SYNTHESIS AND MODELLING OF ZOOPLANKTON AT PAN-ARCTIC SCALES

By

Imme Rutzen, Dipl.-Biol.

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy
in
Marine Biology

University of Alaska Fairbanks
May 2017

APPROVED:

Russell R. Hopcroft, Committee Chair
Falk Huettmann, Committee Member
Kenneth O. Coyle, Committee Member
Rolf R. Gradinger, Committee Member
Thomas J. Weingartner, Committee Member
Sarah M. Hardy, Chair
Department of Marine Biology
S. Bradley Moran, Dean
College of Fisheries and Ocean Sciences
Michael A. Castellini, *Dean of the Graduate School*

Abstract

Zooplankton are an important link between primary producers and higher trophic levels. They are sensitive indicators of change in the Arctic ecosystem due to their relatively short lifespan. To date, the greatest impediment to detect changes in the Arctic zooplankton community at pan-Arctic scales is the absence of a reference baseline. To contribute to baseline data, I taxonomically analyzed zooplankton samples from the Canada Basin collected during August and September of 2003-2006. Over 50 taxonomic groups were identified, but copepods dominated abundance and biomass. Non-copepod abundance was dominated by larvaceans, while non-copepod biomass was dominated by chaetognaths. I applied multivariate analysis to look at patterns in community similarity, finding a tendency to separate the years sampled. The sample analysis served the larger goal of my research: collation of zooplankton data from online databases, reports, papers, and through scientific cooperation with scientists throughout the Arctic. In total, 13,014 zooplankton samples were assembled, containing over 200,000 individual taxonomic records spanning the period from 1921 to 2012. I also assembled 25 environmental layers for variables of possible influence on zooplankton distribution. Using these data, I employed the Geographic Information System ArcMap, as well as the data mining approaches TreeNet and RandomForests to predict the climatological mean distribution and abundance of seven ecologically and numerically important epipelagic copepod species (*Calanus finmarchicus*, *C. glacialis*, *C. hyperboreus*, *Metridia longa*, *M. pacifica*, *N. cristatus*, and *P. glacialis*) on a pan-Arctic scale from 60° - 90°N. The model predicted the overall distribution and abundance characteristics of each species well, but it also predicted potential niches for these or sibling species in areas where they are known to be absent. The model correctly associated species advected to the Arctic with corresponding water masses, while Arctic endemic species were more strongly associated with geographic variables. Continued assimilation of new data, plus rescuing and consolidating older datasets, are critical pathways toward both enhancing this baseline, and building the observational time-series necessary for studying changes in the Arctic zooplankton community.

Table of Contents

List of Figures

List of Tables

Page

List of Appendices

Page

Acknowledgements

I would like to thank my committee chair Russell R. Hopcroft for his continuing support throughout this PhD. I thank my committee member Falk Huettmann for being my first contact in Alaska, and for his support and guidance since the first day I arrived. I further thank my committee members Rolf R. Gradinger, and Thomas J. Weingartner for helpful suggestions and input on my project, and I especially thank Kenneth O. Coyle for the VBA lessons, his patience, and his good eye for English grammar.

I would like to acknowledge the funding agencies and sources: The National Science Foundation Award OPP0909571, the University of Alaska Fairbanks (UAF) for a graduate fellowship, the Graduate School of the University of Alaska Fairbanks for providing me with a thesis completion fellowship, the Dr. Donald Hood Memorial Scholarship for Marine Science, and the Census of Marine Life's Arctic Ocean Diversity project.

I am very thankful to all of the scientists that provided data for my project or made them available through online resources.

Furthermore I want to extend many thanks to the fellow graduate students and lab members at the Hopcroft lab, as well as the students in the EWHALE lab at UAF.

Finally, I am very grateful for my family and friends that were very supportive on many different levels throughout the last years.

Chapter 1 General Introduction

Arctic zooplankton are an important link between primary producers and higher trophic levels (Mauchline *et al.*, 1998; Melle *et al.*, 2004). Zooplankton communities consist of a mixture of species endemic to the Arctic that reproduce there, as well as species that are advected from the Pacific and the Atlantic into the Arctic Ocean (Kosobokova & Hirche, 2000), but that are not viable in the Arctic (Wassmann *et al.*, 2015). The relatively short lifecycles of Arctic zooplankton make them ideal organisms to detect changes in their community composition, biomass, and abundance as a response to environmental forcing (Richardson, 2008). Changes in the zooplankton communities ultimately affect higher trophic levels such as fish, whales, and birds (Masatoshi *et al.*, 2000; Karnovsky *et al.*, 2003; Rogachev *et al.*, 2008). Alterations in zooplankton communities may include a shift away from Arctic endemic species to seasonally expatriated species from the Pacific and the Atlantic Ocean, as well as a shift from larger-bodied zooplankton with longer life cycles toward smaller species with shorter life cycles and less lipid content (Falk-Petersen *et al.*, 2007). A northward shift of species has already been observed in the North Sea (Beaugrand *et al.*, 2002), where a shift from a comparably larger, lipid-rich copepod species (*C. finmarchicus*) to a smaller, nutrient-poor species (*C. helgolandicus*) has occurred.

The Arctic Ocean is a nearly landlocked basin and consists of central deep basins (Canada, Makarov, Amundsen, and Nansen Basin) and the adjacent Seas, all of which are seasonally ice covered to some extent. They include the Beaufort, Chukchi, East Siberian, Laptev, Kara, Barents, Norwegian, and the Greenland Seas (Carmack *et al.*, 2006), and arguably Baffin Bay. Nutrient rich waters from the North Pacific enter the Arctic Ocean through Bering Strait and the Chukchi Sea (Woodgate *et al.*, 2005), while the comparatively saline and nutrient poor Atlantic waters enter through the Norwegian and Greenland Seas (Codispoti & Lowman, 1973; Aagaard, 1989; Emery, 2001; Schauer *et al.*, 2002; Carmack *et al.*, 2006; Seidov *et al.*, 2015). Arctic water exits through the Canadian Archipelago and the eastern shelf of Greenland (Aagaard, 1989; Carmack *et al.*, 2006; Seidov *et al.*, 2015). Strong freshwater discharge into the Arctic Ocean occurs through melting sea ice and several major river systems such as the Kolyma, the Lena and the Mackenzie rivers (Shiklomanov *et al.*, 2000; Yamamoto-Kawai *et al.*, 2009). The central Arctic Ocean is covered in multi-year sea ice and the extent changes seasonally, typically with a maximum ice-extent in March and a minimum in September. This sea ice cover has declined rapidly within the past decades (Comiso, 2012) and has recently had a record minimum summer sea ice extent in 2012 (Parkinson & Comiso, 2013). The Arctic has been predicted to be virtually ice free (sea ice extent 1 million km^2) during September within the next 30 to 40 years (Liu *et al.*, 2013).

Changes in sea ice concentration and extent, as well as the timing of ice retreat and formation, have implications for the Arctic ecosystem on many different trophic levels. Not only does the loss of sea ice impact the higher trophic levels directly, but also indirectly, by changing the timing of the phytoplankton blooms and the lower trophic levels dependent upon them. In the Bering Sea, for example, it was observed that if the early sea ice retreat occurred while the water column was still mixed by winter storms, the spring phytoplankton bloom started later, and in relatively warmer water, once the water column stabilized and stratified. If the ice retreat happened late, the upper water column was stabilized by fresh water from the melting sea ice and an early, ice-associated bloom occurred (Hunt *et al.*, 2011). During the early ice retreat scenario, the abundance of larger, lipid-rich species such as *Calanus marshallae* and *Thysanoessa raschii* markedly declined. These alternatives have implications for higher trophic levels such as pollock and influence their survival rate (Coyle *et al.*, 2011; Hunt *et al.*, 2011). In contrast, Ringuette *et al.* (2002) concluded that a reduced sea-ice cover and longer open water season in the Arctic would enhance population growth of large endemic copepods such as *C. glacialis* and *C. hyperboreus* on Arctic shelves. In the Chukchi Sea, both endemic and expatriated large-bodied species appear to be increasing (Ershova *et al.*, 2015), but in the Canada Basin a decrease in densities of species characteristic for the region was suggested and correlated with the freshening of the Beaufort Gyre due to increased sea ice melt (Hunt *et al.*, 2014).

Considering the changes that are already being observed, it is important that we understand the current range of habitat and the abundance of the key endemic and expatriated species if we are to detect shifts in this ecosystem associated with climate change. We can only detect changes in the Arctic ecosystem if we have a solid baseline of environmental and biological data (Nelson *et al.*, 2014). While zooplankton abundance, distribution, and biomass data have been collected by numerous projects and programs for nearly a century, most data remain unused, buried in stacks of reports, or have been lost over time. In recent years, efforts to rescue such datasets from grey literature have begun, including the Archives of the Arctic Seas by the Zoological Institute of the Russian Academy of Sciences, the BioChem database by the Fisheries and Oceans Canada (DFO), the Coastal and Oceanic Plankton Ecology, Production, and Observation Database (COPEPOD) by the National Oceanic and Atmospheric Administration (NOAA), the Arctic Ocean Biodiversity Project (ArcOD), the Ocean Biogeographic Information System (OBIS), the Pacific Marine Arctic Regional Synthesis (PacMARS), and the European data archive PANGAEA.

In this dissertation, I am presenting an Arctic zooplankton data compilation through the collation of datasets from databases, grey literature, and fellow scientists. This dataset was supplemented by the analysis of new samples collected over several years (2003 – 2006) in the Canada Basin where data was sparse. I then applied a Geographic Information System (GIS) and data-mining approach (Yen *et al.*, 2004; Rutzen, 2007; Humphries *et al.*, 2012) to the collated datasets to model the pan-Arctic distribution and abundance of seven key Arctic copepod species. This dissertation contributes to the creation of a pan-Arctic baseline of zooplankton abundance and distribution that lays the foundation for examining how zooplankton communities vary across space and time.

References

Aagaard, K. (1989) A synthesis of the Arctic Ocean circulation. *Rapp. P-v. Réun. Cons. int. Explor. Mer*, **188**, 11-22.

Beaugrand, G., Ibañez, F., Lindley, J. A. and Reid, P. C. (2002) Diversity of calanoid copepods in the North Atlantic and adjacent seas: species associations and biogeography. *Mar. Ecol.: Prog. Ser.*, **232**, 179-195.

Carmack, E., Barber, D., Christensen, J., Macdonald, R., Rudels, B. and Sakshaug, E. (2006) Climate variability and physical forcing of the food webs and the carbon budget on panarctic shelves. *Progr. Oceanogr.*, **71**, 145-181.

Codispoti, L. and Lowman, D. (1973) A reactive silicate budget for the Arctic Ocean. *Limnol. Oceanogr.*, **18**, 448-456.

Comiso, J. C. (2012) Large Decadal Decline of the Arctic Multiyear Ice Cover. *J. Clim.*, **25**, 1176-1193.

Coyle, K. O., Eisner, L. B., Mueter, F. J., Pinchuk, A. I., Janout, M. A., Cieciel, K. D., Farley, E. V. and Andrews, A. G. (2011) Climate change in the southeastern Bering Sea: impacts on pollock stocks and implications for the oscillating control hypothesis. *Fish. Oceanogr.*, **20**, 139-156.

Emery, W. (2001) Water types and water masses. In: Steele, J. H., Thorpe, S. and Turekian, K. K. (eds), *Encyclopedia of ocean sciences*. Vol. 6, Elsevier Academic Press, London UK, pp 3179-3187

Ershova, E. A., Hopcroft, R. R., Kosobokova, K. N., Matsuno, K., Nelson, R. J., Yamaguchi, A. and Eisner, L. B. (2015) Long-term changes in summer zooplankton communities of the Western Chukchi Sea, 1945-2012. *Oceanography*, **28**, 100-115.

Falk-Petersen, S., Pavlov, V., Timofeev, S. and Sargent, J. R. (2007) Climate variability and possible effects on arctic food chains: The role of *Calanus*. In: Ørbæk, J. B., Kallenborn, R., Tombre, I., Hegseth, E. N., Falk-Petersen, S. and Hoel, A. H. (eds) *Arctic alpine ecosystems and people in a changing environment.* Springer Berlin Heidelberg, pp. 147-166.

Humphries, G. W., Deal, C., Elliott, S. and Huettmann, F. (2012) Spatial predictions of sea surface dimethylsulfide concentrations in the high arctic. *Biogeochemistry*, **110**, 287-301.

Hunt, B. P., Nelson, R. J., Williams, B., McLaughlin, F. A., Young, K. V., Brown, K. A., Vagle, S. and Carmack, E. C. (2014) Zooplankton community structure and dynamics in the Arctic Canada Basin during a period of intense environmental change (2004–2009). *J. Geophys. Res.: Oceans*, **119**, 2518-2538.

Hunt, G. L., Coyle, K. O., Eisner, L. B., Farley, E. V., Heintz, R. A., Mueter, F., Napp, J. M., Overland, J. E., Ressler, P. H., Salo, S. and Stabeno, P. J. (2011) Climate impacts on eastern Bering Sea foodwebs: a synthesis of new data and an assessment of the Oscillating Control Hypothesis. *ICES J. Mar. Sci.*, **68**, 1230-1243.

Karnovsky, N., Kwazniewski, S., Weslawski, J., Walkusz, W. and Beszczynska-Möller, A. (2003) Foraging behavior of little auks in a heterogeneous environment. *Mar. Eol.: Prog. Ser.*, **253**, 289-303.

Kosobokova, K. and Hirche, H.-J. (2000) Zooplankton distribution across the Lomonosov Ridge, Arctic Ocean: species inventory, biomass and vertical structure. *Deep-Sea Res., Part I*, **47**, 2029-2060.

Liu, J., Song, M., Horton, R. M. and Hu, Y. (2013) Reducing spread in climate model projections of a September ice-free Arctic. *Proc. Natl. Acad. Sci. U. S. A.*, **110**, 12571-12576.

Masatoshi, M., Kouichi, K., Hikaru, W. and Akinori, O. (2000) Feeding habits of three dominant myctophid fishes, *Diaphus theta*, *Stenobrachius leucopsarus* and *S. nannochir*, in the subarctic and transitional waters of the western North Pacific. *Mar. Eol.: Prog. Ser.*, **207**, 129-140.

Mauchline, J., (1998) Advances in marine biology - The biology of calanoid copepods. In: Blaxter, J. H. S., Southward, A. J. and Tyler, P. A. (eds) *Advances in marine biology - The biology of calanoid copepods*. Vol. 33. Academic Press Ltd, London, pp. 1-710.

Melle, W., Ellertsen, B. and Skjoldal, H. R. (2004) Zooplankton: The link to higher trophic levels. In: Skjoldal, H.R. (ed) *The Norwegian Sea Ecosystem.* Tapir Academic Press, Trondheim, pp. 137-202.

Nelson, R. J., Ashjian, C. J., Bluhm, B. A., Conlan, K. E., Gradinger, R. R., Grebmeier, J. M., Hill, V. J., Hopcroft, R. R., Hunt, B. P. and Joo, H. M. (2014) Biodiversity and biogeography of the lower trophic taxa of the Pacific Arctic region: sensitivities to climate change. In: Grebmeier, J.M. and Maslowski W. (eds) *The Pacific Arctic Region.* Springer, Dordrecht, pp. 269-336.

Parkinson, C. L. and Comiso, J. C. (2013) On the 2012 record low Arctic sea ice cover: Combined impact of preconditioning and an August storm. *Geophys. Res. Lett.*, **40**, 1356-1361.

Richardson, A. J. (2008) In hot water: zooplankton and climate change. *ICES J. Mar. Sci.*, **65**, 279-295.

Ringuette, M., Fortier, L., Fortier, M., Runge, J. A., Bélanger, S., Larouche, P., Weslawski, J.-M. and Kwasniewski, S. (2002) Advanced recruitment and accelerated population development in Arctic calanoid copepods of the North Water. *Deep-Sea Res., Part II*, **49**, 5081-5099.

Rogachev, K. A., Carmack, E. C. and Foreman, M. G. G. (2008) Bowhead whales feed on plankton concentrated by estuarine and tidal currents in Academy Bay, Sea of Okhotsk. *Cont. Shelf Res.*, **28**, 1811-1826.

Rutzen, I. (2007) Predictive GIS-modeling of arctic mesozooplankton based on available public online data: Experiences from four copepod species *(Calanus glacialis, C. hyperboreus, Metridia longa, and M. pacifica).* University of Veterinary Medicine Hannover, Foundation.

Schauer, U., Loeng, H., Rudels, B., Ozhigin, V. K. and Dieck, W. (2002) Atlantic Water flow through the Barents and Kara Seas. *Deep-Sea Res., Part I*, **49**, 2281-2298.

Seidov, D., Antonov, J. I., Arzayus, K. M., Baranova, O. K., Biddle, M., Boyer, T. P., Johnson, D. R., Mishonov, A. V., Paver, C. and Zweng, M. M. (2015) Oceanography north of 60°N from World Ocean Database. *Prog. Oceanogr.*, **132**, 153-173.

Shiklomanov, I., Shiklomanov, A., Lammers, R., Peterson, B. and Vorosmarty, C. (2000) The dynamics of river water inflow to the Arctic Ocean In: Lewis, E. L., Jones, E.P., Lemke, P., Prowse, T.D. and Wadhams, P. (eds) *The Freshwater Budget of the Arctic Ocean.* Springer, Dordrecht, pp. 281-296.

Wassmann, P., Kosobokova, K. N., Slagstad, D., Drinkwater, K. F., Hopcroft, R. R., Moore, S. E., Ellingsen, I., Nelson, R. J., Carmack, E., Popova, E. and Berge, J. (2015) The contiguous domains of Arctic Ocean advection: Trails of life and death. *Prog. Oceanogr.*, **139**, 42-65.

Woodgate, R. A., Aagaard, K. and Weingartner, T. J. (2005) A year in the physical oceanography of the Chukchi Sea: Moored measurements from autumn 1990–1991. *Deep-Sea Res., Part II*, **52**, 3116-3149.

Yamamoto-Kawai, M., McLaughlin, F. A., Carmack, E. C., Nishino, S., Shimada, K. and Kurita, N. (2009) Surface freshening of the Canada Basin, 2003–2007: River runoff versus sea ice meltwater. *J. Geophys. Res.: Oceans*, **114**, 2156-2202.

Yen, P., Huettmann, F. and Cooke, F. (2004) A large-scale model for the at-sea distribution and abundance of Marbled Murrelets (*Brachyramphus marmoratus*) during the breeding season in coastal British Columbia, Canada. *Ecol. Model.*, **171**, 395-413.

Chapter 2 Distribution, abundance and biomass of epipelagic zooplankton in the Canada Basin[1]

2.1 Abstract

Zooplankton were sampled from the upper 100 m of the Canada Basin during August of the years 2003 to 2006 using a 150-µm mesh net to determine species composition, abundance, and biomass. To describe the zooplankton community and its relation to the environment, we determined Bray-Curtis similarity, and then applied hierarchical clustering, non-parametric multidimensional scaling (nMDS), and the BEST-BioENV routine. The most abundant zooplankton species in all years were the smaller copepod species such as *Oithona similis* and *Microcalanus pygmaeus*, which also contributed the most to the grouping in the cluster analysis. Biomass was dominated by larger copepod species such as *Calanus hyperboreus* and *C. glacialis*. For the non-copepod zooplankton, the pteropod *Limacina helicina*, and the larvacean *Fritillaria borealis* were the most abundant species during most years. The non-copepod biomass was dominated by the chaetognath *Eukrohnia hamata* and *L. helicina*, while *F. borealis* contributed relatively little to the overall biomass despite its high numbers. Zooplankton communities differed between shelf/slope and basin stations, and between years. Mean abundance and biomass increased over time, suggesting that this community may be responding to declining seasonal ice cover.

2.2 Introduction

The Canada Basin is a deep ice-covered basin located in the central Arctic Ocean. Zooplankton within the Arctic basins are intricately tuned to the basins primary production cycle (Smith & Schnack-Schiel, 1990). While the Canada Basin was historically covered year-round by thick multiyear ice, sea ice extent and concentration have declined rapidly within the past decades, reaching a record minimum in summer 2007 only to have it exceeded recently in summer 2012 (Comiso, 2012; Parkinson & Comiso, 2013). This sea ice melt water represents the major freshwater influx to the Arctic that is supplemented by river discharge, with both sources increasing over time (Yamamoto-Kawai *et al.*, 2009). Water temperatures in the Canada Basin already showed a warming trend from 1993-2008 (Jackson *et al.*, 2010) while freshening of the

---

[1] Rutzen, I. and Hopcroft, R.R. Distribution, abundance, and biomass of epipelagic zooplankton in the Canada Basin. Prepared for submission to Journal of Plankton Research

Beaufort Gyre, as well as a deepening of the nutricline and chlorophyll maximum, have also been observed between the years 2003 to 2009 (McLaughlin & Carmack, 2010). Such detectable changes in the environmental conditions typically have an impact on zooplankton communities (Richardson, 2008), and this is may be particularly true in the Arctic (Gradinger *et al.*, 2010; Nelson *et al.*, 2014).

Zooplankton in the Canada Basin has been studied only sporadically due to its traditionally thick, year-round ice cover. Studies during the last century established rudimentary community composition and seasonal cycles, but were seldom synoptic or repeated (Johnson, 1963; Hopkins, 1969; Pautzke, 1979; Thibault *et al.*, 1999; Ashjian *et al.*, 2003). In contrast, the Beaufort Sea slope region experienced more extensive actvities often related to oil and gas exploration (Hufford *et al.*, 1974; Horner & Murphy, 1985; Hopky *et al.*, 1994a; b; c). Even the more recent studies have been concentrated in the more coastal Beaufort Sea and shelf areas (Darnis *et al.*, 2008; Lane *et al.*, 2008; Walkusz *et al.*, 2008; Walkusz *et al.*, 2010; Walkusz *et al.*, 2013; Smoot & Hopcroft, 2017a), with less focus on the central Canada Basin (Hopcroft *et al.*, 2005; Kosobokova & Hopcroft, 2010; Hunt *et al.*, 2014).

At present, indications are that many species are shared between the Arctic's major basins (Kosobokova *et al.*, 2011). The fauna contains a mixture of endemic Arctic species, viable species shared with other ocean basins, and species that were advected from the Pacific such as *Eucalanus bungii*, *Pseudocalanus newmani*, and *Metridia pacifica* (Hopcroft *et al.*, 2005; Kosobokova & Hopcroft, 2010; Hunt *et al.*, 2014) that are not thought to be viable in the Arctic (Wassmann *et al.*, 2015). The abundance and biomass within the epipelagic upper 100 m of the water column are typically dominated by copepods, whereby the smaller- bodied species such as *Oithona similis*, *Microcalanus pygmaeus*, and *Triconia borealis* make up the bulk of the abundance, while larger bodied endemic copepods such as *Calanus hyperboreus*, *C. glacialis*, *M. longa*, and *Paraeuchaeta glacialis* dominate the biomass (Kosobokova & Hopcroft, 2010). Non-copepod abundance is frequently dominated by the larvaceans *Fritillaria borealis* and *Oikopleura vanhoeffeni* as well as the pteropod *Limacina helicina*, which can at times contribute a significant percentage to the biomass (Hopcroft *et al.*, 2005; Kosobokova & Hopcroft, 2010; Hunt *et al.*, 2014). Zooplankton species are also advected from the shelf into the basin within eddies (Carmack & Macdonald, 2002; Llinás *et al.*, 2009), and hence meroplankton typical for shelf communities, such as barnacle cirripedia and echinoderm larvae, can occasionally be found deep into the basin (Hunt *et al.*, 2014).

A freshening and warming of the Canada Basin due to climate change has implications for the zooplankton community. The epipelagic large-bodied endemic copepods are probably at greatest risk of stress or competition from of the advected sub-arctic species. Hunt *et al.* (2014) suggested there has already been a decrease in abundance of species that are typical for the Arctic and subarctic, such as *O. similis*, *L. helicina*, *M. pygmaeus*, and *F. borealis* during 2007 and 2008. An earlier sea ice retreat and a shrinking sea ice extent combined with the freshening of the Beaufort Gyre (McLaughlin & Carmack, 2010) could potentially increase primary productivity (Arrigo & Van Dijken, 2015) and zooplankton biomass in the basins (Hunt *et al.*, 2014) as has already been demonstrated for the Chukchi Sea shelf (Ershova *et al.*, 2015).

Here we present new data describing the epipelagic zooplankton abundance, biomass, and community structure in the Canada Basin for 2003 to 2006 to fill in spatial and temporal gaps. By describing the basin's zooplankton community before the two recent summer sea ice minima of 2007 and 2012, we hope to build a better foundation to compare the zooplankton communities before and after major environmental changes to establish how they are affected.

2.3 Material and Methods

2.3.1 Zooplankton sampling and taxonomic analysis

Our study area encompassed much of the Canada Basin to as far north as 80°N latitude (Fig. 2.1). Samples were collected during August and September of 2003 to 2006 (day and night) aboard the Canadian Coast Guard vessel Louis S. St-Laurent. Bongo nets with a mouth diameter of 60cm and 150-µm mesh size were deployed vertically in the upper 100 m of the water column. During 2003 and 2004, the nets were equipped with General Oceanics flowmeters to measure the volume of filtered water, while during 2005 and 2006 Sea-Gear flowmeters were used. On three occasions flowmeters iced-up and gave false readings. In these cases a filtration efficiency of 100 % was assumed because there was insufficient phytoplankton to impact filtration efficiency (Hunt *et al.*, 2014). A total of 63 samples were analyzed: 23 from 2003, 23 from 2004, only 4 from 2005 (due to wire-time constraints), and 11 from 2006. Upon collection, the samples were preserved in seawater with 10% buffered formalin. In the lab, they were subsampled using a Folsom splitter (Harris *et al.*, 2000), with smaller subsamples (e.g. 1/128th) used to identify very abundant species (such as *Oithona similis*) and larger subsamples used for the rarer species. The full sample was generally analysed for cnidarians, amphipods, chaetognaths, and larger copepods. Animals were enumerated and measured using the ZoopBiom program (Roff & Hopcroft, 1986). For the more abundant species, up to 100 individuals per species were measured with remaining individuals in the aliquot simply enumerated. The developmental stage of larger copepods and the sex of adults was recorded. We used prosome length to differentiate between *Calanus glacialis* and *C. hyperboreus* during early developmental stages (copepodite stages CI-CIII). For species where early life stages could not be distinguished between species, they were grouped according to their genus. In order to calculate dry weights (DW), we applied length/weight relationships according to Hopcroft *et al.* (2005) that were species specific or from morphologically similar species. Samples from two years (2004 and 2006) of our study were collected concurrently with the work by Hunt *et al.* (2014) that employed a coarser mesh-size.

2.3.2 Environmental data

Concurrent temperature (°C), salinity (PSS), and oxygen (mL/L) data were obtained using the SBE 911 plus (Sea-Bird Electronics Inc.) (McLaughlin *et al.*, 2009). Means were calculated for 0-50m and 0-100 m.

We derived the Euclidean distance to the coastline from World Vector Shoreline data from the National Oceanic and Atmospheric Administration (NOAA) by using ArcMap 10.1 and the Geospatial Modelling Environment (GME).

2.3.3 Statistical analysis

Statistical analysis was conducted using the PRIMER (Version6) (Clarke & Warwick, 2001) and ArcMap 10.1 software. Analysis was completed using all species within a sample unless stated otherwise. We applied a 4^{th}-root transformation of abundance and biomass data and calculated the Bray-Curtis similarity index (Bray & Curtis, 1957). Differences and patterns in the zooplankton community between stations or years were detected using weighted average hierarchical cluster analysis and multidimensional scaling (MDS) based on the Bray-Curtis matrix. We used the SIMPER routine to define the similarity percentage between clusters, and which species were driving the grouping/clustering. For the SIMPER analysis, any species that contributed less than 70% to the within-group similarity was excluded.

To test for significant differences between the annual mean abundance and biomass on species level, the data were log transformed and analysis of variance (ANOVA) was conducted in R. If applicable, Tukey tests were used to determine which years were significantly different. The year 2005 was excluded from these analyses because of the small sample size during that year.

To relate environmental data to community patterns, we employed Primer's BEST BioENV routine after normalizing the physical variables.

2.4. Results

2.4.1 Environmental conditions

Overall, 2003 and 2004 seemed more similar in their environmental conditions compared to 2005 and 2006. All years had a core of low salinity towards the center of the study area due to the Beaufort Gyre. In 2003 and 2004, the northern stations were more saline than 2005 and 2006 (Fig. 2.2). A small tongue of elevated salinity occurred at the southern stations in 2003 and 2005.

The average temperature for the upper 100 m decreased with increasing latitude during all years. Overall, temperatures were colder further offshore, and in 2003 on the eastern and western edge. In 2003, 2005, and 2006, the core of the study area had temperatures between -1 to -0.5 °C. This temperature range was further south in 2004 compared to 2003.

Oxygen concentration in 2003 and 2004 was lower in the north relative to the core of the study area. The same pattern was observed in 2006, with the difference that the southern part of the study area showed a lower oxygen concentration compared to the other years. Oxygen concentration in 2005 was relatively high, with the highest concentration in the northeast.

2.4.2 General abundance and biomass

About 50 taxonomic categories were found during 2003 to 2006, of which 26 were copepods (Table 2.1). Of the other categories, three were euphausiids, four amphipods, four hydrozoans, three larvaceans, two pteropods, one ctenophore, an isopod, an annelid, and larvae of several different taxonomic groups.

The mean abundance and biomass varied only 1.5-fold across years, being highest in 2006 (Table 2.2) but there was no significant difference between the years (*p*-value abundance = 0.117; *p*-value biomass = 0.385). Copepods dominated the abundance and biomass (Fig. 2.3) in all years. They made up around 90% of the abundance in 2003, 2005, and 2006, and 83% in 2004, contributing up to 88% (2003) of community biomass. The abundance of non-copepod zooplankton was dominated by larvaceans in 2003, 2004, and 2006 (6.4%, 12.5%, and 8.8%, respectively). In 2005, non-copepod zooplankton abundance was dominated by pteropods (7.02%), but since this observation was based on only four stations, it was not appropriate to statistically compare it to the abundance from the other years. Non-copepod biomass was dominated by chaetognaths in all years (2003 = 6.5%, 2004 = 10.9%, 2005 = 8%, 2006 = 10.2%) (Fig. 2.3). The category of "others" (Fig. 2.3) consisted of polychaete larvae, isopods, cnidarians, meroplankton, amphipods, euphausiids, chaetognaths, and ostracods (Fig. 2.4).

The size spectra of copepod abundance and biomass showed expected patterns with the smaller species being the most abundant, but biomass was highest for large copepods with a series of modes largely dominated by *Calanus* stages (Fig. 2.5). The size spectra were similar throughout all years.

2.4.3 Species-specific abundance

The zooplankton community for most stations consisted mostly of species common in the Arctic such as *Calanus hyperboreus*, *C. glacialis*, *Metridia longa*, *Triconia borealis*, *Oithona similis*, *Microcalanus pygmaeus*, *Paraeuchaeta glacialis*, and *Fritillaria borealis*, with *O. similis* and *Microcalanus pygmaeus* the most abundant in all years (Table 2.1 and Fig. 2.6). The species *C. glacialis*, *Paraheterorhabdus norvegicus*, *M. longa*, *O. similis*, *Spinocalanus* spp., *Oikopleura labradoriensis*, *Eukrohnia hamata*, *Themisto libelula*, *T. abyssorum*, and *A. digitale* were significantly different (*p*-value ≤ 0.05) between some years (Table 2.1). Pacific species that were advected to the Canada Basin, such as *Eucalanus bungii* and *Neocalanus flemingeri*, were observed sporadically, typically as single specimens within a sample.

Microcalanus pygmaeus showed a pattern with higher abundance in the basin than towards the coast. *C. hyperboreus* also generally displayed larger numbers in the basin than towards the coast. In 2003, 2004, and 2006, the lower abundances of *C. hyperboreus*, *C. glacialis*, and *M. longa* coincided with the region of lower salinity and higher oxygen concentration within the Beaufort Gyre (Figs. 2.2 and 2.6). In 2003, 2005, and 2006, *M longa* displayed higher abundances towards the shelf than in the Basin, while in 2004, abundances tended to be lower at a cluster of stations toward the north central basin.

While *C. hyperboreus*, *C. glacialis*, and *M. longa* had higher abundances towards the western Beaufort Sea (especially in 2004), *O. similis* and *M pygmaeus* had elevated abundances towards the eastern Beaufort Sea, where the water was fresher due to the Beaufort Gyre.

In 2003, 2004, and 2006, *F. borealis* was less abundant in the south of our study area, towards the shelf break. Beyond that, no obvious spatial pattern of abundance was observed for *F. borealis* or *L. helicina* (Fig. 2.7). For most species the distance to the coastline, bottom depth, and mean salinity (surface to 100 m) each explained less than eight percent of the variance (Table 2.3). For *C. hyperboreus*, the distance to the coast explained 30 percent of the variance and salinity explained 36 percent of the variance. For *O. similis*, 21 percent of the variance was accounted for by salinity (Table 2.3).

2.4.4 Community structure

The hierarchical cluster analysis using abundance data showed two larger groups (group B: 22 stations; group C: 27 stations) and 4 smaller groups (group A: 5 stations; group D: 2 stations; group E: 3 stations; station 28A clustered by itself) at 68 to 73 percent similarities. The main characteristic of the two larger clusters was that group B was dominated by samples from 2003 and group C was dominated by samples from 2004 (Fig. 2.8). The majority of 2005 and 2006 were divided between both of these clusters. The MDS (2D stress: 0.23; 3D stress: 0.15) reinforced this pattern. SIMPER analysis revealed that most of the similarity within group B was due to *O. similis*, *Microcalanus* copepodites, and calanoid nauplii. The driving species within the group C were *O. similis*, *Microcalanus* copepodites, and *F. borealis*. Most of the dissimilarity between group B and C was due to *O. similis*, *Spinocalanus*, *F. borealis*, and *Microcalanus* copepodites. The abundance of *O. similis* and *Spinocalanus* were significantly different (*p*-value ≤ 0.05) between the years 2003/2006, and 2004/2006, respectively (Table 2.1).

Most of the multi-year stations clustered together, but within different groups (i.e. the CABOS stations clustered in one group while other multiyear stations clustered in a different group than CABOS), indicating that community structure was fairly similar throughout the years

(Fig. 2.8). The CABOS stations were the eastern most-stations and were with depths of less than 1600 m some of the shallowest stations during this study (Fig. 2.1).

Biomass clustered in nine groups at 64 to 70.5 percent similarity. The majority of the stations were again sorted into two larger groups (group C: 34 stations, and D: 17 stations) and four smaller groups (group A: 3 stations; group B: 3 stations; group E: 2 stations; station 28 A clustered by itself) (Fig. 2.9). Group C contained stations from all years, while group D did not include any stations from 2003. Most of the Basin stations from 2004 – 2006 were within groups C and D. Group A consisted of 2 stations from 2003 and one from 2005, all of which were located on the southern part of the study area, but in 2005 the station was closest to the shelf.

The Station 28A (2006) clustered by itself in terms of biomass and abundance (Figs. 2.8 and 2.9). This station was located the closest to the mouth of the Mackenzie River, which influences the species composition and abundance. Compared to other groups, it had a very low abundance of *F. borealis* and *C. glacialis* and was the only station where the neritic *Centropages abdominalis* was observed.

According to the BEST analysis, a combination of mean temperature, oxygen for the upper 50 m, and bottom depth were the best environmental variables to explain the variance for community structure based on abundance ($\rho = 0.379$) and biomass ($\rho = 0.306$) (Table 2.4). Adding salinity and distance to the coastline to the model did not improve the relationship.

2.5 Discussion

Our zooplankton community consisted mostly of species characteristic for the Arctic such as *C. hyperboreus*, *C. glacialis*, *M. longa*, *O. similis*, *M. pygmaeus*, *F. borealis*, and *L. helicina* (Johnson, 1956; Conover & Huntley, 1991; Auel & Hagen, 2002; Hopcroft *et al.*, 2005; Lane *et al.*, 2008; Hunt *et al.*, 2014). Copepods dominated the abundance (83% - 90%) and biomass (up to 88%) in our study. While smaller species such as *O. similis* and *M. pygmaeus* dominated the abundance, the larger bodied *C. glacialis* and *C. hyperboreus* dominated biomass, with notable biomass contributions also made by *M. longa* and *P. glacialis*. These patterns are consistent with previous studies conducted in the Canada Basin and Beaufort Sea slope (Hopcroft *et al.*, 2005; Darnis *et al.*, 2008; Kosobokova & Hopcroft, 2010; Hunt *et al.*, 2014; Smoot & Hopcroft, 2017a). The average abundance of *C. glacialis* and *C. hyperboreus* was at the same order of magnitude as previous observations in the same area (Hunt *et al.*, 2014), and further towards the Chukchi Sea and central Arctic (Thibault *et al.*, 1999). It is notable that we report higher mean abundance but similar biomass in both 2004 and 2006 than Hunt *et al.* (2014), who reported on samples collected concurrently in those years. This is a direct reflection our finer mesh size of 150-µm compared to their 236-µm that catches earlier stages of smaller, abundant species such as *O. similis* and *M. pygmaeus* as well as nauplii (Gallienne & Robins, 2001; Hopcroft *et al.*, 2005), although those species contribute little to biomass. Our mean biomass for all years was slightly higher than the 9.6mg m^{-3} Hopcroft *et al.* (2005) reported for 2002 using a similar methodological approach. The higher abundance of smaller bodied species was reflected in our copepod size spectra (Fig. 2.5), which displayed the pattern typical for Arctic basin copepods (Hopcroft *et al.*, 2005; Kosobokova & Hopcroft, 2010) with the smaller-sized species contributing the most to the abundance, but the larger-sized species contributing the majority to the biomass.

Non-copepod abundance was dominated by larvaceans, mainly *F. borealis* in most years, except in 2005 when pteropods (mostly *L. helicina*) dominated the non-copepod abundance at 7%. However, due to the small sample size in 2005, we cannot be certain whether pteropods were really dominant in the basin, although Kosobokova and Hopcroft (2010) also reported pteropods being more abundant than larvaceans slightly earlier in that same year. Our values of pteropod abundance ranged between 2 to 7% while Thibault *et al.* (1999) reported that pteropods contributed about 8% to the total zooplankton abundance during summer 1994 and Hunt *et al.* (2014) observed a range between 8 -18% of total zooplankton abundance from 2004 to 2008 (excluding 2005). Both these studies had larger mesh sizes than our study (200-µm and 236-µm mesh, respectively), which means that the proportions of taxa caught compared to our 153-µm mesh might be different.

Non-copepod biomass was dominated by chaetognaths with *Eukrohnia hamata* contributing much more than *Parasagitta elegans* except at the most nearshore sites. The other major invertebrate predator of zooplankton in this system was *Aglantha digitale*. The biomass dominance of chaetognaths and hydrozoans is consistent with previous studies (Hopcroft *et al.*, 2005; Kosobokova & Hopcroft, 2010).

Euphausiids were rarely found in our samples due to the limited volumes filtered. Larger zooplankton, like euphausiids, can also avoid our slowly towed plankton nets and were hence underrepresented in our samples. Specimens collected belonged exclusively to the genus *Thysanoessa*, as is typical for the Amerasian Arctic (Kosobokova *et al.*, 2011).

While the abundance of *C. hyperboreus* and *M. longa* seemed to be impacted by the fresher water in the core of our study area, the abundance of *O. similis* did not appear to be substantially influenced. This may reflect the more euryhaline and eurythermic character of *O. similis* (Nishida, 1985; Nielsen *et al.*, 2002) compared to Arctic endemic species. *Oithona similis* showed a significant increase in mean abundance (*p*-value = 0.05) from 2003 to 2006. However, four years of observation is insufficient to convincingly establish whether these trends are persistent or not. If however, we combine our observation with those preceding (Hopcroft *et al.*, 2005) and partially overlapping (Hunt *et al.*, 2014) our observation period, it becomes clear that there has been a steady increase in mean biomass from 2002 to 2008 (except in 2004).

The main patterns observed with the hierarchical clustering analysis and MDS (Figs 2.8 and 2.9) were the groupings of the years 2003 and 2004, and the division of basin stations and shelf/slope stations. The division of most 2003 and 2004 samples into separate groups suggests that there was a difference between these years, which was more pronounced in abundance than biomass. The main differences between these groups were a lower mean abundance of *F. borealis* and *O. similis* in 2003 compared to 2004, but differences were not significant (Table 2.1). This could be due to a difference in sea ice concentration and the environmental conditions that come with it. During 2004, the ice edge was further north than during 2003 and a higher concentration of first year ice was observed in the southern Canada Basin and Beaufort Sea compared to 2004 (National Ice Center: Weekly chart products; http://nsidc.org/data/bist/). A sea ice retreat beyond the shelf break can lead to increased wind-driven upwelling at the shelf break (Carmack & Chapman, 2003), which brings nutrient rich water into the surface layers and leads to increased production.

Our grouping for 2004 and 2006 showed a similar structure to the groupings observed by Hunt *et al.* (2014) during the same years, with the majority of the 2004 and 2006 samples clustering within the same group, except for the 28 A and CABOS stations. That Station 28A built a single cluster for abundance as well as biomass likely reflects influence by river runoff from the Mackenzie, plus it is the station located furthest on the shelf. Shelf communities differ from basin communities due to the hydrography (Walkusz *et al.*, 2010). This was reflected in the observation of the neritic copepod *Centropages abdominalis* and echinoderm larvae (Walkusz *et* *al.*, 2010; Smoot & Hopcroft, 2017a) in the sample. The results of our BEST analysis also confirm the influence of station depth on differences between zooplankton communities, since the best models included bottom depth.

The CABOS stations were represented in an abundance group (Fig. 2.8), for which most samples were located in shallower waters (compared to the rest of our study area) above 1600 meters (Fig. 2.1). As previously mentioned, zooplankton community composition of the basin and shelf differs from shelf stations containing meroplankton such as barnacle nauplii and cyprids, as well as echinoderm larvae and decapod zoea (Smoot & Hopcroft, 2017a). We observed the same distinctions between the shelf/slope and the basin stations as Hunt *et al.* (2014), namely that the basin stations were characterized by the general absence of shelf taxa, although stations in the western basin had low numbers of meroplankton, which suggests transport from the shelf into the basin by eddies (Llinás *et al.*, 2009). The CABOS stations also contained the gymnosome pteropod *C. limacina*, which was only found at seven stations in total. *F. borealis* abundance was relatively low at these shelf stations, which further distinguished them from basin stations.

Historic datasets for comparison to our findings are limited for the Canada Basin due to the remoteness of the area as well as the ice cover. Most older data come from ice-stations such as Drift Station Alpha (Johnson, 1963), T-3 ice islands (Hopkins, 1969), NP-22 (Kosobokova, 1982), and the Surface Heat Budget of the Arctic Ocean (SHEBA) (Ashjian *et al.*, 2003). We cannot directly compare our data to the Drift station Alpha data because there were no samples reported from August. However, samples were reported from June and September. The sampling in September (mesh size 230-µm) was only at one station location and the abundance was probably underestimated due to copepods, such as *C. hyperboreus* (Ashjian *et al.*, 2003), leaving the upper 100 m of the water column and descending to depths. The average mean abundance at this station was only 0.46 ind. m^{-3}. The T-3 ice island data from Hopkins (1969) that were closest to our study area were sampled in March 1964 and are thus not comparable to ours. USS Burton Island mean abundance in 1950 (~52 ind. m^{-3} (Johnson, 1956)) was about 5 times higher than what we observed, but these values include mostly shelf/slope stations, as well as northern Chukchi Sea. Our data were closer to the mean abundance for the zooplankton community reported for SHEBA of 16 ind. m^{-3} (Ashjian *et al.*, 2003) for August and early September in 1998 (mesh size 150-µm, sampling depth: 200 m) and that of Smoot and Hopcroft (2017b) of about 18 ind. m^{-3} along the Beaufort Slope for August and September 2012-2014 (mesh size 150-µm, sampling depth 100 m).

Comparing contemporary data to historical data is difficult due to differing sampling techniques, areas, as well as changing taxonomy (e.g. Johnson (1956) reported *C. finmarchicus* instead of *C. glacialis*, since *C. glacialis* was not described by Jaschnov until 1955). The lack of historic data that are comparable to our study highlights the need for areas such as the Canada Basin to be resampled regularly with consistent methods.

We conclude that there was no obvious interannual change in community structure, but considering the work by Hopcroft *et al.* (2005) and (Hunt *et al.*, 2014), overall biomass and abundance displayed an increasing trend. Other group patterns shown in this paper are mostly due to bathymetric differences such as shelf versus deep basin.

Acknowledgements

This work was funded by the National Science Foundation (NSF) OPP-0909571, The Arctic Ocean Diversity (ArcOD) Project of the Census of Marine Life (CoML), and a University of Alaska Fairbanks graduate fellowship, and contributes to the Circumpolar Biodiversity Monitoring Program (CBMP) through NOAA/CIFAR awards *NA08OAR4320751* and *NA13OAR4320056*.

We thank the Department of Fisheries and Oceans (DFO) and Fiona McLaughlin for providing sampling opportunities, and the crew of the Canadian Coast Guard Icebreaker Louis S. St-Laurent as well as the students and staff collecting the zooplankton samples. We also thank Cheryl Clarke-Hopcroft and Chris Stark for support with the taxonomic questions and Caitlin Smoot for software assistance.

Figures

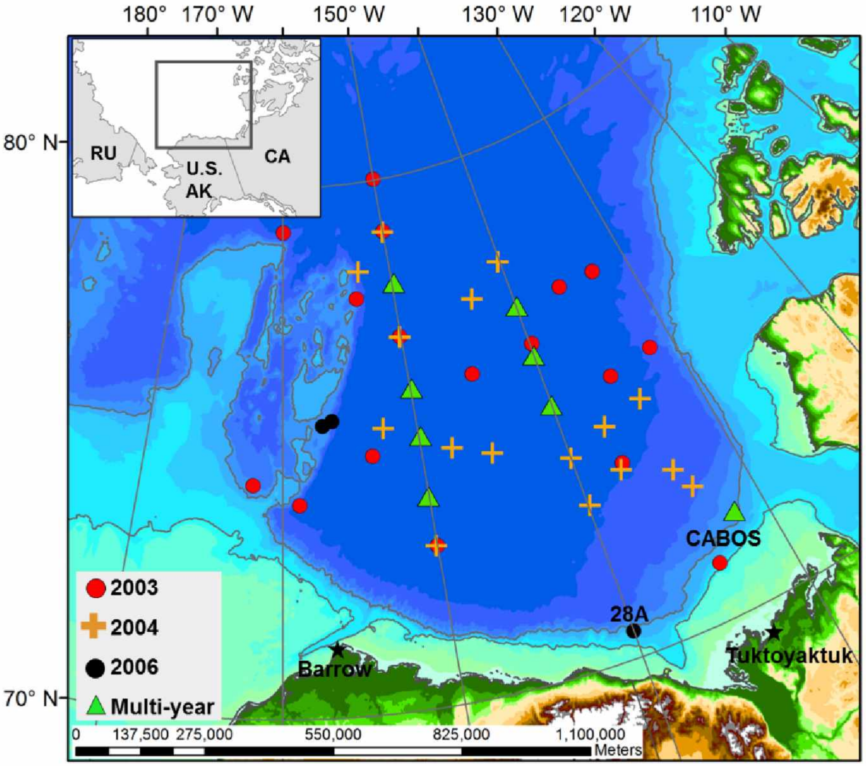


Figure 2.1: Map of the Canada Basin study area. Multi-year stations include 2003, 2004, 2005, and 2006 stations, if applicable

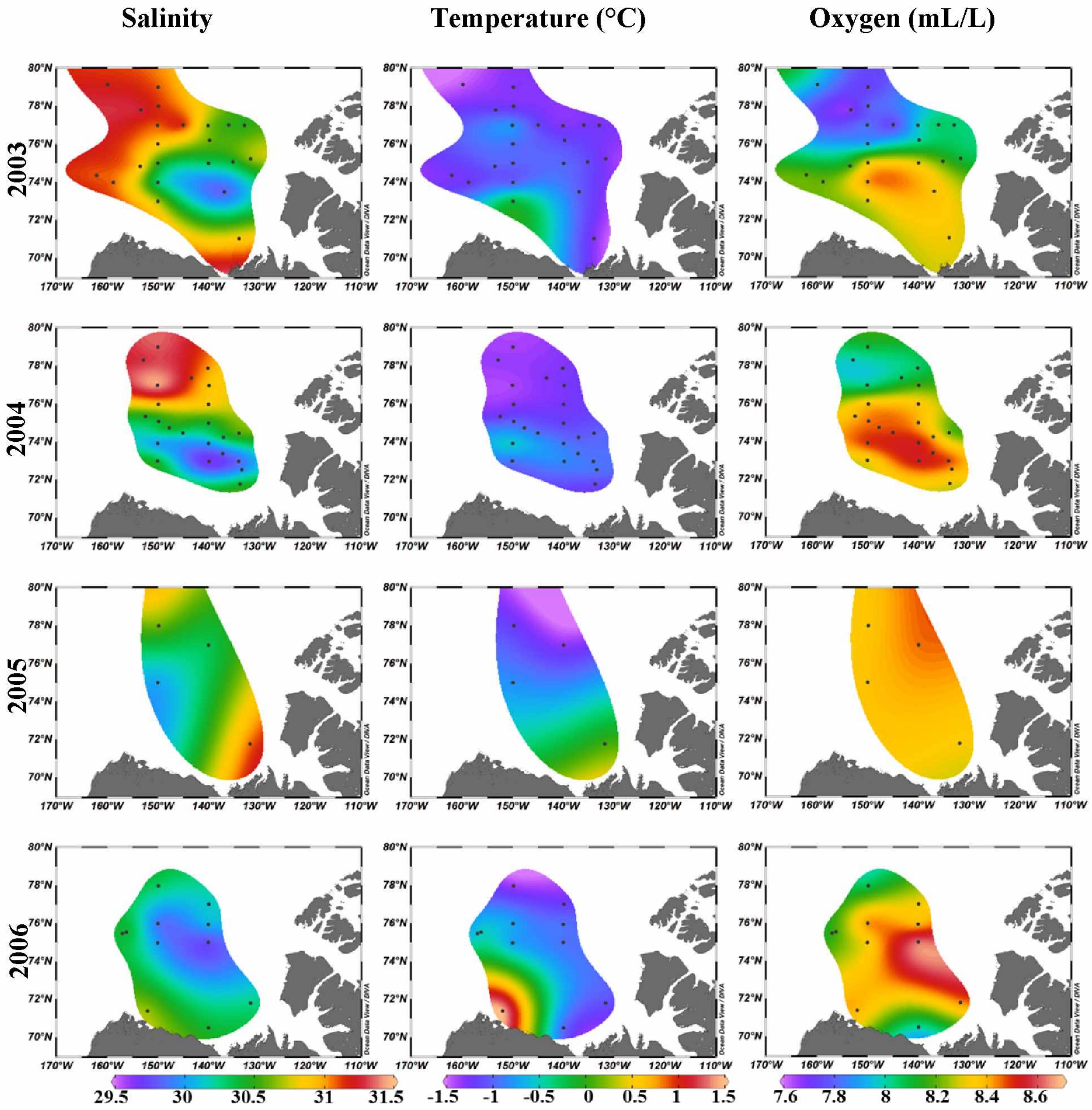


Figure 2.2: Average (surface to 100m) for Salinity, Temperature (°C), and Oxygen (mL/L) plots for the years 2003 - 2006

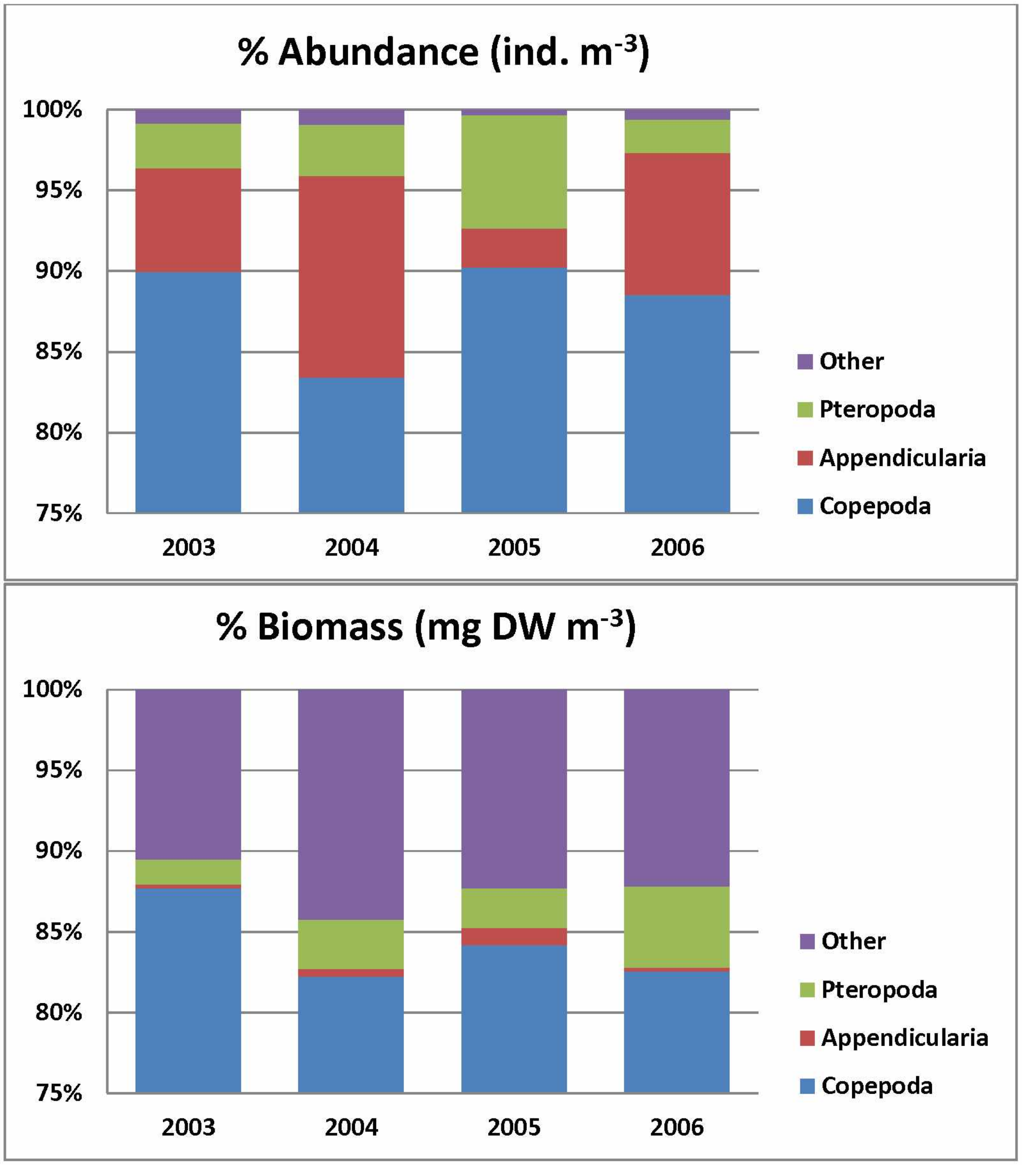


Figure 2.3: Relative abundance (ind. m^{-3}) and biomass (mg DW m^{-3}) of major zooplankton groups in the Canada Basin for 2003-2006. Lower panel range is trimmed to increase resolution

