GREEN VOLT

Technical Appendix 12.4

Comparative analysis of the design-based method and MRSea modelling using Green Volt survey data

Offshore EIA Report: Volume 2

This page is intentionally blank

Green Volt Offshore Wind Farm Ltd

Comparative analysis of the design-based method and MRSea modelling using Green Volt survey data

Dr Rob Catalano, Sean Sweeney & Dr Tim Kasoar

Client: **Green Volt Offshore Wind Farm Ltd**

Address: **Exchange Tower**

                 19 Canning Street

                 Edinburgh

                 EH3 8EH

Project reference: **P00008351**

Date of issue: **12th December 2022**

_____

Project Director: **Sean Sweeney**

Project Manager: **Dr Rob Catalano**

Statistical Modeller: **Dr Tim Kasoar**

_____

APEM Ltd
Riverview
A17 Embankment Business Park
Heaton Mersey
Stockport
SK4 3GN

Tel: 0161 442 8938
Fax: 0161 432 6083

Registered in England No. 02530851

Please cite this report as follows:

"APEM (2022). MRSea for Green Volt – Initial Analysis. APEM Scientific Report P00008351. Flotation Energy PLC, 57 pp."

Revision and Amendment Register

Version Number	Date	Section(s)	Page(s)	Summary of Changes	Approved by
1.0	23/11/22	All	All	First internal draft	RC
1.1	29/11/2022	All	All	Draft issued to client	SS
2.0	09/12/2022	All	All	Updated draft following consultation meeting with SNCBs	SS
2.1	12/12/2022	All	All	Final version submitted to SNCBs	SS

Contents

1. Introduction

Green Volt Offshore Wind Limited (the 'Applicant') is proposing to develop the Green Volt Offshore Wind Farm (OWF) (here on in referred to as 'Green Volt') as a floating offshore wind farm, approximately 75 km northeast of the Aberdeenshire coast in Scottish waters. The Green Volt Scoping Report was issued to stakeholders and the Marine Scotland Licensing Operations Team (MS-LOT) in November 2021 (Green Volt, 2021), with stakeholder consultation responses collated into the formal MS-LOT Scoping Opinion (MS-LOT, 2022) received by the Applicant in April 2022. The MS-LOT Scoping Opinion included a recommendation from both NatureScot and the Royal Society for the Protection of Birds (RSPB) to the use of MRSea modelling to determine seabird spatial distribution, abundance and densities from the 24 months of site-specific aerial digital video survey data. MRSea modelling was recommended on the basis that it may "offer greater facility in understanding the variation in distribution in response to environmental variables." The Scoping Opinion went on to say that "if this is not possible then design-based estimates must be used, but this should be checked and agreed with Marine Scotland Science (MSS) via MS-LOT and NatureScot in advance" (MS-LOT, 2022).

MRSea was introduced by the Centre for Research into Environmental and Ecological Modelling (CREEM) as a tool to implement the Complex Region Spatial Smoother (CReSS) model to marine data. The CReSS model was identified in 2013 as the preferred modelling approach for baseline characterisation and before-after impact assessment compared to Generalized Additive Models (GAMs) and Generalized Additive Mixed Models (GAMMs) (Mackenzie *et al*., 2013; Walker *et al*., 2011). MRSea is a package developed in R to be used for identifying spatially explicit changes in the spatial distribution and abundance of seabirds and marine mammals over time and across an offshore development site (Scott-Hayward *et al*., 2013a, 2013b). MRSea was not specifically designed to model data to produce more accurate estimations of the abundance of a species from individual surveys, but to detect differences in abundance and distribution between offshore renewable energy development phases by accounting for natural variation under baseline conditions. Hence, fitted models may include terms to account for seasonality and interannual variation in abundance. Fitted models may also include terms for any environmental variable which is considered to provide explanatory information regarding species distributions. Examples of environmental variables which could be used include distance to shore or distance to SPA (as proxies of distance to nearest colony); sea depth (as a proxy for foraging suitability for bottom feeders); or sea surface temperature (related to fish density and therefore foraging quality for pursuit feeders).

Where bird distribution is influenced by environmental variables that vary spatially across a given study region, inclusion of those environmental variables as covariates in the modelling approach has two advantages. Firstly, it should make the model results more accurate.

Secondly, it is useful from an ecological perspective to understand the drivers in a species distribution. Whilst it is possible to fit CReSS models purely spatially and without any environmental covariables, doing so will produce a model that does not provide any information on *why* that distribution is observed.

The Applicant has prepared this report to provide a summary of findings relating to the MRSea modelling and to account for its potential use within the Green Volt project. As agreed at a meeting on 18th October 2022 with NatureScot and MSS in advance, this modelling focuses exclusively on guillemot, *Uria aalge*, which is the most abundant species in our survey data. It should be noted that the Green Volt site is approximately 80km from shore and that throughout all analysis and discussion relating to the Green Volt site, count numbers are much lower than found at typical offshore wind farm (OWF) sites in waters closer to shore. Guillemot raw counts for each survey month, the most abundant species recorded, range from 17 to 5,706, with counts of less than 75 in eight months and counts of over 500 in only three of the 24 months of survey data.

Initial findings from this feasibility assessment were shared at a meeting on 30th November 2022 with MS-LOT, Marine Scotland, NatureScot, the RSPB and the model developer Lindsay Scott-Hayward.

1.1   Purpose of this report

The purpose of this report is to set out the results of that feasibility assessment. The Applicant's overarching question is "can outputs from the design-based method be improved upon using the MRSea package for analysis that further inform the Environmental Impact Assessment (EIA) and make meaningful differences to the conclusions reached in the application.?"

In order to inform the EIA, one of the key pieces of information is the total abundance and density of a given bird species across the development area, plus buffers as relevant. This is used to inform quantitative assessments of displacement mortality and collision risk (for species considered vulnerable to those impacts). The Applicant notes that current approaches to both displacement analysis (the matrix approach) and collision risk (the sCRM development by Donovan, 2018) take no account of the spatial distribution of birds within the development area, and therefore quantitative estimates of spatial distribution and spatial uncertainty estimates are unimportant for the assessment methods employed.

Spatial distribution of birds across the development area is used qualitatively in the assessment process. In some cases, it can form a useful narrative, for example when there is a large cluster of birds in close proximity to a fishing vessel or similar. In some cases, an Applicant may consider design refinements to mitigate impacts to ornithological features (e.g. reducing the array area) and a general impression of the distribution of birds can be useful

starting point for that process, although any changes to the design will also need to consider a wide range of other aspects too. Therefore, if spatial distribution patterns are identified that are consistent between months and within seasons they can be explored further to try and understand what may be driving particular distribution patterns.

The models and tools within the MRSea package provide a great deal of functionality and analytical opportunities. One of the main objectives in development of the MRSea package was the ability to carry out before/after impact assessments, which requires quantitative information on spatial uncertainty. The models built using MRSea can also include environmental covariates, which can provide ecologically useful information on the drivers of species' distributions. However, the Applicant notes that these types of analysis are beyond the scope of what is normally expected and required at the EIA stage.

1.2   Methods

Aerial digital video surveys were conducted by HiDef over 24 months from May 2020 to April 2022 across the Green Volt development area plus a 4 km buffer. The full development area plus 4 km buffer data were used to extract locations and counts of birds recorded. The use of these full survey data ensured the models were as accurate as possible, with results subsequently able to be clipped down to any desired area of interest (most notably the array area plus 2 km buffer used for displacement analysis of auk species). Shapefiles of observations and transect lines from each survey were supplied by HiDef. The footprint of each survey was estimated from the transect line shapefile by assuming a 125 m image half-width, as specified by HiDef, and generated using the MMQGIS Create Buffer tool within QGIS (QGIS Version 3.10.5; MMQGIS version 2020.1.16). Observation and transect shapefiles were clipped to the Green Volt development area plus 4 km buffer.

Transects were then split into 1 km segments, with each segment given a unique identification (ID). Note that as transects varied in length, the number of segments per transect varied, and also that as transects were not exact multiples of 1 km in length, all transects had a shorter segment at the boundaries. Each observation was matched joined to the corresponding transect segment, and then the number of birds was summed to give a number per segment. The coordinates (in UTM Zone 30N) of the centroid of the segment were also extracted and added as variables "x.pos" and "y.pos".

For this exploratory analysis, only guillemot was considered, and only six surveys were modelled. Four of these surveys were chosen on the basis of having relatively high numbers of guillemot raw count observations (in the months of August, September and October 2020 and April 2021). One month (May 2021) was chosen as an example with few guillemot raw count observations to understand the limitations of the modelling approach and a final month (August 2021) was chosen as an intermediate. All guillemot behaviours (flying, sitting etc.) were included and not differentiated within the model.

All subsequent analysis was carried out in R version 4.2.1 (R Core Team, 2022) and using MRSea version 1.3.1 (Scott-Hayward et al., 2021).

For each survey, firstly a generalised linear model (GLM) was produced, modelling the bird count as a function of x.pos and y.pos, with an offset of log(area), a quasi-poisson error distribution and a log link. This model is referred to as the "basic GLM".

Then, the function SALSA2D was used to fit a Complex Region Spatial Smoother (CReSS) model, again using bird count as the response, x.pos and y.pos as the spatial coordinates, with an offset of log(area), a quasi-poisson error distribution and a log link. Transect ID was used as a panel identifier. The SALSA2D function trials various models and compares them using a user-specified criterion; in all models the criterion was the 8-fold cross validation score. The starting number of knots was also chosen based on the 8-fold cross validation score. The spatially smoothed model with the lowest 8-fold cross validation score is referred to as the "best fitting 2D model". Other user-specified inputs are detailed in the **Appendix**. Diagnostic tests were then carried out on the best fitting 2D model. Full results of these diagnostic tests are presented in the **Appendix**.

Then, the best fitting 2D model was used to predict abundances in each grid cell of a user-specified prediction grid. In all cases, the prediction grid was a 1x1 km grid covering the Green Volt development area plus 4 km buffer. Then, the total abundance for the Green Volt development area plus 2 km buffer (the Area of Interest) was calculated by summing the abundance in each grid cell within this area. For grid cells only partially included in the Area of Interest, the abundance was multiplied by the proportion of the grid cell that was included.

Finally, a parametric bootstrap was carried out. A parametric bootstrap resamples model coefficients, assuming a multivariate normal distribution of parameters from the best fitting 2D model. In all cases, 1,000 bootstraps were carried out. The bootstrapped predictions were used to estimate the upper and lower 95% confidence limits, and also the standard deviation and coefficient of variation of the extracted total abundance for the Area of Interest. The confidence limits for each grid cell are presented in the **Appendix**.

The abundance estimates produced by the MRSea-based approach are compared with abundance estimates produced by the design-based abundance estimate approach, the methods of which are detailed in full in the Appendix 12.1: Offshore and Intertidal Ornithology Baseline Technical Report.

The spatial distribution maps produced by the MRSea-based approach are compared with "heatmaps" produced using kernel density estimates. Note that the heatmaps do not produce quantitative estimates of density, but can be considered indicative of relative density. The purpose of the heatmaps is not to provide detailed spatial information, but rather to qualitatively visualise the distribution of birds over the development area. The radius used to generate the heatmaps was set to 2-5 km in all cases. This was chosen on the basis of

extensive experience of dealing with digital aerial survey data of a similar scale as being a suitable radius to produce heatmaps that are effective in visualising general distribution trends.

1.3   Results

A summary of the MRSea results for the Green Volt development area plus 2 km buffer for each of the six surveys considered is given in Table 1. The design-based abundance estimates are given for comparison.

Of the six surveys considered, the model diagnostics for all reveal deviations from the underlying assumptions (see **Appendix**). However, in most cases these deviations are minor and are unlikely to significantly alter the robustness of the model conclusions. The main exception is May 2021, for which the diagnostic plots indicate significant deviations from the statistical assumptions of the model and therefore a poor model fit. This is likely to be due to the very low count and overall low density of guillemots in this survey. The runs test for April 2021 also indicated some evidence of residual autocorrelation, which is of some concern, but as the other diagnostics indicated little cause for concern, the model outputs are considered acceptable.

Table 1 Summary of MRSea results compared to design-based estimates in Green Volt development area plus 2 km buffer (for guillemots, all behaviours)

| Survey	Raw count	MRSea			Design-based			Difference in estimated abundance
		Estimated abundance	Lower 95% CI	Upper 95% CI	Estimated abundance	Lower 95% CI	Upper 95% CI	
August 2020	457	1,827	1,397	2,423	1,805	1,340	2,286	+1.21%
September 2020	5,706	23,321	19,051	28,714	22,549	17,084	26,711	+3.42%
October 2020	513	2,010	1,463	2,760	2,025	1,692	2,293	-0.74%
April 2021	1,223	4,958	3,946	6,244	4,832	3,611	5,446	+2.61%
May 2021	24	75	41	140	102	48	203	-26.47%
August 2021	113	408	296	570	478	309	1305	-14.65%

1.3.1   August 2020

Figure 1 shows the MRSea predicted values for each grid cell, and for comparison the heatmaps produced by kernel density estimation from the raw observations, for the Green Volt development area plus 4 km buffer.
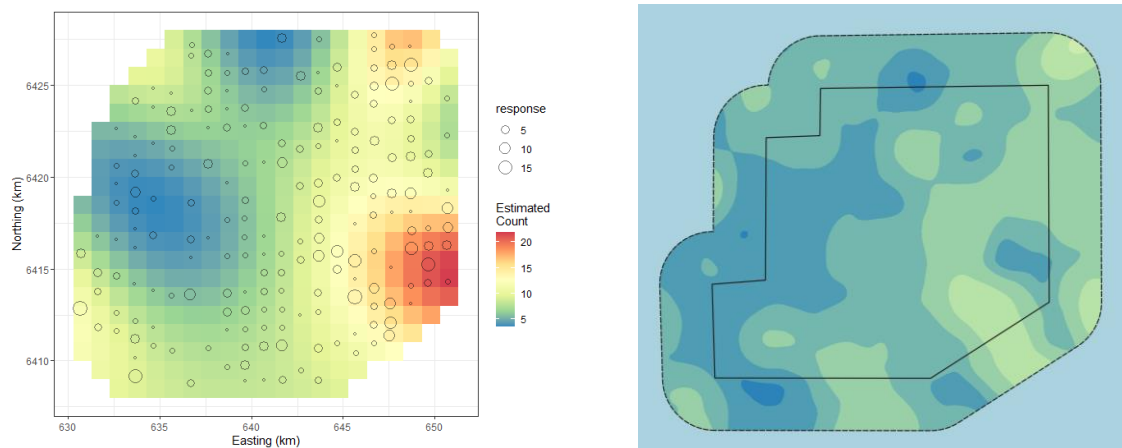
Figure 1 Left panel: MRSea predicted abundance for each grid cell, along with the total raw count for data points within the grid cell ("response"). Right panel: kernel density estimates from raw observations. Note that kernel density estimates are used to visualize patterns in distribution only. They do not produce quantitative estimates of bird counts and cannot be directly compared to MRSea plots or heatmaps for different surveys.
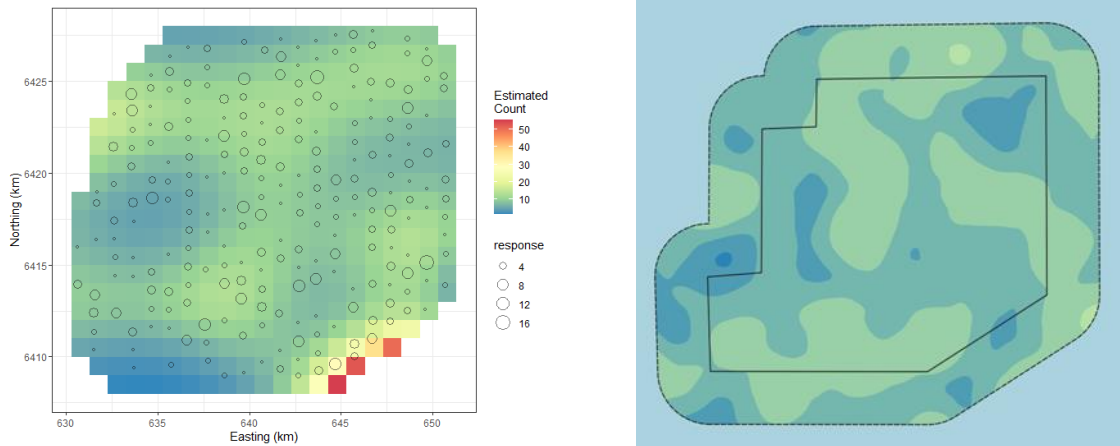
Figure 2 shows the estimated abundance and 95% confidence limits for the Green Volt development area plus 2 km buffer based on MRSea and model-based estimates.
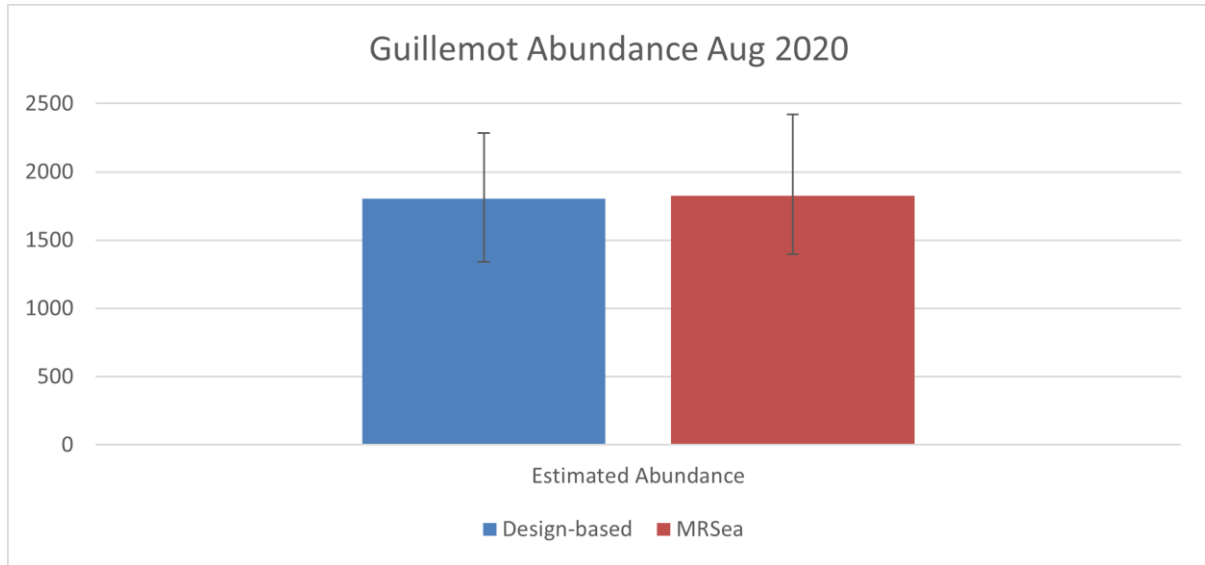
Figure 2 Comparison of estimated abundance and confidence limits for the Green Volt development area plus 2 km buffer based on MRSea and model-based estimates
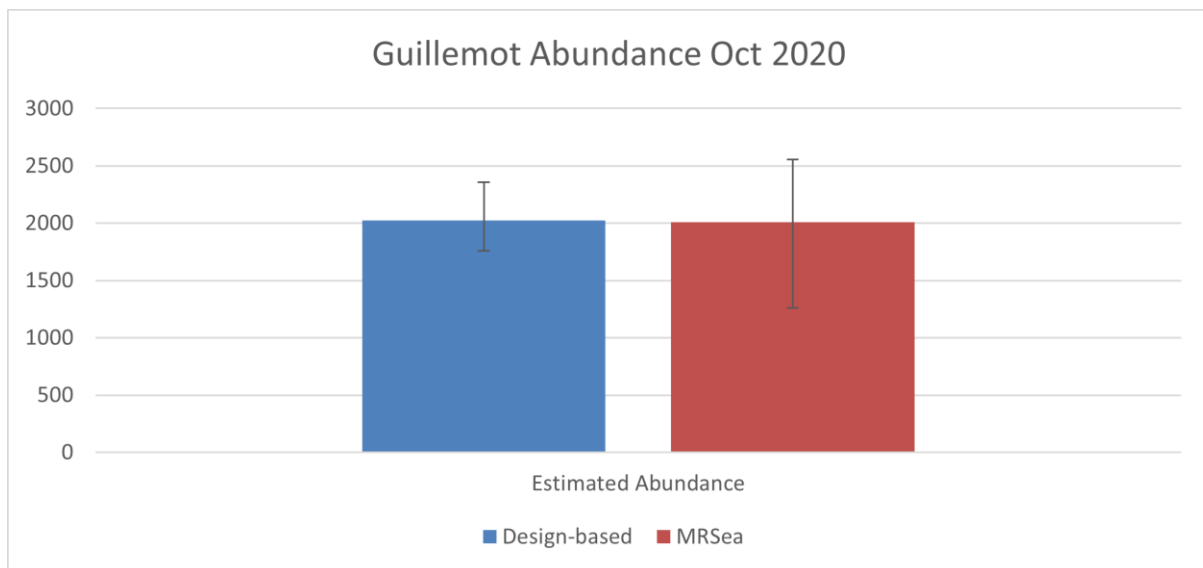
1.3.2   September 2020

