
Intersection–union tests for characterising recent changes in smoothed indicator time series

Verena M. Trenkel^{a,*} and Marie-Joëlle Rochet^a

^a Ifremer, Département EMH, BP 21105, 44311 Nantes cedex 03, France

*: Corresponding author : V. Trenkel, Tel.: +33 2 40374157; fax: +33 2 40374075, email address : Verena.Trenkel@ifremer.fr

Abstract:

Using indicator time series for assessment and management requires methods for characterising recent time trends. We propose an approach where first the indicator time series is smoothed using a generalised additive model with optimal selection of the degree of smoothness. Second an intersection–union test is carried out using two test statistics which are the occurrence of the global maximum (or minimum) within the most recent years and the signs of the estimated annual first derivatives of the smoothed indicator times series during the same period, including years with missing data. The proposed test is applied to fish abundance indices for the North Sea, for which it is they are able to pick up changes happening during the last 3–5 years in contrast to linear regression and the Mann–Kendall test which find much fewer significant recent trends. An additional test for changes in trends using the second derivatives of the smoothed indicator time series provide early warnings for subsequent trends for certain species.

Keywords: Time trend; Metrics; Fisheries management; Assessment; Composite test

1. Introduction

Population and community indicators have become an essential tool for the ecosystem approach to fisheries management. As reference points are not available for many indicators, reference directions rather than absolute reference values have been suggested to be more affordable for assessment purposes (Jennings and Dulvy, 2005; Rochet et al., 2005). Thus the direction of trend in an indicator over time is estimated and then compared to the reference direction expected for the indicator change under the impact of fishing or other drivers. Here we coin 'trend' a change in a given direction that persists over time. Several time horizons might be relevant to decide whether a change is sustained enough to be named a trend that can be compared to a reference direction. Long term trends over one or several decades have been suggested to be the most relevant to fisheries management (van Densen, 2001). However, shorter term trends over three to five years might be worth detecting as well, e.g. for short-lived ecosystem components or for economic indicators that respond on time-scales different from ecological ones. Both long- and short-term trends might be of interest for a particular assessment, the latter relative to longer term dynamics. For example, Caddy and Surette (2005) used three criteria regarding catch trends which had to be fulfilled for a species to be categorised as depleted: 1) a negative time trend in catches over the last five years; 2) an average annual decrease of more than 5% between peak catches and the average catch in the last 3 years; 3) recent catches less than 20% of peak catches. Moreover, detecting changes in trends could be useful as a warning signal that system dynamics are changing, suggesting that a management measure is taking effect or, on the contrary, that new management decisions have to be taken to reverse a trend departing away from a desirable direction.

Many indicators are estimated quantities and thus are prone to various sources of uncertainty and/or variability. For example, survey-based estimates of fish population or community indicators are subject to random sampling variability (Trenkel and Rochet, 2003), but also systematic errors, possibly caused by particular environmental or survey design conditions in a given year (Poulard and Trenkel, 2007). For population indicators, coefficients of variations of 10-30% are not uncommon in many groundfish surveys (V. Trenkel unpublished data). The quantity estimated by the indicator might itself vary under the influence of various factors, many of which show random fluctuations. For example, population abundance will fluctuate randomly from year to year because year-class strength is partly determined by randomly varying environmental conditions. Long-term, recent and changes in trends need to be detected despite these sources of variability.

A commonly used method for determining a trend in estimated indicator time series is fitting linear models and then using the sign of the slope for time if significantly different from zero (e.g. Nicholson and Jennings, 2004; Trenkel and Rochet, 2003). Alternatively non-parametric tests such as the Mann-Kendall test are used (e.g. Conover, 1971). These methods are suitable for determining long term time trends (>10 data points). However, they are less suitable for short term time trends as they lack power (Dulvy et al., 2005). Furthermore, time trends might not be linear, which reduces the power of many tests, and variability in estimated indicators can be as strong as to mask short term changes.

In this paper we propose a method specifically designed to increase testing power for determining and characterising recent time trends in longer indicators time series. To avoid detecting spurious changes caused by outlying years or excessive sampling uncertainty, we do not employ the raw indicator time series but rather a smoothed version of it. The use of smoothed time series is common practice in environmental monitoring, e.g. water quality (Champely and Doledec, 1997; McLeod et al., 1991); phytoplankton (Nicholson et al., 1998), contaminants (Fryer and Nicholson, 1999) or birds (Fewster et al., 2000). Another advantage of smoothed time series is that the method can be applied even if data points are missing in a time series. The extreme points (maximum/minimum), slopes and accelerations of the smoothed time series are used to test for recent trends (first derivatives) and changes (second derivatives) in trends. Immediate management action would then be called for if a

recent time trend in the indicator in an undesirable direction was found. Preparation for future action might be advisable if deceleration (negative second derivative) of a stable indicator or a positive trend occurred. Thus characterisation of recent trends requires testing several hypotheses related to smoothed time series and its derivatives. To combine the tests of several statistics, we use intersection-union tests for which the global null hypothesis is rejected only if the null hypotheses for all test statistics are rejected (Casella and Berger, 1990).

The method is described in details in the next section and applied to North-Sea fish population abundance indices estimated from bottom-trawl survey data. If the method is responsive, the duration of the time period considered will impact test results, as might adding a data point. This is tested by considering different time spans and data periods.

