Investigating effects of shipping on common scoter and red-throated diver distributions in Liverpool Bay SPA

Appendices B, C, D and E

First published September 2022

Natural England Commissioned Report NECR425

www.gov.uk/natural-england

About Natural England

Natural England is here to secure a healthy natural environment for people to enjoy, where wildlife is protected and England's traditional landscapes are safeguarded for future generations.

Further Information

This report can be downloaded from the Natural England Access to Evidence Catalogue. For information on Natural England publications or if you require an alternative format, please contact the Natural England Enquiry Service on 0300 060 3900 or email enquiries@naturalengland.org.uk.

Copyright

Natural England photographs are only available for non-commercial purposes. If any other photographs or information such as maps or data cannot be used commercially this will be made clear within the report.

For information regarding the use of maps or data see our guidance on How to access Natural England's maps and data.

ISBN 978-1-78354-947-4

Catalogue code: NECR425

Report details

Author(s)

ML Burt[1], ML Mackenzie[1], G Bradbury[2] and J Darke[2]

Contractor

1. Centre for Research into Ecological and Environmental Modelling, University of St Andrews, St Andrews, KY16 9LZ

2. Wildfowl and Wetlands Trust (consulting) Limited, Slimbridge, Gloucestershire, GL2 7BT

Natural England Project Manager

Tim Frayling

Keywords

Common scoter, red-throated diver, Liverpool Bay, shipping, disturbance, waterbirds, Special Protected Area (SPA), offshore wind farms (OWF), spatial modelling, digital aerial survey, AIS, MRSea

Citation

Burt, M.L., Mackenzie, M.L., Bradbury, G. and Darke, J. 2022. Investigating effects of shipping on common scoter and red-throated diver distributions in Liverpool Bay SPA. NECR425. Natural England.

Appendix B: Summary of candidate explanatory variables used in the modelling

Digital aerial surveys were flown on five days over Liverpool Bay SPA. The realised search effort along the transect lines was divided into small sections, or segments, approximately 1km in length. Associated with each segment was information about the habitat and shipping traffic. This appendix describes the information available.

For (potentially) continuously varying covariates (e.g. depth) the covariate was measured at each location where birds were detected or, if no birds were detected within the segment, at the midpoint of the segment. Where more than one bird was located per segment a relevant summary value of the covariate was used (e.g. average value) to describe the segment.

Other variables indicated whether, or not, there was an intersection of some anthropogenic activity (e.g. shipping track) with the segment area. Segments were overlain in a GIS with layers containing habitat and shipping information to determine intersections (e.g. to determine if a segment fell within a footprint of a windfarm).

Shipping data obtained from the ShipAIS webservice was extracted for the day of the survey (i.e. 5 dates) and also for the day before the surveys took place. By combining this shipping information, different temporal scales of shipping were incorporated.

Data on infrastructure within the region were obtained from the Crown Estate website (http: www.thecrownestate.co.uk/energy-minerals-and-infrastucture/downloads/maps-and-gis-data/). Data were freely available in shapefile format.

The candidate explanatory variables used in the modelling process are described below. The names in italics are the names used to describe the variable in the modelling process.

- *x.pos*, *y.pos* - location at the mid point of the segment is defined using the Universal Transverse Mercator (UTM) coordinate system and UTM zone 30. Units are in metres.
- *depth* - depth (m) obtained from the UK Hydrographic Office six second bathymetry digital elevation model used under Licence to Natural England. This was measured at the mid point of the segment or, if birds were identified within a segment, at the location of each bird. If the depth differed between birds within a segment, the average value was used. Segments with no associated depth data (over land) were initially assigned as missing (see below).
- *salinity* - salinity (parts per 1000) obtained from the Proudman Oceanographic Laboratories POLCOM 2004 data. This was measured at the mid point of the segment or, if birds were identified within a segment, at the location of each bird. If salinity differed between birds within a segment, the average value was used. Segments with no associated salinity (some inshore areas) were initially assigned as missing (see later). More recent data, including those for 2011 and 2015, were not available.
- $LBspa$ - indicated whether the segment was inside or outside of the Liverpool Bay SPA
- $fish$ - presence (1) or absence of fishing (0) within a segment derived from the total time of fishing in a segment obtained from the MMO fishing boat layer EAIEAEW00170029. This described fishing activity for commercial UK vessels 15 m and over in 2011. Data for 2015 were not available.
- $windfarm$ - the intersection (1) or no intersection (0) of the segment with windfarms where the foundations were constructed by the time of the survey defined in the Crown Estates Offshore wind farms layer.
- *anthrop* - the intersection (1) or no intersection (0) of the segment with anthropogenic activities or effects defined in the Crown Estate layers for NGS pipeline, Lease gas storage, minerals and aggregates and wind export cables
- *windcable* - the intersection (1) or no intersection (0) of the segment with the Crown Estates wind export cable layer

- $shipAIS1$ - intersection (1) or no intersection (0) of the segment with shipping tracks obtained from the ShipAIS web service for the day of the survey only
- $shipAIS5$ - intersection (1) or no intersection (0) of the segment with combined shipping tracks obtained from the ShipAIS web service for all five survey dates (i.e. the intersection with the combined shipping tracks for 5 days in total)
- $shipAIS10$ - intersection (1) or not (0) of the segment with combined shipping tracks obtained from the ShipAIS web service for all five survey dates plus the day before each survey (i.e. the intersection with the combined shipping tracks for 10 days in total)
- $shipMMO$ - intersection (1) or not (0) of the segments with all shipping tracks obtained from the MMO AIS dataset.
- $shipnumf$ - the number of ships intersecting with the segment for day of the survey. This was included as factor variable with three levels, no ships (0), 1-4 ships (1) and >4 ships (2).
- *shiplenav* - the average length (metres) of the ships for shipping intersecting with the segments on the day of the survey. If this differed between birds within a segment, the average length was used. If there were no ships intersecting with the segment, the length was zero.
- *shiplenmax* - the maximum length of the ships for shipping intersecting with the segments on the survey. If this differed between birds within a segment, the maximum length was used. If there were no ships intersecting with the segment, the length was zero.
- *shiplennear* - the maximum length of the nearest ship for shipping intersecting with the segments on the survey. If this differed between birds within a segment, the maximum length was used.
- *shipnear* - nearest distance to ship (km) obtained from the shipping tracks for the day of the survey. This was measured from the mid point of the segment or, if birds were identified within the segment, to each bird. If the distance differed between birds within a segment, the minimum distance was used.

The variable describing anthropgenic activities and effects *anthrop* was a combination of several layers describing NGS pipeline, gas storage, mineral and aggregates and wind export cable but does not include windfarms. With the exception of the windcable, which was also included as a potential explanatory variable, there were not enough intersections with segments to include these other layers as explanatory variables separately.

Two new variables were created, which were a combination of $shipMMO$ and *shiplenav* or *shiplenmax*. These created variables that described segments which never contained a ship (based on the MMO data), sometimes saw a ship (but not on the day of the survey) and saw a ship on the day of the survey and the category was specified by the average or maximum length of the ships.

The quantiles of the length distributions were used to create the categories in order to have roughly equal number of segments in each category (Table B1 and B2).

Table B1. Quantiles of the average length of ships (where ships were present in a segment).

```
 0%  20%  40%  60%  80% 100%
 10   15   20   36   89  274
```

Table B2. Quantiles of the maximum length of ships (where ships were present in a segment).

```
 0%  20%  40%  60%  80% 100%
 10   15   21   36   95  292
```

Using the quantiles of the length distributions, these were classified into 7 levels as follows:

- *shipcatAv* [2] - no ships present (0), ships present but not on the day of the survey (1) and using the average length of ships present on the day of the survey the other factor levels were, small ship 10-15m in length on average (2), ships >15-20m (3), ships >20-36m (4), ships >36-89m and ships >89m in length on average.

- *shipcatMax* [2] - no ships present (0), ships present but not on the day of the survey (1) and using the maximum length of ships present on the day of the survey the other factor levels were, small ship 10-15m in length on average (2), ships >15-21m (3), ships >21-36m (4), ships >36-95m and ships >95m in length on average.

Plots of the explanatory variables are given below. Values are averaged over surveys for segments in the same location.

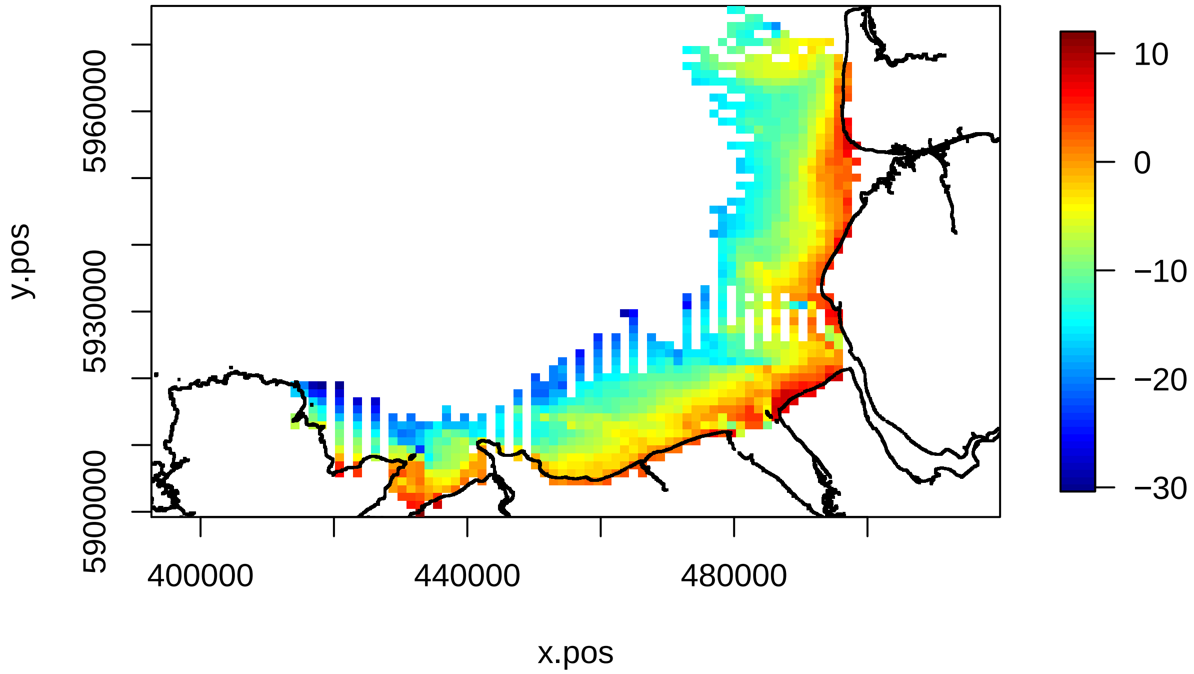


Figure B1. Depth (m) for all surveys.

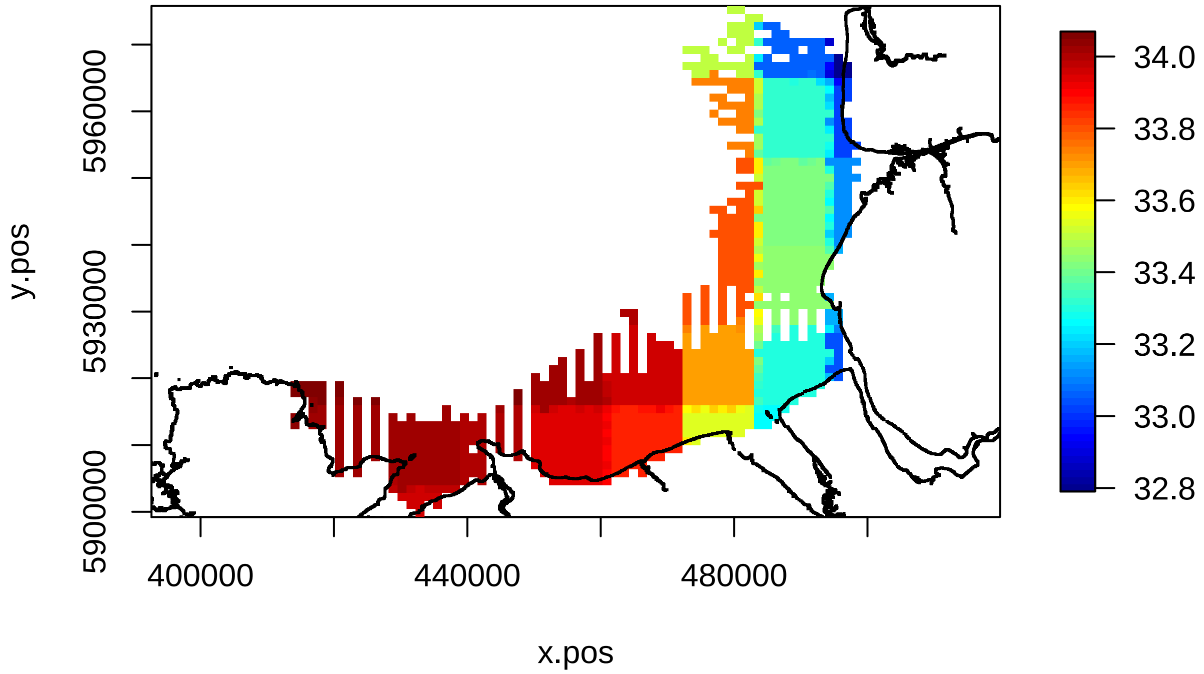


Figure B2. Salinity (ppt) for all surveys.

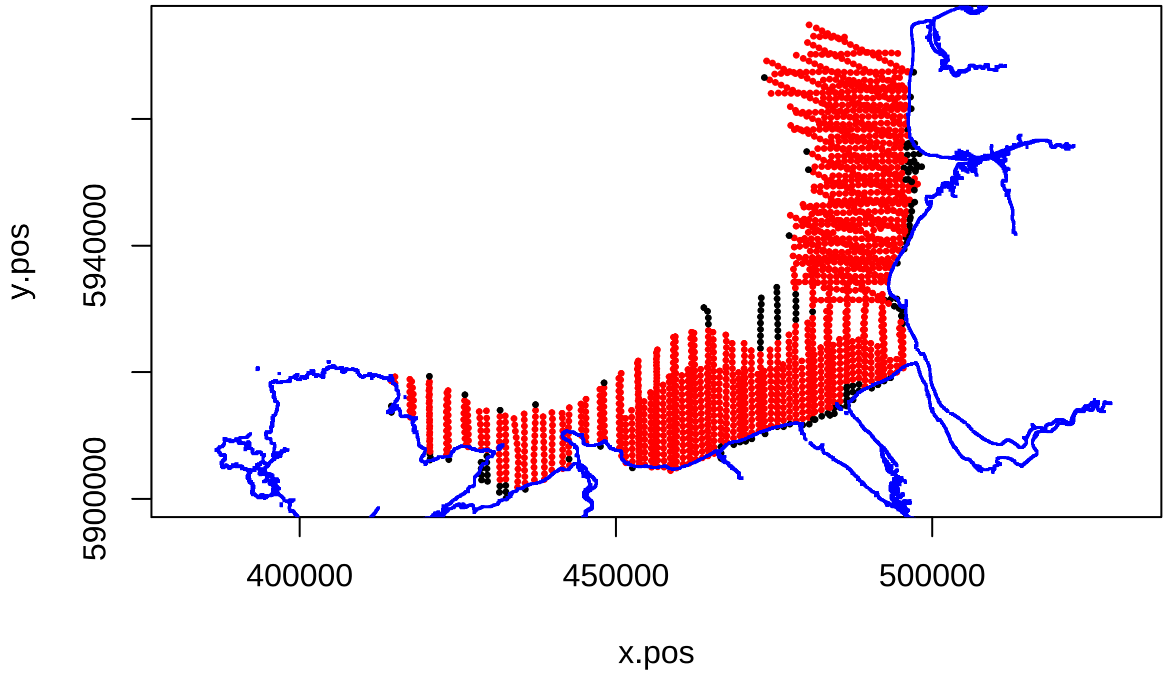


Figure B3. Segments inside (red) and outwith (black) the Liverpool Bay SPA (*LBspa*).

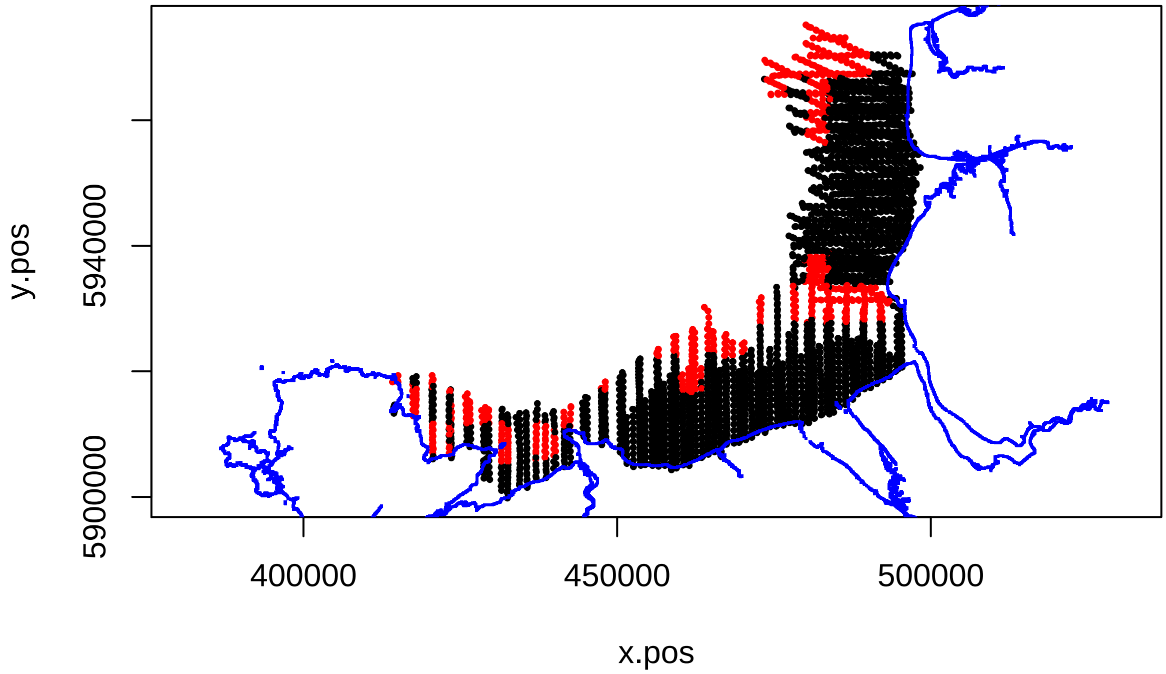


Figure B4. Presence (red) and absence (black) of fishing activity within a segment ($fish$).

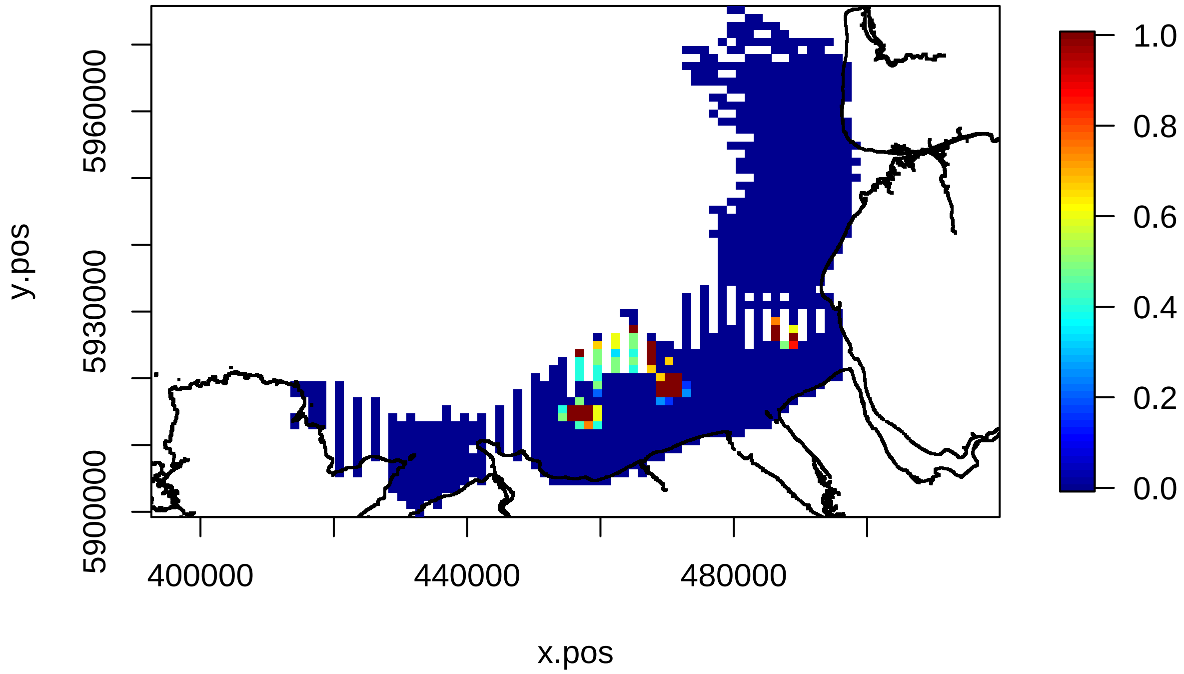


Figure B5. Intersection of windfarm foundations constructed at the time of the survey ($windfarm$). The values are averaged over all surveys and so values in the figure not equal to zero (no windfarm) or one (windfarm) indicate that a windfarm has been built between the dates of the surveys.

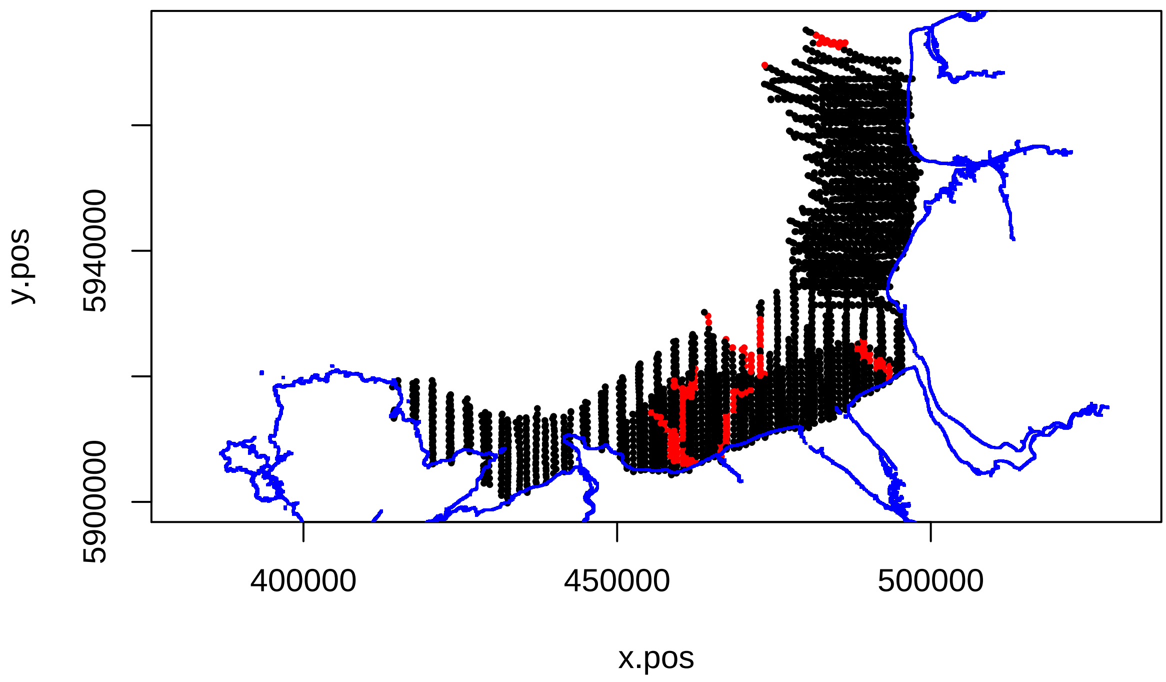


Figure B6. Presence (red) and absence (black) of anthropogenic activities and effects (*anthrop*).

