Methods for assessing the effects of environmental parameters on biological communities in long-term ecological studies - A literature review

Fabien Verniest, Sabine Greulich

Methods for assessing the effects of environmental parameters on biological communities in long-term ecological studies - a literature review

Fabien Verniest[a], Sabine Greulich[a,*]

[a] Interdisciplinary Research Center Cities, Territories, Environment and Society (UMR CNRS 7324 CITERES), University of Tours, 33 allée Ferdinand de Lesseps, 37000 Tours, France

* Corresponding author at: Interdisciplinary Research Center Cities, Territories, Environment and Society (UMR CNRS 7324 CITERES), University of Tours, 33 allée Ferdinand de Lesseps, 37000 Tours, France. E-mail address: sabine.greulich@univ-tours.fr (Sabine Greulich)

Declarations of interest: none

Abstract:

Many ecological processes that play important roles in ecosystems occur over long time periods and can therefore not always be properly studied with short-term studies. However, researchers have to face many challenges while setting up long-term ecological studies, including the choice of relevant data analysis methods and the design of the study (i.e. sampling frequency, number of sites, etc.). This literature review, based on 99 original studies, provides an overview of methodological choices used to analyse the effects of abiotic parameters on biological communities on a long-term scale. To this end, the main characteristics of study design were recorded (e.g. sampling frequency, duration, taxa, variables) and the different data analysis tools summarised and analysed. We found that long-term ecological studies focusing on the effects of environmental factors on biotic parameters mostly concerned aquatic habitats. Studies substantially varied in their design, although many of them had similar aims. Univariate methods, almost entirely performed by means of linear modelling and correlation tests, were used more often than multivariate methods. Finally, constrained and unconstrained ordination methods were used equally, and other data analysis tools were rare. Finally, we created a decision key to help researchers choose the appropriate analysis tools for their specific long-term study.

Key-words (6 max): LTER; data analysis; sampling; multivariate data; modelling; statistics

1. Introduction

The effects of environmental parameters on biological communities can occur at different time scales, from less than a day (e.g. effect of tides) to several millennia (e.g. effect of continental glaciation) and at different spatial scales, from less than a plot (e.g. effect of microtopography) to the entire Earth (e.g. effect of climate change) (Franklin et al., 1990; Magnuson, 1990). Although various phenomena, qualified as long-term, play a key role in ecosystems (Franklin et al., 1990), there is often a difference between the funding duration of ecological studies and the relevant time scale to study these phenomena (Callahan, 1984). This discrepancy, combined with the increasing number of potential drivers of ecosystem change that occur simultaneously (e.g. rising temperature, pollution, habitat destruction), has led to an increasing demand for data from long-term ecological studies and to the development of long-term ecological programs, such as the Long-Term Ecological Research (LTER) network (Callahan, 1984).

There is no consensus about the definition of “long-term” in ecological studies: it can be based either on ecological criteria (e.g. generation time of the studied organism, time scale of ecological processes) or on operational constraints (e.g. funding cycles, human life span) (Knapp et al., 2012; Strayer et al., 1986). Because both strategies have advantages and drawbacks (Lindenmayer et al., 2012; Strayer et al., 1986), a minimum duration threshold, 10 years for example, can be chosen to compromise between the two approaches (Lindenmayer et al., 2012; Wolfe et al., 1987). Several types of short-term study (e.g. retrospective studies, modelling, substitution of space for time, and use of systems of fast dynamics) offer distinct advantages over long-term studies and allow the analysis of long-term ecological phenomena. For that reason, researchers should also consider these less funding-dependent options (Strayer et al., 1986), although these alternatives are not always achievable.

Knowledge of ecosystem behaviour over long time scales is indispensable to gain a deeper understanding of the processes that drive ecosystems and to disentangle anthropogenic and natural changes, as well as short-term fluctuations and long-term trends (Haase et al., 2016). Long-term studies are appropriate to investigate processes that can be classified in four categories: slow processes (e.g. forest succession, vertebrate population cycles), rare events (e.g. fire, flood, disease), subtle processes (i.e. when the magnitude of the long-term trend is small compared to the year-to-year variance), and complex phenomena involving a combination of multiple abiotic parameters that cannot be studied statistically with few observations (Strayer et al., 1986). Long-term ecological studies have thus contributed, and continue to contribute, to many findings in ecological sciences (Franklin et al., 1990; Magurran et al., 2010); this approach makes it possible to quantify ecological responses to environmental change and to understand complex ecosystem phenomena occurring over a prolonged period, in addition to providing ecological data for model development, parameterization and validation (Lindenmayer et al., 2012; Wolfe et al., 1987). It also promotes multidisciplinary research, supports environmental policies and ecosystem management and plays an important role in societal issues (e.g. efficacy of fertilisers, soil acidification, impact of sewage pollution on lakes), education (e.g. students involved in these projects), and communication to the general public (Lindenmayer et al., 2012; Strayer et al., 1986). However, long-term ecological studies have serious disadvantages, the main one being the need of long-term funding, staff and facilities. Thus, these studies are limited to time scales ranging from a few decades to one or two centuries (Strayer et al., 1986). Researchers conducting this type of study also have to face other essential challenges, such as dealing with changing objectives and schemes that can lead to modifications in methodology (Magurran et al., 2010).

Implementing long-term studies on ecosystems and ecological processes requires a series of methodological choices covering, roughly, the following steps: (1) selection of study sites, (2) choice of taxa and abiotic variables to monitor, (3) selection of the appropriate spatial and temporal scales for the monitoring, and (4) selection of data analysis methods. While steps (2) and (3) refer to study design, step (4) can include or be followed by modelling of the monitored system. The methodological choices are largely dependent on the type and extent of the investigated ecosystem, the life cycle duration of the investigated taxa and the dimension of the area where the life cycle takes place (Fig. 1). The main focus of a long-term study that surveys biodiversity is often to identify the main drivers of community and ecosystem dynamics. Datasets resulting from those studies thus typically comprise different kinds of biota and different series of environmental parameters. Although these parameters are also referred to as 'abiotic parameters', they are actually not independent of life but have coevolved with it, so the term 'conbiotic parameters' would be more appropriate (Fath and Müller, 2019). Data analysis that aims to explain biota (i.e. response variables) by environmental parameters (explanatory variables) can be performed with a potentially wide variety of techniques that should be chosen according to the type, number and frequency distribution of data.

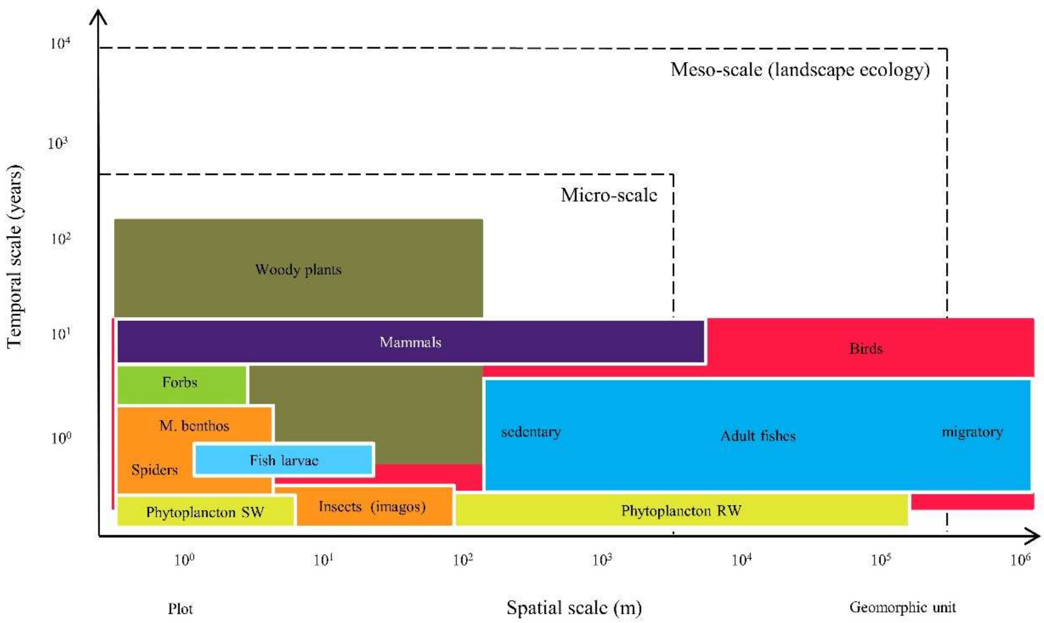

Fig. 1 (2-column): Differences in spatial and temporal life cycle scales in taxa coexisting in an ecosystem (here: a floodplain. Spatial scale: length of the river section, floodplain width: 500m). Scales modified from Delcourt and Delcourt (1988). SW: standing water; RW: running water.

Many authors have discussed the applications of data analysis methods in ecology, most of them focusing on specific methods or approaches (e.g. Bayesian methods, linear and additive modelling) in an ecological framework (e.g. Beninger et al., 2012; Boldina and Beninger, 2016; Dale and Fortin, 2002; Dorazio, 2016; Guisan et al., 2002; Hobbs and Hilborn, 2006; Mukhopadhyay and Banerjee, 2015). By contrast, only a few articles introduce a wide set of analytical tools, often with respect to a particular research area (e.g. Buttigieg and Ramette, 2014; Garamszegi et al., 2009; Paliy and Shankar, 2016; Parker and Arnold, 1999; Ramos et al., 2015). Literature reviews on data analysis techniques applied to ecology (e.g. Crowley, 1992; James and McCulloch, 1990; Jennions, 2003; Ramette, 2007) are even less common. Similarly, many articles and books on ecology focus on study design (e.g. Hurlbert, 1984; Morrison, 2010; Strayer et al., 1986; Yoccoz et al., 2001). However, information about what is actually applied in the field by researchers (i.e. sampling design, field methods, measured variables, etc.) is not easily available, although this could be valuable for researchers designing long-term studies. Examples are Jaeschke et al. (2014), who reviewed aspects including the areas, taxa, and parameters of studies analysing the impact of climate change on ecosystems, and Jackson and Fureder (2006) who summarised the duration and number of sites and sample years of 46 long-term studies of freshwater macroinvertebrates. Literature reviews covering both data analysis and study design in a long-term framework are even rarer, although it is critical (1) to consider the selection of data analysis methods and study design simultaneously, as the former is conditioned by the latter, and (2) to choose the methods carefully, considering the resources necessary to conduct long-term ecological studies and the particularities of their methodology.

In this paper, we aim to give an overview of the main characteristics of study design in existing long-term studies and the methods used to analyse the resulting data. The first part of this article summarises study features with regard to the following questions: (1) what are the main characteristics of these studies (i.e. aims, location and sampling strategy)? (2) Which taxa and habitats are investigated? (3) Which and how many environmental parameters and biological metrics are used? (4) How are data analysed? The second part of this article provides details on the data analysis techniques that were performed in the reviewed articles.

2. Materials and Methods

2.1. Search strategy

We examined the Web of Science (all databases) between May and June 2018, covering all publications available up to that point and exclusively belonging to the Web of Science "Biodiversity and Conservation" and "Environmental Sciences and Ecology" categories. To obtain the most relevant papers, we used various search keywords and operators (Table 1). Only original studies were considered. Because a very extensive number of articles matched these requirements (about 2,500), we restricted the number of potential articles by rejecting studies related to extraneous research areas (e.g. medicine, microbiology, chemistry or molecular biology) or exclusively analysing one very specific effect (e.g. impact of fire) and thus not investigating several environmental parameters simultaneously. The search was further focused by excluding a list of topics (Table A.1). This produced a total of 511 articles for which abstracts and, when necessary, the methods section were read.

Temporal scale	Biotic	Abiotic	Variable	Relationship
"Long term" [1]	Communit*	Abiotic	Factor$	Relation*
LTER	Assemblage$	Habitat	Parameter$	Impact$
ILTER	Assembly	Environmental	Disturbance	Effect$
	Guild$	Variability	Nutrient$	Response$
	"Multi-tax*"		Temperature	Influence$
	"Multi-species"		Driver$	
			Gradient$	

[1] Keyword only searched in title

Table 1: List of terms searched in title, abstract and keywords to select relevant research articles. The OR operator was used between each term of the same group (rows), whereas the AND operator was added between each group (columns). "*" allows more letters, and "$" allows only one more letter.

Among the 511 articles, we selected those that explicitly analysed (i.e. using data analysis methods) the effects of two or more abiotic factors (excluding the drivers shown in Table S1) on two or more biotic variables (or one explained variable based on several taxa) in a long-term framework. This choice was motivated by the specificity and infrequency of the other topics (i.e. biotic interactions and impacts of the biological compartment on environmental characteristics). However, the presence of biotic variables among abiotic predictors was not a reason for rejecting articles. Given the various definitions of "long-term", and because it was not our purpose to discuss them in this review, we considered as long-term the studies qualified as such by their authors and characterised by a long-term sampling design. We thus excluded, in line with Lindenmayer et al. (2012), retrospective investigations (*sensu* Likens, 1989) (e.g. studies based on tree rings), studies using simulated data and studies with extended gaps between sampling.

In total, the relevant sections (i.e. the ones dealing with effects of abiotic parameters on biocoenosis) of 99 articles were analysed in depth for this study.

2.2. Study characterisation

The following characteristics were recorded: (1) investigated ecosystem (i.e. taxa and habitat), (2) geographical location, (3) LTER involvement, (4) aim of the study, (5) explained variables (i.e. total number and type), (6) explanatory variables (i.e. total number and type), (7) sampling strategy (i.e. study duration, sampling frequency, number of sites, and type of study: observational or experimental), and (8) data analysis methods. Only the features associated with the section focusing on the effects of abiotic parameters on biocoenosis were analysed.

When recording the investigated ecosystems, only the biological communities used as explained variables were retained. We followed the taxonomic indications of the authors, whether the authors defined the organisms they studied with monophyletic group names or not (e.g. aquatic invertebrates, benthic communities, plankton). In the first case, we used taxonomic classification (Table 3, lower part), whereas in the second case we based our analyses on the terms given by the authors (Table 3, upper part).

To compute the number of explained variables, we considered the measurement of one taxon (e.g. abundance) as one explained variable, even if the community composition was analysed as a whole (e.g. with ordination methods) and not every taxon separately. The computation of the number of explanatory variables takes into account every single relevant explanatory variable, including variables belonging to the "direct anthropogenic impact" category that were considered as abiotic factors. The calculation of study duration was based on the temporal interval of the dependent variables dataset minus the number of years without sampling. In the few cases in which quantitative data were missing (e.g. concerning time-lag, number of studied species, sampling frequency), semi-quantitative classes (e.g. from one to five variables, more than five variables) were used to characterise the studies. Data analysis methods used to select variables before analysis (e.g. correlation tests among explanatory variables to avoid multicollinearity issues) were not mentioned unless the method led to the creation of new variables used in the analysis process itself (e.g. ordination axes summarising an extensive set of variables). 'Secondary' analysis tools (i.e. resampling techniques and post-hoc tests) associated with the 'main' methods were recorded but were not developed in this review. Spearman's rank order correlation (Spearman, 1907) was performed to test the correlation between the number of articles published per year and the year. More details on study characterisation methodology are available as Supplementary Materials (Table A.2).

3. Results

Publication dates of the 99 reviewed articles range from 1997 to 2018. The number of reviewed publications increased progressively from 1995 to 2017 (Spearman's rank order correlation; ρ = 0.769; S = 306.49; p-value < 0.001). This is shown graphically for data grouped into five-year intervals, with the exception of the most recent data spanning 2.5 years (Fig. 2).

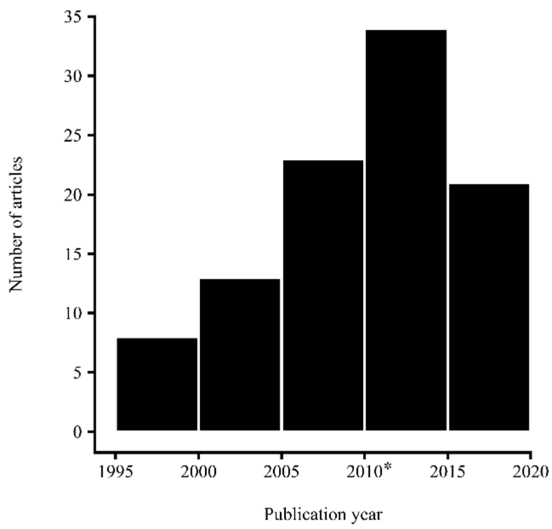

Fig. 2 (single column): Evolution of the number of reviewed articles published over time. The last bar represents two and a half years. The average publication date is indicated by the star.