Method

We assume that an indicator time series, for example a population abundance index has been estimated for the populations or community of interest (see Rochet and Trenkel (2003) for a review of indicators suitable for measuring the impact of fishing). Data from scientific surveys are commonly employed for estimating these indicators. The proposed method for detecting recent changes consists of five steps:

Fit a smoother to the whole available indicator time series to obtain a smoothed series;

Test whether smoother provides a satisfactory fit to the data, if yes proceed with 3., otherwise method is unsuitable for data;

Calculate first and second derivatives for the smoothed time series for all years (including years with no data);

Carry out parametric bootstrap of indicator time series to obtain resampled indicator time series and repeat steps 1.-3.;

Carry out tests.

The steps are described in details below.

2. Smoothing indicator time series

To obtain smoothed indicator time series, generalised additive models (GAM) are fitted with year modelled by a thin plate regression spline (Wood, 2006). Thin plate regression splines are basis functions for estimating smooth functions of one or several covariables. Model parameters are estimated by penalised likelihood, where the penalisation term encapsulates the non-linearity ("wiggleness") of the smooth function of time (years). The wiggleness is measured by the integral of the second derivative of the non-linear function over the range of years. The weight of the penalty term in the likelihood function, called smoothing parameter, is estimated by generalised cross-validation (GCV) which minimises the squared prediction error following the methodology proposed in Wood (2006) and implemented in the `mgcv` package in the statistical software R (R Development Core Team 2007). Given that indicator time series might be short, it has to be noted that the rank of the model cannot be larger than number of observations minus one ($n-1$), which limits the form of the non-linear function for very short series, e.g. a linear function for three years. The resulting smoothed time series is the smoothest possible in terms of second derivatives, i.e. changes in local linear time trends, while still following the general time trend. Furthermore, the lag 1 autocorrelation of residuals is zero (Wood 2006). Thus the smoothed times series encapsulates the time trend while the residuals represent iid random sampling noise. The desired consequence is that the autocorrelation of the smoothed indicator time series is much higher than that of the raw time series.

3. Goodness-of-fit test

The approach assumes that the difference between the observed and smoothed series is due to random sampling variability with no autocorrelation, i.e. the autocorrelation of residuals of the smoothed series is zero. Thus the model fitted to the indicator time series I with t as explanatory variable for year is

$$\hat{I} = E[I] = f(t) \quad (1)$$

The estimated smoothed time series is then \hat{I} (step 1). Assuming I is normally distributed, its goodness-of-fit can be tested by comparing the scaled residual sum of squares of the fitted smoothed model to a χ^2 -distribution with $n-p$ degrees of freedoms, where n are the number of data points and p the number of parameters in the fitted model (Wood, 2006). If the p-value of the test is smaller than some fixed level α , the smoothed model provides a better fit than a model consisting of the average value only. The smoothed indicator time series is then declared providing a satisfactory fit to the data (step 2). In the opposite case the smoothed time series does not give a better description than the mean.

4. Calculating first and second derivatives

As spline models are twice differentiable, first and second derivatives of the smooth series exist and can be calculated for every year of the time series using finite difference intervals. As eq (1) is a linear additive model, it can be written as the ordinary linear model (Wood, 2006)

$$E[I] = \sum_{i=1}^q \beta_i b_i(t) = \mathbf{X}\beta \quad (2)$$

where $b_i(t)$ is the value of the i th thin-plate regression spline basis function for the explanatory variable t and β_i is the corresponding unknown model coefficient. \mathbf{X} is the linear predictor matrix with the number of columns corresponding to the rank q of the thin-plate regression spline model and the number of rows to the number of observations; β is the vector of all model parameters. The local first derivatives of the fitted model can then be estimated as the difference between fitted values at t and $t+\delta$ divided by δ , where δ is a small value, e.g. $1e-7$.

$$\hat{F}'(t) = (\hat{I}(t+\delta) - \hat{I}(t)) / \delta = (\mathbf{X}(t+\delta) - \mathbf{X}(t)) \hat{\beta} / \delta \quad (3)$$

Second derivatives are estimated by taking finite differences of estimates of first derivatives. For multi-normally distributed estimated smoothed indicator time series, the estimates of the first and second derivatives are both multi-normally distributed as they are linear functions of the smoothed indicator values. The distribution question has to be considered for any particular indicator time series.

5. Parametric bootstrap

To estimate the standard deviations of the estimated first and second derivatives, 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

calculated by year. This provides the distribution of first and second derivatives for each year for which the sampling variance can be calculated.

6. Building hypotheses tests

For describing recent temporal changes, the signs of the annual first derivatives of the smoothed indicator series and the location of the maximum or minimum are used to derive an intersection-union test (Casella and Berger, 1990). The null hypothesis in an intersection-union test is a union of type $H_0 : \theta \in \bigcup_{\gamma \in \Gamma} \Theta_\gamma$, where θ is the vector of parameters of interest

and Θ_γ the set of values allowed under the null hypotheses of each of the $\gamma=1, \dots, \Gamma$ tests. The alternative hypothesis is expressed as the intersection $H_1 : \theta \in \bigcap_{\gamma \in \Gamma} \Theta_\gamma^c$ with Θ_γ^c the set of values of the individual alternative hypotheses.

