ICES Journal of Marine Science, 2024, Vol. 00, Issue 0, 1–14
https://doi.org/10.1093/icesjms/fsae103
Received: 30 November 2023; **revised:** 13 June 2024; **accepted:** 16 July 2024
Original Article

ICES CIEM
International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

Poor performance of regime shift detection methods in marine ecosystems

Hannah Haines [1,*], **Benjamin Planque** [2], **Lucie Buttay** [2,3]

[1]Alfred-Wegener-Institut Helmholtz Zentrum für Polar- und Meeresforschung, Am Handelshafen 12 27570 Bremerhaven, Germany
[2]Institute of Marine Research, Postboks 6606 Stakkevollan N-9037 Tromsø, Norway
[3]UiT The Arctic University of Norway, PO Box 6050 Stakkevollan N-9037, Tromsø, Norway

*Corresponding author. Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, Am Handelshafen 12, 27570 Bremerhaven, Germany. E-mail: hannah.haines@awi.de

Abstract

Regime shifts have been reported as ubiquitous features across the world's oceans. Many regime shift detection methods are available, but their performance is rarely evaluated, and the supporting evidence for regime shifts may be thin because of the nature of marine ecological time series that are often short, autocorrelated, and uncertain. In the Norwegian Sea, a regime shift has been reported to have occurred in the mid-2000s, with simultaneous changes in oceanography, plankton, and fish. Here, we evaluate the evidence for this regime shift using four commonly used regime shift detection methods (Strucchange, STARS, EnvCpt, and Chronological Clustering) on 32 annual time series that describe the main components of the Norwegian Sea ecosystem, from hydrography and primary production up to fish population metrics. We quantify the performance of each method by measuring its false-positive rate, i.e. the proportion of times the method detects a regime shift that was not present in simulated control time series. Our results show that all methods have high to very high false-positive rates. This challenges the evidence for a regime shift in the Norwegian Sea and questions earlier reviews of regime shifts across the world's oceans.

Keywords: changepoint; surrogate time series; Norwegian Sea; false-positive rates; STARS; EnvCpt; Strucchange; Chronological Clustering

Introduction

Ecosystems are intricate and ever-fluctuating systems, each possessing a unique collection of species, habitats, and environmental conditions. The notion of ecosystem *state* is often used to characterise an ecosystem at a given point in time, while the idea of ecosystem *regime* embeds the natural variability of an ecosystem (Scheffer and Carpenter 2003, Möllmann et al. 2015). Furthermore, ecosystem regimes are generally considered as being stationary around an average, although they can also exhibit cyclical or chaotic patterns (Scheffer et al. 2009). Transitions between distinct regimes have been termed *regime shifts*. Regime shifts can be defined as sudden and abrupt transitions in community structure (Möllmann et al. 2015) that affect multiple trophic levels (Lees et al. 2006) and result in a rapid reconfiguration of the ecosystem that persists over time (Möllmann et al. 2015). As such, stating the presence of a regime shift requires simultaneous or quasi-simultaneous (i.e. a time lag of a year) breakpoints in multiple time series of an ecosystem's key variables, such as net primary production and the abundance of keystone species. This definition is used here, though definitions vary from study to study (de Young et al. 2004). Well-documented regime shifts include the rapid transition from kelp-dominated to urchin-dominated communities in temperate marine benthic ecosystems (Steneck et al. 2002, Konar and Estes 2003) and the shift from coral-dominated to macroalgal-dominated communities after heatwaves in tropical marine benthic communities (Cheal et al. 2010).

However, beyond this empirical definition and these widely agreed-upon examples, there is much confusion and misinterpretation when it comes to determining the presence of a regime shift. There is an ongoing debate regarding the effective identification of regime shifts from empirical time series. Climatological and ecological time series emerge from the combination of numerous small-scale and large-scale fluctuations (Di Lorenzo and Ohman 2013), resulting in complex layered patterns. These time series often display more variation at low rather than at high frequencies, a property described as red noise. Rudnick and Davis (2003) have shown that regime shifts are likely to be detected in Gaussian red noise with stationary statistics. In other words, regime shift detection methods may wrongly detect shifts when the temporal signal is stationary but dominated by low frequencies. Hsieh et al. (2005) reflect the debate on regime shift detection methods by summarising that one side argues that apparent sudden shifts in physical variables, such as the Pacific Decadal Oscillation (PDO), represent normal statistical deviations or random events (Rudnick and Davis 2003), while the other side argues that sudden environmental variations most likely result from nonlinear phenomena and thus constitute regime shifts (Hare and Mantua 2000, Scheffer et al. 2001, de Young et al. 2004). Both sides seem to agree that true regime shifts are not random features of time series but rather unexpected and, to some degree, unpredictable nonlinear phenomena (Hsieh et al. 2005). However, they seem to disagree on what is defined as *unexpected* and *nonlinear*.

Another common feature of reddened ocean, climate, and ecological time series is autocorrelation, which refers to the dependence of a time series' values on its earlier observations.

When autocorrelation is present, the value of the time series at a given time is related to its previous values. Thus, autocorrelation violates the common assumption of independence between observations, altering statistical inference leading to higher rejection rates of null hypotheses (Pyper and Peterman 1998). This bias in inference can favour the identification of regime shifts. Physical, biological, and ecological processes that lead to autocorrelation and low-frequency variations (reddened signals) render the identification of regime shifts particularly difficult in real-world case studies, as they diverge observational time series from the idealised examples used to develop and test regime shift detection methods. This also means that, unless these time-series properties' are adequately accounted for in the regime shift detection algorithms, detected shifts may be the result of misdiagnoses (Rudnick and Davis 2003, Di Lorenzo and Ohman 2013, Doney and Sailley 2013).

The difficulties in defining what constitutes a regime shift are also reflected in the diversity of detection methods (e.g. Rodionov 2005, Andersen et al. 2009). Methods often rely on somewhat unique definitions of what constitutes a changepoint or regime shift and these definitions are rarely provided explicitly. This makes the comparison between different methods difficult and means that confirming the presence or absence of a regime shift in an ecological system depends, to a large extent, on the methodological choices and associated underlying assumptions. Detection methods can be broadly separated into two groups: those designed for univariate time series and those designed for multivariate time series. In this study, we will refer to the former as *changepoint* and to the latter as *regime shift* detection methods.

In recent years, regime shifts have been reported in a wide range of ecosystems from most regions of the world's oceans (e.g. Scheffer et al. 2001, Biggs et al. 2012, Möllmann et al. 2015): in the Pacific, most notably studied in the seminal paper by Hare and Mantua (2000); in the Atlantic, by Beaugrand et al. (2008); and in the Mediterranean Sea, by Conversi et al. (2010) and Damalas et al. (2021). Beaugrand et al. (2015) have found evidence to suggest quasi-synchronous marine pelagic regime shifts between ocean basins in the Northern Hemisphere. Numerous studies have reported evidence of regime shifts in the North Sea (Reid et al. 2001, Beaugrand 2004, Sguotti et al. 2022), along the South Norwegian coast (Frigstad et al. 2013), and along the Greenland coast (Heide-Jørgensen et al. 2023). Only recently has such evidence been reported for the Norwegian Sea (NoS) ecosystem (Vollset et al. 2022). The NoS, stretching along the west coast of Norway, is a highly productive area of great importance to regional economies. It is characterised by a relatively low species diversity and simple food webs (Skjoldal et al. 2004). The pelagic fish community is dominated by three species: blue whiting (*Micromesistius poutassou*), mackerel (*Scomber scombrus*), and herring (*Clupea harengus*), and the zooplankton community is dominated mainly by copepods, krill, and amphipods (Planque et al. 2022). Atlantic salmon is also present in the area during its marine phase. Recently, Vollset et al. (2022) conducted an analysis of the temporal patterns of marine growth in Atlantic salmon originating from Norwegian rivers, along with physical and other biological time series. Based on these data, they reported a regime shift that occurred in 2005, using a method called EnvCpt, from an R package by the same name. The identified shift was supported by a sudden reduction in salmon growth, concurrent with a warming event and an apparent 50% reduction in zooplankton abundance. Because regime shift detection can be highly dependent on the chosen methods and the intrinsic properties of ecological time series, it is unclear whether this reported shift reflects a genuine catastrophic change in the dynamics of the NoS ecosystem, or whether it is part of the natural variability in climate and ecological components within the same regime. One way to address this question is to assess the performance of changepoint and regime shift detection methods applied to the NoS time series.

The objective of this study is to assess the evidence of a regime shift in the NoS. For this purpose, we evaluate the performance of several methods commonly used to detect regime shifts and changepoints. For each method, we quantify the rate of false detection (i.e. when a regime shift or a changepoint is detected but is not present in the underlying dataset), also known as false-positive rate (FPR). A false positive is a type I error and low FPR is a desirable feature of a detection—or diagnosis—method. FPRs of 5% or less are commonly accepted and here we consider that there is evidence for a regime shift in the NoS if one is detected with a method that has an FPR rate below the acceptable threshold. FPR is prioritised over the false-negative rate (based on the number of times an existent changepoint is missed) due to the prevalence of literature reporting the presence of regime shifts and thus more susceptible to false positivity.

