SEVENTH FRAMEWORK PROGRAMME THEME 6 Environment

Collaborative project (Large-scale Integrating Project)

Project no: 246 933
Project Acronym: EURO-BASIN
Project title: European Basin-scale Analysis, Synthesis and Integration

Deliverable 3.8 Adaptive habitat models on past and present geographical distribution: final model

Due date of deliverable: 31.07.2014
Actual submission date: 31.07.2014
Organisation name of the lead contractor of this deliverable: AZTI-Tecnalia

Start date of project: 31.12.2010  Duration: 48 months
Project Coordinator: Michael St John, DTU Aqua

Project co-funded by the European Commission within the Seventh Framework Programme,
Theme 6 Environment
Dissemination Level

<table>
<thead>
<tr>
<th>PU</th>
<th>PP</th>
<th>RE</th>
<th>CO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Public</td>
<td>Restricted to other programme participants (including the Commission)</td>
<td>Restricted to a group specified by the consortium (including the Commission)</td>
<td>Confidential, only for members of the consortium (including the Commission)</td>
</tr>
</tbody>
</table>

X
Deliverable 3.8 Adaptive habitat models on past and present geographical distribution: final model, is a contribution to

Task 3.5: Development of habitat models

Responsible: AZTI-Tecnalia; Participants: SAHFOS, PML
Start month 12, end month 48

Executive Summary:

In the present deliverable, habitat models developed in D3.6 for four Calanus spp. in the North Atlantic Ocean, using CPR data provided by SAHFOS partner, were refined and extended to other 10 copepod species. We also evaluated the impacts of future climate change in community structure, diversity, distribution, and phenology of 14 copepod species. To this end, historical observations from the Continuous Plankton Recorder (CPR) in the 1970-2004 period and environmental data extracted from the NEMO-ERSEM model, provided by PML partner, have been used. Generalized Additive Models (GAMs) have been applied to relate the species occurrence with environmental variables. The North Atlantic climate regime shift served us to perform a temporal cross-validation of the habitat models: the habitat model built in the cold period (1970-1986) has been validated in the warm period (1987-2004). Thus, selected habitat were projected to future (2080-2099) environmental conditions using coupled HAMOCC (HAMburg Ocean Carbon Cycle) and MPIOM (Max Planck Institute Ocean Model) models under A1B climate scenario, and compared to present (2001-2020) conditions.

The habitat suitability modelling enabled us to identify three key environmental variables (sea surface temperature and salinity, and depth) that determine the present distribution of copepods in the North Atlantic. GAM has proved to be a useful and accurate model to quantify the ecological niche of Calanus spp. This conclusion is based on 1) the accuracy values of random (69-85%) and temporal (74-85%) validation, 2) its flexibility to incorporate seasonal variability, and 3) its performance comparing both accuracy values and spatial distribution maps with MaxEnt and Mahalanobis distance algorithms. The temporal cross-validation showed that model accuracy is relatively good (74-85%) for the models built in cold regime and extrapolated and validated in the warm period. Therefore, we were confident to use the GAM-based habitat models generated for Calanus spp. in the future climate simulation.

Our projections by the end of the century indicated that copepod community is expected to respond substantially to climate change: a poleward latitudinal shift of 8.7 km/decade on average for the overall community with an important species range variation (-15 to 18 km/decade), the species seasonal peak is expected to occur 12-13 days earlier for C. finmarchicus and C. hyperboreus, and important changes in community structure are also expected (high species turnover between 43 and 79% at the south of the Oceanic Polar Front). These changes might lead to alterations of the future North Atlantic pelagic ecosystem.
Relevance to the project & potential policy impact:

This deliverable have used input data from T3.1 consisting of Continuous Plankton Recorder (CPR) survey data of the distribution of key zooplankton groups (copepods) in the North Atlantic basin, which have been prepared by SAHFOS and delivered to AZTI, and environmental data extracted from the NEMO-ERSEM model (WP6), developed by PML and delivered to AZTI. The present deliverable used also the knowledge gained in habitat modelling developed in D3.6 to refine and validate the models, as well as to extend the approach to the most abundant zooplankton species (14 species accounting for the 49.3% of the total occurrences sampled in the community, hence, well representing the overall community in terms of abundance).

The main relevance of this work is the synergism with WP6 (Task 6.3) in coupling the habitat models with hydrographic-biogeochemical models in order to simulate species distribution under future climate change scenarios. Advances in habitat and climate modelling allow us to reduce uncertainties of climate change impacts on species distribution. Thus, we have assessed the potential reduction or increase in the habitat of zooplankton species, which can affect fish populations. A solid base of works indicates that warming can modify the distribution of marine organisms, which in turn, it can be propagated through the upper trophic levels. In a context of rapid alteration of marine ecosystems throughout the world (Pauly et al., 1998), future projections of ocean productivity, based on habitat species distribution, are needed for a detailed assessment of ocean health and benefits and for achieving or maintaining the good environmental status of the North Atlantic (see for instance the environmental status defined by the Marine Strategy Framework Directive, MSFD, European Commission, 2008).
Report:

Deliverable 3.8 Adaptive habitat models on past and present geographical distribution: final model

Contributors: Guillem Chust¹, Ernesto Villarino¹, Priscilla Licandro², Leire Ibaibarriaga¹, Momme Butenschöhn³

¹ AZTI-Tecnalia (Spain)  
² SAHFOS (United Kingdom)  
³ PML (United Kingdom)

Collaborators: Xabier Irigoien⁴

⁴ KAUST (Saudi Arabia)

Introduction

In the present deliverable, habitat models developed in D3.6 for four Calanus spp. (C. finmarchicus, C. glacialis, C. helgolandicus, C. hyperboreus) using CPR data were refined and extended to other 10 copepod species based upon four main aspects: 1) addition of new environmental variables (mixed layer depth, oxygen, pH, and phytoplankton biomass); 2) model selection comparing GAM with other habitat model algorithms (Mahalanobis distance and MaxEnt); 3) validated using randomly independent data sets; 4) validated comparing cold (1970-1986) with warm (1987-2004) climate regimes in order to assess their potential to be extrapolated to future climate; 5) we built habitat models (based upon the selected ones) to 10 other copepod species (Candacia armata, Centropages typicus, Centropages hamatus, Metridia lucens, Paraechaeuta modesta, Paraechaeuta hebes, Pleuromamma borealis, Pleuromamma robusta, Pseudocalanus elongatus, Temora longicornis) to include the main species of the community. The selected habitat models developed and validated here have been projected to future (2080-2099) environmental conditions using coupled HAMOCC (HAMburg Ocean Carbon Cycle) and MPIOM (Max Planck Institute Ocean Model) models under A1B climate scenario, and compared to present (2001-2020) conditions. We evaluated the impacts of future climate change in community structure, diversity, distribution, and phenology of 14 copepod species in the North Atlantic Ocean.

State-of-the-art and problematic

Plankton communities can quickly respond to climatic variability (e.g. Beaugrand et al., 2002). Impacts of global warming affect the whole pelagic ecosystem from plankton to higher trophic levels (Beaugrand & Kirby, 2010; Beaugrand et al., 2012; Richardson & Schoeman, 2004). Such impacts can result in poleward movements in species distribution (Beaugrand et al., 2009; Johns et al., 2001; Perry et al., 2005; Parmesan, 2006, Chust et al., 2014a), shifts in phenology (Edwards & Richardson 2004; Moore et al., 2011) or changes in abundance and community structure (Molinero et al., 2008; Kirby & Beaugrand 2009; Chust et al., 2014b). And these species responses may lead to local extinction and invasions, resulting in changes in the pattern of marine species richness and trophic mismatches (Cheung et al., 2009). Therefore, assessing how these biogeographic processes will change in the future is a key prerequisite to
anticipate consequences of climate change on marine ecosystems.

Sea temperature is one the most important physical variable structuring marine ecosystems. There is overwhelming evidence that the composition, abundance, and phenology of plankton communities are closely linked to water temperature (Richardson, 2008). Throughout the North Atlantic Ocean, a general increase in temperature has been observed in the past century (Beaugrand, 2009) and future ocean temperatures have been forecasted to increase by coupled atmosphere–ocean general circulation models (AOGCMs). In particular, the North Atlantic has warmed faster than all other ocean basins and climate change scenarios project sea surface temperature isotherms to shift up to 600 km northwards by the end of the 21st century (Lee et al., 2011).

