Fisheries Research

Estimating spatio-temporal distribution of fish and gear selectivity functions from pooled scientific survey and commercial fishing data

--Manuscript Draft--

Manuscript Number:	FISH10606R1
Article Type:	Research Paper
Keywords:	Spatio-Temporal, GAM, cod, selectivity, surveys
Corresponding Author:	Guillermo Felipe Martin, M.D Marine Institute Rinville, Rinville IRELAND
First Author:	Guillermo Felipe Martin, MS
Order of Authors:	Guillermo Felipe Martin, MS
	Rodrigo Wiff, PhD
	C. Tara Marshall, PhD
	Thomas Cornulier, PhD
Manuscript Region of Origin:	UNITED KINGDOM
Abstract:	Model-based prediction of fish distribution at fine resolutions in space and time has the potential to inform area-based and dynamic forms of management, such as permanent marine protected areas or real-time temporary closures. A major limitation to the spatial and temporal mapping resolution that is achievable is the amount of high quality, standardised data that can be utilized for fitting statistical models. To achieve an adequate spatio-temporal resolution from sparse data, one option is pooling information from several sources, such as scientific surveys and fisheries data. Because surveys and fisheries data usually use different sampling methods, pooling information from different sources requires cross-calibration of catch rates values across multiple gears. However, the individual gear efficiency and selectivity curves (the ratio between catch and availability at a given length) for all fishing gears and species are typically unknown. Using cod (*Gadus morhua*) in the northern North Sea as a case study, we developed a new formulation of spatio-temporal generalised additive models (GAM) of relative abundance of fish, combining catch data from multiple sources. Differences in gear efficiency and selectivity were internally calibrated within the model by the estimation of the local spatio-temporal variation in abundance. We show that pooling data sources enables the prediction of multi-annual and seasonal spatial variation in cod relative abundance-at-size, at spatio-temporal resolutions that are relevant for informing fishing strategies, e.g., reducing bycatch in real-time, or management objectives, e.g., real-time closed areas. We also show that GAM models fit to catch and effort data can reveal the relative efficiency and selectivity of different survey and commercial gears. The selectivity curve estimates that emerged as a by-product of our analysis are consistent with expert knowledge of the performance of the gears employed for cod. Our analytical approach can therefore serve two useful purposes: to estimate spatio-temporal variation in relative abundance of fish and to estimate relative gear efficiency and selectivity.

Estimating spatio-temporal distribution of fish and gear selectivity functions from pooled scientific survey and commercial fishing data

Guillermo Martin Gonzalez[1,2*], Rodrigo Wiff[1,3,4], C. Tara Marshall[1], Thomas Cornulier[1].

1. School of Biological Sciences Zoology Bldg. University of Aberdeen, Aberdeen, United Kingdom.

2. Marine Institute, Rinville, Oranmore, Co. Galway, Ireland

3. Center of Applied Ecology and Sustainability, Pontificia Universidad Católica de Chile, Avenida Libertador Bernardo O' Higgins 340, Santiago, Chile.

4. Instituto Milenio en Socio-Ecología Costera (SECOS), Chile.

(*) Corresponding Author.

Abstract

Model-based prediction of fish distribution at fine resolutions in space and time has the potential to inform area-based and dynamic forms of management, such as permanent marine protected areas or real-time temporary closures. A major limitation to the spatial and temporal mapping resolution that is achievable is the amount of high quality, standardised data that can be utilized for fitting statistical models. To achieve an adequate spatio-temporal resolution from sparse data, one option is pooling information from several sources, such as scientific surveys and fisheries data. Because surveys and fisheries data usually use different sampling methods, pooling information from different sources requires cross-calibration of catch rates values across multiple gears. However, the individual gear efficiency and selectivity curves (the ratio between catch and availability at a given length) for all fishing gears and species are typically unknown. Using cod (*Gadus morhua*) in the northern North Sea as a case study, we developed a new formulation of spatio-temporal generalised additive models (GAM) of relative abundance of fish, combining catch data from multiple sources. Differences in gear efficiency and selectivity were internally calibrated within the model by the estimation of the local spatio-temporal variation in abundance. We show that pooling data sources enables the prediction of multi-annual and seasonal spatial variation in cod relative abundance-at-size, at spatio-temporal resolutions that are relevant for informing fishing strategies, e.g., reducing bycatch in real-time, or management objectives, e.g., real-time closed areas. We also show that GAM models fit to catch and effort data can reveal the relative efficiency and selectivity of different survey and commercial gears. The selectivity curve estimates that emerged as a by-product of our analysis are consistent with expert knowledge of the performance of the gears employed for cod. Our analytical approach can therefore serve two useful purposes: to estimate spatio-temporal variation in relative abundance of fish and to estimate relative gear efficiency and selectivity.

Keywords: Spatio-Temporal, GAM, cod, selectivity, surveys

1. Introduction

High resolution descriptive modelling of the distribution of harvested fish populations in time and space has been the focus of considerable research interest in recent years (e.g. Maunder et al., 2020; Pinto et al., 2019; Stock et al., 2018; Thorson et al., 2020). Understanding the relative contribution of spatial and temporal components in fish distribution has direct implications for implementing spatially-explicit management objectives, particularly avoiding unwanted or bycatch species. Statistical models that incorporate space, time and other covariates are inherently complex but advances in computational implementation has made fitting these models more feasible. One of the most popular statistical approaches for modelling spatio-temporal dynamics in fish populations are the generalised additive models (GAMs, Hastie and Tibshirani, 1990; Wood, 2006). The popularity of GAMs is due to: 1) flexibility in the non-linear relationship between response and explanatory variables using smoothing techniques; and 2) the range of error distributions to model the response variable. GAMs allow a straightforward estimation of spatio-temporal components through bi-variate smooth functions for two geographical coordinates (longitude and latitude) and time can similarly enter as smooth, or fixed or random factor terms (e.g. Jaureguizar et al., 2016; San Martín et al., 2013). GAMs are also computationally efficient for managing datasets that are large but heterogeneous (in space or time) over broad geographical areas, as is often the case for data obtained from research vessel surveys and the fishing industry.

Recent spatio-temporal models developed for commercial fish populations have often focused on resolving the distributions of unwanted components of the catch, either undersized fish or non-targeted species which are potential bycatch (e.g. Cosandey-Godin et al., 2014; Rezende et al., 2019). Unwanted bycatch and discards have been serious concerns globally, posing a threat to the sustainability of fisheries through economic, biological and ecological losses (Komoroske and Lewison, 2015). Beginning in 2015, a discard ban known as the Landing Obligation has been imposed by the Common Fisheries Policy on commercial fishing operating in European waters. The Landing Obligation requires all catches of regulated commercial species on-board to be landed and counted against quota. The Landing Obligation effectively incentivised the fishing industry to avoid catching unwanted species or size classes given that a single, “lightning strike” haul that exhausts the available quota for that target species can potentially result in that vessel being tied up if there was no further quota available to the vessel for the unwanted catch. Incentivising bycatch avoidance has future potential to changing attitudes of skippers towards sharing catch data and behaviors of skippers with respect to the use of spatio-temporal information, for example maps of bycatch hotspots (e.g. Merrifield et al., 2019).

The data used to develop spatio-temporal models are usually obtained from research vessel surveys which are conducted at fixed time intervals and utilize a statistical survey design intended to produce unbiased estimators of abundance. An alternative source of data describing commercial fish populations is generated by fishing vessels, normally collected for compliance purposes. Combining both data sources together presents several advantages and disadvantages for modelling fish distribution. Commercial fishing data are a rich source of highly resolved spatio-temporal sampling of fish distributions, however, these data are associated with challenging statistical features including a high proportion of zero observations (Kai et al., 2016; Maunder et al., 2020), non-random spatial sampling (bias) (Conn et al., 2017; Diggle et al., 2010; Pennino et al., 2019) and temporal correlation (Ciannelli et al., 2008; Cosandey-Godin et al., 2014). Survey data have several advantages over commercial data because sampling location is determined independently of local abundance using an underlying statistical design, e.g. stratified random sampling, and consistent methods are applied over decadal time scales. However, the relatively low intensity of survey data limits the spatio-temporal resolution of predictions. Recent advances in spatio-temporal modelling have combined data from commercial fishing and survey sources (Pinto et al., 2019), enhancing spatio-temporal coverage and resolution, and resulting in improving the analysis and understanding of population dynamics. Nevertheless, pooling data from several sources, is challenging because requires cross-calibration of catches rates across multiple gears. However, selectivity curves for individual gears (the ratio between catch and availability at a given length) are typically unknown, since availability at the time and location of the fishing operation is not directly observable, and gear efficiency experiments are relatively expensive and difficult to perform.

The clustered nature of fish distribution in space and time (Swartzman et al., 1992) results in localised areas of high abundance or "hotspots" which limits the ability of fishers to manage their portfolio of quotas for different species (Bailey et al., 2010). This clustering also adds to the challenge of developing predictive models that are sufficiently resolved and reliable to be used to inform tactical decisions at sea. The development of such models will aid manage unwanted species of size-classes in the catch. Recently, Pinto et al., 2019 proposed a spatial model to predict occurrence (presence/absence) of data-limited species by combining surveys and commercial fishing data.

The underpinning rationale in the Pinto et al., 2019 approach is to reduce heterogeneity in catchability to a point where it should not affect the resulting inference. This was done by reducing the information to only presence/absence of the species, and by selecting a relatively homogenous subset of the data (based on mesh size). Drawbacks of the approach include the loss of information on stock abundance and size structure due to conversion to presence/absence, and the limited ability to integrate data sources with heterogeneous catch effort and catchability. Furthermore, modelling bycatch so as to be sensitive to the size structure of the stock (e.g., to predict local abundance of juveniles) brings an extra complexity when combining datasets, because size-selectivity of different fishing gears needs to be accommodated into the modelling approach.

