MMath Project:

Chris Martin

2022-03-31

Decleration: *I certify that this project report has been written by me, is a record of work carried out by me, and is essentially different from work undertaken for any other purpose or assessment.*

All of the code discussed and used throughout this report can be found at the following GitHub repository: https://github.com/Christopher-98/MMath-Project

Contents

1 Abstract

2 Introduction

The aim of this report is to detail to extensions made to the *dsims* R package through the inclusion of density surface modelling (DSM) simulations. The intended readership of this report is someone with statistical experience and with an interest in distance sampling for estimating wildlife abundance. Further knowledge on distance sampling survey design would be beneficial to understanding some elements of the report in greater detail however this is not required. Analysis will be undertaken to compare the abundance estimates generated through a design based approach, distance sampling, and the model based approach, density surface modelling, to ensure the validity of the new method. By including the model based approach, this will allow researchers the opportunity to run simulations on distance sampling designs while allowing abundance estimations for specific regions to be generated through the density surface model. This ability to predict abundance estimates over very specific regions is a key reason for the inclusion of DSM, alongside the potential for more accurate error estimation than the existing methods used by distance sampling models. A variety of different simulations will be carried out to establish in which scenarios the each method performs well and potentially uncover opportunities where DSM proves to be a substantial improvement over the existing distance sampling method. As the standard distance sampling simulations are often used to decided on the survey design to use for real world studies, the inclusion of DSM will give an added level of assurance to recommendations based on these simulations. For example, if a particular study was looking to undertake both density surface modelling and distance sampling analysis, the inclusion of both within the simulation at the design stage could provide assurances that if a particular design appears better for DSM, that it is also a better design for distance sampling analysis. Should there be no design which is best for both methods, this would allow the user to mitigate the impact by enabling them to choose which analysis type they wish prioritise and select a design based upon that. While the specifics of good survey design are not covered within this report, Buckland et al. (2015) covers this extensively within Chapters 2 & 3 and is recommended for those looking to delve into those aspects of the simulations further.

3   Background research

Informed from Buckland et al. (2015) One of the key aims in areas of applied ecological research is to determine the abundance of a particular population of interest, such as in a periodic way to monitor its development over time and determine changes, or to evaluate the potential effect of a new factor, such as a human disturbance. The size of the population can determine the importance of any new factors, with a smaller populations more under threat from a given factor compared to an abundant one. One option for determining a populations size is to count every single individual, known as a census, similar to the UK completing a Census of its population every 10 years. However, in the natural world, this is only realistically possible in the simplest instances and therefore a different approach must be used. Researches often use some form of sampling method to conduct a sample of the target population and draw conclusions for the overall population based on this sample.

The two most common methods of sampling for ecological populations are Mark-recapture and distance sampling, where information on the detectability of animals comes from the capture histories of individuals or the distances at which observations are made respectively. Mark and recapture methods, while important, are not the focus of this report and hence will not be further discussed.

Distance sampling was first introduced by Buckland et al. (1993) and includes a variety of techniques such as line and point transect sampling, which can then be used to estimate animal abundance using information on the distances to the individuals or clusters observed. The underpinning theory is that if the probability of animal detection can be estimated based on the sample observed, this can allow for estimates on how many animals were not observed and can therefore correct the abundance estimates to take this information into account.

The two simple techniques of distance sampling are, as mentioned above, line transect sampling and point transect sampling, with subsequent more complex techniques being extensions of these in one aspect or another. Line transect sampling consists of a set of lines being placed over a study area by some predetermined method, for example systematic with regular spacing between each or randomly generated. The observer moves along each line, known as a transect, looking for animals or animal groups, referred to as clusters. These are defined by Buckland et al. (2015) as "a group of animals with a well defined location for the group centre." For any animal or cluster the observer detects as they make their way along each transect, the observer estimates or calculates the perpendicular distance x of the animal or cluster from the nearest point of the line. For point transect sampling, the transects are a set of points placed over the study area, with different placement methods available as with line transects, however, most common is a systematically spaced grid. At each point, the observer records any individuals or clusters observed from the point, along with the distance r from the point at which the observation was made.

Distance sampling can therefore be thought of as a method of plot sampling, with the additional factor of not every animal on the plots being observed. For this, the plots in line transect sampling are rectangles of dimension $2wl$, where l is the length of a given transect and may change between transects depending on the survey design and area shape, and w is the truncation distance. This is the distance from the line, beyond which observations

are not recorded if the truncation distance is determined prior to the study. However, if the truncation distance is determined during the analysis phase, then this is the distance beyond which the observations are excluded from the analysis. For the plots in point transect sampling, these are circles of area πw^2 with the plot radius w being the truncation distance.

These different techniques are then used to calculate the animal density D and therefore the overall abundance $N = DA$, where A represents the area of the study region. This is done by initially taking the total number of animals observed n and dividing by the area of the plots a to give $\hat{D} = n/a$. This then leads to the abundance estimate for the overall study area being $\hat{N} = \frac{nA}{a}$. For line transects, if we have k lines in total, then from above we have that $a = 2wL$ where $L = \sum_{i=1}^{k} l_i$ is the the sum of the individual transect lengths l_i. This leads to the overall abundance estimate being:

$$\hat{N} = \frac{nA}{2wL}$$

And, if we say have k point transects, from above $a = k\pi w^2$ leading to the abundance estimate of:

$$\hat{N} = \frac{nA}{k\pi w^2}$$

3.1 Detection function

To take into account the fact that not all animals or clusters within each transect are observed, the probability of detecting an animal with a transect must be estimated. This is done by using the distances to the animals observed, x for line transects and r for point transects, to fit a detection function $g(x)$, defined as 'the probability of detecting an animal that is distance x $(0 \leq x \leq w)$ from the line.' This can be similarly defined as $g(r)$ for point transects where $r(0 \leq r \leq w)$ is the distance from the point. The normal technique is to assume that all animals on the point or line are definitely observed, such that $g(0) = 1$. Several detection functions can be defined by the user and a model selection criteria can be used to select the most appropriate for the data. However, all good detection function models share a set of properties, namely, they should have a shoulder, be non-increasing, be model robust, have pooling robustness and be efficient. A shoulder is when the probability of detection remains close to 1 as distance from the transect increases, before decreasing at a later point. Non-increasing suggests that the probability of detection at a far distance should not exceed the probability at any shorter distance from the transect. Model robustness is necessary as the true function is never know and as such, any models must be flexible to allow a range of different profiles to be modelled. Pooling robustness is a property whereby it is assumed that the model will not be affected if any covariates which influence detection are not included in the model, however these can be included through the use of multi-covariate detection functions, part of Multiple-covariate distance sampling (MCDS). Efficiency informs that should all other factors be equal, a preferable model is on which gives high precision, although high precision should not outweigh the need for the other properties to be satisfied.

Taking the detection function into account, we can then correct our abundance estimates from above to account for those animals not observed. This is done by defining P_a as the

expected proportion of animals observed. In the case of plot sampling above, $P_a = 1$ while for line transect sampling $0 < P_a < 1$. P_a can then be estimated based on the detection function $g(x)$ using an estimate $\hat{g}(x)$. This gives an estimate of P_a, $\hat{P}_a$ for line transects provided by the formula:

$$\hat{P}_a = \frac{\int_0^w \hat{g}(x)dx}{w}$$

While for point transects:

$$\hat{P}_a = \frac{2}{h(0)w^2}$$

where h(0) is the slope of the probability density function for detection distances r , which is proportional to $g(r)r$, the detection function multiplied by r. This leads to abundance estimates from above becoming:

$$\hat{N} = \frac{nA}{2wL\hat{P}_a}$$

for line transects and

$$\hat{N} = \frac{nA}{k\pi w^2 \hat{P}_a}$$

for point transects.

3.2 Distance Sampling Simulations

All the material in this section is based on Buckland et al. (2015) Prior to the simulation for a particular design being run, a number of objects must be first be defined. The first object is the study region, this can either be the default generated by R or user defined from a shapefile. Following this, a spatial distribution or density surface must be defined, from which animal locations can be generated based on the population description. The desired population size can be user defined and set for a series of simulations or be generated based on the spatial distribution supplied by the user. The desired truncation distance must then be defined and based on this an appropriate design can be generated.

The main considerations when constructing the design are the type, either line or point transects and the desired number or length of transects. If line transects are used, the design angle may be altered from its default of 0. A further parameter that must be set it whether plus or minus sampling should be used. Plus sampling is where the transects extend beyond the survey region into a buffer zone of distance w from the edge, however only observations within the survey region are recorded.With minus sampling, transects end at the edge of the study region and do not extend beyond. Minus sampling will be used for all simulations throughout the remainder of this report. This is because in reality, most studies conduct minus sampling due to the additional cost associated with running a plus sampling survey, in addition to the fact that *dsims* does not support plus sampling simulations at the present time. Based on the design, a set of survey transects can be generated, during which the detection process is simulated. The user can then define a detection function, based on either a half normal ('hn'), hazard rate ('hr') or uniform distribution ('uf') with a defined
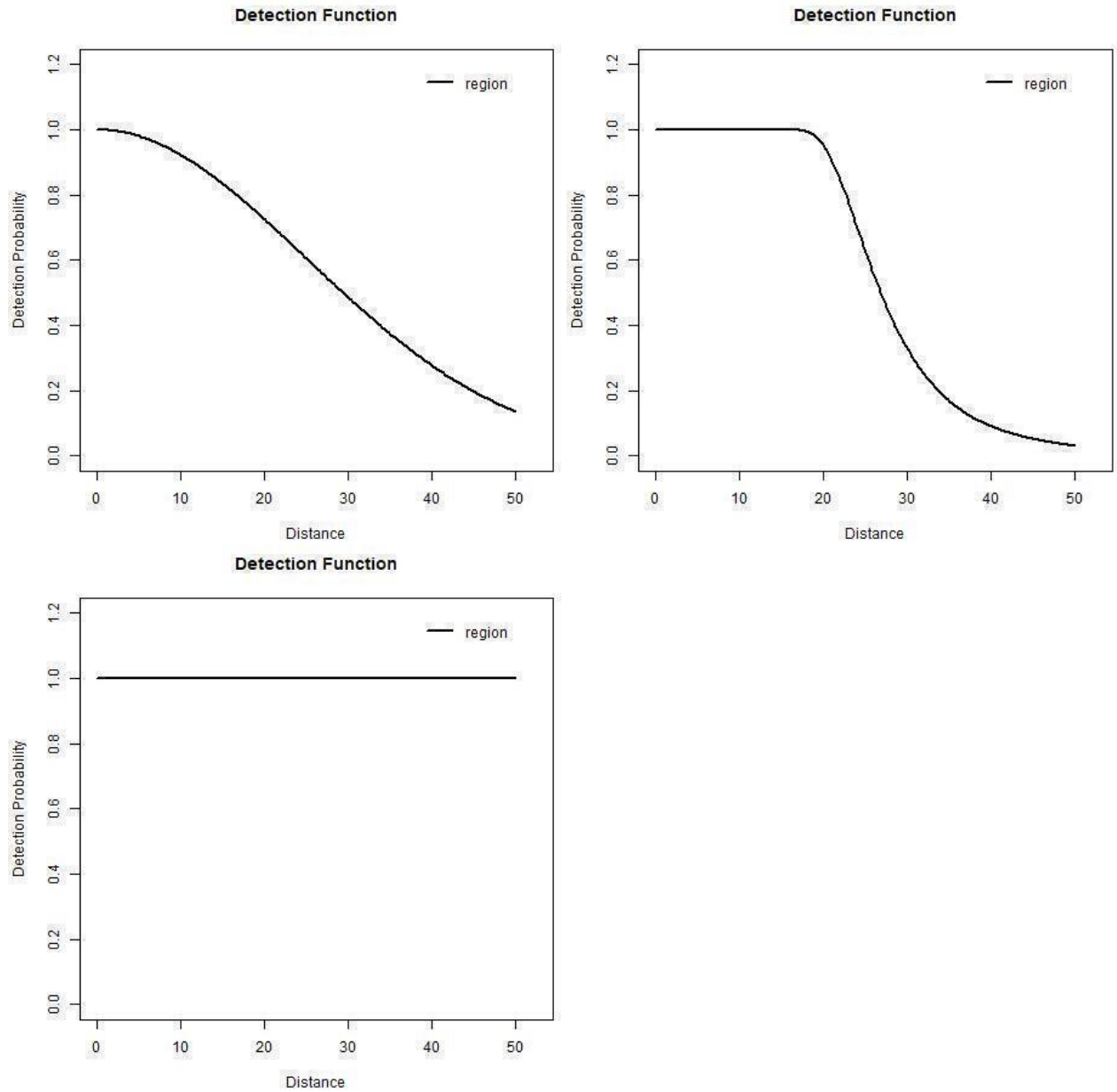
Figure 1: Examples of different detection functions: Top Left: Half Normal, Top Right: Hazard Rate, Bottom Left: Uniform

scale parameter and the desired truncation distance w, examples of which can be seen in Figure 1:

Therefore for an animal at distance x from the closest transect the probability of the animal being detected is given by the detection function evaluated at x, provided x is less than or equal to w. The distance data generated during the survey is then analysed to estimate the abundance N of the study area, with options available for several models to be analysed for each set of distance data, with a model selection criteria used to select the best, with AIC being the default.

