# Pan-Arctic suitable habitat model for Greenland halibut

Supplementary information

Mikko Vihtakari[1], Robinson Holdoir[1], Margaret Treble[2], Meaghan D. Bryan[3], Bjarki Elvarsson[4], Adriana Nogueira[5], Elvar H. Hallfredsson[1,6], Jørgen Schou Christiansen[6,7], and Ole Thomas Albert[1]

[1]Institute of Marine Research, Tromsø/Bergen, Norway
[2]Fisheries and Oceans Canada, Winnipeg, Canada
[3]Resource Ecology and Fisheries Management Division, Alaska Fisheries Science Center, NMFS, NOAA, Seattle, WA, USA
[4]Marine and Freshwater Institute, Reykjavik, Iceland
[5]Greenland Institute of Natural Resources, Nuuk, Greenland
[6]UiT – The Arctic University of Norway, Tromsø, Norway
[7]Åbo Akademi University, Turku, Finland

13 January, 2021

## Supplementary figures

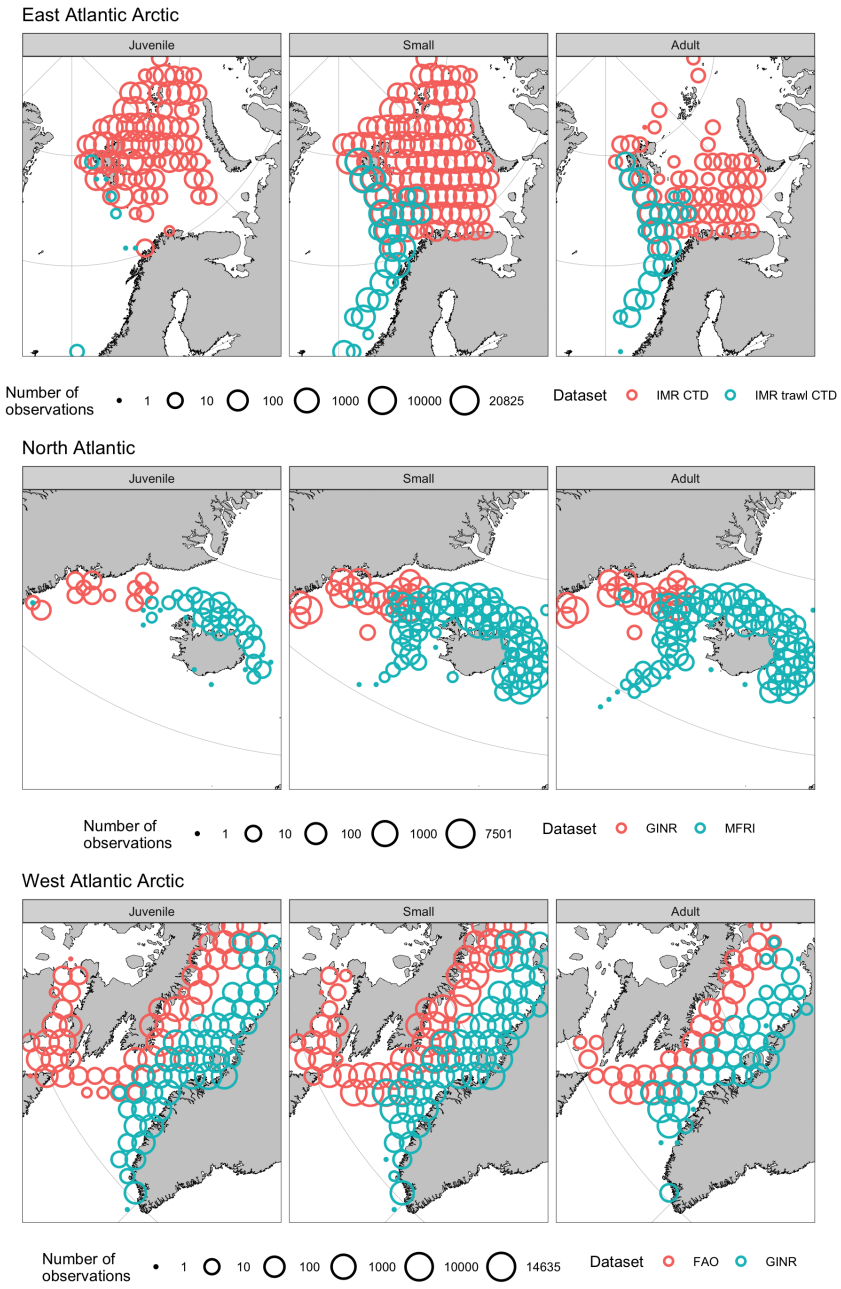



Figure S1: Gridded positions of temperature-depth selection data used in the habitat space estimation. Colored dots indicate fish observations in a given size-class. The color refers to the data source and size to the number of binned fish observed in the grid cell. See Figure S2 for the Bering Sea data.

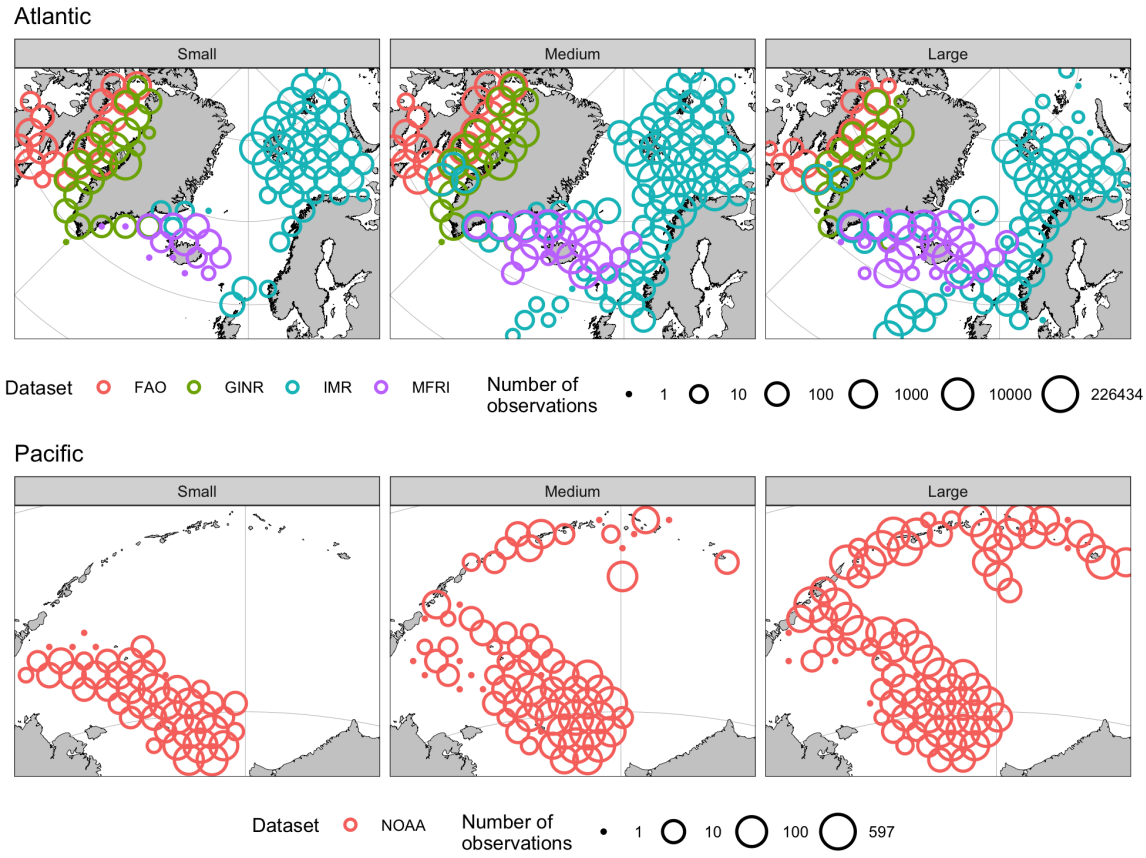



Figure S2: Gridded positions of data used for the size based SHM validation. Colored dots indicate fish observations in a given size-class. The color refers to the data source and size to the number of binned fish observed in the grid cell.