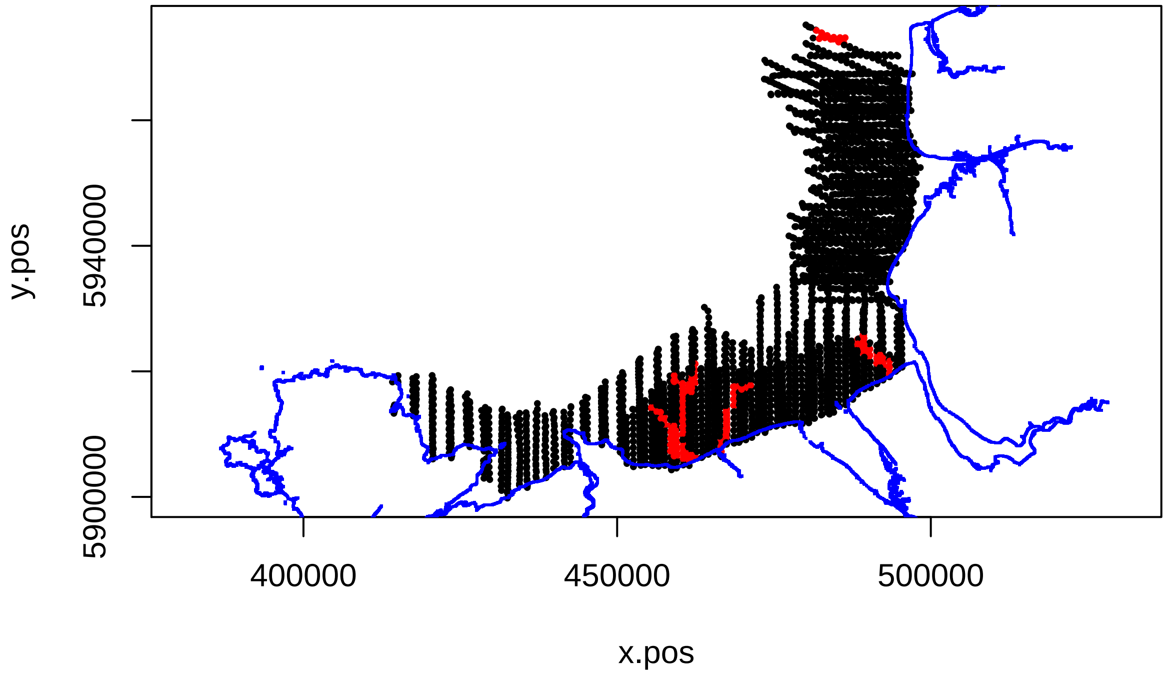


Figure B7. Presence (red) and absence (black) of wind export cable (*windcable*).

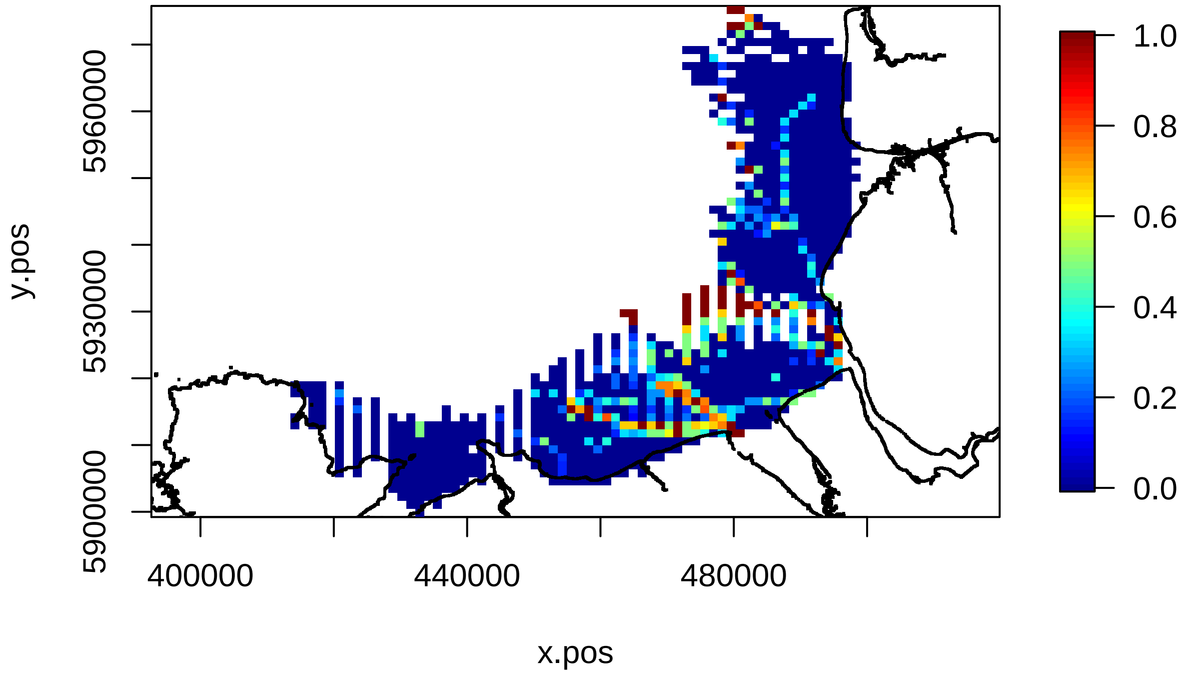


Figure B8. Ship traffic on the day of the survey (*shipAIS*1) averaged over all surveys. Values not equal to zero (no shipping) or one (shipping track) indicate temporal differences between survey dates.

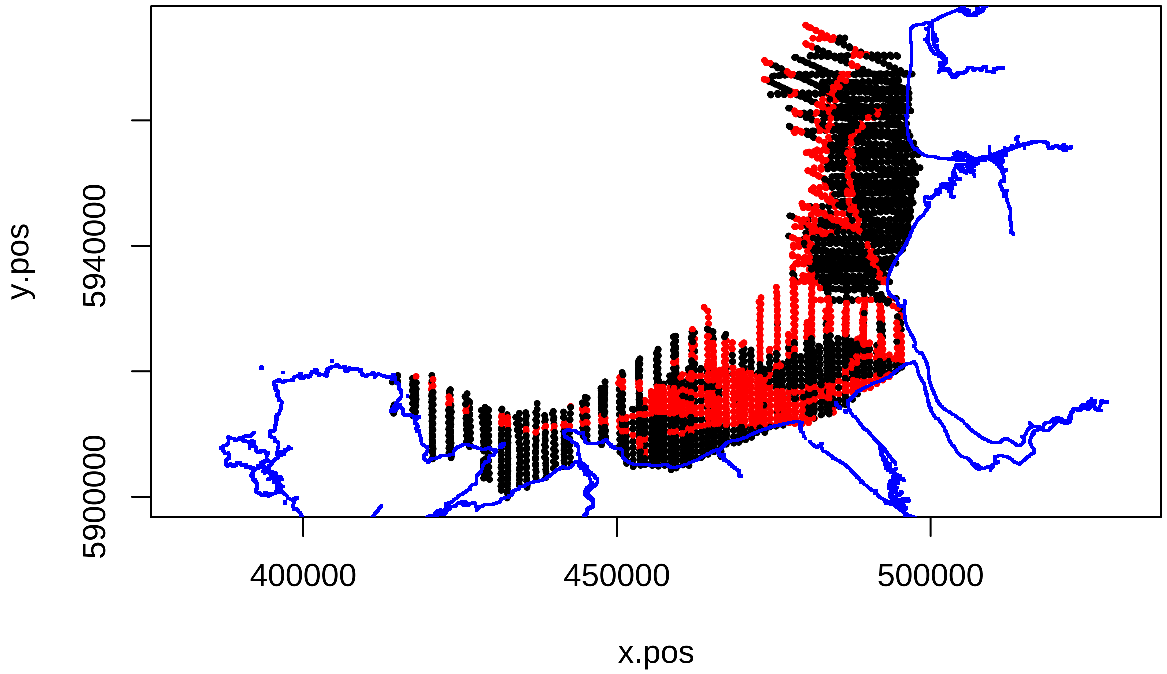


Figure B9. Intersection (red) or no intersection (black) with combined ship tracks from the five survey dates ($shipAIS5$).

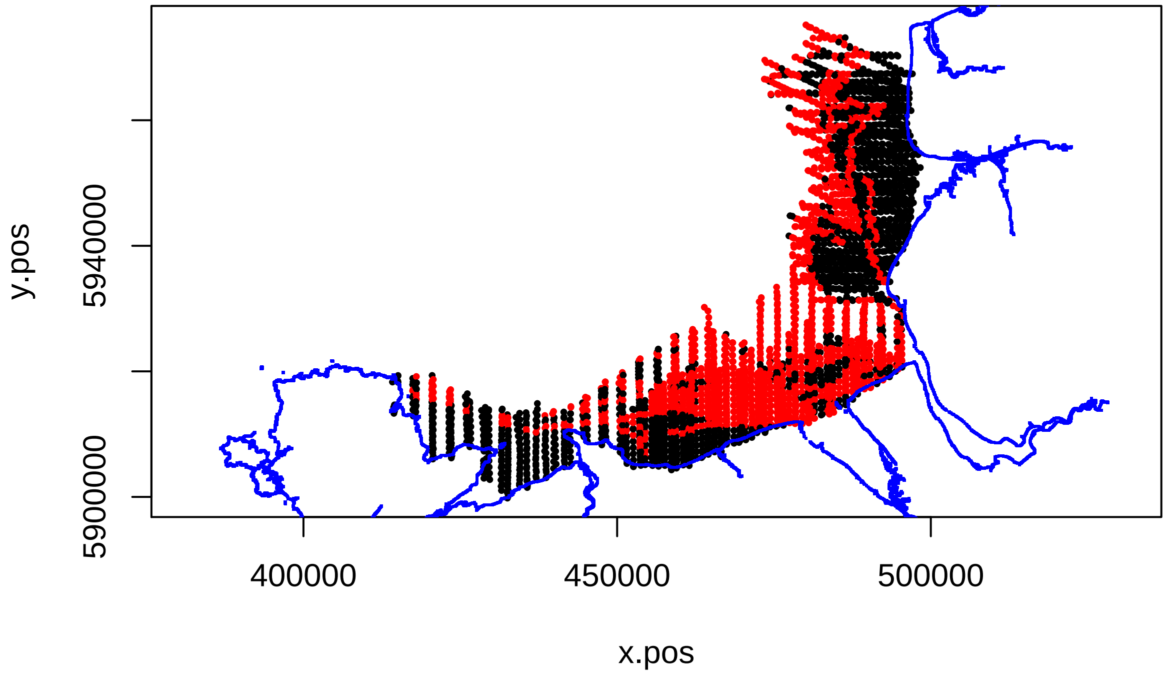


Figure B10. Intersection (red) or no intersection (black) with combined ship tracks from the five survey dates and the five previous days ($shipAIS10$).

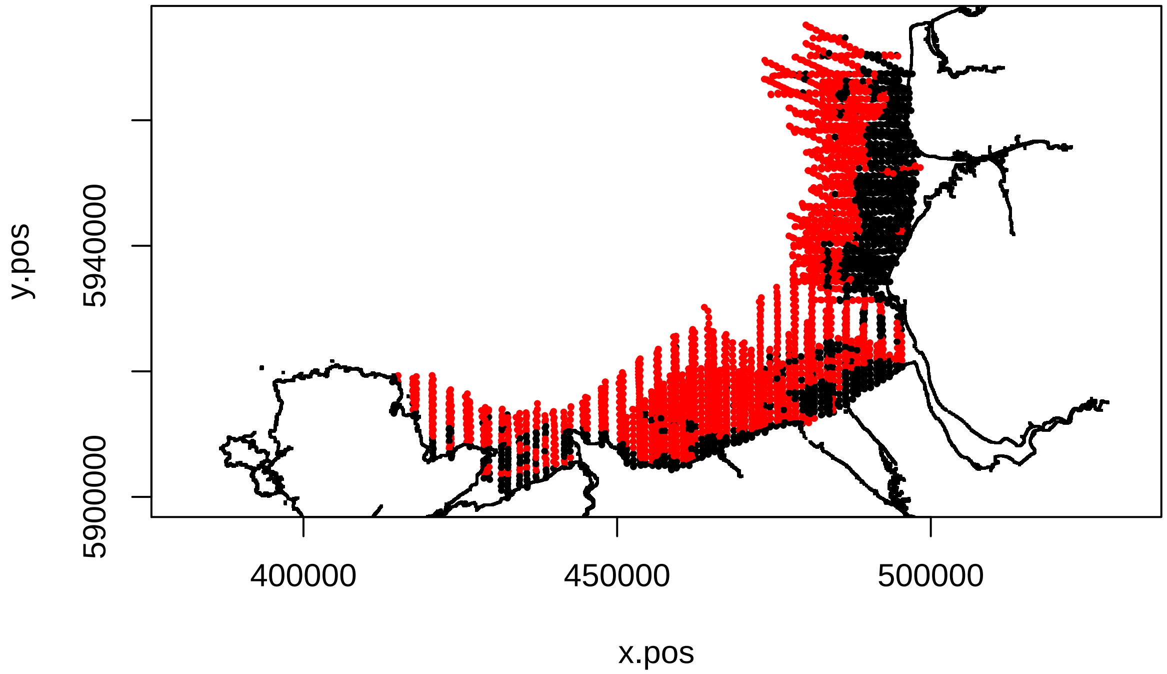


Figure B11. Intersection (red) or no intersection (black) with shipping tracks from the MMO AIS dataset ($shipMMO$).

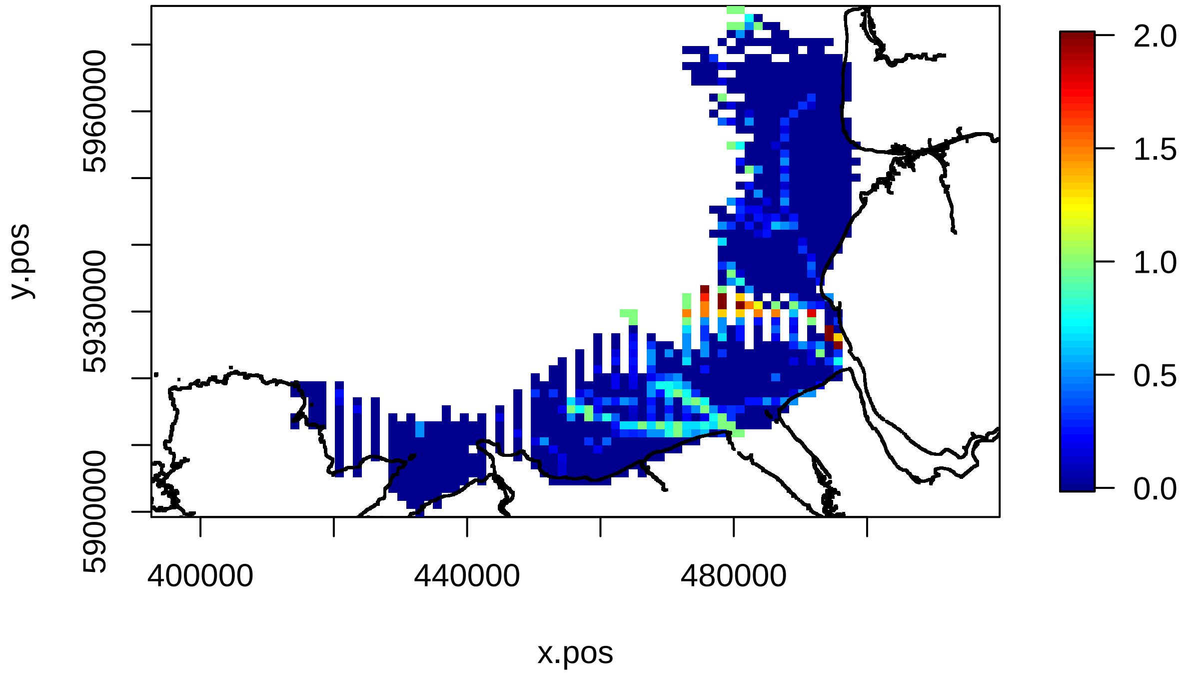


Figure B12. The number of ships (*shipnumf*) averaged over all surveys. This was included in the modelling as a factor type variable and values on the plot not equal to zero (no ships), one (1-4 ships) and two (>4 ships) indicate temporal differences between the surveys.

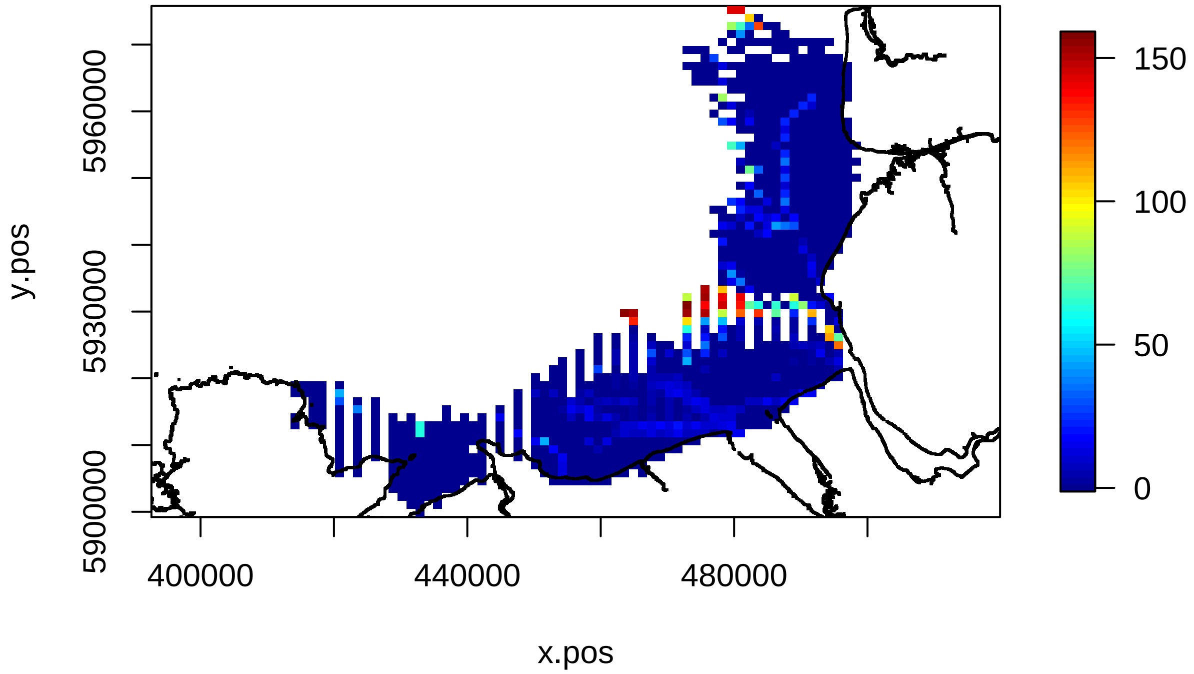


Figure B13. Average length (m) of ships intersecting with the segment (*shiplenav*) with values averaged over all surveys.

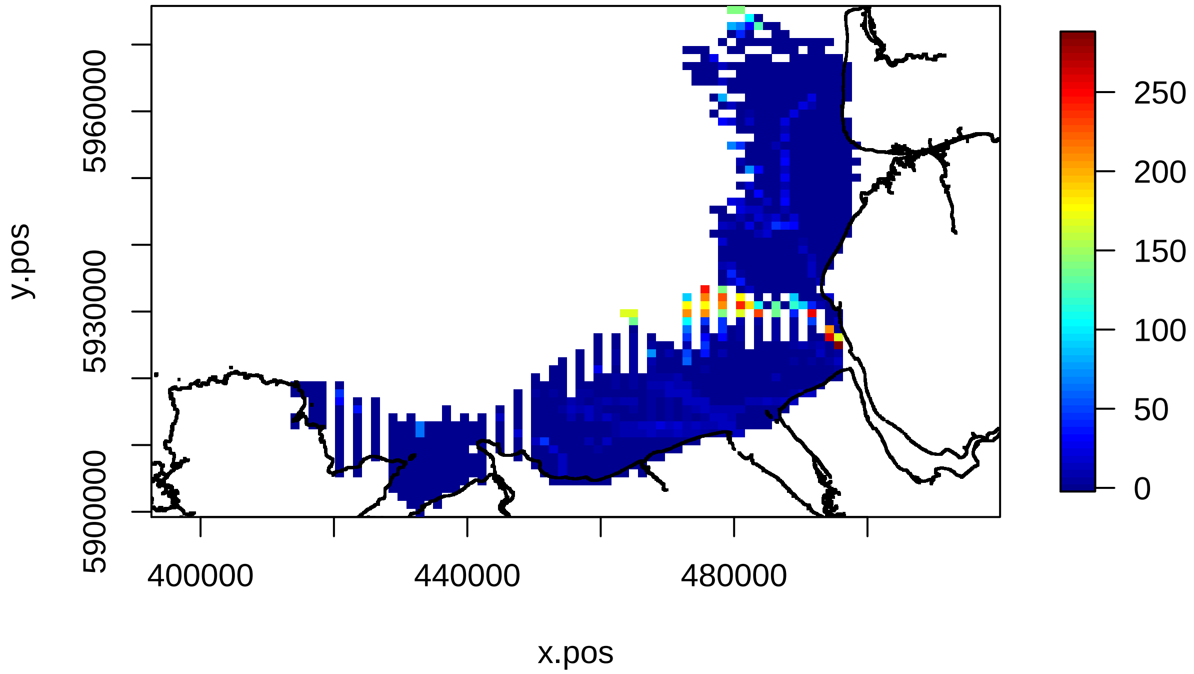


Figure B14. Maximum length (m) of ships intersecting with the segment (*shiplenmax*) with values averaged over all surveys.

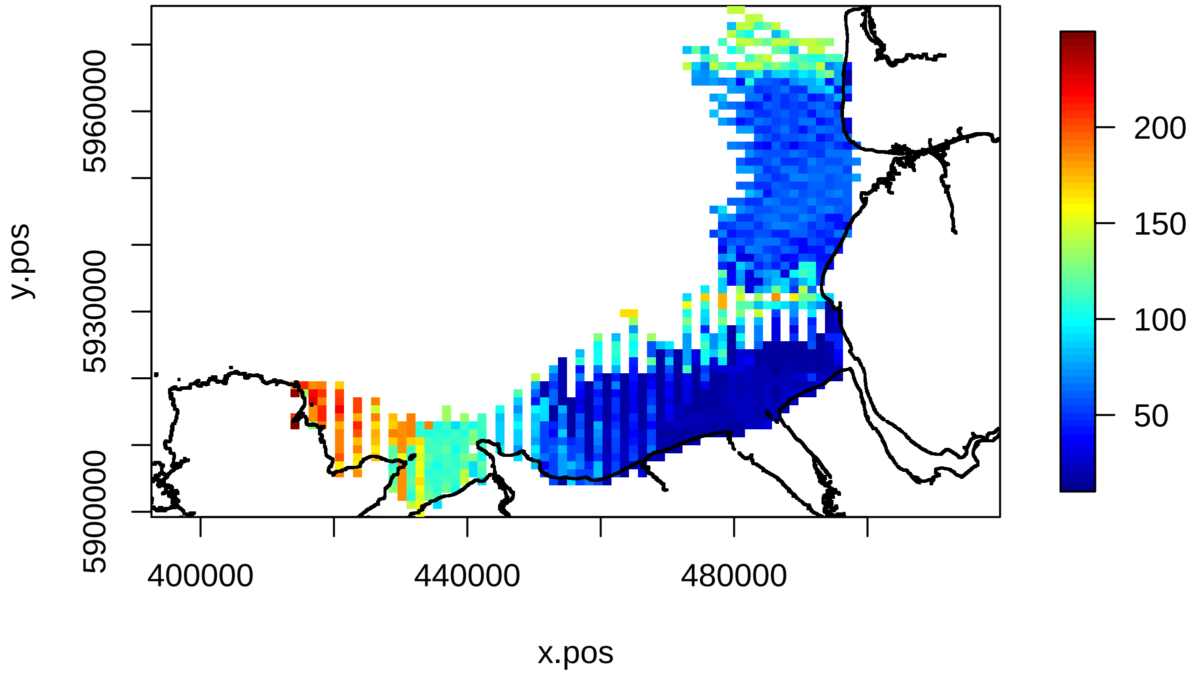


Figure B15. Length of the nearest ship (m) (*shiplennear*) with values averaged over all surveys.

