Estimating uncertainties of diet data for use in Stochastic Multispecies Models (SMS).
Working document to ICES WGSAM October 2023

Morten Vinther, DTU Aqua.

2023-11-10

Summary

Introduction

Diet data are essential for estimating predation mortalities in multi-species models. Diet data may have been obtained from observations of stomach contents or from a qualitative estimate obtained from e.g. expert knowledge. For both types of data, it may be difficult to quantify the observation uncertainty of the diet data and the uncertainty of diet data is often ignored or estimated within the estimation model for estimating predation mortality.

SMS (Lewy and Vinther, 2004) is a stock assessment model including biological interaction estimated from a parametrised size-dependent food selection function. The model is formulated and fitted to observations of total catches, survey cpue and stomach contents (diet) for the North Sea. Parameters are estimated by maximum likelihood and the variance/covariance matrix is obtained from the Hessian matrix.

In the present SMS analysis, the following predator and prey stocks were available: predators and prey (cod, whiting, haddock), prey only (herring, sprat, northern and southern sandeel, Norway pout, plaice), predator only (saithe, mackerel), no predator–prey interactions (sole) and 'external predators' (eight species of seabirds, starry ray, grey gurnard, North Sea horse-mackerel, western horse-mackerel, hake, grey seals and harbour porpoise). The population dynamics of all species except 'external predators' were estimated within the model.

In this analysis diet data is estimated from the default method where the "population" diet is basically calculated from a stratified mean of the individual stomach content samples without an estimate of uncertainties. This diet data set is compared with diet from a new method where diet is estimated from bootstrapping of individual samples and where the uncertainties of the estimated diet are derived from fitting a Dirichlet distribution to the bootstrap replicates. The bootstrap method provides input parameters for the Dirichlet distribution applied for diet observation in the SMS, whereas the default SMS model estimates Dirichlet parameters within the SMS from an assumed relation between sampling level of stomachs and uncertainties. The results, e.g. estimated predation mortality, of the approaches are finally compared.

Data and method.

Input to the SMS includes diet data estimated from observations from around 200,000 fish stomachs primarily sampled in the period 1981-1991 (ICES XXX). Observations from each sample are available from ICES (XX). Diet data for grey seals are obtained from analysis of scats (with fish otoliths). For harbour porpoise, diet data are obtained from the stomach contents of stranded or by-caught animals. For both species of marine mammals, data are not available at sample level such that bootstrapping of samples was not possible. The

same is the case for diet data for seven individual species of sea birds, where diet data are based on expert knowledge, rather than a documented compilation of available observations into a diet composition. compiled diet data are just available for hake.

The likelihood function for diet compares the observed prey weight proportions (x_i) with the within SMS estimated values. x_i is assumed to be stochastic variables subject to sampling and process variations. For each predator entity (combinations of predator species, predator length group, year and quarter) the observations across prey entities i (e.g. sprat, herring and cod) are continuous variables which sum to one. Thus, the probability distribution of the stomach observations for a given predator including all prey groups needs to be a multivariate distribution defined on the simplex. The Dirichlet distribution is fulfilling this requirement. The probability density function for a predator entity with K preys observed in the diet proportions becomes:

$$f(x_1, x_2, ..., x_K|\alpha_1, \alpha_2, ..., \alpha_K) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{K} \Gamma(a_i)} \prod_{i=1}^{K} x_i^{(\alpha_i - 1)}$$

where K is the number of preys, x is the observed prey weight proportion and α is the model parameters such that

$$\sum_{(i=1)}^{K} x_i = 1$$

and

$$\alpha_0 = \sum_{i=1}^{K} \alpha_i$$

The mean and variance of the observations in the Dirichlet distribution are:

$$E[x_i] = \frac{\alpha_i}{\alpha_0}$$

and

$$Var[x_i] = \frac{E[x_i]\ (1 - E[x_i])}{\alpha_0 + 1}$$

Regarding the variance of stomach contents observations unpublished analyses of data from the North Sea stomach-sampling project 1991 (ICES, 1997) indicate that the relationship between variance and the mean of the stomach contents may be formulated in the following way:

$$Var[x_i] = \frac{E[x_i]\ (1 - E[x_i])}{V_j\ U_{j,y,q}}$$

where U_j is a known quantity reflecting the sampling level of stomachs for predator j, predator size class l, in year y, and quarter q and V_j is a predator specific parameter estimated within SMS. The two equations for variance imply that:

$$\alpha_0 = V_j\ U_{j,l,y,q} - 1$$

Estimating α parametres

The compilation of the individual stomach samples from e.g. trawl hauls into the average diet of the North Sea predators basically follows the technique given by ICES (1993). The average "population" diet or food ration is basically calculated from a stratified mean of the individual stomach content samples, weighted by the strata density of the predator and the area of the strata. This seems simple, but incomplete and patchy sampling makes it often necessary to use a series of *ad hoc* solutions. The compilation of stomach

contents for the 2023 keyrun was done using the Fish-Stomachs R-package (available from (https://github.com/MortenVinther/FishStomachs).

The FishStomachs package defines data structures suitable for stomach data and provides the necessary methods to compile observed stomach data into population diet and biomass eaten, used for multispecies models. The methods applied for a set of observations are stored within the data output to document the compilation steps taken.

The stomach contents compilation followed the steps outlined below:

1. Read and check data from the agreed exchange format;
2. Bias correct to take into account variable evacuation rate;
3. Assign size classes for predators and preys;
4. Bias correct to take into account regurgitated stomachs within sample units;
5. Aggregate stomach contents within sample_id and size classes.
6. Allocate unidentified or partly identified prey items;
7. Calculate the population diet and food ration from a weighted average.

The FishStomachs package makes it possible to estimate uncertainties of the estimated diet from bootstrapping of individual samples. Bootstrapping is made between step 4 and 5 in the steps above. First, a set of 500 bootstrap replicates are made from random sampling with replacement of the individual stomach samples (*i.e.* trawl hauls). The diet is then estimated for each replicate (step 5-7 above), such that a set of 500 replicates of diets are produced. The distribution of diet replicates is finally fitted to a Dirichlet distribution (using function diri.est in R-package Compositional) for estimation of the α parameters. Figure 1 shows an example where the bootstrap replicates fit quite well with observations for Dirichlet distribution. Another example figure 2 with much fewer stomach samples shows in some cases a two topped distribution of the bootstrap replicates and a poor fit to the estimated Dirichlet distribution with a low α_0 value.

Effect on SMS results

Three runs with the SMS model were done to explore the effect of using input values for uncertainty on diet data:

- **Default**, diet data are compiled without estimation of uncertainties and SMS estimates α_0 from an assumed relation between number of stomach samples and uncertainty.
- $\alpha_{\mathbf{prey}}$, diet data are estimated from a Dirichlet fit to bootstrap replicates. α_0 and prey proportions (α_{prey}/α_0) are used as input to SMS.
- $\alpha_{\mathbf{0}}$, as above with the use of input α_0, but prey proportions are taken from the default configuration.