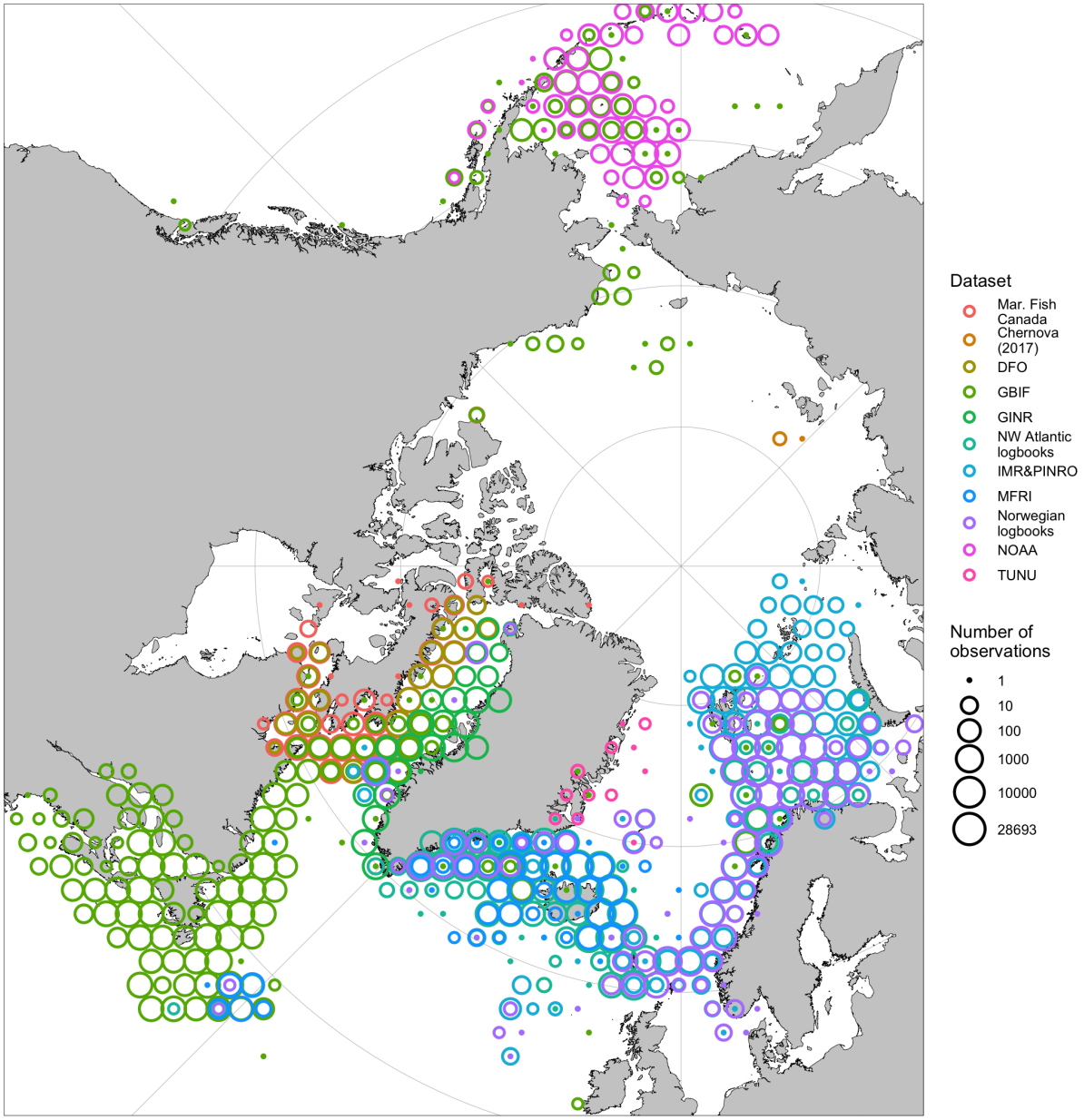



Figure S3: Gridded positions of data used for the general SHM validation. Colored dots indicate occurrences. The color refers to the data source and size to the number of binned observations (not necessarily individual fish) in the grid cell.

## SI text 1. Suitable habitat modeling

### Interpreting habitat spaces

In this paper, we use the niche-based environmental envelope concept, which has been used in various forms in ecological research in the past [e.g. see Hutchinson (1957); Soberón and Nakamura (2009); Araújo and Peterson (2012); and the references therein]. This concept has, however, to our knowledge not been used to study the influence of bottom depth, temperature and salinity on the observed distribution of a marine species similarly to how we use it here. While the mathematics behind our method are relatively simple and the concept can be derived using logic, some further explanation and considerations are required. Since there was not enough space to explain the concept thoroughly in the body text of the article, we explain the concept of indirectly constraining multiple variables and how to apply them using observational data and the associated R package in this supplementary text.

Physiological tolerance curves for a single variable, such as temperature, and a given (ectotherm) species have been established through decades of ecological and physiological research (see e.g. Fry and Hart, 1948; Pörtner, 2002). These curves operate on the assumption that there are physiological processes that constrain the success of an individual given a level of an environmental variable, and that tolerance to this variable can be measured through a physiological response. Such tolerances have excessively been tested using physical or chemical variables, such as temperature (e.g. Pörtner, 2002), pressure (e.g. Harper *et al.*, 1987), salinity (e.g. Kultz, 2015), pO2 (e.g. Domenici *et al.*, 2017), pCO2 (e.g. Enzor *et al.*, 2013) or contaminant levels (e.g. Patra *et al.*, 2015) to mention a few. Recent research has focused on the simultaneous effects of multiple constraining variables on a single measurable response, often by looking at the interaction effects using linear regressions or their generalized variants (Youcef *et al.*, 2013). Similar curves can, however, be formulated in multivariate space using multiple constraining variables as axes and measurable response, probability of occurrence (for binary data), or density of points (for presence-only data) as a response dimension (Figure S4). If the constraining variables are related to the environment and the sample of observations representative for the domain of the study, the constraining factors do not need to stem from physiological tolerances. It is enough that the constraining variables lead to a differential habitat selection through indirect (=unknown) ecological and physiological effects. This realization forms the basis of the environmental niche-based approach as reviewed in Soberón and Nakamura (2009).

In the associated R package (Vihtakari, 2021), we call these environmental niches "habitat spaces" as a contrast to geographical spaces, and because these multivariate spaces define the habitat in the geographical space given the model assumptions. The habitat space concept is a synonym to environmental niche, space or envelope and other terms used in the literature. We further call the resulting geographical spaces "suitable habitat models" instead of "species distribution models" to highlight the fact that these models do not estimate real-world species distributions unless all constraining variables are defined correctly in the habitat space and the oceanographic model projects these variables reliably. Such a scenario is hardly possible given the complexity of the real-world. In this paper, we use bottom temperature and depth as constraining environmental variables to examine whether other unknown factors influence the distribution of Greenland halibut.

