

Intercalibration of survey methods using paired fishing operations and log-Gaussian Cox processes

Thygesen, Uffe Høgsbro; Kristensen, Kasper; Jansen, Teunis; Beyer, Jan E

Published in: ICES Journal of Marine Science

Link to article, DOI: 10.1093/icesjms/fsy191

Publication date: 2019

Document Version Peer reviewed version

Link back to DTU Orbit

Citation (APA): Thygesen, U. H., Kristensen, K., Jansen, T., & Beyer, J. E. (2019). Intercalibration of survey methods using paired fishing operations and log-Gaussian Cox processes. *ICES Journal of Marine Science*, *76*(4), 1189-1199. https://doi.org/10.1093/icesjms/fsy191

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.

- · You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

1	Intercalibration of survey methods using paired fishing
2	operations and log-Gaussian Cox processes
3	Uffe Høgsbro Thygesen
4	Technical University of Denmark, Department of Mathematics and
5	Computer Science, 2800 Lyngby, Denmark
6	uhth@dtu.dk
7	
8	Kasper Kristensen
9	Technical University of Denmark, National Institute of Aquatic
10	Resources (DTU Aqua), Building 201, DK-2800 Lyngby, Denmark
11	kaskr@dtu.dk
12	
10	Teunis Jansen
13	
13	Technical University of Denmark, National Institute of Aquatic Resources
13 14 15	Technical University of Denmark, National Institute of Aquatic Resources (DTU Aqua), Building 201, DK-2800 Lyngby, Denmark, and Greenland
13 14 15 16	Technical University of Denmark, National Institute of Aquatic Resources(DTU Aqua), Building 201, DK-2800 Lyngby, Denmark, and GreenlandInstitute of Natural Resources, Kivioq 2, 3900 Nuuk, Greenland
13 14 15 16 17	Technical University of Denmark, National Institute of Aquatic Resources (DTU Aqua), Building 201, DK-2800 Lyngby, Denmark, and Greenland Institute of Natural Resources, Kivioq 2, 3900 Nuuk, Greenland tei@aqua.dtu.dk
13 14 15 16 17 18	Technical University of Denmark, National Institute of Aquatic Resources (DTU Aqua), Building 201, DK-2800 Lyngby, Denmark, and Greenland Institute of Natural Resources, Kivioq 2, 3900 Nuuk, Greenland tej@aqua.dtu.dk
13 14 15 16 17 18	Technical University of Denmark, National Institute of Aquatic Resources (DTU Aqua), Building 201, DK-2800 Lyngby, Denmark, and Greenland Institute of Natural Resources, Kivioq 2, 3900 Nuuk, Greenland tej@aqua.dtu.dk Jan E. Bever
13 14 15 16 17 18 19 20	Technical University of Denmark, National Institute of Aquatic Resources (DTU Aqua), Building 201, DK-2800 Lyngby, Denmark, and Greenland Institute of Natural Resources, Kivioq 2, 3900 Nuuk, Greenland tej@aqua.dtu.dk Jan E. Beyer Technical University of Denmark, National Institute of Aquatic
13 14 15 16 17 18 19 20 21	Technical University of Denmark, National Institute of Aquatic Resources (DTU Aqua), Building 201, DK-2800 Lyngby, Denmark, and Greenland Institute of Natural Resources, Kivioq 2, 3900 Nuuk, Greenland tej@aqua.dtu.dk Jan E. Beyer Technical University of Denmark, National Institute of Aquatic Besources (DTU Aqua) Building 201 DK-2800 Lyngby Denmark
13 14 15 16 17 18 19 20 21	Technical University of Denmark, National Institute of Aquatic Resources (DTU Aqua), Building 201, DK-2800 Lyngby, Denmark, and Greenland Institute of Natural Resources, Kivioq 2, 3900 Nuuk, Greenland tej@aqua.dtu.dk Jan E. Beyer Technical University of Denmark, National Institute of Aquatic Resources (DTU Aqua), Building 201, DK-2800 Lyngby, Denmark ieb@aqua.dtu.dk
13 14 15 16 17 18 19 20 21 22 22	Technical University of Denmark, National Institute of Aquatic Resources (DTU Aqua), Building 201, DK-2800 Lyngby, Denmark, and Greenland Institute of Natural Resources, Kivioq 2, 3900 Nuuk, Greenland tej@aqua.dtu.dk Jan E. Beyer Technical University of Denmark, National Institute of Aquatic Resources (DTU Aqua), Building 201, DK-2800 Lyngby, Denmark jeb@aqua.dtu.dk
13 14 15 16 17 18 19 20 21 22 23	Technical University of Denmark, National Institute of Aquatic Resources (DTU Aqua), Building 201, DK-2800 Lyngby, Denmark, and Greenland Institute of Natural Resources, Kivioq 2, 3900 Nuuk, Greenland tej@aqua.dtu.dk Jan E. Beyer Technical University of Denmark, National Institute of Aquatic Resources (DTU Aqua), Building 201, DK-2800 Lyngby, Denmark jeb@aqua.dtu.dk
13 14 15 16 17 18 19 20 21 22 23 24	Technical University of Denmark, National Institute of Aquatic Resources (DTU Aqua), Building 201, DK-2800 Lyngby, Denmark, and Greenland Institute of Natural Resources, Kivioq 2, 3900 Nuuk, Greenland tej@aqua.dtu.dk Jan E. Beyer Technical University of Denmark, National Institute of Aquatic Resources (DTU Aqua), Building 201, DK-2800 Lyngby, Denmark jeb@aqua.dtu.dk November 21, 2018

Abstract

We present a statistical method for intercalibration of fishery surveys methods, i.e. 26 determining the difference in catchability and size selectivity of two methods, such 27 as trawl gears or vessels, based on data from paired fishing operations. The model 28 estimates the selectivity ratios in each length class by modelling the size distribution 29 of the underlying population at each station and the size-structured clustering of 30 fish at small temporal and spatial scales. The model allows for overdispersion 31 and correlation between catch counts in neighboring size classes. This is obtained 32 by assuming Poisson distributed catch numbers conditional on unobserved log-33 Gaussian variables, i.e. the catch is modelled using log-Gaussian Cox processes. 34 We apply the method to catches of hake (Merluccius Paradoxus and M. Capensis) 35 in 341 paired trawl hauls performed by two different vessels, viz. the RV Dr. 36 Fridtjof Nansen and the FV Blue Sea, operating off the coast of Namibia. The 37 results demonstrate that it is feasible to estimate the selectivity ratio in each size 38 class, and to test statistically the hypothesis that the selectivity is independent of 39 size or of species. For the specific case, we find that differences between size classes 40 and between species are statistically significant. 41

Keywords: Selectivity; Intercalibration; Mixed-effects models; and Log-Gaussian Cox
 processes