All the SMS configurations were configured with a maximum α_0 at 5 for the seven bird species to constrain the influence of the rather uncertain estimate of bird diets from expert knowledge. The uncertainties of diet data for grey seal and harbour porpoise were estimated within SMS from the assumed sampling level.

Output from SMS is substantial and this document only presents results for cod (predator and prey) and herring (prey only). The results for these two species reflect well the difference in results for the other not shown species.

Results

Effects on estimated diet

The estimated α_0 depends on the sampling level, the predator species and the number of preys for a given predator entity (figure 3. The median value of α_0 and thereby the precision of the diet estimate is highest

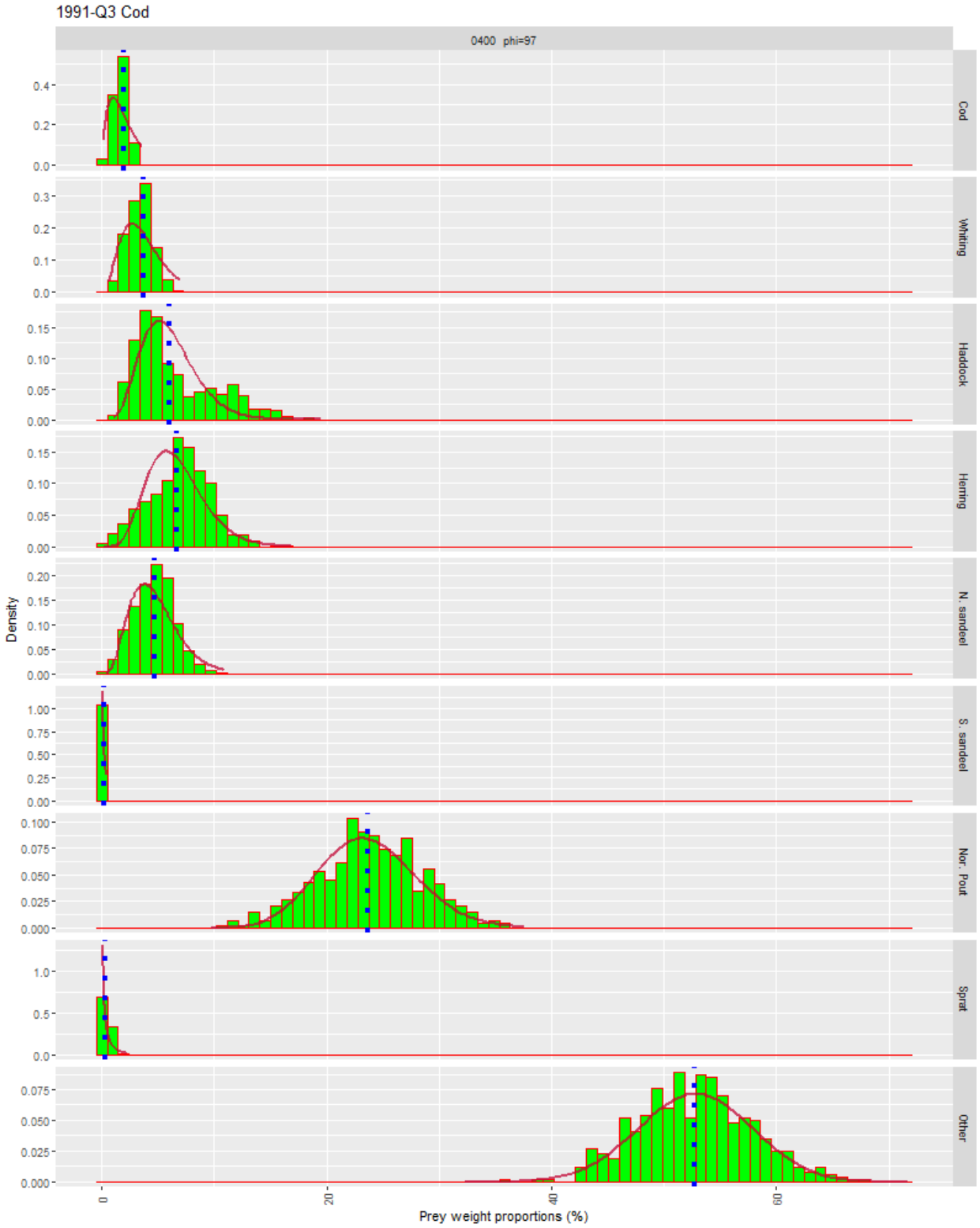

Figure 1: Bootstrap replicates of diet weight proportions for predator cod 40-50 cm in quarter 3 of 1991. The red curve shows the fitted Dirichlet distribution, the blue line shows the average weight proportion of the full (non-bootstrapped) dataset. The fitted concentration parameter (or α_0) is shown at the top panel as phi.

