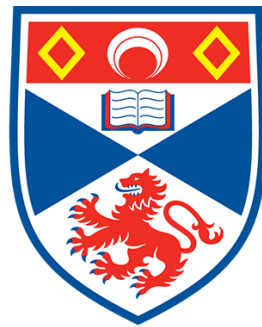# Panic in the Disko Bay: Estimating Greenland Bowhead Whale Detection and Abundance with Distance Sampling

Misha Tseitlin, Louise Blackman, Dometa Tuomaite

15 February 2024

Project 1: Distance Sampling

University of St Andrews

School of Mathematics and Statistics

Abstract

Bowhead whales (*Balaena mysticetus*) aggregate each spring to mate in west Greenland's Disko Bay (Qeqertarsuaq), providing a rare chance to measure this highly mobile species' abundance. Distance sampling (DS) uses likelihood estimation to generate estimates about abundance while accounting for imperfect detectability and has previously estimated bowhead whale numbers around Qeqertarsuaq. We analyse existing aerial survey data with simple DS methods to test the robustness of estimates without incorporating whale surfacing behaviour.[1] As data appears to violate DS assumptions, we compare different data processing approaches and their impact on the size effect and abundance estimates. Our simple DS, even after including all possible covariates, underestimates abundance relative to the original approach. Though DS remains a powerful tool to understand population sizes, we must approach cases of small sample sizes and simplified secondary analyses with caution.

Introduction

Annual springtime bowhead whale aggregations in Qeqertarsuaq allow for analysing both subpopulation stock rebound after whaling moratoria and estimating baseline population before wider climate change impacts in the Arctic. Using existing visual aerial surveys, we conduct a simplified replication of existing DS analysis using only covariates and adjustment terms to test method robustness and data suitability for more accessible methods for understanding bowhead whale detection, abundance, and impacts of group size.

Methods

Distance sampling estimates animal abundance using the perpendicular distance between the transect and observed animals. A detection function $g(x)$ is fitted to calculate the probability of detecting an animal at a distance x and account for imperfect observations. Based on rigid half-normal (HN) and hazard-rate (HR) key functions, we can incorporate covariates to model differences in detectability.[2] When $\sigma > 0$, we use a log link function. So,

$$\sigma(\mathbf{z}_i) = exp(\alpha_0 + \sum_{j=1}^{n} \alpha_j z_{ij})$$

where $\mathbf{z}_i = (z_{i1}, ..., z_{in})$ is the vector of covariates for animal i and $\alpha_0, ..., \alpha_n$ represent relevant covariate parameters. Another approach to flexibility, adjustment terms, can further improve fit.

After splitting Qeqertarsuaq into 16 strata and 41 systematically-placed east-west transects (4445km combined length), surveying used planes with two observation platforms to record declination angles and times of the first abeam whale sighting—generating perpendicular distances for DS—Beaufort sea state, and whale group size. Meaning resolved any differences in size or declination between observers. Authors left-truncated distances by 100 m due to obscured view close to the transect line, so we retained 58 observations of 74 whales.[1]

As-is detection distances demonstrate possible line avoidance, where more whales appear further from the plane than on the transect (Figure 1). DS assumptions—independent animal distribution with respect to the transect, and perfect detection on the transect line—may not hold if whales hid and/or moved.[2] However, a difference of only 4 observations between 0–0.2 and 0.4–0.6 km bins can fall within typical stochastic variation. Initially, we accepted this random variation and fit detection functions as-is. Though we avoid further data loss, this method may overestimate detection probabilities (thus underestimating abundance) without a perfect detectability baseline distance.

Given the risk of assumption violations, we considered two data adjustments. First, we grouped distances into progressively smaller bins with fewer observations in each subsequent grouping
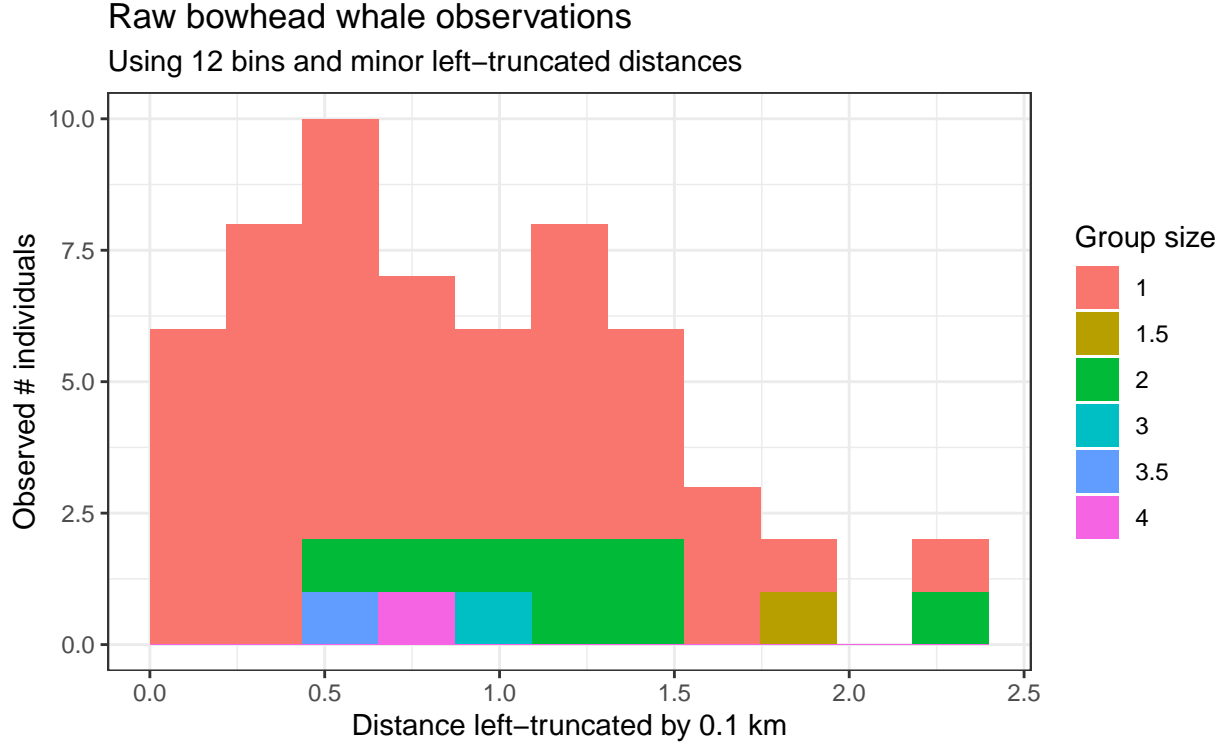
Figure 1: Original bowhead whale detection distances