Habitat suitability (species distribution) models (Guisan & Zimmermann, 2000; Elith & Leathwick, 2009) have been widely used to project how species ranges might change in the future. These models aim to define the species ecological niches by relating the occurrence of species to the environmental variables (e.g., temperature, depth, and phytoplankton) in the same area. Habitat suitability models rely on the environmental niche concept of Hutchinson (1957), where a multi-dimensional hypervolume is defined by the combination of multiple environmental conditions that requires a species population to survive and reproduce. Then, using projections from the Intergovernmental Panel of Climate Change (IPCC), we can analyze how environmental variables will change over the future and project the corresponding future species ranges (Guisan & Thuiller, 2005).

In the last decade, several studies using species distribution models and Continuous Plankton Recorder (CPR) data with future climate change scenarios have been published. For example, Helaouët & Beaugrand (2009) forecasted a poleward movement of *C. finmarchicus* of 1 degree of latitude by the end of the century, Beaugrand et al. (2008) analyzed the reasons behind the climate driven ecosystem future shifts of cod, zooplankton and phytoplankton; Reygondeau & Beaugrand (2011) used the Non Parametric Probabilistic Ecological Niche Model (NPPENM) to project *C. finmarchicus* distribution through the next century, and Beaugrand et al., (2013) to project how climate-induced changes in temperatures will alter marine zooplankton both locally and globally. Most of them rely on the Non Parametric Probabilistic Ecological Niche Model, which is based on the Mahalanobis distance algorithm. A recent work from Chust et al. (2014a) shows that Generalized Additive Models perform well to detect latitudinal shifts of species and identifying its causes.

So far most of the bioclimatological research is concentrated on a single species (Reygondeau & Beaugrand 2011; Helaouët & Beaugrand, 2011; Beaugrand et al., 2007; Beaugrand et al., 2013; Bonnet et al., 2005; Helaouët & Beaugrand, 2007) and there are very few works at community level (Beaugrand et al., 2000; 2002). Yet, some of them analyzed the historic plankton biogeographical shifts in the North Atlantic Ocean (Reygondeau & Beaugrand 2011). However, little is known about the future spatial distribution of copepod biodiversity, seasonal changes and latitudinal shift in the North Atlantic Ocean, despite their importance in marine food webs.
Objectives

The objective of this task is to develop and validate habitat models in key zooplankton species within the North Atlantic (Figure 1), as well as to extrapolate them to future climate scenarios. In particular, we have generated models based upon the prominent climate drivers for the most abundant 14 copepod species to represent the overall community. First, model selection and validation has been evaluated for four species using the entire times series from 1970 to 2004. In particular, we compared GAM with other habitat model algorithms (Mahalanobis distance and MaxEnt), and validated the model using randomly independent data sets and comparing cold (1970-1986) with warm (1987-2004) climate regimes in order to assess their potential to be extrapolated to future climate. Secondly, we built habitat models (based upon the selected ones) to 10 other copepod species. Thirdly, the selected habitat models have been projected to future (2080-2099) environmental conditions using coupled hydrographic-biogeochemical models under A1B climate scenario, and compared to present (2001-2020) conditions. Fourthly, we evaluated the impacts of future climate change in community structure, diversity, distribution, and phenology of 14 copepod species in the North Atlantic Ocean.

Study Area

The North Atlantic Ocean (domain of the grid: 35º to 65º N and 75º W to 2º E, Figure 1).

Material and Methods

Environmental data

A set of seven environmental variables were used to build the N-dimensional ecological niches of copepod species and to predict the probability of occurrence of them over the North Atlantic Ocean: sea surface temperature (SST), sea surface salinity (SSS), Mixed Layer Depth (MLD), oxygen, pH, sea surface phytoplankton biomass, and bathymetry. Sea surface phytoplankton
biomass, oxygen, pH, sea surface temperature (SST), and sea surface salinity data were extracted from a 1960-2004 hindcast of an implementation of the NEMO-ERSEM model forced with atmospheric reanalysis data from the Drakkar Forcing Set 4 (DFS4) composite of NCEP and ECMWF fields (European Centre for Medium-Range Weather Forecasts). Mixed Layer Depth (MLD) data was obtained from the Center for Marine and Atmospheric Sciences (ZMAW, Hamburg). The mixed-layer depth (MLD) is an indicator of water column stability. It is obtained from vertical profiles of temperature and salinity (de Boyer Montégut et al., 2004), using the classical density criterion of 0.125 (Levitus, 1982). Bathymetry was extracted from ETOP01 global model (NOAA). Data are organized in 1° longitude and 1° latitude grid resolution available for every month of the period 1970–2004.

SST and salinity (salinity especially in coastal environments) are essential factors because of its recognized influence on spatial distribution of *Calanus* spp. (Mauchline, 1998; Helaouët & Beaugrand, 2009; Helaouët et al., 2011; Reygondeau & Beaugrand, 2011). Bathymetry was selected because it has been suggested that it influences the distribution of some copepod species in regions such as the southern North Sea (Reygondeau & Beaugrand, 2011). Phytoplankton is an important food source for *Calanus* spp. that dominates zooplankton biomass in the North Atlantic (Melle et al., 2014). MLD is an important parameter for phytoplankton production and controls the spatial distribution of many plankton species (Longhurst, 2007). Oceanic pH influences calcifying organisms such as coccolithophorids, foraminifers, corals and pteropods (Orr et al., 2005; Kroeker et al., 2010).

**Biological data**

Data on the abundance (mean density (ind./m$^3$)) of four species (*C. finmarchicus*, *C. glacialis*, *C. helgolandicus*, *C. hyperboreus*) was obtained from the CPR database. The CPR survey is an upper layer plankton monitoring programme that has regularly collected samples, at monthly intervals, in the North Atlantic and adjacent seas since 1946 (Warner & Hays, 1994). In the supplementary material (Fig.S1) it is shown the number of occurrences from CPR survey for 1970-2004 period of the four copepods.

These calanoids are key species in subarctic (*C. finmarchicus*) and temperate shelf-edge (*C. helgolandicus*) regions of the North Atlantic Ocean (Bonnet et al., 2005; Speirs et al., 2005). *C. helgolandicus* is considered to be a pseudo-oceanic species (i.e., a species that can be found in oceanic and neritic waters, but it is mostly abundant above the shelf edge (Beaugrand et al., 2002). *C. glacialis* and *C. hyperboreus* are arctic species, while *C. finmarchicus* is a subarctic species that overlaps in size range with *C. helgolandicus*. *C. hyperboreus* is the largest in size among them.

In order to have a better representation of copepod community at basin scale, data on other 10 copepod species (ind/m$^3$) (*Candacia armata*, *Centropages typicus*, *Centropages hamatus*, *Metridia lucens*, *Paraechaeuta modesta*, *Paraechaeuta hebes*, *Pleuromamma borealis*, *Pleuromamma robusta*, *Pseudocalanus elongatus*, *Temora longicornis*) for the same area were obtained from the NMFS-COPEPOD global plankton data base (http://www.st.nmfs.noaa.gov/copepod/data/sahfosatl/index.html) between 1995 and 1999, which is also based on CPR survey. Those species were selected as they were the most
abundant taxon identified at species level (with more than 100 occurrence records in the data set). We selected 14 species out of 66 copepod sampled species that represents 49.3% of the total occurrences sampled in the community, hence, well representing the overall community in terms of abundance. All CPR data used in the present study were analysed monthly, gridded at 1 by 1º spatial resolution, and within 35º to 65ºN and -75ºW to 2ºE.

**Habitat modelling**

Prior to model building, we tested for collinearity between explanatory variables by calculating variance inflation factors (VIF) with the AED package in R (Zuur et al., 2009). We excluded any variable that had a VIF>3, and then recalculated VIF for the remaining variables. We iterated this process until all variables had a VIF<3. The variable most often thrown out was oxygen, which highly correlated with temperature. Therefore we excluded it from the analysis.

Species distribution models assume that observations represent species at equilibrium with its environment. Here, Generalized Additive Models (GAMs) (Hastie & Tibshirani, 1990) have been used to model occurrences for each of the 4 *Calanus* spp. as a function of environmental factors (SST, salinity, MLD, pH, and bathymetry) and potential food resource (phytoplankton biomass). The strength of GAMs is due to its capacity to deal with highly non-linear relationships between the response and the set of explanatory variables, allowing asymmetrical unimodal distributions since interaction between species and extreme environmental gradients may cause skewed responses (Oksanen & Minchin, 2002). GAMs also enable us to model the seasonal response of the species. A GAM using the binomial error distribution and logit function of the mgcv package in R (Wood, 2006) was used to relate copepods presence-absence data and the explanatory environmental variables, following Chust et al., (2014a). To prevent overfitting, we first restricted the degrees of smoothness to ecologically interpretable responses according to niche theory; second we analysed the response of species occurrence to each environmental predictors; and third we used cross-validation methods to evaluate the reliability of the models (see section below). The CPR dataset used here includes 112161 samples across the spatial domain and irregularly distributed at year and monthly basis.