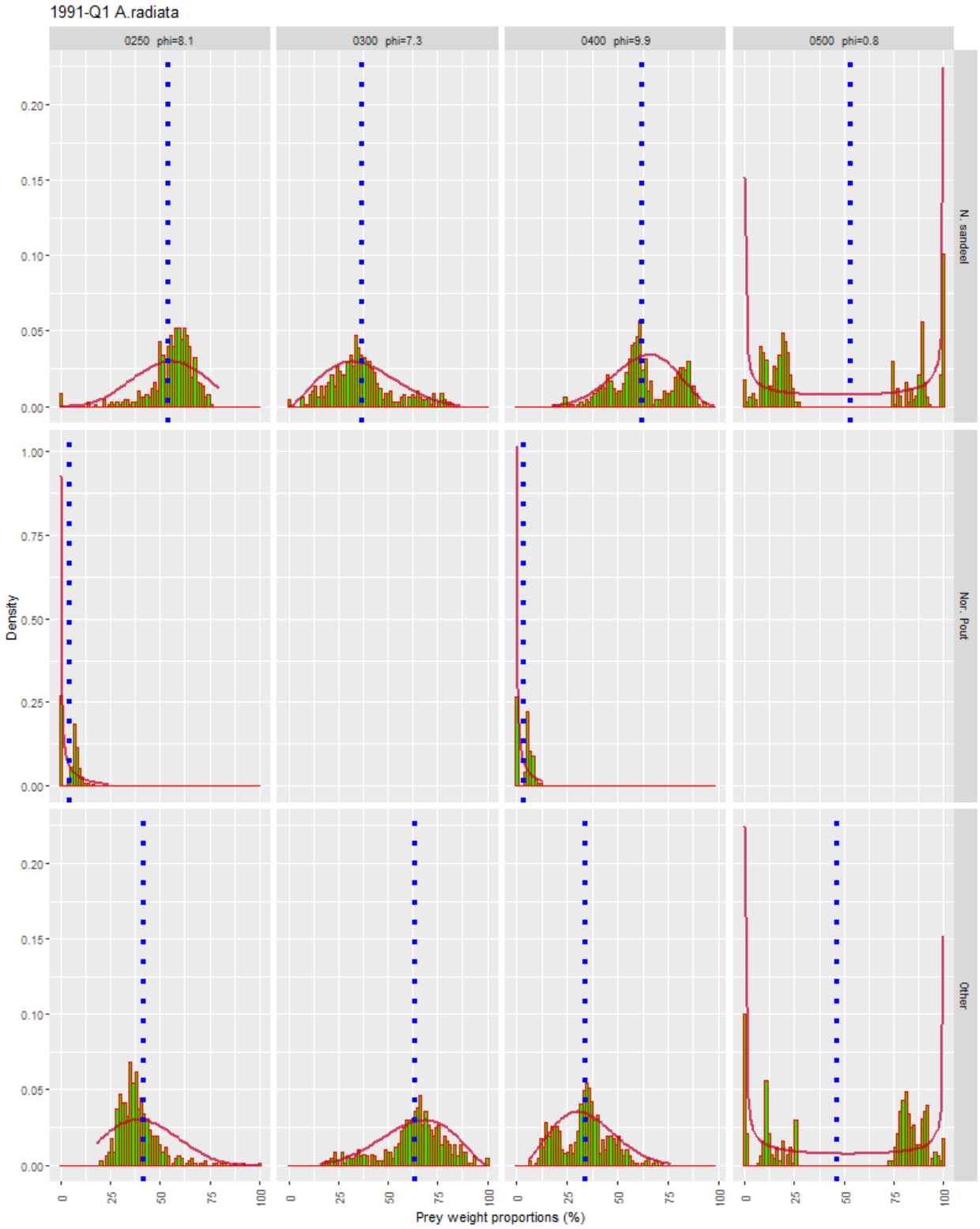

Figure 2: Bootstrap replicates of diet weight proportions for predator *Amblyraja radiata* in quarter 1 of 1991. The red curve shows the fitted Dirichlet distribution, the blue lines show the average weight proportion of the full (non-bootstrapped) dataset. The fitted concentration parameters (α_0) are shown at the top of the panel as value phi.

for cod and whiting followed by haddock and saithe. This corresponds well to the sampling level for these predators,

The bootstrap estimates of α_0 (figure 3) are in general higher than the values estimated within the SMS model (figure 4). A considerable increase in the median α_0 from bootstrap is seen for cod and whiting (table 1), while the bootstrap method gives a smaller α_0 for saithe and mackerel.

The prey weight proportions estimated from the two methods are, as they should be, highly correlated (figure 5). There are however examples where the two estimates differ quite at lot. The "other food" prey constitutes a high diet proportion for most predators and there is a tendency that the proportions estimated from α_{prey} are higher than the simple approach (figure 6). As the diet proportions sum up to one, the weight proportion from named preys become smaller when estimated from α_{prey}. This can also be seen for some predator-prey combinations in figure 5, even though the bias is not large. A closer look at diet data from cod for diet data shows the difference in prey proportion seems to depend on the prey proportion. A small prey proportion (e.g. 1-2 %) estimated from α_{prey} is in general higher than the simple estimate (figure 7).

Effects on SMS results

The overall effect of the choice of diet data and method seems limited based on the assessment output, recruitment, average fishing mortality (F) and spawning stock biomass (SSB). For both cod (figure 8) and herring (figure 9). the largest difference is for recruitment. For both cod and herring, and the other species not presented, there is a tendency that the "alpha prey" SMS configurations provide the lowest estimate of recruitment. Estimated recruitment is influenced by the predation mortality (M2) at age. A closer look at the M2 (figure 10) and (figure 11) reveals quite similar results for the three SMS configurations. It is also seen that the ratio of M2 at age between configurations is not the same for all ages, such that e.g. the "alpha prey" configuration for herring provides the lowest M2 at age 0, but the highest M2 for ages 2-4.

The SMS likelihood statistics (table 2) for the three configurations show that the "default" configuration gives the best (lowest negative log likelihood value) followed by the "alpha 0" and "alpha prey" SMS. The largest differences in log-likelihood are for diet data, where the individual likelihood contributions by predator are best for the "default" configuration. The same pattern is seen for the likelihood values for catch and CPUE, even though there are few exceptions, e.g. the likelihood of cod CPUE is best when the uncertainties of diet data are given as input ("alpha prey" and "alpha 0").

Discussion

SMS is a model with likelihood contributions from both catch, cpue, stock-recruitment and diet observation. The stock-recruitment likelihoods are down-weighted (factor 0.1) within SMS as both recruitment and SSB are estimated within the model and as such not observations to the model. The remaining three likelihood components have no *a priori* weighting, such that the overall model fit and weighting of the data sources are done from the total likelihood of the model. A catch-at-age observation fits, in general, better than a CPUE observation. Diet observations have in general the poorest fit which might explain the rather stable estimates of F and SSB and to some extent also recruitment, even though diet data are changed considerably.

The initial testing of the SMS model with artificial input data with known variance and known model parameters showed that the model is able to estimate model parameters to the correct values, if the variance of input data was not too high. The were however problems in estimating the parameters that link α_0 to the sampling level of diet data (the V_j parameters in the relation between sampling level and variance). The same is seen in several SMS runs, where this parameter in some cases only can be estimated if the parameter reaches an input bound for one of the predators. The bound for cod and two bird species was e.g. reached in the 2020 Key run for the North Sea. This suggests that input values for α_0 are advantageous to fix the variance of diet data.

The prey proportions estimated from the Dirichlet α values differ in some cases quite a lot from the prey proportions estimated the default way. This difference seems largest for poorly sampled diets, however for

even a species like cod with large sample sizes, there seems to be a bias, where e.g. small (around <2%) prey proportions become higher when the bootstrap method is applied. If this bias is due to the bootstrap itself or due to the estimation of the α parameters needs to be investigated. A way to circumvent this bias is to use the estimated α_0 to scale the default prey proportions (estimated without bootstrapping). Likelihood statistics from the "alpha prey" and "alpha 0" configurations are however quite the same, even though the estimated M2 values may vary slightly between the two configurations.

The bootstrap method provides higher α_0 and for e.g. cod and whiting than the default method (table 1, and figure 3 and 4). Likewise, the bootstrap method provides higher α_{prey} figure 12), but the diet likelihoods (figure 13) are not better. The default method estimates an α_0 (or actually a parameter, V_{pred}, to estimate α_0 from sampling level) that gives the best total likelihood for all diet observations from the given predator. To handle the in general rather poor fit between observed and estimated diet, it seems like the optimization ends up with a low V_{pred} and thereby a lower α_0 and α_{prey} for all diet data for the predator. The approach where α_0 is provided as input for each predator entity, will in some cases where bootstrapping estimates a low observation variance, resulting in a poorer likelihood either because the bootstrap estimate of variance is biased (too low) or due to process uncertainties (the model for diet is not adequate to model predation, and produces large residuals for diets observation). Providing α_0 estimated externally to SMS reduces the number of estimated parameters and is a first step in separating observation and process uncertainties.

Some diet entities are based on only a few samples which creates a two-topped distribution of the bootstrap replicates (see figure 2). The α_0 value estimated becomes very low for these cases such the diet input gets a low weight in minimizing the total model likelihood. It could be argued that such diet based on only a few samples should not be used by SMS, however providing input values for the accuracy of the diet entity limits the risk of overfitting.