Table 1: Estimated abundance from best-performing models

Key Function	Dataset	Size Incl?	Abundance	SE	Lower CI	Upper CI	AIC
Half Normal	Base	Yes	212.07	65.62	112.13	401.06	86.03
Half Normal	Base	No	228.82	71.45	122.00	429.14	86.32
Half Normal	Binned	Yes	212.84	66.14	112.32	403.31	197.58
Half Normal	Binned	No	229.54	72.22	121.94	432.07	197.91
Hazard Rate	Truncated	Yes	1803.36	1346.22	461.56	7045.88	27.20
Half Normal	Truncated	No	519.51	90.74	363.25	743.00	28.43
Half Normal	Truncated	Yes	608.75	154.26	356.93	1038.20	28.73

(Figure 2). This presumes that whales moved away from transect lines and were instead seen further away. Though accounting for potential assumption violations, our grouping copiously discards information.

Our second option left-truncated distances until peak detectability at 0.74 km. Thus, we ensure a strict decline in detectability with increasing distance in exchange for discarding 20 more observations (25.7% of our total). Now accounting for potential hiding behaviours (e.g., diving), we still must assume perfect detectability at our new baseline distance—0.74 km. Interestingly, this data better fit complex models with multiple covariates and adjustment terms at the risk of over-fitting and poor generalisability.

