Species richness in North Atlantic fish: Process concealed by pattern

Gislason, Henrik; Collie, Jeremy; MacKenzie, Brian; Nielsen, Anders; Borges, Maria de Fátima; Bottari, Teresa; Chaves, Corina; Dolgov, Andrey V.; Dulčić, Jakov; Duplisea, Daniel

Total number of authors:
30

Published in:
Global Ecology and Biogeography

Link to article, DOI:
10.1111/geb.13068

Publication date:
2020

Document Version
Peer reviewed version

Appendix S1. Description of models

Metabolic model

The metabolic model of species richness forms part of the Metabolic Theory of Ecology (Brown et al., 2004) and was first described by Allen et al. (2002). The model uses speciation and extinction rates to predict species richness. Speciation rate is assumed to be influenced by temperature. An increase in temperature will shorten generation time and increase per capita mutation rate, leading to an overall increase in speciation rate. The rate of extinction is assumed to be a function of the average abundance per species. In the equilibrium situation the model predicts that the species richness, *S*, in a community within an area of size *A*, can be described by:

$$S = \left(\frac{J}{A}\right) * \frac{b_0}{B_T} * M^{\frac{3}{4}} * \exp(-\frac{\beta_2}{k * T})$$

Where J is the total number of individuals in the area of size A, B_T is the average per species energy use in the community, b_0 is a scaling constant relating to metabolism and varies with taxa and other variables, M is the average body mass in the community raised to 3⁄4 to predict its influence on total metabolic rate, β_2 is the average activation energy of metabolism, T is the average temperature in A measured in Kelvin (K), and k is Boltzmann's constant ($k = 8.62 \times 10^{-5}$ eV/K). Furthermore, β_2 is assumed to be a constant (~0.6 - 0.7eV) (Gillooly & Allen 2007). In the original formulation of the model (Allen et al., 2002) the Energy Equivalence Rule of Damuth (1987) was used to demonstrate that B_T is independent of both average body mass and temperature. Also ignoring the potential contributions of changes in density, (J/A), and average body mass, M, in the above equation, the

exponential effect of temperature on metabolism was assumed to dominate the richness response (Allen et al., 2002). Later attempts to fit the model to empirical data often produced significant deviations from the predicted linear relationship between $ln(S)$ and $^{1}/_{T}$ (e.g. Hawkins et al., 2007), suggesting that one or several of these assumptions could have been violated. It was therefore proposed that further tests at least should consider differences in individual density and average body mass as determinants of species richness (Allen et al., 2007; Cassemiro & Diniz-Filho, 2010; Segura et al., 2015). In the implementation used in this paper we account for differences in body size and use data on the density of individuals in log maximum length intervals (lml). We also account for differences in sampling effort, catchability, and mesh size in the survey trawls. According to the metabolic theory we expect our estimates of activation energy, β_2, to be close to 0.65 eV, and the body size exponent, β_6, to be close to 9/4 = 2.25 ($M^{3/4} \sim ml^{9/4}$).

Neutral model

The basic assumption in the neutral theory is that all individuals within a particular trophic level are functionally equivalent in the sense that they have the same chances of reproduction and death (Hubbell 2001, 2005; Rosindell et al., 2011). This basic assumption is used to model the equilibrium species richness in a spatial setting in which local communities receive individuals from a surrounding meta-community where speciation takes place (see Figure S1.1). When an individual in the local community dies, it is either replaced by an offspring produced by a randomly selected individual from the local community itself with probability (1-λ) or by an offspring from a randomly selected individual in the meta-community with immigration probability λ. The meta-community contains J_M

individuals and undergoes a similar process, but here speciation takes place. In addition to being replaced by the offspring of a randomly selected individual from the meta-community with probability (1- ν), a dead individual in the meta-community may, with probability ν, be replaced by an individual of a new species generated by speciation. The whole procedure is repeated until the number of species in the local communities and in the meta-community has reached a stochastic equilibrium where introduction of new species in the local communities caused by meta-community speciation and emigration balance with local species extinctions caused by random death and replacement events.

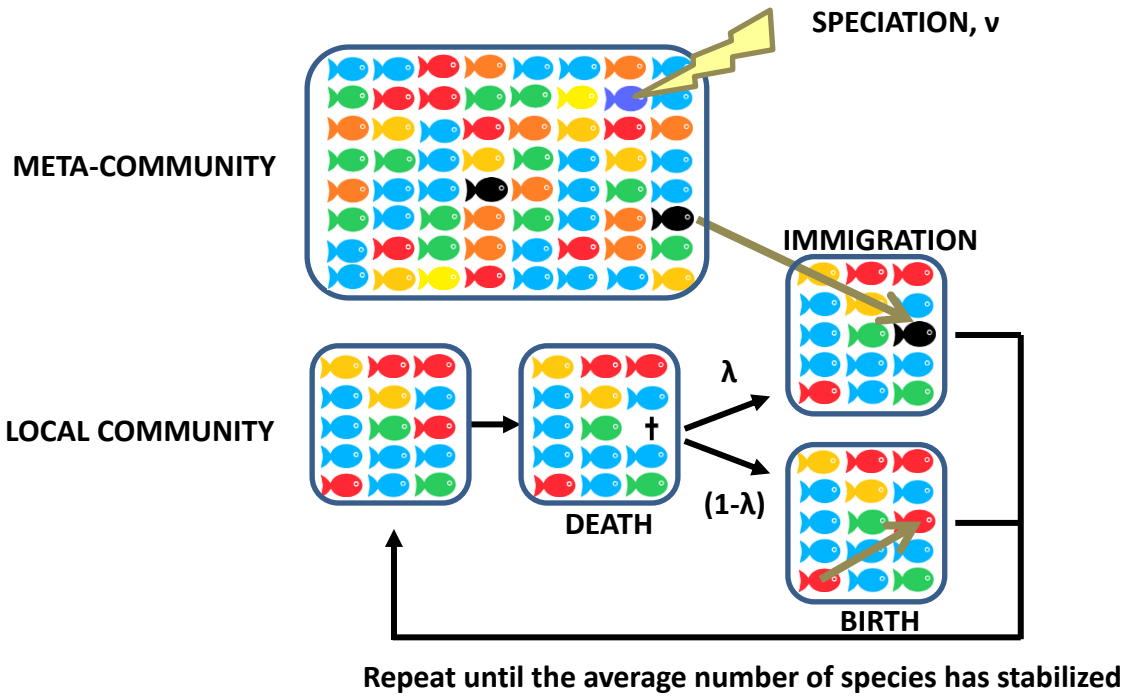

Figure S1.1 Basic mechanisms in the neutral theory of biodiversity of Hubbell (2001). Adapted from Rosindell et al. (2011).

Body mass is considered the 'master trait' of aquatic communities because of its influence on metabolism, respiration, movement, fecundity and natural mortality (Blanchard et al., 2017). For fish, the body mass of an individual of a given age and species can conveniently be predicted by the von Bertalanffy equation if the asymptotic body mass, W_∞, and the growth parameter, K, are known. Asymptotic body mass is a species or population specific parameter defined as the average body mass of infinitely old individuals, but is often close to the maximum observed body mass. Pelagic marine communities are generally highly size-structured and plots of log abundance versus log body mass – how abundance depends on individual body mass - frequently reveal linear size spectra with predictable slopes (Blanchard et al., 2017). This led Andersen and Beyer (2006) to demonstrate that the slope of the equivalent asymptotic size spectrum (log abundance versus log asymptotic species body mass – how abundance depends on asymptotic species mass) also can be predicted. Reuman et al. (2014) developed this idea further by combining asymptotic size spectra theory with the neutral theory. They derived relative abundance in each log asymptotic body mass interval from the asymptotic size spectrum and used the neutral model to derive the corresponding relative number of species in the intervals. Using the Moran version of the neutral model, for which Etienne and Olff (2004) had derived an approximate solution, and species richness data and maximum body mass data from Fishbase (Froese & Pauly, 2016), they were able to compare the relative predicted and observed log of the number of fish species in different log maximum mass classes, and found good agreement for fish species with a maximum body mass larger than 1 kg ($ml\sim$ 50 cm), the size above which fish dominate the biomass spectrum (Reuman et al., 2014). Note, that no observations of relative

abundance were used in these predictions. All relative abundance data were generated by the asymptotic size spectrum.

Here we use the neutral model as applied by Reuman et al. (2014) to predict the number of species in log maximum length bins, but replace the size spectrum predictions of abundance at size with survey observations of the abundance in the bins, assuming, as in metabolic theory, that speciation rate depends on temperature, while correcting for survey trawl efficiency, sampling effort and differences in trawl mesh size.

References

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

Andersen, K.H. and Beyer, J.E. (2006). Asymptotic size determines species abundance in the marine size spectrum. *The American Naturalist*, 168, 54-61.

Blanchard, J.L., Heneghan, R.F., Everett, J.D., Trebilco, R. and Richardson, A.J. (2017). From bacteria to whales: using functional size spectra to model marine ecosystems. *Trends in Ecology & Evolution*, 32, 174-186.

Brown, J.H., Gillooly, J.F., Allen, A.P., Savage, V.M. & West, G.B. (2004). Toward a metabolic theory of ecology. *Ecology*, 85, 1771-1789.

