Density Model for Kogia Whales (*Kogia spp.*) for the U.S. Gulf of Mexico: Supplementary Report

Duke University Marine Geospatial Ecology Lab*

Model Version 3.3 - 2015-10-07

Citation

When referencing our methodology or results generally, please cite our open-access article:

Roberts JJ, Best BD, Mannocci L, Fujioka E, Halpin PN, Palka DL, Garrison LP, Mullin KD, Cole TVN, Khan CB, McLellan WM, Pabst DA, Lockhart GG (2016) Habitat-based cetacean density models for the U.S. Atlantic and Gulf of Mexico. Scientific Reports 6: 22615. doi: 10.1038/srep22615

To reference this specific model or Supplementary Report, please cite:

Roberts JJ, Best BD, Mannocci L, Fujioka E, Halpin PN, Palka DL, Garrison LP, Mullin KD, Cole TVN, Khan CB, McLellan WM, Pabst DA, Lockhart GG (2015) Density Model for Kogia Whales (*Kogia spp.*) for the U.S. Gulf of Mexico Version 3.3, 2015-10-07, and Supplementary Report. Marine Geospatial Ecology Lab, Duke University, Durham, North Carolina.

Copyright and License

Revision History

Version	Date	Description of changes
1	2014-08-06	Initial version.
2	2014-10-24	Updated distance to eddy predictors using Chelton et al.'s 2014 database. Removed distance to canyon and wind speed predictors; they were not ecologically justified. Fixed missing pixels in several climatological predictors, which led to not all segments being utilized.
3	2015-01-10	Added a missing sighting and refitted detection functions and models.
3.1	2015-02-02	Updated the documentation. No changes to the model.
3.2	2015-05-14	Updated calculation of CVs. Switched density rasters to logarithmic breaks. No changes to the model.
3.3	2015-10-07	Updated the documentation. No changes to the model.

*For questions, or to offer feedback about this model or report, please contact Jason Roberts (jason.roberts@duke.edu)

Survey Data

Survey	Period	Length (1000 km)	Hours	Sightings
SEFSC GOMEX92-96 Aerial Surveys	1992-1996	27	152	0
SEFSC Gulf of Mexico Shipboard Surveys, 2003-2009	2003-2009	19	1156	32
SEFSC GulfCet I Aerial Surveys	1992-1994	50	257	34
SEFSC GulfCet II Aerial Surveys	1996-1998	22	124	21
SEFSC GulfSCAT 2007 Aerial Surveys	2007-2007	18	95	0
SEFSC Oceanic CetShip Surveys	1992-2001	49	3102	132
SEFSC Shelf CetShip Surveys	1994-2001	10	707	0
Total		195	5593	219

Table 2: Survey effort and sightings used in this model. Effort is tallied as the cumulative length of on-effort transects and hours the survey team was on effort. Sightings are the number of on-effort encounters of the modeled species for which a perpendicular sighting distance (PSD) was available. Off effort sightings and those without PSDs were omitted from the analysis.

Period	Length (1000 km)	Hours	Sightings
1992-2009	195	5592	219
1998-2009	62	2679	75
% Lost	68	52	66

Table 3: Survey effort and on-effort sightings having perpendicular sighting distances. % Lost shows the percentage of effort or sightings lost by restricting the analysis to surveys performed in 1998 and later, the era in which remotely-sensed chlorophyll and derived productivity estimates are available. See Figure 1 for more information.

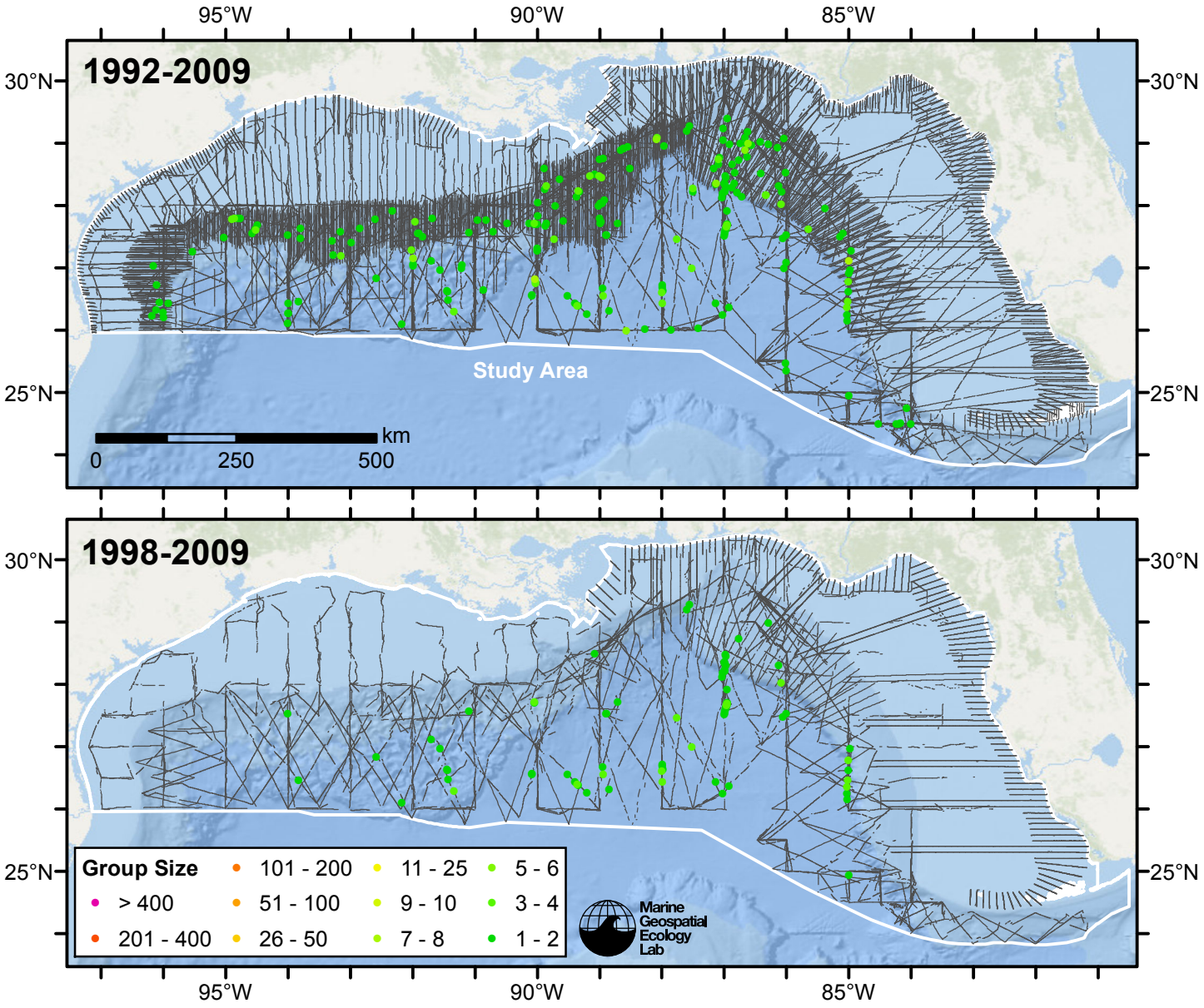


Figure 1: Kogia whales sightings and survey tracklines. The top map shows all surveys. The bottom map shows surveys performed in 1998 or later. the era in which remotely-sensed chlorophyll and derived productivity estimates are available. Models fitted to contemporaneous (day-of-sighting) estimates of those predictors only utilize these surveys. These maps illustrate the survey data lost in order to utilize those predictors. Models fitted to climatogical estimates of those predictors do not suffer this data loss.

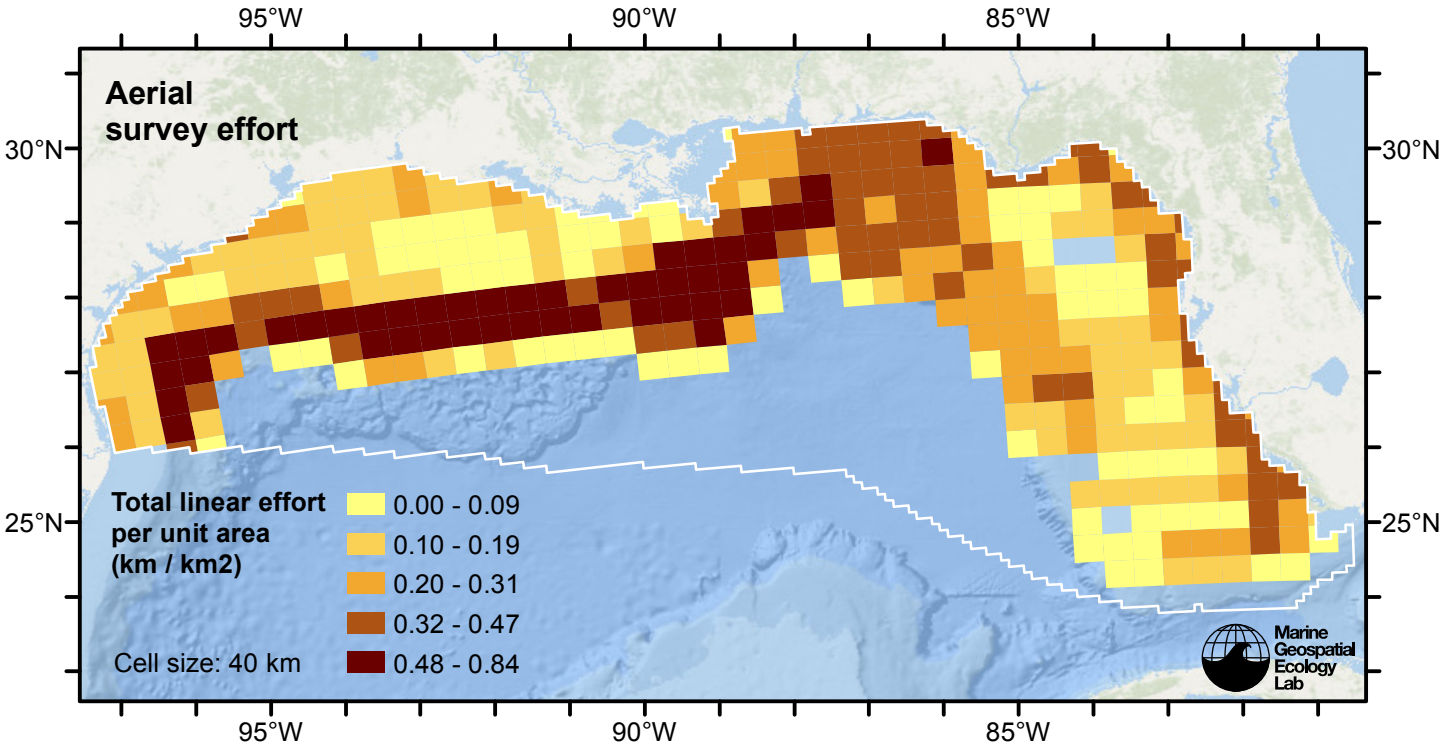


Figure 2: Aerial linear survey effort per unit area.

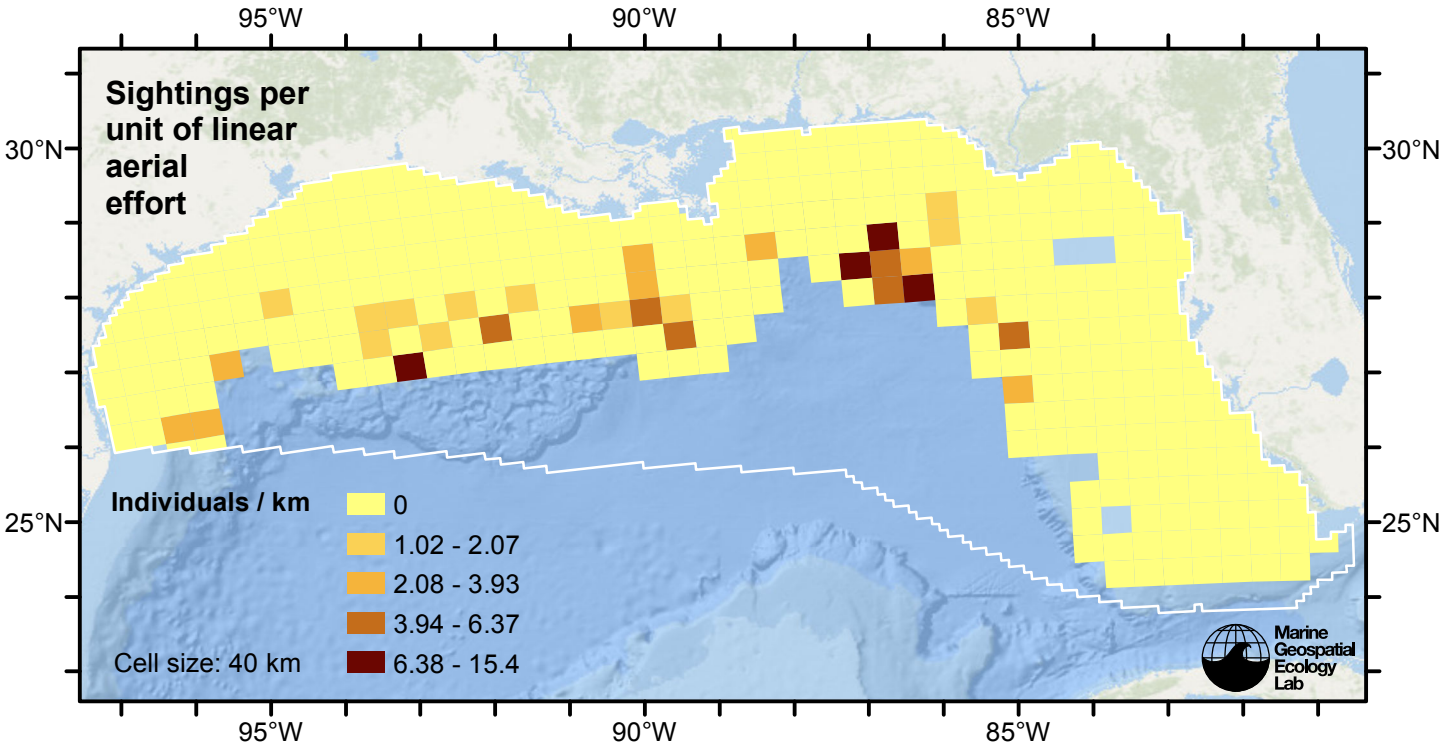


Figure 3: Kogia whales sightings per unit aerial linear survey effort.

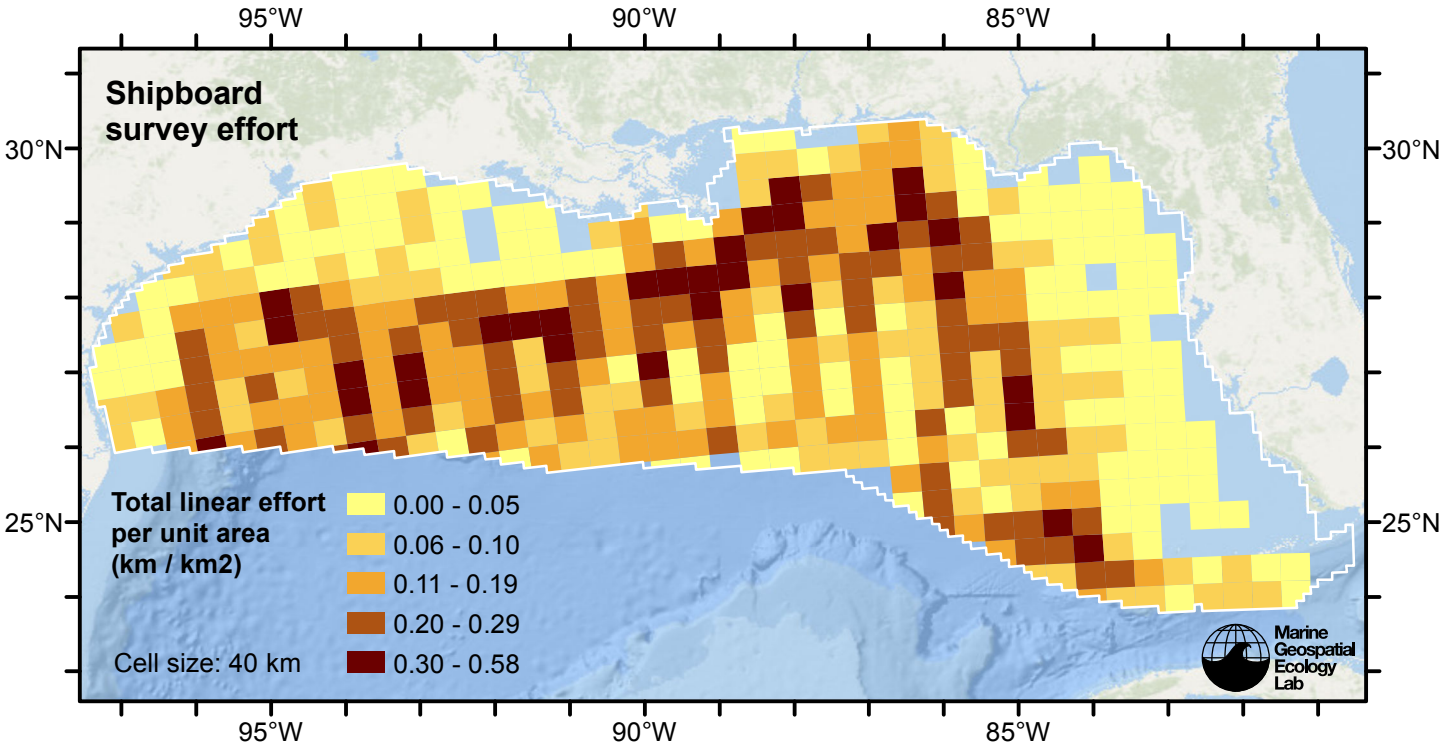


Figure 4: Shipboard linear survey effort per unit area.

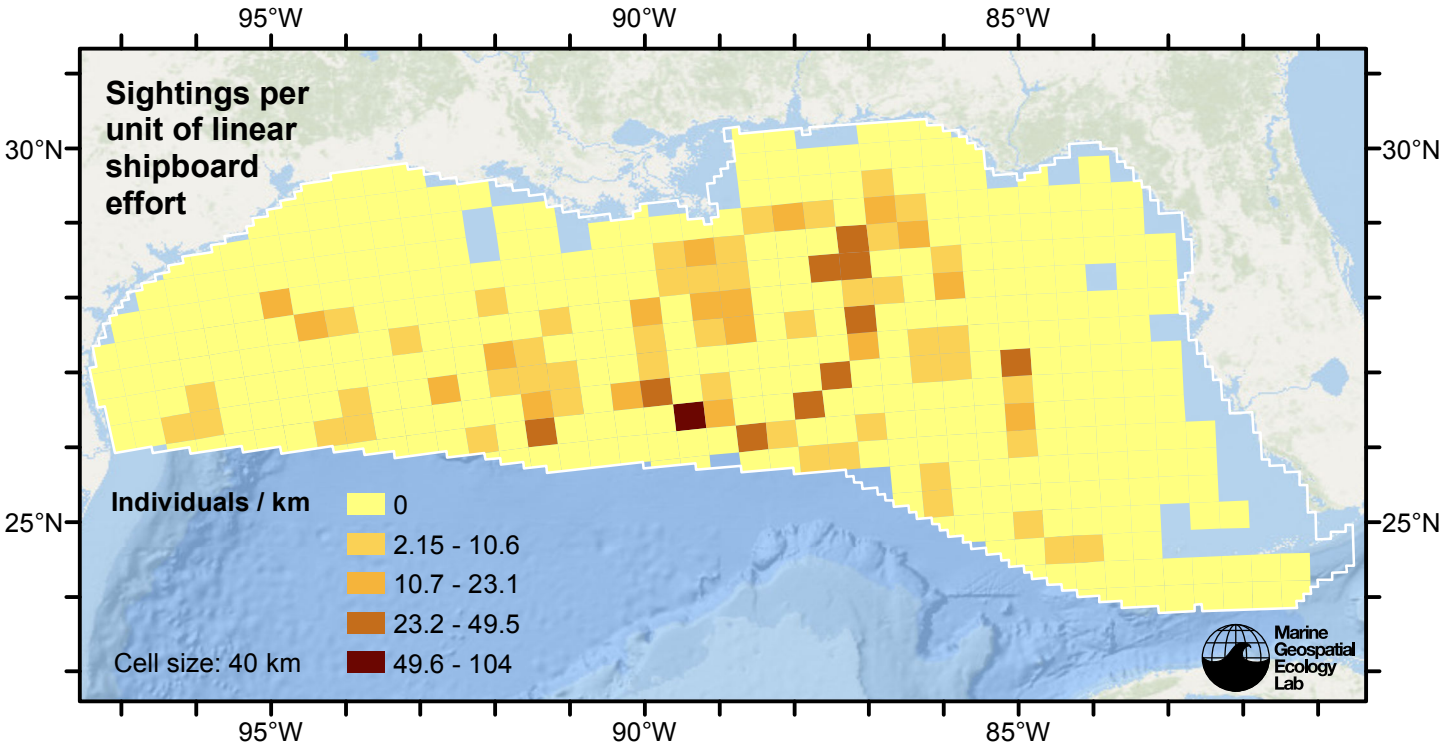


Figure 5: Kogia whales sightings per unit shipboard linear survey effort.

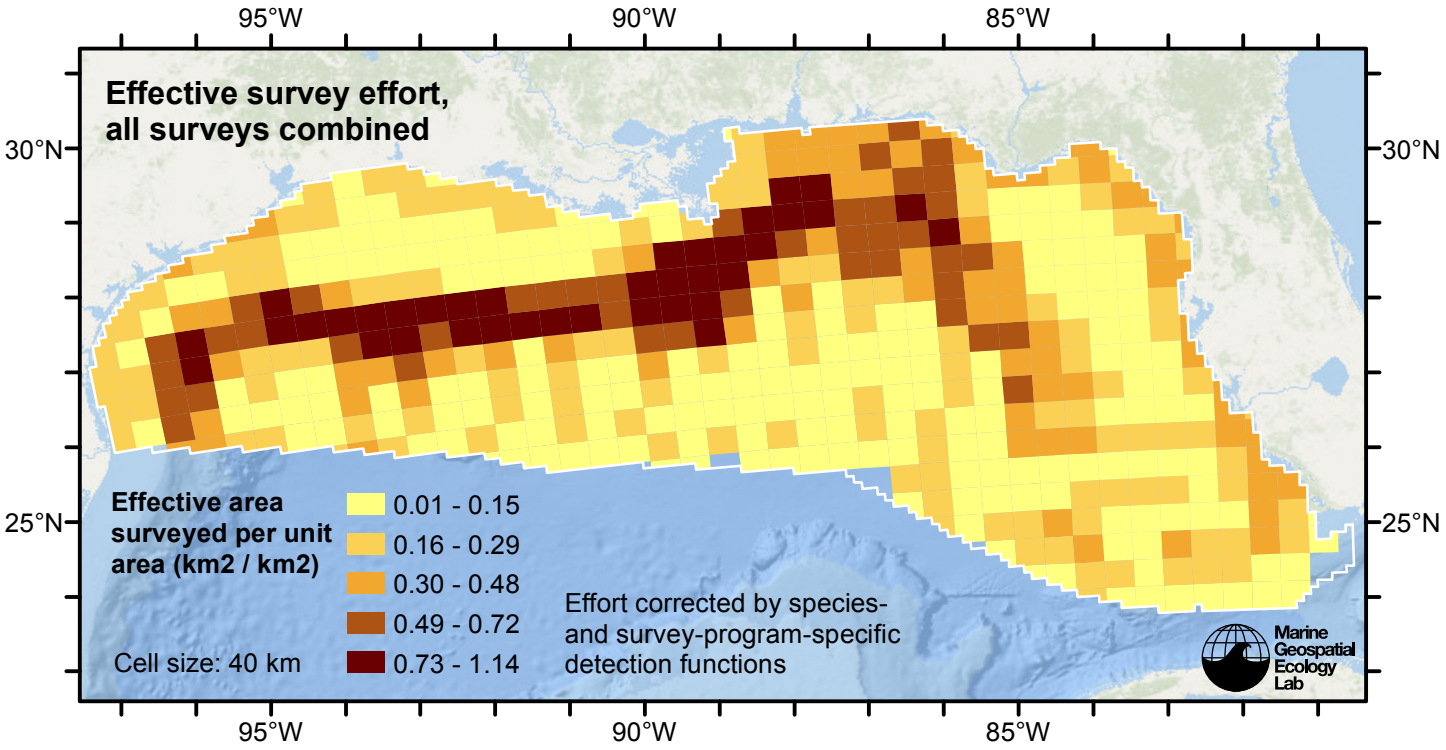


Figure 6: Effective survey effort per unit area, for all surveys combined. Here, effort is corrected by the species- and survey-program-specific detection functions used in fitting the density models.

