

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

# Ecological Indicators

journal homepage: [www.elsevier.com/locate/ecolind](http://www.elsevier.com/locate/ecolind)

## Analyses of structural changes in ecological time series (ASCETS)

Örjan Östman<sup>a,\*</sup>, Lena Bergström<sup>a</sup>, Kjell Leonardsson<sup>b</sup>, Anna Gårdmark<sup>a</sup>, Michele Casini<sup>c,d</sup>, Ylva Sjöblom<sup>a</sup>, Fredrik Haas<sup>e</sup>, Jens Olsson<sup>a</sup>

<sup>a</sup> Dept. Aquatic Resources, Swedish University of Agricultural Sciences, Skolgatan 6, S-742 42 Öregrund, Sweden

<sup>b</sup> Dept. Wildlife, Fish and Environmental Studies, Swedish University of Agricultural Sciences, S-901 83 Umeå, Sweden

<sup>c</sup> Dept. Aquatic Resources, Swedish University of Agricultural Sciences, Turistgatan 5, S-453 30 Lysekil, Sweden

<sup>d</sup> Department of Biological, Geological and Environmental Sciences, University of Bologna, Via Selmi 3, 40126 Bologna, Italy

<sup>e</sup> Dept. of Biology, Biodiversity and Conservation Science Lund University, Sölvegatan 37, S-223 62 Lund, Sweden



### ARTICLE INFO

#### Keywords:

Assessments  
Confidence  
Ecosystem based management  
Indicators  
Integrated ecosystem approach  
Risk analysis

### ABSTRACT

Assessing status of natural resources and ecosystem components is pivotal for management, where indicators or indices often are used as proxies of ecological state. Many indicators, however, lack reference points and are associated with sampling errors and environmental noise, limiting their usefulness in management. Here we present a method for assessing state changes in ecological indicator from time-series: *Analyses of Structural Changes in Ecological Time Series (ASCETS)*. ASCETS enables both quantitative boundary levels for changes in indicator states (e.g. for management targets), and the confidence for a change in state during an assessment period. Thereby it can be used in risk assessments and is suitable for aggregation or integration of different indicator states across sites, or for an ecosystem based approach to management. With extended information about ecological state during a reference period, ASCETS can support reference levels for defining ecological status of an indicator. ASCETS first identifies structural changes in time-series to determine reference periods with coherent indicator dynamics. Next, from the observed indicator values during the reference period, a distribution of resampled median values is used to set boundary levels as a tolerable range of indicator variation reflecting the same state as during the reference period. Finally, a confidence of a change in indicator state is evaluated during an assessment period as the proportion of resampled median values of the assessment period overlapping the boundary levels of the reference period. Simulations indicate ASCETS correctly detects changes in indicator state when changes in indicator values are at least twice as large as the coefficient of variation, with a false rate of changes around 5%. We apply ASCETS to indicators for bird and fish communities used within the Marine Strategy Framework Directive to illustrate how indicator boundary levels can be set where reference levels may be ecologically and analytically troublesome. An R-script is provided for further use and modification. We propose ASCETS as a flexible and generic method for assessing changes in ecological states from time-series to support identification of management targets.

### 1. Introduction

Reference points or levels for assessing the state of ecosystem components are important for environmental and natural resources management. Assessments using analytical reference points, such as Biomass at Maximum Sustainable Yield ( $B_{MSY}$ ), or Minimum Viable Population Size (MVP) require high availability and quality of data, and are currently possible only for a small fraction of species, typically those most exploited or endangered (Kempf, 2010; Flather et al., 2011; Borja et al., 2013). To support an ecosystem-based management it is necessary to assess include a wider range of ecosystem components in

assessments. This requires developing assessments of data-poor species and ecosystem components, including setting corresponding boundary levels and management targets (Gregory et al., 2005; Shin & Shannon, 2010; Borja et al., 2013).

Assessments of data-poor species and ecosystem components typically involve use of different indicators or indices, like indicators of species abundance, reproduction success, size/age distributions, trophic structure, and functional diversity. However, the lack of quantitative reference points limits the usefulness of such indicators in management (Borja et al., 2012; Samhoury et al., 2012). Three approaches have been proposed to set boundary levels of indicators (reviewed in Samhoury

\* Corresponding author.

E-mail addresses: [orjan.ostman@slu.se](mailto:orjan.ostman@slu.se) (Ö. Östman), [lena.bergstrom@slu.se](mailto:lena.bergstrom@slu.se) (L. Bergström), [kjell.leonardsson@slu.se](mailto:kjell.leonardsson@slu.se) (K. Leonardsson), [anna.gardmark@slu.se](mailto:anna.gardmark@slu.se) (A. Gårdmark), [michele.casini@slu.se](mailto:michele.casini@slu.se) (M. Casini), [fredrik.haas@biol.lu.se](mailto:fredrik.haas@biol.lu.se) (F. Haas), [jens.olsson@slu.se](mailto:jens.olsson@slu.se) (J. Olsson).

<https://doi.org/10.1016/j.ecolind.2020.106469>

Received 10 October 2019; Received in revised form 6 April 2020; Accepted 25 April 2020

Available online 07 May 2020

1470-160X/ © 2020 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

et al., 2012; Borja et al., 2013): 1) *Functional indicator-pressure relationships* where the state of the indicator can be linked to a known pressure level, 2) *Spatial comparisons* where boundary levels from pristine or less disturbed areas can be applied to other areas (borrowing), or 3) *Historic approach* where time-series data can be compared to an “internal standard” (Samhouri et al., 2012). The *Historic approach* can be divided into the ‘reference period’ and the ‘moving window’-approaches (Samhouri et al., 2012). The reference period approach focuses on current state relative to a previous (reference) period. It is useful when a desired or unwanted state is known, or at least assumed, to have occurred based on prior knowledge (Modica et al., 2014). The moving window approach addresses short-term fluctuations (Samhouri et al., 2012), so that management boundaries are set by comparing the current state of an indicator relative to a time lag. Nominal reference levels will change over time and this approach is common in advice and assessment of for example data poor fish stocks (ICES, 2012).

All sampling of ecological indicators is associated with sampling uncertainty, i.e., observation errors, which is more pronounced in cases when data is poor or scarce (Thorpe et al., 2015). Indicator levels are also affected by natural processes (process errors), e.g. internal dynamics and variations in abiotic drivers (Samhouri et al., 2012). It is therefore necessary to encompass natural variation and observation errors when setting boundary levels for management, as well as to provide an estimate of the confidence or certainty of the assessment (Greenstreet et al., 2012; Thorpe et al., 2015; Soldaat et al., 2017).

Here we outline a method for assessing changes in the state of ecological indicators from time-series data, *Analyses of Structural Changes in Ecological Time Series (ASCETS)*, encompassing both sampling and process errors for changes in indicator state for any time-series, and we test its performance using simulated data. The method is based on breakpoint analysis (Zeileis et al., 2002) and resampling of observed indicator values (Wolodzko, 2018) during a reference period that provides quantitative boundary levels of the indicator state. Based on indicator values during an assessment period, *ASCETS* assesses both changes in state and the confidence of a change, which can be used for risk assessments. *ASCETS* also allows for an integration and aggregation of indicator states across different indicators based on the estimated confidence of each indicator assessment. This enables a unified assessment of indicators derived from different types of data sources, as well as combining indicator assessments into an ‘*Integrated status assessment*’ (Borja et al., 2010).

Assessing the status of ecological indicators is required in EU’s Water Framework Directive (WFD) and Marine Strategy Framework Directive (MSFD) (EC 2000, 2008), but challenging when analytical assessments are not available. We illustrate the applicability of *ASCETS* to address changes in indicator state of fish and waterbird indicators used in the MSFD. We provide the R-script (Supplementary Material) for *ASCETS* to be applied to other data, and foresee further modifications and input to evaluate and develop the framework further.

## 2. Material and methods

The first step of *ASCETS* is a breakpoint analysis to identify periods with coherent indicator dynamics. Second, from a reference period with coherent dynamics, a smoothed median distribution is derived from the observed values with added random error (Fig. 1). The percentiles from this distribution are then used as boundary levels of indicator state during reference period. Finally, if there is an assessment period, the smoothed median distribution of observed values during the assessment period can be used to assess changes in indicator state (Fig. 1). The confidence for that indicator state is the same during reference- and assessment periods is estimated as the proportional overlap of smoothed median distribution of the assessment period within boundary levels derived from the reference period.

