THIRD WORKSHOP ON INTEGRATED TREND ANALYSIS TO SUPPORT INTEGRATED ECOSYSTEM ASSESSMENT (WKINTRA-3)

VOLUME 4 | ISSUE 32

ICES SCIENTIFIC REPORTS

RAPPORTS SCIENTIFIQUES DU CIEM

ICES INTERNATIONAL COUNCIL FOR THE EXPLORATION OF THE SEA
CIEM CONSEIL INTERNATIONAL POUR L'EXPLORATION DE LA MER

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

H.C. Andersens Boulevard 44-46
DK-1553 Copenhagen V
Denmark
Telephone (+45) 33 38 67 00
Telefax (+45) 33 93 42 15
www.ices.dk
info@ices.dk

ISSN number: 2618-1371

ICES Scientific Reports

Volume 4 | Issue 32

THIRD WORKSHOP ON INTEGRATED TREND ANALYSIS TO SUPPORT INTEGRATED ECOSYSTEM ASSESSMENT (WKINTRA-3)

Recommended format for purpose of citation:

ICES. 2022. Third workshop on Integrated Trend Analysis to support Integrated Ecosystem Assessment (WKINTRA-3).
ICES Scientific Reports. 4:32. 41 pp. https://doi.org/10.17895/ices.pub.19398317

Editors

Saskia Otto • Benjamin Planque

Authors

Per Arneberg • Andrea Belgrano • Lucie Buttay • Mårten Erlandsson • Bérengère Husson
Jed Kempf • Sigrid Lehuta • Marcos Llope • Hiroko Solvang • Nigel Yoccoz

Contents

i Executive summary

The overall aim of the workshops on Integrated Trend Analysis to support Integrated Ecosystem Assessment is to develop good practices in the application of integrated trend analyses (ITA) and interpretation of their results for integrated ecosystem assessment (IEA). This series of workshops follow a simulation-based evaluation approach.

The third workshop (WKINTRA3) was dedicated to the review of simulated multivariate ecological "control" datasets, which were further used to evaluate the following selection of ITA methods: heatmaps, principal component analysis (PCA), integrated resilience analysis (IRA), multivariate autoregressive trend models (MAR-T), trend estimation and classification (TREC), minimum/maximum autocorrelation factor analysis (MAFA), redundancy discriminant analysis (RDA) and dynamics factor analysis (DFA).

Based on the ITA evaluation results, several recommendations are made for further development and use of ITA to support the work of ICES IEA groups. These recommendations include 1) clear specification of the objective when applying ITA methods, 2) increased transparency and traceability of the methods used, 3) explicit consideration of input data uncertainties, 4) methods for detecting extreme events (such as heatwaves), 5) harmonisation in the reporting of ITA outputs, 6) generalisation of the evaluation of ITA method performance and 7) peer-reviewing of ITA methods across IEA groups.

ii Expert group information

Expert group name	Third workshop on Integrated Trend Analysis to support Integrated Ecosystem Assessment (WKINTRA-3)
Expert group cycle	Annual
Year cycle started	2021
Reporting year in cycle	1/1
Chair(s)	Saskia Otto, Germany
	Benjamin Planque, Norway
Meeting venue(s) and dates	21–24 September 2021, online, 12 participants

1 Terms of reference, agenda and participation

The general objective of the workshop series (WKINTRA) is to develop good practices in the application of integrated trend analyses (ITA) and interpretation of their results for integrated ecosystem assessment.

The Terms of Reference of the workshop WKINTRA-3 were:

a) Review the simulated multivariate ecological datasets prepared during and following WKINTRA2 (Science plan codes 1.3 and 1.9).

b) Evaluate a selection of Integrated Trend Analysis (ITA) methods (Science plan codes 1.3 and 1.9).
For this, the following steps were to be taken,

- select a set of ITA methods,
- provide the R code to run the analyses,
- define method-specific qualitative or quantitative criteria that allow for an objective comparison across simulated datasets,
- apply the ITA methods on relevant simulated datasets, and assess outcomes on a case study- and approach-specific basis.

c) Develop guidelines for IEA groups to evaluate ITA methods, including a comprehensive documentation of data generation and method application using the R environment (Science plan code 6.5).

In 2018, a first workshop (WKINTRA-1) was held in Hamburg, Germany (ICES, 2018). It was then recommended that two further workshops should be conducted. A second workshop to generate simulated datasets for few contrasted ecosystems (WKINTRA-2) and a third workshop to perform the evaluation of selected ITA methods on the simulated datasets (WKINTRA-3).

WKINTRA-3 took place as an online meeting on September 21st-24th, 2021, and was attended by 12 participants from five countries (see annex 1).

Benjamin Planque and Saskia Otto opened the meeting. The agenda was adopted (see annex 2). Participants were informed about ICES code of conduct. No conflicts of interest were identified by the participants or the chairs. The introduction to the meeting was followed by a brief introductory round table of the participants.

The following ITA methods were agreed to be evaluated:

- Heatmaps or so-called Traffic Light Plots (TLP)
- Principal Component Analysis (PCA)
- Integrated Resilience Analysis (IRA)
- Multivariate Autoregressive Trend Analysis (MAR-T)
- Trend Estimation and Classification (TREC)
- Minimum/Maximum Autocorrelation Factors (MAFs)
- Redundancy Discriminant Analysis (RDA)
- Dynamic Factor Analysis (DFA)

2 Review the simulated datasets

2.1 Surrogate time-series using phase randomisation

The method of surrogate time-series based on phased randomisation is designed to simulate time-series with specified mean, variance, and temporal autocorrelation. The method is described in Schreiber and Schmitz (2000) and was applied in Planque and Arneberg (2018), where the authors assessed the robustness of Principal Component Analysis (PCA) when applied to multivariate time-series. The general principle is that a time-series in the time-domain can be fully described by its equivalent in the frequency domain, by specifying the values for the phase and amplitude for each frequency. If the amplitudes remain unchanged but the phases are shuffled, the resulting time-series share the same power-spectrum (and therefore the same temporal autocorrelation) as the original time-series, but the values in the time-domain are random. The method is used here to simulate multivariate time-series datasets in which each individual time-series simulation preserves specific properties of the original data. Because the time-series are generated independently from each other, it is assumed that relationships between time-series can only emerge by chance. Such multivariate dataset produced with phase randomisation is analogue to an experimental 'control' or 'blank' or 'null model'.

The phase randomisation requires that the original time-series be stationary, i.e. without a monotonic trend and that the data be normally distributed. For this purpose, on each time-series, the method was implemented as follows:

1. Transform non-normal data (usually with double square root transformation)
2. Estimate the linear time-trend, and compute the residuals
3. Apply the phase randomisation on the residuals
4. Add the linear time-trend to the randomised residuals
5. Back-transform (with e.g. double square transformation).

The issue of the linear temporal trend is not easy to deal with because the trends in the datasets contain information and, at the same time, constitute a statistical nuisance. When a linear trend dominates (i.e. the % variance explained by the trend is high), the information content of the time-series is close to 2 degrees of freedom (the slope and intercept). The statistical value of the observed relationships between 2 series dominated by trends is similar to that of a correlation based on 2 observations. Although it is tempting to associate time-series that share similar or opposite trends, the statistical support for the association between the time-series is very weak.

In practice, surrogates can be generated in the R language using the function `surrogate` of the library `tseries` (Trapletti and Hornik, 2020).

During the previous workshop and this workshop, surrogate time-series were generated for the following datasets: Barents Sea, Bay of Biscay, Central Baltic Sea, Celtic Sea, Norwegian Sea, North Sea Skagerrak, NoBa simulations (see section 2.4) and Baltic Sea simulations (see section 2.5).

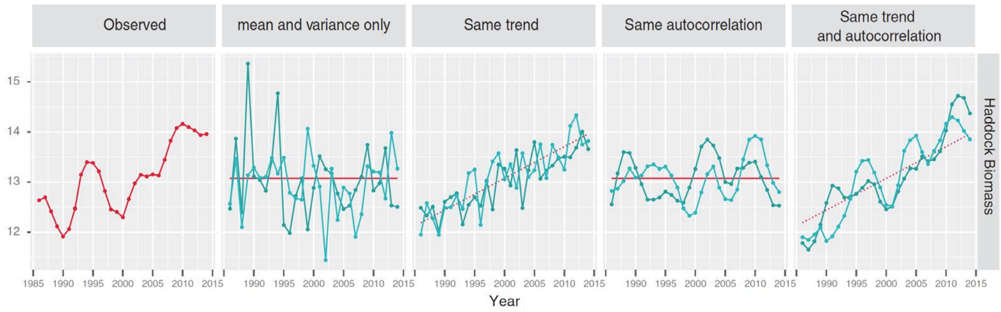

Figure 1. An illustration of surrogate time-series. The left panel shows the original dataseries for Haddock biomass in the Barents Sea from 1985 to 2015. The rightmost panel shows two surrogates with the same trend and residual autocorrelation as the original observations. The 3 panels in the middle show surrogates with only part of the original data structure preserved: only mean and variance (middle left), trend and residual variance but not autocorrelation (middle), or residual variance and autocorrelation but not trend (middle right).

2.2 Multivariate autoregressive simulations

The observed time-series data display stationary and non-stationary mean components. In statistical time-series modelling, the stationary and non-stationary mean parts are considered separately in the model because the statistical property of non-stationary mean process (mean is not constant over time) is different from the property of stationary process (mean, variance are constant over time and covariance is consistent for a fixed lag). Once the trend component is estimated by a trend model (see next paragraph), MAR (if the data are multivariate) and AR (if the data are univariate) models are applied to the residual parts that are obtained by extracting trend from the observations. MAR model expresses the state of all variables at present and past time-steps, and the appropriate effect from the past are evaluated by the statistical criterion. In MAR, the relationships between variables are explicit in the autoregressive (AR) coefficient matrix. AR model is the univariate version. The physical meaning of the estimated AR coefficients is obtained by considering the frequency domain. The frequency response function is given by the estimate AR coefficients and the (cross) power spectrum is calculated by the variance (- covariance matrix) of prediction error of (M)AR model and the frequency response function.

MAR model is used to simulate multivariate time-series data, or to estimate the likely value of model parameters given observational data. It is expected that results of ITA performed on MAR outputs should recover some information about the relationships between variables, that were explicitly included via the model parameters. Interestingly, it is possible to estimate MAR model parameters from observational data, simulate new datasets with the fitted MAR model and recover the parameters by fitting new MAR to the simulated data. If the data presents non-stationary trend, the trend model can be considered in addition to MAR model. Furthermore, effects from abiotic factors can be considered in the model, expressed as exogenous variables (X). The formula is given by MARX or MARX+TREND. In any case, this multistage fitting-simulation-fitting process can be insights into the performance of MAR(X) + Trend as a tool for ITA.

2.3 ISIS-Fish simulations

ISIS-Fish is a deterministic simulation model designed to explore the dynamics of mixed fisheries (Mahevas and Pelletier, 2004; Pelletier *et al.*, 2009). It is spatially explicit with a monthly time-step. Catches result from the interaction between the spatial distribution of population abundance and the spatial distribution of fishing effort standardised per gear, métier, and fleet, both dynamically updated each month. The ISIS-Fish application to the mixed demersal fishery of the

Eastern English Channel simulates the response of fleets and stocks to the implementation of the landing obligation regulation. It comprises 10 commercial species and 17 fleets (Lehuta *et al.*, 2015; Lehuta and Vermard, submitted). It evidences the effect of choke species on the fishery, that is an early closure of the fishery in the year, due to the limiting TAC for plaice and sole and its impact on effort, catches and biomass of all the target species. For the purpose of the workshop, simulations were ran for 30 years, assuming stock recruitment relationships for five of the stocks. To illustrate the impact of random noise on the perceived dynamics, a deterministic run and 10 additional runs including interannual noise around the stock-recruitment relationship were carried out. Outputs of interest are typical from fishery monitoring and consist of 48 annual time-series of stock biomass, catches, mean weight in catch and effort per fleet. MAFA were applied to the time-series.

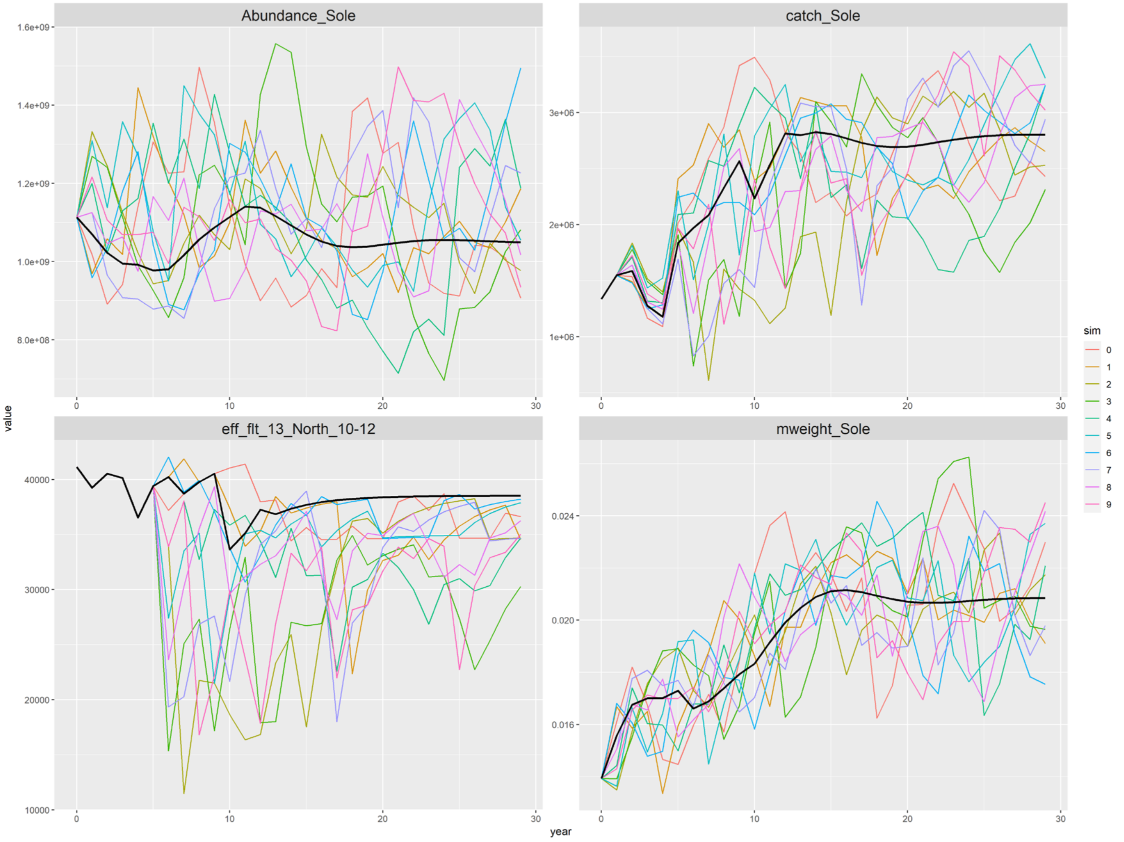

Figure 2. Examples of ISIS-Fish outputs for sole abundance, catch and mean weight of catch and the effort of one of the netter fleets. Black line is the deterministic run, coloured lines are the runs with random noise on the stock recruitment relationship.