We used the `Distance` package in `R` to fit models using a maximum likelihood approach.[3] For each data set (unedited, binned, truncated) we fitted HN and HR detection functions with all combinations of adjustment terms (i.e., cosine, Hermite, polynomial) and covariates. Aikake Information Criterion (AIC) compared between models using the same manipulated data.
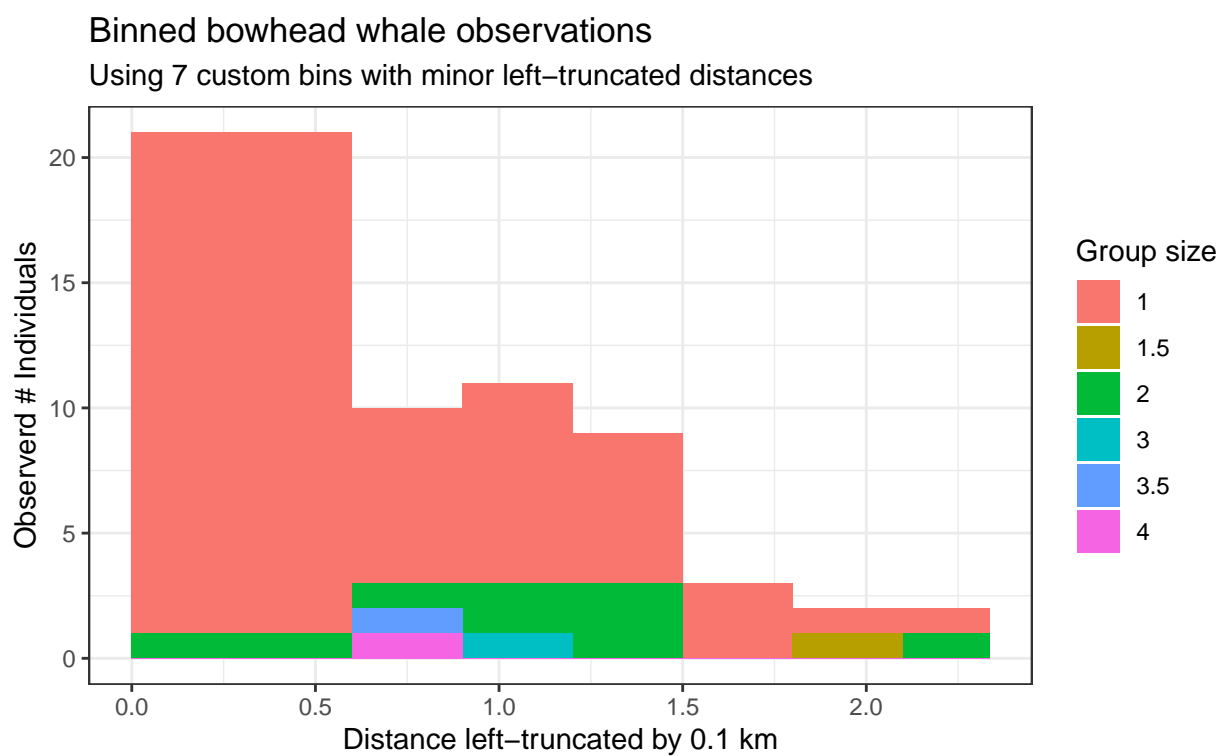
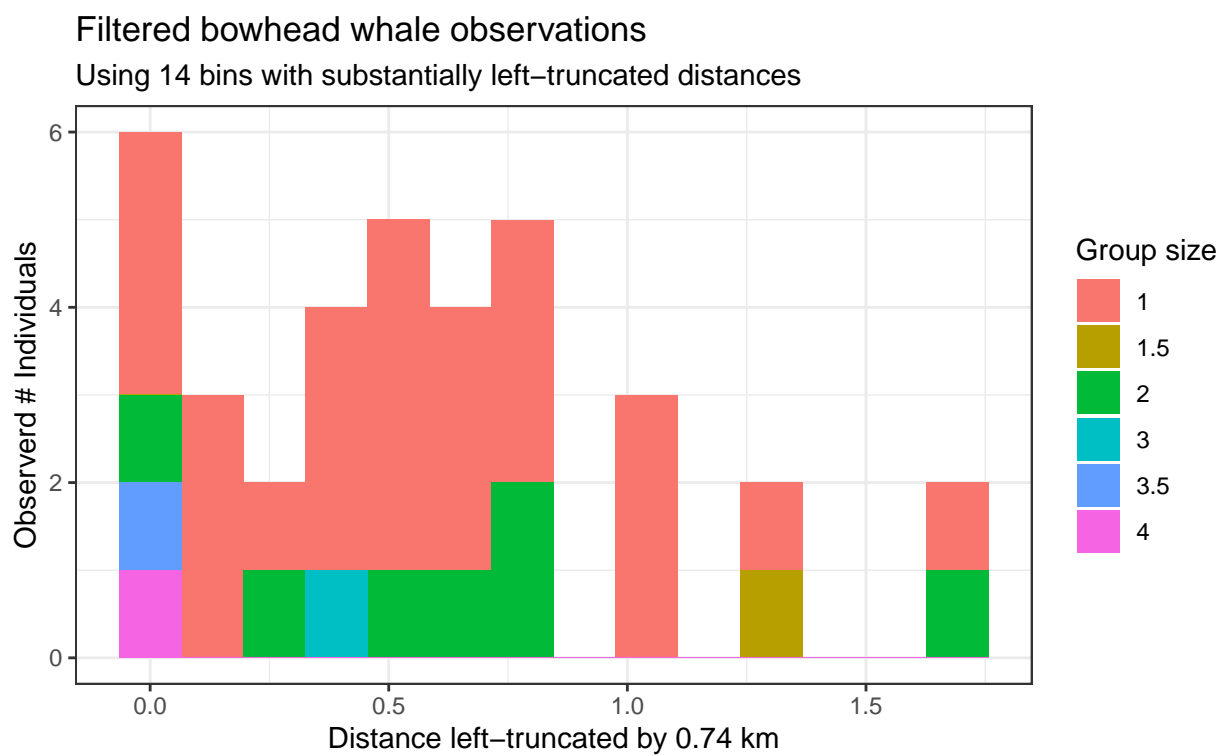Figure 2: Original bowhead whale detection distances with custom bins

Figure 3: More truncated bowhead whale detection distances

Table 2: Size effects on detectability

Model	Size Effect Magnitude	Standard Error
Base HN + size	6.16	5046.40
Binned HN + size	5.64	305.31
Truncated HN + size	-0.51	0.26

Results

For the base dataset, AIC preferred a HN detection function with a size covariate and no adjustments which estimated abundance as 212 individuals, only slightly lower than the non-covariate model after accounting for stratified survey design. Binned data produced near-identical results, with a preferred HN function including size. In contrast, truncated data produced substantially higher estimates both for HR and HN models; despite large standard errors, these better matched primary analysis estimates (Table 1). The effect of size also remains unclear—original data suggests larger group sizes improve detectability while truncated data implies the converse (Table 2). Because truncated modelling with HN detection suggests worse fit including size, we side with improved detectability after accounting for size bias: higher detectability for larger groups implies artificially-inflated abundance stemming from overconfidence in detectability for single individuals.

Fitted detection functions confirm this; original data models with size assume large groups are perfectly detected while individual detectability is overestimated in base models (Figure 4). As most data was observed in strata 2 which, along with strata 9, were the only areas to exhibit varied group size, we cannot comment too much given extreme heterogeneity.[1] Binning data results spawned near-identical results to this original analysis with less data thus without robust goodness-of-fit testing (Figure 5).

In contrast, truncated HN detection functions suggest more extreme detectability underestimates with less-extreme size bias. However, the preferred HR function instead displays a highly narrow shoulder that implies a violation of perfect detectability: the actual fit is driven by just eight observations (Figure 6). For cross-data comparability, we thus settle on the suitability of half-normal detection.