Materials and methods

The data: the NoS ecosystem time series

Thirty-two time series relating to the NoS ecosystem were analysed (see Supplementary Information). These series are used by the ICES Working Group on Integrated Assessments of the Norwegian Sea (ICES 2023b). The variables considered range from temperature and salinity measurements, primary production, and zooplankton biomass through to fish population metrics. Together, they provide a comprehensive depiction of variations in the NoS's physical and ecological characteristics over time. In this study, we present six time series that are illustrative of the diversity of the variables available. Additional series were included for the multivariate method and the results of the analyses on all the time series are available in the Supplementary Information. The six selected time series are the Arctic water (AW) index, the North Atlantic Oscillation (NAO) index, temperature in the northwest of Iceland (Temperature), Net primary productivity (NPP), total zooplankton biomass, and total herring biomass.

The AW index is a measure of the volume of water in 10^4 km^3, with a salinity of <34.9 between the depths of 150 and 300 m in the Norwegian Basin. The index reflects the proportion of Arctic versus Atlantic water masses, which affects the abundance and distribution of boreal and Arctic species (ecosystem overview, ICES 2021). The NAO index is an indicator of atmospheric and climatic variability over the North Atlantic Ocean. There are several NAO indices available and we refer here to the index provided by the University of East Anglia (Jones et al. 1997) and spans from 1907 to 2021. The NAO index is known to be connected to a range of biological processes on land and in the ocean (Ottersen et al. 2001). Sea temperature (°C) is measured at the Langanes sampling station (off the north-east coast of Iceland) and calculated as the average temperature between 80 and

120 m depth, in May or June (variable time of sampling). This time series started in 1953, though due to the abundance of absent data, here the time series was taken from 1973 onwards. Net primary production, in $\mathrm{gC\,m^{-2}\,yr^{-1}}$, is calculated based on weekly estimates of photosynthetic rates derived from satellite observations (Behrenfeld and Falkowski 1997) provided by Oregon State University and started in 2003. The total zooplankton biomass time series in the NoS, in $\mathrm{DWg\,m^{-2}}$ is calculated from plankton net sampling conducted during the International ecosystem survey in the Nordic Sea (IESNS, Rybakov et al. 2014) operated annually in May to June (WGINOR, ICES 2023b) since 1995. Total herring biomass in millions of tonnes is taken from yearly stock assessment estimates (ICES 2023a, c), which have been ongoing since 1907.

These six time series have different lengths and properties and thus, represent a diversity of different cases: the AW index and the total herring biomass time series are strongly autocorrelated and have very little noise. In contrast, the NAO index and temperature time series are much noisier and do not appear to be strongly autocorrelated. The latter seems to exhibit a slight upward trend (Fig. 1). The total primary production and total zooplankton biomass time series display intermediate situations and are both relatively short compared to the others, with the total primary production time series spanning only 19 years.

Changepoint and regime shift detection methods

Three commonly used changepoint (Strucchange, STARS, and EnvCpt) and one regime shift (Chronological Clustering) detection methods were selected. The R code and the data used for these analyses on the NoS time series are available on GitHub at https://github.com/hannoo73/ICES_JMS-RS-detection-methods.

STARS: a sequential *t*-test analysis of regime shifts

STARS is a statistical method that examines discontinuities in univariate time series (Rodionov 2004). In this approach, a regime shift index (RSI) is computed and used to evaluate whether each data point belongs to a regime distinct to that of the previous data points. The method requires the input of a cut-off length of regimes, here set to 10. By calculating P-values for each changepoint, the algorithm identifies significant shifts (P-value ≤ 0.05) that reflect meaningful transitions. The *Rshift* R package is used here (Room et al. 2023). Despite criticism of its sensitivity to red noise (Rudnick and Davis 2003, Rodionov 2006), the method has been widely used in ecological research (Lindegren et al. 2010, Morse 2017, Tomczak et al. 2022) and as such selected to be evaluated here.

Strucchange: structural change in linear regression models

The Strucchange method, from the *Strucchange* R package (Zeileis et al. 2002), is a univariate statistical changepoint detection method. This method allows us to compare the fits of both a linear regression model and a piecewise linear regression model to time series, and tests each time point as a potential breakpoint. The best model is selected based on out-of-sample prediction performance using the Bayesian information criterion (BIC). The changepoint is then associated with a confidence level, in the form of a P-value (Damalas et al. 2021); only one changepoint can be identified at a time with this package. A P-value of 0.05 was used to determine the significance of a changepoint. This method is quite often used on ecological time series and has been suggested as a reference method in the review of approaches in identifying regime shifts by Andersen et al. (2009).

EnvCpt: detection of structural changes in climate and environment

The EnvCpt method (*EnvCpt* R package, Killick et al. 2022) is based on the comparison of multiple models fitted to the same time series. The following eight models were fitted here (Fig. 2): (i) constant mean and variance (Mean); (ii) piecewise constant mean and variance (Mean + changepoint); (iii) constant mean with autocorrelated errors [Mean + AR (1)]; (iv) piecewise constant mean with autocorrelated errors [Mean + AR (1) + changepoint]; (v) linear trend (Trend); (vi) piecewise linear trend (Trend + changepoint), similar to that of the Strucchange model (see above); (vii) linear trend with autocorrelated errors [Trend + AR (1)]; and (viii) piecewise linear trend with autocorrelated errors [Trend + AR (1) + changepoint].

This method was originally developed for climate and environmental time series and is now widely used in ecology. This method was used by Vollset et al. (2022), to affirm the presence of a regime shift in the NoS, and the methodology used by this paper is reproduced here. As with the Strucchange method, BIC scores are generated and used to select the best fitting model. A difference of three BIC score units between two models is considered significant here, as in Vollset et al. (2022).

In many cases, the difference in BIC score between the two best fitting models was <3 points, meaning that the BIC score could not separate the two models performances', thus giving an inconclusive result. However, as the question here was not to determine which model best fits the data but rather whether a changepoint model describes the data better than a non-changepoint model, a further investigation was done. Different cases of inconclusive results were separated, those where the conflict arose between non-changepoint models, between changepoint models, and between one non-changepoint and one changepoint model.

To simplify the presentation of the results, the different possibilities are grouped together into three different categories of results: Changepoint, the significantly best fitting model was one with one or more changepoints, to which were added the inconclusive cases where a conflict occurred only between changepoint models (e.g. between the piecewise linear trend and the piecewise constant mean and variance models); No changepoint, the significantly best fitting model was one with no changepoints plus the inconclusive cases where conflicts arose between non-changepoint models (e.g. the linear trend and the constant mean and variance models); and Inconclusive, which refers to the cases where no one model best fits the surrogate time series and the conflict was between changepoint and non-changepoint models. An inconclusive result does not mean that there was no changepoint but only that the time series is equally well (or equally poorly) explained by a model with a changepoint than a model without. As the objective here is to test this method, this uncertainty will be considered here like a potential false positive. Furthermore, an inconclusive result does not contribute to our ability to evaluate the method and high inconclusive rates could translate as a lack of power of the method.

Figure 1. Time series of the six variables used as examples throughout the manuscript. From top to bottom: the Arctic water index, the NAO index, the sea temperature measured at the Langanes station, the total primary production, the total zooplankton biomass, and the total herring biomass time series.

Chronological Clustering: a constrained clustering method

Chronological Clustering (CC) was developed by Legendre et al. (1985) to study ecological succession. The method is derived from standard hierarchical cluster analysis, and is designed for ordered data. By grouping data points into temporally coherent groups, it allows for the identification of temporal discontinuities in multivariate time series (Legendre et al. 1985). This method works by combining observations into clusters based on permutation tests using predefined significance and connectedness levels (Legendre et al. 1985, Damalas et al. 2021).

CC is considered flexible, as it does not impose a functional form between state and driver variables (Perretti et al. 2017) but rather identifies concurrent changes between time series, which matches the definition of a regime shift used here. The