We call these specific habitat spaces "TD-spaces". The constraining variables and how they are plotted influence the interpretation as explained in Figure 7. While temperature (the x-axis) influences the individuals through physiology, the bottom depth range where Greenland halibut occurs (50-2000 m) is unlikely to set physiological limits for a deep-sea adapted fish without a swim-bladder (Sebert, 2002). Nevertheless, centuries of fishing experience demonstrate that different fishes are caught above different bottom depths. The underlying reason for the depth preference is not the depth itself, but other adaptational, ecological and physiological factors that are interwoven into the depth selection. To model the habitat suitability, it is not necessary to know why depth limits the distribution of a species as long as one has strong evidence that it does. Such an assumption could lead to erroneous conclusions. However, so could trying to incorporate all the factors that may influence the depth preferences when data for these parameters are limited.

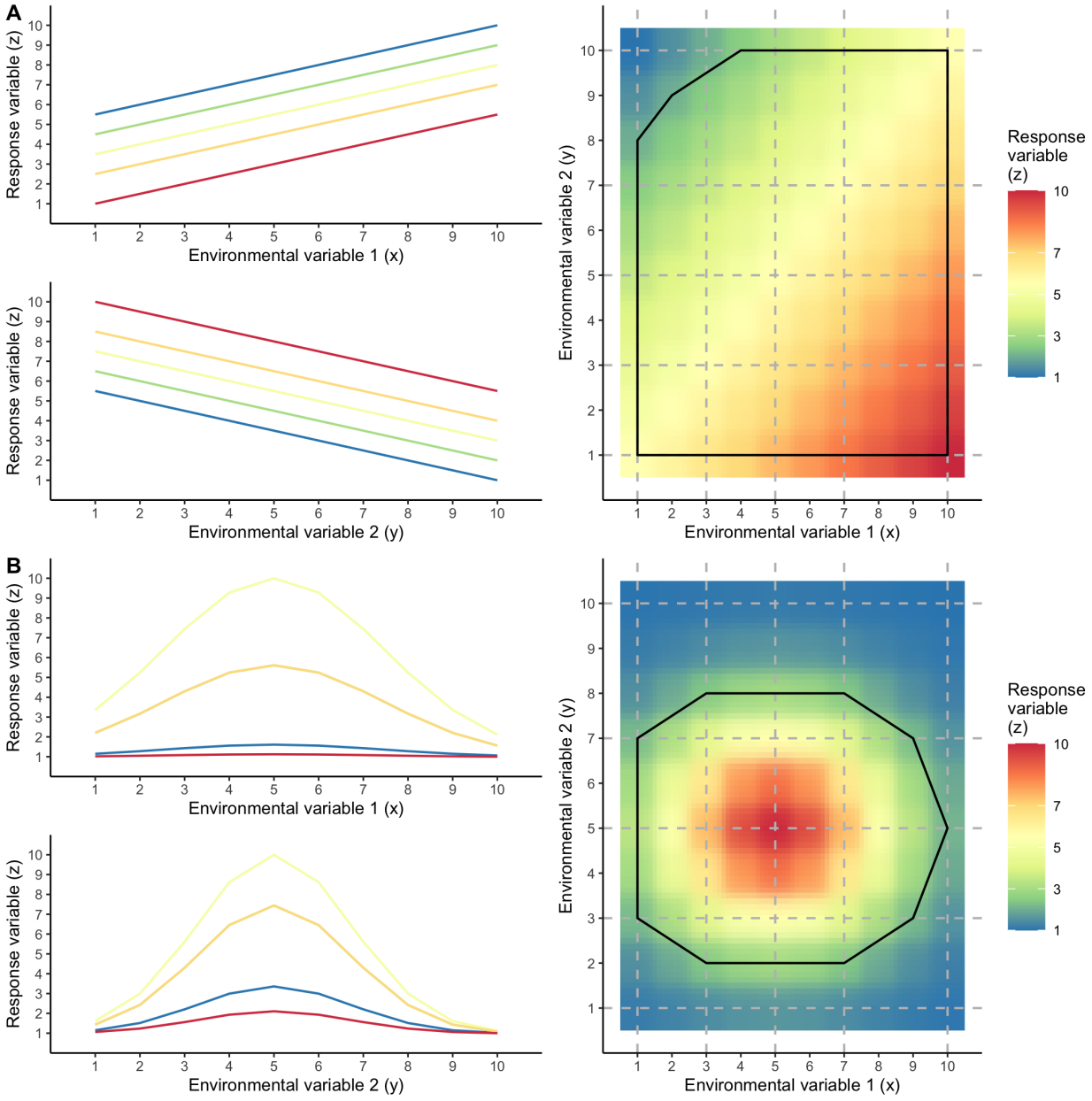



Figure S4: Two examples of how multiple constraining variables (or stressors) can be seen in multivariate space. Environmental variables 1 and 2 are sampled along a gradient or in a controlled experiment and their synergistic effect on a response variable is measured (shown in left panels). The same results can be shown in three dimensions (right panels) where the environmental variables are on x- and y-axes and the response variable is on the z-axis (indicated as color). The dashed lines indicate the sampled values. The polygon indicates the habitat space assuming that z values $< 20$ percent of the maximum are outliers. A) Both environmental variables have a linear effect on z. Linear relationships lead to a plane in the 3D space. The direction of these relationships depends on which way the multivariate plane is tilted. B) Both environmental variables are normally related to z. Nonlinear relationships lead to various surfaces. There are no mathematical limits expanding this concept to as many dimensions as there are data for the constraining variables.

The TD-spaces used in this paper tell their story through multiple considerations. First, if a TD-space was rectangular, one would reach the same solution by limiting the variables separately. Consequently, round corners indicate that environmental variables in that region of the habitat space limit the habitat of a species together (Figure S4). Second, available TD-space can be obtained from the oceanographic model by calculating a TD-space from all values for the entire domain of the study. If an observational TD-space for a species lies outside of the available TD-space, either the oceanographic model values or observational values are wrong. Third, any line, plane, spine shapes, concave edges or other similar multivariate variants in the habitat space represent gradients, which might indicate sampling bias rather than the tolerance of a species (either the entire distribution of a species was not sampled or the concave variable-space areas do not exist in the real world) given that the variables are constraining the habitat of a species. A convex hull can be taken from the TD-spaces to remove such effects. Further, if the observational dataset is not representative of the domain of the study (likely in our study since the survey data did not cover the entire Arctic), sampling biases might unnecessarily constrain the habitat space. Teasing out such sampling biases can be difficult as demonstrated in Figures 3 and 4. Fourth, since the temperature is on the x-axis, the left and right sides of the TD-space are likely constrained by biological processes, either physiology or ecology.

Finally, the cut corners that make the TD-spaces circular are likely due to different reasons: the bottom left corner has low temperature and depth (Figure 7). The survey data were mostly collected using trawlers during summers, and the Bering Sea data demonstrated that Greenland halibut can inhabit this region of the TD-space. It is uncertain whether other variables (competition, predation, salinity) cause Greenland halibut to avoid cold and shallow habitats on the Atlantic side or whether the lack of this corner is due to sampling bias. The bottom right corner is most likely not suitable habitat due to high temperatures (physiology) and competition/predation (Atlantic cod tends to inhabit this zone on the Atlantic side of the Arctic). The top right corner exists West of Iceland and might be a mismatch between the NEMO model and observational data. The top left corner does not exist according to the NEMO model. The interpretation above indicates that bottom depth does not set hard physiological limits for the occurrence of Greenland halibut, but serves as an indirect proxy affecting other factors that limit the distribution.

### Practical implementation

To begin the demonstration, we load the required R packages and example data for the general model presented in the paper. Note that the results here look different from the results in the body text due to data-sharing restrictions.