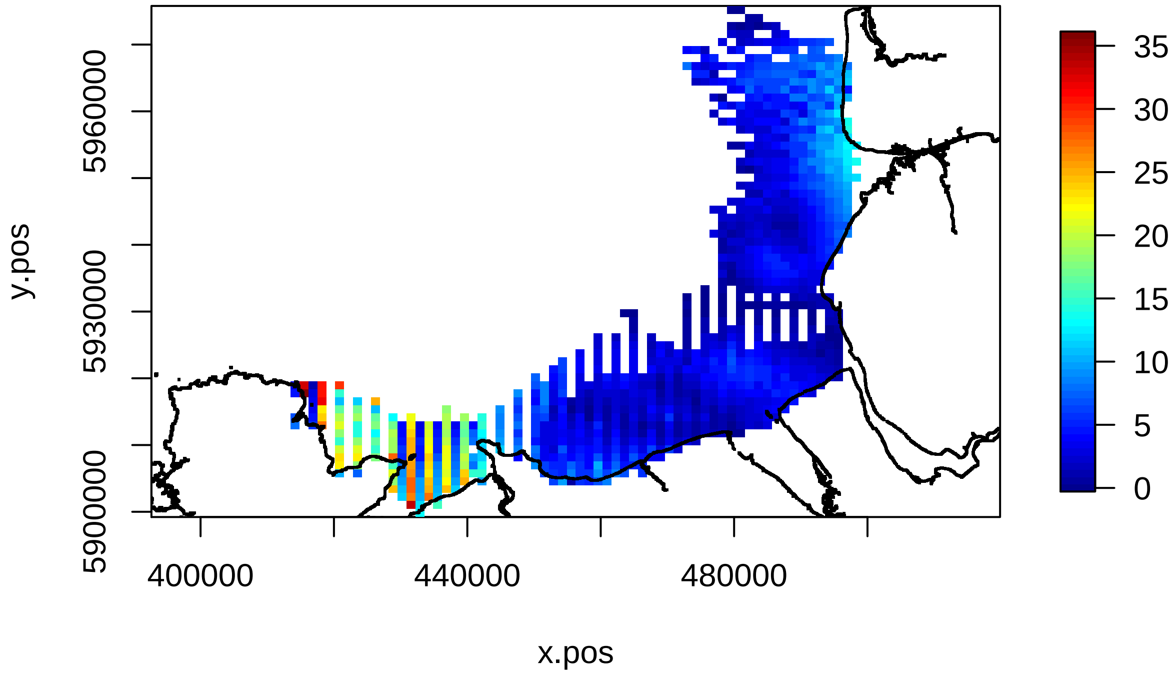


Figure B16. Nearest distance (km) to shipping tracks (*shipnear*) with values averaged over all surveys.

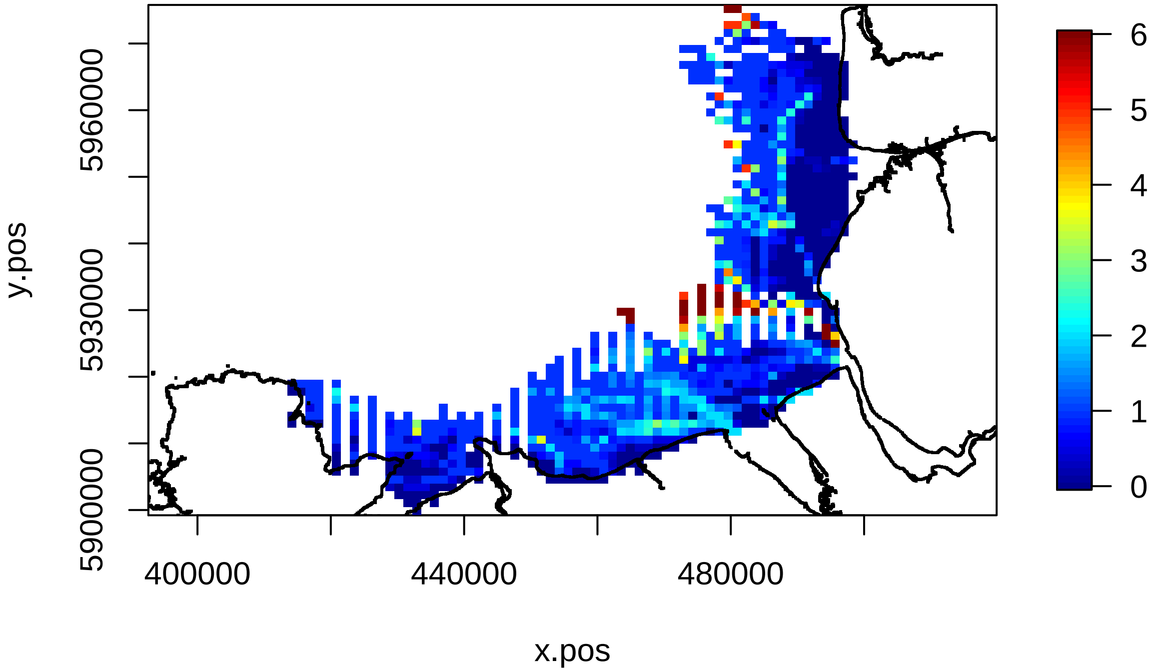


Figure B17. Combination variable *shipcatAv* with values averaged over all surveys. This was included in the modelling as a factor variable (with levels 0 to 6) and so values different from these indicate temporal differences between survey locations.

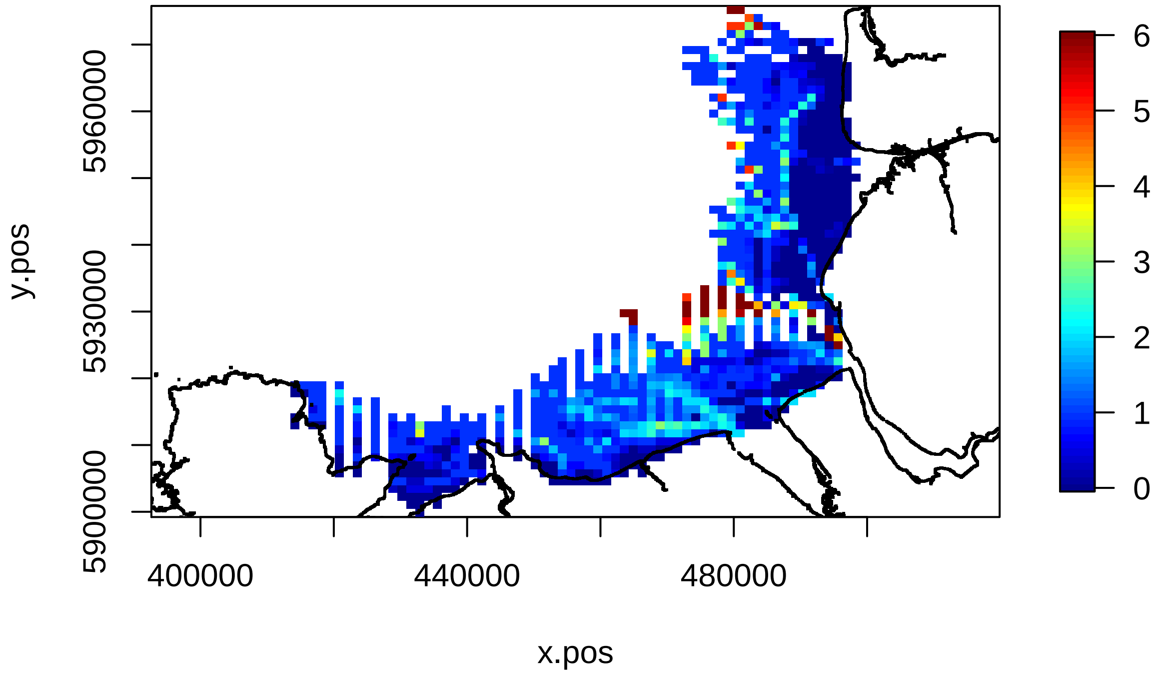


Figure B18. Combination variable $shipcatMax$ with values averaged over all surveys. This was included in the modelling as a factor variable (with levels 0 to 6) and so values different from these indicate temporal differences between survey locations.

Appendix C. Model selection for Common Scoter sitting on the water.

This appendix describes the model selection for Common Scoter sitting on the water.

```
## Loading required package: fields

## Warning: package 'fields' was built under R version 3.3.3

## Loading required package: spam

## Warning: package 'spam' was built under R version 3.3.3

## Loading required package: grid

## Spam version 1.4-0 (2016-08-29) is loaded.
## Type 'help( Spam)' or 'demo( spam)' for a short introduction
## and overview of this package.
## Help for individual functions is also obtained by adding the
## suffix '.spam' to the function name, e.g. 'help( chol.spam)'.

##
## Attaching package: 'spam'

## The following objects are masked from 'package:base':
##
##     backsolve, forwardsolve

## Loading required package: maps

## Warning: package 'maps' was built under R version 3.3.3

## Loading required package: car

## Warning: package 'car' was built under R version 3.3.3

## Loading required package: MRSea

## Loading required package: splines

## Loading required package: mgcv

## Loading required package: nlme

## This is mgcv 1.8-15. For overview type 'help("mgcv-package")'.

## Loading required package: boot

## Warning: package 'boot' was built under R version 3.3.3

##
## Attaching package: 'boot'

## The following object is masked from 'package:car':
##
##     logit
```

Due to a few, very large values of Common Scoter in a segment (Figure C1), the number per segment was capped to 2,150 birds.

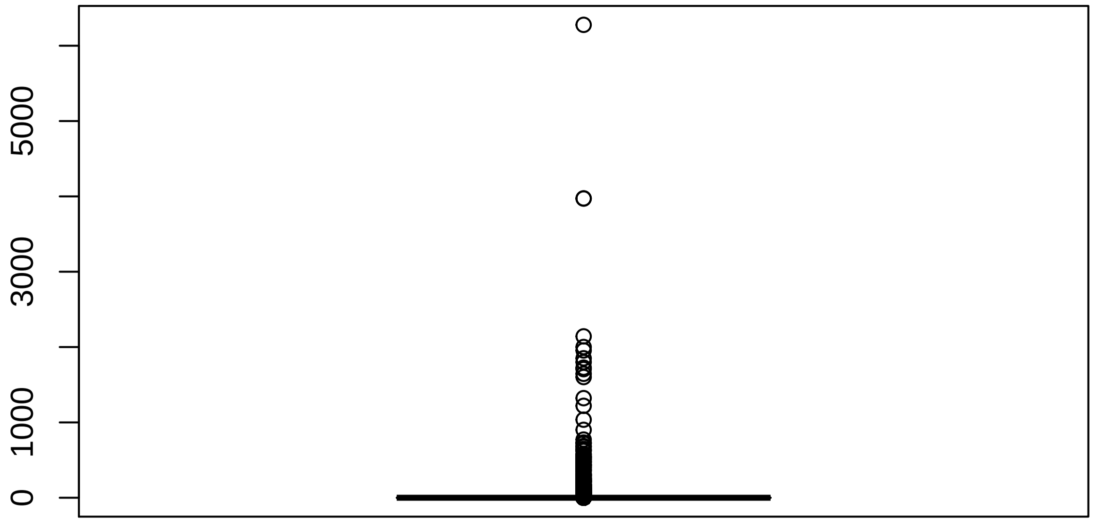


Figure C1. The distribution of the number of Common Scoter identified per segment that were sitting on the water.

Assessing candidate explanatory variables

Each explanatory variable was fitted separately to determine their likely importance and to decide which variables to select for terms which can not be fitted together in the same model (Table 1). The variables *shiplennear* and *shiplenav* were fitted as linear terms because fitting them as one-dimensional smooth terms caused model fitting errors.

Table C1. The pseudo R^2 (R2, a measure of the correlation between the observed values and the fitted values from the model) and probability (p.val) associated with fitting each term seperately. Numbers in the Group column indicate variables which were grouped together and one variable from each group was chosen.

	Variable	meanCV	lowCV	highCV	R2	p.val	Group
17	s(depth)	15676	15618	15761	0.05513	0	
10	as.factor(shipcatAv)	15766	15747	15801	0.04842	1.592e-14	1
11	as.factor(shipcatMax)	15766	15750	15792	0.04841	1.794e-14	1
14	as.factor(shipMMO)	15795	15776	15820	0.04692	1.419e-13	
13	as.factor(shipAIS10)	15803	15790	15819	0.04253	5.094e-18	
18	s(salinity)	15956	15905	16027	0.04004	5.869e-07	
16	s(shipnear)	15995	15943	16050	0.03665	0.0004311	
12	as.factor(shipAIS5)	16013	16001	16037	0.02929	4.364e-08	
8	as.factor(shipnumf)	16047	16038	16059	0.02658	5.131e-07	1
9	as.factor(shipAIS1)	16048	16037	16066	0.02658	1.789e-07	1
2	shiplenav	16055	16045	16069	0.02652	0.009855	

	Variable	meanCV	lowCV	highCV	R2	p.val	Group
15	s(shiplenmax)	16076	16047	16111	0.02658	1.11e-16	
5	as.factor(windfarm)	16087	16079	16100	0.02411	1.324e-05	
3	as.factor(LBspa)	16110	16101	16127	0.02249	0.001553	
6	as.factor(anthrop)	16139	16130	16150	0.02108	0.8328	
1	shiplennear	16140	16126	16160	0.02021	0.1393	
7	as.factor(windcable)	16140	16129	16154	0.02088	0.3196	
4	as.factor(fish)	16143	16130	16161	0.02178	0.2867	

The variable chosen from group 1 was *shipcatAv*.

Checking for collinearity

To check for collinearity all the (selected) variables were fitted in a model. The continuous terms were fitted as linear terms (instead of as smooth functions). The generalised variance inflation factors (GVIFs) are shown below and, because a term has more than one degree of freedom (Df), adjusted GVIFs (GVIFˆ(1/(2*Df))) are also shown.

	GVIF	Df	GVIFˆ(1/(2*Df))
as.factor(shipcatAv)	2578	6	1.924
as.factor(shipMMO)	161.2	1	12.7
as.factor(shipAIS10)	4.24	1	2.059
as.factor(shipAIS5)	4.235	1	2.058
as.factor(windfarm)	1.002	1	1.001
as.factor(LBspa)	1.05	1	1.025
as.factor(anthrop)	3775264	1	1943
as.factor(windcable)	3775264	1	1943
as.factor(fish)	1.235	1	1.111
shiplennear	1.46	1	1.208
depth	1.674	1	1.294
salinity	1.292	1	1.137
shipnear	1.325	1	1.151
shiplenav	34.2	1	5.848
shiplenmax	42.49	1	6.518

There are a few large GVIF (>5)

- *shipMMO* - this is likely correlated to *shipcatAv*. The latter is preferred in the model and so *shipMMO* is excluded
- *anthrop* and *windcable* - *anthrop* was chosen because it had a lower CV score (see above)
- *shiplenav* and *shiplenmax* - *shiplenav* was chosen because it had a lower CV score (see above)

Selection of the factor and 1D continuous terms

All selected factor and 1-dimensional terms were included in a model. Non-significant terms were excluded, one at a time, starting with the least significant term. The process of excluding non significant terms and refitting the model was repeated until all terms were significant.

```
## Warning: glm.fit: algorithm did not converge
```

```
## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: algorithm did not converge
## Analysis of 'Wald statistic' Table
## Model: quasipoisson, link: log
## Response: response
## Marginal Testing
## Max Panel Size = 27; Number of panels = 280
##
##                      Df     X2 P(>|Chi|)
## as.factor(shipcatAv)  6 26.770 0.0001599 ***
## as.factor(shipAIS10)  1 38.009  7.04e-10 ***
## as.factor(shipAIS5)   1 14.265 0.0001587 ***
## shiplenav             1 14.026 0.0001803 ***
## as.factor(windfarm)   1 98.769 < 2.2e-16 ***
## as.factor(LBspa)      1  0.410 0.5221164
## as.factor(anthrop)    1  2.862 0.0907063 .
## shiplennear           1  2.082 0.1490599
## as.factor(fish)       1  1.188 0.2758299
## salinity              1  0.133 0.7154275
## s(depth)              3 89.589 < 2.2e-16 ***
## s(shipnear)           3  3.688 0.2971836
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## NumIter      Model       CV   meanCV    lowCV   highCV p.value
## 2.5%       1 Full model 15319.25 15621.21 15464.94 15813.69      NA
##      numvalid
## 2.5%      100
```

Despite reducing the maximum number of knots for *salinity* there was still a problem and so *salinity* was included as a linear rather than as a smooth term.

However, *salinity* was the first term to be excluded.

```
## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: algorithm did not converge
## Analysis of 'Wald statistic' Table
## Model: quasipoisson, link: log
## Response: response
## Marginal Testing
## Max Panel Size = 27; Number of panels = 280
##
##                      Df     X2 P(>|Chi|)
## as.factor(shipcatAv)  6 26.000 0.0002226 ***
## as.factor(shipAIS10)  1 38.188 6.424e-10 ***
## as.factor(shipAIS5)   1 14.030 0.0001799 ***
## shiplenav             1 13.983 0.0001845 ***
## as.factor(windfarm)   1 99.660 < 2.2e-16 ***
## as.factor(LBspa)      1  0.444 0.5051739
## as.factor(anthrop)    1  2.479 0.1153638
## shiplennear           1  2.118 0.1455489
```

```
## as.factor(fish)       1  1.214 0.2704441
## s(depth)              3 90.233 < 2.2e-16 ***
## s(shipnear)           3  3.735 0.2915908
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##      NumIter      Model       CV   meanCV    lowCV   highCV   p.value
## 2.5%       2 - salinity 15307.55 15578.74 15441.34 15760.44 0.7154275
##      numvalid
## 2.5%      100
```

The next term to be removed was *LBspa*.

```
## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: algorithm did not converge
```

```
## Analysis of 'Wald statistic' Table
## Model: quasipoisson, link: log
## Response: response
## Marginal Testing
## Max Panel Size = 27; Number of panels = 280
##
##                      Df      X2 P(>|Chi|)
## as.factor(shipcatAv)  6  26.149 0.0002088 ***
## as.factor(shipAIS10)  1  37.976 7.162e-10 ***
## as.factor(shipAIS5)   1  14.072 0.0001759 ***
## shiplenav             1  14.270 0.0001584 ***
## as.factor(windfarm)   1 100.073 < 2.2e-16 ***
## as.factor(anthrop)    1   2.530 0.1117067
## shiplennear           1   2.079 0.1492923
## as.factor(fish)       1   1.237 0.2660118
## s(depth)              3  84.993 < 2.2e-16 ***
## s(shipnear)           3   3.751 0.2896249
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##      NumIter              Model       CV   meanCV    lowCV   highCV
## 2.5%       3 - as.factor(LBspa) 15299.62 15574.54 15437.57 15757.38
##        p.value numvalid
## 2.5% 0.5051739      100
```

The term *shipnear* was excluded next.

```
## Analysis of 'Wald statistic' Table
## Model: quasipoisson, link: log
## Response: response
## Marginal Testing
## Max Panel Size = 27; Number of panels = 280
##
##                      Df      X2 P(>|Chi|)
## as.factor(shipcatAv)  6  49.229 6.708e-09 ***
## as.factor(shipAIS10)  1  38.027 6.977e-10 ***
## as.factor(shipAIS5)   1  13.553 0.0002319 ***
## shiplenav             1  14.585 0.0001340 ***
## as.factor(windfarm)   1 104.271 < 2.2e-16 ***
## as.factor(anthrop)    1   2.691 0.1009355
```

```
## shiplennear           1   1.715 0.1903371
## as.factor(fish)       1   0.238 0.6259240
## s(depth)              4  85.218 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##      NumIter         Model       CV   meanCV    lowCV   highCV   p.value
## 2.5%       4 - s(shipnear) 15396.42 15585.35 15485.08 15750.34 0.2896249
##      numvalid
## 2.5%      100
```

The term *fish* was removed.

```
## Analysis of 'Wald statistic' Table
## Model: quasipoisson, link: log
## Response: response
## Marginal Testing
## Max Panel Size = 27; Number of panels = 280
##
##                      Df      X2 P(>|Chi|)
## as.factor(shipcatAv)  6  49.064 7.238e-09 ***
## as.factor(shipAIS10)  1  37.592 8.722e-10 ***
## as.factor(shipAIS5)   1  13.556 0.0002315 ***
## shiplenav             1  15.427 8.575e-05 ***
## as.factor(windfarm)   1 104.057 < 2.2e-16 ***
## as.factor(anthrop)    1   2.430 0.1190383
## shiplennear           1   1.811 0.1783990
## s(depth)              4  89.744 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##      NumIter             Model       CV   meanCV    lowCV   highCV
## 2.5%       5 - as.factor(fish) 15374.73 15519.72 15416.72 15653.13
##       p.value numvalid
## 2.5% 0.625924      100
```

The term *shiplennear* was excluded next.

```
## Analysis of 'Wald statistic' Table
## Model: quasipoisson, link: log
## Response: response
## Marginal Testing
## Max Panel Size = 27; Number of panels = 280
##
##                      Df      X2 P(>|Chi|)
## as.factor(shipcatAv)  6  50.871 3.144e-09 ***
## as.factor(shipAIS10)  1  43.770 3.693e-11 ***
## as.factor(shipAIS5)   1  13.772 0.0002064 ***
## shiplenav             1  18.201 1.987e-05 ***
## as.factor(windfarm)   1 107.661 < 2.2e-16 ***
## as.factor(anthrop)    1   2.051 0.1521507
## s(depth)              4 198.577 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##      NumIter         Model       CV   meanCV    lowCV   highCV  p.value
## 2.5%       6 - shiplennear 15310.39 15446.56 15342.86 15624.11 0.178399
```

```
##      numvalid
## 2.5%      100
```

The term *anthrop* was excluded next.

```
## Analysis of 'Wald statistic' Table
## Model: quasipoisson, link: log
## Response: response
## Marginal Testing
## Max Panel Size = 27; Number of panels = 280
##
##                      Df      X2 P(>|Chi|)
## as.factor(shipcatAv)  6  52.110 1.772e-09 ***
## as.factor(shipAIS10)  1  44.022 3.247e-11 ***
## as.factor(shipAIS5)   1  13.899 0.0001929 ***
## shiplenav             1  18.223 1.965e-05 ***
## as.factor(windfarm)   1 101.032 < 2.2e-16 ***
## s(depth)              4 200.416 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##      NumIter                Model       CV   meanCV    lowCV   highCV
## 2.5%       7 - as.factor(anthrop) 15307.72 15437.54 15338.07 15582.55
##        p.value numvalid
## 2.5% 0.1521507      100
```

All terms in the above model were significant.

Inclusion of a 2D term for location

Having selected all the factor and 1-dimensional terms, a 2-dimensional term for location was added.

```
## Analysis of 'Wald statistic' Table
## Model: quasipoisson, link: log
## Response: response
## Marginal Testing
## Max Panel Size = 27; Number of panels = 280
##
##                      Df      X2 P(>|Chi|)
## as.factor(shipcatAv)  6  45.846 3.178e-08 ***
## as.factor(shipAIS10)  1  36.596 1.454e-09 ***
## as.factor(shipAIS5)   1   9.971 0.0015903 **
## shiplenav             1  12.877 0.0003327 ***
## as.factor(windfarm)   1  90.457 < 2.2e-16 ***
## s(depth)              4 163.309 < 2.2e-16 ***
## s(x.pos, y.pos)       5  19.057 0.0018757 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##      NumIter            Model       CV   meanCV    lowCV   highCV
## 2.5%       8 + s(x.pos,y.pos) 15093.06 21018.73 15044.71 15443.56
##         p.value numvalid
## 2.5% 0.00187574      100
```

All terms in this model were significant and so none were excluded at this stage. The CV was as follows.

