Using dynamic Bayesian networks with hidden variables for change inference of the plankton community in the Archipelago Sea

Master's thesis

Rasmus Boman
University of Helsinki
Faculty of biological and environmental sciences
Environmental change and global sustainability

Tiedekunta – Fakultet – Faculty Faculty of biological and environmental sciences	Koulutusohjelma – Utbildningsprogram – Degree Programme Environmental change and global sustainability	
Tekijä – Författare – Author Rasmus Boman		
Työn nimi – Arbetets titel – Title Using dynamic Bayesian networks with hidden variables for change inference of the plankton community in the Archipelago Sea		
Oppiaine/Opintosuunta – Läroämne/Studieinriktning – Subject/Study track Environmental change		
Työn laji – Arbetets art – Level Master's thesis	Aika – Datum – Month and year 09.03.2020	Sivumäärä – Sidoantal – Number of pages 42

Tiivistelmä – Referat – Abstract

The interactions within plankton communities are complex, and realistic modelling of these interactions create a challenge in large-scale environmental models. The objective of this thesis was to evaluate whether Bayesian networks could be a suitable method in the modelling of these communities. Besides observing the interactions between different groups within phyto- and zooplankton communities, another goal was to focus on the potential change on the ecosystem level. To achieve this, dynamic Bayesian networks with hidden variables were used to observe whether structural changes in plankton communities could reveal larger trends in the aquatic ecosystem.

To compare performance and accuracy of the model, two Bayesian food webs with differing causal links between observations were built. Of the two models, the simpler construct utilizing hidden Markov model fared better, and a clear trend was detected in the hidden variable. This trend in the time series signify that the relationships between the observed variables have changed during the study period.

The plankton data set was collected from the Archipelago Sea between 1991 and 2016 and the results from the model were further analyzed alongside with this observational plankton data. In the samples the total biomass of phytoplankton grew throughout the study period, whereas at the same time the total biomass of zooplankton declined. As the Bayesian network considers the observable variables while maximizing the fit of the hidden variable, the observed trend in the hidden variable indicate that some unobservable variables are affecting both phyto- and zooplankton communities. This clear trend detected by the hidden variable might be related to a trend of increasing eutrophication in the study area, but to better understand the drivers causing this change further research is needed. Besides detecting underlying trends, the dynamic Bayesian networks are a promising method to study the interactions within plankton communities.

Avainsanat – Nyckelord – Keywords
Bayesian, Dynamic Bayesian, Hidden variable, Plankton community, Eutrophication, Environmental model, Archipelago sea

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors
Laura Uusitalo (SYKE), Heikki Peltonen (SYKE)

Säilytyspaikka – Förvaringsställe – Where deposited
HELDA - Helsingin yliopiston digitaalinen arkisto / HELDA - Helsingfors universitets digitala publikationsarkiv /HELDA - Digital Repository of the University of Helsinki

Muita tietoja – Övriga uppgifter – Additional information

Tiedekunta – Fakultet – Faculty	Koulutusohjelma – Utbildningsprogram – Degree Programme	
Bio- ja ympäristötieteellinen tiedekunta	Ympäristömuutos ja globaali kestävyys	
Tekijä – Författare – Author		
Rasmus Boman		
Työn nimi – Arbetets titel – Title		
Piilomuuttujilla varustettu Bayes-verkko planktonyhteisöjen muutoksen mallintamisessa Saaristomerellä		
Oppiaine/Opintosuunta – Läroämne/Studieinriktning – Subject/Study track		
Ympäristömuutos		
Työn laji – Arbetets art – Level	Aika – Datum – Month and year	Sivumäärä – Sidoantal – Number of pages
Opinnäytetö	09.03.2020	42

Tiivistelmä – Referat – Abstract

Vuorovaikutukset planktonyhteisön sisällä ovat monimutkaisia ja näiden vuorovaikutusten realistinen mallintaminen on haaste ekosysteemitason mallinnuksessa. Tämän opinnäytetyön tavoitteena oli tutkia soveltuisivatko Bayes-verkot näiden vuorovaikutusten mallintamiseen. Työn toinen tavoite oli tutkia mahdollista muutosta ekosysteemitasolla. Tutkimuksessa käytettiin dynaamisia Bayes-verkkoja piilomuuttujilla ja tarkkailtiin, voisivatko muutokset planktonyhteisöjen rakenteessa heijastua laajempiin muutoksiin akvaattisissa ekosysteemeissä.

Mallien tarkkuuden ja suorituskyvyn vertailua varten luotiin kaksi Bayes-ravintoverkkoa, joissa havaintojen väliset kausaaliset linkit eroavat toisistaan. Yksinkertaisempi rakenne, joka perustuu Markovin piilomalliin, suoriutui paremmin ja havaitsi piilomuuttujassa selkeän trendin. Tämä trendi aikasarjassa viittaa siihen, että tarkasteltavien muuttujien väliset suhteet ovat muuttuneet tutkimusjakson aikana.

Analyyseissä käytetty planktonaineisto oli kerätty vuosien 1991 ja 2016 välillä Saaristomeren tutkimusasemalta ja mallin tuloksia arvioitiin yhdessä näytteistä kerätyn aineiston kanssa. Kasviplanktonin kokonaisbiomassa näytteissä kasvoi tutkimusajanjakson aikana, ja vastaavasti samaan aikaan eläinplanktonin kokonaisbiomassa näytteissä väheni. Bayes-verkko huomioi muutokset tarkastelluissa muuttujissa ja samaan aikaan maksimoi piilomuuttujan sopivuuden. Havaittu muutos piilomuuttujassa viittaa siis siihen, että jotkin muuttujat, jotka eivät ole havaittavissa, vaikuttavat molempien planktonyhteisöjen rakenteeseen. Havaittu kehityssuunta piilomuuttujassa saattaa viitata Saaristomeren rehevöitymiseen tutkimusajanjakson aikana, mutta tarkempien syiden selvittäminen vaatii lisätutkimuksia. Tämän opinnäytetyön perusteella piilomuuttujilla varustettu dynaaminen Bayes-verkko on lupaava menetelmä planktonyhteisön mallintamiseen.

Avainsanat – Nyckelord – Keywords
Bayesilainen mallinnus, Dynaaminen Bayes-verkko, Piilomuuttuja, Planktonyhteisö, Rehevöityminen, Saaristomeri

Ohjaaja tai ohjaajat – Handledare – Supervisor or supervisors
Laura Uusitalo (SYKE), Heikki Peltonen (SYKE)

Säilytyspaikka – Förvaringsställe – Where deposited
HELDA - Helsingin yliopiston digitaalinen arkisto / HELDA - Helsingfors universitets digitala publikationsarkiv /HELDA - Digital Repository of the University of Helsinki

Muita tietoja – Övriga uppgifter – Additional information

Table of contents

1. Introduction

Compared to larger sea basins, the Baltic Sea can be regarded as a relatively closed system with water exchange time ranging from 20 to 30 years (Neumann and Schernewski 2005). This slow turnover rate means that after reaching a certain tipping point some changes, such as a shift from oligotrophic to eutrophic condition, might be virtually irreversible. The latest observed ecological regime shift in Central Baltic Sea occurred between 1987-1989 (Alheit et al. 2005). The change was towards eutrophication, but whether the shift occurred as a result of human impact, climatic changes such as North Atlantic Oscillation, or a combination of both is debatable (Österblom et al. 2007). Regardless of the cause, this shift affected all trophic levels from large sea mammals to fish stocks and phytoplankton blooms. Alheit et al. (2005) detected a general increase in phytoplankton biomass, as well as changes in phyto- and zooplankton community structure. Even though all trophic levels are affected in these shifts, it is often difficult to figure out the broad view from separate variables. In this study I'll expand the scope of these studies to determine whether changes in plankton communities could in turn be used as an indicator of larger trends. In order to achieve this, I'm introducing a statistical model coupling environmental variables and different plankton communities together.

The models that cover the range from physicochemical factors all the way to top predators, or so-called end-to-end models, are still in their infancy (Rose et al. 2010). One of the big challenges in the models that attempt to couple top-down and bottom-up mechanisms, is the accurate submodeling of zooplankton (Carlotti and Poggiale 2010; Rose et al. 2010). In models focusing on higher trophic levels, such as fish stocks, zooplankton has often been treated as a black box, or a singular group (eg. Walters et al. 1997; Daewel et al. 2008). Instead of this simplifying approach Rose et al. (2010) suggest that in end-to-end models the zooplankton submodels should resemble the more complicated modelling of higher trophic levels. One of the goals of this study is to explicate these interactions regarding the different features and traits of zooplankton genera and thus improving the understanding of zooplankton community and its role in the aquatic ecosystems.

Besides aforementioned themes, this study is strongly focusing on testing new methods and their suitability in environmental modelling. To achieve the goals explained above I'm implementing a machine learning method called Bayesian networks with causal links to create a plankton food web of the Archipelago Sea. To complement the examination of large-scale trends, the interactions and temporal changes between different plankton

groups, and the ability to observe these through Bayesian networks, are considered in this study as well.

1.1. Environmental variables and plankton communities

The biomass and community structure of phytoplankton are essentially controlled by bottom-up regulation, via factors such as temperature, light availability and nutrient concentrations (e.g. Eppley 1972; McQueen et al. 1989; Klausmeier et al. 2004). Because of the northern climate, the seasonal succession of phyto- and zooplankton community structure is clearly observed in the Baltic Sea (Kuosa and Kivi 1989; Kivi et al. 1993; Andersson et al. 1996). As my model is built on process-based premises, of special interest are the bottom-up drivers that determine the growth rate of phytoplankton. Of these factors, temperature, light limitation and dissolved inorganic nitrogen (DIN) have been found to be the main regulators for maximum phytoplankton biomass growth in the Baltic Sea (Kivi et al. 1993; Suikkanen et al. 2007).

On the other hand, top-down mechanisms, such as grazing by mesozooplankton (later in text referred to as zooplankton for brevity), have been proved to control the phytoplankton biomass from above (e.g. Müller-Navarra et al. 2004; Strom et al. 2007). From a modelling perspective, the phytoplankton community can thus be considered as a complex entity with hidden information on resource limitation and grazing control. The dynamics of the Baltic Sea ecosystem is based on microbial loop, but still relatively little is known of the exact dynamics within the community (Karjalainen et al. 2007). The species composition of phytoplankton community in turn is decisive in predicting how much of the energy created in primary production is transferred to higher trophic levels (Brett and Müller-Navarra 1997; Müller-Navarra et al. 2004; Ger et al. 2014).

In addition to controlling the total biomass of zooplankton community, the structure of the community is affected by the blooms and dynamics of phytoplankton community (Ojaveer et al. 1998; Karjalainen et al. 2007). These dynamics between zooplankton and their prey are complex and change throughout the year, which can be seen in seasonal succession as well (Kivi et al. 1993; Sommer and Stibor 2002).

Taxonomy-wise the zooplankton community can be divided into rotifers, calanoid copepods and cladocerans, which of the two latter are most abundant in the Baltic Sea (Viitasalo 1992). These groups differ in their feeding habits as well as in their reproductive cycles

(Demott and Kerfoot 1982; Gilbert and Williamson 1983; Sommer and Stibor 2002; Becker et al. 2004). As a result of these differing ecological niches, and especially due to differences in feeding methods, copepods and cladocerans have complementary grazing impacts on phytoplankton (Sommer et al. 2001; Becker et al. 2004; Persson and Vrede 2006; Sommer and Sommer 2006).

A hypothesis originally provided by Porter (1973) suggests that the amount of zooplanktonic grazing depends on the nutritional value of phytoplankton, more specifically the unsaturated fatty acid (HUFA) content. The premise that several zooplankton species favor high-nutrition prey has been confirmed in several studies since. Zooplankton generally favor phylums of *Dinophyta, Ochrophyta, Haptophyta* and *Cryptophyta* and shun phylums of *Chlorophyta* and *Cyanophyta* (Porter 1973; Brett and Müller-Navarra 1997; Müller-Navarra et al. 2004; Persson and Vrede 2006).

Another obvious explanation to selectiveness in feeding is the size of the zooplankton compared to the size of the organisms they feed on (Porter 1973; Engström et al. 2000). Blue-green algae and green algae are not grazed nearly as violently as diatoms, as species of *Cyanophyta* and *Chlorophyta* are believed to be harder to ingest due to their filamentous nature, as well as their ability to flocculate (Porter 1973).

The third possibility to neglection of certain genera is the toxicity of blue-green algae, which inhibits grazing on *Nodularia, Microcystis, Planktothrix, Dolichospermum (*formerly *Anabaena*) and *Aphanizomenon* (Ger et al. 2014). Consequently, the late summer blooms in Baltic Sea tend to be mass-occurrences of *Aphanizomenon* sp.*, Nodularia spumigena* and *Dolichospermum* sp*.* (Suikkanen et al. 2007).