At this stage, the encounter rate variance estimator can be specified, with a comprehensive summary of the various estimators found within Fewster et al. (2009). Unfortunately, not all of the estimators mentioned within Fewster et al. (2009) are currently compatible with the survey realisations produced by *dsims*, with those not compatible being 'R4,' 'S1,' 'S2' and 'O1.' The default estimator is 'R2' for line transects and 'P3' for point transects and unless otherwise stated, these were used within all simulations conducted in this report. Fewster et al. (2009) does however recommend the 'O2' estimator when using systematic parallel line transects, which will be investigated later in this report, however this estimator is only currently implemented for systematic parallel line designs. If we look at how these estimators are constructed, it is possible to identify where one might be an improvement over the other. To begin assessing the variance, we revisit the density estimate as mention above, namely:

$$\hat{D} = \frac{n}{\hat{P}_a} \times \frac{1}{2wL} = \frac{1}{2w} \times \frac{n}{L} \times \frac{1}{\hat{P}_a}$$

Following on from this, the delta method is used to estimate the variance in the density, noting that $1/2w$ is a constant:

$$[cv(\hat{D})]^2 = \left[cv\left(\frac{n}{L}\right)\right]^2 + \left[cv\left(\frac{1}{\hat{P}_a}\right)\right]^2$$

Fewster et al. (2009) notes that the encounter rate component of the variance, var(n/L) usually comprises around 75% of the total variance, while an estimate of var($\hat{P}_a$) can be obtained using standard likelihood theory. Variance estimators are required for n/L, as inference based upon a probability distribution is not likely to be sufficiently robust when dealing with inhomogeneous biological populations. This leads to estimators being used which are constructed around the empirical variability in each transects encounter rate, n_i/l_i. We first look at the construction of the 'R2' estimator. This is a design based estimator for random designs which arises as a result of a Taylor series expansion n/L about ρ, where $\rho = E(\bar{n})/E(\bar{l})$ where $\bar{l}$ denotes L/k. This leads to the estimator becoming:

$$\widehat{var}_{R2}\left(\frac{n}{L}\right) = \frac{k}{L^2} \times \frac{1}{k-1} \sum_{i=1}^{k} (n_i - \hat{\rho}l_i)^2$$

$$= \frac{k}{L^2(k-1)} \sum_{i=1}^{k} l_i^2 \left(\frac{n_i}{l_i} - \frac{n}{L}\right)^2$$

which Fewster et al. (2009) notes is biased within the design framework. When using systematic designs, these often exhibit greater precision than those of random designs, leading to additional information which could be included within the variance estimate. However, this information is not taken advantage of if the estimator assumes a random line placement, as in 'R2,' leading to the likely overestimation of the variance. To account for this increased precision, we can make use of the systematic nature of the design by grouping neighbouring transects in a post-stratification technique and then calculate the individual variances of each strata, before combining to give the overall estimate. To begin explaining this, we look at the 'S2' estimator, while noting this is not currently working within *dsims*. This if formed by splitting the transects into strata, where each transect can only be in one strata, with the strata labels being $h = 1, \ldots,$ H, and letting there be k_h lines within stratum h. This implies:

$$var(\frac{n}{L}) \approx \frac{1}{L^2} \sum_{h=1}^{H} k_h \text{var}_h(n_{hj} - \rho l_{hj})$$

where n_{hj} and l_{hj} are individual random variables n_i and l_i from within stratum h. The 'S2' estimator is then formed by combining the individual stratum estimates in an heuristic manner and adding a weighting factor to each, with this being the total line length for each stratum.

$$\widehat{var}_{S2}(\frac{n}{L}) = \frac{1}{L^2} \sum_{h=1}^{H} L_h^2 \widehat{var}_h(\frac{n_h}{L_h})$$

where n_h and L_h are the observation and line length totals for each strata, with $\widehat{var}_h$ being the within strata variance estimator.

As noted, this is not currently available within the *dsims* package, however there is an analogous estimator which is available, 'O2.' This is constructed in a very similar way to the 'S2' estimator, however it makes uses of the systematic nature of the designs by having overlapping strata where the strata have a natural ordering, often by spatial proximity. In this case, the first strata contains lines 1 and 2, the second lines 2 and 3 and so on. This leads to a total of k-1 strata each with two lines in each and is described as an overlapping strata post-stratification scheme. The individual variances of the strata are then combined to give k-1 variance estimates, with Fewster et al. (2009) noting that while the degrees of freedom are not known, the approximation of k-1 is normally assumed. Due to the systematic construction of the strata, it is known at $k_h = 2$ and strata h contains lines h and *h+1*, leading to the new variance estimator becoming:

$$\widehat{var}_{O2}(\frac{n}{L}) = \frac{2k}{L^2(k-1)} \sum_{i=1}^{k-1} \frac{(l_i l_{i+1})^2}{(l_i + l_{i+1})^2} \left(\frac{n_i}{l_i} - \frac{n_{i+1}}{l_{i+1}} \right)^2$$

Having selected the most appropriate or available variance estimator, all of the simulation elements are in place and the simulation can be run. This is done by generating a realisation of the survey based on the design, population and detection function specified before the analysis above is conducted, These operations are then repeated the specified number of

times, say R, to obtain a set of simulations of animal distribution and survey design, alongside a corresponding set of estimates $\hat{N}$ of N. Typical values for R are between 100 and 1000. The estimates along with other information are then extracted from the survey data allowing mean estimates and confidence intervals alongside other statistics to be calculated.

3.3 Density Surface Modelling

This sections contains material based on Miller et al. (2013). In order to construct a density surface model, initial the approach must be decided upon. The choice is between using a two stage approach, whereby the detection function is fitted first then subsequently fitting a spatial model, while the one stage approach leads to estimating the detection and spatial parameters simultaneously. Miller et al. (2013) states that 'Generally, very little information is lost by taking the two stage approach' as transect width is comparably smaller than that of the study region, therefore, provide the population does not differ spatially within the transect, no information is lost by the two stage approach. This may lead is issues occurring where the density of the species has significant variability at the transect level. However, one drawback of the two stage model is that, to accurately evaluate the model uncertainty, the uncertainty in both the detection function and the spatial models should be suitably combined. For the remainder of this report only the two stage approach will be discussed. Initially, the detection function must be fitted, with the specification being the same as mentioned in the distance sampling section above. Following this, the density surface model can be fitted. To enable this to occur, the data must be separated into segments. This is easily done for point transects with each point being a segment however it more complicated for line transects. With line transects, they must be split up into J segments of length l_j. It is normally from the segments to be approximately square, with dimensions of *2_w_x2_w* where w is the truncation distance of the design. From here, the segment areas enter the model as part of an offset, to allow for non-constant segment areas. This leads the line transect segments to have an area of $2wl_j$ and the point transect segments with an area of πw^2. In the model, the counts or abundances are using a generalised additive model (GAM) using the sum of the smoothed covariates.

3.3.1 Response models

The model used when the count per segment is used as the response is:

$$\mathbb{E}(n_j) = \hat{p}_j A_j exp[\,\beta_0 + \sum_k f_k(z_{jk})]$$

Where f_k are the smoothed functions of the covariates and β_0 is the intercept term. By multiplying the segment area A_j by the estimated probability of detection p_j this gives the effective area of the segment, acting as an offset to account for different segment areas. Where distance is the only covariate in the detection function, p_j is constant across all segments and therefore $\hat{p}_j = \hat{p}\forall j$. The distribution of n_j can then be modeled using an overdispersed Poisson, Negative binomial or Tweedie distribution.

An alternative to using this is to use abundance estimates for each segment generated by distance sampling as the response. To do this, the response n_j is replaced by an estimator of the abundance in each section, $\hat{N}_j$ where this is defined as:

$$\hat{N}_j = \sum_{r=1}^{R_j} \frac{s_{jr}}{\hat{p}_j}$$

Where R_j is the number of observations in the jth segment and s_{jr} is the size of the rth group observed, with this being 1 if only individuals are observed. As identified by Buckland et al. (2015), this is an Horvitz–Thompson-like estimator of the segment abundance, allowing for covariates to be included through $\hat{p}_j$. The fitted model then becomes:

$$\mathbb{E}(\hat{N}_j) = A_j exp[\,\beta_0 + \sum_k f_k(z_{jk})]$$

Where the model follows the same three distributions as before. The main difference between these models is that the offset is now the physical area of each segment, as opposed to the effective area in the first model for n_j.

To allow for a DSM to predict abundance, a series of prediction cells must be defined. These are not necessarily restricted to just the original study region, allowing for regions outside the study area to be predicted over. Each of the prediction cells must include the same covariates as specified in the DSM, including the area of each cell. Predictions can then be made for the abundance in each cell and by summing these over the whole region, an overall abundance estimate can be obtained. The size of the prediction cells may be specified by the user, however cells 'smaller than the resolution of the spatially referenced data' do not have an influence on the abundance estimates produced by DSM.

4 Simulation extentions

Within the *dsims* package Marshall (2021), the current simulation is set-up as explained within the the Distance Sampling Simulations section above. To allow for the inclusion of Density surface modelling within the simulation, a number of changes and additions were required to be made to ensure the compatibility of the two approachs within the simulation. This was a major focus of this project and was where a large portion of time was spent testing to ensure the generality of the code for future use. The initial setup required is the same as with the existing simulation. Namely, the study region, density surface, population and detection function are required. In addition to this, the design in question should also be specified, with any of the choices provided by *dsims* available.

Prior to the simulation beginning, the prediction grid required by the DSM is constructed across the study region, with the resolution of the grid set at half the truncation distance of the design. As Miller et al. (2013) noted, smaller cells sizes could be used but there is a limit since using cells smaller than the spatial data resolution will not have an effect on the abundance estimates provided by DSM, and there is also the computational increase resulting from smaller cell sizes. This can be specified prior to the simulation as it is only dependant on the study region and will not change between iterations of the simulation. Existing code to construct this prediction grid can be found in Souchay (2020) and this was used in early testing and simulations. However, it was noted that some of the R packages used in this code are due to be retired at the end of 2023, namely the *rgeos* package Bivand and Rundel (2021) . Therefore, the decision was made to rewrite this code using the far more widely used Simple Feature (sf) package Pebesma (2022). The prediction grid is currently setup to create the grid over the entire study region, to allow for a comparison with the distance sampling approach. However, this is not a restriction and any prediction grid could be specified, including areas outside of the study region. This would allow for simulations to predict the abundance on areas not specifically studied in the survey, which the distance sampling approach cannot do without increasing the size of the study region itself. This offers a distinct advantage for DSM over the distance sampling approach, however it may be possible to generate unrealistic abundance estimates if the gradient of the density surface is steep near the boundary. This effect will be tested later in this report.

The simulation loop then begin, running for the number of iterations specified within the simulation object. For each iteration, a new survey constructed. From this, the observation data and segmented data is extracted and linked. To do this, a function *generate.dsm.data* was written, taking the study region, survey and type of transects as inputs. This function first extracts the observation data from the survey. Subsequently, the transects are split into the segments required by the DSM approach, using a function *to_segments* which was written for this explicit purpose, taking the same inputs as *generate.dsm.data*.

For Point transect designs, each point is treated as its own segment however, in the case of line transect designs, the lines are split to allow them to be modelled as points. Each line is split into segments of approximate length $2w$ with w being the truncation distance, as suggested by Miller et al. (2013) and each segment assigned its own unique sample label. Subsequently, the length of each segment is recorded as the segments effort value and

the centre if the segment as its location. Polygons of each segment is then created using *st_buffer*, with w being used as the distance. This leads to squares of approximately $2w * 2w$ for line transects and circles of radius w for points. These allow the area of each segment to be calculated, a requirement for the DSM model. This is calculated using *st_area*, on the intersection between the polygons and the outer boundary of the study region. This ensures only areas within the study region are counted towards segment area and internal strata boundary's do not split segments. Failure to do this results in the areas of some segments being larger than they are in the survey, and as a result the DSM abundance estimate is smaller, since the prediction grid is only over the survey area.

Once the segment areas have been calculated, the segments can be linked to the observation data by allocating each observation to the nearest segment and giving this the respective segments sample label in the observation data, overwriting the original allocation. Additionally, the coordinate reference system of the observation data is set as that of the segments to ensure consistency when using different reference systems to the default.

Based on this data, both a distance sampling model and density surface model are constructed, with the DSM modeling the counts against the smooth of spatial locations using a tweedie error distribution. The formulas for both could be changed to include environmental or other covariates such as strata, allowing for different detection functions to be fit to each strata. In the smoothed term, the degrees of freedom is restricted to the total number of transects. The limit for this is the number of segments in the model, which in the case of point transects is the total number of transects. While this could be far larger for line transects, this would increase the computational requirements beyond a reasonable level and not offer a significant improvement for this increase. The abundance estimates are extracted from both models and stored alongside other values from each including the standard errors and confidence intervals for each model.

5 Modelling - Default Region

The simulation was initially tested using the default region generated by *dsims* Marshall (2021) with a truncation distance of 60. Both point and line transect designs were run with an aim of 25 and 12 samplers for the respective designs. A basic test density was then constructed for the region with high and low spots as seen below in Figure 2 with relatively gently gradients.
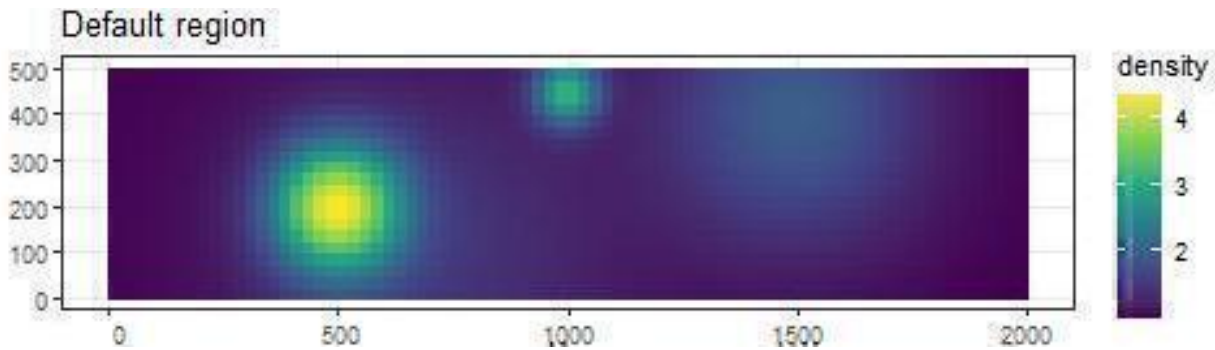

Figure 2: Density surface used for the Default region

A population description was then constructed based on this density surface with a true population of 1000. A detection function was then defined as a half normal with scale parameter of 30 for both designs, producing the detection function seen in Figure 3
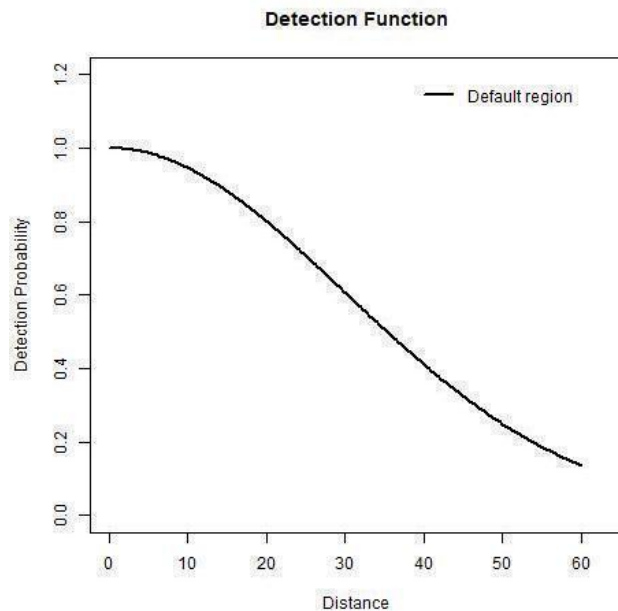

Figure 3: Detection function for the default region, a half normal with scale parameter of 30

