A statistical model for the spatial effort allocation of shrimp fishers in the Dutch coastal area

Author(s): Tony Wilkes

Wageningen University & Research report C041.22

A statistical model for the spatial effort allocation of shrimp fishers in the Dutch coastal area

Author: Tony Wilkes

This research project was carried out by Wageningen Marine Research and subsidized by the ministry of Agriculture, Nature and Food Quality for the purposes of Policy Support Research Theme
'Noordzee natuurbescherming en Herstel' (project no. BO-43-116.01-004).

Wageningen Marine Research
IJmuiden, 22 July 2022

Wageningen Marine Research report C041/22

Keywords: Displacement, MPA, Shrimp, Fishery, North Sea Coastal Zone

Client: Ministry of Agriculture, Nature and Food Quality
Attn.: Mr. J.B.F.Vonk
Bezuidenhoutseweg 73
2594 AC Den Haag

BO-43-116.01-004

This report can be downloaded for free from https://doi.org/10.18174/574072
Wageningen Marine Research provides no printed copies of reports

Wageningen Marine Research is ISO 9001:2015 certified.

Contents

1 Introduction

In 2013 certain areas in the Natura 2000 area North Sea Coastal Zone (NSCZ) were closed for all fishing, with a further addition of areas in 2016 (Fig. 1.1). A preliminary analysis of VMS data (not shown) indicated that it took until 2017 before the closed areas were largely free from fishing activities. Hence, the areas were in practice closed off not since 2013 but only from 2017. These closures raise the question if there has been any displacement of fisheries, and if so, what drives this process and to which areas the effort has been displaced. There is a need to gain a more in-depth understanding of fisheries location choice and ability to predict fisheries displacement when area is closed for nature protection or for example offshore wind purposes.

This study attempts to answer two research questions:

1. Which factors drive where shrimp fishermen decide to fish and where not to fish in the coastal part of the North Sea, before and after fisheries closures?
2. Which factors drive with how much effort (measured in fishing hours) fishermen attempt to fish in their fishing locations before and after some of their fishing locations have been closed (so-called displacement)?

A statistic model was setup to analyse fisheries data and address the research questions.

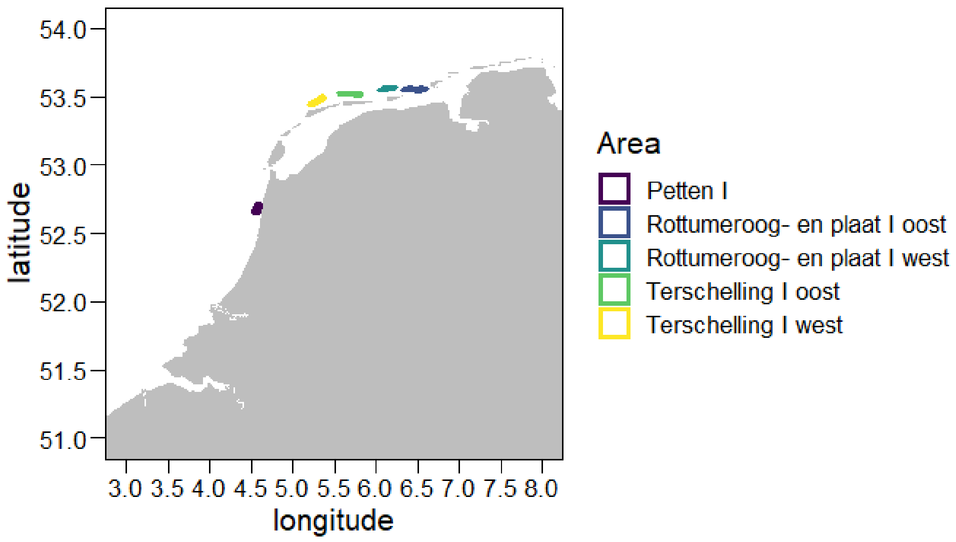


Figure 1.1: Plot of the restricted areas where fishermen were no longer allowed to fish from 2016 onwards.

2 Materials

2.1 Software used

R version 3.6.3 (R Core Team, 2020) and R studio version 1.3.959 (RStudio Team, 2020) were used as the main software for the statistical analyses. The data processing relied in part on the **VMStools** R-package (Hintzen, Bastardie, & Beare, 2017) and all its dependencies. The **INLA** R-package (Bakka et al., 2018; Lindgren & Rue, 2015; Rue, Martino, & Chopin, 2009), henceforth referred to as simply "R-INLA", was used for running the statistical models. Both volumes of the R-INLA book by Zuur *et al.* (2017) were used as a guideline for R-INLA.

2.2 Raw Data

The following datasets were used in the analyses:

- Dutch VMS and logbook data from 2010 to 2020.
- Geotiff maps of the fine and broad bathymetric position index (a measure of relative elevation (Iampietro & Kvitek, 2002; Lundblad et al., 2006; Weiss, 2001)), based on measurements from 2018.
- A geotiff map of the distance to the shore.
- Harbour position data as available in the VMStools package.

2.3 The dataset for modelling

The raw datasets were aligned in space and time and thereafter combined. This dataset contained information on fishing hours, location, time in the year and a number of habitat characteristics as listed in 2.4. Fishing hours were aggregated by year, the location cell on a 0.01 degrees C-squares grid (a grid composed of cells of approximately 1 by 1 km in size; see (Rees, 2003)), and seasons.

A detailed explanation of how the raw data was transformed into a single dataset suitable for statistical modelling, can be found in the Supplementary Materials. In this sub-section, the structure of the dataset suitable for modelling is explained.

The dataset for modelling contains information on location, time, and boat-group.

Location refers to the location of the observation in a certain grid cell on a 0.01 degrees C-squares grid.

Time refers to both years and seasons. The years 2010 to 2020 were available in the dataset. Within a year, season (spring, summer, autumn, winter) is distinguished.

Boat-groups refer to:

- a group of boats that spent at least 1% of their effort trying to catch shrimp in at least one of the restricted areas before 2017.
- a group of boats that never fished any of the restricted areas throughout the study period.

The boat-groups are captured by a binary value referred to as "In Restricted Area" (IRA).

Although the first areas were closed in 2016, only from the beginning of 2017 onwards all restrictions were respected by the fishermen. Therefore, 2017 is being used as the date of closure instead of 2016.

