ICES Journal of
Marine Science

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

ICES Journal of Marine Science (2017), doi:10.1093/icesjms/fsx223

Principal component analyses for integrated ecosystem assessments may primarily reflect methodological artefacts

Benjamin Planque* and Per Arneberg

Institute of Marine Research, Ecosystem processes research group, PO Box 6404, 9294 Tromsø, Norway

**Corresponding author: tel: +47 48 89 30 43; e-mail: benjamin.planque@imr.no.*

Planque, B. and Arneberg, P. Principal component analyses for integrated ecosystem assessments may primarily reflect methodological artefacts – ICES Journal of Marine Science, doi:10.1093/icesjms/fsx223.

Received 5 September 2017; revised 15 November 2017; accepted 17 November 2017.

Multivariate analyses constitute an integral part of today's marine integrated ecosystem assessments (IEAs). Principal component analysis (PCA) is one of the most common of these techniques, and the method has been used repeatedly to summarize the dynamics of marine ecosystems. There seems to be little recognition of the potential pitfalls associated with performing PCA on time-series that are autocorrelated and/or non-stationary. We investigate how the descriptive performance of PCAs may be affected by the structure of the underlying time-series and question whether such analyses can provide useful summaries of ecosystem trajectories. For this purpose, we reanalyse four datasets from the Barents, Norwegian, Baltic, and North Seas. We compare the results with those obtained from simulated datasets that share similar trend and autocorrelation properties, but in which the variables are unrelated. We show that most of the patterns revealed by the PCA can emerge from random time-series and that the fraction of the variance that cannot be accounted for by random processes is minimal. The Norwegian Sea dataset is a pathological case in which the variance explained by the first two components only exceeds what would be expected from randomly simulated time-series by 2%. We conclude that outputs from explorative multivariate analyses provide very little insight into ecosystem status, trajectories and functioning. IEA groups need to be equipped with methods that can provide better insight into how marine ecosystems function, the drivers of their changes and their possible future trajectories.

Keywords: autocorrelation, data mining, marine ecosystem assessment, principal component analysis, spurious correlations, time-series.

Introduction

Describing the status and temporal fluctuations of marine ecosystems is complex. Physical, chemical, biological, and human components may need to be considered simultaneously and, for each of these components, a variety of entities may be observed and interpreted in different ways. For example, sea temperature measurements may be used to determine ocean heat content, thermal stratification, the geography of thermal frontal structures, or anomalies against reference time periods. Similarly, a biological sampling protocol may be used to inform on variations in biodiversity, fish stock size, and demographic structure or biogeographic patterns. The challenge is then to combine the various types of information derived from the collected data into a coherent picture of marine ecosystem status and dynamics.

There is a long tradition of using multivariate numerical analyses to summarize complex data into few entities that capture the main features of interest in marine ecological systems. One of the most commonly used techniques in multivariate analysis is the principal component analysis (PCA) (Vanhatalo and Kulahci, 2016). The central idea of PCA is "to reduce dimensionality of a data set consisting of a large number of interrelated variables" (Jolliffe, 2002, p. 1).

In 1978, Colebrook used PCA to analyse multiple series from the Continuous Plankton Recorder survey, and stated that

> "*the first principal component can be regarded as the best possible single representation of the annual fluctuations in abundance for all the entities included in the analysis.*"

In their seminal paper on marine ecosystem regime shifts, Hare and Mantua (2000) "*assembled 100 physical and biological time-series*" on which they performed a PCA where the goal was "*to concentrate most of the variance of this large dataset into a small* *number of physically interpretable patterns of variability*". Kenny *et al.* (2009) used PCA in the North Sea to

> "*rank objectively the variables that best explain the overall variation in the data [...] and to enable trends in the sample component scores to be objectively compared with trends in selected pressure and state variables*".

More recently, Eriksen *et al.* (2017) used PCA in the Barents Sea

> "*to integrate information across multiple time-series comprising different components of the pelagic community and environmental factors*".

Although there are many examples of methods other than PCA being used for the analysis of multivariate time-series from marine ecosystems, these four examples illustrate the pervasive use of PCA over the last four decades as a tool to summarize complex multivariate temporal datasets. PCA is also part of the suite of multivariate methods used by the International Council for the Exploration of the Sea (ICES) to perform integrated ecosystem assessment (IEA) and is being taught in ICES training courses for that purpose (ICES, 2011).

The assumptions underlying the use of PCA often remain unstated, and when they are, they can be debated. For example, the assumption of multinormality is often disregarded, in particular when there is no statistical inference associated with the PCA analysis, although Comon (1994) recommends an alternative method (independent component analysis) specifically designed to cope with departure from this assumption. Multivariate statistical textbooks rarely mention that independence between observations is an important assumption and that non-independence may cause problems when performing or interpreting PCA. In fact, it is rather the reverse. Jollife (2002, p. 299) states that

> "*when the main objective of PCA is descriptive, not inferential, complications such as non-independence* [in time] *does not seriously affect this objective*".