Besides the general tendency of avoiding certain species of phytoplankton, the species composition within zooplankton community is decisive when observing the interactions between phyto- and zooplankton communities. Copepods are observed to select their prey based on its nutritional value and possible toxicity (Engström et al. 2000; Kozlowsky-Suzuki et al. 2003), whereas cladocerans' ability to choose their prey is limited (DeMott and Kerfoot 1982; Sommer and Stibor 2002). These differences in turn launch several size- and structure related feedback impacts on phytoplankton community (Sommer et al. 2001; Sommer and Stibor 2002). One such feedback mechanism is that zooplankton community dominated by large copepods has been observed to promote the growth of larger phytoplankton species (Bergquist et al. 1985). The relative abundance between cladocerans and copepods have also been registered to change during regime shifts (Alheit et al. 2005). Changes in

phytoplankton community reflect directly to higher levels in the food web and are thus paramount to consider e.g. in the modeling of marine ecosystems.

Besides bottom-up mechanisms, also zooplankton is subject to top-down control. The prey of planktivorous fish, such as sprat and Baltic herring, consists mainly of adult copepods and cladocerans (Flinkman et al. 1992; Viitasalo et al. 2001). Rudstam et al. (1992) estimated that 70 % of annual zooplankton production in northern Baltic is consumed by herring. However, even though the link has been established, the size of herring stock has not shown a clear correlation with observed zooplankton biomass (Rudstam et al. 1994; Kornilovs et al. 2001) implying that the bottom-up mechanisms might remain more forceful predictors when determining zooplankton biomass (Flinkman et al. 1998).

1.2. Complexity and uncertainties in end-to-end models

In large scale end-to-end models of marine ecosystems, the complexities of causal web become significant. Physicochemical - phytoplankton regulation has been studied intensively (e.g. Boynton et al. 1982; Brett and Goldman 1996; Suikkanen et al. 2007) and the lowest level of trophic cascade is relatively well understood. As zooplankton is a vital part of aquatic ecosystems, adding zooplankton observations to models could bridge a central gap between phytoplankton and higher trophic levels.

The main challenges in bottom-up models including zooplankton are the expanded complexity and increasing amount of variables affecting the outcome. Such variables include differing life cycles, changing diets, different stages of development, as well as behavioral patterns (deYoung et al. 2004; Rose et al. 2010). Modelling these variables while considering only biomass of the different genera might prove to be problematic. In purely process-based models the further up one maneuvers from the lowest trophic level, uncertainties accumulate, and the risks of spurious correlations increase (McQueen et al. 1989; deYoung et al. 2004).

An alternative approach to bottom-up modelling is the top-down method used often in ecosystem modelling (Carlotti and Poggiale 2010). Rather than process-based premise, this method takes interactions between species or groups as a starting point and adds complexity stepwise by adding variables. This top-down approach is more common with ecological models where interactions are biological-behavioral, whereas bottom-up

procedures are prevalent in biogeochemical modelling (deYoung et al. 2004). Zooplankton is in the intersection of these two approaches, and whether one approach is superior to another remains debatable. My presumption for the modeling is of probabilistic nature with a bottom-up approach, but the results could be reflected to fish stocks to examine potential correlations with higher food web levels.

A more detailed description of challenges in end-to-end models can be found in Fulton et al. (2003), deYoung et al. (2004), Carlotti and Poggiale (2010) and Fulton (2010).

1.3. Dynamic Bayesian model

The use of Bayesian networks in ecological modelling has grown steadily through the 21st century (e.g. Clark 2005; Aguilera et al. 2011; Pérez-Miñana 2016). Compared to traditional statistical models, Bayesian models can better consider uncertainty and complexity and can accommodate diverse and incomplete sources of information, which is often considered as an advantage in ecosystem models (Clark 2005; Uusitalo 2007; Aguilera et al. 2011). Bayesian networks are also graphically easy to build and understand, making them useful in projects where several stakeholders of different backgrounds are involved (Chan and Pollino 2012). A deeper insight of Bayesian networks for ecosystem modelling can be found in Clark (2005), Uusitalo (2007), Aguilera et al. (2011) and Landuyt et al. (2013).

The major issues with Bayesian networks include the possible discretization of continuous variables, structuring the expert knowledge to a meaningful form and the challenges in processing feedback loops (Uusitalo 2007; Landuyt et al. 2013; McDonald et al. 2015). Large-scale models often assume that the relationships between variables remain unchanged through time, but in some cases, this might be a false assumption. Dynamic Bayesian networks (DBN) attempt to tackle temporal issues by creating a time slice of each individual Bayesian network and linking it to the next one (Murphy and Russell 2002; Robinson and Hartemink 2010). As aquatic ecosystems rarely are in a static state, including a temporal aspect in Bayesian modelling has increased predictive performance in several studies (Tucker and Liu 2004; Trifonova et al. 2017; Trifonova et al. 2019).

Another issue that needs to be addressed are the variables that are either unmeasurable or of which there might not be sufficient data available. So called hidden variables (HV) have been suggested as a solution in modelling uncertainties such as these in physics already in

the late 1960s (Clauser et al. 1969). However, in aquatic ecology they are a novelty with only few studies so far (e.g. Trifonova et al. 2015; Uusitalo et al. 2018; Maldonado et al. 2019).

These hidden variables are variables with no data to begin with, used to measure relative changes in the observations over time (Murphy 2012; Uusitalo et al. 2018). Hidden variables are utilized to maximize the fit of the model to the observed data. Changes in the values of hidden variables might signal changes in the environment and interactions within the system. Hidden variables can be linked to one or several variables within the model, depending on the desired outcome (Murphy 2012). Within DBN, hidden variables act as regular nodes within the model, with links to several other nodes. As in Bayesian modelling generally, links to and from hidden variables can either be determined by the modeler using expert evaluation, or they can be established from the data using Bayesian structure learning (Zhang et al. 2005).

2. Materials and methods

2.1. Collection of the samples and data wrangling

2.1.1. Zooplankton samples

Zooplankton samples were collected by the Archipelago Research Institute of the University of Turku between years 1991 and 2013 at a monitoring station located at 60°15.315´ N and 21°57.174´ E. The samples were collected on average once a month from April to October, winter samples more infrequently. The total amount of yearly samples ranged from 4 to 14, with an average of 10 yearly samples. These were collected with a standard plankton net (mesh size 150 µm, mouth diameter of 33 cm) with a single haul from the depth of 25 meters to the sea surface. The total depth at the station was 50 meters.

The contents of the net were emptied to a 200-mL plastic bottle and later stored in buffered formalin (4%) and analyzed according to standard methods established by HELCOM (1988). The samples were identified to either genus level, or whenever possible, to species level. Copepods were identified into nauplius, copepodite stages CI-CIII, copepodite stages CIV-CV, females and males. Cladocerans and rotifers were identified into juveniles and/or adult stages.

As the samples were counted in individuals m^{-3} I calculated the wet weight biomass in µg m^{-3} using a wet weight table used by the Finnish Environmental Institute in the national monitoring programme. Non-zooplanktonic species (e.g. *Balanus* sp., *Marenzelleria* sp.) were removed from the data. The species were further summed up to family-level and combined to meaningful groups based on their taxonomy.

2.1.2. Phytoplankton samples

Phytoplankton samples were collected at the same monitoring station as the zooplankton samples between years 1991 and 2016. The total amount of yearly samples ranged from 1 to 18, with an average of 9 samples each year. The samples were collected with a single haul from a depth of 2 * secchi depth in order to catch the current productive layer. Samples were conserved in the mixture of Lugol + AA. The results were saved in the BVetRek-register of the Finnish Environment Institute. Whenever possible, the samples were identified to species level. For the Bayesian model, the biomass of phytoplankton species

was further summed up to seasonal and phylum-level.

2.1.3. Water quality data

Water quality data was collected from the same monitoring station as phyto- and zooplankton samples between the years 1991 and 2016. The total amount of yearly samples ranged from 17 to 25, with an average of 20 yearly samples. The environmental data was collected from depths of 1, 10, 15, 20, 40, 49.5 meters. The most crucial measurements for my purposes were temperature, dissolved inorganic phosphorus (PO_4), dissolved inorganic nitrogen (NH_4-N, NO_3-N, NO_2-N) and salinity. As the phytoplankton data was collected from surface layer, the environmental variables used in the model were mean values of observations from 1 and 10 meters respectively.

2.1.4 Wrangling and combining of the data

Data wrangling was executed with R software (https://cran.r-project.org/). In order to hold the number of nodes in the Bayesian model feasible, I divided the observations into meaningful categories. I calculated the seven most abundant phytoplankton classes in terms of biomass. The bio volumes for phytoplankton species were already calculated in the original data table. The seven classes were Diatomophyceae, Dinophyceae, Litostomatea, Cyanophyceae, Cryptophyceae, Chrysophyceae and Prymnesiophyceae. These phytoplankton classes combined resulted in 94.7 % of the total phytoplankton biomass in observations.

Similar calculations were made with zooplankton data in order to identify the most abundant genera and species. All the developmental stages were included in the abundance and wet weight of each species. As there are fewer taxa in zooplankton than phytoplankton, most observations could be identified to species-level. *Acartia* sp., *Daphnia* sp., *Eubosmina longispina*, *Eurytemora affinis*, *Evadne nordmanni*, *Pleopsis polyphemoides* and *Synchaeta* sp. totalled to 96.8 % of the total zooplankton biomass.

The samples were not always taken on the same dates. My criteria were to have environmental variables for every phyto- or zooplankton sample. I included the zooplankton and phytoplankton samples if there was a measure of environmental variables within 14

days to either direction. This resulted in 206 phytoplankton observations and 203 zooplankton observations. In these observations, the mean difference between phytoplankton sampling date and zooplankton sampling date were 3.4 days, and median was 2 days respectively.

201 out of 206 phytoplankton samples were taken on the same date as the environmental samples resulting in mean and median difference of 0 days. Zooplankton samples had a mean and median difference of 4 days when compared to dates of environmental samples. Data combed in aforementioned technique resulted in total of 539 environmental observations, 206 phytoplankton observations and 203 zooplankton observations. These observations were further combined into quarters based on months; December (previous year) -February = Winter, March-May = Spring, June-August = Summer and September-November = Fall. This resulted in 91 distinct observations.

For the Bayesian model, the observational data was logarithmized and standardised (mean value 0, standard deviation 1). The complete script for the data manipulation can be found on https://github.com/RasBoman.

2.2. Bayesian network model

2.2.1. Bayesian network and dynamic Bayesian networks

A Bayesian network consists of separate nodes that are linked to each other with causal or noncausal relations. These nodes and relations create a directed acyclic graph (DAG), which serve as the starting point in model building. The conditional probability tables (CPT) are variables linked to each descendant node and provide the basis for computational Bayesian inference. The variables consist of qualitative and quantitative components. Qualitative part creates the DAG and determines whether there is an assumed connection between the two variables. Quantitative component is a conditional distribution for each variable.

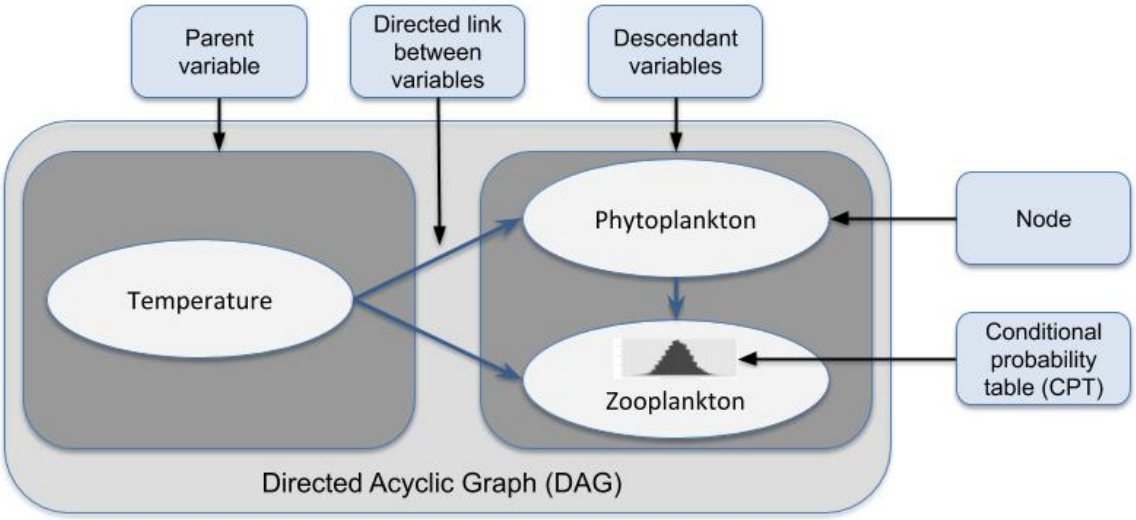

Figure 1. A simplified Bayesian network.

