UNIVERSITY OF SOUTHAMPTON

Faculty of Natural and Environmental Sciences

Ocean and Earth Sciences

Dynamics, control and variability of plankton in the Northeast Atlantic and North Sea between 1958–2014

by

Renata Stella Khouri

A thesis submitted for the degree of Doctor of Philosophy

19th September 2017

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES
OCEAN AND EARTH SCIENCES

Doctor of Philosophy

DYNAMICS, CONTROL AND VARIABILITY OF PLANKTON IN THE NORTHEAST ATLANTIC AND NORTH SEA BETWEEN 1958–2014

Marine plankton provide essential support to all life on the planet, through the production of oxygen, contribution to nutrient cycling in the ocean, and by being the base of the marine food web. Due to their small size and limited motility, plankton are also highly susceptible to changes in the environment. Monitoring and understanding how the plankton responds to different climatic and anthropogenic effects is imperative in order to predict and prevent damaging outcomes to the ecosystem.

In this thesis, I focused on key plankton indicators in the Northeast Atlantic and North Sea, between 1958 and 2014, with aim to understand how and why these populations have been changing through time, what drives the variability observed in the plankton, and which are the best approaches to model plankton abundance from monitoring data.

I first explored the relationship between two major phytoplankton groups, diatoms and dinoflagellates, with a index for chlorophyll concentration. Differences in trend and abrupt transitions, combined with statistical modelling, implied that a non negligible amount of the greenness in the water came from groups of smaller phytoplankton, which were inferred to have increased in abundance in the recent past.

I proceeded to look into bottom-up versus top-down control of phytoplankton variability, from an interannual perspective. Through vector autoregressive modelling, I analysed interactions among the phytoplankton, zooplankton, and environmental indicators. Sea surface temperature emerged as a significant driver of variability, and some evidence for bottom-up control was found from a lack of dependency of diatoms on the other plankton variables. However, the plankton seemed to be mostly regulated by serial correlation and the seasonal cycle.

The serial correlation suggested that there could be nonlinearity in the system, and led to the question of whether linear approximations were suitable for plankton dynamics. I investigated the presence of nonlinearity, stochasticity and deterministic chaos in the plankton. The seasonal cycle was key to stabilise the fluctuations within plankton populations, which appeared to be regulated by a nonlinear stochastic dynamics rather than by chaos. The lack of significant links between the plankton and the two environmental indicators reinforced the complexity of the plankton system, and implied that it is not likely to exist a single factor as the main driver of plankton variability.

The complexity of species interactions and environmental effects, combined with a strong dependency on the seasonal cycle, evidence how plankton communities are directly impacted by a changing climate. Changes to temperature, nutrient availability, ocean currents and mixing, for example, will have a direct impact on plankton phenology, with consequences likely to propagate across the global ecosystem.

Declaration of Authorship

I, RENATA STELLA KHOURI, declare that this thesis, titled 'DYNAMICS, CONTROL AND VARIABILITY OF PLANKTON IN THE NORTHEAST ATLANTIC AND NORTH SEA BETWEEN 1958-2014', and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: _____

Acknowledgements

To my supervisors, Claudie Beaulieu, Stephanie Henson, Adrian Martin, and Martin Edwards, thank you for your unvaluable advice throughout this process. I am forever grateful to your guidance, dedication, and, not least, patience to revise and correct my grammar mistakes.

To the staff at the Sir Alister Hardy Foundation for Ocean Sciences, thank you for keeping the CPR survey going. In particular, thanks to Darren Stevens and David Johns for providing the dataset used in this thesis (SAHFOS DOI No. 10.7487/2014.282.1.45).

To CAPES / Science Without Borders Brazil, thank you for funding this research.

To my family, thank you for supporting my ideas of following a scientific career, for teaching me how to be a better human, and for the many hours of nonsense talking through the internet.

To all of my friends, officemates, and PhD colleagues, thank you for all the the laughs, existential debates, experimental cooking, tea breaks, playlist exchanges, and occasional hunt for free food around the building.

In particular, thanks to Glaucia Fragoso for helping me find my way around when I moved to the UK; Joe Atkins-Turkish for many language discussions and random travels; Barbara Zihlmann for being such an awesome flatmate; Jamie Hizzett for having an appropriate gif for every possible situation; Anna, Sarah and Michelle for the kind words of encouragement when writing was hard; and Edmund King for always being there, and believing in me even when I failed to do so.

To the University of Southampton Light Opera Society, Southampton Choral Society, and University of Southampton Stage Technicians' Society, thank you for so many excellent evenings, and for being my family away from home. I miss you already.

Finally, thanks to you, for being interested in my work. I hope you enjoy the reading.

To infinity, and beyond.

Contents

List of Figures

List of Tables

Acronyms and abbreviations

AR	Autoregressive (model)
ARMA	Autoregressive Moving-Average (model)
CCM	Convergent Cross-Mapping
CPR	Continuous Plankton Recorder
DIAT	Diatoms
DINO	Dinoflagellates
EDM	Empirical Dynamics Modelling
FPT	Fraction of Predicted Trends
kNN	k-Nearest-Neighbours (method)
LCPP	Large Copepods
LB	Ljung-Box (test)
MA	Moving-Average (model)
MAR	Multivariate Autoregressive (model)
MLD	Mixed Layer Depth
MLR	Multivariate Linear Regression (model)
NAO	North Atlantic Oscillation (index)
PCI	Phytoplankton Colour Index
RMSE	Residual Mean Squared Error
SAHFOS	Sir Alister Hardy Foundation for Ocean Sciences
SCPP	Small Copepods
SST	Sea Surface Temperature
VAR	Vector Autoregressive (model)

Chapter 1

Introduction

1.1 Motivation

Plankton, microscopic organisms that drift in the upper ocean, form the base of the marine ecosystem and play key roles in biogeochemical cycles (Doney, 2006, Thomas et al., 2012). Responsible for about half of the global net primary productivity, phytoplankton are essential for the maintenance of the biosphere (Field et al., 1998). Zooplankton provide the link between the primary producers and most of the higher trophic levels in the marine food web (Richardson, 2008). The ecosystems of the Northeast Atlantic and the North Sea, regions of interest in this thesis, are highly susceptible to climate change, which is known to impact the plankton community in various ways depending on different local physical settings (IPCC, Field et al., 2014). These regions are also highly productive, and support many important commercial fisheries (Rouyer et al., 2014).

Changes in community structure, abundance and geographical distribution of plankton have been reported in the Northeast Atlantic and North Sea. Rising temperatures, stronger summer winds and other physical effects from a changing climate have been related to trophic mismatches (Beaugrand, 2009); there is evidence of a northwards migration of warm-water species of zooplankton (Edwards et al., 2013, Harris et al., 2015); and of community structure changes in the phytoplankton (Hinder et al., 2012a). However, the matter of whether any particular variable can be singled out as the main driver of changes in plankton populations, and whether bottom-up or top-down effects are more important in the control of phytoplankton variability in the long term, are still open questions. Both topics are discussed in this thesis.

The long-term variability in marine plankton abundance has been conjectured to be a consequence of nonlinear responses to external forcing (Barton et al., 2015), and can be stochastic or deterministic in its essence. The latter alternative encompasses the possibility of chaotic dynamics, that is, when deterministic mechanisms highly sensitive to perturbations generate oscillations that may appear to arise from a random signal. The presence of deterministic chaos in ecological systems is a topic of great interest from both an applied and theoretical perspective (Sugihara and May, 1990). Knowing the underlying dynamics of an ecological system improves our ability to describe and model the system. In the case of marine plankton, that is particularly relevant to evaluate the feasibility of forecasting and causality analysis (McGowan et al., 2017). Determining the nature of plankton dynamics also brings potential to improve the accuracy of computational models for plankton populations. The different aspects of plankton dynamics are also a topic discussed in this thesis.

1.2 Phytoplankton

Phytoplankton are single-celled autotrophs that produce their energy through photosynthesis and form the base of the oceanic food web (Behrenfeld et al., 2006, Graham et al., 2009, Guiry, 2012). With a rapid life cycle, the turnover time of the global ocean phytoplankton biomass is believed to be less than a week. Marine phytoplankton are responsible for a great portion of the global net primary productivity, and also play key roles in the carbon and nitrogen cycles in the ocean. The main phytoplankton groups considered in this thesis are diatoms and dinoflagellates, which contain some of the largest unicellular marine eukaryotes.

As photosynthetic organisms, light availability is one of the main factors that limit phytoplankton growth, affecting both the amount and rate of photosynthesis (Garrison, 2013, Lalli and Parsons, 1997, Longhurst, 2006, Morán et al., 2010). The amount of irradiance that becomes available for phytoplankton varies naturally with the seasons of the year. Nutrient availability is also a main controlling factor of phytoplankton growth. Nutrients become available when the deep, nutrient-rich waters mix with the upper ocean nutrient-poor surface waters.

The ocean is stratified by density, and the surface layer in which temperature and salinity are fairly uniform is called the mixed layer (Garrison, 2013, Longhurst, 2006). The depth of the mixed layer in the Northeast Atlantic varies with the seasons: shallow in summer, with higher temperatures, less turbulent waters and lower nutrient concentration; and deep in winter, with increased wind stress, cooler surface waters and increased nutrient concentration. Surface temperature is also an important factor controlling phytoplankton metabolism. It can alter growth, photosynthesis and reproduction rates, and can also trigger cell death if it goes above each species' specific thermal tolerance.

1.2.1 Diatoms and dinoflagellates

Diatoms are an abundant and diverse group of key primary producers in the ocean (Garrison, 2011, Graham et al., 2009, Mann, 2010, Reid et al., 1990). They are dominant in cold and nutrient-rich temperate waters, and are responsible for about 20% of the global ocean carbon fixation. Individual cells typically vary between 2 and 200 μm in length, have a regular symmetric shape, and are enclosed by a rigid silica wall. Diatoms can exist as single cells or as colonies suspended in the water column, and are present all over the ocean. Many species have a fast growth rate, and tend to form blooms early in the growing season. With the exception of some gametes, diatoms do not have any flagellae and thus are not motile.

Dinoflagellates can be autotrophic, mixotrophic or heterotrophic (Graham et al., 2009, Hoppenrath and Saldarriaga, 2012, Lalli and Parsons, 1997). About half of the species do not possess plastids, and are sometimes considered as part of the microzooplankton since they feed on other small phytoplankton and bacteria. Dinoflagellates are efficient in light capture and energy transmission. Their cells are motile, with two dissimilar flagellae, and tend to be large (extremes ranging from 5 to 2000 μm in length). Their size can be a disadvantage when competing for nutrients with smaller plankton, which have a larger surface-to-volume ratio. They usually have a lower growth rate than diatoms, but, due to their ability to store nutrients, dinoflagellates can produce blooms even when nutrient levels are low in the water column. Dinoflagellates are also sensitive to, and can be damaged by, small scale turbulence (Berdalet et al., 2007). Many species are toxic and produce harmful algal blooms, which can impact fish, shellfish and their consumers, including humans (Garrison, 2013).

There is no agreement on how many extant species of diatoms exist in the ocean, but some estimate it could be in the order of 10^5 (Vanormelingen et al., 2007). As for oceanic dinoflagellates, the number of living species is estimated to be in the order of 10^3 (Gómez, 2005). The reader can refer to Guiry (2012), for example, for a more detailed discussion on phytoplankton species diversity.

In the Northeast Atlantic, diatoms are dominant in the spring, when high nutrient concentrations are present and the water column is more stratified than during winter (Jordan, 2009, Lalli and Parsons, 1997, Longhurst, 2006). However, by late spring the nutrient concentration is already declining and light becomes more abundant, causing diatoms to be outcompeted by dinoflagellates and sink out of the surface. Small dinoflagellates and other plankton species that do not require the same high concentrations of nutrients as diatoms in order to grow then become more abundant. If new nutrients become available in the stratified surface waters at this time, dinoflagellate blooms can happen. During autumn, decreasing light intensity and occasional mixing favour the return of diatoms, which may produce autumn blooms on a smaller scale in relation to spring. In winter, low water temperatures, low light intensity and absence of stratification due to turbulence prevent significant phytoplankton growth.

1.3 Zooplankton

Zooplankton are heterotrophic planktonic organisms, comprising a variety of sizes and habitats, and are essential for the functioning of the marine ecosystem (Humes, 1994, Johnson and Allen, 2012, Lalli and Parsons, 1997, Melle et al., 2014, Reid et al., 1990). They are the main grazers of phytoplankton, and in turn are consumed by a wide range of animals, from larger zooplankton to baleen whales, providing a main pathway for energy flow through the oceanic food web.

Although zooplankton can swim vertically through the water column, their large-scale geographical mobility is limited and subject to water currents (Johnson and Allen, 2012, Lalli and Parsons, 1997, Richardson, 2008). In addition to prey availability, water temperature is one of the main factors controlling zooplankton populations, regulating growth, respiration, reproduction and mortality rates according to each species' physiological tolerance. Zooplankton distribution in space and time is usually patchy,

depending on factors such as turbulence, mixing, and predator-prey interactions. As a consequence to their short lifetime, there can be a tight coupling between zooplankton dynamics and climate (Richardson, 2008, Taylor et al., 2002).

1.3.1 Copepods

Copepods are a large group of crustaceans, some of the most abundant and well studied oceanic zooplankton (Johnson and Allen, 2012, Melle et al., 2014, Richardson, 2008). The number of known species amounts to the order of 10^4, the order Calanoida alone containing about 1850 different species (Johnson and Allen, 2012, Lalli and Parsons, 1997). Copepods are a dominant group in the Northeast Atlantic. Most copepods feed on phytoplankton, although some species can be omnivorous and feed on small zooplankton as well. Copepods support many populations of fish, sea birds and marine mammals, including commercial stocks such as herring, mackerel, capelin, blue whiting and salmon (Johnson and Allen, 2012, Melle et al., 2014, Richardson, 2008). Copepod populations in general peak between spring and summer, although some species can peak between summer and autumn. The life cycle varies with each different species. Many are seasonally present in the water column and go through a diapause period, usually over the winter months, producing resting eggs and re-emerging in the following year (Baumgartner and Tarrant, 2017).

1.4 Plankton and climate

Global warming has impacts on physical features of the ocean, and consequently on all life forms that inhabit it (Garrison, 2013, Hurrell, 1995, Richardson, 2008). For example, the melting of snow and ice, as well as changes in precipitation and evaporation, can cause a freshening of waters in mid and high latitudes and increased salinity in low latitudes, with direct impact on mixed layer depths. Changes in water properties can have large scale impacts on ocean circulation. Rahmstorf et al. (2015), for instance, present evidence of a slowdown in the Atlantic Meridional Overturning Circulation that seems to be related to the melting of ice sheets in Greenland.

Upper ocean water temperature is directly related to incoming irradiance and stratification of the water column, and sometimes used as a proxy for those variables (Bouman et al., 2003, Lalli and Parsons, 1997, Morán et al., 2010, Richardson, 2008). In

addition to changes in metabolic rates, sea surface temperature impacts the composition, abundance and trophic efficiency of plankton populations. Warmer waters usually imply a more strongly stratified and nutrient-poor mixed layer, and are associated with less efficient food web structures, smaller-sized phytoplankton and zooplankton, and more steps linking primary producers and higher trophic levels such as fish and seabirds. On the other hand, colder water temperatures imply less stratification and more nutrients being mixed from deeper waters. Sea surface temperature is also a good proxy for nutrient enrichment in the ocean, and is sometimes considered to be the main physical covariate in oceanic ecosystems (Richardson, 2008, Richardson and Schoeman, 2004).

Atmospheric events such as the North Atlantic Oscillation (NAO) can also influence plankton dynamics. The NAO index is calculated from the pressure gradient between the Azores and Iceland during winter (Longhurst, 2006). Positive and negative phases of the NAO affect the intensity of westerly winds and storm tracks in the North Atlantic (Hurrell, 1995). Windiness is directly related to the depth of the mixed layer. Therefore wind intensity, and consequently the NAO index as well, can potentially be related to variability in the plankton. The NAO index has been correlated with fluctuations in zooplankton species of the Northeast Atlantic (Fromentin and Planque, 1996), and changes to the timing

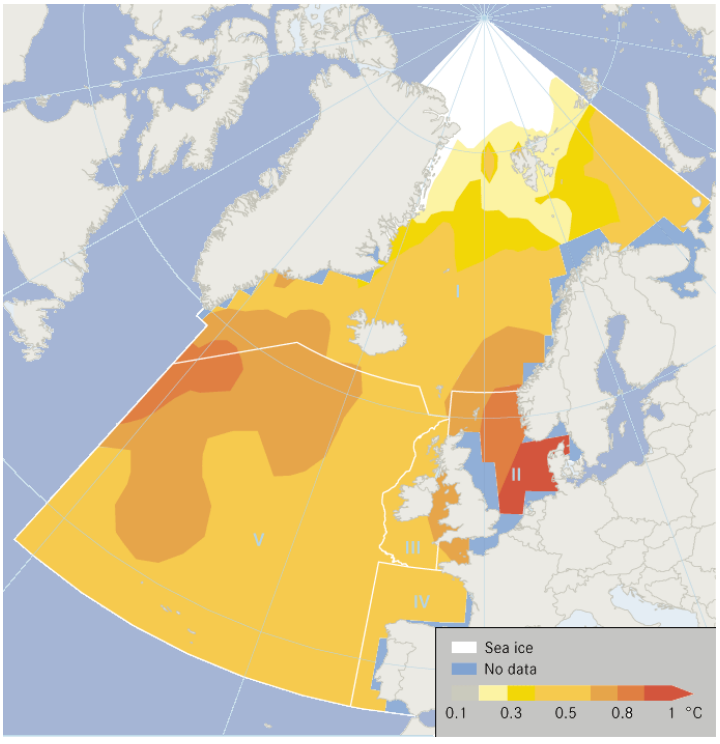

FIGURE 1.1: Annual mean sea surface temperature anomalies in the Northeast Atlantic, for the period of 1999-2008 in relation to 1971-2000. Figure from OSPAR (2010).

of the phytoplankton spring bloom in the central North Atlantic have also been correlated to the NAO and sea surface temperature (Zhai et al., 2013).

It has been conjectured that plankton could serve as indicators of climate change (Taylor et al., 2002), since the signals of changes in the environment could be amplified by the response of the plankton. However, there is also evidence that long datasets, with length of at least 30 years, would be required in order to distinguish a climatic signal from the high interannual variability usually observed in plankton time-series (Henson et al., 2010). Plankton also contribute to the regulation of the global climate (Falkowski, 2012, Falkowski et al., 1998, Richardson, 2008). Carbon dioxide from the atmosphere is fixed by phytoplankton during photosynthesis, then consumed by zooplankton, and finally sinks to the sea bed. This biological pump is one way in which the oceans absorb the excess carbon in the atmosphere.

1.5 The Northeast Atlantic and North Sea

Throughout this thesis, I focus on subregions of the North Atlantic ocean between 45^{o}N - 63^{o}N and 20^{o}W - 10^{o}E. Those temperate waters are part of the Northeast Atlantic and North Sea. They are important from an ecological perspective, with high primary productivity, biodiversity and home to endangered species. They are also home to various commercial fish stocks and important shipping routes (OSPAR, 2010, Rouyer et al., 2014). The region of interest covers shallow coastal waters under the influence of tidal forces, as well as deep oceanic waters mixed by strong winds (Longhurst, 2006). The topography and circulation patterns in the region provide a diversity of habitats for oceanic fauna and flora (Melle et al., 2014).

The ecosystems of the Northeast Atlantic and the North Sea are highly susceptible to climate change (IPCC - Field et al., 2014, Larsen et al., 2015, OSPAR, 2010). Variations in temperature and circulation have been observed in the area, with sea surface temperatures exceeding the long-term historical average by as much as 2^{o}C in some places (figure 1.1), and warming trends observed across the entire region. With increasing temperatures, the northern part of the Northeast Atlantic may be subject to freshening due to the melting of sea ice from the Arctic, and to invasion by non-indigenous species (OSPAR, 2010). The average ocean surface pH of the region is also decreasing and, although the impacts of

acidification are not yet clear, this could have severe consequences for biological processes and carbon sequestration. In the North Sea, overfishing, eutrophication and pollution due to human activity are also believed to have a huge impact in the ecosystem. In the Celtic Sea, there are reports of damage from benthic trawling as well as unsustainable fishing (OSPAR, 2010).

1.5.1 Plankton community and variability

The abundance and distribution of phytoplankton and zooplankton in the North Atlantic are often associated with changes in temperature, mixed layer depth and large-scale atmospheric variability (Edwards and Richardson, 2004, Henson et al., 2013, Martinez et al., 2016, Racault et al., 2012, Richardson, 2008, Thomas et al., 2012, Zhai et al., 2013). Different functional groups do not respond in the same way to changes in the environment, and temporal mismatches could lead to an imbalance in predator-prey relationships, with consequences propagating to higher trophic levels. There is an ongoing debate on whether there is a mismatch in the peaks of phytoplankton and zooplankton in spring, with species that peak during summer being reported to have advanced in their seasonality as well (Edwards and Richardson, 2004). Geographical shifts have also been observed, with warm-water species being found in more temperate regions, such as the notorious example of the recent northwards migration of warm-water copepod *Calanus helgolandicus* into waters previously dominated by cold-water *Calanus finmarchicus* (e.g., Bonnet et al., 2005, and figure 1.2).

Large scale regime shifts have been described in the ecosystems of the North Sea and Northeast Atlantic. Those have been related, for example, to anomalies in temperature and the North Atlantic Oscillation (Beaugrand, 2004, Zhai et al., 2013), temperature and the balance of dissolved nutrients (AlvarezFernandez et al., 2012), deepening of mixed layers and stronger winds (Martinez et al., 2016) and increased sea surface temperatures (Harris et al., 2015). The late 1980s shift in the North Sea has been thoroughly documented (Beaugrand, 2004), however there are also reports of potential shifts in the 1960s and 2000s in the same region (e.g., AlvarezFernandez et al., 2012, Beaugrand et al., 2014). There is no agreement on which is the main driver of the observed shifts, although temperature, either from the sea surface or atmospheric, is a recurrent candidate across multiple studies.

Long-term records of individual phytoplankton groups or species can be scarce, and sometimes a chlorophyll or colour index may be preferred. Satellite-derived chlorophyll time-series are often used to represent phytoplankton biomass (e.g., Behrenfeld et al., 2016b, Doney, 2006). However, comparing satellite datasets is usually challenging due to differences in methodology and possible gaps between missions (Raitsos et al., 2013b). In the North Atlantic, the Continuous Plankton Recorder (CPR) survey has been keeping semi-quantitative records of chlorophyll in the form of an index that measures the greenness of the water since 1946 - the phytoplankton colour index (PCI). The PCI has been shown to be comparable to satellite chlorophyll measurements, and to be a good estimate for phytoplankton biomass (e.g., Batten et al., 2003, Raitsos et al., 2013b). One important question is to what extent does the variation in chlorophyll indices represent changes within the phytoplankton community. I address this issue and discuss the relationship between the PCI and abundance of diatoms and dinoflagellates in chapter 3.

Although phytoplankton and zooplankton are linked by the seasonal predator-prey succession, it is also important to understand the long-term relationships between them, since interannual variability in climatic covariates can potentially become dominant sources of variability for the plankton (Barton et al., 2015). Most previous studies based on *in situ* observations of plankton populations consider the impacts of climate and the environment on either phytoplankton (Hinder et al., 2012b) or zooplankton (Richardson, 2008) separately, but few account for the interactions among the plankton itself. Furthermore, there is no consensus on to what extent the interannual variability in the plankton is due to

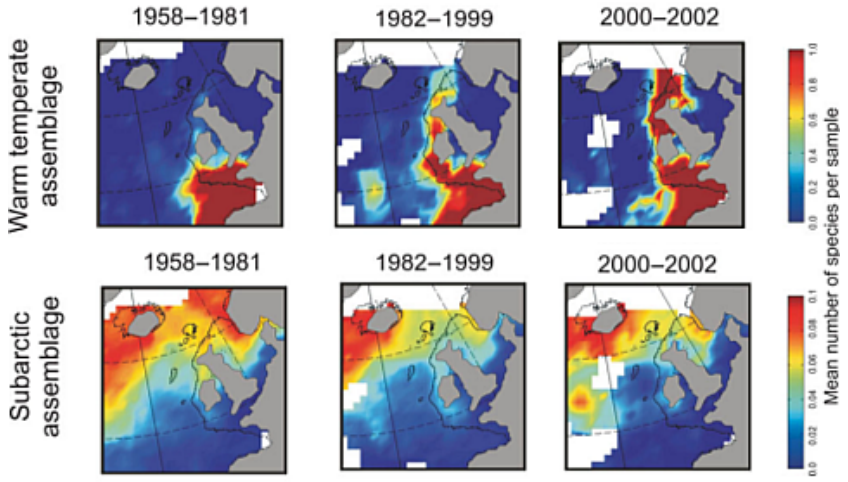

FIGURE 1.2: Averaged geographical distribution of warm-water (containing *C. helgolandicus*) and subartic (containing *C. finmarchicus*) copepod species between 1958 and 2002. Figure adapted from Richardson (2008), based on original from Beaugrand et al. (2002).

bottom-up (climate and environment driven) or top-down (predation driven) processes, although bottom-up control is more frequently reported in the literature (e.g., Chassot et al., 2007, Richardson, 2008). I investigate the presence of bottom-up versus top-down control in phytoplankton of the Northeast Atlantic and North Sea in chapter 4.

1.6 Nonlinearity, stochasticity and chaos in the plankton

The set of statistical methods readily available in the literature and scientific computing softwares has been, in most cases, developed based on hypotheses of linearity of the phenomena being analysed. Ecological systems, however, can present nonlinearity due to complex interactions (e.g., Benincà et al., 2015, McGowan et al., 2017, Sugihara, 1994). Traditional methods, such as linear regression, correlations and analysis of variance, may not be suitable to approximate nonlinear systems, and could potentially lead to inaccurate results if nonlinear effects are significant. On the other hand, nonlinear methods can be more difficult to implement than their linear counterparts, or may require a level of detail or length of dataset that is not readily available. Establishing whether the plankton system is linear or not is important for the identification of appropriate methods for future analysis of plankton data, and also for its potential to improve theoretical models of the plankton dynamics.

When attempting to model natural phenomena, one can choose among a range of stochastic or deterministic models, or even a combination of both. Each model has its own advantages and drawbacks, and ultimately the choice of an ideal model may depend on the data available and the question one is seeking to answer. In the next paragraphs I briefly mention a few examples of stochastic and deterministic models that have been used in the literature to describe plankton systems.

Vector autoregressive (VAR) models are a multivariate extension of stochastic autoregressive models, and can be used to describe the interactions within the plankton, as well as between the plankton and environmental covariates. They have been successfully applied to describe plankton food webs in lakes (Gsell et al., 2016, Ives et al., 2003), and also to identify relationships between chlorophyll and environmental factors in phytoplankton blooms and predict bloom occurences (Lui et al., 2007). Applications of VAR models to oceanic plankton communities are scarce (Hampton et al., 2013) and this type of

modelling has not yet been used to explore the dynamics of plankton populations across the Northeast Atlantic and North Sea. This matter is also discussed in chapter 4.

General additive models (GAMs) are a nonparametric class of stochastic models, and have the advantage of being more flexible than standard linear regressions. GAMs have been used, for example, to model and predict the occurrence of zooplankton species in relation to climate change (Villarino et al., 2015). This type of model has also been applied to investigate the relationship between trophic interactions of phytoplankton and zooplankton in the Bay of Biscay (Stenseth et al., 2006), and to explore bottom-up versus top-down regulation in the ecosystem of the North Sea (Lynam et al., 2017).

Deterministic approaches have also been used to describe plankton dynamics. Differential equations have been used, for example, to construct theoretical models describing the oscillations in deep chlorophyll maxima in relation to mixing (Huisman et al., 2006). Empirical Dynamic Modelling (EDM) is a non-parametric nonlinear forecasting method, particularly good for nonlinear and potentially chaotic deterministic systems. This method has been successfully applied to describe and predict algal blooms in the coast of California (McGowan et al., 2017). Similar methods have also be used to investigate the presence of chaos in plankton populations of the Mid Atlantic Bight (Ascioti et al., 1993) and to identify the nonlinearity of diatom abundance time-series at the Scripps pier (Sugihara, 1994).

Although there is a large amount of data available from the CPR survey in the North Sea and Northeast Atlantic (Richardson et al., 2006), it is still not known whether the plankton of these regions follow nonlinear or potentially chaotic dynamics (Barton et al., 2015). In chapter 5, I use some variations of EDM to investigate if key CPR time-series exhibit nonlinear behaviour, and discuss the presence of chaos versus stochasticity in the plankton of the region.

1.7 Thesis structure

The general objective of this thesis is to improve our understanding of, and describe through statistical modelling, the long-term variability and interactions between key plankton groups and environmental indicators in the Northeast Atlantic and North Sea. This can be summarised in the three following main topics:

- Is it possible to make inferences about individual phytoplankton groups from water greenness data?
- Is the variability observed in phytoplankton more strongly related to changes in the environment, or to zooplankton grazing?
- Is it possible to infer the theoretical dynamics (deterministic or stochastic) of plankton systems from abundance time-series?

In particular, the main topics discussed in each chapter of this thesis are outlined below:

Chapter 2. ***Data and methods***

A description of the CPR survey, of the plankton time-series and of the environmental time-series used in this thesis, as well as an explanation of all of the statistical models and methods used in this thesis.

Chapter 3. ***Water greenness as a function of diatom and dinoflagellate abundance***

- What is the relationship between the water greenness and abundance of diatoms and dinoflagellates?
- Why does the long-term water greenness variability seem so different from those of diatom and dinoflagellate abundances?

Chapter 4. ***Control mechanisms of interannual variability in phytoplankton populations***

- What can we infer about the long-term interactions between phytoplankton and zooplankton from abundance data?
- What are the impacts of environmental variables on those interactions?
- Is it possible to identify whether the phytoplankton are under bottom-up or top-down control, from an interannual perspective?

Chapter 5. ***Nonlinearity, stochasticity and chaos in the plankton ecosystem***

- Is there evidence for deterministic chaos in the plankton?

- How strong is the influence of stochasticity in the plankton?
- Is it possible to identify if the plankton dynamics, and the dynamics of key environmental proxy variables, are linear or nonlinear?

Chapter 6. *Conclusion*

A synthesis of previous chapters, with discussion on why it may not be possible to determine a single driver of plankton variability on interannual timescales, limitations of the present study, and possible directions for future research.

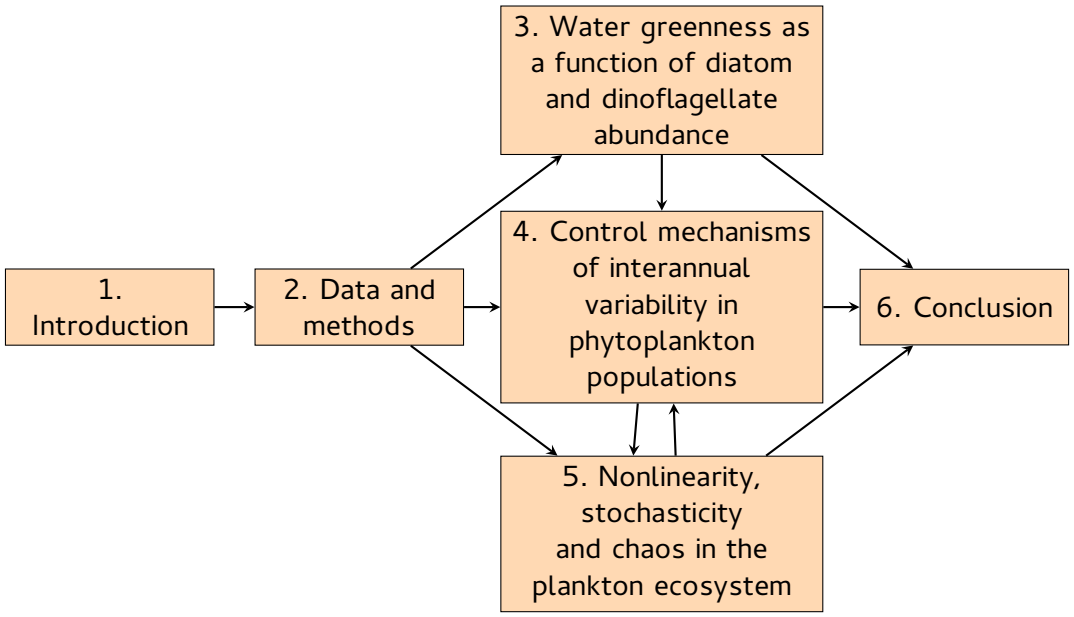

FIGURE 1.3: Structure of the thesis' chapters.

Chapter 2

Data and methods

In this chapter I present the dataset and all statistical methods used in the following chapters of the thesis. The first section (2.1) describes the data: the plankton time-series from the Continuous Plankton Recorder survey and environmental time-series of sea surface temperature, mixed layer depth and the North Atlantic Oscillation index. Unless stated otherwise, throughout this thesis all time-series were standardized into z-scores, to have a mean of zero and a standard deviation of one.

Section 2.2 describes the informational approach for the detection of change-points in time-series, which is used in chapter 3 for investigating the presence of abrupt changes in the series of the phytoplankton colour index and abundance of diatoms and dinoflagellates. Section 2.3 introduces the class of autoregressive moving-average (ARMA) statistical models, which are used in chapter 3 as one possible type of noise process in the model described later in section 2.5 for the plankton system. The stationarity requirements for the usage of autoregressive time-series models are outlined in section 2.4. ARMA models are also extended to a multivariate context in section 2.6, and applied in chapter 4 to describe the interactions among different groups of phytoplankton, zooplankton and environmental forcing. All of these models are stochastic in their conception. It is important to consider the alternative of a deterministic dynamic structure, and section 2.7 presents a method later used in chapter 5 to identify patterns of deterministic chaos and nonlinearity in time-series. This section also describes a deterministic approach to identify causal relationships between time-series, which is used to investigate links between the plankton and potential environmental forcing.

The plankton data used on this thesis was provided by the Sir Alister Hardy Foundation for Ocean Sciences. The data can be obtained from SAHFOS upon request at `www.sahfos.ac.uk/data/our-data/`. The programming codes developed for this thesis will be available at `github.com/skrenata`.

2.1 Data

2.1.1 Continuous Plankton Recorder

The Continuous Plankton Recorder (CPR) survey is maintained by the Sir Alister Hardy Foundation for Ocean Sciences (SAHFOS), and has been monitoring plankton in the upper layer of the ocean in the North Atlantic and North Sea consistently since 1958. CPR machines attached to ships of opportunity filter water through a moving band of silk, providing a picture of the ocean micro fauna and flora at a depth of 5 to 10m below the surface. The mesh size of the silk is 270μm, and was originally designed to collect zooplankton (Batten et al., 2003). Large dinoflagellates and chain-forming diatoms are readily caught in the silk, whereas smaller phytoplankton are often underestimated (Leterme et al., 2006). The silk is divided into sections representing 10 nautical miles of tow, and the geographical position of the midpoint of each section is calculated. Individual species of phytoplankton and zooplankton on the silk are then identified and counted. Additional details on the data collection can be found in Richardson et al. (2006) and references therein.

Throughout this thesis I use monthly abundance time-series of diatom, dinoflagellate, small copepods (all life stages smaller than 2mm, Batten and Welch, 2004) and large copepods from the CPR survey, from 1958 to 2014. Diatoms and dinoflagellates are key phytoplankton groups in the marine food web, and, due to their size, are the most frequently sampled phytoplankton groups in the CPR survey. Copepods are the most abundant zooplankton in the studied region (Richardson, 2008) and thus are used to represent zooplankton populations.

I also use the phytoplankton colour index (PCI), a semi-quantitative indicator of phytoplankton biomass, visually assessed from the greenness of the sampled silk. The silk collected by the CPR machines is assigned a value according to its colour, corresponding to

four categories comparable to water chlorophyll concentration (Batten et al., 2003, Hays and Lindley, 1994). The original PCI categories (0,1,2,3) have been shown to represent a semi-logarithmic scale of increasing colour intensity (Hays and Lindley, 1994), therefore on this thesis the transformed colour intensity scale (0,1,2,6.5), which is the current standard in the CPR dataset, was used as an index to represent chlorophyll concentration in each piece of silk. The final PCI time-series are obtained by averaging these values for each sample over the number of samples taken from each area in each particular month, thus yielding continuous time-series.

The PCI index is believed to be a reliable proxy for phytoplankton biomass, and it has been compared and validated against recent satellite-estimated chlorophyll data (Raitsos et al., 2013a, 2005). Large phytoplankton, such as dinoflagellates, diatoms and their chains, are more frequently captured on the CPR silk. Smaller phytoplankton tend to pass through, or break in the mesh, but can still contribute to the PCI because their chloroplasts remain even when their cells are broken (Batten et al., 2003). In addition, small unarmoured flagellates disintegrate in contact with the formaldehyde used to preserve the silk, but their chloroplasts are not destroyed and can also contribute towards the PCI (Edwards et al., 2001, Leterme et al., 2006).

Given its sampling methodology, the CPR survey tends to underestimate species of microphytoplankton and microzooplankton. However, the length and consistency of the dataset provide a unique framework to investigate long-term, interannual changes in plankton distribution and diversity (Owens et al., 2013).

2.1.2 Environmental variables

In this thesis I use mainly sea surface temperature (SST) and mixed layer depth (MLD) to represent the physical environment, since these variables are most frequently associated with changes in the plankton (AlvarezFernandez et al., 2012, Barton et al., 2015, Harris et al., 2014, Hinder et al., 2012b, Martinez et al., 2016). In some places, the North Atlantic Oscillation (NAO) index was also considered as an environmental predictor (Zhai et al., 2013).

The SST time-series was downloaded from the Hadley Centre Sea Ice and Sea Surface Temperature dataset (HadISST, Rayner et al., 2003, available at

www.metoffice.gov.uk/hadobs/hadisst/). The MLD was calculated from the subsurface ocean temperature and salinity profile data (EN4, version EN.4.1.1), available at: www.metoffice.gov.uk/hadobs/en4/ (Good et al., 2013) using the MATLAB GSW Oceanographic Toolbox library (McDougall and Barker, 2011) and a 0.03kg m^{-3} difference in density from the surface (de Boyer Montégut et al., 2004). The NAO index time-series was downloaded from the NOAA Climate Prediction Centre website.

Nutrient availability and other biochemical variables are also of importance to plankton dynamics, however, to my knowledge, there is no dataset available that spans the same temporal and geographical extent of the CPR survey. SST and MLD are often highlighted as the main drivers of interannual changes in plankton populations (section 1.4 and Irwin and Finkel, 2008, for example). SST can impact the plankton metabolism and is associated to seasonal changes in light availability, and MLD is related to both irradiance and availability of nutrients in the water column (Beaugrand, 2009, Harris et al., 2014, Leterme and Pingree, 2007, Martinez et al., 2016). Therefore, those series should provide a good approximation for the environmental conditions that can impact the plankton.

2.2 Informational change-point detection analysis

The Informational Change-Point Detection (ICPD) method (Beaulieu et al., 2012a,b, Chen and Gupta, 2012) is used to identify abrupt changes in the parameters of a time-series. The final output of the procedure is the selection of a statistical model to describe the time-series. The simplicity of the method allows for comparisons among various different types of linear models.

To understand how the ICPD method works, consider the following example. Suppose one has to decide whether a model with $constant\ mean\ +\ white\ noise$ or $shift\ in\ the\ mean\ +\ white\ noise$ would be more adequate to describe a given series of length n. Each point c is considered as a potential change-point, and the following collection of models are compared:

$$X_0(t) = \mu + \varepsilon(t) \quad \text{and} \quad X_c(t) = \begin{cases} \mu_1 + \varepsilon(t) & , \ t \leq c \\ \mu_2 + \varepsilon(t) & , \ t > c \end{cases} \quad \text{for all } c \text{ in } \{3, \ldots, n-2\} \qquad (2.1)$$

where μ is the sample mean, μ_1, μ_2 are the means after and before the hypothesized shift, σ is the sample variance, $\varepsilon(t)$ is a white noise process with mean zero and variance σ^2, and t in $\{1, 2, \ldots, n\}$.

For each model, the Bayesian Information Criterion (BIC) is calculated (Schwarz, 1978). The BIC rewards models that maximize the likelihood function and explain well the observations, and penalises proportionally to the number of parameters needed in the model. If n is the number of observations, $\Theta = (\theta_1, \ldots, \theta_p)$ are the p parameters of the model, and $L(\Theta)$ is the likelihood function, the BIC is defined as:

$$BIC = -2\log(L(\Theta)) + p\log(n) \tag{2.2}$$

In this example, the first model X_0 has two parameters (the unchanging mean and the variance in the noise term), while the alternative models X_c have three parameters each (the means before and after the change-point and the variance in the noise term) (Beaulieu et al., 2012). The BIC of the initial and alternative models (X_0 and X_c for all c in $\{3, \ldots, n-2\}$) are compared, and the model with the lowest score (that is, with best balance between number of parameters and good fit to the data) is selected.

To minimise the chance of false detections, I compare the results with those from synthetically generated time-series. When a change-point is detected, 10,000 series with no change-point and same parameters μ and σ as the original series are generated. The selection procedure is repeated for each of these synthetic series. The p-value for the significance of the identified shift is calculated from the number of false positive detections on the synthetic series. A similar reasoning can be used to test for change-points in a linear model $X_0(t) = \alpha + \beta\, t + \varepsilon(t)$.

Once a model is selected, I check if there is evidence of serial correlation in its residuals. If that is the case, the analyisis is repeated replacing the white-noise term in the model with a serially correlated (red noise) term. It is then possible to assess if the change-point previously identified remains significant or if it is a consequence of the serial correlation.

In chapter 3, I compare the following models:

$$\text{Constant mean + white noise: } X(t) = \mu + \varepsilon(t) \tag{2.3}$$

$$\text{Shift in the mean + white noise: } X(t) = \begin{cases} \mu_1 + \varepsilon(t) \ , & t \leq c \\ \mu_2 + \varepsilon(t) \ , & t > c \end{cases} \tag{2.4}$$

$$\text{Linear trend + white noise: } X(t) = \alpha + \beta\, t + \varepsilon(t) \tag{2.5}$$

$$\text{Shift in trend and/or intercept + white noise: } X(t) = \begin{cases} \alpha_1 + \beta_1\, t + \varepsilon(t) \ , & t \leq c \\ \alpha_2 + \beta_2\, t + \varepsilon(t) \ , & t > c \end{cases} \tag{2.6}$$

When there was evidence of serial correlation in the series, the following models are also considered:

$$\text{Constant mean + red noise: } X(t) = \mu + R(t) \tag{2.7}$$

$$\text{Shift in the mean + red noise: } X(t) = \begin{cases} \mu_1 + R(t) \ , & t \leq c \\ \mu_2 + R(t) \ , & t > c \end{cases} \tag{2.8}$$

$$\text{Linear trend + red noise: } X(t) = \alpha + \beta\, t + R(t) \tag{2.9}$$

$$\text{Shift in trend and/or intercept + red noise: } X(t) = \begin{cases} \alpha_1 + \beta_1\, t + R(t) \ , & t \leq c \\ \alpha_2 + \beta_2\, t + R(t) \ , & t > c \end{cases} \tag{2.10}$$

where c is the change-point, $\varepsilon(t)$ is a white noise process with mean zero and variance σ^2, and $R(t)$ is a serially correlated AR(1) process (see the following section 2.3).

In this thesis, I consider the possibility of at most one change-point in the phytoplankton time-series studied later in chapter 3. Although multiple change-points may occur in nature, in that chapter I deal with annual averages for the identification of potential change-points in the phytoplankton. Since these time-series are of at most fifty-seven years in length, it would be unreasonable to expect more than one change-point in each of them. In addition, considering multiple change-points in short time-series may lead to misspecification of the model, as discussed for example in Spencer et al. (2011).