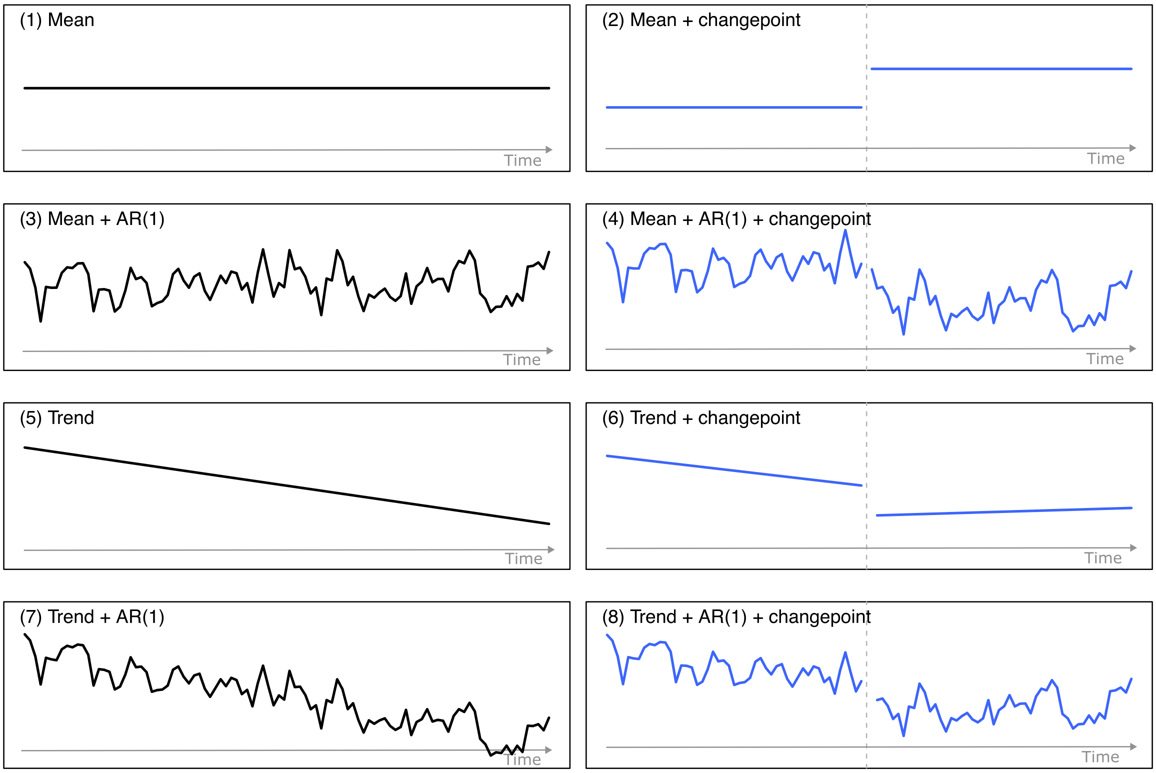

Figure 2. Schema of the eight different models fitted to the time series in the EnvCpt method. In the left column are the models without changepoints and in the right column are the models with changepoints. Model (1) is a constant mean and variance model, model (2) is a piecewise constant mean and variance model, model (3) is a constant mean with autocorrelated errors model, model (4) is a piecewise constant mean with autocorrelated errors model, model (5) is a linear model, model (6) is a piecewise linear model, model (7) is a linear model with autocorrelated errors, and model (8) is a piecewise linear model with autocorrelated errors.

method was not originally designed for regime shift detection but it is commonly used for that purpose (Weijerman et al. 2005, Andersen et al. 2009, Heymans and Tomczak 2016, Morse 2017, Perretti et al. 2017). Being a clustering technique, the method will always identify shifts (or breaks) in the series. Consequently, it is recommended for use in conjunction with other methods (Möllmann and Diekmann 2012) and to select the optimal number of significant clusters with, e.g. a broken stick model approach (MacArthur 1957, Bennett 1996). This is done by comparing the variance explained by a random splitting of the data and the variance explained by different numbers of clusters found by the CC method. If the optimal number of clusters is higher than one, then this suggests the presence of a regime shift in the multivariate time series. In this study, we constructed 6 different types of clustering, based on a selection of the 32 available time series, though a maximum of 19 series were included in the biggest grouping as some time series are redundant. The groupings of time series are, first, defined based on the type of variables included: physical ('Phy'), ecological ('Bio'), or combined physical and ecological time series ('Combi'). Second, they were based on the time-series length: 'Long' (1976–2021), 'Med' (1995–2021), or 'Short' (2003–2021). The time series included in each of the clustering types are detailed in Table 1. Clusterings were performed on euclidean distance matrices using the chclust function from the *rioja* R package (Juggins 2023). The constrained integrated sum of squares agglomeration algorithm ('coniss' method) was used.

Method testing: the null hypothesis and surrogate time series

Changepoint and regime shift detection methods aim to answer the question: 'Is there one or several changepoints or regime shifts?' To which there is a binary response, 'Yes' (there is one or several changepoints or shifts) or 'No' (there are none). This response is provided with a level of confidence, usually estimates of type I and II errors. The two types of error correspond to false positives and false negatives, respectively. False negatives occur when the test fails to detect a changepoint or regime shift that actually exists in the data. Conversely, false positives (type I errors) occur when the test erroneously indicates the presence of a changepoint or regime shift in a time series when there isn't one. The true positive rate refers to the probability of a test correctly detecting a phenomenon. This is also termed sensitivity and is equal to $1 -$ false-negative rate. In contrast, the true negative rate is the probability of a test correctly identifying the absence of a phenomenon. This is also termed specificity and it is calculated as $1 -$ FPR.

There is a bias in scientific publication towards positive results that can arise from researcher psychological bias, encouraged by the competitive publishing environment, or by inappropriate use of statistical methods and interpretation of results (Forstmeier et al. 2017). Thus, false positives are more likely to be reported than false negatives; as such, the focus here is to evaluate the specificity of the previously presented changepoint and regime shift detection methods. One approach to assess the specificity of a changepoint or

Table 1. Summary of the six grouping of variable used to test the Chronological Clustering method.

Time series Start End	PhyLong 1976 2021	PhyMed 1995 2021	BioMed 1995 2021	BioShort 2003 2021	CombiMed 1995 2021	CombiShort 2003 2021
NAO index*	x	x			x	x
Sub-polar gyre index	x	x			x	x
Norwegian-Lofoten gyre index	x	x			x	x
Temperature (Svinoy)	x	x			x	x
Salinity (Svinoy)	x	x			x	x
AW index*	x	x			x	x
Relative heat content	x	x			x	x
Relative freshwater content	x	x			x	x
Temperature (Langanes)*	x	x			x	x
Salinity (Langanes)	x	x			x	x
Total zooplankton biomass*			x	x	x	x
Herring recrutement (age 2)			x	x	x	x
Blue whiting recrutement (age 1)			x	x	x	x
Mackerel recrutement (age 2)			x	x	x	x
Total herring biomass*			x	x	x	x
Total blue whiting biomass			x	x	x	x
Total mackerel biomass			x	x	x	x
Total primary production*				x		x
Day of peak primary production				x		x

The start and end years for the different groups are stated to indicate the variable lengths of the formed groups. The * correspond to the time series presented in Fig. 1 and the Xs indicate the time series included in each grouping..

regime shift method, is to apply them on control time series that exhibit properties similar to real-world time series but are devoid of any changepoints. One such type of simulated time series are surrogate time series, which are simulated time series that have the same number of observations, same linear trend, and same autocorrelation as the original data (ICES 2022). It is important to note that the conservation of the original linear trend does not lead to the transference of changepoints from the original time series to the surrogates, as only a simple non-piecewise trend is conserved no matter the presence of a piecewise trend. In the case of a piecewise trend in the time series, the methodology employed would abstract only an average of these two trends. Surrogate time series have been used to test the null hypothesis, as in Planque and Buffaz (2008), and are generated with a method called phase randomisation (Theiler et al. 1992, Schreiber and Schmitz 2000). In addition to this, our approach necessitated further data transformation (see section below). By design, these surrogates do not contain changepoints and all fluctuations can be attributed to the underlying stochasticity, autoregressive processes, and trend. Surrogates times are sensitive to the length of the time series, and the use of longer series is preferable; however, there is no better alternative that could be used to this analysis. We generated 1000 surrogates, with the *tseries* R package (Trapletti et al. 2023), for each observational time series to evaluate the specificity of changepoint and regime shift detection methods. The proportion of surrogate time series for which a positive result of the test is found, corresponds to the FPR and can be used to estimate the method's specificity. A changepoint or regime shift detection method with a high specificity should not detect changepoints or regime shifts in any of the surrogate time series (FPR = 0, specificity = 1). A 5% FPR is used as a reference to evaluate the acceptability of a method's specificity.

Data transformation: the omnibus normalisation technique

Normal distribution of the data is a necessary precondition for generating surrogate time series, and for several regime shift detection methods, but many of the observational time series used in this study were not normally distributed. To resolve this issue, we used a robust normalisation technique called omnibus normalisation (Legendre and Legendre 2012). The principle of the omnibus transformation is to tie the original observational dataset to a normally distributed dataset of the same size, using a 'random' number generator. Here, we use a modified version of the method in which the normally distributed dataset is formed by a regular sequence of quantiles in a normal distribution (see Supplementary Information for more details). All series were transformed using omnibus normalisation before constructing the surrogates and the tests were applied on the normalised surrogates. For the application of the different changepoint and regime shift detection methods on the original time series, normalised time series were also used.

Results

Application of the changepoint and regime shift detection methods on the NoS time series

The results of the different changepoint detection methods (Table 2) show that for five out of the six series presented here, at least one changepoint was found by one of the three univariate methods. Both the STARS and the Strucchange methods did not identify changepoints for the total primary production time series, the shortest time series here. However, it received an inconclusive result from the EnvCpt method, meaning that changepoint and non-changepoint models were found to fit equally well. Furthermore, for the NAO, herring biomass, and temperature time series, the STARS method found multiple changepoints. In contrast, the EnvCpt method found only inconclusive results, except for the herring biomass time series,

Table 2. Summary of the changepoint years encountered on each of the six normalised time series with STARS, Strucchange, and EnvCpt detection methods.