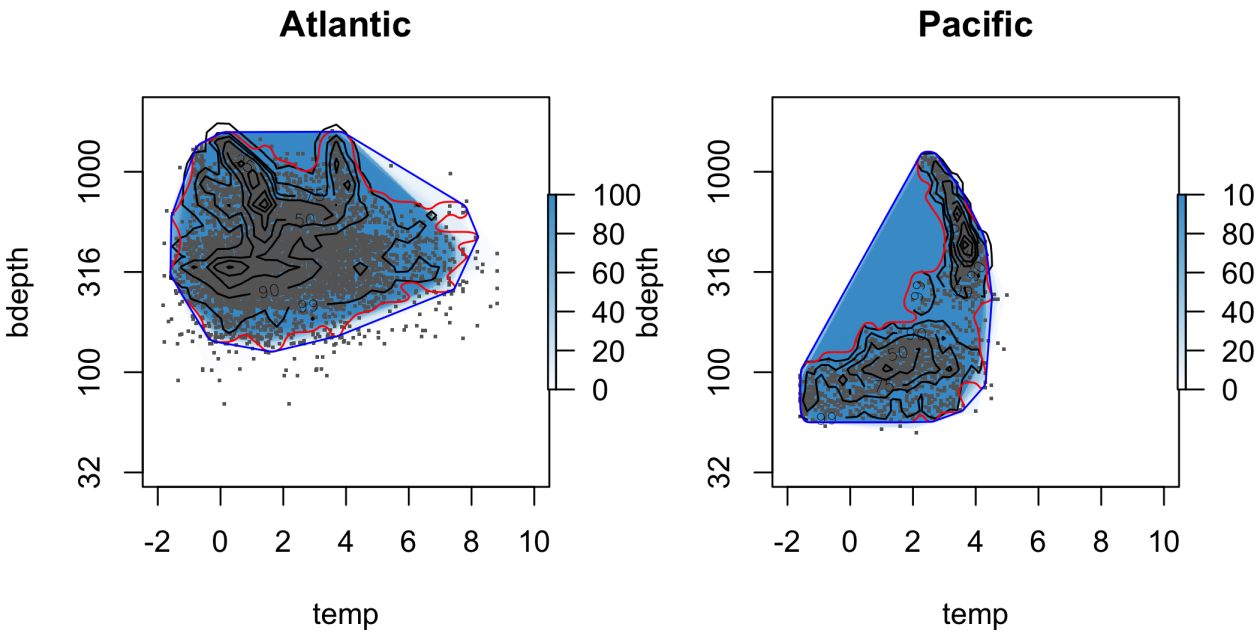
```
# Packages
packages <- c("tidyverse", "cowplot", "sp", "data.table", "devtools", "ggOceanMaps")

## Install packages not yet installed
installed_packages <- packages %in% rownames(installed.packages())

if (any(installed_packages == FALSE)) {
  install.packages(packages[!installed_packages],
                   repos = c("https://mikkovihtakari.github.io/drat", "https://cloud.r-project.org")
  )
}

if(!"SuitableHabitat" %in% rownames(installed.packages())) {
  devtools::install_github("MikkoVihtakari/SuitableHabitat", upgrade = "never")
}

## Load the packages
invisible(lapply(c(packages, "SuitableHabitat"), library, character.only = TRUE))

# Load data
```

```
load("../../Data/SI_size_based_TD_data.rda")
load("../../Data/SI_ghl_distribution_data.rda")
```

We form a cloud of points in equally many dimensions as the number of limiting variables in the TD-space. We, then, calculate a multivariate kernel-density (mKDE) of these points as an extra dimension (Chacón and Duong, 2018). This mKDE is a multivariate analog to the original tolerance curves with variables influencing the density of points synergistically (response variable) given that all variables are constraining the habitat of the species (Figure S4).

```
z <- split(x, list(x$side))

tdsp <- lapply(z, function(k) {
  out <- habitat.space(data = k,
                       cvars = list("temp" = NA, "bdepth" = NA, "sal" = c(29, 37)),
                       log.transform = c(FALSE, TRUE, FALSE),
                       non.suitable.level = 0.001,
                       chull.correction = TRUE
  )
})
#>   |                                                                      |
```

```
par(mfrow = c(1, 2))

for(i in names(tdsp)) {
  plot(tdsp[[i]], ylim = c(1.5,3.3), xlim = c(-2, 10))
  title(i)
}
```

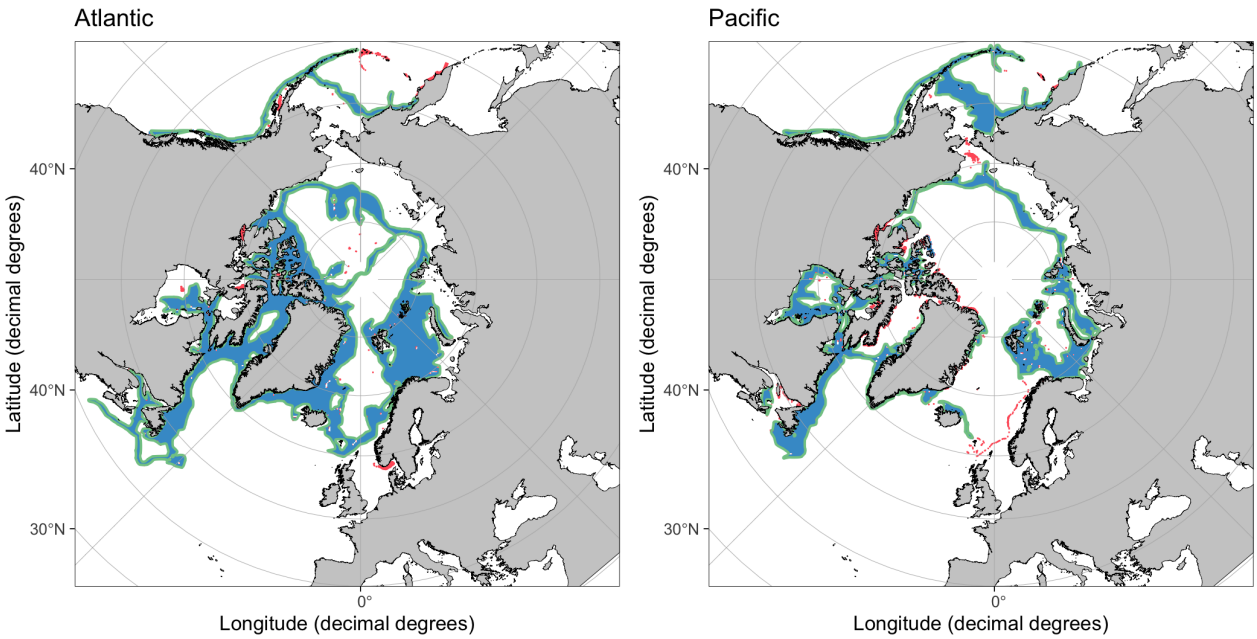



The kernel-density smoothing parameter, bandwidth, binning and variable range all influence the outcome of the mKDE and have to be optimized for each dataset (Chacón and Duong, 2018; Duong, 2019). In the SuitableHabitat package we have optimized these parameters for the data in this paper but also provided the users a possibility to change some of the parameters. The results influence the optimal KDE probability level

threshold for the habitat suitability in the TD-space. Had the dataset used for the mKDE been error-free, one could have assumed that none of the points in TD-space were outside the suitable habitat of Greenland halibut. However, since the dataset was collected from multiple sources over several decades (from 1969 to 2019), it was bound to contain reporting errors. The probability level for outliers was obtained iteratively by making the TD-space to cover most points that were close to each other. The level was set to 0.999 assuming 0.1% of KDE probabilities in the size-based TD-dataset as outliers. In practice, this led to 0.2 to 1.5 % of individual fish observations being defined as outliers depending on the size group.