There is a recent significant decrease (increase) in the smoothed indicator time series if both conditions

C1: the maximum (minimum) smoothed indicator value \hat{I}_{\max} (\hat{I}_{\min}) is not found within the most recent n years, and

C2: all annual slopes $\hat{R}_t \leq 0$ ($\hat{R}_t \geq 0$), for $t = T-n+1, \dots, T$

are met, where T is the final year in the data. Condition C1 allows to integrate the perspective of the whole time series into the test, because in the case where the indicator is at its maximum in the most recent period, it seems most likely that the underlying process is fluctuating around a steady state and hence any temporary negative or positive trends are without consequence.

The two conditions lead to the definition of null and alternative hypotheses for each sub-test. The corresponding hypotheses tests are described for the case of testing a decreasing recent time trend, the test for an increasing time trend is similar. The test is set up in a way that a recent decreasing time trend is declared significant if all null hypotheses of sub-tests T1 and T2 corresponding to conditions C1 and C2 are rejected (intersection condition).

T1 Test on location of maximum

$$H_{01} : \hat{I}_{\max} \in \{\hat{I}_{T-n+1}, K, \hat{I}_T\} \text{ and } H_{11} : \hat{I}_{\max} \in \{\hat{I}_1, K, \hat{I}_{T-n}\}$$

H_{01} is rejected if the maximum value is not situated during the final n years.

T2 Test for values of annual slopes

$$H_{02} : \prod_{t=T-n+1}^T \hat{R}_t > 0 \text{ and } H_{12} : \prod_{t=T-n+1}^T \hat{R}_t \leq 0$$

Assuming a normal distribution for the estimates of the first derivatives, $\hat{R}_t \sim N(\hat{R}_t, \hat{\sigma}_t)$, where the standard deviation is estimated by bootstrapping, H_{02}^t is rejected at the α level if $P_z(\hat{R}_t / \hat{\sigma}_t > 0) < \alpha$. $P_z(\cdot)$ is the standard normal distribution function. The null hypothesis for subtest T1, H_{02} is then rejected at the α level if all H_{02}^t , $t=T-n+1, \dots, T$, are rejected. The overall test level is also α , as the upper bound of the test level of the intersection-union test with sub-tests T1 and T2 is $\alpha = \sup\{\alpha_{T1}, \alpha_{T2}\}$ according to theorem 8.3.5 in Casella and Berger (1990). The same theorem applies to each sub-test. For sub-test T2, where the estimated first derivatives are autocorrelated, the test level is the maximum test level value of n tests each of level α . Thus test level for sub-test T2 is also α , despite the autocorrelation.

A second intersection-union test is carried out to determine if recent time trends have been accelerating, i.e. be getting more negative in case of a negative trend or more positive for a positive trend. In the case of no significant time trend, the test indicates whether a future

negative or positive trend might be expected. Thus, a recent negative (positive) trend has been accelerating or a stable situation has been degrading, if the condition C3: all annual second derivatives over the n recent years are strictly negative (positive) is fulfilled. For deceleration of trends the opposite conditions apply. The condition is translated into a test as follows

T3 Test for acceleration of a negative recent trend or degradation of a stable situation

$$H0_3 : \prod_{t=T-n+1}^T \hat{d}_t \geq 0 \text{ and } H1_3 : \prod_{t=T-n+1}^T \hat{d}_t < 0$$

Again assuming that estimated second derivatives follow a normal distribution, $H0_3$ is rejected if $P_z(\sum_{t=T-n+1}^T \hat{d}_t / \sum_{t=T-n+1}^T \hat{d}_t \geq 0) < \alpha$; $P_z()$ is again the normal distribution function.

In all tests a common time span n is considered. The suitable value for n will depend on the expected changes and the question examined, but it should be at least three because the autocorrelation in derivative estimates would dominate the test result with two years only, and at most $T/2$ to characterize the recent trends relative to a significantly long past.

7. Application to Southern North Sea fish

The proposed tests were applied to log-abundance indices for 33 fish species and families (see taxa list in Table 1). These indices were calculated using International Bottom Trawl Survey (IBTS) data collected in the Southern North Sea (south of 57N) in January during the period 1983-2005 and stored in the DATRAS data base hold by the International Council for the Exploration of the Sea (ICES, 2001). Hauls of 30-mins duration were carried out using a stratified random design. All fish were weighted and a sub-sample or all were counted. A global abundance index and its variance were calculated using a random stratified estimator taking account of the surface swept by each haul and the total stratum surface.

A classical population dynamics model used in wildlife management is $N(t) = N(t-1) \exp(r(t))$, where $N(t)$ is the population abundance index and $r(t)$ the achieved population growth rate in year t . The model is linearised by log-transformation $\ln(N(t)) = \ln(N(t-1)) + r(t)$. Thus the local slope of the linear regression of $\ln(N(t))$ on time is the annual population growth rate $r(t)$. The indicator considered is then $\ln(N(t))$ and its temporal changes indicate changes in the underlying population dynamics.