2.4 NoBa simulations

Atlantis is an end-to-end, mostly deterministic model consisting of several modules that interact: an oceanographic module (physico-chemical environment, hydrology), a biological module (foodweb model), and a harvesting module (with fleets, gears and management schemes). In a more complex form, the model can also include an assessment module (optional, for e.g. management strategy evaluation), and an economic module that can include incentives for fisheries or inform on economical outputs (Audzijonyte *et al.*, 2019). Processes are spatially resolved in a set of boxes and layers where they are considered homogeneous. In the case of the Norwegian and Barents Sea Atlantis model (NoBa, Hansen *et al.*, 2016) the model comprises 60 boxes with up to 7 water column layers and a sediment layer, and 53 biological compartments, with biomass pools for invertebrate species and up to 10 age classes per vertebrate groups. For the workshop, we used readily available and published data from Hansen *et al.* (2019). The simulations aimed at studying the impact of different scenarios of fisheries effort on the ecosystem in a warming Barents Sea (defined by a downscaled forcing under the RCP4.5 climatic scenario). We selected only one scenario with a fishing pressure of 1.0x FMSY ran on 110 years. Selected outputs are the main time-series of integrated ecosystem assessment for the Barents Sea, including environmental conditions (temperature, oxygen, chlorophyll a), population biomass (large zooplankton, capelin, herring, polar cod, cod, Greenland halibut, haddock, saithe, long rough dab) and fisheries catches (capelin, herring, cod, haddock and Greenland halibut). A thousand surrogate time-series were created from those outputs, from which 9 were selected randomly. It was noticed that in the case of very smoothed time-series (e.g. saithe, SAI), the surrogates had a much stronger variability than the original time-series (Figure 3), a problem already identified by Schreiber and Smith (2000). Heatmaps and PCA were applied to the time-series.

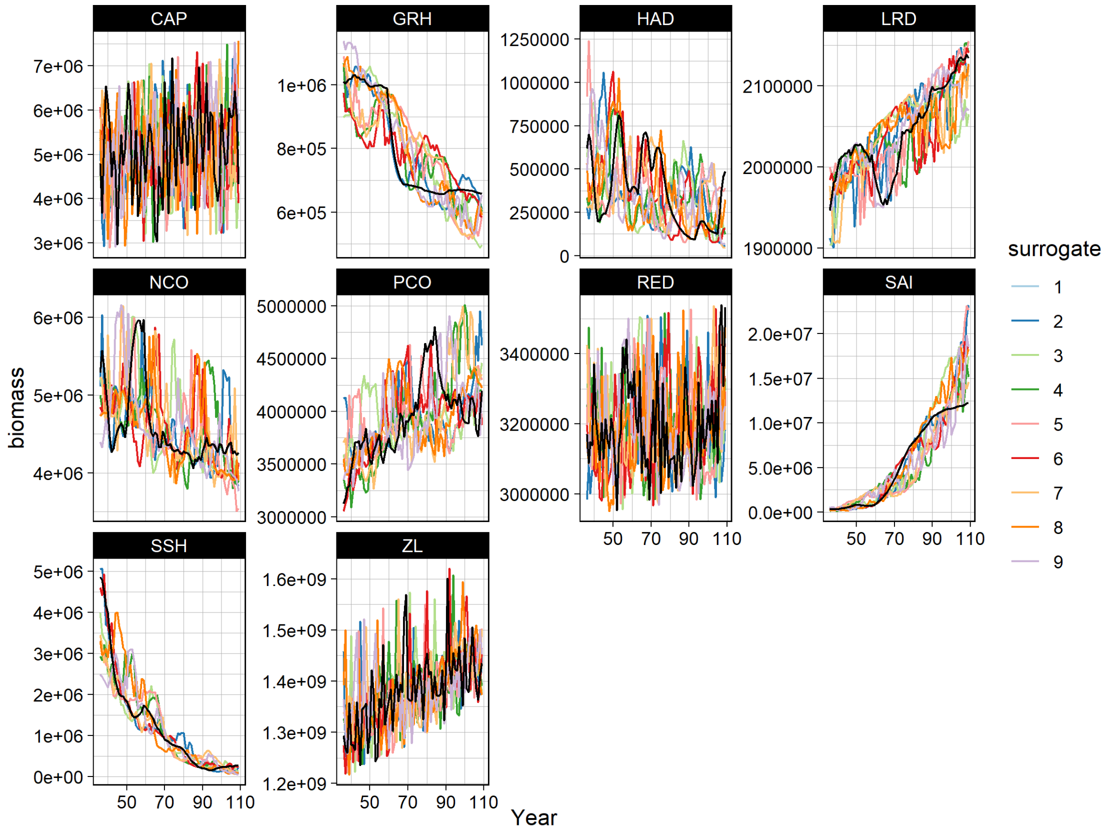

Figure 3. Population biomass simulated time-series from NoBa Atlantis (in black) and corresponding surrogates (colours)

2.5 Baltic Sea

A generalized dynamic foodweb model for the central Baltic Sea was presented and discussed (Blenckner *et al.*, 2015). This model consists of coupled empirical individual models (Generalized Additive Models, GAMs) and allows for regime-dependent dynamics (tGAM, a threshold formulation of a GAM) depending on whether the system is below or above a given value of a threshold variable (see Figure 3 in Blenckner *et al.*, 2015). The modelling approach consists of two steps: first an individual statistical model is fit to each trophic level, then the separate models are coupled together into a joint foodweb model that is able to reproduce the observed population dynamics based on external drivers and the trophic interactions emerging from the individual models. This generalized dynamic foodweb model can be subsequently used to simulate the system under a range of conditions (Figure 2 in Blenckner *et al.,* 2015). We decided to use the Baltic Sea model due to its simple configuration (few foodweb components) and because it accounts for the regime shift that took place in this marine ecosystem in the late 1980s. Regime shifts are an ecological feature that ITA methods are supposed to be able to detect and, therefore, outputs from the Baltic Sea model can be a good candidate dataset to evaluate different ITA methods.

Simulated output time-series consists of 100 realizations of the ecosystem for a series of 81 fishing mortality levels (Flvl), covering the range that goes from 0 to 1.4, every 0.05. Environmental conditions were sampled from values observed during the first or the second regime (before/after 1989). This setup gives a total of 16,200 realizations 8,100 run under pre-regime environmental conditions and 8,100 using those observed after the late-1980s regime shift (Figure 4 in Blenckner *et al.,* 2015).

For the purpose of WKINTRA3 one random realization (81 steps) of the Baltic Sea foodweb for a complete range of Fs (0, 0.05, 0.10, …, 1.30, 1.35, 1.40) was selected. This realization consisted of cod (CODs), sprat (SPRs), herring (HERs), *Pseudocalanus* sp. (PSEs) and cladocerans (CLAs) biomass/abundance as well as the corresponding temperature (TMPsum), salinity (SLNdeep), herring fishing mortality (Fher) and the increasing levels of cod fishing mortality (Flvl). Subsequently, surrogates were created out of the selected simulation.

Heatmaps and PCA (biplot and trajectories) analyses were run on both the simulations and surrogates. The methods were able to pinpoint the cod fishing mortality level at which cod biomass collapsed. The exact level showed some variation when using surrogates.

3 Evaluation of selected Integrated Trend Analysis (ITA) methods

3.1 Heatmaps

Method description

The heatmap (also called Traffic Light Plot, TLP) is a graphical method designed to represent multiple time-series simultaneously. The output of the heatmap method is a colourised matrix in which each row is a variable and each column is a year. The colours reflect the values in the original datasets and the dataseries can be re-ordered (from top to bottom) so that series that share similar properties are plotted near each other.

For this specific application we used the following protocol to generate the heatmaps:

- For each individual time-series:
    - Classify individual observations in 5 categories from low to high. The categories are based on equal intervals from the lowest to the highest recorded value.
    - Colourise the observations from green (low) to red (high).
- Plot all the time-series in an ordered fashion (here, based on the mean in the first 5y).

Data used

The heatmap ITA was applied to observational data and surrogates (section 2.1) from the following six regional IEAs: Barents Sea, Bay of Biscay, Central Baltic Sea, Celtic Sea, Norwegian Sea and North Sea / Skagerrak. A thousand surrogate time-series were available for each region, but only 8 surrogates were analysed. This is because the interpretation of the heatmaps is mostly visual, and it was deemed practical and appropriate to look at a limited number of maps only.

Evaluation approach

The evaluation of the heatmap method was conducted in 3 phases. First, the participants were asked to answer a series of questions regarding their expectations from heatmaps. Second, the participants were provided with 9 heatmaps for each region. For each regional set, one of the heatmaps was generated from the observational data and the 8 others from surrogates. Third, the participants were asked questions similar to those in the first step. In addition, they were asked if they could identify/discriminate the heatmap that was based on the original data, on the regional system of their choice (i.e., the one they knew best). An example of the 9 heatmaps for the Barents Sea is provided as an illustration (Figure 4).

This was followed by a plenary discussion on the same day (day #1 of the meeting) and another discussion in a smaller group on day #3.

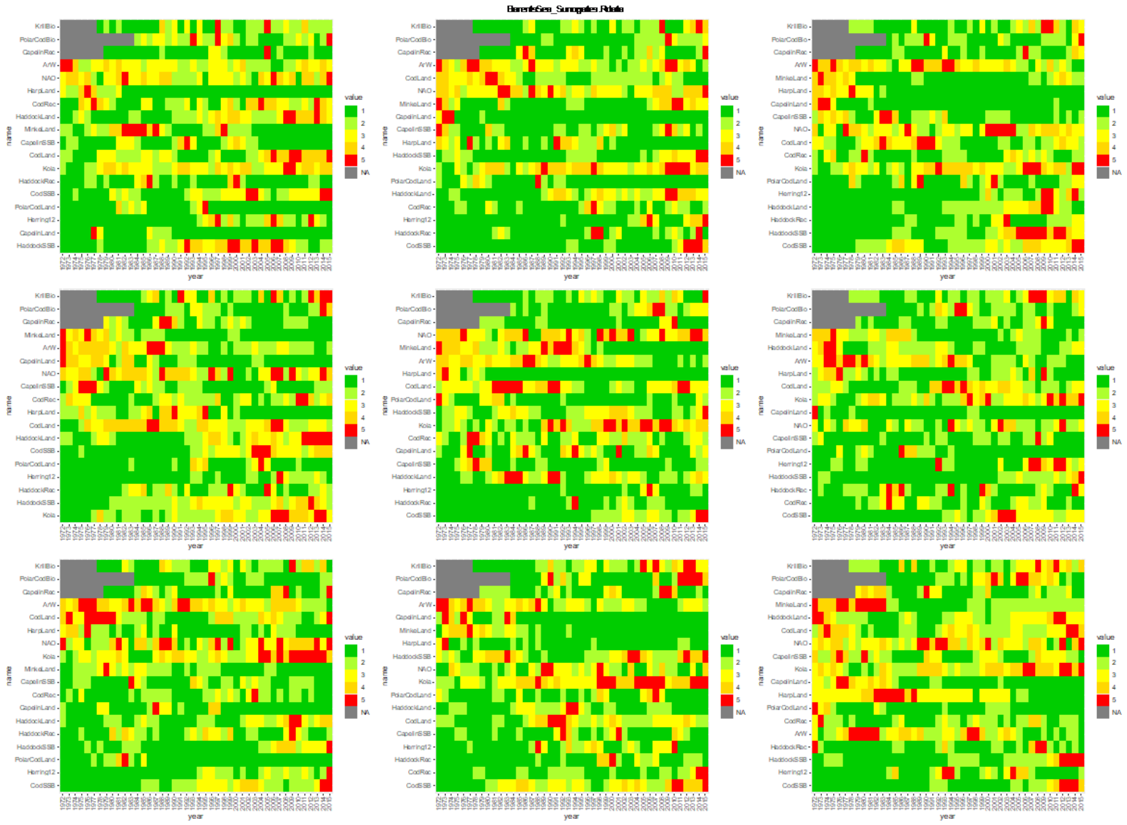

Figure 4. Nine heatmaps for the Barents Sea ecosystem. One map is based on the observational data and the 8 others are based on surrogate time-series.

Results

The surveys before and after looking at the original/simulated heatmaps yielded similar results (Figure 5) regarding objectives that seem legitimate (e.g. get a general visual impression of the long-term changes of many variables simultaneously, split between variables that have increased, decreased or remain stable or identify time-periods that are similar to each other). Similarly, objectives that seemed less legitimate in the first survey remained so in the second survey (single out series that are different from others and single out years that are different from other). Overall, the heatmap is perceived as a method that can be applied with a wide range of objectives in mind.

One noticeable change between the first and the second survey concerns the association between biotic/abiotic variables that have similar temporal patterns. Respondent were more confident in heatmaps achieving this goal after visualising the plots than before.

Five out of 11 respondent were confident in identifying the heatmap derived from original observations against the 8 others derived from surrogates. This was usually achieved by looking at well-known individual time-series (e.g. herring and seabirds in the Norwegian Sea, cod and NAO in the Barents Sea or cod and *Acartia* sp. in the central Baltic Sea) or by inspecting 'blocks' of time-series that were known to co-vary.