The problem of serial correlation is, however, recognized in the statistical process control literature. For example, Huang (2010) notes that

> "*for autocorrelated processes* [...] *PCA fails to take into account the autocorrelation information. Thus, it is doubtful that PCA is the best choice.*".

More recently, Vanhatalo and Kulahci (2016) use time-series simulations to show that "*the descriptive ability of PCA may be seriously affected by autocorrelation*". In time-series analyses, the order of observations is important because past states can influence future states, but not the reverse. Serial autocorrelation is important because it reduces the true number of independent observations and may affect the interpretation of PCA patterns (Vanhatalo and Kulahci, 2016). Stationarity (the absence of trend) is another important assumption, and data should be stationary before performing a PCA (Brillinger, 1975).

In marine ecological research, there seem to be little recognition of the potential risks associated with performing multivariate analyses, such as PCA, on autocorrelated and/or non-stationary time-series. In a recent book on the multivariate analysis of ecological data (Greenacre and Primicerio, 2013), PCA is applied to ecological time-series without explicit consideration of the temporal aspect of the data or of issues such as stationarity or autocorrelation. The IEA groups at ICES (ICES, 2015a, 2015b, 2016a, 2016b) also give little or no consideration to the fact that time-series have specific properties that should be carefully considered before multivariate analyses are performed and when interpreting their results.

In summary, PCA has been and is still widely applied to analyse multivariate ecological time-series, but with little or no consideration of the underlying properties of these temporal datasets, and with the (often implicit) assumption that features such as temporal trends or autocorrelation have minor impact on the results and their interpretation.

In this contribution, we investigate how the descriptive performance of the PCAs used for ecosystem integrated assessment may be affected by the structure of the underlying time-series. We examine whether the results from the PCA analyses emerge from the underlying structure that exists between time-series and can thus can provide useful summaries of ecosystem trajectories. An alternative option is that PCA results are predominantly artefacts of the method, when applied to time-series that display trend and autocorrelation. To evaluate the descriptive performance of the PCA, we re-analyse four well-documented case studies: the North Sea, the Baltic Sea, the Norwegian Sea, and the Barents Sea. We show how autocorrelation and trends in independent time-series can generate the patterns seen in PCA results and discuss how informative these patterns are to describe the dynamics of the ecosystems. We conclude with a call for alternative multivariate approaches to support marine ecosystem integrated assessments.

Material and methods

Empirical data

We base our analysis on the data collated and presented by four ICES integrated assessment working groups: the working group on the integrated assessment of the Barents Sea (WGIBAR, ICES, 2016b), the working group on the integrated assessment of the Norwegian Sea (WGINOR, ICES, 2015a), the working group on the integrated assessment of the North Sea (WGINOSE, ICES, 2016a) and the working group on the integrated assessment of the Baltic Sea (WGIAB, ICES, 2015b). The data consist of multivariate annual time-series and have been prepared by the groups to perform multiple multivariate analyses, including PCA. The datasets are presented in the individual working group reports and a summary of the suite of methods used by each working group can be found in ICES (2016c). For the Norwegian and Barents Sea, the working groups have performed analyses integrated over the entire eco-regions, while in the North and Baltic Seas, the multivariate analyses were conducted in sub-regions. Sub-regions Skagerrak in the North Sea and Central Baltic Sea were selected for the analysis in the current study. The datasets often contain a mixture of abiotic and biotic data that describe changes in the physical, chemical, and biological environment and in fisheries in each region. In addition to these four marine ecosystem datasets, we have used a multivariate dataset of 15 time-series, available from Statistics Norway (https://www.ssb.no/en/), the Office for National Statistics in the United Kingdom (https://www.ons.gov.uk/) and the online publication Our World in Data (https://ourworldindata.org/) (Supplementary Material). The variables in these time-series are *a priori* unrelated (e.g. the ratio of potato-to-cabbage price in the United Kingdom and number of hurricanes making landfall in the United States), and the dataset is used as a control against which the marine ecosystem datasets can be compared.

Simulated time-series

To test the performance of PCA, we constructed simulated time-series that reflect four different null hypotheses. First, we constructed series of similar length, mean and variance to those of the empirical data. In this first simulated dataset, there is no account for the temporal order of the data. Second, we constructed series of similar length, mean, variance, and autocorrelation to those of the empirical data. Third, we constructed series of similar length, and trend to those of the empirical data, and similar mean and variance of the residuals. Finally, we constructed series of similar length, trend and autocorrelation, and similar mean and variance of the residuals to those of the empirical data. To reconstruct series with similar autocorrelation, we use a technique known as phase randomization (Theiler *et al.*, 1992; Schreiber and Schmitz, 2000). For each region and for each null hypothesis, we simulated 999 multivariate datasets.

The most important feature of these simulated datasets is that the individual time-series are generated independently of each other, i.e. there is no underlying structure that relates the time-series to each other and therefore a summary of the overall dynamics of the system is, by definition, meaningless. The features that can arise from the analysis of these datasets result from the structure of the individual time-series only and not from the relationships between them. If such relationships are revealed by the analysis, they have emerged by chance and cannot be used to deduce mechanisms relating various components of the ecosystem. In the same way, the control dataset is made of unrelated variables and any features that are revealed by the PCA must have emerged by chance.