Discussion

Despite multiple adjustments, we remain concerned about perfect detectability and other DS assumptions. Here, alternatives like passive monitoring (e.g., using cameras or hydrophones) could improve future surveys. As whales also continually move—behaviourally demonstrating observer avoidance—animals are certainly undetected at their initial locations. In tandem, we underestimate true whale abundance due to imperfect detection. Whales are often underwater, likely generating availability bias. To address this, the original analysis uses Hidden Markov Models (HMM) to fit a more restrictive model incorporating information on aircraft movement and whale surfacing.[4] Though multi-covariate distance sampling does provide benefits over conventional distance sampling given data heterogeneity and small sample size, we require either larger samples or more relevant covariates to improve estimates.[5]
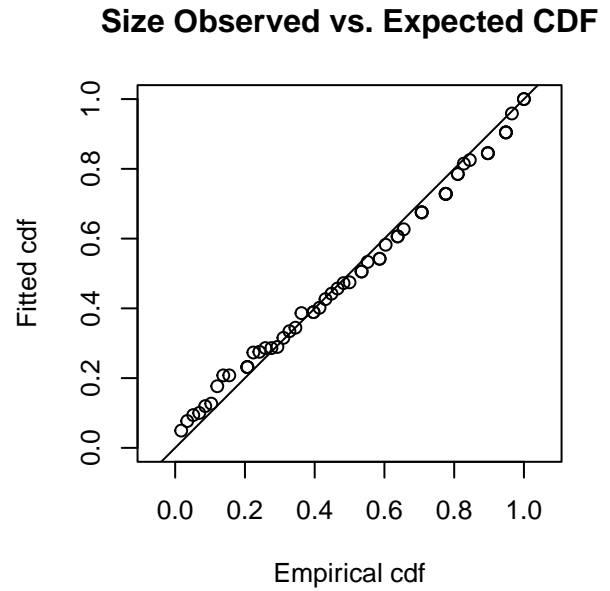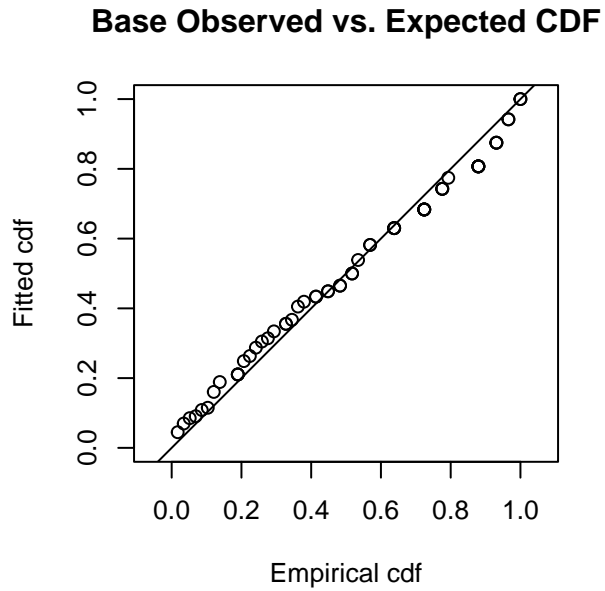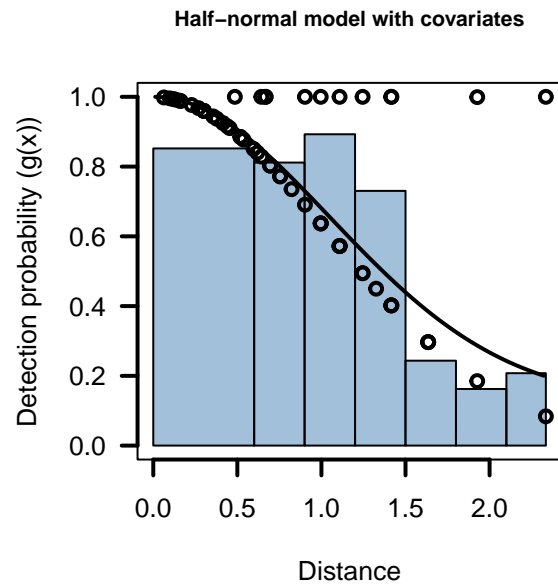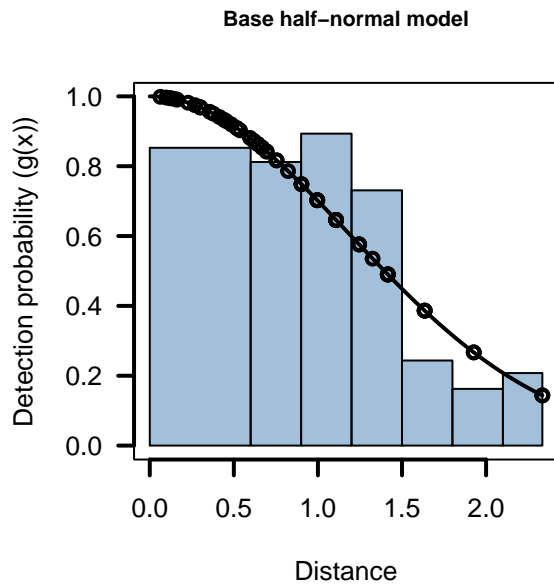
Figure 4: Detection functions using original data

