RESEARCH PAPER

Species richness in North Atlantic fish: Process concealed by pattern

Henrik Gislason[1] 🆔 | Jeremy Collie[2] | Brian R. MacKenzie[1] | Anders Nielsen[1] |
Maria de Fatima Borges[3] | Teresa Bottari[4,5] | Corina Chaves[3] | Andrey V. Dolgov[6,7,8] |
Jakov Dulčić[9] | Daniel Duplisea[10] | Heino O. Fock[11] | Didier Gascuel[12] |
Luís Gil de Sola[13] | Jan Geert Hiddink[14] | Remment ter Hofstede[15] | Igor Isajlović[9] |
Jónas Páll Jonasson[16] | Ole Jørgensen[1] | Kristján Kristinsson[16] |
Gudrun Marteinsdottir[17] | Hicham Masski[18] | Sanja Matić-Skoko[9] | Mark R. Payne[1] |
Melita Peharda[9] | Jakup Reinert[19] | Jón Sólmundsson[16] | Cristina Silva[3] |
Lilja Stefansdottir[18] | Francisco Velasco[20] | Nedo Vrgoč[9]

[1]DTU Aqua, Technical University of Denmark, Lyngby, Denmark

[2]Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island

[3]Instituto Português do Mar e da Atmosfera (IPMA), Lisboa, Portugal

[4]Institute for Biological Resources and Marine Biotechnology (IRBIM), National Research Council (CNR) Section of Messina, Messina, Italy

[5]Stazione Zoologica Anton Dohrn, Centro Interdipartimentale della Sicilia, Messina, Italy

[6]Polar branch of Russian Federal Research Institute of Fisheries and Oceanography (PINRO named after N.M.Knipovich), Murmansk, Russia

[7]Department of Biology and Aquatic Biological Resources, Murmansk State Technical University, Murmansk, Russia

[8]Laboratory of Engineering Surveys and Environmental Management, Tomsk State University, Tomsk, Russia

[9]Institute of Oceanography and Fisheries, Split, Croatia

[10]Fisheries and Oceans Canada, Institut Maurice-Lamontagne, Mont-Joli, Quebec, Canada

[11]Thuenen Institute of Sea Fisheries, Bremerhaven, Germany

[12]Agrocampus Ouest, Universite Europeenne de Bretagne, Rennes, France

[13]Instituto Español de Oceanografía, Málaga, Spain

Abstract

Aim: Previous analyses of marine fish species richness based on presence-absence data have shown changes with latitude and average species size, but little is known about the underlying processes. To elucidate these processes we use metabolic, neutral and descriptive statistical models to analyse how richness responds to maximum species length, fish abundance, temperature, primary production, depth, latitude and longitude, while accounting for differences in species catchability, sampling effort and mesh size.

Data: Results from 53,382 bottom trawl hauls representing 50 fish assemblages.

Location: The northern Atlantic from Nova Scotia to Guinea.

Time period: 1977–2013.

Methods: A descriptive generalized additive model was used to identify functional relationships between species richness and potential drivers, after which nonlinear estimation techniques were used to parameterize: (a) a 'best' fitting model of species richness built on the functional relationships, (b) an environmental model based on latitude, longitude and depth, and mechanistic models based on (c) metabolic and (d) neutral theory.

Results: In the 'best' model the number of species observed is a lognormal function of maximum species length. It increases significantly with temperature, primary production, sampling effort, and abundance, and declines with depth and, for small species, with the mesh size in the trawl. The 'best' model explains close to 90% of the deviance and the neutral, metabolic and environmental models 89%. In all four models, maximum species length and either temperature or latitude account for more than half of the deviance explained.

Main conclusions: The two mechanistic models explain the patterns in demersal fish species richness in the northern Atlantic almost equally well. A better understanding

[14]School of Ocean Sciences, Bangor University, Menai Bridge, Anglesey, United Kingdom

[15]Van Oord Dredging and Marine Contractors, Rotterdam, The Netherlands

[16]Marine and Freshwater Research Institute, Reykjavík, Iceland

[17]Institute of Life and Environmental Sciences, University of Iceland, Reykjavik, Iceland

[18]Institut National de Recherche Halieutique, Casablanca, Morocco

[19]Faroe Marine Research Institute, Tórshavn, Faroe Islands

[20]InstitutoEspañol de Oceanografía, Santander, Spain

Correspondence
Henrik Gislason, DTU Aqua, Technical University of Denmark, Kemitorvet, Building 202, DK2800 Lyngby, Denmark.
Email: hg@aqua.dtu.dk

of the underlying drivers is likely to require development of dynamic mechanistic models of richness and size evolution, fit not only to extant distributions, but also to historical environmental conditions and to past speciation and extinction rates.

KEYWORDS

abundance, biodiversity, density, marine fish, species size, temperature

1 | INTRODUCTION

Although much has been learned about the richness and distribution of marine species, a mechanistic understanding of the processes responsible for generating and maintaining species richness over evolutionary time-scales remains elusive. There is no generally accepted theory to explain the spatial distribution of marine species richness and no general understanding of why some species are more abundant than others (Fine, 2015). This lack of understanding is somewhat surprising. Strong latitudinal gradients in species richness are observed at global and regional scales and these often correlate significantly with environmental variables and life-history traits. Hillebrand (2004) conducted a meta-analysis of gradients in marine biodiversity and found significant relationships between marine species richness, latitude, and species size, while Tittensor et al. (2010) found water temperature to be the main environmental predictor of species richness across a number of marine taxonomic groups. Why latitude, temperature and species size are important is unclear, but size and maximum body size influence the trophic position, mortality, growth, and reproduction of many marine species (Andersen et al., 2016), temperature affects their metabolism and food uptake (Gillooly, Brown, West, Savage & Charnov, 2001), and latitude determines the amplitude of the seasonal changes in solar energy input affecting primary production, average temperature, and annual temperature range (Cullen, Franks, Karl & Longhurst, 2002).

Bony fish and elasmobranchs are among the best taxonomically resolved groups of marine animals and are therefore well suited for studies of marine species richness. Estimates suggest that on a global scale around 79% of the species have now been described (Mora, Tittensor & Myers, 2008) and very few species have been declared extinct due to human activities (Davies & Baum, 2012). However, most inventories of fish species richness are based on single recordings of individuals with little consideration of differences in individual density and sampling effort. Including density and sampling effort is important for at least two reasons. The number of species recorded is known to depend statistically on the number of individuals and number of samples examined (Gotelli & Colwell, 2001), and high-density areas may have higher species richness because they harbour more individuals able to maintain a higher number of viable populations (Brown, 2014). Based on species inventories, MacPherson and Duarte (1994) found fish species richness and average maximum fish species size to increase with depth and decline with latitude in the northern Atlantic and Fisher, Franks, and Leggett (2010) found the geometric mean fish species size to covary with species richness. While Blowes, Belmaker, and Chase (2017) found the latitudinal change in reef fish richness to scale with abundance, no one has so far analysed how species richness of marine fish found on soft or sandy bottoms is related to density or abundance on a basin-wide scale.

To understand how fish species richness in different fish communities is related to density or abundance, species length and

FIGURE 1 Pie charts showing the locations of the surveys and the relative number of species recorded in each of the maximum length groups indicated in the lower right-hand corner of the map (plotted with the R package marmap (Pante & Simon-Bouhet, 2013).
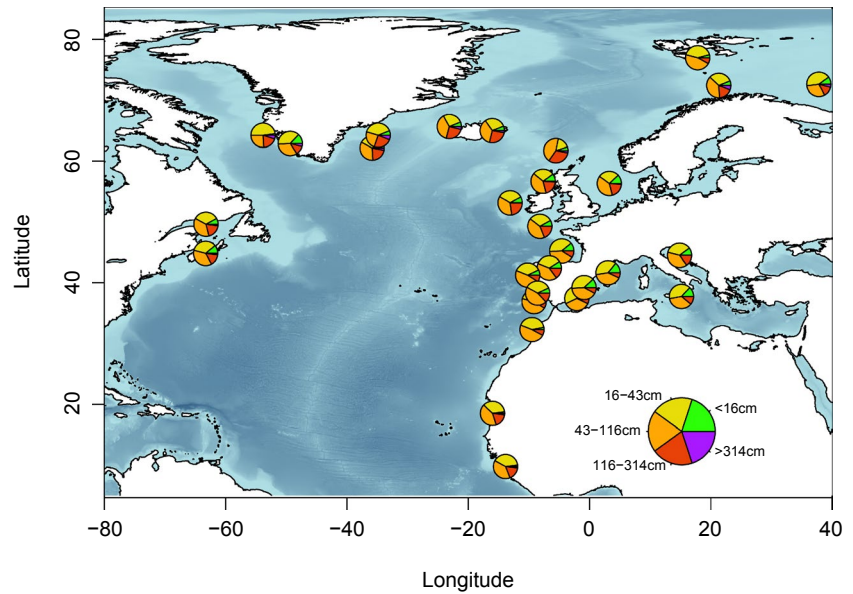
environmental conditions, we analyse an extensive dataset, generated by collating results from 31 standardized bottom trawl surveys from the continental shelves of the northern Atlantic and adjacent areas (Figure 1). Our analysis is based on 123 million individual demersal or benthopelagic fish caught in 53 thousand hauls taken within a total survey area of 3.1 million km². Bottom trawl surveys are often stratified to account for spatial or depth related differences in fish assemblage composition and density. We retain the stratification used in the surveys, correct for differences in catchability, and further stratify species into log maximum species length intervals. Using a generalized additive model (GAM) to identify significant variables and relationships, we construct a 'best' descriptive model of the number of species caught per log maximum species length interval and survey stratum by transforming the significant relationships identified by the GAM into functional relationships. We also fit an environmental model to the data in which latitude, longitude, depth, total catch and mesh-size are used as independent variables without invoking any biological hypotheses. Using the two descriptive models as reference points we investigate how well mechanistic equilibrium models of species richness based on metabolic (Allen, Brown & Gillooly, 2002; Allen & Gillooly, 2007) and neutral theory (Hubbell, 2001) fit the survey data. Both theories explain the present difference in species richness among fish communities from individual density or abundance, and from fundamental evolutionary processes such as speciation, dispersal, and extinction. Recently, they have been combined and used to simulate the latitudinal gradient in species richness in the oceans (Tittensor & Worm, 2016; Worm & Tittensor, 2018).