2.3 Auto-regressive moving-average (ARMA) models

A time-series $X(t)$ is said to be serially correlated, or autocorrelated, if its value in the present time is dependent on some subset of its past. For example, consider a first-order autoregressive process AR(1), figure 2.1. This kind of process arises when the value of the series at time $t+1$ depends directly on the value of the series at time t, plus some added noise. Time-series of natural processes are often serially correlated (Rudnick and Davis, 2006, 2003).

A time-series $X(t)$ is said to be stationary if its expectation and variance $\mathbb{E}(X(t))$ and $Var(X(t))$ are well defined and independent of t, and its covariance $Cov(X(t), X(t+j))$ is also independent of t for any j (Davidson and MacKinnon, 2004). The covariance between two observations should depend only on how far apart they are from each other, and not on the specific time in which they happened. This definition is sometimes called "covariance stationarity" or "weak stationarity". An example of a non-stationary process can once again come from an AR(1) series: if its dependency on the past is strong, the series variance is likely to grow indefinitely as t increases.

One way to model a serially correlated series is with autoregressive moving-average (ARMA) processes. A precise definition of an ARMA model can be found in a standard time-series analysis textbook (for instance, Box and Jenkins, 1970 or Davidson and MacKinnon, 2004). Here I give an informal definition. An $ARMA(p, q)$ process is a combination of two distinct processes: an autoregressive part of order p, $AR(p)$, and a moving-average part of order q, $MA(q)$. The autoregressive term depends explicitely on the past values of the series, while in the moving-average term the dependency comes from smoothed past values of the noise term.

AR and MA processes can be used individually to describe serially correlated time-series, and the difference between AR and MA processes is the type of serially correlated structure that they represent. The order p or q represent how far in the past is the dependence in the serial correlation. This means, for example, that either an $AR(2)$ or an $MA(2)$ processes could be used to describe a time-series in which the present value depends on its two most recent past values. Another example, the processes $AR(0)$, $MA(0)$ and $ARMA(0, 0)$ are all equivalent, and represent a standard white noise process.

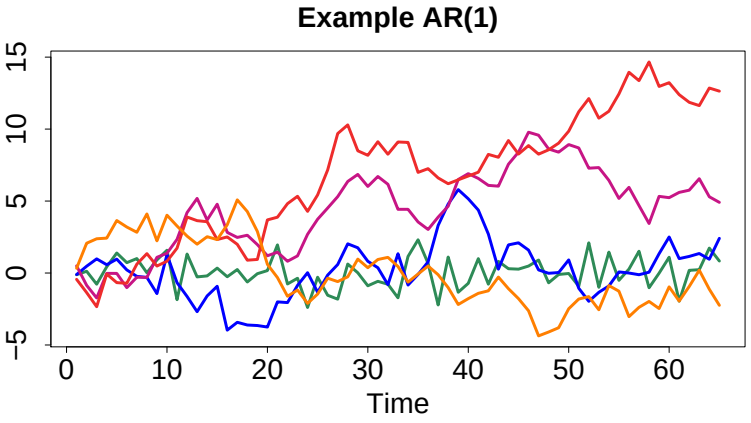

FIGURE 2.1: First-order autoregressive (AR(1)) series, as an example of a process that is serially correlated, and can also be non-stationary depending on how strong is the dependency of the series on its own past.

As an illustration, define $X_1(t)$ as an $AR(1)$ and $X_2(t)$ as an $MA(1)$. That is,

$$X_1(t) = \phi\, X_1(t-1) + \varepsilon(t) \tag{2.11}$$

$$X_2(t) = \varepsilon(t) + \theta\, \varepsilon(t-1) \tag{2.12}$$

where $\varepsilon(t)$ is a white noise process, and ϕ, θ are constants representing how strong is the dependency of the series on its past.

An $ARMA(1,1)$ process $X_3(t)$ would be a combination of the $X_1(t)$ and $X_2(t)$ above:

$$X_3(t) = \phi\, X_1(t-1) + \theta\, \varepsilon(t-1) + \varepsilon(t) \tag{2.13}$$

An ARMA process can be non-stationary, depending on the values of its paramenters ϕ and θ. In this case it is said that the process has an unit root. The term 'unit root' originates from the theoretical definition of the model, which is not the focus here. A more in-depth discussion can be found in a standard time-series analysis textbook (e.g., Box and Jenkins, 1970, Davidson and MacKinnon, 2004).

2.4 Tests for stationarity and first differences

Some tests to check for stationarity of a time-series are the Augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979a) and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS)

(Kwistkowski et al., 1992) test. The ADF test has a null hypothesis of non-stationarity, while the KPSS test has a null hypothesis of stationarity. Using both tests decreases the chance of a false-positive result, since for a series to be considered stationary the null hypothesis must be rejected in the ADF test, and not rejected in the KPSS test. When used in this thesis, the tests were performed using the tseries package, version 0.10-34 (Trapletti and Hornik, 2017), in the R software (www.r-project.org, version 3.4.1).

In addition to the ADF and KPSS tests, the Kim-Perron test can be used to identify the presence of unit roots when there is evidence that a a change-point exists in the series (Kim and Perron, 2009). The null hypothesis of this test is that the time-series is not stationary, and therefore rejecting the null hypothesis implies the stationarity of the series.

When the time-series show evidence of non-stationarity, a usual strategy is to use first differences to stabilize the data. The process of first differencing a time-series is self explanatory: the first-differenced version of a series $X(t)$ is:

$$\Delta X(t) = X(t) - X(t-1) \tag{2.14}$$

First differences represent the rates of change in a variable $X(t)$, and are a common way to stabilize the mean of a time-series by removing trend and seasonality (Hyndman and Athanasopoulos, 2012).

2.5 Multiple linear regression

A standard multiple linear regression (MLR) with normal error (white noise) can be used to describe a single variable as a function of multiple regressors. Suppose a variable $X_0(t)$ is dependent on $X_1(t)$ and $X_2(t)$. In this case, the MLR model would be:

$$X_0(t) = \alpha + \beta t + \gamma_1 X_1(t) + \gamma_2 X_2(t) + \varepsilon(t) \tag{2.15}$$

where $\alpha, \beta, \gamma_1, \gamma_2$ are constants, and $\varepsilon(t)$ is a white noise process.

Autoregressive models, such as the ones described in section 2.3, are useful to describe time-series which have internal variability due to serial correlation. However these models do not consider the influence of other variables on the time-series. On the other hand, traditional linear regression takes into account external variables but not the presence

of serial correlation. By combining these two modelling techniques, one can benefit from both information from external regressors and internal dynamics when describing a time-series.

Including autocorrelated noise in MLR models (sometimes called dynamic regression, Hyndman and Athanasopoulos, 2012) is a simple and intuitive way to describe complex time-series as a function of multiple regressors. For example, if $X_0(t)$ is a serially correlated process and is dependent on $X_1(t)$ and $X_2(t)$ as regressors, a MLR with ARMA errors would be written as:

$$X_0(t) = \alpha + \beta t + \gamma_1 X_1(t) + \gamma_2 X_2(t) + u(t) \tag{2.16}$$

where $\alpha, \beta, \gamma_1, \gamma_2$ are constants, and $u(t)$ is a stationary $ARMA(p, q)$ process (section 2.3).

The order (p, q) of $u(t)$ is determined by empirically varying the parameters of an ARMA process, and then using a Ljung-Box (LB) test (Ljung and Box, 1978) to determine if the residuals present signs of serial correlation after adjusting for the specified ARMA error. The null hypothesis of a LB test is that there is no serial correlation in the residuals. Among the models that pass the LB test, the one with best BIC score is selected (equation 2.2 in section 2.2).

The MLR with white noise error (equation 2.15) represents the hypothesis that $X_0(t)$ could be described directly as a function of $X_1(t)$ and $X_2(t)$ only, whereas the MLR with ARMA error (equation 2.16) accounts for the possibility of a more complex structure or endogenous variability in the system through the serial correlation in the error term.

It is worth noting the difference between "error" and "residual". "Error" refers to the error structure of the model, that is, the theoretical noise component in the model which is not explained by the other terms. For example, in the MLR model (2.16) the error is $u(t)$, an ARMA noise; and in MLR model (2.15) the error is $\varepsilon(t)$, a white noise. "Residual" refers to the difference between the observed data and the values estimated by the model. Ideally, if a model describes well a dataset, the residuals will be small and uncorrelated, no matter what kind of error structure is present in the model. For a more detailed definition, refer to DeGroot and Schervish (2002) or another similar standard statistics textbook.

2.6 Vector autoregressive (VAR) models

Vector autoregressive (VAR) models (Lütkepohl, 2007) are a multivariate linear extension of autoregressive (AR) models, often applied in the econometrics literature to describe and forecast financial time-series. In this thesis, they are used in chapter 4 to describe the interactions between phytoplankton and zooplankton, as well as the influence from environmental covariates, seasonal cycle and serial correlation in the plankton. VAR models are sometimes referred to as Multivariate Autoregressive (MAR) models in the literature. The terms are synonyms, and I use VAR throughout this thesis.

One of the benefits of VAR analysis is its flexibility, usually with few or no transformations required in the original data in order to meet the theoretical assumptions of the model. The necessary conditions for the usage of VAR models are the stationarity of all time-series, and that the relationship between the series be linear. Stationarity means that statistical properties of time-series such as mean and variance do not change with time (see section 2.4).

A limitation of VAR models is that they only capture linear relationships among the variables considered. If nonlinear relationships are present they may not emerge clearly in the analysis and could potentially contribute towards the autocorrelation component, since, for nonlinear data, a good approximation for the nonlinearity is usually the past of the time-series itself.

An example of a theoretical VAR model is given below, for a system consisting of n endogenous variables $X_1(t), \ldots, X_n(t)$, an external (or exogenous) variable $R(t)$ and included seasonal cycle $ssn(t)$. This model is applied to the plankton system in chapter 4, and its description can be found in section 4.1.3.

$$X(t) = \sum_{j=1}^{L} \mathbf{A_j}\, X(t-j) + \sum_{k=0}^{L} b_k\, R(t-k) + ssn(t) + \varepsilon(t) \tag{2.17}$$

- $X(t)$ is a $(n \times 1)$ vector containing the n endogenous variables as its coordinates, and $R(t)$ an exogenous variable.
- L is a constant, representing the extent of the memory in the system. That is, the maximum lag at which the past still influences the present system.

- $\mathbf{A_j}$, for $(j = 1, \cdots, L)$, are $(n \times n)$ coefficient matrices, representing the effects of past values of X to the present of X.
- b_k, for $(k = 0, \cdots, l)$, are $(n \times 1)$ coefficient vectors, representing the effects of the exogenous series and its past to the present of X.
- $ssn(t)$ is a $(n \times 1)$ periodic vector, containing a seasonal constant term for each variable in X, in each month of the year. This term ensures that relationships inferred from the coefficient matrices A_j and b_k in the model are distinct from the influence originated from the common annual seasonal cycle.
- $\varepsilon(t)$ is a $(n \times 1)$ white noise process, with zero mean and constant variance, that accounts for random fluctuations in the system.

2.7 Identification of nonlinearity, causality and chaos

The VAR methods described in the previous section are a class of stochastic models. A system is said to be purely stochastic when all of the variability observed is a consequence of random processes. Identical starting conditions do not yield identical outcomes. On the contrary, it is almost certain that the outcomes would be different for two trajectories with identical starting conditions in a stochastic system. An example of a purely stochastic system is the number of people waiting to be served in a queue.

On the other hand, a system is said to be deterministic when identical starting conditions always yield identical results over time. A deterministic system is said to be chaotic when it produces considerably different outcomes with the slightest difference in initial conditions, that is, it is very sensitive to perturbations, no matter how small. Throughout this text, the word "chaos" refers to this concept of deterministic chaos. The presence of chaos is a consequence of strong nonlinearity in the system, and can generate irregular patterns that may look like random fluctuations, even though they arise from the deterministic nature of the system. An example of a chaotic system is the weather. Chaotic systems are by definition nonlinear, but nonlinear systems are not necessarily chaotic.

Although there is no consensus on the formal theoretical definition of deterministic chaos (Gleick, 1987), it is possible to identify chaotic systems through the irregularities they present. One hallmark of chaotic systems is their sensibility to perturbations: any two

trajectories with initial conditions arbitrarily close to each other will diverge exponentially with time (Ascioti et al., 1993, Cvitanović et al., 2016, Sugihara and May, 1990, Zounemat-Kermani, 2016). This defining characteristic of chaos will be used in this thesis when investigating the presence of chaos in plankton time-series. Returning to the weather example, even though short-term forecasts may be reliable, a long-term weather forecast is practically impossible to be obtained. The accuracy of forecasts in chaotic systems decays exponentially to zero the further ahead in time one attempts to forecast, a consequence of the sensitivity in chaotic systems to perturbations. This does not happen in purely stochastic systems: in general, if a stochastic process is not autocorrelated, a guess for the state of the system tomorrow is just as good and uncertain as a guess for the state of the system in a month's time. When it comes to stochastic systems with serial correlation, the accuracy of forecasts may decay depending on how strong is the dependency on past values, however the speed at which the quality of forecasts decays is still slower than that of the case of deterministic chaos.

2.7.1 Identifying chaos: kNN forecasting

The kNN forecasting method is non-parametric, and only assumes that the time-series of interest have some sort of internal regular structure. This makes the method ideal for series which may be nonlinear or even chaotic. It can be used to investigate the presence of chaos in a time-series, through the assessment of how the accuracy of forecasts decay in relation to how far into the future the forecast is being made. This yields a way of identifying chaotic signatures, since the decay in forecast quality for chaotic systems is exponential. That is, if τ is the last observation in the time-series, and T is how far in the future one wishes to forecast ahead of τ, the quality of the forecast decays exponentially as T increases. For a more detailed description of the kNN forecasting method, refer to Ascioti et al. (1993), Reick and Page (2000) or Sugihara and May (1990). This approach is sometimes called "simplex projection" (McGowan et al., 2017).

The reasoning behind the method is as follows:

- Suppose we have τ observations of a single time-series $X(t)$ and we wish to forecast its next value $X(\tau + 1)$.
- Define a window of length m before observation $X(\tau)$, that is, an array $w_\tau = [X(\tau - m), \ldots, X(\tau - 1)]$ (m is called the embedding dimension).
- Proceed to look for similar windows in the past of the time-series which minimise the Euclidean distance $||w_\tau - w_*||$, and identify the k windows closest to window w_τ in that Euclidean sense.
- Take the average of the values following each of these k windows, and use that as the estimate for $X(\tau + 1)$. (2.18)

In order to use this method to investigate the presence of chaos, it is necessary to divide the series into a training set, which will be used as the available information to perform forecasts, and a testing set, which will be compared with the forecasts obtained from the training set.

In formal notation, let $X(t)$ be a time-series with length n, that is, $t = 1, \ldots, n$. Split the time-series into a training set $X_{train}(t)$, $t = 1, \ldots, \tau$ and a testing set $X_{test}(t)$, $t = \tau + 1, \ldots, n$. To look for the chaos signature, $X_{train}(t)$ is used to perform forecasts of $X_{test}(t)$, and the forecast skill is assessed by comparing the forecasted values with the actual values from $X_{test}(t)$. Following Sugihara and May (1990), I vary the window length m and define the number of neighbours as $k = m + 1$. The varying parameters are m, the window length, and T, how far into the future the forecast is performed with the information currently available.

The fact that arbitrarily close starting points will generate trajectories that diverge exponentially with time in chaotic system implies that, for this type of system, the forecast quality decays exponentially to zero the further ahead one tries to make predictions, regardless of the window length or number of neighbours considered (Ascioti et al., 1993, Sugihara and May, 1990). Figure 2.2 exemplifies patterns produced by the kNN method when applied to different types of time-serie: (a) a purely random white noise, (b) an auto-regressive moving-average (ARMA), (c) a periodic with added noise, and (d) a deterministic chaotic time-series.

2.7.2 Identifying nonlinearity: S–maps

Sugihara (1994) proposed the method of sequential locally weighted global linear maps, or S-maps for short, as a way of identifying the presence of nonlinearity in ecological systems, provided that the time-series are stationary. Similar to kNN, S-maps are also a nearest-neighbours method, but use weighted vectors of the reconstructed system to estimate the forecast.

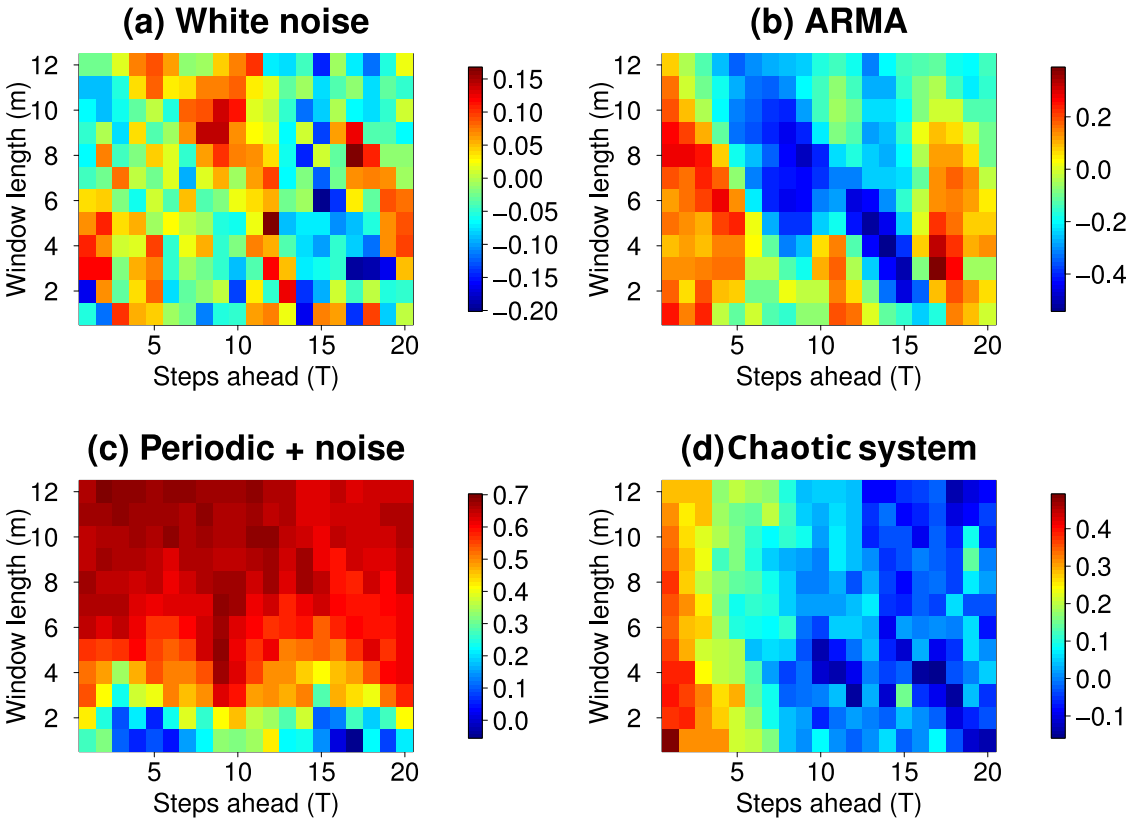

FIGURE 2.2: Results from the kNN method, applied to four example time-series. The colour scale on each panel corresponds to the correlations between the testing set and forecasted values, for different combinations of window length m and time step T. **(a)** Results for a randomly generated standard normal time-series (white noise). The lack of predictability causes no clear pattern to be seen on the map and weak correlations. **(b)** An auto-regressive moving-average (ARMA) stochastic time-series (section 2.3). Even though this is also a random series, the presence of a more defined structure from the ARMA yields a more regular pattern than the white noise. **(c)** A periodic time-series with added noise. The forecast quality improves with the window length m regardless of the value of T. The regular seasonal structure is easier to forecast than a pure noise process, and as a consequence the correlations are stronger in this case. **(d)** A chaotic time-series from the Lorenz attractor. Note the rapid decay in forecasting quality as T increases for all values of m, which is characteristic of deterministic chaos.

The procedure to identify the k nearest windows in S-maps is the same as in the kNN method outlined in 2.18 in the previous section. The difference is that, at the last step, a weighted average of the values is taken, according to the location of these k windows. The furthest away the neighbour, the less it contributes to the forecast estimate, and the averaging weights are proportional to the inverse euclidean distance between the neighbour and the value being forecasted. A specific parameter θ controls the variation of the weights with distance, which reflects how nonlinear the system is. The weighting function is defined as (Sugihara, 1994):

$$weight(d) = e^{-\theta(d/\bar{d})} \tag{2.19}$$

where d is the distance between the present and the neighbour value, and $\bar{d}$ is the average distance between neighbours.

θ equal to zero is optimal for linear time-series, as it represents a uniform weighting over all neighbours. Larger values of θ correspond to nonlinear scenarios, since local dynamics are more important in this scenario (Sugihara, 1994, Ye et al., 2015). If a series is linear, the S-maps show a monotone decrease in forecast quality with increasing values of θ. When the series is nonlinear, the forecast quality increases at least for some subset of values of θ larger than zero. In the particular case of chaotic systems, the S-maps show a monotone increase in forecast quality with θ (Sugihara, 1994).

For more details on the calculation of S-maps, refer to Sugihara (1994) and McGowan et al. (2017). In this thesis I use the algorithms from the "rEDM" package (version 0.4.7) in R to calculate the best embedding dimension m and S-maps for each time-series.

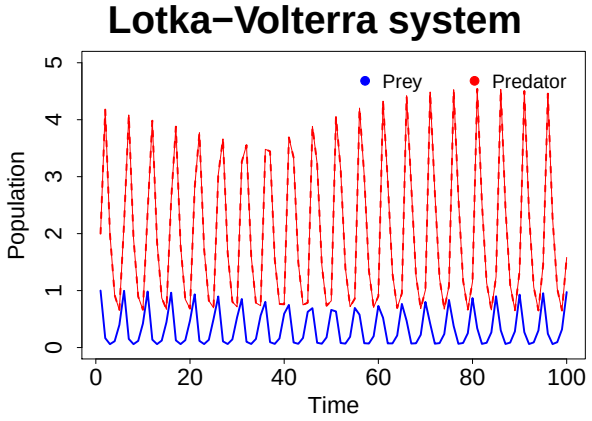

FIGURE 2.3: A fictional example of a predator-prey system generated by equations 2.20.

As an illustrative example, let us consider M a periodic, nonlinear and deterministic Lotka-Volterra predator-prey system (figure 2.3):

$$M \begin{cases} \dfrac{dX}{dt} &= a\,X - b\,X\,Y \\ \dfrac{dY}{dt} &= c\,X\,Y - d\,Y \end{cases} \tag{2.20}$$

where $X(t)$ and $Y(t)$ are the populations of prey and predators at time t, respectively.

This system is not chaotic, and thus it is expected that the forecast skill of the kNN method will not decay with T. However the system is nonlinear, therefore it is expected that the forecast quality of the S-maps will improve for positive values of θ. These results can be clearly seen in figure 2.4.

Now consider a simple variation on the predator-prey system 2.20. Suppose that a white noise signal is added to both the prey and predator components (representing, for example, a measurement error or a response to some external source of variability). The series with added noise are still periodic nonlinear and not chaotic, but are also no longer deterministic. Repeating the kNN and S-map analysis on this system shows once again no evidence of chaos and evidence for nonlinearity (figure 2.5), although the latter presents a different signature than the one obtained from the purely deterministic system.

2.7.3 Identifying causality: convergent cross–mapping

Convergence cross-mapping (CCM, Sugihara et al., 2012) is used in chapter 5 to investigate causal relationships between the plankton and environmental variables. When ecological systems are nonlinear, and standard statistical procedures such as correlations are often misleading and can yield innacurate results due to the assumption of linearity that is a pre-requisite of these methods. Ecological systems are also usually subject to noise, but are seldom purely stochastic. Most systems posess some underlying deterministic dynamic structure, even if said structure is not evident or easily identified.

When nonlinear systems are to some extent governed by a deterministic structure, it is possible to qualitatively reconstruct the dynamics of the system by using lagged versions of a single variable that is part of the system (Takens, 1981). For example, consider again the predator-prey system M (equation 2.20). It is possible to reconstruct a version M_X of

the system using $\{X(t), X(t-1)\}$ that maps one-to-one into M. Similarly, it is possible to reconstruct a version M_Y of the system using $\{Y(t), Y(t-1)\}$ that also maps one-to-one into M. This means that it is possible to define a one-to-one mapping between M_X and M_Y such that local neighbours in M_X are also local neighbours in M_Y. Figure 2.6 exemplifies this process for the Lotka-Volterra system, and figure 2.7 shows the reconstruction of the system with added white noise, illustrating how the reconstruction is still possible when the system has some degree of added stochasticity.

The main implication of this fact is that it is possible to use the reconstructed system M_X to estimate the series $Y(t)$, or the reconstructed system M_Y to estimate the series $X(t)$, in the case when $X(t)$ and $Y(t)$ are mutually related. That is, it is possible to recover information about $Y(t)$ from M_X and vice-versa, and in this case one says that M_X cross maps Y and M_Y cross maps X. This bilateral mapping implies a bilateral relationship between $X(t)$ and $Y(t)$. Figure 2.8 illustrates the CCM between predator and prey in the example system. There are also cases in which information about $Y(t)$ is contained in $X(t)$ (M_X cross maps Y) but $Y(t)$ is insensitive to $X(t)$ (M_Y does not cross map X). This last scenario corresponds to a unilateral relationship and suggest that Y would be an external forcing to X in the system.

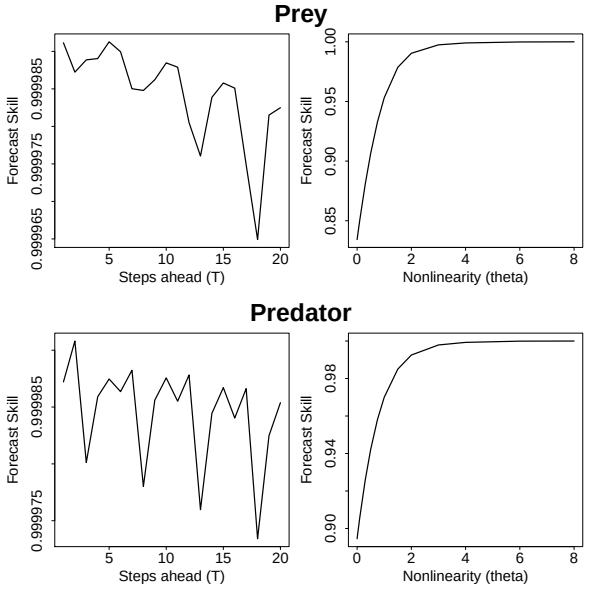

FIGURE 2.4: S-maps for the original predator-prey system. For a fixed optimal value of the embedding dimension m the forecast skill is calculated with varying values of parameters T, how far in the future the prediction is being made (left-hand panel), and θ, the nonlinearity parameter from the S-map analysis (right-hand panel). The forecast skill remained almost constant with T, indicating the lack of a chaotic signature. However, a noticeable improvement in skill can be seen for increasing values of θ, indicating nonlinearity. This leads to the expected conclusion that the predator-prey system defined by the system 2.20 is nonlinear and not chaotic.

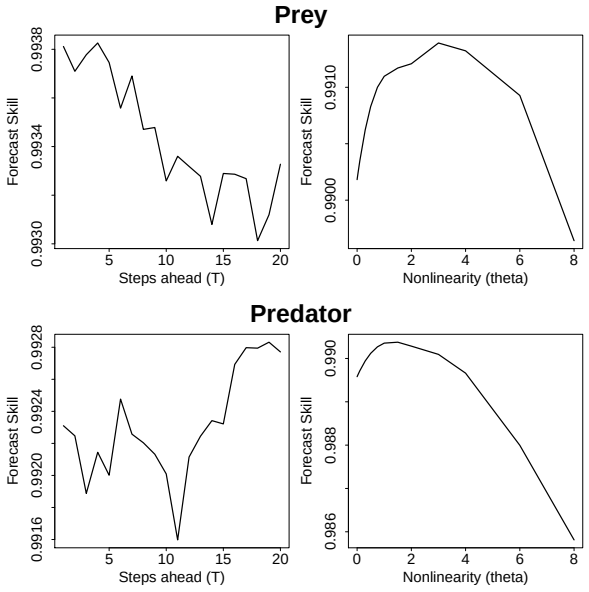

FIGURE 2.5: S-maps for the predator-prey system with added noise. For a fixed optimal value of the embedding dimension m, the forecast skill was calculated with varying values of parameters T (left-hand panel) and θ (right-hand panel). Once again the skill remained fairly constant with T, indicating that the system is not chaotic. The initial improvement of the skill with increasing values of θ also leads to the expected conclusion that the system is nonlinear, even if the skill later decreases for larger values of θ. This eventual decrease is a consequence of the added noise in the system.

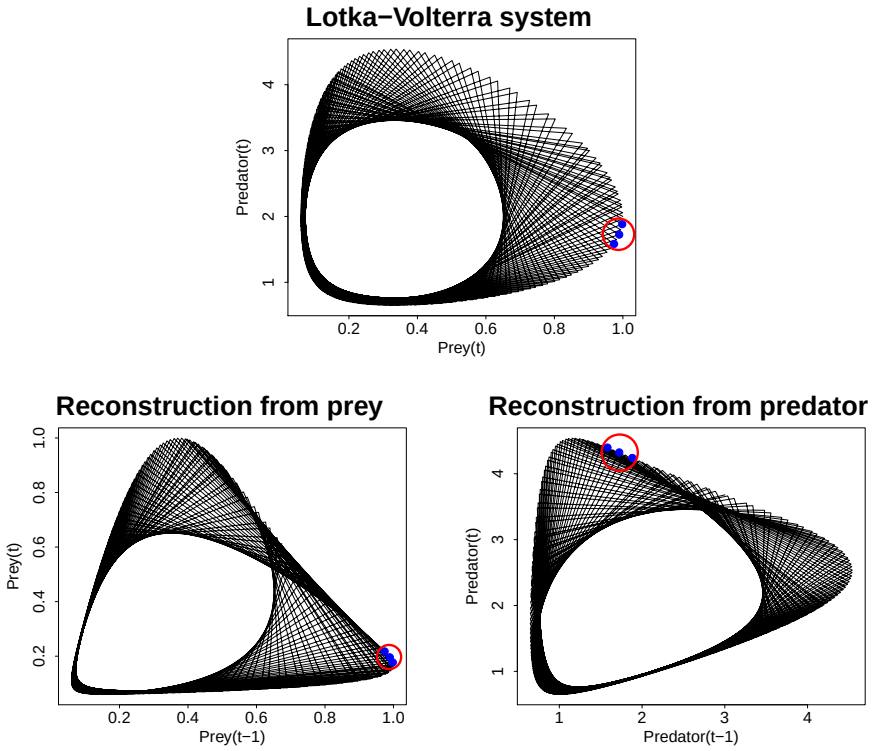

FIGURE 2.6: Original predator-prey system M and its reconstructed versions M_X and M_Y using lags of either preys or predators only. The original and reconstructed systems are homeomorphic, and the reconstructed systems preserve neighbour points.

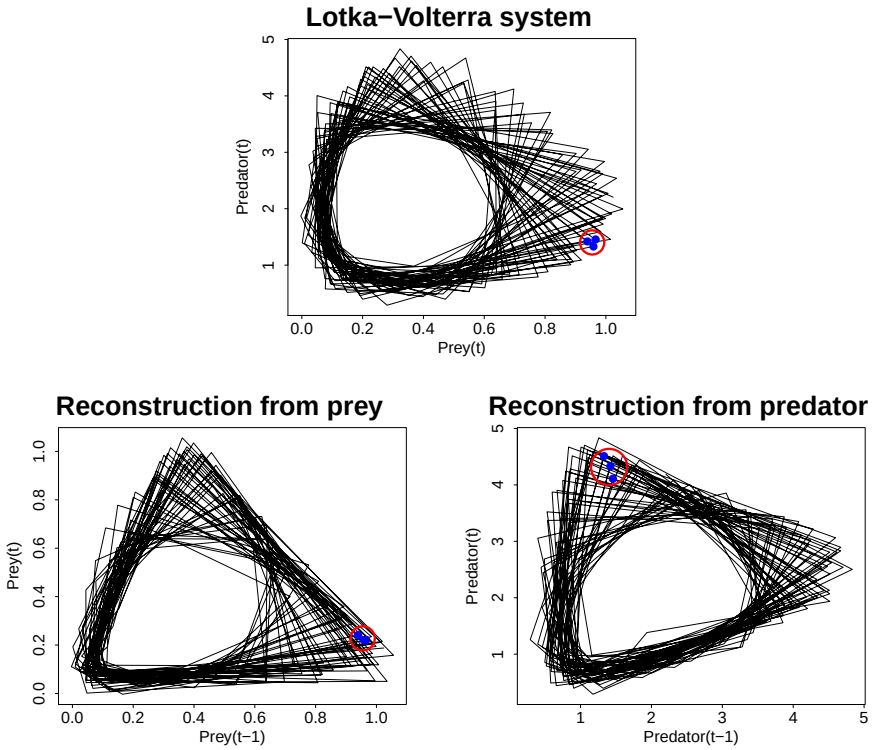

FIGURE 2.7: Predator-prey system with added stochasticity, and its reconstructed versions using lags of either preys or predators only.

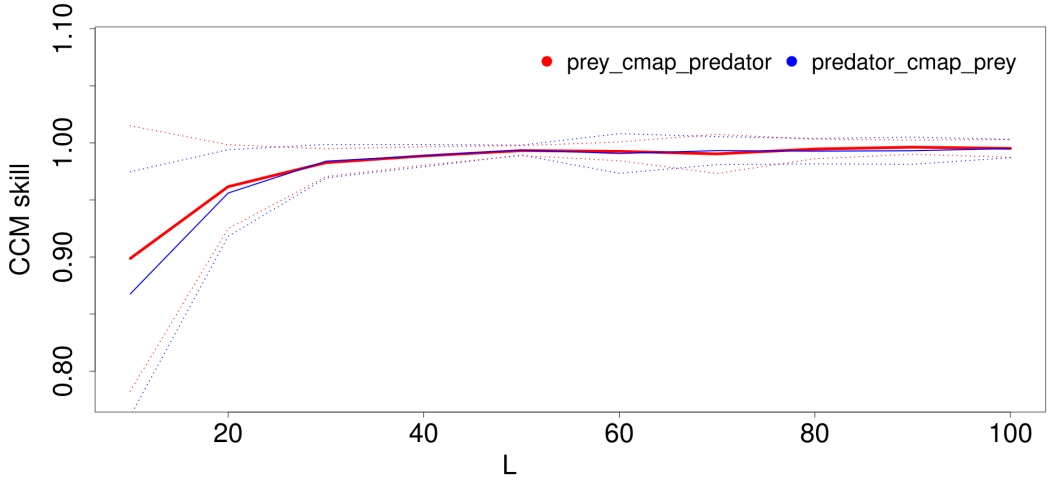

FIGURE 2.8: Convergence cross-mapping (CCM) between predators and preys in the example Lotka-Volterra system. The overlap between the CCM curves from X to Y and from Y to X indicate a bilateral relationship: both predator and prey contain information about one another. The high skill indicate that the variables are strong predictors of one another, which is expected from the definition of the predator-prey system. The parameter L corresponds to the length of the subsection of the series which is used to perform forecasts and determine the convergence.

Chapter 3

Water greenness as a function of diatom and dinoflagellate abundance

The phytoplankton colour index (PCI) from the Continuous Plankton Recorder (CPR) survey is a visual index of water greenness (section 2.1 in chapter 2). It has been used in the literature as a proxy for phytoplankton biomass and chlorophyll concentration (e.g., Batten et al., 2003, Helaouët et al., 2013, Raitsos et al., 2005), including as a component in the identification of marine regime shifts in the North Sea and Northeast Atlantic (Beaugrand et al., 2002, Goberville et al., 2014, McQuatters-Gollop et al., 2007).

Understanding the relationship between the PCI and phytoplankton abundance is important in order to improve our knowledge of how the variability in chlorophyll concentration is reflected in the phytoplankton community. Diatoms and dinoflagellates are the main phytoplankton groups captured by the CPR survey (Batten et al., 2003), and groups of smaller size tend to be underestimated due to the sampling mesh size (270μm, section 2.1.1). Therefore, since the smaller phytoplankton species are undersampled, it would be reasonable to expect diatoms and dinoflagellates to be the main contributors towards the PCI. On the other hand, it is known that smaller phytoplankton can still influence the PCI through their broken chloroplasts that stain the silk (Batten et al., 2003). If the concentration of smaller phytoplankton is large, they could potentially impact the PCI, even though they are not individually counted in the survey.

Regime shifts are known to have occurred in the region of study during the recent past. A few examples are, for instance, the studies of Beaugrand (2004), which reported two ecological shifts in the North Sea during the 1960s and 1980s, related to large scale

hydro-meteorological forcing. Later, AlvarezFernandez et al. (2012) and Beaugrand et al. (2014) identified another shift in the same region in the late 1990s. The 1980s shift is particularly well studied, due to its impact across ecological levels, and has been observed in biological and environmental variables, such as phytoplankton (PCI, diatoms), zooplankton, fish recruitment, temperature, salinity, wind intensity and the North Atlantic Oscillation (NAO) index. Hinder et al. (2012b) detected a shift in the relative abundance of diatoms versus dinoflagellates in the Northeast Atlantic, between 1995 and 2005. The authors reported a decline in dinoflagellate abundance, whereas diatom populations seemed to remain constant. Harris et al. (2015) identified a rapid northward movement of calanoid copepods in the North Sea and Northeast Atlantic as a consequence of rising SST, and suggested there had been a regime shift and reorganisation of the entire ecosystem over the past decades, although the precise timing of the shift varied across the different geographical regions of this study.

In addition to its involvement in the identification of regime shifts, the PCI has been shown to be increasing in the past fifty years in the North Sea and Northeast Atlantic (McQuatters-Gollop et al., 2011, Raitsos et al., 2013b), leading to an inference of increased phytoplankton biomass in this region. However, there are few studies examining the contribution of different phytoplankton species to the PCI. The work by Leterme et al. (2006) is to my knowledge the only work published on this subject. The authors investigated different trends in the PCI and cell concentration of diatoms and dinoflagellates from the CPR time-series. The authors use standard multiple linear regression to investigate the contribution of diatoms and dinoflagellates to the PCI. They proposed three different ecosystem states in the region: stable in the central and southern North Sea, an evolving ecosystem with increased contribution from dinoflagellates in the northern North Sea and southern Northeast Atlantic, and an ecosystem which has switched between two stable states in the northern Northeast Atlantic, with a decrease in the dominance of diatoms. However, the details on how these calculations have been performed are not clear, nor is how they compute the temporal changes in the contribution of the plankton abundance variables. Here I address these issues with a clear statistical modelling, and rigorous change-point detection analysis.

In this chapter, I quantify the relationship between diatom and dinoflagellate abundance and the PCI, and investigate the contribution from diatoms and dinoflagellates towards the PCI in the North Sea and Northeast Atlantic. Using multiple linear regression

models, I argue that those two phytoplankton groups alone cannot explain the recent past changes observed in the PCI, implying and reinforcing that the influence of smaller-sized phytoplankton to this index is not negligible. In addition, I also investigate the presence of trends and abrupt transitions in the time-series of the PCI and abundance of diatoms and dinoflagellates, in order to identify whether the known shifts in the PCI are reflected in diatoms and dinoflagellates as well.

3.1 Methods

3.1.1 Region of study and data

I consider time-series of the phytoplankton colour index (PCI), and abundance of diatoms and dinoflagellates. The series have been collected by the CPR survey, maintained by SAHFOS (chapter 2, section 2.1). The dataset consists of monthly observations from January 1958 to December 2007. The original arithmetic time-series of the PCI and phytoplankton variables have been set to the same scale by standardizing each series into z-scores, with mean zero and standard deviation one.

I used the CPR standard areas covering the waters of the Northeast Atlantic and North Sea, as shown in figure 3.1. The areas were grouped in larger regions, according to table 3.1.

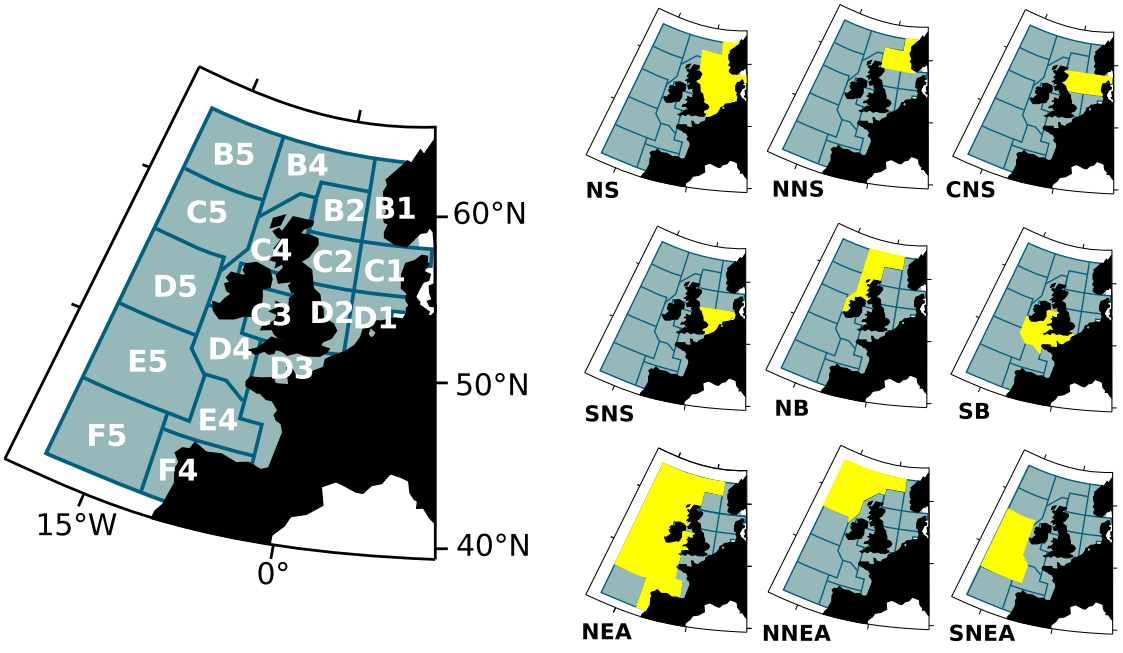

FIGURE 3.1: CPR standard areas, and highlighted grouped regions considered in this chapter.

TABLE 3.1: Grouped standard areas of the CPR survey. Area F5 presented too many missing observations and was not considered in the analysis.

Region name	Grouped CPR areas
NS: North Sea	B1 B2 C1 C2 D1 D2
NNS: Northern NS	B1 B2
CNS: Central NS	C1 C2
SNS: Southern NS	D1 D2
NB: North of Britain	B4 C4
SB: South of Britain	C3 D3 D4
NEA: Northeast Atlantic	B4 B5 C4 C5 D4 D5 E4 E5 F4
NNEA: Northern NEA	B4 B5 C5
SNEA: Southern NEA	D5 E5

The choice of how to group the regions was based on previously suggested divisions from the literature (e.g., Edwards et al., 2001, Leterme et al., 2006), and with the intent of comparing similarities and differences within larger areas. For instance, the North Sea (NS) is considered as a whole unit, and also subdivided in a northern (NNS), central (CNS) and southern (SNS) portions. The Northeast Atlantic (NEA) was subdivided into northern (NNEA) and southern (SNEA) portions, avoiding standard areas which contain coastal regions where possible, and in a region containing the English Channel and waters in the south of the British isles (SB). The waters to the north of the British isles (NB) were also considered separately, as this is a region of transition between the oceanic Northeast Atlantic and the North Sea, and can potentially present different phytoplankton dynamics (Edwards et al., 2001).