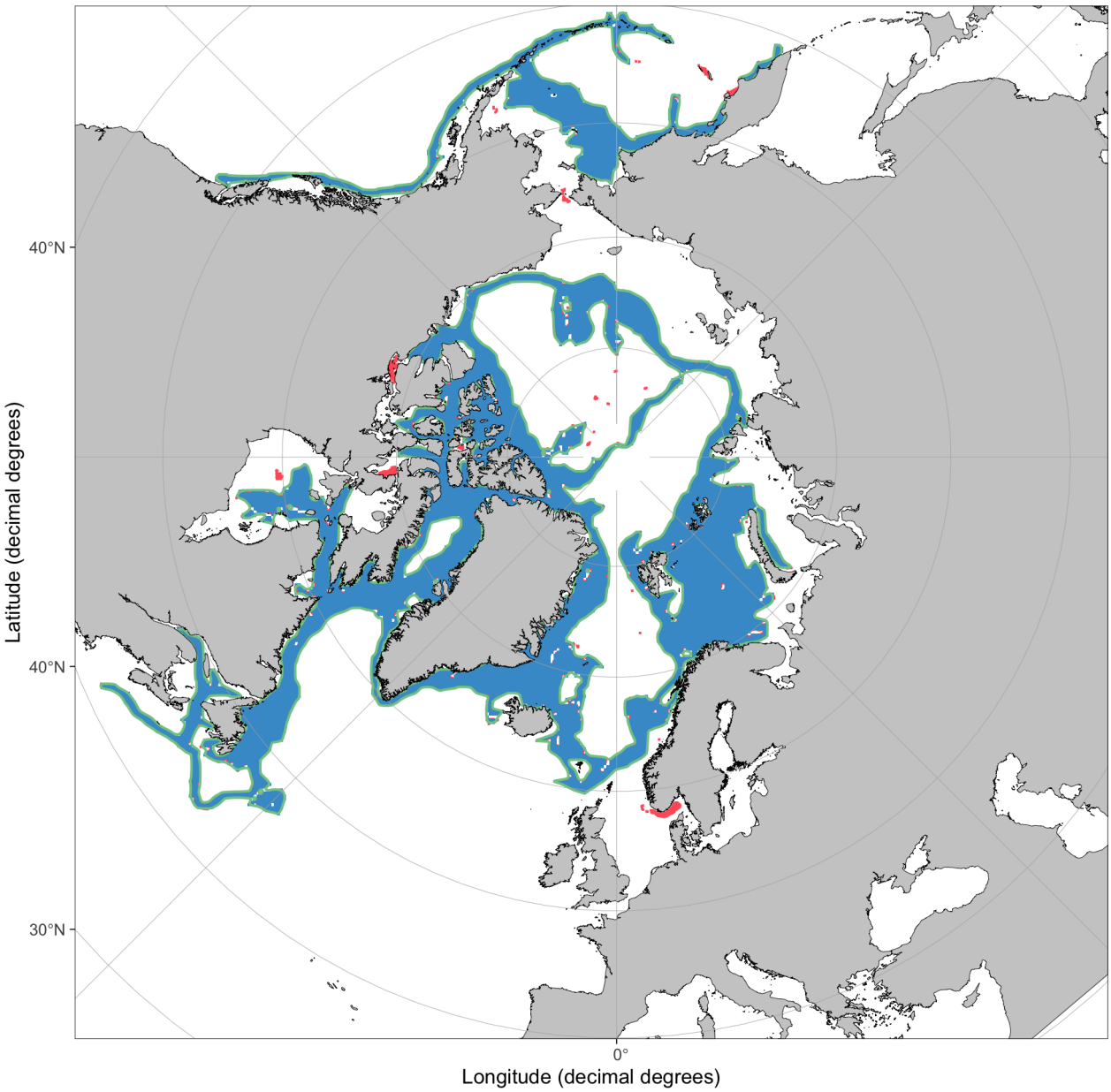
The suitable habitats are estimated from the habitat spaces. The package uses NEMO data for 2000-2009 as default (write `NEMOdata` to see the dataset).

```
## Fit the suitable habitat models

mod1 <- suitable.habitat(tdsp[[1]])
mod2 <- suitable.habitat(tdsp[[2]])

cowplot::plot_grid(plot(mod1) + ggplot2::ggtitle(names(tdsp[1])),
                   plot(mod2) + ggplot2::ggtitle(names(tdsp[2]))
)
```



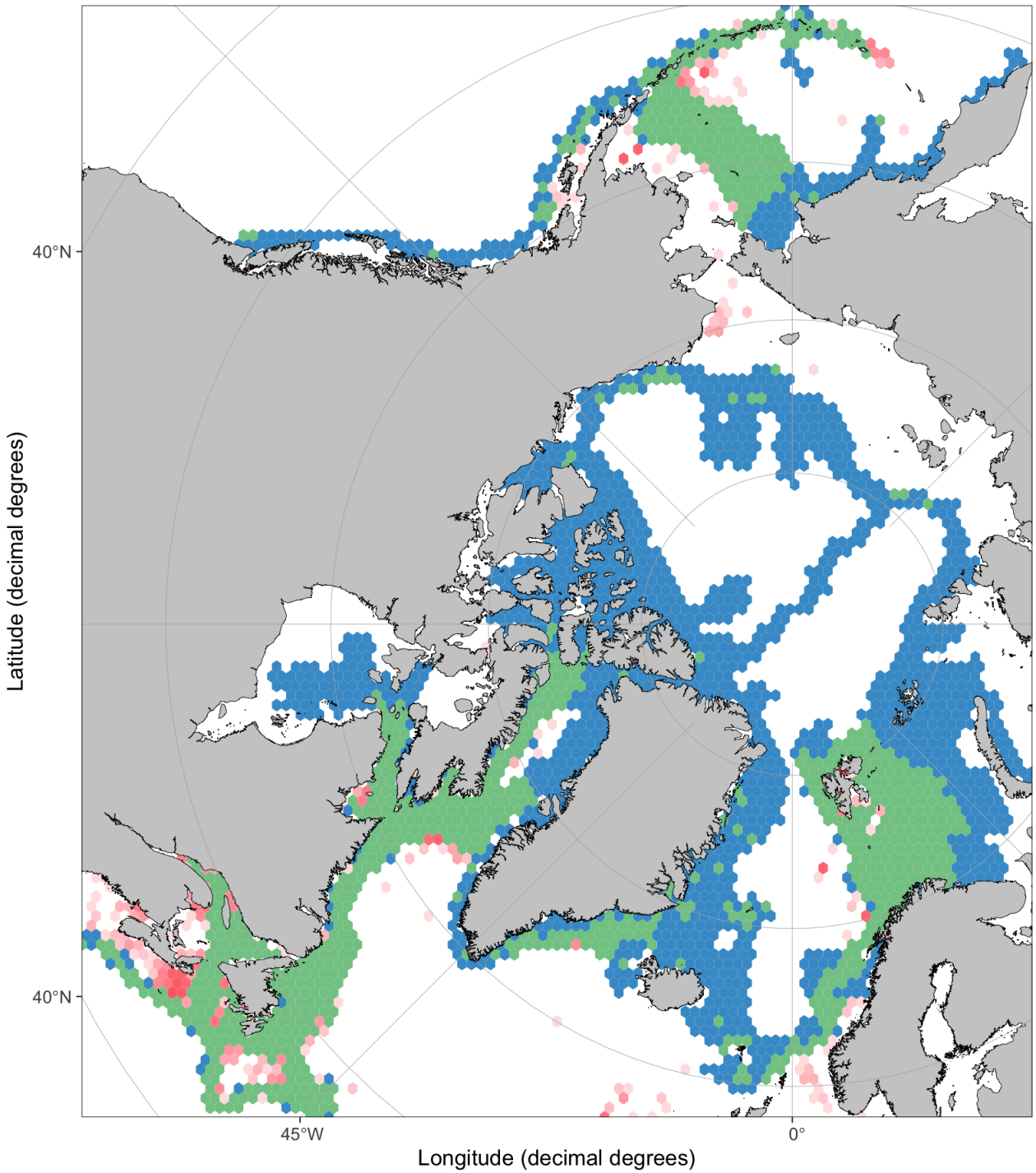

Since the TD-spaces were different for the Atlantic and Pacific sides, they were applied to their respective sides in a combined model. Note that the geographic binning resolution (`res` parameter) influences the connectivity of habitat patches and has to be optimized for each dataset together with the `drop.crumbs` parameter (minimum area of disconnected patches).

```
## Combine the models

mod <- combine.models(mod1 = mod1, mod2 = mod2,
                      y.breaks = c("<2.5e6", ">=2.5e6"),
                      res = 350, drop.crumbs = 3e4
)

plot(mod)
```