3.1. Ecosystems

Most of the articles concentrated on aquatic habitats, with almost half focusing on marine habitats, and about a quarter on terrestrial habitats (Table 2). Most terrestrial studies were conducted in grasslands or woodlands, but five articles involved different kinds of terrestrial habitats at the same time.

Habitat category	**Occurrence**
Aquatic habitats	78
Marine habitats	*42*
Freshwater habitats	*27*
Brackish habitats	*9*
Terrestrial habitats	21
Woodlands	*6*
Grasslands	*6*
Several habitats	*5*
Anthropised habitats	*2*
Desert	*1*
Marshes	*1*

Table 2: List of habitats studied in the reviewed articles, ranked by number of publications. Marine habitats include intertidal habitats; Brackish habitats combine estuaries and lagoons; Grasslands include steppe, scrubland and tundra; Anthropised habitats include crop fields.

Wide and non-monophyletic group names were used in almost half of the articles (n = 47) to describe their biological material (Table 3, upper part). Several groups and taxa were highly represented: Plankton was studied in almost a quarter of the articles; fish, crustaceans and benthic communities were all examined in more than 10 % of the studies. Several phyla, exclusively or mostly associated with aquatic habitats (e.g. Cnidaria, Mollusca and Echinodermata) were rarely investigated but may have been included in articles studying zooplankton, aquatic invertebrates or benthic communities. Similarly, hexapods were mainly studied in terrestrial habitats (n = 6, 85.7 % of hexapod studies), but we assume that insects were often studied as part of aquatic invertebrates and benthic communities. 'Green plants', mammals and birds were mostly studied in terrestrial habitats (Viridiplantae, n = 9, 100 %; Mammalia, n = 6, 85.7 %; Aves, n = 5, 83.3 %).

Several articles investigated many taxa or species groups: for instance, Tian et al. (2006) used fisheries' catch results and included 58 species of fish, molluscs, crustaceans, echinoderms, marine mammals and algae. On the other hand, Clotfelter et al. (2007) focused on 'only' 13 species but at four trophic levels (two oak species, three rodents, one songbird and 7 raptors that were only used as dependent variables), analysing interactions between biotic and abiotic factors at different trophic levels.

Taxon	Occurrence
Planktonic communities	24 [1]
Zooplankton	16
Phytoplankton	11
Benthic communities	15
Aquatic invertebrates	8
ANIMALIA	48 [1]
Chordata	27
Vertebrata	26 [1]
Fishes [2]	*16*
Aves	*6*
Mammalia	*7*
Reptilia [2]	*1*
Tunicata	1
Arthropoda	19
Crustaceans [2]	11
Hexapoda	7
Insecta	*4*
Collembola	*3*
Unspecified	1
Mollusca	3
Cnidaria	3
Syndermata	2
Echinodermata	2
Porifera	1
PLANTAE	10
Viridiplantae	9
Rhodophyta	1
CHROMISTA	2

[1] The occurrence differs from the sum of lower rank occurrences because several lower taxa may be studied in one article.

[2] Paraphyletic group, treated as taxonomically relevant because of its low rank

Table 3: List of taxa and communities studied in the reviewed articles, ranked by number of publications. The upper part concerns communities with no taxonomic relevance and the lower part concerns taxa. The occurrence of one studied organism in a study is incremented in only one part (upper or lower) of the table. For the lower part, Kingdoms are indicated in capital letters, Phyla in standard and Classes in italics.

3.2. Localities

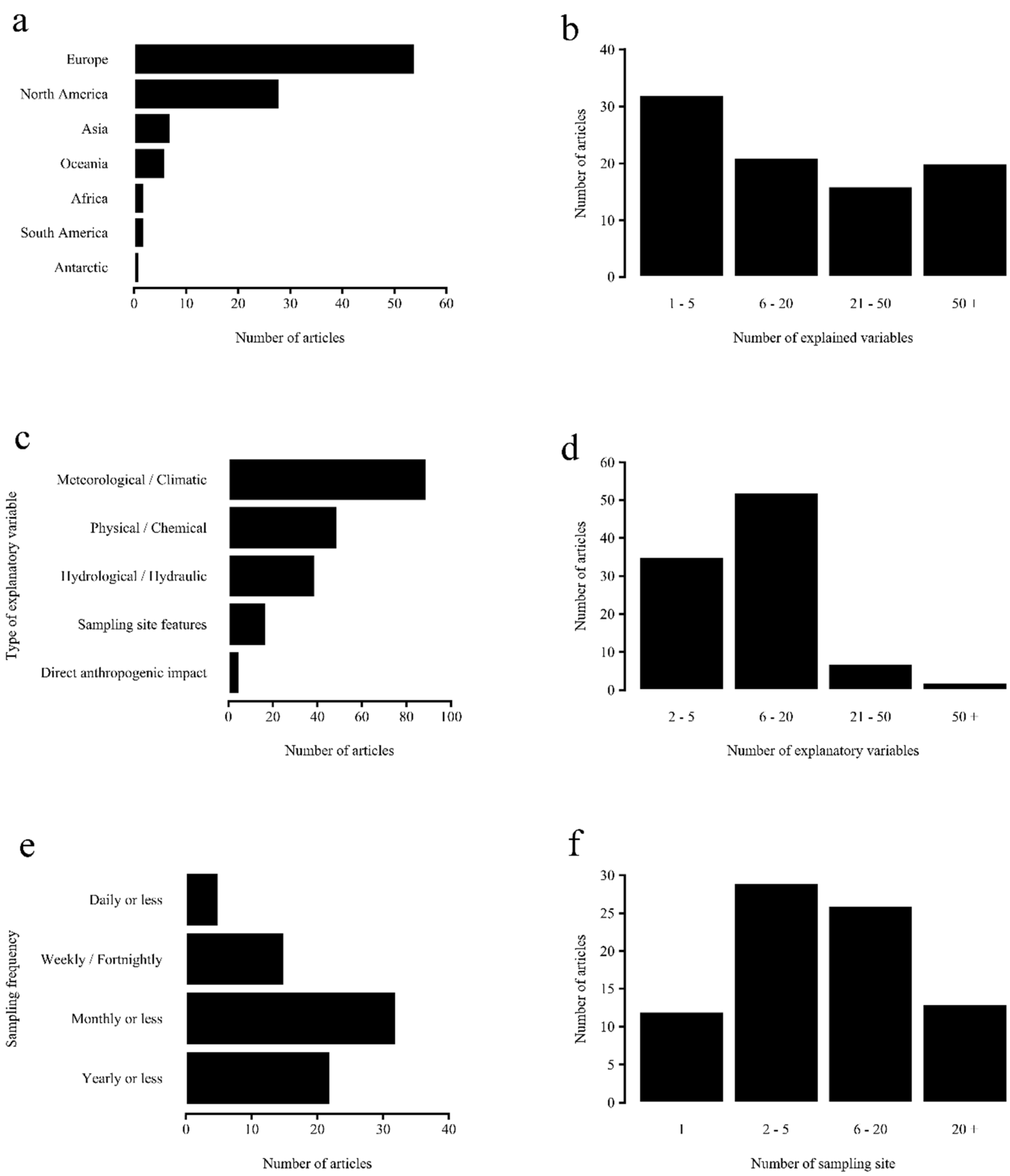

Fig. 3 (2-column): Distribution of the number of reviewed articles by continent (a), number of explained variables (b), type of explanatory variable (c), number of explanatory variables (d), sampling frequency (e), and number of sampling sites (f).

More than half of the reviewed studies concerned Europe (n = 54) and more than a quarter were conducted in North America (n = 28) (Fig. 3.a). The remaining quarter concerned Asia (n = 7), Oceania (n = 6), Africa (n = 2), South America (n = 2) and Antarctica (n = 1). Ershova et al. (2015) counted for both North America and Asia because the study location was the Chukchi Sea. The most commonly featured country was the United States of America, with 23 articles. LTER and similar networks (e.g. eLTER, iLTER) were involved (i.e. funding of the research project or sampling on a study site of this network) in 8 articles: three in the USA, three in Europe, one in Brazil and one in the Arctic.

3.3. Aims

For more than half of the studies (n = 57), the analysis of environmental effects on biological communities was a secondary goal used to understand the temporal (i.e. intra- or inter-annual) or spatial variability of biotic variables by comparing it to spatio-temporal variability of abiotic parameters (e.g. climatic variables or sampling site features). For instance, Kimmel and Roman (2004) assessed monthly abundance variability of two copepod species in relation to water-quality metrics. Consequently, the analysis of abiotic effects on biotic parameters often followed a trend analysis or a comparison of biological measurements between locations in the article structure (e.g. Möllmann et al., 2000).

A second category, comprising 36 articles, involved studies whose primary goal was to relate structure and community composition to abiotic characteristics. These articles aimed (1) to detect differences of ecological preferences between taxa (e.g. comparison of the habitat of native and invasive fish species in Haupt and Phelps, 2016), (2) to understand differences in community assemblages and structures (e.g. Brooker et al., 2012 highlighted the effects of temperature and precipitation on spatial patterns of plant communities), or (3) to compare different effects of environmental parameters on biological communities (e.g. short-term versus long-term climatic effects on bird distribution in Bateman et al., 2016; climatic parameters versus logging effect on zooplankton communities in Lévesque et al., 2017).

Lastly, a few articles (n = 6) aim to evaluate one very specific effect but include several confounding variables in the analyses to control for other effects and isolate the studied one. For instance, to study the long-term effect of an oil spill on a benthic population, Poggiale and Dauvin (2001) created a population dynamics model taking into account not only environmental pollution (i.e. the factor to be studied) but also sea-water temperature and competition (i.e. confounding factors).

3.4. Explained variables

A third of the articles (n = 32) concerned a small number of explained variables (from one to five) (Fig. 3.b). Twenty-one studies used between six and 20 response variables, and 37 articles used more than 20 predictands (from 21 to 50, n = 16; more than 50, n = 20). We were not able to determine the number of dependent variables for 10 articles.

Response variables used in the articles concern either a single species or a group of species making up a considerable proportion of the studied community (e.g. guild, functional group, trait category, size class or taxonomic ranks above species such as genus, family and order). Most metrics, hereafter called ‘species-specific metrics’ (Table 4, upper part) can be applied to both categories. Most of them are abundance-based, but alternative species-specific metrics were also used in the reviewed literature (e.g. biomass, biovolume, covered surface).

On the other hand, several variables, hereafter called ‘community metrics’ (Table 4, lower part), require more than one species to be calculated and meaningful. Consequently, they are only used on groups of species, and mostly on all the organisms studied. These indices summarise community characteristics and usually outline the taxonomic structure and composition of the community, mostly by means of diversity indices (e.g. Species richness, Shannon index, Simpson index, Pielou's evenness). However, community metrics can also focus on alternative community characteristics, such as the mean trophic level or biological traits (e.g. life history, morphology, physiology, behaviour) using fuzzy coding (see Chevenet et al., 1994, for further details on the methodology).

Species-specific metrics were employed in 90 articles, either in combination with community metrics (n = 19, 21.1 % of articles using species-specific metrics) or not (n = 71, 78.9 %). Community metrics were only analysed alone in a few studies (n = 9).

1. Can be applied both on a single species and on a group of species?		2a. Based on the number of individuals?			
		Yes		**No**	
Yes	**'Species-specific metrics'**	Number of individuals	(Fasola et al., 2010; Gutierrez et al., 2016; Obaza et al., 2015)	Biomass	(Dippner and Ikauniece, 2001; Lavaniegos and Ohman, 2003; Wasmund et al., 2011)
		Occurrence	(Bateman et al., 2016; Casey et al., 2015)	Biovolume	(Ayón and Swartzman, 2008; Horn et al., 2011)
		Density	(Aleksandrov et al., 2009)	Tree growth	(Laurance et al., 2009)
		Recruitment	(Menge et al., 2011)	Basal area	(Laurance et al., 2009)
		Mortality	(Laurance et al., 2009)	Surface cover	(Gross and Edmunds, 2015)
		Catch Per Unit Effort (CPUE)	(Haupt and Phelps, 2016; Hurst et al., 2004; James et al., 2008)	Aboveground Net Primary Production	(Childers et al., 2006)
		Phenology	(van Walraven et al., 2017)	Number of nests	(Fasola et al., 2010)
		Reproductive success	(Gauthier et al., 2013)	Spatial Associations	(Brooker et al., 2012)
		2b. Summarises the taxonomic structure and composition of the community?			
		Yes		**No**	
No	**'Community metrics'**	Species Richness	(Bortolini et al., 2014; Szentkirályi et al., 2007; Vaughan and Ormerod, 2012)	Biological Traits (Fuzzy coding)	(Bêche and Resh, 2007; Latli et al., 2017; Lawrence et al., 2010)
		Shannon index	(Carballo et al., 2008; Pitacco et al., 2018; Zettler et al., 2017)	Mean trophic level	(Tian et al., 2006)
		Simpson Index	(Penczak, 2011)	Bird Community Index	(Ladin et al., 2016)
		Pielou Index	(Pitacco et al., 2018; Zettler et al., 2017)		
		Stability of composition	(Marchant and Dean, 2014)		

Table 4: Non-exhaustive list of response variables used in the reviewed articles and up to three related examples. Rare and specific indices with complex names are not shown.

3.5. Explanatory variables

All quantitative predictors can be based on one or multiple measures (e.g. average, sum, variability, minimum, maximum). They can also illustrate previous abiotic conditions (e.g. weather variables, one, two and three years before measurement in Clotfelter et al., 2007). Half of the reviewed articles (n = 52) used from six to 20 independent variables, but a large number used fewer (from 2 to 5, n = 35) (Fig. 3.d). Only 9 articles used more than 20 predictors (from 21 to 50, n = 7; more than 50, n = 2). We were not able to determine the number of explanatory variables for three articles. It is worth noting that the two studies with more than 50 predictors (Kwok et al., 2016, and Ladwig et al. 2016) obtained a large number of explanatory variables because of the extensive use of different time lags on a modest number of measures; they did not include more than 50 unrelated parameters. Meteorological (e.g. temperature, cloudiness, radiation, humidity, precipitation, wind speed, atmospheric pressure) and climatic measures (North Atlantic Oscillation, North Pacific Gyre Oscillation, Pacific Decadal Oscillation, El Niño-Southern Oscillation, Southern Oscillation Index, North Pacific Index, Artic Oscillation Index, etc.) were used as predictors in almost all the reviewed studies (n = 89) (Fig. 3.c), followed by physical and chemical characteristics (e.g. salinity, pH, dissolved oxygen, nutrients and pollutant concentration) and hydrological and hydraulic metrics (e.g. depth, velocity, discharge, turbidity, wave height, sinuosity) in 49 and 39 articles respectively. The effects of sampling site features (e.g. topography, elevation, land cover, habitat type, sedimentary characteristics, distance to the coast) were analysed in 17 studies, while direct anthropogenic impact (e.g. hunting, logging, engineering, oil spill) was studied in only five articles. Sixty-eight articles used two or more categories of abiotic variables.

3.6. Sampling strategy

The average study duration of the 99 papers was 23.46 years (min: 3; max: 114; sd: 15.92; median: 20) (Fig. 4). The sampling frequency occurred monthly or at least several times a year in 32 articles, but a significant number of the studies (n = 22) involved yearly sampling or less (Fig. 3.e). Fieldwork performed more than once a month was less frequent (weekly or fortnightly, n = 15; daily or several times a week, n = 5). Sampling frequency was not indicated in a large number of studies (n = 25).