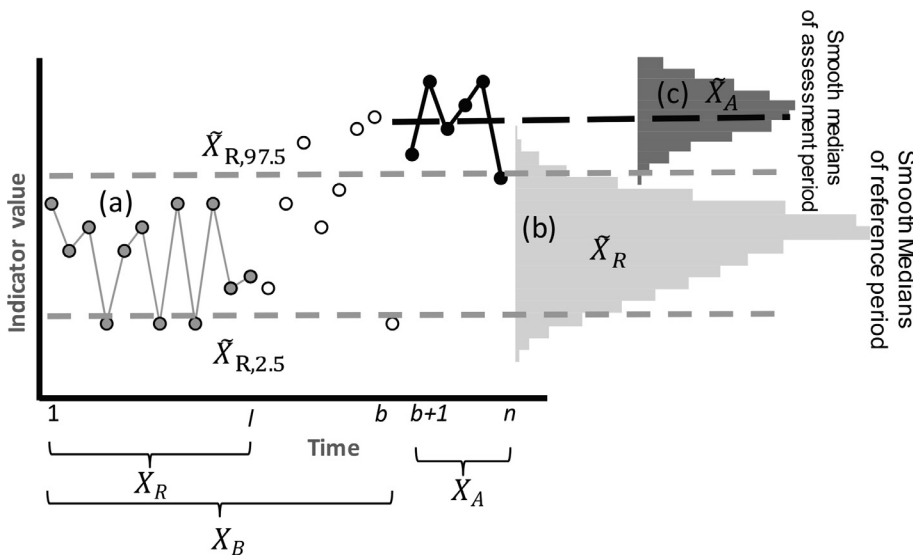
### 2.1. Breakpoint analysis

*ASCETS* can be applied to any time-series,  $X = \{X_1, \dots, X_n\}$ , where  $X_i$  can be either a univariate indicator or a composite index (from hereon we use the term indicator).  $X$  is divided into a baseline period  $X_B = \{X_1, \dots, X_b\}$  and an assessment period  $X_A = \{X_{b+k}, \dots, X_n\}$ , where  $k \geq 1$  is the lag between baseline and assessment periods (Fig. 1a). The division of  $X$  into  $X_B$  and  $X_A$  can either be set *a priori*, as in the MSFD (EC, 2008), or based on data through a breakpoint analysis (Fig. 2). Because structural changes in the indicator values may occur within a predefined baseline period, we define the reference period  $X_R = \{X_{B,a}, \dots, X_{B,l}\}$  as a period with coherent dynamics within the baseline period  $X_B$  where  $a$  is the first and  $l$  is the last observation of the reference period. If there are no structural changes during the baseline period, then  $X_R = X_B$  (Fig. 2). The number of observations in baseline period is  $N_B = \text{length}(X_B)$ , and  $N_R = \text{length}(X_R)$  ( $N_R \leq N_B$ ) is the number of observations in the reference period.  $N_A = \text{length}(X_A)$  is the number of observations in the assessment period. The sampling frequency in  $X$  should be similar over time to not affect temporal autocorrelation between indicator values. If  $X_B$  and  $X_A$  are predefined,  $N_R \geq N_A$  but we recommend  $N_R \geq 2N_A$  to cater for natural variation due to demographic and environmental stochasticity. This version of *ASCETS* cannot handle replicated measures from the same time point, and any multiple samples are therefore combined into an aggregated value. Missing values are removed but not adjusted for. Too many missing values may bias the results, especially if there is autocorrelation in  $X$  and there are several adjacent missing values.

*ASCETS* assumes a reference period with stationary dynamics of the indicator. To investigate the occurrence of structural changes and breakpoints in the dynamics of  $X$ , or  $X_B$  if using a predefined baseline period (Fig. 2), we use the R-package ‘*struchange*’ (Zeileis et al., 2002, 2003). Here we use the conventional probability  $P_{crit} = 0.05$  for identifying a breakpoint, but lower or higher values may be desirable in other situations. To prevent temporal autocorrelative processes, such as cohort-dynamics or density dependence (Östman et al., 2017), from elevating the likelihood to identify breakpoints we use the ‘*ar*’-function in the ‘*stats*’-package to discriminate between an ordinary intercept model and an intercept model with autocorrelation lag 1, i.e., an AR(1) term (see Östman et al., 2017). We use the model with lowest Akaike Information Criteria (AIC-value as it implies a more parsimonious model. The ‘*sctest*’-function is used to calculate the probability for a structural change,  $P_{break}$ , in an empirical fluctuation process of a linear regression model (Zeileis et al., 2003). A structural change in  $X_B$  is revealed if  $P_{break} < P_{crit}$ . Then, we use the “AMOC”-method in the ‘*cpt.mean*’-function in R to identify a single breakpoint (Zeileis et al., 2003, Supplementary Material Fig. A1). We identify only one breakpoint in order to avoid excessive influence of measurement errors and environmental stochasticity. In cases of long time-series (> 30 observations) a method identifying several breakpoints could be considered (e.g., “PELT”, Zeileis et al., 2002).

If  $P_{break} < P_{crit}$ , a sub-period of  $X_B$  should be used as reference period  $X_R$ . *ASCETS* does not identify which sub-period to use and must be judged by the user. It is preferred that  $X_R$  is the longest sub-period of  $X_B$  (Fig. 2). The shorter sub-period could be used instead, for example if there is better prior knowledge on environmental conditions or indicator status from this period. If  $N_B < 2N_A - 1$  there is a risk that  $N_R < N_A$ , which is not feasible condition, and the breakpoint analysis should be suppressed.

In case of a temporal trend in  $X_B$ , the identified breakpoint will split  $X_B$  into sub-periods with different phases of low/high values, and one of the sub-periods can be used as reference period  $X_R$ . In practice, this entails that the indicator state is assessed in relation to a previously occurring phase. This allows an operational assessments to be carried out despite the presence of temporal trends.



**Fig. 1.** Illustration of ‘Analyses of Structural Changes in Ecological Time Series’ (ASCETS): (a) A time-series is divided into a baseline ( $X_B$ , grey and open circles) and an assessment period ( $X_A$ , black circles). Only indicator values during a stationary phase of the baseline period are used as the reference period ( $X_R$ , grey circles) for calculating (b) a bootstrapped smoothed median distribution  $X_R$  (light grey bars). Boundary levels  $X_{R,\alpha}$  for changed status of an indicator is here given by the 95% tolerance interval  $X_{R,2.5}$  and  $X_{R,97.5}$  (hatched grey lines) of  $X_R$ . (c) The proportion of smoothed medians of indicator values during the assessment period ( $X_A$ , dark grey bars) that is within  $X_{R,\alpha}$  (i.e. no change of status) is used as a basis for estimating confidence of a change in indicator status between reference and assessment periods.

**2.2. Calculating boundary levels from reference period**

To account for observed variation in  $X_R$  when calculating boundary levels, ASCETS bootstraps a distribution of median values,  $X_R$ , from  $X_R$  to obtain a parameter range of the indicator state during the reference period (Fig. 1b). As  $N_R$  is preferably larger than  $N_A$ ,  $X_R$  is resampled with  $N_A$  random observations. This ensures that short term fluctuations in indicator values during the assessment period are accounted for in the same way as in the reference period.

Let  $x_i$  be the median of  $N_A$  observed indicator values that are randomly picked with replacement from  $X_R$ . A smoothed distribution for  $x_i$  is derived from a rectangular smoothing distribution using the R-package ‘kernelboot’ (Wolodzko, 2018) with  $T$  (e.g., ten) median values from each random subsample of  $X_R$ . The rectangular smoothing distribution is motivated by that  $N_A$  is typically small and a flat distribution can be assumed.  $x_i$  is resampled  $n$  times to get  $[x_1, \dots, x_{T*n}]$  median values. We recommend  $n \geq 1000$ .

In a smoothed bootstrapped distribution, the bandwidth,  $bw$ , is a free parameter influencing the kernel density (Wolodzko, 2018). As  $bw$  cannot be determined from other parameters, we generate several different  $X_{R,bw}$  distributions using different  $bw = sc^* \epsilon$ , where  $sc$  is a scaling coefficient  $> 0$  and  $\epsilon$  is the Median Absolute Deviation of  $X_R$ . However,  $sc$  must be determined by visual inspection (Supplementary Material Figs. A1, A2); it should be as low as possible to avoid overrepresentation of rare events (fat tails) but large enough to avoid “spikes” in the kernel density (Supplementary Material Fig. A2). We here use the geometric series  $sc = 2^G$ , and  $G = [-4, 4]$  which covers relevant parameter space in our datasets (Supplementary Material Fig. A2). Other dataset may require other parameter space, but  $sc$  should cover from spiky to smooth distributions. By visual inspection of  $X_{R,bw}$  we choose the  $sc$  with the lowest value still providing a smooth kernel distribution and use this  $sc$  for  $X_R = X_{R,bw}$ . Note that  $X_R$  does not have to be unimodal as also multimodal distributions, anticipated for example in case of cohort- or cyclic dynamics, are possible. In an evaluation of different  $sc$  in one of our data sets,  $sc$  had low influence on the confidence intervals relative the uncertainty in the data (Supplementary Material Table A1).

Based on  $X_R$ , ASCETS calculates the boundary levels,  $\hat{X}_i$ , as the tolerance level(s) of  $X_R$ :  $\hat{X}_i = X_{R,\alpha}$  where  $\alpha$  is the tolerance limit of  $X_R$ , i.e., the  $\alpha^{th}$  percentile of  $X_R$  (Fig. 1b, 2; Supplementary Material Table A1). By convention in ecological research,  $p = 0.05$  is often used for statistical significant changes, which corresponds to  $\alpha_1 = 0.025$  and  $\alpha_2 = 0.975$  for a two-sided test of change in indicator status. If one-sided tests or some other level of tolerance of how much indicator

values must differ from the reference period to represent a change in indicator status,  $\alpha$  should change accordingly.  $\alpha$  closer to zero or one means that boundary levels must change more relative the reference period for a change in indicator status to be inferred (Fig. 2).

In cases when indicators show significant temporal autocorrelation during  $X_R$ ,  $X_R$  could be generated with a sliding window of  $N_A$  observations from  $X_R$ . Then,  $x_1, \dots, x_n$  will be  $n$  resampled medians of  $N_A$  randomly picked observed indicator values, with replacement, from the  $N_A$  first observations of  $X_R$ . The procedure is repeated for the next sliding window  $X_R(2: N_A + 1)$ , and so on, until  $X_R(N_R - N_A + 1: N_R)$  (in total  $U = N_R - N_A + 1$  sliding windows over  $X_R$ ) so the resampled smoothed median distribution is  $X_R = [x_1, \dots, x_{U*n}]$ .