Table 1: Median α_0 estimated within SMS and from bootstrapping.

Predator	SMS	bootstrap
A. radiata	9.8	4.7
Grey gurnard	8.9	11.1
Western horse mackerel	3.1	2.4
North Sea horse mackerel	10.4	1.2
Cod	24.4	42.1
Whiting	14.8	30.4
Haddock	12.0	14.2
Saithe	16.0	13.7
Mackerel	15.8	11.1

Table 2: Negative log likelihood from catch, Cpue, stock-recruitment, diet observation and total from SMS where α_0 is estimated by bootstrapping (label bootstrap) or within SMS (label simple). The likelihood contributions from 8 bird species, hake, plaice, sole and marine mammals are not shown as α_0 is estimated within SMS and therefore almost identical between runs, but included in the totals.

SMS conf.	Species	Catch	CPUE	SSB Recruit	Diet	neg log likelihood
default	A.radiata	0.0	0.0	0.0	-44.5	-44.5
default	G.gurnards	0.0	0.0	0.0	-49.3	-49.3
default	W.horse.mac	0.0	0.0	0.0	1.8	1.8
default	N.horse.mac	0.0	0.0	0.0	-11.4	-11.4
default	Cod	-447.9	-137.1	-8.9	-1629.7	-2222.7
default	Whiting	-266.5	-171.4	-33.6	-735.1	-1176.3
default	Haddock	-134.2	-176.3	17.9	-77.5	-386.3
default	Saithe	-326.3	-74.3	-22.5	-84.8	-507.9
default	Mackerel	-457.6	-76.1	-8.2	-102.8	-644.7
default	Herring	266.2	-194.9	-11.1	0.0	70.2
default	N.sandeel	149.6	49.2	13.1	0.0	200.1
default	S.sandeel	100.3	-19.6	1.7	0.0	80.9
default	Nor.pout	269.8	-44.9	-8.9	0.0	224.0
default	Sprat	221.7	-53.4	-5.5	0.0	167.8
default	All	-1472.0	-1018.8	-100.0	-4485.0	-7020.5
alpha prey	A.radiata	0.0	0.0	0.0	-35.0	-35.0
alpha prey	G.gurnards	0.0	0.0	0.0	-15.8	-15.8
alpha prey	W.horse.mac	0.0	0.0	0.0	2.3	2.3
alpha prey	N.horse.mac	0.0	0.0	0.0	-11.3	-11.3
alpha prey	Cod	-443.8	-138.4	-8.5	-1477.2	-2067.1
alpha prey	Whiting	-259.9	-173.0	-31.9	-530.2	-966.4
alpha prey	Haddock	-124.4	-181.0	18.2	-58.8	-362.3
alpha prey	Saithe	-322.5	-73.4	-23.3	-14.1	-433.4
alpha prey	Mackerel	-457.9	-75.7	-8.1	-84.7	-626.4
alpha prey	Herring	266.0	-192.7	-12.1	0.0	72.1
alpha prey	N.sandeel	156.4	54.2	12.8	0.0	211.9
alpha prey	S.sandeel	107.2	-18.9	1.7	0.0	88.4
alpha prey	Nor.pout	283.7	-38.4	-7.8	0.0	244.5
alpha prey	Sprat	223.3	-53.2	-2.9	0.0	169.8
alpha prey	All	-1419.2	-1010.3	-96.0	-3981.9	-6456.1
alpha 0	A.radiata	0.0	0.0	0.0	-32.1	-32.1
alpha 0	G.gurnards	0.0	0.0	0.0	-23.3	-23.3
alpha 0	W.horse.mac	0.0	0.0	0.0	4.9	4.9
alpha 0	N.horse.mac	0.0	0.0	0.0	-5.4	-5.4
alpha 0	Cod	-444.2	-138.1	-8.7	-1399.6	-1989.7
alpha 0	Whiting	-260.3	-172.9	-31.5	-577.4	-1013.7
alpha 0	Haddock	-127.8	-177.6	17.6	-43.8	-347.4
alpha 0	Saithe	-323.0	-73.5	-23.1	3.6	-416.1
alpha 0	Mackerel	-457.8	-75.8	-8.1	-84.6	-626.3
alpha 0	Herring	265.0	-194.9	-11.6	0.0	68.9
alpha 0	N.sandeel	156.3	53.8	12.1	0.0	211.2
alpha 0	S.sandeel	108.2	-20.1	2.0	0.0	88.2
alpha 0	Nor.pout	279.0	-42.6	-8.5	0.0	235.6
alpha 0	Sprat	221.1	-52.6	-3.7	0.0	168.1
alpha 0	All	-1430.6	-1014.4	-97.4	-3910.8	-6400.6