Cassemiro, F.A. and Diniz-Filho, J.A.F. (2010). Deviations from predictions of the metabolic theory of ecology can be explained by violations of assumptions. *Ecology*, 91, 3729-3738.

Damuth, J. (1987). Interspecific allometry of population density in mammals and other animals: the independence of body mass and population energy-use. *Biological Journal of the Linnean Society*, 31, 193-246.

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

Froese, R., & D. Pauly (Editors), (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.

Hawkins, B.A., Albuquerque, F.S., Araujo, M.B., Beck, J., Bini, L.M., Cabrero-Sanudo, F.J., Castro-Parga, I., Diniz-Filho, J.A.F., Ferrer-Castan, D., Field, R., & Gómez, J.F. (2007). A global evaluation of metabolic theory as an explanation for terrestrial species richness gradients. *Ecology*, 88, 1877-1888.

Hubbell, S.P. (2001). *The unified neutral theory of biodiversity and biogeography*. Monographs in Population Biology, Vol. 32. Princeton, Princeton University Press. 375pp.

Hubbell, S.P. (2005). Neutral theory in community ecology and the hypothesis of functional equivalence. *Functional Ecology*, 19, 166-172.

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

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

Appendix S2. Trawl survey data, independent variables and correlations.

Table S2.1. Trawl survey data

Sea area	Region	Av. Lat.	Av. Lon.	Month	First year	Last year	Av. SST C°	Av. SST diff. C°	Prim. prod. gC/m²	Min. depth m	Max. depth m	No of hauls	Mesh size cod end mm	Horiz. open. m	Vert. open. m	Towing speed nm/h	Total stratum area km²	Area swept km²
Mediterranean	Adriatic	43.86	15.74	5-9	96	08	17.9	10.8	224	16	449	657	20	16.5	2.3	3	54363	30.8
Mediterranean	Alboran Sea	36.75	-3.00	5-6	98	07	18.3	7.2	402	200	810	251	20	19.4	3.0	3	8812	27.1
Mediterranean	Alboran Sea	36.75	-3.00	5-6	98	07	18.4	7.6	402	30	200	169	20	16.4	3.0	3	2247	7.7
Mediterranean	Alicante and Ibiza	38.55	0.00	5-6	98	07	19.1	10.9	209	200	788	150	20	19.5	3.0	3	5065	16.2
Mediterranean	Alicante and Ibiza	38.55	0.00	5-6	98	07	19.1	11.3	209	30	200	215	20	16.4	3.0	3	5620	9.8
Atlantic	Bay of Biscay	45.84	-4.58	10-12	97	06	14.7	8.2	520	28	100	486	20	20.5	4.0	4	26316	36.9
Arctic	Bear Island-Spitsbergen	76.98	18.73	10-12	98	07	1.5	3.1	117	36	830	1576	16	25.0	7.0	3.4	258318	214.1
Atlantic	Celtic Sea	49.90	-8.19	10-12	97	06	12.9	6.0	317	65	524	545	20	18.7	4.1	4	141601	37.8
Arctic	Davis Strait	64.23	-54.73	9-11	97	08	1.6	5.0	114	400	1500	591	30	28.0	5.0	3	52602	42.3
Arctic	East Greenland	62.75	-36.75	10-11	82	07	4.2	1.5	112	200	400	1069	30	22.0	4.0	4.5	56488	97.8
Arctic	East Greenland	62.75	-36.75	10-11	82	07	3.3	1.1	112	50	200	366	30	22.0	4.0	4.5	17576	33.7
Arctic	East Greenland	64.23	-35.75	06-09	98	08	4.5	2.0	112	400	1500	476	30	28.0	5.0	3	43458	34.5
Atlantic	Faroe Plateau NE	61.67	-5.50	2-3	94	08	8.7	3.5	215	200	500	283	40	17.5	4.0	3	16393	27.5
Atlantic	Faroe Plateau NE	61.67	-5.50	8	96	08	8.7	3.5	215	200	500	585	40	17.5	4.0	3	16393	56.7
Atlantic	Faroe Plateau SW	61.73	-6.67	8	94	08	8.7	3.2	197	65	200	1033	40	16.0	4.0	3	18178	92.8
Atlantic	Faroe Plateau	61.73	-6.67	2-3	96	08	8.7	3.2	197	65	200	1643	40	16.0	4.0	3	18178	146.2
Atlantic	Faroe Plateau	61.67	-7.92	8	96	08	9.0	3.0	200	200	500	361	40	17.5	4.0	3	8576	35.0
Atlantic	Guinea inshore	9.75	-14.75	1-12	85	08	27.3	1.4	694	5	50	2280	25	13.0	3.0	3	32558	82.1
Atlantic	Guinea offshore	9.75	-14.75	1-12	85	08	27.3	1.4	694	50	200	103	25	13.0	3.0	3	4783	3.7
Atlantic	G. St. Lawrence N	49.60	-62.35	8	90	03	5.1	13.4	324	200	520	2100	44	13.4	5.5	3	59533	62.2
Atlantic	G. St. Lawrence N	49.60	-62.35	8	90	03	5.1	13.4	324	40	200	911	44	13.4	5.5	3	36826	27.0
Atlantic	G. St. Lawrence N	49.60	-62.35	8	04	13	5.1	13.4	324	200	520	1034	13	16.9	4.0	3	59533	24.3
Atlantic	G. St. Lawrence N	49.60	-62.35	8	04	13	5.1	13.4	324	40	200	671	13	16.9	4.0	3	36826	15.8
Arctic	Iceland NE	65.75	-16.63	10-11	96	07	3.9	4.2	188	200	1160	1248	40	17.0	5.5	3.8	106038	118.2
Arctic	Iceland NE	65.52	-17.26	10-11	96	07	5.2	4.7	188	60	200	470	40	17.0	4.5	3.8	40140	44.5
Atlantic	Iceland SW	65.06	-25.77	9-11	96	07	7.7	3.6	262	200	1313	916	40	17.0	5.5	3.8	50278	86.8
Atlantic	Iceland SW	65.01	-20.49	9-11	96	07	7.7	4.5	262	42	200	467	40	17.0	4.5	3.8	39766	44.2
Atlantic	Mauritania inshore	18.50	-16.90	1-12	89	07	20.8	6.9	1886	5	50	1621	40	17.0	4.0	3.5	14127	89.2
Atlantic	Mauritania offshore	18.50	-16.90	1-12	89	07	21.5	7.5	1886	50	200	1285	40	17.0	4.0	3.5	10646	70.7

Atlantic	Morocco N	32.29	-8.59	3-12	0	10	19.1	3.8	380	200	400	178	40	21.4	2.1	3	4876	6.9
Atlantic	Morocco_N	32.29	-8.59	3-12	0	10	19.2	4.0	380	22	200	548	40	21.4	2.1	3	28889	19.9
Atlantic	Morocco_N	32.29	-8.59	3-12	0	10	19.2	4.0	380	400	890	227	40	21.4	2.1	3	10690	12.6
Atlantic	North Sea	56.40	2.45	8-9	91	07	10.1	8.1	387	11	269	3736	20	21.0	5.0	4	493244	289.2
Atlantic	North Sea	56.18	2.27	2-3	77	08	10.1	8.2	387	10	270	9152	20	21.0	5.0	4	499507	765.0
Atlantic	Northern Spain	43.03	-5.74	9-11	85	08	15.5	5.7	379	36	493	2610	20	18.9	2.0	3	20279	137.0
Arctic	Norwegian coast	72.41	22.21	10-12	98	07	6.1	3.7	173	140	803	530	16	25.0	7.0	3.4	118922	66.9
Atlantic	Porcupine Bank	52.48	-13.10	9-10	01	08	12.4	4.5	235	189	763	651	20	20.0	3.5	3.5	53139	42.4
Atlantic	Portugal N	40.60	-9.30	10-11	89	08	16.3	4.7	498	14	678	678	20	15.0	4.6	3.5	12482	47.6
Atlantic	Portugal S	36.85	-8.24	10-11	89	08	18.1	4.5	498	19	690	411	20	15.0	4.6	3.5	2801	29.8
Atlantic	Portugal SW	38.27	-9.46	10-11	89	08	17.0	4.0	498	30	460	567	20	15.0	4.6	3.5	8248	38.5
Arctic	S Barents Sea	73.24	38.71	10-12	98	07	2.3	4.7	128	36	830	2144	16	25.0	7.0	3.4	820387	296.3
Mediterranean	Tyrrhenian Sea S	38.35	14.26	5-8	98	07	19.5	11.2	170	19	200	122	20	18.4	1.9	3	941	6.2
Mediterranean	Tyrrhenian Sea S	38.35	14.26	5-8	98	07	19.5	11.2	170	200	693	127	20	18.4	1.9	3	1821	13.0
Atlantic	Scotian shelf	44.17	-62.40	7-8	84	06	8.0	13.4	306	12	200	4097	19	12.5	4.6	3.5	190955	166.9
Mediterranean	Tramontana	41.00	2.15	5-6	98	07	18.1	10.6	242	200	767	106	20	20.3	3.0	3	5318	31.6
Mediterranean	Tramontana	41.00	2.15	5-6	98	07	17.7	10.2	242	30	200	347	20	16.3	3.0	3	10149	12.0
Arctic	Greenland W	62.88	-50.25	10-11	82	07	1.6	3.9	125	200	400	740	30	22.0	4.0	4.5	13077	67.7
Arctic	Greenland W	62.88	-50.25	10-11	82	07	1.6	3.4	125	50	199	1331	30	22.0	4.0	4.5	47617	121.8
Atlantic	Scotland W	57.29	-7.66	10-11	90	07	10.8	4.6	324	30	500	527	20	20.0	4.6	4	114470	49.1
Atlantic	Scotland W	57.27	-7.66	3	85	07	10.8	4.6	324	10	405	991	20	20.0	4.6	4	114003	99.5

Table S2.2. Independent variables used in the analysis of species richness.