Figure 3 shows the MRSea predicted values for each grid cell, and for comparison the heatmaps produced by kernel density estimation from the raw observations, for the Green Volt development area plus 4 km buffer.
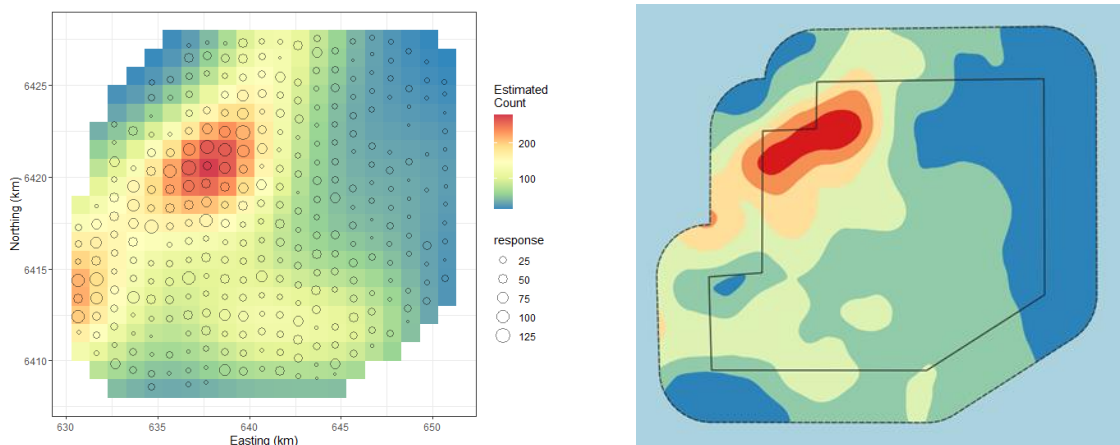
Figure 3 Left panel: MRSea predicted abundance for each grid cell, along with the total raw count for data points within the grid cell ("response"). Right panel: kernel density estimates from raw observations. Note that kernel density estimates are used to visualize patterns in distribution only. They do not produce quantitative estimates of bird counts and cannot be directly compared to MRSea plots or heatmaps for different surveys.

Figure 4 shows the estimated abundance and 95% confidence limits for the Green Volt development area plus 2 km buffer based on MRSea and model-based estimates.
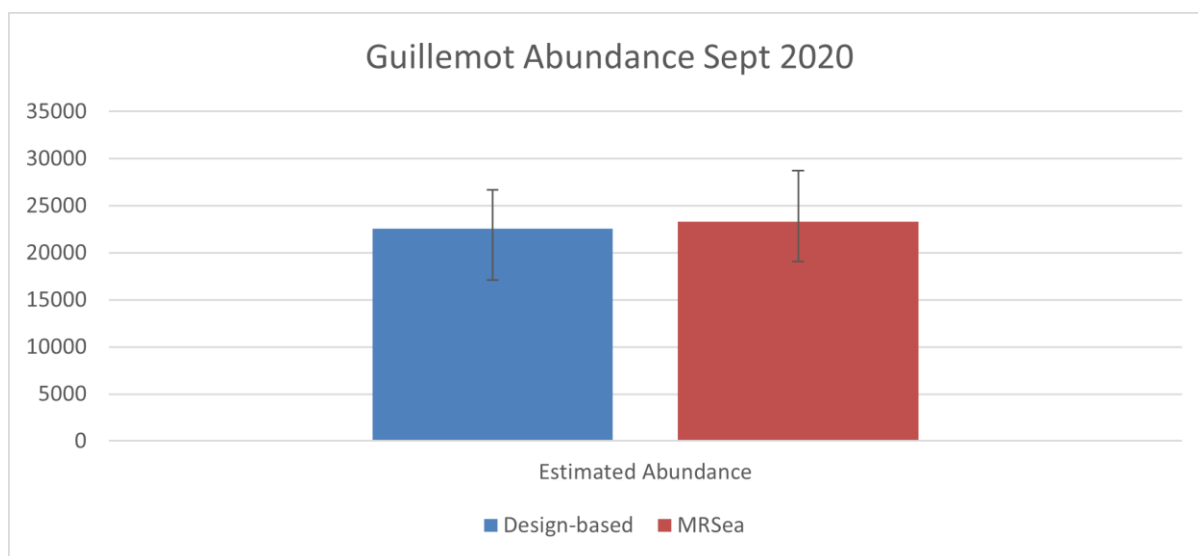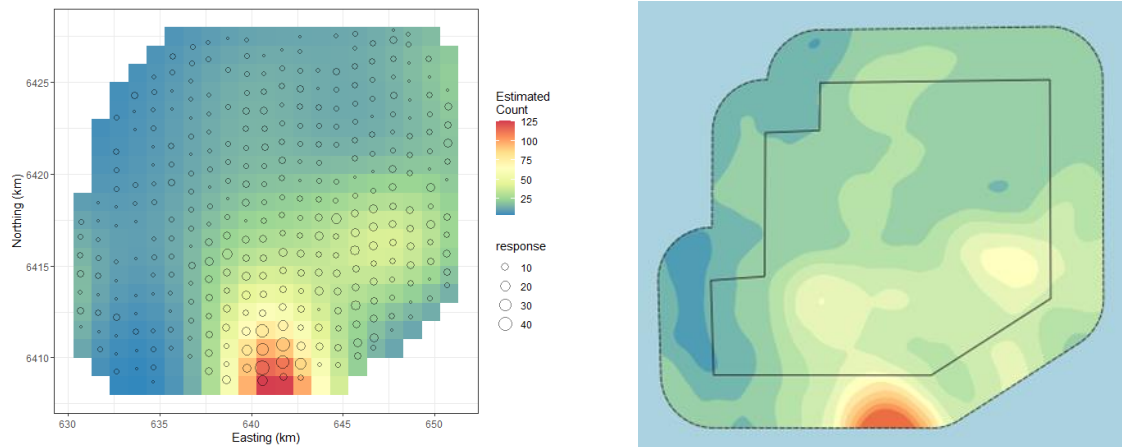

Figure 4 Comparison of estimated abundance and confidence limits for the Green Volt development area plus 2 km buffer based on MRSea and model-based estimates
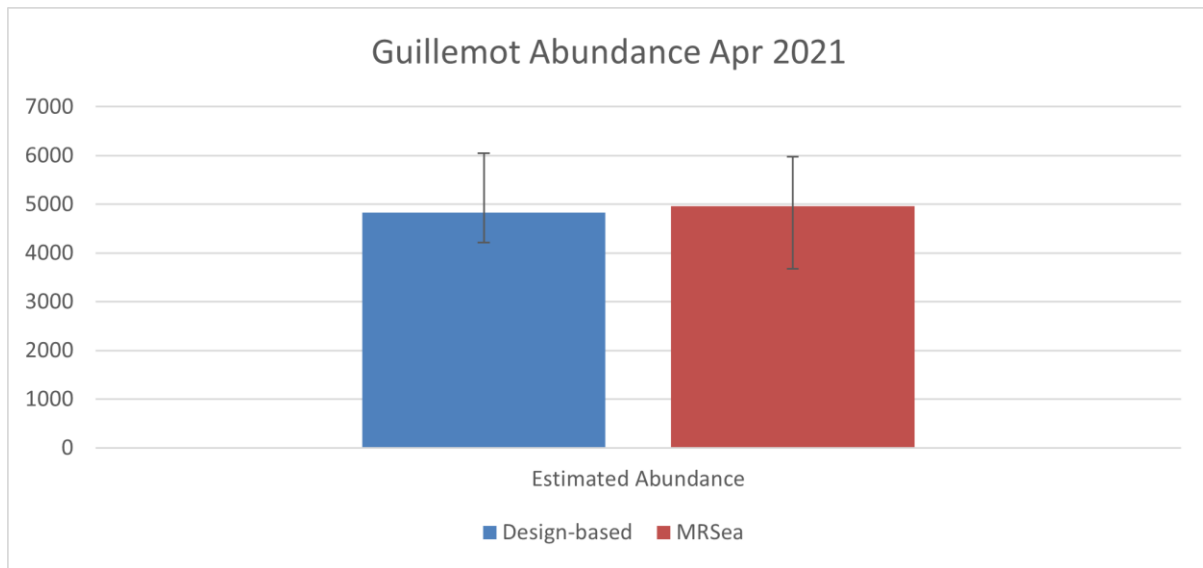
1.3.3   October 2020

Figure 5 shows the MRSea predicted values for each grid cell, and for comparison the heatmaps produced by kernel density estimation from the raw observations, for the Green Volt development area plus 4 km buffer.

Figure 5 Left panel: MRSea predicted abundance for each grid cell, along with the total raw count for data points within the grid cell ("response"). Right panel: kernel density estimates from raw observations. Note that kernel density estimates are used to visualize patterns in distribution only. They do not produce quantitative estimates of bird counts and cannot be directly compared to MRSea plots or heatmaps for different surveys.
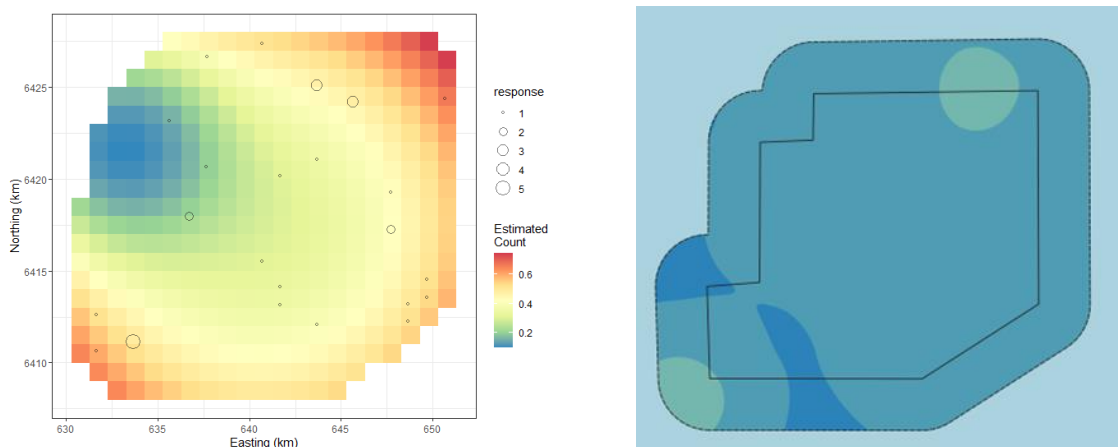
Figure 6 shows the estimated abundance and 95% confidence limits for the Green Volt development area plus 2 km buffer based on MRSea and model-based estimates.

Figure 6 Comparison of estimated abundance and confidence limits for the Green Volt development area plus 2 km buffer based on MRSea and model-based estimates
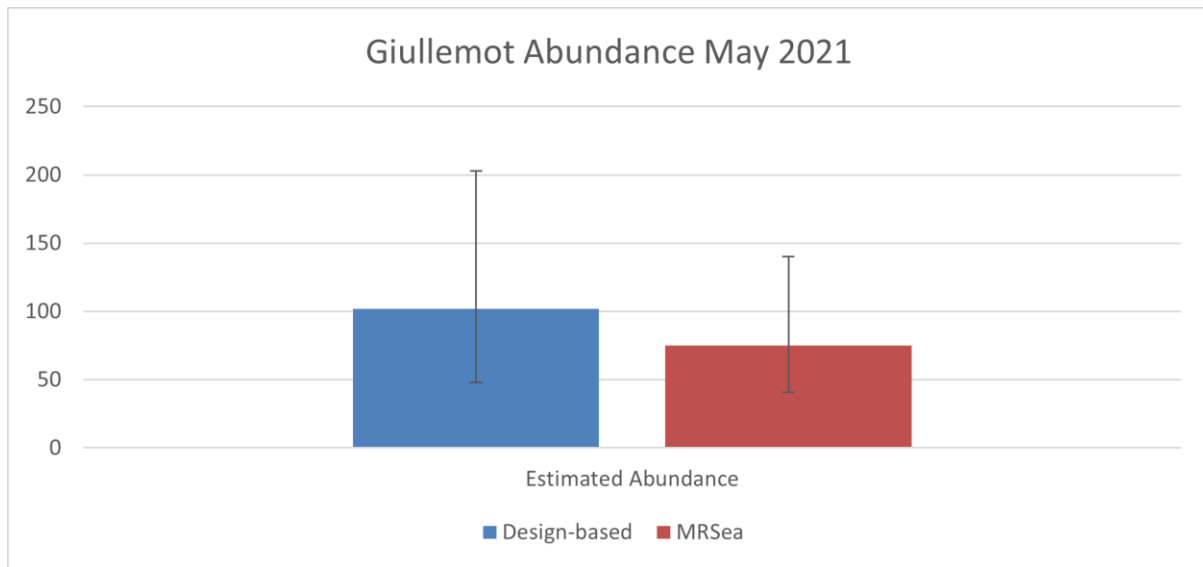
1.3.4   April 2021

Figure 7 shows the MRSea predicted values for each grid cell, and for comparison the heatmaps produced by kernel density estimation from the raw observations, for the Green Volt development area plus 4 km buffer.

Figure 7 Left panel: MRSea predicted abundance for each grid cell, along with the total raw count for data points within the grid cell ("response"). Right panel: kernel density estimates from raw observations. Note that kernel density estimates are used to visualize patterns in distribution only. They do not produce quantitative estimates of bird counts and cannot be directly compared to MRSea plots or heatmaps for different surveys.
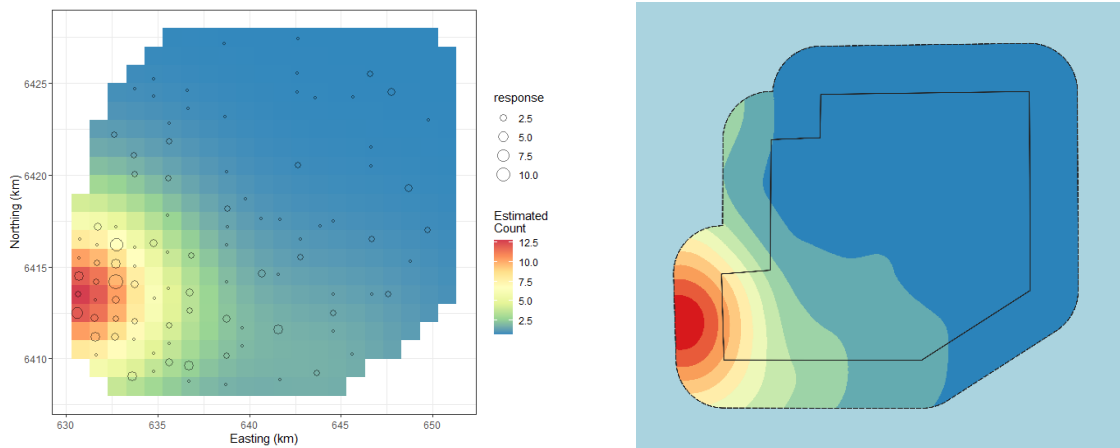
Figure 8 shows the estimated abundance and 95% confidence limits for the Green Volt development area plus 2 km buffer based on MRSea and model-based estimates.

Figure 8 Comparison of estimated abundance and confidence limits for the Green Volt development area plus 2 km buffer based on MRSea and model-based estimates

1.3.5   May 2021

Figure 9 shows the MRSea predicted values for each grid cell, and for comparison the heatmaps produced by kernel density estimation from the raw observations, for the Green Volt development area plus 4 km buffer.

Figure 9 Left panel: MRSea predicted abundance for each grid cell, along with the total raw count for data points within the grid cell ("response"). Right panel: kernel density estimates from raw observations. Note that kernel density estimates are used to visualize patterns in distribution only. They do not produce quantitative estimates of bird counts and cannot be directly compared to MRSea plots or heatmaps for different surveys.

Figure 10 shows the estimated abundance and 95% confidence limits for the Green Volt development area plus 2 km buffer based on MRSea and model-based estimates.

Figure 10 Comparison of estimated abundance and confidence limits for the Green Volt development area plus 2 km buffer based on MRSea and model-based estimates

1.3.6   August 2021

Figure 11 shows the MRSea predicted values for each grid cell, and for comparison the heatmaps produced by kernel density estimation from the raw observations, for the Green Volt development area plus 4 km buffer.

Figure 11 Left panel: MRSea predicted abundance for each grid cell, along with the total raw count for data points within the grid cell ("response"). Right panel: kernel density estimates from raw observations. Note that kernel density estimates are used to visualize patterns in distribution only. They do not produce quantitative estimates of bird counts and cannot be directly compared to MRSea plots or heatmaps for different surveys.

Figure 12shows the estimated abundance and 95% confidence limits for the Green Volt development area plus 2 km buffer based on MRSea and model-based estimates.
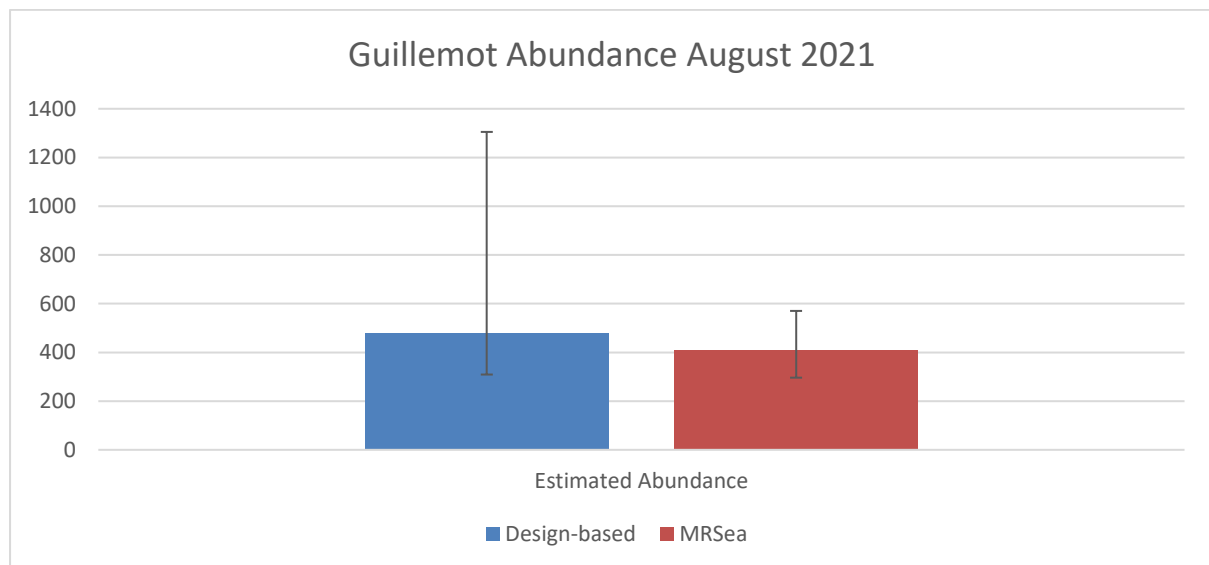
Figure 12 Comparison of estimated abundance and confidence limits for the Green Volt development area plus 2 km buffer based on MRSea and model-based estimates

1.4 Conclusions

This report addresses the application of MRSea and design-based spatial distribution modelling of guillemot data from the Green Volt aerial digital surveys. It should be noted that:

- overall raw count numbers for all species recorded within the Green Volt survey area are considered to be significantly lower when compared to data from other Scottish OWFs, particularly those within near shore waters;
- guillemots were the most abundant species recorded in the Green Volt survey area and are the subject of this analysis;
- the large distance from shore (approximately 80km) and the nature of the North Sea in that region means that there are no informative spatial gradients of environmental variables; and
- surveys were conducted with a relatively high sampling density in line with industry best practice.