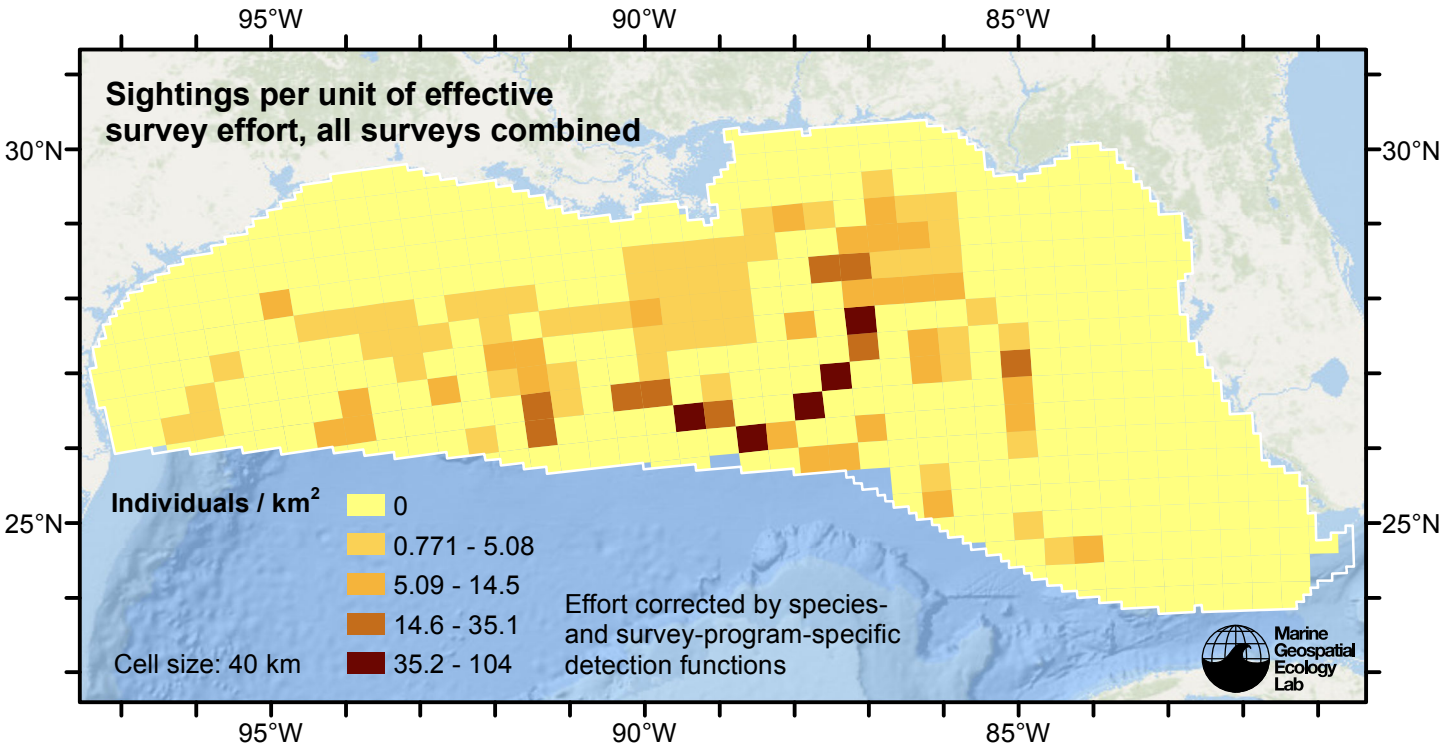


Figure 7: Kogia whales sightings per unit of effective survey effort, for all surveys combined. Here, effort is corrected by the species- and survey-program-specific detection functions used in fitting the density models.

Detection Functions

The detection hierarchy figures below show how sightings from multiple surveys were pooled to try to achieve Buckland et. al's (2001) recommendation that at least 60-80 sightings be used to fit a detection function. Leaf nodes, on the right, usually represent individual surveys, while the hierarchy to the left shows how they have been grouped according to how similar we believed the surveys were to each other in their detection performance.

At each node, the red or green number indicates the total number of sightings below that node in the hierarchy, and is colored green if 70 or more sightings were available, and red otherwise. If a grouping node has zero sightings–i.e. all of the surveys within it had zero sightings–it may be collapsed and shown as a leaf to save space.

Each histogram in the figure indicates a node where a detection function was fitted. The actual detection functions do not appear in this figure; they are presented in subsequent sections. The histogram shows the frequency of sightings by perpendicular sighting distance for all surveys contained by that node. Each survey (leaf node) recieves the detection function that is closest to it up the hierarchy. Thus, for common species, sufficient sightings may be available to fit detection functions deep in the hierarchy, with each function applying to only a few surveys, thereby allowing variability in detection performance between surveys to be addressed relatively finely. For rare species, so few sightings may be available that we have to pool many surveys together to try to meet Buckland's recommendation, and fit only a few coarse detection functions high in the hierarchy.

A blue Proxy Species tag indicates that so few sightings were available that, rather than ascend higher in the hierarchy to a point that we would pool grossly-incompatible surveys together, (e.g. shipboard surveys that used big-eye binoculars with those that used only naked eyes) we pooled sightings of similar species together instead. The list of species pooled is given in following sections.

Shipboard Surveys

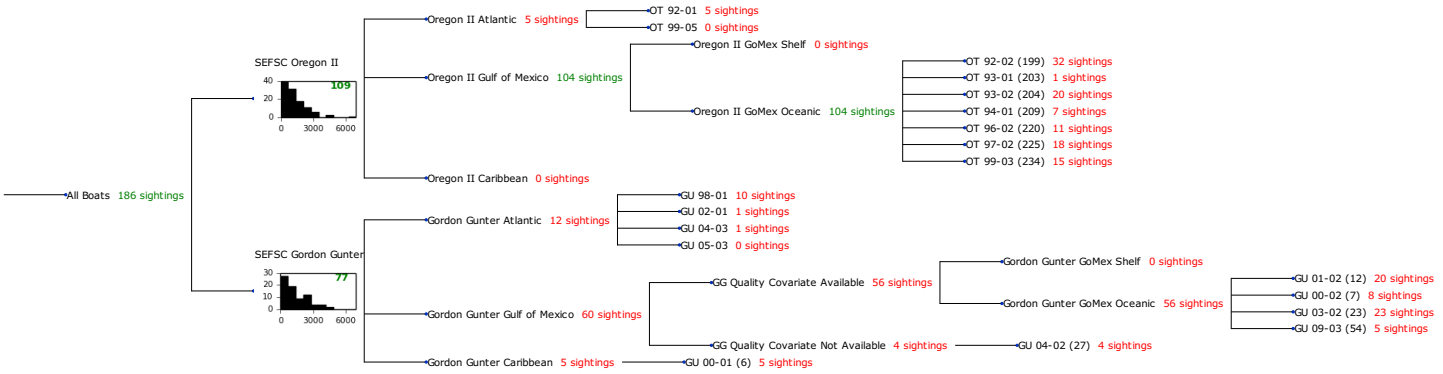


Figure 8: Detection hierarchy for shipboard surveys

SEFSC Oregon II

The sightings were right truncated at 5000m.

Covariate	Description
beaufort	Beaufort sea state.
quality	Survey-specific index of the quality of observation conditions, utilizing relevant factors other than Beaufort sea state (see methods).
size	Estimated size (number of individuals) of the sighted group.

Table 4: Covariates tested in candidate "multi-covariate distance sampling" (MCDS) detection functions.

Key	Adjustment	Order	Covariates	Succeeded	Δ AIC	Mean ESHW (m)
hn			quality, size	Yes	0.00	1934
hn			size	Yes	0.69	1947
hn			beaufort, quality, size	Yes	1.99	1935
hn			beaufort, size	Yes	2.11	1954
hr			quality	Yes	3.53	1962
hn			quality	Yes	4.35	1937
hr				Yes	4.38	1905
hn	cos	2		Yes	4.60	1714
hn				Yes	4.77	1936
hr			quality, size	Yes	4.88	1914
hr			beaufort, quality	Yes	5.53	1962
hr			size	Yes	5.86	1875
hr	poly	2		Yes	6.17	1875
hr	poly	4		Yes	6.32	1894
hr			beaufort	Yes	6.33	1909
hn	cos	3		Yes	6.46	1811
hr			beaufort, quality, size	Yes	6.86	1915
hr			beaufort, size	Yes	7.69	1874
hn	herm	4		No		
hn			beaufort	No		
hn			beaufort, quality	No		

Table 5: Candidate detection functions for SEFSC Oregon II. The first one listed was selected for the density model.

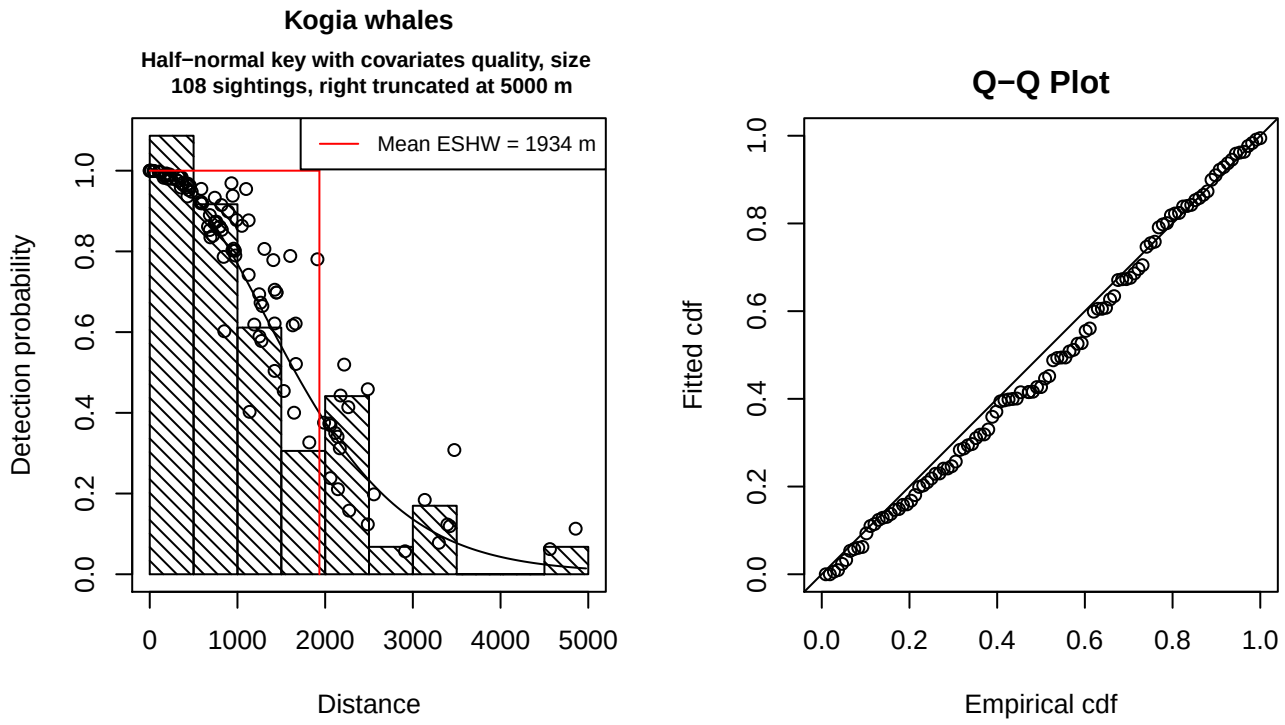


Figure 9: Detection function for SEFSC Oregon II that was selected for the density model

Statistical output for this detection function:

```
Summary for ds object
Number of observations :  108
Distance range         :  0  -  5000
AIC                    :  1738.51

Detection function:
 Half-normal key function

Detection function parameters
Scale Coefficients:
             estimate         se
(Intercept)  7.1308059 0.16525524
quality     -0.1095424 0.07073462
size         0.1551945 0.05746713

                      Estimate          SE         CV
Average p             0.3667174  0.02796449 0.07625624
N in covered region 294.5047099 32.08004956 0.10892882
```

Additional diagnostic plots:

Figure 10: Scatterplots showing the relationship between Beaufort sea state and perpendicular sighting distance, for all sightings (left) and only those not right truncated (right). The line is a simple linear regression.

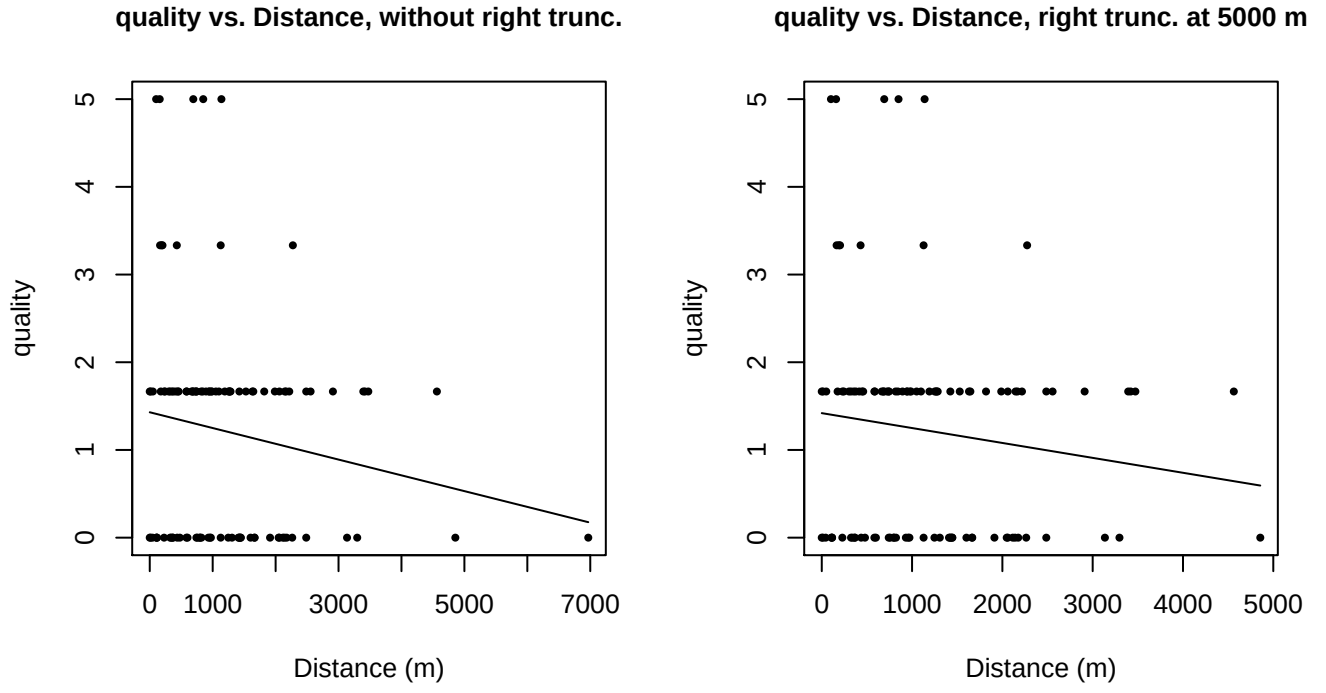


Figure 11: Scatterplots showing the relationship between the survey-specific index of the quality of observation conditions and perpendicular sighting distance, for all sightings (left) and only those not right truncated (right). Low values of the quality index correspond to better observation conditions. The line is a simple linear regression.

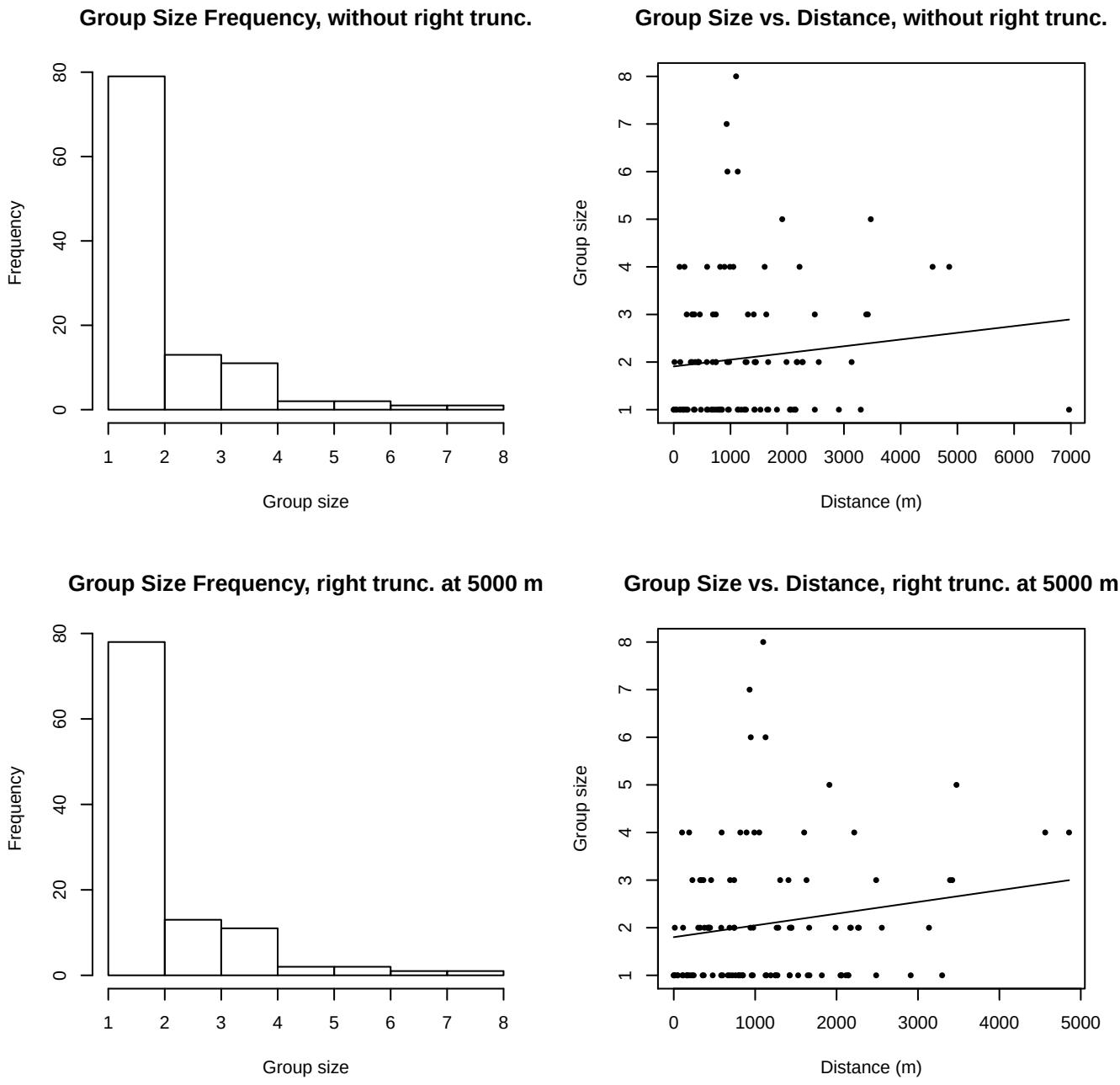


Figure 12: Histograms showing group size frequency and scatterplots showing the relationship between group size and perpendicular sighting distance, for all sightings (top row) and only those not right truncated (bottom row). In the scatterplot, the line is a simple linear regression.

SEFSC Gordon Gunter

The sightings were right truncated at 5000m.

Covariate	Description
beaufort	Beaufort sea state.
quality	Survey-specific index of the quality of observation conditions, utilizing relevant factors other than Beaufort sea state (see methods).
size	Estimated size (number of individuals) of the sighted group.

Table 6: Covariates tested in candidate "multi-covariate distance sampling" (MCDS) detection functions.

Key	Adjustment	Order	Covariates	Succeeded	Δ AIC	Mean ESHW (m)
hn			beaufort, size	Yes	0.00	2361
hn			beaufort, quality, size	Yes	0.85	2323
hr			beaufort, quality, size	Yes	0.89	2297
hr			beaufort, size	Yes	0.92	2312
hn			beaufort, quality	Yes	1.95	2353
hn			beaufort	Yes	4.31	2320
hr			beaufort, quality	Yes	5.06	2449
hn	cos	3		Yes	5.10	1866
hn				Yes	5.35	2332
hr			beaufort	Yes	6.14	2233
hn	cos	2		Yes	6.25	2074
hr			size	Yes	6.34	1976
hn			size	Yes	6.43	2328
hr	poly	4		Yes	6.76	1789
hn			quality	Yes	6.78	2332
hr				Yes	7.01	1894
hn	herm	4		Yes	7.31	2325
hr			quality, size	Yes	7.98	1966
hn			quality, size	Yes	8.35	2326
hr			quality	Yes	8.42	1870
hr	poly	2		Yes	9.01	1894
hn	cos	1		No		

Table 7: Candidate detection functions for SEFSC Gordon Gunter. The first one listed was selected for the density model.

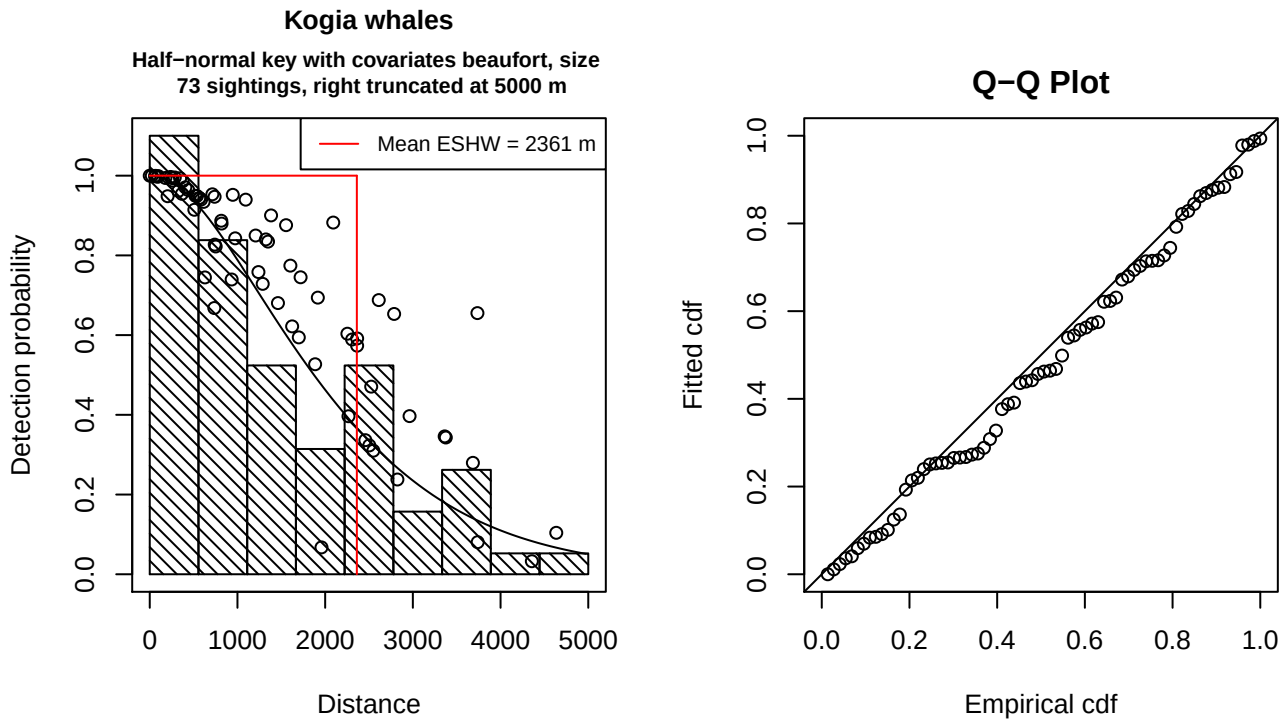


Figure 13: Detection function for SEFSC Gordon Gunter that was selected for the density model

Statistical output for this detection function:

```
Summary for ds object
Number of observations :  73
Distance range         :  0  -  5000
AIC                    :  1197.314

Detection function:
 Half-normal key function

Detection function parameters
Scale Coefficients:
             estimate        se
(Intercept)  7.4469239 0.1988146
beaufort    -0.3259967 0.0902056
size         0.2972217 0.1391281

                      Estimate          SE         CV
Average p             0.4249359  0.03836081 0.09027434
N in covered region 171.7905995 22.18392534 0.12913352
```

Additional diagnostic plots:

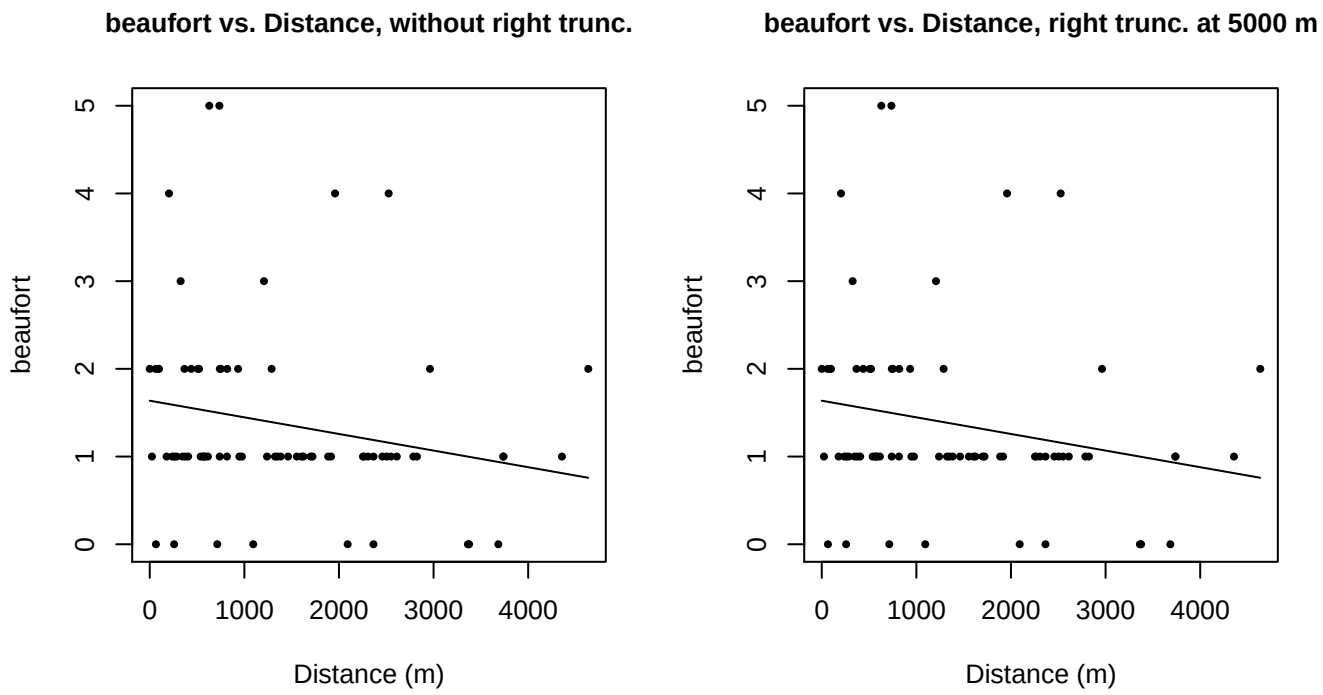


Figure 14: Scatterplots showing the relationship between Beaufort sea state and perpendicular sighting distance, for all sightings (left) and only those not right truncated (right). The line is a simple linear regression.