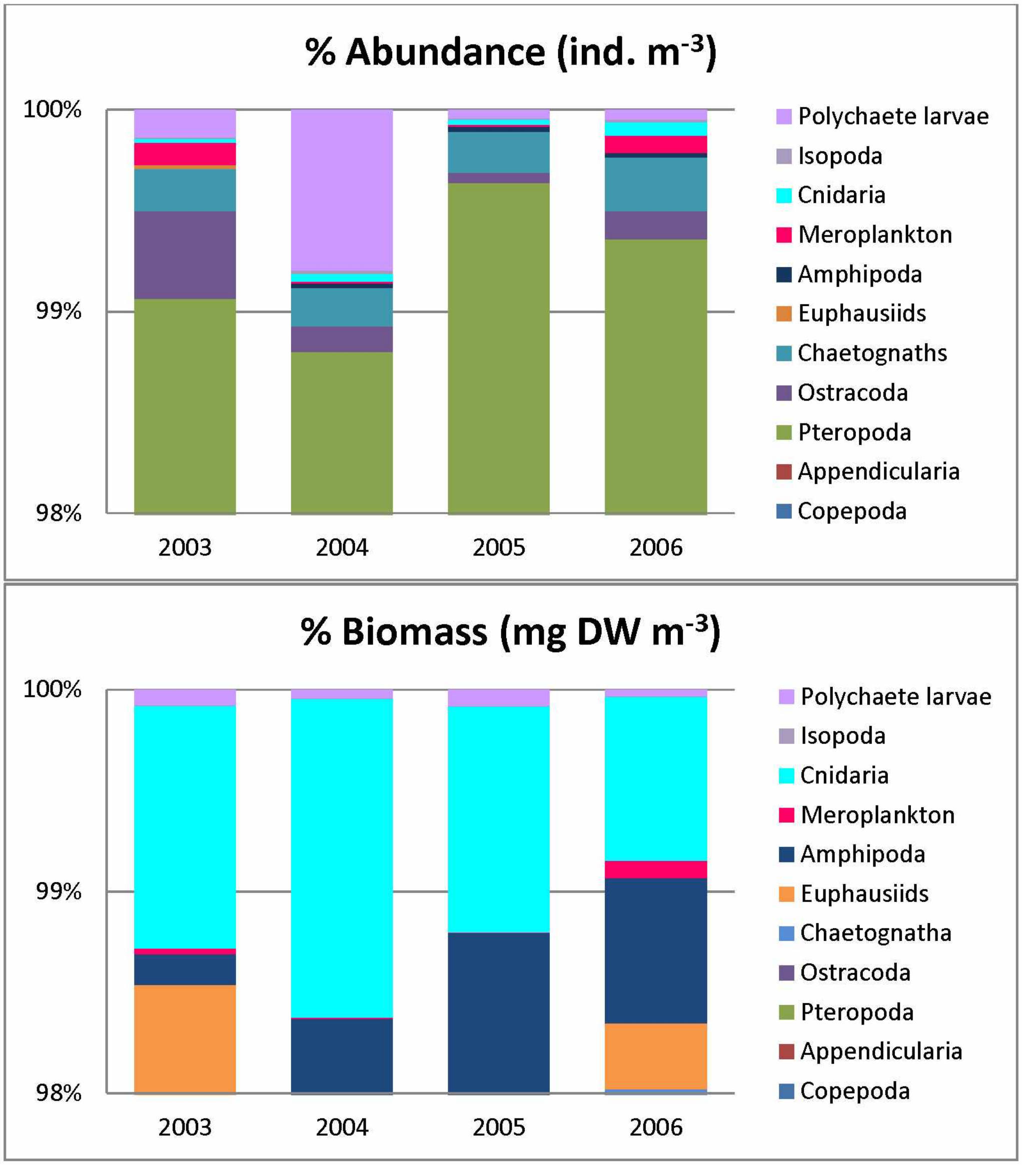


Figure 2.4: Relative abundance (ind. m^{-3}) and biomass (mg DW m^{-3}) of other zooplankton groups in the Canada Basin for 2003-2006 for 2 % of data labeled "other" in Fig. 2.3

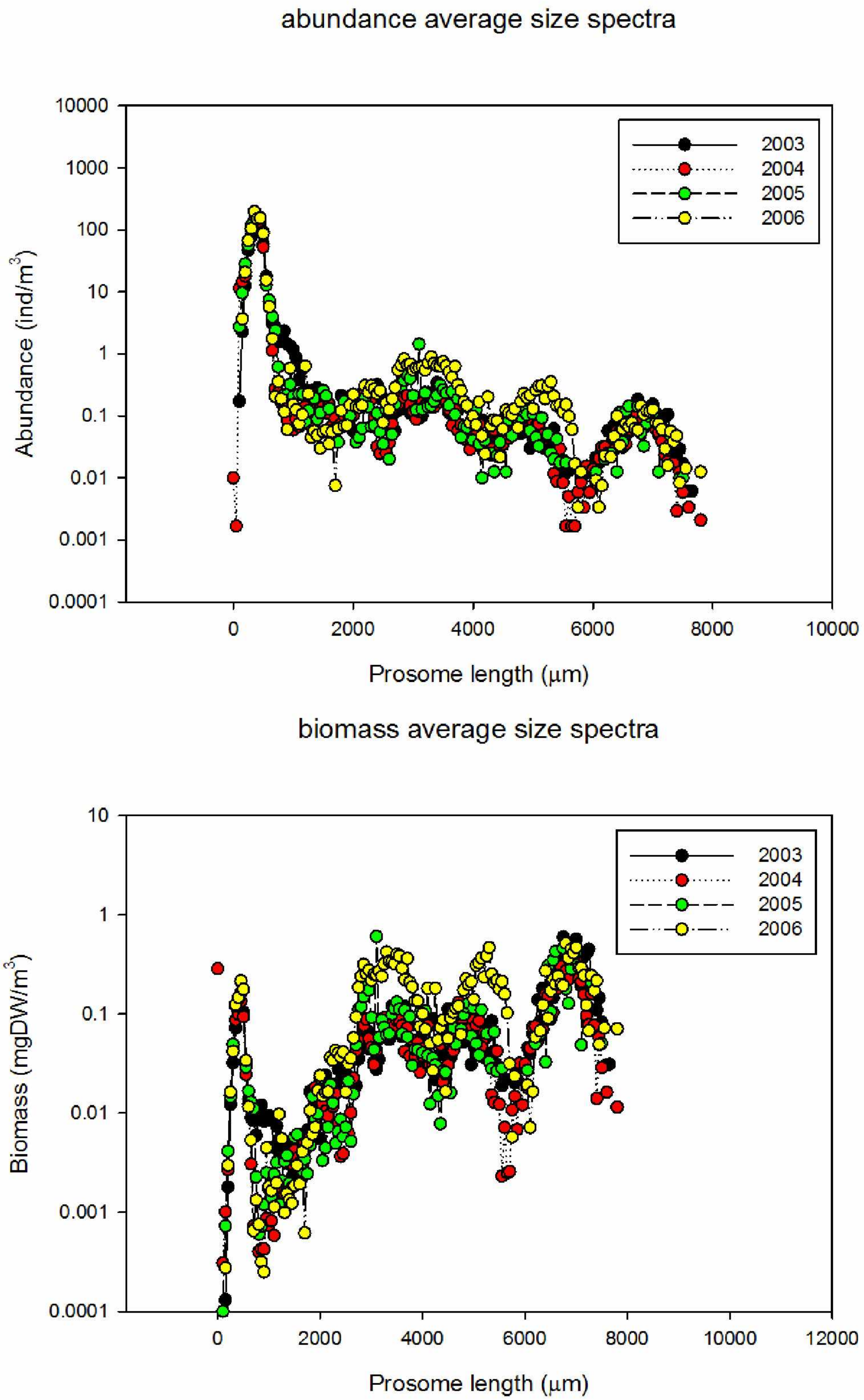


Figure 2.5: Size spectra of copepod prosome length (µm) for abundance and biomass in the Canada Basin during 2003 -2006

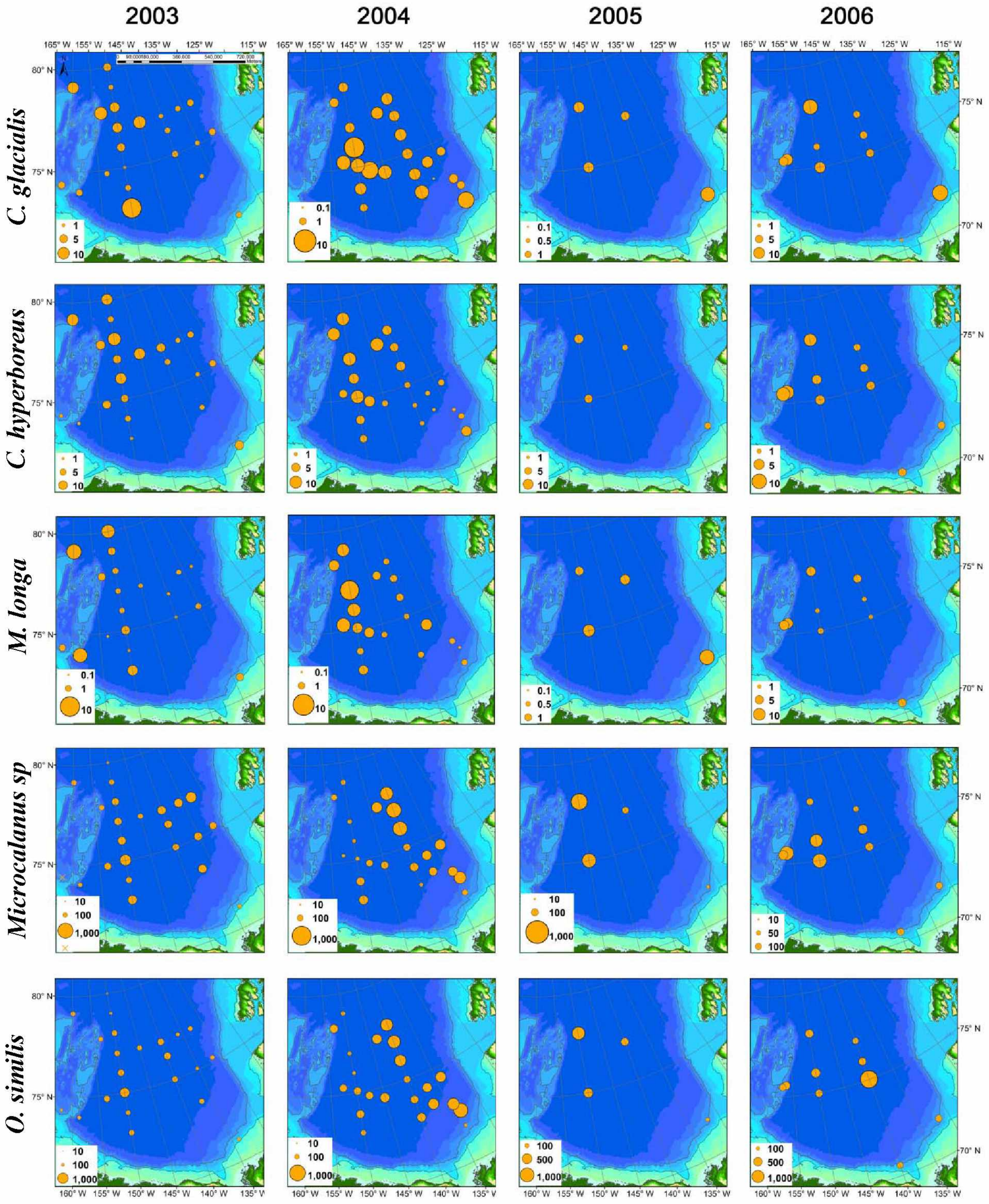


Figure 2.6: Copepod abundance (ind. m^{-3}) proportional plots in the Canada Basin 2003-2006. *Microcalanus* copepodites and adults combined

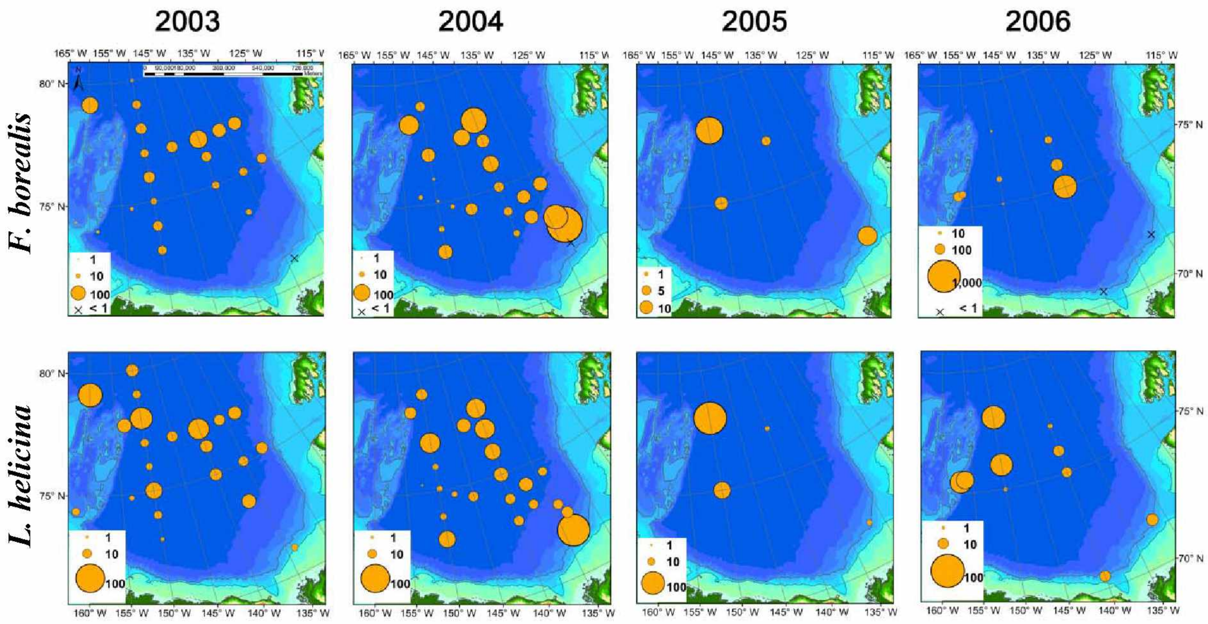


Figure 2.7: Proportional abundance plots (ind. m^{-3}) of *F. borealis* and *L. helicina* in the Canada Basin 2003-2006

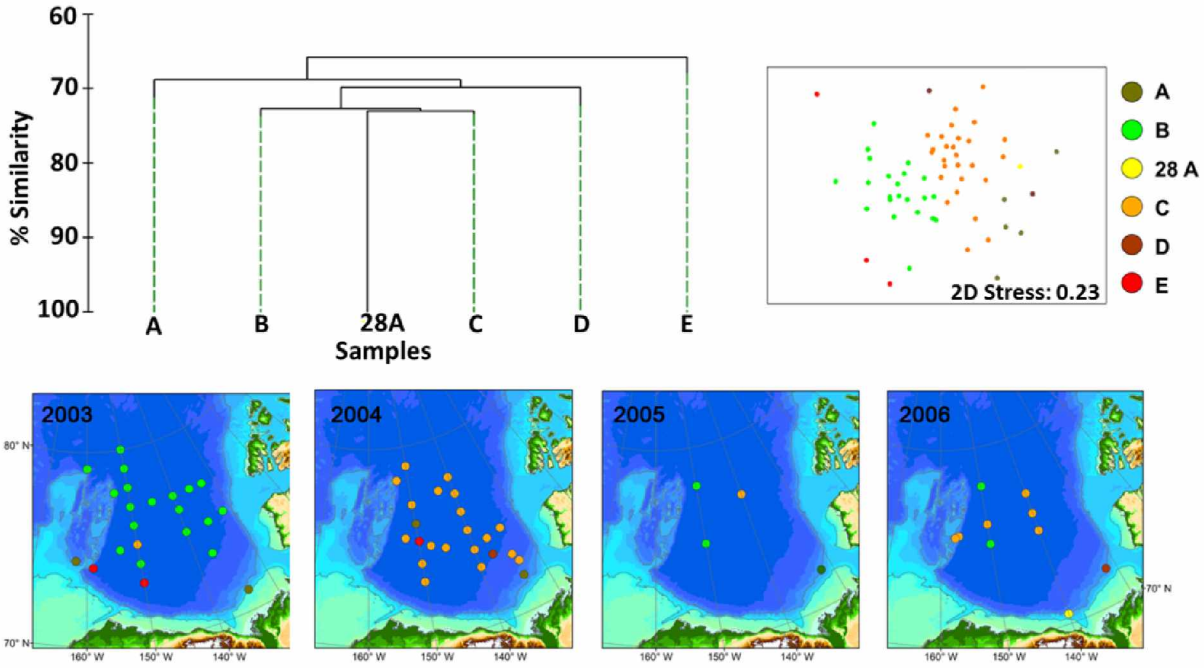


Figure 2.8: Hierarchical clustering analysis (Bray-curtis similarity in %), MDS plot, and spatial distribution of zooplankton abundance groups in the Canada Basin 2003-2006

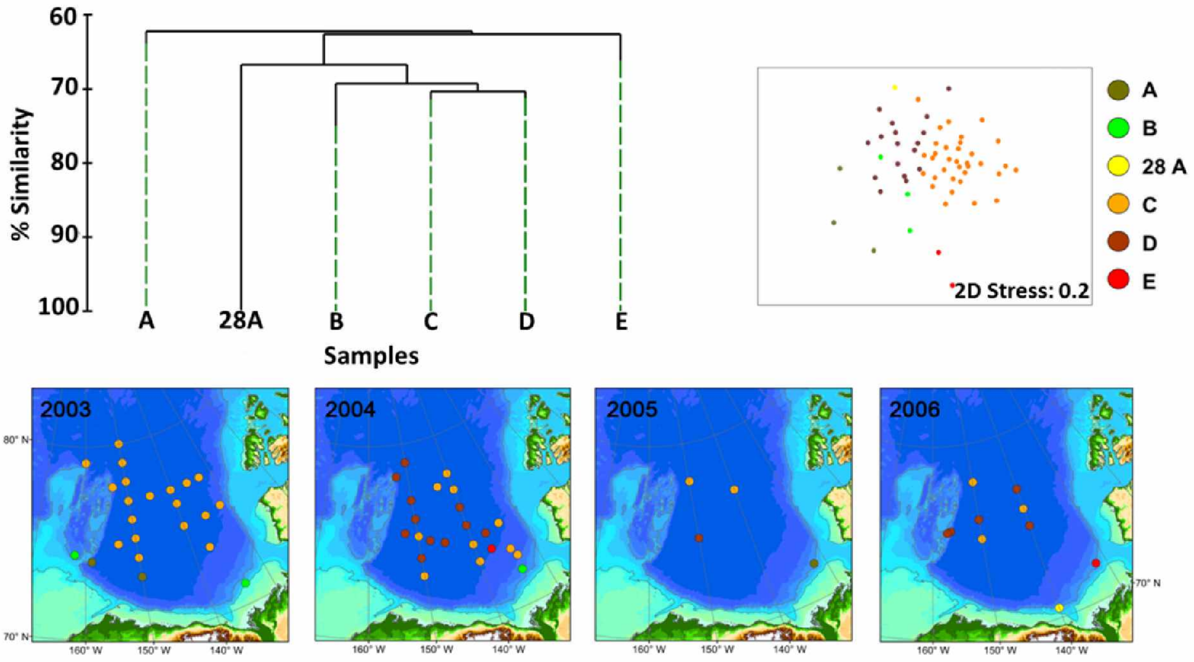


Figure 2.9: Hierarchical clustering analysis (Bray-curtis similarity %), MDS plot, and spatial distribution of zooplankton biomass groups in the Canada Basin 2003-2006

Tables

Table 2.1: Average abundance and biomass of zooplankton in the Canada Basin, 2003-2006. * indicates taxon abundance or biomass <0.01. - indicates species was not present and *p*-value not reported. NC means biomass not calculated. *p*-values from ANOVA for interannual (2005 excluded) differences with a significance level of 0.05. *p*-values are displayed for species that were present in two or more years. The column "year" indicates which years were significantly different if applicable

	Abundance (ind. m^{-3})				***p*-value**	**year**	**Biomass (mg DW m^{-3})**				***p*-value**	**year**
Copepoda	2003	2004	2005	2006			2003	2004	2005	2006		
Acartia longiremis	-	*	-	-	-	-	-	NC	-	-	-	-
Aetideopsis minor	-	-	-	*	-	-	-	-	-	*	-	-
Calanus glacialis	5.34	2.9	3.06	8.4	0.005	04/06	1.52	1.67	1.21	4.3	0.0002	03/06 04/06
Calanus hyperboreus	7.72	5.45	5.58	4.78	0.42	-	8.29	6.33	5.75	6.33	0.64	-
Centropages abdominalis	-	-	-	0.01	-	-	-	-	-	*	-	-
Chiridius obtusifrons	0.01	0.01	0.02	0.02	0.66	-	*	*	*	*	0.74	-
Eucalanus bungii	-	-	-	0.02	-	-	-	-	-	*	-	-
Eurytemora pacifica	-	*	-	-	-	-	-	*	-	-	-	-
Gaetanus sp	-	*	-	-	-	-	-	*	-	-	-	-
Harpacticoida	-	*	-	*	-	-	-	*	-	*	-	-
Paraheterorhabdus norvegicus	-	*	0.04	0.09	0.0002	03/06 04/06	-	*	*	0.02	<0.001	03/06 04/06
Paraheterorhabdus sp	-	0.03	-	-	-	-	-	*	-	-	-	-
Jaschnovia tolli	-	-	0.18	*	-	-	-	-	*	*	-	-
Metridia sp. (cI-III)	1.47	0.09	0.01	2.26	0.51		*	*	*	*	0.61	-
Metridia longa	1.39	2	2.57	3.7	0.01	03/06	0.31	0.59	0.94	1.35	0.0002	03/06 04/06
Microcalanus pygmaeus	1.19	0.14	1.37	1	0.08	-	*	*	*	*	0.11	-
Microcalanus copepodite	221.21	206.1	245.63	231.09	0.78	-	0.17	0.14	0.15	0.15	0.53	-
Microsetella norvegica	0.22	0.14	0.04	0.4	0.29	-	*	*	*	*	NA	-
Neocalanus flemingeri	-	*	-	-	-	-	-	*	-	-	-	-
Oithona similis	255.67	317.08	414.05	520.23	0.05	03/06	0.29	0.32	0.48	0.6	0.03	03/06

Table 2.1 continued…

Oithona spinirostris	-	-	-	2.78	-	-	-	-	-	*	-	-
Paraeuchaeta glacialis	0.66	0.65	0.7	1	0.12	-	0.61	0.6	0.85	0.95	0.16	-
Pseudocalanus sp copepodite	8.47	4.06	2.92	0.6	0.46	-	0.04	*	0.01	*	0.4	-
Pseudocalanus minutus	2.21	0.07	0.07	0.8	0.31	-	0.02	*	*	0.01	0.35	-
Pseudocalanus newmanii	-	-	0.04	0.02	-	-	-	-	*	*	-	-
Scaphocalanus sp	0.02	0.22	1.4	0.07	0.32	-	*	0.08	0.6	0.04	0.55	-
Scolecithricella minor	0.24	0.27	0.22	0.66	0.05	-	*	*	*	0.01	0.01	03/06
Spinocalanus sp copepodite	-	0.02	-	0.03	0.15	-	-	*	-	*	0.32	-
Spinocalanus sp	12.63	0.71	25.08	7.54	0.0005	03/04	0.01	*	0.02	*	0.07	-
Triconia borealis	46.54	32.31	34.4	69.42	0.28	-	0.04	0.03	0.04	0.06	0.02	04/06
Nauplii												
Calanus nauplii	57.48	55.52	94.14	47.39	0.69	-	0.04	0.04	0.1	0.03	0.45	-
Cycplopoida nauplii	1.51	0.03	19.1	0.85	0.19	-	*	*	*	*	0.26	-
Cirripedia nauplii	0.75	0.02	0.04	0.1	0.41	-	*	*	*	*	0.48	-
Appendicularia												
Fritillaria borealis	43.56	81.84	20.05	88.74	0.51	-	0.01	*	*	*	0.54	-
Oikopleura sp	0.34	0.1	1.19	0.31	0.16	-	*	0.01	0.06	*	0.89	-
Oikopleura labradoriensis	-	0.32	1.22	0.23	0.02	03/04	-	*	0.05	*	0.1	-
Oikopleura vanhoeffeni	0.85	0.97	0.24	0.42	0.54	-	0.01	0.05	*	0.03	0.56	-
Pteropoda												
Clione limacina	*	0.02	-	*	0.6	-	0.01	*	-	*	0.54	-
Limacina helicina	19.21	23.88	66.58	21.19	0.82	-	0.16	0.35	0.3	0.84	0.08	-
Chaetognatha		0.01										
Eukrohnia hamata	1.32	1.24	0.96	2.67	0.007	03/06 04/06	0.75	0.97	0.93	1.68	0.006	03/06 04/06
Parasagitta elegans	0.12	0.2	*	0.05	0.38	-	0.09	0.19	0.05	0.03	0.21	-
Ostracoda												
Boroecia maxima	2.65	1.098	0.5	1.27	0.87	-	*	*	*	*	0.68	-

Table 2.1 continued…

Euphausiacea												
Thysanoessa inermis	*	-	-	*	0.57	-	0.02	-	-	0.05	0.37	-
Thysanoessa longicaudata	-	*	-	-	-	-	-	*	-	-	-	-
Thysanoessa raschii	-	*	-	-	-	-	-	*	-	-	-	-
Thysanoessa spinifera	0.13	-	-	-	-	-	0.3	-	-	-	-	-
Euphausiid zoea	-	*	-	-	-	-	-	*	-	-	-	-
Decapoda											-	-
Decapod zoea	-	-	0.02	*	-	-	-	-	*	*	-	-
Paguroid zoea	-	*	-	-	-	-	-	*	-	-	-	-
Amphipoda												
Themisto libellula	*	0.06	0.05	0.08	0.0003	03/04 03/06	0.02	0.17	0.29	0.09	0.08	-
Themisto abyssorum	0.01	0.09	0.16	0.11	0.01	03/04 03/06	*	0.02	0.08	0.03	0.03	-
Gammaridea unk.	-	*	*	0.01	-	-	-	*	*	*	-	-
Hyperiidea unk.	-	*	-	*	-	-	-	*	-	*	-	-
Isopoda	*	0.07	0.01	0.07	0.0003	03/04 03/06	*	*	*	*	0.04	03/06
Hydrozoa												
Aglantha digitale	0.08	0.09	0.24	0.67	0.00006	03/06 04/06	0.09	0.18	0.14	0.11	0.22	-
Aeginopsis laurentii	-	-	-	0.07	-	-	-	-	-	0.25	-	-
Melicertum octopunctata	-	0.2	-	-	-	-	-	*	-	-	-	-
Obelia sp	-	*	-	-	-	-	-	*	-	-	-	-
Ctenophora												
Mertensia ovum	-	*	-	-	-	-	-	*	-	-	-	-
Annelida	-	0.2	-	-	-	-	-	NC	-	-	-	-

Table 2.1 continued…

Larvae												
Cirripedia cyprid	*	0.3	-	0.78	0.09	-	*	*	-	0.01	0.1	-
Echinodermata larvae	-	-	0.05	0.01	0.1	-	-	-	*	*	0.1	-
Echinoidea larvae	-	*	-	-	-	-	-	*	-	-	-	-
Polychaeta larvae	0.96	0.15	0.21	0.41	0.71	-	*	*	*	*	0.67	-
Trochophora larvae	0.83	-	-	-	-	-	*	-	-	-	-	-
Unidentifiable	-	*	-	-	-	-	-	*	-	-	-	-

Table 2.2: Average abundance and biomass of zooplankton in the Canada Basin from 2003-2006. Values rounded to the nearest whole number. SE = Standard error

Year	**Dates**	**# of Samples**	**Abundance (ind. m^{-3}) ± SE**	**Biomass (mg DW m^{-3}) ± SE**
2003	08/11-09/02	23	696 ± 68	13 ± 1.2
2004	08/09-08/30	23	748 ± 90	11 ± 1
2005	08/03-08/25	4	946 ± 322	12 ± 1.6
2006	08/10-09/8	10	1019 ± 203	17 ± 2.5

Table 2.3: Relationship between zooplankton abundance and distance to coast, bottom depth, and mean salinity for the upper 100m of the Canada Basin from 2003-2006. Abundance data were log transformed

Species	**r^2 coast**	**r^2 depth**	**r^2 salinity**
Calanus glacialis	0.05	0.01	0.02
Calanus hyperboreus	0.3	0.006	0.36
Metridia longa	0.11	0.02	0.08
Microcalanus sp	0.004	0.006	0.06
Oithona similis	0.00002	<0.001	0.21
Fritillaria borealis	0.04	0.14	0.002
Limacina helicina	0.08	0.0015	0.004

Table 2.4: BEST BioEnv analysis of zooplankton community structure in the Canada Basin 2003-2006 to Temperature (T), Salinity (S), Oxygen (O), Distance to coastline (C), and Bottomdepth (B). Best combinations explaining clustering for abundance and biomass are in bold. ρ is given in parentheses

	Surface	**0-50m**	**0-100m**
Abundance	T (0.257)	T (0.185)	T (0.217)
	O, B (0.342)	O, B (0.312)	T, B (0.341)
	S, O, B (0.356)	**T, O, B (0.379)**	T, C, B (0.357)
	T, S, O, B (0.366)	T, O, C, B (0.372)	T, O, C, B, (0.342)
	T, S, O, C, B (0.353)	T, S, O, C, B (0.362)	T, S, O, C, B (0.326)
Biomass	T (0.273)	B (0.218)	B (0.218)
	T, O (0.278)	T, B (0.269)	T, B (0.297)
	T, O, B (0.299)	**T, O, B (0.306)**	T, C, B, (0.293)
	T, S, O, B (0.285)	T, O, C, B (0.3)	T, O, C, B (0.281)
	T, S, O, C, B (0.266)	T, S, O, C, B (0.288)	T, S, O, C, B (0.265)

References

Arrigo, K. R. and Van Dijken, G. L. (2015) Continued increases in Arctic Ocean primary production. *Prog. Oceanogr.*, **136**, 60-70.

Ashjian, C. J., Campbell, R. G., Welch, H. E., Butler, M. and Van Keuren, D. (2003) Annual cycle in abundance, distribution, and size in relation to hydrography of important copepod species in the western Arctic Ocean. *Deep-Sea Res., Part I,* **50**, 1235-1261.

Auel, H. and Hagen, W. (2002) Mesozooplankton community structure, abundance and biomass in the central Arctic Ocean. *Mar. Biol.*, **140**, 1013-1021.

Bray, J. R. and Curtis, J. T. (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.*, **27**, 325-349.

Carmack, E. and Chapman, D. C. (2003) Wind-driven shelf/basin exchange on an Arctic shelf: The joint roles of ice cover extent and shelf-break bathymetry. *Geophys. Res. Lett.,* **30**, n/a.

Carmack, E. C. and Macdonald, R. W. (2002) Oceanography of the Canadian Shelf of the Beaufort Sea: A setting for marine life. *Arctic*, **55**, 29-45.

Clarke, K. R. and Warwick, R. M. (2001) *Change in marine communities: an approach to statistical analysis and interpretation.* PRIMER-E Ltd, Plymouth UK.

Comiso, J. C. (2012) Large Decadal Decline of the Arctic Multiyear Ice Cover. *J. Clim.*, **25**, 1176-1193.

Conover, R. and Huntley, M. (1991) Copepods in ice-covered seas -- distribution, adaptations to seasonally limited food, metabolism, growth patterns and life cycle strategies in polar seas. *J. Mar. Sys.*, **2**, 1-41.

Darnis, G., Barber, D. G. and Fortier, L. (2008) Sea ice and the onshore–offshore gradient in pre-winter zooplankton assemblages in southeastern Beaufort Sea. *J. Mar. Sys.*, **74**, 994-1011.

Ershova, E. A., Hopcroft, R. R., Kosobokova, K. N., Matsuno, K., Nelson, R. J., Yamaguchi, A. and Eisner, L. B. (2015) Long-term changes in summer zooplankton communities of the Western Chukchi Sea, 1945-2012. *Oceanography*, **28**, 100-115.

Gallienne, C. P. and Robins, D. B. (2001) Is *Oithona* the most important copepod in the world's oceans? *J. Plankton Res.*, **23**, 1421-1432.

Gradinger, R., Bluhm, B. A., Hopcroft, R. R., Gebruk, A. V., Kosobokova, K., Sirenko, B. and Wesławski, J. M. (2010) Marine life in the Arctic. In: A. McIntyre (ed) *Life in the World's Oceans: Diversity, Distribution and Abundance.* Wiley-Blackwell, Oxford, pp. 183-202.

Harris, R., Wiebe, P., Lenz, J., Skjoldal, H. R. and Huntley, M. (2000) *ICES Zooplankton Methodology Manual.* Elsevier Academic Press, London UK.

Hopcroft, R. R., Clarke, C., Nelson, R. J. and Raskoff, K. A. (2005) Zooplankton communities of the Arctic's Canada Basin: the contribution by smaller taxa. *Polar Biol.*, **28**, 198-206.

Hopkins, T. L. (1969) Zooplankton biomass related to hydrography along the drift track of Arlis II in the Arctic Basin and the East Greenland Current. *J. Fish. Res. Board Can.*, **26**, 305-310.

Hopky, G. E., Lawrence, M. J. and Chiperzak, D. B. (1994a) NOGAP B2; zooplankton data from the Canadian Beaufort Sea shelf, 1984 and 1985. Central and Arctic Region, Department of Fisheries and Oceans, Winnipeg.

Hopky, G. E., Lawrence, M. J. and Chiperzak, D. B. (1994b) NOGAP B2; zooplankton data from the Canadian Beaufort Sea shelf, 1986. Central and Arctic Region, Department of Fisheries and Oceans, Winnipeg.

Hopky, G. E., Lawrence, M. J. and Chiperzak, D. B. (1994c) NOGAP B2; zooplankton data from the Canadian Beaufort Sea shelf, 1987 and 1988. Central and Arctic Region, Department of Fisheries and Oceans, Winnipeg.

Horner, R. and Murphy, D. (1985) Species Composition and Abundance of Zooplankton in the Nearshore Beaufort Sea in Winter-Spring. *Arctic*, **38**, 201-209.

Hufford, G., Fortier, S., Wolfe, D., Doster, J., Noble, D., Barnes, P., Weiss, H., Chew, K., Guttman, M. and Host, A. (1974) WEBSEC 71-72: an ecological survey in the Beaufort Sea: August-September, 1971-1972 US Coast Guard. *Oceanographic Unit, Washington, DC.*

Hunt, B. P., Nelson, R. J., Williams, B., Mclaughlin, F. A., Young, K. V., Brown, K. A., Vagle, S. and Carmack, E. C. (2014) Zooplankton community structure and dynamics in the Arctic Canada Basin during a period of intense environmental change (2004–2009). *J. Geophys. Res.: Oceans*, **119**, 2518-2538.

Jackson, J. M., Carmack, E. C., Mclaughlin, F. A., Allen, S. E. and Ingram, R. G. (2010) Identification, characterization, and change of the near-surface temperature maximum in the Canada Basin, 1993-2008. *J. Geophys. Res.: Oceans*, **115**, 1-16.

Jaschnov, W. (1955) Morphology, distribution and systematics of *Calanus finmarchicus* sl. *Zool. Zh.*, **34**, 1210-1223.

Johnson, M. W. (1956) The plankton of the Beaufort and Chukchi Sea areas of the Arctic and its relation to the hydrography *Technical Paper.* Vol. 1. Arctic Institute of North America Montreal, Quebec, pp. 1-32.

Johnson, M. W. (1963) Zooplankton Collections from the high Polar Basin with special reference to the Copepoda. *Limnol. Oceanogr.*, **8**, 89-102.

Kosobokova, K. N. (1982) Composition and distribution of zooplankton biomass of the central arctic basin. *Okeanologiya*, **22**, 1007-1015.

Kosobokova, K. N. and Hopcroft, R. R. (2010) Diversity and vertical distribution of mesozooplankton in the Arctic's Canada Basin. *Deep-Sea Res., Part II*, **57**, 96-110.

Kosobokova, K. N., Hopcroft, R. R. and Hirche, H.-J. (2011) Patterns of zooplankton diversity through the depths of the Arctic's central basins. *Mar. Biodivers*, **41**, 29-50.

Lane, P. V. Z., Llinás, L., Smith, S. L. and Pilz, D. (2008) Zooplankton distribution in the western Arctic during summer 2002: Hydrographic habitats and implications for food chain dynamics. *J. Mar. Syst.*, **70**, 97-133.

Llinás, L., Pickart, R. S., Mathis, J. T. and Smith, S. L. (2009) Zooplankton inside an Arctic Ocean cold-core eddy: Probable origin and fate. *Deep Sea Res., Part II*, **56**, 1290-1304.

McLaughlin, F. A. and Carmack, E. C. (2010) Deepening of the nutricline and chlorophyll maximum in the Canada Basin interior, 2003–2009. *Geophys. Res. Lett.*, **37**, n/a.

Nelson, R. J., Ashjian, C. J., Bluhm, B. A., Conlan, K. E., Gradinger, R. R., Grebmeier, J. M., Hill, V. J., Hopcroft, R. R., Hunt, B. P. and Joo, H. M. (2014) Biodiversity and biogeography of the lower trophic taxa of the Pacific Arctic region: sensitivities to climate change. In: Grebmeier, J. M. and Maslowski, W. (eds) *The Pacific Arctic Region.* Springer, Dordrecht, pp. 269-336.

Nielsen, T. G., Møller, E. F., Satapoomin, S., Ringuette, M. and Hopcroft, R. R. (2002) Egg hatching rate of the cyclopoid copepod *Oithona similis* in arctic and temperate waters. *Mar. Ecol.: Prog. Ser.*, **236**, 301-306.

Nishida, S. (1985) Taxonomy and distribution of the family Oithonidae (Copepoda, Cyclopoida) in the Pacific and Indian Oceans. *Bulletin of the Ocean Research Institute-University of Tokyo (Japan).*

Parkinson, C. L. and Comiso, J. C. (2013) On the 2012 record low Arctic sea ice cover: Combined impact of preconditioning and an August storm. *Geophys. Res. Lett.*, **40**, 1356-1361.

Pautzke, C. G. (1979) *Phytoplankton primary production below Arctic Ocean pack ice: an ecosystems analysis.* University of Washington.

Richardson, A. J. (2008) In hot water: zooplankton and climate change. *ICES J. Mar. Sci.*, **65**, 279-295.

Roff, J. C. and Hopcroft, R. R. (1986) High precision microcomputer based measuring system for ecological research. *Can. J. Fish. Aquat. Sci.*, **43**, 2044-2048.

Smith, S., L. and Schnack-Schiel, S. B. (1990) Polar Zooplankton. In: Smith, W. O. J. (ed) *Polar Oceanography, Part B: Chemistry, Biology, and Geology.* Vol. 2. Academic Press, San Diego (CA), pp. 527-598.

Smoot, C. A. and Hopcroft, R. R. (2017a) Cross-shelf gradients of epipelagic zooplankton communities of the Beaufort Sea and the influence of localized hydrographic features. *J. Plankton Res.*, **39**, 65-78.

Smoot, C. A. and Hopcroft, R. R. (2017b) Depth-stratified community structure of Beaufort Sea slope zooplankton and its relations to water masses. *J. Plankton Res*, **39**, 79-91.

Thibault, D., Head, E. J. H. and Wheeler, P. A. (1999) Mesozooplankton in the Arctic Ocean in summer. *Deep-Sea Res., Part I*, **46**, 1391-1415.

Walkusz, W., Paulić, J. E., Papst, M. H., Kwasniewski, S., Chiba, S. and Crawford, R. E. (2008) Zooplankton and Ichthyoplankton Data Collected from the Chukchi and Beaufort Seas during the R/V Mirai Cruise, September 2002. *Canadian Data Report of Fisheries and Aquatic Sciences.* Fisheries and Oceans Canada, Winnipeg, Manitoba R3T 2N6, pp. 42.

Walkusz, W., Paulić, J., Kwaśniewski, S., Williams, W., Wong, S. and Papst, M. (2010) Distribution, diversity and biomass of summer zooplankton from the coastal Canadian Beaufort Sea. *Polar Biol.*, **33**, 321-335.

Walkusz, W., Williams, W. J. and Kwasniewski, S. (2013) Vertical distribution of mesozooplankton in the coastal Canadian Beaufort Sea in summer. *J. Mar. Sys.*, **127**, 26-35.

Wassmann, P., Kosobokova, K. N., Slagstad, D., Drinkwater, K. F., Hopcroft, R. R., Moore, S. E., Ellingsen, I., Nelson, R. J., Carmack, E., Popova, E. and Berge, J. (2015) The contiguous domains of Arctic Ocean advection: Trails of life and death. *Prog. Oceanogr.*, **139**, 42-65.

Yamamoto-Kawai, M., McLaughlin, F. A., Carmack, E. C., Nishino, S., Shimada, K. and Kurita, N. (2009) Surface freshening of the Canada Basin, 2003–2007: River runoff versus sea ice meltwater. *J. Geophys. Res.: Oceans*, **114**, 2156-2202.

Chapter 3 A GIS-based approach to model-predicting abundances of the Arctic's dominant copepods at a pan-arctic scale[1]

3.1 Abstract

We present the first modelled spatial distribution and abundance prediction of seven Arctic and Sub-Arctic copepods (*Calanus hyperboreus*, *C. glacialis*, *C. finmarchicus*, *Metridia longa*, *M. pacifica*, *Neocalanus cristatus*, and *Paraeuchaeta glacialis*) in the Arctic Ocean during the summer months. We consolidated 53 zooplankton datasets from 30 sources, subjected it to rigorous quality control, and pooled data regardless of year to achieve greater spatial coverage for the ecological niche. Habitats were characterized using averaged climatologies of environmental variables acquired from online databases. We applied data mining approaches (TreeNet and RandomForests) and present the modelled abundances as GIS maps, covering the area from 60-90°N latitude. Major distribution patterns of the copepods were successfully predicted with good accuracies. Differences between Arctic endemic and advected species reveal that geographic predictors rank higher in importance for endemic species, whereas water mass properties rank higher for advected species. This approach shows it is possible to identify accurate patterns with data pooled from different sources, and different sampling methodology. It lays a thorough foundation for models looking at long-term temporal changes and impacts of climate change on species abundance and distribution.

3.2 Introduction

In the Arctic Ocean, decreased sea ice cover, changes in the timing of sea ice retreat, increased fresh water influx, ocean acidification, and rising water temperatures, are altering ecosystems (Fabry et al. 2008; Mueter et al. 2009; Leu et al. 2011). These ecosystem alterations also affect zooplankton as a major base of the marine food chain, which have been observed to react to changing conditions (Ershova et al. 2015b; Kristiansen et al. 2015). Zooplankton are ideal for detecting the impact of climate change on the Arctic ecosystem, because their relatively

---

[1] Rutzen I, Hopcroft R.R., Huettmann F., Kosobokova K., Coyle K.O., Nelson J.R., Ashjian C., Lischka S., Markhaseva E., Falk-Petersen S., Dalpadado P., Gislason A., Campbell R., Estrada R., Harvey M., Melle W., Fortier L., Darnis G., Yamaguchi A., Matsuno K., Walkusz W., Smoot C.A., Questel J.M., Pinchuk A. A GIS based approach to model-predicting abundances of the Arctic's dominant copepods at a pan-Arctic scale. Prepared for submission to Polar Biology

short life cycles facilitate an observable response to changes in environmental variables, such as rising temperature (e.g. Richardson 2008; Dalpadado et al. 2014). Copepods dominate the zooplankton abundance and biomass in the Arctic Ocean (Kosobokova and Hirche 2000) and they represent a direct link between primary producers and higher trophic levels (Mauchline et al. 1998; Melle et al. 2004; Falk-Petersen et al. 2008). Examples for such trophic connections are found with seabirds (Springer et al. 1987; Karnovsky et al. 2003), fish (Hopkins and Gartner 1992; Masatoshi et al. 2000), and whales (Rogachev et al. 2008). The Arctic zooplankton community consists of a mixture of species endemic to the Arctic that reproduce there, as well as species that are advected and expatriated from the Pacific and the Atlantic Ocean into the Arctic Ocean (Kosobokova and Hirche 2000), but that are otherwise not viable in the Arctic (Wassmann et al. 2015).

Calanus hyperboreus, *C. glacialis*, *Metridia longa*, and *Paraeuchaeta glacialis* are endemic to the Arctic and dominate the biomass in central oceanic waters (Kosobokova and Hirche 2000; Kosobokova and Hopcroft 2010). *C. finmarchicus* is transported with the saline, nutrient-poor North Atlantic water through the Norwegian and Greenland Seas into the Arctic Ocean, where it often dominates community biomass (Hirche and Kosobokova 2007), while *M. pacifica* and *Neocalanus cristatus* are among the large-bodied species, transported with the nutrient-rich water from the North Pacific that enters through Bering Strait and the Chukchi Sea (Conover 1988; Springer et al. 1989; Questel et al. 2013; Ershova et al. 2015a). Arctic water and plankton are transported southward to the North Atlantic through the Canadian Archipelago and the eastern shelf of Greenland (Aagaard 1989; Carmack et al. 2006; Seidov et al. 2015), but lack a southward connection to the Pacific (Aagaard 1989). The seasonally-advected species generally do not reproduce in the Arctic, because the environmental conditions are suboptimal for them. However, in response to climate changes, advected species may survive longer, and in the future might reproduce within the Arctic, leading to pronounced changes in food web dynamics (Nelson et al. 2014).

It has also been proposed that the copepod community composition might shift towards smaller-sized, less energy-rich species, e.g. a shift from *Calanus glacialis* to *C. finmarchicus* (Hirche and Kosobokova 2007), due to a shift in the timing of the phytoplankton bloom (e.g. Falk-Petersen et al. 2007; Hunt et al. 2011). A northward shift of species distribution from a larger, more nutrient-rich species (*C. finmarchicus*) to a smaller sibling species (*C.*

helgolandicus) has already been observed in the North Sea (Beaugrand et al. 2002). In contrast, Ringuette et al. (2002) concluded that a reduced sea-ice cover and longer open water season would simply enhance population growth of large endemic copepods such as *C. glacialis* and *C. hyperboreus* on Arctic shelves. In the Chukchi Sea, both endemic and expatriated large-bodied species appear to be increasing (Ershova et al. 2015b). Thus, it is important that we understand the current range of habitat, and the abundance of the key endemic and expatriated Pacific and Atlantic species if we are to detect shifts in this ecosystem due to climate change.

In order to detect changes in the Arctic ecosystem and the zooplankton community, it is important to collect and compile environmental and biological data over long time periods and establish robust baseline studies (Nelson et al. 2014). Once these baselines are established, we will be able to identify changes occurring in the Arctic Ocean. Numerous projects have focused on observing the Arctic ecosystem, but most of the zooplankton studies are concentrated within the Pacific Arctic (Chukchi and Beaufort Seas) and the Atlantic Arctic. To undertake a comprehensive spatial and temporal analysis of change, the effort to make station-resolved data from programs and studies (historical and ongoing) easily accessible to scientists needs to become a priority. This is already happening; for example, station-resolved zooplankton data from the Chukchi Sea Environmental Science Program (CSESP) were made available for download after the data were published in scientific papers.

Maps and climatologies of species distribution and abundance, and predictive ecological niche modeling can be important tools for monitoring changes in ecosystems, and for explaining the distribution of predators based upon that of their prey. Several data rescuing and archiving projects have compiled oceanographic and biological data in order to make them publicly available, such as the Archives of the Arctic Seas by the Zoological Institute of Russian Academy of Sciences, the BioChem database by the Fisheries and Oceans Canada (DFO), the Coastal and Oceanic Plankton Ecology, Production, and Observation Database (COPEPOD) by the National Oceanic and Atmospheric Administration (NOAA), the Arctic Ocean Biodiversity Project (ArcOD), the Ocean Biogeographic Information System (OBIS), the Pacific Marine Arctic Regional Synthesis (PacMARS), and the European data archive PANGAEA (Table 3.1). We compiled available zooplankton datasets from these and various additional sources into a single database in order to model-predict, for the first time, the abundance of seven prominent large-bodied copepod species on a pan-Arctic scale.

Our primary objective was to create a summer climatology of the spatial distribution and abundance for each of our selected species in the Arctic Ocean. These were to be developed and assessed by employing boosted regression trees (TreeNet) and bagging (RandomForests) which have been shown to be a valuable modelling approach for ecological data (De'ath 2007; Oppel et al. 2009). This modelling approach is algorithmic and can be used for predictions with non-linear data as well as drawing inferences on how the environmental variables (predictors) influence the response variable (e.g. abundance) (Breiman 2001b). These tools have been successfully used in modelling species distributions in a wide variety of habitats and species (Huettmann and Diamond 2001; Rutzen 2007; Hardy et al. 2011; Huettmann et al. 2011; Schmid 2012) and also in predicting abundances (Yen et al. 2004; Oppel et al. 2012), concentrations (Humphries et al. 2012; Oppel et al. 2012), species richness, and spatial niche and biodiversity hotspots for conservation planning and management (e.g. Baltensperger and Huettmann 2015; Miller et al. 2015). Although there are similar modeling efforts already published for zooplankton (e.g. Rutzen 2007; Schmid 2012; Brun et al. 2016; Kaschner et al. 2016), this paper represents the first attempt at modeling the abundance of multiple copepod species on a pan-Arctic scale using machine learning methods.

3.3 Material and Methods

3.3.1 Study area and species

The study area encompasses the Arctic Ocean and the adjacent seas as far south as 60° N latitude (Fig. 3.1). We employed zooplankton abundance data from the upper 200 meters of the water column collected during July, August, and September, because most high Arctic zooplankton data are available only for the summer months when the sea ice cover is reduced. We conducted the modelling at species level for seven copepod species (Table 3.2). These species were chosen primarily because of their prominent role in the arctic ecosystem, they were previously well studied, and because of their larger size, which makes them less subject to biases associated with different mesh sizes (Antacli et al. 2010; Ershova et al. 2015b). Four of these species are endemic (*Calanus hyperboreus*, *C. glacialis*, *Metridia longa*, and *Paraeuchaeta glacialis*) and they reproduce only in the Arctic Ocean and its adjacent seas (Hirche and Bohrer 1987; Conover 1988; Grainger 1989; Hirche and Niehoff 1996; Kosobokova and Hirche 2001; Auel 2004). The remaining three species are passively advected within prevailing currents into the arctic and are not known to reproduce in the high arctic (i.e. *C. finmarchicus* is advected from the Atlantic while *Neocalanus cristatus* and *M. pacifica* are advected from the Pacific Ocean) (Conover 1988; Hirche and Kosobokova 2007). The three *Calanus* species as well as *M. longa* are known to dominate the mesozooplankton biomass in the Arctic-Atlantic (e.g. Smith and Schnack-Schiel 1990; Kosobokova et al. 2011).

3.3.2 Zooplankton data

We collated zooplankton abundance data from several open access databases, previously published but undigitized data (i.e. data rescue), data shared through cooperation by other scientists, and the processing of previously collected field samples (Table A-1). We updated all taxonomy according to the Integrated Taxonomic Information System (ITIS; http://www.itis.gov/) and the World Register of Marine Species (WoRMS; http://www.marinespecies.org/). We evaluated and corrected possible errors in the data such as incorrect sample station coordinates, erroneous sampling metadata, or incorrect definition of abundance units (e.g. individuals 100 m^{-3} wrongly reported as individuals m^{-3}), aided where possible by referring to original publications. We imported the data into a MS Access 2010 database, aligned them to one consistent format using Visual Basic for Applications (VBA), then developed VBA scripts to extract the data used for modelling. We consolidated a total of 53 datasets (Table A-1) representing between 2,179 and 2,581 single spatial positions of abundance and presumed absence (hereafter called data points) for each species (Figs. 3.2 and 3.13). The observations spanned a temporal domain from 1934 to 2012 and were pooled regardless of study year to achieve maximal spatial coverage of the Arctic domain during summer. Several historical datasets were unusable for this study due to missing metadata, a coarse taxonomy (i.e., identification was not conducted to species level), or because the data only indicated presence and not abundance of species. We also excluded some data due to sampling methodology (e.g. water pumps) that made it problematic to compare true abundances.