44 1 Introduction

25

Fishery-independent surveys are of pivotal importance for fish stock assessments, where 45 they provide a relative abundance index, as well as for basic biological research [Millar, 46 1992]. While the objective of a survey is to assess the abundance of the underlying pop-47 ulation, it only provides a filtered view, specified by the selectivity of the operation. The 48 vessels, riggings, and gears applied in these surveys often develop or shift over time, as do 49 fishing methods by captains [Weinberg and Kotwicki, 2008], leading to changes in size se-50 lectivity and overall catch efficiency [Miller, 2013, Thorson and Ward, 2014]. To maintain 51 as long time series as possible, it is often desireable to combine information from differ-52 ent operations. However, differences in selectivity of vessel-gear combinations must be 53 accounted for before time series and spatial distribution data can be combined and syn-54 thesized, which can be problematic [Axelsen and Johnsen, 2015]. To this end, dedicated 55 experiments may be performed, involving two or more vessel-gear combinations, with the 56 objective of calibrating these combinations against each other, i.e. intercalibration. Here, 57 the difference in catch rates are investigated by performing pairwise near-simultaneous 58 hauls in the same area, so as to minimizes the time-space variation of the fished pop-59 ulation between the hauls. With such data, the selectivity ratios, which measure the 60 efficiency of the two vessel-gear combinations against each other, can be estimated for 61

each species and each size class. Then, these selectivity ratios can be used as calibration
factors by adjusting catches from one type of operation so that they are comparable with
the catches from the other operation [Kotwicki et al., 2017].

Multiple calibration procedures have been proposed and applied over time, in partic-65 ular differing in how the size dependency in selectivity ratios is modelled and estimated. 66 When considering the selectivity curve of a single gear, a common choice is to restrict 67 attention to a parametric family of curves; for example logistic functions for towed gear 68 and Gaussian functions for gill nets [Millar and Fryer, 1999]. When comparing two gears, 69 a typical choice has been to use polynomials in length to describe the ratio between 70 the two selectivity curves [Millar et al., 2004, Lewy et al., 2004, Holst and Revill, 2009, 71 Kotwicki et al., 2017]. The coefficients in these polynomials may be estimated in a GLM 72 framework, but a point of particular importance is to allow for overdispersion relative 73 to Poisson counts [Lewy et al., 2004]. This overdispersion arises for many reasons, in-74 cluding between-haul variation in the selectivity [Millar, 1993]. If this effect is ignored, 75 and catches from different hauls are pooled, it will lead to overconfidence in the accuracy 76 of estimates; a remedy is to use a double bootstrap to assess the accuracy of estimates 77 [Millar, 1993, Sistiaga et al., 2016]. An alternative is a GLMM approach where the rela-78 tive selectivity curves are allowed to vary between hauls; either non-parametrically using 79 autoregressive processes [Cadigan et al., 2006] or parametrically in terms of shifting and 80 scaling slope base curves [Cadigan and Dowden, 2010]. Alternatives to fixed polynomi-81 als include orthogonal polynomials, GAMs or Smooth-Curve Mixed Models [Fryer et al., 82 2003, Miller, 2013]. A typical problem of these data is the large number of zero catches; 83 therefore Thorson and Ward [2014] considered delta-GLMM's, where the probability of 84 zero catch is explicitly modeled. Kotwicki et al. [2017] compared three models, two of 85 which included polynomials to account for the dependence on length, and one which used 86 GAM's to this effect, and advocated cross-validation techniques to select the best fitting 87 model for a given data set. 88

When the original assumption is that the catch in each size class and in each haul is Poisson distributed conditional on the abundance, a common approach is to condition on the total catch in each size class. Then, the catch in the individual haul is binomially distributed [Millar, 1992]. Conceptually, a related approach is the beta regression, in which a ratio of Catches Per Unit Effort in each size class is assumed to be beta distributed [Kotwicki et al., 2017].

A common phenomenon for size structures in catches is that not only are the numbers in each length group overdispersed, there is also strong tendency for positive correlations between nearby size classes in the same haul [Pennington and Vølstad, 1994, Kristensen et al., 2014]. If not taken into account, this phenomenon means that fluctuations across size classes in raw selectivity ratios will be over-interpreted. Pragmatically, the consequence of this is that estimated selectivity ratio curves should be smoothed, but ¹⁰¹ preferrably, the size correlations should be included in the statistical model structure.
¹⁰² This ensures that the model describes the fluctuations in data adequately which is a
¹⁰³ prerequisite for the statistical analysis to be valid.

Overdispersion and correlation in count data are, in general, conveniently modeled 104 using compound Poisson distributions. These are hierarchical models, where it is as-105 sumed that the random data is generated through a two-stage procedure: In the first 106 stage, a random *intensity* is generated for each data point. In the second stage, this 107 intensity is used as the mean value for Poisson variables which constitute the count data. 108 With this construction, the variance of the random intensity yields overdispersion rela-109 tive to Poisson data, while the correlation structure of the intensity cascades to the count 110 data. A recent example of such a model structure is Miller et al. [2018]. A particular 111 framework of interest is that of log-Gaussian Cox processes [Diggle et al., 2013], where 112 the log-intensity is a Gaussian process. Since a Gaussian process is fully described by 113 its mean and covariance, this framework is highly operational and lends itself readily to 114 computations. Log-Gaussian Cox processes have previously been applied to the spatio-115 temporal modeling of size structured populations, where it has elucidated distributions 116 of cod (Gadus morhua) in the North Sea [Lewy and Kristensen, 2009, Kristensen et al., 117 2014], of whiting (Merlangius merlangus) in the Baltic [Nielsen et al., 2014], of the larvae 118 and juveniles of mackerel (Scomber scombrus) in the North Sea [Jansen et al., 2012, 2015], 119 and of shallow-water hake (Merluccius capensis) [Jansen et al., 2016] and deep-water hake 120 (*M. paradoxus*) [Jansen et al., 2017] in the Benguela current system. 121

Since log-Gaussian Cox processes proved suitable for these applications, it is natural 122 to ask if the framework is also suitable for the problem of estimating selectivity ratios. 123 The paper addresses this question. When applying the framework of log-Gaussian Cox 124 processes to the selectivity ratios, the unobserved size-dependent phenomena include the 125 selectivity ratios, which is the primary object of inference, but also the local abundance 126 present for each pair of operations, as well as aggregations that are specific to the individ-127 ual operation. Each of these phenomena is characterized by a covariance structure, which 128 describes both the magnitude of fluctuations and their persistence across size ranges. The 129 construction is a fairly simple application of the log-Gaussian Cox framework, and has 130 the appeal that we can specify the properties of the various processes affecting the catch, 131 from which the properties of the log-intensity follow automatically. 132

In this paper we describe the framework and the resulting method. We demonstrate the method using data from a case where the objective was to investigate differences between two vessels which used gear with the same specifications: The RV Dr. Fridtjof Nansen and the FV Blue Sea, which have been used for surveying the stocks of hake in Namibian waters. The objective of the analysis is to estimate the selectivity ratios between the two vessels, including confidence intervals, and to test if the ratios depend on size and the particular hake species. In addition, we perform a simulation experiment to verify the model, test for significance of certain specific model components, and compare
the full model with a simplified model where inference is conditional on total catch at
length for each station.