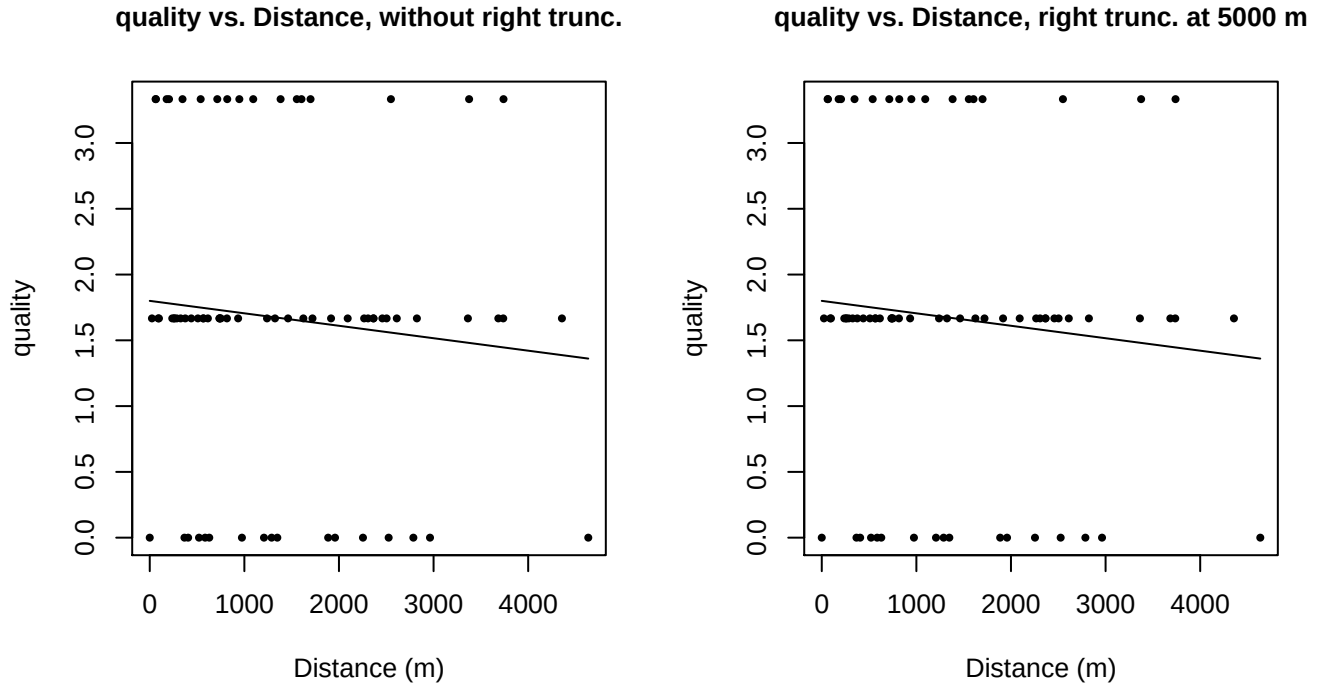


Figure 15: Scatterplots showing the relationship between the survey-specific index of the quality of observation conditions and perpendicular sighting distance, for all sightings (left) and only those not right truncated (right). Low values of the quality index correspond to better observation conditions. The line is a simple linear regression.

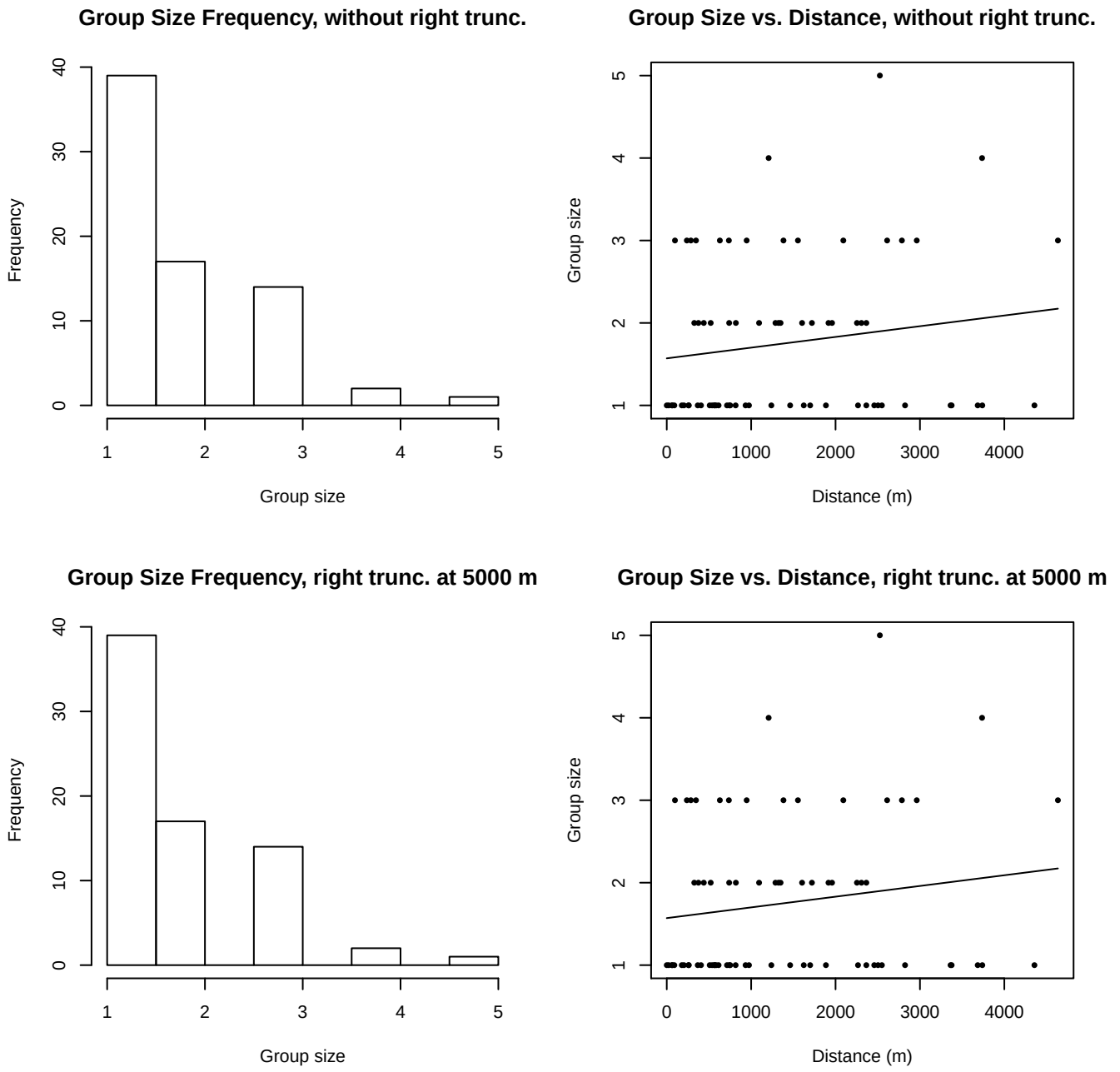


Figure 16: Histograms showing group size frequency and scatterplots showing the relationship between group size and perpendicular sighting distance, for all sightings (top row) and only those not right truncated (bottom row). In the scatterplot, the line is a simple linear regression.

Aerial Surveys

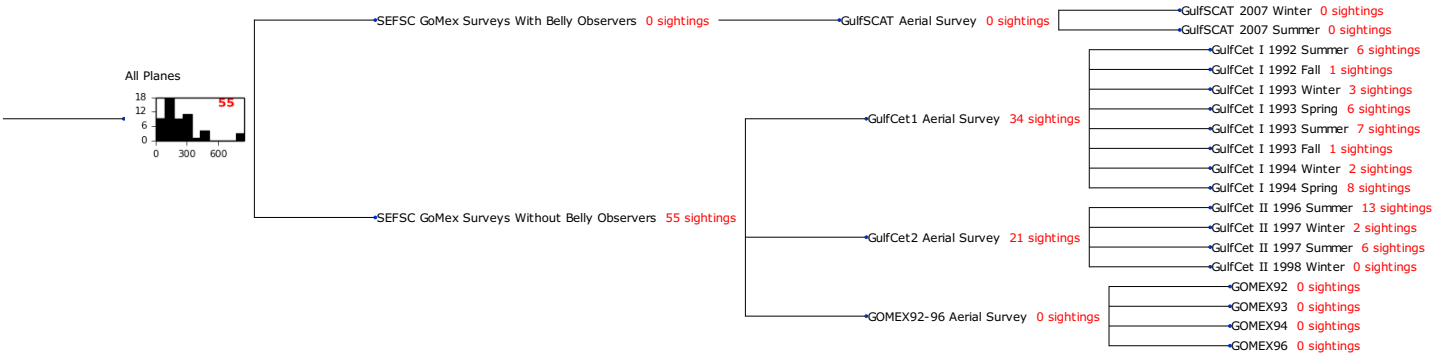


Figure 17: Detection hierarchy for aerial surveys

All Planes

The sightings were right truncated at 628m. Due to a reduced frequency of sightings close to the trackline that plausibly resulted from the behavior of the observers and/or the configuration of the survey platform, the sightings were left truncted as well. Sightings closer than 83 m to the trackline were omitted from the analysis, and it was assumed that the the area closer to the trackline than this was not surveyed. This distance was estimated by inspecting histograms of perpendicular sighting distances. The vertical sighting angles were heaped at 10 degree increments, so the candidate detection functions were fitted using linear bins scaled accordingly.

Key	Adjustment	Order	Covariates	Succeeded	Δ AIC	Mean ESHW (m)
hn				Yes	0.00	194
hn	cos	3		Yes	1.95	171
hn	herm	4		Yes	1.97	203
hn	cos	2		Yes	1.98	203
hr				Yes	2.33	258
hr	poly	4		Yes	4.32	250
hr	poly	2		Yes	4.34	259
hn	cos	1		No		

Table 8: Candidate detection functions for All Planes. The first one listed was selected for the density model.

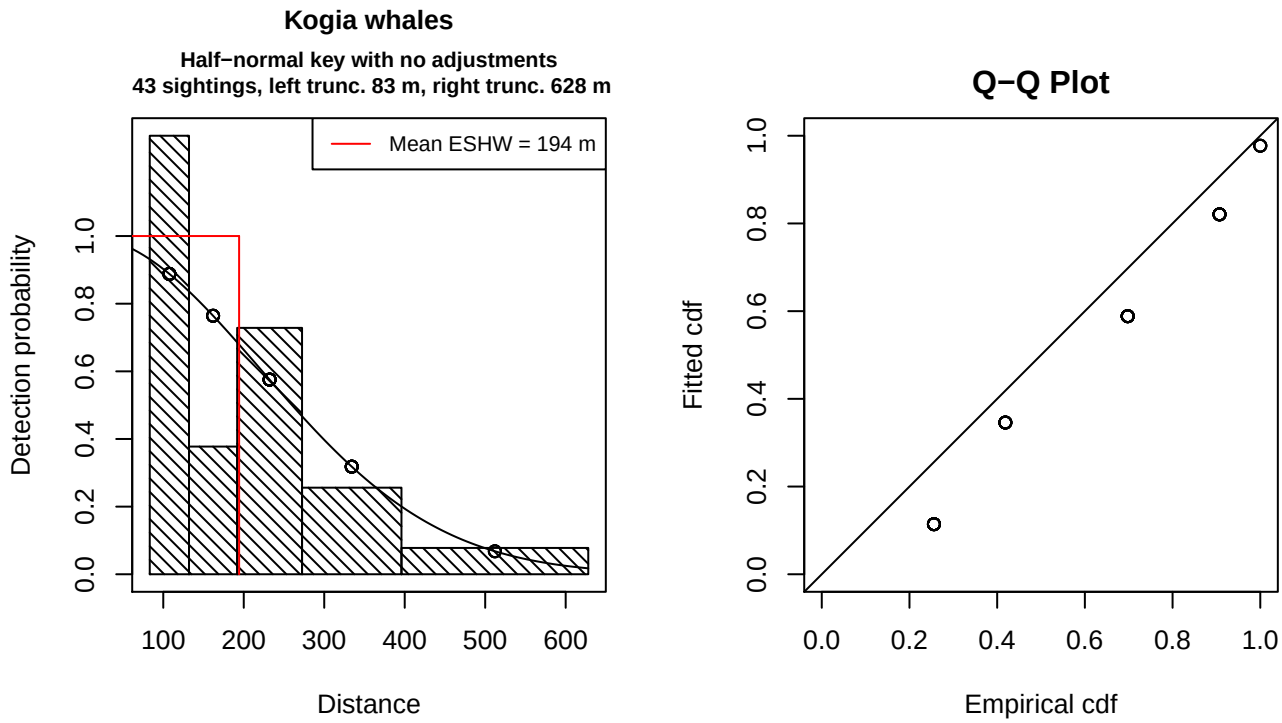


Figure 18: Detection function for All Planes that was selected for the density model

Statistical output for this detection function:

```
Summary for ds object
Number of observations :  43
Distance range         :  83.2036  -  628.0733
AIC                    :  136.7514

Detection function:
 Half-normal key function

Detection function parameters
Scale Coefficients:
            estimate        se
(Intercept) 5.397386 0.1213359

                       Estimate          SE        CV
Average p              0.3092896  0.05038511 0.1629059
N in covered region 139.0282818 28.69557606 0.2064010
```

Additional diagnostic plots:

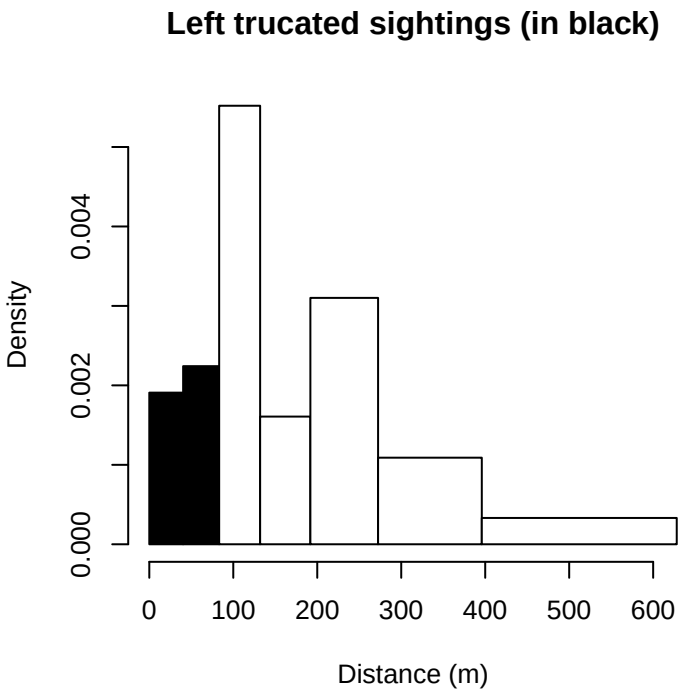


Figure 19: Density of sightings by perpendicular distance for All Planes. Black bars on the left show sightings that were left truncated.

g(0) Estimates

Platform	Surveys	Group Size	*g(0)*	Biases Addressed	Source
Shipboard	All	Any	0.35	Both	Barlow (1999)
Aerial	All	Any	0.12	Availability	Barlow (1999)

Table 9: Estimates of *g(0)* used in this density model.

No survey-specific g(0) estimates were available for shipboard surveys in the Gulf of Mexico, where Kogia are relatively abundant. In our east coast study area, Palka (2006) reported estimates of 0.46 for the 1998 Abel-J cruise and 0.29 for the 2004 Endeavor cruise, but these were based on a limited number of animals and were pooled with beaked whales. Instead of those, we used an estimate from Barlow (1999), who used a simulation-based approach to estimate g(0) for Kogia observed from shipboard surveys that utilized 25x binoculars. This estimate accounted for both availability and perception bias.

No estimate of g(0) was available in the literature for Kogia sighted on aerial surveys. Kogia are long-diving animals (Barlow 1999), thus availability bias is likely to be substantial. Utilizing equation (3) of Carretta et al. (2000) (which follows Barlow et al. 1988), we computed the availability bias component of g(0) from the median duration of surfacing series and long dives (78 s and 10.9 min) for Kogia near California reported by Barlow (1999). We did not incorporate an estimate of perception bias, thus our g(0) estimate is likely to be biased high.

Density Models

The two extant species of Kogia, the dwarf sperm whale (Kogia sima) and the pygmy sperm whale (Kogia breviceps), are very difficult for observers to distinguish at sea (Jefferson and Schiro 1997). Both species occur worldwide in tropical to temperate seas, generally in oceanic waters (Waring et al. 2013; Bloodworth and Odell 2008; Willis and Baird 1998). Although pygmy sperm whales are considered a more temperate species, the habitats and diets of the two species overlap substantially; they are often found over the continental slope, possibly to feed on cephalopods, a staple of their diet (Bloodworth and Odell 2008).

The large majority of sightings reported by the surveys included in our study reported "dwarf or pygmy sperm whale" as the taxonomic identification, and too few fully-identified sightings were reported to fit a habiat-based model for classifying the ambiguous ones. But given the apparent overlap in their habitats, we are uncertain that such an approach would be successful anyway. In any case, we modeled both species as a guild, as NOAA as historically done (Mullin and Fulling 2004; Mullin 2007).

In the Gulf of Mexico study area, all sightings reported over the study period (1992-2009) occurred off the continental shelf, with the shallowest reported at 186m. We found no definitive descriptions in the literature of seasonal movements by Kogia in the Gulf of Mexico. Accordingly, we fitted a year-round model to off-shelf waters, defined here as those deeper than the 100m isobath.

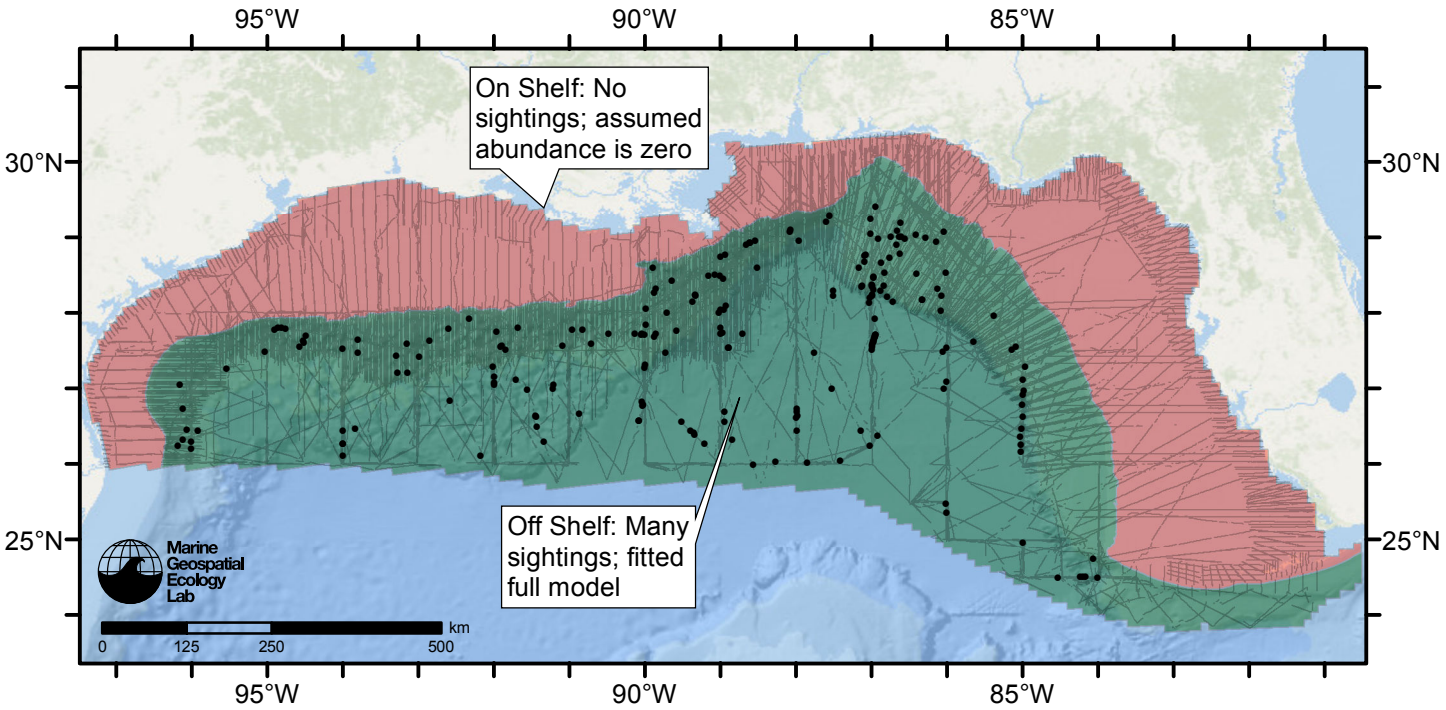


Figure 20: Kogia whales density model schematic. All on-effort sightings are shown, including those that were truncated when detection functions were fitted.

Climatological Model

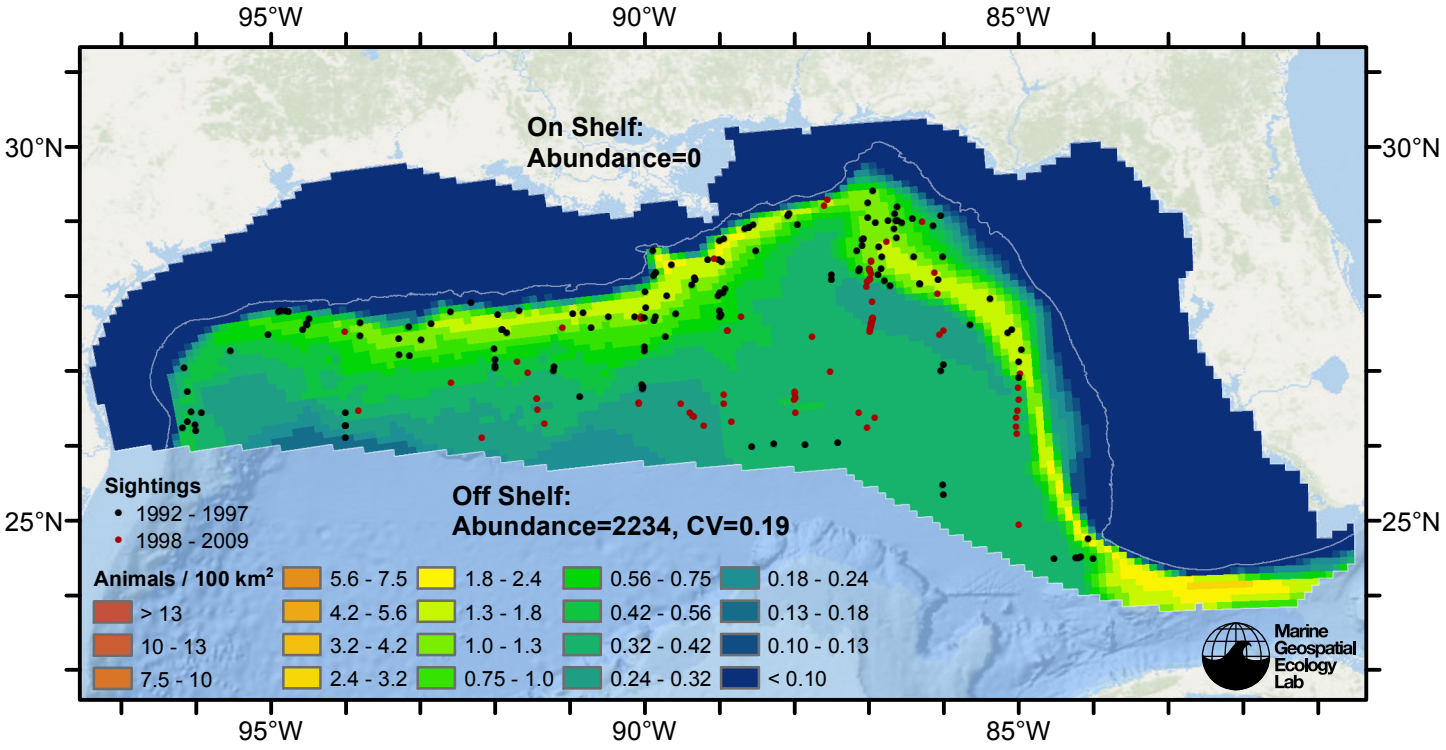


Figure 21: Kogia whales density predicted by the climatological model that explained the most deviance. Pixels are 10x10 km. The legend gives the estimated individuals per pixel; breaks are logarithmic. Abundance for each region was computed by summing the density cells occuring in that region.