Averaging over larger regions also resolved the issue of missing values in the dataset. For each monthly observation in time, if more than half of the cells composing the region were missing, the region-averaged observation was considered as missing as well. Then, annual averaged series were calculated for every region (figure 3.2). When a year had more than five months of missing values, or three consecutive months of missing values, then the whole year was considered as missing. As in all chapters in this thesis, the time-series have been standardized into z-scores, with mean zero and standard deviation one.

3.1.2 Methods

The first goal of this chapter was to investigate whether known shifts in the PCI are reflected in the diatom and dinoflagellate abundance.

As previously mentioned, the PCI has been used as a proxy to represent phytoplankton populations in regime shift analyses (e.g., Beaugrand, 2004). However, with the exception of Leterme et al. (2006), I am not aware of any other studies comparing the trends observed in the PCI with those of individual plankton groups, and none including a discussion on whether the abrupt changes observed in the PCI are also present in diatoms and dinoflagellates. Here I used the Informational Change-Point Detection (ICPD) method to investigate the presence of abrupt changes in the plankton time-series (Beaulieu et al., 2012a,b, and section 2.2), an approach based on information criteria that allows for the identification of change-points in the mean or linear trend of a time-series.

One drawback of the ICPD method is that, to my knowledge, it is not yet possible to calculate confidence intervals for the estimated timing of the change-points. This could lead to some uncertainty regarding the exact timing of a change-point identified in noisy time-series. Therefore, the output of the ICPD analysis was treated as a guideline for the timing of the change-points. For example, if the ICPD identifies a change-point in 1985, I say that

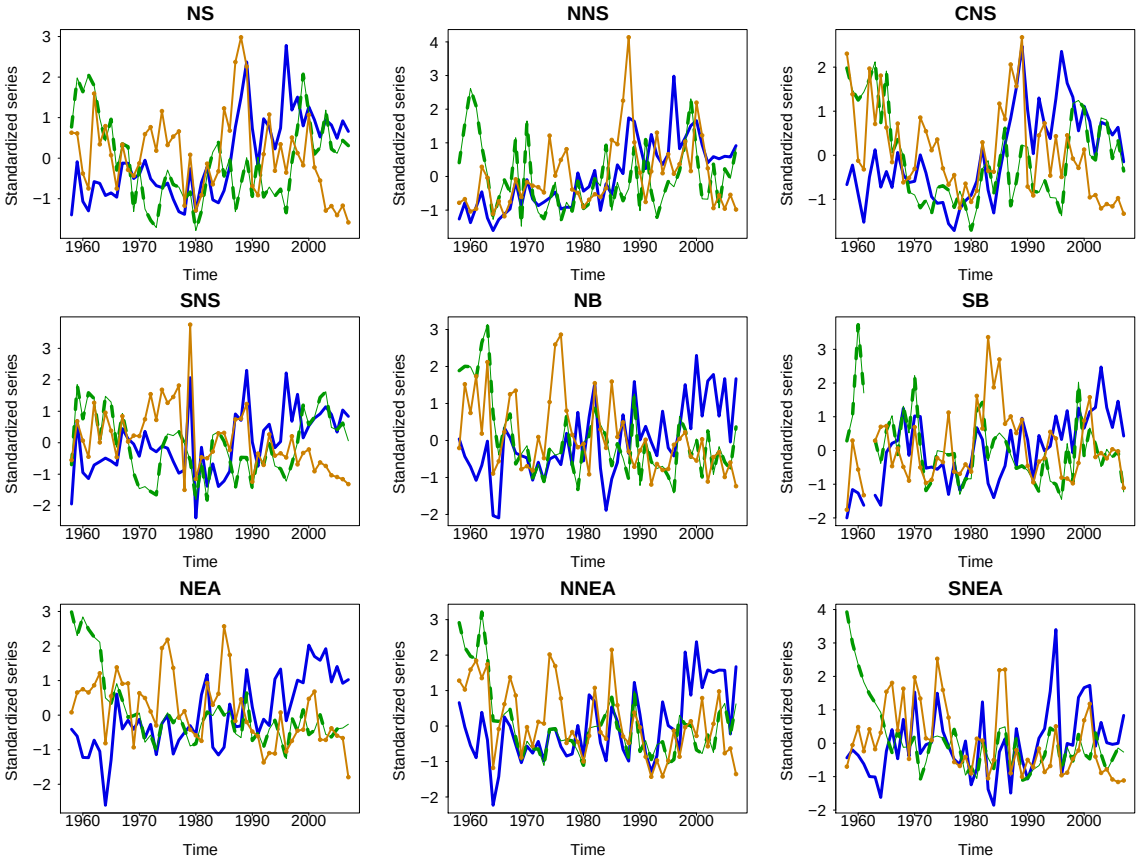

FIGURE 3.2: Annual standardized averaged PCI (blue), diatom (green) and dinoflagellate (dark yellow) time-series. For a definition of the regions, please refer to table 3.1.

the series present a change-point in the mid 1980s.

The second goal of this chapter was to model the PCI as a function of diatom and dinoflagellate abundance, in order to investigate long-term relationships among these variables, and whether it was possible to infer the contribution of each individual plankton variable to the PCI. The time-series are always sampled at the same time and location, since they are measured from the same CPR silk. Therefore, it is reasonable to assume the relationship between those time-series should be direct, and consequently a multiple linear regression (section 2.5) was chosen to model the PCI as a function of the phytoplankton groups.

In order to avoid spurious results in the multiple regression (section 2.4), the first step was to check the stationarity of each time-series with the KPSS (Kwistkowski et al., 1992), ADF and ADF-GLS tests (Dickey and Fuller, 1979a, Elliott et al., 1996), and the Kim-Perron test for the cases when there was a change-point identified in the series (Kim and Perron, 2009). The null hypothesis of ADF, ADF-GLS and Kim-Perron tests is that the time-series are not stationary, while the null hypothesis of a KPSS test is that the series are stationary. Comparing tests with opposite null hypotheses decreases the chance of a biased or false positive conclusion. It is important to check the data for stationarity, since, if a regression is performed on non-stationary time-series, one can obtain significant results suggesting a relationship with a regressor variable when there is none (Hyndman and Athanasopoulos, 2012).

Once it was verified that all series met the stationarity criteria, I compared two different multiple linear regression (MLR) models: a standard MLR with normal (white noise) error term (equation 2.15), and a MLR with serially correlated (ARMA) error term (equation 2.16), as defined in section 2.5 of chapter 2.

The reasoning for comparing two models is the following. If diatoms and dinoflagellates were sufficient regressors to explain the variability in the PCI, the MLR with white noise should yield the best results. On the other hand, if there were missing variables, or some other unexplained source of variability contributing to the PCI, a MLR with ARMA error should perform better. The Ljung-Box (LB) test (Ljung and Box, 1978) was used to investigate if the residuals of both MLR models were uncorrelated. This test has a null hypothesis of no correlation among the observations in a dataset (that is, failing to reject the hypothesis implies no autocorrelation in the residuals), and is often applied to check if

residuals are uncorrelated when modelling time-series (Hyndman and Athanasopoulos, 2012).

3.2 Results

3.2.1 Exploratory analysis of the plankton time–series and identification of change–points

Overall, there was no matching among the timing or nature of the change-points identified in the PCI, diatom and dinoflagellate time-series. The trends observed in the abundance of phytoplankton also did not match those observed in the PCI.

In the North Sea (NS, figure 3.3, first line of table 3.2), there was a decrease in diatom abundance until the late 1960s, followed by a slow increase since. The spring bloom was clearly visible in most years, whereas a weaker autumn bloom seemed to be mostly absent during the years following 1970 until around year 2000. Dinoflagellates had a remarkable peak in the late 1980s, and the sudden decrease after 2000 was identified as a change-point in the mean of the time-series. The PCI, on the other hand, increased considerably in all months of the year after the change-point identified in the late 1980s. The decrease in dinoflagellate and diatom abundance, with the weakening of the diatom bloom in Autumn, did not seem to have any effect on the PCI levels.

Similar patterns were observed across the subdivisions of the North Sea (NNS, CNS and SNS, figures 3.3 - 3.4 and lines 2 to 4 of table 3.2). The PCI presented a change-point in the late 1980s in all three regions, reaching considerable higher levels after that. The change-point observed in the diatom abundance decreasing trend was also identified in CNS and SNS, however in NNS the abundance of diatoms was better represented by a constant mean. The change-point in dinoflagellate abundance seems to have been driven by populations from the central part of the North Sea, whereas in the north (NNS) the abundance of dinoflagellates presented a change-point in the trend (from increasing to decreasing in the late 1980s) and a constant decreasing trend in the south (SNS).

TABLE 3.2: Change-points (c.p.) and trends identified by the ICPD method in the Phytoplankton Colour Index (PCI), diatom and dinoflagellate time-series. The models that best represent each series were chosen based on the Bayesian Information Criterion (BIC, Schwarz, 1978). Yellow cells correspond to models with either constant mean or change-point in the mean, and purple cells correspond to models with either constant trend or change-point in the trend. When a change-point in the trend is identified, the signal of the trend before and after the change-point are indicated in brackets.

	PCI	Diatom	Dinoflagellate
NS:	c.p. mean 1987	c.p. trend 1969 $(-/+)$	c.p. mean 2002
NNS:	c.p. mean 1986	constant mean	c.p. trend 1985 $(+/-)$
CNS:	c.p. mean 1986	c.p. trend 1969 $(-/+)$	c.p. mean 2001
SNS:	c.p. mean 1987	c.p. trend 1969 $(-/+)$	decreasing trend
NB:	increasing trend	c.p. mean 1964	decreasing trend
SB:	c.p. trend 1972 $(+/+)$	c.p. mean 1971	c.p. trend 1981 $(+/-)$
NEA:	increasing trend	c.p. trend 1974 $(-/-)$	decreasing trend
NNEA:	increasing trend	c.p. mean 1964	decreasing trend
SNEA:	increasing trend	c.p. trend 1967 $(-/-)$	c.p. mean 1977

In the Northeast Atlantic region (NEA, figure 3.5 and seventh line of table 3.2), there was an overall decrease in diatom and dinoflagellate abundance since 1958. The diatom autumn bloom weakened through time, especially after the late 1960s. Dinoflagellate abundance had an overall slow decreasing trend. The dinoflagellate summer bloom weakened considerably after 1990, and, in spite of a small increase around year 2000, dinoflagellate abudance was still lower in 2007 than its past levels. The PCI, however, was identified as following a steady increasing trend with time.

Considering the subdivisions of NEA (figures 3.4 - 3.5 and lines 5, 6, 8 and 9 of table 3.2), the PCI presented a constant increasing trend similar to that observed in NEA in the northern (NNEA) and southern (SNEA) oceanic portions and on the north of the British isles (NB). Only in SB was a change-point observed, but still with an increasing trend. Dinoflagellate abundance was better represented by a decreasing trend in all subdivisions except for SB, where a change-point was observed in the early 1980s, when a transition occurred from an increasing to a decreasing trend. Diatom abundance presented the most distinct patterns. The change-point in the decreasing trend observed in the early 1970s for the larger region was not observed in any particular subdivision of the Northeast Atlantic. A change-point in the mean of diatoms was identified in the mid 1960s both in NNEA and NB, and in the early 1970s in SB. Only SNEA presented a change-point in the decreasing trend, identified in the late 1960s.

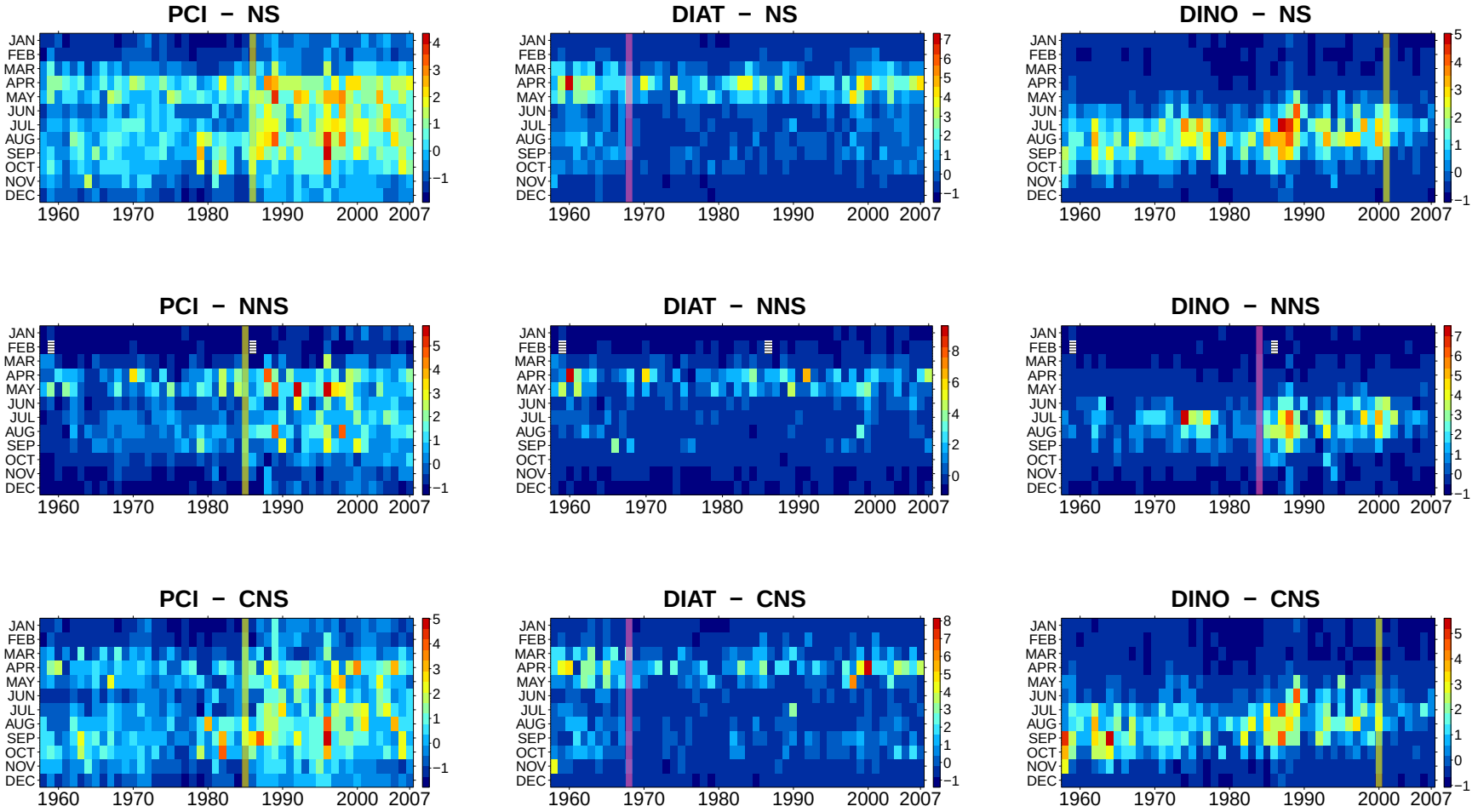

FIGURE 3.3: Standardized time-series of the PCI (left column), and abundance of diatoms (middle column) and dinoflagellates (right column). Change-points in the mean are highlighted in yellow, and change-points in the trend are highlighted in purple. White dashed cells indicate missing values. A change-point in the mean was identified in the PCI time-series of the North Sea (NS) in the late 1980s, which was also observed in the northern and central subdivisions of the region (NNS and CNS). Diatoms presented a change-point in the decreasing trend in the late 1960s in the whole NS, but this shift was not present in the north (NNS). Dinoflagellates had a change-point in the mean in the early 2000s (NS), which was also present in CNS. In NNS, however, had an increasing shift in the trend in the mid 1980s.

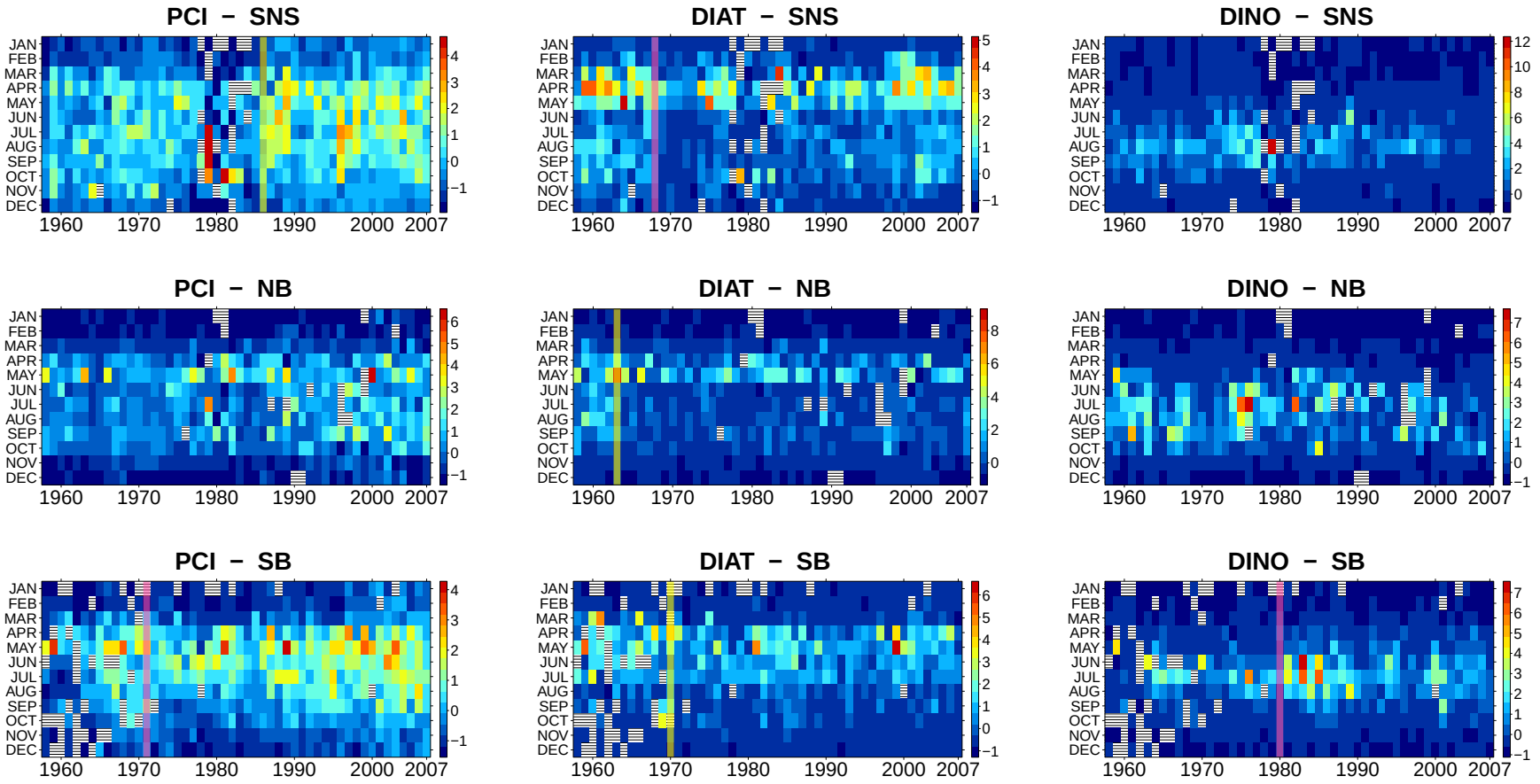

FIGURE 3.4: Standardized time-series of the PCI (left column), and abundance of diatoms (middle column) and dinoflagellates (right column). Change-points in the mean are highlighted in yellow, and change-points in the trend are highlighted in purple. White dashed cells indicate missing values. The PCI series of the southern North Sea (SNS) presented a change-point in the late 1980s, in agreement with what was observed in the other parts of the region. To the north of the British isles (NB) the PCI was better represented by an increasing trend, while to the south (SB) a change-point was detected in the trend. Diatoms had a change-point in the decreasing trend in SNS in the late 1960s, similarly to what was identified in CNS. In NB and SB, a change-poin in the mean was identified in the early 1960s and early 1970s, respectively. Dinoflagellates were better described by a linear decreasing trend in SNS and NB, and presented a shift in the trend in the mid 1980s in SB.

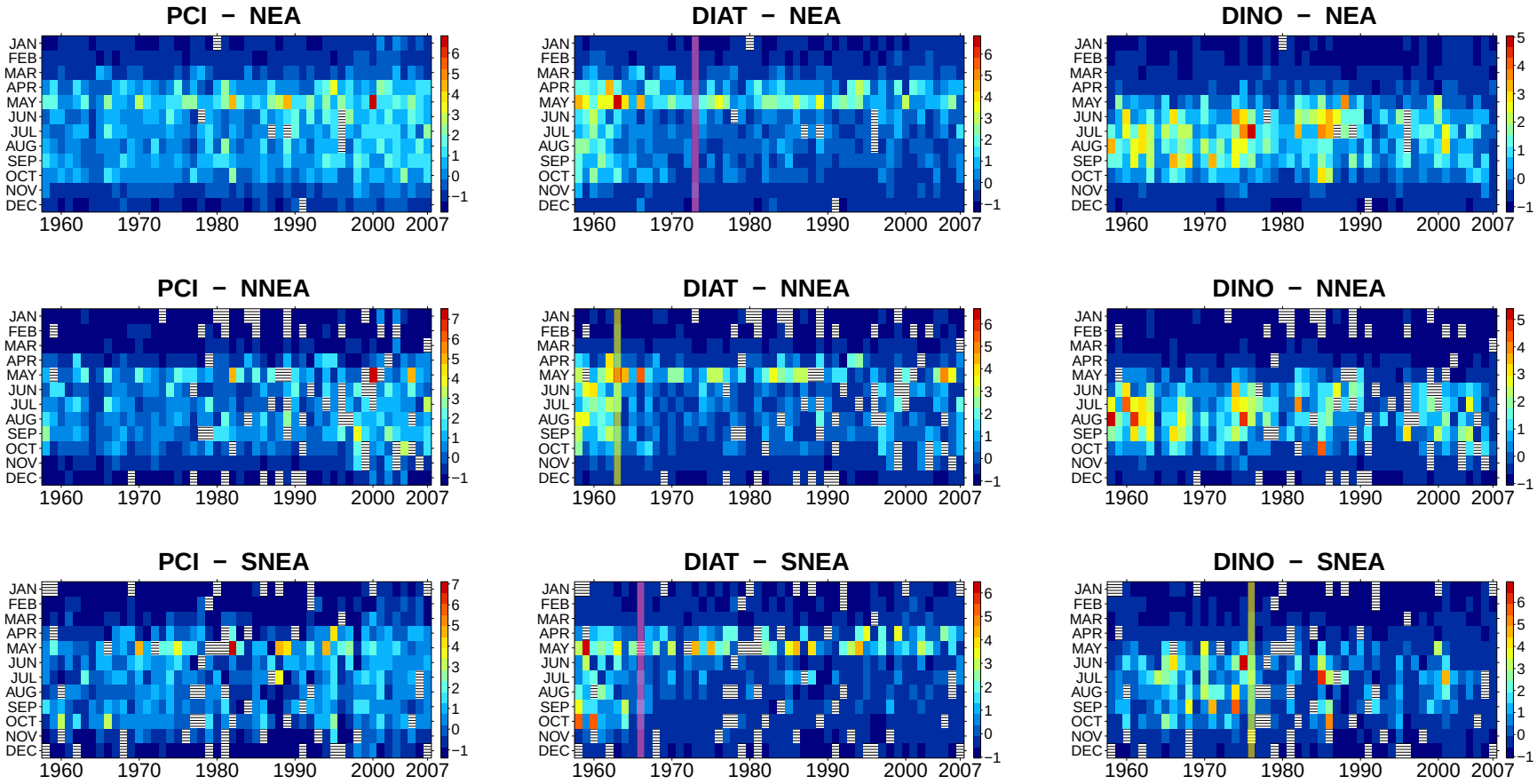

FIGURE 3.5: Standardized time-series of the PCI (left column), and abundance of diatoms (middle column) and dinoflagellates (right column). Change-points in the mean are highlighted in yellow, and change-points in the trend are highlighted in purple. White dashed cells indicate missing values. The PCI was better represented by an increasing trend both in the overall Northeast Atlantic (NEA) and its subdivisions to the north and south (NNEA and SNEA). Diatom abundance presented three distinct change-points. A decrease in the mean was identified in NNEA and a shift in the trend in SNEA, both in the mid 1960s. However, in the overall averaged NEA, a change-point in the trend was identified in the early 1970s.

Dinoflagellates were better described by a constant mean in NEA and NNEA, however they presented a change-point in the mean in SNEA.

3.2.2 The PCI, modelled from diatom and dinoflagellate abundance

Two multiple linear regression (MLR) models were fitted to assess if the PCI could be explained by the diatom and dinoflagellate abundance time-series, in each one of the regions considered. The models were a MLR with normal, white noise error term (equation 2.15) and a MLR with serially correlated, ARMA error term (equation 2.16), as described in section 2.5 of chapter 2.

The best results were obtained from the MLR with ARMA error. Figure 3.6 shows a visual comparison of the observed PCI annual time-series with the averaged outputs from reconstructions made through simulations using the MLR with ARMA error in each region. Each simulation was generated using the parameters estimated for the intercept, trend, diatom and dinoflagellate components and a realisation of the ARMA error component as defined by the MLR. The simulations were then averaged to produce an approximation of the PCI time-series. I varied the number of simulations up to ten thousand, however the convergence of the model was fast, such that there was little difference when using ten or ten thousand simulations for calculating the average approximation.

TABLE 3.3: Comparison of performance of the models with white noise (WN) error term and ARMA error term. Three measures are considered: residual mean squared error (RMSE), correlation (COR) between original and simulated series, and percentage of predicted trends (FPT), that is, the fraction of time steps for which the model correctly estimated an increase or decrease. Correlations marked with * were significant at a $p < 0.05$ level, and ** at a $p < 0.01$ level.

	NS	NNS	CNS	SNS	NEA	NNEA	SNEA	NB	SB
WN (RMSE)	0.14	0.12	0.12	0.14	0.11	0.13	0.12	0.13	0.12
ARMA (RMSE)	0.09	0.09	0.10	0.10	0.07	0.09	0.11	0.09	0.11
WN (COR)	0.19	0.53**	0.49**	0.23	0.57**	0.27	0.48**	0.36*	0.48**
ARMA (COR)	0.79**	0.75**	0.68**	0.68**	0.86**	0.79**	0.64**	0.73**	0.65**
WN (FPT)	0.67	0.49	0.65	0.63	0.65	0.78	0.59	0.67	0.65
ARMA (FPT)	0.65	0.61	0.71	0.65	0.76	0.63	0.65	0.69	0.76

The MLR with ARMA error yielded a good approximation to the original PCI time-series, as seen from table 3.3. The results have a low residual mean square error (RMSE), and there was a high correlation between estimated and original data in all regions considered. The best ARMA structure for the error term in this model was a first-order autoregressive ($AR(1)$) process in all regions except NEA, where a first-order moving-average ($MA(1)$) process was selected by the method. Both the $AR(1)$ and $MA(1)$ represent serially correlated processes dependent on their immediate past (see chapter 2, section 2.3). The residuals of the model also pass the LB test at the 5% significance level (table 3.4), failing to reject the hypothesis of uncorrelated residuals, which means there was no evidence of any remaining unexplained structure in the MLR with ARMA error.

The MLR with white noise error yielded poor results (table 3.5). The residuals were autocorrelated in most regions, with a p-value in the LB test smaller than 5%, rejecting the hypothesis of uncorrelated residuals. This result implies that the MLR with white noise error did not represent all of the variability in the system. The performance was inferior to that of the MLR with ARMA noise. Even though the RMSE is similar to that of the alternative MLR (table 3.3), the correlation between estimated and original data was lower, and even not significant in some regions.

This means that, when modelling the PCI as a function of time, diatom and dinoflagellate abundance, it is more appropriate to use a first-order serially correlated noise term instead of a normal, white noise term. The ARMA noise accounts for the variability that is not explained by the diatom and dinoflagellate time-series alone.

TABLE 3.4: Parameters of fitted MLR models with ARMA error (second to sixt lines) in each region, and $p-$values of the LB test applied to the residuals of the selected model in each region (first line). The null hypothesis of uncorrelated residuals could not be rejected at the 5% significance level in any region when using the $AR(1)$ or $MA(1)$ structures for u.

model (2.16): $PCI(t) = \begin{cases} \alpha + \beta t + \gamma_1 DIAT(t) + \gamma_2 DINO(t) + u(t) \\ \text{where } u(t) = \phi\, u(t-1) + \varepsilon \text{ or } u = \varepsilon + \theta\, \varepsilon(t-1) \end{cases}$									
	NS	NNS	CNS	SNS	NEA	NNEA	SNEA	NB	SB
LB test	0.20	0.08	0.35	0.31	0.35	0.07	0.12	0.10	0.23
α	-1.42	-1.27	-1.42	-1.32	-1.84	-1.59	-1.62	-1.48	-1.47
β	0.05	0.05	0.05	0.05	0.07	0.06	0.06	0.06	0.05
γ_1	0.05	0.24	0.19	0.13	0.32	0.36	0.47	0.42	0.34
γ_2	0.33	0.28	0.36	0.48	0.19	0.35	0.68	0.04	0.03
ϕ or θ	0.43	0.13	0.58	0.19	0.49	0.12	0.41	0.13	0.48

TABLE 3.5: Parameters of fitted MLR models with white noise error (second to fourth lines) in each region, and $p-$values of the LB test applied to the model residuals in each region (first line). The null hypothesis of uncorrelated residuals is rejected at the 5% significance level in almost all regions, with the exceptions being SNEA and NB.

model (2.15): $PCI(t) = \alpha + \beta t + \gamma_1 DIAT(t) + \gamma_2 DINO(t) + \varepsilon(t)$									
	NS	NNS	CNS	SNS	NEA	NNEA	SNEA	NB	SB
LB test	<0.01	0.03	0.03	<0.01	<0.01	0.02	0.12	0.08	<0.01
α	-1.43	-1.68	-1.44	-1.38	-1.27	-1.53	-1.28	-1.58	-1.40
β	0.06	0.06	0.06	0.05	0.05	0.06	0.05	0.06	0.05
γ_1	0.09	0.23	0.38	0.30	0.24	0.10	0.11	0.32	0.35
γ_2	0.37	0.14	0.03	-0.12	0.29	0.53	0.45	0.38	0.60

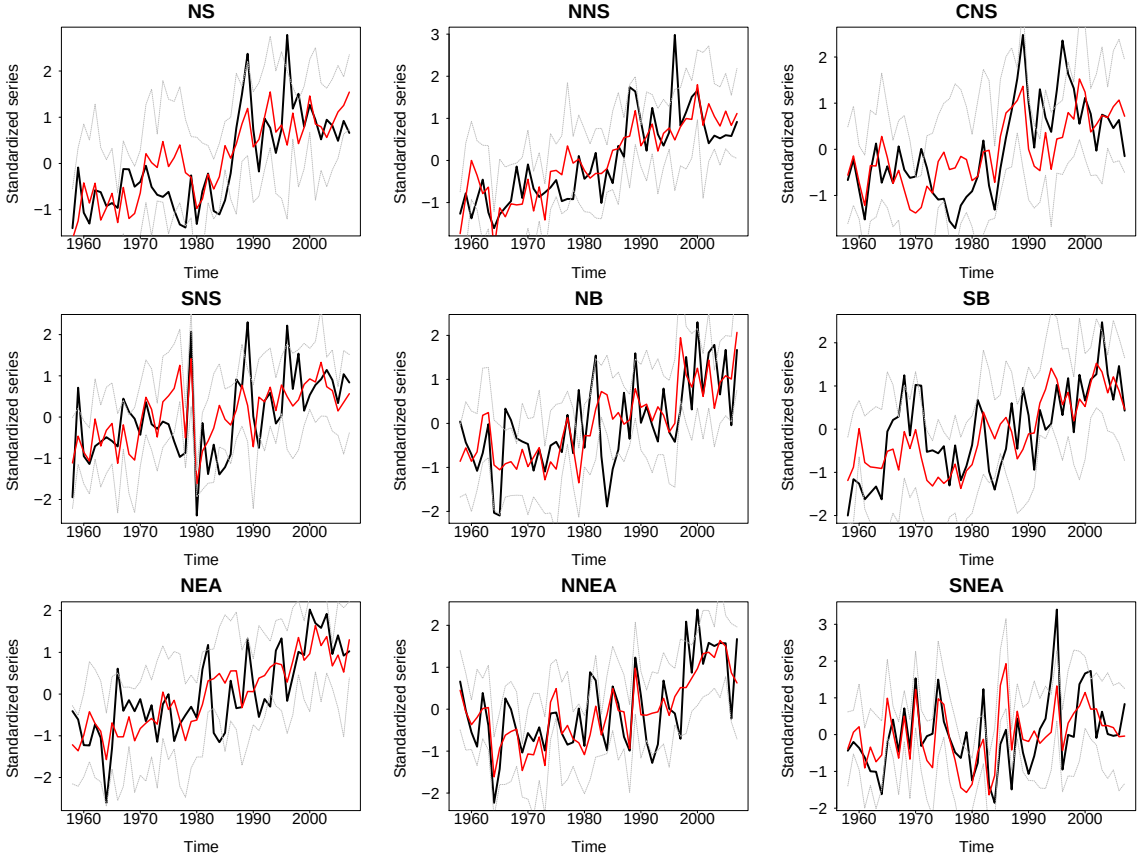

FIGURE 3.6: PCI time-series (black line) and estimated PCI from averaging ten simulations of the index from the multiple linear regression model with ARMA error (red line). Grey lines represent a one standard deviation interval around each simulated point.

3.3 Discussion

The first part of this chapter dealt with the matter of identifying and comparing the timing of change-points and long-term trends in the PCI and abundance of diatoms and dinoflagellates. The methodology applied allowed for the identification of change-points both in the mean and in the linear trend, which has not yet been explored in the literature.

In the North Sea, the timing of the change-points identified in the PCI (table 3.2) is in agreement with the timing of the environmental regime shift mentioned in the introduction, reported to have happened in the late 1980s across multiple throphic levels and associated with large scale hydro-meteorological forcing (e.g., AlvarezFernandez et al., 2012, Beaugrand, 2004, Reid et al., 2001). Furthermore, this change-point in the PCI was identified in all subregions of the North Sea as well. However, I found no evidence that diatom abundance underwent any abrupt changes around the same time (table 3.2, figures 3.3 - 3.4) in either the whole North Sea or any of the subregions considered. Dinoflagellate abundance did present a peak in the late 1980s, coinciding with the known large-scale regime shift in the region. But, after the peak, their abundance decreased again, while the PCI levels remained higher for the following years. One possible hypothesis to explain this discrepancy would be that the driver of the dinoflagellate peak would have also induced smaller-sized phytoplankton species to increase in abundance, and those populations were then responsible for the sustained increased PCI levels after the change-point.

In the Northeast Atlantic, I identified a change-point in the decreasing trend of diatoms in the mid 1970s (table table 3.2), which coincides with one of the regime shifts identified by Goberville et al. (2014) in the ecosystem of this region. However, in most of the subdivisions considered for the Northeast Atlantic, the change-point was identified earlier, in the mid-late 1960s (table 3.2, figures 3.4 - 3.5). The PCI was generally better represented by an increasing trend everywhere, and dinoflagellate abundance by a decreasing trend. Hinder et al. (2012a) showed a gradual change in the relative abundance between diatoms and dinoflagellates in the late 1990s in this area. Although not explicitly identified by the ICPD method, this could be inferred by the results presented here as well, since dinoflagellates exhibited a decreasing trend over time, while diatoms remained fairly constant after the late 1960s / early 1970s.

The different regimes proposed in the work of Leterme et al. (2006) were identified using a cumulative sums method. That is, the cumulative anomalies of each time-series in relation to its average was calculated, and changes in the slope of the resulting series were examined visually. This method is adequate for identifying change-points in the mean only, and might still fail to identify more subtle changes in the mean. The change-points described in this chapter were identified with the ICPD method, which is more robust and flexible than cumulative sums, as it takes into account a collection of possible regimes and trends for the data. Their results coincide with those presented here in regions where an increase or decrease was accentuated, such as in the sharp decrease in diatoms in the southern Northeast Atlantic before 1967 and the sudden increase in the PCI in the North Sea during the mid 1980s. However, Leterme et al. (2006) failed to identify any of the changes in slope or long-term trends which have been presented here.

The second part of this chapter investigated whether it was possible to describe the PCI time-series as a direct function of diatom and dinoflagellate abundances. Two multiple linear regression (MLR) models were considered: one with a normal, white noise error term, and another with a serially correlated error term. The latter proved to be a better model for the PCI.

Diatoms and dinoflagellates were explicitly accounted for as covariates in the multiple linear regression, and yet it was still necessary to include serial correlation in the error term to successfully describe the PCI. In a time-series analysis context, Chaterjee et al. (2000) suggests that the need for this extra structure can be due to (a) some particularity inherent from the nature of the phenomena under consideration, or (b) the absence of some explanatory variable in the regression. The former hypothesis would not be possible in this specific case, as the PCI is derived directly from the greenness that comes from the phytoplankton present in the sampled seawater. It would not be reasonable to expect any other endogenous variability in the PCI, apart from what the index inherits from the phytoplankton that defines it. Consequently, there must be at least one missing variable from the system, that is responsible for the need for an ARMA error structure.

I argue that any missing variables must be of biological origin. It is known that smaller-sized phytoplankton can contribute to the PCI even if their cells are broken (Batten et al., 2003). Even though it could be argued that other inorganic particles may change the water colour locally in coastal waters, the region considered here is large and covers open

oceanic waters as well. Therefore, the results presented here support the hypothesis that the unexplained variability in the PCI, represented in the MLR model by the ARMA error term, is due to smaller-sized phytoplankton groups other than large diatoms and dinoflagellates in the North Sea and Northeast Atlantic. This is in agreement with a similar hypothesis raised by Leterme et al. (2006), that the decrease in contribution from diatoms and dinoflagellates to the PCI would be related to an increase in contribution from other phytoplankton groups of smaller size.

Due to the nature of the CPR structure (section 2.1), small-sized phytoplankton species tend to be underestimated, and many species are not recorded in the CPR survey (Batten et al., 2003). Consequently, it is not possible to directly verify if any specific group or species is more strongly related to the ARMA variability suggested for the unexplained component of the PCI, nor is it possible to compute the exact percentage contribution from the individual plankton groups considered here to the PCI. However, from the present analysis, and comparison among trends and change-points in the PCI, diatoms and dinoflagellates, it is reasonable to infer that the changes in phytoplankton abundance indicated by the increases in the PCI are influenced by changes in the abundance in species of smaller-sized phytoplankton.

There is evidence both from *in situ* (e.g., Hilligsöe et al., 2011, Maranon et al., 2001) and experimental (e.g., Morán et al., 2010) data on how phytoplankton cells of smaller size have advantage over larger cells in warmer, nutrient-poor waters. This has been related to a higher area-to-volume ratio and lower diffusion boundary layer thickness (Peter and Sommer, 2015). In the North Sea and Northeast Atlantic, it is known that sea surface temperature has been rising in the past fifty years (e.g., Beaugrand, 2009, Hinder et al., 2012b, Richardson and Schoeman, 2004, and figure 1.1 in chapter 1). With warming ocean temperatures, and consequently intensified stratification and nutrient limitation, it has been hypothesized that smaller phytoplankton could become dominant in the ocean (Morán et al., 2010). This shift towards smaller phytoplankton would have several consequences to the environment. Smaller cells are less effective in exporting carbon to deep waters since they do not sink as fast as large diatoms (Hilligsöe et al., 2011, Peter and Sommer, 2015). The transfer of energy to higher trophic levels is also likely to become less efficient (Peter and Sommer, 2015) since mesozooplankton, including some copepods, do not seem able to feed efficiently on the smallest phytoplankton. This means that more energy would be lost

to the microbial food web, and organisms such as zooplankton and fish, which occupy higher trophic levels, could reduce in abundance (Sommer et al., 2002).

These results also reinforce the fact that the changes in the PCI may not be reflected in the plankton community structure, as previously suggested in the study of Leterme et al. (2006). It is important that this fact is taken into account when interpreting any analyses that make use of the PCI as an approximation for phytoplankton abundance. Different phytoplankton groups have different niches and roles in the marine environment, and generalizing trends observed in the PCI to the whole phytoplankton community could lead to inaccurate conclusions. A careful choice of suitable phytoplankton variables should be made depending on the nature of the ecological problem under consideration. This conclusion can be extended to studies that make use of satellite chlorophyll data, since these are known to be correlated to, and comparable with, the PCI (e.g., Raitsos et al., 2013a).

It would be redundant to explicitly include exogenous factors in the MLR model presented here, such as changes in climate or in the physical environment. The MLR describes the PCI as a direct function of phytoplankton. Attributing the unexplained variability in the PCI to exogenous variables would be inconsistent with the defining premise and aim of the model. Any effect from exogenous factors would impact the phytoplankton, and consequently be reflected in the PCI, from its definition as a colour index measured from the same silk from which the phytoplankton series are computed. The same can be said of methodological or sampling factors, since those would also impact the recording of individual phytoplankton abundances, and therefore be reflected both in the PCI and the diatom and dinoflagellate terms in the model.

Changes in the intracellular pigmentation of phytoplankton, as a response to changes in light and nutrient availability, have been shown to play an important role in the temporal variability of chlorophyll in the ocean (Behrenfeld et al., 2008, 2016a). It could thus be argued that the increase in the PCI could be a consequence of increased pigmentation in the phytoplankton cells as a response to changes in irradiance. However, it is not clear whether this response would be uniform over different species. Considering the results presented here and previous reports of either a constant or decreasing trend in diatom and dinoflagellate abundance (figures 3.3 - 3.5, and Batten et al., 2003, Harris et al., 2015, Leterme et al., 2006), it would still be reasonable to infer that, even if diatoms and

dinoflagellates were undergoing an increase in pigmentation, that increase alone would not be enough to balance the long-term decreases observed in their abundance. Furthermore, the increase in the PCI was present in all months of the year, meaning that any change in pigmentation, which is usually hypothesized as a consequence of light acclimation (Behrenfeld et al., 2008), would have to happen throughout all of the year and across multiple species and geographical regions. So far there is no evidence that such large-scale adaptation is taking place, and thus the previous conjecture that populations of smaller-sized phytoplankton are increasingly important contributors towards the PCI still holds.

In summary, the results presented here are in line with previous reports regarding the importance of the contribution of groups of smaller-sized phytoplankton to the PCI (Batten et al., 2003, Leterme et al., 2006), and of ecological shifts in the North Sea and Northeast Atlantic (AlvarezFernandez et al., 2012, Beaugrand, 2004, Beaugrand et al., 2014, Reid et al., 2001). However, the differences in trends and change-points identified among the phytoplankton and PCI time-series evidenced an unexpected scenario. The greenness in the ocean water was seen to be increasing, while diatom and dinoflagellate abundances were identified as decreasing with time, in most of the area considered. I demonstrated that it was not possible to accurately describe the PCI as a function of diatom and dinoflagellate abundances alone. It was necessary to include serial correlation explicitly in the model in order to explain all of the variability observed in the PCI. Other biological series of smaller-sized phytoplankton, which are currently unavailable, would be necessary to estimate precisely the individual contribution of different phytoplankton groups to the PCI. Care must be taken when using the PCI as a representation of phytoplankton: while it can be a suitable proxy for total abundance, changes in the PCI are not likely to reflect or be reflected in changes to the phytoplankton community structure.