$_{143}$ 2 Methods

¹⁴⁴ 2.1 Statistical model

Our method for intercalibration is based on a statistical model for the selectivity ratios 145 which explains the size composition of the catch in survey operations, and in particular 146 differences in this composition between operations conducted differently on the same fish 147 population. For ease of reference, we refer to these operations as 'hauls', whether the 148 gear involved is e.g. trawls, longlines or gill nets. Similarly, we refer to differences be-149 tween 'gear', even if the actual differences between operations could also involve different 150 vessels, different personel, or different procedures. The model is a non-linear mixed effect 151 model involving both fixed effects parameters and random effects. We conduct inference 152 in the model using numerical maximum likelihood estimation, employing the Laplace 153 approximation [Kristensen et al., 2016] to integrate out random effects. 154

The observed quantities are count data, N_{ijk} , which represents number of individuals caught at station $i = 1, ..., n_s$, with gear j = 1, 2, and in length group $k = 1, ..., n_l$. Thus, at each station i, two operations have been performed; one with each gear j, and the size distribution of the catch has been measured.

We assume that these catches depend on swept area A_{ij} (or a similar measure of 159 effort) and three sets of random variables, which all depend on the size class k: First, 160 Φ_{ik} which for a given station *i* characterizes the distribution across size of the population 161 encountered by both hauls j. Second, haul-specific fluctuations R_{ijk} in the size compo-162 sition which we will term the "nugget effect" with a reference to geostatistics [Cressie, 163 1993, Petitgas, 2001] and elaborate on in the following. Third, the relative selectivity S_{jk} 164 which is specific to the gear. Given these random variables Φ , R, S, we assume that the 165 count data is Poisson distributed: 166

$$N_{ijk}|\Phi, R, S \sim \text{Poisson}(A_{ij} \cdot \exp(S_{jk} + \Phi_{ik} + R_{ijk}))$$

The swept area A_{ij} is a known input to the model. This is Cox model of catches, also referred to as a doubly stochastic Poisson model, in that the mean values for the Poisson variates are themselves random. The joint distribution of the processes S, Φ , and R is Gaussian, so that the entire model is a log-Gaussian Cox process [Møller et al., 1998, Diggle et al., 2013]. We now describe the details of the processes S, Φ and R.

First, the selectivity (on the log scale) S_{jk} of gear j in size group k is the main object

¹⁷³ of interest. Since we do not know the actual size distribution of the stock, we cannot ¹⁷⁴ estimate the absolute selectivities S_{1k} and S_{2k} of the two types of gear, but only the ¹⁷⁵ relative selectivity, i.e. $S_{1k} - S_{2k}$. We therefore require

$$S_{1k} = -S_{2k},\tag{1}$$

which allows us to focus on S_{1k} . This symmetric choice ensures that N_{i1k} and N_{i2k} are identically distributed, which ultimately implies that the estimated selectivities S_{jk} simply change sign if the gears are relabeled.

¹⁷⁹ We note an alternative would be to enforce $S_{1k} = 0$ and estimate S_{2k} . This would be ¹⁸⁰ reasonable when the first gear is a reference gear that we measure the second gear against. ¹⁸¹ In that case the variance on N_{i1k} would then be smaller than that on N_{i2k} , since N_{i2k} ¹⁸² would contain the extra variance component S_{2k} . This asymmetry would cascade to the ¹⁸³ estimates, so that the estimated relative selectivities depend on which gear is considered ¹⁸⁴ the reference gear. In the present study, we have no reason to consider the one gear a ¹⁸⁵ reference, and therefore we prefer the symmetric choice $S_{1k} = -S_{2k}$.

To interpret the selectivities S_{jk} , it is useful to momentarily disregard the nugget effect *R*. Then, conditional on Φ and *S*, the expected catches at station *i* and in size class *k* with the two types of gear are $A_{i1} \exp(\Phi_{ik} + S_{1k})$ and $A_{i2} \exp(\Phi_{ik} - S_{1k})$, respectively. Thus, $\exp(2S_{1k})$ is the ratio between the expected catch per unit effort with the two types of gear:

$$\exp(2S_{1k}) = \frac{\mathbb{E}\{N_{i1k}/A_{i1}|\Phi, S\}}{\mathbb{E}\{N_{i2k}/A_{i2}|\Phi, S\}} \quad .$$
(2)

This ratio is termed the selectivity ratio [Kotwicki et al., 2017]. Since this ratio must be positive, and since we do not assume a particular parametric form, it is convenient to represent it on the log scale, i.e. in terms of the process S. We model S_{1k} as a random walk in size k, i.e.

$$S_{1(k+1)} - S_{1k} \sim N(0, \sigma_S^2)$$
 for $k = 1, \dots, n_l - 1$

and assume independence between increments. To ensure that the log-selectivity ratio S is a well defined stochastic process, we complement this recursion with initial conditions $S_{j1} \sim N(0, \sigma_1^2)$ where σ_1 is fixed at a "large" value 10, which from a practical point of view implies that the level of the estimated log-sensitivity ratio S is not dictated by the prior model but rather by data.

Next, Φ_{ik} is a log-density which describes the size distribution of the fish caught at station *i*- Specifically, $A_{ij} \exp(\Phi_{ik})$ is the expected number of fish caught in size group k at station *i* with a hypothetical gear which averages the two gears j = 1 and j = 2, in absence of nuggets (R = 0). We assume independence of size distributions at different stations, i.e. Φ_{ik} and $\Phi_{i'k'}$ are independent for $i \neq i'$. At each station *i*, we assume that the log-density of the size distribution is a random walk over size groups, i.e.

$$\Phi_{i(k+1)} - \Phi_{ik} \sim N(0, \sigma_{\Phi}^2)$$
 for $k = 1, \dots, n_l - 1$

²⁰⁷ and that these increments are independent. Thus, the prior on the log-density Φ is a ²⁰⁸ standard random walk which enforces continuity; the most probable density is flat. We ²⁰⁹ add initial conditions

$$\Phi_{i1} \sim N(0, \sigma_1^2)$$

with the same "large" standard deviation $\sigma_1 = 10$, so that the overall level of Φ is not dictated by the prior model but rather by the total catch. The parameter σ_{Φ}^2 is estimated. Since we assume independence between stations, we do not attempt to model any large-scale spatiotemporal structure of the population. We note that this is the main difference between this model and the GeoPop model [Kristensen et al., 2014], where emphasis is exactly on this spatiotemporal structure.