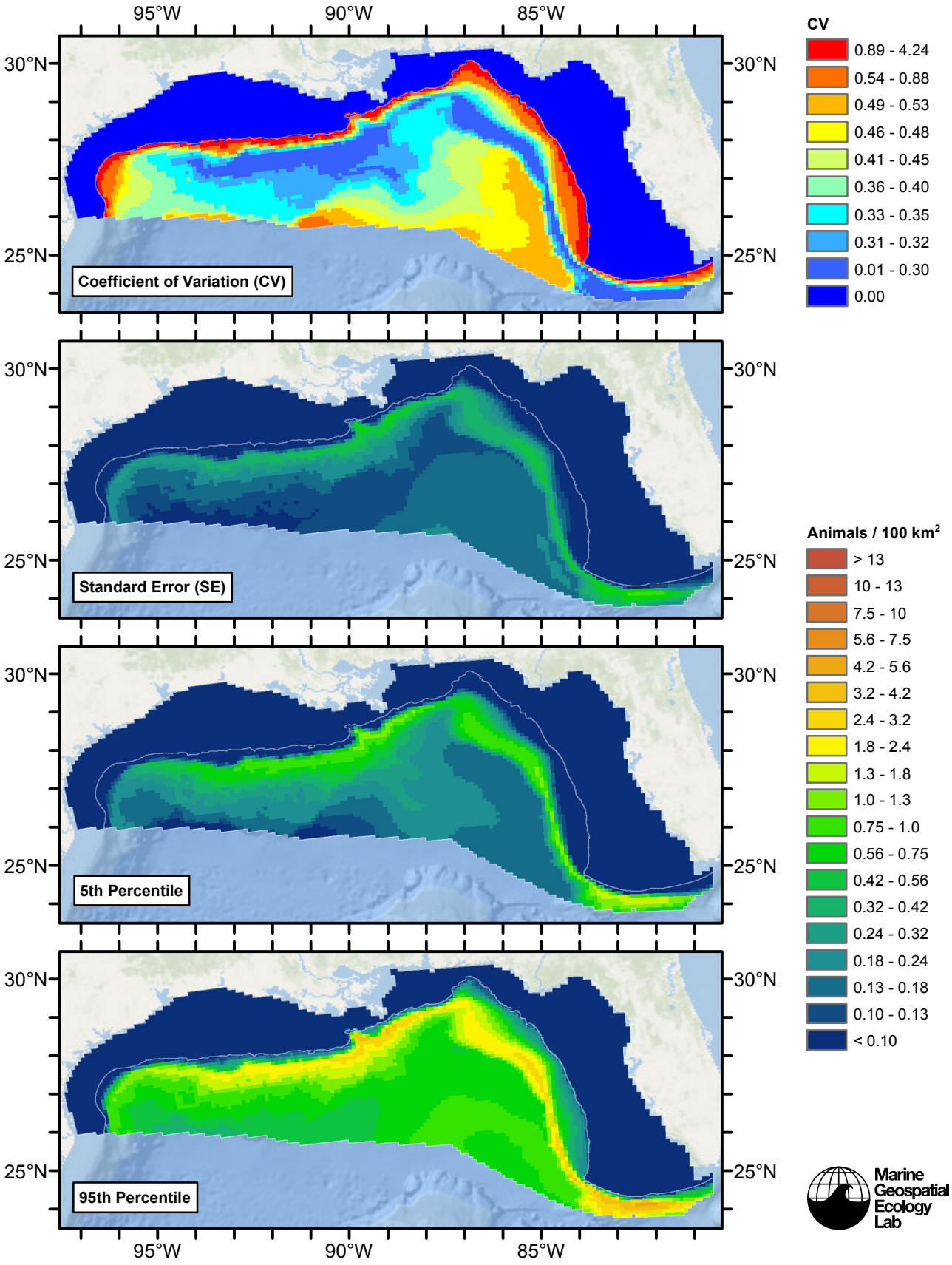


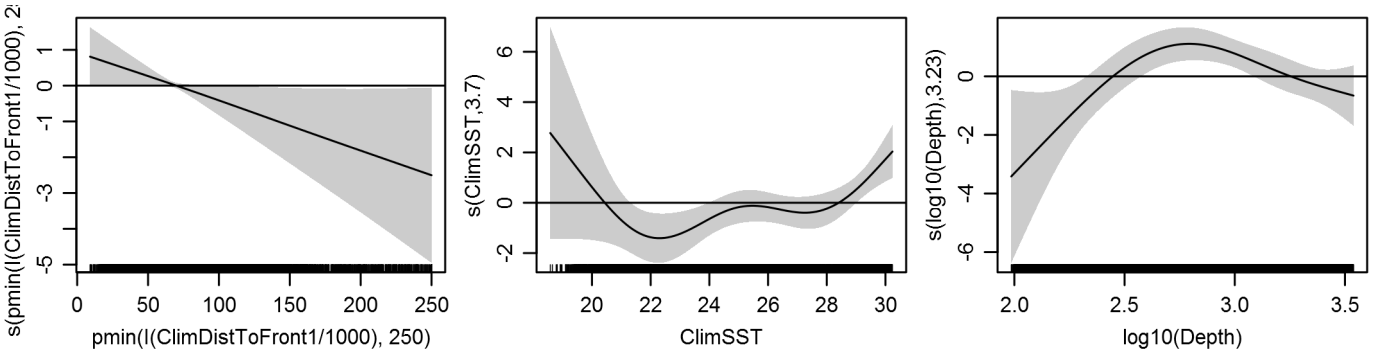
Figure 22: Estimated uncertainty for the climatological model that explained the most deviance. These estimates only incorporate the statistical uncertainty estimated for the spatial model (by the R mgcv package). They do not incorporate uncertainty in the detection functions, g(0) estimates, predictor variables, and so on.

Off Shelf

Statistical output

```
Rscript.exe:   Pearson scale estimate maybe unstable. See ?gam.scale.


Family: Tweedie(1.36)
Link function: log

Formula:
abundance ~ offset(log(area_km2)) + s(log10(Depth), bs = "ts",
    k = 5) + s(ClimSST, bs = "ts", k = 5) + s(pmin(I(ClimDistToFront1/1000),
    250), bs = "ts", k = 5)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -5.4613     0.2222  -24.58   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                       edf Ref.df     F  p-value
s(log10(Depth))                      3.233      4 4.675 0.000207 ***
s(ClimSST)                           3.696      4 4.338 0.000834 ***
s(pmin(I(ClimDistToFront1/1000), 250)) 1.002      4 1.128 0.028609 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  -0.0193   Deviance explained = 13.6%
-REML = 1636.8  Scale est. = 168.71    n = 14455

All predictors were significant. This is the final model.
Creating term plots.
Diagnostic output from gam.check():

Method: REML   Optimizer: outer newton
full convergence after 11 iterations.
Gradient range [-0.0004255749,0.0004117596]
(score 1636.8 & scale 168.7138).
Hessian positive definite, eigenvalue range [0.3923256,466.5722].
Model rank =  13 / 13

Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.

                                       k'   edf k-index p-value
s(log10(Depth))                      4.000 3.233   0.648    0.05
s(ClimSST)                           4.000 3.696   0.654    0.06
s(pmin(I(ClimDistToFront1/1000), 250)) 4.000 1.002   0.669    0.55

Predictors retained during the model selection procedure: Depth, ClimSST, ClimDistToFront1

Predictors dropped during the model selection procedure: Slope
```

Model term plots



Diagnostic plots

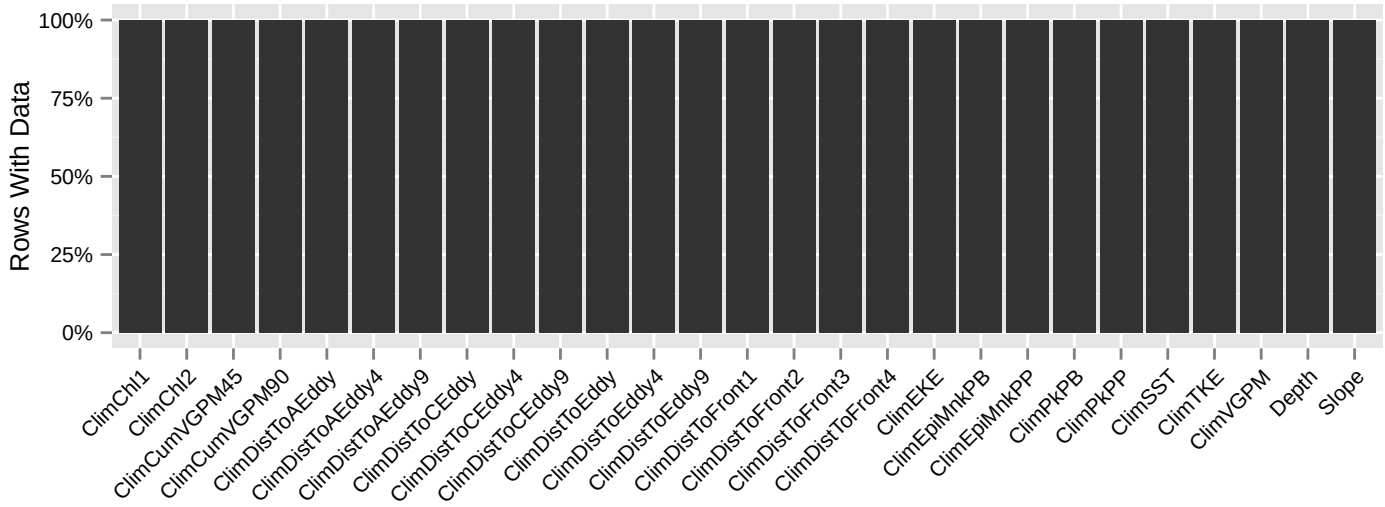


Figure 23: Segments with predictor values for the Kogia whales Climatological model, Off Shelf. This plot is used to assess how many segments would be lost by including a given predictor in a model.

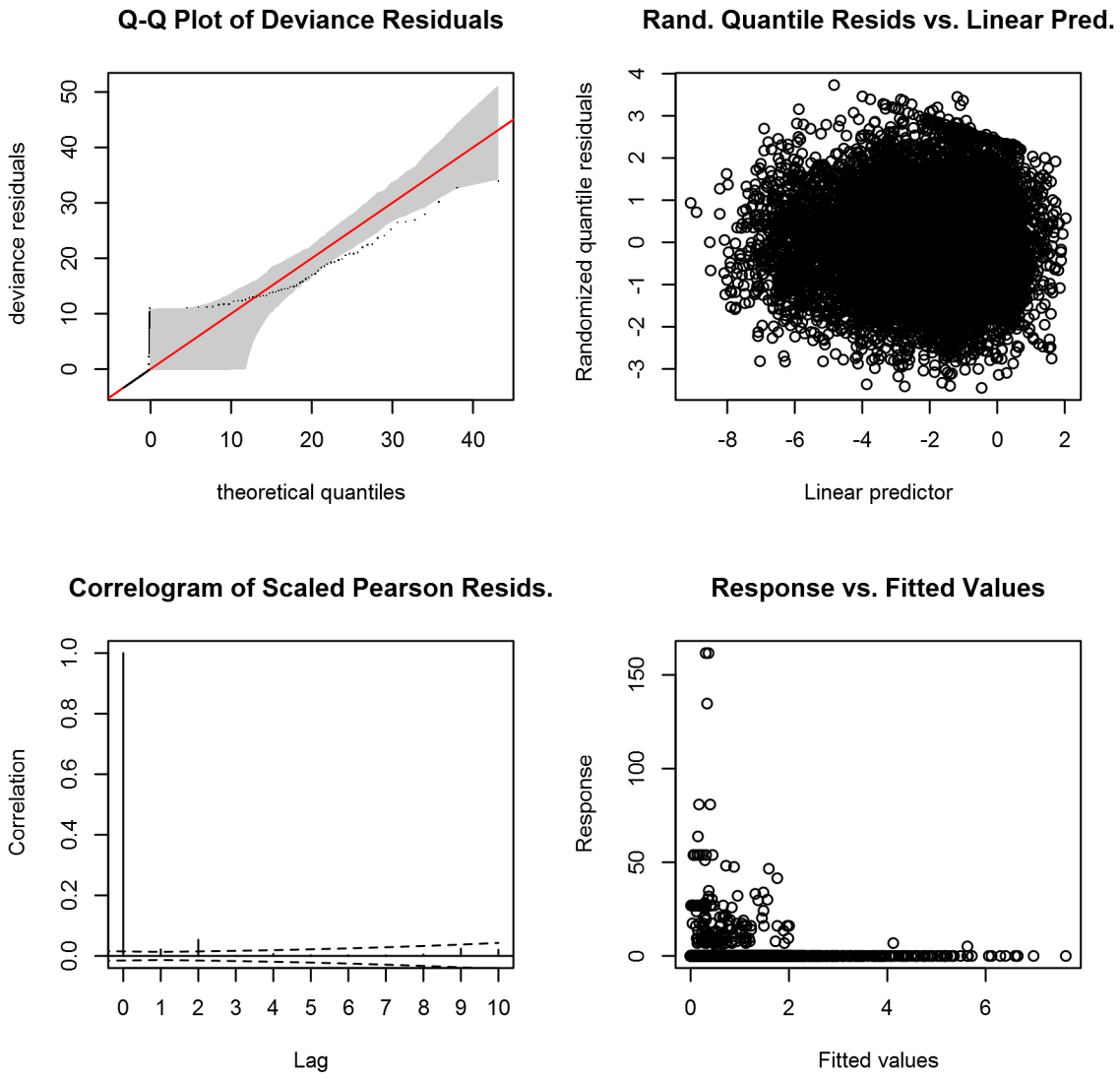


Figure 24: Statistical diagnostic plots for the Kogia whales Climatological model, Off Shelf.

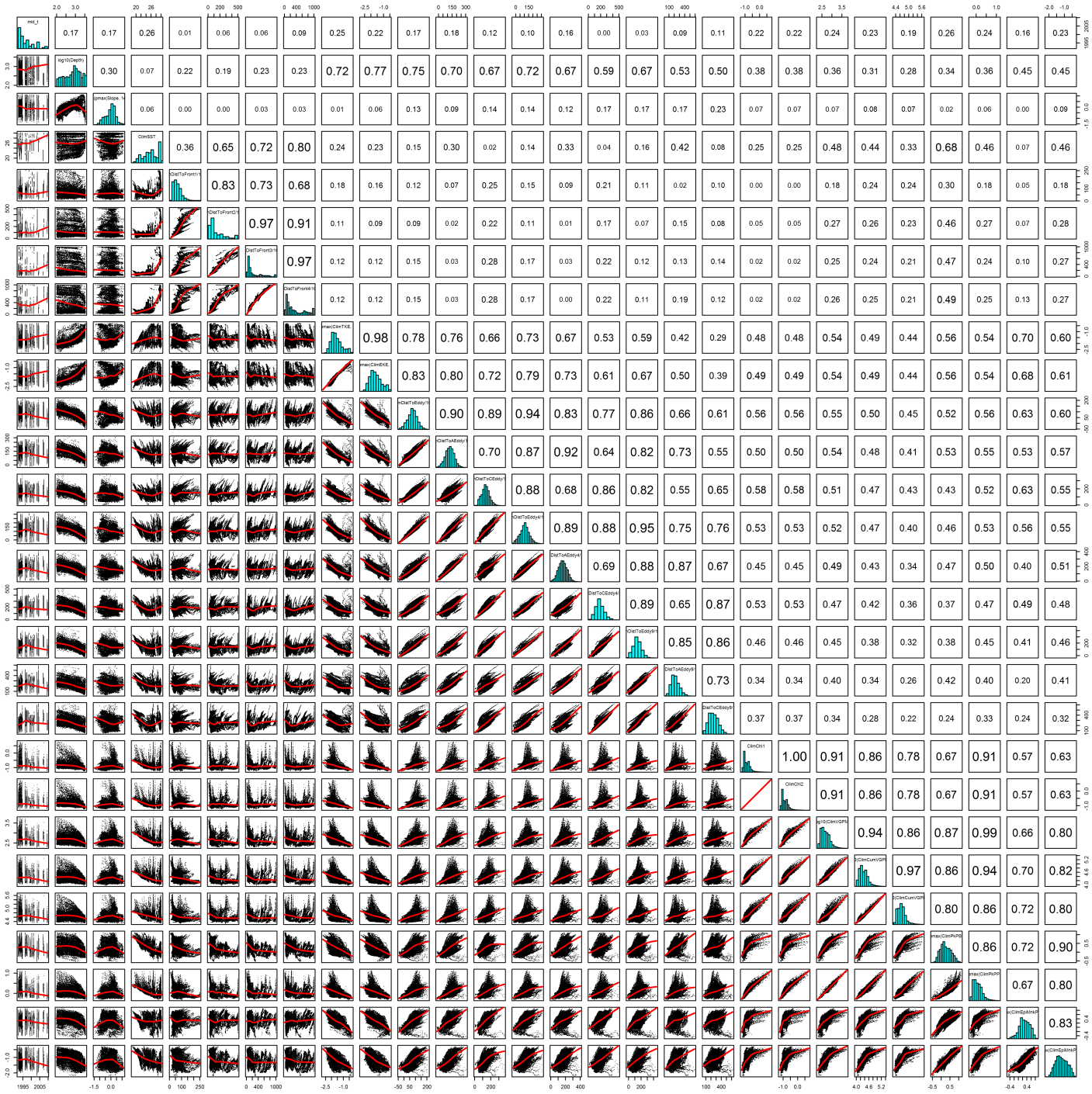


Figure 25: Scatterplot matrix for the Kogia whales Climatological model, Off Shelf. This plot is used to inspect the distribution of predictors (via histograms along the diagonal), simple correlation between predictors (via pairwise Pearson coefficients above the diagonal), and linearity of predictor correlations (via scatterplots below the diagonal). This plot is best viewed at high magnification.

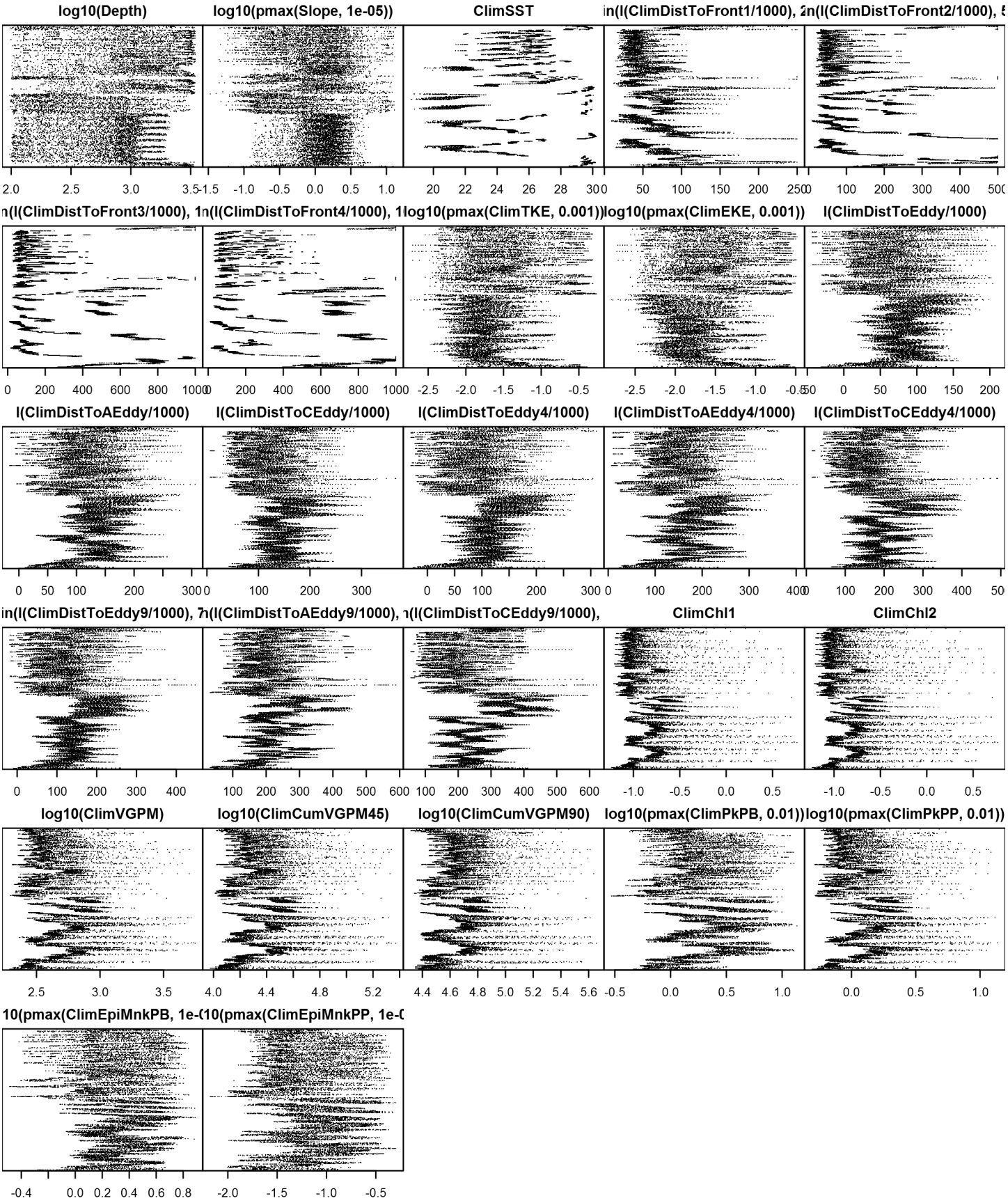


Figure 26: Dotplot for the Kogia whales Climatological model, Off Shelf. This plot is used to check for suspicious patterns and outliers in the data. Points are ordered vertically by transect ID, sequentially in time.

On Shelf

Density assumed to be 0 in this region.

Contemporaneous Model

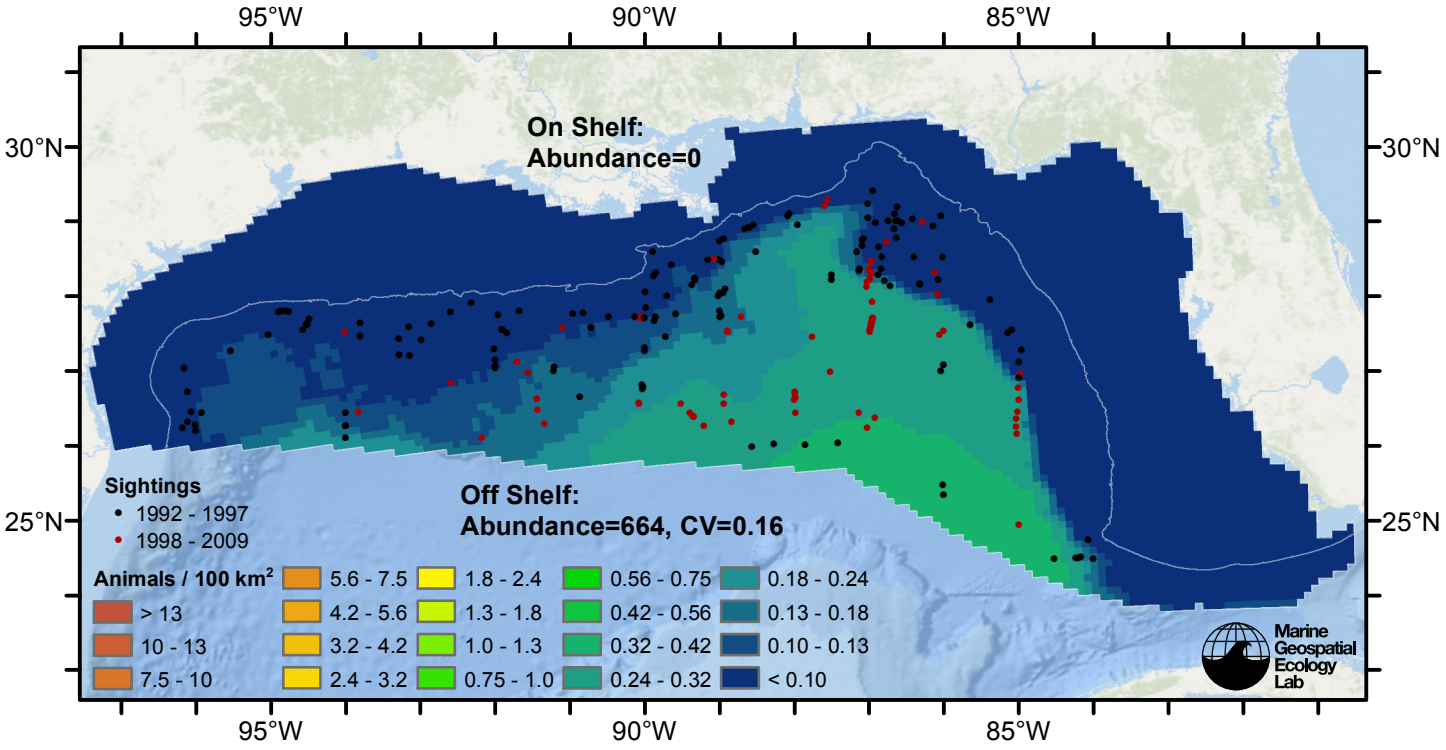


Figure 27: Kogia whales density predicted by the contemporaneous model that explained the most deviance. Pixels are 10x10 km. The legend gives the estimated individuals per pixel; breaks are logarithmic. Abundance for each region was computed by summing the density cells occuring in that region.

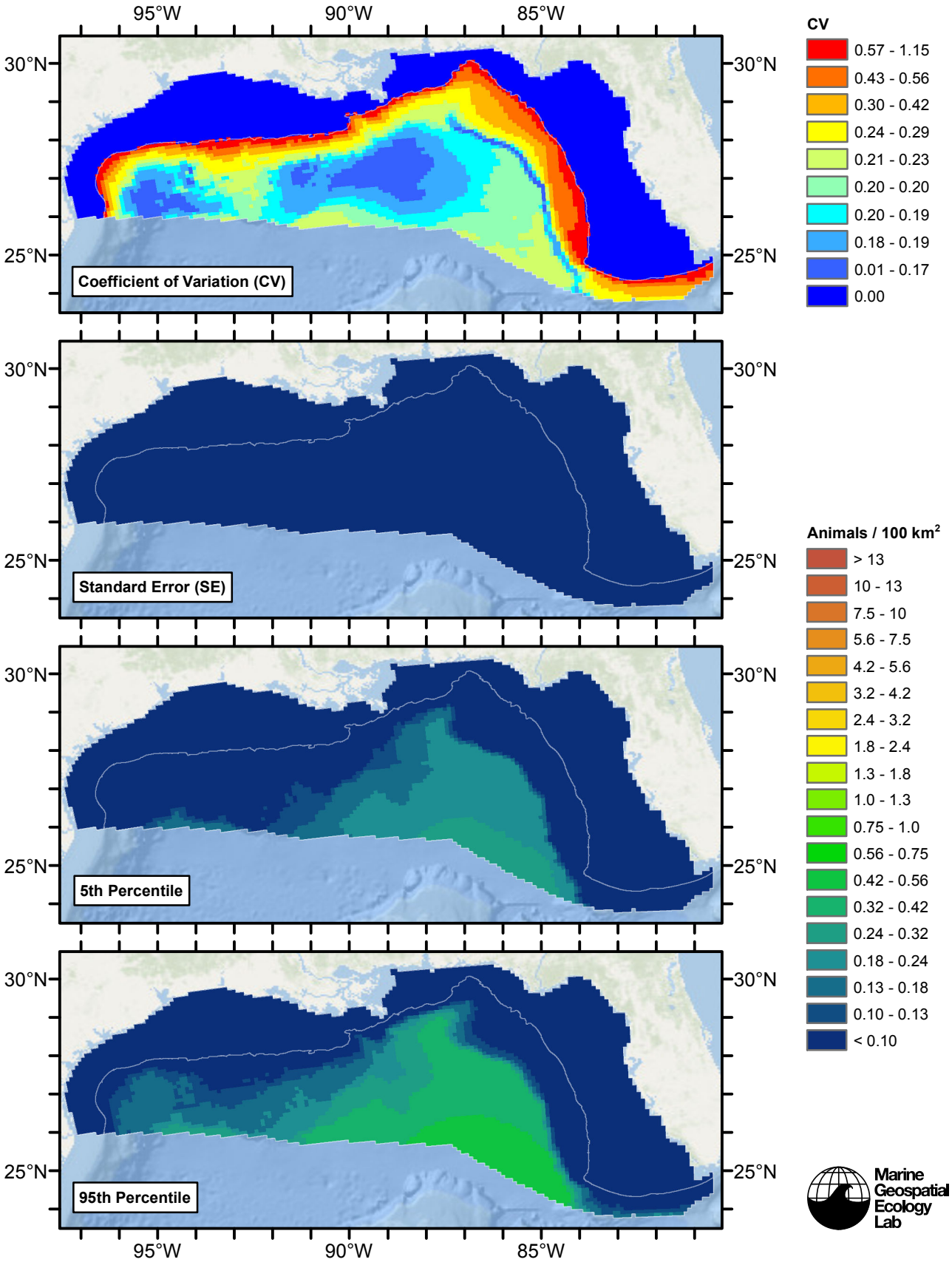


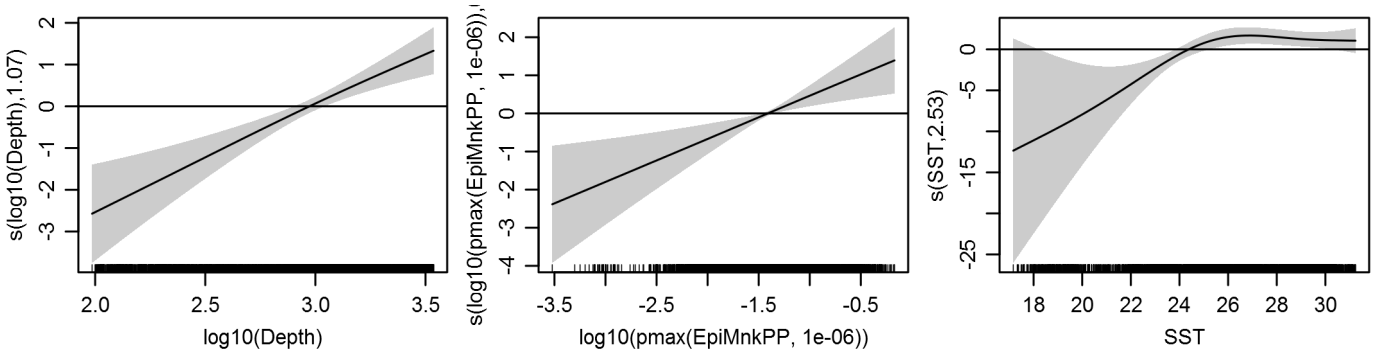
Figure 28: Estimated uncertainty for the contemporaneous model that explained the most deviance. These estimates only incorporate the statistical uncertainty estimated for the spatial model (by the R mgcv package). They do not incorporate uncertainty in the detection functions, g(0) estimates, predictor variables, and so on.