The statistical distribution of $\ln(N(t))$ can be found by first considering the distribution of $N(t)$. Trenkel and Rochet (2003) analysing bottom trawl data similar to here found that both lognormal and Gamma distributions were appropriate for describing the distribution of numbers per haul and as a result for the abundance index $N(t)$. This is a general feature in bottom trawl data. Hence taking logarithms will result in $\ln(N(t))$ being approximately normally distributed and justifies assuming normal distributions for the estimates of first and second derivatives of the smoothed indicator series.

Recent trends in $\ln(N(t))$ and changes in trends were tested using the proposed tests. For comparison, linear models were also fitted and two non-parametric Mann-Kendall tests (one for strictly increasing and one for strictly decreasing trends) was carried out. A one-sided test level of $\alpha=0.05$ was used in all cases. To explore the impact of the chosen time horizon, the number of years considered was varied between $n=3$ and $n=5$, corresponding to recent time trends over the period 2004-2006 and 2002-2006 respectively. To investigate the capacity of the proposed method to pick up changes and guide management action from year to year, recent trends were also calculated for a reduced time series obtained by shifting the time series, i.e. considering recent trends during the period 2003-2005.

For 25 out of 33 species did the GAM model provide a satisfactory fit for the data series 1983-2006 (for 28 species for the shorter period 1983-2005), as shown by small p-values (≤ 0.05) for the χ^2 -tests (Table 1). Using the intersection-union test for trends (T1 & T2), 11

species with recent positive and seven with recent negative population growth rates were identified, compared to none for the Mann-Kendall test and one decreasing and one increasing with linear regression. In addition, for *Hyperoplus lanceolatus*, linear regression indicated a significant decreasing trend over the last three years of the series, while no intersection-union test could be carried out for this species due to lack of explanatory power for the GAM model. Visual inspection of figure 1 indicated that the population has probably been stable over the most recent years. If a test level of $\alpha=0.1$ had been used, linear regression would have indicated significant trends for two additional species, the intersection-union test for six additional and the Mann-Kendall test still for none (table 2). The test for changes in trends (T3) showed for one species (*Callionymus lyra*) that although no increasing or decreasing trend was yet detectable, local slopes were decreasing (Table 1, solid arrows).

When increasing the time horizon considered to five years (period 2002-2006), both linear regression and the Mann-Kendall test picked up more changes, one increasing and four decreasing (Table 1). The Mann-Kendall tests indicated three additional decreasing taxa (Ammodytidae, *Limanda limanda* and *Scomber scombrus*); all cases had a single high value in the early part of the recent period. For three of the cases where linear regression and Mann-Kendal test agreed, *Chelidonichtys gurnardus*, *Merlangius merlangus* and *Microstomus kitt*, the intersection-union was not significant, as the null hypothesis T1 (maximum during period) was not rejected. The case of *Scomber scombrus* is the only case with opposite results. The Mann-Kendall test detected a significant increase while the intersection-union test found a decrease. This species is characterised by a particularly large interannual variability in $\ln(N)$ estimates (Figure 1), which seems to be not well represented by the smoothed index. There was no significant consistent acceleration or deceleration over the five year period for any of the species.

The proposed method seemed to be responsive, as when shifting the time series (removing the last year 2006) before fitting GAM models and then testing for trends over the period 2003-2005, the number of species with short term increasing trends was 16 instead of 11, including three species for which no test could be carried out for the later series (Table 2); seven species were found to decrease which were not the same as for the longer period in four cases. Both linear regression and the Mann-Kendall test did not reveal any significant recent trends. The test for changes in trends (T3) revealed a deceleration of an increasing trend for five species. One species with no time trend over the period (*Gasterosteus aculeatus*) showed an increase in the local slopes. This second test for changes in slopes allowed for two species to anticipate subsequent time trends when one year of data was added (period 2004-2006). The increase in slope indicated by the significant test for *Gasterosteus aculeatus* for the period 2003-2005 become a significant increasing trend for the period 2004-2006. Similarly for *Trisopterus minutes*, where a significant increasing trend during the first period turned into a significant decreasing trend one year later, as indicated by the significant deceleration over the first period.

8. Discussion

We proposed and applied a first intersection-union test for detecting and characterising recent time trends and a second test for changes in trends. In contrast to the two classical methods, the proposed approach includes the uncertainty of indicator estimates. For the case study of North Sea fish, the intersection-union test was able to pick up far more changes than both a linear regression and a Mann-Kendall test, in particular for a short time horizon of three years; unsurprisingly the difference was less when the most recent five years were considered. This confirms the well known fact that both linear regression and the Mann-Kendall test have little power to detect trends in short time series. The investigations of trends over five years bring up the question whether for the intersection-union test the time horizon should be the same for both test statistics (T1 and T2) and what time horizon to

choose for the second test (T3). For the location of the maximum (T1) it might be more suitable to consider a shorter period than for local slopes (T2). Thus the choice of n for each sub-test should be carefully considered and adapted to the particular case study keeping in mind that the aim is to test recent trends in the light of the whole available time series. Alternatively one could consider recent values with respect to a reference level derived from additional information. For example, if the indicator of interest was population abundance, some measure of desirable abundance could be used. It might also be informative to compare recent levels to pre-fishing or early-fishing levels or any other benchmark, using for example the approach by Rosenberg et al. (2005) who estimated 19th century cod population levels based on historic catch-per-unit effort data.