Each observation in the dataset (each row), is a specific combination of a certain location (a single cell on a 0.01 degree c-squares grid), within a certain season, during a certain year, and belonging to a certain boat-group. For those combinations of time and location when no fishing activity was present, fishing effort was set to zero and IRA was set to 0

2.4 List of covariates

It is expected that one or more of the variables listed below drive where shrimp fishermen decide to fish and where not to fish:

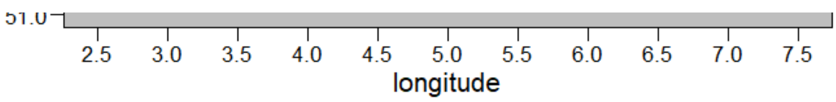
- DS: Distance to Shore, in km.
- DNH: Distance to Nearest Harbour, in km.
- DNRA: Distance to Nearest Restricted Area, in km.
- fine bpi: The fine bathymetric position index (inner radius: 10 cells, or 1 km; outer radius: 20 cells, or 2 km; following Walbridge et al 2018).
- broad bpi: the broad bathymetric position index (inner radius: 5, or 0.5 km; outer radius: 10 cells, or 1 km; following Walbridge et al 2018).
- seasons: the seasons (winter, spring, summer, autumn; winter is reference level).
- R periods: One can distinguish 4 periods in our study. Period 1 were the years when no areas in the NSCZ were closed for shrimp fisheries (years < 2013), period 2 were the years when some areas were closed but these closures were not widely adhered to (years 2013-2015), period 3 was the year when all areas were closed, and this closure was to a certain degree adhered to (2016) and period 4, where all areas were closed, and data indicated no fishing activity in the closed areas either. Period 1 was the reference level.
- IRA: binary variable, indicating the boat-group. For boats that spent at least 1% of their fishing effort in one of the restricted areas before R-period 4, IRA=1. Otherwise, IRA=0.

2.5 Data exploration and pre-model decisions

Data exploration included creating correlograms, pairs plots, distribution plots, summary tables, and covariate-vs-response plots. These are documented in the Supplementary Materials.

Covariates excluded

Bathymetric Position Index (BPI) is a measure of relative depth. There were two BPI covariates available in the data: broad bpi, and fine bpi, and these two were highly correlated, and hence only one of these could be used. Data plots showed that the relation between the fishing hours and broad bpi was more erratic than the relation between fishing hours and fine bpi. Preliminary model runs did not indicate a clear preference for either of these two indices and therefore fine bpi was chosen.



ed to

Figure 2.1: Map with the study area highlighted in yellow.

3 Methods

3.1 Introduction

This section details the statistical method applied for this study. For information on how the model diagnostics were assessed, see appendix 3. A statistical model consisting of two separate parts was constructed (a hurdle model). The first part is a model that estimates what factors drive the presence or absence of fishing activity and the second part estimates what factors drive the intensity of fishing activity (non-zero model). This two-model setup was necessary as the fishing hours data showed signs of zero-inflation which is difficult to fit in a single model.

3.2 Details on the Hurdle model set

To correct for spatial autocorrelation, spatial random effects were included in both the presence/absence model and non-zero models of the hurdle set. The spatial autocorrelation was modelled using a Matèrn correlation function. Constructing the Matèrn correlation function involves making a triangles-based mesh of the area, where each corner of the triangle is a vertex. Each vertex gets its own random effect coefficient. The minimum distance between any 2 vertices was set to 3 km (approximately the size of 3 adjacent 0.01 degrees C-squares grid cells; see (Rees, 2003)), the maximum distance was set to 50km (approximately the size of 50 adjacent 0.01 degrees C-squares grid cells), and the smallest allowed angle in the triangles was 21 degrees. This resulted in a mesh with 1396 vertices. The mesh was used as input to estimate spatial correlation in continuous space (Lindgren, Rue, & Lindström, 2011). The spatial auto-correlation was suspected to be different in different seasons (see maps of the spatial distribution of observed fishing hours in the Supplementary materials). Therefore, the spatial correlation was allowed to vary over the seasons. The spatial random effects were modelled uncorrelated between the season (seasons are not modelled as auto-regressive). Note that the model was limited to the spatial distribution of the study area, but not limited in the study period, as predictions must be made possible for future years. Therefore, no yearly temporal component was added to the spatial random effects.

A Bernoulli Generalized Linear Mixed Model (GLMM) with a logit link function was used as the presence/absence model. For the non-zero models, a Gamma GLMM with a log-link function was used.

As stated in the introduction section, the model should also be able to make predictions outside of the study period. Each year has a certain number of fishing hours in total over all locations and seasons, and years with more total fishing hours should naturally produce generally higher predictions then years with lower total fishing hours. Therefore, the actual number of fishing hours were not used as the response in the non-zero models; instead the response variable was the number of fishing hours as a proportion of the total number of fishing hours in a year divided by 8. The division by 8 is done because each year has 4 seasons and 2 boat-groups. Due to the use of a Gamma GLMM, this proportion was turned into a strictly positive ratio. See Appendix 4 for details.

3.3 Covariate specifications

Some covariates were entered in the models as non-linear, when applicable. When using non-linear terms, special care was taken to prevent overfitting (as the model had to be used for future predictions), perfect separation, and incorrect extrapolation (as the non-zero model was only fitted on non-zero data). Moreover, choice in types of splines used for the non-linear terms was limited. The boat-group variable, IRA, is only applicable when there is actual fishing activity, not when there is an absence of fishing activity. Therefore, the IRA covariate was not present in the presence/absence model. The formulations given here are of the full models; no model covariate selection was performed in this analysis. For a more detailed explanation of the reasoning in the choices made, see Appendix 2.

The linear predictors in the presence/absence model, ignoring all standardizations, transformations, and the coefficients, can be formulated as follows:

$\eta \sim$ Intercept + season + (R periods) + DNRA +
(R period three) × DNRA +
(R period four) × DNRA +
f_{pwls}(DS, knots=(2,4,7)) + (finebpi pos) + (finebpi neg) +
f_{pwls}(DS, knot=5) × (finebpi pos) +
f_{pwls}(DS, knot=5) × (finebpi neg) +
DNH + (year total fishing hours) +
Random effects

where $f_{\text{pwls}}()$ refers to the piece-wise linear spline function.

The linear predictors in the non-zero model, ignoring all coefficients, standardizations, transformations, and the offset term, can be formulated as follows:

$\eta \sim$ Intercept + season + (R periods) + IRA + DNRA +
IRA × DNRA + IRA × (R period three) + IRA × (R period four) +
(R period three) × DNRA + (R period four) × DNRA +
IRA × (R period three) × DNRA +
IRA × (R period four) × DNRA +
DS + poly(finebpi pos, degree=2) + poly(finebpi neg, degree=2) +
DS × poly(finebpi pos, degree=2) + DS × poly(finebpi neg, degree=2) +
DNH + DNH × (year total fishing hours) +
Random effects

where poly() refers to a regular polynomial.

3.4 Post-model adjustment

Model evaluation was done based on both specificity (true positives) and sensitivity (true negatives). In the presence/absence model, due to the strong zero inflation, the specificity was very high, but the sensitivity low. This was fixed by adding a small constant to the linear predictors after the model was fitted. This constant was chosen such that the sensitivity and specificity were maximized, and the 0/1-loss minimized, leading to this constant value added being 0.45.

During R-period 4, no boats were in any of the restricted area. Therefore, when making predictions for a new dataset, the predicted value is necessarily set to zero whenever both R-period=4 and DNRA=0.

4 Results

First, the results from the presence/absence model are shown (Fig. 4.1, Fig. 4.2, Table 4.1), and then the results of the non-zero model (Fig. 4.3, Fig. 4.4, Table 4.2), and finally the results of the Hurdle as a whole are shown (Fig. 4.5 and Fig. 4.6). For the presence/absence model and the non-zero model results, results of the interaction terms and the main effect terms are shown.