Off Shelf

Statistical output

```
Rscript.exe: This is mgcv 1.8-2. For overview type 'help("mgcv-package")'.


Family: Tweedie(1.36)
Link function: log

Formula:
abundance ~ offset(log(area_km2)) + s(log10(Depth), bs = "ts",
    k = 5) + s(SST, bs = "ts", k = 5) + s(log10(pmax(EpiMnkPP,
    1e-06)), bs = "ts", k = 5)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -7.8189     0.4686  -16.68   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                  edf Ref.df     F  p-value
s(log10(Depth))                 1.0712      4 5.867 5.89e-07 ***
s(SST)                          2.5306      4 3.866 0.000473 ***
s(log10(pmax(EpiMnkPP, 1e-06))) 0.9374      4 2.847 0.000356 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.0115   Deviance explained =   17%
-REML = 472.65  Scale est. = 25.654    n = 4219

All predictors were significant. This is the final model.
Creating term plots.
Diagnostic output from gam.check():

Method: REML   Optimizer: outer newton
full convergence after 14 iterations.
Gradient range [-2.428014e-08,-5.748357e-11]
(score 472.6478 & scale 25.65449).
Hessian positive definite, eigenvalue range [0.3371202,152.6771].
Model rank =  13 / 13

Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.

                                  k'   edf k-index p-value
s(log10(Depth))                 4.000 1.071   0.695    0.00
s(SST)                          4.000 2.531   0.676    0.00
s(log10(pmax(EpiMnkPP, 1e-06))) 4.000 0.937   0.740    0.43

Predictors retained during the model selection procedure: Depth, SST, EpiMnkPP

Predictors dropped during the model selection procedure: Slope, DistToFront2, TKE, DistToEddy
```

Model term plots



Diagnostic plots

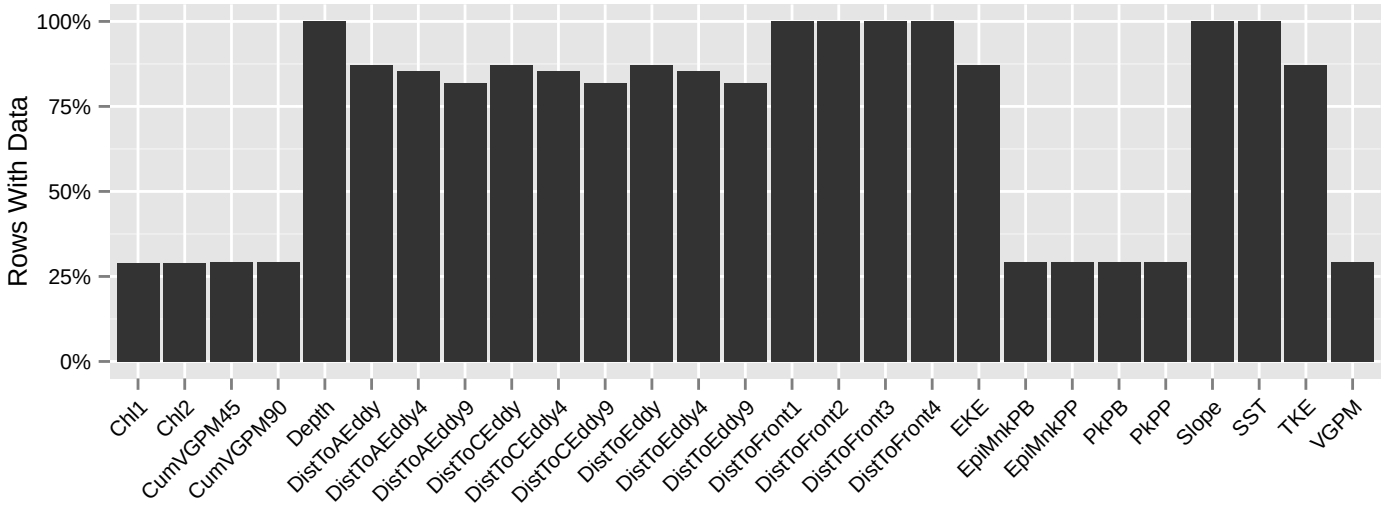


Figure 29: Segments with predictor values for the Kogia whales Contemporaneous model, Off Shelf. This plot is used to assess how many segments would be lost by including a given predictor in a model.

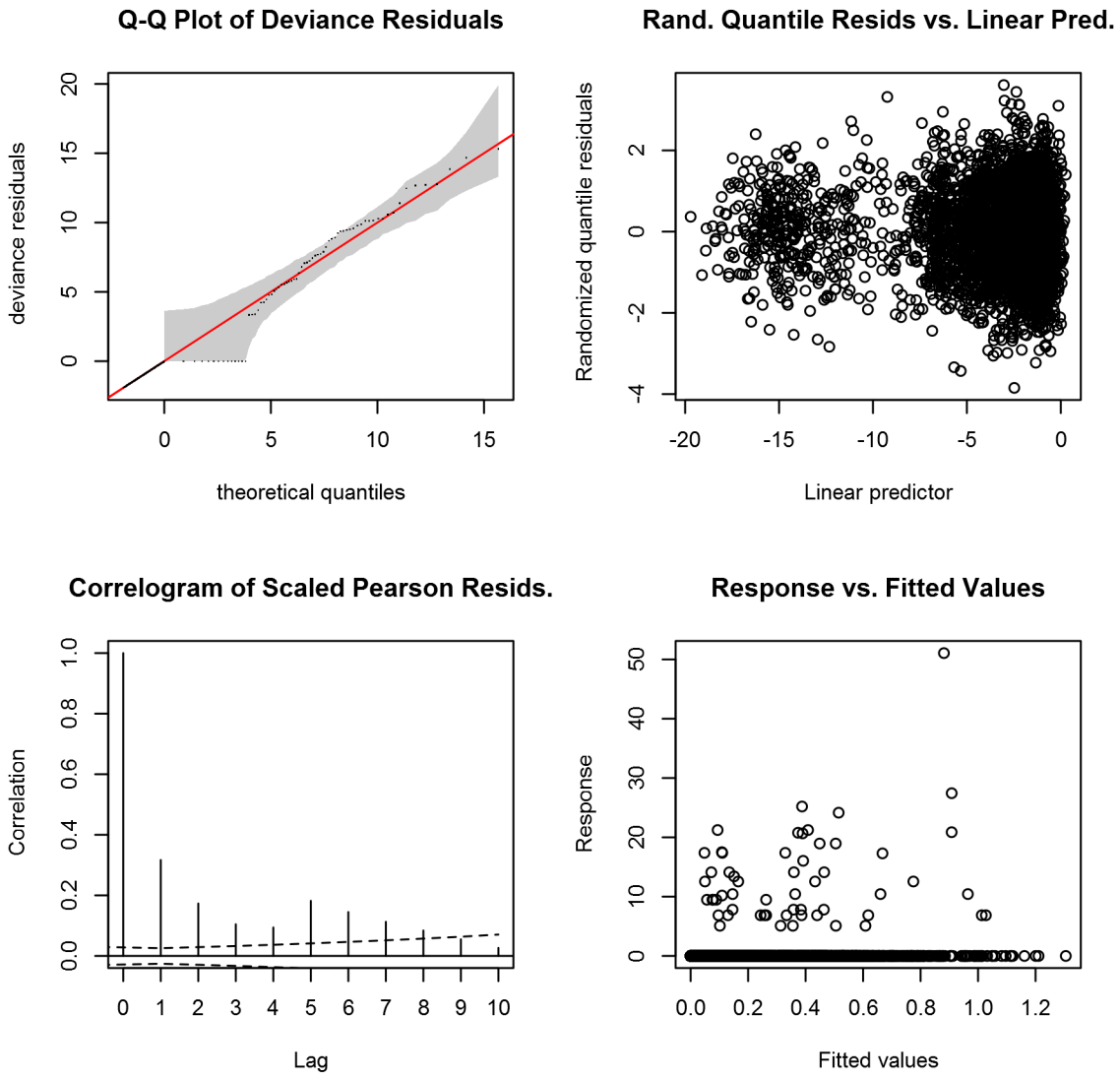


Figure 30: Statistical diagnostic plots for the Kogia whales Contemporaneous model, Off Shelf.

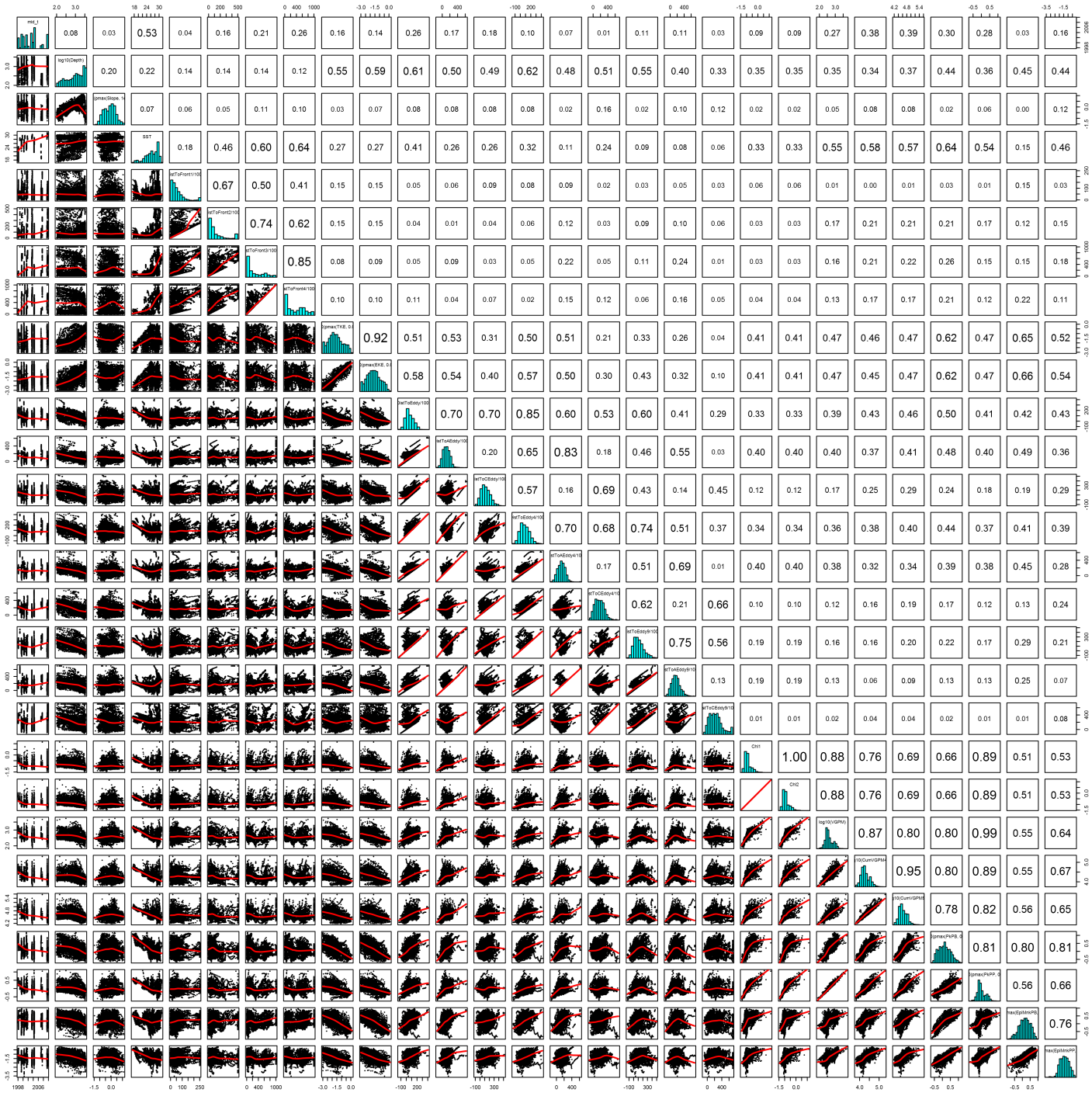


Figure 31: Scatterplot matrix for the Kogia whales Contemporaneous model, Off Shelf. This plot is used to inspect the distribution of predictors (via histograms along the diagonal), simple correlation between predictors (via pairwise Pearson coefficients above the diagonal), and linearity of predictor correlations (via scatterplots below the diagonal). This plot is best viewed at high magnification.

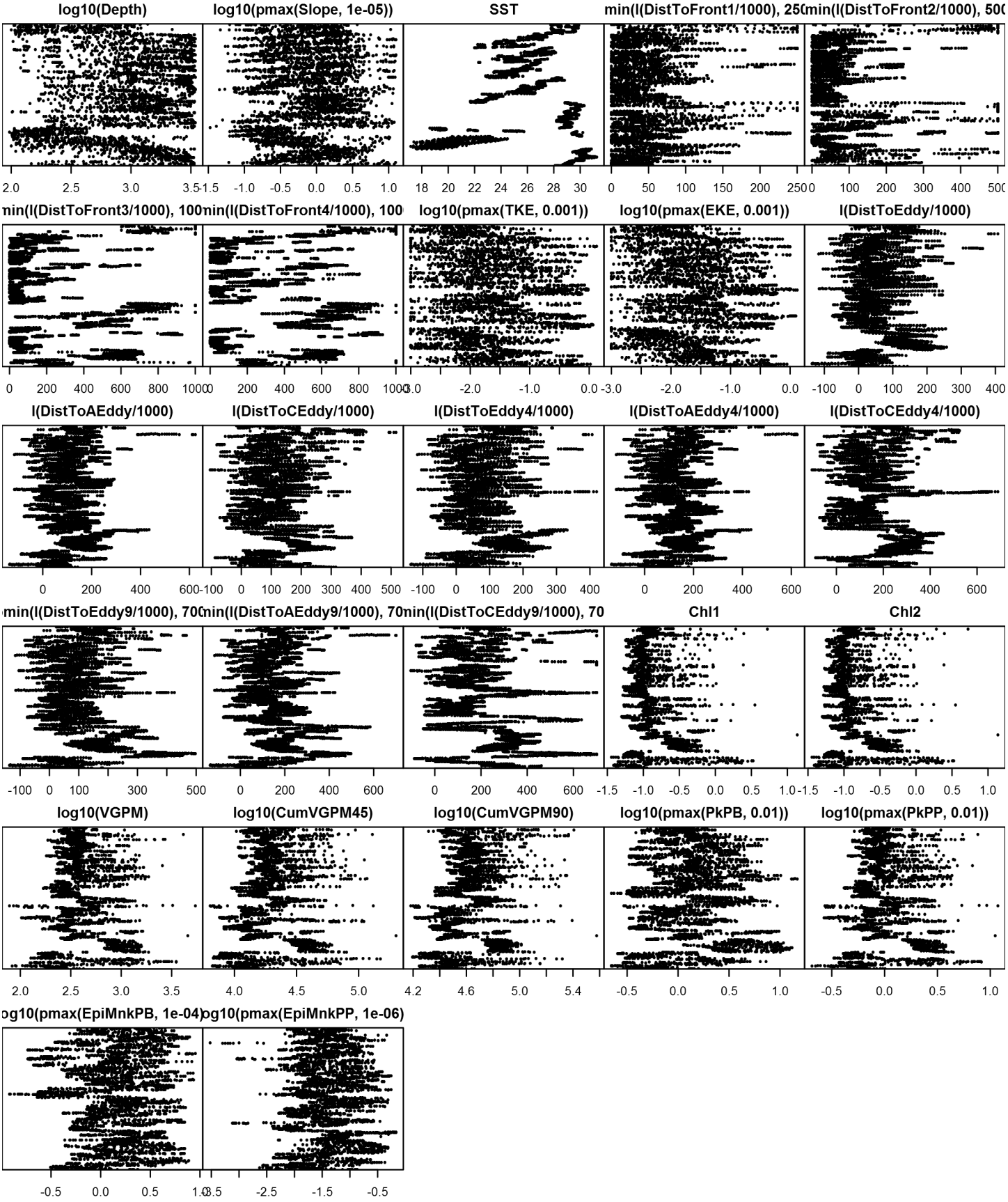


Figure 32: Dotplot for the Kogia whales Contemporaneous model, Off Shelf. This plot is used to check for suspicious patterns and outliers in the data. Points are ordered vertically by transect ID, sequentially in time.

On Shelf

Density assumed to be 0 in this region.

Climatological Same Segments Model

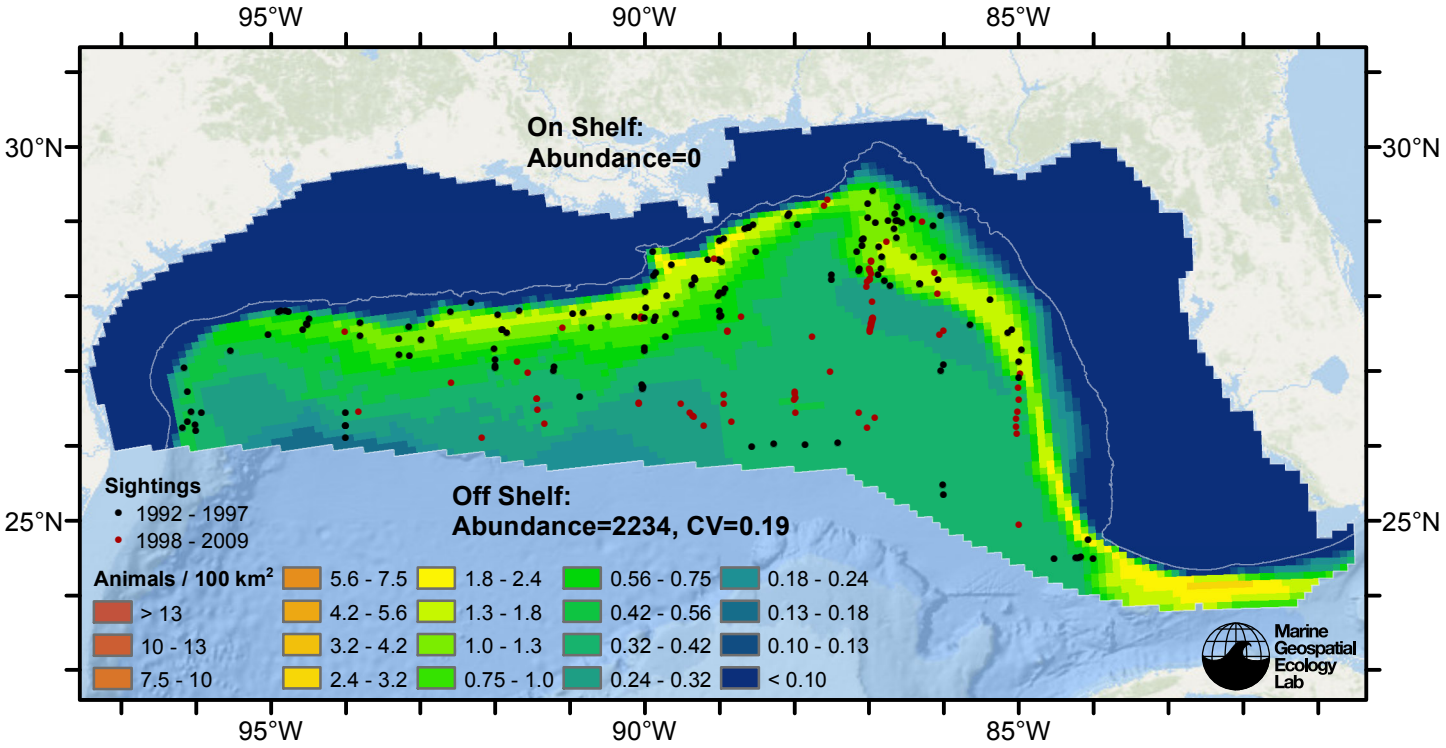


Figure 33: Kogia whales density predicted by the climatological same segments model that explained the most deviance. Pixels are 10x10 km. The legend gives the estimated individuals per pixel; breaks are logarithmic. Abundance for each region was computed by summing the density cells occuring in that region.

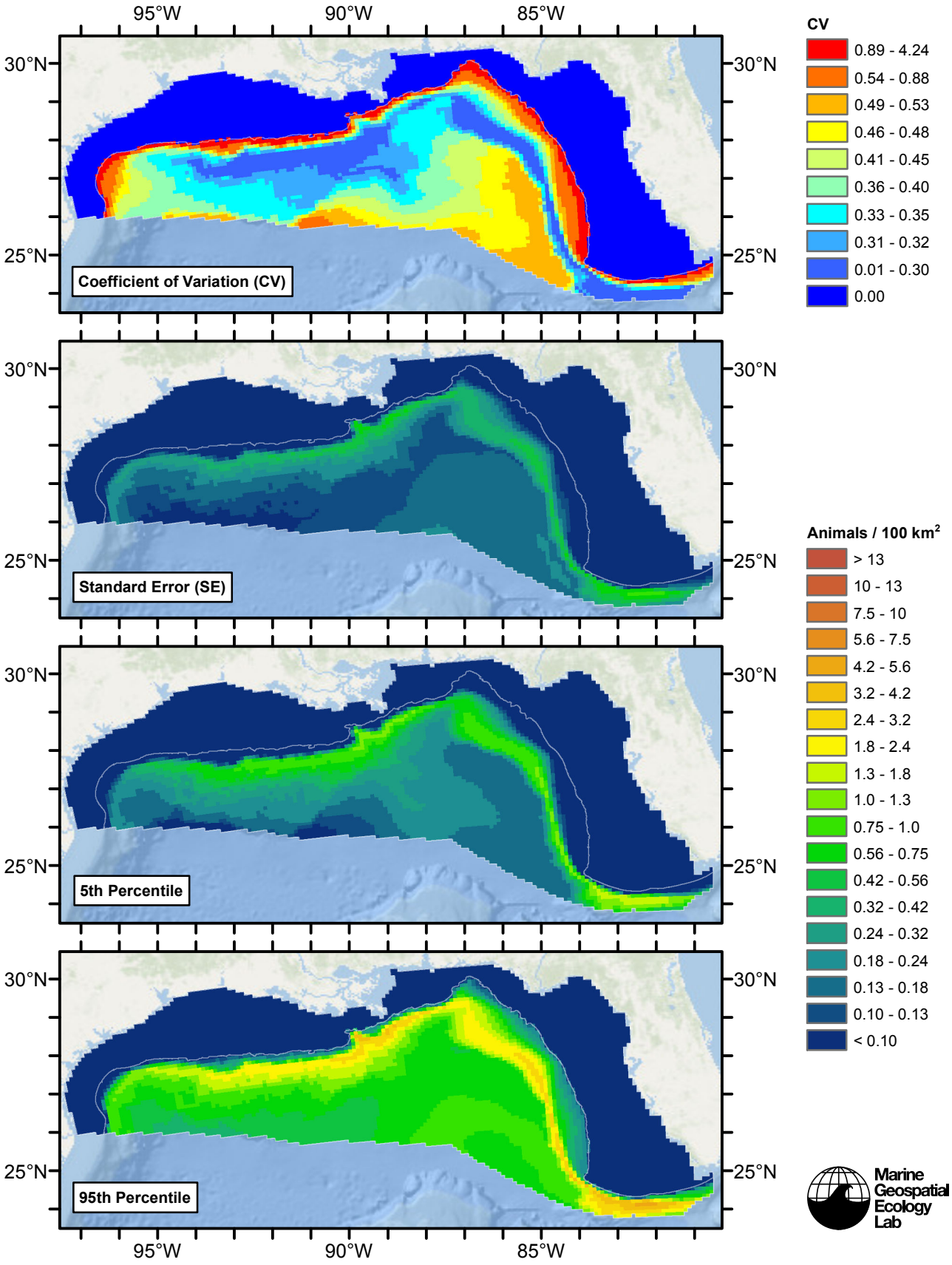


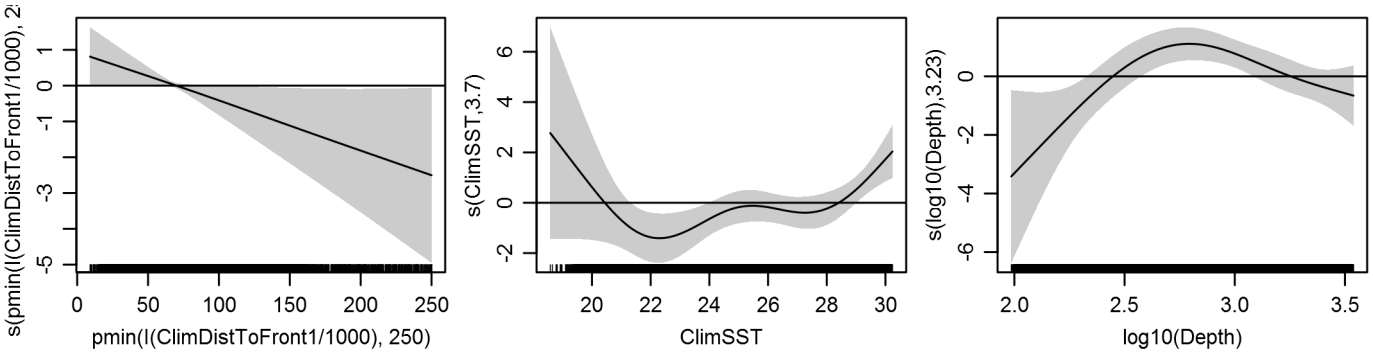
Figure 34: Estimated uncertainty for the climatological same segments model that explained the most deviance. These estimates only incorporate the statistical uncertainty estimated for the spatial model (by the R mgcv package). They do not incorporate uncertainty in the detection functions, g(0) estimates, predictor variables, and so on.