We built and compared different GAMs for each species to find the optimal set of explanatory variables. Variable importance was assessed by adding and removing terms and noting the change in deviance or gain (>1%) in a forward stepwise procedure, and by removing variables that were statistically not significant. For environmental variables, the degree of smoothness of model terms was restricted from 3 to 5 in order to assume a unimodal, ecologically meaningful niche model *sensu* Hutchinson (1957), but allowing asymmetry.

After characterizing the ecological niche of each species, inferred from observed distribution and corresponding values of SST, depth, salinity, phytoplankton biomass and pH, the environmental space was projected into geographical space and the probability of occurrence was calculated.

We have generated models based upon the prominent climate drivers for the most abundant 14 copepod species to represent the overall community. First, model selection and validation has been evaluated for four species (*Calanus finmarchicus*, *Calanus glacialis*, *Calanus*.
**helgolandicus, Calanus hyperboreus** using the entire times series from 1970 to 2004. In particular, we compared GAM with other habitat model algorithms (Mahalanobis distance and MaxEnt), and validated the model using randomly independent data sets and comparing cold (1970-1986) with warm (1987-2004) climate regimes in order to assess their capacity to be extrapolated to future climate. On a second step, we built habitat models (based upon the selected ones) to 10 other copepod species to (1) evaluate the impacts of future climate change in community structure at 2080-2099 period compared to present conditions (2001-2020), (2) quantify the poleward shift of species distribution, and (3) analyse phenological changes of the species in the North Atlantic Ocean at community level, with model outputs corresponding to SRES A1B global warming scenario (IPCC AR4, 2007).

**GAM vs Mahalanobis distance and MaxEnt**

GAM has been also compared with other two habitat models (MaxEnt and Mahalanobis distance algorithm) in order to assess its performance. Both MaxEnt and Mahalanobis are ecological niche models using presence-only species records, although can use absences to model validation. They are implemented in the dismo R package (Hijmans et al., 2013) which is specially designed to model species distribution that do not migrate or shift during seasonal cycle, since it uses static environmental layers. Contrary to Mahalanobis distance and MaxEnt habitat model techniques, GAMs presents the advantage model seasonal response of the species and results in a more “dynamic” habitat modelling technique. Hence, only for the purpose of comparing performances of GAM with MaxEnt and Mahalanobis distance algorithms, we reduced the dataset into a unique spatial layer by accumulating occurrences of all years and months. The Mahalanobis Distance (MD) algorithm technique for a given point expresses the distance between this point and the species optimum in the ecological space. (Farber & Kadmon 2003). MaxEnt uses the principle of maximum entropy to estimate a set of functions that relate environmental variables and habitat suitability in order to approximate the species’ niche and potential geographic distribution (Phillips et al., 2006). MaxEnt model minimizes the relative entropy between two probability densities (presence data and the landscape data) defined in a covariate space (Elith et al., 2011). Although MaxEnt has been widely used in terrestrial species (e.g. Monterroso et al., 2009; Graham & Hijmans, 2006; Yates et al., 2010; Young et al., 2009), applications in pelagic species are still scarce.

**Model validation**

Models were validated using independent data sets for model building and model validation (Burnham and Anderson, 2002). We validate the models in two ways: (1) k-fold random resampling, and (2) temporal cross-validation. In the first procedure, the data is first partitioned into k equally sized segments or folds. Subsequently, k iterations of training and validation are performed such that within each iteration a different fold of the data is held-out for validation while the remaining k-1 folds are used for model fitting (Hijmans, 2012). We used k=5, hence, 80% of the CPR observations were used for model building, and the other 20% (i.e. independent) for model validation in an iterative procedure that was repeated 5 times. Hold-out validation avoids the overlap between training data and test data, yielding a more accurate estimate for the generalization performance of the algorithm. The comparison between the
accuracy of the model (all observations) and that of cross validated also permits the detection of model overfitting, which reduce the usefulness of such models for extrapolation.

The North Atlantic regime shift (Reid et al., 2001; Beaugrand, 2004; deYoung et al., 2004) has been taken into account to perform a temporal cross-validation of the models (i.e. second procedure of model validation). A wide range of studies have investigated the North Atlantic and North Sea climate decadal fluctuations that affect phytoplankton (Edwards et al., 2001; Reid et al., 2001; Beaugrand & Reid, 2003), zooplankton (Beaugrand et al., 2002; Beaugrand, 2003; Beaugrand, 2004) and fish populations (Alheit & Hagen, 1997; Reid et al., 2001). To this end, we performed a Wilcoxon rank sum test (Wilcoxon, 1945) between a cold period from 1970 to 1986 (mean SST: 11.64 ± 0.12°C) and a warm period from 1987 to 2004 (mean SST: 12.10 ± 0.26°C) and defined in our time series two different climatic regimes (p-value < 0.0001, Figure 2). Subsequently, we built the models and compared the four Calanus spp. between cold (1970-1986) and warm (1987-2004) periods. We tested the habitat model predictive capacity built using cold period into warm period and vice versa, using the AUC and confusion matrix accuracy assessment indices (see section below). This approach enabled us not only to explore the model behaviour at different climates but also to see to which degree of reliability we can project the model to a future warmer climate conditions. The temporal cross-validation was undertaken only to the four Calanus spp. since the NMFS-COPEPOD time series (1995-1999) including the other set of ten copepod species is too short.

Figure 2. Mean SST time series; discontinuous line indicates cold period (1970-1896) and warm period (1987-2004) considered.

Model evaluation

We assessed the predictive performance of the overall model and the held-out folds using the area under receiver operating characteristic curve (AUC) (Hanley & McNeil, 1982; Raes & ter Steege, 2007), a measure of the ability of the predictions to discriminate presence from absence, and accuracy indices derived from confusion matrix (VanDerWal et al., 2011). To this end, the species presence modelled probability was converted to either presence or absence using probability thresholds following two criteria: sensitivity (true predicted presences) = specificity (true predicted absences), and maximization of sensitivity plus specificity, as reported in Jimenez-Valverde & Lobo (2007). Thus, the cases above this threshold are assigned to presences, and below to absences. Given the threshold value, a confusion matrix was calculated yielding outputs of correctly identified records of presence and absence to have an
overall accuracy estimate of model performance. Overall accuracy ranges from 0 to 100% and AUC values from 0.5 (random sorting) and 1 (perfect discrimination). Accuracy is a good indicator of model performance since it is the proportion of true results, either true positive or true negative, in a population.

**Climatic scenario for the 21st century**

In order to assess the copepod response to climate change, selected habitat models were projected to future conditions and thresholds were applied to the resulting probability maps. We used modelled environmental predictors (SST, SSS, and phytoplankton) from the DKRZ-CERA at A1B scenario [http://cera-www.dkrz.de](http://cera-www.dkrz.de) for the 2001-2099 period. More in detail, we used the Hamburg Ocean Carbon Cycle (HAMOCC) model for the phytoplankton, and the Max Planck Institute’s Ocean General Circulation Model (MPIOM) (Jungclaus et al., 2013) for the physical set up. HAMOCC, embedded into MPIOM, simulates the oceanic cycles of carbon and other biogeochemical elements (Ilyna et al., 2013). Technical details of the ocean model MPI-OM can be found in Marsland et al., (2003).

**Assessing impacts of climate change to copepods**

The impacts of climate change to copepods were assessed by estimating latitudinal shifts of each species, phenological changes and spatial patterns of biodiversity indices.

The latitudinal shift (km) of the species was calculated by comparing the geographic centres of gravity of the species suitable area for present (2001-2020) and future scenario (2080-2099). The centre of gravity was defined as the mean geographic location of a population (Woillez et al., 2009). Gravity centers of habitat models showing well separated east to west population patches (C. armata, C. hamatus and P. hebes), were calculated separately and then we averaged them. We assumed unlimited copepod dispersal to estimate the extent of gain or loss of suitable space from current to future modelled conditions.

