edn., John Wiley & Sons, New York
- Cochran, W.G. (1950) The comparison of percentages in matched samples. *Biometrika* **37**, 256-266.
- Conover, W.J. (1971) *Practical nonparametric statistics*., John Wiley & Sons, New York
- Cotter, A.J.R. (1985) Water quality surveys: a statistical method based on determinism, quantiles and the binomial distribution. *Water Research* **19**, 1179-1189.
- Cotter, A.J.R. (1994) Compliance testing of two or more water quality determinands using quantiles. *Environmetrics* **5**, 29-45.
- Dietz, E.J. and Killeen, T.J. (1981) A nonparametric multivariate test for monotone trend with pharmaceutical applications. *Journal of the American Statistical Association* **76**, 169-174.
- Edgington, E.S. (1995) *Randomization tests*. 3rd. edn., Marcel Dekker, Inc., New York
- El-Shaarawi, A.H. (1993) Environmental monitoring, assessment and prediction of change. *Environmetrics* **4**, 381-398.
- El-Shaarawi, A.H. and Niculescu, S.P. (1992) On Kendall's Tau as a test of trend in time series data. *Environmetrics* **3**, 385-411.
- Esterby, S.R. (1993) Trend analysis methods for environmental data. *Environmetrics* **4**, 459-481.
- Hirsch, R.M. and Slack, J.R. (1984) A nonparametric test for seasonal data with serial dependence. *Water Resources Research* **20**, 727-732.
- Hirsch, R.M., Slack, J.R. and Smith, R.A. (1982) Nonparametric tests for trend in water quality. *Water Resources Research* **18**, 107-121.
- Hollander, M. and Wolfe, D.A. (1973) *Nonparametric statistical methods*., John Wiley & Sons, New York
- Kendall, M.A. (1976) *Time-series*, Charles Griffin and Co. Ltd., London

- Lettenmaier, D.P. (1976) Detection of trends in water quality data from records with dependent observations. *Water Resources Research* **12**, 1037-1046.
- Loftis, J.C., Taylor, C.H. and Chapman, P.L. (1991a) Multivariate tests for trend in water quality. *Water Resources Research* **27**, 1419-1429.
- Loftis, J.C., Taylor, C.H., Newell, A.D. and Chapman, P.L. (1991b) Multivariate trend testing of lake water quality. *Water Resources Bulletin, American Water Resources Association* **27**, 461-473.
- Mann, H.B. (1945) Nonparametric tests against trend. *Econometrica* **13**, 245-259.
- Nicholson, M.D. and Fryer, R.J. (1992) The statistical power of monitoring programmes. *Marine Pollution Bulletin* **24**, 146-149.
- Pettitt, A.N. (1979) A non-parametric approach to the change-point problem. *Applied Statistics* **28**, 126-135.
- Prins, J. (2006). "NIST/SEMATECH e-Handbook of Statistical Methods." from <http://www.itl.nist.gov/div898/handbook/prc/section4/prc473.htm>.
- Swed, F.S. and Eisenhart, C. (1943) Tables for testing randomness of grouping in a sequence of alternatives. *Annals of mathematical statistics* **14**, 66-87.
- Van Bell, G. and Hughes, J.P. (1984) Nonparametric tests for trend in water quality. *Water Resources Research* **20**, 127-136.
- Yu, Y.-S., Zou, S. and Whittemore, D. (1993) Non-parametric trend analysis of water quality data of rivers in Kansas. *Journal of Hydrology* **150**, 61-80.
- Zetterqvist, L. (1991) Statistical estimation and interpretation of trends in water quality time series. *Water Resources Research* **27**, 1637-1648.

5.4 Construction of multivariate indicators.

5.4.1 Principal Components Analysis (PCA) and biological indicators

Pierre Petitgas,
IFREMER, Nantes,
France

Introduction

For each survey, a variety of indices of stock attributes are usually estimated. A multivariate monitoring procedure is then potentially 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 indices of stock attributes. Here we suggest application of PCA to the Fisboat biological indicators (abundance indices, length indices and mortality index) and represent the evolution of the stock by a multivariate distance to a reference gravity centre. Clearly, abundance and length indices are potentially related and more consistency can be obtained by explicitly using these correlations. But the method will not take into account any correlations in time between indicators, in particular lagged effects of one indicator on another.

Method

Consider an array where for each line (observation) we have a vector of measurements for a variety of parameters (variables). Measured values of variables are in columns and each line is one observation (here in time). The variables can be correlated between each others. PCA constructs linear combinations of variables (factors) that are non correlated between each other and that best account for the variability in the data array. Mathematically, this is done by diagonalising the correlation matrix of the variables. Eigen vectors geometrically support principal components. These are ranked by their decreasing importance of data variance explained. The geometrical properties of the method enables representation of the correlation structure among the variables as well as the position of each observation in the space of the principal components (factorial space). The correlation between two variables (variable - variable or variable - principal component) is represented by the angle between vectors. The similarity between observations is represented by their Euclidean distance in the factorial space. It is usual to analyse correlation between variables and similarity between observations in a factorial sub-space made by a reduced number of principal components. These are the first principal components that account for a large percentage of the data variance (e.g., 80%). Such reduction corresponds to filtering variability in the data assumed to be noise. The few retained non correlated factors then summarise the multivariate structure of the data. PCA (e.g., Lebart et al., 1995) is a widely used technique in many fields including marine ecology that was first developed more than fifty years ago.

Software

The R code available from the FISBOAT web site, <http://www.ifremer.fr/drvecolah/fisboat/index.htm>, is pcachart.R. It is commented. It uses the R library ade4 (Chessel et al., 2006) for performing the PCA. The PCA is applied to the Fisboat Table 2 of biological (non spatial) indicators. A set of reference years is defined, in which the population is considered in acceptable health status. The PCA is performed giving high weight to the set of reference year observations (99.9%). In that way, the factorial space

is not affected by years outside the reference period: years outside the reference period play the role of passive variables that are projected in a reference factorial space. First the decrease in the eigen values is illustrated. Then the correlation structure in the indicators is illustrated using the first three principal axes and the loadings of the indicators on these are provided in a table. This allows to interpret the principal axes. Then the similarity between years is illustrated by positioning the years in the plane made by the first two principal axes. On the figure, the reference period years are marked by a symbol while the non reference years to be monitored are labelled by their number. A multivariate distance in each year is then calculated, which quantifies the deviation of that year to the gravity centre of the reference years, and is saved with the name *mbio*. The multivariate distance is the euclidean distance between the position of any year $p(y)$ and the gravity centre of the reference years

$$c_{ref}: d_y = \sqrt{d^2(p_y; c_{ref})}.$$

Inputs are : Table 2 (from the FISBOAT web-site) of biological non spatial indices, years to consider as reference period, number of principal PCA axes to consider for computing the multivariate distance *mbio*. Outputs are: a figure representing the decrease in eigen values, figures illustrating the correlation structure between variables, correlation table of the variables on the PCA axes (loadings), the time series of the *mbio* multivariate distance as well as its histogram.