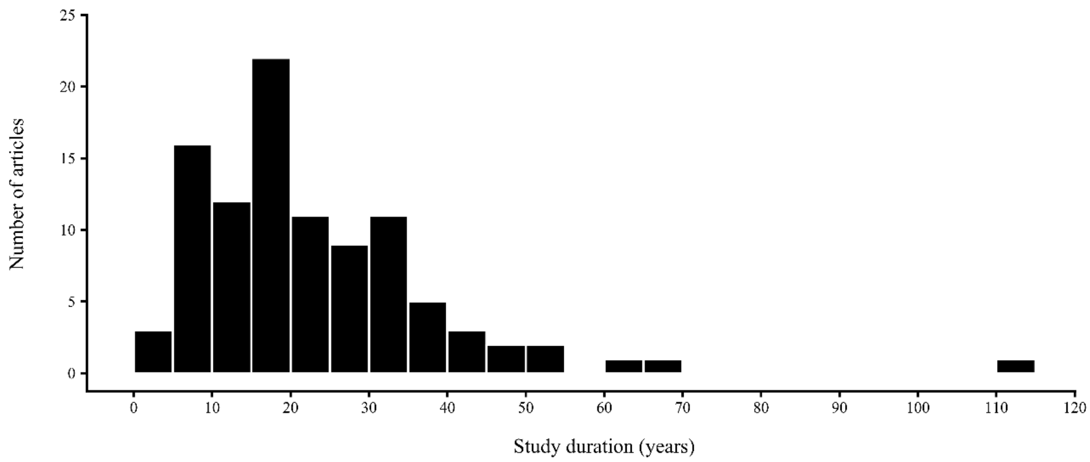

Fig. 4 (2-column): Distribution of study duration of the reviewed articles.

Only three studies can be considered as experimental (i.e. at least one environmental condition was controlled): Daghighi et al. (2017), Gutiérrez-Fonseca et al. (2018) and Waterkeyn et al. (2011) who studied the effects of salinity and hydrological disturbance in mesocosms. The remaining 96 articles were observational studies. The articles studying only one site were similar in number to those studying more than 20 (one site, n = 12; more than 20 sites, n = 13) (Fig. 3.f).

Most papers involved an intermediate number of sampling stations (2 - 5, n = 29; 6 - 20, n = 26). The number of sampling sites in 19 articles could not be determined.

3.7. Data analysis techniques

In all, 40 data analysis methods were recorded. Univariate methods were performed in 73 articles and multivariate techniques were used in 54 articles. Both methods were used in 28 articles. Statistical modelling is the most frequent type of analysis, used in almost half of the reviewed articles (n = 46). In particular, linear models were used in 34 studies (Table 5). Ordination methods are also common, used in 45 articles (constrained ordination, n = 26; unconstrained ordination, n = 25). Statistical tests, mostly correlation tests (n = 27, 73.0 % of all studies with statistical tests), were used in 37 studies. The remaining analysis categories were performed in fewer than four articles. There is a large gap in the use of univariate methods between the two main techniques (i.e. linear modelling and correlation tests) and the others (e.g. additive modelling, Maxent). There is no such gap in multivariate methods. At least two distinct data analysis methods were combined in 43 articles to assess the effects of abiotic factors on biological communities, with up to five methods involved in the reviewed section of one article. Ordination methods were often performed in combination with other analysis methods (Fig. 5). For instance, unconstrained ordination was associated with statistical tests in 13 studies, and constrained ordination was combined with statistical modelling in 8 articles. Statistical modelling is the most common type of analysis used singly (n = 27, 58.7 % of all studies performing statistical modelling).

Method	Acronym	Type of analysis	Examples	Occurrence	
				Total	Lone use
Linear models	LM	Statistical modelling	(Brooker et al., 2012; Jourdan et al., 2018; Lavaniegos and Ohman, 2003)	34	50 %
Correlation test	-	Statistical test	-	27 [1]	37 %
Pearson correlation test	-	Statistical test	(Einarsson and Björk Örnólfsdóttir, 2004; Hall and Rudstam, 1999; Möllmann et al.,	17	41 %
Spearman's rank order correlation	-	Statistical test	(Brown and Edmunds, 2013; Carballo et al., 2008; Sponseller et al., 2010)	6	0 %
Cross-correlation functions	CCF	Statistical test	(Gröger and Rumohr, 2006; Licandro et al., 2012; Szentkirályi et al., 2007)	3	33 %
Partial correlation analysis	-	Statistical test	(van der Wal and Stien, 2014)	1	0 %
unspecified	-	Statistical test	-	2	50 %
Principal Components Analysis	PCA	Unconstrained ordination	-	20	0 %
Ordinary Principal Components Analysis	PCA	Unconstrained ordination	(Bortolini et al., 2014; Buttay et al., 2016; Tian et al., 2006)	17	0 %
3-mode Principal Components Analysis	3-mode PCA	Unconstrained ordination	(Beaugrand et al., 2000)	1	0 %
Between-dates Principal Components Analysis	bPCA	Unconstrained ordination	(Latli et al., 2017)	1	0 %
Eigen Vector Filtering	EVF	Unconstrained ordination	(Licandro et al., 2012)	1	0 %
Redundancy Analysis	RDA	Constrained ordination	-	9 [1]	33 %
Ordinary Redundancy Analysis	RDA	Constrained ordination	(Beuchel et al., 2006; Wasmund et al., 2011; Waterkeyn et al., 2011)	6	17 %
Distance-based Redundancy Analysis	db-RDA	Constrained ordination	(Abonyi et al., 2018; Pitacco et al., 2018; Zettler et al., 2017)	3	67 %
Partial Redundancy Analysis	p-RDA	Constrained ordination	(Horn et al., 2011)	1	0 %
BIO-ENV procedure	BIO-ENV	Statistical test	(Barrio Froján et al., 2008; Kimball et al., 2014; Taylor et al., 2017)	8	75 %
Canonical Correspondence Analysis	CCA	Constrained ordination	(Feike et al., 2007; Henderson et al., 2011; Hurst et al., 2004)	7	71 %
Generalized Linear Model	GLM	Statistical modelling	(Fasola et al., 2010; Gutierrez et al., 2016; Kwok et al., 2016)	6	50 %
Correspondence Analysis	CA	Unconstrained ordination	-	5	20 %
Ordinary Correspondence Analysis	CA	Unconstrained ordination	(Daufresne et al., 2004; Vaughan and Ormerod, 2012)	2	0 %
Fuzzy Correspondence Analysis	FCA	Unconstrained ordination	(Bêche and Resh, 2007; Feio et al., 2015)	2	50 %
Detrended Correspondence Analysis	DCA	Unconstrained ordination	(Bortolini et al., 2014)	1	0 %

Canonical Correlation Analysis	CCorA	Constrained ordination	(Dippner et al., 2000; Kröncke et al., 1998; Molinero et al., 2006)	5	20 %
Process-based Model	-	Mechanistic modelling	(Kerimoglu et al., 2013; Mutshinda et al., 2017; Poggiale and Dauvin, 2001)	3	100 %
Generalized Additive Model	GAM	Statistical modelling	(Latli et al., 2017; Ribic et al., 2011; Silva et al., 2014)	3	67 %
Non-metric Multidimensional Scaling	nMDS	Unconstrained ordination	(Gutiérrez-Fonseca et al., 2018; Horn et al., 2011; Jucevica and Melecis, 2006)	3	0 %
Maximum Entropy Modelling	Maxent	Statistical modelling	(Bateman et al., 2016; Casey et al., 2015)	2	50 %
Self-Organizing Map	SOM	Artificial Neural Network	(Choi et al., 2015; Pfister, 2006)	2	50 %
Path analysis - Causal model	-	Statistical modelling	(Hallett et al., 2014; Irvine et al., 2015)	2	50 %
Additive Model	AM	Statistical modelling	(Carballo et al., 2008; Vaughan and Ormerod, 2012)	2	0 %
Hierarchical partitioning	HP	Statistical modelling	(Latli et al., 2017; Marchant and Dean, 2014)	2	0 %
Classification and Regression Tree	CART	Classification	(Ayón and Swartzman, 2008)	1	100 %
Multiple Discriminant Analysis	MDA	Constrained ordination	(Kodama et al., 2002)	1	100 %
STATICO method	-	Constrained ordination [2]	(Mazzocchi et al., 2012)	1	100 %
Asymmetric Eigenvector map	AEM	Spatial Eigenfunction analysis	(Lévesque et al., 2017)	1	0 %
Co-inertia Analysis	CoIA	Constrained ordination	(Latli et al., 2017)	1	0 %
Min/Max Auto-correlation Factor Analysis	MAFA	Unconstrained ordination	(van Walraven et al., 2017)	1	0 %
Permutational multivariate analysis of variation	PERMANOVA	Statistical test	(Gutiérrez-Fonseca et al., 2018)	1	0 %
Multivariate analysis of variation	MANOVA	Statistical modelling	(Penczak, 2011)	1	0 %
Quantile regression	-	Statistical modelling	(Menge et al., 2011)	1	0 %
Student's t-test	-	Statistical test	(Lawrence et al., 2010)	1	0 %
Variation Partitioning	VP	Constrained ordination & Statistical modelling	(Lévesque et al., 2017)	1	0 %
Cross-wavelet	-	Wavelet analysis	(Menge et al., 2011)	1	0 %
Wavelet coherence	-	Wavelet analysis	(Menge et al., 2011)	1	0 %

[1] The occurrence differs from the sum of alternative methods occurrences because one article may use several alternatives

[2] This method combines both constrained and unconstrained ordination

Table 5: List of data analysis methods performed in the reviewed studies, ranked by their occurrence in publications. When a mixture of methods is used, its occurrence is incremented but the occurrence of the different components of the method is not. The "Example" column contains up to three references of reviewed articles using the method.

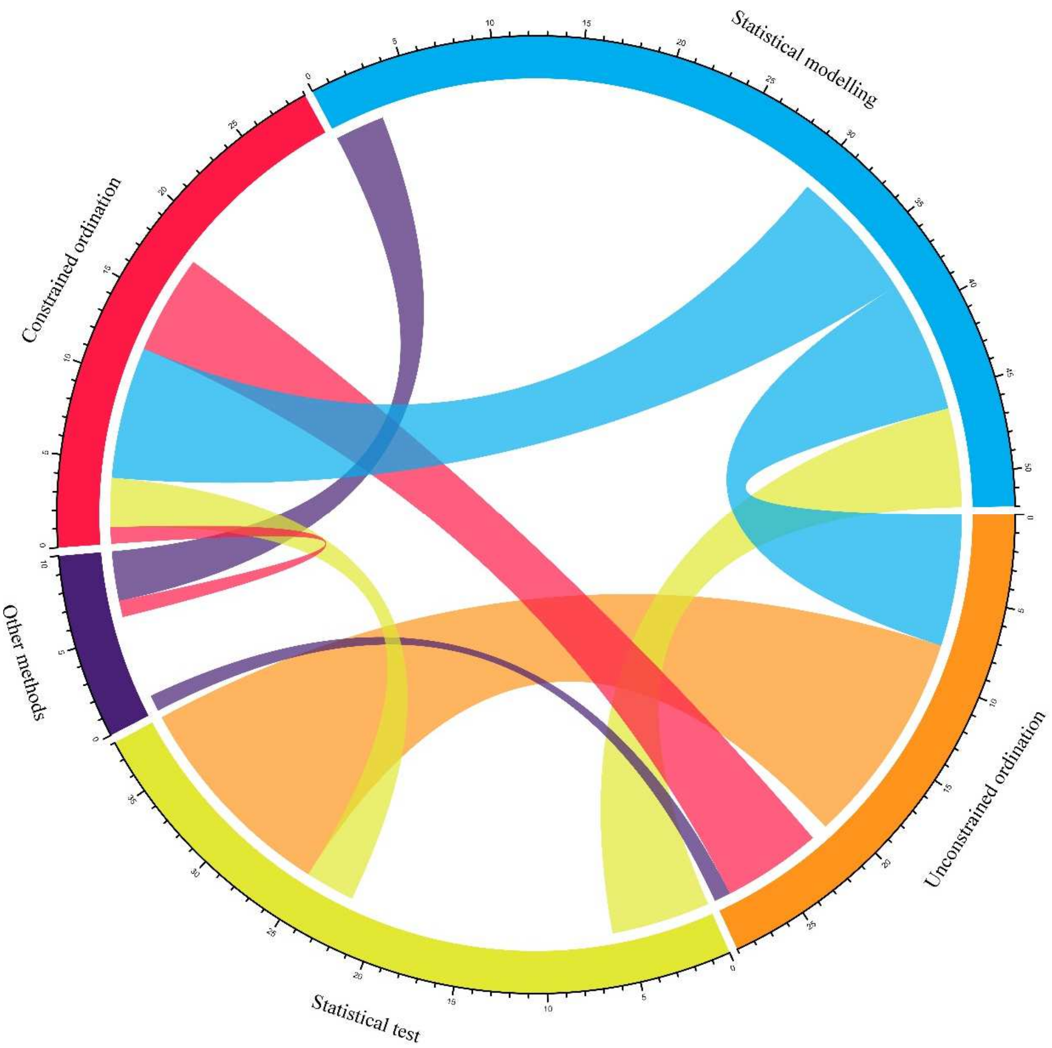

Fig. 5 (2-column): Chord diagram representing combinations of analysis types. Arcs represent pairs of methods used in the same article, with their size proportional to the number of articles in which the combination occurs (the ticks on the outer part of the circle represent articles). Sectors of the circle with no arc represent studies using only one method or a combination of methods belonging to the same group. For studies performing more than two types of analysis, one arc is drawn for each pair. This diagram was drawn with the R package ‘Circlize’ (Gu et al., 2014).

3.7.1. *Statistical modelling*

For univariate analyses, researchers can choose between three main types of model identified by Levins (1966) in population biology: analytical (or mathematical) models, empirical (or statistical or phenomenological) models, and mechanistic models (also called causal models or process-based models) (Guisan and Zimmermann, 2000). Most modelling techniques used in the studies belong to the empirical family, i.e. models that aim to combine precision (accuracy of predicted response) and realism (unrealistic assumptions are limited) (Levins, 1966), but do not describe realistic cause-effect links between predictors and the response variables, or inform about underlying ecological functions and mechanisms (Guisan and Zimmermann, 2000).

In the reviewed literature, statistical modelling was mainly performed by means of linear models (Table 5), mostly with ordinary and simple linear models (n = 23, 50.0 % of articles including modelling approaches, Table 6), e.g. linear regressions, ANOVA, ANCOVA. They were used with other methods half of the time. The purpose of linear modelling is to describe the relationship between a single response variable and a set of explanatory variables, in order to test hypotheses about the model parameters, or to forecast or predict values of the response variable (Legendre and Legendre, 2012). The response variable and the predictors cannot be interchanged. The term 'linear' is misleading: it is possible to model non-linear relationships (e.g. polynomial, trigonometric functions) with linear models. Those models are "linear in the parameters" (Zuur et al., 2009), i.e. each term of the explanatory part of the model is either a constant or the product of a parameter (i.e. coefficient) and a predictor. The random part of an ordinary linear model is only composed of the real random term (i.e. the residuals). However the random part can be extended (see Table 6) with components that allow the model to take into account heterogeneity (i.e. generalized least squares model), nested data (i.e. mixed effect model), and temporal or spatial correlations (e.g. auto-regressive model, auto-regressive moving average model) (Zuur et al., 2009). These extensions were used in many of the studies performing linear modelling (n = 13, 37.1 %). Linear models can also be combined in a hierarchical model (HM, MacKenzie and Kendall, 2002; Royle, 2004), which is a sequence of models ordered by their conditional probability structure (Santoro et al., 2016) and which involves at least one model (i.e. level of the HM model) referring to an unobserved variable (e.g. population abundance). Santoro et al. (2016) used a two-level hierarchical model in order to account for temporal variation in the probability of capture to evaluate changes in abundance.

When the dependent variable is not Gaussian, a generalized linear model (GLM) with a non-normal distribution (e.g. Poisson, Binomial, Gamma) can be used to relate the explanatory variables to the response variable through a link function (e.g. logarithm, logit) (Guisan et al., 2002). However, relatively few articles (n = 6) performed GLM compared to linear models, even though the distribution of the dependent variable is often not Gaussian in ecology (e.g. count data, proportional data, presence-absence data). The same extensions of the random part for linear models are also available in their generalized form.

In additive modelling, in contrast to linear regressions, a smoothing function is used to link an explanatory quantitative variable to the response variable instead of a coefficient (Zuur et al., 2009). This non-parametric method was used in only five articles. It is an appropriate way of evaluating an empirical relationship instead of estimating the parameters of a model (Legendre and Legendre, 2012). It is also possible to generalise this method to non-Gaussian distributions and to extend the random part.