The joint probability for Bayesian network:

$$P(X_1, X_2, \ldots, X_n) = \prod_{k=1}^{n} P(X_k \mid parents(X_k)) \quad (1)$$

Where $X_k, X_1, X_2, \ldots, X_n$ represents all the variables in the model and $parents(X_k)$ the parents of X_k.

In the case of the example in Figure 1 joint probability would translate to:

$$P(Temp, Phyto, Zoopl) = P(Temp)\, P(Phyto \mid Temp)\, P(Zoopl \mid Temp, Phyto)$$

As represented in figure 1, temperature variable is a parent node, phytoplankton node is a descendant of temperature variable and zooplankton is a descendant of temperature as well as phytoplankton variables.

In a dynamic Bayesian network (DBN) each Bayesian network acts as a separate time slice. In this context dynamic refers to considered time dimension, not to changing structure. Some variables of each time slice are connected to the next time slice addressing the connections of temporal relationships between variables.

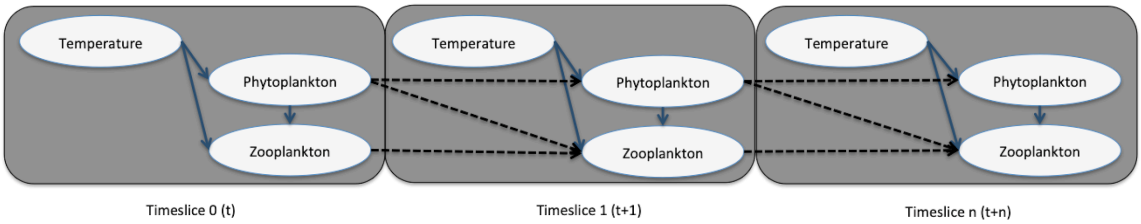

Figure 2. The structure of simplified dynamic Bayesian network

DBN takes into account temporal element linking variables across time. The first time slice (t=0) results in Equation (1) but beginning from the second time slice (t=1) temporal aspect is included. The conditional probability $P(X_n|\mathcal{Y}_n)$ can also be split into the product of conditional probabilities of each variable as

$$P(X_n|\mathcal{Y}_n) = \prod_{k=1}^{n} P(X_k^{(t)} \mid parents(X_k^{(t)}), parents(X_k^{(t-1)})) \quad (2)$$

Where $\mathcal{Y}_n$ represents the $parents(X_k^{(t)})$ and $parents(X_k^{(t-1)})$ of the node $(X_k^{(t)})$ in the same time slice as well as in the previous one. As can be observed in figure 2, the zooplankton variable of time slice *n* is a descendant of parent variables phytoplankton and zooplankton in the previous time slice as well as temperature and phytoplankton in the current one.

No matter how well the model is built, in natural environment there are always additional unaccounted variables and forces affecting the target variables. These so called latent, or hidden variables might be completely unobservable, or one just might not have accurate data of them. In order to uncover these, hidden variables can be implemented into models to expose dynamics that are not represented by the observed variables, but still influence them. These dynamics might further reflect to variables that are provoking the change, and thus help to identify these. When hidden variables are analyzed in the context of additional, observed data, they might reveal larger tendencies in the ecosystem.

2.2.2. The purpose, structure and features of the model

The purpose of the model was to find out whether based on plankton data, there was a trend or a regime shift to be observed in the Archipelago Sea during the study period. The construction of the Bayesian network has been structured utilizing the guidelines offered by Borsuk et al. (2004), Jakeman et al. (2006), Chen and Pollino (2012) and McDonald et al. (2015). The selection of model features and establishing causal links between variables is the most challenging task in creating a Bayesian model. The identification of the nodes and parameters was conducted based on previous studies and expert knowledge. To select the proper variables, I used expert knowledge of Laura Uusitalo, Heikki Peltonen, Veera Norros and Harri Kuosa, as well as profound literature review on previous studies of the subject. The model within one time slice is presented in figure 3.

Temperature, salinity and nutrients represent the parent nodes in the model. Phytoplankton was divided to 7 categories based on their classes, zooplankton was divided into 6 categories based on their genus and each of these categories act as an individual node. As plankton communities and their interactions vary significantly depending on the time of the year, in the analysis of the model each season was examined separately. Connections between time slices were made based on expert elicitation. To quantify the change, two hidden variables were added to DBN to represent non observable changes.

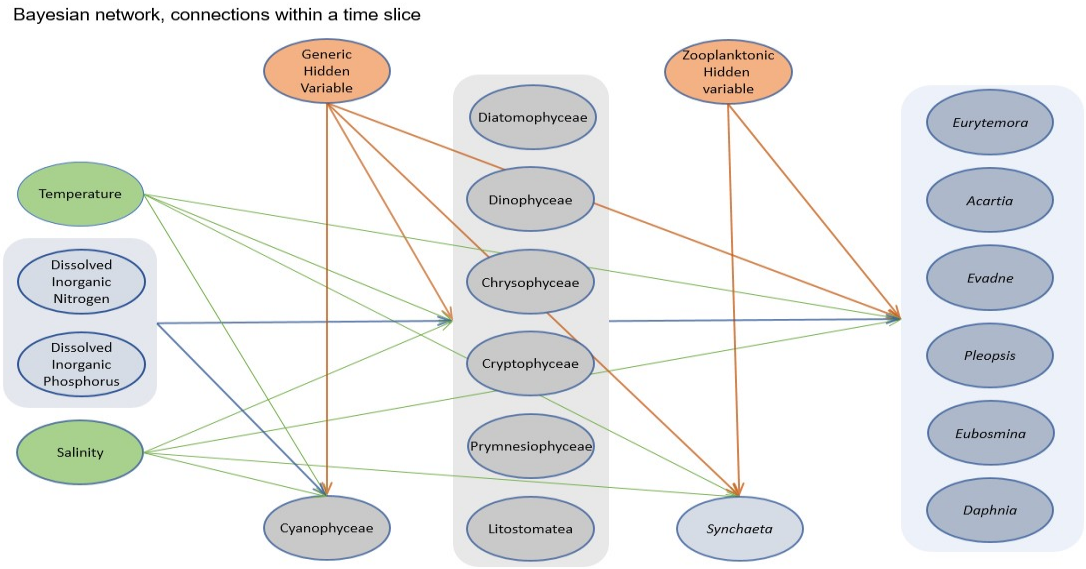

Figure 3. The Bayesian network used as a basis in MATLAB-models. The phytoplankton classes and zooplankton genera have been grouped here for illustrative purposes, but act as individual nodes in the model.

Cyanophyceae contains no causal links to zooplankton genera because of its low nutritional value in the food chain. *Synchaeta* on the other hand receive no causal links from phytoplankton classes, as its traits and feeding habits differ significantly from cladocerans and copepods. A generic hidden variable is assumed to affect the whole biotic system as an explanatory variable, a zooplanktonic hidden variable is connected solely to zooplankton genera.

To be able to better compare the results and their performance, two different dynamic Bayesian networks were created. The first model is based on simple hidden Markov model (Schuster-Böckler and Bateman 2007). With hidden variables in time *t* linked only to the corresponding variables in the next time slices (time *t* + 1), according to Markovian assumption, observations are conditionally independent of all the previous states before time *t* (figure 4).

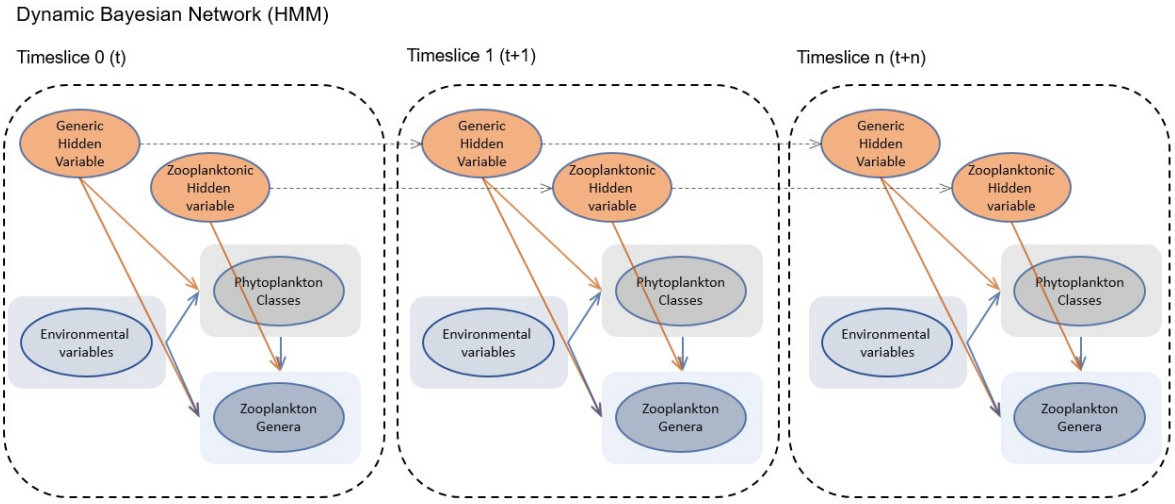

Figure 4. The Hidden Markov model, with only hidden variables linked to themselves across time slices. The causal links between time slices are represented as dotted arrows.

The other dynamic, autoregressive model links environmental and plankton observations across time slices. Some of the variables in the time slice *t* have a direct effect on the time slice *t*+1. In this model the phytoplankton and zooplankton observations are assumed to affect the corresponding biomass variables and thus the community structure in the next time slice. Each phytoplankton class, excluding cyanophyta, has also a causal link to each zooplankton genera in the next time slice along the model seen in figure 3 and 5.

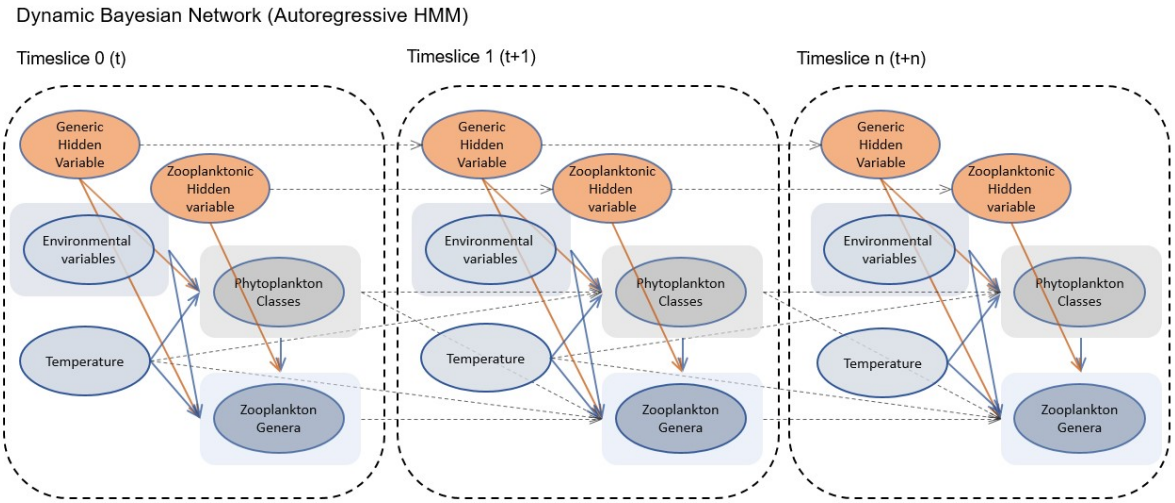

Figure 5. The autoregressive hidden Markov model, with observations linked to next time slice as well. The causal links between time slices are represented as dotted lines.

2.2.3. MATLAB models, evaluation and testing

The models built were executed in MATLAB using BayesNet Toolbox (Murphy and Russell 2002). The model utilizes Expectation-Maximization (EM) algorithm to estimate hidden variables in order to produce the maximum likelihood estimates of parameters (Dempster et al. 1977). To make the model more robust, the observed variables were modelled as Gaussian, the links between these variables were assumed linear and the covariance matrix diagonal. As the structure of the model itself is quite complex, linearity simplifies the model so that every conditional probability table of a variable is assumed to be a linear combination of the CPDs of its parent variables. To increase the chance of finding the global optimum, each model was run 80 times and the model with the best log-likelihood was saved as the result. The log-likelihood is a function that associates probability density to each of the parameters in a given sample.