Selectivity refers to the probability of catching to availability at different age/sizes. The term selectivity is known as direct, if the population age/size structure is known (estimated) as is the case of selectivity resulting from the application of an integrated stock assessment model (Quinn and Deriso, 1999). On the other hand, experimental selectivity usually is indirectly estimated by comparing catches from different variants of the same fishing gear (Millar, 1992). For example, indirect selectivity studies in trawling are based on the covered codend method, where a small mesh cover is attached to the outside of the codend to retain most of the fish that pass through the codend mesh (Madsen and Holst, 2002). Another experimental method is the trouser trawl which consists of a net with two codends, each one constructed from a different size and/or shape of mesh (Cadigan and Millar, 1992). Although commonly undertaken by gear technologists several decades ago, indirect experiments to estimate selectivity are expensive because they require specially designed surveys and modifications of the routinely used commercial fishing gear. Millar, 1992 proposed an alternative statistical method to estimate indirect selectivity from size structures retained in the total catches in fisheries in which variants of the same fishing gears (e.g. different types of trawling nets) are operating simultaneously in the fishing area, without the requirement of a covered codend design. Millar's method of statistical inference relies on estimating the parameters of predetermined selectivity at length functions, using generalised linear models to estimate the probability of being caught by a fishing gear.

Using cod (*Gadus morhua*) in the Northern North Sea as a case study, we develop a new spatio-temporal GAM of fish relative abundance combining survey and commercial fishing data to explore the potential of high-resolution distribution models as bycatch reductions tools. Our model-based approach to gear selectivity calibration has conceptual similarities with that of Millar, 1992, but draws on the opportunities that GAMs provide to estimate catch per unit effort, by jointly modelling the selectivity function of each fishing gear and the spatio-temporal variation in availability. Through internal cross-calibration of observations per unit effort for different sources of data, our model is capable of: 1- estimating high-resolution spatio-temporal trends of cod abundance across length classes, distinguishing repeatable from stochastic components of distribution between and within years/seasons, and 2- approximate the relative efficiency and selectivity of survey and commercial fishing gears, without the need for costly paired gear experiments.

2. Material and Methods

Study Area

The full application of the Landing Obligation in 2019 created a strong incentive for developing effective bycatch avoidance strategies to reduce the probability of catching either choke species or undersized fish that must be counted against the quota but cannot be sold (Needle et al., 2015). In Scotland, there is previous experience with using spatially resolved catch data to reduce catch of juvenile cod in the North Sea (Kraak et al., 2013). For a limited number of years real-time closures were established and had a good level of compliance

(Holmes et al., 2009; Little et al., 2015; Needle and Catarino, 2011). Currently, there is renewed interest in applying spatial management measures to conserve North Sea cod (C. Needle, Marine Scotland Science, personal comm.) which would benefit from combining different sources of data to identify spatial hotspots of unwanted catches (Marshall et al., 2017).

Survey Data.

The North Sea International Bottom Trawl-survey (NS-IBTS) is a demersal trawl survey conducted twice a year (quarter 1 & 3) since 1997 and coordinated by the *International Council for the Exploration of the Sea* (ICES). The NS-IBTS is a multispecies survey with standardised data sampling and processing design and provide data for estimation of relative abundances for fish species in an area within 51^0-62^0 N latitude and 4^0W -9^0E longitude and depths shallower than 300 m (Fig.1). Participating countries use a standardised trawl gear for data collection. The survey area is sampled following a stratified sampling design based on ICES statistical rectangles of approximately 60x30 nautical miles (1-degree longitude x 0.5-degree latitude). Each country is allocated a certain number of rectangles to sample and surveys are organised so that each rectangle has at least two hauls sampled by two different countries (ICES, 2015). Information available for each haul includes georeferenced starting/finishing trawling (latitude and longitude), date, count of cod caught per length class to the nearest centimeter and additional information such as trawling depth in meters (ICES, 2015). In terms of the Landings Obligation, juvenile cod were defined as individuals < 35 cm of total length. In this study, we used cod count data from 2011-2015 downloaded from ICES (https://datras.ices.dk/Data_products/Download/Download_Data_public.aspx; access date: April 2017) which summed up to a total of 2,462 hauls uniformly distributed throughout the North Sea (Fig.1). Fishing effort was 0.5 hours for 99.78% of hauls. For the remaining 0.22% of hauls, the raw counts were standardized to 0.5 hours (a legacy of preliminary analyses)

Commercial Fishing data

The commercial catch data were obtained from the discard monitoring programs conducted by Marine Scotland Science (MSS) and Scottish Fishermen's Federation (SFF). A common sampling protocol is used by onboard scientific observers in both programs. The MSS and SFF programs select vessels to carry observers using a stratified random sampling design by area, gear, and quarter within each year (Jermyn and Robb, 1981). The data collected for each haul includes count of a sample of cod by length-class to the nearest centimeter, trawl duration, depth of trawling, and information regarding operational characteristics such as gear type or mesh size. The total number of individuals at length is then scaled-up to the total amount of catch per hauls. Effort is predefined as trawling time.

Commercial fishing data available for research purposes were anonymised to protect confidentiality. Although the MSS/SFF observer program covers a wider area than the NS-IBTS survey, only data for the North Sea (55^0-61^0 N latitude & 5^0-7^0 W longitude) between 2011-2015 were available for the analysis (Fig.1). Total data used from commercial fishing represented 3,585 hauls covering all quarters, in depths shallower than 200m and included multiple fishing fleets targeting different species using nine different trawl-type gears. These fishing gears are coded as: Seine net (SEN), Light trawl (LTR), Multiple trawl heavy (MTH), Multiple trawl demersal (MTD), Multiple trawl nephrops (MTN), Pair trawl demersal (PTD), Industrial Trawl (ITR), Nephrops trawl single (NTR) and Single trawl demersal (MTR).

Statistical model

The number of cod at length was modelled with a generalised additive model with mixed-effects (GAMs, Hastie and Tibshirani, 1990; Wood, 2006). The proposed model used smooth functions of geographic location, lengths, fishing gear and temporal attributes (month, year) and vessels as random effect. A general form is given by the following expression:

$$E[Y]=g^{-1}[F + \beta_0 + \sum_i s_i(X_i) + \boldsymbol{Z}u] \quad (1)$$

where $E[Y]$ is the expected catch of fish at length, g is the link function which defines the relationship between the response and the linear predictor $[F + \beta_0 + \sum_i s_i(X_i) + Zu]$. β_0 is the intercept, F is an offset (variable with fixed coefficient 1), X_k corresponds to the k-th covariate and $s_k(\circ)$ is a smooth function of the k-th variable, with a shape to be estimated from the data. $\boldsymbol{Z}$ is the design matrix for the random effect u.

The specific model we propose to model cod catch accounts for fixed, random effects and interactions as follows:

$$ln(Y) = F + \beta_0 + ti_1(E, N, M, L) + ti_2(D, M, L) + ti_3(D, L) + ti_4(M, L) + ti_5(M) + ti_6(D) + s_{1,g}(L) + u_v + \beta_{1,y} + \beta_{2,g} \quad (2)$$

Where $ti(\circ)$ define a smooth tensor product interaction applied to the variables month (*M*), length class (*L*), depth (*D*), longitude/eastings (*E*), latitude/northings (*N*). $ti(\circ)$ terms used a combination of thin plate regression splines (*tp*) for depth and length class, thin plate regression splines with smoothing penalties (*ts*) for longitude and latitude and cyclic cubic regression spline (*cc*) (i.e. a penalized cubic regression splines whose ends match) for month. *tp* splines were selected among other options for being considered the optimal default spline for any given dimension/rank (Wood, 2006). ti_1 accounts for the spatio-temporal distribution of fish, assuming continuous variation across length classes. Thus, the model assumes that hauls that are close in space and time will have more similar counts of cod at length class than those that are widely separated. ti_2 defines the pattern of depth use varying continuously with fish length class. ti_3, ti_4, ti_5 and ti_6 are the nested interactions and main effects of t_1 and t_2 and are designed to be orthogonal (Wood 2017). $s_{1,g}(\circ)$ is *tp* smooth term of length class for each fishing gear (g) to capture the length-specific gear selectivity. u_v is a normal random effect of individual vessel (v) with mean zero and variance σ_v^2, and $\beta_{1,y}$ and $\beta_{2,g}$ are fixed effects of year (y) and gear (g) respectively, allowing to account for correlation within groups (Candy, 2004).

Coming back to the more general model formulation in Eq 1, using a natural log-link function, and defining *F* as the natural log of sampling effort, and including the gear effect, we specify the model:

$$ln(Y) = ln(effort_g) + \beta_0 + \sum_i s_i(X_i) + Zu + \beta_{2,g} + s_{1,g}(L) \quad (3)$$

which can be re-written as:

$$\frac{Y}{effort_g} = e^{\beta_0 + \sum_i s_i(X_i) + Zu + \beta_{2,g} + s_{1,g}(L)} \quad (4)$$

Showing the classical log-link model formulation with log-effort as an offset, in which coefficients can be interpreted as the multiplicative effect of predictors on catch per unit effort

(CPUE) (Candy, 2004). In our case a unit effort in one gear type may not be equivalent in CPUE to a unit effort in another gear type. For example, for a given trawling time, a seine and a bottom trawl net are unlikely to yield the same catch due to operational characteristics (in fact in the general case, effort may not even be measured in the same currency for all gear types, with efforts sometimes being described by area covered, or duration or their combination). This leads to a potential difficulty of interpretation of the model.

However, re-parametrization of the gear effects $\beta_{2,g} + s_{1,g}(L)$ yields:

$$\frac{Y}{e^{\beta_{2,g}+s_{1,g}(L)} \times effort_g} = e^{\beta_0 + \sum_i s_i(X_i) + Zu} \quad (5)$$

in which it can be seen that the linear predictor of relative abundance is independent of gear, and the combined gear effects $e^{\beta_{2,g}+s_{1,g}(L)}$ estimate the gear-specific effort adjustment factor for a given fish length. This gear term therefore represents the relative selectivity of survey and commercial fishing gears, and their efficiency relative to the reference gear (here arbitrarily defined as the survey gear).

Several options exist for modelling the distributions of count data, including the Poisson or negative binomial, since all these distributions bound the predictions down to zero. In line with the expectation for schooling fish species, preliminary examination of residual plots indicated significant over-dispersion ($\frac{Residual\ deviance}{n-p\ (df)}$>1) compared to the expectation from a Poisson distribution, and the negative binomial was chosen.

Components of the model in Eqn (2) were chosen to represent the length-specific seasonal variation in the spatial distribution of fish at length which is the primary focus of the study, alongside variation due to vessel and gear reflecting important sampling design features. We used AIC to compare two models with and without the vessel random effect, to evaluate its importance. Models were fitted using Restricted Maximum Likelihood estimation (REML). Other terms in the model were core to the aims of the study and were not subjected to selection.