Data preparation and multivariate analyses

To analyse the empirical and simulated data, we apply a similar approach to the one used in the ICES Integrated assessment working groups when performing PCA. For the Norwegian Sea, the untransformed time-series are used as input to the PCA that is performed on the correlation matrix. For the Barents Sea, all biological time-series are log-transformed. Every time-series is then scaled to zero mean and unit variance before performing the PCA, a procedure that is equivalent to running the PCA on the correlation matrix, as in the Norwegian Sea. For the North Sea, biological time-series of catch per unit effort are log-transformed and then scaled to zero mean and unit variance. For the Baltic Sea, plankton time-series are log-transformed while fish and abiotic time-series are not, and the PCA is performed on the correlation matrix. In several instances, the number of variables exceeded the number of years of observations (e.g. Central Baltic Sea: 37 years and 61 variables, Norwegian Sea: 21 years and 24 variables, Barents Sea: 29 years and 54 variables). In such cases, the maximum number of components is set by the number of years of observation in the dataset. A total of 19,985 PCA were performed ((4 regions + 1 control) × 4 hypotheses × 999 simulations + 5 empirical datasets). Data preparation, generation of simulated time-series, PCAs and graphical outputs were performed in R (R Core Team, 2017). The R-code for the simulation and analysis of the multivariate datasets is provided in the Supplementary Material.

Interpretation of the results

The comparison of the output of the PCA on empirical time-series with the outputs of the PCAs on simulated time-series is analogous to conventional null hypothesis statistical testing. If the ecosystem trajectory and amount of variance explained by the PCA on empirical time-series cannot be differentiated from those obtained with the null model, then the null model cannot be rejected. In such a situation, the empirical results cannot be used to interpret ecosystem structure and trajectory beyond the simple assumptions of the null models, which poses that time-series are independent from each other. On the other hand, if the outputs of the PCA on empirical time-series significantly differ from those of the PCAs on simulated time-series, these results can possibly provide insights on underlying linkages between the various ecosystem components and the summary expressed by the first few components can be meaningful. If the simulation scheme is valid as a null hypothesis, it should also be expected that the results of the PCA on the control dataset should not differ from those on simulated data.

Results

Figure 1 illustrates how the four methods for simulating time-series capture the features of the original data series. When the original data present little trend and autocorrelation (Volume transport, Figure 1, top) all simulation methods generate similar patterns. When temporal fluctuations are dominated by autocorrelation (Proportion mature capelin, Figure 1, middle), random time-series that do not account for autocorrelation do not satisfactorily reproduce the patterns observed in the original data, while simulated series that account for autocorrelation do. Finally, when both trend and autocorrelation are present in the original data (Haddock biomass, Figure 1, bottom), only the full model, which accounts for trend and autocorrelation, can adequately reproduce the original patterns. The full model can be used to simulate independent time-series while preserving most of their original temporal properties, whether these include trend, autocorrelation, or both, and can therefore be considered as the best reference null model, against which the results of empirical analyses should be compared.

PCAs performed on the simulated datasets can sometimes reveal patterns that are similar to those obtained when analysing original data. An example for the Norwegian Sea is illustrated in Figure 2 (illustrations for the Barents Sea, North Sea, Baltic Sea and the control dataset are provided in the Supplementary Material). The analysis of the original data shows a wide dispersion of the variables in the space of the first two components (arrows in Figure 2a), which is also visible for all types of simulated datasets (Figure 2b–e). The temporal trajectory of the original data forms a U-shape which can be interpreted as a temporal trend along the first axis and a swing along the second axis (from high to low and back to high-component values). When only mean and variance are retained in the simulated data, the trajectory of the system seems erratic (Figure 2b). When individual trends are also retained, the trajectory of the system follows a trend along the first axis while displaying random fluctuation along the second axis (Figure 2c). When autocorrelation is kept in the simulated datasets the trajectory is more circular, with out-of-phase long-term swings on the first and second axes (Figure 2d). Finally, when trends and autocorrelation in the original time-series are retained, the temporal trajectory of the system is