Example

North Sea cod

PCA can be used to set up a multivariate monitoring approach of stock status using the many indicators of biological stock attributes available for North Sea cod, see tables 1 and 2 at <http://www.ifremer.fr/drvecohal/fisboat/>. Because the indicators are correlated to each other, PCA is useful to summarise the correlation structure between the parameters and reduce the dimensionality of the monitoring scheme using a small number of non correlated factors. The monitoring approach will then take place in the factorial space composed of the first two (or more) principal axes. A reference domain in that factorial space (in-control domain) can be defined based on the position of reference years in that factorial space. The definition of reference years is analogous to monitoring a process when it is in-control: the reference year period is the set of years where the stock could be considered in acceptable health. The multivariate monitoring approach compares the current year vector of stock indicators to that of the reference period. It is therefore suggested to estimate the gravity centre of the reference years and, for each year, to calculate the distance to that reference gravity centre. The time series of the multivariate distance then summarises the deviation of the population biological characteristics from its reference status.

Fisboat table 2 of biological non spatial indicators comprises the year observations as lines and the columns as variables. It is a typical set for input to a PCA. The PCA will display the correlation structure between indicators (abundance, length, mortality) and will allow quantification of which years depart from the others, not just because of one indicator but as a whole in their multivariate characteristics.

Reference years are 1985-1994. Figure 5.10.1 illustrates the correlation structure in the biological indicators. Length50 at maturity heavily determines the first principal axis. The second is determined by total abundance and the opposition between total abundance and length at the third quartile (L75). The third axis is determined by mortality (Z) which is

hardly correlated with any other indicator. Overall, length indices and abundance indices are correlated. Figure 5.10.2 illustrates the multivariate monitoring approach. Years 1997, 1999, 2003-2005 are well outside the domain defined by the reference years, meaning that they depart largely from the reference status. The direction of departure is on the first and second axes meaning that departure is primarily guided by changes in abundance and length at maturity. This is quantified by the multivariate distance on which a statistical monitoring scheme can then be applied.

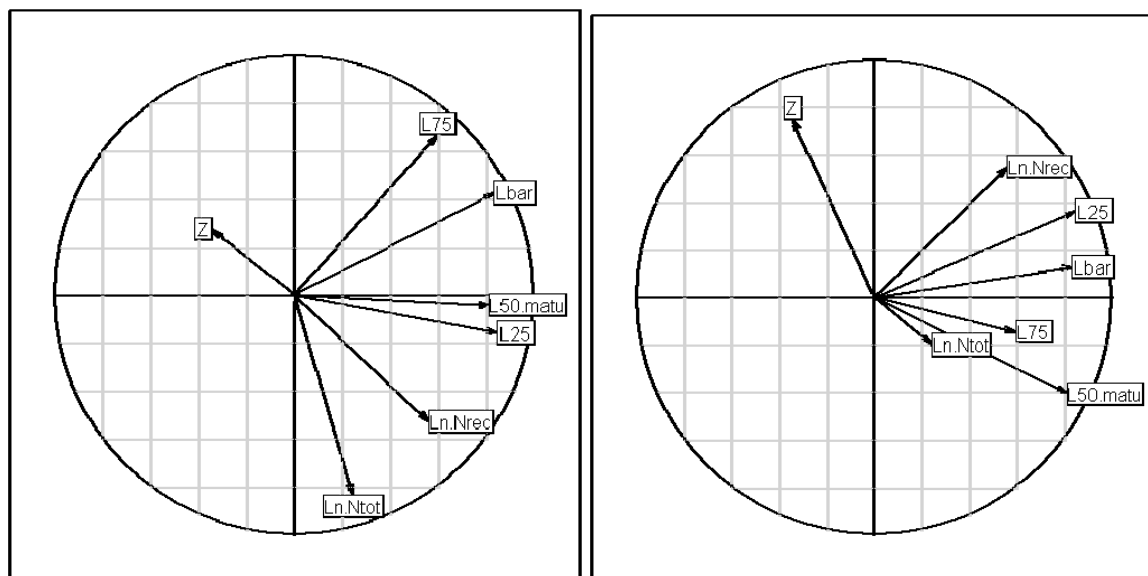


Figure 5.10.1. Correlation structure between the biological non spatial indices (Fisboat Table 2) for North Sea cod. Left: principal axes 1 and 2; right: principal axes 1 and 3.

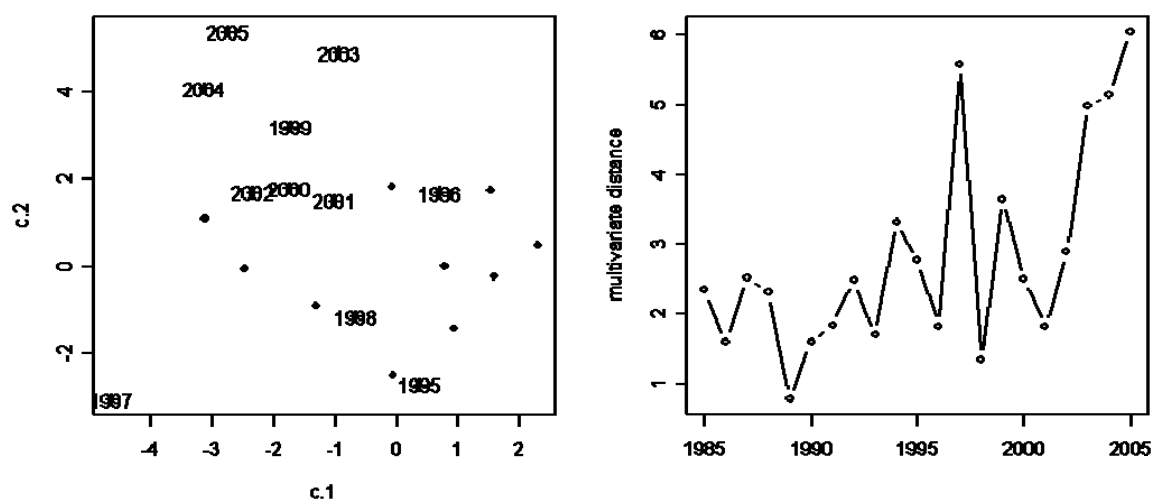


Figure 5.10.2. Monitoring North Sea cod in the factorial sub-space of the two first principal axes using the biological non spatial indicators (Fisboat Table 2). Left: representation of years in the factorial sub-space (the black diamonds are the reference years); right: the time series of the multivariate distance representing the deviation of the stock from its reference status.

References

- Chessel, D., Dufour, A. and Dray, S. 2006. The ade4 package. R documentation. R-project.
- Lebart, L., Morineau, A. and Piron, M. 1995. Statistique exploratoire et multidimensionnelle. Dunod, Paris.

5.4.2 Multi-factorial analysis (MFA) and spatial indicators

Pierre Petitgas, Jean Charles Poulard.
IFREMER, Nantes, France.

Introduction

MFA is a multi-table analysis method (Escoffier and Pagès, 1994 ; Dazy and Le Barzic, 1996) extending PCA methodology (Principal Component Analysis) to the analysis of 3D structured data. In particular, MFA is designed for cases where the same variables (data matrix columns) are measured for the same individuals (e.g., stations : data matrix rows) at various times (third dimension of the data structure). The method allows the construction of a unique factorial space in which to represent each data matrix for each time, each variable and each individual. This unique factorial space is a compromise space that best matches that of each data matrix at each time. It allows extensive tables of data to be represented pictorially in such a way that groupings among variables in space and time are readily identifiable.