While frequentist methods are efficient for model comparison and evaluation, Bayesian frameworks, i.e. approaches that consider that the parameters to be estimated are not fixed and where prior knowledge can be used (Zuur et al., 2009), have many advantages in linear and additive modelling (Dorazio, 2016; Garamszegi et al., 2009; Hobbs and Hilborn, 2006). However, only Gutierrez et al. (2016) took a Bayesian approach by using Markov chain Monte Carlo (MCMC).

		Type of effect			
		Linear		**Additive**	
Distribution of the response variable	**Normal**	Linear model	34 [1]	Additive model	2
		Ordinary (LM)	23	Ordinary (AM)	2
		Random part extended with:	13	Random part extended with:	0
		Term allowing for heterogeneity	0	*Term allowing for heterogeneity*	0
		Correlation structure	6	*Correlation structure*	0
		Random effect (LMM)	6	*Random effect (AMM)*	0
		Correlation structure & random effect	1	*Correlation structure & random effect*	0
	Non-normal	Generalized Linear model	6	Generalized additive model	3
		Ordinary (GLM)	0	Ordinary (GAM)	1
		Random part extended with:	6	Random part extended with:	2
		Term allowing for heterogeneity	0	*Term allowing for heterogeneity*	0
		Correlation structure	2	*Correlation structure*	1
		Random effect (GLMM)	2	*Random effect (GAMM)*	0
		Correlation structure & random effect	1	*Correlation structure & random effect*	0
		Hierarchical Model (HM)	1	*Hierarchical Model (HM)*	1

[1] The occurrence differs from the sum of alternative methods occurrences because one article may use several alternatives.

Table 6: Linear and additive model types used in the reviewed studies and their occurrence

Other statistical modelling methods were rarely performed (n = 8). Hierarchical partitioning (HP, Chevan and Sutherland, 1991) can be used in combination with other modelling techniques (e.g. linear regression) in order to assess the independent contribution of each predictor to the variation of the dependent variable. It also enables the authors to avoid both multi-collinearity issues and overfitting (Chevan and Sutherland, 1991). This method was used in Marchant and Dean (2014) and in Latli et al. (2017) with linear models and GAM respectively.

Maximum entropy modelling (Maxent, Jaynes, 1957) is a type of Species-Distribution Model (SDM) very similar to GLM/GAM, which does not require absence data and has many advantages over other modelling of species distribution using presence-only data (Phillips et al., 2006). It was used by Bateman et al. (2016) and Casey et al. (2015).

Path analysis (Wright, 1960, 1921) is a special case of structural equations modelling (Grace et al., 2012; Legendre and Legendre, 2012) used to test causal hypothesis between multiple variables (Shipley, 2013). Unlike multiple linear regression, from which it derives (Legendre and Legendre, 2012), it allows for more than a simple correlative relationship between two variables by including mediator, moderator and covariate variables in the causal model. Irvine et al. (2015) used this technique to assess the impact of anthropogenic drivers on biological conditions with Bayesian estimation of the parameters. It was also used by Hallett et al. (2014), who called it ‘structural equation modelling’, to relate precipitation to community stability *via* species richness, compensatory dynamics and dominant species stability.

Quantile regression (Koenker and Bassett, 1978) is a method that estimates multiple slopes to describe the relationship between a response variable and predictors. It is useful in cases of linear regression with heterogeneous variances (Cade and Noon, 2003). It was performed by Menge et al. (2011) to evaluate barnacle and mussel recruitment in response to climatic factors.

A multivariate analysis of variance (MANOVA), the equivalent of a univariate analysis of variance (ANOVA) applied to several continuous and independent response variables (Buttigieg and Ramette, 2014; James and McCulloch, 1990), was performed in Penczak (2011).

3.7.2. Ordination methods

Several problems occur when authors have to investigate the link between many explained variables and many drivers: drawing all possible scatterplots is not only laborious but also uninformative, and it is not possible to draw a plot with more than two or three axes, each one representing a descriptor (Legendre and Legendre, 2012). However, ordination methods, widely used in multivariate frameworks, make it possible to project such a multidimensional scatter plot onto bivariate graphs with axes representing a substantial portion of the data matrix variability in a reduced space. A wide variety of methods were used in the reviewed literature, belonging either to the unconstrained family (Table 7, upper part), or to the constrained family (Table 7, lower part). Unconstrained ordination techniques aim to display a maximum amount of a dataset's variability on a few axes without presuming causal relationships (Ramette, 2007). These methods are considered as exploratory and are used for indirect gradient analysis: gradients are unknown *a priori* and are inferred from the response data (Buttigieg and Ramette, 2014). By contrast, in direct gradient analysis, gradients are known, measured and directly related to the response dataset (i.e. the biological variables in our case) (Buttigieg and Ramette, 2014). Direct gradient analysis can be performed with constrained ordination methods, also called canonical analysis, by comparing two or more matrices. Indeed, these techniques aim to display only the dataset variation that can be modelled with constraining variables. We identified two types of constrained ordination in the reviewed articles: asymmetric methods and symmetric methods. The asymmetric forms of constrained ordination involve a response matrix (i.e. biotic variables in our case) and an explanatory matrix (i.e. environmental parameters in our case) that cannot be interchanged without consequences on the analysis. These methods combine ordination and regression: the ordination of the response matrix is constrained to be linearly related to the explanatory matrix (Legendre and

Legendre, 2012). However, asymmetric constrained ordination methods, like all traditional ordination techniques used in the articles, are considered as ‘algorithmic’, because the statistical properties of the data are not taken into account, in contrast to ‘model-based’ methods, which involve the specification of a statistical model (Hui et al., 2015; Warton et al., 2015). In symmetric constrained ordination techniques, the two data sets play the same role (Legendre and Legendre, 2012). Most of these methods are considered as interpretive (e.g. CCA, RDA, CCorA) (Paliy and Shankar, 2016).

Method	**Input data**	**Relation**	**Reference**
PCA	Raw data	Linear	(Hotelling, 1933; Pearson, 1901)
3-mode PCA	Raw data	Linear	(Tucker, 1966)
bPCA	Raw data	Linear	(Dolédec and Chessel, 1987)
EVF	Autocovariance matrix	Linear	(Ibanez and Conversi, 2002)
CA	Raw data	Unimodal	(Benzecri, 1969; Fisher, 1940; Hirschfeld and Wishart, 1935)
FCA	Fuzzy-coded data	Unimodal	(Chevenet et al., 1994)
DCA	Raw data	Unimodal	(Hill and Gauch, 1980)
nMDS	Distance matrix	Any [1]	(Kruskal, 1964a, 1964b; Shepard, 1962)
MAFA	Raw data	Linear	(Shapiro and Switzer, 1989)
Asymmetric methods			
RDA	Raw data	Linear	(Rao, 1964)
db-RDA	Distance matrix	Any [1]	(Legendre and Anderson, 1999)
p-RDA	Raw data	Linear	(Davies and Tso, 1982)
CCA	Raw data	Unimodal	(ter Braak, 1986)
VP	Raw data	Linear	(Borcard et al., 1992)
MDA	Raw data	Linear	(Fisher, 1936; Rao, 1948)
Symmetric methods			
CCorA	Raw data	Linear	(Hotelling, 1936)
CoIA	Ordination output	Any [2]	(Doledec and Chessel, 1994)

STATICO	Raw data	Any [2]	(Thioulouse et al., 2004)

[1] Depends on the type of distance chosen
[2] Depends on the ordination technique chosen

Table 7: Characteristics of ordination methods performed in the reviewed studies, unconstrained analyses in the upper part and constrained analyses in the lower part. “Reference” column contains the bibliographical sources that first introduced the method.

Indirect gradient analysis was performed with unconstrained ordination as exploratory method mostly by means of ordinary Principal Components Analysis (PCA), which was also the most widely used ordination technique in the reviewed articles (n = 17, 37.8 %). This method summarises a large number of quantitative variables in a few dimensions: principal components that are orthogonal to each other and consist of linear combinations of input variables (Legendre and Legendre, 2012). PCA was thus almost always used to limit the number of variables and avoid multicollinearity issues. Only Jahan and Choi (2014) used a single PCA to visualise correlations between abiotic and biotic variables before testing them with a correlation test. Therefore, PCA, in its ordinary form or in a derived form, was not used for analysis strictly speaking and was always performed in combination with other methods (e.g. correlation tests, linear models, Canonical Correlation Analysis). It was applied half the time (n = 9, 47.4 %) on both biotic and abiotic variables with two separate PCAs, only on predictors in 7 studies, and only on predictands in three articles. Different kinds of PCA were not often used: Beaugrand et al. (2000) performed a 3-mode Principal Components Analysis (3-mode PCA) based on three classical PCAs in order to assess variation of species abundance in time and space; a between-date PCA (bPCA) was used in Latli et al. (2017) to maximise the distance between the sampling years along the successive bPCA axes; and Licandro et al. (2012) used EigenVector Filtering adapted to time-series with missing values, which corresponds to a PCA calculated on an autocovariance matrix based on the original time-series lagged with itself.

Correspondence analysis (CA), also called reciprocal averaging, was performed in five articles but in three different forms: ordinary (n = 2), Fuzzy (n = 2) and Detrended (n = 1). Ordinary Correspondence Analysis was first proposed for the analysis of two-way contingency tables, but in ecology it is mostly applied on sites x species (presence / absence or abundance) matrices (Legendre and Legendre, 2012). Although this method is analogous to PCA, it aims at maximising the correspondence between rows and columns. Unlike PCA, the same result is thus produced when transposing the data matrix. Moreover, it preserves χ^2 distance instead of Euclidian distance (i.e. 'ordinary' straight line distance). Vaughan and Ormerod (2012) achieved a CA on biological communities, and then used the main axis of variation as a dependent variable. Fuzzy-coded Correspondence Analysis (FCA, Chevenet et al., 1994) can be applied on fuzzy-coded data (categories and subcategories, e.g. biological traits) and was used to relate community traits to temperature and precipitation in Feio et al. (2015), or to relate FCA axes to environmental parameters in Bêche and Resh (2007). Detrended Correspondence Analysis (DCA, Hill and Gauch, 1980) is used to remove the arch effect due to unimodal responses of communities to environmental gradients (Legendre and Legendre, 2012). It was performed by Bortolini et al. (2014) on biotic variables to produce axes, subsequently used as response variables.

Two other unconstrained ordination techniques were performed: non-Metric Multidimensional Scaling (nMDS) and Min/Max Autocorrelation Factor Analysis (MAFA). nMDS is not an eigenvector-based technique, so, unlike PCA and CA, the input data is a distance matrix. Its aim is to represent the objects in a restricted number of dimensions (i.e. two or three) with all data variance utilised, and it does not preserve the exact distance between objects; the interpretations are thus qualitative and subjective (James and McCulloch, 1990; Legendre and Legendre, 2012; Paliy and Shankar, 2016; Ramette, 2007). Jucevica and Melecis (2006) performed an nMDS on Collembola communities and then used the two axes in correlation tests. Gutiérrez-Fonseca et al. (2018) used nMDS to define groups between macroinvertebrates assemblages, and Horn et al. (2011) applied this method on diatom communities with vector fitting of explanatory variables prior to constrained ordination. MAFA is similar to PCA but the axes represent a measure of autocorrelation. Van Walraven et al. (2017) used this technique on biological data prior to correlation test with MAFA axes.

Redundancy Analysis (RDA) is the most frequently used constrained ordination technique (n = 9, 34.6 %) and was mainly performed in its ordinary form (n = 6, 66.7 %), but distance-based Redundancy Analysis (db-RDA) (n = 3, 33.3 %) and partial Redundancy (p-RDA) (n = 1, 11.1 %) were also used. Ordinary Redundancy Analysis, an asymmetric constrained ordination that allows for different types of explanatory variables (i.e. quantitative and qualitative), is the extension of multiple regression to the modelling of multivariate response data (Legendre and Legendre, 2012). However, as mentioned above, it is not considered to be ‘model-based’ (Warton et al., 2015). This method is also the canonical version of PCA where the components are constrained to be linear combinations of the environmental variables (Paliy and Shankar, 2016; Ramette, 2007). This technique is applied on sites x species matrices (response data set) and on sites x abiotic factors matrices (explanatory data set). For example, Li et al. (2015) conducted Ordinary RDA to analyse the effect of climate factors on vegetation assemblage. Distance-based Redundancy Analysis, also called distance-based linear model (DISTLM), is a particular form of RDA carried out on a distance matrix and thus allows an analysis based on various distance functions. It was performed in three studies (Abonyi et al., 2018; Pitacco et al., 2018; Zettler et al., 2017), for example, to quantify the variation in benthic community explained by climatic parameters in Pitacco et al. (2018). Partial Redundancy Analysis is used to analyse the effect of an explanatory matrix X on the matrix Y adjusted for the effect of covariables in a matrix W. This method was used by Horn et al. (2011) to isolate the effect of climate variables from that of trophic variables on a phytoplankton community. Lévesque et al. (2017) used variation partitioning (VP) by Redundancy Analysis in order to partition the effects of temporal, spatial and environmental parameters on zooplankton communities.

Canonical Correspondence Analysis (CCA), another asymmetric technique similar to RDA in many aspects, was performed in 7 articles, mainly with no other data analysis method (n = 5, 71.4 %). Any data suitable for CA can be used as the response matrix Y (Legendre and Legendre, 2012; Ramette, 2007), and although CCA is suitable for unimodal responses, it seems to be robust for other responses (e.g. bimodal, unequal ranges) (Ramette, 2007). However, its predictive power is inferior to that of GLM, because the same explanatory variables are used for the whole community in constrained ordination, whereas species-specific subsets of predictors can be defined in modelling (Guisan et al., 1999). CCA was used in Garcia et al. (2012) to study the relationship between fish species abundance and environmental variables (water temperature, salinity, rainfall), and also in Pace et al. (2013) to analyse the link between taxa abundance of aquatic insects and environmental variables. Like CA, this method preserves χ^2 distance.

Multiple Discriminant Analysis (MDA), also called Linear Discriminant Analysis (LDA) or Discriminant Function Analysis (DFA), is not interpretive but discriminatory (Paliy and Shankar, 2016). This technique is used to determine the linear combination of explanatory variables that best defines an already known grouping of objects (Legendre and Legendre, 2012). However, this method is often used as an exploratory ordination technique (James and McCulloch, 1990). It was used in Kodama et al. (2002) to explain the groups of fish and invertebrates constituted with nMDS and cluster analysis with environmental variables.

Canonical Correlation Analysis (CCorA) is the symmetric equivalent of RDA and was the most frequently used symmetric constrained analysis method (n = 5, 71.4 %). The difference between these two techniques can be compared to the difference between simple linear regression and linear correlation (Legendre and Legendre, 2012). Studied objects (e.g. sites) in CCorA are described by two matrices containing quantitative parameters (i.e. raw data) and treated symmetrically. Legendre and Legendre (2012) stated that this method has limited applications nowadays for two reasons: (1) the use of Co-inertia Analysis (CoIA), a similar but more flexible technique, and (2) many ecological issues are asymmetric, i.e. variables are defined as explanatory or explained by the study design, and their roles cannot be swapped. Nevertheless, it was used in five articles, once alone in Aßmus et al. (2009), and four times with other methods, for example in Molinero et al. (2006) where PCA was first conducted on both biotic and abiotic variables before the axes were submitted to a CCorA.

As explained above, CoIA is an alternative method to CCorA, presenting many advantages; for example, it allows multicollinearity issues among variables in the same matrix, it preserves Euclidian distance instead of Mahalanobis distance, and the number of species does not have to be less than the number of sampling sites (Legendre and Legendre, 2012). This technique is also based on covariance and not on correlation (Paliy and Shankar, 2016). However, it was only performed in Latli et al. (2017) on principal component axes of environmental and faunal variables.

STATICO, a method that combines Partial Triadic Analysis (PTA, Tucker, 1966) and CoIA, is used to study the dynamics of the relationship between environmental parameters and biological communities by analysing sequences of paired ecological tables (Thioulouse et al., 2004). It was used only in Mazzocchi et al. (2012) to investigate stable patterns and interannual changes in the relationships between copepods and their environment.