Finally, the haul-specific fluctuations R_{ijk} are akin to the nugget effect in spatial statis-216 tics; i.e. they describe variability in the catch data on very small spatial and temporal 217 scales. While the term "nugget" originates in applications to mining, where repeated 218 measurements on the same location may hit or miss a nugget, the envisioned mechanism 219 in survey operations is that the gear may hit or miss aggregations of fish such as schools 220 or shoals, that have limited range in space and quickly form, move, dissolve and regroup. 221 Since the two hauls at one station have been performed at slightly different locations 222 and times, they will encounter different aggregations, and therefore R_{ijk} and $R_{i'j'k'}$ are 223 independent unless (i, j) = (i', j'), i.e. the same haul. Thus, at a given station i and in 224 a given size class k, Φ_{ik} models the population that is common to the two hauls, while 225 R_{ijk} models independent components which are distinct to each haul. We think of the 226 aggregations giving rise to the nugget effect R_{ijk} as size-structured, and therefore, for a 227 given haul (i, j) and as a function of size k, the nugget effect arises as the sum of a white 228 noise process and a zero-mean first order autoregressive process. Specifically 229

$$R_{ijk} = R_{ijk}^{WN} + R_{ijk}^{AR}$$

where $R_{ijk}^{WN} \sim N(0, \sigma_{WN}^2)$ and are independent. In turn $R_{ijk}^{AR} \sim N(0, \sigma_{AR}^2)$ and are independent for different stations *i* or gear *j*, but correlated between size classes at a given station *i* and gear *j* so that $\mathbb{E}(R_{ijk}^{AR}R_{ijk'}^{AR}) = \sigma_{AR}^2\rho^{|k-k'|}$. The white noise component allows overdispersion relative to Poisson without correlation, while the autoregressive component models the size-specific clustering: If a particular size group is more abundant in the haul than expected, we would expect the same to apply to nearby size groups but not necessarily to very different size groups. We note that this same model structure was used by Cadigan et al. [2006] with the same motivation, but also that the effect could equally well represent other differences between the individuals hauls, e.g. differences in the way the gear is deployed, or combinations of such differences.

The model has five fixed effects parameters which are estimated, viz. the variance parameters σ_S^2 , $\sigma_{\Phi}^2 \sigma_{WN}^2$, σ_{AR}^2 , and the correlation ρ . In addition there are a large number of random effects: Φ has $n_s n_l$ variables, S has n_l , and R has $n_s 2n_l$.

243 2.2 Implementation

The statistical model in section 2.1 defines the joint distribution of the count data, N, and the unobserved random variables Φ , R, S, for given parameters σ_S , σ_{Φ} , σ_{WN} , σ_{AR} and ρ . The unobserved Φ , R and S are integrated out using the Laplace approximation, to yield the likelihood as a function of the five parameters. The likelihood function is maximized to yield estimates of the five parameters, after which the posterior modes of the random effects Φ , R, and in particular S are reported.

The computations are performed in R version 3.1.2; we use the Template Model Builder (TMB) package [Kristensen et al., 2016] for evaluating the likelihood function and its derivatives, and in particular for integrating out unobserved random variables using the Laplace approximation. Typical run-times for the models considered in this paper, where there are 77,680 random effects, are 25 seconds on a standard laptop computer. The code is available at GitHub in package github.com/Uffe-H-Thygesen/Intercalibration.

The code and the statistical model is verified by simulation. Briefly, we simulate 1,000 256 realizations of random effects and data sets, adjusting the mean of the size distributions 257 Φ so that the total catch in the simulated data sets are approximately 17,000 fish, which 258 corresponds to the total catch in the case described in the following. For each realization, 259 we re-estimate the parameters in the model and the log-selectivity ratios. The variance 260 parameters σ_S^2 , $\sigma_{\Phi}^2 \sigma_{WN}^2$, and σ_{AR}^2 are estimated on the log scale. We construct 1σ confi-261 dence intervals for each of the five parameters using the estimated standard deviation as 262 computed from the Hessian of the log-likelihood. Theoretically, these confidence intervals 263 should contain the true parameters for 68% of the simulated data sets; we find that they 264 do so for between 66 % and 71 % of the simulated data sets, except for the parameter 265 $\log \sigma_{\Phi}^2$, where the coverage is only 48 %. For this parameter, the low coverage is explained 266 by a bias in the estimates: The mean estimate is 0.07 smaller than the true value, which 267 should be compared with an estimated standard deviation which is also 0.07. While neg-268 ative bias is not uncommon for maximum likelihood estimates of variance parameters, it 269 could possibly be reduced with restricted maximum likelihood (REML) [Pawitan, 2001]. 270 We also constructed 2- σ confidence limits, which should contain the true value in 95 % 271

of the runs, and find that they do so for between 86 % and 96 % of the simulated data 272 sets. The relative uncertainties on the variance parameters σ_S^2 and σ_{AR}^2 (measured from 273 the standard deviation on estimates) are 13 % and 7 %, respectively, with a bias which 274 is an order of magnitude smaller. The relative uncertainty on ρ is 2% with a bias of 0.2 275 %. In roughly half the simulations, the model cannot identify the white noise component 276 in the residuals and consequently estimates σ_{WN}^2 to be very low $(\sigma_{WN}^2/\sigma_{AR}^2 < 10^{-5})$; 277 in these cases, also the estimated variance on $\log \sigma_{WN}^2$ is very large (i.e. > 10) so that 278 the confidence intervals still cover the true value. While the white noise component is 279 effectively removed from the model through the estimation for these simulated data sets, 280 the reduced model is estimated well. We note that such problems of estimating sepa-281 rate variance components in hierarchical models are not uncommon [Auger-Méthé et al., 282 2016]. With this caveat, the simulation experiments verifies the code and the model. 283

284 2.3 Data

We apply the method to a case study involving two vessels, the Norwegian fisheries research vessel Dr Fridtjof Nansen and the commercial trawler F/V Blue Sea, conducting hake surveys in Namibian waters.

Following independence of Namibia in 1990, abundance of Namibias hake stocks was 288 monitored by trawl surveys conducted by the R/V Dr Fridtjof Nansen. From 2000 the 289 Ministry of Fisheries and Marine Resources in Namibia (MFMR) conducted the surveys 290 using the F/V Blue Sea. In 1998 and 1999, before the shift, extensive experiments were 291 performed by completing the entire annual survey in parallel with both vessels. The two 292 vessels used Gisund fishing gear and rigging following the same specifications; neverthe-293 less, some difference in the performance of the gear must be anticipated [Weinberg and 294 Kotwicki, 2008]. The stations are mapped in Figure 2. 295

Catch data collected from these surveys were extracted from the NAN-SIS database in November 2014 [Strømme, 1992]. The analysis was based on 341 of the 365 pairs of trawl hauls. 24 pairs were excluded because the trawl durations were less than 15 minutes and/or the difference in trawl durations exceeded 10 minutes.

Catch in numbers per length group and the hauling distance were available for each haul. Figure 3 shows all catches, summed over all stations, for the two species M. Paradoxus (deep-water hake) and M. Capensis (shallow-water hake). Since the two species have different preferred habitats but are morphologically very similar [Jansen et al., 2016, 2017], a question of particular relevance is if the two species have the same selectivity.

