

Panic in the Disko Bay: Distance Sampling for Identifying Bowhead Whales

14 February 2024

Project 1: Distance Sampling



University of
St Andrews

School of Mathematics and Statistics

in partial fulfilment of the requirements for
MT5751: Estimating Animal Abundance and Biodiversity

Abstract

Introduction

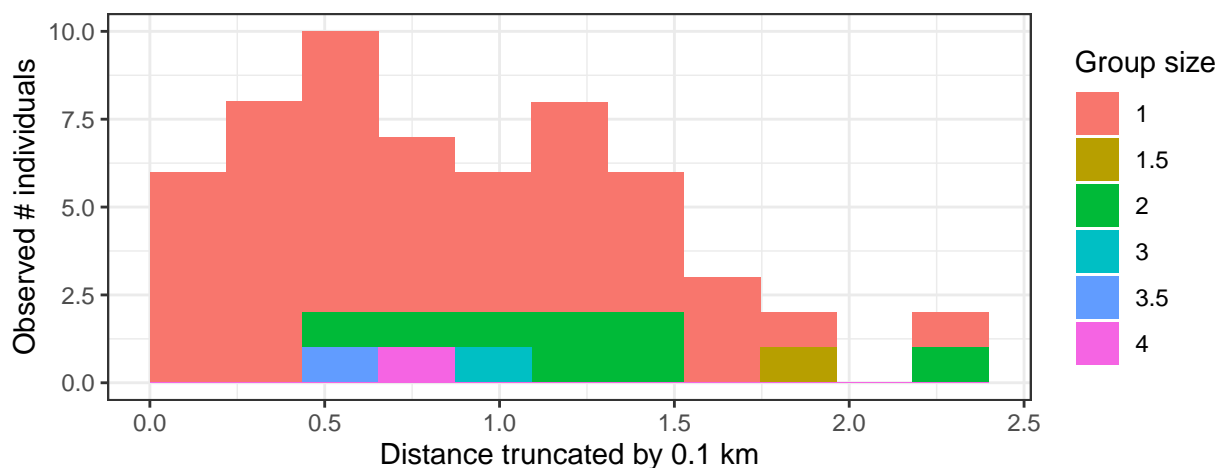
In this report, we aim to use distance sampling methods to estimate the abundance of Bowhead whales in Disko Bay. As the data show some observer avoidance, we also wish to explore different approaches used to deal with this issue.

Methods

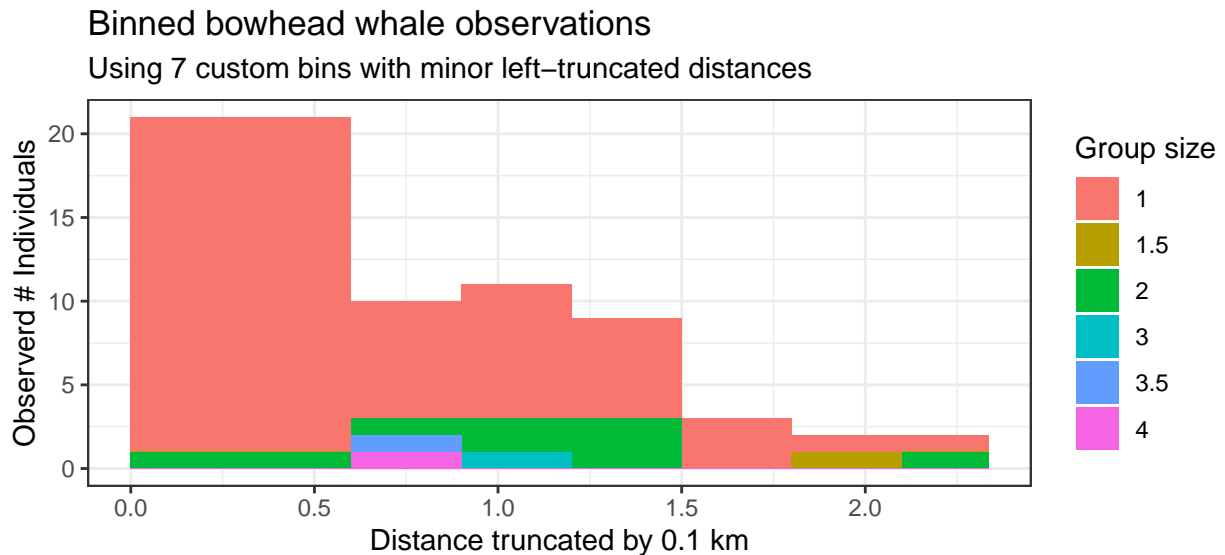
Visual aerial surveys were conducted along 41 systematically placed east-west along the West Greenland coast, with a combined length of 4,445km. There were two observation platforms which each recorded the declination angle using inclinometers and the times of the first sighting and when the whale passed abeam, which were used to calculate the perpendicular and forward distances respectively. Both observers also recorded group size, and if there were any differences in size or declination between the two observers the mean of the two values was taken. The Beaufort sea state was recorded for each survey. The distances were left-truncated to 100m as the view was obscured close to the transect line(Rekdal et al. 2015). In total there were 58 valid whale observations.