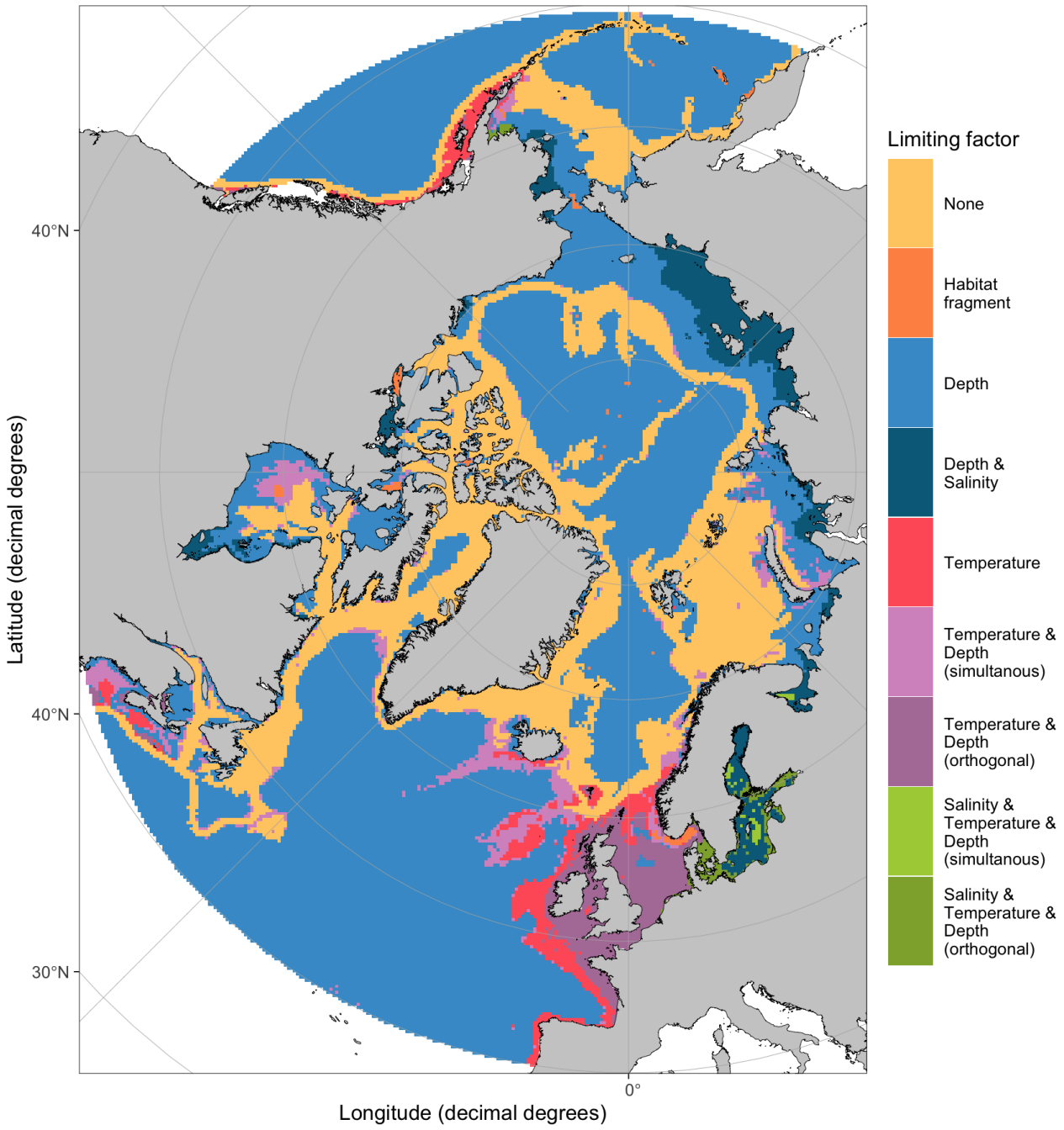

The model fit can be then evaluated using a dataset of spatial point observations. The underlying functions use similar binning than the suitable habitat model fed into the function.

```
## Model fit

modfit <- model.fit(mod = mod,
                    pts = sp::SpatialPoints(pres_dt[,c("lon.proj", "lat.proj")],
                                            proj4string = sp::CRS("+init=epsg:3995")
                    )
)

plot(modfit)
```




Limiting factors can be plotted using the suitable habitat model object.

```
## Limiting factors

plot(mod, type = "lim.fact")
```


Limiting factor
None
Habitat fragment
Depth
Depth & Salinity
Temperature
Temperature & Depth (simultanous)
Temperature & Depth (orthogonal)
Salinity & Temperature & Depth (simultanous)
Salinity & Temperature & Depth (orthogonal)
40°N
40°N
30°N
0°
Latitude (decimal degrees)
Longitude (decimal degrees)

## SI text 2. The NEMO model

We used the Nemo-NAA10km configuration newly developed by the Institute of Marine Research, Norway. Nemo-NAA10km is a regional model for the Arctic and the North Atlantic Oceans based on the NEMO 3.6 ocean engine (Madec and NEMO system team, 2015) with the LIM3 sea-ice model (Vancoppenolle *et al.*, 2009). It presents many similarities with the Nemo-Nordic regional ocean model (Hordoir *et al.*, 2019). Nemo-NAA10km covers the North Atlantic Ocean above a latitude of approximately 39° N, the Arctic Ocean, and part of the North Pacific Ocean to ensure a proper representation of the flow-through Bering Strait (Figures S5 and S6).

The purpose of Nemo-NAA10km is to get a deeper understanding of the effects of climate change on the Arctic Ocean hydrography. To achieve this, it is important that the model can reproduce recent observed changes, such as the Atlantification (Lind *et al.*, 2018), without forcing the model output to observational data. Therefore, in contrast to many regional Arctic Ocean models, Nemo-NAA10km does not use any form of restoring. Instead, it has been designed so that the Arctic freshwater content and temperature follow natural seasonal and inter-annual variabilities, by tuning vertical and horizontal mixing rates in a specific way for freshwater to leave the Arctic basin. Specifically, two types of freshwater circulation are represented: the Arctic estuarine circulation, and the coastally trapped circulation around the coasts of Greenland and Canada. The Arctic estuarine circulation can be properly represented if the halocline depth is realistic, which is achieved through a simple tuning of the background vertical diffusivity following Nguyen *et al.* (2009). Setting the Arctic halocline at the right depth ensures that the amount of available potential energy that drives the Arctic estuarine circulation is high enough. For a given amount of pure freshwater one can demonstrate that this available potential energy is proportional to the depth at which it is mixed.