In brief, the basic assumption of metabolic theory is that temperature enhances species richness by increasing mutation rates and reducing generation times, while extinction rates are inversely related to the average density per species. In contrast to metabolic theory, neutral theory includes a spatial component and assumes that richness is determined by local abundance and random

extinctions among functionally equivalent species counterbalanced by immigration from a surrounding meta-community where speciation takes place. Functionally equivalent species are defined as species that share the same probabilities of death and reproduction (see Supporting Information Appendix S1 for further information on the two models). Because natural mortality and reproductive output depend on body size in fish, we follow Reuman, Gislason, Barnes, Mélin, and Jennings (2014) and assume that functional equivalence primarily applies for species of similar maximum length. We therefore treat each maximum species length group separately. Comparing the results from the neutral and metabolic models with the two descriptive models, our aim is to elucidate the mechanisms behind the richness differences we observe across fish communities in the northern Atlantic.

2 | METHODS

2.1 | Survey data

Average catch in number of individuals per species and haul was provided from 31 scientific bottom trawl surveys. The time period from which data were obtained from each survey was selected to provide temporal overlap between the surveys and as long a time period from each survey as feasible to minimize the influence of random fluctuations in recruitment and population abundance. Surveys with less than eight years of data were hence excluded. Although the earliest trawl hauls were taken in 1977 and the most recent in 2013, the period from 2001 to 2006 was covered by all surveys. Slightly more than half of the surveys took place in the period from October to March, a third in the period from April to September, and the remaining surveys included hauls obtained throughout the year (see Appendix: Data sources, and Supporting Information Appendix S2: Table S2.1). Different bottom trawls were used in the surveys. Cod-end mesh sizes ranged from 13 to 40 mm, horizontal trawl openings

(wing spread) from 13 to 28 m, vertical openings from 1.9 to 7 m and towing speeds from 3 to 4.5 knots. Many of the surveys used a stratified random sampling design to account for spatial and depth related differences in species composition. We retained the major strata used in the surveys, providing us with richness and density data from 50 different strata. The average depth in these strata ranged from 28 to 950 m.

2.2 | Environmental data

Sea surface temperature, average temperature in the upper 200 m of the water column and near bottom temperatures (Kelvin) were obtained from the World Ocean Atlas 2013 (Locarnini et al., 2013) based on decadal average temperature at 0.25° resolution covering the period 1955–2012 for annual, boreal summer (Jul–Sep) and boreal winter (Jan–Mar). Bottom temperatures were defined as the temperature in the layer closest to the bottom. Spatial averages were calculated for each survey stratum, and the seasonal amplitude calculated as the difference between summer and winter values. Estimates of depth integrated pelagic net primary production (npp, $gC/m^2/year$) based on the satellite-derived vertically generalized production model (VGPM; Behrenfeld & Falkowski, 1977) were downloaded from http://www.science.oregonstate. edu/ocean.productivity at 1/12 degree monthly resolution for the period 2002–2012, from which estimates of mean annual npp were derived for each survey area. Latitude and longitude were calculated as the average of the minimum and maximum coordinates of each survey. Average depth was calculated as the midpoint of the depth range of each stratum (see Supporting Information Appendix S2: Table S2.1).

2.3 | Fish species data

Among the fish taxa recorded some individuals had not been identified to species. If possible, we allocated these individuals to species, assuming that their relative species composition would be identical to that of the individuals identified within the same survey stratum, and family or genus. Where no species from the family or genus had been identified in a stratum, the family or genus name was retained. Information about the maximum length of each species was downloaded from FishBase (Froese & Pauly, 2016) and used to bin the observations into 11 log maximum length intervals of equal width (from now on denoted log maximum length groups). In 1% of the species records no maximum species length was available. These records were excluded from further calculations.

To estimate absolute fish density and abundance in a given stratum or area we first calculated swept area density for each species. This was done by dividing the average number of individuals caught per haul by the average area swept per haul, estimated by multiplying the wing spread of the trawl by the average distance covered per haul. Swept area abundance was calculated by multiplying swept area density by the size of the survey area. Swept area density and abundance can be converted to absolute density and abundance if catchability is known. Catchability, the fraction of the population in the path of the trawl that is retained and caught by the gear, can be estimated by dividing the swept area estimate of abundance by the absolute abundance provided by a stock assessment. Catchability is likely to differ between areas and species and depends on a number of factors including the properties of the trawl and species-dependent traits such as the size, behaviour and distribution of the individuals (Arreguín-Sánchez, 1996; Walker, Maxwell, Le Quesne & Jennings, 2017). To account for differences in horizontal and vertical distribution we sorted the species into: (a) species whose main distribution is outside the main depth range of the surveys (species mainly occurring in the infra-littoral zone and bathydemersal or bathypelagic species found mainly at more than 200 m of depth), and species whose main distribution is inside the main depth range of the surveys, but either (b) mostly occur on either untrawlable grounds (species that are mainly found associated with reefs or in rocky areas), (c) are likely to have a low catchability (species that bury in the sediment and pelagic species), or (d) are likely to be regularly retained by the survey gear when available (species resting on the seabed, species found close to but not on the seabed, and midwater species with some bottom contact).

We were able to identify 56 cases where catchability could be derived for the species, time period, and area covered by the survey data (see Supporting Information Appendix S3 and Table S3.1). No catchability estimates could be derived for bathypelagic and bathydemersal stocks, and few estimates could be obtained for infra-littoral species, for species mainly found associated with reefs or in rocky areas, and for burying and pelagic species; species that are likely to be under-sampled by the trawl surveys. The average catchability of these species was only 0.05, while the average catchabilities of the species in group four were 0.34, 1.04 and 0.52 for species that were resting on the seabed, found close to the seabed, or found in midwater, respectively. Note that for some of the species found close to the seabed the estimated catchability exceeded 1.0, probably due to their response to the herding effect of the bridles, sweeps, and doors of the trawl. Due to the few and low catchability estimates available for groups two and three, we decided to use only species from group (d) in our analysis. To extrapolate the 41 catchability estimates available for the 412 species in this group we fitted a log-linear mixed model to the estimates, using the vertical position of the species (resting on seabed, found close to but above seabed, or midwater with some bottom contact) as a fixed variable and species identity and survey area as random factors. Drawing samples at random from the resulting stochastic model we generated 1,000 estimates of catchability for each combination of species and survey stratum (see Supporting Information Appendix S3). The catchabilities were used to calculate average absolute density and abundance in each survey stratum for each of the species found in the surveys. Average absolute density and abundance were finally cumulated

across species within each log maximum length group and survey stratum producing 550 data points as input to the models.

To confirm that the richness of the species in group (d) had been reasonably well sampled by the surveys, we furthermore used the vegan package (Oksanen et al., 2015) to estimate the number of un-observed species in each survey stratum, and found that on average a minimum of 7–8% of the species in a particular stratum may not have been recorded. However, considering all of the species found across the surveys few species appear to have been missed (see Supporting Information Appendix S4: Table S4.1).

2.4 | Selection of independent variables

The number of species recorded in a survey stratum depends on the species richness in the stratum and the bias introduced by the sampling method. Using a bottom trawl to sample species richness is likely to generate a biased estimate of species richness because the number of species recorded depends on the number of individuals caught and identified (the species accumulation curve); the total area swept by the trawl (a measure of sampling effort); the vertical opening of the trawl (potentially influencing the catch of individuals and species trying to escape over the trawl); and its mesh-size (influencing the proportion of small individuals and species recorded in the catch). To account for the bias we included all four variables in the GAM model. We used the total area swept in each survey stratum rather than the total number of hauls to represent sampling effort because the average duration of the trawl hauls ranged from 15 min to 1 hr across surveys.

According to the metabolic and neutral models, richness should depend on temperature, species size, and either density or absolute abundance, but the size of the stratum (the species–area relationship), net primary production, annual temperature range, latitude, longitude and depth may also be important covariates. Temperature may influence richness by affecting fish metabolism, generation time and mutation rate, but varies seasonally depending on latitude, longitude and depth. Identifying the biologically relevant ambient temperature for a fish species is therefore difficult. Average sea surface temperature may be relevant for the pelagic eggs and larvae, average bottom temperature describes the average ambient temperature encountered by the juveniles and adults at the depth where they are caught by the survey trawls, and average temperature in the upper 200 m of the water column may represent the average temperature encountered during the entire life cycle.

We furthermore found more than a third of the pairwise correlations of the potential independent variables to be significant (see Supporting Information Appendix S2: Figure S2.2). Sea surface temperature, bottom temperature and water column temperatures were highly significantly correlated with each other and with both latitude and longitude, while the seasonal temperature range in the upper 200 m of the water column was significantly correlated to the seasonal temperature ranges near the bottom and at the surface. Net primary production decreased with latitude and increased with

temperature, with both correlations highly significant. The vertical opening of the gear was highly significantly correlated to both latitude and to all three temperatures, but not to longitude, reflecting that surveys in high latitudes generally use larger trawls with larger vertical openings than surveys in low latitudes. Total area swept and total stratum area were also highly significantly positively correlated, reflecting that more hauls typically had been taken in large survey strata than in small. Finally, catch in numbers, average abundance, and total swept area were significantly correlated.