Base half−normal model

Half−normal model with covariates
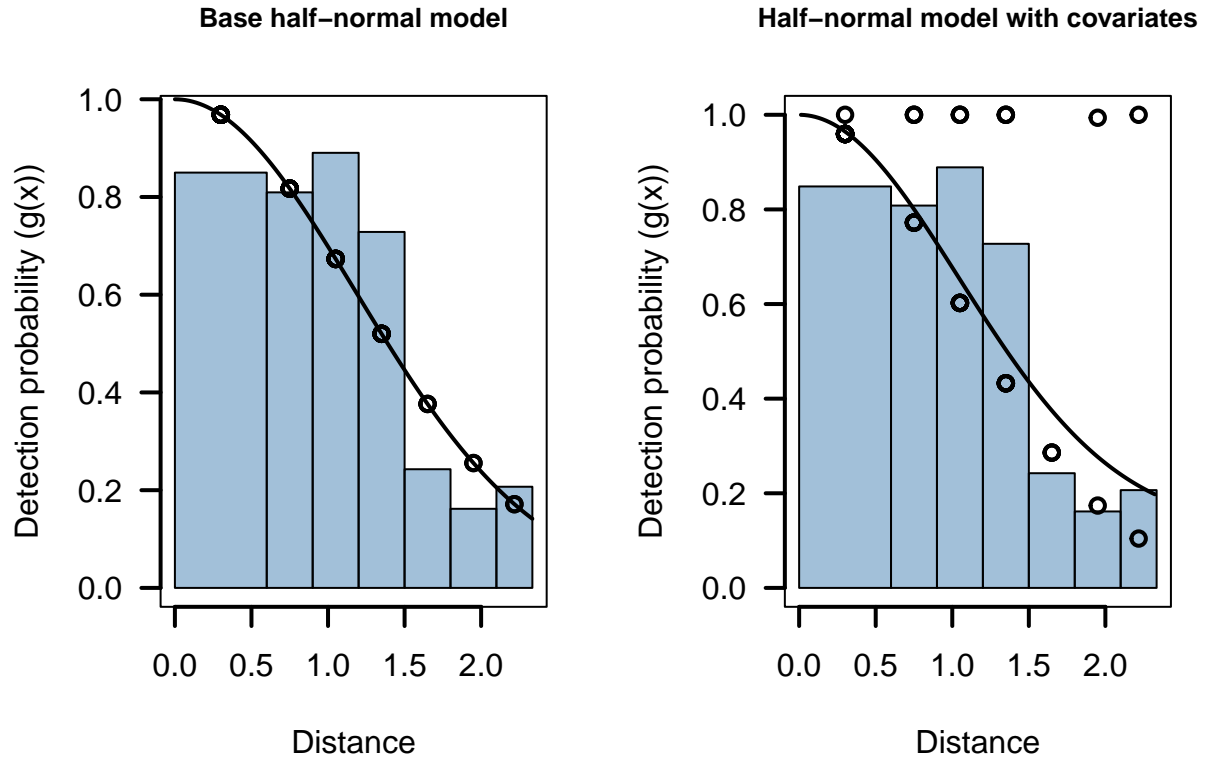
Figure 5: Detection functions using binned data
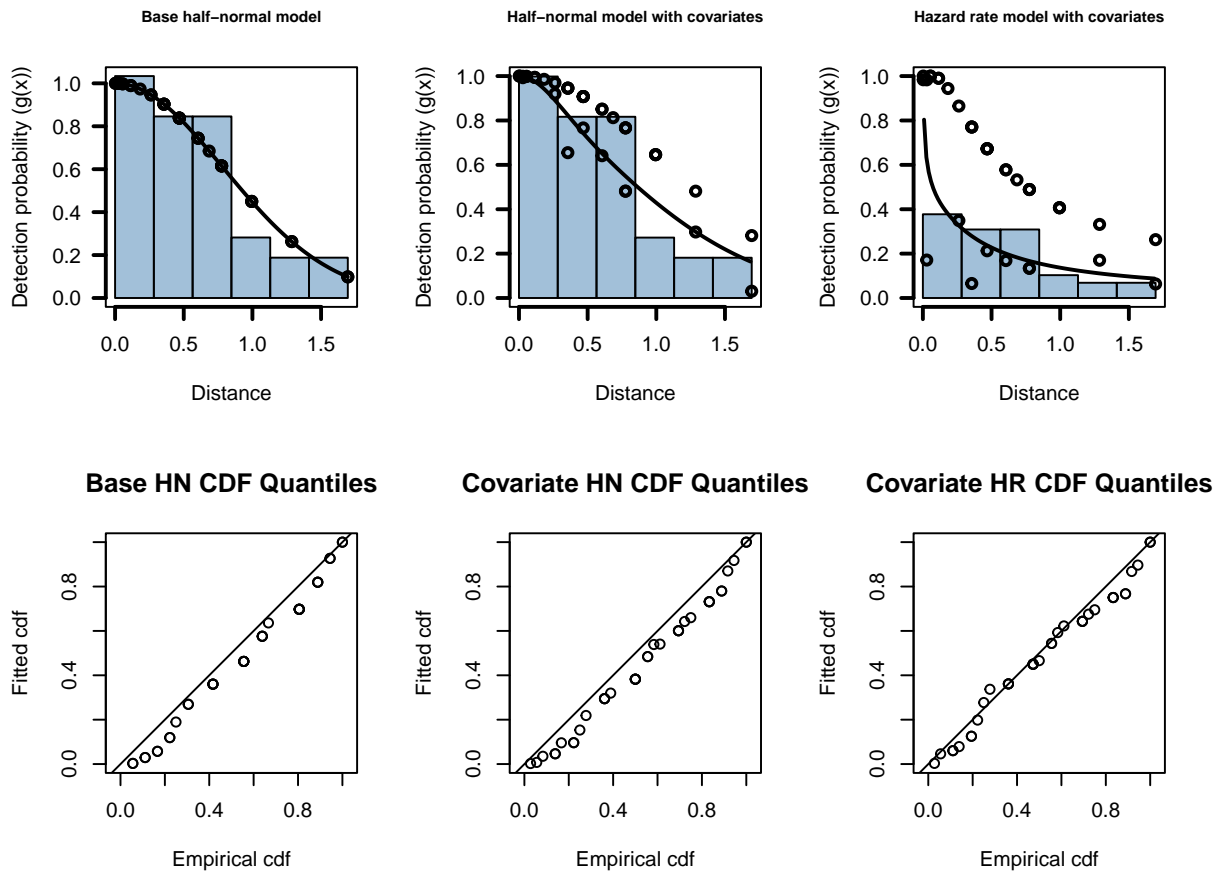
Figure 6: Detection functions using truncated data