3.7.3. Statistical tests

Correlation tests were the most frequently used data analysis method after linear modelling (n = 27), and were mainly used in combination with other techniques (n = 17, 63.0 %). They were mostly performed as Pearson correlation tests (n = 17, 63.0 %), which measures the intensity of the linear relationship between two random variables and does not assume any functional or explanatory response or causal link between them (Legendre and Legendre, 2012). Thus, unlike linear modelling, the two variables play the same role and can be swapped. Spearman's rank order correlation (Spearman, 1907), a non-parametric correlation test based on ranks, was used in 8 studies to measure the strength of non-linear monotonic relationships. Two other correlation methods were used: cross-correlation functions (CCF) to identify the time lag for the predictor that maximises the correlation in Gröger and Rumohr (2006), Licandro et al. (2012), and Szentkirályi et al. (2007), and partial correlation in van der Wal and Stien (2014) where it was used to analyse the correlation between plant biomass and weather parameters (cloud cover and rainfall) after controlling for the effect of temperature. We were not able to determine which correlation technique was used in two articles.

Only three other statistical tests were very occasionally used: the BIO-ENV procedure, PERMANOVA, and Student's t-test. The BIO-ENV procedure (Clarke and Ainsworth, 1993) is a technique that aims to select relevant explanatory variables by performing correlation tests between a dissimilarity matrix derived from a species-specific metric (e.g. abundance) and several dissimilarity abiotic matrices. It was used in 8 articles and six times with no other method. Permutational multivariate analysis of variation (PERMANOVA, Anderson, 2001), a non-parametric method used to perform multivariate ANOVA and test differences between object classes, was used in Gutiérrez-Fonseca et al. (2018) to assess differences in abiotic parameters between groups after non-metric multidimensional scaling. Student's t-test was performed in Lawrence et al. (2010) to detect differences in biotic integrity between two categories of climate parameters.

3.7.4. Other data analysis methods

The five remaining categories of data analysis techniques are found in only 7 articles. Although most of the models reported in the studies can be depicted as statistical models, three articles used mechanistic models. Unlike empirical models, mechanistic models, which are considered to be both realistic and general, are based on real cause-effect relationships, but their predictive power is often lower (Guisan and Zimmermann, 2000). Poggiale and Dauvin (2001) used a discrete population dynamics model that included sea temperature, competition and environmental pollution, calibrated by minimizing the distance between simulated and observed data to estimate the different parameters (e.g. carrying capacity, optimal growth temperature). A similar method was used by Mutshinda et al. (2017) and Kerimoglu et al. (2013).

The Self-Organizing Map (SOM, Kohonen, 1982) is an unsupervised learning algorithm of the Artificial Neural Network (ANN) that identifies clusters and maps high-dimensional data into a two-dimensional representational space. Every input data item selects the best matching model, each one associated with a 'neuron' that is represented on the two-dimensional grid of the SOM, with similar models associated with closer neurons (Kohonen, 2013). In Choi et al. (2015), both environmental and community data were used as input variables, while Penczak (2011) only entered biotic variables and then performed a MANOVA on environmental variables using the SOM clusters as categorical explanatory variable.

Classification and Regression Trees (CART) is a model-based tree classifier that explains the variation of a response variable with one or more predictors by splitting the data into nodes that best distinguish between samples (Ayón and Swartzman, 2008; De'ath and Fabricius, 2000). This technique is suitable for the analysis of complex ecological data and has many strengths: it can deal with non-linear relationships, missing values and categorical or quantitative variables. Nevertheless, this method is univariate, unlike Multivariate Regression Trees (MRT), an extension of CART to multivariate response data (De'ath, 2002; Larsen and Speckman, 2004). Ayón and Swartzman (2008) used this method to determine the parameter with the strongest effect on zooplankton biovolume.

Asymmetric Eigenvector Maps (AEM, Blanchet et al., 2008) is a spatial eigenfunction analysis, a family of methods for multiscale analysis where eigenvectors of spatial configuration matrices are calculated and then used as predictors. AEM specifically considers asymmetric directional physical processes in order to model multivariate spatial distributions and can be extended to time-series analysis (Legendre and Gauthier, 2014). For example, Lévesque et al. (2017) performed AEM to model temporal structure and used it as an explanatory variable of zooplankton taxa abundance.

Finally, wavelet analysis is a time-series analysis method that has many advantages over other similar techniques (e.g. spectral analysis), including robustness to missing values and non-necessity of stationarity of the time-series (Torrence and Compo, 1998). It was performed in Menge et al. (2011) by means of cross-wavelet and wavelet coherence to investigate respectively the covariance and the correlation between recruitment of barnacles and mussels and each environmental variable in a temporal framework.

4. Discussion

4.1. Trends in study characteristics and study design

Our review documented a wide variety of study designs, methods of data analysis and modelling in long-term studies, with an increasing number of studies since the mid-1990s. While this increase may partly be due to difficulty finding older articles, the dominant duration of studies, i.e. up to twenty years, and the average publication date (2010) suggest that many of the studies were initiated in the early 1990s in the context of a growing awareness of biodiversity loss (e.g. the Rio de Janeiro Earth Summit in 1992). More recently, the effect of climate change on biodiversity, which has been the subject of a growing body of published literature (Chapman et al., 2014; Jaeschke et al., 2014), may also contribute to the increasing number of long-term studies.

The fact that 75% of the articles resulting from our search concerned aquatic ecosystems, especially marine systems, was surprising, because none of the key-words in our search was directly linked to this type of habitat or to the biological communities that live in it. By contrast, Jaeschke et al. (2014) reviewed studies on the impact of climate change on organisms and ecosystems published between 2003 and 2013 and found that 44% of the studies focused on aquatic ecosystems. This suggests that aquatic habitats are either the subject of long-term studies, older long-term studies, or studies that focus on parameters not necessarily linked to climate change. Due to the focus on aquatic habitats in our review, only taxa and groups associated with these habitats were widely investigated, taxa linked mainly to terrestrial habitats being underrepresented.

As in Jaeschke et al. (2014), South America, Asia and Africa are underrepresented in the reviewed studies compared to Europe and North America. Jaeschke et al. (2014) demonstrated a positive correlation between gross domestic product and the number of studies per country. Thus, the low number of studies carried out in South America, Asia and Africa might be due to financial reasons, especially as funding is a major challenge in long-term ecological monitoring (Strayer et al., 1986). We expected more articles to be associated with LTER and similar networks, and the relatively low number may be due to our selection criteria.

The use of community indices is less than we expected, as these have many advantages, such as the synthesis of information. However, they also have a number of drawbacks, notably regarding the use of cardinal indices (e.g. Shannon index), which are the most commonly used diversity indices. For instance, all individuals of the same taxon are considered equal (e.g. their body size is not taken into account) and all taxa are assumed to be equally different (Cousins, 1991; Peet, 1974). Researchers working with these widely used indices are also confronted with the problem of knowing which index to use. This in turn raises another equally important question: What is meant by ‘diversity’? This is a critical issue, because diversity may refer to different concepts, namely species richness, equitability and heterogeneity (Peet, 1974), and the absence of a clear-cut definition has led to the development of a large number of ‘diversity’ indices (Hurlbert, 1971). The choice of the appropriate index must also be based on the community to be studied (e.g. the taxa involved) and on knowledge about it (i.e. species richness known or unknown) (Pielou, 1966). At last, specific characteristics of similar indices should be compared. For instance, the Shannon index is more sensitive to rare species than the Simpson index (Peet, 1974). Finally, diversity indices carry specific recommendations; for example, they should only be used for members of a single *taxocene*, in other words, taxa that "are likely to be of about the same size, to have similar life histories, and compete over both evolutionary and ecological time" (Deevey, 1969).

This review documented a wide range of sampling strategies, especially with regard to duration, frequency and number of study sites. This finding is in accordance with Magurran et al. (2010), who observed that long-term ecological studies show considerable variation in their sampling design and that this can be attributed to several factors such as the variety of study aims or the life cycle duration of the monitored taxa. The duration of the reviewed studies (median: 20 years) is longer than that found by Jackson and Fureder (2006) (median: 9 years); this difference could be due to their focus on freshwater macroinvertebrates whose life cycles are shorter than those of many taxa studied in the articles in our review. We also assume that years with no data were not always indicated, which may have led to an overestimation of the study's duration. As in Jaeschke et al. (2014), we observed more field observation than experimental studies, which could be explained by our selection criterion of studies analysing the effects of two or more abiotic parameters on two or more taxa.

4.2. Data analysis methods

Despite the fact that our review focused on long-term ecological studies, few data analysis techniques accounted for temporal correlation in the response data. Therefore, most of the tools described in the studies could also be used with short-term studies. Although we selected studies involving multiple taxa and abiotic parameters, most of the methods used were univariate and not multivariate. This might be due to the fact that community metrics were exclusively analysed using univariate techniques. Moreover, these methods were often used on species-specific metrics applied to groups of species (e.g. total abundance) or when only a few taxa were studies. It may also be because multivariate analyses can be more difficult to understand, perform and interpret, and might be computationally demanding (Paliy and Shankar, 2016). One could argue that the application of two or more methods, which occurred in almost half of the reviewed studies, indicates a lack of accuracy in the study design and its underlying rationale. However, in general, both univariate and multivariate techniques were performed in such situations, in order (1) to analyse the effects of environmental parameters at both community and species level (e.g. constrained ordination was often used in combination with statistical modelling), or (2) to produce synthetic variables before performing the analysis itself (e.g. unconstrained ordination was often performed before statistical modelling and correlation tests). The combination of several data analysis techniques may highlight their complementarity rather than methodological weakness. The preponderance of linear models and correlation tests compared to other univariate methods may be due to the flexibility of linear modelling and the simplicity of both techniques compared to more complex and recent procedures (e.g. Maximum Entropy modelling, Additive modelling). By contrast, the frequency of use of multivariate methods, especially ordination techniques, is more balanced between the techniques, because they appear to be less flexible and more specialised (e.g. techniques suited to short versus long gradients, symmetric versus asymmetric methods, direct versus indirect gradient analysis).

Most reported models belong to the statistical type, with mechanistic models only observed in three articles, and no analytical model. The latter focuses on precision and generality and is thus designed to be used within a limited or simplified reality (Guisan and Zimmermann, 2000). The absence of this type of model could be explained by our selection criteria (e.g. at least two abiotic factors), together with the overwhelming majority of observational studies in the reviewed articles involving many ecological phenomena. Despite multiple criticisms of frequentist methods in the literature (e.g. Beninger et al., 2012; Dorazio, 2016; Garamszegi et al., 2009; Hobbs and Hilborn, 2006; Stephens et al., 2007), analyses performed in a Bayesian framework were very rare. We also observed only a modest use of GLM compared to ordinary linear models. In many cases, the specification of a non-Gaussian distribution (e.g. Poisson, Binomial) is not relevant for most of the community metrics used in the reviewed studies, and linear modelling is therefore more appropriate. Nevertheless, we suspect that GLM (e.g. Poisson regression) was sometimes called linear modelling and sometimes linear regression. The infrequent use of modelling methods other than linear modelling can be explained both by their specificity (e.g. Maxent for presence only data, MANOVA for more than one response variable and a grouping explanatory factor) and by the lack of knowledge. Nevertheless, because these methods are very specific and can be used when linear modelling is not possible, researchers should know about them and when they can be used.

Regarding ordination techniques, PCA was used more than CA, because the former was performed on both biotic and abiotic variables, whereas the latter was used only on biological data, due to the more restricted input data. Legendre and Legendre (2012) suggested that CoIA is used more than CCorA because of its flexibility, but we found CoIA in only one article and CCorA in five. The fact that CoIA is currently used less in ecology than CCorA could be due to its relative novelty. Asymmetric constrained ordination methods were used much more frequently than symmetric ones. This is consistent with the fact that (1) many ecological issues are asymmetric (Legendre and Legendre, 2012), and (2) we only selected studies analysing the effects of environmental parameters on biological communities, i.e. asymmetric topics. However, this preferential use of asymmetric methods is much more marked for constrained ordination methods than for univariate techniques. We assume that this difference is due to the application of correlation tests instead of linear models to study asymmetric topics. This can be appropriate when not only the response variable, but also the predictors, are random (Legendre and Legendre, 2012), even if model II regression is a more generally accepted alternative (Laws and Archie, 1981). The ordination methods used in the reviewed studies are not ‘model-based’ but ‘algorithm-based’ (Hui et al., 2015; Warton et al., 2015), no doubt because the development of multivariate model-based approaches is extremely recent.

Our findings on the relative frequency of multivariate techniques are broadly similar to those of previous studies. For instance, James and McCulloch (1990), who summarised and reviewed the use of multivariate techniques in ecology and systematics, also found that PCA was the most frequently used ordination method and that linear models were widely used. However, they highlighted many differences in the use of analysis methods. In their literature review, asymmetric constrained analyses (i.e. CCA and RDA) were not reported, which is not surprising because these are recent methods that had only been developed a few years previously. Similarly, Ramette (2007) found that exploratory methods performed by means of PCA and cluster analysis were used much more often than interpretive methods.

Most of the methods used to analyse multidimensional ecological data sets mentioned in Legendre and Legendre (2012) were encountered in our study selection. However, we did not find a number of other methods that have been described in similar reviews (i.e. James and McCulloch, 1990; Paliy and Shankar, 2016; Parker and Arnold, 1999; Ramette, 2007) (Table 8), although their use might be appropriate. Descriptions of these methods, many of which have been developed recently, are provided in similar reviews cited above, in particular in Paliy and Shankar (2016).

Method	Acronym	Similar to
Principal Coordinates Analysis	PCoA	nMDS
Factor Analysis	FA	PCA
Hierarchical Clustering Analysis	HCA	SOM
Random Forest	RF	SOM
Orthogonal Projections to Latent Structures Discriminant Analysis	OPLS-DA	MDA
Support Vector Machine	SVM	MDA
Procrustes Analysis	PA	CoIA, CCorA
Mantel test	-	CoIA, CCorA
Principal Response Curves	PRC	RDA
Analysis of Similarity	ANOSIM	MANOVA

Table 8: Methods that were not found in the reviewed studies but mentioned in similar reviews.

4.3. Overview of prospective data analysis techniques

The decision tree (Fig. 6) could provide a rough guide to groups of potential approaches, while descriptions of methods and examples of uses given in the Results section should help researchers to differentiate between techniques. However, the decision tree only introduces potential tools based on broad features and does not assess their relevance, notably for long-term ecological studies. Techniques shown in Table 8 are also suggested, in addition to methods found in the reviewed articles. Paliy and Shankar (2016, Fig. 8) provided a decision table that could also help select multivariate techniques.

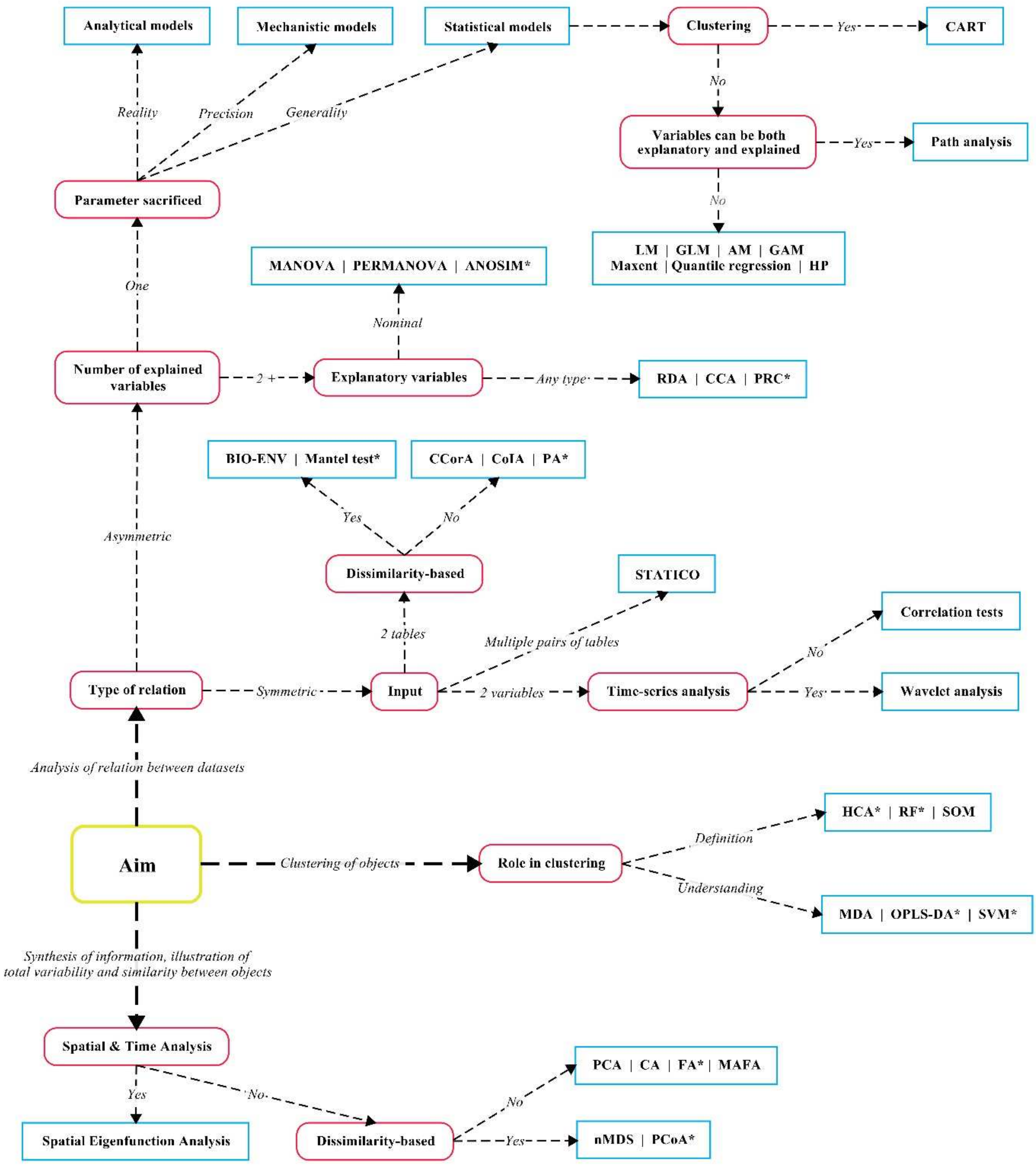

Fig. 6 (2-column): General characteristics of methods encountered in the reviewed studies, and potential methods described in similar reviews (*). Methods are indicated in blue and decision keys in red. Alternative methods (e.g. partial Redundancy Analysis) have not been included to avoid overloading the decision tree. For full names of methods, see Table 6, Table 8 and Results section.