The accuracy of the model was evaluated with two different predictions. The first criteria was how well it could deduce zooplankton succession from environmental and phytoplankton observations inside the observed data set. To achieve this, the last three years, or twelve seasons, of zooplankton observations were removed from the table. The learned model was then used to infer the succession of these zooplankton genera. These inferred means were compared to observed data from these years. *Pleopsis* and *Daphnia* genera were omitted from the predictions as they included missing values in the predicted years.

To observe the ability of predicting future trends independently, the model was trained based on the first 80 observations and the predictions were made for 4 observations, or one year. This second prediction was made in order to see how well the trained model could predict changes in zooplankton community without any observational data available. The accuracy of different models was further evaluated by calculating the natural logarithm of the likelihood, log-likelihood, of the average mean of predicted value compared to observed value. These parameter values try to maximise the likelihood that the estimates produced by the model depict the observed data. and can be used to compare results of the different models (Grossman and Domingos 2004). As the log-likelihood is a monotonically increasing function, higher values signify that the observed result is more likely to occur compared to the alternative. Further analyses and graphs were constructed in R software.

In order to examine the possible link between hidden variables and real observations of fisheries data, I compared fish catches from the area to the Hidden variables of my models. The fish catches were the sum of the yearly catches caught by professional fishermen between July and September in the Archipelago Sea. This data set was collected by natural Resources Institute Finland between 1991 and 2016. These were compared to Hidden variables of summer season (June-August) with Pearson’s correlation.

3. Results

3.1. Plankton succession

The high variability of biomasses between different seasons can be recognized in both phytoplankton (Figure 6) and zooplankton (Figure 7). As expected, phytoplankton biomass is greatest during its springtime bloom and zooplankton increases shortly thereafter causing a biomass peak of zooplankton during summer. Fluctuation inside plankton communities can also be examined during the study period. In phytoplankton community especially the relative abundance of Dinophyceae and Litostomatea have increased since 2010. In 2006-2007 a clear surge in the biomass of springtime Diatomophyceae can be scrutinized. During summers 2000-2012, and especially 2009 and 2010, an increase in cyanobacterial blooms can be observed, but variability remains high.

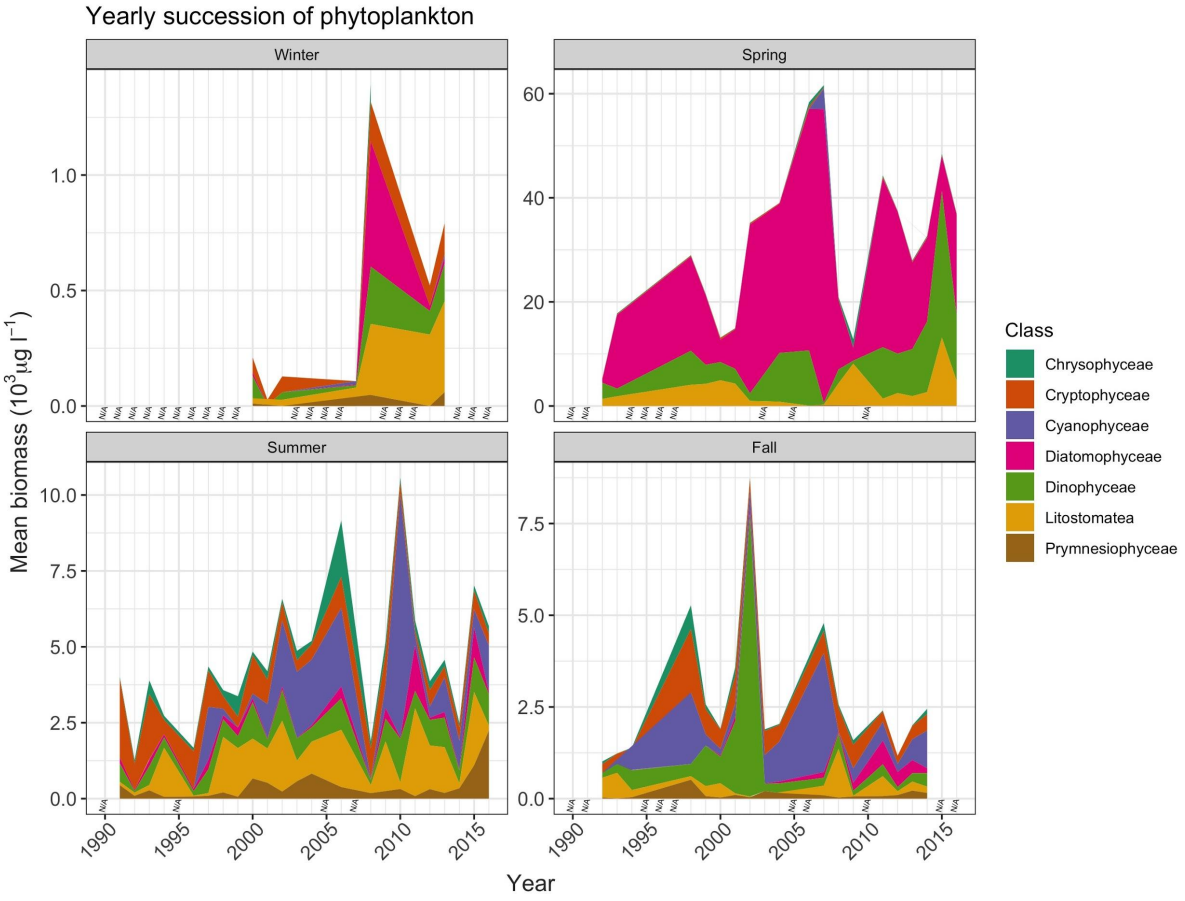

Figure 6. The yearly succession of phytoplankton. Years with no data available are marked with “N/A” on the x-axis and the missing data is extrapolated for visualization purposes. Note the different scales on y-axis.

The overall phytoplankton biomass shows an increasing trend during the study period. The magnitude of diatom (Diatomophyceae) spring bloom is clearly observed in these plots.

The summed biomass of zooplankton shows opposite trends compared to phytoplankton community (Figure 7). It is noteworthy that the main zooplankton bloom happens in the summer, after the diatom and dinophyte bloom in April-May. During springtime, zooplankton biomass has diminished towards the end of observed period, excluding 2008 anomaly of *Synchaeta* and *Pleopsis* biomasses. Summertime observations reveal significant changes inside zooplankton community as *Eurytemora* and *Eubosmina* duel of dominance. Towards fall *Acartia* mainly dominates the colony.

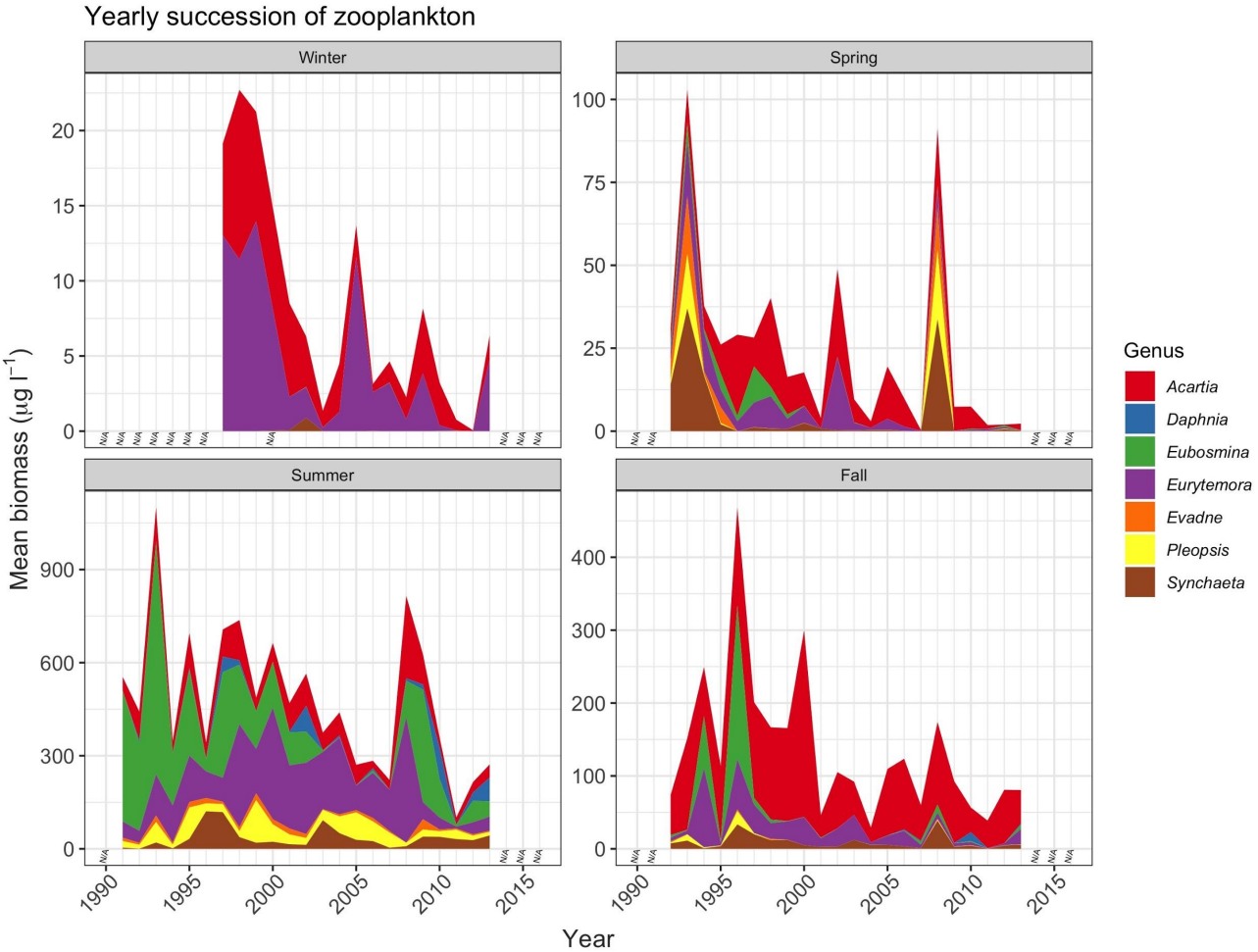

Figure 7. The yearly succession of zooplankton, years with no data available are marked with “N/A” on the x-axis and the missing data is extrapolated for visualization purposes. Note the different scales on y-axis. The mean biomass has declined during the study period, which can be regarded within all seasons.

3.2. Dynamic Bayesian Networks

The results of the dynamic Bayesian Network models showed significant variation, depending on the model structure (Figures 8-11). Some anomalies, such as sudden increase in spring- and summertime zooplankton biomass in 2008, can be identified in the hidden variables as well.

3.2.1. Hidden Markov Model

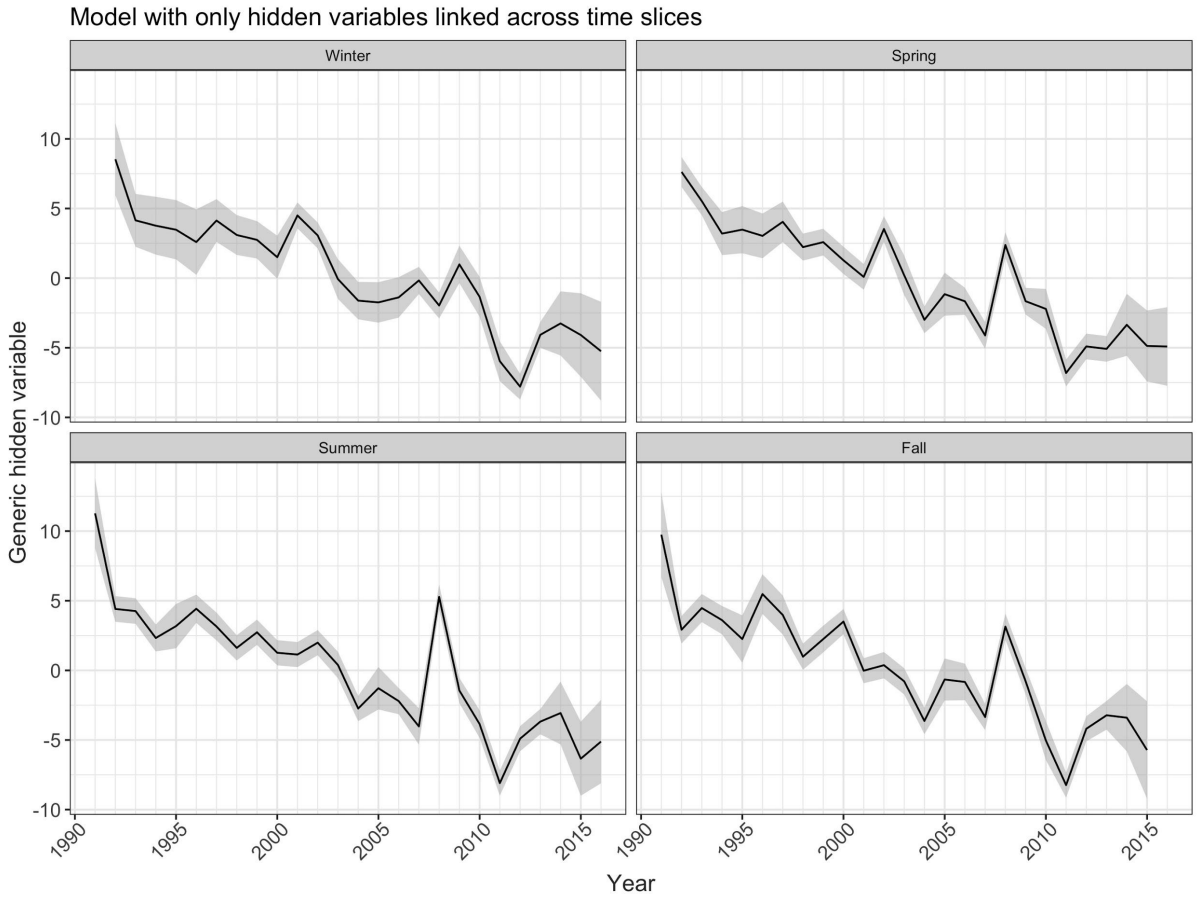

Figure 8. The mean value of generic hidden variable from the model based on figure 4, with only hidden variables linked to themselves across time slices. Standard deviation is represented as grey area.