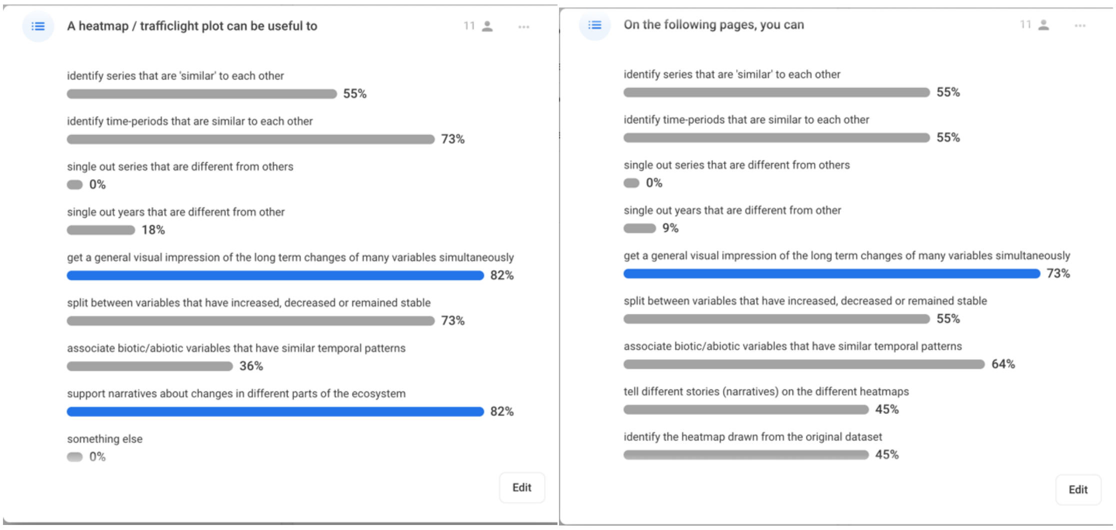

Figure 5. Results from the survey on heatmaps before (left) and after (right) visualising the heatmaps for the 8 regional IEAs.

Important issues

Trends: heatmaps appear efficient at visualising the long-term trends of many series. It is more immediate to read trends from heat maps than to read year-to-year fluctuations. In heatmaps the trends are considered to provide important (maybe the most important) information.

Association between time-series maybe more difficult to identify than trends. This is particularly true when time-series are located far apart on the map, when the association is at short time-scale and when the association is negative.

Recurrent emerging patterns: when many of the dataseries are dominated by trends, the heatmap will display a diagonal pattern, with e.g., declining series at the top and increasing time-series at the bottom.

Colour categorisation: The choice of categories for colouring the dataseries is not neutral. Categories based on equal intervals or quantiles will yield graphically different outputs. At present most categorisation schemes are based on the full range of variability in individual variables. This means that two series, for example one with species biomass varying by 100% and one with species biomass varying by 1%, will both be represented using the full colour range. This could be misleading and in some instances the categorisation scheme needs to be revised.

Choice of colours: The colour scale (green to red) can be mis-interpreted. Some can read it in a normative way, *green = good* and *red = bad* while other associate red with low and green with high. In addition, the scale may not be appropriate for some categories of colour-blind readers.

Ordering of variables: How the different variables are ordered when drawing the heatmap can also greatly influence the reading and interpretation. This is because series that are nearby are easier to visually compare than those that are far apart. Current ordering methods include the use of a prior PCA analysis to order time-series based on their score on the first axis, the use of the average values in the early years of the time-series (series are for example ordered by decreasing mean category values in the first 5y), or the use of an external variable.

Heatmaps are not suitable to represent **uncertainties** in the underlying dataseries, since only the point estimates can be colourised.

The group could not identify obvious **criteria for defining or estimating heatmap performance**.

Although the survey results indicated that heatmaps could be (or had been) used with a range of objectives, **the method appear most useful to provide a graphical summary of the data that support a pre-existing description of changes in an ecosystem**. There exist other methods that are better suited to investigate relationships between time-series, regime shifts, etc. In addition, if prepared in standardised manner, heatmaps could be a way to harmonise the presentation of some of the IEA group results, thereby favouring comparisons between regions and systems.

Recommendations

Use heatmaps primarily as a graphical illustration of the ecosystem changes, not as a tool for investigating these changes.

Pay attention to methodological choices: data categorisation scheme, colour scale, method for ordering variables. Report these choices explicitly.

Use non-normative colours and terminology as much as possible (e.g., the term "traffic light" is normative but the terms "heatmap" or "colormap" are not).

Calm down if you see a diagonal pattern, this is an expected feature when many time-series are dominated by long-term trends.

Harmonise practices between IEA groups as much as possible to promote comparative analyses.

3.2 PCA

Methods description

The evaluation of Principal Component Analysis (PCA) was based on the results published by Planque and Arneberg (2018). PCA is a dimension reduction technique based on the transformation of the original data into a new multidimensional space in which each dimension, i.e., component, is orthogonal to the others. The components are presented by decreasing amount of explained variance. The technique is often used to reduce/summarise multivariate datasets into a smaller number of dimensions and to identify relationships between the original variables. PCA can be performed on the variance-covariance matrix or on the correlation matrix between the original variables. In all IEAs, PCAs are conducted on the correlation matrices because the different variables are often reported in different units. Some IEA groups perform PCAs separately on abiotic and biotic variables while others treat them simultaneously.

Data used

The data used is presented in Planque and Arneberg (2018). It consists of the dataseries provided by the IEA groups for the Norwegian Sea, Barents Sea, North Sea Skagerrak and Central Baltic Sea plus a control dataset constructed by assembling 15 unrelated time-series from large-scale databases. For each dataset, 1000 surrogates were produced using the method outlined in section 2.1.

Evaluation approach

The evaluation was primarily based on two outputs of the PCA analyses. First, the representation of the ecosystem trajectory in the plane formed by the first 2 components. Second, the percentage of variance explained by the first 2 components. The evaluation was done by comparing how these two outputs differ between the analyses performed on the original datasets and those performed on surrogate datasets.

Results

The results show that outputs on the control dataset are not differentiable from outputs based on surrogate time-series (which is the expectation). For the 4 IEA datasets, the "horseshoe" pattern of the ecosystem trajectory is often observed in the surrogate time-series. This is a well-known artefact of the method. When linear trends are dominating in many time-series, the first component of the PCA captures the trend components, and the second component, which is orthogonal to the first one by construction, resemble a second order polynomial. The result is a "horseshoe" pattern in the PC1-2 plane (Figure 6). The percentage variance explained by PC1-2 in the original IEA datasets exceeds, on average – the percentage variance explained in the surrogate datasets. However, the difference is often marginal (Figure 7). Most of the variance can be explained by random (but structured in time) time-series or is explained by other PCs.

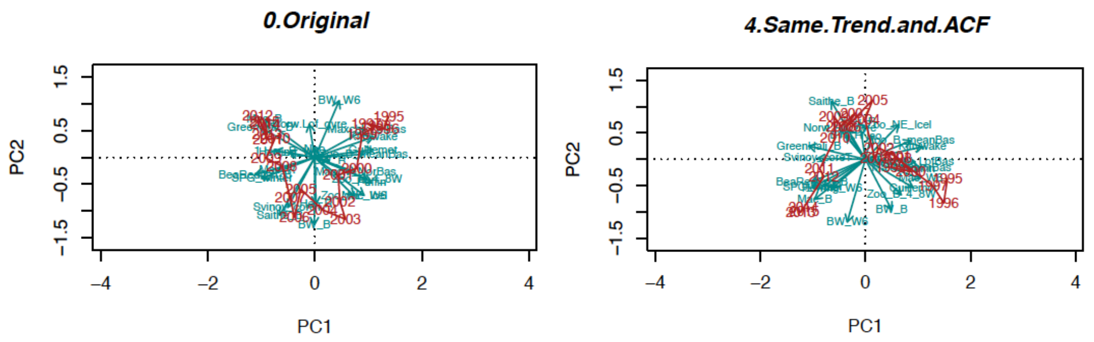

Figure 6. Biplots of the PCA performed for the Norwegian Sea. The arrows show the contribution of each variable in the first 2 components (length of the arrow) as well as the correlation between variables and components (angle between arrows or between arrows and components). The continuous lines indicate the temporal “trajectory” of the system projected on the first two axes of the PCA. The PCA results are shown for the original dataseries (left) and one realization of simulated data with trend and autocorrelation preserved (right). The horseshoe-like trajectory is visible on both plots. (reproduced from Planque and Arneberg, 2018)

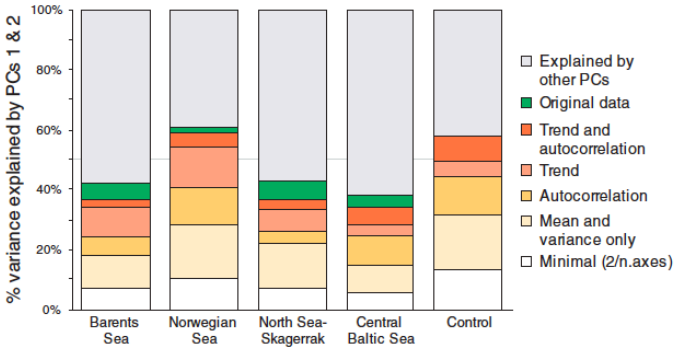

Figure 7. Fraction of the variance explained by the first two axes of the PCAs. From bottom to top: a) minimum possible amount of variance that can be explained by the first two PCS; median variance explained in PCAs performed on simulated data that incorporate b) mean and variance only, c) autocorrelation, d) trend, e) trend and autocorrelation; f) variance explained in the empirical dataset. The upper boxes indicate the residual variance not explained by the first two axes of the PCA. In the control dataset, the variance explained by the PCA on observed data is not greater than that obtained on simulated dataseries with similar trend and autocorrelation. (reproduced from Planque and Arneberg, 2018)

Important issues

Long-term **trends** in the dataseries that are used as input to PCA constitute simultaneously information and nuisance. Correlations between different variables can occur at different time-scales, including long-term, in which case the trends contain important information about how different ecosystem components are related. Temporal changes in the whole system are also expected to occur at multiple time-scales and long-term trends in the dataset contribute to the overall pattern of change. However, PCAs are primarily designed for cross-sectional datasets, not for time-series, and cannot be used to discriminate between changes in variables that can be attributed to time-dependency and those that can be attributed to causal interactions between variables.

Association between time-series (as revealed in the PCA biplots) cannot be attributed to causal relationships and in many cases, may simply reflect that several series display quasi monotonic long-term trends.

In many PCAs performed on IEA dataseries, the high percentages of variance explained as well as the temporal trajectories mainly are **recurrent emerging patterns** which result from the presence of monotonic trends. These are likely to be artefact of the method.

Variable selection is a key step in PCA analysis. If many variables are associated with the same ecosystem process or component (e.g. temperature in January, temperature in May, heat content and ice extent are all related to ocean thermodynamics), then this process/component is likely to dominate the PCA results (i.e. drive the first(s) component(s)).

Dataseries are often preprocessed before entering the PCA analysis. **Preprocessing** may involve data-transformation (e.g. log or double-square root transformation for biotic datasets), or selection of time-series in order to limit the redundancy of variables and avoid upweighting of certain processes (e.g. including sea surface temperature, ice extent and heat content may give too much weight to heat related processes and can be reduced to one variable only). The preprocessing steps, as well as the rationale behind them, should be presented explicitly.

PCA is not suitable to represent **uncertainties** in the underlying dataseries, since only the point estimates are used as input to the analysis.

The group did not discuss in depth the **criteria for assessing the performance of PCA analysis in the context of ITAs.** System trajectories in PC1-2 and percentage of variance have been used as evaluation criteria by Planque and Arneberg (2018).

There appear to be **two main types of objectives** when using PCA for ITA. First, PCAs have been used as an **exploration tool to reveal association between variables** (biotic, abiotic or both). In this case, practitioners should carefully consider the risks of spurious correlations due to trend and autocorrelation in the time-series as these can lead to artefactual high performance of the PCA in terms of variance explained and associations between variables. Second, PCAs have been used **to provide a summary of a dataset** in which the different variables are *a priori* known to be related to each other (a typical example is the use of PCA or EOFs (Empirical Orthogonal Functions) to summarise temporal changes in spatial fields of sea level pressure) if the data can be assumed stationary. In this case, the summary provided by the PCA may be meaningful, although the constraints imposed by orthogonal axes may not always be appropriate.

Additional considerations

PCA is useful for dimensionality reduction, lossy data compression, feature extraction and data visualization however (Bishop 2006); as we demonstrated in this workshop, applying PCA without considering the statistical properties of the data can provide a misleading interpretation to the output – which is common sense in time-series analysis (Kawasaki 2004). PCA is originally used for multivariate analysis, which these data are basically assumed to be sampled independently. Applying PCA directly to the data in the time domain means that we just focus on the correlation of the variables at an identical time point, since the eigen decomposition is performed on the instantaneous covariates, disregarding the information contained in the data's leads and lags. Therefore, it is hard to capture the dynamics behind the data based on the results by PCA by using the time domain. As a possibility for applying PCA to time-series data, Brillinger (1981) proposed the concept of a dynamical PC. The time-series data are converted by discrete Fourier transformation. The procedure guarantees asymptotical independence in the samples, enabling the ordinary PCA technique to be valid in the frequency domain. Brillinger's concept is based on the eigenvalue decomposition of spectral density matrices instead of the eigenvalue decomposition of covariance matrices.

PC captures the fluctuation maximizing the variability of the data, which first and second PCs usually cover over 90 percent of whole variance of the data. If the data includes long-term trend (non-stationary mean), shown in subsection 3.4, PCs just distinguish similar trend pattern. Therefore, nonstationary mean time-series data observed in marine biology, ecology and/or ocean geography should be applied trend model at once and would be better to test time dependency in the component by extracting trend pattern.

Why do we analyse multidimensional time-series data observed in oceans? The scientists engaged in marine biology, ecology and/or ocean geography are originally interested in-dynamical changes or the diagram of relevant biological species and geographical/climate environmental factors in a marine ecosystem. For this challenge, scientists should directly investigate mutual and causal relationships among the data. The first attempt for causal inference between variables goes back to a study on feedback systems by Wiener (1958), where, by his definition, a given time-series is *causal* to another if knowledge of the first series reduces the mean square prediction error of the second-time-series. Granger (1969) followed this notion of causality and applied it to the analysis of economic time-series data using bivariate AR model. A parallel development to Granger's approach was made by Akaike (1968), who provided a feedback system analysis using a MAR model. Akaike's approach was a practical statistical method to investigate mutual relationships among variables from two different angles – the open/closed impulse response calculated in the time domain, and the relative power contribution calculated in the frequency domain. Numerous successful applications of the method have since been found in several fields such as engineering, physics, economics and medical science (Akaike and Kitagawa, 1994). Therefore, applying MAR model is one possible way to analyse multidimensional time-series data observed in oceans. Notice that applying MAR to less than 50-time sampled points data should be done carefully (Hardison *et al.*, 2019; Solvang and Subbey, 2019), and the physical property or precision should be validated by resampling.

Recommendations

IEA groups should explicitly state their objectives when performing a PCA, i.e. exploration of associations between variables vs. summarising of a dataset. Applying PCA for the latter objective is recommended.

Data preprocessing (data selection and transformation) and methodological choices (use of correlation or variance-covariance matrix) must be justified and reported.

High percentage of variance explained by the first components is not advisable as a criteria for justifying the success of the method, unless this has been compared with the same measure on *control* datasets as in e.g., Planque and Arneberg (2018).