Freshwater eventually leaves the Arctic Ocean through the East Greenland Current or the Canadian Archipelago. In each of these cases, the freshwater dynamics are similar to that of typical coastally trapped freshwater plumes, which have a limited mixing with the surrounding saltier water masses. These freshwater masses eventually mix with warmer/saltier waters carried by the Gulf Stream or can be advected southward in the Sub-Tropical gyre. A substantial amount is also advected southward by the Atlantic meridional overturning circulation (AMOC). Evidence shows that the horizontal diffusivity is much larger in the Sub-Tropical gyre than in the Northern Atlantic (Zhurbas and Oh, 2004), and so is the lateral diffusivity in Nemo-NAA10km: it is set at high value of 2000 $m^2$ $s^{-1}$ up to a latitude of 50° N, and at a value of only a few $m^2$ $s^{-1}$ above this latitude. Several tests are being made on this latest diffusivity value as this article is being written, but experiments show that these settings allow for the Arctic and North Atlantic freshwater content to stay stable during a long term simulation.

A long term simulation was conducted from 1958 to 2015 and uses the NFS 5.2 forcing dataset (Brodeau *et al.*, 2010). Open boundary conditions are based on the GLORYS re-analysis (http://marine.copernicus.eu) which exists from 1993, and on which we computed a climatology for the years of the simulation before 1993. Tidal forcing is based on the model described by Egbert and Erofeeva (2002) at the open boundary conditions, and tidal potential built-in the NEMO ocean engine is also used. A full assessment of the model results and their accuracy is to be published soon, but a recent analysis shows Nemo-NAA10km can reproduce the observations made by (Lind *et al.*, 2018).

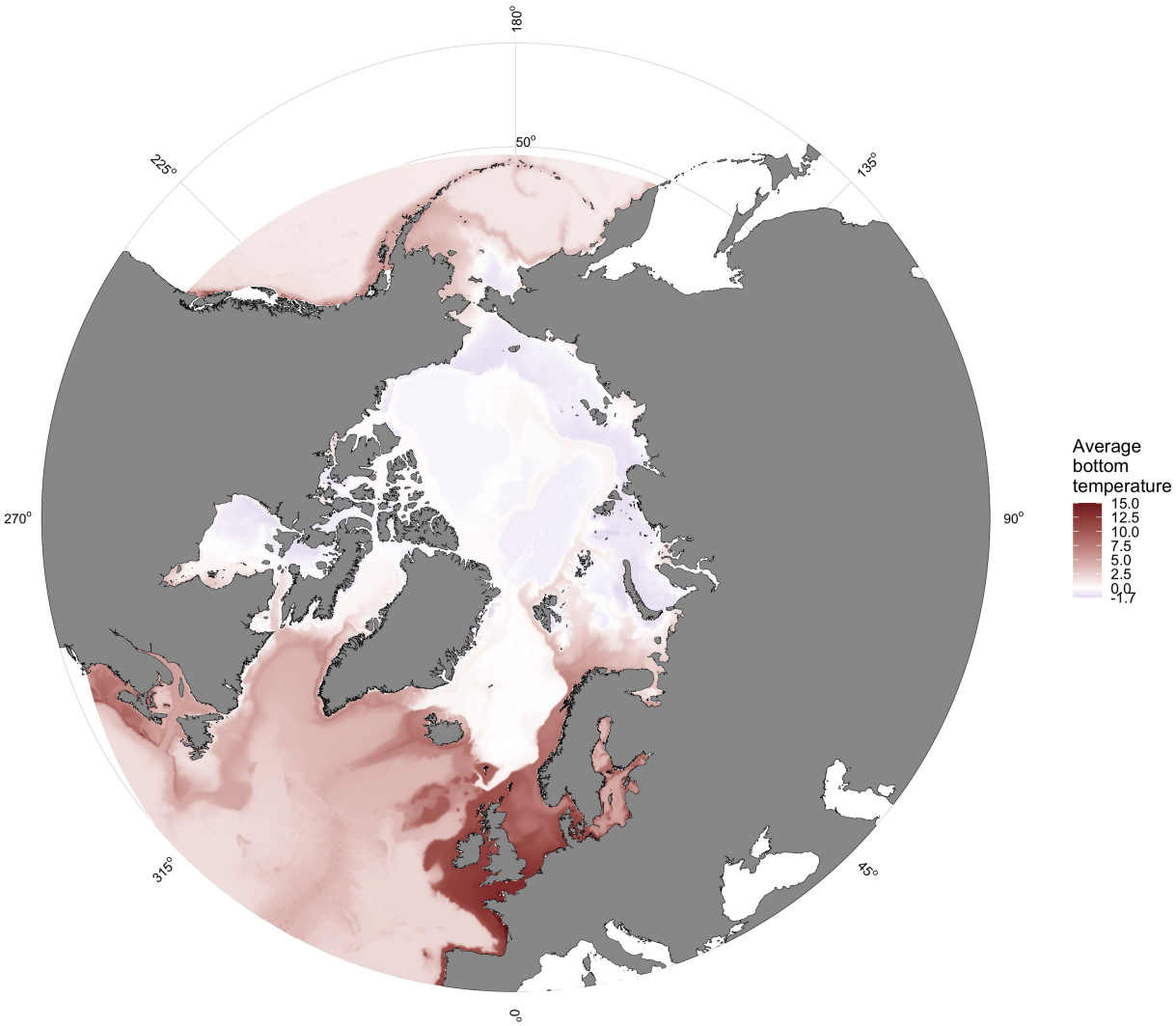



Figure S5: Mean bottom temperature from the oceanographic NEMO model used for geographic projections of habitat spaces to acquire the habitat models.