Authors' Contributions

MT, LB, and DT all separately conducted analysis in R and collectively decided to use MT's (edited) models for the final report. For initial drafting, MT wrote the abstract; LB wrote up the methods; and DT wrote results and discussions sections. Final edits for submission are unattributable and were performed by all group members. For more detail on individual contributions, please reference GitHub.

Code Appendix

```r
# import packages and data
library(statsecol)
library(tidyverse)
library(viridisLite)
library(ggridges)
library(Distance)
data(bowhead_LT)

# raw data plot
# looks like hazard rate would be the best fit here
ggplot(data = bowhead_LT) +
  geom_histogram(aes(x = distance, fill = as.factor(size)), bins= 21) +
  labs(x = "Distance truncated by 0.1 km",
       y = "Observerd # individuals",
       title = "Raw bowhead whale observations",
       subtitle = "Using 21 bins and minor left-truncated distances",
       fill = "Group size") +
  theme_bw()

# aggressive binning
# definitely half-normal is the better fit here
histBreaks2 <- c(0, 0.6, 0.9, 1.2, 1.5, 1.8, 2.1, max(bowhead_LT$distance,
                                                      na.rm = TRUE))
ggplot(data = bowhead_LT) +
  geom_histogram(aes(x = distance, fill = as.factor(size)),
                 breaks = histBreaks2) +
  labs(x = "Distance truncated by 0.1 km",
       y = "Observerd # Individuals",
       title = "Binned bowhead whale observations",
       subtitle = "Using 7 custom bins with minor left-truncated distances",
       fill = "Group size") +
  theme_bw()

#truncate to get decreasing observations with distance
bowhead_LT_Trunc = bowhead_LT %>% filter(distance >= 0.64) %>%
  mutate(distance = distance - 0.64)

# plot the new data
# both half-normal and hazard rate appear possible here
ggplot(data = bowhead_LT_Trunc) +
  geom_histogram(aes(x = distance, fill = as.factor(size)), bins = 14) +
  labs(x = "Distance left-truncated by 0.74 km",
       y = "Observerd # Individuals",
       title = "Filtered bowhead whale observations",
       subtitle = "Using 14 bins with substantially left-truncated distances",
       fill = "Group size") +
  theme_bw()

conversion.factor <- convert_units(distance_units = "kilometre",
```

```r
                                        effort_units = "kilometre",
                                        area_units = "square kilometre")

# using model fits from separate alternatives,
# we'll only fit no-adjustment HN, HR, and size covariates
# ds() will automatically pick the low AIC version (adjustment = NULL)
baseRawHN <- ds(data = bowhead_LT, key = "hn",
                convert_units = conversion.factor)
baseRawHR <- ds(data = bowhead_LT, key = "hr",
                convert_units = conversion.factor)
# sometimes does not converge
sizeRawHN <- ds(data = bowhead_LT, key = "hn", adjustment = NULL,
                convert_units = conversion.factor,
                formula = ~size)
sizeRawHR <- ds(data = bowhead_LT, key = "hr", adjustment = NULL,
                convert_units = conversion.factor,
                formula = ~size)

# size + half-normal is the best fit on the raw data
summarize_ds_models(baseRawHN,
                    baseRawHR,
                    sizeRawHN,
                    sizeRawHR, output = "plain")

par(mfrow = c(2,2))
plot(baseRawHN, which=2, pl.col = adjustcolor("steelblue",0.5), border=NULL,
     lwd = 2,
     breaks = histBreaks2,
     ylab = "Detection probability (g(x))", xlab = "Distance", las=1,
     main = "Base half-normal model")
# covariate model assumes perfect detection of large-sized groups
# this is a quirk of the limited data
# only Region 2 has any within-group variation in size
plot(sizeRawHN, which=2, pl.col = adjustcolor("steelblue",0.5),border=NULL,
     lwd = 2,
     breaks = histBreaks2,
     ylab = "Detection probability (g(x))", xlab = "Distance", las=1,
     main = "Half-normal model with covariates")
# both models have solid GOF
# covariate model is slightly better from visual QQ inspection
gof_ds(baseRawHN, main="Base Observed vs. Expected CDF", ks = TRUE)
gof_ds(sizeRawHN, main="Covariate Observed vs. Expected CDF", ks = TRUE)

# repeat for binned data
baseBinHN <- ds(data = bowhead_LT, key = "hn",
                cutpoints = histBreaks2,
                convert_units = conversion.factor)
baseBinHR <- ds(data = bowhead_LT, key = "hr",
                cutpoints = histBreaks2,
                convert_units = conversion.factor)
# often does not converge
```

```r
sizeBinHN <- ds(data = bowhead_LT, key = "hn", adjustment = NULL,
                cutpoints = histBreaks2,
                convert_units = conversion.factor,
                formula = ~size)
sizeBinHR <- ds(data = bowhead_LT, key = "hr", adjustment = NULL,
                cutpoints = histBreaks2,
                convert_units = conversion.factor,