Both MRSea and design-based analysis were applied to guillemot data from six separate months of surveys with various levels of counts.

In all surveys considered, the estimated abundance based on MRSea fell within the 95% confidence limits of the design-based results, and the design-based estimate falls within the 95% confidence limit of the MRSea results.

For months with relatively higher raw counts, the MRSea package is able to produce a model which appears to be relatively robust and in such cases the estimated abundance is virtually indistinguishable from the design-based estimated abundance. With higher raw counts (>400 individuals), the estimated abundances differ by less than 4%, and in some cases by less than 1%. In all cases, the 95% confidence limits tend to be of a similar magnitude when compared between the MRSea and design-based approached.

With very low raw counts (i.e. May 2021) it was evident that the modelling approach used by the MRSea package struggled to create a reliable distribution model. As discussed in consultation with Dr Scott-Hayward, the MRSea model may struggle to provide useful distribution models when count numbers are low and there are few environmental features in the dataset.

These results provide confidence that the design-based abundance estimates are robust and reliable for Green Volt. Whilst the Applicant appreciates that the MRSea modelling approach offered similar spatial and density distribution mapping to the design-based approach for some of the higher count data analysed in this report, it is noted that the vast majority of species data from the Green Volt site identifies fewer than 100 birds across the survey area. At these levels of bird count and with few features to cause population density to vary across

the survey area, the Applicant found that the MRSea modelling was unable to provide any further understanding of bird spatial and density distribution and in some cases struggled to produce meaningful results. Therefore, the Applicant does not consider that it is helpful to continue to apply MRSea modelling across further data from additional survey months for guillemot or for any other species to improve our assessment of potential impacts from Green Volt on seabirds with regards to either displacement analysis or assessment of collision risk.

We set out an example below:

> In the assessment of displacement effects, the mean peak abundance estimate is used to calculate how many birds may be subject to displacement consequent mortality.
>
> For guillemot the highest numbers are observed in the non-breeding season when peak abundance occurred in September 2020 with an estimated 22,540 birds and October 2021 with an estimated 1,807 birds as calculated using the design-based method. This gives a mean peak abundance estimate of 12,173.5 taken forward for assessment of displacement in the non-breeding season.
>
> If the abundance estimate derived using the design-based method of 22,540 is substituted for the MRSea modelled estimate of 23,321, the mean peak abundance estimate taken forward for assessment is 12,564.
>
> The Applicant's approach for auk displacement uses a displacement rate of 50% and a consequent mortality rate of 1%.
>
> The predicted additional mortalities using the design-based method is 60.8 birds compared to 62.8 birds using the MRSea modelled estimate, a difference of 2 birds.
>
> Even when using the worst-case scenario advocated by SNCBs of 60% displacement and 3% consequent mortality the difference is 7 birds (219 vs 226 birds, respectively). Considering the mean peak abundances in the breeding season are considerably lower than the non-breeding season differences between MRSea and design-based methods are likely to be less than 1 bird.

We further note that:

- MRSea modelling can require a greater degree of judgement from the modeller in selecting parameters, and greater understanding in order to scrutinise the outputs.
- MRSea modelling approach cannot cope with low raw counts and relatively even spatial distribution, which the majority of Green Volt monthly data for all seabirds would be classified as being. The modelling and analysis within this report suggests that the MRSea approach can only be used for a few select species and individual

survey months of data, whilst the design-based abundance estimates can be used in all cases.

- A hybrid approach (using both MRSea modelling and design-based estimates) would add further complexity and potential for confusion.
- In the case of Green Volt, the design-based estimates also benefit from a higher than normal survey coverage of 25%.

The Applicant, therefore, proposes that the potential benefits of minor refinements to abundance estimates from MRSea modelling are outweighed by the potential complexity of trying to apply the model to a sample data set that is not well suited.

The Applicant also considers that although the MRSea approach could be used for certain species and in certain months, doing so offers no tangible benefit to the EIA or key decision-making processes at this stage for this proposed project.

Carrying out partial MRSea application also adds complexity to both the Applicant carrying out the EIA and to the review of the EIA documents carried out by MS-LOT and other interested parties.

Running the models within the MRSea package with no environmental covariates does not offer additional understanding of species' responses to environmental variables, and therefore does not fulfil one of the aims of running MRSea as per the scoping opinion from MS-LOT.

The Applicant proposes that kernel density estimate heatmaps provide qualitative information of spatial distribution that is sufficient for the requirements of EIA.

Based on these findings and discussion with relevant parties, the Applicant continues to propose to rely on the design-based abundance estimates for the Green Volt dataset baseline characterisation and / or impact assessments. The Applicant notes that every development is different, so this finding for Green Volt should not be considered to set a precedent that MRSea analysis can be disregarded for other OWF projects.

1.5   References

Donovan, C. (2018) Stochastic Band CRM – GUI User Manual, Draft V1.0, 31/03/2017.

Green Volt (2021) Green Volt Offshore windfarm - Offshore Environmental Impact Assessment Scoping Report. Royal HaskoningDHV, Ref PC2483-RHD-ZZ-XX-RP-Z-0001, November 2021.

Mackenzie, M.L, Scott-Hayward,L.A.S., Oedekoven, C.S., Skov, H., Humphreys, E., and Rexstad E. (2013). Statistical Modelling of Seabird and Cetacean data: Guidance Document. University of St. Andrews contract for Marine Scotland; SB9 (CR/2012/05).

Marine Scotland – Licensing Operations Team. Scoping Opinion adopted by the Scottish Ministers for Green Volt Offshore Windfarm. Scottish Government, April 2022.

Scott-Hayward, L., Mackenzie, M. L., Donovan, C. R., Walker, C. G., and Ashe, E. (2013a). Complex Region Spatial Smoother (CReSS). Journal of Computational and Graphical Statistics.

Scott-Hayward, L.A.S., Oedekoven, C.S., Mackenzie, M.L., Walker, C.G., and Rexstad E. (2013b). User Guide for the MRSea Package: Statistical Modelling of bird and cetacean distributions in o shore renewables development areas. University of St. Andrews contract for Marine Scotland; SB9 (CR/2012/05).

Scott-Hayward, L.A.S., Mackenzie, M.L. and Walker, C.G. (2021). Vignette for the MRSea Package v1.3: Statistical Modelling of bird and cetacean distributions in offshore renewables development areas. Centre for Research into Ecological and Environmental Modelling, University of St Andrews.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Walker, C., Mackenzie, M., Donovan, C., and O'Sullivan, M. (2011). SALSA – a Spatially Adaptive Local Smoothing Algorithm. Journal of Statistical Computation and Simulation 81, 179-191.

Appendix

1.1   August 2020

```
initial_guillemot_model <- glm(response ~ offset(log(area)), family = "quasipoisson", data = guillemot_model_data)

knot_grid <- getKnotgrid(guillemot_model_data[, c("x.pos", "y.pos")])
distMats <-makeDists(cbind(guillemot_model_data$x.pos, guillemot_model_data$y.pos), knot_grid)

salsa2dlist <-list(fitnessMeasure = "cv.gamMRSea",
                   cv.opts = list(cv.gamMRSea.seed = 1, K=8),
                   knotgrid = knot_grid,
                   startKnots =8,
                   minKnots = 2,
                   maxKnots = 20,
                   gap = 0)

salsa2dOutput <-runSALSA2D(initial_guillemot_model,
                           salsa2dlist,
                           distMats$dataDist,
                           distMats$knotDist,
                           panels = guillemot_model_data$TransectID)
```

Figure 13 Code snippet of input parameters to SALSA2D function.

```
> summary(best_guillemot_model)

Call:
gamMRSea(formula = response ~ LRF.g(radiusIndices, dists, radii,
    aR) + offset(log(area)), family = "quasipoisson", data = guillemot_model_data,
    splineParams = splineParams)

Deviance Residuals:
    Min       1Q    Median       3Q       Max
-3.2946   -1.8133   -0.8525    0.6827    5.9445

Coefficients:
                                       Estimate Std. Error Robust S.E. t value Pr(>|t|)
(Intercept)                              2.1204     0.2035      0.2237   9.477  < 2e-16 ***
LRF.g(radiusIndices, dists, radii, aR)b1  7.3310     3.0674      3.1038   2.362 0.018704 *
LRF.g(radiusIndices, dists, radii, aR)b2  4.4113     1.3613      1.0052   4.389 1.50e-05 ***
LRF.g(radiusIndices, dists, radii, aR)b3 -8.6796     3.3343      3.3420  -2.597 0.009783 **
LRF.g(radiusIndices, dists, radii, aR)b4 -8.0968     2.6687      1.9124  -4.234 2.91e-05 ***
LRF.g(radiusIndices, dists, radii, aR)b5  8.2057     2.7267      2.1290   3.854 0.000137 ***
LRF.g(radiusIndices, dists, radii, aR)b6  5.9676     2.2795      2.2929   2.603 0.009630 **
LRF.g(radiusIndices, dists, radii, aR)b7 -8.6997     3.0299      2.7331  -3.183 0.001583 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 3.874733)

    Null deviance: 1376.3  on 370  degrees of freedom
Residual deviance: 1243.0  on 363  degrees of freedom
AIC:  NA

Max Panel Size = 20; Number of panels = 21
Number of Fisher Scoring iterations: 6
```

**Figure 14 Summary outputs of the best fitting 2D model**
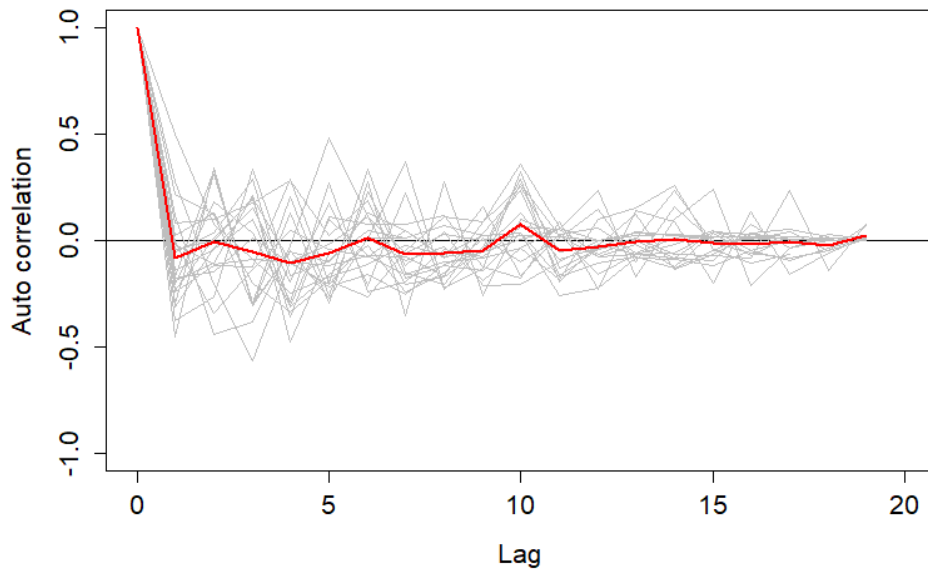
Figure 15 Autocorrelation function plot of best fitting 2D model. Auto correlation drops rapidly to zero for most runs, indicating a lack of auto correlation in the model residuals.

```
> runsTest(residuals(best_guillemot_model, type = "pearson"), emp.distribution=empdist)

        Runs Test - Two sided; Empirical Distribution

data:  residuals(best_guillemot_model, type = "pearson")
Standardized Runs Statistic = 1.4038, p-value = 0.1
```

Figure 16 Results of runs test using empirical distribution. Non-significant p-value (>0.05) indicates no evidence of autocorrelation using this test.
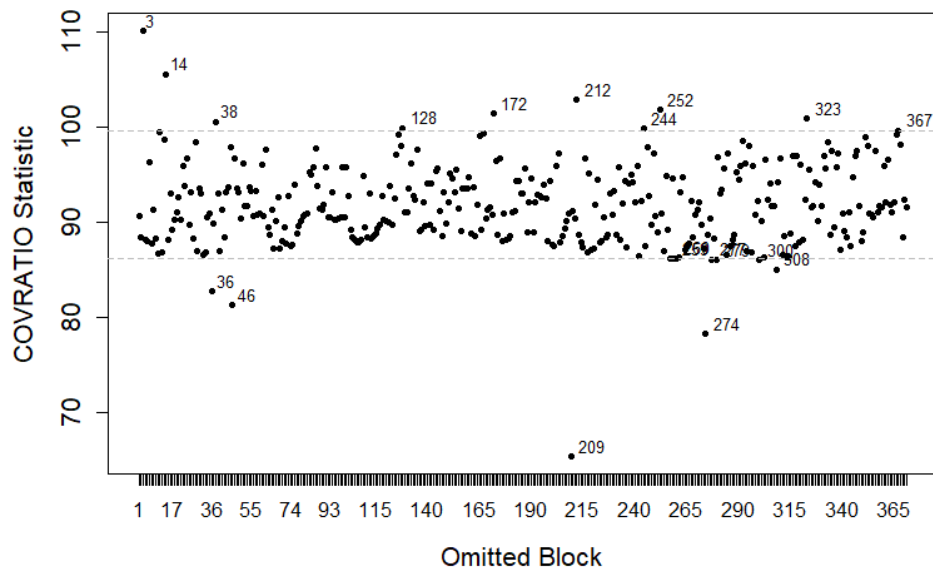
Figure 17 The "COVRATIO" (covariance ratio) measures the change in covariance as a result of removing each data point in turn, which is indicative of how influential that data point is in the precision of parameter estimates.
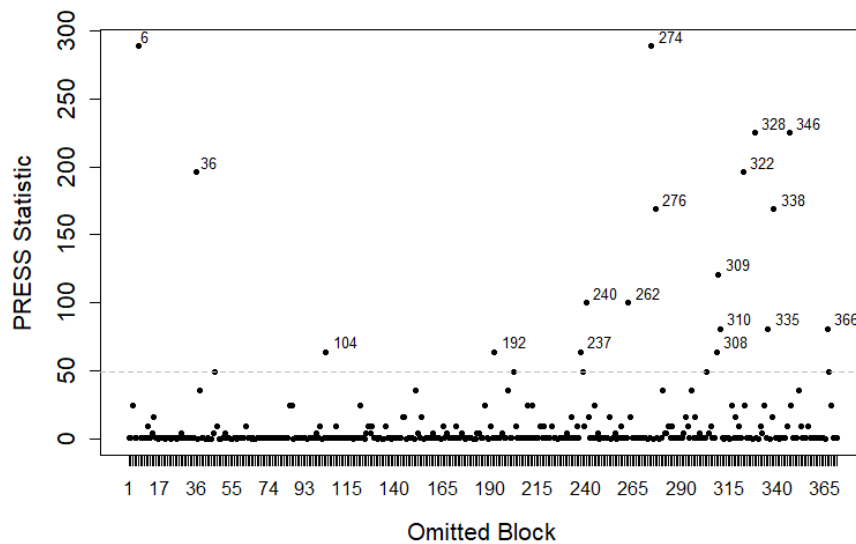
Figure 18 This plot shows the change in the PRESS (Predicted Residual Sum of Squares) as a result of omitting each data point, which is indicative of the influence of that data point on the sensitivity of model predictions.
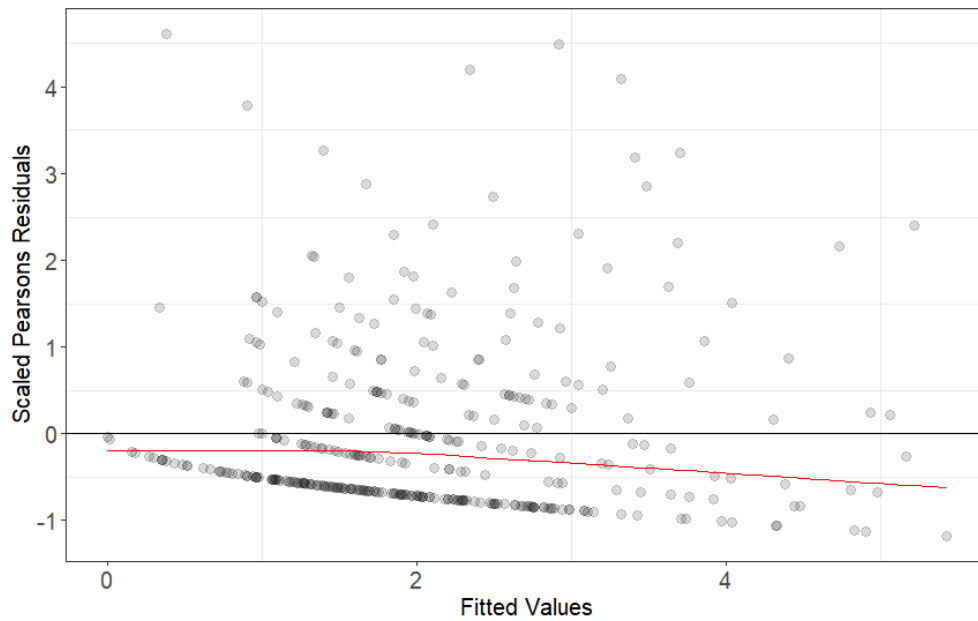
Figure 19 The plot of scaled Pearson residuals against fitted values provides information on the extent to which the model applies an appropriate error structure. If the model was perfect, there would be no discernible patterns in this plot, with residuals scattered about a mean of zero. Given that some pattern is evident, the model does not fully fit; however, overall this plot does not indicate a problem.
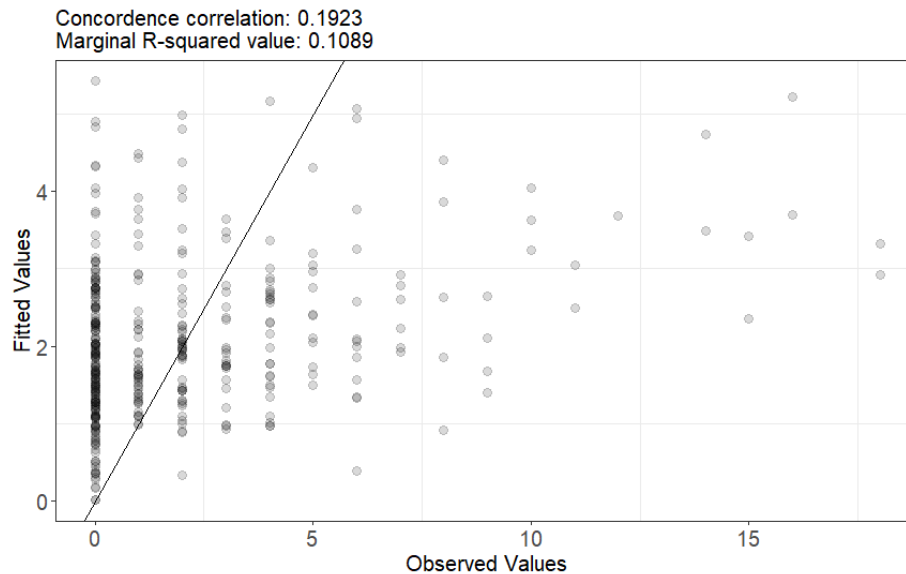
Figure 20 Plot of fitted versus observed values. Ideally points would be randomly scattered about the 1:1 line (shown). It is evident that the model is underpredicting variation in the data – many points with an observed value of zero are predicted to be above zero by the model, while the small number of points with high observed values tend to be underestimated in the model fit.
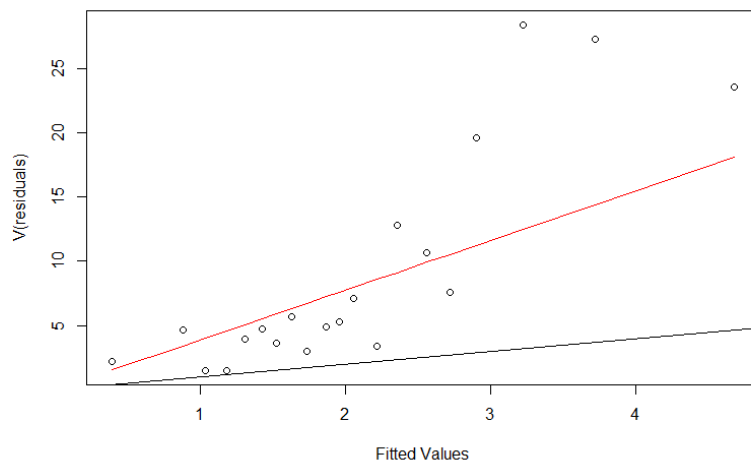
Figure 21 Plot of mean residual variance against mean fitted for each 5% quantile of fitted values. The black line is the 1:1 line while the red line has a gradient equal to the estimated dispersion parameter. A quasi-poisson model (as used) assumes variance is proportional to the mean and therefore ideally the points would be scattered about the red line. This figure therefore also indicates that the data are not truly quasi-poisson, with greater residual variance than expected at higher fitted values.
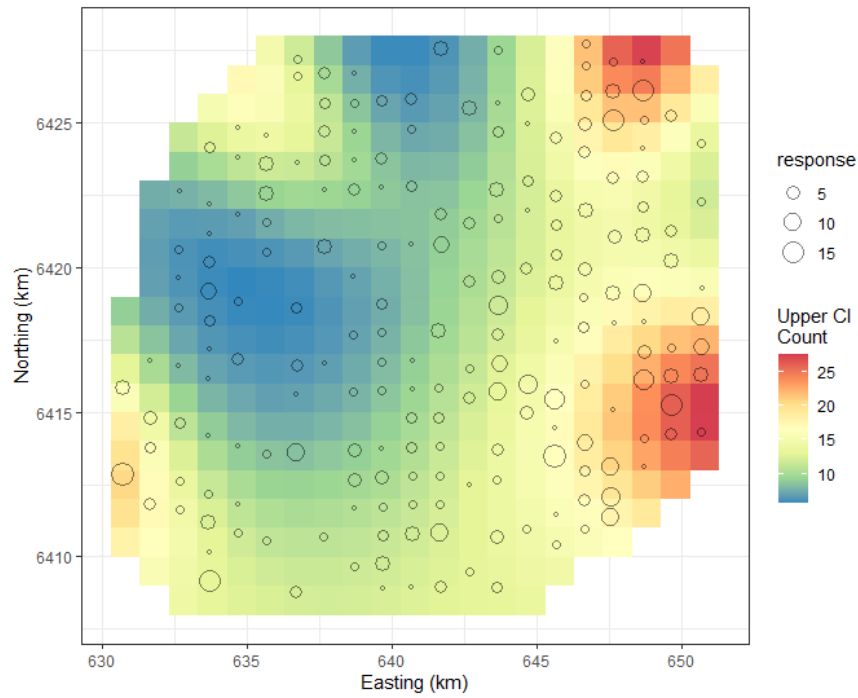
Figure 22 Upper 95% confidence limit of estimated abundance from parametric bootstrapping
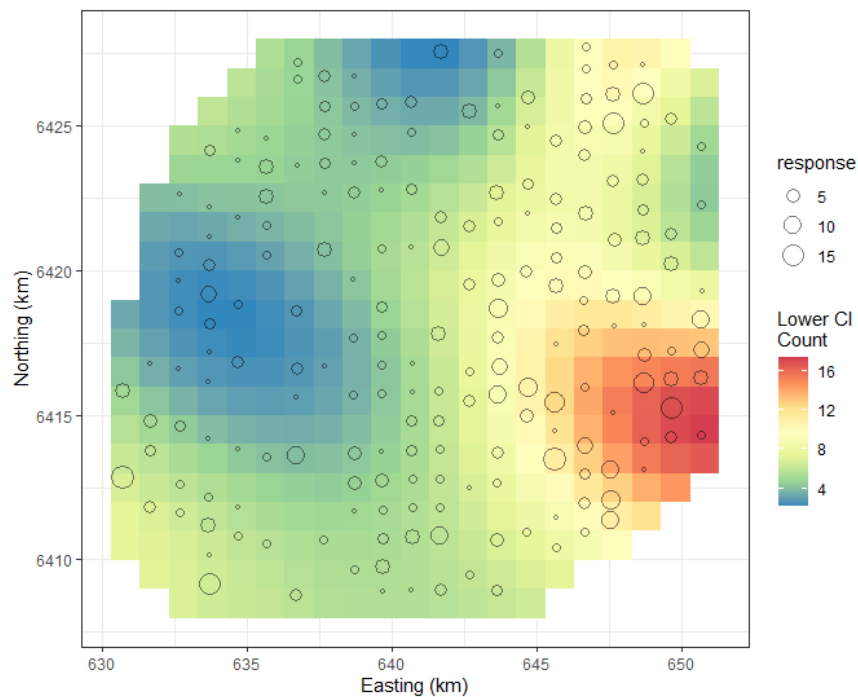
Figure 23 Lower 95% confidence limit of estimated abundance from parametric bootstrapping