2.5 | Identifying functional relationships

To find the 'best' descriptive model we used a GAM (Wood, 2006) to identify the functional form and error structure of the relationship between the number of species caught per log maximum length group and the independent variables using the R package mgcv v. 1.8.22 (Wood & Säfken, 2016). In the GAM the log of the expected mean number of species caught, $\mu_{i,j}$, in survey stratum i, maximum length group j, was described using:

$$\begin{aligned}
\log\left(\mu_{ij}\right) = {} & \alpha + s_1\left(\text{temp\_range}_i\right) + s_2\left(\text{temp}_i\right) + s_3\left(\text{abundance}_{ij}\right) \\
& + s_4\left(\text{depth}_i\right) + s_5\left(\text{npp}_i\right) + s_6\left(\text{asurv}_i\right) + s_7\left(\text{lml}_j\right) + s_8\left(\text{catch}_{ij}\right) \\
& + s_9\left(\text{aswept}_i\right) + s_{10}\left(\text{vertop}_i\right) + s_{11,j}\left(\text{mesh}_i\right)
\end{aligned} \tag{1}$$

where α is a proportionality constant; suffixes i and j signify survey stratum and maximum length group, respectively, temp_range_i is the intra-annual temperature range in the stratum (Kelvin); temp_i is temperature (Kelvin); abundance_{ij} is the average absolute abundance of fish of maximum length group j in stratum i; depth_i is depth (m); npp_i is annual net primary production (gC/m^2/year); asurv_i is the total stratum area (km^2); lml_j is the midpoint of the log maximum length group; catch_{ij} is the total number of fish caught in stratum i, maximum length group j over the time period of the survey; aswept_i is area swept by the survey trawl (km^2); mesh_i is mesh-size (mm); and vertop_i is the vertical opening of the trawl (m). The $s_1, \ldots, s_{10}$ are general spline smoothers, while $s_{11,j}$ denotes that for each log maximum length group, j, a separate spline smoother was applied to describe the effect of mesh-size on the number of species caught. The temp_i and temp_range_i variables were either sea surface, average upper 200 m water column or bottom temperature or were replaced by latitude, lat_i, and longitude, lon_i, when the effect of geographical location was examined, and abundance_{ij} was changed to density_{ij} to examine which of the two would provide the best fit.

We used thin plate regression splines with a basis dimension of four as smoothers and a log link. To account for the correlation between many of the independent variables, we analysed the effect of including these in separate model versions using residual plots and estimates of concurvity (a nonlinear analogue of multicollinearity) to select the best fitting parameter combinations, and the Akaike information criterion (AIC) to identify the most parsimonious model. Model terms were selected by backwards removal of insignificant variables, after which covariates generating an estimated concurvity

larger than .80 were sequentially removed to reduce variance inflation and avoid bias. Distributions of residuals were visually inspected for normality and plotted against each covariate to reveal heteroscedasticity. We compared models with Poisson and negative binomial error distributions, and found the two to provide an almost equally good fit to the data based on AIC values and comparisons of the observed and theoretically expected variance, where the importance of over-dispersion was assessed by dividing the sum of squared residuals by the sample size minus the number of parameters estimated (Hilbe, 2011). For the negative binomial model this produced a variance ratio of 0.94, confirming the appropriateness of a negative binomial assumption.

We simplified the GAM model and further reduced its AIC value by inserting the functional relationships indicated by the significant GAM smoothers (see Figure 3). To model the effect of temperature, we assumed that the relationship between species richness and temperature would follow the Arrhenius equation (Gillooly et al., 2001) and consequently used the inverse of temperature in the model. The functional relationships included logarithmic transformations of several of the other independent variables and the addition of a second-order polynomial to capture the change in log species richness with log maximum length. All log transformations used natural logarithms. Using log transformations meant either that zero observations had to be excluded, or that a small positive number had to be added to avoid having to calculate the log of zero. When zero individuals had been caught in a given stratum and log maximum length group, we therefore used the inverse of the total area swept in the stratum to provide a tentative estimate of the maximum density in the stratum and group. As evidenced by the residuals, this introduced a small bias in the fit (see Supporting Information Appendix S5: Figure S5.3). Because the neutral model cannot easily be linearized, we used nonlinear techniques to estimate the parameters of the four models presented below. This also allowed us to retain the zeros and removed the source of the bias in the linearized GAM model.

2.6 | Best descriptive model

The significant independent variables in the linearized GAM model were used to construct a 'best' nonlinear descriptive model of the number of species caught. The 'best' nonlinear model followed the simplified GAM equation and contained an Arrhenius expression where β_2, the 'activation energy of metabolism' (Gillooly et al., 2001) was divided by average water column temperature (Kelvin) multiplied by Boltzmann's constant, k (8.62×10^{-5} eV/K). It also contained total abundance, depth, annual net primary production and a quadratic log maximum length term, $\exp\left(\mathrm{lml}_j + \beta_7 \mathrm{lml}_j^2\right)$, to capture the unimodal relationship between species richness and log maximum length, as well as the catch in numbers, the area swept by the trawl, and a mesh-size/log maximum length interaction to account for sampling bias:

$$\mu_{i,j} = \alpha * \exp\left(\frac{-\beta_2}{k * \mathrm{temp}_i}\right) * \mathrm{abundance}_{i,j}^{\beta_3} * \mathrm{depth}_i^{\beta_4} * \mathrm{npp}_i^{\beta_5}$$
$$* \exp\left(\mathrm{lml}_j + \beta_7 \mathrm{lml}_j^2\right) * \mathrm{catch}_{i,j}^{\beta_8} * \mathrm{aswept}_i^{\beta_9} * \mathrm{mesh}_i^{\beta_{11,j}} \tag{2}$$

where α, the proportionality constant, subsumes the combined effects of the standardization of the Arrhenius expression to a reference temperature, and other pre-factors related to abundance, depth, net primary production, the maximum length term, catch in numbers, area swept, and mesh-size.

2.7 | Environmental model

The environmental model assumes that the number of species observed in survey stratum i log maximum length group j can be calculated from species richness, described by a simple function of latitude, longitude, depth and log maximum length, corrected for differences in catch in numbers, area swept and mesh-size:

$$\mu_{i,j} = \alpha * \mathrm{lat}_i^{\beta_0} * \mathrm{lon}_i^{\beta_1} * \mathrm{depth}_i^{\beta_4}$$
$$* \exp\left(\mathrm{lml}_j + \beta_7 \mathrm{lml}_j^2\right) * \mathrm{catch}_{i,j}^{\beta_8} * \mathrm{aswept}_i^{\beta_9} * \mathrm{mesh}_i^{\beta_{11,j}} \tag{3}$$

2.8 | Metabolic model

In the metabolic theory of ecology, temperature and body size influence the rate of per capita speciation in the same way as they influence metabolism (Gillooly & Allen, 2007; see Supporting Information Appendix S1). Combining absolute density with a per capita rate of speciation determined by maximum length and temperature provides the speciation rate. In the equilibrium situation, speciation is counterbalanced by extinction, assumed to decline linearly with the average density per species. We added the effect of differences in number of individuals caught, area swept and trawl mesh-size to the model of Segura et al. (2015) to describe the number of species caught:

$$\mu_{i,j} = \alpha * \exp\left(\frac{-\beta_2}{k * \mathrm{temp}_i}\right) * \mathrm{density}_{i,j}^{\beta_3} * \mathrm{ml}_j^{\beta_6} * \mathrm{catch}_{i,j}^{\beta_8} * \mathrm{aswept}_i^{\beta_9} * \mathrm{mesh}_i^{\beta_{11,j}} \tag{4}$$

where β_2 is the 'activation energy of metabolism' (Gillooly & Allen, 2007), k is Boltzmann's constant, ml_j is the median maximum length of the species in log maximum length group j, and α, the proportionality constant, accounts for the combined effects of the standardization of the Arrhenius expression to a reference temperature, as well as other pre-factors related to the density term, and to the maximum length, catch in numbers, area swept and mesh-size terms.

2.9 | Neutral model

According to the neutral theory of biodiversity and biogeography, the number of functionally equivalent species in a local community is determined by random extinctions caused by ecological drift,

counterbalanced by immigration of species from a larger surrounding meta-community where random speciation takes place (Hubbell, 2001; Rosindell, Hubbell & Etienne, 2011; see Supporting Information Appendix S1).

Following Reuman et al. (2014), we assume that species of similar log maximum length are functionally equivalent and model each log maximum length group separately, using the approximate formula derived by Etienne and Olff (2004) and Reuman et al. (2014) to describe the relative number of species in each survey stratum and log maximum length group. We also assume that the probability of immigration, λ, is independent of stratum area, but allow it to vary with log maximum length. To account for the effect of differences in the number of individuals examined, effort and sampling gear on the number of species caught, we add number of individuals caught, total area swept and mesh-size terms to the species richness model of Reuman et al. (2014) providing the following equation:

$$\mu_{i,j} \approx J_{M_j} * \left(\frac{v_i}{1-v_i} \right) * \log \left[1 - \frac{\lambda_j \log(\lambda_j)}{1-\lambda_j} * \left(\frac{abundance_{i,j}}{J_{M_j} * (v_i/(1-v_i))} \right) \right] \quad (5)$$
$$* catch_{i,j}^{\beta_8} * aswept_j^{\beta_9} * mesh_i^{\beta_{11j}}$$

where J_{M_j} is absolute abundance in log maximum length group j in the meta-community, $abundance_{i,j}$ is the absolute abundance of group j in the local community and v_i is the per capita speciation rate in area i. Note also that J_{M_j} and v_i are confounded in the $J_{M_j}(v_i/(1-v_i))$ term. However, as the speciation rate is likely to be very small, the term can be approximated by the fundamental biodiversity number, $\theta_{i,j} = J_{M_j} v_i$ (Rosindell et al., 2011). Because fish evolution is affected by temperature (Wright, Ross, Keeling, McBride & Gillman, 2011), we follow Tittensor and Worm (2016) and make v_i temperature dependent by adding the Arrhenius equation. Finally, we approximate the change in J_{M_j} with log maximum length by a quadratic term as found in the 'best' descriptive model and thus end up with:

$$J_{M_j} \left(\frac{v_{i,j}}{1-v_{i,j}} \right) \approx \theta_{i,j} = \alpha * \exp \left(lml_j + \beta_7 lml_j^2 \right) * \exp \left(\frac{-\beta_2}{k * temp_i} \right) \quad (6)$$

where α again is an overall proportionality constant. Hence

$$\mu_{i,j} = \theta_{ij} * \log \left[1 - \frac{\lambda_j \log(\lambda_j)}{1-\lambda_j} * \frac{abundance_{i,j}}{\theta_{ij}} \right] * catch_{i,j}^{\beta_8} * aswept_j^{\beta_9} * mesh_i^{\beta_{11j}}$$
$$(7)$$

where $j = 1 \ldots 11$, is log maximum length group, i is stratum and $abundance_{i,j}$ is the total number of individuals in stratum i group j estimated by multiplying the size of stratum i with the absolute density of fish in i belonging to log maximum length group j.