The downward trend detected by the model is clearly observed in all of the seasons. There is an anomaly in 2008, where the mean value of hidden variable surges from negative to positive. Similar anomaly can be observed in zooplankton data. The variation increases

towards the end of the study period, likely because of the missing values in zooplankton data.

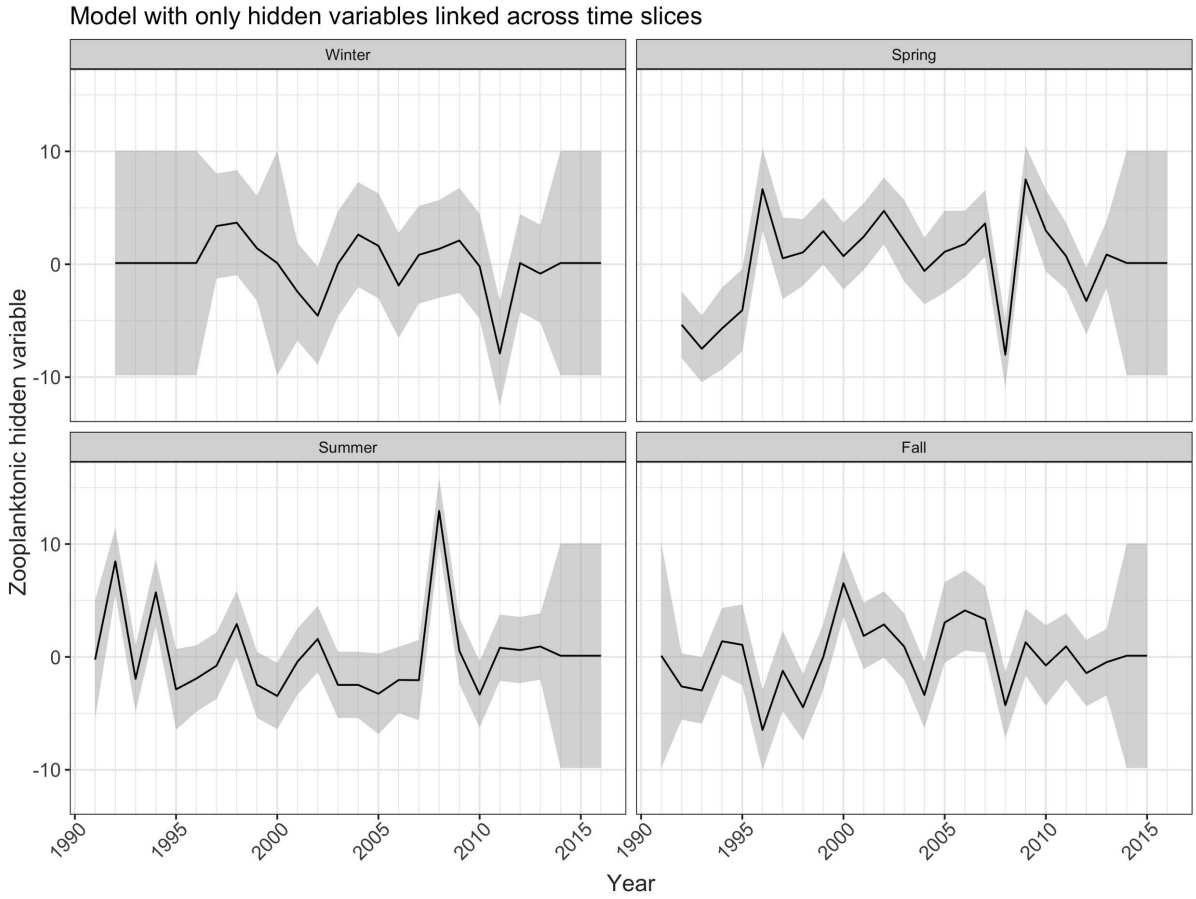

Figure 9. The mean value of zooplanktonic hidden variable from the model based on figure 4, with only hidden variables linked to themselves across time slices.

Compared to generic hidden variable, there isn't a clear pattern to be observed in the zooplanktonic hidden variable, and excluding year 2008, deviation remains high throughout the seasons. Generic hidden variable (figure 8) seems to mainly explain the changes observed in both phyto- and zooplankton. The anomaly of 2008 recognized in generic hidden variable can be detected in zooplanktonic hidden variable as well.

3.2.2. Autoregressive hidden Markov model

Compared to the simple HMM above, the autoregressive model showed much larger variation in the values of generic hidden variable. The clear trend observed in the simple HMM (Figure 8.) is not seen in the more complex model (figure 10).

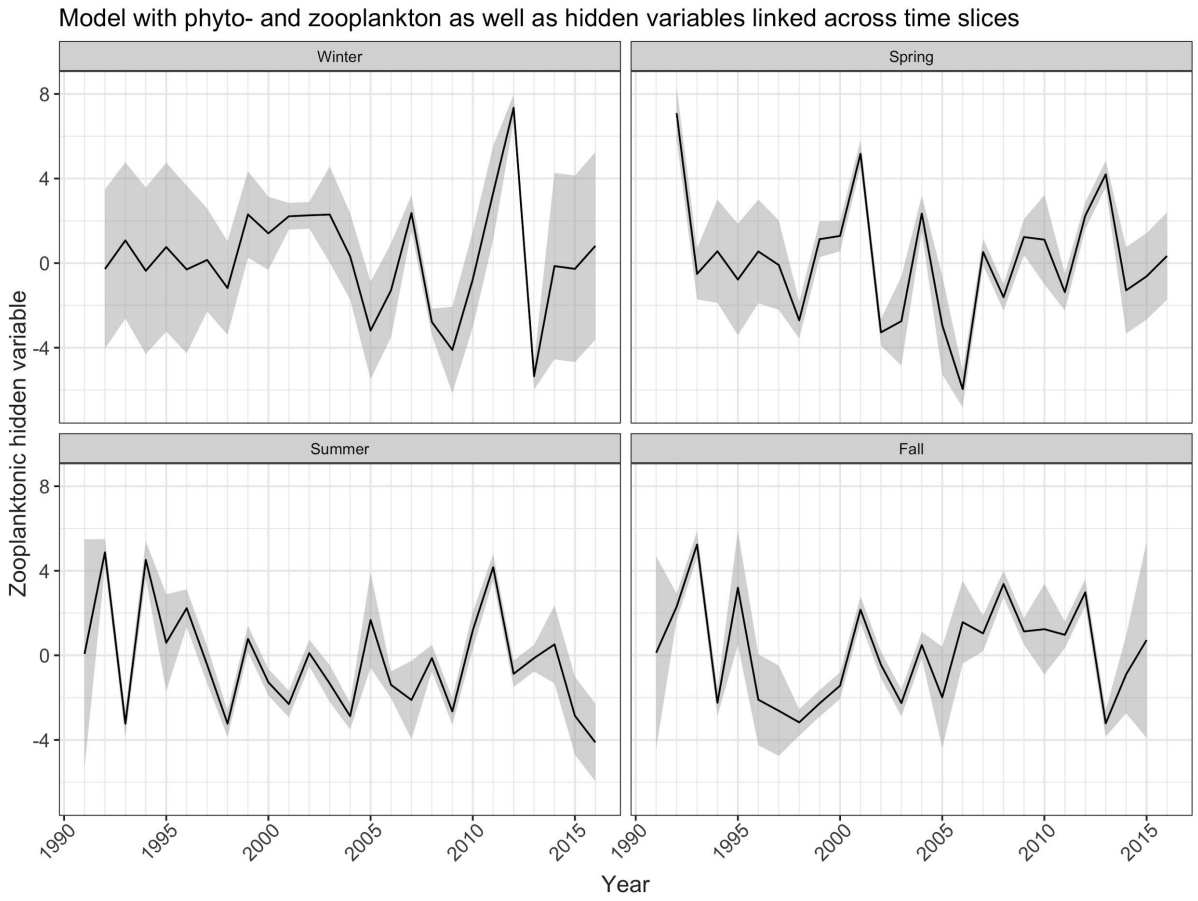

Figure 10. The mean value of generic hidden variable in autoregressive model. In addition to the hidden variables, also the plankton variables are causally linked between time slices.

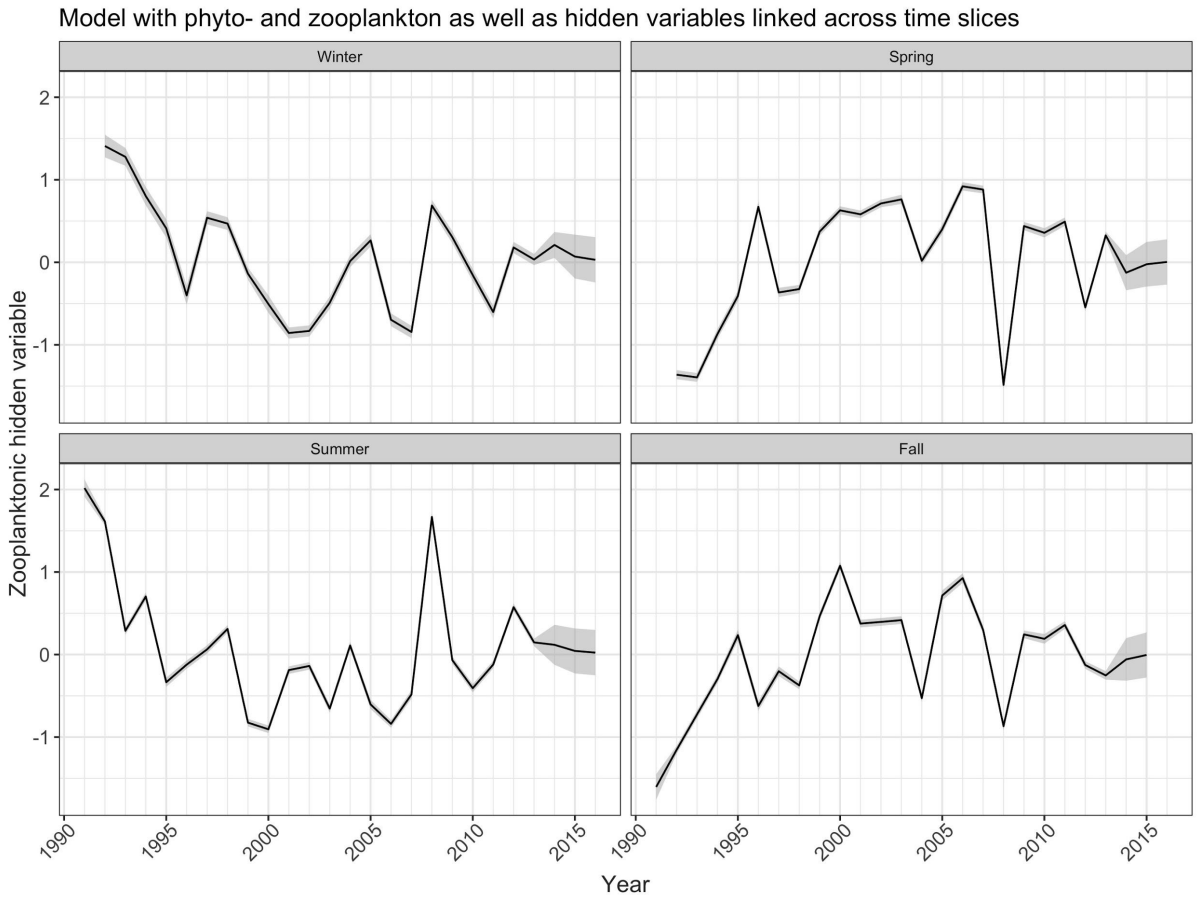

Figure 11. The zooplanktonic hidden variable in the autoregressive model. The deviation is very modest compared to other hidden variables.