1.2   September 2020

```
initial_guillemot_model <- glm(response ~ offset(log(area)), family = "quasipoisson", data = guillemot_model_data)

knot_grid <- getKnotgrid(guillemot_model_data[, c("x.pos", "y.pos")])
distMats <-makeDists(cbind(guillemot_model_data$x.pos, guillemot_model_data$y.pos), knot_grid)

salsa2dlist <-list(fitnessMeasure = "cv.gamMRSea",
                   cv.opts = list(cv.gamMRSea.seed = 1, K=8),
                   knotgrid = knot_grid,
                   startKnots = 15,
                   minKnots = 2,
                   maxKnots = 20,
                   gap = 0)

salsa2dOutput <-runSALSA2D(initial_guillemot_model,
                           salsa2dlist,
                           distMats$dataDist,
                           distMats$knotDist,
                           panels = guillemot_model_data$TransectID)
```

Figure 24 Code snippet of input parameters to SALSA2D function.

```
> summary(best_guillemot_model)

Call:
gamMRSea(formula = response ~ LRF.g(radiusIndices, dists, radii,
    aR) + offset(log(area)), family = "quasipoisson", data = guillemot_model_data,
    splineParams = splineParams)

Deviance Residuals:
    Min      1Q   Median      3Q      Max
-8.2884  -3.1229  -0.9658   1.8916  11.2865

Coefficients:
                                       Estimate Std. Error Robust S.E. t value Pr(>|t|)
(Intercept)                             1.4539     0.5696      0.6286    2.313 0.021289 *
LRF.g(radiusIndices, dists, radii, aR)b1  -35.4868     8.0403      8.3480   -4.251 2.72e-05 ***
LRF.g(radiusIndices, dists, radii, aR)b2   24.2736     5.1644      5.4590    4.447 1.16e-05 ***
LRF.g(radiusIndices, dists, radii, aR)b3   -4.8995     1.0568      0.9119   -5.373 1.40e-07 ***
LRF.g(radiusIndices, dists, radii, aR)b4   13.6056     3.4646      3.6375    3.740 0.000214 ***
LRF.g(radiusIndices, dists, radii, aR)b5    7.6070     1.3180      1.2292    6.188 1.66e-09 ***
LRF.g(radiusIndices, dists, radii, aR)b6    2.5335     0.9435      0.6795    3.728 0.000224 ***
LRF.g(radiusIndices, dists, radii, aR)b7    2.7550     0.5443      0.5851    4.709 3.56e-06 ***
LRF.g(radiusIndices, dists, radii, aR)b8   -4.5163     0.9078      0.8418   -5.365 1.45e-07 ***
LRF.g(radiusIndices, dists, radii, aR)b9    2.2361     0.4531      0.4874    4.588 6.19e-06 ***
LRF.g(radiusIndices, dists, radii, aR)b10   5.0510     0.7513      0.8156    6.193 1.61e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 13.17284)

    Null deviance: 7685.7  on 370  degrees of freedom
Residual deviance: 4648.0  on 360  degrees of freedom
AIC:  NA

Max Panel Size = 20; Number of panels = 21
Number of Fisher Scoring iterations: 5
```

Figure 25 Summary outputs of the best fitting 2D model
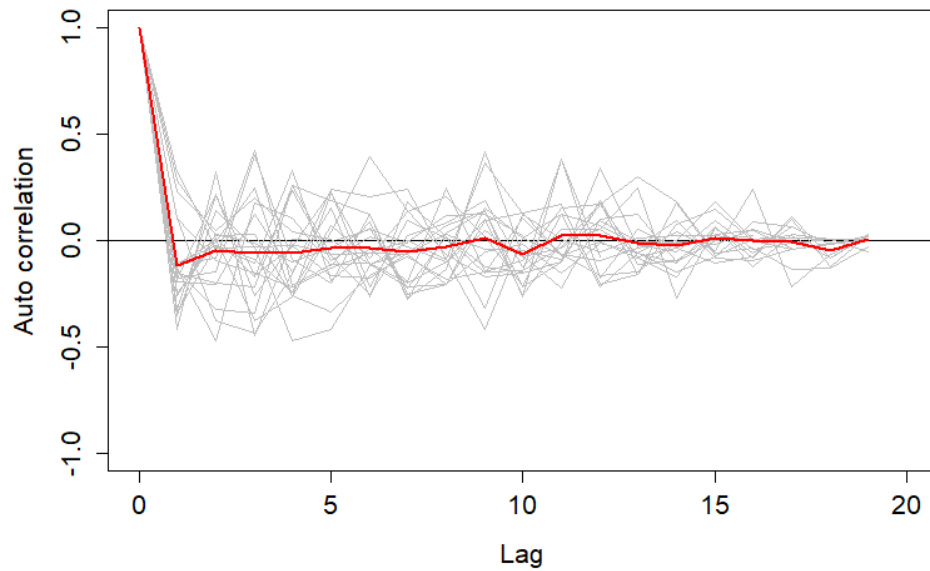
Figure 26 Autocorrelation function plot of best fitting 2D model. Auto correlation drops rapidly to zero for most runs, indicating a lack of auto correlation in the model residuals.

```
> runsTest(residuals(best_guillemot_model, type = "pearson"), emp.distribution=empdist)

        Runs Test - Two sided; Empirical Distribution

data:  residuals(best_guillemot_model, type = "pearson")
Standardized Runs Statistic = -0.2598, p-value = 0.646
```

Figure 27 Results of runs test using empirical distribution. Non-significant p-value (>0.05) indicates no evidence of autocorrelation using this test.
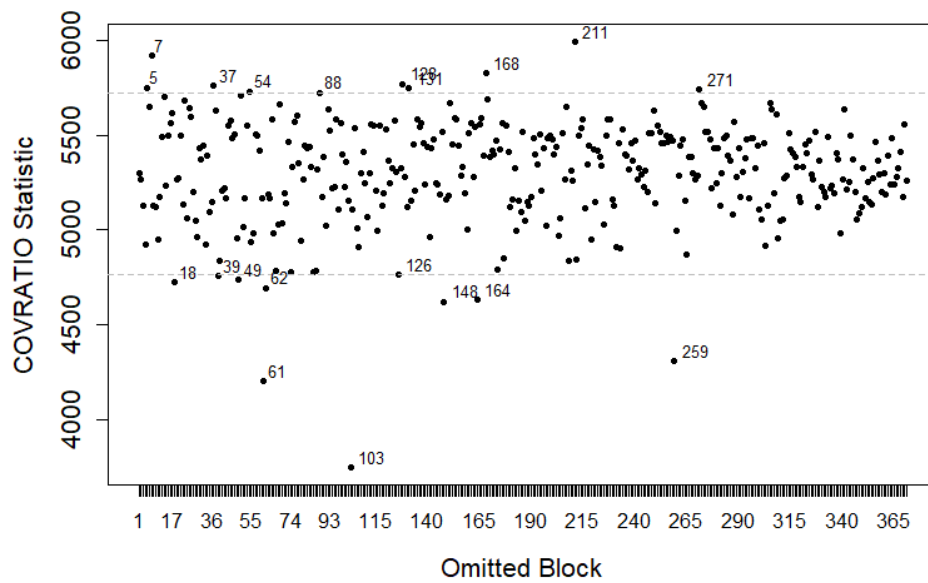
Figure 28 The "COVRATIO" (covariance ratio) measures the change in covariance as a result of removing each data point in turn, which is indicative of how influential that data point is in the precision of parameter estimates.
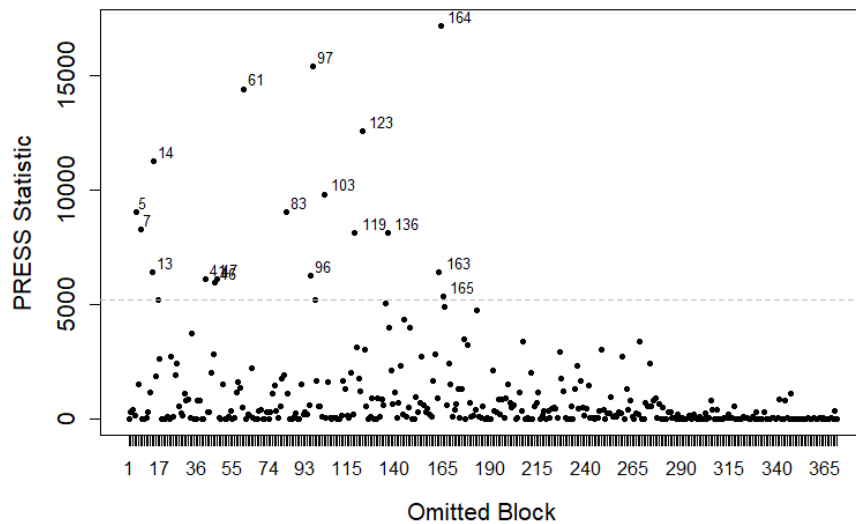
Figure 29 This plot shows the change in the PRESS (Predicted Residual Sum of Squares) as a result of omitting each data point, which is indicative of the influence of that data point on the sensitivity of model predictions.
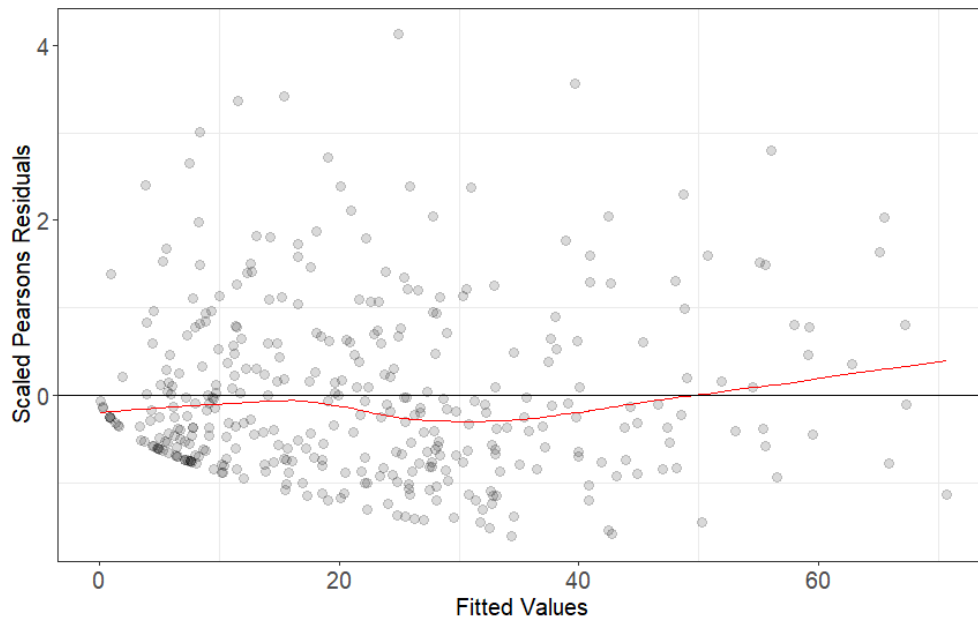
Figure 30 The plot of scaled Pearson residuals against fitted values provides information on the extent to which the model applies an appropriate error structure. If the model was perfect, there would be no discernible patterns in this plot, with residuals scattered about a mean of zero. Given that some pattern is evident, the model does not fully fit; however, overall this plot does not indicate a problem.