Chapter 4

Control mechanisms of interannual variability in phytoplankton populations

As discussed in section 1.5 of chapter 1, changes in the plankton distribution and community structure are known to have happened in the recent past in the North Sea and Northeast Atlantic. Most previous studies based on *in situ* observations of plankton populations consider the impacts of climate and the environment on either phytoplankton or zooplankton separately (e.g., Hinder et al., 2012b, Richardson, 2008), but few account for the interactions among the plankton itself. Furthermore, there is no consensus on to what extent the changes in the phytoplankton are controlled by bottom-up or top-down processes, driven by climate and environmental forcing or pressure from predation, respectively. Bottom-up control is more frequently reported in the literature (Chassot et al., 2007, Richardson, 2008), possibly because comprehensive datasets of environmental drivers are readily available, whereas similar biological datasets to account for predatory effects are harder to find.

In this chapter, I investigate whether the phytoplankton of the Northeast Atlantic and North Sea are under bottom-up control driven by physical forcing, or top-down control driven by zooplankton grazing, from an interannual perspective. I apply a vector autoregressive analysis to monthly time-series of phytoplankton and zooplankton to simultaneously model their interactions, as well as the influence of SST and MLD (as indicators of environmental conditions), seasonality and autocorrelation. This type of multivariate statistical models has been successfully applied to describe plankton food webs in lakes (e.g. Gsell et al., 2016, Ives et al., 2003), but applications to oceanic plankton communities are scarce (Hampton

et al., 2013). This method allowed me to evaluate the interactions within the plankton and between the plankton and the environment, even in the absence of grazing data, in order to identify which type of mechanism is more likely to control the phytoplankton dynamics in the region.

4.1 Methods

4.1.1 Region of study

I focused on four subregions of the waters between 63^{o}N - 45^{o}N and 20^{o}W - 10^{o}E, which consist of the North Sea and a large portion of the Northeast Atlantic (figure 4.1). The subregions provide good geographical coverage of the area, and are frequently sampled by CPR transects, with fewer missing data than the individual CPR standard areas considered

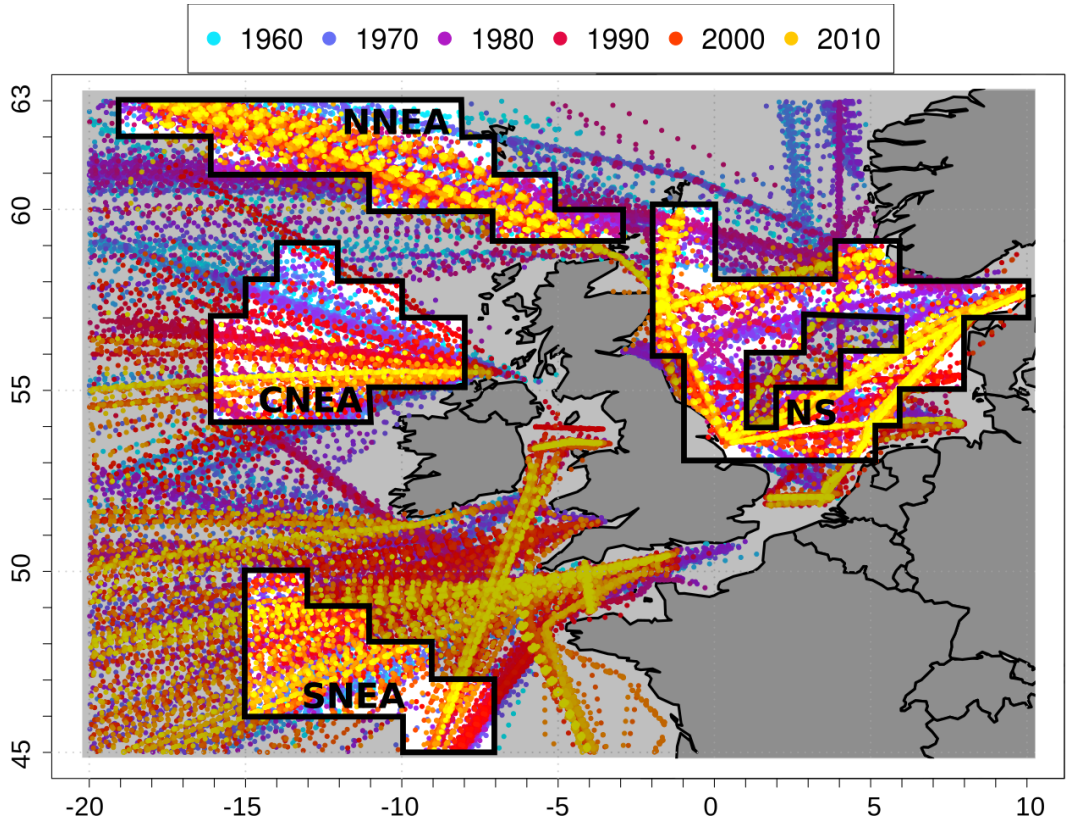

FIGURE 4.1: Subregions considered in this chapter: North Sea (NS), northern Northeast Atlantic (NNEA), central Northeast Atlantic (CNEA) and southern Northeast Atlantic (SNEA). The coloured dots represent sampling points of the CPR survey between 1958 and 2014.

TABLE 4.1: Length of time-series and percentage of missing values in the monthly data in the North Sea (NS) and northern (NNEA), central (CNEA) and southern (SNEA) Northeast Atlantic.

	NS	NNEA	CNEA	SNEA
Period	1958 - 2014	1960 - 2014	1958 - 1993	1958 - 2014
Missing months	2.95%	8.62%	4.97%	5.99%

previously (see table 4.1). The North Sea subregion (NS) consists of 684 monthly observations, between 1958 and 2014. The Northern Northeast Atlantic subregion (NNEA) consists of 660 monthly observations, between 1960 and 2014 in the waters between the north of the British Isles and the south of Iceland. The Central Northeast Atlantic subregion (CNEA) consists of 432 monthly observations, between 1958 and 1993 in the waters west of Ireland. The Southern Northeast Atlantic subregion (SNEA) consists of 684 monthly observations, between 1958 and 2014 in the waters south of the Celtic Sea, off the Bay of Biscay. The shorter length of the series in the NNEA and CNEA regions are due to long periods with multiple missing observations, and had to be removed.

4.1.2 Data

To represent the plankton community, I used monthly abundance time-series of diatom, dinoflagellate, small copepods (all life stages smaller than 2mm, Batten and Welch, 2004) and large copepods from the Continuous Plankton Recorder (CPR) survey. Refer to section 2.1.1 in chapter 2 for more details on the plankton dataset. Diatom, dinoflagellate and copepod abundance provide a wide picture of the plankton community, since those are the main phytoplankton and zooplankton groups captured by the CPR survey in the area. I also consider the phytoplankton colour index (PCI) as an alternative proxy for total phytoplankton abundance, since it has been shown in the previous chapter to contain information about groups other than diatoms and dinoflagellates. When three or less consecutive observations were missing, I linearly interpolated the values in order to obtain a complete series to work with.

I used sea surface temperature (SST) and mixed layer depth (MLD) as covariates to represent the physical environment, as these variables are most frequently associated with changes in the plankton (chapter 1 and AlvarezFernandez et al., 2012, Barton et al., 2015, Hinder et al., 2012b, Martinez et al., 2016). In a preliminary analysis, the North Atlantic Oscillation (NAO) index was also considered as an environmental covariate, but it was

removed since it had little power as a predictor variable. Although other covariates, such as nutrient availability and irradiance, are of importance to phytoplankton control, to my knowledge there is no dataset on these variables available spanning the same temporal and geographical extent of the CPR survey. SST and MLD are often highlighted as the main drivers of interannual changes in plankton populations (Beaugrand, 2009, Harris et al., 2014, Leterme and Pingree, 2007, Martinez et al., 2016), and both can be linked to variations in light and nutrient availability (Irwin and Finkel, 2008). Therefore they should provide a good, even if partially indirect, representation of environmental conditions that can impact the plankton. Refer to section 2.1.2 in chapter 2 for more details on the environmental data.

4.1.3 Vector autoregressive analysis

I used vector autoregressive (VAR) models to estimate the interactions between phytoplankton and zooplankton time-series, as well as interactions between the plankton and environmental covariates, seasonality and autocorrelation. The main advantage of VAR models is the simultaneous estimation of all interactions, allowing for inferences on the importance of bottom-up versus top-down processes. Due to their linear structure, VAR models are also closed under spatial averaging, that is, if neighbour regions can be approximated by a VAR structure, their combined average can also be approximated by a VAR structure (Lütkepohl, 2007). This ensures that any common results in the subregions of the Northeast Atlantic would still hold if the entire area had been considered as a whole. These models are sometimes referred to as Multivariate Autoregressive (MAR) models in the literature. MAR and VAR are synonyms, and I use VAR in this thesis as it is more common in the statistical literature. For a detailed introduction to VAR models and how they can be applied to model the plankton, refer to section 2.6 in chapter 2.

VAR models have been successfully applied to describe phytoplankton blooms (Lui et al., 2007) and food webs of freshwater plankton communities (e.g., Gsell et al., 2016, Hampton et al., 2008, Ives et al., 1999, Kratina et al., 2014), but have rarely been used to describe marine plankton systems (Francis et al., 2012, Scheef et al., 2012). To my knowledge, VAR models have not yet been applied to CPR data except at the vicinity of the L4 Station in the Western English Channel (Scheef et al., 2012).

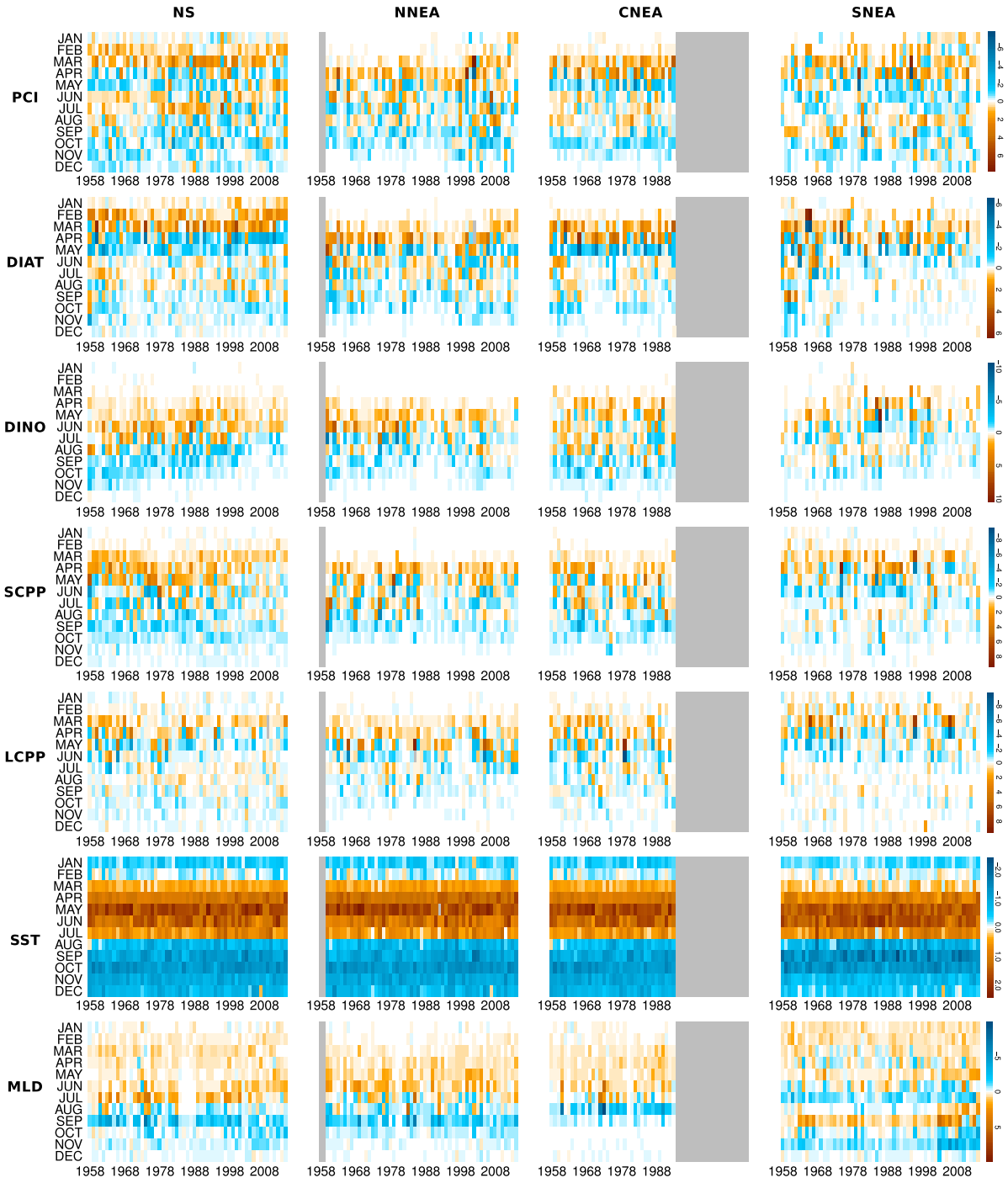

FIGURE 4.2: First-differenced standardised time-series of the phytoplankton colour index (PCI, first row), diatom abundance (second row), dinoflagellate abundance (third row), small copepod abundance (fourth row), large copepod abundance (fifth row), sea surface temperature (sixt row) and mixed layer depth (seventh row); in the North Sea (NS, first column), northern Northeast Atlantic (NNEA, second column), central Northeast Atlantic (CNEA, third column) and southern Northeast Atlantic (SNEA, fourth column).

TABLE 4.2: Maximum lags (L), selected using the BIC (Schwarz, 1978), to fit the VARs models considered in the North Sea (NS) and northern (NNEA), central (CNEA) and southern (SNEA) Northeast Atlantic.

	NS	NNEA	CNEA	SNEA
VAR diatoms & dinoflagellates	3	3	2	4
VAR with the PCI	3	4	3	5

VAR models require the stationarity of all variables in the analysis, that is, that some statistical properties of time-series, such as mean and variance, do not change with time. In order to verify stationarity, I applied the ADF (Dickey and Fuller, 1979b) and the KPSS (Kwistkowski et al., 1992) tests to each plankton and environmental time-series (see section 2.4 for more details). The tests were performed using the 'tseries' package, version 0.10-34 (Trapletti and Hornik, 2017), in the R software (www.r-project.org, version 3.4.1). The results of these tests were not conclusive for the original series, but after taking first differences all series were classified as stationary according to both tests (p-value of the ADF-GLS ≤ 0.01 and p-value of the KPSS ≥ 0.1 for all series and regions). The first difference transformation is linear, and thus preserves the qualitative interpretation of the VAR results (detailed proof in appendix A, page 131).

The VAR model was thus estimated for a system composed of first-differenced time-series of four plankton variables (abundance of diatoms, dinoflagellates, small copepods and large copepods) with two external regressors (sea surface temperature, SST, and mixed layer depth, MLD). Figure 4.2 shows the first-differenced time-series of these variables.

The choice of using first differences of the original time-series rather than a log scale as a strategy to stabilize the data was made for a few reasons. The first-difference operation is linear and thus the VAR results are easier to interpret than if they were performed in log scaled data. In addition, there is no indication that the original data would follow any sort of exponential law, and therefore there is no warranty that a log transformation would bring stationarity to the data, which was achieved by first differences.

In matrix notation, the VAR system can be written as:

$$\Delta PKT(t) = \sum_{j=1}^{L} \mathbf{A_j}\, \Delta PKT(t-j) + \sum_{k=0}^{L} (u_k\, \Delta SST(t-k) + v_k\, \Delta MLD(t-k)) + ssn(t) + \varepsilon(t)$$

- $\Delta PKT(t)$ is a (4×1) vector containing the first differences of diatoms, dinoflagellates, small copepods and large copepods, and $\Delta SST(t)$, $\Delta MLD(t)$ are the first-differenced

SST and MLD time-series.

- L the maximum lag at which the past still influences the present system, that is, the extent of the memory in the system.
- $\mathbf{A_j}$, for $j = 1, \cdots, L$, are (4×4) coefficient matrices for the effects of past plankton variability to current variability, and can be interpreted to represent linear biological interactions among the plankton (Ives et al., 2003).
- u_k, v_k, for $k = 0, \cdots, l$, are (4×1) coefficient vectors representing the effects of the environmental series and their past to the plankton variability.
- $ssn(t)$ is a (4×1) periodic vector, representing the average seasonal cycle throughout a year, and is a constant for each month of the year. This term ensures that relationships inferred from the coefficient matrices A_j, u_k and v_k in the model are distinct from the influence originated from the annual seasonal cycle.
- $\varepsilon(t)$ is a (4×1) white noise process, with zero mean and constant variance, that accounts for random fluctuations in the system.

All parameters were estimated using the 'vars' package in R, version 1.5-2 (Pfaff, 2008). The maximum lag, L, was chosen for each region using the Bayesian information criterion (chapter 2, equation 2.2). The constant parameters corresponding to the plankton interactions (matrices $\mathbf{A_j}$), influence of external variables (u_k and v_k) and the seasonal and white noise terms ($ssn(t)$ and $\varepsilon(t)$) were estimated using ordinary least squares. Finally, the significant parameters estimated were identified using a critical level of 5%, and any non-significant parameters were set to zero.

For comparison, I estimated an alternative model using the PCI first-differenced series instead of diatoms and dinoflagellates to represent the phytoplankton. I also attempted to repeat the analysis using annually averaged data, however this presented two challenges. Firstly, the annually averaged series did not pass the ADF-GLS and KPSS tests, even after taking first differences. Since both the annual averages and their first differences failed to meet the theoretical assumptions of the VAR model, it would be necessary to further differentiate the data, or to find another transformation to stabilize the data, in order to avoid spurious results. Secondly, the length of the annual averaged time-series (maximum of 57 points, corresponding to each year) is likely to be too short to yield robust

results in this type of analysis (Lütkepohl, 2007). Therefore, the following sections discuss the results obtained using monthly data only.

VAR models only capture linear relationships among the variables considered. If nonlinear relationships are present they may not emerge clearly in the analysis and could potentially contribute towards the autocorrelation component, since, for nonlinear data, a good approximation for the nonlinearity is usually the past of the time-series itself.

4.2 Results

4.2.1 VAR with diatoms and dinoflagellates

A summary of the relative contribution of the VAR parameters estimated for the system with diatoms, dinoflagellates, small and large copepods can be found in figure 4.4, and a detailed breakdown of the estimated parameters can be found in appendix B, page 133. Figure 4.3 gives an example of a VAR summary output for a system with two synthetic series and one exogenous variable.

Table 4.2 indicates the lags selected for each model in each region.

The results showed that the variability of diatoms was not affected by the other plankton groups in the North Sea, northern and central Northeast Atlantic. This can be seen by the lack of significant relationships in the VAR model, indicated in the blank barplots of figure 4.4. The exception was the southern Northeast Atlantic subregion (4.4, panel D), where large copepods had some small negative influence on diatom variability. Temporal autocorrelation and effects from the seasonal cycle were significant for diatom variability, particularly in late winter and spring (4.4, panels A - D). SST was a statistically significant covariate for diatom variability in all subregions, with different lags of SST exerting a small inverse influence on diatoms. Meanwhile, MLD emerged as a significant positive covariate only in the northern Northeast Atlantic (4.4, panel B).

Dinoflagellate variability was also related to the seasonality (mostly in late spring and summer), autocorrelation and SST in all subregions. SST influence as a covariate was both positive and inverse, depending on the different lags and regions; while MLD was a significant covariate only in the northern Northeast Atlantic (4.4, panel B). There were

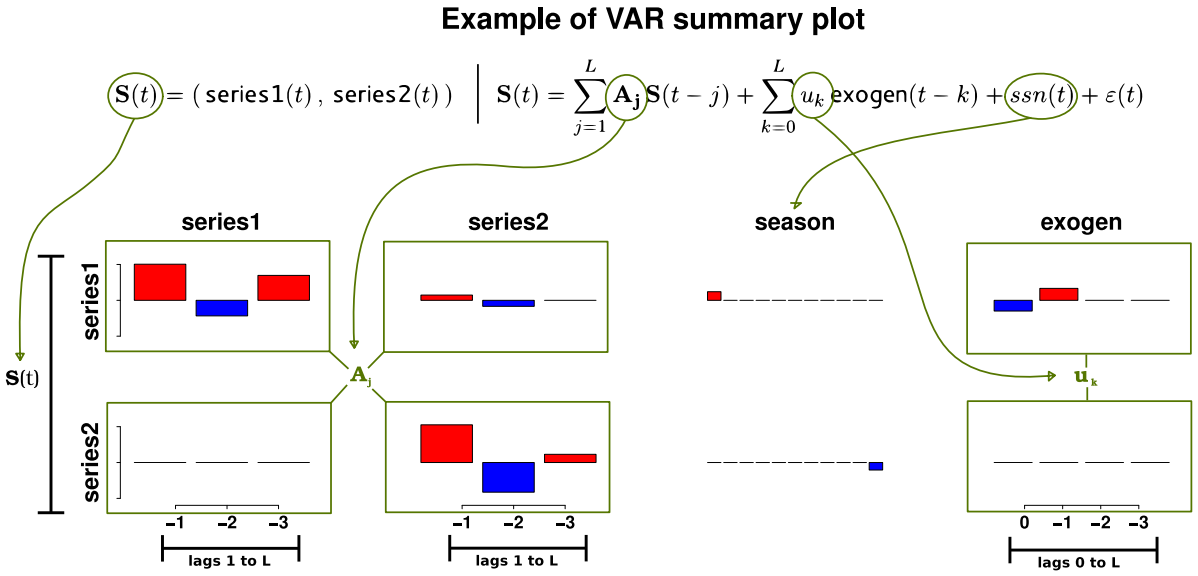

FIGURE 4.3: Example of a VAR model fitted to a synthetic system composed of an exogenous periodic variable ("exogen"), an autoregressive process ("series2") and a time-series generated from a quadratic combination of both processes ("series1"). From the construction of this example, the following expected features can be observed. "Series1" is related to "series2" and "exogen" but also contains some level of serial correlation due to the nonlinearity on its definition; while "series2" was defined independently of the other processes and presents virtually only serial correlation. Neither series present a strong seasonal component; as any seasonality in "series1" is inherited directly from "exogen", and "series2" is not periodic. The empty bar plots correspond to a non statistically significant relationship between the variables in the corresponding line and column.

significant interactions with different combinations of the other plankton variables: diatoms in the northern Northeast Atlantic (panel B); small copepods in the North Sea, central and southern Northeast Atlantic (panels A, C and D); and large copepods in the North Sea, northern and southern Northeast Atlantic (panels A, B and D).

Small copepod variability in the North Sea (4.4, panel A) was influenced by seasonality, autocorrelation, diatom variability (positive relationship) and large copepods (inverse relationship) as well. In the Northeast Atlantic regions, small copepods were under the influence of seasonality (with peak in different timings depending on the region), autocorrelation and SST, with the addition of a positive influence from dinoflagellates in the southern Northeast Atlantic (panel D) and a positive influence from large copepods in the central Northeast Atlantic (panel C). In the northern Northeast Atlantic (panel A), the model suggested a positive relationship with diatoms and dinoflagellates, but an inverse relationship with large copepods.

Large copepod variability was influenced by seasonality (with the model highlighting late winter and spring as having a high importance), autocorrelation and diatom variability

(positive relationship) in all subregions. SST was also a significant covariate, with the model indicating a positive relationship in the North Sea (4.4, panel A), but an inverse relationship in the Northeast Atlantic (panels B, C and D). In the northern Northeast Atlantic, dinoflagellates and small copepods also emerged as significant to large copepod variability (panel B).

4.2.2 VAR with the PCI

It has been seen in chapter 3 that there is a significant component of the plankton in this region which is not represented by diatoms and dinoflagellates alone, therefore I repeated the analysis in an alternative model, using the PCI instead of diatom and dinoflagellate abundance to represent phytoplankton abundance. The summary of the estimated parameters can be found in figure 4.5. It can be seen that, once again, all three plankton variables (PCI, small copepods and large copepods) were under the influence of autocorrelation and seasonality.

SST was a statistically significant covariate for the PCI variability in the northern and central Northeast Atlantic only (4.5, panels B and C). Large and small copepods presented a weak inverse relationship with the PCI in the southern Northeast Atlantic (panel D), similar to what was observed in diatoms and dinoflagellates (4.4, panel D). Large copepods were also inversely related to the PCI variability in the North Sea (4.5, panel A).

For small copepod variability, SST was a statistically significant covariate in all four regions considered, exerting a direct or inverse influence depending on the different lags identified as significant by the model. MLD was also a significant covariate in the northern and central Northeast Atlantic (4.5, panels B and C). Small copepods were weakly positively influenced by the PCI in the northern Northeast Atlantic only (panel B), and by large copepods in the North Sea and central Northeast Atlantic (panels A and C). Large copepods were positively influenced by the PCI in the central and southern Northeast Atlantic (panels C and D), and had SST as a significant covariate in all three subregions of the Northeast Atlantic (panels B, C and D).

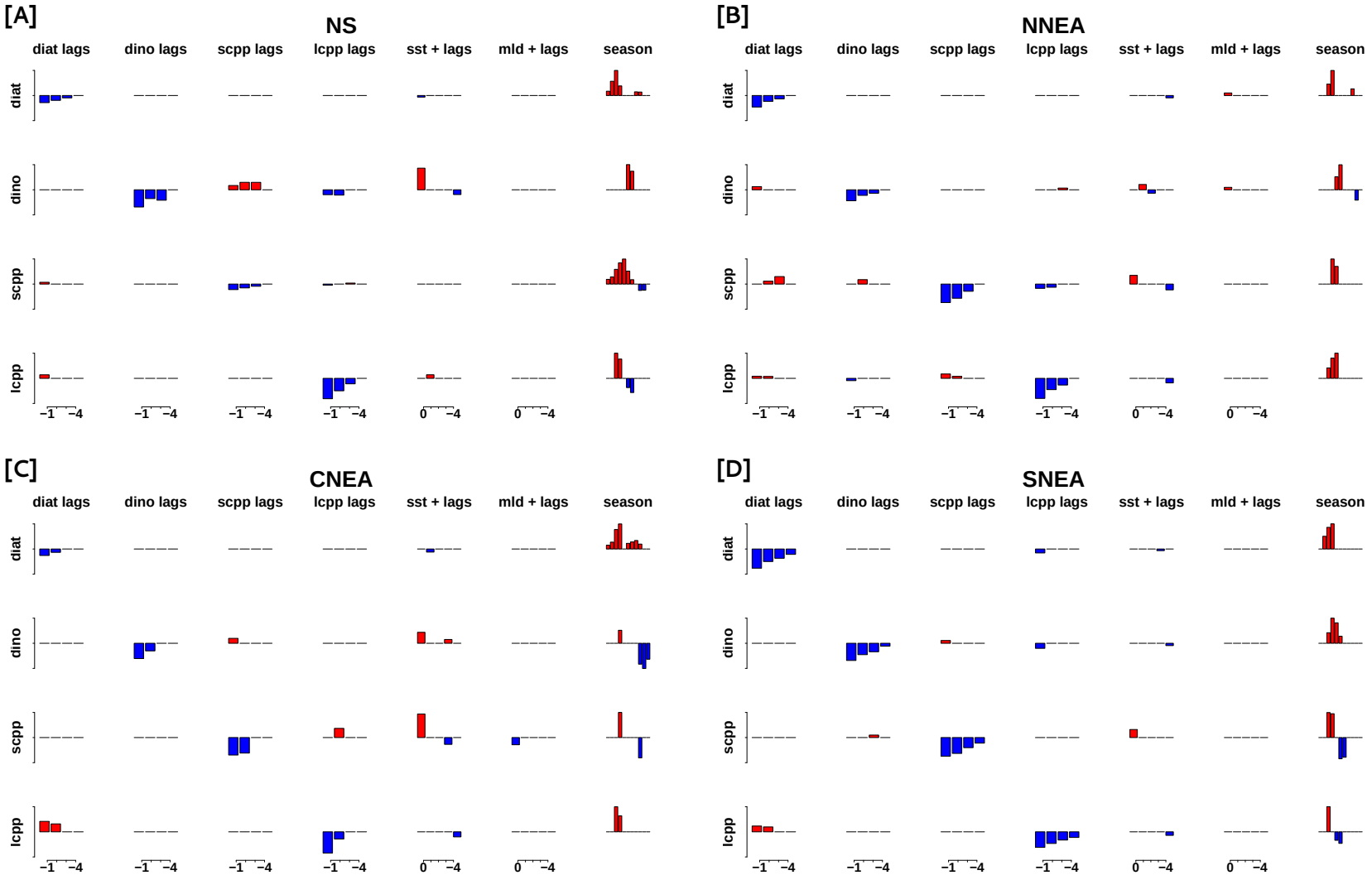

FIGURE 4.4: Graphical representation of parameters estimated by the vector autoregressive (VAR) models in monthly resolved time-series considering diatoms (DIAT) and dinoflagellates (DINO) as indicators of phytoplankton abundance and small (SCPP) and large (LCPP) copepods as indicator of zooplankton abundance. VAR models are fitted in the North Sea (NS), and northern, central and southern NE Atlantic (NNEA, CNEA and SNEA). Blank plots indicate non-significant relationships. Each row is scaled relative to the maximum and minimum parameters estimated. The detailed values of the parameters estimated can be found in tables B.1 to B.4 in appendix B.

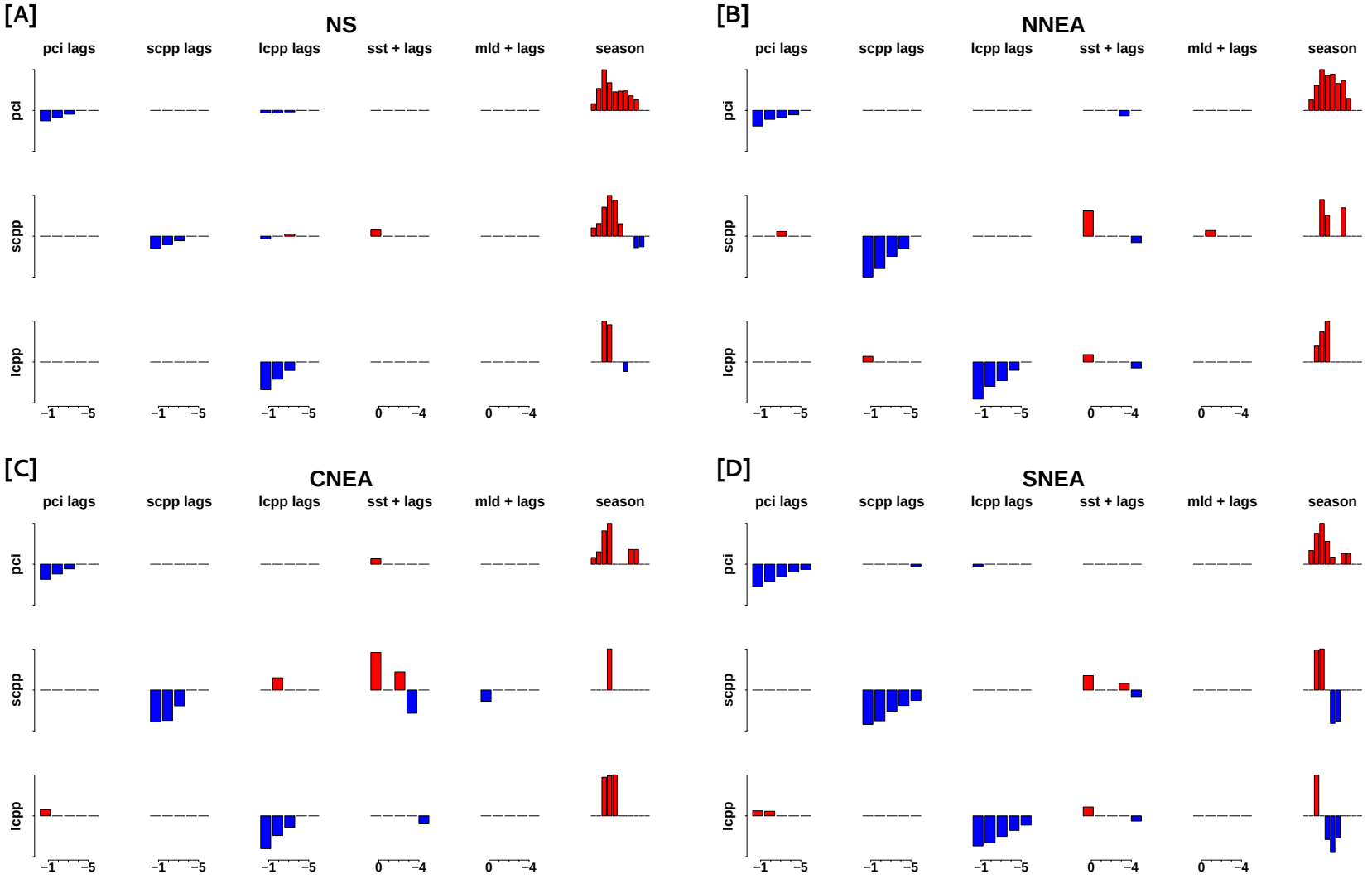

FIGURE 4.5: Graphical representation of parameters estimated by the vector autoregressive (VAR) models in monthly resolved time-series of the phytoplankton colour index (PCI) as an indicator of phytoplankton abundance and small (SCPP) and large (LCPP) copepods as indicator of zooplankton abundance in the North Sea (NS), and northern, central and southern NE Atlantic (NNEA, CNEA and SNEA). Blank plots indicate non-significant relationships. Each row is scaled relative to the maximum and minimum parameters estimated. The detailed values of the parameters estimated can be found in tables B.5 to B.8 in appendix B.

4.3 Discussion

Interpreting the coefficients of a VAR model is not straightforward, since all parameters are estimated simultaneously, and different lags of the same variable can have either a positive or negative effect on the other variables. Francis et al. (2012) suggest that a qualitative view of the VAR results may be more appropriate than a strict interpretation of the estimated coefficients, and Gsell et al. (2016) remark that interactions which are not consistent through time, or are important only during a short period, tend to be eliminated in the VAR estimation. Hence, I focus the present discussion mostly on the consequences of the presence or absence of relationships among the plankton.

Diatom variability (figure 4.4) appeared to be linearly independent of the other plankton variables considered in all regions except for SNEA. In addition, SST emerged as a significant covariate for diatoms in all regions. This is consistent with a bottom-up scenario, where interannual diatom variability would be more strongly regulated by environmental factors than by zooplankton grazing. Diatoms were negatively related to SST in all regions, and the MLD had a significant positive effect on diatom variability only in the northern Northeast Atlantic. Even when repeating the analysis without the SST, the MLD was still not a significant covariate for diatom variability. This contrasts with previous reports of a large influence from the MLD (e.g., Martinez et al., 2016) or from atmospheric forcing represented by the NAO index (e.g., Zhai et al., 2013) as drivers of diatom variability, and could suggest that SST has a more direct metabolic influence to diatom variability in the study region.

Dinoflagellates (figure 4.4) also had SST as an important covariate in all four regions considered, and, like diatoms, were positively related to MLD in the northern Northeast Atlantic only. However, dinoflagellates exhibited different interactions with the other plankton variables in each region, showing no common pattern across the regions. These results may be due to the fact that the total dinoflagellate abundance from the CPR survey combines autotrophic, mixotrophic and heterotrophic species. A more refined, species-level analysis would be necessary to fully understand the dinoflagellate dynamics, which falls outside the scope of this study.

Small and large copepods (figure 4.4) had SST as a significant covariate, in agreement with previous results reporting that temperature can have effects on the zooplankton

seasonal cycle (e.g., Richardson, 2008) and spatial distribution, with cold water species moving northwards with increasing SST (Beaugrand et al., 2002, Harris et al., 2015). There was evidence supporting a bottom-up control of copepod variability, since copepod abundance, either small or large, was related to phytoplankton in all subregions considered. A positive association between large copepods and diatoms was identified, and also positive links between small copepod variability and either diatoms or dinoflagellates, depending on the region.

When considering the PCI as a proxy for phytoplankton (figure 4.5), I found that the index was more strongly influenced by autocorrelation and seasonality than by zooplankton variability, and that the environmental proxies considered were rarely significant covariates in the VAR model. The PCI seemed more dissociated from copepod and environmental variables than diatoms and dinoflagellates. This result reinforces the discussion from chapter 3, that the PCI may not capture the full complexity of the dynamics within the phytoplankton community (Batten et al., 2003, Leterme et al., 2006).

Temperature has the potential to affect the metabolism, physiology, species composition, abundance and trophic efficiency of both phytoplankton and zooplankton (section 1.4, chapter 1). In the Northeast Atlantic and North Sea, the consequences of warming surface waters are not clear. Martinez et al. (2016) argue that, despite higher SST, primary productivity has the potential to increase as simultaneously increased winds and mixing make nutrients more available for phytoplankton. On the other hand, Barton et al. (2015) associate stratification and warmer SST with increased phytoplankton biomass, while deep mixing is associated with a decrease in phytoplankton biomass in the North Atlantic subpolar gyre. The effects of warming waters may also not be the same for every plankton group. A decline in dinoflagellate populations in relation to diatoms was associated with increasing SST and increased summer windiness in the Northeast Atlantic and North Sea (Hinder et al., 2012b), with suggestion that the increased turbulence has been detrimental to dinoflagellates in the region. As mentioned previously, warmer waters could also favour smaller-sized phytoplankton species (e.g., Hilligsöe et al., 2011).

Here I found both positive and negative links between the plankton variables and different lags of SST in each subregion considered (figures 4.4 and 4.5). The sign of the estimated VAR parameters differed depending on the time lags and location. For example, dinoflagellate and small copepod variabilities were directly and inversely related to

different lags of SST in the northern Northeast Atlantic. Although some interactions seem easier to interpret, such as the inverse relationship between diatom variability and SST lags in all subregions, the complexity of the interactions and absence of a general pattern are consistent with previous results that report different impacts of climate change on plankton in different geographical regions of the North Atlantic. For example, McGinty et al. (2011) reported how copepods in the Northeast Atlantic respond differently to environmental conditions depending on how local hydrography and advection alters climate forcing, Beaugrand (2009) associated changes in the seasonal cycle and geographical distribution of plankton in the North Atlantic with warmer SST, and Martinez et al. (2016) suggested that the deepening of the mixed layer can have a positive impact on phytoplankton abundance.

Factors other than SST, MLD and copepod grazing are known to influence the bottom-up and top-down control of phytoplankton populations. For example, light and nutrient availability are essential for the growth and bloom of phytoplankton. However they cannot be explicitly accounted for in the present analysis since, to my knowledge, there is no dataset spanning the same spatio-temporal resolution of the CPR series for those variables. Nevertheless, the SST and MLD should be good proxies for these (Beaugrand et al., 2002, Hinder et al., 2012b, Irwin and Finkel, 2008, Martinez et al., 2016, Richardson, 2008). SST and irradiance follow similar annual seasonal cycles, and nutrient availability is linked to the mixing of the water column. For a more precise analysis of bottom-up versus top-down control, it would also have been useful to include data on the temporal variability of the reproduction and mortality rates of phytoplankton, and other loss terms such as grazing or viral lysis. However, once again, such data are not available at an appropriate temporal and spatial scale. Taking all into account, the present chapter supports the hypothesis that temperature is one of the main physical drivers of interannual phytoplankton variability in the ocean (Behrenfeld et al., 2006).

The results also showed evidence of strong autocorrelation in the plankton, evidenced by the almost diagonal structure in the plankton interactions matrix, usually dominating the signal relative to the interactions between the plankton and the environmental covariates. This serial correlation was often negative, which could indicate an internal regulation in the plankton, that is, periods of increase followed by decrease in abundance, and vice-versa. The presence of this correlation in marine ecosystems can arise from the integration of stochastic atmospheric forcing (Lorenzo and Ohman, 2013). However, this could also be due to the presence of nonlinear relationships not accounted for in the analysis, or from

some missing covariate in the model (Barton et al., 2015). Given the complexity of the plankton ecosystem and of its interactions with physical variables, both hypotheses are plausible, and both would yield a similar diagonal autocorrelated structure to the one observed here. Since the VAR does not account for nonlinear relationships, I cannot pinpoint which are the specific causes of the autocorrelation observed in the plankton system. Barton et al. (2015) argue that, although direct (that is, linear) relationships between plankton abundance and physical variables are not expected due to the pronounced interannual variability of the plankton, the lack of any widespread correlations indicates that the drivers of ecological change in the plankton of the North Atlantic may be more uncertain than usually assumed. Further investigation is needed to understand the interactions and general nature of the dynamics between the plankton and the environment. These issues are addressed in the following chapter.

A few studies have used a similar approach to describe plankton communities in other regions of the ocean. For instance, Francis et al. (2012) applied a VAR model to describe the zooplankton community structure in the Northern California Current, and suggested phytoplankton abundance or chlorophyll-a as covariates for zooplankton abundance. They also found evidence that oceanic temperature and the ENSO index were more significant as drivers of zooplankton variability in their model, supporting a bottom-up control mechanism in the Northern California Current. I found similar results in the North Sea and Northeast Atlantic subregions: weak links between copepods and phytoplankton, and SST emerging as the main environmental driver of plankton variability, while the large scale atmospheric index initially considered (the NAO) did not present any statistically significant linear links with the plankton. Scheef et al. (2012), is, so far, the only study in the Northeast Atlantic that uses the VAR methodology. The authors compared VAR models fitted to data from the L4 station in the English Channel and CPR data in the same area, reporting considerably different outcomes between the different datasets, likely to be a consequence of the differences in sampling method (fixed point versus spatially distributed). Although the variables considered were different from the present study, autocorrelation was also present in many of the different models considered by these authors.

Most of the previous analyses using VAR models on plankton populations restrict their application to a single temporal unit lag (e.g., Francis et al., 2012, Gsell et al., 2016, Hampton et al., 2013, Ives et al., 2003, Scheef et al., 2012). Here I chose the number of lags according to an information criterion, in order to select the best model to describe the

data. A single lag in the VAR coefficient matrix has the advantage of being easier to interpret, but a mis-specified number of lags could yield misleading results. Furthermore, although most of those studies performed transformations to stabilize the data (e.g. log(x+1)), some formal check for stationarity should be used on the transformed series to confirm that they meet the requirement of the VAR model. The reliability of the VAR parameters depends on the assumption of stationarity, which was accounted for and met through first-differencing in this chapter.

In summary, autocorrelation and the seasonal cycle were found to be the strongest drivers of plankton variability in the North Sea and Northeast Atlantic, when considering linear relationships among them. Between the environmental drivers considered, SST was a dominating covariate over MLD. I found supporting evidence for the hypothesis that bottom-up mechanisms would be important for the interannual variability of phytoplankton in this region, while top-down control from grazers appeared as less significant in an interannual time scale. The long and detailed plankton datasets required for performing VAR analysis are rare, and continued, consistent monitoring of marine plankton, such as the CPR survey, is essential to improve our knowledge about climate-driven variability in the plankton, and to predict future impacts on marine ecosystems.

Chapter 5

Nonlinearity, stochasticity and chaos in the plankton ecosystem

The interannual variability in monthly observations of marine plankton abundance can be a consequence of a strong stochastic signal, or of a nonlinear response to environmental forcing. The latter alternative encompasses the possibility of chaotic dynamics, introduced in section 2.7. The presence of deterministic chaos in ecological systems is a topic of great interest from both an applied and theoretical perspective, since chaos can impose as much difficulty to analyse a dataset as a strong stochastic signature (Sugihara et al., 1990). Accurately characterising the dynamics of an ecosystem is important, for instance, for a better development of numerical models, statistical analyses or performing forecasts based on previous observations of the ecosystem state.