For all model diagnostic plots, the map visualizing the spatial distribution of the predicted and observed fishing hours, and a table of the fit of the hurdle (in terms of the fit-slope, specificity, sensitivity, 0/1-loss, and mean absolute deviation), the reader is referred to the Supplementary Materials.

The Hurdle fit reasonably well. Only the specificity for the IRA-boat group is rather low, while the specificity for the non-IRA group is high. This means that the hurdle does not always properly predict an absence of fishing activity for the IRA boats while it predicts rather well for the non-IRA boats.

The model diagnostic plots for the presence/absence model showed no violations of the model assumptions. For the non-zero model, some reduction of the residual variance as the predicted values increase (a light form of heteroskedasticity) was found, though this was considered to be of minor impact.

4.1 Presence/absence model

The following patterns were observed in the results:

- As the distance to the nearest restricted area increases, the probability of fishing activity decreases across all R-periods, though the significance of this effect reduces as the R-periods advances (Fig. 4.1).
- A bpi of zero has the highest probability of fishing activity across all distances from the shore (though around 6km from the shore the probability of fishing activity is a straight line, decreasing as the bpi increases).
- The probability of fishing activity increases in the first 2 km, and then is somewhat stable and flat until around 7km; at >7km the probability of fishing activity decreases as the distance to shore increases (Fig. 4.1).
- The second R-period has the highest probability of fishing activity (Fig. 4.2).
- The probability of fishing hour decreases as the distance from the nearest harbour increases (Fig. 4.2).
- The plots show the interaction effects. The main effects are also available (Table 4.1).

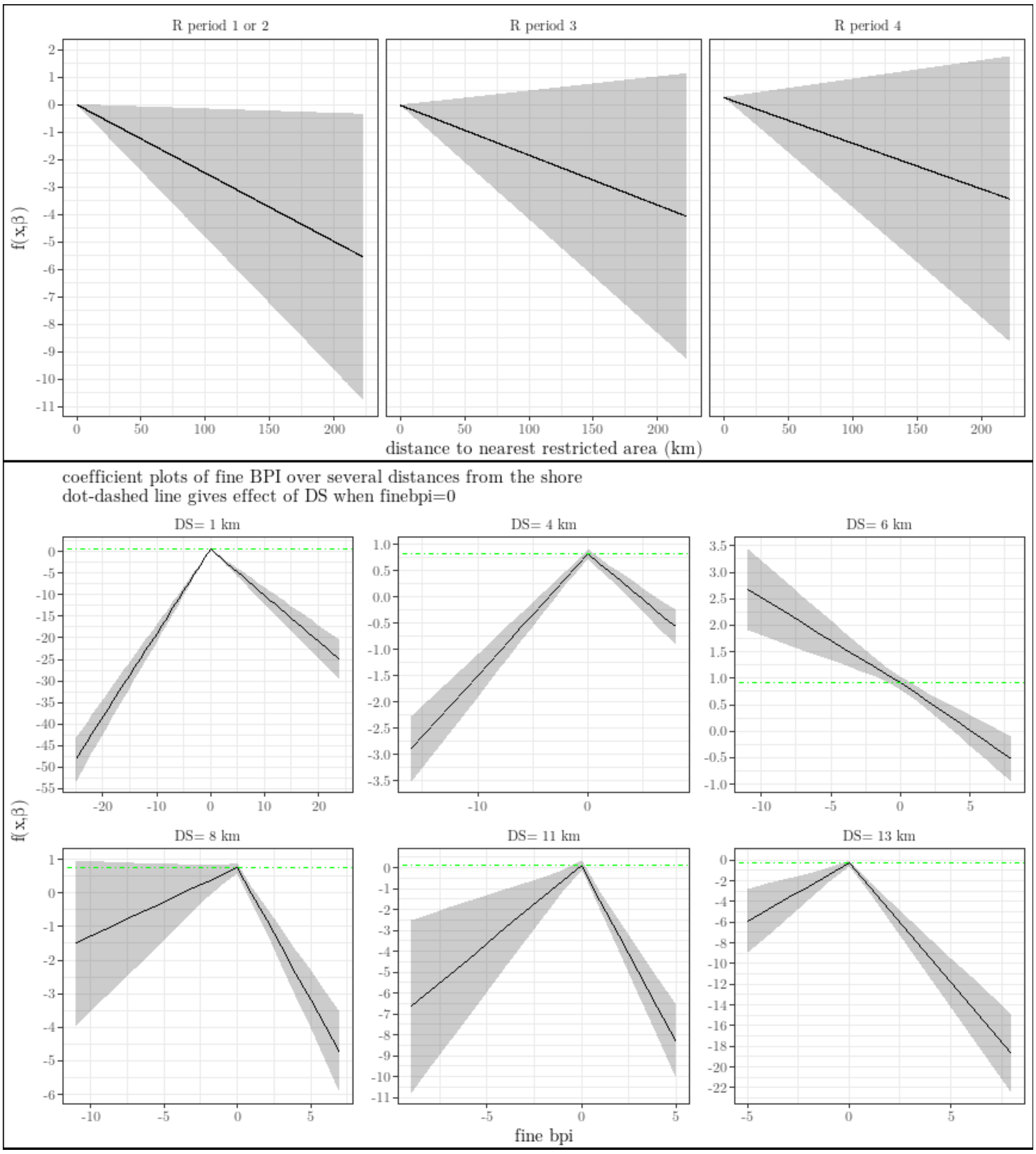


Figure 4.1: Coefficient plots of the interaction terms, of the presence/absence model. The plots were made on the linear scale. The y-axis represents the relative preference for the covariate value. Higher y-values indicate more preference, lower indicate less preference. The ribbon (the shaded band) gives the 95 percent credible interval. For the full quantitative interpretation, see Appendix 1.