Variable	**Name**	**Unit**
Average latitude of stratum i	lat_i	Degrees
Average longitude of stratum i	lon_i	Degrees
Temperature in stratum i	$temp_i$	Kelvin
Intra-annual temperature range in stratum i	$temp\_range_i$	Kelvin
Sea surface temperature in stratum i	sst_i	Kelvin
Intra-annual range in sea surface temperature in stratum i	$sstdif_i$	Kelvin
Temperature (0-200 m) in stratum i	ult_i	Kelvin
Intra-annual range in temperature (0-200 m) in stratum i	$ultdif_i$	Kelvin
Bottom temperature in stratum i	sbt_i	Kelvin
Intra-annual range in bottom temperature in stratum i	$sbtdif_i$	Kelvin
Net primary production in stratum i	npp_i	gC/m^2/ year
Average depth in stratum i	$depth_i$	m
Total number of valid hauls taken in stratum i	$nhauls_i$	No
Total area of stratum i	$asurv_i$	km^2
Total area swept in stratum i	$aswept_i$	km^2
Vertical opening of trawl used in stratum i	$vertop_i$	m
Horizontal opening of trawl used in stratum i	$horop_i$	m
Average towing speed in stratum i	$towsp_i$	nm/hour
Mesh size used in stratum i	$mesh_i$	mm
Mid length of maximum species length group j	ml_j	cm
Mid log length of maximum species length group j	lml_j	ln(cm)
Mid log length of maximum species length group j as factor	$mlgr_j$	
Total number of species recorded in stratum i and max. species length group j	$nsp_{i,j}$	No
No. of fish caught in stratum i and max. species length group j	$catch_{i,j}$	No
Absolute fish density in stratum i and max. species length group j	$density_{i,j}$	No/km
Absolute fish abundance in stratum i and max. species length group j	$abundance_{i,j}$	No

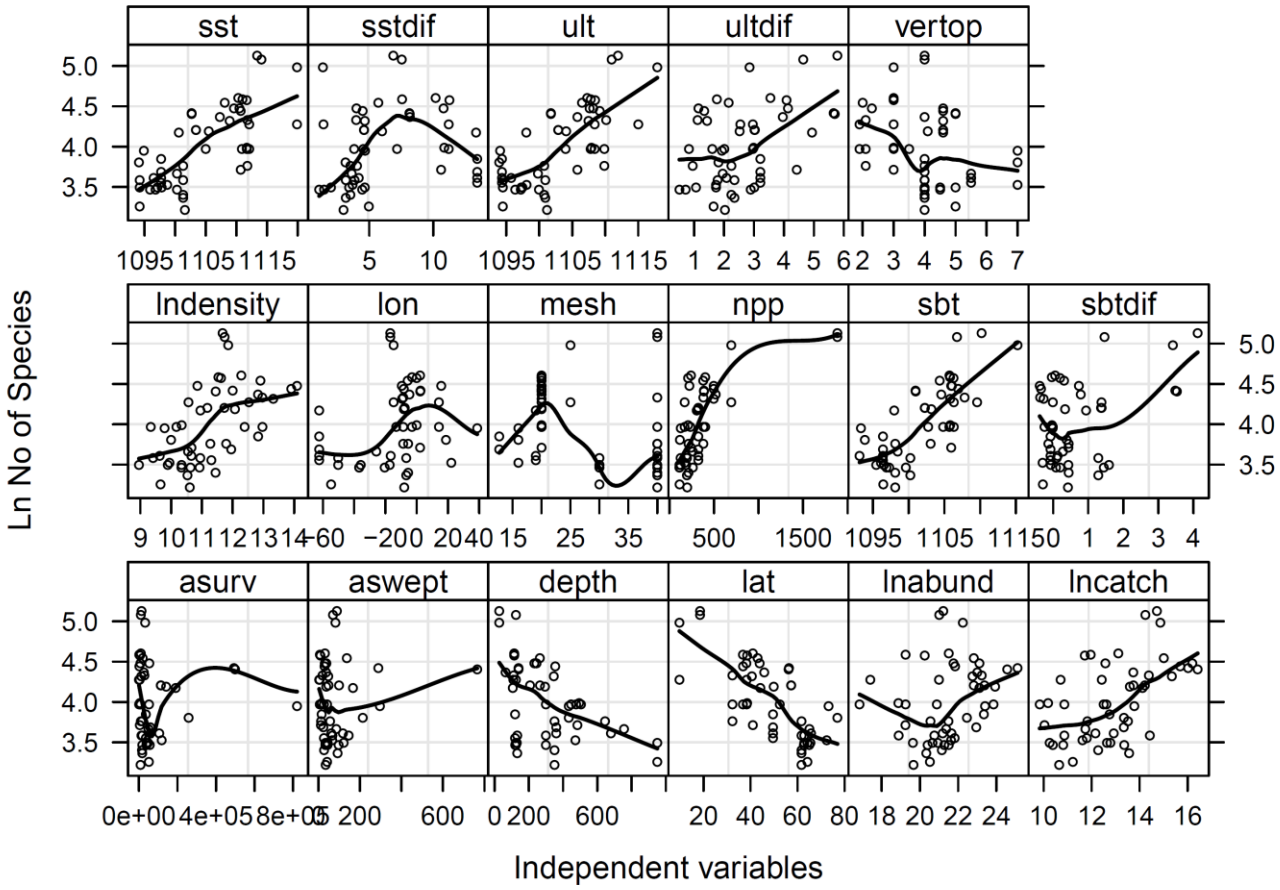

Figure S2.1 Log number of species versus independent variables (see Table S2.2 for variables and units. Logged values shown for density, abundance and catch).

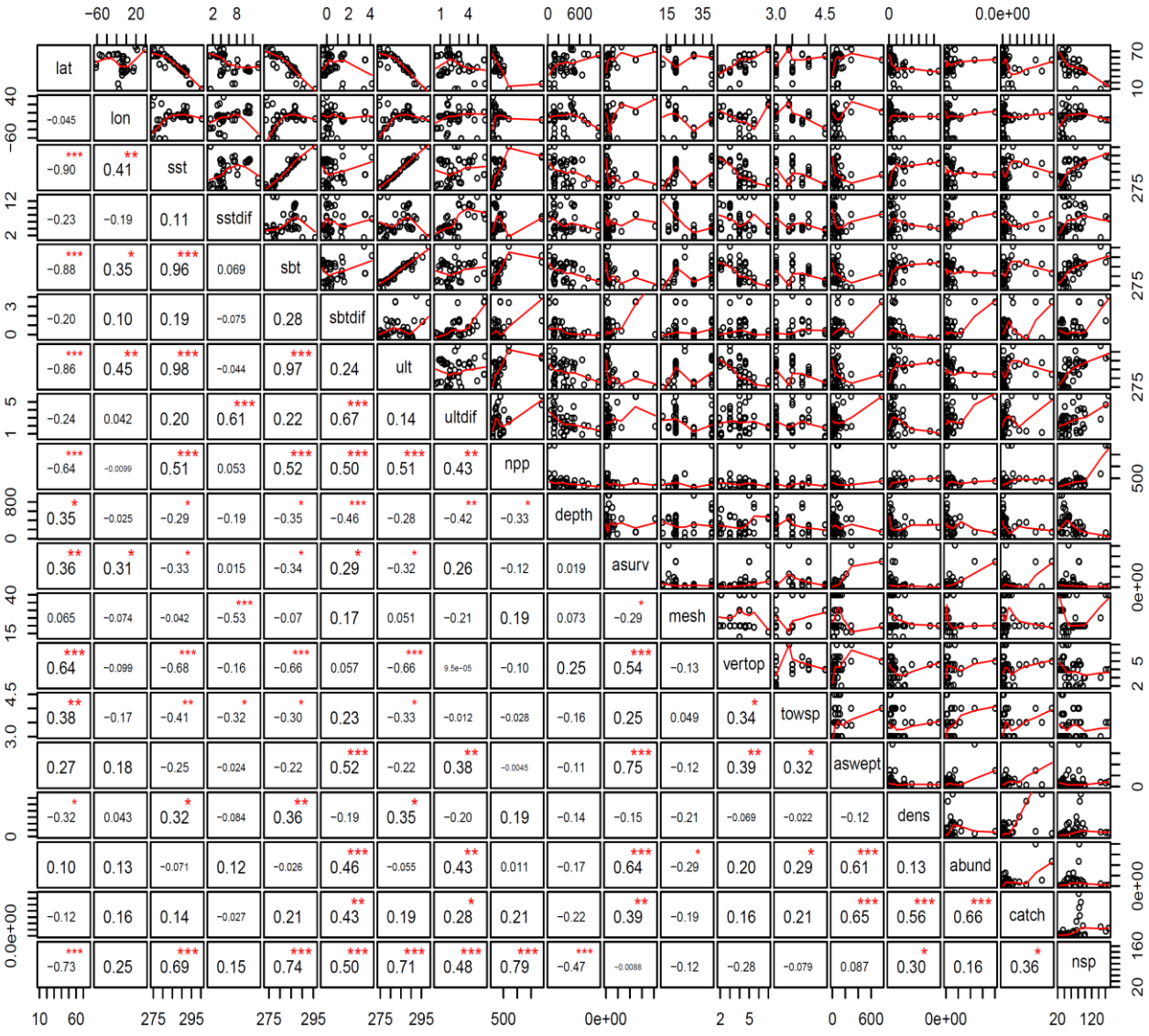

Figure S2.2. Pairwise plots of independent variables. Numbers below diagonal are Pearson correlation coefficients and stars indicate significance levels (***0.001, **0.01, *0.05), red lines are lowess smoothers. Variables are defined in Table S2.2.