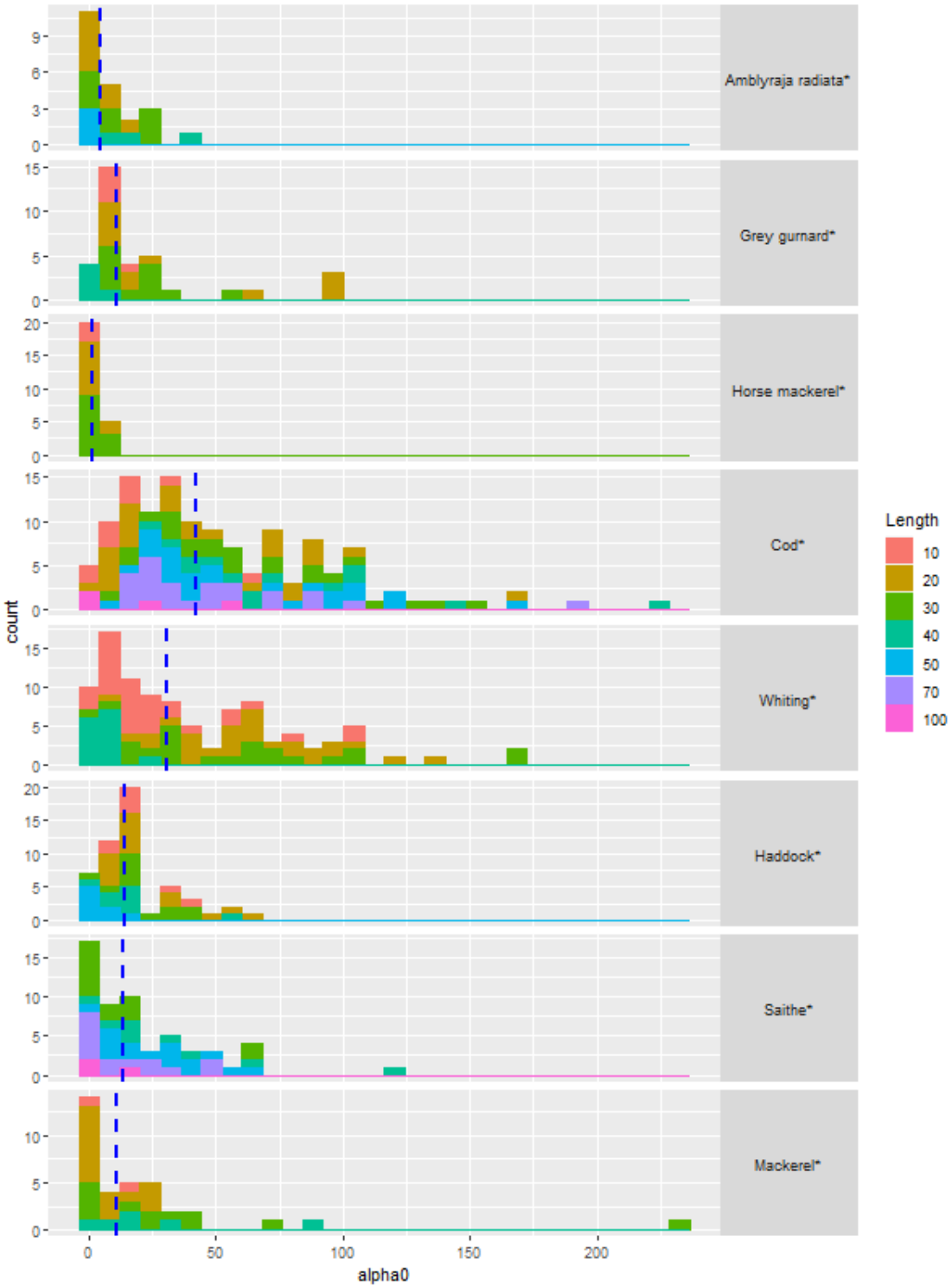

Figure 3: Histogram of estimated α_0 from bootstrapping by predator species and predator length classes. The blue lines show the median α_0.

Figure 4: Histogram of within SMS estimated α_0 by predator species and predator length classes. The blue line shows the median α_0.

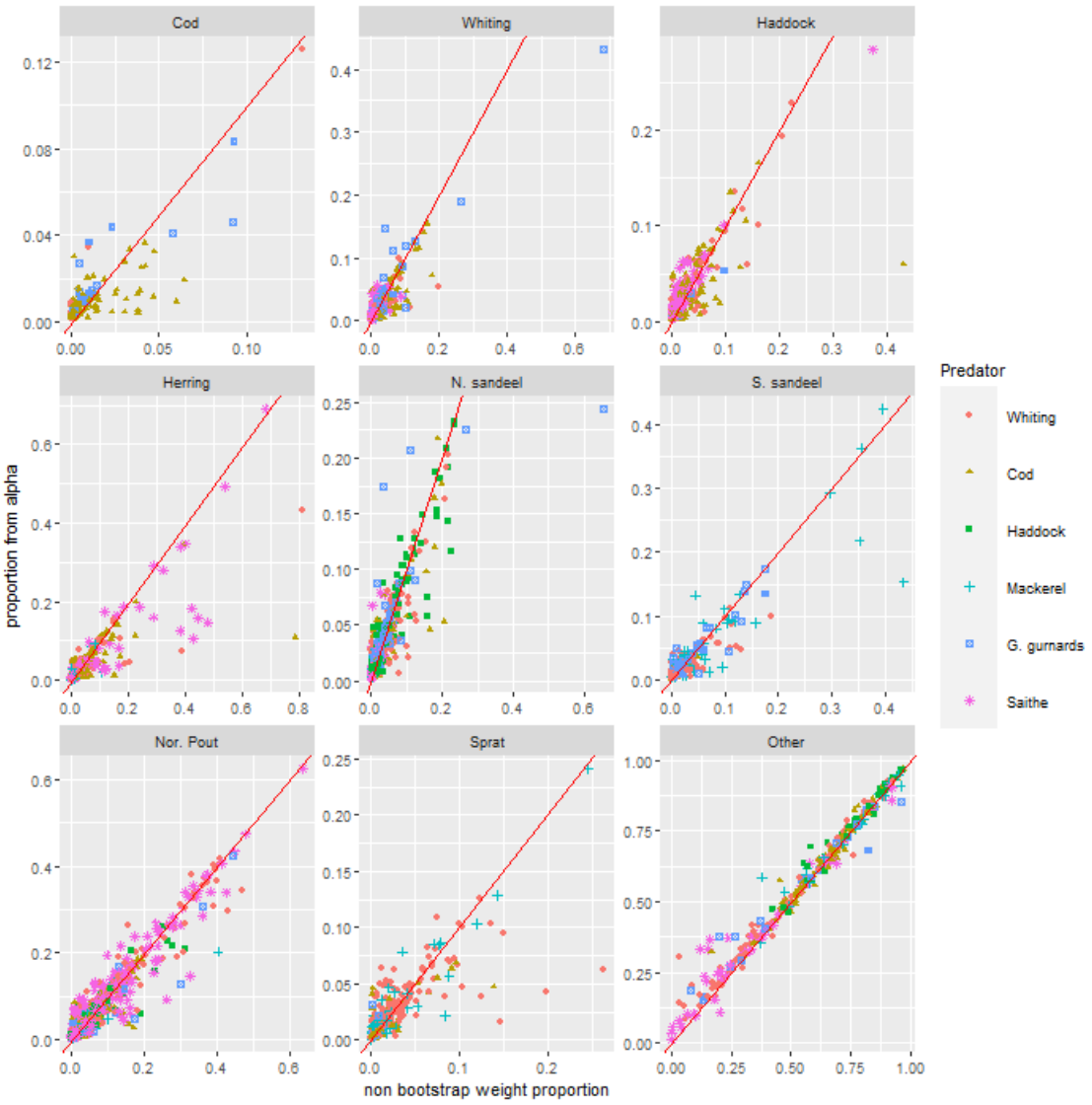

Figure 5: Weight proportion in the diet by prey species and predator estimated from non-bootstrapped data (x-axis) against prey proportion estimated from bootstrapping derived from Dirichlet α_{prey} values. The red lines have slope 1.