We computed the changes in the seasonal cycle or phenology of *Calanus* spp. by analysing the difference in terms of days on the annual maxima of the copepods probabilities of occurrence on both present and future conditions. The timing of the peak was determined as the date where the modelled species occurrences reached the annual maximum. Monthly mean species occurrences were used to build a GAM fitted function (with a Gaussian link and cyclic cubic regression spline) in order to infer the predicted seasonal peaks, enabling us to quantify the phenological shifts in days.

Changes in local biodiversity were assessed in terms of species turnover, colonization and extinction. In particular, we mapped four biodiversity components of change: a) stability, i.e. the number of species that were present or absent at both present and future scenarios at each pixel; b) extinction, i.e. the number of species that were present at present and were absent in the future; c) colonization, i.e. the number of species that were absent at present and present in the future; and d) turnover, i.e. the number of species that suffer either colonization or extinction.
Species assemblages were categorized following Beaugrand et al., 2009 in two main groups: (1) the “cold water species assemblage”, including the cold-temperate mixed water (Pleurommama robusta), subarctic (Calanus finmarchicus) and arctic (Calanus hyperboreus and Calanus glacialis) species assemblages and (2) the “warm water species assemblage”, including the warm-temperate oceanic and pseudo-oceanic (Pleurommama borealis, Paraeuchaeta modesta and Metridia lucens and Paraeuchaeta hebles), the temperate pseudo-oceanic (Centropages typicus, Candacia armata and Calanus helgolandicus) and continental shelf (Pseudocalanus elongatus, Temora longicornis and Centropages typicus). This simplified way of proceeding enabled us to understand which set of species were more affected by environmental change.

Results

Future environmental changes

MPIOM SST model projections have forecasted an average increase of 1.54 °C ± 0.35 °C (Wilcoxon rank sum test, p-value < 0.0001) in the study area of the North Atlantic by 2100 (Figure 3). Our spatial examination of changes in SST revealed regional differences. For instance, SST increased in 4-6 °C in areas of the Gulf Stream extension and the Newfoundland Continental Shelf, south of the Oceanic Polar Front (i.e. 60°W-45°W and 43°N-48°N). On the other hand, in some areas of the subarctic region south of Labrador current in the North Atlantic gyre (i.e. 40°W-30°W and 55°N-60°N), SST is expected to decrease between 1 and 0 °C degrees.

The HAMOCC biogeochemical model projects a general phytoplankton biomass decrease by the end of century in the North Atlantic. Results showed a clear east to west asymmetry on phytoplankton biomass changes, with strong negative anomalies (-30 to -20 mgC/m³) along the east of the Oceanic Polar Front, from the subarctic region south of Iceland down to the Bay of Biscay and the Southern European Shelf edge (i.e. 25°W -5°W and 38°N-60°N). In turn, a slight increase in phytoplankton biomass (0-10 mgC/m³) is projected along the Flemish Cup area and extending thought the Oceanic Polar Front (i.e. 50°W-40°W and 43°N-50°N) (Figure 3).
Figure 3. SST (A) and phytoplankton (B) difference models for present (2001-2020) and future (2080-2099) periods.

GAM habitat models

We evaluated the response of the *Calanus* spp. to each explanatory variable individually using GAMs (Table 1). SST was the most important environmental driver in the *Calanus* ssp. environmental space. Oceanic pH and MLD were explaining less deviance of species occurrence, although pH was considered in the model selection as it accounted for more than 1% of deviance explained for all models.

Subsequently, habitat suitability models were constructed for the four *Calanus* spp (Figure 4). All the environmental variables, i.e. temperature, salinity, depth and pH and phytoplankton biomass, were included in all models except for *C. glacialis* (without pH) and surface phytoplankton biomass was only kept in *C. hyperboreus* and *C. finmarchicus* models. Therefore, surface phytoplankton biomass seems to be a controlling factor in the probability of occurrence and distribution of these two species, and not on *C. glacialis* and *C. helgolandicus*, according to models.
Table 1. Explained deviance of Calanus spp. occurrence according to each environmental factor.

<table>
<thead>
<tr>
<th></th>
<th>C. finmarchicus</th>
<th>C. helgolandicus</th>
<th>C. glacialis</th>
<th>C. hyperboreus</th>
</tr>
</thead>
<tbody>
<tr>
<td>SST</td>
<td>14.2</td>
<td>11.3</td>
<td>29.6</td>
<td>21.5</td>
</tr>
<tr>
<td>SSS</td>
<td>15.1</td>
<td>5.4</td>
<td>25.5</td>
<td>13.0</td>
</tr>
<tr>
<td>O₂</td>
<td>18.9</td>
<td>3.4</td>
<td>25.9</td>
<td>30.8</td>
</tr>
<tr>
<td>pH</td>
<td>4.4</td>
<td>1.8</td>
<td>1.9</td>
<td>6.6</td>
</tr>
<tr>
<td>Phytoplankton biomass</td>
<td>6.2</td>
<td>0.4</td>
<td>4.2</td>
<td>16.5</td>
</tr>
<tr>
<td>Chl-a</td>
<td>6.1</td>
<td>0.3</td>
<td>3.2</td>
<td>12.0</td>
</tr>
<tr>
<td>MLD</td>
<td>0.6</td>
<td>2.8</td>
<td>0.7</td>
<td>0.3</td>
</tr>
<tr>
<td>Depth</td>
<td>5.5</td>
<td>14.1</td>
<td>0.1</td>
<td>4.8</td>
</tr>
</tbody>
</table>

Figure 4. Occurrence models of Calanus spp. in 1970-2004 period. Map key: Orange is presence and grey is absence. A) C. finmarchicus, B) C. helgolandicus, C) C. glacialis, D) C. hyperboreus.

The random cross-validation of models is shown in Table 2. The habitat models in the 4 Calanus spp. showed a slight drop in the accuracy measure if we compare the model with all observations (74-85%) vs the k-fold cross-validation (69-85%); this is owing to a slight signal of model overfitting. Here, C. helgolandicus showed low overall deviance explained (25.4%) in the habitat suitability models, whilst the other species deviance explained was higher: C. finmarchicus (46.9%), C. glacialis (34.1%) and C. hyperboreus (42.3%).

The temporal cross-validation enabled us to assess the model reliability to be extrapolated to different climates. Results have shown that model accuracy is relatively good (75-84%) for the models built in cold regime and extrapolated and validated in the warm period (Table 3). Therefore, species models can be used to be projected in future climate simulations with relative confidence.
Table 2. Evaluation of GAM models (yearly accumulated) with k-fold cross-validation. Variables entered: Sea Surface Temperature, Salinity, Bathymetry, Surface Phytoplankton Biomass (Pc), and pH. EDF: Estimated degrees of freedom. Threshold for conversion of probability of species presence to either presence or absence in model validation: 0.08 (C. hyperboreus), 0.11 (C. glacialis), 0.48 (C. helgolandicus), 0.63 (C. finmarchicus).

| Species          | Variables selected | EDF   | p-value     | Overall deviance explained (%) | AUC (model with all observations/ Mean k-fold cross-validation) | Accuracy (model with all observations/ Mean k-fold cross-validation) |
|------------------|--------------------|-------|-------------|-------------------------------|---------------------------------------------------------------|-----------------------------------------------------------------
| C. hyperboreus   | SST                | 1.99  | <2e-16      | 42.3                          | 0.685/0.698                                                   | 85.34/69.84                                                       |
|                  | SSS                | 1.00  |             |                               |                                                               |                                                                  |
|                  | Depth              | 2.17  |             |                               |                                                               |                                                                  |
|                  | pH                 | 1.96  |             |                               |                                                               |                                                                  |
|                  | Pc                 |       |             |                               |                                                               |                                                                  |
| C. glacialis     | SST                | 1.95  | <2e-16      | 34.1                          | 0.816/0.642                                                   | 81.90/71.25                                                       |
|                  | SSS                | 2.88  |             |                               |                                                               |                                                                  |
|                  | Depth              | 1.00  |             |                               |                                                               |                                                                  |
| C. helgolandicus | SST                | 1.99  | <2e-16      | 25.4                          | 0.749/0.754                                                   | 74.94/75.40                                                       |
|                  | SSS                | 2.99  |             |                               |                                                               |                                                                  |
|                  | Depth              | 2.85  |             |                               |                                                               |                                                                  |
|                  | pH                 | 1.97  |             |                               |                                                               |                                                                  |
| C. finmarchicus  | SST                | 1.97  | <2e-16      | 46.9                          | 0.852/0.851                                                   | 85.19/85.13                                                       |
|                  | SSS                | 2.74  |             |                               |                                                               |                                                                  |
|                  | Depth              | 2.88  |             |                               |                                                               |                                                                  |
|                  | pH                 | 1.83  |             |                               |                                                               |                                                                  |