2.10 | Estimating model parameters

We used the nonlinear model fitting R package TMB (Kristensen, Nielsen, Berg, Skaug & Bell, 2015) to estimate the parameters of the four nonlinear models. Fitting each model to the number of species

observed we removed any insignificant variables, except if they were important for the theoretical underpinning of a model. We visually inspected the Pearson residuals of each model for normality and plotted them against each covariate to reveal potential heteroscedasticity. To compare the models we calculated AIC values (Burnham & Anderson, 2002), R^2 from observed and predicted number of species, and proportion of deviance explained. The latter was estimated by fixing the estimated scale parameter, κ, of the negative binomial distribution used in each of the models, comparing the difference in deviance between a saturated model (with one parameter for each of the 550 observations) and the actual model, to the difference in deviance between a saturated model and a model with only one parameter (Cameron & Windmeijer, 1996). To also illustrate how much of the overall deviance each model term explained, we consecutively replaced each of the independent variables by its overall average and calculated the relative increase in the proportion of deviance explained when the observations were used instead of the average. Having identified the four most parsimonious models, we examined their sensitivity to the uncertainty in the abundance and density data by fitting them to the 1,000 separate estimates of density and abundance obtained from the mixed effects catchability model, and calculated the mean and variance of the resulting parameter estimates. We
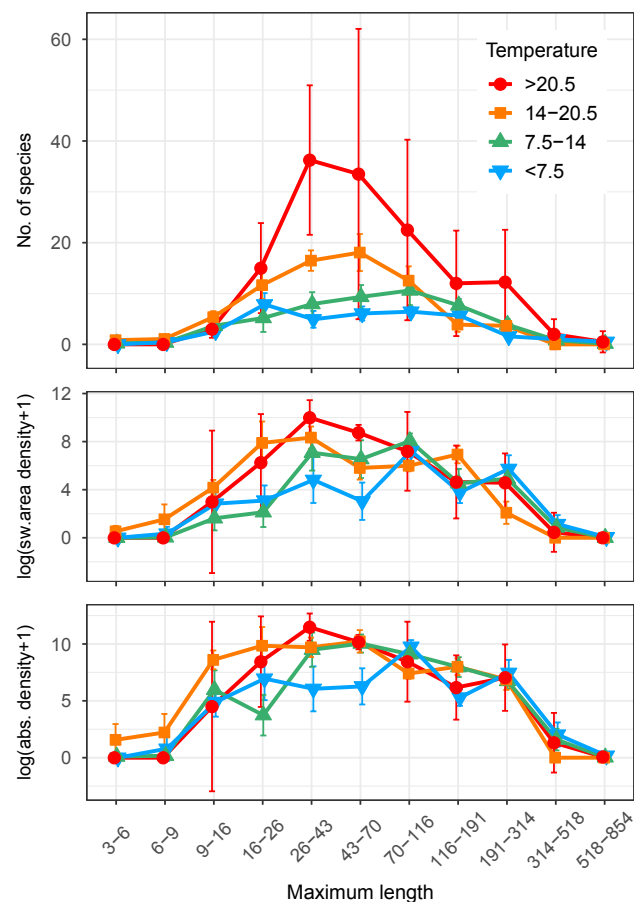
FIGURE 2 Average number of species, log swept (sw.) area density (no./km^2) and log absolute (abs.) density (no./km^2) ($\pm$95% confidence limits) versus maximum length (cm) in four different sea surface temperature intervals (°C).

plotted the proportion of the deviance explained by each of the model variables in the 1,000 runs, and used these to illustrate the sensitivity of our results to the uncertainty in the catchability estimates. All analyses were undertaken in R version 3.4.4 (R Core Team, 2019).

3 | RESULTS

3.1 | Observed number of species and density

The number of observed species, log average swept area density and log average absolute density follow almost symmetrical distributions

when plotted against log maximum length (Figure 2). As expected, the average number of species observed increases with temperature, while log average swept area density and log average absolute density change little except in areas with a mean annual sea temperature below 7.5 °C where the densities are significantly lower in the intermediate length range.

3.2 | GAM model

Fitting the GAM to the survey data reveals a strong and highly significant unimodal effect of log maximum length on log number of
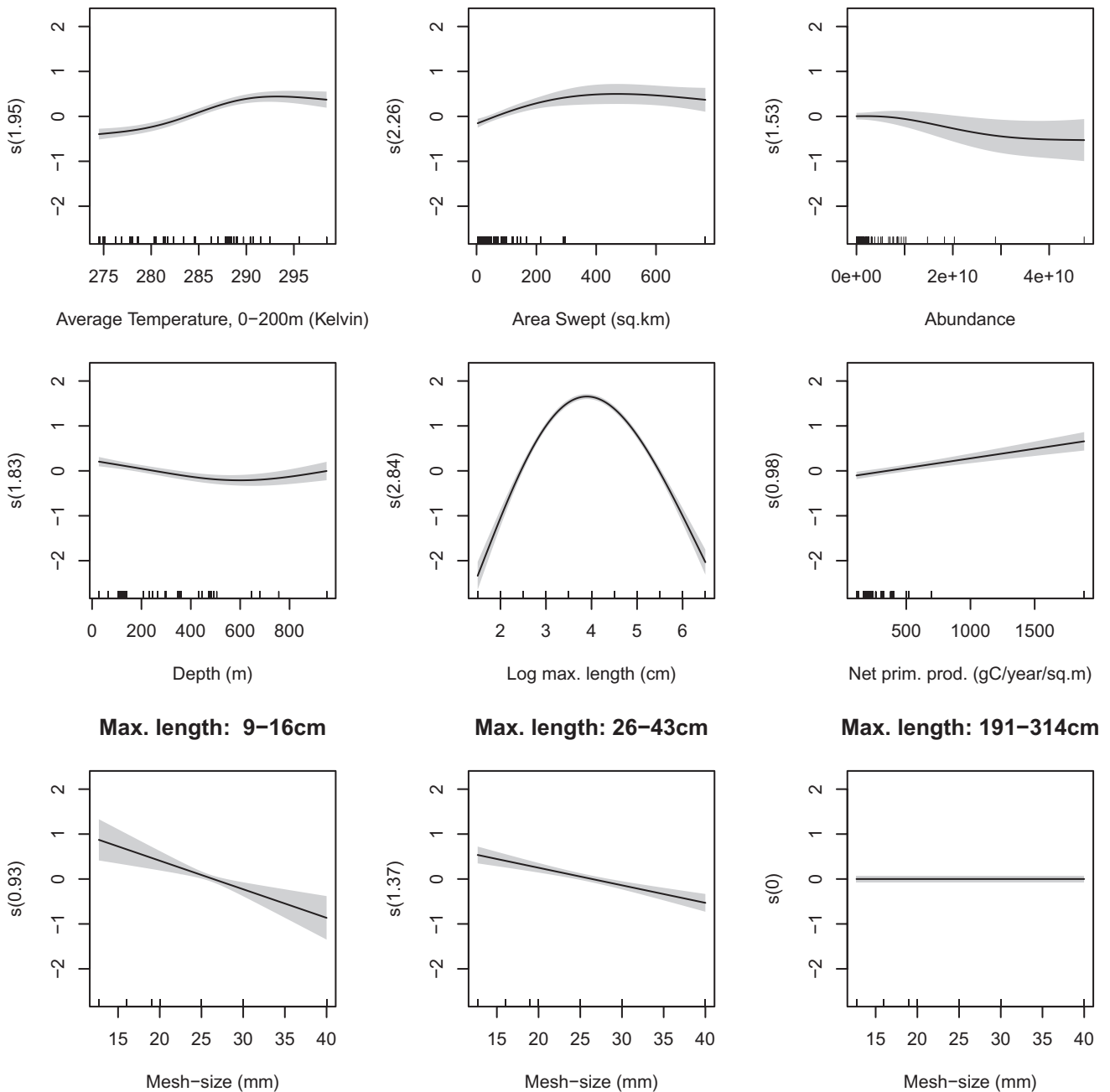
FIGURE 3 Estimated smoothing curves from the generalized additive model (GAM) using average sea temperature and other covariates to model the number of species observed by log maximum length group. Estimated degrees of freedom in parentheses on the *y* axis labels. Shaded area: 2*SE. Mesh-size smooths in bottom row only shown for three numerically abundant maximum length groups.
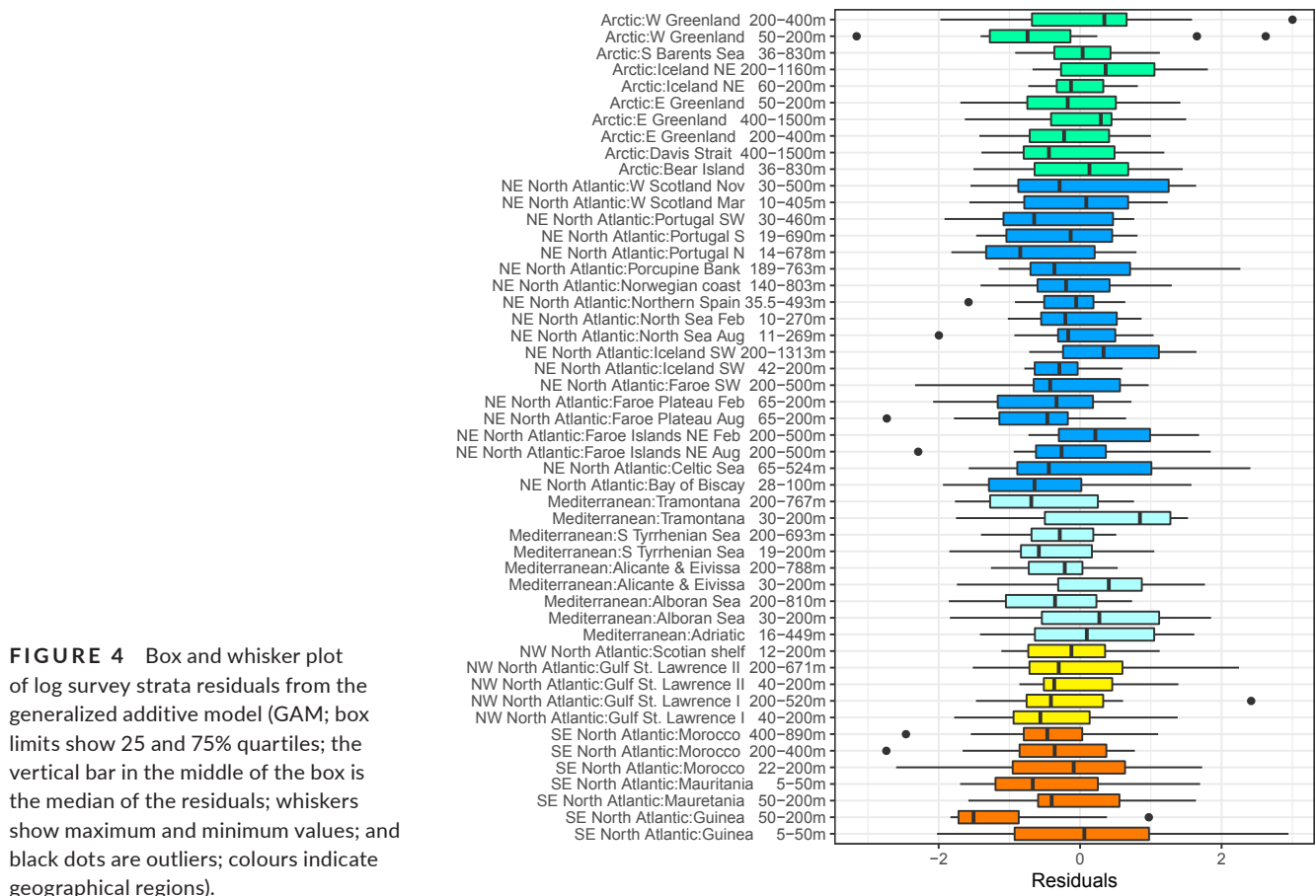
species observed, a significant effect of absolute fish abundance, significant nonlinear positive effects of average temperature in the upper 200 m of the water column and area swept, and a significant positive linear effect of net primary production. Log number of species caught declined significantly with depth and, for the smaller length groups, with increasing mesh-size (Figure 3). Vertical opening, temperature range, and catch in numbers were all insignificant, while stratum area was significant ($p = .02$), but generated a too high concurvity to be retained due to its high correlation with area swept. The model explained 85% of the deviance and had a lower AIC than model versions in which abundance was replaced by density. Using average temperature produced a lower AIC than bottom temperature and latitude and longitude, and a slightly higher AIC than sea surface temperature, but the best variance ratio. Although there are survey strata that produce significant negative residuals, such as the 50–200 m stratum in Guinea, which features the lowest number of hauls of all strata, there are no clear patterns in the residuals across survey strata. This suggests that the model provides an equally good description of fish species richness in the Atlantic, Arctic and Mediterranean Seas (Figure 4). Further model diagnostics are shown in Supporting Information Appendix S5: Figures S5.1 and S5.2).