Raw bowhead whale observations

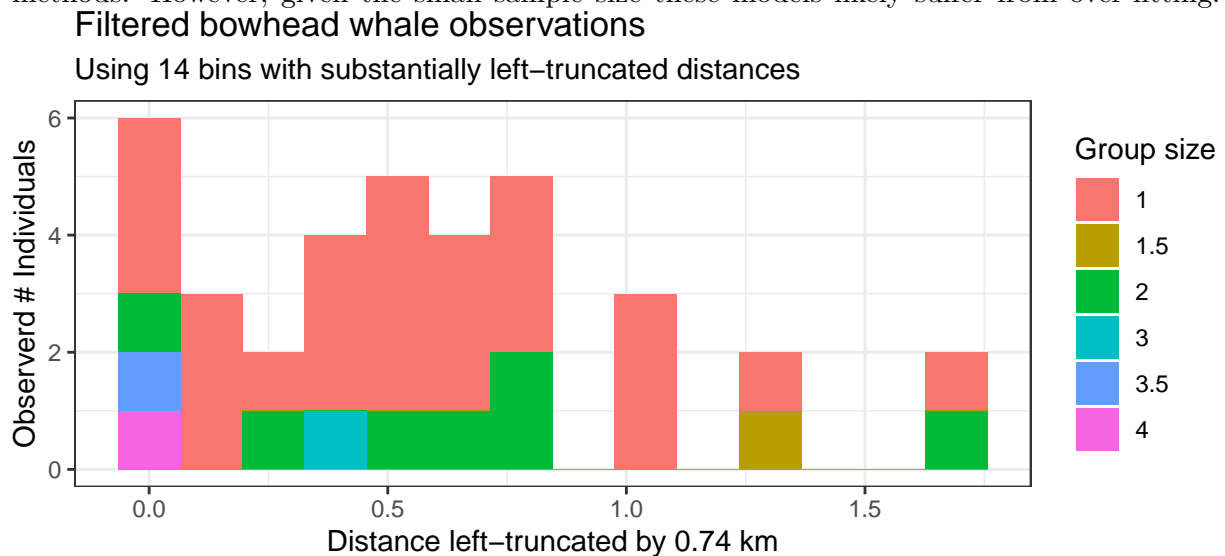
Using 12 bins and minor left-truncated distances



An initial plot of the detection distances suggests there could be some line avoidance, as more whales were detected further away from the plane than on the transect, even after the left-truncation. This could violate the assumption in distance sampling that animals are distributed independently with respect to the transect, or the assumption that all animals on the line are detected, depending on whether the whales are hiding or moving away in response to the plane (Buckland et al, 2015). Equally, this could just be stochastic variation as there is only a difference of 4 observations between the 0-0.2 bin and the 0.4-0.6 bin. We considered 3 approaches to deal with this problem. Firstly, treating the difference in detection as random variation and fitting detection functions as normal. This preserves all the observations but will result in overestimating the detection probability (and therefore underestimating abundance) if detection is not certain on the line. Another option is to bin the distances into 0-0.6km, 0.6-0.9km, ..., 2.1-2.4km categories so that there are more observations in the first group than the others. This assumes the whales moved away from the observers, so those that would have been detected on the transect were instead seen up to 0.6km away. This method deals well with the potentially violated assumptions, but in grouping the data we lose a lot of information.