The optimal shape and degree of smoothing for the non-linear terms was automatically estimated from the data as part of the model fitting procedure using the REML method implemented in the ‘gam’ function (Wood, 2006). Correct specification of the number of knots for each smoothing term was checked after initial model fitting following the protocol described in Wood, 2006. Trials to increase the number of knots in smoothing terms resulted in a significant intensification of computational power and convergence fails (approximately

12 hours of computing time with default settings, in a 6th Gen Intel Skylake Core i7-6700HQ 256GB SSD, 16GB RAM processor type for a total of 134,221 fishing operations and 192 knots for ti_1). The goodness-of-fit was visually assessed using residuals and fitted values versus observed values and model covariates to explore heterogeneity and independence assumptions (Supplementary materials; Fig.1).

Predictions from the model were computed on a grid with a total of 2,141 cells, each representing 1/16 of an ICES statistical rectangle. Cod length was grouped in 2cm bins to reduce computational requirements. For each haul, zeros were included when no cod was captured in a given size class. For visualization purposes, lengths classes were further grouped in 10 cm bins. Thus, model predictions represent the average counts of cod in 10 cm groups within each grid cell for a given year/month and by a specific fishing gear. Average depth within each grid cell was used for model predictions. Bathymetry data was downloaded from General Bathymetric Chart of the Oceans (http://www.gebco.net/data_and_products/gridded_bathymetry_data/; access date: April 2017). Coordinates of the center of each grid cell were used as latitude and longitude predictors.

To understand the contribution of each data source in model inference and emphasize the benefits of data pooling, a second model excluding the survey data was implemented using Eq.2. Predictions of both models, with and without scientific data were then computed, using the grid described above, to assess the correlation of the predictions between both data sources. As we suspect the quality of the available data is strongly affected by the different strata of our model, predictions were compared in four months and five different length classes.

All analyses were conducted using R (R Core Team, 2021), and code is available at https://github.com/gfmg/Spatiotemporal-Cod

3. Results

A comparison of the two models in Table 1 indicated that the vessel effect is relevant, as it considerably lowered the value of AIC. All the smoothing components of the model were found to have a significant effect ($p<0.05$, Table 2). Other terms in the model, such as gear effects, relate to the model or study design and had to be included irrespective of their statistical significance. Their statistical significance is irrelevant to the objectives of this study but were reported (along with confidence intervals) in the interest of illustrating some of the features of the data and of the corresponding estimates. Fixed year effects were all different from 2011, except for year 2013. While the reality of differences in catch efficiency across the variety of gears studied is unquestionable (and not part of the research questions of interest for the present study), differences between scientific ("GOV") and commercial gears (all others) in mean efficiency across all lengths were not significant (p>0.05, Table 2), reflecting the limited precision of the estimates relative to the size of the true differences. The model performed well, with a total explained deviance of 60.8% and a coefficient of determination R^2 between log-observed and log-predicted values of 0.371 (Table 2, Supplementary Materials; Fig.2).

Figure 2 provides an indication of the amount of spatio-temporal overlap between the deployments of different gears, which is key to enabling the estimation of relative gear efficiencies in the model. Thanks to an extensive spatial coverage in two quarters of the year, the IBTS scientific survey has a good volume overlap with most commercial gears. One commercial gear (MTR) overlapped with a large proportion of other gears, whereas other pairs of commercial gears ranged from a strong spatio-temporal overlap to an almost complete segregation.

Our analysis confirms that the data contain sufficient information to estimate relative gear-specific selectivity at length, as indicated by the significant non-linear effect of length on catch rates per unit effort, for all gear types. The smooth selectivity functions in Figure 3 showed different patterns across fishing gears, with uncertainty increasing for lengths smaller than 10 cm, as expected given the high escape rate and resulting usual lower sample size for small fish. Most gears show an asymptotic behavior after 30 cm, except for MTD, MTN and NTR were a decrease in the effectivity of the fishing gear to catch these size classes is suggested (Fig. 3). Comparison across fishing gears show a broadly similar pattern among them, except for GOV and ITR (Fig 3). GOV (survey) gear is the least selective gear for fish <30 cm, due to the attachment of a codend in survey nets. Among commercial vessels, ITR, MTD, PTD and SEN gears tend to be more effective at excluding fish <30cm. In particular, ITR gear shows considerably high selectivity against fish <30cm.

Figure 4 illustrates the marginal main effects of months (ti_5 in Eq 2), depth (ti_6) and years on CPUE of juvenile fish. The marginal effect of month showed a peak between July and August. Afterward, between September and February CPUE of cod decreased with a minimum found during March-April. Higher relative abundances of cod were found at depths between 50 and

100 m with low numbers of data and higher levels of uncertainty beyond 150 m (Fig. 4). Overall cod relative abundance was maximal in 2015 and 2014 and lowest in 2012 (Fig. 4).

In order to evaluate variation in the effects of depth across months and lengths, we used the sum of marginal predictions of terms 'ti_2', 'ti_3' *and* 'ti_6' in Eqn 2. Standardised marginal effect of depth across months and lengths is presented in Figure 5, where a clear separation is found between fish below and above a length of 25 cm. Fish smaller than 25cm are generally found in shallower depths. A general pattern can be observed in autumn/winter months (November to March) when most of fish < 25 cm appear at depths ranging from 80 to 10 m. On the other hand, during late spring and summer months (May-August) the separation of pattern moves around 35 cm with an increase of length with depth, with the majority of fish larger than 40 cm are in depth of around 150 m. Fish below 10cm were only found during July, August and September at very shallow depths below 50m. Conclusions from other periods of the year cannot be drawn for these smaller fish as few of these sizes classes are present in the data.

Model predictions of cod abundance on the log scale for 10 cm length intervals are shown in Figure 6 for February, May, August and November. The proposed model captured clear differences in the spatial distribution of CPUE across length strata and time of year. Markedly different distribution patterns were found for cod under and above the 20-30cm length stratum (Fig. 6), reinforcing evidence for a bimodal spatial behavior of cod, with a split around a length of 25 cm, previously described in Figure 5. Likewise, differences were found among months in which February and May were characterized by the lack of 0-10 cm (0-group) cod. During these months the model also predicted a hotspot near the southeast limit of the study area, while the distribution of higher length classes (20-30, 30-40 & 40-50 cm) appeared more homogeneously distributed in northern zones, yet changing seasonally (Fig. 6). Depth was an important predictor of the relative abundance of cod. For example, smaller fish (<20 cm) were found closer to the coastline, while larger fish were located in deeper waters matching the bathymetry contours of -50 and -100m depth (Fig.6). Results indicated that the model was able to capture both large and small scale patterns of relative abundance distribution of specific lengths of cod, in each month of the year.

Figure 7 shows predictions of models fitted with and without the scientific survey data, across four different months and five different length classes. The plots show two contrasting behaviors. For the larger length classes of fish (30-50 cm) there was a good correlation between both models, whereas the smallest fish (10 cm) showed a strong divergence, with high predicted values in the model fitted with only the commercial data. The 20 cm fish show a transition pattern between the two behaviors described above, with good correlation between the model predictions only in February and May, months were the IBTS survey occurs (Feb) or where commercial gears tend to catch these size classes.

4. **Discussion**

We proposed a modelling approach which combines commercial fishing and survey data to predict the spatio-temporal distribution of cod. One of the main strengths of this approach is the ability to combine information about catch rates coming from gears with different selectivities. Such combination of information is achieved by implementing an additive model structure where heterogeneity caused by factors affecting catch rates were incorporated into a model via fixed or random effects. Combining information from different sources is firmly rooted in fisheries sciences, particularly in the area of integrated stock assessment modelling (Maunder and Punt, 2013). However, when dealing with spatial and temporal modelling the combination of data sources is less common because heterogeneity across fishing methods (i.e. commercial fishing vs surveys) needs to be incorporated into the modelling approach. Recently, Pinto et al., 2019 combined survey and commercial fishing data to model the spatio-temporal dynamics with application to data-limited fisheries. The model in Pinto et al., 2019, is based only on presence/absence of fish in space and the critical assumption is that variability in selectivity and catchability across sampling methods is negligible in informing the presence/absence of fish. In other words, if the studied fish species is present in the area, any of the sampling methods should be able to detect it with the same probability. Our modelling framework follows a similar philosophy as the analysis in Pinto et al., 2019, in the sense that all sampling methods (different fishing gears) can detect cod. However, in our case gear efficiency needs to be accounted to accommodate differences in catch rates among gears, and because we aim to model sizes, we made extra allowance for differential selectivity of different fishing gears into the model. We incorporated length effect differences between sampling methods using spline-based models which proved to be a flexible framework to accommodate differences in gear selectivity. The shape of these selectivities estimated within the model broadly agreed with expert knowledge for cod in the North Sea (B. O'Neill, Marine Scotland Science, personal comm).

In our study we expect sampling bias by commercial fishers to be reduced by their limited ability to trigger a fishing operation following detection of demersal fish, but also due to our focus on undersized fish which are not directly targeted by the fleets (although some degree of active avoidance may occur). The model approach further makes estimates more robust to spatially-biased sampling effort. Indeed, the use of flexible smoothing splines serves the same function as the spatial random fields in Pinto et al., 2019, making the estimation of model parameters more local spatially and temporally. Intuitively, a more local estimation means a higher reliance on a smaller pool of local data, and a lesser influence of more distant and potentially over-sampled areas of low or high fish abundance (depending on the nature of the sampling bias). Nevertheless, model estimation benefits hugely from highly resolved commercial data being complemented by systematic surveys, to fill potential data gaps particularly in the size classes of fish that are not targeted by commercial gears, as the comparison of predictions after excluding the scientific sources suggests. Bias in the abundance estimates due to non-random sampling by the commercial fleet may not always be possible to eliminate effectively, particularly where sampling is very heavily biased towards high fish densities, as would be the case in pelagic detect-and-catch fisheries without complementary survey data. In this case, jointly modelling the sampling process and abundance may be required (Conn et al., 2017; Diggle et al., 2010; Pennino et al., 2019).