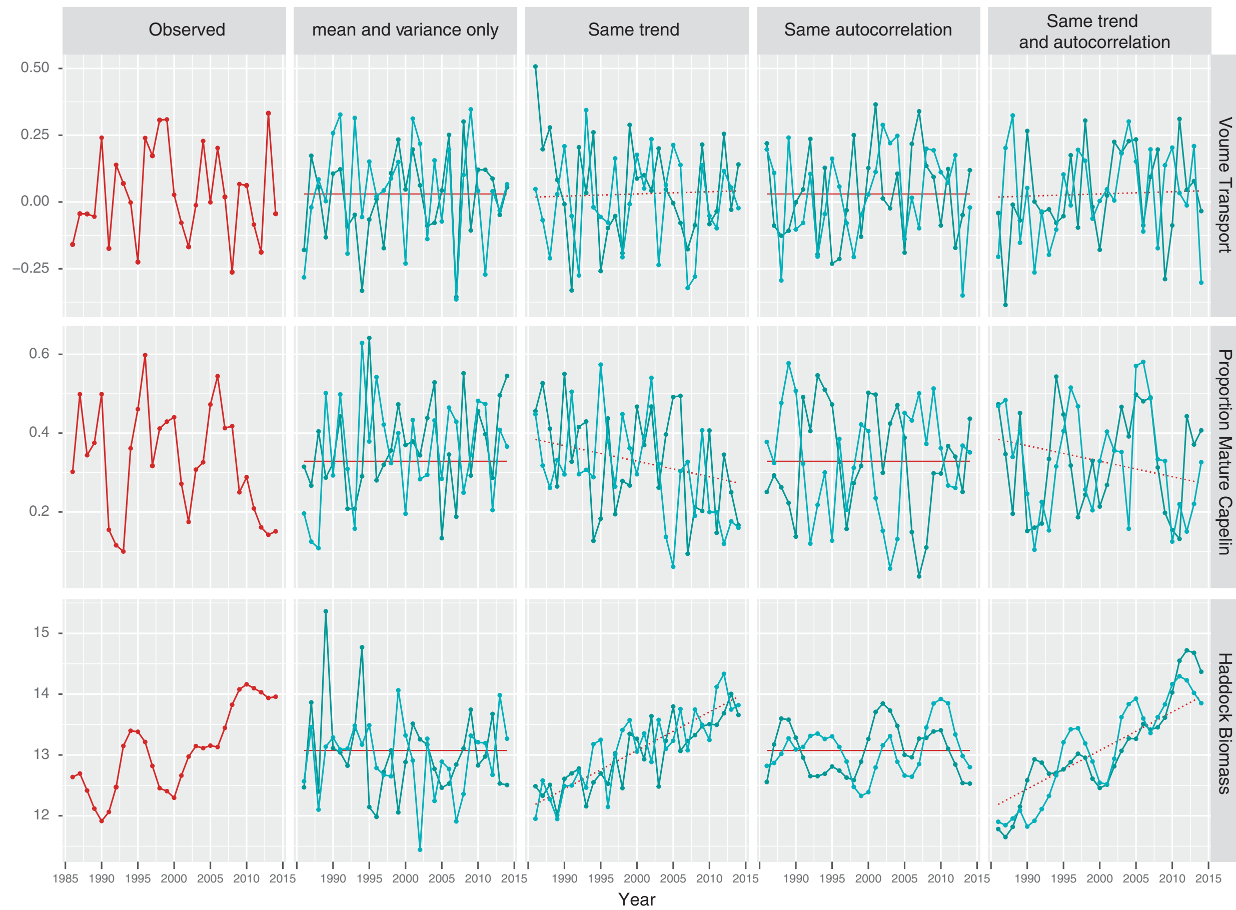

Figure 1. Interannual variations in three selected indices in the Barents Sea: (a) the net southward volume transport between Svalbard and Franz Josef Land (top), (b) the proportion of mature capelin at age 2 years (middle), and (c) the total stock biomass of Haddock (*Melanogrammus aeglefinus*) age 3 years and above (bottom). The observed time-series display the empirical data provided to the expert ICES working group. For each time-series simulation scheme, two realizations are shown. The horizontal lines indicate the mean of the original and simulated time-series. The dotted lines indicate the linear trend in the original and simulated time-series (when used).

U-shaped and resembles the PCA results obtained on the original data (Figure 2e).

A key element of the PCA interpretation lies in the amount of variance that can be accounted for by the first components. Often (as in Figure 2), only the first two components are explicitly considered and their interpretation can be considered informative only if they account for a substantial fraction of the total variance. When interpreting patterns of variance, a common and implicit assumption is that if all variables in the original datasets were independent (i.e. orthogonal), each Principal Component should explain the minimal possible fraction of the total variance (equal to 1 divided by the number of components). In the case of the Skagerrak sub-region, where 28 components are computed, this minimum variance explained by the first two components would amount to $2/28 = 7.1\%$. Outputs from the PCA on the Skagerrak sub-region show that the first two components account for 42.7% of the total variance (respectively 23.5 and 19.2%) which is substantially greater than the minimum expectation. However, PCA performed on simulated time-series reveal that the expected amount of variance when all variables are independent can be much greater than the theoretical 7.1%. When simulated data only keep the mean and variance from the original data, the first two PCs account for 22.2% (Figure 3a) of the total variance. This proportion raises to 26.4% when autocorrelation is considered (Figure 3b), to 33.2% when trends are considered (Figure 3c) and finally to 36.5% when trends and autocorrelation are jointly considered (Figure 3d). The difference between the percentage of variance explained by the first two PCs on the empirical observations and the percentage from a PCA on random independent time-series is only 6.2%. Similar results are found for the Barents Sea, Norwegian Sea and Baltic Sea (Figure 4 and Supplementary Material). For the control dataset, the first two Principal Components account for 57.2% of the variance, which is similar to the variance explained from simulated data when trends and autocorrelation are jointly considered (median: 57.6%, 95% *CI*: 50.6–64.3%). This shows that the simulation method that incorporates trend and autocorrelation mimics unrelated data series well and can serve as a null model against which the results of empirical PCA analyses can be compared.