<table>
<thead>
<tr>
<th>Species</th>
<th>Variables selected</th>
<th>EDF</th>
<th>p-value</th>
<th>Overall deviance explained (%)</th>
<th>AUC (model with all observations/ Mean k-fold cross-validation)</th>
<th>Accuracy (model with all observations/ Mean k-fold cross validation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. hyperboreus</td>
<td>SST</td>
<td>1.99</td>
<td>&lt;2e-16</td>
<td>48.1</td>
<td>0.845/0.807</td>
<td>85.34/80.80</td>
</tr>
<tr>
<td></td>
<td>SSS</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>2.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pH</td>
<td>1.98</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. glacialis</td>
<td>SST</td>
<td>1.73</td>
<td>&lt;2e-16</td>
<td>30.7</td>
<td>0.816/0.691</td>
<td>81.90/74.33</td>
</tr>
<tr>
<td></td>
<td>SSS</td>
<td>2.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pH</td>
<td>0.0143</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. helgolandicus</td>
<td>SST</td>
<td>1.99</td>
<td>&lt;2e-16</td>
<td>27.3</td>
<td>0.749/0.745</td>
<td>74.94/74.83</td>
</tr>
<tr>
<td></td>
<td>SSS</td>
<td>2.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>2.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pH</td>
<td>1.98</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. finmarchicus</td>
<td>SST</td>
<td>1.95</td>
<td>&lt;2e-16</td>
<td>45.2</td>
<td>0.852/0.857</td>
<td>85.19/85.73</td>
</tr>
<tr>
<td></td>
<td>SSS</td>
<td>2.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>2.80</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pH</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pc</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The GAM habitat models for the non Calanus spp. set of copepod species were generated using the NMFS-COPEPOD database in the 1995-1999 period. Models results (Table 4) showed moderate accuracy (64-74%) and moderate deviance explained for Centropages typicus (25.4%), Centropages hamatus (28.7%), Paraeuchaeta modesta (27.5%), Paraeuchaeta hebes (35.4%) and Temora longicornis (24.9%). On the other hand, in the
remaining species the deviance explained was lower (17.5-7.7%) as well as the accuracy values. The difference in the accuracy values of the model using all the observations (76-63%) and that cross-validated (53-57%), indicate a slight overfitting in these latter models (Table 4).

**Table 4.** GAM models 1995-1999. All observation based model vs k-fold cross validated models. Threshold for conversion of probability of species presence to either presence or absence in model validation: 0.09 (Candacia armata), 0.24 (Centropages typicus), 0.07 (Centropages hamatus), 0.35 (Metridia lucens), 0.14 (Paraechaeuta modesta), 0.1 (Paraechaeuta hebes), 0.1 (Pleuromamma borealis), 0.08 (Pleuromamma robusta), 0.2 (Pseudocalanus elongatus), 0.26 (Temora longicornis). E.D.: explained deviance.

<table>
<thead>
<tr>
<th>Species</th>
<th>Variables selected</th>
<th>EDF</th>
<th>p-value</th>
<th>Overall E.D. (%)</th>
<th>AUC (model with all observations/ mean k-fold cross-validation)</th>
<th>Accuracy (model with all observations/ mean k-fold cross-validation) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Candacia armata</td>
<td>SST</td>
<td>1.80</td>
<td>6.49E-08</td>
<td>15.4</td>
<td>0.738/0.558</td>
<td>75.1/55.9</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>1</td>
<td>0.000589</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>1.79</td>
<td>4.60E-10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Centropages typicus</td>
<td>SST</td>
<td>1.002</td>
<td>&lt;2e-15</td>
<td>25.4</td>
<td>0.761/0.730</td>
<td>76.2/73.0</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>1</td>
<td>0.000587</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>2.413</td>
<td>&lt;2e-15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pH</td>
<td>1.965</td>
<td>2.12E-08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Centropages hamatus</td>
<td>SST</td>
<td>1.95</td>
<td>1.67E-05</td>
<td>28.7</td>
<td>0.846/0.747</td>
<td>80.3/74.7</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>1.96</td>
<td>2.46E-05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>2.62</td>
<td>7.84E-06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pH</td>
<td>1.88</td>
<td>9.46E-06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metridia lucens</td>
<td>SST</td>
<td>1.98</td>
<td>6.74E-11</td>
<td>7.65</td>
<td>0.637/0.574</td>
<td>63.3/57.4</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>2.70</td>
<td>1.87E-07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>2.92</td>
<td>2.06E-06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pH</td>
<td>1.89</td>
<td>1.17E-05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paraechaeuta modesta</td>
<td>SST</td>
<td>1.99</td>
<td>6.44E-07</td>
<td>27.5</td>
<td>0.783/0.644</td>
<td>78.2/64.4</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>2.71</td>
<td>5.68E-16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>2.95</td>
<td>&lt;2e-16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pH</td>
<td>1.86</td>
<td>2.67E-07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Paraechaeuta hebes</td>
<td>SST</td>
<td>1.96</td>
<td>1.12E-08</td>
<td>35.4</td>
<td>0.854/0.70</td>
<td>84.5/70.0</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>2.88</td>
<td>0.0896</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>2.75</td>
<td>1.80E-15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pH</td>
<td>1.44</td>
<td>6.87E-11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pleuromamma borealis</td>
<td>SST</td>
<td>1.91</td>
<td>1.68E-05</td>
<td>23.3</td>
<td>0.770/0.672</td>
<td>78.3/67.3</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>2.29</td>
<td>9.25E-12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>1.81</td>
<td>0.0231</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pleuromamma robusta</td>
<td>SST</td>
<td>1.93</td>
<td>1.20E-04</td>
<td>15.7</td>
<td>0.767/0.530</td>
<td>76.7/53.0</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>2.74</td>
<td>9.54E-09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>1</td>
<td>0.0028</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pseudocalanus elongatus</td>
<td>SST</td>
<td>1.35</td>
<td>2.78E-02</td>
<td>11.4</td>
<td>0.713/0.530</td>
<td>71.3/52.9</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>2.88</td>
<td>3.93E-05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>2.49</td>
<td>&lt;2e-16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pH</td>
<td>1.79</td>
<td>1.01E-08</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pH</td>
<td>1</td>
<td>1.52E-06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temora longicornis</td>
<td>SST</td>
<td>1.00</td>
<td>8.93E-04</td>
<td>17.5</td>
<td>0.723/0.631</td>
<td>71.8/63.1</td>
</tr>
<tr>
<td></td>
<td>SST</td>
<td>1.26</td>
<td>6.83E-15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Depth</td>
<td>2.61</td>
<td>&lt;2e-16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>pH</td>
<td>1.40</td>
<td>1.52E-06</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Model comparison: GAM vs Mahalanobis distance and Maxent**

When comparing the performance of the habitat suitability models, MaxEnt has ranked first in terms of accuracy, followed by GAM and Mahalanobis, with similar values for *C. glacialis* and *C.*
helgolandicus but higher values in for GAM in C. hyperboreus and C. finmarchicus (Figure 5).

However, GAM, which is a presence-absence based model, predicts correctly the potential distribution of C. glacialis along the Labrador Sea, Newfoundland shelf and the Davis Strait where it is abundant according to Head et al., (2003) and Pomerleau et al. (2011), whilst both MaxEnt and Mahalanobis predicted absence (Fig.S2).

**Latitudinal shift under climate change scenarios**

Distribution centroids of most of our studied species were projected to shift poleward under climate change (Table 5). All copepod assemblages showed northward shift of 0.1 to 13.5 km/decade for the shelf sea association species (Paraeuchaeta hebes and Paraeuchaeta modesta, Temora longicornis), of 3.7-11.3 km/decade in the arctic and subarctic association (Calanus hyperboreus, Calanus glacialis and Calanus finmarchicus) and of 1.9-17.8 km of temperate or warm-water species association (Metridia lucens, Pleuromamma robusta, Pleuromamma borealis, Calanus helgolandicus, Centropages typicus, Candacia armata). A southward migration in their center of gravity of about ca. 11-15.4 km/decade was found in other shelf sea and temperate association species (Pseudocalanus elongatus and Centropages hamatus). On average, the community poleward shift resulted in 8.7 km/decade. These shift rates were generally associated with a reduction located at the southern edge of the species spatial distribution. Such changes could have been linked to regional sea surface temperature warming. Despite the different thermal window of each of the 14 species analyzed, all centers of gravity have been located in the central temperate part (45-55º N) of the North Atlantic Ocean, both at present and future periods (Figure 6, Figure 7).
Table 5. Latitudinal and longitudinal shift of the species in the North Atlantic by taking into account the gravity centre of each of them. The shift is calculated as the distance (in km) between gravity centre of each species at present (2001-2020) and future (2080-2099).* Candacia armata, Centropages hamatus and Pleuromamma hebes gravity centres are calculated taking into account each species’ subpopulation patches individually and averaging them.