The final option we considered is to left-truncate the distances up to the peak in detections at 0.64km. This method ensures we have the monotonic decline in detections with increasing distance, but again loses a lot of information and assumes that detection is certain 0.64km away from the plane. The monotonic decrease in detection allowed complex models with multiple covariates and adjustment terms to be fitted successfully, which didn't happen for the other methods. However, given the small sample size these models likely suffer from over-fitting.



We used the Distance package to fit models using a maximum likelihood approach. For each data set (unedited, binned and truncated) we fitted half normal and hazard rate detection functions with all combinations of adjustment terms (cos, Hermite and polynomial) and covariates (group size and Beaufort sea state). We used AIC to compare between models fitted to the same data set.

Results

Given raw data set the best model selected by AIC was half-normal detection function with size covariate and no adjustments. The model estimates whale abundance as 212 with confidence interval between 112.13 and 401.06. However, covariate included is not statistically significant, therefore, model without covariate gives similar estimate of whale abundance (229) with confidence interval between 112.13 and 401.06.

When using binning, the best model selected by AIC is half-normal with no adjustments and size

Table 1: Model Summary Table

Model	Estimate	SE	LowerCI	UpperCI	AIC
Base HN + size	212.07	65.62	112.13	401.06	86.03
Base HN	228.82	71.45	122.00	429.14	86.32
Binned HN + size	212.84	66.14	112.32	403.31	197.58
Binned HN	229.54	72.22	121.94	432.07	197.91
Truncated HR + size	1803.36	1346.22	461.56	7045.88	27.20
Truncated HN	519.51	90.74	363.25	743.00	28.43
Truncated HN + size	608.75	154.26	356.93	1038.20	28.73

Table 2: Significance of 'Size' covariate

Model	CoefficientofSize	SE	SignificanceofSize
Base HN + size	6.16	5046.40	Not Significant
Binned HN + size	5.64	305.31	Not Significant
Truncated HR + size	-1.53	0.60	Significant
Truncated HN + size	-0.51	0.26	Significant at 0.1 level