3.3 IRA

Method description

The Integrated Resilience Analysis (IRA) framework is a multivariate, 3-step approach developed by Vasilakopoulos and Marshall (2015). In the first step, the ecosystem components of interest are reduced to a single (or few) system indicator variable(s) using a dimension reduction method such as the Principal Component Analysis (PCA). The ecosystem state, e.g. the first principal component (PC1) of the biotic variables, is then regressed against a single pressure or stressor (Vasilakopoulos *et al.*, 2017) or a composite indicator, such as the PC1 of a pressure-based PCA (Vasilakopoulos and Marshall, 2015). The presence of a bifurcation pattern is identified if the response curve is not only non-linear but can be best described by two response functions separated in time, i.e., if a threshold (non-additive) formulation of a Generalized Additive Model (aka tGAM) performs better in terms of predictive performance than an ordinary fully additive GAM. In the case of a discontinuous system response, the lines of the fitted (two or more) GAMs represent the alternative attractors or basins. The position of the tipping points can then be approximated based on the thresholds identified by the non-additive models, and one or more fold-bifurcations can be revealed.

Once a fold bifurcation pattern is detected, annual resilience values are estimated based on the position of each year in relation to the fitted attractors and assumed tipping points of the fold bifurcation. By interpolating these annual resilience values, a folded stability landscape can then be fitted.

Data used

The IRA was applied to observational data and surrogates (section 2.1) from the following five regional IEAs: Barents Sea, Bay of Biscay, Central Baltic Sea, Norwegian Sea and North Sea / Skagerrak. Out of the 1000 available surrogate datasets 100 datasets were randomly chosen to shorten the computation time. All biotic time-series were considered as ecosystem components of interest and *a priori* fourth root-transformed to normalize the data. All environmental variables were considered as stressors.

Evaluation approach

Two IRAs were conducted for each case study and the observational time-series: one in which the biotic PC1 was regressed against the environmental PC1 (following Vasilakopoulos and Marshall, 2015) and one where the pressure with the strongest correlation was used as explanatory variable (see Vasilakopoulos *et al.*, 2017). This dual approach was then adopted for the 100 surrogate datasets, here using the same pressure as in the observation-based analysis for comparability. Since the direction of the PC1 has little meaning, the axis was allowed to rotate in the last step of the resilience estimation to better reproduce a potential fold bifurcation pattern.

The evaluation of the IRA was then conducted in two phases. First, a more quantitative approach was used in which the observation-based results from the first two steps of the IRA, i.e. the PCA and the GAM/tGAM modelling, were directly compared with the distribution of surrogate-based results. The assumptions here were that

- the biotic and environmental PC1 of the observation-based PCAs explain significantly more variance in the multivariate dataset than when using surrogate data.
- the fold bifurcation is more pronounced in the observational data, expressed as higher explanatory power in the GAM and tGAM and a lower probability of tGAMs outperforming the corresponding GAMs in the surrogate datasets.

Second, a visual comparison of the resilience landscape was made. Participants were provided with six shuffled resilience landscapes for each case study: one based on the true observations and five based on the surrogates. Participants were then asked to identify the observation-based assessment result.

Results

The results of the quantitative evaluation (phase 1) show that outputs from the observational datasets are not necessarily different from the surrogate datasets. Explained variances of the observation-based PCAs were only found for the Barents Sea (environmental PC1; see also Figure 8, left) and the Central Baltic Sea (both PC1) to be significantly higher. Also, the fold bifurcation pattern was not more pronounced in the observational time-series as one would have expected. In fact, the tGAMs outperformed the GAM not only in all observational datasets but in 90-100% of all surrogates (Figure 8, top-right). An exception is here the North Sea with 58% when using the environmental PC1 as covariate. In five out of the six case studies, the surrogate-based tGAMs explained more variance than the control, i.e. based on observations (see Figure 8).

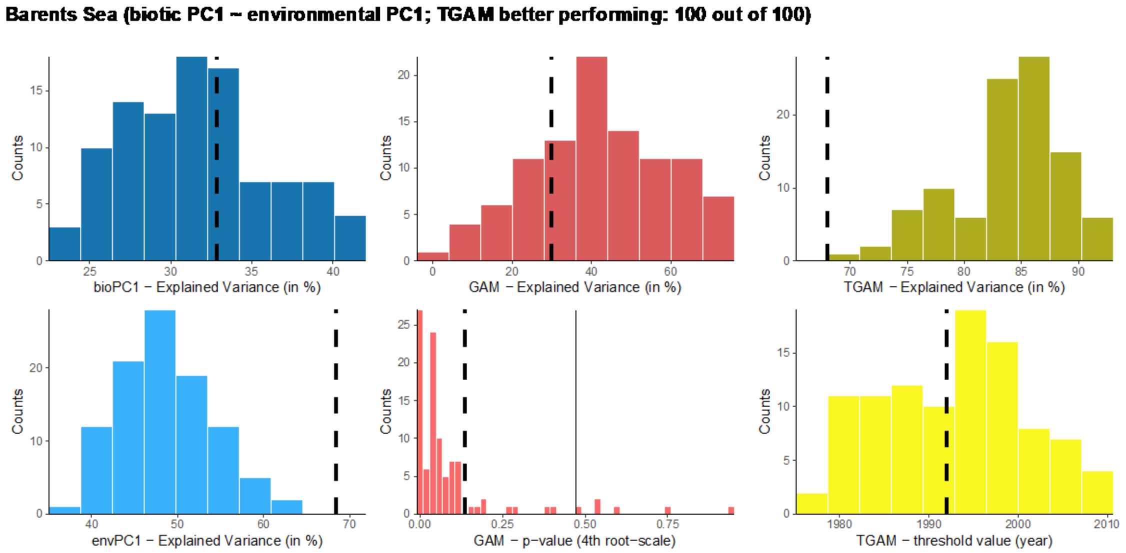

Figure 8. Distribution of surrogate-based analyses outputs in contrast to observational-based output (dashed black line): explained variance of the first component of the biotic and environmental PCA (left column), the explained variance and p-value of the GAM (middle column), as well as the explained variance of the TGAM and the identified time threshold (right column).

The identification of the observation-based resilience assessment from the six plots shown to the participants revealed challenging. A potential fold bifurcation pattern was often found for multiple datasets, both for the observations and surrogates (see Figure 9). However, none of the studied assessment plots showed such a distinct pattern as demonstrated in the original papers.

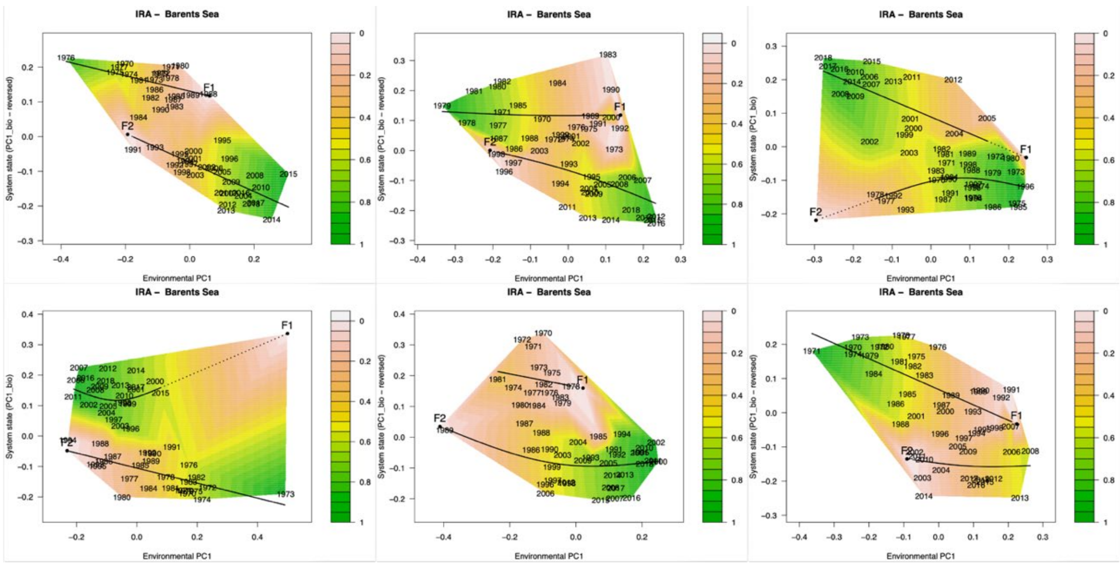

Figure 9. Folded stability landscape and resilience assessment for the observation-based analysis (upper middle panel) and five surrogate-based analyses for the Barents Sea. On the empirical folded stability landscapes continuous black lines indicate the linear attractors, dotted black line indicates the possible extension of the lower and upper branch and F1 and F2 indicate the tipping points. Colors represent the relative resilience contour interpolated from the relative resilience of each year.

Important issues

The **main objective** of the IRA framework is to test the theory of critical transitions, which suggests that complex natural systems impacted by multiple stressors may exhibit fold bifurcations featuring folded response curves, tipping points and alternate attractors. As confirmative method, however, there is the danger of misuse particularly as the method tends to easily generate the pattern looked for.

Trends: the PCA-based IRA framework faces a similar dilemma as the PCA long-term trends in biotic time-series might not necessarily be driven by the same exogenous drivers but rather by internal auto-regressive processes, as often found for fish species or other long-living species. However, the method cannot differentiate between the underlying processes and instead interprets similar trends as system-wide dynamics that can be related to the same stressor or stressor combination (in the second step).

Emerging systematic patterns: the modelling approach in the IRA framework shows the tendency for inflated type I error and overfit in the GAMs and particularly tGAMs. Splitting the response curve into two or more time periods (i.e. using 'time' as the threshold variable) leads generally to a better fit to the data on which the model is trained. Flexible modelling approaches such as the tGAM (but also GAMs) have the tendency to overfit the data and have higher variances, i.e. small changes in the training data can result in large changes in the estimated coefficients (James *et al.*, 2013). When then allowing the biotic PC1 (i.e., the y-axis) in the resilience landscape plot to rotate, a fold bifurcation pattern is easily emerging.

Methodological choices: as with many multivariate analyses, variable selection is a key step also in the IRA. Any overrepresentation of a specific component group is likely to dominate the system-wide analysis. Similarly crucial is the data preprocessing, such as transformation to normalize the data for the PCA, although GAMs and tGAMs allow for any distribution as long as they are clearly defined in the modelling function. In general, all preprocessing steps, as well as the rationale behind them, should be presented explicitly.

The IRA framework is neither suitable to represent **uncertainties** in the underlying dataseries nor in the estimated resilience of the system.

The group did not discuss in-depth further **criteria for defining or estimating IRA performance** as done in the evaluation here.

Recommendations

As recommended for the PCA, data preprocessing (data selection and transformation) and methodological choices (use of correlation or variance-covariance matrix) must be justified and reported.

The decision for selecting the tGAM over the ordinary GAM should not only be made using the genuine Cross-Validation (gCV) score but also visually:

- Based on the development of the Generalized Cross-Validation (GCV) score estimated for the sequence of models with a different threshold value ('valley plots', see Figure 2 in Llope *et al.*, 2011 for examples). Only if the selected tGAM has also a distinct lower GCV than all other theoretical tGAMs and if both regimes have sufficient data points to allow the estimation of the response curve, should the tGAM be considered optimal.
- Based on the confidence intervals of the two (or more) GAMs that describe the two (or more) basins of attraction. Only if there is no (or not much) overlap, i.e. if the basins of

attraction are clearly distinguishable, should the tGAM be retained over the fully additive GAM.

Following the evaluation scheme outlined here, we recommend that the performance of the IRA should be assessed by comparing observational-based results with simulated results, e.g. based on surrogates. This helps to identify the sensitivity of the method to generate the theoretical pattern when applied on these data.

3.4 MAR-T

Method description

Multivariate autoregressive models with trends (MAR-T) are constructed with the following procedures: 1. Apply the polynomial trend model to the observed time-series data. The order of polynomial trend model is set up to three and the optimum order is selected by an information criterion; 2. Apply MAR model to the residual obtained by extracting the estimated trend from the observation; 3. Using the estimated MAR coefficients and trend, the simulation data are generated by the time-series model with random noise that obeys to normal distribution with zero mean and estimated variance (covariance); 4. Combine the generated MAR data and the trend that was estimated in procedure 2. Using these procedures, 100 MAR-T time-series dataset is generated.

Data used

The observed time-series data includes HERSSB, SPRSSB, Acrtia_Spr, Temora_Spr, Pseudo_Spr, Chla-GBSpr, dia_GB_spr and dino_GB_spr from the dataset for Central Baltic Sea. By excluding missing values, the data includes 32 time points (1975-2013). In addition to the 100 MAR-T dataset, two kinds of surrogates, 1000 datasets for phase randomization and 1000 datasets for time-domain randomization with the same mean and variance of the estimated MAR part (from original data) are prepared. Finally, these data are also combined with the estimated polynomial trend parts.

Evaluation approach

To 100 MAR-T datasets, 1000 phase randomized datasets and 1000 time-domain randomized datasets, procedures 1 and 2 explained in 'Method description' are applied again. Using the estimated AR coefficients and variance-covariance matrix, power spectra for the three different datasets are calculated to confirm the physical properties. Then, PCA is applied to the three different datasets, and the first and second PC scores are compared.

Results

The power spectra for the 100 MAR-T datasets and 1000 phase-randomized datasets are similar to the power spectra for the original data. However, the power spectra for 1000 time-domain-randomized datasets indicate white noise.

The first PC scores show similar patterns for all cases while the second PC scores are varied. The patterns of the first PC scores correspond to the estimated trend patterns as shown in Figure 10.