The test for changes in recent trends (T3) was intended to allow to determine current changes in underlying dynamics, i.e. acceleration or deceleration, which would become detectable trends only in the future and thus to add a degree of anticipation to the diagnosis as suggested by Buckland et al. (2005). In the application to Southern North Sea fish species, significant results for this second test were mainly found when considering three years time horizons. When comparing changes in trends found for the period 2003-2005 with actual trends for 2004-2006, in two cases (*Trisopterus minutus* and *Gasterosteus aculeatus*) was a significant change in local trends followed by a significant trend. In three other cases (*Agonus cataphractus*, *Arnoglossus laterna* and *Platichthys flesus*), a deceleration in trend was not yet reflected in the smoothed time series when the additional year was added.

When fitting a smoother to the indicator time series, we simultaneously estimated the optimal degree of smoothing. In the past before the methods developed by Wood (2006) were available, many authors have used empirical rules for fixing the degree of smoothing which were selected based on visual (Fewster et al., 2000) or theoretical considerations (Fryer and Nicholson, 1999). We believe that using a statistical criterion for optimal smoothness makes the resulting smoothed time series more objective as it avoids any arbitrary user decisions. Nevertheless, in the presence of large interannual variability, either due to true population variations, e.g. variable recruitment, or high sampling variability, the smoothed series will hopefully reflect the long term trends but might not well represent the most recent years. Indeed, the goodness-of-fit test we carried out indicates the global suitability of the GAM model compared to a simple linear model represented by the mean, not its local fit at the end of the time series. For example, for *Enchelopus cimbrius* it might be argued that the smoothed series did not describe well the time trend in the final years. In contrast, for *Scomber scombrus* or even *Trachurus trachurus* estimated indices varied widely around the smoothed indices (Figure 1). Both species are pelagic and form large schools, hence they are difficult to sample reliably using a bottom trawl as done by the IBTS survey. Thus for these two species it is probably the indices that are unsuitable for a short term analysis rather than an indication of failure of the proposed approach.

We applied the proposed test for recent time trends using a classical false detection rate of $\alpha=0.05$. However, as Field et al. (2004) pointed out, the costs of restoring a deteriorated population compared to taking overcautious measures are unequal. Hence they advocate much higher α -levels in trend tests. Transferring this principle to the test proposed here means that it might be better to detect too many than too few changes, but given the sensitivity of the proposed union-intersection test, using larger α -levels does not seem to be warranted. However, as the p-values for the linear regression tests showed (Table 1), α -levels of more than 0.1 would have to be used to make linear regression as sensitive as the proposed method.

Evaluation guidelines for the suitability of a metric as an ecological indicator often include prescriptions about the required power of the indicator to detect a certain degree of change over a specified number of years, for example a 20% change in ecological condition over a 10-year period with 90% confidence (Kurtz et al., 2001). While measurement of pollutants might be precise and hence the power of monitoring data to detect small changes satisfactorily high, population indicator estimates derived from annual sampling surveys are generally imprecise leading to large interannual variability which can drown trends. Sampling for marine species is particularly affected by this problem, due to large survey areas, small

scale as well as large scale temporal and spatial variability and generally low sample sizes. For example, for population metrics such as mean length in the population, Trenkel and Rochet (2003) found that tests using Western IBTS survey data had a power of 0.8 for detecting a significant linear change over a four year period for 15 out of 25 tested fish populations only. Similarly for community mean length, use of IBTS data provided tests with a power of less than 0.1 to detect a linear 0.2 cm change per year within five years (Nicholson and Jennings, 2004). In contrast, Hughes et al. (1998) found that with their index of biological integrity of a river fish assemblage they could detect a 8% year-to-year difference with a testing power of 0.8. Thus short term linear trend tests have proven to be rather blunt instruments for informing fisheries management decisions. We expect that the proposed tests have higher testing power for the same α -level compared to linear regression, thus making indicators more useful for fisheries management (Rice, 2000; Rice and Rochet, 2005). Incidentally this suggests that the absolute power of an indicator is not a property of the indicator itself but rather depends on the statistical method used for detecting changes over time.

One possible solution to the problem of high signal-to-noise ratios in marine population and community (state) indicators is to use them only for monitoring long term changes, and rely on hopefully more precisely estimated pressure indicators, e.g. fishing effort, and response, e.g. fishing quota, for short term assessment and management as suggested by Jennings (2005). The proposed method could then be applied to detect recent trends in pressure indicators. However, few firm direct links between pressure and response indicators and resulting states have so far been established (but see Link et al. 2002). Thus in order to assess how the exploited ecosystem is doing, it still seems a good idea to monitor short term changes of the system as well. The capacity to take timely management actions based on population indicators (Trenkel et al., 2007) crucially depends on the availability of methods for rapidly and reliably detecting short term changes. The proposed tests hopefully allow to fill this gap in the field of fisheries management but also in other fields where indicators are used for monitoring and management. In other fields data collection might be on longer or shorter time scales than the annual level in the case study presented here, which changes the meaning of short term trends but not the applicability of the method.

Acknowledgement

This study received financial support from the EU funded projects FISBOAT (FP5-502572) and IMAGE (FP6 – 044227) and the project CHALOUPE funded by the French Agence Nationale de la Recherche. We thank all survey staff for collecting the data and ICES for making them available in the DATRAS database.