The method has been applied in fisheries science to characterise seasonal and inter-annual variation in fish community structure (Gaertner et al., 1998), fishing activity (Poulard and Léauté, 2002), as well as common structure between trophic levels (Petitgas et al., 2006). It has been used in FISBOAT for summarising the average life cycle spatial organisation (Woillez et al., in press). Here we suggest a measure of inter-annual variation in that pattern and an R code for doing so.

Method

The method proceeds in two steps. First a PCA is performed on each data matrix. Then each variable at each time is weighted by the inverse of the first eigen value of that matrix. Then a general matrix is constructed that contains all the weighted variables in columns and the individuals as rows. The PCA of that general matrix constructs the MFA compromise factorial space. Its principal axes are interpretable using the correlation of the variables to them. The interest in the method is the construction of a compromise factorial space in which to represent the 3D structure of the data : each individual is represented by n points (n repetitions in time) as well as by its gravity centre (average position in the compromise factorial space).

Software

The R code available from the FISBOAT web site, <http://www.ifremer.fr/drvecohal/fisboat/index.htm>, is `dmul_mfa.R`. It is commented. It uses the R library `ade4` (Chessel et al., 2006) for performing the MFA. Using the Fisboat Table 1 of spatial indicators [<http://www.ifremer.fr/drvecohal/fisboat/>] as input, the code will produce a figure representing the life cycle spatial pattern in the two first principal axes of the MFA space. On the figure, each point represents a particular age in a particular year. The average position for each age is labelled. A multivariate distance is built quantifying in each year the deviation from the average. This distance, named `dmul`, is the sum over the ages of the distance between the yearly position of each age $p(a,y)$ and its gravity centre $c(a)$:

$$d_y = \sqrt{\sum_a d^2(p_{a,y}; c_a)}$$

The code also provides as output a table of the number of times the correlation between each spatial indicator and the MFA axes is greater than 0.5, allowing interpretation of the axes. The output result is the times series of the multivariate `dmul`

distance, which characterises the inter-annual variation around the average spatial pattern of the life cycle.

The position of each age in each year $p(a,y)$ is computed by applying MFA to all years. The age-specific gravity centres relative to which deviations are referenced are calculated considering reference years only, that must be provided as input. The gravity centre for each age is the average position of each age specific point in the MFA space for the years of the reference period.

Inputs are : Table 1 of spatial indices, years to consider as the reference period, and the number of principal MFA axes to consider for computing the distance $dmul$. Outputs are a figure representing the life cycle pattern in the MFA space, correlation tables of the variables with the MFA axes, the time series of the $dmul$ distance as well as its histogram.

Example

North Sea cod

The life cycle of any fish population is organised in space because the fish will occupy different habitats at different life stages. Therefore, the young and the old can be expected to show different characteristics in their spatial distribution. The spatial indicators developed in Fisboat have been applied to survey numbers-at-age for a series of years. They thus characterise the fish distributions at age in different years. In each year, the matrix made of the spatial indicators in columns and the ages in rows describes the life cycle pattern of the population in that year. In the MFA compromise space, the cloud of points of the compromise individuals represents the average life cycle pattern. In general, structure in the data is strong and the first two principal axes are explained at least by location and aggregation indicators, meaning that there is a change with age in the location and aggregation of the fish. MFA is then used to describe these patterns and their inter-annual variation. In particular, the departure of individual points in each year from the average cloud represents inter-annual variation in the life cycle organisation. An R code has been developed to calculate such variation.

Reference years are 1985-1994. Fig. 5.11.1 illustrates the structure in the spatial organisation and fig. 5.11.2 illustrates the monitoring of the deviations from the mean reference structure. Each point on fig. 5.11.1 (left) represents the position of a given age and year in the MFA factorial sub-space of the first two principal components. The age labels identify the mean position for each age for the reference years, materialising the reference life cycle spatial organisation of North Sea cod. In effect, marked and progressive differences exist across ages. The table on fig. 5.11.1 (right) summarises the correlation structure in the spatial indicators. The first component is determined by having a larger area occupied, a centre of gravity more to the north and west, higher inertia (dispersion), and a lower nugget effect. The second component is determined by having a centre of gravity more to the north, higher anisotropy and a smaller positive area. Marked and progressive differences in the spatial distribution of ages are shown on fig. 5.11.1 (left), which characterise the life cycle spatial pattern of North Sea cod. Young (A1) and old ages (A5-6) differ on the first component from intermediate ages (A2-4). Spatial distributions of young and old ages are more to the east, are less dispersed, occupy a smaller area, and have a higher nugget effect than that of intermediate ages, which are more to the west, more dispersed, occupying larger areas and with smoother correlation. Age 1 and Ages 5-6 differ on the second principal component by

the location of their centre of gravity and anisotropy. The spatial distribution of old ages is more to the north, is more anisotropic, and occupies a smaller area than that of the age 1 fish.

In order to monitor the departure of all the ages from the reference mean spatial organisation, the multivariate distance $dmul$ was calculated (fig. 5.11.2): in each year the distance between the age reference point (labelled) and the current year's point (black dot) was calculated and summed over the ages. Ages 4 and 5 show some elongation in the direction of departure from their reference, meaning that these ages tend to show a systematic change, namely a reduction in area occupied and a more northerly centre of gravity. This is quantified by the multivariate distance on which a statistical monitoring scheme could be applied.