4.4. Limitations of this study

The main purpose of this study is to provide an overview of the study design features and data analysis techniques that are currently used in long-term ecological studies. Its purpose is not to assess whether the features and analyses used in each study are the most suitable or whether the techniques' assumptions are satisfied. Within the framework of this review, only a few general indications can be given (e.g. Fig. 6). Techniques that are common in ecological studies are frequently misused; for example, (1) misinterpretation of p-values and overlooking statistical power in the frequentist approach and null hypothesis testing (Beninger et al., 2012), (2) temporal autocorrelation in time-series, i.e. observations that are closer in time are more similar (or less in the case of negative temporal autocorrelation) than observations paired at random (Zuur et al., 2009), (3) model specifications and validation. Essential assumptions (e.g. for linear modelling: linearity in parameters, predictors not correlated with the error term, non-collinearity between predictors, non-auto-correlation in residuals, homoscedasticity and normality in distribution of residuals) are frequently violated or not checked (Boldina and Beninger, 2016). Alternative or better fitting methods often exist (e.g. multivariate techniques specifically designed for time-series, such as AEM, PRC and STATICO, Bayesian methods as alternative frameworks for frequentist techniques, likelihood and information theoretic approaches instead of null hypothesis testing). Readers who are considering an analysis technique based on our decision tree (Fig. 6) are strongly advised to refer to specialized literature on the technique, to look carefully at its underlying assumptions, and to consider possible alternatives.

Acknowledgements

We would like to thank Céline Kowalczyk for her helpful comments and grammar correction. We also thank the two anonymous reviewers for their comments and suggestions on the manuscript, and Elizabeth Yates for linguistic improvements. This article was written as part of the BPO Loire project, funded by the Région Centre-Val de Loire. The authors' laboratory is supported by the French National Centre for Scientific Research (CNRS) and the University of Tours. This study was carried out within the framework of the Loire LTSER and the observation network "biodiversity of the Loire hydrosystem" (OBLA).

References

Abonyi, A., Ács, É., Hidas, A., Grigorszky, I., Várbíró, G., Borics, G., Kiss, K.T., 2018. Functional diversity of phytoplankton highlights long-term gradual regime shift in the middle section of the Danube River due to global warming, human impacts and oligotrophication. Freshwater. Biol. 63, 456–472. https://doi.org/10.1111/fwb.13084

Aleksandrov, S.V., Zhigalova, N.N., Zezera, A.S., 2009. Long-term dynamics of zooplankton in the southeastern Baltic Sea. Russ. J. Mar. Biol. 35, 296–304. https://doi.org/10.1134/S106307400904004X

Anderson, M.J., 2001. A new method for non-parametric multivariate analysis of variance. Austral. Ecol. 26, 32–46. https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x

Aßmus, J., Melle, W., Tjøstheim, D., Edwards, M., 2009. Seasonal cycles and long-term trends of plankton in shelf and oceanic habitats of the Norwegian Sea in relation to environmental variables. Deep. Sea. Res. Part. 2. Top. Stud. Oceanogr. 56, 1895–1909. https://doi.org/10.1016/j.dsr2.2008.11.004

Ayón, P., Swartzman, G., 2008. Changes in the long-term distribution of zooplankton in the Humboldt Current Ecosystem off Peru, 1961-2005, and its relationship to regime shifts and environmental factors. Fish. Oceanogr. 17, 421–431. https://doi.org/10.1111/j.1365-2419.2008.00488.x

Barrio Froján, C.R.S., Boyd, S.E., Cooper, K.M., Eggleton, J.D., Ware, S., 2008. Long-term benthic responses to sustained disturbance by aggregate extraction in an area off the east coast of the United Kingdom. Estuar. Coast. Shelf. S. 79, 204–212. https://doi.org/10.1016/j.ecss.2008.03.023

Bateman, B.L., Pidgeon, A.M., Radeloff, V.C., Flather, C.H., VanDerWal, J., Akçakaya, H.R., Thogmartin, W.E., Albright, T.P., Vavrus, S.J., Heglund, P.J., 2016. Potential breeding distributions of U.S. birds predicted with both short-term variability and long-term average climate data. Ecol. Appl. 26, 2720–2731. https://doi.org/10.1002/eap.1416

Beaugrand, G., Ibañez, F., Reid, P., 2000. Spatial, seasonal and long-term fluctuations of plankton in relation to hydroclimatic features in the English Channel, Celtic Sea and Bay of Biscay. Mar. Ecol. Prog. Ser. 200, 93–102. https://doi.org/10.3354/meps200093

Bêche, L.A., Resh, V.H., 2007. Biological traits of benthic macroinvertebrates in California mediterranean-climate streams: long-term annual variability and trait diversity patterns. Fund. Appl. Limnol. 169, 1–23. https://doi.org/10.1127/1863-9135/2007/0169-0001

Beninger, P.G., Boldina, I., Katsanevakis, S., 2012. Strengthening statistical usage in marine ecology. J. Exp. Mar. Biol. Ecol. 426, 97–108. https://doi.org/10.1016/j.jembe.2012.05.020

Benzecri, J.-P., 1969. Statistical analysis as a tool to make patterns emerge from data, in: Watanabe, S. (Ed.), Methodologies of Pattern Recognition. Academic Press, New York, pp. 35–74.

Beuchel, F., Gulliksen, B., Carroll, M.L., 2006. Long-term patterns of rocky bottom macrobenthic community structure in an Arctic fjord (Kongsfjorden, Svalbard) in relation to climate variability (1980–2003). J. Marine. Syst. 63, 35–48. https://doi.org/10.1016/j.jmarsys.2006.05.002

Blanchet, F.G., Legendre, P., Borcard, D., 2008. Modelling directional spatial processes in ecological data. Ecol. Model. 215, 325–336. https://doi.org/10.1016/j.ecolmodel.2008.04.001

Boldina, I., Beninger, P.G., 2016. Strengthening statistical usage in marine ecology: Linear regression. J. Exp. Mar. Biol. Ecol. 474, 81–91. https://doi.org/10.1016/j.jembe.2015.09.010

Borcard, D., Legendre, P., Drapeau, P., 1992. Partialling out the spatial component of ecological variation. Ecology. 73, 1045–1055. https://doi.org/10.2307/1940179

Bortolini, J.C., Bovo-Scomparin, V.M., Paula, A.C.M. de, Moresco, G.A., Reis, L.M., Jati, S., Rodrigues, L.C., 2014. Composition and species richness phytoplankton in a subtropical floodplain lake: a long-term study. Acta. Limnol. Bras. 26, 296–305. https://doi.org/10.1590/S2179-975X2014000300009

Brooker, R.W., Matesanz, S., Valladares, F., Klotz, S., 2012. Long-term spatial pattern change in a semi-arid plant community: The role of climate and composition. Acta. Oecol. 45, 8–15. https://doi.org/10.1016/j.actao.2012.08.002

Brown, D., Edmunds, P.J., 2013. Long-term changes in the population dynamics of the Caribbean hydrocoral Millepora spp. J. Exp. Mar. Biol. Ecol. 441, 62–70. https://doi.org/10.1016/j.jembe.2013.01.013

Buttay, L., Miranda, A., Casas, G., González-quirós, R., Nogueira, E., 2016. Long-term and seasonal zooplankton dynamics in the northwest Iberian shelf and its relationship with meteo-climatic and hydrographic variability. J. Plankton. Res. 38, 106–121. https://doi.org/10.1093/plankt/fbv100

Buttigieg, P.L., Ramette, A., 2014. A guide to statistical analysis in microbial ecology: a community-focused, living review of multivariate data analyses. FEMS. Microbiol. Ecol. 90, 543–550. https://doi.org/10.1111/1574-6941.12437

Cade, B.S., Noon, B.R., 2003. A gentle introduction to quantile regression for ecologists. Front. Ecol. Environ. 1, 412–420. https://doi.org/10.1890/1540-9295(2003)001[0412:AGITQR]2.0.CO;2

Callahan, J.T., 1984. Long-term ecological research. BioScience. 34, 363–367. https://doi.org/10.2307/1309727

Carballo, J.L., Vega, C., Cruz-Barraza, J.A., Yáñez, B., Nava, H., Ávila, E., Wilson, M., 2008. Short- and long-term patterns of sponge diversity on a rocky tropical coast: evidence of large-scale structuring factors. Mar. Ecol. 29, 216–236. https://doi.org/10.1111/j.1439-0485.2008.00228.x

Casey, L.M., Rebelo, H., Rotheray, E., Goulson, D., 2015. Evidence for habitat and climatic specializations driving the long-term distribution trends of UK and Irish bumblebees. Divers. Distrib. 21, 864–875. https://doi.org/10.1111/ddi.12344

Chapman, S., Mustin, K., Renwick, A.R., Segan, D.B., Hole, D.G., Pearson, R.G., Watson, J.E.M., 2014. Publishing trends on climate change vulnerability in the conservation literature reveal a predominant focus on direct impacts and long time-scales. Divers. Distrib. 20, 1221–1228. https://doi.org/10.1111/ddi.12234

Chevan, A., Sutherland, M., 1991. Hierarchical partitioning. Am. Stat. 45, 90. https://doi.org/10.2307/2684366

Chevenet, F., Dolédec, S., Chessel, D., 1994. A fuzzy coding approach for the analysis of long-term ecological data. Freshwater. Biol. 31, 295–309. https://doi.org/10.1111/j.1365-2427.1994.tb01742.x

Childers, D.L., Iwaniec, D., Rondeau, D., Rubio, G., Verdon, E., Madden, C.J., 2006. Responses of sawgrass and spikerush to variation in hydrologic drivers and salinity in Southern Everglades marshes. Hydrobiologia. 569, 273–292. https://doi.org/10.1007/s10750-006-0137-9

Choi, J.-Y., Jeong, K.-S., Joo, G.-J., 2015. Rainfall as dominant driver of rotifer dynamics in shallow wetlands: Evidence from a long-term data record (Upo Wetlands, South Korea). Int. Rev. Hydrobiol. 100, 21–33. https://doi.org/10.1002/iroh.201401745

Clarke, K., Ainsworth, M., 1993. A method of linking multivariate community structure to environmental variables. Mar. Ecol. Prog. Ser. 92, 205–219. https://doi.org/10.3354/meps092205

Clotfelter, E.D., Pedersen, A.B., Cranford, J.A., Ram, N., Snajdr, E.A., Nolan, V., Ketterson, E.D., 2007. Acorn mast drives long-term dynamics of rodent and songbird populations. Oecologia. 154, 493–503. https://doi.org/10.1007/s00442-007-0859-z

Cousins, S.H., 1991. Species diversity measurement: Choosing the right index. Trends. Ecol. Evol. 6, 190–192. https://doi.org/10.1016/0169-5347(91)90212-G

Crowley, P.H., 1992. Resampling methods for computation-intensive data analysis in ecology and evolution. Annu. Rev. Ecol. Syst. 23, 405–447. https://doi.org/10.1146/annurev.es.23.110192.002201

Daghighi, E., Koehler, H., Kesel, R., Filser, J., 2017. Long-term succession of Collembola communities in relation to climate change and vegetation. Pedobiologia. 64, 25–38. https://doi.org/10.1016/j.pedobi.2017.06.001

Dale, M.R.T., Fortin, M.-J., 2002. Spatial autocorrelation and statistical tests in ecology. Écoscience. 9, 162–167. https://doi.org/10.1080/11956860.2002.11682702

Daufresne, M., Roger, M.C., Capra, H., Lamouroux, N., 2004. Long-term changes within the invertebrate and fish communities of the Upper Rhone River: effects of climatic factors. Glob. Change. Biol. 10, 124–140. https://doi.org/10.1046/j.1529-8817.2003.00720.x

Davies, P.T., Tso, M.K.-S., 1982. Procedures for reduced-rank regression. Appl. Stat. J. Roy. St. C. 31, 244. https://doi.org/10.2307/2347998

De'ath, G., 2002. Multivariate regression trees: A new technique for modeling species-environment relationships. Ecology. 83, 1105–1117. https://doi.org/10.1890/0012-9658(2002)083[1105:MRTANT]2.0.CO;2

De'ath, G., Fabricius, K.E., 2000. Classification and regression trees: A powerful yet simple technique for ecological data analysis. Ecology. 81, 3178–3192. https://doi.org/10.1890/0012-9658(2000)081[3178:CARTAP]2.0.CO;2

Deevey, E.S., 1969. Coaxing history to conduct experiments. BioScience. 19, 40–43. https://doi.org/10.2307/1294629

Delcourt, H.R., Delcourt, P.A., 1988. Quaternary landscape ecology: Relevant scales in space and time. Landscape. Ecol. 2, 23–44. https://doi.org/10.1007/BF00138906

Dippner, J.W., Ikauniece, A., 2001. Long-term zoobenthos variability in the Gulf of Riga in relation to climate variability. J. Marine. Syst. 30, 155–164. https://doi.org/10.1016/S0924-7963(01)00055-0

Dippner, J.W., Kornilovs, G., Sidrevics, L., 2000. Long-term variability of mesozooplankton in the Central Baltic Sea. J. Marine. Syst. 25, 23–31. https://doi.org/10.1016/S0924-7963(00)00006-3

Dolédec, S., Chessel, D., 1987. Rythmes saisonniers et composantes stationnelles en milieu aquatique. I: Description d'un plan d'observation complet par projection de variables. Acta. Oecol. Oec. Gen. 8, 403–426.

Dolédec, S., Chessel, D., 1994. Co-inertia analysis: an alternative method for studying species-environment relationships. Freshwater. Biol. 31, 277–294. https://doi.org/10.1111/j.1365-2427.1994.tb01741.x

Dorazio, R.M., 2016. Bayesian data analysis in population ecology: Motivations, methods, and benefits. Popul. Ecol. 58, 31–44. https://doi.org/10.1007/s10144-015-0503-4

Einarsson, A., Björk Örnólfsdóttir, E., 2004. Long-term changes in benthic Cladocera populations in Lake Myvatn, Iceland. Aquat. Ecol. 38, 253–262. https://doi.org/10.1023/B:AECO.0000032060.29256.95

Ershova, E., Hopcroft, R., Kosobokova, K., Matsuno, K., Nelson, R.J., Yamaguchi, A., Eisner, L., 2015. Long-term changes in summer zooplankton communities of the Western Chukchi Sea, 1945–2012. Oceanography. 28, 100–115. https://doi.org/10.5670/oceanog.2015.60

Fasola, M., Rubolini, D., Merli, E., Boncompagni, E., Bressan, U., 2010. Long-term trends of heron and egret populations in Italy, and the effects of climate, human-induced mortality, and habitat on population dynamics. Popul. Ecol. 52, 59–72. https://doi.org/10.1007/s10144-009-0165-1

Fath, B.D., Müller, F., 2019. Conbiota, in: Fath, B.D. (Ed.), Encyclopedia of Ecology, second ed. Elsevier, pp. 274–280.