Appendix S3. Estimating absolute density by correcting for differences in catchability.

Not all sizes and species of fish are caught equally efficiently by a survey trawl and differences in catchability across surveys and species may bias swept area estimates of fish density and abundance. Catchability is a species-specific parameter and is here defined as the average proportion of the individuals in the path of the trawl that are retained. Catchability depends, among other things, on the size, behavior and vertical position of the species and individuals relative to the gear, as well as on trawl characteristics such as mesh size and vertical opening (Arreguín-Sánchez 1996; Cadrin et al. 2016). It can be estimated by experiments where the catch per unit area swept is compared to absolute estimates of local density derived e.g. from stock assessments, or from visual observations, depletion experiments, tagging experiments or modelling of herding and escapement (Somerton et al. 1999, Walsh 1992). Here we use wing spread to characterise the width of the path swept by the trawl and estimate catchability as the swept area estimate of average abundance in the survey stratum over the time span of the survey divided by the total average abundance obtained from a stock assessment over the same time span, month and area.

However, for the majority of the species and survey areas no stock assessments are available and no direct estimate of catchability can therefore be made. To predict the missing catchabilities we followed the approach of Walker et al. (2017), who assumed that differences in the distribution and vertical position of the species would be responsible for much of the across-species differences in catchability (see also Aglen et al. 1999). We thus divided the species into: (a) those whose main distribution is outside the main depth range of the surveys (bathydemersal or bathypelagic species

found mainly at more than 200 m of depth, and species mainly occurring in the infra-littoral zone); and those whose main distribution is inside the main depth range of the surveys, but either: (b) occur on un-trawlable grounds (species that are mainly found on 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 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 bottom contact).

We identified 56 cases for which the spatio-temporal coverage of one of our surveys matched a stock assessment. In most of these cases we calculated catchability as the average swept area estimate of total stock size divided by the corresponding average stock assessment abundance of fish age 1 and older, adjusted, if necessary, by mortality to match the timing of the survey, Table S3.1. In a few cases we used available catchability estimates from surplus production models. For haddock (*Melanogrammus aeglefinus*) and whiting (*Merlangius merlangus*) some catchability estimates exceeded 1.0, probably because the herding effect of the bridles, sweeps and doors for these two species widened the effective path swept by the trawl (Fraser et al. 2007).

The results show that average catchability is lower and more variable for pelagic (mean catchability = 0.099) and burying (mean catchability = 0.050) species, and for species occurring in reefs, rocky and shallow areas (mean catchability = 0.012), than for midwater (0.518), near bottom (1.043) and bottom resting species (0.339). No data are available for bathypelagic and bathydemersal species, of which many were identified only to the family or genus level. Only species found on the seabed, near the

seabed or in midwater with bottom contact (group d) were therefore retained in the analysis. Plotting log assessment stock size against log swept area stock size for these species revealed a good correspondence between log abundance from swept area calculations and age-based stock assessments (R^2=0.70, n=36), see Figure S3.1.

We fitted a Gaussian model to the logged catchabilities, $logQ_i$, to predict catchability for the demersal and benthopelagic species in group (d). In the model the vertical position of the species was considered a fixed factor and species identity and survey area were random factors:

$$logQ_i = \alpha(group_i) + B(species_i) + D(area_i) + \varepsilon_i$$

Where i=1,…,41 is a sample identifier, $logQ_i$ is the log-catchability of $species_i$ in survey $area_i$, $\alpha(group_i)$ is a fixed factor characterising each of the three vertical groups $species_i$ may belong to (on bottom, close to bottom, and midwater); $B(species_i)$ and $D(area_i)$ are normally distributed random factors associated with species and survey area, respectively; and ε_i is a normally distributed random error with zero mean. The log catchability of a species will thus depend on whether a catchability estimate for the same species is available from another survey area, in which case the overall variance is reduced by subtracting the between species variance from the overall variance; or whether an estimate for another species is available from the same survey area, in which case the overall variance is reduced by the between survey area variance.

We drew random estimates of trawl efficiency from the model for each combination of vertical group, species and survey stratum, and used these to calculate absolute density and abundance for each species and area. We repeated this procedure 1000 times for each of the 412 species in group (d), generating a dataset with 1000 estimates of absolute density and abundance for each species and survey stratum. The average of the estimates for each species and survey stratum were finally cumulated for each maximum length group and used as input to the GAM model and the four non-linear models. To illustrate the sensitivity of the models to the uncertainty in the catchability estimates we also fitted the four non-linear models to each of the 1000 datasets and calculated the average and standard deviation of the resulting parameter estimates, Table S5.1.

Stock assessments were only available from northern areas, where high opening survey trawls are used, and for this reason we could not include vertical opening in the catchability model. The lack of length composition data for the non-commercial fish species from many of the surveys meant that the effect of individual size also could not be included directly in the model. However, catchability is known to depend on the lengths of the individual fish and mesh size, and may also depend on the vertical opening for species escaping above the headline. To account for these dependencies, we inserted vertical opening and a mesh size $\times$ log maximum length interaction as separate variables in all the species richness models. We found mesh size to explain a small but significant part of the deviance, while vertical opening was insignificant.

References

Aglen, A., Engås, A., Huse, I., Michalsen, K. and Stensholt, B.K. (1999). How vertical fish distribution may affect survey results. *ICES Journal of Marine Science*, 56, 345-360.

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

Cadrin, S. X., DeCelles, G. R., & Reid, D. (2016). Informing fishery assessment and management with field observations of selectivity and efficiency. *Fisheries Research*, 184, 9-17.

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.

Somerton, D., Ianelli, J., Walsh, S., Smith, S., Godø, O.R. and Ramm, D. (1999). Incorporating experimentally derived estimates of survey trawl efficiency into the stock assessment process: a discussion. *ICES Journal of Marine Science*, 56, 299-302.

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.

Walsh, S. (1992). Size-dependent selection at the footgear of a groundfish survey trawl. *North American Journal of Fisheries Management*, 12:3, 625-633.

Table S3.1. Estimates of catchability by species, area and season derived by dividing swept area abundance estimates with abundance estimates from stock assessments.

Species\Area	Northern Spain	Bay of Biscay	Celtic Sea	North Sea		West of Scotland		Faroe Islands		Iceland	Barents Sea	East Greenland	Scotian shelf	Northern Gulf of St. Lawrence	Average	Overall mean of species values	Overall stdev of species values
Quarter	4	4	3	1	3	1	4	1	3	4	3	4	3	3			
Pelagic																0.099	0.13744
Sprattus sprattus				0.0386											0.0386		
Clupea harengus				0.4598		0.4549	0.2640			0.0260			0.0767		0.2563		
Sardina pilchardus	0.0021														0.0021		
Midwater																0.518	0.2734
Trisopterus esmarkii				0.5599											0.5599		
Mallotus villosus											0.5660				0.5660		
Pollachius virens				0.1547				0.0522	0.1210	0.1036	0.0270				0.0917		
Sebastes fasciatus														0.8500	0.8500		
Sebastes norwegicus										0.3230					0.3230		
Sebastes mentella											0.0846			1.3500	0.7173		
Near bottom																1.043	0.84985
Merluccius bilinearis													0.3500		0.3500		
Gadus morhua			0.2620	0.1056	0.2441	0.1347	0.2025	0.7325	0.2272	0.1144	0.1280	0.2257	0.2949	0.5127	0.2653		
Melanogrammus aeglefinus			3.5081	0.2439	3.2069			1.2267	1.6827	1.3782	1.0147				1.7516		
Merlangius merlangus				0.2335		2.2912	2.8883								1.8043		
On bottom																0.339	0.27386
Pleuronectes platessa				0.0429						0.0821					0.0625		
Scophthalmus maximus				0.0284											0.0284		
Lepidorhombus whiffiagonis	0.5533														0.5533		
Glyptocephalus cynoglossus														0.5100	0.5100		
Reinhardtius hippoglossoides													0.2000		0.2000		
Hippoglossus hippoglossus													0.6800		0.6800		
Burried																0.050	0.06836
Solea solea		0.18912		0.0066											0.0979		
Ammodytes sp.				0.0003	0.0021										0.0012		
Mostly occurring on reefs or in shallow or rocky areas																0.012	0.01048
Brosme brosme										0.0095					0.0095		
Molva molva										0.0267					0.0267		
Anarhichas lupus										0.0109					0.0109		
Dicentrachus labrax				0.0017											0.0017		

Table S3.2. Results of fitting a mixed model to the log catchability data.