Off Shelf

Statistical output

```
Rscript.exe:   Pearson scale estimate maybe unstable. See ?gam.scale.


Family: Tweedie(1.36)
Link function: log

Formula:
abundance ~ offset(log(area_km2)) + s(log10(Depth), bs = "ts",
    k = 5) + s(ClimSST, bs = "ts", k = 5) + s(pmin(I(ClimDistToFront1/1000),
    250), bs = "ts", k = 5)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -5.4613     0.2222  -24.58   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                                       edf Ref.df     F  p-value
s(log10(Depth))                      3.233      4 4.675 0.000207 ***
s(ClimSST)                           3.696      4 4.338 0.000834 ***
s(pmin(I(ClimDistToFront1/1000), 250)) 1.002      4 1.128 0.028609 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  -0.0193   Deviance explained = 13.6%
-REML = 1636.8  Scale est. = 168.71    n = 14455

All predictors were significant. This is the final model.
Creating term plots.
Diagnostic output from gam.check():

Method: REML   Optimizer: outer newton
full convergence after 11 iterations.
Gradient range [-0.0004255749,0.0004117596]
(score 1636.8 & scale 168.7138).
Hessian positive definite, eigenvalue range [0.3923256,466.5722].
Model rank =  13 / 13

Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.

                                       k'   edf k-index p-value
s(log10(Depth))                      4.000 3.233   0.621    0.04
s(ClimSST)                           4.000 3.696   0.628    0.10
s(pmin(I(ClimDistToFront1/1000), 250)) 4.000 1.002   0.627    0.08

Predictors retained during the model selection procedure: Depth, ClimSST, ClimDistToFront1

Predictors dropped during the model selection procedure: Slope
```

Model term plots



Diagnostic plots

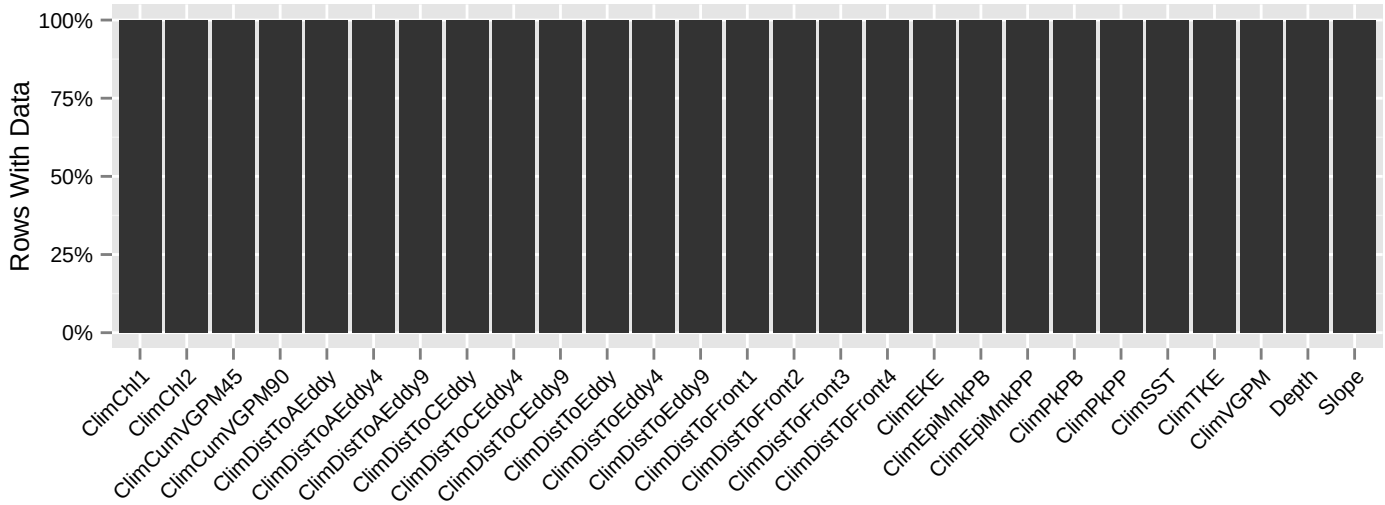


Figure 35: Segments with predictor values for the Kogia whales Climatological model, Off Shelf. This plot is used to assess how many segments would be lost by including a given predictor in a model.

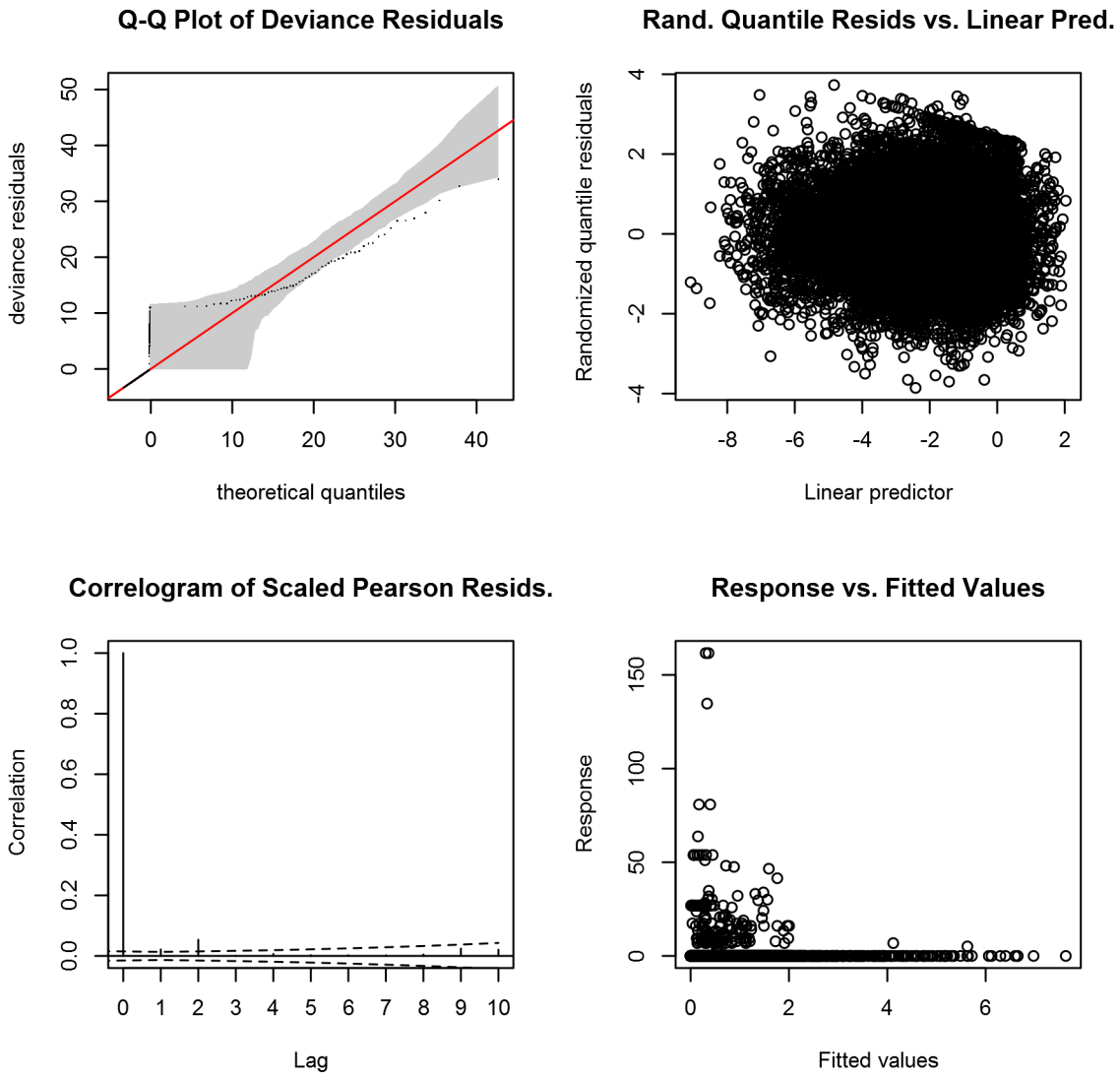


Figure 36: Statistical diagnostic plots for the Kogia whales Climatological model, Off Shelf.

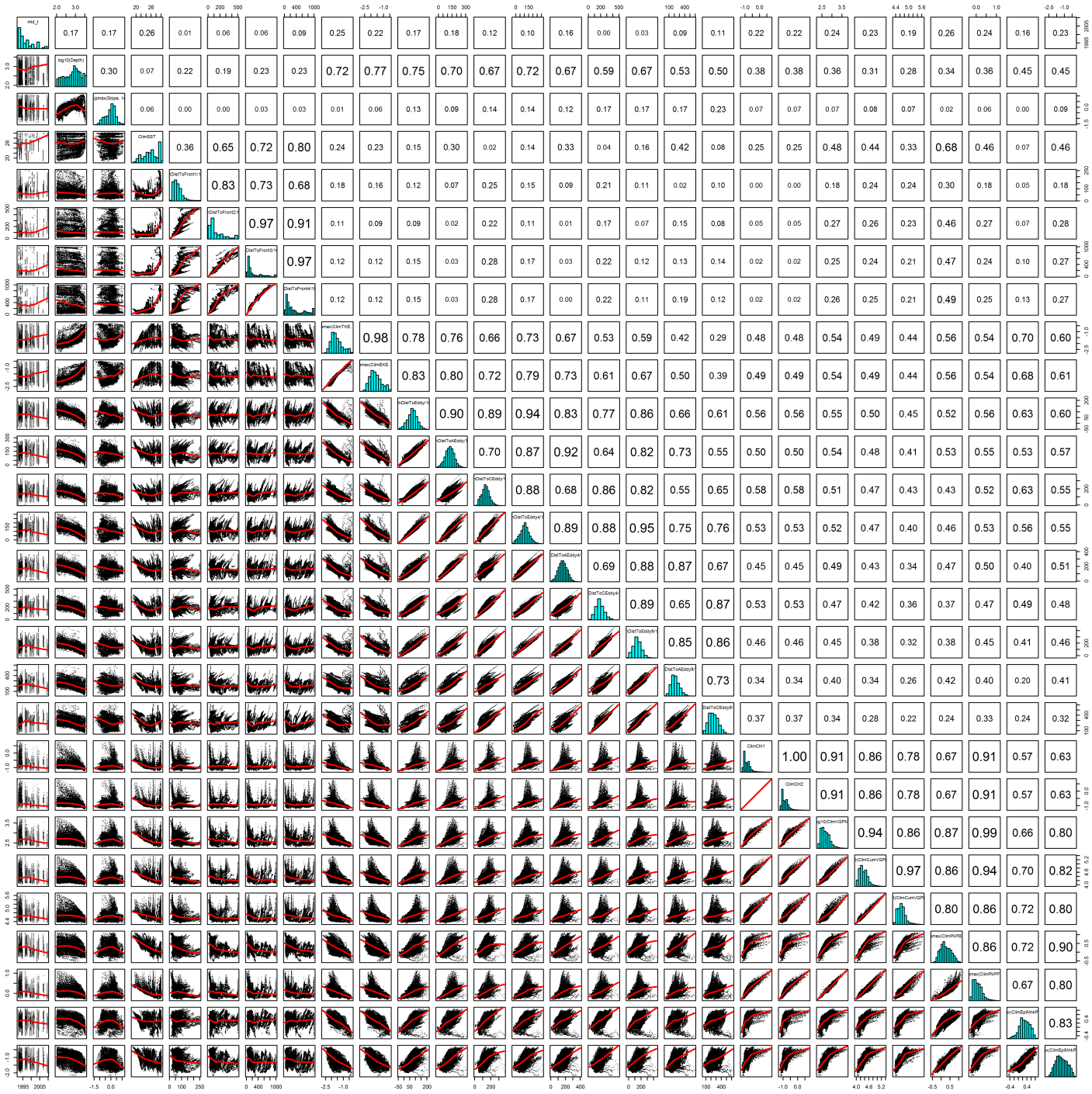


Figure 37: Scatterplot matrix for the Kogia whales Climatological model, Off Shelf. This plot is used to inspect the distribution of predictors (via histograms along the diagonal), simple correlation between predictors (via pairwise Pearson coefficients above the diagonal), and linearity of predictor correlations (via scatterplots below the diagonal). This plot is best viewed at high magnification.

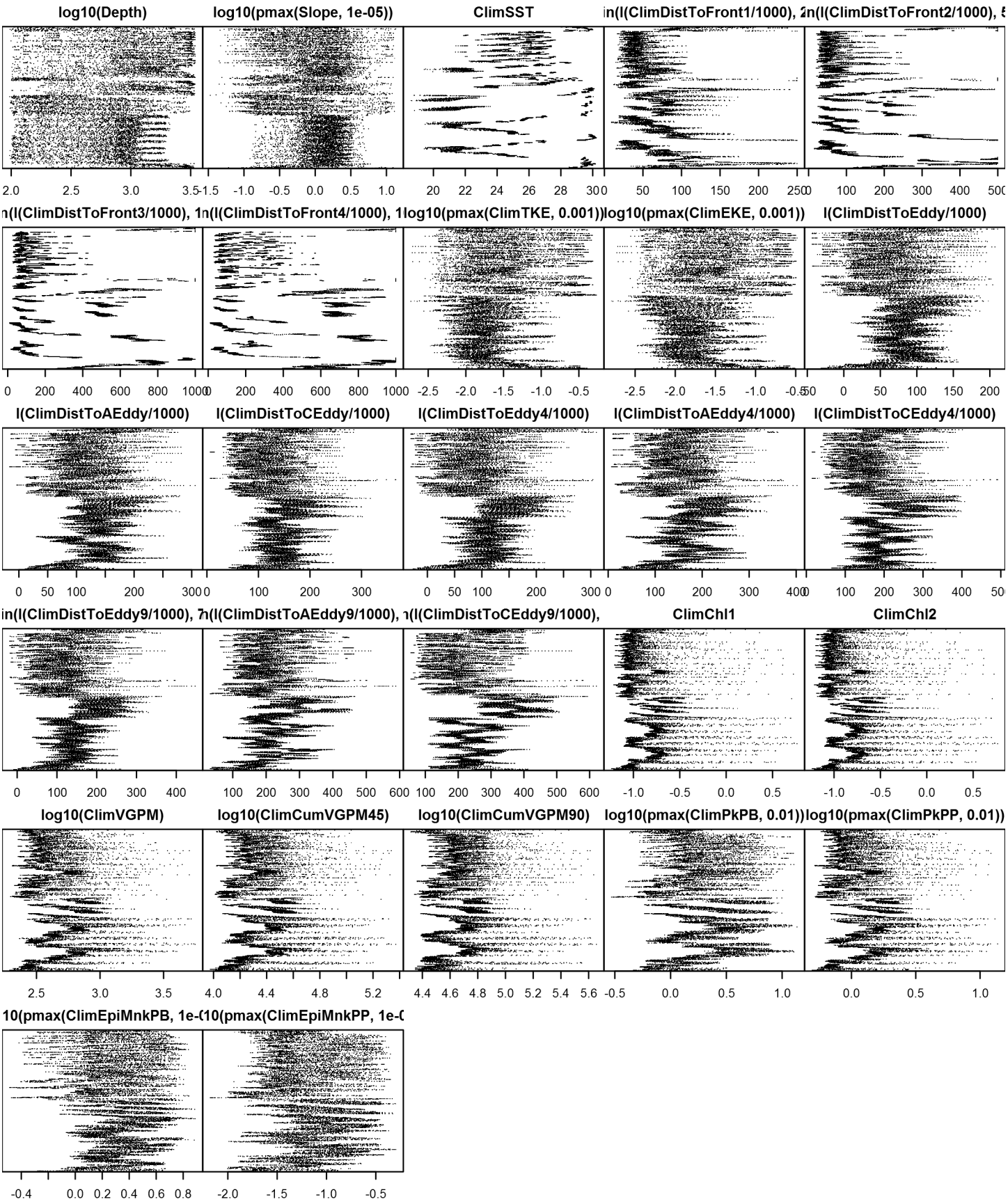


Figure 38: Dotplot for the Kogia whales Climatological model, Off Shelf. This plot is used to check for suspicious patterns and outliers in the data. Points are ordered vertically by transect ID, sequentially in time.

On Shelf

Density assumed to be 0 in this region.

Model Comparison

Spatial Model Performance

The table below summarizes the performance of the candidate spatial models that were tested. The first model contained only physiographic predictors. Subsequent models added additional suites of predictors of based on when they became available via remote sensing.

For each model, three versions were fitted; the % Dev Expl columns give the % deviance explained by each one. The "climatological" models were fitted to 8-day climatologies of the environmental predictors. Because the environmental predictors were always available, no segments were lost, allowing these models to consider the maximal amount of survey data. The "contemporaneous" models were fitted to day-of-sighting images of the environmental predictors; these were smoothed to reduce data loss due to clouds, but some segments still failed to retrieve environmental values and were lost. Finally, the "climatological same segments" models fitted climatological predictors to the segments retained by the contemporaneous model, so that the explantory power of the two types of predictors could be directly compared. For each of the three models, predictors were selected independently via shrinkage smoothers; thus the three models did not necessarily utilize the same predictors.

Predictors derived from ocean currents first became available in January 1993 after the launch of the TOPEX/Poseidon satellite; productivity predictors first became available in September 1997 after the launch of the SeaWiFS sensor. Contemporaneous and climatological same segments models considering these predictors usually suffered data loss. Date Range shows the years spanned by the retained segments. The Segments column gives the number of segments retained; % Lost gives the percentage lost.

Predictors	Climatol % Dev Expl	Contemp % Dev Expl	Climatol Same Segs % Dev Expl	Segments	% Lost	Date Range
Phys	6.9			14455		1992-2009
Phys+SST	13.6	13.5	13.6	14455	0.0	1992-2009
Phys+SST+Curr	13.6	13.5	13.6	14455	0.0	1992-2009
Phys+SST+Curr+Prod	13.6	17.0	13.0	4219	70.8	1998-2009

Table 10: Deviance explained by the candidate density models.

Abundance Estimates

The table below shows the estimated mean abundance (number of animals) within the study area, for the models that explained the most deviance for each model type. Mean abundance was calculated by first predicting density maps for a series of time steps, then computing the abundance for each map, and then averaging the abundances. For the climatological models, we used 8-day climatologies, resulting in 46 abundance maps. For the contemporaneous models, we used daily images, resulting in 365 predicted abundance maps per year that the prediction spanned. The Dates column gives the dates to which the estimates apply. For our models, these are the years for which both survey data and remote sensing data were available.

The Assumed g(0)=1 column specifies whether the abundance estimate assumed that detection was certain along the survey trackline. Studies that assumed this did not correct for availability or perception bias, and therefore underestimated abundance. The In our models column specifies whether the survey data from the study was also used in our models. If not, the study provides a completely independent estimate of abundance.

Dates	Model or study	Estimated abundance	CV	Assumed g(0)=1	In our models

1992-2009	Climatological model*	2234	0.19	No	
1998-2009	Contemporaneous model	664	0.16	No	
1992-2009	Climatological same segments model	2234	0.19	No	
2009	Oceanic waters, Jun-Aug (Waring et al. 2013)	186	1.04	Yes	Yes
2003-2004	Oceanic waters, Jun-Aug (Mullin 2007)	453	0.35	Yes	Yes
1996-2001	Oceanic waters, Apr-Jun (Mullin and Fulling 2004)	742	0.29	Yes	Yes
1991-1994	Oceanic waters, Apr-Jun (Hansen et al. 1995)	547	0.28	Yes	Yes

Table 11: Estimated mean abundance within the study area. We selected the model marked with * as our best estimate of the abundance and distribution of this taxon. For comparison, independent abundance estimates from NOAA technical reports and/or the scientific literature are shown. Please see the Discussion section below for our evaluation of our models compared to the other estimates. Note that our abundance estimates are averaged over the whole year, while the other studies may have estimated abundance for specific months or seasons. Our coefficients of variation (CVs) underestimate the true uncertainty in our estimates, as they only incorporated the uncertainty of the GAM stage of our models. Other sources of uncertainty include the detection functions and g(0) estimates. It was not possible to incorporate these into our CVs without undertaking a computationally-prohibitive bootstrap; we hope to attempt that in a future version of our models.

Density Maps

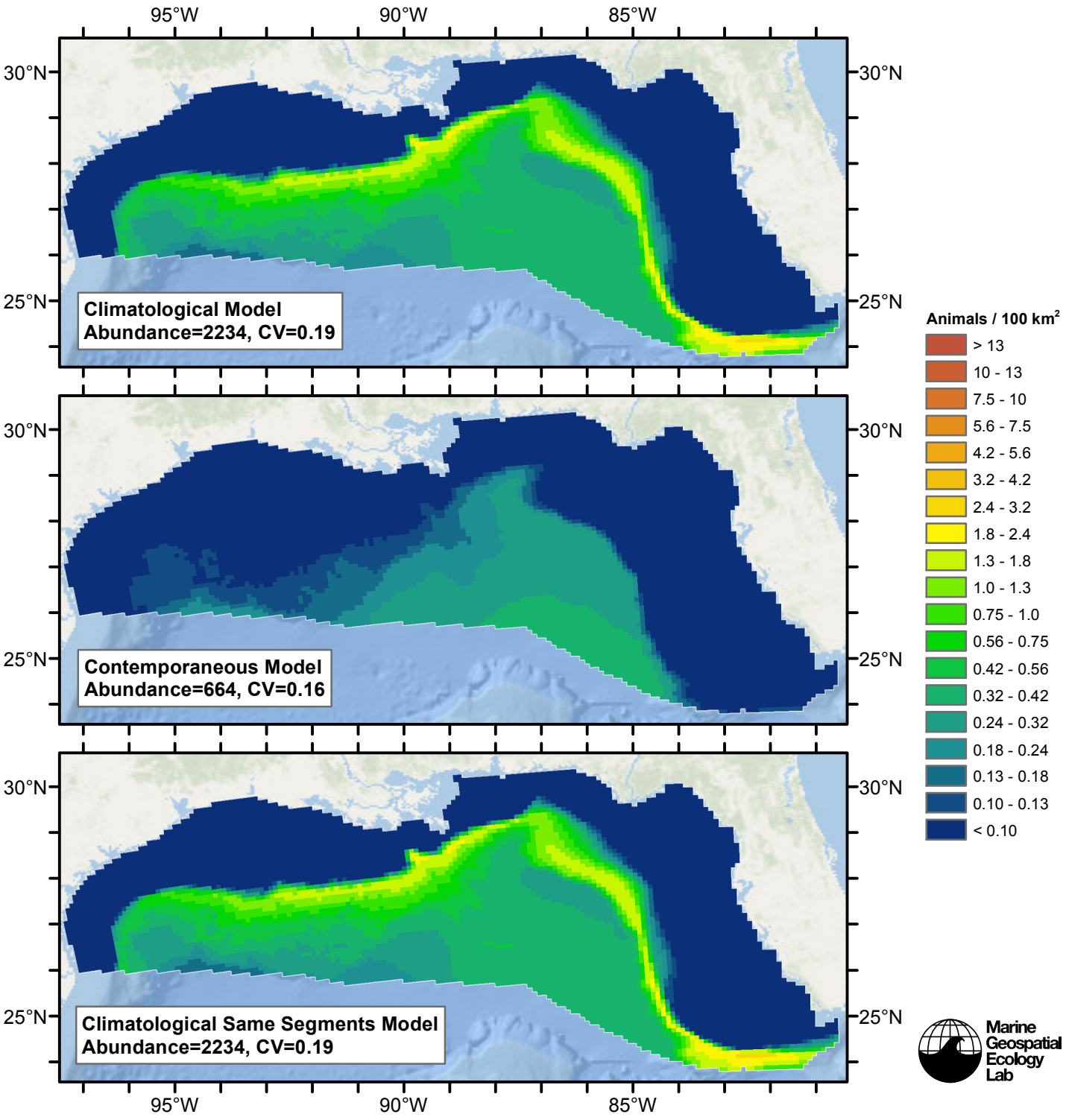


Figure 39: Kogia whales density and abundance predicted by the models that explained the most deviance. Regions inside the study area (white line) where the background map is visible are areas we did not model (see text).