Feike, M., Heerkloss, R., Rieling, T., Schubert, H., 2007. Studies on the zooplankton community of a shallow lagoon of the Southern Baltic Sea: long-term trends, seasonal changes, and relations with physical and chemical parameters. Hydrobiologia. 577, 95–106. https://doi.org/10.1007/s10750-006-0420-9

Feio, M.J., Dolédec, S., Graça, M.A.S., 2015. Human disturbance affects the long-term spatial synchrony of freshwater invertebrate communities. Environ. Pollut. 196, 300–308. https://doi.org/10.1016/j.envpol.2014.09.026

Fisher, R.A., 1936. The use of multiple measurements in taxonomic problems. Ann. Eugenic. 7, 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x

Fisher, R.A., 1940. The precision of discriminant functions. Ann. Eugenic. 10, 422–429. https://doi.org/10.1111/j.1469-1809.1940.tb02264.x

Franklin, J.F., Bledsoe, C.S., Callahan, J.T., 1990. Contributions of the long-term ecological research program. BioScience. 40, 509–523. https://doi.org/10.2307/1311319

Garamszegi, L.Z., Calhim, S., Dochtermann, N., Hegyi, G., Hurd, P.L., Jørgensen, C., Kutsukake, N., Lajeunesse, M.J., Pollard, K.A., Schielzeth, H., Symonds, M.R.E., Nakagawa, S., 2009. Changing philosophies and tools for statistical inferences in behavioral ecology. Behav. Ecol. 20, 1363–1375. https://doi.org/10.1093/beheco/arp137

Garcia, A., Vieira, J., Winemiller, K., Moraes, L., Paes, E., 2012. Factoring scales of spatial and temporal variation in fish abundance in a subtropical estuary. Mar. Ecol. Prog. Ser. 461, 121–135. https://doi.org/10.3354/meps09798

Gauthier, G., Bety, J., Cadieux, M.-C., Legagneux, P., Doiron, M., Chevallier, C., Lai, S., Tarroux, A., Berteaux, D., 2013. Long-term monitoring at multiple trophic levels suggests heterogeneity in responses to climate change in the Canadian Arctic tundra. Philos. T. Roy. Soc. B. 368, 20120482. https://doi.org/10.1098/rstb.2012.0482

Grace, J.B., Schoolmaster, D.R., Guntenspergen, G.R., Little, A.M., Mitchell, B.R., Miller, K.M., Schweiger, E.W., 2012. Guidelines for a graph-theoretic implementation of structural equation modeling. Ecosphere. 3, art73. https://doi.org/10.1890/ES12-00048.1

Gröger, J., Rumohr, H., 2006. Modelling and forecasting long-term dynamics of Western Baltic macrobenthic fauna in relation to climate signals and environmental change. J. Sea. Res. 55, 266–277. https://doi.org/10.1016/j.seares.2005.11.005

Gross, K., Edmunds, P.J., 2015. Stability of Caribbean coral communities quantified by long-term monitoring and autoregression models. Ecology. 96, 1812–1822. https://doi.org/10.1890/14-0941.1

Gu, Z., Gu, L., Eils, R., Schlesner, M., Brors, B., 2014. Circlize implements and enhances circular visualization in R. Bioinformatics. 30, 2811–2812. https://doi.org/10.1093/bioinformatics/btu393

Guisan, A., Edwards, T.C., Hastie, T., 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecol. Model. 157, 89–100. https://doi.org/10.1016/S0304-3800(02)00204-1

Guisan, A., Weiss, S.B., Weiss, A.D., 1999. GLM versus CCA spatial modeling of plant species distribution. Plant. Ecol. 143, 107–122. https://doi.org/10.1023/A:1009841519580

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135, 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9

Gutierrez, M., Devercelli, M., Brucet, S., Lauridsen, T., Søndergaard, M., Jeppesen, E., 2016. Is recovery of large-bodied zooplankton after nutrient loading reduction hampered by climate warming? A long-term study of shallow hypertrophic Lake Søbygaard, Denmark. Water. 8, 341. https://doi.org/10.3390/w8080341

Gutiérrez-Fonseca, P.E., Ramírez, A., Pringle, C.M., 2018. Large-scale climatic phenomena drive fluctuations in macroinvertebrate assemblages in lowland tropical streams, Costa Rica: The importance of ENSO events in determining long-term (15y) patterns. PLOS ONE. 13, e0191781. https://doi.org/10.1371/journal.pone.0191781

Haase, P., Frenzel, M., Klotz, S., Musche, M., Stoll, S., 2016. The long-term ecological research (LTER) network: Relevance, current status, future perspective and examples from marine, freshwater and terrestrial long-term observation. Ecol. Indic. 65, 1–3. https://doi.org/10.1016/j.ecolind.2016.01.040

Hall, S.R., Rudstam, L.G., 1999. Habitat use and recruitment: A comparison of long-term recruitment patterns among fish species in a shallow eutrophic lake, Oneida Lake, NY, U.S.A., in: Walz, N., Nixdorf, B. (Eds.), Shallow Lakes ’98. Springer Netherlands, Dordrecht, pp. 101–113.

Hallett, L.M., Hsu, J.S., Cleland, E.E., Collins, S.L., Dickson, T.L., Farrer, E.C., Gherardi, L.A., Gross, K.L., Hobbs, R.J., Turnbull, L., Suding, K.N., 2014. Biotic mechanisms of community stability shift along a precipitation gradient. Ecology. 95, 1693–1700. https://doi.org/10.1890/13-0895.1

Haupt, K., Phelps, Q., 2016. Mesohabitat associations in the Mississippi River Basin: A long-term study on the catch rates and physical habitat associations of juvenile silver carp and two native planktivores. Aquat. Invasions. 11, 93–99. https://doi.org/10.3391/ai.2016.11.1.10

Henderson, P.A., Seaby, R.M.H., Somes, J.R., 2011. Community level response to climate change: The long-term study of the fish and crustacean community of the Bristol Channel. J. Exp. Mar. Biol. Ecol. 400, 78–89. https://doi.org/10.1016/j.jembe.2011.02.028

Hill, M.O., Gauch, H.G., 1980. Detrended correspondence analysis: An improved ordination technique, in: van der Maarel, E. (Ed.), Classification and ordination: Symposium on advances in vegetation science, Nijmegen, The Netherlands, May 1979. Springer Netherlands, Dordrecht, pp. 47–58.

Hirschfeld, H.O., Wishart, J., 1935. A connection between correlation and contingency. Math. Proc. Cambridge. 31, 520. https://doi.org/10.1017/S0305004100013517

Hobbs, N.T., Hilborn, R., 2006. Alternatives to statistical hypothesis testing in ecology: A guide to self teaching. Ecol. Appl. 16, 5–19. https://doi.org/10.1890/04-0645

Horn, H., Paul, L., Horn, W., Petzoldt, T., 2011. Long-term trends in the diatom composition of the spring bloom of a German reservoir: is Aulacoseira subarctica favoured by warm winters?. Freshwater. Biol. 56, 2483–2499. https://doi.org/10.1111/j.1365-2427.2011.02674.x

Hotelling, H., 1933. Analysis of a complex of statistical variables into principal components. Jpn. J. Educ. Psychol. 24, 417–441. https://doi.org/10.1037/h0071325

Hotelling, H., 1936. Relations between two sets of variates. Biometrika. 28, 321–377. https://doi.org/10.1093/biomet/28.3-4.321

Hui, F.K.C., Taskinen, S., Pledger, S., Foster, S.D., Warton, D.I., 2015. Model-based approaches to unconstrained ordination. Methods. Ecol. Evol. 6, 399–411. https://doi.org/10.1111/2041-210X.12236

Hurlbert, S.H., 1971. The nonconcept of species diversity: A critique and alternative parameters. Ecology. 52, 577–586. https://doi.org/10.2307/1934145

Hurlbert, S.H., 1984. Pseudoreplication and the design of ecological field experiments. Ecol. Monogr. 54, 187–211. https://doi.org/10.2307/1942661

Hurst, T.P., McKown, K.A., Conover, D.O., 2004. Interannual and long-term variation in the nearshore fish community of the mesohaline Hudson River estuary. Estuaries. 27, 659–669. https://doi.org/10.1007/BF02907651

Ibanez, F., Conversi, A., 2002. Prediction of missing values and detection of ‘exceptional events’ in a chronological planktonic series: A single algorithm. Ecol. Model. 154, 9–23. https://doi.org/10.1016/S0304-3800(02)00033-9

Irvine, K.M., Miller, S.W., Al-Chokhachy, R.K., Archer, E.K., Roper, B.B., Kershner, J.L., 2015. Empirical evaluation of the conceptual model underpinning a regional aquatic long-term monitoring program using causal modelling. Ecol. Indic. 50, 8–23. https://doi.org/10.1016/j.ecolind.2014.10.011

Jackson, J.K., Fureder, L., 2006. Long-term studies of freshwater macroinvertebrates: A review of the frequency, duration and ecological significance. Freshwater. Biol. 51, 591–603. https://doi.org/10.1111/j.1365-2427.2006.01503.x

Jaeschke, A., Bittner, T., Jentsch, A., Beierkuhnlein, C., 2014. The last decade in ecological climate change impact research: Where are we now? Naturwissenschaften. 101, 1–9. https://doi.org/10.1007/s00114-013-1132-4

Jahan, R., Choi, J.K., 2014. Climate regime shift and phytoplankton phenology in a macrotidal estuary: Long-term surveys in Gyeonggi Bay, Korea. Estuar. Coast. 37, 1169–1187. https://doi.org/10.1007/s12237-013-9760-7

James, F.C., McCulloch, C.E., 1990. Multivariate analysis in ecology and systematics: Panacea or Pandora's box? Annu. Rev. Ecol. Syst. 21, 129–166. https://doi.org/10.1146/annurev.es.21.110190.001021

James, N.C., Whitfield, A.K., Cowley, P.D., 2008. Long-term stability of the fish assemblages in a warm-temperate South African estuary. Estuar. Coast. Shelf. S. 76, 723–738. https://doi.org/10.1016/j.ecss.2007.07.036

Jaynes, E.T., 1957. Information theory and statistical mechanics. Phys. Rev. 106, 620–630. https://doi.org/10.1103/PhysRev.106.620

Jennions, M.D., 2003. A survey of the statistical power of research in behavioral ecology and animal behavior. Behav. Ecol. 14, 438–445. https://doi.org/10.1093/beheco/14.3.438

Jourdan, J., O'Hara, R.B., Bottarin, R., Huttunen, K.-L., Kuemmerlen, M., Monteith, D., Muotka, T., Ozoliņš, D., Paavola, R., Pilotto, F., Springe, G., Skuja, A., Sundermann, A., Tonkin, J.D., Haase, P., 2018. Effects of changing climate on European stream invertebrate communities: A long-term data analysis. Sci. Total. Environ. 621, 588–599. https://doi.org/10.1016/j.scitotenv.2017.11.242

Jucevica, E., Melecis, V., 2006. Global warming affect Collembola community: A long-term study. Pedobiologia. 50, 177–184. https://doi.org/10.1016/j.pedobi.2005.10.006

Kerimoglu, O., Straile, D., Peeters, F., 2013. Seasonal, inter-annual and long term variation in top-down versus bottom-up regulation of primary production. Oikos. 122, 223–234. https://doi.org/10.1111/j.1600-0706.2012.20603.x

Kimball, M.E., Eash-Loucks, W.E., Petrinec, K.M., 2014. Long-term changes in an estuarine mud crab community: Evaluating the impact of non-native species. J. Crustacean. Biol. 34, 731–738. https://doi.org/10.1163/1937240X-00002287

Kimmel, D., Roman, M., 2004. Long-term trends in mesozooplankton abundance in Chesapeake Bay, USA: Influence of freshwater input. Mar. Ecol. Prog. Ser. 267, 71–83. https://doi.org/10.3354/meps267071

Knapp, A.K., Smith, M.D., Hobbie, S.E., Collins, S.L., Fahey, T.J., Hansen, G.J.A., Landis, D.A., La Pierre, K.J., Melillo, J.M., Seastedt, T.R., Shaver, G.R., Webster, J.R., 2012. Past, present, and future roles of long-term experiments in the LTER network. BioScience. 62, 377–389. https://doi.org/10.1525/bio.2012.62.4.9

Kodama, K., Aoki, I., Shimizu, M., Taniuchi, T., 2002. Long-term changes in the assemblage of demersal fishes and invertebrates in relation to environmental variations in Tokyo Bay, Japan. Fisheries. Manag. Ecol. 9, 303–313. https://doi.org/10.1046/j.1365-2400.2002.00313.x

Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica. 46, 33. https://doi.org/10.2307/1913643

Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. Biol. Cybern. 43, 59–69. https://doi.org/10.1007/BF00337288

Kohonen, T., 2013. Essentials of the self-organizing map. Neural Networks. 37, 52–65. https://doi.org/10.1016/j.neunet.2012.09.018

Kröncke, I., Dippner, J., Heyen, H., Zeiss, B., 1998. Long-term changes in macrofaunal communities off Norderney (East Frisia, Germany) in relation to climate variability. Mar. Ecol. Prog. Ser. 167, 25–36. https://doi.org/10.3354/meps167025

Kruskal, J.B., 1964a. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika. 29, 1–27. https://doi.org/10.1007/BF02289565

Kruskal, J.B., 1964b. Nonmetric multidimensional scaling: A numerical method. Psychometrika. 29, 115–129. https://doi.org/10.1007/BF02289694

Kwok, A.B.C., Wardle, G.M., Greenville, A.C., Dickman, C.R., 2016. Long-term patterns of invertebrate abundance and relationships to environmental factors in arid Australia. Austral. Ecol. 41, 480–491. https://doi.org/10.1111/aec.12334

Ladin, Z.S., Higgins, C.D., Schmit, J.P., Sanders, G., Johnson, M.J., Weed, A.S., Marshall, M.R., Campbell, J.P., Comiskey, J.A., Shriver, W.G., 2016. Using regional bird community dynamics to evaluate ecological integrity within national parks. Ecosphere. 7, e01464. https://doi.org/10.1002/ecs2.1464

Ladwig, L.M., Ratajczak, Z.R., Ocheltree, T.W., Hafich, K.A., Churchill, A.C., Frey, S.J.K., Fuss, C.B., Kazanski, C.E., Muñoz, J.D., Petrie, M.D., Reinmann, A.B., Smith, J.G., 2016. Beyond arctic and alpine: The influence of winter climate on temperate ecosystems. Ecology. 97, 372–382. https://doi.org/10.1890/15-0153.1

Larsen, D.R., Speckman, P.L., 2004. Multivariate regression trees for analysis of abundance data. Biometrics. 60, 543–549. https://doi.org/10.1111/j.0006-341X.2004.00202.x

Latli, A., Descy, J.-P., Mondy, C.P., Floury, M., Viroux, L., Otjacques, W., Marescaux, J., Depiereux, E., Ovidio, M., Usseglio-Polatera, P., Kestemont, P., 2017. Long-term trends in trait structure of riverine communities facing predation risk increase and trophic resource decline. Ecol. Appl. 27, 2458–2474. https://doi.org/10.1002/eap.1621

Laurance, S.G.W., Laurance, W.F., Nascimento, H.E.M., Andrade, A., Fearnside, P.M., Rebello, E.R.G., Condit, R., 2009. Long-term variation in Amazon forest dynamics. J. Veg. Sci. 20, 323–333. https://doi.org/10.1111/j.1654-1103.2009.01044.x

Lavaniegos, B.E., Ohman, M.D., 2003. Long-term changes in pelagic tunicates of the California Current. Deep. Sea. Res. Part. 2. Top. Stud. Oceanogr. 50, 2473–2498. https://doi.org/10.1016/S0967-0645(03)00132-2

Lawrence, J.E., Lunde, K.B., Mazor, R.D., Bêche, L.A., McElravy, E.P., Resh, V.H., 2010. Long-term macroinvertebrate responses to climate change: implications for biological assessment in mediterranean-climate streams. J. N. Am. Benthol. Soc. 29, 1424–1440. https://doi.org/10.1899/09-178.1

Laws, E.A., Archie, J.W., 1981. Appropriate use of regression analysis in marine biology. Mar. Biol. 65, 13–16. https://doi.org/10.1007/BF00397062

Legendre, P., Anderson, M.J., 1999. Distance-based redundancy analysis: Testing multispecies responses in multifactorial ecological experiments. Ecol. Monogr. 69, 1–24. https://doi.org/10.1890/0012-9615(1999)069[0001:DBRATM]2.0.CO;2

Legendre, P., Gauthier, O., 2014. Statistical methods for temporal and space-time analysis of community composition data. P. R. Soc. B. 281, 20132728. https://doi.org/10.1098/rspb.2013.2728

Legendre, P., Legendre, L., 2012. Numerical ecology, third English edition. Elsevier, Amsterdam.