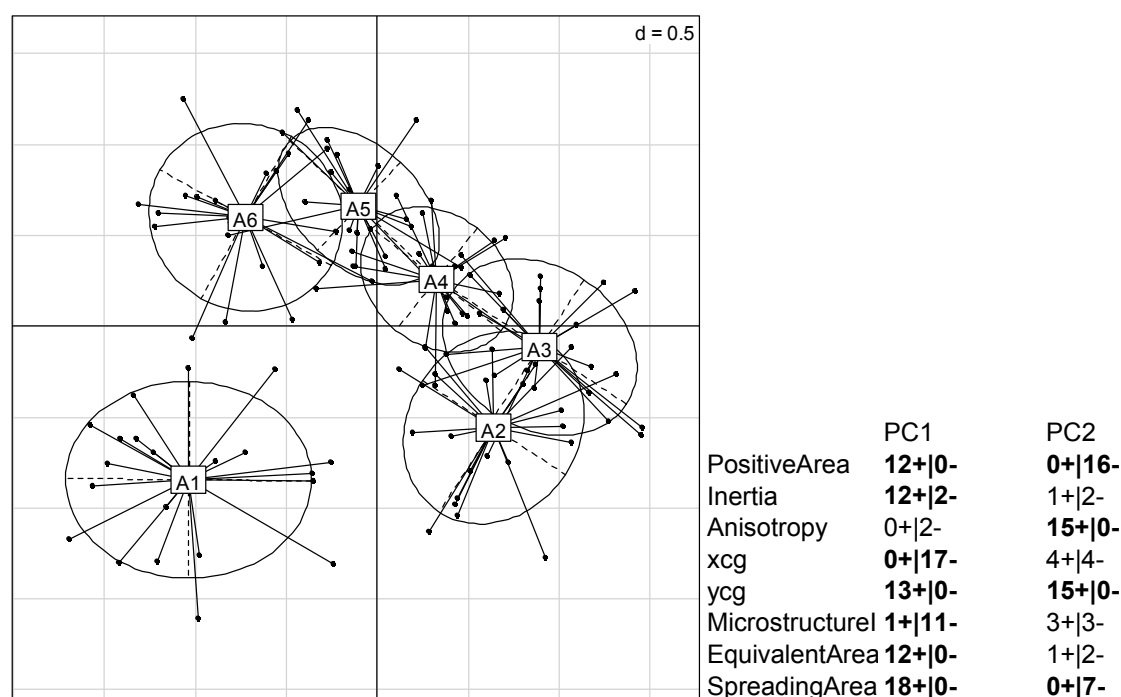


Figure 5.11.1. Left: MFA representation of each age in each year (points) relative to the age gravity centre. Right: Spatial indicators that are the most correlated to the first two principal axes

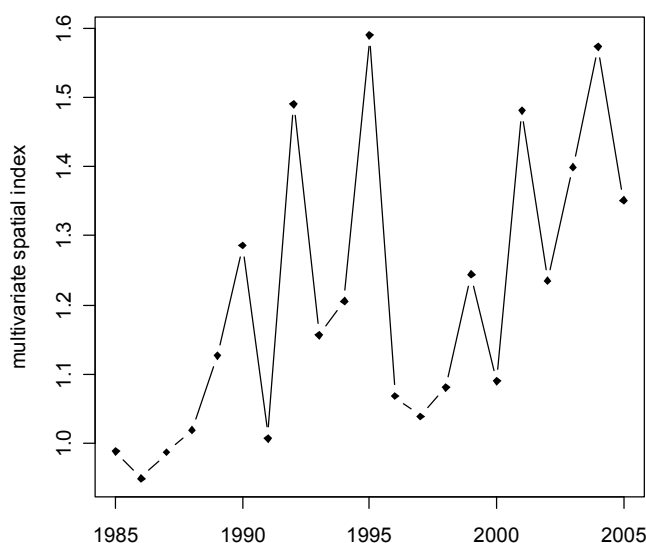


Figure 5.11.2. Monitoring North Sea cod in the sub-space of the two first principal MFA axes using the spatial indicators (Fisboat Table 1) : time series of the multivariate distance representing the deviation of the stock from its reference spatial distribution all ages considered.

References

- Chessel, D., Dufour, A. and Dray, S. 2006. The ade4 package. R documentation. R-project.
- Dazy, F. and Le Barzic, J.-F. 1996. L'analyse des données évolutives : méthodes et applications. Editions Technip, Paris, 227p.
- Escoffier, B. and Pagès, J. 1994. Multiple factor analysis (AFMULT package). Computational Statistics and Data Analysis, 18 : 121-140.
- Gaertner, J.-C., Chessel, D. and Bertrand, J. 1998. Stability of spatial structures of demersal assemblages. Aquatic Living Resources, 11 : 75-85.
- Petitgas, P., Massé, J., Bourriau, P., Beillois, P. et al. 2006. Hydro-plankton characteristics and their relationship with sardine and anchovy distributions on the French shelf of the Bay of Biscay. Scientia Marina, 70(S1): 161-172.
- Poulard, J.-C. and Léauté, J.-P. 2002. Interaction between marine populations and fishing activities: temporal patterns of landings of La Rochelle trawlers in the Bay of Biscay. Aquatic Living Resources, 15: 197-210.
- Wuillez, M., Poulard, J.-C., Rivoirard, J., Petitgas, P. and Bez, N. (accepted February 2007). Indices for capturing spatial patterns and their evolution in time, with application to European hake (*Merluccius merluccius*) in the Bay of Biscay. ICES Journal of Marine Science.

5.4.3 Min/Max autocorrelation factors (MAFs) and time continuity

Wuillez M., Rivoirard J.

Centre de Géostatistique, Armines/Ecole des Mines de Paris,
Fontainebleau, France.

What are MAFs?

Min/max autocorrelation factors (MAFs) are a multivariate statistical method, having similarities with the classical Principal Components of PCA when analyzing repeated values taken by a set of variables. When applied to a time series, the MAFs allow the set of initial variables to be decomposed into factors, the autocorrelation of which decreases from the first factors to the last ones (or more generally, the variogram – half variance of increments – of which increases from the first factors to the last ones). Hence the very first factors extract the part of the variables which is the most continuous in time. A recent application of MAFs in a fishery context was reported by Erzini et al. (2005).

Some details are now provided, beginning with PCs. The PCs are linear combinations of the original variables, each of them explaining a decreasing part of the variability present in the values (these values can be seen as a cloud of points in the space having the initial variables as coordinates; the cloud is centered on the means of the variables, and we are interested in the variability of the cloud around its center, which corresponds to the centered variables). The PCs are uncorrelated with each other. The 1st component explains the highest part of the variability (it corresponds to the direction of maximal variability of the cloud). The 2nd component explains the second highest part of the variability, while being uncorrelated to the 1st one (it corresponds to the direction of maximal variability of the cloud, while being orthogonal to the first direction), and so on. Then the set of variables can be decomposed and represented by these uncorrelated PCs, and they can be summarized by selecting the often few PCs that explain most of the variability of the cloud.

Note that the PCs depend on the magnitude of the values taken by the different variables, and then on the unit used for each of them. Because the different variables may be of different nature, with different and conventional units, PCA is very often performed on the normed variables (i.e. having the variance of each variable set to 1).

PCA is well suited to the case where the repeated values of the variables are independent. In case of repetition in time (time series) or in space, the PCs are uncorrelated with each other at the same time (or location), but may be correlated between different times (or locations), making the representation of the variables by the factors (the PCs) less appropriate.

The MAFs, which are also linear combinations of the original variables, and have a variance of 1, offer a better representation of variables distributed in time or space: in addition to being uncorrelated with each other at the same time (or location), they are uncorrelated with each other for a given time (or space) lag (taken equal to the sampling lag in practice). Moreover they are computed with the aim of: (1) presenting the highest autocorrelation (or smallest variogram) at this lag for the 1st MAF; (2) then presenting the second highest autocorrelation at this lag, while being uncorrelated with the 1st MAF, for the 2nd MAF. Etc. Hence, in a time series, the MAFs offer a way to build the combinations of variables which present the maximal continuity in time (as measured at the lag) for the first MAFs, and the minimal continuity for the last MAFs.

Note that the MAFs depend on the chosen computation lag and may be correlated with each other for other lags. Note also that the MAFs do not depend on a possible normalisation of the initial variables.

From a technical point of view, the MAFs are the solution of a generalized eigenvalues problem. This can be simplified into a simple eigenvalues system in the case where the variables are uncorrelated with each other at the same time (or location). Hence MAFs can be obtained by solving two simple eigenvalue systems: one to transform the initial variables into PCs, the other to obtain the MAFs from the increments of PCs at the computational lag (by maximizing/minimizing the variance of increments, i.e. minimizing/maximizing the autocorrelation).