Some of the smooth relationships suggested that the AIC value could be further reduced by using the logarithm or the inverse of the independent variable, and for log maximum length, in particular, that the smoother could be replaced by a second-order term, corresponding to a log-normal like distribution of richness versus maximum length. Replacing the independent variables in the GAM by inverse temperature, log abundance, log depth, net primary production, log area swept, an interaction between mesh-size and log maximum length, and the exponential of a second-order polynomial in log maximum length, reduced the AIC value from 2090 to 1860 and increased the percentage of deviance explained to 91%.

3.3 | Nonlinear models

We use nonlinear estimation techniques to compare the 'best' descriptive model identified by the GAM to the three other models. Fitting the four models to the average absolute densities and abundances we initially used variance ratio tests to determine whether the bias correcting terms ($\text{catch}_{i,j}^{\beta_8}$, $\text{aswept}_j^{\beta_9}$ and $\text{mesh}_i^{\beta_{11,j}}$) contributed significantly to the fit. We found that $\text{catch}_{i,j}^{\beta_8}$ did not improve the fit of the 'best' and neutral models significantly, improved the metabolic model marginally, but contributed highly significantly to the fit of the environmental model. The total area swept, $\text{aswept}_j^{\beta_9}$, contributed significantly to all models, except the environmental, while the term reflecting the interaction between mesh size and maximum length, $\text{mesh}_i^{\beta_{11,j}}$, was significant in all four models. In the neutral model the per capita immigration rates, λ_j, were not significantly different from zero; and were therefore replaced by a single overall λ for all log maximum length groups (see Supporting Information Appendix S5: Table S5.1).

FIGURE 4 Box and whisker plot of log survey strata residuals from the generalized additive model (GAM; box limits show 25 and 75% quartiles; the vertical bar in the middle of the box is the median of the residuals; whiskers show maximum and minimum values; and black dots are outliers; colours indicate geographical regions).

TABLE 1 Parameter estimates from TMB-model fits using average absolute density and abundance.

| Parameter	Best descriptive model	Neutral	Metabolic	Environmental
Constant $(\log \alpha)$	16.90 (1.63)***	22.95 (1.35)***	24.89 (1.65)***	3.093 (0.687)***
Latitude (β_0)				−0.518 (0.055)***
Longitude (β_1)				0.426 (0.073)***
Temperature (β_2)	0.322 (0.035)***	0.521 (0.029)***	0.466 (0.029)***	
Abundance (β_3)	0.034 (0.009)***			
Density (β_3)			0.056 (0.011)***	
Depth (β_4)	−0.115 (0.029)***			−0.167 (0.034)***
Net primary production (β_5)	0.217 (0.045)***			
Max. length (β_6)			−1.000 (0.246)***	
Log. max. length^2 (β_7)	−0.131 (0.028)***	−0.131 (0.031)***		−0.235 (0.029)***
Immigration (λ)		NS		
Catch (β_8)	NSR	NSR	NSR	0.067 (0.010)***
Area swept (β_9)	0.079 (0.023)***	0.129 (0.022)***	0.176 (0.022)***	NSR
Mesh:mlgr$_{1.5}$ $(\beta_{11,1})$	−1.537 (0.181)***	−1.351 (0.184)***	−1.675 (0.242)***	−1.070 (0.187)***
Mesh:mlgr$_{2.0}$ $(\beta_{11,2})$	−1.378 (0.162)***	−1.205 (0.164)***	−1.421 (0.202)***	−1.021 (0.168)***
Mesh:mlgr$_{2.5}$ $(\beta_{11,3})$	−0.977 (0.120)***	−0.875 (0.124)***	−0.972 (0.143)***	−0.755 (0.125)***
Mesh:mlgr$_{3.0}$ $(\beta_{11,4})$	−0.598 (0.099)***	−0.509 (0.103)***	−0.552 (0.108)***	−0.458 (0.103)***
Mesh:mlgr$_{3.5}$ $(\beta_{11,5})$	−0.401 (0.078)***	−0.335 (0.081)***	−0.347 (0.082)***	−0.340 (0.082)***
Mesh:mlgr$_{4.0}$ $(\beta_{11,6})$	−0.222 (0.066)***	−0.165 (0.070)**	−0.167 (0.070)*	−0.192 (0.070)***
Mesh:mlgr$_{4.5}$ $(\beta_{11,7})$	NS	NS	NS	NS
Mesh:mlgr$_{5.0}$ $(\beta_{11,8})$	NS	NS	NS	NS
Mesh:mlgr$_{5.5}$ $(\beta_{11,9})$	NS	NS	NS	NS
Mesh:mlgr$_{6.0}$ $(\beta_{11,10})$	NS	NS	NS	NS
Mesh:mlgr$_{6.5}$ $(\beta_{11,11})$	NS	NS	NS	NS
Scale parameter $(\log \kappa)$	3.752 (0.402)***	3.058 (0.239)***	3.085 (0.247)***	3.049 (0.239)***
Proportion of deviance explained	.900	.892	.891	.890
Pearson's R^2 (observed versus predicted)	.838	.787	.792	.789
AIC	1,891	1,929	1,931	1,937
ΔAIC		38	40	46

Note: AIC = Akaike information criterion; NS = non-significant term retained in the model fit; NSR = non-significant term removed from the model. Standard error in parentheses and significance levels indicated by stars (*** .001, ** .01, * .05; one-sided *t* test, *n* = 550).