Intercept		
Habitat	$\alpha(group_i)$	Standard error
On bottom	-1.687***	0.4334
Near bottom	-0.4559	0.4358
Midwater	-0.9347*	0.4241

Variance component	
	Variance
$B(species_i)$	0.5278
$D(area_i)$	0.2488
ε_i	0.4172

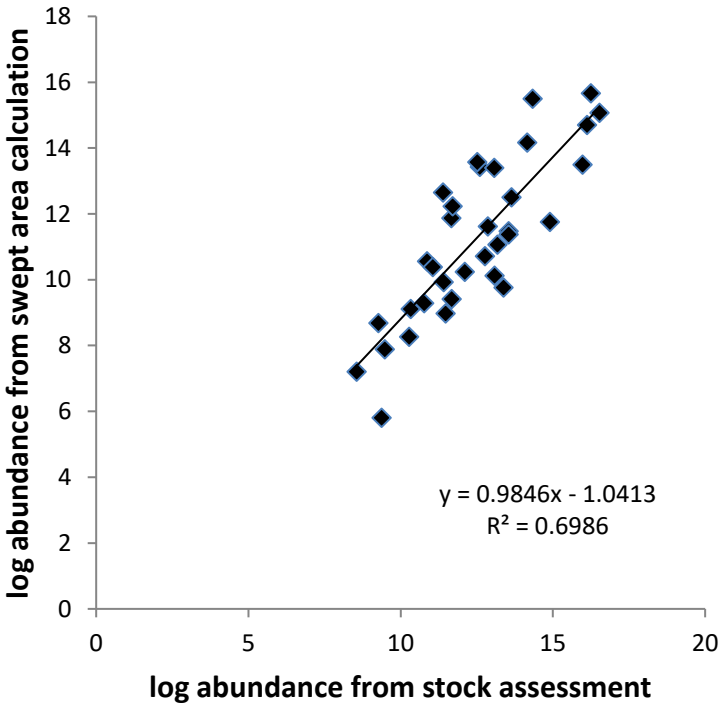

Figure S3.1 Log swept area abundance versus log abundance from stock assessments.

Appendix S4. Estimates of the minimum number of unobserved species

To estimate how well the richness of bottom dwelling, near bottom and midwater fish was sampled by the surveys we used the Chao1 and the ACE estimators (see Chao & Chiu, 2014) from the specpool function in the R-package vegan ver. 2.5-6 (Oksanen *et al.* 2015) to provide minimum estimates of the number of unobserved species in each survey stratum. These estimators use the number of species that have been caught once, twice, etc. to generate a minimum prediction of the number of unobserved species. Based on the Chao1 and ACE estimates an overall minimum of 7 to 8% of the demersal fish species available to the survey gear were predicted not to have been recorded in a particular stratum (Table S4.1). In seven survey strata no singletons (a species represented by a single individual) or doubletons (a species represented by two individuals) at all were found, suggesting that factors other than sampling intensity, i.e. species identification, may have influenced the recording of the rarer species.

References

Chao, A. & Chiu, C.H. (2014). Species richness: estimation and comparison. Wiley StatsRef: *Statistics Reference Online*, 1-26.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., ..., & Wagner, H. (2015). Vegan: Community Ecology Package. R package version 2.5-6. http://www.pelagicos.net/MARS6910_spring2015/manuals/R_vegan.pdf

Table S4.1. Total number of hauls taken, species in group (d) (on bottom + over bottom + midwater) identified and individuals caught. Number of species recorded once (singletons) or twice (doubletons) and Chao1 and ACE extrapolations of the minimum total number of species present in each stratum.

Region	Min depth m	Max depth M	No. of hauls	No of individuals millions	No of species	No of doubletons	No of singletons	Minimum no of species (Chao1)	Minimum no of species (ACE)
Adriatic	16	449	657	0.353	88	3	8	95	98
Alboran Sea	200	810	251	0.154	53	3	12	70	70
Alboran Sea	30	200	169	0.210	98	7	15	111	119
Alicante and Eivissa	200	788	150	0.116	54	2	10	69	64
Alicante and Eivissa	30	200	215	0.268	98	2	12	120	108
Bay of Biscay	28	100	486	3.475	80	8	11	86	93
Bear Island-Spitsbergen	36	830	1576	0.985	45	0	2	46	46
Celtic Sea	65	524	545	1.788	66	1	4	69	69
Davis Strait	400	1500	591	0.075	26	1	4	29	29
East Greenland	200	400	1069	0.639	33	0	0	33	33
East Greenland	50	200	366	0.347	33	9	0	33	33
East Greenland	400	1500	476	0.037	33	0	4	39	37
Faroe Plateau NE	200	500	283	2.081	43	0	3	46	46
Faroe Plateau NE	200	500	585	0.812	30	0	2	31	31
Faroe Plateau	65	200	1033	3.050	36	0	1	36	36
Faroe Plateau	65	200	1643	1.279	29	3	2	29	30
Faroe SW	200	500	361	0.263	25	0	0	25	25
Guinea	5	50	2280	2.997	147	3	6	151	150
Guinea	50	200	103	0.253	72	1	4	75	73
Gulf of St. Lawrence N	200	520	2100	0.525	37	0	3	40	39
Gulf of St. Lawrence N	40	200	911	0.845	36	1	4	39	46
Gulf of St. Lawrence N	200	520	1034	0.797	40	1	5	45	46
Gulf of St. Lawrence N	40	200	671	1.993	47	0	6	62	54
Iceland NE	200	1160	1248	0.333	37	7	0	37	37
Iceland NE	60	200	470	0.447	33	0	3	36	34
Iceland SW	200	1313	916	0.826	39	2	1	39	39
Iceland SW	42	200	467	0.899	32	2	1	32	32
Mauritania	5	50	1621	2.687	169	2	5	172	171
Mauritania	50	200	1285	1.568	161	4	6	164	164
Morocco	200	400	178	0.344	53	0	0	53	NA
Morocco	22	200	548	1.849	76	0	0	76	NA
Morocco	400	890	227	0.218	43	0	0	43	NA
North Sea	11	269	3736	16.400	83	3	4	85	85
North Sea	10	270	9152	23.368	83	1	8	97	90
Northern Spain	36	493	2610	10.567	95	5	12	106	111
Norwegian coast	140	803	530	0.287	34	3	4	36	37
Porcupine Bank	189	763	651	3.885	53	0	0	53	53
Portugal N	14	678	678	6.175	78	5	4	79	80
Portugal S	19	690	411	7.394	89	3	7	94	93
Portugal SW	30	460	567	11.914	92	5	8	97	98
Southern Barents Sea	36	830	2144	1.176	52	3	3	53	54
Tyrrhenian Sea S	19	200	122	0.057	72	5	12	83	86
Tyrrhenian Sea S	200	693	127	0.033	53	4	5	55	56
Scotian Shelf	12	200	4097	2.086	66	7	2	66	67
Tramontana	200	767	106	0.115	41	1	7	51	47
Tramontana	30	200	347	1.566	100	13	0	100	100
West Greenland	200	400	740	0.141	37	2	0	37	37
West Greenland	50	199	1331	0.411	34	0	0	34	34
West of Scotland	30	500	527	2.347	67	2	5	70	70
West of Scotland	10	405	991	2.999	67	1	2	68	68
All Regions			53382	123.435	412	8	7	414	415

Appendix S5. Model diagnostics.

The figures in this appendix display the diagnostic plots for the GAM model, the simplified GAM model and the four non-linear TMB models.

Interpreting plots of raw residuals from regression models with negative binomially distributed data is difficult, because the dependent variable undertakes distinct integer values while the predicted responses are real numbers. For small predicted values the deviance residuals will therefore appear as curves or lines of points in a residual plot, each curve responding to a particular observed count, making visual checks of bias difficult or impossible. Dunn and Smythe (1995) suggested to present the residuals in a more easily interpretable way by using quantile residuals instead of Pearson or deviance residuals and to randomize the discrete observations assuming a uniform distribution of probability within each step in the cumulative probability distribution. Using randomized quantile residuals removes the lines and provide normally distributed residuals except if the underlying model is biased.

Figure S5.1a presents the standard plots from the gam.check routine of the mgcv package showing deviance residuals versus theoretical quantiles and linear predictors as well as a histogram of the distribution of deviance residuals and a plot of predicted versus observed values. Note the curved lines of points formed by the deviance residuals in the upper right panel. Figure S5.1.b shows the logged randomized quantile residuals, demonstrating that the residuals are normally distributed with constant variance and a small and negligible model bias.

Figure S5.2 displays plots of deviance residuals against each independent variable in the model, demonstrating that heteroscedasticity is not a problem.

Figure S5.3 displays the smoothing curves from the simplified GAM demonstrating that inserting logged values for the independent parameters and replacing zero abundances with arbitrary small numbers introduce a bias in the relationship between richness and log abundance (upper right hand panel).

Figure S5.4 shows observed versus predicted species richness from each of the four non-linear TMB-models and Figure S5.5 displays the randomized quantile residuals and the Q-Q plots from the models.

Dunn, P.K. & Smythe, G.K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5, 236-244.