Figure 31 Plot of fitted versus observed values. Ideally points would be randomly scattered about the 1:1 line (shown). It is evident that the model is underpredicting variation in the data – many points with an observed value of zero are predicted to be above zero by the model, while the small number of points with high observed values tend to be underestimated in the model fit.
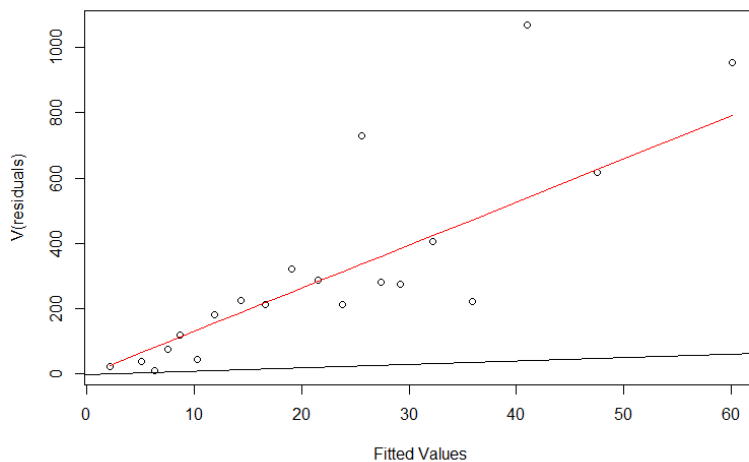
Figure 32 Plot of mean residual variance against mean fitted for each 5% quantile of fitted values. The black line is the 1:1 line while the red line has a gradient equal to the estimated dispersion parameter. A quasi-poisson model (as used) assumes variance is proportional to the mean and therefore ideally the points would be scattered about the red line. This figure therefore also indicates that the data are not truly quasi-poisson, with greater residual variance than expected at higher fitted values.
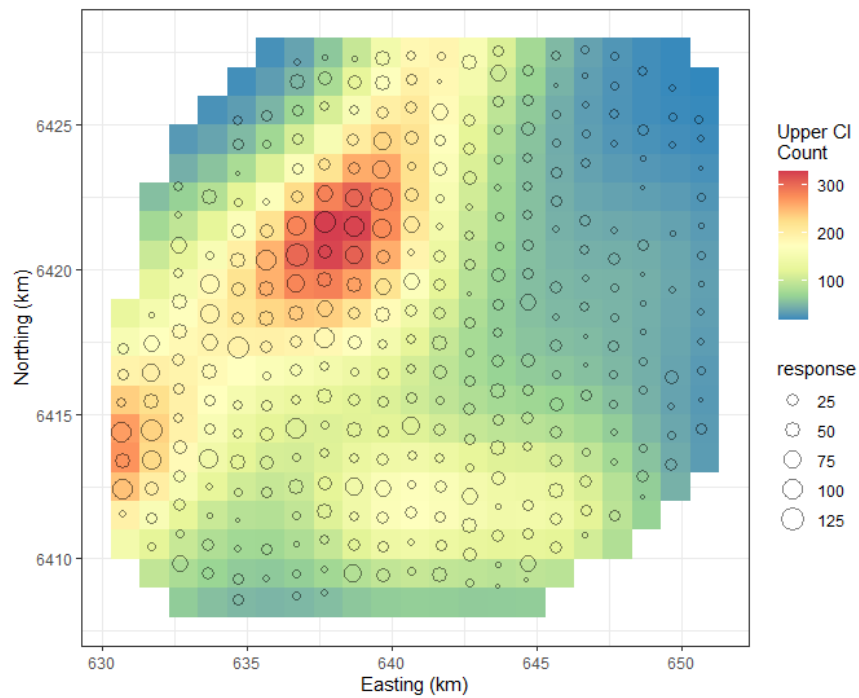
Figure 33 Upper 95% confidence limit of estimated abundance from parametric bootstrapping
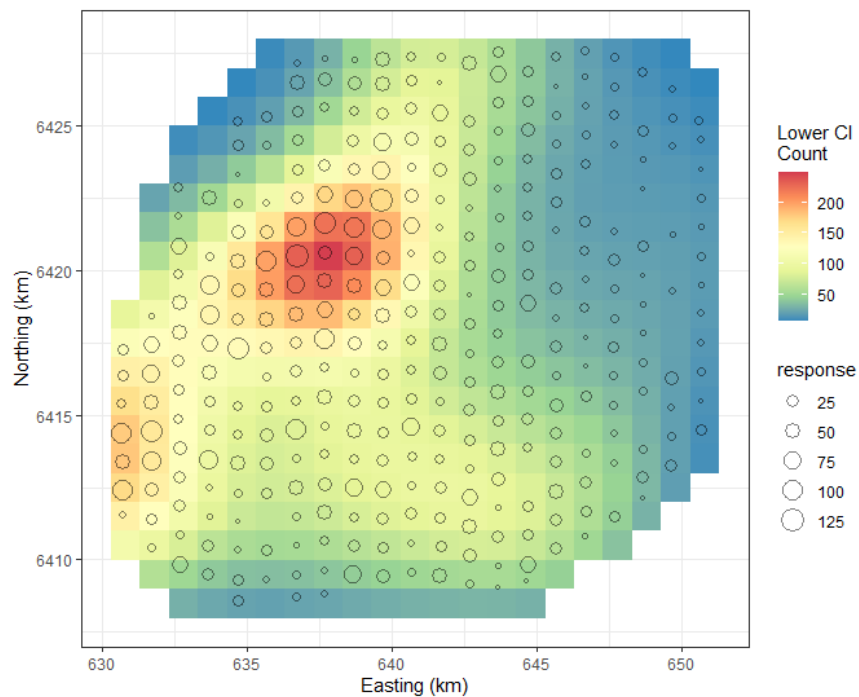
Figure 34 Lower 95% confidence limit of estimated abundance from parametric bootstrapping

1.3 October 2020

```
initial_guillemot_model <- glm(response ~ offset(log(area)),  family = "quasipoisson", data = guillemot_model_data)

knot_grid <- getKnotgrid(guillemot_model_data[, c("x.pos", "y.pos")])
distMats <-makeDists(cbind(guillemot_model_data$x.pos, guillemot_model_data$y.pos), knot_grid)

salsa2dlist <-list(fitnessMeasure = "cv.gamMRSea",
                   cv.opts = list(cv.gamMRSea.seed = 1, K=8),
                   knotgrid = knot_grid,
                   startKnots = 15,
                   minKnots = 2,
                   maxKnots = 20,
                   gap = 0)

salsa2dOutput <-runSALSA2D(initial_guillemot_model,
                           salsa2dlist,
                           distMats$dataDist,
                           distMats$knotDist,
                           panels = guillemot_model_data$TransectID)
```

Figure 35 Code snippet of input parameters to SALSA2D function.

```
> summary(best_guillemot_model)

Call:
gamMRSea(formula = response ~ LRF.g(radiusIndices, dists, radii,
    aR) + offset(log(area)), family = "quasipoisson", data = guillemot_model_data,
    splineParams = splineParams)

Deviance Residuals:
    Min      1Q   Median      3Q      Max
-2.7712  -1.6467  -0.5264   0.6169   5.2688

Coefficients:
                                     Estimate Std. Error Robust S.E. t value Pr(>|t|)
(Intercept)                            1.5130     0.5493      0.4866   3.109  0.00203 **
LRF.g(radiusIndices, dists, radii, aR)b1   15.9713     3.8864      2.9213   5.467 8.61e-08 ***
LRF.g(radiusIndices, dists, radii, aR)b2  -85.7055    22.0322     17.0951  -5.013 8.44e-07 ***
LRF.g(radiusIndices, dists, radii, aR)b3  -17.1663     4.3338      3.6856  -4.658 4.52e-06 ***
LRF.g(radiusIndices, dists, radii, aR)b4    6.9295     1.8161      1.5974   4.338 1.87e-05 ***
LRF.g(radiusIndices, dists, radii, aR)b5   -3.3129     1.2405      1.2385  -2.675  0.00782 **
LRF.g(radiusIndices, dists, radii, aR)b6   -8.2289     2.1756      1.6955  -4.853 1.82e-06 ***
LRF.g(radiusIndices, dists, radii, aR)b7  -58.9437    17.0647     12.1538  -4.850 1.85e-06 ***
LRF.g(radiusIndices, dists, radii, aR)b8   19.0195     5.3164      3.7577   5.061 6.67e-07 ***
LRF.g(radiusIndices, dists, radii, aR)b9    7.9408     2.1892      1.7728   4.479 1.01e-05 ***
LRF.g(radiusIndices, dists, radii, aR)b10 -31.5696     8.8326      7.3057  -4.321 2.01e-05 ***
LRF.g(radiusIndices, dists, radii, aR)b11  35.0609     8.6460      6.5243   5.374 1.39e-07 ***
LRF.g(radiusIndices, dists, radii, aR)b12 -67.3025    20.0835     16.9057  -3.981 8.31e-05 ***
LRF.g(radiusIndices, dists, radii, aR)b13 189.5051    51.6756     41.2878   4.590 6.15e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 2.84911)

    Null deviance: 1024.69  on 370  degrees of freedom
Residual deviance:  947.68  on 357  degrees of freedom
AIC:  NA

Max Panel Size = 21; Number of panels = 21
Number of Fisher Scoring iterations: 5
```

**Figure 36 Summary outputs of the best fitting 2D model**
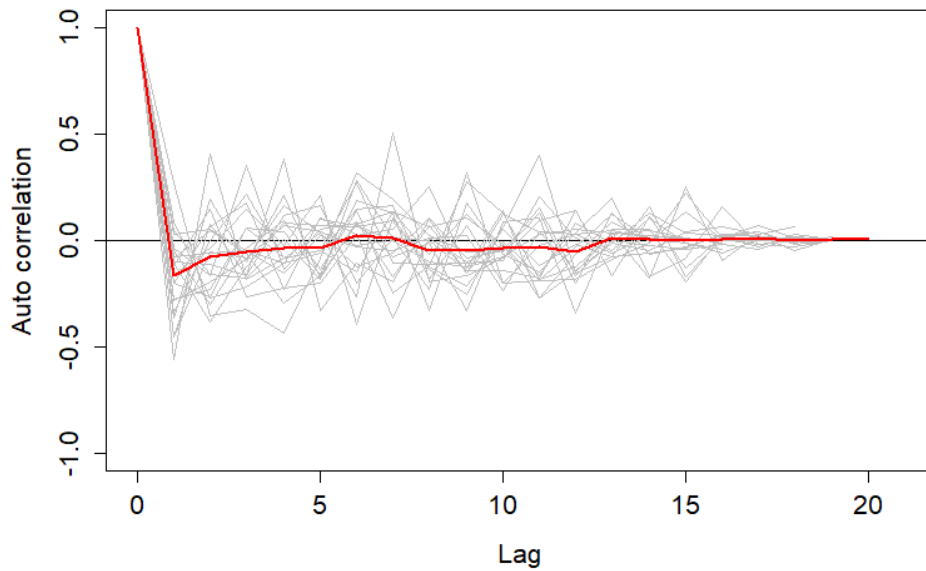
Figure 37 Autocorrelation function plot of best fitting 2D model. Auto correlation drops rapidly to zero for most runs, indicating a lack of auto correlation in the model residuals.

```
> runsTest(residuals(best_guillemot_model, type = "pearson"), emp.distribution=empdist)

        Runs Test - Two sided; Empirical Distribution

data:  residuals(best_guillemot_model, type = "pearson")
Standardized Runs Statistic = -0.051848, p-value = 0.78
```

Figure 38 Results of runs test using empirical distribution. Non-significant p-value (>0.05) indicates no evidence of autocorrelation using this test.
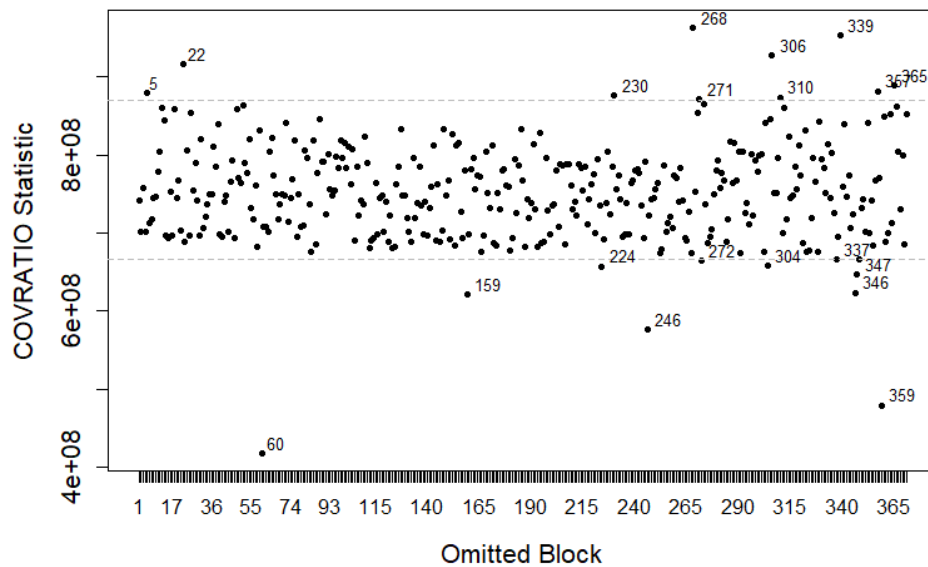
Figure 39 The "COVRATIO" (covariance ratio) measures the change in covariance as a result of removing each data point in turn, which is indicative of how influential that data point is in the precision of parameter estimates.
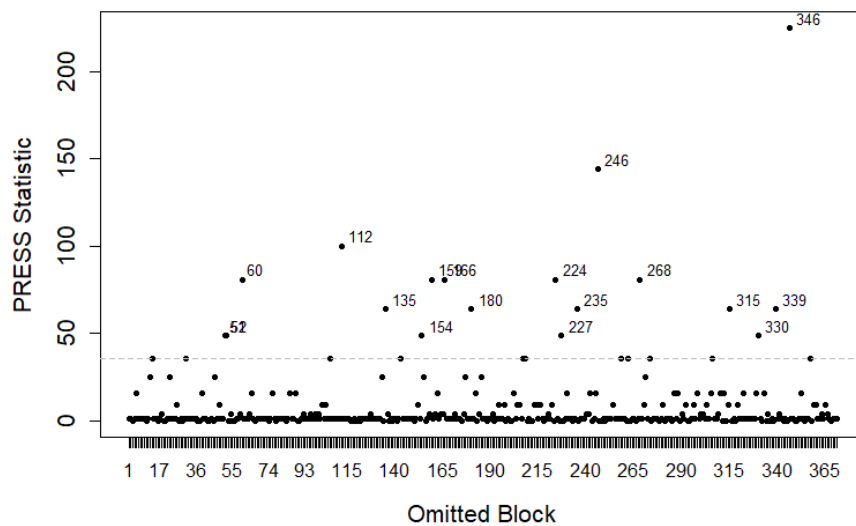
Figure 40 This plot shows the change in the PRESS (Predicted Residual Sum of Squares) as a result of omitting each data point, which is indicative of the influence of that data point on the sensitivity of model predictions.
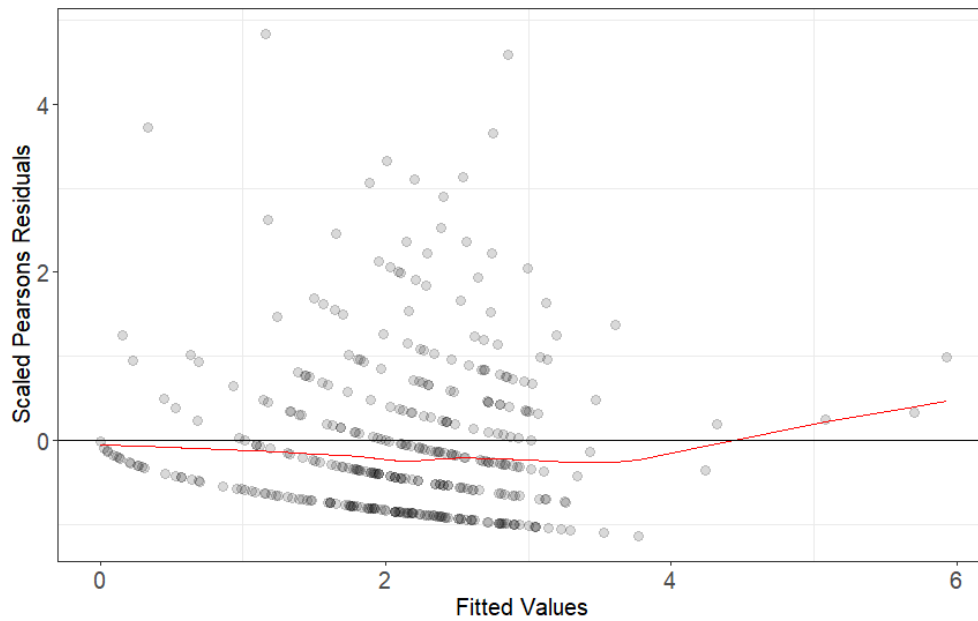
Figure 41 The plot of scaled Pearson residuals against fitted values provides information on the extent to which the model applies an appropriate error structure. If the model was perfect, there would be no discernible patterns in this plot, with residuals scattered about a mean of zero. Given that some pattern is evident, the model does not fully fit; however, overall this plot does not indicate a problem.

Concordance correlation: 0.1871
Marginal R-squared value: 0.1025
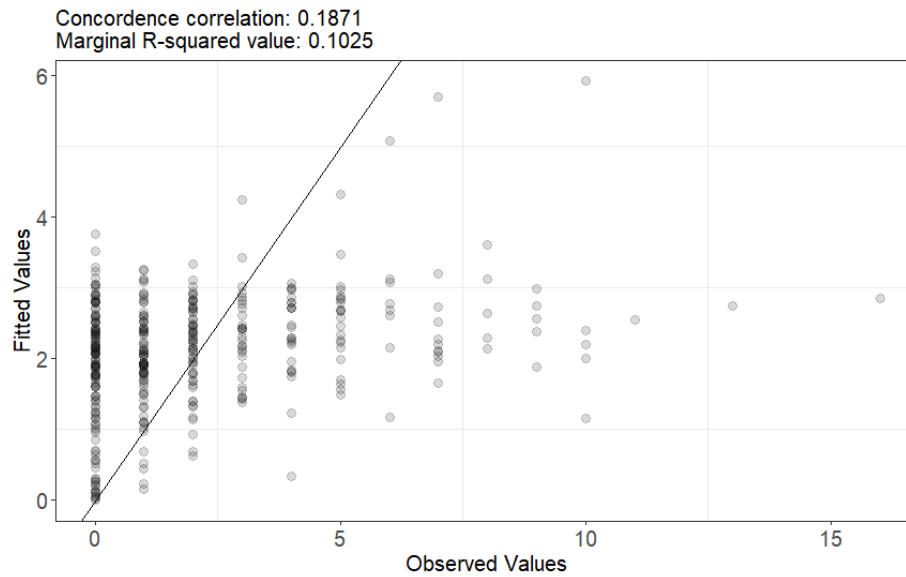
Figure 42 Plot of fitted versus observed values. Ideally points would be randomly scattered about the 1:1 line (shown). It is evident that the model is underpredicting variation in the data – many points with an observed value of zero are predicted to be above zero by the model, while the small number of points with high observed values tend to be underestimated in the model fit.
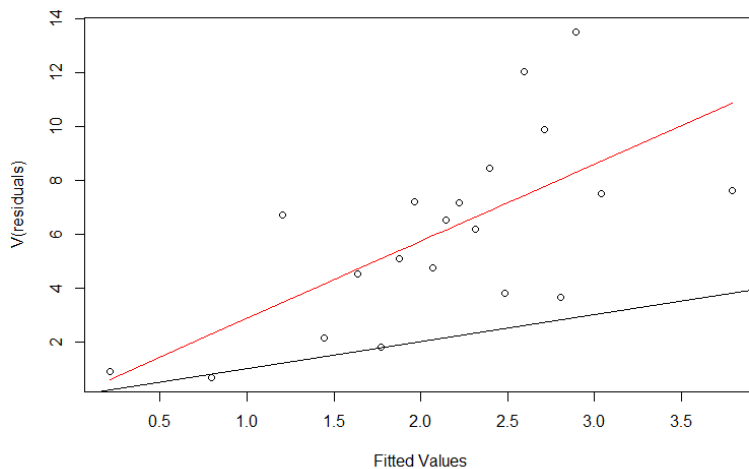
Figure 43 Plot of mean residual variance against mean fitted for each 5% quantile of fitted values. The black line is the 1:1 line while the red line has a gradient equal to the estimated dispersion parameter. A quasi-poisson model (as used) assumes variance is proportional to the mean and therefore ideally the points would be scattered about the red line. This figure therefore also indicates that the data are not truly quasi-poisson, with greater residual variance than expected at higher fitted values.
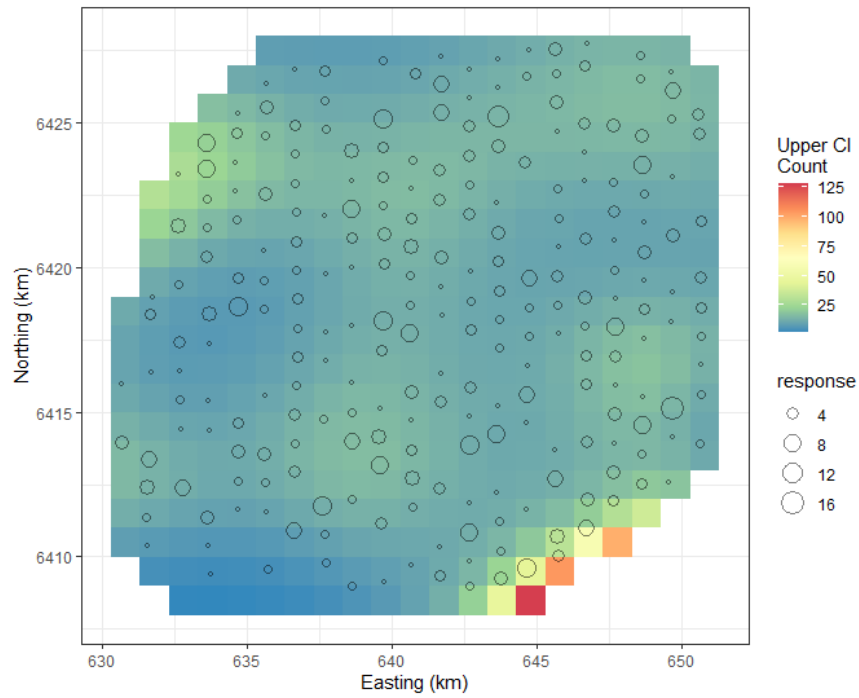
Figure 44 Upper 95% confidence limit of estimated abundance from parametric bootstrapping
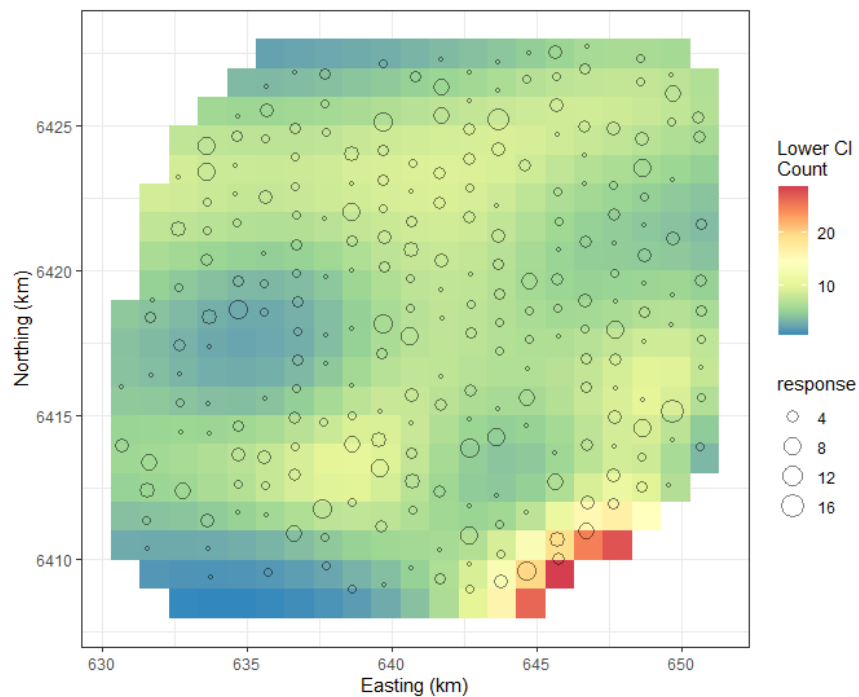
Figure 45 Lower 95% confidence limit of estimated abundance from parametric bootstrapping