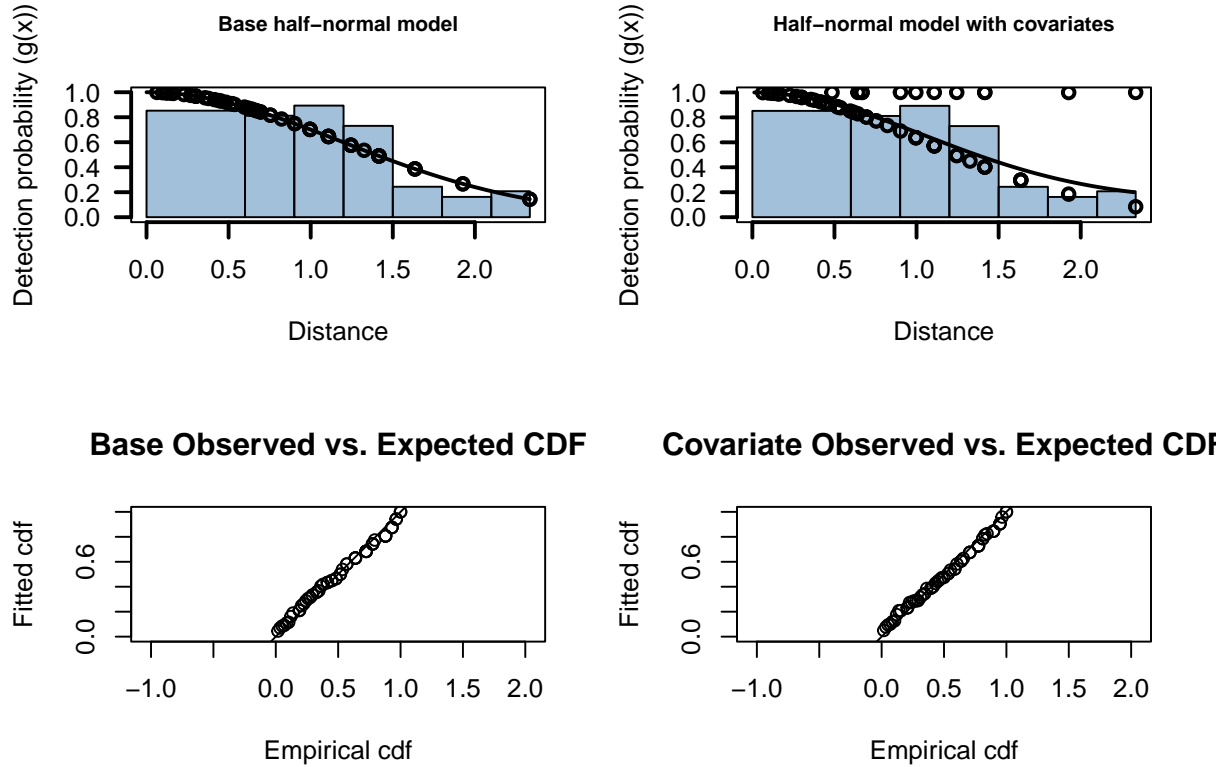


Figure 1: Detection functions using raw data

covariate. Similar to model using raw data, it gives estimate of abundance of 213 and confidence interval between 112.32 and 403.31. Covariate is not statistically significant.