<table>
<thead>
<tr>
<th>Species</th>
<th>Latitudinal shift / decade</th>
<th>Longitudinal shift / decade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calanus finmarchicus</td>
<td>3.7</td>
<td>8.1</td>
</tr>
<tr>
<td>Calanus glacialis</td>
<td>11.3</td>
<td>15.1</td>
</tr>
<tr>
<td>Calanus helgolandicus</td>
<td>17.8</td>
<td>0.8</td>
</tr>
<tr>
<td>Calanus hyperboreus</td>
<td>7.8</td>
<td>-11.9</td>
</tr>
<tr>
<td>*Candacia armata (subpop.)</td>
<td>1.9</td>
<td>-10.8</td>
</tr>
<tr>
<td>*Centropages hamatus (subpop.)</td>
<td>-15.4</td>
<td>-3.7</td>
</tr>
<tr>
<td>Centropages typicus</td>
<td>2.2</td>
<td>-0.8</td>
</tr>
<tr>
<td>Metridia lucens</td>
<td>7.3</td>
<td>-14.7</td>
</tr>
<tr>
<td>*Paraeuchaeta hebes (subpop.)</td>
<td>13.5</td>
<td>-5.6</td>
</tr>
<tr>
<td>Paraeuchaeta modesta</td>
<td>12.8</td>
<td>9.6</td>
</tr>
<tr>
<td>Pleuromamma borealis</td>
<td>7.2</td>
<td>-4.7</td>
</tr>
<tr>
<td>Pleuromamma robusta</td>
<td>11.2</td>
<td>-3.4</td>
</tr>
<tr>
<td>Pseudocalanus elongatus</td>
<td>-11.0</td>
<td>-20.6</td>
</tr>
<tr>
<td>Temora longicornis</td>
<td>0.1</td>
<td>14.6</td>
</tr>
</tbody>
</table>


The habitat suitability of the 4 Calanus spp. is expected to be partially lost along their present southern distribution boundary ranges by 2080-2099 under climate change scenarios (Figure 7). The average northward retreat is more clearly seen in C. finmarchicus, C. glacialis and C. hyperboreus, with local projected shifts of up to 25 - 70km/decade in the southern limits of their distribution. We observed that the probabilities of C. finmarchicus for occupying large areas of Labrador Sea and Buffin Bay will increase considerably by the end of the century, as well as in the Northern North Atlantic gyre and the Irminger current. C. hyperboreus and C. glacialis map showed a similar pattern: their distribution will be mainly reduced north-westward, from Labrador to Newfoundland and Greenland Sea. The ecological niche of C. finmarchicus and C.


*C. glacialis* will suffer an important habitat reduction in the warm temperate waters of the central North Atlantic around the Gulf Stream and the North Atlantic Drift provinces, with potential local extinctions. A mean poleward migration of *C. glacialis* of 11.3 km/decade is estimated, much more acute than in *C. finmarchicus* (3.7 km/decade). Its habitat suitability will probably respond to future warmer SST shifting northward to the Irminger Current, Faroe-Shetland Channel and Norwegian trench (Figure 6, Figure 7). A noticeable habitat gain in *C. glacialis* has been observed in the northern subarctic region, south of Iceland and northern European shelf edge; this has contributed to a higher poleward mean shift. *C. hyperboreus* habitat suitability has also been projected to be reduced in the south-western edge of its spatial distribution where the species might face extinction. It will have a straightforward climatic response with a poleward mean latitudinal shift of ca. 8 km/decade. The species might disappear also from the southward flow of the East Greenland Current on the East Greenland. Results also have shown that *C. helgolandicus* might disappear from the warm temperate subtropical areas of the North Atlantic and some areas of the Bay of Biscay and southern European shelf edge, and that it will colonize the North Atlantic drift province and the east of the Oceanic Polar Front, with a relatively high net northward movement of ca. 18 km/decade (Figure 6, Figure 7).

The ecological niche models also have also projected a habitat gain in the temperate North Atlantic and along the Gulf Stream and North Atlantic current, in warm-temperate pseudooceanic (*Candacia armata* and *Centropages typicus*) and continental shelf assemblages (*Pseudocalanus elongatus*). The Warm temperate *Pleuromamma borealis* and the cold mixed water *Pleuromamma robusta* also will gain habitat in the Oceanic Polar Front and subarctic region of the North Atlantic. The warm temperate *Paraechaeta modesta* might face local extinction along the southern limits of its distribution. The overall suitable habitat is therefore expected to increase in these temperate-warm and shelf species assemblages. The habitat suitability of other set of studied species did not show latitudinal shift, but an east to west asymmetry: *Centropages hamatus* might colonize the Bay of Biscay and the southern European shelf edges, and *Paraechaeta hebes* would extinct from the North Sea and the southern European shelf edge. The GAMs deviance explained for *M. lucens* is too low (7.7%) to draw conclusions on its habitat suitability change.

*Species turnover under climate change scenarios*

Results have shown a high species turnover area (4 to 11 species) south of the Oceanic Polar Front (42.8-78.5%) compared with the global North Atlantic turnover (ca. 10%), covering vast areas of the center of North Atlantic drift and extending up to the northern boundary of the influence of the Mediterranean water (Figure 8D). Another moderate turnover rate (2-4 species, 21.4%) has been found in coastal zones of southern Bay of Biscay and in the Continental shelf current. High intensity of species invasion (3-5 new species) was projected to be concentrated along through the Oceanic Polar Front (Figure 8C). Local extinctions were projected to be most common (3-6 species lost) in temperate waters of the North Atlantic, south of the Oceanic Polar Front and by the northern boundary of the influence of Mediterranean water (Figure 8B). High turnover areas overlap with areas of both highest SST and phytoplankton biomass changes between present and future periods (Figure 3), and also they correspond relatively well with the southern edge of the cold-temperate, subarctic and arctic species assemblage (*C. hyperboreus, C. glacialis, C. finmarchicus*), and the northern biogeographic boundaries of the warm
temperate species assemblage (C. armata, C. typicus) (Figure 7). Additionally, we compared present and future species richness on a pixel basis (Figure 8A).

![Species richness (2001-2020)](image1)

![Extinction](image2)

![Colonization](image3)

![Species turnover](image4)

**Figure 8.** A) Species richness model at present (2001-2020). B) Colonization model: number of new species that will occur at each pixel by 2080-2099. C) Extinction model: number of species that will disappear at each pixel by 2080-2099. D) Turnover model: Number of species that will either colonize or gone extinct at each pixel by 2080-2099.

**Phenology changes under climate change scenarios**

We carried out a seasonal quantitative analysis on C. hyperboreus and C. finmarchicus since their predicted phenological patterns have been similar to observed ones (Figure 9A, Figure 9B). C. helgolandicus and C. glacialis have been omitted from such analysis owing to different observed and predicted seasonal signals. We have not performed any phenology analysis in the NMFS-COPEPOD set of species either, due to time series shortness.

Our projections showed an earlier timing of copepods predicted occurrences annual maxima of 12 days in both species (C. finmarchicus annual maxima at present was 15 April, and future: 3 April; for C. hyperboreus: 27 April (at present), 15 April (future)) (Figure 9C, Figure 9D).
Figure 9. A) and B) *C. finmarchicus* and *C. hyperboreus* CPR number of observations versus the habitat model predictions for the CPR sampling area in 1970-2004. Predicted occurrence phenology of *C. finmarchicus* and *C. hyperboreus* at present (2001-2020) and future (2080-2099) in the whole study area. Vertical arrows represent the seasonal peak of each species at each period.
Discussion

Habitat suitability models

The habitat suitability modelling enabled us to identify three key environmental variables (sea surface temperature and salinity, and depth) that determine the present distribution of Calanus spp. Sea surface temperature is, in general, the environmental driver explaining most of the variance of species occurrence in the four Calanus spp. (especially in C. glacialis) habitat models. Previous niche model related works (Beaugrand et al., 2013; Chust et al., 2014a) have shown similar results. Brown et al., (2004) showed that temperature is the most broadly influential factor controlling biological processes. Salinity and depth were also identified as key environmental variables that determine Calanus spp. distribution. It is interesting to pinpoint that the models included surface phytoplankton biomass and pH. These variables have not been used frequently in zooplankton habitat modelling up to date, since they explain low variance of Calanus spp. occurrence, as in previous attempts (e.g. Reygondeau and Beaugrand, 2011). This is probably because it represents only a part of the food available and because food is not a limiting factor above the mixed layer depth, regardless of its key importance (Irigoien, 2004; Moller et al., 2012).