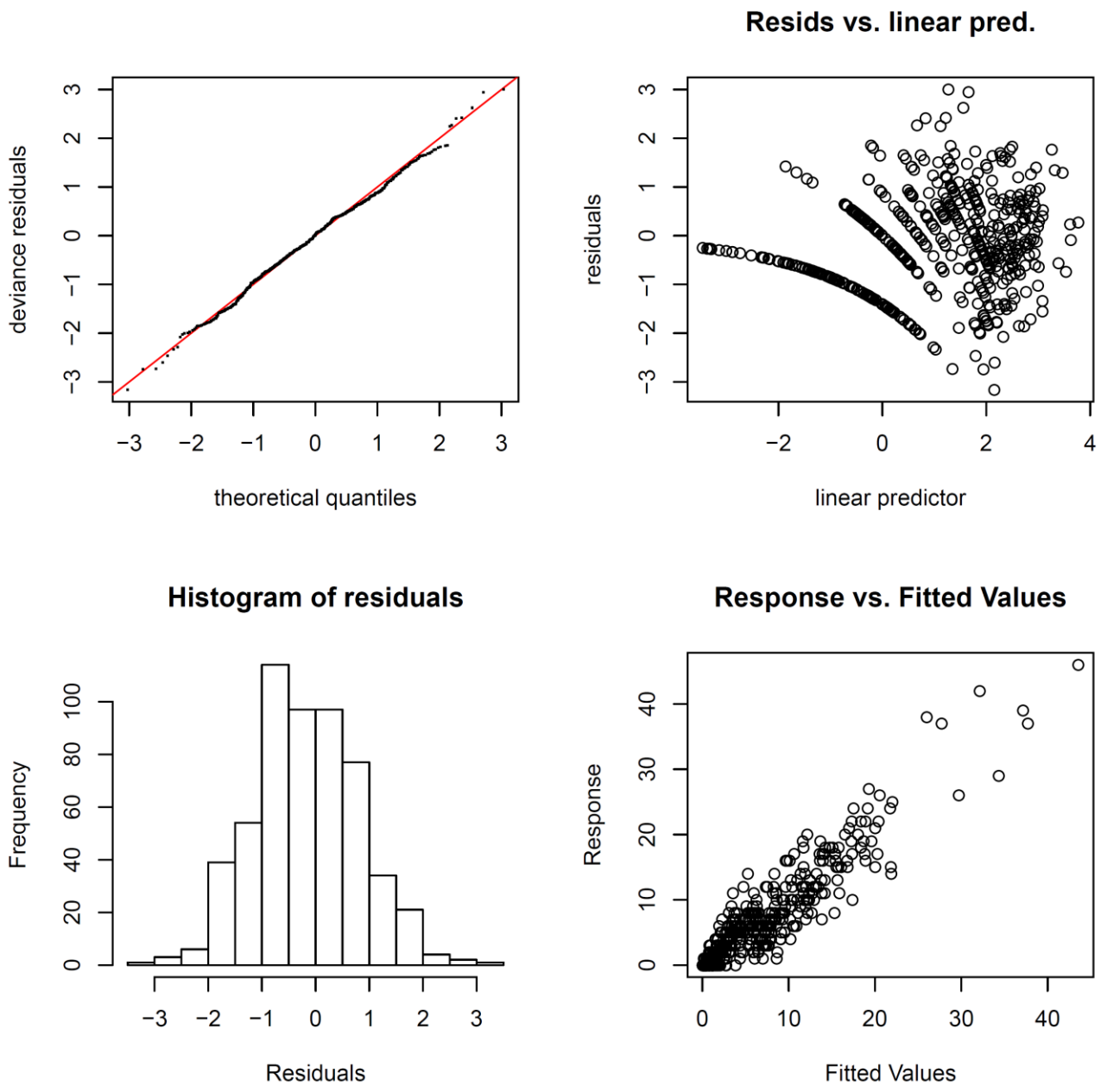

Figure S5.1a. QQ-plot, log residuals versus predicted values, histogram of log residuals and response versus fitted values from GAM model with average water column (0-200m) temperature (n=550). Note that the upper right hand plot estimates of deviance residuals from richness observations (integers) and predictors (real numbers). This produces the curved relationships.

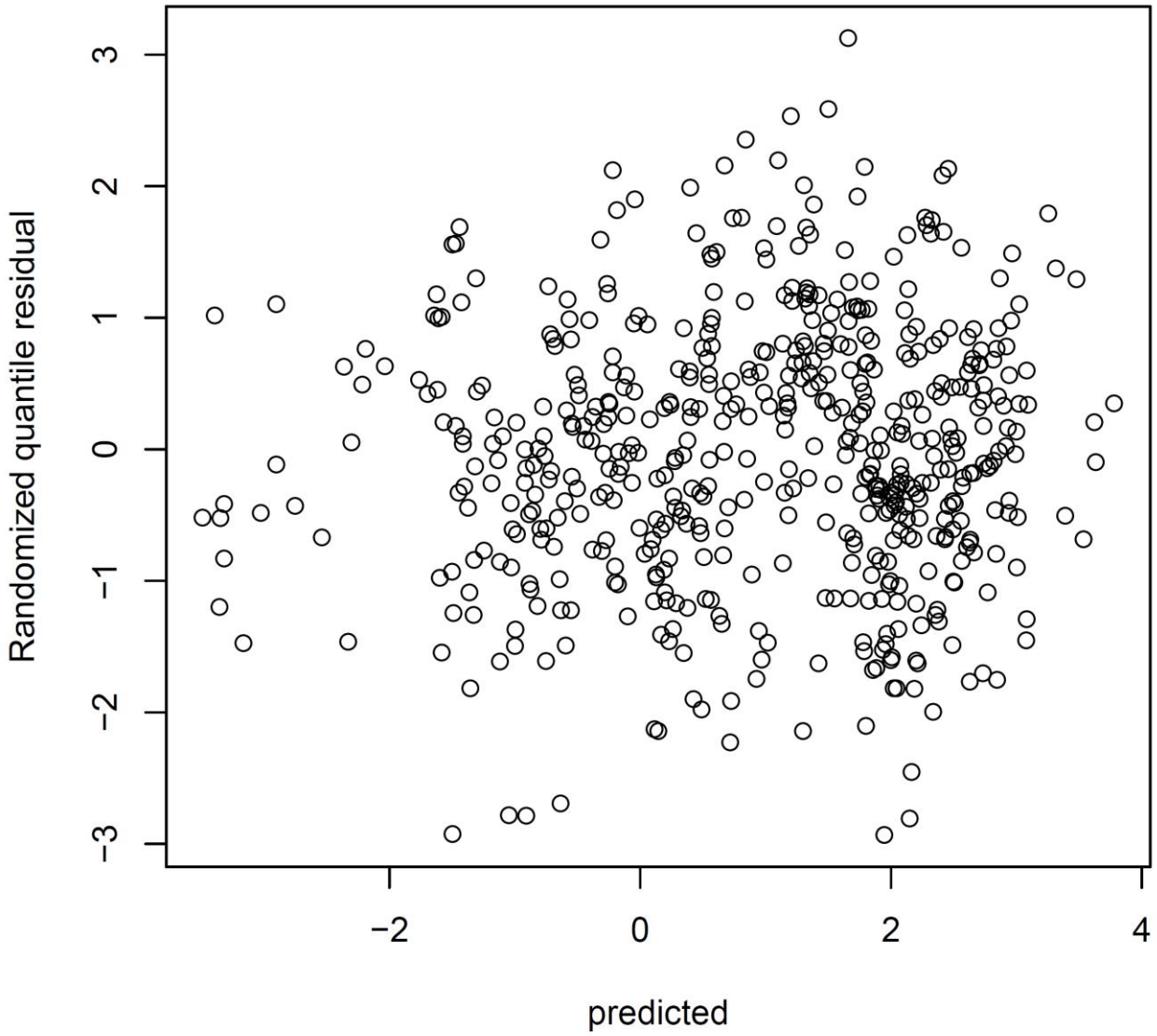

Figure S5.1b. Randomized quantile residuals versus predicted values from GAM model using average water column temperature (0-200m, n=550).