Deterministic chaos has been previously identified in plankton models. For example, Moroz et al. (2016) showed that simple models of plankton food chains can present chaotic dynamics depending on the type of zooplankton grazing strategy. The authors also specified seasonal forcing as an inducer of chaos, which was amplified through the plankton interactions in their model. Huisman et al. (2006) proposed a model for deep chlorophyll maxima in oligotrophic waters that reproduced main features observed *in situ*, such as the depth of the chlorophyll maxima, the seasonality induced in chlorophyll concentration by the light cycle and the vertical zonation of plankton species. In their model, chaotic dynamics could be induced in plankton biomass and species composition when the vertical mixing in the water column was reduced, indicating that variability in the physical environment could be a source of destabilisation in the plankton dynamics.

Chaos has also been observed in experiments with plankton ecosystems. Benincà et al. (2008) isolated a plankton community from the Baltic Sea and cultivated it in laboratory, under constant external conditions of temperature, salinity, aeration and irradiance, with a constant daily light/dark cycle. The authors collected data from ten plankton functional groups, two nutrients and one detritus pool, sampled for over six years, to obtain time-series of 690 equidistant observations. The predictability of the system decayed exponentially with time, a characteristic of deterministic chaos. The calculation of Lyapunov exponents, a rate of divergence between nearby trajectories, also indicated that species interactions in the plankton food web would induce chaos in their system.

However, the discussion of whether chaotic dynamics is present in *in situ* oceanic plankton systems is still open. Ascioti et al. (1993) investigated the presence of deterministic chaos in phytoplankton and zooplankton biomass of the Mid-Atlantic Bight, but could not confirm the presence of chaos, or rule out the presence of a stochastic signal inherited from physical disturbances. Sugihara and May (1990) suggested that diatom abundance at the Scripps Pier, although partially consistent with a chaotic attractor, was still predominantly under the influence of additive noise. It is therefore still not clear whether oceanic plankton are governed by deterministic chaos or not, and some degree of stochasticity is thought to be responsible for the variability in the plankton.

In this chapter I discuss the presence of chaos in plankton abundance time-series of the Northeast Atlantic and North Sea in the past decades, a question never before addressed in this region or time scale. I use the empirical dynamics modelling analysis, outlined in McGowan et al. (2017), Ye et al. (2015) and originally proposed by Sugihara and May (1990), to search for signatures of deterministic chaos in the dataset. This powerful, non-parametric method is suitable for most types of time-series data, and bases its diagnosis on identifying key known features of chaotic systems (section 2.7.1). In spite of being in general highly suitable for ecological datasets, this approach has seldom been applied to study plankton dynamics, and has never before been explored in the Northeast Atlantic or North Sea, nor applied in the way proposed here to investigate long-term relationships among the plankton.

These methods also carry the ability to check for nonlinearity, a hypothesis raised in the previous chapter from the autocorrelated component estimated in the VAR model. I use S-maps analysis (section 2.7.2 and Sugihara, 1994) to investigate the presence of

nonlinearity in the plankton time-series. I also discuss the role of the annual seasonal cycle, which was identified in chapter 4 as one of the main regulators of plankton variability, as a potential stabilizer of long-term plankton fluctuations. Another important factor to revisit, often reported in the literature and identified by the VAR models in chapter 4, is that of sea surface temperature being identified as a main driver of plankton variability. In the previous chapter, this result was based on a linear, stochastic premise. Unless the plankton system is proved to be linear, it is important to verify which environmental covariates remain related to the plankton in a nonlinear scenario. Therefore, I use a convergent cross-mapping analysis (section 2.7.3 and McGowan et al., 2017, Sugihara et al., 2012) to investigate the relationships between the plankton and environmental time-series from a nonlinear perspective.

5.1 Methods

5.1.1 Region of study and data

As in chapter 4, I focus on the four sub-regions of the North Sea and Northeast Atlantic waters defined in figure 4.1 and section 4.1.1. I consider time-series of the phytoplankton colour index (PCI) and abundance of diatoms, dinoflagellates, small and large copepods to represent the plankton system. I also consider sea surface temperature (SST) and mixed layer depth (MLD) as proxies to represent environmental conditions. Refer to sections 2.1 and 4.1.1 for more details on the dataset and region of study.

5.1.2 Methods

Following Ascioti et al. (1993), Reick and Page (2000), Sugihara and May (1990), Sugihara et al. (1990) and McGowan et al. (2017), I use the non-parametric k-Nearest-Neighbours (kNN) forecasting method (sometimes called "simplex projection") to search for a signature of chaotic dynamics in the plankton and environmental time-series in the regions considered (chapter 2, section 2.7.1). This is a powerful methodology, readily applicable to most types of data, that has been successfully used to model many complex systems (e.g., incidence of measles in different geographical regions, Sugihara et al., 1990; influenza outbreaks, Deyle et al., 2016a; variability of infant heart rhythms, Sugihara et al., 1996). In spite of its success,

this approach has seldom been applied to plankton systems, and this is the first time that this methodology is used in the North Sea and Northeast Atlantic, and the first time they are applied to such long time-series as those from the CPR dataset as well.

The fact that the method is non-parametric removes the risk of uncertainty from mis-specifying a model to the data. kNN forecasting can be applied to any time-series, although it will perform better in cases when the data presents some internal regularity or periodicity. The method is not as efficient for non-correlated stochastic data, since it relies on identifying similar patterns in the past of the series, which would not be informative in the case of pure independent noise. For a detailed description of the kNN method and its assumptions, please refer back to section 2.7.1 in chapter 2. A summarized overview of the method is given here.

The procedure starts by splitting the time-series into a training set, which will be used as the base information, and a testing set, which will be used to assess the quality of the forecasts. Suppose we wish to forecast one step in the future. That is, if the last observation was at time n, the forecast will be for time $n + 1$.

Define a window length m, and look back in the training set for other windows of length m that are similar to $(n - m + 1,\ \ldots\ ,\ n - 1,\ n)$. Use the next value of these windows to forecast the value at time $n + 1$. Repeat these steps for the whole length of the testing set, and compare the forecasted values with the original values from the testing set by calculating the correlation between them.

Note that we could have chosen to forecast taking larger steps into the future, that is, instead of forecasting $n+1$ one could have chosen to forecast $n+T$. By repeating this process with different combinations of m and T, the result is a map of the correlation between the forecast series and the testing set.

As previously mentioned, a key feature of a chaotic system is that arbitrarily close starting points will generate predictions that diverge exponentially with time (Ascioti et al., 1993, Cvitanović et al., 2016, Sugihara and May, 1990, Zounemat-Kermani, 2016). From this it stems that the quality of a forecast made for one time step ahead will be exponentially better than the quality of a forecast made for, say, twenty time steps ahead; since the uncertainties on the forecast values will lead to an exponentially fast divergence from the actual values of the process the further ahead one attempts to forecast. The kNN method makes use of this to look for a signature of chaos in time-series observations.

From this map constructed by repeating the kNN procedure for different combinations of m and T, it is possible to assess how feasible it is to forecast a particular variable, and to infer characteristics of its dynamical behaviour. Deterministic chaotic series, for example, will produce maps that present an exponential decay to zero in the forecast quality with T for all values of m.Figure 2.2 in chapter 2 illustrates some of the patterns produced by different types of time-series.

The window length parameter m is called the "embedding dimension". In general, embedding procedures are useful when a single variable is used to approximate an unknown system. Here it is a suitable approach to check for chaos and nonlinearity, since even though data is available on a group of variables which are part of the plankton system, the precise relationship between the plankton variables is unknown, as seen from the VAR results from chapter 4. Consequently, it is best to first understand the particularities and dynamics of each time-series individually, before considering them together as a group for further analysis, a topic not covered in this thesis.

In chapter 4, the seasonal cycle was identified in the linear VAR analysis as one of the main sources for plankton variability. In order to verify the influence of seasonality in the system, and if this relationship still persists from a nonlinear perspective, I decompose each series as a sum of a seasonal, trend and noise components, using the seasonal-trend decomposition from Cleveland et al. (1990). This decomposition is a filtering technique, implemented as part of the standard library in R, and considered to be one of the most versatile and robust methods for time-series decomposition (Hyndman and Athanasopoulos, 2012). By repeating the kNN analysis on the series with the removed seasonal component and on the trend-only component, it is possible to identify whether the different components of the series follow a similar dynamics, or if any specific component is dominant in the kNN map obtained from the original series. Figure 5.1 shows the output of this decomposition for the PCI time-series, and the decomposition for the other plankton and environmental series can be found in appendix C.

I then use the S-maps method proposed by Sugihara (1994) to investigate the presence of nonlinearity in the system (more details in section 2.7.2). This is a variation of the kNN method, that uses different weights when averaging the contribution of each window, inversely proportional to the distance to the point that is being forecasted. A

specific parameter θ controls the strength of these weights, and allows the assessment of the presence of nonlinearity in the series.

In order to calculate the S-maps, the first step is to identify the best window length (parameter m in the kNN algorithm) for each time-series. This is done by using the kNN method with varying m and $T = 1$, and selecting the m that yields the forecasts with best skill, that is, best correlation between the testing set and the forecasts. For that m, the forecast of the testing set is performed repeatedly with varying θ.

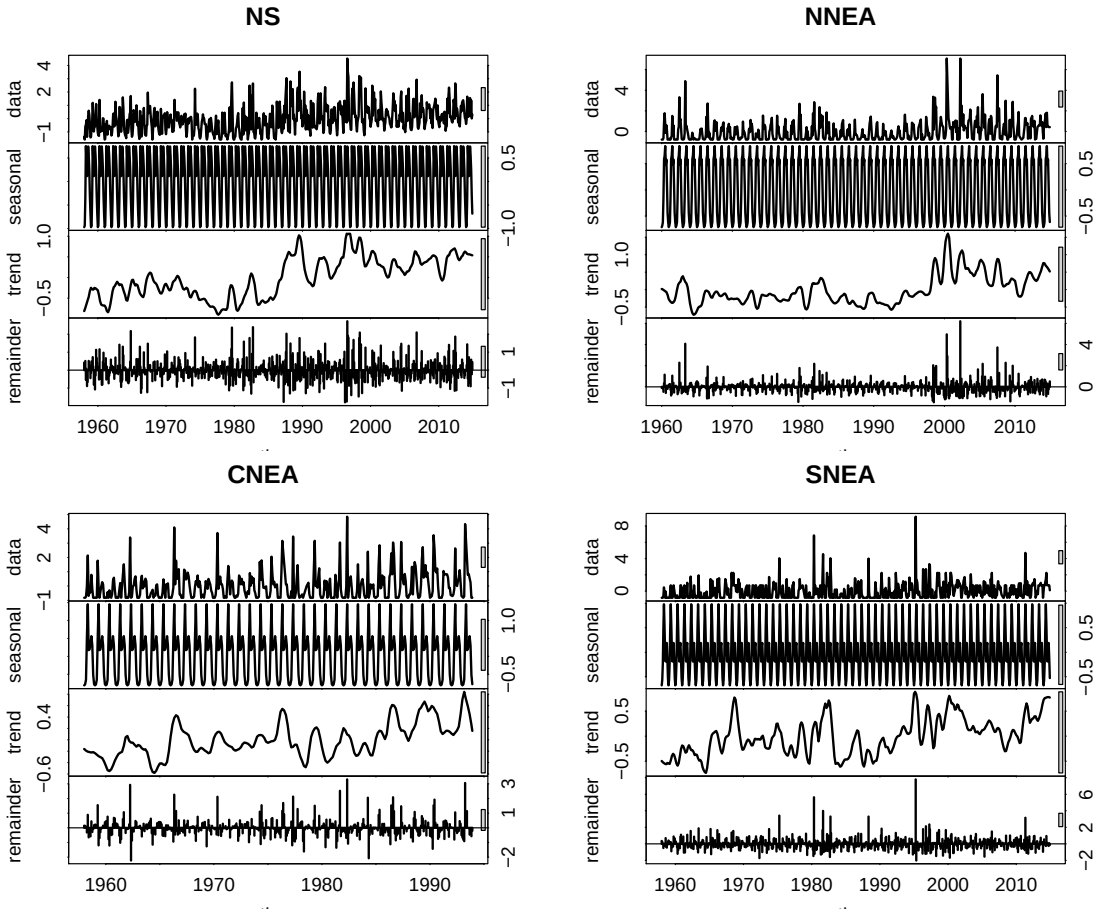

FIGURE 5.1: Decomposition of the PCI time-series into a seasonal, trend and noise components, using the STL method of Cleveland et al. (1990). The decomposition of the other plankton and environmental time-series can be found in appendix C, page 143.

The weighting function is defined as (Sugihara, 1994; and equation 2.19 in chapter 2):

$$weight(d) = e^{-\theta(d/\bar{d})} \tag{5.1}$$

where d is the distance between the present and the neighbour value, and $\bar{d}$ is the average distance between neighbours.

When θ is equal to zero all weights are equal to one, which corresponds to a globally linear scenario. Increasing values of θ give more weight to local influences and correspond to a nonlinear scenario (Sugihara et al., 2012, Ye et al., 2015). By calculating the correlation between the testing and forecast sets for different values of θ it is possible to identify if nonlinearity is present in the data. For linear series $\theta = 0$ will yield the best correlation, and a monotone decrease in forecast skill will be observed as θ increases. On the other hand, nonlinear series will present an increase in skill for some subset of positive values of θ.

S-maps require that the series are stationary, which is not always the case for the plankton. Therefore, for this part of the analysis, I ensured that the stationarity criterion is met by taking first differences of the plankton time-series, as previously done in chapter 4.

Finally, I use convergent cross-mapping (CCM) (section 2.7.3 and Sugihara et al., 2012, Ye et al., 2015) to investigate the presence of causal relationships between SST, MLD and plankton abundance. This method is more efficient than correlations to identify links between variables in nonlinear systems (McGowan et al., 2017), and could also provide a different insight in the relationship between the plankton and the environment than what was indicated by the VAR models considered in chapter 4. Identifying causality with CCM requires two steps: (1) the CCM skill (that is, the correlation between the predicted and testing sets) must increase with the length of the training set, and (2) the CCM skill must be statistically significant. Although the CCM presented an increase in skill, which could indicate SST and MLD as causal drivers of variability in plankton abundance, the correlations were not statistically significant (p larger than 0.1 in all cases, calculated from 1000 repetitions of surrogate data generated for each target time-series, following McGowan et al., 2017), thus rendering the analysis inconclusive. The CCM analysis was repeated including the North Atlantic Oscillation (NAO) index as another potential causal variable, but again no statistically significant relationship with the plankton was found.

5.2 Results

5.2.1 Stochasticity versus chaos

Chaotic systems are characterised by an exponential decay to zero of the correlation between the testing and forecast sets with increasing values of T for all m. From the first column of figures 5.2 - 5.6, it can be seen that the quality of the forecasts did not decrease with T for any of the plankton variables considered, in all regions. Therefore, it seems unlikely that the interannual variability in phytoplankton and zooplankton abundance arises from deterministic chaos. In addition, the plankton time-series presented a non-uniform structure that seems to depend on the window length m rather than on T. This was especially evident in the PCI, dinoflagellate and small copepod series. This structure seemed to be related to the twelve-month periodicity of the data, and is consistent with the signature displayed by periodic time-series with additive noise (panel (c) of figure 2.2). The environmental time-series considered, SST and MLD, also presented a signature that is consistent with a periodic series with added noise (first column in figures 5.7 and 5.8).

Some specific features can be observed in the kNN output for the original time-series, first column of figures 5.2 - 5.6. The North Sea (NS) is the region where the plankton present the most regular seasonal signature, with the exception of large copepods, which did not have a good forecasting skill in this region, and could almost seem to be drawn from a random distribution (panels (a)-(b) of figure 2.2). The consistency in PCI, diatoms, dinoflagellates and small copepods is likely to be a consequence of the fact that the NS region is well covered by CPR transects through time. The poor predictability of large copepods is likely due to the patchy distribution of zooplankton, which can present large irregular abundance peaks in the horizontal sampling of the CPR (Reid et al., 1990).

In the northern Northeast Atlantic (NNEA), however, large copepods presented a reasonable predictability, and a signature consistent with that of a seasonal process. The same was observed in the other plankton time-series, although the signature was weaker in diatoms this time. In the central Northeast Atlantic (CNEA) again the seasonal signature was present in the PCI, dinoflagellate and small copepods, and weaker in diatoms, while large copepods once again have a signature that seems to be closer to a random process.

In the southern Northeast Atlantic (SNEA) we can observe a curious feature. A clear periodic structure is not so evident when considering the entirety of the map, however for a subset of possible values of T the seasonal signal could be present. This is the only region in which there is an irregularity in the environmental maps, as well as the predictability of the MLD is weaker for T smaller than 12. This phenomenon could indicate some lagged effect of MLD on its own predictability, and this effect could be propagating to the plankton. From the annual series of MLD in this region (figure D.5 in appendix D, page 149), notice an abnormal increasing trend in MLD since the early 2000s, which is not present in the other regions. In order to identify the causes of this irregularity in the MLD, a deeper study of the physicochemical properties of the region would be required.

Since the series have such a clear seasonal signature, and considering the results from the previous chapter which indicated a strong influence from the seasonal cycle on interannual plankton variability, I repeated the analysis after removing the seasonal component from the series (second column of figures 5.2 - 5.6). A clear stochastic signature emerged in the kNN map. The correlations between the testing and predicted sets were poor across all plankton variables, with an absolute value below 0.4 in almost all cases. Once again there was no evidence of a decrease in forecast quality with T, which implies that chaos was also not present in the underlying dynamics of the plankton, after the removal of the seasonal signal. The fact that a stronger stochastic signature appeared with the removal of the seasonal cycle indicates that the natural periodicity would be acting as a stabilizer of the plankton variability, adding a layer of predictability to the system.

The SST and MLD series were also examined after removing their seasonal cycle (second column of figures 5.7 and 5.8). Their kNN maps presented an irregular decay in forecast skill with increasing T. This decay, although more regular than that observed in the plankton, was still not as uniform on m as it would be expected of chaos (, and section 2.7.1), and could also correspond to that of a structured stochastic process , such as an autoregressive moving-average (ARMA) process (section 2.3, and panels (b) versus (d) in figure 2.2), although with the current analysis alone it is not possible to determine which is the precise structure of the data.

Finally, I considered the trend-only component of each time-series, that is, the series with removed seasonal cycle and noise signal (third column of figures 5.2 - 5.6). Working with the trend-only component is not appropriate from a forecasting perspective, since it

is not possible to accurately forecast the noise component separately from the rest of the series, but it can provide insights on the inherent nature of the smoothed component of the series. This would correspond to an ideal experimental situation, free of any influence or variability due to external factors.

In the trend-only component of the plankton series, the kNN maps presented an irregular decay in predictability with increasing T. This pattern could ultimately be consistent with chaos, in the cases where the decrease is uniform for all m, such as the map of diatoms in the NNEA. However, in many of the regions and plankton variables, the decrease is too irregular (e.g., dinoflagellates in the southern Northeast Atlantic) to be unequivocally classified as chaotic; and the patterns seen in the maps could also be a consequence of strong autocorrelated structures (Sugihara, 1994, and panel (b) of figure 2.2). Furthermore, this decreasing signature did not emerge unless core characteristics of the data (that is, seasonality and the noise component) were completely filtered out. Therefore, even if a signature of chaos were absolutely clear, it would still be misleading to claim that this decay in the trend-only component as sufficient evidence for chaos.

In the SST and MLD series, the removal of the seasonal and noise components revealed kNN maps with a much more uniform decay of forecast skill with increasing T than that observed in the plankton. Especially in SST, this decay is rapid and independent of m, which resembles the signature of chaos. This result is not entirely unexpected: the variations in these physical conditions are less subject to fluctuations than the biological variables, since their variability depends only on the physical environment. Nevertheless, this tentative deterministic signature is still dominated by the stochastic and seasonal components.

5.2.2 Nonlinearity of the plankton

Another objective of this chapter was to investigate the presence of nonlinearity in the plankton time-series. This was motivated from the strong autocorrelation component identified in the VAR analysis from chapter 4, which was discussed as potentially being a consequence of nonlinearity present in the system (Barton et al., 2015).

In figure 5.9 the first panel presents some examples of synthetic series, for ease of comparison and interpretation of the results. The first case is that of the highly deterministic

series (chaotic and chaotic + noise). Observe that there is a considerable increase in forecast skill with the increase of the nonlinearity parameter θ.

The second case is that of the less irregular nonlinear processes and non-stationary processes (chaos + cycle and first-differenced ARMA), which present an increase in skill for some values of θ, but with the skill eventually decreasing. This signature is characteristic of nonlinear, but non-chaotic, systems. Another example is that of the predator-prey system discussed in section 2.7.2 (figure 2.5). This was a deterministic, nonlinear periodic system with added noise, and also presented an increase in forecast skill only for a subset of values of θ.

The third case is that of linear stationary stochastic processes (white noise and ARMA). Both present a decay in forecast skill with θ, since the best forecasts are obtained when the S-map model is linear ($\theta = 0$). Note, however, that the skill for a white noise is practically zero, a consequence of the unpredictability of the white noise. On the other hand, the serial correlation in the ARMA makes it easier to predict, and thus the forecast skill is much higher than that of the pure white noise.

In the following panels in figure 5.9, it can be seen that the S-maps indicate a nonlinear structure present in all of the plankton time-series, and that this structure would not be chaotic. All maps present an initial increase in forecasting skill with θ, followed by an eventual gradual decrease. This is compatible with the second case described above, and seen in the example panel. Combined with the result from the kNN maps, this analysis indicates that the plankton are best described as a nonlinear, periodic system with a stochastic component.

It is interesting to observe that, for PCI, diatoms, dinoflagellates and small copepods, the best forecast skills were obtained in the North Sea (NS), and the worst skills in the southern Northeast Atlantic (SNEA). This is coherent with the results from the kNN analysis, where the most and least regular maps were observed in the NS and SNEA, respectively. It is also worth noting that large copepods presented a weak forecast skill in all four regions, again a likely consequence of the irregularities present in their time-series due to the patchiness of zooplankton in the ocean (Lalli and Parsons, 1997).

The environmental variables, SST and MLD, were also compatible with a nonlinear structure. SST presented a high forecast skill for all regions, a consequence of its regular seasonality also observed in the kNN maps. MLD also presented a good forecast skill in all

regions. It is worth noting that the MLD presents a similar nonlinear signature across the Northeast Atlantic subregions, which is different from the nonlinear signature observed in the North Sea, indicating a potentially different dynamics in these two areas.

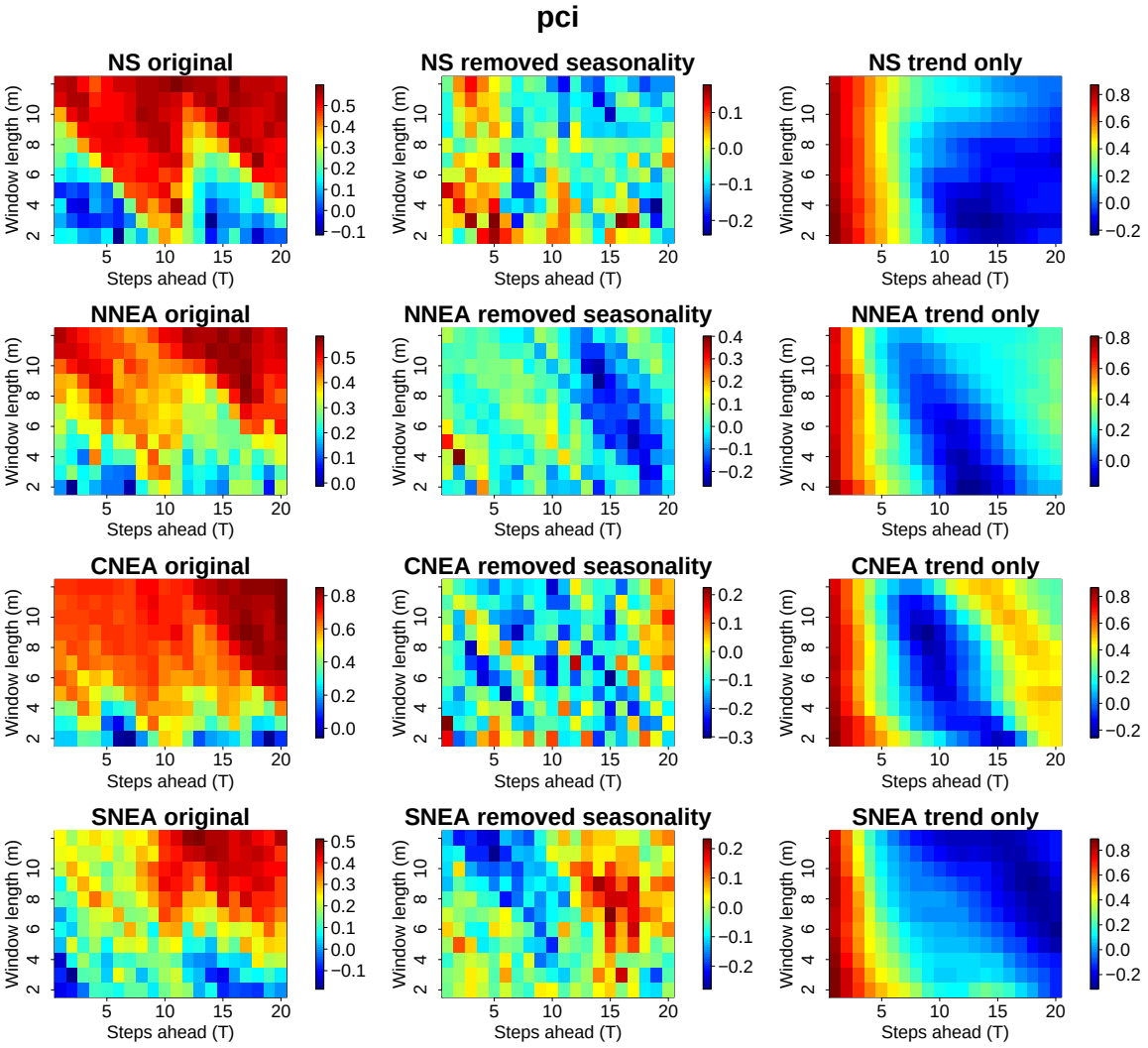

FIGURE 5.2: kNN method applied to the PCI original time-series (left column), series with removed seasonal cycle (middle column), and trend-only series (right column) in the North Sea (NS, first line), Northern Northeast Atlantic (NNEA, second line), Central Northeast Atlantic (CNEA, third line) and Southern Northeast Atlantic (SNEA, fourth line). The colour scale corresponds to the correlation between the forecasted and testing sets for each combination of m and T. Refer to figure 2.2 for a guide on possible patterns.

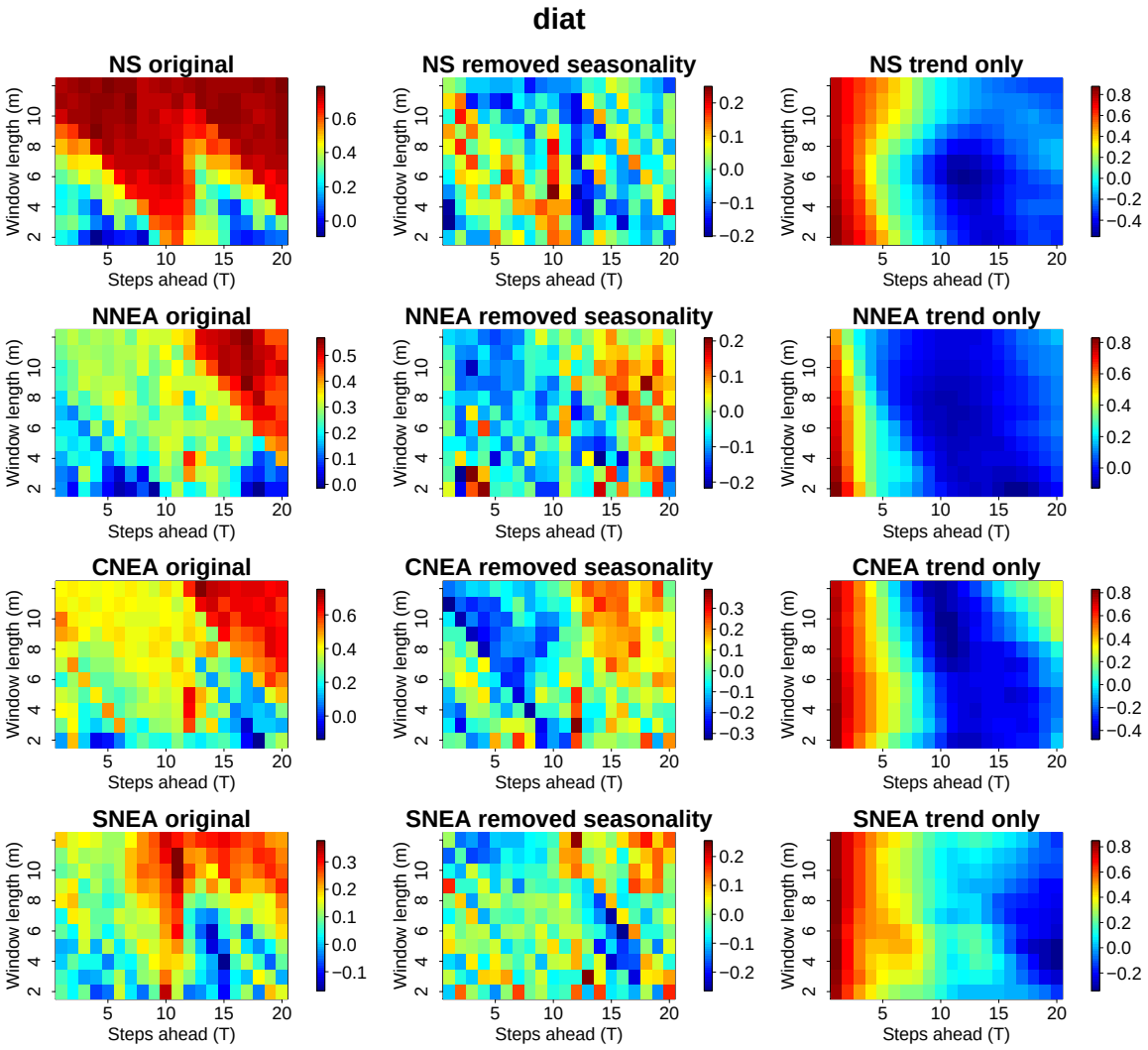

FIGURE 5.3: kNN method applied to diatom (diat) abundance original time-series (left column), series with removed seasonal cycle (middle column), and trend-only series (right column) in the North Sea (NS, first line), Northern Northeast Atlantic (NNEA, second line), Central Northeast Atlantic (CNEA, third line) and Southern Northeast Atlantic (SNEA, fourth line). The colour scale corresponds to the correlation between the forecasted and testing sets for each combination of m and T. Refer to figure 2.2 for a guide on possible patterns.

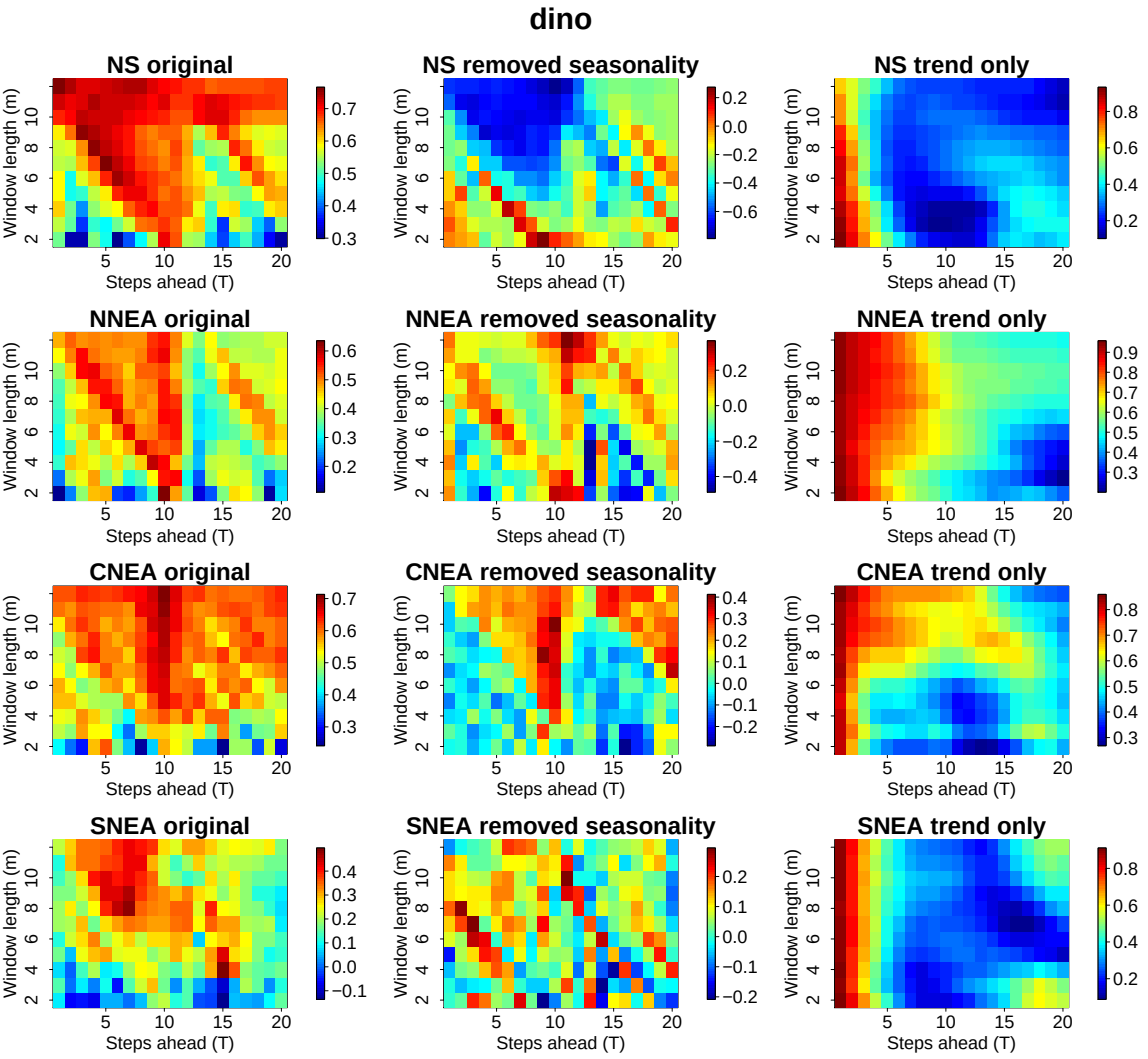

FIGURE 5.4: kNN method applied to dinoflagellate (dino) abundance original time-series (left column), series with removed seasonal cycle (middle column), and trend-only series (right column) in the North Sea (NS, first line), Northern Northeast Atlantic (NNEA, second line), Central Northeast Atlantic (CNEA, third line) and Southern Northeast Atlantic (SNEA, fourth line). The colour scale corresponds to the correlation between the forecasted and testing sets for each combination of m and T. Refer to figure 2.2 for a guide on possible patterns.

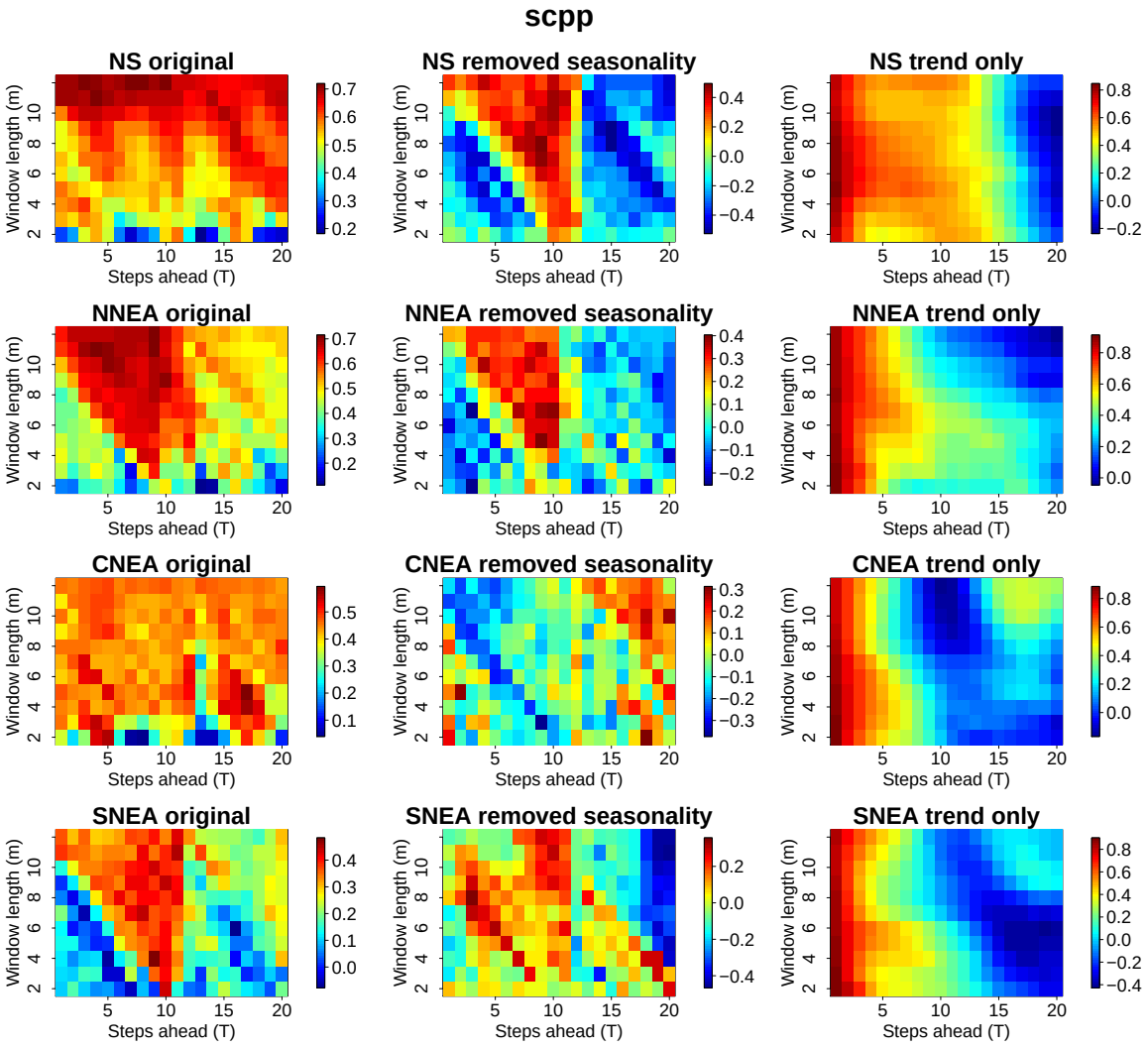

FIGURE 5.5: kNN method applied to small copepods (scpp) abundance original time-series (left column), series with removed seasonal cycle (middle column), and trend-only series (right column) in the North Sea (NS, first line), Northern Northeast Atlantic (NNEA, second line), Central Northeast Atlantic (CNEA, third line) and Southern Northeast Atlantic (SNEA, fourth line). The colour scale corresponds to the correlation between the forecasted and testing sets for each combination of m and T. Refer to figure 2.2 for a guide on possible patterns.

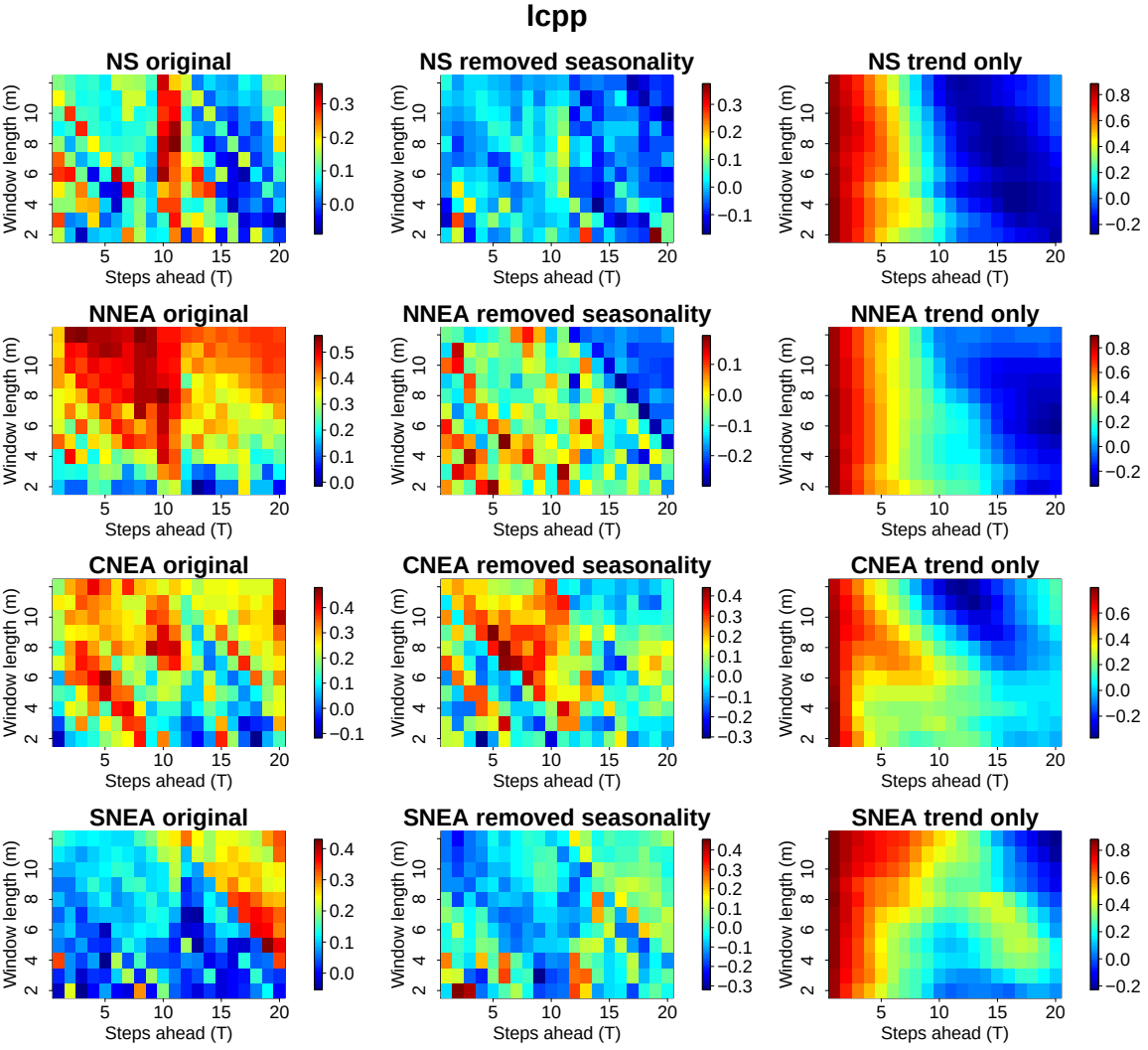

FIGURE 5.6: kNN method applied to large copepods (lcpp) abundance original time-series (left column), series with removed seasonal cycle (middle column), and trend-only series (right column) in the North Sea (NS, first line), Northern Northeast Atlantic (NNEA, second line), Central Northeast Atlantic (CNEA, third line) and Southern Northeast Atlantic (SNEA, fourth line). The colour scale corresponds to the correlation between the forecasted and testing sets for each combination of m and T. Refer to figure 2.2 examples of possible patterns.