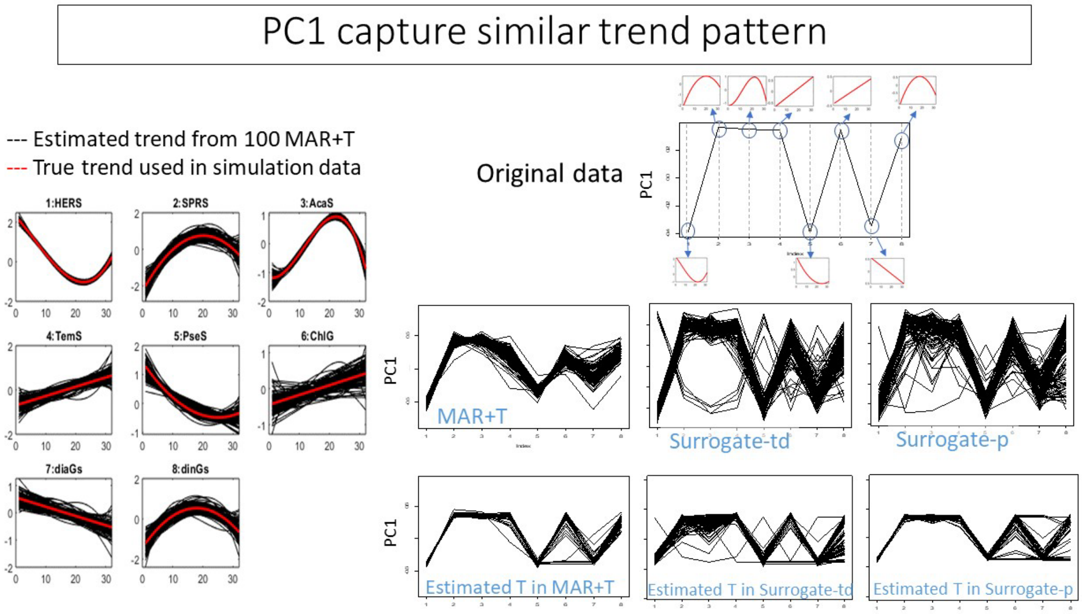

Figure 10. The eight panels in left-hand side indicate the estimated trend patterns (black lines) when applying procedure 1 in Method Description to 100 MAR+T data. The true trends estimated from observation is illustrated by red lines. In right-hand side, top panel indicates the first PC score for original data. Larger scores correspond to rising trend and lower scores correspond to declined trend. The middle three panels show plots of the first PC scores in the case of 100 MAR+T, 1000 time-domain-randomized and 1000 phase-randomized surrogates. The lower three panes show plots of the first PC scores in the case of estimated trends given by procedure 1 to 100 MAR+T, 1000 time-domain-randomized and 1000 phase-randomized surrogates. Plots of the first PC score in all cases show similar tendency to the case of original data.

Important issues

Even if the statistical properties for time domain randomization data are different, the first PC sores of all types show similarity. When applying PCA to the estimated trend component, the first PC demonstrates similar tendency; that is, it distinguishes between upward and downward configuration.

Recommendations

The observed time-series data usually present non-stationary mean process, and time-series trend model should be applied to estimate trend pattern. Based on the above results, it should be obvious that PCA is not suited to distinguish trend pattern. Moreover, theoretically the method is not appropriate to model non-stationary mean process.

3.5 TREC

Method description

Common trends refer to trends that are similar across ecosystem components time-series. Identifying common trends can be useful as a diagnostic tool to reveal past changes and to explore the relationships among biological communities, as well as between these communities and environmental conditions. In the present investigation, trend estimation and classification (TREC) has been proposed in Solvang and Planque (2020). The method is based on two statistical procedures that includes trend modelling and discriminant analysis for classifying similar trend (common trend) classes. TREC includes two different kinds of parametric trend models, a polynomial regression model and a stochastic *d*th order difference equation model. The optimum order of the polynomial trend model or difference equation model is selected by an information criterion (Akaike information criterion, AIC, Akaike, 1974). For the estimated trends, a two-category discrimination procedure is applied to roughly divided them into three groups representing configurations related to upward, flat and downward. If it is necessary to classify them into groups of more concrete patterns from the three groups, multiple-category discrimination is applied to the reference trend that analyser is interested in. The reference trend is assigned as a general icon.

Data used

We applied TREC to the time-series complied by the ICES integrated assessment working groups for the Barents Sea (WGIBAR) and the Norwegian Sea (WGINOR), including abiotic, biotic and human impact variables.

Evaluation approach

A simulation clarified the performances of two different trend models and their flexibility for use with several representative patterns, which can be predefined as icons.

Results

The trends for abiotic, biotic and human impact data were estimated and classified into common trend groups, e.g. Table 1 summarized classified groups of WGINOR dataset consisting of 24 annual time-series for the period 1995–2015.

Abiotic	1: ArW	Area of Arctic water
	2: AW	Area of Atlantic water
	3: BSO	Barents Sea opening
	4: FB	Fugløya-Bear Island section
	5: KolaS	Salinity level in Kola section
	6: KolaT	Atlantic water temperature in Kola section
	7: NAO	Winter North Atlantic Oscillation index
Biotic	8: Jel	Jellyfish biomass, mostly *Cyanea capillata*
	9: Krill	Krill biomass
	10: Shr-a	Relative shrimp stock biomass from assessment
	11: Cap0n	Abundance 0-group capelin
	12: CapS	Capelin SSB, spawning-stock biomass
	13: CapT	Capelin TSB, stock biomass (age 1+)
	14: Her0n	Abundance 0-group herring
	15: Her1	Herring stock biomass (ages 1 and 2)
	16: Lump1	Number of lumpfish age 1 and older
	17: LumpJ	Number of juvenile lumpfish
	18: Polcd0n	Abundance 0-group polar cod
	19: Cod0n	0-group cod abundance
	20: CodRe	Recruitment of cod at age 3
	21: Cod3 + b	Cod stock biomass (age 3+)
	22: GH0n	0-group Greenland halibut abundance
	23: Had0n	0-group haddock abundance
	24: HadR3	Recruitment of herring at age 3
	25: HadSSB	Spawning-stock biomass of haddock, ages 6–8
Human impact	26: CdF510	Fishing mortality of cod, ages 5–10
	27: HadF47	Fishing mortality of haddock, ages 4–7
	28: RelFc	Relative fishing mortality calculated as sum of catches of capelin in fall and next spring divided by biomass in August/September
	29: RelFS	Relative fishing mortality of shrimp
	30: HarpSL	Landings of harp seals
	31: MinL	Landings of minke whales
	32: PolcdL	Landings of Barents Sea polar cod
	33: ShrL	Landings of shrimp in Barents Sea

Common trend configuration	Category	Classified data	Icon
Upward		19:Cod0n, 21:Cod3+b, 25:HadSSB	
		2:AW, 4:FB, 5:KolaS, 6:KolaT, 8:Jel, 17:LumpJ, 24:HadR3	
		14:Her0n, 20:CodRe	
		9:Kril, 16:Lump1	
Flat		23:Had0n, 32:PolcdL	
		3:BSO, 15:Her1, 18:Polcd0n	
		7:NAO, 11:Cap0n, 12:CapS, 13:CapT, 32:PolcdL	
Downward		10: Shr-a, 22:GH0n, 28:RelFc, 31:MinL	
		1:ArW, 29:RelFS, 30:HarpSL, 33:ShrL	
		26:CdF51, 27:HadF47	

Table 1. Results of discrimination analysis for WGIBAR data: assigned icons by multi-category discriminates.

Important issues

For trend estimation, polynomial trend and stochastic difference trend models are available in TREC. Polynomial trend model gives simple configuration as the polynomial order is set up to three. Simple trends are easily classified into understandable configuration groups and easily assign to general icons. On the other hand, higher order polynomial trend or stochastic difference trend model gives more complicated trend pattern, which becomes difficult to assign to simple general icons simply. Depending on the aim, analyser can choose whether focusing on getting precise trend estimates or focusing on classification of the trend.

Recommendations

TREC method focuses on long-term trends in time-series data, it works for any length for sampled time points and is robust for application to short time-series. It directly identifies and classifies the dominant trends underlying observations.

In contrast to ordinal multivariate analysis, assumption of Gaussian or identical distribution is not necessary for the data to apply TREC.

The identified common trend patterns are simple comparing with common trend obtained by Dynamic Factor Analysis. Based on classified trend groups by TREC, communication among stakeholders can be enhanced by showing the common tendency between a biological community in a marine ecosystem and the environmental factors as well as the icons allocated (Table 1) by generalizing common trend patterns. Furthermore, the classified groups could help to consider further numerical analysis for investigation of precise ecosystem function.

3.6 MAFA

Method description

Minimum/maximum autocorrelation factors (MAFs) is a multivariate statistical method that aims to maximise the autocorrelation between neighbouring observations (Shapiro and Switzer, 1989; Woillez *et al.*, 2009; Doray *et al.*, 2018). When applied to multivariate time-series, minimum/maximum autocorrelation factors analysis (MAFA) decomposes the original suite of variables into a series of MAFs, in which autocorrelation decreases from the first factor to the last. Selecting the first MAFs enables to extract the time-series components that are the most continuous in time, in a way akin to selecting the first sinusoid cycles in a Fourier transformation. MAFA is based on the implementation of 2 successive PCAs: the first one transforms the multivariate time-series into principal components (PCs). The second PCA is performed on the PCs increments, so as to maximise/minimise their variance, and hence the PCs autocorrelation at a chosen time-lag.

Data used

MAFA was applied to the ISIS-Fish time-series in order to evidence the main trends in the evolution of the Eastern English Channel fishery following the landing obligation implementation. The analysis was performed comparatively on the random runs and the deterministic one, in order to evaluate the sensitivity of the method to random noise. Because MAFA relies on PCA, the number of time-series should not exceed the number of observations in each series. A selection of the variables must therefore be done prior to the analysis. Woillez *et al.* (2009) advise to include the variables which normalised one-lag variogram values are below 1.

Evaluation approach

The evaluation concerned two aspects: First the ability of the method to retrieve the most important changes in the fishery following the landing obligation implementation. Second the robustness of the detection of these signals to random noise in the stock recruitment relationships. This second aspect is of particular interest because of the very nature of MAFA that tends to discard the most chaotic variables. It must be noted however that the introduction of noise only on certain species and on biological variables, may bias variable selection toward the deterministic species and toward the variables that are the most loosely related to biomass.

Results

Variable selection: As expected with modelled data, normalised one-lag variogram values are very low and almost always below 1. We decided to keep the maximum number of variables allowed (29) with the lowest normalised variogram values. These variables changed depending on the run. In particular, they greatly differ between the deterministic run and the random runs (Figure 11).

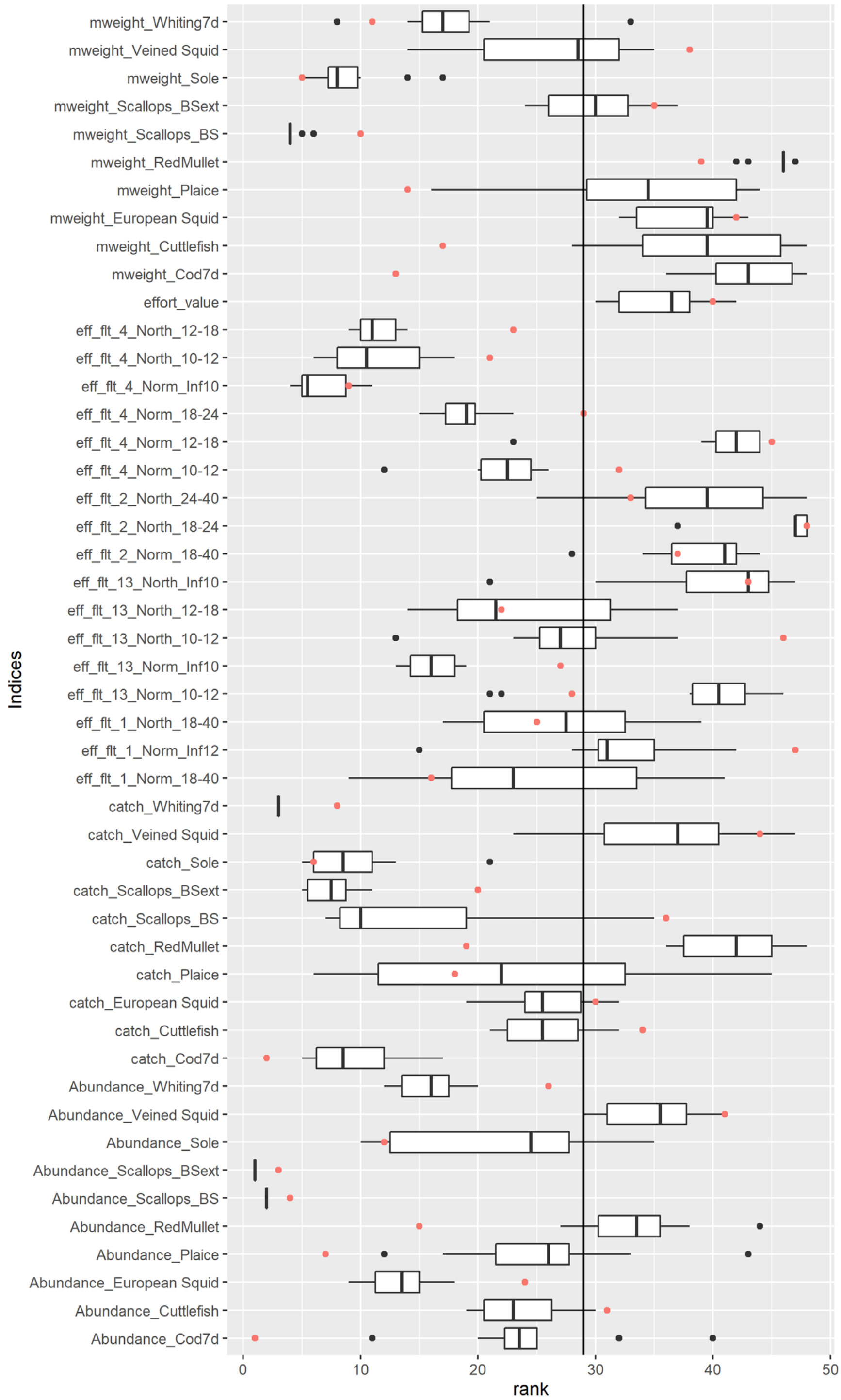

Figure 11. Ranks of the variables according to their normalised one-lag variogram values. The boxplots represent the variation of rank across random runs and the red dot the rank according to the deterministic run.

It was decided to keep the 29 most continuous variables of each run. Therefore, variables selected differed across MAFA. Despite a relatively different selection of variables across runs, the MAFs obtained are similar in pattern at least for the first three. However, the MAFs obtained from random runs are less smooth than the deterministic MAFs.