References

- Buckland, S.T., Magurran, A.E., Green, R.E., Fewster, R.M., 2005. Monitoring change in biodiversity through composite indices. *Philosophical Transactions of the Royal Society London Series B* 360, 243-254.
- Caddy, J.F., Surette, T., 2005. In retrospect the assumption of sustainability for Atlantic fisheries has proved an illusion. *Reviews in Fish Biology and Fisheries* 15, 313-337.
- Casella, G., Berger, R.L., 1990. *Statistical inference*. Duxbury Press, Belmont, California, p. 650.
- Champely, S., Doledec, S., 1997. How to separate long-term trends from periodic variation in water quality monitoring. *Water Research* 31, 2849-2857.
- Conover, W.J., 1971. *Practical nonparametric statistics*. John Wiley & Sons, New York
- Dulvy, N.K., Jennings, S., Goodwin, N.B., Grant, A., Reynolds, J.D., 2005. Comparison of threat and exploitation status in North-East Atlantic marine populations. *Journal of Applied Ecology* 42, 883-891.

- Fewster, R.M., Buckland, S.T., Siriwardena, G.M., Baillie, S.R., Wilson, J.D., 2000. Analysis of population trends for farmland birds using generalized additive models. *Ecology* 81, 1970-1984.
- Field, S.A., Tyre, A.J., Jonzén, N., Rhodes, J.R., Possingham, P., 2004. Minimizing the cost of environmental management decisions by optimizing statistical threshold. *Ecology Letters* 7, 669-675.
- Fryer, R.J., Nicholson, M.D., 1999. Using smoothers for comprehensive assessments of contaminant time series in marine biota. *ICES Journal of Marine Science* 56, 779-790.
- Hughes, R.M., Kaufmann, P.R., Herlihy, A.T., Kincaid, T.M., Reynolds, L., Larsen, D.P., 1998. A process for developing and evaluating indices of fish assemblage integrity. *Canadian Journal of Fisheries and Aquatic Sciences* 55.
- ICES, 2001. Development of a central database for European trawl survey data. DATRAS Database TRAWL Surveys.
- Jennings, S., 2005. Indicators to support an ecosystem approach to fisheries. *Fish and Fisheries* 6, 212-232.
- Jennings, S., Dulvy, K., 2005. Reference points and reference directions for size-based indicators of community structure. *ICES Journal of Marine Science* 62, 397-404.
- Kurtz, J.C., Jackson, L.E., Fisher, W.S., 2001. Strategies for evaluating indicators based on guidelines from the Environmental Protection Agency's Office of research and development. *Ecological Indicators* 1, 49-60.
- Link, J.S., Brodziak, J.K.T., Edwards, S.F., Overholtz, W.J., Mountain, D., Jossi, J.W., Smith, T.D., Fogarty, M.J., 2002. Marine ecosystem assessment in a fisheries management context. *Canadian Journal of Fisheries and Aquatic Sciences* 59, 1429-1440.
- McLeod, A.I., Hipel, K.W., Bodo, B.A., 1991. Trend analysis methodology for water quality time series. *1991* 2, 169-200.
- Nicholson, M., Fryer, R., Maxwell, D., 1998. Multivariate trends in phytoplankton species groups in the southern North Sea. *ICES Journal of Marine Science* 55, 581-586.
- Nicholson, M.D., 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.
- Poulard, J.-C., Trenkel, V.M., 2007. Do survey design and wind conditions influence survey indices ? *Canadian Journal of Fisheries and Aquatic Sciences* 64, 1551-1562.
- R Development Core Team, 2007. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Rice, J.C., 2000. Evaluating fishery impacts using metrics of community structure. *ICES Journal of Marine Science* 57, 682-688.
- Rice, J.C., Rochet, M.-J., 2005. A framework for selecting a suite of indicators for fisheries management. *ICES Journal of Marine Science* 62, 516-527.
- Rochet, M.-J., Trenkel, V.M., Bellail, R., Coppin, F., Le Pape, O., Mahé, J.-C., Morin, J., Poulard, J.-C., Schlaich, I., Souplet, A., Vérin, Y., Bertrand, J., 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., 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.
- Rosenberg, A.A., Bolster, W.J., Alexander, K.E., Leavenworth, W.B., Cooper, A.B., McKenzie, M.G., 2005. The history of ocean resources: modeling cod biomass using historical records. *Frontiers in Ecology and the Environment* 3, 84-90.
- Trenkel, V.M., 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 and Aquatic Sciences* 60, 67-85.

- Trenkel, V.M., Rochet, M.-J., Mesnil, B., 2007. From model-based prescriptive advice to indicator-based interactive advice. *ICES Journal of Marine Science* 64, 768-774.
- van Densen, L.T., 2001. On the perception of time trends in resource outcome. Its importance in fisheries co-management, agriculture and whaling. PhD. Twente University, Enschede, the Netherlands, p. 299.
- Wood, S.N., 2006. Generalized additive models. An introduction with R. Chapman & Hall/CRC, Boca Raton, p. 391.