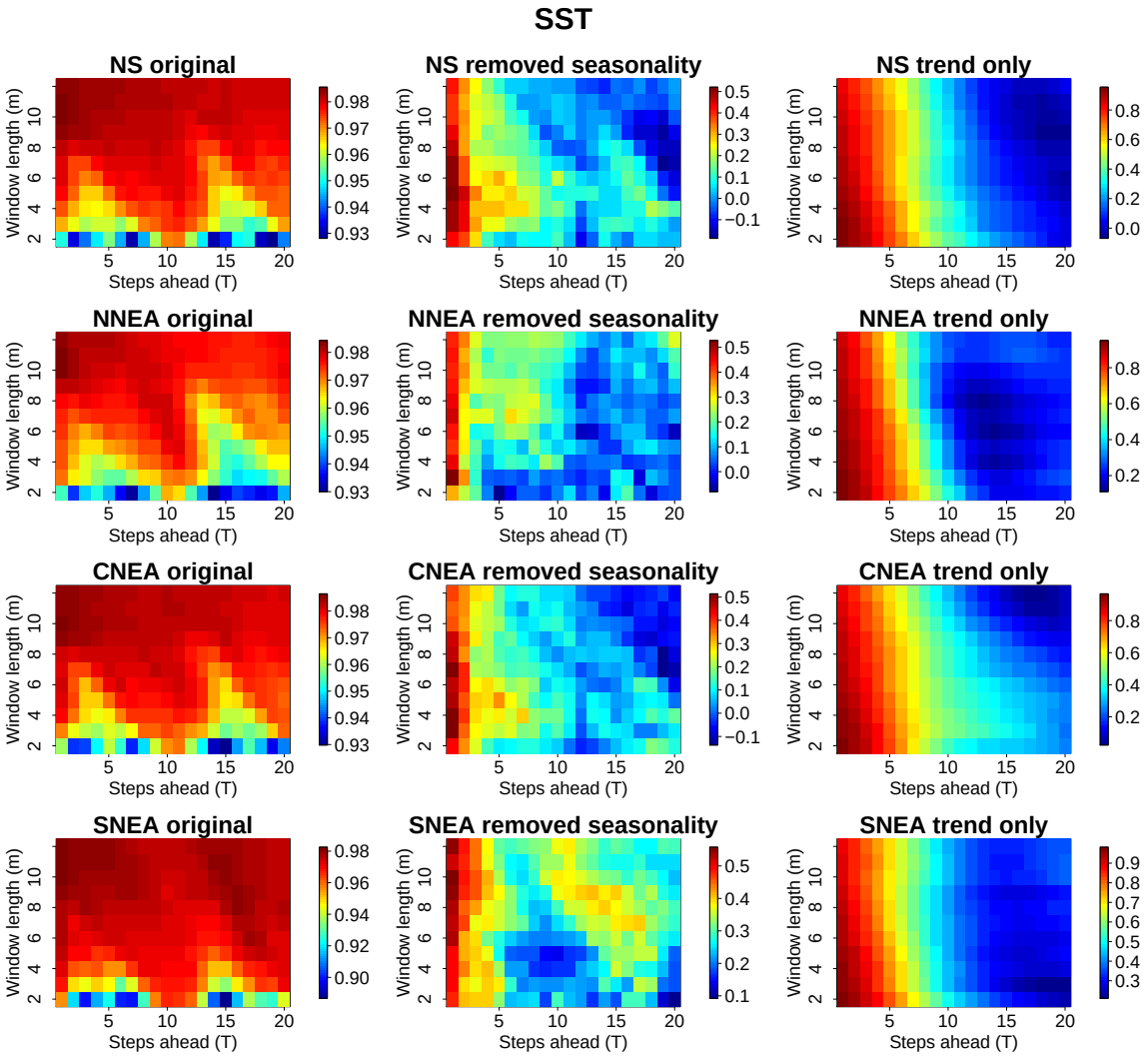

FIGURE 5.7: kNN method applied to the sea surface temperature original time-series (left column), series with removed seasonal cycle (middle column), and trend-only series (right column) in each of the four subregions considered. The colour scale corresponds to the correlation between the forecasted and testing sets for each combination of m and T. The high quality of the forecasts for all combinations of parameters evidences the regular cyclical structure in the original data.

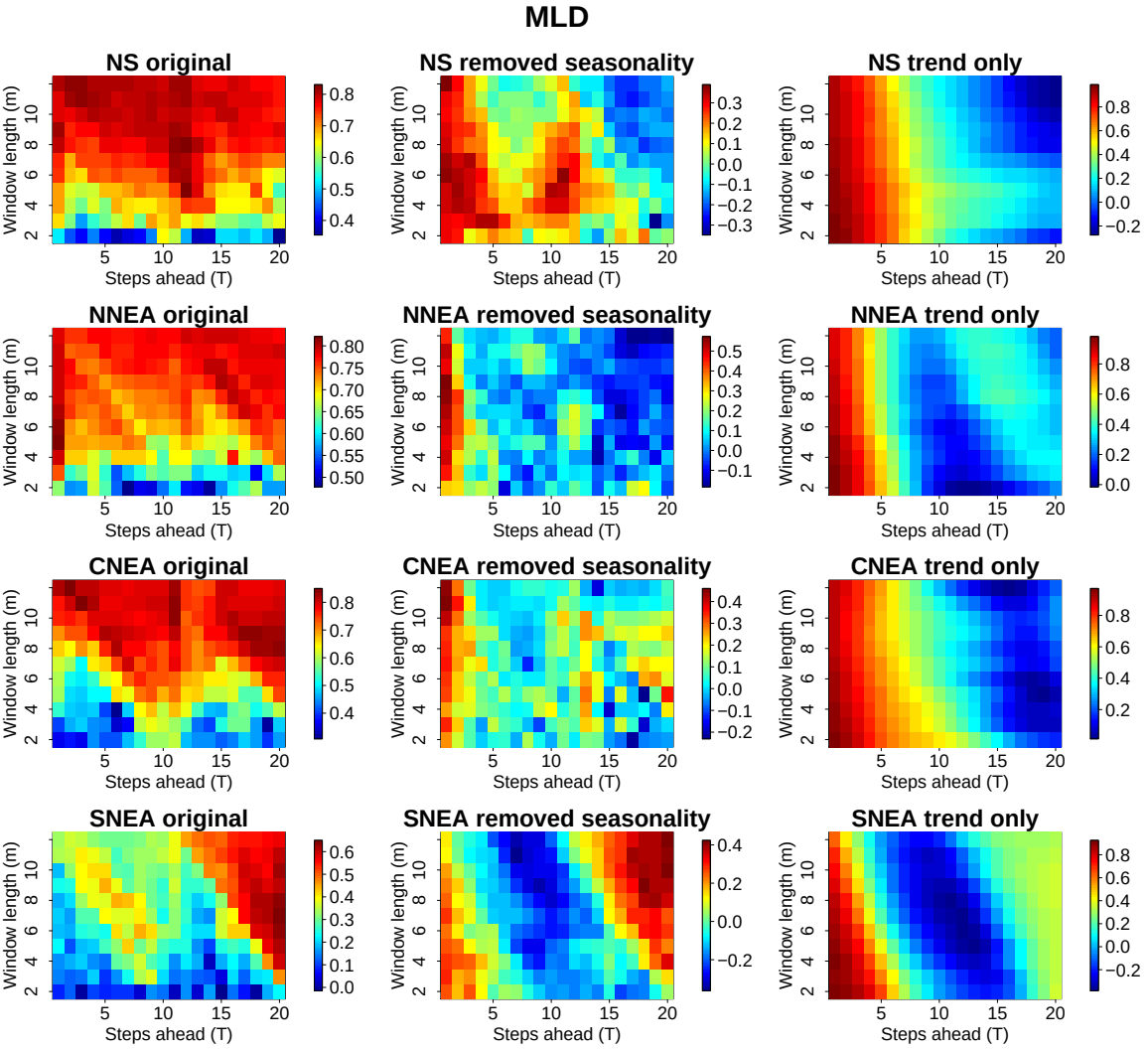

FIGURE 5.8: kNN method applied to the mixed layer depth original time-series (left column), series with removed seasonal cycle (middle column), and trend-only series (right column) in each of the four subregions considered. The colour scale corresponds to the correlation between the forecasted and testing sets for each combination of m and T. Although not as much as in the sea surface temperature series, the high quality of predictions for most combinations of parameters also indicate a regular structure in the original data.

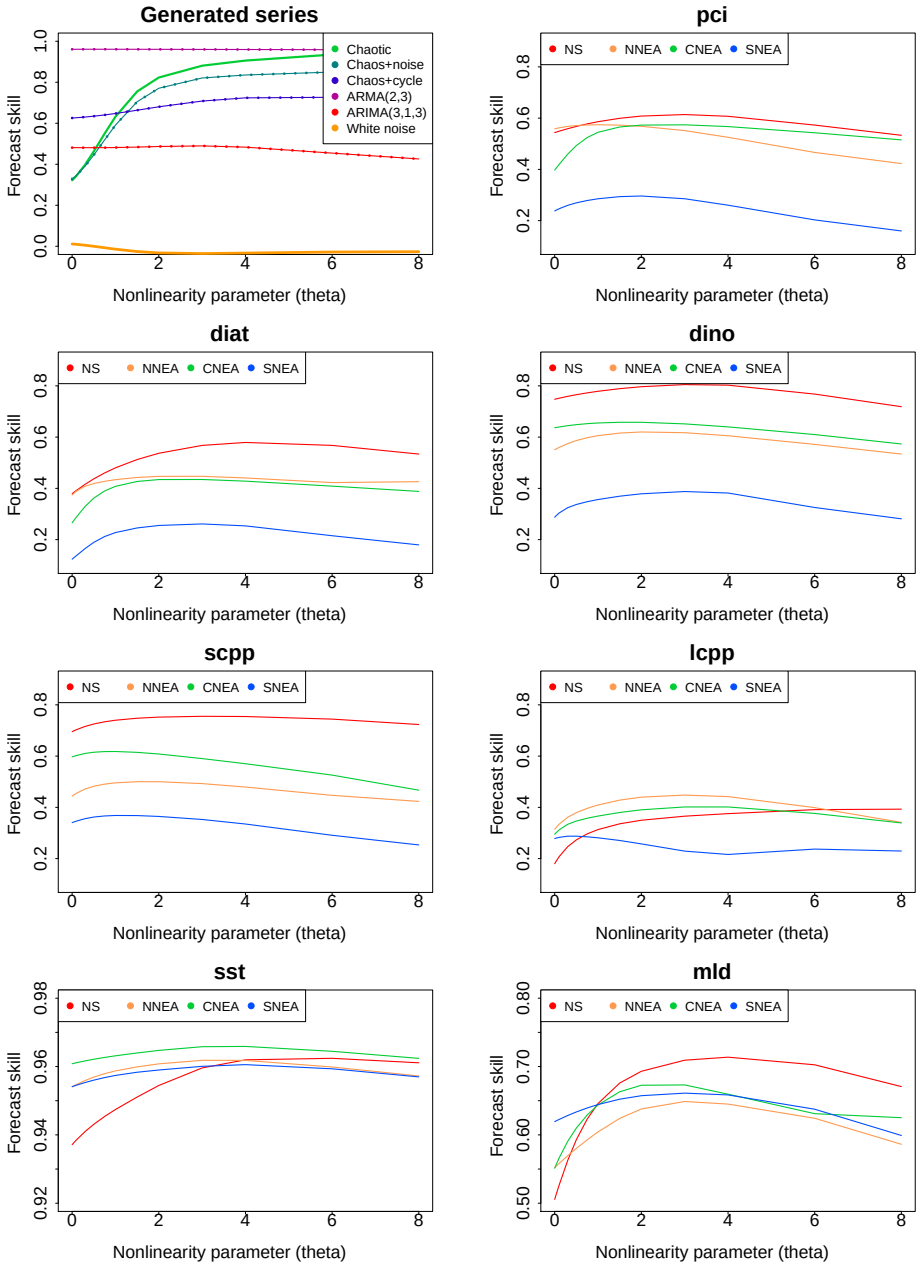

FIGURE 5.9: The first panel shows the S-maps for a set of synthetic series: a chaotic series, the same series with added white noise, the same series with added noise and seasonal cycle, a stationary autoregressive moving-average process (ARMA, section 2.3), a first-differenced stationary ARMA process, and a pure white noise series. The following panels show the S-maps calculated for the first-differenced time-series of PCI and abundance of diatoms, dinoflagellates, small and large copepods in the four regions considered. The last two panels show the S-maps for the first-differenced time-series of SST and MLD, the two environmental variables considered in this chapter.

5.3 Discussion

Determining whether deterministic chaos is present or absent in natural ecosystems is a challenging task (Zimmer, 1999). The work of Benincà et al. (2015) is, to my knowledge, the closest case in which deterministic chaos was positively identified in a marine community. Their case-study consisted of a cyclic succession in an intertidal community with periodicity of about two years, located in a fairly stable environment in the north of New Zealand. Bare rock was colonised by barnacles, which were invaded by crustose algae, and later dominated by mussels. When the mussels detached from the rock, the cycle started again. The cyclic succession produced irregular species fluctuations, and their results indicated that the community alternated between chaotic and stabilising dynamics, with seasonal environmental forcing as a main driver of the system dynamics.

In the plankton literature, Ascioti et al. (1993) analysed phytoplankton and zooplankton biomass time-series from a station in the Mid Atlantic Bight and could not conclude if chaos was present, although there was support for nonlinear trophic dynamics present in the system. Sugihara and May (1990) analysed diatom abundance data from the Scripps Pier and concluded that, although partially consistent with a chaotic attractor, at least half of the variability observed could be explained by additive noise.

I found no evidence for deterministic chaos in the CPR time-series of the PCI, diatom, dinoflagellate, small or large copepod abundances. The signature seen in the first column of figures 5.2 - 5.6 is mostly consistent with that of a periodic series, rather than with that of a chaotic one (panel (c) in figure 2.2). In the cases where a periodic signature is not so clear, the signature is still not consistent with that of chaos, but rather with that of a structured stochastic process (panel (b) in figure 2.2). Seasonality was previously identified by the linear VAR analysis in chapter 4 as one of the main drivers of plankton variability, and this fact is now confirmed by a different methodology based on a nonlinear premise.

To gain further insights on the structure of the series, the analysis was repeated in the series with removed seasonal cycle, in order to check if any underlying structure was present underneath the periodic signal. The results indicated a signature which resembled that of random noise, with poor forecast skill across all plankton variables and regions. The regular signature in the original series, together with the random patterns observed when seasonality is removed, imply that the plankton seasonal cycle, which depends on a

combination of changes in light, temperature, nutrient concentration and mixing, add an important layer of regularity and predictability to the plankton dynamics.

As mentioned in the opening of this chapter, Benincà et al. (2008) found evidence of deterministic chaos in an experiment that lasted for a little over six years, sampling a large number of plankton species twice a week during that time. The authors used two different methods, based on Lyapunov exponents, to positively identify a signature of chaos in their community. Laboratory experiments can create an artificial environment with removed external disturbances, to reveal the underlying dynamics in plankton interactions.

The equivalent of holding exogenous conditions constant in the present results would be to consider the time-series with removed seasonal cycle and noise components. The signature present in the trend-only component of the plankton time-series (third column of figures 5.2 - 5.6) does present a decay that could be indicative of chaos, which would be consistent with the experimental results of Benincà et al. (2008). However, in most cases, the decay was not uniform for all values of m, which would be expected from pure deterministic chaos. In their original work, Sugihara and May (1990) point out that differentiating between autocorrelated noise and chaos in natural systems is considerably challenging, since serial correlation can yield signatures similar to those of a chaotic system in the decay of forecast skill. Therefore, it is not entirely possible to determine whether the trend-only plankton series corresponds to a strongly autocorrelated or chaotic process.

In order to directly compare my results with the work of Benincà et al. (2008), I also considered the use of Lyapunov exponents to investigate the presence of chaotic dynamics. Lyapunov exponents are a way of quantifying the sensitivity of a system to perturbations in initial conditions, by measuring the divergence of different trajectories in time. The calculation of Lyapunov exponents assumes that the system can be written as an underlying noise-free dynamics with an overlapped noise component, and the estimation of the exponent is done on the theoretical noise-free component. Rosenstein et al. (1993), for instance, present an in-depth background and definition of Lyapunov exponents, as well as a method to implement its calculation. However, Sugihara (1994) argue that Lyapunov exponents may not be appropriate for ecological time-series, since their theoretical assumptions often are not met by short and noisy datasets, which are common in ecological systems. The author suggests that forecast methods, such as the ones applied here, would be more appropriate for ecological data, as they do not make any distinction or assumption

about the process noise. With that in mind, I still attempted to compute the Lyapunov exponent for the CPR plankton time-series. However, one of the local linearity criteria needed to complete the calculation of the exponent was not met by any of the plankton time-series in any of the regions considered here. Thus, the calculation of the maximum Lyapunov exponent could not be performed, hindering the possibility of a direct comparison of my results with the ones of Benincà et al. (2008).

It is important to keep in mind that the removal of external perturbations, either in laboratory or through a decomposition of observational data, can in itself alter the plankton dynamics to an artificial state. Theoretical experiments have shown that competition and predation can induce chaos in a plankton system even in the absence of external perturbations (e.g., the theoretical model of Huisman and Weissing, 1999, and laboratory results of Benincà et al., 2008). Sugihara (1994) and Zimmer (1999) suggest that, in addition to the propagation of random fluctuations, the natural response of ecosystems to external conditions has the potential to alter the whole underlying structure of the process. Sugihara (1994) states that natural systems will always have some unexplained noise component that is inherently part of the system, and therefore attempting to completely isolate an ecosystem from noise might lead to artificial results.

It is more common for deterministic models to present chaotic behaviour when they are defined by few variables. Zimmer (1999) provides some examples such as that of trophic interactions, which can present deterministic chaos when modelled as a linear chain (primary producers in one level, herbivores in the next and so on). That is because oscillations in one trophic level can be reflected or amplified as they propagate on to other levels. However, if the trophic model is designed to incorporate more realistic aspects of food webs, such as same trophic level interactions or omnivorous predators who feed on multiple lower levels, these extra connections diminish the effect of chaotic oscillations, and the final model is often not chaotic. This does not imply that theoretical models which present chaos are wrong. Every theoretical model provides an approximation of reality, which is useful in interpreting and understanding the natural phenomena.

For instance, theoretical plankton models often find indications of chaotic dynamics (Huisman et al., 2006, Moroz et al., 2016), which is not surprising since these models tend to be based on a few key parameters. But it is important to acknowledge that, in observational data, it is difficult to find a clear separation between noise and chaos (Scheffer et al., 2003).

In natural systems, it is hard to identify or measure an appropriate subset of parameters to describe an ecosystem, and it is also necessary to account for the influence of external noise. The complex interactions in ecological systems, such as the capability of predator species to switch diets in the absence of a preferred prey, can yield strongly nonlinear, yet non-chaotic, dynamical systems (Zimmer, 1999). Evidence for nonlinearity is precisely what was found here in all of the plankton time-series (figure 5.9), even if no chaotic signal could be identified in the plankton. The results presented are thus coherent with what is expected from complex natural systems.

CCM is a powerful technique to infer causal relationships in nonlinear systems, where simple correlations are often spurious. A classic example of when this method was applied successfully is that of Sugihara et al. (2012). The authors investigated the Sardine-Anchovy system at the Scripps Pier and Newport Pier in California, and demonstrated that, although there was no direct causal relationship between the two species, both were subject to external forcing from sea surface temperature. Another recent example is that of McGowan et al. (2017), who successfully applied the CCM method to identify potential drivers of phytoplankton blooms along the coast of California. The authors used S-maps to demonstrate that phytoplankton blooms at the Scripps pier are nonlinear, and CCM analysis to positively identify causal relationships between chlorophyll-a and a set of variables that indicate water column stability and nutrient concentration. The authors also hypothesize that different bloom events may have different underlying causes in the region.

Here, the CCM method failed to identify a statistically significant causal link between SST, MLD and the NAO index with the plankton. This does not necessarily mean that no causation is present, since the importance of SST, MLD and the NAO to the plankton community is well known (e.g., Edwards and Richardson, 2004, Henson et al., 2013, Longhurst, 2006, Richardson, 2008). Instead this indicates that these environmental variables alone may not be sufficient to represent the dominant environmental conditions driving the plankton. According to Ye et al. (2015), the lack of a statistically significant CCM relationship can mean that other external factors may also be necessary to be positively identified by the method. Since the CCM analysis is based on a nonlinear, and consequently non-additive premise, it may not be possible to identify the individual contributions of interacting external drivers. This means that a single variable on itself may not improve the skill of the model, even if it becomes significant when combined with other

drivers (in the example of Ye et al., 2015, temperature and river discharge as causal for fish recruitment).

Plankton populations are subject to a large set of environmental controls (sections 1.2 to 1.4 and references thererin), and it could be that an incomplete set of these interacting drivers, such as the one explored here, is not sufficient to improve the forecasts, and consequently do not return a significant CCM correlation (Ye et al., 2015). For example, it could be that a combination of MLD and nutrient concentration would emerge as a significant cause for plankton variability, while MLD alone might not significantly improve the model. Unfortunately, the absence of long-term data on other physicochemical factors prevent the further investigation on this front at present. Nevertheless, the lack of a significant causal link between SST, MLD or the NAO and the plankton imply that it would not be possible to point at any of these factors alone as being the main driver of plankton variability. The field of empirical dynamics modelling is still growing, and the development and implementation of novel techniques such as multivariate CCM (Deyle et al., 2016b) could help resolve these issues in the near future.

The fact that no significant causal relationship was found between the environmental variables and the plankton, allied with a significant stabilising effect from the seasonal cycle, imply that in order to thoroughly assess the impact of the environment in the plankton it would be necessary to consider a wider set of environmental variables in the analysis. The nonlinear nature of the plankton series and the results from the CCM analysis indicate that complex interactions from the environment are more likely to drive the interannual variability in the plankton than individual environmental factors. It may not be possible to single out one variable as the main driver of interannual plankton dynamics, contrary to what has been suggested in previous studies (such as, for example, Beaugrand et al., 2002, Zhai et al., 2013 and Goberville et al., 2014).

In summary, in this chapter I evaluated the presence of nonlinearity, stochasticity and chaos in the dynamics of the plankton of four subregions of the Northeast Atlantic and North Sea. The presence of a nonlinear structure, and evidence supporting the presence of stochasticity rather than chaos in the system, could limit the possibility of performing short-term forecasts of plankton abundance. However, the annual seasonality brings a layer of regularity to the system, which could be explored from a predictive perspective. The fact that the CCM analysis failed to identify causal links between the SST, MLD and the

plankton indicates that a broader combination of external factors is likely to be driving the plankton of this region. The complex relationships that govern plankton dynamics render it unlikely that a single environmental factor can be specified as the main driver of the variability in the plankton dynamics of the region. I suggest that a nonlinear seasonal structure with a stochastic component would be the most appropriate to model the plankton of the Northeast Atlantic and North Sea. The feasibility of short-term, data-based forecasts of plankton abundance from the CPR time-series could help in the monitoring of the plankton system, and should be investigated in future studies.

Chapter 6

Conclusion

The general objective of this thesis was to understand and describe the long-term variability and interactions between key plankton groups and environmental indicators in the North Sea and Northeast Atlantic through the use of statistical models. Referring back to the questions raised in section 1.7, each chapter has provided answers as follows.

6.1 Is it possible to make inferences about individual phytoplankton groups from water greenness data?

6.1.1 Motivation

Due to their rapid life cycles and tight coupling to environmental variability, plankton have been proposed as potential indicators of ecosystem conditions in the ocean. Phytoplankton are of particular interest, due to their key contribution for the maintenance of life on the planet. However, long-term detailed data on phytoplankton are scarce on a global scale, and often proxy variables related to chlorophyll concentration are used to represent phytoplankton abundance. In the Northeast Atlantic, there is a unique opportunity to investigate the relationship between chlorophyll and phytoplankton abundance thanks to the Continuous Plankton Recorder (CPR) survey, which has been consistently recording both individual phytoplankton species and a water greenness index since at least 1958.

6.1.2 Chapter summary

In chapter 3, I aimed to better understand the relationship between the phytoplankton colour index (PCI) and the abundance of major phytoplankton groups, and to answer the question of whether it was possible to infer information about the community composition from the water greenness and vice-versa. I investigated the contrasting patterns in the time-series of the PCI and diatom and dinoflagellate abundances, all captured by the CPR survey. The visual intuition of an increase in the PCI over the past fifty years was confirmed by the informational change-point analysis, which identified either sudden increases in the mean of the PCI or positive trends across all regions, in agreement with previous reports from the literature. Meanwhile, the abundance of diatoms and dinoflagellates exhibited either a decrease or almost constant pattern with time.

This was a counter-intuitive result: diatoms and dinoflagellates are known to be the phytoplankton groups most accurately counted by the CPR survey, since the large mesh size tends to underestimate smaller species of phytoplankton. Some missing biological variable was conjectured to be actively contributing towards the increased water greenness. This hypothesis was reinforced by a multiple linear regression (MLR) model, fitted to describe the PCI as a function of diatom and dinoflagellate abundances. The results from the MLR model reinforced that the most likely scenario is that there must be at least one missing explanatory variable for the PCI.

Considering the sampling method and assessment of the PCI from the CPR silks, and the definition of the MLR model, it would be unreasonable to attribute the missing variability to environmental or methodological factors. Phytoplankton abundance was used in the MLR to describe the greenness of the same samples from which phytoplankton species had been counted. Changes in intracellular pigmentation are also unlikely to be the cause of the missing variability, since that would imply a large-scale adaptation across multiple species, geographical regions and time scales. Under this scenario, a hypothesis of a missing biological component is the most reasonable.

What could this missing link be? It is known that the CPR tends to underestimate populations of smaller-sized phytoplankton, but those may still contribute to the greenness index, as their chloroplasts may stain the mesh even if the cells get broken. The different trends in abundance, and different timings and types of change-points that were observed in

diatoms and dinoflagellates in this chapter support the hypothesis that groups of smaller-sized phytoplankton are likely to be growing in abundance in the region of study (Leterme et al., 2006), which is also compatible with the hypothesis that warmer waters may favour smaller-sized phytoplankton (Hilligsöe et al., 2011, Morán et al., 2010).

Higher SST is known to favour smaller species of plankton, and SST is also known to be increasing in the area considered (e.g., Hinder et al., 2012b, Richardson and Schoeman, 2004, and figure 1.1 in chapter 1). This change in the base trophic level of the ocean can have consequences for higher trophic levels, as the feeding preferences of growing copepods, the main herbivores in the region, are for larger diatoms. This raises the question of how much exogenous environmental factors and grazing effects from zooplankton contribute towards the phytoplankton variability. This was addressed in the following chapter.

6.1.3 Challenges and limitations

The model presented in this chapter to describe the PCI was an approximation reflecting the data currently available. Without more information about the abundance of phytoplankton smaller than diatoms and dinoflagellates, it is hard to infer exactly what percentage of the PCI comes from nano and picoplankton. The analysis would have benefited greatly from data on other key groups for the region of study, such as for example coccolithophores, which would have allowed me to better confirm the hypothesis that these groups also contribute significantly to the PCI. The CPR survey has started to account for some smaller-sized phytoplankton species since 1993, and once these series are long enough they should be used to provide a more accurate analysis on this matter.

Another possible route to explore this topic is through the use of hyperspectral methods on satellite data. These methods estimate information regarding individual plankton functional groups from satellite reflectance data (Nair et al., 2008, Xi et al., 2015), and have been successfully applied, for example, to identify and monitor phytoplankton blooms (red tides, Ryan et al., 2014; and cyanobacteria, Kudela et al., 2015). A similar analysis could be applied to compare the estimates of functional groups with the chlorophyll-a signal in the Northeast Atlantic area. The hypothesis of the model would be equivalent to those presented here: both the colour series and the species series would be computed from the same mechanism, and thus could be modelled directly from one another. The major challenge would be once again the length of the dataset (1998 -

present). Some exploratory time-series and residual analysis would be required to determine the necessary length of the series, confirm whether it is possible to use satellite measurements in this way, or if longer periods of observations would be required.

Without further information on smaller-sized phytoplankton, one could use the autoregressive moving-average component of the model presented here as a potential model for the dynamics of these phytoplankton. This approximation however may not be entirely accurate, since there is not yet confirmation that this signal consists only of the variability due to smaller phytoplankton, or how much of the smaller phytoplankton it represents. The results from this thesis suggest that the PCI may be a useful representation of the total phytoplankton biomass, although not necessarily reflecting nuances of the phytoplankton community. At present, it is not possible to extract information on the variability of any specific phytoplankton groups from the PCI, or to attribute the variability in the PCI to diatoms and dinoflagellates alone.

6.2 Is the variability observed in phytoplankton more dependent on changes in the environment, or on zooplankton grazing?

6.2.1 Motivation

Different trophic levels in an ecosystem can be linked through different types of control mechanism. Some of the most common are bottom-up control, when changes to the environment affect populations of primary producers and propagate up to higher levels of the food web, and top-down control, when changes to the populations of top predators propagate down to lower levels of the food web. Knowing the type of ecological control of a community is relevant, for instance, in ecosystems management and development of conservation strategies. In the case of phytoplankton, although there is evidence that grazing plays a major role on a seasonal scale (Barton et al., 2015), there is no agreement on whether bottom-up control from physical variables or top-down control from zooplankton grazing would be dominant from an interannual perspective. Although grazing is known to be important on a seasonal time scale, studies suggest that a bottom-up control

from the environment would prevail in the long term (e.g., Barton et al., 2015, Chassot et al., 2007, Richardson, 2008).

6.2.2 Chapter summary

In chapter 4, I investigated whether the phytoplankton of the Northeast Atlantic and North Sea were under bottom-up or top-down control from an interannual perspective. I used a vector autoregressive (VAR) model to describe the variability in monthly time-series of phytoplankton and zooplankton simultaneously, including their interactions, seasonality, autocorrelation, and the influence of sea surface temperature (SST) and mixed layer depth (MLD) as covariates indicating environmental conditions.

The VAR analysis allowed me to evaluate the interactions within the phytoplankton and zooplankton indicator variables, and between the plankton and the environmental variables. VAR models have been successfully applied to describe plankton blooms and plankton food webs in lakes, but their applications are still limited when it comes to oceanic plankton communities. This chapter thus provided a novel, large scale approach to this matter in the region of study.

I found little evidence for statistically significant interannual top-down control of phytoplankton in this regions. Diatoms especially were mostly unrelated to the other biological variables, and a similar pattern was observed when using the PCI to represent the phytoplankton community. Bottom-up control, in this study driven mostly by variations in SST, appeared as more likely from an interannual perspective. Other physical variables which are often correlated to SST, such as irradiance and nutrient concentration, could also be playing an important part in plankton variability on an interannual time scale. However, the absence of datasets over the same geographical and temporal scales as the CPR hinders the confirmation of this hypothesis.

A main finding of this chapter was that, according to the VAR analysis, the most important contributors towards interannual plankton variability were the seasonal cycle and the dependency of each plankton group on its own past. This serial correlation suggests the plankton system of the region is nonlinear, raising the question of whether there could be some chaotic dynamics in the plankton. These issues were addressed in the following chapter.

6.2.3 Challenges and limitations

The analysis of bottom-up versus top-down control presented here is limited by the lack of available data on other potential drivers of these controls. Although sea surface temperature and mixed layer depth should provide a good picture of the environmental factors that impact phytoplankton growth, more explicit information on nutrient concentrations (such as nitrate, nitrite and silicate, McGowan et al., 2017) or turbulence (Barton et al., 2015) would have benefited the investigation of the bottom-up hypothesis. In addition, explicit estimates of zooplankton grazing would have improved the evaluation of top-down control from an interannual perspective. Finally, further information about viral activity and the mortality of phytoplankton attributed to viral infection could bring a new insight to this discussion, although it remains uncertain how important viral lysis is to the interannual variability of the phytoplankton of the region (Riegman and Winter, 2003).

An extension of the current work would be to re-evaluate the CPR dinoflagellate dataset, and consider time-series of autotrophic, mixotrophic and hetetotrophic dinoflagellates separately. It is known that some species of dinoflagellates can feed on small diatoms (e.g., Lalli and Parsons, 1997) and thus dinoflagellates could not only be subject to bottom-up effects from the environment, but also contribute to the grazing of diatoms. The VAR analysis did not find statistically significant links from dinoflagellates to diatom variability, which could be interpreted as an absence of effects from negative interactions such as competition or heterotrophy; nor links from diatoms to dinoflagellate variability, which could be expected in the case of dinoflagellates grazing on diatoms. A further investigation in this direction would nevertheless be interesting.

However, when analysing larger sets of plankton variables VAR models may not be the best strategy. Since the methodology is parametric, a wider set of variables would also introduce more uncertainty in the model (Shmueli, 2010). The approximation used here for the plankton community (diatoms, dinoflagellates, small and large copepods), although arguably incomplete, comprises the best sampled species from the CPR survey, which are known key groups for the maintenance of the ecosystem in the region of study. Keeping the analysis to this small number of variables was a parsimonious choice, while providing a picture of the plankton ecosystem containing the major plankton variables available.

In the present analysis, small copepods (all individuals smaller than 2mm, including nauplii) were used as a representation of the microzooplankton. Although other groups measured by the CPR could have been included, such as radiolarians and foramniferans, these were not explicitly counted before 1993. Considering this difference in time-series length and the parsimony argument, small copepods were the best choice in this case, since they are part of the dominant zooplankton group in the North Sea and Northeast Atlantic (Melle et al., 2014). Possible effects from microzooplankton not accounted for in this thesis could be related to grazing by large copepods. The positive influence from diatoms to large copepod variability identified by the VAR in all regions could be a consequence of indirect abundance of microzooplankton, as the growth of the latter can be associated with that of the primary producers. Their dynamics is tightly coupled with that of microzooplankton. Therefore when diatom abundance is high, so would be that of microzooplankton, which in turn would be consumed by large copepods (Schmoker et al., 2013).

6.3 Is it possible to infer the theoretical dynamics of plankton systems from abundance time-series?

6.3.1 Motivation

Rather than just a mathematical exercise, determining whether an ecological system follows a specific theoretical dynamics is relevant for various practical reasons. For example, it can improve our ability to model an ecosystem, and consequently, in the case of plankton, our understanding of global biogeochemical cycles. It is also relevant from a forecasting perspective, since knowledge about the mathematical structure of the ecosystem will provide insights on the feasibility and accuracy of forecasts.

6.3.2 Chapter summary

In chapter 5, I started by investigating the presence of chaos in the time-series of PCI, diatom, dinoflagellate, small and large copepods from the CPR survey in the Northeast Atlantic and North Sea. The methodology used was a non-parametric nearest-neighbours (kNN) forecasting method, which is suitable for most types of data, including systems with

potential for chaotic behaviour. For testing nonlinearity, a variation of the kNN method known as S-maps was applied. I identified nonlinearity in all of the plankton time-series: the PCI and abundance of diatoms, dinoflagellates, small and large copepods. This confirmed the hypothesis of nonlinearity that was raised in chapter 4, which had been inferred from the autocorrelation component in the VAR model.

The results indicated no evidence for deterministic chaos in the plankton, not even upon repeating the analysis in the series with removed seasonal cycle. On the contrary, the removal of the seasonality indicated a stochastic, rather than deterministic, signature in the plankton series. The seasonal cycle seemed to bring stability and regularity to the series, as the kNN analysis yielded better forecasts for the original series than for the series that had the seasonality removed. This non-parametric analysis, which does not make any *a priori* assumptions about the data, led to the same conclusion as the VAR analysis from the previous chapter: that the plankton seasonal cycle, which is regulated by environmental conditions such as irradiance, mixing and temperature, is responsible for a large part of the interannual variability observed in the plankton, rather than a single specific environmental variable or trophic interaction.

These results imply that it may be possible to forecast plankton abundance, at least in the short term of a few years, mainly thanks to the seasonal features of the plankton system. The final quality and possible length of such forecasts will depend on the stochastic structure behind the seasonality, as regions with evidence of serial correlation tend to yield better forecasts than those with a more random stochastic signal.

The seasonality of the plankton is regulated by a large set of exogenous variables, such as light, nutrient availability, sea surface temperature, mixed layer depth, among others. All of those factors are, on some scale, subject to variations under a changing climate and rising temperatures. The analysis presented in this thesis corroborates that it is unlikely to be possible to single out a main driver of plankton variability. Considering as well the evidence for bottom-up mechanisms regulating the interannual variability of the plankton (chapter 4 and Barton et al., 2015, Richardson, 2008, for instance), I conclude that any environmental changes that may impact the plankton seasonal cycle are likely to propagate on to higher trophic levels, with consequences for the entire marine ecosystem.

6.3.3 Challenges and limitations

Although chaos is known to exist in nature, such as in the weather system and fluid dynamics, evidence of its presence on ecological systems is rare, even in cases where theoretical models suggest that a chaotic dynamics should be present. Models are useful tools to help us understand and approximate the dynamics observed in nature, and deterministic models which present chaos are no exception. In practice, however, randomness is always present in observational data to some extent, be it from measurement errors or from unknown endogenous features of the system being monitored. Plankton models which present chaotic dynamics are valuable and important, just as linear models can also provide useful insights. However, the presence of a specific feature in a model should not be taken as proof that this feature will be present in nature, but instead that the feature can be useful to approximate nature, provided that the model is rigorously validated with observational data.

The methodology used in this chapter stems from an expanding field of nonlinear dynamics analysis. Although the results were derived from current state-of-the-art techniques, it may be that even better nonlinear methods will become available in the future. For instance, the issue of identifying statistically significant causal relationships between the plankton and environmental drivers could be solved with improved, multivariate extensions of the CNN methods. Nevertheless, the diagnosis for nonlinearity in the plankton was clear, and should remain the same even under a more refined analysis.

A main finding of this chapter was that, underneath the regularity of the seasonal cycle, the plankton are subject to different degrees of stochasticity. The current analysis did not aim to identify the source of this, and is not capable of differentiating among endogenous stochasticity (fluctuations that arise from internal regulations), exogenous stochasticity (fluctuations inherited from noise originated in external drivers) and randomness due to measurement errors. The serially correlated structure suggested by some of the kNN maps could indicate an endogenous stochasticity, but it could also be argued that this structure is a consequence of integrated noise inherited from the atmosphere (e.g., from storms or large scale oscillators, Lorenzo and Ohman, 2013). A different type of analysis would be necessary to test the hypotheses, however, to my knowledge, there is no reliable way of doing so at present.

6.4 Implications and future work

In chapter 5, I mentioned a four-stage intertidal community described in Benincà et al. (2015), consisting of a cyclic succession of barnacles, algae and mussels. In this study, the authors found evidence to support the hypothesis of nonlinear and perhaps even chaotic behaviour in the individual species fluctuations. This is a good example of how apparently simple ecosystems can be challenging to understand, model and analyse. In the case of marine plankton, the complexity stems from various sources, from the distinct and sometimes unknown response of different species to environmental variability to the difficulty of maintaining long-term, consistent monitoring programs. In this thesis, I have strived to use suitable approaches to answer the questions raised in chapter 1, using a set of varied statistical techniques. The research conducted addressed its aims, but raised further questions and possible lines of future investigation.

In particular, this thesis has improved the understanding of the plankton dynamics in the North Sea and Northeast Atlantic. After a detailed data-driven analysis, I propose that the plankton of the region follow a nonlinear periodic dynamics with added stochastic noise. This result stems from an analysis conducted in key plankton groups of the region, and, considering that complex trophic interactions exist over global plankton communities, I suggest that a similar dynamics may hold across different plankton systems in other temperate environments with similar biochemical characteristics, since, although different particularities are expected, the seasonal dynamics and controlling environmental factors would be similar to that considered here.

Although this thesis has focused on a specific area of the ocean, the analyses presented have the potential to work in other areas of the global ocean, since the techniques applied do not depend on any region-specific features. All of the procedures were data-driven, and thus reproducible as long as time-series are available. For instance, similar studies could be conducted in the Australian and Southern Ocean CPR databases, which have been collecting monthly samples since 1991, and thus are potentially long enough to meet all of the necessary criteria to avoid spurious results. The Northeast Atlantic and North Sea are highly productive areas of the ocean, and it would be interesting to evaluate whether the plankton dynamics and control in different regions of the ocean are similar to what was identified here. The results from such studies could improve the

development of numerical modelling and forecasting of plankton in a global scale, by providing a strong, observation-based test of such models, which is essential for the management of the marine ecosystem.

The confirmation of nonlinearity in plankton time-series should be taken as a warning for ecosystem monitoring. Nonlinear processes can respond in a rapid or amplified manner to changes in their controlling drivers. In the case of plankton, even with a vast literature available, it is still not known for sure how different communities would respond to changes in environmental factors, such as increasing water temperatures. The complexity of the interactions and nonlinear mechanisms regulating the plankton make it unlikely that a single factor can be identified as responsible for the majority of the changes in the plankton. In this thesis I have demonstrated that the seasonal component, rather than a single exogenous driver, is one of the main controlling factors of plankton variability, and also responsible for increasing the predictability of the plankton dynamics. The plankton seasonal cycle is regulated by a sort of environmental factors, and consequently any alteration on these factors will impact the plankton community. The different drivers suggested by different studies in the literature (e.g., northern hemisphere temperature and the North Atlantic Oscillation (NAO) index in Beaugrand, 2004; sea surface temperature and NAO in Zhai et al., 2013; temperature and the Atlantic Multidecadal Oscillation index in Goberville et al., 2014 and Harris et al., 2015) could be a consequence of this non-uniform response in different locations and conditions. Ultimately, the results presented here imply that any activity with impacts on factors controlling the plankton seasonality are dangerous to the upkeeping of plankton ecosystems.

Lastly, I present some more specific directions for future investigation. In chapter 3, I analysed the relationship between the PCI and abundance of diatoms and dinoflagellates under the hypothesis that said relationship would be linear. This is a reasonable hypothesis in this case, if we consider the CPR sampling and counting methodology. Nevertheless, further investigation on the possibility of nonlinearity could improve our ability to infer details about the plankton community from the PCI or chlorophyll measurements, and vice versa. Another possible extension of this work would be to refine the phytoplankton data by including individual species, instead of the broader diatoms and dinoflagellates groups. The CPR survey has also been counting some individual species of coccolithophores and phaeocystis since 1993. In 6-10 years time these series will be over 30 years long and more likely to be long enough to provide statistically significant results. It would be

interesting to repeat a similar analysis including those groups, to further discuss their contribution towards the greenness index.

In chapters 4 and 5, I discussed the relationships between major phytoplankton and zooplankton groups and key environmental variables, under different premises of stochasticity and determinism. The hypothesis of nonlinearity in the plankton was confirmed with the analysis in chapter 5, but it did not invalidate the results based on the linear models from chapter 4. On the contrary, the importance of the seasonal cycle for the stability in plankton fluctuations was reinforced with the non-parametric analysis of chapter 5. One natural extension of the results presented here, which I consider to be a major topic to be explored, would be to investigate the feasibility of data-driven forecasts of plankton abundance in the North Sea and Northeast Atlantic. The nonlinearity result indicates that the non-parametric methods such as the ones applied in chapter 5 would be more suitable for this purpose, although linear approximations sometimes suffice to perform predictions. Comparing the performance of different forecasting methods would also indicate how effective these forecasts can be in relation to numerical simulations, for example. Data-based forecasts have the potential to provide more accurate predictions than numerical models and contribute to a more effective monitoring of the ecosystem.