Compared to fluctuation in Zooplanktonic HV of simple HMM, the scale of changes is much smaller in the autoregressive model (Figure 11) and the standard deviation remains remarkably low. After the initial high start in winter and summer, and low start of spring and fall, the mean value seems to settle near a value of 0. Some events, such as sudden increase in zooplankton biomass in 2008 can be seen in both the hidden variables of simple HMM (Figures 8 and 9) but caught only by the zooplanktonic HV in the autoregressive model (Figure 11).

3.3. Predictions

Predictions made within the model recognized large-scale trends, but mainly failed to acknowledge sudden changes (Figure 12). Predictions made without any supportive data failed to forecast reliably changes in zooplankton community (Figure 13). According to table of log-likelihoods of the two models, the simple hidden Markov model fared better in both occasions.

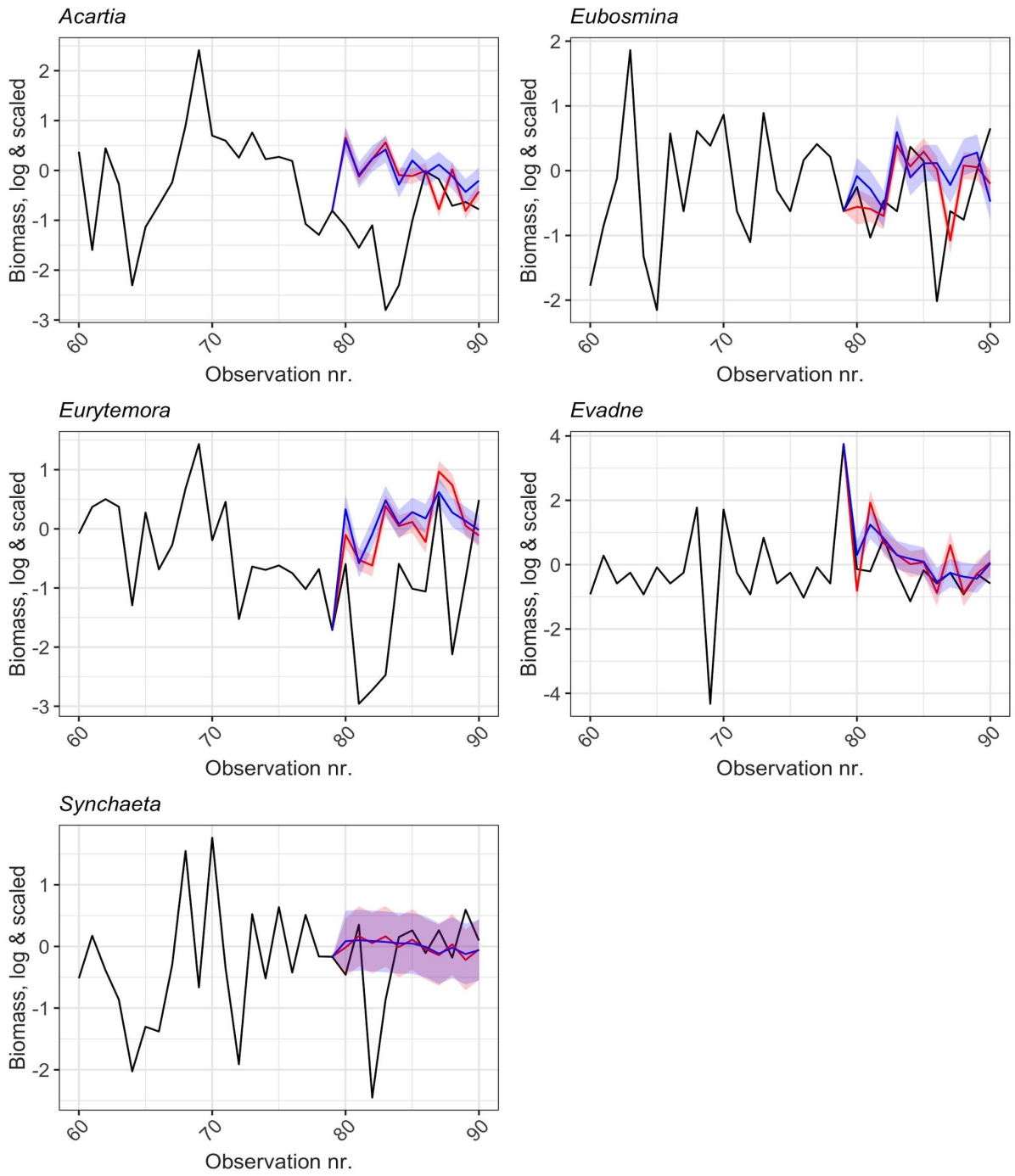

Figure 12. The inference of zooplankton biomasses within the model for the last 12 observations, or 3 years. Blue line represents the simple hidden Markov model, red line the autoregressive model. Predictions determine how well the model could predict the zooplanktonic variables given the environmental and phytoplanktonic data being available.

In inferences within the data set the autoregressive model made predictions with smaller deviation. However, according to log-likelihoods these assumptions were not as accurate as the ones made by the simpler hidden Markov model (Table 2).

Table 2. Log-likelihoods of the two predictive models with phytoplanktonic and environmental observational data available. The predictions were made for the last 3 years, or 12 seasons.

Species	Autoregressive hidden Markov model	Simple hidden Markov model
Acartia	-5.00	-2.03
Eubosmina	-6.55	-3.20
Eurytemora	-6.26	-1.31
Evadne	-12.05	-6.78
Synchaeta	8.96	13.96

The predictions made without any additional data navigate quickly towards a mean value of 0 (figure 13.). As such, the possible correlations between predicted values and observed values seem to be coincidental.

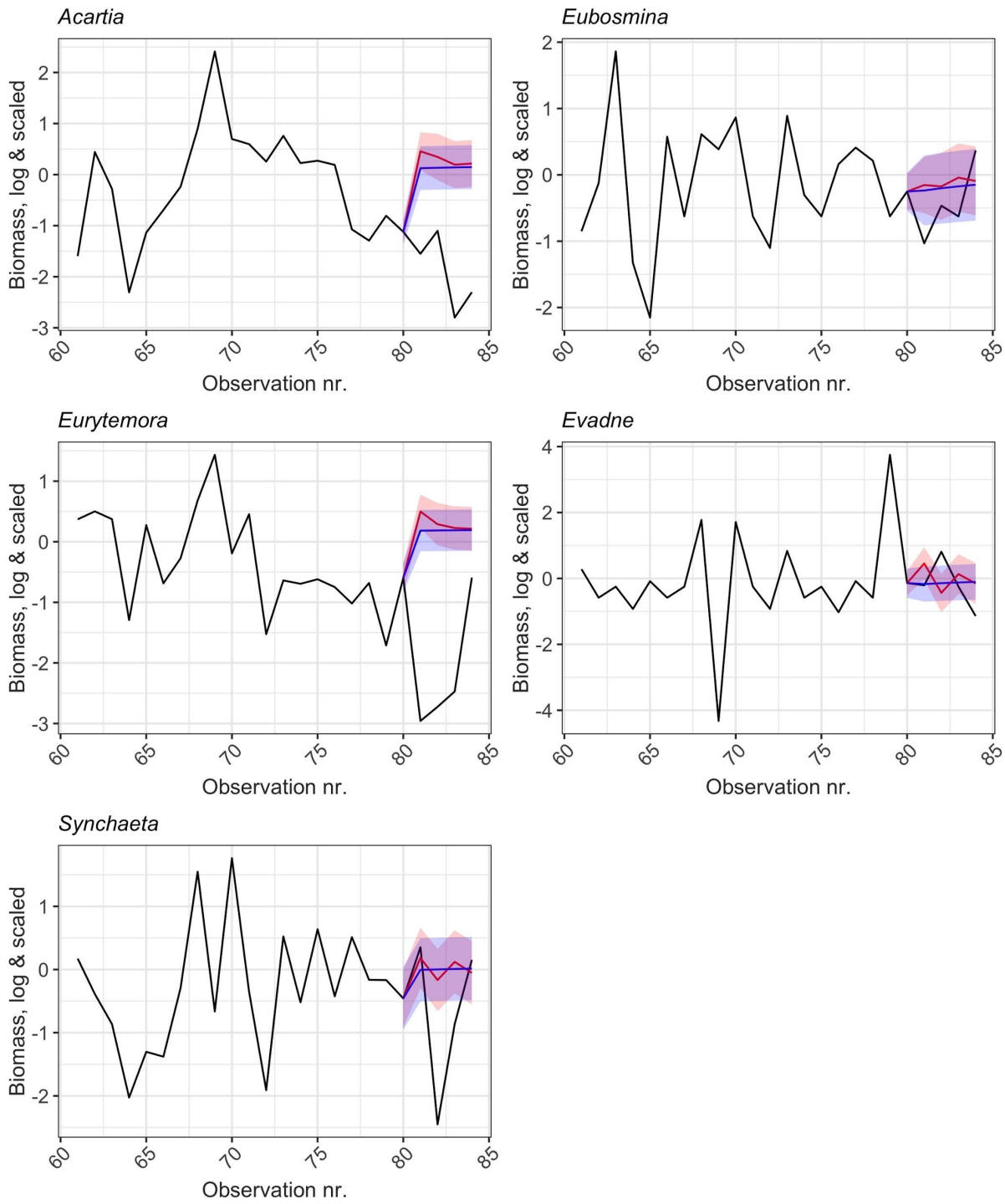

Figure 13. The predictions made without other observational variables. Blue line represents simple HMM, red line autoregressive HMM.

Table 3. Log-likelihoods of the two predictive models with no observational data available. The predictions were made for one year, or 4 seasons.

Species	Autoreg. Model 1 year prediction	Simple HMM 1 year prediction
Acartia	5.32	14.93
Eubosmina	6.29	10.87
Eurytemora	3.96	20.27
Evadne	-1.00	9.83
Synchaeta	3.50	13.97

As can be seen in Table 3, the log-likelihoods are mainly positive values. As the likelihood is the outcome of the density assessed at the observations, high positive values indicate small dispersion parameters. This can be observed in Figure 13 as well.

Table 3. Correlations between yearly fish catches in the Archipelago Sea (Jul-Sep) and hidden variables of summer season (Jun-Aug) from the different models.

Model and Hidden Variable	Herring	Whitefish	Roach	Ide	Zander	Perch	Pike
Autoreg. Generic HV	0.16	0.04	-0.30	0.05	0.03	-0.12	0.03
Autoreg. Zoopl. HV	0.30	-0.09	**-0.39***	-0.19	-0.25	-0.38	0.15
Simple HMM Generic HV	-0.14	**0.61*****	0.08	0.30	**0.46***	0.33	**0.72********
Simple HMM Zoopl. HV	0.26	-0.25	-0.20	0.00	-0.18	-0.26	-0.05

**** Significant at level p < .0001 *** p < .001 ** p < .01 * p < .05

The Pearson correlation of fish data discovered some significant positive correlations with the generic hidden variable of simple HMM. With other hidden variables the correlations were mainly non-significant.

4. Discussion

4.1. Performance of different models

The simple HMM performed better in inferences with, as well as without, additional observational data (Figures 8 & 9) and I'm going to focus on these results in further analyses. This outcome is similar to the result of Maldonado et al. (2019), in which the simplest model applying naïve Bayes inference made the most accurate predictions. It is noteworthy however, that in my study the results of the different models varied significantly and the reasons behind these differences deserve a closer examination. From the autoregressive model it's hard to isolate any meaningful trends, whereas the simpler HMM shows clear downward tendency throughout the study period. No meaningful trend can be seen in the simple HMM zooplanktonic HV, whereas the zooplanktonic HV in autoregressive model is less noisy and shows a steadier trajectory. This might signify that in the autoregressive model most of the deviation and uncertainty is caught in the generic hidden variable, whereas in simple HMM the uncertainty seems to have transferred to zooplanktonic HV.

The inference of zooplankton genera worked reasonably well when the rest of the observational data was available (Figure 12, Table 2). This might signify that the dynamics of zooplankton community could be inferred moderately well from lower level food web data through process-based premises. However, in predictions, where no environmental or phytoplankton data were available, the inference worked poorly (Figure 13, Table 3.). This was somewhat expected, as the Bayesian model was built in a way which requires additional information as input for accurate inference.