The Mahalanobis distance and Maxent species distribution models failed in predicting the spatial distribution of C. glacialis along the Labrador Sea, Newfoundland shelf and the Davis Strait, probably because the CPR routes do not regularly cover the Labrador Sea area and this type of models are based only on the presence points while not accounting for absence points. GAMs instead, has proved to be a useful and accurate model to quantify the ecological niche of Calanus spp. in the North Atlantic. This conclusion is based on 1) the accuracy values of random (69-85%) and temporal (74-85%) validation, 2) its flexibility to incorporate seasonal variability, and 3) its performance comparing both accuracy values and spatial distribution maps with MaxEnt and Mahalanobis distance algorithms. The temporal cross-validation showed that model accuracy is relatively good (74-85%) for the models built in cold regime and extrapolated and validated in the warm period. Therefore, we were confident to use the GAM-based habitat models generated for Calanus spp. in the future climate simulation.

Latitudinal shift

Projections of 14 main copepod species in the North Atlantic by the end of the century indicated a prevailing poleward shift of most of the studied species, with poleward community shift of 8.7 km/decade on average, and an important species range variation (-15 to 18 km per decade). Poleward shifts of warm temperate copepod assemblage were more important than the range contraction of the subarctic and arctic species assemblage.

At species levels, projections revealed a poleward shift with a slight contraction of the southern limit of habitat suitability distribution of the C. finmarchicus, C. glacialis and C. hyperboreus, and a shelfward constriction of C. helgolandicus, disappearing from oceanic warm waters south of the Oceanic Polar Front. This study has shown that the Hutchinson’s ecological niche of C. finmarchicus and that of C. helgolandicus will kept well separated in the future. This species
niche separation was well described for the historic CPR dataset (1942-2002) in Helaouët & Beaugrand (2007). Results have also shown that the biogeographical range of *C. glacialis* and *C. hyperboreus* will be rather similar. The subarctic species *C. finmarchicus* has a broader tolerance interval than its congener *C. helgolandicus* (Helaouët & Beaugrand, 2007), and it is therefore able to support larger environmental variations. The modelled spatial distribution of *C. finmarchicus* showed that this species mainly occurred in areas above the Oceanic Polar Front (Dietrich, 1964). *C. finmarchicus* is adapted to a cold oceanic environment, with high mixing during the winter, and it is indicative of the Atlantic Arctic Biome and especially the Atlantic Arctic and Subarctic provinces defined by Longhurst (1998). At the northern limits of its distribution, i.e. north of Iceland, it occurs along with the Arctic species *C. glacialis* and *C. hyperboreus*, while in the northeastern North Atlantic, in the North Sea, and in the southern part of the Norwegian Sea, it co-occurs with *C. helgolandicus* (Conover, 1988). *C. hyperboreus* and *C. glacialis* are both known to be associated with water of Arctic Ocean origin, with a main distribution in the northern North Atlantic Ocean, the Greenland Sea and the Arctic Ocean (Hirche, 1991). The present time biogeographic features of *C. finmarchicus*, *C. hyperboreus* and *C. glacialis* have been relatively well resolved by our habitat models. *C. helgolandicus*, instead, is more adapted to the temperate waters of the Atlantic Westerly Winds Biome (Longhurst 1998), although our projections have shown that it will be mainly present along shelf edges of the Mediterranean, the northern and southern European shelf edges, and the North Sea and south of Icelandic water. The habitat suitability of this species will respond fast to climate change, from local extinction in the warming waters of temperate west North Atlantic to colonization in waters of the Newfoundland continental shelf. *C. helgolandicus* can be considered as a climatic sensible species since SST contributes to explain half of the deviance explained of the GAM.

The warm-temperate and continental shelf set of species assemblages have shown the highest local northward shifts. Southern temperate regions are becoming warmer and are expected to provide suitable habitat for the warm-temperate and temperate pseudoceanic species assemblages. Thus, overall, warm species assemblages will respond faster to climatic change and cold species assemblages will retract their core distribution northward at a slower pace.

Our GAM models projected a plankton community latitudinal shift of 8.7 km per decade. Cheung et al., (2009) estimated a global-warming-related shift of marine species (fishes and invertebrates) of 1.4-28 km/decade, which is substantially less than the estimations of Sorte et al., (2010) for 129 marine species (190 km/decade). At species level, the rate of northward movement projected in *C. finmarchicus* (3.7 km/decade) is considerably lower than the change in distribution suggested by Helaouet & Beaugrand (2009) for nearly the same area and period (1 degree/decade).

In this study, the use of GAMs to predict the habitat suitability of the species has been limited to a geographical subset in the North Atlantic, hence, the biogeographic range of the species are not fully represented. This drawback may have affected to the estimates of the gravity center of the species because in some cases the southern colonization area overcomes in number the northern extinction area which results in an underestimation of the poleward mean latitudinal shift of the species. We found a southward migration projected in *Centropages hamatus* (15.4 km/decade) and *Pseudocalanus elongatus* (11 km/decade) by the end of century. This is
because the southern colonized area of these two species is larger compared to the north area,
yielding a net southward migration of the species. Moreover, SST seems not to have a driving
effect in the distribution of these two continental shelf species assemblages (6.6% out of the
total 28.7% deviance explained in C. hamatus and 1.0% out of 17.6% in P. elongatus), and
other environmental variables appear to be more important such as phytoplankton. Although
covering the whole biogeographic range of the species is preferable, the estimation of gravity
center considered here is relatively reliable to capture population shifts.

Species turnover

We have identified a high turnover area south of the Oceanic Polar Front covering large areas
of the Centre of North Atlantic drift and extending up to the northern boundary of the influence of
the Mediterranean water. This high turnover matches with the southern distribution of arctic and
subarctic species assemblages and with the northern limits of the warm-temperate copepod
assemblages. Hence, the boundaries of species biogeographic domains are prone to suffer
higher extinction or colonization rates, resulting in a high turnover of species. High turnover
areas accordingly coincide with an important predicted SST increase by the end of century
where warm species assemblage could benefit to settle their populations there, while the
southern limits of the cold subarctic and arctic species assemblages will retract.

These projections, that agree with the eco-physiology of animals (Tewksbury et al., 2008),
suggest that marine communities at the extreme ends of their environmental space are
especially at local risk due to climate change. The poleward expansion of the species in the
subpolar regions is limited by the availability of suitable habitats. Thus, the retreat of the
southern biogeographical limits of species leads to a general range constriction, having a
potential impact in the species interactions with higher trophic levels.

It is difficult to identify the causes of diversity changes we projected. At the core of most of the
hypothesis that try to explain contemporary patterns of diversity, there is a link between the
abiotic environment and the species diversity (Turner, 2004). Features of marine environment
that have affected patterns of pelagic diversity in our studied copepods seems to be related to
(1) local rapid temperature and phytoplankton changes expected by the end of century in the
Newfoundland Continental Shelf and along the Oceanic Polar Front southern edges (Figure 3),
that may have contributed in a different manner to species colonization or extinctions, (2) the
seasonality of these environmental variables (Woodd-Walker et al., 2002), and (3) the chemical
and physical properties of the ocean (Ruddiman, 1969), all of which are influenced by climate.
These factors are often interwoven, also act at different scales and their contribution varies
geographically, as appeared in our modelled species richness map (Figure 8A).

In high species turnover areas, species characterising the warm-temperate oceanic and
pseudoceanic assemblages are moving poleward and will overlap with the southern distribution
of subarctic and arctic species assemblage. This might lead to trophic mismatches and alter the
prey-predator relationships. The high turnover areas match with an important predicted SST
increase by the end of century (4 to 6 °C, see Figure 3) from which warm species assemblage
could benefit to settle their populations there. The Oceanic Polar Front (Dietrich, 1964) has
acted as a sharp boundary for shelf edges and warm temperate species associations limiting
dispersal northwards (Figure 8D). Our models have identified a high-turnover area along the Newfoundland Continental shelf and along the western part of the Oceanic Polar fronts that matches with the southern distribution of the arctic and subarctic species assemblages and with the northern distribution of warm-temperate species assemblage. These changes in copepod community structure, which are key species at the base of the marine food webs and transfer energy from primary producers to higher trophic levels, may propagate through higher trophic levels (Kirby & Beaugrand, 2009; Chust et al., 2014b), having an ecosystem wide effect on the North Atlantic marine provinces.