Tables

Table 1. Test results for recent trends and changes in trends over a three or five year period in population growth rates in Southern North Sea fish ($\alpha=0.05$) using linear regression (LR), Mann-Kendall test (MK) and intersection-union test (IU). p-values for χ^2 -test for GAM model fit to entire time series and for linear regression. Significant time trends (T1 & T2): decreasing (\searrow), increasing (\nearrow) or stable. Changes in time trends (T3): decelerating trend (d) or negative second derivatives (\swarrow); positive derivatives (\nearrow).

Scientific name	1984-2006	2004-2006				2002-2006				1984-2005	2003-2005			
	Chi2 P-value	LR P-value	LR	MK	IU	LR P-value	LR	MK	IU	Chi2 P-value	LR P-value	LR	MK	IU
<i>Agonus cataphractus</i>	0.001	0.06	\leftrightarrow	\leftrightarrow	\nearrow	0.52	\leftrightarrow	\leftrightarrow	\nearrow	0.001	0.45	\leftrightarrow	\leftrightarrow	\searrow d
<i>Arnoglossus laterna</i>	0.051	0.51	\leftrightarrow	\leftrightarrow		0.15	\leftrightarrow	\leftrightarrow		0.049	0.18	\leftrightarrow	\leftrightarrow	\searrow d
<i>Buglossidium luteum</i>	0.039	0.64	\leftrightarrow	\leftrightarrow	\nearrow	0.14	\leftrightarrow	\leftrightarrow	\nearrow	0.040	0.6	\leftrightarrow	\leftrightarrow	\searrow d
<i>Callionymus lyra</i>	<0.001	0.92	\leftrightarrow	\leftrightarrow	\swarrow	0.81	\leftrightarrow	\leftrightarrow	\leftrightarrow	<0.001	0.66	\leftrightarrow	\leftrightarrow	\leftrightarrow
<i>Chelidonichthys gurnardus</i>	0.001	0.07	\leftrightarrow	\leftrightarrow	\leftrightarrow	0.03	\nearrow	\nearrow	\leftrightarrow	0.002	0.28	\leftrightarrow	\leftrightarrow	\leftrightarrow
<i>Clupea harengus</i>	<0.001	0.42	\leftrightarrow	\leftrightarrow	\searrow	0.01	\searrow	\searrow	\searrow	<0.001	0.17	\leftrightarrow	\leftrightarrow	\searrow
<i>Echiichthys vipera</i>	0.052	0.40	\leftrightarrow	\leftrightarrow		0.24	\leftrightarrow	\leftrightarrow		0.020	0.068	\leftrightarrow	\leftrightarrow	\nearrow
<i>Enchelyopus cimbrius</i>	0.002	0.19	\leftrightarrow	\leftrightarrow	\nearrow	0.35	\leftrightarrow	\leftrightarrow	\nearrow	0.002	0.81	\leftrightarrow	\leftrightarrow	\nearrow
<i>Engraulis encrasicolus</i>	<0.001	0.31	\leftrightarrow	\leftrightarrow	\nearrow	0.42	\leftrightarrow	\leftrightarrow	\nearrow	<0.001	0.96	\leftrightarrow	\leftrightarrow	\nearrow
<i>Ammodytidae</i>	<0.001	0.96	\leftrightarrow	\leftrightarrow	\searrow	0.11	\leftrightarrow	\searrow	\searrow	<0.001	0.96	\leftrightarrow	\leftrightarrow	\searrow
<i>Gobiidae</i>	0.07	0.66	\leftrightarrow	\leftrightarrow		0.12	\leftrightarrow	\leftrightarrow		0.071	0.67	\leftrightarrow	\leftrightarrow	
<i>Gadus morhua</i>	<0.001	0.68	\leftrightarrow	\leftrightarrow	\searrow	0.39	\leftrightarrow	\leftrightarrow	\searrow	<0.001	0.81	\leftrightarrow	\leftrightarrow	\searrow
<i>Gasterosteus aculeatus</i>	0.002	0.64	\leftrightarrow	\leftrightarrow	\nearrow	0.1	\leftrightarrow	\leftrightarrow	\leftrightarrow	0.002	0.34	\leftrightarrow	\leftrightarrow	\leftrightarrow \nearrow
<i>Hippoglossoides platessoides</i>	<0.001	0.32	\leftrightarrow	\leftrightarrow	\searrow	0.045	\searrow	\searrow	\searrow	<0.001	0.31	\leftrightarrow	\leftrightarrow	\nearrow
<i>Hyperoplus lanceolatus</i>	0.07	0.002	\searrow	\leftrightarrow		0.19	\leftrightarrow	\leftrightarrow		0.073	0.82	\leftrightarrow	\leftrightarrow	
<i>Limanda limanda</i>	<0.001	0.69	\leftrightarrow	\leftrightarrow	\leftrightarrow	0.09	\leftrightarrow	\searrow	\leftrightarrow	<0.001	0.06	\leftrightarrow	\leftrightarrow	\leftrightarrow
<i>Liparis liparis</i>	0.02	0.22	\leftrightarrow	\leftrightarrow	\nearrow	0.36	\leftrightarrow	\leftrightarrow	\nearrow	0.023	0.52	\leftrightarrow	\leftrightarrow	\nearrow