                formula = ~size)


#binned data also prefers models in the same order as unbinned
summarize_ds_models(baseBinHN,
                    baseBinHR,
                    sizeBinHN,
                    sizeBinHR, output = "plain")


# binned models have a very similar fit (if not identical) to the raw data
par(mfrow = c(1,2))
plot(baseBinHN, which=2, pl.col = adjustcolor("steelblue",0.5),border=NULL,
     lwd = 2,
     ylab = "Detection probability (g(x))", xlab = "Distance", las=1,
     main = "Base half-normal model")
# covariate model still assumes perfect detection of large-sized groups
# still a quirk of the data due to between-strata variation
plot(sizeBinHN, which=2, pl.col = adjustcolor("steelblue",0.5),border=NULL,
     lwd = 2,
     ylab = "Detection probability (g(x))", xlab = "Distance", las=1,
     main = "Half-normal model with covariates")
# GOF produce output but cannot generate plots
# bootstrapping impossible so KS test doesn't happen
gof_ds(baseBinHN, main="Base Observed vs. Expected CDF", ks = TRUE)
gof_ds(sizeBinHN, main="Covariate Observed vs. Expected CDF", ks = TRUE)


# finally repeat for truncated data
# truncated data prefers complex fits w/ more covariates and adjustments
baseTruncHN <- ds(data = bowhead_LT_Trunc, key = "hn",
                convert_units = conversion.factor)
baseTruncHR <- ds(data = bowhead_LT_Trunc, key = "hr",
                convert_units = conversion.factor)
sizeTruncHN <- ds(data = bowhead_LT_Trunc, key = "hn", adjustment = NULL,
                convert_units = conversion.factor,
                formula = ~size)
sizeTruncHR <- ds(data = bowhead_LT_Trunc, key = "hr", adjustment = NULL,
                convert_units = conversion.factor,
                formula = ~size)


# truncated data shows very different rankings
# hazard rate is preferred w/ size covariate
# then basic hanf-normal, then size + HN, finally base HR
summarize_ds_models(baseTruncHN,
                    baseTruncHR,
                    sizeTruncHN,
```

```r
                        sizeTruncHR, output = "plain")

# looks like the binned models have exactly the same fit as the raw data
par(mfrow = c(2,3))
plot(baseTruncHN, which=2, pl.col = adjustcolor("steelblue",0.5),border=NULL,
     lwd = 2,
     ylab = "Detection probability (g(x))", xlab = "Distance", las=1,
     main = "Base half-normal model")
# covariate model no longer assumes perfect detection of large-sized groups
plot(sizeTruncHN, which=2, pl.col = adjustcolor("steelblue",0.5),border=NULL,
     lwd = 2,
     ylab = "Detection probability (g(x))", xlab = "Distance", las=1,
     main = "Half-normal model with covariates")
plot(sizeTruncHR, which=2, pl.col = adjustcolor("steelblue",0.5),border=NULL,
     lwd = 2,
     ylab = "Detection probability (g(x))", xlab = "Distance", las=1,
     main = "Hazard rate model with covariates")
# covariate model still slightly better from visual QQ plot
gof_ds(baseTruncHN, main="Base Half-Normal Observed vs. Expected CDF",
       ks = TRUE)
gof_ds(sizeTruncHN, main="Covariate Half-Normal Observed vs. Expected CDF",
       ks = TRUE)
gof_ds(sizeTruncHR, main="Covariate Hazard Rate Observed vs. Expected CDF",
       ks = TRUE)


# output the models for consistency and processing speed
save(baseRawHN, baseRawHR, sizeRawHN, sizeRawHR,
     baseBinHN, baseBinHR, sizeBinHN, sizeBinHR,
     baseTruncHN, baseTruncHR, sizeTruncHN, sizeTruncHR,
     file="df-models.RData")


# break out abundance under both models by region
# generate abundance estimates for the best raw data models
region_table <- unique(bowhead_LT[,c("Region.Label", "Area")])
sample_table <- unique(bowhead_LT[,c("Region.Label",
                                     "Sample.Label",
                                     "Effort")])
observation_table <- unique(bowhead_LT[,c("object",
                                          "Region.Label",
                                          "Sample.Label")])

# raw model
baseRawHN_N <- dht(model = baseRawHN$ddf,
                   region.table = region_table,
                   sample.table = sample_table,
                   obs.table = observation_table)
# covariate model
# lower overall SE driven by Regions 2 and 9
# Regions 3, 11, 12, and 15 have increased SE
# abundance decreases in regions 2 and 9 while increasing elsewhere
# Regions 2 and 9 are the only sub-regions with size > 1 observations
sizeRawHN_N <- dht(model = sizeRawHN$ddf,
```

```r
                      region.table = region_table,
                      sample.table = sample_table,
                      obs.table = observation_table)
# repeat abundance estimation process for remaining models of interest
# base binned model
baseBinHN_N <- dht(model = baseBinHN$ddf,
                   region.table = region_table,
                   sample.table = sample_table,
                   obs.table = observation_table)