There were three shipping factor variables in the model which could potentially be combined with location in an interaction term. Each term was tried in turn and the CVs for each of these models are shown below. 'None' indicates the model with no interaction term fitted.

```
## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: algorithm did not converge
```

Variable	meanCV	lowCV	highCV	p.value
None	21019	15045	15444	0.001876
as.factor(shipcatAv)	8.124e+16	14578	1.204e+10	0
as.factor(shipAIS5)	14822	14584	15263	0.0003929
as.factor(shipAIS10)	15194	14988	15509	0.01796
shiplenav	15585	14794	15298	0.0001553

The model with the lowest CV included an interaction between location and $shipAIS5$ (see below).

```
## Analysis of 'Wald statistic' Table
## Model: quasipoisson, link: log
## Response: response
## Marginal Testing
## Max Panel Size = 27; Number of panels = 280
##
##                                      Df      X2 P(>|Chi|)
## as.factor(shipcatAv)                  6  57.096 1.747e-10 ***
## as.factor(shipAIS10)                  1  36.549 1.489e-09 ***
## as.factor(shipAIS5)                   1   5.533 0.0186635 *
## shiplenav                             1  11.335 0.0007608 ***
## as.factor(windfarm)                   1  64.302 1.110e-15 ***
## s(depth)                              4 137.945 < 2.2e-16 ***
## s(x.pos, y.pos)                       5  21.264 0.0007220 ***
## s(x.pos, y.pos):as.factor(shipAIS5)  5  22.655 0.0003929 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##      NumIter                     Model       CV   meanCV    lowCV   highCV
## 2.5%       9 + s(x.pos,y.pos):shipAIS5 14685.75 14791.29 14567.03 15122.72
##           p.value numvalid
## 2.5% 0.0003929289      100
```

All terms were significant. Steps in model selection are summarised as follows:

Table 4: Table continues below

	NumIter	Model	CV	meanCV	lowCV	highCV
2.5%	1	Full model	15319	15621	15465	15814
2.5%1	2	- salinity	15308	15579	15441	15760
2.5%2	3	- as.factor(LBspa)	15300	15575	15438	15757
2.5%3	4	- s(shipnear)	15396	15585	15485	15750
2.5%4	5	- as.factor(fish)	15375	15520	15417	15653

	NumIter	Model	CV	meanCV	lowCV	highCV
2.5%5	6	- shiplennear	15310	15447	15343	15624
2.5%6	7	- as.factor(anthrop)	15308	15438	15338	15583
2.5%7	8	+ s(x.pos,y.pos)	15093	21019	15045	15444
2.5%8	9	+ s(x.pos,y.pos):shipAIS5	14686	14791	14567	15123

	p.value	numvalid
2.5%	NA	100
2.5%1	0.7154	100
2.5%2	0.5052	100
2.5%3	0.2896	100
2.5%4	0.6259	100
2.5%5	0.1784	100
2.5%6	0.1522	100
2.5%7	0.001876	100
2.5%8	0.0003929	100

Selected model

A summary of the selected model is given below.

```
##
## Call:
## gamMRSea(formula = response ~ as.factor(shipcatAv) + as.factor(shipAIS10) +
##     as.factor(shipAIS5) + shiplenav + as.factor(windfarm) + bs(depth,
##     knots = splineParams[[2]]$knots, degree = splineParams[[2]]$degree,
##     Boundary.knots = splineParams[[2]]$bd) + LRF.g(radiusIndices,
##     dists, radii, aR) + as.factor(shipAIS5):LRF.g(radiusIndices,
##     dists, radii, aR) + offset(log(area)), family = quasipoisson(link = log),
##     data = all.seg, splineParams = splineParams)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -25.285   -6.038   -2.549   -0.392  108.063
##
## Coefficients:
##                                          Estimate Std. Error Robust S.E.
## (Intercept)                             13.318264  28.872264    8.160914
## as.factor(shipcatAv)1                    0.056355   0.140563    0.351325
## as.factor(shipcatAv)2                   -0.700252   0.698531    0.761293
## as.factor(shipcatAv)3                   -2.822715   1.743648    0.679977
## as.factor(shipcatAv)4                   -1.670900   0.899997    0.634943
## as.factor(shipcatAv)5                   -3.483316   1.830179    0.920977
## as.factor(shipcatAv)6                   -5.833697   4.026122    1.458324
## as.factor(shipAIS10)1                   -2.003240   0.353390    0.331358
## as.factor(shipAIS5)1                    10.752285   2.439754    4.571192
## shiplenav                                0.020970   0.019687    0.006229
## as.factor(windfarm)1                    -4.134402   3.027880    0.515584
## s(depth)1                              -33.909457  29.553585   10.467961
## s(depth)2                              -13.749900  28.833040    7.685294
## s(depth)3                              -11.906719  28.903388    8.380264
```

```
## s(depth)4                               -21.962187  28.769228    7.744709
## s(x.pos, y.pos)b1                          6.008795   1.447492    2.862600
## s(x.pos, y.pos)b2                          5.436844   1.270566    2.855823
## s(x.pos, y.pos)b3                          5.141147   2.927681    5.038923
## s(x.pos, y.pos)b4                          4.026986   1.180155    2.219774
## s(x.pos, y.pos)b5                         -3.944780   2.942156    5.190429
## as.factor(shipAIS5)1:s(x.pos, y.pos)b1 -13.258509   2.937851    5.340090
## as.factor(shipAIS5)1:s(x.pos, y.pos)b2 -13.469510   2.843769    5.708004
## as.factor(shipAIS5)1:s(x.pos, y.pos)b3 -30.976502   6.525914   12.323147
## as.factor(shipAIS5)1:s(x.pos, y.pos)b4  -6.374550   2.526031    4.519662
## as.factor(shipAIS5)1:s(x.pos, y.pos)b5  32.924009   6.337933   11.881839
##                                         t value Pr(>|t|)
## (Intercept)                               1.632 0.102767
## as.factor(shipcatAv)1                     0.160 0.872569
## as.factor(shipcatAv)2                    -0.920 0.357723
## as.factor(shipcatAv)3                    -4.151 3.38e-05 ***
## as.factor(shipcatAv)4                    -2.632 0.008532 **
## as.factor(shipcatAv)5                    -3.782 0.000158 ***
## as.factor(shipcatAv)6                    -4.000 6.44e-05 ***
## as.factor(shipAIS10)1                    -6.046 1.62e-09 ***
## as.factor(shipAIS5)1                      2.352 0.018712 *
## shiplenav                                 3.367 0.000768 ***
## as.factor(windfarm)1                     -8.019 1.39e-15 ***
## s(depth)1                                -3.239 0.001208 **
## s(depth)2                                -1.789 0.073671 .
## s(depth)3                                -1.421 0.155451
## s(depth)4                                -2.836 0.004594 **
## s(x.pos, y.pos)b1                         2.099 0.035873 *
## s(x.pos, y.pos)b2                         1.904 0.057011 .
## s(x.pos, y.pos)b3                         1.020 0.307654
## s(x.pos, y.pos)b4                         1.814 0.069730 .
## s(x.pos, y.pos)b5                        -0.760 0.447293
## as.factor(shipAIS5)1:s(x.pos, y.pos)b1  -2.483 0.013075 *
## as.factor(shipAIS5)1:s(x.pos, y.pos)b2  -2.360 0.018334 *
## as.factor(shipAIS5)1:s(x.pos, y.pos)b3  -2.514 0.011987 *
## as.factor(shipAIS5)1:s(x.pos, y.pos)b4  -1.410 0.158498
## as.factor(shipAIS5)1:s(x.pos, y.pos)b5   2.771 0.005615 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 237.1477)
##
##     Null deviance: 454836  on 4038  degrees of freedom
## Residual deviance: 309430  on 4014  degrees of freedom
## AIC:  NA
##
## Max Panel Size = 27; Number of panels = 280
## Number of Fisher Scoring iterations: 10
```

Diagnostics

The diagnostics of this selected model were investigated to ensure that the model was valid.

The diagnostics for this model are as follows:

- Figure C3 shows the position of the fitted knots for the 2D smooth term.
- Figure C4 indicates that the blocking structure was appropriate, the correlation in all blocks declined to zero as the distance increased.
- Figure C5 shows the partial plots on the scale of the link function. Values above/below zero on the y-axis indicate that the number of birds would increase/decrease at that value of the covariate shown by the x-axis (given values for other covariates remain the same).
- Figure C6 shows the predicted values obtained from the model for the survey data. Predictions for the whole of the prediction grid are shown in the main report.
- Figure C7 shows the residuals from the model. The large residuals occurred where there were very high observed values.

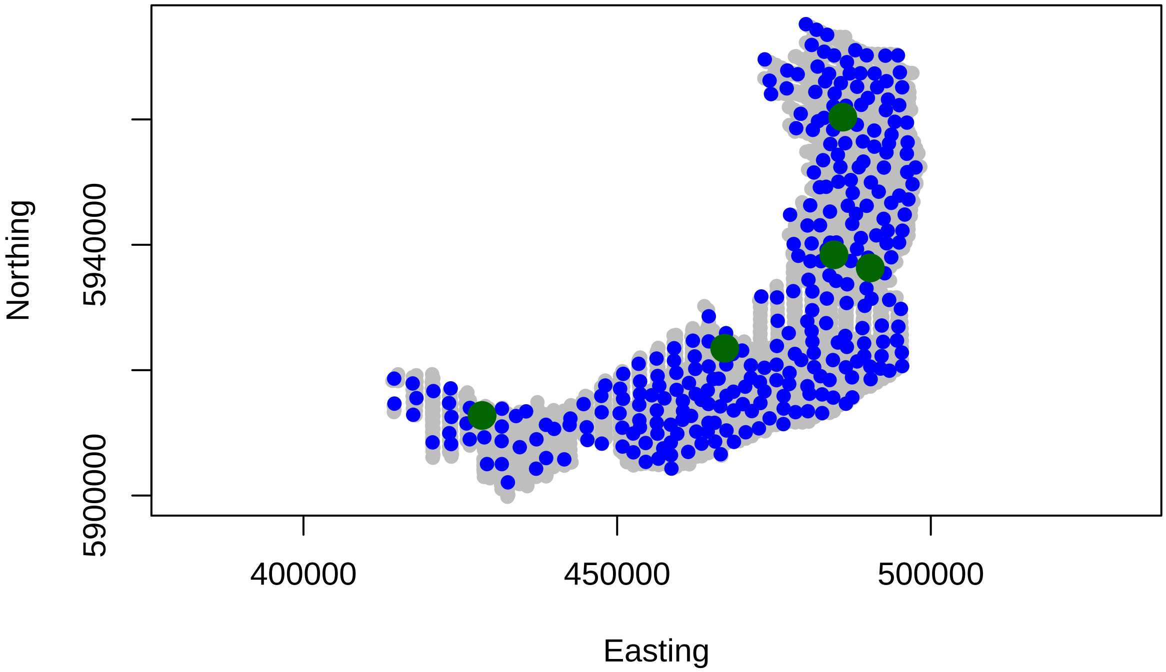


Figure C3. Location of the knots in the chosen model.

A runs test was conducted to determine whether the residuals were correlated. The significant test result (shown below) indicated that the residuals were correlated and so a blocking structure was required during the model fitting.

```
## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: algorithm did not converge

## Warning: glm.fit: algorithm did not converge
##
##  Runs Test - Two sided; Empirical Distribution
```

```
##
## data:  residuals(css.2dOutput$bestModel, type = "pearson")
## Standardized Runs Statistic = -28.751, p-value < 2.2e-16
```

The ACF plot indicated that the blocking structure (transects) was appropriate because the correlation declined to zero as the distance, or lag, between segments in the same block increased.

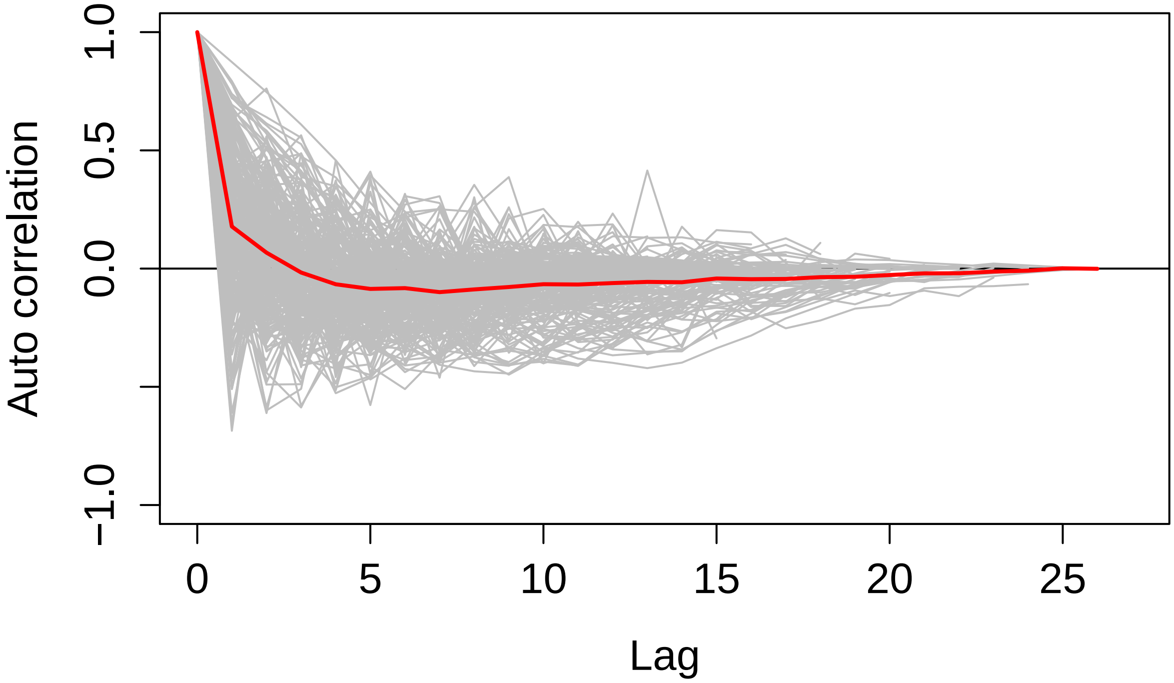
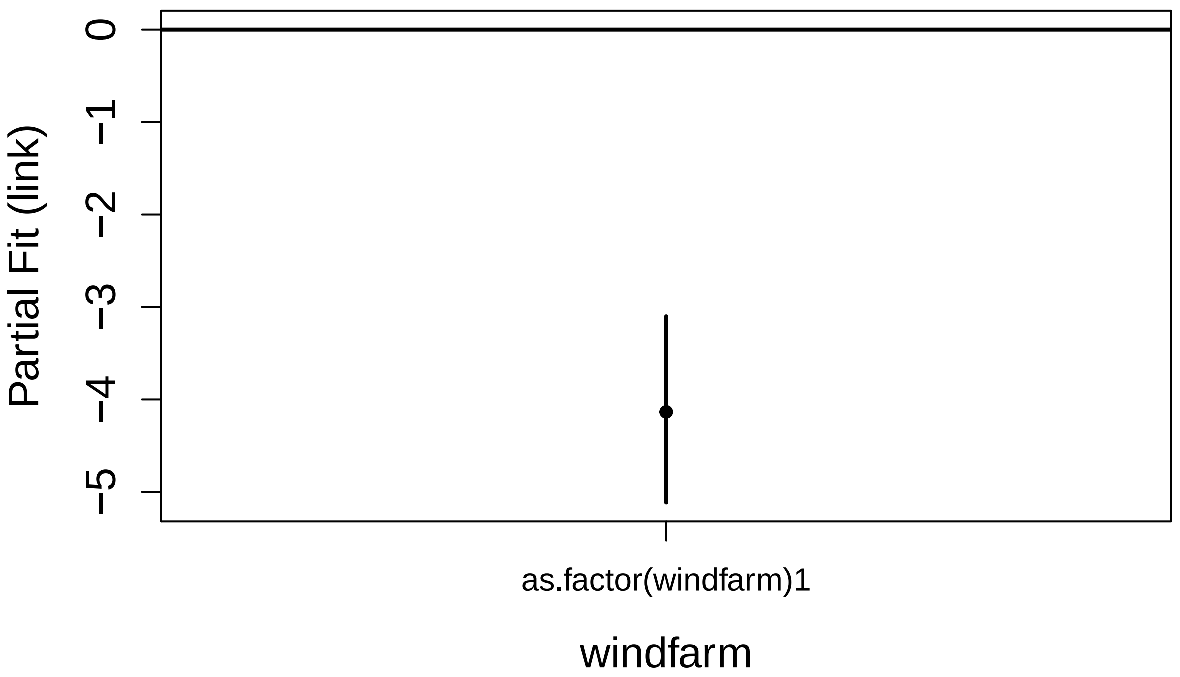


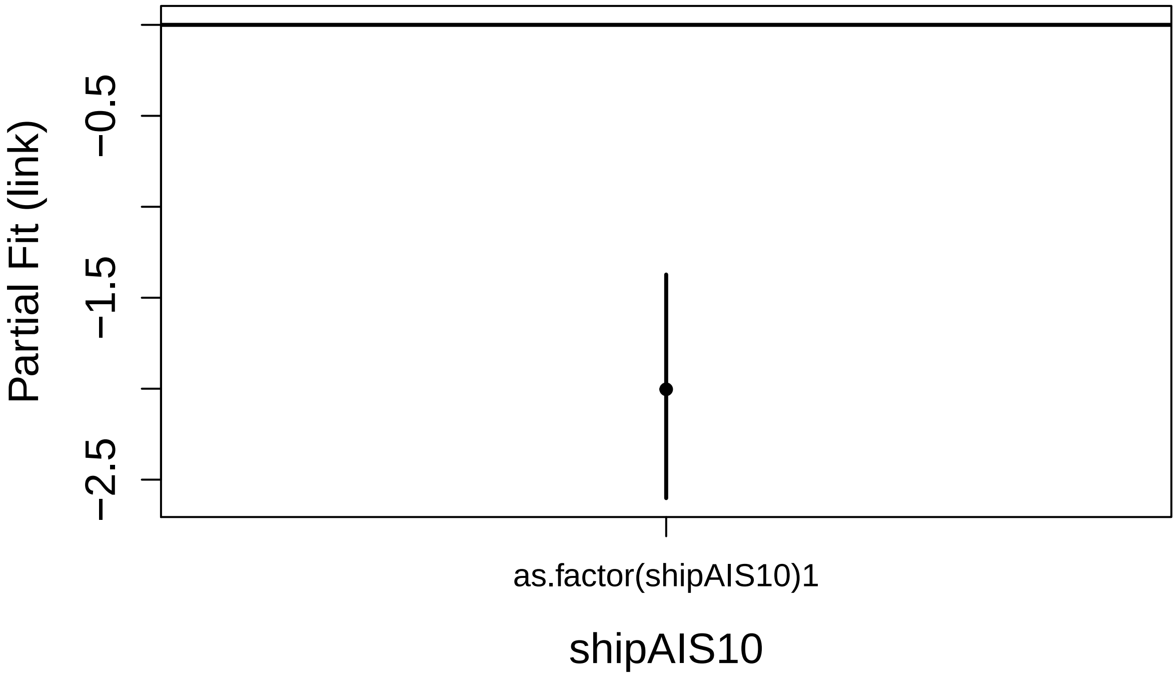
Figure C4. Plot of the correlation in the residuals for each block (grey lines). The mean correlation at each lag is indicated in red.

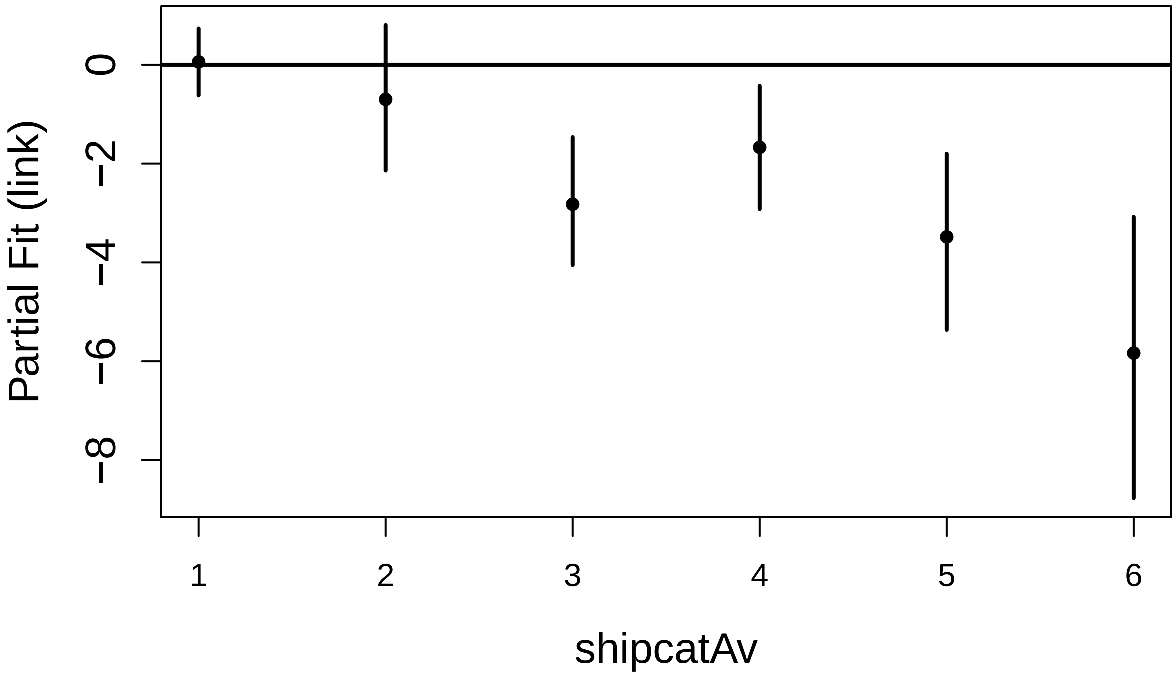
```
## [1] "Making partial plots"
```

```
## Loading required package: mvtnorm
```

```
## Loading required package: Matrix
```


0
−1
−2
−3
−4
−5
Partial Fit (link)
as.factor(windfarm)1
windfarm


Partial Fit (link)
−0.5
−1.5
−2.5
as.factor(shipAIS10)1
shipAIS10



Partial Fit (link)
0
−2
−4
−6
−8
1
2
3
4
5
6
shipcatAv

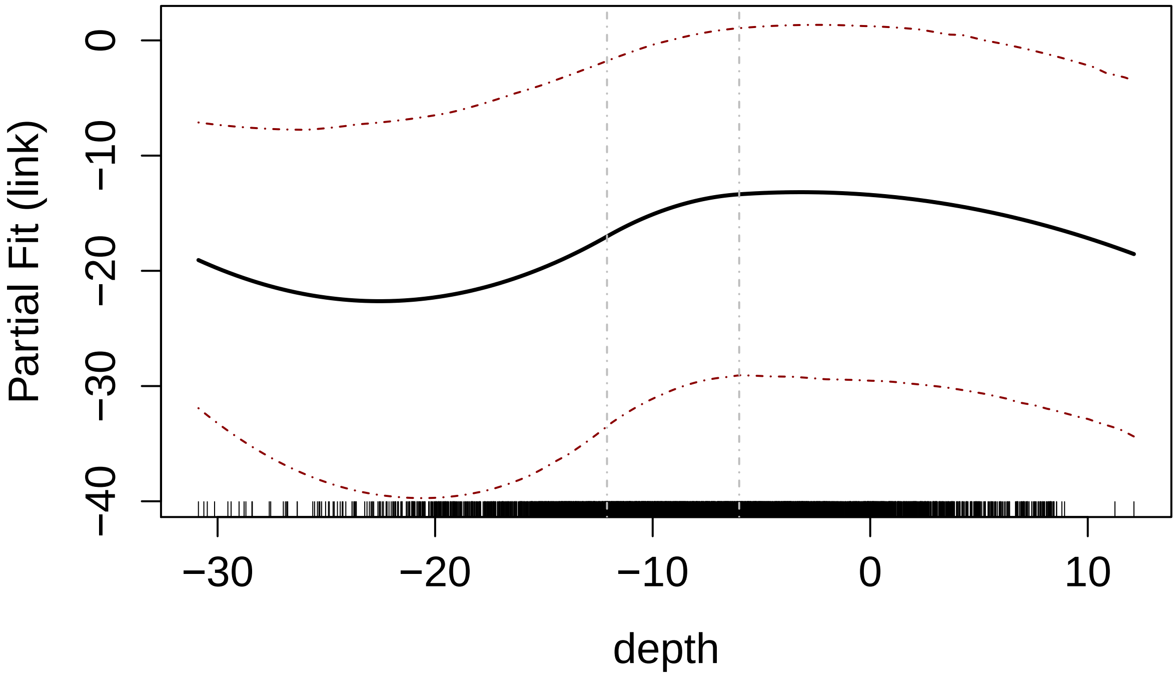


Figure C5. Partial plots of the terms in the model (on a logarithmic scale). For factor terms, level 0 is used as a baseline, or reference, level.