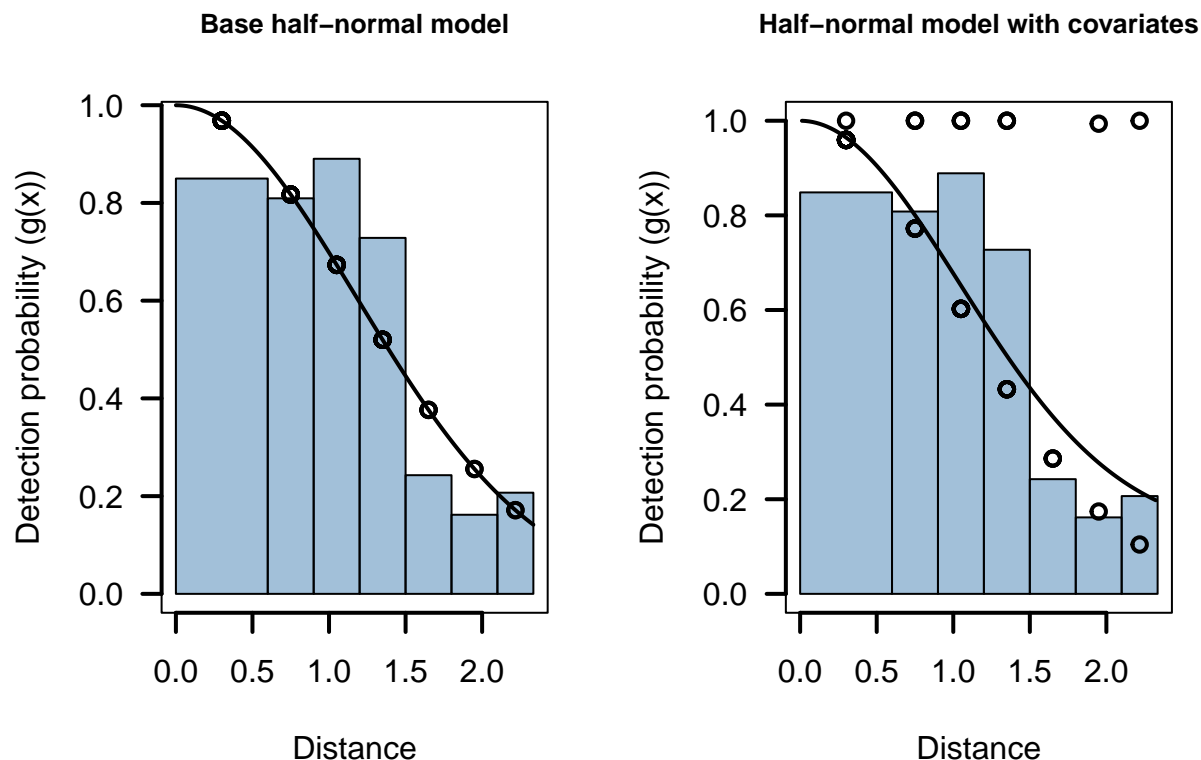


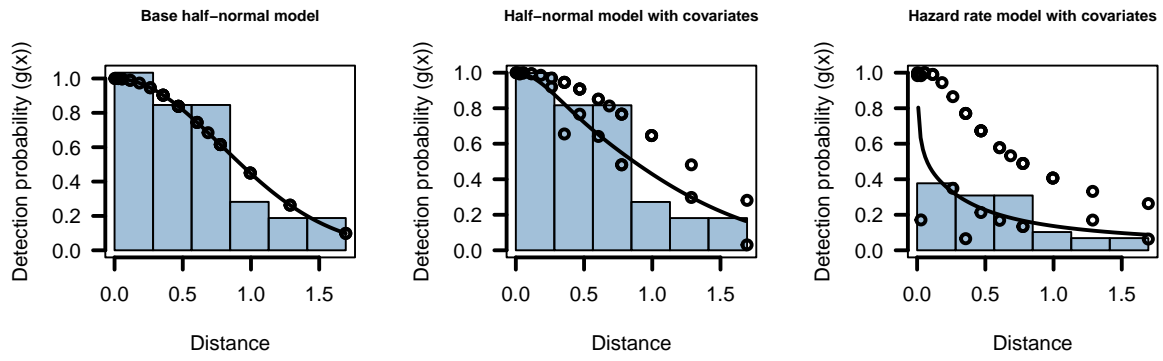
Figure 2: Detection functions using binned data

However, preferred model chosen using truncated data is quite different. The best model is hazard rate with size covariate, which gives an estimate of 1803 whales with confidence interval between 461.56 and 7045.88. It is worth noting, that size covariate is significant here. However, given huge standard error, there's a lot of uncertainty, therefore, it is likely that the model largely overestimates. It is then followed by half normal model with no covariates and half normal model with size covariate.

Even though detection functions using truncated data has lowest AIC, it overestimates whale abundance. The reason might be limited to sample size and extrapolation based on the area we have no data about. Therefore, it might be best to use half normal detection function with size as a covariate to get the best estimate of whale abundance.

Discussion

Given that whale count was conducted from the aircraft, there are several assumptions that might be violated. Since the area under the aircraft is not visible to surveyors, it creates a blind strip and, therefore, we cannot assume that objects on the transect line are detected with certainty. To account for this, passive monitoring such as cameras to check for the animals on the transect line could be used. Moreover, assumption that objects are detected at their initial locations is violated as well. Due to avoidance response, as whales notice observers, they might feel threat and are more likely to run away or hide which results in less observations within close distance. Since whales are never stationary, we are likely underestimating the true whale population size due to imperfect detection. To address this availability bias, hidden Markov model could be used (Borchers et al, 2013). HMM copes with a small data by fitting a more restrictive model, which incorporates information on how the plane moves. Violation of these assumptions bias our estimates of whale abundance – models will underestimate the real number



Half-Normal Observed vs. Expected Half-Normal Observed vs. Expected Hazard Rate Observed vs. Expected

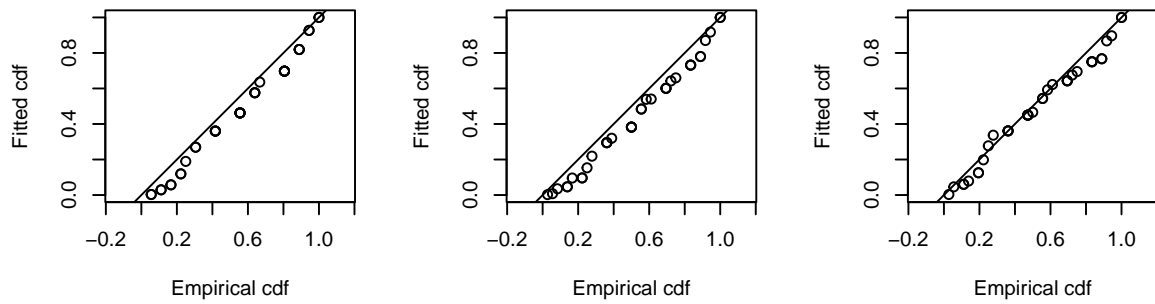


Figure 3: Detection functions using truncated data

of whales; therefore, truncation of data is needed. However, truncation reduces survey area by 30% and since data is quite unbalanced and sample size is small, it is not a great approximation of how whale abundance would look like in real life and any variation seen in estimates is due to stochasticity.