

Supplementary Information:

Poor performance
of regime shift detection methods
in marine ecosystems

Hannah Haines
Alfred Wegener Institute,
Bremerhaven, Germany
hannah.haines@awi.de

Benjamin Planque
Institute of Marine Research,
Tromsø, Norway
benjamin.planque@hi.no

Lucie Buttay
Institute of Marine Research, Tromsø, Norway
UiT the Arctic University of Norway, Tromsø, Norway
lucie.buttay@uit.no

November 29, 2023

1 Summary of all tables included in the Supplementary Information

Table 1: Summary of results from STARS application on all 32 available Norwegian Sea original time series, a 0 appears in the chgpt years columns when no changepoints were found.

Table 2: Summary of results from STARS surrogate testing on all 32 available Norwegian Sea variables, each columns gives the number of changepoints found.

Table 3: Summary of results from Strucchange method application on all 32 available Norwegian Sea original time series.

Table 4: Summary of results from Strucchange surrogate testing on all 32 available Norwegian Sea variables.

Table 5: Summary of results from EnvCpt surrogate testing on all 32 available Norwegian Sea variables, including additional information into the origin of the conflicts.

Table 6: Summary of results from EnvCpt application on all 32 available Norwegian Sea original time series, an x indicates that this model was selected according to its BIC score, if more than one X occurs for each variable/row, a conflict arose.

Table 7: Summary of results from EnvCpt surrogate testing on all 32 available Norwegian Sea variables.

Table 8: Summary of results from Chronological clustering surrogate testing on 6 different groupings, giving here the percentage of each 1000 multivariate surrogate set finding between 1 and 10 significantly different clusters with the chclust method.

2 Data and methods: additional information

2.1 Omnibus transformation

As a majority of the time series did not exhibit a normal distribution, which is a necessary criterion in generating surrogate time series, multiple transformations were tested. These included logarithmic transformations and raising the data to the power of 0.25. However, these transformations are not suitable for data containing both negative and positive numbers. To address this, a lesser-known robust normalisation technique was employed, called the omnibus normalisation technique which was initially established by Legendre et al, 2012. Given the absence of an R package or an established R function, a function was formulated following these steps.

- 1- Determine the ranks of each data point in the time-series with the `rank()` function, resolving ex aequo (tied) values with the ‘first’ method to avoid subsequent issues.
- 2- Generate a new series of random data using the `rnorm()` function, of the same

length as the original series.

- 3- Determine the ranks of the random series (with the same `rank()` function and the ‘first’ method),
- 4- By associating the ranks between both the original and random normal series, reorder the random normal data using the original sequence of ranks to obtain the new normal series exhibiting a similar overall pattern as the original time series.

However, this technique has an element of randomness and can impact the results across iterations. Consequently, a modified version of this transformation was applied instead, dubbed OmnibusQ, which replaces the `rnorm()` function with the `qnorm()` function. Nonetheless, this approach has a drawback: for the time series of the same length, the same values will always be generated since it is based on quantiles and not the normal distribution curve itself. The `qnorm` series are generated with a sequence of variables going from $\frac{1}{n+1}$ to $1 - \frac{1}{n+1}$, by $\frac{1}{n+1}$, where n is the length of the original series. To facilitate back-transformation, a corresponding function was also developed, named `deOmnibusQ`, which uses the original ranks and sequence of the original time-series to reassociate the transformed data with the original data. It is important to remark that this robust method has a second limitation: repeating values in the time series are transformed into similar but distinct values. Unfortunately, addressing this flaw would compromise the normality of the transformed data and thus the effectiveness of this technique.

2.2 Surrogate time series

Surrogate time series are simulated time-series that conserve the same trend and autoregressive processes as the original data by relying on the principal of phase randomisation. To generate these surrogate time series, the original data is de-trended (by extracting the residuals after applying the `lm()` R function) and normalised (using the omnibus transformation). Surrogates are then generated using the `surrogate()` function of the `tseries` R package, with both the arguments `fft` and `amplitude` set to `true`, to ensure phase-randomisation and amplitude-adjustment surrogates. The trend is then added back on to the surrogates and the normalisation is reversed. Due to this reversed normalisation, all surrogates will simply be a reordering of the original data. However, because for example for a time series of 30 data points, there are !30 possible reorderings of the data, it is unlikely that any of the 1000 surrogates are identical.

By design these surrogates do not contain changepoints and all fluctuations can be solely attributed to autoregressive processes and the trend. These surrogates are generated in high number for each individual time series. Consequently, they correspond to a means to test the null hypothesis: "There are no changepoints or regime shifts". If a changepoint or regime shift detection method performs perfectly, all surrogates should

yield negative results. The number of surrogates for which a positive result of the test is found, can be transformed to get the false positive rate and consequently the test's specificity. 1000 surrogates were generated for each time series and conserved so that the same surrogates were used between methods, to avoid any possible iterative errors.

Each univariate changepoint detection method is applied iteratively to the surrogates of each time series. The result of each test is saved and the results are summarised across the 1000 surrogates and used to deduce the proportion of false positives and true negatives. Due to the distinction posed between a changepoint and a regime shift, a multivariate method was also evaluated by combining the surrogates from all constituting time series grouping.