**2.3. Confidence of changes in indicator state**

In order to account for observation errors and variability of indicator values also during the assessment period we propose to use the same resampling procedure for  $X_A$  as for  $X_R$  (Fig. 1c, 2). As ASCETS addresses the indicator state during the whole assessment period, temporal pattern during the assessment period is ignored. A distribution of resampled  $X_A$  of  $X_A$  is calculated by  $X_{A,t} = median(resample(X_A)^t)$ , resampled, for example,  $t = 1000$  times with the same  $bw$ , i.e. using the same  $sc^* \epsilon$ , as for  $X_R$  (Fig. 1c). The proportion of  $X_{A,t}$  within the parameter space ( $X_{R,\alpha_1}, X_{R,\alpha_2}$ ) is used as the estimate of the confidence of indicator values during assessment period belonging to indicator state of the reference period ( $X_A \in X_R$ ), denoted  $C(S)$  (Figs. 1 and 2). If  $X_{A,t}$  is completely within ( $X_{R,\alpha_1}, X_{R,\alpha_2}$ ),  $C(S) = 1$ , there is high confidence of no change in indicator state. If  $X_{A,t}$  is completely outside ( $X_{R,\alpha_1}, X_{R,\alpha_2}$ ),  $C(S) = 0$ , meaning there is high confidence of a change in indicator state. Hence, if  $C(S) < 0.5$  a change in state is more likely than not but less confident closer to 0.5, whereas if  $C(S) > 0.5$  a change in state is less likely than not but less confident closer to 0.5.

The variation of biological processes often scales with the mean (Taylor, 1961). As  $N_A$  often is low, the variation estimation would become uncertain and we prefer to use the same  $\epsilon$  to calculate  $X_A$  as for  $X_R$ , even when  $X_A \neq X_R$ . That means a possible overdispersion of  $X_A$  when  $X_A < X_R$ , and under-dispersion when  $X_A > X_R$ , which can affect the tails of the  $X_A$  distribution, and hence  $C(S)$ .

We denote  $\hat{C}(S)$  as the aggregated confidence of change in indicator state. For example, if  $C(S)$  for three different indicators are 0.8, 0.1 and 0.5, an arithmetic mean  $\hat{C}(S) = (0.8 + 0.1 + 0.5)/3 = 0.47$ . The value of  $\hat{C}(S)$  can be interpreted in the same way as  $C(S)$ .

Note that  $\hat{C}(S)$  does not provide any information on direction of

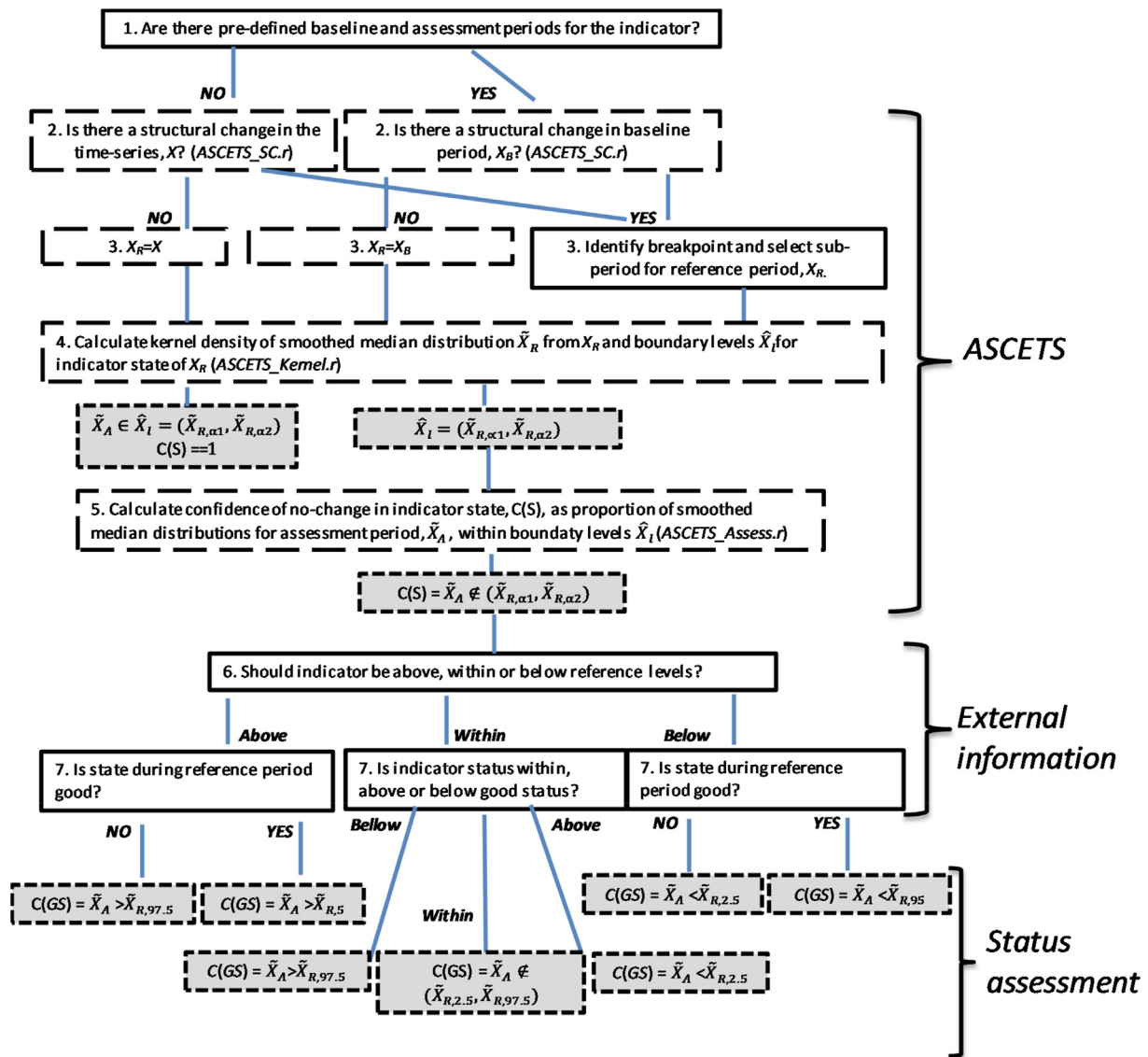


Fig. 2. Flow chart of ASCETS. Solid boxes indicate decision steps by the user, i.e. outside ASCETS, hatched boxes computer steps done by ASCETS, and grey boxes model outcomes.  $X_A$  is the distribution of bootstrapped smoothed medians during assessment period and  $X_{R,\alpha}$  is the  $\alpha$  percentile of  $X_R$  providing boundary levels  $\hat{X}_l$  of indicator state. Note that we have here only indicated potential  $\alpha$  for reference levels, other boundaries can be applied to represent the desired statistical certainty. The R-code for each step is indicated in italics.

change. If one indicator has increasing values and another is declining  $\hat{C}(S) = 0$ . This can be adjusted by for example using  $1 - C(S)$  when  $X_A > X_R$  if increasing indicator values are of no concern, or vice versa.

Other aggregation methods are possible. A geometric mean puts higher weight on  $C(S)$  close to zero (one  $C(S) = 0$  causes  $\hat{C}(S) = 0$ ). Weighted or stratified averages could up-weight results for specific indicators, such as those covering larger spatial ranges, assessed based on data of higher quality, or reflecting prioritized aspects. If  $\hat{C}(S)$  is the aggregated confidence of many ( $> 20$ ) indicator states, some will by chance show change in indicator state, why for example  $\hat{C}(S) = 0.45$  (95% tolerance interval for each indicator) could be used as threshold for changes in indicator state. Regardless of how  $\hat{C}(S)$  is calculated, it allows for transparent quantitative integrations and aggregations of different indicators.

#### 2.4. Simulations

To study how well ASCETS performs we evaluated it using simulated data. We use an assessment period of six observations, in analogy

with the six year assessment period of the MSFD (EC, 2008), using reference periods with 6, 12, 18 and 24 observations, respectively. Random data were generated from Gaussian distributions with mean 100 and standard deviations (SD) of 0.1, 1, 5, 10, 15, 20, 30, 40, or 50 (the coefficient of variation,  $CV = SD/Mean = SD/100$  in all simulations). Values during assessment period were generated by a reduction of on average  $[0, 0.1, \dots, 0.9]$  using the same SD as the reference period. For each parameter combination of length of reference period (four levels), SD (nine levels), and reduction (ten levels), we simulated 100 replicates, in total 36 000 time-series. For each simulation we calculated  $C(S)$ , and recorded the number of breakpoints during the reference periods to assess the type 1 error rate in the breakpoint analysis.

#### 2.5. Application to ecological assessments

With extended information ASCETS can support ecological assessments within the EU's Marine Strategy Framework directive (EC, 2008) and Water Framework Directive (EC, 2000). The MSFD and WFD require that the status of the assessment period  $X_A$  should be classified

into two or five categories, respectively. This is not part of the *ASCETS* but can be provided by external information, such as independent data, prior knowledge or other types of “expert” judgements (Fig. 2). We here focus on implementation for a binary status classification as in the MSFD (EC, 2008), representing either success ( $X_A \in GS$ ) or failure ( $X_A \notin GS$ ) of achieving good ecological status (GS). A prerequisite to classify status with *ASCETS* is that the status of  $X_R$  can be defined based on external information. Boundary levels  $\hat{X}_i$  from *ASCETS* in this case would represent threshold levels for status classification (Fig. 2), which depend on type of indicator (above, below or within boundary levels), and whether  $X_R$  is considered to be in good status or not (Fig. 2). In this case the boundary levels represent the limit for a significant improvement or deterioration in indicator state of assessment period relative the reference period. To adequately signify improvement of status, it can be necessary to use more conservative percentiles when  $X_R \notin GS$  than when  $X_R \in GS$  (Fig. 2). It is also possible to use a higher number of status classes ( $c$ ), but the status during  $X_R$  has to be assigned to a specific status class  $c_i$  and additional boundary levels  $\hat{X}_i$  are determined as additional percentiles of  $X_R$  ( $X_{R,\alpha_1}, X_{R,\alpha_2}, \dots, X_{R,\alpha_{c-1}}$ ).