Time series	STARS	Strucchange	EnvCpt
AW index	2001	2002	2015
NAO index	1962/1989/1996	–	Inconclusive
Temperature (Langanes)	1984/1988/2003	2002	Inconclusive
Total primary production	–	–	Inconclusive
Zooplankton biomass	2004	2004	Inconclusive
Herring biomass	1924/1929/1943/1951/1959/1967/1988/1997	1960	1976/2006

'–' indicates that no changepoints were found. For the EnvCpt results, 'Inconclusive' indicates that the method cannot distinguish between a model with or without a changepoint.

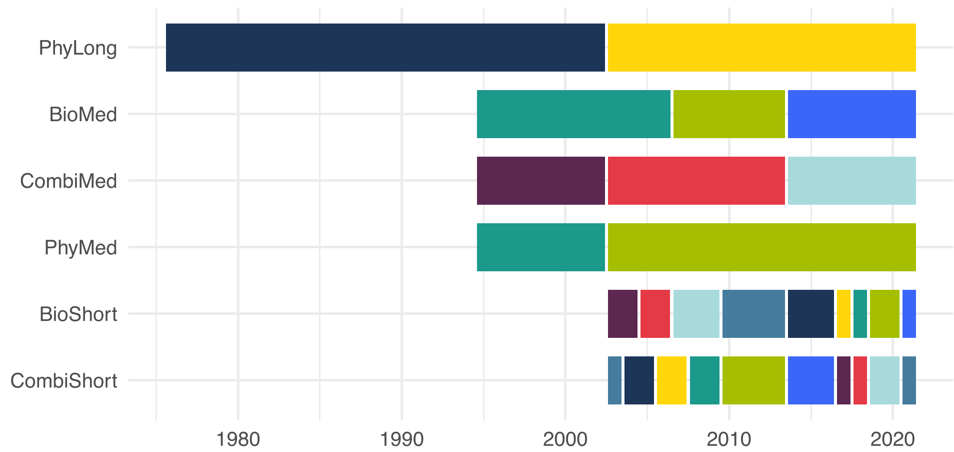

Figure 3. Representation of the different clusters (i.e. regimes) found by the Chronological Clustering method in the six different time-series groupings. The different colours indicate the different regimes.

where two changepoints were found and the AW index, where only one was found. As for the Strucchange method, where the number of changepoints identified in a time series is limited to one, changepoints were identified in four out of the six series, the timings of which are similar to those of the STARS method. Overall, changepoints were most frequently found in the early 2000s, though the exact timing differs, which would indicate a possible regime shift present in the physical time series and felt across ecological compartments; however, the results from the subsequent specificity testing will give indications about the validity of these results. The results of the testing of all 32 NoS time series can be found in the Supplementary Information (Supplementary Tables S1, S3, and S6).

For the multivariate CC approach (Fig. 3), at least 2 significantly different clusters and up to 10 different clusters were found in the 6 different groupings, which suggests the presence of at least 1 regime shift in all groupings. Within the PhyLong, CombiMed, and PhyMed groupings, a concomitant shift was found in 2003, which was not detected with BioMed grouping. Within the BioMed and CombiMed groupings, a concomitant shift in 2014 was detected, which does not appear in the PhyLong and PhyMed groupings. For the shorter groupings, BioShort and CombiShort, which span only 19 years, 10 clusters were found, most of which contain only 1 year, rendering the results uninterpretable. Here, the optimal number of clusters found seems to increase with decreasing time-series lengths.

Testing changepoint and regime shift detection methods with surrogate time series

STARS

With the STARS method, changepoints were identified at a rate exceeding the acceptable threshold of 5% of false positives within the surrogates simulated for all the 32 NoS time series tested (Supplementary Table S2). Among the six example series (Fig. 4), this rate exceeded 70% in most cases, with the exception of primary production, in which 23% of the surrogates presented a changepoint. The STARS method also found, in some cases, a large number of changepoints by surrogate time series. This is the case for the surrogate time series of total herring biomass where up to 13 changepoints were found (Supplementary Table S2).

Strucchange

With the Strucchange method, changepoints were also identified at a rate exceeding the acceptable threshold of 5% of false positives within the surrogates simulated for all the 32 NoS time series tested (Supplementary Table S4). Among the six example time series (Fig. 5), the lowest rate of false positives was found for the shortest time series (primary production, 35% of FPR) and the highest for the longest series (total herring biomass, 99% of FPR).

EnvCpt

The results for EnvCpt are classified here, into the three categories (Fig. 6): Changepoint, the cases where the signifi-

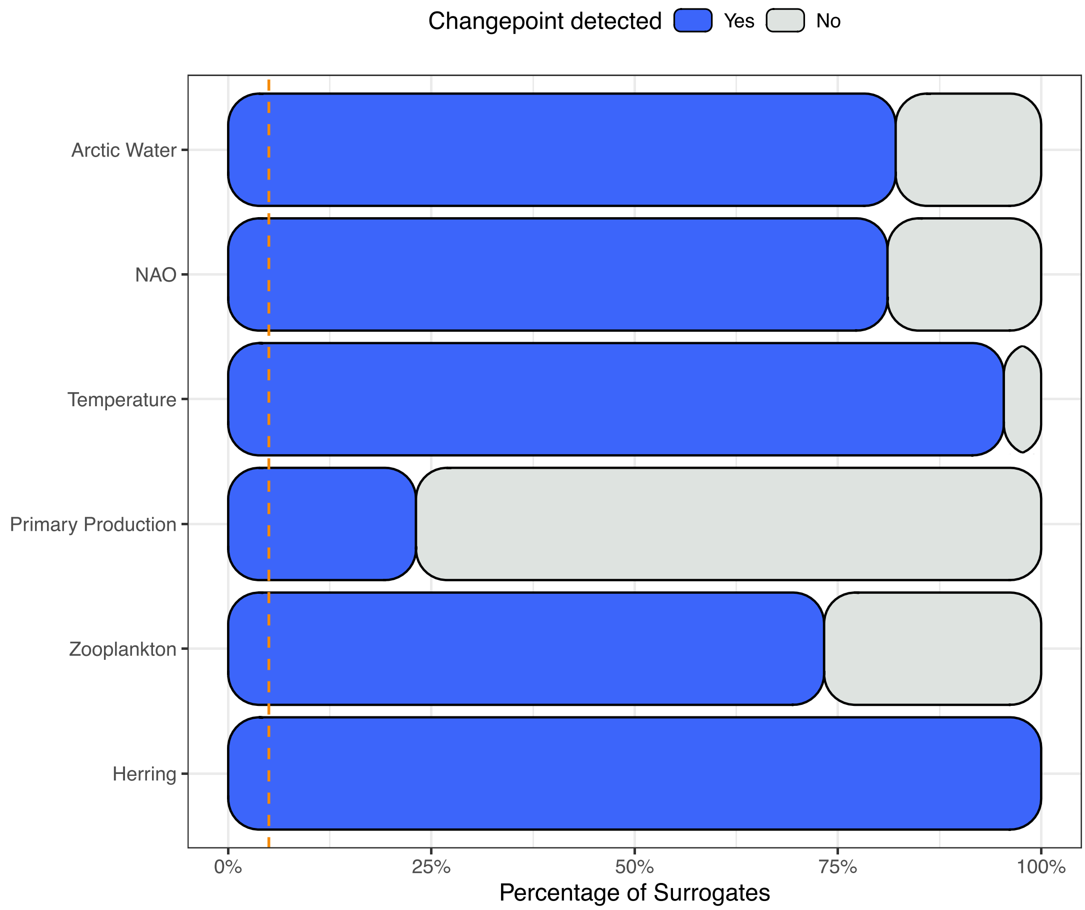

Figure 4. The STARS method's specificity: percentage of surrogates, for each of the six surrogate time-series sets, in which at least one changepoint is detected (i.e. false positive, in dark blue, left) and percentage of surrogates in which no changepoint is detected (i.e. true negative, in light grey, right). The broader the blue bar, the lower the specificity. The orange dotted line depicts the 5% threshold of acceptable false positives.

cantly best fitting model or models to the surrogate time series were those with changepoints, indicating a false positive; No changepoint, the significantly best fitting model or models were those with no changepoints; and Inconclusive, where models with and without changepoint were judged to fit equally well (ΔBIC < 3). Among the six example time series, two (NAO and primary production) presented a rate of false positives below the acceptable threshold of 5%, while the rest of the series presented rates between 5 and 25%. However, the rate of inconclusive results is notably high for NAO and primary production (97 and 90%, respectively). Indeed, the inconclusive category represents the highest rate in all 32 series, with an average of 81% and a minimum value of 55%. The results for the other NoS time series are available in the Supplementary Information (Supplementary Tables S5-S7).

Chronological Clustering

With the CC method, an optimum number of clusters is defined. A number of clusters higher than one suggests the presence of at least one regime shift. As is shown in Fig. 7, the percentage of cases for which the optimal number of clusters was higher than one was between 10 and 98% of the multivariate surrogates time series, respectively for the CombiMed and the CombiShort time-series groupings. Thus, for all groupings, FPRs are well above the acceptable error rate threshold of 5%, with only CombiMed <20% FPR. However, it is worth noting that longer time-series groupings seem to fare better than the shorter time-series groupings, as is evidenced by the two Short groupings having both >90% FPRs, though CombiMed and PhyMed have much lower FPRs than PhyLong, despite the fact they have 21 fewer data points in their time series.