1.4   April 2021

```
initial_guillemot_model <- glm(response ~ offset(log(area)), family = "quasipoisson", data = guillemot_model_data)

knot_grid <- getKnotgrid(guillemot_model_data[, c("x.pos", "y.pos")])
distMats <-makeDists(cbind(guillemot_model_data$x.pos, guillemot_model_data$y.pos), knot_grid)

salsa2dlist <-list(fitnessMeasure = "cv.gamMRSea",
                   cv.opts = list(cv.gamMRSea.seed = 1, K=8),
                   knotgrid = knot_grid,
                   startKnots = 15,
                   minKnots = 2,
                   maxKnots = 20,
                   gap = 0)

salsa2dOutput <-runSALSA2D(initial_guillemot_model,
                           salsa2dlist,
                           distMats$dataDist,
                           distMats$knotDist,
                           panels = guillemot_model_data$TransectID)
```

Figure 46 Code snippet of input parameters to SALSA2D function.

```
> summary(best_guillemot_model)

Call:
gamMRSea(formula = response ~ LRF.g(radiusIndices, dists, radii,
    aR) + offset(log(area)), family = "quasipoisson", data = guillemot_model_data,
    splineParams = splineParams)

Deviance Residuals:
    Min      1Q  Median      3Q     Max
-4.1981  -1.0602  -0.1894  0.8067  3.9591

Coefficients:
                                          Estimate Std. Error Robust S.E.  t value Pr(>|t|)
(Intercept)                                 1.1776     0.4571      0.5409    2.177 0.030121 *
LRF.g(radiusIndices, dists, radii, aR)b1  -30.5032     4.0526      6.5476   -4.659 4.50e-06 ***
LRF.g(radiusIndices, dists, radii, aR)b2  -21.6700     3.0028      3.9811   -5.443 9.76e-08 ***
LRF.g(radiusIndices, dists, radii, aR)b3   -4.2854     1.0389      1.2415   -3.452 0.000624 ***
LRF.g(radiusIndices, dists, radii, aR)b4    0.9647     0.4577      0.4517    2.136 0.033385 *
LRF.g(radiusIndices, dists, radii, aR)b5    4.7507     1.1583      1.4080    3.374 0.000822 ***
LRF.g(radiusIndices, dists, radii, aR)b6   16.7679     2.5135      3.2454    5.167 3.98e-07 ***
LRF.g(radiusIndices, dists, radii, aR)b7   22.1310     2.7154      4.5370    4.878 1.62e-06 ***
LRF.g(radiusIndices, dists, radii, aR)b8    2.2418     0.7033      0.9054    2.476 0.013755 *
LRF.g(radiusIndices, dists, radii, aR)b9   22.7693     3.8126      5.2455    4.341 1.85e-05 ***
LRF.g(radiusIndices, dists, radii, aR)b10  18.5919     3.0258      3.9011    4.766 2.75e-06 ***
LRF.g(radiusIndices, dists, radii, aR)b11 -13.4625     2.6856      3.3534   -4.015 7.26e-05 ***
LRF.g(radiusIndices, dists, radii, aR)b12  -9.2828     2.0343      2.3243   -3.994 7.90e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 2.18924)

    Null deviance: 1694.56  on 368  degrees of freedom
Residual deviance:  797.22  on 356  degrees of freedom
AIC:  NA

Max Panel Size = 20; Number of panels = 21
Number of Fisher Scoring iterations: 5
```

**Figure 47 Summary outputs of the best fitting 2D model**

APEM
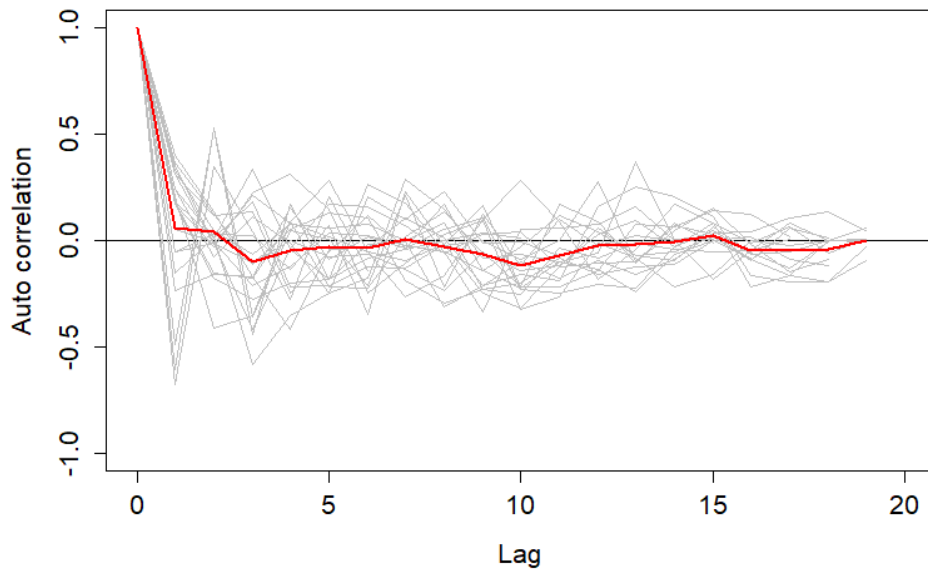
Figure 48 Autocorrelation function plot of best fitting 2D model. Auto correlation drops rapidly to zero for most runs, indicating a lack of auto correlation in the model residuals.

```
> runsTest(residuals(best_guillemot_model, type = "pearson"), emp.distribution=empdist)

        Runs Test - Two sided; Empirical Distribution

data:  residuals(best_guillemot_model, type = "pearson")
Standardized Runs Statistic = -2.5542, p-value = 0.004
```

Figure 49 Results of runs test using empirical distribution. Significant p-value (<0.05) indicates evidence of autocorrelation using this test.
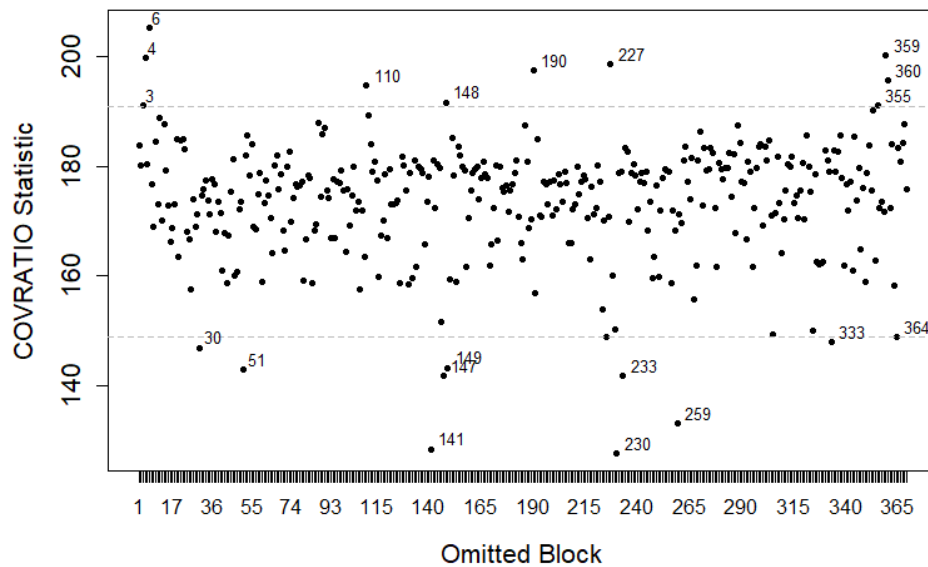
Figure 50 The "COVRATIO" (covariance ratio) measures the change in covariance as a result of removing each data point in turn, which is indicative of how influential that data point is in the precision of parameter estimates.
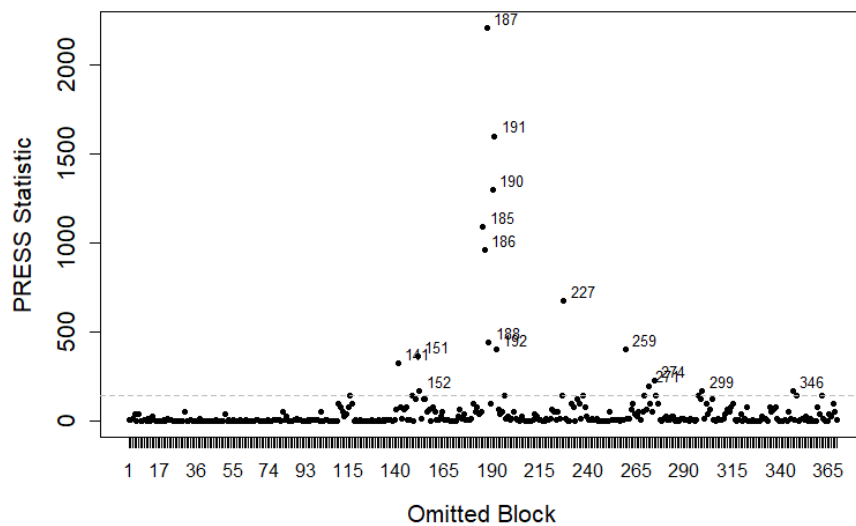
Figure 51 This plot shows the change in the PRESS (Predicted Residual Sum of Squares) as a result of omitting each data point, which is indicative of the influence of that data point on the sensitivity of model predictions.
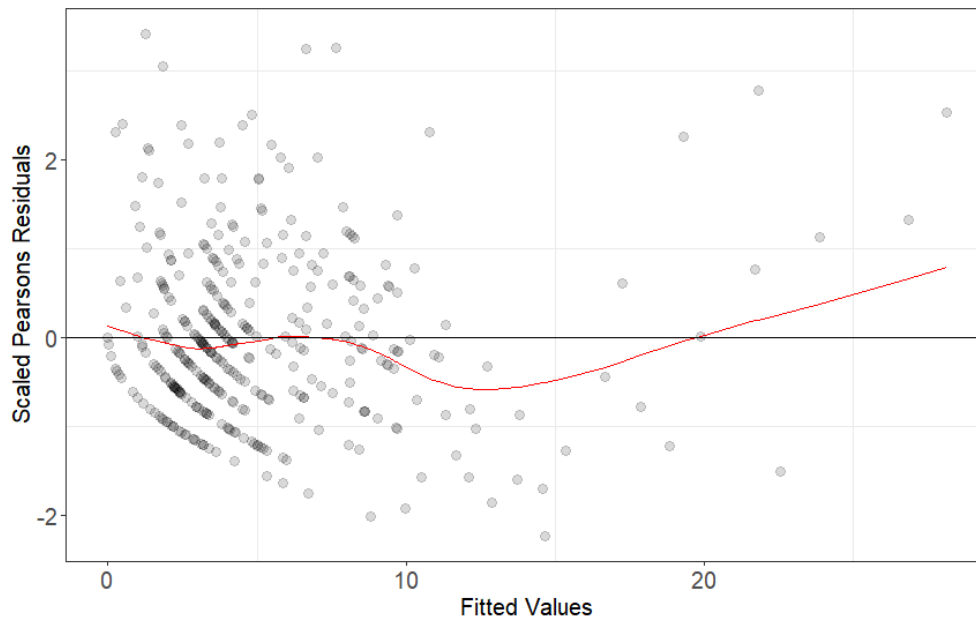
Figure 52 The plot of scaled Pearson residuals against fitted values provides information on the extent to which the model applies an appropriate error structure. If the model was perfect, there would be no discernible patterns in this plot, with residuals scattered about a mean of zero. Given that some pattern is evident, the model does not fully fit; however, overall this plot does not indicate a problem.
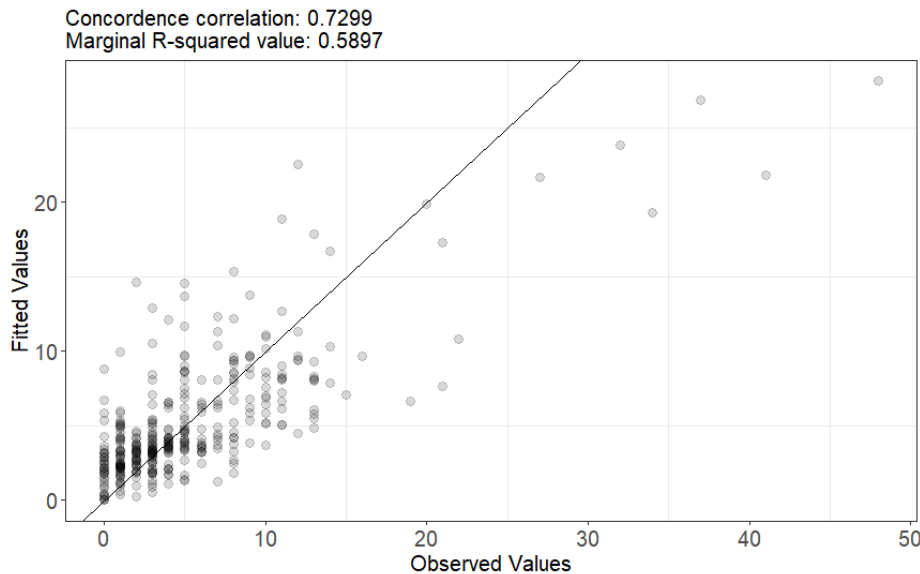
Figure 53 Plot of fitted versus observed values. Ideally points would be randomly scattered about the 1:1 line (shown). It is evident that the model is underpredicting variation in the data – many points with an observed value of zero are predicted to be above zero by the model, while the small number of points with high observed values tend to be underestimated in the model fit.
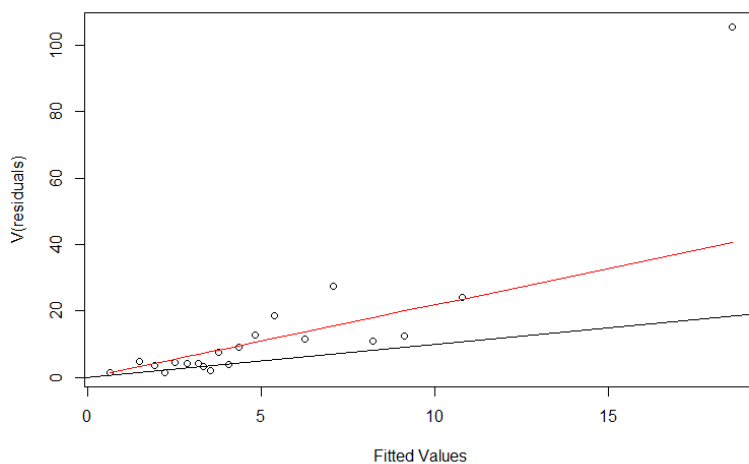
Figure 54 Plot of mean residual variance against mean fitted for each 5% quantile of fitted values. The black line is the 1:1 line while the red line has a gradient equal to the estimated dispersion parameter. A quasi-poisson model (as used) assumes variance is proportional to the mean and therefore ideally the points would be scattered about the red line. This figure therefore also indicates that the data are not truly quasi-poisson, with greater residual variance than expected at higher fitted values.
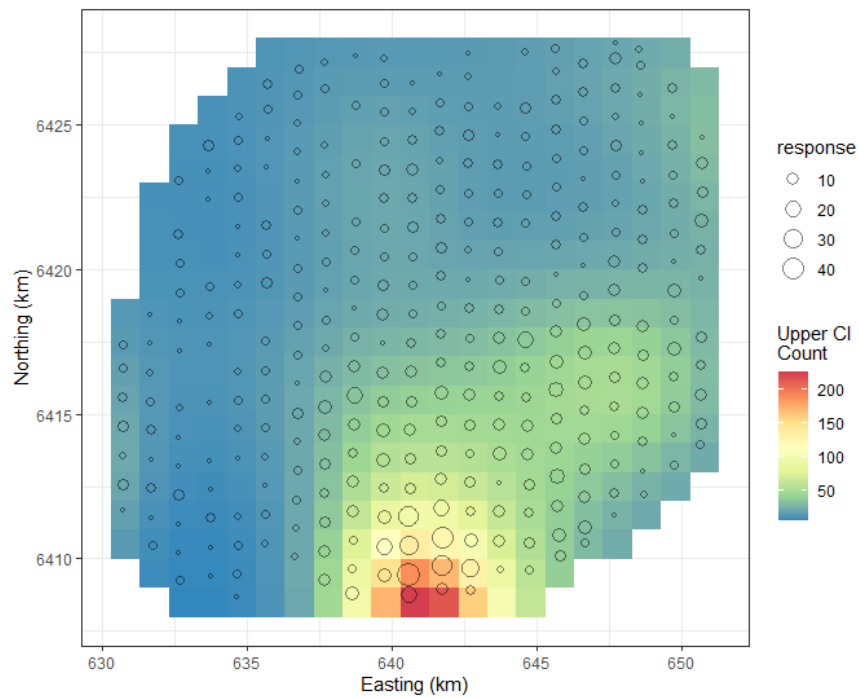
Figure 55 Upper 95% confidence limit of estimated abundance from parametric bootstrapping
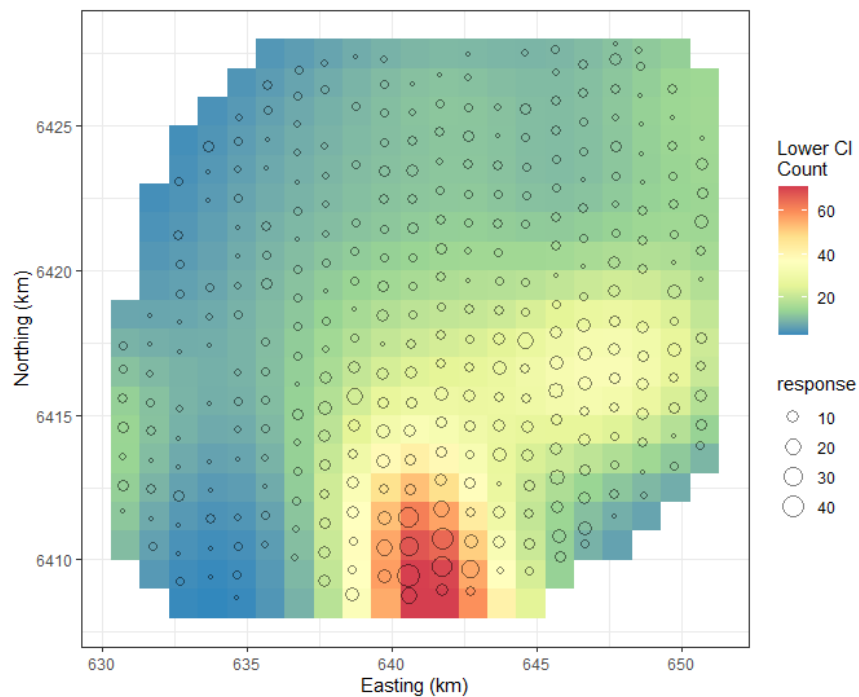
Figure 56 Lower 95% confidence limit of estimated abundance from parametric bootstrapping