The percentage of variance explained by the first two components is generally <50%, with the exception of the Norwegian Sea (60.8%), so the majority of the variance remains unaccounted for by the first two PCs. In addition, the amount of variance in excess of the expected value from the full null models (with mean, variance, autocorrelation, and trend) is small: 6.0% (Barents Sea), 2.0% (Norwegian Sea), 6.2% (North Sea—Skagerrak), and 4.5% (Central Baltic Sea). Despite the results from null models and empirical data being significantly different ($p < 0.001$ in most cases), the small

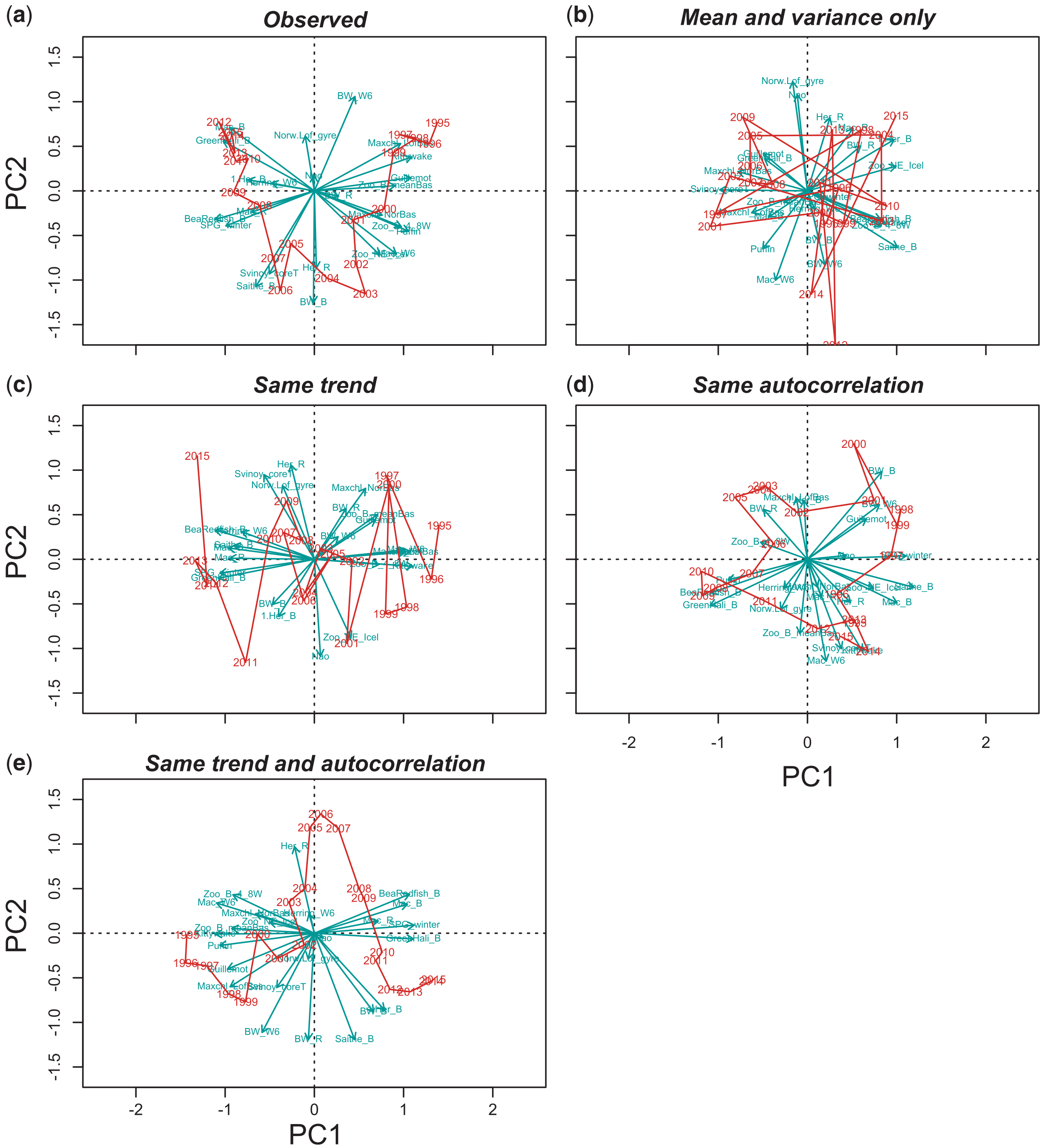

Figure 2. Biplots of the PCA performed for the Norwegian Sea. The arrows show the contribution of each variable in the first 2 components (length of the arrow) as well as the correlation between variables and components (angle between arrows or between arrows and components). The continuous lines indicate the temporal "trajectory" of the system projected on the first two axes of the PCA. The PCA results are shown for (a) the original data series and one realization of simulated data with (b) only mean and variance preserved, (c) trend preserved, (d) autocorrelation preserved, and (e) trend and autocorrelation preserved.

amount of additional variance explained in the original analysis compared with what is expected from null models raises questions about the value of the interpretation of such analyses, beyond giving a description of individual time-series considered to be independent.