The MAFs in Fisboat

In Fisboat (work package WP2A), a set of spatial indices were selected to represent a target spatial population over its time series. The estimated indices presented notable variations in time. These may have been due to actual variations but also to various errors. MAFs can be used to extract from the set of indices the very first factors, that present the maximal continuity in time, and that can be thought to be used for a follow-up of the population in time.

Note: Here the continuity in time is measured at the lag of the time series: for instance a lag of 1 year if there is a survey every year, or the varying lag between successive surveys if there are gaps.

The interpretation of the MAFs

The very first MAFs (typically MAF1 and 2) allow us to extract trends in the multivariate time series of a set of indices. A jump upwards or downwards in the trends can be interpreted as a change in the spatial pattern of the considered age or functional group. Loadings informs us about the contribution of each index in the observed trends. Such an index can be used in an indicator approach to qualify a component of stock status (e.g. the spatial component). Moreover, correlation and delayed correlations allow us to put in relation MAFs and the abundance.

The limits of the MAFs

The number of MAFs cannot exceed the number of variables, nor the number of year increments (number of sampled years - 1). If the number of variables tends to be larger than the number of sampled years, the MAF $n^{\circ} i$ ($i = 1, 2, \dots$) tends to have a period (number of years - 1)*2/ i . In particular there will be evidence of a high continuity with period (number of years - 1)*2 for MAF1, (number of years - 1) for MAF2, etc, whatever the data, which may not be significant outside this series e.g. for additional years.

To prevent such an overfitting to the very detailed values of the variables, and so to increase the significance of the MAFs, these are computed while adding a repeated random noise to the variables. The noise, which is function of the number of variables and years, vanishes when the time series is longer.

Software

An R script has been written, MAF_noise_script_WP5.R, in which the solution of MAFs is obtained from a double call to the PCA standard routine prcomp(). In the first call, PCs are

determined from the initial variables. In the second call, MAFs are computed from the increments of the PCs at the time lag.

MAFs are computed for each of 100 realizations where a random noise is added to the centered and normed indices. The random noise follows a Gaussian distribution with mean 0 and variance $0.1 * (\text{nb.indices} / (\text{nb.years} - 1))$. For each resulting MAF, the 100 values of the loading of each index are available. Their distribution is symmetric. The median of the loading values (which is more robust than the mean) is used to determine the median MAFs profile.

The MAFs that present the lowest variogram (half variance of increments) at the lag are retained finally. By default the two first MAFs are retained. (Note: since each MAF has a variance of 1, it can be shown that a small value for the variogram at short time lags is compensated by some large values at larger time lags. This explains the large variogram values at large time lags that can be found for the first MAFs).

Output per age or functional group for a target spatial population:

For each retained MAF, coming from the 100 realizations:

- the median MAF time series;
- the median MAF variogram;
- the contribution (median of the loading coefficients) of each initial index into the median MAF.

For each retained MAF:

- the median MAF time series;
- the times series of the log of the abundance;
- the regression between the median MAF and the log of the abundance;
- the delayed correlation between the median MAF and the log of the abundance.

Important remark on the sign of MAFs:

Like a PC, a MAF is equivalent to its opposite (= the MAF with changed sign, that would be obtained by changing the sign of each coefficient of its linear combination), since the unit variance and the variogram at the computation lag would be unchanged. Then a MAF that would be monotonic over a time series can indifferently appear to be increasing as well as decreasing. Similarly a MAF with an extremum in the middle of a time series could indifferently present a maximum or a minimum.

References

- Switzer, P. and Green, A. A. 1984. Min/max autocorrelation factors for multivariate spatial imaging. Technical report no 6, Department of Statistics, Stanford University, 14 p.
- 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**, 353-359.

5.5 Diagnosing stock status from indicator series.

5.5.1 Combining trend signals using a cause-effects table

Verena M. Trenkel, Marie-Joëlle Rochet and Benoît Mesnil
Ifremer, Nantes, France

Introduction

This method provides guidance for an interpretation of time trends in indicators of the biological status of a stock derived from survey data (Gangl and Pereira 2003; Rochet and Trenkel 2003; Ault *et al.* 2005), and indicators of fishing pressure derived from knowledge of the fishery. Trends are interpreted relative to a past reference period. The aim is to estimate the current state of a stock of interest with respect to management objectives. A fuller account of the method together with a practical example of its application is given by Trenkel *et al.* (2007).

Method

At the preliminary stage, managers define operational objectives, for example, to obtain landings at a certain level, or individuals of a certain average size. Meanwhile, scientists and stakeholders decide upon a suite of indicators suitable for monitoring the stock and the fishery, possibly using the multi-stage framework proposed by Rice and Rochet (2005). Secondly, scientists examine estimated values of the selected indicators together with reports of any other studies that pertain to the stock at some time in the past, called the reference state, most probably the starting year of a survey time series. They then categorise the stock as having been either ‘satisfactory’ or ‘unsatisfactory’ for each indicator at that time. Thirdly, changes in indicator values after this reference time are estimated, interpreted, and combined into a diagnostic that highlights possible causes of the changes observed. Finally, this diagnostic is considered with the management objectives, indicators of fishing pressure, and past experience of managing the fishery in order to decide appropriate managerial actions.

In more detail, having selected a suite of suitable indicators, there are several steps:

1. Select the reference time and calculate time series of indicators.
2. Determine time trends and status for each indicator in the current year.
3. Evaluate any other relevant information and combine the results of different indicators to provide an interpretation of the changes observed.
4. Agree a final diagnostic, including possible causes.
5. Determine trends in fishing pressure and propose appropriate management actions in the light of the diagnosis and stated management objectives.

For step 2, time trends in population indices from the reference year to the current year allow the current population dynamics to be assessed with respect to the reference situation. Uncertainty and natural variability in the survey data are accommodated through a hypothesis testing framework. A hypothesis test involves two risks of error, the type-I error of detecting a trend when none occurs, and the type-II error of not detecting a real trend. Whereas the α -level of the type-I error can be selected, the probability of type-II errors is generally not known, but it increases as the α -level decreases. Clearly there is a trade-off here between the type of error that is to be most avoided, so the selection of α -levels in hypothesis tests for time trends is a task for managers of the fishery. Time trends over the whole series provide

estimates of longer term changes while trends over a more restricted number of years can inform on recent changes. Generally at least five years of data are required to be able to detect linear trends, in many cases, however, much longer series (>20 years) are necessary due to the high uncertainty and large random inter-annual variations in population indices (Nicholson and Jennings, 2004). The choice of the management time scale is clearly a decision for managers. Even with a multi-annual approach it is desirable to detect drastic changes in order to take rapid measures if necessary. For this, a choice of methods is available, as described elsewhere in this FISBOAT manual, e.g. CUSUM charts (Mesnil and Petitgas), and the method of second derivatives (Trenkel). For certain indicators reference points may exist, for example Z^* for the total mortality rate (Die and Caddy, 1997) which can be used to evaluate directly the status of an index in the current year. The final aim of step 2 is to determine the direction of the most recent changes for each indicator, i.e. decreasing, stable or increasing.