According to log-likelihoods (table 2.), the simpler model performed better in both accounts. Regarding the superiority of the simple model, there might be several reasons for this. Firstly, in the simple HMM the observations between different time slices are not directly linked, but rather only through hidden variables. As the lifespan of plankton species are relatively short, the seasonal time frame might be too long to accurately describe real interactions within the community. Possibly inaccurate links in turn might create spurious correlations and increase the amount of noise within the model. As the direct causal links between observations were dropped, the accuracy of the model increased. The links between hidden variables might increase the model performance as well, as large-scale

environmental trends usually happen in much longer time spans compared to plankton life cycles (Carlotti and Poggiale 2010).

Another issue that creates difficulties in such large-scale models are the complex, and often not so well-known, interactions within and between different plankton groups (Carlotti and Poggiale 2010; Rose et al. 2010). One concrete example in my models is *Mesodinium rubrum*, the most abundant species of Litostomatea, which is a mixotrophic species. Their growth is possible by using smaller phytoplankton (e.g. cryptophytes) as energy source or by photosynthesis (Gustafson Jr et al. 2000), representing an example of phytoplankton species hard to categorize for modelling purposes. The abundance of this species varies significantly between different years and thus their diet alone might have substantial effect on other groups in the same ecological niche. As the causal links created by expert evaluation have a significant weight, the accuracy of the model suffers along with compromises that must be made. Whether a more complex structure would give more accurate results remains to be uncovered.

Third larger topic that needs to be addressed is the effect of consumed resources compared to observed resources in the samples. As my model relies on bottom-up theory, this paradox is relevant throughout the food web. Nutrients are the single most decisive bottom-up indicator of phytoplankton biomass. During growth season the observed amount in the samples stays minimal as phytoplankton actively utilize all the available nutrients. If the nutrient levels stay low while at the same time phytoplankton biomass varies significantly, the correlation between nutrients and phytoplankton is not expected to be satisfactory. The same fallacy applies to phytoplankton - zooplankton interactions. The autoregressive dynamic model addresses this issue by taking into account the size of previous stock, but as can be observed in the hidden variables of autoregressive model, the time gap between observations would need to be considerably shorter in order to get reliable results.

4.2. Implications of the results

A clear downward trend was noticeable in the simple hidden Markov model (figure 8), but a regime shift couldn’t be detected based on my models. The hidden variable in the model reacts mainly to changes between different variables, as illustrated in figure 3. As data-based model fitting procedures assume that the interactions between model variables remain unchanged throughout the study period, the HV adapts to possible changes and tries to maximize the fit of the data set. A trend in the HV time series suggests that the

relationship between the observed variables in the model have changed. In case there were no change in these relationships, the HV time series would just fluctuate randomly. This random fluctuation could also imply some false presumption or causal link in the structure of the Bayesian model, which would further distort the results.

When HV is analyzed in context with observational data, it is clear that both plankton groups have distinct, if opposite, trends throughout the study period (Figures 6 & 7). In eutrophicated conditions the general abundance of phytoplankton tend to increase and zooplankton decrease (e.g. McQueen et al. 1989; Suikkanen et al. 2007), and thus the trend caught by the hidden variable might signify increased eutrophication of the Archipelago Sea between 1991 and 2014. This trend can be witnessed in all of the results (Figures 6,7,8 and Table 3). On the other hand, Suikkanen et al. (2013) concluded that the observed changes in phytoplankton community were mainly driven by changes in temperature and didn't directly correlate with trends in nutrient levels. In other studies of the Bothnian sea the eutrophication status has been observed to increase between 1990-2005 and decrease between 2005-2012, but the confidence of the result assessment has declined during this time period (Andersen et al. 2017).

These changes can be observed within the different plankton groups as well. Cladocerans are filter feeders generally consuming smaller prey, whereas copepods hunt larger prey and are able to better select their target (Porter 1973; Becker et al. 2004). Thus, the possible alterations in phytoplankton community are likely to reflect to higher trophic levels as well. One implication of the change within zooplankton community can be observed in summer composition of the zooplankton genera. During 1990s cladoceran *Eubosmina* dominated the community, whereas the dominance switched to favor copepod *Eurytemora* during the 2000s. After 2010 both of these genera declined considerably (Figure 7).

At the same time as the dominance of zooplankton genera switched from *Eubosmina* to *Eurytemora*, from 2000 to 2010 the summertime biomass in the samples was mainly dominated by cyanophyta (Figure 6). During summer, grazing pressure created by zooplankton is high and most available resources are consumed quickly. As discussed earlier zooplankton tend to avert cyanophyta in their diet whenever possible (e.g. Porter 1973; Ger et al. 2014). During summer the growth of zooplankton community, especially copepods, might be food-limited in the Baltic Sea (Karjalainen et al. 2007). As they actively select other resources than cyanobacteria, large copepod population might even encourage the cyanobacterial blooms by reducing the abundance of their competitors (Karjalainen et al. 2007). Eutrophicated conditions have been further noted to benefit species favored by

microbial loop, including Litostomatea and selectively feeding copepods (Karjalainen et al. 2007), which can be acknowledged in my study as well (Figures 6 & 7).

Hence, it is reasonable to assume that even though a clear increase of cyanophyta can be seen, it's harder to deduce the real biomasses of other phytoplankton classes as they are consumed more aggressively, and their turnover rate is likely to be much higher. If the abundance of cyanophyta within the phytoplankton community has increased in the expense of other phylums, this could be one reason for the decline of zooplankton biomass, especially during the last years of the study period.

On the other hand, phytoplankton species composition is merely one factor when determining the composition of zooplankton community. Ojaveer et al. (1998) conclude that the dynamics of copepod species are highly influenced by environmental variables, especially salinity. The abundance of copepods on the other hand correlate positively with temperature changes (Viitasalo et al. 1994). As these are both bottom-up regulators for copepod biomass, it's challenging to single out the effects of separate components. Besides determining independently zooplankton community interactions, these aforementioned variables might have cumulative responses as well.

To examine the food web further, the mean value of summertime hidden variables was compared to annual July - September fish catches from the area (Table 3.). Changes in fish catches could roughly correspond to dynamics of the fish populations if the exploitation of the stocks stays more or less constant. The correlations between different hidden variables and fish catches were mainly insignificant, but the generic hidden variable of simple HMM showed some significant correlations. Rather than trying to infer direct causality between HV and the fish stocks, the correlations between HV and the fish catches should be interpreted as an indicator of change. Whitefish and pike stocks have been observed to suffer from eutrophication, whereas zander favors eutrophic conditions (Willemsen 1980; Olin et al. 2002; Winkler 2002). As all of these species tend to react to eutrophication in one way or another, the correlation observed between Hidden variable and these catches might support the theory of widespread effects of the increased eutrophication in the area.

Previous studies have established a link between herring catches and larger zooplankton taxa (Flinkman et al. 1992; Arrhenius 1997; Flinkman et al. 1998), and abundance of roach and other cyprinids could be interpreted as a sign of eutrophication (Olin et al. 2002; Tammi et al. 1999), but no correlations were found between the catch sizes of these fish species and hidden variables. It is noteworthy that the catches are not necessarily directly linked to

fish stock sizes and thus possible correlations, or the lack thereof, need to be treated with caution. This is especially true with species that are considered as low valued ones, such as roach and ide, with low economic value in human consumption.

4.3. Further research topics and conclusions

When deducing natural phenomena from samples, the effect of chance needs to be considered. Zooplankton are known to migrate both vertically, as well as horizontally (DeMott and Kerfoot 1982; Ojaveer et al. 1998). Hence the composition of species in the sample might vary considerably depending on the date, and time of the day. Another variable to be regarded is the phytoplankton's, especially cyanophytes', tendency to flocculate and form colonies (De Bernardi and Giussani 1990). Depending on the odds, the number of individuals caught in the sample might vary highly as the sampling gear might, or might not, pass through a colony.

Yet another challenge is the short lifespan of phytoplankton and the brevity of blooms. During spring blooms, the diatom and dinophyte abundance might change considerably within days (Karjalainen et al. 2007). As the sampling in this study was done on average monthly, many of the nuances are likely to remain undiscovered. All these concerns are lessened by using mean values of several observations and thus decreasing the risk of coarse outliers. This method is a double-edged sword, as some meaningful anomalies might go undetected when blended in with other observations. When considering plankton in models, ideally the sampling would be done more frequently. As the sampling and analysis of the samples must be done manually and is rather costly, with current technology this hardly is feasible.

The model itself could be further enhanced by addressing the issues discussed above. In addition to these considerations, the variables could be discretized. This approach might detect some complex and non-linear changes that go easily undetected with continuous scales (Myllymäki et al. 2002). Results as few distinct states might be easier to analyze and tweaking the model would be faster, and changes thus easier to implement. The potential discretization would have to be done with extreme care as the variation and scales between different observations differ significantly and the number of bins used in discretization have a significant impact on the results as well (Uusitalo 2007).

The Bayesian inference has potential in the modelling of plankton food web, but attention must be paid especially concerning the temporal scale. The challenges presented earlier should also be kept in mind when exploring the Bayesian models further. This said, dynamic Bayesian networks present capability to detect underlying larger trends and simplify hugely complex systems in a reproducible method. In obscure models it becomes increasingly important to be able to quantify uncertainty and for that end Bayesian inference fills a purpose. Compared to some other machine learning methods, such as neural networks and random forests, Bayesian networks provide transparent and easy-to-approach method to model aquatic ecosystems. The hidden variables studied in my models were able to pick up clear trends, but plenty of work remains in the exploration of the interactions within both of the plankton communities.

5. Acknowledgements

This work was done in collaboration with Finnish Environment Institute, the BlueAdapt project (the Strategic Research Council of Academy of Finland, Contract No. 312650) and BONUS BLUEWEBS project. I'd like to thank Laura Uusitalo, Heikki Peltonen and Veera Norros for inspiring me to tackle the Bayesian networks headfirst and guiding me through it all. I'd also like to thank Allan Tucker for providing guidance in model building and Harri Kuosa for helping me understand the dynamics within plankton communities. I'd also like to express my gratitude for Katja Mäkinen, who kindly provided the zooplankton data for this study. I'm also grateful for all the people and institutes who have gathered and analyzed the data used in this thesis and allowed me to utilize it in this work. Without your effort, this thesis wouldn't have been possible.

6. References

Aguilera PA, Fernández A, Fernández R, Rumí R, Salmerón A. 2011. Bayesian networks in environmental modelling. Environmental Modelling & Software. 26(12):1376-88.

Alheit J, Möllmann C, Dutz J, Kornilovs G, Loewe P, Mohrholz V, Wasmund N. 2005. Synchronous ecological regime shifts in the central baltic and the north sea in the late 1980s. ICES J Mar Sci. 62(7):1205-15.

Andersson A, Hajdu S, Haecky P, Kuparinen J, Wikner J. 1996. Succession and growth limitation of phytoplankton in the gulf of bothnia (baltic sea). Mar Biol. 126(4):791-801.

Andersen JH, Carstensen J, Conley DJ, Dromph K, Fleming-Lehtinen V, Gustafsson BG, Josefson AB, Norkko A, Villnäs A, Murray C. 2017. Long-term temporal and spatial trends in eutrophication status of the baltic sea. Biological Reviews. 92(1):135-49.

Arrhenius F. 1997. Top-down controls by young-of-the-year herring (clupea harengus) in the northern baltic proper. Forage fishes in marine ecosystems Fairbanks, Alaska: University of Alaska Sea Grant College Program AK-SG-97-01. 77 p.

Becker C, Feuchtmayr H, Brepohl D, Santer B, Boersma M. 2004. Differential impacts of copepods and cladocerans on lake seston, and resulting effects on zooplankton growth. Hydrobiologia. 526(1):197-207.

Bergquist AM, Carpenter SR, Latino JC. 1985. Shifts in phytoplankton size structure and community composition during grazing by contrasting zooplankton assemblages 1. Limnol Oceanogr. 30(5):1037-45.

Borsuk ME, Stow CA, Reckhow KH. 2004. A bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. Ecol Model. 173(2-3):219-39.

Boynton WR, Kemp WM, Keefe CW. 1982. A comparative analysis of nutrients and other factors influencing estuarine phytoplankton production. In: Estuarine comparisons. Elsevier. 69 p.

Brett M, Müller-Navarra D. 1997. The role of highly unsaturated fatty acids in aquatic foodweb processes. Freshwater Biol. 38(3):483-99.

Brett MT, Goldman CR. 1996. A meta-analysis of the freshwater trophic cascade. Proceedings of the National Academy of Sciences. 93(15):7723-6.