Due to their extensive spatial coverage, survey data also ensured a consistent spatio-temporal overlap with most other sources of data, and thus played an important role in enabling the cross-calibration of CPUE between gears. Although in the case of the North Sea cod, dropping the survey data generated some noise in the estimates, there was still enough information in the commercial data alone for the model to perform well (Fig 7), thanks to the partial overlap between pairs of commercial gears, despite none of them covering the entire study area (Fig. 2 and Supplementary materials).

Our model estimates reveal important aspects of the biology and life cycle of North Sea cod, including the size-specific bathymetric preferences and seasonal changes in distribution. In the North Sea, cod spawning occurs between February and April, while larvae settling during the late summer/ autumn months (Brander, 2005). The increase in relative abundance for cod between 0-10cm predicted by the model during the August/November months in the shallower South-Eastern limit of the North Sea and coastal areas of Scotland, agrees with previously reported nursery grounds (Brander, 2005; Fox et al., 2008; Lelièvre et al., 2014). Larger size classes of cod are known to appear in deeper waters and broadly geographically distributed throughout the North Sea. In addition, our model shows a clear pattern of sizes and depth in which cod smaller than 25 cm rarely appearing beyond 100 m of depth. This is consistent with Neat et al., 2006, who reported that cod generally stayed deeper during winter/spring from mark-recapture experiments. Our results describing the distribution of very small fish (<10 cm) should be interpreted with caution due to the limited sample sizes. However, results of distribution of smaller size classes can be seen as suggesting the behaviours of smaller fish which should be considered when designing nursery areas or closed areas for fisheries management purposes.

A major strength of the modelling framework is to combine commercial and survey data, accommodating high resolution variations of cod at length, season, depth, space and fishing gear. High spatio-temporal resolution is enabled by combining interaction smoothing terms giving account of the space (latitude, longitude), depth, month effects, using a (very) large number of knots, with the achieved resolution being determined statistically by the amount of data available locally and/or practically by increasing computational costs. The modelling of a fine scale spatial pattern in many cases will prove more useful for the fishing fleet in terms of fishing suitability maps. In fisheries with large amount of bycatch, the location of areas that are temporarily closed to fishing in order to avoid unwanted catch is determined empirically from real-time data provided by fishing vessels and shared among other members of the fleet (Little et al., 2015). Several examples worldwide have shown successful outcomes as the finer spatial and temporal scale of this approach better addresses fisheries management issues (e.g. Bailey et al., 2010; Cosandey-Godin et al., 2014). Compared to richer mechanistic population models (e.g. Kristensen et al., 2014), the high resolution, simplicity and computational speed of our descriptive modelling method mean that it has potential for informing spatial fisheries management in near-real time.

The proposed approach aims at modelling spatio-temporal variation in fish relative abundance across size classes and could aid fisheries managers to tackle bycatch problems or inform marine spatial planning. Moreover, recent advances in CPUE standardization have recognized the importance of explicitly including spatio-temporal correlation in catch rates (e.g. Grüss et al., 2019; Maunder et al., 2020; Thorson, 2019; Zhou et al., 2019). We suggest that our spatio-temporal modelling approach may be used for CPUE standardization, adding the ability to standardize CPUE across multiple commercial or scientific gears at once. Indeed, abundance or recruitment indices may be derived by integrating our model predictions over areas and length classes of interest, for example in the context of integrated stock assessment. In doing so, practitioners should pay attention to selecting an appropriate time of the year as a reference. Where a biomass index is required, weight-at-length relationship may be used for conversion, although it could be advantageous to directly model the log-biomass of the catch in place of fish count, when fitting the model.

An emergent property of the proposed model, not formally noted in previous investigations of cross-calibration studies (e.g. Punt et al., 2000), is the potential for generating selectivity curves from different fishing gears from routine commercial and observational data. This can be seen as a by-product of the spatial model, because these selectivity curves are needed to accommodate differences between fishing gear efficiencies into the modelling approach. Our approach to estimating selectivity functions within the model has some similarities with that of Millar, 1992. In this study, Millar used generalised linear models to estimate the parameters of predetermined selectivity at length functions, using the contrast between the size structures retained in two fishing gears operating experimentally in close proximity. The greater flexibility of GAMs allows us to relax two major constraints of Millar's method. First, smoothing splines of fish size effects remove the need to assume a parametric form of the selectivity function. Second, the smooth spatio-temporal terms effectively act as a soft statistical pairing device, performing "on the fly" matching of fishing operations with different gear types based on spatio-temporal proximity. The approach therefore relies on having partial spatial and temporal mixing of different sampling gears, as was the case in the North Sea cod data. It draws on a similar idea as Punt et al., 2000, but introducing size-dependence of relative catch efficiency, allowing arbitrary catch effort currencies, and relaxing reliance on pre-defined arbitrary regions. The method appears to work well in a data-rich context like the North-Sea cod, producing results consistent with the empirical knowledge of the stock and of the gears in use. It is unclear yet the extent to which reducing the resolution of the spatio-temporal trends in the model would lower the precision of relative gear efficiency estimates, for example in the case of more data-poor species. Simulation studies and further application of the proposed method would be required in this regard, but earlier work by Punt et al., 2000 suggests that much coarser spatial resolutions would still afford useful estimates. More generally, empirical and simulation studies would also be useful to better understand factors affecting the accuracy of the relative abundance estimates, including the use of more environmental predictors, and the levels of spatio-temporal overlap between gears that are likely to be sufficient for effective CPUE cross-calibration. Our statistical method for jointly estimating spatio-temporal fish distributions and relative selectivity of any number and type of gears is promising, because the shape of the selectivity is usually unknown in most fisheries. It synthesises and considerably generalises previous methods by Millar, 1992; Punt et al., 2000, and Pinto et al., 2019, by providing a flexible and generic approach to estimating indirect selectivity without the need to undertake costly gear research.

5. Credit authorship contribution statement

Guillermo Martin Gonzalez: concept, coding, writing original draft, visualisation and editing**. Rodrigo Wiff:** concept, coding, writing original draft, editing. **C. Tara Marshall:** concept, draft preparation, project administrator, funding acquisition, editing. **Thomas Cornulier:** concept, model design, draft preparation, funding acquisition, editing.

6. Acknowledgements

This work was part-funded by FISA project Number 01/15 and Fisheries Innovation Scotland (FIS011B). G. Martin Gonzalez is grateful to the scholarship program 'Becas de La Caixa' for sponsoring his MSc, where part of this work was developed. R. Wiff was part-funded by CONICYT Project CAPES FB 0002 and by ANID – Programa Iniciativa Científica Milenio – Código ICN2019_015. We thank Marine Scotland Science for providing the anonymized commercial fishing data. Coby Needle initiated interest in the problem. Barry O'Neill and Rui Catarino assisted with interpreting the selectivity findings. Max Lindmark conducted preliminary analyses of the data as part of his MSc thesis. We are sincerely grateful to two anonymous reviewers whose comments and suggestions greatly improved an early version of this manuscript.

7. References

Bailey, N., Campbell, N., Holmes, S., Needle, C., Wright, P., 2010. Real-time closures of fisheries, Study PE 438. 598 for the European Parliament. External report.

Brander, K.M., 2005. Spawning and life history information for North Atlantic cod stocks, ICES Cooperative Research Report, 274. ICES: Copenhagen.

Cadigan, N.G., Millar, R.B., 1992. Reliability of selection curves obtained from trouser trawl or alternate haul experiments. Can. J. Fish. Aquat. Sci. 49, 1624–1632. https://doi.org/10.1139/f92-181

Candy, S.G., 2004. Modelling catch and effort data using generalised linear models, the Tweedie distribution, random vessel effects and random stratum-by-year effects, CCAMLR Science.

Ciannelli, L., Fauchald, P., Chan, K.S., Agostini, V.N., Dingsør, G.E., 2008. Spatial fisheries ecology: Recent progress and future prospects. J. Mar. Syst. 71, 223–236. https://doi.org/10.1016/j.jmarsys.2007.02.031

Conn, P.B., Thorson, J.T., Johnson, D.S., 2017. Confronting preferential sampling when analysing population distributions: diagnosis and model-based triage. Methods Ecol. Evol. 8, 1535–1546. https://doi.org/10.1111/2041-210X.12803

Cosandey-Godin, A., Krainski, E.T., Worm, B., Flemming, J.M., 2014. Applying Bayesian spatiotemporal models to fisheries bycatch in the Canadian Arctic. Can. J. Fish. Aquat. Sci. 72, 186–197. https://doi.org/10.1139/cjfas-2014-0159

Diggle, P.J., Menezes, R., Su, T. li, 2010. Geostatistical inference under preferential sampling. J. R. Stat. Soc. Ser. C Appl. Stat. 59, 191–232. https://doi.org/10.1111/j.1467-9876.2009.00701.x

Fox, C.J., Taylor, M., Dickey-Collas, M., Fossum, P., Kraus, G., Rohlf, N., Munk, P., van Damme, C.J.., Bolle, L.J., Maxwell, D.L., Wright, P.J., 2008. Mapping the spawning grounds of North Sea cod (*Gadus morhua*) by direct and indirect means. Proc. R. Soc. B Biol. Sci. 275, 1543–1548. https://doi.org/10.1098/rspb.2008.0201

Grüss, A., Walter, J.F., Babcock, E.A., Forrestal, F.C., Thorson, J.T., Lauretta, M. V., Schirripa, M.J., 2019. Evaluation of the impacts of different treatments of spatio-temporal variation in catch-per-unit-effort standardization models. Fish. Res. 213, 75–93. https://doi.org/10.1016/j.fishres.2019.01.008

Hastie, T.J., Tibshirani, R.J., 1990. Generalized Additive Models. https://doi.org/10.4135/9781412950589.n368

Holmes, S.J., Campbell, N., Aires, C., Fernandes, P.G., Catarino, R., Bailey, N., Barratt, & K., 2009. Using VMS and Fishery Data in a Real Time Closure Scheme as a Contribution to Reducing Cod Mortality and Discards. ICES C. 2009/M13.