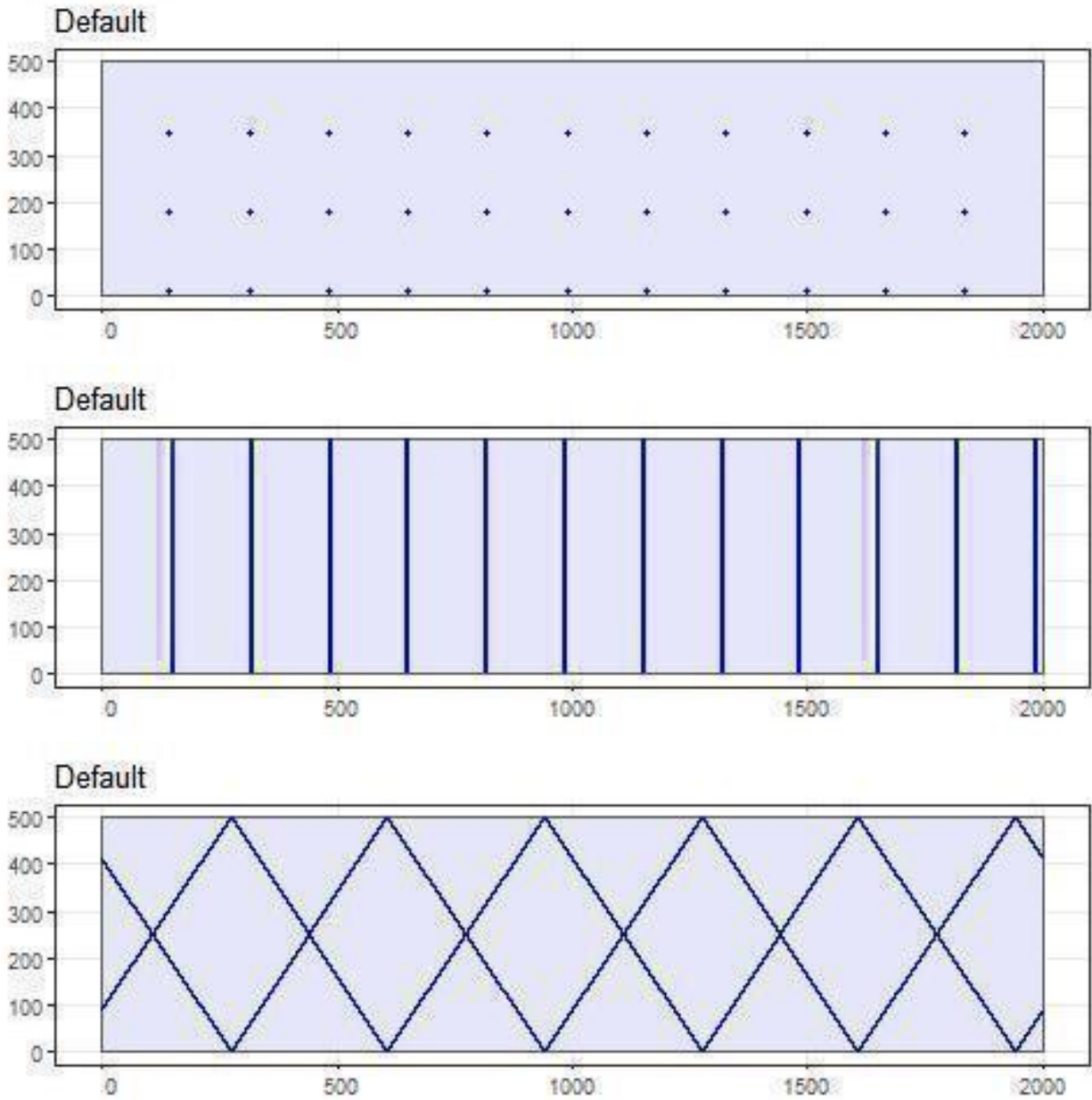An example survey for each of the designs is displayed below in Figure 4.

Figure 4: Example Surveys for the default region. Top: Point Transect design, Middle: Parallel line design, Bottom: Zigzag line design

5.1 Results

Having completed 1000 simulations for both the distance sampling and density surface models with each design, we can now examine and compare these to give us an insight into the circumstances under which a particular model is better of worse than the other.

5.1.1 Default Region Point design

For the initial default region with the point transect design, the histograms of both the distance sampling and DSM abundance estimates are displayed below in Figure 5:
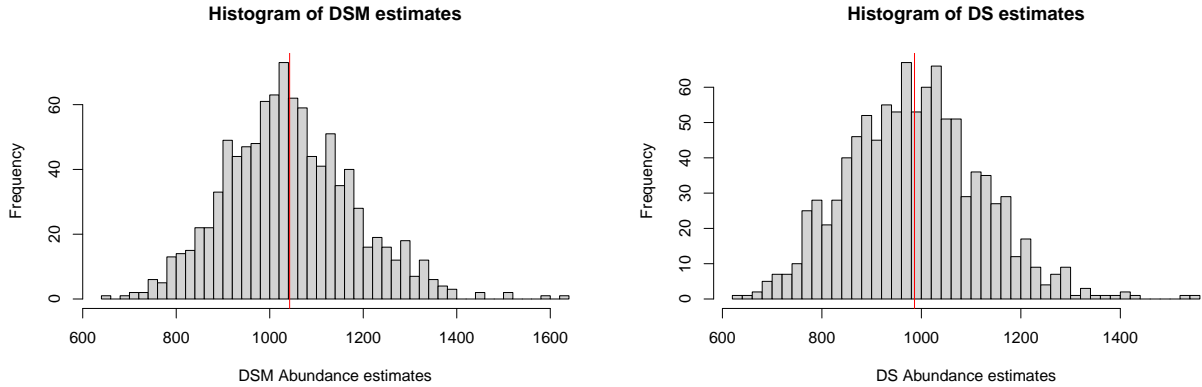

Figure 5: Histograms of Point Design estimates for Default region

These plots show somewhat similar data since both estimates are generated by the same data set. If we now compare the means of the two approaches, the mean of DSM estimates was 1042.55 and the mean of DS estimates was 986.21. It can be seen that the mean of the DSM estimates appears a good distance from the true population of 1000 while the distance sampling method is closer.

Table 1: Default region results - Point

Statistic	Point Design	
	DSM results	DS results
Mean	1050.3465198	989.0635222
Bias	5.0346520	-1.0936478
Mean Std error of estimates	160.7477256	152.7249734
Std Dev of estimates	167.6863061	156.2742970
Mean CV	0.1545738	0.1556531
Confidence Interval Coverage	0.9210000	0.9480000

Table 1 above provides statistics which allow us to compare the variability of the models as well as their ability to capture truth within their confidence intervals. We see in this

case that ths DSM method has a slightly larger positve bias than that of the DS method, indicating the DS method produced more accurate results in this case. This trend continues with the Mean Standard errors and estimate standard deviations being similar but again the DS method producing a smaller error. The DSM does produce a smaller Coefficient of variation, however the difference is marginal. Next, we can evaluate how many times the true abundance was within the 95% CI for every model computed. The results show that truth was captured by the CI with a probability of above 0.9 for both models, with the DS again performing better at 0.948 compared with 0.921 for the DSM method.

5.1.2   Default region parallel Line design

Now examining the results of the line transect design, on the same density surface, we see the histograms of the two estimates below in Figure 6:
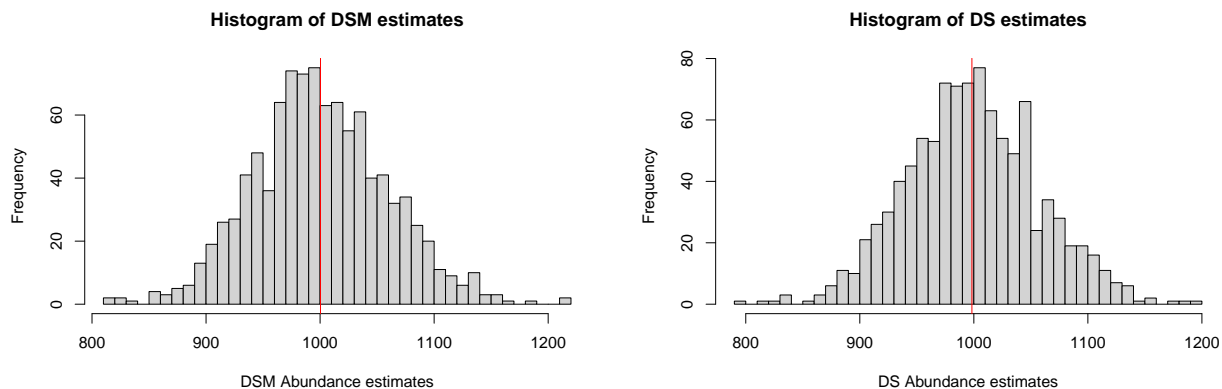

Figure 6: Histograms of Line Design estimates for Default region

These plots are very similar to their point transect counterparts, with the mean for each signified by the red line appearing very close to truth. Investigating this further, we find the mean of DSM estimates to be 1000.35 and the mean of DS estimates to be 998.31. These estimates are both very close to the true value, possibly as a result a large portion of the survey area being covered by the transects.

From table 2 above, it can be observed that in this case, the DSM method overall performed better than the DS method, having a slightly smaller bias as well as a dramatically smaller mean standard error and mean coefficient of variation. However, for the DS method, the standard deviation of the estimates is markedly less than the mean standard error, indicating this method provides better precision than the standard errors suggest. This is also the case with the DSM method however the difference is far smaller. The DS method also has a higher probability of truth being contained within its confidence intervals, at 0.999 compared to 0.972.

Table 2: Default region results - Line

Statistic	Parallel Line Design	
	DSM results	DS results
Mean	1000.3464022	998.3144253
Bias	0.0346402	-0.1685575
Mean Std error of estimates	68.9012076	106.1785893
Std Dev of estimates	58.8779025	57.5569613
Mean CV	0.0689178	0.1064057
Confidence Interval Coverage	0.9720000	0.9990000

5.1.3 Default region Zigzag Line design

Now examining the results of the zigzag line transect design, on the same density surface, we see the histograms of the two estimates below in Figure 7 :
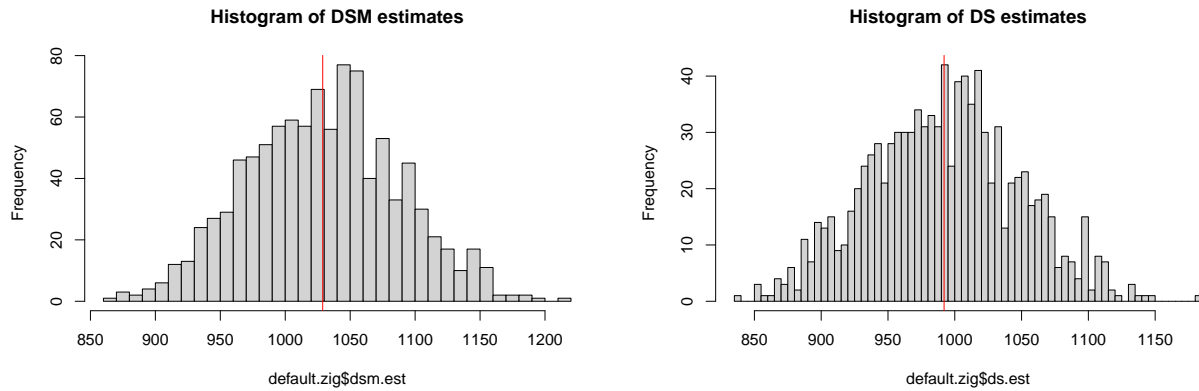

Figure 7: Histograms of Zigzag Design estimates for Default region

These plots differ slightly to their point and parallel line counterparts, with the mean appearing further away from truth than previously, however there is a known error in the code that may cause this and is being investigated. Investigating this further, we find the means of DSM estimates to be 1028.85 and the mean of DS estimates to be 991.99. These estimates are both very close to the true value, possibly as a result a large portion of the survey area being covered by the transects.

Table 3: Default region results - Zigzag

Statistic	Zigzag Line Design	
	DSM results	DS results
Mean	1028.8483718	991.9923373
Bias	2.8848372	-0.8007663
Mean Std error of estimates	70.4895927	90.2284558
Std Dev of estimates	58.3785323	55.9582394
Mean CV	0.0685151	0.0907944
Confidence Interval Coverage	0.9680000	0.9970000

From Table 3, despite the known issue, the DSM model performs relatively well, having a reasonably small bias. The mean standard error is in fact lower than that of the DS model, indicating higher precision on average across the models, with the standard deviation of the estimates being lower in both cases, suggesting greater precision for both modes than the standard errors imply. The mean coefficient of variation is slightly lower for the DSM model however the confidence interval coverage is higher with the DS model, indicating truth has a higher probability of being captured within the models confidence interval.

6   Modeling - North Sea

Having gone through the process and results for the default region produced by *dsims*, a more complex example was created to further test the capabilities of the simulations with a real world example. This region with be in the area of the North sea off the east coast of the UK. As this would likely be a shipborne or airborne study, point transects are not realistic in this scenario, however they will be included for completeness. To assess the performance of both the distance sampling and density surface modelling approaches, point and line transect designs will be examined. These will be two point transect designs and four line transect designs, with one of each type using a stratified design with the region split in two. The standard designs will be compared first before examining the stratified designs. All designs used a truncation distance of 10 with the number of transects varying between the point and line transect designs. This difference is ensure the covered area of each study is similar to allow the results to be compared, as well as ensure sufficient observations to fit the detection function can be obtained. The density maps for stratified and non-stratified will be different to justify the use of different strata and examine the models ability to pick this up. These density surfaces can be seen in Figure 8 below.
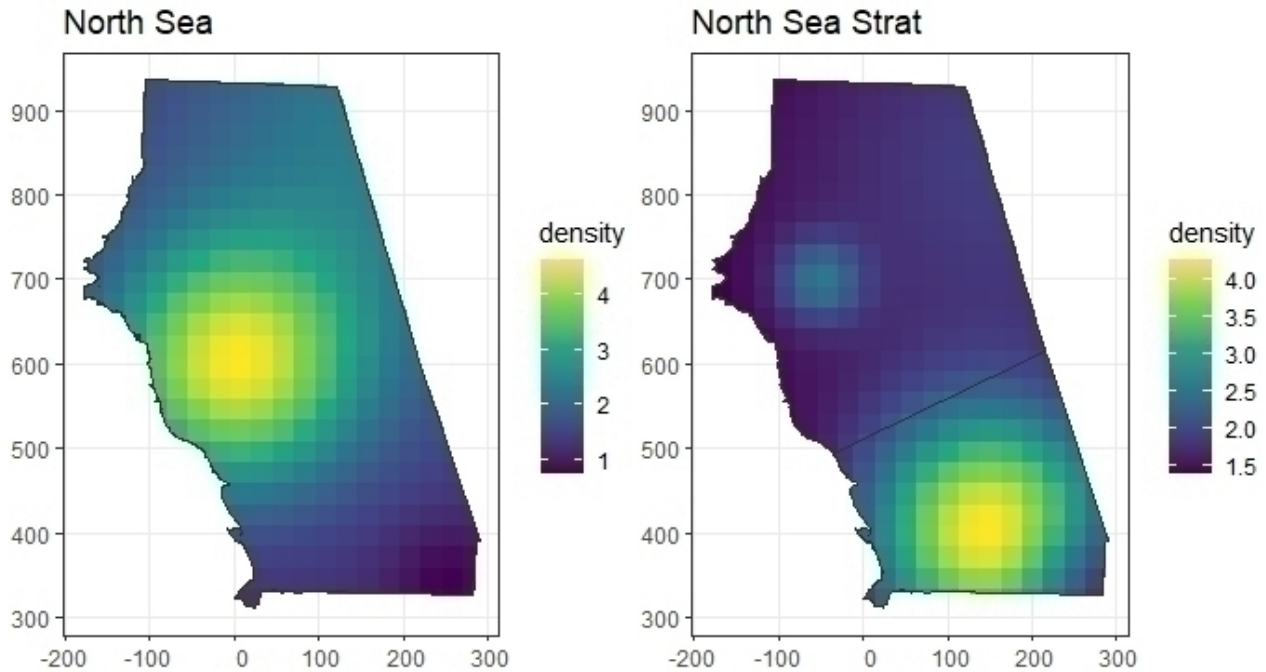