Step 3 involves combining the results of several population indices. One well known method is the traffic light approach put forward by Caddy and others (Halliday *et al.* 2001; Caddy 2002). Depending on how many indicators are red, i.e. in an undesirable state, the overall assessment is set. For this approach the different indices take equal weights but they could just as well be weighted based on some *a priori* criteria. Rochet *et al.* (2005) proposed an alternative approach based on combining population and community indicators according to their biological meaning. Step 3 should also involve consideration of biological information additional to that provided by the indicator series (e.g. recruitment estimates, mean weight-at-age) in order to clarify the likely causes of the changes observed. Investigation of time-trends in indicators for fishing pressure (Piet *et al.* 2007) such as days-at-sea, fishing mortality, and quantities landed and discarded will allow corroboration of whether changes in fishing pressure could have been the major cause, before stating the final diagnosis (step 4).

The last step is to propose possible managerial measures linked to the diagnoses. The proposals should depend on whether the reference state was considered satisfactory or not, and whether fishing pressure increased since the reference year. The advice provides the direction of appropriate measures rather than prescribing them in quantitative terms, leaving the final decision to managers who should be guided by past experience.

Example

Here we apply this approach to an imaginary case of five population indicators: total mortality (Z), log-transformed abundance $\ln N$, mean length \bar{L} , and length quartiles $L_{25\%}$ and $L_{75\%}$. Table 5.4.1 shows the increasing or decreasing effects (shown with arrows) that the most relevant factors are expected to have on each indicator. We assume that its long term time trend from the reference year to the current year has been categorised as either increasing, stable or decreasing. Using the method proposed by Trenkel (2006), the sign of the recent time trend for $\ln N$ is also obtained. Table 5.4.2 is then used to determine which time trends, short or long term, are used in table 5.4.1. Depending on agreement or not between long and short term trends in $\ln N$, the estimated category of the short or long term trends determines which trend is used. The latter case occurs if a recent increase in $\ln N$ can be explained by recent management measures that reduce fishing while the long term trend is decreasing. Note that table 5.4.1 only considers single causes and not the expected effects of combined causes such as reduced recruitment and increased fishing. In certain cases additional information such as recruitment estimates, should be sought in order to clarify the causes of the observed changes. Thus, in the proposed approach, biological knowledge is

used to point at the possible causes behind the observed changes, at least as far as being able to say whether increased fishing pressure could have contributed.

The last step consists of proposing possible managerial measures. The last column of table 5.4.1 makes suggestions for mitigating measures based on whether the reference state is considered to be impacted by fishing (1), or satisfactory (2). As the diagnosis is qualitative, so are the managerial measures. The direction rather than a prescribed amount of treatment is proposed. The quantitative decision is left to the managers. The adaptive management approach advocated in the 1980's by Walters (1986) seems a natural choice for implementing qualitative management advice, with a careful monitoring of response to the chosen policy being fed into the process for the next time step. The listed measures are by no means exhaustive and are only intended to give a flavour of the kind of advice that could be proposed.

Table 5.4.1: Expected effects of different causes on selected indicators and possible mitigation measures for counterbalancing changes. Δ stands for change, — for no effect, \nearrow for increasing and \searrow for decreasing. Population indicators: Z = total mortality, $\ln N$ = log-transformed total abundance, L_{bar} = mean length, $L_{25\%}$ and $L_{75\%}$ = length distribution quartiles.

Cause	Z	$\ln N$	L_{bar}	$L_{25\%}$	$L_{75\%}$	Mitigation measure 1 = impacted reference state 2 = satisfactory reference state
\nearrow fishing mortality	\nearrow	\searrow	\searrow	—	\searrow	1 & 2 : $-\Delta F$: reduction in overall fishing mortality
\searrow fishing mortality	\searrow	\nearrow	\nearrow	—	\nearrow	1: status quo 2: $+\Delta F$ allowed
\nearrow recruitment	—	\nearrow	\searrow	\searrow	—	1: status quo 2: $+\Delta TAC$: increase in TAC
\searrow recruitment	—	\searrow	\nearrow	\nearrow	—	1 & 2: $-\Delta TAC$: reduction in TAC
Faster growth	—	—	\nearrow	—	\nearrow	1: status quo 2 : Δ selection pattern: decrease selectivity to smaller sizes
Slower growth	—	—	\searrow	—	\searrow	1 & 2: Δ selection pattern: increase selectivity to larger sizes
Larger fish caught (Δ fishing area, stock distribution or gear)	\searrow	\nearrow	\nearrow	—	\nearrow	1: status quo 2: Δ selection pattern: decrease selectivity to smaller sizes
Smaller fish caught (Δ fishing area, stock distribution or gear)	\nearrow	\searrow	\searrow	\searrow	—	1 & 2: Δ selection pattern: increase selectivity to larger sizes

Table 5.4.2. Table for determining whether 'long term' or 'recent term' trends are used in table 5.4.1. If recent time trends are used, the reference state at the beginning of the recent period needs to be used!

Long term trend in $\ln N$	Recent trend in $\ln N$		
	\nearrow	—	\searrow
\nearrow	long term	long term	recent term
—	if recent management measures intended to decrease fishing: continue	long term	recent term
\searrow	else: as for long term	long term	long term

References

- Ault, J. S., Smith, S. G., and Bohnsack, J. A. 2005. Evaluation of average length as an estimator of exploitation status for the Florida coral-reef fish community. *ICES Journal of marine Science*, 62: 417-423.
- Caddy, J. F. 2002. Limit reference points, traffic lights, and holistic approaches to fisheries management with minimal stock assessment input. *Fisheries Research*, 56: 133-137.
- Die, D. J., and Caddy, J. F. 1997. Sustainable yield indicators from biomass: are there appropriate reference points for use in tropical fisheries? *Fisheries Research*, 32: 69-79.
- Gangl, R. S., and Pereira, D. L. 2003. Biological performance indicators for evaluating exploitation of Minnesota's large-lake walleye fisheries. *North American Journal of Fisheries Management*, 23: 1303-1311.
- Halliday, R. G., Fanning, L. P., and Mohn, R. K. 2001. Use of the Traffic Light method in fishery management planning. CSAS Research Document, 2001/18, 41 pp.
- Rochet, M. J. 2000. Does the concept of spawning per recruit make sense? *ICES Journal of marine Science*, 57: 1160-1174.
- Rochet, M.-J., Trenkel, V., Bellail, R., Coppin, F., Le Pape, O., Mahé, J.-C., Morin, J., Poulard, J. C., Schlaich, I., Souplet, A., Vérin, Y., and Bertrand, J. A. 2005. Combining indicator trends to assess ongoing changes in exploited fish communities: diagnostic of communities off the coasts of France. *ICES Journal of marine Science*, 62: 1647-1664.
- Rochet, M. J., and Trenkel, V. M. 2003. Which community indicators can measure the impact of fishing? a review and proposals. *Canadian Journal of Fisheries and Aquatic Sciences*, 60: 86-99.
- Trenkel, V. M. 2006. Nonparametric method for determining recent trends in indicator time series. Working document for Fisboat project. June 2006, 7 pp.
- Walters, C. 1986. Adaptive management of renewable resources. Biological resource management, Macmillan Publishing Company, New York. 374 pp.