Discussion

In this study, we have evaluated four commonly used methods of changepoint and regime shift detection for their specificity. For this purpose, we have used a dataset of 32 time series describing different components of the NoS ecosystem, though the results of only 6 are presented here. The performance of these methods is evaluated, simulating 1000 surrogate time series that serve as a null hypothesis for the absence of a regime shift or changepoint as they share mean, variance, and autocorrelation with the original time series, but do not contain changepoints. The four changepoint/regime shift detection methods, STARS, Strucchange, EnvCpt, and CC, were applied to the NoS time series and to all surrogate time series. The probability of detecting a regime shift under the null hypothesis (type I error) is interpreted as a measure of the specificity of the method (1 − FPR). Our results indicate support for a regime shift in the NoS in the early 2000s based on the direct application of these methods on the time series, but all four methods have FPRs far exceeding the acceptable 5% er-

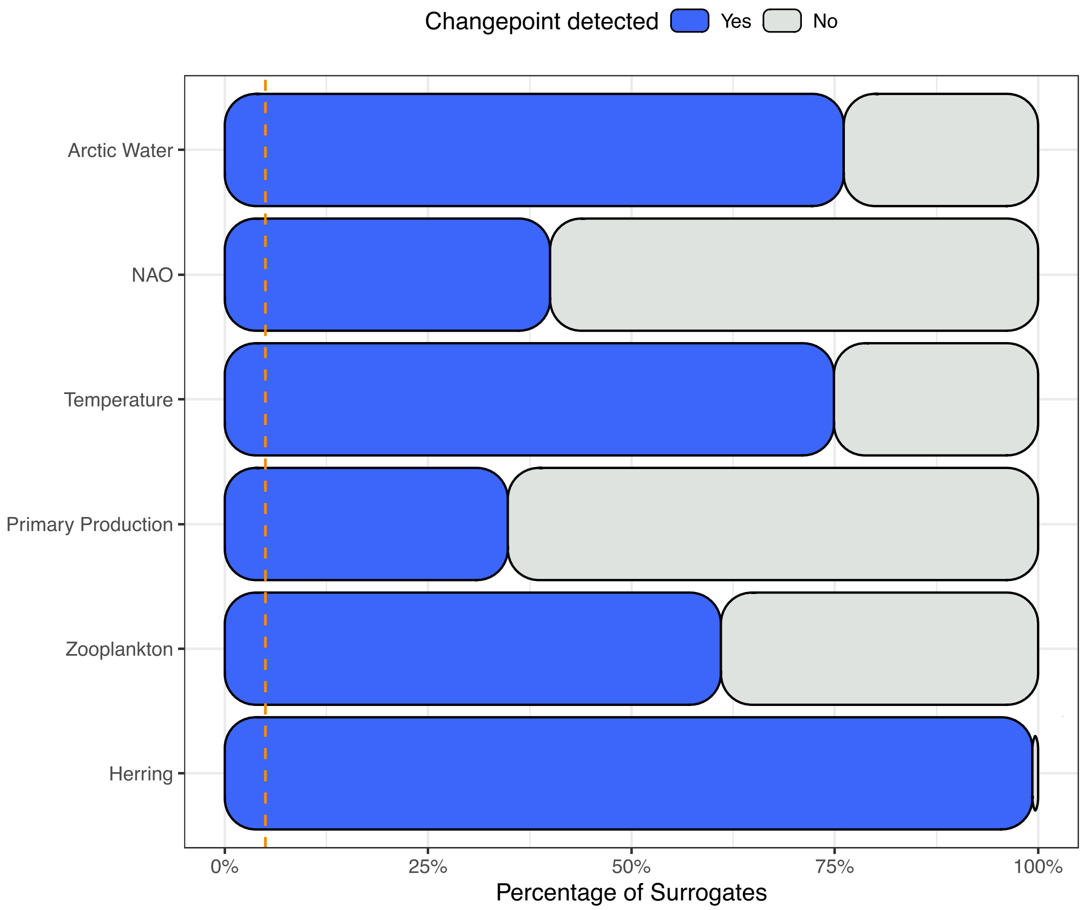

Figure 5. The Strucchange method's specificity: percentage of surrogates, for each of the six surrogate time-series sets, in which a changepoint is detected (i.e. false positive, in dark blue, left) and percentage of surrogates in which no changepoint is detected (i.e. true negative, in light grey, right). The broader the blue bar, the lower the specificity. The orange dotted line depicts the 5% threshold of acceptable false positives.

ror rate; for most, even accepting 20% error rates would not be sufficient to approve these methods. These low specificities mean that erroneous detections of changepoints or regime shifts are highly likely. Indeed, in some cases, a method will find changepoints 100% of the time. Consequently, it is impossible to draw any firm conclusions regarding the presence of a regime shift in the NoS. We cannot definitively confirm the presence of a regime shift, nor can we completely dismiss the possibility, based on these methods. These findings align with the results of the workshop on integrated trend analysis (ICES 2022), which reviewed several time-series analysis methods and found that many of these can have high FPRs.

While none of the tested methods achieve satisfactory FPRs, at first glance, the EnvCpt approach outperforms the others. Across all 32 available time series, it exhibits an average FPR of 10.8% (Supplementary Table S5), which increases slightly to 12.6% for the 6 time series included here. Considering false positives as only the cases where changepoint models outperform non-changepoint models, we can say that the EnvCpt method is less likely to produce false positives than the other three methods analysed here, though on average it still produces twice the acceptable amount of false positives. However, far outnumbering the false positives and the true negatives, with an average of 81% across the 32 time series available (Supplementary Table S5), the inconclusive results must nuance this conclusion. As the predominant result of applying the EnvCpt method to all surrogate test sets, the inconclusive cases must be taken into account and means this method cannot reliably detect the presence or absence of regime shifts. This means that this method is more likely than not to give an uncertain and unactionable result, answering 'I don't know' to the question 'Is there a changepoint?'. While it is comforting that EnvCpt can explicitly handle uncertainty in changepoint detection, it does not help answer the question posed. A closer examination of the types of models that often enter into conflict (i.e. that have overlapping BIC scores) shows that AR (1) and changepoint models of both the mean and trend model types are often indistinguishable. Incorporating autocorrelation into the fitted models is a key strength of the EnvCpt method, but distinguishing between autoregressive processes and actual changepoints remains challenging, an observation i.e. in line with earlier results (Overland et al. 2006).

In addition, it is plausible that the issue lies in the way the models are compared, specifically through the BIC scores. This model selection criterion has been criticised for overpenalising complex models compared to simpler ones and being too sensitive to the sample size, or here the length of the time series, and selecting simpler models for longer time series, whereas overly complex models may be favoured for shorter time series (called overfitting) (Burnham and Anderson 2004). Furthermore, BIC scores only compare models among those tested; as such, the lowest scoring model may still inade-