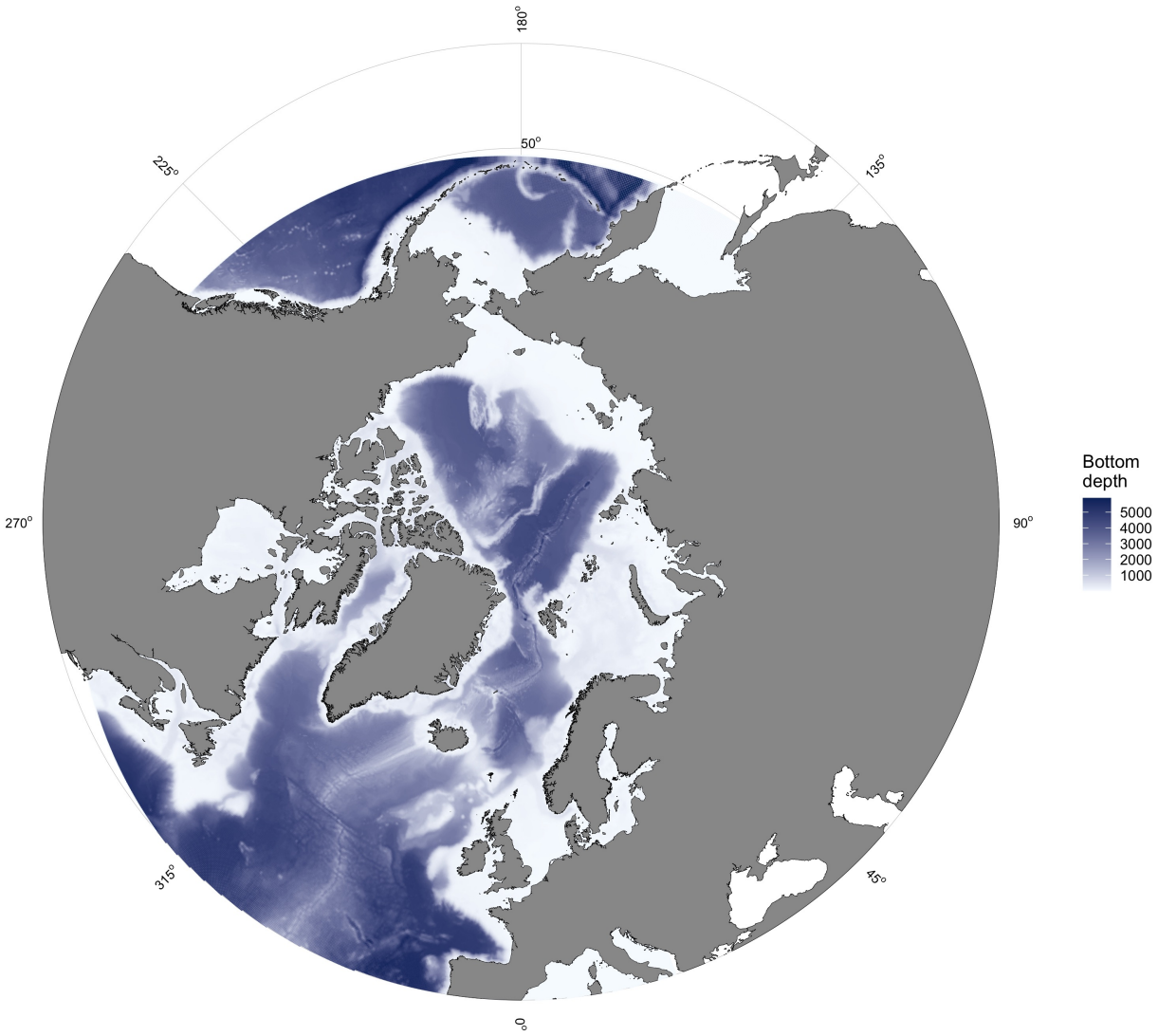



Figure S6: Bottom depth from the oceanographic NEMO model used for geographic projections of habitat spaces to acquire the habitat models.

## References


Araújo, M. B., and Peterson, A. T. 2012. Uses and misuses of bioclimatic envelope modeling. Ecology, 93: 1527–1539. http://doi.wiley.com/10.1890/11-1930.1.

Brodeau, L., Barnier, B., Treguier, A.-M., Penduff, T., and Gulev, S. 2010. An ERA40-based atmospheric forcing for global ocean circulation models. Ocean Modelling, 31: 88–104. https://linkinghub.elsevier.com/retrieve/pii/S1463500309002017.

Chacón, J. E., and Duong, T. 2018. Multivariate kernel smoothing and Its applications. CRC Press, Boca Raton, FL, USA. http://www.mvstat.net/mvksa/mvksa.pdf.

Domenici, P., Steffensen, J. F., and Marras, S. 2017. The effect of hypoxia on fish schooling. Philosophical Transactions of the Royal Society B: Biological Sciences, 372: 20160236. https://royalsocietypublishing.org/doi/10.1098/rstb.2016.0236.

Duong, T. 2019. ks: Kernel Smoothing. R package version 1.11.5. https://cran.r-project.org/package=ks.

Egbert, G. D., and Erofeeva, S. Y. 2002. Efficient Inverse Modeling of Barotropic Ocean Tides. Journal of Atmospheric and Oceanic Technology, 19: 183–204.

Enzor, L. A., Zippay, M. L., and Place, S. P. 2013. High latitude fish in a high CO2 world: Synergistic effects of elevated temperature and carbon dioxide on the metabolic rates of Antarctic notothenioids. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology, 164: 154–161. https://linkinghub.elsevier.com/retrieve/pii/S1095643312004370.

Fry, F. E. J., and Hart, J. S. 1948. The Relation of Temperature to Oxygen Consumption in the Goldfish. Biological Bulletin, 94: 66–77. Marine Biological Laboratory. http://www.jstor.org/stable/1538211.

Harper, A. A., Macdonald, A. G., Wardle, C. S., and Pennec, J.-P. 1987. The pressure tolerance of deep sea fish axons: Results of challenger cruise 6B/85. Comparative Biochemistry and Physiology Part A: Physiology, 88: 647–653. https://linkinghub.elsevier.com/retrieve/pii/0300962987906773.

Hordoir, R., Axell, L., Höglund, A., Dieterich, C., Fransner, F., Gröger, M., and Liu, Y. *et al.* 2019. Nemo-Nordic 1.0: a NEMO-based ocean model for the Baltic and North seas – research and operational applications. Geoscientific Model Development, 12: 363–386. https://www.geosci-model-dev.net/12/363/2019/.

Hutchinson, G. E. 1957. Concluding Remarks. Cold Spring Harbor Symposia on Quantitative Biology. *In* Cold spring harbor symposia on quantitative biology, pp. 415–427. http://symposium.cshlp.org/cgi/doi/10.1101/SQB.1957.022.01.039.

Kultz, D. 2015. Physiological mechanisms used by fish to cope with salinity stress. Journal of Experimental Biology, 218: 1907–1914. http://jeb.biologists.org/cgi/doi/10.1242/jeb.118695.

Lind, S., Ingvaldsen, R. B., and Furevik, T. 2018. Arctic warming hotspot in the northern Barents Sea linked to declining sea-ice import. Nature Climate Change, 8: 634–639. https://doi.org/10.1038/s41558-018-0205-y.

Madec, G., and NEMO system team. 2015. NEMO Ocean Engine, Version 3.6 Stable. IPSL, http://www.nemo-ocean.eu/.

Nguyen, A. T., Menemenlis, D., and Kwok, R. 2009. Improved modeling of the Arctic halocline with a subgrid-scale brine rejection parameterization. Journal of Geophysical Research, 114: C11014. http://doi.wiley.com/10.1029/2008JC005121.

Patra, R. W., Chapman, J. C., Lim, R. P., Gehrke, P. C., and Sunderam, R. M. 2015. Interactions between water temperature and contaminant toxicity to freshwater fish. Environmental Toxicology and Chemistry, 34: 1809–1817. https://setac.onlinelibrary.wiley.com/doi/abs/10.1002/etc.2990.

Pörtner, H. O. 2002. Climate variations and the physiological basis of temperature dependent biogeography: systemic to molecular hierarchy of thermal tolerance in animals. Comparative biochemistry and physiol-

ogy. Part A, Molecular & integrative physiology, 132: 739–61. http://www.ncbi.nlm.nih.gov/pubmed/12095860.

Sebert, P. 2002. Fish at high pressure: a hundred year history. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology, 131: 575–585. https://linkinghub.elsevier.com/retrieve/pii/S1095643301005098.

Soberón, J., and Nakamura, M. 2009. Niches and distributional areas: Concepts, methods, and assumptions. Proceedings of the National Academy of Sciences of the United States of America, 106: 19644–19650.

Vancoppenolle, M., Fichefet, T., Goosse, H., Bouillon, S., Madec, G., and Maqueda, M. A. M. 2009. Simulating the mass balance and salinity of Arctic and Antarctic sea ice. 1. Model description and validation. Ocean Modelling, 27: 33–53.

Vihtakari, M. 2021. SuitableHabitat: Estimate Habitat Spaces for Geographic Suitable Habitat Models. R package version 0.4.0. https://github.com/MikkoVihtakari/SuitableHabitat.

Youcef, W. A. I. T., Lambert, Y., and Audet, C. 2013. Spatial distribution of Greenland halibut Reinhardtius hippoglossoides in relation to abundance and hypoxia in the estuary and Gulf of St . Lawrence. Fisheries Oceanography, 22: 41–60. https://onlinelibrary.wiley.com/doi/abs/10.1111/fog.12004.

Zhurbas, V., and Oh, I. S. 2004. Drifter-derived maps of lateral diffusivity in the Pacific and Atlantic Oceans in relation to surface circulation patterns. Journal of Geophysical Research, 109: C05015. http://doi.wiley.com/10.1029/2003JC002241.