# size covariate binned data model
sizeBinHN_N <- dht(model = sizeBinHN$ddf,
                   region.table = region_table,
                   sample.table = sample_table,
                   obs.table = observation_table)
# base truncated data model
baseTruncHN_N <- dht(model = baseTruncHN$ddf,
                     region.table = region_table,
                     sample.table = sample_table,
                     obs.table = observation_table)
# size covariate truncated data model
sizeTruncHN_N <- dht(model = sizeTruncHN$ddf,
                     region.table = region_table,
                     sample.table = sample_table,
                     obs.table = observation_table)


# check variance using bootstraps rather than delta method
# delta-method approximation assumes independence between
# uncertainty in the detection function and variability in encounter rate

# takes ~3 minutes on my computer, likely longer for most
# check or delete the "cores = " if you don't have 10 cores on your computer
est.boot <- bootdht(model=sizeRawHN, flatfile=bowhead_LT,
                    summary_fun=bootdht_Nhat_summarize,
                    convert_units=conversion.factor, nboot=999, cores=10)
alpha <- 0.05
# the lower bound of our estimate is very poorly constrained
# 95% CI between 0 and 276
bootci <- quantile(est.boot$Nhat, probs = c(alpha/2, 1-alpha/2),
                   na.rm=TRUE)

# plot generated bootstraps
par(mfrow = c(1,1))
hist(est.boot$Nhat, nc=30,
     main="Distribution of bootstrap estimates\nwithout model uncertainty",
     xlab="Estimated abundance")
abline(v=bootci, lwd=2, lty=2)

# repeat boostraps w/ resample by region rather than by transect
# our upper CI increases slightly but the lower CI stays at 0
#est.bootStrata <- bootdht(model=bowhead.hn.null.size, flatfile=bowhead_LT,
#                  summary_fun=bootdht_Nhat_summarize,
```

```
#                    resample_strata = TRUE,
#                    convert_units=conversion.factor, nboot=999, cores=10)
#bootciStrata <- quantile(est.bootStrata$Nhat, probs = c(alpha/2, 1-alpha/2),
#                    na.rm=TRUE)


# this following section is for transparency on full model fitting
# attempted with the original, minorly truncated data
# play around with some adjustment terms with the size model
# adjustment terms generally aren't included in covariates for whatever reason
# Cosine(2) is preferred but never monotonic so that's not it
#bowhead.hn.cos.size <- ds(data = bowhead_LT, key = "hn", adjustment = "cos",
#                    convert_units = conversion.factor,
#                    formula = ~size,
#                    nadj = c(1,2,3,4))
# Hermite(4) is preferred but is non-monotonic; Hermite(4,6) is not as bad
#bowhead.hn.herm.size <- ds(data = bowhead_LT, key = "hn", adjustment = "herm",
#                    convert_units = conversion.factor,
#                    formula = ~size,
#                    nadj = c(2,3,4))
# Poly(4,6) ends up being non-monotonic, so leaving it at Polynomial(4)
#bowhead.hn.poly.size <- ds(data = bowhead_LT, key = "hn", adjustment = "poly",
#                    convert_units = conversion.factor,
#                    formula = ~size,
#                    nadj = c(1,2,3,4))
# now repeat for half-normal
# non-monotonic and g(x) > 1 with a cosine adjustment
#bowhead.hr.cos.size <- ds(data = bowhead_LT, key = "hr", adjustment = "cos",
#                    convert_units = conversion.factor,
#                    formula = ~size,
#                    nadj = c(1,2,3,4))
#Hermite(4,6) surprisingly matches everything
#bowhead.hr.herm.size <- ds(data = bowhead_LT, key = "hr", adjustment = "herm",
#                    convert_units = conversion.factor,
#                    formula = ~size,
#                    nadj = c(2,3,4))
#Polynomial adjustments don't match monotonicity
#bowhead.hr.poly.size <- ds(data = bowhead_LT, key = "hr", adjustment = "poly",
#                    convert_units = conversion.factor,
#                    formula = ~size,
#                    nadj = c(1,2,3,4))
```

References

1. Rekdal, S. L. *et al.* Trends in bowhead whales in west greenland: Aerial surveysvs. Genetic capture-recapture analyses. *Marine Mammal Science* **31**, 133–154 (2014).

2. Buckland, S. T., Rexstad, E. A., Marques, T. A. & Oedekoven, C. S. *Distance Sampling: Methods and Applications.* (Springer International Publish, 2015).

3. Miller, D. L., Rexstad, E., Thomas, L., Marshall, L. & Laake, J. L. Distance sampling in R. *Journal of Statistical Software* **89**, 1–28 (2019).

4. Borchers, D. L., Zucchini, W., Heide-Jørgensen, M. P., Cañadas, A. & Langrock, R. Using hidden markov models to deal with availability bias on line transect surveys. *Biometrics* **69**, 703–713 (2013).

5. Marques, T. A., Thomas, L., Fancy, S. G. & Buckland, S. T. Improving Estimates of Bird Density Using Multiple- Covariate Distance Sampling. *The Auk* **124**, 1229–1243 (2007).