Lévesque, D., Pinel-Alloul, B., Méthot, G., Steedman, R., 2017. Effects of climate, limnological features and watershed clearcut logging on long-term variation in zooplankton communities of Boreal Shield lakes. Water. 9, 733. https://doi.org/10.3390/w9100733

Levins, R., 1966. The strategy of model building in population biology. Am. Sci. 54, 421–431.

Li, Z., Ma, W., Liang, C., Liu, Z., Wang, W., Wang, L., 2015. Long-term vegetation dynamics driven by climatic variations in the Inner Mongolia grassland: Findings from 30-year monitoring. Landscape. Ecol. 30, 1701–1711. https://doi.org/10.1007/s10980-014-0068-1

Licandro, P., Souissi, S., Ibanez, F., Carré, C., 2012. Long-term variability and environmental preferences of calycophoran siphonophores in the Bay of Villefranche (north-western Mediterranean). Progr. Oceanogr. 97, 152–163. https://doi.org/10.1016/j.pocean.2011.11.004

Likens, G.E., 1989. Long-term studies in ecology: Approaches and alternatives. Springer, New York.

Lindenmayer, D.B., Likens, G.E., Andersen, A., Bowman, D., Bull, C.M., Burns, E., Dickman, C.R., Hoffmann, A.A., Keith, D.A., Liddell, M.J., Lowe, A.J., Metcalfe, D.J., Phinn, S.R., Russell-Smith, J., Thurgate, N., Wardle, G.M., 2012. Value of long-term ecological studies. Austral. Ecol. 37, 745–757. https://doi.org/10.1111/j.1442-9993.2011.02351.x

MacKenzie, D.I., Kendall, W.L., 2002. How should detection probability be incorporated into estimates of relative abundance? Ecology. 83, 2387–2393. https://doi.org/10.1890/0012-9658(2002)083[2387:HSDPBI]2.0.CO;2

Magnuson, J.J., 1990. Long-term ecological research and the invisible present. BioScience. 40, 495–501. https://doi.org/10.2307/1311317

Magurran, A.E., Baillie, S.R., Buckland, S.T., Dick, J.M., Elston, D.A., Scott, E.M., Smith, R.I., Somerfield, P.J., Watt, A.D., 2010. Long-term datasets in biodiversity research and monitoring: Assessing change in ecological communities through time. Trends. Ecol. Evol. 25, 574–582. https://doi.org/10.1016/j.tree.2010.06.016

Marchant, R., Dean, J., 2014. A long-term study of the factors that influence compositional stability of stream invertebrates. Inland Waters. 4, 113–120. https://doi.org/10.5268/IW-4.2.658

Mazzocchi, M.G., Dubroca, L., García-Comas, C., Capua, I.D., d'Alcalà, M.R., 2012. Stability and resilience in coastal copepod assemblages: The case of the Mediterranean long-term ecological research at Station MC (LTER-MC). Progr. Oceanogr. 97–100, 135–151. https://doi.org/10.1016/j.pocean.2011.11.003

Menge, B.A., Gouhier, T.C., Freidenburg, T., Lubchenco, J., 2011. Linking long-term, large-scale climatic and environmental variability to patterns of marine invertebrate recruitment: Toward explaining "unexplained" variation. J. Exp. Mar. Biol. Ecol. 400, 236–249. https://doi.org/10.1016/j.jembe.2011.02.003

Molinero, J.C., Anneville, O., Souissi, S., Balvay, G., Gerdeaux, D., 2006. Anthropogenic and climate forcing on the long-term changes of planktonic rotifers in Lake Geneva, Europe. J. Plankton. Res. 28, 287–296. https://doi.org/10.1093/plankt/fbi110

Möllmann, C., Kornilovs, G., Sidrevics, L., 2000. Long-term dynamics of main mesozooplankton species in the central Baltic Sea. J. Plankton. Res. 22, 2015–2038. https://doi.org/10.1093/plankt/22.11.2015
Möllmann, C., Köster, F., Kornilovs, G., Sidrevics, L., 2002. Long-term trends in abundance of cladocerans in the Central Baltic Sea. Mar. Biol. 141, 343–352. https://doi.org/10.1007/s00227-002-0822-9
Morrison, M.L. (Ed.), 2010. Wildlife study design, second ed. Springer, New York.
Mukhopadhyay, N., Banerjee, S., 2015. Sequential negative binomial problems and statistical ecology: A selected review with new directions. Stat. Methodol. 26, 34–60. https://doi.org/10.1016/j.stamet.2015.02.006
Mutshinda, C., Finkel, Z.V., Widdicombe, C.E., Irwin, A.J., 2017. Phytoplankton traits from long-term oceanographic time-series. Mar. Ecol. Prog. Ser. 576, 11-25. https://doi.org/10.1101/148304
Obaza, A., Hoffman, R., Clausing, R., 2015. Long-term stability of eelgrass fish assemblages in two highly developed coastal estuaries. Fisheries. Manag. Ecol. 22, 224–238. https://doi.org/10.1111/fme.12119
Pace, G., Bonada, N., Prat, N., 2013. Long-term effects of climatic-hydrological drivers on macroinvertebrate richness and composition in two Mediterranean streams. Freshwater. Biol. 58, 1313–1328. https://doi.org/10.1111/fwb.12129
Paliy, O., Shankar, V., 2016. Application of multivariate statistical techniques in microbial ecology. Mol. Ecol. 25, 1032–1057. https://doi.org/10.1111/mec.13536
Parker, W.C., Arnold, A.J., 1999. Quantitative methods of data analysis in foraminiferal ecology, in: Modern Foraminifera. Springer Netherlands, Dordrecht, pp. 71–89. https://doi.org/10.1007/0-306-48104-9_5
Pearson, K., 1901. LIII. On lines and planes of closest fit to systems of points in space. Philos. Mag. 2, 559–572. https://doi.org/10.1080/14786440109462720
Peet, R.K., 1974. The measurement of species diversity. Annu. Rev. Ecol. Syst. 5, 285–307. https://doi.org/10.1146/annurev.es.05.110174.001441
Penczak, T., 2011. Fish assemblages composition in a natural, then regulated, stream: A quantitative long-term study. Ecol. Model. 222, 2103–2118. https://doi.org/10.1016/j.ecolmodel.2011.03.032
Pfister, C.A., 2006. Concordance between short-term experiments and long-term censuses in tide pool fishes. Ecology. 87, 2905–2914. https://doi.org/10.1890/0012-9658(2006)87[2905:CBSEAL]2.0.CO;2
Phillips, S.J., Anderson, R.P., Schapire, R.E., 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190, 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026
Pielou, E.C., 1966. The measurement of diversity in different types of biological collections. J. Theor. Biol. 13, 131–144. https://doi.org/10.1016/0022-5193(66)90013-0
Pitacco, V., Mistri, M., Munari, C., 2018. Long-term variability of macrobenthic community in a shallow coastal lagoon (Valli di Comacchio, northern Adriatic): Is community resistant to climate changes? Mar. Environ. Res. 137, 73–87. https://doi.org/10.1016/j.marenvres.2018.02.026
Poggiale, J., Dauvin, J., 2001. Long-term dynamics of three benthic Ampelisca (Crustacea-Amphipoda) populations from the Bay of Morlaix (western English Channel) related to

their disappearance after the “Amoco Cadiz” oil spill. Mar. Ecol. Prog. Ser. 214, 201–209. https://doi.org/10.3354/meps214201

Ramette, A., 2007. Multivariate analyses in microbial ecology. FEMS. Microbiol. Ecol. 62, 142–160. https://doi.org/10.1111/j.1574-6941.2007.00375.x

Ramos, M.R., Oliveira, M.M., Borges, J.G., McDill, M.E., 2015. Statistical models for categorical data: Brief review for applications in ecology. AIP. Conf. Proc. 1648, 840015. https://doi.org/10.1063/1.4913055

Rao, C.R., 1948. The utilization of multiple measurements in problems of biological classification. J. Roy. Stat. Soc. B. Met. 10, 159–203.

Rao, C.R., 1964. The use and interpretation of principal component analysis in applied research. Sankhya. Ser. A. 26, 329–358.

Ribic, C.A., Ainley, D.G., Glenn Ford, R., Fraser, W.R., Tynan, C.T., Woehler, E.J., 2011. Water masses, ocean fronts, and the structure of Antarctic seabird communities: Putting the eastern Bellingshausen Sea in perspective. Deep. Sea. Res. Part. 2. Top. Stud. Oceanogr. 58, 1695–1709. https://doi.org/10.1016/j.dsr2.2009.09.017

Royle, J.A., 2004. N-mixture models for estimating population size from spatially replicated counts. Biometrics. 60, 108–115. https://doi.org/10.1111/j.0006-341X.2004.00142.x

Santoro, S., Sanchez-Suarez, C., Rouco, C., Palomo, L.J., Fernández, M.C., Kufner, M.B., Moreno, S., 2016. Long-term data from a small mammal community reveals loss of diversity and potential effects of local climate change. Curr. Zool. 63, 515-523. https://doi.org/10.1093/cz/zow109

Shapiro, D., Switzer, P., 1989. Minimum/maximum auto-correlation factor analysis. Technical Report 132. Department of Statistics, Stanford University, USA.

Shepard, R.N., 1962. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. Psychometrika. 27, 125–140. https://doi.org/10.1007/BF02289630

Shipley, B., 2013. The AIC model selection method applied to path analytic models compared using a d-separation test. Ecology. 94, 560–564. https://doi.org/10.1890/12-0976.1

Silva, T., Gislason, A., Licandro, P., Marteinsdóttir, G., Ferreira, A.S.A., Gudmundsson, K., Astthorsson, O.S., 2014. Long-term changes of euphausiids in shelf and oceanic habitats southwest, south and southeast of Iceland. J. Plankton. Res. 36, 1262–1278. https://doi.org/10.1093/plankt/fbu050

Spearman, C., 1907. Demonstration of formulæ for true measurement of correlation. Am. J. Psychol. 18, 161–169. https://doi.org/10.2307/1412408

Sponseller, R.A., Grimm, N.B., Boulton, A.J., Sabo, J.L., 2010. Responses of macroinvertebrate communities to long-term flow variability in a Sonoran Desert stream. Glob. Change. Biol. 16, 2891–2900. https://doi.org/10.1111/j.1365-2486.2010.02200.x

Stephens, P.A., Buskirk, S.W., Del Rio, C.M., 2007. Inference in ecology and evolution. Trends. Ecol. Evol. 22, 192–197. https://doi.org/10.1016/j.tree.2006.12.003

Strayer, D., Glitzenstein, J.S., Jones, C.G., Kolasa, J., Likens, G.E., McDonnell, M.J., Parker, G.G., Pickett, S.T.A., 1986. Long-term ecological studies: an illustrated account of their design, operation, and importance to ecology. Occasional Publication of the Institute of Ecosystem Studies. Institute of Ecosystem Studies, Millbrook.

Szentkirályi, F., Leskó, K., Kádár, F., 2007. Climatic effects on long-term fluctuations in species richness and abundance level of forest macrolepidopteran assemblages in a Hungarian moutainous region. Carpath. J. Earth. Env. 2, 73–82.

Taylor, J., Krumpen, T., Soltwedel, T., Gutt, J., Bergmann, M., 2017. Dynamic benthic megafaunal communities: Assessing temporal variations in structure, composition and diversity at the Arctic deep-sea observatory Hausgarten between 2004 and 2015. Deep. Sea. Res. Part. 1. Oceanogr. Res. Pap. 122, 81–94. https://doi.org/10.1016/j.dsr.2017.02.008

ter Braak, C.J.F., 1986. Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. Ecology. 67, 1167–1179. https://doi.org/10.2307/1938672

Thioulouse, J., Simier, M., Chessel, D., 2004. Simultaneous analysis of a sequence of paired ecological tables. Ecology. 85, 272–283. https://doi.org/10.1890/02-0605

Tian, Y., Kidokoro, H., Watanabe, T., 2006. Long-term changes in the fish community structure from the Tsushima warm current region of the Japan/East Sea with an emphasis on the impacts of fishing and climate regime shift over the last four decades. Progr. Oceanogr. 68, 217–237. https://doi.org/10.1016/j.pocean.2006.02.009

Torrence, C., Compo, G.P., 1998. A practical guide to wavelet analysis. B. Am. Meteorol. Soc. 79, 61–78. https://doi.org/10.1175/1520-0477(1998)079<0061:APGTWA>2.0.CO;2

Tucker, L.R., 1966. Some mathematical notes on three-mode factor analysis. Psychometrika. 31, 279–311. https://doi.org/10.1007/BF02289464

van der Wal, R., Stien, A., 2014. High-arctic plants like it hot: a long-term investigation of between-year variability in plant biomass. Ecology. 95, 3414–3427. https://doi.org/10.1890/14-0533.1

van Walraven, L., Dapper, R., Nauw, J.J., Tulp, I., Witte, J.I., van der Veer, H.W., 2017. Long-term patterns in fish phenology in the western Dutch Wadden Sea in relation to climate change. J. Sea. Res. 127, 173–181. https://doi.org/10.1016/j.seares.2017.04.001

Vaughan, I.P., Ormerod, S.J., 2012. Large-scale, long-term trends in British river macroinvertebrates. Glob. Change. Biol. 18, 2184–2194. https://doi.org/10.1111/j.1365-2486.2012.02662.x

Warton, D.I., Foster, S.D., De'ath, G., Stoklosa, J., Dunstan, P.K., 2015. Model-based thinking for community ecology. Plant. Ecol. 216, 669–682. https://doi.org/10.1007/s11258-014-0366-3

Wasmund, N., Tuimala, J., Suikkanen, S., Vandepitte, L., Kraberg, A., 2011. Long-term trends in phytoplankton composition in the western and central Baltic Sea. J. Marine. Syst. 87, 145–159. https://doi.org/10.1016/j.jmarsys.2011.03.010

Waterkeyn, A., Vanschoenwinkel, B., Vercampt, H., Grillas, P., Brendonck, L., 2011. Long-term effects of salinity and disturbance regime on active and dormant crustacean communities. Limnol. Oceanogr. 56, 1008–1022. https://doi.org/10.4319/lo.2011.56.3.1008

Wolfe, D.A., Champ, M.A., Flemer, D.A., Mearns, A.J., 1987. Long-term biological data sets: Their role in research, monitoring, and management of estuarine and coastal marine systems. Estuaries. 10, 181. https://doi.org/10.2307/1351847

Wright, S., 1921. Correlation and causation. J. Agric. Res. 20, 557–585.

Wright, S., 1960. Path coefficients and path regressions: Alternative or complementary concepts? Biometrics. 16, 189. https://doi.org/10.2307/2527551

Yoccoz, N.G., Nichols, J.D., Boulinier, T., 2001. Monitoring of biological diversity in space and time. Trends. Ecol. Evol. 16, 446–453. https://doi.org/10.1016/S0169-5347(01)02205-4

Zettler, M.L., Friedland, R., Gogina, M., Darr, A., 2017. Variation in benthic long-term data of transitional waters: Is interpretation more than speculation? PLOS ONE. 12, e0175746. https://doi.org/10.1371/journal.pone.0175746

Zuur, A.F., Ieno, E.N., Walker, N., Saveliev, A.A., Smith, G.M., 2009. Mixed effects models and extensions in ecology with R. Springer New York, New York. https://doi.org/10.1007/978-0-387-87458-6