1.5   May 2021

```
initial_guillemot_model <- glm(response ~ offset(log(area)),  family = "quasipoisson", data = guillemot_model_data)

knot_grid <- getKnotgrid(guillemot_model_data[, c("x.pos", "y.pos")])
distMats <-makeDists(cbind(guillemot_model_data$x.pos, guillemot_model_data$y.pos), knot_grid)

salsa2dlist <-list(fitnessMeasure = "cv.gamMRSea",
                   cv.opts = list(cv.gamMRSea.seed = 1, K=8),
                   knotgrid = knot_grid,
                   startKnots =10,
                   minKnots = 2,
                   maxKnots = 20,
                   gap = 0)

salsa2dOutput <-runSALSA2D(initial_guillemot_model,
                           salsa2dlist,
                           distMats$dataDist,
                           distMats$knotDist,
                           panels = guillemot_model_data$TransectID)
```

Figure 57 Code snippet of input parameters to SALSA2D function.

```
> summary(best_guillemot_model)

Call:
gamMRSea(formula = response ~ LRF.g(radiusIndices, dists, radii,
    aR) + offset(log(area)), family = "quasipoisson", data = guillemot_model_data,
    splineParams = splineParams)

Deviance Residuals:
    Min       1Q    Median       3Q       Max
-0.5698  -0.4734  -0.4243  -0.3471    5.2074

Coefficients:
                                        Estimate Std. Error Robust S.E. t value Pr(>|t|)
(Intercept)                                0.590      1.667       1.354   0.436    0.663
LRF.g(radiusIndices, dists, radii, aR)b1  -1.628      1.592       1.211  -1.344    0.180
LRF.g(radiusIndices, dists, radii, aR)b2  -1.766      2.100       1.686  -1.048    0.295

(Dispersion parameter for quasipoisson family taken to be 1.757522)

    Null deviance: 191.35  on 368  degrees of freedom
Residual deviance: 187.32  on 366  degrees of freedom
AIC:  NA

Max Panel Size = 21; Number of panels = 21
Number of Fisher Scoring iterations: 6
```

**Figure 58 Summary outputs of the best fitting 2D model**
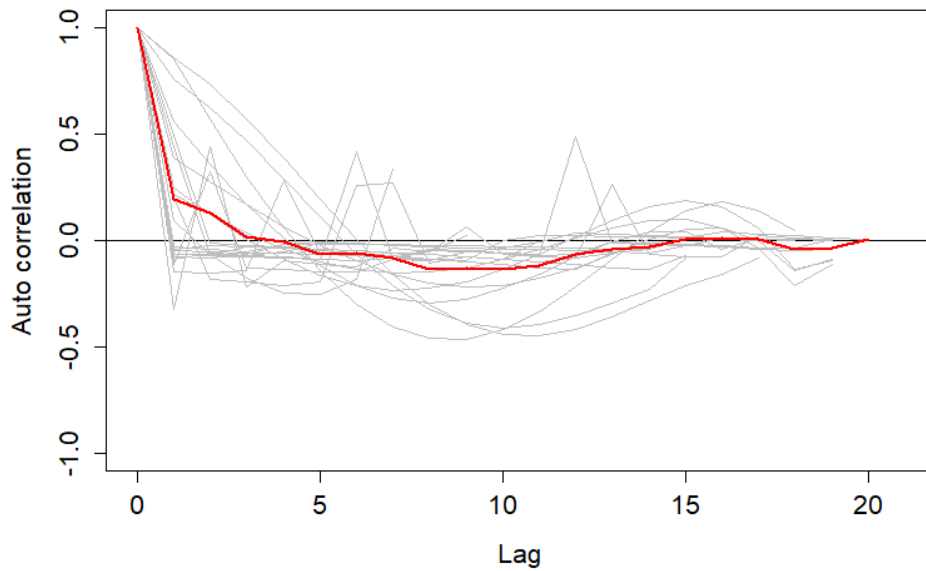
Figure 59 Autocorrelation function plot of best fitting 2D model. Auto correlation drops eventually to zero for most runs, some evidence of auto correlation in the model residuals.

```
> runsTest(residuals(best_guillemot_model, type = "pearson"), emp.distribution=empdist)

        Runs Test - Two sided; Empirical Distribution

data:  residuals(best_guillemot_model, type = "pearson")
Standardized Runs Statistic = -11.937, p-value = 0.67
```

Figure 60 Results of runs test using empirical distribution. Non-significant p-value (>0.05) indicates no evidence of autocorrelation using this test.
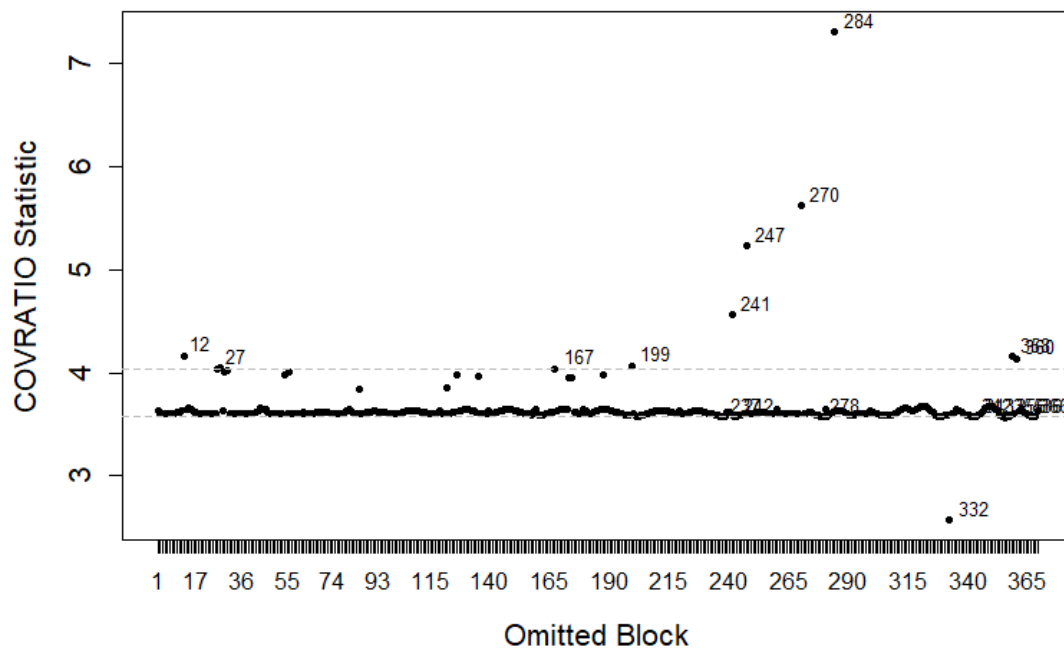
Figure 61 The "COVRATIO" (covariance ratio) measures the change in covariance as a result of removing each data point in turn, which is indicative of how influential that data point is in the precision of parameter estimates. All small number of very influential data points are a concern in this model.
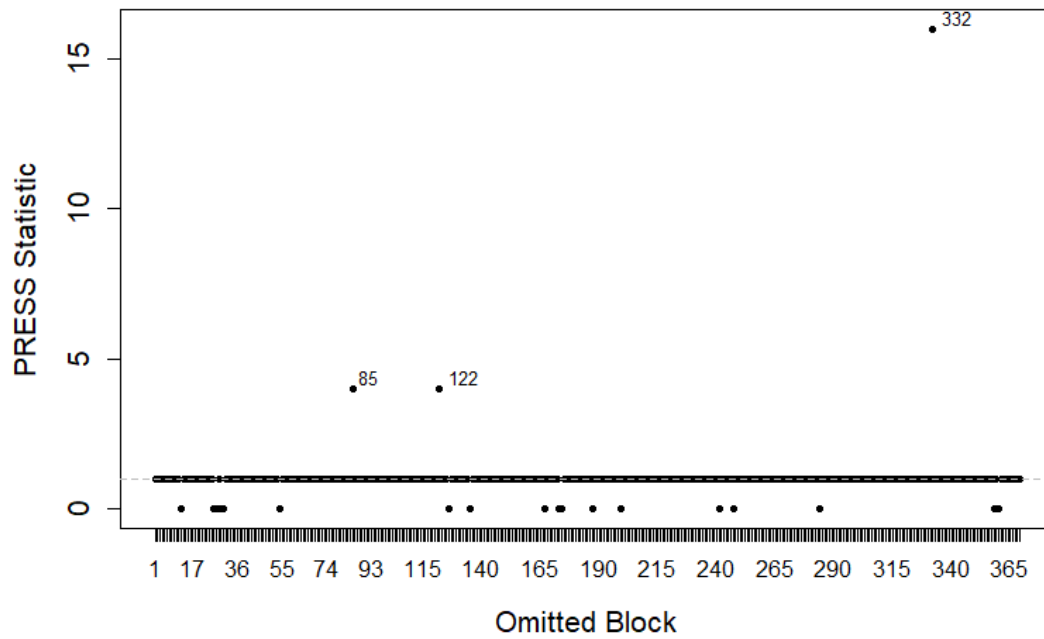
Figure 62 This plot shows the change in the PRESS (Predicted Residual Sum of Squares) as a result of omitting each data point, which is indicative of the influence of that data point on the sensitivity of model predictions. All small number of very influential data points are a concern in this model.
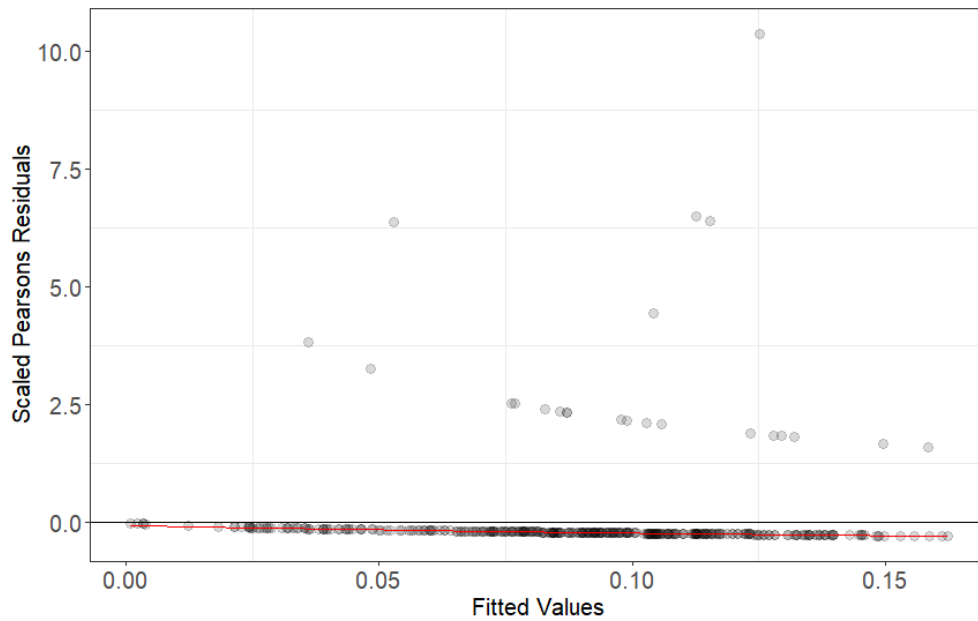
Figure 63 The plot of scaled Pearson residuals against fitted values provides information on the extent to which the model applies an appropriate error structure. If the model was perfect, there would be no discernible patterns in this plot, with residuals scattered about a mean of zero. Given that some pattern is evident, the model does not fully fit.
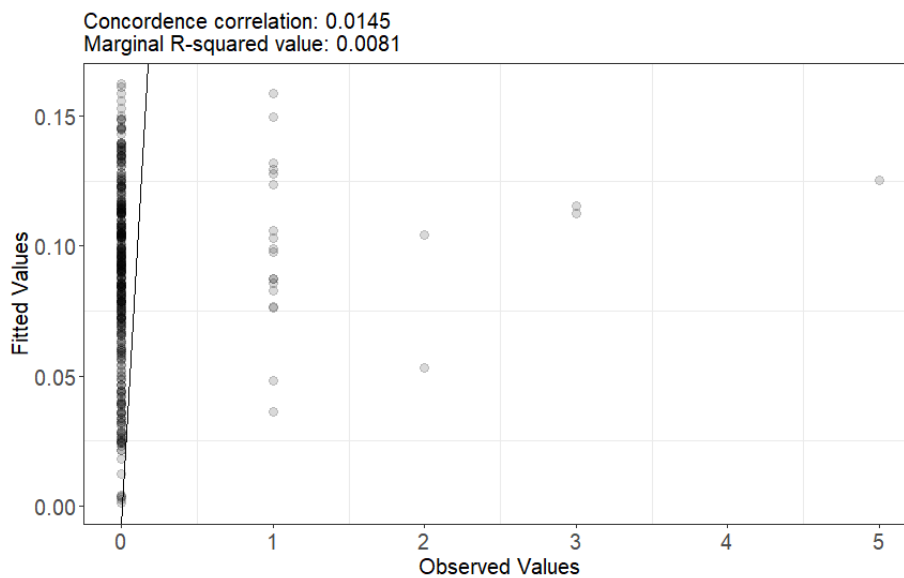

Figure 64 Plot of fitted versus observed values. Ideally points would be randomly scattered about the 1:1 line (shown). It is evident that the model is underpredicting variation in the data – many points with an observed value of zero are predicted to be above zero by the model, while the small number of points with high observed values tend to be underestimated in the model fit.
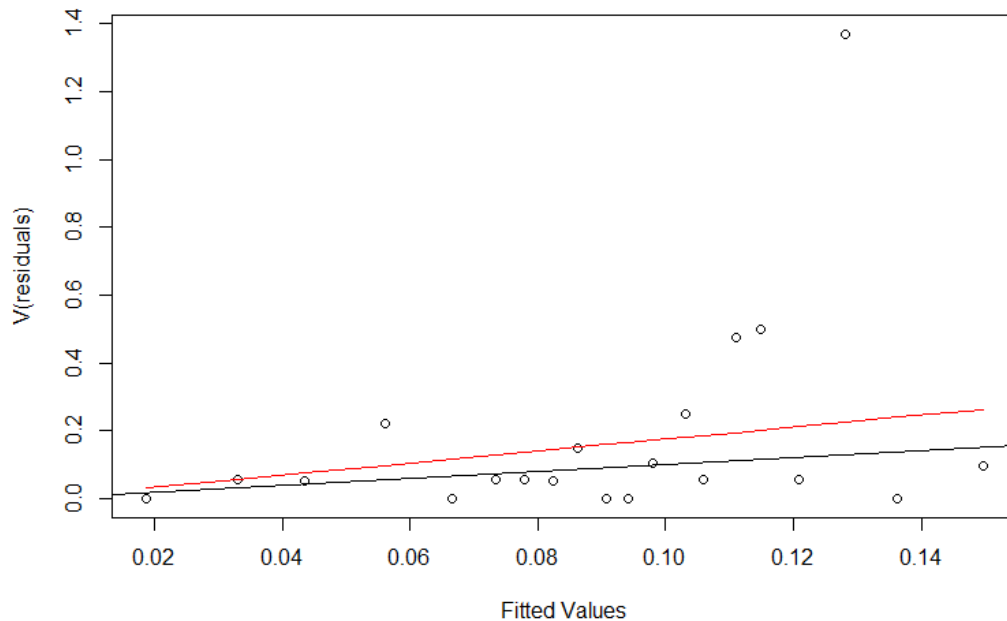
Figure 65 Plot of mean residual variance against mean fitted for each 5% quantile of fitted values. The black line is the 1:1 line while the red line has a gradient equal to the estimated dispersion parameter. A quasi-poisson model (as used) assumes variance is proportional to the mean and therefore ideally the points would be scattered about the red line. This figure therefore also indicates that the data are not truly quasi-poisson, with greater residual variance than expected at higher fitted values.
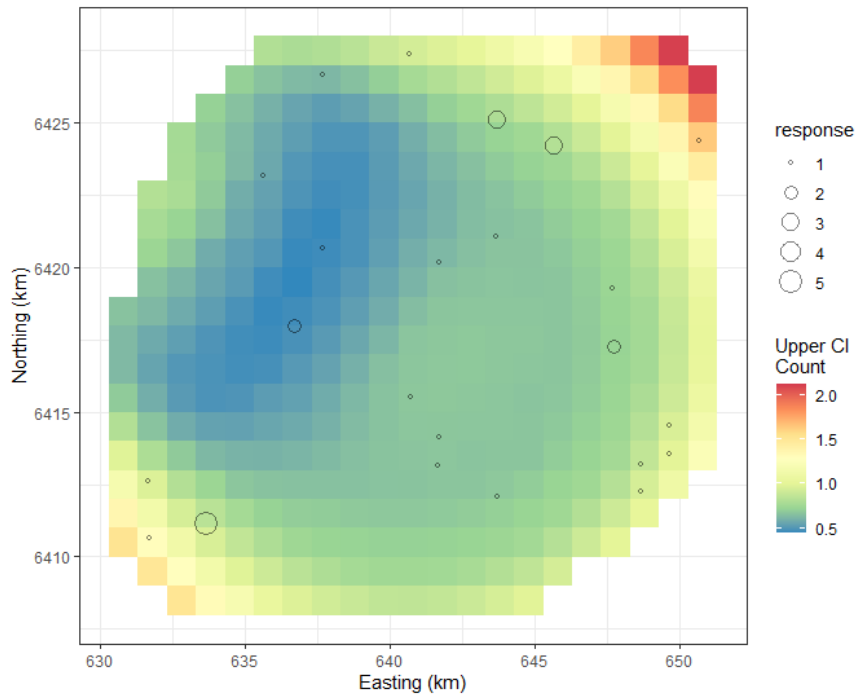
Figure 66 Upper 95% confidence limit of estimated abundance from parametric bootstrapping
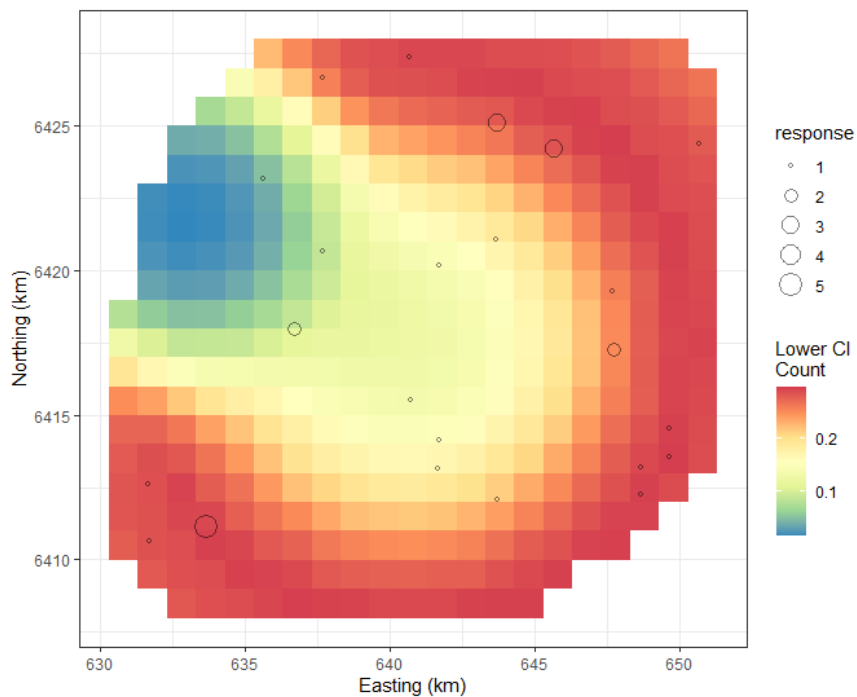
Figure 67 Lower 95% confidence limit of estimated abundance from parametric bootstrapping