Discussion

Our results show that PCA results can be driven, to a large extent, by the temporal structure of the individual time-series that underlie the analysis. Most of the variance explained by PCAs and the patterns revealed in the first two principal components can emerge from randomly generated time-series, when these share the properties of trend and autocorrelation found in the original datasets. The results of PCA conducted on multiple and unrelated time-series can be misleading. High percentages of variance explained by the first components and apparent coherent temporal patterns can emerge even in the absence of relationships

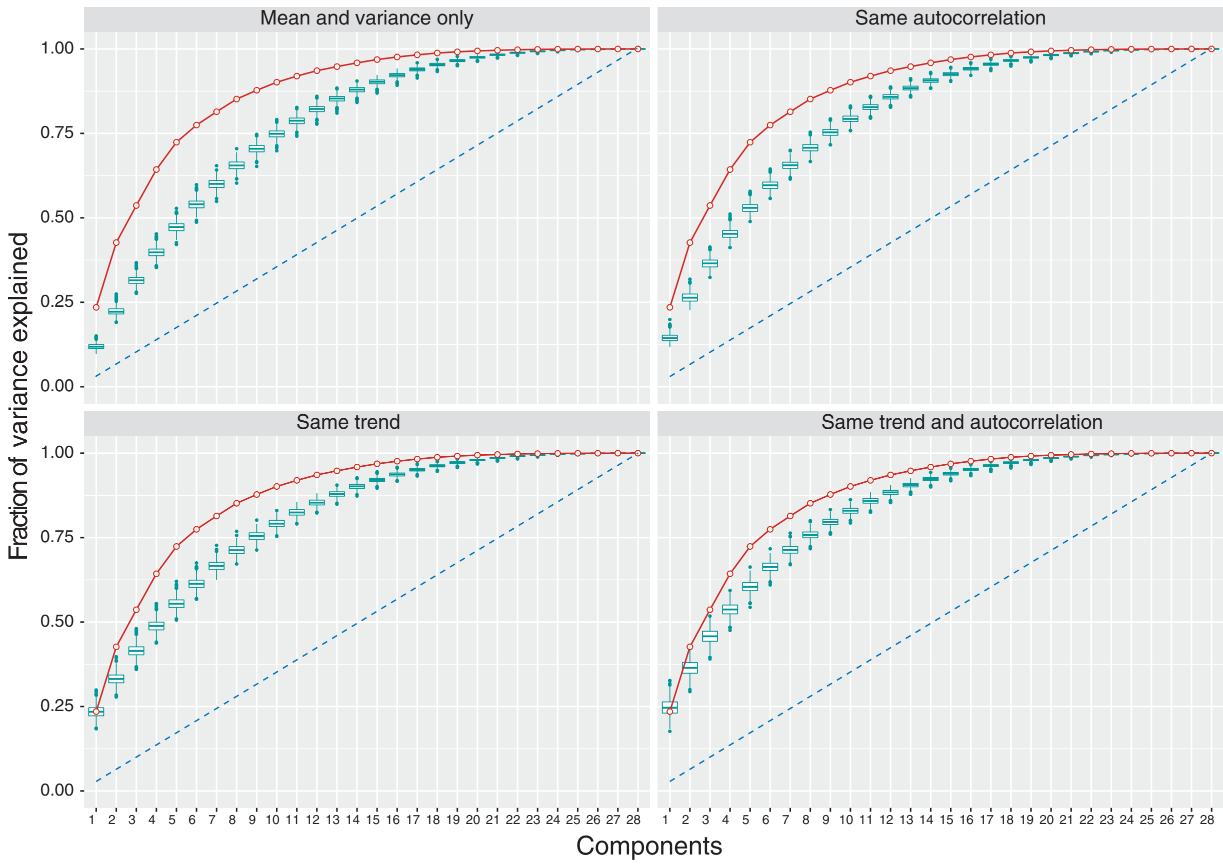

Figure 3. Cumulated proportion of variance explained by the principal components in the North Sea Skagerrak. Plain lines and open dots show the results of the PCA performed on original data. Boxplots show the distribution of cumulated variance proportions from the 999 simulations with only mean and variance preserved (top left), autocorrelation preserved (top right), trend preserved (bottom left) and trend and autocorrelation preserved (bottom right). Dotted lines show the theoretical minimum cumulated proportion of variance explained by the principal components.