Fitting the 'best' model to the average of the absolute abundances explains 90% of the deviance (Table 1). The neutral model provides the second-best fit (ΔAIC = 38) followed by the metabolic model (ΔAIC = 40) and the environmental model (ΔAIC = 46). Note that the difference between the metabolic and neutral models can be explained by the additional parameter included in the former. Many of the parameter estimates are similar across models. The interaction between log maximum length and mesh-size, $\beta_{11,j}$, is thus negative for the smaller species in all models, implying a general decline in the number of small species caught as mesh-size increases. In all models log maximum length and either temperature or latitude account for most of the deviance explained (Figure 5). The parameter estimates of the 'best', metabolic and neutral models are robust to the uncertainty in the modelled catchabilities as shown by

the limited distribution of deviance around the mean value of the 1,000 estimates. The standard deviations of the parameter estimates are also small (see Supporting Information Appendix S5: Table S5.1). Additional model diagnostics are presented in Supporting Information Appendix S5: Figures S5.4 and S5.5.

4 | DISCUSSION

Our study reveals strong consistent patterns in the number of demersal and benthopelagic fish species across the northern Atlantic. As in previous investigations, we find body size, depth and either temperature or latitude to be important, but our analysis is the first in which differences in the number of individuals caught, area swept
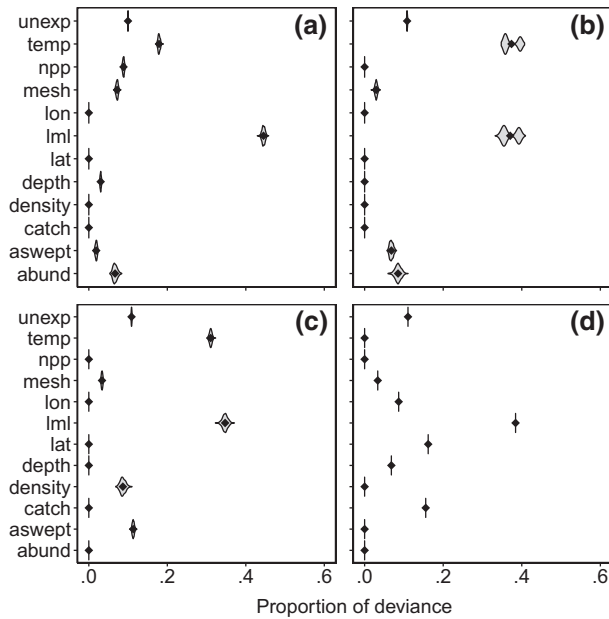
FIGURE 5 Violin plots of the relative contribution of the variables in each of the four models to the total deviance explained by each model. Results from 1,000 nonlinear model runs with stochastic catchabilities. unexp. = unexplained deviance, temp = average water column temperature, npp = net primary production, mesh = mesh size, lon = longitude, lml = log maximum species length, lat = latitude, aswept= area swept, abund = abundance . Models: (a) 'best' descriptive, (b) neutral, (c) metabolic, (d) environmental.

and mesh-size are considered, and where net primary production and absolute fish abundance or density are used as covariates. We find fish species richness to increase with temperature, fish abundance and net primary production, but to decline with depth and latitude. Adjusting for differences in area swept and mesh-size, the 'best' descriptive model explains 90% of the deviance in the number of species caught by log maximum length, temperature, fish abundance, depth and primary production (Table 1). The neutral model in which inverse temperature, a parabolic relationship with log maximum length, area swept and mesh-size are significant, explains 89% of the deviance, and so does the metabolic model. Our analyses furthermore show that both the neutral and metabolic models provide significantly better fits than the environmental model in which local richness is described as a function of log maximum length, catch, latitude, longitude and depth.

In all four nonlinear models more than half of the deviance is explained by a combination of log maximum length and either temperature or latitude (Figure 5). In the data the distribution of the number of species observed across maximum length groups is approximately lognormal (Figure 2). Similar distributions have been obtained for marine bivalves (Roy, Jablonsky & Martien, 2000), terrestrial snakes (Boback & Guyer, 2003), and insects (Siemann, Tilman & Haarstad, 1996), while more right-skewed distributions have been found for birds and mammals (Purvis, Orme & Dolphin, 2003; Smith & Lyons, 2013). A lognormal distribution also provided a highly significant fit in the best, neutral and environmental models

(Table 1). Contrary to this, metabolic theory predicts that species richness should scale with body mass raised to a power of 0.75, hence maximum length to a power of 2.25. This prediction was not confirmed by our analysis in which the power was estimated to be −1.00 (±0.48 confidence limits) and thus highly significantly different from the expected.

The average water column temperature from 0 to 200 m is only a marginally better predictor of the observed number of fish species than surface temperature, but much better than bottom temperature, and latitude and longitude. Latitude and average temperature are negatively correlated, but the correlation breaks down at intermediate latitudes, where average temperature generally is higher in the eastern part of the northern Atlantic due to the influence of the Gulf Stream. The increase in the number of fish species caught with temperature seems to be well described by the Arrhenius equation. Metabolic theory emphasizes the influence of temperature and body size on mutation rate and generation time, and it is interesting that the Arrhenius constant, β_2, is 0.47 (±0.06 conf. lim.) and 0.52 eV (±0.06 conf. lim.), respectively, in the metabolic and neutral models. This range is not far from the average activation energy of metabolism of 0.65 eV predicted by metabolic theory (Bailly, Cassemiro, Agostinho, Marques, & Agostinho, 2014; Gillooly & Allen, 2007), and close to empirical estimates of the activation energy of fish metabolism. Clarke and Johnston (1999) and Gillooly et al. (2001) both used the Arrhenius equation to describe the relationship between the resting metabolism of fish and temperature, and independently estimated the activation energy as 0.43 eV. Barneche et al. (2014) used a model with a temperature optimum to account for metabolic inactivation at high temperatures and found an activation energy of 0.59 eV. How temperature influences the rates of speciation and extinction is not completely known and other covarying factors may be involved (see e.g. Rabosky et al., 2018).

The 'best' and neutral models contain positive relationships between abundance and the number of species observed. The 'best' model also includes a significant positive relationship with net primary production. Areas of high productivity have been hypothesized to have higher species richness because they harbour more individuals and thus may be able to maintain a higher number of viable populations (Brown, 2014), although a recent review by Storch, Bodhalková, and Okie (2018) found the empirical evidence in favour of this hypothesis to be mixed. However, in areas where abundance has been significantly reduced by fishing, primary production may better reflect fish abundance and density in the unexploited state and hence be a better predictor of richness. Without primary production included in the 'best' model, the three largest positive differences between the observed and predicted number of species were generated by the data from the shelf off Mauretania, which features the highest primary production, but has been subject to marked overexploitation (Meissa & Gascuel, 2014). Note however, that abundance or density never accounted for more than 10% of the total deviance in the 'best', neutral and metabolic models, explaining the robustness of these models to the uncertainty in the catchabilities (Figure 5).

Tittensor and Worm (2016) and Worm and Tittensor (2018) used a neutral model to simulate species richness in the oceans and allowed speciation rate and generation time to depend on temperature. Thermal effects on speciation rate generated a stable but weak latitudinal richness gradient in their model, while thermal effects on generation time produced a transient latitudinal richness gradient that eventually disappeared. Combining the effect of an increase in abundance caused by the increase in ocean area towards the Equator and a temperature-dependent speciation rate produced the most realistic gradient in richness. Fitting a neutral model to the survey data we found a strong effect of temperature on species richness and a weaker influence of fish abundance. Furthermore, the shelf areas in the eastern Atlantic down to 200 m, the depth range where our fish species have their maximum abundance, increases with latitude from the Equator to the Arctic (Pilson & Seitzinger, 1996). A consistent decline in habitat area with latitude is therefore unlikely to explain our results.

The parameter describing the probability of immigration in the neutral model could not be estimated with sufficient precision. The known functional dependency between per capita immigration probability and the speciation rate in the surrounding meta-community makes it difficult to estimate both parameters simultaneously (Jabot & Chave, 2011). The immigration probability may depend on temperature and size, as assumed by Reuman et al. (2014), but the evidence for temperature related differences in larval dispersal is lacking (Leis et al., 2013) and when immigration probability was assumed to be size dependent, none of the estimates of λ_j was significant. Additional analysis of species distributions and information on the genetic divergence of subpopulations is necessary to fully understand the relationship. The neutral model has been criticized for predicting unrealistically long species ages for common species and too short species ages for new species with few individuals (Chisholm & O'Dwyer, 2014). Recent work has shown that more realistic species ages are generated when protracted speciation and weak selection caused by small differences in hereditary fitness are incorporated in the model (Rosindell et al., 2015), but no approximate solution for the number of species in each community is yet available for this model.

Despite the large sample sizes and good geographical coverage of the survey data, several problems may be associated with using bottom trawl survey data to study fish species richness and density patterns. The main aim of a scientific bottom trawl survey is often to provide reliable estimates of the relative abundance and year-class strength of commercially important fish species, and less attention may therefore be given to identifying species that are rare or of little or no commercial value. Trawl-survey catches may furthermore provide biased estimates of the species composition and density due to species and size-specific differences in the probability of the individuals being retained by the trawl (Arreguín-Sánchez, 1996). Some species and sizes are herded into the path of the trawl by the action of the otter doors and trawl sweeps, others escape under the fishing line or over the headline, while yet others are able to outswim the trawl. Among those entering the trawl the smaller individuals and

species may escape through the meshes. Factors that have been reported to influence the catch efficiency of survey trawls include time of day, light intensity, turbidity, current strength and direction, depth, sweep length, net spread, vertical opening, trawl speed, haul duration, and the size and type of the ground gear (Arreguín-Sánchez, 1996; Fraser, Greenstreet & Piet, 2007). Although we corrected our analysis for differences in species catchability, we were unable to fully account for all of the factors that may lead to species and size specific differences in catchability. This was due to the sparsity of spatially and temporally overlapping stock assessments, the absence of individual length measurements for many of the non-commercial species, and our use of average catch rates rather than individual hauls. However, as seen in Figure 5, density or abundance only explains less than 10% of the deviance. The sensitivity of our overall conclusions to the uncertainty in the catchabilities is therefore modest, and the parameter estimates and the relative importance of the variables change little in the different models. Finally, our use of a single estimate of maximum length for each species hides the fact that maximum body length in fish is likely to vary from area to area (Rypel, 2013). However, the maximum length of a species in a given area is difficult to estimate as it depends on local fishing mortality and sampling effort.