305 **3** Results

Figure 4 shows the selectivity ratio from (2), i.e. $\exp(2S_{1k})$, between the RV Dr. Fridtjof 306 Nansen and the FV Blue Sea. Index 1 corresponding to FV Blue Sea, so that a ratio above 307 1 indicates that the FV Blue Sea has higher expected catch than the RV Dr. Fridtjof 308 Nansen. Estimated parameters, including standard errors derived from the Hessian of 309 the log-likelihood function, are shown in Table 1. Since the gears used on the two vessels 310 have the same specifications, a reasonable hypothesis is that there is no size structure in 311 these calibration factors. This hypothesis could be accepted for *M. Capensis* (a likelihood 312 ratio test of the hypothesis $\sigma_S = 0$ has critical significance level $p \sim 0.08$) but is rejected 313 strongly for *M. Paradoxus* $(p < 10^{-9})$. These *p*-values have been computed with the 314 standard asymptotic χ^2 -distribution of the log-likelihood ratio, which does not strictly 315 apply since the null hypothesis $\sigma_S = 0$ is on the boundary of the parameter space, so 316 that the correct p-values may be somewhat smaller. It holds for both species that the 317 FV Blue Sea is more efficient at catching larger hakes than the RV Dr. Fridtjof Nansen. 318 The size dependency is more pronounced for M. Paradoxus, where the FV Blue Sea is 319 less efficient in the small size classes. The selection of small M. Capensis is similar for 320 the two vessels. The estimated relative selectivity appears to fluctuate more between 321 neighboring size classes for *M. paradoxus* than for *M. capensis*. This may be because the 322 smaller catches of *M. capensis* imply less statistical certainty, so that the smooth prior 323 is more visible in the estimates. It could also be connected to the observation that the 324 estimated correlation ρ is closer to 1 for M. paradoxus than for M. capensis so that small 325 scale fluctuations in the data is attributed to the nugget effect for M. capensis but, to a 326 larger degree, to fluctuations in the relative selectivity for *M. paradoxus*. 327

Since there is no clear prior explanation why the selectivity curves for the two species 328 would differ, a reasonable hypothesis is that they are identical. This hypothesis appears 329 to be strengthened by the qualitative similarity between the estimated curves in Figure 330 4. This suggests to estimate a combined selectivity ratios for the two species, see Figure 331 5. In this combined model, we assume that the size distribution and the nugget effect 332 applies to the two species separately, i.e. the small-scale clustering of fish is species-333 specific. Since each fit yields a likelihood, it is possible to select between the two models 334 (i.e., the two species have the same relative selectivity curve, or two different curves) 335 using an information criterion such as that of Akaike, the AIC. The log-likelihood of the 336 combined model is 258 less than that of the original model; this decrease results from 337 the reduction of the number of parameters (fixed effects) from 10 to 5. Thus, the AIC 338 will prefer strongly the model where the two species have separate selectivity ratios; for a 339 likelihood ratio test, the critical p-value would be 10^{-108} . We note that since the primary 340 objective of inference is on the relative selectivity curves, which are random effects in the 341 model, one could argue that model selection should be performed with the *conditional* 342

Species	$\log \sigma_{\Phi}$	φ	$\log \sigma_{WN}$	$\log \sigma_{AR}$	$\log \sigma_S$	$-\log L$	DF
M. Capensis	-0.15 ± 0.02	0.95 ± 0.01	-0.41 ± 0.02	-0.05 ± 0.04	-4.17 ± 0.50	44940	5 C
$M. \ Paradoxus$	-0.25 ± 0.02	0.98 ± 0.01	-0.95 ± 0.03	0.06 ± 0.05	-3.29 ± 0.24	34607	J.
Sum						79547	10
Combined	-0.19 ± 0.01	0.96 ± 0.01	-0.61 ± 0.01	0.01 ± 0.03	-3.68 ± 0.24	79805	υ
Combined w/o ρ	-0.18 ± 0.01		-0.10 ± 0.01		-3.87 ± 0.27	82417	က

Table 1: Parameter estimate for the two species separately and combined, with estimated standard deviations. Included is also the negative log-likelihood and the number of parameters (fixed effects) of the model.

AIC [Vaida and Blanchard, 2005]. While the computation of the conditional AIC is a 343 non-trivial task in our settings, a bound can be obtained by including the random effects 344 in the degrees of freedom; this holds because each random effect in the cAIC framework 345 is associated with a non-integer degree of freedom between 0 and 1. Then, the difference 346 in log-likelihood should be compared with a maximum difference of 76 in the degrees of 347 freedom, which would still favour strongly separate selectivity ratios for the two species. 348 We conclude that the differences between the two species are statistically significant, even 349 if the relative selectivity curves for the two species show similar qualitative features. 350

To illustrate the importance of the correlation between the different size classes, we 351 fit a new model to this combined data set, in which the autoregressive component of 352 the nugget effect has been removed, so that the nugget effect acts independently at each 353 size class (Figure 5, right panel). Removing this component from the model results in 354 a decrease in the maximum log-likelihood of 2612 while decreasing the numbers of pa-355 rameters by 2; thus this autoregressive component is extremely significant ($p \approx 10^{-1134}$). 356 Nevertheless, the estimates from this reduced model agree qualitatively with the those 357 from the model that includes autocorrelation in the nugget effect (compare Figure 5, left 358 panel), although some minor differences are noticeable. Moreover, omitting the autocor-359 relation decreases the estimated variance associated with the selectivity ratio curves, so 360 that the simpler model indicates higher accuracy than warranted. 361

Since several previous studies including [Millar, 1992, Lewy et al., 2004, Cadigan and Dowden, 2010] have considered a conditional approach, where inference is conditional on the total catch at length at each station, appendix A compares such a conditional model with the model as described in section 2.1. The two models give qualitatively similar results, but the estimated selectivity ratios from the unconditional model are generally closer to 1. The unconditional model has slightly narrower confidence intervals and is slightly more demanding in terms of computing time.

369 4 Discussion

We developed a statistical method for intercalibrating survey gear and vessels, based 370 on estimating the selectivity ratios from paired hauls. The method is directly available 371 through an R package on GitHub. The envisioned application of our method is to adjust 372 data obtained from multiple surveys, thus allowing them to be combined to yield a longer 373 time series which may enter into a stock assessment. The adjustment would take place by 374 multiplying the one series with the estimated selectivity ratios. The uncertainties on the 375 estimated selectivity ratios would then propagate to the adjusted time series, for example 376 using the delta method as implemented in TMB [Kristensen et al., 2016]. While one could 377 envision integrated stock assessment models that use multiple raw survey indices as well 378 as data from paired fishing operations, the preliminary step of adjusting and combining 379

³⁸⁰ surveys appears to be preferable at least in the foreseeable future.