ICES, 2015. Manual for the International Bottom Trawl Surveys. https://doi.org/10.17895/ices.pub/7583

Jaureguizar, A.J., Wiff, R., Clara, M.L., 2016. Role of the preferred habitat availability for small shark (Mustelus schmitti) on the interannual variation of abundance in a large Southwest Atlantic Coastal System (El Rincón, 39◦–41◦S). Aquat. Living Resour. https://doi.org/10.1051/alr/2016031

Jermyn, A., Robb, A., 1981. Review of the cod, haddock, and whiting discarded in the North Sea by Scottish fishing vessels for the period 1975–1980. ICES 47.

Kai, M., Thorson, J.T., Piner, K.R., Maunder, M.N., 2016. Spatiotemporal variation in size-structured populations using fishery data: An application to shortfin mako (Isurus oxyrinchus) in the Pacific Ocean. Can. J. Fish. Aquat. Sci. 74, 1765–1780. https://doi.org/10.1139/cjfas-2016-0327

Komoroske, L.M., Lewison, R.L., 2015. Addressing fisheries bycatch in a changing world. Front. Mar. Sci. 2, 83. https://doi.org/10.3389/fmars.2015.00083

Kraak, S.B.M., Bailey, N., Cardinale, M., Darby, C., De Oliveira, J.A.A., Eero, M., Graham, N., Holmes, S., Jakobsen, T., Kempf, A., Kirkegaard, E., Powell, J., Scott, R.D., Simmonds, E.J., Ulrich, C., Vanhee, W., Vinther, M., 2013. Lessons for fisheries management from the EU cod recovery plan. Mar. Policy 37, 200–213. https://doi.org/10.1016/j.marpol.2012.05.002

Kristensen, K., Thygesen, U.H., Andersen, K.H., Beyer, J.E., 2014. Estimating spatio-temporal dynamics of size-structured populations. Can. J. Fish. Aquat. Sci. 71, 326–336. https://doi.org/10.1139/cjfas-2013-0151

Lelièvre, S., Vaz, S., Martin, C.S., Loots, C., 2014. Delineating recurrent fish spawning habitats in the North Sea. J. Sea Res. 91, 1–14. https://doi.org/10.1016/j.seares.2014.03.008

Little, A.S., Needle, C.L., Hilborn, R., Holland, D.S., Marshall, C.T., 2015. Real-time spatial management approaches to reduce bycatch and discards: experiences from Europe and the United States. Fish Fish. 16, 576–602. https://doi.org/10.1111/faf.12080

Madsen, N., Holst, R., 2002. Assessment of the cover effect in trawl codend selectivity experiments. Fish. Res. 56, 289–301. https://doi.org/10.1016/S0165-7836(01)00330-7

Marshall, C.T., Wiff, R., Cornulier, T., 2017. Using Commercial and Survey Data to Infer Real-Time Fish Distribution in the North Sea at High Resolution. https://doi.org/10.7489/1973-1

Maunder, M.N., Punt, A.E., 2013. A review of integrated analysis in fisheries stock assessment. Fish. Res. 142, 61–74. https://doi.org/10.1016/j.fishres.2012.07.025

Maunder, M.N., Thorson, J.T., Xu, H., Oliveros-Ramos, R., Hoyle, S.D., Tremblay-Boyer, L., Lee, H.H., Kai, M., Chang, S.K., Kitakado, T., Albertsen, C.M., Minte-Vera, C. V., Lennert-Cody, C.E., Aires-da-Silva, A.M., Piner, K.R., 2020. The need for spatio-temporal modeling to determine catch-per-unit effort based indices of abundance and associated composition data for inclusion in stock assessment models. Fish. Res. 229. https://doi.org/10.1016/j.fishres.2020.105594

Merrifield, M., Gleason, M., Bellquist, L., Kauer, K., Oberhoff, D., Burt, C., Reinecke, S., Bell, M., 2019. eCatch: Enabling collaborative fisheries management with technology. Ecol. Inform. 52, 82–93. https://doi.org/10.1016/j.ecoinf.2019.05.010

Millar, R.B., 1992. Estimating the size-selectivity of fishing gear by conditioning on the total catch. J. Am. Stat. Assoc. 87, 962–968. https://doi.org/10.1080/01621459.1992.10476250

Neat, F.C., Wright, P.J., Zuur, A.F., Gibb, I.M., Gibb, F.M., Tulett, D., Righton, D.A., Turner, R.J., 2006. Residency and depth movements of a coastal group of Atlantic cod (Gadus morhua L.). Mar. Biol. 148, 643–654. https://doi.org/10.1007/s00227-005-0110-6

Needle, C.L., Catarino, R., 2011. Evaluating the effect of real-time closures on cod targeting. ICES J. Mar. Sci. 68, 1647–1655. https://doi.org/10.1093/icesjms/fsr092

Needle, C.L., Dinsdale, R., Buch, T.B., Catarino, R.M.D., Drewery, J., Butler, N., 2015. Scottish science applications of Remote Electronic Monitoring. ICES J. Mar. Sci. 72, 1214–1229. https://doi.org/10.1093/icesjms/fsu225

Pennino, M.G., Paradinas, I., Illian, J.B., Muñoz, F., Bellido, J.M., López-Quílez, A., Conesa, D., 2019. Accounting for preferential sampling in species distribution models. Ecol. Evol. 9, 653–663. https://doi.org/10.1002/ece3.4789

Pinto, C., Travers-Trolet, M., Macdonald, J.I., Rivot, E., Vermard, Y., 2019. Combining multiple data sets to unravel the spatiotemporal dynamics of a data-limited fish stock. Can. J. Fish. Aquat. Sci. 76, 1338–1349. https://doi.org/10.1139/cjfas-2018-0149

Punt, A.E., Walker, T.I., Taylor, B.L., Pribac, F., 2000. Standardization of catch and effort data in a spatially-structured shark fishery. Fish. Res. 45, 129–145. https://doi.org/10.1016/S0165-7836(99)00106-X

Quinn, T., Deriso, R., 1999. Quantitative Fish Dynamics.

R Core Team, 2021. R: A Language and Environment for Statistical Computing.

Rezende, G.A., Rufener, M.C., Ortega, I., Ruas, V.M., Dumont, L.F.C., 2019. Modelling the spatio-temporal bycatch dynamics in an estuarine small-scale shrimp trawl fishery. Fish. Res. 219, 105336. https://doi.org/10.1016/j.fishres.2019.105336

San Martín, M.A., Wiff, R., Saavedra-Nievas, J.C., Cubillos, L.A., Lillo, S., 2013. Relationship between Chilean hake (Merluccius gayi gayi) abundance and environmental conditions in the central-southern zone of Chile. Fish. Res. 143, 89–97. https://doi.org/10.1016/j.fishres.2013.01.010

Stock, B.C., Ward, E.J., Thorson, J.T., Jannot, J.E., Semmens, B.X., 2018. The utility of spatial model-based estimators of unobserved bycatch. https://doi.org/10.1093/icesjms/fsy153

Swartzman, G., Huang, C., Kaluzny, S., 1992. Spatial analysis of Bering Sea groundfish survey data using generalized additive models. Can. J. Fish. Aquat. Sci. 49, 1366–1378. https://doi.org/10.1139/f92-152

Thorson, J.T., 2019. Guidance for decisions using the Vector Autoregressive Spatio-

Temporal (VAST) package in stock, ecosystem, habitat and climate assessments. Fish. Res. 210, 143–161. https://doi.org/10.1016/j.fishres.2018.10.013

Thorson, J.T., Adams, C.F., Brooks, E.N., Eisner, L.B., Kimmel, D.G., Legault, C.M., Rogers, L.A., Yasumiishi, E.M., 2020. Seasonal and interannual variation in spatio-temporal models for index standardization and phenology studies. ICES J. Mar. Sci. 77, 1879–1892. https://doi.org/10.1093/icesjms/fsaa074

Wood, S.N., 2006. Generalized Additive Models: An Introduction with R. https://doi.org/https://doi.org/10.1201/9781315370279

Zhou, S., Campbell, R.A., Hoyle, S.D., 2019. Catch per unit effort standardization using spatio-temporal models for Australia's Eastern Tuna and Billfish Fishery. ICES J. Mar. Sci. https://doi.org/10.1093/icesjms/fsz034

CAPTIONS

Table 1: Model selection, effective degrees of freedom, and AIC value.

Table 2: model estimate, standard error (Std.error), effective degree of freedom (edf), value of Chi square statistic (Chi.sq) and p-values for fixed and smooth terms in model m01.

Figure 1: North Sea study area for the GAM model with locations of hauls between 2011-2015 for observer data (grey triangles) and NS-IBTS survey data (black circles).

Figure 2: Overlap between the spatio-temporal distributions of gears in the North Sea cod data, on a [0-1] scale (left panel). Overlap was approximated using the default parameters in the R package "hypervolume". The panels on the right show examples of calculated spatio-temporal volumes around the data for four pairs of gears, including the IBTS survey ("GOV"). Hypervolumes are materialized by random points within each hypervolume (raw data not

shown). The 3 axes are: scaled latitude (bottom to top, “ShootLat.sc”); scaled longitude (back to front-right, “ShootLon.sc”) and scaled time of year (right to left, “J.day.sc”)

Figure 3: Relative gear selectivity marginal effects according to linear predictor for m01 with 95% confident intervals for: a) GOV; b) ITR; c) LTR; d) MTD; e) MTH; f) MTN; g) MTR; h) NTR; i) PTD; j) SEN and k) showing comparison between different relative gear selectivity functions.

Figure 4: Marginal cod abundance smooth effects from m01 for a) Month; b) Depth and c) marginal cod abundance fixed effect for Year. The covariates are centered, i.e., zero corresponds to the mean of the covariate. Shade around smoother corresponds to 95% confidence bands.

Figure 5. Standardised marginal effects by length class according to the combined length and depth smooth interaction. Values calculated as the sum of marginal effects for terms ti2, ti3 and ti6 in Eq. 2. Marginal predictions for 2015 were standardised to the interval (0-1) by length classes (within each month).

Figure 6: Spatio-temporal model predictions of counts of cod (log scale) for 0-10, 10-20, 20-30, 30-40 and 40-50 cm length groups (columns) during February, May, August, and November (rows). Prediction for 2015 based on counts per 1h trawling effort. Bathymetry contours for -50, -100 and -200 meters in blue scale colors.