Figure 8: Density surfaces for the basic and stratified North Sea regions

As before, the detection function was kept identical between each design for both stratified and non-stratified. In this case it was a Half Normal detection function with a scale parameter of 5, which can be seen in Figure 9 below:
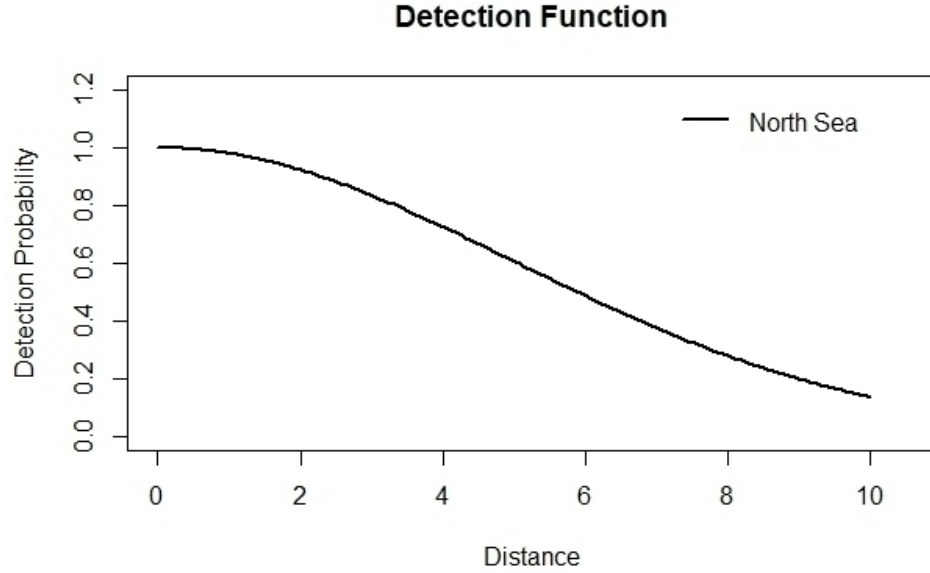
Figure 9: Plot of the detection function used in both the basic and stratified North Sea regions: A half Normal with scale parameter 5 and truncation 10

6.1 Non-Stratified Designs

The point transect designs aimed for 70 samplers and the line transects for 25 samplers. The standard line designs consisted of a systematic parallel line design with transects perpendicular to the coast and a zigzag line design originating in the North of the region. An example of each design can be seen in Figure 10 below:

The simulation was run 5000 times with the DSM model using a tweedie error distribution and the only covariates being spatial coordinates.
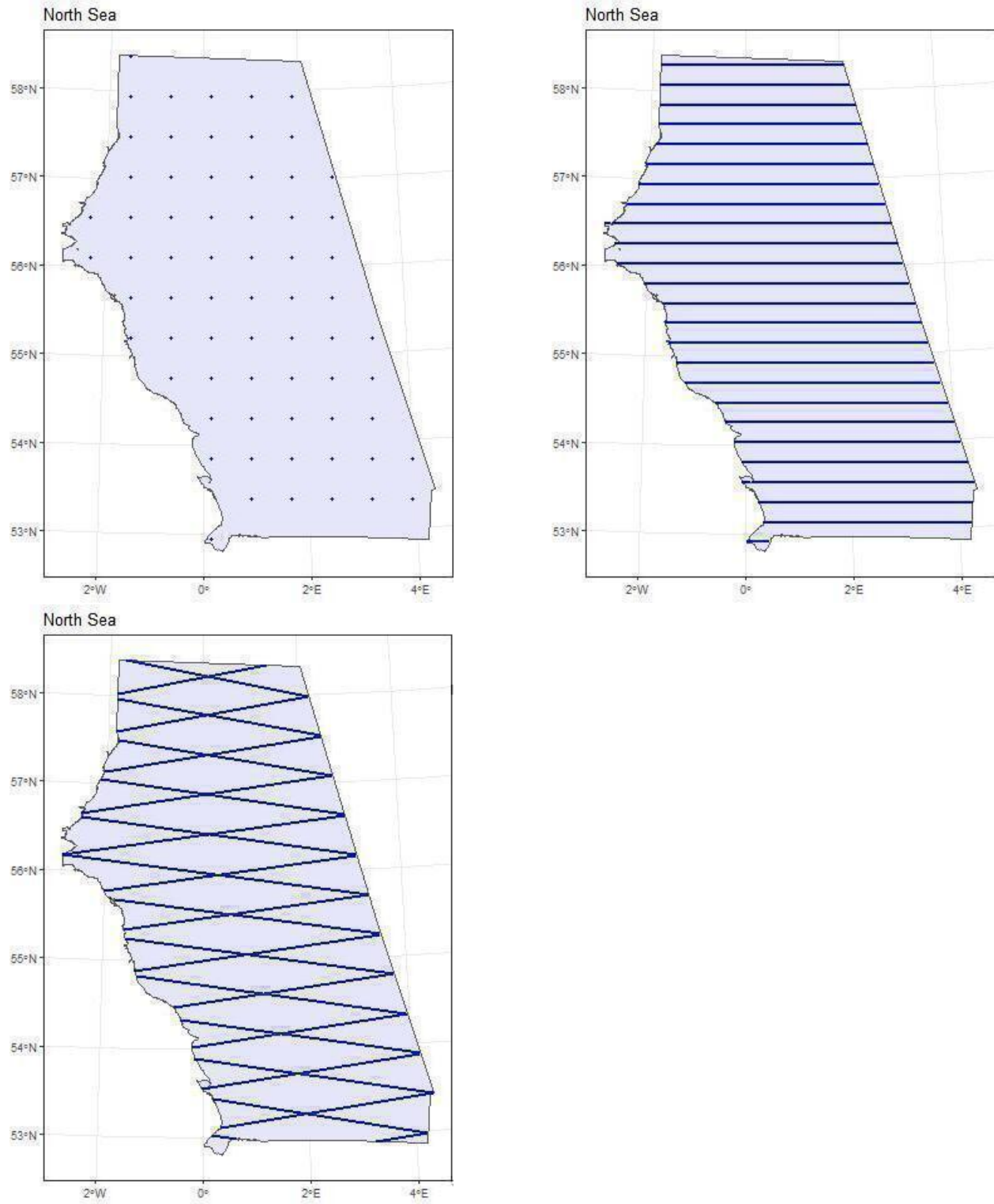
Figure 10: Examples of surveys for the North Sea region. Top Left: Point Transect design, Top Right: Parallel line design, Bottom Left: Zigzag line design

6.1.1   Non Stratifed Results - Point Design

For the point transect design, the histograms of the abundance estimates from each simulation run are displayed below in Figure 11:
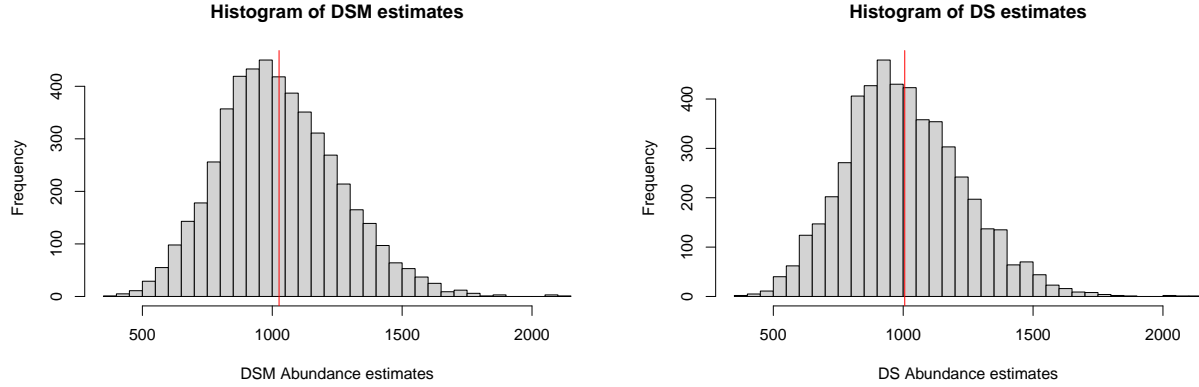

Figure 11: Histograms of the abundance estimates for the North Sea region using a point transect design

These plots suggest both methods are relatively good at estimating the true abundance of 1000, with both methods appearing to overestimate, with the DSM method overestimating slightly more than the DS method. There also appears to be a slight right skewness to both plots which will be looked at analytically. Investigating the means further, we find the mean of DSM estimates to be 1026.21 and the mean of DS estimates to be 1005.97. These estimates are both close to the true abundance, with the overestimation present as mentioned above. Overall, it was found that the estimates produced by DSM were greater than those produced by DS from the same survey data 92.8 % of the time, indicating it is highly likely DSM will produce an estimate greater than that of DS. A Wilcoxon test was performed on both sets of estimates with the alternative being that DSM was greater than DS.

This returned a p-values of 0 indicating at any significance level, the null hypothesis of the estimates being zero is rejected in favour of the alternate, that the DSM estimates are greater than those of DS. If we now compare the other variables extracted from the models in the below, we can examine the accuracy and variability of each method:

From the results table above, it can be observed that in this case, the DS method overall performed better than the DSM method, having a smaller bias as well as higher confidence interval converge, despite having a smaller Mean Std error. Additionally, the standard deviation of the estimates is slightly less for the DS method, however both methods do a relatively good job at estimating this. Additionally, we observe from the skewness statistics that both methods returned a positive statistic indicating the right-skewed as identified above, which suggests the right tail is longer and this is supported by some of the estimates being over 2000, with the maximums for DSM and DS being 2113.2 and 2112.52 respectively.

Table 4: North Sea Non-Stratified results - Point

Statistic	Point Design	
	DSM results	DS results
Mean	1026.2081430	1005.9650684
Bias	2.6208143	0.5965068
Mean Std error of estimates	234.5771639	234.0227893
Std Dev of estimates	230.8981167	226.7679405
Mean CV	0.2321996	0.2362012
Confidence Interval Coverage	0.9484000	0.9582000
Estimate Skewness	0.4034455	0.4097919
Estimate Kurtosis	3.2729606	3.3024713

The kurtosis figures for both methods are greater than 3, indicating we would expect more frequent outliers than a normal distribution While the Mean CV for the DSM method is lower, the difference is marginal and does not out-way the bias or variance issues of this method. Overall this clearly seems to be a case where the DS method performs better than the DSM estimation method.

6.1.2 Non Stratifed Results - Parallel Line Design

Now examining the results of the line transect design, on the same density surface, we see the histograms of the two estimates below:
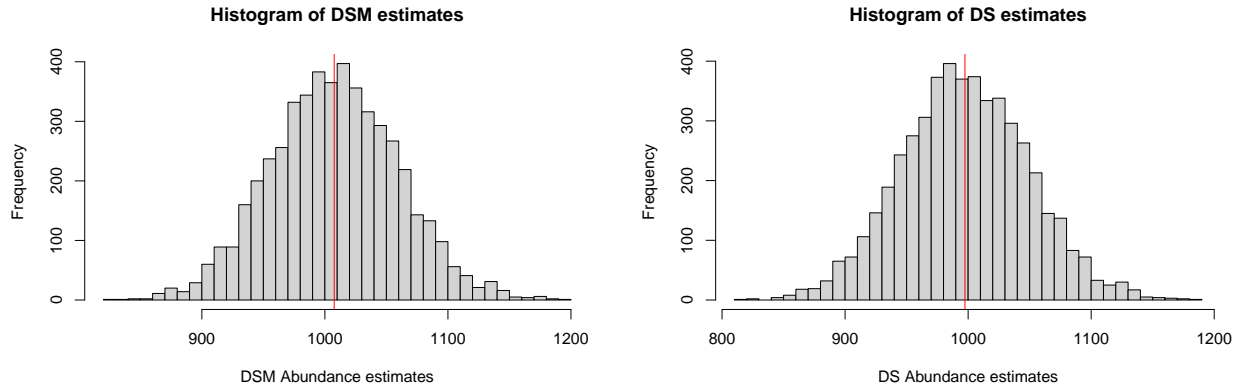
These plots are very similar to their point transect counterparts, with the mean for each signified by the red line appearing very close to truth.However, there appears to be less skewness for these than is present with the point transect results. Investigating this further, we find the mean of DSM estimates to be 1007.57 and the mean of DS estimates to be 997.41. These estimates are very close to the true Overall, it was found that the estimates produced by DSM were greater than those produced by DS from the same survey data 100 % of the time, indicating it is highly likely DSM will produce an estimate greater than that of DS. A Wilcoxon test was performed on both sets of estimates with the alternative being that DSM

was greater than DS.

This returned a p-values of 0 indicating at any significance level, the null hypothesis of the estimates being zero is rejected in favour of the alternate, that the DSM estimates are greater than those of DS. Examining the other statistics extracted from the simulation.

Table 5: North Sea Non-Stratified results - Line

Statistic	Parallel Line Design	
	DSM results	DS results
Mean	1007.5691041	997.4080181
Bias	0.7569104	-0.2591982
Mean Std error of estimates	61.0407009	86.7850841
Std Dev of estimates	51.8726020	51.7437163
Mean CV	0.0606351	0.0870698
Confidence Interval Coverage	0.9748000	0.9996000
Estimate Skewness	0.0607668	0.0591685
Estimate Kurtosis	3.0058358	2.9856027

From the results table above, it can be seen that in this case, the performance of both models was similar and each performed well in different areas compared to the each other. The mean estimate was again a slight overestimation and as a result, bias was slightly higher for DSM, however this difference was comparatively small. The Mean Std errors were lower for DSM and far closer to the Std Dev of the estimates, highlighting the better error estimation of DSM. However the Std Dev of the estimates was slightly smaller for the DS approach. The DS method also has a higher probability of truth being contained within its confidence intervals, at 0.9996 compared to 0.9748, however this difference is likely due to the higher Std errors from the DS method wideing the confidence interval. Despite this, both methods achieved truth in their confidence intervals over the 95% significance level. The skewness scores for both are very similar and suggest again a slight right skew however this is very small. The kurtosis scores are very close to 3, suggesting the kurtosis is very similar to that of a normal distribution. In the Jarque-Bera test for normality, DSM and DS estimates return p-values of 0.2139247 and 0.2275685 respectively. These indicate we can not reject the null hypothesis that these estimates are normally distributed. Overall, while the mean estimate appears in favour of the DS method, the DSM approach can be said to have performed better in this simulation, due to the small bias, lower Mean Std error and coefficients of variation.