C. finmarchicus and *C. glacialis* were not distinguished as separate species until 1955 (Jaschnov 1955; Jaschnov 1970), so we adjusted older taxonomic records where it was applicable (Fig. A-1). For example, the “Archives of the Arctic Seas” dataset includes data from 1934, before the species were taxonomically separated. We made adjustments based on known species distribution and expert knowledge, excluding data in areas where it was impossible to assign the species with certainty (e.g. Barents Sea).

The sampling methodology varied considerably between the datasets. Some collections were performed vertically from target depths continuously upward to the surface, others were towed obliquely and several were separated into contiguous strata using multiple opening-closing nets. In the case of contiguous data, the data were integrated to reflect the upper 200 m of the water column. If the sampling depth did not have a stratum at exactly 200 m, the closest depth was used for integration, but never exceeded 200m. The abundance data were standardized to reflect the individuals per m^2. In the case of time series data, the abundances over our target seasonal window (i.e. July-September) were averaged across years at each station.

The data were collected with several gear types such as Nansen nets, Juday nets, and Bongo nets (Table A-1). The mesh sizes included in the analysis ranged from 20 µm to 500 µm. Previous studies have shown that the gear and mesh size used to collect zooplankton influences abundance estimates, especially of smaller bodied species (Hopcroft et al. 2005; Antacli et al. 2010; Skjoldal et al. 2013; Ershova et al. 2015b). Hence, we selected species large enough to be caught by coarser mesh sizes and also common enough to be sufficiently represented with smaller-mouthed nets (Ershova et al. 2015b).

We pooled all copepodite life-stages and sex, because many datasets were lacking such information. Nauplii data and eggs were excluded from the analysis because they were not available for most datasets, are highly influenced by mesh size, and seldom reported to species.

In order to incorporate absence data, and due to the lack of "confirmed absence", we used sampling stations where the species was not mentioned in the datasets, excluding stations that said "Copepoda", "Calanoida", or "*Calanus*", and also excluding datasets that had a biomass for the target species but did not list their abundance.

3.3.3 Environmental data

The environmental data used to describe the habitat were averaged climatologies (Table 3.3), because matching measurements were not available for many older datasets. We selected environmental variables believed to directly or indirectly influence copepod abundance or distribution as predictors of their niche. The data for the environmental variables were derived from open access online databases (Table 3.3). Most of them (mostly older versions) have been applied before in comparable modelling approaches, on pan-arctic scales (> 66° N latitude) (Rutzen 2007; Huettmann et al. 2011; Humphries et al. 2012) as well as globally (Wei et al. 2011). Here we updated those GIS layers with the best available datasets having a spatial coverage from 60 - 90° N latitude.

The data derived from the World Ocean Atlas (WOA13 (Seidov et al. 2015)) were interpolated using the inverse distance weighting tool (IDW) in ArcMap 10.1 for 15 different depth levels starting at the surface with a 10 meter depth increase until 100 meters and then every 25 meters until 200 meters. We employed the statistical mean during the summer months (July to September, average of 6 decadal means (1955-2012)) for modelling. The data from WOA13 were averaged for two different depth strata: surface to 50 m depth to describe the polar mixed layer (Jones 2001) and 60 to 200 m to describe the arctic halocline water (Rudels et al. 1996). For bathymetry, we applied the General Bathymetric Chart of the Oceans with a resolution of 30 arc-seconds (GEBCO_08, http://www.gebco.net/data_and_products/gridded_bathymetry_data/). The bathymetric slope was derived from the bathymetry layer within ArcMap 10.1 using the slope command in spatial analyst tools. We imported a seasonal climatology for chlorophyll a data (summer, 2002-2014), obtained from NASA's Aqua/MODIS satellite data, into ArcMap using the Marine Geospatial Ecology Tools (MGET; http://mgel.env.duke.edu/mget) version 0.8a56. The chlorophyll a dataset is limited north of 76 degree latitude due to seasonal ice coverage. We used the freshwater discharge from the major rivers into the Arctic Ocean and adjacent seas from R-ArcticNET version 4.0. We calculated the Euclidean distance to rivers (hereafter called distance to hydrology) from Version 2 of the Global lakes and wetlands database (GLWD). The Euclidean distance to glaciers was based on glacier distribution data from National Snow and Ice Data Center (NSIDC) World Glacier Inventory. We obtained the mean sea ice concentration climatology from NSIDC and converted it to a raster following their instructions. The sea ice climatology shows the mean sea ice concentration percentages for July to September for the time period of 1979 through 2013. The data were averaged over the months and years, so that a single raster for sea ice concentration was employed in the model. Sea ice data were lacking in an area around the North Pole and we therefore used IDW to interpolate. We calculated the Euclidean distance to the coastline/shore based on the World Vector Shoreline data from the National Oceanic and Atmospheric Administration (NOAA). For modelling purposes, we defined the shelf break from the bathymetry layer at 200 m. In order to incorporate a variable that would pick up on shelf species vs deep basin species, we created a raster showing the distance between the 200 m bathymetric isoline and the coast as negative values, and between the 200 m isoline and North Pole as positive values. All environmental variables were fitted to the North Pole Stereographic projection using geographic datum WGS-84.

3.3.4 GIS overlays and processing

To test and account for different mesh sizes, we incorporated mesh as a predictor into the model to determine if the model's accuracy was influenced by the mesh size. For this, we sorted mesh sizes into bins depending on the size of the species (Table A-2). We determined the bins by considering prosome length size- frequency distributions from various Arctic studies conducted by our lab (Hopcroft et al. 2005; Questel et al. 2013; Smoot 2015) as well as additional studies that consider size for efficient copepod retention by the mesh size (Nichols and Thompson 1991; Hopcroft et al. 2001; Antacli et al. 2010; Skjoldal et al. 2013). Each bin was then assigned a numerical value. Similarly, we assigned a numerical value used to identify each taxonomist as a unique categorical predictor in the model. If the taxonomist was unknown, then a unique value was assigned for the dataset.

We incorporated the year of the observation, the taxonomist, and the binned mesh size as predictor variables during exploratory model runs. However, these variables could not be applied for pan-arctic predictions, because they could not be meaningfully assigned to a pan-arctic spatial grid covering the entire study area. Nonetheless, this approach is still rather informative for affect inference.

We spatially overlaid the abundance data with the previously described environmental data (in total 25 environmental variables, Table 3.3) by using the Geospatial Modelling Environment (GME; http://www.spatialecology.com/gme/). This approach links each datapoint to the environmental variables that fall into the same geospatial pixel.

3.3.5 Modelling

For our zooplankton abundance models, we applied regression trees using TreeNet and RandomForests (Salford Predictive Modeler (SPM) Version 7). These machine-learning algorithms can handle erroneous data and outliers as well as missing values (Breiman 2001a;

De'ath 2007), which are frequently encountered in large real-world spatial datasets (Craig and Huettmann 2008; Oppel et al. 2009). RandomForests and TreeNet have already been applied effectively to data with uneven spatial distribution and data gaps (Wei et al. 2011; Humphries et al. 2012). In TreeNet, "many weak learners are combined and averaged to create a strong learner" with a low predictive error (De'ath 2007). The prediction of the target variable (in our case copepod abundance) is optimised by selecting trees with the lowest error value of internal cross-validation data. RandomForests have the advantage that they usually do not overfit and are successfully used for predictions (Breiman 2001a). Here, the prediction of the target variable is optimised and validated by averaging the result over all trees in the 'forest' (De'ath 2007). During model development, the data were split into learn and test data. The learn data were a subset from the entire dataset used for training the model while the test data were withheld from modelling and applied in order to test, or assess the models performance.

In order to find the best model settings (e.g. number of cross validations, learn rate, number of nodes of the regression trees) we applied the "battery" command (see Humphries et al. 2012), that allows us to run several models with different settings and compare their performance using Mean Squared Errors (MSE). We chose the models with the lowest MSE for the final predictions (Tables A3 and A4). After fitting the model to the available data, we applied the abundance prediction to a regular spatial grid that was created in ArcGIS 10.1 and covered the entire study area (Huettmann and Diamond 2001; Yen et al. 2004; Wei et al. 2011; Humphries et al. 2012). We assigned a predicted abundance value to each gridpoint, then combined and averaged the RandomForests and TreeNet models for a more accurate prediction with reduced variance (De'ath 2007). We produced a map of the predicted abundance in individuals m^{-2} by interpolating the grid on the North Pole Stereographic Projection using inverse distance weighting.

TreeNet provides a relative importance score for each predictor variable derived through a permutation procedure (Friedman 2001). Sequentially, the values of each predictor variable are randomly permuted, and together with the unpermuted variables, are used to predict the response (in our case predicted abundance). The importance of the predictor variable is assessed by the decrease in model accuracy after the permutation process (Strobl et al. 2007). The relative importance score is standardized and ranges from 0 to 100, with 100 being the most important to predict the species abundance (Oppel and Huettmann 2010). Based on the relative importance score, we examined the partial dependence plots of the 3 most important variables. Partial dependence plots were created by applying the model to the abundance data with the target predictor (e.g. bathymetry) being the only predictor that changed, while all other predictors were averaged and kept stable. The outcome created a response curve for every learn data record (learn data = subset from the entire dataset used for training the model), that was averaged to create a partial dependence plot showing the partial contribution of a predictor to the abundance prediction (Friedman 2001; Elith et al. 2008).

3.3.6 Model evaluation and calibration

To establish the best fitting models for each species 60 or 50 percent of the data were randomly selected within TreeNet for training and the remaining 40 or 50 percent were used for testing the TreeNet models (Table A-3). For the RandomForests an "Out of bag" (OOB) approach was applied, where the randomly chosen OOB data are withheld from constructing the model and instead used for testing it (Oppel et al. 2009). We calculated the Mean Additive Deviation (MAD), MSE, and r^2 based on the withheld data. For evaluation of the model performance and for calibration, we plotted true values versus predicted values and fitted a simple linear regression to the data in R (Version 3.2.3). If most of the data fell above the 1:1 line, the model was overestimating abundance, while if most data fell below, the model was underestimating abundance. The resultant slope indicated the model consistency and the intercept represented the bias (Potts and Elith 2006; Piñeiro et al. 2008; Oppel et al. 2012). This derived formula was then applied to the predicted data to correct the final prediction. We also spatially overlaid the observed abundances with the values that were predicted onto the lattice and applied the same regression and 1:1 line as described before. We calculated the root-mean-square error (RMSE) and the r^2 values in order to assess the model performance. This was done for the predicted abundance derived from averaging the RandomForests and TreeNet model output.

3.4 Results

3.4.1 Predictive climatologies of pan-arctic copepod abundance and partial dependence plots

The overall predicted abundance for *C. finmarchicus* was highest in parts of the Arctic Ocean that are influenced by Atlantic water masses, as well as in the sub-arctic Atlantic where values reached as much as 107,190 m^{-2} (Fig. 3.3). Towards the high Arctic and through the Pacific Arctic, the abundance decreased, with large areas of low abundance. *C. finmarchicus* also had an extremely low abundance throughout most of the White Sea. According to the partial dependence plots for the three best-scoring predictors (Fig. 3.4), *C. finmarchicus* preferred areas with silicate concentrations between 3-6 µM in the surface to 50 m depth interval (Fig. 3.5), and salinity of 34 and higher (Fig. A-3). According to our environmental layers, these characteristics are associated with water masses in the Atlantic inflow.

C. glacialis displayed the highest abundance over most of the shelf areas throughout the Arctic Ocean and its adjacent Seas, with hotspots in the Barents Sea and north of the Chukchi Sea (Fig. 3.6). In these areas, the predicted abundance ranged from about 1,000 to over 29,000 m^{-2}. In most parts of the central Arctic Ocean, the abundance was lower and ranged from absent to 1,400 m^{-2}, as was the case north of Iceland. The partial dependence plots showed that *C. glacialis* preferred water masses that had silicate concentrations below 5 µM (surface to 50 m average) (Fig. A-5). It occurred in higher numbers at sites that had a bottom depth deeper than 1,000 m (Fig. A-6). The partial dependence was high with low freshwater discharge values (Fig. A-7).

The prediction for *C. hyperboreus* showed a relatively high abundance throughout most of the Arctic. The highest values were in the Canada Basin, Baffin Bay, the Greenland Sea, and south of Iceland (Fig. 3.7), where the values ranged from 1,000 to 6,731 m^{-2}. *C. hyperboreus* had the lowest abundance in areas with broad, relatively shallow shelves, such as large parts of the East Siberian, White, and Laptev Seas, and parts of the Kara, Barents, Chukchi and Bering Seas. According to the partial dependence plots, *C. hyperboreus* was described as a species that had a high abundance in waters deeper than 1,000 m (Fig A-8) and away from the shelf break towards the central Arctic (Fig. A-9). Based on the predictor for distance to hydrology, *C. hyperboreus* favored areas further away from the major rivers (Fig. A-10).

The prediction for *M. longa* showed the highest abundance between 1,000 to over 8,300 m^{-2} along the shelf break, parts of the Barents and White Seas, and south of Iceland as well as deeper areas such as the Norwegian Sea (Fig. 3.8). It was widely distributed and had a high abundance throughout most of the Arctic Ocean with the exceptions of shallow parts of the Laptev, Kara, and White Seas as well as the Bering Sea, parts of the Chukchi Sea, and the Beaufort Shelf. The partial dependence plots indicate that *M. longa* had a higher abundance in areas with a bottom depth of 200 m and deeper (Fig. A-11), a mean apparent oxygen utilization (surface to 50 m depth) of 0.2 and higher (Fig. A-12), and in areas with moderate to high freshwater discharge between 4,000 and 8,000 km^3 $year^{-1}$ (Fig. A-13).

For *M. pacifica*, we predicted very low abundance throughout most of the Arctic Ocean (Fig. 3.9), with highest abundances in the Chukchi and Bering Seas and the northern Sea of Okhotsk where abundances reached as high as 15,540 m^{-2}. It also showed a relatively high abundance in the North Atlantic, south of Iceland and along the Norwegian shelf with values ranging between 100 and 1,000 m^{-2}. According to the partial dependence plots, *M. pacifica* showed a higher abundance in areas with a larger distance from the shelf break towards the coast (Fig. A-14), and a mean phosphate concentration (surface to 50 m) larger than 1 µM (Fig. A-15), as well as a salinity of 31.5 and higher (surface to 50 m) (Fig. A-16). According to our environmental layers these characteristics are mostly found in the Bering and Chukchi Seas, as well as in between the transition from Atlantic to Arctic water masses.

The abundance pattern for *N. cristatus* looked similar to *M. pacifica* with the highest abundance found in the Chukchi, and Bering Seas, as well as in the northern Sea of Okhotsk (between 3 to 133 m^{-2}) (Fig. 3.10). It showed very low (< 0-1) abundance throughout most of the Arctic Ocean and adjacent Seas with higher abundance of 3-30 m^{-2} south of Iceland and in the southern Barents Sea. The partial dependence plots showed a higher abundance with increasing distance from the shelf break towards the coast with some low partial dependence values at about 500 km and around 600 km distance (Fig. A-17). It also showed a higher abundance for mean salinity (surface to 50 m) (Fig. A-18) of ~31 and then again at salinities ~ 32 PSU, which, as mentioned for *M. pacifica*, are mostly found in the Bering and Chukchi Seas, as well as in between the transition from Atlantic to Arctic water masses. It also showed a higher partial dependence at chlorophyll a concentrations between 0.5 and around 1 mg m^{-3} (Fig. A-19).

The prediction for *P. glacialis* showed highest abundance over the deeper areas of the Arctic Ocean such as the Canada and Makarov Basins (between 30 and 315 m^{-2}), but not the Nansen, Amundsen, and Norwegian Basins (Fig. 3.11). The predicted abundance decreased towards the coast in most areas and was low on broad, shallow shelfs (between 1 and 10 m^{-2}) with the lowest abundance found in most of the Chukchi, Bering, Laptev, East Siberian, and White Seas. The partial dependence plots confirmed the observation that *P. glacialis* had the highest abundance over deeper waters, and showed an increasing abundance with increasing distance to the coast (Fig. A-20), as well as in areas where the bottom depth was deeper than 500 m (Fig. A-21). These results are all consistent with the classification of *P. glacialis* as a deep-water species. The abundance was also higher in areas with a sea ice concentration of over 60% (Fig. A-22).

Comparing of the maximum abundance between species, *C. finmarchicus* had the highest maximum abundance, followed in descending order by *C. glacialis*, *M. pacifica*, *M. longa*, *C. hyperboreus*, *P. glacialis*, and *N. cristatus*. This ranking was reflective of the underlying observational abundance data. In general, the abundance for advected species (except *N. cristatus*) was more described by water properties (Fig. 3.4) in the three most important predictors, such as salinity, silicate, and phosphate. In contrast, the endemic species *C. glacialis*, *C. hyperboreus*, *M longa*, and *P. glacialis* tended to be defined by variables such as the distance to coast, distance to shelf break, bathymetry, distance to hydrology, and freshwater discharge. Bathymetry was within the two most important predictors for all endemic species, but for none of the expatriates. Some endemics (*C. glacialis* and *M. longa*) also have water properties such and apparent oxygen utilization within their top three predictors.

3.4.2 Overall modelling statistics

Most species displayed large ranges in both observed and modelled abundances (e.g. *C. finmarchicus* observations ranged from 0 to 336,896 ind. m^{-2}), thus a high statistical variance was to be expected. Higher MADs and MSEs (Table 3.4) were associated with a high range in observed abundance values and high mean abundance (e.g. *C. finmarchicus* and *C. glacialis*). Lower MADs and MSEs coincided with species that typically had lower observed abundances (e.g. *N. cristatus* and *P. glacialis*). The same pattern is seen in the RMSE (Table 3.5), where the highest RMSE's coincided with the species that had the highest mean abundance (*C. finmarchicus* and *C. glacialis*).

3.4.3 Confounding factors

Although we focused on environmental predictors, it is notable that when the year and taxonomist were incorporated into the models, the taxonomist was the most important predictor for all species. For most species year ranked among the top five variables. It was unclear if this was related to true changes over time or the higher availability of data after 2003 in several regions. In contrast, the effect mesh size employed by the sampling gear appeared negligible in the predictions. It was not an important predictor variable in most of the model runs and had no relative importance score assigned, except for *C. finmarchicus*, where the mesh size was the least important predictor with a relative importance score of 8.24%.

3.5 Discussion

3.5.1 Predictive maps of pan-arctic copepod distribution, partial dependence plots, and variable importance

Several studies have modelled zooplankton distribution on species level including the ecological niche (Speirs et al. 2006; Rutzen 2007; Reygondeau and Beaugrand 2011; Beaugrand et al. 2013), future distribution (Schmid 2012; Brun et al. 2016), and individual based modelling (Ji et al. 2012). Some of these studies applied machine-learning approaches (Rutzen 2007; Schmid 2012; Brun et al. 2016). Brun et al. (2016) hypothesized that it might be difficult to model zooplankton distribution due to their short life span and dispersion through ocean currents, a position in complete contrast to Richardson (2008). By applying machine-learning in our current study, we demonstrated that even with abundance data from numerous sources, and averaged environmental data, we were able to detect the major patterns of species distribution and abundance for our selected copepod species. Our model correctly detected whether a species was associated with shelf or deep-water habitat, and distinguished between Arctic endemic and seasonally advected species. Thus this approach opens up many opportunities, even when data quality is poor (Drew et al. 2011).

3.5.2 *Calanus finmarchicus*

The overall pattern of abundance for *C. finmarchicus* (Fig. 3.3) showed a good fit with the underlying abundance data. Its abundance was highest in the subarctic Atlantic and the Barents Sea, consistent with the characterization of *C. finmarchicus* being an Atlantic species that is advected to the Arctic (Kosobokova and Hirche 2000; Hirche and Kosobokova 2007). Modelled abundance showed general agreement with values obtained from publications that were not used in the modelling (Table 3.6), with most studies that estimated higher abundances encompassing earlier parts of the season when *C. finmarchicus* is typically more abundant (Arashkevich et al. 2004; Broms et al. 2009). Our model correctly predicted low abundance in the North Pacific and the Canada Basin (Grice 1962; Kosobokova and Hirche 2000), albeit with some small scattered hot spots. These small areas of high abundance in the Pacific and Canada Basin possibly point to a potential habitat for a sibling species (e.g. *C. marshallae*) or a niche that would be available if *C. finmarchicus* was to be introduced in that area. Notably, *C. finmarchicus* also had a low abundance in the White Sea, in agreement with independent findings (Jaschnov 1970), as well as the data employed in our modelling (Pertsova and Kosobokova 2003).

The predicted occurrence of the native distribution map for *C. finmarchicus* from SeaLifeBase (Kaschner et al. 2016) was similar to ours, with higher index of occurrences within the Barents Sea, Iceland Sea, Norwegian Sea, and along the Norwegian coast, as well as up the south Greenland coast. The low index of occurrence in the White Sea, Hudson and Baffin Bay is also comparable to ours, but SeaLifeBase showed an area of occurrence in the Chukchi Sea, that our prediction did not show. We suspect this reflects poor quality control of older datasets that list *C. finmarchicus* within the Chukchi prior to the erection of *C. glacialis* (Jaschnov 1955) and *C. marshallae* (Frost 1974).

The environmental predictors indicate that *C. finmarchicus* was more influenced by water mass characteristics than by geographical environmental variables such as bathymetry. *C. finmarchicus* was found to have the highest abundance at a salinity of ~35, which is representative of subarctic Atlantic water masses (Emery 2001). Similarly, the abundance was higher with lower silicate concentrations, which is also a reflection of Atlantic water that has lower silicate concentrations than Pacific water entering the Arctic through Bering Strait (Codispoti and Lowman 1973). Taken together, these confirm that our model was correctly associating *C. finmarchicus* with Atlantic water (Melle et al. 2004).

3.5.3 *Calanus glacialis*

The predicted abundance and distribution pattern of *C. glacialis* (Fig. 3.6) agreed with the pattern of most observational data. The prediction of the highest abundance over the shelf areas and lower abundance over the deep basins of the Greenland, Norwegian, and Iceland Seas was accurate, and is consistent with its characterization as a shelf species (Conover 1988; Hirche and Kwasniewski 1997; Melle and Skjoldal 1998; Melle et al. 2004; Falk-Petersen et al. 2007). Our prediction also correctly showed *C. glacialis* to be distributed throughout Hudson Bay, the

Canadian Archipelago, Foxe Basin, Baffin Bay, and Davis Strait as well as the White Sea (Buchanan and Sekerak 1982; Rochet and Grainger 1988; Harvey et al. 2001). Comparison of the abundance with independent values showed that our prediction was mostly within the same order of magnitude, or displayed the same patterns as the independent values (Table 3.6). Our prediction that *C. glacialis* occurs in the North Pacific Ocean, the Sea of Okhotsk, the Bering Sea, and the Chukchi Sea is supported by the literature (Jaschnov 1970; Frost 1974; Conover 1988; Nelson et al. 2009) as well as by our data for these regions (Table A-1), reflecting the genetically distinct North Pacific population of *C. glacialis* that is advected northward within Anadyr water entering through Bering Strait (Coachman et al. 1975; Nelson et al. 2009). In Western Baffin Bay the predicted abundance supported the observation that *C. glacialis* is more abundant there than *C. finmarchicus* (Table 3.6).

High abundance of *C. glacialis* tended to be associated with lower mean silicate concentration (surface to 50 m) (Fig. 3.16), which is typical for Atlantic water as well as Arctic surface water (Codispoti and Lowman 1973). Partial dependence increased rapidly with bathymetry at a depth of about 800 m with a peak at 1,000 m depth (Fig. 3.17). Predicted abundance seemed to be associated with lower freshwater discharge (Fig. 3.18), although the areas *C. glacialis* has been observed in cover a wide range of salinities (OBIS; www.iobis.org).

3.5.4 *Calanus hyperboreus*

The abundance pattern (Fig. 3.7) of *C. hyperboreus* largely fits with the underlying observed abundance data. *C. hyperboreus* is predominantly a deep-water species and known to be most abundant in the Greenland Sea and Arctic Ocean (Hirche 1991; Hirche and Mumm 1992; Thibault et al. 1999; Falk-Petersen et al. 2007). The abundance is low on the broad and shallow sections of the Chukchi, East Siberian, and parts of the Laptev and Kara Seas (Carmack et al. 2006). The low abundance in the White Sea was also coherent with the underlying observations (Pertsova and Kosobokova 2003). The high abundance in the Western Bering Sea as well as the Sea of Okhotsk was likely incorrect (Grice 1962) and, as with *C. finmarchicus*, would be pointing towards a possible niche or sibling species. The model fit observations reasonably well in Hudson Bay, Baffin Bay, the Canadian Archipelago (Table 3.6), but overestimates abundances south of Iceland (Gislason et al. 2009). In the Norwegian and Barents Sea, our abundance correctly reflected the pattern of *C. hyperboreus* having the lowest abundance in the Norwegian coastal water, and the highest in the Arctic water (Broms et al. 2009). Northeast of Spitsbergen, our model overestimated the abundance (Hirche and Kosobokova 2003). For the Western Baffin Bay, our predicted abundance fit very well with the independently observed data (Buchanan and Sekerak 1982).

The most important predictor for *C. hyperboreus* was bathymetry (Fig. 3.4), and the abundance increased with increasing depth (Fig. 3.19). This is reasonable, since *C. hyperboreus* is an oceanic species and more abundant in the basin areas instead of shelf areas. Also, bathymetry is known to be an important predictor for copepod distribution (Reygondeau and Beaugrand 2011), and to influence population growth (Smith and Schnack-Schiel 1990). The higher abundance with increasing distance from the shelf break towards the central Basin (Fig. 3.20) as well as increasing distance to hydrology (Fig. 3.21) also strengthens the assumption that the model correctly classified *C. hyperboreus* as a deep-water species.

3.5.5 *Metridia longa*

The overall prediction of *M. longa* (Fig. 3.8) showed a reasonable fit with the underlying data. *M. longa* is defined as an Arctic deep-water species with the Northern North Atlantic and the Arctic as the major areas of its distribution (Grice 1962). Our abundance fit well with the observed mean abundance north of Iceland (Table 3.6) (Gislason and Astthorsson 1998). In Baffin Bay, the abundance pattern of highest values occurring over the deep basin in Davis Strait, and the lowest in Greenland shelf water coincided with independent observations (Kjellerup et al. 2015). In Western Baffin Bay our abundance was only slightly lower than the observed abundance (Buchanan and Sekerak 1982). In the Northeast Atlantic our abundance of 1,000 to over 4,000 seemed to be elevated and driven by a lack of data in that area or missing environmental variables to describe the niche more accurately. However, the higher abundance in the Northwestern North Atlantic as well as in the Norwegian Sea coincided with Continuous Plankton Recorder (CPR) data from these regions (Hays 1995).

Bathymetry was clearly the most important predictor for *M. longa* (Fig. 4) with relatively high and stable partial dependence values from the surface to almost 4,000 m depth (Fig. 3.22). This is reasonable, since *M. longa* has been observed to be abundant in the upper 50 m of the water column as well as occurring at greater depths (Buchanan and Sekerak 1982). The second most important variable was apparent oxygen utilization (surface to 50 m), with higher partial dependence at medium to higher apparent oxygen utilization values (Fig. 3.23). The partial dependence for freshwater discharge was high with high freshwater discharge. This was displayed in the model with high predicted abundance partially coinciding with discharge of the major rivers. Fresh water discharge is known to have an influence on pelagic copepods (Smith and Schnack-Schiel 1990).

3.5.6 *Metridia pacifica*

The abundance of *M. pacifica* (Fig. 3.9) showed a good fit with the underlying data for the Bering and Chukchi Seas, but it did not reflect the observed data within the Atlantic very well. *M. pacifica* is described as an "oceanic copepod of the subarctic North Pacific" (Batchelder 1985), which is also distributed in the Bering and Chukchi Seas (Sirenko 2001; Hopcroft et al. 2010; Questel et al. 2013; Ershova et al. 2015a). Our model showed higher abundance in the Western Bering and Chukchi shelf, which are influenced by Anadyr water, and lower abundance in the Eastern Bering Sea, which agrees with earlier observations (Springer et al. 1989) (Table 3.6).

In the North Atlantic our model should have predicted the lowest abundance, but instead it indicated a suitable habitat based on environmental attributes. The high abundance in this case could be indicating the potential niche for its sibling species *M lucens* (Thorp 1980) that occurs in the North Atlantic.

The relative importance score and partial dependence plots showed that *M. pacifica* had higher abundance in water masses with moderate to high phosphate concentrations (Fig. 3.26), as well as salinities that are representative of inflowing Pacific waters (Fig. 3.27), which are known to be fresher and richer in nutrients than Atlantic water (McLaughlin et al. 1996; Ekwurzel et al. 2001; Emery 2001). This model finding is consistent with previous observations that *M. pacifica* is found in Pacific water masses (Ashjian et al. 2003).

3.5.7 *Neocalanus cristatus*

Neocalanus cristatus showed a distribution mostly as expected with the highest abundance in the subarctic Pacific, Bering Strait, and Chukchi Sea and very low abundance in the central Arctic Ocean (Fig. 3.10) (Sirenko 2001; Questel et al. 2013; Ershova et al. 2015a). The prediction for the Pacific Arctic was very good and mostly coincided with our underlying data. Our prediction confirmed that *N. cristatus* was most abundant in Anadyr waters along the Bering Sea and Chukchi shelf (Table 3.6) (Springer et al. 1989). The high predicted abundance of 133 m^{-2} in the Eastern Chukchi Sea appeared in an area that is influenced by Anadyr water. Our prediction suggested an area of higher abundance in Atlantic waters, similar to *M. pacifica*, but since *N. cristatus* is a Pacific species (Conover 1988), this just indicated that the environmental conditions in parts of the North Atlantic are such that *N. cristatus* could in principle thrive there, or that the niche should be occupied by a sibling species.

It was reasonable that *N. cristatus* showed elevated abundance with increasing distance from the shelf break towards the coast (away from the North Pole) (Fig. 3.28), since *N. cristatus* has higher observed abundances in these areas. As with *M. pacifica*, it is associated with lower salinities characteristic of Pacific waters (Fig. 3.29). These two factors correctly described *N. cristatus* as a Pacific species.

3.5.8 *Paraeuchaeta glacialis*

The overall pattern of our *P. glacialis* prediction (Fig. 3.11) reflected the trend of the underlying data. In general, our predicted distribution of *P. glacialis* also coincided with the literature: a broad distribution with hotspots in the deeper Arctic basins instead of shelf areas (Park 1993; Auel and Hagen 2002; Dvoretsky and Dvoretsky 2015). The low abundance in the Amundsen and Nansen Basin could be due to a lack of sufficient data in that region and because the few data that we had showed either very low abundance or absence. Our predicted abundance for Western Baffin Bay and between Svalbard and Greenland indicated a good fit with the literature (Table 3.6), while values south of Iceland are likely overestimated, since it is rare in that area (Park 1993). In the Western Kara Sea, our model potentially underestimated *P. glacialis* abundance, although the mean Dvoretsky and Dvoretsky (2015) reported was for *P. glacialis* and *P. norvegica* combined. Even though the majority of adult *Paraeuchaeta* spp. were *glacialis*, we cannot be certain how many were *norvegica*, and hence our model might not be underestimating by much. Since *P. glacialis* is absent from the Pacific Ocean (Grice 1962), the high abundance in the North Pacific indicated a potential habitat of *P. glacialis* or for a corresponding sibling species (e.g. *P. elongata*).

The model confirmed *P. glacialis* as a deep-water oceanic species (Kosobokova et al. 2011), since it showed an increasing abundance with increasing distance to the coast (Fig. 3.31), and increasing depth (Fig. 3.32). The high abundance with deeper bathymetry is reasonable since earlier developmental stages have been found as far down as 1 km in the central Greenland Sea (Auel 2004). Dvoretsky and Dvoretsky (2015) reported the highest abundances of *P. glacialis* in the Kara Sea at the stations with the deepest bottom depth. Considering this, our model made a correct association between the abundance and bathymetry.

3.5.9 Concluding discussion

Our approach succeeded in capturing the major patterns of zooplankton distribution at the scale of the circumpolar region, demonstrating what can be achieved with our approach even with “messy data” from numerous sources (and with a wide temporal coverage), and averaged environmental variables. Overall, the model performed well for most species. Our model correctly distinguished between seasonally advected species versus endemic species by assigning the highest scores to water mass properties for most advected species. It also correctly described oceanic, deep-water, and shelf species. Compared to independent literature, our model performed well and captured the major trends in the species distributions. Abundance for most species and regions were similar in magnitude to independent literature values. Even though the model performed well for the expatriated species in their native range, it frequently predicted hot spots of their distribution within the Pacific (*C. finmarchicus*) or Atlantic (*M. pacifica*, *N. cristatus*) where they are not currently observed. We encountered the same issue for the endemic species, where the model predicted the native range well, but also predicted high abundance in areas of the North Atlantic for example, where the species should show lower abundance. These areas suggest potential niches for the studied species or the niche of a sibling species. To further enhance the predictive success of our model, we would require more high quality abundance and environmental data. Additional environmental variables that could possibly improve our model performance are residence time or “time of entry”, location of polynyas, ocean currents, fronts, upwelling areas, dispersal time, predation, and competition, as well as a detailed distribution and abundance of the prey species such as diatoms and other phytoplankton.

Improved spatial coverage of the underlying abundance data might be the best approach to improve our models. Especially for the Arctic endemics, additional data in the North Atlantic should increase model performance. Over-concentrated data in a particular region could lead to the association of numerous abundance data with only a few environmental variables

(Humphries et al. 2012), and thus not correctly represent the niche, although ideally machine learning should be able to handle such situations (Drew and Collazo 2012). A way to better characterize the species-habitat link, as well as to obtain a better spatial coverage might, have been to reduce our abundance model to a presence/absence model. Since confirmed absence data and accurate abundance estimates are often hard to obtain (Huettmann and Gottschalk 2011), presence/absence or presence/random modeling is a good way to utilize distribution data (e.g. Hardy et al. 2011; Huettmann et al. 2011), and is also less influenced by the accuracy of abundance estimates. Those models usually serve as good indicators for bounding conditions of species occurrences.

We were not surprised that the taxonomist received a high relative importance score, since other studies have also shown that the taxonomist can be a very important factor in such questions (e.g. Bluhm et al. 2011). We find this indicates that strong bias occurs between different laboratories and taxonomists, however it is equally plausible that it is a confounded effect, because most taxonomists only work in certain regions of the Arctic Ocean and over a finite period of time. Since the taxonomist is a categorical predictor, it could also be a “greedy” predictor (Steinberg 2016), meaning that the categorical predictor is ranked as more important than it actually is due to confounding spatial and temporal associations. We found that sample year was estimated as an important predictor for most species. It is possible that interannual or decadal differences and variation are driving this and will need further investigation. Notably, the partial dependences for *C. hyperboreus* and *M. longa* switched from negative to positive in the mid-1970s, the same time as the shift in the Pacific Decadal Oscillation (Mantua and Hare 2002). To better ascertain such changes, we would need to look on a finer scale in time and in space (e.g. Ershova et al. 2015b) with concurrently measured environmental parameters, or with better spatial and temporal data coverage of plankton as well as the GIS layers. Ideally, our modelling should have been conducted for each month separately, as well as divided temporally by decades, but our ecological niche is to be robust overall. Also, the data became too sparse and too patchy for such a modeling approach, hence we required compression of months and years to generate a good spatial coverage. Sample coverage could be improved particularly in the North Atlantic if we extended our seasonal sampling window forward because many surveys are focused on the earlier spring bloom at lower latitudes (e.g. Melle and Skjoldal 1989; Melle and Skjoldal 1998).

Frequently, it is argued that datasets cannot be combined due to differences in collecting gear, especially the mesh size employed. It is notable that for the larger-bodied species examined in this study the mesh size did not receive a relative importance score, indicating a negligible effect compared to other variables. This confirms our assumption that the mesh size is unimportant to detecting patterns with larger copepods (Ershova et al. 2015b) and when presence-only approaches are used. Loss of animals between mesh sizes is clearly a concern for smaller-bodied species (Hopcroft et al. 2005; Skjoldal et al. 2013; Miloslavić et al. 2014), but of little importance when estimating abundance of medium and larger bodied zooplankton (Antacli et al. 2010). Even though smaller stages of copepods (CI and CII) are not captured by larger mesh sizes (Nichols and Thompson 1991), the main pattern of high and low abundance still emerges from our data when applying machine learning approaches. Our finding has large implications, beyond plankton and the Arctic.

Combining abundance data from various sources poses numerous challenges beyond differing mesh sizes. The data exist in different formats, some only in the form of microfiche or paper reports. Different sampling techniques and different taxonomists could introduce errors into the data. The level of identification often varied widely between working groups and taxonomists, although the dominant copepods were most consistently well-documented. Missing or incomplete metadata made it difficult and problematic to use some datasets, for example missing gear, mesh size, time of sampling, and some cases missing units of the reported abundance (e.g. m^3 vs m^2) (see also Huettmann 2009; Bluhm et al. 2010; Zuckerberg et al. 2011).

This paper highlights the need for more comprehensive metadata, data sharing, structured sampling strategy, and standardized sampling techniques in order to detect long term changes in the Arctic Ocean ecosystem beyond gross patterns. The issue of missing or inadequate metadata and the need for inter-comparable data is well known, and finding solutions has become critically important (see Wiebe et al. 2015). Often, projects collect ample data, but the data are not used beyond initial publications or reports, and are never broadly distributed – this practice must be replaced with one of data permanency and availability, as well as open access data sharing. A repeatable and transparent science must be the goal and is best professional practice.

This study's success in detecting patterns within major zooplankton species demonstrates the ability of our approach to gain useful and new information from somewhat imperfect data.

Using machine learning algorithms, we predicted the zooplankton abundance well, and described the ecological niche by drawing inference from the model output on how the environmental variables were associated with the abundance (Breiman 2001b). The maps we presented in this paper can now be used as first quantitative base layers to predict the distribution of whales and other planktivorous animals (e.g. Best et al. 2012). These maps can be further improved by incorporating the considerable reservoirs of data still in other researchers' hands, while rescuing older data and unprocessed samples. Incorporation of such data, and improvement of environmental layers, may allow this model to resolve long-term temporal shifts, and later even predict future species distributions when used in conjunction with climate projected environmental layers. Lastly, the methods we employed can and should be applied to different Oceanic regions beyond the Arctic, and to a variety of species other than zooplankton as has been done before in numerous studies (e.g. Yen et al. 2004; Huettmann et al. 2011; Humphries et al. 2012; Oppel et al. 2014; Baltensperger and Huettmann 2015). Here we offer a template and data to be applied and tested further.

Acknowledgements

This work was funded through National Science Foundation Award OPP0909571, a University of Alaska Fairbanks (UAF) graduate fellowship and a UAF Graduate School thesis completion fellowship to Imme Rutzen, and the Census of Marine Life's Arctic Ocean Diversity project.

We especially thank all data providers without whom an effort such as this would not have been possible. We thank C. Clarke-Hopcroft for data assistance, S. Oppel for input on early drafts of the manuscript and modelling expertise, and G. Humphries, A. Baltensperger, and M. Schmid for specific modelling expertise.

Figures

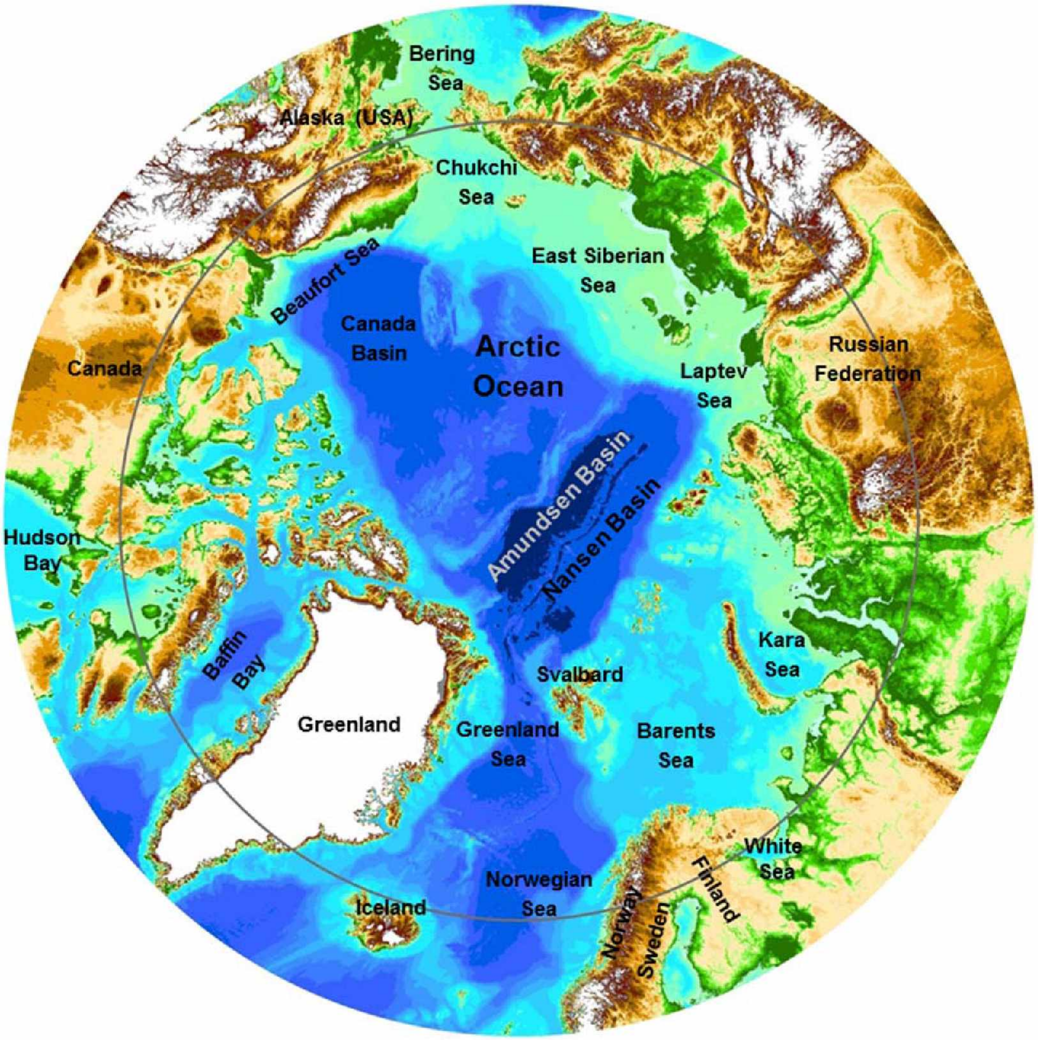


Figure 3.1: Study Area (Arctic Ocean and adjacent Seas) with Arctic Circle noted. Bathymetry derived from GEBCO

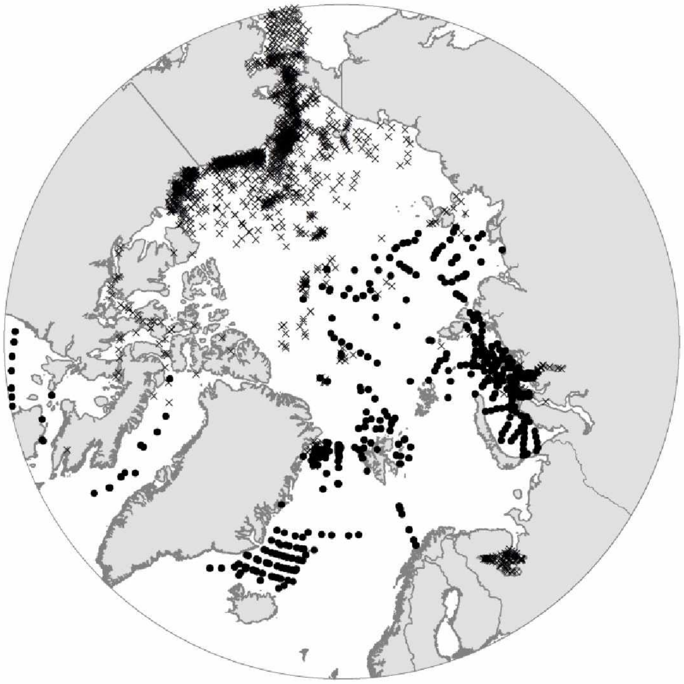


Figure 3.2: Station location distribution (datapoints) for abundance data (black dots) and absence data (black crosses) for *Calanus finmarchicus*

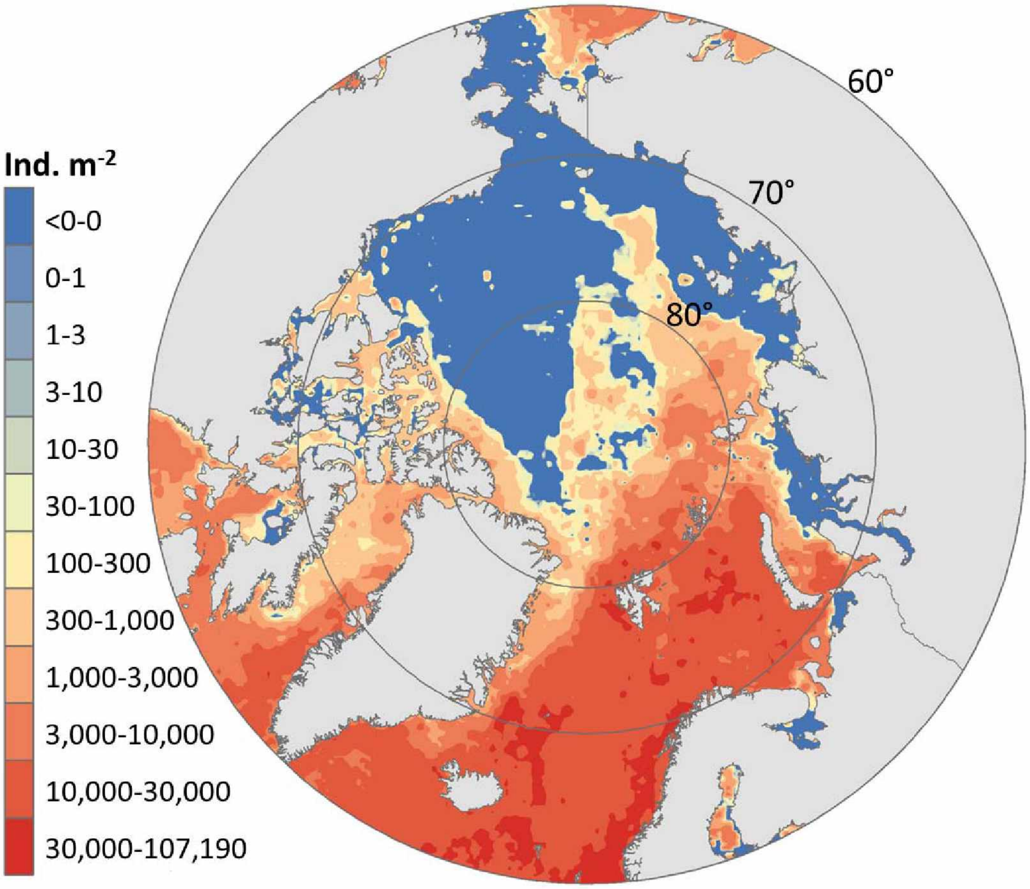
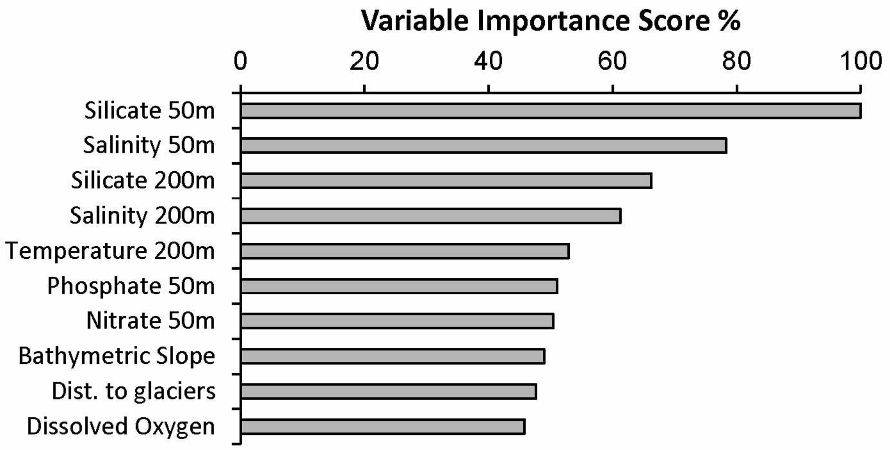


Figure 3.3: Predicted summer pan-arctic abundance for *Calanus finmarchicus* displayed on logarithmic scale

Calanus finmarchicus



Calanus glacialis

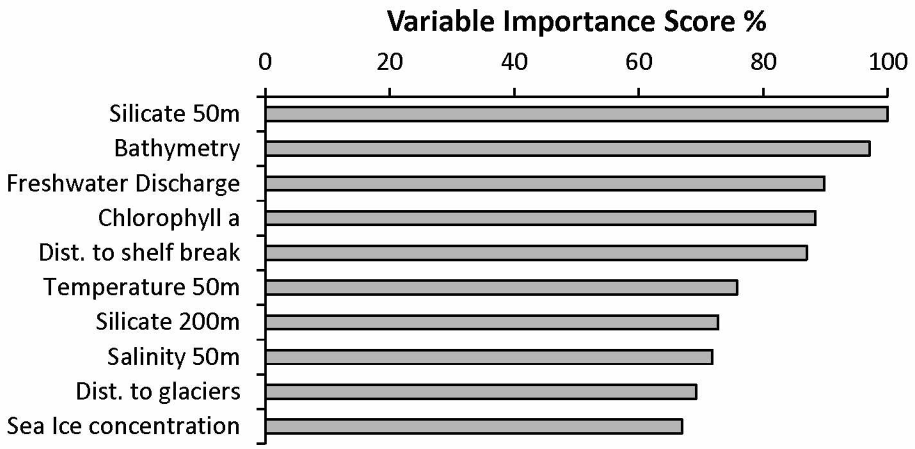
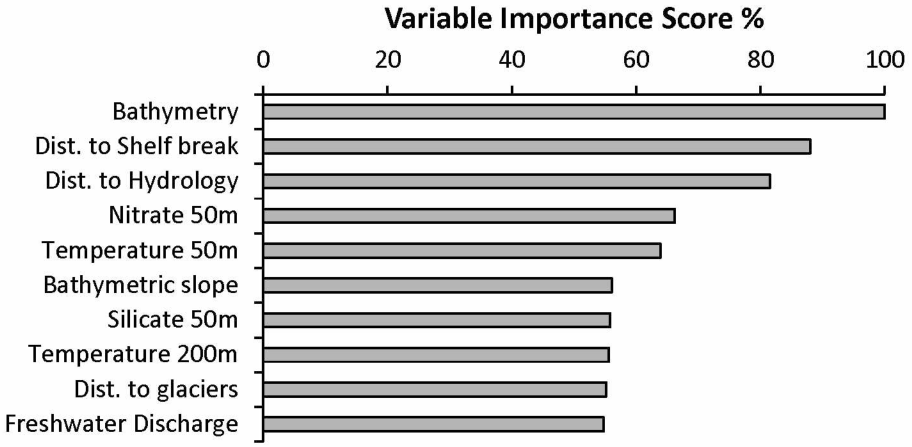


Figure 3.4: TreeNet relative importance scores for the 10 most important environmental variables per species (for models with taxonomist and year excluded) in percent

Figure 3.4: cont.

Calanus hyperboreus



Metridia longa

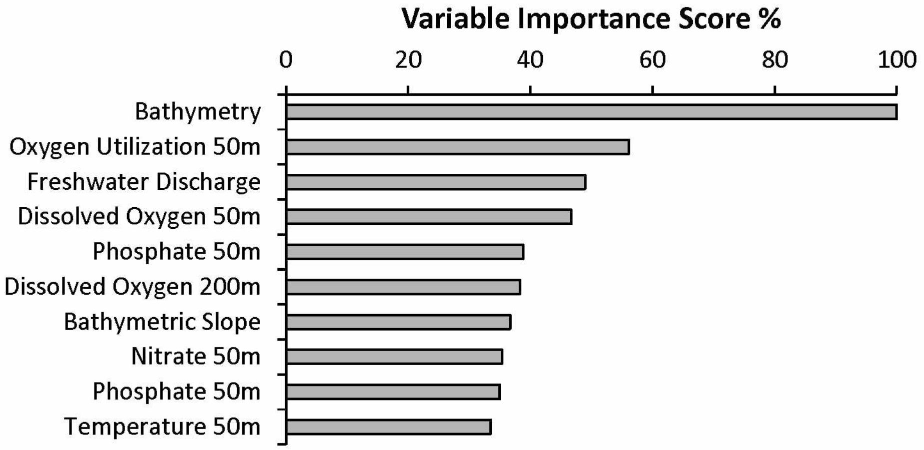
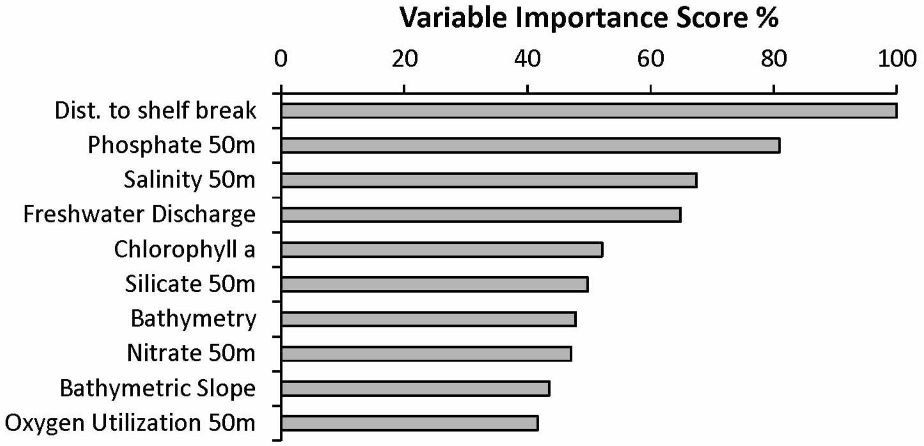


Figure 3.4: cont.