We base our analysis on the number of fish species and individuals observed over a recent period of time in different regions of the northern Atlantic, Arctic and Mediterranean Seas. It is now well documented that changes in fish distributions have occurred over the last decade or two in many regions of the North Atlantic and that these are significantly associated with changes in temperature, (Perry, Low, Ellis & Reynolds, 2005; Hiddink & Ter Hofstede, 2008; Fossheim et al., 2015; Batt, Morley, Selden, Tingley & Pinsky, 2017). We have fitted our models to data from a period when temperatures have been increasing, but where regulatory processes generally seem to maintain existing patterns in species richness (Gotelli et al., 2017). Future analyses should investigate whether these patterns will persist over longer time periods and how our model parameters will be modified by temperature change, for example by conducting the analyses on different time periods characterized by different mean temperatures. Such analyses could provide insight into the relative importance of temperature having a direct effect on metabolic processes versus its effects on other ecosystem features that affect species richness. For example, Marbá, Jordà, Augustí, Girard, and Duarte (2015) showed that the activation energy for many biological responses in the Mediterranean Sea is far higher than the reported activation energy for metabolism, suggesting that temperature increases are having additional ecosystem effects on biotic responses beyond their effect on metabolic processes and speciation rates. The effects of global warming on fish communities have been predicted from stacked species distribution models (SSDMs; e.g. Jones & Cheung, 2015), but these models have so far largely ignored the regularity in the distribution of fish species richness and abundance with log maximum length. This regularity accounts for a third or more of the deviance explained by our models (Figure 5) and may thus be used to improve the predictive capability of the SSDMs

significantly. But while the right-hand side of the richness versus log maximum length distribution, consisting of species with a maximum length larger than *c.* 50 cm, has been explained by size spectrum theory (Reuman et al., 2014), little is known about the processes shaping the left-hand side.

Numerous hypotheses have been put forward to explain the latitudinal pattern in species richness (Brown, 2014; Fine, 2015). Finding log maximum length, temperature, absolute fish abundance, depth and net primary production to explain 90% of the deviance in the distribution of demersal fish species richness across the northern Atlantic, and both neutral and metabolic equilibrium models to explain close to 89%, conveys an important message. When 89% of the deviance in the extant species richness can be explained by two competing mechanistic hypotheses, and by a model based on latitude, longitude and depth, and when many of the independent variables are significantly correlated, it seems relevant to question how much the present patterns in species richness and abundance can tell us about the underlying environmental, ecological and evolutionary processes (Gotelli et al., 2009). We probably need dynamic mechanistic models with more realistic descriptions of speciation, dispersal and extinction plus additional data to reveal how past changes in environmental (e.g. temperature, currents, ice cover, shelf area) and biotic (e.g. primary production) variables may have contributed to shaping the present distribution of species richness and the strong lognormal relationship between richness and maximum length (Fine, 2015; Descombes et al., 2018). Such data should include information from palaeo-geographical and climatological reconstructions of past environmental conditions as well as information about body size evolution, diversification rates and species lifetimes from molecular phylogenetics and the fossil record (Romano et al., 2016; Alfaro et al., 2018). In addition to providing a baseline from which we can evaluate future change, our data and results point to new possibilities for understanding demersal fish species biogeography in the northern Atlantic.

DATA AVAILABILITY STATEMENT

The richness and abundance data that support the findings of this study are deposited in DRYAD. They are also available together with the R code for the four models on GitHub (https://github.com) in the repository 'DTUAqua/biodiversity'.

ORCID

Henrik Gislason [ID] https://orcid.org/0000-0003-0242-3333

REFERENCES

Alfaro, M. E., Faircloth, B. C., Harrington, R. C., Sorenson, L., Friedman, M., Thacker, C. E., … Near, T. J. (2018). Explosive diversification of marine fishes at the Cretaceous-Palaeogene boundary. *Nature Ecology and Evolution*, *2*, 688–696.

Allen, A. P., Brown, J. H., & Gillooly, J. F. (2002). Global biodiversity, biochemical kinetics, and the energetic-equivalence rule. *Science*, *297*, 1545–1548.

Allen, A. P., & Gillooly, J. F. (2007). The mechanistic basis of the metabolic theory of ecology. *Oikos*, *116*, 1073–1077.

Andersen, K. H., Berge, T., Gonçalves, R. J., Hartvig, M., Heuschele, J., Hylander, S., … Kiørboe, T. (2016). Characteristic sizes of life in the oceans, from bacteria to whales. *Annual Review of Marine Science*, *8*, 217–241.

Arreguín-Sánchez, F. (1996). Catchability: A key parameter for fish stock assessment. *Reviews in Fish Biology and Fisheries*, *6*, 221–242.

Bailly, D., Cassemiro, F. A., Agostinho, C. S., Marques, E. E., & Agostinho, A. A. (2014). The metabolic theory of ecology convincingly explains the latitudinal diversity gradient of Neotropical freshwater fish. *Ecology*, *95*, 553–562.

Barneche, D. R., Kulbicki, M., Floeter, S. R., Friedlander, A. M., Maina, J., & Allen, A. P. (2014). Scaling metabolism from individuals to reef-fish communities at broad spatial scales. *Ecology Letters*, *17*, 1067–1076.

Batt, R. D., Morley, J. W., Selden, R. L., Tingley, M. W., & Pinsky, M. L. (2017). Gradual changes in range size accompany long-term trends in species richness. *Ecology Letters*, *20*, 1148–1157.

Behrenfeld, M. J., & Falkowski, P. G. (1977). Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and Oceanography*, *42*, 1–20.

Blowes, S. A., Belmaker, J., & Chase, J. M. (2017). Global reef fish richness gradients emerge from divergent and scale-dependent component changes. *Proceedings of the Royal Society B: Biological Sciences*, *284*, 20170947.

Boback, S. M., & Guyer, C. (2003). Empirical evidence for an optimal body size in snakes. *Evolution*, *57*, 345–351.

Brown, J. H. (2014). Why are there so many species in the tropics? *Journal of Biogeography*, *41*, 8–22.

Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information theoretic approach* (2nd ed.). New York, NY: Springer.

Cameron, A. C., & Windmeijer, F. A. (1996). R-squared measures for count data regression models with applications to health-care utilization. *Journal of Business and Economic Statistics*, *14*, 209–220.

Chisholm, R. A., & O'Dwyer, J. P. (2014). Species ages in neutral biodiversity models. *Theoretical Population Biology*, *93*, 85–94.

Clarke, A., & Johnston, N. M. (1999). Scaling of metabolic rate with body mass and temperature in teleost fish. *Journal of Animal Ecology*, *68*, 893–905.

Cullen, J. J., Franks, P. J., Karl, D. M., & Longhurst, A. (2002). Physical influences on marine ecosystem dynamics. In A. R. Robinson, J. J. McCarthy, & B. J. Rothschild (eds.). *The sea* (Vol. 12, pp. 297–336). New York: John Wiley & Sons, Inc.

Davies, T. D., & Baum, J. K. (2012). Extinction risk and overfishing: reconciling conservation and fisheries perspectives on the status of marine fishes. *Scientific Reports*, *2*, 561. https://doi.org/10.1038/srep00561

Descombes, P., Gaboriau, T., Albouy, C., Heine, C., Leprieur, F., & Pellissier, L. (2018). Linking species diversification to palaeo-environmental changes: A process-based modelling approach. *Global Ecology and Biogeography*, *27*, 233–244.

Etienne, R. S., & Olff, H. (2004). How dispersal limitation shapes species body size distributions in local communities. *The American Naturalist*, *163*, 69–83.

Fine, P. V. (2015). Ecological and evolutionary drivers of geographic variation in species diversity. *Annual Review of Ecology, Evolution, and Systematics*, *46*, 369–392.

Fisher, J. A., Frank, K. T., & Leggett, W. C. (2010). Global variation in marine fish body size and its role in biodiversity–ecosystem functioning. *Marine Ecology Progress Series*, 405, 1–13.

Fossheim, M., Primicerio, R., Johannesen, E., Ingvaldsen, R. B., Aschan, M. M., & Dolgov, A. V. (2015). Recent warming leads to a rapid borealization of fish communities in the Arctic. *Nature Climate Change*, 5, 673–677.

Fraser, H. M., Greenstreet, S. P., & Piet, G. J. (2007). Taking account of catchability in groundfish survey trawls: implications for estimating demersal fish biomass. *ICES Journal of Marine Science*, 64, 1800–1819.

Froese, R., & Pauly, D. (eds.), (2016). FishBase. World Wide Web electronic publication. www.fishbase.org, version (01/2016).

Gillooly, J. F., & Allen, A. P. (2007). Linking global patterns in biodiversity to evolutionary dynamics using metabolic theory. *Ecology*, 88, 1890–1894.

Gillooly, J. F., Brown, J. H., West, G. B., Savage, V. M., & Charnov, E. L. (2001). Effects of size and temperature on metabolic rate. *Science*, 293, 2248–2251.

Gotelli, N. J., Anderson, M. J., Arita, H. T., Chao, A., Colwell, R. K., Connolly, S. R., … Willig, M. R. (2009). Patterns and causes of species richness: a general simulation model for macroecology. *Ecology Letters*, 12, 873–886.

Gotelli, N. J., & Colwell, R. K. (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, 4, 379–391.

Gotelli, N. J., Shimadzu, H., Dornelas, M., McGill, B., Moyes, F., & Magurran, A. E. (2017). Community-level regulation of temporal trends in biodiversity. *Science Advances*, 3, e1700315.

Hiddink, J. G., & Ter Hofstede, R. (2008). Climate induced increase in species richness of marine fishes. *Global Change Biology*, 13, 453–460.