Our model is based on log-Gaussian Cox processes, which have been used earlier in 381 the context of fisheries surveys to map spatiotemporal dynamics of stocks [Kristensen 382 et al., 2014, Jansen et al., 2016], but not in the present way for comparing selectivities. 383 The framework uses a non-parametric model for the relative selectivity and allows for 384 overdispersion relative to the Poisson distribution, as well as correlations between size 385 groups in paired trawl catches. These features all contribute to larger variability in data, 386 and the Gaussian structure of the components simplifies analysis and computations. If the 387 statistical analysis is based on models which fail to include such variance contributions, 388 there is a risk that the confidence in the results are inflated, e.g. in the sense that 389 confidence intervals appear narrower than justified. Such phenomena of overconfidence 390 are well know, both in general statistics and in the specific context of selectivity studies 391 [Fryer, 1991]. They can be seen as a manifestation of the general bias-variance trade-off. 392 Previous methods to address between-haul and within-haul variation include bootstrap 393 [Millar, 1993, Sistiaga et al., 2016] in addition to mixed effects models [Cadigan et al., 394 2006]. In the present study, an example of such overconfidence is seen in Figure 5, 395 comparing the two panels, where the right panel is based on a simplified model in which 396 the autoregressive component of the nugget effect has been removed. Recalling that a 397 hypothesis test rejected this simplification, and noticing that the reduced model produces 398 estimated confidence intervals which are considerably narrower, we can conclude that 399 these confidence intervals give an overoptimistic view on the accuracy of estimates. This 400 overoptimism can be attributed to the omission of an important variance component. 401

As another example of possible overconfidence, selectivity ratios can be modeled as 402 constants which apply to all size classes, as size-dependent functions using parametric 403 forms, or non-parametrically as we have done here. While specific parametric families 404 of functions are convenient in the analysis, it is difficult to hypothesize a reasonable 405 functional form prior to seeing the data. If a specific functional form is postulated, then 406 it is likely that parameters in this form can be estimated with seemingly high accuracy. 407 However, the sensitivity of the results to mis-specification of the functional form needs to 408 be taken into consideration which is not straightforward. As a result, we would be prone 409 to overestimate our confidence in estimated selectivity ratio curves, by the same reasoning 410 as in the previous paragraph. Thus nonparametric curves, such as the ones we provide 411 in this study, involve the smallest number of assumptions and are the most conservative 412 choice in the sense of not risking overinterpretation of data. For some applications it 413 is convenient to report parametric forms. This would be a minor extension, technically, 414 but a subsequent step of model validation needs to ensure that the parametric family 415 is suitable. On the other hand, non-parametric estimates require some regularization to 416 avoid erratic fluctuations in the estimated curves. Here, we have obtained this smoothness 417 by using a random walk prior on the relative selectivity curve, which is a minimal way of 418

enforcing continuity. An alternative is to use smooth basis functions or smoothing splines
[Miller, 2013].

The core of our approach is to take into consideration the covariance between different 421 size classes, both in the selectivity ratio curves that we aim to estimate, and in the catch 422 data. Neglecting this covariance would require that data is binned into large size bins 423 with sufficiently high catch numbers, so that we can estimate the selectivity in each 424 bin without borrowing information from neighboring bins. If the true selectivity ratios 425 vary with size, this would lead to a classical trade-off between bias and variance of the 426 estimates. Specifying the fluctuations between size classes, as we have done, bypasses 427 this trade-off and will give consistent results regardless of how small size bins are chosen. 428 The crux of this approach is the correct specification of the covariance structure. Here, 429 we have taken a conservative approach in that we model the log-densities Φ and the 430 relative selectivity S as random walks across size, which amounts to enforcing continuous 431 dependency on size. In turn, the nugget effect is an autoregressive process. The effect 432 of this structure is that large catches across size groups in a specific haul is attributed 433 to high selectivity (S) or to high density at the station (Φ), whereas an isolated peak in 434 catch numbers at a given size range in a specific haul is attributed to size-specific shoaling 435 aggregations, i.e. the nugget effect R. 436

In our model, the random walks have unbiased and identically distributed steps. One 437 would expect that the selectivity ratios fluctuate more in those size classes, where the 438 selectivity curve of each gear changes the most, and less for the large size classes where 439 both gears have full selectivity. Similarly, we would expect that the size distributions are 440 skewed towards the smaller size classes. Thus, our model structure relies on simplifying 441 assumptions, and we do not expect the model to fully describe all variability in the data. 442 Nevertheless, our simulation study indicates that the model structure allows estimation 443 of the selectivity ratios which is the objective of the model. 444

Inspecting the appearance of the nugget effect in the model, we see that it could 445 equally well be interpreted as a factor that modifies the selectivity of the gear in the 446 operation, although we interpret it as a factor affecting the local abundance. Such ran-447 dom fluctuations in selectivity have been considered previously [Fryer, 1991, Miller, 2013]. 448 Based on the information in data sets such as the present, the two effects are confounded 449 [Cadigan and Dowden, 2010]: It is not possible to tell if a high catch in one particular op-450 eration was because the gear encountered an aggregation, or because the gear functioned 451 better than average in that operation. In both cases, the net effect is a larger variability 452 between repeated hauls. 453

⁴⁵⁴ A key question that the model aims to answer is if the gear (or vessel) effect can ⁴⁵⁵ be assumed to be identical for all size classes, and it is interesting to notice that this ⁴⁵⁶ does not appear to be the case for *M. Paradoxus*. Similarly, it is interesting that the ⁴⁵⁷ two species appear to have different selectivity ratios. Although there is no single clear

biological explanation for this, there will always be several minor differences in the nets, 458 the rigging, and the way the hauls are performed, which can contribute to such differences 459 [Weinberg and Kotwicki, 2008], keeping in mind the numerous processes that interact 460 and influence the catchability. At the same time caution most be exercised: The results 461 indicate that the size structure in the catches would be extremely improbable if the gear 462 effect acted identically on all size classes, or identically to the two species, under the 463 assumptions in the model. The result therefore hinges on the model representing the 464 variability in catches correctly. While informal model checks suggest that this is the 465 case, we have not performed a stringent model validation using e.g. the techniques in 466 [Thygesen et al., 2017], as the computations would be prohibative. Thus, there is a risk 467 that some overdispersion in the data is not included in the model, and that the apparent 468 differences between size classes and species are artifacts of this overdispersion. 469

While our main motivation for investigating the relative selectivity is scientific surveys, another important area of application is the selectivity of commercial gear. Here, trade-offs between efficiency and environmental impact is one concern that motivates comparative studies of the selectivity of different gear [Sistiaga et al., 2015, Vogel et al., 2017].

An underlying assumption behind our analysis is that the two operations at a given 475 station do not affect each other. This assumption conflicts somewhat with the require-476 ment that the two operations are performed close to each other, both in space and time, 477 so that it is plausible that they encounter the same population. In contrast, Lewy et al. 478 [2004] focused on the disturbance effect that a first haul has on the local fished popula-479 tion, and the implications for the second haul. In the present study, none of the pairs 480 in the available dataset are exceedingly close, so it would be superfluous to include such 481 effects. Nevertheless, when applying the method to other data sets, it would be possible 482 to parametrize such an effect and include it in the model. A logical extension would be 483 to let the variance on the nugget effect increase with the distance between the two oper-484 ations in space and time; however, it may be difficult to identify such structures reliably. 485 The limiting case of unpaired fishing operations [Sistiaga et al., 2016] is straightforward 486 to analyze with our present framework but we have not investigated the quality of the 487 resulting estimates. 488

Several previous similar studies have used a conditional approach along the lines in 489 appendix A. In the present study, we found that the estimates from the conditional and 490 unconditional model differed somewhat with estimates from the unconditional model gen-491 erally being closer to 1. The conditional model has fewer random effects, but computing 492 times are becoming less important thanks to the efficiency of Template Model Builder. 493 The unconditional model has the advantage that it is applicable also to data sets with 494 unpaired, or partially paired, hauls, but it is conceivable that the prior model for the size 495 distribution in the population (Φ) is more critical in such situations and would require 496

⁴⁹⁷ further scrutiny.