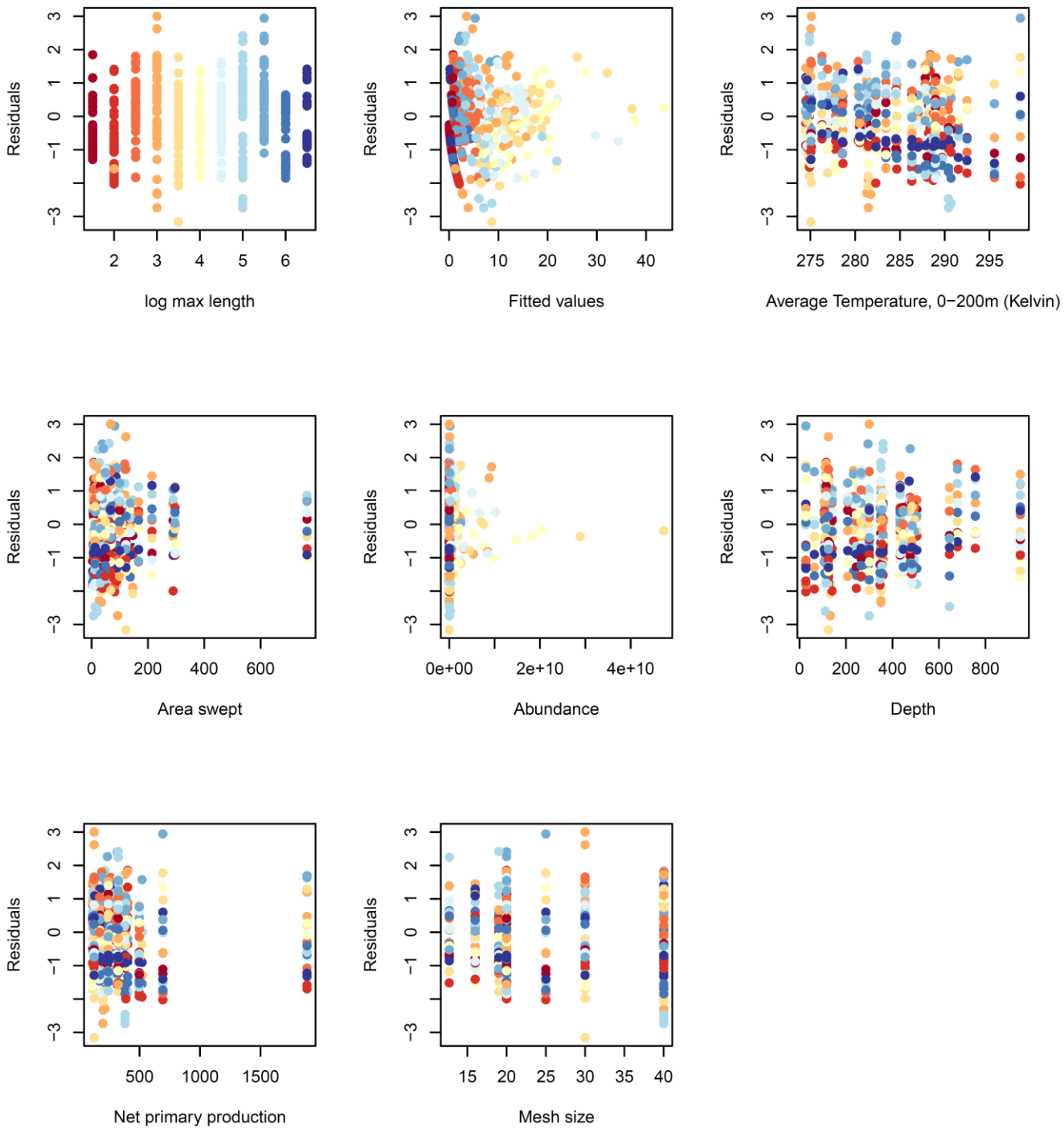

Figure S5.2. Deviance residuals from GAM model. Colours indicate maximum length group as shown in upper left panel.

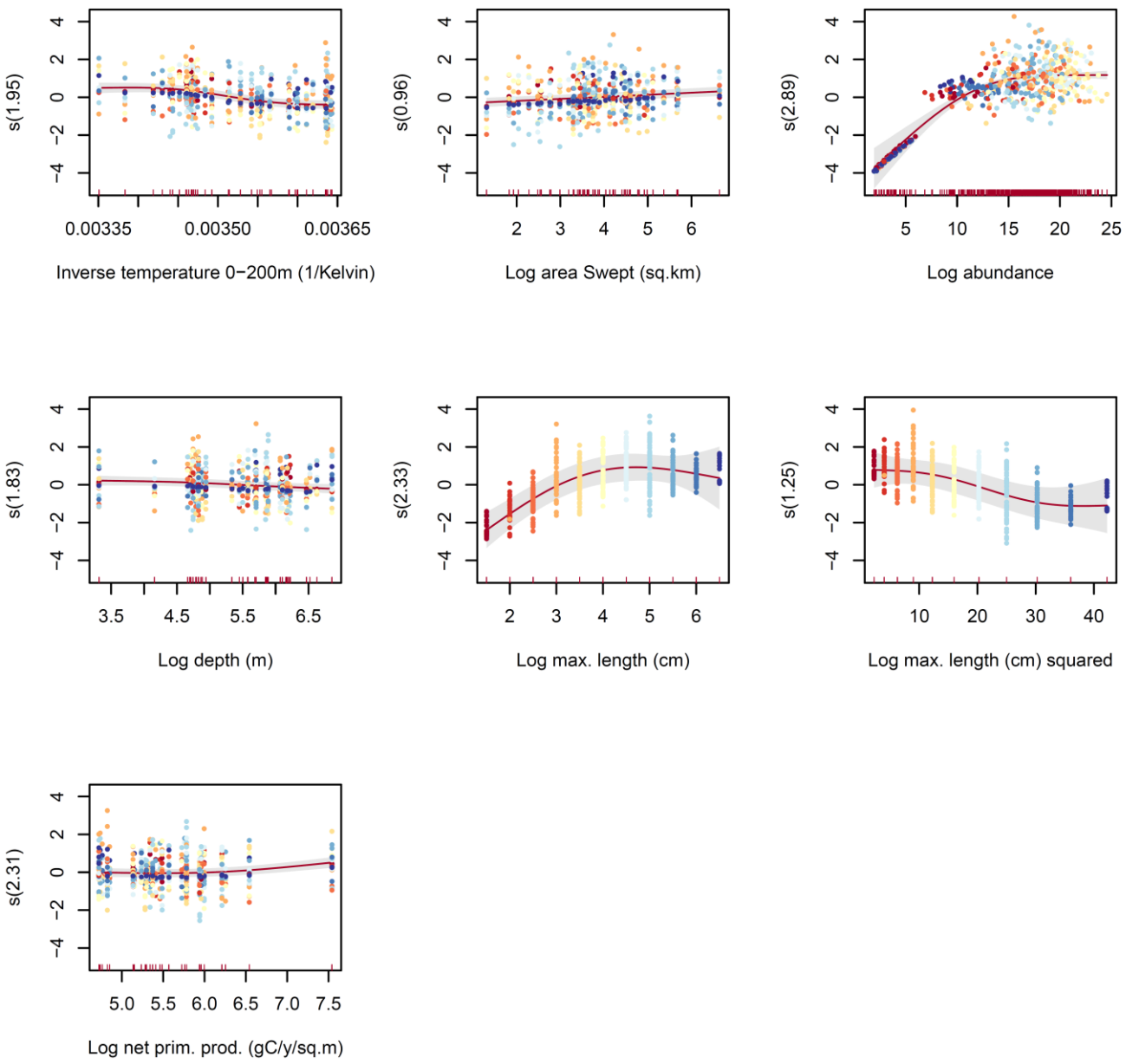

Figure S5.3 Estimated smoothing curves from the simplified GAM using the functional relationships identified in the initial GAM. Note the bias in the relationship between richness and log abundance in the panel in the upper right hand corner, where the tentative estimates of log density and log richness for the zero observations generates a strong, but false, positive relationship at low log abundance. Colours indicate maximum length group as in Figure S5.2. Gray area shows 2*standard error.

Table S5.1. Average parameter estimates from 1000 TMB-model runs of the 'best', neutral and metabolic models using randomly selected catchabilities to calculate density and abundance. Standard error in parentheses and significance levels indicated by stars (***0.001, **0.01, *0.05) (one-sided t-test, n=550). Note that the environmental model does not contain abundance or density and that the standard errors therefore are zero. NS= Non Significant term still included in the model. NSR= Non Significant term removed from the model.

Parameter		Best descriptive model	Neutral	Metabolic	Environmental
Constant	$(log\alpha)$	16.93 (0.12)***	22.91 (0.140)***	25.01 (0.24)***	3.084
Latitude	(β_0)				-0.517
Longitude	(β_1)				0.427
Temperature	(β_2)	0.322 (0.002)***	0.520 (0.003)***	0.467 (0.004)***	
Abundance	(β_3)	0.033 (0.003)***			
Density	(β_3)			0.054 (0.004)***	
Depth	(β_4)	-0.113 (0.002)***			-0.167
Net prim. prod.	(β_5)	0.217 (0.004)***			
Max. length	(β_6)			-1.001 (0.026)***	
Log. max. length^2	(β_7)	-0.131 (0.008)***	-0.128 (0.003)***		-0.235
Immigration	(λ)		NS		
Catch	(β_8)	NSR	NSR	NSR	0.067
Area swept	(β_9)	0.079 (0.003)***	0.127 (0.002)***	0.174 (0.002)***	NSR
Mesh:mlgr$_{1.5}$	$(\beta_{11,1})$	-1.538 (0.021)***	-1.320 (0.019)***	-1.682 (0.031)***	-1.070
Mesh:mlgr$_{2.0}$	$(\beta_{11,2})$	-1.380 (0.019)***	-1.179 (0.017)***	-1.428 (0.026)***	-1.021
Mesh:mlgr$_{2.5}$	$(\beta_{11,3})$	-0.979 (0.013)***	-0.860 (0.013)***	-0.976 (0.017)***	-0.754
Mesh:mlgr$_{3.0}$	$(\beta_{11,4})$	-0.601 (0.010)***	-0.500 (0.011)***	-0.557 (0.012)***	-0.458
Mesh:mlgr$_{3.5}$	$(\beta_{11,5})$	-0.404 (0.007)***	-0.330 (0.008)***	-0.352 (0.008)***	-0.340
Mesh:mlgr$_{4.0}$	$(\beta_{11,6})$	-0.225 (0.006)***	-0.162 (0.007)***	-0.170 (0.007)***	-0.192
Mesh:mlgr$_{4.5}$	$(\beta_{11,7})$	-0.097 (0.005)***	-0.040 (0.007)***	-0.053 (0.008)***	-0.074
Mesh:mlgr$_{5.0}$	$(\beta_{11,8})$	-0.064 (0.006)***	-0.011 (0.010)	-0.047 (0.009)***	-0.024
Mesh:mlgr$_{5.5}$	$(\beta_{11,9})$	-0.001 (0.008)	0.034 (0.013)**	-0.043 (0.012)***	0.052
Mesh:mlgr$_{6.0}$	$(\beta_{11,10})$	-0.050 (0.011)***	0.024 (0.008)***	-0.119 (0.014)***	0.074
Mesh:mlgr$_{6.5}$	$(\beta_{11,11})$	0.254 (0.014)***	0.302 (0.023)***	-0.070 (0.019)***	0.419
Scale parameter	$(log\kappa)$	3.765 (0.037)***	3.065 (0.036)***	3.096 (0.036)***	3.045
Proportion of deviance explained		0.900 (0.001)***	0.892 (0.001)***	0.891 (0.001)***	0.890
Pearson's R^2 (observed vs. predicted)		0.839 (0.002)***	0.788 (0.003)***	0.793 (0.003)***	0.789
AIC		1890 (2.639)***	1928 (4.010)***	1930 (3.885)***	1937