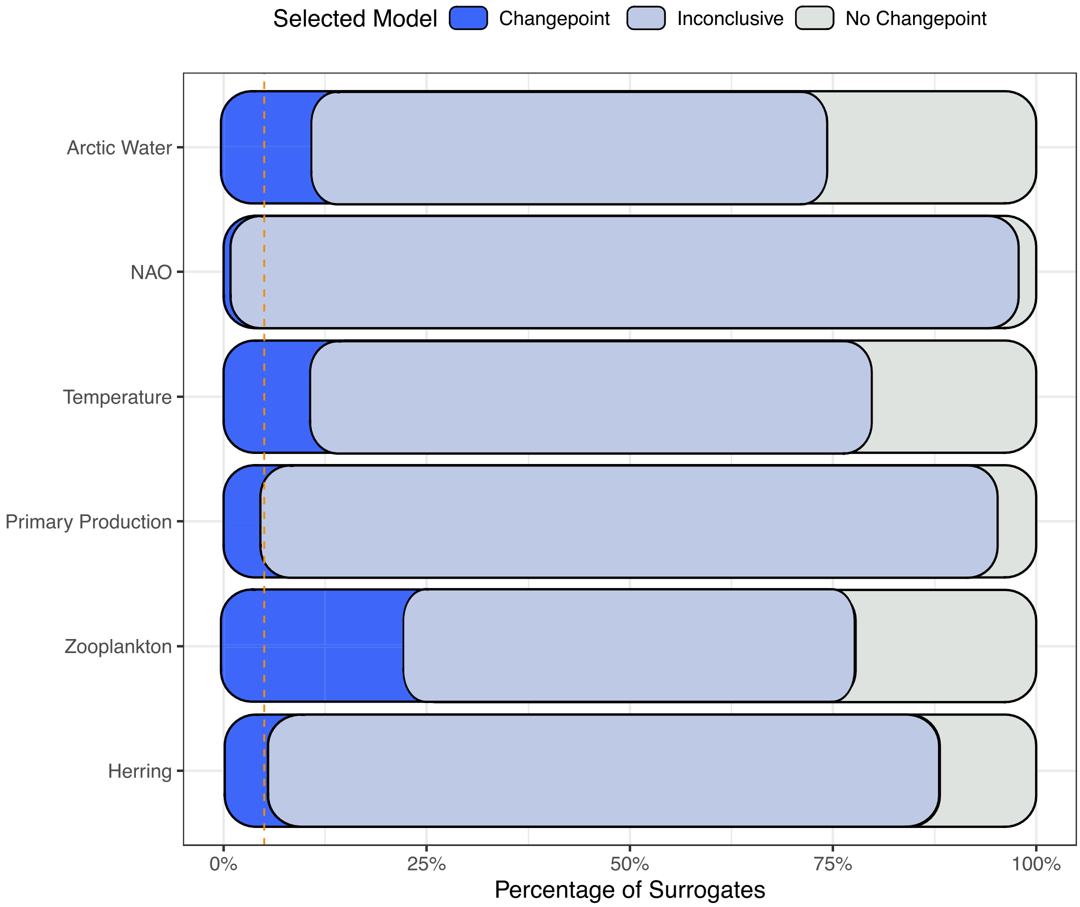

Figure 6. The EnvCpt method's specificity: percentage of surrogates, for each of the six surrogate time-series sets, for which a changepoint model is the best fitting model, including cases where no one changepoint model among changepoint models could be distinguished (i.e. false positive, in dark blue, left) and percentage of surrogates in which no changepoint is detected (i.e. true negative, in light grey, right). The remaining surrogates are those where no one model could be selected and the equally best fitting models were both changepoint and non-changepoint models (i.e. inconclusive results, in pale blue, middle). The orange dotted line depicts the 5% threshold of acceptable false positives.

quately reflect the true patterns of the modelled time series but is still nonetheless selected (Burnham and Anderson 2004). However, this approach was chosen to be comparable to that of Vollset et al. (2022). Finally, the difficulty in distinguishing between autoregressive processes and actual changepoints may reside with the time series themselves, which are, in the case of ecological time series, generally too short and too noisy and thus may contain too little information to be able to robustly detect changepoints.

The performance of the remaining methods, STARS, Strucchange, and CC, is notably poor, as evidenced by the presence of a 100% false-positive error rate for certain combinations of methods and sets of surrogate time series. The STARS method, in particular, exhibits the worst performance with on average across the 32 NoS time series, a 75% false-positive error rate. It is worth noting that the STARS method is widely employed in ecological research (Daskalov et al. 2007, Lindegren et al. 2010, Seddon et al. 2014, Heide-Jørgensen et al. 2023), and it seems quite likely that these high FPRs are present in most of these cases. This is also true of the Strucchange and CC methods, though only when CC is applied in regime shift research. Noisy time series such as the NAO and the temperature time series also seem to have high numbers of detected changepoints, indicating that STARS has difficulties distinguishing changepoints from white noise. STARS is also reportedly sensitive to red noise and autocorrelated processes, which would explain the extraordinary number of changepoints found in time series such as the total herring biomass time series. A solution has been suggested to correct for this called 'prewhitening' of time series, which removes red noise (low-frequency patterns) and thus dampens autocorrelation in the time series (Rodionov 2006); however, this process could dramatically modify the original time series and alter the overall question posed as the test would no longer look at changes in the time-series values, but rather changes in the year-to-year variations of the time series, i.e. the time series' derivative.

Recently, Stirnimann et al. (2019) tested the STARS method's performance using Autoregressive Integrated Moving Average (ARIMA) and normally distributed simulated time series. True positive rates were evaluated by quantifying STARS's detection of artificially included changepoints in the simulated time series and FPRs were evaluated by quantifying STARS's detection of changepoints that were not artificially added to the series. However, the evaluation did not include the type of negative controls used here and therefore did not quantify FPRs in series containing strictly no changepoints. Thus, the results are difficult to compare to ours. They did nevertheless note a better performance of STARS in the nor-

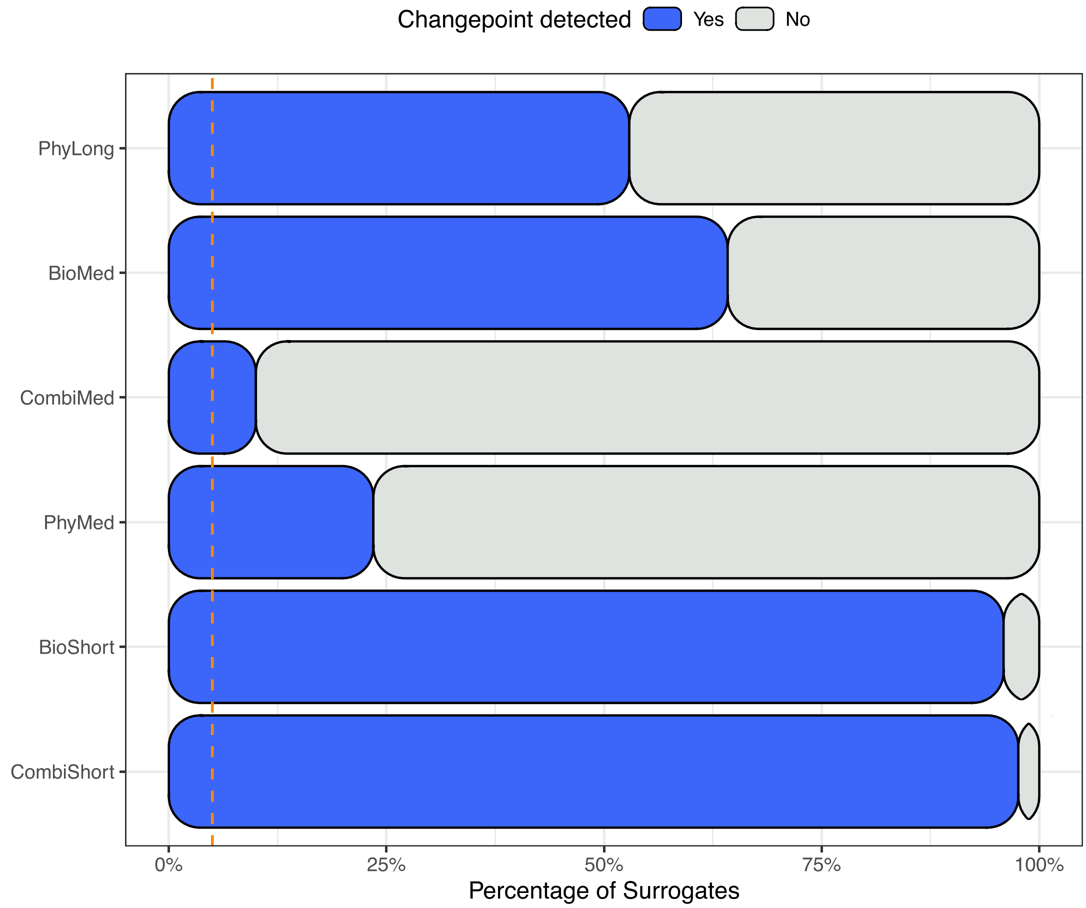

Figure 7. The Chronological Clustering method's specificity: percentage of surrogates, for each of the six surrogate time-series sets, in which two or more optimal clusters were selected and thus containing at least one regime shift (i.e. false positive, in dark blue, left) and percentage of surrogates in which only one cluster was optimal (i.e. true negative, in light grey, right) and, as such, no regime shifts were found. The broader the blue bar, the lower the specificity. The orange dotted line depicts the 5% threshold of acceptable false positives.

mal compared to the ARIMA time series, despite a prewhitening of the ARIMA time series, confirming that autocorrelation can disrupt this method.

While our analyses show that we can neither confirm nor deny the presence of a regime shift in the NoS on the basis of commonly used regime shift detection methods, this result is significant for our understanding of the NoS ecosystem dynamics. Furthermore, we hope to stimulate similar work to test the performance of other regime shift detection methods for other data series and other marine systems. One can reasonably expect that a substantial fraction of previously reported regime shifts, based on the method analysed here, may in fact be false positives, though determining the exact proportion would require dedicated analyses.

Misleading wording in the field of regime shifts is in line with the propensity to overdramatise results. Articles publishing negative results or simply the absence of positive results are virtually nonexistent, and with the necessity to publish regularly in the scientific community, the use of 'persuasive communication devices' is dangerously tempting and thus inescapable in scientific literature (Corneille et al. 2023). Regime shifts as a scientific area of research are unfortunately rife with confusion and inconsistencies of definitions and seem to demonstrate most of the problems presented by Corneille et al. (2023), especially the mischaracterisation of the state of the art, overselling by the use of excessive titles, and overgeneralisation. The complexity of the topic of regime shifts means that this may be done unconsciously.

A major issue with issuing such statements without explicitly studying and presenting associated uncertainties is that they can be subsequently taken up by others, overlooking the potential uncertainty (Hellenbrecht et al. 2023).

The next steps of this investigation would be to assess additional changepoint and regime shift detection methods in the hopes of finding a suitably reliable method to apply to the NoS time series. While we used the NoS ecosystem time series as examples here, the nature of these time series is not dissimilar to most ecological time series around the world and thus could also potentially present equally high FPRs. A future investigation into this assumption would be interesting. Additionally, for this research, we were and had to be rather specific in methodology but there are many other approaches available that also need to be evaluated in a similar manner as above, before being applied.