5.5.2 A 'traffic light' procedure based on Cusum out-of-control tables.

Pierre Petitgas,
IFREMER, Nantes,
France

Introduction

Usually, more than one indicator is used to assess the status of a fish stock. The important question then arises of how to combine the various results into a single assessment. One simple, illustrative way of doing so is to use a 'traffic light approach' (Caddy et al., 2005). An alternative approach is that of multivariate statistical process control. Both are complementary as the CUSUM diagnostic table with all indicators may assist interpretation of multivariate alarm signals. Part II (section 5.9) of this 2-part section illustrates the multivariate statistical approach.

Example

To illustrate, a CUSUM procedure (see Mesnil and Petitgas elsewhere in this manual) was applied separately to time-series of each of several attributes (indicators) for the North Sea cod stock studied under FISBOAT and available at <http://www.ifremer.fr/drvecohal/fisboat/>. The CUSUM procedure comprises three steps (Hawkins and Olwell, 1997; Montgomery, 2005). First, a reference period is defined as a period when the health of the stock was considered acceptable (in-control). This serves to estimate the in-control reference mean and variance for each indicator. The same reference period was applied to all indicators with the results shown in table 5.8.1. Second, the CUSUM was tuned for each indicator to signal important deviations from the reference mean in years outside the reference period. The tuning of the CUSUM results in statistically defining the false alarm rate, and the no alarm rate associated with the in-control limits that are set to enclose acceptable deviations from the reference mean. The application of the CUSUM to each indicator for the cod stock resulted in an array of deviations from the reference mean vector expressed in standard deviation units. This is the CUSUM diagnostic table shown in table 5.8.2. Each column of the array corresponds to each indicator time series of deviations. Because the deviations being expressed in units of standard deviation, comparisons between indicators are immediate. Setting the non-alerting deviations to zero, the diagnostic CUSUM table provides the quantitative values of the deviations from the reference means with a + or – sign which trigger alarm signals. Clearly, in some years only a small number of indicators signal alarms, some perhaps with high deviations, whereas, in other years, many will signal. Here, we used expert judgement in assigning each year as 'in-control' or 'out-of-control' based on which indicator signalled alarm and how many of them there were (table 5.8.2). Cells in the table may be coloured red, orange, or green, as for traffic lights, to show at a glance the perceived seriousness of the state indicated.

Table 5.8.1: Parameters of the CUSUM control scheme for North Sea cod for the biological indicators. Reference period is 1985-1994. Parameters are: μ , σ : mean and standard deviation in reference period; k : allowance in σ units; h : decision interval in σ units; $ic.arl$ (years): in-control ARL (average run length to a false alarm) also noted $ARL(0)$; $ic.rl.25$ (years): RL value at the first quartile of the RL distribution; $oc.arl$ (years): out-of-control ARL (average run length to signal real change) also noted $ARL(2k)$. Indicators are: Survey index: abundance index for ages 1 to 6; Recruit index: abundance at age 2; $Lbar$: average length in the population; $L25$: length value of the first quartile; $L75$: length value of the third quartile; $L50$ maturity: length value at which 50% of the population is mature; Z : apparent total mortality.

	Survey index	Recruit index	Lbar	L25	L75	L50 maturity	Z
μ	19.12	18.00	34.77	20.69	41.70	65.44	1.12
σ	0.26	0.77	4.80	5.16	6.45	5.24	0.44
k	1.3	0.9	1.2	0.9	0.8	1.1	1.0
h	1.0	1.0	1.0	1.0	1.2	1.1	1.0
$ic.arl$	79.3	27.5	60.0	27.5	30.0	56.2	35.3
$ic.rl.25$	23.0	8.0	17.0	8.0	9.0	16.0	10.0
$oc.arl$	1.5	1.9	1.6	1.9	2.3	1.8	1.8

Table 5.8.2: CUSUM diagnostics table for North Sea cod using biological population indicators. Values are the deviation from the reference mean for each indicator in standard deviation units. Reference period is 1985-1994. The procedure signals an alarm from 1999. Indicators are: Survey index: abundance index for ages 1 to 6; Recruit index: abundance at age 2; Lbar: average length in the population; L25: length value of the first quartile; L75: length value of the third quartile; L50 maturity: length value at which 50% of the population is mature; Z: apparent total mortality.

Year	Survey Index	Recruit index	Lbar	L25	L75	L50 mat-urity	Z	diag-nostics
1985	0.00	0.00	0.00	0.00	0.00	0.00	0.00	ref
1986	0.00	0.00	0.00	0.00	0.00	0.00	0.00	ref
1987	0.00	0.00	0.00	0.00	0.00	0.00	0.00	ref
1988	0.00	0.00	0.00	0.00	0.00	0.00	0.00	ref
1989	0.00	0.00	0.00	0.00	0.00	0.00	0.00	ref
1990	0.00	0.00	0.00	0.00	0.00	0.00	0.00	ref
1991	0.00	0.00	0.00	0.00	0.00	0.00	0.00	ref
1992	0.00	0.00	0.00	0.00	0.00	0.00	0.00	ref
1993	0.00	0.00	0.00	0.00	0.00	0.00	0.00	ref
1994	0.00	0.00	0.00	0.00	0.00	0.00	0.00	ref
1995	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
1996	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
1997	0.00	0.00	-1.84	-1.03	-2.21	-1.85	0.00	
1998	0.00	0.00	0.00	0.00	-2.00	-3.44	0.00	
1999	-1.27	-1.30	0.00	0.00	0.00	-7.36	0.00	alarm
2000	-1.65	0.00	0.00	0.00	0.00	-9.34	0.00	alarm
2001	-3.04	0.00	0.00	0.00	0.00	-9.84	0.00	alarm
2002	-3.96	0.00	0.00	0.00	0.00	-12.78	0.00	alarm
2003	-7.48	0.00	0.00	0.00	0.00	-15.95	0.00	alarm
2004	-10.50	-1.18	0.00	0.00	0.00	-19.33	0.00	alarm
2005	-14.97	-2.02	0.00	0.00	1.23	-23.10		alarm

References

- Caddy, J., Wade, E., Surette, T., Hebert, M. and Moriyasu, M. 2005. Using an empirical traffic light procedure for monitoring and forecasting in the Gulf of St. Lawrence fishery for the snow crab, *Chionoecetes opilio*. Fisheries Research, 76: 123-145.
- Hawkins, D. and Olwell, D. 1997. Cumulative Sum Charts and Charting for Quality Improvement. Springer Verlag, New York.
- Montgomery, D. 2005. Introduction to statistical quality control. Wiley, New York. 5th Edition.

5.5.3 A multivariate statistical procedure.

Pierre Petitgas,
IFREMER, Nantes,
France

Introduction

Part I of this presentation (see section 5.8) illustrated the 'traffic light approach' for combining results from time-series of several indicators into a single assessment of a fish stock. A complimentary approach is to use multivariate statistical methods. An illustrative example is given here. More detail on the PCA and MFA approaches discussed can be found in sections 5.10 and 5.11 of this manual, respectively.