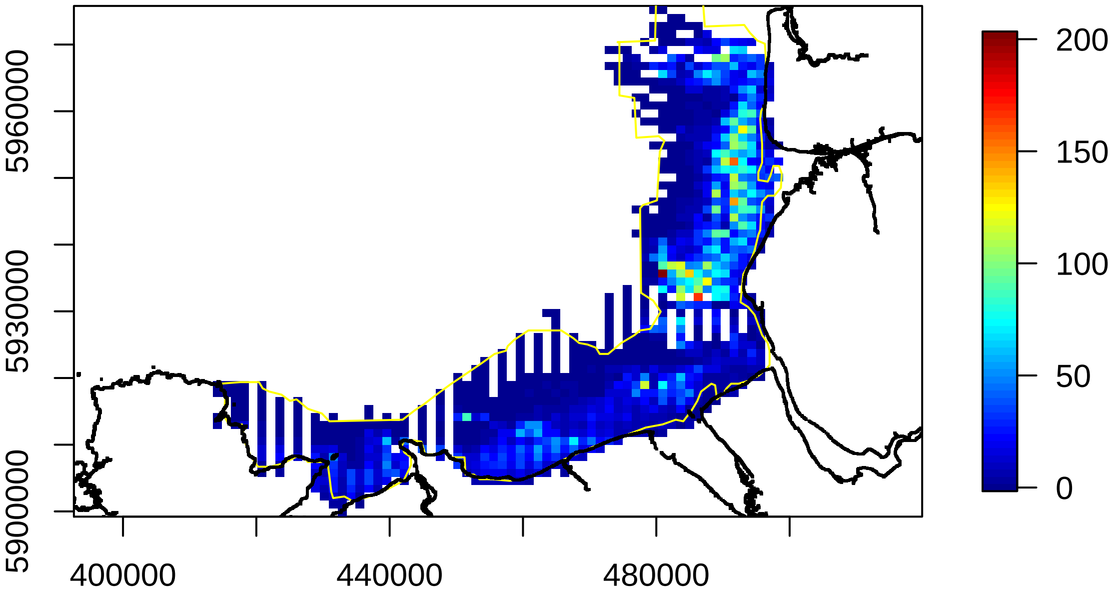


Figure C6. Plot of the predicted values (numbers of CS sitting on water) for the survey data and averaged over all surveys. The yellow line indicates the approximate boundary of Liverpool Bay SPA.

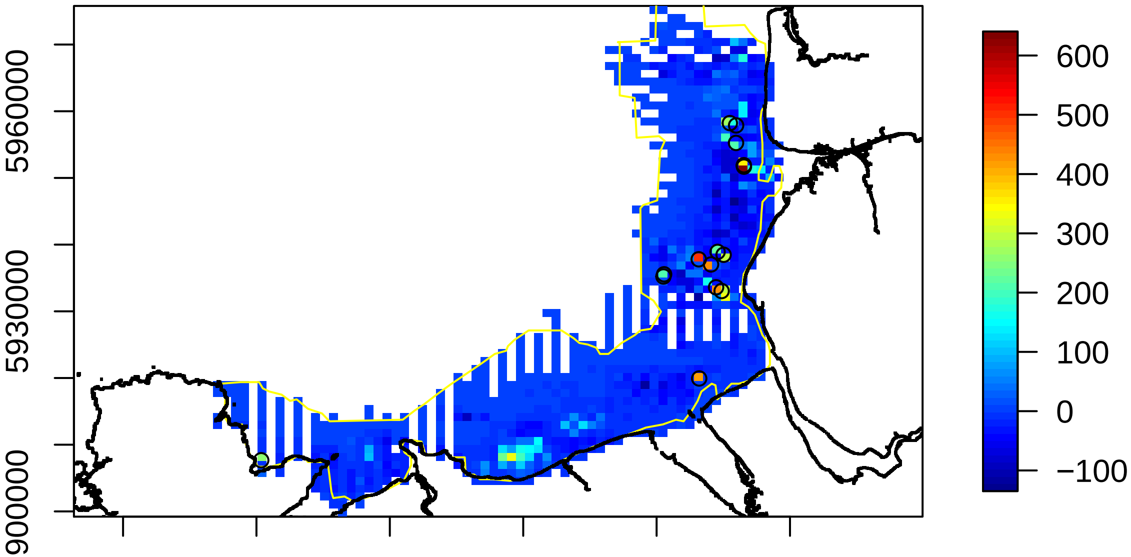
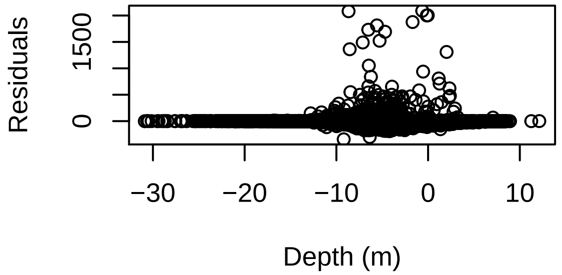
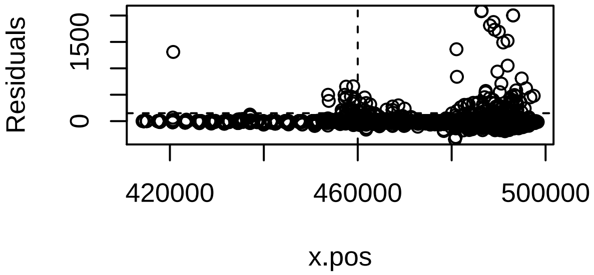
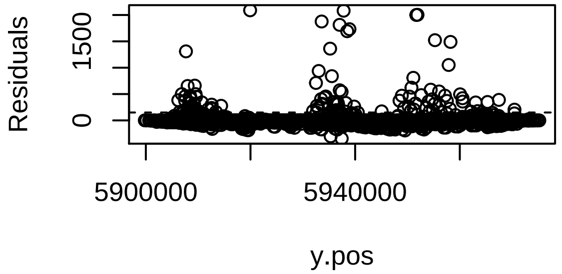


Figure C7. Plot of the residuals (difference between the observed number of birds and predicted number) averaged over all surveys. The large positive residuals (>200) occurred in segments where large numbers of birds (>1000) were observed. The yellow line indicates the approximate boundary of Liverpool Bay SPA.

To investigate the patch of residuals that were similar, then residuals were plotted against the explanatory variables. The dashed lines approximately indicate the 'locations' of the 'patch' of residuals.


Residuals
0
1500
−30
−20
−10
0
10
Depth (m)


Residuals
0
1500
420000
460000
500000
x.pos


Residuals
0
1500
5900000
5940000
y.pos

Appendix D. Model selection for Common Scoter identified as flying

This document describes the model selection for CS identified as flying.

Due to a few, very large values of CS in a segment (Figure D1), the number per segment was capped to 110 birds.

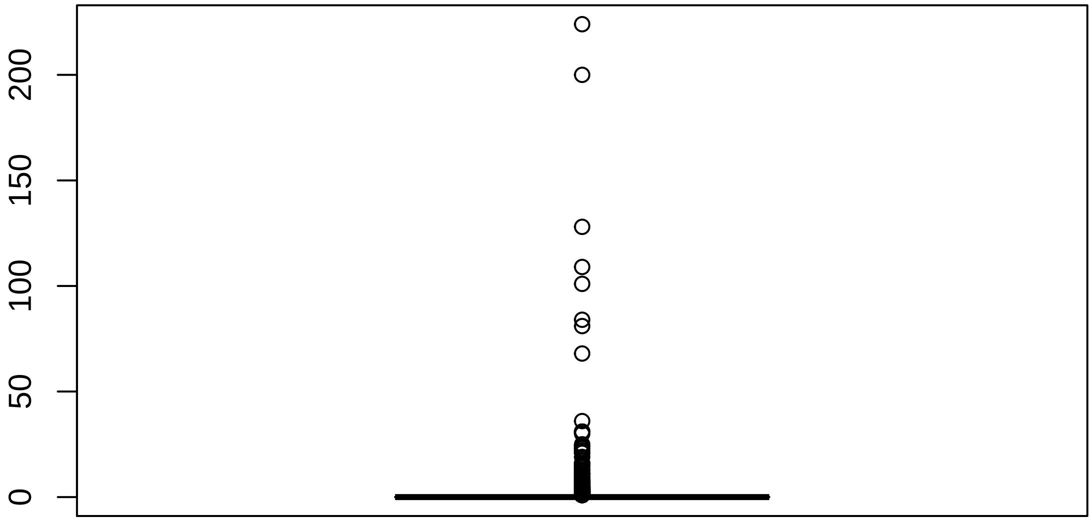


Figure D1. The distribution of the numbers of CS identified as flying per segment.

Assessing candidate explanatory variables

Each explanatory variable was fitted separately to determine their likely importance and to decide which variables to select for terms which can not be fitted in the same model (Table D1).

Table D1. The CV scores and percentile-based CI, pseudo-R^2 (R2, a measure of the correlation between the observed values and the fitted values from the model) and probability (p.val) associated with fitting each term seperately. Numbers in the Group column indicate variables which are grouped together and one variable from each group was chosen.

```
                Variable   meanCV    lowCV   highCV          R2      p.val
17              s(depth) 22.09740 22.05122 22.17153 0.016974603 0.00000000
13  as.factor(shipAIS10) 22.21160 22.19701 22.22889 0.008706787 0.22683026
5    as.factor(windfarm) 22.22053 22.20873 22.23720 0.006916032 0.04711510
2              shiplenav 22.22268 22.21102 22.23957 0.007063417 0.13940676
6     as.factor(anthrop) 22.22517 22.21426 22.24362 0.006649672 0.08875194
7   as.factor(windcable) 22.22993 22.21982 22.24770 0.006386762 0.16306920
14    as.factor(shipMMO) 22.24280 22.22736 22.26415 0.006610358 0.62170874
12   as.factor(shipAIS5) 22.24828 22.23365 22.26982 0.006415953 0.73611007
```

```
1           shiplennear 22.25121 22.23922 22.26949 0.006151314 0.65083232
9    as.factor(shipAIS1) 22.25206 22.23966 22.27023 0.005974084 0.98602508
10  as.factor(shipcatAv) 22.25221 22.22756 22.32062 0.008710965 0.01768984
8    as.factor(shipnumf) 22.25397 22.24100 22.27320 0.005920368 0.55728187
4        as.factor(fish) 22.25417 22.23907 22.27404 0.006060635 0.54153812
11 as.factor(shipcatMax) 22.25553 22.23013 22.31877 0.008722463 0.01673501
15         s(shiplenmax) 22.26318 22.23909 22.29450 0.006508916 0.51077440
3       as.factor(LBspa) 22.28496 22.27464 22.30172 0.006028818 0.96950622
18           s(salinity) 22.29111 22.25482 22.34033 0.007214814 0.30703982
16           s(shipnear) 22.32110 22.27436 22.39061 0.006479560 0.19786913
   Group
17
13
5
2
6
7
14
12
1
9      1
10     1
8      1
4
11     1
15
3
18
16
```

The variable chosen from group 1 based on the lowest CV score in the group of variables was *shipAIS*1. However, this would be excluded at the next stage based on the GVIF score (see below). Therefore, the variable in group 1 with the next lowest CV was selected (i.e. *shipcatAv*). This variable also had a significant *p*-value compared to *shipAIS*1.

Table 1: Table continues below

as.factor(shipAIS10)	as.factor(windfarm)	as.factor(anthrop)
3.788	1.003	218621

Table 2: Table continues below

as.factor(windcable)	as.factor(shipMMO)	as.factor(shipAIS5)
218621	1.703	3.943

Table 3: Table continues below

as.factor(shipAIS1)	as.factor(fish)	as.factor(LBspa)	depth	shiplenav
6.749	1.242	1.183	1.95	26.68

shiplennear	shiplenmax	salinity	shipnear
1.789	16.61	1.388	1.766

Checking for collinearity

```
# Fit chosen factors

csf.linear <- glm(response ~ as.factor(shipAIS10) + as.factor(windfarm) + as.factor(anthrop) + as.facto

csf.vifs2 <- vif(csf.linear)
save(csf.vifs2,file="csf_vifs2.RData")

pander::pander(csf.vifs2)
```

	GVIF	Df	GVIF^(1/(2*Df))
as.factor(shipAIS10)	3.835	1	1.958
as.factor(windfarm)	1.004	1	1.002
as.factor(anthrop)	219340	1	468.3
as.factor(windcable)	219340	1	468.3
as.factor(shipMMO)	91.63	1	9.572
as.factor(shipAIS5)	3.979	1	1.995
as.factor(shipcatAv)	4002	6	1.996
as.factor(fish)	1.27	1	1.127
as.factor(LBspa)	1.181	1	1.087
depth	1.964	1	1.402
shiplenav	15.61	1	3.951
shiplennear	1.807	1	1.344
shiplenmax	22.94	1	4.789
salinity	1.386	1	1.177
shipnear	1.763	1	1.328

The collinearity of the terms in the mode was checked. Variance inflation factors (vifs) of >5 indicate collinearity - there were a few values indicating this.

- *windcable* and *anthrop* - *anthrop* chosen based on CV score above and *windcable* excluded.
- *shiplenav* and *shiplenmax* - these values are high but not >5 and so at present will be retained.
- *shipMMO* - excluded

Selection of the factor and 1D continuous terms

The starting model was as follows, with *shiplenav* and *shiplennear* in as linear terms due to problems when assessing terms individually in a model. The term *shiplenmax* was also included as a linear term due to warning messages when including it as a smooth term (eben with maximum knots = 1). The maximum number of knots for salinity was reduced to 1 (again due to problems when assessing terms individually); 2 for other smooth terms.

```
Analysis of 'Wald statistic' Table
Model: quasipoisson, link: log
Response: response
```

```
Marginal Testing
Max Panel Size = 27; Number of panels = 280

                     Df      X2 P(>|Chi|)
as.factor(shipAIS10)  1   0.298 0.5853520
as.factor(windfarm)   1  16.425 5.061e-05 ***
as.factor(anthrop)    1  15.215 9.595e-05 ***
as.factor(shipcatAv)  6  12.330 0.0550090 .
as.factor(shipAIS5)   1   0.025 0.8744315
as.factor(fish)       1   4.624 0.0315343 *
as.factor(LBspa)      1   0.817 0.3659635
shiplenav             1   1.307 0.2530111
shiplenmax            1   0.260 0.6101543
shiplennear           1   0.180 0.6717366
s(depth)              4 138.194 < 2.2e-16 ***
s(shipnear)           4   9.763 0.0446240 *
s(salinity)           3  20.676 0.0001229 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
     NumIter      Model          CV       meanCV    lowCV      highCV
2.5%       1 Full model 51167334062 3.096055e+25 22.40848 3.07761e+20
     p.value numvalid
2.5%      NA      100
```

The term *shipAIS*5 was first to be excluded.

```
Analysis of 'Wald statistic' Table
Model: quasipoisson, link: log
Response: response
Marginal Testing
Max Panel Size = 27; Number of panels = 280

                     Df      X2 P(>|Chi|)
as.factor(shipAIS10)  1   0.916 0.3385435
as.factor(windfarm)   1  15.644 7.646e-05 ***
as.factor(anthrop)    1  15.364 8.867e-05 ***
as.factor(shipcatAv)  6  12.208 0.0574948 .
as.factor(fish)       1   4.946 0.0261506 *
as.factor(LBspa)      1   0.820 0.3653131
shiplenav             1   1.319 0.2507841
shiplenmax            1   0.259 0.6105207
shiplennear           1   0.207 0.6494252
s(depth)              4 162.804 < 2.2e-16 ***
s(shipnear)           4  10.259 0.0362775 *
s(salinity)           3  20.497 0.0001339 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
     NumIter                 Model          CV       meanCV    lowCV
2.5%       1 - as.factor(shipAIS5) 38526957499 2.215245e+24 22.29869
           highCV   p.value numvalid
2.5% 2.680513e+20 0.8744315      100
```

The term *shiplennear* was the next to be excluded.

```
Analysis of 'Wald statistic' Table
```

```
Model: quasipoisson, link: log
Response: response
Marginal Testing
Max Panel Size = 27; Number of panels = 280

                     Df      X2 P(>|Chi|)
as.factor(shipAIS10)  1   0.212   0.64553
as.factor(windfarm)   1  18.393 1.797e-05 ***
as.factor(anthrop)    1  15.601 7.821e-05 ***
as.factor(shipcatAv)  6  12.647   0.04900 *
as.factor(fish)       1   2.758   0.09674 .
as.factor(LBspa)      1   1.030   0.31019
shiplenav             1   1.177   0.27799
shiplenmax            1   0.585   0.44436
s(depth)              4 251.940 < 2.2e-16 ***
s(shipnear)           4   6.633   0.15658
s(salinity)           3  54.135 1.050e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
     NumIter         Model           CV       meanCV    lowCV       highCV
2.5%       3 - shiplennear 2.175574e+13 4.443243e+26 426.0062 5.860032e+22
       p.value numvalid
2.5% 0.6494252      100
```

The term *shipAIS*10 went next.

```
Analysis of 'Wald statistic' Table
Model: quasipoisson, link: log
Response: response
Marginal Testing
Max Panel Size = 27; Number of panels = 280

                     Df      X2 P(>|Chi|)
as.factor(windfarm)   1  17.725 2.552e-05 ***
as.factor(anthrop)    1  14.881 0.0001145 ***
as.factor(shipcatAv)  6  11.889 0.0644809 .
as.factor(fish)       1   5.200 0.0225836 *
as.factor(LBspa)      1   0.842 0.3587779
shiplenav             1   1.305 0.2533588
shiplenmax            1   0.267 0.6051072
s(depth)              4 145.626 < 2.2e-16 ***
s(shipnear)           4  10.520 0.0325227 *
s(salinity)           3  19.960 0.0001730 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
     NumIter                  Model           CV       meanCV    lowCV
2.5%       4 - as.factor(shipAIS10) 530296895992 2.341029e+27 22.54333
           highCV   p.value numvalid
2.5% 5.378384e+22 0.6455338      100
```

The term *shiplenmax* went next.

```
Analysis of 'Wald statistic' Table
Model: quasipoisson, link: log
Response: response
```

```
Marginal Testing
Max Panel Size = 27; Number of panels = 280

                     Df      X2 P(>|Chi|)
as.factor(windfarm)   1  17.502  2.87e-05 ***
as.factor(anthrop)    1  14.773 0.0001213 ***
as.factor(shipcatAv)  6  12.085 0.0601048 .
as.factor(fish)       1   5.174 0.0229278 *
as.factor(LBspa)      1   0.823 0.3642874
shiplenav             1   1.560 0.2115924
s(depth)              4 144.801 < 2.2e-16 ***
s(shipnear)           4  10.052 0.0395565 *
s(salinity)           3  19.953 0.0001736 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
     NumIter        Model           CV      meanCV    lowCV       highCV
2.5%       5 - shiplenmax 25498583305 1.86865e+27 23.71814 1.172106e+23
       p.value numvalid
2.5% 0.6051072      100
```

The term *LBspa* went next.

```
Analysis of 'Wald statistic' Table
Model: quasipoisson, link: log
Response: response
Marginal Testing
Max Panel Size = 27; Number of panels = 280

                     Df      X2 P(>|Chi|)
as.factor(windfarm)   1  14.613 0.0001320 ***
as.factor(anthrop)    1  15.976 6.416e-05 ***
as.factor(shipcatAv)  6  14.385 0.0256223 *
as.factor(fish)       1   4.462 0.0346498 *
shiplenav             1   1.906 0.1673742
s(depth)              4 192.249 < 2.2e-16 ***
s(shipnear)           4   7.315 0.1201282
s(salinity)           3  18.181 0.0004036 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
     NumIter              Model           CV       meanCV    lowCV
2.5%       6 - as.factor(LBspa) 157609609448 7.565239e+27 22.51357
           highCV   p.value numvalid
2.5% 1.768084e+25 0.3642874      100
```

The term *shiplenav* went next.

```
Analysis of 'Wald statistic' Table
Model: quasipoisson, link: log
Response: response
Marginal Testing
Max Panel Size = 27; Number of panels = 280

                     Df      X2 P(>|Chi|)
as.factor(windfarm)   1  26.400 2.775e-07 ***
as.factor(anthrop)    1  17.895 2.334e-05 ***
```

```
as.factor(shipcatAv)  6  35.600 3.297e-06 ***
as.factor(fish)       1   5.475   0.01929 *
s(depth)              4 146.518 < 2.2e-16 ***
s(shipnear)           4  13.102   0.01079 *
s(salinity)           3  39.969 1.082e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
     NumIter       Model       CV   meanCV    lowCV   highCV   p.value
2.5%       7 - shiplenav 21.92364 22.22604 21.94831 22.70061 0.1673742
     numvalid
2.5%      100
```

Inclusion of a 2D term for location

A two-dimensional term for location was added.

```
Analysis of 'Wald statistic' Table
Model: quasipoisson, link: log
Response: response
Marginal Testing
Max Panel Size = 27; Number of panels = 280

                     Df     X2 P(>|Chi|)
as.factor(windfarm)   1 21.413 3.703e-06 ***
as.factor(anthrop)    1 13.061 0.0003016 ***
as.factor(shipcatAv)  6 44.079 7.131e-08 ***
as.factor(fish)       1  0.637 0.4247131
s(depth)              4 95.110 < 2.2e-16 ***
s(shipnear)           4 31.199 2.789e-06 ***
s(salinity)           3 15.696 0.0013091 **
s(x.pos, y.pos)       5 60.936 7.784e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
     NumIter            Model       CV   meanCV    lowCV   highCV
2.5%       8 + s(x.pos,y.pos) 20.11574 20.33893 19.80832 21.21387
          p.value numvalid
2.5% 7.783996e-12      100
```

The term $fish$ is no longer significant and so this was excluded.

```
Analysis of 'Wald statistic' Table
Model: quasipoisson, link: log
Response: response
Marginal Testing
Max Panel Size = 27; Number of panels = 280

                     Df      X2 P(>|Chi|)
as.factor(windfarm)   1  21.511 3.518e-06 ***
as.factor(anthrop)    1  10.670  0.001089 **
as.factor(shipcatAv)  6  46.205 2.695e-08 ***
s(depth)              4 121.675 < 2.2e-16 ***
s(shipnear)           3  13.394  0.003858 **
s(salinity)           3  10.521  0.014622 *
s(x.pos, y.pos)       5  53.317 2.895e-10 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
     NumIter              Model       CV   meanCV    lowCV   highCV
2.5%       9 - as.factor(fish) 20.05771 20.34079 19.81753 21.34348
       p.value numvalid
2.5% 0.4247131      100
```

Check whether an interaction with *shipcatAv* is required.

There was a problem fitting an interaction term. A model with just the terms for the interaction were fitted to check whether an interaction term is possible to be fitted.

This also created an error and so an interaction will not be included.

The steps in the model selection are summarised below.

Table 6: Table continues below

	NumIter	Model	CV	meanCV	lowCV
2.5%	1	Full model	5.117e+10	3.096e+25	22.41
2.5%1	1	- as.factor(shipAIS5)	3.853e+10	2.215e+24	22.3
2.5%2	3	- shiplennear	2.176e+13	4.443e+26	426
2.5%3	4	- as.factor(shipAIS10)	5.303e+11	2.341e+27	22.54
2.5%4	5	- shiplenmax	2.55e+10	1.869e+27	23.72
2.5%5	6	- as.factor(LBspa)	1.576e+11	7.565e+27	22.51
2.5%6	7	- shiplenav	21.92	22.23	21.95
2.5%7	8	+ s(x.pos,y.pos)	20.12	20.34	19.81
2.5%8	9	- as.factor(fish)	20.06	20.34	19.82

	highCV	p.value	numvalid
2.5%	3.078e+20	NA	100
2.5%1	2.681e+20	0.8744	100
2.5%2	5.86e+22	0.6494	100
2.5%3	5.378e+22	0.6455	100
2.5%4	1.172e+23	0.6051	100
2.5%5	1.768e+25	0.3643	100
2.5%6	22.7	0.1674	100
2.5%7	21.21	7.784e-12	100
2.5%8	21.34	0.4247	100

Selected model

The anova table for the selected model is shown below.

```
Analysis of 'Wald statistic' Table
Model: quasipoisson, link: log
Response: response
Marginal Testing
Max Panel Size = 27; Number of panels = 280

                     Df      X2 P(>|Chi|)
as.factor(windfarm)   1  21.511 3.518e-06 ***
as.factor(anthrop)    1  10.670  0.001089 **
```

```
as.factor(shipcatAv)  6  46.205 2.695e-08 ***
s(depth)              4 121.675 < 2.2e-16 ***
s(shipnear)           3  13.394  0.003858 **
s(salinity)           3  10.521  0.014622 *
s(x.pos, y.pos)       5  53.317 2.895e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The summary of the selected model is given below.