6.1.3 Non Stratifed Results - Zigzag Line Design

Now examining the results of the zigzag line design, on the same density surface, we see the histograms of the two estimates below:
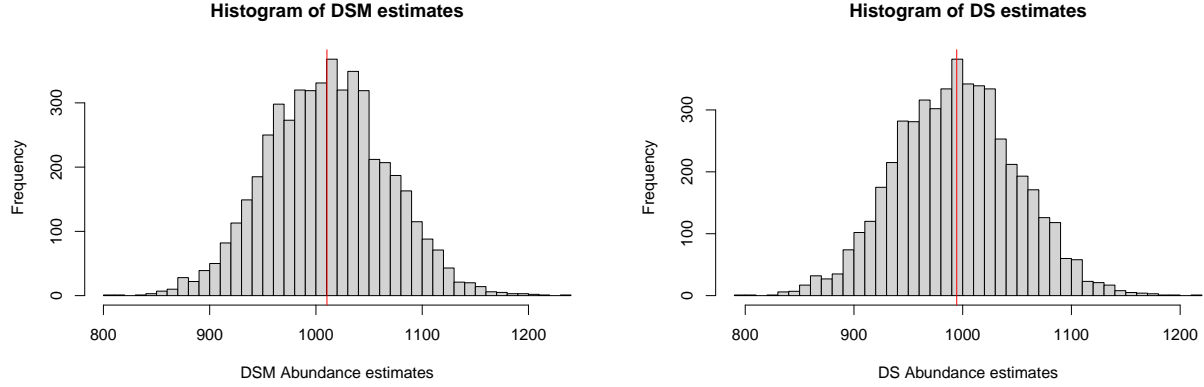
These plots are very similar to their point transect counterparts, with the mean for each signified by the red line appearing very close to truth.However, there appears to be less skewness for these than is present with the point transect results. Investigating this further, we find the mean of DSM estimates to be 1010.34 and the mean of DS estimates to be 994.51. These estimates are very close to the true Overall, it was found that the estimates produced by DSM were greater than those produced by DS from the same survey data 99.96 % of the time, indicating it is highly likely DSM will produce an estimate greater than that of DS. A Wilcoxon test was performed on both sets of estimates with the alternative being that DSM was greater than DS.

This returned a p-values of 0 indicating at any significance level, the null hypothesis of the estimates being zero is rejected in favour of the alternate, that the DSM estimates are greater than those of DS. Examining the other statistics extracted from the simulation.

Table 6: North Sea Non-Stratified results - Zigzag

Statistic	Parallel Line Design	
	DSM results	DS results
Mean	1010.3404728	994.5099646
Bias	1.0340473	-0.5490035
Mean Std error of estimates	61.1878286	81.8307222
Std Dev of estimates	55.9827256	55.3082445
Mean CV	0.0606281	0.0823353
Confidence Interval Coverage	0.9660000	0.9950000
Estimate Skewness	0.0640797	0.0614198
Estimate Kurtosis	2.9645959	2.9557928

From the results table above, it can be seen that in this case, the performance of both models was similar and each performed well in different areas compared to the each other. The mean estimate was again a slight overestimation and as a result, bias was slightly higher for DSM, however this difference was comparatively small. The Mean Std errors were lower for DSM and far closer to the Std Dev of the estimates, highlighting the better error estimation of DSM. However the Std Dev of the estimates was slightly smaller for the DS

approach. The DS method also has a higher probability of truth being contained within its confidence intervals, at 0.995 compared to 0.966, however this difference is likely due to the higher Std errors from the DS method wideing the confidence interval. Despite this, both methods achieved truth in their confidence intervals over the 95% significance level. The skewness scores for both are very similar and suggest again a slight right skew however this is very small. The kurtosis scores are very close to 3, suggesting the kurtosis is very similar to that of a normal distribution. In the Jarque-Bera test for normality, DSM and DS estimates return p-values of 0.1585816 and 0.1694159 respectively. These indicate we can not reject the null hypothesis that these estimates are normally distributed. Overall, while the mean estimate appears in favour of the DS method, the DSM approach can be said to have performed better in this simulation, due to the small bias, lower Mean Std error and coefficients of variation.

6.2 Stratified Designs

For the stratified designs, it was decided to split the region into two strata, a large northern strata 'North' and a smaller southern strata 'South' to test how the simulation and both methods cope with a stratified design, alongside ensuring th . The point transect designs aimed for 60 samplers, with 30 in each strata, and the line transects for 25 samplers with 10 in the North strata and 15 in the South strata. The standard line designs consisted of a systematic parallel line design with transects perpendicular to the coast for both the North and South Stratum. The zigzag line design selected has the transects in the North strata originating in the North of the Strata, with a design angle of $90°$, while in the South Strata, transects originate on the Western boundary with a design angle of $0°$. An example of each design, including the point design, can be seen in Figure 12 below:

The simulation was run 5000 times with the DSM model using a tweedie error distribution and the only covariates being spatial coordinates.
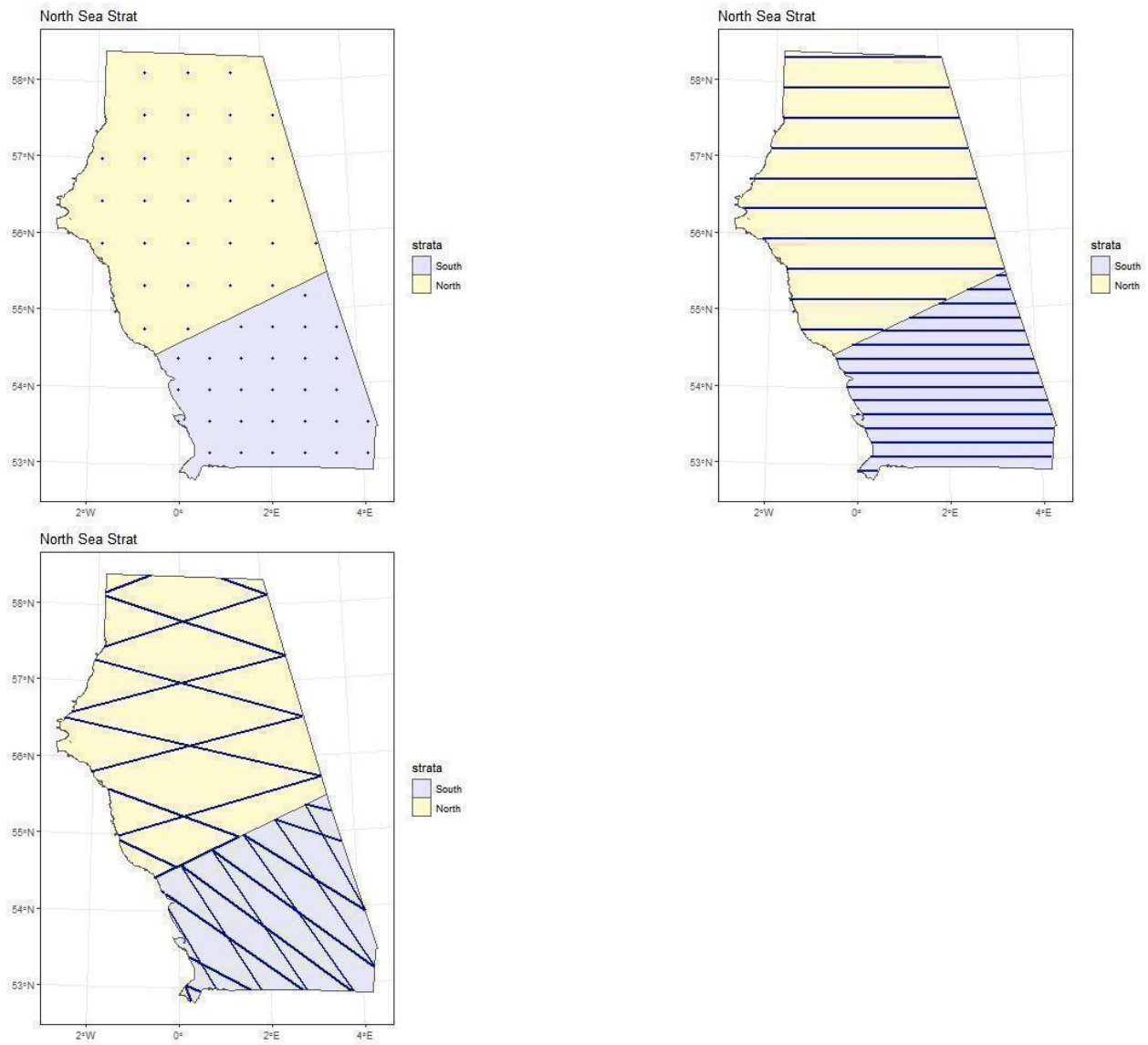
Figure 12: Example surveys for each Stratified Designs. Top Left: Point Transect design, Top Right: Parallel line design, Bottom Left: Zigzag line design

6.2.1 Stratifed Results - Point Design

For the point transect design, the histograms of the abundance estimates from each simulation run are displayed below in Figure 13:
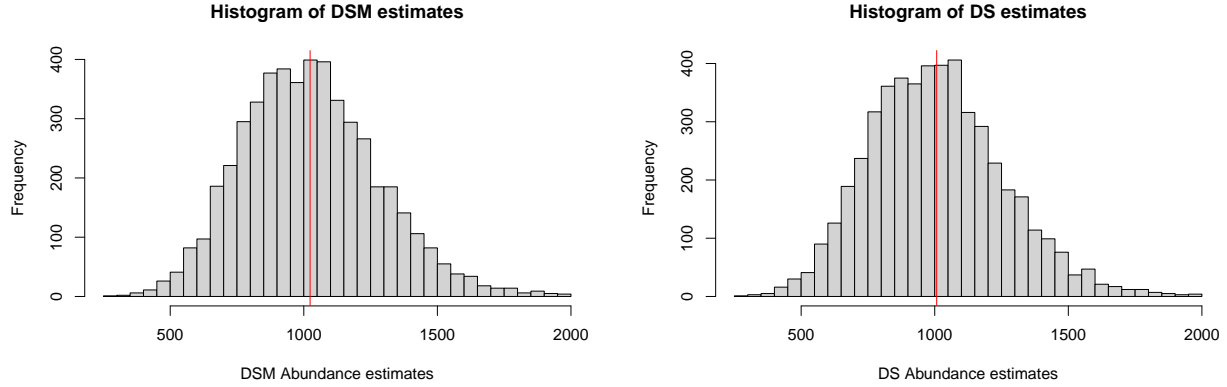

Figure 13: Histograms of Point Design abundance estimates for Stratified North Sea Region

These plots suggest both methods are relatively good at estimating the true abundance of 1000, with both methods appearing to overestimate, with the DSM method overestimating slightly more than the DS method. There again appears to be a slight right skewness to both plots which will be looked at analytically below. Investigating the means further, we find the mean of DSM estimates to be 1023.56 and the mean of DS estimates to be 1007.28. These estimates are both close to the true abundance, with the overestimation present as mentioned above and to a very similar level to the Non-Stratified point results. Overall, it was found that the estimates produced by DSM were greater than those produced by DS from the same survey data 74.08 % of the time, indicating it is highly likely DSM will produce an estimate greater than that of DS. A Wilcoxon test was performed on both sets of estimates with the alternative being that DSM was greater than DS.

This returned a p-values of 0 indicating at any significance level, the null hypothesis of the difference between the DSM and DS estimates being zero is rejected in favour of the alternate, that the DSM estimates are greater than those of DS. If we now compare the other variables extracted from the models below, we can examine the accuracy and variability of each method:

From Table 7, it can be observed that in this case, the DS method overall performed better than the DSM method, having a smaller bias as well as higher confidence interval coverage, despite having a smaller Mean Std error. Additionally, the standard deviation of the estimates is slightly less for the DS method, however both methods do a relatively good job at estimating this. Additionally, we observe from the skewness statistics that both methods returned a positive statistic indicating the right-skewed as identified above, which suggests the right tail is longer and this is supported by some of the estimates being close to 2000, with the maximums for DSM and DS being 1979.15 and 1990.98 respectively. The kurtosis figures

Table 7: North Sea Stratified results - Point

Statistic	Point Design	
	DSM results	DS results
Mean	1023.5635274	1007.2782907
Bias	2.3563527	0.7278291
Mean Std error of estimates	256.9842564	253.9268810
Std Dev of estimates	253.8727239	250.7267820
Mean CV	0.2558409	0.2569336
Confidence Interval Coverage	0.9500000	0.9534000
Estimate Skewness	0.4033869	0.4093011
Estimate Kurtosis	3.2215680	3.2570919

for both methods are greater than 3, indicating we would expect more frequent outliers than a normal distribution. In the Jarque-Bera test for normality, DSM and DS estimates return p-values of 0 and 0 respectively. These indicate we can reject the null hypothesis that these estimates are normally distributed, therefore our use of the Wilcoxon test above is justified. While the Mean CV for the DSM method is lower, the difference is marginal and does not out-way the bias or variance issues of this method. Overall this clearly seems to be a case where the DS method performs better than the DSM estimation method.

6.2.2 Stratifed Results - Parallel Line Design

Now examining the results of the line transect design, on the same density surface, we see the histograms of the estimates below in Figure 14:
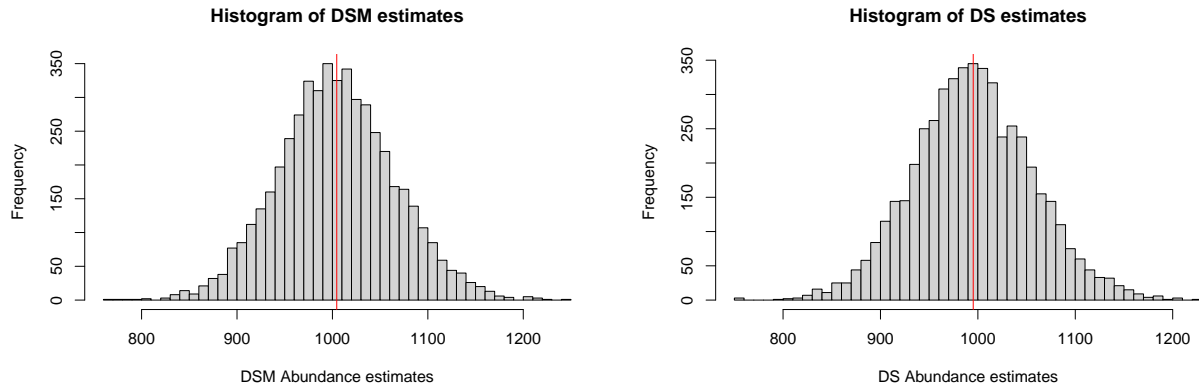

Figure 14: Histograms of Parallel Design estimates for North Sea Region

These plots are similar to their point transect counterparts, with the mean for each signified by the red line appearing very close to truth. However, there appears to be less skewness for these than is present with the point transect results, exhibiting a slightly sharper point at

the mean. Investigating this further, we find the mean of DSM estimates to be 1004.53 and the mean of DS estimates to be 995.22. These estimates are very close to the truth, with DSM again overestimating in comparison to DS. Overall, it was found that the estimates produced by DSM were greater than those produced by DS from the same survey data 81.86 % of the time, indicating it is highly likely DSM will produce an estimate greater than that of DS. A Wilcoxon test was performed on both sets of estimates with the alternative being that DSM was greater than DS. This returned a p-values of 0 indicating at any significance level, the null hypothesis of the difference between the DSM and DS estimates being zero is rejected in favour of the alternate, that the DSM estimates are greater than those of DS. Examining the other statistics extracted from the simulation.