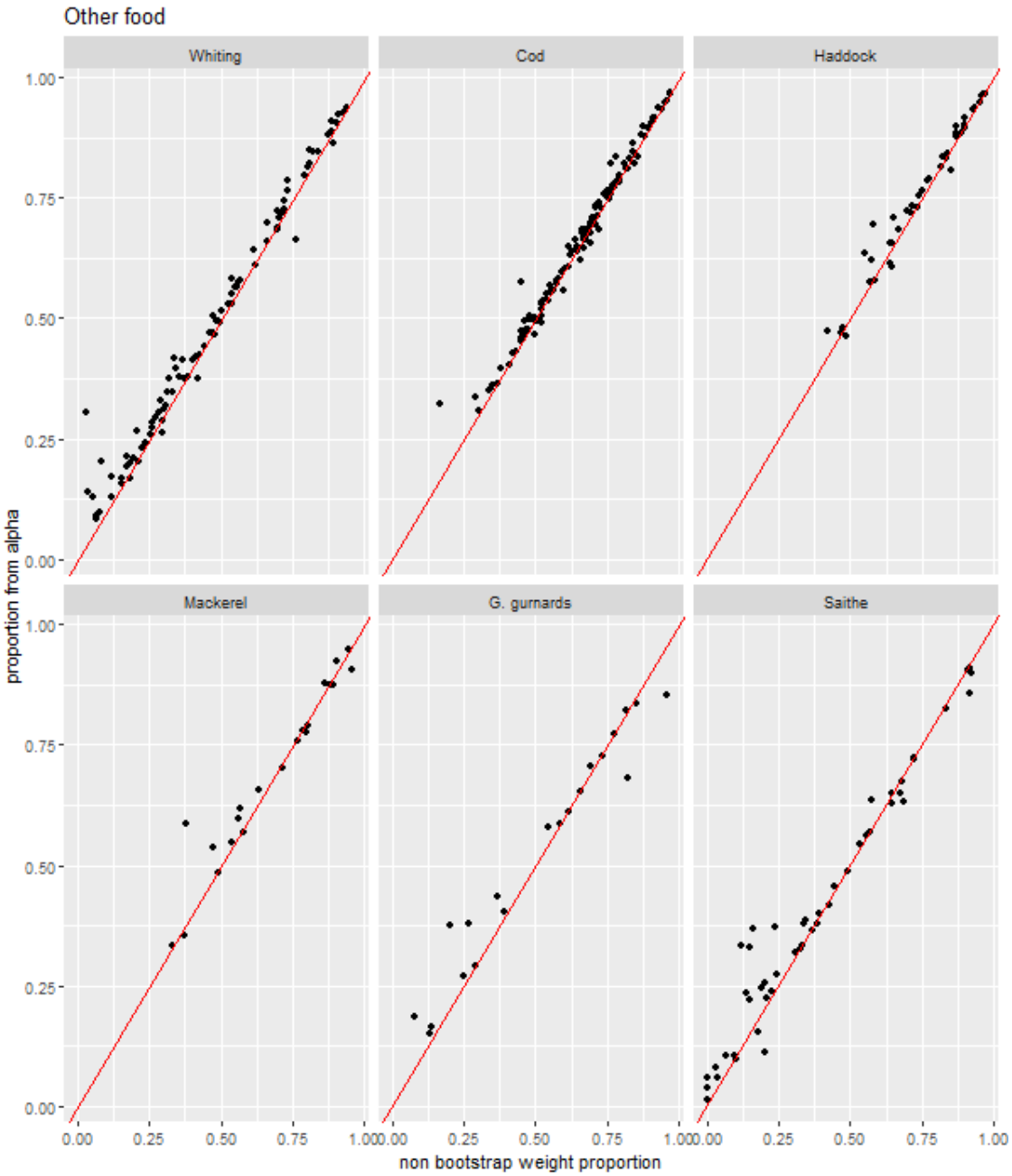

Figure 6: Weight proportion in the diet of 'other food' by predator estimated from non-bootstrapped data (x-axis) against prey proportion estimated from bootstrapping derived from Dirichlet α_{prey} values. The red lines have slope 1.

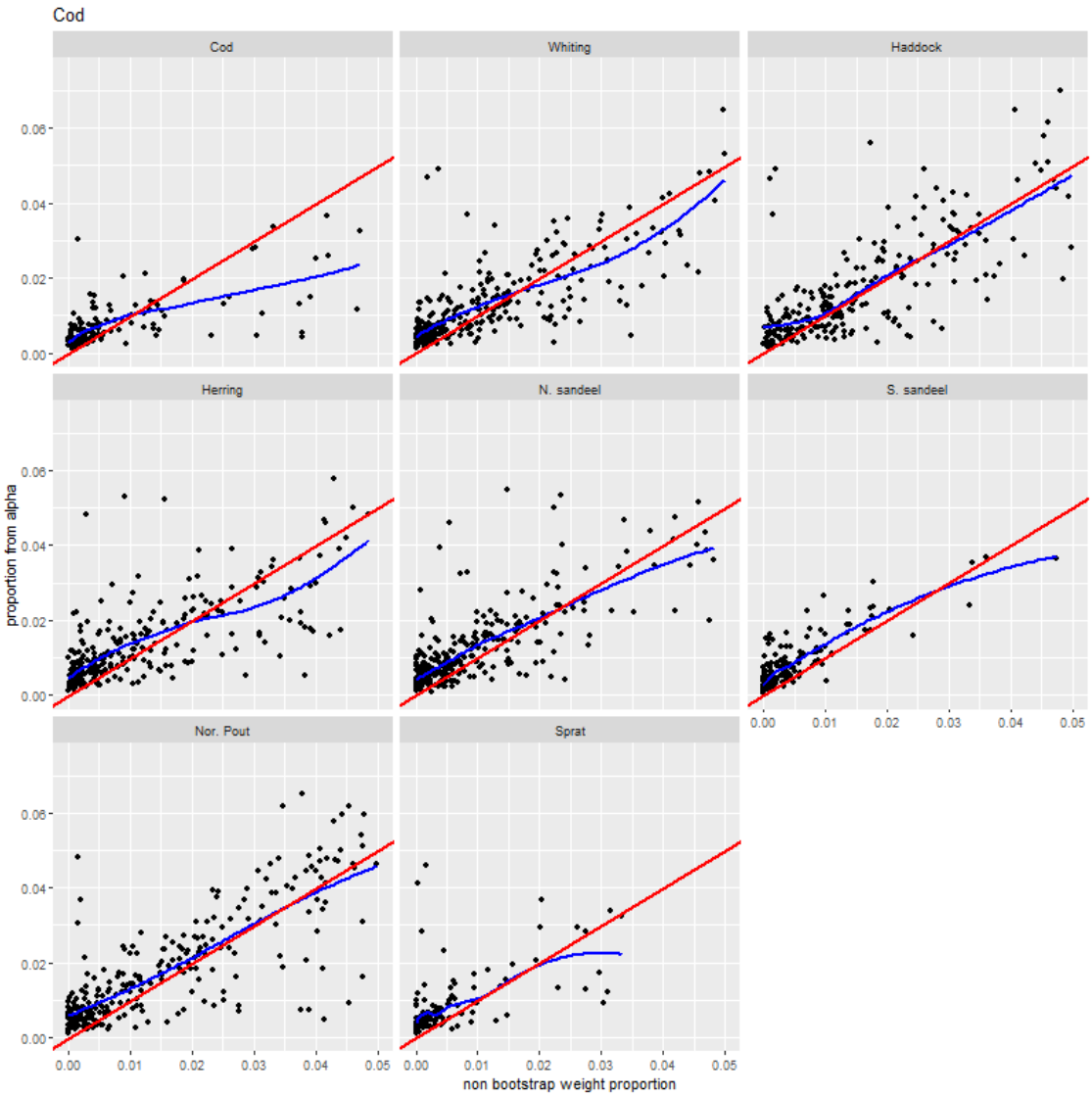

Figure 7: Weight proportion in the diet by prey species of cod estimated from non-bootstrapped data (x-axis) against prey proportion estimated from bootstrapping derived from Dirichlet α_{prey} values, for preys with less than 0.05 observed weight proportion in the diet. The red lines have slope 1.