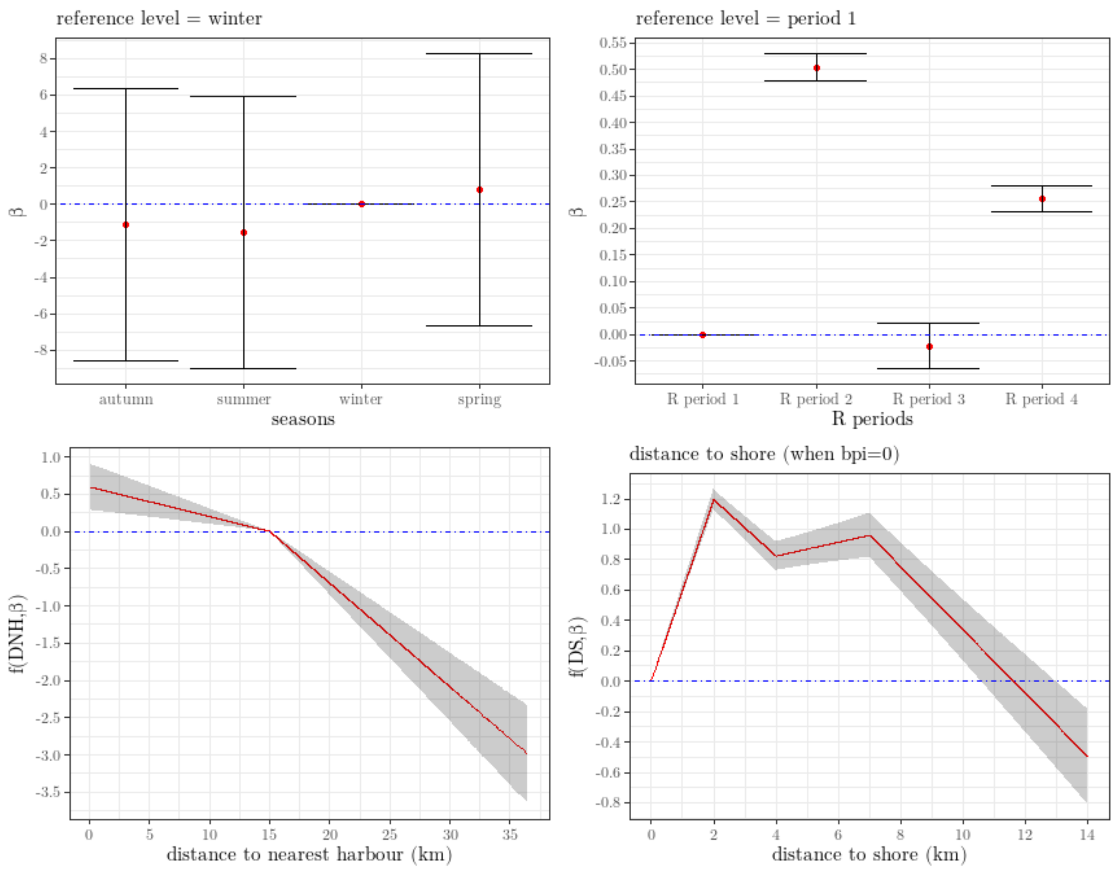


Figure 4.2: Coefficient plots of the non-linear main effects, of the presence/absence model. The plots were made on the linear scale. The y-axis represents the relative preference for the covariate value. Higher y-values indicate more preference, lower indicate less preference. The ribbon (the shaded band) gives the 95 percent credible interval. For the full quantitative interpretation, see Appendix 1.

Table 4.1: Fixed effects coefficient summary table of the linear main effects of the presence/absence model for. The column 'relevant' indicates if a variable is important, based on the 95 percent credible interval. Column 'st.dev' gives the standard deviation of $\boldsymbol{\beta}$, and 'exp(cred.int)' gives the exponent of its 95 percent credible interval. The interpretation of a covariate coefficient $\boldsymbol{\beta}$ is as follows: Consider a covariate $\boldsymbol{x_1}$ and its corresponding coefficient $\boldsymbol{\beta_1}$. Suppose $\boldsymbol{x_1}$ increases by a single unit (see column 'unit'), and all other variables do not change. Then the odds of fishing activity will be multiplied by exp($\boldsymbol{\beta_1}$).

term	**unit**	$\boldsymbol{\beta}$	**st.dev**	**exp($\boldsymbol{\beta}$)**	**exp(cred.int)**	**relevant**
Intercept	NA	-3.389	2.817	0.03373	0.0001336, 8.472	No
DNRA	100 km	-2.489	1.191	0.08298	0.008003, 0.8584	YES
fhr year total	15000 hours	0.341	0.006669	1.406	1.388, 1.425	YES

4.2 Non-zero model

The following patterns were observed in the results:

- As the distance to the nearest restricted area (DNRA) increases, the amount of expected effort decreases. This effect of DNRA is strong among the IRA boats, but very weak among the non-IRA boats. This effect of DNRA is similar across all R-periods, though in R-period 4 the effect is slightly reduced for both boat groups (Fig. 4.3).
- As the distance to nearest harbour increases, the amount of expected effort also decreases, and this effect becomes stronger as the yearly total fishing hours increases (Fig. 4.3).
- Generally speaking, areas with a negative bpi have a higher expected effort than areas with positive bpi, though areas with a strongly negative bpi have less expected effort. This effect is consistent as one moves away from the shore (Fig. 4.3). Note, however, that as the distance from the shore increases, the range of bpi in the areas also decrease somewhat. Distance to shore itself has a negative effect on the fishing effort: the larger the distance, the lower the expected effort.
- The first R-period has the highest expected effort (Fig. 4.4).
- The plots show the interaction effects. The main effects are also available (Table 4.2).