Other forecasting methods should also be considered, such as the multivariate version of kNN (e.g., Reick and Page, 2000). The potential for predictability is not likely to be identical across all variables, since each plankton group or species responds differently to its surroundings. Therefore, there is potential for improving forecasts through the inclusion of simultaneous information about other variables in the system. Simple univariate methods should also be considered for the sake of parsimony, providing a paradigm against which the other methodologies can be compared. Another path that can be explored is to consider the different species or functional groups derived from the CPR time-series. It is important, however, to keep in mind that increasing the number of variables when forecasting could be problematic, as it often brings more uncertainty to the final results (Shmueli, 2010). This comparative analysis would bring novel methodology to assess the possible future scenarios for the plankton system in the North Sea and Northeast Atlantic, and could contribute to the planning and management of environmental conservation strategies.

Another key topic that stems from this analysis is to apply the empirical dynamics modelling (EDM) procedure, presented in chapters 2 and 5, to marine ecosystems across

the globe. This methodology is promising to the study of ocean ecosystems (McGowan et al., 2017, Sugihara et al., 2012), and could be explored in the context of forecasting the occurrence of harmful algal blooms, or to investigate causal relationships between competing populations and environmental factors, for example. In a global scale, forecasting of satellite-derived chlorophyll measurements can be tested and validated with EDM techniques, potentially providing insights on global trends of phytoplankton biomass and primary productivity. Investigating multivariate causal relationships among global chlorophyll and environmental variables such as sea surface temperature and climatic oscillators (e.g., NAO, ENSO, PDO) may also be possible with the development of multivariate convergent cross-mapping analysis. This new branch of statistical analysis is growing fast, and research in ocean sciences can benefit greatly, as these techniques are particularly suitable for complex ecosystems such as those of the marine environment.

6.5 Concluding remark

This thesis has provided an in-depth perspective on long-term changes in plankton populations of the North Sea and Northeast Atlantic, exploring and proposing models for the interactions among key plankton and environmental variables from a data-driven context. Through the use of a rigorous statistical framework, I described the relationship between water greenness and the abundance of diatoms and dinoflagellates in the North Sea and Northeast Atlantic; and also examined long-term interactions among SST, MLD, diatoms, dinoflagellates, small and large copepods. I investigated the presence of nonlinearity and chaos in these phytoplankton and zooplankton groups, and concluded that the plankton system of this region is best described as a nonlinear, periodic stochastic process. Considering the complexity of the plankton dynamics and its interactions with external variables, this thesis suggests that determining a single variable as the main driver of long-term changes in the plankton is unlikely to be possible. Considering as well the proposed likelihood of bottom-up effects in the plankton, the results presented indicate that changes to the environmental factors controlling the plankton seasonal cycle can propagate through the marine food web and impact the entire ocean ecosystem. Further analyses exploring the possibilities for data-driven forecasting of plankton can significantly improve our understanding of the plankton dynamics, contributing to the modelling and management of marine ecosystems.

Bibliography

AlvarezFernandez, S., Lindeboom, H., and Meesters, E. Temporal changes in plankton of the North Sea: community shifts and environmental drivers. *Marine Ecology Progress Series*, 462:21–38, 2012.

Ascioti, F. A., Beltrami, E., Carroll, T. O., and Wirick, C. Is there chaos in plankton dynamics? *Journal of Plankton Research*, 15(6):603–617, 1993.

Barton, A. D., Lozier, M. S., and Williams, R. G. Physical controls of variability in North Atlantic phytoplankton communities. *Limnology and Oceanography*, 60(1):181–197, 2015.

Batten, S. D. and Welch, D. W. Changes in oceanic zooplankton populations in the north-east Pacific associated with the possible climatic regime shift of 1998/1999. *Deep Sea Research Part II: Topical Studies in Oceanography*, 51(6-9):863–873, 2004.

Batten, S. D., Walne, A. W., Edwards, M., and Groom, S. B. Phytoplankton biomass from continuous plankton recorder data: an assessment of the phytoplankton colour index. *Journal of Plankton Research*, 25(7):697–702, 2003.

Baumgartner, M. F. and Tarrant, A. M. The physiology and ecology of diapause in marine copepods. *Annual Review of Marine Science*, 9(1):387–411, 2017. PMID: 27814030.

Beaugrand, G. The North Sea regime shift: Evidence, causes, mechanisms and consequences. *Progress in Oceanography*, 60(2-4):245–262, 2004.

Beaugrand, G. Decadal changes in climate and ecosystems in the North Atlantic Ocean and adjacent seas. *Deep Sea Research Part II: Topical Studies in Oceanography*, 56(8-10): 656–673, 2009.

Beaugrand, G., Reid, P. C., Ibañez, F., Lindley, J. A., and Edwards, M. Reorganization of North Atlantic Marine Copepod Biodiversity and Climate. *Science*, 296(5573):1692–1694, 2002.

Beaugrand, G., Harlay, X., and Edwards, M. Detecting plankton shifts in the North Sea: a new abrupt ecosystem shift between 1996 and 2003. *Marine Ecology Progress Series*, 502:85–104, 2014.

Beaulieu, C., Chen, J., and Sarmiento, J. L. Change-point analysis as a tool to detect abrupt climate variations. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 370(1962):1228–1249, 2012.

Beaulieu, C., Chen, J., and Sarmiento, J. L. Change-point analysis as a tool to detect abrupt climate variations. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 370(1962):1228–1249, 2012a.

Beaulieu, C., Sarmiento, J. L., Mikaloff Fletcher, S. E., Chen, J., and Medvigy, D. Identification and characterization of abrupt changes in the land uptake of carbon. *Global Biogeochem. Cycles*, 26(1):1007, 2012b.

Behrenfeld, M. J., O'Malley, R. T., Siegel, D. A., McClain, C. R., Sarmiento, J. L., Feldman, G. C., Milligan, A. J., Falkowski, P. G., Letelier, R. M., and Boss, E. S. Climate-driven trends in contemporary ocean productivity. *Nature*, 444(7120):752–755, 2006.

Behrenfeld, M. J., Halsey, K. H., and Milligan, A. J. Evolved physiological responses of phytoplankton to their integrated growth environment. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 363(1504):2687–2703, 2008.

Behrenfeld, M. J., O'Malley, R. T., Boss, E. S., Westberry, T. K., Graff, J. R., Halsey, K. H., Milligan, A. J., Siegel, D. A., and Brown, M. B. Revaluating ocean warming impacts on global phytoplankton. *Nature Climate Change*, 6(3):323–330, 2016a.

Behrenfeld, M. J., O'Malley, R. T., Boss, E. S., Westberry, T. K., Graff, J. R., Halsey, K. H., Milligan, A. J., Siegel, D. A., and Brown, M. B. Revaluating ocean warming impacts on global phytoplankton. *Nature Climate Change*, 6(3):323–330, 2016b.

Benincà, E., Huisman, J., Heerkloss, R., Jöhnk, K. D., Branco, P., Van Nes, E. H., Scheffer, M., and Ellner, S. P. Chaos in a long-term experiment with a plankton community. *Nature*, 451 (7180):822–825, 2008.

Benincà, E., Ballantine, B., Ellner, S. P., and Huisman, J. Species fluctuations sustained by a cyclic succession at the edge of chaos. *Proceedings of the National Academy of Sciences*, 112(20):6389–6394, 2015.

Berdalet, E., Peters, F., Koumandou, V. L., Roldán, C., Guadayol, O., and Estrada, M. Species-specific physiological response of dinoflagellates to quantified small-scale turbulence1. *Journal of Phycology*, 43(5):965–977, 2007.

Bonnet, D., Richardson, A., Harris, R., Hirst, A., Beaugrand, G., Edwards, M., Ceballos, S., Diekman, R., López-Urrutia, A., Valdes, L., Carlotti, F., Molinero, J. C., Weikert, H., Greve, W., Lucic, D., Albaina, A., Yahia, N. D., Umani, S. F., Miranda, A., Santos, A. d., Cook, K., Robinson, S., and Fernandez de Puelles, M. L. An overview of Calanus helgolandicus ecology in European waters. *Progress in Oceanography*, 65(1):1–53, 2005.

Bouman, H. A., Platt, T., Sathyendranath, S., Li, W. K. W., Stuart, V., FuentesYaco, C., Maass, H., Horne, E. P. W., Ulloa, O., Lutz, V., and Kyewalyanga, M. Temperature as indicator of optical properties and community structure of marine phytoplankton: implications for remote sensing. *Marine Ecology Progress Series*, 258:19–30, 2003.

Box, G. E. P. and Jenkins, G. M. *Time series analysis: forecasting and control*. Holden-Day series in time series analysis. Holden-Day, San Francisco, USA, 1 edition, 1970.

Chassot, E., Mlin, F., Pape, O. L., and Gascuel, D. Bottom-up control regulates fisheries production at the scale of eco-regions in European seas. *Marine Ecology Progress Series*, 343:45–55, 2007.

Chaterjee, S., Hadi, A., and Price, B. *Regression analysis by example*. Wiley Series in Probability and Statistic. Wiley, 2000.

Chen, J. and Gupta, A. K. *Parametric Statistical Change Point Analysis*. BirkhÃd'user Boston, Boston, 2012. ISBN 978-0-8176-4800-8 978-0-8176-4801-5.

Cleveland, R. B., Cleveland, J. E., and I, T. Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6:3–73, 1990.

Cvitanović, P., Artuso, R., Mainieri, R., Tanner, G., and Vattay, G. *Chaos: Classical and Quantum*. Niels Bohr Inst., Copenhagen, 2016.

Davidson, R. and MacKinnon, J. *Econometric theory and methods*. Oxford University Press, 2004. ISBN 978-0-19-512372-2.

de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., and Iudicone, D. Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *Journal of Geophysical Research: Oceans*, 109(C12):C12003, 2004.

DeGroot, M. and Schervish, M. *Probability and Statistics*. Addison-Wesley series in statistics. Addison-Wesley, 2002. ISBN 9780201524888.

Deyle, E. R., Maher, M. C., Hernandez, R. D., Basu, S., and Sugihara, G. Global environmental drivers of influenza. *Proceedings of the National Academy of Sciences*, 113(46):13081–13086, 2016a.

Deyle, E. R., May, R. M., Munch, S. B., and Sugihara, G. Tracking and forecasting ecosystem interactions in real time. *Proc. R. Soc. B*, 283(1822):20152258, 2016b.

Dickey, D. and Fuller, W. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 1979a.

Dickey, D. A. and Fuller, W. A. Distribution of the Estimators for Autoregressive Time Series With a Unit Root. *Journal of the American Statistical Association*, 74(366):427–431, 1979b.

Doney, S. C. Oceanography: Plankton in a warmer world. *Nature*, 444(7120):695–696, 2006.

Edwards, M. and Richardson, A. J. Impact of climate change on marine pelagic phenology and trophic mismatch. *Nature*, 430(7002):881–884, 2004.

Edwards, M., Reid, P., and Planque, B. Long-term and regional variability of phytoplankton biomass in the Northeast Atlantic (1960-1995). *ICES Journal of Marine Science: Journal du Conseil*, 58(1):39–49, 2001.

Edwards, M., Bresnan, E., Cook, K., Heath, M., Helaouët, P., Lynam, C., Raine, R., and Widdicombe, C. Impacts of climate change on plankton. *MCCIP Science Review*, pages 98–112, 2013.

Elliott, G., Rothenberg, T. J., and Stock, J. H. Efficient tests for an autoregressive unit root. *Econometrica*, 64(4):813–836, 1996. ISSN 00129682, 14680262.

Falkowski, P. Ocean Science: The power of plankton. *Nature*, 483(7387):S17–S20, 2012.

Falkowski, P. G., Barber, R. T., and Smetacek, V. Biogeochemical controls and feedbacks on ocean primary production. *Science*, 281(5374):200–206, 1998.

Field, C., Barros, V., Dokken, D., Mach, K., Mastrandrea, M., Bilir, T., Chaterjee, M., Ebi, K., Estrada, Y., Genova, R., Girma, B., Kissel, E., Levy, A., MacCracken, S., Mastrandrea, P., and White, L. *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Summaries, Frequently Asked Questions, and Cross-Chapter Boxes. A Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, 2014.

Field, C. B., Behrenfeld, M. J., Randerson, J. T., and Falkowski, P. Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components. *Science*, 281(5374):237–240, 1998.

Francis, T. B., Scheuerell, M. D., Brodeur, R. D., Levin, P. S., Ruzicka, J. J., Tolimieri, N., and Peterson, W. T. Climate shifts the interaction web of a marine plankton community. *Global Change Biology*, 18(8):2498–2508, 2012.

Fromentin, J. and Planque, B. Calanus and environment in the eastern North Atlantic. II. Influence of the North Atlantic Oscillation on C. finmarchicus and C. helgolandicus. *Marine Ecology Progress Series*, 134:111–118, 1996.

Garrison, T. *Essentials of Oceanography*. Cengage Learning, 2011. ISBN 0-8400-6155-2.

Garrison, T. *Oceanography: An Invitation to Marine Science*. Cengage Learning, 2013. ISBN 1-111-99085-9.

Gleick, J. *Chaos: Making a New Science*. Penguin Books, New York, NY, USA, 1987. ISBN 0-14-009250-1.

Goberville, E., Beaugrand, G., and Edwards, M. Synchronous response of marine plankton ecosystems to climate in the Northeast Atlantic and the North Sea. *Journal of Marine Systems*, 129:189–202, 2014.

Gómez, F. A list of free-living dinoflagellate species in the world's oceans. *Acta Botanica Croatica*, 64(25):129–212, 2005.

Good, S. A., Martin, M. J., and Rayner, N. A. EN4: Quality controlled ocean temperature and salinity profiles and mmonthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, 118(12):6704–6716, 2013.

Graham, L., Graham, J., and Wilcox, L. *Algae*. Benjamin Cummings, 2009. ISBN 9780321559654.

Gsell, A. S., özkundakci, D., Hébert, M.-P., and Adrian, R. Quantifying change in pelagic plankton network stability and topology based on empirical long-term data. *Ecological Indicators*, 65:76–88, 2016.

Guiry, M. D. How Many Species of Algae Are There? *Journal of Phycology*, 48(5):1057–1063, 2012.

Hampton, S. E., Izmest'eva, L. R., Moore, M. V., Katz, S. L., Dennis, B., and Silow, E. A. Sixty years of environmental change in the world's largest freshwater lake - Lake Baikal, Siberia. *Global Change Biology*, 14(8):1947–1958, 2008.

Hampton, S. E., Holmes, E. E., Scheef, L. P., Scheuerell, M. D., Katz, S. L., Pendleton, D. E., and Ward, E. J. Quantifying effects of abiotic and biotic drivers on community dynamics with multivariate autoregressive (MAR) models. *Ecology*, 94(12):2663–2669, 2013.

Harris, V., Edwards, M., and Olhede, S. C. Multidecadal Atlantic climate variability and its impact on marine pelagic communities. *Journal of Marine Systems*, 2014.

Harris, V., Olhede, S. C., and Edwards, M. Multidecadal spatial reorganisation of plankton communities in the North East Atlantic. *Journal of Marine Systems*, 142:16–24, 2015.

Hays, G. and Lindley, J. Estimating chlorophyll a abundance from the 'phytoplankton colour' recorded by the continuous plankton recorder survey: validation with simultaneous fluorometry. *Journal of Plankton Research*, 16(1):23–34, 1994.

Helaouët, P., Beaugrand, G., and Edwards, M. Understanding Long-Term Changes in Species Abundance Using a Niche-Based Approach. *PLoS ONE*, 8(11):e79186, 2013.

Henson, S., Cole, H., Beaulieu, C., and Yool, A. The impact of global warming on seasonality of ocean primary production. *Biogeosciences*, 10(6):4357–4369, 2013.

Henson, S. A., Sarmiento, J. L., Dunne, J. P., Bopp, L., Lima, I., Doney, S. C., John, J., and Beaulieu, C. Detection of anthropogenic climate change in satellite records of ocean chlorophyll and productivity. *Biogeosciences*, 7(2):621–640, 2010.

Hilligsöe, K. M., Richardson, K., Bendtsen, J., Sörensen, L.-L., Nielsen, T. G., and Lyngsgaard, M. M. Linking phytoplankton community size composition with temperature, plankton food

web structure and sea-air co2 flux. *Deep Sea Research Part I: Oceanographic Research Papers*, 58(8):826 – 838, 2011.

Hinder, S. L., Hays, G. C., Edwards, M., Roberts, E. C., Walne, A. W., and Gravenor, M. B. Changes in marine dinoflagellate and diatom abundance under climate change. *Nature Climate Change*, 2(4):271–275, 2012a.

Hinder, S. L., Manning, J. E., Gravenor, M. B., Edwards, M., Walne, A. W., Burkill, P. H., and Hays, G. C. Long-term changes in abundance and distribution of microzooplankton in the NE Atlantic and North Sea. *Journal of Plankton Research*, 34(1):83–91, 2012b.

Hoppenrath, M. and Saldarriaga, J. F. Dinoflagellates. Tree of Life Web Project - `http://tolweb.org/Dinoflagellates/2445/2012.12.15`, 2012. (version 15-December-2012).

Huisman, J. and Weissing, F. J. Biodiversity of plankton by species oscillations and chaos. *Nature*, 402(6760):407–410, 1999.

Huisman, J., Pham Thi, N. N., Karl, D. M., and Sommeijer, B. Reduced mixing generates oscillations and chaos in the oceanic deep chlorophyll maximum. *Nature*, 439(7074):322–325, 2006.

Humes, A. G. How many copepods? *Hydrobiologia*, 292(1):1–7, 1994.

Hurrell, J. W. Decadal trends in the north atlantic oscillation: Regional temperatures and precipitation. *Science*, 269(5224):676–679, 1995.

Hyndman, R. and Athanasopoulos, G. *Forecasting: principles and practice*. https://www.otexts.org/fpp, 2012.

Irwin, A. J. and Finkel, Z. V. Mining a sea of data: Deducing the environmental controls of ocean chlorophyll. *PLOS ONE*, 3(11):1–6, 2008.

Ives, A. R., Dennis, B., Cottingham, K. L., and Carpenter, S. R. Estimating Community Stability and Ecological Interactions from Time-Series Data. *Ecological Monographs*, 73(2):301–330, 2003.

Ives, A. R., Carpenter, S. R., and Dennis, B. Community Interaction Webs and Zooplankton Responses to Planktivory Manipulations. *Ecology*, 80(4):1405–1421, 1999.

Johnson, W. and Allen, D. *Zooplankton of the Atlantic and Gulf coasts: a guide to their identification and ecology*. The Johns Hopkins University Press, 2nd edition, 2012. ISBN 1-4214-0746-9.

Jordan, R. W. Coccolithophores. In Schaechter, M., editor, *Encyclopedia of Microbiology (Third Edition)*, pages 593–605. Academic Press, Oxford, 2009. ISBN 978-0-12-373944-5.

Kim, D. and Perron, P. Unit roots tests allowing for a break in the trend function at an unknown time under both the null and alternative hypotheses. *Journal of Econometrics*, 2009.

Kratina, P., Mac Nally, R., Kimmerer, W. J., Thomson, J. R., and Winder, M. Human-induced biotic invasions and changes in plankton interaction networks. *Journal of Applied Ecology*, 51(4):1066–1074, 2014.

Kudela, R. M., Palacios, S. L., Austerberry, D. C., Accorsi, E. K., Guild, L. S., and Torres-Perez, J. Application of hyperspectral remote sensing to cyanobacterial blooms in inland waters. *Remote Sensing of Environment*, 167(Supplement C):196 – 205, 2015. Special Issue on the Hyperspectral Infrared Imager (HysplRI).

Kwistkowski, D., Phillips, P., Schmidt, P., and Shin, Y. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 1992.

Lalli, C. M. and Parsons, T. R. Biological oceanography: An introduction. In *Biological Oceanography: An Introduction*. Butterworth-Heinemann, Oxford, second edition edition, 1997. ISBN 978-0-7506-3384-0.

Larsen, K., Gonzalez-Pola, C., Fratantoni, P., Beszczynska-Möller, A., and Hughes, S. Ices report on ocean climate. ICES Cooperative Research Report - `www.ices.dk`, 2015. No. 331, 79 pp.

Leterme, S. C. and Pingree, R. D. Structure of phytoplankton (Continuous Plankton Recorder and SeaWiFS) and impact of climate in the Northwest Atlantic Shelves. *Ocean Sci.*, 3(1): 105–116, 2007.

Leterme, S. C., Seuront, L., and Edwards, M. Differential contribution of diatoms and dinoflagellates to phytoplankton biomass in the NE Atlantic Ocean and the North Sea. *Marine Ecology Progress Series*, 312:57–65, 2006.

Ljung, G. M. and Box, G. E. P. On a Measure of Lack of Fit in Time Series Models. *Biometrika*, 65(2):297–303, 1978.

Longhurst, A. *Ecological georgaphy of the sea*. Elsevier Academic Press, 2nd edition, 2006. ISBN 0-12-455521-7.

Lorenzo, E. D. and Ohman, M. D. A double-integration hypothesis to explain ocean ecosystem response to climate forcing. *Proceedings of the National Academy of Sciences*, 110(7): 2496–2499, 2013.

Lui, G. C. S., Li, W. K., Leung, K. M. Y., Lee, J. H. W., and Jayawardena, A. W. Modelling algal blooms using vector autoregressive model with exogenous variables and long memory filter. *Ecological Modelling*, 200(1-2):130–138, 2007.

Lütkepohl, H. *New Introduction to Multiple Time Series Analysis*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. ISBN 978-3-540-40172-8 978-3-540-27752-1.

Lynam, C. P., Llope, M., Möllmann, C., Helaouët, P., Bayliss-Brown, G. A., and Stenseth, N. C. Interaction between top-down and bottom-up control in marine food webs. *Proceedings of the National Academy of Sciences*, 114(8):1952–1957, 2017.

Mann, D. G. Diatoms. Tree of Life Web Project - `http://tolweb.org/Diatoms/21810/2010.02.07`, 2010. (version 07-February-2010).

Maranon, E., Holligan, P., Barciela, R., Gonzalez, N., Mourino, B., Pazo, M., and Varela, M. Patterns of phytoplankton size structure and productivity in contrasting open-ocean environments. *Marine Ecology Progress Series*, 216:43–56, 2001.

Martinez, E., Raitsos, D. E., and Antoine, D. Warmer, deeper, and greener mixed layers in the North Atlantic subpolar gyre over the last 50Âăyears. *Global Change Biology*, 22(2): 604–612, 2016.

McDougall, T. J. and Barker, P. M. *Getting started with TEOS-10 and the Gibbs Seawater (GSW) Oceanographic Toolbox*. Trevor J McDougall, Battery Point, Tas., 2011. ISBN 978-0-646-55621-5. OCLC: 724024071.

McGinty, N., Power, A. M., and Johnson, M. P. Variation among northeast Atlantic regions in the responses of zooplankton to climate change: Not all areas follow the same path. *Journal of Experimental Marine Biology and Ecology*, 400(1-2):120–131, 2011.

McGowan, J. A., Deyle, E. R., Ye, H., Carter, M. L., Perretti, C. T., Seger, K. D., de Verneil, A., and Sugihara, G. Predicting coastal algal blooms in southern California. *Ecology*, 98 (5):1419–1433, 2017.

McQuatters-Gollop, A., Raitsos, D. E., Edwards, M., Pradhan, Y., Mee, L. D., Lavender, S. J., and Attrill, M. J. A long-term chlorophyll dataset reveals regime shift in north sea phytoplankton biomass unconnected to nutrient levels. *Limnology and Oceanography*, 52 (2):635–648, 2007.

McQuatters-Gollop, A., Reid, P. C., Edwards, M., Burkill, P. H., Castellani, C., Batten, S., Gieskes, W., Beare, D., Bidigare, R. R., Head, E., Johnson, R., Kahru, M., Koslow, J. A., and Pena, A. Is there a decline in marine phytoplankton? *Nature*, 472(7342):E6–E7, 2011.

Melle, W., Runge, J., Head, E., Plourde, S., Castellani, C., Licandro, P., Pierson, J., Jonasdottir, S., Johnson, C., Broms, C., Debes, H., Falkenhaug, T., Gaard, E., Gislason, A., Heath, M., Niehoff, B., Nielsen, T. G., Pepin, P., Stenevik, E. K., and Chust, G. The north atlantic ocean as habitat for calanus finmarchicus: Environmental factors and life history traits. *Progress in Oceanography*, 129:244 – 284, 2014. North Atlantic Ecosystems, the role of climate and anthropogenic forcing on their structure and function.

Morán, X. A., López-Urrutia, Á., Calvo-Díaz, A., and Li, W. Increasing importance of small phytoplankton in a warmer ocean. *Global Change Biology*, 16(3):1137–1144, 2010.

Moroz, I. M., Cropp, R., and Norbury, J. Chaos in plankton models: Foraging strategy and seasonal forcing. *Ecological Modelling*, 332:103–111, 2016.

Nair, A., Sathyendranath, S., Platt, T., Morales, J., Stuart, V., Forget, M.-H., Devred, E., and Bouman, H. Remote sensing of phytoplankton functional types. *Remote Sensing of Environment*, 112(8):3366 – 3375, 2008. Earth Observations for Marine and Coastal Biodiversity and Ecosystems Special Issue.

OSPAR. Quality status report. Convention for the Protection of the Marine Environment of the North-East Atlantic - `www.ospar.org`, 2010. 176 pp.

Owens, N. J. P., Hosie, G. W., Batten, S. D., Edwards, M., Johns, D. G., and Beaugrand, G. All plankton sampling systems underestimate abundance: Response to "Continuous plankton recorder underestimates zooplankton abundance" by J.W. Dippner and M. Krause. *Journal of Marine Systems*, 128:240–242, 2013.

Peter, K. H. and Sommer, U. Interactive effect of warming, nitrogen and phosphorus limitation on phytoplankton cell size. *Ecology and Evolution*, 5(5):1011–1024, 2015.

Pfaff, B. Var, svar and svec models: Implementation within r package vars. *Journal of Statistical Software*, 27(4):1–32, 2008.

Racault, M.-F., Le Quéré, C., Buitenhuis, E., Sathyendranath, S., and Platt, T. Phytoplankton phenology in the global ocean. *Ecological Indicators*, 14(1):152–163, 2012.

Rahmstorf, S., Box, J. E., Feulner, G., Mann, M. E., Robinson, A., Rutherford, S., and Schaffernicht, E. J. Exceptional twentieth-century slowdown in Atlantic Ocean overturning circulation. *Nature Climate Change*, 5(5):475–480, 2015.

Raitsos, D., Pradhan, Y., Lavender, S., Hoteit, I., McQuatters-Gollop, A., Reid, P. C., and Richardson, A. J. From silk to satellite: half a century of ocean colour anomalies in the Northeast Atlantic. *Global Change Biology*, 2013a.

Raitsos, D. E., Reid, P. C., Lavender, S. J., Edwards, M., and Richardson, A. J. Extending the SeaWiFS chlorophyll data set back 50 years in the northeast Atlantic. *Geophysical Research Letters*, 32(6):L06603, 2005.

Raitsos, D. E., Walne, A., Lavender, S. J., Licandro, P., Reid, P. C., and Edwards, M. A 60-year ocean colour data set from the continuous plankton recorder. *Journal of Plankton Research*, 2013b.

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., and Kaplan, A. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, 108(D14):4407, 2003.

Reick, C. H. and Page, B. Time series prediction by multivariate next neighbor methods with application to zooplankton forecasts. *Mathematics and Computers in Simulation*, 52(3-4): 289–310, 2000.

Reid, P. C., Lancelot, C., Gieskes, W. W. C., Hagmeier, E., and Weichart, G. Phytoplankton of the North Sea and its dynamics: A review. *Netherlands Journal of Sea Research*, 26 (2-4):295–331, 1990.

Reid, P. C., de Fatima Borges, M., and Svendsen, E. A regime shift in the north sea circa 1988 linked to changes in the north sea horse mackerel fishery. *Fisheries Research*, 50(1):163 – 171, 2001.

Richardson, A. J., Walne, A. W., John, A. W. G., Jonas, T. D., Lindley, J. A., Sims, D. W., Stevens, D., and Witt, M. Using continuous plankton recorder data. *Progress in Oceanography*, 68(1):27–74, 2006.

Richardson, A. J. In hot water: zooplankton and climate change. *ICES Journal of Marine Science: Journal du Conseil*, 65(3):279–295, 2008.

Richardson, A. J. and Schoeman, D. S. Climate impact on plankton ecosystems in the northeast atlantic. *Science*, 305(5690):1609–1612, 2004.

Riegman, R. and Winter, C. Lysis of plankton in the non-stratified southern north sea during summer and autumn 2000. *Acta Oecologica*, 24, 05 2003.

Rosenstein, M. T., Collins, J. J., and De Luca, C. J. A practical method for calculating largest Lyapunov exponents from small data sets. *Physica D: Nonlinear Phenomena*, 65(1):117–134, 1993.

Rouyer, T., Fromentin, J.-M., Hidalgo, M., and Stenseth, N. C. Combined effects of exploitation and temperature on fish stocks in the Northeast Atlantic. *ICES Journal of Marine Science: Journal du Conseil*, 71(7):1554–1562, 2014.

Rudnick, D. and Davis, R. Comment on "Regime shifts and red noise in the North Pacific". *Deep-Sea Research Part I-Oceanographic Research Papers*, 53:589–590, 2006.

Rudnick, D. L. and Davis, R. E. Red noise and regime shifts. *Deep Sea Research Part I: Oceanographic Research Papers*, 50(6):691–699, 2003.

Ryan, J. P., Davis, C. O., Tufillaro, N. B., Kudela, R. M., and Gao, B.-C. Application of the hyperspectral imager for the coastal ocean to phytoplankton ecology studies in monterey bay, ca, usa. *Remote Sensing*, 6:1007–1025, 2014.

Scheef, L. P., Pendleton, D. E., Hampton, S. E., Katz, S. L., Holmes, E. E., Scheuerell, M. D., and Johns, D. G. Assessing marine plankton community structure from long-term monitoring data with multivariate autoregressive (MAR) models: a comparison of fixed station versus spatially distributed sampling data. *Limnology and Oceanography: Methods*, 10(1):54–64, 2012.

Scheffer, M., Rinaldi, S., Huisman, J., and Weissing, F. J. Why plankton communities have no equilibrium: solutions to the paradox. *Hydrobiologia*, 491(1-3):9–18, 2003.

Schmoker, C., Hernández-León, S., and Calbet, A. Microzooplankton grazing in the oceans: impacts, data variability, knowledge gaps and future directions. *Journal of Plankton Research*, 35(4):691–706, 2013.

Schwarz, G. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464, 1978.

Shmueli, G. To explain or to predict? *Statistical Science*, 25(3):289–310, 2010.

Sommer, U., Stibor, H., Katechakis, A., Sommer, F., and Hansen, T. Pelagic food web configurations at different levels of nutrient richness and their implications for the ratio fish production:primary production. *Hydrobiologia*, 484(1):11–20, 2002.

Spencer, M., Birchenough, S. N. R., Mieszkowska, N., Robinson, L. A., Simpson, S. D., Burrows, M. T., Capasso, E., Cleall-Harding, P., Crummy, J., Duck, C., Eloire, D., Frost, M., Hall, A. J., Hawkins, S. J., Johns, D. G., Sims, D. W., Smyth, T. J., and Frid, C. L. J. Temporal change in UK marine communities: trends or regime shifts? *Marine Ecology*, 32: 10–24, 2011.

Stenseth, N. C., Llope, M., Anadón, R., Ciannelli, L., Chan, K.-S., Hjermann, D., Bagöien, E., and Ottersen, G. Seasonal plankton dynamics along a cross-shelf gradient. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1603):2831–2838, 2006.

Sugihara, G., Allan, W., Sobel, D., and Allan, K. D. Nonlinear control of heart rate variability in human infants. *Proceedings of the National Academy of Sciences*, 93(6):2608–2613, 1996.

Sugihara, G. Nonlinear Forecasting for the Classification of Natural Time Series. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 348(1688):477–495, 1994.

Sugihara, G. and May, R. M. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, 344(6268):734–741, 1990.

Sugihara, G., Grenfell, B., May, R. M., Chesson, P., Platt, H. M., and Williamson, M. Distinguishing Error from Chaos in Ecological Time Series [and Discussion]. *Philosophical*

Transactions of the Royal Society of London B: Biological Sciences, 330(1257):235–251, 1990.

Sugihara, G., May, R., Ye, H., Hsieh, C.-h., Deyle, E., Fogarty, M., and Munch, S. Detecting Causality in Complex Ecosystems. *Science*, 338(6106):496–500, 2012.

Takens, F. Detecting strange attractors in turbulence. *Lecture Notes in Mathematics, Berlin Springer Verlag*, 898:366, 1981.

Taylor, A. H., Allen, J. I., and Clark, P. A. Extraction of a weak climatic signal by an ecosystem. *Nature*, 416(6881):629–632, 2002.

Thomas, M. K., Kremer, C. T., Klausmeier, C. A., and Litchman, E. A Global Pattern of Thermal Adaptation in Marine Phytoplankton. *Science*, 338(6110):1085–1088, 2012.

Trapletti, A. and Hornik, K. *tseries: Time Series Analysis and Computational Finance*, 2017. R package version 0.10-42.

Vanormelingen, P., Verleyen, E., and Vyverman, W. The diversity and distribution of diatoms: from cosmopolitanism to narrow endemism. *Biodiversity and Conservation*, 17(2):393–405, 2007.

Villarino, E., Chust, G., Licandro, P., Butenschön, M., Ibaibarriaga, L., Larrañaga, A., and Irigoien, X. Modelling the future biogeography of North Atlantic zooplankton communities in response to climate change. *Marine Ecology Progress Series*, 2015.

Xi, H., Hieronymi, M., Röttgers, R., Krasemann, H., and Qiu, Z. Hyperspectral differentiation of phytoplankton taxonomic groups: A comparison between using remote sensing reflectance and absorption spectra. *Remote Sensing*, 7:14781–14805, 2015.

Ye, H., Beamish, R. J., Glaser, S. M., Grant, S. C. H., Hsieh, C., Richards, L. J., Schnute, J. T., and Sugihara, G. Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. *Proceedings of the National Academy of Sciences*, 112(13):E1569–E1576, 2015.

Zhai, L., Platt, T., Tang, C., Sathyendranath, S., and Walne, A. The response of phytoplankton to climate variability associated with the North Atlantic Oscillation. *Deep Sea Research Part II: Topical Studies in Oceanography*, 93:159–168, 2013.

Zimmer, C. Life After Chaos. *Science*, 284(5411):83–86, 1999.

Zounemat-Kermani, M. Investigating chaos and nonlinear forecasting in short term and mid-term river discharge. *Water Resources Management*, 30(5):1851–1865, Mar 2016. ISSN 1573-1650.

Appendix A

VAR model for first-differenced data

Let $X(t)$ and $Y(t)$ be two periodic time-series, which we want to describe in a VAR system:

$$\begin{cases} X(t) = a_{11}\, X(t-1) + a_{12}\, Y(t-1) + a_X\, Z(t) + SSN_X(t) + \varepsilon_X(t) \\ Y(t) = a_{21}\, X(t-1) + a_{22}\, Y(t-1) + a_Y\, Z(t) + SSN_Y(t) + \varepsilon_Y(t) \end{cases} \tag{A.1}$$

where $Z(t)$ is an external covariate, $SSN_*(t)$ is a seasonal constant for each month of the year (that is, $SSN_*(t+12) = SSN_*(t)$ for all $k = 1, 2, \dots$), and $\varepsilon_*(t)$ is a white noise term.

Say that it is necessary to take first-differences in order to make the series stationary. The first-differenced series are:

$$\begin{cases} \Delta X(t) = X(t) - X(t-1) \\ \Delta Y(t) = Y(t) - Y(t-1) \end{cases} \tag{A.2}$$

Consider the equation for $X(t)$ in the system (A.1), and subtract $X(t-1)$ from both sides of the equation:

$$X(t) - X(t-1) = a_{11}\, X(t-1) + a_{12}\, Y(t-1) + a_X\, Z(t) + SSN_X(t) + \varepsilon_X(t) - X(t-1)$$

$$\begin{aligned}\Delta X(t) =& a_{11}\, X(t-1) + a_{12}\, Y(t-1) + a_X\, Z(t) + SSN_X(t) + \varepsilon_X(t) \\ & - a_{11}\, X(t-2) - a_{12}\, Y(t-2) - a_X\, Z(t-1) - SSN_X(t-1) - \varepsilon_X(t-1)\end{aligned}$$

$$\Delta X(t) = a_{11}\, \Delta X(t-1) + a_{12}\, \Delta Y(t-1) + a_X\, \Delta Z(t) + \Delta SSN_X(t) + \Delta\varepsilon_X(t) \quad \text{(A.3)}$$

Applying the same reasoning for $Y(t)$, we have:

$$\Delta Y(t) = a_{21}\, \Delta X(t-1) + a_{22}\, \Delta Y(t-1) + a_Y\, \Delta Z(t) + \Delta SSN_Y(t) + \Delta\varepsilon_Y(t) \quad \text{(A.4)}$$

But $\Delta SSN_*(t)$ is also a seasonal constant, so we can write $\xi_*(t) = \Delta SSN_*(t)$ and it is true that $\xi_*(t+12) = \xi_*(t)$. Also, if $\varepsilon(t)$ is a white noise process, then $\epsilon(t) = \Delta\varepsilon(t)$ is also a white noise process. Therefore, if we fit a VAR model to the first-differenced series:

$$\begin{cases} \Delta X(t) = & a_{11}\, \Delta X(t-1) + a_{12}\, \Delta Y(t-1) + a_X\, \Delta Z(t) + \xi_X(t) + \epsilon_X(t) \\ \Delta Y(t) = & a_{21}\, \Delta X(t-1) + a_{22}\, \Delta Y(t-1) + a_Y\, \Delta Z(t) + \xi_Y(t) + \epsilon_Y(t) \end{cases} \quad \text{(A.5)}$$

We can see that the set of parameters a_*, corresponding to the interactions between variables $X(t)$ and $Y(t)$, are equivalent to the ones for the interactions between $\Delta X(t)$ and $\Delta Y(t)$. This implies that, even if the final model is different for original versus first-differenced data, the interpretation of the parameters corresponding to the interactions among variables and influence of external covariates is equivalent in both cases.