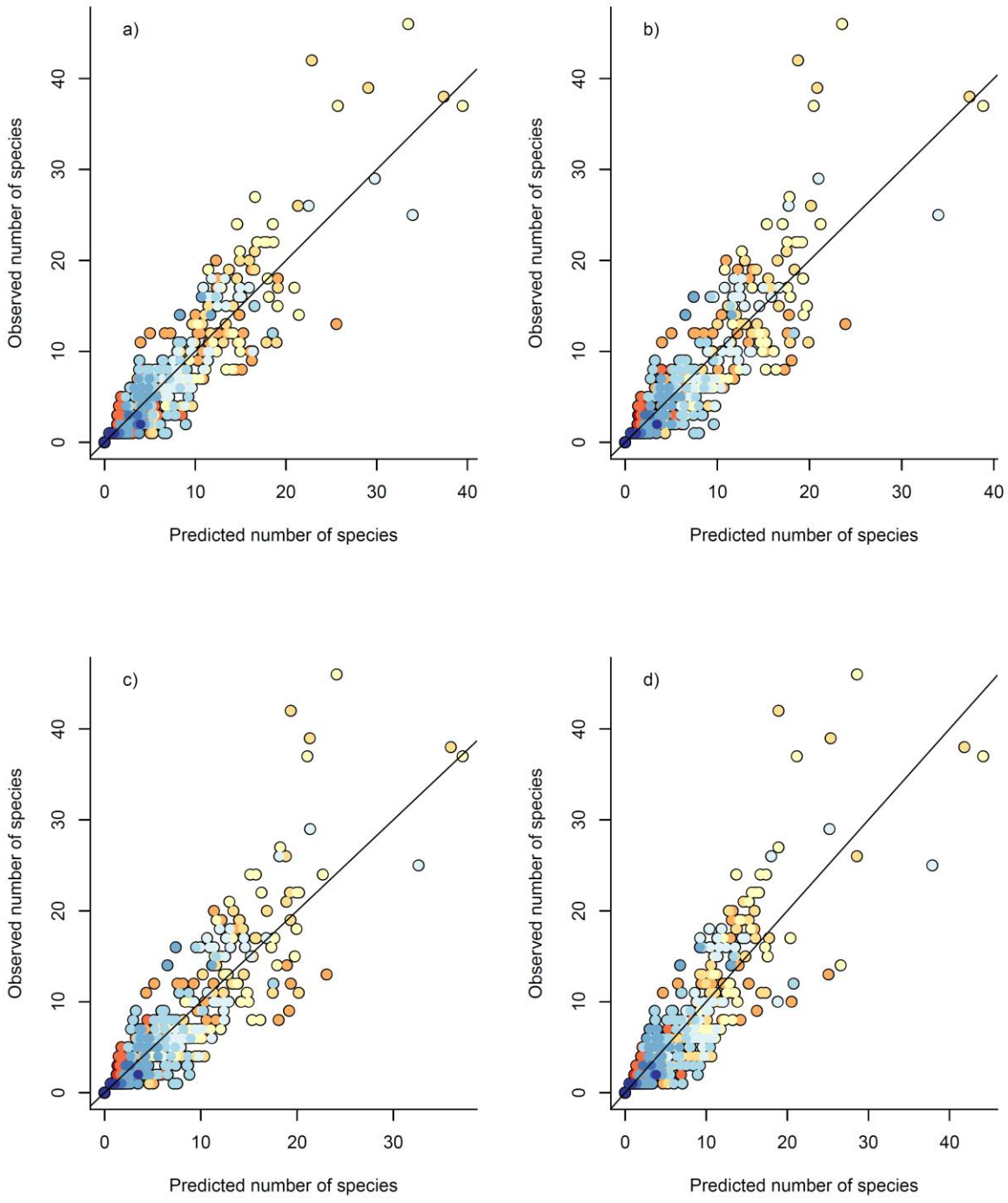

Figure S5.4. Predicted versus observed number of species per maximum length group for the TMB-fit of the four models. a) 'best' descriptive model, b) neutral model, c) metabolic model, d) environmental model. Colours indicate maximum length group (see Fig. S4.2).

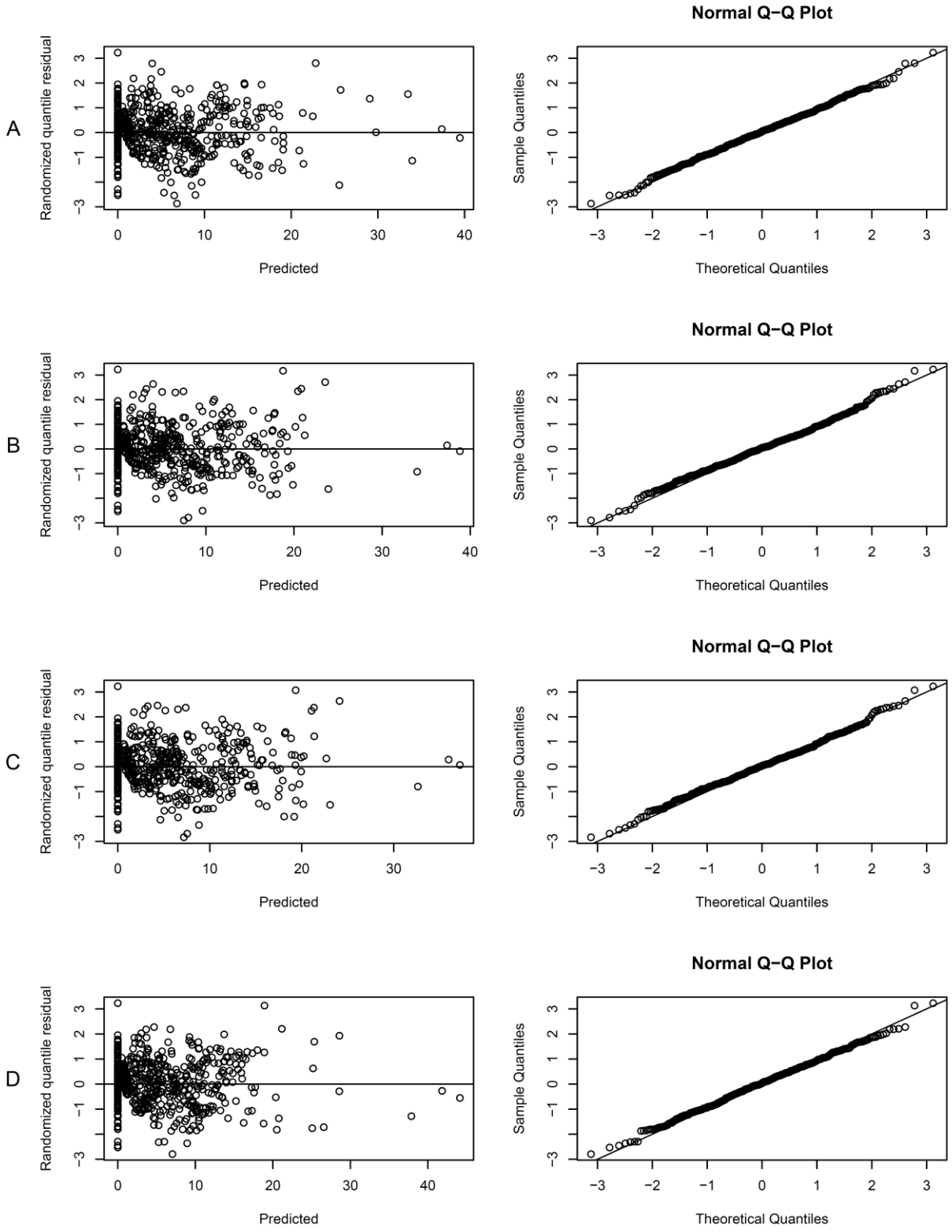

Figure S5.5. Randomised quantile residuals and Q-Q plots from TMB model fits. A) 'best' descriptive model, B) neutral model, C) metabolic model, D) environmental model.