If status  $X_R$  can be defined, the confidence of  $X_A \in GS$ , denoted  $C(GS)$ , can be calculated in the same way as  $C(S)$ , i.e., as the proportion of  $X_A \in (X_{R,\alpha_1}, X_{R,\alpha_2})$ .  $C(GS) = 1$  means high confidence of achieving good status and  $C(GS) = 0$  high confidence of failure of achieving good status, and  $C(GS) = 0.5$  means no confidence of status classification (success as probable as failure). Also  $C(GS)$  can be aggregated across different indicators or sites as  $C(S)$  to obtain an integrated assessments,  $\hat{C}(GS)$ .

### 3. Results

#### 3.1. Evaluation of *ASCETS* based on simulations

Applied to simulated data, *ASCETS* correctly identified changes of indicator state  $p < 0.05$  when the proportional change in the indicator values between reference and assessment periods is at least twice the CV (Fig. 3a). *ASCETS* fails to detect a true change in indicator state, type II error, with increasing CV (Fig. 3a), and when the proportional change in indicator values are similar to CV,  $C(S) \approx 0.5$ . With greater CV, *ASCETS* fails to detect a true change in state more often than not, i.e. type II error dominates. Thus, based on random data, *ASCETS* will on average indicate changes in indicator state when the proportional change is larger than CV, and with high confidence when the change is twice the CV.

The type I error rate, indicating a change in indicator state when there is none, was around 6% so there is no obvious inflated type I error (Fig. 3b). At low CV (< 5%) and few (six) observations during the reference period, false indications of a change in state (type I error), occurred in 13% of the runs (Fig. 3b).

As *ASCETS* is intended for assessing changes in data poor situations, it is important to note that the number of observations during the reference period only have marginal influence on error rates (Fig. 3b, c). Failure to detect a change in status (type II error) instead mainly depends on the CV in relation to the proportional change between reference and assessment period, not the number of observations in reference period (Fig. 3c). Also, there is a moderate occurrence of type I error in the breakpoint analysis, on average 5% and highest for the largest number of observations (24 observations, Fig. 3d).

#### 3.2. *ASCETS* applied to coastal fish community indicators

For coastal fish indicators we use 2011–2016 as an assessment period and a baseline period of ten years or more ( $N_B \geq 10$ ) in line with EC (2008) and Helcom (2018b). Fig. 4a, b shows the indicator *Abundance of key coastal fish species*, *NPERCH* (in this case perch; *Perca fluviatilis* (L.)), at two Swedish coastal sites (Vinö and Forsmark). A structural change is identified at Vinö after the first year, which were

removed (Fig. 4a).  $C(S) = 0.4$  indicates a change in indicator state, but with low confidence. At Forsmark  $X_A$  is well within boundaries for a change in indicator state and  $C(S) = 1$  (Fig. 4b) suggesting a high confidence that the indicator state is the same during assessment and reference period. Aggregating the  $C(GS)$  for *NPERCH* gives, with an arithmetic mean,  $\hat{C}(S) = (0.4 + 1)/2 = 0.7$  and no overall change in indicator state across these two sites.

The indicator *Abundance of coastal fish key functional groups* is here the abundance of cyprinids, *NCYPR*. We show the indicator development at two Swedish coastal sites, Holmön and Forsmark. At both sites  $C(S) > 0.5$ , indicating no change in indicator state between assessment and reference periods, but the confidence for no-change is lower  $C(S) = 0.67$  at Holmön than at Forsmark  $C(S) = 0.97$ , with an aggregated confidence of change in status  $\hat{C}(S) = (0.67 + 0.97)/2 = 0.82$ .

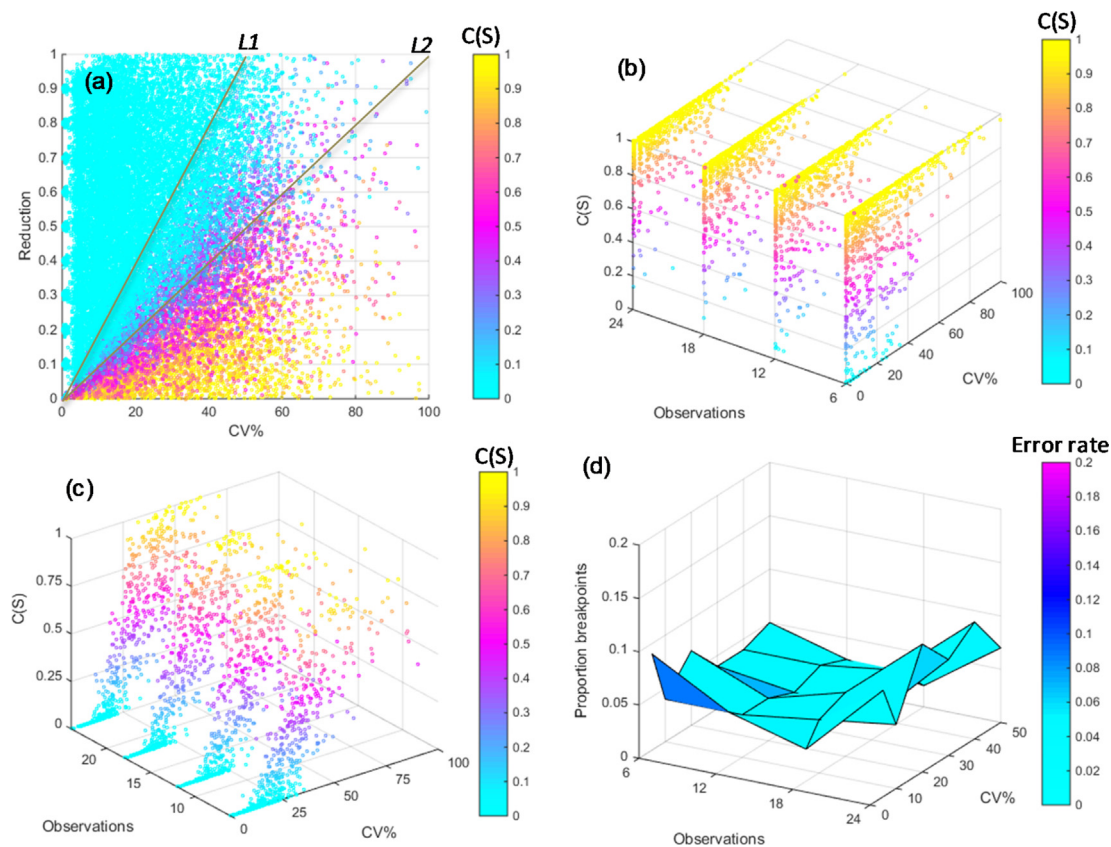
The coefficients of variation (CV) for *NPERCH* and *NCYPR* were around 50–60% (Fig. 4), which according to the simulations would require proportional changes in indicator values between reference and assessments periods to be around 50% or larger for indications of change in indicator states (Fig. 3a). This is also what is observed for *NPERCH* at Vinö (Fig. 4a) and *NCYPR* at Holmön (Fig. 4c) resulting in low confidence assessments. In contrast, at Forsmark *NPERCH* and *NCYPR* (Fig. 4b,d) have smaller changes in indicator values and not likely to be detected by *ASCETS*, or rephrased, we can be confident changes in indicator values are so low relative CV that changes in indicator state cannot be detected using *ASCETS*.

#### 3.2.1. Example of status classification

To apply *ASCETS* to an assessment for the MSFD we need to decide what boundary levels represent in this specific management context, and classify the indicator status during reference periods. According to Helcom (2018b), the *NPERCH* indicator should be above a reference level for good environmental status, and the *NCYPR* indicator should be within an upper and a lower reference level. There is lack of independent data to define status during the reference periods. For illustrative purpose only, as there is no obvious deterioration of indicator values during baseline periods we consider reference periods as a sustainable state and classify them as representing good status. With these assumptions the lower tolerance level of the *NPERCH* would represent the boundary for good status, and the range between lower and upper tolerance levels of *NCYPR* would represent the boundary levels for good status of this indicator. It is worth pointing out that for indicators with only one reference level, like *NPERCH*, two-sided tests are not relevant, and one-sided tests are more suitable.