<i>Melanogrammus aeglefinus</i>	0.06	0.55	\leftrightarrow	\leftrightarrow		0.94	\leftrightarrow	\leftrightarrow		0.027	0.06	\leftrightarrow	\leftrightarrow	\searrow
<i>Merlangius merlangus</i>	<0.001	0.03	\leftrightarrow	\leftrightarrow	\leftrightarrow	0.003	\searrow	\searrow	\leftrightarrow	<0.001	0.11	\leftrightarrow	\leftrightarrow	\searrow
<i>Microstomus kitt</i>	<0.001	0.06	\leftrightarrow	\leftrightarrow	\searrow	0.017	\searrow	\searrow	\leftrightarrow	<0.001	0.10	\leftrightarrow	\leftrightarrow	\leftrightarrow
<i>Myoxocephalus scorpius</i>	0.02	0.16	\leftrightarrow	\leftrightarrow	\nearrow	0.46	\leftrightarrow	\leftrightarrow	\nearrow	0.022	0.31	\leftrightarrow	\leftrightarrow	\nearrow
<i>Platichthys flesus</i>	<0.001	0.8	\leftrightarrow	\leftrightarrow	\nearrow	0.89	\leftrightarrow	\leftrightarrow	\leftrightarrow	<0.001	0.27	\leftrightarrow	\leftrightarrow	\searrow d
<i>Pleuronectes platessa</i>	<0.001	0.30	\leftrightarrow	\leftrightarrow	\leftrightarrow	0.45	\leftrightarrow	\leftrightarrow	\leftrightarrow	<0.001	0.35	\leftrightarrow	\leftrightarrow	\searrow
<i>Pomatoschistus spp</i>	0.08	0.45	\leftrightarrow	\leftrightarrow		0.26	\leftrightarrow	\leftrightarrow		0.068	0.44	\leftrightarrow	\leftrightarrow	
<i>Raja clavata</i>	0.062	0.14	\leftrightarrow	\leftrightarrow		0.83	\leftrightarrow	\leftrightarrow		0.064	0.99	\leftrightarrow	\leftrightarrow	
<i>Scomber scombrus</i>	<0.001	0.25	\leftrightarrow	\leftrightarrow	\nearrow	0.19	\leftrightarrow	\searrow	\nearrow	<0.001	0.25	\leftrightarrow	\leftrightarrow	\nearrow
<i>Scyliorhinus canicula</i>	0.01	0.03	\nearrow	\leftrightarrow	\leftrightarrow	0.59	\leftrightarrow	\leftrightarrow	\leftrightarrow	0.013	0.75	\leftrightarrow	\leftrightarrow	\leftrightarrow
<i>Solea solea</i>	0.005	0.93	\leftrightarrow	\leftrightarrow	\nearrow	0.76	\leftrightarrow	\leftrightarrow	\nearrow	0.007	0.69	\leftrightarrow	\leftrightarrow	\nearrow
<i>Sprattus sprattus</i>	0.004	0.67	\leftrightarrow	\leftrightarrow	\searrow	0.42	\leftrightarrow	\leftrightarrow	\searrow	0.004	0.4	\leftrightarrow	\leftrightarrow	\nearrow
<i>Trachurus trachurus</i>	<0.001	0.8	\leftrightarrow	\leftrightarrow	\nearrow	0.324	\leftrightarrow	\leftrightarrow	\nearrow	<0.001	0.78	\leftrightarrow	\leftrightarrow	\nearrow
<i>Trisopterus esmarkii</i>	0.027	0.31	\leftrightarrow	\leftrightarrow	\leftrightarrow	0.55	\leftrightarrow	\leftrightarrow	\leftrightarrow	0.014	0.39	\leftrightarrow	\leftrightarrow	\searrow
<i>Trisopterus luscus</i>	0.058	0.87	\leftrightarrow	\leftrightarrow		0.94	\leftrightarrow	\leftrightarrow		0.061	0.16	\leftrightarrow	\leftrightarrow	
<i>Trisopterus minutus</i>	0.005	0.68	\leftrightarrow	\leftrightarrow	\searrow	0.44	\leftrightarrow	\leftrightarrow	\searrow	0.005	0.069	\leftrightarrow	\leftrightarrow	\searrow d

Table 2. Number of species for recent trend tests over a three or five year period in population growth rates in Southern North Sea fish using linear regression (LR), Mann-Kendall test (MK) and intersection-union test (IU) and test level α . Direction of significant time trends: decreasing (\searrow), increasing (\nearrow) or stable (\leftrightarrow).

Trend direction	2004-2006 $\alpha=0.05$			2004-2006 $\alpha=0.1$			2003-2005 $\alpha=0.05$			2002-2006 $\alpha=0.05$		
	LR	MK	IU	LR	MK	IU	LR	MK	IU	LR	MK	IU
\searrow	2	0	7	4	0	12	0	0	7	4	7	6
\leftrightarrow	30	33	7	28	33	1	33	33	5	28	25	10
\nearrow	1	0	11	1	0	12	0	0	16	1	1	9

Figures

Figure 1. Time series of ln-abundance of some fish species in the Southern North Sea based on IBTS bottom trawl data. The continuous line is a generalised additive model (GAM) fit. The broken lines are 95% confidence bands for this fit based on a parametric bootstrap of annual indicator estimators. Full species names in table 1.

Figure 1