498 5 Conclusion

We have demonstrated the feasibility of estimating size-specific selectivity ratios from 499 paired fishing operations, using conditional Poisson distributions while overdispersion 500 and the covariance structure is modelled using unobserved random fields. These fields 501 represent stock size composition, small scale size structured clustering, and gear selec-502 tivity. The Laplace approximation, implemented in TMB, allows us to integrate out the 503 many unobserved random variables so that the model is computationally feasible. The 504 model allows testing of various hypotheses using the likelihood ratio principle, and model 505 selection using for example AIC. The model, of which an R implementation is publically 506 available, yields non-parametric selectivity ratios, including confidence regions, which can 507 be used to integrate survey catches obtained with different vessels or gear configurations. 508

Acknowledgements

This work was supported by EuropeAid through the EcoFish project (CRIS Number C-222387).

512 **References**

Marie Auger-Méthé, Chris Field, Christoffer M Albertsen, Andrew E Derocher, Mark A
Lewis, Ian D Jonsen, and Joanna Mills Flemming. State-space models dirty little
secrets: Even simple linear gaussian models can have estimation problems. *Scientific reports*, 6:26677, 2016.

⁵¹⁷ Bjørn Erik Axelsen and Espen Johnsen. An evaluation of the bottom trawl surveys in
⁵¹⁸ the Benguela Current Large Marine Ecosystem. *Fisheries Oceanography*, 24(1):74–87,
⁵¹⁹ 2015.

- Noel G Cadigan and Jeff J Dowden. Statistical inference about the relative efficiency of
 a new survey protocol, based on paired-tow survey calibration data. *Fishery Bulletin*,
 108(1):15–30, 2010.
- Noel G Cadigan, Stephen J Walsh, and WB Brodie. Relative efficiency of the Wilfred
 Templeman and Alfred Needler research vessels using a Campelen 1800 shrimp trawl in
 NAFO Subdivision 3Ps and Divisions 3LN, 2006. Department of Fisheries and Oceans
 Canadian Science Advisory Secretariat Research Document 2006/085.

- Noel A.C. Cressie. *Statistics for Spatial Data*. Wiley Series in Probability and Statistics.
 Wiley, New York, second edition, 1993.
- Peter J. Diggle, Paula Moraga, Barry Rowlingson, and Benjamin M. Taylor. Spatial and
 spatio-temporal log-Gaussian Cox processes: Extending the geostatistical paradigm.
 Statistical Science, 28(4):542–563, 2013.
- R.J. Fryer. A model of between-haul variation in selectivity. *ICES Journal of Marine Science: Journal du Conseil*, 48(3):281–290, 1991.
- R.J. Fryer, Alain F Zuur, and Norman Graham. Using mixed models to combine smooth
 size-selection and catch-comparison curves over hauls. *Canadian Journal of Fisheries and Aquatic Sciences*, 60(4):448–459, 2003.
- ⁵³⁷ René Holst and Andrew Revill. A simple statistical method for catch comparison studies.
 ⁵³⁸ Fisheries Research, 95(2-3):254–259, 2009.
- Teunis Jansen, Kasper Kristensen, Mark Payne, Martin Edwards, Corinna Schrum, and
 Sophie Pitois. Long-term retrospective analysis of mackerel spawning in the North Sea:
 a new time series and modeling approach to CPR data. *PLoS One*, 7(6):e38758, 2012.
- Teunis Jansen, Paulus Kainge, Larvika Singh, Margit Wilhelm, Deon Durholtz, Tore
 Strømme, John Kathena, and Victoria Erasmus. Spawning patterns of shallow-water
 hake (Merluccius capensis) and deep-water hake (M. paradoxus) in the Benguela Current Large Marine Ecosystem inferred from gonadosomatic indices. Fisheries Research,
 172:168–180, 2015.
- Teunis Jansen, Kasper Kristensen, Paulus Kainge, Deon Durholtz, Tore Strømme,
 Uffe Høgsbro Thygesen, Margit R Wilhelm, John Kathena, Tracey P Fairweather,
 Sarah Paulus, et al. Migration, distribution and population (stock) structure of shallowwater hake (Merluccius capensis) in the Benguela Current Large Marine Ecosystem inferred using a geostatistical population model. Fisheries research, 179:156–167, 2016.
- Teunis Jansen, K Kristensen, TP Fairweather, P Kainge, JN Kathena, MD Durholtz,
 JE Beyer, and Uffe Høgsbro Thygesen. Geostatistical modelling of the spatial life
 history of post-larval deepwater hake *Merluccius paradoxus* in the benguela current
 large marine ecosystem. *African Journal of Marine Science*, 39(3):349–361, 2017.
- Stan Kotwicki, Robert R Lauth, Kresimir Williams, and Scott E Goodman. Selectivity
 ratio: A useful tool for comparing size selectivity of multiple survey gears. *Fisheries Research*, 191:76–86, 2017.

 Kasper Kristensen, Uffe Høgsbro Thygesen, Ken Haste Andersen, and Jan E. Beyer.
 Estimating spatio-temporal dynamics of size-structured populations. *Canadian Journal* of Fisheries and Aquatic Sciences, 71:326–336, 2014.

Kasper Kristensen, Anders Nielsen, Casper W. Berg, and Hans Skaug. TMB: Automatic
 differentiation and Laplace approximation. *Journal of Statistical Software*, 70, 2016.

Peter Lewy and Kasper Kristensen. Modelling the distribution of fish accounting for
 spatial correlation and overdispersion. *Canadian Journal of Fisheries and Aquatic Sciences*, 66(10):1809–1820, 2009.

Peter Lewy, J Rasmus Nielsen, and Holger Hovgård. Survey gear calibration independent
 of spatial fish distribution. *Canadian Journal of Fisheries and Aquatic Sciences*, 61(4):
 636–647, 2004.

⁵⁷⁰ Russell B Millar. Estimating the size-selectivity of fishing gear by conditioning on the ⁵⁷¹ total catch. *Journal of the American Statistical Association*, 87(420):962–968, 1992.

Russell B Millar. Incorporation of between-haul variation using bootstrapping and nonparametric estimation of selection curves. *Fishery Bulletin*, 91:564–572, 1993.

⁵⁷⁴ Russell B Millar and Robert J Fryer. Estimating the size-selection curves of towed gears,
⁵⁷⁵ traps, nets and hooks. *Reviews in Fish Biology and Fisheries*, 9(1):89–116, 1999.

⁵⁷⁶ Russell B Millar, Matt K Broadhurst, and William G Macbeth. Modelling between-haul
⁵⁷⁷ variability in the size selectivity of trawls. *Fisheries Research*, 67(2):171–181, 2004.

Timothy J Miller. A comparison of hierarchical models for relative catch efficiency based
on paired-gear data for US Northwest Atlantic fish stocks. *Canadian Journal of Fish- eries and Aquatic Sciences*, 70(9):1306–1316, 2013.