```
Call:
gamMRSea(formula = response ~ as.factor(windfarm) + as.factor(anthrop) +
    as.factor(shipcatAv) + bs(depth, knots = splineParams[[2]]$knots,
    degree = splineParams[[2]]$degree, Boundary.knots = splineParams[[2]]$bd) +
    bs(shipnear, knots = splineParams[[3]]$knots, degree = splineParams[[3]]$degree,
        Boundary.knots = splineParams[[3]]$bd) + bs(salinity,
    knots = splineParams[[4]]$knots, degree = splineParams[[4]]$degree,
    Boundary.knots = splineParams[[4]]$bd) + LRF.g(radiusIndices,
    dists, radii, aR) + offset(log(area)), family = quasipoisson(link = log),
    data = all.seg, splineParams = splineParams)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-5.9511  -0.8202  -0.3732  -0.1241  27.3039

Coefficients:
                      Estimate Std. Error Robust S.E. t value Pr(>|t|)
(Intercept)           -52.5449    51.8852     34.2973  -1.532  0.12559
as.factor(windfarm)1   -3.9394     2.9975      0.8494  -4.638 3.63e-06 ***
as.factor(anthrop)1    -1.7921     1.1476      0.5486  -3.266  0.00110 **
as.factor(shipcatAv)1   0.8470     0.2715      0.5540   1.529  0.12640
as.factor(shipcatAv)2   1.6133     0.5219      1.0057   1.604  0.10876
as.factor(shipcatAv)3  -1.2903     1.2938      0.9935  -1.299  0.19413
as.factor(shipcatAv)4  -1.6438     1.4622      0.9145  -1.798  0.07233 .
as.factor(shipcatAv)5  -1.6577     1.7788      0.8565  -1.935  0.05300 .
as.factor(shipcatAv)6  -2.8583     3.0161      1.2016  -2.379  0.01742 *
s(depth)1              23.0364    55.3562     37.2627   0.618  0.53647
s(depth)2              37.7205    51.5714     33.6150   1.122  0.26187
s(depth)3              39.0076    51.9206     33.8904   1.151  0.24980
s(depth)4              35.6818    51.2410     33.4927   1.065  0.28678
s(shipnear)1           -2.2546     0.5642      0.7991  -2.821  0.00480 **
s(shipnear)2            2.5694     1.0391      1.6702   1.538  0.12403
s(shipnear)3           -0.7085     2.3369      2.6808  -0.264  0.79158
s(salinity)1           -0.7404     1.9513      2.6695  -0.277  0.78154
s(salinity)2            3.6743     1.3768      1.8044   2.036  0.04178 *
s(salinity)3           -1.3098   744.9676      2.2803  -0.574  0.56573
s(x.pos, y.pos)b1      18.3500     6.1885      6.2948   2.915  0.00358 **
s(x.pos, y.pos)b2      11.7464     2.4711      2.6014   4.515 6.50e-06 ***
s(x.pos, y.pos)b3     -18.4286     9.1371     10.3946  -1.773  0.07632 .
s(x.pos, y.pos)b4      -1.7947     0.8814      1.0955  -1.638  0.10145
s(x.pos, y.pos)b5      19.7299     7.4030      7.3780   2.674  0.00752 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for quasipoisson family taken to be 17.82734)

    Null deviance: 13731.0  on 4038  degrees of freedom
Residual deviance:  9175.7  on 4015  degrees of freedom
AIC:  NA

Max Panel Size = 27; Number of panels = 280
Number of Fisher Scoring iterations: 12
```

Diagnostics

The diagnostics of this selected model were investigated to ensure that the model was valid.

The diagnostics for this model were as follows:

- Figure D3 shows the position of the fitted knots for the 2D smooth term.
- Figure D4 indicates that the blocking structure was appropriate, the correlation in all blocks declines to zero.
- Figure D5 shows the partial plots on the scale of the link function.
- Figure D6 shows the fitted values for the survey data. Predictions for the whole of the prediction grid are shown in a separate document.
- Figure D7 shows the residuals (observed numbers - predicted numbers). The large residuals occurred where there were high observed values.

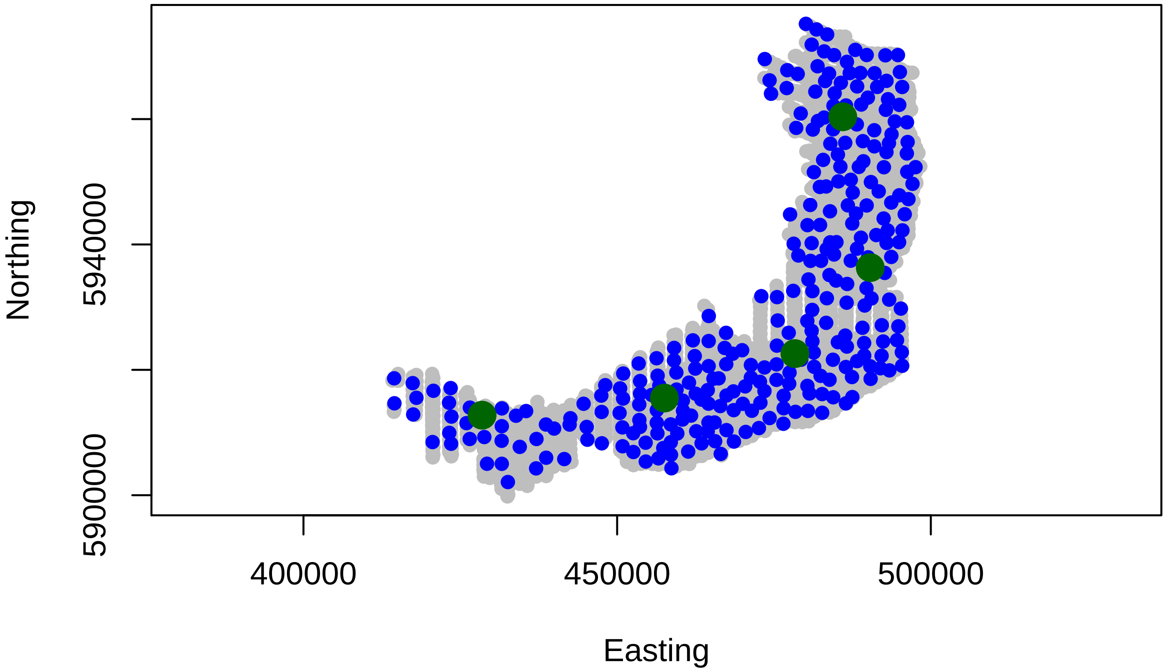


Figure D3. Location of the knots in the chosen model.

The runs test was performed to determine if the residuals were correlated. The *p*-value associated with this test did not indicate that the residuals were correlated. Thus, the standard errors and robust standard errors shown in the model summary above were very similar.

```
	Runs Test - Two sided; Empirical Distribution

data:  residuals(csf.2dOutput$bestModel, type = "pearson")
Standardized Runs Statistic = -29.569, p-value = 0.2
```

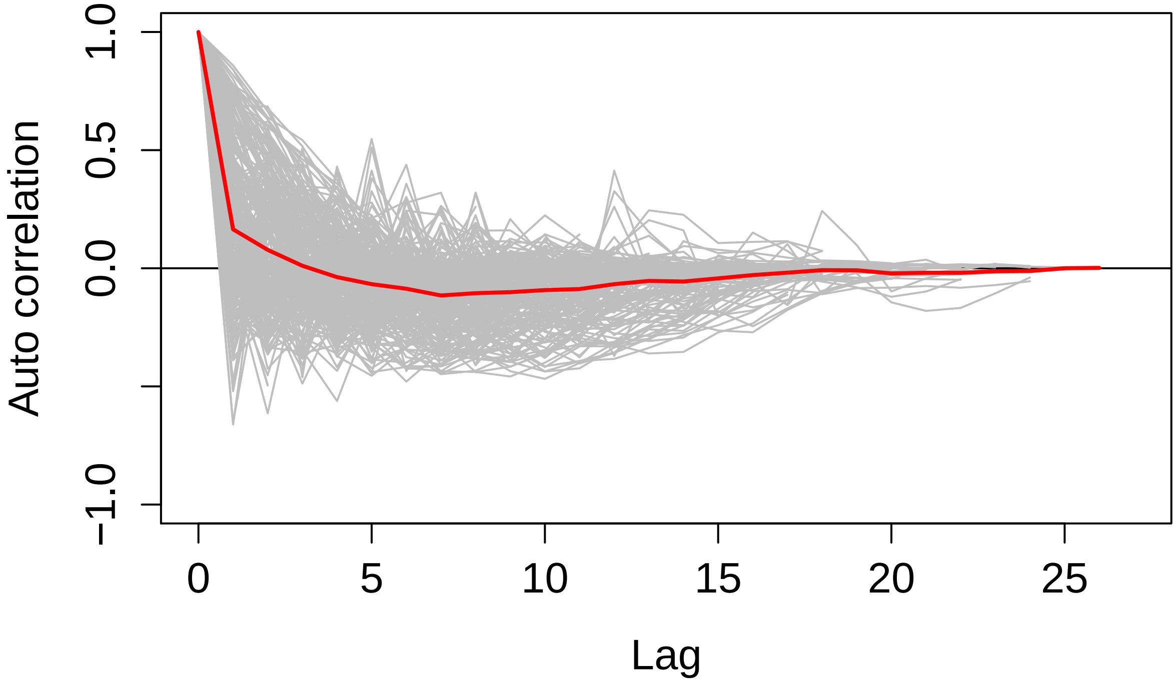
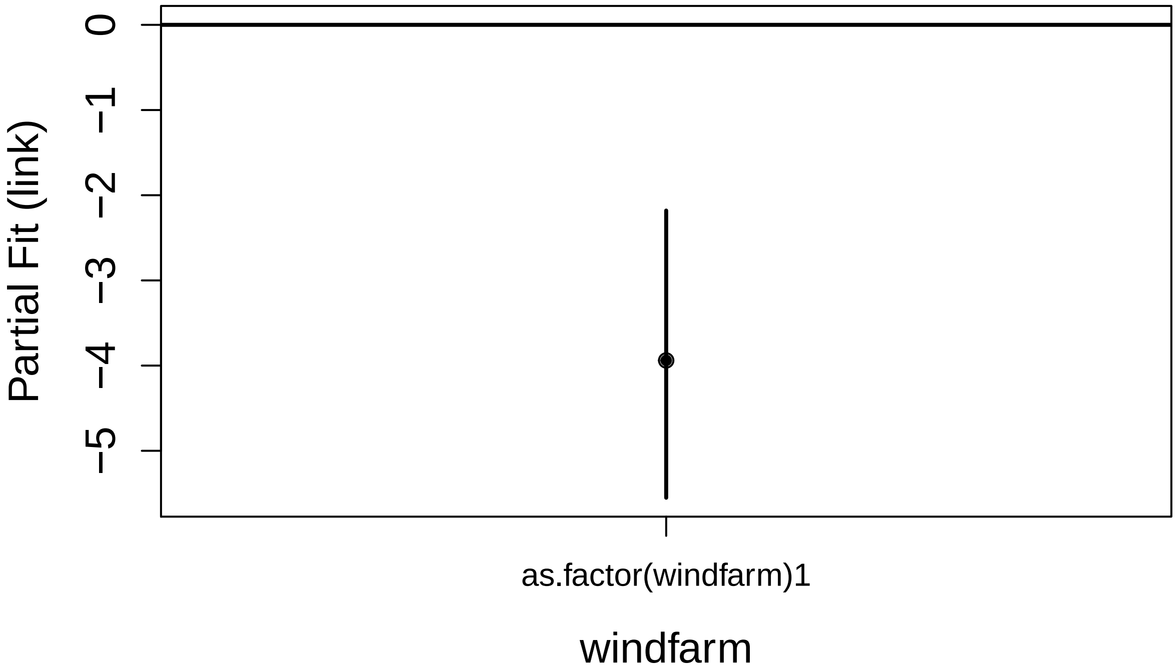
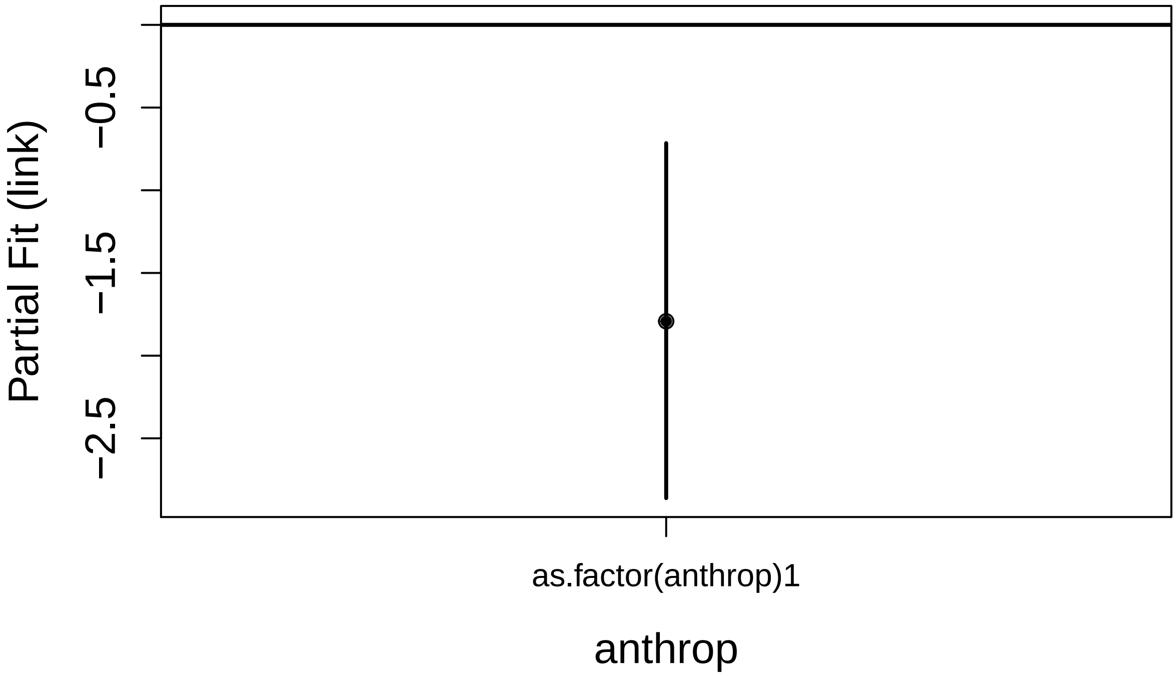


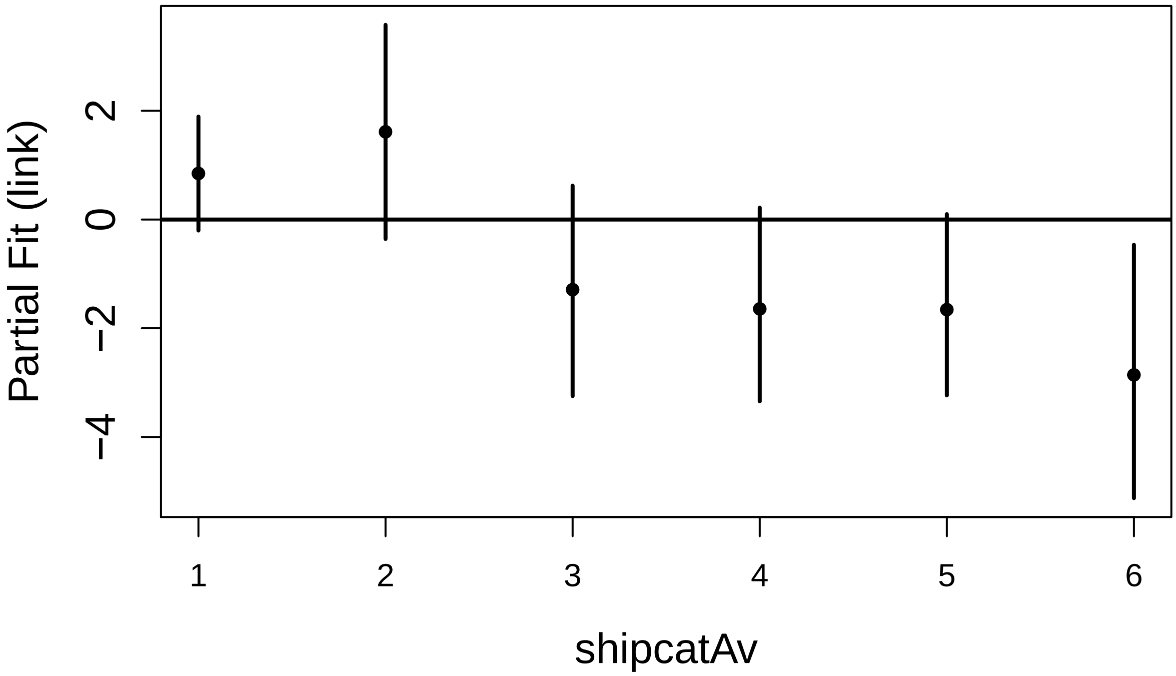
Figure D4. Plot of the correlation in the residuals for each block (grey lines). The mean correlation at each lag is indicated in red. The correlation should decay to zero (as in this case) which indicates that the correlation between residuals within a block (transect) reduces as the distance (or lag) between the segments increases.

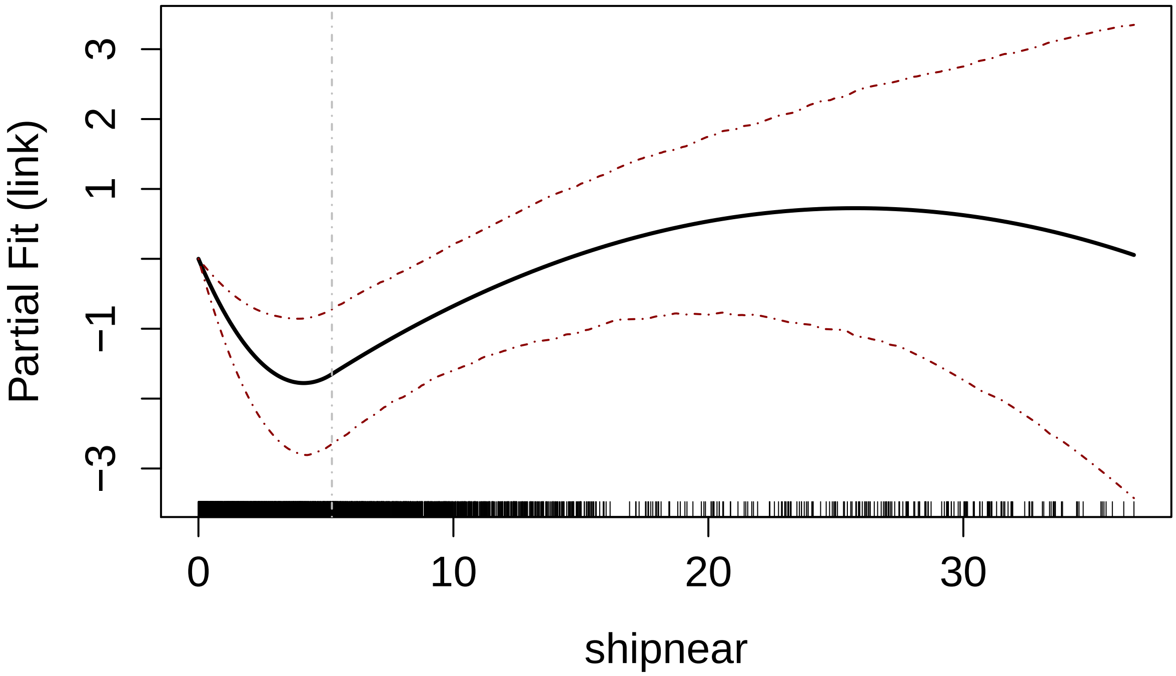
```
[1] "Making partial plots"
```

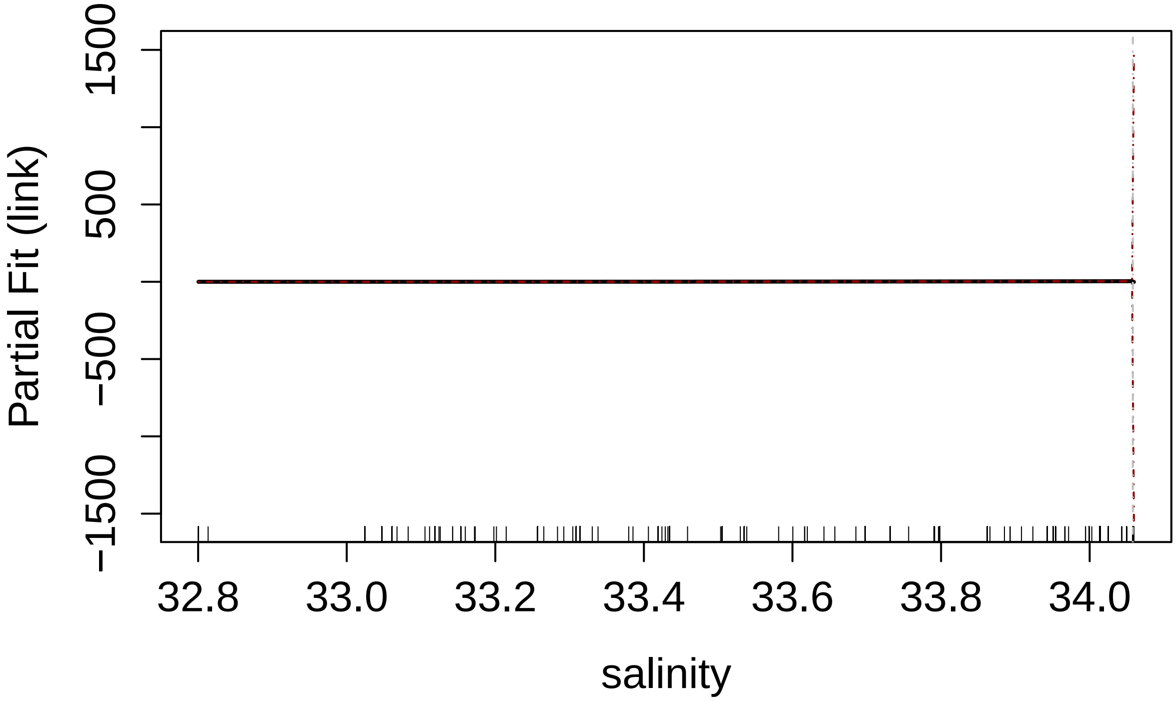

Partial Fit (link)
0
−1
−2
−3
−4
−5
as.factor(windfarm)1
windfarm



Partial Fit (link)
−0.5
−1.5
−2.5
as.factor(anthrop)1
anthrop


Partial Fit (link)
2
0
−2
−4
1
2
3
4
5
6
shipcatAv


Partial Fit (link)
3
2
1
−1
−3
0
10
20
30
shipnear


1500
500
−500
−1500
Partial Fit (link)
32.8
33.0
33.2
33.4
33.6
33.8
34.0
salinity

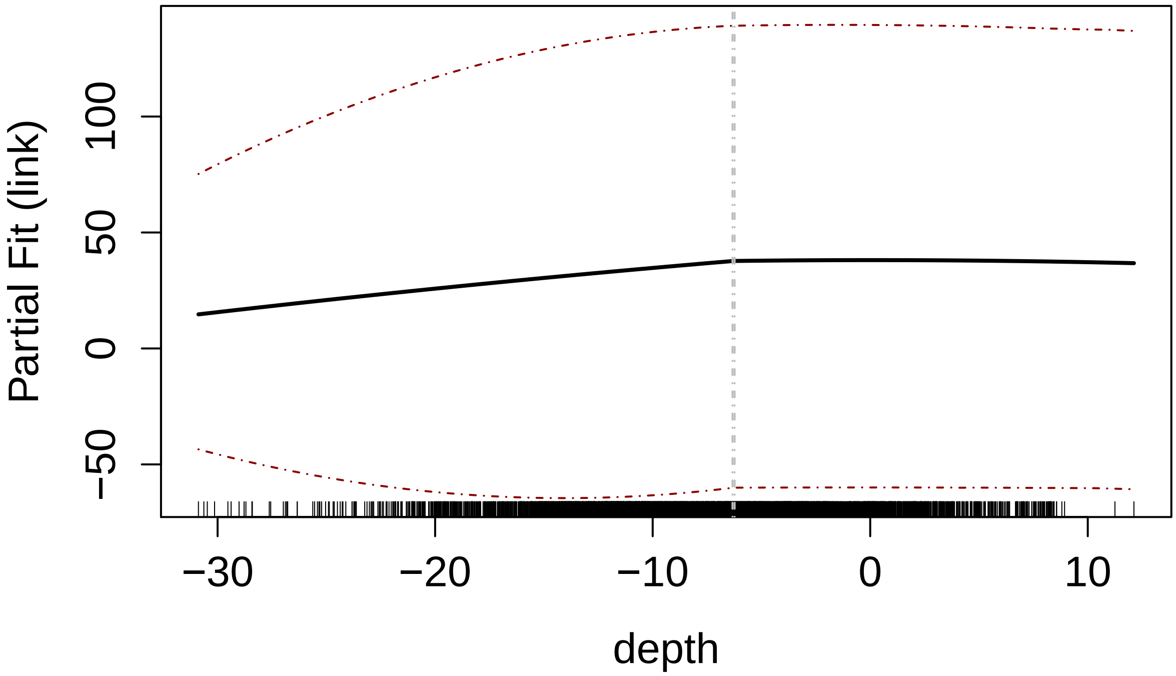


Figure D5. Partial plots of the terms in the model (on a logarithmic scale). For factor terms, level 0 is used as a baseline, or reference, level.

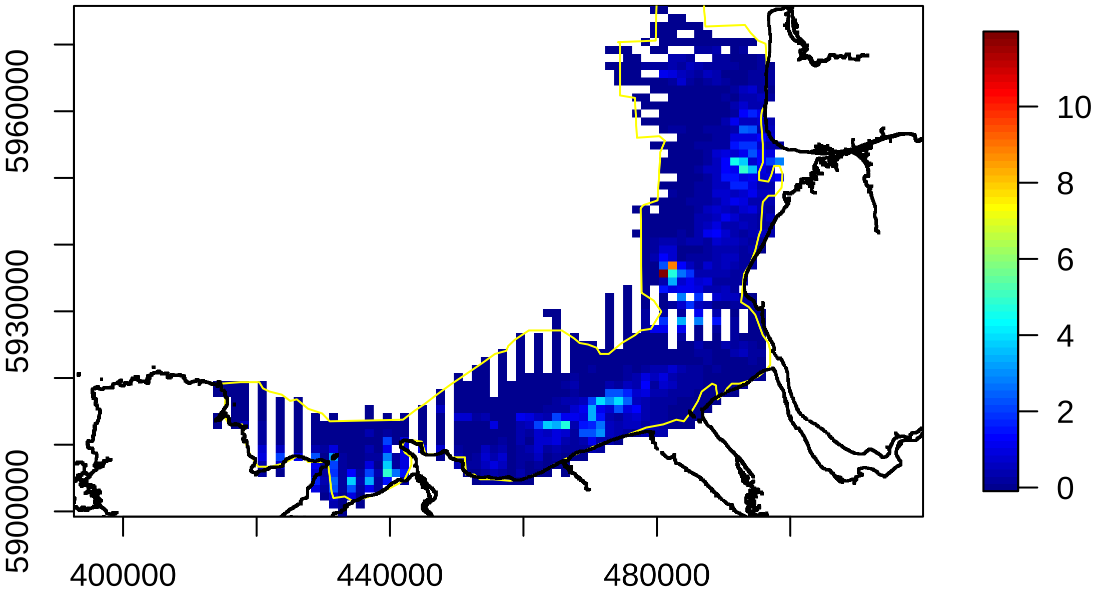


Figure D6. Plot of the fitted values (numbers of CS flying) for the survey data and averaged over all surveys.

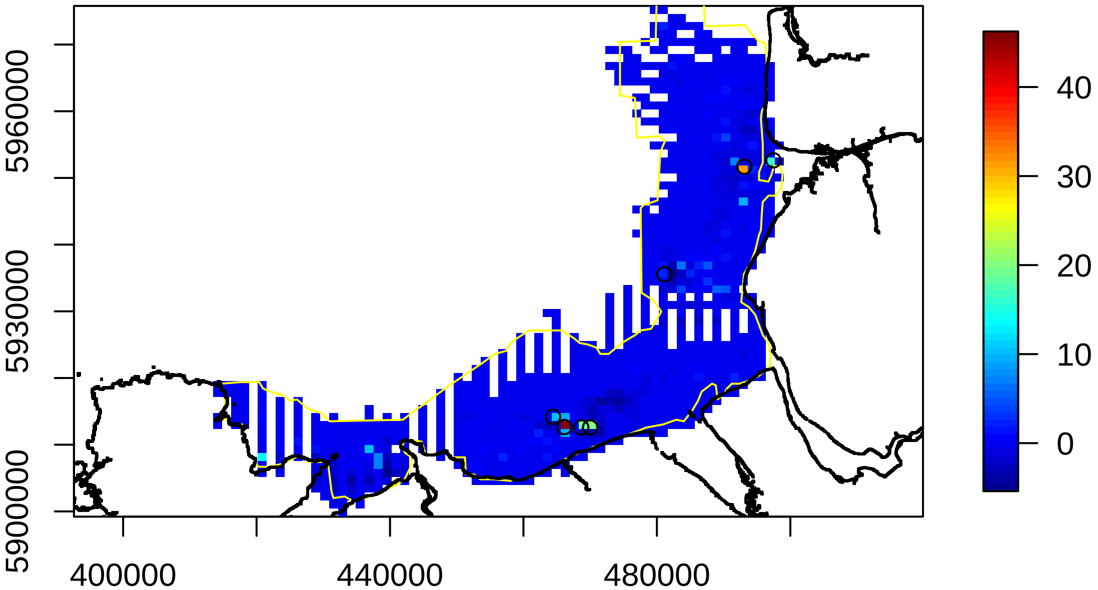


Figure D7. Plot of the residuals (difference between the observed number of birds and predicted number) averaged over all surveys. The black circles indicate segments where large numbers (>80) of CS were observed flying. The yellow line indicates the approximate boundary of Liverpool Bay SPA.

Appendix E. Model selection for Red-throated Diver sitting on the water.

This appendix describes the model selection for red-throated divers sitting on the water.

Assessing candidate explanatory variables

Each candidate explanatory variable was fitted separately to determine their likely importance and to decide which variables to select for terms which can not be fitted in the same model (Table 1).

Table 1. Fit statistics including each term seperately: average CV, and 2.5 and 97.5 percentile confidence limits, pseudo R^2 (R2, a measure of the correlation between the observed values and the fitted values from the model) and probability (p.val) associated with fitting each term seperately. Numbers in the Group column indicate variables which were grouped together and one variable from each group was chosen.

	Variable	meanCV	lowCV	highCV	R2	p.val	Group
17	s(depth)	0.856	0.854	0.8584	0.02831	0	
1	shiplennear	0.8604	0.8597	0.8613	0.01884	1.412e-11	
9	as.factor(shipcatAv)	0.87	0.8691	0.8712	0.0118	0.02009	1
13	as.factor(shipMMO)	0.87	0.8692	0.8711	0.01035	0.01305	
3	as.factor(fish)	0.8702	0.8697	0.8712	0.01004	0.0002626	
10	as.factor(shipcatMax)	0.8704	0.8694	0.8718	0.01149	0.02385	1
2	as.factor(LBspa)	0.8707	0.8702	0.8712	0.01031	9.847e-07	
4	as.factor(windfarm)	0.8713	0.8709	0.872	0.0092	0.0003464	
15	s(shipnear)	0.8727	0.8705	0.8756	0.01242	4.687e-05	
8	as.factor(shipAIS1)	0.8732	0.8727	0.8738	0.007602	0.1262	1
7	as.factor(shipnumf)	0.8735	0.8729	0.8742	0.007577	0.2692	1
16	s(salinity)	0.8736	0.8718	0.8763	0.01269	9.292e-08	
14	s(shiplenmax)	0.8738	0.8723	0.8758	0.009091	0.0001043	
12	as.factor(shipAIS10)	0.874	0.8734	0.8748	0.007162	0.4525	
11	as.factor(shipAIS5)	0.8741	0.8736	0.8747	0.007171	0.4382	
6	as.factor(windcable)	0.8745	0.874	0.8752	0.007219	0.01743	
5	as.factor(anthrop)	0.8748	0.8743	0.8755	0.006726	0.1903	
18	s(shiplenav)	0.8776	0.872	0.901	0.009173	0	

Not all factor terms can be included in a model together as the coefficients can not be estimated,

The variables *shipcatAv* was chosen.

Checking for collinearity

To check for collinearity all the (selected) variables were fitted in a model. The continuous terms were fitted as linear terms (instead of as smooth functions). The generalised variance inflation factors (GVIFs) are shown below.