Temporal Variability

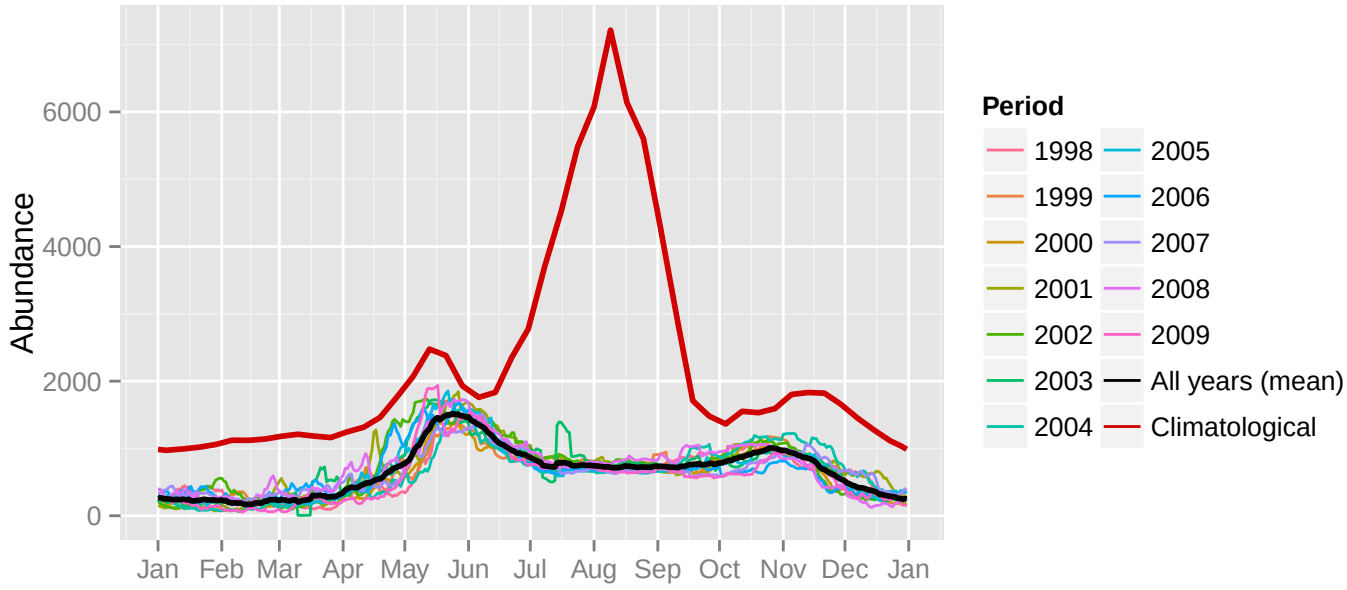


Figure 40: Comparison of Kogia whales abundance predicted at a daily time step for different time periods. Individual years were predicted using contemporaneous models. "All years (mean)" averages the individual years, giving the mean annual abundance of the contemporaneous model. "Climatological" was predicted using the climatological model. The results for the climatological same segments model are not shown.

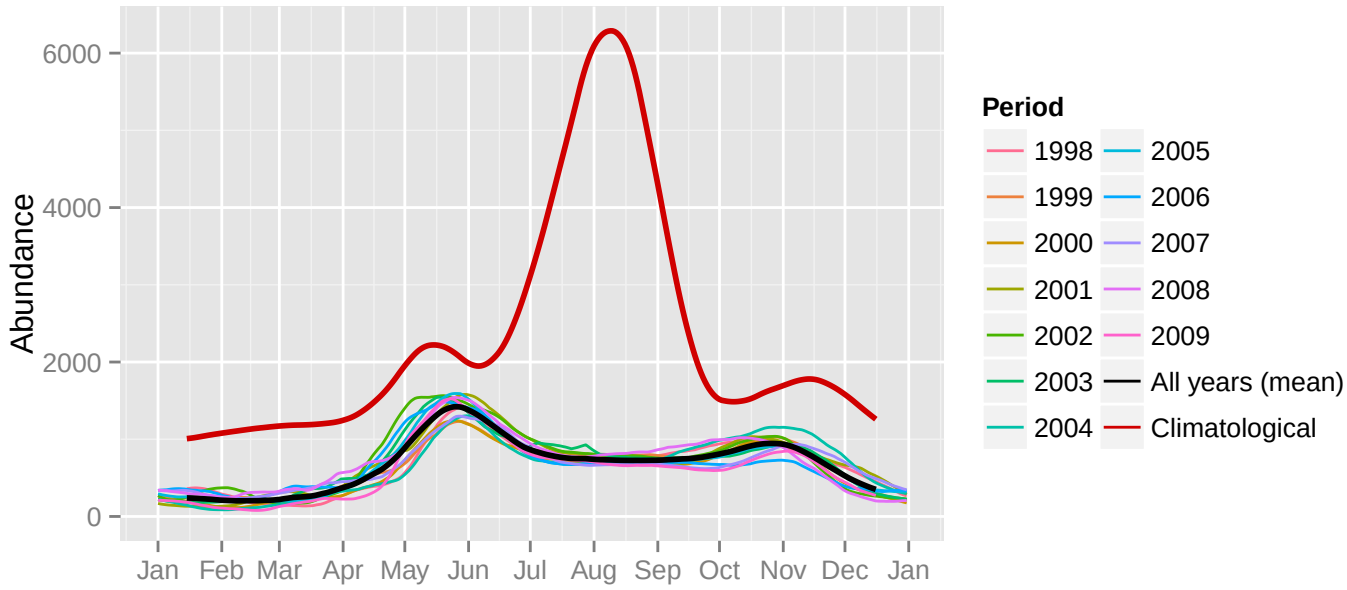
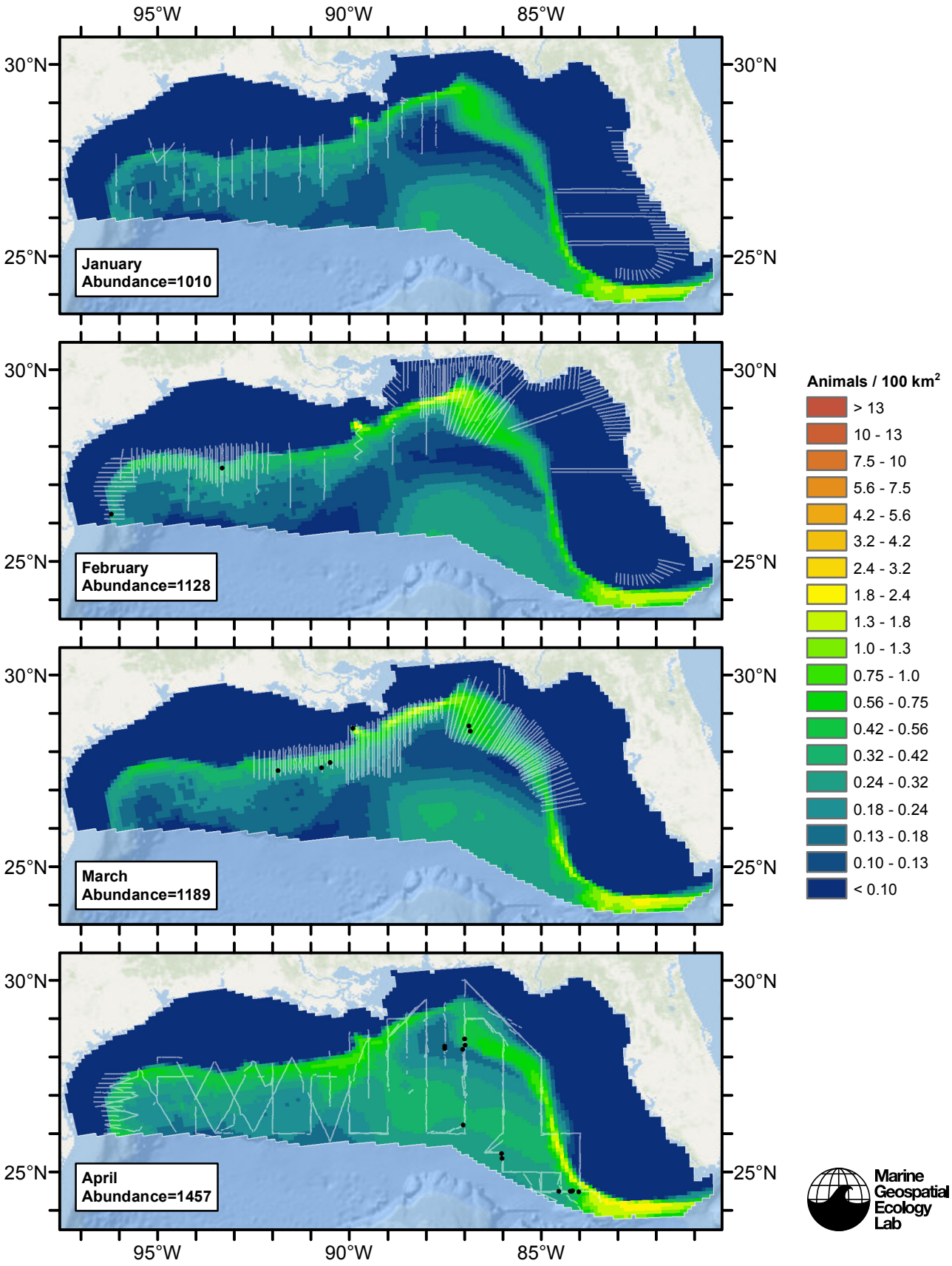
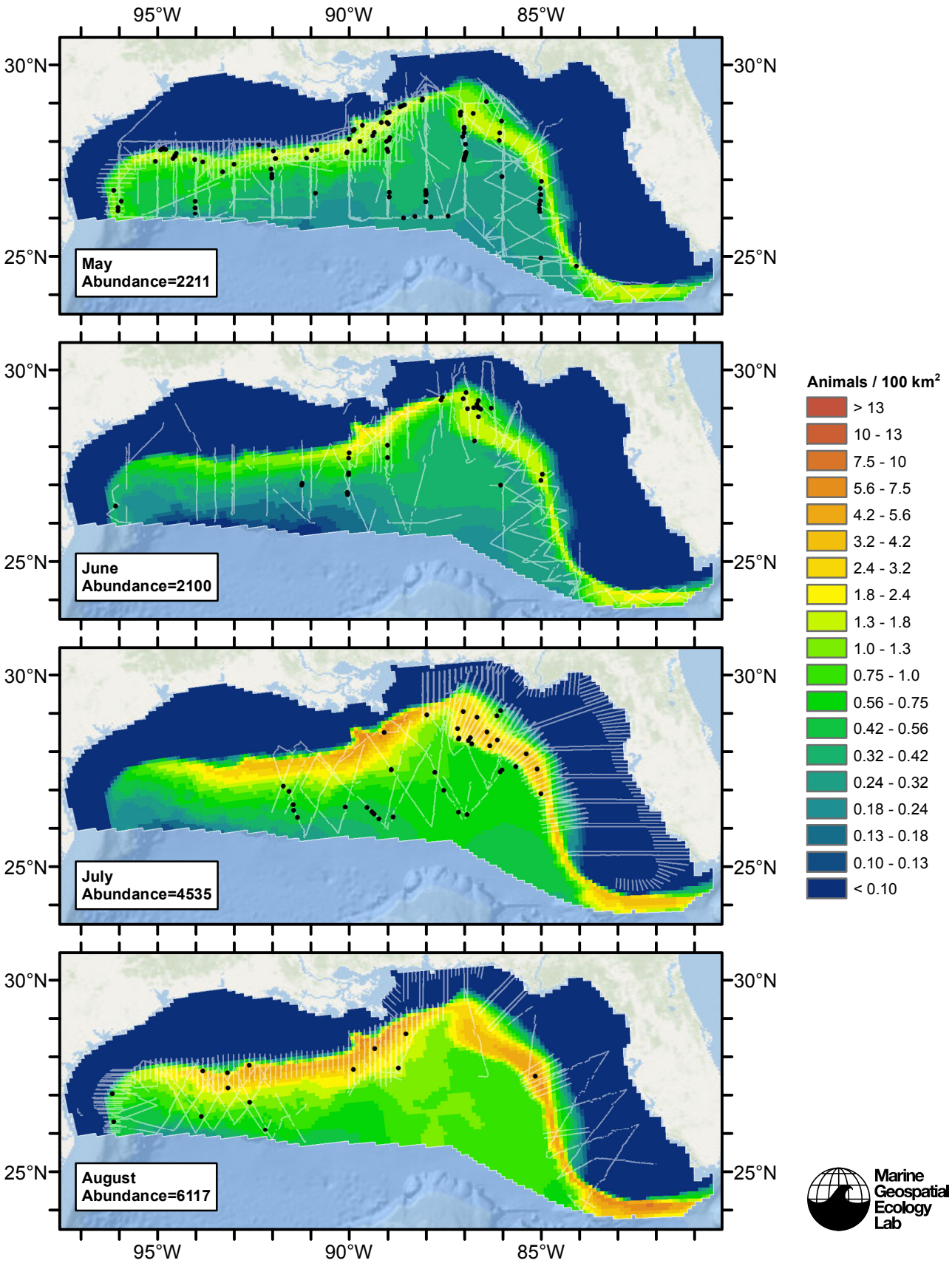
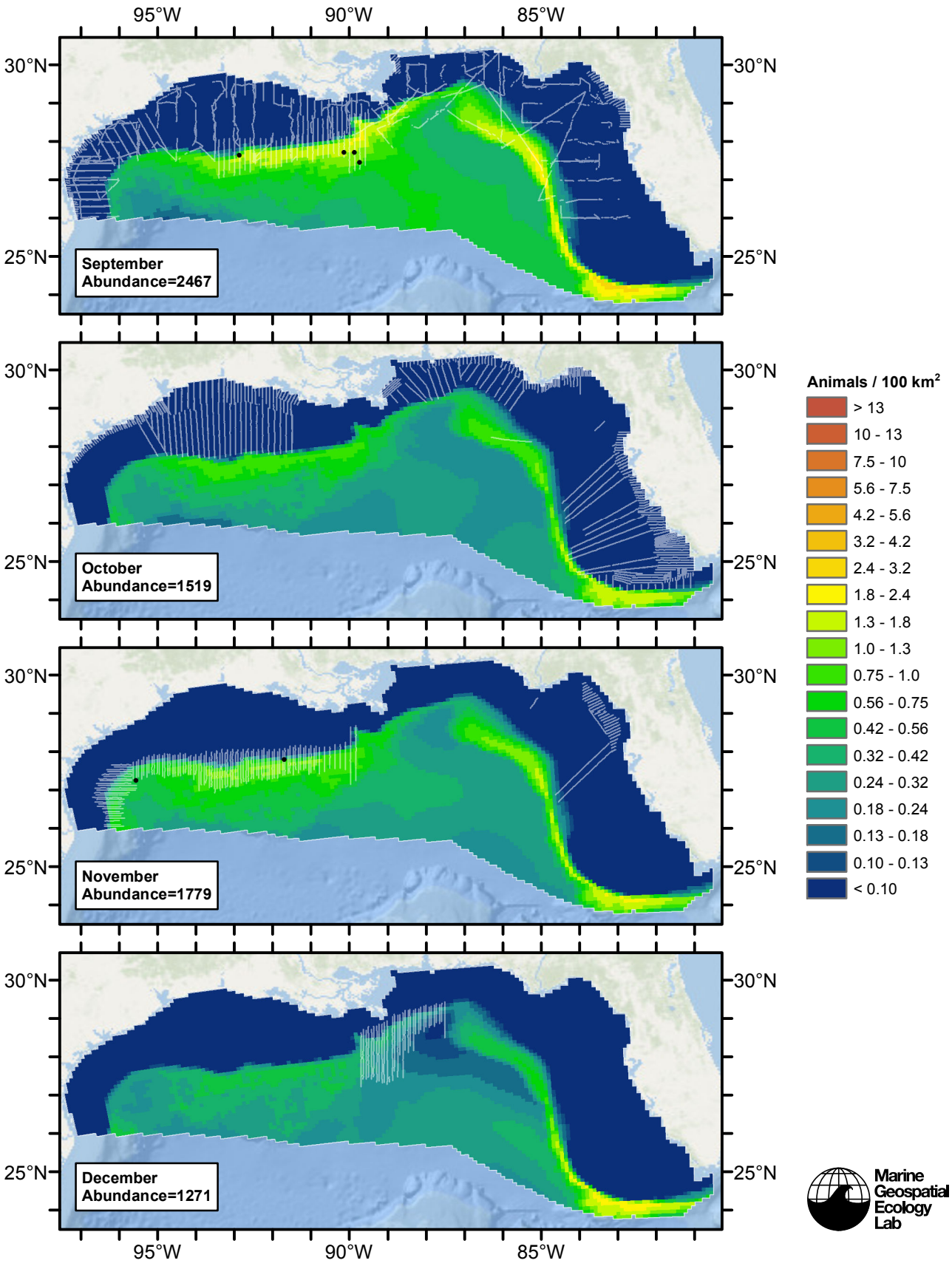


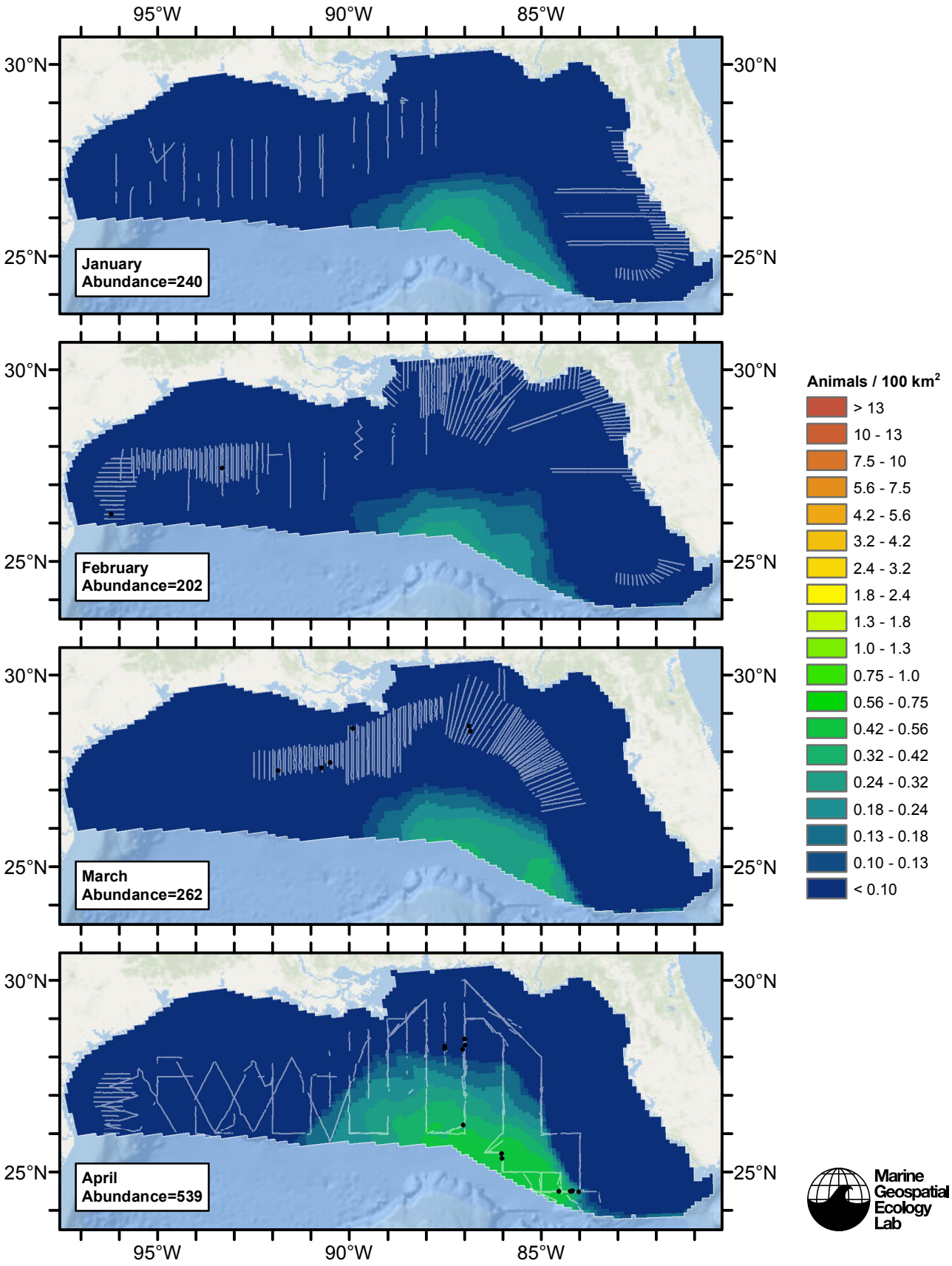
Figure 41: The same data as the preceding figure, but with a 30-day moving average applied.

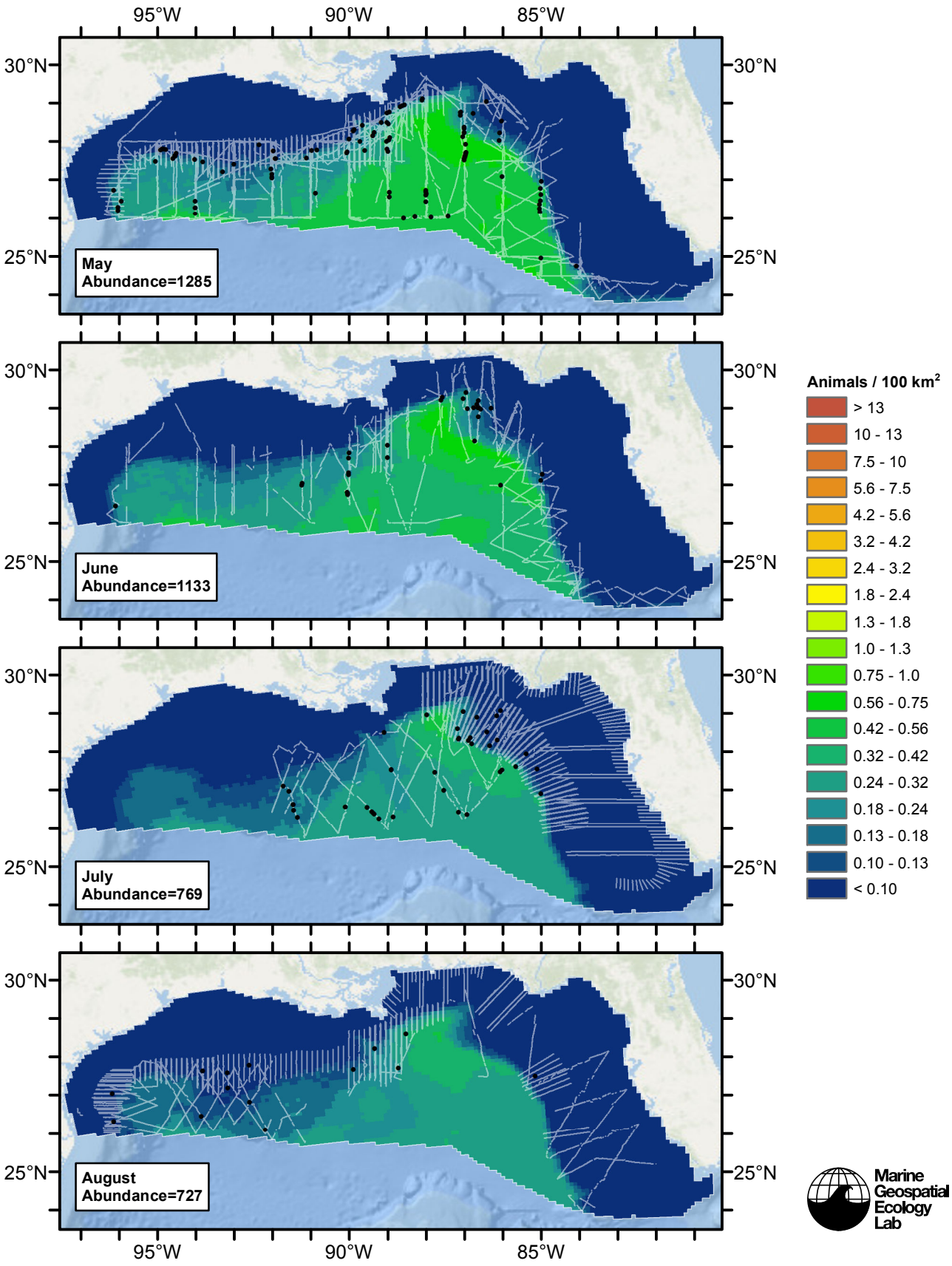

95°W
90°W
85°W
30°N
25°N
January
Abundance=1010
February
Abundance=1128
March
Abundance=1189
April
Abundance=1457
Animals / 100 km²
> 13
10 - 13
7.5 - 10
5.6 - 7.5
4.2 - 5.6
3.2 - 4.2
2.4 - 3.2
1.8 - 2.4
1.3 - 1.8
1.0 - 1.3
0.75 - 1.0
0.56 - 0.75
0.42 - 0.56
0.32 - 0.42
0.24 - 0.32
0.18 - 0.24
0.13 - 0.18
0.10 - 0.13
< 0.10
Marine
Geospatial
Ecology
Lab

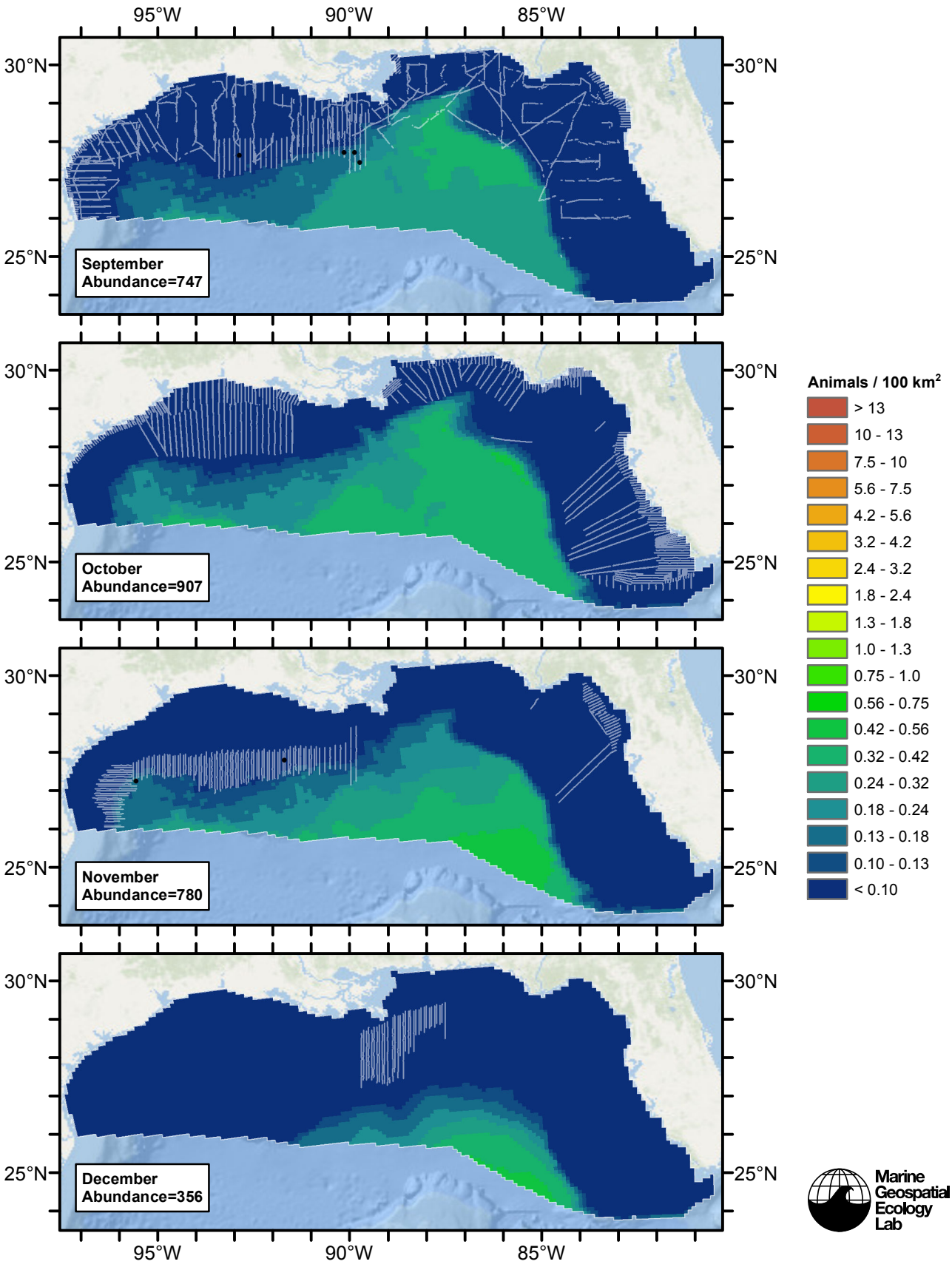

95°W
90°W
85°W
30°N
25°N
May
Abundance=2211
June
Abundance=2100
July
Abundance=4535
August
Abundance=6117
Animals / 100 km²
> 13
10 - 13
7.5 - 10
5.6 - 7.5
4.2 - 5.6
3.2 - 4.2
2.4 - 3.2
1.8 - 2.4
1.3 - 1.8
1.0 - 1.3
0.75 - 1.0
0.56 - 0.75
0.42 - 0.56
0.32 - 0.42
0.24 - 0.32
0.18 - 0.24
0.13 - 0.18
0.10 - 0.13
< 0.10
Marine
Geospatial
Ecology
Lab