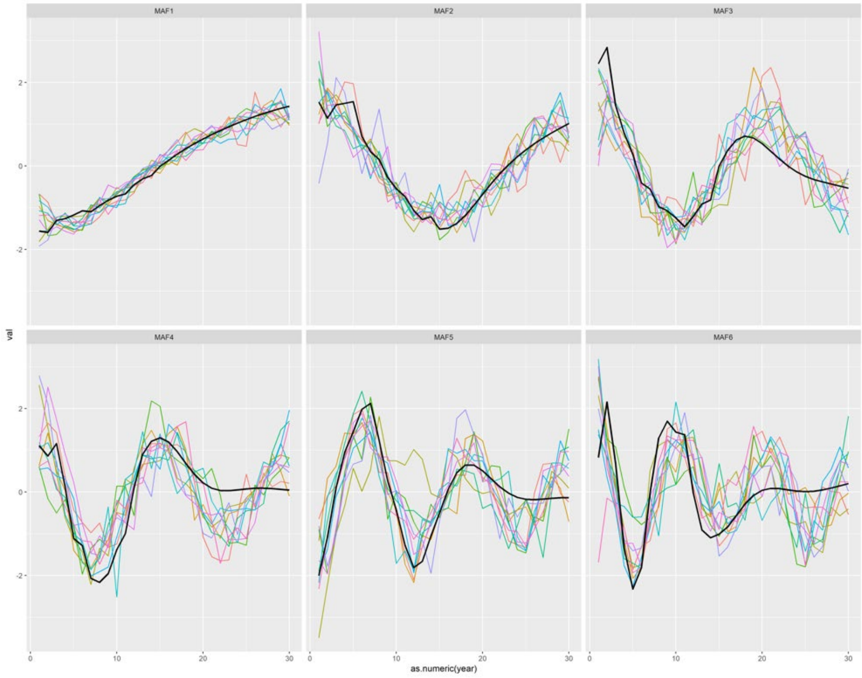

Figure 12. First six MAFs obtained with the deterministic run (black line) and the 10 randomised runs (thin coloured lines).

As for variable contributions to the first two MAFs, they are similar across runs (the top 4 are the same) (Figure 12). However out of the eleven most contributing variables to MAF 1 and 2, six are not common to the deterministic run and the random runs. Finally, these most contributing time-series only loosely resemble the MAFs. In terms of interpretation, MAF1 and MAF2 point to variables related to cod and scallops abundance, which display a spectacular increase over the period. However, it also selected plaice abundance which stayed almost constant and discarded whiting abundance despite an abrupt change.

Conversely MAFs 3 to 6 display more complex shapes to which less variables contributed and therefore resemble more to the original data. Interestingly the most contributing variables to these MAFs were related to sole, which is the species responsible for the choke of the system following the landing obligation implementation. This suggests that MAFs were successful at catching the informative patterns and select the variables supporting the dynamics.

The method aims at retrieving the most continuous patterns present in the datasets. In the present application, the system experienced a brutal change in dynamics following the implementation of a new management measure. The method might therefore not be the most appropriate to evidence this shift and should probably be reserved to the demonstration of more progressive changes. Nonetheless, higher MAFs were effective at pointing out the interesting dynamics of sole variables.

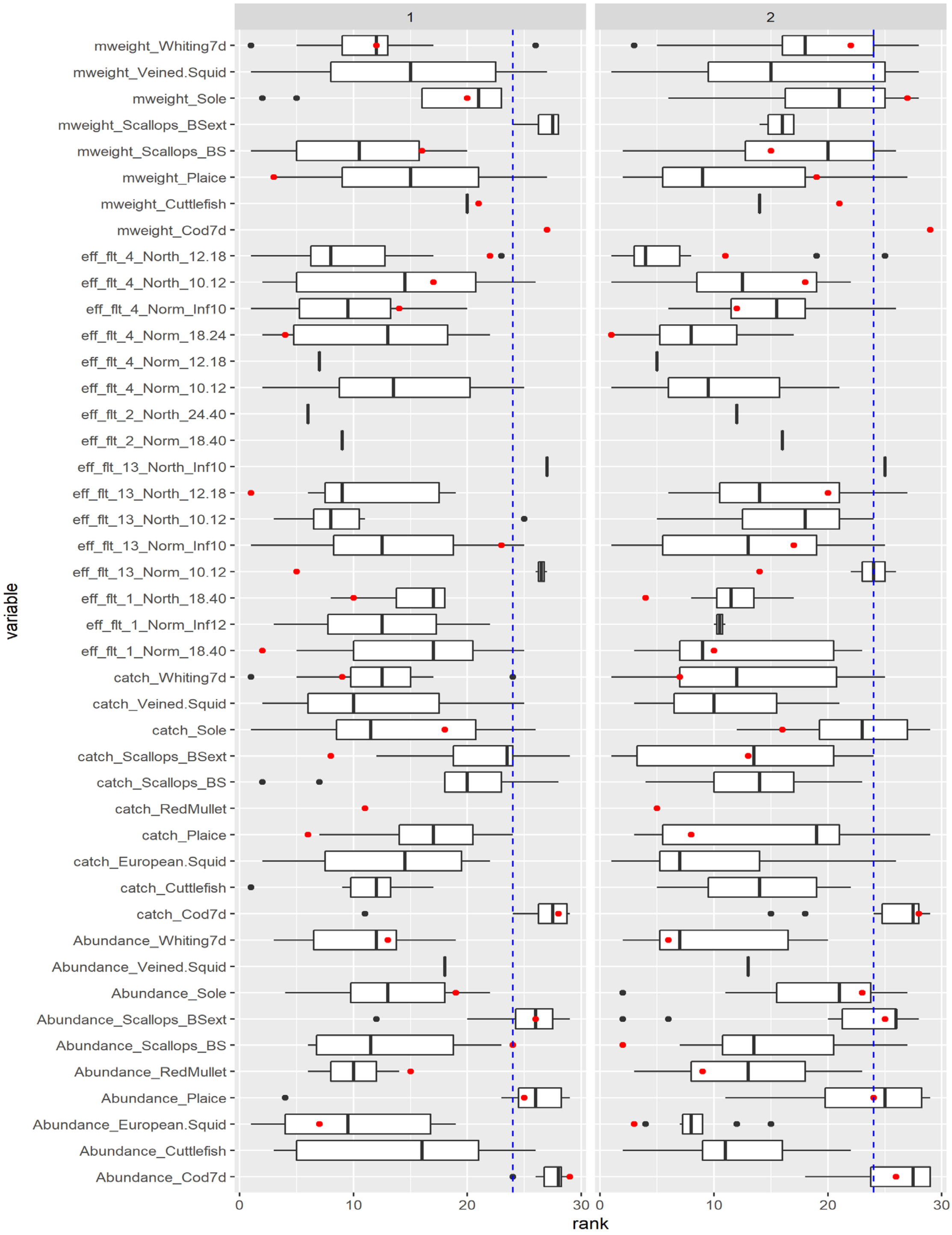

Figure 13. Ranks of variable contribution to MAFs 1 (left) and 2 (right). The box summarize the ranks of a variable over the random runs and the red dot indicates the rank of the variable in the deterministic run. Variables with ranks over 23 (on the right of the dashed blue line) are considered the most contributing.

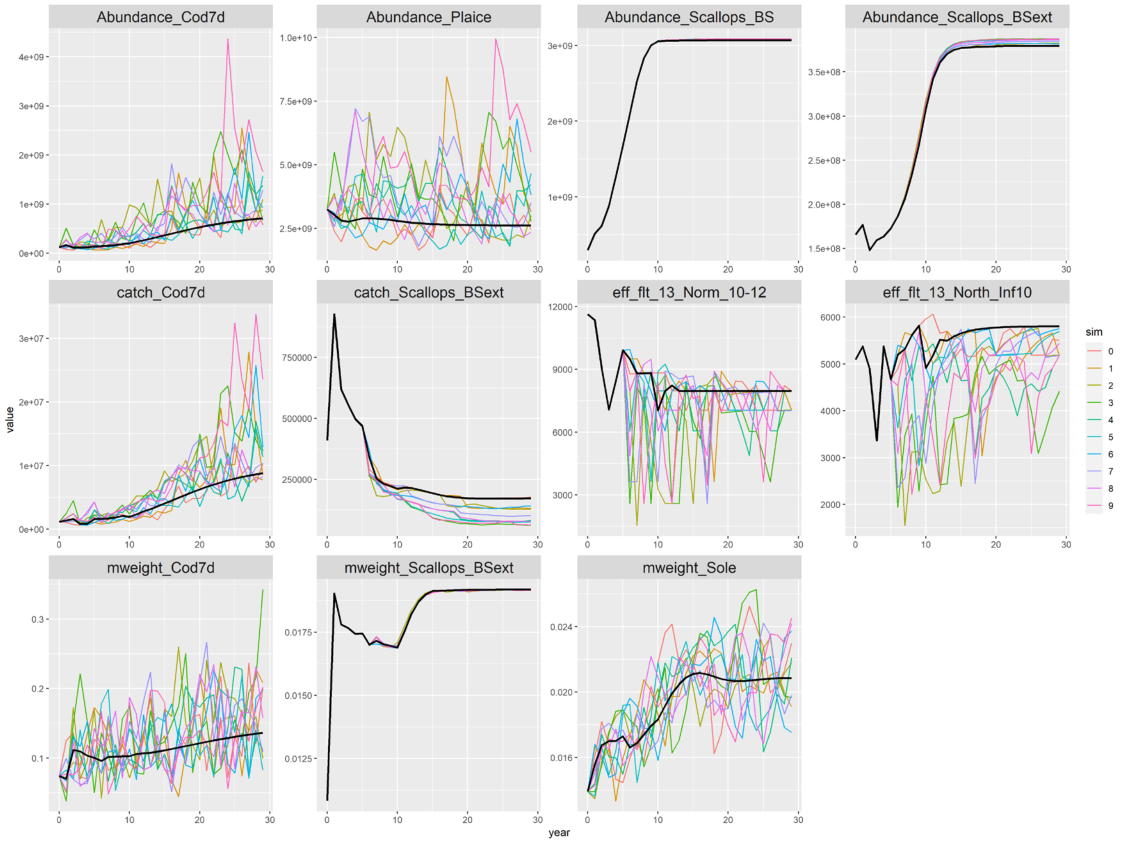

Figure 14. Time-series contributing the most to MAFs 1 and 2, obtained with the deterministic run (black line) and the 10 randomised runs (thin coloured lines).

Important issues

As for PCA, by construction the method tends to extract the linear trend and then cycles whatever the variables included. Therefore, the method can almost be seen more as a projection on orthogonal functions like Fourier transformation does on sinusoids. The difference lies in the orthogonal functions being built using the dataset and thus representing the smoothest trend orthogonal to the others but not a necessarily a perfect line or cycle which could be informative of breaking points or asymmetries.

The method requires that a limited number of time-series are considered which requires a selection. Different methods are proposed in the literature for the selection based on keeping the smoothest series (Woillez *et al.*, 2009; Doray *et al.*, 2018) which were unsuccessful in our case because of the time-series being model outputs they were all very smooth and the selection was not conclusive. An *ad hoc* method had to be chosen.

The application proves the robustness of the method to variable selection and random noise but also questions its relevance because it provides the same answer for datasets that are substantially different.

The method is meant at selecting the smoothest variables, therefore it is important to avoid mixing data of different nature that are expected to have contrasted levels of autocorrelation (such as model vs. observations or aggregated data (in time or space) vs. local measurement).

Recommendations

MAFA is an interesting tool to explore complex datasets composed of time-series. It has the advantage of explicitly accounting for autocorrelation in the decomposition of the variance of the dataset. MAFA allows selecting variables based on their level of autocorrelation or according to how much they resemble to temporal trends built using the full datasets.

MAFA seems to be robust to variable selection and random noise around temporal patterns. However, it could be of interest to try to project the variables left aside in the selection phase on the MAFs to allow a more throughout interpretation.

Depending on the question MAFA is intended to address, it might be relevant to stop at the first MAFs, if looking for progressive changes or to explore higher MAFs, when more complex dynamics or shifts are looked for.

The method presents the same issues than PCA regarding the orthogonal patterns that can be extracted from time-series and one should not expect the first MAFs to resemble time-series. Instead, not only the first MAFs should be explored to have an overview of the patterns present in the data, and the contributions should be exploited to select variables and evidence the mix of patterns they display.

It is recommended to use the MAFs and contribution to select the smoothest variables but to go back to original data for interpretation because the MAF can differ significantly from the time-series even when they contribute greatly.

As for PCA, MAFA should not be used to draw conclusions on causal relationships between variables because these would only rely on correlations.

3.7 RDA

Method description

Redundancy Analysis (RDA) is a constrained canonical ordination technique that combines multiple regression and principal component analysis (van den Wollenberg, 1977; Legendre and Gallagher, 2001; Borcard *et al.,* 2018). RDA computes orthogonal axes (e.g. RDA 1 and RDA2) whose linear combinations of explanatory variables best explain the variation in the response data. Unlike PCA, RDA can formally test statistical hypotheses about the significance of those relationships which provides opportunities for model selection techniques.

Using RDA as an ITA method is useful when the user intends to identify the most relevant explanatory variables, e.g. before a Dynamic Factor Analysis (DFA), from a rich dataset but poor *a priori* hypotheses. RDA can identify the most influential explanatory variables using model selection techniques such as forward selection, backward elimination and stepwise selection. After model selection the subset of explanatory variables identified in the reduced and more parsimonious RDA model can then be used as covariates in DFA.

Data used

RDA was applied to the observational Celtic Sea data and five surrogates simulated using phase randomisation (Section 2.1). The Celtic Sea data are relatively short (1997–2019) and contained missing values. The missing values for states in 2017 was imputed using a 5 year moving average (i.e. the mean of the two years before and after the missing value) and missing values at the end of the time-series (i.e. 2019) were estimated using a 3-year trailing moving average. Response data were fourth-root transformed to have a symmetric distribution and stabilise the variances. The transformed response variables and the explanatory variables were standardized to ensure all units were comparable. The transformed and standardised data used for this section is the same data used for the DFA section (section 3.8).

Evaluation approach

Full models were reduced to more parsimonious model using backward elimination for the observational and five surrogate datasets. The purpose was to compare the most influential explanatory variables identified in the reduced models of the observational and surrogate datasets. In the absence of dominant linear trends, we would expect to observe zero or at least fewer explanatory variables in the reduced model and different linear combinations for each surrogate.

Results

Many of the response and explanatory variables in the Celtic Sea observational dataset had clear linear trends. Mean fishing exploitation rate (Fbar) had a clear negative trend whereas many of the response variables tended to have a positive trend. The reintroduction of these linear trends in the phase-randomisation process makes the evaluation of RDA difficult because the time-series are autocorrelated, particularly fishing pressure variables. After backward elimination the reduced RDA model on the observational data identified the linear combinations of Fbar, Fcod, TLC, NPP and TSC as the explanatory variables that best explained the variation in the response data. The surrogates had fewer explanatory variables which would be expected for randomised data. However, 'Fbar' was the dominant explanatory variable on the first and most important RDA axis for the observational and most of the surrogates (Figure 15). The time trajectory was similar for all triplots which moved across the RDA1 axis from high to low fishing pressure indicating the presence of dominant linear trend (i.e. Fbar) in the set of explanatory variables. Similarly, response variables (e.g. plaice, john dory, rays and hake) retained their negative relationship with fishing pressure in the surrogate time-series. Except for fishing pressure, the linear combination of explanatory variables changed for simulation 2 and simulation 5 indicating that the method may be prone to type 1 errors possibly due to a combination of time-dependency structures in the data and small sample size. Overall, RDA appears to suffer from the same statistical artefacts as PCA. The first RDA axes captures the dominant trend in the set of explanatory variables and the first PC of the PCA on the matrix of fitted response variables captures the dominant linear trend in the response variables.