Appendix B

Estimated coefficients from the VAR model

TABLE B.1: VAR coefficients for the model with diatoms and dinoflagellates in the North Sea (NS). Fields marked with * were not statistically significant.

	diat	dino	scpp	lcpp
diat lag -1	-0.72831	*	0.137356	0.115583
dino lag -1	*	-0.33566	*	*
scpp lag -1	*	0.083525	-0.42421	*
lcpp lag -1	*	-0.09515	-0.08056	-0.66562
diat lag -2	-0.49387	*	*	*
dino lag -2	*	-0.17141	*	*
scpp lag -2	*	0.145881	-0.29462	*
lcpp lag -2	*	-0.1039	*	-0.40677
diat lag -3	-0.24018	*	*	*
dino lag -3	*	-0.20357	*	*
scpp lag -3	*	0.14546	-0.16777	*
lcpp lag -3	*	*	0.074985	-0.17906
diat lag -4	*	*	*	*
dino lag -4	*	*	*	*
scpp lag -4	*	*	*	*
lcpp lag -4	*	*	*	*
ssnl 1	0.441998	*	0.363839	*
ssnl 2	1.459702	*	0.548168	*
ssnl 3	2.568098	*	1.133316	0.822119
ssnl 4	0.980963	*	1.630258	0.630193
ssnl 5	*	*	1.925087	*
ssnl 6	*	0.491283	0.998342	-0.3007
ssnl 7	*	0.362342	0.336644	-0.46295
ssnl 8	0.374613	*	*	*
ssnl 9	0.355523	*	-0.48957	*
ssnl 10	*	*	-0.4679	*
ssnl 11	*	*	*	*
SST lag 0	-0.16035	0.422271	*	*
SST lag -1	*	*	*	0.116292
SST lag -2	*	*	*	*
SST lag -3	*	*	*	*
SST lag -4	*	-0.09279	*	*
MLD lag 0	*	*	*	*
MLD lag -1	*	*	*	*
MLD lag -2	*	*	*	*
MLD lag -3	*	*	*	*
MLD lag -4	*	*	*	*

TABLE B.2: VAR coefficients estimated for the model with diatoms and dinoflagellates in the northern Northeast Altantic (NNEA). Fields marked with * were not statistically significant.

	diat	dino	scpp	lcpp
diat lag -1	-0.56776	0.144039	*	0.081451
dino lag -1	*	-0.50237	*	-0.0874
scpp lag -1	*	*	-0.53669	0.167083
lcpp lag -1	*	*	-0.12387	-0.76673
diat lag -2	-0.287	*	0.084457	0.07364
dino lag -2	*	-0.25525	0.128562	*
scpp lag -2	*	*	-0.41183	0.077501
lcpp lag -2	*	*	-0.0926	-0.43144
diat lag -3	-0.16601	*	0.218273	*
dino lag -3	*	-0.15755	*	*
scpp lag -3	*	*	-0.20952	*
lcpp lag -3	*	0.079314	*	-0.25518
diat lag -4	*	*	*	*
dino lag -4	*	*	*	*
scpp lag -4	*	*	*	*
lcpp lag -4	*	*	*	*
ssnl 1	*	*	*	*
ssnl 2	*	*	*	*
ssnl 3	0.56346	*	*	0.399407
ssnl 4	1.237303	*	0.72941	0.775077
ssnl 5	*	0.596534	0.513894	0.964167
ssnl 6	*	1.146426	*	*
ssnl 7	*	*	*	*
ssnl 8	*	*	*	*
ssnl 9	0.329836	*	*	*
ssnl 10	*	-0.47434	*	*
ssnl 11	*	*	*	*
SST lag 0	*	*	0.252651	*
SST lag -1	*	0.242703	*	*
SST lag -2	*	-0.15583	*	*
SST lag -3	*	*	*	*
SST lag -4	-0.11408	*	-0.16834	-0.17352
MLD lag 0	0.127054	0.112346	*	*
MLD lag -1	*	*	*	*
MLD lag -2	*	*	*	*
MLD lag -3	*	*	*	*
MLD lag -4	*	*	*	*

TABLE B.3: VAR coefficients estimated for the model with diatoms and dinoflagellates in the central Northeast Altantic (CNEA). Fields marked with * were not statistically significant.

	diat	dino	scpp	lcpp
diat lag -1	-0.53473	*	*	0.262318
dino lag -1	*	-0.46316	*	*
scpp lag -1	*	0.148959	-0.30628	*
lcpp lag -1	*	*	*	-0.53193
diat lag -2	-0.27474	*	*	0.199621
dino lag -2	*	-0.23015	*	*
scpp lag -2	*	*	-0.27216	*
lcpp lag -2	*	*	0.160728	-0.18064
diat lag -3	*	*	*	*
dino lag -3	*	*	*	*
scpp lag -3	*	*	*	*
lcpp lag -3	*	*	*	*
diat lag -4	*	*	*	*
dino lag -4	*	*	*	*
scpp lag -4	*	*	*	*
lcpp lag -4	*	*	*	*
ssnl 1	0.313506	*	*	*
ssnl 2	0.564276	*	*	*
ssnl 3	1.606417	*	*	0.629517
ssnl 4	2.067233	0.393332	0.440312	0.402601
ssnl 5	*	*	*	*
ssnl 6	0.465561	*	*	*
ssnl 7	0.580636	*	*	*
ssnl 8	0.700329	*	*	*
ssnl 9	0.405521	-0.63185	-0.35678	*
ssnl 10	*	-0.76453	*	*
ssnl 11	*	-0.48294	*	*
SST lag 0	*	0.331698	0.412931	*
SST lag -1	-0.2543	*	*	*
SST lag -2	*	*	*	*
SST lag -3	*	0.115243	-0.11912	*
SST lag -4	*	*	*	-0.12645
MLD lag 0	*	*	-0.12623	*
MLD lag -1	*	*	*	*
MLD lag -2	*	*	*	*
MLD lag -3	*	*	*	*
MLD lag -4	*	*	*	*

TABLE B.4: VAR coefficients estimated for the model with diatoms and dinoflagellates in the southern Northeast Altantic (SNEA). Fields marked with * were not statistically significant.

	diat	dino	scpp	lcpp
diat lag -1	-0.74142	*	*	0.20338
dino lag -1	*	-0.67787	*	*
scpp lag -1	*	0.108346	-0.57129	*
lcpp lag -1	-0.15224	-0.19914	*	-0.5448
diat lag -2	-0.48335	*	*	0.171921
dino lag -2	*	-0.44868	*	*
scpp lag -2	*	*	-0.48606	*
lcpp lag -2	*	*	*	-0.40712
diat lag -3	-0.35781	*	*	*
dino lag -3	*	-0.33433	0.075718	*
scpp lag -3	*	*	-0.31005	*
lcpp lag -3	*	*	*	-0.28776
diat lag -4	-0.20752	*	*	*
dino lag -4	*	-0.11249	*	*
scpp lag -4	*	*	-0.17086	*
lcpp lag -4	*	*	*	-0.1966
ssnl 1	*	*	*	*
ssnl 2	0.490986	*	*	*
ssnl 3	0.836703	0.415623	0.766455	0.881353
ssnl 4	0.972241	0.99039	0.731616	*
ssnl 5	*	0.805685	*	-0.29595
ssnl 6	*	0.278795	-0.65142	-0.40426
ssnl 7	*	*	-0.60449	*
ssnl 8	*	*	*	*
ssnl 9	*	*	*	*
ssnl 10	*	*	*	*
ssnl 11	*	*	*	*
SST lag 0	*	*	0.246567	*
SST lag -1	*	*	*	*
SST lag -2	*	*	*	*
SST lag -3	-0.063	*	*	*
SST lag -4	*	-0.09141	*	-0.12295
MLD lag 0	*	*	*	*
MLD lag -1	*	*	*	*
MLD lag -2	*	*	*	*
MLD lag -3	*	*	*	*
MLD lag -4	*	*	*	*

TABLE B.5: VAR coefficients estimated for the model with the PCI in the North Sea (NS). Fields marked with * were not statistically significant.

	pci	scpp	lcpp
pci lag -1	-0.58724	*	*
scpp lag -1	*	-0.42442	*
lcpp lag -1	-0.12667	-0.09269	-0.67053
pci lag -2	-0.40102	*	*
scpp lag -2	*	-0.29475	*
lcpp lag -2	-0.1417	*	-0.41737
pci lag -3	-0.21386	*	*
scpp lag -3	*	-0.15608	*
lcpp lag -3	-0.08686	0.072549	-0.20358
pci lag -4	*	*	*
scpp lag -4	*	*	*
lcpp -lag 4	*	*	*
pci lag -5	*	*	*
scpp lag -5	*	*	*
lcpp lag -5	*	*	*
ssnl 1	0.377451	0.289305	*
ssnl 2	1.233835	0.441636	*
ssnl 3	2.296494	1.010417	0.993634
ssnl 4	1.56041	1.419067	0.904671
ssnl 5	1.046847	1.247826	*
ssnl 6	1.085636	0.434822	*
ssnl 7	1.095828	*	-0.22892
ssnl 8	0.819151	*	*
ssnl 9	0.60439	-0.39893	*
ssnl 10	*	-0.36491	*
ssnl 11	*	*	*
SST lag 0	*	0.216252	*
SST lag -1	*	*	*
SST lag -2	*	*	*
SST lag -3	*	*	*
SST lag -4	*	*	*
MLD lag 0	*	*	*
MLD lag -1	*	*	*
MLD lag -2	*	*	*
MLD lag -3	*	*	*
MLD lag -4	*	*	*

TABLE B.6: VAR coefficients estimated for the model with the PCI in the northern Northeast Altantic (NNEA). Fields marked with * were not statistically significant.

	pci	scpp	lcpp
pci lag -1	-0.64308	*	*
scpp lag -1	*	-0.62827	0.12226
lcpp lag -1	*	*	-0.79976
pci lag -2	-0.36817	*	*
scpp lag -2	*	-0.49843	*
lcpp lag -2	*	*	-0.52444
pci lag -3	-0.29863	0.073301	*
scpp lag -3	*	-0.3108	*
lcpp lag -3	*	*	-0.40186
pci lag -4	-0.18108	*	*
scpp lag -4	*	-0.18549	*
lcpp -lag 4	*	*	-0.17828
pci lag -5	*	*	*
scpp lag -5	*	*	*
lcpp lag -5	*	*	*
ssnl 1	*	*	*
ssnl 2	0.431103	*	*
ssnl 3	1.030184	*	0.344077
ssnl 4	1.683884	0.565282	0.649238
ssnl 5	1.439262	0.3234	0.879226
ssnl 6	1.490918	*	*
ssnl 7	1.108961	*	*
ssnl 8	1.214757	0.439008	*
ssnl 9	0.49264	*	*
ssnl 10	*	*	*
ssnl 11	*	*	*
SST lag 0	*	0.391885	0.160724
SST lag -1	*	*	*
SST lag -2	*	*	*
SST lag -3	-0.22107	*	*
SST lag -4	*	-0.09768	-0.12676
MLD lag 0	*	*	*
MLD lag -1	*	0.088045	*
MLD lag -2	*	*	*
MLD lag -3	*	*	*
MLD lag -4	*	*	*

TABLE B.7: VAR coefficients estimated for the model with the PCI in the central Northeast Altantic (CNEA). Fields marked with * were not statistically significant.

	pci	scpp	lcpp
pci lag -1	-0.66324	*	0.10919
scpp lag -1	*	-0.35183	*
lcpp lag -1	*	*	-0.59397
pci lag -2	-0.4291	*	*
scpp lag -2	*	-0.33543	*
lcpp lag -2	*	0.134818	-0.35706
pci lag -3	-0.20492	*	*
scpp lag -3	*	-0.17403	*
lcpp lag -3	*	*	-0.20972
pci lag -4	*	*	*
scpp lag -4	*	*	*
lcpp -lag 4	*	*	*
pci lag -5	*	*	*
scpp lag -5	*	*	*
lcpp lag -5	*	*	*
ssnl 1	0.296666	*	*
ssnl 2	0.547463	*	*
ssnl 3	1.469378	*	0.699209
ssnl 4	1.80087	0.453266	0.727565
ssnl 5	*	*	0.741846
ssnl 6	*	*	*
ssnl 7	*	*	*
ssnl 8	0.649296	*	*
ssnl 9	0.647914	*	*
ssnl 10	*	*	*
ssnl 11	*	*	*
SST lag 0	0.23834	0.414508	*
SST lag -1	*	*	*
SST lag -2	*	0.200206	*
SST lag -3	*	-0.25572	*
SST lag -4	*	*	-0.14554
MLD lag 0	*	-0.12425	*
MLD lag -1	*	*	*
MLD lag -2	*	*	*
MLD lag -3	*	*	*
MLD lag -4	*	*	*

TABLE B.8: VAR coefficients estimated for the model with the PCI in the southern Northeast Altantic (SNEA). Fields marked with * were not statistically significant.

	pci	scpp	lcpp
pci lag -1	-0.76864	*	0.103271
scpp lag -1	*	-0.60558	*
lcpp lag -1	-0.06663	*	-0.6182
pci lag -2	-0.60066	*	0.094764
scpp lag -2	*	-0.54206	*
lcpp lag -2	*	*	-0.552
pci lag -3	-0.42643	*	*
scpp lag -3	*	-0.37688	*
lcpp lag -3	*	*	-0.42335
pci lag -4	-0.27292	*	*
scpp lag -4	*	-0.27307	*
lcpp -lag 4	*	*	-0.29818
pci lag -5	-0.17983	*	*
scpp lag -5	-0.06757	-0.18249	*
lcpp lag -5	*	*	-0.18965
ssnl 1	*	*	*
ssnl 2	0.476077	*	*
ssnl 3	1.07636	0.708408	0.83689
ssnl 4	1.422433	0.724539	*
ssnl 5	0.792998	*	-0.48567
ssnl 6	0.246503	-0.58852	-0.75395
ssnl 7	*	-0.55179	-0.45248
ssnl 8	0.373999	*	*
ssnl 9	0.37109	*	*
ssnl 10	*	*	*
ssnl 11	*	*	*
SST lag 0	*	0.254085	0.182096
SST lag -1	*	*	*
SST lag -2	*	*	*
SST lag -3	*	0.118134	*
SST lag -4	*	-0.11594	-0.10684
MLD lag 0	*	*	*
MLD lag -1	*	*	*
MLD lag -2	*	*	*
MLD lag -3	*	*	*
MLD lag -4	*	*	*

Appendix C

Seasonal–trend decomposition of plankton and environmental time–series

STL decomposition of the North Sea time−series

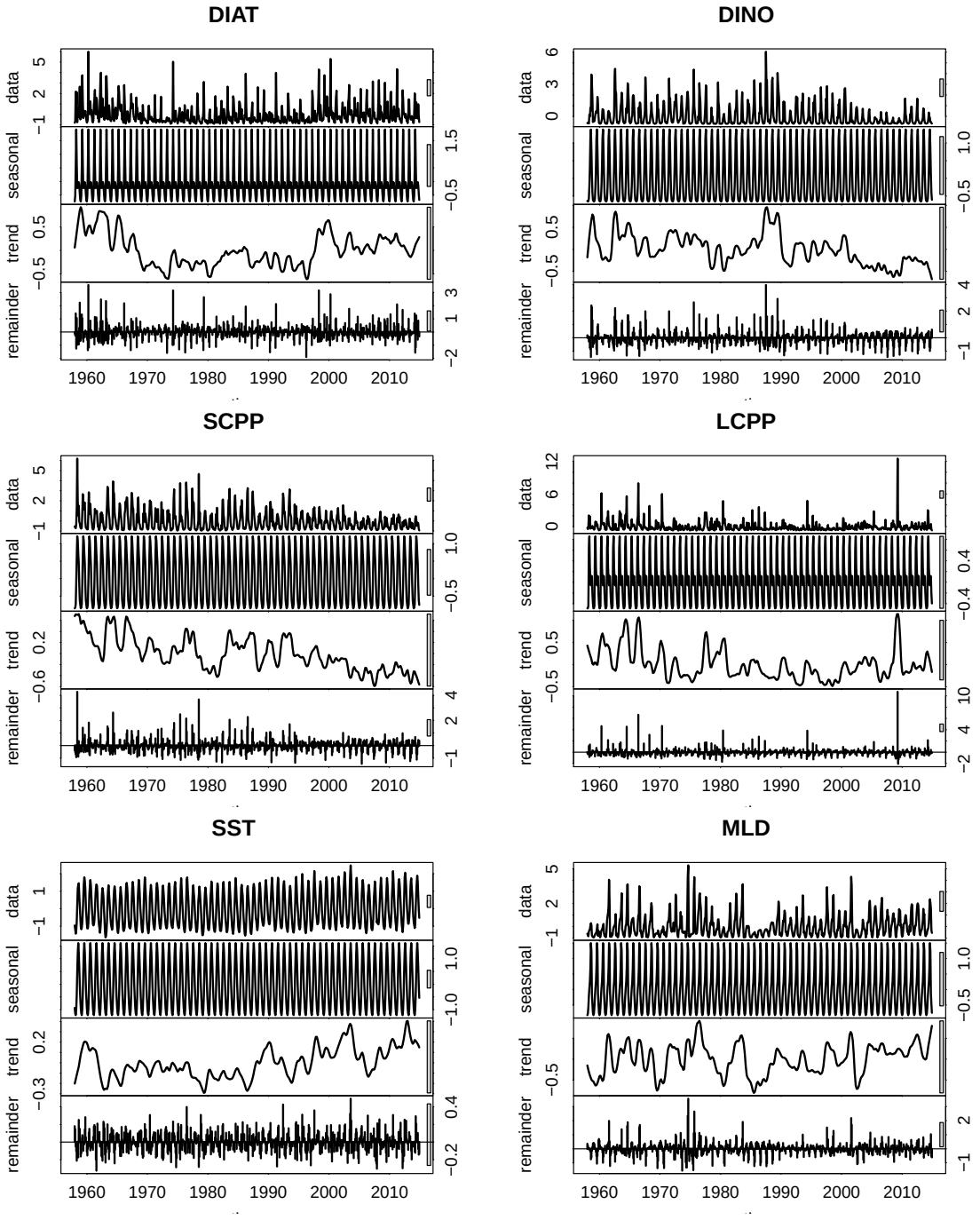

FIGURE C.1: Decomposition of the North Sea (NS, figure 4.1) time-series into a seasonal, trend and noise components, using the STL method of Cleveland et al. (1990), section 5.1.2.

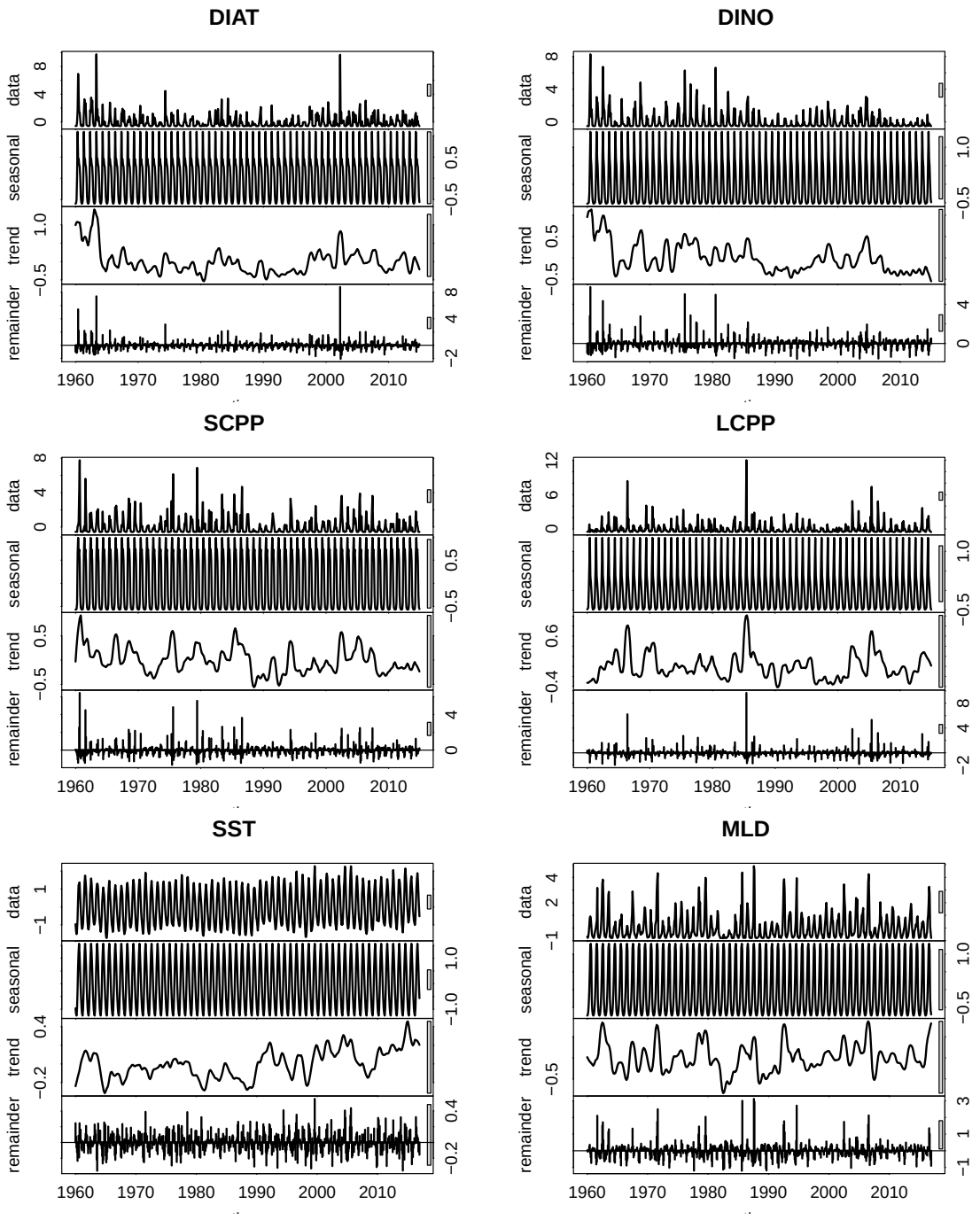

FIGURE C.2: Decomposition of the northern Northeast Atlantic (NNEA, figure 4.1) time-series into a seasonal, trend and noise components, using the STL method of Cleveland et al. (1990), section 5.1.2.

STL decomposition of the central Northeast Atlantic time-series

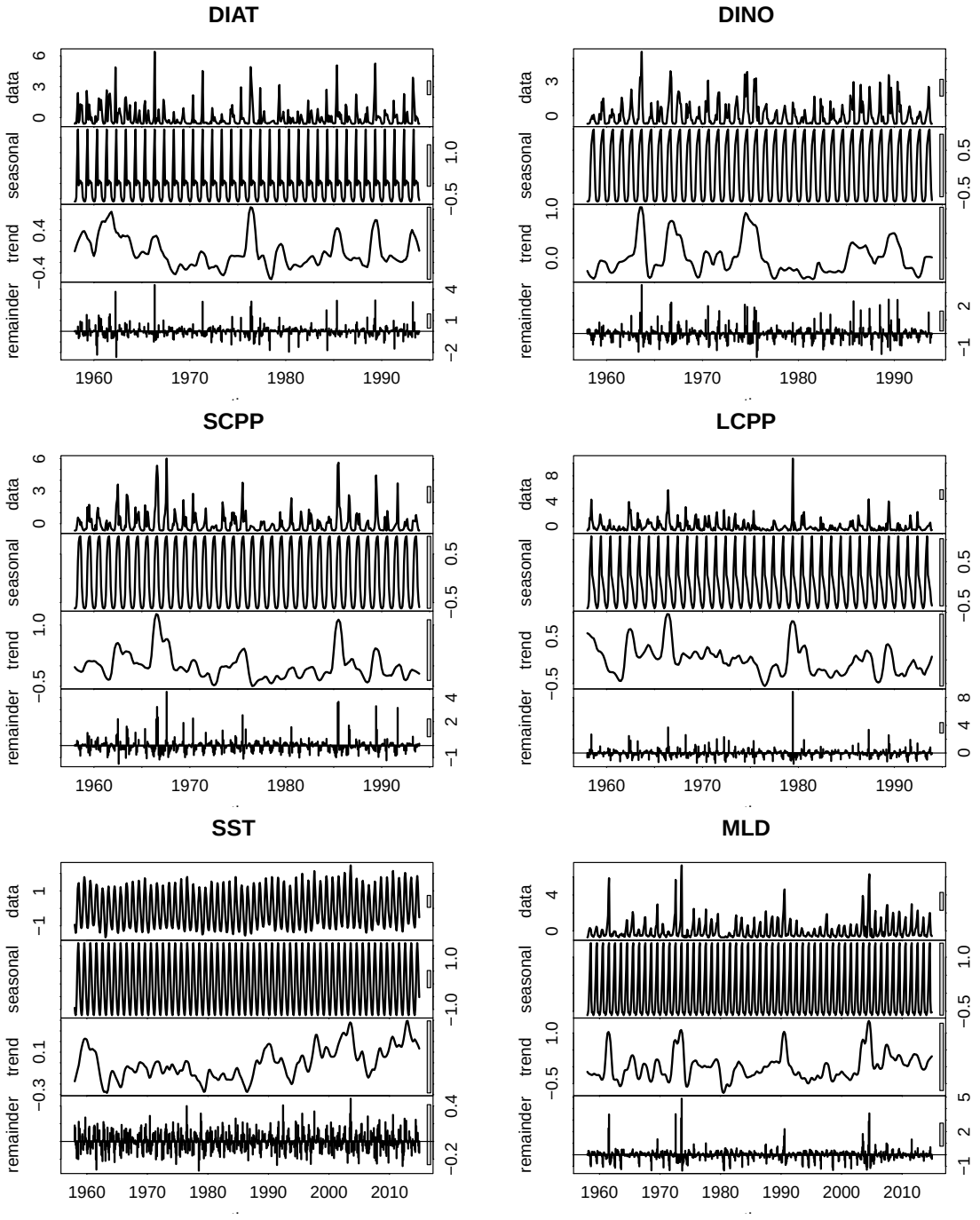

FIGURE C.3: Decomposition of the central Northeast Atlantic (CNEA, figure 4.1) time-series into a seasonal, trend and noise components, using the STL method of Cleveland et al. (1990), section 5.1.2.

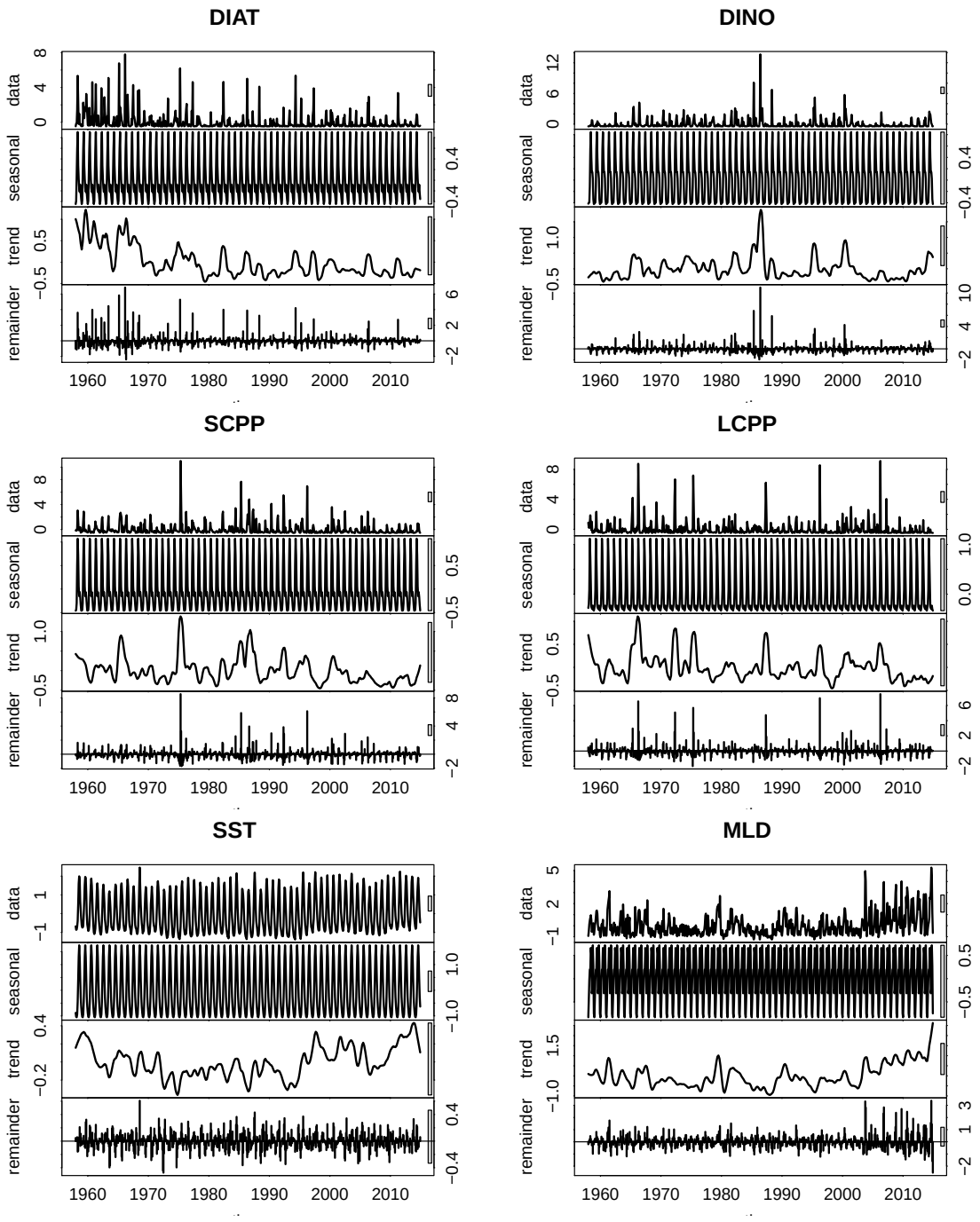

FIGURE C.4: Decomposition of the southern Northeast Atlantic (SNEA, figure 4.1) time-series into a seasonal, trend and noise components, using the STL method of Cleveland et al. (1990), section 5.1.2.

Appendix D

Standardized annual time-series of plankton, SST and MLD

Annual-averaged PCI time-series

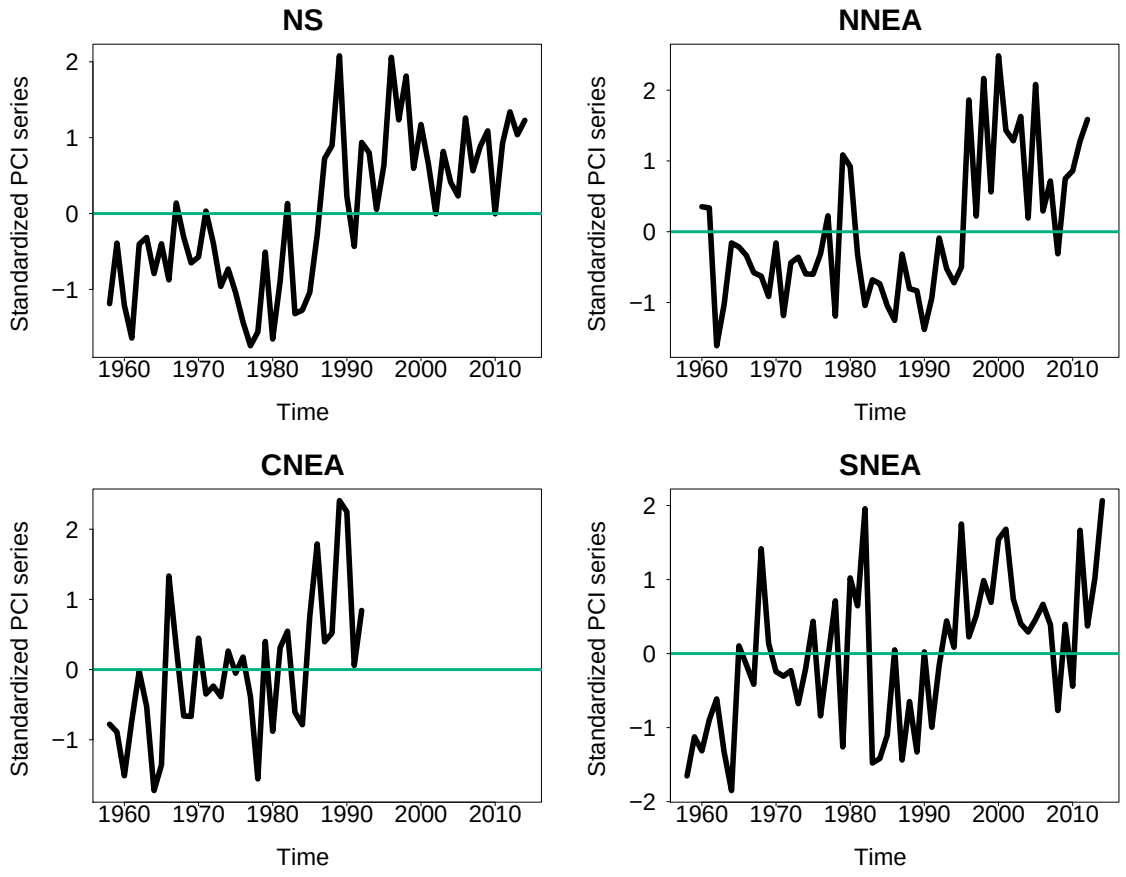

FIGURE D.1: Standardized annual time-series of the PCI in the North Sea, northern, central and southern Northeast Atlantic regions (figure 4.1). Details about the data in section 2.1.

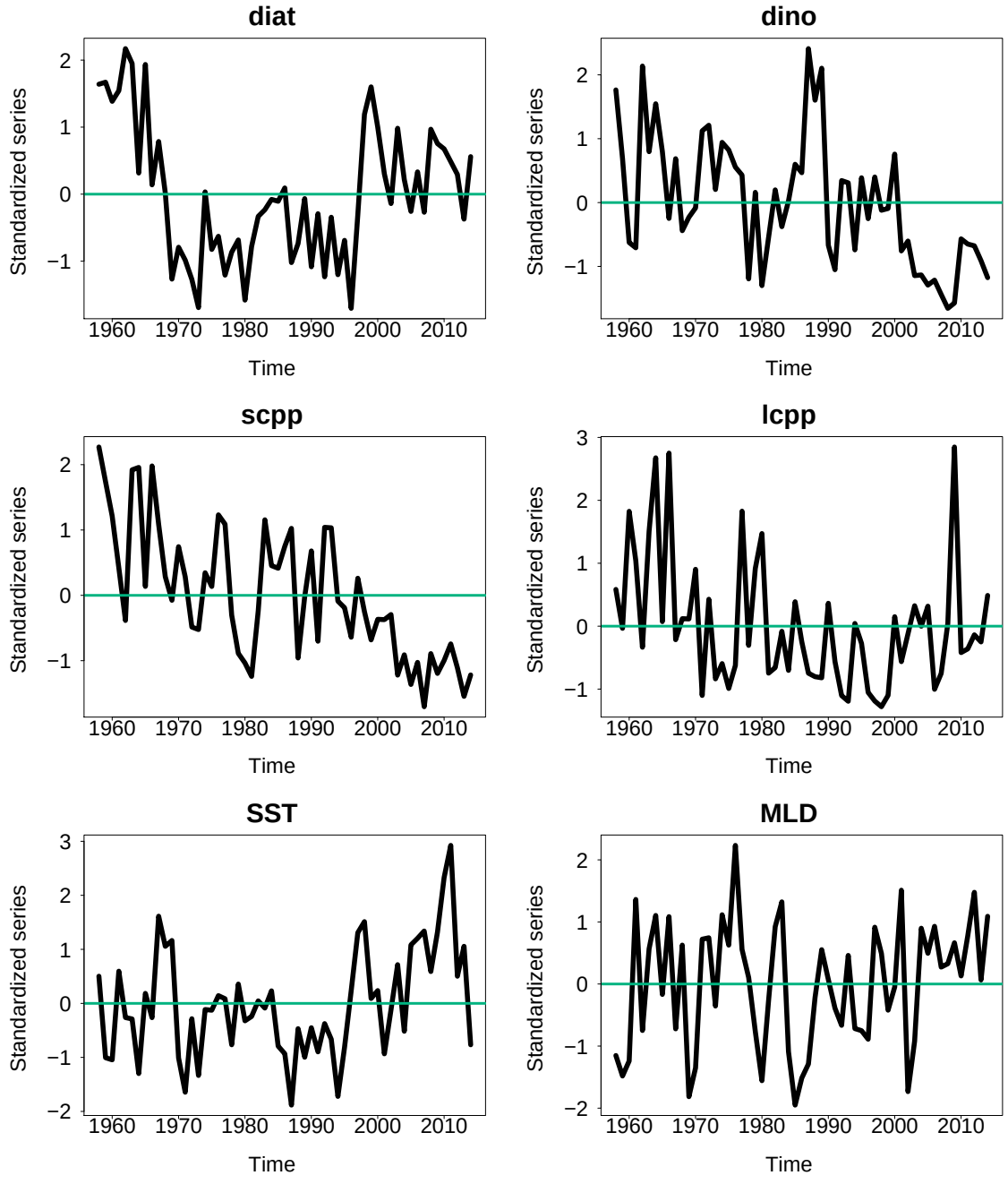

FIGURE D.2: Standardized annual time-series of diatom, dinoflagellate, small copepod and large copepod abundances, SST and MLD in the North Sea (NS, figure 4.1). Details about the data in section 2.1.

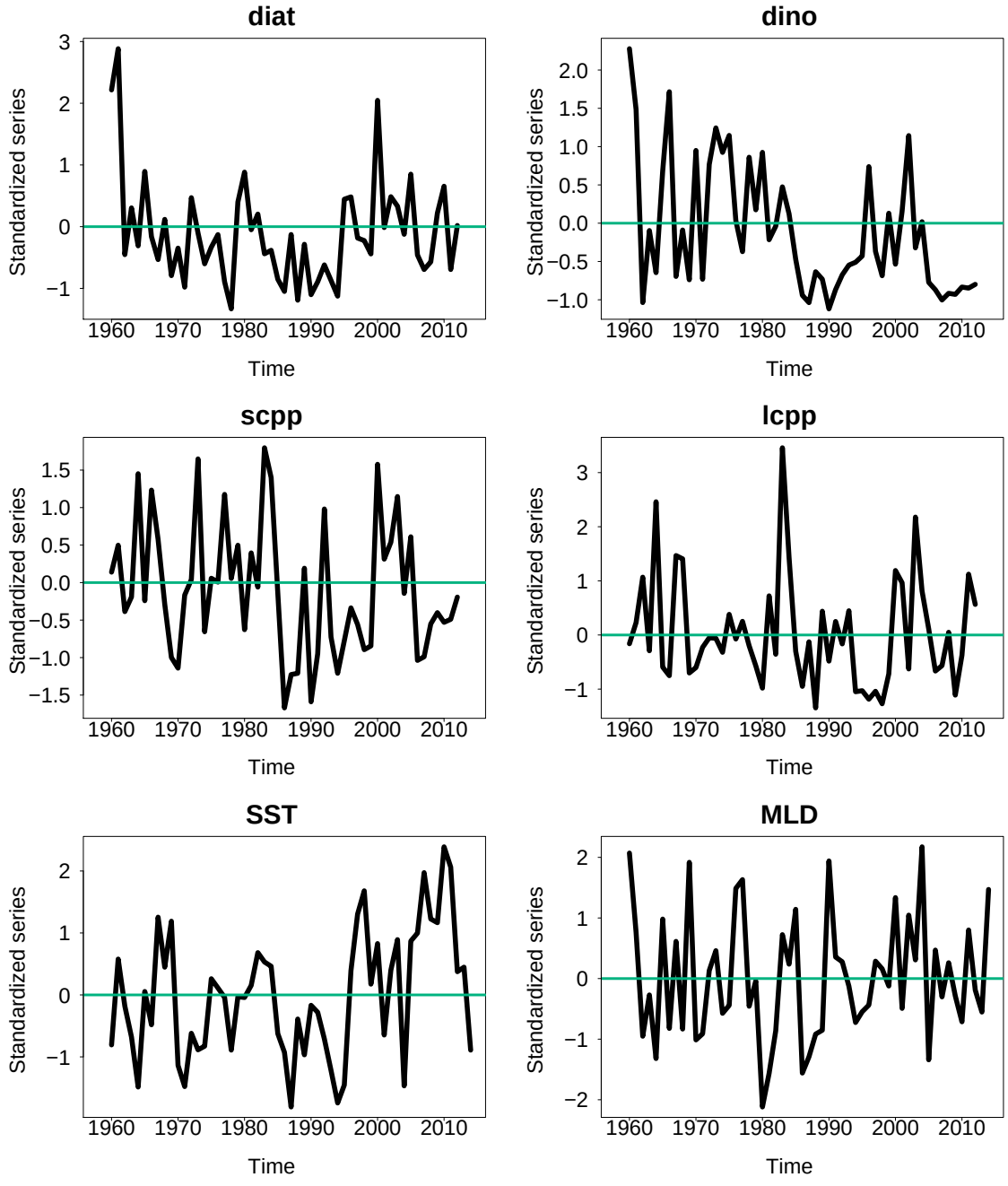

FIGURE D.3: Standardized annual time-series of diatom, dinoflagellate, small copepod and large copepod abundances, SST and MLD in the northern Northeast Atlantic (NNEA, figure 4.1). Details about the data in section 2.1.

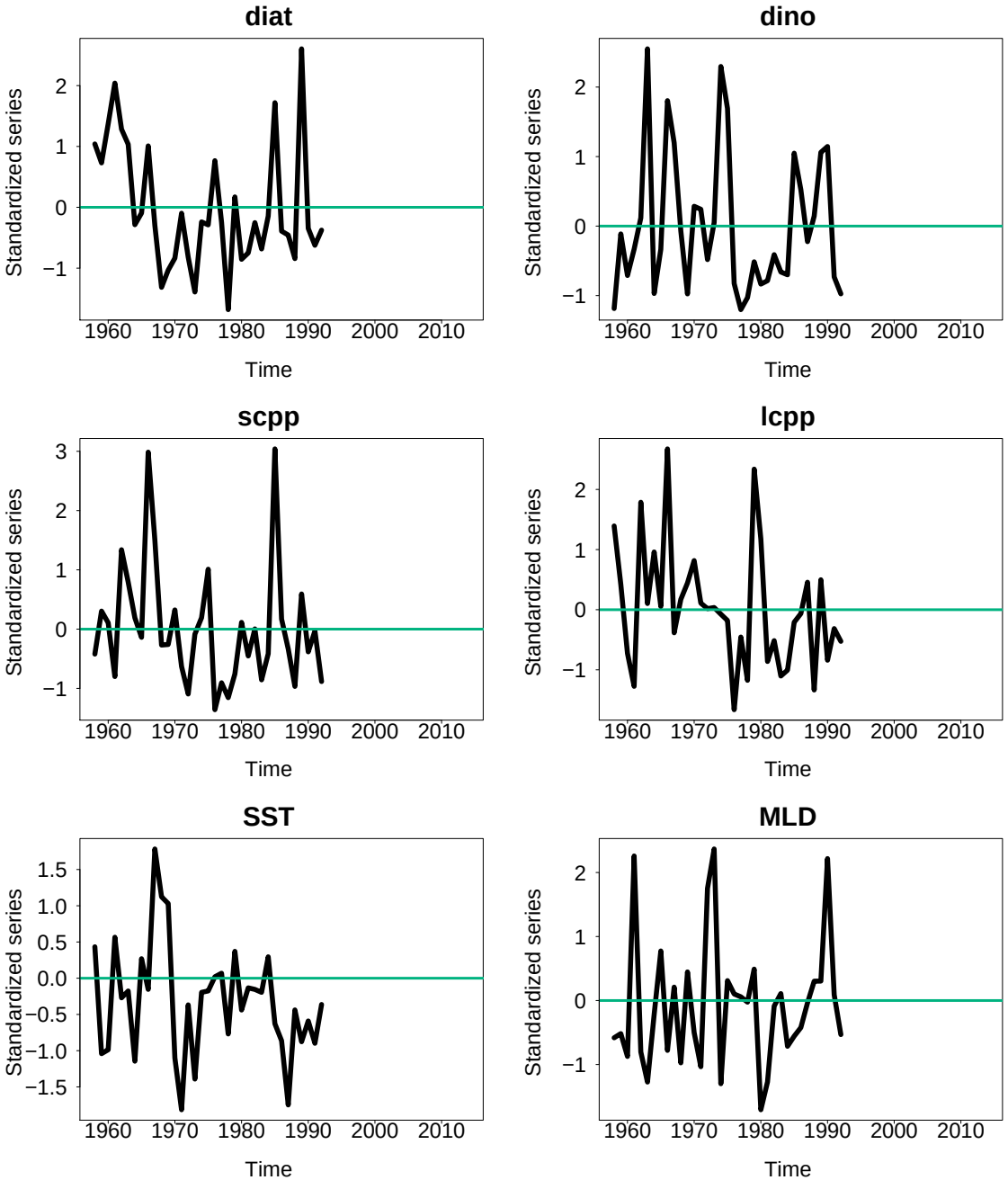

FIGURE D.4: Standardized annual time-series of diatom, dinoflagellate, small copepod and large copepod abundances, SST and MLD in the central Northeast Atlantic (CNEA, figure 4.1). Details about the data in section 2.1.

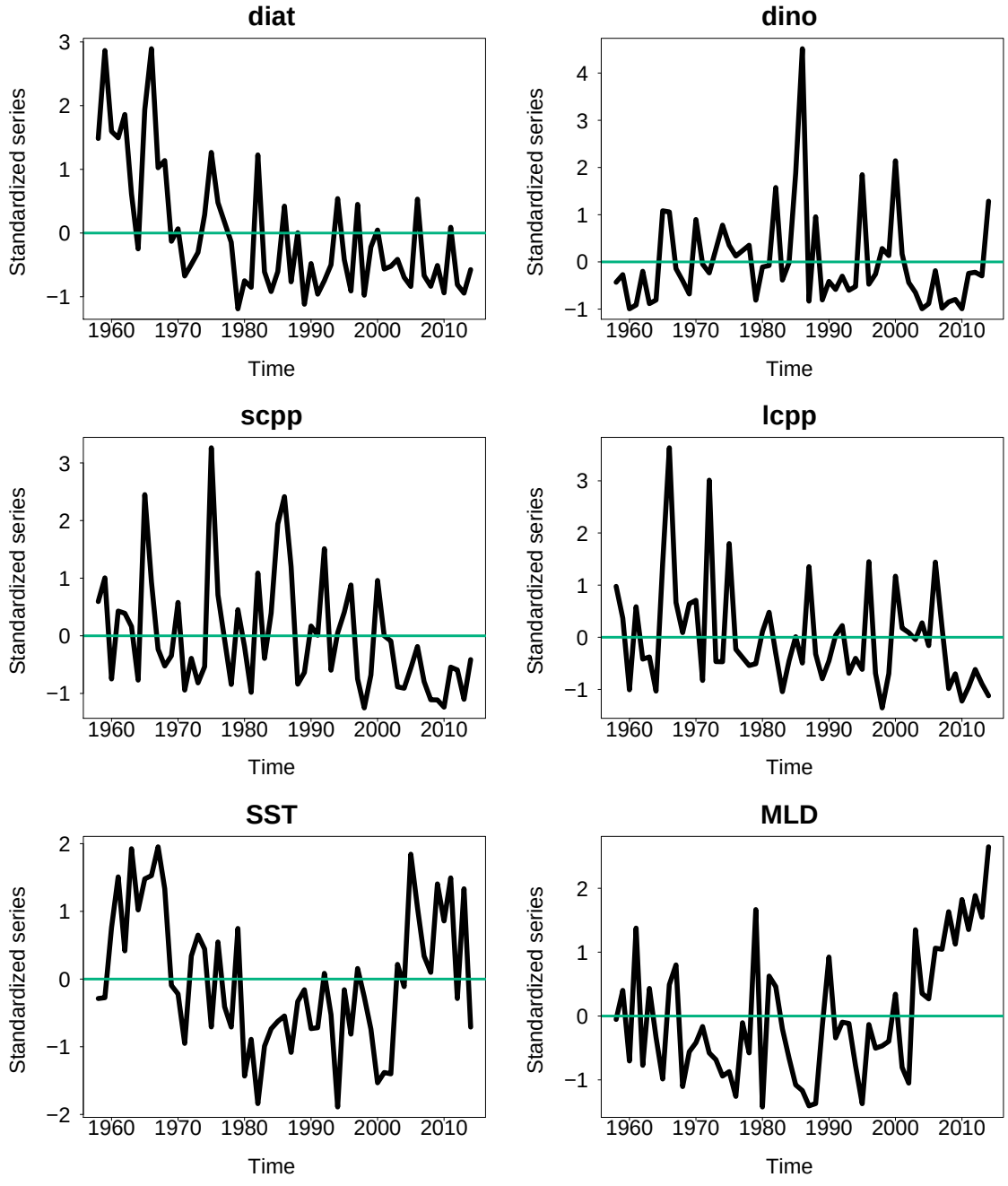

FIGURE D.5: Standardized annual time-series of diatom, dinoflagellate, small copepod and large copepod abundances, SST and MLD in the southern Northeast Atlantic (SNEA, figure 4.1). Details about the data in section 2.1.