Metridia pacifica



Neocalanus cristatus

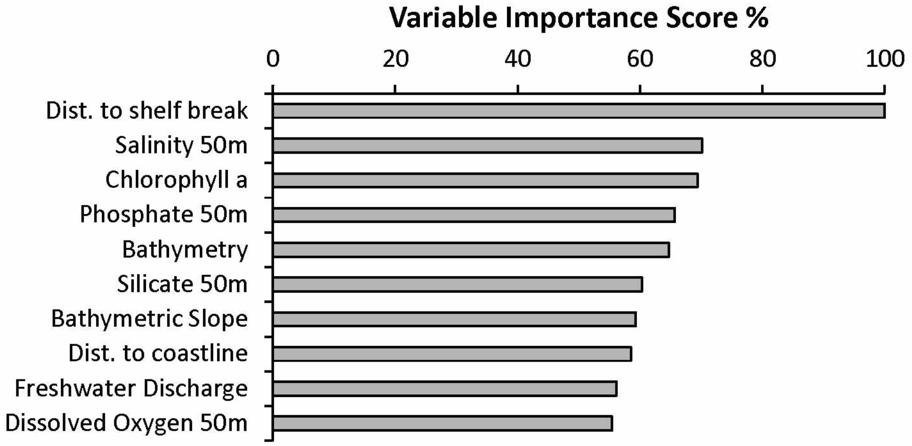
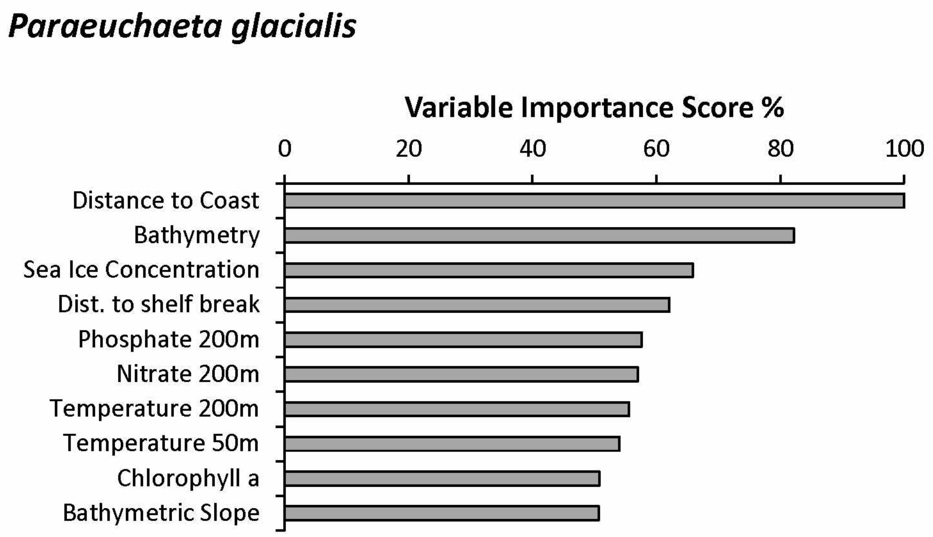


Figure 3.4: cont.



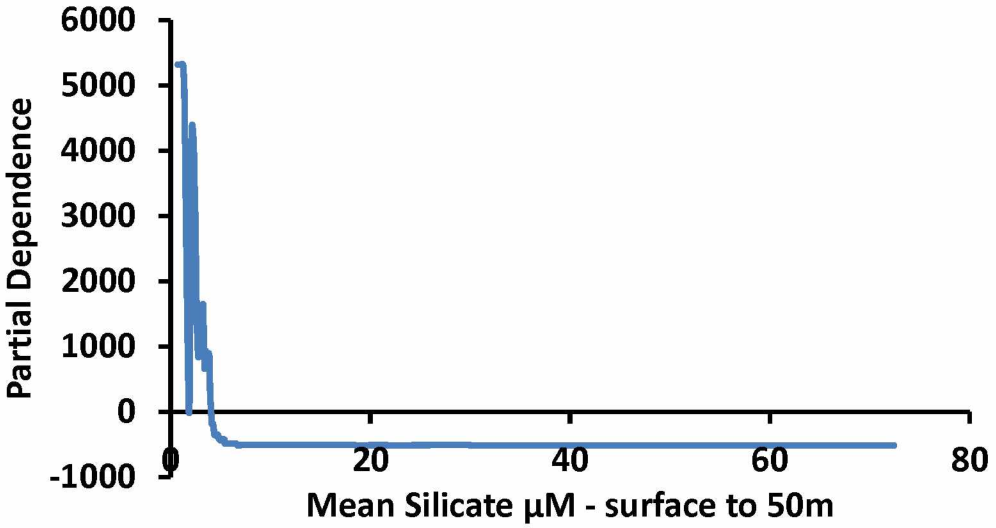


Figure 3.5: Partial dependence plot for mean silicate concentration (surface to 50 m depth) in µM for *Calanus finmarchicus*

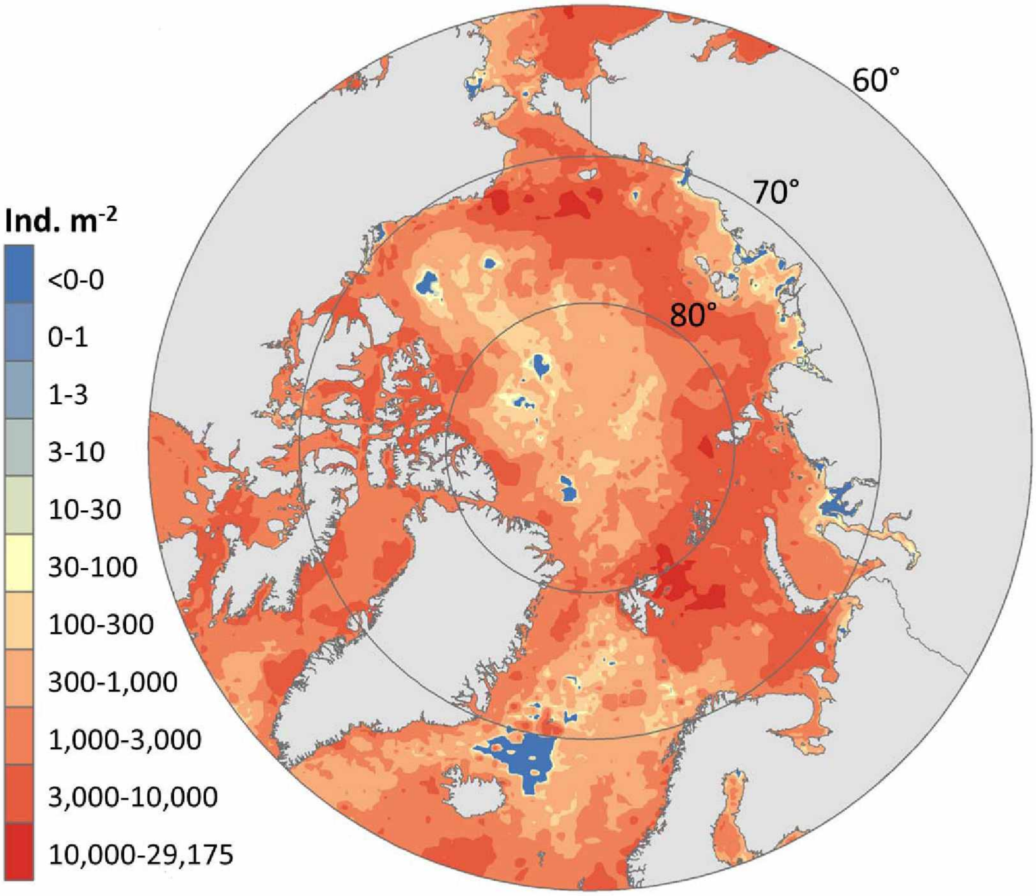


Figure 3.6: Predicted summer pan-arctic abundance for *Calanus glacialis* displayed on logarithmic scale

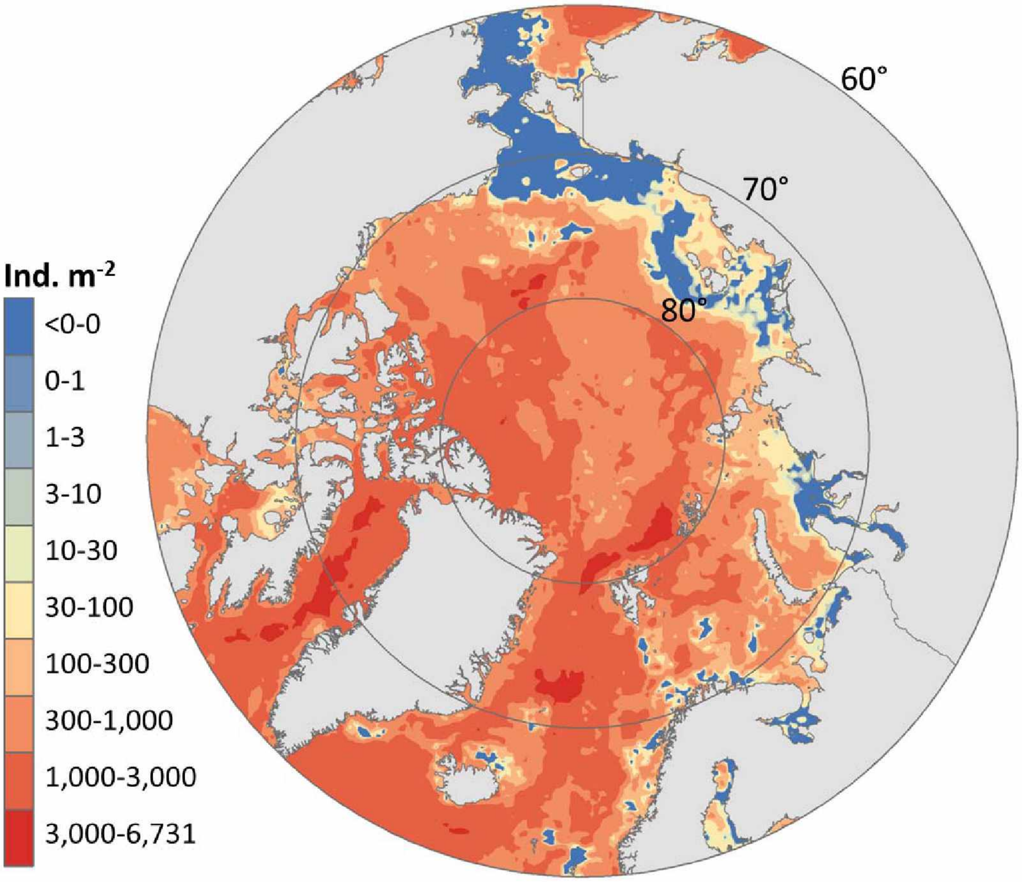


Figure 3.7: Predicted summer pan-arctic abundance for *Calanus hyperboreus* displayed on logarithmic scale

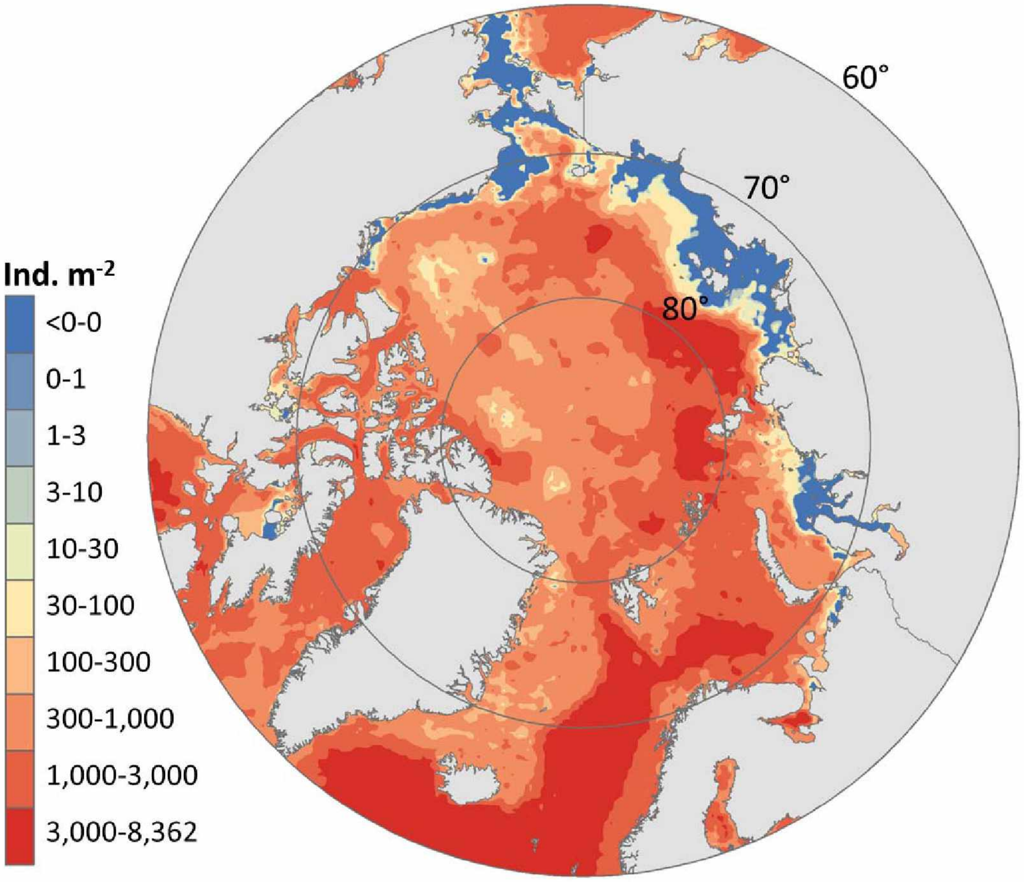


Figure 3.8: Predicted summer pan-arctic abundance for *Metridia longa* displayed on logarithmic scale

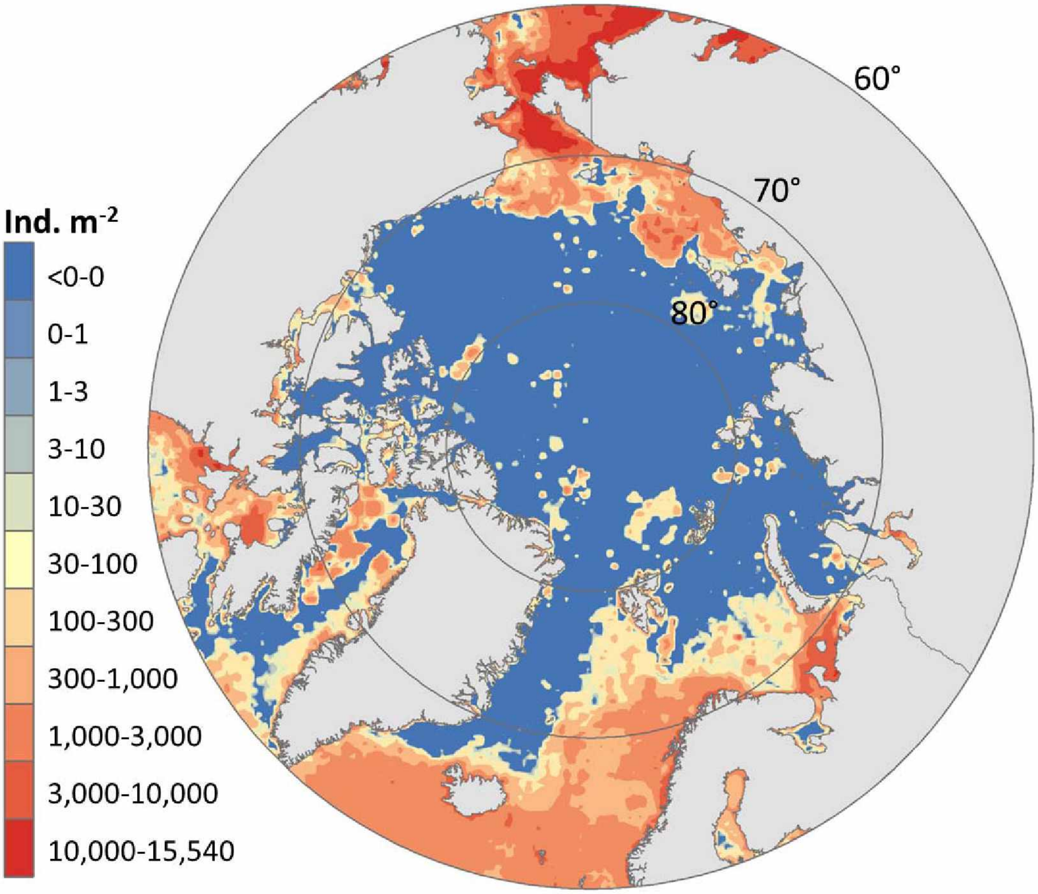


Figure 3.9: Predicted summer pan-arctic abundance for *Metridia pacifica* displayed on logarithmic scale

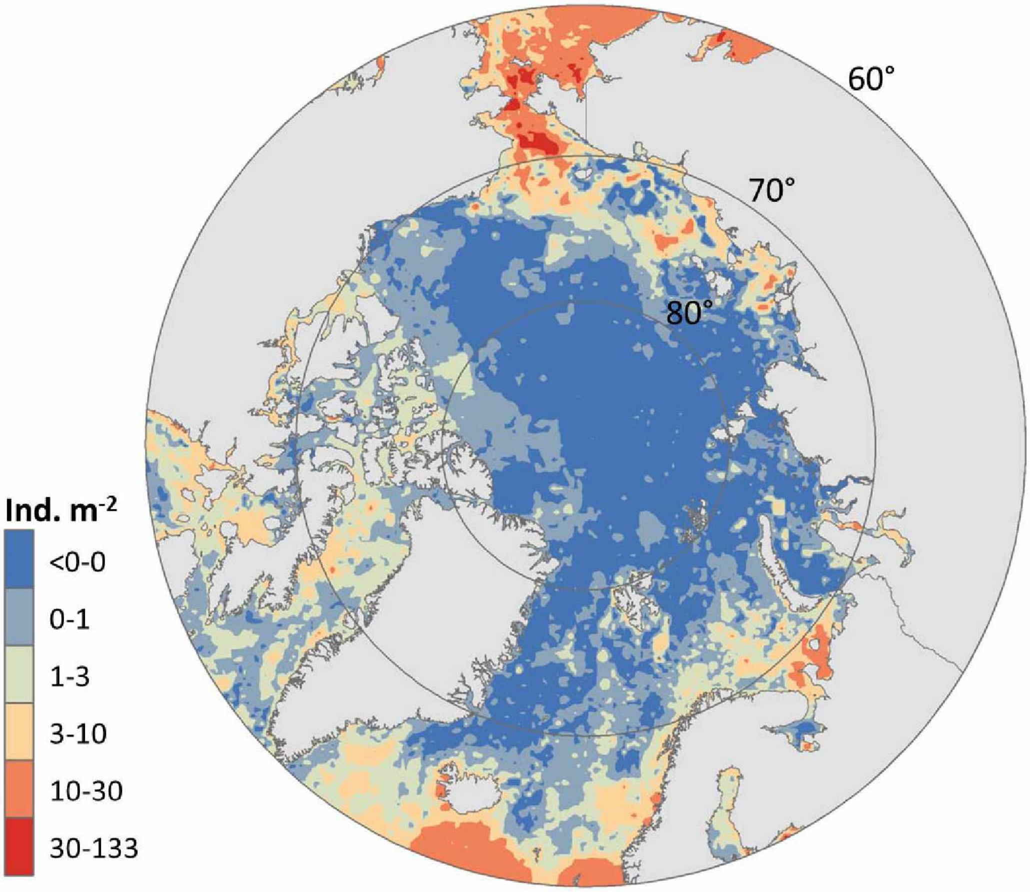


Figure 3.10: Predicted summer pan-arctic abundance for *Neocalanus cristatus* displayed on logarithmic scale

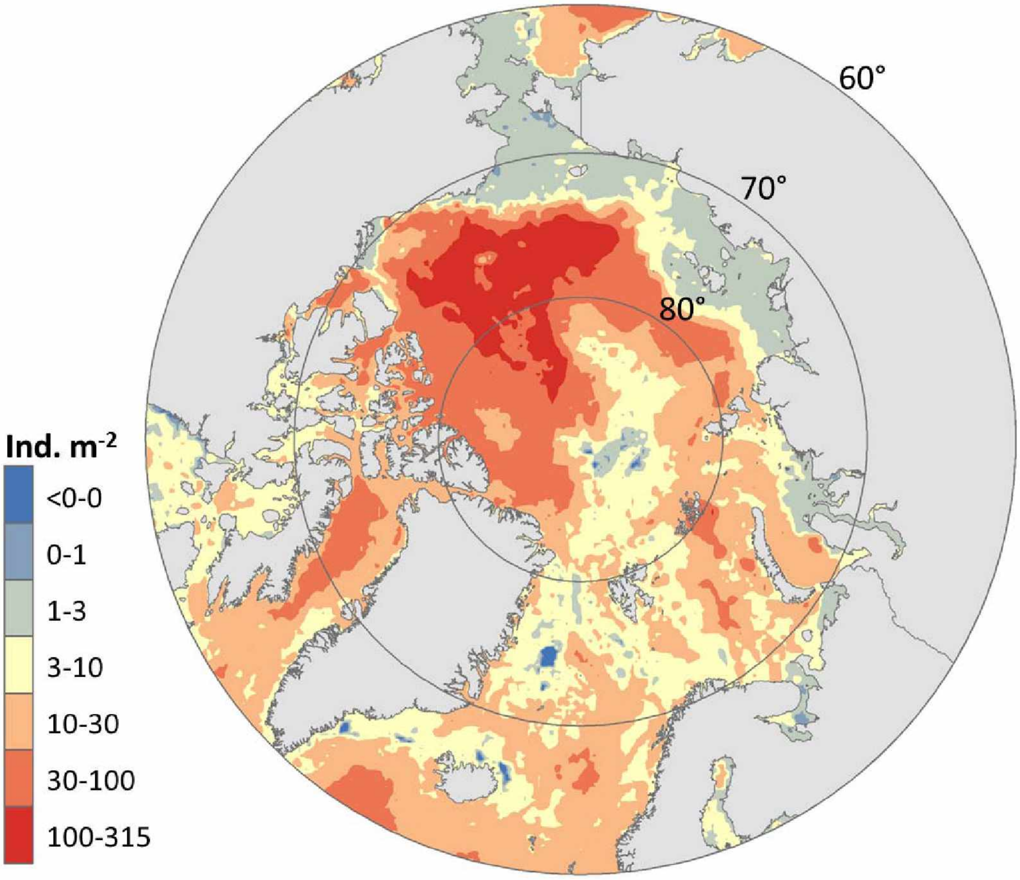


Figure 3.11: Predicted summer pan-arctic abundance for *Paraeuchaeta glacialis* displayed on logarithmic scale

Tables

Table 3.1: Overview of data archiving projects with corresponding URL or reference

Data archiving project	URL or reference
Archives of the Arctic Seas - Zoological Institute of Russian Academy of Sciences	Contributions from the Zoological Institute RAS. 2005. No 8. 44 p.
BioChem archive - Fisheries and Oceans Canada (DFO)	http://www.dfo-mpo.gc.ca/science/data-donnees/biochem/index-eng.html
Coastal and Oceanic Plankton Ecology, Production, and Observation Database (COPEPOD) by the National Oceanic and Atmospheric Administration (NOAA)	http://www.st.nmfs.noaa.gov/copepod/
Arctic Ocean Biodiversity Project (ArcOD) – a Census of Marine Life project	http://www.arcodiv.org/
Ocean Biogeographic Information System (OBIS)	http://www.iobis.org/
Pacific Marine Arctic Regional Synthesis (PacMARS)	http://pacmars.cbl.umces.edu/
PANGAEA – Data publisher for Earth and Environmental science	https://www.pangaea.de/

Table 3.2: Overview of copepod species used for modelling pan-arctic abundances

Scientific name	**ITIS taxonomic serial number (TSN)**	**WORMS AphiaID**	**Arctic endemic species**	**Known Occurrence**	**Reproducing in Arctic**	**Selected citations**
Calanus finmarchicus	85272	104464	no	North Atlantic, Norwegian, Barents, Kara, Laptev, East Siberian, and White Seas.	no	Conover (1988), Hirche and Kosobokova (2007), Sirenko (2001)
Calanus glacialis	85267	104465	yes	Arctic Ocean and adjacent Seas	yes	Conover (1988), Sirenko (2001)
Calanus hyperboreus	85266	104467	yes	Arctic Ocean and adjacent Seas	yes	Conover (1988), Sirenko (2001)
Metridia longa	85746	104632	yes	Arctic Ocean and adjacent Seas	yes	Daase et al. (2008), Sirenko (2001)
Metridia pacifica	85748	196784	no	North Pacific, Bering Sea, Chukchi Sea	no	Hopcroft et al. (2010), Sirenko (2001)
Neocalanus cristatus	667083	104470	no	North Pacific, Bering Sea, Chukchi Sea	no	Conover (1988), Sirenko (2001)
Paraeuchaeta glacialis	85539	104560	yes	Arctic Ocean and adjacent Seas (excluding White Sea)	yes	Auel and Hagen (2002), Sirenko (2001)

Table 3.3: Source and URL of the Environmental variables

Environmental Variable	Source	URL	Spatial resolution	Temporal resolution	Units
Salinity	NOAA World Ocean Atlas 2013 (WOA13)	http://www.nodc.noaa.gov/OC5/woa13/	1°	July - September 1955-2012	PSU
Temperature			1°	July - September 1955-2012	°C
Nitrate concentration			1°	July - September 1955-2012	µM
Phosphate concentration			1°	July - September 1955-2012	µM
Silicate concentration			1°	July - September 1955-2012	µM
Dissolved Oxygen (amount of gaseous O_2 dissolved in water)			1°	July - September 1955-2012	mL/L
Apparent Oxygen Utilization (remineralized component of O_2)			1°	July - September 1955-2012	mL/L
Percent Oxygen Saturation (% of dissolved O_2 relative to complete saturation at measured temperature and salinity)			1°	July - September 1955-2012	%

Table 3.3 continued…

Euclidean Distance to Glacier	Calculated from National Snow and Ice Data Center (NSIDC)-World Glacier Inventory in ArcGIS	http://nsidc.org/data/docs/noaa/g01130_glacier_inventory/		1900 - 2003	Meters
Euclidean Distance to Coastline	Calculated from World Vector Shoreline in ArcGIS	http://rimmer.ngdc.noaa.gov/coast/ http://shoreline.noaa.gov/data/datasheets/wvs.html		Invariable	Meters
Euclidean Distance to Hydrology	Calculated from Global lakes and wetlands database Level 2	http://www.worldwildlife.org/publications/global-lakes-and-wetlands-database-small-lake-polygons-level-2		Invariable	Meters
Bathymetry	General Bathymetric Chart of the Oceans (GEBCO_08)	http://www.gebco.net/	30 arc-seconds	Invariable	Meters
Slope	Calculated from Bathymetry in ArcGIS		30 arc-seconds	Invariable	Degrees
Freshwater Discharge	R-ArcticNET – Version 4	http://www.r-arcticnet.sr.unh.edu/	10 km	1877 - 2003	km^3 /year
Sea Ice concentration	National Snow and Ice Data Center	http://nsidc.org/data/smmr_ssmi_ancillary/monthly_means.html		1979 - 2013	%
Chlorophyll a	Aqua MODIS	http://oceancolor.gsfc.nasa.gov/cgi/l3	4 km	Seasonal climatology Summer (July-September) 2002-2014	mg/m^3

Table 3.4: Evaluation values from RandomForests (RF) and TreeNet (TN) models showing Mean Absolute Deviation (MAD), Mean Squared Error (MSE), r^2, and percent of data randomly selected for testing (%test) per species. MAD and MSE values were rounded to the nearest whole number

Species	**MAD RF**	**MAD TN**	**MSE RF**	**MSE TN**	**r^2 RF**	**r^2 TN**	**%test RF**	**%test TN**
C. finmarchicus	1456	1595	57489564	120567605	-0.08	0.35	50	40
C. glacialis	2658	2300	34725626	35529348	0.003	0.12	40	50
C. hyperboreus	375	316	1100788	1291980	0.2	0.27	30	40
M. longa	524	422	2586528	2803432	0.19	0.21	30	50
M. pacifica	499	326	8741603	4564663	0.13	0.15	30	50
N. cristatus	6	4	395	486	-0.08	0.03	50	60
P. glacialis	11	9	1132	1342	0.2	0.25	40	50

Table 3.5: Root-mean-square error (RMSE) and r^2 for spatial overlay of observed abundances and predicted abundances (from RandomForests and TreeNet model output combined after calibration). RMSE rounded to the nearest whole number

Species	**r^2**	**RMSE**
Calanus finmarchicus	0.40	9913
Calanus glacialis	0.14	8192
Calanus hyperboreus	0.46	997
Metridia longa	0.18	2162
Metridia pacifica	0.25	1916
Neocalanus cristatus	0.18	20
Paraeuchaeta glacialis	0.4	35

Table 3.6: Comparison of predicted abundance and independent observed abundances from literature resources for each species. Predicted abundance values are rounded. Abundance values from the literature were converted to ind. m^{-2}, averaged if necessary, and rounded

Species	**Region**	**Predicted Abundance ind. m^{-2}**	**Abundance ind. m^{-2} from literature**	**Temporal coverage**	**Citation**
Calanus finmarchicus	North of Iceland	11,000 - 21,000	7,600	July-September 1994	Gislason and Astthorsson (1998)
	Northeast of Iceland	18,000 - 28,000	16,000	July 1995	Astthorsson and Gislason (2003)
	Norwegian and Barents Sea – Coastal Water	9,000 - > 50,000	77,959	March – August 1995	Broms *et al.* (2009)
	Norwegian and Barents Sea – Atlantic Water	16,000 - > 36,000	109,950	March – August 1995	Broms *et al.* (2009)
	Norwegian and Barents Sea – Arctic Water	14,000 - > 27,000	51,690	March – August 1995	Broms *et al.* (2009)
	Northeast of Svalbard	14,500	~5,000	July 1st 1997	Hirche and Kosobokova (2003)
	Western Baffin Bay – Eastern Lancaster Sound	293 - > 2,000	169	July – October 1978	Buchanan and Sekerak (1982)
	Pacific	< 0 - > 6,000	absent	August – September 1960	Grice (1962)
Calanus glacialis	Northeast of Iceland	~30 - 1,900	380	June to July 1995	Astthorsson and Gislason (2003)
	Norwegian and Barents Sea – Coastal Water	112 - > 6,000	26	March – August 1995	Broms *et al.* (2009)
	Norwegian and Barents Sea – Atlantic Water	240 - 1,800	92	March – August 1995	Broms *et al.* (2009)
	Norwegian and Barents Sea – Arctic Water	42 - ~ 1,800	198	March – August 1995	Broms *et al.* (2009)
	Northeast of Svalbard	~ 9,000	3,000	July 1st 1997	Hirche and Kosobokova (2003)
	Western Baffin Bay – Eastern Lancaster Sound	2,000 - ~2,500	4,154	July – October 1978	Buchanan and Sekerak (1982)
Calanus hyperboreus	Eastern Coast of Hudson Bay	200 - 1,700	0-1,704	September 1993	Harvey *et al.* (2001)
	North of Iceland	102 - ~900	242	July – September 1994	Gislason and Astthorsson (1998)
	Northeast of Iceland	< 0 - ~400	max ~ 8,000	June – July 1996	Astthorsson and Gislason (2003)
	Norwegian and Barents Sea – Coastal Water	< 0 - ~700	51	March – August 1995	Broms *et al.* (2009)

Table 3.6 continued…

Calanus hyperboreus	Norwegian and Barents Sea – Atlantic Water	200 – ~1,800	1,074	March – August 1995	Broms *et al.* (2009)
	Norwegian and Barents Sea – Arctic Water	1,800 - ~6,000	9,048	March – August 1995	Broms *et al.* (2009)
	Northeast of Svalbard	600 - ~1,000	100	July 1st 1997	Hirche and Kosobokova (2003)
	Western Baffin Bay – Eastern Lancaster Sound	1,600 - 3,500	2,946	July – October 1978	Buchanan and Sekerak (1982)
Metridia longa	North of Iceland	600 - ~1,000	600	July – September 1994	Gislason and Astthorsson (1998)
	Eastern Coast of Hudson Bay and Hudson Strait	600 - ~ 4,000	4,331 Range: 0 – 16,823	September 1993	Harvey *et al.* (2001)
	Western Baffin Bay – Eastern Lancaster Sound	1,300 – 2,500	2,961.5	July – October 1978	Buchanan and Sekerak (1982)
	Canadian Shelf Water	500 - ~ 2,000	1,850	September 2009	Kjellerup *et al.* (2015)
	Davis Strait – Deep Basin	1,000 - > 1,500	2,750	September 2009	Kjellerup *et al.* (2015)
	Greenland Shelf Water	250 - ~1,000	795	September 2009	Kjellerup *et al.* (2015)
Metridia pacifica	Western Bering and Chukchi shelf	1,000 - ~6,000	1,360	July and August 1985 and 1986	Springer *et al.* (1989)
	Eastern Bering and Chukchi shelf	< 0 - > 3,000	19	July and August 1985 and 1986	Springer *et al.* (1989)
Neocalanus cristatus	Western Bering and Chukchi shelf	3 - > 130	36	July and August 1985 and 1986	Springer *et al.* (1989)
	Eastern Bering and Chukchi shelf	< 0 – ~80	2	July and August 1985 and 1986	Springer *et al.* (1989)
Paraeuchaeta *glacialis*	Western Kara Sea	5 - ~25	104	August 2012	Dvoretsky and Dvoretsky (2015)
	Western Baffin Bay – Eastern Lancaster Sound	25 - ~50	42	July – October 1978	Buchanan and Sekerak (1982)
	North Pacific	< 0 - > 25	absent	August – September 1960	Grice (1962)

References

Aagaard K (1989) A synthesis of the Arctic Ocean circulation. Rapp P-v Réun Cons int Explor Mer 188:11-22

Antacli J, Hernández D, Sabatini M (2010) Estimating copepods' abundance with paired nets: Implications of mesh size for population studies. J Sea Res 63:71-77

Arashkevich EG, Tande KS, Pasternak AF, Ellertsen B (2004) Seasonal moulting patterns and the generation cycle of *Calanus finmarchicus* in the NE Norwegian Sea, as inferred from gnathobase structures, and the size of gonads and oil sacs. Mar Biol 146:119-132 doi:10.1007/s00227-004-1416-5.

Ashjian CJ, Campbell RG, Welch HE, Butler M, Van Keuren D (2003) Annual cycle in abundance, distribution, and size in relation to hydrography of important copepod species in the western Arctic Ocean. Deep-Sea Res., Part I 50:1235-1261

Astthorsson OS, Gislason A (2003) Seasonal variations in abundance, development and vertical distribution of Calanus finmarchicus, C. hyperboreus and C. glacialis in the East Icelandic Current. J. Plankton Res 25: 843-854

Auel H (2004) Egg size and reproductive adaptations among Arctic deep-sea copepods (Calanoida, *Paraeuchaeta*). Helgol Mar Res 58:147-153 doi:10.1007/s10152-004-0179-1.

Auel H, Hagen W (2002) Mesozooplankton community structure, abundance and biomass in the central Arctic Ocean. Mar Biol 140:1013-1021 doi:10.1007/s00227-001-0775-4.

Baltensperger A, Huettmann F (2015) Predictive spatial niche and biodiversity hotspot models for small mammal communities in Alaska: applying machine-learning to conservation planning. Landscape Ecol 30:681-697 doi:10.1007/s10980-014-0150-8.

Batchelder HP (1985) Seasonal abundance, vertical-distribution, and life-history of *Metridia pacifica* (Copepoda, Calanoida) in the oceanic subarctic Pacific. Deep-Sea Res., Part A 32:949-964

Beaugrand G, Mackas D, Goberville E (2013) Applying the concept of the ecological niche and a macroecological approach to understand how climate influences zooplankton: Advantages, assumptions, limitations and requirements. Prog Oceanogr 111:75-90 doi:http://dx.doi.org/10.1016/j.pocean.2012.11.002.

Beaugrand G, Reid PC, Ibañez F, Lindley JA, Edwards M (2002) Reorganization of North Atlantic marine copepod biodiversity and climate. Science 296:1692-1694 doi:10.1126/science.1071329.

Best BD et al. (2012) Online cetacean habitat modeling system for the US East Coast and Gulf of Mexico. Endanger Species Res 18:1-15

Bluhm B, Watts D, Huettmann F (2010) Free database availability, metadata and the internet: An example of two high latitude components of the Census of Marine Life. In: Cushman S, Huettmann F (eds) Spatial Complexity, Informatics, and Wildlife Conservation. Springer Japan, pp 233-243. doi:10.1007/978-4-431-87771-4_13.

Bluhm BA et al. (2011) Diversity of the arctic deep-sea benthos. Mar Biodiv 41:87-107 doi:10.1007/s12526-010-0078-4.

Breiman L (2001a) Random Forests. Machine Learning 45:5-32 doi:10.1023/a:1010933404324.

Breiman L (2001b) Statistical modeling: The two cultures. Statistical science 16:199-231

Broms C, Melle W, Kaartvedt S (2009) Oceanic distribution and life cycle of *Calanus* species in the Norwegian Sea and adjacent waters. Deep-Sea Res., Part II 56:1910-1921 doi:http://dx.doi.org/10.1016/j.dsr2.2008.11.005.

Brun P, Kiørboe T, Licandro P, Payne MR (2016) The predictive skill of species distribution models for plankton in a changing climate. Glo Change Biol 22:3170-3181 doi:10.1111/gcb.13274.

Buchanan RA, Sekerak AD (1982) Vertical distribution of zooplankton in Eastern Lancaster Sound and Western Baffin Bay, July - October 1978. Arctic 35:15

Carmack E, Barber D, Christensen J, Macdonald R, Rudels B, Sakshaug E (2006) Climate variability and physical forcing of the food webs and the carbon budget on panarctic shelves. Progr Oceanogr 71:145-181 doi:http://dx.doi.org/10.1016/j.pocean.2006.10.005.

Coachman LK, Aagaard K, Tripp RB (1975) Bering Strait: the regional physical oceanography. University of Washington Press, Seattle

Codispoti L, Lowman D (1973) A reactive silicate budget for the Arctic Ocean. Limnol Oceanogr 18:448-456

Conover RJ (1988) Comparative life histories in the genera *Calanus* and *Neocalanus* in high-latitudes of the northern hemisphere. Hydrobiologia 167:127-142

Craig E, Huettmann F (2008) Using "blackbox" algorithms such as TreeNet and Random Forests for data-mining and for finding meaningful patterns, relationships and outliers in complex ecological data: an overview, an example using golden eagle satellite data and an outlook for a promising future. In: Wang H-F (ed) Intelligent Data Analysis: Developing New Methodologies Through Pattern Discovery and Recovery. IGI Global, Hershey, PA, USA, pp 65-84. doi:10.4018/978-1-59904-982-3.ch004.

Daase M, Eiane K, Aksnes DL, Vogedes D (2008) Vertical distribution of *Calanus* spp. and *Metridia longa* at four Arctic locations. Mar Biol Res 4:193-207 doi:10.1080/17451000801907948.

Dalpadado P et al. (2014) Productivity in the Barents Sea-response to recent climate variability. PLoS ONE 9: e95273 doi.org/10.1371/journal.pone.0095273

De'ath G (2007) Boosted Trees for ecological modeling and prediction. Ecology 88:243-251 doi:10.1890/0012-9658(2007)88[243:btfema]2.0.co;2.

Drew CA, Collazo JA (2012) Expert knowledge as a foundation for the management of secretive species and their habitat. In: Perera AH, Drew CA, Johnson CJ (eds) Expert Knowledge and Its Application in Landscape Ecology. Springer New York, New York, pp 87-107. doi:10.1007/978-1-4614-1034-8_5.

Drew CA, Wiersma YF, Huettmann F (2011) Conclusion: An attempt to describe the state of habitat and species modeling today. In: Drew CA, Wiersma YF, Huettmann F (eds) Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications. Springer New York, New York, NY, pp 291-298. doi:10.1007/978-1-4419-7390-0_15.

Dvoretsky VG, Dvoretsky AG (2015) Summer population structure of the copepods *Paraeuchaeta* spp. in the Kara Sea. J Sea Res 96:18-22 doi:http://dx.doi.org/10.1016/j.seares.2014.10.010.

Ekwurzel B, Schlosser P, Mortlock RA, Fairbanks RG, Swift JH (2001) River runoff, sea ice meltwater, and Pacific water distribution and mean residence times in the Arctic Ocean. J Geophys Res-Oceans 106:9075-9092 doi:10.1029/1999jc000024.

Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. J Anim Ecol 77:802-813 doi:10.1111/j.1365-2656.2008.01390.x.

Emery W (2001) Water types and water masses. In: Steele JH, Thorpe S, Turekian KK (eds) Encyclopedia of ocean sciences, vol 6. Elsevier San Diego, Calif, pp 3179-3187

Ershova EA, Hopcroft RR, Kosobokova KN (2015a) Inter-annual variability of summer mesozooplankton communities of the western Chukchi Sea: 2004–2012. Polar Biol 38:1461-1481 doi:10.1007/s00300-015-1709-9.

Ershova EA, Hopcroft RR, Kosobokova KN, Matsuno K, Nelson RJ, Yamaguchi A, Eisner LB (2015b) Long-term changes in summer zooplankton communities of the Western Chukchi Sea, 1945-2012. Oceanography 28:100-115

Fabry VJ, Seibel BA, Feely RA, Orr JC (2008) Impacts of ocean acidification on marine fauna and ecosystem processes. ICES J Mar Sci 65:414-432 doi:10.1093/icesjms/fsn048

Falk-Petersen S, Mayzaud P, Kattner G, Sargent JR (2008) Lipids and life strategy of Arctic *Calanus*. Mar Biol Res 5:18-39 doi:10.1080/17451000802512267.

Falk-Petersen S, Pavlov V, Timofeev S, Sargent J (2007) Climate variability and possible effects on arctic food chains: The role of *Calanus*. In: Ørbæk J, Kallenborn R, Tombre I, Hegseth E, Falk-Petersen S, Hoel A (eds) Arctic alpine ecosystems and people in a changing environment. Springer Berlin Heidelberg, pp 147-166. doi:10.1007/978-3-540-48514-8_9.

Friedman J (2001) Greedy Function Approximation: A Gradient Boosting Machine. Ann Stat 29:1189-1232 doi:10.1214/aos/1013203451.

Frost BW (1974) *Calanus marshallae*, a new species of calanoid copepod closely allied to the sibling species *C. finmarchicus* and *C. glacialis*. Mar Biol 26:77-99 doi:10.1007/bf00389089.

Gislason A, Astthorsson O (1998) Seasonal variations in biomass, abundance and composition of zooplankton in the subarctic waters north of Iceland. Polar Biol 20:85-94

Gislason A, Petursdottir H, Astthorsson OS, Gudmundsson K, Valdimarsson H (2009) Inter-annual variability in abundance and community structure of zooplankton south and north of Iceland in relation to environmental conditions in spring 1990-2007. J Plankton Res 31:541-551 doi:10.1093/plankt/fbp007.

Grainger E (1989) Vertical distribution of zooplankton in the central Arctic Ocean. In: Proceedings of the sixth conference of the Comité Arctique International, 1989 1989. Brill, pp 48-60

Grice G (1962) Copepods collected by the nuclear submarine Seadragon on a cruise to and from the North Pole, with remarks on their geographic distribution. J Mar Res 20:97-109

Hardy SM, Lindgren M, Konakanchi H, Huettmann F (2011) Predicting the distribution and ecological niche of unexploited Snow Crab (*Chionoecetes opilio*) populations in Alaskan Waters: a first Open-Access Ensemble Model. Integr Comp Biol 51:608-622 doi:10.1093/icb/icr102.

Harvey M, Therriault J-C, Simard N (2001) Hydrodynamic control of late summer species composition and abundance of zooplankton in Hudson Bay and Hudson Strait (Canada). J Plankton Res 23:481-496 doi:10.1093/plankt/23.5.481.

Hays GC (1995) Ontogenetic and seasonal variation in the diel vertical migration of the copepods *Metridia lucens* and *Metridia longa*. Limnol Oceanogr 40:1461-1465

Hirche H-J (1991) Distribution of dominant calanoid copepod species in the Greenland Sea during late fall. Polar Biol 11:351-362

Hirche H-J, Bohrer RN (1987) Reproduction of the Arctic copepod *Calanus glacialis* in Fram Strait. Mar Biol 94:11-17

Hirche H-J, Kosobokova K (2003) Early reproduction and development of dominant calanoid copepods in the sea ice zone of the Barents Sea - need for a change of paradigms? Mar Biol 143:769-781

Hirche H-J, Kosobokova K (2007) Distribution of *Calanus finmarchicus* in the northern North Atlantic and Arctic Ocean—Expatriation and potential colonization. Deep-Sea Res., Part II 54:2729-2747 doi:http://dx.doi.org/10.1016/j.dsr2.2007.08.006.

Hirche H-J, Kwasniewski S (1997) Distribution, reproduction and development of *Calanus* species in the northeast water in relation to environmental conditions. J Marine Syst 10:299-317 doi:http://dx.doi.org/10.1016/S0924-7963(96)00057-7.

Hirche H-J, Niehoff B (1996) Reproduction of the Arctic copepod *Calanus hyperboreus* in the Greenland Sea-field and laboratory observations. Polar Biol 16:209-219 doi:10.1007/bf02329209.

Hirche H-J, Mumm N (1992) Distribution of dominant copepods in the Nansen Basin, Arctic-Ocean, in summer. Deep-Sea Res., Part A 39:S485-S505

Hopcroft RR, Roff JC, Chavez FP (2001) Size paradigms in copepod communities: a re-examination. Hydrobiologia 453:133-141

Hopcroft RR, Clarke C, Nelson RJ, Raskoff KA (2005) Zooplankton communities of the Arctic's Canada Basin: the contribution by smaller taxa. Polar Biol 28:198-206

Hopcroft RR, Kosobokova KN, Pinchuk AI (2010) Zooplankton community patterns in the Chukchi Sea during summer 2004. Deep-Sea Res., Part II 57:27-39 doi:http://dx.doi.org/10.1016/j.dsr2.2009.08.003.

Hopkins TL, Gartner JV (1992) Resource-partitioning and predation impact of a low-latitude myctophid community. Mar Biol 114:185-197 doi:10.1007/bf00349518

Huettmann F (2009) The global need for, and appreciation of, high-quality metadata in biodiversity. In: Spehn E, Körner C (eds) Data Mining for Global Trends in Mountain Biodiversity. CRC Press, Boca Raton, pp 25-28

Huettmann F, Artukhin Y, Gilg O, Humphries G (2011) Predictions of 27 Arctic pelagic seabird distributions using public environmental variables, assessed with colony data: a first digital IPY and GBIF open access synthesis platform. Mar Biodiv 41:141-179 doi:10.1007/s12526-011-0083-2.

Huettmann F, Diamond AW (2001) Seabird colony locations and environmental determination of seabird distribution: a spatially explicit breeding seabird model for the Northwest Atlantic. Ecol Model 141:261-298

Huettmann F, Gottschalk T (2011) Simplicity, model fit, complexity and uncertainty in spatial prediction models applied over time: We are quite sure, aren't we? In: Drew CA, Wiersma YF, Huettmann F (eds) Predictive species and habitat modeling in landscape ecology, 1st edn. Springer, New York, pp 189-208

Humphries GW, Deal C, Elliott S, Huettmann F (2012) Spatial predictions of sea surface dimethylsulfide concentrations in the high arctic. Biogeochem 110:287-301 doi:10.1007/s10533-011-9683-y.

Hunt GL et al. (2011) Climate impacts on eastern Bering Sea foodwebs: a synthesis of new data and an assessment of the Oscillating Control Hypothesis. ICES J Mar Sci doi:10.1093/icesjms/fsr036.

Jaschnov W (1955) Morphology, distribution and systematics of *Calanus finmarchicus* sl. Zool Zh 34:1210-1223

Jaschnov WA (1970) Distribution of *Calanus* species in the Seas of the Northern Hemisphere. Int Revue ges Hydrobiol 55:197-212 doi:10.1002/iroh.19700550203.

Ji R et al. (2012) Life history and biogeography of *Calanus* copepods in the Arctic Ocean: An individual-based modeling study. Prog Oceanogr 96:40-56 doi:http://dx.doi.org/10.1016/j.pocean.2011.10.001.

Jones EP (2001) Circulation in the arctic ocean. Polar Res 20:139-146

Karnovsky N, Kwazniewski S, Weslawski J, Walkusz W, Beszczynska-Möller A (2003) Foraging behavior of little auks in a heterogeneous environment. Mar Ecol: Prog Ser 253:289-303

Kaschner K, Kesner-Reyes K, Garilao C, Rius-Barile J, Rees T, Froese R (2016) Computer generated distribution maps for *Calanus finmarchicus*, with modelled year 2100 native range map based on IPCC A2 emissions scenario. version of Aug. 2013, www.aquamaps.org

Kjellerup S et al. (2015) Vertical and horizontal distribution of zooplankton and polar cod in southern Baffin Bay (66–71°N) in September 2009. Polar Biol 38:699-718 doi:10.1007/s00300-014-1633-4.

Kosobokova K, Hirche H-J (2000) Zooplankton distribution across the Lomonosov Ridge, Arctic Ocean: species inventory, biomass and vertical structure. Deep-Sea Res., Part I 47:2029-2060

Kosobokova KN, Hirche H-J (2001) Reproduction of *Calanus glacialis* in the Laptev Sea, Arctic Ocean. Polar Biol 24:33-43

Kosobokova KN, Hopcroft RR (2010) Diversity and vertical distribution of mesozooplankton in the Arctic's Canada Basin. Deep-Sea Res., Part II 57:96-110 doi:http://dx.doi.org/10.1016/j.dsr2.2009.08.009.

Kosobokova KN, Hopcroft RR, Hirche H-J (2011) Patterns of zooplankton diversity through the depths of the Arctic's central basins. Mar Biodivers 41:29-50

Kristiansen I, Gaard E, Hátún H, Jónasdóttir S, Ferreira ASA (2015) Persistent shift of *Calanus* spp. in the southwestern Norwegian Sea since 2003, linked to ocean climate. ICES J Mar Sci doi:10.1093/icesjms/fsv222.

Leu E, Søreide JE, Hessen DO, Falk-Petersen S, Berge J (2011) Consequences of changing sea-ice cover for primary and secondary producers in the European Arctic shelf seas: Timing, quantity, and quality. Prog Oceanogr 90:18-32 doi:http://dx.doi.org/10.1016/j.pocean.2011.02.004.

Mantua NJ, Hare SR (2002) The Pacific Decadal Oscillation. J Oceanogr 58:35-44 doi:10.1023/a:1015820616384.

Masatoshi M, Kouichi K, Hikaru W, Akinori O (2000) Feeding habits of three dominant myctophid fishes, *Diaphus theta*, *Stenobrachius leucopsarus* and *S. nannochir*, in the subarctic and transitional waters of the western North Pacific. Mar Eol: Prog Ser 207:129-140

Mauchline J (1998) Advances in marine biology - The biology of calanoid copepods - Introduction In: Blaxter JHS, Southward AJ, Tyler PA (eds) Advances In Marine Biology. Academic Press Ltd, London, pp 1-710

McLaughlin FA, Carmack EC, Macdonald RW, Bishop JK (1996) Physical and geochemical properties across the Atlantic/Pacific water mass front in the southern Canadian Basin. J Geophys Res: Oceans 101:1183-1197

Melle W, Ellertsen B, Skjoldal HR (2004) Zooplankton: The link to higher trophic levels. In: Skjoldal HR (ed) The Norwegian Sea Ecosystem. Tapir Academic Press, Trondheim, pp 137-202

Melle W, Skjoldal HR (1989) Zooplankton reproduction in the Barents Sea: vertical distribution of eggs and nauplii of *Calanus finmarchicus* in relation to spring phytoplankton development. Olsen Int S:137-145

Melle W, Skjoldal HR (1998) Reproduction and development of *Calanus finmarchicus*, *C. glacialis* and *C. hyperboreus* in the Barents Sea. Mar Ecol: Prog Ser 169:211-228

Miller KB, Huettmann F, Norcross BL (2015) Efficient spatial models for predicting the occurrence of subarctic estuarine-associated fishes: implications for management. Fisheries Manag Ecol 22:501-517 doi:10.1111/fme.12148.

Miloslavić M, Lučić D, Gangai B, Onofri I (2014) Mesh size effects on mesozooplankton community structure in a semi-enclosed coastal area and surrounding sea (South Adriatic Sea). Mar Ecol 35:445-455 doi:10.1111/maec.12101.

Mueter FJ et al. (2009) Ecosystem responses to recent oceanographic variability in high-latitude Northern Hemisphere ecosystems. Prog Oceanogr 81:93-110 doi:http://dx.doi.org/10.1016/j.pocean.2009.04.018.