#### 3.3. *ASCETS* applied to the large fish indicator

The fish community indicator *Large Fish Indicator (LFI)* has been suggested for assessing the status of fish communities (functioning and resilience), with a greater proportion of large fish indicating a better status (Shephard et al., 2011; Oesterwind et al., 2013; Modica et al., 2014). For the demersal fish community in the Baltic Sea, *LFI* is the proportion biomass of fish > 40 cm (Oesterwind et al., 2013). Here we apply standardised data of *LFI* from a GAM-model to pelagic trawl hauls from fishery-independent surveys in the Baltic Sea Proper (Oesterwind et al., 2013; see Casini et al., 2019 for the GAM modelling framework). Cod (*Gadus morhua* (L. 1758)) is a piscivore that dominates the fish community > 40 cm while the planktivores herring (*Clupea harengus* (L. 1758)), sprat (*Sprattus sprattus* (L. 1758)) and three-spined stickleback (*Gasterosteus aculeatus* (L. 1758)) dominate < 40 cm. We use 2010–2015 as the assessment period ( $N_A = 6$ ) and 1979–2009 as baseline period. The structural change test of the baseline period indicates a breakpoint in 1990, and as 1990–2009 covers a longer time-period without any trend we use it as the reference period (Fig. 5a). *LFI* during the assessment period has declined more than twice the standard deviation of the reference period resulting in high confidence for a



**Fig. 3.** Results from applying ASCETS to simulated time-series. (A) Estimated confidence of change in state,  $C(S)$ , from ASCETS in relation to coefficient of variation (CV) and proportion reduction of indicator values between reference and assessment period, across reference periods with 6, 12, 18, and 24 number of observations. In general,  $C(S) < 0.05$  (cyan, ASCETS indicates high confidence of change) when CV is half of the change in indicator values (line L1). When the reduction in indicator values is similar to CV (line L2),  $C(S) \approx 0.5$ , hence, detecting a change with similar probability as not (magenta). As CV becomes larger than the reduction in indicator values, ASCETS fails to identify a change in indicator state ( $C(S) > 0.5$ , yellow). (B) When there is no change in indicator values between reference and assessment period, false detection rate (type I error) is below 5% ( $C(S) > 0.95$ , yellow) for  $CV > 40\%$  for all length of reference periods (numbers of observations). False detection of change in state mainly occurs when CV is close to zero ( $C(S) < 0.05$ , cyan) and the reference period contains few observations. (C) The number of observations during the reference period has similar influence on the failure (type II error) rate of ASCETS to detect true changes irrespective of CV. Graph shows the example of 30% reduction in indicator values between reference and assessment periods. When  $CV < 30\%$  ASCETS detects the change with high confidence ( $C(S) < 0.05$ , cyan), and the increase in  $C(S)$  (reduced confidence of change) with CV is similar for all number of observations in reference period. (D) The breakpoint analysis did not show an inflated false detection rate of breakpoints (error rate  $< 0.05$ , cyan) for reference periods of 6–18 observations, but a higher false detection rate is observed at 24 observations of the reference period ( $\approx 10\%$ , blue).

change in indicator state,  $C(S) = 0.007$ .

### 3.3.1. Status classification using external data

Cod, herring and sprat in the Baltic Sea have analytical single species reference points from stock assessment models during the reference period (ICES, 2014) that can be used to assess status of  $LFI$  during the reference period. (Fig. 5). The combined spawning stock biomass of herring and sprat,  $B_{HS} = \text{mean}(B_H/B_{H,pa} B_S/B_{S,pa})$ , 1990–2009 has been above  $B_{pa}$  (i.e. biomass assuring full reproductive capacity accounting for uncertainty, corresponding in the stocks analysed here to  $MSY$  Brigger, that is the biomass that should trigger a management action) for most the reference period, and their combined fishing mortality has been around  $F_{msy}$  (fishing mortality assuring long term maximum yield) (Fig. 5b). For cod, however,  $B_{cod}$  was below  $B_{msy}$  during 1999–2009 and  $F_{cod}$  well above  $F_{msy}$  (Fig. 5c). Therefore we can set the status of  $LFI$  during the reference period 1990–2009 as poor. That means the upper boundary level could be used as management target for success of achieving good status,  $LFI = 1.05$ , far from the median of the assessment period of  $-2.18$ .

There is an evident positive correlation between  $B_{cod}$  and  $LFI$  (Fig. 5d), suggesting  $LFI$  is related to  $B_{cod}$ . Although substantial variation around the relationship  $B_{cod} > B_{pa}$  for most  $LFI > 1$  (management target derived from ASCETS). The opposite does not apply as

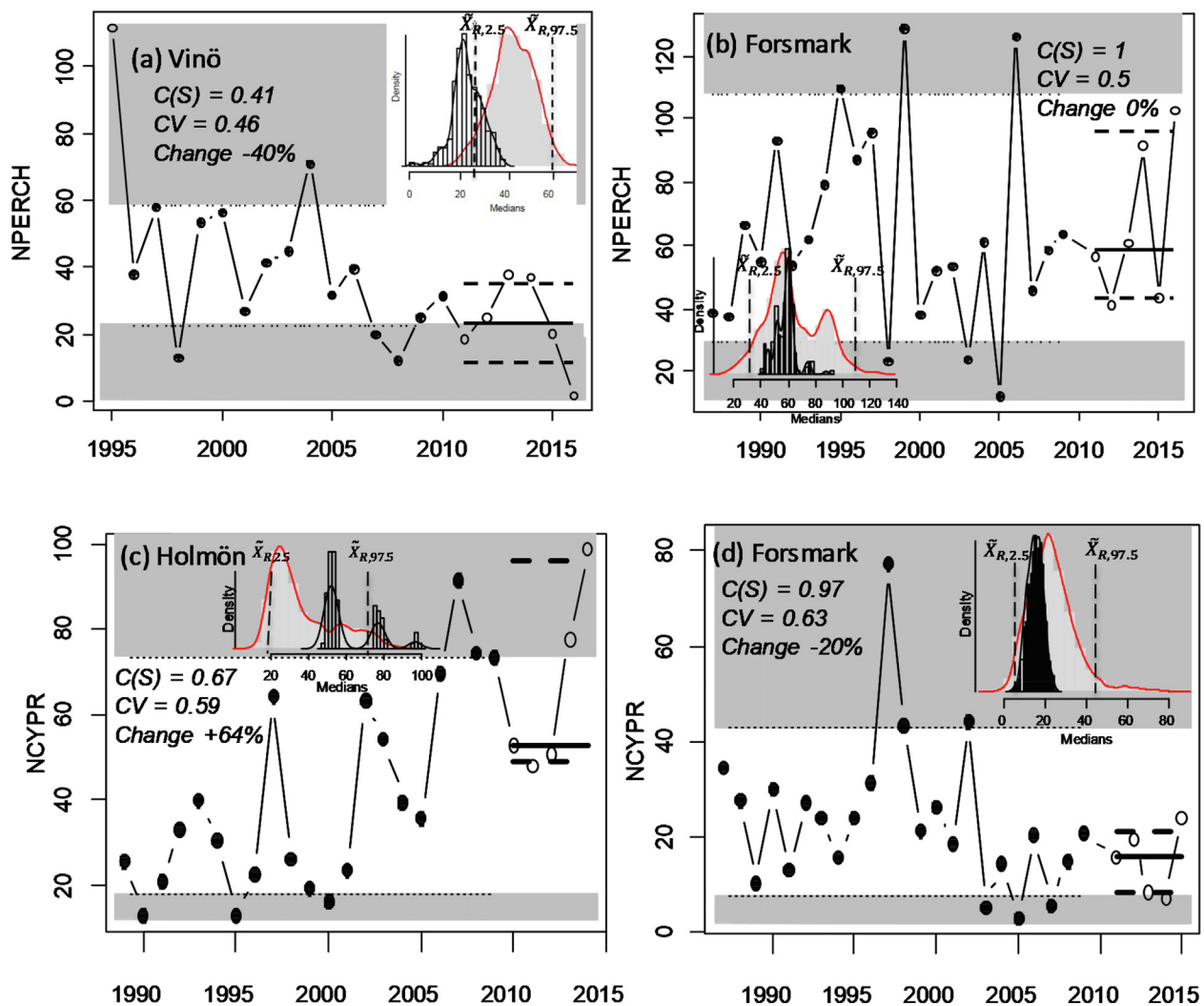
$$B_{cod} > B_{pa} \text{ for } LFI < 1.$$

### 3.4. ASCETS applied to waterbird indicators

The indicator *Abundance of waterbirds in the breeding season* is used to assess the status of 29 waterbird species at the Baltic Sea scale, and are done for all bird species combined and five functional groups based on feeding behaviour (surface, pelagic, benthic, grazer and wader feeders; Supplementary Material Table A2; Helcom, 2018c). We here apply ASCETS to bird count data of breeding pairs of respective functional group from the Småland archipelago in south east of Sweden (Johansson and Larsson, 2008). In this areas, breeding birds have been monitored on 53 islands 1990–2015, and we use the same baseline period 1991–2000 as in Helcom (2018c) and 2010–2015 as a six-year assessment period. The ‘All’ species indicator was calculated as the average annual z-score (mean = 0, SD = 1) of each functional group, as some species had so low abundances that stochasticity is likely to impact the results.

The structural change tests indicate structural changes in the baseline periods for surfer feeding birds, waders, and the ‘All’ groups (Fig. 6) and we consequently use the longer parts of the time-series as reference period (to avoid that the  $N_R < N_A$ ).