Carlotti F, Poggiale J. 2010. Towards methodological approaches to implement the zooplankton component in “end to end” food-web models. Prog Oceanogr. 84(1-2):20-38.

Chen SH, Pollino CA. 2012. Good practice in bayesian network modelling. Environmental Modelling & Software. 37:134-45.

Clark JS. 2005. Why environmental scientists are becoming bayesians. Ecol Lett. 8(1):2-14.

Clauser JF, Horne MA, Shimony A, Holt RA. 1969. Proposed experiment to test local hidden-variable theories. Phys Rev Lett. 23(15):880.

Daewel U, Peck MA, Kuehn W, ST. John MA, Alekseeva I, Schrum C. 2008. Coupling ecosystem and individual-based models to simulate the influence of environmental variability on potential growth and survival of larval sprat (sprattus sprattus L.) in the north sea. Fish Oceanogr. 17(5):333-51.

De Bernardi Rd, Giussani G. 1990. Are blue-green algae a suitable food for zooplankton? an overview. Hydrobiologia. 200(1):29-41.

DeMott WR, Kerfoot WC. 1982. Competition among cladocerans: Nature of the interaction between bosmina and daphnia. Ecology. 63(6):1949-66.

Dempster AP, Laird NM, Rubin DB. 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (Methodological). 39(1):1-22.

DeYoung B, Heath M, Werner F, Chai F, Megrey B, Monfray P. 2004. Challenges of modeling ocean basin ecosystems. Science. 304(5676):1463-6.

Engström J, Koski M, Viitasalo M, Reinikainen M, Repka S, Sivonen K. 2000. Feeding interactions of the copepods eurytemora affinis and acartia bifilosa with the cyanobacteria nodularia sp. J Plankton Res. 22(7):1403-9.

Eppley RW. 1972. Temperature and phytoplankton growth in the sea. Fish Bull. 70(4):1063-85.

Flinkman J, Vuorinen I, Aro E. 1992. Planktivorous baltic herring (clupea harengus) prey selectively on reproducing copepods and cladocerans. Can J Fish Aquat Sci. 49(1):73-7.

Flinkman J, Aro E, Vuorinen I, Viitasalo M. 1998. Changes in northern baltic zooplankton and herring nutrition from 1980s to 1990s: Top-down and bottom-up processes at work. Mar Ecol Prog Ser. 165:127-36.

Fulton EA. 2010. Approaches to end-to-end ecosystem models. J Mar Syst. 81(1-2):171-83.

Fulton EA, Smith AD, Johnson CR. 2003. Effect of complexity on marine ecosystem models. Mar Ecol Prog Ser. 253:1-16.

Ger KA, Hansson L, Lürling M. 2014. Understanding cyanobacteria-zooplankton interactions in a more eutrophic world. Freshwat Biol. 59(9):1783-98.

Gilbert JJ, Williamson CE. 1983. Sexual dimorphism in zooplankton (copepoda, cladocera, and rotifera). Annu Rev Ecol Syst. 14(1):1-33.

Grossman D, Domingos P. 2004. Learning bayesian network classifiers by maximizing conditional likelihood. Proceedings of the twenty-first international conference on machine learning. 46 p.

Gustafson Jr DE, Stoecker DK, Johnson MD, Van Heukelem WF, Sneider K. 2000. Cryptophyte algae are robbed of their organelles by the marine ciliate mesodinium rubrum. Nature. 405(6790):1049.

HELCOM. 1988. Guidelines for the baltic monitoring programme for the third stage; part D biological determinants. Baltic sea environment proceedings. 166 p.

Jakeman AJ, Letcher RA, Norton JP. 2006. Ten iterative steps in development and evaluation of environmental models. Environmental Modelling & Software. 21(5):602-14.

Karjalainen M, Engström-Öst J, Korpinen S, Peltonen H, Pääkkönen J, Rönkkönen S, Suikkanen S, Viitasalo M. 2007. Ecosystem consequences of cyanobacteria in the northern baltic sea. AMBIO: A Journal of the Human Environment. 36(2):195-203.

Kivi K, Kaitala S, Kuosa H, Kuparinen J, Leskinen E, Lignell R, Marcussen B, Tamminen T. 1993. Nutrient limitation and grazing control of the baltic plankton community during annual succession. Limnol Oceanogr. 38(5):893-905.

Klausmeier CA, Litchman E, Levin SA. 2004. Phytoplankton growth and stoichiometry under multiple nutrient limitation. Limnol Oceanogr. 49(4part2):1463-70.

Kornilovs G, Sidrevics L, Dippner JW. 2001. Fish and zooplankton interaction in the central baltic sea. ICES J Mar Sci. 58(3):579-88.

Kozlowsky-Suzuki B, Karjalainen M, Lehtiniemi M, Engström-Öst J, Koski M, Carlsson P. 2003. Feeding, reproduction and toxin accumulation by the copepods acartia bifilosa and eurytemora affinis in the presence of the toxic cyanobacterium nodularia spumigena. Mar Ecol Prog Ser. 249:237-49.

Kuosa H, Kivi K. 1989. Bacteria and heterotrophic flagellates in the pelagic carbon cycle in the northern baltic sea. Marine Ecology Progress Series.Oldendorf. 53(1):93-100.

Landuyt D, Broekx S, D'hondt R, Engelen G, Aertsens J, Goethals PL. 2013. A review of bayesian belief networks in ecosystem service modelling. Environmental Modelling & Software. 46:1-11.

Maldonado AD, Uusitalo L, Tucker A, Blenckner T, Aguilera PA, Salmerón A. 2019. Prediction of a complex system with few data: Evaluation of the effect of model structure and amount of data with dynamic bayesian network models. Environmental Modelling & Software. 118:281-97.

McDonald KS, Ryder DS, Tighe M. 2015. Developing best-practice bayesian belief networks in ecological risk assessments for freshwater and estuarine ecosystems: A quantitative review. J Environ Manage. 154:190-200.

McQueen DJ, Johannes MR, Post JR, Stewart TJ, Lean DR. 1989. Bottom-up and top-down impacts on freshwater pelagic community structure. Ecol Monogr. 59(3):289-309.

Müller-Navarra DC, Brett MT, Park S, Chandra S, Ballantyne AP, Zorita E, Goldman CR. 2004. Unsaturated fatty acid content in seston and tropho-dynamic coupling in lakes. Nature. 427(6969):69.

Murphy KP. 2012. Machine learning: A probabilistic perspective. MIT press.

Murphy KP, Russell S. 2002. Dynamic bayesian networks: Representation, inference and learning.

Myllymäki P, Silander T, Tirri H, Uronen P. 2002. B-course: A web-based tool for bayesian and causal data analysis. Int J Artif Intell Tools. 11(03):369-87.

Neumann T, Schernewski G. 2005. An ecological model evaluation of two nutrient abatement strategies for the baltic sea. J Mar Syst. 56(1-2):195-206.

Ojaveer E, Lumberg A, Ojaveer H. 1998. Highlights of zooplankton dynamics in estonian waters (baltic sea). ICES J Mar Sci. 55(4):748-55.

Olin M, Rask M, Ruuhljärvi J, Kurkilahti M, Ala-Opas P, Ylönen O. 2002. Fish community structure in mesotrophic and eutrophic lakes of southern finland: The relative abundances of percids and cyprinids along a trophic gradient. J Fish Biol. 60(3):593-612.

Pérez-Miñana E. 2016. Improving ecosystem services modelling: Insights from a bayesian network tools review. Environmental Modelling & Software. 85:184-201.

Persson J, Vrede T. 2006. Polyunsaturated fatty acids in zooplankton: Variation due to taxonomy and trophic position. Freshwat Biol. 51(5):887-900.

Porter KG. 1973. Selective grazing and differential digestion of algae by zooplankton. Nature. 244(5412):179.

Robinson JW, Hartemink AJ. 2010. Learning non-stationary dynamic bayesian networks. Journal of Machine Learning Research. 11(Dec):3647-80.

Rose KA, Allen JI, Artioli Y, Barange M, Blackford J, Carlotti F, Cropp R, Daewel U, Edwards K, Flynn K. 2010. End-to-end models for the analysis of marine ecosystems: Challenges, issues, and next steps. Marine and Coastal Fisheries. 2(1):115-30.

Rudstam LG, Aneer G, Hildén M. 1994. Top-down control in the pelagic baltic ecosystem. Dana. 10:105-29.

Rudstam LG, Hansson S, Johansson S, Larsson U. 1992. Dynamics of planktivory in a coastal area of the northern baltic sea. Marine Ecology Progress Series.Oldendorf. 80(2):159-73.

Schuster-Böckler B, Bateman A. 2007. An introduction to hidden markov models. Current Protocols in Bioinformatics. 18(1):A. 3A. 1,A. 3A. 9.

Sommer U, Sommer F. 2006. Cladocerans versus copepods: The cause of contrasting top–down controls on freshwater and marine phytoplankton. Oecologia. 147(2):183-94.

Sommer U, Stibor H. 2002. Copepoda–Cladocera–Tunicata: The role of three major mesozooplankton groups in pelagic food webs. Ecol Res. 17(2):161-74.

Sommer U, Sommer F, Santer B, Jamieson C, Boersma M, Becker C, Hansen T. 2001. Complementary impact of copepods and cladocerans on phytoplankton. Ecol Lett. 4(6):545-50.

Strom SL, Macri EL, Olson MB. 2007. Microzooplankton grazing in the coastal gulf of alaska: Variations in top-down control of phytoplankton. Limnol Oceanogr. 52(4):1480-94.

Suikkanen S, Laamanen M, Huttunen M. 2007. Long-term changes in summer phytoplankton communities of the open northern baltic sea. Estuar Coast Shelf Sci. 71(3-4):580-92.

Suikkanen S, Pulina S, Engström-Öst J, Lehtiniemi M, Lehtinen S, Brutemark A. 2013. Climate change and eutrophication induced shifts in northern summer plankton communities. PLoS One. 8(6).

Tammi J, Lappalainen A, Mannio J, Rask M, Vuorenmaa J. 1999. Effects of eutrophication on fish and fisheries in finnish lakes: A survey based on random sampling. Fish Manage Ecol. 6(3):173-86.

Trifonova N, Karnauskas M, Kelble C. 2019. Predicting ecosystem components in the gulf of mexico and their responses to climate variability with a dynamic bayesian network model. PloS One. 14(1):e0209257.

Trifonova N, Maxwell D, Pinnegar J, Kenny A, Tucker A. 2017. Predicting ecosystem responses to changes in fisheries catch, temperature, and primary productivity with a dynamic bayesian network model. ICES J Mar Sci. 74(5):1334-43.

Trifonova N, Kenny A, Maxwell D, Duplisea D, Fernandes J, Tucker A. 2015. Spatio-temporal bayesian network models with latent variables for revealing trophic dynamics and functional networks in fisheries ecology. Ecological Informatics. 30:142-58.

Tucker A, Liu X. 2004. A bayesian network approach to explaining time series with changing structure. Intelligent Data Analysis. 8(5):469-80.

Uusitalo L. 2007. Advantages and challenges of bayesian networks in environmental modelling. Ecol Model. 203(3-4):312-8.

Uusitalo L, Tomczak MT, Müller-Karulis B, Putnis I, Trifonova N, Tucker A. 2018. Hidden variables in a dynamic bayesian network identify ecosystem level change. Ecological Informatics. 45:9-15.

Viitasalo M. 1992. Mesozooplankton of the gulf of finland and northern baltic proper a review of monitoring data. Ophelia. 35(2):147-68.

Viitasalo M, Flinkman J, Viherluoto M. 2001. Zooplanktivory in the baltic sea: A comparison of prey selectivity by clupea harengus and mysis mixta, with reference to prey escape reactions. Mar Ecol Prog Ser. 216:191-200.

Viitasalo M, Katajisto T, Vuorinen I. 1994. Seasonal dynamics of acartia bifilosa and eurytemora affinis (copepoda: Calanoida) in relation to abiotic factors in the northern baltic sea. In: Ecology and morphology of copepods. Springer. 415 p.

Walters C, Christensen V, Pauly D. 1997. Structuring dynamic models of exploited ecosystems from trophic mass-balance assessments. Rev Fish Biol Fish. 7(2):139-72.

Willemsen J. 1980. Fishery-aspects of eutrophication. Hydrobiological Bulletin. 14(1-2):12-21.

Winkler HM. 2002. Effects of eutrophication on fish stocks in baltic lagoons. In: Baltic coastal ecosystems. Springer. 65 p.

Zhang H, Jiang L, Su J. 2005. Hidden naive bayes. Aaai. 919-924.

Österblom H, Hansson S, Larsson U, Hjerne O, Wulff F, Elmgren R, Folke C. 2007. Human-induced trophic cascades and ecological regime shifts in the baltic sea. Ecosystems. 10(6):877-89.