1.6   August 2021

```
initial_guillemot_model <- glm(response ~ offset(log(area)),  family = "quasipoisson", data = guillemot_model_data)

knot_grid <- getKnotgrid(guillemot_model_data[, c("x.pos", "y.pos")])
distMats <-makeDists(cbind(guillemot_model_data$x.pos, guillemot_model_data$y.pos), knot_grid)

salsa2dlist <-list(fitnessMeasure = "cv.gamMRSea",
                   cv.opts = list(cv.gamMRSea.seed = 1, K=8),
                   knotgrid = knot_grid,
                   startKnots =8,
                   minKnots = 2,
                   maxKnots = 20,
                   gap = 0)

salsa2dOutput <-runSALSA2D(initial_guillemot_model,
                   salsa2dlist,
                   distMats$dataDist,
                   distMats$knotDist,
                   panels = guillemot_model_data$TransectID)
```

Figure 68 Code snippet of input parameters to SALSA2D function.

```
> summary(best_guillemot_model)

Call:
gamMRSea(formula = response ~ LRF.g(radiusIndices, dists, radii,
    aR) + offset(log(area)), family = "quasipoisson", data = guillemot_model_data,
    splineParams = splineParams)

Deviance Residuals:
    Min      1Q   Median      3Q      Max
-2.0319  -0.8085  -0.6313  -0.3155   3.9084

Coefficients:
                                          Estimate Std. Error Robust S.E. t value Pr(>|t|)
(Intercept)                                -0.3276     0.2163      0.2444  -1.340   0.1810
LRF.g(radiusIndices, dists, radii, aR)b1    0.8639     0.4520      0.4365   1.979   0.0486 *
LRF.g(radiusIndices, dists, radii, aR)b2    2.8593     0.3008      0.3823   7.479 5.59e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 1.59926)

    Null deviance: 561.14  on 368  degrees of freedom
Residual deviance: 406.83  on 366  degrees of freedom
AIC:  NA

Max Panel Size = 20; Number of panels = 21
Number of Fisher Scoring iterations: 6
```
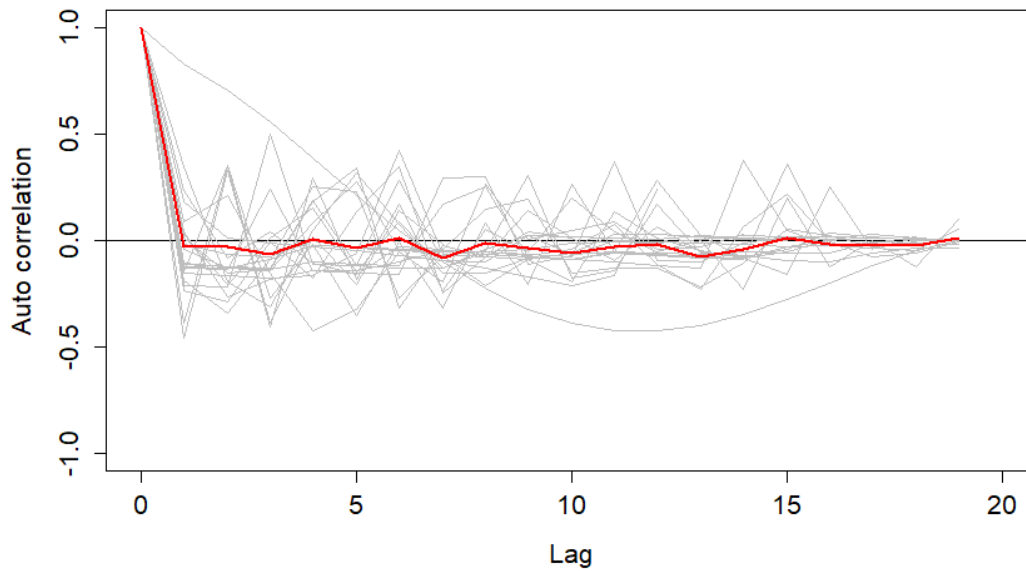
**Figure 69 Summary outputs of the best fitting 2D model**

Figure 70 Autocorrelation function plot of best fitting 2D model. Auto correlation drops rapidly to zero for most runs, indicating a lack of auto correlation in the model residuals.

```
> runsTest(residuals(best_guillemot_model, type = "pearson"), emp.distribution=empdist)

        Runs Test - Two sided; Empirical Distribution

data:  residuals(best_guillemot_model, type = "pearson")
Standardized Runs Statistic = -6.7245, p-value = 0.816
```

Figure 71 Results of runs test using empirical distribution. Non-significant p-value (>0.05) indicates no evidence of autocorrelation using this test.
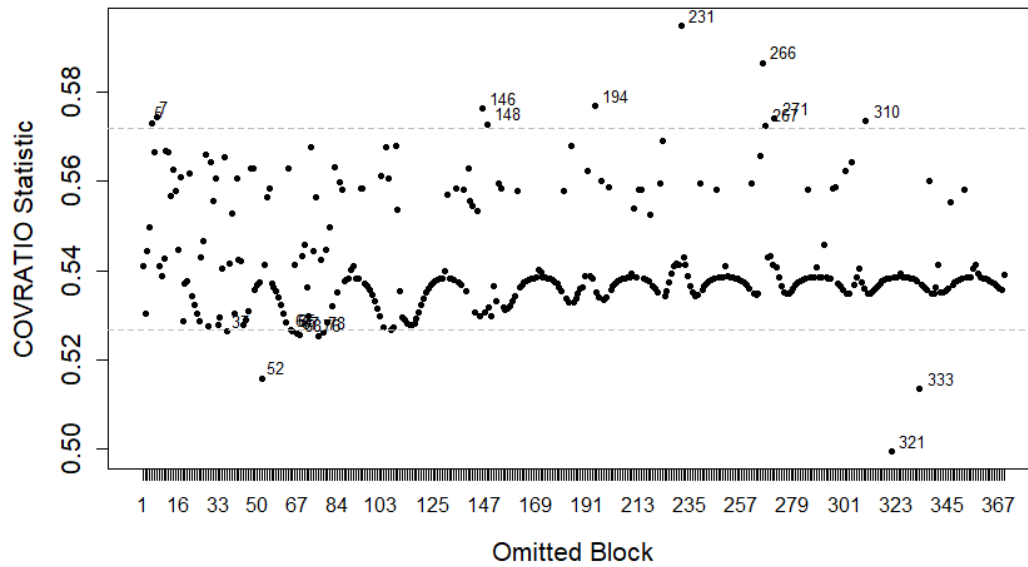
Figure 72 The "COVRATIO" (covariance ratio) measures the change in covariance as a result of removing each data point in turn, which is indicative of how influential that data point is in the precision of parameter estimates.
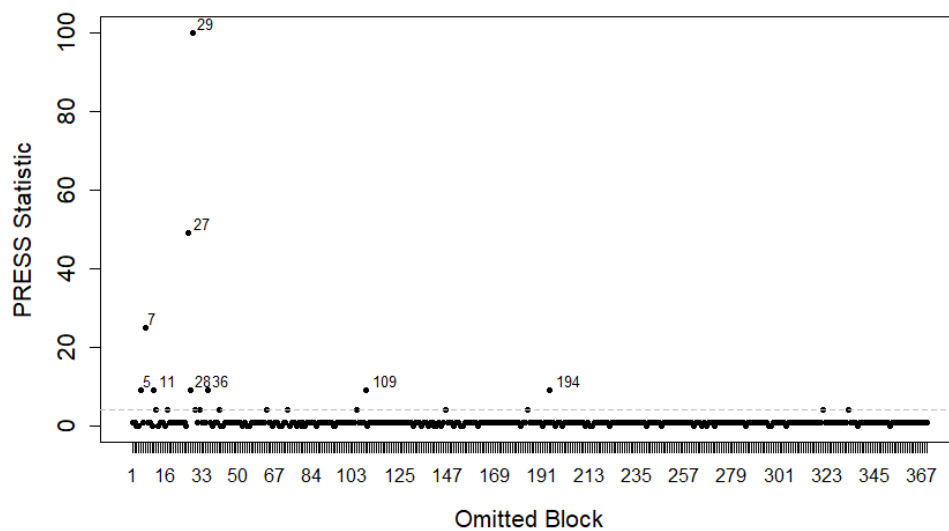

Figure 73 This plot shows the change in the PRESS (Predicted Residual Sum of Squares) as a result of omitting each data point, which is indicative of the influence of that data point on the sensitivity of model predictions.
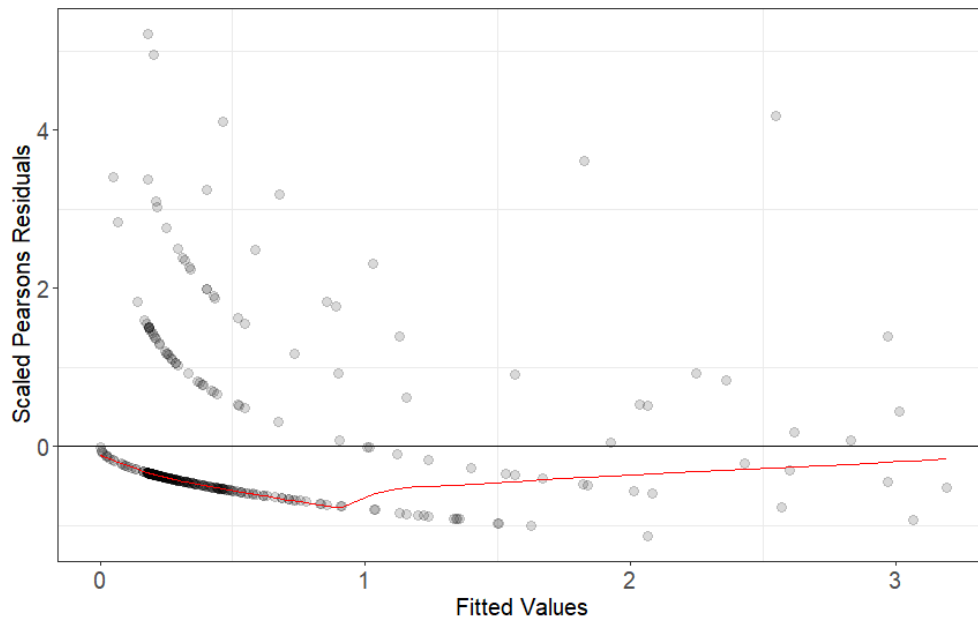
Figure 74 The plot of scaled Pearson residuals against fitted values provides information on the extent to which the model applies an appropriate error structure. If the model was perfect, there would be no discernible patterns in this plot, with residuals scattered about a mean of zero. Given that some pattern is evident, the model does not fully fit; however, overall this plot does not indicate a problem.
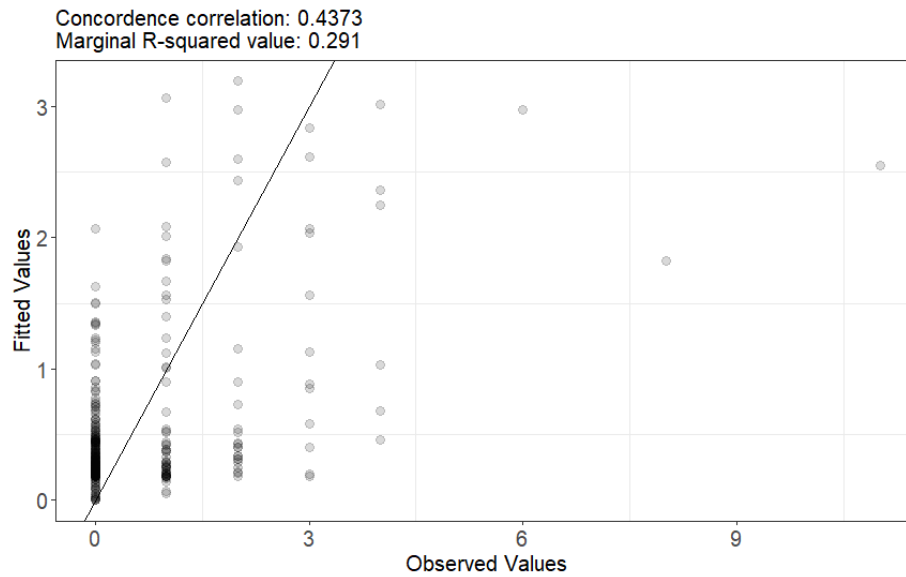
Figure 75 Plot of fitted versus observed values. Ideally points would be randomly scattered about the 1:1 line (shown). It is evident that the model is underpredicting variation in the data – many points with an observed value of zero are predicted to be above zero by the model, while the small number of points with high observed values tend to be underestimated in the model fit.
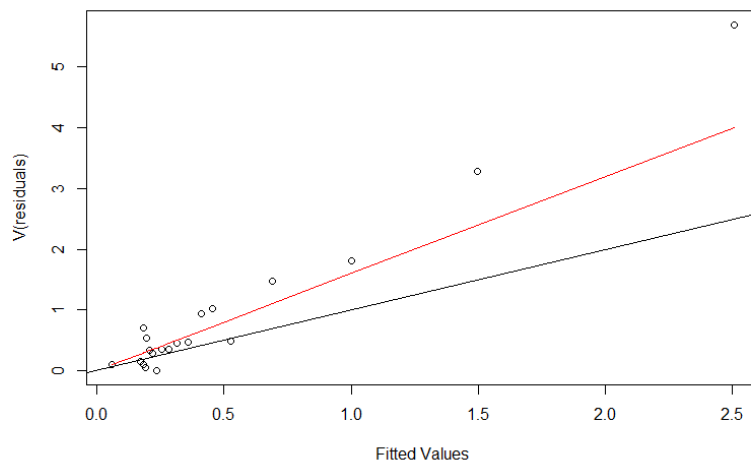
Figure 76 Plot of mean residual variance against mean fitted for each 5% quantile of fitted values. The black line is the 1:1 line while the red line has a gradient equal to the estimated dispersion parameter. A quasi-poisson model (as used) assumes variance is proportional to the mean and therefore ideally the points would be scattered about the red line. This figure therefore also indicates that the data are not truly quasi-poisson, with greater residual variance than expected at higher fitted values.
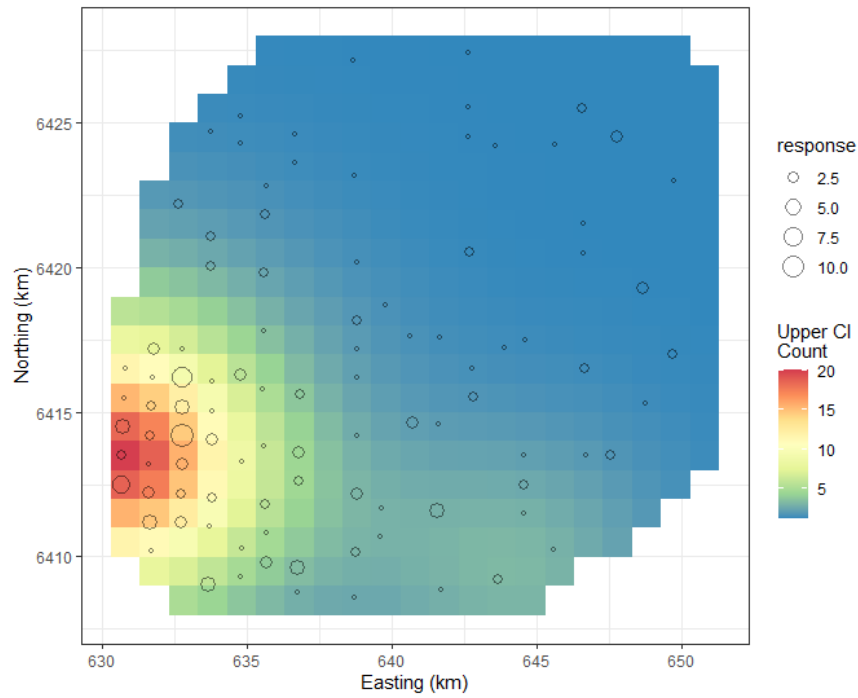
Figure 77 Upper 95% confidence limit of estimated abundance from parametric bootstrapping
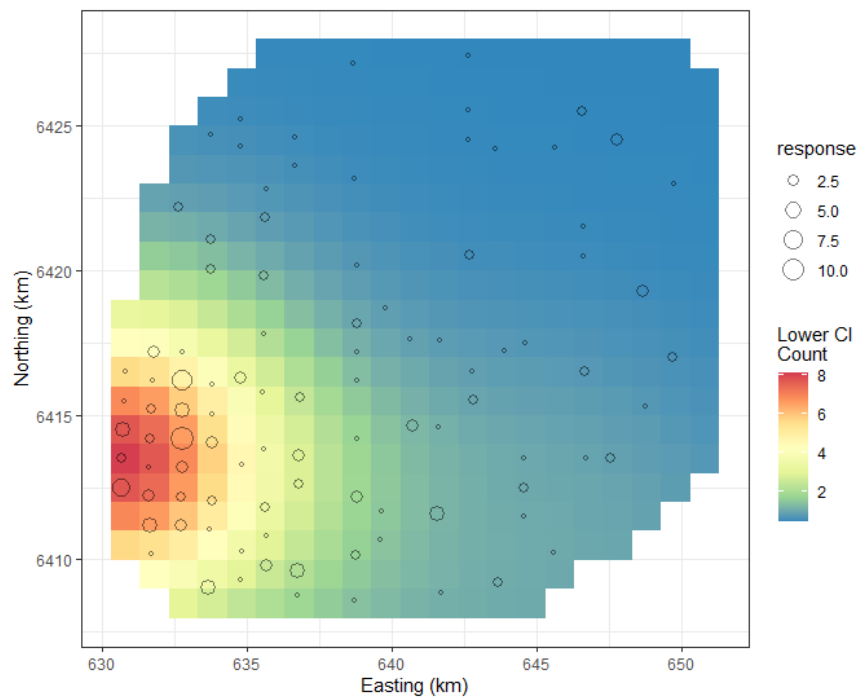
Figure 78 Lower 95% confidence limit of estimated abundance from parametric bootstrapping