Multivariate statistical process control methods use the relationships existing among control variables to prevent having to deal with false alarms in the many individual control charts. Thus multivariate process control methods are potentially more efficient than control methods based on the analysis of a collection of univariate charts (e.g., Hawkins and Olwell, 1997). Various multivariate control methods are available. Hotelling's T^2 statistic (Hotelling, 1947) is the analogue in the multivariate case of the Shewart chart in the univariate case (e.g., Hawkins and Olwell, 1997). It is best suited to detect large shifts in the mean as it uses the current sample only at each time step. To detect rapidly small shifts in the mean, consecutive samples need to be considered for which multivariate CUSUM methods (Crozier, 1988) and multivariate exponentially weighted moving average (EWMA) methods (Lowry et al., 1992) have been developed. Scranton et al. (1996) used multivariate EWMA on a reduced number of principal components with increased shift detection capability. In environmental monitoring, Manly and MacKenzie (2000) made use of Principal Component Analysis (PCA) to project multivariate observations in a factorial space and assess whether they were inside or outside the area within which the process could be considered in control.

Example

To illustrate a multivariate approach, the same CUSUM results for North Sea cod discussed previously in relation to the Traffic Light approach (section 5.8) will be used with the addition of parallel time series for some spatial indicators. The approach considered is inspired by the PCA procedure of Manly and MacKenzie (2000). First, the centre of gravity in factorial space was estimated for the in-control reference years. Next, the distance in factorial space of the observation in each year to that centre of gravity was computed and a new time series constructed from them. Then a CUSUM procedure was applied to that distance. PCA-based distances were applied to the biological (non-spatial) indices, while, for the spatial indices, MFA-based distances (see section 5.11) were used. Evolution of the stock can then be summarised with two distances, one for the spatial and one for the non-spatial indicators. A traffic light type table with two columns (one for each distance) of CUSUM deviations from the in-control domain represents simply the multivariate process control scheme for all the attributes of the stock (tables 5.9.1, 5.9.2).

Table 5.9.1: Parameters of the CUSUM control scheme for North Sea cod for the multivariate indicators. Reference period is 1985-1994. Parameters are: μ , σ : mean and standard deviation in reference period; k : allowance in σ units; h : decision interval in σ units; $ic.arl$ (years): in-control ARL (average run length to a false alarm) also noted $ARL(0)$; $ic.rl.25$ (years): RL value at the first quartile of the RL distribution; $oc.arl$ (years): out-of-control ARL (average run length to signal real change) also noted $ARL(2k)$. Indicators are: Spatial.mul.mfa: MFA-based multivariate distance for the spatial indices; Biol.mul.pca: PCA-based multivariate distance for the biological indices.

Parameter	Spatial.mul.mfa	Biol.mul.pca
μ	1.35	1.67
σ	0.23	0.76
k	0.9	1.5
h	1.2	1.0
$ic.arl$	39.2	142.2
$ic.rl.25$	11.0	41.0
$oc.arl$	2.1	1.4

Table 5.9.2: CUSUM diagnostics table for North Sea cod using multivariate indicators, one for all the spatial indices and the other for all the biological indices. Values are the deviation from the reference mean for each indicator in standard deviation units. Reference period is 1985-1994. The procedure signals an alarm from year 2000. Indicators are: Spatial.mul.mfa: MFA-based multivariate distance for the spatial indices; Biol.mul.pca: PCA-based multivariate distance for the biological indices.

Year	Spatial.mul.mfa	Biol.mul.pca	diagnostics
1985	0.00	0.00	ref
1986	0.00	0.00	ref
1987	0.00	0.00	ref
1988	0.00	0.00	ref
1989	0.00	0.00	ref
1990	0.00	0.00	ref
1991	0.00	0.00	ref
1992	0.00	0.00	ref
1993	0.00	0.00	ref
1994	0.00	0.00	ref
1995	1.40	0.00	
1996	0.00	0.00	
1997	0.00	1.46	
1998	0.00	0.00	
1999	0.00	3.46	
2000	1.77	3.80	alarm
2001	2.66	4.06	alarm
2002	2.00	4.63	alarm
2003	1.69	8.36	alarm
2004	2.42	11.96	alarm
2005	2.85	17.13	alarm

References

- Crosier, R. 1988. Multivariate Generalizations of Cumulative Sum Quality-Control Schemes. *Technometrics*, 30: 291-303.
- Hawkins, D. and Olwell, D. 1997. *Cumulative Sum Charts and Charting for Quality Improvement*. Springer Verlag, New York.
- Hotelling, H. 1947. Multivariate Quality Control. *In: Techniques of Statistical Analysis*, Eds. C. Eisenhart, M. Hastay and W. Wallis, New York, McGraw-Hill, pp.111-184
- Lowry, C., Woodall, W., Champ, C., Rigdon, S. 1992. A Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics*, 34: 46-53.
- Manly, B. and MacKenzie, D. 2000. A cumulative sum type of methods for environmental monitoring. *Environmetrics*, 11: 151-166.
- Scranton, R., Runger, G., Keats, B. and Montgomery, D. 1996. Efficient shift detection using multivariate exponentially weighted moving average control charts and principal components. *Quality and Reliability Engineering International*, 12: 165-171.

Appendix 1

Format for documenting indicators

The format adopted here for presenting indicators is derived from Halliday, R.G. and Mohn, R. (2001). Proceedings of the Fisheries Management Studies WG, 8-11 January 2001. Canadian Science Advisory Secretariat, Proceedings Series 2001/08 (Appendix 5, pp. 45-48).

Description of indicators/indices

INDEX : descriptive name + acronym

Description : short description of what it is (and for case studies in second stage, survey in which it is measured, e.g. "mean length of NEA cod caught during winter Barents Sea surveys")

Stock attribute :

- attribute(s) that the indicator is deemed to reflect (e.g. abundance, productivity, recruitment, mortality, ecosystem, ...)

Derivation :

- document briefly how the index is derived from raw data for each station, and then integrated for the whole survey year (+ ref. published manuals for details).
- and how variance of total index is obtained

Reference points :

- bases for setting the RPs (as target, limit or trigger)
- alternative choices and their rationale

Interpretability :

- how does the indicator reflect stock status or the identified attribute?
- what caveats exist regarding interpretation? (e.g., whole stock vs. population in survey area, if they differ)
- processes involved in changes (is indicator's response specific?)

Measurability : (*keyword here is*: confidence in estimates of the indicator)

- statistical properties of estimator (variability, bias, skewness, ...)
- transformations required before use
- alternative formulations for the same estimators
- alternative estimators of same indicators ; pros & cons

Sensitivity :

- how rapidly & accurately does indicator respond to changes in stock status?
- does natural variability likely masks real changes?

Review of performance :

- performance of the indicator in hindsight, to infer stock status
- document the adequacy of the indicator, its estimator and its RPs (or problems encountered for the specific case study)

References :

(e.g., for case studies, references to published manuals of procedures, articles in which the indicator has been applied in advisory context, ...)