Figure 7: model predictions of counts of cod (response scale) with both data sources (X-axis) and only commercial data (Y-axis) for 10, 20, 30, 40 and 50 cm length bins (columns) during February, May, August and November (rows). Range of both axes limited from 0-2 to aid in the visualization of the main cloud of points due to the presence of outliers. Transparency of predictions proportional to its standard error.

Table 1:

Fixed structure	Random intercepts	Df	AIC
ti(x,y,m,L)+ti(D,m,L)+ ti(D,L)+ti(m,L)+s(L, by=Gear)+ti(m)+ti(D)+Year+Gear	-	315.70	833953
ti(x,y,m,L)+ti(D,m,L)+ ti(D,L)+ti(m,L)+s(L, by=Gear)+ti(m)+ti(D)+Year+Gear	Vessel	467.09	244964

Table 2

Fixed terms	Estimate	Std.error	p-value	Smooth terms	edf	Chi.sq	p-value
GOV:2011	-7.527	1.066	$1.36e^{-12}$	ti(x,y,m,L)	142.37	9961.99	$< 2e^{-16}$
ITR	-5.081	3.436	0.139	ti(m,L,D)	41.81	1329.24	$< 2e^{-16}$
LTR	-1.822	1.234	0.140	ti(L,D)	14.59	2828.22	$< 2e^{-16}$
MTD	-2.371	1.246	0.057	ti(m,L)	11.83	1257.71	$< 2e^{-16}$
MTH	-1.848	1.096	0.091	s(L):GOV	8.91	1845.70	$< 2e^{-16}$
MTN	-1.352	1.074	0.208	s(L):ITR	5.48	201.54	$< 2e^{-16}$
MTR	-1.455	1.093	0.183	s(L):LTR	6.34	92.16	$< 2e^{-16}$
NTR	-1.089	1.079	0.313	s(L):MTD	5.38	146.31	$< 2e^{-16}$
PTD	-1.082	1.107	0.329	s(L):MTH	6.89	2813.43	$< 2e^{-16}$
SEN	-2.745	1.453	0.059	s(L):MTN	8.57	2483.16	$< 2e^{-16}$
2012	-0.308	0.042	$1.1e^{-13}$	s(L):MTR	7.28	3037.90	$< 2e^{-16}$
2013	-0.037	0.040	0.358	s(L):NTR	7.15	753.85	$< 2e^{-16}$
2014	0.122	0.040	0.002	s(L):PTD	6.51	2444.64	$< 2e^{-16}$
2015	0.352	0.040	<0.05	s(L):SEN	5.46	660.57	$< 2e^{-16}$
				ti(D)	3.77	138.00	$< 2e^{-16}$
				ti(m)	2.95	337.35	$< 2e^{-16}$
				Vessel (random effect)	152.75	4007.94	$< 2e^{-16}$
Total deviance explained			60.80%				
Shape parameter Theta (θ)			0.179				