```
rtd.linear <- glm(response ~ depth + shiplennear + as.factor(shipMMO) + as.factor(shipcatAv) +  as.facto

#rtd.linear <- glm(response ~ as.factor(shipcatMax) + as.factor(shipnumf) + offset(log(area)), family =

rtd.vifs <- vif(rtd.linear)
save(rtd.vifs,file="rtd_vifs.RData")
pander::pander(rtd.vifs)
```

	GVIF	Df	GVIF^(1/(2*Df))
depth	1.845	1	1.358
shiplennear	1.719	1	1.311
as.factor(shipMMO)	32.69	1	5.718
as.factor(shipcatAv)	663.7	6	1.719
as.factor(fish)	1.179	1	1.086
as.factor(LBspa)	1.025	1	1.012
as.factor(windfarm)	1.038	1	1.019
shipnear	1.774	1	1.332
salinity	1.373	1	1.172
shiplenmax	15.85	1	3.981
shiplenav	11.8	1	3.435
as.factor(shipAIS10)	2.399	1	1.549
as.factor(shipAIS5)	2.425	1	1.557
as.factor(windcable)	5800155	1	2408
as.factor(anthrop)	5800155	1	2408

There were a few large GVIFs (>5)

- *shipMMO* - this is perhaps not surprising since this information is also in *shipcatAv*. The CV scores are very similar between these two variables and since *shipcatAv* contains information on both permanent and transient shipping information, it is retained over *shipMMO*. *shipcatAv* also has a larger psuedo-R^2 value.
- *windcable* and *anthrop* - again not surprising. *windcable* has a higher CV score and so is chosen over *anthrop*
- *shiplenav* and *shiplenmax* - although these GVIFs are not >5, they are correlated to each other. At this stage they will be retained.

Selection of the factor and 1D continuous terms

The significance of terms in the initial or `full` model is shown below. Including *shiplenav* and *shiplenmax* as smooth terms in the full mode caused 'prediction from a rank deficient fit may be misleading errors' and so were included as linear terms.

```
Analysis of 'Wald statistic' Table
Model: quasipoisson, link: log
Response: response
Marginal Testing
Max Panel Size = 27; Number of panels = 280

                      Df     X2 P(>|Chi|)
shiplennear            1 13.473 0.0002420 ***
as.factor(shipcatAv)   6 10.722 0.0973421 .
as.factor(fish)        1  1.199 0.2734378
as.factor(LBspa)       1 12.882 0.0003318 ***
as.factor(windfarm)    1 26.288 2.942e-07 ***
as.factor(shipAIS10)   1  1.209 0.2715696
as.factor(shipAIS5)    1  0.001 0.9696217
as.factor(windcable)   1  0.169 0.6808544
shiplenav              1 27.583 1.505e-07 ***
shiplenmax             1  0.006 0.9385494
s(depth)               3 71.113 2.442e-15 ***
```

```
s(shipnear)           3 32.305 4.512e-07 ***
s(salinity)           4 59.893 3.055e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
     NumIter      Model        CV    meanCV     lowCV    highCV p.value
2.5%       1 Full model 0.8301184 0.8454574 0.8315617 0.8857272      NA
     numvalid
2.5%      100
```

Terms were excluded which were not significant, starting with the least significant. The first variable to be excluded was *shipAIS*5. The model was refitted and the significance of terms was checked.

```
Analysis of 'Wald statistic' Table
Model: quasipoisson, link: log
Response: response
Marginal Testing
Max Panel Size = 27; Number of panels = 280

                     Df     X2 P(>|Chi|)
shiplennear           1 13.668 0.0002181 ***
as.factor(shipcatAv)  6 10.809 0.0944709 .
as.factor(fish)       1  1.192 0.2748608
as.factor(LBspa)      1 12.881 0.0003318 ***
as.factor(windfarm)   1 26.487 2.653e-07 ***
as.factor(shipAIS10)  1  1.249 0.2637537
as.factor(windcable)  1  0.167 0.6827991
shiplenav             1 27.496 1.574e-07 ***
shiplenmax            1  0.007 0.9330567
s(depth)              3 72.716 1.110e-15 ***
s(shipnear)           3 32.518 4.070e-07 ***
s(salinity)           4 60.765 2.004e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
     NumIter                 Model        CV    meanCV     lowCV    highCV
2.5%       2 - as.factor(shipAIS5) 0.8289285 0.8446231 0.8311358 0.8838468
       p.value numvalid
2.5% 0.9696217      100
```

This process of excluding non significant terms was repeated until all terms were significant.

The term *shiplenmax* went next.

```
Analysis of 'Wald statistic' Table
Model: quasipoisson, link: log
Response: response
Marginal Testing
Max Panel Size = 27; Number of panels = 280

                     Df     X2 P(>|Chi|)
shiplennear           1 13.609 0.0002251 ***
as.factor(shipcatAv)  6 10.100 0.1205239
as.factor(fish)       1  1.289 0.2561875
as.factor(LBspa)      1 12.762 0.0003537 ***
as.factor(windfarm)   1 27.137 1.895e-07 ***
as.factor(shipAIS10)  1  1.340 0.2470139
```

```
as.factor(windcable)  1  0.280 0.5970170
shiplenav             1 27.178 1.855e-07 ***
s(depth)              3 70.051 4.108e-15 ***
s(shipnear)           4 29.242 6.982e-06 ***
s(salinity)           3 50.606 5.934e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
     NumIter        Model        CV   meanCV     lowCV   highCV   p.value
2.5%       3 - shiplenmax 0.8371605 0.874528 0.8327445 1.035246 0.9330567
     numvalid
2.5%      100
```

The term *windcable* was excluded next.

```
Analysis of 'Wald statistic' Table
Model: quasipoisson, link: log
Response: response
Marginal Testing
Max Panel Size = 27; Number of panels = 280

                     Df     X2 P(>|Chi|)
shiplennear           1 13.844 0.0001986 ***
as.factor(shipcatAv)  6 10.191 0.1168509
as.factor(fish)       1  1.225 0.2683960
as.factor(LBspa)      1 12.762 0.0003537 ***
as.factor(windfarm)   1 26.224 3.041e-07 ***
as.factor(shipAIS10)  1  1.352 0.2449911
shiplenav             1 27.151 1.881e-07 ***
s(depth)              3 77.242 < 2.2e-16 ***
s(shipnear)           4 30.087 4.698e-06 ***
s(salinity)           4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

There was a problem fitting salinity and so the max knots for that variable was reduced to 1.

```
Analysis of 'Wald statistic' Table
Model: quasipoisson, link: log
Response: response
Marginal Testing
Max Panel Size = 27; Number of panels = 280

                     Df     X2 P(>|Chi|)
shiplennear           1 13.844 0.0001986 ***
as.factor(shipcatAv)  6 10.191 0.1168509
as.factor(fish)       1  1.225 0.2683960
as.factor(LBspa)      1 12.762 0.0003537 ***
as.factor(windfarm)   1 26.224 3.041e-07 ***
as.factor(shipAIS10)  1  1.352 0.2449911
shiplenav             1 27.151 1.881e-07 ***
s(depth)              3 77.242 < 2.2e-16 ***
s(shipnear)           4 30.087 4.698e-06 ***
s(salinity)           3 53.788 1.245e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
     NumIter                  Model        CV    meanCV     lowCV   highCV
2.5%       4 - as.factor(windcable) 0.8377701 0.8778593 0.8317088 1.048431
      p.value numvalid
2.5% 0.597017      100
```

The term *fish* was excluded next.

```
Analysis of 'Wald statistic' Table
Model: quasipoisson, link: log
Response: response
Marginal Testing
Max Panel Size = 27; Number of panels = 280

                     Df     X2 P(>|Chi|)
shiplennear           1 16.963 3.811e-05 ***
as.factor(shipcatAv)  6 10.210 0.1160957
as.factor(LBspa)      1 12.453 0.0004173 ***
as.factor(windfarm)   1 26.368 2.822e-07 ***
as.factor(shipAIS10)  1  1.435 0.2308763
shiplenav             1 26.305 2.915e-07 ***
s(depth)              3 78.239 < 2.2e-16 ***
s(shipnear)           4 30.630 3.642e-06 ***
s(salinity)           3 55.314 5.886e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
     NumIter             Model        CV    meanCV     lowCV   highCV
2.5%       5 - as.factor(fish) 0.8345556 0.8716411 0.8316427 1.014325
      p.value numvalid
2.5% 0.268396      100
```

The term *shipAIS*10 was excluded next.

```
Analysis of 'Wald statistic' Table
Model: quasipoisson, link: log
Response: response
Marginal Testing
Max Panel Size = 27; Number of panels = 280

                     Df     X2 P(>|Chi|)
shiplennear           1 16.265 5.507e-05 ***
as.factor(shipcatAv)  6  9.410 0.1517802
as.factor(LBspa)      1 12.849 0.0003377 ***
as.factor(windfarm)   1 24.880 6.101e-07 ***
shiplenav             1 26.367 2.823e-07 ***
s(depth)              3 66.550 2.343e-14 ***
s(shipnear)           4 30.786 3.386e-06 ***
s(salinity)           3 52.726 2.098e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
     NumIter                  Model        CV    meanCV     lowCV   highCV
2.5%       6 - as.factor(shipAIS10) 0.8345652 0.8716539 0.8315752 1.018748
       p.value numvalid
2.5% 0.2308763      100
```

The term *shipcatAv* was excluded next.

```
Analysis of 'Wald statistic' Table
Model: quasipoisson, link: log
Response: response
Marginal Testing
Max Panel Size = 27; Number of panels = 280

                    Df     X2 P(>|Chi|)
shiplennear          1 18.051 2.151e-05 ***
as.factor(LBspa)     1 13.998  0.000183 ***
as.factor(windfarm)  1 25.122 5.382e-07 ***
shiplenav            1  4.646  0.031133 *
s(depth)             3 84.988 < 2.2e-16 ***
s(shipnear)          4 27.741 1.407e-05 ***
s(salinity)          3 54.295 9.709e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
     NumIter                  Model        CV    meanCV     lowCV
2.5%       7 - as.factor(shipcatAv) 0.8251952 0.8322885 0.8280589
        highCV   p.value numvalid
2.5% 0.8379216 0.1517802      100
```

All terms in the model were significant.

Inclusion of a 2D term for location

A 2D term for location was then added.

```
Analysis of 'Wald statistic' Table
Model: quasipoisson, link: log
Response: response
Marginal Testing
Max Panel Size = 27; Number of panels = 280

                    Df     X2 P(>|Chi|)
shiplennear          1 11.902 0.0005607 ***
as.factor(LBspa)     1 16.697 4.385e-05 ***
as.factor(windfarm)  1 22.166 2.501e-06 ***
shiplenav            1  4.955 0.0260213 *
s(depth)             3 77.928 < 2.2e-16 ***
s(shipnear)          4 23.245 0.0001131 ***
s(salinity)          3  7.772 0.0509740 .
s(x.pos, y.pos)      8 49.524 5.045e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
     NumIter            Model        CV    meanCV     lowCV    highCV
2.5%       8 + s(x.pos,y.pos) 0.7899974 0.7942346 0.7874972 0.8037324
     p.value numvalid
2.5%      NA      100
```

There were two shipping variables in the model which could potentially be combined with location in an interaction term. Each term was tried in turn and the CVs for each of these models are shown below. ‘None’ indicates the model with no interaction term fitted.

Variable	meanCV	lowCV	highCV	p.value
None	0.7942	0.7875	0.8037	NA
shiplenav	0.8004	0.795	0.8065	0.0004607
shiplennear	0.8061	0.8011	0.8123	0.2274

The CV score increased with adding interaction terms and so no interaction term was included.

A summary of the iterations from the model selection process are given below.

Table 4: Table continues below

	NumIter	Model	CV	meanCV	lowCV	highCV
2.5%	1	Full model	0.8301	0.8455	0.8316	0.8857
2.5%1	2	- as.factor(shipAIS5)	0.8289	0.8446	0.8311	0.8838
2.5%2	3	- shiplenmax	0.8372	0.8745	0.8327	1.035
2.5%3	4	- as.factor(windcable)	0.8378	0.8779	0.8317	1.048
2.5%4	5	- as.factor(fish)	0.8346	0.8716	0.8316	1.014
2.5%5	6	- as.factor(shipAIS10)	0.8346	0.8717	0.8316	1.019
2.5%6	7	- as.factor(shipcatAv)	0.8252	0.8323	0.8281	0.8379
2.5%7	8	+ s(x.pos,y.pos)	0.79	0.7942	0.7875	0.8037

	p.value	numvalid
2.5%	NA	100
2.5%1	0.9696	100
2.5%2	0.9331	100
2.5%3	0.597	100
2.5%4	0.2684	100
2.5%5	0.2309	100
2.5%6	0.1518	100
2.5%7	NA	100

Selected model

The anova of the selected model is given below.

```
Analysis of 'Wald statistic' Table
Model: quasipoisson, link: log
Response: response
Marginal Testing
Max Panel Size = 27; Number of panels = 280

                    Df     X2 P(>|Chi|)
shiplennear          1 11.902 0.0005607 ***
as.factor(LBspa)     1 16.697 4.385e-05 ***
as.factor(windfarm)  1 22.166 2.501e-06 ***
shiplenav            1  4.955 0.0260213 *
s(depth)             3 77.928 < 2.2e-16 ***
s(shipnear)          4 23.245 0.0001131 ***
s(salinity)          3  7.772 0.0509740 .
s(x.pos, y.pos)      8 49.524 5.045e-08 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The summary is given below

```

Call:
gamMRSea(formula = response ~ shiplennear + as.factor(LBspa) +
    as.factor(windfarm) + shiplenav + bs(depth, knots = splineParams[[2]]$knots,
    degree = splineParams[[2]]$degree, Boundary.knots = splineParams[[2]]$bd) +
    bs(shipnear, knots = splineParams[[3]]$knots, degree = splineParams[[3]]$degree,
        Boundary.knots = splineParams[[3]]$bd) + bs(salinity,
    knots = splineParams[[4]]$knots, degree = splineParams[[4]]$degree,
    Boundary.knots = splineParams[[4]]$bd) + LRF.g(radiusIndices,
    dists, radii, aR) + offset(log(area)), family = quasipoisson(link = log),
    data = all.seg, splineParams = splineParams)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6849  -0.7242  -0.4753  -0.2337   8.8705

Coefficients:
                       Estimate Std. Error Robust S.E. t value Pr(>|t|)
(Intercept)          -15.466148   8.711942    5.327308  -2.903 0.003714
shiplennear           -0.004245   0.001283    0.001231  -3.450 0.000566
as.factor(LBspa)1      1.300013   0.451721    0.318146   4.086 4.47e-05
as.factor(windfarm)1  -1.435850   0.416752    0.304976  -4.708 2.59e-06
shiplenav              0.006960   0.003481    0.003127   2.226 0.026077
s(depth)1              7.416161   8.836083    5.306659   1.398 0.162334
s(depth)2             13.757889   8.304407    4.915624   2.799 0.005154
s(depth)3              6.216248   8.825321    5.369780   1.158 0.247082
s(shipnear)1          -0.344164   0.557025    0.537391  -0.640 0.521926
s(shipnear)2           1.029420   0.291634    0.333715   3.085 0.002051
s(shipnear)3          -1.707781   0.572281    0.725306  -2.355 0.018592
s(shipnear)4           2.339594   0.721972    1.006649   2.324 0.020168
s(salinity)1           2.835475   1.711869    1.699699   1.668 0.095350
s(salinity)2           1.613581   1.504935    1.256139   1.285 0.199021
s(salinity)3           3.368724   1.695718    1.585564   2.125 0.033679
s(x.pos, y.pos)b1     31.875040   4.356424    6.934438   4.597 4.43e-06
s(x.pos, y.pos)b2      3.330521   1.222648    1.759172   1.893 0.058399
s(x.pos, y.pos)b3      8.340681   2.061783    3.200610   2.606 0.009195
s(x.pos, y.pos)b4      4.457728   0.870828    1.030316   4.327 1.55e-05
s(x.pos, y.pos)b5    -42.757291   5.995289    9.702934  -4.407 1.08e-05
s(x.pos, y.pos)b6      6.807058   1.479939    2.021673   3.367 0.000767
s(x.pos, y.pos)b7    -45.907432   9.914132   16.033072  -2.863 0.004214
s(x.pos, y.pos)b8     38.496701   8.493416   13.503830   2.851 0.004383

(Intercept)          **
shiplennear          ***
as.factor(LBspa)1    ***
as.factor(windfarm)1 ***
shiplenav            *
s(depth)1
s(depth)2            **
s(depth)3
```

```
s(shipnear)1
s(shipnear)2        **
s(shipnear)3        *
s(shipnear)4        *
s(salinity)1        .
s(salinity)2
s(salinity)3        *
s(x.pos, y.pos)b1   ***
s(x.pos, y.pos)b2   .
s(x.pos, y.pos)b3   **
s(x.pos, y.pos)b4   ***
s(x.pos, y.pos)b5   ***
s(x.pos, y.pos)b6   ***
s(x.pos, y.pos)b7   **
s(x.pos, y.pos)b8   **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 2.220622)