All indicators show changes in state between reference and



**Fig. 4.** Application of ASCETS to coastal fish community indicators. Solid circles represent observations used for reference period, and open circles observations during the assessment period. Smoothed resampled median distributions of the reference periods are shown in the inserted histogram. Grey bars are smoothed bootstrapped distribution from the reference period and open bars from the assessment period. White areas correspond to indicator values representing same status as reference period and grey areas representing changes in indicator status. Inserted grey histograms show smoothed resampled median distribution of indicators during the reference period and black bars show of the assessment period. Solid lines are observed median indicator values during assessment periods and the hatched lines are the 95% percentile interval of  $X_A$ . The confidence for change in indicator status is given by  $C(S)$ . CV and Change indicate coefficient of variation during reference periods and proportional mean change between assessment and reference period, respectively. NPERCH is the catch per unit effort of perch in fish monitoring sites at (a) Vinö 1995–2016 and (b) Forsmark 1987–2016. NCYPR is the Abundance of cyprinids in (c) Holmön 1989–2014 and (d) Forsmark 1991–2016.

assessment period (Fig. 6), and the proportional change between assessment and reference period is much higher than CV resulting in overall high confidence of changes in indicator states (Fig. 6). Grazing feeders is the only group showing an increase in indicator state with moderate confidence,  $C(S) = 0.32$ , whereas all other groups show a decline in indicator state. The integrated assessment based on the confidence of the single group assessments would use an arithmetic mean be  $\hat{C}(S) = (0 + 0.32 + 0.06 + 0 + 0)/5 = 0.08$ , indicating high confidence of change in state of breeding waterbirds in this area.

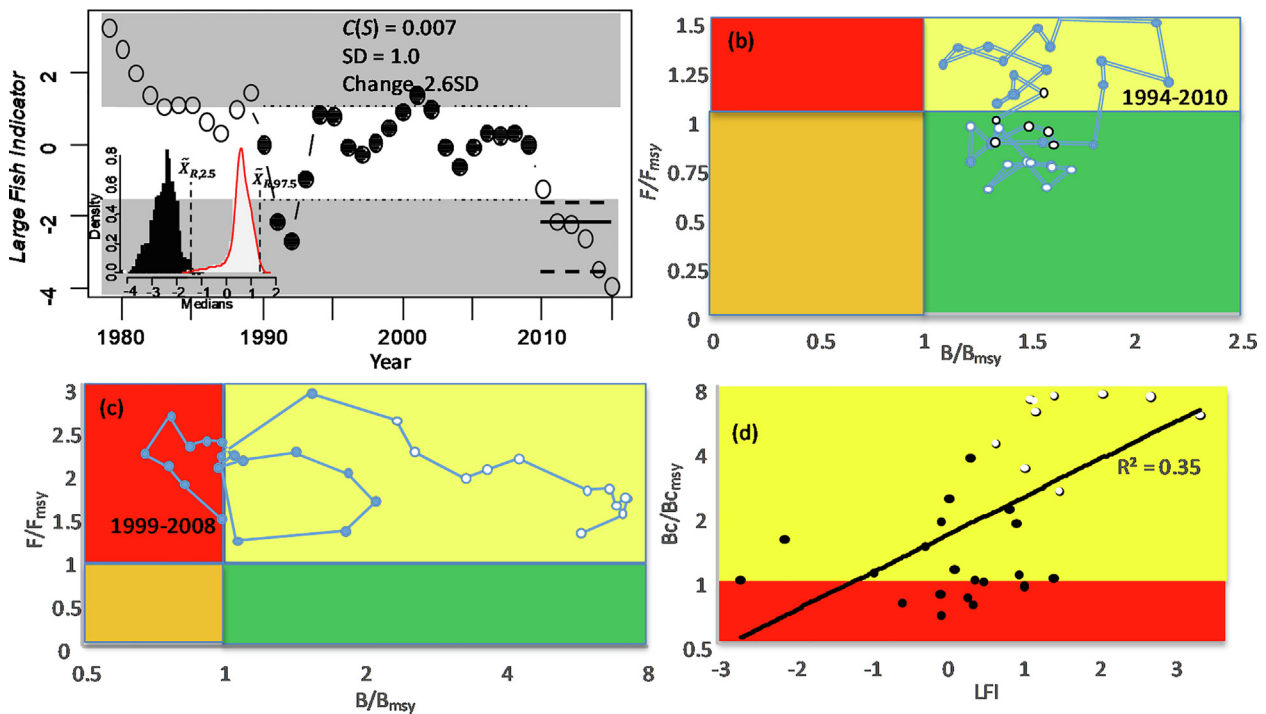
It is difficult to assess the status of waterbird indicators during the reference period at this spatial scale, but waterbird indicators should be above a threshold level for good status (Helcom, 2018c). We therefore use the lower boundary as management target for non-deteriorated waterbird community, which gives  $\hat{C}(GS) = (0 + 0.96 + 0.06 + 0 + 0)/5 = 0.2$ , i.e. the integrated assessment suggests a decline in waterbird status with a moderately confidence. The difference between  $\hat{C}(GS)$  and  $\hat{C}(S)$  is that Grazing feeders has increased during the assessment period and therefore has high confidence of non-deterioration (0.96) in  $\hat{C}(GS)$ , but lower confidence of being in

the same state as during reference period.

We also applied ASCETS on data from a subset of 31 sampling areas from a regional bird monitoring scheme along the Swedish west coast 2001–2013 (Alexandersson, 2011). This is a too short time period to divide time-series into baseline and assessment periods. Instead, ASCETS is applied to identify breakpoints that would indicate a change in the dynamics of indicators (Fig. 7). For grazers, benthic and pelagic-feeding birds there was no significant structural change. Three groups: “All group”, surface and wader feeding birds, had significant breakpoints with declining indicator values after the breakpoint (Fig. 7). As time-series are short a quantitative assessment of confidence is not feasible but instead we can conclude that two out of five functional groups of waterbirds have declining indicator values during this time period.

#### 4. Discussion

We propose ASCETS as a generic tool for assessing changes in indicator state of ecosystem indicators from time-series. It supports



**Fig. 5.** Application of ASCETS to the Large Fish Indicator (LFI) for pelagic fish in the Baltic Sea. (a) Standardised data of LFI 1979–2015 where 1990 is identified as a breakpoint. 1990–2009 is used as reference period (blue circles; 1979–1989 white circles). The confidence for change in status during the assessment period is high,  $C(S) = 0.007$ . SD and Change indicate standard variation during reference periods and proportional change between assessment and reference period, respectively. The inserted histogram shows smoothed resampled median distributions from reference period (black) and assessment period (white). In (b, c) coloured areas indicate different status from single stock assessments (ICES, 2014) depending on stock biomass (B) and fishing mortality (F). Green is sustainable state ( $B > B_{pa}$ ,  $F < F_{msy}$ ), red non-sustainable ( $B < B_{pa}$ ,  $F > F_{msy}$ ), and yellow/orange represent states that require actions for retaining sustainable use. White circles are estimates 1979–1989, solid circles the reference period 1990–2009, and in (b) black circles estimates during the assessment period 2010–2015. Based on these single stock assessment the status of LFI is assessed as poor 1990–2009. Albeit the average fishing pressure of herring and sprat (b) has been above  $F_{msy}$ , average spawning stock biomass has been above  $B_{pa}$ . In contrast, for cod (c) both fishing pressure has been above  $F_{msy}$  and biomass below  $B_{pa}$ , especially during the second half of the reference period. LFI is positively associated with stock biomass estimates of cod ( $B_c$ ) in 1979–2008 (d).

derivation of quantitative boundary levels for assessing changes in indicator state and management targets when other objective reference levels are lacking, as is typical for many ecological indicators. ASCETS can provide confidence for changes in indicator state, which may support risk assessments across different indicators as well as enable aggregated or integrated assessment based on the confidence of state changes of each indicator. It is developed mainly for time-series with more than ten observations, and not suitable for indicator assessments from few or single observations, but can be adjusted for temporal trends, breakpoints and autocorrelation that makes it applicable to almost any time-series of ecological indicators, as shown by the examples here. Based on random data ASCETS on average identify changes in indicator values larger than the coefficient of variation (CV) as a change in indicator state, and with high confidence when the change is twice the CV or more. If the indicator status during reference period can be defined, which requires additional information, the boundary levels can be used to assess ecological/environmental status during the assessment period.

The evaluation of ASCETS against simulated data shows that ASCETS correctly assess changes in indicator state with 95% probability or more when the change between reference and assessment periods is twice as large as the CV (Fig. 3a). When the change between reference and assessment periods is less than the CV, ASCETS more often fails to assess the changes in indicator state than not. This evaluation was done on random data with a Gaussian distribution, which however, may rarely apply to real data where skewed distributions and autocorrelation are common features. Future studies are needed to investigate how well ASCETS performs under such situations. Importantly however, the number of observations during the reference period has marginal

impact on the likelihood of correct identifying changes in indicator state, which means that ASCETS works also for shorter time series. In our simulated data, the number of observations during reference period had little effect on the likelihood for detecting false breakpoints, typically 3–5% and maximum around 10% for longer time-series (Fig. 3d). ASCETS seems robust with respect to the number of observations in the reference period, and is more sensitive to the variation in time-series (CV) in relation to the degree of change between reference and assessment periods for assessments of changes in indicator state.

If there is no predefined baseline period and assessment period or short time-series, 'strucchange' can be used to identify breakpoints in indicator time-series (Fig. 7). A significant breakpoint indicates a change in the indicator dynamics that could correspond to a change in indicator state. Boundary levels for indicator state may then be set from resampled median distributions during a sub-period of the time-series. When no breakpoint is identified the whole time-series can be used to calculate boundary levels. Thus, ASCETS can be used to set boundary levels of indicator state also without predefined baseline and assessment periods.