Hilbe, J. M. (2011). *Negative binomial regression* (2nd ed.). Cambridge: Cambridge University Press.

Hillebrand, H. (2004). Strength, slope and variability of marine latitudinal gradients. *Marine Ecology Progress Series*, 273, 251–267.

Hubbell, S. P. (2001). *The unified neutral theory of biodiversity and biogeography.* Monographs in Population Biology (Vol. 32). Princeton: Princeton University Press.

Jabot, F., & Chave, J. (2011). Analyzing tropical forest tree species abundance distributions using a non-neutral model and through approximate Bayesian inference. *The American Naturalist*, 178, E37–E47.

Jones, M. C., & Cheung, W. W. L. (2015). Multi-model ensemble projections of climate change effects on global marine biodiversity. *ICES Journal of Marine Science*, 72, 741–752.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., & Bell, B. (2015). TMB: Automatic differentiation and Laplace approximation. *arXiv preprint arXiv*: 1509.00660.

Leis, J. M., Caselle, J. E., Bradbury, I. R., Kristiansen, T., Llopiz, J. K., Miller, M. J., … Warner, R. R. (2013). Does fish larval dispersal differ between high and low latitudes? *Proceedings of the Royal Society B: Biological Sciences*, 280(1759), 20130327.

Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., Garcia, H. E., Baranova, O. K., … Seidov, D. (2013). Temperature. In S. Levitus (Ed.), A. Mishonov (Technical Ed.), *World Ocean Atlas 2013* (Vol 1, p. 40). NOOA, Silver Spring, MD, USA: NOAA Atlas NESDIS 73.

Macpherson, E., & Duarte, C. M. (1994). Patterns in species richness, size, and latitudinal range of East Atlantic fishes. *Ecography*, 17, 242–248.

Marbà, N., Jordà, G., Agustí, S., Girard, C., & Duarte, C. M. (2015). Footprints of climate change on Mediterranean Sea biota. *Frontiers in Marine Science*, 2, 56. 1–11.

Meissa, B., & Gascuel, D. (2014). Overfishing of marine resources: some lessons from the assessment of demersal stocks off Mauritania. *ICES Journal of Marine Science*, 72, 414–427.

Mora, C., Tittensor, D. P., & Myers, R. A. (2008). The completeness of taxonomic inventories for describing the global diversity and distribution of marine fishes. *Proceedings of the Royal Society B: Biological Sciences*, 275(1631), 149–155.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … Wagner, H. (2015). Vegan: Community ecology package. R package version 2.2-1. http://www.pelagicos.net/MARS6910_spring2015/manuals/R_vegan.pdf

Pante, E., & Simon-Bouhet, B. (2013). marmap: A Package for Importing, Plotting and Analyzing Bathymetric and Topographic Data in R. *PLoS ONE*, 8, e73051.

Perry, A. L., Low, P. J., Ellis, J. R., & Reynolds, J. D. (2005). Climate change and distribution shifts in marine fishes. *Science*, 308, 1912–1915.

Pilson, M. E., & Seitzinger, S. P. (1996). Areas of shallow water in the North Atlantic. *Biogeochemistry*, 35, 227–233.

Purvis, A., Orme, C. D. L., & Dolphin, K. (2003). Why are most species small-bodied? In Blackburn, T. M., & Gaston, K. J. (Eds.). *Macroecology: Concepts and consequences* (pp. 155–173). 43rd Annual Symposium of the British Ecological Society, Oxford, Blackwell Science Ltd., UK.

R Core Team (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., … Alfaro, M. E. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*, 55, 392–395.

Reuman, D. C., Gislason, H., Barnes, C., Mélin, F., & Jennings, S. (2014). The marine diversity spectrum. *Journal of Animal Ecology*, 83, 963–979.

Romano, C., Koot, M. B., Kogan, I., Brayard, A., Minikh, A. V., Brinkmann, W., … Kriwet, J. (2016). Permian-Triassic Osteichthyes (bony fishes): Diversity dynamics and body size evolution. *Biological Reviews*, 91, 106–147.

Rosindell, J., Harmon, L. J., & Etienne, R. S. (2015). Unifying ecology and macroevolution with individual-based theory. *Ecology Letters*, 18, 472–482.

Rosindell, J., Hubbell, S. P., & Etienne, R. S. (2011). The unified neutral theory of biodiversity and biogeography at age ten. *Trends in Ecology and Evolution*, 26, 340–348.

Roy, K., Jablonski, D., & Martien, K. K. (2000). Invariant size-frequency distributions along a latitudinal gradient in marine bivalves. *Proceedings of the National Academy of Sciences USA*, 97, 13150–13155.

Rypel, A. L. (2013). The cold-water connection: Bergmann's rule in North American freshwater fishes. *The American Naturalist*, 183, 147–156.

Segura, A. M., Calliari, D., Kruk, C., Fort, H., Izaguirre, I., Saad, J. F., & Arim, M. (2015). Metabolic dependence of phytoplankton species richness. *Global Ecology and Biogeography*, 24, 472–482.

Siemann, E., Tilman, D., & Haarstad, J. (1996). Insect species diversity, abundance and body size relationships. *Nature*, 380, 704–706.

Smith, F. A., & Lyons, S. K. (2013). *Animal body size: Linking pattern and process across space, time, and taxonomic group*. Chicago: University of Chicago Press..

Storch, D., Bodhalková, E., & Okie, J. (2018). The more-individuals hypothesis revisited: the role of community abundance in species richness regulation and the productivity-diversity relationship. *Ecology Letters*, 21, 920–937.

Tittensor, D. P., Mora, C., Jetz, W., Lotze, H. K., Ricard, D., Berghe, E. V., & Worm, B. (2010). Global patterns and predictors of marine biodiversity across taxa. *Nature*, 466, 1098–1101.

Tittensor, D. P., & Worm, B. (2016). A neutral-metabolic theory of latitudinal biodiversity. *Global Ecology and Biogeography*, 25, 630–641.

Walker, N. D., Maxwell, D. L., Le Quesne, W. J., & Jennings, S. (2017). Estimating efficiency of survey and commercial trawl gears from comparisons of catch-ratios. *ICES Journal of Marine Science*, 74, 1448–1457.

Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Boca Raton, FL: Chapman and Hall/CRC.

Wood, S. N., Pya, N., & Säfken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association*, 111, 1548–1563.

Worm, B., & Tittensor, D. P. (2018). *A theory of global biodiversity.* Monographs in Population Biology 60. Princeton: Princeton University Press.

Wright, S. D., Ross, H. A., Keeling, D. J., McBride, P., & Gillman, L. N. (2011). Thermal energy and the rate of genetic evolution in marine fishes. *Evolutionary Ecology, 25*, 525–530.

BIOSKETCH

Henrik Gislason is Professor Emeritus in fisheries ecology and interested in fish community models, species interaction, fish life history theory, marine biodiversity, and the impacts of fishing on marine ecosystems.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Gislason H, Collie J, MacKenzie BR, et al. Species richness in North Atlantic fish: Process concealed by pattern. *Global Ecol Biogeogr.* 2020;00:1–15. https://doi.org/10.1111/geb.13068

APPENDIX A

DATA SOURCES

Anon. (2006). *Methodic manual on instrumental surveys of commercial hydrobiont stocks in PINRO investigations areas* (p. 163). Murmansk, Russia: PINRO Press. (in Russian).

Benchoucha, S., Berraho, A., Bazairi, H., Katara, I., Benchrifi, S., & Valavanis, V. D. (2008). Salinity and temperature as factors controlling the spawning and catch of Parapenaeus longirostris along the Moroccan Atlantic Ocean. *Hydrobiologia, 612*, 109–123.

Bertrand, J. A., Gil de Sola, L., Papaconstantinou, C., Relini, G., & Souplet, A. (2002). The general specification of the MEDITS surveys. *Scientia Marina, 66*(Suppl. 2), 9–17.

Bourdages H., & Oullet, S. (2011). Geographic distribution and abundance indices of marine fish in the northern Gulf of St. Lawrence (1990–2009). *Canadian Technical Report of Fisheries and Aquatic Sciences, 2963*.

Domain, F., Chavance, P., & Diallo, A. (Eds.). (1999). La pêche côtière en Guinée: ressources et exploitation. IRD édition, Paris, *Collection colloques et séminaires*, 394 p.

Fowler, G. M., & Showell, M. A. (2009). Calibration of bottom trawl survey vessels: Comparative fishing between the Alfred Needler and Teleost on the Scotian Shelf during the summer of 2005. *Canadian Technical Report of Fisheries and Aquatic Sciences, 2824*, iv + 25.

Gascuel, D., Labrosse, P., Meissa, B., Taleb Sidi, M. O., & Guénette, S. (2007). Decline of the demersal resources in the North-West Africa: an analysis of Mauritanian trawl-survey data over the past 25 years. *African Journal of Marine Science, 29*, 331–345.

ICES. (2017). Manual of the IBTS North Eastern Atlantic Surveys. Series of ICES Survey Protocols SISP 15. 92 pp. http://doi.org/10.17895/ices.pub.3519

Jørgensen O. A. (1998). Survey for Greenland Halibut in NAFO Division 1C-1D. *NAFO Sci. Council Reports.* Doc. 98/25. Serial No. N3010, 26 pp.

Magnussen, E. (2002). Demersal fish assemblages of Faroe Bank: species composition, distribution, biomass spectrum and diversity. *Marine Ecology Progress Series, 238*, 211–225.

Marine Research Institute. (2010). Manuals for the Icelandic bottom trawl surveys in spring and autumn. Marine Research in Iceland: 125 pp. https://www.hafogvatn.is/static/research/files/fjolrit-156.pdf

Rätz, H.-J. (1999). Structures and changes of the demersal fish assemblage off Greenland, 1982-96. *NAFO Science Council Studies, 32*, 1–15.