Nelson RJ et al. (2014) Biodiversity and biogeography of the lower trophic taxa of the Pacific Arctic region: sensitivities to climate change. In: Grebmeier JM, Maslowski W (eds) The Pacific Arctic Region. Springer, Dordrecht; New York, pp 269-336

Nelson RJ, Carmack EC, McLaughlin FA, Cooper GA (2009) Penetration of Pacific zooplankton into the western Arctic Ocean tracked with molecular population genetics. Mar Ecol:Prog Ser 381:129-138

Nichols J, Thompson A (1991) Mesh selection of copepodite and nauplius stages of four calanoid copepod species. J Plankton Res 13:661-671

Oppel S et al. (2014) Estimating population size of a nocturnal burrow-nesting seabird using acoustic monitoring and habitat mapping. Nat Conserv 7:1-13 doi:10.3897/natureconservation.7.6890.

Oppel S, Huettmann F (2010) Using a Random Forest model and public data to predict the distribution of prey for marine wildlife management. In: Cushman SA, Huettmann F (eds) Spatial Complexity, Informatics, and Wildlife Conservation. Springer Japan, Tokyo, pp 151-163. doi:10.1007/978-4-431-87771-4_8.

Oppel S, Meirinho A, Ramírez I, Gardner B, O'Connell AF, Miller PI, Louzao M (2012) Comparison of five modelling techniques to predict the spatial distribution and abundance of seabirds. Biol Conserv 156:94-104 doi:http://dx.doi.org/10.1016/j.biocon.2011.11.013

Oppel S, Strobl C, Huettmann F (2009) Alternative methods to quantify variable importance in ecology. Department of Statistics, University of Munich

Park T (1993) Taxonomy and distribution of the marine calanoid copepod family Euchaetidae. vol 29. University of California Press, pp1-203

Pertsova NM, Kosobokova KN (2003) Zooplankton of the White Sea: Features of the composition and structure, seasonal dynamics, and the contribution to the formation of matter fluxes. Oceanology 43:S108-S122

Piñeiro G, Perelman S, Guerschman JP, Paruelo JM (2008) How to evaluate models: observed vs. predicted or predicted vs. observed? Ecol Model 216:316-322

Potts JM, Elith J (2006) Comparing species abundance models. Ecol Model 199:153-163 doi:http://dx.doi.org/10.1016/j.ecolmodel.2006.05.025.

Questel JM, Clarke C, Hopcroft RR (2013) Seasonal and interannual variation in the planktonic communities of the northeastern Chukchi Sea during the summer and early fall. Cont Shelf Res 67:23-41 doi:http://dx.doi.org/10.1016/j.csr.2012.11.003.

Reygondeau G, Beaugrand G (2011) Future climate-driven shifts in distribution of *Calanus finmarchicus*. Glob Change Biol 17:756-766 doi:10.1111/j.1365-2486.2010.02310.x.

Richardson AJ (2008) In hot water: zooplankton and climate change. ICES J Mar Sci 65:279-295 doi:10.1093/icesjms/fsn028.

Ringuette M et al. (2002) Advanced recruitment and accelerated population development in Arctic calanoid copepods of the North Water. Deep-Sea Res., Part II 49:5081-5099

Rochet M, Grainger EH (1988) Community structure of zooplankton in eastern Hudson Bay. Can J Zool 66:1626-1630 doi:10.1139/z88-237.

Rogachev KA, Carmack EC, Foreman MGG (2008) Bowhead whales feed on plankton concentrated by estuarine and tidal currents in Academy Bay, Sea of Okhotsk. Cont Shelf Res 28:1811-1826 doi:http://dx.doi.org/10.1016/j.csr.2008.04.014.

Rudels B, Anderson LG, Jones EP (1996) Formation and evolution of the surface mixed layer and halocline of the Arctic Ocean. J Geophys Res: Oceans 101:8807-8821 doi:10.1029/96jc00143.

Rutzen I (2007) Predictive GIS-modeling of arctic mesozooplankton based on available public online data: Experiences from four copepod species (*Calanus glacialis*, *C. hyperboreus*, *Metridia longa*, and *M. pacifica*). University of Veterinary Medicine Hannover, Foundation

Schmid M (2012) Model-predicting the effect of freshwater inflow on saltwater layers, migration and life history of zooplankton in the Arctic Ocean: Towards scenarios and future trends. Georg-August-Universitaet Goettingen

Seidov D et al. (2015) Oceanography north of 60°N from World Ocean Database. Prog Oceanogr 132:153-173 doi:http://dx.doi.org/10.1016/j.pocean.2014.02.003.

Sirenko BI (2001) List of species of free-living invertebrates of Eurasian Arctic seas and adjacent deep waters. vol 51. Explorations of the Fauna of the Seas. Russian Academy of Science, Zoological Institute, St. Petersburg

Skjoldal HR, Wiebe PH, Postel L, Knutsen T, Kaartvedt S, Sameoto DD (2013) Intercomparison of zooplankton (net) sampling systems: Results from the ICES/GLOBEC sea-going workshop. Prog Oceanogr 108:1-42

Smith SL, Schnack-Schiel SB (1990) Polar Zooplankton. In: Smith WOJ (ed) Polar Oceanography, Part B: Chemistry, Biology, and Geology, vol 2. Academic Press, San Diego (CA), pp 527-598

Smoot CA (2015) Contemporary Mesozooplankton Communities of the Beaufort Sea. University of Alaska, Fairbanks

Speirs DC, Gurney WS, Heath MR, Horbelt W, Wood SN, De Cuevas BA (2006) Ocean-scale modelling of the distribution, abundance, and seasonal dynamics of the copepod *Calanus finmarchicus*. Mar Ecol:Prog Ser 313

Springer AM, McRoy CP, Turco KR (1989) The paradox of pelagic food webs in the northern Bering Sea—II. Zooplankton communities. Cont Shelf Res 9:359-386 doi:http://dx.doi.org/10.1016/0278-4343(89)90039-3.

Springer AM, Murphy EC, Roseneau DG, McRoy CP, Cooper BA (1987) The paradox of pelagic food webs in the northern Bering Sea—I. Seabird food habits. Cont Shelf Res 7:895-911 doi:http://dx.doi.org/10.1016/0278-4343(87)90005-7.

Steinberg D (2016) Modeling tricks with TreeNet: Treating Categorical Variables as Continuous. vol 2017. Salford Systems, Inc., San Diego, CA, USA

Strobl C, Boulesteix A-L, Zeileis A, Hothorn T (2007) Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics 8:1-21 doi:10.1186/1471-2105-8-25.

Thibault D, Head EJH, Wheeler PA (1999) Mesozooplankton in the Arctic Ocean in summer. Deep-Sea Res., Part I 46:1391-1415

Thorp AC (1980) Comparative morphology of sibling species of *Metridia* (Copepoda: Calanoida). *M. lucens*, *M. pacifica* and a species indeterminata from Indian Arm, British Columbia. University of British Columbia

Wassmann P et al. (2015) The contiguous domains of Arctic Ocean advection: Trails of life and death. Prog Oceanogr 139:42-65 doi:http://dx.doi.org/10.1016/j.pocean.2015.06.011.

Wei C-L et al. (2011) Global patterns and predictions of seafloor biomass using Random Forests. PLoS ONE 5:e15323 doi:10.1371/journal.pone.0015323.

Wiebe PH, Allison D, Kennedy M, Moncoiffé G (2015) A vocabulary for the configuration of net tows for collecting plankton and micronekton. J Plankton Res 37:21-27 doi:10.1093/plankt/fbu101.

Yen P, Huettmann F, Cooke F (2004) A large-scale model for the at-sea distribution and abundance of Marbled Murrelets (*Brachyramphus marmoratus*) during the breeding season in coastal British Columbia, Canada. Ecol Model 171:395-413

Zuckerberg B, Huettmann F, Frair J (2011) Proper data management as a scientific foundation for reliable species distribution modeling. In: Drew CA, Wiersma YF, Huettmann F (eds) Predictive Species and Habitat Modeling in Landscape Ecology: Concepts and Applications. Springer New York, New York, NY, pp 45-70. doi:10.1007/978-1-4419-7390-0_4.

Appendix A

Supplementary Figures and Tables

Figures

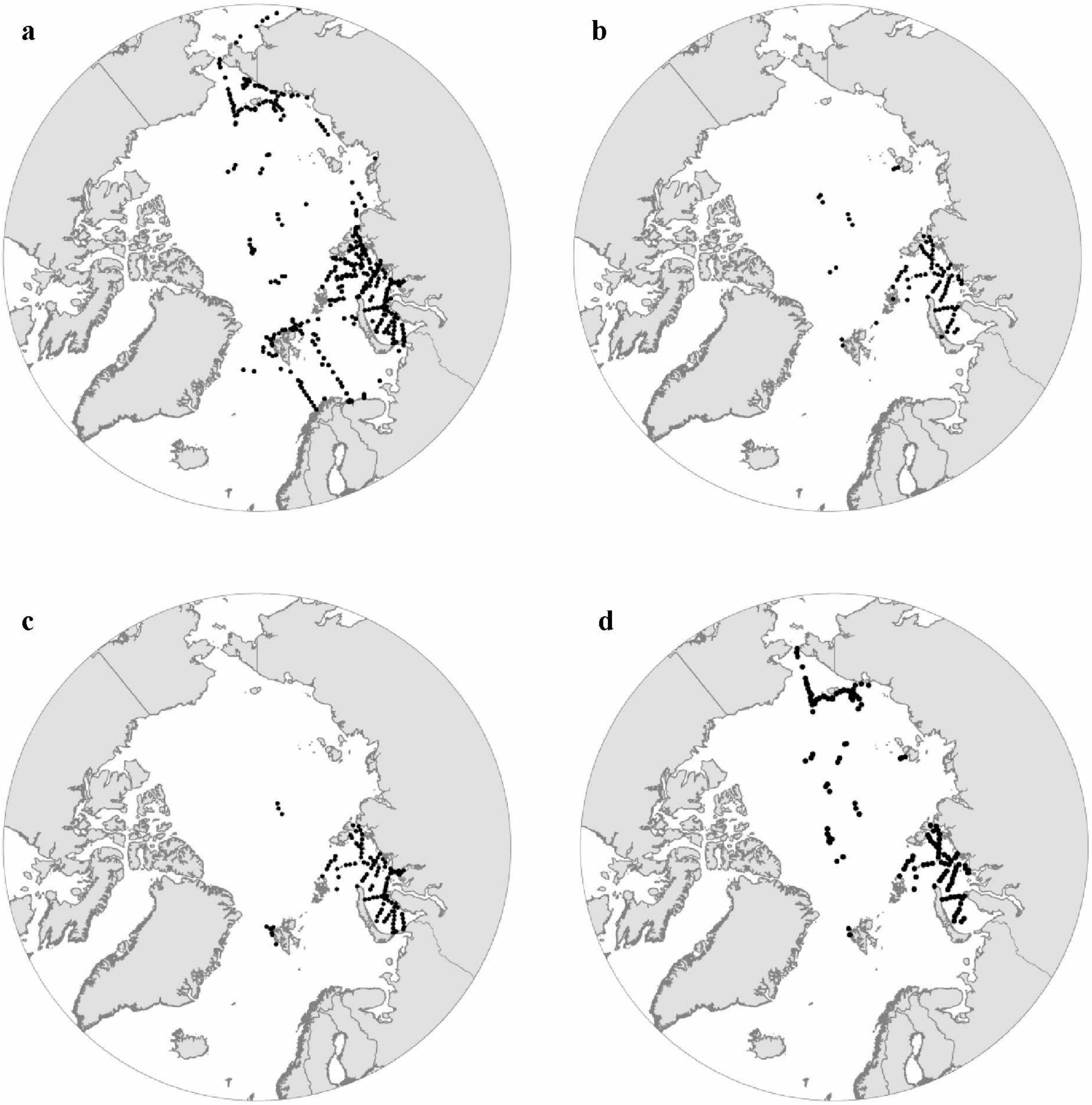


Figure A-1: Distribution of *Calanus finmarchicus* (a and c) and *C. glacialis* (b and d) in the "Archive of the Arctic Seas" dataset before taxonomic adjustment (a and b) and after (c and d)

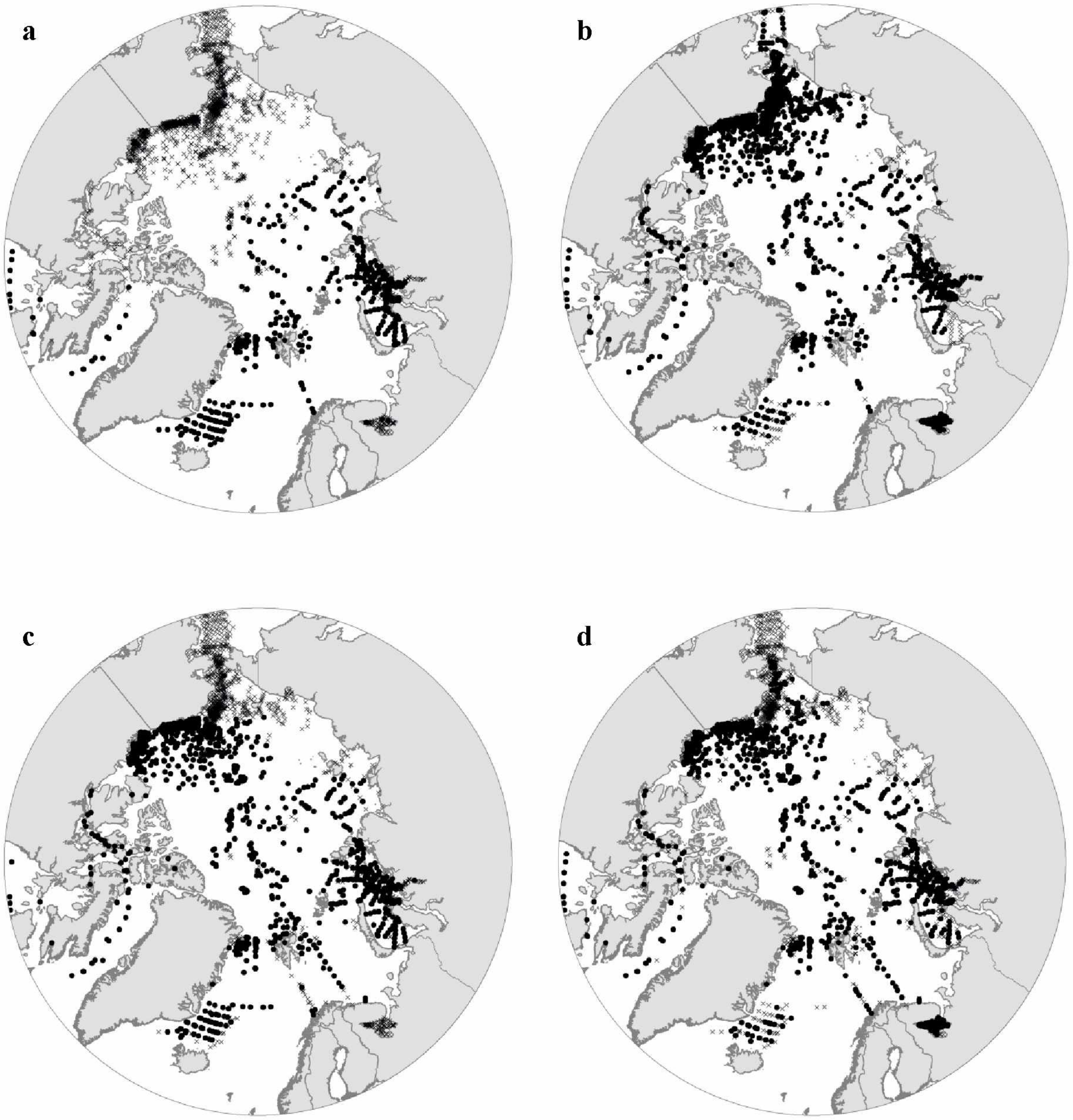
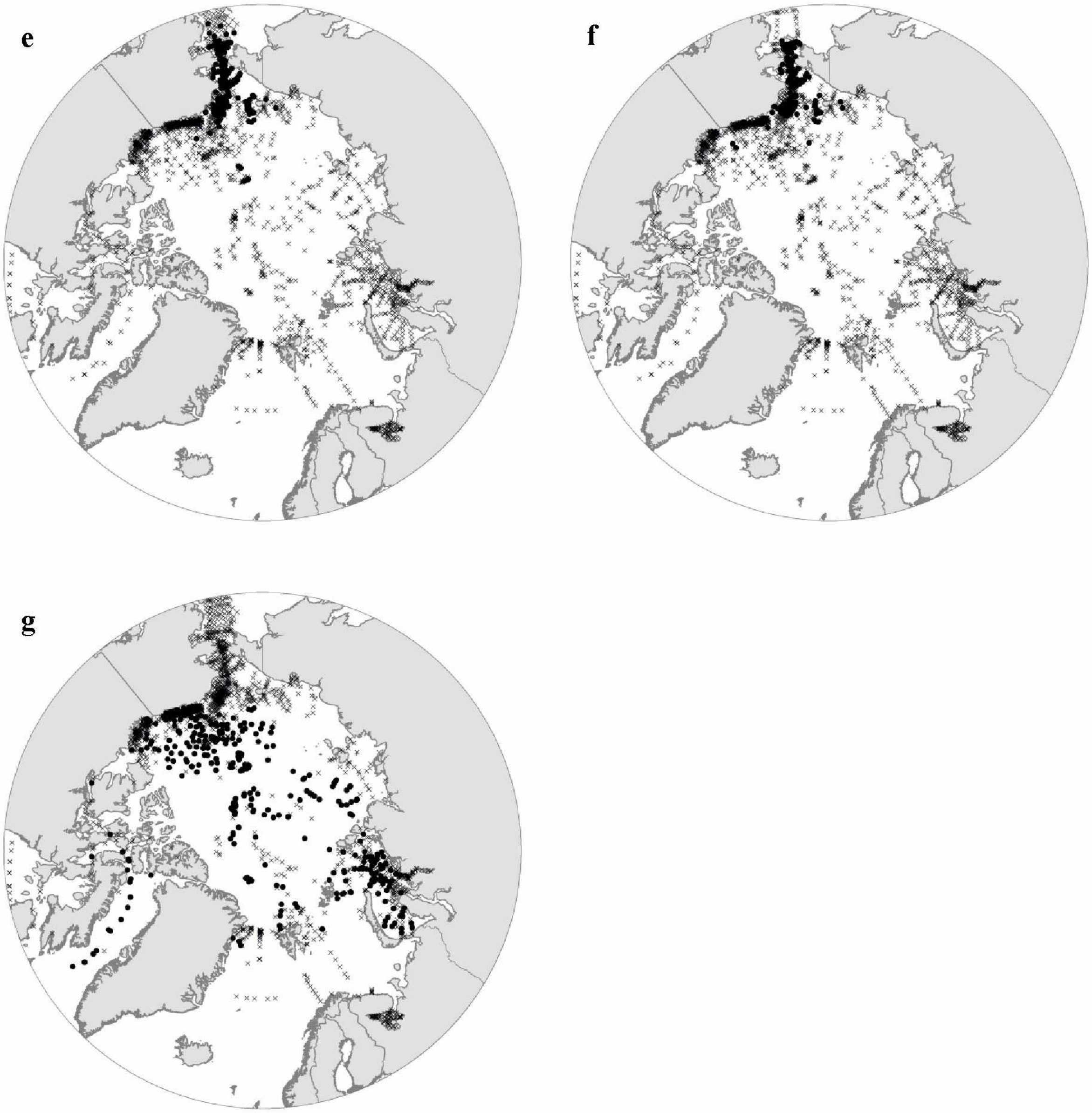


Figure A-2: Distribution of data points for *Calanus finmarchicus* (a), *C. glacialis* (b), *C. hyperboreus* (c), *Metridia longa* (d), *M. pacifica* (e), *Neocalanus cristatus* (f), and *Paraeuchaeta glacialis* (g). Black dots: abundance. Black crosses: absence

Figure A-2: cont.



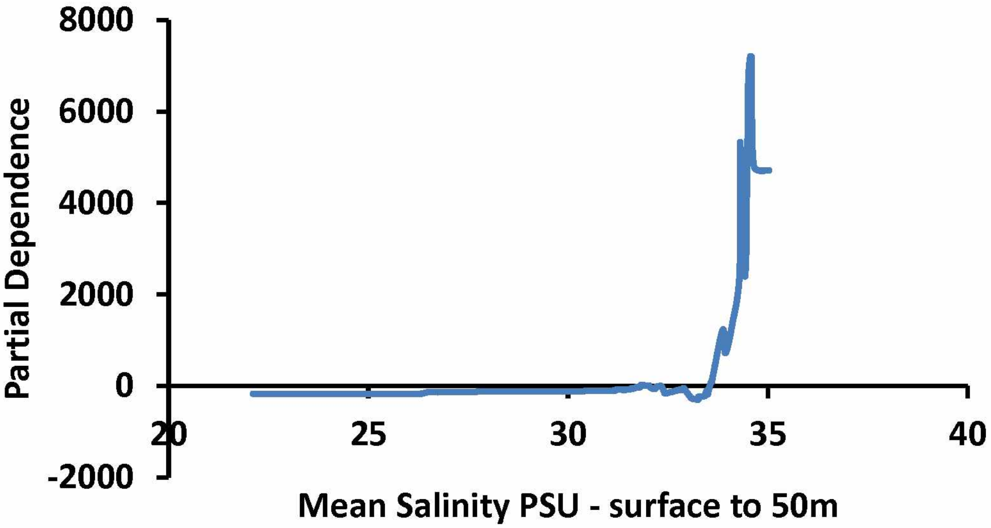


Figure A-3: Partial dependence plot of mean salinity (surface to 50 m depth) in PSU for *Calanus finmarchicus*

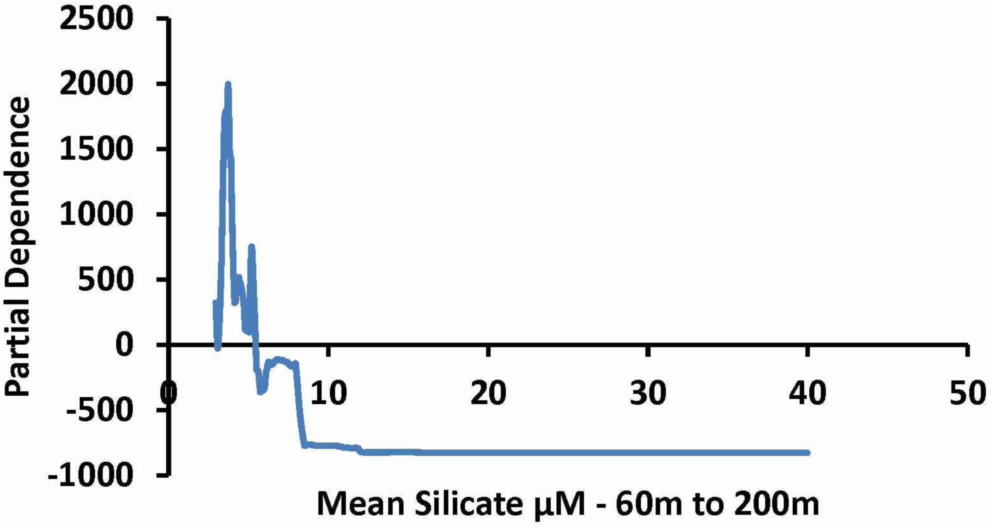


Figure A-4: Partial dependence plot of mean silicate (60 m to 200 m depth) in µM for *Calanus finmarchicus*

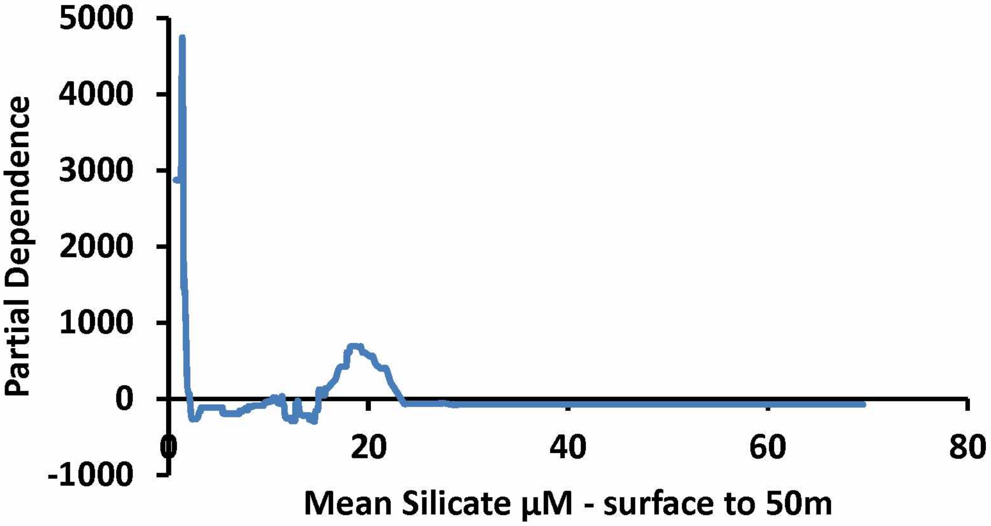


Figure A-5: Partial dependence plot of mean silicate (surface to 50 m depth) in µM for *Calanus glacialis*

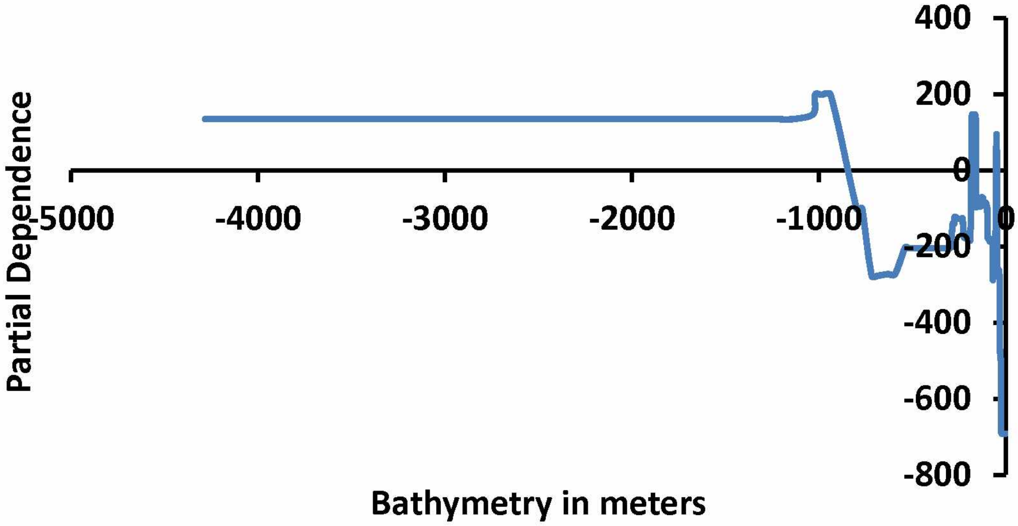


Figure A-6: Partial dependence plot of bathymetry in meters for *Calanus glacialis*

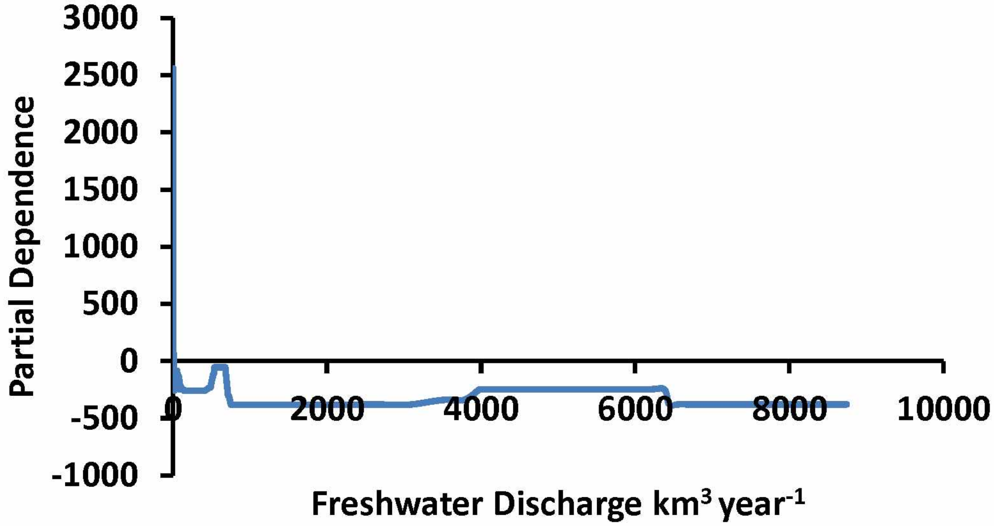


Figure A-7: Partial dependence plot of freshwater discharge in km^3 year^{-1} for *Calanus glacialis*

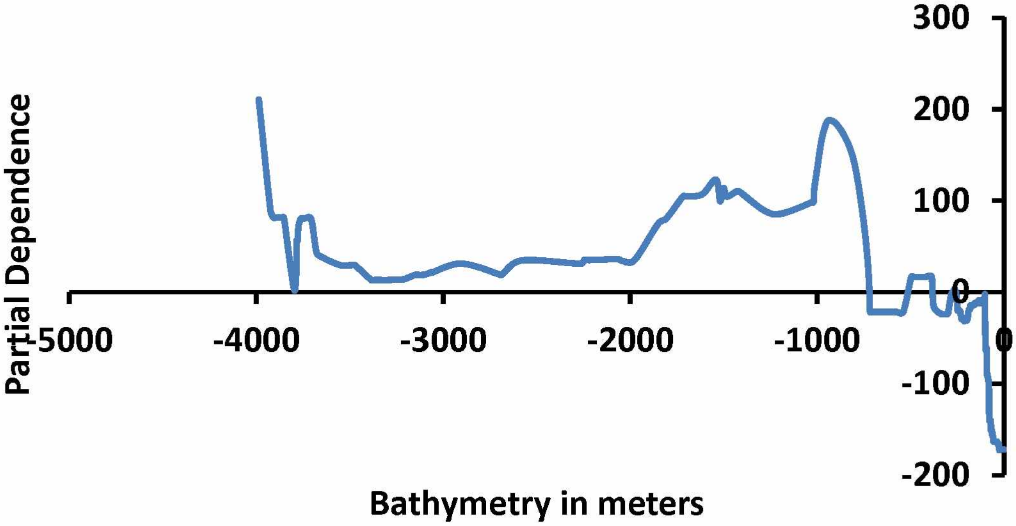


Figure A-8: Partial dependence plot of bathymetry in meters for *Calanus hyperboreus*

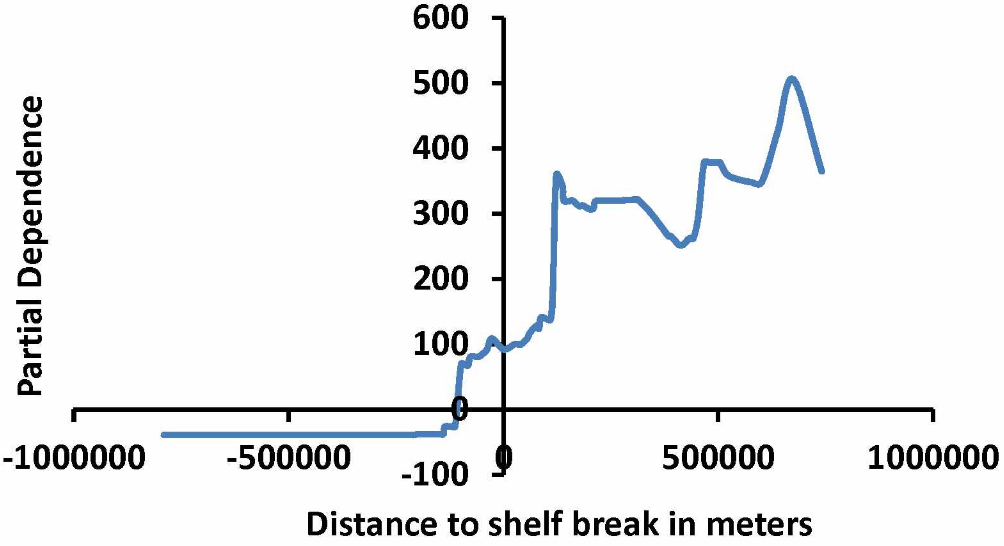


Figure A-9: Partial dependence plot of Euclidean distance to the shelf break in meters for *Calanus hyperboreus*

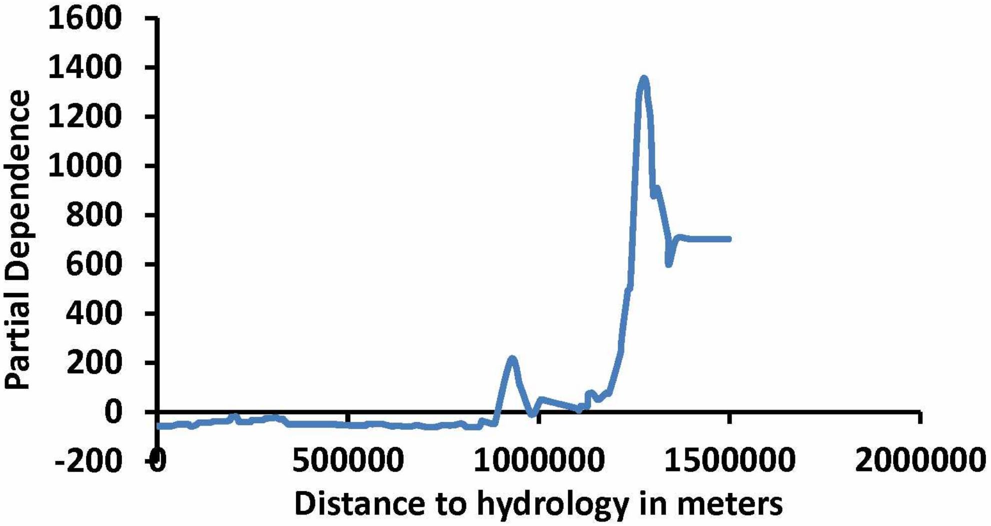


Figure A-10: Partial dependence plot of distance to hydrology in meters for *Calanus hyperboreus*

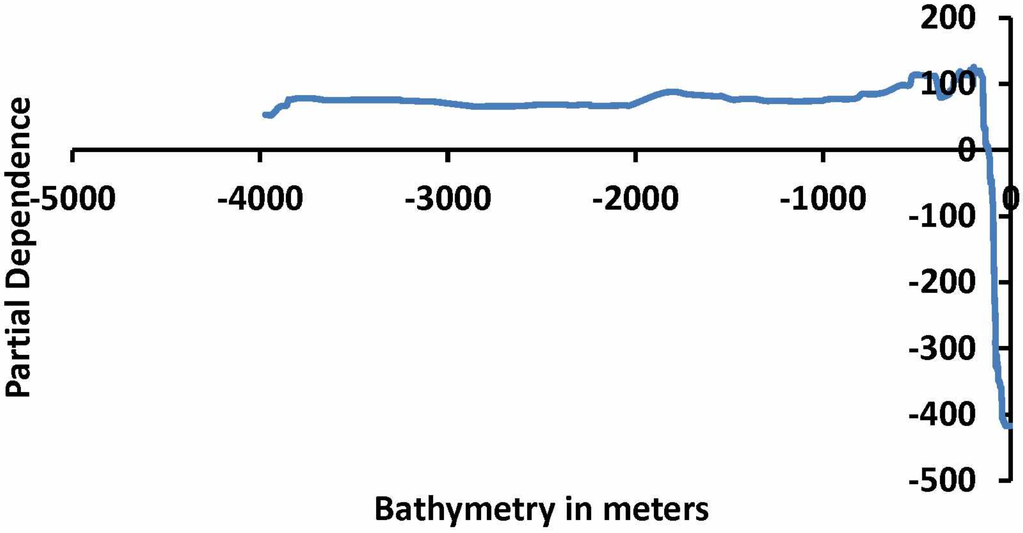


Figure A-11: Partial dependence plot of bathymetry in meters for *Metridia longa*

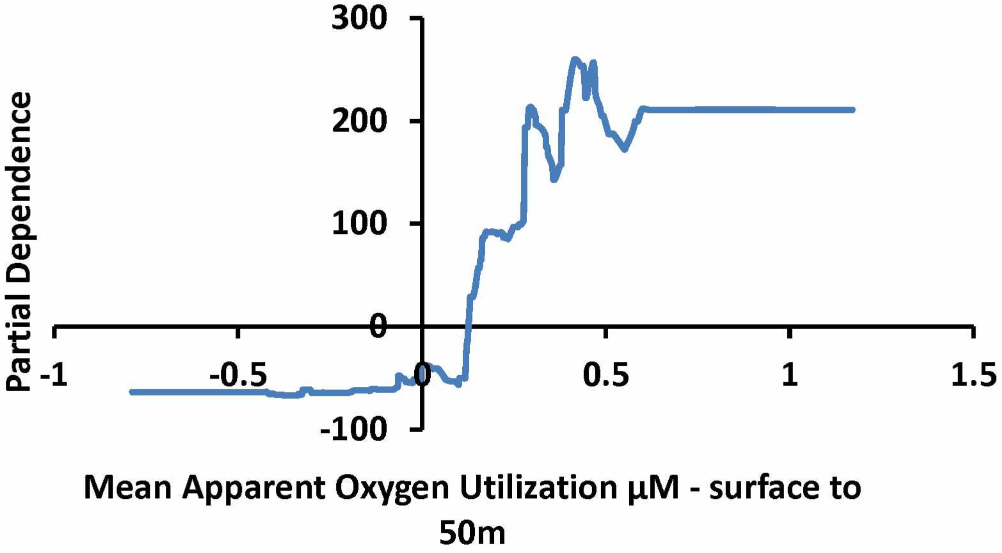


Figure A-12: Partial dependence plot of apparent oxygen utilization (surface to 50 m depth) in µM for *Metridia longa*

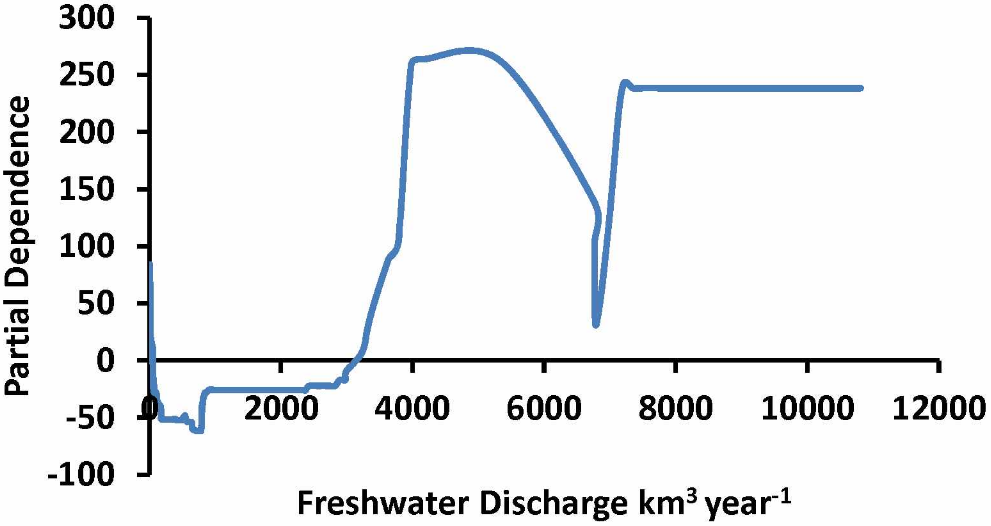


Figure A-13: Partial dependence plot of freshwater discharge in km^3 year^{-1} for *Metridia longa*

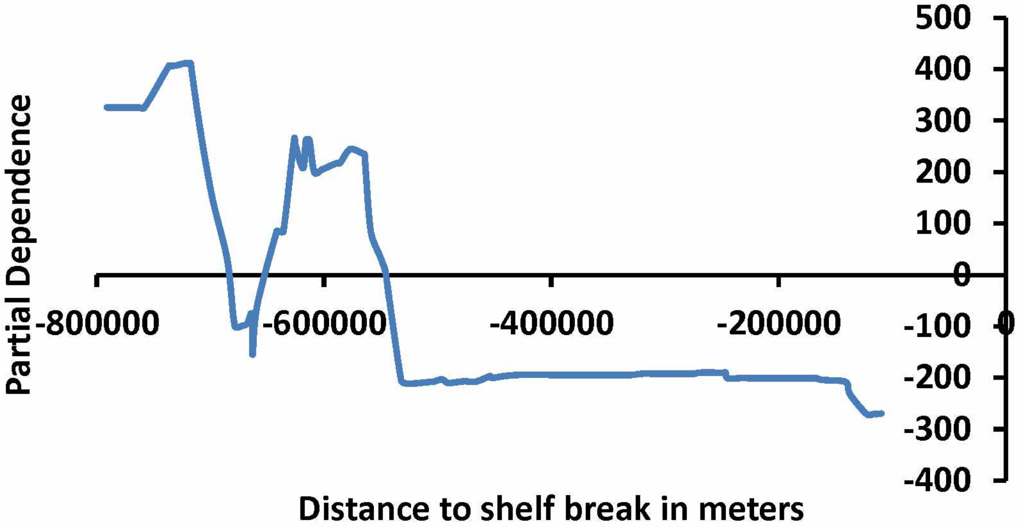


Figure A-14: Partial dependence plot of distance to shelf break in meters for *Metridia pacifica*

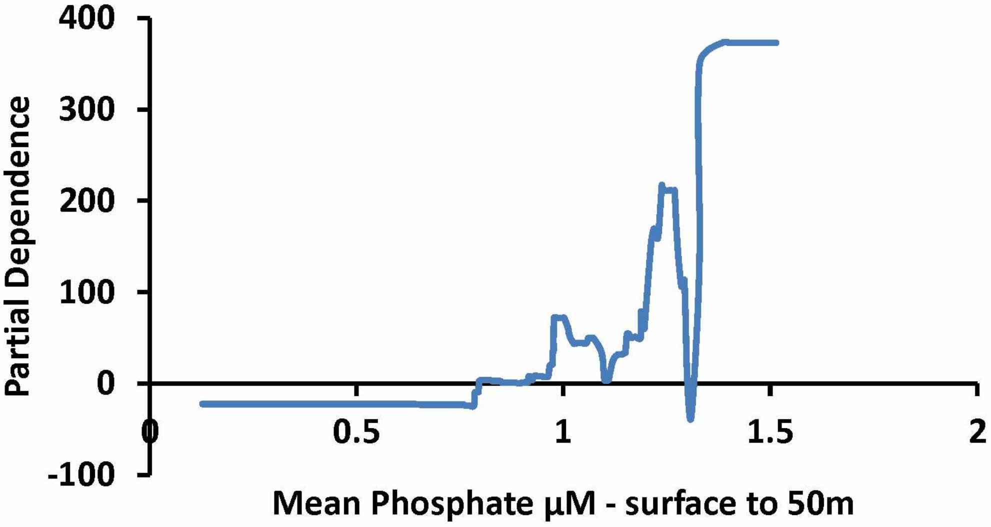


Figure A-15: Partial dependence plot of mean phosphate (surface to 50 m depth) in µM for *Metridia pacifica*

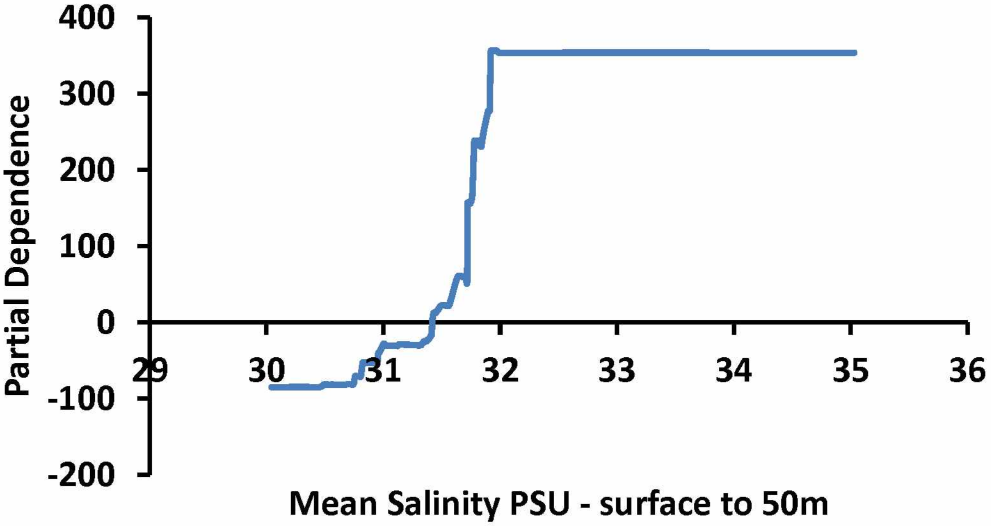


Figure A-16: Partial dependence plot of mean salinity (surface to 50 m depth) in PSU for *Metridia pacifica*

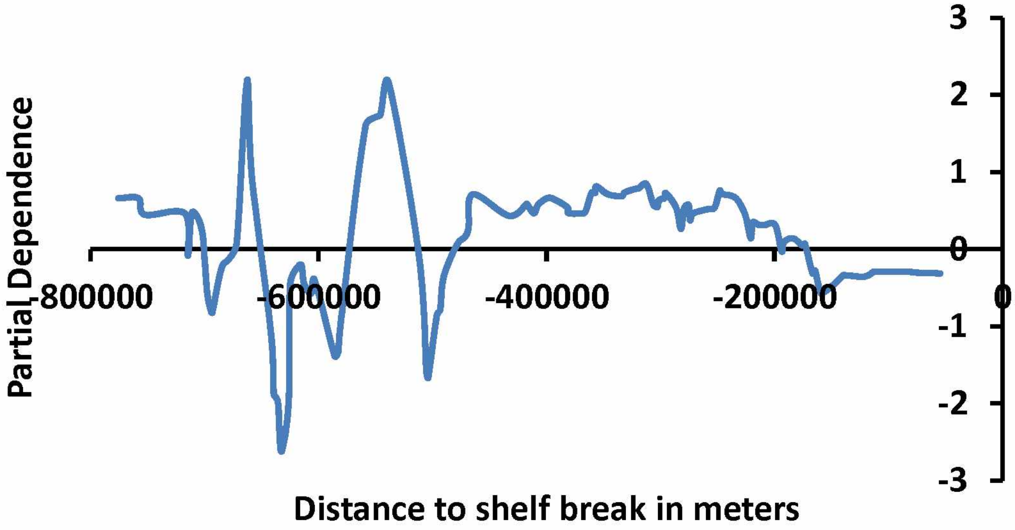


Figure A-17: Partial dependence plot of distance to shelf break in meters for *Neocalanus cristatus*

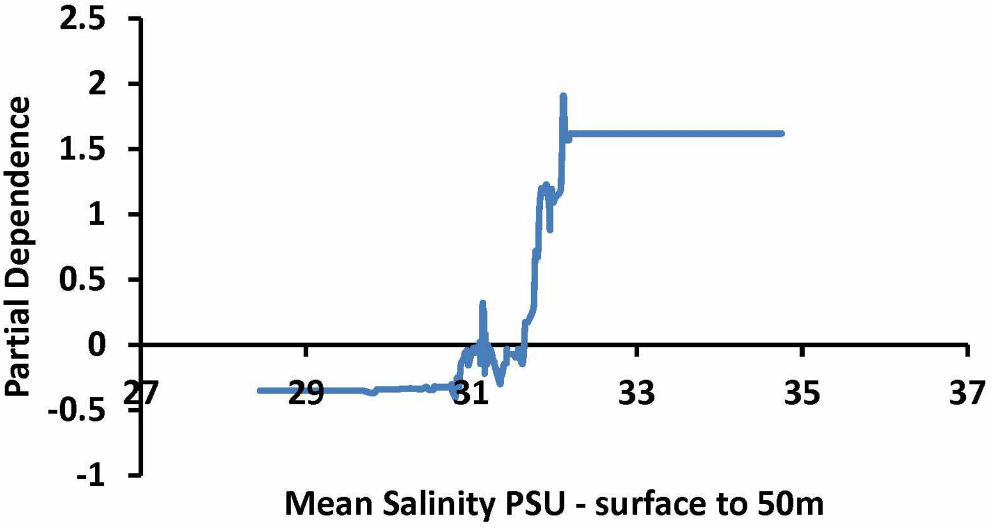


Figure A-18: Partial dependence plot of salinity (surface to 50 m depth) in PSU for *Neocalanus cristatus*

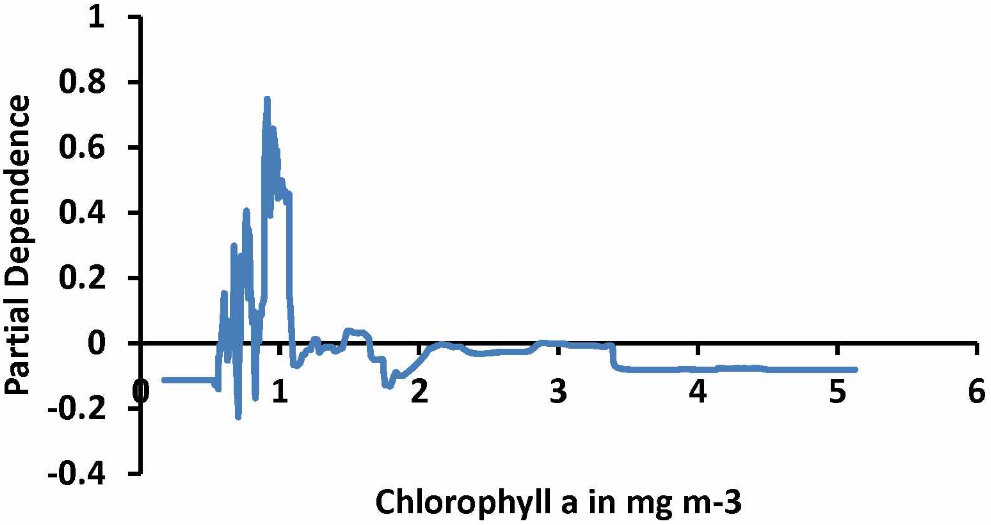


Figure A-19: Partial dependence plot of chlorophyll a in mg m^{-3} for *Neocalanus cristatus*

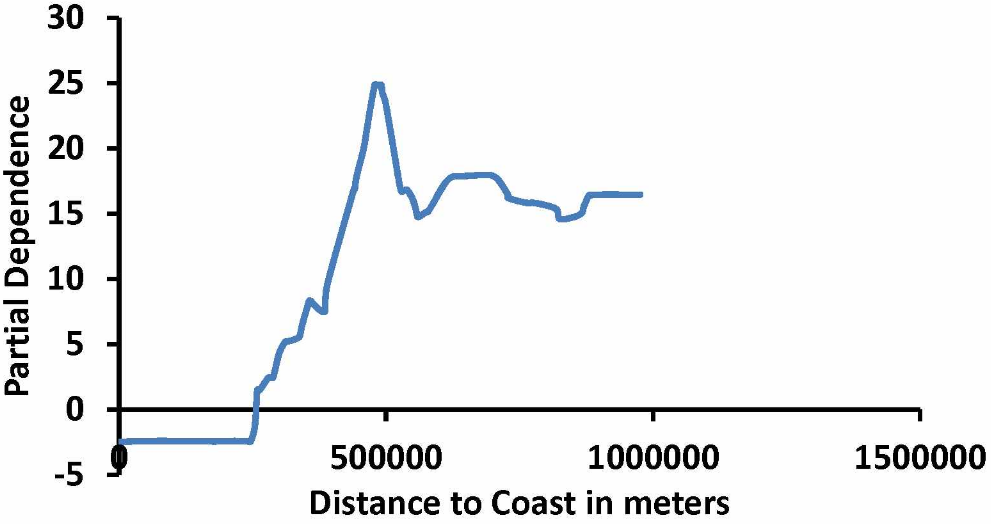


Figure A-20: Partial dependence plot of Euclidean distance to the coast in meters for *Paraeuchaeta glacialis*

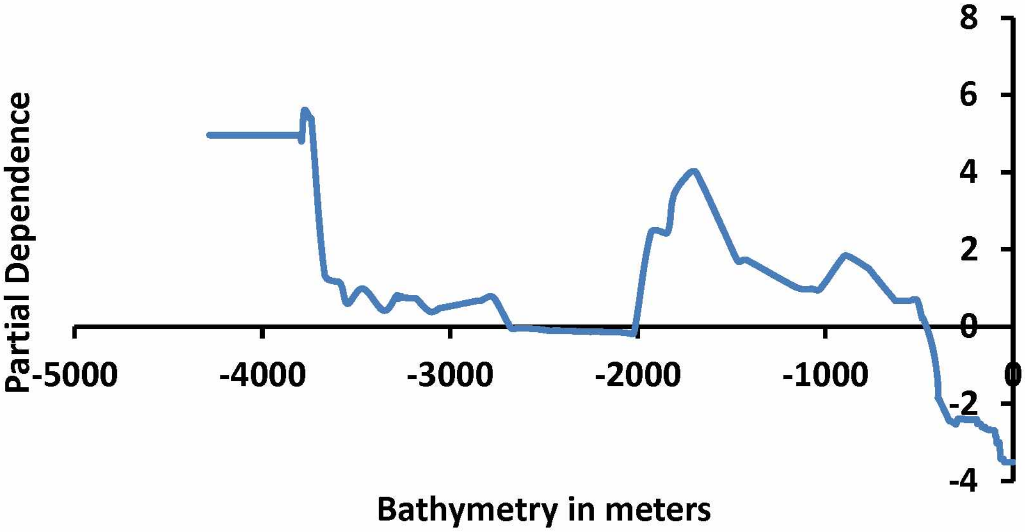


Figure A-21: Partial dependence plot of bathymetry in meters for *Paraeuchaeta glacialis*