Other factors that contribute to maintain copepod pelagic diversity are the prevailing currents and ocean circulation. Our spatial examination of the effect of climate induced changes in SST in the temperate western North Atlantic revealed marked regional disparities in the effect of climate change (Figure 3). For example, in some areas of the subarctic region south of Labrador Current in North Atlantic gyre (i.e. -40°W-30°W and 55°N-60°N) is expected to cool by the end of the century due to the reduction in Meridional Overturning Circulation (MOC) (Rahmstorf, 1999), whereas in most of the other northern regions (north of the Oceanic Polar Front, subarctic region above 50°N, south Iceland waters) a slight warming is likely to occur (1-2°C). Despite the MPIOM SST Model projected a mean increase of 1.54°C±0.35°C in the overall study area of the North Atlantic by 2100, regional climate induced changes in SST will have strong effects on local populations; a decrease of cold water copepods is expected where warming takes places, which coincides with the southern edge of the species spatial distribution.

**Phenology**

In agreement with previous studies highlighting the advance in spring seasonal peaks of zooplankton times series (e.g. Edwards & Richardson (2004) reported a 10 days advance in annual maxima in North Sea copepods on the 1958-2002; Greeve et al., (2004) estimated an annual peak occurring 37 days earlier in Helgoland Road cladocerans on the 1975-1999), we expected similar phenology changes (12-13 days earlier in the year) between present and future periods in *C. finmarchicus* and *C. hyperboreus* (Figure 9).

Zooplankton timing variability is often linked with temperature and/or with phytoplankton biomass during the preceding weeks or month. For zooplankton, taxa that have their maximum occurrences or abundances and activity in spring-summer like for example *C. finmarchicus* and *C. hyperboreus*, the usual pattern is “earlier when warmer” (Mackas et al., 2007; Edwards & Richardson 2004). Our modelled projected seasonal peak of the two species is also occurring earlier, responding to a climate warming trend by the end of the century; these changes may propagate higher up in the food web.

Results of the phenology model have shown the following sequence: After the SST seasonal peak by mid-April zooplankton maxima will occur: *C. finmarchicus* and *C. hyperboreus* have shown a positive relationship with a “sudden” increase in SST by the beginning of April (Figure 9). One or two weeks later, the phytoplankton bloom will take place. Modelled phytoplankton seasonal peaks have remained rather fixed in time - a week earlier by 2100 comparing present to future periods (data not shown). Modelled phytoplankton blooms will be presumably more
dependent on day length and light intensity (Eilersten & Wyatt 2000) rather than on temperature mediated physiological responses. *C. finmarchicus* and *C. hyperboreus* annual peaks, therefore, seem to be more dependent on temperature rather than on food availability. Melle et al., (2014) reported also positive relationships between maximum abundances of *C. finmarchicus* and maximum temperatures in the North Atlantic, with no clear relationship with phytoplankton biomass maximum.

Although organisms may respond to aspects of climate change other than to the environmental window of our study, our aim was to generate projections for shifts in phenology from physical and biological descriptors of the climate related variables. Seasonal peak timing variability cannot be only explained by the acceleration of the physiological rates owing to temperature increase or food availability. Other factors that have not been included in our habitat models and might be relevant: the reproductive timing of parent generation, the onset from dormancy, the advection processes or changes in flow patterns, and the species specific prey-predator relationships.

The aforementioned phenology studies as well as our modelled phenology approach are spatially limited because they take into account only a subset of the entire geographic range of the species. If the overall distribution area had been analysed, we would not expect to obtain substantial shifts in phenology since the species would shift poleward in the future to a similar thermal window where it could succeed. On the contrary, local studies of zooplankton phenology (e.g. Mackas et al., (1998) in a Subarctic Pacific station; Bornhold et al., (1998) in the Strait of Georgia), reported higher shifts (30-60 days), since local environmental changes are subjected to more variability and the climatic response of the species will be more pronounced. Our spatial scale is in between local and entire biogeographic range studies, which might explain the intermediate mean shifts values found.

**Model uncertainties and implications**

Our projections assume no thermal adaptation of the species (*sensu* population fitness) to a changing environment. This assumption is supported by a recent study (Hinder et al., 2013) revealing that *C. finmarchicus* and *C. helgolandicus* show a lack of thermal adaptation to rising temperatures. Species that fail to acclimatize physiologically or evolve genetically to increasing temperature will either move northwards following their habitats (Walther et al., 2002; Parmesan & Yohe, 2003; Parmesan, 2005, 2006; Hickling et al., 2006; Thomas, 2010) or become extinct (Thomas et al., 2004).

GAM occurrence models of *Calanus* spp. in 1970-2004 period are in agreement with the observed spatial distribution from CPR surveys (Helaouët & Beaugrand, 2007), conventional sampling datasets (Heath et al., 2004, Reygondeau & Beaugrand, 2011), CPR data reconstruction studies (Chust et al., 2014a) and with other distribution studies inferred from models (Speirs et al., 2005; Hinder et al., 2013; Durbin & Kane 2007, Kane & Prezioso, 2008). Therefore, the general agreement of our models with occurrence records and other modelled distribution studies supports the view that climatic (mainly SST), and also SSS, depth, and biological (phytoplankton) factors are enough to provide a first approximation of niche shifts under climate warming (Pearson & Dawson, 2003, Araujo & Guisan, 2006).
Our study projects the spatial distribution of a representative subset of the North Atlantic copepod community. We have gained new insights on where the species are potentially able to expand or extinct locally. There are uncertainties related to our projections; first to the climate model itself, second to the habitat model, and third to the coupling of both. The accuracy of our habitat suitability projections will depend on both the intrinsic characteristics of the studied species as well as the extrinsic biotic and abiotic factors. Our model (i.e. GAM) does not incorporate other ecological processes such as dispersal limitation and population dynamics. It has been shown that such ecological processes play an important role in the plankton spatial distribution (Irigoien et al., 2011; Chust et al., 2013). Research efforts should focus on including this two mechanisms in the habitat modelling frame, in the same way as has already been done for fish and invertebrates (e.g. Cheung et al., 2009). The application of the combined analytical methods beyond those traditionally used by ecologists will shed new light on the understanding of climate impacts on plankton communities.

Conclusions

The habitat suitability modelling enabled us to identify three key environmental variables (sea surface temperature and salinity, and depth) that determine the present distribution of copepods in the North Atlantic. GAM has proved to be a useful and accurate model to quantify the ecological niche of Calanus spp. This conclusion is based on 1) the accuracy values of random (69-85%) and temporal (74-85%) validation, 2) its flexibility to incorporate seasonal variability, and 3) its performance comparing both accuracy values and spatial distribution maps with MaxEnt and Mahalanobis distance algorithms. The temporal cross-validation showed that model accuracy is relatively good (74-85%) for the models built in cold regime and extrapolated and validated in the warm period. Therefore, we were confident to use the GAM-based habitat models generated for Calanus spp. in the future climate simulation.

Projections of 14 main copepod species in the North Atlantic by the end of the century under climate change scenarios indicate: 1) a prevailing poleward shift of most of the studied species, with poleward community shift of 8.7 km/decade on average, with an important species range variation from -15 to 18 km per decade; 2) a high species turnover area of local colonization and extinction south of the Oceanic Polar Front has been identified and coincide with an important projected SST increase by the end of century; 3) the seasonal peak of copepods will occur earlier in the year in response to the ocean warming trend. All these changes may propagate higher up in the food web.

Acknowledgements

We acknowledge CliSAP-Integrated Climate Data Center and NOAA for providing climate and bathymetry data, respectively. E. Villarino has benefited from a PhD Scholarship granted by the Iñaki Goenaga–Technology Centres Foundation. We are grateful to present and past staff of SAHFOS who have contributed to the maintenance of the CPR time series. The authors wish to acknowledge also to Markus Kreus in providing data from HAMOCC and MPIOM models through DKRZ [http://www.dkrz.de/].
References


Supplementary information

Fig. S1. Number of occurrences from CPR survey for 1970-2004 period. A) *C. finmarchicus*, B) *C. helgolandicus*, C) *C. glacialis*, D) *C. hyperboreus*. 