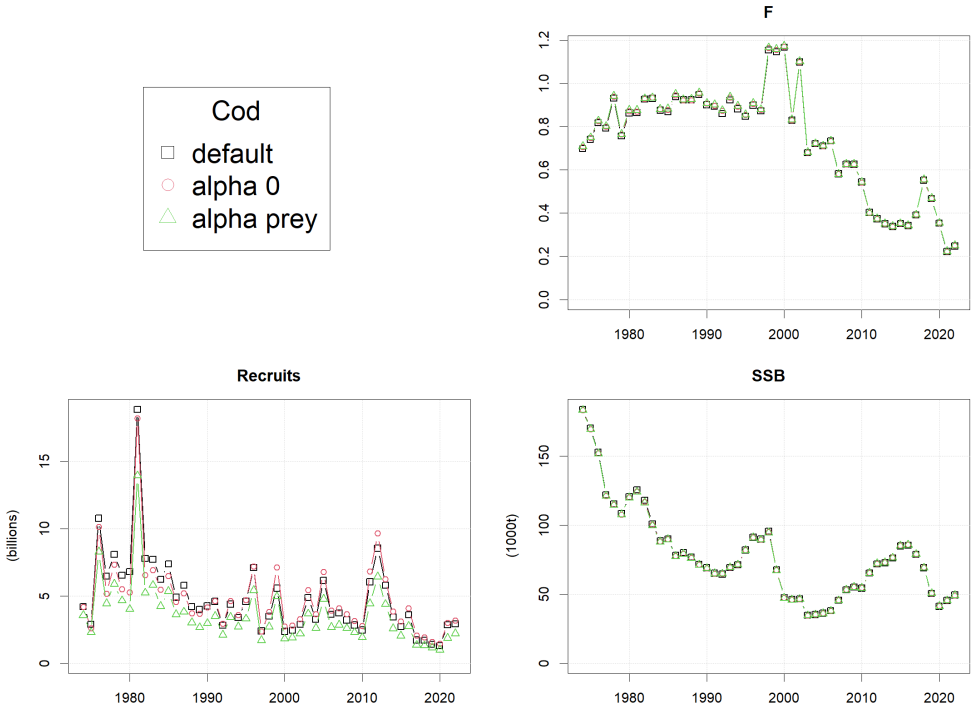

Figure 8: Main assessment results for cod from SMS configurations (see text for labels).

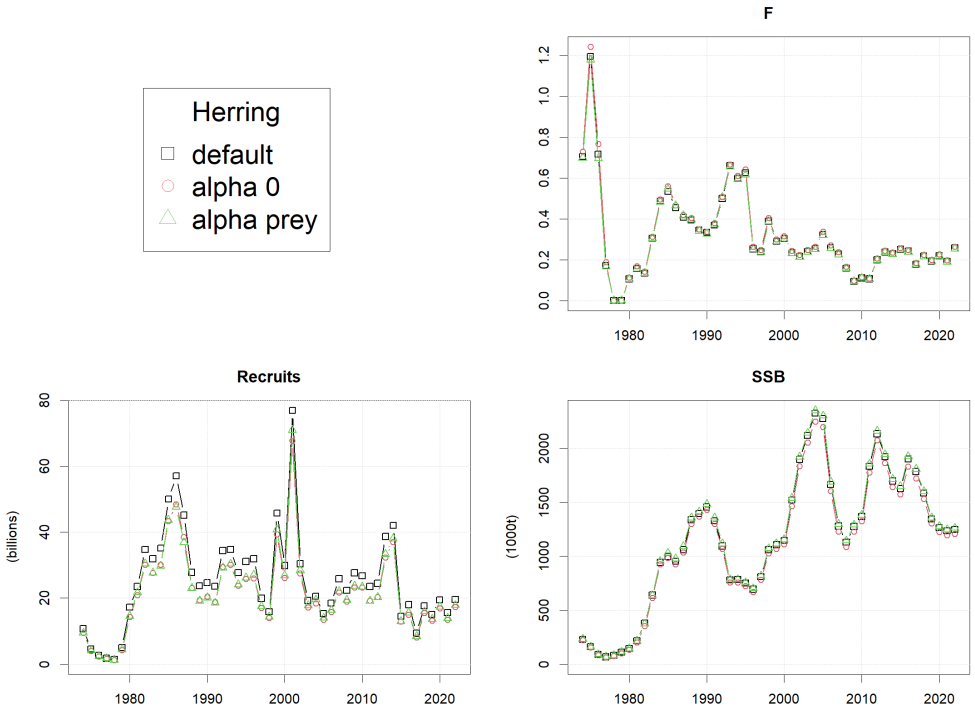

Figure 9: Main assessment results for herring from SMS configurations.

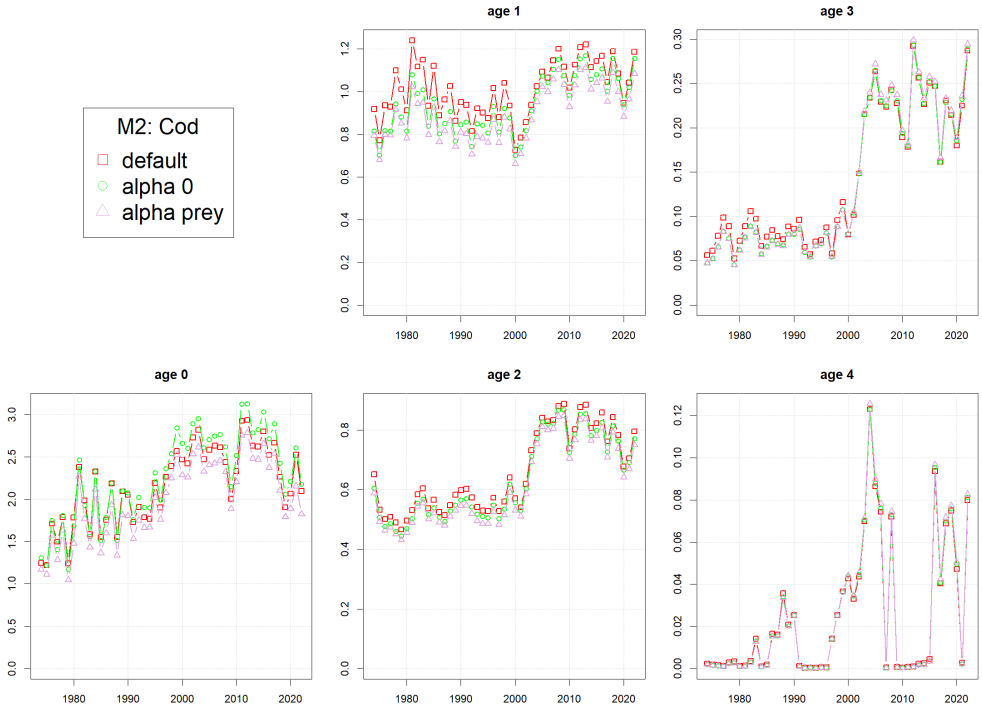

Figure 10: M2 by age for cod from SMS configurations.