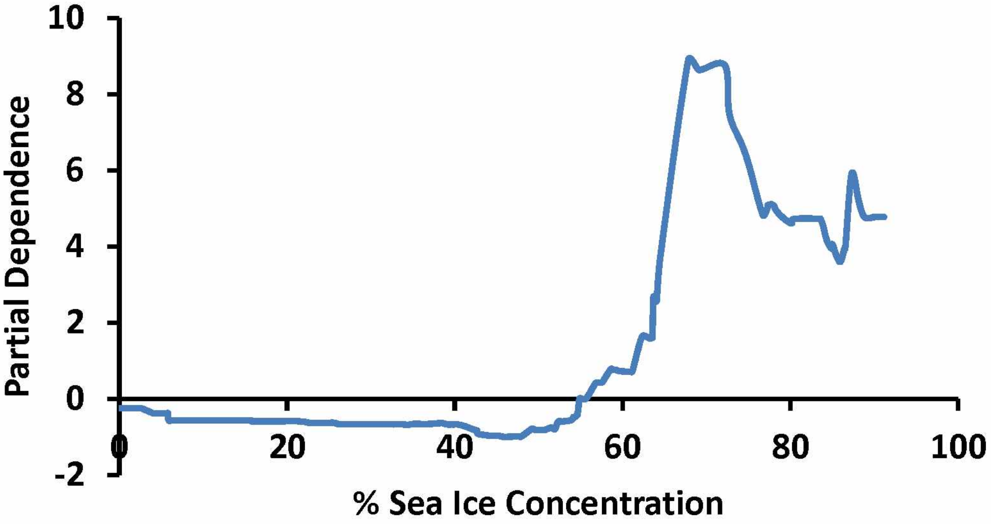


Figure A-22: Partial dependence plot of sea ice concentration in percent for *Paraeuchaeta glacialis*

Table A-1: Dataset description including Dataset name, source, years covered, waterbody where samples were taken, mesh size in µm, sampling gear, citation, and link to the data source/doi if applicable. The data are alphabetically sorted by the waterbody they were sampled in

Dataset	**Datasource**	**Years**	**Waterbody**	**Mesh in µm**	**Gear**	**Citation/Link**
Food web uptake of persistent organic pollutants in the Arctic Marginal Ice Zone of the Barents Sea (FAMIZ)	E. Arashkevich, A. Pasternak	2001	Amundsen and Nansen Basins	180	WP II Net	Olli, K. et al. 2007: The fate of production in the central Arctic Ocean–top–down regulation by zooplankton expatriates? Progress in Oceanography, 72(1). p.84-113
Archives of the Arctic Seas Zooplankton	E. Markhaseva/ Zoological Institute of Russian Academy of Sciences, Saint Petersburg	1921, 1930, 1934, 1935, 1938, 1939,1946-1948, 1950, 1952, 1954-1956, 1970-1973	Arctic Ocean, East Siberian Sea, Laptev Sea, Kara Sea	168, 333	Juday Net, Nansen closing, Nansen surface, Closing Plankton Net	Markhaseva, E. et al. 2005: Archives of the Arctic Seas Zooplankton: Contributions from the Zoological Institute RAS. No 8. 44p.
ARK XXVI	K. Kosobokova	2011	Arctic Ocean	150	MultiNet	Unpublished data
North Pole Station, NP22 & NP23	ArcOD/ K. Kosobokova	1975-1977	Arctic Ocean	180	Juday Net	Kosobokova, K.N. 1989: Vertical distribution of plankton animals in the eastern part of the central Arctic Basin. Explorations of the Fauna of the Seas, Marine Plankton, Leningrad 41:24 31 [*in Russian*]

Table A-1 continued…

Driftalpha	NMFS COPEPOD	1957, 1958	Arctic Ocean	230, 550	Ring Net	Johnson, M.W. 1963: Zooplankton collections from the high polar basin with special reference to the copepoda. Limnology and Oceanography, 8(1). p.89-102 http://www.st.nmfs.noaa.gov/copepod/data/driftalpha/index.html
Ice1, Ice2	Pangaea/ S. Falk-Petersen	2004	Arctic Ocean/North of Svalbard	180	MultiNet	Falk-Petersen, S. et al. 2008: Vertical migration in high Arctic waters during autumn 2004. Deep Sea Research II, 55(20-21). p. 2275-2284
Minoda, Drift Station Arlis II	NMFS COPEPOD	1964	Arctic Ocean; North of Greenland	330	NORPAC	Minoda, T. 1967: Seasonal distribution of Copepoda in the Arctic Ocean from June to December, 1964. Recent Oceanic Works in Japan, 9. p.161-168 http://www.st.nmfs.noaa.gov/copepod/data/minoda1967/index.html
Polarstern ARK XI-1, ARK XII, and ARK XIV	K. Kosobokova	1995-1998	Arctic Ocean	150	MultiNet	Kosobokova, K.N. et al. 2010: Patterns of zooplankton diversity through the depths of the Arctic's central basins. Marine Biodiversity, 41(1). p.29-50
Barents Sea Ecosystem Program - FB section	IMR/P. Dalpadado	1995-2012	Barents Sea	180	WP II Net	Gjøsaeter, H. et al. 2002: Growth of Barents Sea capelin (Mallotus villosus) in relation to zooplankton abundance. ICES Journal of Marine Science, 59(5). p. 959-967
ICEBAR	Norwegian Polar Institute/S. Falk-Petersen	1996	Barents Sea	180	MultiNet	Falk-Petersen, S. et al. 2000: Physical and ecological processes in the marginal ice zone of the northern Barents Sea during the summer melt period. Journal of Marine Systems, 27(1). p.131-159
On thin Ice (OTI)	Norwegian Polar Institute/ S. Falk-Petersen	2003	Barents Sea/Svalbard Archipelago	180	MultiNet	Blachowiak-Samolyk, K. et al. 2007: Trophic structure of zooplankton in the Fram Strait in spring and autumn 2003. Deep Sea Research II, 54(23). p. 2716-2728
BeauFish	C. A. Smoot	2011	Beaufort Sea	150	Bongo Net	Smoot, C.A. 2015: Contemporary Mesozooplankton Communities of the Beaufort Sea. M.S. Thesis, University of Alaska, Fairbanks
Mackenzie	W. Walkusz	2005, 2006	Beaufort Sea	153	Conical & Bongo Net combined	Walkusz, W. et al. 2010: Distribution, diversity and biomass of summer zooplankton from the coastal Canadian Beaufort Sea. Polar Biology, 33(3). p. 321-335

Table A-1 continued…

NOGAP B2	ArcOD/ Hopky, Lawrence and Chiperzak	1984-1988	Beaufort Sea	63, 85, 500, 763	Wisconsin, Neuston, Bongo Net,	Hopky, G.E. et al. 1994: NOGAP B2; Zooplankton data from the Canadian Beaufort Sea Shelf, 1984 and 1985. DFO Canada Hopky, G.E. et al. 1994: NOGAP B2; Zooplankton data from the Canadian Beaufort Sea Shelf, 1986. DFO Canada Hopky, G.E. et al. 1994: NOGAP B2; Zooplankton data from the Canadian Beaufort Sea Shelf, 1987 and 1988. DFO Canada
Ocean Exploration	R.R.Hopcroft, K. Kosobokova	2002, 2005, 2008	Beaufort Sea	53, 236, 150	Bongo Net, Bongo/MARMAP-style, MultiNet	Hopcroft, R.R. et al. 2005: Zooplankton communities of the Arctic's Canada Basin: the contribution by smaller taxa. Polar Biology, 28(3). p. 198-206
Study of the Alaskan Coastal System (SNACS)	C. Ashjian	2005, 2006	Beaufort Sea	150	Bongo Net	Ashjian, C.J. et al. 2010: Climate variability, oceanography, bowhead whale distribution, and Iñupiat subsistence whaling near Barrow, Alaska. Arctic, 63(2). p. 179-194
Western Beaufort Sea Ecological Cruises (WEBSEC)	NMFS COPEPOD/ C. Clarke Hopcroft	1970-1972	Beaufort Sea	570	NORPAC	Kinds and Abundance of Zooplankton collected by the USCG Icebreaker GLACIER in the Eastern Chukchi Sea, September-October 1970. NOAA Tech. Rep. NMFS SSRF-679. http://www.st.nmfs.noaa.gov/copepod/data/websec/index.html Hufford, G.L. et al. 1974: WEBSEC 71-72 an ecological survey in the Beaufort Sea. Oceanographic Report No. CG 373-64. United States Coast Guard. Oceanographic Unit, Washington, D.C.
Zooplankton Data from the Beaufort Sea	entered from report/ E.H. Grainger and K. Grohe	1951-1975	Beaufort Sea	73, 79, 233, 282, 569, 579, 1000	NA	Grainger, E.H. and Grohe, K. 1975: Zooplankton Data from the Beaufort Sea, 1951 to 1975. Technical Report No. 591, Environment Canada, Fisheries and Marine Service, 54 pp. https://www.aoncadis.org/dataset/Beaufort_Sea_Zooplankton_1951_1975_Grainger_Grohe.html doi:10.18739/A2430F

Table A-1 continued…

Bering-Aleutian Salmon International Survey (BASIS)	L. Eisner/K. Coyle	2002-2007, 2009-2011	Bering Sea	150, 335, 505	Bongo Net	Helle, J. et al. 2007. The Bering-Aleutian Salmon International Survey (BASIS). AFSC Quarterly Report Feature (January-February-March 2007), 5 p. http://www.afsc.noaa.gov/ABL/MESA/archives/mesa_occ_basis.htm
Bering Strait Zooplankton	ArcOD, A. Pinchuk & K. Coyle, UAF	1991	Bering Strait	333	Ring Net	Coyle, K.O. et al. 1996 Zooplankton of the Bering Sea: A review of Russian-language literature. Alaska Sea Grant. 97-133 Piatt et al. 1992 Foraging distribution and feeding ecology of seabirds at the Diomede Islands, Bering Strait. U.S. Fish and Wildlife. Serv. Final Report for Minerals Management Service OCS Study MMS 92-0041, Anchorage, Alaska: 133 pp
LSSL	I. Rutzen & R.R. Hopcroft	2003-2006	Canada Basin	150	Bongo Net	https://www.aoncadis.org/dataset/Canada_Basin_Zooplankton_2003_2006.html doi:10.18739/A27S4S
Surface Heat Budget of the Arctic (SHEBA)	C. Ashjian & R. Campbell	1997, 1998	Canada Basin	53, 150	Ring Net	Ashjian, C.J. et al. 2003: Annual cycle in abundance, distribution, and size in relation to hydrography of important copepod species in the western Arctic Ocean. Deep Sea Research I, 5 (10-11). p. 1235-1261
T-3 Ice Island & Arctic Ice Dynamics Joint Experiment (AIDJEX)	ArcOD	1970-1972, 1975	Canada Basin	73, 215, 223	Ring Net, English Net	Pautzke C.G. 1979: Copepoda collected from the Canada Basin Arctic Ocean; Fletcher's Ice Island (T-3) 1970-1972 and AIDJEX, 1975
T-3 Ice Island	ArcOD	1966-1969	Canada Basin/ Arctic Ocean	110, 215, 223, 300	Ring Net, Pump	Scott D.A.: Copepoda collected from Fletcher's Ice Island (T-3) in the Canadian Basin of the Arctic Ocean, 1966-1970. Technical Report No. 240 Reference M69-62
Canada's Three Oceans (C3O)	J. Nelson	2007-2009	Canada Basin and Canadian Archipelago	236	Bongo Net, SCOR Net	Pomerleau, C. et al. 2011: Spatial patterns in zooplankton communities across the eastern Canadian sub-Arctic and Arctic waters: insights from stable carbon (δ13C) and nitrogen (δ15N) isotope ratios. Journal of Plankton Research, 33(12). p.1779-1792

Table A-1 continued…

Canadian Arctic Archipelago	entered from report/ A.A. Mohammed and E.H. Grainger	1962	Canadian Arctic Archipelago	233	30 cm diameter, vertical	Mohammed, A.A. and Grainger, E.H. 1974: Zooplankton data from the Canadian Arctic Archipelago, 1962. Technical report No. 460, Fisheries and Marine Service, Environment Canada
Zooplankton Canada Basin and Chukchi Borderlands, R/V Mirai	A. Yamaguchi, K. Matsuno/ Hokkaido University	2008	Canada Basin and Chukchi Sea	333	NORPAC	Matsuno, K. et al. 2012: Horizontal distribution of calanoid copepods in the western Arctic Ocean during the summer of 2008. Polar Science, 6(1). p.105-119
Cape Bathurst NWP Lancaster Sound	L. Fortier, G. Darnis	2007	Cape Bathurst and Lancaster Sound/ Baffin Bay	200	Hydrobios MultiNet	Darnis, G. and Fortier, L. 2014: Temperature, food and the seasonal vertical migration of key arctic copepods in the thermally stratified Amundsen Gulf (Beaufort Sea, Arctic Ocean). Journal of Plankton Research, 36(4).p.1092-1108 doi:10.18739/A2H890
Chaun Bay	A. Pinchuk	1986	Chaun Bay/East Siberian Sea	163	Juday Net	Pinchuk, A. 1994: On the zooplankton of the Chaun Bay (East Siberian Sea). Explorations of the Fauna of the Seas, 47(55).p. 121-127 https://www.aoncadis.org/dataset/Chaun_Bay_Zooplankton_1986.html doi:10.18739/A2CK5B
Chukchi Sea Environmental Studies Program (CSESP)	ACADIS/R.R. Hopcroft, J.M. Questel	2008-2012	Chukchi Sea	150, 505	Bongo Net	Questel, J.M. et al. 2013: Seasonal and interannual variation in the planktonic communities of the northeastern Chukchi Sea during the summer and early fall. Continental Shelf Research, 67 p. 23-41 https://www.aoncadis.org/dataset/ucar.ncar.eol.dataset.255_038.html doi:10.5065/D63R0QX7
Chukchi Sea Zooplankton/R V Lomonosov	ArcOD/ A. Pinchuk	1953,1954	Chukchi Sea	333	Nansen Net	UAF - Arctic Ocean Chukchi Sea August 1953-1954 Zooplankton vertical stratified collections on board of the Russian R/V Lomonosov, program ANII A-65
Chukchi Sea Zooplankton	Retrieved from Paper/ E. A. Pavshtik	1976	Chukchi Sea	169	Juday Net	Pavshtiks, E.A. 1984: Zooplankton of the Chukchi Sea as indices of water origins. Trudy Arkticheskogo i Antarkticheskogo Nauchno-Issledovatel′skogo Institute, 368. p. 140-153 [*in Russian*] https://www.aoncadis.org/dataset/Chukchi_Zooplankton_1976_Pavshtiks.html doi:10.18739/A2N31G

Table A-1 continued…

Oshoro Maru	Matsuno et al.	1991, 1992, 2007, 2008	Chukchi Sea	333	NORPAC	Matsuno, K. et al. 2011: Year-to-year changes of the mesozooplankton community in the Chukchi Sea during summers of 1991, 1992 and 2007, 2008. Polar Biology, 34(9). p. 1349-1360
Shelf Basin Interactions (SBI)	Lane et al.	2004	Chukchi and Beaufort Seas	153	Bongo Net, 1-meter MOCNESS	Lane, P. et al. 2008: Zooplankton Data Report: Distribution of Zooplankton in the Western Arctic during Summer 2004. RSMAS Technical Report Number 2008-02.
R/V Mirai	W. Walkusz	2002	Chukchi and Beaufort Seas	333	Bongo	Walkusz, W. et al. 2008: Zooplankton and Ichthyoplankton Data Collected from the Chukchi and Beaufort Seas during the R/V Mirai Cruise, September 2002. Can. Data Rep. Fish. Aquat. Sci. 1211: vi + 34 p.
USS Burton Island	NMFS COPEPOD	1950-1951	Chukchi and Beaufort Seas	120	Nansen type	Johnson, M. W. 1956: The plankton of the Beaufort and Chukchi Sea areas of the Arctic and its relation to the Hydrography. Arctic Institute of North America, Technical Paper no. 1, 32pp. http://www.st.nmfs.noaa.gov/copepod/data/ussburton/index.html
Russian American Longterm Census of the Arctic (RUSALCA)	R.R. Hopcroft, K. Kosobkova, A. Pinchuk, E. Ershova	2004 and 2009	Chukchi Sea/Bering Strait	150	Bongo Net	Hopcroft, R.R. et al. 2010: Zooplankton community patterns in the Chukchi Sea during summer 2004. DSR II: Topical Studies in Oceanography. 57(1 -2). p. 27-39 Ershova, E.A. et al. 2015: Inter-annual variability of summer mesozooplankton communities of the western Chukchi Sea: 2004–2012. Polar Biology, 38(9). p.1-21
USCGC Chelan	NMFS COPEPOD	1934	Chukchi and Bering Seas	76	Plankton net (silk)	http://www.st.nmfs.noaa.gov/copepod/data/uscgchelan/index.html
East Siberian Sea	Retrieved from paper/ E.A. Pavshtik	1948 and 1973	East Siberian Sea	168	Juday Net	Pavshtik, E.A. 1994: Composition and quantitative distribution of the zooplankton in the East Siberian Sea. [*in Russian*] https://www.aoncadis.org/dataset/Zooplankton_East_Siberian_Sea_1948_and_1973_Pavshtiks.html doi:10.18739/A2VK5P
ARK III	entered from Dissertation / S. Diel	1985	Fram Strait	200	MultiNet	Diel, S. 1991: On the life history of dominant copepod species (*Calanus finmarchicus*, *C. glacialis*, *C. hyperboreus*, *Metridia longa*) in the Fram Strait. Berichte zur Polarforschung, 88

Table A-1 continued…

Frobisher Bay Station 5	entered from report/ E.H. Grainger	1967-1970	Frobisher Bay	73	NA	Fisheries Research Board of Canada, Report 266; Biological oceanographic observations in Frobisher Bay. II Zooplankton Data, 1967-1970, E.H. Grainger https://www.aoncadis.org/dataset/Frobisher_Bay_Zooplankton_1967-1971_Grainger.html doi:10.18739/A2RS32
North East Water Polynya (NEWP)	C. Ashjian	1992 and 1993	Greenland Sea	149, 153	MOCNESS, Bongo Net	Ashjian, C.J. et al. 1997: Distribution of zooplankton in the Northeast Water Polynya during summer 1992. Journal of Marine Systems, 10(1-4).p.279-298
ARK VIII-1	NMFS COPEPOD/ Hirche	1991	Greenland and Norwegian Sea	200, 335, 4500	Bongo Net, RMT-1, RMT-8	Hirche, H.-J. et al. 1994: The Northeast Water Polynya, Greenland Sea .3. Meso- And Macrozooplankton Distribution And Production Of Dominant Herbivorous Copepods During Spring. Polar Biology, 14(7).p. 491-503 http://www.st.nmfs.noaa.gov/copepod/data/arkpolarstern/index.html
INFERNO	Melle, W.	1997, 2005-2008	Greenland and Norwegian Sea	180	WP II Net	Hjøllo, S.S. et al. 2012: Modelling secondary production in the Norwegian Sea with a fully coupled physical/primary production/individual-based Calanus finmarchicus model system. Marine Biology Research, 8(5-6). p. 508-526
Estrada et al. 2012	M. Harvey	2003-2006	Hudson Bay and Fox Basin	73, 202	Ring Net	Estrada, R. et al. 2012: Late-summer zooplankton community structure, abundance, and distribution in the Hudson Bay system (Canada) and their relationships with environmental conditions, 2003–2006. Progress in Oceanography, 101 (1). p.121-145
Kara Sea data	ArcOD/ K. Kosobokova	1997, 1999 - 2001	Kara Sea	150	Nansen Net, Niskin bottle	Hirche, H.J. et al. 2006: Structure and function of contemporary food webs on Arctic shelves: A panarctic comparison: The pelagic system of the Kara Sea – Communities and components of carbon flow. Progress in Oceanography,71(2-4).p.288-313 https://arcticdata.io/catalog/#view/doi:10.18739/A2089B
ARK IX-4	K. Kosobokova	1993	Laptev Sea	150	MultiNet	Kosobokova, K.N. et al. 1998: Composition and distribution of zooplankton in the Laptev Sea and adjacent Nansen Basin during summer, 1993. Polar Biology, 19(1). p. 63-76 https://www.aoncadis.org/dataset/Laptev_Sea_Nansen_Basin_Zooplankton_1993_Polarstern_ARKIX4.html doi:10.18739/A2WK6C

Table A-1 continued…

Transdrift I & III	S. Lischka	1993	Laptev Sea	20, 200, 335	Hand Net, Bongo Net	Lischka, S. et al. 2001: Mesozooplankton assemblages in the shallow Arctic Laptev Sea in summer 1993 and autumn 1995. Polar Biology, 24(3). p. 186-199
ARK IV-3	NMFS COPEPOD/ N. Mumm	1987	Nansen Basin	300	MultiNet	Mumm, N. 1993: Composition and distribution of mesozooplankton in the Nansen Basin, Arctic-Ocean, during summer. Polar Biology,13(7).p. 451-461 http://www.st.nmfs.noaa.gov/copepod/data/arkpolarstern/index.html
Iceland Sea Ecosystem Project	A. Gislason	2006-2008	Subarctic Iceland Sea	200	MultiNet & WP II Net	Gislason, A., and Silva, T. 2012: Abundance, composition, and development of zooplankton in the Subarctic Iceland Sea in 2006, 2007, and 2008. ICES Journal of Marine Science, 69(7). p. 1263-1276.
White Sea	ArcOD/ K. Kosobokova	1972, 1998-2003	White Sea	180	Juday Net	Pertsova, N.M. and Kosobokova, K.N. 2003: Zooplankton of the White Sea: Features of the composition and structure, seasonal dynamics, and the contribution to the formation of matter fluxes. Oceanology, 43. p.108-122 https://arcticdata.io/catalog/#view/doi:10.18739/A2089B
Zackenberg Research Station	Zackenberg Research Station/ Greenland	2003-2008	Young Sound/ East Greenland	50	WP II Net	http://zackenberg.dk/monitoring/marinebasis/ Jensen, L.M. and Rasch, M (eds.) 2009: Zackenberg Ecological Research Operations, 14th Annual Report, 2008. Roskilde, National Environmental Research Institute, Aarhus University, 2009, 116 pp.

Table A-2: Mesh size in µm and bins for the selected species

Species	**Mesh size**	**Bin**
Calanus finmarchicus	20-169	1
	180-335	2
	500-579	3
Calanus glacialis	20-579	1
Calanus hyperboreus	20-579	1
Metridia longa	20-335	1
	500-579	2
Metridia pacifica	20-169	1
	180-335	2
	500-579	3
Neocalanus cristatus	20-579	1
Paraeuchaeta glacialis	20-579	1

Table A-3: Final model settings for TreeNet for *Calanus finmarchicus*, *C. glacialis*, *C. hyperboreus*, *Metridia longa*, *M. pacifica*, *Neocalanus cristatus*, and *Paraeuchaeta glacialis*

Species	**Trees**	**Nodes per tree**	**Learnrate**	**# of Cross-validation cycles**	**%data randomly selected for testing**
C. finmarchicus	1000	9	0.1	10	40
C. glacialis	1000	6	0.1	20	50
C. hyperboreus	1000	4	0.1	50	50
M. longa	1000	6	auto	30	50
M. pacifica	1000	10	0.1	40	50
N. cristatus	1000	15	0.1	20	50
P. glacialis	1000	10	auto	30	50

Table A-4: Final model settings for RandomForests for *Calanus finmarchicus*, *C. glacialis*, *C. hyperboreus*, *Metridia longa*, *M. pacifica*, *Neocalanus cristatus*, and *Paraeuchaeta glacialis*

Species	**Trees**	**#Predictors considered for each**	**#Proximal cases**	**Bootstrap sample size**	**Parent node minimum case**	**%data randomly selected for testing**
C. finmarchicus	1000	3	auto	auto	2	30
C. glacialis	1000	3	auto	auto	2	30
C. hyperboreus	1000	3	auto	auto	2	40
M. longa	1000	5	auto	auto	2	50
M. pacifica	1000	5	auto	auto	2	30
N. cristatus	1000	5	auto	auto	2	30
P. glacialis	1000	5	auto	auto	2	30

Chapter 4 General Conclusion

The impacts of climate change are especially pronounced in the Arctic and the life cycles of Arctic zooplankton make them "sensitive beacons of climate change" (Richardson, 2008). Hence, it is important to study changes in the Arctic zooplankton community in order to describe the impact of global warming. Although this study demonstrated that four years of zooplankton data sampled in the same region (i.e. Canada Basin) did not detect major shifts in zooplankton community structure, when combined with other studies (i.e. Hopcroft *et al.*, 2005; Hunt *et al.*, 2014) it contributes to an 8-year trend of increasing community abundance and biomass during a period when sea ice extent has been in continual decline (Parkinson & Comiso, 2013).

It is now clear that zooplankton communities in the Arctic (as elsewhere) are heavily influenced by water mass types and their distribution (e.g. Walkusz *et al.*, 2010; Ershova *et al.*, 2015a; Kristiansen *et al.*, 2015; Smoot & Hopcroft, 2017a) as well as by depth (e.g. Kosobokova & Hirche, 2000; Lane *et al.*, 2008; Kosobokova & Hopcroft, 2010; Walkusz *et al.*, 2013; Smoot & Hopcroft, 2017b). The underlying physical conditions are changing as the Arctic shifts into a warmer and less ice-covered state (Yamamoto-Kawai *et al.*, 2009; McLaughlin & Carmack, 2010) with earlier sea ice retreat and shifts in timing and magnitude of phytoplankton blooms (Hunt *et al.*, 2011). A northward shift in species distribution of *Calanus*, from a larger (*C. finmarchicus*) to a smaller species (*C. helgolandicus*), has already been observed in the North Sea and linked to changes in hydrographic conditions (Beaugrand *et al.*, 2002). Kristiansen *et al.* (2015) reported that *C. finmarchicus* was able to produce two generations instead of one generation, after an increase in temperature and salinity occurred in the Iceland-Faroe front, while at the same time the occurrence of the larger *C. hyperboreus* decreased. In contrast, Ringuette *et al.* (2002) concluded that a reduced sea-ice cover and longer open water season would enhance population growth of large copepods such as *C. glacialis* and *C. hyperboreus* on Arctic shelves. In the Chukchi Sea, both endemic and expatriated large-bodied species appear to be increasing (Ershova *et al.*, 2015b). Changes in zooplankton distribution and abundance as a result of hydrographic change have already been observed, but it remains unclear which species will ultimately emerge as the winners and losers, especially in the central basins.

Efforts to compare and combine recent data with multiyear (Questel *et al.*, 2013; Hunt *et al.*, 2014; Ershova *et al.*, 2015a) or historical datasets (Ershova *et al.*, 2015b) have so far been restricted to single regions or species (Rutzen, 2007; Schmid, 2012) instead of a pan-Arctic community approach. To assemble data that can be used for pan-Arctic community analysis, it is important to have adequate metadata and sample data that are comparable; hence, the use and application of comparable sampling and analyzing techniques is important so that data can be compared on broad spatial and temporal scales. It is a difficult undertaking to implement standardized sampling techniques across various laboratories and nationalities, not only because most laboratories have their own established preferences for gear and mesh-size, but because these different mesh-sizes and gear catch different sized zooplankton (stages and species) and thus yield different abundance and biomass estimates (Hopcroft *et al.*, 2005; Skjoldal *et al.*, 2013) and apparent community structure. What may be achievable is that at least one standardized gear and mesh-size is applied to all forthcoming studies (along with gear that is needed for the specific study question). This way at least a part of the zooplankton community would be sampled over a large spatial and temporal domain.

Thus, to make use of the array of existing data with its inherent biases, it is important to have analysis techniques that can handle various shortcomings and find patterns within "messy" data. This is where data mining techniques such as multiple regression tree analysis (e.g. TreeNet and RandomForests) become essential (Craig & Huettmann, 2008; Oppel *et al.*, 2009). By applying machine-learning in the current study, I demonstrated that even with abundance data from numerous different sources, and averaged environmental data as predictors, it was possible to recover the major patterns of species distribution and abundance for the Arctic's key copepod species. The model correctly distinguished between Arctic endemic and seasonally advected species, and appropriately characterized species as associated with shelf or deep-water areas. This approach was again species based, but analysis at the community level using RandomForests has already been conducted (Miller *et al.*, 2014) and can be applied to the dataset assembled here.

This dissertation highlights shortcomings in the zooplankton data assembled thus far such as missing metadata, changing taxonomy, and different taxonomic resolution. Given that taxonomist was a very important variable in the prediction of species distribution and abundance, concerns are raised about how comparable zooplankton data are between laboratories. There are several approaches to erase biases occurring through different taxonomic identification. One would be a regular meeting or exchanges of taxonomists working in the Arctic, where guidelines are set and taxonomic training can occur. Another common practice is that identified specimens are sent to experienced colleagues to confirm identifications, although this has become uncommon in recent decades. Online resources are serving part of this role, but the real breakthrough has been the use of molecular identification (e.g. Bucklin *et al.*, 2010a; Bucklin *et al.*, 2010b) that allows anyone to prepare sequences in house or commercially and compare them to existing expertly-identified databases. Nonetheless, there is no substitute for long-term archival of samples so that material can be re-examined in light of changes in taxonomy, or as new techniques allow increased information to be recovered from them.

This work contributes to a baseline of Arctic zooplankton data and provides the largest source of consolidated pan-Arctic zooplankton abundance to date. By presenting newly analyzed zooplankton community data for the Canada Basin, as well as species distribution maps for seven ecologically important copepod species and the description of their niches, this work contributes to baseline information required to monitor impacts of climate change. The 136 environmental layers that were compiled here can be applied to a large variety of species, while our distribution and abundance maps can be used to describe and explain the distribution of planktivorous animals.

The next steps of this research should be to expand and maintain the database as a common resource, as well as to extend modelling efforts to encompass more species. This dissertation applied averaged climatologies as environmental predictors, but assembling the environmental measurements that were taken concurrently with the zooplankton samples, should be attempted and a new model built upon them. Relating the zooplankton abundances and spatial distribution with the actual environmental measurements should improve the predicted niche, and will be critical to explore changes over time, which our analysis already suggests have occurred. A community analysis should also be attempted to detect patterns in the zooplankton community structure across Arctic seas and basins. Various climate change scenarios should be explored at the species level, using for example the Canadian Earth System model CanESM2 (see Schmid, 2012), to predict the future distribution of Arctic endemic and advected Pacific and Atlantic Ocean species. The predicted future distribution could provide insight into how expatriate species may expand their territory and whether the lipid-rich endemic species will show a decreased abundance and displacement due to environmental conditions.

References

Beaugrand, G., Reid, P. C., Ibañez, F., Lindley, J. A. and Edwards, M. (2002) Reorganization of North Atlantic marine copepod biodiversity and climate. *Science*, **296**, 1692-1694.

Bucklin, A., Hopcroft, R. R., Kosobokova, K. N., Nigro, L. M., Ortman, B. D., Jennings, R. M. and Sweetman, C. J. (2010a) DNA barcoding of Arctic Ocean holozooplankton for species identification and recognition. *Deep-Sea Res., Part II*, **57**, 40-48.

Bucklin, A., Ortman, B. D., Jennings, R. M., Nigro, L. M., Sweetman, C. J., Copley, N. J., Sutton, T. and Wiebe, P. H. (2010b) A "Rosetta Stone" for metazoan zooplankton: DNA barcode analysis of species diversity of the Sargasso Sea (Northwest Atlantic Ocean). *Deep-Sea Res., Part II*, **57**, 2234-2247.

Craig, E. and Huettmann, F. (2008) Using "Blackbox" algorithms such as TreeNet and Random Forests for data-mining and for finding meaningful patterns, relationships and outliers in complex ecological data: an Overview, an example using golden eagle satellite data and an outlook for a promising future. In: Wang, H.-F. (ed) *Intelligent Data Analysis: Developing New Methodologies Through Pattern Discovery and Recovery.* IGI Global, Hershey, PA, USA, pp. 65-84.

Ershova, E., Hopcroft, R. and Kosobokova, K. (2015a) Inter-annual variability of summer mesozooplankton communities of the western Chukchi Sea: 2004–2012. *Polar Biol.*, **38**, 1-21.

Ershova, E. A., Hopcroft, R. R., Kosobokova, K. N., Matsuno, K., Nelson, R. J., Yamaguchi, A. and Eisner, L. B. (2015b) Long-term changes in summer zooplankton communities of the Western Chukchi Sea, 1945-2012. *Oceanography*, **28**, 100-115.

Hopcroft, R. R., Clarke, C., Nelson, R. J. and Raskoff, K. A. (2005) Zooplankton communities of the Arctic's Canada Basin: the contribution by smaller taxa. *Polar Biol.*, **28**, 198-206.

Hunt, B. P., Nelson, R. J., Williams, B., McLaughlin, F. A., Young, K. V., Brown, K. A., Vagle, S. and Carmack, E. C. (2014) Zooplankton community structure and dynamics in the Arctic Canada Basin during a period of intense environmental change (2004–2009). *J. Geophys. Res.: Oceans*, **119**, 2518-2538.

Hunt, G. L., Coyle, K. O., Eisner, L. B., Farley, E. V., Heintz, R. A., Mueter, F., Napp, J. M., Overland, J. E., Ressler, P. H., Salo, S. and Stabeno, P. J. (2011) Climate impacts on eastern Bering Sea foodwebs: a synthesis of new data and an assessment of the Oscillating Control Hypothesis. *ICES J. Mar. Sci.*, **68**, 1230-1243.

Kosobokova, K. and Hirche, H.-J. (2000) Zooplankton distribution across the Lomonosov Ridge, Arctic Ocean: species inventory, biomass and vertical structure. *Deep-Sea Res., Part I*, **47**, 2029-2060.

Kosobokova, K. N. and Hopcroft, R. R. (2010) Diversity and vertical distribution of mesozooplankton in the Arctic's Canada Basin. *Deep-Sea Res., Part II*, **57**, 96-110.

Kristiansen, I., Gaard, E., Hátún, H., Jónasdóttir, S. and Ferreira, A. S. A. (2015) Persistent shift of *Calanus* spp. in the southwestern Norwegian Sea since 2003, linked to ocean climate. *ICES J Mar Sci*, **73**, 1319-1329.

Lane, P. V. Z., Llinás, L., Smith, S. L. and Pilz, D. (2008) Zooplankton distribution in the western Arctic during summer 2002: Hydrographic habitats and implications for food chain dynamics. *J. Mar. Syst.*, **70**, 97-133.

McLaughlin, F. A. and Carmack, E. C. (2010) Deepening of the nutricline and chlorophyll maximum in the Canada Basin interior, 2003–2009. *Geophys. Res. Lett.*, 37, 1944-8007.

Miller, K., Huettmann, F., Norcross, B. and Lorenz, M. (2014) Multivariate random forest models of estuarine-associated fish and invertebrate communities. *Mar. Ecol.: Prog. Ser.*, **500**, 159-174.

Oppel, S., Strobl, C. and Huettmann, F. (2009) Alternative methods to quantify variable importance in ecology. *Technical Report.* Department of Statistics, University of Munich.

Parkinson, C. L. and Comiso, J. C. (2013) On the 2012 record low Arctic sea ice cover: Combined impact of preconditioning and an August storm. *Geophys. Res. Lett.*, **40**, 1356-1361.

Questel, J. M., Clarke, C. and Hopcroft, R. R. (2013) Seasonal and interannual variation in the planktonic communities of the northeastern Chukchi Sea during the summer and early fall. *Cont. Shelf Res.*, **67**, 23-41.

Richardson, A. J. (2008) In hot water: zooplankton and climate change. *ICES J. Mar. Sci.*, **65**, 279-295.

Ringuette, M., Fortier, L., Fortier, M., Runge, J. A., Bélanger, S., Larouche, P., Weslawski, J.-M. and Kwasniewski, S. (2002) Advanced recruitment and accelerated population development in Arctic calanoid copepods of the North Water. *Deep-Sea Res., Part II*, **49**, 5081-5099.

Rutzen, I. (2007) *Predictive GIS-modeling of arctic mesozooplankton based on available public online data: Experiences from four copepod species (Calanus glacialis, C. hyperboreus, Metridia longa, and M. pacifica).* University of Veterinary Medicine Hannover, Foundation.

Schmid, M. (2012) *Model-predicting the effect of freshwater inflow on saltwater layers, migration and life history of zooplankton in the Arctic Ocean: Towards scenarios and future trends.* Georg-August-Universitaet Goettingen.

Skjoldal, H. R., Wiebe, P. H., Postel, L., Knutsen, T., Kaartvedt, S. and Sameoto, D. D. (2013) Intercomparison of zooplankton (net) sampling systems: Results from the ICES/GLOBEC sea-going workshop. *Prog. Oceanogr.*, **108**, 1-42.

Smoot, C. A. and Hopcroft, R. R. (2017a) Cross-shelf gradients of epipelagic zooplankton communities of the Beaufort Sea and the influence of localized hydrographic features. *J. Plankton Res.*, **39**, 65-78.

Smoot, C. A. and Hopcroft, R. R. (2017b) Depth-stratified community structure of Beaufort Sea slope zooplankton and its relations to water masses. *J. Plankton Res.*, **39**, 79-91.

Walkusz, W., Paulić, J., Kwaśniewski, S., Williams, W., Wong, S. and Papst, M. (2010) Distribution, diversity and biomass of summer zooplankton from the coastal Canadian Beaufort Sea. *Polar Biol.*, **33**, 321-335.

Walkusz, W., Williams, W. J. and Kwasniewski, S. (2013) Vertical distribution of mesozooplankton in the coastal Canadian Beaufort Sea in summer. *J. Mar. Sys.*, **127**, 26-35.

Yamamoto-Kawai, M., McLaughlin, F. A., Carmack, E. C., Nishino, S., Shimada, K. and Kurita, N. (2009) Surface freshening of the Canada Basin, 2003–2007: River runoff versus sea ice meltwater. *J. Geophys. Res.: Oceans*, **114**, 2156-2202.

Appendix B

Description of zooplankton abundance and presence/absence shapefiles

B-1 General Information

The shapefiles used for the abundance modelling (chapter 3) of this dissertation were derived from datasets that were assembled from online sources, papers, reports, fellow scientists, and new sample analyzing (Table B-1). The data were checked and mistakes (e.g. wrong units reported) corrected if necessary and imported into a MS Access 2010 database. The data were then brought into the same format and exported as .csv files using a Visual Basic for Applications script. Afterwards they were imported into ArcMap 10.1 and exported as shapefiles. The shapefiles were projected in the North Pole Stereographic projection, the geographic datum was WGS-84. X and y coordinates were added to the shapefiles using the Geospatial Modeling Environment (GME, http://www.spatialecology.com/gme/) ("addxy" command). The shapefiles were then spatially overlayed with environmental layers also using GME ("isectpntrst" command).

B-2 Column names and explanation of shapefiles

StationID = Database specific unique identifier for each station

Latitude = Northern Latitude in decimal degrees

Longitude = Western Longitude in decimal degrees

Mesh = Mesh size in µm used for sampling

Year = Year in which zooplankton sample was taken

Species = Zooplankton species name

Abun_m3 = Zooplankton abundance in individuals per m^{-3}

Abun_m2 = Zooplankton abundance in individuals per m^{-2}

PA = Presence/Absence; 1 = presence, 0 = absence

DatasetID = unique identifier per dataset (see Table B-1)

Taxonomist = unique identifier per taxonomist or dataset if information about individual taxonomist was unavailable (see Table B-2)

Mesh_Bin = Mesh sizes sorted into bins according to copepod size (see table A-2)

X = X coordinate added using GME

Y = Y coordinate added using GME

Ice_conc = concentration of Sea Ice in percent

Idw_Dis = Freshwater discharge from Global lakes and wetlands database level 2 in km^3 $year^{-1}$

Euc_Hydr = Euclidean Distance to hydrology/rivers in meters

Euc_coast = Euclidean Distance to coastline in meters

gbco_dgslp = bathymetric slope in percent

Gebco_ras3 = bathymetry (General Bathymetric Chart of the Oceans, GEBCO) in meters

Chla_4k = Chlorophyll a in mg m^{-3}

Euc_shelf = Euclidean Distance to the shelf break (at 200 m) in meters

Euc_gla = Euclidean Distance to glacier locations in meters

idw_ao_50 = mean apparent oxygen utilization averaged from the surface to 50 m depth in mL/L

idw_ao_200 = mean apparent oxygen utilization averaged from 60 m to 200 m depth in mL/L

idw_do_50 = mean dissolved oxygen concentration averaged from the surface to 50 m depth in mL/L

idw_do_200 = mean dissolved oxygen concentration averaged from 60 m to 200 m depth in mL/L

idw_ni_50 = mean nitrate concentration averaged from the surface to 50 m depth in µM

idw_ni_200 = mean nitrate concentration averaged from 60 m to 200 m depth in µM

idw_po_50 = mean percent oxygen saturation in percent averaged from the surface to 50 m depth

idw_po_200 = mean percent oxygen saturation in percent averaged from 60 m to 200 m depth

idw_ph_50m = mean phosphate concentration averaged from the surface to 50 m depth in µM

idw_ph_200m = mean phosphate concentration averaged from 60 m to 200 m depth in µM

idw_sa_50m = mean salinity averaged from the surface to 50 m depth (PSU)

idw_sa_200m = mean salinity averaged from 60 m to 200 m depth (PSU)

idw_si_50m = mean silicate concentration averaged from the surface to 50 m depth in µM

idw_si_200m = mean silicate concentration averaged from 60 m to 200 m depth in µM

idw_tp_50m = mean temperature in degrees Celsius averaged from the surface to 50 m depth

idw_tp_200m = mean temperature in degrees Celsius averaged from 60 m to 200 m depth

B-3 Shapefile names:

General naming of shapefiles: In_Species_Date_exp_proj.shp

In = input for modeling

Species = short name for the species

Date = Date the shapefile was created

exp = indicating that the shapefile has been exported

proj = indicating that the shapefile has been projected in North Pole Stereographic projection

In_Cfin_02022017_exp_proj.shp = shape file of abundance records for *Calanus finmarchicus*

In_Cgla_02032017_exp_proj.shp = shape file of abundance records for *Calanus glacialis*

In_Chy_02032017_exp_proj.shp = shape file of abundance records for *Calanus hyperboreus*

In_Mlo_02032017_exp_proj.shp = shape file of abundance records for *Metridia longa*

In_MePa_02032017_exp_proj.shp = shape file of abundance records for *Metridia pacifica*

In_NCris_02032017_exp_proj.shp = shape file of abundance records for *Neocalanus cristatus*

In_PGla_02032017_exp_proj.shp = shape file of abundance records for *Paraeuchaeta glacialis*

Table B-1: Dataset ID’s with corresponding datasets, source, and references

Dataset ID	Dataset	Datasource	Years	Citation/Link
2002: ID 1 2005: ID 69 2008: ID 72	Ocean Exploration	R.R.Hopcroft, K. Kosobokova	2002, 2005, 2008	Hopcroft et al. 2005: Zooplankton communities of the Arctic's Canada Basin: the contribution by smaller taxa. Polar Biology, 28(3). p. 198-206
2	Chaun Bay	A. Pinchuk	1986	Pinchuk 1994: On the zooplankton of the Chaun Bay (East Siberian Sea). Explorations of the Fauna of the Seas, 47 (55). p. 121-127 https://www.aoncadis.org/dataset/Chaun_Bay_Zooplankton_1986.html doi:10.18739/A2CK5B
3	East Siberian Sea	Retrieved from paper/ E.A. Pavshtik	1948 and 1973	Pavshtik, E.A. 1994: Composition and quantitative distribution of the zooplankton in the East Siberian Sea. [*in Russian*] https://www.aoncadis.org/dataset/Zooplankton_East_Siberian_Sea_1948_and_1973_Pavshtiks.html doi:10.18739/A2VK5P
1950: ID 4 1951: ID5	USS Burton Island	NMFS COPEPOD	1950-1951	Johnson, Martin W. 1956: The plankton of the Beaufort and Chukchi Sea areas of the Arctic and its relation to the Hydrography. Arctic Institute of North America, Technical Paper no. 1, 32pp. http://www.st.nmfs.noaa.gov/copepod/data/ussburton/index.html
6	Chukchi Sea Zooplankton	Retrieved from Paper/ E. A. Pavshtik	1976	Pavshtiks, E.A. 1984: Zooplankton of the Chukchi Sea as indices of water origins. Trudy Arkticheskogo i Antarkticheskogo Nauchno-Issledovatel′skogo Institute, 368. p. 140-153 [*in Russian*] https://www.aoncadis.org/dataset/Chukchi_Zooplankton_1976_Pavshtiks.html doi:10.18739/A2N31G
7	Kara Sea data	ArcOD/ K. Kosobokova	1997, 1999 - 2001	Hirche et al 2006: Structure and function of contemporary food webs on Arctic shelves: A panarctic comparison: The pelagic system of the Kara Sea – Communities and components of carbon flow. Progress in Oceanography,71(2-4).p.288-313 https://arcticdata.io/catalog/#view/doi:10.18739/A2089B
8	R/V Mirai	W. Walkusz	2002	Walkusz, W. et al. 2008: Zooplankton and Ichthyoplankton Data Collected from the Chukchi and Beaufort Seas during the R/V Mirai Cruise, September 2002. Can. Data Rep. Fish. Aquat. Sci. 1211: vi + 34 p.

Table B-1 continued…

1972: ID 9 1998-2003: ID 14	White Sea	ArcOD/ K. Kosobokova	1972, 1998-2003	Pertsova, N.M. and Kosobokova, K.N. 2003: Zooplankton of the White Sea: Features of the composition and structure, seasonal dynamics, and the contribution to the formation of matter fluxes. Oceanology, 43. p.108-122 https://arcticdata.io/catalog/#view/doi:10.18739/A2089B
10	Cape Bathurst NWP Lancaster Sound	L. Fortier, G. Darnis	2007	doi:10.18739/A2H890
12	North Pole Station, NP22 & NP23	ArcOD/ K. Kosobokova	1975-1977	Kosobokova, K.N. 1989: Vertical distribution of plankton animals in the eastern part of the central Arctic Basin. Explorations of the Fauna of the Seas, Marine Plankton, Leningrad 41:24 31 [*in Russian*]
13	Polarstern ARK XI-1, ARK XII, and ARK XIV	K. Kosobokova	1995-1998	Kosobokova, K.N. et al. 2010: Patterns of zooplankton diversity through the depths of the Arctic's central basins. Marine Biodiversity, 41(1), p.29-50
15	T-3 Ice Island & Arctic Ice Dynamics Joint Experiment (AIDJEX)	ArcOD	1970-1972, 1975	Pautzke C.G. 1979: Copepoda collected from the Canada Basin Arctic Ocean; Fletcher's Ice Island (T-3) 1970-1972 and AIDJEX, 1975
17	Archives of the Arctic Seas Zooplankton	E. Markhaseva/ Zoological Institute of Russian Academy of Sciences, Saint Petersburg	1921, 1930, 1934, 1935, 1938, 1939,1946-1948, 1950, 1952, 1954-1956, 1970-1973	Markhaseva, E. et al. 2005: Archives of the Arctic Seas Zooplankton: Contributions from the Zoological Institute RAS. No 8. 44p.

Table B-1 continued…

18	Bering Strait Zooplankton	ArcOD, A. Pinchuk & K. Coyle, UAF	1991	Coyle, K.O., Chavtur, V.G., Pinchuk, A.I. 1996 Zooplankton of the Bering Sea: A review of Russian-language literature. Alaska Sea Grant. 97-133 Piatt J.F., A.I. Pinchuk, A.S. Kitaisky, A.M. Springer, S.A. Hatch 1992 Foraging distribution and feeding ecology of seabirds at the Diomede Islands, Bering Strait. U.S. Fish and Wildlife. Serv. Final Report for Minerals Management Service OCS Study MMS 92-0041, Anchorage, Alaska: 133 pp
19	Chukchi Sea Zooplankton/R V Lomonosov	ArcOD/ A. Pinchuk	1953,1954	UAF - Arctic Ocean Chukchi Sea August 1953-1954 Zooplankton vertical stratified collections on board of the Russian R/V Lomonosov, program ANII A-65
20	Food web uptake of persistent organic pollutants in the Arctic Marginal Ice Zone of the Barents Sea (FAMIZ)	E. Arashkevich, A. Pasternak	2001	Olli et al 2007: The fate of production in the central Arctic Ocean–top–down regulation by zooplankton expatriates? Progress in Oceanography, 72(1). p.84-113
1971: ID 21 1970: ID 22 1972: ID 85	Western Beaufort Sea Ecological Cruises (WEBSEC)	NMFS COPEPOD/ C. Clarke Hopcroft	1970-1972	Kinds and Abundance of Zooplankton collected by the USCG Icebreaker GLACIER in the Eastern Chukchi Sea, September-October 1970. NOAA Tech. Rep. NMFS SSRF-679. http://www.st.nmfs.noaa.gov/copepod/data/websec/index.html Hufford et al. 1974: WEBSEC 71-72 an ecological survey in the Beaufort Sea. Oceanographic Report No. CG 373-64. United States Coast Guard. Oceanographic Unit, Washington, D.C.
23	ARK IX-4	K. Kosobokova	1993	Kosobokova et al. 1998: Composition and distribution of zooplankton in the Laptev Sea and adjacent Nansen Basin during summer, 1993. Polar Biology, 19(1). p. 63-76 https://www.aoncadis.org/dataset/Laptev_Sea_Nansen_Basin_Zooplankton_1993_Polarstern_ARKIX4.html doi:10.18739/A2WK6C

Table B-1 continued…

25	Surface Heat Budget of the Arctic (SHEBA)	C. Ashjian & R. Campbell	1997, 1998	Ashjian et al 2003: Annual cycle in abundance, distribution, and size in relation to hydrography of important copepod species in the western Arctic Ocean. Deep Sea Research I, 50(10-11). p. 1235-1261
44	ARK III	entered from Dissertation / S. Diel	1985	S. Diel 1991: On the life history of dominant copepod species (*Calanus finmarchicus*, *C. glacialis*, *C. hyperboreus*, *Metridia longa*) in the Fram Strait. Berichte zur Polarforschung, 88
45	Canadian Arctic Archipelago	entered from report/ A.A. Mohammed and E.H. Grainger	1962	Mohammed, A.A. and Grainger, E.H. 1974: Zooplankton data from the Canadian Arctic Archipelago, 1962. Technical report No. 460, Fisheries and Marine Service, Environment Canada
46	Study of the Alaskan Coastal System (SNACS)	C. Ashjian	2005, 2006	Ashjian et al. 2010: Climate variability, oceanography, bowhead whale distribution, and Iñupiat subsistence whaling near Barrow, Alaska. Arctic, 63(2). p. 179-194
49	Mackenzie	W. Walkusz	2005, 2006	Walkusz, W. et al. 2010: Distribution, diversity and biomass of summer zooplankton from the coastal Canadian Beaufort Sea. Polar Biology, 33(3). p. 321-335
50	Oshoro Maru	Matsuno et al.	1991, 1992, 2007, 2008	Matsuno et al. 2011: Year-to-year changes of the mesozooplankton community in the Chukchi Sea during summers of 1991, 1992 and 2007, 2008. Polar Biology, 34(9). p. 1349-1360
53	Zooplankton Canada Basin and Chukchi Borderlands, R/V Mirai	A. Yamaguchi, K. Matsuno/ Hokkaido University	2008	Matsuno, K. et al. 2012: Horizontal distribution of calanoid copepods in the western Arctic Ocean during the summer of 2008. Polar Science, 6(1). p. 105-119
56	North East Water Polynya (NEWP)	C. Ashjian	1992 and 1993	Ashjian et al. 1997: Distribution of zooplankton in the Northeast Water Polynya during summer 1992. Journal of Marine Systems. 10(1-4).p. 279-298
58	Shelf Basin Interactions (SBI)	Lane et al.	2004	Lane et al. 2008: Zooplankton Data Report: Distribution of Zooplankton in the Western Arctic during Summer 2004. RSMAS Technical Report Number 2008-02.

