**SUPPORTING INFORMATION**

This document provides supporting information on the formulation of the detection function.

**Appendix S1: Detection Function**

For our detection function we used information from the double platform survey method. Information collected from this survey design allowed us to apply mark-recapture distance sampling (mrds) methods to estimate detection probability in each grid cell (Laake and Borchers 2004). To model the sightings data from the dual observers we adopt the formulation for point independence outlined by Laake and Borchers (2004). This estimator combines a mark-recapture analysis with conventional distance sampling (CDS) to estimate detection probability such that *g(0)* can be estimated directly, and therefore, is not assumed to be 1. The estimator is

Where represents the estimate of detection probability on the trackline and is estimated from the mark-recapture data, represents the detection function and is estimated from the distance data, y is a vector of observed distances and is a matrix of detection covariates that influence perception bias. By assuming that the two observer platforms are independent at *y*=0 this method allows for correlation among observers at distances greater than 0.

*DS Component*

To model the likelihood for the distance data we implementedboth half-normal and hazard rate detection functions. The likelihood for the half normal detection function can be written as

Where σ represents the scale parameter. The likelihood for the hazard rate detection function is

where α represents the scale parameter and *b* represent the shape parameter.

The probability density function (pdf) is

where *W* represents the right truncation distance. The integral in the equation is referred to as the effective strip width (ESW). Under the assumptions of CDS (i.e. *g(0)* =1), when the pdf is evaluated at distance y=0

We model the variance of the half normal detection function, , as a log-linear combination of detection covariates

where *B0* is an intercept term,  represent the k linear coefficients of the detection function and *Zk*  represents a matrix of k detection covariates (e.g. beaufort sea state, glare, group size, etc.)

*MR Component*

The DS component described above assumes detection on the trackline equals 1. To estimate g(0) directly we can use the MR data collected from the double observer platform. To do so, we model the binary outcome of whether or not observer *m* successfully detects an animal cluster that was present at distance *y* as the outcome of a Bernoulli trial. To model each p we use a logit link function such that

 Logit(pm) = αm + αD\*y+ Σ αk \*Z.Detk

where αm represents the probability of detection on the trackline for observer *m*, αD represents the change in detection probability with distance *y*, and the sum terms represents the linear effect of all other detection covariates (e.g. beaufort sea state, visibility etc.).

Each sighting can be assigned to one of three possible capture histories ()that are conditional on a group being sighted. The group can be seen by the primary observer and not by the secondary observer (=1, 0), not seen by the primary observer but seen by the secondary observer (=0,1) or seen by both observers (=1,1). Because the probability of detection is conditional on a group being sighted, we write the conditional likelihood of detection given that an animal is sighted as

where

P(ω = (1,0) | detected) = p1 \* (1-p2)

 P(ω = (0,1) | detected) = (1-p1) \* p2

 P(ω = (1,1) | detected) = p1 \* p2

We model each one of these capture histories as the outcome of a Bernoulli trial such that

 ω*h* ~Bern(

where *c* indicates the capture history (i.e. 10, 01 or 11).

For a given set of detection covariates, the estimator for detectability on the trackline (y=0) is computed as

= p1(0,***Z***)+p2(0,***Z***)- p1(0,***Z***)\* p2(0,***Z***)

where

pm(0,***Z***)=invlogit(αm + Σ αk \* *Z* k)

Our final correction for detection probability can be written as:

*Aerial surveys*

For the aerial surveys, the secondary team consisted of a single observer at the back of the plane with an obstructed view complicating a direct implementation of the mrds approach. To estimate *g(0)* for aerial surveys we used the observation of these secondary observers in a trial configuration (sensu Burt et al. 2014) where we could use the observations of the back team as “trials” for the front team assuming full independence between the two teams (Burt et al. 2014). Because we only used sighting that were available to the back team and therefore only a limited distance from the trackline, the full independence assumption is less likely to be violated. We then used the estimate of *g(0,****Z****)*and its standard error to develop an informed beta prior using the same method as Pardo et al. (2015). We can now implement the mrds formulation to calculate detection probability where *g(0,****Z****)* is sampled from the informed prior and *g(y,****Z****)* is estimated using the observed distances recorded by the front team. More details on how *g(0,****Z****)* was estimated for the aerial surveys are provided in Palka et al.(2017).