Table 8: North Sea Stratified results - Line

Statistic	Parallel Line Design	
	DSM results	DS results
Mean	1004.5256879	995.2226076
Bias	0.4525688	-0.4777392
Mean Std error of estimates	69.2225719	70.0728537
Std Dev of estimates	61.0506879	61.0934785
Mean CV	0.0689899	0.0704740
Confidence Interval Coverage	0.9710000	0.9720000
Estimate Skewness	0.0779817	0.0779652
Estimate Kurtosis	3.1668925	3.1843860

From Table 8 above, it can be seen that in this case, the performance of both models was very similar throuhgout all areas, with the exception being the mean estimates and bias. The mean estimate for DSM was again a slight overestimation and as a result, bias was positive, while DS underestimated by almost the same amount, leading to an almost identical but negative bias. The Mean Std errors and Std Deviation of estimates were slightly lower for DSM, however the difference in both cases is very small. The Mean CV for each method is very similar, being slightly in favour of the DSM method although the difference is negligible. The DS method has a slightly higher probability of truth being contained within its confidence intervals, at 0.972 compared to 0.971, however this small difference is likely due to the slightly higher Std errors from the DS method widening the confidence interval. Despite this, both methods achieved truth in their confidence intervals over the 95% significance level. The skewness scores for both are very similar and suggest again a slight right skew however this is small. The kurtosis scores are close to 3 but larger, suggesting the distribution of the estimates are is similar to that of a normal distribution, with a slight trend towards a leptokurtic distribution. In the Jarque-Bera test for normality, DSM and DS estimates return p-values of 0.0043605 and 0.0023014 respectively. These indicate we can reject the null hypothesis that these estimates are normally distributed and our use of the Wilcoxon test above is justified, since the normality assumption would have been violated for a parametric test. Overall, while the mean estimate appears in favour of the DS method, the DSM approach can be said to have performed better in this simulation, due to the small bias, lower Mean Std error and coefficients of variation.

6.2.3   Stratifed Results - Zigzag Line Design

Unfortunately, for the stratified zigzag design, an error intermittently occurred within the simulation which caused the fitting of the density surface model to fail. Upon further inspection, this error did not occur within any of the code written as part of this project. After consultation with L Marshall, the author of the *dsims* package, it was determined the error likely occurred as a result of how some of the survey realisations within *dsims* were constructed. This likely led to issues with the coordinates for the transects and segments, with this not being an issue within the DS model however the density surface model would not accept these, resulting in the error. It is outside the scope of this project to investigate the root cause of this error, therefore L Marshall was notified of the bug and given access to the code to determine the root cause. As mentioned, this error occurred intermittently, resulting in a full set of 5000 iterations not being completed in a single run. However, multiple runs were completed with the exact same conditions until a total of 5000 were completed, with these results being examined below. Examining these results, we see the histograms of the two abundance estimates below in Figure 15:
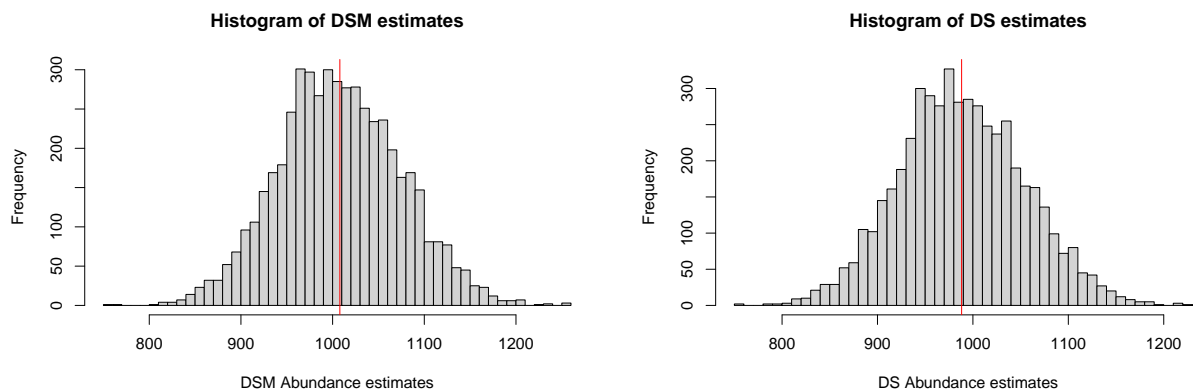

Figure 15: Histograms of Abundance estimates generated by a zigzag line design for the Stratified North Sea region.

These plots are very similar to their point transect counterparts, with the mean for each signified by the red line appearing very close to truth.However, there appears to be less skewness for these than is present with the point transect results. Investigating this further, we find the mean of DSM estimates to be 1007.89 and the mean of DS estimates to be 988.17. These estimates are very close to the true Overall, it was found that the estimates produced by DSM were greater than those produced by DS from the same survey data 99.46 % of the time, indicating it is highly likely DSM will produce an estimate greater than that of DS. A Wilcoxon test was performed on both sets of estimates with the alternative being that DSM was greater than DS.

This returned a p-values of 0 indicating at any significance level, the null hypothesis of the estimates being zero is rejected in favour of the alternate, that the DSM estimates are greater than those of DS. Examining the other statistics extracted from the simulation.

Table 9: North Sea Non-Stratified results - Zigzag

Statistic	Parallel Line Design	
	DSM results	DS results
Mean	1007.8906091	988.1679758
Bias	0.7890609	-1.1832024
Mean Std error of estimates	71.3800598	71.1709390
Std Dev of estimates	66.5563054	65.6280487
Mean CV	0.0709281	0.0721369
Confidence Interval Coverage	0.9634000	0.9666000
Estimate Skewness	0.1105532	0.1130020
Estimate Kurtosis	2.9413753	2.9488149

From the results in Table 9 above, it can be seen that in this case, the performance of both models was similar and each performed well in different areas compared to the each other. The mean estimate was again a slight overestimation and as a result, bias was slightly higher for DSM, however this difference was comparatively small. The Mean Std errors were lower for DSM and far closer to the Std Dev of the estimates, highlighting the better error estimation of DSM. However the Std Dev of the estimates was slightly smaller for the DS approach. The DS method also has a higher probability of truth being contained within its confidence intervals, at 0.9666 compared to 0.9634, however this difference is likely due to the higher Std errors from the DS method wideing the confidence interval. Despite this, both methods achieved truth in their confidence intervals over the 95% significance level. The skewness scores for both are very similar and suggest again a slight right skew however this is very small. The kurtosis scores are very close to 3, suggesting the kurtosis is very similar to that of a normal distribution. In the Jarque-Bera test for normality, DSM and DS estimates return p-values of 0.0042941 and 0.0037219 respectively. These indicate we can not reject the null hypothesis that these estimates are normally distributed. Overall, while the mean estimate appears in favour of the DS method, the DSM approach can be said to have performed better in this simulation, due to the small bias, lower Mean Std error and coefficients of variation.

7 Modeling - Buckland 2015 Example

For our next set of simulations to examine the capabilities of the DSM approach in respect to the DS method, I will revisit the the example simulation provided in Section 2.5.2 of Buckland et al. (2015) and use both a DS and DSM approach to repeat these simulations. It must be noted that the original analysis was part of a survey design case study by Laura Marshall and used the R package *DSsim* Marshall (2020) . For the purpose of this study, each of the designs and simulations will be translated into the *dsims* Marshall (2021) package used throughout the remainder of the simulations to ensure comparability with the DSM approach developed. All details from the original designs will be kept constant, with the population being fixed at 1500 individuals and using a truncation distance of 1000m. For each of the surveys, a half-normal detection function with a scale parameter of 500 will be assumed. The density surfaced used is displayed below alongside the detection function in Figure 16.
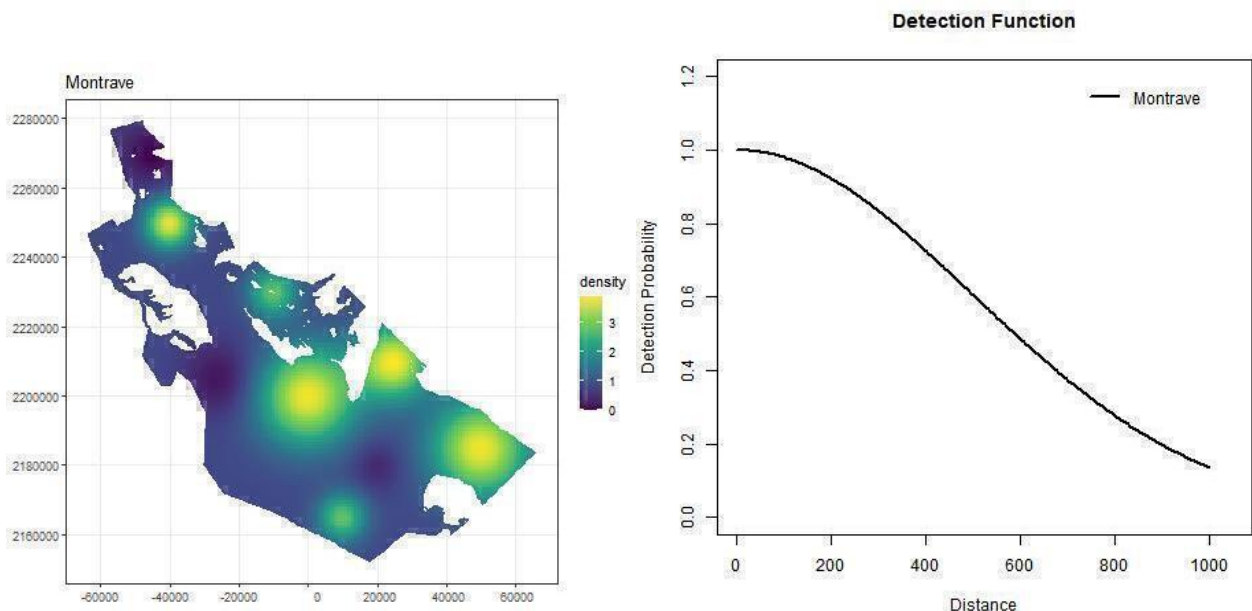

Figure 16: Density surface and detection function for Buckland 2015 example

By using this study region, it is possible to verify the simulation method developed for DSM can be run on study regions with internal holes, such as islands on a seaborne survey. The three different designs will be used, being the same as those tested in the case study in Buckland et al. (2015). These will be a systematic parallel line design, a random parallel line design and a zigzag design. Examples of each design are displayed in Figure 17 below:

As with each of the simulations run as part of the North Sea section, 5000 iterations will be run for each design. While this is different from the 100 iterations originally conducted, the increase in the number of iterations will increase the reliability of the results.
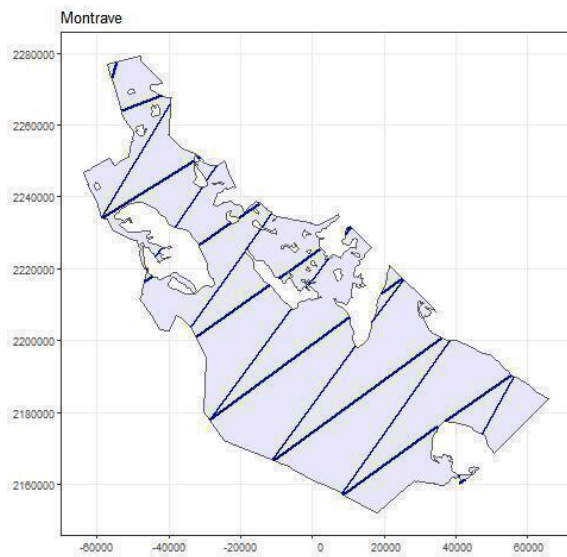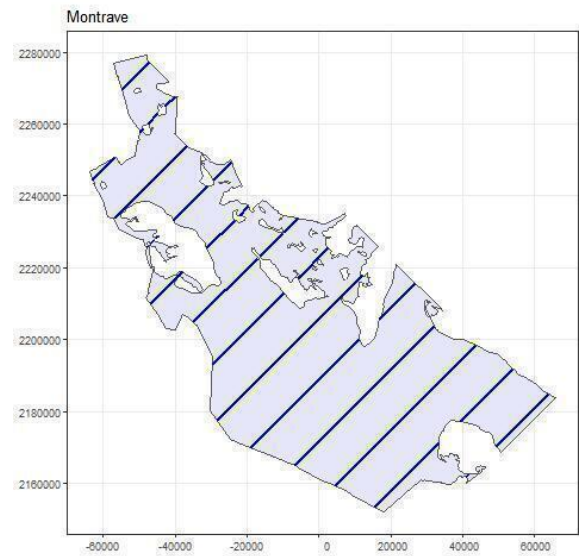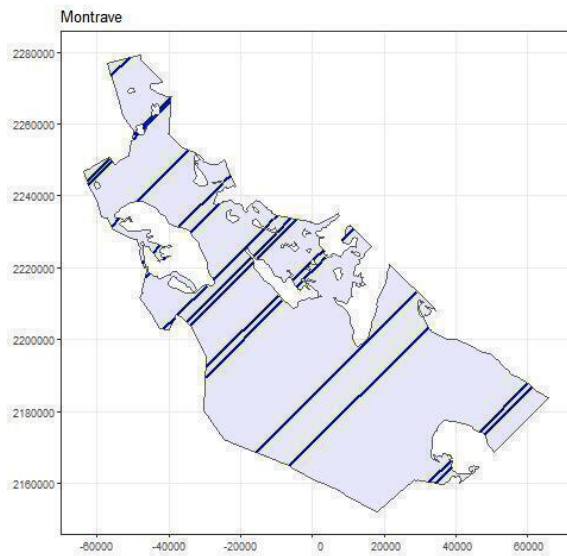
Figure 17: Example surveys for each Design: Top Left: Random Parallel Line, Top Right: Systematic Parallell line, Bottom Left: Zigzag line

7.1 Results - Random Parallel Line

We can now examine the results of the Random parallel line design by initially plotting histograms of the estimates from both the DS and DSM methods below in Figure 18:
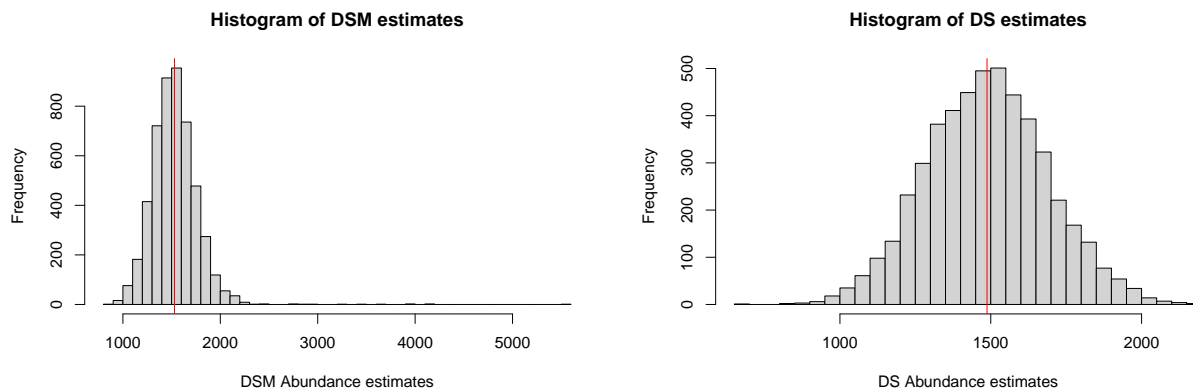

Figure 18: Histograms of Random Parallel Line Design estimates for Buckland 2015 example region

While the plot for the DS estimates appears similar to those we have seen before, the estimates for the DSM method appear markedly different. This is due to the prescence of very extreme overestimates with the largest estimate being 5552, over 3 times the true population. It initially appears that the majority of the DSM estimates are centred around the true population of 1500, however this will be investigated further below. As with the previous simulations, it was found that the estimates produced by DSM were greater than those produced by DS from the same survey data 61.68 % of the time, indicating it is likely DSM will produce an estimate greater than that of DS. A Wilcoxon test was performed on both sets of estimates with the alternative being that DSM was greater than DS.