    Null deviance: 4556.7  on 4038  degrees of freedom
Residual deviance: 3609.9  on 4016  degrees of freedom
AIC:  NA

Max Panel Size = 27; Number of panels = 280
Number of Fisher Scoring iterations: 6
```

Diagnostics

The diagnostics of the final model were investigated to ensure that the model was valid. The diagnostic plots for this model are as follows:

- Figure E1 shows the position of the fitted knots for the 2D smooth term.
- Figure E2 indicates that the blocking structure was appropriate, the correlation in all blocks declined to zero as distance increased.
- Figure E3 contains the partial plots on the scale of the link function for selected terms in the model. Values above/below zero on the y-axis indicate that the number of birds would increase/decrease at that value of the covariate shown by the x-axis (given values for other covariates remain the same).
- Figure E4 shows the predicted values obtained from the model for the survey data. Predictions for the whole of the prediction grid are shown in the main report.
- Figure E5 shows the residuals from the model. The large residuals occurred where there were very high observed numbers of birds in segments.

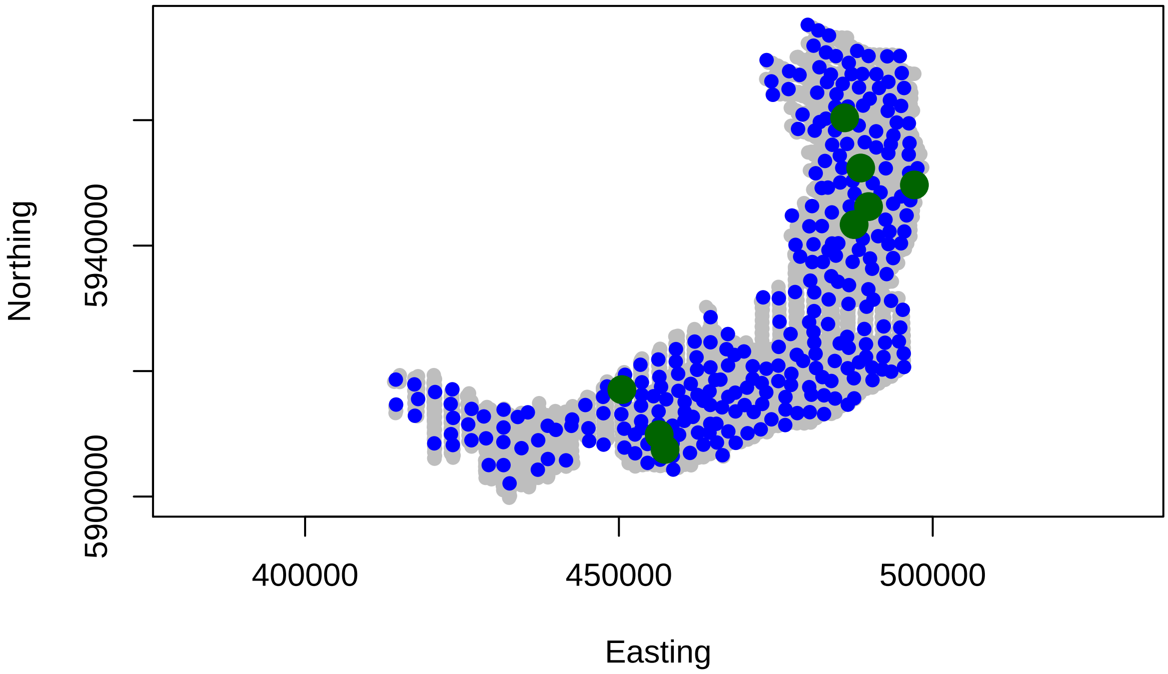


Figure E1. Location of the knots in the chosen model.

The runs test was performed to determine if the residuals were correlated. The *p*-value associated with this test did not indicate that the residuals were correlated. Thus, the standard errors and robust standard errors shown in the model summary above were very similar.

```
    Runs Test - Two sided; Empirical Distribution

data:  residuals(rtd.2dOutput$bestModel, type = "pearson")
Standardized Runs Statistic = -21.386, p-value = 0.512
```

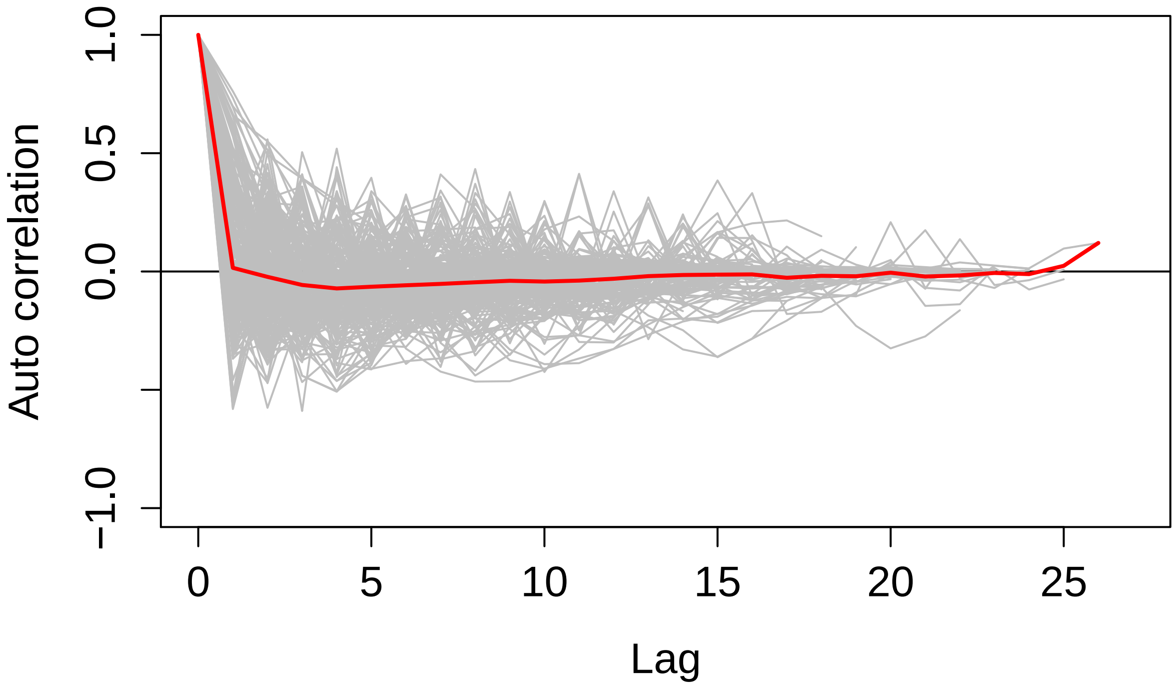
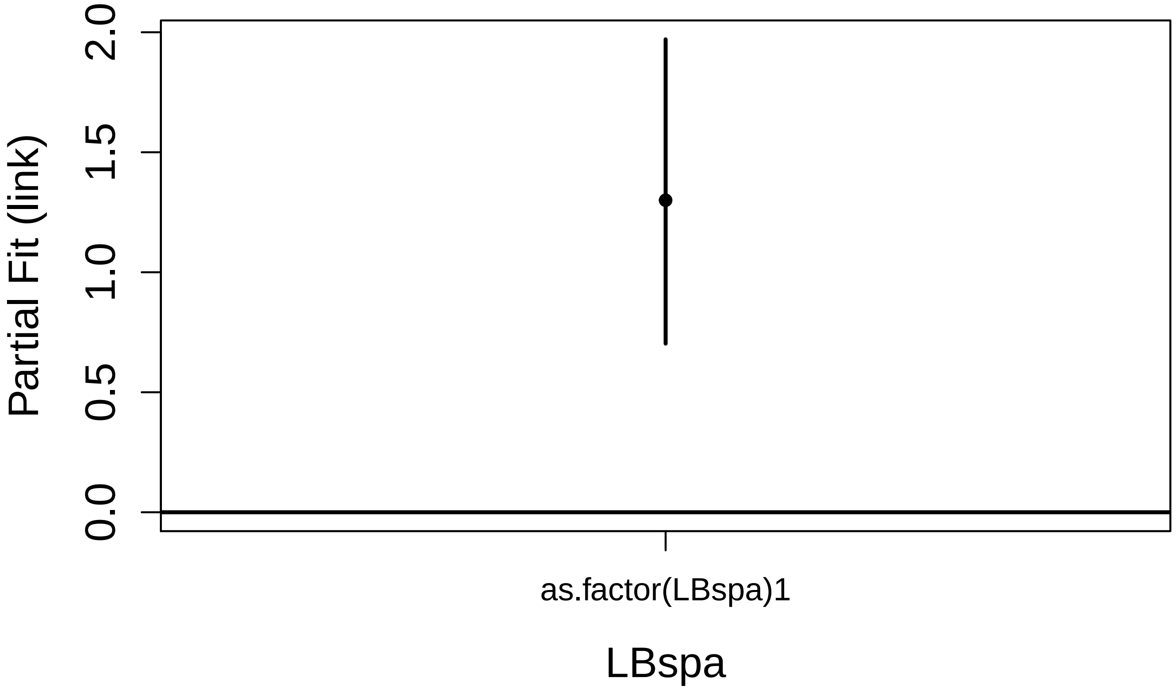
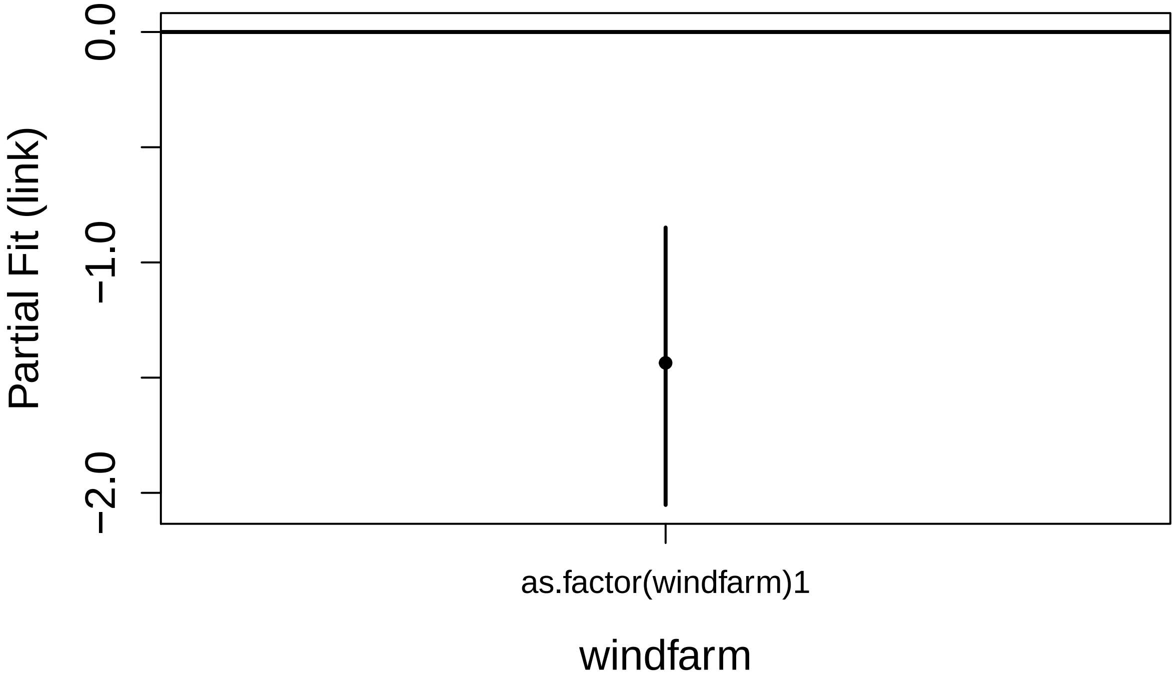
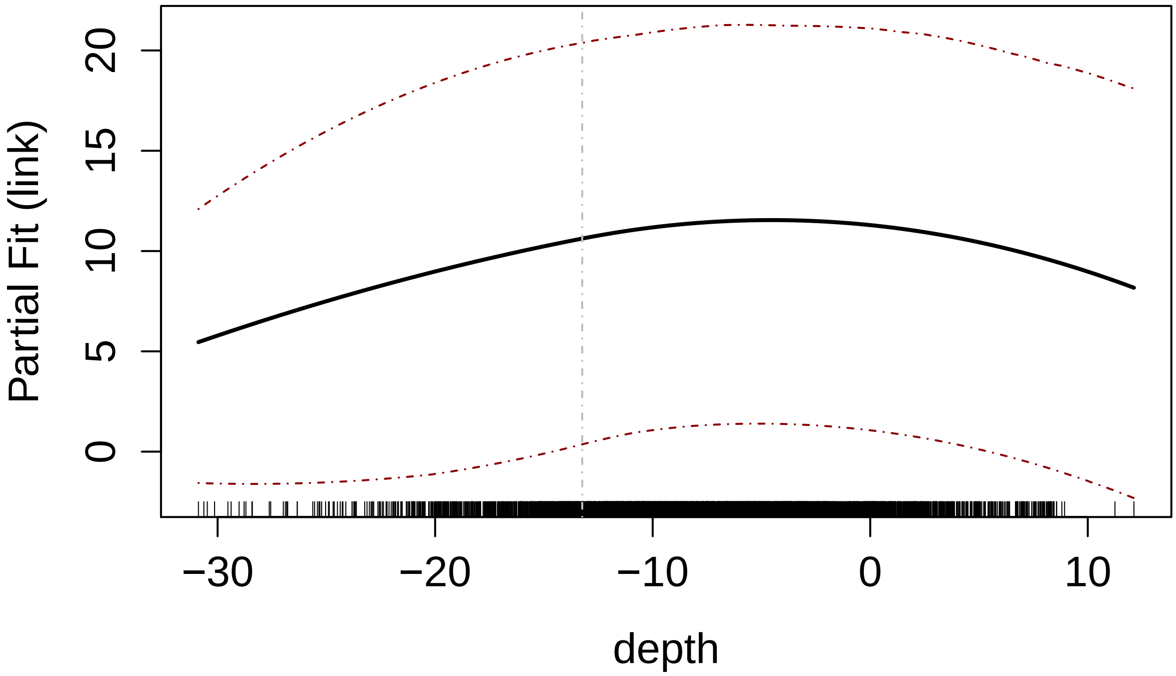


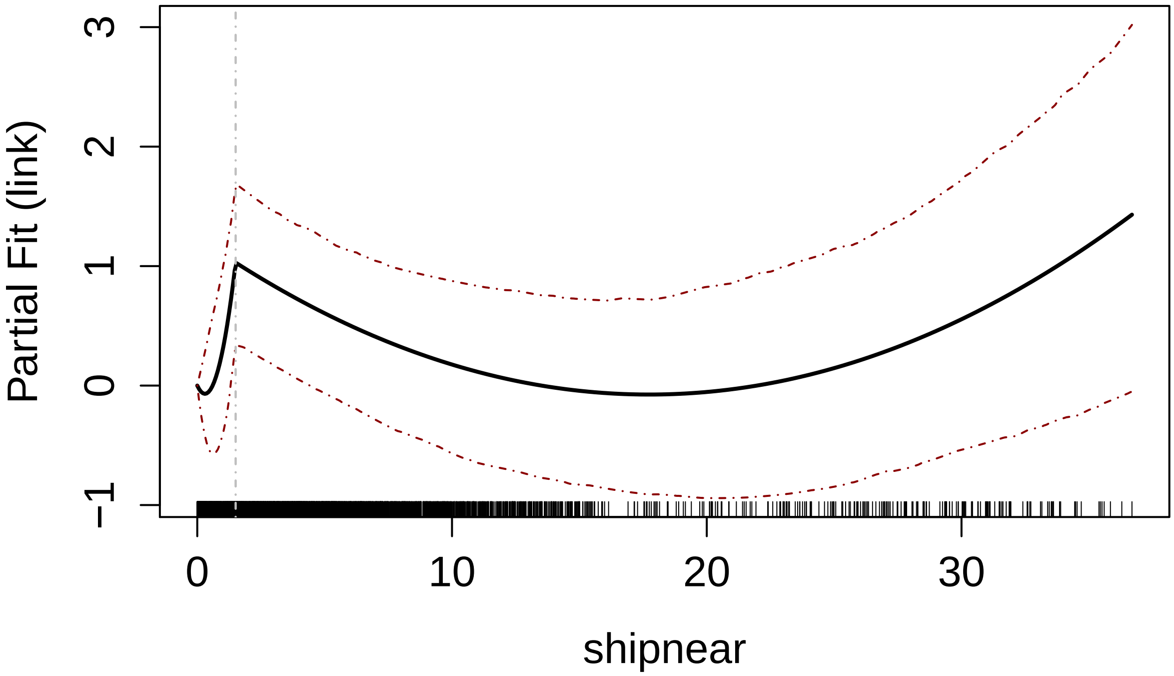
Figure E2. Plot of the correlation in the residuals for each block (grey lines). The mean correlation at each lag is indicated in red. The correlation should decay to zero (as in this case) which indicates that the correlation between residuals within blocks (transects) reduces as the distance (or lag) between the segments increases.

```
[1] "Making partial plots"
```


Partial Fit (link)
0.0
0.5
1.0
1.5
2.0
as.factor(LBspa)1
LBspa


Partial Fit (link)
0.0
−1.0
−2.0
as.factor(windfarm)1
windfarm


Partial Fit (link)
0
5
10
15
20
−30
−20
−10
0
10
depth


Partial Fit (link)
3
2
1
0
−1
0
10
20
30
shipnear

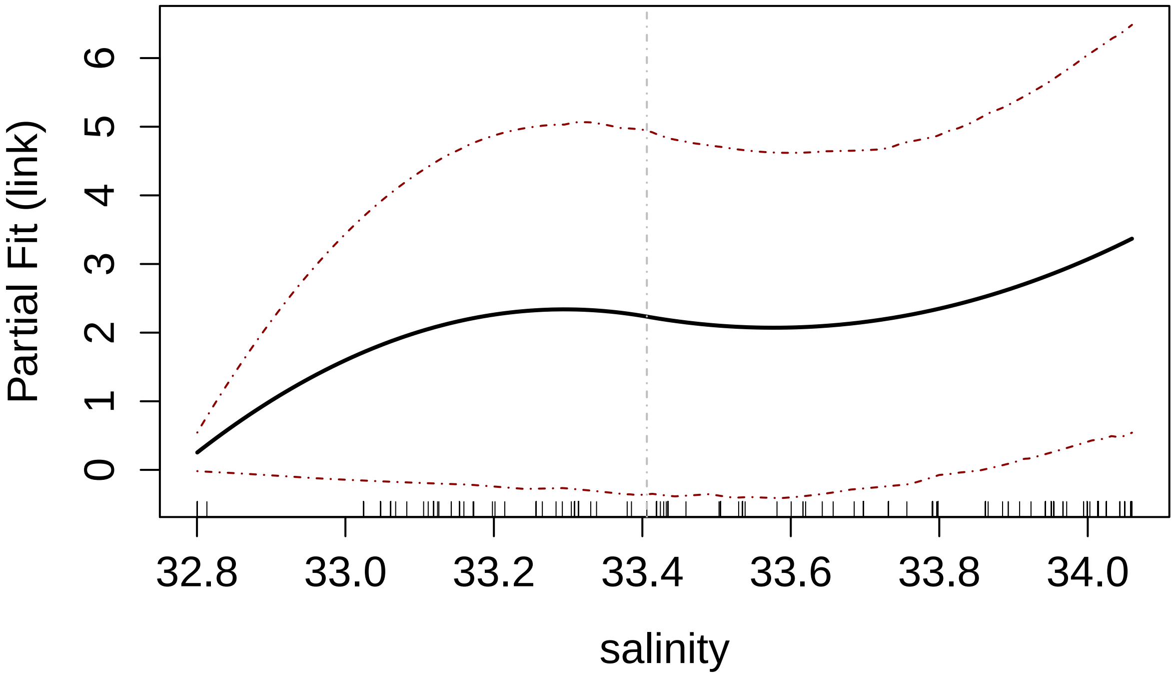


Figure E3. Partial plots of the terms in the model (on a logarithmic scale). For factor terms, level 0 is used as a baseline, or reference, level.

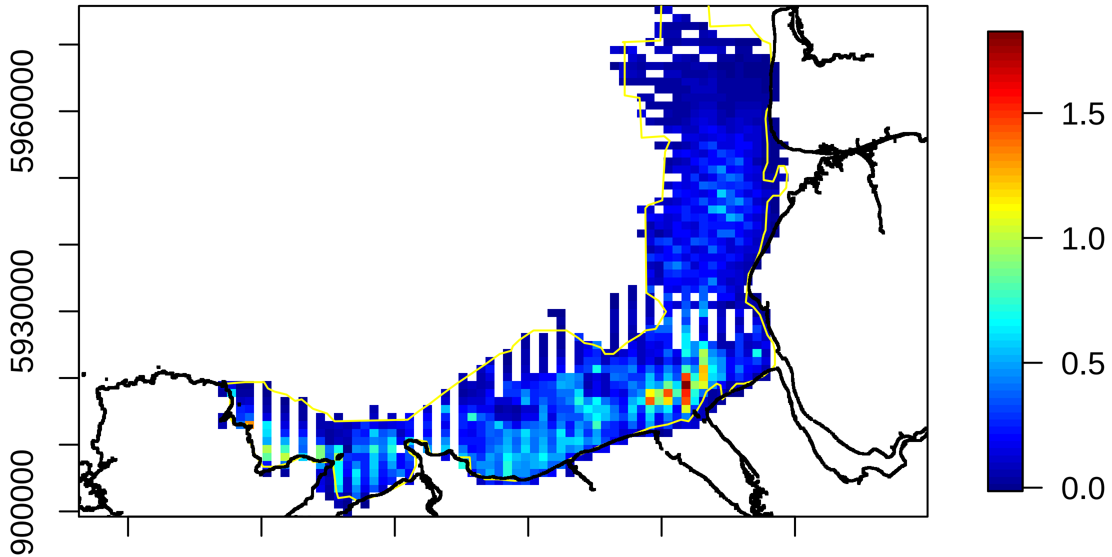


Figure E4. Plot of the predicted values (numbers of RTD sitting) for the survey data averaged over all surveys. The yellow line indicates the approximate boundary of Liverpool Bay SPA.

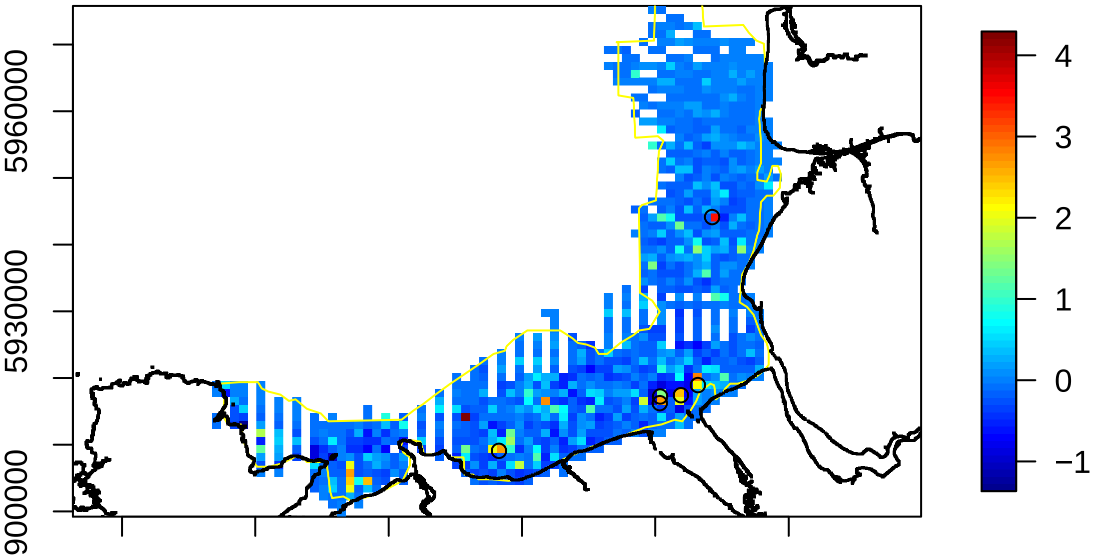


Figure E5. Plot of the residuals (difference between the observed number of birds and predicted number) averaged over all surveys. The yellow line indicates the approximate boundary of Liverpool Bay SPA. The black circles indicate the location of segments where large numbers of birds (>10 per segment) were observed.

www.gov.uk/natural-england