Table B-1 continued…

2004: ID 60 2009: ID 71	Russian American Longterm Census of the Arctic (RUSALCA)	R.R. Hopcroft, K. Kosobkova, A. Pinchuk, E. Ershova	2004 and 2009	Hopcroft et al. 2010: Zooplankton community patterns in the Chukchi Sea during summer 2004. DSR II: Topical Studies in Oceanography. 57(1 -2). p. 27-39 Ershova, E. et al. 2015: Inter-annual variability of summer mesozooplankton communities of the western Chukchi Sea: 2004–2012. Polar Biology, 38(9). p.1461-1481
61	Transdrift I-III	S. Lischka		Lischka et al. 2001: Mesozooplankton assemblages in the shallow Arctic Laptev Sea in summer 1993 and autumn 1995. Polar Biology, 24(3). p. 186-199
62	Estrada et al. 2012	M. Harvey	2003-2006	Estrada, R. et al. 2012: Late-summer zooplankton community structure, abundance, and distribution in the Hudson Bay system (Canada) and their relationships with environmental conditions, 2003–2006. Progress in Oceanography, 101(1). p.121-145
63	Zooplankton Data from the Beaufort Sea	entered from report/ E.H. Grainger and K. Grohe	1951-1975	Grainger, E.H. and K. Grohe 1975: Zooplankton Data from the Beaufort Sea, 1951 to 1975. Technical Report No. 591, Environment Canada, Fisheries and Marine Service, 54 pp. https://www.aoncadis.org/dataset/Beaufort_Sea_Zooplankton_1951_1975_Grainger_Grohe.html doi:10.18739/A2430F
65	NOGAP B2	ArcOD/ Hopky, Lawrence and Chiperzak	1984-1988	Hopky et al. 1994: NOGAP B2; Zooplankton data from the Canadian Beaufort Sea Shelf, 1984 and 1985. DFO Canada Hopky et al. 1994: NOGAP B2; Zooplankton data from the Canadian Beaufort Sea Shelf, 1986. DFO Canada Hopky et al. 1994: NOGAP B2; Zooplankton data from the Canadian Beaufort Sea Shelf, 1987 and 1988. DFO Canada
66	Zackenberg Research Station	Zackenberg Research Station/ Greenland	2003-2008	http://zackenberg.dk/monitoring/marinebasis/ Jensen, L.M. and Rasch, M (eds.) 2009: Zackenberg Ecological Research Operations, 14th Annual Report, 2008. Roskilde, National Environmental Research Institute, Aarhus University, 2009, 116 pp.
73	LSSL	I. Rutzen & R.R. Hopcroft	2003-2006	https://www.aoncadis.org/dataset/Canada_Basin_Zooplankton_2003_2006.html doi:10.18739/A27S4S
74	Canada's Three Oceans (C3O)	J. Nelson	2007-2009	Pomerleau, C. et al. 2011: Spatial patterns in zooplankton communities across the eastern Canadian sub-Arctic and Arctic waters: insights from stable carbon (δ13C) and nitrogen (δ15N) isotope ratios. Journal of Plankton Research, 33(12).p. 1779-1792

Table B-1 continued…

75	T-3 Ice Island	ArcOD	1966-1969	Scott D.A.: Copepoda collected from Fletcher's Ice Island (T-3) in the Canadian Basin of the Arctic Ocean, 1966-1970. Technical Report No. 240 Reference M69-62
76	USCGC Chelan	NMFS COPEPOD	1934	http://www.st.nmfs.noaa.gov/copepod/data/uscgchelan/index.html
77	Driftalpha	NMFS COPEPOD	1957, 1958	Johnson, M.W. 1963: Zooplankton collections from the high polar basin with special reference to the copepoda. Limnology and Oceanography, 8(1). p.89-102 http://www.st.nmfs.noaa.gov/copepod/data/driftalpha/index.html
78	Minoda, Drift Station Arlis II	NMFS COPEPOD	1964	Minoda, T. 1967: Seasonal distribution of Copepoda in the Arctic Ocean from June to December, 1964. Recent Oceanic Works in Japan, 9, p.161-168 http://www.st.nmfs.noaa.gov/copepod/data/minoda1967/index.html
79	ARK IV-3	NMFS COPEPOD/ N. Mumm	1987	Mumm, N. 1993: Composition and distribution of mesozooplankton in the Nansen Basin, Arctic-Ocean, during summer; Polar Biology,13(7).p. 451-461
82	ICEBAR	Norwegian Polar Institute/S. Falk-Petersen	1996	Falk-Petersen et al 2000: Physical and ecological processes in the marginal ice zone of the northern Barents Sea during the summer melt period. Journal of Marine Systems, 27(1).p.131-159
84	INFERNO	Melle, W.	1997, 2005-2008	Hjøllo, S.S. et al. 2012: Modelling secondary production in the Norwegian Sea with a fully coupled physical/primary production/individual-based Calanus finmarchicus model system. Marine Biology Research, 8(5-6). p. 508-526
86	On thin Ice (OTI)	Norwegian Polar Institute/ S. Falk-Petersen	2003	Blachowiak-Samolyk et al. 2007: Trophic structure of zooplankton in the Fram Strait in spring and autumn 2003. Deep Sea Research II, 54(23). p. 2716-2728
87	Barents Sea Ecosystem Program - FB section	IMR/P. Dalpadado	1995-2012	Gjøsaeter, H. et al. 2002: Growth of Barents Sea capelin (Mallotus villosus) in relation to zooplankton abundance. ICES Journal of Marine Science, 59(5). p. 959-967
89	Ice1, Ice2	Pangaea/ S. Falk-Petersen	2004	Falk-Petersen et al. 2008: Vertical migration in high Arctic waters during autumn 2004. Deep Sea Research II, 55(20-21). p. 2275-2284
91	BeauFish	C. A. Smoot	2011	Smoot, C.A. 2015: Contemporary Mesozooplankton Communities of the Beaufort Sea. M.S. Thesis, University of Alaska, Fairbanks

Table B-1 continued…

92	Frobisher Bay Station 5	entered from report/ E.H. Grainger	1967-1970	Fisheries Research Board of Canada, Report 266; Biological oceanographic observations in Frobisher Bay. II Zooplankton Data, 1967-1970, E.H. Grainger https://www.aoncadis.org/dataset/Frobisher_Bay_Zooplankton_1967-1971_Grainger.html doi:10.18739/A2RS32
93	Chukchi Sea Environmental Studies Program (CSESP)	ACADIS/R.R. Hopcroft, J.M. Questel	2008-2012	Questel et al. 2013: Seasonal and interannual variation in the planktonic communities of the northeastern Chukchi Sea during the summer and early fall. Continental Shelf Research, 67. p. 23-41 https://www.aoncadis.org/dataset/ucar.ncar.eol.dataset.255_038.html doi:10.5065/D63R0QX7
94	ARK XXVI	K. Kosobokova	2011	Unpublished data
95	Bering-Aleutian Salmon International Survey (BASIS)	L. Eisner/K. Coyle	2002-2007, 2009-2011	Helle, J., E. Farley, J. Murphy, A. Feldmann, K. Cieciel, J. Moss, L. Eisner, J. Pohl, and M. Courtney. 2007. The Bering-Aleutian Salmon International Survey (BASIS). AFSC Quarterly Report Feature (January-February-March 2007), 5 p. http://www.afsc.noaa.gov/ABL/MESA/archives/mesa_occ_basis.htm
96	Iceland Sea Ecosystem Project	A. Gislason	2006-2008	Gislason, A., and Silva, T. 2012: Abundance, composition, and development of zooplankton in the Subarctic Iceland Sea in 2006, 2007, and 2008. ICES Journal of Marine Science, 69(7). p. 1263-1276.

Table B-2: Taxonomist ID. In some cases taxonomists are assigned several taxonomist IDs if they were appearing in combination with other taxonomists in one dataset. If no taxonomist was available, a negative number was assigned for the dataset

Taxonomist	TaxonomistID
C. Clarke Hopcroft	1
A. Pinchuk	2
-1	3
-2	4
M. W. Johnson	5
E.A. Pavshtiks	6
E. Kolosova	7
-3	8
K.N. Kosobokova	9
-4	10
-5	11
-6	12
Barkalova, G.V. Boldovskoy, M.K. Jashnova, Khromova, D. Potemkina, M.A. Virketis	13
-7	14
-8	15
Arashkevich, Pasternak	16
-9	17
-10	18
C. Ashjian, R. Campbell	19
S. Diehl	20
Mohammed, E.H.Grainger	21
C. Ashjian	22
-11	23
-11	23
-12	24
-13	25
-14	26
-15	27
K.A. Brodskii	28
-16	29
K.N. Kosobokova, R.R. Hopcroft, A. Pinchuk	30
S. Lischka	31
-17	32
E.H. Grainger, Grohe	33
-18	34
-19	35
-20	36
K.N. Kosbokova, R.R. Hopcroft	37
E.A. Ershova	38

Table B-2 continued…

-21	39
I. Rutzen	40
-22	41
-23	42
-24	43
-25	44
-26	45
N. Mumm	46
-27	47
-28	48
-29	49
-30	50
-31	51
-32	52
J. Ranning	53
C.S. Smoot	54
E.H. Grainger	55
T. Bernstein	56
V.G. Bogorov	57
Dzen'	58
L.L. Rossolimo	59
Tchislenko	60
M.A. Virketis	61
-34	62
-35	63
-36	64
-37	65
-38	66
-39	67
-40	68
-41	69
-42	70
L. Eisner	71
J. Questel	72
J. Questel, P. Hariharan	73
J. Lam	74
C. Clarke Hopcroft, Chris Stark	75
J. Questel, P. Hariharan, C. Stark, E.A. Ershova	76
K.N. Kosobokova, R.R. Hopcroft, A. Pinchuk, E.A. Ershova	77
AR	78
Berit Endresen	79
Bjarnar Ellertsen	80
Laura Rey	81

Table B-2 continued…

Penny Lee Liebig	82
A. Gislason	83

Appendix C

Environmental variables and sources

C-1 General Description

The here presented environmental layers were produced to describe the habitat of zooplankton and to model their distribution and abundance (see chapter 3). They can be applied to a variety of analysis. All rasters cover the area between the North Pole and 60° northern latitude and were all projected in the North Pole Stereographic projection in ArcMap 10.1. The geodetic datum is WGS-84. All layers were clipped to their current extent using the extract by mask tool (Spatial Analyst >> Extraction >> Extract by Mask). Clipping masks were generated from the bathymetry layer (Spatial Analyst >> Reclass >> Reclassify). To create a clipping mask that accounts for the surface area, values below 400 m were reclassified as zero and other values were assigned to "noData". The reason 400 m was used instead of 0 m for the surface extend was that with anything below 400 m, small areas in estuaries were excluded. Table 3.3 (chapter 3) gives an overview of the environmental layers described in this appendix.

C-2 World Ocean Atlas 2013

Temperature, salinity, oxygen and nutrient variables were downloaded as shapefiles (.shp) from the World Ocean Atlas 2013 (WOA13) from the National Oceanic and Atmospheric Administration (NOAA). The grids presented here are for two different averaged depth levels. The first one extends from the surface to 50 meters depth, the second one is an average from 60 m to 200 m depth. The grids display the statistical mean during the summer months (July to September, average of 6 decadal means (1955-2012)).

The data have a resolution of 1 degree. Values were integrated in ArcMap 10.1 using inverse distance weighting tool of the spatial analyst tools (Spatial Analyst >> Interpolation >> IDW). Clipping masks were derived from bathymetry layer of the General Bathymetric Chart of the Oceans (GEBCO) using the reclassify tool in the spatial analyst toolbox in ArcMap 10.1. Depth levels for clipping masks started at the surface, five meters, followed by a 10 m depth increase until 100 meters and then every 25 m until 200 m. Afterwards, clipping masks were used to derive data for the different depth levels from the WOA13 files. For this, the spatial analyst tools, extract by mask tool was applied. The data were averaged to the surface to 50 m, and 60 m to 200 m depth levels using the cell statistics tool (Spatial analyst >> local >> cell statistics), where the "ignore NoData in calculations" was checked.

Source: https://www.nodc.noaa.gov/OC5/indprod.html
Version: WOA 13 Version 1

File names:
aomn_50m = Mean apparent oxygen utilization surface to 50 m (Fig. C-1)
aomn_200m = Mean apparent oxygen utilization 60 m to 200 m (Fig. C-2)
domn_50m = Mean dissolved oxygen surface to 50 m (Fig. C-3)
domn_200m = Mean dissolved oxygen 60 m to 200 m (Fig. C-4)
nimn_50m = Mean nitrate concentration surface to 50 m (Fig. C-.5)
nimn_200m= Mean nitrate concentration 60 m to 200 m (Fig. C-6)
phomn_50m = Mean phosphate concentration surface to 50 m (Fig. C-7)
phomn_200m = Mean phosphate concentration 60 m to 200 m (Fig. C-8)
posmn_50m = Mean percent oxygen saturation surface to 50 m (Fig. C-9)
posmn_200m = Mean percent oxygen saturation 60 m to 200 m (Fig. C-10)
samn_50m = Mean salinity surface to 50 m (Fig. C-11)
samn_200m = Mean salinity 60 m to 200 m (Fig. C-12)
simn_50m = Mean silicate concentration surface to 50 m (Fig. C-13)
simn_200m = Mean silicate concentration 60 m to 200 m (Fig. C-14)
tpmn_50m = Mean temperature surface to 50 m (Fig. C-15)
tpmn_200m = Mean silicate concentration 60 m to 200 m (Fig. C-16)

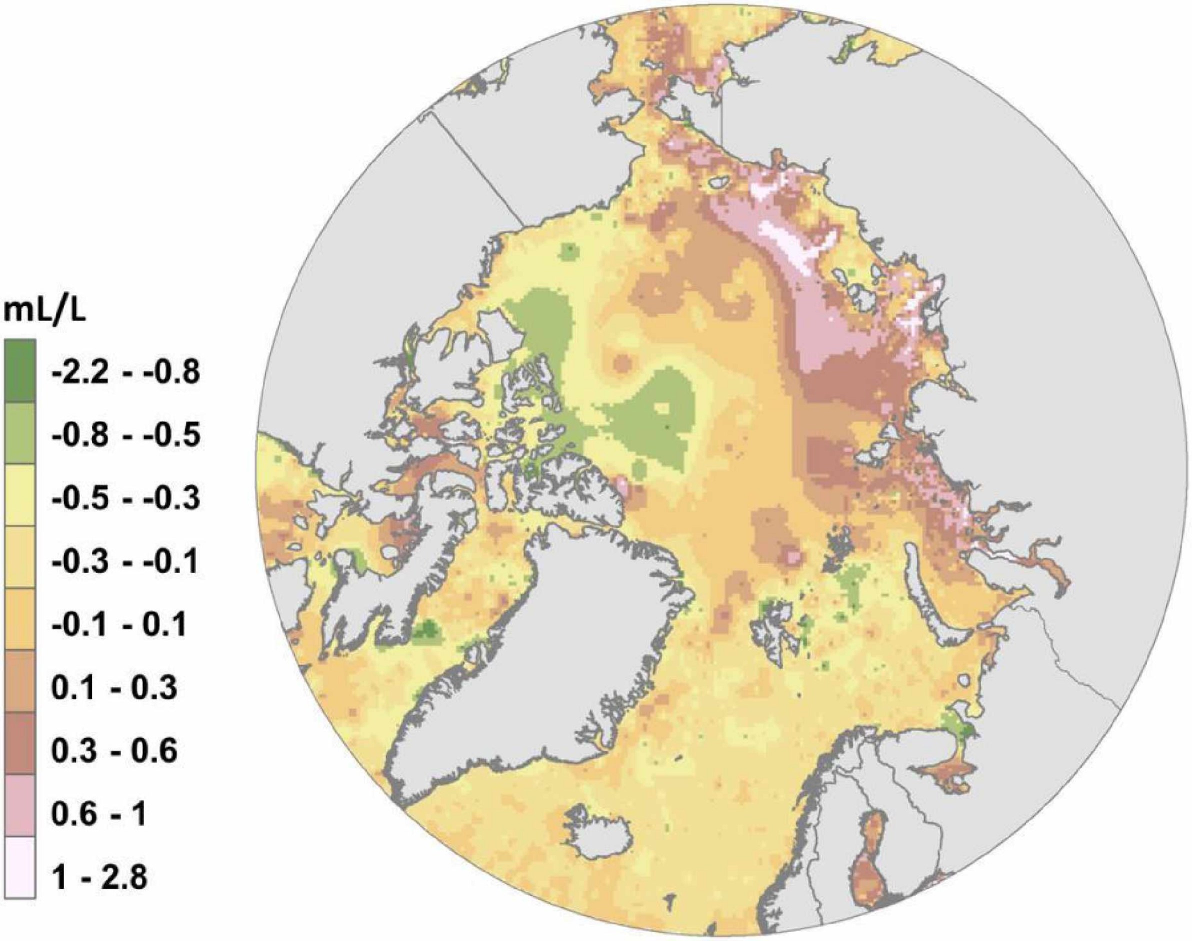


Figure C-1: Mean apparent oxygen utilization from the surface to 50 m in mL/L

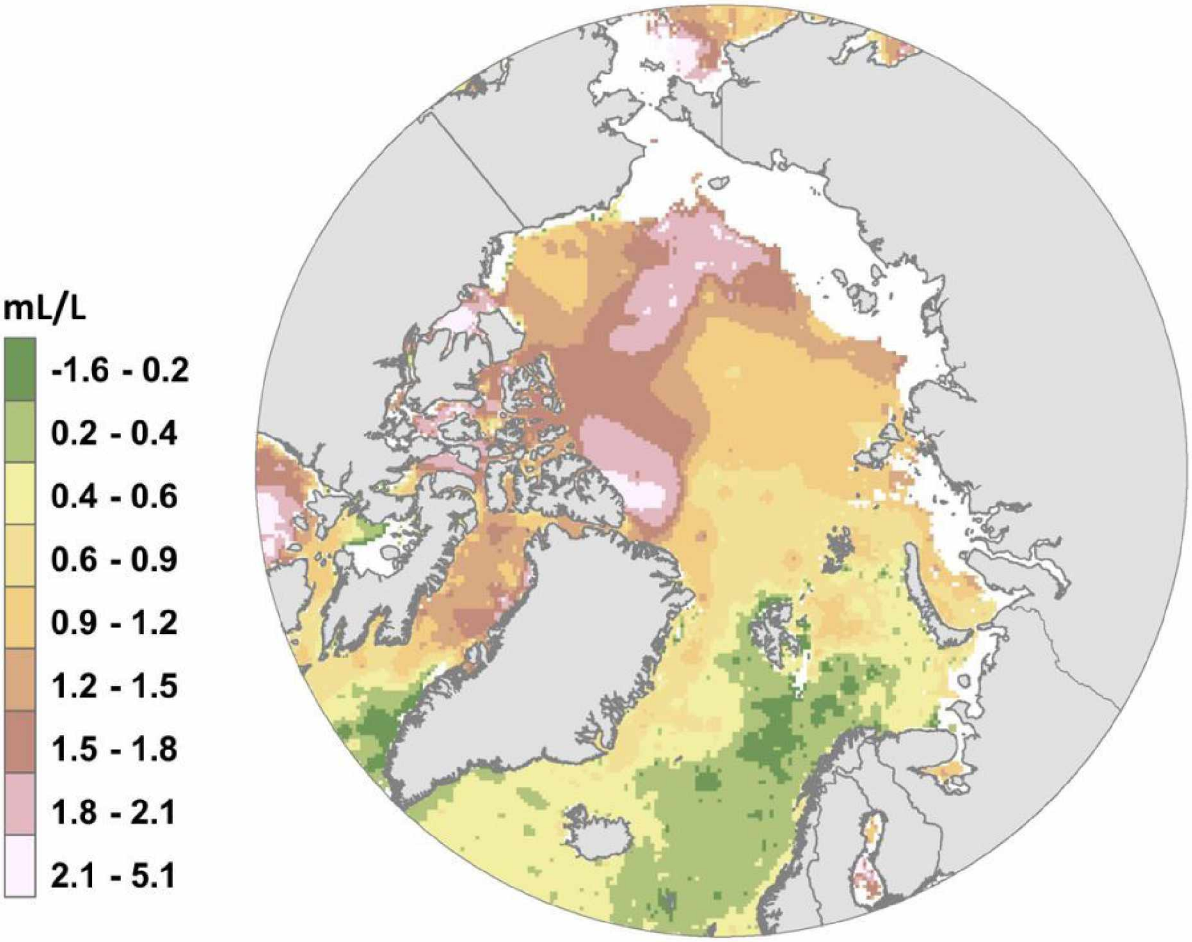


Figure C-2: Mean apparent oxygen utilization from 60 m to 200 m in mL/L. White areas = NoData

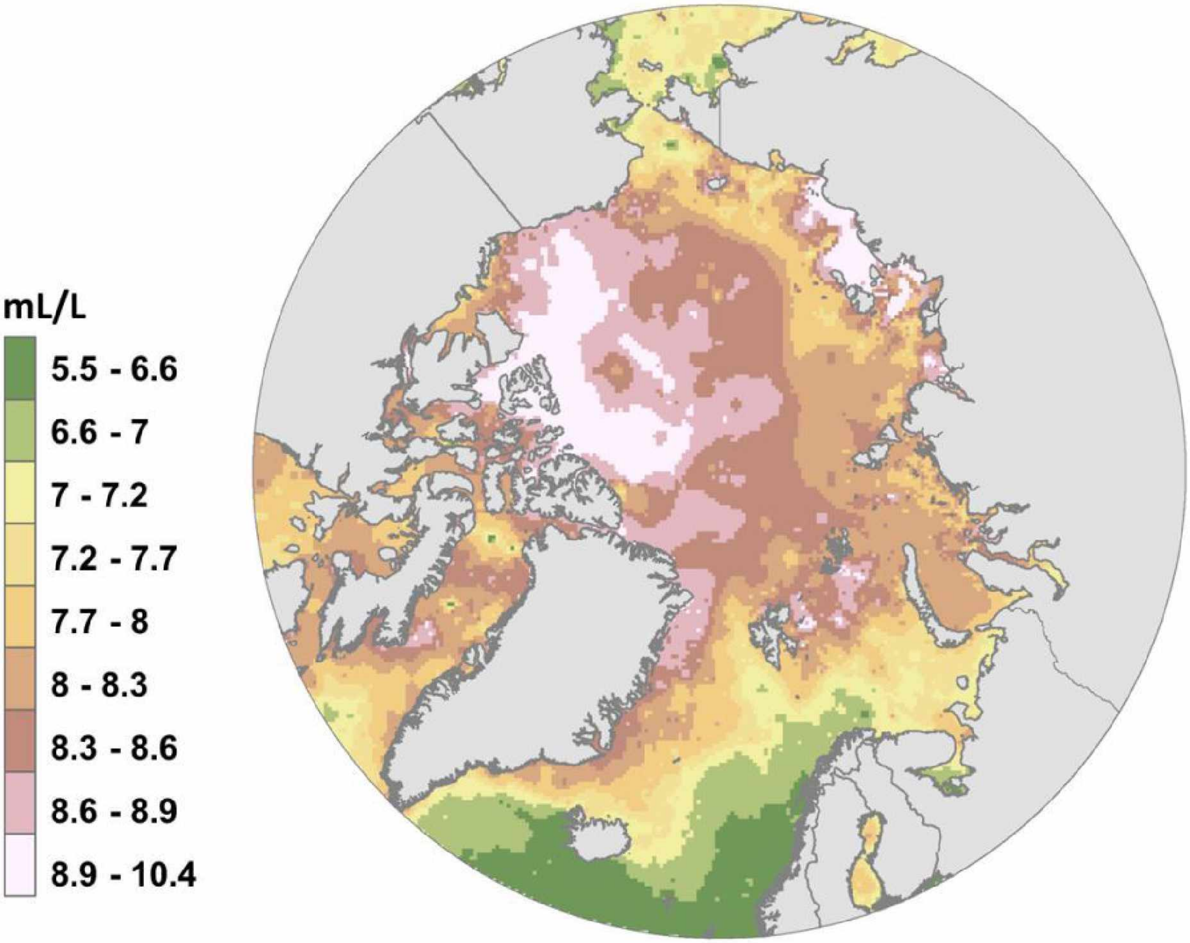


Figure C-3: Mean dissolved oxygen concentration from the surface to 50 m in mL/L

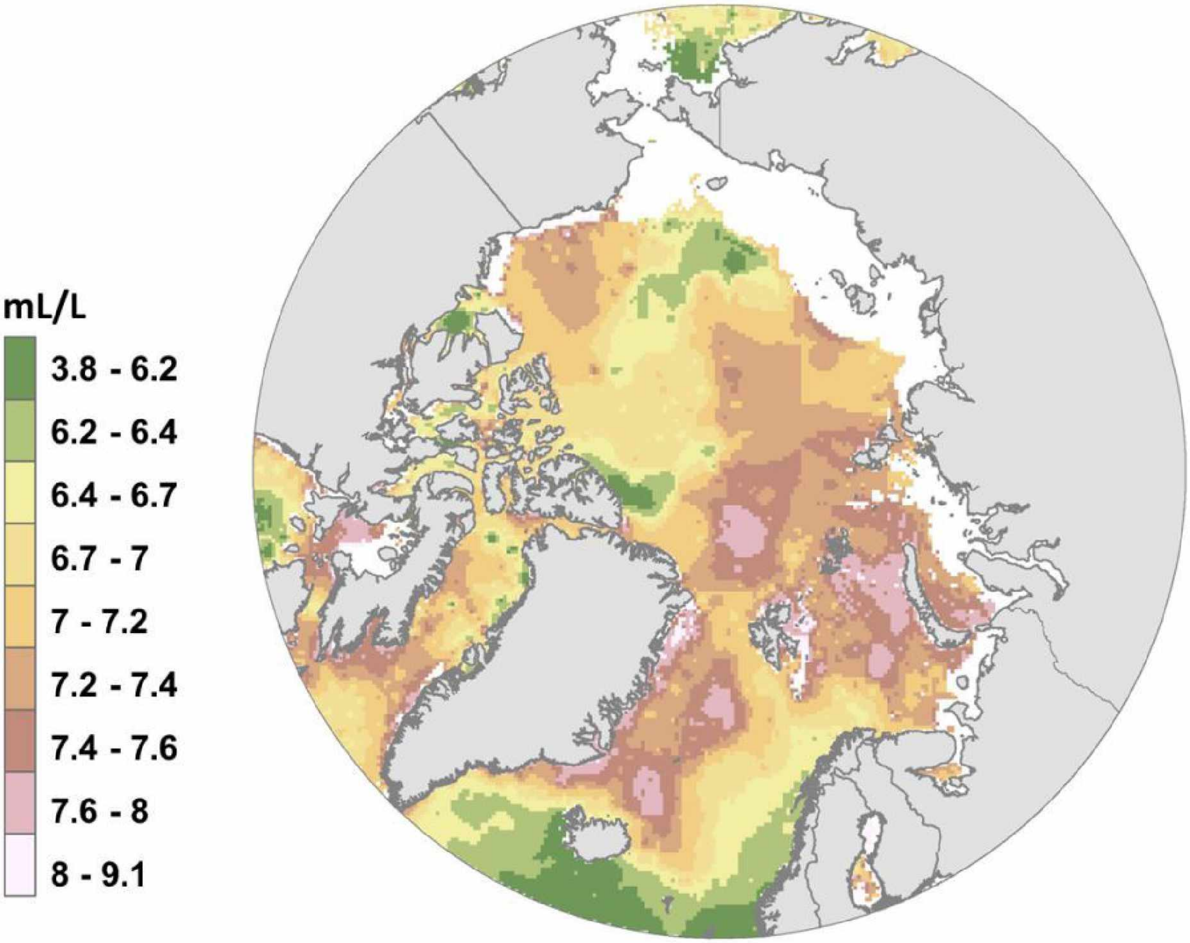


Figure C-4: Mean dissolved oxygen concentration from 60 m to 200 m in mL/L. White areas = NoData

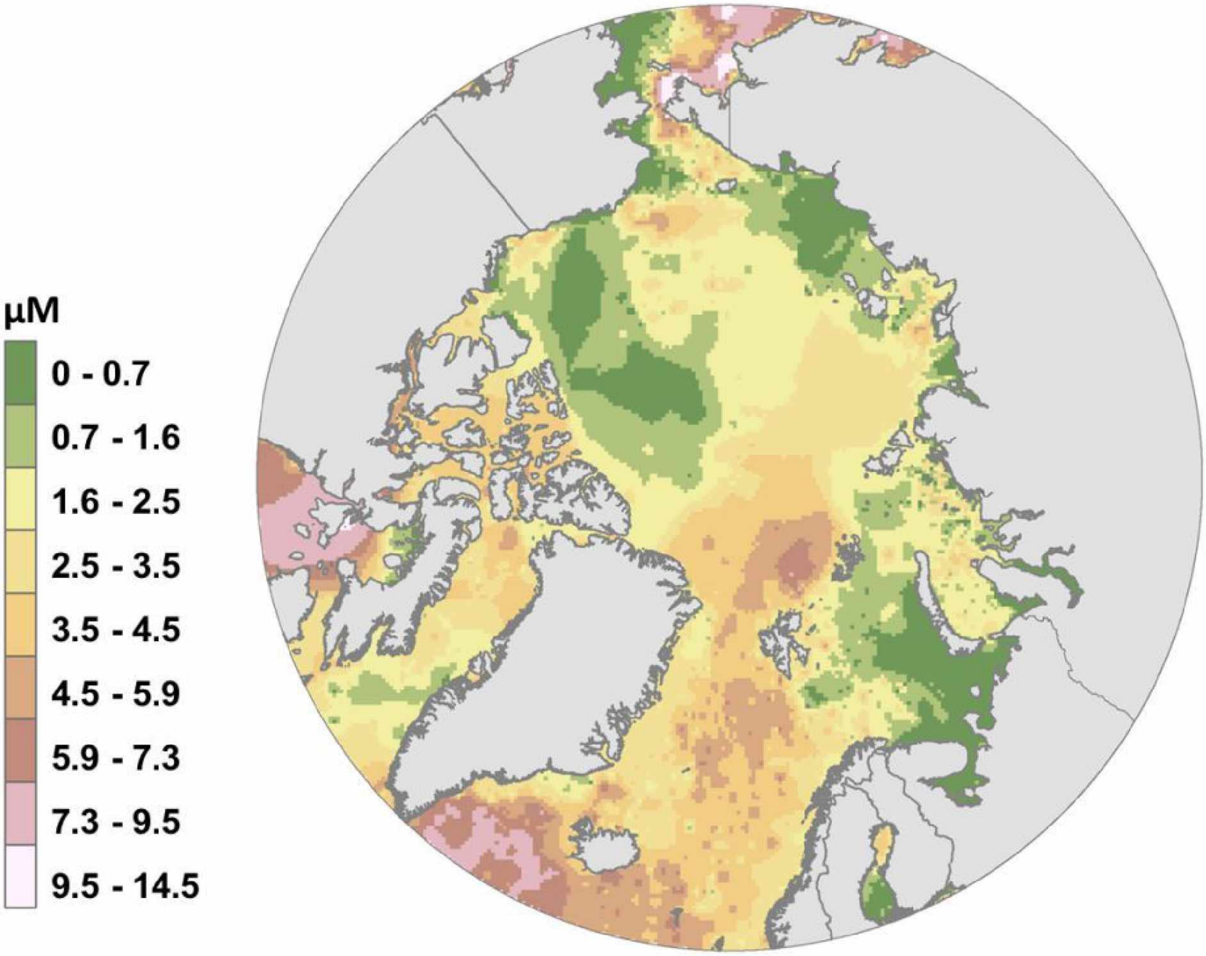


Figure C-5: Mean nitrate concentration from the surface to 50 m in µM

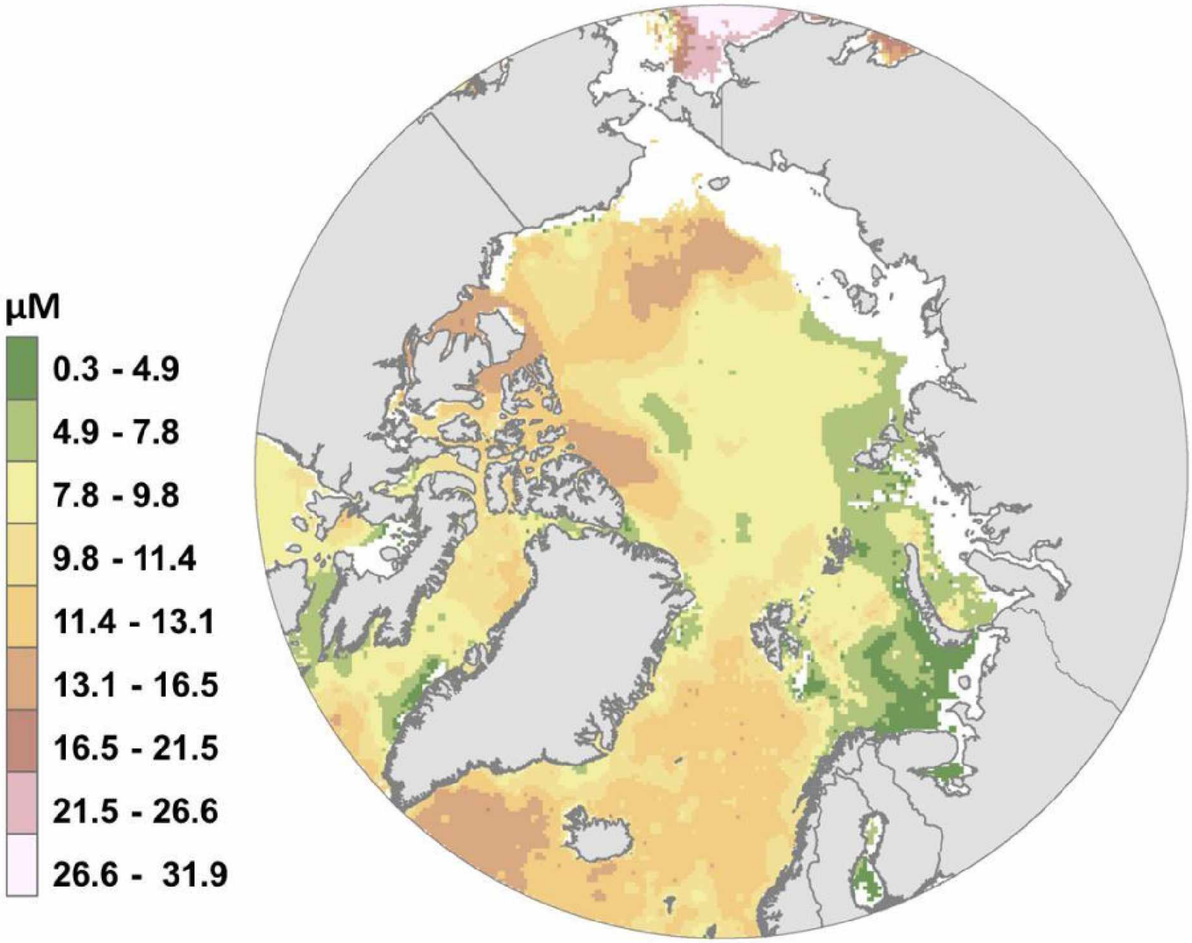


Figure C-6: Mean nitrate concentration from 60 m to 200 m in µM. White areas = NoData

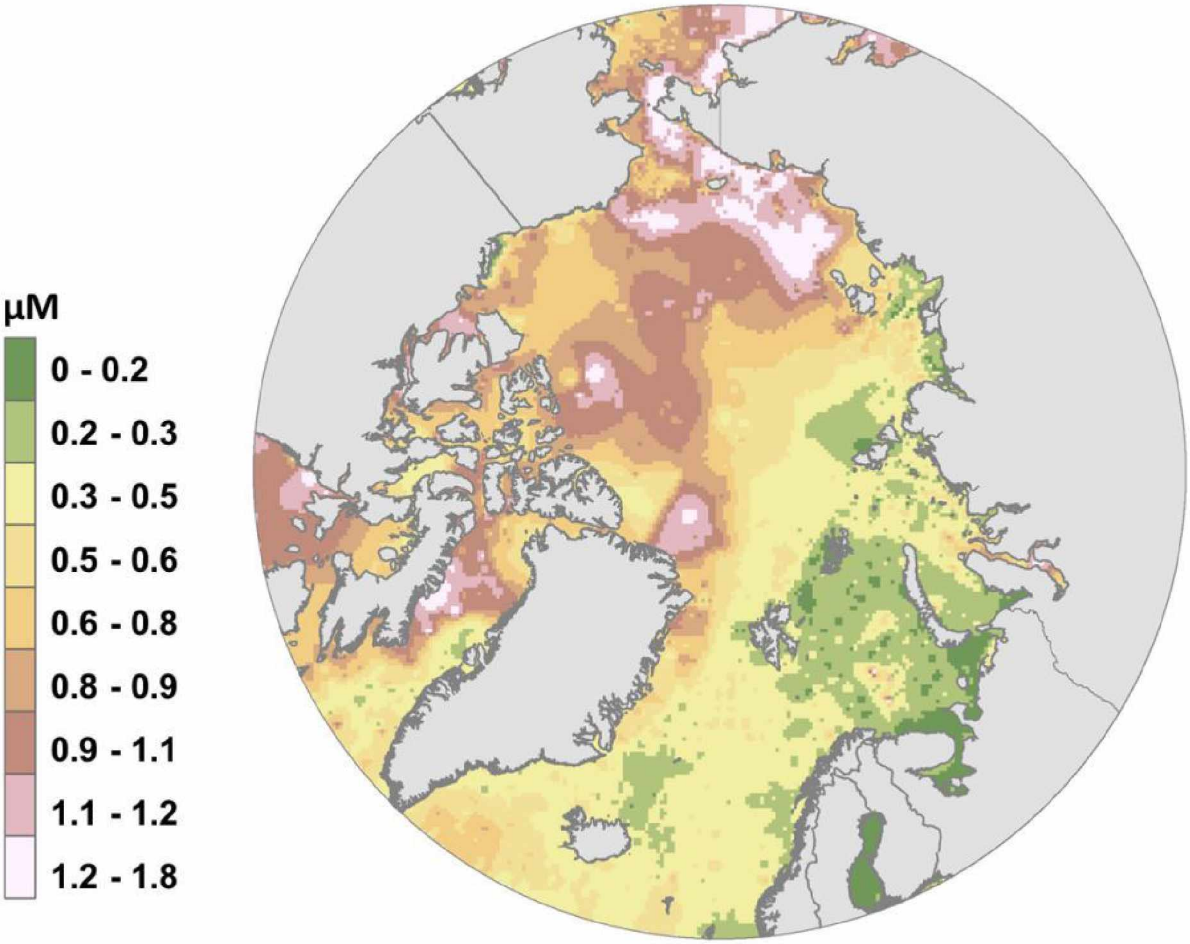


Figure C-7: Mean phosphate concentration from the surface to 50 m in µM

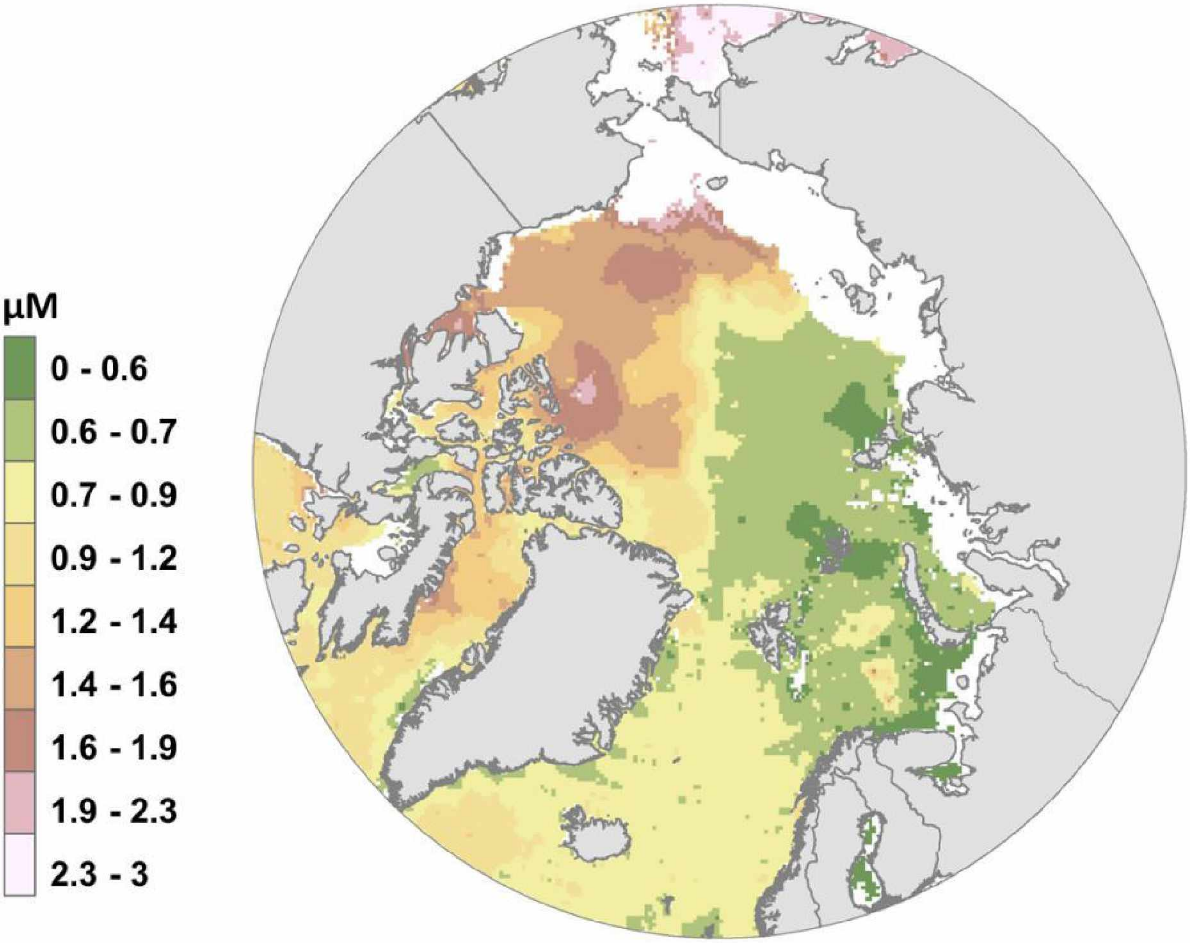


Figure C-8: Mean phosphate concentration from 60 m to 200 m in µM. White areas = NoData

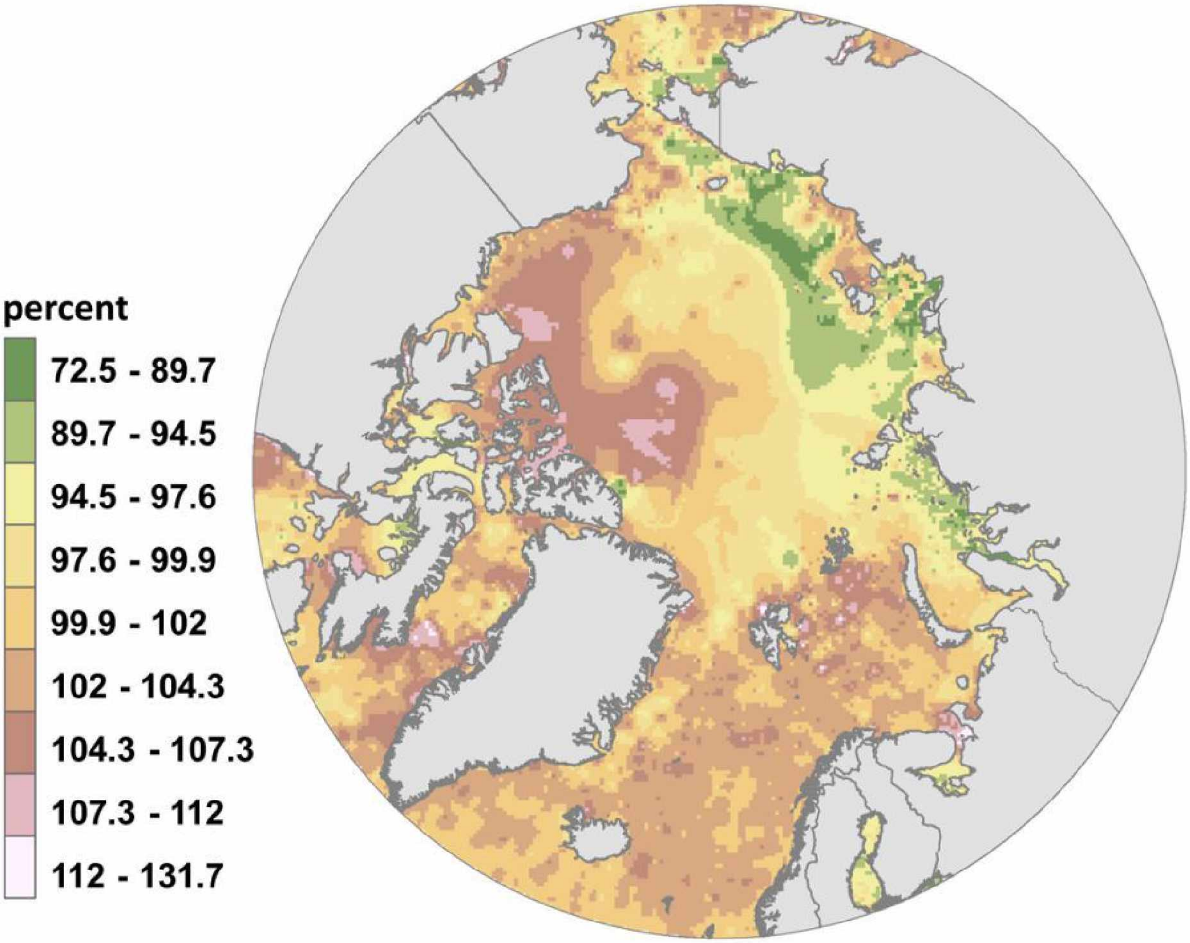


Figure C-9: Mean percent oxygen saturation in percent from the surface to 50 m

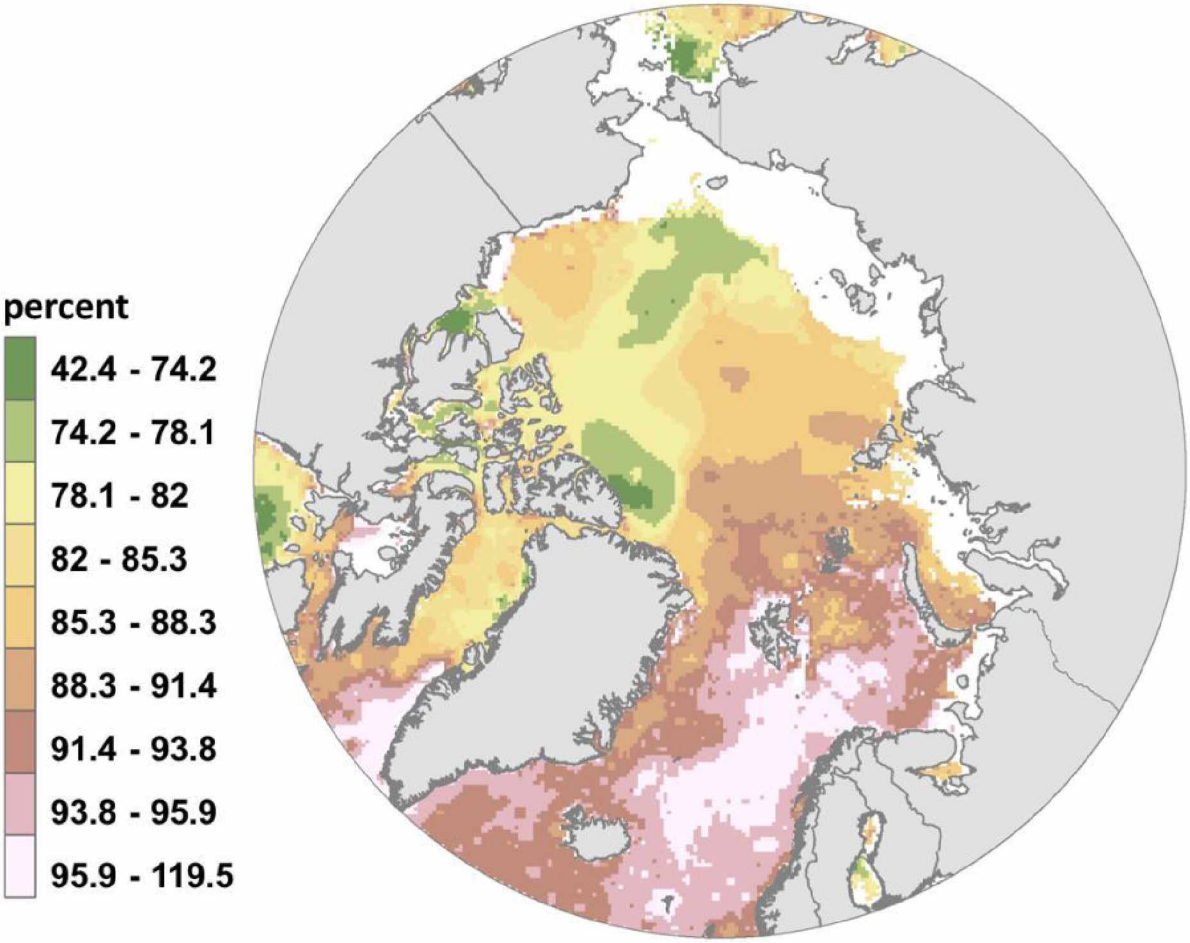


Figure C-10: Mean percent oxygen saturation in percent from 60 m to 200 m. White areas = NoData