Figure 1

Click here to access/download;Figure;Figure 1.tiff

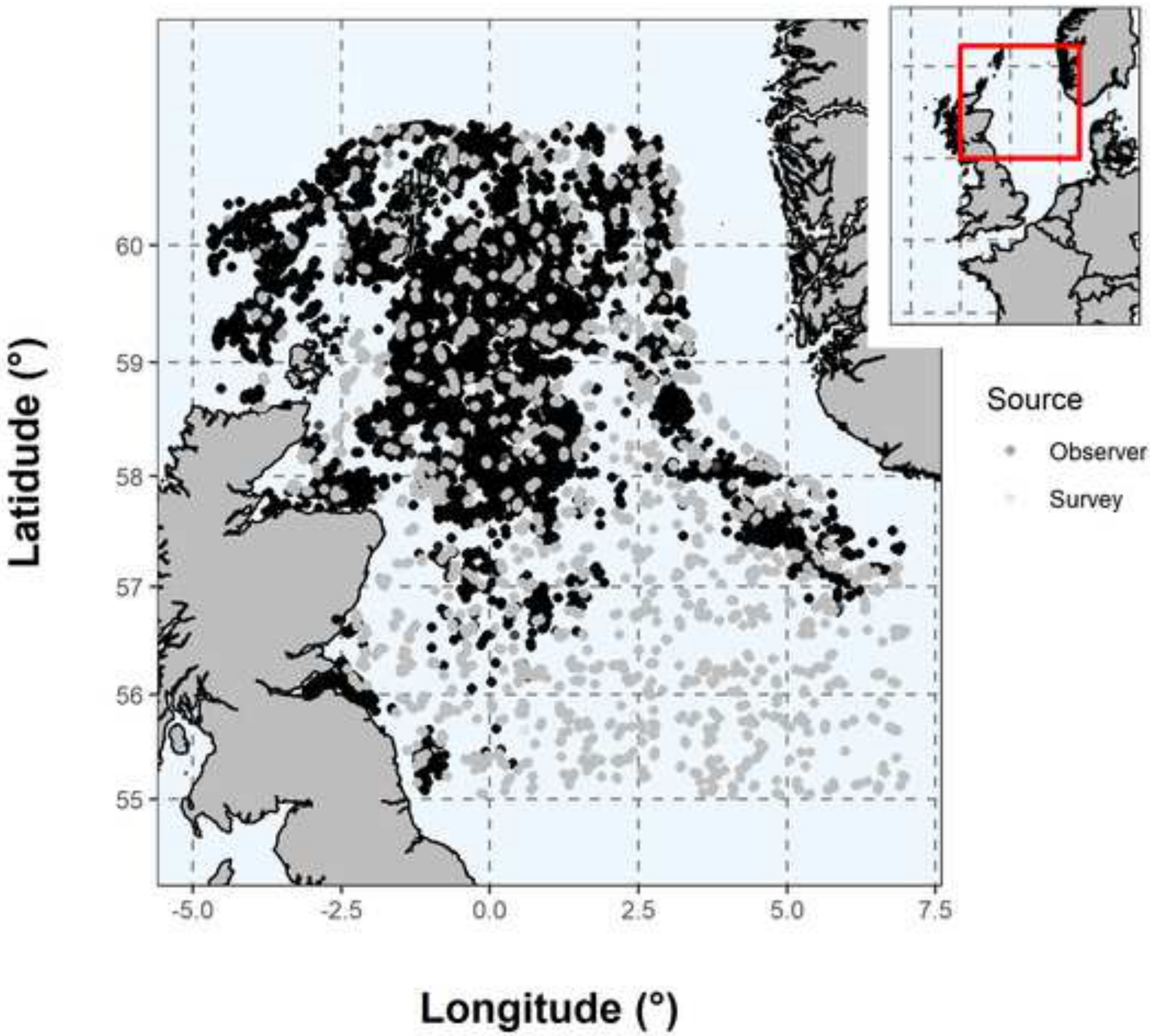


Figure 2

Click here to access/download;Figure;Figure 2extended.png

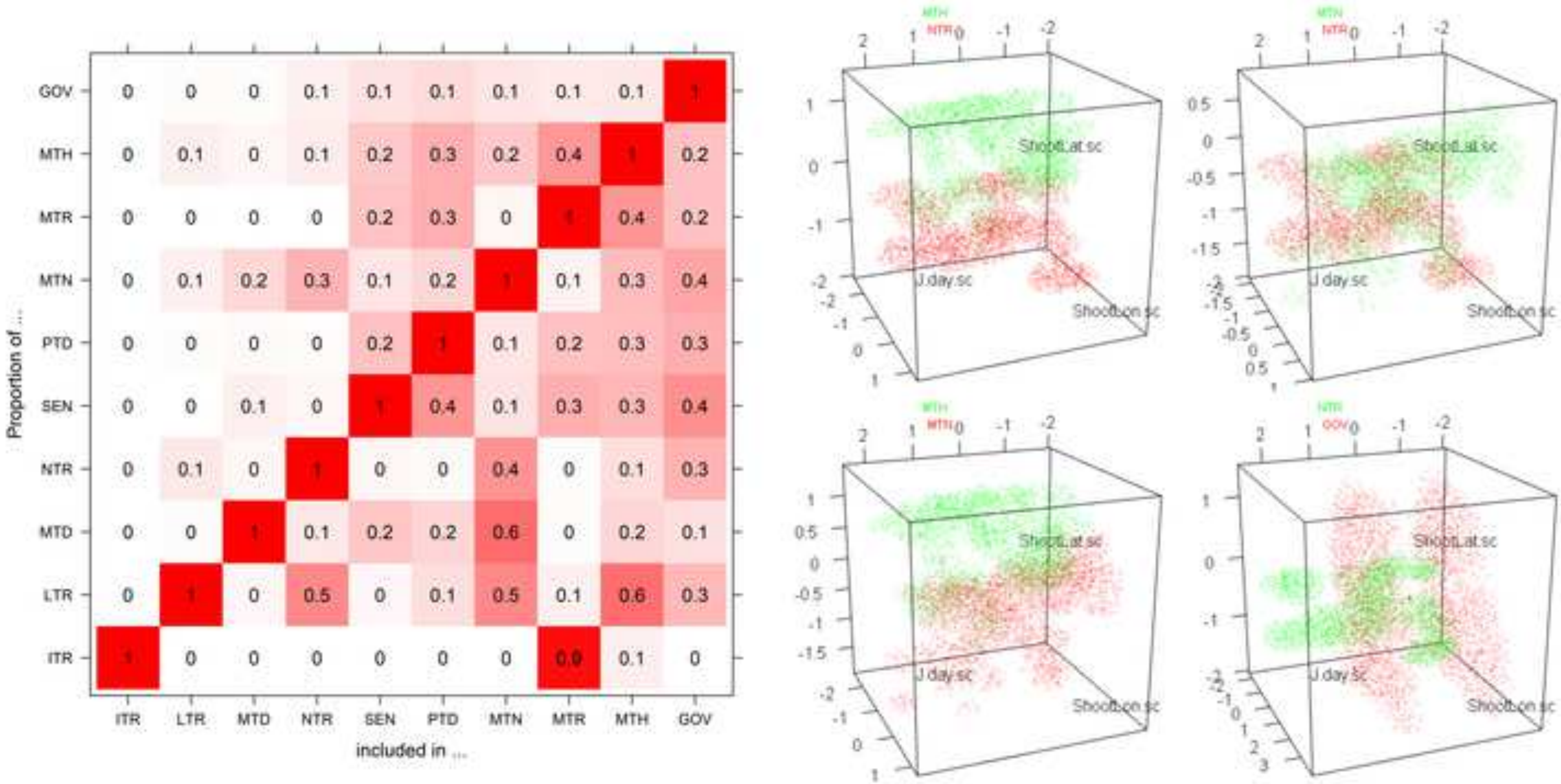


Figure 3

Click here to access/download;Figure;Figure 3.tiff

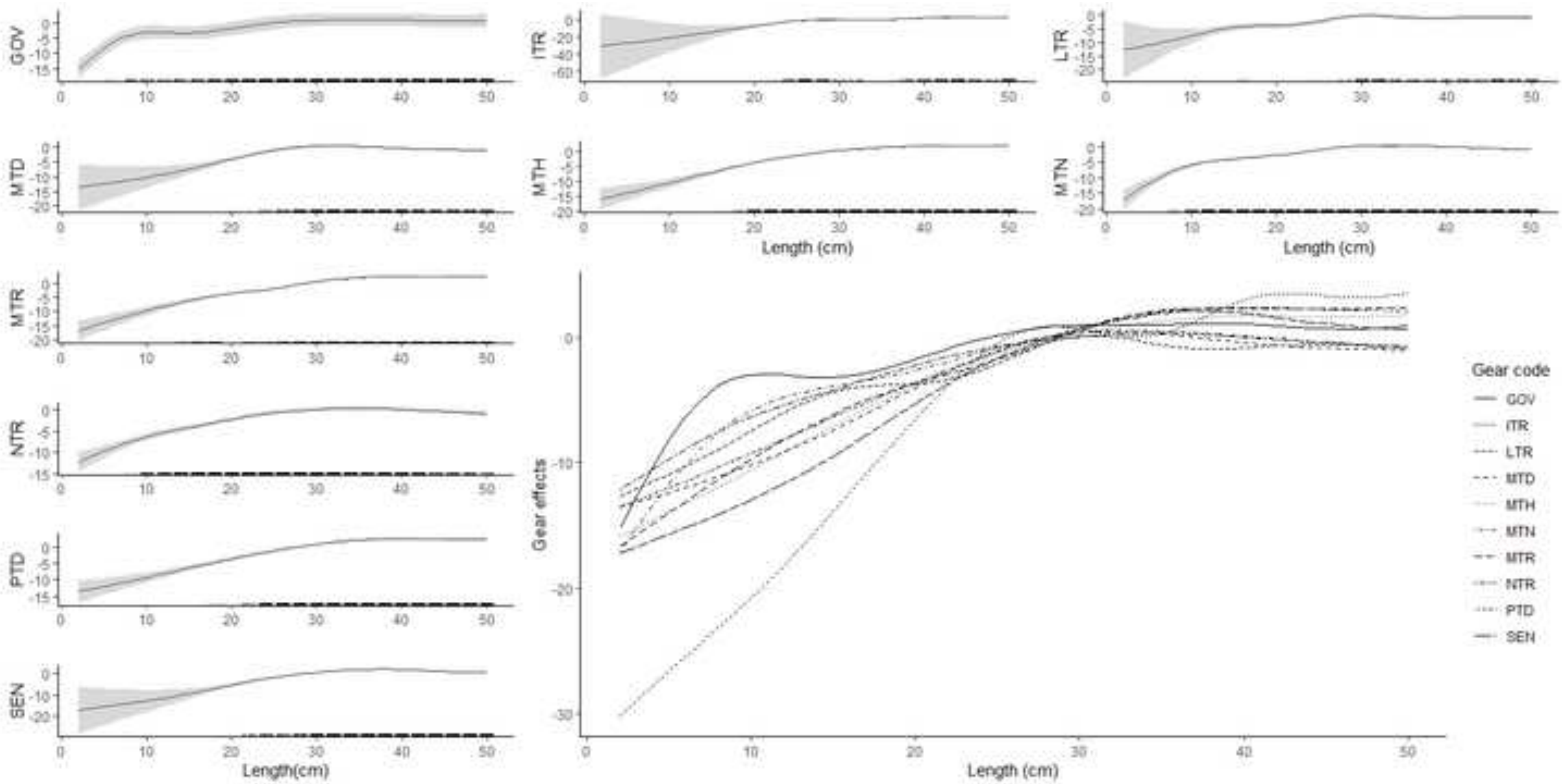


Figure 4

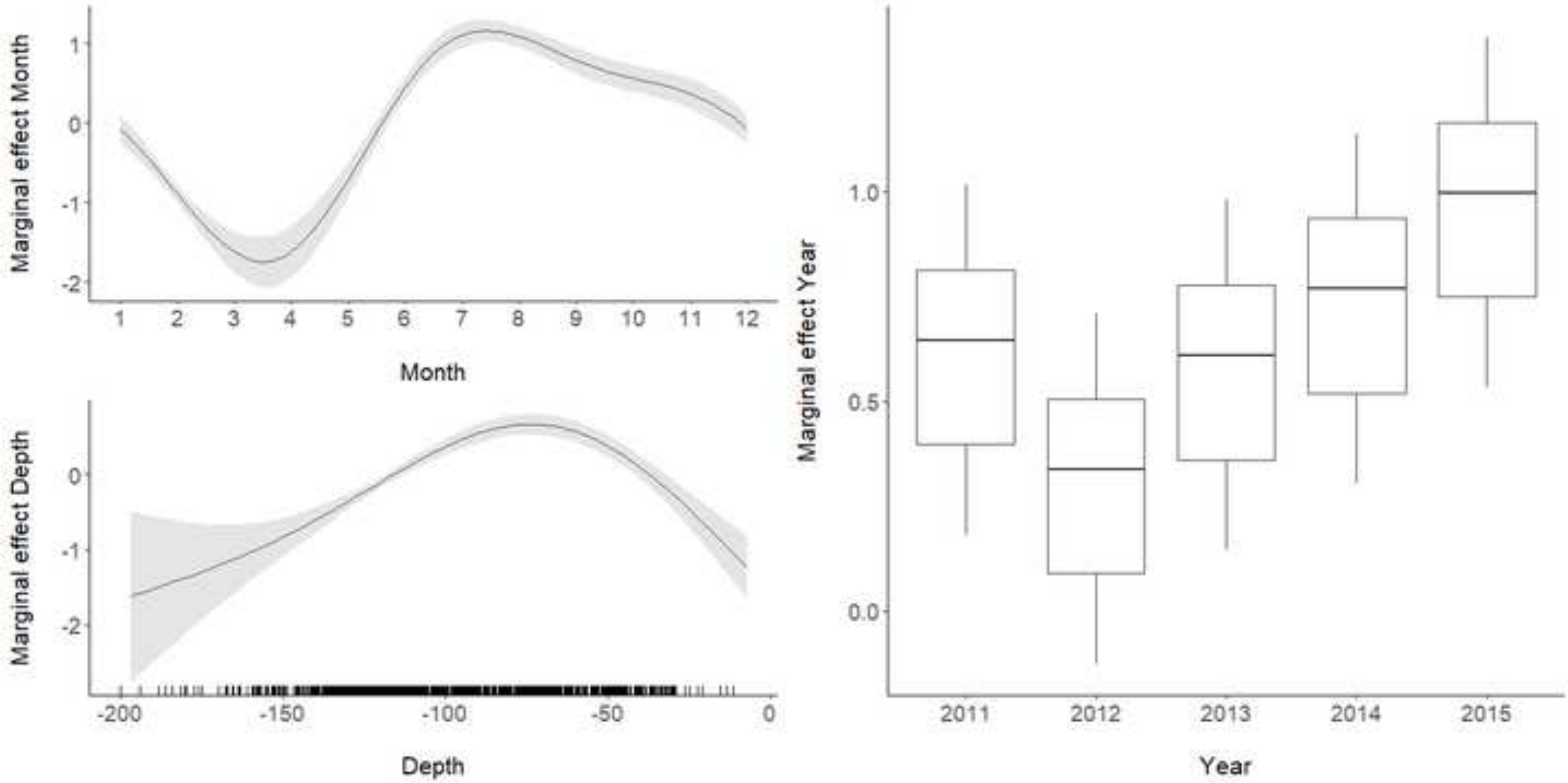

Marginal effect Month
1
0
-1
-2
1 2 3 4 5 6 7 8 9 10 11 12
Month
Marginal effect Depth
0
-1
-2
-200 -150 -100 -50 0
Depth
Marginal effect Year
1.0
0.5
0.0
2011 2012 2013 2014 2015
Year