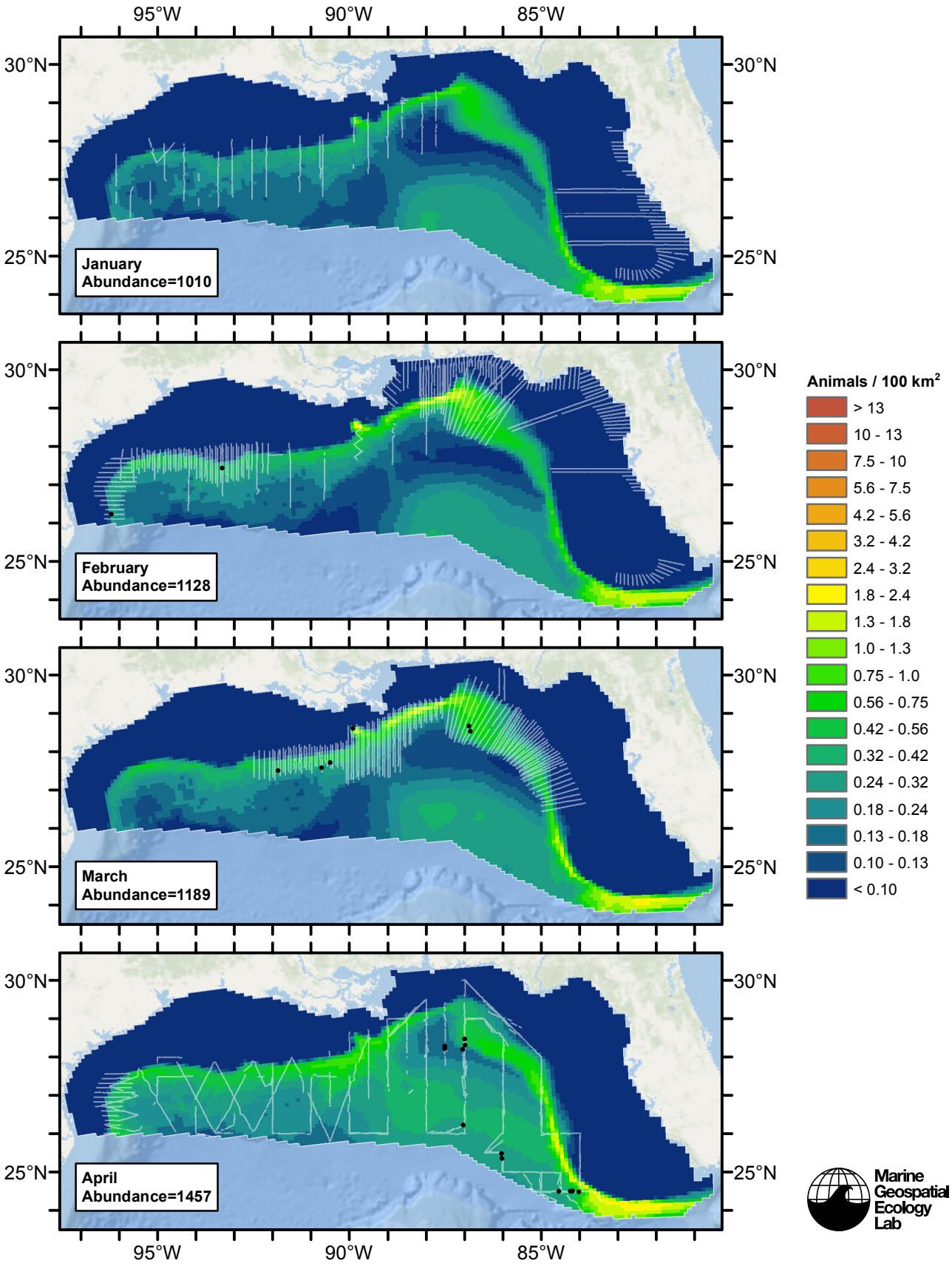

95°W
90°W
85°W
30°N
25°N
September
Abundance=2467
October
Abundance=1519
November
Abundance=1779
December
Abundance=1271
Animals / 100 km²
> 13
10 - 13
7.5 - 10
5.6 - 7.5
4.2 - 5.6
3.2 - 4.2
2.4 - 3.2
1.8 - 2.4
1.3 - 1.8
1.0 - 1.3
0.75 - 1.0
0.56 - 0.75
0.42 - 0.56
0.32 - 0.42
0.24 - 0.32
0.18 - 0.24
0.13 - 0.18
0.10 - 0.13
< 0.10
Marine
Geospatial
Ecology
Lab

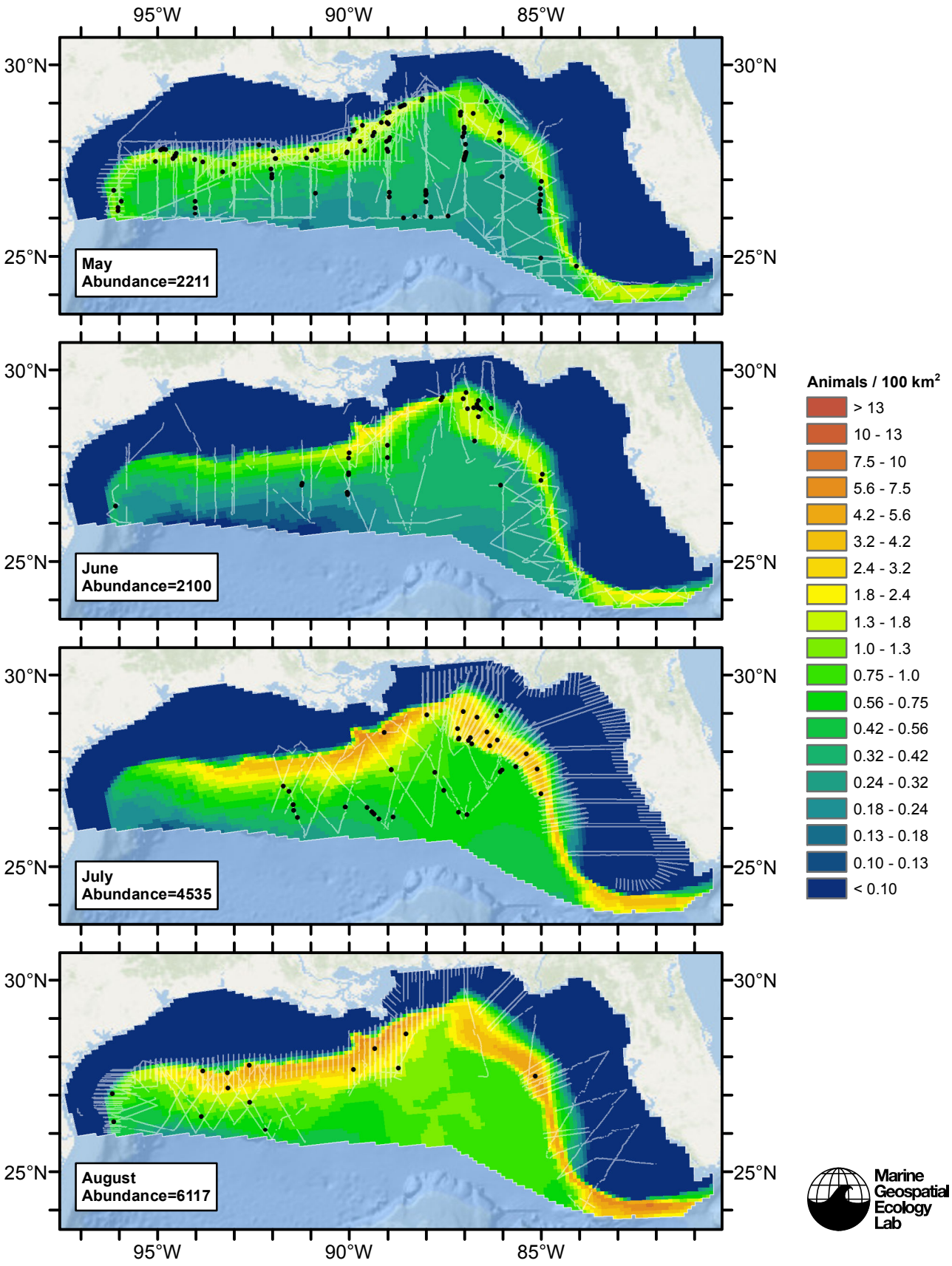
Contemporaneous Model

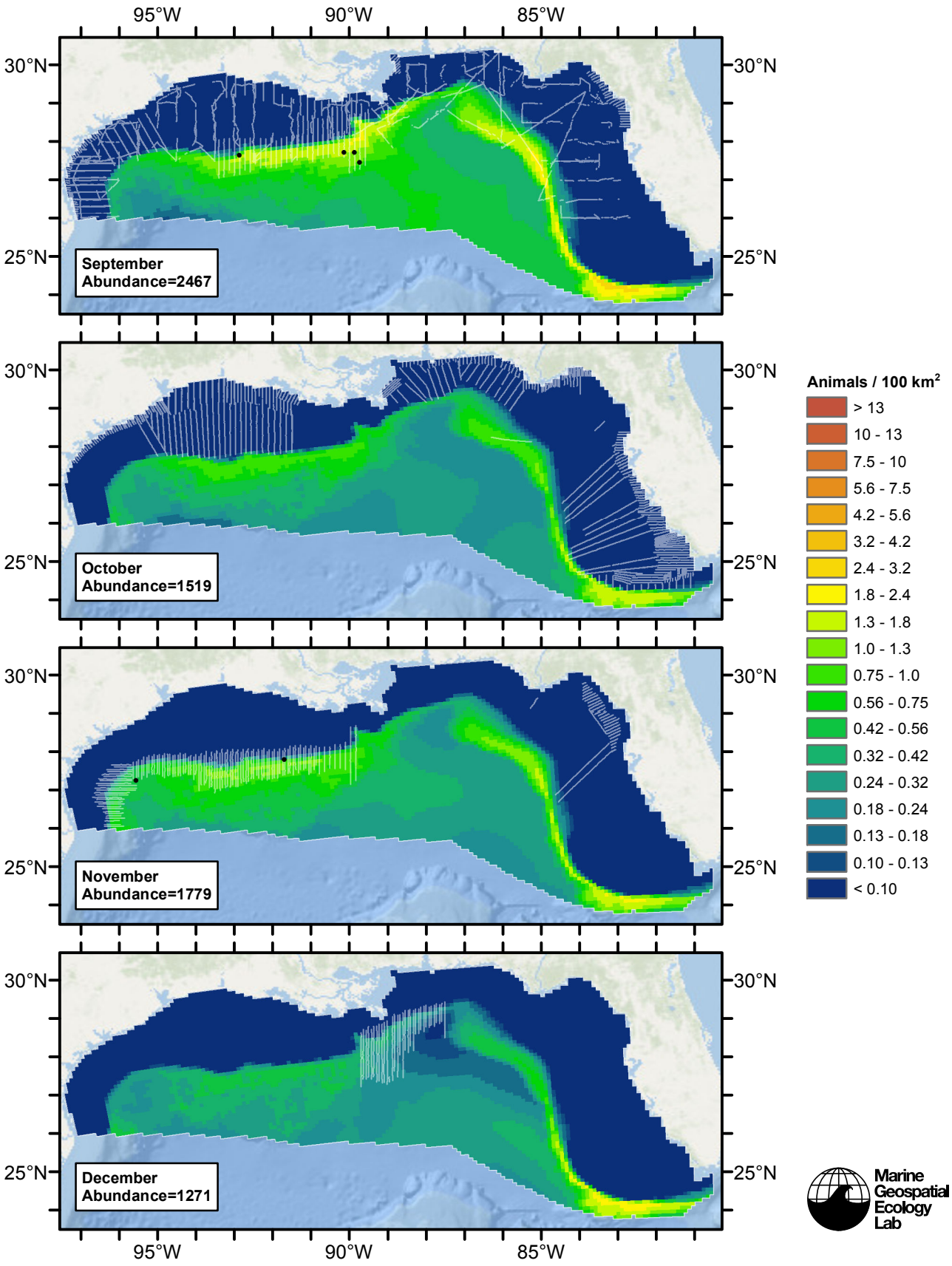



95°W
90°W
85°W
30°N
25°N
May
Abundance=1285
June
Abundance=1133
July
Abundance=769
August
Abundance=727
Animals / 100 km²
> 13
10 - 13
7.5 - 10
5.6 - 7.5
4.2 - 5.6
3.2 - 4.2
2.4 - 3.2
1.8 - 2.4
1.3 - 1.8
1.0 - 1.3
0.75 - 1.0
0.56 - 0.75
0.42 - 0.56
0.32 - 0.42
0.24 - 0.32
0.18 - 0.24
0.13 - 0.18
0.10 - 0.13
< 0.10
Marine
Geospatial
Ecology
Lab


95°W
90°W
85°W
30°N
25°N
September
Abundance=747
October
Abundance=907
November
Abundance=780
December
Abundance=356
Animals / 100 km²
> 13
10 - 13
7.5 - 10
5.6 - 7.5
4.2 - 5.6
3.2 - 4.2
2.4 - 3.2
1.8 - 2.4
1.3 - 1.8
1.0 - 1.3
0.75 - 1.0
0.56 - 0.75
0.42 - 0.56
0.32 - 0.42
0.24 - 0.32
0.18 - 0.24
0.13 - 0.18
0.10 - 0.13
< 0.10
Marine
Geospatial
Ecology
Lab


95°W
90°W
85°W
30°N
25°N
January
Abundance=1010
February
Abundance=1128
March
Abundance=1189
April
Abundance=1457
Animals / 100 km²
> 13
10 - 13
7.5 - 10
5.6 - 7.5
4.2 - 5.6
3.2 - 4.2
2.4 - 3.2
1.8 - 2.4
1.3 - 1.8
1.0 - 1.3
0.75 - 1.0
0.56 - 0.75
0.42 - 0.56
0.32 - 0.42
0.24 - 0.32
0.18 - 0.24
0.13 - 0.18
0.10 - 0.13
< 0.10
Marine
Geospatial
Ecology
Lab


95°W
90°W
85°W
30°N
25°N
May
Abundance=2211
June
Abundance=2100
July
Abundance=4535
August
Abundance=6117
Animals / 100 km²
> 13
10 - 13
7.5 - 10
5.6 - 7.5
4.2 - 5.6
3.2 - 4.2
2.4 - 3.2
1.8 - 2.4
1.3 - 1.8
1.0 - 1.3
0.75 - 1.0
0.56 - 0.75
0.42 - 0.56
0.32 - 0.42
0.24 - 0.32
0.18 - 0.24
0.13 - 0.18
0.10 - 0.13
< 0.10
Marine
Geospatial
Ecology
Lab


95°W
90°W
85°W
30°N
25°N
September
Abundance=2467
October
Abundance=1519
November
Abundance=1779
December
Abundance=1271
Animals / 100 km²
> 13
10 - 13
7.5 - 10
5.6 - 7.5
4.2 - 5.6
3.2 - 4.2
2.4 - 3.2
1.8 - 2.4
1.3 - 1.8
1.0 - 1.3
0.75 - 1.0
0.56 - 0.75
0.42 - 0.56
0.32 - 0.42
0.24 - 0.32
0.18 - 0.24
0.13 - 0.18
0.10 - 0.13
< 0.10
Marine
Geospatial
Ecology
Lab

Discussion

When models included only physiographic covariates and covariates related to SST and ocean currents, the models fitted to climatological estimates of dynamic predictors explained slightly more deviance than models fitted to contemporaneous estimates. But when covariates related to biological productivity were introduced, the contemporaneous model explained substantially more deviance than the climatological model fitted to the same segments (Table 10). These models, however, resulted in a loss of 71% of the survey segments (Table 10) and 66% of the sightings (Table 3, Fig. 1), including many reported over the continental slope during the extensive surveying that occurred there prior to 1998. We consider this an unacceptable loss of data, thus we selected the climatological model fitted to all segments as our best estimate of Kogia distribution and abundance.

In an analysis of sightings that occurred over the continental slope of the western Gulf of Mexico in 1992-1994, Davis et al. (1998) noted that deep-diving species, including Kogia, were sighted in waters with the steepest SST gradient and may have been foraging along thermal fronts associated with eddy systems. Our model is consistent with this observation; it predicted that Kogia density increases as distance to SST fronts decreases. Baumgartner et al. (2001) analyzed data collected Gulf-wide during spring of the same years and reported that the distributions of Kogia with respect to depth and zooplankton biomass were significantly different from a uniform distribution, with peak Kogia density occurring over the upper continental slope and in waters of high zooplankton biomass. The model we selected as best–the one fitted to climatological covariates–agreed with this study's finding of high Kogia density over the continental slope. Our modeling procedure also tested zooplankton biomass (the ClimPkPB covariate, from the SEAPODYM ocean model) but observed no statistically significant effect and dropped the predictor. In the models fitted to contemporaneous covariates, which were limited to surveys conducted since late 1997 (due to the unavailability of satellite ocean color measurements prior to late 1997), the candidate model that included zooplankton biomass did retain that covariate and fitted a positive correlation between it and Kogia density. This model did not explain as much deviance as the best contemporaneous model, however, which fitted a positive correlation to epipelagic micronekton productivity, a covariate not tested by Baumgartner et al. (2001). Together, these results indicate that physical and biological covariates believed correlate with the presence of prey may be important predictors of Kogia density at a contemporaneous time scale.

Because the survey effort used as input to our models was biased toward spring and summer and was spatiotemporally patchy (see maps in the Temporal Variability section above), we were not confident that any of our models could produce realistic predictions at a monthly temporal resolution. Both the climatological models and the contemporaneous model predicted a large range of abundance, with a low in January and a peak in August and May, respectively, that was 6 times larger than the January low. The effort bias problem affected all species that we modeled in the Gulf of Mexico, and we recommend that year-round average predictions be used for all Gulf of Mexico species.

Our model predicted an abundance of 2234, substantially higher than NOAA's series of estimates (Table 11) which ranged from a low of 186 in 2009 (estimated from 5 sightings (unpublished data)) to a high of 742 in 1996-2001 (estimated from 58 sightings (Mullin and Fulling 2004)). Like our model, NOAA's estimates included ambiguous "dwarf or pygmy sperm whale" sightings. The main reason that our estimate is so much higher than NOAA's may concern availability and perception bias, represented by in density models by the g(0) parameter, the probability of detecting an animal that occurs along the trackline (at distance 0). NOAA's estimates assumed that g(0)=1, indicating that animals present on the survey trackline would always be sighted, while we assumed much lower values (see g(0) Estimates section above). For shipboard surveys, we assumed g(0)=0.35. NOAA's estimates from 2009 (186), 2003-2004 (435), and 1996-2001 (742) were all based exclusively on shipboard surveys. Had NOAA used the g(0)=0.35 estimate that we used, their estimates would have been roughly three times higher and much closer to ours. NOAA's estimate from 1991-1994 was based on both shipboard and aerial surveys. For aerial surveys, we assumed g(0)=0.12, based on the very low availability of Kogia to aerial observers due to their long dive times. Applying this correction to NOAA's 1991-1994 estimate would have boosted it by an even greater degree than the later, shipboard-only estimates.

Another possibility is that the abundance of Kogia in the northern Gulf of Mexico has decreased since the 1990s. In our test of contemporaneous vs. climatological covariates, we tested models fitted to surveys conducted since late 1997, when contemporaneous ocean color data became available. For these surveys, the best performing model predicted an abundance of only 664 (it happened to use contemporaneous covariates). In models that considered all survey data since 1992, the best performing model predicted an abundance of 2334 (it necessarily used climatological covariates, as contemporaneous covariates were not available). While the since-1997 surveys tallied roughly 1/3 of the effort of the since-1992 surveys and contained a much different ratio of aerial-to-shipboard effort as well as a different spatiotemporal distribution of effort, our habitat-based models were designed to correct for these differences. Assuming they did so effectively, they suggest that Kogia abundance has greatly diminished along the continental slope since the mid 1990s, while remaining more stable in abyssal waters (Fig. 39, compare top panel to middle panel). Be that as it may, NOAA's stock assessment report for Kogia asserts that that the extant data are insufficient for determining population trends (Waring et al. 2013).

At the time of this writing, NOAA's most recent abundance estimate of 186 is what NOAA used to estimate stock-level parameters important to management, including the Minimum Population Estimate (Nmin) and the Potential Biological Removal (PBR). Because this estimate is very low relative to the abundance we estimated, it is likely that if our results are used to estimate population-level impacts from potentially harmful human activities (i.e. "takes", as defined by the Marine Mammal Protection Act), the estimated impacts will be very high relative NOAA's estimated stock size (i.e. the estimated takes will greatly exceed PBR).

There is no easy solution to this problem. One possibility is that NOAA could recalculate stock-level parameters such as Nmin and PBR using our results. But this would violate NOAA's guideline that data older than 8 years not be used to estimate stock-level parameters (Moore et al. 2011). Alternatively, impacts could be estimated using NOAA's abundance estimate of 186, computing density by dividing this number by the total area of the off-shelf portion of the U.S. Exclusive Economic Zone in the Gulf of Mexico. But this would fail to account for the non-uniform distribution of Kogia predicted by our study. Finally, in a hybrid approach, a new density surface could be obtained by apportioning NOAA's abundance estimate of 186 proportionally according to the density surface predicted by our models. To do that, divide our density surface by our total estimated abundance (2234), then multiply every cell by 186. To check that the result computed correctly, sum up all of the cells; the result should equal 186. This new density surface would reflect the distribution pattern predicted by our study but use the total abundance estimate from NOAA.

Interested parties should consult with NOAA about the best way to proceed with this problem.

References

Barlow J (1999) Trackline detection probability for long diving whales. In: Marine Mammal Survey and Assessment Methods (Garner GW, Amstrup SC, Laake JL, Manly BFJ, McDonald LL, Robertson DG, eds.). Balkema, Rotterdam, pp. 209-221.

Barlow J, Oliver CW, Jackson TD, Taylor BL (1988) Harbor Porpoise, Phocoena phocoena, Abundance Estimation for California, Oregon, and Washington: II. Aerial Surveys. Fishery Bulletin 86: 433-444.

Baumgartner MF, Mullin KD, May LN, Leming TD (2001) Cetacean habitats in the northern Gulf of Mexico. Fishery Bulletin 99: 219-239.

Bloodworth BE, Odell DK (2008) Kogia breviceps (Cetacea: Kogiidae). Mammalian Species 819: 1-12.

Carretta JV, Lowry MS, Stinchcomb CE, Lynn MS, Cosgrove RE (2000) Distribution and abundance of marine mammals at San Clemente Island and surrounding offshore waters: results from aerial and ground surveys in 1998 and 1999. Administrative Report LJ-00-02, available from Southwest Fisheries Science Center, P.O. Box 271, La Jolla, CA USA 92038. 44 p.

Davis RW, Fargion GS, May N, Leming TD, Baumgartner M, Evans WE, et al. (1998) Physical Habitat of Cetaceans Along the Continental Slope in the Northcentral and Western Gulf of Mexico. Marine Mammal Science 14: 490-507

Hansen LJ, Mullin KD, Roden CL (1995) Estimates of cetacean abundance in the northern Gulf of Mexico from vessel surveys. Southeast Fisheries Science Center, Miami Laboratory, Contribution No. MIA-94/95-25, 9 pp.

Jefferson TA, Schiro AJ (1997) Distribution of cetaceans in the offshore Gulf of Mexico. Mammal Rev. 27(1): 27-50.

Moore JE, Merrick RL, Angliss R, Barlow J, Bettridge S, Caretta J, et al. (2011) Guidelines for Assessing Marine Mammal Stocks: Report of the GAMMS III Workshop, February 15-18, 2011, La Jolla, California. US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Office of Protected Resources.

Mullin KD (2007) Abundance of cetaceans in the oceanic Gulf of Mexico based on 2003-2004 ship surveys. 26 pp.

Mullin KD, Fulling GL (2004) Abundance of cetaceans in the oceanic northern Gulf of Mexico. Mar. Mamm. Sci. 20(4): 787-807.

Palka DL (2006) Summer Abundance Estimates of Cetaceans in US North Atlantic Navy Operating Areas. US Dept Commer, Northeast Fish Sci Cent Ref Doc. 06-03: 41 p.

Waring GT, Josephson E, Maze-Foley K, Rosel PE, eds. (2013) U.S. Atlantic and Gulf of Mexico Marine Mammal Stock Assessments – 2012. NOAA Tech Memo NMFS NE 223; 419 p.

Willis PM, Baird RW (1998) Status of the dwarf sperm whale, Kogia simus, with special reference to Canada. Canadian Field-Naturalist 112: 114-125