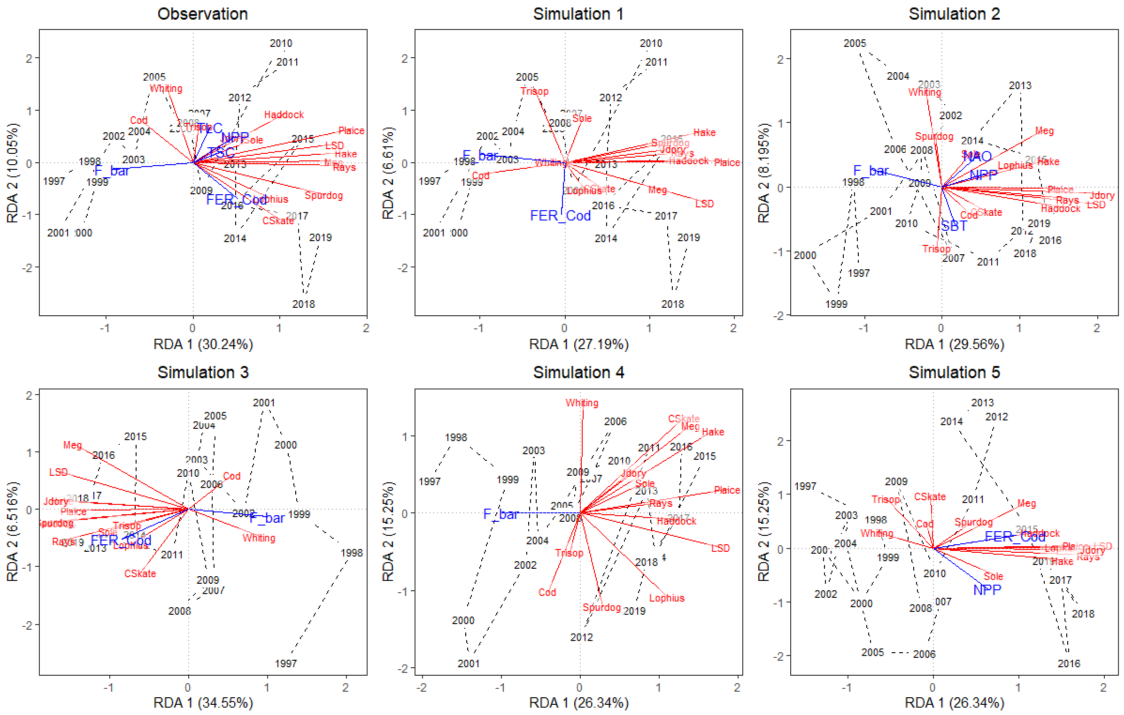

Figure 15. Triplots of the reduced RDA model for the observational and five surrogate dataseries. Blue lines indicate explanatory variables, red lines indicate response variables and the angles between the lines represent the strength and direction of their linear relationship.

Important issues

In the context of an IEA, the main **objective** of RDA is to **explore linear relationships** between multivariate data and multiple explanatory variables and then, via model selection, **identify the most influential explanatory variables**. However, relationships in ecological data are not always linear and the user needs to identify whether a linear or unimodal gradient analysis is appropriate by checking the gradient lengths of the response variables. Axes lengths greater than two to three standard deviations indicate that states are likely to be unimodal and Canonical Correspondence Analysis (CCA) may be more appropriate than RDA.

RDA is an **extension of PCA and therefore suffers similar issues** (see section 3.2). An advantage of RDA over PCA is that the user can state that the linear combination of explanatory variables on the axes *explain* (in the statistical sense) the variation in the response dataset. However, similar to PCA, the user needs to exercise caution when inferring causal relationships between variables due to time-dependency structures.

Variable selection to identify the most influential explanatory variables can be achieved using forward, backward and stepwise selection however strong linear dependencies in the explanatory variables need to be explored by computing variance inflation factors (VIF) to ensure stable regression coefficients.

Point estimates are used as the input into RDA so the method cannot represent **uncertainties** in the datasets.

RDA often use dataseries that are **preprocessed** by transforming the response variables to achieve normal (or at least symmetric) distributions and stabilise variances. Preprocessing steps should be justified and explicitly reported by the user.

Recommendations

IEA groups may have rich datasets but poor *a priori* hypotheses. In this context RDA can be used to explore and identify linear relationships and then search for parsimony. Due to time-dependency structures in the dataseries it is recommended that the method is used to reduce a large set of explanatory variables into a smaller subset that characterise the key pressures in the ecosystem. The subset of key pressures can then be used as covariates for methods that are computationally demanding but better suited to time-series analysis (e.g. MAFA and DFA).

Transformation to stabilise variances and standardization to achieve dimensionally homogenous variables must be justified and stated. Hellinger transformation may be more appropriate when using raw abundances to avoid double-zero issues (Legendre and Gallagher, 2001).

Permutation tests to assess the significance of global and reduced models are likely to be affected by time-dependencies structures in the dataseries.

3.8 DFA

Method description

Dynamic Factor Analysis (DFA) is a dimension reducing technique used to detect fewer common trends in a larger set of short non-stationary time-series and has been modified for fisheries ecology (Zuur *et al.*, 2003b, 2003a). It has two applications of which the second application is an extension on the first. The first application involves identifying *M* common trends in *N* time-series without considering the effects of explanatory variable(s) on the response variables/states. This application is useful when the objective is to identify groupings of states that share an underlying pattern. The second application introduces a linear regression component into the model which estimates state-specific relationships with covariates. The *M* common trends are now interpreted as underlying patterns in the temporal variation not captured by the effects of the covariates (i.e. underlying patterns in the residuals). The state-specific relationships with the covariate(s) in the model is quantified by regression parameters, standard errors and *t*-values. The DFA model with the linear regression component is useful when the objective is to explore relationships between pressures and states and the common trend(s) may represent latent variables which were not included in the original model. For both applications each state has a factor loading and a canonical correlation coefficient whose sign and magnitude indicate the direction and strength of the relationship between the *M* common trend(s) and each state. This allows the user to identify groupings of response variables/states which share an underlying pattern. The resulting *M* trends, factor loadings, canonical correlations and the effects of the covariates characterise the temporal variability of the states which can then be used by an IEA group to identify potential drivers of change in the ecosystem.

Data used

The same transformed and standardized Celtic Sea data used for RDA (section 3.7) was also used for the DFA section.

Evaluation approach

To gain a better understanding of emerging patterns, the structure of the DFA model was kept constant for each dataset and the linear regression component of the DFA model that accounts for the effects of the explanatory variables was omitted. The objective was to identify common trends in the states without taking into account the effect of covariates. The number of common trends to be estimated was set to two (M = 2) and to reflect that relationships between variables were removed in the surrogates we set the error covariance matrix to a diagonal matrix. It was expected that DFA on each surrogate dataset would estimate different common trends and also inconsistent groupings of states on the different common trends.

Results

The most dominant common trend (trend 1) was positive for both the observational and surrogate time-series (Figure 16). This is likely to be a result of the process used to generate surrogate time-series which preserved linear trends in the observational time-series. Many of observational time-series exhibited a positive trend and because the linear time-trend is reintroduced into the surrogates it is likely that DFA identified the reintroduced trends as an underlying pattern in each surrogate states time-series. It is unlikely that this is an emerging pattern of DFA but rather a consequence of the method used to generate the surrogate time-series. Interestingly, the less dominant trend (trend 2) varied for each dataset indicating that once the dominant linear trend (trend 1) was accounted for the remaining temporal variation did not have a consistent underlying pattern/trend and groupings of states which would be expected for the randomised data.

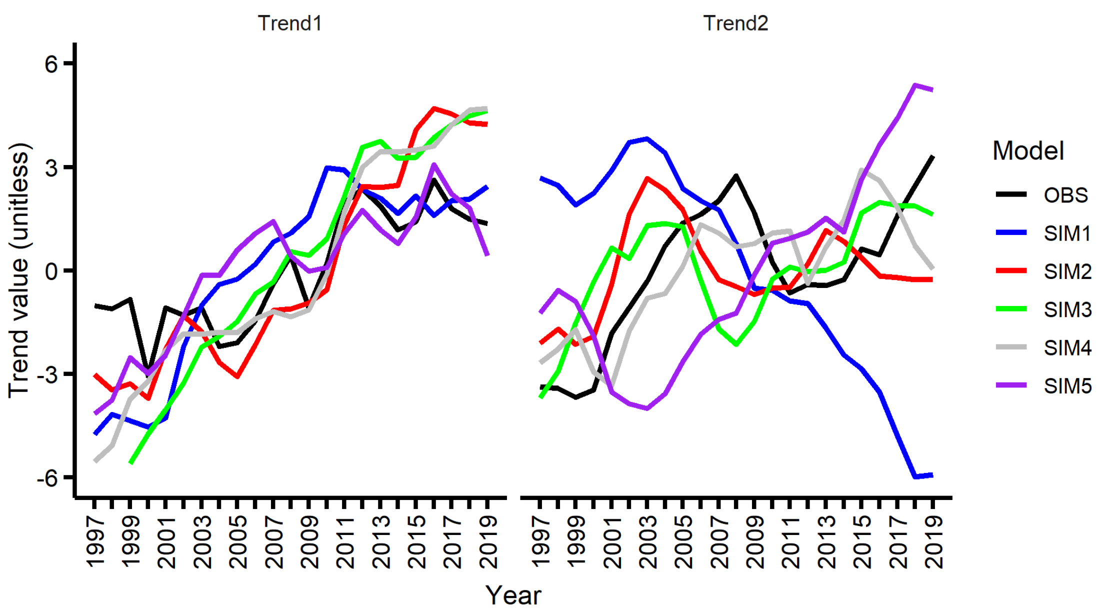

Figure 16. Common trends identified in the observational (black line) and surrogate time-series (coloured lines) using a DFA model with M=2 and R=diagonal.

The canonical correlations which quantify the association between the states/response variables and the common trends was consistent for trend 1 but varied for trend 2 (Figure 17). Once again trend 1 detected the reintroduced linear time-trends of states whose time-series were increasing with time whereas trend 2 and the groupings of species on that trend was less consistent which would be expected for randomised data.

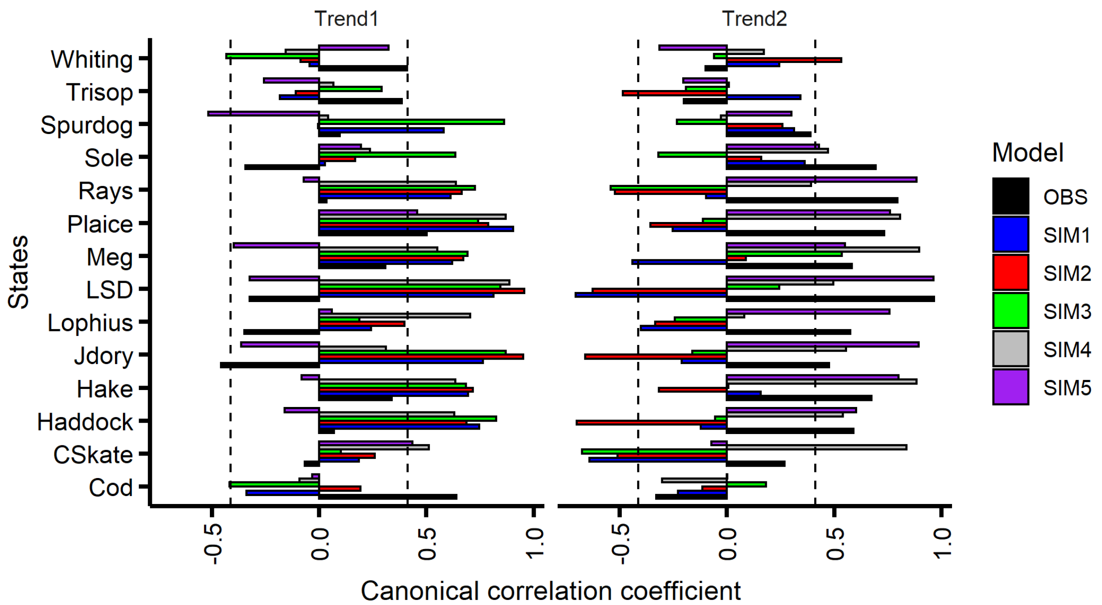

Figure 17. Canonical correlations between each state's time-series and the common trends. The horizontal line denotes the threshold for a statistically significant association.

Important issues

DFA is **useful for exploring and identifying common trends** within a set of time-series but by adding the linear regression component into the model it can **quantify relationships between each state and pressure** in the form of regression parameters.

The **common trends** are interpreted differently when the linear regression component (that accounts for causal linear relationships between states and pressures) is included in the model. Without the linear regression component, the common trend(s) describe underlying pattern(s) and the canonical correlations indicate the groupings of states that follow the underlying pattern(s). Whereas the addition of the regression component into the DFA model is used to quantify the effect of a pressure on each state and the **common trends are estimated in the residuals**. The common trends can then be used to identify variable(s) that were not included in the original model.

Trends provide useful information because the method is designed for short non-stationary time-series however, similar to MAFA, the number of variables included in the analysis cannot exceed the length of the time-series.

To get an idea of the **importance of each common trend** the user should apply the model with one common trend then apply a model with two common trends and compare the trends of both models. The dominant pattern is the trend in the model with two common trends which looks like the common trend in the first model.

Interpretation of a DFA model is easier when the covariates explain most of the variation in the data, otherwise the user has the task of explaining the meaning of the common trend(s) and groupings of states.

Model evaluation is similar to linear regression. Useful tools include residuals versus fitted values, QQ-plots and histograms of residuals.

DFA cannot handle **uncertainties** in the input data as it uses point estimates and **data needs to be standardized or normalised** prior to analysis.

Sometimes a common trend perfectly fits one of the response variables, this is known as a Heywood case and occurs in other forms of factor analysis. Using a non-diagonal matrix for the error covariance matrix often solves this problem. Using a non-diagonal error matrix can also reduce the number of common trends required for an adequate model fit albeit at the cost of increasing the number of model parameters and computing time.

Recommendations

IEA groups should identify the most influential pressures in the system prior to DFA to avoid overfitting the model, modelling crude variance and to reduce computational time. It is recommended that the user identifies the dominant pressures by exploring the data using other technique such as PCA (section 3.2) or using the model selection process in RDA (section 3.7).

The user should clearly indicate what the common trends describes when the linear regression component is added or omitted from the DFA model.