Figure 5 Click here to access/download;Figure;Figure 5.tiff

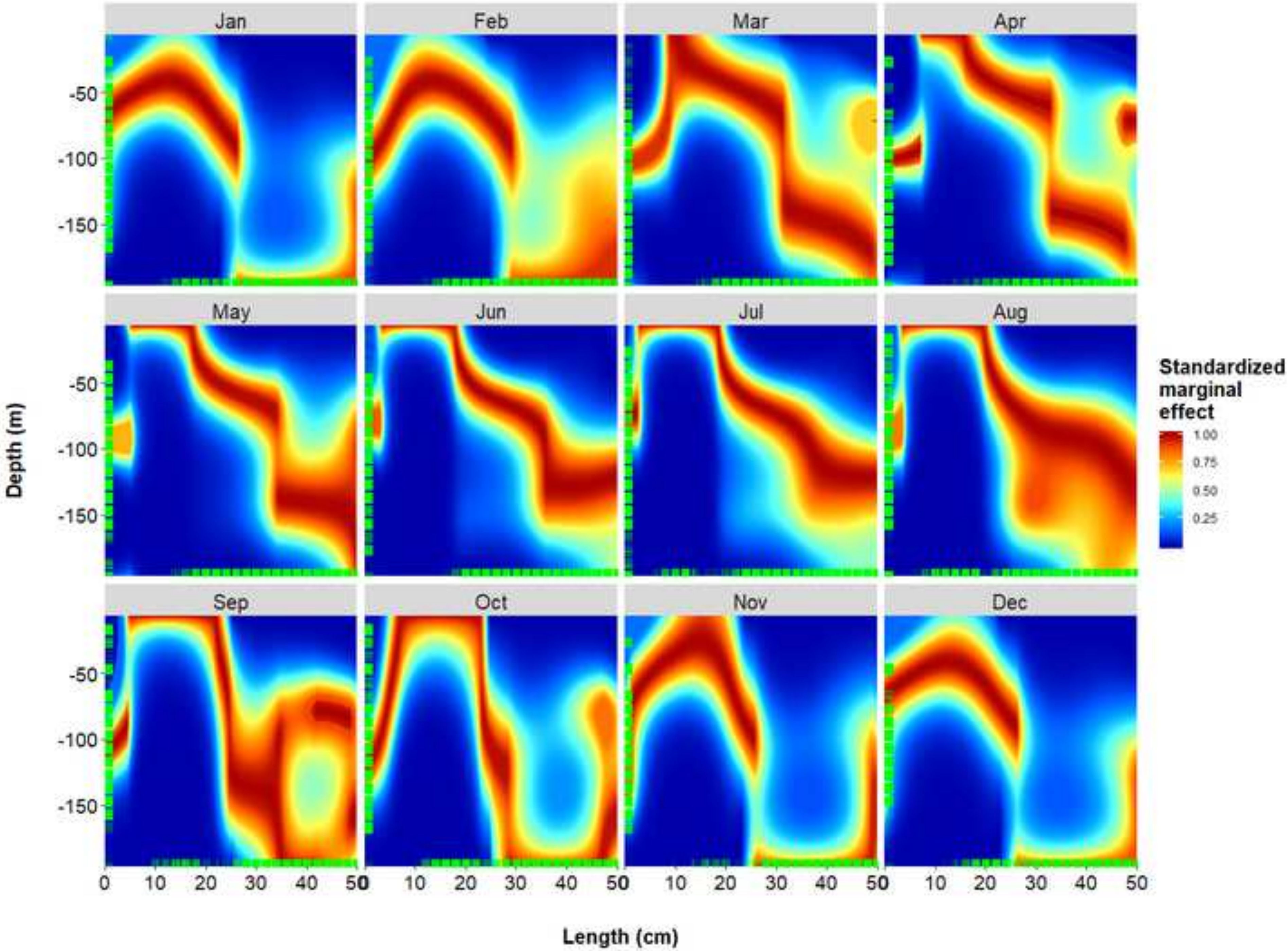


Figure 6

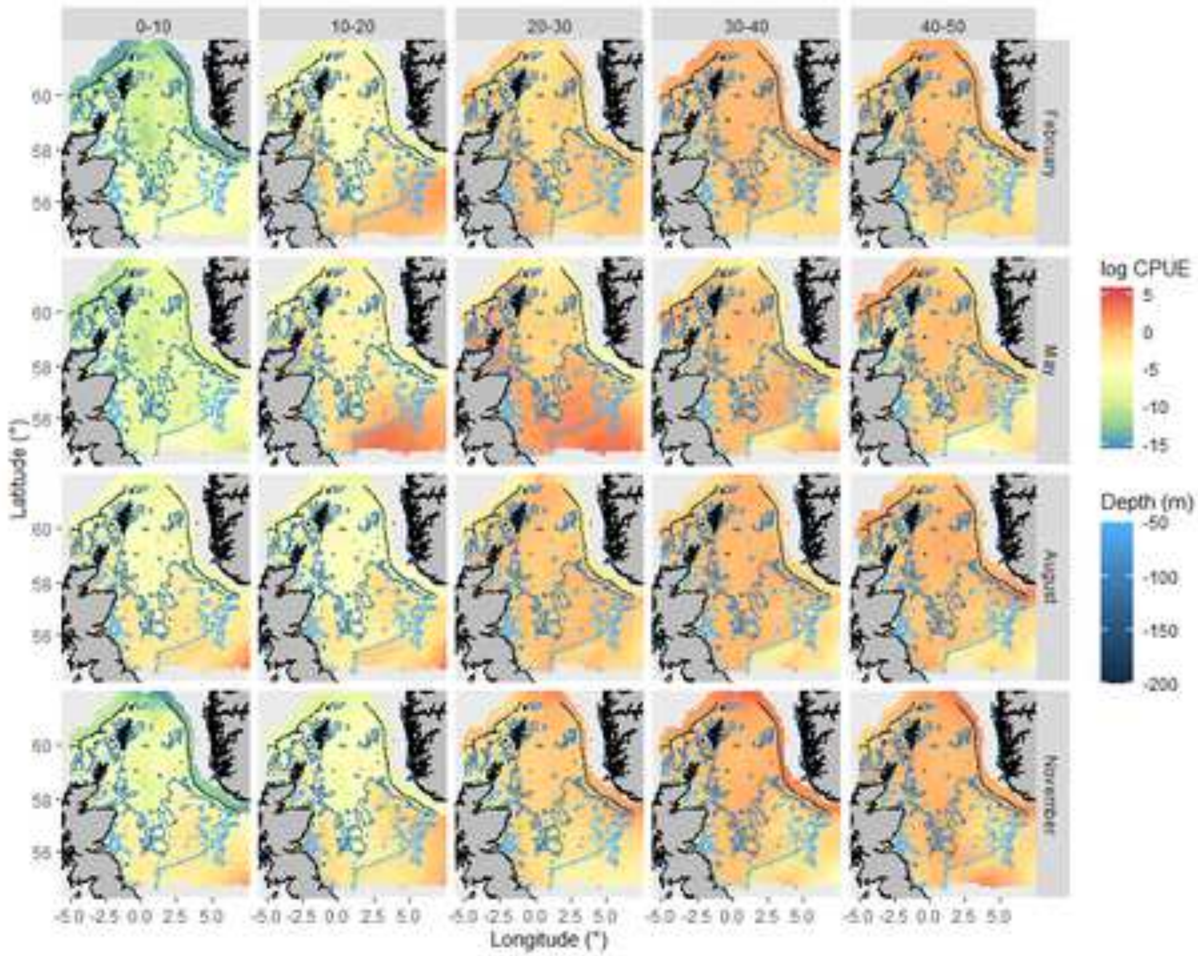
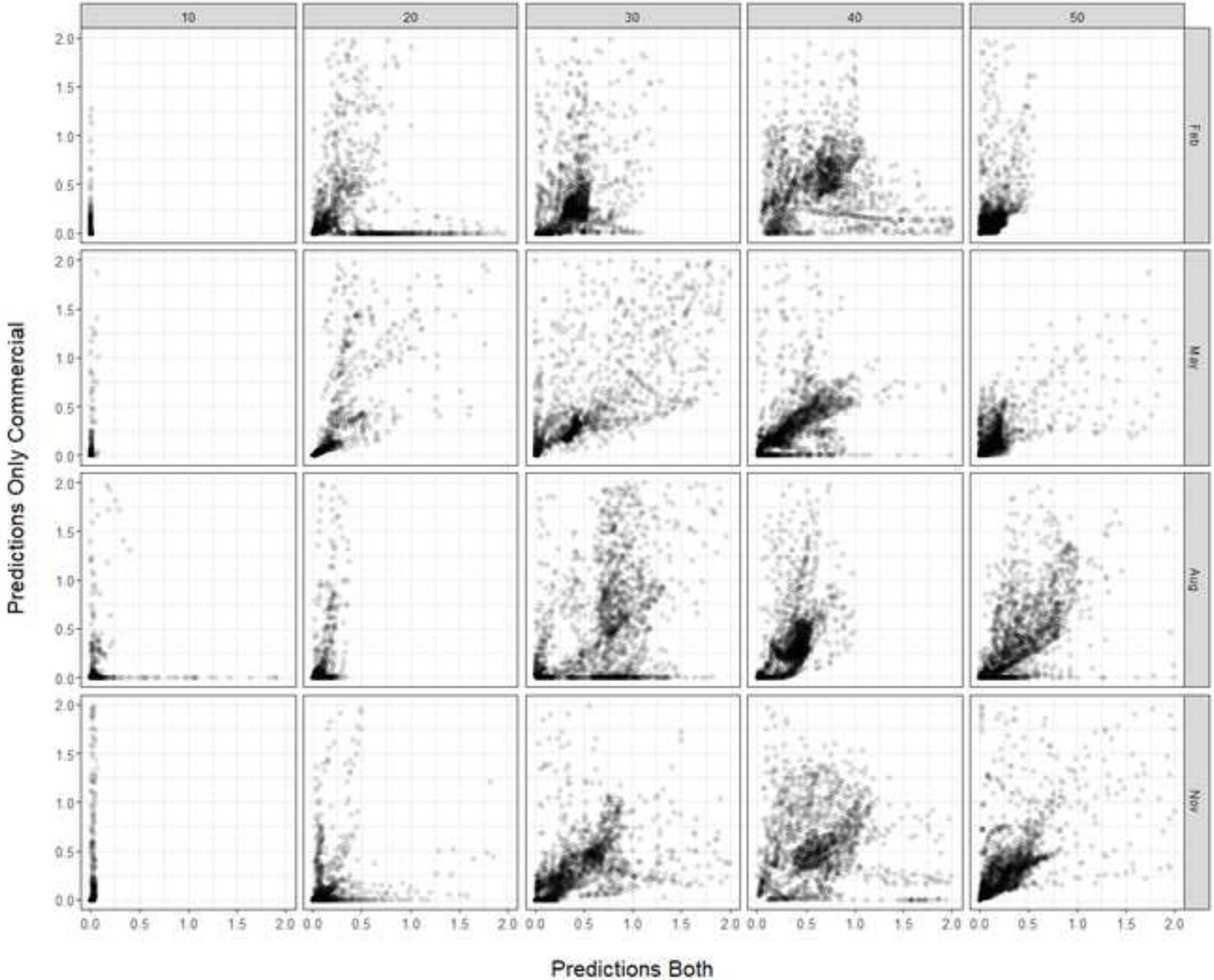

0-10
10-20
20-30
30-40
40-50
February
May
August
November
Latitude (°)
Longitude (°)
log CPUE
Depth (m)

Figure 7


10
20
30
40
50
Feb
May
Aug
Nov
Predictions Only Commercial
Predictions Both
0.0
0.5
1.0
1.5
2.0

Guillermo Martin Gonzalez: concept, coding, writing original draft, visualisation and editing**.** **Rodrigo Wiff:** concept, coding, writing original draft, editing. **C. Tara Marshall:** concept, draft preparation, project administrator, funding acquisition, editing. **Thomas Cornulier:** concept, model design, draft preparation, funding acquisition, editing.