Biological data are associated with both observational and environmental noise, which may affect the confidence of change (Soldaat et al., 2017). ASCETS provides a quantitative measure of the confidence specified as a proportion (% resampled median distribution during an assessment period within boundary levels from the reference period), which enable comparisons between sites or across indicators also when nominal reference levels differ between sites or indicators. This allows for a quantitative evaluation of which ecosystem component, or site, may be most likely to have changed state between reference and assessment period. The confidence level also enable aggregation of



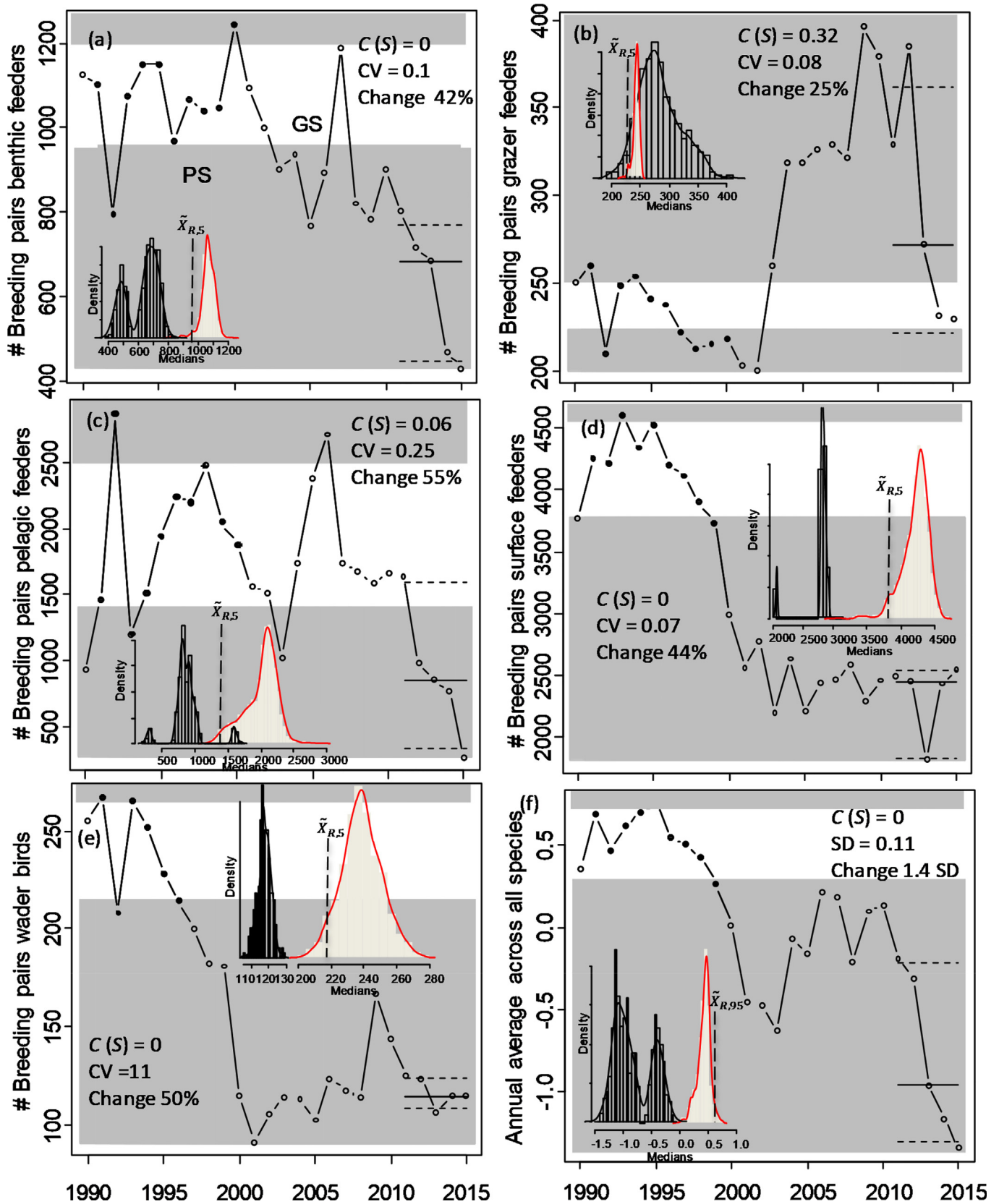


Fig. 6. Observed number of breeding waterbirds at 53 islands in Småland archipelago in southeast Sweden for different functional groups of waterbirds. The period 1991–2000 is used as baseline period for all groups, but because of significant breakpoints, the reference period (solid circles) is shorter for some groups (e, f). The assessment period was 2010–2015 (open circles) for all groups. Smoothed resampled median distributions (grey bars = reference period, black bars = assessment period) are shown in the inserted histogram. The estimated confidence for good status during the assessment period is given by  $C(GS)$ ,  $CV$  and Change indicate coefficient of variation during reference periods and proportional change between assessment and reference period, respectively.

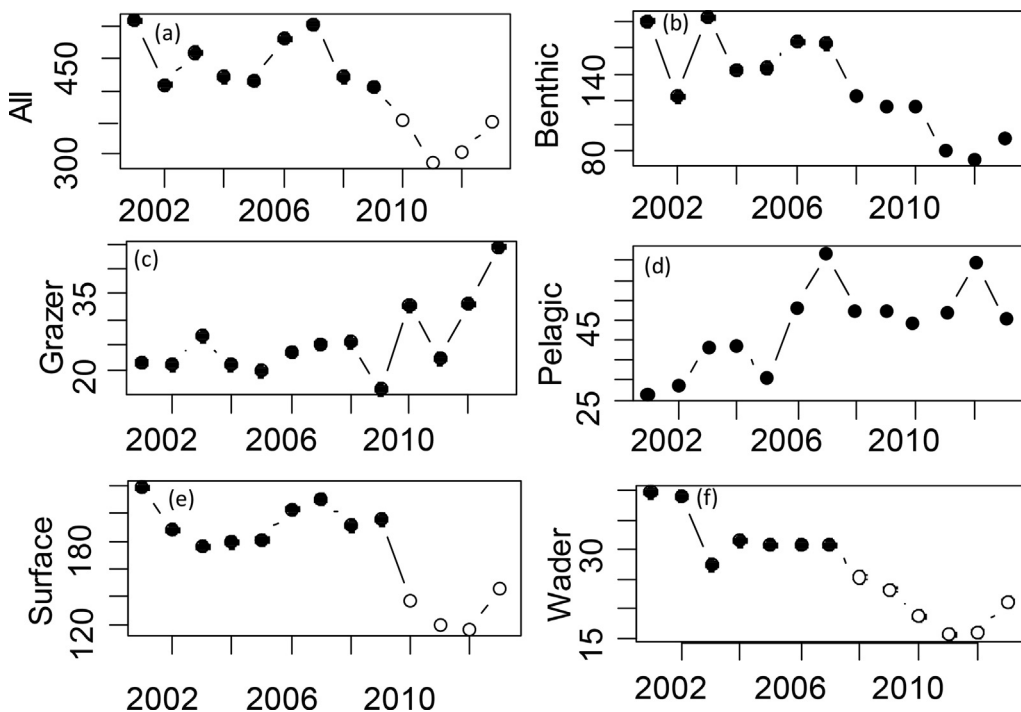


Fig. 7. Observed number of breeding waterbirds 2001–2013 at 31 islands in an archipelago in western Sweden for different groups of waterbirds. Because of the short time-series there is no pre-defined baseline/reference or assessment period. Breakpoint analysis suggests significant breakpoints in time-series for the All-group (a), surface (c) and wader feeding birds (d), indicating a structural change in indicator dynamics. Breakpoints are indicated as a shift from black to white dots. For the other groups there was no significant breakpoint during time-series.

estimates of changes in indicator state across different indicator time-series.

ASCETS has some resemblance to assessment models for data limited fish stocks lacking analytical reference points, e.g. Status-quo harvest control, Depletion-Corrected Average Catch (reviewed by ICES, 2012), in that the current state is assessed relative a reference period. These fishery models aim to set year-to-year fishing levels or effort based on the last years’ observations (often last 2–3 years, ICES, 2012) and are prone to observation errors and environmental stochasticity. ASCETS focuses on larger time scales (preferably > 10 observations) and may respond to an array of both natural and human induced environmental drivers as well as observation errors. The pressures or drivers impacting indicators need to be discerned and appropriate actions taken when ASCETS indicates a change in indicator state.

Other approaches used to assess changes of ecosystem components based on temporal trends include ‘TRIM’, used for assessments of birds (Gregory et al., 2005; Soldaat et al., 2017) and ‘Criteria A’ used for categorising species for the IUCN ‘Red list’ (IUCN, 2012). Criteria A considers species-specific changes in abundance during a ten-year period, or three generations. Different estimated levels (in %) of decrease result in different status classes (from “extinct” to “least concern”), but uncertainty is not explicitly handled other than that there is a data-deficient class (IUCN, 2012). TRIM uses standardised values of count data over a large number of sites or species to calculate an index and the statistical significance of a temporal change, and uncertainty can be handled by resampling of data to provide confidence intervals (Soldaat et al., 2017), hence, similar to ASCETS. Both ASCETS and TRIM can integrate assessments across species, indicators and sites and, thus, share some features such as stochasticity and uncertainty and can be used to integrate across different indicators. The main difference lays in their objectives. Whereas TRIM is designed for assessing statistical deviations of linear (linearized) long-term trends, ASCETS is designed to identify breakpoints and compare reference and assessment periods.

Greenstreet et al. (2012) suggested an ecosystem approach using a large number of indicators. To assess overall changes in ecosystem status they used the number of indicators departing from expected from a binomial model. ASCETS resembles their approach in that indicators are assessed relative to a distribution of observed values during a time-series, but the approach by Greenstreet et al. (2012) focuses on

assessment at a single occasion (year) and require an aggregated assessment of several (> 10) indicators simultaneously. ASCETS works better for longer assessment periods and is based on assessments of single indicator states that can be aggregated.