The user should standardise the explanatory variables to ease the interpretation of the regression parameters and avoid introducing highly correlated explanatory variables into the model.

AIC is initially used to determine the goodness-of-fit however the user should also compare patterns in residuals and fitted versus observed values. Sometimes the addition of a common trend to the model may lower the AIC but the additional trend is only important for a relatively small number of states. This makes interpretation difficult, and the user will need to decide whether to include the additional trend or use the less optimal model which is easier to interpret.

4 Guidelines to IEA groups

It is recommended that IEA groups explicitly **specify the objective of the ITA method** they employ and check that method is appropriate to address this objective. This issue was already identified in WKINTRA-2 where it was found that different IEA groups had different motivations for using the same ITA, or used different ITAs to address similar questions. Sometimes the match between objective and method is not clear or suboptimal (e.g. heatmaps used to investigate synchrony between time-series).

Special attention should be given to **transparency and traceability** of the ITA analysis and of the data supporting it. Reference to data sources, within or outside ICES databases should be provided. Data processing, including data transformation or interpolation should be explicitly justified and reported. The detail of methodological choices (e.g. category definitions for heatmaps, correlation vs. covariance matrix for PCA, and so on) should also be reported exhaustively, if necessary in supplementary documents or appendices.

WKINTRA ToR c aims to develop guidelines to evaluate ITA methods, and in collaboration with the IEA groups, could provide the incorporation of a new topic for the (EO) for each ecoregion on ITA evaluation following the ICES Technical Guidelines (ICES 2021).

None of the methods reviewed during WKINTRA-3 could account for uncertainties in the input data. The dataseries that support IEA are provided with variable degrees of certainty (e.g. sea surface temperature time-series are much less uncertain than zooplankton biomass time-series) and therefore provide information with different levels of quality/certainty. It is recommended that IEA groups and future WKINTRA-like workshops pay attention to **methods that can explicitly handle uncertainties in input dataseries**. It is also recommended that IEA groups report the uncertainties associated with the dataseries that support their IEAs. Presently, most IEA groups don't report these uncertainties.

Many methods currently used in IEAs focus on temporal trends and on the detection of ruptures in time-series (regime shifts). Methods for the **detection of extreme events** (such as heatwaves) would constitute an additional useful contribution to IEAs.

When appropriate, different IEA groups should seek for **harmonisation of the presentation of the ITA outputs**. Without being prescriptive, such harmonisation can be achieved by providing guidelines common to all IEA groups on e.g., colour schemes, time-series representations, icons to depict ecosystem components, and so on.

IEA groups are encouraged to **assess the sensitivity of ITA methods** to the selected input data and to **carefully consider the interpretation of ITA results** in light if the method's strength, weaknesses and possible misuse. This can be done by applying evaluation schemes like the ones used in the present report, based on simulated "control" datasets.

Methods that are recurrently used by one or several IEA groups will benefit from **peer-evaluation** within ICES. This will ensure that IEAs are supported by high quality science through internal peer-reviews, increased transparency, and replicability. There is currently no mechanism at ICES to achieve this for IEAs (although it exists for stock assessments). Future WKINTRA workshops could contribute to the process, but other mechanisms can be envisaged at the IEASG level.

5 References

Akaike, H. 1968. On the use of a linear model for the identification of feedback systems. Annals of the Institute of Statistical Mathematics, 20: 425–439.

Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19: 716–723.

Akaike, H., and Kitagawa, G. 1994. The Practice of Time-series Analysis. springer.

Audzijonyte, A., Pethybridge, H., Porobic, J., Gorton, R., Kaplan, I., and Fulton, E. A. 2019. Atlantis: A spatially explicit end-to-end marine ecosystem model with dynamically integrated physics, ecology and socio-economic modules. Methods in Ecology and Evolution, 10: 1814–1819. https://onlinelibrary.wiley.com/doi/10.1111/2041-210X.13272 (Accessed 29 October 2021).

Blenckner, T., Llope, M., Möllmann, C., Voss, R., Quaas, M. F., Casini, M., Lindegren, M., *et al.* 2015. Climate and fishing steer ecosystem regeneration to uncertain economic futures. Proceedings of the Royal Society B: Biological Sciences, 282: 20142809. https://royalsocietypublishing.org/doi/10.1098/rspb.2014.2809 (Accessed 29 September 2021).

Borcard, D., Gillet, F., and Legendre, P. 2018. Canonical Ordination. *In* Numerical Ecology with R, pp. 203–297. Springer International Publishing, Cham. http://link.springer.com/10.1007/978-3-319-71404-2_6 (Accessed 29 October 2021).

Doray, M., Petitgas, P., Huret, M., Duhamel, E., Romagnan, J. B., Authier, M., Dupuy, C., *et al.* 2018. Monitoring small pelagic fish in the Bay of Biscay ecosystem, using indicators from an integrated survey. Progress in Oceanography, 166: 168–188. https://linkinghub.elsevier.com/retrieve/pii/S0079661117300915 (Accessed 29 October 2021).

Granger, C. W. J. 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica, 37: 424–439.

Hansen, C., Skern-Mauritzen, M., van der Meeren, G. I., Jähkel, A., and Drinkwater, K. 2016. Set-up of the Nordic and Barents Seas (NoBa) Atlantis model. Fisken og Havet. Institute of Marine Research, Norway, Bergen, Norway.

Hansen, C., Nash, R. D. M., Drinkwater, K. F., and Hjøllo, S. S. 2019. Management Scenarios Under Climate Change – A Study of the Nordic and Barents Seas. Frontiers in Marine Science, 6: 668. https://www.frontiersin.org/article/10.3389/fmars.2019.00668/full (Accessed 29 October 2021).

Hardison, S., Perretti, C. T., DePiper, G. S., and Beet, A. 2019. A simulation study of trend detection methods for integrated ecosystem assessment. ICES Journal of Marine Science.

James, G., Witten, D., Hastie, T., and Tibshirani, R. 2013. An Introduction to Statistical Learning. Springer Texts in Statistics. Springer New York, New York, NY. http://link.springer.com/10.1007/978-1-4614-7138-7 (Accessed 29 October 2021).

Legendre, P., and Gallagher, E. D. 2001. Ecologically meaningful transformations for ordination of species data. Oecologia, 129: 271–280.

Lehuta, S., Vermard, Y., and Marchal, P. 2015. A spatial model of the mixed demersal fisheries in the Eastern Channel ». In 187-95., 2015. *In* Marine Productivity: Perturbations and Resilience of Socio-ecosystems. Proc. 15th French-Japan. Oceanogr. Symposium., pp. 187–195. H.-J. Ceccaldi et al.

Lehuta, S., and Vermard, Y. submitted. Contrasted impacts of the Landing obligation at fleet scale: a simulation study of the impacts of mitigation scenarios in the Eastern English Channel. ICES J. Mar. Sci.

Llope, M., Daskalov, G. M., Rouyer, T. A., Mihneva, V., Chan, K.-S., Grishin, A. N., and Stenseth, N. C. 2011. Overfishing of top predators eroded the resilience of the Black Sea system regardless of the climate and anthropogenic conditions. Global Change Biology, 17: 1251–1265.

Mahevas, S., and Pelletier, D. 2004. ISIS-Fish, a generic and spatially explicit simulation tool for evaluating the impact of management measures on fisheries dynamics. Ecological Modelling, 171: 65–84.

Pelletier, D., Mahevas, S., drouineau, H., Vermard, Y., Thebaud, O., Guyader, O., and Poussind, B. 2009. Evaluation of the bioeconomic sustainability of multi-species multi-fleet fisheries under a wide range of policy options using ISIS-Fish. Ecological Modelling, 220: 1013–1033.

Planque, B., and Arneberg, P. 2018. Principal component analyses for integrated ecosystem assessment may primarily reflect methodological artefacts. ICES Journal of Marine Science, 75: 1021–102.

Schreiber, T., and Schmitz, A. 2000. Surrogate time-series. Physica D, 142: 346–382.

Shapiro, D. E., and Switzer, P. 1989. Extracting time trends from multiple monitoring sites. Tech. Rep., 132. Department of Statistics, Stanford University. 132.

Solvang, H. K., and Subbey, S. 2019. An improved methodology for quantifying causality in complex ecological systems. PLoS ONE, 14: e0208078.

Solvang, H. K., and Planque, B. 2020. Estimation and classification of temporal trends to support integrated ecosystem assessment. ICES Journal of Marine Science, 77: 2529–2450.

Trapletti, A., and Hornik, K. 2020. tseries: Time-series Analysis and Computational Finance. https://CRAN.R-project.org/package=tseries.

van den Wollenberg, A. L. 1977. Redundancy analysis an alternative for canonical correlation analysis. Psychometrika, 42: 207–219. http://link.springer.com/10.1007/BF02294050 (Accessed 29 October 2021).

Vasilakopoulos, P., and Marshall, C. T. 2015. Resilience and tipping points of an exploited fish population over six decades. Global Change Biology, 21: 1834–1847.

Vasilakopoulos, P., Raitsos, D. E., Tzanatos, E., and Maravelias, C. D. 2017. Resilience and regime shifts in a marine biodiversity hotspot. Scientific Reports, 7: 13647.

Wiener, N. 1958. The theory of prediction. Modern Mathematics for Engineers. McGraw-Hill.

Woillez, M., Rivoirard, J., and Petitgas, P. 2009. Using min/max autocorrelation factors of survey-based indicators to follow the evolution of fish stocks in time. Aquatic Living Resources, 22: 193–200.

Zuur, A. F., Fryer, R. J., Jolliffe, I. T., Dekker, R., and Beukema, J. J. 2003a. Estimating common trends in multivariate time-series using dynamic factor analysis. Environmetrics, 14: 665–685.

Zuur, A. F., Tuck, I. D., and Bailey, N. 2003b. Dynamic factor analysis to estimate common trends in fisheries time-series. Canadian Journal of Fisheries and Aquatic Sciences, 60: 542–552.

Annex 1: List of participants

Name	Institute	Country (of institute)	Email
Per Arneberg	Institute of Marine Research	Norway	per.arneberg@hi.no
Andrea Belgrano	Swedish University of Agricultural Sciences, Department of Aquatic Resources, Institute of Marine Research, (SLU), and Swedish Institute for the Marine Environment (SIME), University of Gothenburg	Sweden	andrea.belgrano@slu.se
Lucie Buttay	Institute of Marine Research	Norway	lucie.buttay@hi.no
Mårten Erlandsson	Swedish University of Agricultural Sciences, Department of Aquatic Resources	Sweden	marten.erlandsson@slu.se
Bérengère Husson	Institute of Marine Research	Norway	berengere.husson@hi.no
Jed Kempf	Marine Institute and University College Cork	Ireland	jed.kempf@marine.ie
Sigrid Lehuta	Ifremer	France	slehuta@ifremer.fr
Marcos Llope	Instituto Español de Oceanografía (IEO, CSIC), Centro Oceanográfico de Cádiz	Spain	marcos.llope@ieo.es
Saskia Otto	Institute of Marine Ecosystem and Fishery Science (IMF), University of Hamburg	Germany	saskia.otto@uni-hamburg.de
Benjamin Planque	Institute of Marine Research	Norway	benjamin.planque@hi.no
Hiroko Solvang	Institute of Marine Research	Norway	hiroko.solvang@hi.no
Nigel Yoccoz	UiT The Arctic University of Norway	Norway	nigel.yoccoz@uit.no

Annex 2: WKINTRA-3 resolution

WKINTRA3 - The third workshop on integrated trend analyses in support to integrated ecosystem assessment

2019/WK/IEASG09 **The third workshop on integrated trend analyses in support to integrated ecosystem assessment (WKINTRA3)**, chaired by Saskia Otto, Germany, and Benjamin Planque, Norway, will meet in 20-24 September 2021 online.

The general objective of the workshop series is to develop good practices in the application of integrated trend analyses (ITA) and interpretation of their results for integrated ecosystem assessment. The third workshop will:

a) Review the simulated multivariate ecological datasets prepared during and following WKINTRA2 (Science plan codes 1.3 and 1.9)

b) Evaluate a selection of Integrated Trend Analysis (ITA) methods (Science plan codes 1.3 and 1.9).

   For this:

   - a set of ITA methods will be selected,
   - the R code to run the analyses will be provided,
   - method-specific qualitative or quantitative criteria will be defined that allow for an objective comparison across simulated datasets
   - the ITA methods will be applied on relevant simulated datasets outcomes will be assessed on a case study- and approach-specific basis

c) Develop guidelines for IEA groups to evaluate ITA methods, including a comprehensive documentation of data generation and method application using the R environment (Science plan code 6.5)

WKINTRA3 will report by 29 October 2021 for the attention of IEASG.

Supporting Information

Priority	The use of ITA is widespread in the ICES integrated ecosystem assessment community, and recent publications have challenged the interpretation of its results. Thus, the priority should be considered medium to high.
Scientific justification	The first workshop on integrated trend analyses in support to integrated ecosystem assessment (WKINTRA) recognized some of the limitations in the ITA methods currently used as a standard tool by ICES IEA groups. It was recommended to approach the evaluation problem through simulation studies, in a way similar to that used earlier in ICES for stock assessment models (ICES, 1993). The second workshop (WKINTRA2) developed and compared numerical simulation protocols and algorithms, with the aim of simulating few contrasted ecosystem datasets. These will form the basis of ITA methods evaluation for the intended WKINTRA-3 workshop.

Resource requirements	No major resourcing
Participants	Statisticians and researchers from across the IEASG network.
Secretariat facilities	None.
Financial	No financial implications for ICES.
Linkages to advisory committees	Link to ACOM through the development of ecosystem overviews
Linkages to other committees or groups	Links across all ICES IEA working groups
Linkages to other organizations	Links to IEA groups in the Arctic and PICES Working Groups working on similar topics.

Annex 3: Agenda

We met every day from 9:00 to 13:00 with breaks at about 10:15-10:30 and 11:45-12:00.

21st September

- Opening of the meeting
- Round table presentation
- Summary of WKINTRA-1&2, objectives of WKINTRA-3 (Benjamin & Saskia)
- Questions/Discussion
- Review of the datasets, observations and simulations (focus on new ones since last meeting, ToR a)
- ToR b: ITA methods evaluations:
    - Heatmaps and PCA (Benjamin)

22nd September

- ToR b: ITA methods evaluations:
    - Integrated Resilience Analysis (Saskia)
    - RDA/DFA (Jed)
    - MAR-X (Hiroko)
    - MAFA (Sigrid)
    - Other (MARCOS?)

23rd September

- Criteria used to perform an objective evaluation of the different methods (discussion)
- Group discussions on ITA methods

24th September

- Development of guidelines for IEA groups (discussion, ToR c)
- Pending issues, writing assignments.