Here, the focus was on methods' specificity, or FPR detection. Evaluating false-negative rates, or sensitivity, would also be relevant in cases when regime shift may occur but would remain undetected due to a low sensitivity. To accomplish this, a clear mathematical definition of a changepoint or a regime shift would need to be considered, as done by Stirnimann et al. (2019), who use varying magnitudes of increased or decreased standard deviation (SD). Being able to assess both false-positive and false-negative rates would give a comprehensive picture of a method's overall performance. However, it is likely that the two will be inversely linked, meaning the fewer false changepoints detected, the more changepoints missed, and vice versa.

Determining false-negative detection rates is arguably more important from a conservation and a resource management perspective. Using a method with a high FPR and falsely finding a regime shift would logically, in a managed ecosystem, only lead to the allocation of funds to a relatively healthy or stable ecosystem and hopefully have no negative effect so long as the preventive measures themselves do not cause lastly damage. However, the possibility of missing a regime shift due to a method with a low sensitivity could have devastating and irreversible effects. This is perhaps a moot point though as it is nigh on impossible to accurately predict regime shifts before they happen.

While we do not deny the existence of abrupt regime shifts, our results indicate that the methods commonly used for detecting these abrupt shifts in marine ecological time series are unreliable. This is in line with the results of Hillebrand et al. (2020), who conclude that regime shifts or more precisely threshold transgressions are rarely detectable from empirical ecological observations.

Conclusion

We have assessed the specificity of four commonly used changepoint and regime shift detection methods for marine ecosystems. Our approach is based on the quantification of false-positive rates (FPR, the proportion of time a shift is detected when it actually does not exist) of selected time series for the Norwegian S ecosystem. We found none of the methods to be fit for purpose. All methods displayed FPRs that are far beyond the usually accepted 5% rate. We advise to rigorously evaluate the performance of any regime shift detection method in the context of individual case studies, before drawing robust conclusions. Earlier reviews of regime shifts across world's oceans should also be considered carefully given that the specificity of the methods used for these reviews has generally not been assessed.

Acknowledgements

This output reflects only the authors' view and the Research Executive Agency (REA) cannot be held responsible for any use that may be made of the information contained therein. The authors are grateful to the ICES Working Group on the Integrated Assessments of the Norwegian Sea (WGINOR) for providing the time series data. ChatGPT was used in the writing and editing of this paper and in the writing of the code for the analyses to help formulate ideas and streamline coding pipelines.

Author contributions

H.H.: writing – original draft and review & editing, conceptualisation and methodology, investigation, software, and visualisation. B.P. and L.B.: supervision, conceptualisation and methodology, writing – review and editing, and visualisation.

Supplementary data

Supplementary material is available at the *ICES Journal of Marine Science* online.

Conflict of interest: None declared.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 862428 (MISSION ATLANTIC).

Data availability

All the original data used in this study have already been published elsewhere. All data and R code used are available on GitHub at https://github.com/hannoo73/ICES_JMS-RS-detection-methods.

References

Andersen T, Carstensen J, Hernández-García E *et al.* Ecological thresholds and regime shifts: approaches to identification. *Trends Ecol Evol* 2009;**24**:49–57. https://doi.org/10.1016/j.tree.2008.07.014

Beaugrand G, Conversi A, Chiba S *et al.* Synchronous marine pelagic regime shifts in the Northern Hemisphere. *Philos Trans R Soc B Biol Sci* 2015;**370**:20130272. https://doi.org/10.1098/rstb.2013.0272

Beaugrand G, Edwards M, Brander K *et al.* Causes and projections of abrupt climate-driven ecosystem shifts in the North Atlantic. *Ecol Lett* 2008;**11**:1157–68. https://doi.org/10.1111/j.1461-0248.2008.01218.x

Beaugrand G. The North Sea regime shift: evidence, causes, mechanisms and consequences. *Prog Oceanogr* 2004;**60**:245–62. https://doi.org/10.1016/j.pocean.2004.02.018

Behrenfeld MJ, Falkowski PG. Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnol Oceanogr* 1997;**42**:1–20. https://doi.org/10.4319/lo.1997.42.1.0001

Bennett KD. Determination of the number of zones in a biostratigraphical sequence. *New Phytologist* 1996;**132**:155–70. https://doi.org/10.1111/j.1469-8137.1996.tb04521.x

Biggs R, Blenckner T, Folke C *et al.* Regime shifts. In: *Encyclopedia of Theoretical Ecology*. Ewing, NJ: University of California Press, 2012, pp. 609–17.

Burnham KP, Anderson DR. Multimodel inference: understanding AIC and BIC in model selection. *Sociol Method Res* 2004;**33**:261–304. https://doi.org/10.1177/0049124104268644

Cheal AJ, MacNeil MA, Cripps E *et al.* Coral-macroalgal phase shifts or reef resilience: links with diversity and functional roles of herbivorous fishes on the Great Barrier Reef. *Coral Reefs* 2010;**29**:1005–15. https://doi.org/10.1007/s00338-010-0661-y

Conversi A, Fonda Umani S, Peluso T *et al.* The Mediterranean Sea regime shift at the end of the 1980s, and intriguing parallelisms with other European basins. *PLoS One* 2010;**5**:e10633. https://doi.org/10.1371/journal.pone.0010633

Corneille O, Havemann J, Henderson EL *et al.* Beware 'persuasive communication devices' when writing and reading scientific articles. *eLife* 2023;**12**:e88654. https://doi.org/10.7554/eLife.88654

Damalas D, Sgardeli V, Vasilakopoulos P *et al.* Evidence of climate-driven regime shifts in the Aegean Sea's demersal resources: a study spanning six decades. *Ecol Evol* 2021;**11**:16951–71. https://doi.org/10.1002/ece3.8330

Daskalov GM, Grishin AN, Rodionov S *et al.* Trophic cascades triggered by overfishing reveal possible mechanisms of ecosystem regime shifts. *Proc Natl Acad Sci USA* 2007;**104**:10518–23. https://doi.org/10.1073/pnas.0701100104 de Young B, Harris R, Alheit J *et al.* Detecting regime shifts in the ocean: data considerations. *Prog Oceanogr* 2004;**60**:143–64. https://doi.org/10.1016/j.pocean.2004.02.017

Di Lorenzo E, Ohman MD. A double-integration hypothesis to explain ocean ecosystem response to climate forcing. *Proc Natl Acad Sci* 2013;**110**:2496–9. https://doi.org/10.1073/pnas.1218022110

Doney SC, Sailley SF. When an ecological regime shift is really just stochastic noise. *Proc Natl Acad Sci* 2013;**110**:2438–39. https://doi.org/10.1073/pnas.1222736110

Forstmeier W, Wagenmakers EJ, Parker TH. Detecting and avoiding likely false-positive findings—a practical guide. *Biol Rev* 2017;**92**:1941–68. https://doi.org/10.1111/brv.12315

Frigstad H, Andersen T, Hessen DO *et al.* Long-term trends in carbon, nutrients and stoichiometry in Norwegian coastal waters: Evidence of a regime shift. *Prog Oceanogr* 2013;**111**:113–24. https://doi.org/10.1016/j.pocean.2013.01.006

Hare SR, Mantua NJ. Empirical evidence for North Pacific regime shifts in 1977 and 1989. *Prog Oceanogr* 2000;**47**:103–45. https://doi.org/10.1016/S0079-6611(00)00033-1

Heide-Jørgensen MP, Chambault P, Jansen T *et al.* A regime shift in the southeast Greenland marine ecosystem. *Glob Change Biol* 2023;**29**:668–85. https://doi.org/10.1111/gcb.16494

Hellenbrecht LM, Utne KR, Karlsen r *et al.* Diet analysis of Atlantic salmon (*Salmo salar*) post-smolts after the ecological regime shift in the northeast Atlantic. *Fish Res* 2023;**262**:106672. https://doi.org/10.1016/j.fishres.2023.106672

Heymans JJ, Tomczak MT. Regime shifts in the northern Benguela ecosystem: challenges for management. *Ecol Model* 2016;**331**:151–9. https://doi.org/10.1016/j.ecolmodel.2015.10.027

Hillebrand H, Donohue I, Harpole WS *et al.* Thresholds for ecological responses to global change do not emerge from empirical data. *Nat Ecol Evol* 2020;**4**:1502–9. https://doi.org/10.1038/s41559-020-1256-9

Hsieh Ch, Glaser SM, Lucas AJ *et al.* Distinguishing random environmental fluctuations from ecological catastrophes for the North Pacific Ocean. *Nature* 2005;**435**:336–40. https://doi.org/10.1038/nature03553

ICES. 2021. Norwegian Sea ecoregion – Ecosystem overview. ICES Advice: Ecosystem Overviews. Report. https://doi.org/10.17895/ices.advice.8188

ICES. 2022. Third Workshop on Integrated Trend Analysis to Support Integrated Ecosystem Assessment (WKINTRA3). ICES Scientific Reports, 4:32. 41 pp. https://doi.org/10.17895/ices.pub.19398317.v1

ICES. 2023a. Report of the Herring Assessment Working Group for the Area South of 62°N (HAWG). ICES Scientific Reports, 5:23. 837 pp.