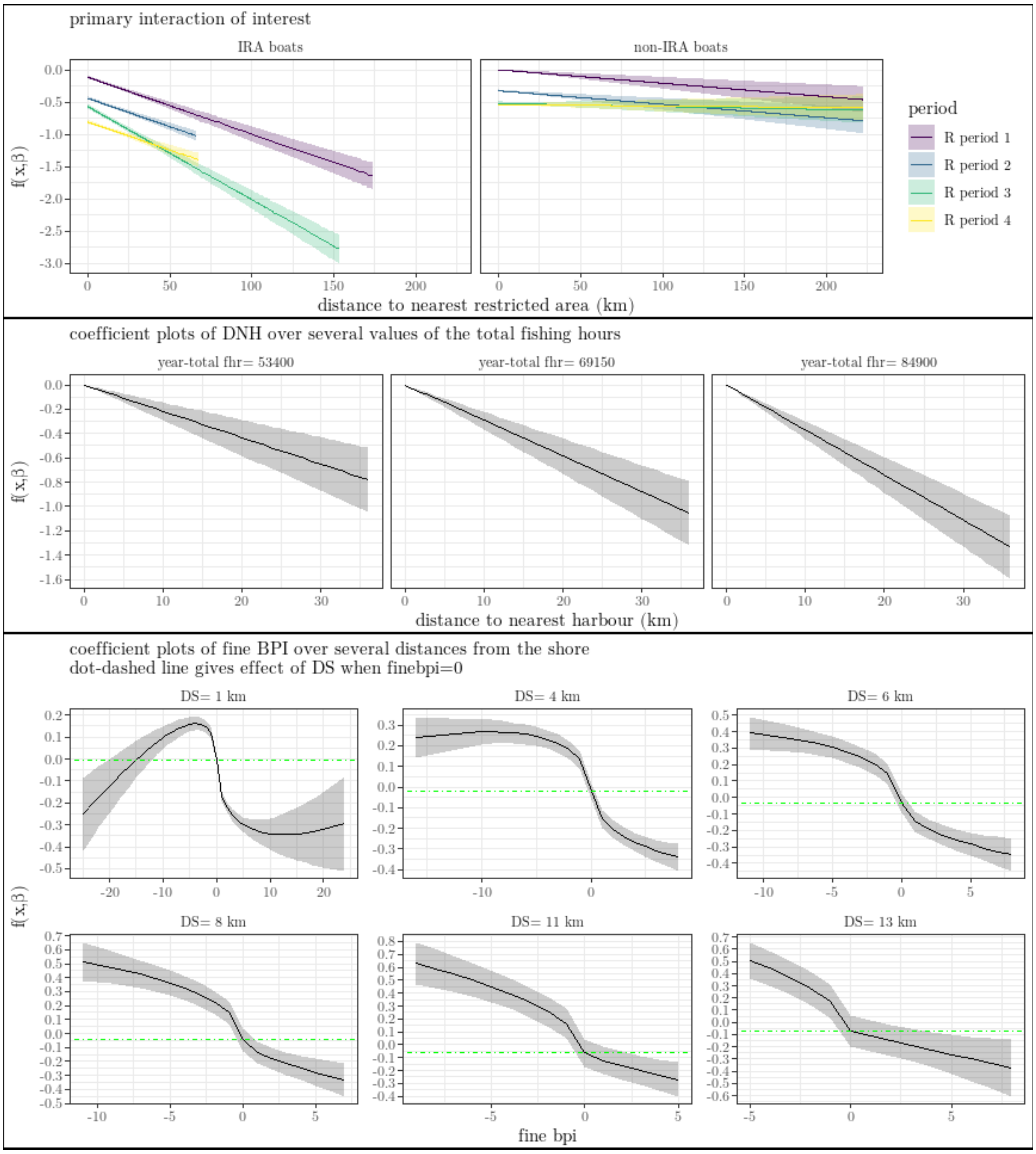


Figure 4.3: Coefficient plots of the interaction terms, of the non-zero model. The plots were made on the linear scale. The y-axis represents the relative preference for the covariate value. Higher y-values indicate more preference, lower indicate less preference. The ribbon (the shaded band) gives the 95 percent credible interval. For the full quantitative interpretation, see Appendix 1.

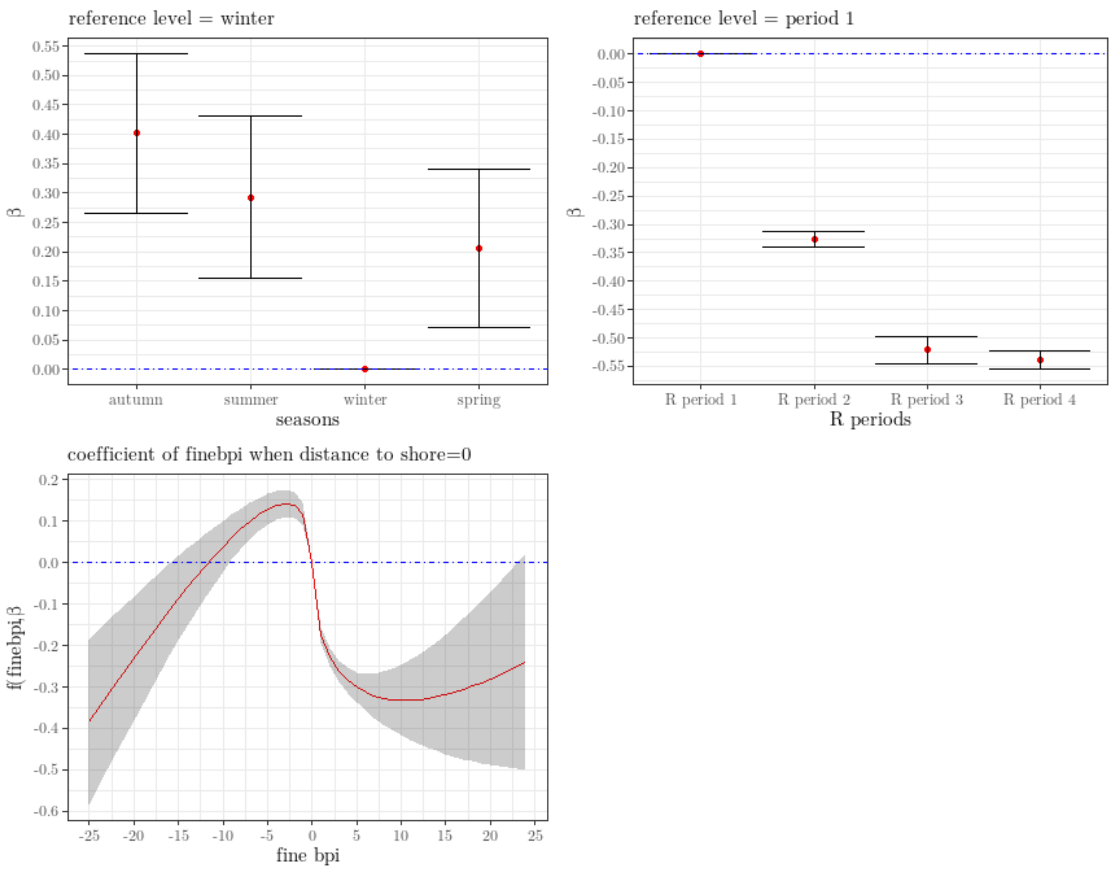


Figure 4.4: Coefficient plots of the non-linear main effects, of the non-zero model. The plots were made on the linear scale. The y-axis represents the relative preference for the covariate value. Higher y-values indicate more preference, lower indicate less preference. The ribbon (the shaded band) gives the 95 percent credible interval. For the full quantitative interpretation, see Appendix 1.

Table 4.2: Fixed effects coefficient summary table of the linear main effects of the presence/absence model for. The column 'relevant' indicates if a variable is important, based on the 95 percent credible interval. Column 'st.dev' gives the standard deviation of $\boldsymbol{\beta}$, and 'exp(cred.int)' gives the exponent of its 95 percent credible interval. The interpretation of a covariate coefficient $\boldsymbol{\beta}$ is as follows. Consider a covariate x_1 and its corresponding coefficient $\boldsymbol{\beta}_1$. Suppose x_1 increases by a single unit (see column 'unit'), and all other variables do not change. Then the number of fishing hours (relative to the yearly total fishing hours) will be multiplied by $\exp(\boldsymbol{\beta}_1)$. Note that the offset term is shown in this table as well, and should be very close to 1. When computing fitted model values, the offset coefficient is set to be exactly 1. For the meaning of the offset term, the reader is referred to Appendix 4.

term	unit	$\boldsymbol{\beta}$	st.dev	$\exp(\boldsymbol{\beta})$	exp(cred.int)	relevant
Intercept	NA	-7.401	0.08421	0.0006109	0.0005178, 0.0007207	YES
DNRA	100 km	-0.2082	0.04608	0.812	0.7418, 0.8889	YES
DNH	1 km	-0.02976	0.00372	0.9707	0.9636, 0.9778	YES
DS	1 km	-0.005725	0.004972	0.9943	0.9846, 1.004	No
offset term	NA	0.9814	0.00464	2.668	2.644, 2.692	offset term