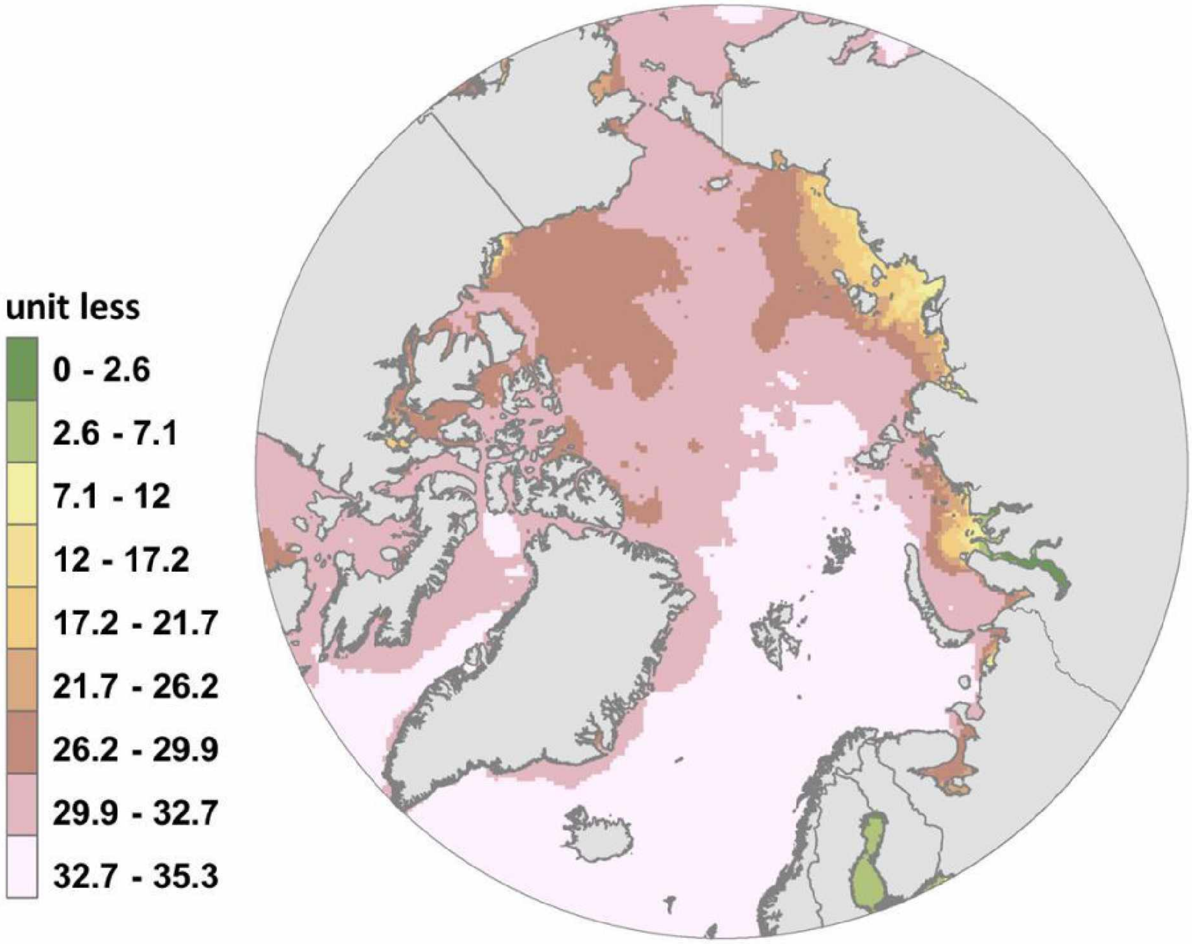


Figure C-11: Mean salinity (PSU) from the surface to 50 m

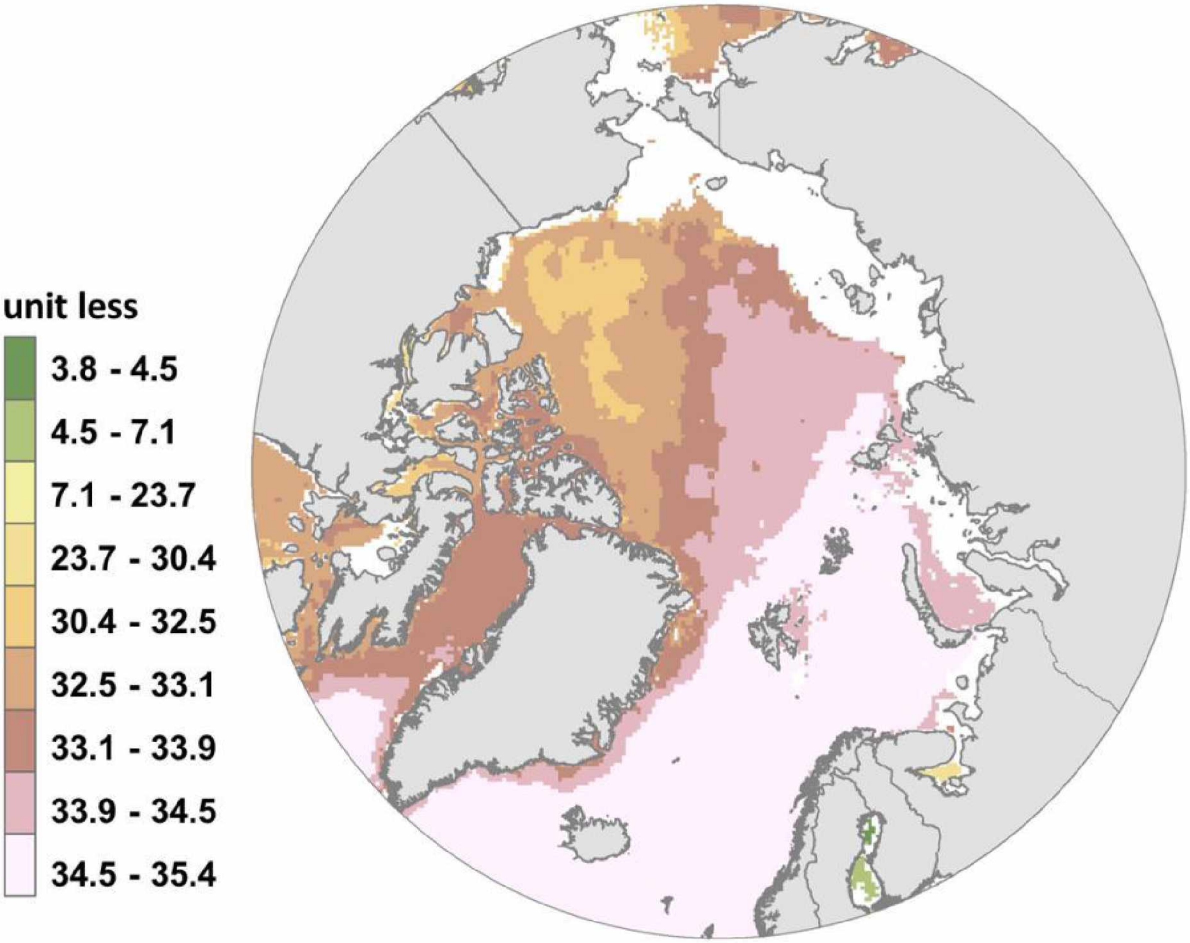


Figure C-12: Mean salinity (PSU) from 60 m to 200 m. White areas = NoData

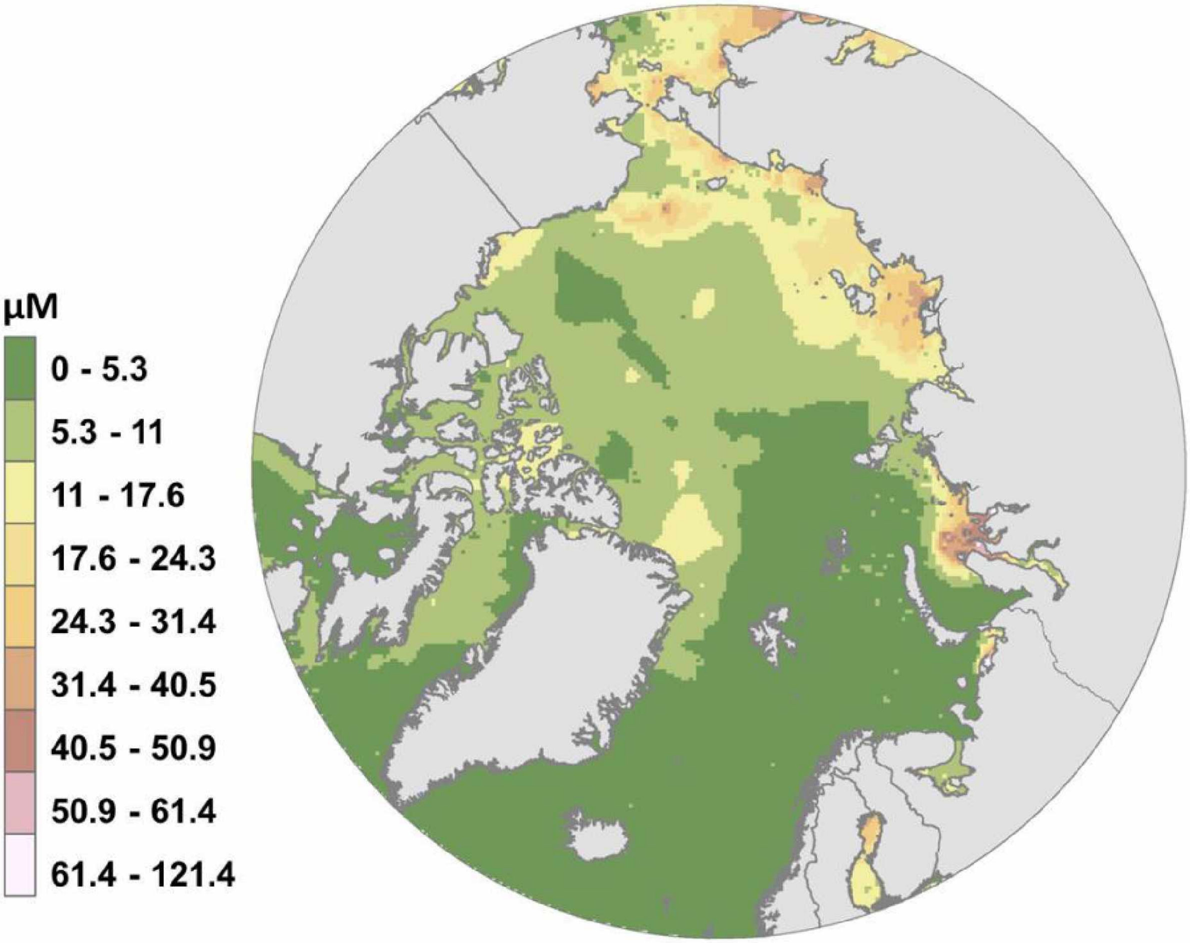


Figure C-13: Mean silicate concentration from the surface to 50 m in µM

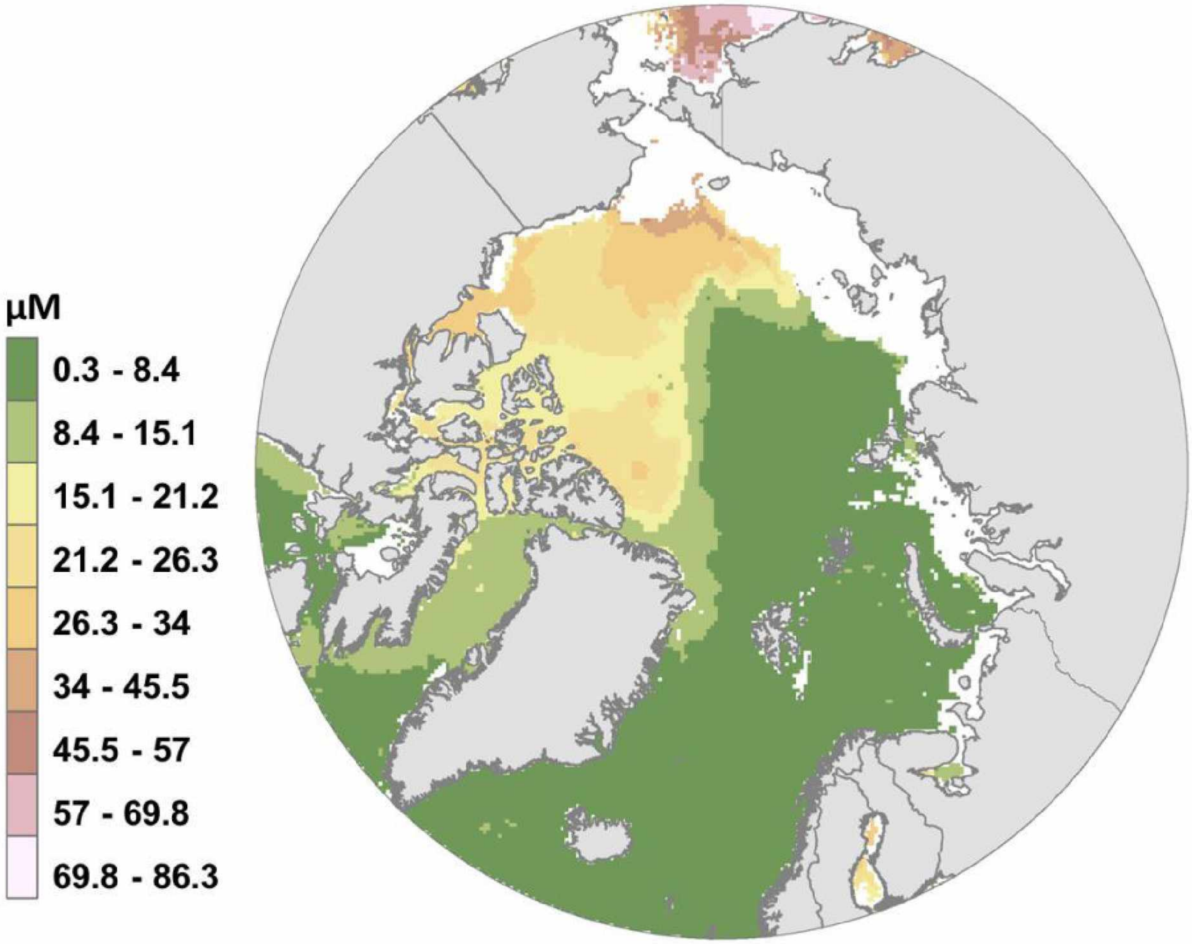


Figure C-14: Mean silicate concentration from 60 m to 200 m in µM. White areas = NoData

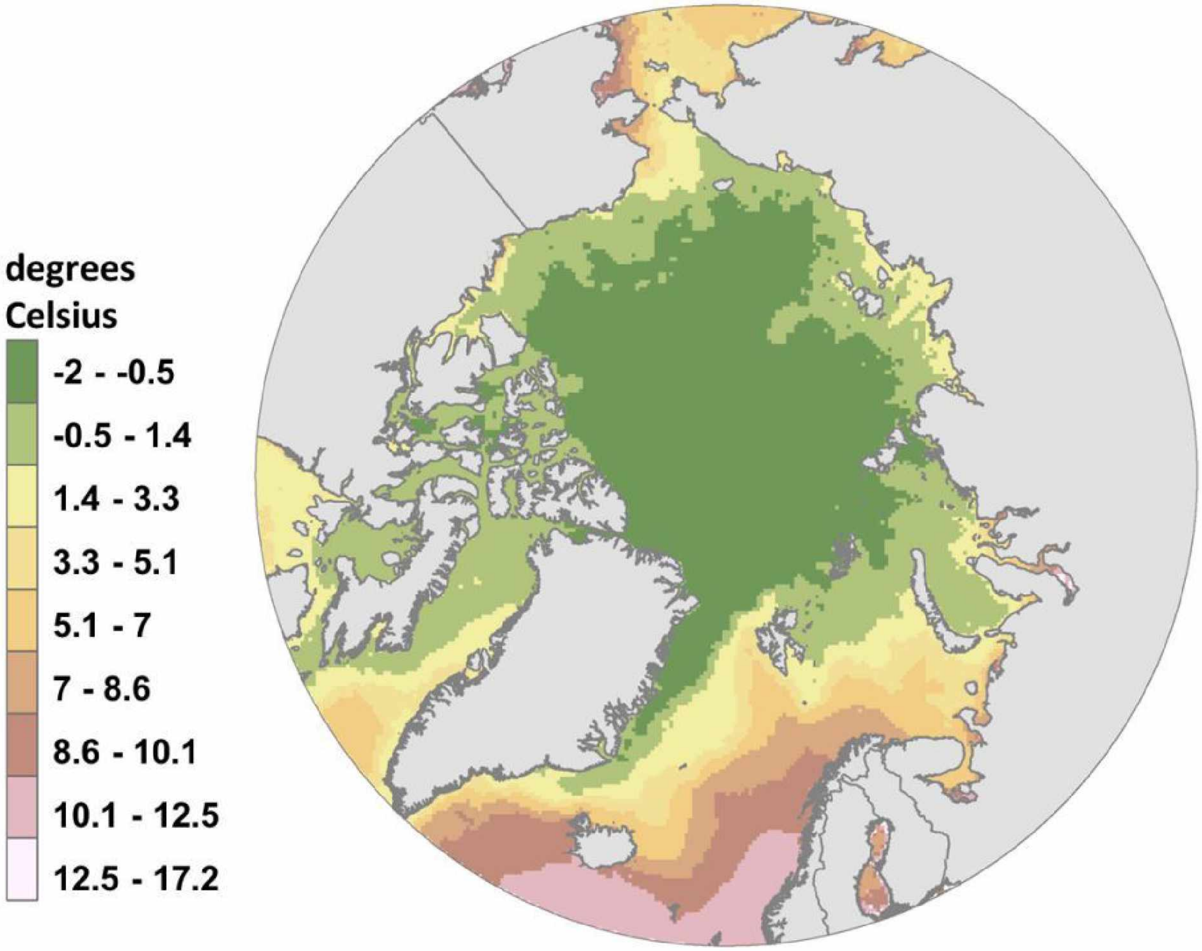


Figure C-15: Mean temperature in degrees Celsius from the surface to 50 m

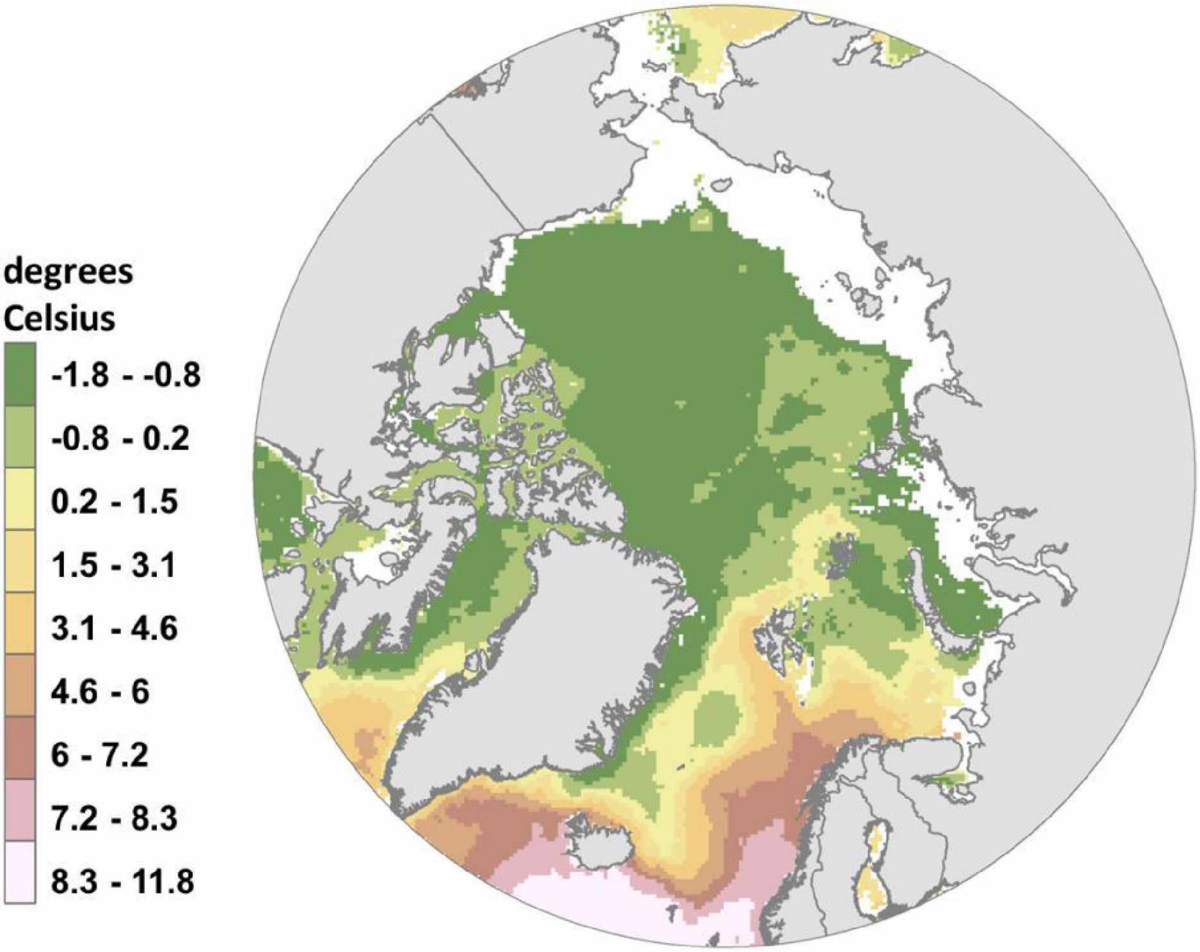


Figure C-16: Mean temperature in degrees Celsius from 60 m to 200 m. White areas = NoData

C-3 General Bathymetric Chart of the Oceans (GEBCO)

For the bathymetry, The General Bathymetric Chart of the Oceans (GEBCO) version gebco_08 with a 30 arc-seconds resolution was downloaded as a netCDF file. It was then displayed using the GEBCO grid display software from the British Oceanographic Data Center (BODC) (https://www.bodc.ac.uk/about/news_and_events/gebco_grid_software.html). Within the software coordinates were selected (90°N to 60°N and 180°W to 180°E) and exported as an ASCII file. The ASCII file was converted to a raster (floating point), using the ArcMap conversion tools, ASCII to raster.

Source: http://www.gebco.net/data_and_products/gridded_bathymetry_data/
Grid version: The GEBCO_08 Grid, version 1396469250786, http://www.gebco.net

Filename: gebco_proj2

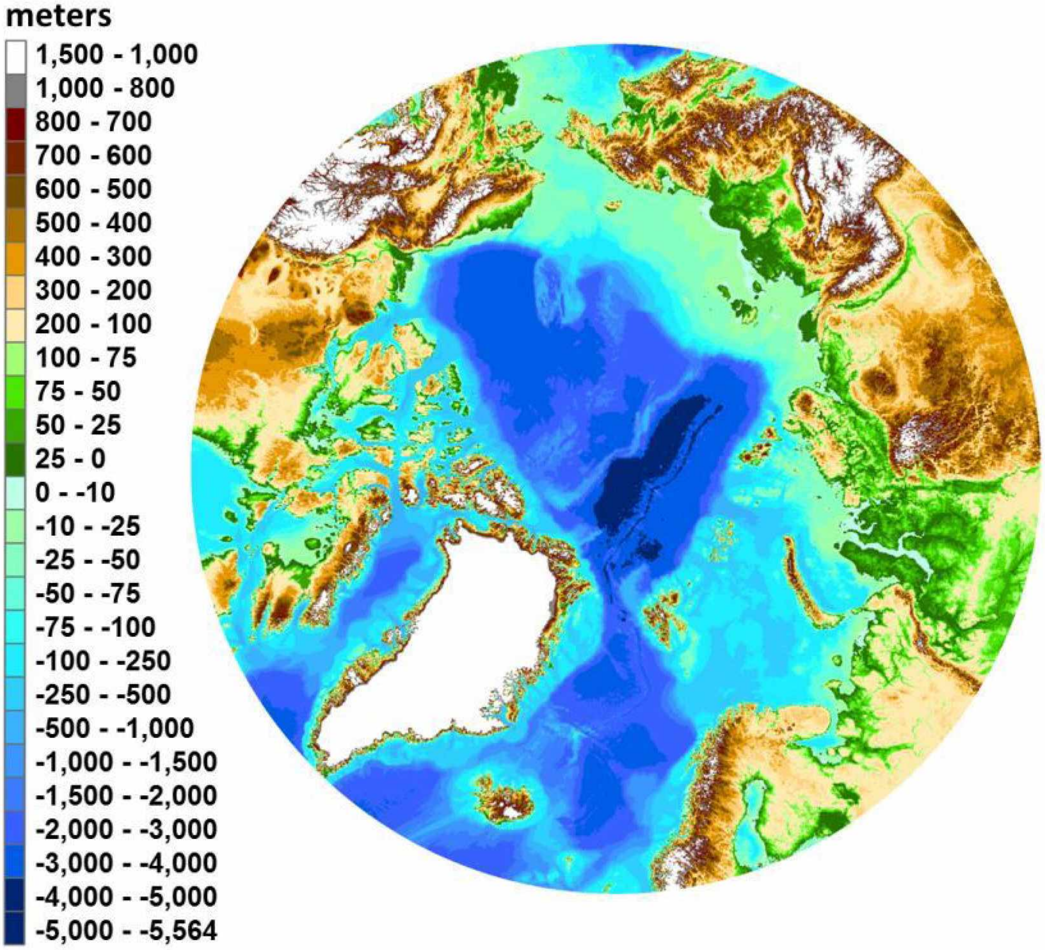


Figure C-17: GEBCO Bathymetry in meters

C-4 Bathymetric Slope

The bathymetric slope in degrees (Fig. C-18) was generated from the GEBCO bathymetry (see C-17), using the slope tool in the spatial analyst toolbox (Spatial Analyst >> Surface >> Slope). The resolution of the grid is 30 arc-seconds.

Filename: cl_dgslp400

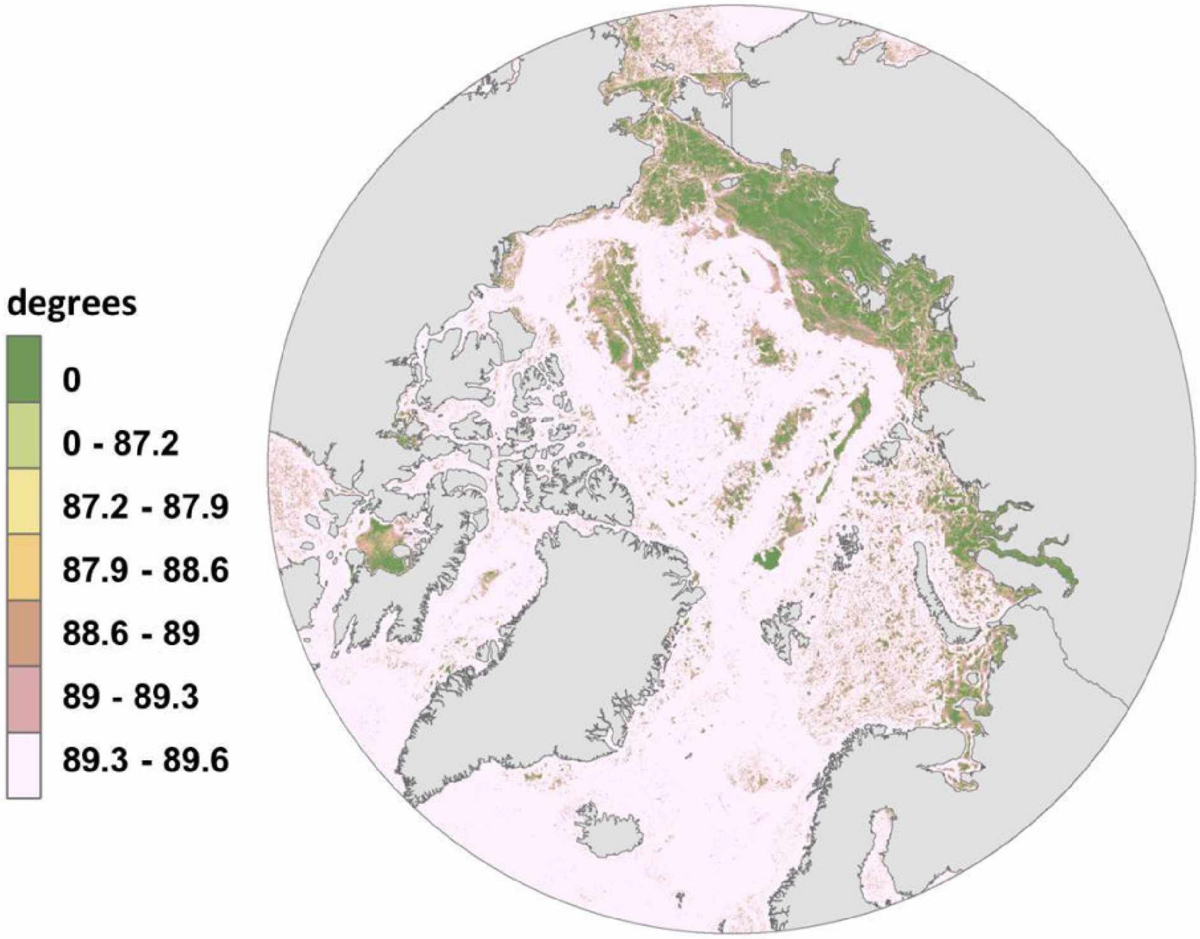


Figure C-18: Bathymetric slope

C-5 Euclidean Distance to glaciers

Location data of glaciers were downloaded from NSIDC as a shapefile. The temporal coverage of the data is from 1900 to 2003. The shapefile clipped using a polygon covering the study area (Analysis Tools >> Extract >> Clip). Afterwards, the Euclidean distance tool (Spatial Analyst >> Distance >> Euclidean Distance) was applied to calculate the Euclidean distance to the glacier locations (Fig. C-19) with a cell size of 27411.505 x 27411.505 meters.

Source: http://nsidc.org/data/docs/noaa/g01130_glacier_inventory/
Version: wgi_shapefile_feb2012.shp
Filename: cl_gl4

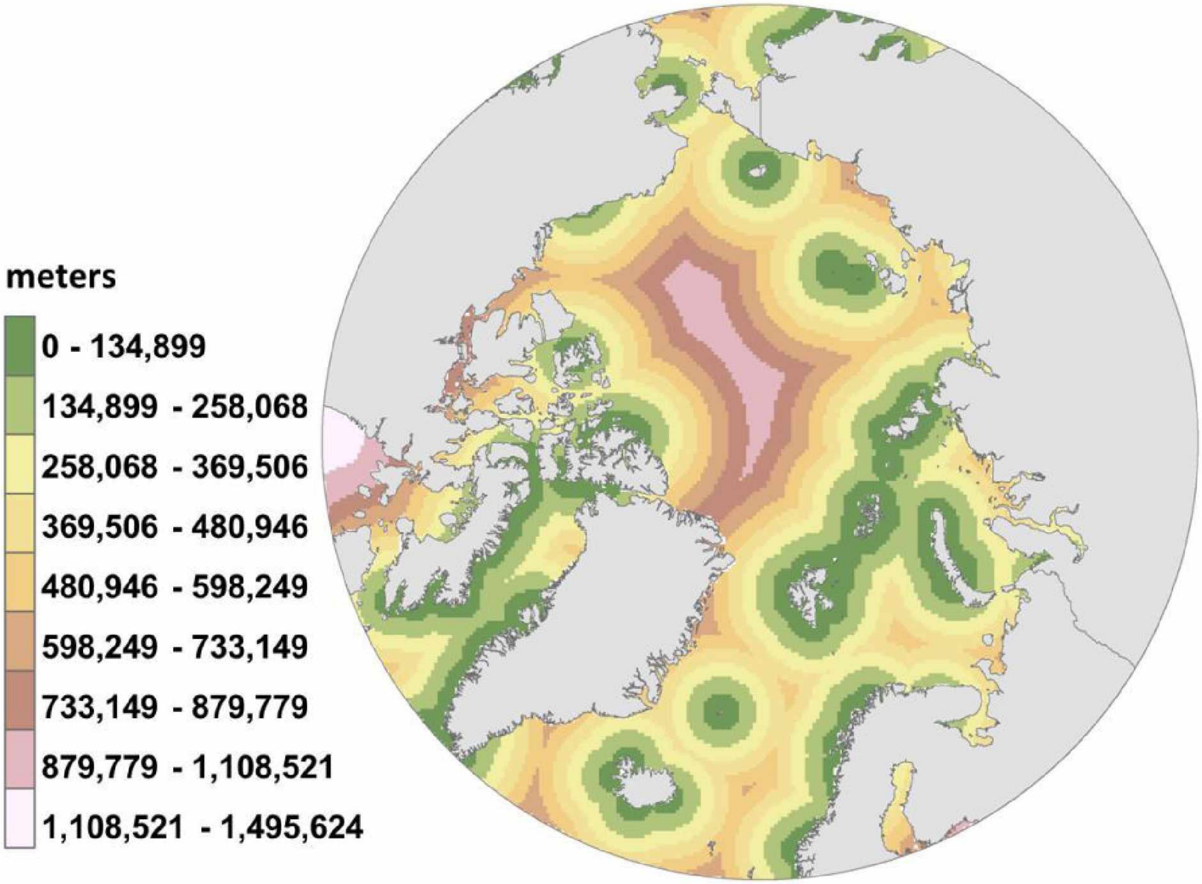


Figure C-19: Euclidean Distance to glaciers in meters

C-6 Euclidean Distance to coastline

The coastline with a scale of 1:250,000 was extracted from the NOAA's World Vector Shoreline for the study area in the Arc/Info Ungenerate format. Due to the size of the extracted area, it had to be downloaded in eleven files. Every file ranged from 90°N to 60°N latitude and the longitude increased in steps of 30 starting at 180°W (180°W-150°W; 150°W-120°W; etc.). The script gen2shap1.ave (https://pubs.usgs.gov/of/1998/of98-502/chapt6/shape_files/gen2shap1.ave) was applied in ArcView 3.0 to convert Arc/Info Ungenerate formats into shapefiles. Next, the shapefiles were imported into ArcGIS 9.2 and merged using the merge tool (Data Management Tools >> Merge) to construct a single coastline file. Based on this file, the Euclidean distance to the coastline was calculated with a cell size of 27411.505 x 27411.505 meters (Fig. C-20).

Source: https://www.ngdc.noaa.gov/mgg/shorelines/
https://shoreline.noaa.gov/data/datasheets/wvs.html

Filename: cleucoast400

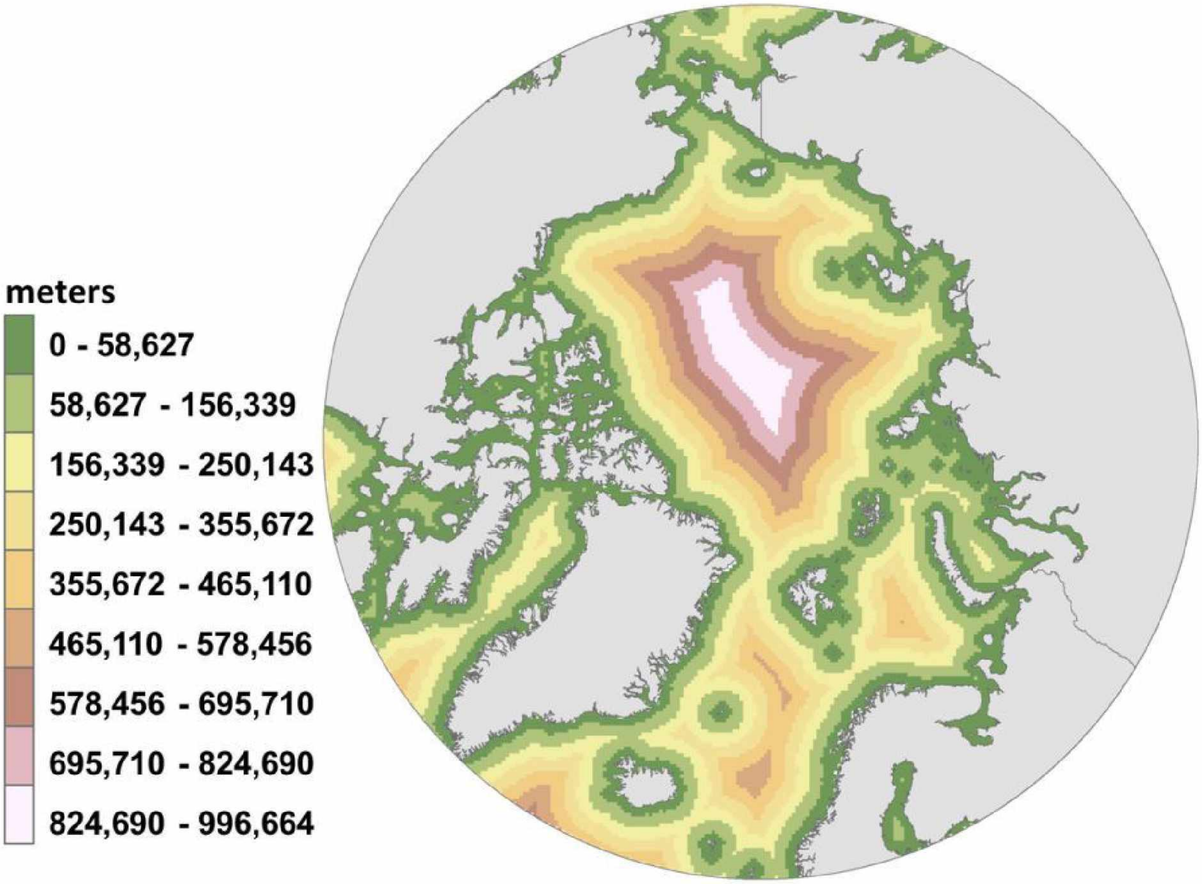


Figure C-20: Euclidean Distance to the coastline in meters

C-7 Freshwater Discharge from rivers

Monthly freshwater discharge (km^3 year $^{-1}$) of the major rivers into the Arctic Ocean was downloaded from R-ArcticNet version 4.0 as a .xlsx file. The discharge data cover a temporal window from 1877 to 2003. The data for the months of July, August, and September were averaged in MS Excel 2010. Afterwards, they were turned into shapefiles in ArcMap 10.1, and the Inverse Distance Weighted tool was applied (Spatial Analyst >> Interpolation >> IDW) to produce a raster with a cell size of 27411.505 x 27411.505 meters, covering the entire study area (Fig. C-21).

Source: http://www.r-arcticnet.sr.unh.edu/v4.0/index.html

Version: Vers. 4

Filename: cl_dchrg

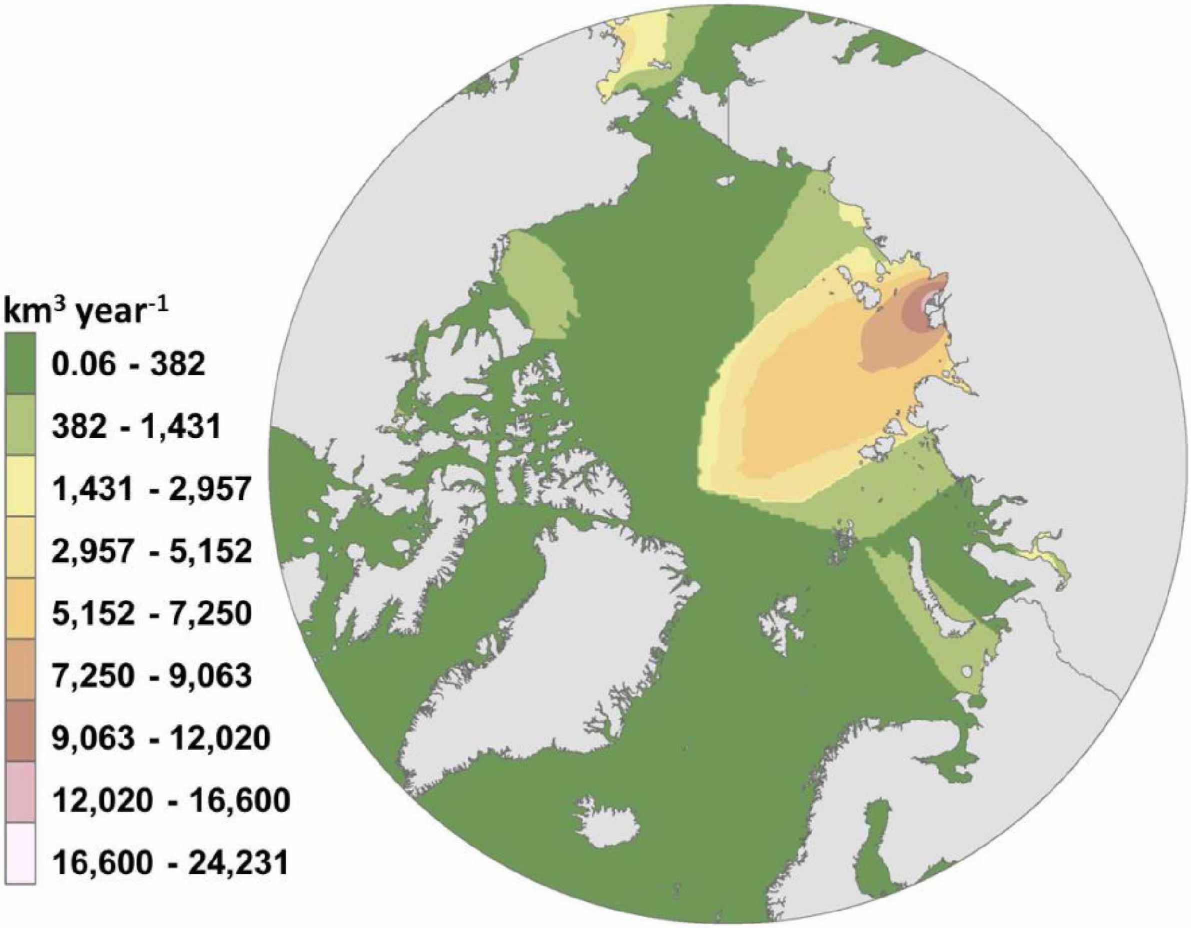


Figure C-21: Freshwater discharge of the major rivers in km^3 year^{-1}

C-8 Sea Ice concentration

Sea ice concentration in percent was downloaded from the National Snow and Ice Data center (NSIDC) as Flat binary (1-byte scaled, unsigned integers) files for the months July, August, and September from NSIDC. The temporal coverage is from 1979 to 2012 and the spatial resolution is 25 km. The data were converted in ArcMap 10.1 to a raster using the steps explained in the support protocol: https://support.nsidc.org/entries/22256950-How-do-I-import-the-data-into-ArcGIS. The monthly rasters were then averaged using the cell statistics tool

(Spatial Analyst >> Local >> Cell Statistics) to create one raster averaged over July, August, and September (Fig. C-22).

Source: http://nsidc.org/data/docs/daac/nsidc0051_gsfc_seaice.gd.html

Import protocol: https://support.nsidc.org/entries/22256950-How-do-I-import-the-data-into-ArcGIS

Filename: clice_400m

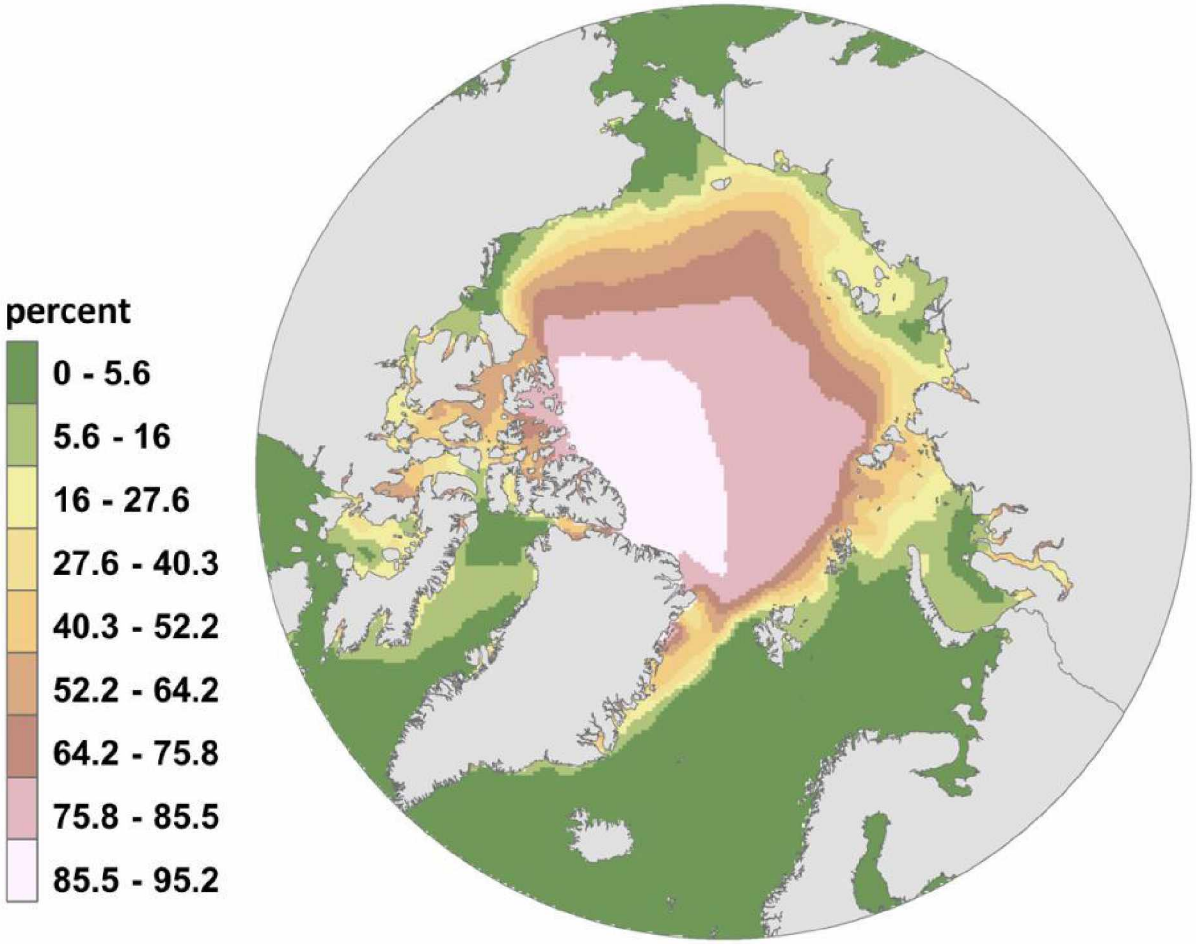


Figure C-22: Mean sea ice concentration in percent

C-9 Chlorophyll a

Data were downloaded from NASA Aqua MODIS satellite data as a monthly climatology for July to September 2002 to 2014. The resolution is 4 km. The data were imported into ArcMap 10.1 and the Marine Geospatial Ecology tools version 0.8a56 (http://mgel.env.duke.edu/mget) were applied and a tutorial https://code.env.duke.edu/projects/mget/wiki/Converting%20HDFs%20to%20ArcGIS%20rasters) was followed to convert the data from .hdf files to a raster (Fig. C-23) (MGET >> Conversion >> To ArcGIS Raster >> Convert SDS in HDF to ArcGIS Raster).

Source: https://oceancolor.gsfc.nasa.gov/cgi/l3
Filename: chla_4kclp

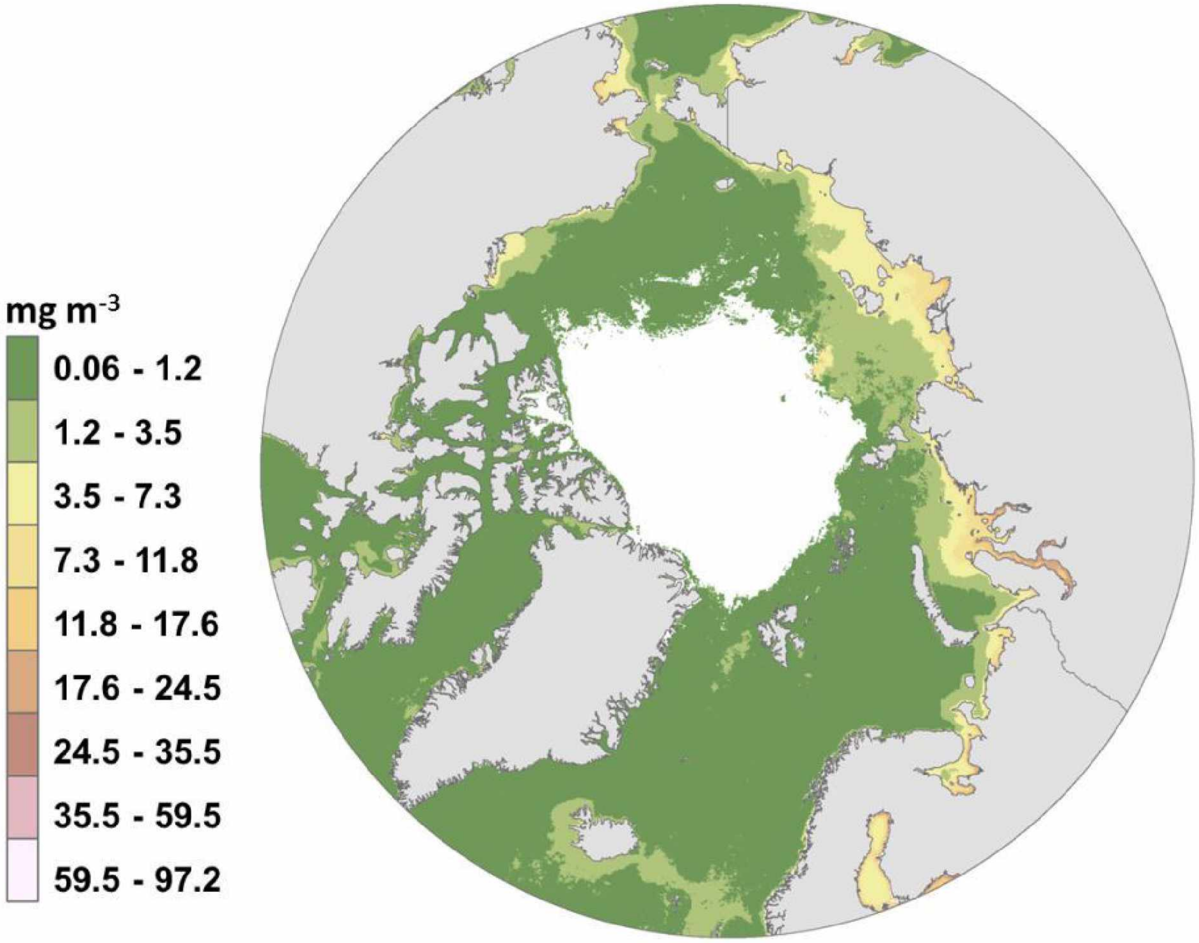


Figure C-23: Mean chlorophyll a concentration in mg m^{-3}

C-10 Distance to hydrology

The data for the distance to hydrology were derived from the Global Lakes and Wetlands Database (GLWD) (Lehner & Döll, 2004). The level 2 data consist of polygons of rivers, lakes, and reservoirs, and were applied to calculate the Euclidean distance to rivers with a cell size of 27411.505 x 27411.505 meters (Fig. C-24).

Source: http://www.worldwildlife.org/pages/global-lakes-and-wetlands-database
Version: Level 2 data
Filename: cl_euchydr3

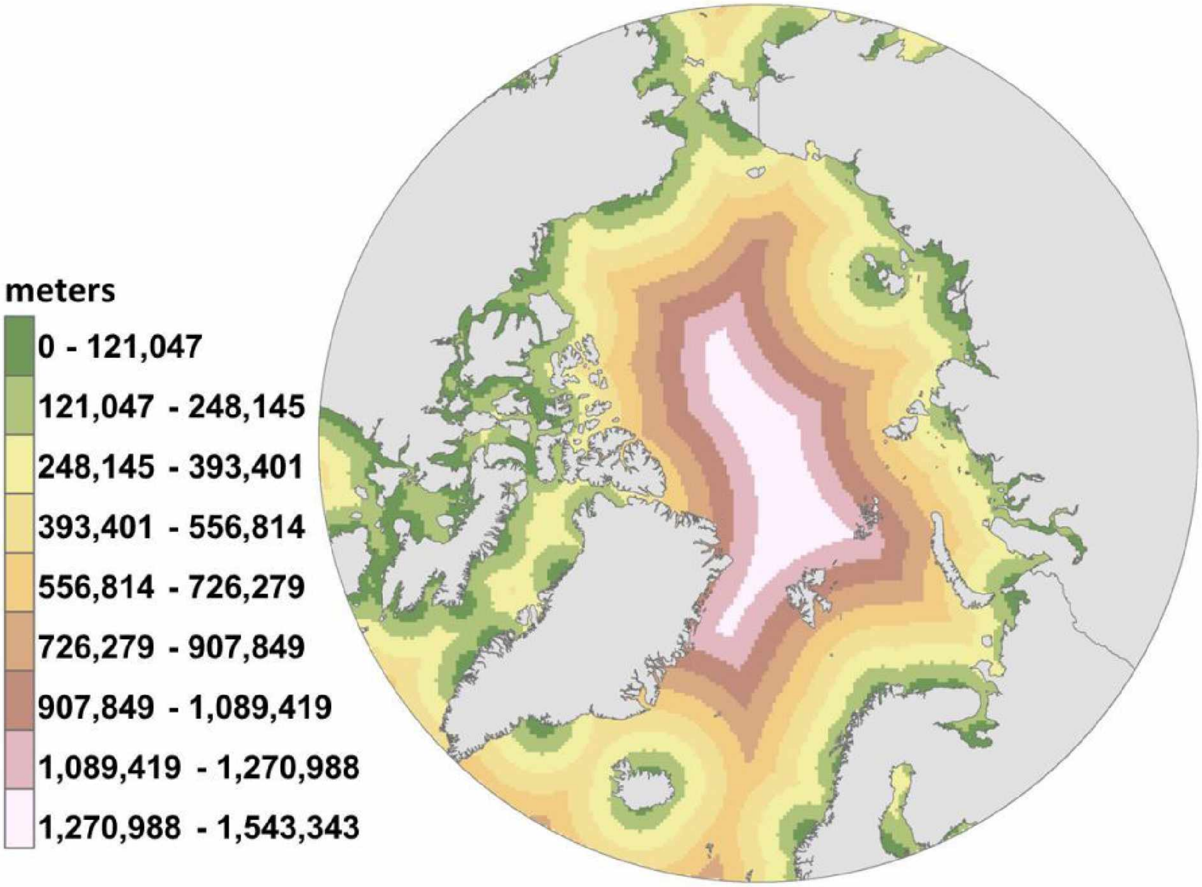


Figure C-24: Euclidean Distance to hydrology based on Global Lakes and Wetlands Database level 2 data

C-11 Euclidean Distance to shelf break

For the Euclidean distance to the shelf break, the shelf break was defined to be at a depth of 200 m. First, isolines were created from the GEBCO bathymetry layer for 0 m and for 200 m depth (3D Analyst >> Raster surface >> Contour list). The isolines were then used to create a polygon. The area between 0 and 200 m isolines of the newly created polygon was selected and the value set to 1 in the attribute table. After that, the areas that had the value 1 were selected and a new layer was created (Selection >> Create Layer from Selected Features).

The Euclidean distance to the 0 m isoline was calculated and clipped using the polygon between 0 m and 200 m that was created in the step before. The Euclidean distance values between 0 m and 200 m were then turned to negative numbers (Spatial Analyst >> Math >> Negate). A second polygon was created for the area between the 200 m isoline and the North Pole and the Euclidean Distance was extracted for this area as well. The Euclidean distance between 200 m and the North Pole was kept positive. Both extracted Euclidean distance rasters (0 m to 200 m, and 200 m to the North Pole) were then combined to one raster (Data Management Tools >> Raster >> Raster Dataset >> Mosaic to new Raster) (Fig. C-25).

Filename: EuD_po_neg_mo

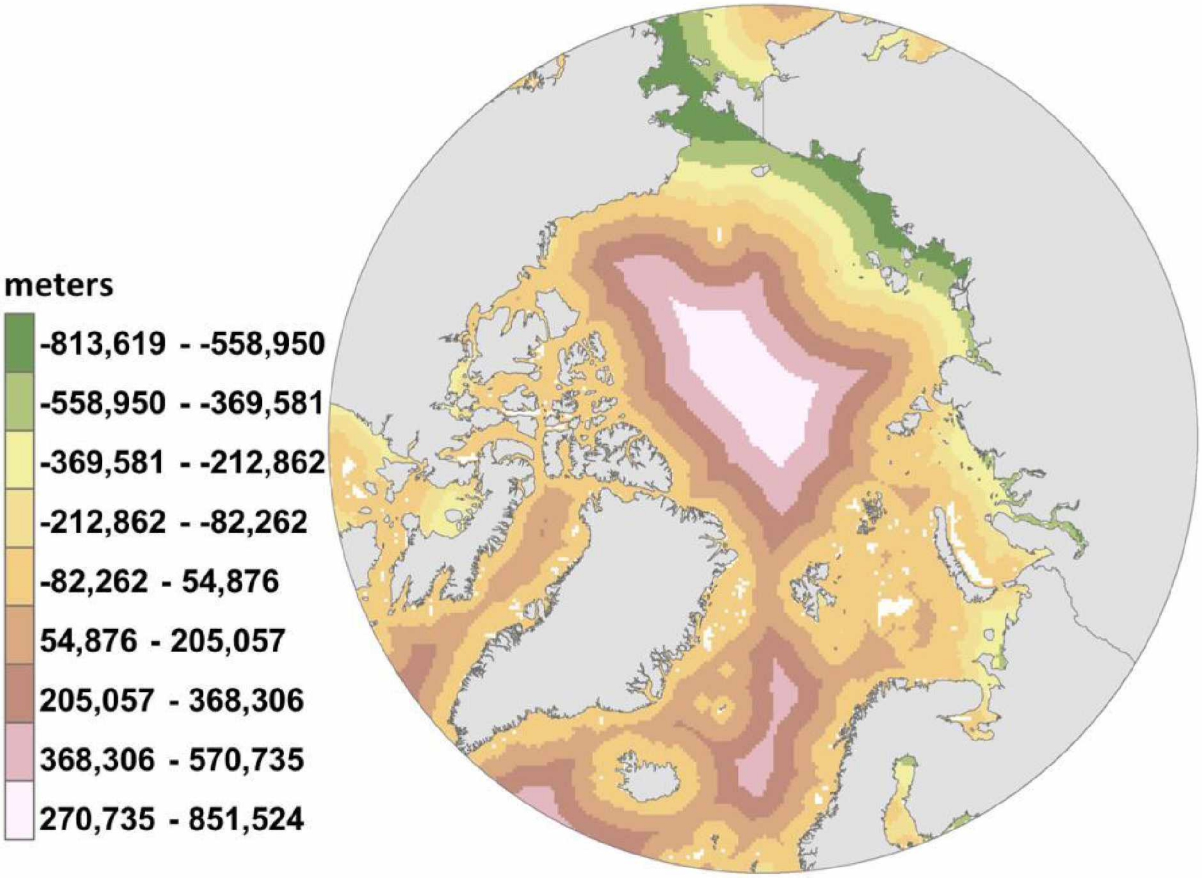


Figure C-25: Euclidean Distance to shelf break at 200 m

References

Lehner, B. and Döll, P. (2004) Development and validation of a global database of lakes, reservoirs and wetlands. *J Hydrol (Amst)*, **296**, 1-22.