This returned a p-values of $6.5547013 \times 10^{-84}$ indicating at any significance level, the null hypothesis of the diffeence between the estimates being zero is rejected in favour of the alternate, that the DSM estimates are greater than those of DS.

Table 10: Buckland 2015 example results - Random Line

Statistic	Random Parallel Line Design	
	DSM results	DS results
Mean	1529.7525902	1487.6347184
Bias	1.9835060	-0.8243521
Mean Std error of estimates	175.4654981	196.5362110
Std Dev of estimates	238.4670400	201.3611866
Mean CV	0.1139636	0.1334070
Confidence Interval Coverage	0.8880000	0.9420000
Estimate Skewness	2.4886720	0.0578036
Estimate Kurtosis	31.3798293	2.9516809

In Table 10 above, we can see the various statistics extracted during the course of the simulations. As we have seen throughout throughout this report, the mean for the DSM is a slight overestimates and the DS is a slight underestimates. This is transferred into the bias, with the DSM having a positve bias and DS having a negative bias, however both of these are below the 5% level so no concerns are raised for either method. Unfortunately, there are major concerns regarding the Mean Std errors of the estimates, which are the estimator of the Std deviations. For both methods, it can be seen that they are underestimates of the Std deviations and very severely in the case of the DSM method. This is problematic as not capturing all of the variance can lead to the the results assuming they are more accurate than they actually are. These underestimations of variance are translated into the confidence interval coverage, where both fail to contain truth at the 95% confidence level. Additionally, the skewness and kurtosis statistics are very different for the DSM method, likely as a result of the extreme overestimations produce by some survey realisation. Overall, it could be concluded that in this scenario, the DS method performed better, by virtue of underestimating the variance less and having less extreme estimates, however neither method should be considered adequate in this case.

7.2 Results - Systematic Parallel Line

Moving on the systematic parallel line design, histograms of the abundance estimates are plotted below in Figure 19.
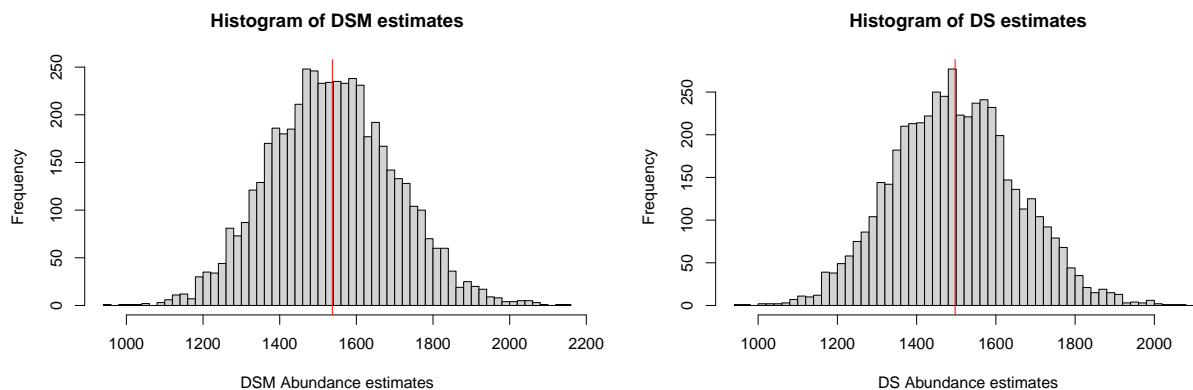

Figure 19: Histograms of Systematic Parallel Line Design estimates for Buckland 2015 example region

These plots in Figure 19 appear to be a significant improvement over those seen with the random line design. This can be determined based on the significant reduction in the range of estimates, especially in the case of the DSM approach. The mean of the estimates, also appears very close to truth at 1500, with some indication that the DSM approach has slightly overestimated. A more detailed analysis of the simulation results can be seen in Table 11.

Table 11: Buckland 2015 example results - Parallel Line

Statistic	Systematic Parallel Line Design	
	DSM results	DS results
Mean	1537.3035256	1497.1665768
Bias	2.4869017	-0.1888949
Mean Std error of estimates	167.4557120	217.2829383
Std Dev of estimates	162.7215057	157.1575653
Mean CV	0.1093682	0.1455074
Confidence Interval Coverage	0.9516000	0.9912000
Estimate Skewness	0.1282707	0.1328383
Estimate Kurtosis	2.9807139	2.9845720

From the results in Table 11 the overestimation implied from the histograms is clearly seen, with the mean of the abundance estimates having a bias of almost 2.5%. However, when it comes to the variance estimation, the DSM approach is far superior in this case. While the true variance, the Std Deviation of the estimates, is lower for the DS approach, the estimate of this, the Mean Std error of the estimates is dramatically larger than what it is trying to

estimate. In contrast the DSM variance estimate is very close to the true variance. These differences in variance estimation the propagate into the coefficients of variation, with again the DSM method having a lower CV. These also have an effect on the confidence interval coverage, with the DS results capturing truth over 99% of the time. However, this is most likely due to the high variance estimate for this approach widening the confidence intervals. Despite the DSM method having a lower confidence interval coverage, both methods achieved coverage at the 95% level resulting in no cause for concern with either method. The Skewness and Kurtosis statistics are near identical for both sets of abundance estimates and appear close to those of a normal distribution. In the Jarque-Bera test for normality, DSM and DS estimates return p-values of 0.0010135 and 6.252889×10^{-4} respectively. These indicate that we can reject the null hypothesis of normality for both sets of estimates in favour of the alternate that they are not normally distributed.

7.3   Results - Zigzag Line

The final simulation run as part of the case study in Buckland et al. (2015) used a zigzag design, the results of which are displayed below:
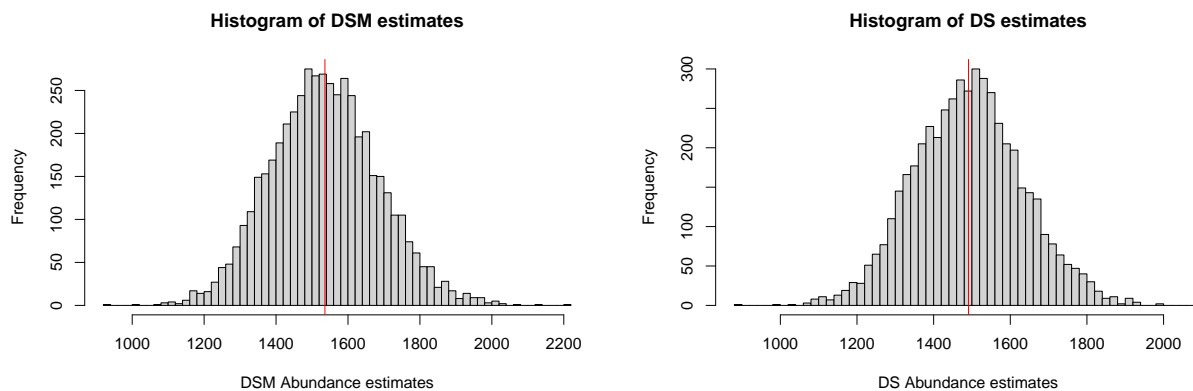

Figure 20: Histograms of Zigzag Line Design estimates for Buckland 2015 example region

In Figure 20 above the histograms of the abundance estimates for the DSM and DS approaches can be observed. These appear to follow a similar pattern to those seen in the systematic parallel line simulation above. Namely both histograms look fairly pointed, with the mean for the DSM estimates appearing to be a slight overestimate while the mean for the DS abundance estimates looks to be close to truth. The results are examined more in-depth in Table 12 below:

The results seen within Table 12 appear very similar to those seen for the systematic parallel line design in Table 11. Once again, the DSM approach has slightly overestimates the abundance, with a bias of over 2%, however this is below the level of 5% at which concerns could be raised. In contrast, the DS approach slightly underestimates the true abundance, having a bias of below 1%. The similarities continue into the variance estimates, where DSM

42

Table 12: Buckland 2015 example results - Zigzag Line

Statistic	Zigzag Line Design	
	DSM results	DS results
Mean	1535.8679659	1490.9859160
Bias	2.3911977	-0.6009389
Mean Std error of estimates	154.0152961	201.4087609
Std Dev of estimates	149.6840030	142.3056900
Mean CV	0.1005728	0.1353724
Confidence Interval Coverage	0.9488000	0.9936000
Estimate Skewness	0.1562206	0.0825283
Estimate Kurtosis	3.1318763	3.0696502

does a far better job of estimating the true variance than the DS method. This leads to the DSM results having a lower coefficient of variation, a desirable attribute. However, when investigating the confidence interval coverage, the DS approach, as usual, perform better, with truth being within the confidence interval over 99% of the time, although, as has been noted previously several times, this is likely due to the large variance estimate. In contrast, the coverage from the DSM method is just below 95% raising some elements of concern about its ability to cover truth within the confidence interval. The skewness and kurtosis statistics agree with the initial observations of the histograms, with the kurtosis suggesting a slightly leptokurtic distribution, one with larger tails and more pointed in the centre than that of the normal distribution, while the skewness suggests a slight right skew to the estimates, indicating a bias towards higher estimates. Overall, it could be concluded that the DSM approach proved superior in this case, mainly due to the accuracy of the variance estimation over the DS approach. While the Bias and confidence interval coverage were not as good as those for the DS method, I would not consider them sufficiently concerning to out-way the improvement in variance estimation.

Additionally, both the Systemic parallel line design and the zigzag design were selected and specified such that they had a comparable cost to the survey. This cost is the trackline length, which is to total distance required to travel along each transect and includes the distances between transects. This is often termed the cost of a study as the distance travelled during a study is a major factor in the overall cost of a study, such as through fuel costs for air or seaborne studies as well as the time and resources associated with additional distances travelled. In these designs, the mean trackline length was around 700km for both survey designs. This allows for the results of the two designs to be compared and would be one of the steps conducted when selecting an appropriate survey design in the lead up to a real work survey. Comparing the results in both Tables 11 and 12, it can be seen that the mean estimates from the DSM method are consistent accross the two survey designs, with a similar level of positive bias present in both. With the DS results, there appears to be a reduction in bias when using the parallel line design. Looking at variance estimation, the variance estimates from the zigzag designs are lower for both methods, indicating an increase in the precision as a result of using the zigzag design. The coefficients of variation stayed

similar across the different designs as did the confidence interval coverage, with the skewness and kurtosis statistics also changing little between the two designs. Overall these results are fairly consistent with those seen in Buckland et al. (2015) with it being noted that the zigzag design offered a higher level of precision than the Mean Std error suggests. This has been confirmed by both the original DS method as well as the DSM approach, with this approach in fact giving improved estimates for the variance. It is however noted within Buckland et al. (2015) that while the trackline length was similar for both designs, in situations where travelling on effort, i.e on a transect, is far more costly than off effort, the advantage gained from the zigzag design are negated by the addition cost.

8 Modeling - Variance estimation tests

As has been noted through the analysis of previous simulations, the variance estimation of the DS method has often performed poorly in comparison with that of the DSM approach. As noted within Section 3.2, the default 'R2' encounter rate variance estimator has been used throughout all of the simulations so far, as according to Fewster et al. (2009) this shows good performance for random line designs. However, many of the simulations run have used systematic line designs, for which it is noted that 'R2' may not perform well. In this case, Fewster et al. (2009) recommends using the 'O2' estimator. To investigate if changing the variance estimator improves the performance of the DS method, we return to the non stratified North Sea region in Section 6. Here, the simulation will follow the same systematic line survey used previously, as seen in Figure 10 alongside the density surface seen in Figure 8. The detection function was also kept constant, using a half normal function with a scale parameter of 5, as seen in Figure 9. The only difference between the two simulations is the specification of the encounter rate variance estimator, with one simulation using the 'R2' estimator and the other using the more appropriate 'O2' estimator as recommended by Fewster et al. (2009).

8.1 Results

Table 13: North Sea Line transect results with "R2" encounter rate variance estimator

Statistic	"R2" estimator	
	DSM results	DS results
Mean	1009.6565694	998.4464616
Bias	0.9656569	-0.1553538
Mean Std error of estimates	61.1480511	86.7946268
Std Dev of estimates	51.6240111	51.4407921
Mean CV	0.0606163	0.0869709
Confidence Interval Coverage	0.9772000	0.9990000
Estimate Skewness	0.0757837	0.0585542
Estimate Kurtosis	2.9440610	2.9392117

From Table 13 the same pattern as has been seen throughout this report is evident. The variance estimate, the Mean std error, for the DS method is far larger than the true variance, the Std Dev of the estimates, by over 60%. In comparison, the variance estimate for the DSM approach is only a slight overestimate at only around 10% higher. These then transfer into the coefficients of variation with the DSM having a respectively lower CV than that of the DS method. The confidence interval coverage for the DS method is very impressive, however this is likely due to the large variance estimates widening the confidence interval appreciably to contain truth. While the coverage for the DSM method is slightly lower, it is still above the 95% confidence limit and does not raise concerns. Overall in this case the

DSM method can be said to have performed better, mainly due to the substantially better variance estimate and resulting coefficient of variation improvements over the DS method.