4.3 Hurdle model set as a whole

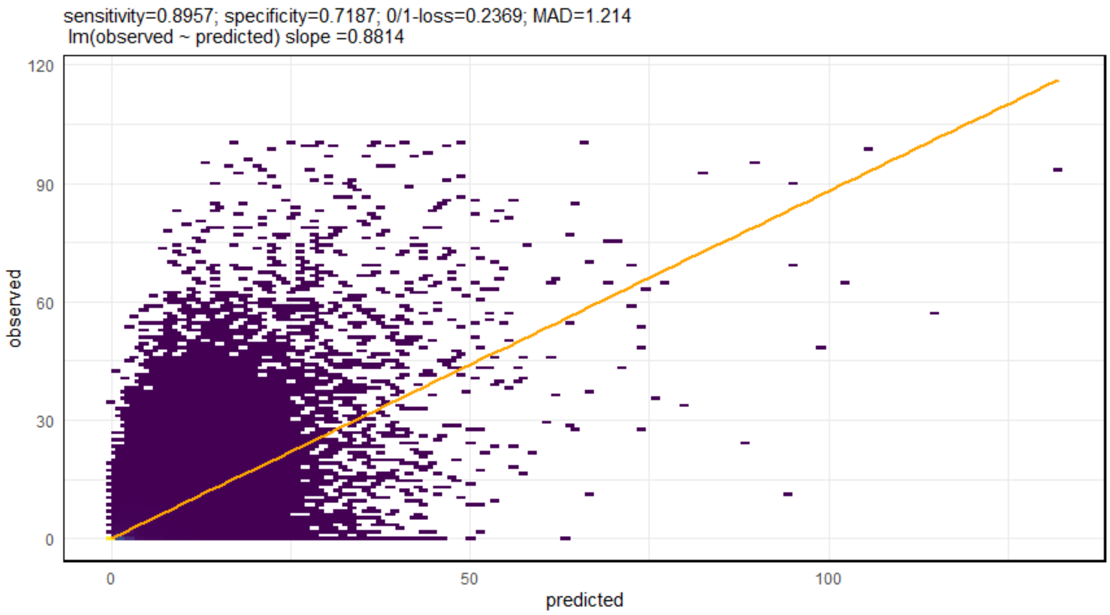


Figure 4.5: Observed versus predicted fishing hours of the whole hurdle. The closer the sensitivity and specificity are to 1, the better. The close the 0/1 loss is to zero, the better. The lower the mean absolute deviation (MAD), the better. The closer the fit slope (slope of the fitted yellow straight line) is to 1, the better. (Due to the large number of data points, the points on the plot were binned.)

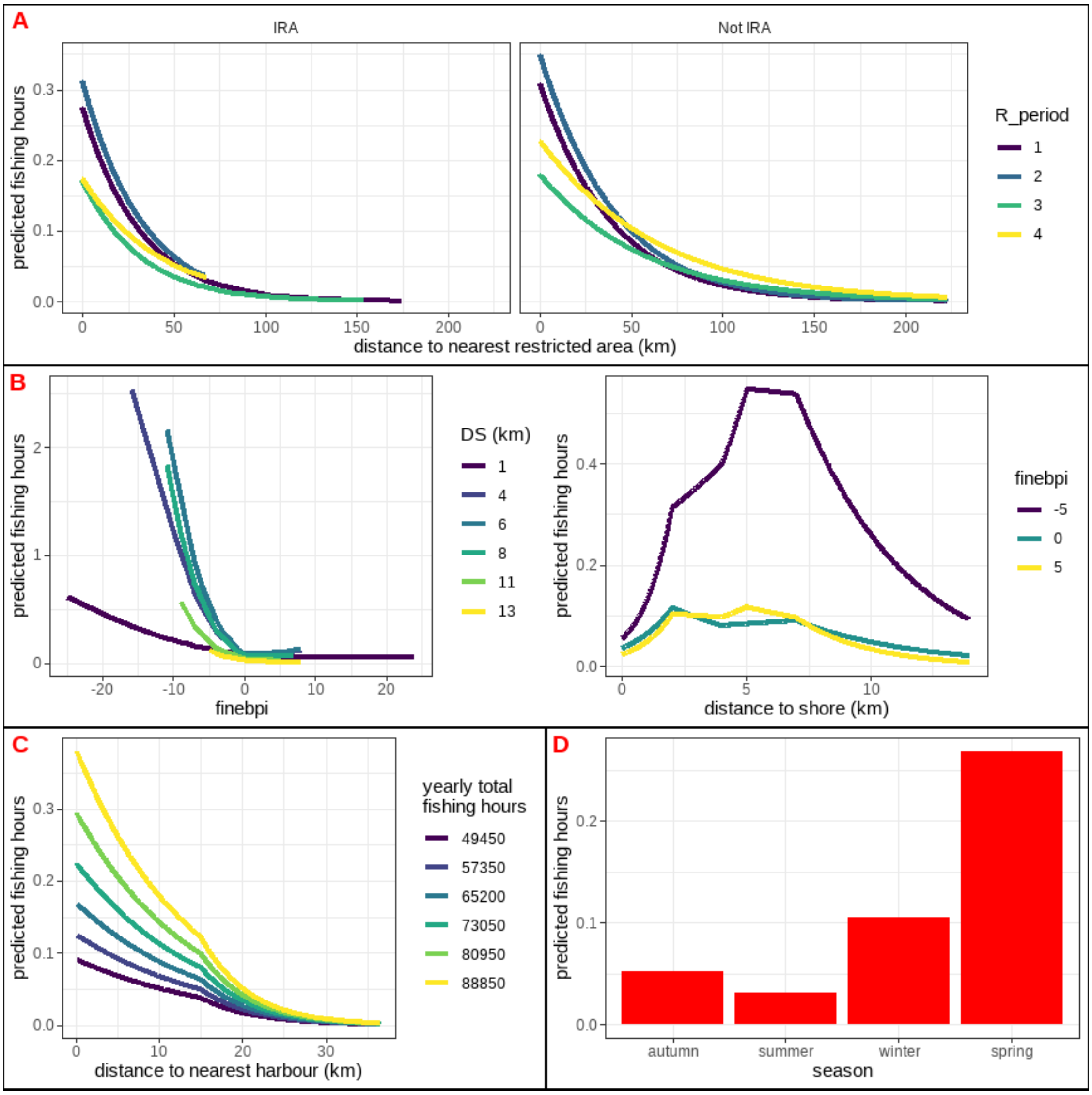


Figure 4.6: Predictor-effect plots of the whole hurdle model set. In all predictor effect plots, the random effects are kept at zero. Note that, due to the covariates going through 2 non-linear link functions, the relation between the varying covariates and predictions might change a bit if different values are chosen for the covariates that remain fixed. There are 4 sets of predictor effect plots shown here. Set A shows the effect of the combinations of IRA, the R-periods, and the distance to the nearest restricted area, on the fishing hours predicted by the hurdle as a whole, while all other covariates remain fixed at their mean values. Set B consists of 2 plots that both show the effect of the combination of fine bpi and the distance to the shore on the fishing hours predicted by the hurdle as a whole, while all other covariates remain fixed at their mean values. Plot C shows the effect of the combinations of the yearly total fishing hours and the distance to the nearest harbour on the fishing hours predicted by the hurdle as a whole, while all other covariates remain fixed at their mean values. Plot D shows the effect of the seasons on the fishing hours predicted by the hurdle as a whole, while all other covariates remain fixed at their mean values.