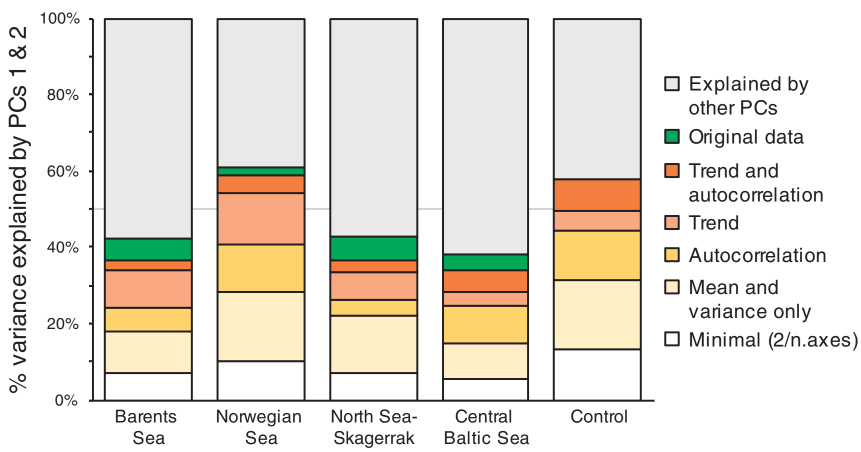

Figure 4. Fraction of the variance explained by the first two axes of the PCAs. From bottom to top: a) minimum possible amount of variance that can be explained by the first two PCS; median variance explained in PCAs performed on simulated data that incorporate b) mean and variance only, c) autocorrelation, 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 data series with similar trend and autocorrelation.

between the original series. In the control case, consisting of unrelated time-series, more than half of the variance is explained in the first two components, the trajectory of the system is a clear swing and stabilization over the last three decades, and the analysis reveals strong associations between time-series. For example, trust in United States government is negatively correlated with divorce rate in Norway, number of UN peacekeeping operations is negatively related with the per capita calorie supply in Afghanistan, number of genocides worldwide is strongly correlated with fatality rates due to lightning in the United States and the ratio of nuclear warheads in United States and Russia follows the ratio of sugar-to-fish price in the United Kingdom (Supplementary Material).

Strong correlations between time-series can possibly reflect true underlying causal relationships, confounding effects (two time-series are related to a third unobserved variable) or the presence of trend and autocorrelation that inflate correlation values. The common assumption, explicit or implicit, that temporal structures in ecological data series have little influence on the interpretation of the results from PCAs, should be revised. It is necessary here to remember that the absence of trend and autocorrelation are underlying assumptions of the method. When these are not adequately addressed, it is safer to assume by default that the patterns of variation revealed by PCAs can be spurious (i.e. they emerge by chance) and subsequently test if this assertion can be rejected.

For all cases reanalysed here (except the control case), the proportion of variance explained by the PCA on original data is always greater than that obtained from simulated data. Statistical inference shows that the probability of observing such amount of variance by chance is often extremely low ($p < 0.001$). However, despite the inference being "statistically significantly", the difference between the variance explained in the observed and simulated data is of very small magnitude. The Norwegian Sea dataset is a pathological case in this respect, where the first two components account for >60% of the variance but the difference between the observed and simulated datasets amounts to only 2%. In such a situation, one must be careful not to over-interpret statistical inference, as happens too often in ecological studies (Yoccoz, 1991).

In this study, we focussed specifically on the PCA because the method has been used for a long period of time and the analyses are performed in a standard manner by IEA groups. Other multivariate methods may suffer similar types of problems, if temporal properties such as trend and autocorrelation are not considered explicitly. For example, the graphical outputs of the analysis known as "heatplots" can easily be misinterpreted in a way similar to the mis-interpretation of PCA results (Supplementary Material).

There is a need to perform robust evaluation of existing methods, even when these have been used for long periods of time. Because marine ecological time-series are often short (<50 years) and display trends and autocorrelation, it is likely that most descriptive multivariate analyses will have very low statistical power and will reveal patterns that have possibly emerged by chance. Unfortunately, the persistence of methods and practices in literature does not guarantee their performance, as a research community can continually misuse methods that are easy to implement even when the results have been shown to be erroneous (Smaldino and McElreath, 2016).

Results of the PCA analyses conducted here are based on the correlation matrix between variables. There is a plethora of examples that support the assertion that "correlation does not entail causation" (many of which are nicely summarized in Vigen, 2015) and PCA results may therefore provide little insight into causal mechanisms. Other multivariate data-mining approaches which argue for "letting the data speak" often rely on identifying association between data sets in a similar way to that of PCA, albeit often relaxing some of the underlying assumptions such as linear relationships or normality. These are known to suffer from similar problems (Rexstad *et al.*, 1988) and unless the performance of such methods to analyse multivariate ecological time-series is properly evaluated, they may provide little insight into ecosystem status, trajectories, and functioning.

Methods that are dedicated to the study of multivariate time-series such as dynamic factor analysis (Molenaar *et al.*, 1992) should be preferred to those which ignore temporal processes. More importantly, investigation of the dynamics of complex ecosystems should be conducted with tools that can provide plausible models of causality behind the observed dynamics, whether these are linear or not (see e.g. Ives *et al.*, 2003; Pearl, 2010; Sugihara *et al.*, 2012; Shipley, 2016). Not only can such methods enlighten IEA groups by revealing the essential pathways behind ecosystem dynamics, but they can provide a way to evaluate forecasting possibilities at time horizons that are relevant for marine management (typically 1–5 years).

There is a need for IEA groups to critically evaluate the numerical methods at their disposal and, when necessary, to change practice and move away from low-power explorative analyses towards methods that can provide better insight into how marine ecosystems function, the drivers of their changes and their possible future trajectories.