Timothy J. Miller, Deborah R. Hart, Karen Hopkins, Norman H. Vine, Richard Taylor,
Amber D. York, and Scott M. Gallager. Estimation of the capture efficiency and
abundance of Atlantic sea scallops *Placopecten magellanicus* from paired photographicdredge tows using hierarchical models. *Canadian Journal of Fisheries and Aquatic Sciences*, 2018. In press.

 J. Møller, A.R. Syversveen, and R.P. Waagepetersen. Log Gaussian Cox processes. Scandinavian Journal of Statistics, 25:451–482, 1998.

J Rasmus Nielsen, Kasper Kristensen, Peter Lewy, and Francois Bastardie. A statistical model for estimation of fish density including correlation in size, space, time and between species from research survey data. *PloS one*, 9(6):e99151, 2014.

- Yudi Pawitan. In all likelihood: Statistical modelling and inference using likelihood. Ox ford University Press, Oxford, UK, 2001.
- Michael Pennington and Jon Helge Vølstad. Assessing the effect of intra-haul correlation
 and variable density on estimates of population characteristics from marine surveys.
 Biometrics, 50(3):725-732, 1994.
- ⁵⁹⁶ Pierre Petitgas. Geostatistics in fisheries survey design and stock assessment: models,
 ⁵⁹⁷ variances and applications. *Fish and Fisheries*, 2(3):231–249, 2001.
- Manu Sistiaga, Bent Herrmann, Eduardo Grimaldo, Roger B Larsen, and Ivan Tatone.
 Effect of lifting the sweeps on bottom trawling catch efficiency: A study based on the
 Northeast arctic cod (Gadus morhua) trawl fishery. Fisheries research, 167:164–173,
 2015.
- Manu Sistiaga, Bent Herrmann, Eduardo Grimaldo, and Finbarr O'Neill. Estimating the
 selectivity of unpaired trawl data: A case study with a pelagic gear. *Scientia Marina*,
 80:321–327, 09 2016.
- T Strømme. Software for Fishery Survey Data Logging and Analysis: User's Manual,
 volume 4. Food & Agriculture Org., 1992.
- James T Thorson and Eric J Ward. Accounting for vessel effects when standardizing catch rates from cooperative surveys. *Fisheries Research*, 155:168–176, 2014.
- ⁶⁰⁹ Uffe Høgsbro Thygesen, Christoffer Moesgaard Albertsen, Casper Willestofte Berg,
 ⁶¹⁰ Kasper Kristensen, and Anders Nielsen. Validation of ecological state space models
 ⁶¹¹ using the Laplace approximation. *Environmental and Ecological Statistics*, pages 1–23,
 ⁶¹² 2017.
- ⁶¹³ Florin Vaida and Suzette Blanchard. Conditional akaike information for mixed-effects
 ⁶¹⁴ models. *Biometrika*, 92(2):351–370, 2005.
- ⁶¹⁵ Camille Vogel, Dorothée Kopp, Fabien Morandeau, Marie Morfin, and Sonia Méhault.
 ⁶¹⁶ Improving gear selectivity of whiting (Merlangius merlangus) on board French demersal
 ⁶¹⁷ trawlers in the English Channel and North Sea. *Fisheries Research*, 193:207–216, 2017.
- Kenneth L. Weinberg and Stan Kotwicki. Factors influencing net width and sea floor
 contact of a survey bottom trawl. *Fisheries Research*, 93(3):265 279, 2008.

⁶²⁰ A Conditioning on the total catch at length and ⁶²¹ station

We compare the model as described in section 2.1 with a variant where we condition on the total catch at length and station. Specifically, let $N_{i\cdot k} = N_{i1k} + N_{i2k}$ be the total catch at station *i* in length group *k*. Then the conditional distribution of the catch in the first haul, N_{i1k} given this total catch $N_{i\cdot k}$ is binomial:

$$N_{i1k}|\Phi, R, S, N_{i\cdot k} \sim \text{Binom}(N_{i\cdot k}, \frac{A_{i1}\exp(S_{1k} + R_{i1k})}{A_{i1}\exp(S_{1k} + R_{i1k}) + A_{i2}\exp(S_{2k} + R_{i2k})})$$
(3)

In turn, the probabilities of the total catches $N_{i\cdot k}$ are

$$N_{i\cdot k} | \Phi, R, S \sim \text{Poisson}(A_{i1} \exp(\Phi_{ik} + S_{1k} + R_{i1k}) + A_{i2} \exp(\Phi_{ik} + S_{2k} + R_{i2k}))$$
(4)

The joint density as developed in section 2.1 could therefore alternatively be written 627 as a product of these binomial probabilities (3), the Poisson probabilities (4), and the 628 prior density of the Gaussian processes Φ , R, S. We may now condition the inference on 629 the total catch $N_{i\cdot k}$ and thus remove the term in the joint density that originates from 630 the total catches $N_{i\cdot k}$, i.e. the terms (4). Since the size distributions Φ do not enter 631 into the conditional probabilities (3), they only appear in the joint density through their 632 prior distribution. Thus, the size distributions Φ vanish after integration, so they can be 633 removed from the model. 634

Figure 6 shows the result from this modified model. The Figure should be compared 635 with Figure 4, which shows the corresponding results for the original model. Notice that 636 the estimates of the selectivity curves are not completely identical, since the omitted 637 term (4) does depend on the selectivities, but still fairly similar. The estimates from the 638 unconditional model are, in general, closer to 1, and the marginal confidence intervals are 639 somewhat wider when conditioning on the total catches. This is not surprising, since this 640 model excludes the information in the total catches. The conditional models has 20 %641 less random effects, which allows faster computations, although our code does not fully 642 exploit this. 643



Figure 1: Example of data and model components. Estimated density $\exp(\Phi)$ of the size distribution at one particular station (thick solid lines). Different nugget effects R apply to the two hauls and results in different size structures encountered by the two hauls (thin solid and dashed lines). The relative selectivity S modifies the expected catch in each size group and for each haul (not shown). Observed counts N in each size group and in each haul are shown with "o" and "+", respectively. Note log scale on the count axis; zero catches are not shown.



Figure 2: Map of the study area.



Figure 3: Density (total catch divided by swept area) by size, summed over all hauls. Left panel: *M. Capensis.* Right panel: *M. Paradoxus.*



Figure 4: Relative selectivity (vessel calibration factor), comparing catches of M. Capensis (left) and M. Paradoxus (right) with Gisund gear on RV Dr. Fridtjof Nansen and FV Blue Sea. Large values indicate that the FV Blue Sea has higher selectivity. Solid curve: Estimated relative selectivity (posterior mode). Grey region: Marginal 95 % confidence intervals for the relative selectivity, computed as $1.96-\sigma$ -intervals on the log scale.



Figure 5: Left panel: Relative selectivity, as in Figure 4, for the two species combined. Right panel: Same, but without the autoregressive component in the nugget effect.



Figure 6: As figure 4, but based on the model where we condition on the total catch in each length group, i.e. without the terms (4) in the likelihood.