5 Discussion and conclusions

For the probability of a fishing activity, the preferred distance to the nearest restricted area does not change much between the restriction periods, and fishers prefer to fish nearby the restricted areas, even after their closure. For the amount of fishing activity, fishers tend to fish with more effort when nearby the restricted areas after their closure, especially if those fishers used to fish in the restricted areas before the closure. In Restriction periods 3 and 4 there was generally less fishing effort compared to the earlier periods.

As the distance to the nearest harbour increases, both the probability and amount of fishing activity decreases. This effect does not change much as the total fishing hours in a year changes.

The results relating the probability and amount of fishing activity to bpi and distance to shore are not entirely consistent. But the results of the hurdle model as a whole indicate that areas with a bpi between 0 and -10 and a distance to shore between 2.5 and 7.5km are most preferred by the fishers.

For both the presence and amount of fishing activity, the seasons have a noticeable impact on the spatial distribution of the fisheries (see the random effect mesh maps in the Supplementary Materials). None of the seasons seem to be clearly favoured with respect to the presence of fishing activity. For the number of fishing hours, however, it was found that fishermen prefer the winter the least of all seasons. When taking the whole hurdle into account, it appears one can conclude that the spring is the most preferred season overall. These seasonal patterns generally correspond to shrimp availability as a result of their life cycle.

Which factors drive where shrimp fishermen decide to fish and where not to fish in the coastal part of North Sea, before and after fisheries closures? And which factors drive with how much effort (measured in fishing hours) fishermen attempt to fish in their fishing locations before and after some of their fishing locations have been closed-off (i.e. displacement)? Both before and after the closures of certain restricted fishing areas, the distance to those areas is generally similarly relevant for both boat groups. We have found, after correcting for other covariates (such as BPI, distance to coast, season, etc.) and after correcting for spatial autocorrelation, that fishers tend to fish close to those areas even after their closure, and they tend to fish with more effort when close to those areas, and less when further away from those areas.

6 References

Allaire, J., Xie, Y., McPherson, J., Luraschi, J., Ushey, K., Atkins, A., … Iannone, R. (2019). *Rmarkdown: Dynamic documents for r*. Retrieved from https://github.com/rstudio/rmarkdown

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A. I., Bolin, D., Illian, J., … Lindgren, F. K. (2018). Spatial modelling with R-INLA: A review. *WIRES (Invited Extended Review)*, (Feb). Retrieved from http://arxiv.org/abs/1802.06350

Dunn, P. K., & Smyth, G. K. (1996). Randomized Quantile Residuals. *Journal of Computational and Graphical Statistics*, *5*(3), 236–244. Retrieved from https://www.jstor.org/stable/1390802

Garnier, S. (2018). *Viridis: Default color maps from 'matplotlib'*. Retrieved from https://CRAN.R-project.org/package=viridis

Hintzen, N., Bastardie, F., & Beare, D. (2017). *Vmstools: For analysing fisheries vms (vessel monitoring system) data*. Retrieved from https://CRAN.R-project.org/package=vmstools

Iampietro, P., & Kvitek, R. (2002). Quantitative seafloor habitat classification using gis terrain analysis: Effects of data density, resolution, and scale. In *Proceedings of the 22nd annual esri user conference* (pp. 8–12).

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86.

Lindgren, F., & Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, *63*(19), 1–25. Retrieved from http://www.jstatsoft.org/v63/i19/

Lindgren, F., Rue, H., & Lindström, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(4), 423–498.

Lundblad, E. R., Wright, D. J., Miller, J., Larkin, E. M., Rinehart, R., Naar, D. F., … Battista, T. (2006). A benthic terrain classification scheme for american samoa. *Marine Geodesy*, *29*(2), 89–111.

R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rees, T. (2003). "C-squares," a new spatial indexing system and its applicability to the description of oceanographic datasets. *Oceanography*, *16*. doi:10.5670/oceanog.2003.52

RStudio Team. (2020). *RStudio: Integrated development environment for r*. Boston, MA: RStudio, PBC. Retrieved from http://www.rstudio.com/

Rue, H., Martino, S., & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society B*, *71*, 319–392.

Tierney, L. (2018). *Codetools: Code analysis tools for r*. Retrieved from https://CRAN.R-project.org/package=codetools

Walbridge, S.; Slocum, N.; Pobuda, M.; Wright, D.J. 2018. Unified Geomorphological Analysis Workflows with Benthic Terrain Modeler. Geosciences 8, 94. doi:10.3390/geosciences8030094

Weiss, A. (2001). Topographic position and landforms analysis. In *Poster presentation, esri user conference, san diego, ca* (Vol. 200).

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from https://ggplot2.tidyverse.org

Xie, Y. (2016). *Bookdown: Authoring books and technical documents with R markdown*. Boca Raton, Florida: Chapman; Hall/CRC. Retrieved from https://bookdown.org/yihui/bookdown

Zuur, A. F., Ieno, E. N., & Saveliev, A. A. (2017). Spatial, temporal and spatial-temporal ecological data analysis with r-inla. *Highland Statistics Ltd*, *1*.

Quality Assurance

Wageningen Marine Research utilises an ISO 9001:2015 certified quality management system. The organisation has been certified since 27 February 2001. The certification was issued by DNV.

Justification

Report C041/22
Project Number: 4318100322

The scientific quality of this report has been peer reviewed by a colleague scientist and a member of the Management Team of Wageningen Marine Research

Approved: Niels Hintzen
Fisheries researcher

Signature:

Date: 22-07-2022

Approved: Drs. J. Asjes
Manager Integration

Signature:

Date: 22-07-2022

Appendix 1 Quantitative interpretation of Coefficient plots

Let x_1 be some covariate, and let $f(x_1)$ the function of the effect of thus covariate. (i.e. for a piecewise linear spline, $f(x_1)$ would be the linear combination variables and their coefficients that make up a piecewise linear spline; for a categorical covariate, $f(x_1)$ would be the set of dummy variables and their coefficients corresponding to that categorical covariate).

The quantitative interpretation of the coefficient plots for the presence/absence models is as follows. Consider a covariate x_1. Suppose x_1 **changes** from value a to value b, and all other variables do not change. Then the **odds** of the presence of fishing activity will be **multiplied** by $\exp\left(\frac{f(x_1=b)}{f(x_1=a)}\right)$.

The quantitative interpretation of the coefficient plots for the non-zero models is as follows. Consider a covariate x_1. Suppose x_1 **changes** from value a to value b, and all other variables do not change. Then the response variable will be **multiplied** $\exp\left(\frac{f(x_1=b)}{f(x_1=a)}\right)$.