Supplementary data

Supplementary material is available at the *ICESJMS* online version of the article.

Acknowledgements

We are grateful to members of ICES working groups who helped accessing the data used in this work and understanding the details of the original analyses performed by the groups, in particular Erik Olsen, Elena Eriksen and Saskia Otto. We thank Hiroko Solvang, Sandra Hamel and Nigel Yoccoz for stimulating discussions and for bringing methodological insights to our attention and Kirsteen MacKenzie for her precious proof reading and editing of the original article.

References

Brillinger, D. R. 1975. Time series: data analysis and theory. Holt, Rinehart and Winston, New York. 512 pp.

Colebrook, J. M. 1978. Continuous Plankton Records: zooplankton and environment, North-East Atlantic and North Sea, 1948-1975. Oceanologica Acta, 1: 9–23.

Comon, P. 1994. Independent component analysis, a new concept? Signal Processing, 36: 287–314.

Eriksen, E., Rune Skjoldal, H., Gjøsæter, H., and Primicerio, R. 2017. Spatial and temporal changes in the Barents Sea pelagic compartment during the recent warming. Progress in Oceanography, 151: 206–226.

Greenacre, M., and Primicerio, R. 2013. Multivariate Analysis of Ecological Data. Fundación BBVA, Bilbao, Spain.

Hare, S. R., and Mantua, N. 2000. Empirical evidence for North Pacific regime shifts in 1977 and 1989. Progress in Oceanography, 47: 103–145.

Huang, X. 2010. Topics in multivariate time series analysis: statistical control, dimension reduction visualization and their business applications. *In* Management Science. University of Massachusettss, Amherst.

ICES. 2011. Report of the Training Course: Approaches to the Integrated Assessment of Status and Trends in Marine Ecosystems (TCIEA), 17-19 October 2011, ICES Headquarters, Copenhagen. 9 pp.

ICES. 2015a. Final report of the working group on the integrated assessment of the Norwegian Sea (WGINOR). ICES CM, 2015/SSGIEA: 10: 149 p.

ICES. 2015b. Report of the ICES/HELCOM working group on integrated assessments of the Baltic Sea (WGIAB). ICES CM, 2015/SSGIEA: 08: 34.

ICES. 2016a. Final report of the working group on integrated assessments of the North Sea (WGINOSE) ICES CM, 2016/SSGIEA: 06: 52.

ICES. 2016b. Final report of the working group on the integrated assessment of the Barents Sea (WGIBAR). ICES CM, 2016/SSGIEA: 04: 123 p.

ICES. 2016c. Report of the workshop on integrated ecosystem assessment methods (WKIDEA). ICES CM, 2016/SSGIEA: 15 pp.

Ives, A. R., Dennis, B., Cottingham, K. L., and Carpenter, S. R. 2003. Estimating community stability and ecological interactions from time-series data. Ecological Monographs, 73: 301–330.

Jolliffe, I. T. 2002. Principal Component Analysis, 2nd edn. Springer-Verlag: New York, NY.

Kenny, A. J., Skjoldal, H. R., Engelhard, G. H., Kershaw, P. J., and Reid, J. B. 2009. An integrated approach for assessing the relative significance of human pressures and environmental forcing on the status of Large Marine Ecosystems. Progress in Oceanography, 81: 132–148.

Molenaar, P. C. M., De Gooijer, J. G., and Schmitz, B. 1992. Dynamic factor analysis of nonstationary multivariate time series. Psychometrika, 57: 333–349.

Pearl, J. 2010. An introduction to causal inference. International Journal of Biostatistics, 6: article 7.

R Core Team 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.r-project.org/.

Rexstad, E. A., Miller, D. D., Flather, C. H., Anderson, E. M., Hupp, J. W., and Anderson, D. R. 1988. Questionable multivariate statistical inference in wildlife habitat and community studies. The Journal of Wildlife Management, 52: 794–798.

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

Shipley, B. 2016. Cause and correlation in biology. A User's Guide to Path Analysis, Structural Equations and Causal Inference in R, 2nd edn, Cambridge University Press, Cambridge, UK.

Smaldino, P. E., and McElreath, R. 2016. The natural selection of bad science. Royal Society Open Science, 3: 160384.

Sugihara, G., May, R., Ye, H., Hsieh, C-h., Deyle, E., Fogarty, M., and Munch, S. 2012. Detecting causality in complex ecosystems. Science, 338: 496–500.

Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., and Doyne Farmer, J. 1992. Testing for nonlinearity in time series: the method of surrogate data. Physica D, 58: 77–94.

Vanhatalo, E., and Kulahci, M. 2016. Impact of autocorrelation on Principal Components and their use in statistical process control. Quality and Reliability Engineering International, 32: 1483–1500.

Vigen, T. 2015. Spurious Correlations. Hachette books, New York, Boston.

Yoccoz, N. G. 1991. Use, overuse, and misuse of significance tests in evolutionary biology and ecology. Bulletin of the Ecological Society of America, 72: 106–111.

Handling editor: James Watson