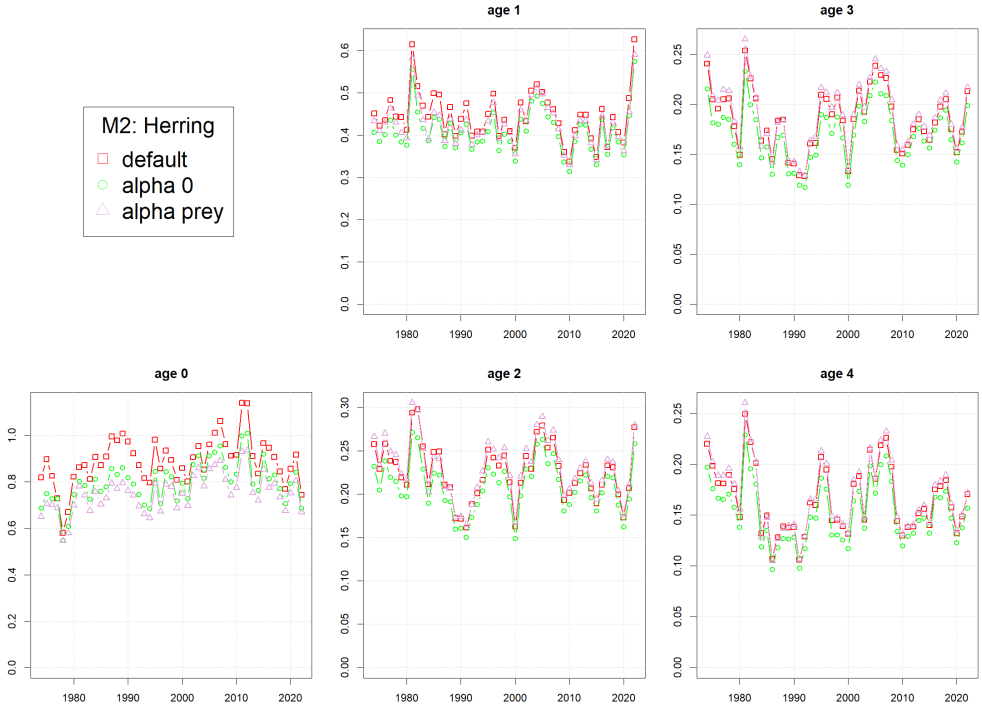

Figure 11: M2 by age for herring from SMS configurations.

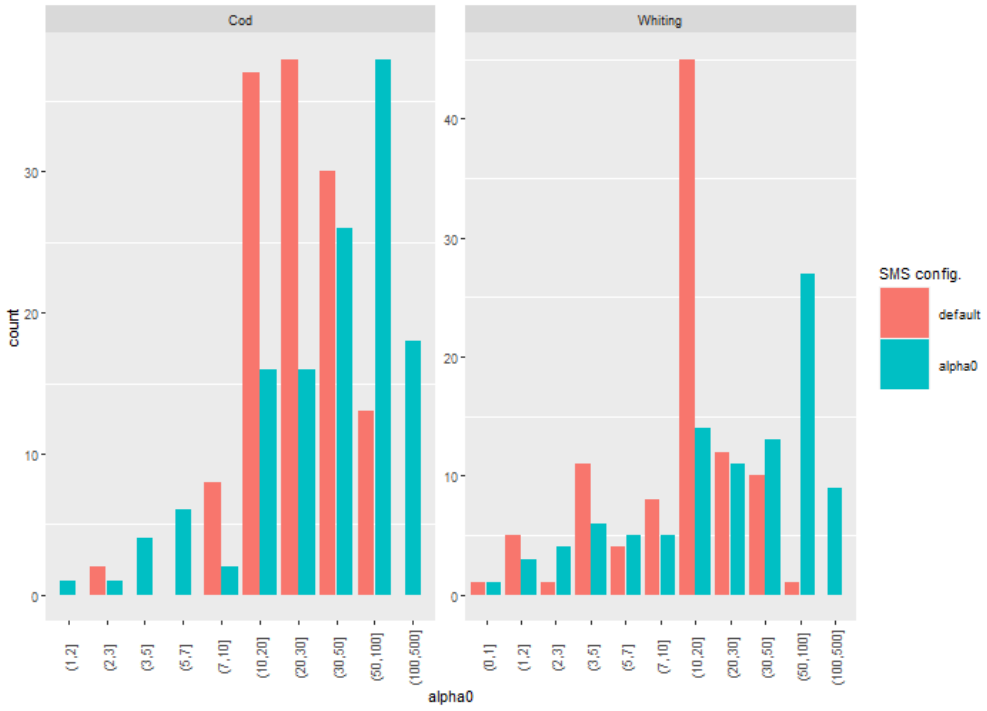

Figure 12: Histogram of the α_{prey} for diet observation used by SMS configurations.

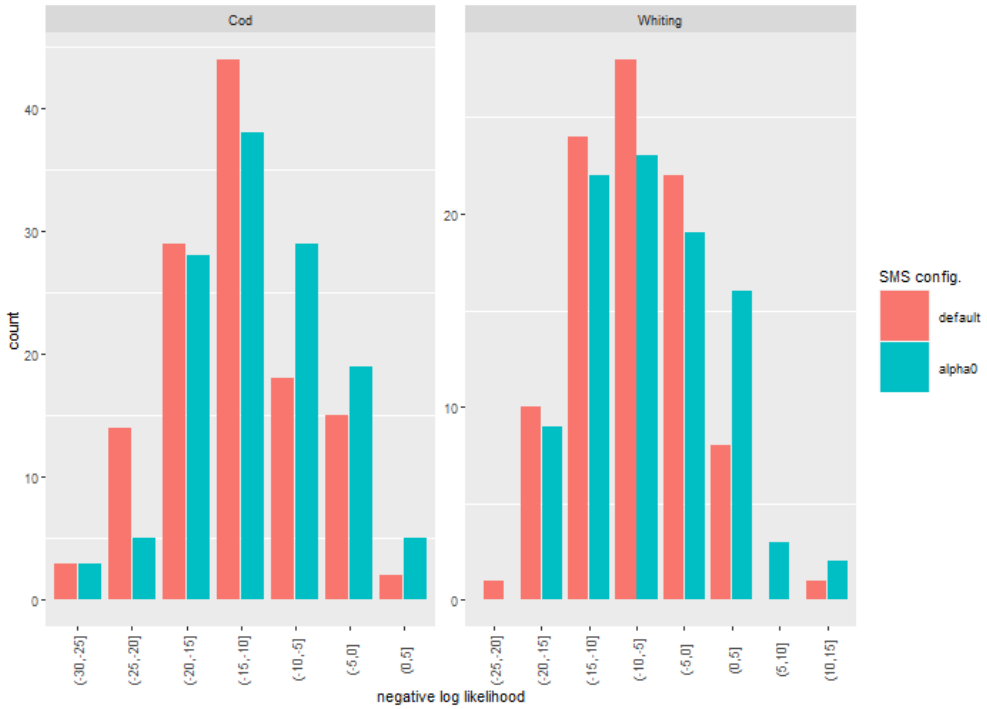

Figure 13: Histogram of negative log liklihood contibutions for diet observation estimated by SMS configurations.