ICES. 2023b. Working Group on the Integrated Assessments of the Norwegian Sea (WGINOR, outputs from 2022 meeting). ICES Scientific Reports, 5:15. 57 pp. https://doi.org/10.17895/ices.pub.22110260.v1

ICES. 2023c. Working Group on Widely Distributed Stocks (WGWIDE). ICES Scientific Reports, 5:82. 980 pp.

Jones PD, Jonsson T, Wheeler D. Extension to the North Atlantic oscillation using early instrumental pressure observations from Gibraltar and south-west Iceland. *Int J Climatol* 1997;**17**:1433–50. https://doi.org/10.1002/(SICI)1097-0088(19971115)17:13⟨1433::AID-JOC203⟩3.0.CO;2-P

Juggins S. Package ‘rioja’. 2023. https://cran.r-project.org/web/packages/rioja/rioja.pdf (June 2024, date last accessed).

Killick R, Beaulieu C, Taylor S, *et al.* **EnvCpt:** Detection of Structural Changes in Climate and Environment Time Series, R package version 0.1. 1. 2022. https://cran.r-project.org/web/packages/EnvCpt/EnvCpt.pdf (January 2024, date last accessed).

Konar B, Estes JA. The stability of boundary regions between kelp beds and deforested areas. *Ecology* 2003;**84**:174–85. https://doi.org/10.1890/0012-9658(2003)084[0174:TSOBRB]2.0.CO;2

Lees K, Pitois S, Scott C *et al.* Characterizing regime shifts in the marine environment. *Fish Fish* 2006;**7**:104–27. https://doi.org/10.1111/j.1467-2979.2006.00215.x

Legendre P, Dallot S, Legendre L. Succession of species within a community: Chronological Clustering, with applications to marine and freshwater zooplankton. *Am Nat* 1985;**125**:257–88. https://doi.org/10.1086/284340

Legendre P, Legendre L. *Numerical Ecology*. Amsterdam: Elsevier, 2012.

Lindegren M, Diekmann R, Möllmann C. Regime shifts, resilience and recovery of a cod stock. *Mar Ecol Prog Ser* 2010;**402**:239–53. https://doi.org/10.3354/meps08454

MacArthur RH. On the relative abundance of bird species. *Proc Natl Acad Sci* 1957;**43**:293–95. https://doi.org/10.1073/pnas.43.3.293

Möllmann C, Diekmann R. Marine ecosystem regime shifts induced by climate and overfishing. *Adv Ecol Res* 2012;**47**:303–47. https://doi.org/10.1016/B978-0-12-398315-2.00004-1

Möllmann C, Folke C, Edwards M *et al.* Marine regime shifts around the globe: theory, drivers and impacts. *Philos Trans R Soc B Biol Sci* 2015;**370**:20130260. https://doi.org/10.1098/rstb.2013.0260

Morse RE, Friedland KD, Tommasi D *et al.* Distinct zooplankton regime shift patterns across ecoregions of the US northeast continental shelf large marine ecosystem. *J Mar Syst* 2017;**165**:77–91.

Ottersen G, Planque B, Belgrano A *et al.* Ecological effects of the North Atlantic Oscillation. *Oecologia* 2001;**128**:1–14. https://doi.org/10.1007/s004420100655

Overland JE, Percival DB, Mofjeld HO. Regime shifts and red noise in the North Pacific. *Deep Sea Res 1 Oceanogr Res Pap* 2006;**53**:582–8. https://doi.org/10.1016/j.dsr.2005.12.011

Perretti C, Fogarty M, Friedland K *et al.* Regime shifts in fish recruitment on the northeast US Continental Shelf. *Mar Ecol Prog Ser* 2017;**574**:1–11. https://doi.org/10.3354/meps12183

Planque B, Buffaz L. Quantile regression models for fish recruitment–environment relationships: four case studies. *Mar Ecol Prog Ser* 2008;**357**:213–23. https://doi.org/10.3354/meps07274

Planque B, Favreau A, Husson B *et al.* Quantification of trophic interactions in the Norwegian Sea pelagic food-web over multiple decades. *ICES J Mar Sci* 2022;**79**:1815–30. https://doi.org/10.1093/icesjms/fsac111

Pyper BJ, Peterman RM. Comparison of methods to account for autocorrelation in correlation analyses of fish data. *Can J Fish Aquatic Sci* 1998;**55**:2127–40. https://doi.org/10.1139/f98-104

Reid PC, de Fatima Borges M, Svendsen E. A regime shift in the North Sea circa 1988 linked to changes in the North Sea horse mackerel fishery. *Fish Res* 2001;**50**:163–71. https://doi.org/10.1016/S0165-7836(00)00249-6

Rodionov S. A brief overview of the regime shift detection methods. In: *Large-Scale Disturbances (Regime Shifts) and Recovery in Aquatic Ecosystems: Challenges for Management Toward Sustainability*. Varna: UNESCO-ROSTE/BAS Workshop on Regime Shifts, 2005, 17–24.

Rodionov SN. A sequential algorithm for testing climate regime shifts. *Geophys Res Lett* 2004;**31**: https://doi.org/10.1029/2004GL019448

Rodionov SN. Use of prewhitening in climate regime shift detection. *Geophys Res Lett* 2006;**33**: https://doi.org/10.1029/2006GL025904

Room AH, Franco-Gaviria F, Urrego DH. rshift STARS manual—regime shift analysis for paleoecological data v2.2.0. 2023.

Rudnick DL, Davis RE. Red noise and regime shifts. *Deep Sea Res 1 Oceanogr Res Pap* 2003;**50**:691–9. https://doi.org/10.1016/S0967-0637(03)00053-0

Rybakov M, Firsov Y, Nosov M *et al.* International Ecosystem Survey in Nordic Sea (IESNS) in April–June 2016. Vol. **21**. Copenhagen: ICES Working Group on International Pelagic Surveys, 2016.

Scheffer M, Bascompte J, Brock WA *et al.* Early-warning signals for critical transitions. *Nature* 2009;**461**:53–9. https://doi.org/10.1038/nature08227

Scheffer M, Carpenter S, Foley JA *et al.* Catastrophic shifts in ecosystems. *Nature* 2001;**413**:591–6. https://doi.org/10.1038/35098000

Scheffer M, Carpenter SR. Catastrophic regime shifts in ecosystems: linking theory to observation. *Trends Ecol Evol* 2003;**18**:648–56. https://doi.org/10.1016/j.tree.2003.09.002

Schreiber T, Schmitz A. Surrogate time series. *Physica D* 2000;**142**:346–82. https://doi.org/10.1016/S0167-2789(00)00043-9

Seddon AWR, Froyd CA, Witkowski A *et al.* A quantitative framework for analysis of regime shifts in a Galápagos coastal lagoon. *Ecology* 2014;**95**:3046–55. https://doi.org/10.1890/13-1974.1

Sguotti C, Blöcker AM, Färber L *et al.* Irreversibility of regime shifts in the North Sea. *Front Mar Sci* 2022;**9**:945204. https://doi.org/10.3389/fmars.2022.945204

Skjoldal HR, Sætre R, Færnö A *et al.* The Norwegian sea ecosystem. 2003.

Steneck RS, Graham MH, Bourque BJ *et al.* Kelp forest ecosystems: biodiversity, stability, resilience and future. *Environ Conserv* 2002;**29**:436–59. https://doi.org/10.1017/S0376892902000322

Stirnimann L, Conversi A, Marini S. Detection of regime shifts in the environment: testing 'STARS' using synthetic and observed time series. *ICES J Mar Sci* 2019;**76**:2286–96. https://doi.org/10.1093/icesjms/fsz148

Theiler J, Eubank S, Longtin A *et al.* Testing for nonlinearity in time series: the method of surrogate data. *Physica D* 1992;**58**:77–94. https://doi.org/10.1016/0167-2789(92)90102-S

Tomczak MT, Müller-Karulis B, Blenckner T *et al.* Reference state, structure, regime shifts, and regulatory drivers in a coastal sea over the last century: the Central Baltic Sea case. *Limnol Oceanogr* 2022;**67**:S266–84. https://doi.org/10.1002/lno.11975

Trapletti A, Hornik K, LeBaron B. Package 'tseries'. Technical report, 2023.

Vollset KW, Urdal K, Utne K *et al.* Ecological regime shift in the northeast Atlantic Ocean revealed from the unprecedented reduction in marine growth of Atlantic salmon. *Sci Adv* 2022;**8**:eabk2542. https://doi.org/10.1126/sciadv.abk2542

Weijerman M, Lindeboom H, Zuur A. Regime shifts in marine ecosystems of the North Sea and Wadden Sea. *Mar Ecol Prog Ser* 2005;**298**:21–39. https://doi.org/10.3354/meps298021

Zeileis A, Leisch F, Hornik K *et al.* **Strucchange**: an *R* package for testing for structural change in linear regression models. *J Stat Softw* 2002;**7**:1–38.

Handling Editor: Mary Hunsicker