Table 14: North Sea Line transect results with "O2" encounter rate variance estimator

Statistic	"O2" estimator	
	DSM results	DS results
Mean	1009.5655104	998.3978929
Bias	0.9565510	-0.1602107
Mean Std error of estimates	61.1544936	61.6984808
Std Dev of estimates	51.7560955	51.5678540
Mean CV	0.0606277	0.0618436
Confidence Interval Coverage	0.9758000	0.9786000
Estimate Skewness	0.0537371	0.0395134
Estimate Kurtosis	3.0159802	3.0367704

Moving on now to look at the results in Table 14, from when the variance estimator is set as the 'O2' estimator. It can be seen that there has been a significant improvement in the variance estimation in comparison to the 'R2' results above. The variance estimates for both methods are near identical now, with both being an overestimate of the true variance by around 10%. It appears that while the variance estimate for the DSM method is slightly closer than the DS approach, the difference is very marginal, highlighting the improvements to the estimation in the DS method. The confidence interval coverage for DS is slightly lower than previously, with this being the only statistic which appears worse for the 'O2' estimator than the 'R2' estimator. However this is likely due to the more representative variance estimate shrinking the confidence interval. Despite this, both methods acheived a confidence interval coverage above the 95% level. Overall in this case there is very little to differentiate between the two methods. It could be considered that the DS method performed better, as given that all other statistics are near identical, the mean estimate is closer to truth and as a result the bias is slightly smaller. However, the DSM method still performed well, with its bias being below 1%.

These simulations have highlighted the improvements which can be made to the DS method to improve its variance estimation and make it comparable or possibly better than the DS method, if the particular design allows.

9   Modeling - North Sea Predictions outside study area:

As mentioned within the simulation extensions section, the prediction grid for DSM is not limited to the area of the study region, it can be smaller or larger and does not even need to include any part of the study region. However, predicting populations outside of the original study region may be subject to large variations dependant on the gradient of the density surface at the edge of the region. This will be tested by using the same North Sea region we have explored before, however with two very contrasting density surfaces to show the two extremes of what could happen in this scenario. As this simulation will focus on predicting the abundance outside of the study region, the estimates from the DS method will not be compared here as the predictions will be made over a far larger area. Additionally, there is no true population for the estimates to be compared against as the population of 1000 was only specified for the study region. As a simple example, the prediction grid used will be based on the bounding box of the North Sea region, in essence a large rectangle encompassing the entirety of the region.

The two contrasting density surfaces can be seen in Figure 21
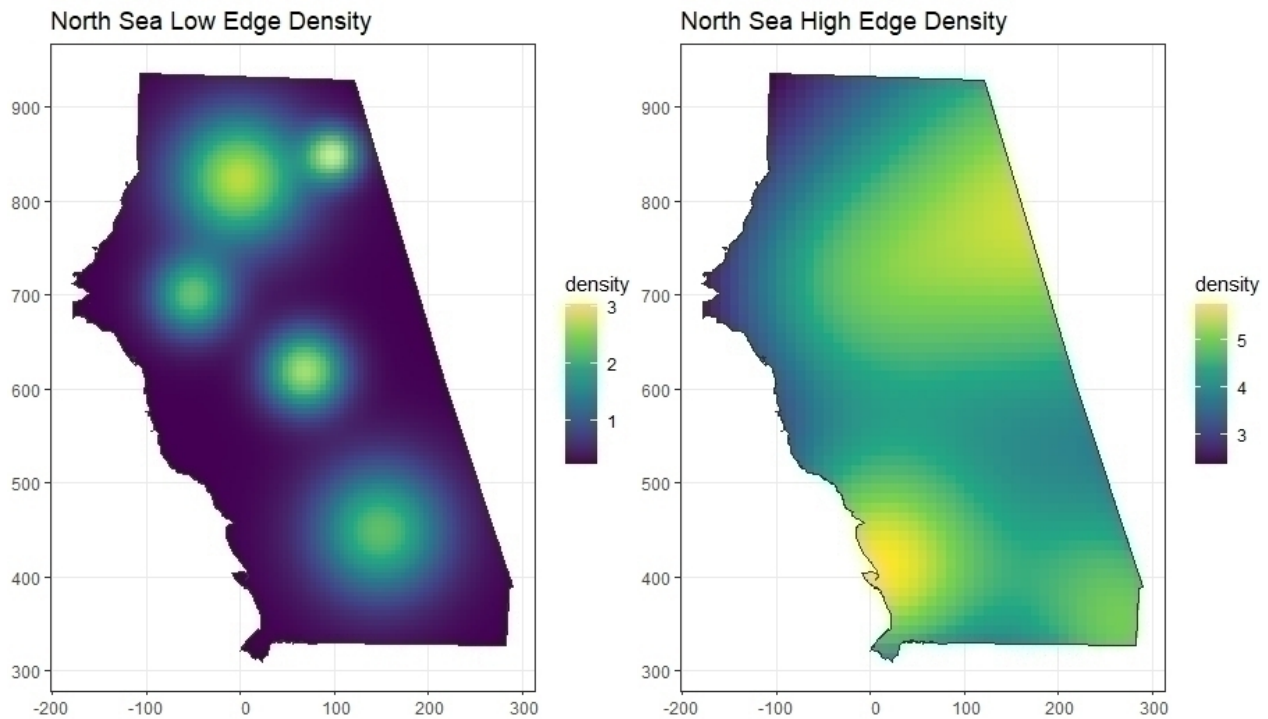

Figure 21: Density surface and detection function for Montrave region

In a slight difference to the original simulations, the number of transects was reduced to 20 to reduce the amount of area covered by each study. All other aspects of the original simulation were kept the same, with the detection function being the same half normal function with a

scale parameter of 5 with a truncation distance of 10 seen in Figure 9.
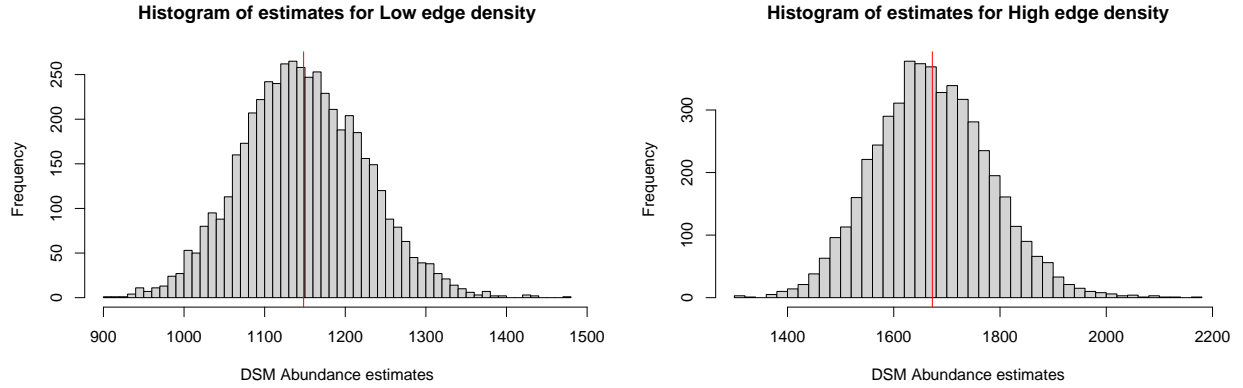
9.1 Results

Figure 22: Histograms of DSM abundance estimates for the North Sea region when predicting outside the study region using different density surfaces

In Figure 22 it can be observed that while the estimates exhibit a very similar shape, the location of the plots are significantly different, with the low edge density located around 1150 in contrast to 1700 for the high edge density approach. This implies that despite the same region being predicted over and the study area containing 1000 for each simulation, the predictions outside the study area are greatly dependant on the underlying density surface. A more detailed comparison of the results can be seen in Table 15:

Table 15: North Sea region with different densities when predicting outside the study region

Statistic	DSM results	
	Low Edge Density	High Edge Density
Mean	1148.3549197	1672.6246878
Mean Std error of estimates	82.8996866	125.3147025
Std Dev of estimates	75.9493672	108.5915038
Mean CV	0.0722341	0.0749656
Estimate Skewness	0.1509434	0.2247339
Estimate Kurtosis	3.0347758	3.2674359

From the results in Table 15 the substantial difference in the two density estimates can be observed, with the high edge density (HED) estimate being almost 50% greater than the low edge density (LED) estimate. This is likely due to the model predicting that the density for the HED will continue to rise outside the study area but still within the prediction region, massively increasing the abundance estimates. The opposite is likely true for the LED case.

48

When variance estimation is considered, the LED simulation appears very good at estimating the variance, possibly as a result of the areas outside the study area contributing less to the predicted estimate, reducing the variance within the model. in contrast, the variance estimate for the HED simulation shows an overestimate, though not a significantly concerning one. The overall variance in this case is increased compared to the LED simulation, likely due to the steeper density gradients increasing the models variance estimates. Despite this, the coefficients of variation are very similar for the two different density surfaces, highlighting that while the variance did increase for the HED simulation, when compared to the mean estimate it is proportional to the LED simulation. The skewness and kurtosis statistics indicate that the distribution of the estimates has not changed markedly between the two density surfaces, although the kurtosis is slightly higher for the HED simulation. This indicates these estimates are slightly leptokurtic and have heavier tails than that of a normal distribution, implying extreme estimates are more likely than would be expected under a normal distribution.

Overall these simulations have highlighted that when predicting over regions outside the original study area, care should be taken as the underlying density surface can have a significant influence on the abundance estimates. As we have seen, the same survey design with different underlying density surface has generated vastly different estimate. This does raise cause for concern as should these be used to perform analysis on real world survey data, there may be the potential to seriously under or overestimate the true population of an area, which could result in dire consequences for the populations management plan and survival.

10 Notes

I hope to include maps of the different density surfaces fitted under the outside region models to highlight the differences once I have the code working.

bad design: potential simulation with bad design where detection functions are different for each strata and not giving this info to the dsm model we discussed.

11   Discussion

While an number of different simulation have been conducted throughout this report, it is noted that the majority of these are fairly basic in nature and aim to show that the simulation method developed is able to sufficiently deal with issues encountered early on in the development process. These issues were namely the creation of the segments required for the DSM, with the Buckland 2015 example introducing the most errors as a result of the islands presents which had to be overcome. To that end, there are a number of other areas which I wished to investigate, had time allowed, which can be split into two separate and almost distinct areas, variance estimators within *dsims* and further more in depth simulations to determine the limits of the model and simulations.

11.1   Variance estimators within *dsims*.

As noted within Section 3.2, there are a variety of different encounter rate variance estimators available for distance sampling data, as detailed within Fewster et al. (2009). However, at present, only a select few are available within the *dsims* package and even these are limited based on the particular design chosen. For example, while the 'O2' estimator tested during Section 8 proved to markedly improve the variance estimation, this estimator is only available within the existing simulations when using systematic parallel line designs, and not with the very common zigzag design. While distance sampling analysis can be conducted with variance estimators not available within the simulation, this allows for a potentially unfair comparison between survey designs when the best available estimator is used, as opposed to the best overall and may results in an non-ideal design being selected. As identified throughout the majority of the simulations, the DS approach did not perform as well as the DSM method with regards to variance estimation, highlighting this is an area where substantial improves could be made.

11.2   Further simulations

At the current stage, none of the simulations have looked at covariates within either the detection function or the models themselves. As seen within the example study conducted within Miller et al. (2013), an example of this is using depth within seaborne surveys as a covariate within the models. Due to the likelihood of covariates being considered within both distance sampling and density surface modelling approached, an interesting set of simulations would be to look at how the use of covariates influences the results generates by each method and indeed if the coefficients of these vary between the approaches. At the current moment, the DSM simulations developed are not capable of implementing this, however the additional work required would not be major and would involved ensuring any covariate information is appropriately passed to the relevant segments. The model could then be generalised to account for the additional information and allow these additional factors to be implemented.

Additionally, when discussing covariates and the segments, Miller et al. (2013) suggests that the segment size for DSM should be chosen such that neither the density nor covariates should

change substantially within the segment. A potentially interesting comparison could arise by running simulations in which this was not the case, where a covariate significantly varies within segments. This would establish how badly the DSM performs when the segments are not sufficiently small and if distance sampling offers a viable alternative in this case.

Leading on from this, as noted when generating predictions outside the original study area, the estimates were very different based upon the density prescribed at the edge. Miller et al. (2013) highlights that at edges of the study area, both internal and external, models may link these areas and align the density between the two, without taking account for the area for which no animals may occur. They term this as 'smoothed across,' implying that predictions could in theory generate unphysical abundance estimates, i.e. dolphins present on an island. When constructing the simulations, care was taken to ensure that the prediction area was cropped to the study region in an attempt to remove this effect. However, it is noted that a soap film smoother, a function of spatial location, could be used to account for these complex areas. Also mentioned within Miller et al. (2013) is that even with smoothers, the density profile at the edge can still cause issues. There is the potential for smoothers to have covariates which cause the models fitted density surface to behave unrealistically when predicting the surface a substantial distance from the survey effort. This is likely what occurred when conducting the simulations in Section 9. Miller et al. (2013) does note that this issue may be reduces by selecting an alternate smoother, with a generalisation of Duchon Splines suggested. A further area of study could focus on investigating these effects further, with the potential aim of constructing smoothers which are less prone to exhibiting this behaviour.

12 Conclusion

- overall dsm performs well in comparison to ds
- point transect ds better
- line transects mixed results based on survey design
- for comparable designs, DSM proved more precise
- random designs lead to severe variance underestimation
- using better variance estimators makes DS more comparable to DSM
- predictions outside the study region can be very dependant on edge effects of the region.

References

Bivand, Roger, and Colin Rundel. 2021. *Rgeos: Interface to Geometry Engine - Open Source (GEOS).* https://CRAN.R-project.org/package=rgeos.

Buckland, S. T., D. R. Anderson, K. P. Burnham, and J. L. Laake. 1993. *Distance Sampling: Estimating Abundance of Biological Populations.* Chapman; Hall.

Buckland, S. T., E. A. Rexstad, T. A. Marques, and C. S. Oedekoven. 2015. *Distance Sampling: Methods and Applications.* Springer.

Fewster, Rachel M., Stephen T. Buckland, Kenneth P. Burnham, David L. Borchers, Peter E. Jupp, Jeffrey L. Laake, and Len Thomas. 2009. "Estimating the Encounter Rate Variance in Distance Sampling." *Biometrics* 65 (1): 225–36. https://doi.org/https://doi.org/10.1111/j.1541-0420.2008.01018.x.

Marshall, Laura. 2020. *DSsim: Distance Sampling Simulations.* https://CRAN.R-project.org/package=DSsim.

———. 2021. *Dsims: Distance Sampling Simulations.* https://github.com/DistanceDevelopment/dsims.

Miller, David L., M. Louise Burt, Eric A. Rexstad, and Len Thomas. 2013. "Spatial Models for Distance Sampling Data: Recent Developments and Future Directions." *Methods in Ecology and Evolution* 4: 1001–10. https://doi.org/https://doi.org/10.1111/2041-210X.12105.

Pebesma, Edzer. 2022. *Sf: Simple Features for r.* https://CRAN.R-project.org/package=sf.

Souchay, Guillaume. 2020. *Density Surface Model for Hare Point Transect Survey.* https://examples.distancesampling.org/dsm-point/hare_point_transect_dsm-distill.html#:~:text=Density%20surface%20model%20for%20hare%20point%20transect%20survey,is%20calculated%20properly.%20...%204%20Making%20predictions.