In the Baltic Sea, a decline in the state of waterbirds is currently assessed as if < 75% of the species during the assessment period have an abundance index  $\leq 70\%$  of the abundance index during the baseline period (Helcom, 2018c). For assessments on a smaller spatial scale as applied herein there may be few species in each functional group. Reference levels based on a proportion of species with low abundance can result in stochastic and uncertain assessment. ASCETS can therefore serve as a complement for bird communities in cases of lower data availability (fewer species, smaller ranges, and shorter time-series).

The MSFD and WFD require EU member states to assess ecological status of all waterbodies, with a management cycle of six years (EC, 2000, 2008). ASCETS cannot assess ecological status, but it derives quantitative boundary levels for assessing changes in indicator state and therefore be a supportive tool for assessing ecological status from indicators. If the state of the indicator during the reference period can be decided, the boundary levels from ASCETS imply thresholds for significant improvement or deterioration in indicator state that can act as assessing criteria for the current ecological status. It should be noted that an indicator showing significant deviation from another period does not necessarily imply a change in ecological/environmental status. Given the many aspects of ecosystem components to be evaluated under ecosystem-based management, site-specific management targets, and in many cases data limitations, the approach to identify boundary levels by ASCETS can be one pragmatic method. Especially as it is possible to quantitatively aggregate and integrate indicator assessments derived from ASCETS, ecological assessments over larger spatial scales or many different ecosystem components is supported.

All examples presented here apply to ecological indicators. It is also possible to apply ASCETS to time-series of pressure or driver indicators. In such cases boundary levels should be more conservative. Although a pressure indicator is within a specific distribution (e.g. 95% of the median distribution from a reference period), changes in its average value may inflict significant effects on state variables. Applying ASCETS to pressure indicators, such as nutrients and toxin concentrations, fishing or hunting effort, may nevertheless serve as a way to assess long-

term changes in pressures incorporating stochasticity.

## 5. Conclusions

ASCETS offers a generic quantitative tool be applicable to a wide variety of ecological time-series for assessing changes in indicator state, and can be used both to set site and indicator specific management targets and for risk assessments. The framework is primarily useful in situations when analytical or local reference levels are missing, as a result of environmental variation or differences in sampling methodology across time-series. As ASCETS allows for a scale independent integration of indicators, it is useful in many cases for implementation of an ecosystem-based approach to management, where different ecosystem components should be considered together.

## CRedit authorship contribution statement

**Örjan Östman:** Methodology, Software, Formal analysis, Writing - original draft. **Lena Bergström:** Conceptualization, Writing - review & editing. **Kjell Leonardsson:** Conceptualization, Methodology. **Anna Gårdmark:** Conceptualization, Methodology, Writing - review & editing. **Michele Casini:** Data curation, Writing - review & editing. **Ylva Sjöblom:** Data curation. **Fredrik Haas:** Data curation, Writing - review & editing. **Jens Olsson:** Writing - review & editing, Project administration.

## Acknowledgement

This study has been partly funded by the Swedish Agency for Marine and Water Management projects #1121-208 and #1411-19, as well as Quantitative Fish and Fisheries Ecology at SLU.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ecolind.2020.106469>.

## References

- Alexandersson, H., 2011. Bohuskustens häckfågelfauna 2001-2009. Förekomst, reproduktion och habitat. Rapport 2011:70. Länsstyrelsen i Västra Götalands län. ISSN: 1403-168X.
- Borja, A., Elliott, M., Carstensen, J., Heiskanen, A.-S., van de Bund, W., 2010. Marine management – towards an integrated implementation of the European Marine Strategy Framework and the Water Framework Directives. *Mar. Pollut. Bull.* 60, 2175–2186.
- Borja, A., Dauer, D.M., Grémare, A., 2012. The importance of setting targets and reference conditions in assessing marine ecosystem quality. *Ecol. Ind.* 12, 1–7.
- Borja, A., Elliott, M., Andersen, J.H., Cardoso, A.C., Carstensen, J., Ferreira, J.G., Uusitalo, L., 2013. Good environmental status of marine ecosystems: what is it and how do we know when we have attained it? *Mar. Pollut. Bull.* 76, 16–27.
- Casini, M., Tian, H., Hansson, M., Grygiel, W., Strods, G., Statkus, R., Larson, N., 2019. Spatio-temporal dynamics and behavioural ecology of a “demersal” fish population as detected using research survey pelagic trawl catches: the Eastern Baltic Sea cod (*Gadus morhua*). *ICES J. Mar. Sci.* 76, 1591–1600.
- EC, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. *Official Journal of the European Communities*, L327, 22/12/2000.
- EC, 2008. Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy. *Official Journal of the European Communities*, L164, 19-40.
- Flather, C.H., Hayward, G.D., Beissinger, S.R., Stephens, P.A., 2011. Minimum viable populations: is there a ‘magic number’ for conservation practitioners? *Trends Ecol. Evol.* 26, 307–316.
- Greenstreet, S.P.R., Rossberg, A.G., Fox, C.J., Le Quesne, W.J.F., Blasdale, T., Boulcott, P., Mitchell, I., Millar, C., Moffat, C.F., 2012. Demersal fish biodiversity: species-level indicators and trends-based targets for the Marine Strategy Framework Directive. *ICES J. Mar. Sci.* 69, 1789–1801.
- Gregory, R.D., van Strien, A., Vorisek, P., Meyling, A.W.G., Noble, D.G., Foppen, R.P.B., Gibbons, D.W., 2005. Developing indicators for European birds. *Philos. Trans. R. Soc. B: Biol. Sci.* 360, 269–288.
- Helcom, 2018a. State of the Baltic Sea – second HELCOM holistic assessment 2011–2016. *Baltic Sea Environment Proceedings* 155.
- Helcom, 2018b. Status of coastal fish communities in the Baltic Sea during 2011–2016 – the third thematic assessment. *Baltic Sea Environment Proceedings* 161.
- Helcom, 2018c. Abundance of waterbirds in the breeding season. HELCOM core indicator report.
- ICES, 2012. ICES Implementation of Advice for Data-limited Stocks in 2012 in its 2012 Advice. *ICES CM* 2012/ACOM, 68.
- ICES, 2014. Report of the Baltic Fisheries Assessment Working Group (WGBFAS), 3–10 April 2014, ICES HQ, Copenhagen, Denmark. 919 pp.
- IUCN, 2012. IUCN Red List Categories and Criteria: Version 3, 2nd ed. Gland, Switzerland and Cambridge, UK.
- Johansson, T., Larsson, T., 2008. Häckfågelfaunan i östra Smålands ytterskärgård 1990–2008. Länsstyrelsens meddelandeserie 2008, 13.
- Kempf, A., 2010. Ecosystem approach to fisheries in the European context – history and future challenges. *J. Appl. Ichthyol.* 26, 102–109.
- Modica, L., Velasco, F., Preciado, I., Soto, M., Greenstreet, S.P.R., 2014. Development of the large fish indicator and associated target for a Northeast Atlantic fish community. *ICES J. Mar. Sci.* 71, 2403–2415.
- Oesterwind, D., Psuty, I., Pachur, M., Dorrien, C. von, Lejk, A.M., Casini, M., Larson, N., 2013. Proportion of large fish in the community. In: Helsinki Commission/Baltic Marine Environment Protection Commission. Helsinki Commission, Baltic Marine Environment Protection Commission, Helsinki.
- Östman, Ö., Lingman, A., Bergström, L., Olsson, J., 2017. Temporal development and spatial scale of coastal fish indicators in reference ecosystems: hydroclimate and anthropogenic drivers. *J. Appl. Ecol.* 54, 557–566.
- Samhuri, J.F., Lester, S.E., Selig, E.R., Halpern, B.S., Fogarty, M.J., Longo, C., McLeod, K.L., 2012. Sea sick? Setting targets to assess ocean health and ecosystem services. *Ecosphere* 3, 1–18.
- Shepherd, S., Reid, D.G., Greenstreet, S.P.R., 2011. Interpreting the large fish indicator for the Celtic Sea. *ICES J. Mar. Sci.* 68, 1963–1972.
- Shin, Y.-J., Shannon, L.J., 2010. Using indicators for evaluating, comparing, and communicating the ecological status of exploited marine ecosystems. 1. The IndiSeas Project. *ICES J. Mar. Sci.* 67, 686–691.
- Soldaat, L.L., Pannekoeka, J., Verweija, R.J.T., van Turnhout, C.A.M., van Striena, A.J., 2017. A Monte Carlo method to account for sampling error in multi-species indicators. *Ecol. Ind.* 81, 340–347.
- Taylor, L.R., 1961. Aggregation, variance and the mean. *Nature* 189, 732–735.
- Thorpe, R.B., Le Quesne, W.J.F., Luxford, F., Collie, J.S., Jennings, S., 2015. Evaluation and management implications of uncertainty in a multispecies size-structured model of population and community responses to fishing. *Methods Ecol. Evol.* 6, 49–58.
- Wolodko, T., 2018. kernelboot: Smoothed Bootstrap and Random Generation from Kernel Densities. R package version 0.1.4. <https://CRAN.R-project.org/package=kernelboot>.
- Zeileis, A., Leisch, F., Hornik, K., Kleiber, C., 2002. strucchange: an R Package for testing for structural change in linear regression models. *J. Stat. Softw.* 7, 1–38.
- Zeileis, A., Kleiber, C., Kraemer, W., Hornik, K., 2003. Testing and dating of structural changes in practice. *Comput. Stat. Data Anal.* 44, 109–123.