Appendix 2 Covariate specifications

Presence / absence model

It would not make sense to include the "IRA" variable in the presence/absence model, as IRA is always zero when the fishing hours are zero; thus IRA is excluded (otherwise the problem of complete separation is introduced). The other 2 variables of primary interest, namely "DNRA" (the distance to the nearest restricted area), and the R-periods, are included in the model, and so are their interactions.

The total number of fishing hours in a year was considered a relevant covariate for the presence/absence models: If the total number of fishing hours in a year is larger there may be an increase in the probability of fishing activity.

As the distance to the shore increases, one may expect the depth to change (generally increasing). So an interaction between fine bpi (positive and negative) and the "DS" covariate was included, allowing the fine bpi slope to change linearly as one moves away or towards the shore. It was suspected that the effect of the distance to the shore may be different when one is in shore rather than away from it. Therefore, in the aforementioned interaction term, the distance to shore was split into a "in shore" part (distance to the shore < 5km) and an "out shore" part (distance to the shore > 5km).

Based on the data plots, the distance tot the shore main effect was entered in the model as a piecewise linear spline with knots at 2, 4, and 7km.

Non-zero model

Of primary interest for this study were the variables "IRA" (whether a boat spent at least 1% of their effort in one of the restricted areas before 2017, or not), "DNRA" (the distance to the nearest restricted area), and R periods 3 and 4. Especially important are 2 sets of interactions:

- all interactions between IRA, DNRA, and R period three, including their 3-way interaction
- all interactions between IRA, DNRA, and R period four, including their 3-way interaction

How far away from the harbours fishermen might be willing to travel to fish might change as the total number of fishing hours in a year (partly an indicator of competition) increases. The total number of fishing hours in a year should not be entered as a main effect in the model, as it might interfere with the offset term (see Appendix 4). Therefore, the total number of fishing hours was only entered in an interaction term with the distance to the nearest harbour (DNH) covariate.

The square root of the positive and negative bpi variables were entered into the non-zero models as second degree regular polynomials. As the distance to the shore increases, one may expect the depth to change (generally increasing). So interaction terms between these polynomials and the distance to shore was also added.

All other covariates were also present, as main fixed effects.

Appendix 3 Model diagnostics

The fit for the non-zero model (gamma) were checked in 2 ways:

- A scatter plot was produced with the observed response on the y-axis, and the predictions on the x-axis. A straight line was fitted through these points, and the slope of this line (the fit slope) was determined. The closer the fit slope is to 1, the better.
- The mean absolute deviation ("MAD") was calculated between the observed response and the predictions, again both without binning or logarithm. The lower the MAD, the better.

The fit for the presence/absence models were checked in 3 ways:

- A scatter plot was produced with the observed response on the y-axis, and the predictions on the x-axis. A straight line was fitted through these points, and the slope of this line (the fit slope) was determined. The closer the fit slope is to 1, the better.
- The 0/1-loss was calculated between the observed presence/absence of fishing activity, and the predicted presence/absence of fishing activity. The lower the 0/1-loss, the better.
- Two complementary probabilities were calculated. One is the Sensitivity, which is the probability of correctly predicting a true 1 (presence), or P(prediction=1|observation=1). The other is the Specificity, which is the probability of correctly predicting a true 0 (absence), or P(prediction=0|observation=0). For both probabilities it holds that the higher (closer to 1), the better.

Dunn-Smyth residuals (Dunn & Smyth, 1996) were used for all models. Residual diagnostic plots were produced to check for violations of the model assumptions.

Convergence of the fixed effects coefficient estimates was checked using the Kullback-Leibler divergence (Kullback & Leibler, 1951), which R-INLA provides.

Appendix 4 Details on the Gamma GLMM formulation

As stated in the introduction section, the model should also be able to make predictions outside of the study period. Each year has a certain number of fishing hours in total over all locations and seasons, and years with more total fishing hours should naturally produce generally higher predictions then years with lower total fishing hours. Therefore, the actual number of fishing hours were not used as the response in the non-zero models; instead the response variable was the number of fishing hours as a proportion of the total number of fishing hours in a year divided by 8. The division by 8 is done because each year has 4 seasons and 2 boat-groups. Due to the use of a Gamma GLMM, this proportion was turned into a strictly positive ratio. See Appendix 4 for details.

Let y_i be the number of fishing hours at observation i (each observation is the specific combination of a year, location, season, and boat-group). The proportion of the non-zero fishing hours at observation i can be expressed as follows:

$$\text{proportion}(y_i) = \frac{y_i}{1/8 \times y_{\text{yeartotal}}} \qquad (3.1)$$

where $y_{\text{yeartotal}}$ is the sum of the fishing hours in a specific year over all locations, seasons, and boat-groups.

For ease of modelling, this proportion was turned into the following strictly positive ratio:

$$\text{ratio}(y_i) = \frac{y_i}{1/8 \times y_{\text{yeartotal}} - y_i} \qquad (3.2)$$

As a Gamma model with log-link function was used, one can formulate the relation between the linear predictors (the linear combination of all fixed and random effects), η, and the response as follows:

$$\log\big(E(y_i|y > 0)\big) = \eta_i + \text{offset}_i \qquad (3.3)$$

where $\text{offset}_i = \log\big(1/8 \times y_{\text{yeartotal}} - y_i\big)$ is the offset term: a fixed effect where the coefficient is forced to be equal to 1. We have found some problems using R-INLA's "regular" offset methods, so the offset term was instead introduced as a fixed effects with a coefficient whose prior is defined as Norm($\mu = 1$, $\sigma^2 = \frac{1}{40000}$), which forces the coefficient to be estimated close to 1.

Predicting the non-zero fishing hours, whether it be on the training data (data used for fitting the model) or new data (for the future), is done as follows:

$$\text{predict}\big(E(y|y > 0)\big) = \frac{E(\text{ratio}(y)|y > 0)}{1 + E(\text{ratio}(y)|y > 0)} \times \frac{y_{\text{yeartotal}}}{8} \qquad (3.4)$$

where $E(\text{ratio}(y)|y > 0)$ are the predictions from the gamma model when ignoring the offset term.

Appendix 5 Other noteworthy R packages that were used

All R-functions we wrote ourselves were checked for scoping with the help of the **codetools** R-package (Tierney, 2018). This report was written with the **Bookdown** (Xie, 2016) extension of R-Markdown (Allaire et al., 2019) of R-studio. Most of the figures shown in this report relied primarily on the **ggplot2** (Wickham, 2016) R-package, accompanied by **viridis** (Garnier, 2018)

Wageningen Marine Research
T +31 (0)317 48 7000
E: marine-research@wur.nl
www.wur.eu/marine-research

Visitors' address
- Ankerpark 27 1781 AG Den Helder
- Korringaweg 7, 4401 NT Yerseke
- Haringkade 1, 1976 CP IJmuiden

With knowledge, independent scientific research and advice, **Wageningen Marine Research** substantially contributes to more sustainable and more careful management, use and protection of natural riches in marine, coastal and freshwater areas.

Wageningen Marine Research is part of Wageningen University & Research. Wageningen University & Research is the collaboration between Wageningen University and the Wageningen Research Foundation and its mission is: 'To explore the potential for improving the quality of life'